

(19)



Евразийское
патентное
ведомство

(21) 202291730 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2022.08.10

(51) Int. Cl. *A61K 31/7088* (2006.01)
C12N 15/11 (2006.01)
C12N 9/22 (2006.01)

(22) Дата подачи заявки
2020.10.09

(54) КОМПОЗИЦИИ И СПОСОБЫ ДЛЯ IN VIVO СИНТЕЗА ПОЛИПЕПТИДОВ НЕ ПРИРОДНОГО ПРОИСХОЖДЕНИЯ

(31) 62/913,664; 62/988,882

(32) 2019.10.10; 2020.03.12

(33) US

(86) PCT/US2020/054947

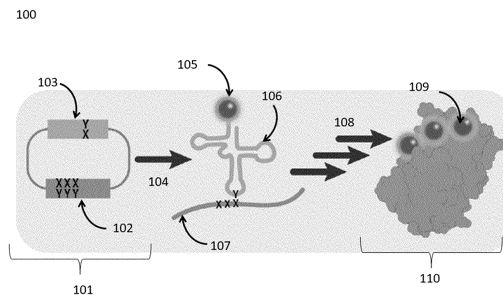
(87) WO 2021/072167 2021.04.15

(71) Заявитель:
ДЗЕ СКРИППС РИСЕРЧ
ИНСТИТЮТ (US)

(72) Изобретатель:
Ромесберг Флойд И., Фишер Эмил К.,
Хасимото Кодзи, Фельдман Аарон У.,
Дин Вивиан Т., Чжан Йорк (US)

(74) Представитель:
Медведев В.Н. (RU)

(57) В изобретении описаны композиции, способы и наборы для включения в клетку аминокислот не природного происхождения в полипептид не природного происхождения. В изобретении также описаны композиции, способы и наборы для повышения активности и выхода полипептида не природного происхождения, синтезируемого клеткой.



202291730

A1

A1

202291730

ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-573832EA/085

КОМПОЗИЦИИ И СПОСОБЫ ДЛЯ *IN VIVO* СИНТЕЗА ПОЛИПЕПТИДОВ НЕ ПРИРОДНОГО ПРОИСХОЖДЕНИЯ

ПЕРЕКРЕСТНАЯ ССЫЛКА НА РОДСТВЕННЫЕ ЗАЯВКИ

По настоящей заявке испрашивается приоритет предварительной заявки США № 62/913,664, поданной 10 октября 2019 г. и предварительной заявки США № 62/988,882, поданной 12 марта 2020 г., каждая из которых включена посредством ссылки в настоящий документ во всей полноте.

ПЕРЕЧЕНЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Настоящая заявка содержит Перечень последовательностей, который был представлен в электронном виде в формате ASCII и полностью включен в настоящий документ посредством ссылки. Указанная копия ASCII, созданная 6 октября 2010 г., называется 36271-809_601_SL.txt и имеет размер 21 килобайт.

ЗАЯВЛЕНИЕ О ФЕДЕРАЛЬНО СПОНСИРУЕМОМ ИССЛЕДОВАНИИ

Это изобретение было сделано при поддержке правительства по гранту № GM118178, выданному Национальными институтами здравоохранения. Правительство имеет определенные права на это изобретение.

ВКЛЮЧЕНИЕ ПОСРЕДСТВОМ ССЫЛКИ

Все публикации, патенты и патентные заявки, упомянутые в данном описании, включены в настоящее описание посредством ссылки в той же степени, как если бы каждая отдельная публикация, патент или патентная заявка были конкретно и отдельно указаны для включения посредством ссылки. В той степени, в которой публикации и патенты или патентные заявки, включенные посредством ссылки, противоречат описанию, содержащемуся в спецификации, спецификация предназначена для замены и/или приоритета над любым таким противоречащим материалом.

УРОВЕНЬ ТЕХНИКИ

Естественный генетический код состоит из 64 кодонов, обозначенных четырьмя буквами генетического алфавита. Три кодона используют в качестве стоп-кодонов, оставляя 61 смысловой кодон, которые распознаются транспортной РНК (тРНК), заряженной родственной аминокислотой тРНК-синтетазой (также называемой в настоящем документе тРНК-синтетаза) с одной из 20 протеогенных аминокислот. Хотя канонические аминокислоты сделали возможным удивительное разнообразие живых организмов, существует множество химических функций и связанных с ними реакционных способностей, которые они не обеспечивают. Способность расширять генетический код, включающий не природного происхождения или не канонические аминокислоты (ncAA), вероятно, придает белку желаемую функцию или активность и значительно облегчает многие известные и появляющиеся применения белков, такие как терапевтическая разработка. Современные способы синтеза белков не природного происхождения или не полипептидов природного происхождения, содержащих аминокислоты не природного

происхождения, имеют ограничения. Примечательно, что большинство способов позволяют вводить только отдельные аминокислоты не природного происхождения или нескольких копий аминокислоты не природного происхождения в полипептид не природного происхождения. Кроме того, полипептид не природного происхождения, синтезированный доступными в настоящее время способами, часто обладает пониженной ферментативной активностью, растворимостью или выходом.

Одним из альтернативных решений для преодоления этих ограничений является синтез полипептидов не природного происхождения с помощью бесклеточной системы экспрессии или системы экспрессии *in vitro*. Однако такая система экспрессии подходит для создания среды пост-трансляционной модификации, где окислительно-восстановительные свойства полипептида не природного происхождения и других пост-трансляционных модификаций синтезированного полипептида не природного происхождения полностью реализованы. Поэтому остается потребность в композициях и способах синтеза *in vivo* полипептидов не природного происхождения, содержащих аминокислоты не природного происхождения.

СУЩНОСТЬ ИЗОБРЕТЕНИЯ

В настоящем документе описаны композиции, способы, клетки (как не сконструированные, так и сконструированные), полусинтетические организмы (SSO), реагенты, генетический материал, плазмиды и наборы для синтеза *in vivo* полипептидов не природного происхождения или белков не природного происхождения, где каждый полипептид не природного происхождения или белок не природного происхождения включает две или несколько аминокислот не природного происхождения, которые декодируются клетками.

В настоящем документе описаны способы синтеза полипептида не природного происхождения *in vivo*, включающие: получение, по меньшей мере, одной молекулы дезоксирибонуклеиновой кислоты (ДНК) не природного происхождения, содержащей, по меньшей мере, четыре пары оснований не природного происхождения; транскрипцию, по меньшей мере, одной молекулы ДНК не природного происхождения с получением молекулы информационной рибонуклеиновой кислоты (иРНК), содержащей, по меньшей мере, два кодона не природного происхождения; транскрибирование, по меньшей мере, одной молекулы ДНК не природного происхождения с получением, по меньшей мере, двух молекул транспортной РНК (тРНК), каждая из которых содержит, по меньшей мере, один антикодон не природного происхождения, где, по меньшей мере, две пары оснований не природного происхождения в соответствующей ДНК находятся в контекстах последовательности таким образом, чтобы кодоны не природного происхождения молекул иРНК были комплементарны антикодону не природного происхождения каждой из молекул тРНК; и синтез полипептида не природного происхождения путем трансляции молекулы иРНК не природного происхождения с использованием, по меньшей мере, двух молекул тРНК не природного происхождения, где каждый антикодон не природного происхождения направляет сайт-специфическое включение аминокислоты не природного происхождения

происхождения в полипептид не природного происхождения. В некоторых вариантах осуществления, по меньшей мере, две пары оснований не природного происхождения включают пары оснований, выбранные из dCNMO-dTPT3, dNaM-dTPT3, dCNMO-dTAT1 или dNaM-dTAT1.

В некоторых вариантах осуществления, представлен способ синтеза полипептида не природного происхождения, включающий: предоставление, по меньшей мере, одной молекулы дезоксирибонуклеиновой кислоты (ДНК) не природного происхождения, содержащей, по меньшей мере, четыре пары оснований не природного происхождения, где, по меньшей мере, одна молекула ДНК не природного происхождения кодирует (i) молекулу информационной рибонуклеиновой кислоты (иРНК), содержащую, по меньшей мере, первый и второй кодоны не природного происхождения и (ii) по меньшей мере, первую и вторую молекулу транспортной РНК (тРНК), где первая молекула тРНК содержит первый антикодон не природного происхождения, и вторая молекула тРНК содержит второй антикодон не природного происхождения, и, по меньшей мере, четыре пары оснований не природного происхождения, по меньшей мере, в одной молекуле ДНК находятся в таких контекстах последовательности, что первый и второй кодоны не природного происхождения молекулы иРНК комплементарны первому и второму антикодонам не природного происхождения, соответственно; транскрипцию, по меньшей мере, одной молекулы ДНК не природного происхождения с получением иРНК; транскрипцию, по меньшей мере, одной молекулы ДНК не природного происхождения с получением, по меньшей мере, первую и вторую молекул тРНК; и синтез полипептида не природного происхождения путем трансляции молекулы иРНК не природного происхождения с использованием, по меньшей мере, первой и второй молекул тРНК не природного происхождения, где каждый из, по меньшей мере, первого и второго антикодонов не природного происхождения направляет сайт-специфическое включение аминокислоты не природного происхождения в полипептид не природного происхождения.

В некоторых вариантах осуществления, способы включают, по меньшей мере, два кодона не природного происхождения, каждый из которых содержит первый нуклеотид не природного происхождения, расположенный в первом положении, втором положении или третьем положении кодона, необязательно, где первый нуклеотид не природного происхождения расположен во втором положении или третьем положении кодона. В некоторых случаях, способы включают, по меньшей мере, два кодона не природного происхождения, каждый из которых содержит последовательность нуклеиновой кислоты NNX или NXN, и антикодон не природного происхождения, содержащий последовательность нуклеиновой кислоты XNN, YNN, NXN или NYN, для образования пары кодон-антикодон не природного происхождения, содержащей NNX-XNN, NNX-YNN или NXN-NYN, где N представляет собой любой природный нуклеотид, X представляет собой первый нуклеотид не природного происхождения, и Y представляет собой второй нуклеотид не природного происхождения, отличный от первого нуклеотида не природного происхождения, где X-Y образует пару нуклеотидов не природного происхождения (UBP)

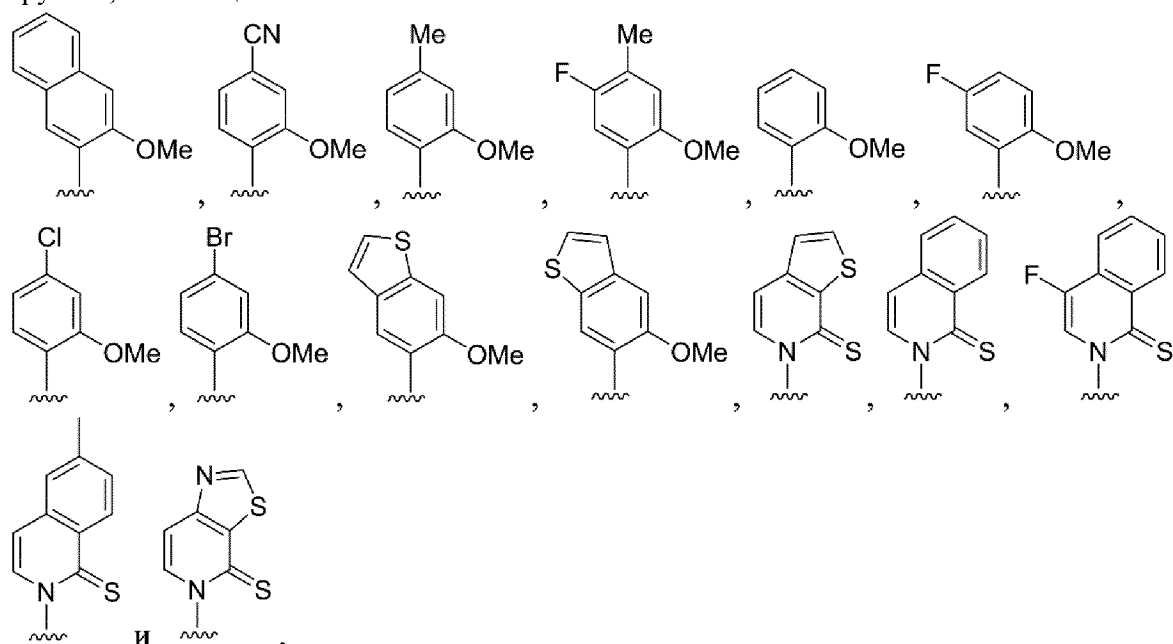
в ДНК.

В некоторых вариантах осуществления, UBP образуются между последовательностью кодона иРНК и последовательностью антикодона тРНК для облегчения трансляции иРНК в полипептид не природного происхождения. UBP кодон-антикодон содержит, в некоторых случаях, последовательность кодона, содержащую три непрерывных прочтения нуклеиновых кислот 5'-3' иРНК (например, UUX) и последовательность антикодона, содержащую три непрерывных прочтения нуклеиновых кислот 5'-3' от тРНК (например, YAA или XAA). В некоторых вариантах осуществления, когда кодон иРНК представляет собой UUX, антикодон тРНК представляет собой YAA или XAA. В некоторых вариантах осуществления, когда кодон иРНК представляет собой UGX, антикодон тРНК представляет собой YCA или XCA. В некоторых вариантах осуществления, когда кодон иРНК представляет собой CGX, антикодон тРНК представляет собой YCG или XCG. В некоторых вариантах осуществления, когда кодон иРНК представляет собой AGX, антикодон тРНК представляет собой YCU или XCU. В некоторых вариантах осуществления, когда кодон иРНК представляет собой GAX, антикодон тРНК представляет собой YUC или XUC. В некоторых вариантах осуществления, когда кодон иРНК представляет собой CAH, антикодон тРНК представляет собой YUG или XUG. В некоторых вариантах осуществления, когда кодоном иРНК является GXU, антикодон тРНК представляет собой AYC. В некоторых вариантах осуществления, когда кодон иРНК представляет собой CXU, антикодон тРНК представляет собой AYG. В некоторых вариантах осуществления, когда кодон иРНК представляет собой GXG, антикодон тРНК представляет собой CYC. В некоторых вариантах осуществления, когда кодон иРНК представляет собой AXG, антикодон тРНК представляет собой CYU. В некоторых вариантах осуществления, когда кодон иРНК представляет собой GXC, антикодон тРНК представляет собой GYC. В некоторых вариантах осуществления, когда кодон иРНК представляет собой AXC, антикодон тРНК представляет собой GYU. В некоторых вариантах осуществления, когда кодон иРНК представляет собой GXA, антикодон тРНК представляет собой UYC. В некоторых вариантах осуществления, когда кодон иРНК представляет собой CXC, антикодон тРНК представляет собой GYG. В некоторых вариантах осуществления, когда кодон иРНК представляет собой UXC, антикодон тРНК представляет собой GYA. В некоторых вариантах осуществления, когда кодон иРНК представляет собой AUX, антикодон тРНК представляет собой YAU или XAU. В некоторых вариантах осуществления, когда кодон иРНК представляет собой CUX, антикодон тРНК представляет собой XAG или YAG. В некоторых вариантах осуществления, когда кодон иРНК представляет собой UUX, антикодон тРНК представляет собой XAA или YAA. В некоторых вариантах осуществления, когда кодон иРНК представляет собой GUX, антикодон тРНК представляет собой XAC или YAC. В некоторых вариантах осуществления, когда кодон иРНК представляет собой UAX, антикодон тРНК представляет собой XUA или YUA. В некоторых вариантах осуществления, когда кодон иРНК представляет собой GGX, антикодон тРНК представляет собой XCC или YCC.

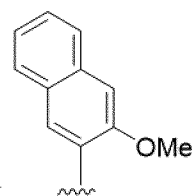
В некоторых вариантах осуществления, по меньшей мере, одна молекула ДНК не природного происхождения транскрибируется в информационную РНК (иРНК), содержащую основания не природного происхождения, описанные в настоящем документе (например, d5SICS, dNaM, dTPT3, dMTMO, dCNMO, dTAT1). Типовые кодоны иРНК кодируются типовыми областями ДНК не природного происхождения, содержащими три последовательных дезоксирибонуклеотида (NNN), включая TTX, TGX, CGX, AGX, GAX, CAX, GXT, CXT, GXG, AXG, GXC, AXC, GXA, CXC, TXC, ATX, CTX, TTX, GTX, TAX или GGX, где X представляет собой основание не природного происхождения, присоединенное к 2'-деоксирибозильной группе. Типовые кодоны иРНК, полученные в результате транскрипции типовой ДНК не природного происхождения, содержат три непрерывных рибонуклеотида (NNN), содержащих UUX, UGX, CGX, AGX, GAX, CAX, GXU, CXU, GXG, AXG, GXC, AXC, GXA, CXC, UXC, AUX, CUX, UUX, GUX, UAX или GGX, соответственно, где X представляет собой основание не природного происхождения, присоединенное к рибозильной группе. В некоторых вариантах осуществления, основание не природного происхождения находится в первом положении последовательности кодона (X-N-N). В некоторых вариантах осуществления, основание не природного происхождения находится во втором (или среднем) положении последовательности кодона (N-X-N). В некоторых вариантах осуществления, основание не природного происхождения находится в третьем (последнем) положении последовательности кодона (N-N-X).

В некоторых вариантах осуществления, способы включают кодон, содержащий, по меньшей мере, один G, и антикодон содержит, по меньшей мере, один C. В некоторых случаях, способы включают X и Y, где X и Y независимо выбраны из группы, состоящей из: (i) 2-тиоурацила, 2'-дезоксиуридина, 4-тиоурацила, урацил-5-ила, гипоксантин-9-ила (I), 5-галоурацила, 5-пропинилурацила, 6-азоурацила, 5-метиламинометилурацила, 5-метоксиаминометил-2-тиоурацила, псевдоурацила, метилового эфира урацил-5-оксуксусной кислоты, урацил-5-оксуксусной кислоты, 5-метил-2-тиоурацила, 3-(3-амино-3-N-2-карбоксыпропил)урацила, 5-метил-2-тиоурацила, 4-тиоурацила, 5-метилурацила, 5'-метоксикарбоксиметилурацила, 5-метоксиурацила, урацил-5-оксиуксусной кислоты, 5-(карбоксихидроксиметил)урацила, 5-карбоксиметиламинометил-2-тиоуридина, 5-карбоксиметиламинометилурацила или дигидроурацила; (ii) 5-гидроксиметилцитозина, 5-трифторметилцитозина, 5-галоцитозина, 5-пропинилцитозина, 5-гидроксицитозина, циклоцитозина, цитозина арабинозида, 5,6-дигидроцитозина, 5-нитроцитозина, 6-азоцитозина, азацитозина, N4-этилцитозина, 3-метилцитозина, 5-метилцитозина, 4-ацетилцитозина, 2-тиоцитозина, феноксазина цитидина ([5,4-b][1,4]бензоксазин-2(3H)-она), фенотиазина цитидин (1H-пиримидо[5,4-b][1,4]бензотиазин-2(3H)-она), феноксазина цитидина (9-(2-аминоэтокси)-H-пиримидо[5,4-b][1,4]бензоксазин-2(3H)-она), карбазола цитидина (2H-пиримидо[4,5-b]индол-2-она) или пиридоиндола цитидина (H-пиридо[3',2':4,5]пирроло[2,3-d]пиримидин-2-она); (iii) 2-амиoadенина, 2-пропиладенина, 2-амиoadенина, 2-F-аденина, 2-аминопропиладенина, 2-амино-2'-дезоксаденозина, 3-дезааденина, 7-метиладенина, 7-дезааденина, 8-азааденина, 8-галоген-, 8-амино-, 8-тиол-

, 8-тиоалкил- и 8-гидрокси-замещенных аденинов, N6-изопентениладенина, 2-метиладенина, 2,6-диаминопурина, 2-метилтио-N6-изопентениладенина или 6-азааденина; (iv) 2-метилгуанина, 2-пропила и алкильных производных гуанина, 3-дезагуанина, 6-тиогуанина, 7-метилгуанина, 7-дезагуанина, 7-дезагуанозина, 7-деаза-8-азагуанина, 8-азагуанина, 8-галоген-, 8-амино-, 8-тиол-, 8-тиоалкил- и 8-гидрокси-замещенных гуанинов, 1-метилгуанина, 2,2-диметилгуанина, 7-метилгуанина или 6-азагуанина и (v) гипоксантина, ксантина, 1-метилюридина, кеозина, бета-D-галактозилкеозина, инозина, бета-D-маннозилкеозина, вибутоксозина, гидроксимочевины, (аср3)w, 2-аминопиридина или 2-пиридона. В некоторых вариантах осуществления, каждый из X и Y независимо выбран из группы, состоящей из:

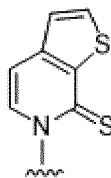


В некоторых случаях, X представляет собой



. В некоторых вариантах

осуществления, Y представляет собой



В некоторых вариантах осуществления способы, описанные в настоящем документе, включают пару кодон-антикодон не природного происхождения NNX-XNN, где NNX-XNN выбран из группы, состоящей из UUX-XAA, UGX-XCA, CGX-XCG, AGX-XCU, GAX-XUC, CAX-XUG, AUX-XAU, CUX-XAG, GUX-XAC, UAX-XUA и GGX-XCC. В некоторых вариантах осуществления способы, описанные в настоящем документе, включают пару кодон-антикодон не природного происхождения NNX-YNN, где NNX-YNN выбран из группы, состоящей из UUX-YAA, UGX-YCA, CGX-YCG, AGX-YCU, GAX-YUC, CAX-

YUG, AUX-YAU, CUX-YAG, GUX-YAC, UAX-YUA и GGX-YCC. В некоторых случаях способы, описанные в настоящем документе, включают пару кодон-антикодон не природного происхождения NXN-NYN, где NXN-NYN выбран из группы, состоящей из GXU-AYC, CXU-AYG, GXG-CYC, AXG-CYU, GXC-GYC, AXC-GYU, GXA-UYC, CXC-GYG и UXC-GYA. В некоторых вариантах осуществления способы, описанные в настоящем документе, включают, по меньшей мере, две молекулы тРНК не природного происхождения, каждая из которых содержит другой антикодон не природного происхождения. В некоторых случаях, по меньшей мере две молекулы тРНК не природного происхождения содержат пирролизил тРНК из рода *Methanosarcina* и тирозил тРНК из *Methanocaldococcus jannaschii*, или их производные. В некоторых вариантах осуществления, способы включают зарядку по меньшей мере, двух молекул тРНК не природного происхождения с помощью аминоацил-тРНК синтетазы. В некоторых случаях, тРНК синтетаза выбрана из группы, состоящей из химерных PylRS (chPylRS) и *M. jannaschii* AzFRS (*MjpAzFRS*). В некоторых вариантах осуществления, способы, описанные в настоящем документе, включают зарядку, по меньшей мере, двух молекул тРНК не природного происхождения, по меньшей мере, двумя различными тРНК синтетазами. В некоторых случаях, по меньшей мере, две различные тРНК синтетазы включают химерную PylRS (chPylRS) и *M. jannaschii* AzFRS (*MjpAzFRS*).

В некоторых вариантах осуществления, в настоящем документе описаны способы синтеза полипептидов не природного происхождения *in vivo*. В некоторых вариантах осуществления, полипептид не природного происхождения содержит две, три или несколько аминокислот не природного происхождения. В некоторых случаях, полипептид не природного происхождения содержит, по меньшей мере, две аминокислоты не природного происхождения, которые являются одинаковыми. В некоторых вариантах осуществления, полипептид не природного происхождения содержит, по меньшей мере, две различные аминокислоты не природного происхождения. В некоторых случаях, аминокислота не природного происхождения содержит:

аналог лизина; ароматическую боковую цепь; азидогруппу; алкиновую группу; альдегидную или кетонную группу. В некоторых случаях, аминокислота не природного происхождения не содержит ароматическую боковую цепь. В некоторых вариантах осуществления, одна или несколько аминокислот не природного происхождения выбрана из группы, состоящей из N6-((азидоэтокси)-карбонил)-L-лизина (AzK), N6-пропаргилэтокси-карбонил-L-лизина (PraK), N6-(пропаргилокси)карбонил-L-лизина (PrK), п-азидофенилаланина (pAzF), BCN-L-лизина, норборненилизина, TCO-лизина, метилтетразинлизина, аллилоксикарбониллизина, 2-амино-8-оксононановой кислоты, 2-амино-8-оксооктановой кислоты, п-ацетил-L-фенилаланина, п-азидометил-L-фенилаланина (pAMF), п-йодо-L-фенилаланина, м-ацетилфенилаланина, 2-амино-8-оксононаноевой кислоты, п-пропаргилоксифенилаланина, п-пропаргилфенилаланина, 3-метилфенилаланина, L-Дора, фторированного фенилаланина, изопропил-L-фенилаланина, п-азидо-L-фенилаланина, п-ацил-L-фенилаланина, п-бензоил-L-фенилаланина, п-

бромфенилаланина, п-амино-L-фенилаланина, изопропил-L-фенилаланина, О-аллилтирозина, О-метил-L-тирозина, О-4-аллил-L-тирозина, 4-пропил-L-тирозина, фосфонтирозина, три-О-ацетил-GlcNAc α -серина, L-фосфосерина, фосфоносерина, L-3-(2-нафтил)аланина, 2-амино-3-(((2-((3-(бензилокси)-3-оксопропил)амино)этил)селанил)пропановой кислоты, 2-амино-3-(фенилселанил)пропановой кислоты, селеноцистеина, N6-(((2-азидобензил)окси)карбонил)-L-лизина, N6-(((3-азидобензил)окси)карбонил)-L-лизина и N6-(((4-азидобензил)окси)карбонил)-L-лизина.

В некоторых вариантах осуществления, способы синтеза *in vivo* полипептидов не природного происхождения, как описано в настоящем документе, включают, по меньшей мере, одну молекулу ДНК не природного происхождения в форме плазмиды. В некоторых случаях, по меньшей мере, одну молекулу ДНК не природного происхождения интегрируют в геном клетки. В некоторых вариантах осуществления, по меньшей мере, одна молекула ДНК не природного происхождения кодирует полипептид не природного происхождения. В некоторых вариантах осуществления, описанные в настоящем документе способы включают репликацию и транскрипцию *in vivo* молекулы ДНК не природного происхождения, и трансляцию *in vivo* транскрибированной иРНК молекулы в клеточном организме. В некоторых вариантах осуществления, клеточный организм представляет собой микроорганизм. В некоторых вариантах осуществления, клеточный организм представляет собой прокариот. В некоторых вариантах осуществления, клеточный организм представляет собой бактерию. В некоторых случаях, клеточный организм представляет собой грамположительную бактерию. В некоторых вариантах осуществления, клеточный организм представляет собой грамотрицательную бактерию. В некоторых вариантах осуществления, клеточный организм представляет собой *Escherichia coli*. В некоторых вариантах осуществления, клеточный организм содержит транспортер нуклеозидтрифосфата. В некоторых случаях, транспортер нуклеозидтрифосфата содержит аминокислотную последовательность PtNTT2. В некоторых вариантах осуществления, транспортер нуклеозидтрифосфата содержит усеченную аминокислотную последовательность PtNTT2. В некоторых альтернативах, усеченная аминокислотная последовательность PtNTT2 на, по меньшей мере, 80% идентична PtNTT2, кодируемой SEQ ID NO. 1. В некоторых вариантах осуществления, клеточный организм содержит, по меньшей мере, одну молекулу ДНК не природного происхождения. В некоторых вариантах осуществления, по меньшей мере, одна молекула ДНК не природного происхождения содержит, по меньшей мере, одну плазмиду. В некоторых вариантах осуществления, по меньшей мере, одна молекула ДНК не природного происхождения интегрирована в геном клетки. В некоторых случаях, по меньшей мере, одна молекула ДНК не природного происхождения кодирует полипептид не природного происхождения. В некоторых случаях, способы, описанные в настоящем описании, могут быть способом *in vitro*, включающим синтез полипептида не природного происхождения с бесклеточной системой.

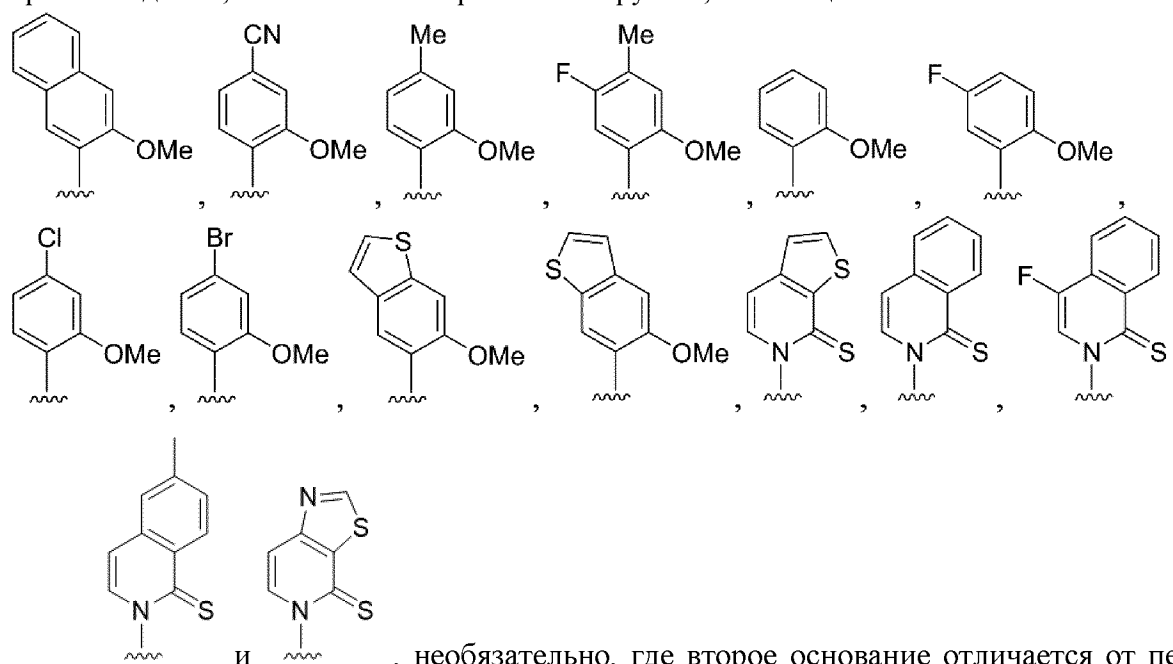
В некоторых вариантах осуществления, в настоящем документе описаны способы

синтеза *in vivo* полипептидов не природного происхождения, где полипептиды не природного происхождения содержат сахарную группу не природного происхождения. В некоторых вариантах осуществления, пары оснований не природного происхождения содержат, по меньшей мере, один нуклеотид не природного происхождения, содержащий сахарную группу не природного происхождения. В некоторых вариантах осуществления, сахарная группа не природного происхождения выбрана из группы, состоящей из: OH, замещенного низшего алкила, алкарила, аралкила, O-алкарила или O-аралкила, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂F; O-алкила, S-алкила, N-алкила; O-алкенила, S-алкенила, N-алкенила; O-алкинила, S-алкинила, N-алкинила; O-алкил-O-алкила, 2'-F, 2'-OCH₃, 2'-O(CH₂)₂OCH₃, где алкил, алкенил и алкинил могут быть замещены или не замещены C₁-C₁₀ алкилом, C₂-C₁₀ алкенилом, C₂-C₁₀ алкинилом, O[(CH₂)_nO]_mCH₃, -O(CH₂)_nOCH₃, -O(CH₂)_nNH₂, -O(CH₂)_nCH₃, -O(CH₂)_n-NH₂ и -O(CH₂)_nON[(CH₂)_nCH₃]₂, где n и m составляют от 1 до примерно 10; и/или модификации в положении 5': 5'-винила, 5'-метила (R или S); модификации в положении 4': 4'-S, гетероциклоалкила, гетероциклоалкарила, аминокламино, полиалкиламино, замещенного силила, группы, расщепляющей РНК, репортерной группы, интеркалятора, группы для улучшения фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида, и любой их комбинации.

В некоторых вариантах осуществления, в настоящем документе описана клетка для синтеза *in vivo* полипептидов не природного происхождения, содержащая: по меньшей мере, две различные пары кодон-антикодон не природного происхождения, где каждая пара кодон-антикодон не природного происхождения содержит кодон не природного происхождения из матричной РНК (мРНК) не природного происхождения и антикодон не природного происхождения из транспортной рибонуклеиновой кислоты (тРНК) не природного происхождения, где указанный кодон не природного происхождения содержит первый нуклеотид не природного происхождения, и указанный антикодон не природного происхождения содержит второй нуклеотид не природного происхождения и, по меньшей мере, две разные аминокислоты не природного происхождения, каждая из которых ковалентно связана с соответствующей тРНК не природного происхождения. В некоторых случаях, клетка дополнительно содержит, по меньшей мере, одну молекулу ДНК не природного происхождения, содержащую, по меньшей мере, четыре пары оснований (УВР) не природного происхождения. В настоящем документе, в некоторых вариантах осуществления, описана клетка для *in vivo* синтеза полипептидов не природного происхождения, где клетка содержит: по меньшей мере, одну молекулу ДНК не природного происхождения, содержащую, по меньшей мере, четыре пары оснований не природного происхождения, где, по меньшей мере, одна молекула ДНК не природного происхождения кодирует (i) молекулу информационной рибонуклеиновой кислоты (мРНК), кодирующую полипептид не природного происхождения и содержащую, по меньшей мере, первый и второй кодоны не природного происхождения, и (ii) по меньшей мере, первую и вторую молекулы транспортной РНК (тРНК), где первая молекула тРНК содержит первый

антикодон не природного происхождения, и вторая молекула тРНК, содержащая второй антикодон не природного происхождения, и, по меньшей мере, четыре пары оснований не природного происхождения в, по меньшей мере, одной молекуле ДНК находятся в таком контексте последовательности, что первый и второй кодоны не природного происхождения молекулы мРНК комплементарны первому и второму антикодонам не природного происхождения, соответственно. В некоторых вариантах осуществления, клетка дополнительно содержит молекулу мРНК и, по меньшей мере, первую и вторую молекулы тРНК. В некоторых вариантах осуществления клетки, по меньшей мере, первая и вторая молекулы тРНК ковалентно связаны с аминокислотами не природного происхождения. В некоторых вариантах осуществления, клетка дополнительно содержит полипептид не природного происхождения.

В некоторых вариантах осуществления, первый нуклеотид не природного происхождения расположен во втором или третьем положении кодона не природного происхождения и представляет собой комплементарную пару оснований со вторым нуклеотидом не природного происхождения антикодона не природного происхождения. В некоторых случаях, первый нуклеотид не природного происхождения и второй нуклеотид не природного происхождения содержат первое и второе основания не природного происхождения, независимо выбранные из группы, состоящей из



основания. В некоторых вариантах осуществления, клетки дополнительно содержат, по меньшей мере, одну молекулу ДНК не природного происхождения, содержащую, по меньшей мере, четыре пары оснований не природного происхождения (UBP). В некоторых случаях, по меньшей мере, четыре пары оснований не природного происхождения независимо выбраны из группы, состоящей из dCNMO/dTPT3, dNaM/dTPT3, dCNMO/dTAT1 или dNaM/dTAT1. В некоторых случаях, по меньшей мере, одна молекула ДНК не природного происхождения содержит, по меньшей мере, один плазмид. В

некоторых вариантах осуществления, по меньшей мере одна молекула ДНК не природного происхождения интегрирована в геном клетки. В некоторых вариантах осуществления, по меньшей мере одна молекула ДНК не природного происхождения кодирует полипептид не природного происхождения. В некоторых вариантах осуществления, клетки, как описано в настоящем документе, экспрессируют транспортер нуклеозидтрифосфата. В некоторых вариантах осуществления, транспортер нуклеозидтрифосфата содержит аминокислотную последовательность *PtNTT2*. В некоторых случаях, транспортер нуклеозидтрифосфата содержит усеченную аминокислотную последовательность *PtNTT2*, необязательно, где усеченная аминокислотная последовательность *PtNTT2* на, по меньшей мере, 80% идентична *PtNTT2*, кодируемой SEQ ID NO. 1. В некоторых вариантах осуществления, клетки экспрессируют, по меньшей мере, две тРНК синтетазы. В некоторых вариантах осуществления, по меньшей мере, две тРНК синтетазы представляют собой химерные *PyIRS* (*chPyIRS*) и *M. jannaschii* *AzFRS* (*MjpAzFRS*). В некоторых вариантах осуществления, клетки содержат нуклеотиды не природного происхождения, содержащие сахарную группу не природного происхождения. В некоторых случаях, сахарная группа не природного происхождения выбрана из группы, состоящей из: модификации в положении 2': OH, замещенного низшего алкила, алкарила, аралкила, O-алкарила или O-аралкила, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂F; O-алкила, S-алкила, N-алкила; O-алкенила, S-алкенила, N-алкенила; O-алкинила, S-алкинила, N-алкинила; O-алкил-O-алкила, 2'-F, 2'-OCH₃, 2'-O(CH₂)₂OCH₃, где алкил, алкенил и алкинил могут быть замещены или не замещены C₁-C₁₀ алкилом, C₂-C₁₀ алкенилом, C₂-C₁₀ алкинилом, O[(CH₂)_nO]_mCH₃, -O(CH₂)_nOCH₃, -O(CH₂)_nNH₂, -O(CH₂)_nCH₃, -O(CH₂)_n-NH₂ и -O(CH₂)_nON[(CH₂)_nCH₃]₂, где n и m составляют от 1 до примерно 10; и/или модификации в положении 5': 5'-винила, 5'-метила (R или S); модификации в положении 4': 4'-S, гетероциклоалкила, гетероциклоалкарила, аминоалкиламино, полиалкиламино, замещенного силила, группы, расщепляющей РНК, репортерной группы, интеркалятора, группы для улучшения фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида, и любой их комбинации. В некоторых вариантах осуществления, клетки содержат, по меньшей мере, одно нуклеотидное основание не природного происхождения, которое распознается РНК полимеразой во время транскрипции. В некоторых вариантах осуществления, клетки, описанные в настоящем документе, транслируют, по меньшей мере, один полипептид не природного происхождения, содержащий, по меньшей мере, две аминокислоты не природного происхождения. В некоторых случаях, по меньшей мере, две аминокислоты не природного происхождения выбраны из группы, состоящей из N6-((азидоэтокси)-

карбонил)-L-лизина (AzK), N6-пропаргилэтокси-карбонил-L-лизина (PraK), N6-(пропаргилокси)карбонил-L-лизина (PrK), п-азидофенилаланина (pAzF), BCN-L-лизина, норборненлизина, TCO-лизина, метилтетразинлизина, аллилоксикарбониллизина, 2-амино-8-оксононановой кислоты, 2-амино-8-оксооктановой кислоты, п-ацетил-L-фенилаланина, п-азидометил-L-фенилаланина (pAMF), п-йодо-L-фенилаланина, м-ацетилфенилаланина, 2-амино-8-оксононаноевой кислоты, п-пропаргилоксифенилаланина, п-пропаргилфенилаланина, 3-метилфенилаланина, L-Дора, фторированного фенилаланина, изопропил-L-фенилаланина, п-азидо-L-фенилаланина, п-ацил-L-фенилаланина, п-бензоил-L-фенилаланина, п-бромфенилаланина, п-амино-L-фенилаланина, изопропил-L-фенилаланина, О-аллилтирозина, О-метил-L-тирозина, О-4-аллил-L-тирозина, 4-пропил-L-тирозина, фосфонтирозина, три-О-ацетил-GlcNAc-серины, L-фосфосерина, фосфоносерины, L-3-(2-нафтил)аланина, 2-амино-3-((2-((3-(бензилокси)-3-оксопропил)амино)этил)селанил)пропановой кислоты, 2-амино-3-(фенилселанил)пропановой кислоты, селеноцистеина, N6-(((2-азидобензил)окси)карбонил)-L-лизина, N6-(((3-азидобензил)окси)карбонил)-L-лизина и N6-(((4-азидобензил)окси)карбонил)-L-лизина. В некоторых вариантах осуществления, клетки, описанные в настоящем документе, представляют собой выделенные клетки. В некоторых альтернативах, клетки, описанные в настоящем документе, представляют собой прокариоты. В некоторых случаях, клетки, описанные в настоящем документе, содержат клеточную линию.

КРАТКОЕ ОПИСАНИЕ ЧЕРТЕЖЕЙ

Различные аспекты настоящего описания подробно изложены в прилагаемой формуле изобретения. Лучшее понимание особенностей и преимуществ настоящего описания будет получено путем обращения к следующему подробному описанию, в котором представлены иллюстративные варианты осуществления, в которых используются принципы настоящего описания, и прилагаемым чертежам, в которых:

На **фиг. 1** показан рабочий процесс с использованием пар оснований не природного происхождения (UBP) для сайт-специфического включения не канонических аминокислот (ncAA) в полипептид не природного происхождения или белок не природного происхождения с использованием пары оснований не природного происхождения X-Y. Включение трех ncAA в полипептид не природного происхождения или белок не природного происхождения показано только в качестве примера; может быть включено любое количество ncAA.

На **фиг. 2** изображены типовые пары оснований не природного происхождения (UBP).

На **фиг. 3** показаны аналоги дезоксирибы X. Дезоксирибоза и фосфаты не показаны для ясности.

На **фиг. 4А-В** показаны аналоги рибонуклеотидов. **Фиг. 4А** представляет собой изображение аналогов рибонуклеотида X, где рибоза и фосфаты не показаны для ясности. **Фиг. 4В** представляет собой изображение аналогов рибонуклеотида Y, где рибоза и фосфаты не показаны для ясности.

На **фиг. 5А-Г** показаны типовые аминокислоты не природного происхождения. **Фиг. 5А** взята из **фиг. 2** книги Young *et al.*, “Beyond the canonical 20 amino acids: expanding the genetic lexicon,” *J. of Biological Chemistry* 285(15): 11039-11044 (2010). На **фиг. 5В** представлены типовые лизиновые производные аминокислот не природного происхождения. На **фиг. 5С** представлены типовые фенилаланиновые производные аминокислот не природного происхождения. На **фиг. 5D-5G** представлены типовые аминокислоты не природного происхождения. Эти аминокислоты не природного происхождения (UAA) были генетически кодированы в белках (**фиг. 5D** - UAA № 1-42; **фиг. 5E** - UAA № 43-89; **фиг. 5F** - UAA № 90-128; **фиг. 5G** - UAA № 129-167). **Фиг. 5D-5G** взяты из таблицы Dumas *et al.*, *Chemical Science* 2015, 6, 50-69.

На **фиг. 6А-Д** показано продуцирование белка в не клональных SSO с использованием кодонов и антикодонов не природного происхождения. Кодоны не природного происхождения и антикодоны не природного происхождения записаны с точки зрения кодирующей их ДНК последовательности. На **фиг. 6А** представлена химическая структура UBP dNaM-dTPT3. На **фиг. 6В** представлены собой химические структуры ncAA, AzK, PrK и pAzF. На **фиг. 6С** представлена схематическая иллюстрация генной кассеты, используемой для экспрессии sfGFP¹⁵¹(NNN) и *M. mazei* тРНК^{Pro}(NNN), где NNN относится к любому указанному кодону или антикодону. **Фиг. 6D** изображает нормализованную флуоресценцию не клональных культур SSO в конечной точке экспрессии белка (т. е. $t=180$ мин после добавления аТс) с использованием определенных кодонов и антикодонов, с или без AzK, в среде (е.п., произвольные единицы). Каждая повторная культура происходит из другой партии компетентных стартовых клеток SSO, трансформированных плазмидой, несущей UBP ($n=3$, биологические репликаты). Показано среднее значение с отдельными точками данных. Один типовой вырезанный вестерн-блот очищенного sfGFP, подвергнутого SPAAC с TAMRA-PEG₄-DBCO из культур SSO, показанных над каждым кодоном и антикодоном (только канал α -GFP). Вставка на **фиг. 6D** представляет собой диаграмму рассеяния средней флуоресценции в конечной точке в присутствии AzK (из **фиг. 6D**) по сравнению со средним значением количественно оцененного относительного белкового сдвига, индуцированного SPAAC ($n=3$; биологические репликаты). Семь верхних кодонов, выбранных для дальнейшего анализа, обведены кружком.

На **фиг. 7А-В** показано продуцирование белка и анализ ортогональности кодонов в клональных SSO. Кодоны не природного происхождения и антикодоны не природного происхождения записаны с точки зрения кодирующей их последовательности ДНК. На **фиг. 7А** изображена нормализованная флуоресценция клональных SSO в конечной точке экспрессии белка (т.е. $t=180$ мин после добавления аТс) для семи верхних кодонов и антикодонов (*слева*), а также четырех других выбранных кодонов (*справа*), с и без AzK.

Каждую повторную культуру размножают из отдельной колонии SSO (слева: $n=3$, справа: $n=[5, 4, 3, 3]$; биологические репликаты). Показано среднее значение с отдельными точками данных. Показан один типовой вырезанный вестерн-блот очищенного sfGFP, подвергнутого SPAAC с TAMRA-PEG₄-DBCO из культур SSO (только канал α -GFP). На **фиг. 7B** представлена нормализованная флуоресценция клональных культур SSO в конечной точке экспрессии для кодонов AXC, GXT и AGX и антикодонов GYT, AYC и XCT. Исследуют все парные комбинации, с и без AzK, в среде, в также без рибонуклеозидтрифосфатов NaMTP и TPTЗTP в среде. Каждую культуру размножают из одной колонии, и указано среднее значение \pm стандартное отклонение (черный текст; $n=3$; биологические репликаты).

На **фиг. 8A-F** показано одновременное декодирование двух кодонов не природного происхождения. Кодоны не природного происхождения и антикодоны не природного происхождения записаны с точки зрения кодирующей их последовательности ДНК. На **фиг. 8A** схематически показана генная кассета, содержащая *sfGFP*^{190,200}(GXT, AXC), *M. mazei* tRNA^{Pyl}(AYC) и *M. jannaschii* tRNA^{pAzF}(GYT). **Фиг. 8B-C**, график временной динамики нормализованной флуоресценции во время экспрессии *sfGFP* в присутствии обозначенных псАА. IPTG добавляют при $t=-60$ мин, и аТс добавляют при $t=0$. Экспрессию каждого репликата проводят в культурах, размноженных из отдельной колонии SSO ($n=3$, биологические репликаты). Показаны средние и отдельные точки данных. На **фиг. 8B** показана клональная экспрессия SSO кассеты с **фиг. 8A**, а также контроли, демонстрирующие экспрессию кассет, содержащих только одиночные кодоны с соответствующей тРНК. **Фиг. 8C** иллюстрирует клональную экспрессию кассеты, содержащей *sfGFP*^{190,200}(TAA, TAG), *M. mazei* тРНК^{Pyl}(TТА) и *M. jannaschii* тРНК^{pAzF}(СТА) также показаны, а также контрольный кассеты, содержащие одиночные стоп-кодоны с соответствующей супрессорной тРНК. На **фиг. 8D** показаны псевдоокрашенные вестерн-блоты α -GFP и флуоресцентные сканирования TAMRA очищенного sfGFP из SSO с **фиг. 8B-C** с и без конъюгации с TAMRA-PEG₄-DBCO с помощью SPAAC. Изображения вырезаны из тех же блотов (конструкции UBP и супрессоры стоп-кодона), но расположены так, чтобы выровнять несмещенную полосу, чтобы облегчить сравнение электрофоретической миграции. На **фиг. 8E** показан график временной динамики нормализованной флуоресценции во время клональной экспрессии кассет с двойным кодоном/тРНК с **фиг. 8B-C** с добавлением PrK и *pAzF*. Показаны средние и индивидуальные точки данных ($n=3$, биологические репликаты). На **фиг. 8F** показаны псевдоокрашенные вестерн-блоты α -GFP и сканы флуоресценции TAMRA очищенного sfGFP из SSO на **фиг. 8E**, с и без конъюгации с TAMRA-PEG₄-DBCO с помощью SPAAC и с TAMRA-PEG₄-азидом с помощью CuAAC.

На **фиг. 9A-C** показано одновременное декодирование трех кодонов не природного происхождения. Кодоны не природного происхождения и антикодоны не природного происхождения записаны с точки зрения кодирующей их последовательности ДНК. **Фиг. 9A** представляет собой схематическую иллюстрацию генной кассеты, содержащей

sfGFP^{151,190,200}(AXC, GXT, AGX), *M. mazei* tRNA^{Pyl}(XCT), *M. jannaschii* tRNA^{pAzF}(GYT) и *E. coli* tRNA^{Ser}(AYC). **Фиг. 9B** представляет собой график временной динамики нормализованной флуоресценции во время экспрессии *sfGFP* в отсутствие или в присутствии AzK и /или *pAzF*. IPTG добавляют при $t=-60$ мин, и аТс добавляют при $t=0$. Каждую повторную экспрессию проводят в культурах, размноженных из отдельной колонии SSO ($n=3$, биологические повторы). Показаны средние и отдельные точки данных. На **фиг. 9C** представлен типовой масс-спектр после деконволюции, полученный в результате анализа HRMS интактного *sfGFP*, очищенного от SSO на **фиг. 9B**. Метки пиков обозначают молекулярную массу, а также количественную оценку каждого пика по отношению к другим соответствующим видам. Использован стандартный однобуквенный код аминокислоты. Показано среднее значение \pm стандартное отклонение для каждого из этих видов ($n=3$).

На **фиг. 10** показан первоначальный скрининг кодонов не природного происхождения в не клональных SSO. Кодоны не природного происхождения и антикодоны не природного происхождения записаны с точки зрения их кодирующей последовательности ДНК. Парные ленточные диаграммы нормализованной флуоресценции клеток SSO в конечной точке экспрессии белка (т. е. $t=180$ мин после добавления аТс) для выбранных пар кодон/антикодон, несущих UBP в первом, втором или третьем положении кодона. Плюс/минус обозначает добавление 20 mM AzK к среде. Каждый репликат происходит из другой партии компетентных стартовых клеток SSO ($n=3$, биологические репликаты).

На **фиг. 11A-B** показаны вестерн-блоты и сканы флуоресценции для не клональной экспрессии SSO. Кодоны не природного происхождения и антикодоны не природного происхождения записаны с точки зрения их кодирующей последовательности ДНК. На **фиг. 11A** показаны псевдоокрашенные вестерн-блоты α -GFP и сканы флуоресценции TAMRA очищенного *sfGFP* из культур на **фиг. 6D** с конъюгацией с TAMRA-PEG4-DBCO с помощью SPAAC. Знак плюс/минус означает, что был проведен SPAAC. Проведено три испытания (обозначены 1, 2, 3; биологические репликаты). Три испытания каждого набора (NXN/NYN и NNX/XNN) проводят параллельно. **Фиг. 11B**, количественные оценки относительного сдвига в вестерн-блотах (на **фиг. 11A**) для определенных пар кодон/антикодон (т.е. сигнал сдвинутой полосы, деленный на общий сигнал как сдвинутых, так и не сдвинутых полос). Знак плюс/минус означает, что была проведена SPAAC. Показаны среднее значение \pm стандартное отклонение, а также отдельные точки данных ($n=3$).

Фиг. 12A-B иллюстрируют вестерн-блоты и сканы флуоресценции для клональной экспрессии SSO. Кодоны не природного происхождения и антикодоны не природного происхождения записаны с точки зрения их кодирующей последовательности ДНК. **Фиг. 12A**, псевдоокрашенные вестерн-блоты α -GFP и сканы флуоресценции TAMRA очищенного *sfGFP* из культур на **ФИГ. 7A** с конъюгацией с TAMRA-PEG4-DBCO с помощью SPAAC. Отображаемая (обрезанная) область мигрирует между стандартными

белковыми маркерами 32 кДа и 25 кДа. **Фиг. 12В**, количественные оценки относительных сдвигов в вестерн блотах (на **фиг. 12А**) для указанных кодонов. Показано среднее \pm стандартное отклонение, а также отдельные точки данных ($n=3$, за исключением n СХС=5 и n GXG=4)

На **фиг. 13** показана экспрессия клональных SSO в отсутствие ТРТЗТР. Кодоны не природного происхождения и антикодоны не природного происхождения записаны с точки зрения их кодирующей последовательности ДНК. Нормализованная флуоресценция клональных SSO в конечной точке экспрессии белка (т.е. $t=180$ мин после добавления аТс) для четырех верхних самоспаривающихся кодонов/антикодонов. Каждую экспрессию репликата проводят в культурах, размноженных из отдельной колонии, как это сделано на **фиг. 7А** ($n=3$, биологические повторы). Показано среднее значение \pm стандартное отклонение как для флуоресценции, так и для количественно оцененного вестерн блота сдвига белка (т. е. относительный сдвиг; гели не показаны), а также отдельные точки данных для флуоресценции.

На **фиг. 14** показаны контроли экспрессии двойных кодонов. Кодоны не природного происхождения и антикодоны не природного происхождения записаны с точки зрения их кодирующей последовательности ДНК. График временной динамики нормализованной флуоресценции во время экспрессии sfGFP определенных генотипов, с или без обозначенных псАА в среде. IPTG добавляют при $t=-60$ мин, и аТс добавляют при $t=0$. Каждую повторную экспрессию проводят в культурах, размноженных из отдельной колонии ($n=3$, биологические репликаты). Показаны средние и отдельные точки данных.

На **фиг. 15А-В** показан HRMS анализ белка из экспрессии двойного кодона. HRMS анализ интактного sfGFP, очищенного от SSO, экспрессирующих sfGFP^{151,190,200}(GXT, АХС), tRNA^{PyI}(АУС) и tRNA^{pAzF}(GYT) с AzK и pAzF в среде, как показано на **фиг. 8В** ($n=3$, биологические репликаты). Использован стандартный однобуквенный код аминокислот. На **фиг. 15А** представлены спектры после деконволюции с аннотацией соответствующих пиков и их относительного содержания друг к другу. На **фиг. 15В** показана назначение и интерпретация пиков.

На **фиг. 16А-В** показан HRMS анализ белка из экспрессии тройного кодона. HRMS анализ интактного sfGFP, очищенного от SSO, экспрессирующих sfGFP^{151,190,200}(АХС, GXT, AGX), tRNA^{PyI}(ХСТ), tRNA^{pAzF}(GYT) и tRNA^{Ser}(АУС) с AzK и pAzF в среде, как показано на **фиг. 9В** ($n=3$, биологические репликаты). Использован стандартный однобуквенный код аминокислоты. На **фиг. 16А** представлены спектры после деконволюции с аннотацией соответствующих пиков и их относительного содержания друг к другу. На **фиг. 16В** показана назначение и интерпретация пиков.

ПОДРОБНОЕ ОПИСАНИЕ

Определенная терминология

Если не указано иное, все технические и научные термины, используемые в настоящем документе, имеют то же значение, которое обычно понимается специалистом в области техники, к которой относится заявленный объект изобретения. Следует понимать,

что предшествующее общее описание и следующее подробное описание являются только примерными и объясняющими, и не ограничивают какой-либо заявленный объект. В этой заявке использование единственного числа включает множественное число, если специально не указано иное. Следует отметить, что, как используется в описании и прилагаемой формуле изобретения, формы единственного числа «a», «an» и «the» включают ссылки во множественном числе, если контекст явно не указывает на иное. В этой заявке использование «или» означает «и/или», если не указано иное. Кроме того, использование термина «включая» а также другие формы, такие как «включать», «включает» и «включается», не является ограничивающим.

Используемые в настоящем документе диапазоны и количества могут быть выражены как «приблизительно» для конкретного значения или диапазона. Примерно также включает точное количество. Следовательно, «примерно 5 мкл» означает «примерно 5 мкл», а также «5 мкл». В общем, термин «примерно» включает количество, которое, как ожидается, будет находиться в пределах ошибки эксперимента.

Такие фразы, как «в условиях, подходящих для получения» или «в условиях, достаточных для получения» и подобные, в контексте способов синтеза, используемые в настоящем документе, относятся к условиям реакции, таким как время, температура, растворитель, концентрации реагентов и подобные, которые экспериментатор может варьировать в пределах своей квалификации, которые дают полезное количество или выход продукта реакции. Необязательно, чтобы желаемый продукт реакции был единственным продуктом реакции, или чтобы исходные материалы были полностью израсходованы, при условии, что желаемый продукт реакции может быть выделен или иным образом использован в дальнейшем.

Под «химически возможным» подразумевается расположение связей или соединение, в котором не нарушаются общепонятные правила органической структуры; например, структуру в рамках определения пункта формулы изобретения, которая в определенных ситуациях может содержать пятивалентный атом углерода, которого не существует в природе, следует понимать как не подпадающую под формулу изобретения. Структуры, описанные в настоящем документе, во всех их вариантах осуществления, предназначены для включения только «химически возможных» структур и любых перечисленных структур, которые не являются химически возможными, например, в структуре, показанной с переменными атомами или группами не предназначены для описания или заявки в настоящем документе.

«Аналог» химической структуры, как этот термин используется в настоящем документе, относится к химической структуре, которая сохраняет существенное сходство с исходной структурой, хотя ее нельзя легко получить синтетическим путем из исходной структуры. В некоторых вариантах осуществления, аналог нуклеотида представляет собой нуклеотид не природного происхождения. В некоторых вариантах осуществления, аналог нуклеозида представляет собой нуклеозид не природного происхождения. Родственная химическая структура, которая легко получается синтетическим путем из исходной

химической структуры, называется «производной».

Соответственно, термин полинуклеотид, используемый в настоящем документе, относится к ДНК, РНК, ДНК- или РНК-подобным полимерам, таким как пептидные нуклеиновые кислоты (ПНК), закрытые нуклеиновые кислоты (ЗНК), фосфоротиоаты, основания не природного происхождения и подобные, которые хорошо известны в данной области техники. Полинуклеотиды могут быть синтезированы в автоматических синтезаторах, например, с использованием химии фосфорамидитов или других химических подходов, адаптированных для использования в синтезаторах.

ДНК включает, но не ограничен ими, кДНК и геномную ДНК. ДНК может быть присоединена ковалентными или не ковалентными способами к другой биомолекуле, включая, но не ограничиваясь ими, РНК и пептид. РНК включает кодирующую РНК, например информационную РНК. (иРНК). В некоторых вариантах осуществления, РНК представляет собой рРНК, РНКи, мякРНК, микроРНК, киРНК, мяРНК, эксРНК, риРНК, длинную нкРНК или любую их комбинацию или гибрид. В некоторых случаях РНК представляет собой компонент рибозима. ДНК и РНК может быть в любой форме, включая, помимо прочего, линейную, кольцевую, сверхспиральную, одноцепочечную и двухцепочечную.

Пептидная нуклеиновая кислота (ПНК) представляет собой синтетический аналог ДНК/РНК, в котором пептидоподобный остов заменяет сахаро-фосфатный остов ДНК или РНК. Олигомеры ПНК демонстрируют более высокую силу связывания и большую специфичность в связывании с комплементарными ДНК, где несоответствие оснований ПНК/ДНК является более дестабилизирующим, чем аналогичное несоответствие в дуплексе ДНК/ДНК. Эта сила связывания и специфичность также применимы к дуплексам ПНК/РНК. ПНК не легко распознаются ни нуклеазами, ни протеазами, что делает их устойчивыми к расщеплению ферментами. ПНК также стабильны в широком диапазоне pH. См. также Nielsen PE, Egholm M, Berg RH, Buchardt O (December 1991). "Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide", *Science* **254** (5037): 1497-500. doi:10.1126/science.1962210. PMID 1962210; и Egholm M, Buchardt O, Christensen L, Behrens C, Freier SM, Driver DA, Berg RH, Kim SK, Nordén B, and Nielsen PE (1993), "PNA Hybridizes to Complementary Oligonucleotides Obeying the Watson-Crick Hydrogen Bonding Rules". *Nature* **365** (6446): 566-8. doi:10.1038/365566a0. PMID 7692304

Запертая нуклеиновая кислота (ЗНК) представляет собой модифицированный нуклеотид РНК, в котором рибозная группа нуклеотида ЗНК модифицирована дополнительным мостиком, соединяющим 2'-кислород и 4'-углерод. Мост «запирает» рибозу в конформации 3'-эндо (North), которая часто встречается в дуплексах А-формы. Нуклеотиды ЗНК можно смешивать с остатками ДНК или РНК в олигонуклеотиде, когда это необходимо. Такие олигомеры можно синтезировать химическим путем, и они коммерчески доступны. Заблокированная конформация рибозы улучшает укладку оснований и предварительную организацию остова. См., например, Kaur, H; Aroga, A; Wengel, J; Maiti, S (2006), "Thermodynamic, Counterion, and Hydration Effects for the

Incorporation of Locked Nucleic Acid Nucleotides into DNA Duplexes", *Biochemistry* **45** (23): 7347-55. doi:10.1021/bi060307w. PMID 16752924; Owczarzy R.; You Y., Groth C.L., Tataurov A.V. (2011), "Stability and mismatch discrimination of locked nucleic acid-DNA duplexes.", *Biochem.* **50** (43): 9352-9367. doi:10.1021/bi200904e. PMC 3201676. PMID 21928795; Alexei A. Koshkin; Sanjay K. Singh, Poul Nielsen, Vivek K. Rajwanshi, Ravindra Kumar, Michael Meldgaard, Carl Erik Olsen, Jesper Wengel (1998), "LNA (Locked Nucleic Acids): Synthesis of the adenine, cytosine, guanine, 5-methylcytosine, thymine and uracil bicyclonucleoside monomers, oligomerisation, and unprecedented nucleic acid recognition", *Tetrahedron* **54** (14): 3607-30. doi:10.1016/S0040-4020(98)00094-5; и Satoshi Obika; Daishu Nanbu, Yoshiyuki Hari, Ken-ichiro Morio, Yasuko In, Toshimasa Ishida, Takeshi Imanishi (1997), "Synthesis of 2'-O,4'-C-methyleneuridine and -cytidine. Novel bicyclic nucleosides having a fixed C3'-endo sugar pucker", *Tetrahedron Lett.* **38** (50): 8735-8. doi:10.1016/S0040-4039(97)10322-7.

Молекулярный маяк или зонд молекулярного маяка представляет собой зонд гибридизации олигонуклеотидов, который может обнаруживать присутствие определенной последовательности нуклеиновой кислоты в гомогенном растворе. Молекулярные маяки представляют собой молекулы в форме шпильки с внутренне погашенным флуорофором, флуоресценция которых восстанавливается, когда они связываются с последовательностью нуклеиновой кислоты-мишени. См., например, Tyagi S, Kramer FR (1996), "Molecular beacons: probes that fluoresce upon hybridization", *Nat Biotechnol.* **14** (3): 303-8. PMID 9630890; Täpp I, Malmberg L, Rennel E, Wik M, Syvänen AC (2000 Apr), "Homogeneous scoring of single-nucleotide polymorphisms: comparison of the 5'-nuclease TaqMan assay and Molecular Beacon probes", *Biotechniques* **28** (4): 732-8. PMID 10769752; и Akimitsu Okamoto (2011), "ECHO probes: a concept of fluorescence control for practical nucleic acid sensing", *Chem. Soc. Rev.* **40**: 5815-5828.

В некоторых вариантах осуществления, нуклеотидное основание обычно представляет собой часть гетероциклического основания нуклеозиды. Нуклеотидные основания могут встречаться в природе, могут быть модифицированы, могут не иметь сходства с природными основаниями и могут быть синтезированы, например, с помощью органического синтеза. В определенных вариантах осуществления, нуклеотидное основание содержит любой атом или группу атомов, способных взаимодействовать с основанием другой нуклеиновой кислоты с или без использования водородных связей. В некоторых вариантах осуществления, нуклеотидное основание не природного происхождения не происходит из природного нуклеотидного основания. Необходимо отметить, что нуклеотидные основания не природного происхождения не обязательно обладают основными свойствами, однако для простоты называются нуклеотидными основаниями. В некоторых вариантах осуществления, когда речь идет о нуклеотидных основаниях, «(d)» указывает, что нуклеотидные основания могут быть присоединены к дезоксирибозе или рибозе.

В некоторых вариантах осуществления, нуклеозид представляет собой соединение, содержащее группу нуклеотидного основания и группу сахара. Нуклеозиды включают, но

не ограничены ими, встречающиеся в природе нуклеозиды (такие, как обнаружены в ДНК и РНК), нуклеозиды с удаленным азотистым основанием, модифицированные нуклеозиды и нуклеозиды, имеющие миметические основания и/или сахарные группы. Нуклеозиды включают нуклеозиды, содержащие любое разнообразие заместителей. Нуклеозид может быть гликозидным соединением, образованным посредством гликозидной связи между нуклеиновым основанием и восстанавливающей группой сахара.

В некоторых вариантах осуществления кодоны не природного происхождения мРНК и антикодоны не природного происхождения тРНК, как описано в настоящем описании, могут быть записаны в терминах кодирующей их последовательности ДНК, например, антикодоны не природного происхождения тРНК могут быть записаны как GYU или GYT.

Заголовки разделов, используемые в настоящем документе, предназначены только для организационных целей и не должны толковаться как ограничивающие описываемый объект.

Композиции и способы *in vivo* синтеза полипептидов не природного происхождения

В настоящем документе описаны композиции и способы *in vivo* синтеза полипептидов не природного происхождения с расширенным генетическим алфавитом. В некоторых случаях композиции и способы, описанные в настоящем документе, содержат молекулу нуклеиновой кислоты не природного происхождения, кодирующую полипептид не природного происхождения, где полипептид не природного происхождения содержит аминокислоту не природного происхождения. В некоторых случаях, полипептид не природного происхождения содержит, по меньшей мере, две аминокислоты не природного происхождения. В некоторых случаях, полипептид не природного происхождения содержит, по меньшей мере, три аминокислоты не природного происхождения. В некоторых случаях, полипептид не природного происхождения содержит две аминокислоты не природного происхождения. В некоторых случаях, полипептид не природного происхождения содержит три аминокислоты не природного происхождения. В некоторых случаях, по меньшей мере, две аминокислоты не природного происхождения, которые включаются в полипептид не природного происхождения, могут быть одинаковыми или разными аминокислотами не природного происхождения. В некоторых случаях, аминокислоты не природного происхождения включаются в полипептид не природного происхождения сайт-специфическим образом. В некоторых случаях, полипептид не природного происхождения представляет собой белок не природного происхождения.

В некоторых случаях композиции и способы, описанные в настоящем документе, включают полусинтетический организм (SSO). В некоторых случаях, способы включают, по меньшей мере, одну пару оснований не природного происхождения (UBP) в, по меньшей мере, одной молекуле нуклеиновой кислоты не природного происхождения. В некоторых вариантах осуществления, способы включают включение одного UBП, по меньшей мере, в

одну молекулу нуклеиновой кислоты не природного происхождения. В некоторых вариантах осуществления, способы включают включение двух UBP, по меньшей мере, в одну молекулу нуклеиновой кислоты не природного происхождения. В некоторых вариантах осуществления, способы включают включение трех UBP, по меньшей мере, в одну молекулу нуклеиновой кислоты не природного происхождения. Пары оснований UBP образуются путем спаривания между нуклеотидными основаниями не природного происхождения двух нуклеозидов не природного происхождения. В некоторых вариантах осуществления, молекула нуклеиновой кислоты не природного происхождения представляет собой молекулу ДНК не природного происхождения.

В некоторых вариантах осуществления, по меньшей мере, одна молекула нуклеиновой кислоты не природного происхождения представляет собой или содержит одну молекулу (например, плазмиду или хромосому). В некоторых вариантах осуществления, по меньшей мере, одна молекула нуклеиновой кислоты не природного происхождения представляет собой или содержит две молекулы (например, две плазмиды, две хромосомы или хромосому и плазмиду). В некоторых вариантах осуществления, по меньшей мере, одна молекула нуклеиновой кислоты не встречающаяся в природе представляет собой или содержит три молекулы (например, три плазмиды, две плазмиды и хромосому, плазмиду и две хромосомы или три хромосомы). Примеры хромосом включают геномные хромосомы, в которые была интегрирована UBP, и искусственные хромосомы (например, бактериальные искусственные хромосомы), содержащие UBP. В некоторых вариантах осуществления, где используют, по меньшей мере, одну молекулу ДНК не природного происхождения, содержащую, по меньшей мере, четыре пары оснований не встречающиеся в природе, и, по меньшей мере, одна молекула ДНК не природного происхождения представляет собой две или несколько молекул, по меньшей мере, четыре пары оснований не природного происхождения могут быть распределены между двумя или несколькими молекулами любым возможным способом (например, одна в первой и три во второй, две в первой и две во второй и т. д.).

В некоторых случаях, по меньшей мере, одна молекула нуклеиновой кислоты не природного происхождения, необязательно включающая UBP, транскрибируется с получением молекулы информационной РНК, содержащей, по меньшей мере, один кодон не природного происхождения, содержащий, по меньшей мере, один нуклеотид не природного происхождения. В некоторых вариантах осуществления, транскрипция относится к созданию одной или нескольких молекул РНК, комплементарных части молекулы ДНК. В некоторых случаях, нуклеотид не природного происхождения занимает первое, второе или третье положение кодона в кодоне не природного происхождения, например, второе или третье положение кодона. В некоторых случаях, два нуклеотида не природного происхождения занимают первое и второе, первое и третье, второе и третье или первое и третье положения кодона в кодоне не природного происхождения. В некоторых случаях, три нуклеотида не природного происхождения занимают все три положения кодона в кодоне не природного происхождения. В некоторых случаях, мРНК, содержащая

нуклеотиды не природного происхождения, содержит, по меньшей мере, два кодона не природного происхождения (в некоторых вариантах осуществления, выражение «по меньшей мере, два кодона не природного происхождения» является взаимозаменяемым с «по меньшей мере, первый и второй кодон не природного происхождения»). В некоторых случаях, мРНК, содержащая нуклеотиды не природного происхождения, содержит два кодона не природного происхождения. В некоторых случаях, мРНК, содержащая нуклеотиды не природного происхождения, содержит три кодона не природного происхождения.

В некоторых вариантах осуществления, молекула нуклеиновой кислоты не природного происхождения, необязательно включающая UBP, транскрибируется с образованием, по меньшей мере, одной молекулы тРНК, где молекула тРНК содержит антикодон не природного происхождения, содержащий, по меньшей мере, один нуклеотид не природного происхождения. В некоторых случаях, нуклеотид не природного происхождения занимает первое, второе или третье положение антикодона в антикодоне не природного происхождения. В некоторых случаях, два нуклеотида не природного происхождения занимают первое и второе, первое и третье, второе и третье или первое и третье положение антикодона в антикодоне не природного происхождения. В некоторых случаях, три нуклеотида не природного происхождения занимают все три положения антикодона в антикодоне не природного происхождения. В некоторых случаях, молекула не нуклеиновой кислоты природного происхождения, необязательно включающая UBP, транскрибируется с получением, по меньшей мере, двух тРНК, содержащих, по меньшей мере, два антикодона не природного происхождения. В некоторых случаях, по меньшей мере, два антикодона не природного происхождения могут быть одинаковыми или разными. В некоторых случаях, молекула нуклеиновой кислоты не природного происхождения, необязательно включающая UBP, транскрибируется с получением двух тРНК, содержащих антикодона не природного происхождения, которые могут быть одинаковыми или разными. В некоторых случаях, молекула нуклеиновой кислоты не природного происхождения, необязательно включающая UBP, транскрибируется с получением трех тРНК, содержащих три антикодона не природного происхождения, которые могут быть одинаковыми или разными.

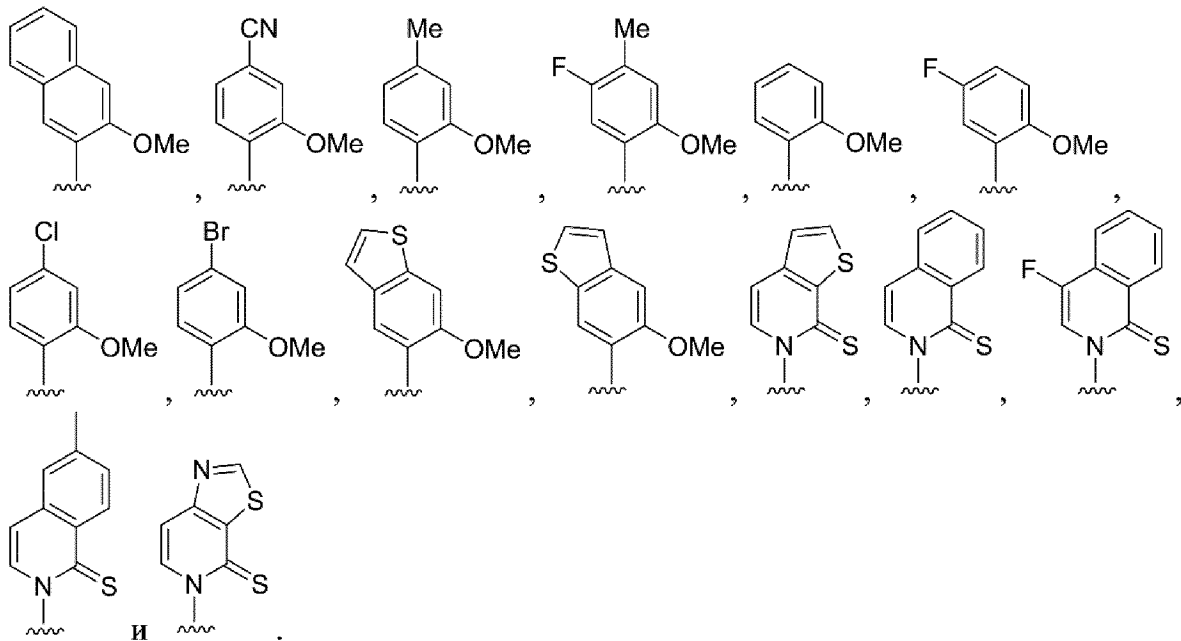
В некоторых вариантах осуществления, по меньшей мере, один кодон не природного происхождения, кодируемый мРНК, может быть комплементарным, по меньшей мере, антикодону не природного происхождения тРНК с образованием пары кодон-антикодон не природного происхождения. В некоторых случаях, композиции и способы, описанные в настоящем документе, включают синтез полипептида не природного происхождения с одной, двумя, тремя или несколькими парами кодон-антикодон не встречающимися в природе. В некоторых случаях, композиции и способы, описанные в настоящем документе, включают синтез полипептида не природного происхождения с двумя парами кодон-антикодон не природного происхождения. В некоторых случаях, композиции и способы, описанные в настоящем документе, включают синтез полипептида не природного

происхождения с тремя парами кодон-антикодон не природного происхождения.

В некоторых случаях, композиции и способы, описанные в настоящем документе, включают синтез полипептида не природного происхождения с одной, двумя, тремя или несколькими аминокислотами не природного происхождения с использованием одной, двух, трех или нескольких пар кодон-антикодон не природного происхождения. В некоторых случаях, композиции и способы, описанные в настоящем документе, включают синтез полипептида не природного происхождения с двумя аминокислотами не природного происхождения с использованием двух пар кодон-антикодон не природного происхождения. В некоторых случаях, композиции и способы, описанные в настоящем документе, включают синтез полипептида не природного происхождения с тремя аминокислотами не природного происхождения с использованием трех пар кодон-антикодон не природного происхождения.

В некоторых случаях, кодон не природного происхождения содержит последовательность нуклеиновой кислоты XNN, NXN, NNX, XXN, XNX, NXX или XXX, и антикодон не природного происхождения содержит последовательность нуклеиновой кислоты XNN, YNN, NXN, NYN, NNX, NNY, NXX, NYU, XNX, YNY, XXN, YYN или YYY для образования пары кодон-антикодон не природного происхождения. В некоторых случаях, пара кодон-антикодон не природного происхождения состоит из NNX-XNN, NNX-YNN или NXN-NYN, где N представляет собой любой природный нуклеотид, X представляет собой первый нуклеотид не природного происхождения, и Y представляет собой второй нуклеотид не природного происхождения. В некоторых вариантах осуществления, любой природный нуклеотид включает нуклеотиды, имеющие стандартное основание, такое как аденин, тимин, урацил, гуанин или цитозин, и нуклеотиды, имеющие природное модифицированное основание, такое как псевдоуридин, 5-метилцитозин и т. д. В некоторых вариантах осуществления, пара кодон-антикодон не природного происхождения содержит, по меньшей мере, один G в кодоне и, по меньшей мере, один C в антикодоне. В некоторых вариантах осуществления, пара кодон-антикодон не природного происхождения содержит, по меньшей мере, один G или C в кодоне и, по меньшей мере, один комплементарный C или G в антикодоне. Каждый из X и Y независимо выбран из группы, состоящей из: (i) 2-тиоурацила, 2'-дезоксидуридина, 4-тиоурацила, урацил-5-ила, гипоксантин-9-ила (I), 5-галоурацила, 5-пропинилурацила, 6-азоурацила, 5-метиламинометилурацила, 5-метоксиаминометил-2-тиоурацила, псевдоурацила, метилового эфира урацил-5-оксуксусной кислоты, урацил-5-оксуксусной кислоты, 5-метил-2-тиоурацила, 3-(3-амино-3-N-2-карбоксивпропил)урацила, 5-метил-2-тиоурацила, 4-тиоурацила, 5-метилурацила, 5'-метоксикарбоксиметилурацила, 5-метоксиурацила, урацил-5-оксиуксусной кислоты, 5-(карбоксихидроксиметил)урацила, 5-карбоксиметиламинометил-2-тиоуридина, 5-карбоксиметиламинометилурацила, дигидроурацила, 5-гидроксиметилцитозина, 5-трифторметилцитозина, 5-галоцитозина, 5-пропинилцитозина, 5-гидроксцитозина, циклоцитозина, цитозина арабинозида, 5,6-дигидроцитозина, 5-нитроцитозина, 6-азоцитозина, азацитозина, N4-этилцитозина, 3-

метилцитозина, 5-метилцитозина, 4-ацетилцитозина, 2-тиоцитозина, феноксазина цитидина ([5,4-b][1,4]бензоксазин-2(3H)-она), фенотиазина цитидин (1H-пиримидо[5,4-b][1,4]бензотиазин-2(3H)-она), феноксазина цитидина (9-(2-аминоэтокси)-H-пиримидо[5,4-b][1,4]бензоксазин-2(3H)-она), карбазола цитидина (2H-пиримидо[4,5-b]индол-2-она), пиридоиндола цитидина (H-пиридо[3',2':4,5]пирроло[2,3-d]пиримидин-2-она), 2-аминоаденина, 2-пропиладенина, 2-аминоаденина, 2-F-аденина, 2-аминопропиладенина, 2-амино-2'-дезоксаденозина, 3-дезааденина, 7-метиладенина, 7-дезааденина, 8-азааденина, 8-галоген-, 8-амино-, 8-тиол-, 8-тиоалкил- и 8-гидрокси-замещенных аденинов, N6-изопентениладенина, 2-метиладенина, 2,6-диаминопурина, 2-метилтио-N6-изопентениладенина, 6-азааденина, 2-метилгуанина, 2-пропила и алкильных производных гуанина, 3-дезагуанина, 6-тиогуанина, 7-метилгуанина, 7-дезагуанина, 7-дезагуанозина, 7-деза-8-азагуанина, 8-азагуанина, 8-галоген-, 8-амино-, 8-тиол-, 8-тиоалкил- и 8-гидрокси-замещенных гуанинов, 1-метилгуанина, 2,2-диметилгуанина, 7-метилгуанина или 6-азагуанина, гипоксантина, ксантина, 1-метилюридина, кеозина, бета-D-галактозилкеозина, инозина, бета-D-маннозилкеозина, вибутоксозина, гидроксимочевины, (аср3)w, 2-аминопиридина или 2-пиридона. В некоторых вариантах осуществления, каждый из X и Y независимо выбран из группы, состоящей из:



В некоторых вариантах осуществления, пара кодон-антикодон не природного происхождения содержит NNX-XNN, где NNX-XNN выбран из группы, состоящей из AAX-XUU, AUX-XAU, ACX-XGU, AGX-XCU, UAX-XUA, UUX-XAA, UCX-XGA, UGX-XCA, SAX-XUG, CUX-XAG, CCX-XGG, CGX-XCG, GAX-XUC, GUX-XAC, GCX-XGC и GGX-XCC. В некоторых вариантах осуществления, пара кодон-антикодон не природного происхождения содержит NNX-YNN, где NNX-YNN выбран из группы, состоящей из AAX-YUU, AUX-YAU, ACX-YGU, AGX-YCU, UAX-YUA, UUX-YAA, UCX-YGA, UGX-YCA, SAX-YUG, CUX-YAG, CCX-YGG, CGX-YCG, GAX-YUC, GUX-YAC, GCX-YGC и GGX-YCC. В некоторых случаях способы, пара кодон-антикодон не природного

происхождения содержит NXN-NXN, где NXN-NXN выбран из группы, состоящей из AXA-UXU, AXU-AXU, AXC-GXU, AXG-CXU, UXA-UXA, UXU-AXA, UXC-GXA, UXG-CXA, CXА-UXG, CXU-AXG, CXC-GXG, CXG-CXG, GXA-UXC, GXU-AXC, GXC-GXC и GXG-CXC. В некоторых случаях способы, пара кодон-антикодон не природного происхождения содержит NXN-NYN, где NXN-NYN выбран из группы, состоящей из AXA-UYU, AXU-AYU, AXC-GYU, AXG-CYU, UXA-UYA, UXU-AYA, UXC-GYA, UXG-CYA, CXА-UYG, CXU-AYG, CXC-GYG, CXG-CYG, GXA-UYC, GXU-AYC, GXC-GYC и GXG-CYC.

В некоторых вариантах осуществления, пара кодон-антикодон не природного происхождения содержит XNN-NNX, где XNN-NNX выбран из группы, состоящей из XAA-UUX, XAU-AUX, XAC-AGX, XAG-CUX, XUA-UAX, XUУ-ААХ, XUC-GAX, XUG-CAX, XCA-UGX, XCU-AGX, XCC-GGX, XCG-CGX, XGA-UCX, XGU-ACX, XGC-GCX и XGG-CCX. В некоторых вариантах осуществления, пара кодон-антикодон не природного происхождения содержит XNN-NNY, где XNN-NNY выбран из группы, состоящей из XAA-UUY, XAU-AUY, XAC-AGY, XAG-CUY, XUA-UAY, XUУ-ААУ, XUC-GAY, XUG-CAY, XCA-UGY, XCU-AGY, XCC-GGY, XCG-CGY, XGA-UCY, XGU-ACY, XGC-GCY и XGG-CCY.

В некоторых вариантах осуществления, пара кодон-антикодон не природного происхождения содержит XXN-NXX, где XXN-NXX выбран из группы, состоящей из XXA-UXX, XXU-AXX, XXC-GXX и XXG-CXX. В некоторых вариантах осуществления, пара кодон-антикодон не природного происхождения содержит XXN-NYY, где XXN-NYY выбран из группы, состоящей из XXA-UYY, XXU-AYY, XXC-GYY и XXG-CYY. В некоторых альтернативах, пара кодон-антикодон не природного происхождения содержит XNX-XNX, где XNX-XNX выбран из группы, состоящей из XAX-XUX, XUX-XAX, XCX-XGX и XGX-XCX. В некоторых вариантах осуществления, пара кодон-антикодон не природного происхождения содержит XNX-YNУ, где XNX-YNУ выбран из группы, состоящей из XAX-YUY, XUX-YAY, XCX-YGY и XGX-YCY. В некоторых случаях, пара кодон-антикодон не природного происхождения содержит NXX-XXN, где NXX-XXN выбран из группы, состоящей из AXX-XXU, UXX-XXA, CXX-XXG и GXX-XXC. В некоторых случаях, пара кодон-антикодон не природного происхождения содержит NXX-YYN, где NXX-YYN выбран из группы, состоящей из AXX-YYU, UXX-YYA, CXX-YYG и GXX-YYC. В некоторых случаях, пара кодон-антикодон не природного происхождения содержит XXX-XXX или XXX-YYY.

На типовом технологическом процессе 100 (**фиг. 1**) способа получения полипептида не природного происхождения с расширенным генетическим алфавитом (**фиг. 2**), ДНК 101, кодирующая белок 102, и тРНК 103, каждая из которых содержит комплементарные нуклеотидные основания не природного происхождения (X, Y), транскрибируется 104 с образованием тРНК 106 и иРНК 107. X является первым нуклеотидом не природного происхождения и Y является вторым нуклеотидом не природного происхождения. После зарядки тРНК аминокислотой не природного происхождения 105, иРНК 107 транслируется

108 с образованием белка 110, содержащего одну или несколько аминокислот не природного происхождения 109. Способы и композиции, описанные в настоящем документе, в некоторых случаях позволяют встраивать аминокислоты не природного происхождения в сайты с высокой точностью и выходом. В настоящем документе также описаны полусинтетические организмы, включающие расширенный генетический алфавит, способы использования организмов для продуцирования белковых продуктов, включая такие, которые содержат, по меньшей мере, один аминокислотный остаток не природного происхождения.

Выбор нуклеотидных оснований не природного происхождения позволяет оптимизировать одну или несколько стадий в описанных в настоящем документе способах. Например, нуклеотидные основания выбирают для высокоэффективной репликации, транскрипции и/или трансляции. В некоторых случаях, более одной пары нуклеотидных оснований не природного происхождения используют для способов, описанных в настоящем документе. Например, первый набор нуклеотидных оснований, содержащих группу дезоксирибо, используют для репликации ДНК (такой как первое нуклеотидное основание и второе нуклеотидное основание, сконфигурированные для образования первой пары оснований), и второй набор нуклеотидных оснований (такой как третье нуклеотидное основание и четвертое нуклеотидное основание, где третье и четвертое нуклеотидные основания присоединены к рибозе, сконфигурированы с образованием второй пары оснований) используют для транскрипции/трансляции. Комплементарное спаривание между нуклеотидным основанием первого набора и нуклеотидным основанием второго набора в некоторых случаях позволяет транскрипцию генов для создания тРНК или белков из матрицы ДНК, содержащей нуклеотидные основания из первого набора. Комплементарное спаривание между нуклеотидными основаниями второго набора (вторая пара оснований) в некоторых случаях позволяет трансляцию путем сопоставления тРНК, содержащих нуклеиновые кислоты не природного происхождения и иРНК. В некоторых случаях, нуклеотидные основания в первом наборе присоединены к группе дезоксирибозы. В некоторых случаях, нуклеотидные основания в первом наборе присоединены к группе рибозы. В некоторых случаях, нуклеотидные основания обоих наборов уникальны. В некоторых случаях, по меньшей мере, одно нуклеотидное основание одинаково в обоих наборах. В некоторых случаях, первое нуклеотидное основание и третье нуклеотидное основание являются одинаковыми. В некоторых вариантах осуществления, первая пара оснований и вторая пара оснований не являются одинаковыми. В некоторых случаях, первая пара оснований, вторая пара оснований и третья пара оснований не являются одинаковыми.

В некоторых вариантах осуществления, выход полипептида не природного происхождения или белка не природного происхождения, синтезированного композициями и способами, описанными в настоящем документе, выше по сравнению с выходом того же полипептида не природного происхождения или белка не природного происхождения, синтезированного другими способами. В некоторых случаях, выход полипептида не

природного происхождения или белка не природного происхождения, синтезированного композициями и способами, описанными в настоящем документе, по меньшей мере, на 10%, по меньшей мере, на 20%, по меньшей мере, на 30%, по меньшей мере, на 40% или, по меньшей мере, на 50% выше, чем выход того же полипептида не природного происхождения или белка не природного происхождения, синтезированного с помощью других способов. Пример других способов включает способы, использующие супрессию янтарного кодона.

В некоторых случаях, растворимость полипептида не природного происхождения или белка не природного происхождения, синтезированного с помощью композиций и способов, описанных в настоящем документе, выше по сравнению с растворимостью того же полипептида не природного происхождения или белка не природного происхождения, синтезированного другими способами. В некоторых случаях, растворимость полипептида не природного происхождения или белка не природного происхождения, синтезированного с помощью композиций и способов, описанных в настоящем документе, по меньшей мере, на 10%, по меньшей мере, на 20%, по меньшей мере, на 30%, по меньшей мере, на 40% или, по меньшей мере, на 50% выше, чем для того же полипептида не природного происхождения или белка не природного происхождения, синтезированного другими способами. В некоторых случаях, биологическая активность белка не природного происхождения, синтезированного с помощью композиций и способов, описанных в настоящем документе, выше по сравнению с биологической активностью того же белка не природного происхождения, синтезированного другими способами. В некоторых случаях, биологическая активность белка не природного происхождения, синтезированного с помощью композиций и способов, описанных в настоящем документе, по меньшей мере, на 10%, по меньшей мере, на 20%, по меньшей мере, на 30%, по меньшей мере, на 40% или по меньшей мере, на 50% выше, чем биологическая активность того же белка не природного происхождения, синтезированного другими способами.

В некоторых вариантах осуществления композиции и способы *in vivo* синтеза неприродных полипептидов, как описано в настоящем документе, используют или содержат полусинтетический организм (SSO). В некоторых вариантах осуществления, SSO размножается клонированием во время синтеза полипептидов не природного происхождения. В некоторых случаях, SSO не размножается клонированием во время синтеза полипептидов не природного происхождения. В некоторых случаях, SSO может быть остановлен на любой фазе клеточного цикла во время синтеза полипептидов не природного происхождения. В некоторых вариантах осуществления, композиции и способы, как описано в настоящем документе, позволяют синтезировать полипептиды не природного происхождения *in vitro*. В некоторых случаях, композиции и способы, описанные в настоящем документе, могут включать бесклеточную систему для синтеза полипептидов не природного происхождения.

Молекулы нуклеиновой кислоты

В некоторых вариантах осуществления, нуклеиновая кислота (*например*, также

называемая в настоящем документе представляющей интерес молекулой нуклеиновой кислоты) происходит из любого источника или композиции, такой как ДНК, кДНК, гДНК (геномная ДНК), РНК, киРНК (короткая ингибиторная РНК), РНКи, тРНК, иРНК или рРНК (рибосомная РНК), например, и находится в любой форме (*например*, линейной, кольцевой, сверхспиральной, одноцепочечной, двухцепочечной и подобной). В некоторых вариантах осуществления, нуклеиновые кислоты содержат нуклеотиды, нуклеозиды или полинуклеотиды. В некоторых случаях, нуклеиновые кислоты также содержат нуклеиновые кислоты не природного происхождения, такие как аналоги ДНК или РНК (*например*, содержащие аналоги оснований, аналоги сахаров и/или не нативный остов и подобные). Понятно, что термин «нуклеиновая кислота» не относится к конкретной длине полинуклеотидной цепи или не подразумевает ее, поэтому в определение также включены полинуклеотиды и олигонуклеотиды. Примеры природных нуклеотидов включают, без ограничения, АТФ, УТФ, СТФ, ГТФ, АДП, УДП, СДП, ГДП, АМФ, УМФ, СМФ, ГМФ, дАТФ, дТТФ, дСТФ, дГТФ, дАДП, дТДП, дСДП, дГДП, дАМФ, дТМФ, дСМФ и дГМФ. Примеры природных рибонуклеотидов включают дАТФ, дТТФ, дСТФ, дГТФ, дАДП, дТДП, дСДП, дГДП, дАМФ, дТМФ, дСМФ и дГМФ. Для природной РНК урациловым основанием является уридин. Иногда нуклеиновая кислота представляет собой вектор, плазмиду, фагмид, автономно реплицирующуюся последовательность (ARS), центромер, искусственную хромосому, искусственную хромосому дрожжей (*например*, YAC) или другую нуклеиновую кислоту, способную реплицироваться или быть реплицированной в клетке-хозяине. В некоторых случаях, нуклеиновая кислота не природного происхождения представляет собой аналог нуклеиновой кислоты. В дополнительных случаях, нуклеиновая кислота не природного происхождения происходит из внеклеточного источника. В других случаях, нуклеиновая кислота не природного происхождения доступна для внутриклеточного пространства организма, представленного в настоящем документе, *например*, генетически модифицированного организма. В некоторых вариантах осуществления, нуклеотид не природного происхождения не является природным нуклеотидом. В некоторых вариантах осуществления, нуклеотид, который не содержит природное основание, содержит нуклеотидное основание не природного происхождения.

Нуклеиновые кислоты не природного происхождения

Нуклеотидный аналог или нуклеотид не природного происхождения включает нуклеотид, который содержит некоторый тип модификации основания, сахарной или фосфатной группы. В некоторых вариантах осуществления, модификация содержит химическую модификацию. В некоторых случаях, модификации происходят в 3'ОН или 5'ОН группе в основной цепи, в сахарном компоненте или в нуклеотидном основании. Модификации, в некоторых случаях, необязательно включают линкерные молекулы не природного происхождения и/или межцепочечные или внутрицепочечные поперечные связи. В одном аспекте, модифицированная нуклеиновая кислота содержит модификацию одной или нескольких групп 3'ОН или 5'ОН, остова, сахарного компонента или нуклеотидного основания, и/или добавление линкерных молекул не природного

происхождения. В одном аспекте, модифицированный остов включает остов, отличный от фосфодизфирного остова. В другом аспекте, модифицированный сахар содержит сахар, отличный от дезоксирибозы (в модифицированной ДНК) или отличный от рибозы (модифицированная РНК). В другом аспекте, модифицированное основание содержит аденин, гуанин, цитозин или тимин (в модифицированной ДНК) или основание, отличное от аденина, гуанина, цитозина или урацила (в модифицированной РНК).

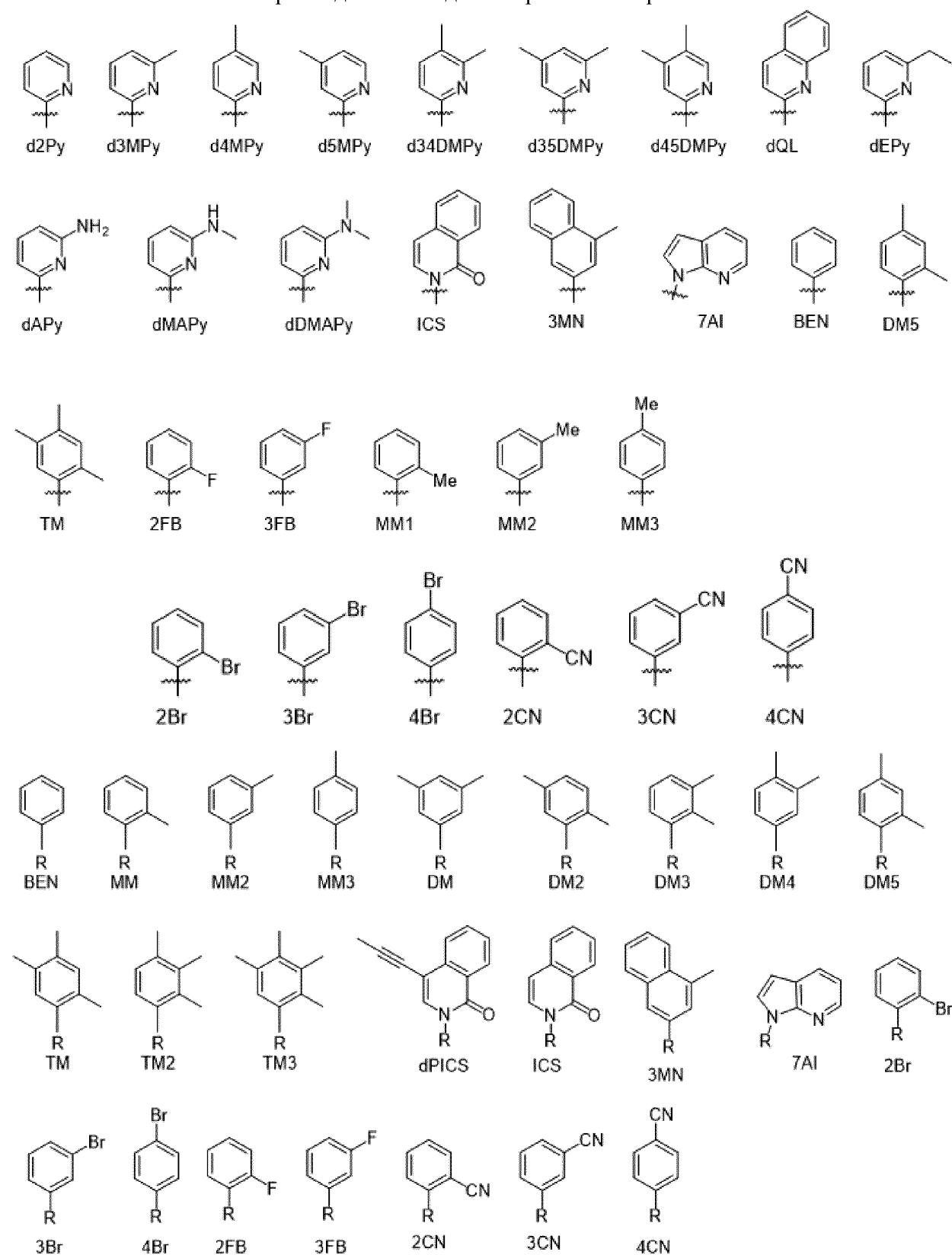
В некоторых вариантах осуществления, нуклеиновая кислота содержит, по меньшей мере, одно модифицированное основание. В некоторых случаях, нуклеиновая кислота содержит 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20 или более модифицированных оснований. В некоторых случаях, модификации основной группы включают природные и синтетические модификации А, С, G и T/U, а также различные пуриновые или пиримидиновые основания. В некоторых вариантах осуществления, модификация представляет собой модифицированную форму аденина, гуанина, цитозина или тимина (в модифицированной ДНК) или модифицированную форму аденина, гуанина, цитозина или урацила (в модифицированной РНК).

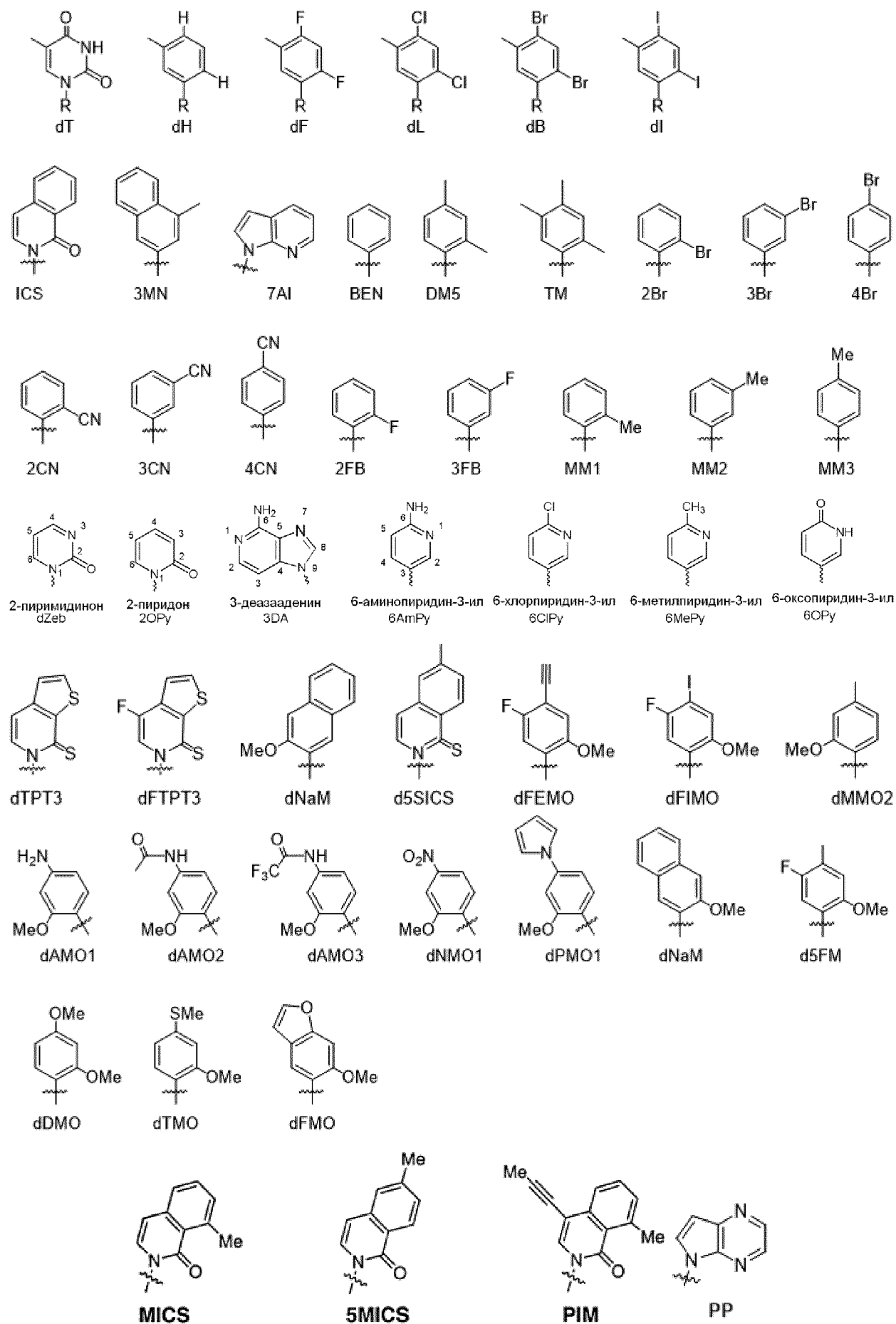
Модифицированное основание не природного происхождения нуклеиновой кислоты включает, но не ограничено ими, урацил-5-ил, гипоксантин-9-ил (I), 2-аминоаденин-9-ил, 5-метилцитозин (5-me-C), 5-гидроксиметилцитозин, ксантин, гипоксантин, 2-аминоаденин, 6-метил и другие алкильные производные аденина и гуанина, 2-пропил и другие алкильные производные аденина и гуанина, 2-тиоурацил, 2-тиотимин и 2-тиоцитозин, 5-галоурацил и -цитозин, 5-пропилиурацил и -цитозин, 6-азоурацил, -цитозин и -тимин, 5-урацил (псевдоурацил), 4-тиоурацил, 8-галоген-, 8-амино-, 8-тиол-, 8-тиоалкил-, 8- гидроксил- и другие 8-замещенные аденины и гуанины, 5-галоген-, особенно 5-бром-, 5-трифторметил- и другие 5-замещенные урацилы и цитозины, 7-метилгуанин и 7-метиладенин, 8-азагуанин и 8-азааденин, 7-деазагуанин и 7-деазааденин и 3-деазагуанин и 3-деазааденин. Некоторые нуклеиновые кислоты не природного происхождения, такие как 5-замещенные пиримидины, 6-азапиримидины и N-2 замещенные пурины, N-6 замещенные пурины, O-6 замещенные пурины, 2-аминопропиладенин, 5-пропилиурацил, 5-пропилилцитозин, 5-метилцитозин, которые повышают стабильность образования дуплексов, универсальные нуклеиновые кислоты, гидрофобные нуклеиновые кислоты, смешанные нуклеиновые кислоты, нуклеиновые кислоты увеличенного размера, фторированные нуклеиновые кислоты, 5-замещенные пиримидины, 6-азапиримидины и N-2, N-6 и O-6 замещенные пурины, включая 2-аминопропиладенин, 5-пропилиурацил и 5-пропилилцитозин, 5-метилцитозин (5-me-C), 5-гидроксиметилцитозин, ксантин, гипоксантин, 2-аминоаденин, 6-метил, другие алкильные производные аденина и гуанина, 2-пропил и другие алкильные производные аденина и гуанина, 2-тиоурацил, 2-тиотимин и 2-тиоцитозин, 5-галоурацил, 5-галоцитозин, 5-пропил (-C≡C-CH₃)урацил, 5-пропилилцитозин, другие алкильные производные пиримидиновых нуклеиновых кислот, 6-азоурацил, 6-азоцитозин, 6-азотимин, 5-урацил (псевдоурацил), 4-тиоурацил, 8-галоген-, 8-амино-, 8-тиол-, 8-тиоалкил-, 8-гидроксил- и другие 8-замещенные аденины и гуанины, 5-галоген-, особенно 5-бром-, 5-

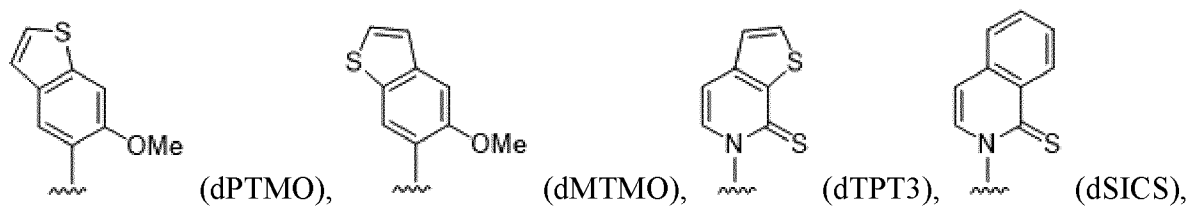
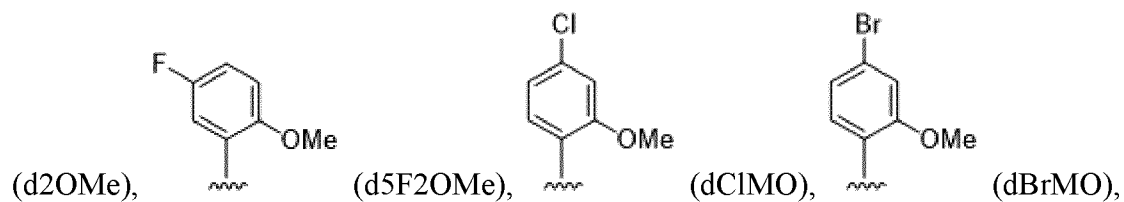
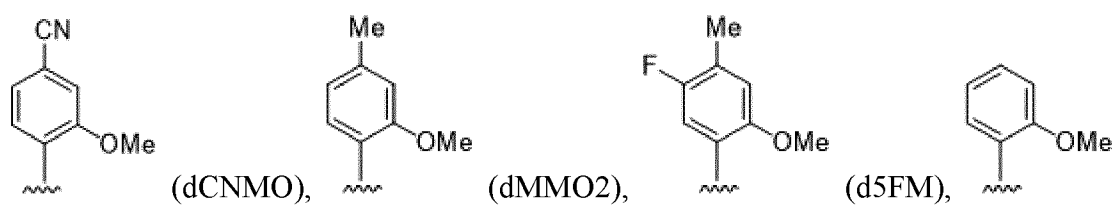
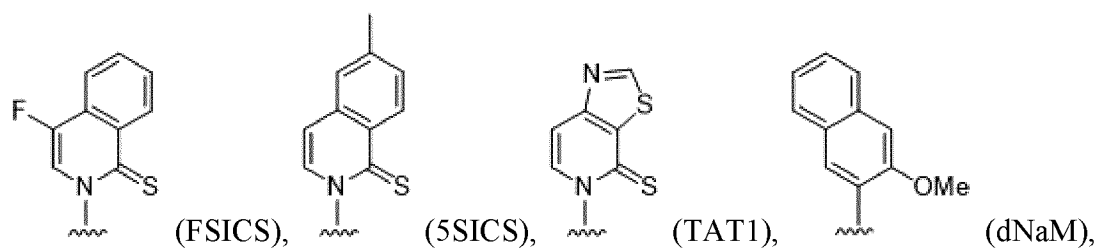
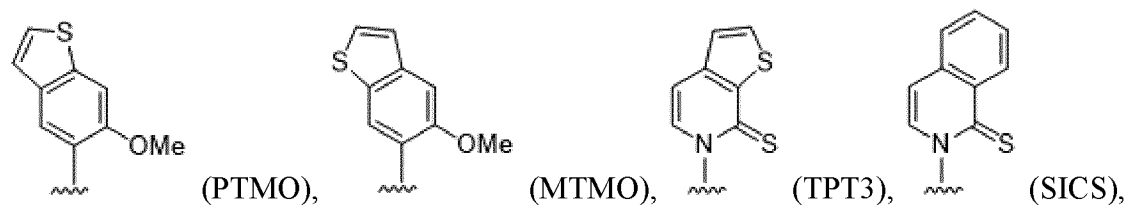
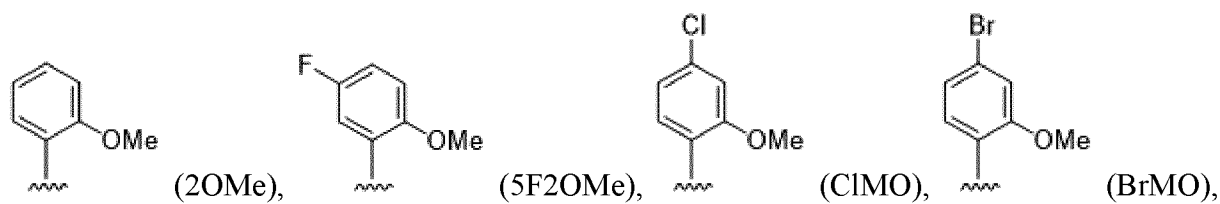
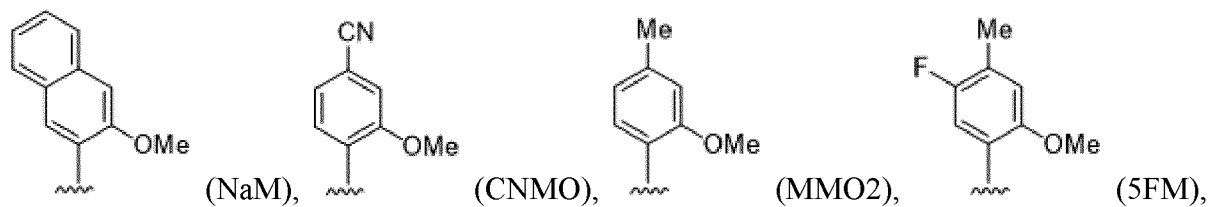
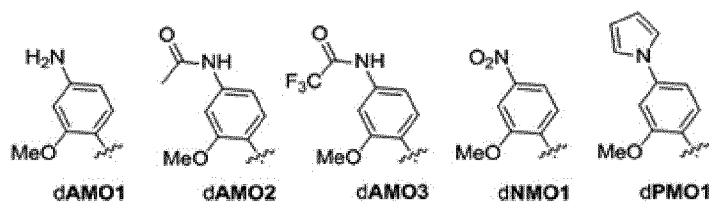
трифторметил-, другие 5-замещенные урацилы и цитозины, 7-метилгуанин, 7-метиладенин, 2-F-аденин, 2-аминоаденин, 8-азагуанин, 8-азааденин, 7-деазагуанин, 7-дезааденин, 3-деазагуанин, 3-дезааденин, трициклические пиримидины, феноксазин цитидин ([5,4-b][1,4]бензоксазин-2(3H)-он), фенотиазин цитидин (1H-пиримидо[5,4-b][1,4]бензотиазин-2(3H)-он), G-фиксирующие основания, феноксазин цитидин (*например*, 9-(2-аминоэтокси)-H-пиримидо[5,4-b][1,4]бензоксазин-2(3H)-он), карбазол цитидин (2H-пиримидо[4,5-b]индол-2-он), пиридоиндолцитидин (H-пиридо[3',2':4,5]пирроло[2,3-d]пиримидин-2-он), в которых пуриновое или пиримидиновое основание заменено другими гетероциклами, 7-дезааденин, 7-деазагуанозин, 2-аминопиридин, 2-пиридон, азацитозин, 5-бромцитозин, бромурацил, 5-хлорцитозин, хлорированный цитозин, циклоцитозин, цитозин арабинозид, 5-фторцитозин, фторпиримидин, фторурацил, 5,6-дигидроцитозин, 5-йодоцитозин, гидроксимочевина, йодоурацил, 5-нитроцитозин, 5-бромурацил, 5-хлорурацил, 5-фторурацил и 5-йодоурацил, 2-аминоаденин, 6-тиогуанин, 2-тиотимин, 4-тиотимин, 5-пропинилурацил, 4-тиоурацил, N4-этилцитозин, 7-деазагуанин, 7-деза-8-азагуанин, 5-гидроксицитозин, 2'-дезоксуридин, 2-амино-2'-дезоксаденозин и те, которые описаны в патентах США №№ 3,687,808; 4,845,205; 4,910,300; 4,948,882; 5,093,232; 5,130,302; 5,134,066; 5,175,273; 5,367,066; 5,432,272; 5,457,187; 5,459,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540; 5,587,469; 5,594,121; 5,596,091; 5,614,617; 5,645,985; 5,681,941; 5,750,692; 5,763,588; 5,830,653 и 6,005,096; Kandimalla et al., (2001) Bioorg. Med. Chem. 9:807-813; The Concise Encyclopedia of Polymer Science and Engineering, Kroschwitz, J.I., Ed., John Wiley & Sons, 1990, 858- 859; Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613; и Sanghvi, Chapter 15, Antisense Research and Applications, Crooke and Lebleu Eds., CRC Press, 1993, 273-288. Дополнительные модификации оснований можно найти, например, в патенте США № 3,687,808, Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613. В некоторых случаях, нуклеиновая кислота не природного происхождения содержит нуклеотидное основание с **ФИГ. 3**. В некоторых случаях, нуклеиновая кислота не природного происхождения содержит нуклеотидное основание с **ФИГ. 4А**. В некоторых случаях, нуклеиновая кислота не природного происхождения содержит нуклеотидное основание с **ФИГ. 4В**.

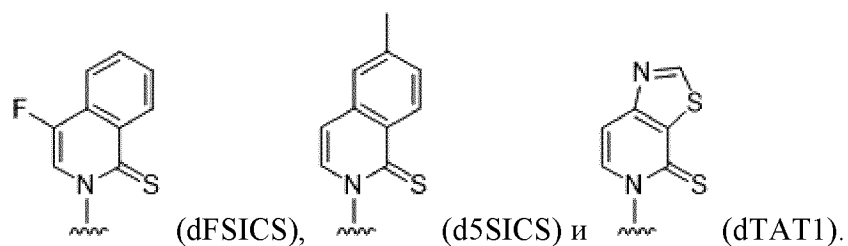
Нуклеиновые кислоты не природного происхождения, содержащие различные гетероциклические основания и различные группы сахаров (и аналоги сахаров), доступны в данной области техники, и нуклеиновые кислоты в некоторых случаях включают одно или несколько гетероциклических оснований, отличных от пяти основных компонентов природных нуклеиновых кислот. Например, гетероциклическое основание включает в некоторых случаях группы урацил-5-ил, цитозин-5-ил, аденин-7-ил, аденин-8-ил, гуанин-7-ил, гуанин-8-ил, 4-аминопирроло[2,3-d]пиримидин-5-ил, 2-амино-4-оксопирроло[2,3-d]пиримидин-5-ил, 2-амино-4-оксопирроло[2,3-d]пиримидин-3-ил, где пурины присоединены к сахарной группе нуклеиновой кислоты через положение 9, пиримидины через положение 1, пирролопиримидины через положение 7 и пиразолопиримидины через положение 1.

В некоторых вариантах осуществления, модифицированное основание не природного происхождения нуклеиновой кислоты изображено ниже, где волнистой линией или R обозначена точка присоединения к дезоксирибозе или рибозе.









В некоторых вариантах осуществления, аналоги нуклеотидов также модифицированы на фосфатной группе. Модифицированные фосфатные группы включают, но не ограничены ими, соединения с модификацией на связи между двумя нуклеотидами и содержат, например, фосфоротиоат, хиральный фосфоротиоат, фосфородитиоат, фосфотриэфир, аминокилфосфотриэфир, метил и другие алкилфосфонаты, включая 3'-алкиленфосфонат и хиральные фосфонаты, фосфинаты, фосфорамидаты, включая 3'-аминофосфорамидат и аминокилфосфорамидаты, тионофосфорамидаты, тионоалкилфосфонаты, тионоалкилфосфотриэфиры и боранофосфаты. Понятно, что эти фосфатные или модифицированные фосфатные связи между двумя нуклеотидами проходят через связь 3'-5' или связь 2'-5', и связь имеет обратную полярность, такую как 3'-5' к 5'-3' или 2'-5' к 5'-2'. Также включены различные соли, смешанные соли и формы свободных кислот. Многочисленные патенты США учат тому, как получать и использовать нуклеотиды, содержащие модифицированные фосфаты, и включают, но не ограничены ими, 3,687,808; 4,469,863; 4,476,301; 5,023,243; 5,177,196; 5,188,897; 5,264,423; 5,276,019; 5,278,302; 5,286,717; 5,321,131; 5,399,676; 5,405,939; 5,453,496; 5,455,233; 5,466,677; 5,476,925; 5,519,126; 5,536,821; 5,541,306; 5,550,111; 5,563,253; 5,571,799; 5,587,361; и 5,625,050.

В некоторых вариантах осуществления, нуклеиновые кислоты не природного происхождения включают 2',3'-дидезокси-2',3'-дидегидронуклеозиды (PCT/US2002/006460), 5'-замещенные производные ДНК и РНК (PCT/US2011/033961; Saha et al., *J. Org Chem.*, 1995, 60, 788-789; Wang et al., *Bioorganic & Medicinal Chemistry Letters*, 1999, 9, 885-890; и Mikhailov et al., *Nucleosides & Nucleotides*, 1991, 10(1-3), 339-343; Leonid et al., 1995, 14(3-5), 901-905; и Eppacher et al., *Helvetica Chimica Acta*, 2004, 87, 3004-3020; PCT/JP2000/004720; PCT/JP2003/002342; PCT/JP2004/013216; PCT/JP2005/020435; PCT/JP2006/315479; PCT/JP2006/324484; PCT/JP2009/056718; PCT/JP2010/067560), или 5'-замещенные мономеры, сделанные в виде монофосфата с модифицированными основаниями (Wang et al., *Nucleosides Nucleotides & Nucleic Acids*, 2004, 23 (1 & 2), 317-337).

В некоторых вариантах осуществления, нуклеиновые кислоты не природного происхождения включают модификации в 5'-положении и 2'-положении сахарного кольца (PCT/US94/02993), такие как 5'-CH₂-замещенные 2'-О-защищенные нуклеозиды (Wu et al., *Helvetica Chimica Acta*, 2000, 83, 1127-1143 and Wu et al., *Bioconjugate Chem.* 1999, 10, 921-924). В некоторых случаях, нуклеиновые кислоты не природного происхождения включают амид-связанные нуклеозидные димеры, которые были получены для введения в олигонуклеотиды, где 3'-связанный нуклеозид в димере, (5'-3') содержит 2'-OCH₃ и 5'-(S)-

CH₃ (Mesmaeker et al., Synlett, 1997, 1287-1290). Нуклеиновые кислоты не природного происхождения могут включать 2'-замещенные 5'-CH₂ (или O) модифицированные нуклеозиды (PCT/US92/01020). нуклеиновые кислоты не природного происхождения могут включать 5'-метиленфосфонатные ДНК и РНК мономеры и димеры (Bohringer et al., Tet. Lett., 1993, 34, 2723-2726; Collingwood et al., Synlett, 1995, 7, 703-705; and Hutter et al., Helvetica Chimica Acta, 2002, 85, 2777-2806). Нуклеиновые кислоты не природного происхождения могут включать 5'-фосфонатные мономеры с 2'-заменой (US 2006/0074035) и другие модифицированные 5'-фосфонатные мономеры (WO 1997/35869). Нуклеиновые кислоты не природного происхождения могут включать 5'-модифицированные метиленфосфонатные мономеры (EP614907 и EP629633). Нуклеиновые кислоты не природного происхождения могут включать аналоги 5'- или 6'-фосфонатных рибонуклеозидов, содержащие гидроксильную группу в 5'- и/или 6'-положении (Chen et al., Phosphorus, Sulfur and Silicon, 2002, 777, 1783-1786; Jung et al., Bioorg. Med. Chem., 2000, 8, 2501-2509; Gallier et al., Eur. J. Org. Chem., 2007, 925-933; and Hampton et al., J. Med. Chem., 1976, 19(8), 1029-1033). Нуклеиновые кислоты не природного происхождения могут включать 5'-фосфонатдезоксирибонуклеозидные мономеры и димеры, имеющие 5'-фосфатную группу (Nawrot et al., Oligonucleotides, 2006, 16(1), 68-82). Нуклеиновые кислоты не природного происхождения могут включать нуклеозиды, имеющие 6'-фосфонатную группу, где 5'- и/или 6'-положение не замещено или замещено тио-трет-бутильной группой (SC(CH₃)₃) (и их аналоги); метиленаминогруппой (CH₂NH₂) (и их аналоги) или цианогруппой (CN) (и их аналоги) (Fairhurst et al., Synlett, 2001, 4, 467-472; Kappler et al., J. Med. Chem., 1986, 29, 1030-1038; Kappler et al., J. Med. Chem., 1982, 25, 1179-1184; Vrudhula et al., J. Med. Chem., 1987, 30, 888-894; Hampton et al., J. Med. Chem., 1976, 19, 1371-1377; Geze et al., J. Am. Chem. Soc, 1983, 105(26), 7638-7640; и Hampton et al., J. Am. Chem. Soc, 1973, 95(13), 4404-4414).

В некоторых вариантах осуществления, нуклеиновые кислоты не природного происхождения также включают модификации сахарной группы. В некоторых случаях, нуклеиновые кислоты содержат один или несколько нуклеозидов, в которых сахарная группа была модифицирована. Такие сахар-модифицированные нуклеозиды могут придавать повышенную нуклеазную стабильность, повышенную аффинность связывания или некоторые другие полезные биологические свойства. В некоторых вариантах осуществления, нуклеиновые кислоты содержат химически модифицированную рибофуранозную кольцевую группу. Примеры химически модифицированных рибофуранозных колец включают, без ограничений, добавление замещающих групп (включая 5'- и/или 2'-замещающие группы; связывание мостиком двух кольцевых атомов с образованием бициклических нуклеиновых кислот (БНК); замену атома кислорода рибозильного кольца с S, N(R) или C(R₁)(R₂) (R=H, C₁-C₁₂ алкил или защитная группа); и их комбинации. Примеры химически модифицированных сахаров могут быть найдены в WO 2008/101157, US 2005/0130923 и WO 2007/134181.

В некоторых случаях, модифицированная нуклеиновая кислота включает

модифицированные сахара или аналоги сахаров. Таким образом, в дополнение к рибозе и дезоксирибозе, сахарная группа может представлять собой пентозу, дезоксипентозу, гексозу, дезоксигексозу, глюкозу, арабинозу, ксилозу, ликсозу или сахарную «аналог» циклопентильную группу. Сахар может быть в пиранозильной или фуранозильной форме. Сахарная группа может представлять собой фуранозид рибозы, дезоксирибозы, арабинозы или 2'-О-алкилрибозы, и сахар может быть присоединен к соответствующим гетероциклическим основаниям либо в [альфа], либо в [бета] аномерной конфигурации. Сахарные модификации включают, но не ограничены ими, аналоги 2'-алкокси-РНК, аналоги 2'-амино-РНК, 2'-фтор-ДНК и химеры 2'-алкокси- или amino-РНК/ДНК. Например, модификация сахара может включать 2'-О-метилуридин или 2'-О-метилцитидин. Сахарные модификации включают 2'-О-алкилзамещенные дезоксирибонуклеозиды и 2'-О-этиленгликоль-подобные рибонуклеозиды. Получение этих сахаров или аналоги сахаров и соответствующие «нуклеозиды», где такие сахара или аналоги присоединены к гетероциклическому основанию (основанию нуклеиновой кислоты) известны. Можно также производить сахарные модификации и комбинировать их с другими модификациями.

Модификации группы сахара включают природные модификации рибозы и дезоксирибозы, а также модификации не природного происхождения. Сахарные модификации включают, но не ограничены ими, следующие модификации в положении 2': ОН, F, O-, S- или N-алкил; O-, S- или N-алкенил; O-, S- или N-алкинил; или O-алкил-O-алкил, где алкил, алкенил и алкинил могут быть замещенным или незамещенным C₁-C₁₀ алкилом или C₂-C₁₀ алкенилом и алкинилом. 2' сахарные модификации также включают, но не ограничены ими, -O[(CH₂)_nO]_mCH₃, -O(CH₂)_nOCH₃, -O(CH₂)_nNH₂, -O(CH₂)_nCH₃, -O(CH₂)_nONH₂ и -O(CH₂)_nON[(CH₂)_nCH₃]₂, где n и m составляют от 1 до примерно 10.

Другие модификации в положении 2' включают, но не ограничены ими: C₁-C₁₀ низший алкил, замещенный низший алкил, алкарил, аралкил, O-алкарил, O-аралкил, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂, гетероциклоалкил, гетероциклоалкарил, аминоалкиламино, полиалкиламино, замещенный силлил, расщепляющую РНК группу, репортерную группу, интеркалятор, группу для улучшения фармакокинетических свойств олигонуклеотида или группу для улучшения фармакодинамических свойств олигонуклеотида и другие заместители, обладающие аналогичными свойствами. Аналогичные модификации могут быть также сделаны в других положениях сахара, в частности в положении 3' сахара на 3' концевом нуклеотиде или в 2'-5'-связанных олигонуклеотидах, и в положении 5' на 5' концевом нуклеотиде. Модифицированные сахара также включают такие, которые содержат модификации в кислороде мостикового кольца, такие как CH₂ и S. Нуклеотидные аналоги сахара могут также иметь миметики сахара, такие как циклобутильные группы вместо пентофуранозилового сахара. Существует множество патентов США, которые описывают получение таких модифицированных сахарных структур, и которые подробно описывают ряд модификаций оснований, такие как патенты США №№ 4,981,957; 5,118,800; 5,319,080; 5,359,044; 5,393,878; 5,446,137; 5,466,786; 5,514,785; 5,519,134; 5,567,811; 5,576,427;

5,591,722; 5,597,909; 5,610,300; 5,627,053; 5,639,873; 5,646,265; 5,658,873; 5,670,633; 4,845,205; 5,130,302; 5,134,066; 5,175,273; 5,367,066; 5,432,272; 5,457,187; 5,459,255; 5,484,908; 5,502,177; 5,525,711; 5,552,540; 5,587,469; 5,594,121, 5,596,091; 5,614,617; 5,681,941; и 5,700,920, каждый из которых включен в настоящий документ посредством ссылки.

Примеры нуклеиновых кислот, содержащих модифицированные группы сахара, включают, без ограничения, нуклеиновые кислоты, содержащие 5'-винил, 5'-метил (R или S), 4'-S, 2'-F, 2'-OCH₃ и 2'-O(CH₂)₂OCH₃ группы заместителей. Заместители в положении 2' также могут быть выбраны из аллила, амина, азидо, тио, O-аллила, O-(C₁-C₁₀ алкила), OCF₃, O(CH₂)₂SCH₃, O(CH₂)₂-O-N(R_m)(R_n) и O-CH₂-C(=O)-N(R_m)(R_n), где каждый R_m и R_n независимо представляют собой H или замещенный или незамещенный C₁-C₁₀ алкил.

В некоторых вариантах осуществления, нуклеиновые кислоты, описанные в настоящем документе, включают одну или несколько бициклических нуклеиновых кислот. В некоторых таких вариантах осуществления, бициклические нуклеиновые кислоты содержат мостик между 4' и 2' атомами рибозильного кольца. В некоторых вариантах осуществления, нуклеиновые кислоты, представленные в настоящем документе, включают одно или несколько бициклических нуклеиновых кислот, где мостик содержит 4'-2' бициклическую нуклеиновую кислоту. Примеры таких 4'-2' бициклических нуклеиновых кислот включают, но не ограничены ими, одну из формул: 4'-(CH₂)-O-2' (LNA); 4'-(CH₂)-S-2'; 4'-(CH₂)₂-O-2' (ENA); 4'-CH(CH₃)-O-2' и 4'-CH(CH₂OCH₃)-O-2' и их аналоги (см. патент США №7,399,845), 4'-C(CH₃)(CH₃)-O-2' и ее аналоги (см. WO2009/006478, WO2008/150729, US 2004/0171570, патент США № 7,427,672, Chattopadhyaya et al., J. Org. Chem., 209, 74, 118-134, и WO 2008/154401). См. также, например: Singh et al., Chem. Commun., 1998, 4, 455-456; Koshkin et al., Tetrahedron, 1998, 54, 3607-3630; Wahlestedt et al., Proc. Natl. Acad. Sci. U. S. A., 2000, 97, 5633-5638; Kumar et al., Bioorg. Med. Chem. Lett., 1998, 8, 2219-2222; Singh et al., J. Org. Chem., 1998, 63, 10035-10039; Srivastava et al., J. Am. Chem. Soc., 2007, 129(26) 8362-8379; Elayadi et al., Curr. Opinion Invens. Drugs, 2001, 2, 558-561; Braasch et al., Chem. Biol, 2001, 8, 1-7; Oram et al., Curr. Opinion Mol. Ther., 2001, 3, 239-243; патенты США №№ 4,849,513; 5,015,733; 5,118,800; 5,118,802; 7,053,207; 6,268,490; 6,770,748; 6,794,499; 7,034,133; 6,525,191; 6,670,461; и 7,399,845; международные публикации №№ WO2004/106356, WO1994/14226, WO2005/021570, WO2007/090071, и WO2007/134181; заявки на патент США №№ US2004/0171570, US2007/0287831 и US2008/0039618; предварительные заявки США №№ 60/989,574, 61/026,995, 61/026,998, 61/056,564, 61/086,231, 61/097,787 и 61/099,844; и международные заявки №№ PCT/US2008/064591, PCT US2008/066154, PCT US2008/068922 и PCT/DK98/00393.

В некоторых вариантах осуществления, нуклеиновые кислоты содержат связанные нуклеиновые кислоты. Нуклеиновые кислоты могут быть связаны друг с другом с помощью любой связи между нуклеиновыми кислотами. Два основных класса связывающих групп между нуклеиновыми кислотами определяются присутствием или отсутствием атома фосфора. Типовые фосфорсодержащие связи между нуклеиновыми кислотами включают,

но не ограничены ими, фосфодиэфиры, фосфотриэфиры, метилфосфонаты, фосфорамидаты и фосфоротиоаты (P=S). Типовые не содержащие фосфор связывающие группы между нуклеиновыми кислотами включают, но не ограничены ими, метиленметилямино (-CH₂-N(CH₃)-O-CH₂-), тиодифир (-OC(O)-S-), тионокарбамат (-OC(O)(NH)-S-), силоксан (-O-Si(H)₂-O-); и N, N*-диметилгидразин (-CH₂-N(CH₃)-N(CH₃)). В некоторых вариантах осуществления, связи между нуклеиновыми кислотами, имеющие хиральный атом, могут быть получены в виде рацемической смеси, в виде отдельных энантиомеров, *например*, алкилфосфонатов и фосфоротиоатов. Нуклеиновые кислоты не природного происхождения могут содержать одну модификацию. Нуклеиновые кислоты не природного происхождения могут содержать несколько модификаций в пределах одной из групп или между различными группами.

Фосфатные модификации остова нуклеиновой кислоты включают, но не ограничены ими, метилфосфонат, фосфоротиоат, фосфорамидат (мостиковый или не мостиковый), фосфотриэфир, фосфородитиоат, фосфодитиоат и боранофосфат, и могут использоваться в любой комбинации. Другие не фосфатные связи также могут использоваться.

В некоторых вариантах осуществления, модификации остова (*например*, метилфосфонатные, фосфоротиоатные, фосфорамидатные и фосфородитиоатные межнуклеотидные связи) могут придавать модифицированной нуклеиновой кислоте иммуномодулирующую активность и/или повышать ее стабильность *in vivo*.

В некоторых случаях, производное фосфора (или модифицированная фосфатная группа) присоединено к группе сахара или аналога сахара, и может представлять собой монофосфат, дифосфат, трифосфат, алкилфосфонат, фосфоротиоат, фосфородитиоат, фосфорамидат и подобные. Примеры полинуклеотидов, содержащих модифицированные фосфатные связи или не фосфатные связи можно найти у Peyrottes et al., 1996, *Nucleic Acids Res.* 24: 1841-1848; Chaturvedi et al., 1996, *Nucleic Acids Res.* 24:2318-2323; и Schultz et al., (1996) *Nucleic Acids Res.* 24:2966-2973; Matteucci, 1997, "Oligonucleotide Analogs: an Overview" in *Oligonucleotides as Therapeutic Agents*, (Chadwick and Cardew, ed.) John Wiley and Sons, New York, NY; Zon, 1993, "Oligonucleoside Phosphorothioates" in *Protocols for Oligonucleotides and Analogs, Synthesis and Properties*, Humana Press, pp. 165-190; Miller et al., 1971, *JACS* 93:6657-6665; Jager et al., 1988, *Biochem.* 27:7247-7246; Nelson et al., 1997, *JOC* 62:7278-7287; патент США № 5,453,496; и Micklefield, 2001, *Curr. Med. Chem.* 8: 1157-1179.

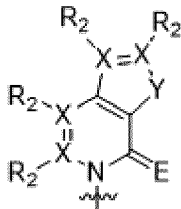
В некоторых случаях, модификация остова включает замену фосфодиэфирной связи альтернативной группой, такой как анионная, нейтральная или катионная группа. Примеры таких модификаций включают: анионную межнуклеозидную связь, N3'-P5' фосфорамидатную модификацию; боранофосфатную ДНК; проолигонуклеотиды; нейтральные межнуклеозидные связи такие как метилфосфонаты; ДНК, связанную амидом; метиленовые (метилямино) связи; формацетальные и тиоформацетальные связи; остовы, содержащие сульфонильные группы; олигонуклеотиды морфолино; пептидные нуклеиновые кислоты (ПНК); и положительно заряженные олигонуклеотиды дезоксирибонуклеинового гуанидина (DNG) (Micklefield, 2001, *Current Medicinal Chemistry*

8: 1157-1179). Модифицированная нуклеиновая кислота может содержать химерный или смешанный остов, содержащий одну или несколько модификаций, *например*, комбинацию фосфатных связей, такую как комбинация фосфодиэфирных и тиофосфатных связей.

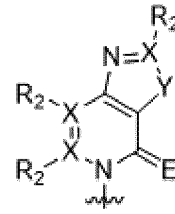
Заменители фосфата включают, например, короткоцепочечные алкильные или циклоалкильные межнуклеозидные связи, смешанные гетероатомные и алкильные или циклоалкильные межнуклеозидные связи или одну или несколько короткоцепочечных гетероатомных или гетероциклических межнуклеозидных связей. Они включают такие, которые имеют морфолино связи (образованные частично из сахарной части нуклеозида); силоксановые остовы; сульфидные, сульфоксидные и сульфоновые остовы; формацетильные и тиоформацетильные остовы; метиленформацетильные и тиоформацетильные остовы; алкенсодержащие остовы; сульфаматные остовы; метиленимино и метиленгидразино остовы; сульфонатные и сульфонамидные остовы; амидные остовы; и другие, имеющие смешанные N, O, S и CH₂ части компонентов. Многочисленные патенты США описывают как производить и использовать эти типы заменителей фосфатов, и включают, но не ограничены ими, патенты США 5,034,506; 5,166,315; 5,185,444; 5,214,134; 5,216,141; 5,235,033; 5,264,562; 5,264,564; 5,405,938; 5,434,257; 5,466,677; 5,470,967; 5,489,677; 5,541,307; 5,561,225; 5,596,086; 5,602,240; 5,610,289; 5,602,240; 5,608,046; 5,610,289; 5,618,704; 5,623,070; 5,663,312; 5,633,360; 5,677,437; и 5,677,439. Также необходимо понимать, что при замещении нуклеотида и сахар, и фосфатные группы нуклеотида может быть замещены, например, связью амидного типа (аминоэтилглицин) (ПНК). В патентах США №№ 5,539,082; 5,714,331 и 5,719,262 описано, как получать и использовать молекулы ПНК, каждый из которых включен в настоящий документ в качестве ссылки. См. также Nielsen et al., Science, 1991, 254, 1497-1500. Также можно связать другие типы молекул (конъюгаты) с нуклеотидами или аналогами нуклеотидов для усиления, например, клеточного поглощения. Конъюгаты могут быть химически связаны с нуклеотидом или аналогами нуклеотидов. Такие конъюгаты включают, но не ограничены ими, липидные группы, такие как холестериновая группа (Letsinger et al., Proc. Natl. Acad. Sci. USA, 1989, 86, 6553-6556), холевую кислоту (Manoharan et al., Bioorg. Med. Chem. Lett., 1994, 4, 1053-1060), тиозфир, *например*, гексил-S-тримилтиол (Manoharan et al., Ann. KY. Acad. Sci., 1992, 660, 306-309; Manoharan et al., Bioorg. Med. Chem. Lett., 1993, 3, 2765-2770), тиохолестерин (Oberhauser et al., Nucl. Acids Res., 1992, 20, 533-538), алифатическую цепь, *например*, додекандиольные или ундецильные остатки (Saison-Behmoaras et al., EMBOJ, 1991, 10, 1111-1118; Kabanov et al., FEBS Lett., 1990, 259, 327-330; Svinarchuk et al., Biochimie, 1993, 75, 49-54), фосфолипид, *например*, ди-гексадецил-рац-глицерин или триэтиламмоний 1-ди-О-гексадецил-рац-глицеро-SH-фосфонат (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654; Shea et al., Nucl. Acids Res., 1990, 18, 3777-3783), цепь полиамина или полиэтиленгликоля (Manoharan et al., Nucleosides & Nucleotides, 1995, 14, 969-973), или адамантануксусную кислоту (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654), пальмитиновую группу (Mishra et al., Biochem. Biophys. Acta, 1995, 1264, 229-237) или октадециламин- или

гексиламинокарбонил-оксихолестериновую группу (Crooke et al., J. Pharmacol. Exp. Ther., 1996, 277, 923-937). Многочисленные патенты США описывают получение таких конъюгатов и включают, но не ограничены ими, патенты США №№ 4,828,979; 4,948,882; 5,218,105; 5,525,465; 5,541,313; 5,545,730; 5,552,538; 5,578,717; 5,580,731; 5,580,731; 5,591,584; 5,109,124; 5,118,802; 5,138,045; 5,414,077; 5,486,603; 5,512,439; 5,578,718; 5,608,046; 4,587,044; 4,605,735; 4,667,025; 4,762,779; 4,789,737; 4,824,941; 4,835,263; 4,876,335; 4,904,582; 4,958,013; 5,082,830; 5,112,963; 5,214,136; 5,082,830; 5,112,963; 5,214,136; 5,245,022; 5,254,469; 5,258,506; 5,262,536; 5,272,250; 5,292,873; 5,317,098; 5,371,241; 5,391,723; 5,416,203; 5,451,463; 5,510,475; 5,512,667; 5,514,785; 5,565,552; 5,567,810; 5,574,142; 5,585,481; 5,587,371; 5,595,726; 5,597,696; 5,599,923; 5,599,928 и 5,688,941.

В настоящем документе описаны нуклеотидные основания, используемые в композициях и способах репликации, транскрипции, трансляции и включения аминокислот не природного происхождения в белки. В некоторых вариантах осуществления, нуклеотидное основание, описанное в настоящем документе, имеет структуру:



, где каждый X независимо представляет собой углерод или азот; R₂ является необязательным, и когда присутствует, независимо представляет собой водород, алкил, алкенил, алкинил, метокси, метантиол, метанселено, галоген, циано или азидную группу; где каждый Y независимо представляет собой серу, кислород, селен или вторичный амин; где каждый E независимо представляет собой кислород, серу или селен; и где волнистая линия указывает на точку связывания с рибозильной, дезоксирибозильной или дидезоксирибозильной группой или их аналогом, где рибозильная, дезоксирибозильная или дидезоксирибозильная группа или их аналог находятся в свободной форме, соединены с монофосфатной, дифосфатной или трифосфатной группой, необязательно содержащей α-тиотрифосфатную, β-тиотрифосфатную или γ-тиотрифосфатную группу, или включенную в РНК или ДНК, или в аналог РНК, или аналог ДНК. В некоторых вариантах осуществления, R₂ представляет собой низший алкил (например, C₁-C₆), водород или галоген. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, R₂ представляет собой фтор. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, X представляет собой углерод. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, E представляет собой серу. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, Y представляет собой серу. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем



документе, нуклеотидное основание имеет структуру: . В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, E представляет собой серу, и Y представляет собой серу. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, волнистая линия указывает точку связывания с рибозильной или дезоксирибозильной группой. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, волнистая линия указывает точку связывания с рибозильной или дезоксирибозильной группой, соединенной с трифосфатной группой. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, нуклеотидное основание представляет собой компонент тРНК. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, нуклеотидное основание представляет собой компонент антикодона в тРНК. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, нуклеотидное основание представляет собой компонент иРНК. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, нуклеотидное основание является компонентом кодона иРНК. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, нуклеотидное основание представляет собой компонент РНК или ДНК. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, нуклеотидное основание представляет собой компонент кодона ДНК. В некоторых вариантах осуществления нуклеотидного основания, описанного в настоящем документе, нуклеотидное основание образует пару нуклеотидных оснований с другим комплементарным нуклеотидным основанием.

Свойства спаривания оснований нуклеиновой кислоты

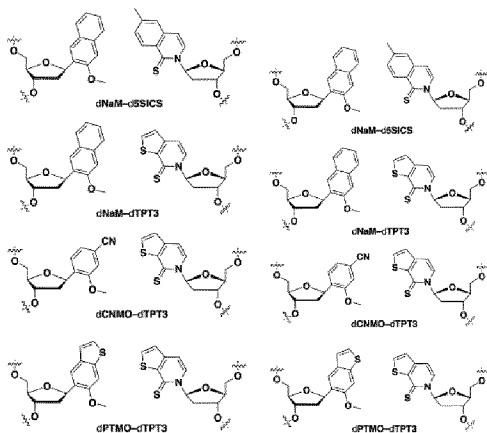
В некоторых вариантах осуществления, нуклеотид не природного происхождения образует пару оснований (пару оснований не природного происхождения; УВР) с другим нуклеотидом не природного происхождения, например, во время или после введения ДНК или РНК. В некоторых вариантах осуществления, стабильно интегрированная нуклеиновая кислота не природного происхождения представляет собой нуклеиновую кислоту не природного происхождения, которая может образовывать пару оснований с другой нуклеиновой кислотой, например, нуклеиновой кислотой природного или не природного происхождения. В некоторых вариантах осуществления, стабильно интегрированная нуклеиновая кислота не природного происхождения представляет собой нуклеиновую

кислоту не природного происхождения, которая может образовывать пару оснований с другой нуклеиновой кислотой не природного происхождения (парой оснований нуклеиновой кислоты не природного происхождения (UBP)). Например, первая нуклеиновая кислота не природного происхождения может образовывать пару оснований с второй нуклеиновой кислотой не природного происхождения. Например, одна пара нуклеозидтрифосфатов не природного происхождения, которые могут образовывать пары оснований во время и после введения в нуклеиновые кислоты, включает трифосфат (d)5SICS ((d)5SICSTP) и трифосфат (d)NaM ((d)NaMTP). Другие примеры включают, но не ограничены ими: трифосфат (d)CNMO ((d)CNMOTP) и трифосфат (d)TPT3 ((d)TPT3TP). Такие нуклеотиды не природного происхождения могут иметь группу сахара рибозы или дезоксирибозы (обозначается «(d)»). Например, одна пара нуклеозидтрифосфатов не природного происхождения, которые могут образовывать пары оснований при включении в нуклеиновые кислоты, включает трифосфат TAT1 (TAT1TP) и трифосфат NaM (NaMTP). В некоторых вариантах осуществления, одна пара нуклеозидтрифосфатов не природного происхождения, которые могут образовывать пары оснований при включении в нуклеиновые кислоты, включает трифосфат dCNMO (dCNMOTP) и трифосфат TAT1 (TAT1TP). В некоторых вариантах осуществления, одна пара нуклеозидтрифосфатов не природного происхождения, которые могут образовывать пары оснований при включении в нуклеиновые кислоты, включает трифосфат (d)TPT3 ((d)TPT3TP) и трифосфат NaM (NaMTP). В некоторых вариантах осуществления, нуклеиновая кислота не природного происхождения по существу не образует пару оснований с природной нуклеиновой кислотой (A, T, G, C). В некоторых вариантах осуществления, нуклеиновая кислота не природного происхождения может образовывать пару оснований с природной нуклеиновой кислотой.

В некоторых вариантах осуществления, стабильно интегрированный (дезокси)рибонуклеотид не природного происхождения представляет собой (дезокси)рибонуклеотид не природного происхождения, который может образовывать UBP, но по существу не образует пару оснований с каждым из природных (дезокси)рибонуклеотидов. В некоторых вариантах осуществления, стабильно интегрированный (дезокси)рибонуклеотид не природного происхождения представляет собой (дезокси)рибонуклеотид не природного происхождения, который может образовывать UBP, но по существу не образует пару оснований с одной или несколькими природными нуклеиновыми кислотами. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с A, T и C, но может образовывать пару оснований с G. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с A, T и G, но может образовывать пару оснований с C. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с C, G и A, но может образовывать пару оснований с T. Например, стабильно интегрированная нуклеиновая

кислота не природного происхождения может по существу не образовывать пару оснований с С, G и T, но может образовывать пару оснований с А. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с А и T, но может образовывать пару оснований с С и G. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с А и С, но может образовывать пару оснований с T и G. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с А и G, но может образовывать пару оснований с С и T. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с С и T, но может образовывать пару оснований с А и G. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с С и G, но может образовывать пару оснований с T и G. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с T и G, но может образовывать пару оснований с А и G. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с G, но может образовывать пару оснований с А, T и С. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с А, но может образовывать пару оснований с G, T и С. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с T, но может образовывать пару оснований с G, А и С. Например, стабильно интегрированная нуклеиновая кислота не природного происхождения может по существу не образовывать пару оснований с С, но может образовывать пара оснований с G, T и А.

Примеры нуклеотидов не природного происхождения, способных образовывать пару оснований не природного происхождения (UBP) ДНК или РНК в условиях *in vivo* включают, но не ограничены ими, 5SICS, d5SICS, NaM, dNaM, dTPT3, dMTMO, dCNMO, TAT1 и их комбинации. В некоторых вариантах осуществления, пары нуклеотидных оснований не природного происхождения включают, но не ограничены ими:



Сконструированные организмы

В некоторых вариантах осуществления, способы и плазмиды, описанные в настоящем документе, дополнительно используются для создания сконструированных организмов, например организма, который включает и реплицирует нуклеотид не природного происхождения или пару оснований нуклеиновой кислоты не природного происхождения (UBP), и может также использовать нуклеиновую кислоту, содержащую нуклеотид не природного происхождения для транскрипции мРНК и тРНК, которые используются для трансляции полипептидов не природного происхождения или белков не природного происхождения, содержащих, по меньшей мере, один остаток аминокислоты не природного происхождения. В некоторых случаях, аминокислотный остаток не природного происхождения вводится в полипептид не природного происхождения или белок не природного происхождения сайт-специфическим образом. В некоторых случаях, организм представляет собой не человеческий полусинтетический организм (SSO). В некоторых случаях, организм представляет собой полусинтетический организм (SSO). В некоторых случаях, SSO представляет собой клетку. В некоторых случаях, *in vivo* способы включают полусинтетический организм (SSO). В некоторых случаях, полусинтетический организм содержит микроорганизм. В некоторых случаях, организм включает бактерию. В некоторых случаях, организм содержит грамотрицательную бактерию. В некоторых случаях, организм содержит грамположительную бактерию. В некоторых случаях, организм содержит *Escherichia coli*. Такие модифицированные организмы по-разному содержат дополнительные компоненты, такие как механизм репарации ДНК, модифицированные полимеразы, транспортеры нуклеотидов или другие компоненты. В некоторых случаях SSO содержит штамм *E. coli* YZ3. В некоторых случаях, SSO содержит штамм *E. coli* ML1 или ML2, например штаммы, описанные на фигуре 1 (B-D) из Ledbetter, et al. *J. Am Chem. Soc.* 2018, 140(2), 758. В некоторых случаях, SSO представляет собой клеточную линию. В некоторых случаях, клеточная линия представляет собой иммортализованную клеточную линию. В некоторых случаях, клеточная линия содержит первичные клетки. В некоторых случаях, клеточная линия включает стволовые клетки. В некоторых случаях, SSO представляет собой органоид.

В некоторых случаях, используемая клетка генетически трансформируется кассетой экспрессии, кодирующей гетерологичный белок, например транспортер нуклеозидтрифосфата, способный транспортировать нуклеозидтрифосфаты не природного происхождения в клетку, и, необязательно, системой CRISPR/Cas9 для устранения ДНК, которая потеряла нуклеотид не природного происхождения (например, штамм *E. coli* YZ3, ML1 или ML2). В некоторых случаях, клетки дополнительно проявляют повышенную активность в отношении поглощения нуклеиновой кислоты не природного происхождения. В некоторых случаях, клетки дополнительно проявляют повышенную активность в отношении импорта нуклеиновой кислоты не природного происхождения.

В некоторых вариантах осуществления, Cas9 и соответствующая направляющая РНК (енРНК) кодируются на отдельных плаزمидях. В некоторых случаях, Cas9 и енРНК

кодируются на одной и той же плазмиде. В некоторых случаях, молекула нуклеиновой кислоты, кодирующая Cas9, енРНК или молекулу нуклеиновой кислоты, содержащую нуклеотид не природного происхождения, расположена на одной или нескольких плаزمидах. В некоторых случаях, Cas9 кодируется на первой плазмиде, и енРНК и молекула нуклеиновой кислоты, содержащая нуклеотид не природного происхождения, кодируются на второй плазмиде. В некоторых случаях, Cas9, енРНК и молекула нуклеиновой кислоты, содержащая нуклеотид не природного происхождения, кодируется на одной и той же плазмиде. В некоторых случаях, молекула нуклеиновой кислоты содержит два или несколько нуклеотидов не природного происхождения. В некоторых случаях, Cas9 встраивается в геном организма-хозяина, и енРНК кодируются на плазмиде или в геноме организма.

В некоторых случаях, в сконструированный микроорганизм вводят первую плазмиду, кодирующую Cas9 и енРНК, и вторую плазмиду, кодирующую молекулу нуклеиновой кислоты, содержащую нуклеотид не природного происхождения. В некоторых случаях, первую плазмиду, кодирующую Cas9, и вторую плазмиду, кодирующую енРНК, и молекулу нуклеиновой кислоты, содержащую нуклеотид не природного происхождения, вводят в сконструированный микроорганизм. В некоторых случаях, в сконструированный микроорганизм вводят плазмиду, кодирующую Cas9, енРНК и молекулу нуклеиновой кислоты, содержащую нуклеотид не природного происхождения. В некоторых случаях, молекула нуклеиновой кислоты содержит два или несколько нуклеотидов не природного происхождения.

В некоторых вариантах осуществления, создается живая клетка, которая включает в своей ДНК (плазмиде или геноме), по меньшей мере, одну молекулу нуклеиновой кислоты не природного происхождения, содержащую, по меньшей мере, одну пару оснований (УВР) не природного происхождения. В некоторых случаях, по меньшей мере, одна молекула нуклеиновой кислоты не природного происхождения содержит одну, две, три, четыре или более УВР. В некоторых случаях, по меньшей мере, одна молекула нуклеиновой кислоты не природного происхождения представляет собой плазмиду. В некоторых случаях, по меньшей мере, одна молекула нуклеиновой кислоты не природного происхождения интегрирована в геном клетки. В некоторых вариантах осуществления, по меньшей мере, одна молекула нуклеиновой кислоты не природного происхождения кодирует полипептид не природного происхождения или белок не природного происхождения. В некоторых случаях, по меньшей мере, одна молекула нуклеиновой кислоты не природного происхождения транскрибируется с образованием кодона не природного происхождения мРНК и антикодона не природного происхождения тРНК. В некоторых вариантах осуществления, по меньшей мере, одна молекула нуклеиновой кислоты не природного происхождения представляет собой молекулу ДНК не природного происхождения.

В некоторых случаях, пара оснований не природного происхождения включает пару взаимно спаривающихся нуклеотидов не природного происхождения, способных образовывать пары не природного происхождения в условиях *in vivo*, когда взаимно

спаренные нуклеотиды не природного происхождения в виде соответствующих им трифосфатов переносятся в клетку под действием транспортера нуклеотидтрифосфата. Клетка может быть генетически трансформирована с помощью кассеты экспрессии, кодирующей транспортер нуклеотидтрифосфата, так что транспортер нуклеотидтрифосфата экспрессируется и доступен для транспорта нуклеотидов не природного происхождения в клетку. Клетка может быть прокариотической или эукариотической клеткой, и пара нуклеотидов не природного происхождения со спаривающимися между собой основаниями, в виде их соответствующих трифосфатов, может быть трифосфат dTPT3 (dTPT3TP) и трифосфат dNaM (dNaMTP) или dCNMO (dCNMOTP).

В некоторых вариантах осуществления, клетки представляют собой генетически трансформированные клетки с нуклеиновой кислотой, например, кассетой экспрессии, кодирующей транспортер нуклеотидтрифосфата, способный транспортировать такие нуклеотиды не природного происхождения в клетку. Клетка может содержать транспортер гетерологичного нуклеозидтрифосфата, где транспортер гетерологичного нуклеозидтрифосфата может транспортировать нуклеозидтрифосфаты природного и не природного происхождения в клетку.

В некоторых случаях, описанные здесь способы также включают контактирование генетически трансформированной клетки с соответствующими трифосфатами, в присутствии фосфата калия и/или ингибитора фосфатаз или нуклеотидаз. Во время или после такого контакта, клетка может быть сохраняться в помещена в жизнеобеспечивающую среду, подходящую для роста и репликации клетки. Клетку можно сохранять в жизнеобеспечивающей среде таким образом, чтобы соответствующие трифосфатные формы нуклеотидов не природного происхождения включались в нуклеиновые кислоты внутри клеток, и через, по меньшей мере, один цикл репликации клетки. Пара взаимно спаренных оснований нуклеотидов не природного происхождения в виде соответствующего трифосфата может включать трифосфат dTPT3 или (dTPT3TP) и трифосфат dCNMO или dNaM (dCNOM или dNaMTP), клетка может быть *E. coli*, и dTPT3TP и dNaMTP может быть импортирован в *E. coli* с помощью транспортера *PtNTT2*, где полимеразы *E. coli*, такая как Pol III или Pol II, может использовать трифосфаты не природного происхождения для репликации ДНК, содержащей UBP, тем самым включая нуклеотиды не природного происхождения и/или пары оснований не природного происхождения в клеточные нуклеиновые кислоты в клеточной среде. Кроме того, рибонуклеотиды, такие как NaMTP и TAT1TP, 5FMTP и TPT3TP, в некоторых случаях импортируются в *E. coli* транспортером *PtNTT2*. В некоторых случаях, *PtNTT2* для импорта рибонуклеотидов представляет собой усеченный *PtNTT2*, где усеченный *PtNTT2* имеет аминокислотную последовательность, которая на, по меньшей мере, 60%, по меньшей мере, 65%, по меньшей мере, 70%, по меньшей мере, 75%, по меньшей мере, 80%, по меньшей мере, 85% или, по меньшей мере, 90% идентична аминокислотной последовательности не усеченного *PtNTT2*. Пример не усеченного *PtNTT2* (номер доступа NCBI EEC49227.1, GI:

217409295) имеет аминокислотную последовательность (SEQ ID NO: 1):

1 MRPYPTIALI SVFLSAATRI SATSSHQASA LPVKKGTHVP
 41 DSPKLSKLYI MAKTKSVSSS FDPFRGGSTV APTTPLATGG
 81 ALRKVRQAVF PIYGNQEVTK FLLIGSIKFF IILALTLTRD
 121 TKDTLIVTQC GAEAI AFLKI YGVLPAATAF IALYSKMSNA
 161 MGKKMLFYST CIPFFTFGL FDVFIYPNAE RLHPSLEAVQ
 201 AILPGGAASG GMAVLAKIAT HWTSALFYVM AEIYSSVSVG
 241 LLFWQFANDV VNVDQAKRFY PLFAQMSGLA PVLAGQYVVR
 281 FASKAVNFEA SMHRLTAAVT FAGIMICIFY QLSSSYVERT
 321 ESAKPAADNE QSIKPKKKKP KMSMVESGKF LASSQYLRLI
 361 AMLVLGYGLS INFTEIMWKS LVKKQYPDPL DYQRFMGNFS
 401 SAVGLSTCIV IFFGVHVIRL LGWKVGALAT PGIMAILALP
 441 FFACILLGLD SPARLEIAVI FGTIQSLLSK TSKYALFDPT
 481 TQMAYIPLDD ESKVKGKAAI DVLGSRIGKS GGSLIQQGLV
 521 FVFGNIINAA PVVGVVYYSV LVAWMSAAGR LSGLFQAQTE
 561 MDKADKMEAK TNKEK

В настоящем документе описаны композиции и способы, включающие использование трех или нескольких нуклеотидов со спариванием оснований не природного происхождения. Такие нуклеотиды со спариванием оснований в некоторых случаях попадают в клетку с помощью транспортеров нуклеотидов или с помощью стандартных способов трансформации нуклеиновых кислот, известных в данной области техники (например, электропорации, химической трансформации или других способов). В некоторых случаях, нуклеотид со спариванием оснований не природного происхождения проникает в клетку как часть полинуклеотида, такого как плазида. Один или несколько нуклеотидов со спариванием оснований не природного происхождения, которые могут попадать в клетку как часть полинуклеотида (РНК или ДНК), сами по себе не обязательно реплицируются *in vivo*. Например, двухцепочечная ДНК-плазида или другая нуклеиновая кислота, содержащая первый дезоксирибонуклеотид не природного происхождения и второй дезоксирибонуклеотид не природного происхождения с основаниями, сконфигурированными для образования первой пары оснований не природного происхождения, электропорировать в клетку. Клеточную среду обрабатывают третьим дезоксирибонуклеотидом не природного происхождения, четвертым дезоксирибонуклеотидом не природного происхождения с основаниями, сконфигурированными для образования второй пары оснований не природного происхождения друг с другом, где первое основание не природного происхождения дезоксирибонуклеотида и третье основание дезоксирибонуклеотида не природного происхождения образуют вторую пару оснований не природного происхождения, и где второе основание дезоксирибонуклеотида не природного происхождения и четвертое основание дезоксирибонуклеотида не природного происхождения образуют третью пару оснований не природного происхождения. В некоторых случаях, репликация *in vivo*

первоначально трансформированной двухцепочечной ДНК плазмиды приводит к последующим реплицированным плазидам, содержащим третий дезоксирибонуклеотид не природного происхождения и четвертый дезоксирибонуклеотид не природного происхождения. Альтернативно или в комбинации, рибонуклеотидные варианты третьего дезоксирибонуклеотида не природного происхождения и четвертого дезоксирибонуклеотида не природного происхождения добавляются в клеточную среду. Эти рибонуклеотиды в некоторых случаях включаются в РНК, такие как мРНК или тРНК. В некоторых случаях, первый, второй, третий и четвертый дезоксирибонуклеотиды содержат разные основания. В некоторых случаях, первый, третий, и четвертый дезоксирибонуклеотиды содержат разные основания. В некоторых случаях, первый и третий дезоксирибонуклеотиды содержат одинаковое основание.

Применяя на практике способы по настоящему изобретению, специалист в данной области техники может получить популяцию живых и размножающихся клеток, которая содержит, по меньшей мере, один нуклеотид не природного происхождения и/или, по меньшей мере, одну пару оснований (УВР) не природного происхождения, по меньшей мере, в одной нуклеиновой кислоте, содержащейся в, по меньшей мере, некоторых из отдельных клеток, где, по меньшей мере, одна нуклеиновая кислота стабильно размножается внутри клетки, и где клетка экспрессирует транспортер нуклеотидтрифосфата, подходящий для обеспечения клеточного поглощения трифосфатных форм одного или нескольких нуклеотидов не природного происхождения при контакте с (например, выращивании в присутствии) нуклеотидов не природного происхождения в жизнеобеспечивающей среде, подходящей для роста и репликации организма.

После транспорта в клетку транспортером нуклеотидтрифосфата, нуклеотиды со спариванием оснований не природного происхождения включаются в нуклеиновые кислоты внутри клетки с помощью клеточного механизма, например, собственной ДНК и/или РНК полимеразы клетки, гетерологичной полимеразы или полимеразы, которая была получена с помощью направленной эволюции (Chen T, Romesberg FE, FEBS Lett. 2014 Jan 21;588(2):219-29; Betz K et al., J Am Chem Soc. 2013 Dec 11;135(49):18637-43). Нуклеотиды не природного происхождения могут быть включены в клеточные нуклеиновые кислоты, такие как геномная ДНК, геномная РНК, мРНК, тРНК, структурная РНК, микроРНК и автономно реплицирующиеся нуклеиновые кислоты (например, плазмиды, вирусы или векторы).

В некоторых случаях, генетически сконструированные клетки получают путем введения в клетки нуклеиновых кислот, например, гетерологичных нуклеиновых кислот. В некоторых случаях, нуклеиновые кислоты, вводимые в клетки, находятся в форме плазмиды. В некоторых случаях, нуклеиновые кислоты, вводимые в клетки, интегрируются в геном клетки. Любая клетка, описанная в настоящем документе, может быть клеткой-хозяином и может содержать вектор экспрессии. В одном варианте осуществления, клетка-хозяин представляет собой прокариотическую клетку. В другом варианте осуществления,

клетка-хозяин представляет собой *E. coli*. В некоторых вариантах осуществления, клетка содержит один или несколько гетерологичных полинуклеотидов. Реагенты нуклеиновой кислоты можно вводить в микроорганизмы с использованием различных методов. Неограничивающие примеры способов, используемых для введения гетерологичных нуклеиновых кислот в различные организмы, включают: трансформацию, трансфекцию, трансдукцию, электропорацию, ультразвуковую трансформацию, конъюгацию, бомбардировку частицами и подобные. В некоторых случаях, добавление молекулоносителей (например, соединений бис-бензоимидазолила, например, см. патент США № 5,595,899) могут увеличить поглощение ДНК клетками, хотя обычно их трудно трансформировать обычными способами. Обычные способы трансформации легко доступны для специалистов в данной области техники и могут быть найдены в Maniatis, T., E. F. Fritsch and J. Sambrook (1982) *Molecular Cloning: a Laboratory Manual*; Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

В некоторых вариантах осуществления, генетическая трансформация достигается путем прямого переноса кассеты экспрессии, включая, но не ограничиваясь ими, плазмиды, вирусные векторы, вирусные нуклеиновые кислоты, фаговые нуклеиновые кислоты, фаги, космиды и искусственные хромосомы, или путем переноса генетического материала в клетки или носители, такие как катионные липосомы. Такие способы доступны в данной области техники и легко адаптируются для использования в способах, описанных в настоящем документе. Векторы переноса могут представлять собой любую нуклеотидную конструкцию, используемую для доставки генов в клетки (например, плазмиду), или быть частью общей стратегии для доставки генов, например, в составе рекомбинантного ретровируса или аденовируса (Ram et al. *Cancer Res.* 53:83-88, (1993)). Подходящие средства для трансфекции, включая вирусные векторы, химические трансфектанты или физико-механические способы, такие как электропорация и прямая диффузия ДНК, описаны, например, у Wolff, J. A., et al., *Science*, 247, 1465-1468, (1990); and Wolff, J. A. *Nature*, 352, 815-818, (1991).

Например, ДНК, кодирующую транспортер нуклеозидтрифосфата, или полимеразную кассету экспрессии и/или вектор, можно вводить в клетку любыми способами, включая, но не ограничиваясь ими, кальций-опосредованную трансформацию, электропорацию, микроинъекцию, липофекцию, бомбардировку частицами и подобные.

В некоторых случаях, клетка содержит нуклеозидтрифосфаты не природного происхождения, вставленные в одну или несколько нуклеиновых кислот внутри клетки. Например, клетка может быть живой клеткой, способной включать, по меньшей мере, один нуклеотид не природного происхождения в ДНК или РНК, содержащиеся в клетке. Клетка также может включать, по меньшей мере, одну пару оснований не природного происхождения (UBP), содержащую пару нуклеотидов со спаренными основаниями не природного происхождения, в нуклеиновых кислотах внутри клетки в условиях *in vivo*, где нуклеотиды со спаренными основаниями не природного происхождения, например, их соответствующие трифосфаты, попадают в клетку под действием транспортера

нуклеозидтрифосфата, ген которого присутствует (например, введен) в клетку путем генетической трансформации. Например, при включении в нуклеиновую кислоту, содержащуюся в клетке, dTPT3 и dCNMO могут образовывать стабильную пару оснований не природного происхождения, которая может стабильно размножаться механизмом репликации ДНК организма, например, при выращивании в жизнеобеспечивающей среде, содержащей dTPT3TP и dCNMOTP.

В некоторых случаях, клетки способны реплицировать нуклеиновую кислоту, содержащую нуклеотид не природного происхождения. Такие способы могут включать генетическую трансформацию клетки кассетой экспрессии, кодирующей транспортер нуклеозидтрифосфата, способный транспортировать в клетку в виде соответствующего трифосфата один или несколько нуклеотидов не природного происхождения в условиях *in vivo*. Альтернативно, можно использовать клетку, которая ранее была генетически трансформирована кассетой экспрессии, которая может экспрессировать кодируемый переносчик нуклеозидтрифосфата. Способы также могут включать контакт или воздействие генетически трансформированной клетки на фосфат калия и соответствующие трифосфатные формы, по меньшей мере, одного не природного происхождения нуклеотида (например, двух взаимно спаривающихся нуклеотидов, способных образовывать пару оснований не природного происхождения (UBP)) в жизнеобеспечивающей среде, подходящей для роста и репликации клетки, и сохранен трансформированной клетки в жизнеобеспечивающей среде в присутствии соответствующих трифосфатных форм, по меньшей мере, одного нуклеотида не природного происхождения (например, двух взаимно спаривающихся нуклеотидов, способных образовывать пару оснований не природного происхождения (UBP)) в условиях *in vivo* в течение, по меньшей мере, одного цикла репликации клетки.

В некоторых вариантах осуществления, клетка содержит стабильно встроенную нуклеиновую кислоту не природного происхождения. Некоторые варианты осуществления включают клетку (например, *E. coli*), которая стабильно включает нуклеотиды, отличные от А, G, Т и С, в нуклеиновые кислоты, находящиеся в клетке. например, нуклеотиды, отличные от А, G, Т и С, могут представлять собой d5SICS, dCNMO, dNaM и/или dTPT3, которые при включении в нуклеиновые кислоты клетки, могут образовывать стабильную пару оснований не природного происхождения в нуклеиновых кислотах. В одном аспекте, нуклеотиды не природного происхождения и пары оснований не природного происхождения могут стабильно размножаться репликационным аппаратом организма, когда организм, трансформированный геном транспортера трифосфата, выращивают в жизнеобеспечивающей среде, которая включает фосфат калия и трифосфатные формы d5SICS, dNaM, dCNMO и/или dTPT3.

В некоторых случаях, клетка содержит расширенный генетический алфавит. Клетка может содержать стабильно включенную нуклеиновую кислоту не природного происхождения. В некоторых вариантах осуществления, клетка с расширенным генетическим алфавитом содержит нуклеиновую кислоту не природного происхождения,

которая содержит нуклеотид не природного происхождения, который может спариваться с другим нуклеотидом не природного происхождения. В некоторых вариантах осуществления, клетка с расширенным генетическим алфавитом содержит нуклеиновую кислоту не природного происхождения, которая связана водородной связью с другой нуклеиновой кислотой. В некоторых вариантах осуществления, клетка с расширенным генетическим алфавитом содержит нуклеиновую кислоту не природного происхождения, которая не связана водородной связью с другой нуклеиновой кислотой, с которой она спарена основаниями. В некоторых вариантах осуществления, клетка с расширенным генетическим алфавитом содержит нуклеиновую кислоту не природного происхождения, которая содержит нуклеотид не природного происхождения с нуклеотидным основанием, которое спаривается с нуклеотидным основанием или другим нуклеотидом не природного происхождения посредством гидрофобных и/или упаковочных взаимодействий. В некоторых вариантах осуществления, клетка с расширенным генетическим алфавитом содержит нуклеиновую кислоту не природного происхождения, которая образует пары оснований с другой нуклеиновой кислотой посредством взаимодействий, не относящихся к водородным связям. Клетка с расширенным генетическим алфавитом может быть клеткой, которая может копировать гомологичную нуклеиновую кислоту с образованием нуклеиновой кислоты, содержащей нуклеиновую кислоту не природного происхождения. Клеткой с расширенным генетическим алфавитом может быть клетка, содержащая основание нуклеиновой кислоты не природного происхождения, спаренное с другой нуклеиновой кислотой не природного происхождения (парой оснований нуклеиновой кислоты не природного происхождения (UBP)).

В некоторых вариантах осуществления, клетки образуют пары оснований ДНК не природного происхождения (UBP) из импортированных нуклеотидов не природного происхождения в условиях *in vivo*. В некоторых вариантах осуществления, фосфат калия и/или ингибиторы активности фосфатазы и/или нуклеотидазы могут способствовать транспорту нуклеотидов не природного происхождения. Способы включают использование клетки, экспрессирующей гетерологичный транспортер нуклеозидтрифосфатов. Когда такая клетка контактирует с одним или несколькими нуклеозидтрифосфатами, нуклеозидтрифосфаты транспортируются в клетку. Клетка может находиться в присутствии фосфата калия и/или ингибиторов фосфатаз и нуклеотидаз. Нуклеозидтрифосфаты не природного происхождения могут быть включены в нуклеиновые кислоты внутри клетки с помощью естественного механизма клетки (т.е. полимеразы) и, например, взаимных пар оснований с образованием пар оснований не природного происхождения в нуклеиновых кислотах клетки. В некоторых вариантах осуществления, UBP образуются между нуклеотидами ДНК и РНК, несущими основания не природного происхождения.

В некоторых вариантах осуществления, UBP может быть включен в клетку или популяцию клеток при воздействии трифосфатов не природного происхождения. В некоторых вариантах осуществления, UBP может быть включен в клетку или популяцию

клеток при по существу постоянном воздействии на трифосфаты не природного происхождения.

В некоторых вариантах осуществления индукция экспрессии гетерологичного гена, например, переносчика нуклеозидтрифосфата (НТТ), в клетке может приводить к замедлению роста клеток и повышенному поглощению трифосфатов не природного происхождения по сравнению с ростом и поглощением одного или нескольких трифосфатов не природного происхождения в клетках без индукции экспрессии гетерологичного гена. Поглощение по разному включает транспорт нуклеотидов в клетку, например, посредством диффузии, осмоса или действия транспортеров. В некоторых вариантах осуществления, индукция экспрессии гетерологичного гена, например, НТТ в клетке может привести к повышенному росту клеток и повышенному поглощению нуклеиновой кислоты не природного происхождения по сравнению с ростом и поглощением клетки без индукции экспрессии гетерологичного гена.

В некоторых вариантах осуществления, УВР вводится во время логарифмической фазы роста. В некоторых вариантах осуществления, УВР вводится во время не логарифмической фазы роста. В некоторых вариантах осуществления, УВР вводится во время по существу линейной фазы роста. В некоторых вариантах осуществления, УВР стабильно вводятся в клетку или популяцию клеток после роста в течение определенного периода времени. Например, УВР может стабильно вводиться в клетку или популяцию клеток после роста в течение, по меньшей мере, примерно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 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 или 50 или более дупликаций. Например, УВР может быть стабильно включен в клетку или популяцию УВР клеток после роста в течение, по меньшей мере, примерно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23 или 24 часов роста. Например, УВР может стабильно включаться в клетку или популяцию клеток после роста в течение, по меньшей мере, примерно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30 или 31 день роста. Например, УВР может стабильно включаться в клетку или популяцию клеток после роста в течение, по меньшей мере, примерно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11 или 12 месяцев роста. Например, УВР может стабильно встраиваться в клетку или популяцию клеток после роста в течение, по меньшей мере, примерно 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 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, 50 лет роста.

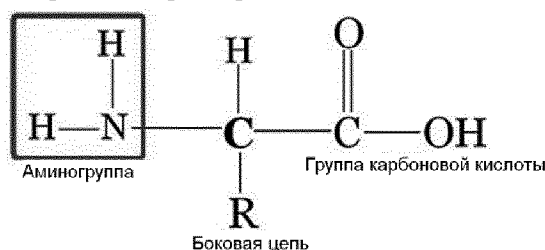
В некоторых вариантах осуществления, клетка дополнительно использует РНК полимеразу для создания мРНК, которая содержит один или несколько нуклеотидов не природного происхождения. В некоторых случаях, клетка дополнительно использует полимеразу для создания тРНК, которая содержит антикодон, содержащий один или несколько нуклеотидов не природного происхождения. В некоторых случаях, тРНК заряжена аминокислотой не природного происхождения. В некоторых случаях, антикодон не природного происхождения тРНК соединяется с кодоном не природного происхождения

мРНК во время трансляции для синтеза полипептида не природного происхождения или белка не природного происхождения, содержащего, по меньшей мере, одну аминокислоту не природного происхождения.

Аминокислоты природного и не природного происхождения

В настоящем документе, аминокислотный остаток может относиться к молекуле, содержащей как аминогруппу, так и карбоксильную группу. Подходящие аминокислоты включают, без ограничения, как D-, так и L-изомеры встречающихся в природе аминокислот, а также аминокислот не природного происхождения, полученных органическим синтезом или любыми другими способами. Термин аминокислота, используемый в настоящем документе, включает, без ограничения, α -аминокислоты, природные аминокислоты, аминокислоты не природного происхождения и аналоги аминокислот.

Термин « α -аминокислота» может относиться к молекуле, содержащей как аминогруппу, так и карбоксильную группу, связанную с углеродом, который обозначается как α -углерод. Например:



Термин « β -аминокислота» может относиться к молекуле, содержащей как аминогруппу, так и карбоксильную группу в β -конфигурации.

«Встречающаяся в природе аминокислота» может относиться к любой из двадцати аминокислот, обычно встречающихся в пептидах, синтезированных в природе и известных под однобуквенными сокращениями A, R, N, C, D, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y и V.

В следующей таблице приведены краткие сведения о свойствах природных аминокислот:

Аминокислота	3-буквенный код	1-буквенный код	Полярность боковой цепи	Заряд боковой цепи (pH 7,4)	Индекс гидропатии
Аланин	Ala	A	неполярная	нейтральный	1,8
Аргинин	Arg	R	полярная	положительный	-4,5
Аспарагин	Asn	N	полярная	нейтральный	-3,5

Аспарагиновая кислота	Asp	D	полярная	отрицательный	-3,5
Цистеин	Cys	C	полярная	нейтральный	2,5
Глутаминовая кислота	Glu	E	полярная	отрицательный	-3,5
Глютамин	Gln	Q	полярная	нейтральный	-3,5
Глицин	Gly	G	неполярная	нейтральный	-0,4
Гистидин	His	H	полярная	положительный(10%) нейтральный(90%)	-3,2
Изолейцин	Ile	I	неполярная	нейтральный	4,5
Лейцин	Leu	L	неполярная	нейтральный	3,8
Лизин	Lys	K	полярная	положительный	-3,9
Метионин	Met	M	неполярная	нейтральный	1,9
Фенилаланин	Phe	F	неполярная	нейтральный	2,8
Пролин	Pro	P	неполярная	нейтральный	-1,6
Серин	Ser	S	полярная	нейтральный	-0,8
Треонин	Thr	T	полярная	нейтральный	-0,7
Триптофан	Trp	W	неполярная	нейтральный	-0,9
Тирозин	Tyr	Y	полярная	нейтральный	-1,3
Валин	Val	V	неполярная	нейтральный	4,2

«Гидрофобные аминокислоты» включают малые гидрофобные аминокислоты и большие гидрофобные аминокислоты. «Малые гидрофобные аминокислоты» могут представлять собой глицин, аланин, пролин и их аналоги. «Большие гидрофобные

аминокислоты» могут представлять собой валин, лейцин, изолейцин, фенилаланин, метионин, триптофан и их аналоги. «Полярными аминокислотами» могут быть серин, треонин, аспарагин, глутамин, цистеин, тирозин и их аналоги. «Заряженными аминокислотами» могут быть лизин, аргинин, гистидин, аспартат, глутамат и их аналоги.

«Аналог аминокислоты» может представлять собой молекулу, которая структурно подобна аминокислоте и которая может быть заменена аминокислотой при образовании пептидомиметического макроцикла. Аналоги аминокислот включают, без ограничения, β -аминокислоты и аминокислоты, где амино- или карбоксигруппа замещена аналогично реакционноспособной группой (например, замещение первичного амина вторичным или третичным амином или замещение карбоксигруппы сложным эфиром).

Не каноническая аминокислота (ncAA) или «аминокислота не природного происхождения» может быть аминокислотой, которая не является одной из двадцати аминокислот, обычно встречающихся в пептидах, синтезированных в природе, и известна под однобуквенными сокращениями A, R, N, C, D, Q, E, G, H, I, L, K, M, F, P, S, T, W, Y и V. В некоторых случаях, аминокислоты не природного происхождения представляют собой подмножество не канонических аминокислот.

Аналоги аминокислот могут включать аналоги β -аминокислот. Примеры аналогов β -аминокислот включают, но не ограничены ими, следующие: циклические аналоги β -аминокислот; β -аланин; (R)- β -фенилаланин; (R)-1,2,3,4-тетрагидроизохинолин-3-уксусную кислоту; (R)-3-амино-4-(1-нафтил)масляную кислоту; (R)-3-амино-4-(2,4-дихлорфенил)масляную кислоту; (R)-3-амино-4-(2-хлорфенил)масляную кислоту; (R)-3-амино-4-(2-цианофенил)масляную кислоту; (R)-3-амино-4-(2-фторфенил)масляную кислоту; (R)-3-амино-4-(2-фурил)масляную кислоту; (R)-3-амино-4-(2-метилфенил)масляную кислоту; (R)-3-амино-4-(2-нафтил)масляную кислоту; (R)-3-амино-4-(2-тиенил)масляную кислоту; (R)-3-амино-4-(2-трифторметилфенил)масляную кислоту; (R)-3-амино-4-(3,4-дихлорфенил)масляную кислоту; (R)-3-амино-4-(3,4-дифторфенил)масляную кислоту; (R)-3-амино-4-(3-бензотиенил)масляную кислоту; (R)-3-амино-4-(3-хлорфенил)масляную кислоту; (R)-3-амино-4-(3-цианофенил)масляную кислоту; (R)-3-амино-4-(3-фторфенил)масляную кислоту; (R)-3-амино-4-(3-метилфенил)масляную кислоту; (R)-3-амино-4-(3-пиридил)масляную кислоту; (R)-3-амино-4-(3-тиенил)масляную кислоту; (R)-3-амино-4-(3-трифторметилфенил)масляную кислоту; (R)-3-амино-4-(4-бромфенил)масляную кислоту; (R)-3-амино-4-(4-хлорфенил)масляную кислоту; (R)-3-амино-4-(4-цианофенил)масляную кислоту; (R)-3-амино-4-(4-фторфенил)масляную кислоту; (R)-3-амино-4-(4-йодофенил)масляную кислоту; (R)-3-амино-4-(4-метилфенил)масляную кислоту; (R)-3-амино-4-(4-нитрофенил)-масляную кислоту; (R)-3-амино-4-(4-пиридил)масляную кислоту; (R)-3-амино-4-(4-трифторметилфенил)масляную кислоту; (R)-3-амино-4-пентафторфенилмасляную кислоту; (R)-3-амино-5-гексеновую кислоту; (R)-3-амино-5-гексиновую кислоту; (R)-3-амино-5-фенилпентановую кислоту; (R)-3-амино-6-фенил-5-гексеновую кислоту; (S)-1,2,3,4-тетрагидроизохинолин-3-уксусную кислоту; (S)-3-амино-4-(1-нафтил)масляную кислоту;

(S)-3-амино-4-(2,4-дихлорфенил)масляную кислоту; (S)-3-амино-4-(2-хлорфенил)масляную кислоту; (S)-3-амино-4-(2-цианофенил)масляную кислоту; (S)-3-амино-4-(2-фторфенил)масляную кислоту; (S)-3-амино-4-(2-фурил)масляную кислоту; (S)-3-амино-4-(2-метилфенил)масляную кислоту (S)-3-амино-4-(2-нафтил)масляную кислоту; (S)-3-амино-4-(2-тиенил)масляную кислоту; (S)-3-амино-4-(2-трифторметилфенил)масляную кислоту; (S)-3-амино-4-(3,4-дихлорфенил)масляную кислоту; (S)-3-амино-4-(3,4-дифторфенил)масляную кислоту, (S)-3-амино-4-(3-бензотиенил)масляную кислоту; (S)-3-амино-4-(3-хлорфенил)масляную кислоту; (S)-3-амино-4-(3-цианофенил)масляную кислоту; (S)-3-амино-4-(3-фторфенил)масляную кислоту; (S)-3-амино-4-(3-метилфенил)масляную кислоту; (S)-3-амино-4-(3-пиридил)масляную кислоту; (S)-3-амино-4-(3-тиенил)масляную кислоту; (S)-3-амино-4-(3-трифторметилфенил)масляную кислоту; (S)-3-амино-4-(4-бромфенил)масляную кислоту; (S)-3-амино-4-(4-хлорфенил)масляную кислоту; (S)-3-амино-4-(4-цианофенил)масляную кислоту; (S)-3-амино-4-(4-фторфенил)масляную кислоту; (S)-3-амино-4-(4-йодофенил)масляную кислоту; (S)-3-амино-4-(4-метилфенил)масляную кислоту; (S)-3-амино-4-(4-нитрофенил)масляную кислоту, (S)-3-амино-4-(4-пиридил)масляную кислоту; (S)-3-амино-4-(4-трифторметилфенил)масляную кислоту; (S)-3-амино-4-пентафторфенилмасляную кислоту; (S)-3-амино-5-гексеновую кислоту; (S)-3-амино-5-гексиновую кислоту; (S)-3-амино-5-фенилпентановую кислоту; (S)-3-амино-6-фенил-5-гексеновую кислоту; 1,2,5,6-тетрагидропиридин-3-карбоновую кислоту, 1,2,5,6-тетрагидропиридин-4-карбоновую кислоту, 3-амино-3-(2-хлорфенил)пропионовую кислоту, 3-амино-3-(2-тиенил)пропионовую кислоту; 3-амино-3-(3-бромфенил)пропионовую кислоту; 3-амино-3-(4-хлорфенил)пропионовую кислоту; 3-амино-3-(4-метоксифенил)пропионовую кислоту; 3-амино-4,4,4-трифтормасляную кислоту; 3-аминоадипиновую кислоту; D-β-фенилаланин; β-лейцин; L-β-гомоаланин; γ-бензиловый эфир L-β-гомоаспарагиновой кислоты; δ-бензиловый эфир L-β-гомоглутаминовой кислоты; L-β-гомоизолейцин; L-β-гомолейцин; L-β-гомометионин; L-β-гомофенилаланин; L-β-гомопролин; L-β-гомотриптофан; L-β-гомовалин; L-Nω-бензилоксикарбонил-β-гомолизин; Nω-L-β-гомоаргинин; O-бензил-L-β-гомогидроксипролин; O-бензил-L-β-гомосерин; O-бензил-L-β-гомотреонин; O-бензил-L-β-гомотирозин; γ-третил-L-β-гомоаспарагин; (R)-β-фенилаланин; γ-трет-бутиловый эфир L-β-гомоаспарагиновой кислоты; δ-трет-бутиловый эфир L-β-гомоглутаминовой кислоты; L-Nω-β-гомолизин; Nδ-третил-L-β-гомоглутамин; Nω-2,2,4,6,7-пентаметил-дигидробензофуран-5-сульфонил-L-β-гомоаргинин; O-т-бутил-L-β-гомогидроксипролин; O-т-бутил-L-β-гомосерин; O-т-бутил-L-β-гомотреонин; O-т-бутил-L-β-гомотирозин; 2-аминоциклопентанкарбоновую кислоту и 2-аминоциклогексанкарбоновую кислоту.

Аналоги аминокислот могут включать аналоги аланина, валина, глицина или лейцина. Примеры аминокислотных аналогов аланина, валина, глицина и лейцина включают, но не ограничены ими, следующие: α-метоксиглицин; α-аллил-L-аланин; α-аминоизомасляную кислоту; α-метиллейцин; β-(1-нафтил)-D-аланин; β-(1-нафтил)-L-

аланин; β -(2-нафтил)-D-аланин; β -(2-нафтил)-L-аланин; β -(2-пиридил)-D-аланин; β -(2-пиридил)-L-аланин; β -(2-тиенил)-D-аланин; β -(2-тиенил)-L-аланин; β -(3-бензотиенил)-D-аланин; β -(3-бензотиенил)-L-аланин; β -(3-пиридил)-D-аланин; β -(3-пиридил)-L-аланин; β -(4-пиридил)-D-аланин; β -(4-пиридил)-L-аланин; β -хлор-L-аланин; β -циано-L-аланин; β -циклогексил-D-аланин, β -циклогексил-L-аланин, β -циклопентен-1-илаланин, β -циклопентилаланин, β -циклопропил-L-Ala-OH.дициклогексиламмониевая соль, β -трет-бутил-D-аланин; β -трет-бутил-L-аланин; γ -аминомасляную кислоту; L- α,β -диаминопропионовую кислоту; 2,4-динитрофенилглицин; 2,5-дигидро-D-фенилглицин; 2-амино-4,4,4-трифтормасляную кислоту; 2-фторфенилглицин; 3-амино-4,4,4-трифтормасляную кислоту; 3-фторвалин; 4,4,4-трифторвалин; 4,5-дегидро-L-leu-OH.дициклогексиламмониевую соль, 4-фтор-D-фенилглицин, 4-фтор-L-фенилглицин, 4-гидрокси-D-фенилглицин, 5,5,5-трифторлейцин, 6-аминогексановую кислоту, циклопентил-D-Gly-OH.дициклогексиламмониевую соль; циклопентил-Gly-OH.дициклогексиламмониевую соль; D- α,β -диаминопропионовую кислоту; D- α -аминомасляную кислоту; D- α -трет-бутилглицин; D-(2-тиенил)глицин; D-(3-тиенил)глицин; D-2-аминокапроновую кислоту; D-2-инданилглицин; D-аллилглицин.дициклогексиламмониевую соль; D-циклогексилглицин; D-норвалин; D-фенилглицин; β -аминомасляную кислоту; β -аминоизомасляную кислоту; (2-бромфенил)глицин; (2-метоксифенил)глицин; (2-метилфенил)глицин; (2-тиазол)глицин; (2-тиенил)глицин; 2-амино-3-(диметиламино)пропионовую кислоту; L- α,β -диаминопропионовую кислоту; L- α -аминомасляную кислоту; L- α -трет-бутилглицин; L-(3-тиенил)глицин; L-2-амино-3-(диметиламино)пропионовую кислоту; дициклогексиламмониевую соль L-2-аминокапроновой кислоты; L-2-инданилглицин; дициклогексиламмониевую соль L-аллилглицина; L-циклогексилглицин; L-фенилглицин; L-пропаргилглицин; L-норвалин; N- α -аминометил-L-аланин; D- α,γ -диаминомасляную кислоту; L- α,γ -диаминомасляную кислоту; β -циклопропил-L-аланин; (N- β -(2,4-динитрофенил))-L- α,β -диаминопропионовую кислоту; (N- β -1-(4,4-диметил-2,6-диоксоциклогекс-1-илиден)этил)-D- α,β -диаминопропионовую кислоту; (N- β -1-(4,4-диметил-2,6-диоксоциклогекс-1-илиден)этил)-L- α,β -диаминопропионовую кислоту; (N- β -4-метилтретил)-L- α,β -диаминопропионовую кислоту; (N- β -аллилоксикарбонил)-L- α,β -диаминопропионовую кислоту; (N- γ -1-(4,4-диметил-2,6-диоксоциклогекс-1-илиден)этил)-D- α,γ -диаминомасляную кислоту; (N- γ -1-(4,4-диметил-2,6-диоксоциклогекс-1-илиден)этил)-L- α,γ -диаминомасляную кислоту; (N- γ -4-метилтретил)-D- α,γ -диаминомасляную кислоту; (N- γ -4-метилтретил)-L- α,γ -диаминомасляную кислоту; (N- γ -аллилоксикарбонил)-L- α,γ -диаминомасляную кислоту; D- α,γ -диаминомасляную кислоту; 4,5-дегидро-L-лейцин; циклопентил-D-Gly-OH; циклопентил-Gly-OH; D-аллилглицин, D-гомоциклогексилаланин, L-1-пиренилаланин, L-2-аминокапроновую кислоту, L-аллилглицин, L-гомоциклогексилаланин и N-(2-гидрокси-4-метокси-Bzl)-Gly-OH.

Аналоги аминокислот могут включать аналоги аргинина или лизина. Примеры аналогов аминокислот аргинина и лизина включают, но не ограничены ими, следующие:

цитруллин; L-2-амино-3-гуанидинопропионовую кислоту; L-2-амино-3-уреидопропионовую кислоту; L-цитруллин; Lys(Me)₂-OH; Lys(N₃)-OH; N δ -бензилоксикарбонил-L-орнитин; N ω -нитро-D-аргинин; N ω -нитро-L-аргинин; α -метилорнитин; 2,6-диаминогептандионовую кислоту; L-орнитин; (N δ -1-(4,4-диметил-2,6-диоксоциклогекс-1-илиден)этил)-D-орнитин; (N δ -1-(4,4-диметил-2,6-диоксоциклогекс-1-илиден)этил)-L-орнитин; (N δ -4-метилтретил)-D-орнитин; (N δ -4-метилтретил)-L-орнитин; D-орнитин; L-орнитин; Arg(Me)(Pbf)-OH; Arg(Me)₂-OH (асимметричный); Arg(Me)₂-OH (симметричный); Lys(ivDde)-OH; Lys(Me)₂-OH.HCl, Lys(Me₃)-OH хлорид, N ω -нитро-D-аргинин и N ω -нитро-L-аргинин.

Аналоги аминокислот могут включать аналоги аспарагиновой или глутаминовой кислот. Примеры аминокислотных аналогов аспарагиновой и глутаминовой кислот включают, но не ограничены ими, следующие: α -метил-D-аспарагиновую кислоту; α -метилглутаминовую кислоту; α -метил-L-аспарагиновую кислоту; γ -метилэнглутаминовую кислоту; (N- γ -этил)-L-глутамин; [N- α -(4-аминобензоил)]-L-глутаминовую кислоту; 2,6-диаминопимелиновую кислоту; L- α -аминомасляную кислоту; D-2-аминоадипиновую кислоту; D- α -аминосубериновую кислоту; α -аминопимелиновую кислоту; иминодиуксусную кислоту; L-2-аминоадипиновую кислоту; трео- β -метиласпарагиновую кислоту; γ , γ -ди-трет-бутиловый эфир γ -карбоксо-D-глутаминовой кислоты; γ , γ -ди-трет-бутиловый эфир γ -карбоксо-L-глутаминовой кислоты; Glu(OAll)-OH; L-Asu(OtBu)-OH; и пироглутаминовую кислоту.

Аналоги аминокислот могут включать аналоги цистеина и метионина. Примеры аналогов аминокислот цистеина и метионина включают, но не ограничены ими, Цис(фарнезил)-OH, Цис(фарнезил)-OMe, α -метилметионин, Cys(2-гидроксиэтил)-OH, Cys(3-аминопропил)-OH, 2-амино-4-(этилтио)масляную кислоту, бутионин, бутионинсульфоксимин, этионин, хлорид метионин метилсульфония, селенометионин, цистеиновую кислоту, [2-(4-пиридил)этил]-DL-пеницилламин, [2-(4-пиридил)этил]-L-цистеин, 4-метоксибензил-D-пеницилламин, 4-метоксибензил-L-пеницилламин, 4-метилбензил-D-пеницилламин, 4-метилбензил-L-пеницилламин, бензил-D-цистеин, бензил-L-цистеин, бензил-DL-гомоцистеин, карбамоил-L-цистеин, карбоксиэтил-L-цистеин, карбоксиметил-L-цистеин, дифенилметил-L-цистеин, этил-L-цистеин, метил-L-цистеин, трет-бутил-D-цистеин, третил-L-гомоцистеин, третил-D-пеницилламин, цистатионин, гомоцистеин, L-гомоцистеин, (2-аминоэтил)-L-цистеин, селено-L-цистин, цистатионин, Cys(StBu)-OH и ацетамидометил-D-пеницилламин.

Аналоги аминокислот могут включать аналоги фенилаланина и тирозина. Примеры аналогов аминокислот фенилаланина и тирозина включают β -метилфенилаланин, β -гидроксифенилаланин, α -метил-3-метокси-DL-фенилаланин, α -метил-D-фенилаланин, α -метил-L-фенилаланин, 1,2,3,4-тетрагидроизохинолин-3-карбоновую кислоту, 2,4-дихлорфенилаланин, 2-(трифторметил)-D-фенилаланин, 2-(трифторметил)-L-фенилаланин, 2-бром-D-фенилаланин, 2-бром-L-фенилаланин, 2-хлор-D-фенилаланин, 2-хлор-L-фенилаланин, 2-циано-D-фенилаланин, 2-циано-L-фенилаланин, 2-фтор-D-фенилаланин, 2-

фтор-L-фенилаланин, 2-метил-D-фенилаланин, 2-метил-L-фенилаланин, 2-нитро-D-фенилаланин, 2-нитро-L-фенилаланин, 2;4;5-тригидроксифенилаланин, 3,4,5-трифтор-D-фенилаланин, 3,4,5-трифтор-L-фенилаланин, 3,4-дихлор-D-фенилаланин, 3,4-дихлор-L-фенилаланин, 3,4-дифтор-D-фенилаланин, 3,4-дифтор-L-фенилаланин, 3,4-дигидрокси-L-фенилаланин, 3,4-диметокси-L-фенилаланин, 3,5,3'-трийод-L-тиронин, 3,5-дийод-D-тирозин, 3,5-дийод-L-тирозин, 3,5-дийод-L-тиронин, 3-(трифторметил)-D-фенилаланин, 3-(трифторметил)-L-фенилаланин, 3-амино-L-тирозин, 3-бром-D-фенилаланин, 3-бром-L-фенилаланин, 3-хлор-D-фенилаланин, 3-хлор-L-фенилаланин, 3-хлор-L-тирозин, 3-циано-D-фенилаланин, 3-циано-L-фенилаланин, 3-фтор-D-фенилаланин, 3-фтор-L-фенилаланин, 3-фтортирозин, 3-йод-D-фенилаланин, 3-йод-L-фенилаланин, 3-йод-L-тирозин, 3-метокси-L-тирозин, 3-метил-D-фенилаланин, 3-метил-L-фенилаланин, 3-нитро-D-фенилаланин, 3-нитро-L-фенилаланин, 3-нитро-L-тирозин, 4-(трифторметил)-D-фенилаланин, 4-(трифторметил)-L-фенилаланин, 4-амино-D-фенилаланин, 4-амино-L-фенилаланин, 4-бензоил-D-фенилаланин, 4-бензоил-L-фенилаланин, 4-бис(2-хлорэтил)амино-L-фенилаланин, 4-бром-D-фенилаланин, 4-бром-L-фенилаланин, 4-хлор-D-фенилаланин, 4-хлор-L-фенилаланин, 4-циано-D-фенилаланин, 4-циано-L-фенилаланин, 4-фтор-D-фенилаланин, 4-фтор-L-фенилаланин, 4-йод-D-фенилаланин, 4-йод-L-фенилаланин, гомофенилаланин, тироксин, 3,3-дифенилаланин, тиронин, этилтирозин и метилтирозин.

Аналоги аминокислот могут включать аналоги пролина. Примеры аналогов аминокислот пролина включают, но не ограничены ими, 3,4-дегидропролин, 4-фторпролин, цис-4-гидроксипролин, тиазолидин-2-карбоновую кислоту и транс-4-фторпролин.

Аналоги аминокислот могут включать аналоги серина и треонина. Примеры аналогов аминокислот серина и треонина включают, но не ограничены ими, 3-амино-2-гидрокси-5-метилгексановую кислоту, 2-амино-3-гидрокси-4-метилпентановую кислоту, 2-амино-3-этоксипутановую кислоту, 2-амино-3-метоксипутановую кислоту, 4-амино-3-гидрокси-6-метилгептановую кислоту, 2-амино-3-бензилоксипропионовую кислоту, 2-амино-3-бензилоксипропионовую кислоту, 2-амино-3-этоксипропионовую кислоту, 4-амино-3-гидроксибутановую кислоту и α -метилсерин.

Аналоги аминокислот могут включать аналоги триптофана. Примеры аналогов аминокислоты триптофана включают, но не ограничены ими, следующие: α -метилтриптофан; β -(3-бензотиенил)-D-аланин; β -(3-бензотиенил)-L-аланин, 1-метилтриптофан, 4-метилтриптофан, 5-бензилокситриптофан, 5-бромтриптофан, 5-хлортриптофан, 5-фтортриптофан, 5-гидрокситриптофан, 5-гидрокси-L-триптофан, 5-метокситриптофан, 5-метокси-L-триптофан, 5-метилтриптофан, 6-бромтриптофан, 6-хлор-D-триптофан, 6-хлортриптофан, 6-фтортриптофан; 6-метилтриптофан; 7-бензилокситриптофан; 7-бромтриптофан; 7-метилтриптофан; D-1,2,3,4-тетрагидроноргарман-3-карбоновую кислоту; 6-метокси-1,2,3,4-тетрагидроноргарман-1-карбоновая кислота, 7-азатриптофан, L-1,2,3,4-тетрагидроноргарман-3-карбоновую кислоту, 5-метокси-2-метилтриптофан и 6-хлор-L-триптофан.

Аналоги аминокислот могут быть рацемическими. В некоторых случаях,

используется D-изомер аналога аминокислоты. В некоторых случаях, используется L-изомер аналога аминокислоты. В некоторых случаях, аналог аминокислоты содержит хиральные центры, имеющие конфигурацию R или S. Иногда аминогруппа(ы) аналога β-аминокислоты замещена защитной группой, например, трет-бутилоксикарбонилем (группа BOC), 9-флуоренилметилоксикарбонилем (FMOC), тозилем и подобными. Иногда функциональную группу карбоновой кислоты аналога β-аминокислоты защищают, например, в виде ее сложноэфирного производного. В некоторых случаях, используют соль аналога аминокислоты.

В некоторых вариантах осуществления, аминокислота не природного происхождения представляет собой аминокислоту не природного происхождения, описанную в Liu C.C., Schultz, P.G. *Annu. Rev. Biochem.* 2010, 79, 413. В некоторых вариантах осуществления, аминокислота не природного происхождения содержит N⁶(2-азидоэтокси)карбонил-L-лизин.

В некоторых вариантах осуществления, аминокислотный остаток, описанный в настоящем документе (например, в белке), мутирован до аминокислоты не природного происхождения до связывания с конъюгирующей группой. В некоторых случаях, мутация до аминокислоты не природного происхождения предотвращает или сводит к минимуму аутоантигенный ответ иммунной системы. Используемый в настоящем документе термин «аминокислота не природного происхождения» относится к аминокислоте, отличной от 20 аминокислот, которые встречаются в белке в природе. Неограничивающие примеры аминокислот не природного происхождения включают: п-ацетил-L-фенилаланин, п-йод-L-фенилаланин, п-метоксифенилаланин, O-метил-L-тирозин, п-пропаргилоксифенилаланин, п-пропаргилфенилаланин, L-3-(2-нафтил)аланин, 3-метилфенилаланин, O-4-аллил-L-тирозин, 4-пропил-L-тирозин, три-O-ацетил-GlcNAc₆-серин, L-Dopa, фторированный фенилаланин, изопропил-L-фенилаланин, п-азидо-L-фенилаланин, п-азидо-L-фенилаланин п-азидофенилаланин, п-бензоил-L-фенилаланин, *n*-боронофенилаланин, O-пропаргилтирозин, L-фосфосерин, фосфоносерин, фосфотирозин, п-бромфенилаланин, селеноцистеин, п-амино-L-фенилаланин, изопропил-L-фенилаланин, N⁶-(пропаргилокси)карбонил-L-лизин (PrK), азидо-лизин (N⁶-азидоэтоксикарбонил-L-лизин, AzK), N⁶-(((2-азидобензил)окси)карбонил)-L-лизин, N⁶-(((3-азидобензил)окси)карбонил)-L-лизин и N⁶-(((4-азидобензил)окси)карбонил)-L-лизин, аналог аминокислоты тирозина не природного происхождения; аналог глутаминовой аминокислоты не природного происхождения; аналог фенилаланиновой аминокислоты не природного происхождения; аналог сериновой аминокислоты не природного происхождения; аналог треониновой аминокислоты не природного происхождения; алкил, арил, ацил, азидо, циано, гало, гидразин, гидразид, гидроксил, алкенил, алкинил, простой эфир, тиол, сульфонил, селено, сложный эфир, тиокислоту, борат, боронат, фосфо, фосфоно, фосфин, гетероцикл, енон, имин, альдегид, гидроксилламин, кето или аминозамещенную аминокислоту или их комбинацию; аминокислоту с фотоактивируемым сшивающим агентом, спин-меченую аминокислоту; флуоресцентную аминокислоту; металл-связывающую аминокислоту;

металлсодержащую аминокислоту; радиоактивную аминокислоту; аминокислоту с фотолabileй временной защитой и/или фотоизомеризуемую аминокислоту; аминокислоту, содержащую биотин или аналог биотина; кетосодержащую аминокислоту; аминокислоту, содержащую полиэтиленгликоль или простой полиэфир; аминокислоту, замещенную тяжелым атомом; химически расщепляемую или фоторасщепляемую аминокислоту; аминокислоту с удлиненной боковой цепью; аминокислоту, содержащую токсическую группу; замещенную сахаром аминокислоту; углерод-связанную аминокислоту, содержащую сахар; редокс-активную аминокислоту; а-гидроксисодержащую кислоту; аминотиокислоту; α,α -дизамещенную аминокислоту; β -аминокислоту; циклическую аминокислоту, отличную от пролина или гистидина, и ароматическую аминокислоту, отличную от фенилаланина, тирозина или триптофана.

В некоторых вариантах осуществления, аминокислота не природного происхождения содержит селективную реакционноспособную группу или реакционноспособную группу для сайт-селективного мечения белка-мишени или полипептида. В некоторых случаях, химия представляет собой биоортогональную реакцию (например, биосовместимые и селективные реакции). В некоторых случаях, химия представляет собой реакцию образования алкин-азидтриазола, катализируемую Cu (I) или «без меди», лигирование Штаудингера, реакцию Дильса-Альдера с обратными электронными требованиями (IEDDA), химию «фото-клик», или процесс, опосредованный металлами, такой как метатезис олефинов и перекрестное сочетание Сузуки-Мияура или Соногашира. В некоторых вариантах осуществления, аминокислота не природного происхождения содержит фотореактивную группу, которая поперечно сшивается при облучении, например, УФ. В некоторых вариантах осуществления, аминокислота не природного происхождения включает аминокислоту с фотолabileй временной защитой. В некоторых случаях, аминокислота не природного происхождения представляет собой *пара*-замещенное, *мета*-замещенное или *орто*-замещенное производное аминокислоты.

В некоторых случаях, аминокислота не природного происхождения включает п-ацетил-L-фенилаланин, п-азидометил-L-фенилаланин (pAMF), п-йодо-L-фенилаланин, О-метил-L-тирозин, п-метоксифенилаланин, п-пропаргиллоксифенилаланин, п-пропаргилфенилаланин, L-3-(2-нафтил)аланин, 3-метилфенилаланин, О-4-аллил-L-тирозин, 4-пропил-L-тирозин, три-О-ацетил-GlcNAc-серин, L-Dopa, фторированный фенилаланин, изопропил-L-фенилаланин, п-азидо-L-фенилаланин, п-ацил-L-фенилаланин, п-бензоил-L-фенилаланин, L-фосфосерин, фосфоносерин, фосфотирозин, п-бромфенилаланин, п-амино-L-фенилаланин или изопропил-L-фенилаланин.

В некоторых случаях, аминокислота не природного происхождения представляет собой 3-аминотирозин, 3-нитротирозин, 3,4-дигидроксифенилаланин или 3-йодтирозин. В некоторых случаях, аминокислота не природного происхождения представляет собой фенилселеноцистеин. В некоторых случаях, аминокислота не природного происхождения представляет собой производное фенилаланина, содержащее бензофенон, кетон, йодид, метокси, ацетил, бензоил или азид. В некоторых случаях, аминокислота не природного

происхождения представляет собой производное лизина, содержащее бензофенон, кетон, йодид, метокси, ацетил, бензоил или азид. В некоторых случаях, аминокислота не природного происхождения содержит ароматическую боковую цепь. В некоторых случаях, аминокислота не природного происхождения не содержит ароматическую боковую цепь. В некоторых случаях, аминокислота не природного происхождения содержит азидогруппу. В некоторых случаях, аминокислота не природного происхождения содержит акцепторную группу Михаэля. В некоторых случаях, акцепторные группы Михаэля содержат ненасыщенную группу, способную образовывать ковалентную связь через реакцию 1,2-присоединения. В некоторых случаях, акцепторные группы Михаэля содержат электрон-дефицитные алкины или алкины. В некоторых случаях, акцепторные группы Михаэля включают, но не ограничены ими, альфа-, бета-ненасыщенные: кетоны, альдегиды, сульфоксиды, сульфоны, нитрилы, имины или ароматические соединения. В некоторых случаях, аминокислота не природного происхождения представляет собой дегидроаланин. В некоторых случаях, аминокислота не природного происхождения содержит альдегидную или кетонную группу. В некоторых случаях, аминокислота не природного происхождения представляет собой производное лизина, содержащее альдегидную или кетонную группу. В некоторых случаях, аминокислота не природного происхождения представляет собой производное лизина, содержащее одну или несколько атомов O, N, Se или S в бета-, гамма- или дельта-положении. В некоторых случаях, аминокислота не природного происхождения представляет собой производное лизина, содержащее атомы O, N, Se или S в гамма-положении. В некоторых случаях, аминокислота не природного происхождения представляет собой производное лизина, в котором атом эпсилон N заменен атомом кислорода. В некоторых случаях, аминокислота не природного происхождения представляет собой производное лизина, которое не является посттрансляционно модифицированным не природного происхождения.

В некоторых случаях, аминокислота не природного происхождения представляет собой аминокислоту, содержащую боковую цепь, в которой шестой атом от альфа-положения содержит карбонильную группу. В некоторых случаях, аминокислота не природного происхождения представляет собой аминокислоту, содержащую боковую цепь, в которой шестой атом от альфа-положения содержит карбонильную группу, и пятый атом от альфа-положения представляет собой азот. В некоторых случаях, аминокислота не природного происхождения представляет собой аминокислоту, содержащую боковую цепь, в которой седьмой атом от альфа-положения представляет собой атом кислорода.

В некоторых случаях, аминокислота не природного происхождения представляет собой производное серина, содержащее селен. В некоторых случаях, аминокислота не природного происхождения представляет собой селеносерин (2-амино-3-гидроселенопропановую кислоту). В некоторых случаях, аминокислота не природного происхождения представляет собой 2-амино-3-((2-((3-(бензилокси)-3-оксопропил)амино)этил)селанил)пропановую кислоту. В некоторых случаях, аминокислота не природного происхождения представляет собой 2-амино-3-

(фенилселанил)пропановую кислоту. В некоторых случаях, аминокислота не природного происхождения содержит селен, где окисление селена приводит к образованию аминокислоты не природного происхождения, содержащей алкен.

В некоторых случаях, аминокислота не природного происхождения содержит циклооктинильную группу В некоторых случаях, аминокислота не природного происхождения содержит трансциклоктенильную группу В некоторых случаях, аминокислота не природного происхождения содержит норборненильную группу В некоторых случаях, аминокислота не природного происхождения содержит циклопропенил В некоторых случаях, аминокислота не природного происхождения содержит диазириновую группу В некоторых случаях, аминокислота не природного происхождения содержит тетразиновую группу.

В некоторых случаях, аминокислота не природного происхождения представляет собой производное лизина, где азот боковой цепи карбамилирован. В некоторых случаях, аминокислота не природного происхождения представляет собой производное лизина, в котором азот боковой цепи ацилирован. В некоторых случаях, аминокислота не природного происхождения представляет собой 2-амино-6-{{(трет-бутокси)карбонил]амино}гексановую кислоту. В некоторых случаях, аминокислота не природного происхождения представляет собой 2-амино-6-{{(трет-бутокси)карбонил]амино}гексановую кислоту. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-Вос-N6-метиллизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-ацетиллизин. В некоторых случаях, аминокислота не природного происхождения представляет собой пирролизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-трифторацетиллизин. В некоторых случаях, аминокислота не природного происхождения представляет собой 2-амино-6-{{(бензилокси)карбонил]амино}гексановую кислоту. В некоторых случаях, аминокислота не природного происхождения представляет собой 2-амино-6-{{(п-иодобензилокси))карбонил]амино}гексановую кислоту. В некоторых случаях, аминокислота не природного происхождения представляет собой 2-амино-6-{{(п-нитробензилокси)карбонил]амино}гексановую кислоту. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-пролилизин. В некоторых случаях, аминокислота не природного происхождения представляет собой 2-амино-6-{{(циклопентилокси)карбонил]амино}гексановую кислоту. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-(циклопентанкарбонил)лизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-(тетрагидрофуран-2-карбонил)лизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-(3-этинилтетрагидрофуран-2-карбонил)лизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-((проп-2-ин-1-илокси)карбонил)лизин. В некоторых случаях, аминокислота не природного происхождения представляет собой 2-

происхождения представляет собой цистеиниллизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-((1-(6-нитробензо[d][1,3]диоксол-5-ил)этокси)карбонил)лизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-((2-(3-метил-3H-диазирин-3-ил)этокси)карбонил)лизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-((3-(3-метил-3H-диазирин-3-ил)пропокси)карбонил)лизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-((метанитробенилокси)N6-метилкарбонил)лизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-((бицикло[6.1.0]нон-4-ин-9-илметокси)карбонил)лизин. В некоторых случаях, аминокислота не природного происхождения представляет собой N6-((циклопепт-3-ен-1-илокси)карбонил)-L-лизин.

В некоторых случаях, аминокислота не природного происхождения представляет собой 2-амино-3-((((бензилокси)карбонил)амино)метил)селанил)пропановую кислоту. В некоторых вариантах осуществления, аминокислота не природного происхождения вводится в полипептид не природного происхождения или белок не природного происхождения с помощью переориентированного стоп-кодона янтаря, опала или охры. В некоторых вариантах осуществления, аминокислота не природного происхождения вводится в полипептид не природного происхождения или белок не природного происхождения кодоном из 4 оснований. В некоторых вариантах осуществления, аминокислота не природного происхождения вводится в белок с помощью переориентированного редкого смыслового кодона.

В некоторых вариантах осуществления, аминокислота не природного происхождения вводится в полипептид не природного происхождения или белок не природного происхождения с помощью кодона не природного происхождения, содержащего нуклеотид не природного происхождения.

В некоторых случаях, включение не природного происхождения аминокислоты в белок опосредуется ортогональной модифицированной парой синтетаза/тРНК. Такие ортогональные пары содержат природную или мутированную синтетазы, которые способны заряжать не природного происхождения тРНК специфической аминокислотой не природного происхождения, часто при этом минимизируя зарядку а) других эндогенных аминокислот или альтернативных аминокислот не природного происхождения в тРНК не природного происхождения и б) любые другие (включая эндогенные) тРНК. Такие ортогональные пары включают тРНК, которые способны заряжаться синтетазой, при этом избегая зарядки другими эндогенными аминокислотами через эндогенные синтетазы. В некоторых вариантах осуществления, такие пары идентифицируют из различных организмов, таких как бактерии, дрожжи, археи или человеческие источники. В некоторых вариантах осуществления, пара ортогональная синтетаза/тРНК содержит компоненты из одного организма. В некоторых вариантах осуществления, пара ортогональная синтетаза/тРНК содержит компоненты из двух разных организмов. В некоторых вариантах

ортогональная тРНК представляет собой модифицированную тРНК цистеина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК глутамина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК глутаминовой кислоты. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированный аланин глицин. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК гистидина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК лейцина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК изолейцина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК лизина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК метионина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК фенилаланина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК пролина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК серина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК треонина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК триптофана. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК тирозина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК валина. В некоторых вариантах осуществления, ортогональная тРНК представляет собой модифицированную тРНК фосфосерина.

В некоторых вариантах осуществления, аминокислота не природного происхождения может быть включена в полипептид не природного происхождения или белок не природного происхождения с помощью пары аминоксил (aaRS или RS)-тРНК синтетаза-тРНК. Примеры пар aaRS-тРНК включают, но не ограничены ими, пары *Methanococcus jannaschii* (*Mj-Tyr*) aaRS/тРНК, *Methanococcus jannaschii* (*M. jannaschii*) TyrRS вариант *pAzFRS* (*MjpAzFRS*), пары *E. coli* TyrRS (*Ec-Tyr*)/*B. stearothermophilus* tRNA_{CUA}, пары *E. coli* LeuRS (*Ec-Leu*)/*B. stearothermophilus* tRNA_{CUA} и пары пирролизил-тРНК. В некоторых случаях, кислота не природного происхождения включается в полипептид не природного происхождения или белок не природного происхождения с помощью пары *Mj-TyrRS*/тРНК. Примеры аминокислот не природного происхождения (UAA), которые могут быть включены с помощью пары *Mj-TyrRS*/тРНК, включают, но не ограничены ими, пара-замещенные производные фенилаланина, такие как *p*-Азидо-L-Фенилаланин (*pAzF*), N6-(((2-азидобензил)окси)карбонил)-L-лизин, N6-(((3-азидобензил)окси)карбонил)-L-лизин, N6-(((4-азидобензил)окси)карбонил)-L-лизин, *n*-аминофенилаланин и *n*-метойфенилаланин; мета-замещенные производные тирозина, такие как 3-аминотирозин, 3-нитротирозин, 3,4-дигидроксифенилаланин и 3-йодтирозин;

фенилселенцистеин; *n*-боронофенилаланин; и *o*-нитробензилтирозин.

В некоторых вариантах осуществления, аминокислота не природного происхождения может быть включена в полипептид не природного происхождения или белок не природного происхождения с помощью пары *Ec-Tyr/tRNA_{CUA}* или *Ec-Leu/tRNA_{CUA}*. Примеры аминокислот не природного происхождения (UAA), которые могут быть включены с помощью пары *Ec-Tyr/tRNA_{CUA}* или *Ec-Leu/tRNA_{CUA}*, включают, но не ограничены ими, производные фенилаланина, содержащие бензофеноновые, кетоновые, йодидные или азидные заместители; *O*-пропаргилтирозин; α -аминокаприловую кислоту, *O*-метилтирозин, *O*-нитробензилцистеин; и 3-(нафталин-2-иламино)-2-аминопропановую кислоту.

В некоторых вариантах осуществления, аминокислота не природного происхождения может быть включена в полипептид не природного происхождения или белок не природного происхождения с помощью пары пирролизил-тРНК. В некоторых случаях, PylRS может быть получен из архебактериальных видов, например, из метаногенных архебактерий. В некоторых случаях, PylRS может быть получен из *Methanosarcina barkeri*, *Methanosarcina mazei* или *Methanosarcina acetivorans*. В некоторых случаях, PylRS может быть химерным PylRS. Примеры UAA, которые могут быть включены в пару пирролизил-тРНК, включают, но не ограничены ими, амид- и карбамат-замещенные лизины, такие как N⁶-(2-азидоэтокси)карбонил-L-лизин (AzK), N⁶-(((2-азидобензил)окси)карбонил)-L-лизин, N⁶-(((3-азидобензил)окси)карбонил)-L-лизин, N⁶-(((4-азидобензил)окси)карбонил)-L-лизин, 2-амино-6-((R)-тетрагидрофуран-2-карбоксамидо)гексановая кислота, *N*- ϵ -D-пролил-L-лизин и *N*- ϵ -циклопентилоксикарбонил-L-лизин, *N*- ϵ -акрилоил-L-лизин, *N*- ϵ -[(1-(6-нитробензо[d][1,3]диоксол-5-ил)этокси)карбонил]-L-лизин и *N*- ϵ -(1-метилциклопро-2-енкарбоксамидо)лизин.

В некоторых случаях, композиции и способы, описанные в настоящем документе, включают использование, по меньшей мере, двух тРНК-синтетаз для включения, по меньшей мере, двух аминокислот не природного происхождения в полипептид не природного происхождения или белок не природного происхождения. В некоторых случаях, по меньшей мере, две тРНК синтетазы могут быть одинаковыми или разными. В некоторых случаях, по меньшей мере, две аминокислоты не природного происхождения могут быть одинаковыми или разными. В некоторых случаях, по меньшей мере, две аминокислоты не природного происхождения, входящие в состав полипептида не природного происхождения, являются разными. В некоторых случаях, по меньшей мере, две разные аминокислоты не природного происхождения могут быть введены в полипептид не природного происхождения или белок не природного происхождения сайт-специфическим образом.

В некоторых случаях, аминокислота не природного происхождения может быть включена в полипептид не природного происхождения или белок не природного происхождения, описанный в настоящем документе, синтетазой, описанной в US 9,988,619 и US 9,938,516. Примеры UAA, которые могут быть включены такими синтетазами,

включают пара-метилазидо-L-фенилаланин, аралкил, гетероциклил, гетероаралкильные аминокислоты не природного происхождения и другие. В некоторых вариантах осуществления, такие UAA включают пиридил, пиразинил, пиразолил, триазолил, оксазолил, тиазолил, тиофенил или другой гетероцикл. Такие аминокислоты в некоторых примерах включают азиды, тетразины или другую химическую группу, способную конъюгироваться с партнером по связыванию, таким как водорастворимая группа. В некоторых вариантах осуществления, такие синтетазы экспрессируются и используются для включения UAA в белки *in vivo*. В некоторых вариантах осуществления, такие синтетазы используются для включения UAA в белки с использованием бесклеточной системы трансляции.

В некоторых случаях, аминокислота не природного происхождения может быть включена в полипептид не природного происхождения или белок не природного происхождения, описанный в настоящем документе, с помощью природной синтетазы. В некоторых вариантах осуществления, аминокислота не природного происхождения включена в полипептид не природного происхождения или белок не природного происхождения организмом, который является ауксотрофным для одной или нескольких аминокислот. В некоторых вариантах осуществления, синтетазы, соответствующие ауксотрофной аминокислоте, способны заряжать соответствующую тРНК аминокислотой не природного происхождения. В некоторых вариантах осуществления, аминокислота не природного происхождения представляет собой селеноцистеин или его производное. В некоторых вариантах осуществления, аминокислота не природного происхождения представляет собой селенометионин или его производное. В некоторых вариантах осуществления, аминокислота не природного происхождения представляет собой ароматическую аминокислоту, где ароматическая аминокислота содержит арилгалогенид, такой как йодид. В некоторых вариантах осуществления, аминокислота не природного происхождения структурно подобна ауксотрофной аминокислоте.

В некоторых случаях, аминокислота не природного происхождения содержит аминокислоту не природного происхождения, показанную на **фиг. 5A**.

В некоторых случаях, аминокислота не природного происхождения содержит производное или аналог лизина или фенилаланина. В некоторых случаях, аминокислота не природного происхождения содержит производное лизина или аналог лизина. В некоторых случаях, аминокислота не природного происхождения содержит пирролизин (Pyl). В некоторых случаях, аминокислота не природного происхождения содержит производное фенилаланина или аналог фенилаланина. В некоторых случаях, аминокислота не природного происхождения представляет собой аминокислоту не природного происхождения, описанную в Wan, et al., "Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool," *Biochem Biophys Acta* 1844(6): 1059-4070 (2014). В некоторых случаях, аминокислота не природного происхождения содержит аминокислоту не природного происхождения, показанную на **фиг.5B** и **фиг.5C**.

В некоторых вариантах осуществления, аминокислота не природного

происхождения содержит аминокислоту не природного происхождения, показанную на **фиг. 5D - фиг. 5G** (взято из таблицы 1 из Dumas *et al.*, *Chemical Science* 2015, 6, 50-69).

В некоторых вариантах осуществления, аминокислота не природного происхождения, включенная в белок, описанный в настоящем документе, описана в US 9,840,493; US 9,682,934; US 2017/0260137; US 9,938,516; или US 2018/0086734. Типовые UAA, которые могут быть включены такими синтетазами, включают пара-метилазидо-L-фенилаланин, аралкил, гетероциклил и гетероаралкил и производные лизина. В некоторых вариантах осуществления, такие UAA включают пиридил, пиразинил, пиразолил, триазолил, оксазолил, тиазолил, тиофенил или другой гетероцикл. Такие аминокислоты в некоторых вариантах осуществления, включают азиды, тетразины или другую химическую группу, способную конъюгироваться с партнером по связыванию, таким как водорастворимая группа. В некоторых вариантах осуществления, UAA содержит азид, присоединенный к ароматической группе через алкильный линкер. В некоторых вариантах осуществления, алкильный линкер представляет собой C₁-C₁₀ линкер. В некоторых вариантах осуществления, UAA содержит тетразин, присоединенный к ароматической группе через алкильный линкер. В некоторых вариантах осуществления, UAA содержит тетразин, присоединенный к ароматической группе через аминогруппу. В некоторых вариантах осуществления, UAA содержит тетразин, присоединенный к ароматической группе через алкиламиногруппу. В некоторых вариантах осуществления, UAA содержит азид, присоединенный к концевому атому азота (например, N6 производного лизина, или N5, N4 или N3 производного, содержащего более короткую алкильную боковую цепь) боковой цепи аминокислоты через алкильную цепь. В некоторых вариантах осуществления, UAA содержит тетразин, присоединенный к концевому атому азота боковой цепи аминокислоты через алкильную цепь. В некоторых вариантах осуществления, UAA содержит азид или тетразин, присоединенный к амиду через алкильный линкер. В некоторых вариантах осуществления, UAA представляет собой азид- или тетразин-содержащий карбамат или амид 3-аминоаланина, серина, лизина или их производного. В некоторых вариантах осуществления, такие UAA включаются в белки *in vivo*. В некоторых вариантах осуществления, такие UAA включаются в белки в бесклеточной системе.

Типы клеток

В некоторых вариантах осуществления, используются многие типы клеток/микроорганизмов, например, для трансформации или генной инженерии. В некоторых вариантах осуществления, клетка представляет собой прокариотическую или эукариотическую клетку. В некоторых случаях, клетка представляет собой микроорганизм, такой как бактериальная клетка, клетка гриба, дрожжи или одноклеточные простейшие. В некоторых случаях, клетка представляет собой эукариотическую клетку, такую как культивируемое животное, растение или клетка человека. В дополнительных случаях, клетка присутствует в организме, таком как растение или животное.

В некоторых вариантах осуществления, сконструированный микроорганизм представляет собой одноклеточный организм, часто способный к делению и пролиферации.

Микроорганизм может иметь один или несколько из следующих признаков: аэробный, анаэробный, нитчатый, не нитчатый, моноплоидный, диплоидный, ауксотрофный и/или не ауксотрофный. В некоторых вариантах осуществления, сконструированный микроорганизм представляет собой прокариотический микроорганизм (например, бактерию), и в некоторых вариантах осуществления, сконструированный микроорганизм представляет собой не прокариотический микроорганизм. В некоторых вариантах осуществления, сконструированный микроорганизм представляет собой эукариотический микроорганизм (*например*, дрожжи, грибки, амебу). В некоторых вариантах осуществления, сконструированный микроорганизм представляет собой гриб. В некоторых вариантах осуществления, сконструированный организм представляет собой дрожжи.

Любые подходящие дрожжи могут быть выбраны в качестве микроорганизма-хозяина, сконструированного микроорганизма, генетически модифицированного организма или источника гетерологичного или модифицированного полинуклеотида. Дрожжи включают, но не ограничены ими, дрожжи *Yarrowia* (*например*, *Y. lipolytica* (ранее классифицированный как *Candida lipolytica*)), дрожжи *Candida* (*например*, *C. revkaufi*, *C. viswanathii*, *C. pulcherrima*, *C. tropicalis*, *C. utilis*), дрожжи *Rhodotorula* (*например*, *R. glutinus*, *R. graminis*), дрожжи *Rhodospiridium* (*например*, *R. toruloides*), дрожжи *Saccharomyces* (*например*, *S. cerevisiae*, *S. bayanus*, *S. pastorianus*, *S. carlsbergensis*), дрожжи *Cryptococcus*, дрожжи *Trichosporon* (*например*, *T. pullans*, *T. cutaneum*), дрожжи *Pichia* (*например*, *P. pastoris*) и дрожжи *Lipomyces* (*например*, *L. starkeyii*, *L. lipoferus*). В некоторых вариантах осуществления, подходящие дрожжи относятся к роду *Arachniotus*, *Aspergillus*, *Aureobasidium*, *Auxarthron*, *Blastomyces*, *Candida*, *Chrysosporium*, *Chrysosporium*, *Debaryomyces*, *Coccidiodes*, *Cryptococcus*, *Gymnoascus*, *Hansenula*, *Histoplasma*, *Issatchenkia*, *Kluveromyces*, *Lipomyces*, *Issatchenkia*, *Microsporum*, *Мухотрихум*, *Мухозыма*, *Oidiodendron*, *Pachysolen*, *Penicillium*, *Pichia*, *Rhodospiridium*, *Rhodotorula*, *Rhodotorula*, *Saccharomyces*, *Schizosaccharomyces*, *Scopulariopsis*, *Sepedonium*, *Trichosporon* или *Yarrowia*. В некоторых вариантах осуществления, подходящие дрожжи принадлежат к видам *Arachniotus flavoluteus*, *Aspergillus flavus*, *Aspergillus fumigatus*, *Aspergillus niger*, *Aureobasidium pullulans*, *Auxarthron thaxteri*, *Blastomyces dermatitidis*, *Candida albicans*, *Candida dubliniensis*, *Candida famata*, *Candida glabrata*, *Candida guilliermondii*, *Candida kefyri*, *Candida krusei*, *Candida lambica*, *Candida lipolytica*, *Candida lusitanae*, *Candida parapsilosis*, *Candida pulcherrima*, *Candida revkaufi*, *Candida rugosa*, *Candida tropicalis*, *Candida utilis*, *Candida viswanathii*, *Candida xestobii*, *Chrysosporium keratinophilum*, *Coccidiodes immitis*, *Cryptococcus albidus* var. *diffluens*, *Cryptococcus laurentii*, *Cryptococcus neoformans*, *Debaryomyces hansenii*, *Gymnoascus dugwayensis*, *Hansenula anomala*, *Histoplasma capsulatum*, *Issatchenkia occidentalis*, *Issatchenkia orientalis*, *Kluveromyces lactis*, *Kluveromyces marxianus*, *Kluveromyces thermotolerans*, *Kluveromyces waltii*, *Lipomyces lipoferus*, *Lipomyces starkeyii*, *Microsporum gypseum*, *Мухотрихум deflexum*, *Oidiodendron echinulatum*, *Pachysolen tannophilis*, *Penicillium notatum*, *Pichia anomala*, *Pichia pastoris*, *Pichia stipitis*, *Rhodospiridium toruloides*, *Rhodotorula glutinus*, *Rhodotorula graminis*, *Saccharomyces*

cerevisiae, *Saccharomyces kluyveri*, *Schizosaccharomyces pombe*, *Scopulariopsis acremonium*, *Sepedonium chrysospermum*, *Trichosporon cutaneum*, *Trichosporon pullans*, *Yarrowia lipolytica*, or *Yarrowia lipolytica* (ранее классифицирован как *Candida lipolytica*). В некоторых вариантах осуществления дрожжами является штамм *Y. lipolytica*, который включает, но не ограничен ими, штаммы ATCC20362, ATCC8862, ATCC18944, ATCC20228, ATCC76982 и LGAM S(7)1 (Papanikolaou S., and Aggelis G., *Bioresour. Technol.* 82(1):43-9 (2002)). В некоторых вариантах осуществления, дрожжи представляют собой дрожжи *Candida* (т.е. *Candida* spp.). Любые подходящие виды *Candida* могут быть использованы и/или генетически модифицированы для получения жирной дикарбоновой кислоты (*например*, октандиикислоты, декандиикислоты, додекандиикислоты, тетрадекандиикислоты, гексадекандиикислоты, октадекандиикислоты, эйкозандиикислоты). В некоторых вариантах осуществления, подходящие виды *Candida* включают, но не ограничены ими, *Candida albicans*, *Candida dubliniensis*, *Candida famata*, *Candida glabrata*, *Candida guilliermondii*, *Candida kefyr*, *Candida krusei*, *Candida lambica*, *Candida lipolytica*, *Candida lusitanae*, *Candida parapsilosis*, *Candida pulcherrima*, *Candida revkaufi*, *Candida rugosa*, *Candida tropicalis*, *Candida utilis*, *Candida viswanathii*, *Candida xestobii* и любые другие дрожжи *Candida* spp., описанные в настоящем документе. Неограничивающие примеры штаммов *Candida* spp. включают, но не ограничены ими, штаммы sAA001 (ATCC20336), sAA002 (ATCC20913), sAA003 (ATCC20962), sAA496 (US2012/0077252), sAA106 (US2012/0077252), SU-2 (*ura3-/ura3-*), H5343 (блокирован бета-окислением; патент США № 5,648,247). Любые подходящие штаммы дрожжей *Candida* spp. можно использовать в качестве родительских штаммов для генетической модификации.

Роды, виды и штаммы дрожжей часто настолько тесно связаны по генетическому содержанию, что их бывает трудно различить, классифицировать и/или назвать. В некоторых случаях, штаммы *C. lipolytica* и *Y. lipolytica* трудно различить, классифицировать и/или назвать, и в некоторых случаях их можно рассматривать как один и тот же организм. В некоторых случаях, различные штаммы *C. tropicalis* и *C. viswanathii* могут быть трудно различимы, классифицированы и/или названы (*например*, см. Arie et.al., *J. Gen. Appl. Microbiol.*, 46, 257-262 (2000)). Некоторые штаммы *C. tropicalis* и *C. viswanathii*, полученные от ATCC, а также из других коммерческих или академических источников, можно считать эквивалентными и в равной степени подходящими для вариантов осуществления, описанных в настоящем документе. В некоторых вариантах осуществления, считается, что некоторые родительские штаммы *C. tropicalis* и *C. viswanathii* отличаются только названием.

Любой подходящий гриб может быть выбран в качестве микроорганизма-хозяина, сконструированного микроорганизма или источника гетерологичного полинуклеотида. Неограничивающие примеры грибов включают, но не ограничены ими, грибы *Aspergillus* (*например*, *A. parasiticus*, *A. nidulans*), грибы *Thraustochytrium*, грибы *Schizochytrium* и грибы *Rhizopus* (*например*, *R. arrhizus*, *R. oryzae*, *R. nigricans*). В некоторых вариантах осуществления, гриб представляет собой штамм *A. parasiticus*, который включает, но не

ограничен им, штамм ATCC24690, и в некоторых примерах, гриб представляет собой штамм *A. nidulans*, который включает, но не ограничен им, штамм ATCC38163.

Любой подходящий прокариот может быть выбран в качестве микроорганизма-хозяина, сконструированного микроорганизма или источника гетерологичного полинуклеотида. Могут быть выбраны грамотрицательные или грамположительные бактерии. Примеры бактерий включают, но не ограничены ими, бактерии *Bacillus* (например, *B. subtilis*, *B. megaterium*), бактерии *Acinetobacter*, бактерии *Nocardia*, бактерии *Xanthobacter*, бактерии *Escherichia* (например, *E. coli* (например, штаммы DH10B, Stb12, DH5-альфа, DB3, DB3.1), DB4, DB5, JDP682 и *ccdA-over* (например, заявка США № 09/518,188))), бактерии *Streptomyces*, бактерии *Erwinia*, бактерии *Klebsiella*, бактерии *Serratia* (например, *S. marcescens*), бактерии *Pseudomonas* (например, *P. aeruginosa*), бактерии *Salmonella* (например, *S. typhimurium*, *S. typhi*), бактерии *Megasphaera* (например, *Megasphaera elsdenii*). Бактерии также включают, но не ограничены ими, фотосинтезирующие бактерии (например, зеленые бессерные бактерии (например, бактерии *Chloroflexus* (например, *C. aurantiacus*), бактерии *Chloronema* (например, *C. giganteum*)), зеленые серные бактерии (например, бактерии *Chlorobium* (например, *C. limicola*), бактерии *Pelodictyon* (например, *P. luteolum*), пурпурные серные бактерии (например, бактерии *Chromatium* (например, *C. okenii*)) и пурпурные бессерные бактерии (например, бактерии *Rhodospirillum* (например, *R. rubrum*), бактерии *Rhodobacter* (например, *R. sphaeroides*, *R. capsulatus*) и бактерии *Rhodomicrobium* (например, *R. vanellii*)).

Клетки не микробных организмов можно использовать в качестве микроорганизма-хозяина, сконструированного микроорганизма или источника гетерологичного полинуклеотида. Примеры таких клеток включают, но не ограничены ими, клетки насекомых (например, *Drosophila* (например, *D. melanogaster*), *Spodoptera* (например, клетки Sf9 или Sf21 *S. frugiperda*) и *Trichoplusia* (например, клетки High-Five); клетки нематод (например, клетки *C. elegans*); клетки птиц; клетки земноводных (например, клетки *Xenopus laevis*); клетки рептилий; клетки млекопитающих (например, клетки NIH3T3, 293, CHO, COS, VERO, C127, ВНК, Per-C6, клетки меланомы Боуэса и HeLa); и клетки растений (например, *Arabidopsis thaliana*, *Nicotiana tabacum*, *Cuphea acinifolia*, *Cuphea aequipetala*, *Cuphea angustifolia*, *Cuphea appendiculata*, *Cuphea avigera*, *Cuphea avigera* var. *pulcherrima*, *Cuphea axilliflora*, *Cuphea bahiensis*, *Cuphea baillonis*, *Cuphea brachypoda*, *Cuphea bustamanta*, *Cuphea calcarata*, *Cuphea calophylla*, *Cuphea calophylla* subsp. *mesostemon*, *Cuphea carthagenensis*, *Cuphea circaeoides*, *Cuphea confertiflora*, *Cuphea cordata*, *Cuphea crassiflora*, *Cuphea cyanea*, *Cuphea decandra*, *Cuphea denticulata*, *Cuphea disperma*, *Cuphea epilobiifolia*, *Cuphea ericoides*, *Cuphea flava*, *Cuphea flavisetula*, *Cuphea fuchsiiifolia*, *Cuphea gaumeri*, *Cuphea glutinosa*, *Cuphea heterophylla*, *Cuphea hookeriana*, *Cuphea hyssopifolia* (вереск Мексиканский), *Cuphea hyssopoides*, *Cuphea ignea*, *Cuphea ingrata*, *Cuphea jorullensis*, *Cuphea lanceolata*, *Cuphea linarioides*, *Cuphea llavea*, *Cuphea lophostoma*, *Cuphea lutea*, *Cuphea lutescens*, *Cuphea melanium*, *Cuphea melvilla*, *Cuphea micrantha*, *Cuphea micropetala*, *Cuphea mimuloides*, *Cuphea nitidula*, *Cuphea palustris*, *Cuphea parsonsia*, *Cuphea pascuorum*, *Cuphea*

paucipetala, Cuphea procumbens, Cuphea pseudosilene, Cuphea pseudovaccinium, Cuphea pulchra, Cuphea racemosa, Cuphea repens, Cuphea salicifolia, Cuphea salvadorensis, Cuphea schumannii, Cuphea sessiliflora, Cuphea sessilifolia, Cuphea setosa, Cuphea spectabilis, Cuphea spermacoce, Cuphea splendida, Cuphea splendida var. viridiflava, Cuphea strigulosa, Cuphea subuligera, Cuphea teleandra, Cuphea thymoides, Cuphea tolucana, Cuphea urens, Cuphea utriculosa, Cuphea viscosissima, Cuphea watsoniana, Cuphea wrightii, Cuphea lanceolata).

Микроорганизмы или клетки, используемые в качестве организмов-хозяев или источника гетерологичного полинуклеотида, являются коммерчески доступными. Описанные в настоящем документе микроорганизмы и клетки, и другие подходящие микроорганизмы и клетки доступны, например, от Invitrogen Corporation, (Carlsbad, CA), American Type Culture Collection (Manassas, Virginia) и Agricultural Research Culture Collection (NRRL; Peoria, Illinois). Микроорганизмы-хозяева и сконструированные микроорганизмы могут быть предоставлены в любой подходящей форме. Например, такие микроорганизмы могут быть предоставлены в виде жидкой культуры или твердой культуры (*например*, среды на основе агара), которая может представлять собой первичную культуру или может быть пассирована (*например*, разведена и культивирована) один или несколько раз. Микроорганизмы также могут быть предоставлены в замороженной или сухой форме (*например*, лиофилизированными). Микроорганизмы могут быть предоставлены в любой подходящей концентрации.

Полимеразы

Особенно полезной функцией полимеразы является катализация полимеризации цепи нуклеиновой кислоты с использованием существующей нуклеиновой кислоты в качестве шаблона. Другие полезные функции описаны в настоящем документе в другом месте. Примеры полезных полимераз включают ДНК полимеразы и РНК полимеразы.

Способность улучшать специфичность, процессивность или другие характеристики полимераз нуклеиновых кислот не природного происхождения была бы крайне желательной в различных контекстах, где, например, желательно включение нуклеиновых кислот не природного происхождения, включая амплификацию, секвенирование, мечение, обнаружение, клонирование и многие другие.

В некоторых случаях, описанные раскрытые в настоящем документе включают полимеразы, которые включают нуклеиновые кислоты не природного происхождения в растущую шаблонную копию, например, во время амплификации ДНК. В некоторых вариантах осуществления, полимеразы могут быть модифицированы так, чтобы активный сайт полимеразы был модифицирован для снижения ингибирования входа нуклеиновой кислоты не природного происхождения в активный сайт. В некоторых вариантах осуществления, полимеразы могут быть модифицированы для обеспечения комплементарности с одной или несколькими характеристиками не природного происхождения нуклеиновых кислот не природного происхождения. Такие полимеразы могут быть экспрессированы или сконструированы в клетках для стабильного включения UBP в клетки. Следовательно, настоящее изобретение включает композиции, которые

включают гетерологичную или рекомбинантную полимеразу, и способы ее применения.

Полимеразы можно модифицировать с помощью способов, относящихся к белковой инженерии. Например, можно провести молекулярное моделирование на основе кристаллических структур, для идентификации локаций полимераз, где могут быть проведены мутации для изменения активности мишени. Остаток, идентифицированный как мишень для замены, может быть заменен остатком, выбранным с использованием моделирования минимизации энергии, моделирования гомологии и/или консервативных аминокислотных замен, таких как описано в Bordo, et al. *J Mol Biol* 217: 721-729 (1991) and Hayes, et al. *Proc Natl Acad Sci, USA* 99: 15926- 15931 (2002).

Любая из множества полимераз может быть использована в способах или композициях, изложенных в настоящем документе, включая, например, ферменты на основе белка, выделенные из биологических систем, и их функциональные варианты. Подразумевается, что ссылка на конкретную полимеразу, такую как приведена в качестве примера ниже, включает ее функциональные варианты, если не указано иное. В некоторых вариантах осуществления, полимеразы представляет собой полимеразу дикого типа. В некоторых вариантах осуществления, полимеразы представляет собой модифицированную или мутантную полимеразу.

Полимеразы, обладающие характеристиками для улучшения проникновения нуклеиновых кислот не природного происхождения в области активного сайта и для координации с нуклеотидами не природного происхождения в области активного сайта. В некоторых вариантах осуществления модифицированная полимеразы имеет модифицированный сайт связывания нуклеотидов.

В некоторых вариантах осуществления, модифицированная полимеразы имеет специфичность в отношении нуклеиновой кислоты не природного происхождения, которая составляет, по меньшей мере, примерно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа по отношению к нуклеиновой кислоте не природного происхождения. В некоторых вариантах осуществления, модифицированная полимеразы или полимеразы дикого типа имеет специфичность в отношении нуклеиновой кислоты не природного происхождения, которая составляет, по меньшей мере, примерно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа по отношению к природной нуклеиновой кислоте и/или нуклеиновой кислоте не природного происхождения без модифицированного сахара. В некоторых вариантах осуществления, модифицированная полимеразы или полимеразы дикого типа обладает специфичностью к нуклеиновой кислоте не природного происхождения, содержащей модифицированное основание, которая составляет, по меньшей мере, примерно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа к природной нуклеиновой кислоте и/или нуклеиновой кислоте не природного происхождения без модифицированного основания. В некоторых вариантах осуществления, модифицированная полимеразы или полимеразы дикого типа обладает

специфичностью в отношении нуклеиновой кислоты не природного происхождения, содержащей трифосфат, которая составляет, по меньшей мере, примерно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа в отношении нуклеиновой кислоты, содержащей трифосфат, и/или нуклеиновой кислоты не встречающейся в природе без трифосфата. Например, модифицированная полимеразы дикого типа может иметь специфичность в отношении нуклеиновой кислоты не природного происхождения, содержащей трифосфат, составляющую, по меньшей мере, примерно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа в отношении нуклеиновой кислоты не природного происхождения с дифосфатом или монофосфатом, или без фосфата, или их комбинации.

В некоторых вариантах осуществления, модифицированная полимеразы дикого типа имеет ослабленную специфичность в отношении нуклеиновой кислоты не природного происхождения. В некоторых вариантах осуществления, модифицированная полимеразы дикого типа обладает специфичностью к нуклеиновой кислоте не природного происхождения и специфичностью к природной нуклеиновой кислоте, которая составляет, по меньшей мере, примерно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа по отношению к природной нуклеиновой кислоте. В некоторых вариантах осуществления, модифицированная полимеразы дикого типа обладает специфичностью к нуклеиновой кислоте не природного происхождения, содержащей модифицированный сахар, и специфичностью к природной нуклеиновой кислоте, которая составляет, по меньшей мере, примерно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа по отношению к природной нуклеиновой кислоте. В некоторых вариантах осуществления, модифицированная полимеразы дикого типа обладает специфичностью к нуклеиновой кислоте не природного происхождения, содержащей модифицированное основание, и специфичностью к природной нуклеиновой кислоте, которая составляет, по меньшей мере, примерно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа по отношению к природной нуклеиновой кислоте.

Отсутствие экзонуклеазной активности может быть характеристикой дикого типа или характеристикой, приданной вариантной или сконструированной полимеразой. Например, фрагмент экзо-минуса Клену представляет собой мутированную версию фрагмента Клену, у которого отсутствует 3'-5' коррекция экзонуклеазной активности.

Способы по настоящему описанию могут быть использованы для расширения диапазона субстратов любой ДНК полимеразы, у которой отсутствует внутренняя 3-5' корректирующая экзонуклеазная активность, или у которой 3-5' корректирующая экзонуклеазная активность была выключена, например, посредством мутации. Примеры ДНК полимеразы включают polA, polB (см., например, Parrel & Loeb, Nature Struc Biol

2001), polC, polD, polY, polX и обратные транскриптазы (RT), но предпочтительно представляют собой процессивные высокоточные полимеразы (PCT/GB2004/004643). В некоторых вариантах осуществления, модифицированная полимеразы или полимеразы дикого типа по существу не имеет 3'-5' корректирующую экзонуклеазную активность. В некоторых вариантах осуществления, модифицированная полимеразы или полимеразы дикого типа по существу не имеет 3'-5' корректирующую экзонуклеазную активность для нуклеиновой кислоты не природного происхождения. В некоторых вариантах осуществления, модифицированная полимеразы или полимеразы дикого типа имеет 3'-5' корректирующую экзонуклеазную активность. В некоторых вариантах осуществления, модифицированная полимеразы или полимеразы дикого типа имеет 3'-5' корректирующую экзонуклеазную активность для природной нуклеиновой кислоты и по существу не имеет 3'-5' корректирующую экзонуклеазную активность для нуклеиновой кислоты не природного происхождения.

В некоторых вариантах осуществления, модифицированная полимеразы имеет 3'-5' корректирующую экзонуклеазную активность, которая составляет, по меньшей мере, примерно 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% корректирующей экзонуклеазной активности полимеразы дикого типа. В некоторых вариантах осуществления, модифицированная полимеразы имеет 3'-5' корректирующую экзонуклеазную активность для кислоты не природного происхождения, которая составляет, по меньшей мере, примерно 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% корректирующей экзонуклеазной активности полимеразы дикого типа к природной нуклеиновой кислоте. В некоторых вариантах осуществления, модифицированная полимеразы имеет 3'-5' корректирующую экзонуклеазную активность для нуклеиновой кислоты не природного происхождения и имеет 3'-5' корректирующую экзонуклеазную активность для природной нуклеиновой кислоты, которая составляет, по меньшей мере, примерно 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% корректирующей экзонуклеазной активности полимеразы дикого типа к природной нуклеиновой кислоте. В некоторых вариантах осуществления, модифицированная полимеразы имеет 3'-5' корректирующую экзонуклеазную активность для природной нуклеиновой кислоты, которая составляет, по меньшей мере, примерно 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% корректирующей экзонуклеазной активности полимеразы дикого типа по отношению к природной нуклеиновой кислоте.

В некоторых вариантах осуществления, полимеразы характеризуются скоростью диссоциации из нуклеиновых кислот. В некоторых вариантах осуществления, полимеразы имеет относительно низкую скорость диссоциации для одной или нескольких нуклеиновых кислот природного и не природного происхождения. В некоторых вариантах осуществления, полимеразы имеет относительно высокую скорость диссоциации для одной или нескольких нуклеиновых кислот природного и не природного происхождения. Скорость диссоциации представляет собой активность полимеразы, которую можно корректировать для настройки скорости реакции в способах, изложенных в настоящем

документе.

В некоторых вариантах осуществления, полимеразы характеризуются в соответствии с их точностью при использовании с конкретной нуклеиновой кислотой природного и/или не природного происхождения или коллекциями нуклеиновой кислоты природного и/или не природного происхождения. Точность обычно относится к точности, с которой полимеразы включает правильные нуклеиновые кислоты в растущую цепь нуклеиновой кислоты при создании копии матрицы нуклеиновой кислоты. Точность ДНК полимеразы может быть измерена как отношение правильных включений нуклеиновых кислот природного и не природного происхождения к неправильным, когда нуклеиновые кислоты природного и не природного происхождения присутствуют, например, в равных концентрациях, для конкуренции за синтез цепи в одном и том же месте в бинарном комплексе нуклеиновой кислоты полимеразы-цепь-матрица. Точность ДНК полимеразы можно рассчитать как отношение (k_{cat}/K_m) для нуклеиновой кислоты природного и не природного происхождения и (k_{cat}/K_m) для нуклеиновой кислоты природного и не природного происхождения, где k_{cat} и K_m представляет собой параметры Михаэлиса-Ментен в кинетике стационарного фермента (Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, 2nd ed., p 350, W. H. Freeman & Co., New York., включен в настоящий документ посредством ссылки). В некоторых вариантах осуществления, полимеразы имеет значение точности, по меньшей мере, примерно 100, 1000, 10000, 100000 или 1×10^6 , с или без корректирующей активности.

В некоторых вариантах осуществления, полимеразы из природных источников или их варианты подвергают скринингу с использованием анализа, который выявляет включение нуклеиновой кислоты не природного происхождения, имеющей особую структуру. В одном примере, полимеразы можно подвергать скринингу на способность включать нуклеиновую кислоту не природного происхождения или UBP; например, d5SICSTP, dCNMOTP, dTPT3TP, dNaMTP, dCNMOTP-dTPT3TP или d5SICSTP-dNaMTP UBP. Может быть использована полимеразы, например, гетерологичная полимеразы, которая проявляет модифицированное свойство для нуклеиновой кислоты не природного происхождения по сравнению с полимеразой дикого типа. Например, модифицированным свойством могут быть, например, K_m , k_{cat} , V_{max} , процессивность полимеразы в присутствии нуклеиновой кислоты не природного происхождения (или природного нуклеотида), средняя длина считывания матрицы полимеразой в присутствии нуклеиновой кислоты не природного происхождения, специфичность полимеразы к нуклеиновой кислоте не природного происхождения, скорость связывания нуклеиновой кислоты не природного происхождения, скорость высвобождения продукта (пирофосфата, трифосфата и т. д.), скорость ветвления, или любая их комбинация. В одном варианте осуществления, модифицированное свойство представляет собой пониженное значение K_m для нуклеиновой кислоты не природного происхождения и/или повышенное значение k_{cat}/K_m или V_{max}/K_m для нуклеиновой кислоты не природного происхождения. Точно так же, полимеразы необязательно имеет повышенную скорость связывания нуклеиновой кислоты

не природного происхождения, повышенную скорость высвобождения продукта и/или пониженную скорость ветвления по сравнению с полимеразой дикого типа.

В то же время, полимеразы могут включать в растущую копию нуклеиновой кислоты природные нуклеиновые кислоты, например, А, С, G и Т. Например, полимеразы необязательно проявляют специфическую активность для природной нуклеиновой кислоты, которая, по меньшей мере, примерно на 5% выше (например, 5%, 10%, 25%, 50%, 75%, 100% или выше), чем у соответствующей полимеразы дикого типа, и процессивность с природными нуклеиновыми кислотами в присутствии матрицы, которая, по меньшей мере, на 5% выше (например, 5%, 10%, 25%, 50%, 75%, 100% или выше) полимеразы дикого типа в присутствии природной нуклеиновой кислоты. Необязательно, полимеразы демонстрируют k_{cat}/K_m или V_{max}/K_m для природного нуклеотида, которые, по меньшей мере, примерно на 5% выше (например, примерно на 5%, 10%, 25%, 50%, 75% или 100% или выше) чем у полимеразы дикого типа.

Полимеразы, используемые в настоящем документе, которые могут обладать способностью включать нуклеиновую кислоту не природного происхождения определенной структуры, также могут быть получены с использованием подхода направленного развития. Анализ синтеза нуклеиновой кислоты можно использовать для скрининга вариантов полимеразы, обладающих специфичностью к любому из множества нуклеиновых кислот не природного происхождения. Например, варианты полимеразы могут быть подвергнуты скринингу на способность включать нуклеозидтрифосфат не природного происхождения напротив нуклеотида не природного происхождения в матрицу ДНК; например, dTPTЗTP напротив dCNMO, dCNMOTP напротив dTPTЗ, NaMTP напротив dTPTЗ или TAT1TP напротив dCNMO или dNaM. В некоторых вариантах осуществления, такой анализ представляет собой анализ *in vitro*, например, с использованием варианта рекомбинантной полимеразы. В некоторых вариантах осуществления, такой анализ представляет собой анализ *in vivo*, например, экспрессию варианта полимеразы в клетке. Такие методы направленного развития могут использоваться для скрининга вариантов любой подходящей полимеразы на активность в отношении любой из нуклеиновых кислот не природного происхождения, указанных в настоящем документе. В некоторых случаях, полимеразы, используемые в настоящем документе, обладают способностью включать рибонуклеотиды не природного происхождения в нуклеиновую кислоту, такую как РНК. Например, NaM или TAT1 рибонуклеотиды включены в нуклеиновые кислоты с помощью описанных в настоящем документе полимераз.

Модифицированные полимеразы описанных композиций необязательно могут представлять собой модифицированную и/или рекомбинантную ДНК полимеразу типа Ф29. Необязательно, полимеразы могут представлять собой модифицированную и/или рекомбинантную Ф29, B103, GA-1, PZA, Ф15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722 или L17 полимеразу.

Модифицированные полимеразы описанных композиций необязательно могут представлять собой модифицированную и/или рекомбинантную прокариотическую ДНК

полимеразу, например, ДНК полимеразу II (Pol II), ДНК полимеразу III (Pol III), ДНК полимеразу IV (Pol IV), ДНК полимеразу V (Pol V). В некоторых вариантах осуществления, модифицированные полимеразы включают полимеразы, которые опосредуют синтез ДНК через не обучающие поврежденные нуклеотиды. В некоторых вариантах осуществления, гены, кодирующие Pol I, Pol II (*polB*), Pol IV (*dinB*) и/или Pol V (*umuCD*), конститутивно экспрессируются или сверхэкспрессируются в сконструированной клетке, или SSO. В некоторых вариантах осуществления, увеличение экспрессии или сверхэкспрессии Pol II способствует увеличению удержания пар оснований не природного происхождения (UBP) в сконструированной клетке или SSO.

Полимеразы нуклеиновых кислот, обычно применяемые в настоящем изобретении, включают ДНК полимеразы, РНК полимеразы, обратные транскриптазы и их мутантные или измененные формы. ДНК полимеразы и их свойства подробно описаны, среди прочего, в DNA Replication 2nd edition, Kornberg and Baker, W. H. Freeman, New York, N. Y. (1991). Известные традиционные ДНК полимеразы, применяемые в настоящем изобретении, включают, но не ограничены ими, ДНК полимеразу *Pyrococcus furiosus* (Pfu) (Lundberg et al., 1991, Gene, 108: 1, Stratagene), ДНК полимеразу *Pyrococcus woesei* (Pwo) (Hinnisdaels et al., 1996, Biotechniques, 20:186-8, Boehringer Mannheim), ДНК полимеразу *Thermus thermophilus* (Tth) (Myers and Gelfand 1991, Biochemistry 30:7661), ДНК полимеразу *Bacillus stearothermophilus* (Stenesh and McGowan, 1977, Biochim Biophys Acta 475:32), ДНК полимеразу *Thermococcus litoralis* (Tli) (также называемую ДНК полимеразой VentTM, Cariello et al., 1991, Polynucleotides Res, 19: 4193, New England Biolabs), ДНК полимеразу 9 NmTM (New England Biolabs), фрагмент Стоффеля, Thermo Sequenase® (Amersham Pharmacia Biotech UK), TherminatorTM (New England Biolabs), ДНК полимеразу *Thermotoga maritima* (Tma) (Diaz and Sabino, 1998 Braz J Med. Res, 31:1239), ДНК полимеразу *Thermus aquaticus* (Taq) (Chien et al., 1976, J. Bacteriol, 127: 1550), ДНК полимеразу, ДНК полимеразу KOD *Pyrococcus kodakaraensis* (Takagi et al., 1997, Appl. Environ. Microbiol. 63:4504), ДНК полимеразу JDF-3 (из *thermococcus* sp. JDF-3, патентная заявка WO 0132887), ДНК полимеразу *Pyrococcus GB-D* (PGB-D) (также называемую ДНК полимеразой Deep VentTM, Juncosa-Ginesta et al., 1994, Biotechniques, 16:820, New England Biolabs), ДНК полимеразу U1Tma (из термофилы *Thermotoga maritima*; Diaz and Sabino, 1998 Braz J. Med. Res, 31:1239; PE Applied Biosystems), ДНК полимеразу Tgo (из *thermococcus gorgonarius*, Roche Molecular Biochemicals), ДНК полимеразу I *E. coli* (Lecomte and Doubleday, 1983, Polynucleotides Res. 11: 7505), ДНК полимеразу T7 (Nordstrom et al., 1981, J Biol. Chem. 256:3112) и ДНК полимеразу II архей DP11/DP2 (Cann et al., 1998, Proc. Natl. Acad. Sci. USA 95:14250). Рассматриваются и мезофильные полимеразы, и термофильные полимеразы. Термофильные ДНК полимеразы включают, но не ограничены ими, ThermoSequenase®, 9 NmTM, TherminatorTM, Taq, Tne, Tma, Pfu, Tfl, Tth, Tli, фрагмент Стоффеля, VentTM и Deep VentTM ДНК полимеразу, ДНК полимеразу KOD, Tgo, JDF-3 и их мутанты, варианты и производные. Также рассматривается полимеразы, которая представляет собой мутант с дефицитом 3' экзонуклеазы. Обратные транскриптазы, применимые в настоящем описании,

включают, но не ограничены ими, обратные транскриптазы из HIV, HTLV-I, HTLV-II, FeLV, FIV, SIV, AMV, MMTV, MoMuLV и других ретровирусов (см. Levin, Cell 88:5-8 (1997); Verma, Biochim Biophys Acta. 473:1-38 (1977); Wu et al, CRC Crit Rev Biochem. 3:289-347(1975)). Дополнительные примеры полимераз включают, но не ограничены ими, ДНК полимеразу 9 NTM, ДНК полимеразу Taq, ДНК полимеразу Phusion®, ДНК полимеразу Pfu, ДНК полимеразу RB69, ДНК полимеразу KOD и ДНК полимеразу VentR® Gardner et al. (2004) «Comparative Kinetics of Nucleotide Analog Incorporation by Vent DNA Polymerase (J. Biol. Chem., 279(12), 11834-11842; Gardner and Jack «Determinants of nucleotide sugar recognition in an archaeon DNA polymerase» Nucleic Acids Research, 27(12) 2545-2553). Полимеразы, выделенные из не термофильных организмов, могут инактивироваться нагреванием. Примеры включают ДНК полимеразы из фага. Следует понимать, что полимеразы из любого множества источников можно модифицировать для повышения или снижения их устойчивости к условиям высокой температуры. В некоторых вариантах осуществления, полимеразы могут быть термофильными. В некоторых вариантах осуществления, термофильная полимеразы могут быть инактивируемой нагреванием. Термофильные полимеразы обычно применимы в условиях высокой температуры или в условиях термоциклирования, таких, которые используются в методах полимеразной цепной реакции (ПЦР).

В некоторых вариантах осуществления, полимеразы содержат Ф29, B103, GA-1, PZA, Ф15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, L17, ThermoSequenase®, 9 NmTM, ДНК полимеразу TherminatorTM, Tne, Tma, Tfl, Tth, Tii, фрагмент Стоффеля, ДНК полимеразу VentTM и Deep VentTM, ДНК полимеразу KOD, Tgo, JDF-3, Pfu, Taq, ДНК полимеразу T7, РНК полимеразу T7, PGB-D, ДНК полимеразу UITma, ДНК полимеразу I *E. coli*, ДНК полимеразу III *E. coli*, ДНК полимеразу II архей DP1I/DP2, ДНК полимеразу 9 NTM, ДНК полимеразу Taq, ДНК полимеразу Phusion®, ДНК полимеразу Pfu, РНК полимеразу SP6, ДНК полимеразу RB69, обратную транскриптазу вируса миелобластома птиц (AMV), обратную транскриптазу вируса лейкоза Молони мыши (MMLV), обратную транскриптазу SuperScript® II и обратную транскриптазу SuperScript® III.

В некоторых вариантах осуществления, полимеразы представляют собой ДНК полимеразу I (или фрагмент Кленова), полимеразу Vent, ДНК полимеразу Phusion®, ДНК полимеразу KOD, полимеразу Taq, ДНК полимеразу T7, РНК полимеразу T7, ДНК полимеразу TherminatorTM, полимеразу POLB, РНК полимеразу SP6, ДНК полимеразу I *E. coli*, ДНК полимеразу III *E. coli*, обратную транскриптазу вируса миелобластома птиц (AMV), обратную транскриптазу вируса лейкоза Молони мыши (MMLV), обратную транскриптазу SuperScript® II или обратную транскриптазу SuperScript® III.

Транспортер нуклеотидов

Транспортеры нуклеотидов (NT) представляют собой группу мембранных транспортных белков, которые облегчают перенос нуклеотидных субстратов через клеточные мембраны и везикулы. В некоторых вариантах осуществления, существует два

типа NT, концентрационные транспортеры нуклеозидов и уравнивающие транспортеры нуклеозидов. В некоторых случаях, NT также включают транспортеры органических анионов (OAT) и транспортеры органических катионов (OCT). В некоторых случаях, транспортер нуклеотидов представляет собой транспортер нуклеозидтрифосфата (NTT).

В некоторых вариантах осуществления, транспортер нуклеозидтрифосфата (NTT) происходит из бактерий, растений или водорослей. В некоторых вариантах осуществления, транспортер нуклеотида нуклеозидтрифосфата представляет *TpNTT1*, *TpNTT2*, *TpNTT3*, *TpNTT4*, *TpNTT5*, *TpNTT6*, *TpNTT7*, *TpNTT8* (*T. pseudonana*), *PtNTT1*, *PtNTT2*, *PtNTT3*, *PtNTT4*, *PtNTT5*, *PtNTT6* (*P. tricornutum*), *GsNTT* (*Galdieria sulphuraria*), *AtNTT1*, *AtNTT2* (*Arabidopsis thaliana*), *CtNTT1*, *CtNTT2* (*Chlamydia trachomatis*), *PamNTT1*, *PamNTT2* (*Protochlamydia amoebophila*), *CeNTT* (*Caedibacter caryophilus*) или *RpNTT1* (*Rickettsia prowazekii*). В некоторых вариантах осуществления, NTT представляет собой CNT1, CNT2, CNT3, ENT1, ENT2, OAT1, OAT3 или OCT1. В некоторых случаях, NTT представляет собой *PtNTT1*, *PtNTT2*, *PtNTT3*, *PtNTT4*, *PtNTT5* или *PtNTT6*.

В некоторых вариантах осуществления, NTT импортирует нуклеиновые кислоты не природного происхождения в организм, например в клетку. В некоторых вариантах осуществления, NTT могут быть модифицированы таким образом, что сайт связывания нуклеотидов NTT модифицируется для снижения ингибирования пространственного проникновения нуклеиновой кислоты не природного происхождения в сайт связывания нуклеотида. В некоторых вариантах осуществления, NTT могут быть модифицированы для обеспечения повышенного взаимодействия с одним или несколькими природными или не природными признаками нуклеиновых кислот не природного происхождения. Такие NTT могут быть экспрессированы или сконструированы в клетках для стабильного импорта UBP в клетки. Соответственно, настоящее описание включает композиции, которые включают гетерологичный или рекомбинантный NTT, и способы их применения.

NTT можно модифицировать с помощью способов, относящихся к белковой инженерии. Например, можно проводить молекулярное моделирование на основе кристаллических структур, чтобы определить расположение NTT, где можно произвести мутации для изменения таргетной активности или сайта связывания. Остаток, идентифицированный как мишень для замены, может быть заменен остатком, выбранным с использованием моделирования минимизации энергии, моделирования гомологии и/или консервативных аминокислотных замен, таких как описано в Bordo, et al. J Mol Biol 217: 721-729 (1991) and Hayes, et al. Proc Natl Acad Sci, USA 99: 15926- 15931 (2002).

Любой из множества NTT можно использовать в способах или композициях, изложенных в настоящем документе, включая, например, белковые ферменты, выделенные из биологических систем, и их функциональные варианты. Будет понятно, что ссылка на конкретный NTT, такой как примеры, приведенные ниже, включает его функциональные варианты, если не указано иное. В некоторых вариантах осуществления, NTT представляет собой NTT дикого типа. В некоторых вариантах осуществления, NTT представляет собой

модифицированный или мутантный NTT.

В некоторых вариантах осуществления модифицированные или мутированные NTT, используемые в настоящем документе, представляют собой NTT, усеченные на N-конце, на C-конце или на N и C-конце. В некоторых вариантах осуществления, усеченный NTT, по меньшей мере, на 60%, по меньшей мере, на 65%, по меньшей мере, на 70%, по меньшей мере, на 75%, по меньшей мере, на 80%, по меньшей мере, на 85% или, по меньшей мере, на 90% идентичны не усеченному NTT. В некоторых случаях, NTT, используемые в настоящем документе, представляют собой *PtNTT1*, *PtNTT2*, *PtNTT3*, *PtNTT4*, *PtNTT5* или *PtNTT6*. В некоторых случаях, используемые в настоящем документе *PtNTT* усечены на N-конце, на C-конце или на N и на C-конце. В некоторых вариантах осуществления, усеченные *PtNTT*, по меньшей мере на 60%, по меньшей мере, на 65%, по меньшей мере, на 70%, по меньшей мере, на 75%, по меньшей мере, на 80%, по меньшей мере, на 85% или, по меньшей мере, на 90% идентичны не усеченным *PtNTT*. В некоторых случаях, используемый в настоящем документе NTT представляет собой усеченный *PtNTT2*, где усеченный *PtNTT2* имеет аминокислотную последовательность, которая по меньшей мере, на 60%, по меньшей мере, на 65%, по меньшей мере, на 70%, по меньшей мере, на 75%, по меньшей мере, на 80%, по меньшей мере, на 85%, или, по меньшей мере на 90% идентична аминокислотной последовательности не усеченного *PtNTT2*. Пример не усеченного *PtNTT2* (номер доступа NCBI EEC49227.1, GI: 217409295) имеет аминокислотную последовательность SEQ ID NO: 1.

Также можно использовать NTT, обладающие свойствами для улучшения входа нуклеиновых кислот не природного происхождения в клетки и для координации с нуклеотидами не природного происхождения в области связывания нуклеотидов. В некоторых вариантах осуществления, модифицированный NTT имеет модифицированный сайт связывания нуклеотидов. В некоторых вариантах осуществления, модифицированная нуклеиновая кислота или NTT дикого типа обладают ослабленной специфичностью в отношении нуклеиновой кислоты не природного происхождения. Например, NTT необязательно проявляет специфическую активность по импорту нуклеотида не природного происхождения, которая составляет, по меньшей мере, примерно 0,1% (например, около 0,1%, 0,2%, 0,5%, 0,8%, 1%, 1,1%, 1,2%, 1,5%, 1,8%, 2%, 3%, 4%, 5%, 10%, 25%, 50%, 75%, 100% или выше) от активности соответствующего NTT дикого типа. Необязательно, NTT демонстрирует k_{cat}/K_m или V_{max}/K_m для нуклеотида не природного происхождения, который, по меньшей мере, примерно на 0,1% выше (например, примерно 0,1%, 0,2%, 0,5%, 0,8%, 1%, 1,1%, 1,2%, 1,5%, 1,8%, 2%, 3%, 4%, 5%, 10%, 25%, 50%, 75% или 100% или выше) значений NTT дикого типа.

NTT можно охарактеризовать в соответствии с их аффинностью к трифосфату (т.е. K_m) и/или скоростью импорта (т.е. V_{max}). В некоторых вариантах осуществления, NTT имеет относительную K_m или V_{max} для одного или нескольких трифосфатов природного и не природного происхождения. В некоторых вариантах осуществления, NTT имеет относительно высокие значения K_m или V_{max} для одного или нескольких трифосфатов

природного и не природного происхождения.

NTT из нативных источников или их варианты можно подвергать скринингу с помощью анализа, определяющего количество трифосфата (либо с помощью масс-спектра, либо с помощью радиоактивности, если трифосфат соответствующим образом помечен). В одном примере, NTT могут быть подвергнуты скринингу на способность импортировать трифосфат не природного происхождения, например, dTPTЗTP, dCNMOTP, d5SICSTP, dNaMTP, NaMTP и/или TPT1TP. Можно использовать NTT, например, гетерологичный NTT, который проявляет модифицированное свойство нуклеиновой кислоты не природного происхождения по сравнению с NTT дикого типа. Например, модифицированным свойством может быть, например, K_m , k_{cat} , V_{max} , для импорта трифосфата. В одном варианте, модифицированное свойство представляет собой пониженное значение K_m для трифосфата не природного происхождения и/или повышенное значение k_{cat}/K_m или V_{max}/K_m для трифосфата не природного происхождения. Точно так же, NTT необязательно имеет повышенную скорость связывания трифосфата не природного происхождения, повышенную скорость внутриклеточного высвобождения и/или повышенную скорость импорта клеток по сравнению с NTT дикого типа.

В то же время NTT может импортировать в клетку природные трифосфаты, например, dATP, dCTP, dGTP, dTTP, ATP, CTP, GTP и/или TTP. В некоторых вариантах осуществления, NTT необязательно демонстрирует специфическую активность импорта природной нуклеиновой кислоты, которая способна поддерживать репликацию и транскрипцию. В некоторых вариантах осуществления, NTT необязательно демонстрирует k_{cat}/K_m или V_{max}/K_m для природной нуклеиновой кислоты, которая способна поддерживать репликацию и транскрипцию.

Используемые в настоящем документе NTT, которые могут обладать способностью импортировать трифосфат не природного происхождения определенной структуры, также могут быть получены с использованием подхода направленного развития. Анализ синтеза нуклеиновой кислоты можно использовать для скрининга вариантов NTT, обладающих специфичностью к любому из множества трифосфатов не природного происхождения. Например, варианты NTT могут быть проверены на способность импортировать трифосфат не природного происхождения, например, d5SICSTP, dNaMTP, dCNMOTP, dTPTЗTP, NaMTP и/или TPT1TP. В некоторых вариантах осуществления, такой анализ представляет собой анализ *in vitro*, например, с применением рекомбинантного варианта NTT. В некоторых вариантах осуществления, такой анализ представляет собой анализ *in vivo*, например экспрессию варианта NTT в клетке. Такие методики можно использовать для скрининга вариантов любого подходящего NTT на активность в отношении любого трифосфата не природного происхождения, указанного в настоящем документе.

Реагенты и инструменты для нуклеиновых кислот

Нуклеотид и/или реагент нуклеиновой кислоты (или полинуклеотид) для применения в способах, клетках или сконструированных микроорганизмах, описанных в настоящем документе, содержит одну или несколько ORF с или без нуклеотида не

природного происхождения. ORF может быть из любого подходящего источника, иногда из геномной ДНК, иРНК, обратно транскрибируемой РНК или комплементарной ДНК (кДНК) или библиотеки нуклеиновых кислот, содержащей одну или несколько из вышеперечисленных, и из любого вида организма, который содержит представляющую интерес последовательность нуклеиновой кислоты, представляющий интерес белок или представляющую интерес активность. Неограничивающие примеры организмов, из которых можно получить ORF, включают, например, бактерии, дрожжи, грибы, человека, насекомых, нематод, крупный рогатый скот, лошадей, собак, кошек, крыс или мышей. В некоторых вариантах осуществления, нуклеотид или реагент нуклеиновой кислоты или другой реагент, описанный в настоящем документе, выделяют или очищают. Могут быть созданы ORF, которые включают нуклеотиды не природного происхождения, с помощью опубликованных способов *in vitro*. В некоторых случаях, нуклеотидный или нуклеиновый реагент содержит нуклеиновое основание не природного происхождения.

Реагент нуклеиновой кислоты иногда содержит нуклеотидную последовательность, примыкающую к ORF, которая транслируется вместе с ORF и кодирует аминокислотную метку. Последовательность нуклеотидов, кодирующая метку, расположена на 3' и/или 5' ORF реагента нуклеиновой кислоты, тем самым кодируя метку на С-конце или N-конце белка или пептида, кодируемого ORF. Можно использовать любую метку, которая не аннулирует транскрипцию и/или трансляцию *in vitro*, и она может быть надлежащим образом выбрана специалистом в данной области техники. Метки могут облегчить выделение и/или очистку желаемого продукта ORF из культуральной или ферментационной среды. В некоторых случаях, библиотеки реагентов нуклеиновых кислот используют с описанными в настоящем документе способами и композициями. Например, библиотека, по меньшей мере, 100, 1000, 2000, 5000, 10000 или более 50000 уникальных полинуклеотидов присутствуют в библиотеке, где каждый полинуклеотид содержит, по меньшей мере, одно нуклеотидное основание не природного происхождения.

Нуклеиновая кислота или реагент нуклеиновой кислоты, с или без нуклеотидом не природного происхождения, могут содержать определенные элементы, *например*, регуляторные элементы, часто выбранные в соответствии с предполагаемым использованием нуклеиновой кислоты. Любой из следующих элементов может быть включен или исключен из реагента нуклеиновой кислоты. Реагент нуклеиновой кислоты, например, может включать один или несколько или все из следующих нуклеотидных элементов: один или несколько промоторных элементов, одну или несколько 5' нетранслируемых областей (5'UTR), одну или несколько областей в которую может быть вставлена последовательность нуклеотида-мишени («элемент вставки»), одна или несколько последовательностей нуклеотидов-мишеней, одна или несколько 3' нетранслируемых областей (3'UTR) и один или несколько элементов селекции. Реагент нуклеиновой кислоты может быть представлен с одним или несколькими такими элементами, и другие элементы могут быть вставлены в нуклеиновую кислоту до того, как нуклеиновая кислота будет введена в желаемый организм. В некоторых вариантах

осуществления, представленный реагент нуклеиновой кислоты содержит промотор, 5'UTR, необязательно, 3'UTR и элемент(ы) вставки, с помощью которых нуклеотидная последовательность-мишень встраивается (т.е. клонируется) в реагент нуклеотидной кислоты. В некоторых вариантах осуществления, представленный реагент нуклеиновой кислоты содержит промотор, элемент(ы) вставки и, необязательно, 3'UTR, и 5'UTR/нуклеотидная последовательность-мишень вставляется с необязательной 3'UTR. Элементы могут быть расположены в любом порядке, подходящем для экспрессии в выбранной системе экспрессии (*например*, экспрессии в выбранном организме или экспрессии в бесклеточной системе, *например*), и в некоторых вариантах осуществления, реагент нуклеиновой кислоты, содержащий следующие элементы в направлении 5'-3': (1) промоторный элемент, 5'UTR, и элемент(ы) вставки; (2) промоторный элемент, 5'UTR и нуклеотидную последовательность-мишень; (3) промоторный элемент, 5'UTR, элемент(ы) вставки и 3'UTR; и (4) промоторный элемент, 5'UTR, нуклеотидную последовательность-мишень и 3'UTR. В некоторых вариантах осуществления, UTR можно оптимизировать для изменения или увеличения транскрипции или трансляции ORF, которые либо полностью являются существующими в природе, либо содержат нуклеотиды не природного происхождения.

Реагенты нуклеиновой кислоты, *например*, кассеты экспрессии и/или векторы экспрессии, могут включать различные регуляторные элементы, включая промоторы, энхансеры, последовательности инициации трансляции, последовательности терминации транскрипции и другие элементы. «Промотор» обычно представляет собой последовательность или последовательности ДНК, которые функционируют, когда находятся в относительно фиксированной локации по отношению к сайту начала транскрипции. *Например*, промотор может располагаться выше сегмента нуклеиновой кислоты транспортера нуклеотидтрифосфата. «Промотор» содержит коровые элементы, необходимые для основного взаимодействия РНК полимеразы и факторов транскрипции, и могут содержать вышерасположенные элементы и элементы ответа. «Энхансер» обычно относится к последовательности ДНК, которая функционирует на не фиксированном расстоянии от сайта начала транскрипции, и может быть либо 5', либо 3" к единице транскрипции. Кроме того, энхансеры могут находиться как внутри интрона, так и внутри самой кодирующей последовательности. Они обычно имеют длину от 10 до 300 и функционируют в *цис*-положении. Энхансеры усиливают транскрипцию из близлежащих промоторов. Энхансеры, как и промоторы, также часто содержат элементы ответа, которые опосредуют регуляцию транскрипции. Энхансеры часто определяют регуляцию экспрессии и могут использоваться для изменения или оптимизации экспрессии ORF, включая ORF, которые являются полностью природными или содержат нуклеотиды не природного происхождения.

Как отмечалось выше, реагенты нуклеиновой кислоты могут также содержать одну или несколько 5'UTR и одну или несколько 3'UTR. *Например*, векторы экспрессии, используемые в эукариотических клетках-хозяевах (*например*, дрожжах, грибах,

насекомых, растениях, животных, человеческом или ядерных клетках) и прокариотических клетках-хозяевах (например, вирусе, бактерии) могут содержать последовательности, которые сигнализируют об окончании транскрипции, что может влиять на экспрессию иРНК. Эти области могут быть транскрибированы как полиаденилированные сегменты в нетранслируемой части иРНК, кодирующей белок фактора ткани. 3' нетранслируемые области также включают сайты терминации транскрипции. В некоторых предпочтительных вариантах осуществления, единица транскрипции содержит область полиаденилирования. Одно из преимуществ этой области заключается в том, что она увеличивает вероятность того, что транскрибируемая единица будет процессироваться и транспортироваться подобно иРНК. Идентификация и использование сигналов полиаденилирования в конструкциях экспрессии хорошо известны. В некоторых предпочтительных вариантах осуществления, гомологичные сигналы полиаденилирования можно использовать в трансгенных конструкциях.

5' UTR может содержать один или несколько элементов, эндогенных по отношению к нуклеотидной последовательности, из которой она происходит, и иногда содержит один или несколько экзогенных элементов. 5'UTR может происходить из любой подходящей нуклеиновой кислоты, такой как геномная ДНК, плазмидная ДНК, РНК или иРНК, например, из любого подходящего организма (*например*, вируса, бактерии, дрожжей, гриба, растения, насекомого или млекопитающего). Специалист в данной области техники может выбрать подходящие элементы для 5' UTR на основе выбранной системы экспрессии (*например*, экспрессии в выбранный организм или экспрессии в бесклеточной системе, *например*). 5'UTR иногда содержит один или несколько из следующих элементов, известных специалисту в данной области техники: энхансерные последовательности (*например*, транскрипционные или трансляционные), сайт инициации транскрипции, сайт связывания фактора транскрипции, сайт регуляции трансляции, сайт инициации трансляции, сайт связывания фактора трансляции, сайт связывания акцессорного белка, сайты связывания агента регуляции с обратной связью, бокс Прибнова, бокс ТАТА, -35 элемент, Е-бокс (спираль-петля-спираль связывающий элемент), сайт связывания рибосомы, репликон, сайт внутренней посадки рибосомы (IRES), ослабляющий элемент и подобные. В некоторых вариантах осуществления, промоторный элемент может быть выделен таким образом, чтобы все 5' UTR элементы, необходимые для надлежащей условной регуляции, содержались в фрагменте промоторного элемента или внутри функциональной субпоследовательности фрагмента промоторного элемента.

5'UTR в реагенте нуклеиновой кислоты может содержать нуклеотидную последовательность энхансера трансляции. Нуклеотидная последовательность энхансера трансляции часто расположена между промотором и последовательностью нуклеотидов-мишеней в реагенте нуклеиновой кислоты. Последовательность энхансера трансляции часто связывается с рибосомой, иногда представляет собой рибонуклеотидную последовательность, связывающую 18S рРНК (т. е. последовательность связывания рибосомы 40S), и иногда представляет собой последовательность внутренней посадки

рибосомы (IRES). IRES обычно образует каркас РНК с точно расположенными третичными структурами РНК, которые контактируют с субъединицей 40S рибосомы через ряд специфических межмолекулярных взаимодействий. Примеры последовательностей рибосомных энхансеров известны и могут быть идентифицированы специалистом в данной области техники (*например*, Mignone et al., *Nucleic Acids Research* 33: D141-D146 (2005); Paulous et al., *Nucleic Acids Research* 31: 722-733 (2003); Akbergenov et al., *Nucleic Acids Research* 32: 239-247 (2004); Mignone et al., *Genome Biology* 3(3): reviews0004.1-0001.10 (2002); Gallie, *Nucleic Acids Research* 30: 3401-3411 (2002); Shaloiko et al., DOI: 10.1002/bit.20267; и Gallie et al., *Nucleic Acids Research* 15: 3257-3273 (1987)).

Последовательность энхансера трансляции иногда представляет собой эукариотическую последовательность, такую как консенсусная последовательность Козака, или другую последовательность (*например*, последовательность гидроидного полипа, номер доступа в GenBank U07128). Последовательность энхансера трансляции иногда является прокариотической последовательностью, такой как консенсусная последовательность Шайна-Далгарно. В некоторых вариантах осуществления, последовательность энхансера трансляции представляет собой вирусную нуклеотидную последовательность. Последовательность энхансера трансляции иногда происходит из 5' UTR вируса растений, такого как вирус табачной мозаики (TMV), вирус мозаики люцерны (AMV); вирус гравировки табака (ETV); вирус Y картофеля (PVY); вирус мозаики репы (poty) и вирус мозаики, переносимый семенами гороха, например. В некоторых вариантах осуществления, омега последовательность длиной около 67 оснований из TMV включена в реагент нуклеиновой кислоты в качестве последовательности энхансера трансляции (*например*, лишенной гуанозиновых нуклеотидов и включающей поли (CAA) центральную область длиной 25 нуклеотидов).

3' UTR может содержать один или несколько элементов, эндогенных по отношению к нуклеотидной последовательности, из которой она происходит, и иногда включает один или несколько экзогенных элементов. 3' UTR может происходить из любой подходящей нуклеиновой кислоты, такой как геномная ДНК, плазмидная ДНК, РНК или иРНК, например, из любого подходящего организма (*например*, вируса, бактерии, дрожжей, гриба, растения, насекомого или млекопитающего). Специалист в данной области техники может выбрать подходящие элементы для 3' UTR на основе выбранной системы экспрессии (*например*, экспрессии в выбранный организм, например). 3'UTR иногда содержит один или несколько из следующих элементов, известных специалисту в данной области техники: сайт регуляции транскрипции, сайт инициации транскрипции, сайт терминации транскрипции, сайт связывания фактора транскрипции, сайт регуляции трансляции, сайт терминации трансляции, сайт инициации трансляции, сайт связывания фактора трансляции, сайт связывания рибосомы, репликон, энхансерный элемент, сайленсерный элемент и полиаденозиновый хвост. 3' UTR часто включает полиаденозиновый хвост, и иногда нет, и, если присутствует полиаденозиновый хвост, в него могут быть добавлены или удалены одна или несколько аденозиновых групп (*например*, примерно 5, примерно 10, примерно

15, примерно 20, примерно 25, примерно 30, примерно 35, примерно 40, примерно 45 или примерно 50 аденозиновых групп могут быть добавлены или удалены).

В некоторых вариантах осуществления, модификация 5' UTR и/или 3' UTR используется для изменения (например, увеличения, добавления, уменьшения или по существу устранения) активности промотора. Изменение активности промотора, в свою очередь, может изменить активность пептида, полипептида или белка (например, активность фермента) путем изменения транскрипции представляющей интерес нуклеотидной последовательности(ей) из функционально связанного промоторного элемента, содержащего модифицированную 5' или 3' UTR. Например, микроорганизм может быть сконструирован путем генетической модификации для экспрессии реагента нуклеиновой кислоты, содержащего модифицированную 5' или 3' UTR, которая может добавить новую активность (например, активность, обычно не обнаруживаемую в организме-хозяине) или увеличить экспрессию существующей активности путем повышения транскрипции из гомологичного или гетерологичного промотора, функционально связанного с представляющей интерес нуклеотидной последовательностью (например, гомологичной или гетерологичной представляющей интерес нуклеотидной последовательностью), в некоторых вариантах осуществления. А некоторых вариантах осуществления, микроорганизм может быть сконструирован генетической модификацией для экспрессии реагента нуклеиновой кислоты, содержащего модифицированную 5' или 3' UTR, который может снижать экспрессию активности за счет снижения или по существу устранения транскрипции из гомологичного или гетерологичного промотора, функционально связанного с представляющей интерес нуклеотидной последовательностью, в некоторых вариантах осуществления.

Экспрессия транспортера нуклеотидтрифосфата из кассеты экспрессии или вектора экспрессии может контролироваться любым промотором, способным к экспрессии в прокариотических или эукариотических клетках. Промоторный элемент обычно необходим для синтеза ДНК и/или синтеза РНК. Промоторный элемент часто содержит участок ДНК, который может способствовать транскрипции определенного гена, обеспечивая сайт начала для синтеза РНК, соответствующей гену. Промоторы обычно расположены рядом с генами, которые они регулируют, расположены выше гена (*например*, 5' от гена), и находятся на той же цепи ДНК, что и смысловая цепь гена, в некоторых вариантах осуществления. В некоторых вариантах осуществления, промоторный элемент может быть выделен из гена или организма и встроен в функциональную связь с полинуклеотидной последовательностью, обеспечивающей измененную и/или регулируемую экспрессию. Не нативный промотор (*например*, промотор, обычно не связанный с данной последовательностью нуклеиновой кислоты), используемый для экспрессии нуклеиновой кислоты, часто называют гетерологичным промотором. В некоторых вариантах осуществления, гетерологичный промотор и/или 5'UTR могут быть вставлены в функциональную связь с полинуклеотидом, который кодирует полипептид, обладающий желаемой активностью, как описано в настоящем документе. Термины «функционально

связанный» и «в функциональной связи с», используемые в настоящем документе в отношении промоторов, относятся к взаимосвязи между кодирующей последовательностью и промоторным элементом. Промотор функционально связан или находится в функциональной связи с кодирующей последовательностью, когда экспрессия из кодирующей последовательности через транскрипцию регулируется или контролируется промоторным элементом. Термины «функционально связанный» и «в функциональной связи с» используются в настоящем документе взаимозаменяемо в отношении промоторных элементов.

Промотор часто взаимодействует с РНК полимеразой. Полимераза представляет собой фермент, который катализирует синтез нуклеиновых кислот с использованием ранее существовавшего реагента нуклеиновой кислоты. Когда матрица представляет собой матрицу ДНК, молекула РНК транскрибируется до синтеза белка. Ферменты, обладающие полимеразной активностью, подходящей для применения в настоящих способах, включают любую полимеразу, активную в выбранной системе, с выбранной матрицей для синтеза белка. В некоторых вариантах осуществления, промотор (*например*, гетерологичный промотор), также называемый в настоящем документе промоторным элементом, может быть функционально связан с нуклеотидной последовательностью или открытой рамкой считывания (ORF). Транскрипция из промоторного элемента может катализировать синтез РНК, соответствующей нуклеотидной последовательности или ORF, функционально связанной с промотором, что, в свою очередь, приводит к синтезу желаемого пептида, полипептида или белка.

Промоторные элементы иногда демонстрируют реакцию на регуляторный контроль. Промоторные элементы также иногда могут регулироваться селективным агентом. То есть, транскрипция из промоторных элементов иногда может включаться, выключаться, усиливаться или подавляться в ответ на изменение в окружающей среде, питании или внутренних условиях или сигналах (*например*, промоторы, индуцируемые нагреванием, промоторы, регулируемые светом, промоторы, регулируемые с обратной связью, промоторы, влияющие на гормоны, тканеспецифические промоторы, промоторы, влияющие на кислород и рН, промоторы, которые реагируют на селективные агенты (*например*, канамицин) и подобные, например). Промоторы, находящиеся под влиянием сигналов окружающей среды, питания или внутренних сигналов, часто подвергаются влиянию сигнала (прямого или косвенному), который связывается на или рядом с промотором, и увеличивает или уменьшает экспрессию последовательности-мишени при определенных условиях. Что касается всех описанных в настоящем документе способов, включение природных или модифицированных промоторов можно использовать для изменения или оптимизации экспрессии полностью природной ORF (*например*, NTT или aaRS) или ORF, содержащей нуклеотид не природного происхождения (*например*, иРНК или тРНК).

Неограничивающие примеры селективных или регуляторных агентов, которые влияют на транскрипцию из промоторного элемента, используемого в вариантах

осуществления, описанных в настоящем документе, включают, без ограничения, (1) сегменты нуклеиновых кислот, которые кодируют продукты, которые представляют резистентность к другим токсичным соединениям (*например*, антибиотикам); (2) сегменты нуклеиновой кислоты, которые кодируют продукты, которые иначе отсутствуют в клетке-реципиенте (*например*, основные продукты, гены тРНК, ауксотрофные маркеры); (3) сегменты нуклеиновых кислот, которые кодируют продукты, подавляющие активность генного продукта; (4) сегменты нуклеиновой кислоты, которые кодируют продукты, которые можно легко идентифицировать (*например*, фенотипические маркеры, такие как антибиотики (*например*, β -лактамаза), β -галактозидаза, зеленый флуоресцентный белок (GFP).), желтый флуоресцентный белок (YFP), красный флуоресцентный белок (RFP), голубой флуоресцентный белок (CFP) и белки клеточной поверхности); (5) сегменты нуклеиновой кислоты, которые связывают продукты, которые в противном случае вредны для выживания и/или функции клеток; (6) сегменты нуклеиновой кислоты, которые иным образом ингибируют активность любого из сегментов нуклеиновой кислоты, описанных выше в пп. 1-5 (*например*, антисмысловые олигонуклеотиды); (7) сегменты нуклеиновой кислоты, которые связывают продукты, модифицирующие субстрат (*например*, рестрикционные эндонуклеазы); (8) сегменты нуклеиновой кислоты, которые можно использовать для выделения или идентификации желаемой молекулы (*например*, сайты связывания специфического белка); (9) сегменты нуклеиновой кислоты, которые кодируют специфическую нуклеотидную последовательность, которая в противном случае может быть не функциональной (*например*, для ПЦР амплификации субпопуляций молекул); (10) сегменты нуклеиновой кислоты, которые, при отсутствии, прямо или косвенно придают устойчивость или чувствительность к конкретным соединениям; (11) сегменты нуклеиновой кислоты, которые кодируют продукты, которые либо являются токсичными, либо превращают относительно нетоксичное соединение в токсичное соединение (*например*, тимидинкиназа Herpes simplex, цитозиндезаминаза) в клетках-реципиентах; (12) сегменты нуклеиновой кислоты, которые ингибируют репликацию, разделение или наследуемость молекул нуклеиновых кислот, которые их содержат; (13) сегменты нуклеиновой кислоты, которые кодируют условные функции репликации, *например*, репликацию в определенных хозяевах или штаммах клеток-хозяев или в определенных условиях окружающей среды (*например*, температура, условия питания и подобные); и/или (14) нуклеиновые кислоты которые кодируют одну или несколько иРНК или тРНК, содержащие нуклеотиды не природного происхождения. В некоторых вариантах осуществления, регуляторный или селективный агент может быть добавлен для изменения существующих условий роста, которым подвергается организм (*например*, роста в жидкой культуре, роста в ферментере, роста на твердых питательных чашках и подобное, *например*).

В некоторых вариантах осуществления, регулирование промоторного элемента можно использовать для изменения (*например*, увеличения, добавления, снижения или по существу устранения) активности пептида, полипептида или белка (*например*, активности

фермента). Например, микроорганизм может быть сконструирован путем генетической модификации для экспрессии реагента нуклеиновой кислоты, который может добавить новую активность (например, активность, обычно не обнаруживаемую в организме-хозяине) или увеличить экспрессию существующей активности путем усиления транскрипции из гомологичного или гетерологичного промотора, функционально связанного с представляющей интерес последовательностью нуклеотидов (например, гомологичной или гетерологичной представляющей интерес последовательностью нуклеотидов) в некоторых вариантах осуществления. В некоторых вариантах осуществления, микроорганизм может быть сконструирован путем генетической модификации для экспрессии реагента нуклеиновой кислоты, который может снижать экспрессию активности путем уменьшения или по существу удаления транскрипции с гомологичного или гетерологичного промотора, функционально связанного с интересующей нуклеотидной последовательностью, в определенных вариантах осуществления.

Нуклеиновые кислоты, кодирующие гетерологичные белки, например, транспортеров нуклеотидтрифосфата, могут быть встроены или использованы с любой подходящей системой экспрессии. В некоторых вариантах осуществления, реагент нуклеиновой кислоты иногда стабильно интегрирован в хромосому организма-хозяина, или реагент нуклеиновой кислоты может быть делецией части хромосомы-хозяина, в некоторых вариантах осуществления, (например, генетически модифицированных организмов, где изменение генома-хозяина дает возможность селективно или предпочтительно поддерживать желаемый организм, несущий генетическую модификацию). Такие реагенты нуклеиновой кислоты (например, нуклеиновые кислоты или генетически модифицированные организмы, чей измененный геном придает организму селективный признак) могут быть отобраны по их способности направлять продуцирование желаемого белка или молекулы нуклеиновой кислоты. При желании, реагент нуклеиновой кислоты может быть изменен так, что кодоны кодируют (i) ту же аминокислоту, с использованием другой тРНК, отличной от той, которая указана в нативной последовательности, или (ii) отличающуюся от обычной аминокислоту, включающую нестандартные аминокислоты или аминокислоты не природного происхождения (включая определяемые меченые аминокислоты).

Рекомбинантную экспрессию полезно осуществлять с использованием кассеты экспрессии, которая может быть частью вектора, такого как плазида. Вектор может включать промотор, функционально связанный с нуклеиновой кислотой, кодирующей транспортер нуклеотидтрифосфата. Вектор также может включать другие элементы, необходимые для транскрипции и трансляции, как описано в настоящем документе. Кассета экспрессии, вектор экспрессии и последовательности в кассете или векторе могут быть гетерологичны клетке, с которой контактируют нуклеотиды не природного происхождения. Например, последовательность транспортер нуклеотидтрифосфата может быть гетерологичной к клетке.

Могут быть получены различные прокариотические и эукариотические векторы экспрессии, подходящие для переноса, кодирования и/или экспрессии транспортеров нуклеотидтрифосфата. Такие векторы экспрессии включают, например, pET, pET3d, pCR2.1, pBAD, pUC и дрожжевые векторы. Векторы можно использовать, например, в различных ситуациях *in vivo* и *in vitro*. Неограничивающие примеры прокариотических промоторов, которые можно использовать, включают SP6, T7, T5, *lac*, *bla*, *trp*, *gal*, *lac* или мальтозные промоторы. Неограничивающие примеры эукариотических промоторов, которые можно использовать, включают конститутивные промоторы, *например*, вирусные промоторы, такие как промоторы CMV, SV40 и RSV, а также регулируемые промоторы, *например*, индуцируемый или репрессируемый промотор, такой как *tet* промотор, *hsp70* промотор и синтетический промотор, регулируемый CRE. Векторы для бактериальной экспрессии включают pGEX-5X-3, и для эукариотической экспрессии включают pCIneo-CMV. Вирусные векторы, которые можно использовать, включают векторы, относящиеся к лентивирусам, аденовирусам, ассоциированный вирус, вирус герпеса, вирус коровьей оспы, вирус полиомиелита, вирус AIDS, нейрональный трофический вирус, вирус Синдбис и другие вирусы. Также пригодны любые вирусные семейства, обладающие свойствами этих вирусов, которые делают их подходящими для использования в качестве векторов. Ретровирусные векторы, которые могут применяться, включают описанные в Verma, American Society for Microbiology, pp. 229-232, Washington, (1985). Например, такие ретровирусные векторы могут включать вирус лейкоза Малони мыши, MMLV и другие ретровирусы, которые проявляют желательные свойства. Обычно, вирусные векторы содержат неструктурные ранние гены, структурные поздние гены, транскрипт РНК полимеразы III, инвертированные концевые повторы, необходимые для репликации и инкапсуляции, и промоторы для контроля транскрипции и репликации вирусного генома. При разработке в качестве векторов вирусы обычно имеют один или несколько удаленных ранних генов, и ген или кассету гена/промотора вставляют в вирусный геном вместо удаленной вирусной нуклеиновой кислоты.

Клонирование

Для включения элемента, такого как ORF, в реагент нуклеиновой кислоты можно использовать любую удобную стратегию клонирования, известную в данной области техники. Известные способы можно использовать для вставки элемента в матрицу независимо от элемента вставки, например (1) расщепление матрицы по одному или нескольким существующим сайтам рестрикционных ферментов и лигирование представляющего интерес элемента, и (2) добавление сайтов рестрикционных ферментов к матрице путем гибридизации олигонуклеотидных праймеров, которые включают один или несколько подходящих сайтов рестрикционных ферментов и амплификацию с помощью полимеразной цепной реакции (описанной в более подробно в настоящем документе). В других стратегиях клонирования используется преимущество одного или нескольких сайтов вставки, присутствующих или встроенных в реагент нуклеиновой кислоты, например, сайта гибридизации олигонуклеотидного праймера для ПЦР, и других,

описанных в настоящем документе. В некоторых вариантах осуществления, стратегия клонирования может быть объединена с генетическими манипуляциями, такими как рекомбинация (например, рекомбинация реагента нуклеиновой кислоты с последовательностью нуклеиновой кислоты, представляющей интерес, в геном модифицированного организма, как описано далее в настоящем документе). В некоторых вариантах осуществления, клонированная ORF может продуцировать (прямо или косвенно) модифицированные или дикого типа транспортеры нуклеотидтрифосфата и/или полимеразы путем конструирования микроорганизма с одной или несколькими представляющими интерес ORF, где микроорганизмы содержат измененную активность активности транспортера нуклеотидтрифосфата или полимеразной активности.

Нуклеиновая кислота может быть специфически расщеплена при контакте нуклеиновой кислоты с одним или несколькими специфическими расщепляющими агентами. Специфические расщепляющие агенты часто будут специфически расщеплять в соответствии с определенной последовательностью нуклеотидов на конкретном участке. Примеры фермент-специфических расщепляющих агентов включают, без ограничения, эндонуклеазы (*например*, ДНКазу (*например*, ДНКазу I, II); РНКазу (*например*, РНКазу E, F, H, P); фермент Cleavase™; Taq ДНК полимеразу; ДНК полимеразу I *E. coli* и эукариотические структурно-специфические эндонуклеазы; эндонуклеазы FEN-1 мыши; рестрикционные эндонуклеазы типа I, II или III, такие как Acc I, Afl III, Alu I, Alw44 I, Apa I, Asn I, Ava I, Ava II, BamH I, Ban II, Bcl I, Bgl I, Bgl II, Bln I, BsaI, Bsm I, BsmBI, BssH II, BstE II, Cfo I, Cla I, Dde I, Dpn I, Dra I, EcIX I, EcoR I, EcoR II, EcoR V, Hae II, Hae II, Hind II, Hind III, Hpa I, Hpa II, Kpn I, Ksp I, Mlu I, MluN I, Msp I, Nci I, Nco I, Nde I, Nde II, Nhe I, Not I, Nru I, Nsi I, Pst I, Pvu I, Pvu II, Rsa I, Sac I, Sal I, Sau3A I, Sca I, ScrF I, Sfi I, Sma I, Spe I, Sph I, Ssp I, Stu I, Sty I, Swa I, Taq I, Xba I, Xho I); гликозилазы (*например*, урацил-ДНК-гликозилазу (UDG), 3-метиладенин-ДНК-гликозилазу, 3-метиладенин-ДНК-гликозилазу II, пиримидингидрат-ДНК-гликозилазу, FaPy-ДНК-гликозилазу, тимин-несовместимую ДНК гликозилазу, гипоксантин-ДНК гликозилазу, 5-гидроксиметилурацил ДНК гликозилазу (HmUDG), 5-гидроксиметилцитозин ДНК гликозилазу или 1,N6-этенаденин ДНК гликозилазу); экзонуклеазы (*например*, экзонуклеазу III); рибозимы и ДНКзимы. Образец нуклеиновой кислоты можно обработать химическим агентом или синтезировать с использованием модифицированных нуклеотидов, и модифицированная нуклеиновая кислота может быть расщеплена. В неограничивающих примерах, образец нуклеиновой кислоты можно обработать (i) алкилирующими агентами, такими как метилнитрозомочевина, которые образуют несколько алкилированных оснований, включая N3-метиладенин и N3-метилгуанин, которые распознаются и расщепляются алкилпуриновой ДНК-гликозилазой; (ii) бисульфитом натрия, который вызывает дезаминирование цитозиновых остатков в ДНК с образованием урацильных остатков, которые могут быть расщеплены урацил N-гликозилазой, и (iii) химическим агентом, который превращает гуанин в его окисленную форму, 8-гидроксигуанином, который может быть расщеплен формамидопиримидин ДНК N-гликозилазой. Примеры процессов

химического расщепления включают, без ограничения, алкилирование (*например*, алкилирование фосфоротиоат-модифицированной нуклеиновой кислоты), расщепление кислотнеустойчивой нуклеиновой кислоты, содержащей Р3'-N5'-фосфорамидат; и обработку нуклеиновой кислоты тетроксидом осмия и пиперидином.

В некоторых вариантах осуществления, реагент нуклеиновой кислоты включает один или несколько сайтов встраивания рекомбиназы. Сайт встраивания рекомбиназы представляет собой последовательность распознавания на молекуле нуклеиновой кислоты, которая участвует в реакции интеграции/рекомбинации рекомбинационными белками. Например, сайт рекомбинации для рекомбиназы Cre представляет собой loxP, который представляет собой последовательность из 34 пар оснований, состоящую из двух инвертированных повторов из 13 пар оснований (служащих сайтами связывания рекомбиназы), фланкирующих коровую последовательность из 8 пар оснований (*например*, Sauer, Curr. Opin. Biotech. 5:521-527 (1994)). Другие примеры сайтов рекомбинации включают последовательности attB, attP, attL и attR, и их мутанты, фрагменты, варианты и производные, которые распознаются рекомбинационным белком λ Int и вспомогательными белками фактора интеграции хозяина (IHF), FIS и эксцизионазой (Xis) (*например*, патенты США №№ . 5,888,732; 6,143,557; 6,171,861; 6,270,969; 6,277,608; и 6,720,140; заявки на патент США №№ 09/517,466 и 09/732,914; публикацию патента США № 002/0007051; и Landy, Curr. Opin. Biotech., 3:699-707 (1993)).

Примерами рекомбиназного клонирования нуклеиновых кислот являются системы Gateway® (Invitrogen, California), которые содержат, по меньшей мере, один сайт рекомбинации для клонирования желаемых молекул нуклеиновых кислот *in vivo* или *in vitro*. В некоторых вариантах осуществления, в системе используются векторы, которые содержат, по меньшей мере, два разных сайт-специфических сайтов рекомбинации, часто основанных на лямбда системе бактериофага (*например*, att1 и att2), и мутируют из сайтов дикого типа (att0). Каждый мутированный сайт имеет уникальную специфичность для att сайта его родственного партнера (т.е. сайта рекомбинации его партнера по связыванию) того же типа (*например*, attB1 с attP1 или attL1 с attR1) и не будет перекрестно реагировать с сайтами рекомбинации другого мутантного типа или с сайтом att0 дикого типа. Разная сайт-специфичность позволяет направленное клонирование или связывание желаемых молекул, таким образом обеспечивая желаемую ориентацию клонированных молекул. Фрагменты нуклеиновой кислоты, фланкированные сайтами рекомбинации, клонируют и субклонируют с использованием системы Gateway® путем замены селективируемого маркера (*например*, ccdB), фланкированного att сайтами, на молекуле плазмиды-реципиента, иногда называемой принимающим вектором. Затем нужные клоны отбирают путем трансформации штамма-хозяина, чувствительного к ccdB, и положительной селекции по маркеру на молекуле-реципиенте. Аналогичные стратегии отрицательной селекции (*например*, использование токсичных генов) можно использовать и в других организмах, таких как тимидинкиназа (ТК) у млекопитающих и насекомых.

Реагент нуклеиновой кислоты иногда содержит один или несколько элементов

начала репликации (ORI). В некоторых вариантах осуществления, матрица содержит два или несколько ORI, где один эффективно функционирует в одном организме (*например*, бактерии), и другой эффективно функционирует в другом организме (*например*, эукариоте, таком как, например, дрожжи). В некоторых вариантах осуществления, ORI может эффективно функционировать у одного вида (*например*, *S. cerevisiae*), и другой ORI может эффективно функционировать у других видов (*например*, *S. pombe*). Реагент нуклеиновой кислоты также иногда включает один или несколько сайтов регуляции транскрипции.

Реагент нуклеиновой кислоты, *например*, кассета или вектор экспрессии, может включать последовательность нуклеиновой кислоты, кодирующую маркерный продукт. Маркерный продукт используют для определения того, был ли ген доставлен в клетку и экспрессируется ли он после доставки. Примеры маркерных генов включают ген *lacZ E. coli*, который кодирует β -галактозидазу и зеленый флуоресцентный белок. В некоторых вариантах осуществления, маркер может быть селективируемым маркером. Когда такие селективируемые маркеры успешно переносятся в клетку-хозяин, трансформированная клетка-хозяин может выжить при воздействии селективного давления. Существуют две широко используемые отдельные категории селективных схем. Первая категория основана на клеточном метаболизме и использовании мутантной клеточной линии, которая не способна расти независимо от дополненной среды. Вторая категория представляет собой доминантный отбор, который относится к схеме селекции, используемой для любого типа клеток, и не требует использования мутантной клеточной линии. В этих схемах обычно используется лекарственное средство для остановки роста клетки-хозяина. Те клетки, которые имеют новый ген, будут экспрессировать белок, передающий лекарственную резистентность, и выживут при селекции. Примеры такой доминантной селекции используют лекарственные препараты неомицин (Southern et al., J. Molec. Appl. Genet. 1: 327 (1982)), микофеноловую кислоту (Mulligan et al., Science 209: 1422 (1980)) или гигромицин (Sugden, et al., Mol. Cell. Biol. 5: 410-413 (1985)).

Реагент нуклеиновой кислоты может включать один или несколько элементов селекции (*например*, элементов для селекции присутствия реагента нуклеиновой кислоты, а не для активации промоторного элемента, который можно селективно регулировать). Элементы селекции часто используют с использованием известных процессов для определения того, включен ли реагент нуклеиновой кислоты в клетку. В некоторых вариантах осуществления, реагент нуклеиновой кислоты включает два или несколько элементов селекции, где один эффективно функционирует в одном организме, а другой эффективно функционирует в другом организме. Примеры элементов селекции включают, но не ограничиваются ими: (1) сегменты нуклеиновых кислот, которые кодируют продукты, обеспечивающие резистентность к токсичным соединениям (*например*, антибиотикам); (2) сегменты нуклеиновых кислот, которые кодируют продукты, которых нет в клетке-реципиенте (*например*, жизненно важные продукты, гены тРНК, ауксотрофные маркеры); (3) сегменты нуклеиновых кислот, которые кодируют продукты, которые подавляют активность генного продукта; (4) сегменты нуклеиновых кислот, которые

кодируют легко идентифицируемые продукты (*например*, фенотипические маркеры, такие как антибиотики (*например*, β -лактамаза), β -галактозидаза, зеленый флуоресцентный белок (GFP), желтый флуоресцентный белок (YFP), красный флуоресцентный белок (RFP), голубой флуоресцентный белок (CFP) и белки клеточной поверхности); (5) сегменты нуклеиновых кислот, которые связывают продукты, которые в противном случае вредны для выживания и/или функции клеток; (6) сегменты нуклеиновых кислот, которые иначе ингибируют активность любого из сегментов нуклеиновой кислоты, описанных выше в пп. 1-5 (*например*, антисмысловые олигонуклеотиды); (7) сегменты нуклеиновых кислот, которые связывают продукты, модифицирующие субстрат (*например*, рестрикционные эндонуклеазы); (8) сегменты нуклеиновых кислот, которые можно использовать для выделения или идентификации желаемой молекулы (*например*, специфические сайты связывания с белком); (9) сегменты нуклеиновых кислот, которые кодируют специфическую нуклеотидную последовательность, которая в противном случае может быть нефункциональной (*например*, для ПЦР-амплификации субпопуляций молекул); (10) сегменты нуклеиновых кислот, которые, при отсутствии, прямо или косвенно придают резистентность или чувствительность к определенным соединениям; (11) сегменты нуклеиновых кислот, которые кодируют продукты, которые либо являются токсичными, либо превращают относительно нетоксичное соединение в токсичное соединение (*например*, тимидинкиназа Herpes simplex, цитозиндезаминаза) в клетках-реципиентах; (12) сегменты нуклеиновых кислот, которые ингибируют репликацию, разделение или наследуемость молекул нуклеиновых кислот, которые их содержат; и/или (13) сегменты нуклеиновых кислот, которые кодируют условные функции репликации, *например*, репликацию в определенных хозяевах или штаммах клеток-хозяев или при определенных условиях окружающей среды (*например*, температуре, условиях питания и подобных).

Реагент нуклеиновой кислоты может быть в любой форме, пригодной для транскрипции и/или трансляции *in vivo*. Нуклеиновая кислота иногда представляет собой плазмиду, такую как сверхспиральная плазида, иногда представляет собой искусственную хромосому дрожжей (*например*, YAC), иногда представляет собой линейную нуклеиновую кислоту (*например*, линейную нуклеиновую кислоту, полученную ПЦР или рестриктазными фрагментами), иногда она является одноцепочечной, и иногда является двухцепочечной. Реагент нуклеиновой кислоты иногда получают с помощью процесса амплификации, такого как процесс полимеразной цепной реакции (ПЦР) или процесс амплификации, опосредованный транскрипцией (ТМА). В ТМА два фермента используются в изотермической реакции для продуцирования продуктов амплификации, определяемых испусканием света (*например*, Biochemistry 1996 Jun 25;35(25):8429-38). Стандартные ПЦР процессы известны (*например*, патенты США №№ 4,683,202; 4,683,195; 4,965,188; и 5,656,493) и, как правило, выполняются циклами. Каждый цикл включает денатурацию нагреванием, при которой диссоциируют гибридные нуклеиновые кислоты, охлаждение, при котором олигонуклеотиды праймера гибридизируются; и удлинение олигонуклеотидов полимеразой (т.е. полимеразой Taq). Примером циклического процесса

ПЦР является обработка образца при 9°C в течение 5 минут; повторение сорока пяти циклов при 95°C в течение 1 минуты, 59°C в течение 1 минуты, 10 секунд и 72°C в течение 1 минуты 30 секунд; и затем обработка образца при 72°C в течение 5 минут. Несколько циклов часто проводят с использованием коммерчески доступного термоциклера. Иногда продукты ПЦР амплификации хранят в течение время при более низкой температуре (*например*, при 4°C) и иногда перед анализом замораживают (*например*, при -20°C).

Стратегии клонирования, аналогичные описанным выше, могут быть использованы для получения ДНК, содержащей нуклеотиды не природного происхождения. Например, олигонуклеотиды, содержащие нуклеотиды не природного происхождения в нужных положениях, синтезируют с использованием стандартного твердофазного синтеза и очищаются с помощью ВЭЖХ. Затем олигонуклеотиды встраиваются в плазмиду, содержащую требуемый контекст последовательности (т.е. UTR и кодирующую последовательность) с использованием способа клонирования (такого как Golden Gate Assembly) с сайтами клонирования, такими как сайты BsaI (хотя можно использовать и другие, обсуждавшиеся выше).

Наборы/готовые изделия

В некоторых вариантах осуществления, в настоящем документе описаны наборы и готовые изделия для использования с одним или несколькими способами, описанными в настоящем документе. Такие наборы включают носитель, упаковку или контейнер, разделенный на отсеки для приема одного или нескольких контейнеров, таких как флаконы, пробирки и подобные, где каждый из этих контейнеров содержит один из отдельных элементов, применяемых в способе, описанном в настоящем документе. Подходящие контейнеры включают, например, бутылки, флаконы, шприцы и пробирки. В одном варианте осуществления, контейнеры сформированы из различных материалов, таких как стекло или пластик.

В некоторых вариантах осуществления, набор включает подходящий упаковочный материал для размещения содержимого набора. В некоторых случаях, упаковочный материал предпочтительно изготавливается с использованием хорошо известных методов, чтобы обеспечить стерильную, свободную от загрязнений среду. Упаковочные материалы, используемые в настоящем документе, могут включать, например, те, которые обычно используются в коммерческих наборах, продаваемых для использования с системами секвенирования нуклеиновых кислот. Примеры упаковочных материалов включают, без ограничений, стекло, пластик, бумагу, фольгу и подобные, способные удерживать в фиксированных пределах набор компонентов, представленный в настоящем документе.

Упаковочный материал может включать этикетку, на которой указано конкретное использование компонентов. Использование набора, указанное на этикетке, может представлять собой один или несколько способов, изложенных в настоящем документе, в зависимости от конкретной комбинации компонентов, присутствующих в наборе. Например, на этикетке может быть указано, что набор можно использовать для способа синтеза полинуклеотида или для способа определения последовательности нуклеиновой

кислоты.

Инструкции по использованию упакованных реагентов или компонентов также могут быть включены в набор. Инструкции, как правило, включают материальное выражение, описывающее параметры реакции, такие как относительное количество компонентов набора и смешиваемого образца, периоды сохранения смесей реагента/образца, температуру, буферные условия и подобные.

Следует понимать, что не все компоненты, необходимые для конкретной реакции, должны присутствовать в конкретном наборе. Скорее, один или несколько дополнительных компонентов могут быть предоставлены из других источников. В инструкциях, прилагаемых к набору, могут быть указаны дополнительные компоненты, которые должны быть предоставлены, и где их можно получить.

В некоторых вариантах осуществления, представлен набор, пригодный для стабильного включения нуклеиновой кислоты не природного происхождения в клеточную нуклеиновую кислоту, *например*, с использованием способов, представленных настоящим описанием, для получения генетически сконструированных клеток. В одном варианте осуществления, набор, описанный в настоящем документе, включает генетически сконструированную клетку и одну или несколько нуклеиновых кислот не природного происхождения.

В дополнительных вариантах осуществления, набор, описанный в настоящем документе, представляет клетку и молекулу нуклеиновой кислоты, содержащую гетерологичный ген для введения в клетку, чтобы тем самым получить генетически сконструированную клетку, такую как векторы экспрессии, содержащие нуклеиновую кислоту из любого из вариантов осуществления, описанных выше в этом параграфе.

Пронумерованные варианты осуществления. Настоящее описание включает следующие неограничивающие пронумерованные варианты осуществления:

Вариант осуществления 1. Способ синтеза полипептида не природного происхождения, включающий:

а. получение, по меньшей мере, одной молекулы дезоксирибонуклеиновой кислоты (ДНК) не природного происхождения, содержащей, по меньшей мере, четыре пары оснований не природного происхождения;

б. транскрипцию, по меньшей мере, одной молекулы ДНК не природного происхождения с получением молекулы информационной рибонуклеиновой кислоты (иРНК), содержащей, по меньшей мере, два кодона не природного происхождения;

с. транскрибирование, по меньшей мере, одной молекулы ДНК не природного происхождения с получением, по меньшей мере, двух молекул транспортной РНК (тРНК), каждая из которых содержит, по меньшей мере, один антикодон не природного происхождения, где, по меньшей мере, две пары оснований не природного происхождения в соответствующей ДНК находятся в контекстах последовательностей так, что кодоны не природного происхождения молекулы мРНК комплементарны антикодону не природного происхождения каждой из молекул тРНК; и

d. синтез полипептида не природного происхождения путем трансляции молекулы мРНК не природного происхождения с использованием, по меньшей мере, двух молекул тРНК не природного происхождения, где каждый антикодон не природного происхождения направляет сайт-специфическое включение аминокислоты не природного происхождения в полипептид не природного происхождения.

Вариант осуществления 1.1. Способ синтеза полипептида не природного происхождения, включающий:

a. получение, по меньшей мере, одной молекулы дезоксирибонуклеиновой кислоты (ДНК) не природного происхождения, содержащей, по меньшей мере, четыре пары оснований не природного происхождения;

b. транскрипцию, по меньшей мере, одной молекулы ДНК не природного происхождения с получением молекулы информационной рибонуклеиновой кислоты (иРНК), содержащей, по меньшей мере, два кодона не природного происхождения;

c. транскрибирование, по меньшей мере, одной молекулы ДНК не природного происхождения с получением, по меньшей мере, двух молекул транспортной РНК (тРНК), каждая из которых содержит, по меньшей мере, один антикодон не природного происхождения, где, по меньшей мере, две пары оснований не природного происхождения в соответствующей ДНК находятся в контекстах последовательностей так, что кодоны не природного происхождения молекулы мРНК комплементарны антикодону не природного происхождения каждой из молекул тРНК и, по меньшей мере, один или несколько других кодонов не природного происхождения комплементарны антикодону не природного происхождения, по меньшей мере, других молекул тРНК; и

d. синтез полипептида не природного происхождения путем трансляции молекулы мРНК не природного происхождения с использованием, по меньшей мере, двух молекул тРНК не природного происхождения, где каждый антикодон не природного происхождения направляет сайт-специфическое включение аминокислоты не природного происхождения в полипептид не природного происхождения.

Вариант осуществления 2. Способ синтеза полипептида не природного происхождения, включающий:

a. предоставление, по меньшей мере, одной молекулы дезоксирибонуклеиновой кислоты (ДНК) не природного происхождения, содержащей, по меньшей мере, четыре пары оснований не природного происхождения, где, по меньшей мере, одна молекула ДНК не природного происхождения кодирует (i) молекулу информационной рибонуклеиновой кислоты (иРНК), содержащую, по меньшей мере, первый и второй кодоны не природного происхождения, и (ii) по меньшей мере, первую и вторую молекулы транспортной РНК (тРНК), где первая молекула тРНК содержит первый антикодон не природного происхождения, и вторая молекула тРНК содержит второй антикодон не природного происхождения, и, по меньшей мере, четыре пары оснований не природного происхождения в, по меньшей мере, одной молекуле ДНК находятся в таком контексте последовательности, что первый и второй кодоны не природного происхождения молекулы

мРНК комплементарны первому и второму антикодонам не природного происхождения, соответственно;

b. транскрипцию, по меньшей мере, одной молекулы ДНК не природного происхождения с получением мРНК;

c. транскрипцию, по меньшей мере, одной молекулы ДНК не природного происхождения с получением, по меньшей мере, первой и второй молекул тРНК; и

d. синтез полипептида не природного происхождения путем трансляции молекулы мРНК не природного происхождения с использованием, по меньшей мере, первой и второй молекул тРНК не природного происхождения, где каждый из, по меньшей мере, первого и второго антикодонов не природного происхождения направляет сайт-специфическое включение аминокислоты не природного происхождения в полипептид не природного происхождения.

Вариант осуществления 3. Способ по варианту осуществления 1, 1.1. или 2, где каждый из, по меньшей мере, двух кодонов не природного происхождения содержит первый нуклеотид не природного происхождения, расположенный в первом положении, втором положении или третьем положении кодона, необязательно, где первый нуклеотид не природного происхождения расположен во втором или третьем положении кодона.

Вариант осуществления 4. Способ по любому из предыдущих вариантов осуществления, где, по меньшей мере, два кодона не природного происхождения содержат последовательность нуклеиновой кислоты NNX или NXN, и антикодон не природного происхождения содержит последовательность нуклеиновой кислоты XNN, YNN, NXN или NYN для формирования пары кодон-антикодон не природного происхождения, содержащую NNX-XNN, NNX-YNN или NXN-NYN, где N представляет собой любой природный нуклеотид, X представляет собой первый нуклеотид не природного происхождения, и Y представляет собой второй нуклеотид не природного происхождения, отличный от первого нуклеотида не природного происхождения, с X-Y или X-X, образующими пару оснований не природного происхождения в ДНК.

Вариант осуществления 4.1. Способ по любому из предыдущих вариантов осуществления, где, по меньшей мере, два кодона не природного происхождения содержат последовательность нуклеиновой кислоты XNN, NXN, NNX, и антикодон не природного происхождения содержит последовательность нуклеиновой кислоты NNX, NNY, NXN, NYN, NNX, или NNY, для формирования пары кодон-антикодон не природного происхождения, включающей XNN-NNX, XNN-NNY, NXN-NXN, NXN-NYN, NNX-XNN или NNX-YNN, где N представляет собой любой природный нуклеотид, X представляет собой первый нуклеотид не природного происхождения, и Y представляет собой второй нуклеотид не природного происхождения, отличный от первого нуклеотида не природного происхождения, где X-X или X-Y образуют пару оснований не природного происхождения в ДНК.

Вариант осуществления 5. Способ по варианту осуществления 4, где кодон содержит, по меньшей мере, один G или C, и антикодон содержит, по меньшей мере, один

комплементарный С или G.

Вариант осуществления 6. Способ по варианту осуществления 4 или 5, где X и Y независимо выбраны из группы, состоящей из:

(i) 2-тиоурацила, 2'-дезоксуридина, 4-тиоурацила, урацил-5-ила, гипоксантин-9-ила (I), 5-галоурацила, 5-пропилиурацила, 6-азоурацила, 5-метиламинометилурацила, 5-метоксиаминометил-2-тиоурацила, псевдоурацила, метилового эфира урацил-5-оксуксусной кислоты, урацил-5-оксуксусной кислоты, 5-метил-2-тиоурацила, 3-(3-амино-3-N-2-карбоксивпропил)урацила, 5-метил-2-тиоурацила, 4-тиоурацила, 5-метилурацила, 5'-метоксикарбоксиметилурацила, 5-метоксиурацила, урацил-5-оксуксусной кислоты, 5-(карбоксихидроксиметил)урацила, 5-карбоксиметиламинометил-2-тиоуридина, 5-карбоксиметиламинометилурацила или дигидроурацила;

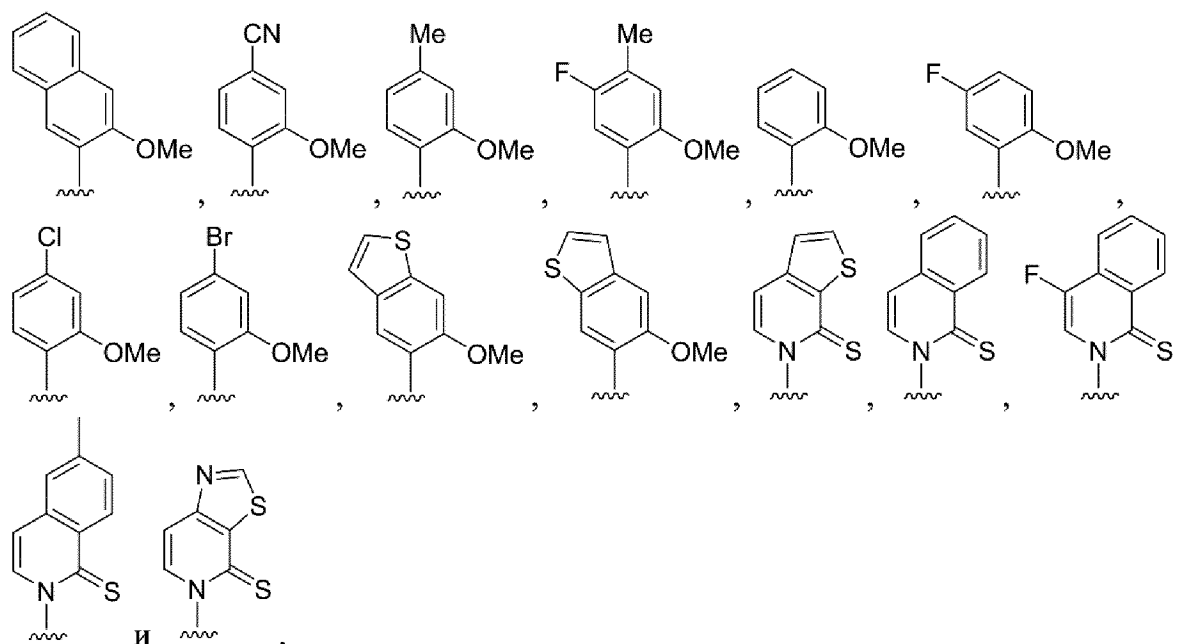
(ii) 5-гидроксиметилцитозина, 5-трифторметилцитозина, 5-галоцитозина, 5-пропилилцитозина, 5-гидроксицитозина, циклоцитозина, цитозина арабинозида, 5,6-дигидроцитозина, 5-нитроцитозина, 6-азоцитозина, азацитозина, N4-этилцитозина, 3-метилцитозина, 5-метилцитозина, 4-ацетилцитозина, 2-тиоцитозина, феноксазина цитидина ([5,4-b][1,4]бензоксазин-2(3H)-она), фенотиазина цитидин (1H-пиримидо[5,4-b][1,4]бензотиазин-2(3H)-она), феноксазина цитидина (9-(2-аминоэтокси)-H-пиримидо[5,4-b][1,4]бензоксазин-2(3H)-она), карбазола цитидина (2H-пиримидо[4,5-b]индол-2-она) или пиридоиндола цитидина (H-пиридо[3',2':4,5]пирроло[2,3-d]пиримидин-2-она);

(iii) 2-аминоаденина, 2-пропиладенина, 2-аминоаденина, 2-F-аденина, 2-аминопропиладенина, 2-амино-2'-дезоксаденозина, 3-дезааденина, 7-метиладенина, 7-дезааденина, 8-азааденина, 8-галоген-, 8-амино-, 8-тиол-, 8-тиоалкил- и 8-гидрокси-замещенных аденинов, N6-изопентениладенина, 2-метиладенина, 2,6-диаминопурина, 2-метилтио-N6-изопентениладенина или 6-азааденина;

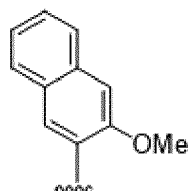
(iv) 2-метилгуанина, 2-пропила и алкильных производных гуанина, 3-дезагуанина, 6-тиогуанина, 7-метилгуанина, 7-дезагуанина, 7-дезагуанозина, 7-деза-8-азагуанина, 8-азагуанина, 8-галоген-, 8-амино-, 8-тиол-, 8-тиоалкил- и 8-гидрокси-замещенных гуанинов, 1-метилгуанина, 2,2-диметилгуанина, 7-метилгуанина или 6-азагуанина; и

(v) гипоксантина, ксантина, 1-метилюозина, кеозина, бета-D-галактозилкеозина, инозина, бета-D-маннозилкеозина, вибутоксозина, гидроксимочевины, (аср3)w, 2-аминопиридина или 2-пиридона.

Вариант осуществления 7. Способ по варианту осуществления 4 или 5, где основания, содержащие каждый из X и Y, независимо выбраны из группы, состоящей из:

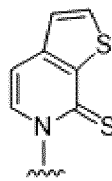


Вариант осуществления 8. Способ по варианту осуществления 7, где основание,



содержащее каждый X, представляет собой

Вариант осуществления 9. Способ по варианту осуществления 7 или 8, где



основание, содержащее каждый Y, представляет собой

Вариант осуществления 10. Способ по любому из вариантов осуществления 4-9, где NNX-XNN выбран из группы, состоящей из UUX-XAA, UGX-XCA, CGX-XCG, AGX-XCU, GAX-XUC, CAX-XUG, AUX-XAU, CUX-XAG, GUX-XAC, UAX-XUA и GGX-XCC.

Вариант осуществления 11. Способ по любому из вариантов осуществления 4-9, где NNX-YNN выбран из группы, состоящей из UUX-YAA, UGX-YCA, CGX-YCG, AGX-YCU, GAX-YUC, CAX-YUG, AUX-YAU, CUX-YAG, GUX-YAC, UAX-YUA и GGX-YCC.

Вариант осуществления 12. Способ по любому из вариантов осуществления 4-9, где NXN-NYN выбран из группы, состоящей из GXU-AYC, CXU-AYG, GXG-CYC, AXG-CYU, GXC-GYC, AXC-GYU, GXA-UYC, CXC-GYG и UXC-GYA.

Вариант осуществления 13. Способ по варианту осуществления 12, где NXN-NYN выбран из группы, состоящей из AXG-CYU, GXC-GYC, AXC-GYU, GXA-UYC, CXC-GYG и UXC-GYA.

Вариант осуществления 13.1. Способ по любому из вариантов осуществления 4.1-9, где XNN-NNY выбран из группы, состоящей из XUU-AAAY, XUG-CAAY, XCG-CGY, XAG-CUY, XGA-UCY, XCA-UGY, XAU-AUY, XCU-AGY, XGU-ACY, XUA-UAY, XUC-GAY,

XCC-GGY, XAA-UUY, XAC-GUY, XGC-GCY, XGG-CCY и XGG-CCY.

Вариант осуществления 13.2. Способ по любому из вариантов осуществления 4.1-9, в котором XNN-NNX выбран из группы, состоящей из XUU-AAX, XUG-CAX, XCG-CGX, XAG-CUX, XGA-UCX, XCA-UGX, XAU-AUX, XCU-AGX, XGU-ACX, XUA-UAX, XUC-GAX, XCC-GGX, XAA-UUX, XAC-GUX, XGC-GCX, XGG-CCX и XGG-CCX.

Вариант осуществления 14. Способ по любому из предшествующих вариантов осуществления, где, по меньшей мере, две молекулы тРНК не природного происхождения содержат разные антикодоны не природного происхождения.

Вариант осуществления 15. Способ по варианту осуществления 14, где, по меньшей мере, две молекулы тРНК не природного происхождения содержат пирролизил тРНК из рода *Methanosarcina* и тирозил тРНК из *Methanocaldococcus jannaschii* или их производные.

Вариант осуществления 16. Способ по любому из вариантов осуществления 13, 14 или 15, включающий зарядку, по меньшей мере, двух молекул тРНК не природного происхождения с помощью аминоксил тРНК синтетазы.

Вариант осуществления 17. Способ по варианту осуществления 16, где аминоксил тРНК синтетаза выбрана из группы, состоящей из химерного PylRS (chPylRS) и *M. jannaschii* AzFRS (*MjpAzFRS*).

Вариант осуществления 18. Способ по варианту осуществления 14 или 15, включающий зарядку, по меньшей мере, двух молекул тРНК не природного происхождения, по меньшей мере, двумя тРНК синтетазами.

Вариант осуществления 19. Способ по варианту осуществления 18, где, по меньшей мере, две тРНК синтетазы включают химерную PylRS (chPylRS) и *M. jannaschii* AzFRS (*MjpAzFRS*).

Вариант осуществления 20. Способ по любому из вариантов осуществления 1-19, где полипептид не природного происхождения содержит две, три или несколько аминокислот не природного происхождения.

Вариант осуществления 21. Способ по любому из вариантов осуществления 1-20, где полипептид не природного происхождения содержит, по меньшей мере, две одинаковые аминокислоты не природного происхождения.

Вариант осуществления 22. Способ по любому из вариантов осуществления 1-20, где полипептид не природного происхождения содержит, по меньшей мере, две различные аминокислоты не природного происхождения.

Вариант осуществления 23. Способ по любому из вариантов осуществления 1-22, где аминокислота не природного происхождения содержит

аналог лизина;

ароматическую боковую цепь;

азидогруппу;

алкиновую группу; или

альдегидную или кетонную группу.

Вариант осуществления 24. Способ по любому из вариантов осуществления 1-22,

где аминокислота не природного происхождения не содержит ароматическую боковую цепь.

Вариант осуществления 25. Способ по любому из вариантов осуществления 1-22, где аминокислота не природного происхождения выбрана из N6-((азидоэтокси)-карбонил)-L-лизина (AzK), N6-пропаргилэтокси-карбонил-L-лизина (PraK), N6-(пропаргилокси)карбонил-L-лизина (PrK), п-азидофенилаланина (pAzF), BCN-L-лизина, норборненилизина, TCO-лизина, метилтетразинлизина, аллилоксикарбониллизина, 2-амино-8-оксононановой кислоты, 2-амино-8-оксооктановой кислоты, п-ацетил-L-фенилаланина, п-азидометил-L-фенилаланина (pAMF), п-йодо-L-фенилаланина, м-ацетилфенилаланина, 2-амино-8-оксононаноевой кислоты, п-пропаргилоксифенилаланина, п-пропаргилфенилаланина, 3-метилфенилаланина, L-Dopa, фторированного фенилаланина, изопропил-L-фенилаланина, п-азидо-L-фенилаланина, п-ацил-L-фенилаланина, п-бензоил-L-фенилаланина, п-бромфенилаланина, п-амино-L-фенилаланина, изопропил-L-фенилаланина, O-аллилтирозина, O-метил-L-тирозина, O-4-аллил-L-тирозина, 4-пропил-L-тирозина, фосфонтирозина, три-O-ацетил-GlcNAc-серина, L-фосфосерина, фосфоносерина, L-3-(2-нафтил)аланина, 2-амино-3-((2-((3-(бензилокси)-3-оксопропил)амино)этил)селанил)пропановой кислоты, 2-амино-3-(фенилселанил)пропановой кислоты, селеноцистеина, N6-(((2-азидобензил)окси)карбонил)-L-лизина, N6-(((3-азидобензил)окси)карбонил)-L-лизина и N6-(((4-азидобензил)окси)карбонил)-L-лизина.

Вариант осуществления 26. Способ по любому из предшествующих вариантов осуществления, где, по меньшей мере, одна молекула ДНК не природного происхождения находится в форме плазмиды.

Вариант осуществления 27. Способ по любому из вариантов осуществления 1-26, где, по меньшей мере, одна молекула ДНК не природного происхождения интегрируется в геном клетки.

Вариант осуществления 28. Способ согласно варианту осуществления 26 или 27, где, по меньшей мере, одна молекула ДНК не природного происхождения кодирует полипептид не природного происхождения.

Вариант осуществления 29. Способ по любому из предыдущих вариантов осуществления, включающий репликацию и транскрипцию *in vivo* молекулы ДНК не природного происхождения и трансляцию *in vivo* транскрибированной молекулы мРНК в клеточном организме.

Вариант осуществления 30. Способ по варианту осуществления 29, в котором клеточный организм представляет собой микроорганизм.

Вариант осуществления 31. Способ по варианту осуществления 30, где клеточный организм представляет собой прокариот.

Вариант осуществления 32. Способ по варианту осуществления 31, где клеточный организм представляет собой бактерию.

Вариант осуществления 33. Способ по варианту осуществления 32, где клеточный

организм представляет собой грамположительную бактерию.

Вариант осуществления 34. Способ по варианту осуществления 32, где клеточный организм представляет собой грамотрицательную бактерию.

Вариант осуществления 35. Способ по варианту осуществления 34, где клеточный организм представляет собой *Escherichia coli*.

Вариант осуществления 36. Способ по любому из предыдущих вариантов осуществления, где, по меньшей мере, две пары оснований не природного происхождения включают пары оснований, выбранные из dCNMO-dTPT3, dNaM-dTPT3, dCNMO-dTAT1 или dNaM-dTAT1.

Вариант осуществления 37. Способ по любому из вариантов осуществления 29-36, где клеточный организм содержит транспортер нуклеозидтрифосфата.

Вариант осуществления 38. Способ по варианту осуществления 37, где транспортер нуклеозидтрифосфата содержит аминокислотную последовательность *PtNTT2*.

Вариант осуществления 39. Способ по варианту осуществления 38, где переносчик нуклеозидтрифосфата содержит усеченную аминокислотную последовательность *PtNTT2*.

Вариант осуществления 40. Способ по варианту осуществления 39, отличающийся тем, что усеченная аминокислотная последовательность *PtNTT2*, по меньшей мере, на 80% идентична *PtNTT2*, кодируемой SEQ ID NO. 1.

Вариант осуществления 41. Способ по любому из вариантов осуществления 29-40, где клеточный организм содержит, по меньшей мере, одну молекулу ДНК не природного происхождения.

Вариант осуществления 42. Способ по варианту осуществления 41, где, по меньшей мере, одна молекула ДНК не природного происхождения содержит, по меньшей мере, одну плазмиду.

Вариант осуществления 43. Способ по варианту осуществления 42, где, по меньшей мере, одна молекула ДНК не природного происхождения интегрируется в геном клетки.

Вариант осуществления 44. Способ по варианту осуществления 42 или 43, где, по меньшей мере, одна молекула ДНК не природного происхождения кодирует полипептид не природного происхождения.

Вариант осуществления 45. Способ по любому из вариантов осуществления 1-26, где способ представляет собой способ *in vitro*, включающий синтез полипептида не природного происхождения с помощью бесклеточной системы.

Вариант осуществления 46. Способ по любому из предыдущих вариантов осуществления, где пары оснований не природного происхождения содержат, по меньшей мере, один нуклеотид не природного происхождения, содержащий сахарную группу не природного происхождения.

Вариант осуществления 47. Способ по варианту осуществления 46, где фрагмент сахара не природного происхождения содержит группу, выбранную из группы, состоящей из: OH, замещенного низшего алкила, алкарила, аралкила, O-алкарила или O-аралкила, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂F;

O-алкила, S-алкила, N-алкила;

O-алкенила, S-алкенила, N-алкенила;

O-алкинила, S-алкинила, N-алкинила;

O-алкил-O-алкила, 2'-F, 2'-OCH₃, 2'-O(CH₂)₂OCH₃, где алкил, алкенил и алкинил могут быть замещены или не замещены C₁-C₁₀ алкилом, C₂-C₁₀ алкенилом, C₂-C₁₀ алкинилом, O[(CH₂)_nO]_mCH₃, -O(CH₂)_nOCH₃, -O(CH₂)_nNH₂, -O(CH₂)_nCH₃, -O(CH₂)_n-NH₂ и -O(CH₂)_nON[(CH₂)_nCH₃]₂, где n и m составляют от 1 до примерно 10;

и/или модификации в положении 5':

5'-винила, 5'-метила (R или S);

модификации в положении 4':

4'-S, гетероциклоалкила, гетероциклоалкарила, аминокламинами, полиалкиламинами, замещенного силила, группы, расщепляющей РНК, репортерной группы, интеркалятора, группы для улучшения фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида, и любой их комбинации.

Вариант осуществления 48. Клетка, содержащая, по меньшей мере, одну молекулу ДНК не природного происхождения, по меньшей мере, из четырех пар оснований не природного происхождения, где, по меньшей мере, одна молекула ДНК не природного происхождения кодирует (i) молекулу информационной рибонуклеиновой кислоты (иРНК), кодирующую полипептид не природного происхождения и содержащую, по меньшей мере, первый и второй кодоны не природного происхождения и (ii) по меньшей мере, первую и вторую молекулы транспортной РНК (тРНК), где первая молекула тРНК содержит первый антикодон не природного происхождения, и вторая молекула тРНК содержит второй антикодон не природного происхождения, и, по меньшей мере, четыре пары оснований не природного происхождения, по меньшей мере, в одной молекуле ДНК находятся в таком контексте последовательности, что первый и второй кодоны не природного происхождения молекулы мРНК комплементарны первому и второму антикодонам не природного происхождения, соответственно.

Вариант осуществления 49. Клетка по варианту осуществления 48, дополнительно содержащая молекулу мРНК и, по меньшей мере, первую и вторую молекулы тРНК.

Вариант осуществления 50. Клетка по варианту осуществления 49, где, по меньшей мере, первая и вторая молекулы тРНК ковалентно связаны с аминокислотами не природного происхождения.

Вариант осуществления 51. Клетка по варианту осуществления 50 дополнительно содержит полипептид не природного происхождения.

Вариант осуществления 52. Клетка, содержащая:

а. по меньшей мере, две разные пары кодон-антикодон не природного происхождения, где каждая пара кодон-антикодон не природного происхождения содержит кодон не природного происхождения из матричной РНК (мРНК) не природного происхождения и антикодон не природного происхождения из транспортной рибонуклеиновой кислоты (тРНК) не природного происхождения, где указанный кодон не

природного происхождения содержит первый нуклеотид не природного происхождения, и указанный антикодон не природного происхождения, содержащий второй нуклеотид не природного происхождения; и

в. по меньшей мере, две разные аминокислоты не природного происхождения, каждая из которых ковалентно связана с соответствующей тРНК не природного происхождения.

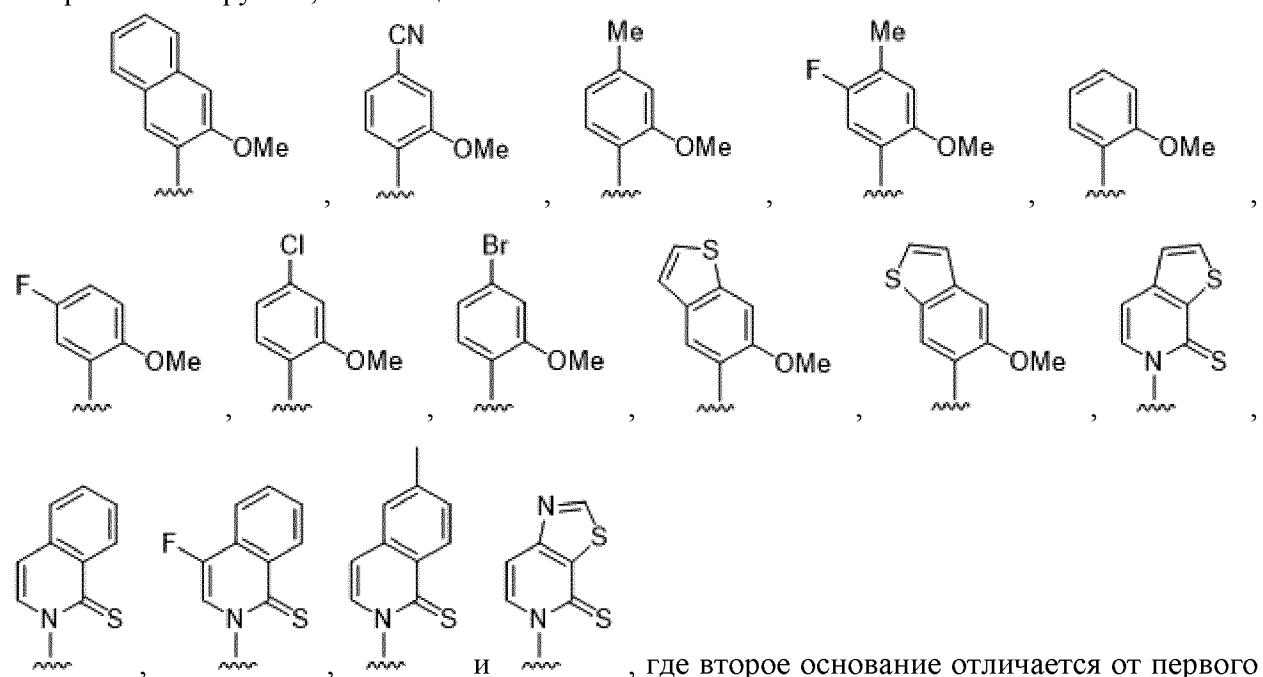
Вариант осуществления 53. Клетка по варианту осуществления 52, дополнительно содержащая, по меньшей мере, одну молекулу ДНК не природного происхождения, содержащую, по меньшей мере, четыре пары оснований не природного происхождения (UBP).

Вариант осуществления 54. Клетка по любому из вариантов осуществления 48-53, где первый нуклеотид не природного происхождения расположен во втором или третьем положении кодона не природного происхождения.

Вариант осуществления 54.1 Клетка по любому из вариантов осуществления 48-53, где первый нуклеотид не природного происхождения расположен в первом, втором или третьем положении кодона не природного происхождения.

Вариант осуществления 55. Клетка по варианту осуществления 54 или 54.1, где первый нуклеотид не природного происхождения представляет собой комплементарное основание, спаренное с вторым нуклеотидом не природного происхождения антикодона не природного происхождения.

Вариант осуществления 56. Клетка по любому из вариантов осуществления 48-55, где первый нуклеотид не природного происхождения и второй нуклеотид не природного происхождения содержат первое и второе основания соответственно, независимо выбранные из группы, состоящей из



, где второе основание отличается от первого основания.

Вариант осуществления 57. Клетка по любому из вариантов осуществления 48 или 50-56, где, по меньшей мере, четыре пары оснований не природного происхождения независимо выбраны из группы, состоящей из dCNMO-dTPT3, dNaM-dTPT3, dCNMO-dTAT1 или dNaM-dTAT1.

Вариант осуществления 58. Клетка по любому из вариантов осуществления 48 или 50-57, где, по меньшей мере, одна молекула ДНК не природного происхождения содержит, по меньшей мере, одну плазмиду.

Вариант осуществления 59. Клетка по любому из вариантов осуществления 48 или 50-58, где, по меньшей мере, одна молекула ДНК не природного происхождения интегрирована в геном клетки.

Вариант осуществления 60. Клетка по любому из вариантов осуществления 50-59, где, по меньшей мере, одна молекула ДНК не природного происхождения кодирует полипептид не природного происхождения.

Вариант осуществления 61. Клетка по любому из вариантов реализации 48-60, где клетка экспрессирует переносчик нуклеозидтрифосфата.

Вариант осуществления 62. Клетка по варианту осуществления 61, где транспортер нуклеозидтрифосфата содержит аминокислотную последовательность *PtNTT2*.

Вариант осуществления 63. Способ по варианту осуществления 62, где транспортер нуклеозидтрифосфатов содержит усеченную аминокислотную последовательность *PtNTT2*.

Вариант осуществления 64. Способ по варианту осуществления 63, где усеченная аминокислотная последовательность *PtNTT2*, по меньшей мере, на 80% идентична *PtNTT2*, кодируемой SEQ ID NO. 1.

Вариант осуществления 65. Клетка по любому из вариантов осуществления 48-64, где клетка экспрессирует, по меньшей мере, две тРНК синтетазы.

Вариант осуществления 66. Клетка по варианту осуществления 65, где, по меньшей мере, две тРНК синтетазы представляют собой химерную PylRS (chPylRS) и *M. jannaschii* AzFRS (*MjpAzFRS*).

Вариант осуществления 67. Клетка по любому из вариантов осуществления 48-66, где клетка содержит нуклеотиды не природного происхождения, включающие сахарную группу не природного происхождения.

Вариант осуществления 68. Клетка по варианту осуществления 67, где сахарная группа не природного происхождения выбрана из группы, состоящей из:

модификации в положении 2':

ОН, замещенного низшего алкила, алкарила, аралкила, О-алкарила или О-аралкила, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂F;

О-алкила, S-алкила, N-алкила;

О-алкенила, S-алкенила, N-алкенила;

О-алкинила, S-алкинила, N-алкинила;

О-алкил-О-алкила, 2'-F, 2'-OCH₃, 2'-O(CH₂)₂OCH₃, где алкил, алкенил и алкинил

могут быть замещены или не замещены C₁-C₁₀ алкилом, C₂-C₁₀ алкенилом, C₂-C₁₀ алкинилом, O[(CH₂)_nO]_mCH₃, -O(CH₂)_nOCH₃, -O(CH₂)_nNH₂, -O(CH₂)_nCH₃, -O(CH₂)_n-NH₂ и -O(CH₂)_nON[(CH₂)_nCH₃]₂, где n и m составляют от 1 до примерно 10;

и/или модификации в положении 5':

5'-винила, 5'-метила (R или S);

модификации в положении 4':

4'-S, гетероциклоалкила, гетероциклоалкарила, аминоквалиамино, полиалкиламино, замещенного силила, группы, расщепляющей РНК, репортерной группы, интеркалятора, группы для улучшения фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида, и любой их комбинации.

Вариант осуществления 69. Клетка по любому из вариантов осуществления 48-68, где, по меньшей мере, одно нуклеотидное основание не природного происхождения распознается РНК полимеразой во время транскрипции.

Вариант осуществления 70. Клетка по любому из вариантов осуществления 48-69, где клетка транслирует, по меньшей мере, один полипептид не природного происхождения, содержащий, по меньшей мере, две аминокислоты не природного происхождения.

Вариант осуществления 71. Клетка по любому из вариантов осуществления 48-70, где, по меньшей мере, две аминокислоты не природного происхождения независимо выбраны из группы, состоящей из N6-((азидоэтокси)-карбонил)-L-лизина (AzK), N6-пропаргилэтокси-карбонил-L-лизина (PraK), N6-(пропаргилокси)карбонил-L-лизина (PrK), п-азидофенилаланина (pAzF), BCN-L-лизина, норборненилизина, TCO-лизина, метилтетразинлизина, аллилоксикарбониллизина, 2-амино-8-оксононановой кислоты, 2-амино-8-оксооктановой кислоты, п-ацетил-L-фенилаланина, п-азидометил-L-фенилаланина (pAMF), п-йодо-L-фенилаланина, м-ацетилфенилаланина, 2-амино-8-оксононаноевой кислоты, п-пропаргиллоксифенилаланина, п-пропаргилфенилаланина, 3-метилфенилаланина, L-Дора, фторированного фенилаланина, изопропил-L-фенилаланина, п-азидо-L-фенилаланина, п-ацил-L-фенилаланина, п-бензоил-L-фенилаланина, п-бромфенилаланина, п-амино-L-фенилаланина, изопропил-L-фенилаланина, O-аллилтирозина, O-метил-L-тирозина, O-4-аллил-L-тирозина, 4-пропил-L-тирозина, фосфонтирозина, три-O-ацетил-GlcNAc-серина, L-фосфосерина, фосфоносерина, L-3-(2-нафтил)аланина, 2-амино-3-((2-((3-(бензилокси)-3-оксопропил)амино)этил)селанил)пропановой кислоты, 2-амино-3-(фенилселанил)пропановой кислоты, селеноцистеина, N6-(((2-азидобензил)окси)карбонил)-L-лизина, N6-(((3-азидобензил)окси)карбонил)-L-лизина и N6-(((4-азидобензил)окси)карбонил)-L-лизина.

Вариант осуществления 72. Клетка по любому из вариантов осуществления 48-71, отличающаяся тем, что клетка выделена.

Вариант осуществления 73. Клетка по любому из вариантов осуществления 48-72, где клетка представляет собой прокариот.

Вариант осуществления 74. Линия клеток, содержащая клетку по любому из

вариантов осуществления 48-73.

ПРИМЕРЫ

Пример 1. Исходный скрининг кодона

Зеленый флуоресцентный белок и его варианты, такие как sfGFP, используют в качестве модельных систем для исследования включения ncAA, особенно в положении Y151, которое, как было показано, допускает множество природных и ncAA замен. Плазмиды сконструированы так, чтобы содержать два dNaM-dTPT3 UBP, один из которых расположен в кодоне 151 sfGFP, и другой расположен для кодирования антикодона *M. mazei* тРНК^{Pyl} (фиг. 6С), который был селективно заряжен с помощью PylRS с ncAA N6-((2-азидоэтокси)карбонил)-L-лизина (AzK) (фиг. 6В). Плазмиды конструируют для изучения декодирования шести кодонов, включая два кодона не природного происхождения в первом положении (ХТС и ХТГ; Х относится к dNaM), два кодона не природного происхождения во втором положении (АХС и GХА), и два кодона не природного происхождения в третьем положении (АGХ и САХ), а также кодоны контекста противоположной цепи (УТС, УТГ, АУС, GУА, АGУ и САУ; У относится к dTPT3).

В то время как клональные популяции SSO способны продуцировать большие количества чистого белка не природного происхождения, вероятно, из-за устранения плазмид, которые были неправильно собраны во время конструирования *in vitro*, для облегчения первоначальной экспрессии белка скрининга кодонов впервые исследуют с не клональной популяцией клеток, и продуцирование белка анализируют сразу после трансформации. Плазмиды используют для трансформации *E. coli* ML2 (BL21(DE3) lacZYA:*P*tNTT2(66-575) ΔrecA polB⁺⁺), которая содержит дополнительную плазмиду, кодирующую химерную пирролизил-тРНК синтетазу (chPylRS^{IPYE}) и после роста до ранней стационарной фазы в селективных средах с добавлением dNaMTP и dTPT3TP, клетки переносят на свежие среды. После роста до средней экспоненциальной фазы, в культуру добавляют NaMTP, TPT3TP, AzK, и изопропил-β-D-тиогалактозид (IPTG) добавляют для индукции экспрессии РНК полимеразы T7 (T7 RNAP), chPylRS^{IPYE} и тРНК^{Pyl}. Через 1 ч дополнительного роста, ангидротетрациклин (aTc) добавляют для индукции экспрессии sfGFP, которую отслеживают флуоресценцию.

Кодоны первого положения не показывают значительную флуоресценцию в отсутствие или в присутствии AzK, независимо от того, была ли предпринята попытка декодирования с помощью гетероспаривания или самоспаривания антикодонов (например, тРНК^{Pyl}(CAУ) или тРНК^{Pyl}(CAХ), соответственно, для ХТГ) (фиг. 10). Кодоны с dNaM во втором положении показывают незначительную флуоресценцию в отсутствие AzK, но в его присутствии показывают значительную флуоресценцию при декодировании с тРНК^{Pyl}, перекодированной с помощью гетероспаренных антикодонов тРНК^{Pyl}(GYT) или тРНК^{Pyl}(TУС), но не с самоспаривающимися антикодонами тРНК^{Pyl}(GYT) или тРНК^{Pyl}(ТХС). С dTPT3 во втором положении, флуоресценция не наблюдается с или без добавления AzK, независимо от того, была ли предпринята попытка декодирования с помощью гетероспаривания или самоспаривания тРНК. Кодоны в третьем положении САХ

и САУ показывают высокую флуоресценцию в отсутствие AzK и неожиданно показывают меньшую флуоресценцию при ее добавлении, независимо от того, была ли предпринята попытка декодирования с помощью гетероспаривания или самоспаривания тРНК^{Pyl}. Этот результат позволяет предположить, что соответствующее третье положение тРНК не природного происхождения непродуктивно связывается на рибосоме и блокирует прочтение кодона не природного происхождения природной тРНК. В отсутствие AzK, AGX и AGY показывают незначительную флуоресценцию, и AGX с тРНК^{Pyl}(ХСТ) показывает увеличение флуоресценции с добавлением AzK.

Поскольку кодоны в первом положении не представляются многообещающими, проводят более полный скрининг кодонов второго положения. Поскольку первоначальный анализ показал потенциальное декодирование только с NaM в кодоне и с TPT3 в антикодоне, исследуют кодоны NXN и родственную тРНК^{Pyl}(NYN). Из 16 возможных кодонов исключают CXА, CXG и TXG, так как соответствующий контекст последовательности плохо сохраняется в ДНК SSO. В соответствии с предыдущими результатами, в отсутствие AzK, использование кодонов АХС и GXC дает незначительное или не дает флуоресценцию, в то время как в присутствии AzK они дают значительное повышение флуоресценции (**фиг. 6D**). Аналогично, с кодонами GXT, CXС, TXС, GXG, GXA, СХТ и АХG, добавление AzK приводит к значительному увеличению флуоресценции по сравнению с отсутствием AzK. Остальные четыре кодона, АХА, АХТ, ТХА и ТХТ, вызывают незначительную флуоресценцию независимо от того, был ли добавлен AzK или нет, что указывает на строгое требование, по меньшей мере, одной пары G-C.

Для скрининга продукции белка не природного происхождения, sfGFP очищают с помощью С-концевой аффинной метки StrepII и подвергают штамм-стимулированной реакции азид-алкинового циклоприсоединения (SPAAC) с дибензоциклооктином (DBCO), связанным с родаминовым красителем (TAMRA) четырьмя PEG единицами (DBCO-PEG4-TAMRA). Как было показано ранее, успешная конъюгация не только метит белки, содержащие psAA, с обнаруживаемым флуорофором, но также вызывает определяемый сдвиг в электрофоретической подвижности, что позволяет количественно определить белок, содержащий AzK, по отношению к общему продуцированному белку (т.е. точности включения psAA; **фиг. 6D**). В соответствии с предыдущими результатами, использование кодонов GXC и АХС дает продуцирование значительных количеств sfGFP с остатком AzK. Примечательно, семь дополнительных кодонов не природного происхождения, GXT, CXС, TXС, GXG, GXA, СХТ и АХG, также дают значительные уровни белка не природного происхождения (**фиг. 6D, фиг. 11**).

Наконец, проводят более полный скрининг кодонов в третьем положении. Поскольку в начальном скрининге оказалось, что декодируется только AGX, и только затем с помощью самоспаривающейся тРНК^{Pyl}(ХСТ), далее исследуют кодоны с dNaM в третьем положении кодона с родственной самоспаривающейся тРНК^{Pyl}(XNN) (**фиг. 6C**). Кодоны NCX исключают, поскольку они приводят к контекстам последовательностей NCXA, которые, как отмечалось выше, плохо сохраняются в ДНК SSO. В соответствии с

первоначальным анализом, в отсутствие AzK эти кодоны обычно дают большую флуоресценцию, чем наблюдают с кодонами во втором положении, но в присутствии AzK наблюдается переменное увеличение флуоресценции (**фиг. 6D**). Независимо от того, когда белок был выделен и проанализирован, как описано выше, использование CGX, ATX, CAX, AGX, GAX, TGX, CTX, TTX, GTX или TAX приводит к значительным уровням продуцирования белка не природного происхождения (**фиг. 6D, фиг. 11**). Кодон GGX продуцирует множественные смещенные виды, позволяющие предположить, что тРНК^{Pyl}(XCC) декодирует один или несколько природных кодонов. При использовании кодона AAX не определяют никакой белок не природного происхождения.

Пример 2. Характеристика кодонов в клональных SSO

Чтобы выбрать наиболее многообещающие пары кодон/антикодон, идентифицированные в описанном выше скрининге кодонов, сравнивают наблюдаемую флуоресценцию в присутствии AzK и индуцированный сдвиг подвижности в выделенном белке (**фиг. 6D, вставка**). На основании этого анализа, семь пар кодонов/антикодонов не природного происхождения, GXC/GYC, GXT/AYC, AXC/GYT, AGX/XCT, CGX/XCG, TTX/XAA, TGX/XCA, выбирают для дальнейшей характеристики. Эти пары кодон/антикодон исследуют в клональных SSO, что исключает клетки, которые были трансформированы неправильно собранными плазмидами или плазмидами, утратившими UBP во время конструирования *in vitro*. Клональные SSO получают «штрихованием» трансформантов в твердую среду для роста, содержащую dNaMTP и dTPTЗTP, выбирая отдельные колонии и подтверждая целостность плазмиды и высокое удержание UBP. Клоны с высоким удержанием повторно выращивают и индуцируют для продуцирования белка, как описано выше. Примечательно, что наблюдаемая флуоресценция указывает на то, что каждая из семи пар кодон/антикодон продуцирует белок на уровне, который благоприятно сравнивается с контролем подавления янтаря и, кроме того, анализ сдвига в геле демонстрирует, что практически весь sfGFP содержит ncAA (**фиг. 7A, фиг. 12**). Декодирование с применением кодонов/антикодонов AGX/XCT, CGX/XCG, TTX/XAA и TGX/XCA зависит только от NaMTP в среде экспрессии и продуцирует sfGFP с аналогичным содержанием AzK, с и без добавления TPTЗTP (**фиг. 13**).

Семь пар кодон/антикодон не природного происхождения, проанализированных выше, явно опосредуют эффективное декодирование на рибосоме; однако возможно, что другие кодоны из предварительного не клонального скрининга показывают эффективное декодирование при анализе в клональных SSO. Таким образом, исследуют продуцирование белка не природного происхождения в клональном SSO с четырьмя дополнительными парами кодон/антикодон TXC/GYA, GXG/CYC, CXC/GYG и AXT/AYT. Несмотря на высокое удержание UBP (таблица 1), AXT не показывает сигнал флуоресценции с или без AzK, что еще раз подтверждает необходимость пары G-C с кодонами во втором положении. Флуоресценция с добавлением AzK для TXC, CXC и GXG сравнима с флуоресценцией семи первоначально охарактеризованных кодонов, хотя она несколько выше в отсутствие AzK (**фиг. 7A**). Анализ сдвига геля SPAAC показал, что CXC явно приводит к значительно

большему сдвигу белка в клональном SSO, чем наблюдается при предварительном скрининге с не клональными SSO, и также вероятно TXC и GXG, хотя относительно большая ошибка данных из предварительного скрининга не позволяет провести количественное сравнение (**фиг. 7B**). Данные позволяют предположить, что для некоторых кодонов субоптимальная эффективность в скрининге является результатом, по меньшей мере, частично, зависящих от последовательности различий в конструкции плазмиды *in vitro*. Несмотря на это, результаты идентифицируют два дополнительных кодона высокой точности, TXC и CXС, и позволяют предположить, что еще могут быть идентифицированы более жизнеспособные кодоны.

Чтобы начать оценку ортогональности пар кодонов/антикодонов не природного происхождения, AXC/GYT, GXT/AYC и AGX/XCT выбирают и исследуют на продуцирование белка в клональных SSO со всеми попарными комбинациями кодонов и антикодонов не природного происхождения. При добавлении AzK, значительная флуоресценция наблюдается, когда каждый кодон не природного происхождения соединен с родственным антикодоном не природного происхождения, и практически не наблюдается увеличение по сравнению с фоном при соединении с антикодоном не природного происхождения (**фиг. 7B**). Таким образом, AXC/GYT, GXT/AYC и AGX /XCT являются ортогональными и допускают одновременное использование в SSO.

Пример 3. Одновременное декодирование двух кодонов не природного происхождения.

Для исследования одновременного декодирования нескольких кодонов, сначала конструируют плазмиду, в которой нативные кодоны *sfGFP* в положениях 190 и 200 заменены на GXT и AXC, соответственно (*sfGFP*^{190,20} (GXT, AXC)). Кроме того, плаزمида, кодирующая обеими тРНК^{Pyl}(AYC) и *M. jannaschii* тРНК^{pAzF}, которая селективно заряжена *M. jannaschii* TyrRS (*MjTyrRS*) с *p*-азидо-L-фенилаланином (*pAzF*; **фиг. 6B**), и чей антикодон перекодирован для распознавания AXC (тРНК^{pAzF}(GYT); **фиг. 8A**). *E. coli* ML2, несущая вспомогательную плазмиду, кодирующую как *chPylRS*^{IPYE}, так и *MjpAzFRS*, трансформируют UBP-содержащей плазмидой и получают, выращивают и индуцируют клональные SSO с образованием *sfGFP*, как описано выше. При наличии как AzK, так и *pAzF*, повышенная флуоресценция клеток наблюдается в том же временном масштабе, что и экспрессия с конструкциями с одним кодоном (**фиг. 8B, фиг. 14**). Хотя уровень флуоресценции с экспрессией из *sfGFP*^{109,200}(GXT, AXC) несколько меньше, чем половина того, что наблюдается при *sfGFP*¹⁹⁰(GXT) или *sfGFP*²⁰⁰(AXC), он значительно выше, чем тот, который наблюдается в контроле янтарь-охрана (*sfGFP*^{190,200}(TAA, TAG)), декодированном соответствующими супрессорными тРНК (**фиг. 8C, фиг. 14**). В обоих случаях, при анализе сдвигом геля SPAAC не видно никакой несмещенной полосы, и подвижность основной полосы еще более замедлена по сравнению той, которая наблюдается при включении одной псАА, что свидетельствует о том, что действительно включены две псАА (**фиг. 8D**). Чтобы подтвердить, что и *pAzF*, и AzK включены, очищенный белок анализируют с использованием количественной масс-спектрометрии

интактного белка (HRMS ESI-TOF). В соответствии с анализом сдвига геля, этот анализ показывает, что $91 \pm 1,1\%$ выделенного белка содержит как *pAzF* и *AzK*, при этом $1,7 \pm 0,4\%$ содержит отдельно *pAzF* и $7,5 \pm 0,78\%$ содержит отдельно *AzK* (фиг. 15). В обоих случаях, масса идентифицированных примесей соответствует аминокислотной замене, соответствующей мутации dX в dT, позволяя предположить, что большая часть потери точности включения псАА связана с потерей dNaM или dTPT3 во время репликации, и не из-за ошибок во время транскрипции или трансляции. Удержание UBP на основе анализа стрептавидин-биотинового сдвига. Удержание включает относительный сдвиг (т.е. сигнал сдвинутой полосы, разделенный на общий сигнал сдвинутой и несдвинутой полос), нормализованный к относительному сдвигу контроля матрицы оцДНК, за исключением тРНК^{*pAzF*} и тРНК^{*Ser*}, где нормализация невозможна. Показано среднее значение \pm стандартное отклонение (таблица 1).

Таблица 1. Сохранение пары оснований (BP) в описанных SSO

Конструкция						
Эксперименты с одним кодоном	Показано на	<i>n</i>	Кодоны	Кодоны удержания UBP	Антикодоны	Антикодоны удержания UBP
<i>sfGFP¹⁵¹ M. mazei</i> тРНК ^{<i>Py1</i>}	ФИГ. 6А	3	АХС	94 \pm 3	GYT	92 \pm 4
<i>sfGFP¹⁵¹ M. mazei</i> тРНК ^{<i>Py1</i>}	ФИГ. 6А	3	GXC	94 \pm 3	GYC	96 \pm 5
<i>sfGFP¹⁵¹ M. mazei</i> тРНК ^{<i>Py1</i>}	ФИГ. 6А	3	GXT	99 \pm 1	AYC	99 \pm 1
<i>sfGFP¹⁵¹ M. mazei</i> тРНК ^{<i>Py1</i>}	ФИГ. 6А	3	AGX	89 \pm 3	XCT	61 \pm 18
<i>sfGFP¹⁵¹ M. mazei</i> тРНК ^{<i>Py1</i>}	ФИГ. 6А	3	CGX	89 \pm 3	XCG	83 \pm 8
<i>sfGFP¹⁵¹ M. mazei</i> тРНК ^{<i>Py1</i>}	ФИГ. 6А	3	TGX	91 \pm 2	XCA	78 \pm 13
<i>sfGFP¹⁵¹ M. mazei</i> тРНК ^{<i>Py1</i>}	ФИГ. 6А	3	TTX	95 \pm 3	XAA	76 \pm 37
<i>sfGFP¹⁵¹ M. mazei</i> тРНК ^{<i>Py1</i>}	ФИГ. 6А	5	CXC	67 \pm 8	GYG	91 \pm 4
<i>sfGFP¹⁵¹ M. mazei</i> тРНК ^{<i>Py1</i>}	ФИГ. 6А	4	GXG	58 \pm 2	CYC	60 \pm 10

<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6А	3	ТХС	87±6	GYA	94±11
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6А	3	АХТ	97±3	АУТ	95±1
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6В	3	AGX	91±1	АУС	101±1
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6В	3	AGX	92±1	GYT	99±6
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6В	3	AGX	82±3	ХСТ	100±4
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6В	3	АХС	96±3	АУС	99±2
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6В	3	АХС	98±1	GYT	94±8
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6В	3	АХС	99±2	ХСТ	84±12
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6В	3	GXT	99±4	АУС	97±2
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6В	3	GXT	100±1	GYT	100±1
<i>sfGFP¹⁵¹ M. mazei</i> ТРНК ^{Py1}	ФИГ. 6В	3	GXT	99±1	ХСТ	101±1
Multicodon codons experiments (including controls)						
<i>sfGFP¹⁹⁰ M. mazei</i> ТРНК ^{Py1}	Фиг. 7В	3	GXT	103±4	АУС	101±4
<i>sfGFP²⁰⁰ M. jannaschii</i> ТРНК ^{pAzF}	Фиг. 7В	3	АХС	96±2	GYT	>94±1
<i>sfGFP^{190,200} M. mazei</i>	Фиг. 7В	3	GXT, АХС	98±3, 86±2	АУС, GYT	96±1, >88±1

ТРНК ^{Py1} <i>M. jannaschii</i> ТРНК ^{pAzF}						
<i>sfGFP</i> ^{151,190,20} ⁰ <i>M. mazei</i> ТРНК ^{Py1} <i>M. jannaschii</i> ТРНК ^{pAzF} <i>E. coli</i> ТРНК ^{Ser}	Фиг. 7В	3	АХС, GXT, AGX	92±1, 101±2, 96±3	ХСТ, GYT, AУС	93±3, >87±3, >94±2

SSO дает $16 \pm 3,2$ мкг•мл⁻¹ очищенного белка, тогда как контроль подавления янтаря, охры дает $6,8 \pm 1,1$ мкг•мл⁻¹. Однако отмечено, что культура SSO растет до более низкой плотности, чем контрольные клетки янтаря, охры, и при нормализации для OD₆₀₀ SSO дает $13 \pm 1,6$ мкг•мл⁻¹ очищенного белка, тогда как подавление янтаря, охры дает $2,8 \pm 0,28$ мкг•мл⁻¹, демонстрируя, что SSO продуцирует в 4,5 раза больше белка на OD₆₀₀. Все выходы определяют по захвату sfGFP с использованием избытка гранул Strep-Tactin XT во время аффинной очистки. Выход нормализуют до конечной OD₆₀₀ при t=180 мин экспрессии. Показано среднее ± стандартное отклонение (таблица 2). Таким образом, SSO эффективно продуцирует белок не природного происхождения с двумя псАА.

Таблица 2. Выход белка экспрессии sfGFP

Конструкция	<i>n</i>	Кодоны/антикодоны	Выход белка (мкг/мл)	Нормализованный выход белка (мкг/мл/OD600)
<i>sfGFP</i> ¹⁵¹ <i>M. mazei</i> ТРНК ^{Py1}	3	TAC/-	66±13	23±1,9
<i>sfGFP</i> ¹⁵¹ <i>M. mazei</i> ТРНК ^{Py1}	3	TAG/CTA	52±11	18±3,0
<i>sfGFP</i> ¹⁵¹ <i>M. mazei</i> ТРНК ^{Py1}	3	AХС/GYT	28±6,3	19±2,1
<i>sfGFP</i> ¹⁵¹ <i>M. mazei</i> ТРНК ^{Py1}	3	GXC/GYC	31±0,32	18±2,9
<i>sfGFP</i> ¹⁵¹ <i>M. mazei</i> ТРНК ^{Py1}	3	GXT/AУС	29±3,3	21±0,22
<i>sfGFP</i> ¹⁵¹ <i>M. mazei</i> ТРНК ^{Py1}	3	AGX/XCT	34±4,7	19±1,7
<i>sfGFP</i> ¹⁵¹ <i>M. mazei</i> ТРНК ^{Py1}	3	CGX/XCG	29±2,8	19±5,2

<i>sfGFP¹⁵¹ M. mazei</i> тPHK ^{PyI}	3	TGX/XCA	27±3,2	18±4,8
<i>sfGFP¹⁵¹ M. mazei</i> тPHK ^{PyI}	3	TTX/XAA	27±4,1	19±4,6
<i>sfGFP^{190,200} M. mazei</i> тPHK ^{PyI} , <i>M. jannaschii</i> тPHK ^{pAzF}	3	TAA, TAG/TTA, СТА	5,6±1,0	5,0±0,24
<i>sfGFP^{190,200} M. mazei</i> тPHK ^{PyI} , <i>M. jannaschii</i> тPHK ^{pAzF}	3	TAA, TAG/TTA, СТА	6,8±1,1	2,8±0,28
<i>sfGFP^{190,200} M. mazei</i> тPHK ^{PyI} , <i>M. jannaschii</i> тPHK ^{pAzF}	3	GXT, AXC/AYC, GYT	16±3,2	13±1,6
<i>sfGFP^{151,190,200} M. mazei</i> тPHK ^{PyI} , <i>M. jannaschii</i> тPHK ^{pAzF} , <i>E. coli</i> тPHK ^{Ser}	3	AXC, GXT,AGX/XCT, GYT, AYC	12±1,9	7,8±1,1

Чтобы охарактеризовать экспрессию белков с псАА с различными функциональными группами, *sfGFP^{190,200}*(GXT, AXC) экспрессируют в SSO, как описано выше, но с добавлением в среду для выращивания *N*⁶-(пропаргилокси)-карбонил-L-лизина (PrK, **фиг. 6B**), который также распознается chPyIRS^{IPYE}, вместо AzK. Не наблюдается существенное влияние на экспрессию флуоресценцией ни для SSO, ни для контроля янтаря, охры (**фиг. 8E**). В каждом случае подтверждено, что правильное включение как PrK, так и *pAzF* с помощью SPAAC с TAMRA-PEG₄-DBCO с последующим катализируемым медью алкин-азидным циклоприсоединением (CuAAC) с использованием TAMRA-PEG₄-азида, поскольку оба вызывают наблюдаемый сдвиг в электрофоретической подвижности. SSO, а также контроль янтарь, охра показывают ожидаемые сдвиги геля и сигнал TAMRA (**фиг. 8F**).

Пример 4. Одновременное декодирование трех кодонов не природного происхождения

Для исследования одновременного декодирования трех ортогональных кодонов не природного происхождения, используют эндогенную сериновую тPHK^{Ser}, *E. coli* SerT, которая заряжена эндогенным SerRS без распознавания антикодона, и которая ранее была перекодирована для декодирования кодона не природного происхождения. *E. coli* ML2, имеющая необязательную плазмиду, кодирующую chPyIRS^{IPYE} и *MjpAzFRS*, трансформируют плазмидой, экспрессирующей *sfGFP^{151,190,200}*(AXC, GXT,AGX), а также тPHK^{PyI}(XCT), тPHK^{pAzF}(GYT) и тPHK^{Ser}(AYC) (**фиг. 9A**), и клональные SSO получают,

выращивают и индуцируют для продуцирования белка, как описано выше. При добавлении в среду AzK и *pAzF*, наблюдают значительную флуоресценцию, сходную с результатами, полученными выше для одновременного декодирования двух кодонов (фиг. 9В, фиг. 14). Эти клетки дают $12,1 \pm 1,9$ мкг•мл⁻¹ ($7,8 \pm 1,1$ мкг•мл⁻¹ OD⁻¹) выделенного белка, что лишь немного меньше, чем количество, выделенное при декодировании двух кодонов не природного происхождения (Таблица 2). Для подтверждения того, что *pAzF*, AzK и Ser включены, очищенный белок анализируют с помощью количественной масс-спектрометрии интактного белка (HRMS ESI-TOF) и обнаруживают, что $96 \pm 0,63\%$ выделенного белка содержат *pAzF*, AzK и Ser, в то время как основной примесью является sfGFP, содержащий только AzK и Ser ($3,5 \pm 0,63\%$). Белок без включения Ser почти не определяется ($0,20 \pm 0,087\%$), тогда как масса, соответствующая белку, содержащему только *pAzF* и Ser, не может быть обнаружена (фиг. 9С, фиг. 16). Дополнительно, какие-либо примеси, соответствующие множественной вставке либо Ser, AzK, либо *pAzF* не обнаружены.

Пример 5. Способы экспрессии *in vivo* полипептидов не природного происхождения

Материалы

Полный список использованных олигонуклеотидов и плазмид приведен в таблице 3. Природные олигонуклеотиды одноцепочечной ДНК и блоки gBlock покупают у IDT (San Diego, CA). Секвенирование проводит компания Genewiz (San Diego, CA). Всю очистку ДНК проводят с использованием наборов для колонок с диоксидом кремния Zymo Research. Все клонирующие ферменты и полимеразы покупают у New England Biolabs (Ipswich, MA). Все реагенты для биоконъюгации покупают у Click Chemistry Tools (Scottsdale, AZ). Все нуклеозидтрифосфаты и нуклеозидфосфорамидиты не природного происхождения, используемые в этом исследовании, получают из коммерческих источников. Все матрицы ssDNA dNaM также получают из коммерческих источников, за исключением *sfGFP*²⁰⁰(AGX), который синтезирован, как описано в литературе.

Таблица 3. Олигонуклеотиды одноцепочечной ДНК, используемые в ПЦР и анализе стрептавидин-биотинового сдвига

ID	Применение	Последовательность (5'-3')	SEQ ID NO:
Праймеры для ПЦР UBP			
Efo309	sfGFP Y151 вставка F	ATGGGTCTCACACAAACTCGAGTACA ACTTTAACTCACAC	2
Efo310	sfGFP Y151 вставка R	ATGGGTCTCGATTCCATTCTTTTGT GTCTGC	3

Efo296	sfGFP Y200 вставка F	CATAATGGTCTCGCTGCTGCCCGATA ACCAC	4
Efo297	sfGFP Y200 вставка R	TGATATTGGTCTCGGTCTTTCGATAA AACACTCTGAGTAGAG	5
Efo311	<i>M. mazei</i> тPHK ^{Py1} вставка F	ATGGGTCTCGAAACCTGATCATGTAG ATCGAACGG	6
Efo312	<i>M. mazei</i> тPHK ^{Py1} вставка R	ATGGGTCTCATCTAACCCGGCTGAAC GG	7
Efo313	<i>M. jannaschii</i> тPHK ^{pAzF} вставка F	ATGGGTCTCCGGTAGTTCAGCAGGGC AGAACG	8
Efo314	<i>M. jannaschii</i> тPHK ^{pAzF} вставка R	ATGGGTCTCGGAGGGGATTTGAACCC CTGCCATG	9
Efo294	sfGFP D190 вставка	ATATTCGGTCTCGTCAGCAGAATACG CCGATTGG	10
Efo295	sfGFP D190 вставка	ACGCGTTGGTCTCGGTTATCGGGCAG CAGCACC	11
YZ401	<i>E. coli</i> тPHK ^{Ser} вставка F	ATTGGTCTCGGCCGAGCGGTTGAAGG CAC	12
YZ403	<i>E. coli</i> тPHK ^{Ser} вставка R	ATTGGTCTCTCTGGAACCCTTTCGGG TCG	13
Праймеры для анализа стрептавидин-био тинового сдвига			
Efo251	Положение Y151 вставка F	CTCGAGTACAACCTTAACTCACAC	14
Efo252	Положение Y151 вставка R	GATTCATTCTTTTGTGTTGTCTGC	15
Efo294	Положение D190 вставка F	ATATTCGGTCTCGTCAGCAGAATACG CCGATTGG	10

Efo295	Положение D190 вставка R	ACGCGTTGGTCTCGGTTATCGGGCAG CAGCACC	11
Efo347	Положение Y200 вставка F	GCTGCTGCCCGATAACCAC	16
Efo348	Положение Y200 вставка R	GGTCTTTCGATAAAACACTCTGAGTA GAG	17
Efo343	<i>M. mazei</i> тРНК ^{Py1} вставка F	GAAACCTGATCATGTAGATCGAACG G	18
Efo344	<i>M. mazei</i> тРНК ^{Py1} вставка R	ATCTAACCCGGCTGAACGG	19
Efo313	<i>M. jannaschii</i> тРНК ^{pAzF} вставка F	ATGGGTCTCCGGTAGTTCAGCAGGGC AGAACG	8
Efo305	<i>M. jannaschii</i> тРНК ^{pAzF} вставка R	CCGCTGCCACTAGGAAGCTTATG	20
Efo119	<i>E. coli</i> тРНК ^{Ser} вставка F	CCTCTAGAAAATCATTCCGGAAGTGT G	21
Efo162	<i>E. coli</i> тРНК ^{Ser} вставка R	CTCTGGAACCCTTTCGGGTCGCCGGT TTGXTAGACCGGTGCCTTCAACCGCT CGGC	22
Матрица для ПЦР UBP ([NNN] означает любой указанный триплет кодон/ антикодон)			
GFP151_[NNN]	sfGFP Y151 вставка	CTCGAGTACAACCTTTAACTCACACAA TGTA[NNN]ATCACGGCAGACAAACA AAAGAATGGAATC	23
GFP190_GXT	sfGFP D190 вставка	CAGCAGAATACGCCGATTGGCGXTG GCCCGGTGCTGCTGCCCGATAACC	24

GFP200_AXС	sfGFP Y200 вставка F	GCTGCTGCCCGATAACCACAXCCTCT CTACTCAGAGTGTTTTATCGAAAGAC C	25
GFP200_opt_AGX	sfGFP Y200 вставка R	GCTGCCCGATAACCACAGXTTGTCTA CTCAGAGTGTTTTATCG	26
тРНК_Pyl_[NNN]	<i>M. mazei</i> тРНК ^{Pyl} вставка	GAATCTAACCCGGCTGAACGGATT[N NN]AGTCCGTTTCGATCTACATGATCA GG	27
тРНК_Mj_GYT	<i>M. mazei</i> тРНК ^{Pyl} вставка	GATTTGAACCCCTGCCATGCGGATTA XCAGTCCGCCGTTCTGCCCTGCTGAA	28
тРНК_Eser_AYC	<i>E. coli</i> тРНК ^{Ser} вставка	CTCTGGAACCCTTTCGGGTCGCCGGT TTGXTAGACCGGTGCCTTCAACCGCT CGGC	22

Условия выращивания

Все эксперименты с бактериями проводят в 300 мкл среды 2xYT (Fisher Scientific) с добавлением фосфата калия. (50 мМ рН 7). Выращивание проводят в плоскодонных 48-луночных планшетах (CELLSTAR, Greiner Bio-One) при встряхивании при 200 об/мин при 37°C (Infors HT Minitron). Антибиотики используют в следующих концентрациях (если не указано иное): хлорамфеникол (5 мкг /мл), карбенициллин (100 мкг /мл) и зеоцин (50 мкг /мл). Нуклеозидтрифосфаты не природного происхождения используют в следующих концентрациях (если не указано иное): dNaMTP (150 мкМ), dTPTЗTP (10 мкМ), NaMTP (250 мкМ), TPTЗTP (30 мкМ). Среду UBP определяют как указанную среду 2xYT, содержащую dNaMTP и dTPTЗTP.

Плазмидная конструкция

Большие вставки (>100 п.н.), вставки *MjpAzFRS*, тРНК или кассеты резистентности к антибиотикам получают с помощью сборки Gibson ампликонов ПЦР или gBlocks. Ампликоны обрабатывают DpnI в течение ночи при КТ перед сборкой в течение 1,5 ч при 50°C. Делеции или небольшие вставки (<50 п.н., например, мутагенез кодонов или антикодонов, удаление сайтов рестрикции или введение сайтов назначения Golden gate) конструируют путем внесения желаемых изменений в липкие концы праймеров для ПЦР, сконструированные для амплификации всей плазмиды. Праймеры фосфорилируют с использованием T4 PNK перед ПЦР, и полученный ампликон ПЦР обрабатывают DpnI в течение ночи при комнатной температуре и рециркулируют с использованием T4 ДНК лигазы. После первоначальной сборки/лигирования плазмиды трансформируют в электрокомпетентные клетки XL-10 Gold и выращивают на селективном агаре LB Lennox (BP Difco). Плазмиды выделяют из отдельных колоний и перед использованием проверяют секвенированием по Сэнгеру. Все плазмиды, используемые в этом исследовании, можно

найти в **таблице 4**. Все рамки считывания *sfGFP* контролируют P_{T7-tetO}, и все тРНК контролируют P_{T7-lacO}. Каркас pSYN содержит: ori(p15A) *bleoR*. Каркас pGEX содержит: ori(pBR322) *ampR*. Сайты назначения Golden gate (*dest*) состоят из последовательностей распознавания BsaI-KpnI-BsaI.

Таблица 4. Плазмиды, использованные в примерах

Остов	Источник	Применение	Соответствующие свойства
Superfolder GFP плазмиды экспрессии			
pSYN	Zhang et al. ¹	Природная плаزمида экспрессии	<i>sfGFP¹⁵¹</i> (TAG), <i>M. mazei</i> тРНК ^{Py1} (СТА)
pSYN	Zhang et al. ¹	Природная плазмида экспрессии	<i>sfGFP¹⁵¹</i> (ТАС)
pSYN	Эта работа	Природная плазмида экспрессии	<i>sfGFP¹⁹⁰</i> (ТАА), <i>M. mazei</i> тРНК ^{Py1} (ТТА), стоп-кодон опала
pSYN	Эта работа	Природная плазмида экспрессии	<i>sfGFP²⁰⁰</i> (TAG), <i>M. jannaschii</i> тРНК ^{pAzF} (СТА)
pSYN	Эта работа	Природная плазмида экспрессии	<i>sfGFP^{190,200}</i> (ТАА, TAG), <i>M. mazei</i> тРНК ^{Py1} (ТТА) <i>M. jannaschii</i> тРНК ^{pAzF} (СТА); стоп-кодон опала
pSYN	Zhang et al. ¹	Плазмида назначения UBP	<i>sfGFP¹⁵¹</i> (<i>dest</i>), <i>M. mazei</i> тРНК ^{Py1} (<i>dest</i>)
pSYN	Эта работа	Плазмида назначения UBP	<i>sfGFP¹⁹⁰</i> (<i>dest</i>), <i>M. mazei</i> тРНК ^{Py1} (<i>dest</i>)
pSYN	Эта работа	Плазмида назначения UBP	<i>sfGFP²⁰⁰</i> (<i>dest</i>), <i>M. jannaschii</i> тРНК ^{pAzF} (<i>dest</i>)

pSYN	Эта работа	Плазмида назначения UBP	<i>sfGFP¹⁹⁰⁻²⁰⁰</i> (dest), <i>M. mazei</i> тРНК ^{PyI} (dest), <i>M. jannaschii</i> тРНК ^{pAzF} (dest)
pSYN	Эта работа	Плазмида назначения UBP	<i>sfGFP^{151,190-200}</i> (dest, dest), <i>M. mazei</i> тРНК ^{PyI} (dest), <i>M. jannaschii</i> тРНК ^{pAzF} (dest), <i>E. coli</i> тРНК ^{Ser} (dest)
Необязательные плазмиды			
pGEX	Эта работа	Необязательная плазмида	P_{AmpR} - <i>tetR</i> , P_{lacIq} - <i>lacI</i> , $P_{tac-lacO}$ - <i>chPyIRS^{IPYE}</i>
pGEX	Эта работа	Необязательная плазмида	P_{AmpR} - <i>tetR</i> , P_{lacIq} - <i>lacI</i> , $P_{lacUV5-lacO}$ - <i>MjpAzFRS</i> , $P_{tac-lacO}$ - <i>chPyIRS^{IPYE}</i>

¹ Zhang, Y. et al. A semi-synthetic organism that stores and retrieves increased genetic information. *Nature* 551, 644-647 (2017)

ПЦР олигонуклеотидов UBP

Вставки двухцепочечной ДНК с UBP-содержащей последовательностью получают с помощью ПЦР (стандартный буфер OneTaq 1x, 0,025 единиц/мкл OneTaq, 0,2 мМ dNTPs, 0,1 мМ dTPTЗTP, 0,1 мМ dNaMTP, 1,2 мМ MgSO₄, 1x SYBR Green, 1,0 мкМ праймеров, ~20 пМ матрицы, циклирование: 96°C 0:30 мин, 96°C 0:30 мин, 54°C 0:30 мин, 68°C 4:00 мин, прочтения флуоресценции, перейти к стадии 2 <24 раза) с праймерами (в списке А) с использованием химически синтезированного dNaM, содержащего олигонуклеотиды оцДНК (в списке В) в качестве матрицы. Вставки для положения *sfGFP¹⁹⁰* и *sfGFP²⁰* объединяют путем перекрытия удлинения с использованием идентичных условий, как указано выше, но с обеими матрицами при 1 нМ. Амплификации отслеживают, и реакционные смеси ставят на лед по мере того, как зеленая кривая SYBR выходит на плато. Продукты анализируют с помощью нативного PAGE (6% акриламид:бисакриламид 29:1; окрашивание SYBR Gold в 1x TBE) для проверки одиночных ампликонов, очищенных на спин-колонке (Zymo Research) и количественно определяют с помощью Qubit дцДНК HS (ThermoFisher).

Сборка Golden Gate векторов экспрессии SSO

Вставки, содержащие UBP, включают в каркас входного вектора pSYN (таблица 4) с помощью сборки Golden Gate (буфер Cutsmart 1x, 1 мМ АТФ, 6,67 единиц/мкл Т4 ДНК лигазы, 0,67 единиц/мкл BsaI-HFv2, 20 нг/мкл ДНК вектора входа; циклирование: 37°C 10:00 мин, 37°C 5:00 мин, 16°C 5:00 мин, 22°C 2:00 мин, повторить со стадии 2 39 раз, 37°C 20:00 мин, 55°C 15:00 мин, 80°C 30:00 мин) с молярным соотношением каждой вставки к

вектору входа 3:1. *BsaI*-HF используют для экспериментов на **фиг.6**. Остаточная линейная ДНК и непереваренный вектор входа переваривают сначала *KpnI*-HF (0,33 единицы/мкл, 1 ч при 37°C), и затем экзонуклеазой T5 (0,17 единицы/мкл, 30 мин при 37°C). Продукт очищают на спин-колонке и количественно определяют, используя Qubit дцДНК HS (ThermoFisher).

Приготовление компетентных стартерных клеток

Штамм ML2 (BL21(DE3) *lacZYA::PtNTT2 (66-575) Δ recA polB⁺⁺*) трансформируют дополнительной плазмидой pGEX (таблица 4) и высевают на агар LB Lennox с хлорамфениколом и карбенициллином. Отдельные колонии отбирают и верифицируют на активность *PtNTT2* через поглощение радиоактивного [α -³²P]dATP, как описано ранее (Zhang et al. 2017). Компетентные клетки для репликации и трансляции UBP получают путем выращивания в среде 2xYT при 37°C, 250 об/мин в культуральной колбе с дефлекторами до OD₆₀₀ 0,25-0,30. Культуры переносят в предварительно охлажденные 50 мл пробирки Falcon и осторожно встряхивают на бане со льдом и водой в течение 2 мин. Клетки осаждают центрифугированием (10 мин, 3200 об/мин) и промывают в холодной стерильной воде, осаждают и снова промывают, затем, наконец, осаждают и суспендируют в 50 мкл 10% глицерина на 10 мл культуры. Клетки либо используют сразу, либо замораживают при -80°C для последующего использования.

Эксперименты в не клональной популяции

Свежеприготовленные компетентные клетки подвергают электропорации (2,5 кВ) с ~0,4 нг продукта сборки Golden Gate и немедленно суспендируют в 950 мкл 2xYT с добавлением фосфата калия (50 мМ pH 7), где 10 мкл разводят в 40 мкл среды UBP, содержащей 1,25X dNaMTP и dTPT3TP без зеоцина. После восстановления клеток в течение 1 ч при 37°C, 15 мкл клеток суспендируют в 285 мкл среды UBP с зеоцином и выращивают при встряхивании при 37°C в 48-луночном планшете. Культуры переносят на лед до достижения стационарной фазы, при OD₆₀₀ ~1, и хранят в течение ночи для экспрессии белка.

Эксперименты с клональной SSO

Компетентные клетки подвергают электропорации с продуктом сборки Golden Gate (1-20 нг) и восстанавливают как для экспериментов с не клональными популяциями. Посев проводят путем нанесения 10 мкл восстановительной культуры (и ее разведений) на капли агара (250 мкл 2xYT 2% агар 50 мМ фосфат калия), содержащий хлорамфеникол, карбенициллин, зеоцин, dNaMTP и dTPT3TP. Колонии диаметром примерно 0,5 мм собирают и суспендируют в среде UBP (300 мкл) после роста на планшете (12-20 ч; 37°C). Каждую культуру переносят в предварительно охлажденные пробирки на льду до достижения стационарной фазы, при OD ~1, и хранят в течение ночи для экспрессии белка. Каждую культуру предварительно подвергают скринингу для 1) сохранения UBP с использованием анализа стрептавид-биотинового сдвига (как описано ниже) и 2) качественной экспрессии *sfGFP* путем смешивания культуры в соотношении 1:4 со средой, уже содержащей компоненты для экспрессии (рибонуклеозидтрифосфаты, psAA, IPTG и

ангидротетрациклин). Колонии выбрасывают, если они не продуцируют флуоресцентный сигнал, когда добавляют подходящие псАА через 2 ч инкубации при 37°C или в течение ночи при комнатной температуре. Кроме того, колонии с <80% удержанием UBP в *sfGFP* выбрасывают. Если более трех колоний удовлетворяют этим критериям, то только три с самым высоким удержанием UBP выбирают для ограничения материальных затрат. Данные справа от пунктирной линии на **фиг. 7А** были получают с помощью слегка модифицированных способов. Вместо предварительного скрининга колоний, как описано выше, экспрессию проводят на многочисленных колониях, но анализ белка проводят только для культур, которые демонстрируют обнадеживающую флуоресценцию во время экспрессии. Во время экспрессии используют 10 мМ AzK. Дополнительно, во время очистки белка используют буфер W2 вместо буфера W.

Предварительно клонированные векторы экспрессии SSO

В экспериментах на **фиг. 7В**, **фиг. 8** и **фиг. 9**, плазмиды из предварительно подвергнутых скринингу колоний выделяют (Zymo Research Miniprep), чтобы они служили исходными плаزمидами для (предварительно клонированной) трансформации, чтобы облегчить предварительный скрининг колоний. Плазмиды подвергают предварительному скринингу (как описано выше) для качественной флуоресценции из экспрессии *sfGFP* с соответствующими псАА. Колонии для данных на **фиг. 7В** вместо этого подвергали предварительному скринингу с и без *rNaMTP* и *rTPT3TP* в присутствии AzK для качественного получения темного и флуоресцентного сигнала, соответственно. Все предварительно клонированные плазмиды подвергают предварительному скринингу на удержание UBP в *sfGFP* (> 80%). Более того, эти плазмиды ПЦР амплифицируют с применением стандартного протокола OneTaq (New England Biolabs), без нуклеозидтрифосфатов не природного происхождения для усиления мутаций dX до dN, и ампликон секвенируют по Сэнгеру для верификации целостности природной последовательности в плазмиде. В последовательностях, кодирующих белок, допускаются молчащие мутации.

Экспрессия белка UBP

Культуры обновляют в среде UBP до OD₆₀₀ 0,10-0,15 и встряхивают при 37°C до OD 0,5-0,8, когда рибонуклеотидтрифосфаты добавляют к 250 мкМ NaMTP и 30 мкМ TPT3TP вместе с псАА при 5 мМ *pAzF*, 20 мМ AzK или 10 мМ PrK. Только 10 мМ AzK используют в экспериментах с двойным/тройным кодоном или их контролях (**фиг. 8**, **фиг. 9**). Через 20 минут дополнительной инкубации, преиндукцию инициируют добавлением IPTG (1 мМ) и культуры инкубируют в течение еще 1 ч. Наконец, экспрессию *sfGFP* индуцируют дерепрессией *tetO* путем добавления ангидротетрациклина (100 нг/мкл). OD₆₀₀ и флуоресценцию GFP контролируют (каждые 30 минут) с использованием Perkin Elmer Envision 2103 Multilabel Reader (OD: фильтр 590/20 нм; *sfGFP*: возб. 485/14 нм, исп. 535/25 нм). Через 3 ч экспрессии, культуры осаждают и хранят при -80°C для последующего анализа.

Анализ стрептавидин-биотинового сдвига для удержания UBP

Удержание UBP в плазмидной ДНК определяют с помощью ПЦР амплификации с использованием нуклеозидтрифосфата не природного происхождения d5SICSTP, а также биотинилированного аналога dNaM dMMO2^{Bio}TP. Плазмиды из SSO выделяют с помощью стандартного miniprep, получая смесь плазмид экспрессии SSO (pSYN) и необязательных плазмид (pGEX). В общей сложности 2 нг смеси плазмид используют в качестве матрицы в 15 мкл ПЦР реакции (стандартный буфер OneTaq 1x, 0,018 единиц/мкл OneTaq, 0,007 единиц/мкл DeepVent, 0,4 mM dNTPs, 0,1 mM d5SICSTP, 0,1 mM dMMO2^{Bio}TP, 2,2 mM MgSO₄, 1X SYBR Green, 1,0 мкМ праймеров, циклирование: 96°C 2:00 мин, 96°C 0:30 мин, 50°C 0:10 мин, 68°C 4:00 мин, прочтения флуоресценции, 68°C 0:10 мин, переход к стадии 2 <24 раз). Отдельные образцы удаляют на последнем этапе каждого цикла, поскольку кривая SYBR Green I показывает амплификацию до плато. В полученный биотинилированный ампликон добавляют 10 мкг стрептавидина (Promega) на 1,5-2,0 мкл неочищенной ПЦР реакции. Связанную с стрептавидином фракцию визуализируют как сдвиг на 6% нативного PAGE, и как сдвинутые, так и не сдвинутые полосы количественно определяют с помощью ImageStudioLite или Fiji, чтобы получить относительный необработанную долю сдвига. Нормализацией неочищенного сдвига до контрольного сдвига, созданного шаблонизацией ПЦР реакции с химически синтезированным олигонуклеотидом, оценивают общее удержание UBP. Нормализация была невозможна для tPHK^{pAzF} или tPHK^{Ser}, поскольку достоверная амплификация возможна только с отжигом праймеров вне вставки Golden Gate и, таким образом, не отжигается с соответствующим контрольным олигонуклеотидом.

Очистка белка

Дебрисы клеток из экспериментов по экспрессии белка (200 мкл) лизируют с использованием BugBuster (100 мкл; EMD Millipore; 15 мин; КТ; 220 об/мин). Затем лизаты клеток разводят в Buffer W (50 mM HEPES pH 8, 150 mM NaCl, 1 mM EDTA) до конечного объема, равного 500 мкл минус объем используемых аффинных гранул. Магнитные гранулы Strep-Tactin XT (5% (об./об.) суспензия гранул MagStrep «type3» XT, IBA Lifesciences) используют в количестве 20 мкл для обычной очистки и 100 мкл для оценки общего выхода экспрессии. Белок связывают с гранулами (30 мин, 4°C, осторожное вращение) перед тем, как гранулы стягивают вниз и промывают Buffer W (2x500 мкл). Для очистки белка для анализа HRMS используют Buffer W2 (50 mM HEPES, pH 8, 1 mM EDTA). Наконец, белок элюируют с использованием 25 мкл буфера BXT (50 mM HEPES pH 8, 150 mM NaCl, 1 mM EDTA, 50 mM d-биотина) в течение 10 мин при комнатной температуре с периодическим встряхиванием. Белок элюируют буфером BXT2 (50 mM HEPES pH 8, 1 mM EDTA, 50 mM d-биотина) для анализа HRMS. Для количественного определения используют Qubit Protein Assay Kit (ThermoFisher).

Вестерн-блоттинг sfGFP, конъюгированного с TAMRA

SPAAC проводят путем инкубации 33 нг/мкл чистого белка с 0,1 mM TAMRA-PEG₄-DBCO (Click Chemistry Tools) в течение ночи при КТ в темноте. Реакционные смеси смешивают 2:1 с загрузочным красителем SDS-PAGE (250 mM Tris-HCl pH 6, 30%

глицерина, 5% βМЕ, 0,02% бромфенолового синего) и денатурируют в течение 5 мин при 95°C. Гели SDS-PAGE представляют собой 5% акриламидные концентрирующие гели и 15% акриламидный разрешающий гель при анализе положения sfGFP¹⁵¹ и 17% при анализе sfGFP^{190,200} (разрешающий гель: 15% или 17% акриламид:бисакриламид 29:1, 0,1% (масс./об.) APS, 0,04% TEMED, 0,375 М Tris-HCl pH 8,8, 0,1% (масс./об.) SDS; концентрация: 5% акриламид:бисакриламид 29:1, 0,1% (масс./об.) APS, 0,1% TEMED, 0,125 М Tris-HCl, pH 6,8, 0,1% (масс./об.) SDS). Проводят электрофорез в течение 15 мин при 40 В, и затем запускают в течение ~5 ч при 120 В для 15% гелей и ~6,5 ч для 17% гелей. Подвижный буфер (25 мМ основание Tris, 200 мМ глицин, 0,1% (масс./об.) SDS) меняют каждые 2 ч. Полученный гель наносят на PVDF (EMD Millipore 0,45 мкм PVDF-FL) с использованием влажного переноса в буфере для холодного переноса (20% (об./об.) MeOH, 50 мМ основания Tris, 400 мМ глицина, 0,0373% (масс./об.) SDS) в течение 1 ч при 90 В. Мембрану блокируют 5% раствором нежирного молока в PBS-T (PBS pH 7,4, 0,01% (об./об.) Tween-20) в течение ночи при 4°C при осторожном перемешивании. Первичные антитела (α-Nterm-GFP кролика, Sigma Aldrich #G1544) применяют в PBS-T (1:3000) в течение 1 ч (КТ; осторожное перемешивание). Блот промывают в PBS-T (5 мин) перед тем, как вторичные антитела (α-кроличье-Alexa Fluor 647-конъюгированное антитело козы, ThermoFisher #A32733) применяют в PBS-T (1:20000) в течение 45 мин (КТ; легкое перемешивание). Блот промывают PBS-T перед (3x5 мин) визуализацией с использованием лазерного сканера Typhoon 9410 (Typhoon Scanner Control v5 GE Healthcare Life) при разрешении 50-100 мкм, сканируют сначала для AlexaFluor 647 (возб. 633 нм; исп. 670/30 нм; PMT 500 В), и затем TAMRA (возб. 532 нм; исп. 580/30 нм; PMT 400 В).

Двойная биоконъюгация PrK-pAzF меченого белка

Дебрисы клеток из 1 мл культуры лизируют использованием BugBuster (100 мкл; EMD Millipore; 15 мин при комнатной температуре; 220 об/мин). Лизат разводят в Buffer W (600 мкл) и добавляют гранулы MagStrep (200 мкл) и позволяют связывание (30 мин, 4°C, осторожное вращение). Гранулы стягивают вниз с помощью магнита и промывают холодным Buffer W (2x1000 мкл), и затем суспендируют в Buffer W (200 мкл). SPAAC проводят с применением половины этой суспензии с TAMRA-PEG₄-DBCO (0,5 мМ) в течение 12-16 ч (КТ; осторожное вращение). Гранулы промывают Buffer W без EDTA (2x500 мкл; HEPES 50 мМ, pH 7,4, 150 мМ NaCl) перед суспендированием в Buffer W без EDTA (100 мкл). Проводят CuAAC (1,5 ч; КТ; осторожное вращение) с использованием половины этой суспензии с азидо-PEG₄-ТАМРА (0,2 мМ), а также сульфатом меди(II) (0,5 мМ), трис(бензилтриазолилметил)амином (2 мМ; ТНРТА) и аскорбатом натрия (15 мМ). Гранулы промывают Buffer W (2x500 мкл) перед проведением элюирования буфером ВХТ (10 мин; КТ; иногда встряхивание).

Масс-спектрометрия высокого разрешения интактного белка

Очищенный белок (5 мкг) обессоливают в воде для ВЭЖХ (4x500 мкл) с помощью четырех циклов центрифугирования через фильтры 10K Amicon Ultra Centrifugal (EMD Millipore) при 14000xg (3x10 мин, и затем 1x18 мин), как описано выше. После

восстановления белка, 6 мкл белка вводят в Waters I-Class LC, соединенный с Waters G2-XS TOF. Условия потока: 0,4 мл/мин 50:50 вода:ацетонитрил плюс 0,1% муравьиная кислота. Ионизацию проводят с помощью ИЭР+ и собирают данные для m/z 500-2000. Проводят спектральное объединение над основной частью массового пика, и объединенный спектр подвергают деконволюции с использованием Waters MaxEnt1. Анализ проводят путем автоматического интегрирования пиков, а также ручной идентификации пиков (**фиг. 15, фиг. 16**). Точность рассчитывают как интеграл ожидаемой массы относительно интегралов всех масс, идентифицированных как продукт или примесь не принимая во внимание технические примеси (например, солевые аддукты, окисление аргинина).

Хотя в настоящем документе были показаны и описаны предпочтительные варианты осуществления настоящего изобретения, специалистам в данной области техники будет очевидно, что такие варианты осуществления, представлены только в качестве примера. Специалистам в данной области техники будут очевидны многочисленные вариации, изменения и замены без отступления от описания. Следует понимать, что различные альтернативы вариантам осуществления описания, описанным в настоящем документе, могут быть использованы при практическом применении описания. Подразумевается, что следующая формула изобретения определяет объем изобретения, и что способы и структуры в пределах объема этих пунктов формулы изобретения и их эквиваленты покрываются этим.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ синтеза полипептида не природного происхождения, включающий:

a. предоставление, по меньшей мере, одной молекулы дезоксирибонуклеиновой кислоты (ДНК) не природного происхождения, содержащей, по меньшей мере, четыре пары оснований не природного происхождения, где, по меньшей мере, одна молекула ДНК не природного происхождения кодирует (i) молекулу информационной рибонуклеиновой кислоты (иРНК), содержащую, по меньшей мере, первый и второй кодоны не природного происхождения, и (ii) по меньшей мере, первую и вторую молекулы транспортной РНК (тРНК), где первая молекула тРНК содержит первый антикодон не природного происхождения, и вторая молекула тРНК содержит второй антикодон не природного происхождения, и, по меньшей мере, четыре пары оснований не природного происхождения в, по меньшей мере, одной молекуле ДНК находятся в таком контексте последовательности, что первый и второй кодоны не природного происхождения молекулы мРНК комплементарны первому и второму антикодонам не природного происхождения, соответственно;

b. транскрипцию, по меньшей мере, одной молекулы ДНК не природного происхождения с получением мРНК;

c. транскрипцию, по меньшей мере, одной молекулы ДНК не природного происхождения с получением, по меньшей мере, первой и второй молекул тРНК; и

d. синтез полипептида не природного происхождения путем трансляции молекулы мРНК не природного происхождения с использованием, по меньшей мере, первой и второй молекул тРНК не природного происхождения, где каждый из, по меньшей мере, первого и второго антикодонов не природного происхождения направляет сайт-специфическое включение аминокислоты не природного происхождения в полипептид не природного происхождения.

2. Способ по п. 1, отличающийся тем, что каждый из, по меньшей мере, двух кодонов не природного происхождения содержит первый нуклеотид не природного происхождения, расположенный в первом положении, втором положении или третьем положении кодона, необязательно, где первый нуклеотид не природного происхождения расположен во втором или третьем положении кодона.

3. Способ по любому из предыдущих пп., отличающийся тем, что, по меньшей мере, два кодона не природного происхождения содержат последовательность нуклеиновой кислоты NNX или NXN, и антикодон не природного происхождения содержит последовательность нуклеиновой кислоты XNN, YNN, NXN или NYN для формирования пары кодон-антикодон не природного происхождения, содержащую NNX-XNN, NNX-YNN или NXN-NYN, отличающийся тем, что N представляет собой любой природный нуклеотид, X представляет собой первый нуклеотид не природного происхождения, и Y представляет собой второй нуклеотид не природного происхождения, отличный от первого нуклеотида не природного происхождения, с X-Y, образующими пару оснований не природного происхождения в ДНК.

4. Способ по п. 4, отличающийся тем, что кодон содержит, по меньшей мере, один G или C, и антикодон содержит, по меньшей мере, один комплементарный C или G.

5. Способ по п. 3 или 4, отличающийся тем, что X и Y независимо выбраны из группы, состоящей из:

(i) 2-тиоурацила, 2'-дезоксуридина, 4-тиоурацила, урацил-5-ила, гипоксантин-9-ила (I), 5-галоурацила, 5-пропинилурацила, 6-азоурацила, 5-метиламинометилурацила, 5-метоксиаминометил-2-тиоурацила, псевдоурацила, метилового эфира урацил-5-оксуксусной кислоты, урацил-5-оксуксусной кислоты, 5-метил-2-тиоурацила, 3-(3-амино-3-N-2-карбоксивпропил)урацила, 5-метил-2-тиоурацила, 4-тиоурацила, 5-метилурацила, 5'-метоксикарбоксиметилурацила, 5-метоксиурацила, урацил-5-оксиуксусной кислоты, 5-(карбоксихидроксиметил)урацила, 5-карбоксиметиламинометил-2-тиоуридина, 5-карбоксиметиламинометилурацила или дигидроурацила;

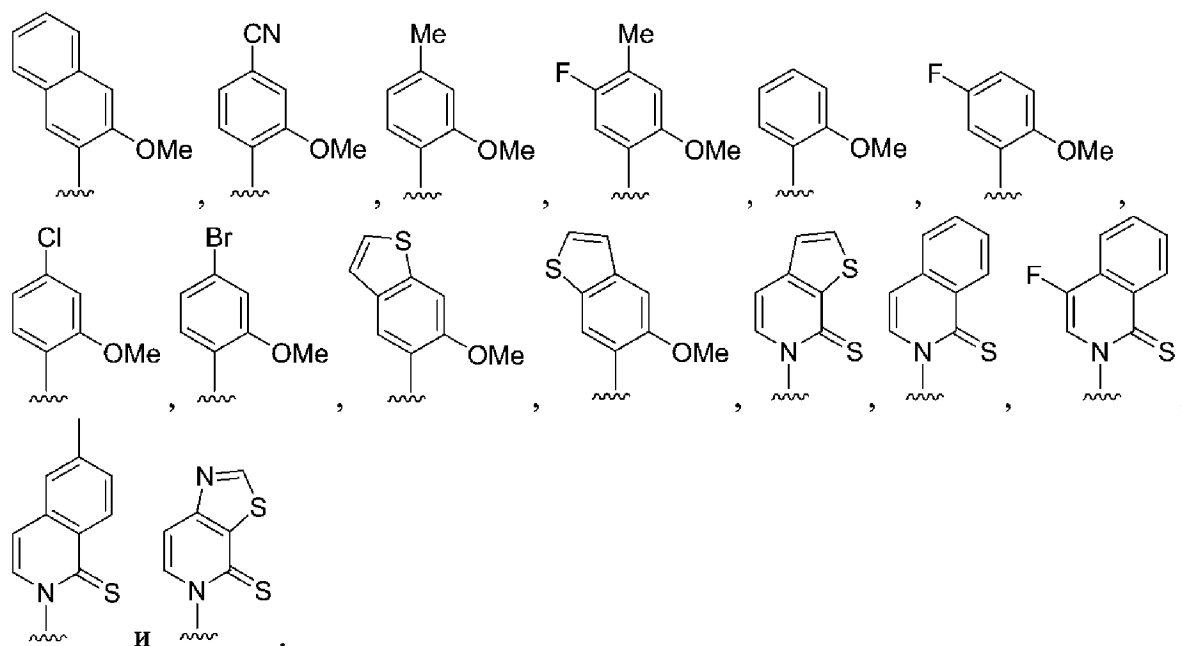
(ii) 5-гидроксиметилцитозина, 5-трифторметилцитозина, 5-галоцитозина, 5-пропинилцитозина, 5-гидроксицитозина, циклоцитозина, цитозина арабинозида, 5,6-дигидроцитозина, 5-нитроцитозина, 6-азоцитозина, азацитозина, N4-этилцитозина, 3-метилцитозина, 5-метилцитозина, 4-ацетилцитозина, 2-тиоцитозина, феноксазина цитидина ([5,4-b][1,4]бензоксазин-2(3H)-она), фенотиазина цитидин (1H-пиримидо[5,4-b][1,4]бензотиазин-2(3H)-она), феноксазина цитидина (9-(2-аминоэтокси)-H-пиримидо[5,4-b][1,4]бензоксазин-2(3H)-она), карбазола цитидина (2H-пиримидо[4,5-b]индол-2-она) или пиридоиндола цитидина (H-пиридо[3',2':4,5]пирроло[2,3-d]пиримидин-2-она);

(iii) 2-аминоаденина, 2-пропиладенина, 2-аминоаденина, 2-F-аденина, 2-аминопропиладенина, 2-амино-2'-дезоксаденозина, 3-дезааденина, 7-метиладенина, 7-дезааденина, 8-азааденина, 8-галоген-, 8-амино-, 8-тиол-, 8-тиоалкил- и 8-гидрокси-замещенных аденинов, N6-изопентениладенина, 2-метиладенина, 2,6-диаминопурина, 2-метилтио-N6-изопентениладенина или 6-азааденина;

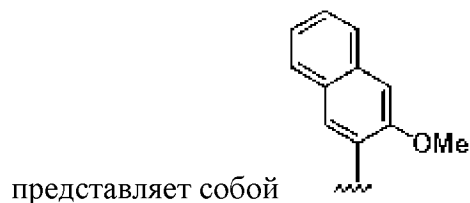
(iv) 2-метилгуанина, 2-пропила и алкильных производных гуанина, 3-дезагуанина, 6-тиогуанина, 7-метилгуанина, 7-дезагуанина, 7-дезагуанозина, 7-деза-8-азагуанина, 8-азагуанина, 8-галоген-, 8-амино-, 8-тиол-, 8-тиоалкил- и 8-гидрокси-замещенных гуанинов, 1-метилгуанина, 2,2-диметилгуанина, 7-метилгуанина или 6-азагуанина; и

(v) гипоксантина, ксантина, 1-метилюридина, кеозина, бета-D-галактозилкеозина, инозина, бета-D-маннозилкеозина, вибутоксозина, гидроксимочевина, (аср3)w, 2-аминопиридина или 2-пиридона.

6. Способ по п. 4 или 5, отличающийся тем, что основания, содержащие каждый из X и Y, независимо выбраны из группы, состоящей из:



7. Способ по п. 6, отличающийся тем, что основание, содержащее каждый X,



8. Способ по п. 6 или 7, отличающийся тем, что основание, содержащее каждый Y,



9. Способ по любому из пп. 3-8, отличающийся тем, что NNX-XNN выбран из группы, состоящей из UUX-XAA, UGX-XCA, CGX-XCG, AGX-XCU, GAX-XUC, CAH-XUG, AUX-XAU, CUX-XAG, GUX-XAC, UAX-XUA и GGX-XCC.

10. Способ по любому из пп. 3-8, отличающийся тем, что NNX-YNN выбран из группы, состоящей из UUX-YAA, UGX-YCA, CGX-YCG, AGX-YCU, GAX-YUC, CAH-YUG, AUX-YAU, CUX-YAG, GUX-YAC, UAX-YUA и GGX-YCC.

11. Способ по любому из пп. 3-8, отличающийся тем, что NXN-NYN выбран из группы, состоящей из GXU-AYC, CXU-AYG, GXG-CYC, AXG-CYU, GXC-GYC, AXC-GYU, GXA-UYC, CXC-GYG и UXC-GYA.

12. Способ по любому из предыдущих пп., отличающийся тем, что, по меньшей мере, две молекулы тРНК не природного происхождения содержат разные антикодоны не природного происхождения.

13. Способ по п. 12, отличающийся тем, что, по меньшей мере, две молекулы тРНК не природного происхождения содержат пирролизил тРНК из рода *Methanosarcina* и тирозил тРНК из *Methanocaldococcus jannaschii* или их производные.

14. Способ по любому из пп. 11-13, включающий зарядку, по меньшей мере, двух молекул тРНК не природного происхождения с помощью аминоксил тРНК синтетазы.

15. Способ по п. 14, отличающийся тем, что аминоксил тРНК синтетаза выбрана из группы, состоящей из химерного PylRS (chPylRS) и *M. jannaschii* AzFRS (*MjpAzFRS*).

16. Способ по п. 12 или 13, включающий зарядку, по меньшей мере, двух молекул тРНК не природного происхождения, по меньшей мере, двумя разными тРНК синтетазами.

17. Способ по п. 16, отличающийся тем, что, по меньшей мере, две тРНК синтетазы включают химерную PylRS (chPylRS) и *M. jannaschii* AzFRS (*MjpAzFRS*).

18. Способ по любому из пп. 1-17, отличающийся тем, что полипептид не природного происхождения содержит две, три или несколько аминокислот не природного происхождения.

19. Способ по любому из пп. 1-18, отличающийся тем, что полипептид не природного происхождения содержит, по меньшей мере, две аминокислоты не природного происхождения, которые являются одинаковыми.

20. Способ по любому из пп. 1-18, отличающийся тем, что полипептид не природного происхождения содержит, по меньшей мере, две различные аминокислоты не природного происхождения.

21. Способ по любому из пп. 1-20, отличающийся тем, что аминокислота не природного происхождения содержит

аналог лизина;

ароматическую боковую цепь;

азидогруппу;

алкиновую группу; или

альдегидную или кетонную группу.

22. Способ по любому из пп. 1-20, отличающийся тем, что аминокислота не природного происхождения не содержит ароматическую боковую цепь.

23. Способ по любому из пп. 1-20, отличающийся тем, что аминокислота не природного происхождения выбрана из N⁶-((азидоэтокси)-карбонил)-L-лизина (AzK), N⁶-пропаргилэтокси-карбонил-L-лизина (PraK), N⁶-(пропаргилокси)карбонил-L-лизина (PrK), п-азидофенилаланина (*pAzF*), BCN-L-лизина, норборненилизина, TCO-лизина, метилтетразинлизина, аллилоксикарбониллизина, 2-амино-8-оксононановой кислоты, 2-амино-8-оксооктановой кислоты, п-ацетил-L-фенилаланина, п-азидометил-L-фенилаланина (*pAMF*), п-йодо-L-фенилаланина, м-ацетилфенилаланина, 2-амино-8-оксононаноевой кислоты, п-пропаргилоксифенилаланина, п-пропаргилфенилаланина, 3-метилфенилаланина, L-Дора, фторированного фенилаланина, изопропил-L-фенилаланина, п-азидо-L-фенилаланина, п-ацил-L-фенилаланина, п-бензоил-L-фенилаланина, п-бромфенилаланина, п-амино-L-фенилаланина, изопропил-L-фенилаланина, O-аллилтирозина, O-метил-L-тирозина, O-4-аллил-L-тирозина, 4-пропил-L-тирозина, фосфонтирозина, три-O-ацетил-GlcNAc α -серина, L-фосфосерина, фосфоносерина, L-3-(2-нафтил)аланина, 2-амино-3-(((2-((3-(бензилокси)-3-

оксопропил)амино)этил)селанил)пропановой кислоты, 2-амино-3-(фенилселанил)пропановой кислоты, селеноцистеина, N6-(((2-азидобензил)окси)карбонил)-L-лизина, N6-(((3-азидобензил)окси)карбонил)-L-лизина и N6-(((4-азидобензил)окси)карбонил)-L-лизина.

24. Способ по любому из предыдущих пп., отличающийся тем, что, по меньшей мере, одна молекула ДНК не природного происхождения находится в форме плазмиды.

25. Способ по любому из пп. 1-26, отличающийся тем, что, по меньшей мере, одна молекула ДНК не природного происхождения интегрируется в геном клетки.

26. Способ согласно варианту осуществления 24 или 25, отличающийся тем, что, по меньшей мере, одна молекула ДНК не природного происхождения кодирует полипептид не природного происхождения.

27. Способ по любому из предыдущих пп., где способ включает *in vivo* репликацию и транскрипцию молекулы ДНК не природного происхождения и *in vivo* трансляцию транскрибированной молекулы мРНК в клеточном организме.

28. Способ по п. 27, в котором клеточный организм представляет собой микроорганизм.

29. Способ по п. 28, отличающийся тем, что клеточный организм представляет собой прокариот.

30. Способ по п. 29, отличающийся тем, что клеточный организм представляет собой бактерию.

31. Способ по п. 30, отличающийся тем, что клеточный организм представляет собой грамположительную бактерию.

32. Способ по п. 30, отличающийся тем, что клеточный организм представляет собой грамотрицательную бактерию.

33. Способ по п. 32, отличающийся тем, что клеточный организм представляет собой *Escherichia coli*.

34. Способ по любому из предыдущих пп., отличающийся тем, что, по меньшей мере, две пары оснований не природного происхождения включают пары оснований, выбранные из dCNMO-dTPT3, dNaM-dTPT3, dCNMO-dTAT1 или dNaM-dTAT1.

35. Способ по любому из пп. 27-34, отличающийся тем, что клеточный организм содержит транспортер нуклеозидтрифосфата.

36. Способ по п. 35, отличающийся тем, что транспортер нуклеозидтрифосфата содержит аминокислотную последовательность *PtNTT2*.

37. Способ по п. 36, отличающийся тем, что переносчик нуклеозидтрифосфата содержит усеченную аминокислотную последовательность *PtNTT2*, необязательно, где усеченная аминокислотная последовательность *PtNTT2*, по меньшей мере, на 80% идентична *PtNTT2*, кодируемой SEQ ID NO. 1.

38. Способ по любому из пп. 27-37, отличающийся тем, что клеточный организм содержит, по меньшей мере, одну молекулу ДНК не природного происхождения.

39. Способ по п. 38, отличающийся тем, что, по меньшей мере, одна молекула ДНК

не природного происхождения содержит, по меньшей мере, одну плазмиду.

40. Способ по п. 38, отличающийся тем, что, по меньшей мере, одна молекула ДНК не природного происхождения интегрируется в геном клетки.

41. Способ по п. 39 или 40, отличающийся тем, что, по меньшей мере, одна молекула ДНК не природного происхождения кодирует полипептид не природного происхождения.

42. Способ по любому из пп. 1-24, отличающийся тем, что способ представляет собой *in vitro* способ, включающий синтез полипептида не природного происхождения с помощью бесклеточной системы.

43. Способ по любому из предыдущих пп., отличающийся тем, что пары оснований не природного происхождения содержат, по меньшей мере, один нуклеотид не природного происхождения, содержащий сахарную группу не природного происхождения.

44. Способ по п. 43, отличающийся тем, что фрагмент сахара не природного происхождения содержит группу, выбранную из группы, состоящей из:

модификации в положении 2', включающей:

ОН, замещенный низший алкил, алкарил, аралкил, О-алкарил или О-аралкил, SH, SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃, NH₂F;

О-алкил, S-алкил, N-алкил;

О-алкенил, S-алкенил, N-алкенил;

О-алкинил, S-алкинил, N-алкинил;

О-алкил-О-алкил, 2'-F, 2'-OCH₃, 2'-O(CH₂)₂OCH₃, отличающийся тем, что алкил, алкенил и алкинил могут быть замещены или не замещены C₁-C₁₀ алкилом, C₂-C₁₀ алкенилом, C₂-C₁₀ алкинилом, O[(CH₂)_nO]_mCH₃, -O(CH₂)_nOCH₃, -O(CH₂)_nNH₂, -O(CH₂)_nCH₃, -O(CH₂)_n-NH₂ и -O(CH₂)_nON[(CH₂)_nCH₃]₂, отличающийся тем, что n и m составляют от 1 до примерно 10;

модификации в положении 5', включающей:

5'-винил, 5'-метил (R или S); или

модификации в положении 4', 4'-S, гетероциклоалкила, гетероциклоалкарила, аминоалкиламино, полиалкиламино, замещенного силила, группы, расщепляющей РНК, репортерной группы, интеркалятора, группы для улучшения фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида; или

любой их комбинации.

45. Клетка, содержащая, по меньшей мере, одну молекулу ДНК не природного происхождения, по меньшей мере, из четырех пар оснований не природного происхождения, отличающийся тем, что, по меньшей мере, одна молекула ДНК не природного происхождения кодирует (i) молекулу информационной рибонуклеиновой кислоты (иРНК), кодирующую полипептид не природного происхождения и содержащую, по меньшей мере, первый и второй кодоны не природного происхождения и (ii) по меньшей мере, первую и вторую молекулы транспортной РНК (тРНК), отличающийся тем, что первая молекула тРНК содержит первый антикодон не природного происхождения, и

вторая молекула тРНК содержит второй антикодон не природного происхождения, и, по меньшей мере, четыре пары оснований не природного происхождения, по меньшей мере, в одной молекуле ДНК находятся в таком контексте последовательности, что первый и второй кодоны не природного происхождения молекулы мРНК комплементарны первому и второму антикодонам не природного происхождения, соответственно.

46. Клетка по п. 45, дополнительно содержащая молекулу мРНК и, по меньшей мере, первую и вторую молекулы тРНК.

47. Клетка по п. 46, отличающийся тем, что, по меньшей мере, первая и вторая молекулы тРНК ковалентно связаны с аминокислотами не природного происхождения.

48. Клетка по п. 47, дополнительно содержит полипептид не природного происхождения.

49. Клетка, содержащая:

а. по меньшей мере, две разные пары кодон-антикодон не природного происхождения, отличающийся тем, что каждая пара кодон-антикодон не природного происхождения содержит кодон не природного происхождения из матричной РНК (мРНК) не природного происхождения и антикодон не природного происхождения из транспортной рибонуклеиновой кислоты (тРНК) не природного происхождения, отличающийся тем, что указанный кодон не природного происхождения содержит первый нуклеотид не природного происхождения, и указанный антикодон не природного происхождения, содержащий второй нуклеотид не природного происхождения; и

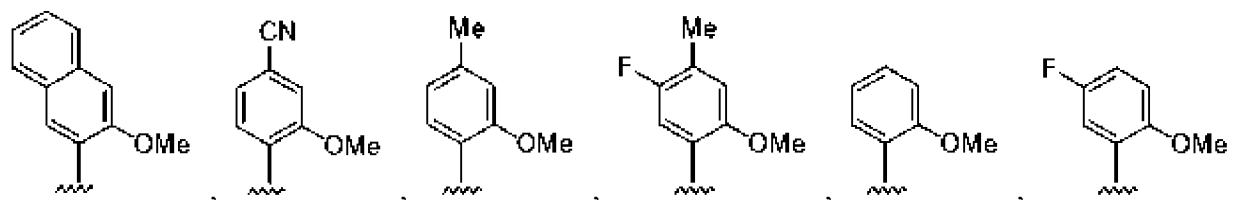
б. по меньшей мере, две разные аминокислоты не природного происхождения, каждая из которых ковалентно связана с соответствующей тРНК не природного происхождения.

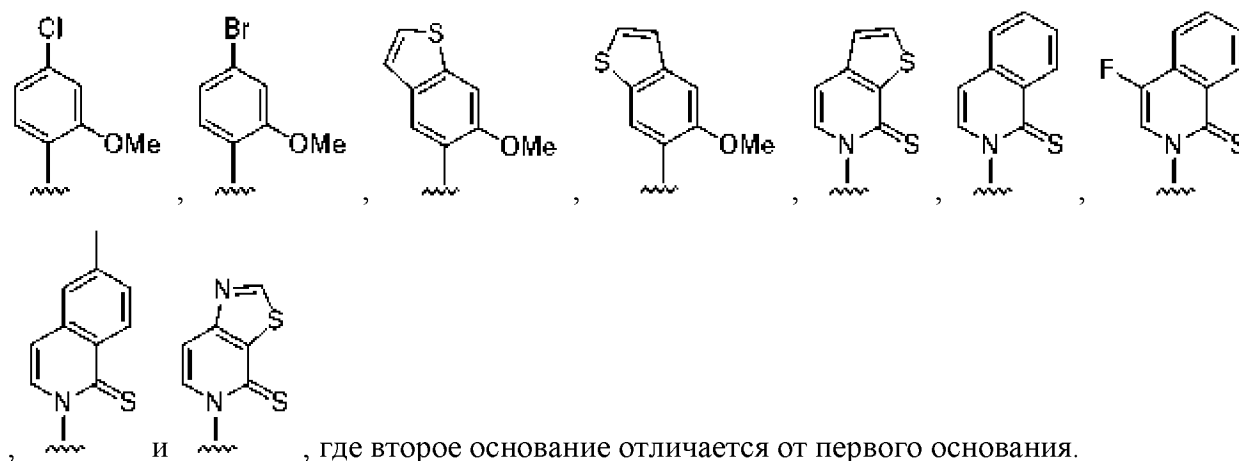
50. Клетка по п. 49, дополнительно содержащая, по меньшей мере, одну молекулу ДНК не природного происхождения, содержащую, по меньшей мере, четыре пары оснований не природного происхождения (UBP).

51. Клетка по любому из пп. 45-50, отличающаяся тем, что первый нуклеотид не природного происхождения расположен во втором или третьем положении кодона не природного происхождения.

52. Клетка по п. 51, отличающаяся тем, что первый нуклеотид не природного происхождения представляет собой комплементарное основание, спаренное со вторым нуклеотидом не природного происхождения антикодона не природного происхождения.

53. Клетка по любому из пп. 45-52, отличающаяся тем, что первый нуклеотид не природного происхождения и второй нуклеотид не природного происхождения содержат первое и второе основания соответственно, независимо выбранные из группы, состоящей из





54. Клетка по любому из пп. 45 или 47-53, отличающаяся тем, что, по меньшей мере, четыре пары оснований не природного происхождения независимо выбраны из группы, состоящей из dCNMO/dTPT3, dNaM/dTPT3, dCNMO/dTAT1 или dNaM/dTAT1.

55. Клетка по любому из пп. 45 или 47-54, отличающаяся тем, что, по меньшей мере, одна молекула ДНК не природного происхождения содержит, по меньшей мере, одну плазмиду.

56. Клетка по любому из пп. 45 или 47-54, отличающаяся тем, что, по меньшей мере, одна молекула ДНК не природного происхождения интегрирована в геном клетки.

57. Клетка по любому из пп. 47-56, отличающаяся тем, что, по меньшей мере, одна молекула ДНК не природного происхождения кодирует полипептид не природного происхождения.

58. Клетка по любому из вариантов реализации 45-57, отличающаяся тем, что клетка экспрессирует переносчик нуклеозидтрифосфата.

59. Клетка по п. 58, отличающаяся тем, что транспортер нуклеозидтрифосфата содержит аминокислотную последовательность *PtNTT2*.

60. Способ по п. 59, отличающийся тем, что транспортер нуклеозидтрифосфатов содержит усеченную аминокислотную последовательность *PtNTT2*, необязательно, где усеченная аминокислотная последовательность *PtNTT2*, по меньшей мере, на 80% идентична *PtNTT2*, кодируемой SEQ ID NO. 1.

61. Клетка по любому из пп. 45-60, отличающаяся тем, что клетка экспрессирует, по меньшей мере, две тРНК синтетазы.

62. Клетка по п. 61, отличающаяся тем, что, по меньшей мере, две тРНК синтетазы представляют собой химерную *PyIRS* (ch*PyIRS*) и *M. jannaschii* *AzFRS* (*MjPAzFRS*).

63. Клетка по любому из пп. 45-62, отличающаяся тем, что клетка содержит нуклеотиды не природного происхождения, включающие сахарную группу не природного происхождения.

64. Клетка по п. 63, отличающаяся тем, что сахарная группа не природного происхождения выбрана из группы, состоящей из:

модификации в положении 2', содержащей:

ОН, замещенный низший алкил, алкарил, аралкил, О-алкарил или О-аралкил, SH,

SCH₃, OCN, Cl, Br, CN, CF₃, OCF₃, SOCH₃, SO₂CH₃, ONO₂, NO₂, N₃ или NH₂F;

O-алкил, S-алкил, N-алкил;

O-алкенил, S-алкенил, N-алкенил;

O-алкинил, S-алкинил, N-алкинил;

O-алкил-O-алкил, 2'-F, 2'-OCH₃, 2'-O(CH₂)₂OCH₃, отличающийся тем, что алкил, алкенил и алкинил могут быть замещены или не замещены C₁-C₁₀ алкилом, C₂-C₁₀ алкенилом, C₂-C₁₀ алкинилом, O[(CH₂)_nO]_mCH₃, -O(CH₂)_nOCH₃, -O(CH₂)_nNH₂, -O(CH₂)_nCH₃, -O(CH₂)_n-NH₂ и -O(CH₂)_nON[(CH₂)_nCH₃]₂, отличающийся тем, что n и m составляют от 1 до примерно 10;

модификации в положении 5', содержащей:

5'-винил, 5'-метил (R или S); или

модификации в положении 4', 4'-S, гетероциклоалкила, гетероциклоалкарила, аминокислотамино, полиалкиламино, замещенного силила, группы, расщепляющей РНК, репортерной группы, интеркалятора, группы для улучшения фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида; или

любой их комбинации.

65. Клетка по любому из пп. 45-64, отличающийся тем, что, по меньшей мере, одно нуклеотидное основание не природного происхождения распознается РНК полимеразой во время транскрипции.

66. Клетка по любому из пп. 45-65, отличающийся тем, что клетка транслирует, по меньшей мере, один полипептид не природного происхождения, содержащий, по меньшей мере, две аминокислоты не природного происхождения.

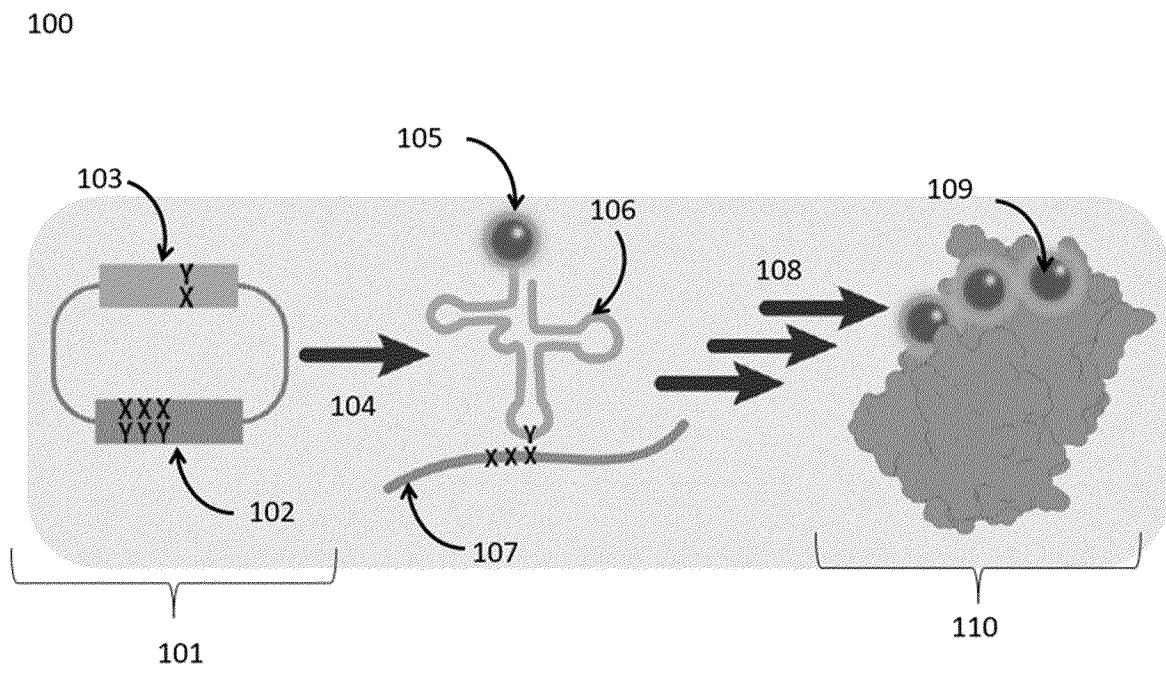
67. Клетка по любому из пп. 45-66, отличающийся тем, что, по меньшей мере, две аминокислоты не природного происхождения независимо выбраны из группы, состоящей из N⁶-((азидоэтокси)-карбонил)-L-лизина (AzK), N⁶-пропаргилэтокси-карбонил-L-лизина (PraK), N⁶-((пропаргилокси)карбонил)-L-лизина (PrK), п-азидофенилаланина (pAzF), BCN-L-лизина, норборненилизина, TCO-лизина, метилтетразинлизина, аллилкарбониллизина, 2-амино-8-оксононановой кислоты, 2-амино-8-оксооктановой кислоты, п-ацетил-L-фенилаланина, п-азидометил-L-фенилаланина (pAMF), п-йодо-L-фенилаланина, м-ацетилфенилаланина, 2-амино-8-оксононаноевой кислоты, п-пропаргилоксифенилаланина, п-пропаргилфенилаланина, 3-метилфенилаланина, L-Dopa, фторированного фенилаланина, изопропил-L-фенилаланина, п-азидо-L-фенилаланина, п-ацетил-L-фенилаланина, п-бензоил-L-фенилаланина, п-бромфенилаланина, п-амино-L-фенилаланина, изопропил-L-фенилаланина, O-аллилтирозина, O-метил-L-тирозина, O-4-аллил-L-тирозина, 4-пропил-L-тирозина, фосфонтитрозина, три-O-ацетил-GlcNAc-серина, L-фосфосерина, фосфоносерина, L-3-(2-нафтил)аланина, 2-амино-3-((2-((3-(бензилокси)-3-оксопропил)амино)этил)селанил)пропановой кислоты, 2-амино-3-(фенилселанил)пропановой кислоты, селеноцистеина, N⁶-(((2-азидобензил)окси)карбонил)-L-лизина, N⁶-(((3-азидобензил)окси)карбонил)-L-лизина и

№6-(((4-азидобензил)окси)карбонил)-L-лизина.

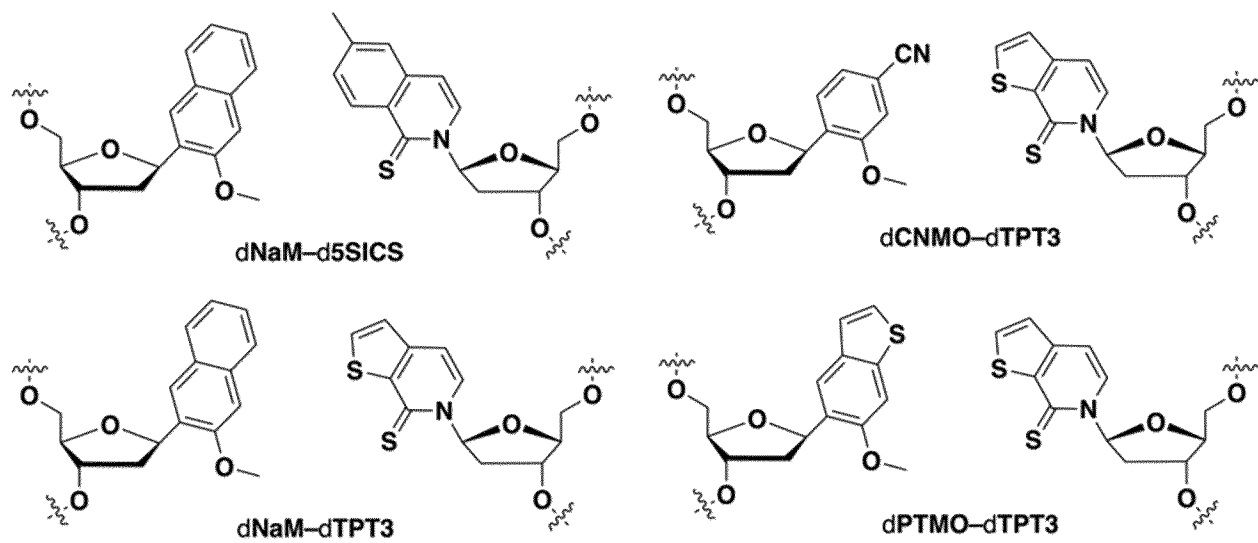
68. Клетка по любому из пп. 45-67, отличающаяся тем, что клетка выделена.
69. Клетка по любому из пп. 45-68, отличающийся тем, что клетка представляет собой прокариот.
70. Линия клеток, содержащая клетку по любому из пп. 45-69.

1/24

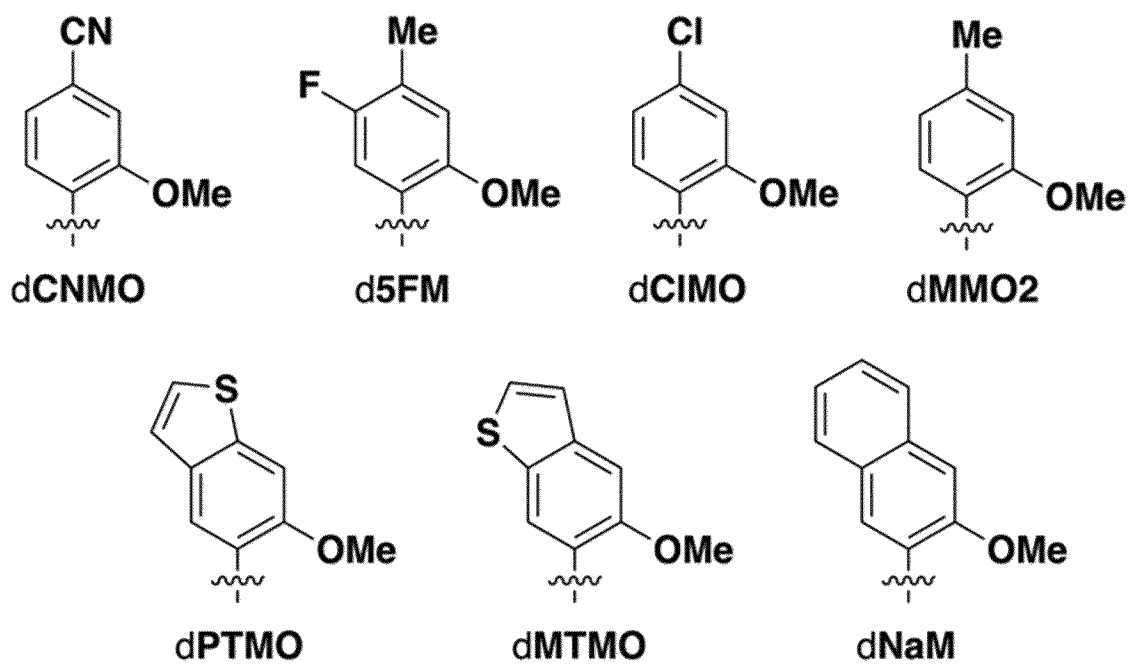
ФИГ. 1



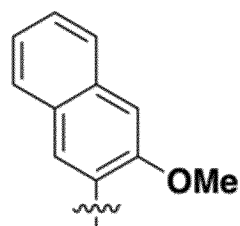
ФИГ. 2



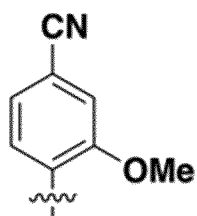
ФИГ. 3



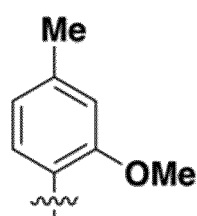
ФИГ. 4А



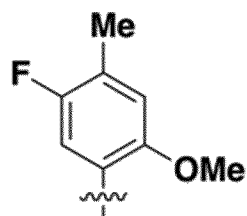
NaM



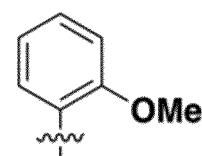
CNMO



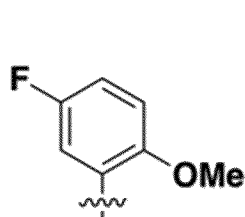
MMO2



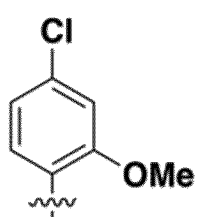
5FM



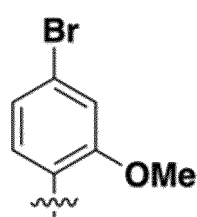
2OMe



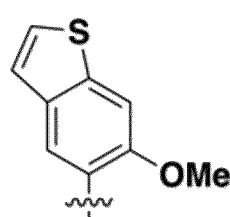
5F2OMe



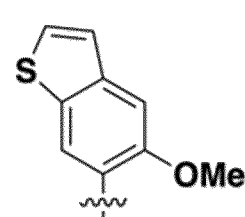
ClMO



BrMO

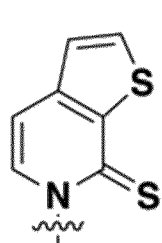


PTMO

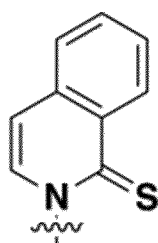


MTMO

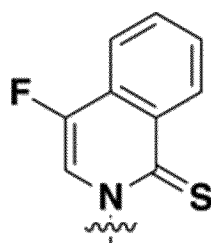
ФИГ. 4В



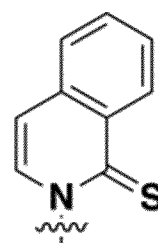
TPT3



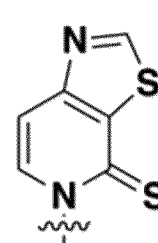
SICS



FSICS

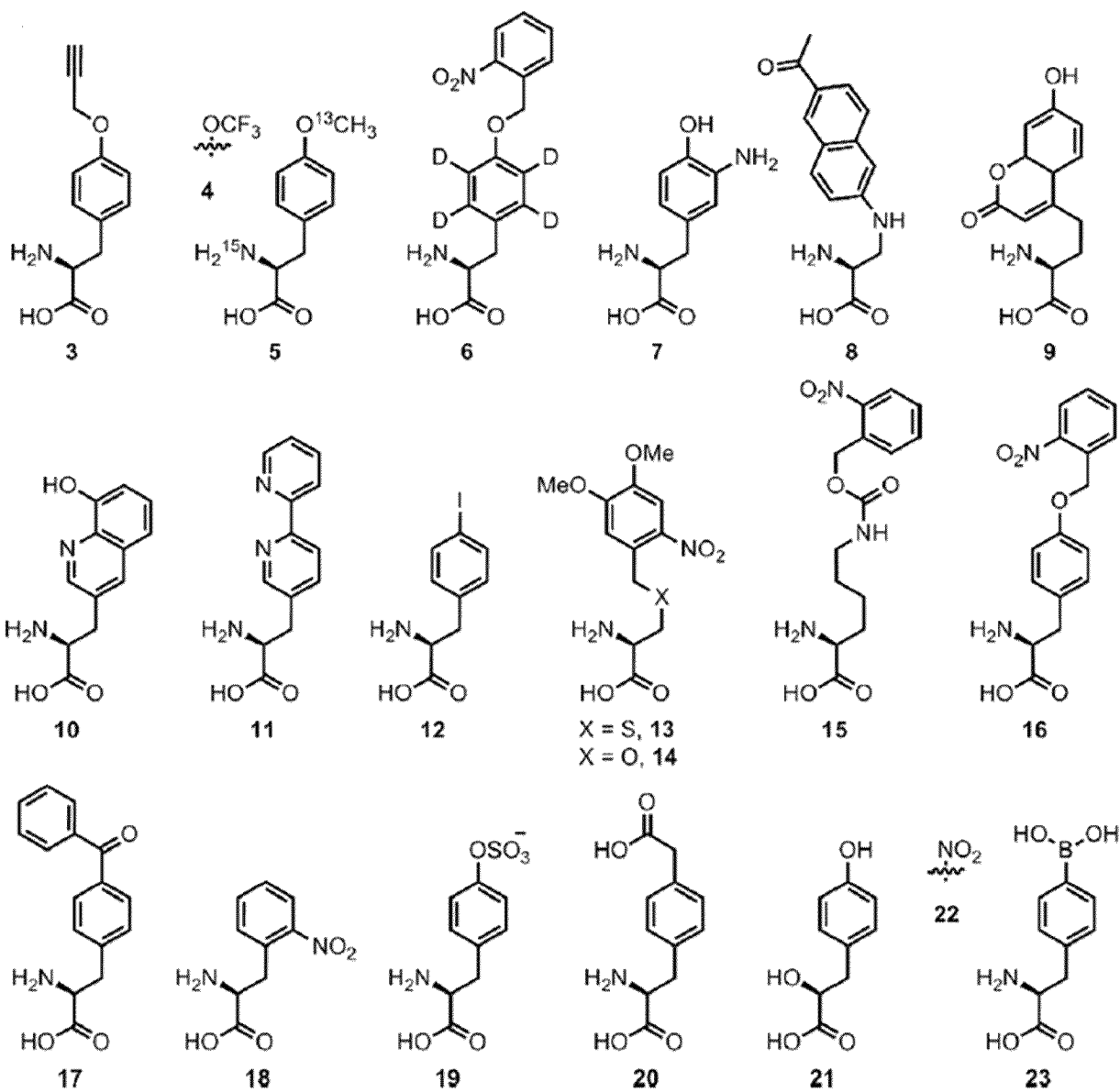


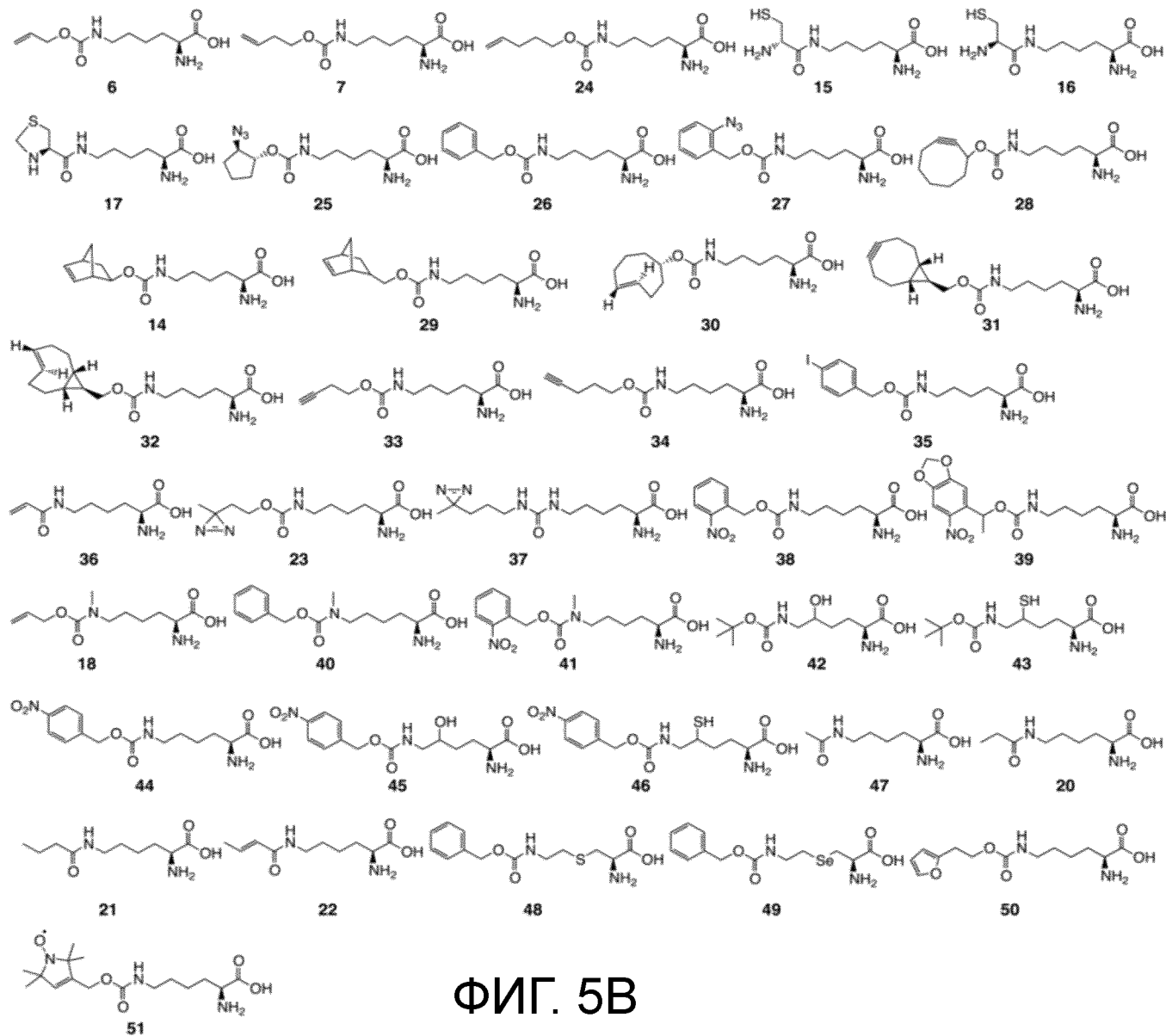
5SICS



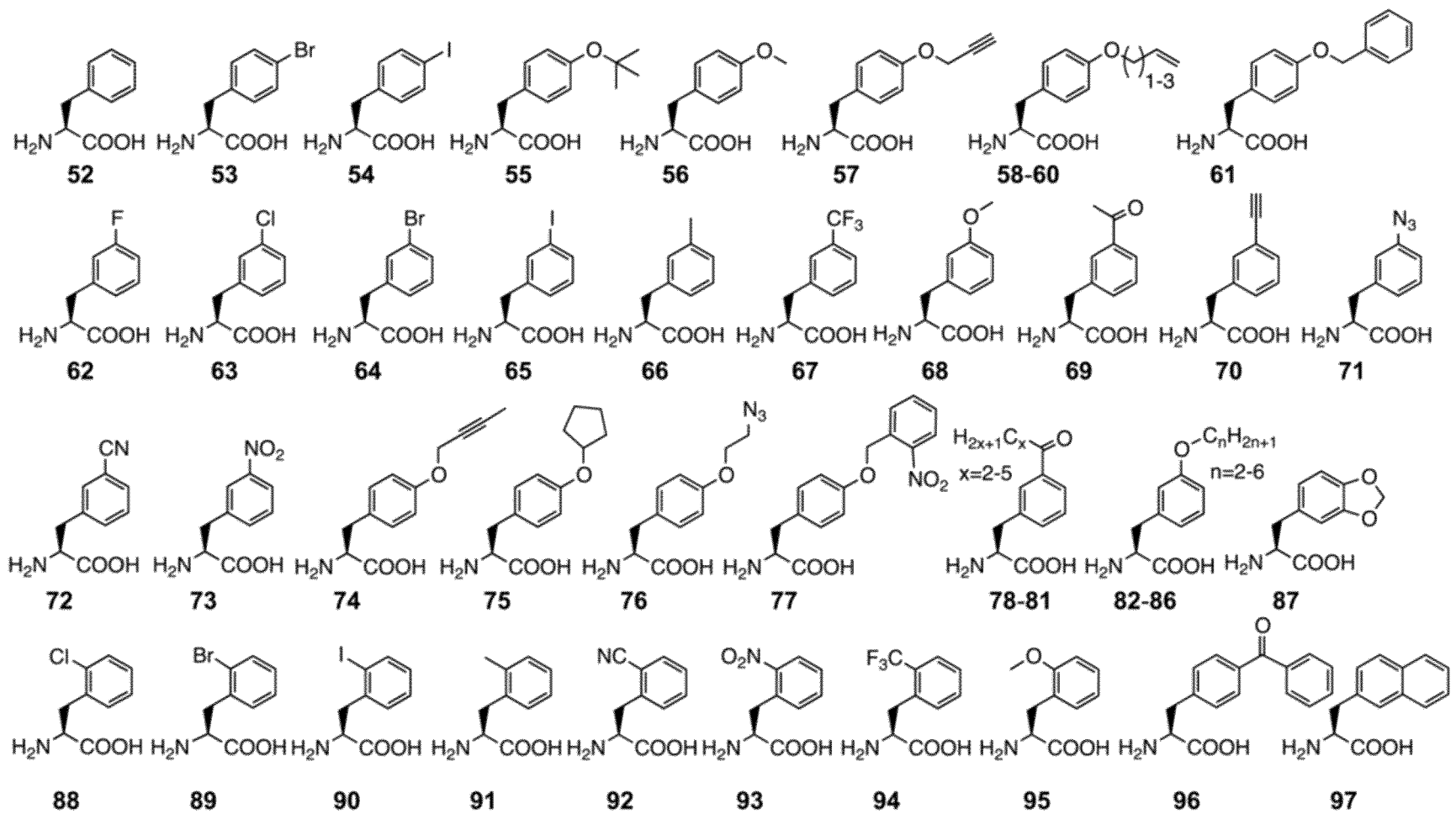
TAT1

ФИГ. 5А



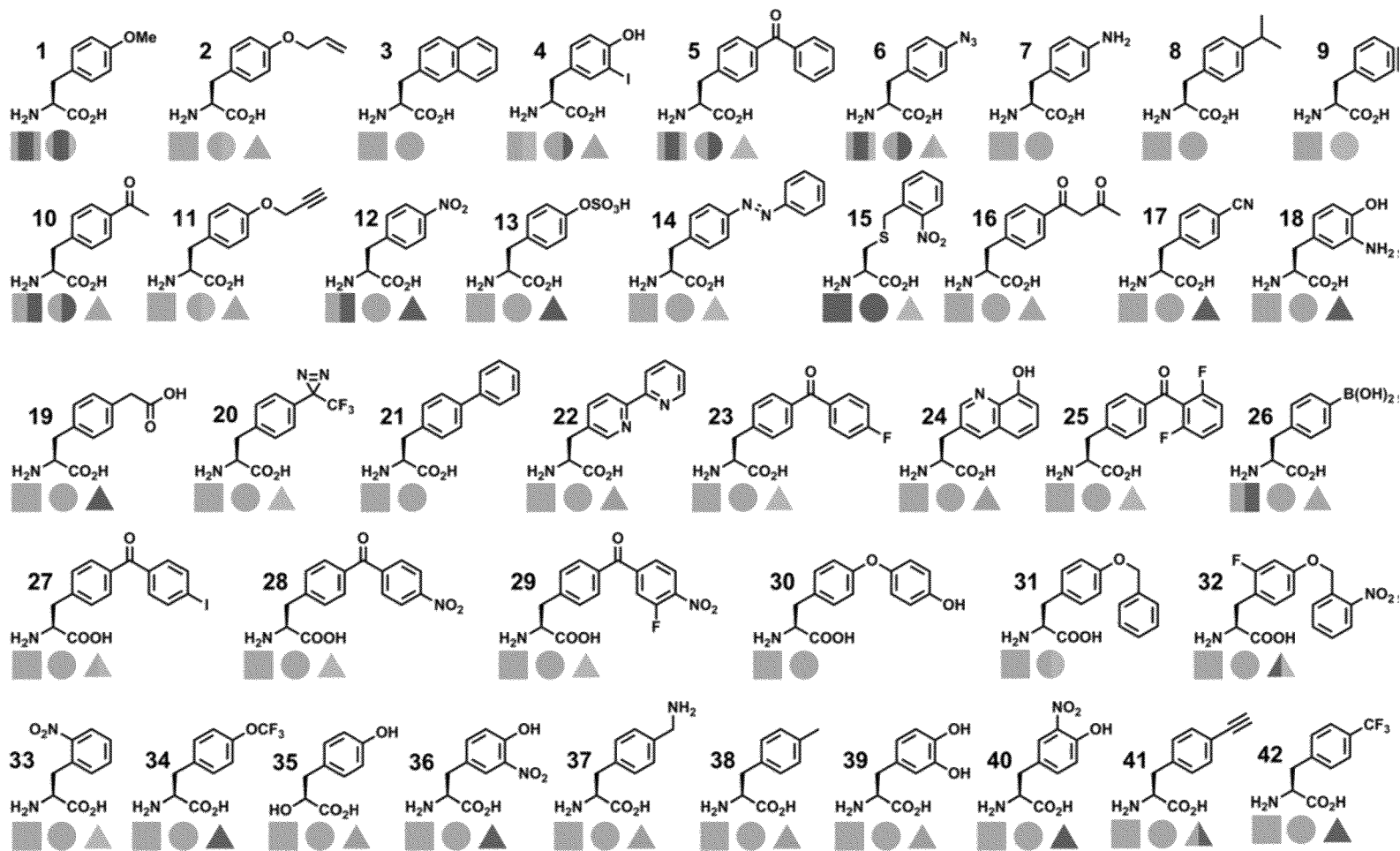


ФИГ. 5В

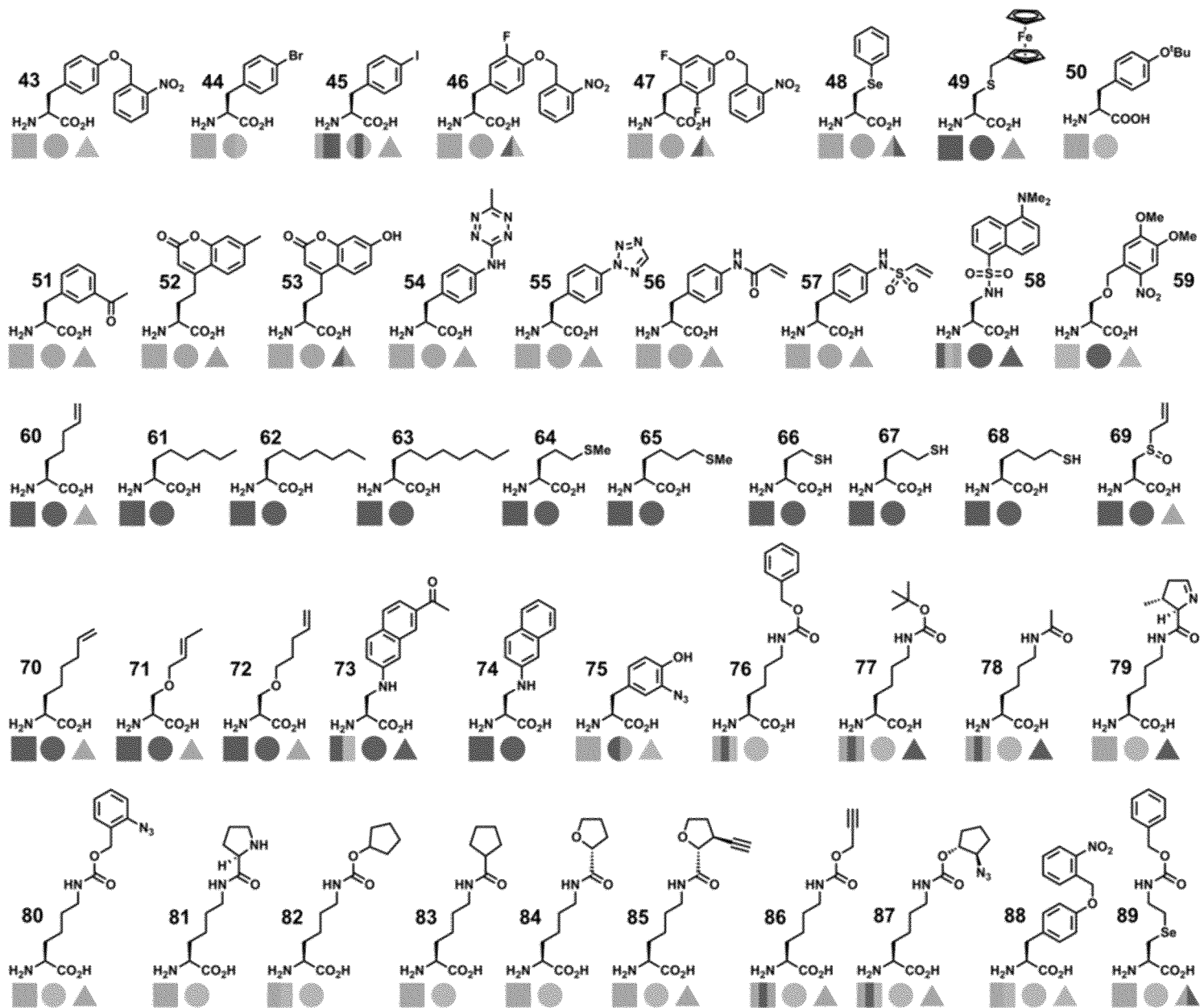


7/24

ФИГ. 5С

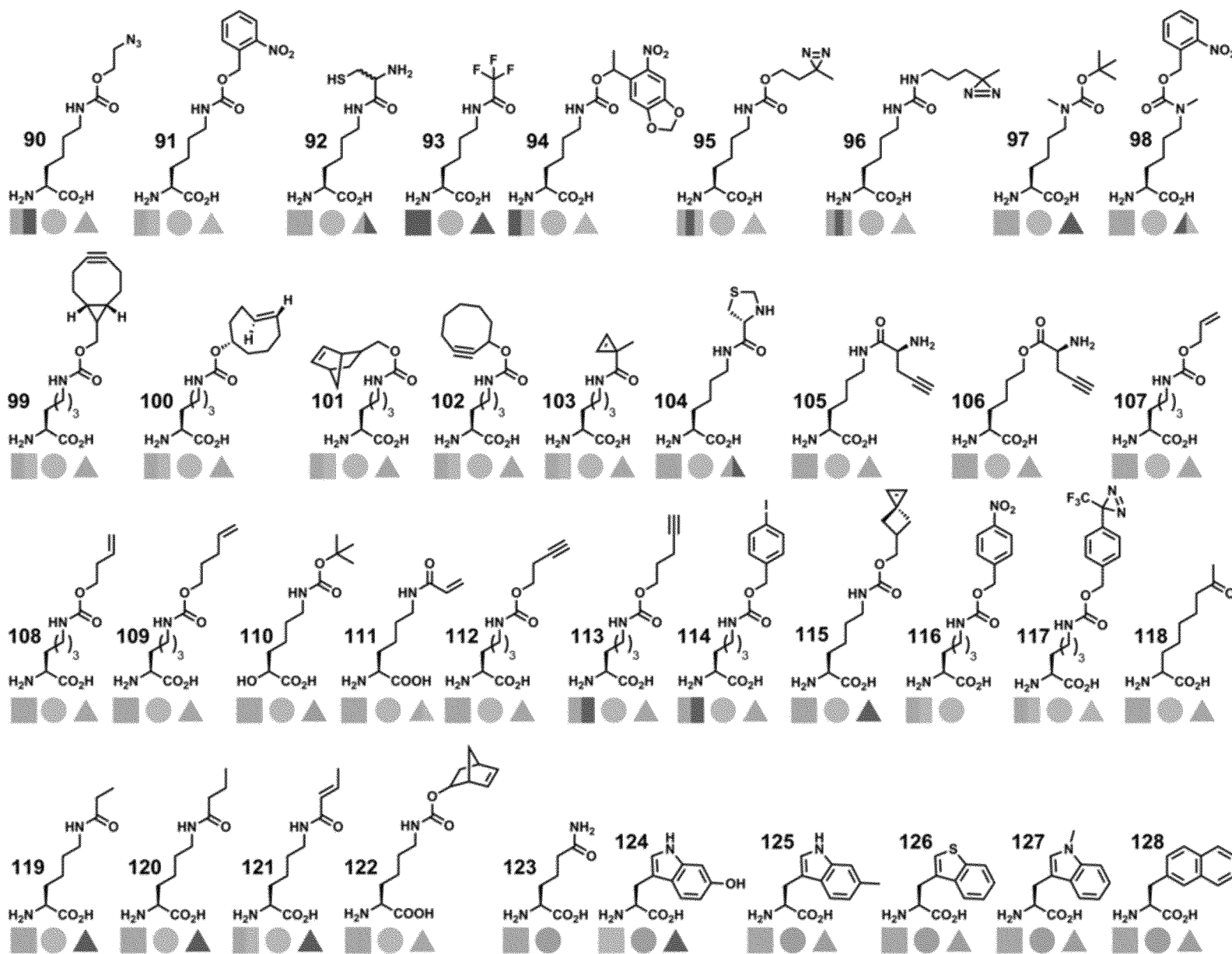


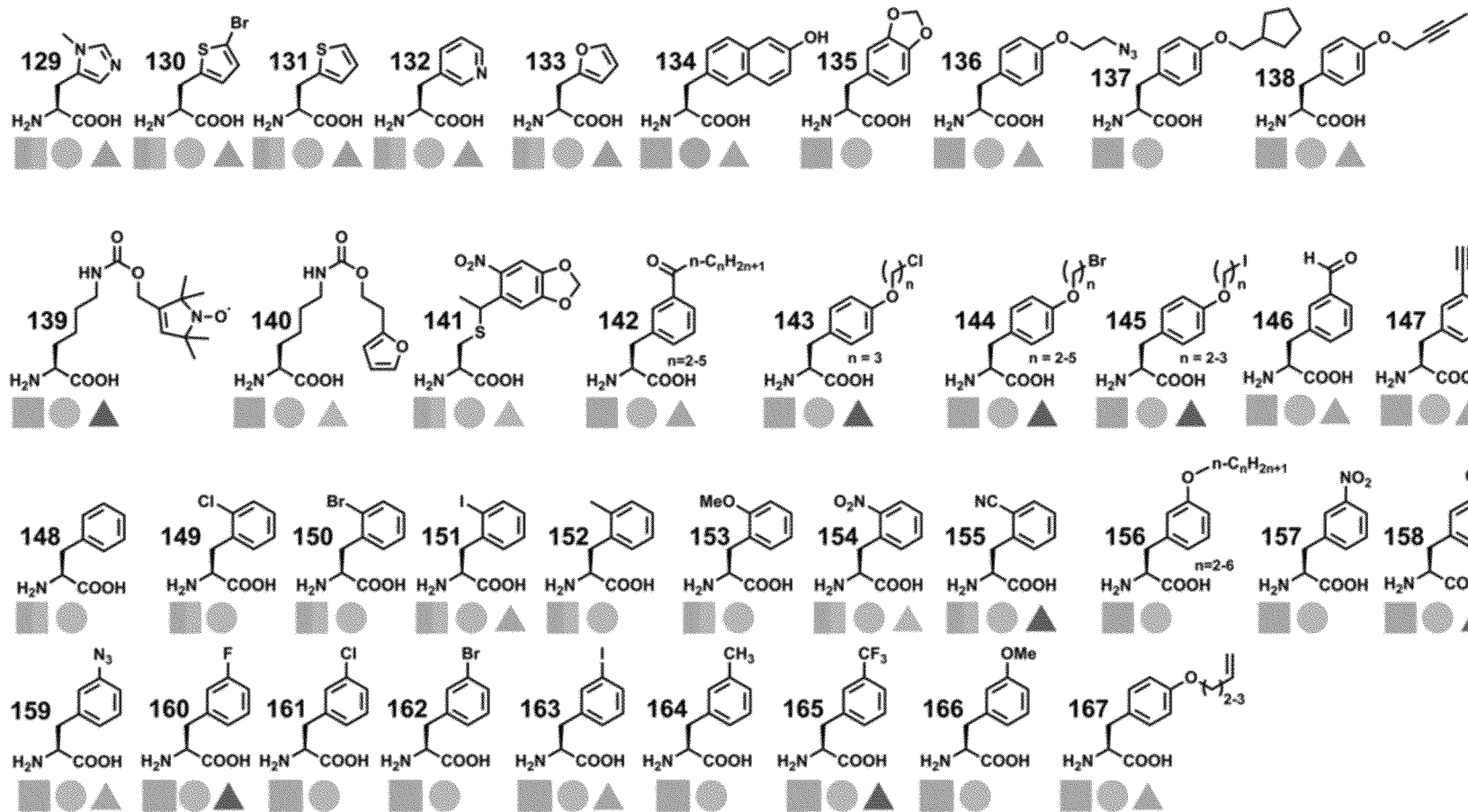
ФИГ. 5D



ФИГ. 5Е

ФИГ. 5F





Организм

пара аатрансфераза/тРНК

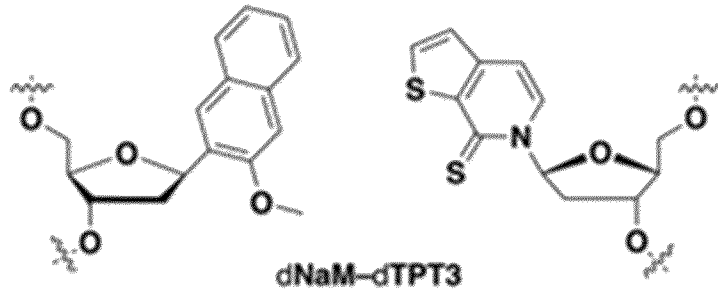
Функция

- E. Coli
- другие прокариоты
- Дрожжи
- эукариотические (клеточная культура)
- эукариотические (организм)
- Mj Tyr
- Ec Leuc
- Ec Tyr
- Mm/Mb tyr
- другие

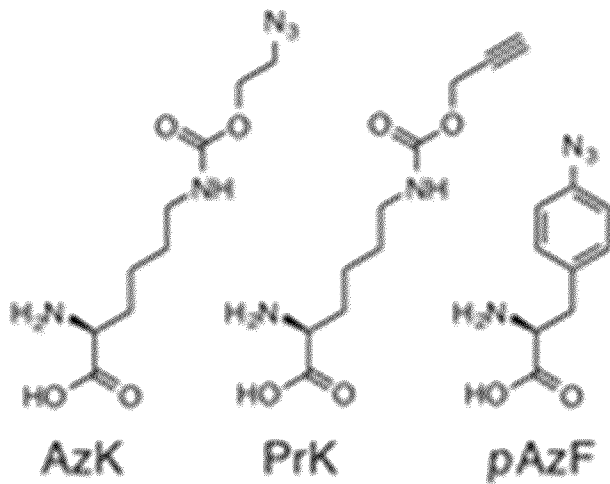
- ▲ Биоконъюгация
- ▲ Репортер
- ▲ Природный РТМ (или предшественник)
- ▲ Фотореактивный (поперечное сшивание/заклучение)
- ▲ Новая активность (хелатирующая, окислительно-восстановительная)

ФИГ. 5G

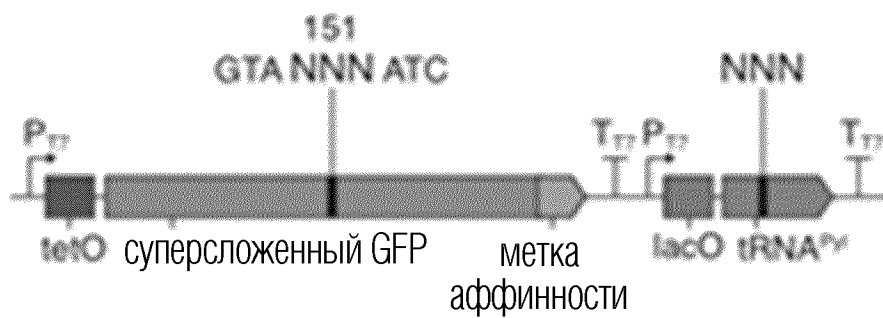
ФИГ. 6А

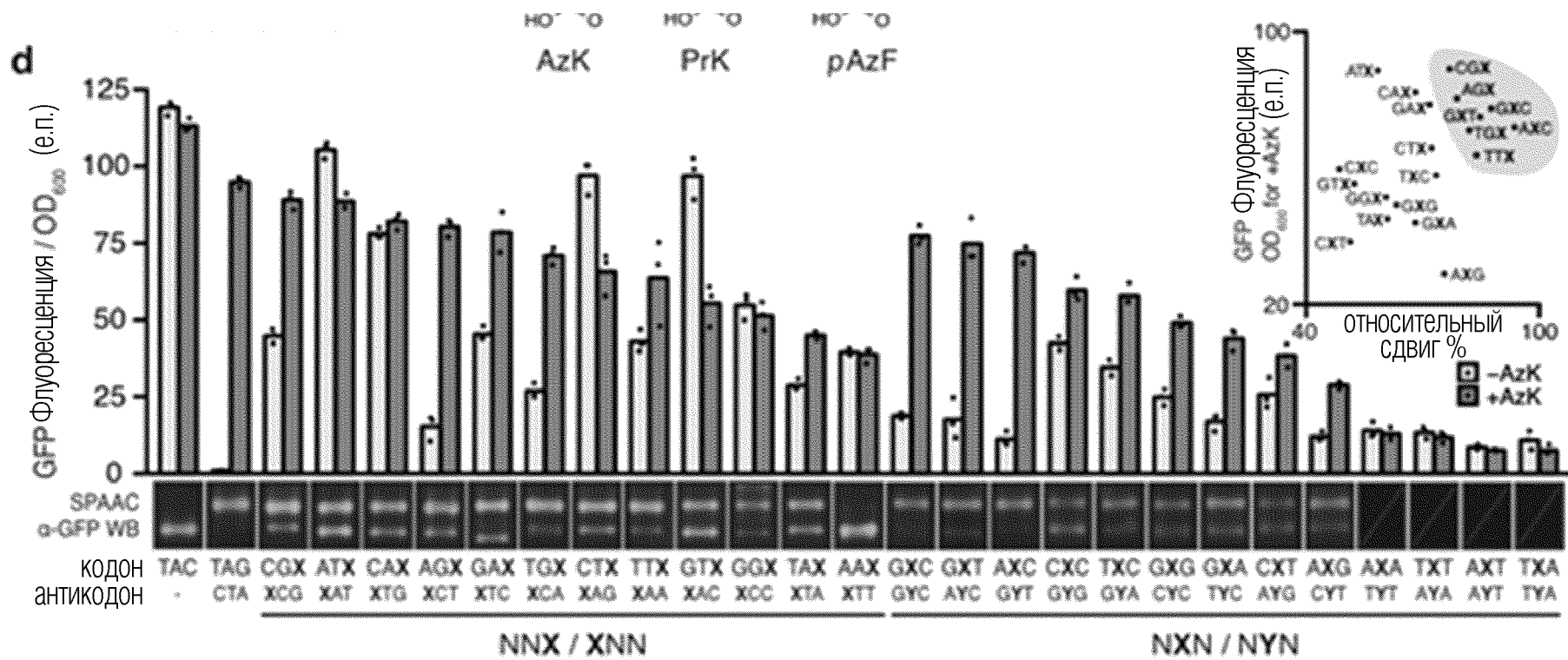


ФИГ. 6В



ФИГ. 6С

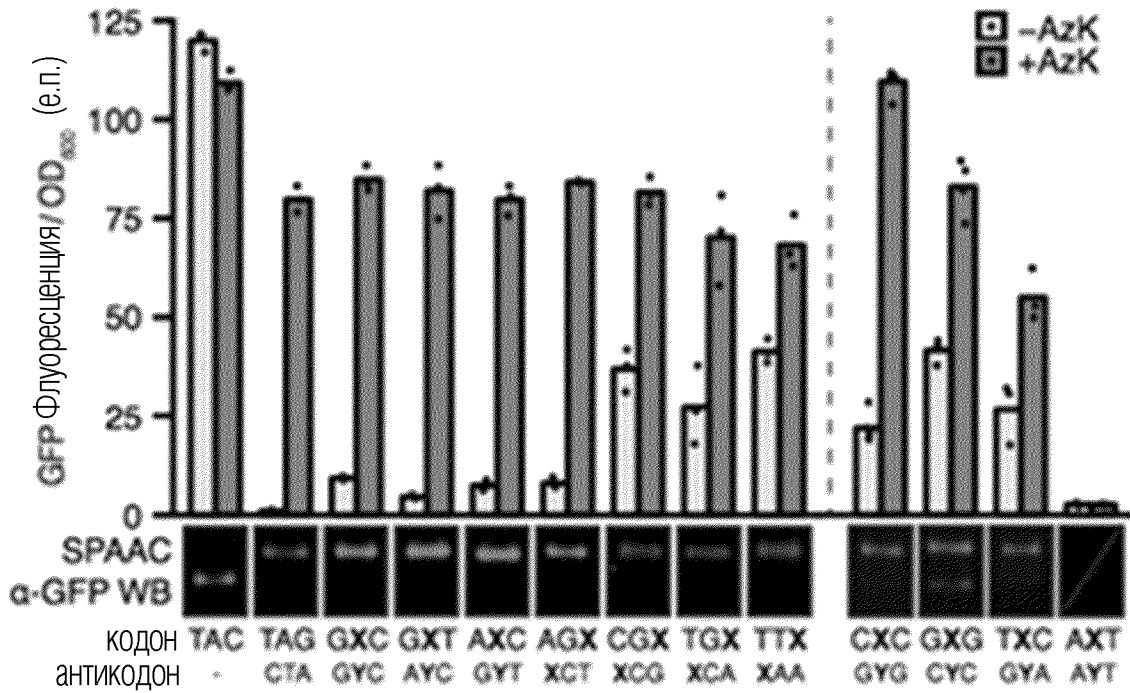




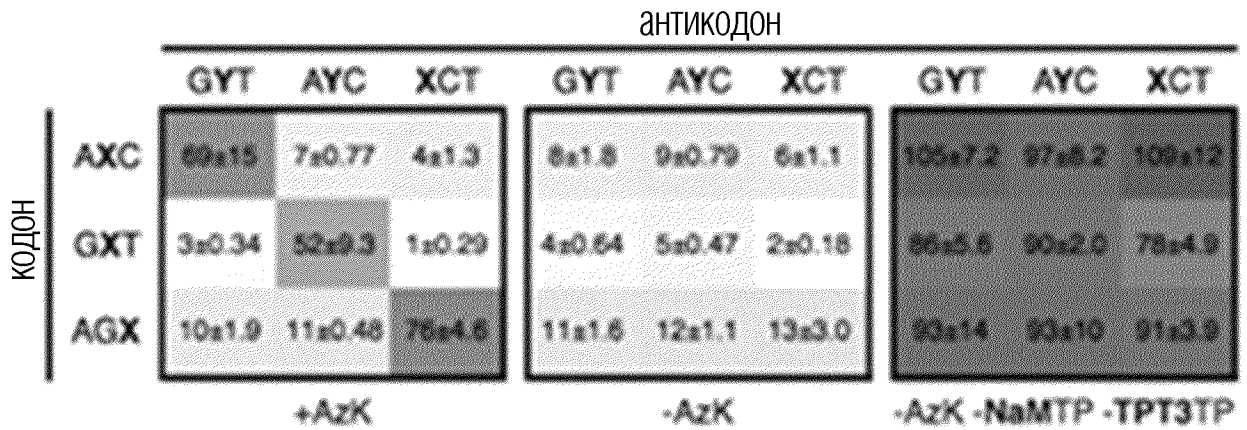
13/24

ФИГ. 6D

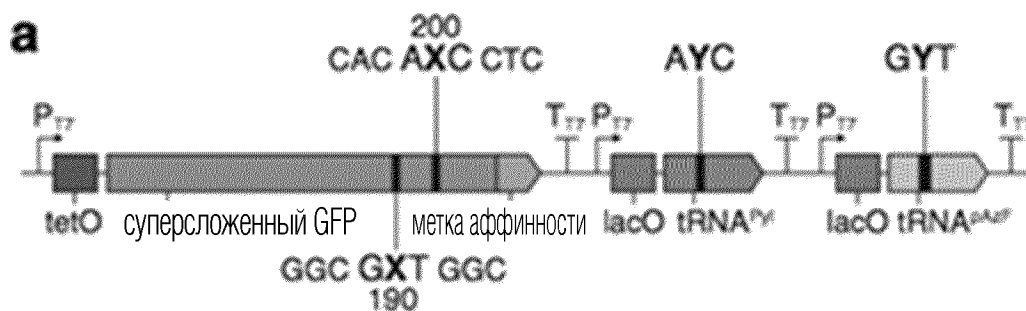
ФИГ. 7А



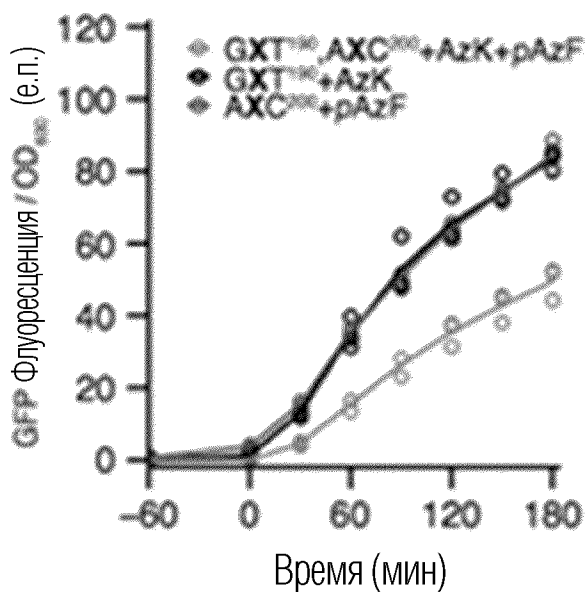
ФИГ. 7В



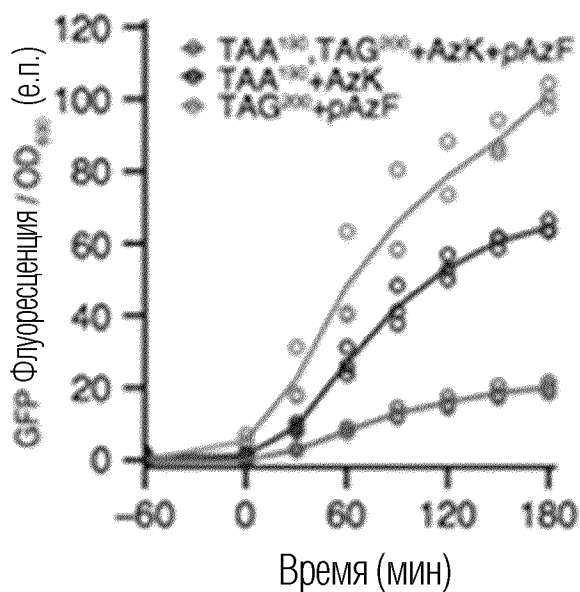
15/24
ФИГ. 8А



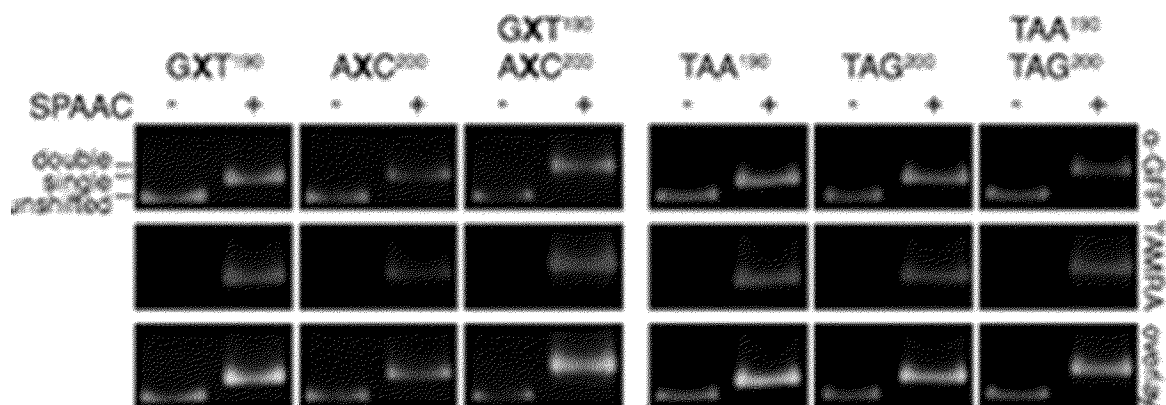
ФИГ. 8В



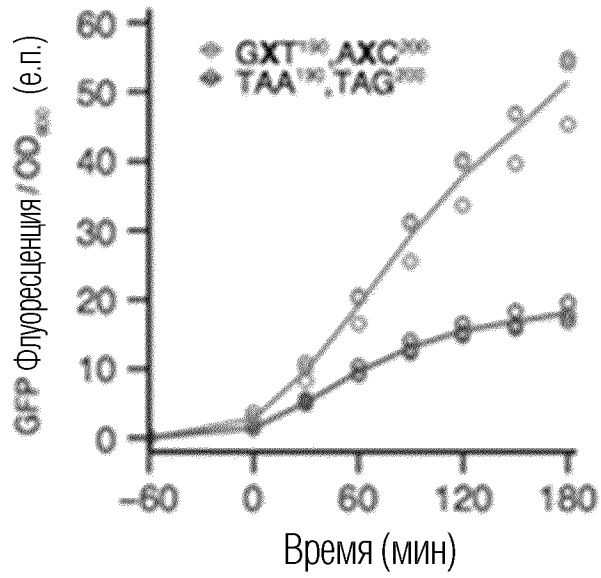
ФИГ. 8С



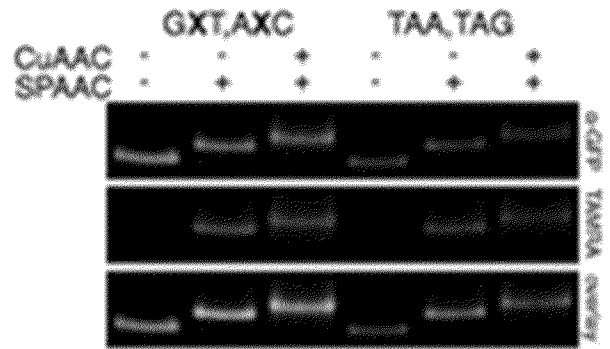
ФИГ. 8D



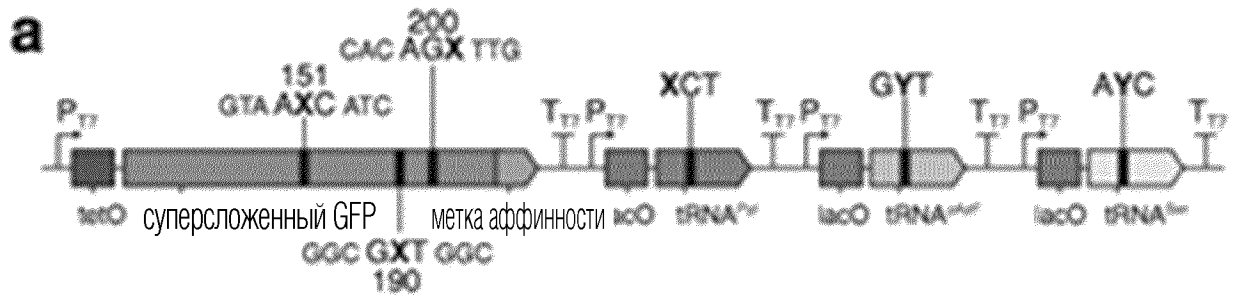
ФИГ. 8Е



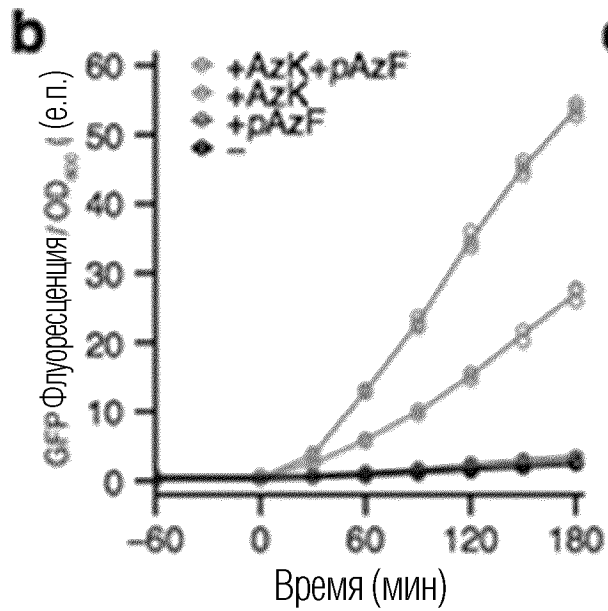
ФИГ. 8F



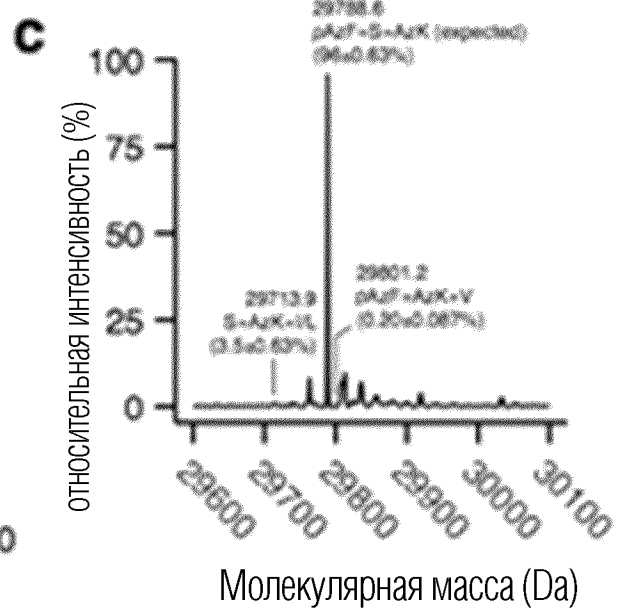
ФИГ. 9А



ФИГ. 9В

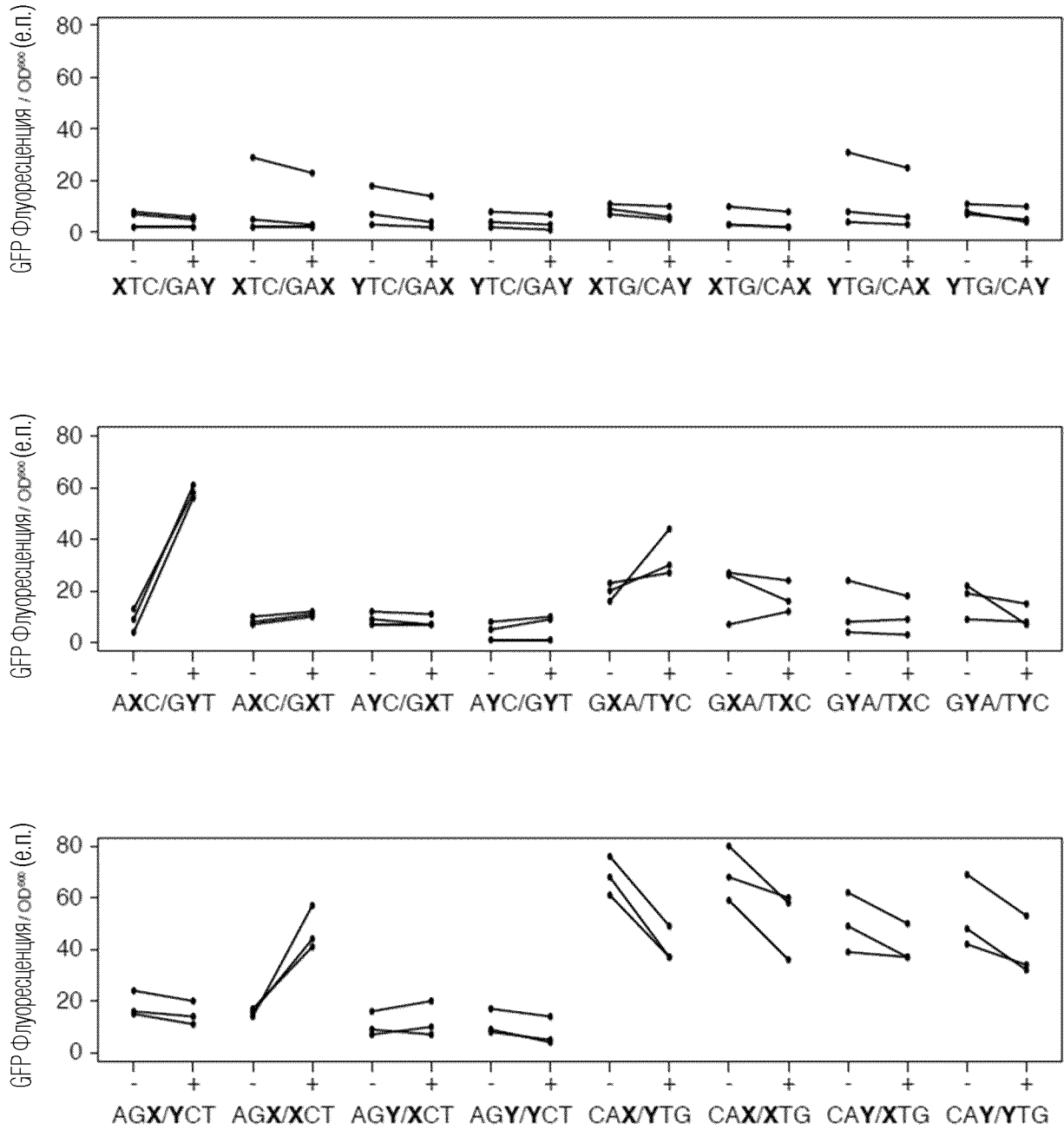


ФИГ. 9С

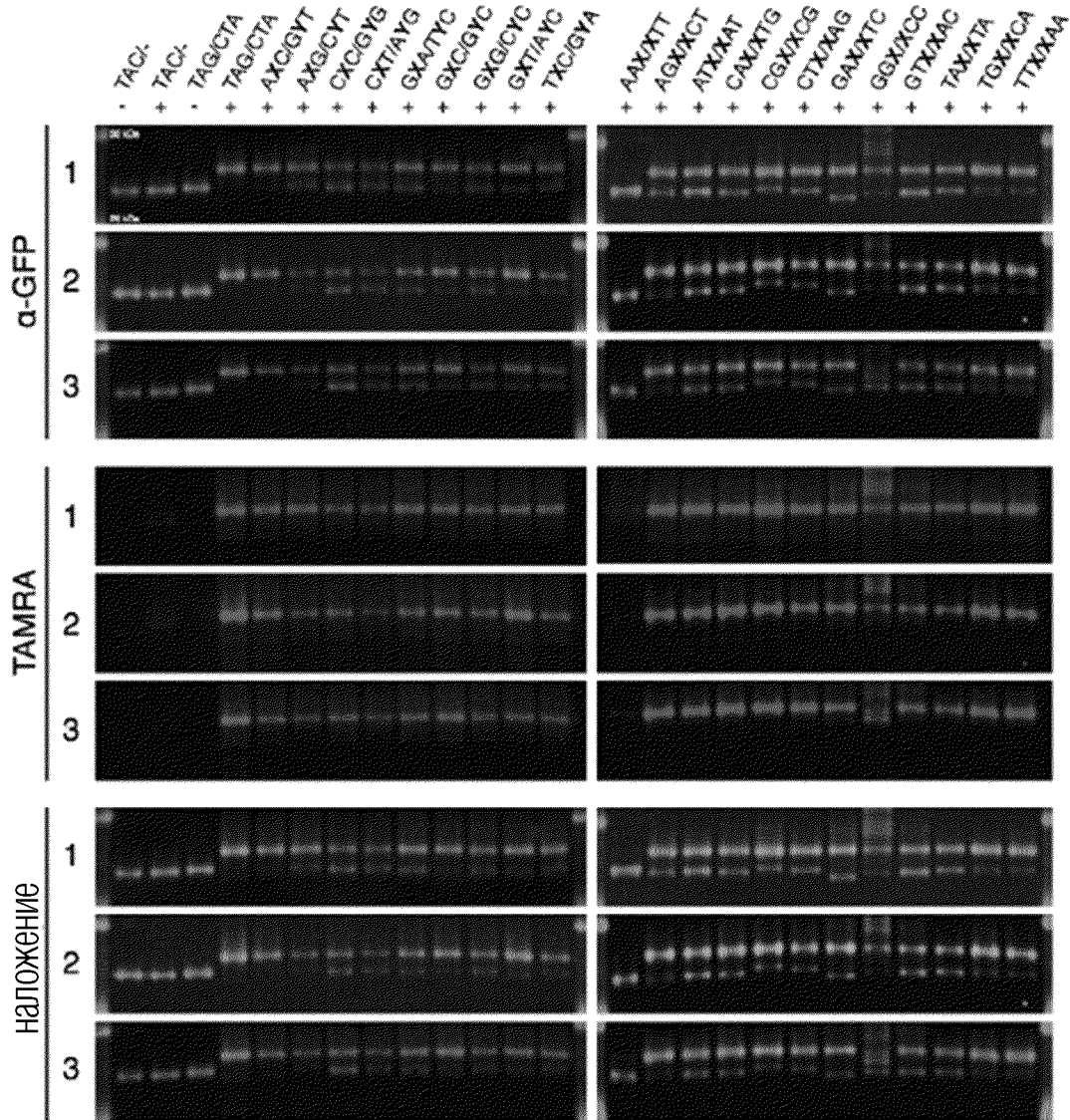


18/24

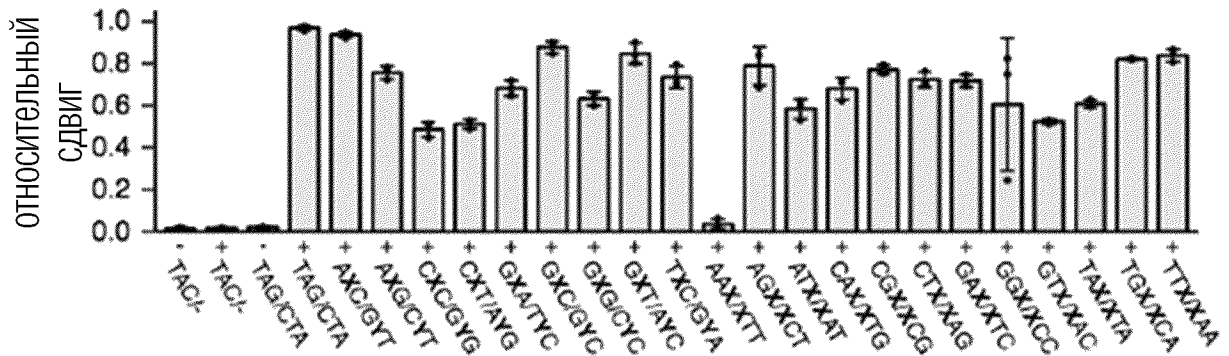
ФИГ. 10



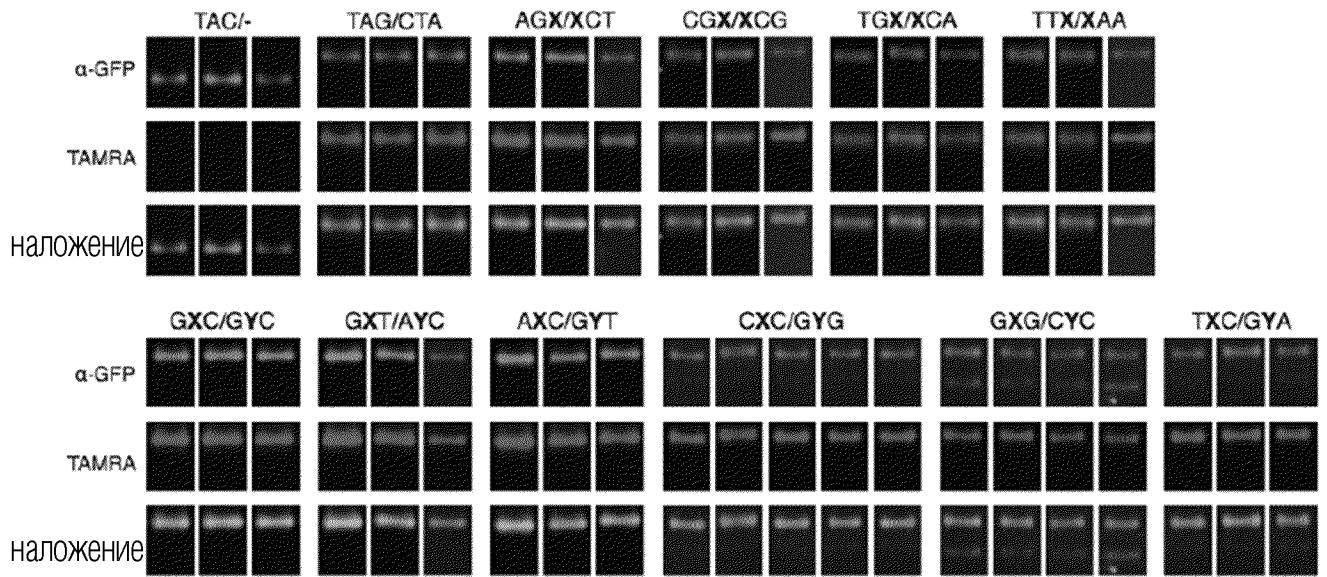
ФИГ. 11А



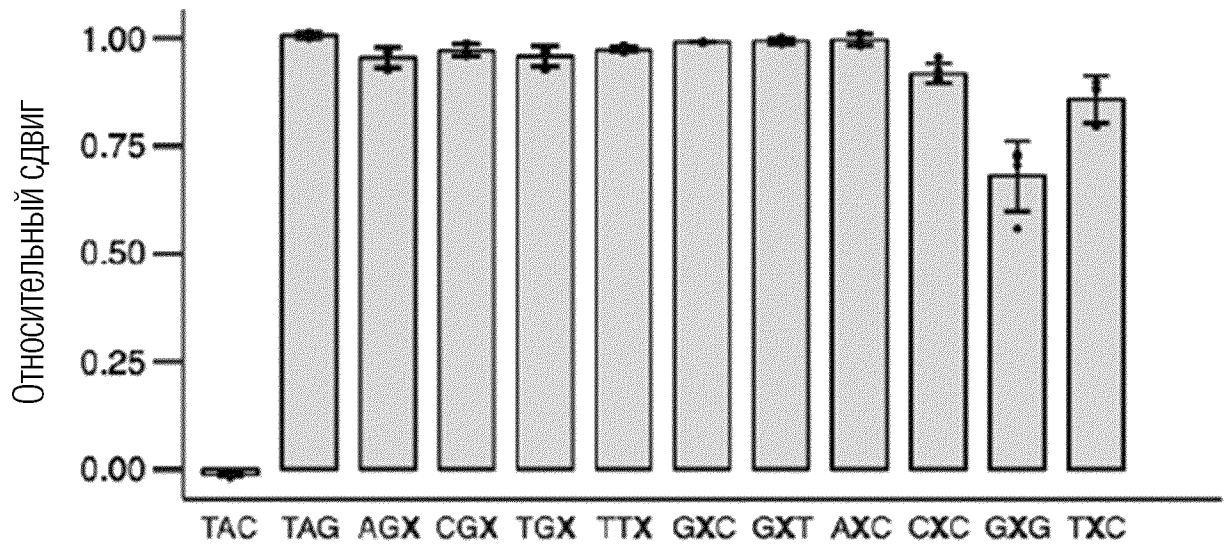
ФИГ. 11В



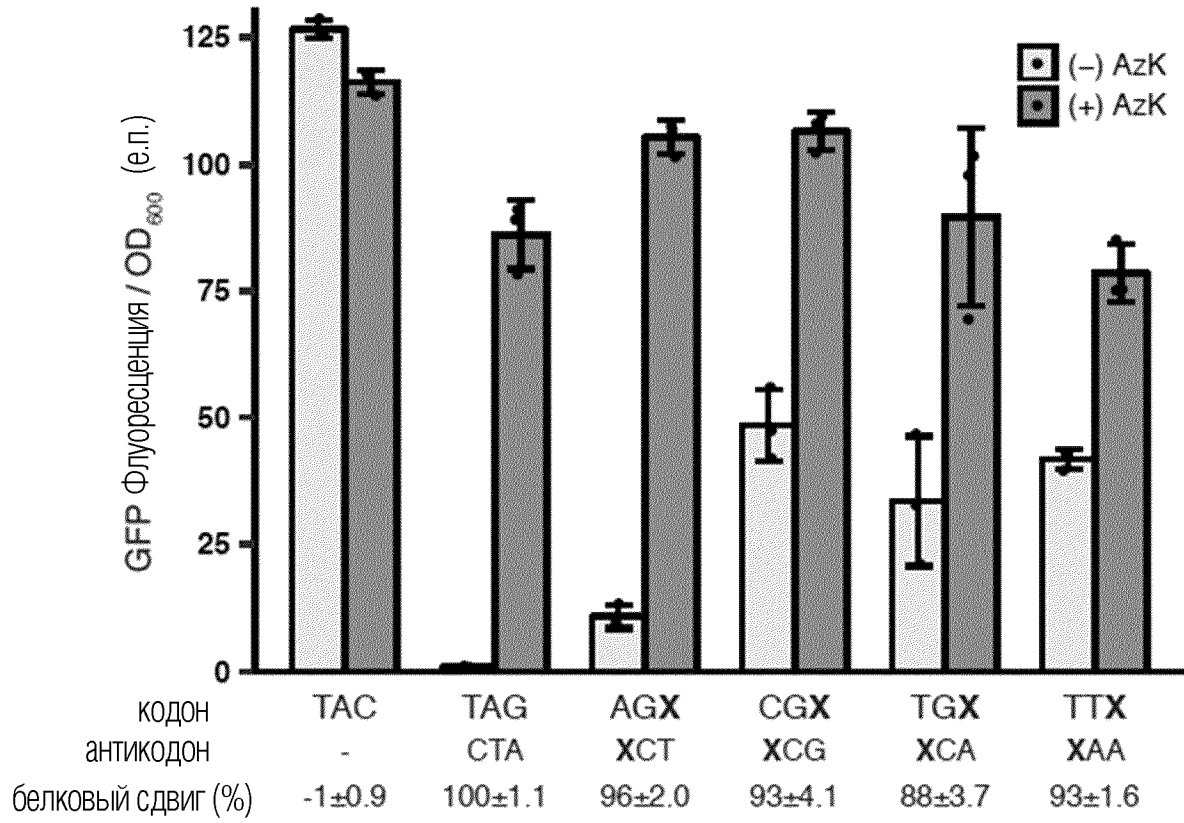
ФИГ. 12А



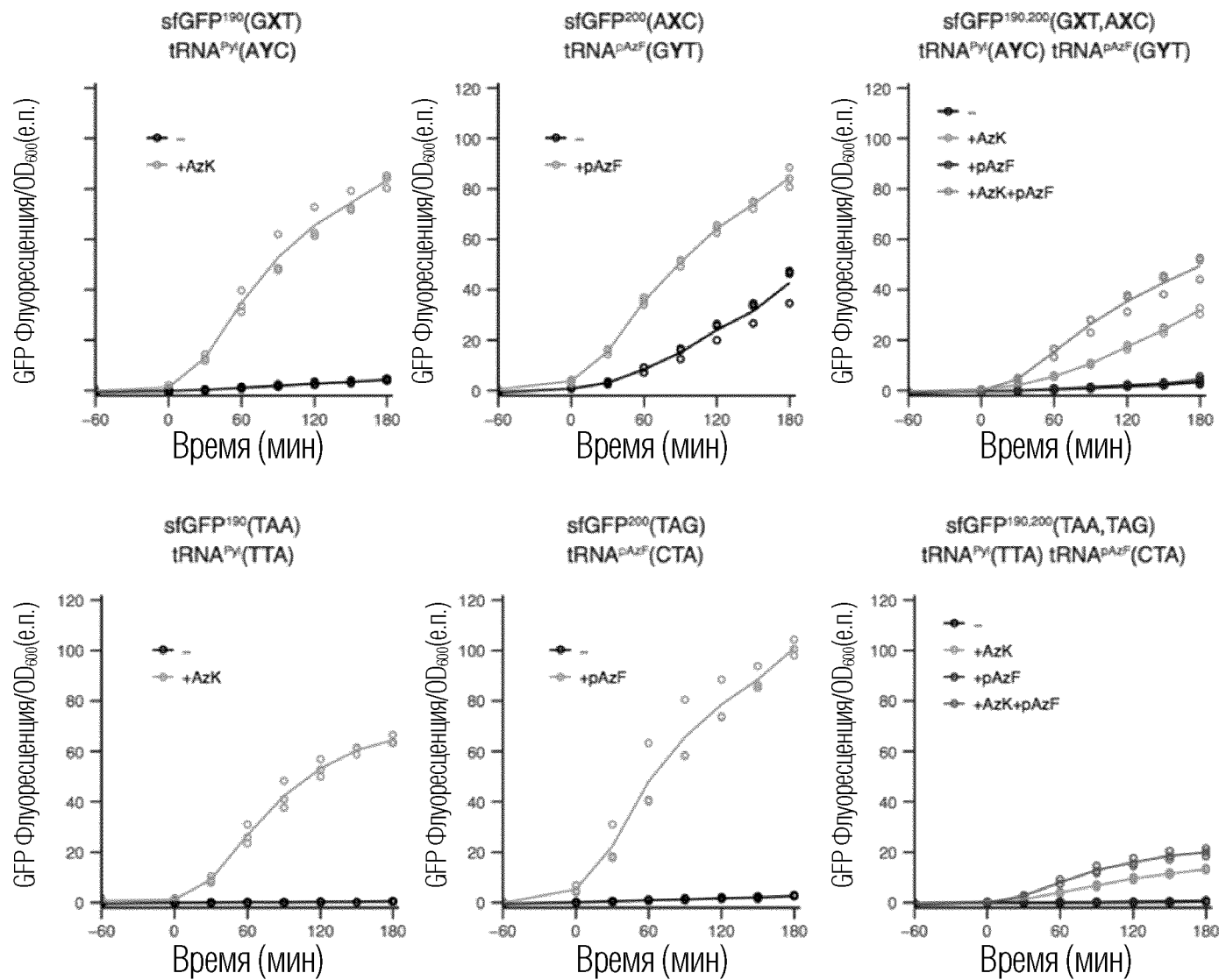
ФИГ. 12В



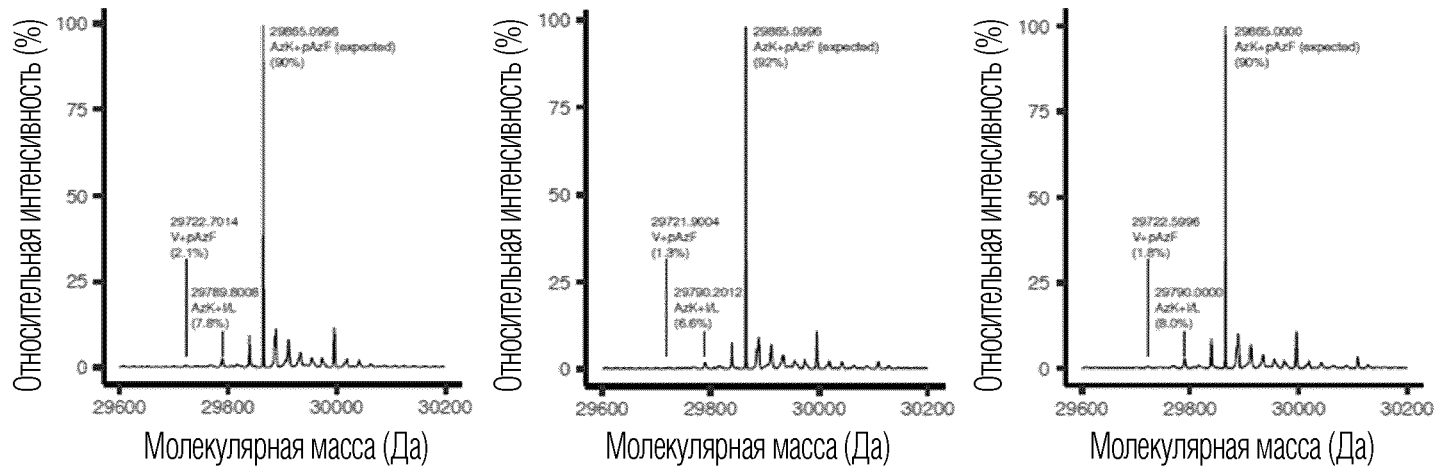
ФИГ. 13



ФИГ. 14



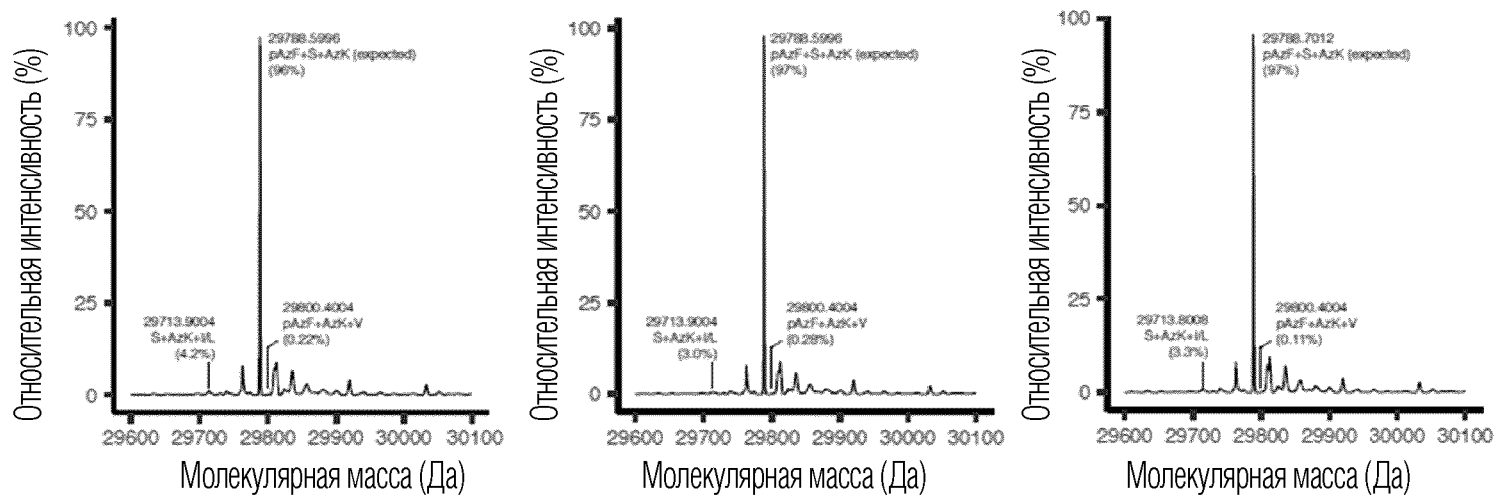
ФИГ. 15А



ФИГ. 15В

MW (Da)	Δ MW Относительно эталона, Da	Назначение	Интерпретация
29865.1	0.0	D190AzK Y200pAzF	ожидаемый продукт
29722.4	-142.7	D190V Y200pAzF	GXT→GTT
29790.0	-75.1	D190AzK Y200IL	AXC→ATC
29839.4	-25.7	Окисление аргининов	технический
29888.6	23.5	Na аддукт	технический
29911.7	46.6	2xNa аддукт	технический
29933.5	68.4	3xNa аддукт	технический
29995.9	130.8	+Met	непроцессированный N-конец
30108.7	243.6	+Biotin	технический

ФИГ. 16А



ФИГ. 16В

MW (Да)	Δ MW Относительно эталона, Да	Назначение	Интерпретация
29788.6	0.0	Y151pAzF D190S Y200AzK	ожидаемый продукт
29713.9	-74.7	Y151ML D190S Y200AzK	AXC→ATC
29763.0	-25.6	Окисление аргининов	технический
29801.2	12.6	Y151pAzF D190V Y200AzK	GXT→GTT
29812.3	23.7	Na аддукт	технический
29835.6	47.0	2xNa аддукт	технический
29857.1	68.5	3xNa аддукт	технический
29919.4	130.8	+Met	непроцессированный N-конец
30033.0	244.4	+Biotin	технический