

(19)



**Евразийское  
патентное  
ведомство**

(11) **042937**

(13) **B1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОМУ ПАТЕНТУ**

(45) Дата публикации и выдачи патента  
**2023.04.06**

(51) Int. Cl. *C12N 15/11* (2006.01)  
*C12P 21/00* (2006.01)

(21) Номер заявки  
**202090090**

(22) Дата подачи заявки  
**2018.07.10**

---

(54) **ВКЛЮЧЕНИЕ НЕПРИРОДНЫХ НУКЛЕОТИДОВ И СПОСОБЫ С НИМИ**

---

(31) **62/531,325**

(56) WO-A1-2016115168  
US-A1-20170029829  
US-B2-7045337

(32) **2017.07.11**

(33) **US**

(43) **2020.06.09**

(86) **PCT/US2018/041509**

(87) **WO 2019/014267 2019.01.17**

(71)(73) Заявитель и патентовладелец:  
**СИНТОРКС, ИНК.; ДЗЕ СКРИППС  
РИСЕРЧ ИНСТИТЮТ (US)**

(72) Изобретатель:  
**Птацин Джерод, Каффаро Каролина,  
Аэрни Ханс, Чжан Йорке, Фишер  
Эмиль К., Фельдман Аарон У., Диев  
Вивиан Т., Ромесберг Флойд Е. (US)**

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

---

(57) В изобретении раскрыты способы синтеза белков, которые включают в себя неприродные аминокислоты, которые используют мутантную tRNA.

**B1**

**042937**

**042937**

**B1**

## Ссылка

По настоящей заявке испрашивается приоритет в соответствии с предварительной заявкой на патент США № 62/531325, поданной 11 июля 2017 г., которая полностью включена в настоящий документ посредством ссылки.

**Предшествующий уровень техники настоящего изобретения**

Олигонуклеотиды и их применение произвели революцию в биотехнологии. Однако каждый из олигонуклеотидов, включая в себя как ДНК, так и РНК, содержит только четыре природных нуклеотида: аденозин (А), гуанозин (G), цитозин (С), тимин (Т) для ДНК и четыре природных нуклеотида: аденозин (А), гуанозин (G), цитозин (С) и уридин (U) для РНК, и это значительно ограничивает потенциальные функции и применения олигонуклеотидов.

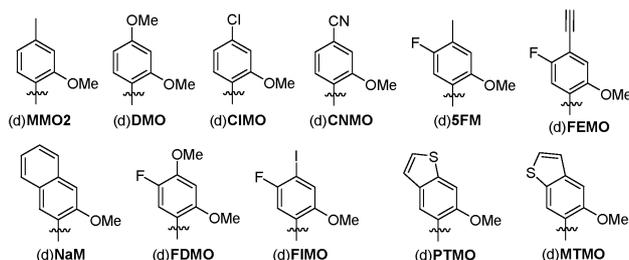
Способность специфически к последовательности синтезировать/амплифицировать олигонуклеотиды (ДНК или РНК) с полимеразами, например, с помощью ПЦР или систем изотермической амплификации (например, транскрипции с помощью РНК-полимеразы T7), произвела революцию в биотехнологии. В дополнение ко всем потенциальным применениям в нанотехнологиях, это позволило использовать широкий спектр новых технологий, таких как эволюция *in vitro* с помощью SELEX (систематическая эволюция лигандов экспоненциальным обогащением) РНК и ДНК-аптамеров и ферментов. См., например, Oliphant AR, Brandl CJ & Struhl K (1989), Defining the sequence specificity of DNA-binding proteins by selecting binding sites from random-sequence oligonucleotides: analysis of yeast GCN4 proteins, *Mol. Cell Biol.*, 9:2944-2949; Tuerk C & Gold L (1990), Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase, *Science*, 249:505-510; Ellington AD & Szostak JW (1990), *In vitro* selection of RNA molecules that bind specific ligands, *Nature*, 346:818-822.

Согласно некоторым аспектам эти применения ограничены характеризующимся пределом химическим/физическим разнообразием, присутствующим в природном генетическом алфавите (четыре природных нуклеотида А, С, G и Т в ДНК и четыре природных нуклеотида А, С, G и U в РНК). В настоящем документе раскрыт дополнительный способ получения нуклеиновых кислот, который содержит расширенный генетический алфавит.

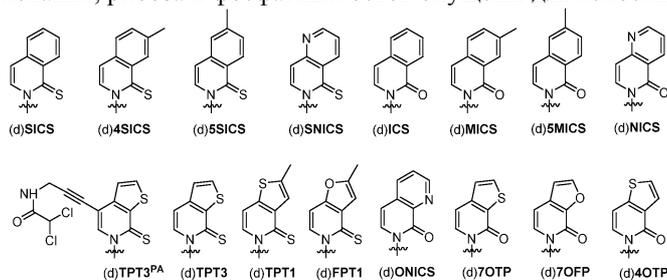
**Краткое раскрытие настоящего изобретения**

В настоящем документе согласно некоторым вариантам осуществления раскрыты способы получения белка, содержащего неприродную аминокислоту, причем способ предусматривает: получение мутантной tRNA, причем мутантная tRNA содержит последовательность мутантных антикодонов, выбранных из табл. 1 или 2; получение мутантной mRNA, причем мутантная mRNA содержит последовательность мутантных кодонов, выбранных из табл. 1 или 2; и синтез белка, содержащего неприродную аминокислоту, с использованием мутантной tRNA и мутантной mRNA. В некоторых случаях белок синтезируется в бесклеточной системе трансляции. В некоторых случаях белок синтезируется в клетке (полусинтетическом организме или SSO). В некоторых случаях полусинтетический организм включает в себя бактерию. В некоторых случаях полусинтетический организм включает в себя *Escherichia coli*. В некоторых случаях мутантный антикодон мутантной tRNA образует пару с мутантным кодоном, выбранным из табл. 1-3. В некоторых случаях неприродная аминокислота содержит по меньшей мере один неприродный нуклеотид. В некоторых случаях неприродный нуклеотид содержит неприродное нуклеотидное основание. В некоторых случаях неприродное основание неприродного нуклеотида выбирают из группы, состоящей из 2-аминоаденин-9-ила, 2-аминоаденина, 2-F-аденина, 2-тиоурацила, 2-тиотимина, 2-тиоцитозина, 2-пропила и алкильных производных аденина и гуанина, 2-аминоаденина, 2-аминопропиладенина, 2-аминопиридина, 2-пиридона, 2'-дезоксинуридина, 2-амино-2'-дезоксаденозина, 3-дезагуанина, 3-дезааденина, 4-тиоурацила, 4-тиотимина, урацил-5-ила, гипоксантин-9-ила (I), 5-метилцитозина, 5-гидроксиметилцитозина, ксантина, гипоксантина, 5-бром- и 5-трифторметилурацилов и цитозинов; 5-галоурацила, 5-галоцитозина, 5-пропилиурацила, 5-пропилцитозина, 5-урацила, 5-замещенных, 5-гало, 5-замещенных пиримидинов, 5-гидроксицитозина, 5-бромцитозина, 5-бромурацила, 5-хлорцитозина, хлорированного цитозина, циклоцитозина, цитозинарабинозида, 5-фторцитозина, фторпиримидина, фторурацила, 5,6-дигидроцитозина, 5-йодцитозина, гидроксимочевины, йодурацила, 5-нитроцитозина, 5-бромурацила, 5-хлорурацила, 5-фторурацил и 5-йодурацила, 6-алкильных производных аденина и гуанина, 6-азапиримидинов, 6-азоурацила, 6-азоцитозина, азацитозина, 6-азотимина, 6-тиогуанина, 7-метилгуанина, 7-метиладенина, 7-дезагуанина, 7-дезагуанозина, 7-дезааденина, 7-деза-8-азагуанина, 8-азагуанина, 8-азааденина, 8-галоген, 8-амино, 8-тиол, 8-тиоалкил и 8-гидроксилзамещенных аденинов и гуанинов; N-4-этилцитозина, N-2-замещенных пуринов, N-6-замещенных пуринов, O-6-замещенных пуринов, тех, которые увеличивают стабильность образования дуплекса, универсальных нуклеиновых кислот, гидрофобных нуклеиновых кислот, смешанных нуклеиновых кислот, нуклеиновых кислот с увеличенным размером, фторированных нуклеиновых кислот, трициклических пиримидинов, феноксазинцитидин([5,4-b][1,4]бензоксазин-2(3H)-она), фенотиазинцитидин(1H-пиримидо[5,4-b][1,4]бензотиазин-2(3H)-она), G-образных зажимов, феноксазинцитидин(9-(2-аминоэтокси)-N-пиримидо[5,4-b][1,4-бензоксазин-2(3H)-она), карбазолцитидин(2H-пиримидо[4,5-b]индол-2-она), пиридоиндолцитидин(N-пиридо[3',2':4,5]пирроло[2,3-d]пиримидин-2-она), 5-фторурацила, 5-

бром урацила, 5-хлорурацила, 5-йодурацила, гипоксантина, ксантина, 4-ацетилцитозина, 5-(карбоксигидроксилометил)урацила, 5-карбоксиметиламинометил-2-тиоуридина, 5-карбоксиметиламинометилурацила, дигидроурацила, бета-D-галактозилквеуозина, инозина, N6-изопентениладенина, 1-метилгуанина, 1-метилюридина, 2,2-диметилгуанина, 2-метиладенина, 2-метилгуанина, 3-метилцитозина, 5-метилцитозина, N6-аденина, 7-метилгуанина, 5-метиламинометилурацила, 5-метоксиаминометил-2-тиоурацила, бета-D-маннозилквеуозина, 5'-метоксикарбоксиметилметоксиурацила, 2-метилтио-N6-изопентениладенина, урацил-5-оксиуксусной кислоты, вибутоксозина, псевдоурацила, квеуозина, 2-тиоцитозина, 5-метил-2-тиоурацила, 2-тиоурацила, 4-тиоурацила, 5-метилурацила, метилового эфира урацил-5-оксиуксусной кислоты, урацил-5-оксиуксусной кислоты, 5-метил-2-тиоурацила, 3-(3-амино-3-N-2-карбоксипропил)урацила, (аср3)w и 2,6-диаминопурина и тех, в которых пуриновое или пиримидиновое основание заменено гетероциклом. В некоторых случаях неприродный нуклеотид выбран из группы, состоящей из (показана только часть нуклеотидного основания, рибоза и фосфатный остов опущены для ясности)



В некоторых случаях неприродный нуклеотид выбран из группы, состоящей из (показана только часть нуклеотидного основания, рибоза и фосфатный остов опущены для ясности)



В некоторых случаях неприродный нуклеотид дополнительно содержит неприродный фрагмент сахара. В некоторых случаях неприродный фрагмент сахара неприродного нуклеотида выбирают из группы, состоящей из модификации в положении 2': OH; замещенный низший алкил, алкарил, аралкил, O-алкарил или O-аралкил, SH, SCH<sub>3</sub>, OCN, Cl, Br, CN, CF<sub>3</sub>, OCF<sub>3</sub>, SOCH<sub>3</sub>, SO<sub>2</sub>CH<sub>3</sub>, ONO<sub>2</sub>, NO<sub>2</sub>, N<sub>3</sub>, NH<sub>2</sub>F; 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'-OCH<sub>3</sub>, 2'-O(CH<sub>2</sub>)<sub>2</sub>OCH<sub>3</sub>, где алкил, алкенил и алкинил могут быть замещенным или незамещенным C<sub>1</sub>-C<sub>10</sub>-алкилом, C<sub>2</sub>-C<sub>10</sub>-алкенилом, C<sub>2</sub>-C<sub>10</sub>-алкинилом, -O[(CH<sub>2</sub>)<sub>n</sub>O]<sub>m</sub>CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>OCH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>NH<sub>2</sub>, -O(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>-ONH<sub>2</sub> и -O(CH<sub>2</sub>)<sub>n</sub>ON[(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>]<sub>2</sub>, где n и m равны от 1 до приблизительно 10; и/или модификации в положении 5': 5'-винил, 5'-метил (R или S), модификации в положении 4', 4'-S, гетероциклоалкила, гетероциклоалкиларила, аминоалкиламино, полиалкиламино, замещенного силила, группы расщепления РНК, репортерной группы, интеркалятора, группы для улучшения фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида и любой их комбинации. В некоторых случаях мутантный антикодон или мутантный кодон дополнительно содержит неприродный остов. В некоторых случаях мутантный антикодон и мутантный кодон дополнительно содержат неприродный остов. В некоторых случаях неприродные нуклеотиды распознаются ДНК-полимеразой, РНК-полимеразой или обратной транскриптазой. В некоторых случаях неприродный нуклеотид включается РНК-полимеразой в mRNA во время транскрипции для производства мутантной mRNA, содержащей мутантный кодон. В некоторых случаях неприродный нуклеотид включается РНК-полимеразой в tRNA во время транскрипции для производства мутантной tRNA, содержащей мутантный антикодон. В некоторых случаях неприродный нуклеотид включается РНК-полимеразой в mRNA во время транскрипции для производства мутантной mRNA. В некоторых случаях неприродный нуклеотид включается РНК-полимеразой в tRNA во время транскрипции для производства мутантной tRNA. В некоторых случаях мутантная tRNA заряжена неприродным аминокислотным остатком. В некоторых случаях белок, содержащий неприродную аминокислоту, производится во время трансляции с использованием мутантной tRNA и мутантной mRNA.

#### Краткое описание графических материалов

Различные аспекты настоящего раскрытия изложены с особым вниманием в прилагаемой формуле изобретения. Лучшее понимание особенностей и преимуществ настоящего раскрытия будет получено со ссылкой на следующее подробное описание, в котором приведены иллюстративные варианты осуществ-

ления, в которых используются принципы настоящего раскрытия, и прилагаемые графические материалы:

На фиг. 1A показана химическая структура UBP dNaM-dTPT3 и природной пары оснований dA-dT.

На фиг. 1B показана генная кассета, используемая для экспрессии sfGFP(AXC)<sup>151</sup> и tRNA(GYT)<sup>Ser</sup>. P<sub>T7</sub> и T<sub>T7</sub> обозначают промотор и терминатор RNAP T7 соответственно. В контролях, где sfGFP экспрессируется в отсутствие serT, последовательность, следующая за терминатором sfGFP T7, отсутствует.

На фиг. 1C показан график флуоресценции клеток, экспрессирующих sfGFP и tRNA<sup>Ser</sup> с указанными кодоном и антикодоном положения 151 соответственно. Знак минус обозначает отсутствие serT в кассете экспрессии, t=0 соответствует добавлению IPTG для индукции экспрессии RNAP T7 и tRNA<sup>Ser</sup> (если присутствует); aTc добавляли в t=0,5 ч для индукции экспрессии sfGFP. AGT, природный кодон Ser; TAG, амбер стоп-кодон; CTA, антикодон амбер-супрессора. Данные представлены как среднее ± стандартное отклонение, n=4 культуры, каждая размножена из отдельной колонии.

На фиг. 1D показан график роста клеток, экспрессирующих sfGFP и tRNA<sup>Ser</sup> с указанными кодоном и антикодоном положения 151 соответственно. Знак минус обозначает отсутствие serT в кассете экспрессии, t=0 соответствует добавлению IPTG для индукции экспрессии RNAP T7 и tRNA<sup>Ser</sup> (если присутствует); aTc добавляли в t=0,5 ч для индукции экспрессии sfGFP. AGT, природный кодон Ser; TAG, амбер стоп-кодон; CTA, антикодон амбер-супрессора. Данные представлены как среднее ± стандартное отклонение, n=4 культуры, каждая размножена из отдельной колонии.

На фиг. 1E показан вестерн-блот лизатов (нормированный по OD<sub>600</sub>) из клеток, собранных в последний момент времени, показанный на фиг. 1C и 1D, зондированных антителом α-GFP (N-концевой эпитоп).

На фиг. 1F показан график относительного содержания аминокислот (обозначенных их однобуквенными кодами в условных обозначениях на фигуре), обнаруженных в положении 151 sfGFP, очищенного из клеток, экспрессирующих sfGFP(AGT)<sup>151</sup> или sfGFP(AXC)<sup>151</sup> и tRNA<sup>Ser</sup>(GYT), как определено с помощью LC-MS/MS и количественного определения интенсивности ионов-предшественников (аминокислоты, обнаруженные при <0,1% (в среднем, для обоих кодонов), не показаны; подробности см. в разделе Способы и в табл. 4 полного перечня обнаруженных аминокислот). Данные представлены в виде среднего значения для отдельных точек данных, n=4 очищенных образца sfGFP, каждый из культуры, размноженной из отдельной колонии и собранной в последний момент времени, показанный на фиг. 1C и фиг. 1D.

На фиг. 2A показан график флуоресценции клеток, экспрессирующих sfGFP с указанным кодоном положения 151, в присутствии (+) или в отсутствие (-) tRNA<sup>Pyl</sup> с родственным антикодоном, PylRS, или 20 mM PrK (N<sup>6</sup>-[(2-пропинилокси)карбонил]-L-лизин) в среде. Флуоресценцию определяли в последний момент времени, показанный на фиг. 2B. Звездочкой обозначено отсутствие tRNA<sup>Pyl</sup> в клетках, экспрессирующих sfGFP(TAC)<sup>151</sup>. TAC, природный кодон Tug; TAG, амбер стоп-кодон; n.d., не определено. Данные показаны как средние с отдельными точками данных, каждая из которых размножена из отдельной колонии.

На фиг. 2B показан анализ времени подмножества данных, показанных на фиг. 2A. Знаки плюс и минус обозначают наличие или отсутствие, соответственно, 20 mM PrK в средах, t=0 соответствует добавлению IPTG для индукции экспрессии PylRS, RNAP T7 и tRNA<sup>Pyl</sup>; aTc добавляли в t=1 ч для индукции экспрессии sfGFP. Данные представлены как среднее ± стандартное отклонение, n=4 культуры, каждая размножена из отдельной колонии.

На фиг. 2C показаны вестерн-блоты sfGFP, очищенные из клеток, экспрессирующих sfGFP и tRNA<sup>Pyl</sup> с указанными кодоном и антикодоном положения 151, соответственно, с клик-конъюгацией TAMRA или без нее и/или добавлением 20 mM PrK в среду. tRNA<sup>Pyl</sup> отсутствует (-) в клетках, экспрессирующих sfGFP(TAC)<sup>151</sup>. sfGFP очищали от культур, собранных в последний момент времени, показанный на фиг. 2B. Вестерн-блоты зондировали антителом α-GFP и визуализировали для обнаружения как sfGFP, так и конъюгированным TAMRA.

На фиг. 2D показан график относительного содержания аминокислот (обозначенных их однобуквенными кодами в условных обозначениях фигуры) в положении 151 sfGFP, очищенного из клеток (собранных в последний момент времени, показанный на фиг. 2B), экспрессирующих sfGFP(TAC)<sup>151</sup> или sfGFP и tRNA<sup>Pyl</sup> с указанным кодоном и родственным антикодоном положения 151 соответственно, как определено с помощью LC-MS/MS и количественного определения на основе интенсивности ионов-предшественников (аминокислоты, обнаруженные при < 0,1% (в среднем для всех кодонов) не показаны; смотрите в разделе Способы для подробностей и в табл. 4 приведен полный перечень обнаруженных аминокислот). Данные представлены в виде среднего значения с отдельными точками данных, n=4 очищенных образца sfGFP, каждый из культуры, размноженной из отдельной колонии.

На фиг. 3A показан график флуоресценции клеток, экспрессирующих sfGFP(TAC)<sup>151</sup> или sfGFP и tRNA<sup>pAzF</sup> с указанным кодоном и родственным антикодоном положения 151 соответственно, в присутствии (+) или в отсутствие (-) 5 mM pAzF в среде, t=0 соответствует добавлению IPTG для индукции экспрессии pAzFRS, RNAP T7 и tRNA<sup>pAzF</sup>; aTc добавляли в t=0,5 ч для индукции экспрессии sfGFP. TAC,

природный кодон Тут; TAG, амбер стоп-кодон. Данные представлены как среднее  $\pm$  стандартное отклонение,  $n=4$  культуры, каждая размножена из отдельной колонии. Флуоресценция, наблюдаемая с sfGFP(AXC)<sup>151</sup> в отсутствие pAzF, объясняется зарядом tRNA<sup>pAzF</sup>(GYT) с природной аминокислотой (вероятно, Тут).

На фиг. 3В показан вестерн-блот sfGFP, очищенный из клеток, экспрессирующих sfGFP и tRNA<sup>pAzF</sup> с указанным кодоном и антикодоном положения 151, соответственно, с клик-конъюгацией TAMRA или без нее и/или добавлением 5 мМ pAzF в среду. Там, где указано, знак минус означает отсутствие tRNA<sup>pAzF</sup> в клетках, экспрессирующих sfGFP(TAC)<sup>151</sup>. sfGFP очищали от культур, собранных в последний момент времени, показанный на фиг. 3А. Вестерн-блоты зондировали антителом  $\alpha$ -GFP и визуализировали для обнаружения как sfGFP, так и конъюгированного TAMRA.

На фиг. 4 показана флуоресценция клеток, экспрессирующих sfGFP с различными кодонами в положении 151. Клетки, несущие плазмиду sfGFP с указанным кодоном положения 151, выращивали до OD<sub>600</sub>  $\sim$  0,5 и индуцировали с помощью IPTG и аТс. Измерения флуоресценции проводили через 3 ч после индукции. Данные показаны как средние с отдельными точками данных,  $n=3$  культуры, отделенные от одной колонии и выращенные параллельно.

На фиг. 5А показано декодирование кодона АХС с природными почти родственными антикодонами с графиком флуоресценции клеток, экспрессирующих sfGFP(AXC)<sup>151</sup>, с tRNA<sup>Ser</sup> или без него с указанным антикодоном. Клетки индуцировали, как описано на фиг. 1С и фиг. 1D, и измерения флуоресценции соответствуют последней временной точке, показанной на фиг. 1С. Значения для антикодона GYT и в отсутствие tRNA<sup>Ser</sup> (-tRNA) соответствуют тем же значениям на фиг. 1с, d. Данные представлены как среднее  $\pm$  стандартное отклонение,  $n=4$  культуры, каждая размножена из отдельной колонии.

На фиг. 5В показано декодирование кодона АХС с природными почти родственными антикодонами с графиком роста клеток, экспрессирующих sfGFP(AXC)<sup>151</sup> с tRNA<sup>Ser</sup> или без него с указанным антикодоном. Клетки индуцировали, как описано на фиг. 1С и фиг. 1D, и измерения флуоресценции соответствуют последней временной точке, показанной на фиг. 1С. Значения для антикодона GYT и в отсутствие tRNA<sup>Ser</sup> (-tRNA) соответствуют одинаковым значениям на фиг. 1С и 1D. Данные представлены как среднее  $\pm$  стандартное отклонение,  $n=4$  культуры, каждая размножена из отдельной колонии.

На фиг. 6А показаны вестерн-блоты и рост клеток, декодирующих кодоны АХС и GХС с tRNA<sup>Pyl</sup>. Вестерн-блот лизатов (нормированных по OD<sub>600</sub>) из клеток, экспрессирующих sfGFP с указанным кодоном положения 151, в присутствии (+) или в отсутствие (-) tRNA<sup>Pyl</sup> с родственными антикодонами, PyIRS или 20 мМ PrK в среде. Блоты зондировали антителом  $\alpha$ -GFP (N-концевой эпитоп). Клетки индуцировали и собирали в эквивалентный момент времени, как описано на фиг. 2В.

На фиг. 6В показан рост культур, проанализированных на фиг. 6А. Кратное изменение OD<sub>600</sub> между индукцией sfGFP (t=1 ч) и конечным моментом времени является наибольшим, когда присутствуют все компоненты, необходимые для аминокислотирования tRNA<sup>Pyl</sup>. Изменения в абсолютном значении OD<sub>600</sub> обусловлены небольшими колебаниями плотности клеток в начале индукции RNAP T7 (и, если присутствует, tRNA<sup>Pyl</sup>) (t=0). Данные представлены как среднее  $\pm$  стандартное отклонение,  $n=4$  культуры, каждая размножена из отдельной колонии.

На фиг. 7А показано декодирование кодонов АХС и GХС с tRNA<sup>Pyl</sup> и ростом клеток в зависимости от добавленных неприродных риботрифосфатов. Флуоресценция очищенного sfGFP (нижняя панель) от клеток, экспрессирующих sfGFP и tRNA<sup>Pyl</sup> с указанным кодоном/антикодоном положения 151, в присутствии (+) или в отсутствие (-) каждого неприродного риботрифосфата в среде и с 20 мМ PrK или без него. Клетки индуцировали, как описано на фиг. 2В, и измерения флуоресценции проводили в конце индукции ( $\sim$  3,5 ч) перед сбором клеток и очисткой белка sfGFP для конъюгации TAMRA с помощью клик-конъюгации и вестерн-блоттинга.

На фиг. 7В показан гель декодирующих кодонов АХС и GХС с tRNA<sup>Pyl</sup> в зависимости от добавленных неприродных риботрифосфатов. Вестерн-блоты зондировали антителом  $\alpha$ -GFP и визуализировали для обнаружения как sfGFP, так и конъюгированного TAMRA; все дорожки соответствуют sfGFP, очищенному из клеток, выращенных с добавлением PrK. Данные показаны как средние с отдельными точками данных,  $n=3$  культуры, каждая размножена из отдельной колонии; n.d., не определено.

На фиг. 7С показаны графики флуоресценции и роста клеток, экспрессирующих sfGFP(TAC)<sup>151</sup>, в присутствии (+) или в отсутствие (-) как неприродных дезоксириботрифосфатов, так и каждого неприродного риботрифосфата. t=0 соответствует добавлению IPTG для индукции экспрессии RNAP T7; аТс добавляли в t=1 ч для индукции экспрессии sfGFP. Данные представлены в виде среднего значения  $\pm$  стандартная ошибка,  $n=3$  культуры, каждая размножена из отдельной колонии. При используемых концентрациях (см. Способы) dNaMTP и dTPT3TP не ингибируют рост клеток, тогда как оба неприродных риботрифосфата, в частности TPT3TP, демонстрируют некоторое ингибирование роста.

На фиг. 7D показан график роста клеток, соответствующих культурам с добавленным PrK (20 мМ), флуоресценция которых показана на фиг. 2В. Клетки, экспрессирующие sfGFP с природными кодонами, выращивали без каких-либо неприродных трифосфатов, тогда как клетки, экспрессирующие sfGFP с неприродными кодонами, выращивали с неприродными как дезокси-, так и с риботрифосфатами. Данные

представлены как среднее  $\pm$  стандартное отклонение,  $n=4$  культуры, каждая размножена из отдельной колонии.

На фиг. 8А показан гель декодирующих кодонов АХС и GХС с tRNA<sup>Pyl</sup> в зависимости от концентрации PrK в среде. Вестерн-блоты sfGFP очищают от клеток, экспрессирующих sfGFP и tRNA<sup>Pyl</sup> с указанным кодоном/антикодоном положения 151, с клик-конъюгацией TAMRA и добавлением PrK к среде в указанных концентрациях. sfGFP индуцировали и очищали от клеток, собранных, как описано на фиг. 2В. Вестерн-блоты зондировали антителом  $\alpha$ -GFP и визуализировали для обнаружения как sfGFP, так и конъюгированного TAMRA.

На фиг. 8В показан график декодирования кодонов АХС и GХС с tRNA<sup>Pyl</sup> в зависимости от концентрации PrK в среде. Флуоресценция клеток (измеренная в последний момент времени, показанный на с), экспрессирующих sfGFP и tRNA<sup>Pyl</sup> с указанным кодоном и антикодоном положения 151, соответственно, в зависимости от концентрации PrK в среде. Значения флуоресценции для 0 и 20 мМ PrK такие же, как значения (-) и (+) PrK, соответственно, показанные на фиг. 2В. Данные представлены как среднее  $\pm$  стандартное отклонение,  $n=4$  культуры, каждая размножена из отдельной колонии.

На фиг. 8С показан анализ флуоресценции во времени. Для ясности показана только одна репрезентативная культура для каждой пары кодон/антикодон и концентрации PrK. Без ограничения теорией авторы настоящего изобретения связывают низкий уровень sfGFP, производимого в отсутствие PrK, с декодированием эндогенными tRNA и потерей удерживания UBP в sfGFP (табл. 5). Однако относительное количество sfGFP, которое содержит PrK (фиг. 8А), и абсолютное количество экспрессированного sfGFP (фиг. 8В и 8С) увеличивалось в зависимости от дозы с увеличением PrK в среде, что в конечном итоге приводило к почти полному включению PrK, предполагая, что эндогенное считывание кодонов АХС и GХС может быть эффективно подавлено при достаточных концентрациях заряженных PrK-tRNA<sup>Pyl</sup> (GYT) или PrK-tRNA<sup>Pyl</sup> (GYC).

На фиг. 8D показан анализ времени роста клеток при различных концентрациях PrK для эксперимента, показанного на фиг. 8С.

На фиг. 9 показан клеточный рост культур, флуоресценция которых показана на фиг. 3А. Данные представлены в виде среднего значения  $\pm$  стандартная ошибка,  $n=4$  культуры, каждая размножена из отдельной колонии.

Таблица 4.| Относительное содержание аминокислот в положении 151 в sfGFP для экспериментов, описанных на фиг. 1F и 2D. sfGFP, очищенный из клеток, экспрессирующих sfGFP, с tRNA или без нее с указанным кодоном и антикодоном положения 151, соответственно, анализировали способом LC-MS/MS. Извлеченные интенсивности ионов MS1 для репортерных пептидов LEYNFNSHNVX<sup>151</sup>ITADK (X = PrK или любая идентифицированная природная аминокислота, кроме K или R) и LEYNFNSHNVX<sup>151</sup> (если X=K или R) выражены в процентах от суммы интенсивностей ионов для всех наблюдаемых репортерных пептидов. Таблица значений соответствует средней относительной численности и 95% CI всех аминокислот, обнаруженных в положении 151 sfGFP,  $n=4$  очищенных образца sfGFP, каждый из культуры, размноженной из отдельной колонии. Значения <0,1% (в среднем для кодонов, указанных на соответствующих фигурах) исключены из данных, представленных на фиг. 1F и 2D.

Таблица 5. Удерживание UBP.

Удерживание UBP в плаزمиде с указанными кодонами положения 151 sfGFP и антикодонами указанных tRNA определяли на момент времени до индукции sfGFP и в конце индукции, как описано в разделе Способы. Указанные значения представляют собой среднее удерживание UBP в течение индукции (рассчитывается по удерживаниям в эти два момента времени)  $\pm$  95% CI,  $n=4$  культуры, каждая размножена из отдельной колонии, за исключением значений, отмеченных звездочкой, для которых  $n=3$ . n/a, не применимо (потому что соответствующая последовательность является природной или отсутствует). Все плазмиды были выделены из культур, выращенных в присутствии 20 мМ PrK или 5 мМ pAzF (за исключением экспериментов по декодированию Ser). SerRS указывает на зарядку эндогенной синтетазой E.coli. Знак минус означает отсутствие PylRS в клетках с tRNA<sup>Pyl</sup> или отсутствие эктопически экспрессируемой tRNA. Удерживания в строках, обозначенных символом §, соответствуют культурам, из которых sfGFP также был очищен и проанализирован с помощью LC-MS/MS и/или вестерн-блоттинга конъюгированного с TAMRA sfGFP (см. фиг. 1F (Ser), фиг. 2D (PrK) и фиг. 3B (pAzF)); строки со звездочкой соответствуют культурам, проанализированным на фиг. 7A-D. Несмотря на тот факт, что все четыре неприродных трифосфата попадают в клетку через один и тот же транспортер и, таким образом, конкурентно ингибируют импорт друг друга, различий в удерживании UBP не наблюдалось при наличии (+) или отсутствии (-) NaMTP и/или TPT3TP в среде. Эти данные, а также потребность в обоих неприродных риботрифосфатах для высоких уровней экспрессии sfGFP с высокоточным включением PrK (фиг. 7A-D) в совокупности демонстрируют, что уровень экспрессии транспортера PtNTT2 в YZ3 импортирует необходимые уровни неприродных трифосфатов, необходимые для поддержания репликации и транскрипции UBP.

Таблица 6.| Выходы белка sfGFP, экспрессированного в экспериментах по включению Ser, Prk и pAzF. Выходы рассчитывали из общего количества очищенного белка и объема культуры, использованного для очистки (см. Способы). Данные представляют собой среднее  $\pm$  стандартное отклонение, ( $n=4$

образца sfGFP, каждый очищенный из культуры, размноженной из отдельной колонии и определенный из тех же культур, проанализированных на фиг. 1F (для SerRS) и фиг. 2D (для PylRS), а также культур, соответствующих образцам (+) pAzF на фиг. 3A (для pAzFRS)). Выходы очищенного sfGFP сравнимы со средней суммарной флуоресценцией (не нормированной к OD<sub>600</sub>) культур, из которых они были очищены. Значения флуоресценции соответствуют моменту времени, в который клетки собирали для очистки sfGFP; см. фиг. 1C (Ser), фиг. 2B (PrK) и фиг. 3A (pAzF).

### **Подробное описание настоящего изобретения**

#### **Определенная терминология**

Если не указано иное, все технические и научные термины, используемые в настоящем документе, характеризуются тем же значением, которое обычно понимается специалистом в области техники, к которой относится заявленный объект изобретения. Следует понимать, что предшествующее общее описание и последующее подробное описание являются только иллюстративными и пояснительными и не ограничивают любой заявленный предмет. В настоящей заявке использование единственного числа включает в себя множественное число, если специально не указано иное. Следует отметить, что, используемые в описании и прилагаемой формуле изобретения формы единственного числа включают в себя множественные ссылки, если контекст явно не предписывает иное. В настоящей заявке использование "или" означает "и/или", если не указано иное. Кроме того, использование термина "включающий в себя", а также других форм, таких как "включает в себя", "включают в себя" и "включенный", не является ограничивающим.

Используемые в настоящем документе диапазоны и количества могут быть выражены как "приблизительно" конкретное значение или диапазон. Приблизительно также включает в себя точное количество. Следовательно, "приблизительно 5 мкл" означает "приблизительно 5 мкл", а также "5 мкл". Как правило, термин "приблизительно" включает в себя количество, которое, как ожидается, будет в пределах ошибки эксперимента.

Заголовки разделов, используемые в настоящем документе, предназначены только для организационных целей и не должны рассматриваться как ограничивающие предмет описания.

#### **Обзор**

Информация о жизни кодируется четырехбуквенным генетическим алфавитом, что стало возможным благодаря избирательному образованию двух пар оснований: (d)G-(d)C и (d)A-dT/U. Третья неприродная пара оснований (UBP), образованная между двумя синтетическими нуклеотидами, расширяет эту систему, увеличивая тем самым потенциал для хранения информации, и имеет глубокие академические и практические последствия. Из большого разнообразия аналогов синтетических нуклеотидов, о которых сообщалось, несколько стабильно образуют пару друг с другом в природном ДНК-дуплексе, но не распознаются полимеразой, что указывает на то, что силы, регулирующие стабильное спаривание в дуплексной ДНК, не такие же, как те, что регулируют опосредованную полимеразой репликацию. В результате использовали различные подходы для разработки реплицируемых UBPs, например UBPs, которые предназначены для взаимодействия посредством структур комплементарных водородных связей (H-связей), которые не используются природными нуклеотидами. Хотя природные пары оснований образуются посредством H-связи, нет оснований предполагать априори, что H-связь является единственной силой, достаточной для того, чтобы лежать в основе хранения (или извлечения) генетической информации. Например, было продемонстрировано, что фрагмент Кленова ДНК-полимеразы I (Kf) E.coli соединяет dA с неприродным нуклеотидом dF, чье дифтортолуоловое нуклеотидное основание является имитатором формы тимина, который не способен к значительной H-связи. Это поддерживает механизм "геометрического отбора" репликации ДНК и предполагает, что силы, отличные от H-связи, также способствуют репликации.

Разработка UBPs, которые реплицируются, транскрибируются и транслируются в белок *in vitro*, дают представление о силах, лежащих в основе хранения и поиска природной информации, а также дают возможность широкого применения в химической и синтетической биологии. Однако конечной целью многих усилий по разработке UBPs является их применение *in vivo* в качестве основы полусинтетического организма (SSO)-организма, который стабильно хранит и получает увеличенную (неприродную или синтетическую, т.е. созданную человеком) информацию. Более того, такая SSO имеет революционное практическое применение, в том числе и для здоровья человека. В частности, SSO революционизирует растущую область белковой терапии. Однако по сравнению с традиционными низкомолекулярными терапевтическими средствами белковые терапевтические средства строго ограничены по своим молекулярным свойствам из-за конечного химического разнообразия, доступного для двадцати природных аминокислот.

Недавно авторы настоящего изобретения сообщили о создании SSO E.coli, которая с помощью транспортера нуклеозидтрифосфата из *Phaeodactylum tricornutum* (PtNTT2) импортирует необходимые неприродные трифосфаты из среды и затем использует их для репликации плазмиды, содержащей dBaM-dTPT3 UBPs. С тех пор авторы настоящего изобретения показали, что ДНК, содержащая UBPs, может транскрибироваться в SSO с помощью РНК-полимеразы T7, и что, когда неприродный нуклеотид включен в кодон mRNA, разные tRNA, заряженные pcAA и содержащие родственные неприродные нуклео-



Таблица 2

GGX	GXG	XGG
GAX	GXA	XGA
GCX	GXC	XGC
GUX	GXU	XGU
CAX	CXA	XCA
CGX	CXG	XCG
CUX	CXU	XCU
CCX	CXC	XCC
AAx	AXA	XAA
AGX	AXG	XAG
ACX	AXC	XAC
AUX	AXU	XAU
UUX	UXU	XUU
UAX	UXA	XUA
UGX	UXG	XUG
UCX	UXC	XUC
GXX	XGX	XXG
CXX	XCX	XXC
AXX	XAX	XXA
UXX	XUX	XXU
XXX		

Таблица 3

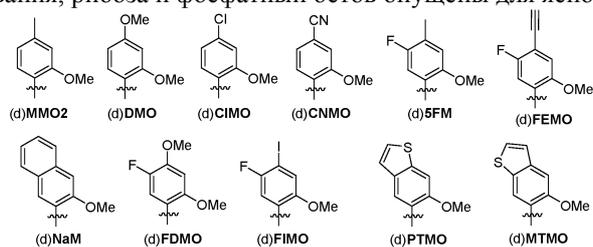
GXY	GYX	XYG
YXG	XGY	YGX
AXY	AYX	XYA
YXA	XAY	YAX
CXY	CYX	XYC
YXC	XCY	YCX
UXY	UYX	XYU
YXU	XUY	YUX
XYY	XXY	YXX
YXX	YXY	XYX

В некоторых случаях мутантный антикодон мутантной tRNA соединяется с мутантным кодоном. Согласно некоторым вариантам осуществления мутантный кодон представляет собой мутантный кодон, представленный в табл. 1. Согласно некоторым вариантам осуществления мутантный кодон представляет собой мутантный кодон, представленный в табл. 2. Согласно некоторым вариантам осуществления мутантный кодон представляет собой мутантный кодон, представленный в табл. 3.

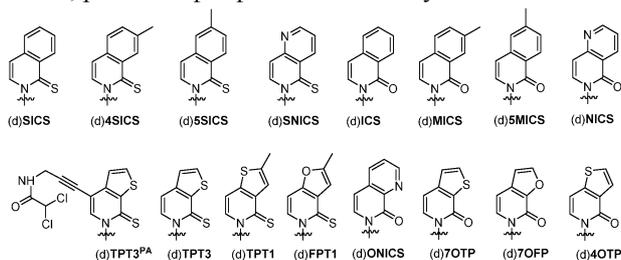
Согласно некоторым вариантам осуществления Y и X, представленные в табл. 1, табл. 2 и табл. 3, представляют собой неприродные основания неприродного нуклеотида. Согласно некоторым вариантам осуществления неприродное основание выбрано из группы, состоящей из 2-аминоаденин-9-ила, 2-аминоаденина, 2-F-аденина, 2-тиоурацила, 2-тиотимина, 2-тиоцитозина, 2-пропила и алкильных производных аденина и гуанина, 2-аминоаденина, 2-аминопропиладенина, 2-аминопиридина, 2-пиридона, 2'-дезоксигуанидина, 2-амино-2'-дезоксиаденозина, 3-дезагуанина, 3-дезааденина, 4-тиоурацила, 4-тиотимина, урацил-5-ила, гипоксантин-9-ила (I), 5-метилцитозина, 5-гидроксиметилцитозина, ксантина, гипоксантина, 5-бром- и 5-трифторметилурацилов и цитозинов; 5-галоурацила, 5-галоцитозина, 5-пропилиурацила, 5-пропинилцитозина, 5-урацила, 5-замещенных, 5-гало, 5-замещенных пиримидинов, 5-гидроксицитозина, 5-бромцитозина, 5-бромуррацила, 5-хлороцитозина, хлорированного цитозина, циклоцитозина, цитозинарабинозида, 5-фторцитозина, фторпиримидина, фторурацила, 5,6-дигидроцитозина, 5-йодцитозина, гидроксимочевины, йодурацила, 5-нитроцитозина, 5-бромуррацила, 5-хлорурацила, 5-фторурацил и 5-йодурацила, 6-алкильных производных аденина и гуанина, 6-азапиримидинов, 6-азоурацила, 6-азоцитозина, азацитозина, 6-азотимина, 6-тиогуанина, 7-метилгуанина, 7-метиладенина, 7-дезагуанина, 7-дезагуанозина, 7-дезааденина, 7-деза-8-азагуанина, 8-азагуанина, 8-азааденина, 8-галоген, 8-амино, 8-тиол, 8-тиоалкил и 8-гидроксил замещенных аденинов и гуанинов; N4-этилцитозина, N-2-замещенных пуринов, N-6-замещенных пуринов, O-6-замещенных пуринов, тех, которые увеличивают стабильность образования дуплекса, универсальных нуклеиновых кислот, гидрофобных нуклеиновых кислот, смешанных нуклеиновых кислот, нуклеиновых кислот с увеличенным размером, фторированных нуклеиновых кислот, трициклических пиримидинов, феноксазинцитидин([5,4-

b][1,4]бензоксазин-2(3H)-она), фенотиазинцитидин(1H-пиримидо[5,4-b][1,4]бензотиазин-2(3H)-она), G-образных зажимов, феноксазинцитидин(9-(2-аминоэтокси)-H-пиримидо[5,4-b][1,4-бензоксазин-2(3H)-она), карбазолцитидин(2H-пиримидо[4,5-b]индол-2-она), пиридоиндолцитидин(H-пиридо[3',2':4,5]пироло[2,3-d]пиримидин-2-она), 5-фторурацила, 5-бром урацила, 5-хлорурацила, 5-йодурацила, гипоксантина, ксантина, 4-ацетилцитозина, 5-(карбоксигидроксиметил)урацила, 5-карбоксиметиламинометил-2-тиоуридина, 5-карбоксиметиламинометилурацила, дигидроурацила, бета-D-галактозилквеузона, инозина, N6-изопентениладенина, 1-метилгуанина, 1-метилюрина, 2,2-диметилгуанина, 2-метиладенина, 2-метилгуанина, 3-метилцитозина, 5-метилцитозина, N6-аденина, 7-метилгуанина, 5-метиламинометилурацила, 5-метоксиаминометил-2-тиоурацила, бета-D-маннозилквеузона, 5'-метоксикарбоксиметилметоксиурацила, 2-метилтио-N6-изопентениладенина, урацил-5-оксиуксусной кислоты, вибутоксозина, псевдоурацила, квеузона, 2-тиоцитозина, 5-метил-2-тиоурацила, 2-тиоурацила, 4-тиоурацила, 5-метилурацила, метилового эфира урацил-5-оксиуксусной кислоты, урацил-5-оксиуксусной кислоты, 5-метил-2-тиоурацила, 3-(3-амино-3-N-2-карбоксыпропил)урацила, (аср3)w и 2,6-диаминопурина и тех, в которых пуриновое или пиримидиновое основание заменено гетероциклом.

В некоторых случаях неприродный нуклеотид выбран из группы, состоящей из (показана только часть нуклеотидного основания, рибоза и фосфатный остов опущены для ясности)



В некоторых случаях неприродный нуклеотид выбран из группы, состоящей из (показана только часть нуклеотидного основания, рибоза и фосфатный остов опущены для ясности)



В некоторых случаях неприродный нуклеотид дополнительно содержит неприродный сахарный фрагмент. В некоторых случаях неприродный сахарный фрагмент неприродного нуклеотида выбирают из группы, состоящей из модификации в положении 2': OH; замещенный низший алкил, алкарил, аралкил, O-алкарил или O-аралкил, SH, SCH<sub>3</sub>, OCN, Cl, Br, CN, CF<sub>3</sub>, OCF<sub>3</sub>, SOCH<sub>3</sub>, SO<sub>2</sub>CH<sub>3</sub>, ONO<sub>2</sub>, NO<sub>2</sub>, N<sub>3</sub>, NH<sub>2</sub>F; 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'-OCH<sub>3</sub>, 2'-O(CH<sub>2</sub>)<sub>2</sub>OCH<sub>3</sub>, где алкил, алкенил и алкинил могут быть замещенным или незамещенным C<sub>1</sub>-C<sub>10</sub>-алкилом, C<sub>2</sub>-C<sub>10</sub>-алкенилом, C<sub>2</sub>-C<sub>10</sub>-алкинилом, -O[(CH<sub>2</sub>)<sub>n</sub>O]<sub>m</sub>CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>OCH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>NH<sub>2</sub>, -O(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>ONH<sub>2</sub> и -O(CH<sub>2</sub>)<sub>n</sub>ON[(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>]<sub>2</sub>, где n и m равны от 1 до приблизительно 10; и/или модификации в положении 5': 5'-винил, 5'-метил (R или S), модификации в положении 4', 4'-S, гетероциклоалкила, гетероциклоалкиларила, аминоалкиламино, полиалкиламино, замещенного силила, группы расщепления РНК, репортерной группы, интеркалятора, группы для улучшения фармакокинетических свойств олигонуклеотида или группы для улучшения фармакодинамических свойств олигонуклеотида и любой их комбинации.

В некоторых случаях мутантный антикодон или мутантный кодон дополнительно содержит неприродный остов. В некоторых случаях мутантный антикодон дополнительно содержит неприродный остов. В некоторых случаях мутантный кодон дополнительно содержит неприродный остов. В некоторых случаях, неприродный остов выбран из группы, состоящей из фосфоротиоата, хирального фосфоротиоата, фосфородитиоата, фосфотриэфира, аминоалкилфосфотриэфира, C<sub>1</sub>-C<sub>10</sub>-фосфонатов, 3'-алкиленфосфоната, хиральных фосфонатов, фосфинатов, фосфорамидатов, 3'-аминофосфорамидата, аминоалкилфосфорамидатов, тионофосфорамидатов, тионоалкилфосфонатов, тионоалкилфосфотриэфиров и боранофосфатов.

В некоторых случаях неприродные нуклеотиды распознаются с помощью полимеразы. В некоторых случаях полимеразы представляет собой ДНК-полимеразу, РНК-полимеразу или обратную транскриптазу. В некоторых случаях полимеразы содержит Ф29, В103, GA-1, PZA, Ф15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722, L17, ThermoSequenase®, 9°Nm™, ДНК-полимеразу Terminator™, Tne, Tma, Tfl, Tth, Tti, фрагмент Стоффеля, ДНК-полимеразу Vent™ и Deep Vent™, ДНК-полимеразу KOD, Tgo, JDF-3, Pfu, Taq, ДНК-полимеразу T7, РНК-полимеразу T7, PGB-D, ДНК-

полимеразу UTMa, ДНК-полимеразу I E.coli, ДНК-полимеразу III E.coli, археальную ДНК-полимеразу DP11/DP2 II, ДНК-полимеразу 9°N, ДНК-полимеразу Taq, ДНК-полимеразу Phusion®, ДНК-полимеразу Pfu, РНК-полимеразу SP6, ДНК-полимеразу RB69, обратную транскриптазу вируса птичьего миелобластога (AMV), обратную транскриптазу вируса мышинного лейкоза (MMLV), обратную транскриптазу SuperScript® II и обратную транскриптазу SuperScript® III.

В некоторых случаях полимеразы представляет собой ДНК-полимеразу I-фрагмент Кленова, полимеразу Vent, ДНК-полимеразу Phusion®, ДНК-полимеразу KOD, полимеразу Taq, ДНК-полимеразу T7, РНК-полимеразу T7, ДНК-полимеразу Terminator™, полимеразу POLB, РНК-полимеразу SP6, ДНК-полимеразу I E.coli, ДНК-полимеразу III E.coli, обратную транскриптазу вируса птичьего миелобластога (AMV), обратную транскриптазу вируса мышинного лейкоза (MMLV), обратную транскриптазу SuperScript® II или обратную транскриптазу SuperScript® III.

В некоторых случаях неприродный нуклеотид включается полимеразой в mRNA во время транскрипции, чтобы производить мутантную mRNA, содержащую мутантный кодон. В некоторых случаях неприродный нуклеотид включается полимеразой в mRNA во время транскрипции, чтобы производить мутантную mRNA.

В некоторых случаях неприродный нуклеотид включается полимеразой в tRNA во время транскрипции, чтобы производить мутантную tRNA, содержащую мутантный антикодон. В некоторых случаях неприродный нуклеотид включается полимеразой в tRNA во время транскрипции, чтобы производить мутантную tRNA.

В некоторых случаях мутантная tRNA представляет собой неприродный аминокислотный остаток. В некоторых случаях неприродный аминокислотный остаток представляет собой неприродную аминокислоту, такую как описанная в публикации Liu C.C., Schultz, P.G. Annu. Rev. Biochem. 2010, 79, 413.

В некоторых случаях белок, содержащий неприродную аминокислоту, производится во время трансляции с использованием мутантной tRNA и мутантной mRNA. В некоторых случаях белок, содержащий неприродную аминокислоту, производится в бесклеточной системе трансляции. В некоторых случаях белок синтезируется в клетке или полусинтетическом организме (SSO). В некоторых случаях полусинтетический организм включает в себя микроорганизм. В некоторых случаях полусинтетический организм включает в себя бактерию. В некоторых случаях полусинтетический организм включает в себя Escherichia coli.

### Нуклеиновые кислоты

Нуклеиновая кислота (например, также называемая в настоящем документе целевой нуклеиновой кислотой, целевой нуклеотидной последовательностью, представляющей интерес последовательностью нуклеиновой кислоты или представляющей интерес областью нуклеиновой кислоты) может быть из любого источника или композиции, такой как ДНК, cDNA, gDNA (геномная ДНК), РНК, siRNA (короткая ингибирующая РНК), например RNAi, tRNA или mRNA, и может быть в любой форме (например, линейной, кольцевой, суперспиральной, одноцепочечной, двухцепочечной и т.п.). Нуклеиновые кислоты могут содержать нуклеотиды, нуклеозиды или полинуклеотиды. Нуклеиновые кислоты могут включать в себя природные и неприродные нуклеиновые кислоты. Нуклеиновая кислота также может включать в себя неприродные нуклеиновые кислоты, такие как аналоги ДНК или РНК (например, содержащие аналоги оснований, аналоги сахаров и/или неприродный остов и т.п.). Понятно, что термин "нуклеиновая кислота" не относится к конкретной длине полинуклеотидной цепи или не подразумевает ее, таким образом, полинуклеотиды и олигонуклеотиды также включены в определение. Иллюстративные природные нуклеотиды включают в себя, без ограничения, ATP, UTP, CTP, GTP, ADP, UDP, CDP, GDP, AMP, UMP, CMP, GMP, dATP, dTTP, dCTP, dGTP, dADP, dTDP, dCDP, dGDP, dAMP, dTMP, dCMP и dGMP. Иллюстративные природные дезоксирибонуклеотиды включают в себя dATP, dTTP, dCTP, dGTP, dADP, dTDP, dCDP, dGDP, dAMP, dTMP, dCMP и dGMP. Иллюстративные природные рибонуклеотиды включают в себя ATP, UTP, CTP, GTP, ADP, UDP, CDP, GDP, AMP, UMP, CMP и GMP. Для РНК урациловым основанием является уридин. Нуклеиновая кислота иногда представляет собой вектор, плазмиду, фаг, автономно реплицирующуюся последовательность (ARS), центромеру, искусственную хромосому, дрожжевую искусственную хромосому (например, YAC) или другую нуклеиновую кислоту, способную реплицироваться или реплицироваться. Неприродная нуклеиновая кислота может представлять собой аналог нуклеиновой кислоты.

### Неприродные нуклеиновые кислоты

Нуклеотидный аналог или неприродный нуклеотид включает в себя нуклеотид, который содержит некоторый тип модификации по отношению к основанию, фрагментам сахара или фосфата. Модификация может включать в себя химическую модификацию. Модификации могут быть, например, из группы 3'ОН или 5'О остова, компонента сахара или нуклеотидного основания. Модификации могут включать в себя добавление не встречающихся в природе линкерных молекул и/или межцепочечных или внутрицепочечных поперечных связей. Согласно одному аспекту модифицированная нуклеиновая кислота содержит модификацию одного или нескольких из группы 3'ОН или 5'ОН, остова, компонента сахара или нуклеотидного основания и/или добавление не встречающихся в природе линкерных молекул. Согласно од-

ному аспекту модифицированный остов включает в себя остов, отличающийся от фосфодиэфирного остова. Согласно одному аспекту модифицированный сахар включает в себя сахар, отличный от дезоксирибозы (в модифицированной ДНК) или отличный от рибозы (в модифицированной РНК). Согласно одному аспекту модифицированное основание включает в себя основание, отличное от аденина, гуанина, цитозина или тимина (в модифицированной ДНК) или основание, отличное от аденина, гуанина, цитозина или урацила (в модифицированной РНК).

Нуклеиновая кислота может содержать по меньшей мере одно модифицированное основание. Модификации фрагмента основания будут включать в себя природные и синтетические модификации А, С, G и T/U, а также различные пуриновые или пиримидиновые основания. Согласно некоторым вариантам осуществления модификация относится к модифицированной форме аденина, гуанина, цитозина или тимина (в модифицированной ДНК) или модифицированной форме аденина, гуанина, цитозина или урацила (в модифицированной РНК).

Модифицированное основание не природной нуклеиновой кислоты включает в себя, без ограничения, 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-пропинилурацил и цитозин, 6-азоурацил, цитозин и тимин, 5-урацил (псевдоурацил), 4-тиоурацил, 8-гало, 8-амино, 8-тиол, 8-тиоалкил, 8-гидроксил и другие 8-замещенные аденины и гуанины, 5-галоген, в частности 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-дезааденин и 3-деазагуанин и 3-дезааденин. Некоторые не природные нуклеиновые кислоты, такие как 5-замещенные пиримидины, 6-азапиримидины и N-2-замещенные пурины, N-6-замещенные пурины, O-6-замещенные пурины, 2-аминопропиладенин, 5-пропинилурацил, 5-пропинилцитозин, 5-метилцитозин, те, которые повышают стабильность образования дуплекса, универсальные нуклеиновые кислоты, гидрофобные нуклеиновые кислоты, смешанные нуклеиновые кислоты, нуклеиновые кислоты с увеличенным размером, фторированные нуклеиновые кислоты, 5-замещенные пиримидины, 6-азапиримидины и N-2, N-6 и O-6 замещенные пурины, включая в себя 2-аминопропиладенин, 5-пропинилурацил и 5-пропинилцитозин, 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-пропинил ( $-C\equiv C-CI^{1/4}$ ) урацил, 5-пропинилцитозин, другие алкильные производные пиримидиновых нуклеиновых кислот, 6-азоурацил, 6-азоцитозин, 6-азотимин, 5-урацил (псевдоурацил), 4-тиоурацил, 8-гало, 8-амино, 8-тиол, 8-тиоалкил, 8-гидроксил и другие 8-замещенные аденины и гуанины, 5-галоген, в частности 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, феноксазинцитидин ([5,4-b][1,4]бензоксазин-2(3H)-он), фенотиазинцитидин (1H-пиримидо[5,4-b][1,4]бензотиазин-2(3H)-он), G-образные зажимы, феноксазинцитидин (например, 9-(2-аминоэтокси)-N-пиримидо[5,4-b][1,4]бензоксазин-2(3H)-он), карбазолцитидин (2H-пиримидо[4,5-b]индол-2-1), пиридоиндолцитидин (H-пиридо[3',2':4,5]пирроло[2,3-d]пиримидин-2-он), в котором пуриновое или пиримидиновое основание заменено другими гетероциклами, 7-дезааденин, 7-дезагуанозин, 2-аминопиридин, 2-пиридон, азацитозин, 5-бромцитозин, бромурацил, 5-хлорцитозин, хлорированный цитозин, циклоцитозин, цитозинарабинозид, 5-фторцитозин, фторпиримидин, фторурацил, 5,6-дигидроцитозин, 5-йодцитозин, гидроксимочевина, йодурацил, 5-нитроцитозин, 5-бромурацил, 5-хлорурацил, 5-фторурацил и 5-йодурацил, 2-аминоаденин, 6-тиогуанин, 2-тиотимин, 4-тиотимин, 5-пропинилурацил, 4-тиоурацил, N4-этилцитозин, 7-деазагуанин, 7-деза-8-азагуанин, 5-гидроксицитозин, 2'-деоксиуридин, 2-амино-2'-деоксиаденозин и те, которые описаны в патентах США № 3687808; 4845205; 4910300; 4948882; 5093232; 5130302; 5134066; 5175273; 5367066; 5432272; 5457187; 5459255; 5484908; 5502177; 5525711; 5552540; 5587469; 5594121; 5596091; 5614617; 5645985; 5681941; 5750692; 5763588; 5,830,653 и 6,005,096; WO 99/62923; 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, Y.S., Chapter 15, Antisense Research and Applications, Crooke, S.T. and Lebleu, B., Eds., CRC Press, 1993, 273-288. Дополнительные модификации оснований можно найти, например, в патенте США № 3687808, Englisch et al., Angewandte Chemie, International Edition, 1991, 30, 613 и Sanghvi, Y. S., Chapter 15, Antisense Research and Applications, pages 289-302, Crooke, S. T. and Lebleu, B. ed., CRC Press, 1993.

В настоящей области техники доступны не природные нуклеиновые кислоты, содержащие различные гетероциклические основания и различные фрагменты сахара (и аналоги сахара), и нуклеиновая кислота может включать в себя одно или несколько гетероциклических оснований, отличных от основных пяти компонентов оснований природных нуклеиновых кислот. Например, гетероциклическое основание может включать в себя 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-оксопиррол[2,3-d]пиримидин-5-ил, 2-амино-4-оксо-

пирроло[2,3-d]пиримидин-3-ил, где пурины присоединяются к фрагменту сахара нуклеиновой кислоты через 9-положение, пиримидины через 1-положение, пирролопиримидины через 7-положение и пиразолопиримидины через 1-положение.

Аналоги нуклеотидов также могут быть модифицированы в фосфатном фрагменте. Модифицированные фосфатные фрагменты включают в себя, без ограничения, те, которые могут быть модифицированы таким образом, что связь между двумя нуклеотидами содержит фосфоротиоат, хиральный фосфоротиоат, фосфородитиоат, фосфотриэфир, аминоалкилфосфотриэфир, метил и другие алкилфосфонаты, включая в себя 3'-алкиленфосфонат и хиральные фосфонаты, фосфинаты фосфорамидаты, включая в себя 3'-аминофосфорамидаты и аминоалкилфосфорамидаты, тионофосфорамидаты, тионоалкилфосфонаты, тионоалкилфосфотриэфиры и боранофосфаты. Понятно, что эти фосфатные или модифицированные фосфатные связи между двумя нуклеотидами могут быть через 3'-5' связь или 2'-5' связь, и эта связь может содержать обратную полярность, такую как от 3'-5' к 5'-3' или от 2'-5' к 5'-2'. Различные соли, смешанные соли и формы свободных кислот также включены. Многочисленные патенты США описывают, как производить и использовать нуклеотиды, содержащие модифицированные фосфаты, и включают в себя, без ограничения, 3687808; 4469863; 4476301; 5023243; 5177196; 5188897; 5264423; 5276019; 5278302; 5286717; 5321131; 5399676; 5405939; 5453496; 5455233; 5466677; 5476925; 5519126; 5536821; 5541306; 5550111; 5563253; 5571799; 5587361 и 5625050, каждый из которых включен в настоящий документ посредством ссылки.

Неприродные нуклеиновые кислоты могут включать в себя 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/JP 2000/004720; PCT/JP 2003/002342; PCT/JP 2004/013216; PCT/JP 2005/020435; PCT/JP 2006/315479; PCT/JP 2006/324484; PCT/JP 2009/056718; PCT/JP2 010/067560) или 5'-замещенные мономеры, полученные в виде монофосфата с модифицированными основаниями (Wang et al., *Nucleosides Nucleotides & Nucleic Acids*, 2004, 23 (1 и 2), 317-337).

Неприродные нуклеиновые кислоты могут включать в себя модификации в 5'-положении и 2'-положении сахарного кольца (PCT/US 94/02993), такие как 5'-CH<sub>2</sub>-замещенные 2'-O-защищенные нуклеозиды (Wu et al. al., *Helvetica Chimica Acta*, 2000, 83, 1127-1143 и Wu et al. *Bioconjugate Chem.* 1999, 10, 921-924). Неприродные нуклеиновые кислоты могут включать в себя амидосвязанные нуклеозидные димеры, полученные для включения в олигонуклеотиды, причем 3'-связанный нуклеозид в димере (от 5' к 3') содержит 2'-OCH<sub>3</sub> и 5'-(S)-CH<sub>3</sub> (Mesmaeker et al., *Synlett*, 1997, 1287-1290). Неприродные нуклеиновые кислоты могут включать в себя 2'-замещенные 5'-CH<sub>2</sub> (или O) модифицированные нуклеозиды (PCT/US92/01020). Неприродные нуклеиновые кислоты могут включать в себя 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 и Hutter et al. al, *Helvetica Chimica Acta*, 2002, 85, 2777-2806). Неприродные нуклеиновые кислоты могут включать в себя 5'-фосфонатные мономеры, имеющие 2'-замену (US 2006/0074035), и другие модифицированные 5'-фосфонатные мономеры (WO 97/35869). Неприродные нуклеиновые кислоты могут включать в себя 5'-модифицированные метиленфосфонатные мономеры (EP 614907 и EP 629633). Неприродные нуклеиновые кислоты могут включать в себя аналоги 5'- или 6'-фосфонатных рибонуклеозидов, содержащих гидроксильную группу в положении 5' или 6' (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 и Hampton et al., *J. Med. Chem.*, 1976, 19(8), 1029-1033). Неприродные нуклеиновые кислоты могут включать в себя 5'-фосфонатные дезоксирибонуклеозидные мономеры и димеры, имеющие 5'-фосфатную группу (Nawrot et al., *Oligonucleotides*, 2006, 16(1), 68-82). Природные нуклеиновые кислоты могут включать в себя нуклеозиды, имеющие 6'-фосфонатную группу, в которой положение 5' или/и 6' является незамещенным или замещенным тио-трет-бутильной группой (SC(CH<sub>3</sub>)<sub>3</sub>) (и ее аналогами); метиленаминогруппой (CH<sub>2</sub>NH<sub>2</sub>) (и ее аналогами) или цианогруппой (CN) (и ее аналогами) (Fairhurst et al, *Synlett*, 2001, 4, 467-472; Kappler et al, *J. Med. Chem.*, 1986, 29, 1030-1038 и *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).

Неприродные нуклеиновые кислоты могут также включать в себя модификации фрагмента сахара. Нуклеиновые кислоты по настоящему изобретению могут необязательно содержать один или несколько нуклеозидов, в которых сахарная группа была модифицирована. Такие нуклеозиды с модифицированным сахаром могут придавать повышенную стабильность нуклеазы, повышенную аффинность связывания или некоторые другие полезные биологические свойства. Согласно некоторым вариантам осуществления нуклеиновые кислоты содержат химически модифицированный фрагмент рибофуранозного кольца. Примеры химически модифицированных рибофуранозных колец включают в себя, без ограничения, добавление групп заместителей (включая в себя 5' и/или 2' группы заместителей, мостик двух атомов кольца с образованием бициклических нуклеиновых кислот (BNA), замещение атома кислорода рибозильно-

го кольца на S, N(R) или C(R<sub>1</sub>)(R<sub>2</sub>) (R=H, C<sub>1</sub>-C<sub>12</sub>-алкил или защитная группа) и их комбинации. Примеры химически модифицированных сахаров можно найти в WO 2008/101157, US 2005/0130923 и WO 2007/134181.

Модифицированная нуклеиновая кислота может содержать модифицированные сахара или аналоги сахаров. Таким образом, в дополнение к рибозе и дезоксирибозе фрагмент сахара может представлять собой пентозу, дезоксипентозу, гексозу, дезоксигексозу, глюкозу, арабинозу, ксилозу, ликсозу и циклопентильную группу "аналога" сахара. Сахар может быть в пиранозильной или фуранозильной форме. Фрагмент сахара 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. Модификации сахара включают в себя, без ограничения, 2'-алкокси-РНК-аналоги, 2'-амино-РНК-аналоги, 2'-фтор-ДНК и 2'-алкокси- или amino-РНК/ДНК-химеры. Например, модификация сахара может включать в себя 2'-О-метилуридин и 2'-О-метилцитидин. Модификации сахара включают в себя 2'-О-алкилзамещенные дезоксирибонуклеозиды и 2'-О-этиленгликольподобные рибонуклеозиды. Получение этих сахаров или аналогов сахара и соответствующих "нуклеозидов", в которых такие сахара или аналоги присоединены к гетероциклическому основанию (основанию нуклеиновой кислоты), известно. Модификации сахара также могут быть сделаны и объединены с другими модификациями.

Модификации фрагмента сахара включают в себя природные модификации рибозы и дезоксирибозы, а также неприродные модификации. Модификации сахара включают в себя, без ограничения, следующие модификации в положении 2': OH; F; O-, S- или N-алкил; O-, S- или N-алкенил; O-, S- или N-алкинил или O-алкил-O-алкил, причем алкил, алкенил и алкинил могут представлять собой замещенный или незамещенный C<sub>1</sub>-C<sub>10</sub> алкил или C<sub>2</sub>-C<sub>10</sub>-алкенил и алкинил. 2' Модификации сахара также включают в себя, без ограничения, -O[(CH<sub>2</sub>)<sub>n</sub>O]<sub>m</sub>CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>OCH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>NH<sub>2</sub>, -O(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>, -O(CH<sub>2</sub>)<sub>n</sub>-ONH<sub>2</sub> и -O(CH<sub>2</sub>)<sub>n</sub>ON[(CH<sub>2</sub>)<sub>n</sub>CH<sub>3</sub>]<sub>2</sub>, где n и m равны от 1 до приблизительно 10.

Другие модификации в положении 2' включают в себя, без ограничения: C<sub>1</sub>-C<sub>10</sub> низший алкил, замещенный низший алкил, алкарил, аралкил, O-алкарил или O-аралкил, SH, SCH<sub>3</sub>, OCN, Cl, Br, CN, CF<sub>3</sub>, OCF<sub>3</sub>, SOCH<sub>3</sub>, SO<sub>2</sub> CH<sub>3</sub>, ONO<sub>2</sub>, NO<sub>2</sub>, N<sub>3</sub>, NH<sub>2</sub>, гетероциклоалкил, гетероциклоалкарил, аминоалкиламино, полиалкиламино, замещенный силлил, группу расщепления РНК, репортерную группу, интеркалятор, группу для улучшения фармакокинетических свойств олигонуклеотида или группу для улучшения фармакодинамических свойств олигонуклеотида и других заместителей, имеющих сходные свойства. Подобные модификации могут быть также выполнены в других положениях на сахаре, особенно в положении 3' сахара на 3'-концевом нуклеотиде или в 2'-5' связанных олигонуклеотидах и в положении 5'-концевого нуклеотида. Модифицированные сахара также включают в себя те, которые содержат модификации в кислороде мостикового кольца, такие как CH<sub>2</sub> и S. Аналоги нуклеотидного сахара также могут иметь миметики сахара, такие как циклобутильные фрагменты, вместо сахара пентофуранозила. Существует множество патентов США, в которых описывается получение таких модифицированных структур сахара, как, например, 4981957; 5118800; 5319080; 5359044; 5393878; 5446137; 5466786; 5514785; 5519134; 5567811; 5576427; 5591722; 5597909; 5610300; 5627053; 5639873; 5646265; 5658873; 5670633; 4845205; 5130302; 5134066; 5175273; 5367066; 5432272; 5457187; 5459255; 5484908; 5502177; 5525,711; 5552540; 5587469; 5594121; 5596091; 5614617; 5681941 и 5700920, каждый из которых полностью включен в настоящий документ посредством ссылки, который детализирует и описывает ряд модификаций оснований. Каждый из этих патентов включен в настоящий документ посредством ссылки.

Примеры нуклеиновых кислот, содержащих модифицированные фрагменты сахара, включают в себя, без ограничения, нуклеиновые кислоты, содержащие 5'-винил, 5'-метил (R или S), 4'-S, 2'-F, 2'-OCH<sub>3</sub> и 2'-O(CH<sub>2</sub>)<sub>2</sub>OCH<sub>3</sub> замещающие группы. Заместитель в положении 2' также может быть выбран из аллила, amino, азида, тии, O-аллила, O-C C<sub>10</sub>алкила, OCF<sub>3</sub>, O(CH<sub>2</sub>)<sub>2</sub>SCH<sub>3</sub>, O(CH<sub>2</sub>)<sub>2</sub>-O-N(R<sub>m</sub>)(R<sub>n</sub>) и O-CH<sub>2</sub>-C(=O)-N(R<sub>m</sub>)(R<sub>n</sub>), где каждый R<sub>m</sub> и R<sub>n</sub> независимо представляет собой H или замещенный или незамещенный C<sub>1</sub>-C<sub>10</sub>алкил.

Согласно некоторым вариантам осуществления нуклеиновые кислоты по настоящему изобретению включают в себя одну или несколько бициклических нуклеиновых кислот. Согласно некоторым таким вариантам осуществления бициклическая нуклеиновая кислота содержит мостик между 4'- и 2'-атомами рибозильного кольца. Согласно определенным вариантам осуществления нуклеиновые кислоты, представленные в настоящем документе, включают в себя одну или несколько бициклических нуклеиновых кислот, причем мостик содержит 4'-2'-бициклическую нуклеиновую кислоту. Примеры таких 4'-2'-бициклических нуклеиновых кислот включают в себя, без ограничения, одну из формул: 4'-(CH<sub>2</sub>)-O-2' (LNA); 4'-(CH<sub>2</sub>)-S-2'; 4'-(CH<sub>2</sub>)<sub>2</sub>-O-2' (ENA); 4'-CH(CH<sub>3</sub>)-O-2' и 4'-CH(CH<sub>2</sub>OCH<sub>3</sub>)-O-2' и их аналоги (см. патент США № 7399845, выданный 15 июля 2008 г.); 4'-C(CH<sub>3</sub>)(CH<sub>3</sub>)-O-2' и его аналоги (см. WO 2009/006478, WO 2008/150729, US 2004/0171570, патент США 7427672, (Chattopadhyaya et al., J. Org. Chem., 2 09, 74, 118-134) и WO 2008/154401, опубликованной 8 декабря 2008 г.). Также смотрите, 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,

129(26) 8362-8379 (Jul. 4, 2007); Elayadi et al., *Curr. Opin. Inven. Drugs*, 2001, 2, 558-561; Braasch et al., *Chem. Biol.*, 2001, 8, 1-7; Oram et al., *Curr. Opin. Mol. Ther.*, 2001, 3, 239-243; патенты США № 7053207, 6268490, 6770748, 6794499, 7034133, 6525191, 6670461 и 7399845; международные заявки WO 2004/106356, WO 1994/14226, WO 2005/021570 и WO 2007/134181; патенты США №№ US 2004/0171570, US 2007/0287831 и US 2008/0039618; патенты США, серийные номера 12/129154, 60/989574, 61/026995, 61/026998, 61/056564, 61/086231, 61/097787 и 61/099844 и международные заявки PCT №№ PCT/US 2008/064591, PCT US 2008/066154 и PCT US 2008/068922, PCT/DK 98/00393 и патенты США №№ 4849513; 5015733; 5118800 и 5118802.

Согласно определенным вариантам осуществления нуклеиновые кислоты могут содержать связанные нуклеиновые кислоты. Нуклеиновые кислоты могут быть связаны вместе, используя любую связь между нуклеиновыми кислотами. Два основных класса связывающих групп между нуклеиновыми кислотами определяются наличием или отсутствием атома фосфора. Типичные фосфорсодержащие связи между нуклеиновыми кислотами включают в себя, без ограничения фосфодиэфиры, фосфотриэфиры, метилфосфонаты, фосфорамидаты и фосфоротиоаты (P=S). Типичные нефосфорсодержащие связывающие группы между нуклеиновыми кислотами включают в себя, без ограничения, метиленимелимино (-CH<sub>2</sub>-N(CH<sub>3</sub>)-O-CH<sub>2</sub>-), тиодифосфат (-O-C(O)-S-), тиокарбамат ((-O-C(O)-S-); силоксан (-O-Si(H)<sub>2</sub>-O-) и N,N\*-диметилгидразин (-CH<sub>2</sub>-N(CH<sub>3</sub>)-N(CH<sub>3</sub>-). Согласно некоторым вариантам осуществления связи между нуклеиновыми кислотами, содержащие хиральный атом, могут быть получены в виде рацемической смеси в виде отдельных энантиомеров, например, алкилфосфонатов и фосфоротиоатов. Неприродные нуклеиновые кислоты могут содержать одну модификацию. Неприродные нуклеиновые кислоты могут содержать несколько модификаций в пределах одной из групп или между различными фрагментами.

Модификации фосфата остова нуклеиновой кислоты включают в себя, без ограничения, метилфосфонат, фосфоротиоат, фосфорамидат (мостиковый или не мостиковый), фосфотриэфир, фосфородитиоат, фосфодитиоат и боранофосфат и могут использоваться в любой комбинации. Другие нефосфатные связи также могут быть использованы.

Согласно некоторым вариантам осуществления модификации остова (например, метилфосфонатные, фосфоротиоатные, фосфоамидатные и фосфородитиоатные межнуклеотидные связи) могут придавать иммуномодулирующую активность модифицированной нуклеиновой кислоте и/или повышать их стабильность *in vivo*.

Фосфорное производное (или модифицированная фосфатная группа) может быть присоединено к фрагменту сахара или аналога сахара и может представлять собой монофосфат, дифосфат, трифосфат, алкилфосфонат, фосфоротиоат, фосфородитиоат, фосфорамидат или т.п. Иллюстративные полинуклеотиды, содержащие модифицированные фосфатные связи или нефосфатные связи, можно найти в 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*, (DJ. Chadwick and G. Cardew, ed.) John Wiley and Sons, New York, NY; (Zon (1993) "Oligonucleoside Phosphorothioates" in *Protocols for Oligonucleotides and Analogs, Synthesis and Properties* (Agrawal, ed.) 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 № 5453496); Micklefield, J. 2001, *Current Medicinal Chemistry* 8: 1157-1179.

Модификация остова может включать в себя замену фосфодиэфирной связи альтернативной группой, такой как анионная, нейтральная или катионная группа. Примеры таких модификаций включают в себя: анионную межнуклеозидную связь; модификацию фосфорамидата с N3' по P5'; боранофосфатную ДНК; проолигонуклеотиды; нейтральные межнуклеозидные связи, такие как метилфосфонаты; ДНК с амидной связью; метиленовые (метилено) связи; формацетальные и тиоформацетальные связи; остовы, содержащие сульфонильные группы; морфолино олигонуклеотиды; пептидные нуклеиновые кислоты (PNA) и положительно заряженные олигонуклеотиды дезоксирибонуклеинового гуанидина (DNG), Micklefield, J. 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 фосфоротиоатных связей.

Заместителями для фосфата могут быть, например, алкильные или циклоалкильные межнуклеозидные связи с короткой цепью, смешанные гетероатомные и алкильные или циклоалкильные межнуклеозидные связи или одна или несколько гетероатомных или гетероциклических межнуклеозидных связей с короткой цепью. К ним относятся те, которые имеют морфолино-связи (образованные частично из части сахара нуклеозида); силоксановые остовы; сульфидные, сульфоксидные и сульфоновые остовы; формацетальный и тиоформацетальный остовы; метиленимелимино и метиленимелимино остовы; алкенилсодержащие остовы; сульфаматные остовы; метиленимелимино и метиленимелимино остовы; сульфонатные и сульфониламидные остовы; амидные остовы и другие, содержащие смешанные N, O, S и CH<sub>2</sub> составные части. Многочисленные патенты Соединенных Штатов раскрывают, как производить и применять эти типы заместителей фосфатов, и включают в себя, без ограничения, 5034506; 5166315; 5185444; 5214134; 5216141; 5235033; 5264562; 5264564; 5405938; 5434257; 5466677; 5470967; 5489677; 5541307;

5561225; 5596086; 5602240; 5610289; 5602240; 5608046; 5610289; 5618704; 5623070; 5663312; 5633360; 5677437 и 5677439, каждый из которых включен в настоящий документ посредством ссылки. Понятно также, что при замене нуклеотида можно заменить как сахарную, так и фосфатную части нуклеотида, например, связью амидного типа (аминоэтилглицин) (PNA). Патенты США 5539082; 5714331 и 5719262 учат, как создавать и применять молекулы PNA, каждая из которых включена в настоящий документ посредством ссылки. (Смотрите также Nielsen et al., Science, 1991, 254, 1497-1500). Конъюгаты могут быть химически связаны с нуклеотидами или нуклеотидными аналогами. Такие конъюгаты включают в себя, без ограничения, липидные фрагменты, такие как фрагмент холестерина (Letsinger et al., Proc. Natl. Acad. Sci. USA, 1989, 86, 6553-6556), желчная кислота (Manoharan et al., Bioorg. Med. Chem. Lett., 1994, 4, 1053-1060), тиоэфир, например, гексил-S-тримитиол (Manoharan et al., Ann. NY. 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), алифатическая цепь, например, додекандиол 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, дигексадецил-рац-глицерин или триэтилammоний-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), полиаминовая или полиэтиленгликолевая цепь (Manoharan et al., Nucleosides & Nucleotides, 1995, 14, 969-973) или адамантан-уксусная кислота (Manoharan et al., Tetrahedron Lett., 1995, 36, 3651-3654), пальмитильный фрагмент (Mishra et al., Biochem. Biophys. Acta, 1995, 1264, 229-237) или октадециламиновый или гексиламино-карбонил-оксихолестериновый фрагмент (Crooke et al., J. Pharmacol. Exp. Ther., 1996, 277, 923-937). Многочисленные патенты США описывают получение таких конъюгатов и включают в себя, без ограничения, патенты США № 4828979; 4948882; 5218105; 5525465; 5541313; 5545730; 5552538; 5578717, 5580731; 5580731; 5591584; 5109124; 5118802; 5138045; 5414077; 5486603; 5512439; 5578718; 5608046; 4587044; 4605735; 4667025; 4762779; 4789737; 4824941; 4835263; 4876335; 4904582; 4958013; 5082830; 5112963; 5214136; 5082830; 5112963; 5214136; 5245022; 5254469; 5258506; 5262536; 5272250; 5292873; 5317098; 5371241 5391723; 5416203 5451463; 5510475; 5512667; 5514785; 5565552; 5567810; 5574142; 5585481; 5587371; 5595726; 5597696; 5599923; 5599928 и 5688941, каждый из которых включен в настоящий документ посредством ссылки.

### Полимераза

Особенно полезная функция полимеразы заключается в том, чтобы катализировать полимеризацию цепи нуклеиновой кислоты с использованием существующей нуклеиновой кислоты в качестве матрицы. Другие полезные функции описаны в настоящем документе в другом месте. Примеры полезных полимераз включают в себя ДНК-полимеразы и РНК-полимеразы.

Способность улучшать специфичность, процессивность или другие свойства полимераз неприродных нуклеиновых кислот была бы весьма желательна во множестве контекстов, где, например, желательно включение неприродных нуклеиновых кислот, включая в себя амплификацию, секвенирование, мечение, обнаружение, клонирование и много других. В настоящем изобретении предусмотрены полимеразы с модифицированными свойствами для неприродных нуклеиновых кислот, способы получения таких полимераз, способы применения таких полимераз и многие другие признаки, которые станут очевидными после полного обзора следующего.

В некоторых случаях раскрытый в настоящем документе термин включает в себя полимеразы, которые включают неприродные нуклеиновые кислоты в растущую копию матрицы, например, во время амплификации ДНК. Согласно некоторым вариантам осуществления полимеразы могут быть модифицированы таким образом, что активный сайт полимеразы модифицируется для уменьшения подавления стерического проникновения неприродной нуклеиновой кислоты в активный сайт. Согласно некоторым вариантам осуществления полимеразы могут быть модифицированы для обеспечения комплементарности с одним или несколькими неприродными признаками неприродных нуклеиновых кислот. Соответственно, настоящее изобретение включает в себя композиции, которые включают в себя гетерологичную или рекомбинантную полимеразу и способы их применения.

Полимеразы могут быть модифицированы с использованием способов, относящихся к белковой инженерии. Например, молекулярное моделирование может быть выполнено на основе кристаллических структур, чтобы идентифицировать местоположения полимераз, где могут быть сделаны мутации для модификации целевой активности. 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) и Hayes, et al. Proc Natl Acad Sci, USA 99: 15926-15931 (2002).

Любая из множества полимераз может быть использована в способе или композиции, представленных в настоящем документе, включая в себя, например, ферменты на основе белка, выделенные из биологических систем, и их функциональные варианты. Следует понимать, что ссылка на конкретную полимеразу, такую как приведенная ниже в качестве примера, включает ее функциональные варианты, если не указано иное. Согласно некоторым вариантам осуществления полимеразы представляет собой полимеразу дикого типа. Согласно некоторым вариантам осуществления полимеразы представляет собой мо-

дифицированную или мутантную полимеразу.

Полимеразы, имеющие признаки для улучшения проникновения неприродных нуклеиновых кислот в области активного сайта и для координации с неприродными нуклеотидами в области активного сайта, также могут быть использованы. Согласно некоторым вариантам осуществления модифицированная полимеразы имеет модифицированный сайт связывания нуклеотидов.

Согласно некоторым вариантам осуществления модифицированная полимеразы обладает специфичностью к неприродной нуклеиновой кислоте, которая составляет по меньшей мере приблизительно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа по отношению к неприродной нуклеиновой кислоте. Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа обладает специфичностью к неприродной нуклеиновой кислоте, содержащей модифицированный сахар, которая составляет по меньшей мере приблизительно 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 неприродной нуклеиновой кислоте без модифицированного сахара. Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа обладает специфичностью к неприродной нуклеиновой кислоте, содержащей модифицированное основание, которое составляет по меньшей мере приблизительно 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 неприродной нуклеиновой кислоте без модифицированного основания. Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа обладает специфичностью к неприродной нуклеиновой кислоте, содержащей трифосфат, которая составляет по меньшей мере приблизительно 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, содержащей трифосфат, и/или неприродной нуклеиновой кислоте без трифосфата. Например, модифицированная полимеразы или полимеразы дикого типа может обладать специфичностью к неприродной нуклеиновой кислоте, содержащей трифосфат, которая составляет по меньшей мере приблизительно 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% специфичности полимеразы дикого типа по отношению к неприродной нуклеиновой кислоте с дифосфатом или монофосфатом, или без фосфата, или их комбинацией.

Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа обладает ослабленной специфичностью к неприродной нуклеиновой кислоте. Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа обладает специфичностью к неприродной нуклеиновой кислоте и специфичностью к природной нуклеиновой кислоте, которая составляет по меньшей мере приблизительно 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. Согласно некоторым вариантам осуществления модифицированная полимеразы дикого типа обладает специфичностью к неприродной нуклеиновой кислоте, содержащей модифицированный сахар, и специфичностью к природной нуклеиновой кислоте, которая составляет по меньшей мере приблизительно 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. Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа обладает специфичностью к неприродной нуклеиновой кислоте, содержащей модифицированное основание, и специфичностью к природной нуклеиновой кислоте, которая составляет по меньшей мере приблизительно 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.

Отсутствие экзонуклеазной активности может быть характеристикой дикого типа или характеристикой, придаваемой вариантной или сконструированной полимеразой. Например, экзо-минус фрагмент Кленова является мутированной версией фрагмента Кленова, в котором отсутствует 3'-5'-корректирующая экзонуклеазная активность.

Способ по настоящему изобретению может быть использован для расширения диапазона субстратов любой ДНК-полимеразы, которая не обладает собственной 3'-5'-корректирующей экзонуклеазной активностью или когда 3'-5'-корректирующая экзонуклеазная активность отключена, например, через мутацию. Примеры ДНК-полимераз включают в себя polA, polB (см., например, Parrel & Loeb, Nature Struc Biol 2001) polC, polD, polY, polX и обратные транскриптазы (RT), но предпочтительно представляют собой процессивные полимеразы высокой точности (PCT/GB 2004/004643). Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа по существу не обладает 3'-5'-корректирующей экзонуклеазной активностью. Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа по существу не обладает 3'-5'-корректирующей экзонуклеазной активностью в отношении неприродной нуклеиновой кислоты. Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа обладает 3'-5'-корректирующей экзонуклеазной активностью. Согласно некоторым вариантам осуществления модифицированная полимеразы или полимеразы дикого типа обладает 3'-5'-корректирующей экзо-

нуклеазной активностью в отношении природной нуклеиновой кислоты и по существу не обладает 3'-5'-корректирующей экзонуклеазной активностью в отношении неприродной нуклеиновой кислоты.

Согласно некоторым вариантам осуществления модифицированная полимераза обладает 3'-5'-корректирующей экзонуклеазной активностью, которая составляет по меньшей мере приблизительно 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% корректирующей экзонуклеазной активности полимеразы дикого типа. Согласно некоторым вариантам осуществления модифицированная полимераза обладает 3'-5'-корректирующей экзонуклеазной активностью в отношении неприродной нуклеиновой кислоты, которая составляет по меньшей мере приблизительно 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% корректирующей экзонуклеазной активности полимеразы дикого типа в отношении природной нуклеиновой кислоты. Согласно некоторым вариантам осуществления модифицированная полимераза обладает 3'-5'-корректирующей экзонуклеазной активностью в отношении неприродной нуклеиновой кислоты и 3'-5'-корректирующей экзонуклеазной активностью в отношении природной нуклеиновой кислоты, которая составляет по меньшей мере приблизительно 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% корректирующей экзонуклеазной активности полимеразы дикого типа в отношении природной нуклеиновой кислоты. Согласно некоторым вариантам осуществления модифицированная полимераза обладает 3'-5'-корректирующей экзонуклеазной активностью в отношении природной нуклеиновой кислоты, которая составляет по меньшей мере приблизительно 60%, 70%, 80%, 90%, 95%, 97%, 98%, 99%, 99,5%, 99,99% корректирующей экзонуклеазной активности полимеразы дикого типа в отношении природной нуклеиновой кислоты.

Согласно связанному аспекту в настоящем изобретении предусмотрены способы получения модифицированной полимеразы, которые предусматривают структурное моделирование исходной полимеразы, например, ДНК-полимеразы, идентификацию одной или нескольких характеристик комплексной стабильности или взаимодействия нуклеотидов, влияющих на комплексную стабильность или доступ или связывание нуклеотидов в активном сайте или признак комплементарности для нуклеотидного аналога в активном сайте и мутирование исходной полимеразы для включения или удаления этих признаков. Например, полимераза может быть мутирована для улучшения стерического доступа неприродного нуклеотида к активному сайту или для улучшения зарядового или гидрофобного взаимодействия между неприродным нуклеотидом и полимеразой. Способы также предусматривают определение того, проявляет ли полученная модифицированная полимераза повышенное включение нуклеотида или неприродного нуклеотида в растущую копию нуклеиновой кислоты по сравнению с исходной полимеразой.

Полимеразы могут быть охарактеризованы в соответствии с их скоростью диссоциации от нуклеиновых кислот. Согласно некоторым вариантам осуществления полимеразы характеризуется относительно низкой скоростью диссоциации для одной или нескольких природных и неприродных нуклеиновых кислот. Согласно некоторым вариантам осуществления полимеразы характеризуется относительно высокой скоростью диссоциации для одной или нескольких природных и неприродных нуклеиновых кислот. Скорость диссоциации представляет собой активность полимеразы, которую можно регулировать для настройки скорости реакции в способах, изложенных в настоящем документе.

Полимеразы могут быть охарактеризованы в зависимости от их правильности при использовании с конкретной природной и/или неприродной нуклеиновой кислотой или коллекциями природной и/или неприродной нуклеиновой кислоты. Под правильностью, как правило, понимают точность, с которой полимераза включает правильные нуклеиновые кислоты в растущую цепь нуклеиновой кислоты при создании копии матрицы нуклеиновой кислоты. Правильность ДНК-полимеразы может быть измерена как отношение правильных и неправильных включений природных и неприродных нуклеиновых кислот, когда присутствуют природная и неприродная нуклеиновая кислота, например, в равных концентрациях, чтобы конкурировать за синтез цепи в одном и том же сайте в бинарном комплексе цепь полимеразы-матрица нуклеиновой кислоты. Правильность ДНК-полимеразы можно рассчитать как отношение ( $k_{cat}/K_m$ ) для природной и неприродной нуклеиновой кислоты и ( $k_{cat}/K_m$ ) для неправильной природной и неприродной нуклеиновой кислоты; где  $k_{cat}$  и  $K_m$  представляют собой параметры Михаэлиса-Ментена в кинетике ферментативного состояния (Fersht, A. R. (1985) *Enzyme Structure and Mechanism*, 2nd ed., p. 350, W. H. Freeman & Co., New York., включенные в настоящий документ посредством ссылки). Согласно некоторым вариантам осуществления полимеразы характеризуется значением правильности по меньшей мере приблизительно 100, 1000, 10000, 100000 или  $1 \times 10^6$ , с корректирующей активностью или без нее.

Полимеразы из нативных источников или их вариантов могут быть подвергнуты скринингу с использованием анализа, который обнаруживает включение неприродной нуклеиновой кислоты, характеризующейся конкретной структурой. В одном примере полимеразы могут быть подвергнуты скринингу на способность включать неприродную нуклеиновую кислоту или UBP; например, d5SICSTP, dNaMTP или d5SICSTP-dNaMTP UBP. Можно использовать полимеразу, например, гетерологичную полимеразу, которая проявляет модифицированное свойство неприродной нуклеиновой кислоты по сравнению с полимеразой дикого типа. Например, модифицированным свойством может быть, например,  $K_m$ ,  $k_{cat}$ ,  $V_{max}$ , полимеразная процессивность в присутствии неприродной нуклеиновой кислоты (или встречающегося в природе нуклеотида), средняя длина считывания матрицы полимеразой в присутствии неприродной нук-

леиновой кислоты, специфичность полимеразы к неприродной нуклеиновой кислоте, скорость связывания неприродной нуклеиновой кислоты, скорость высвобождения продукта (пирофосфат, трифосфат и т.д.), скорость разветвления или любая их комбинация. Согласно одному варианту осуществления модифицированное свойство представляет собой пониженную  $K_m$  для неприродной нуклеиновой кислоты и/или повышенную  $k_{cat}/K_m$  или  $V_{max}/K_m$  для неприродной нуклеиновой кислоты. Подобным образом, полимеразы необязательно характеризуются повышенной скоростью связывания неприродной нуклеиновой кислоты, повышенной скоростью высвобождения продукта и/или пониженной скоростью ветвления по сравнению с полимеразой дикого типа.

В то же время, полимеразы могут включать природные нуклеиновые кислоты, например, А, С, G и Т, в растущую копию нуклеиновой кислоты. Например, полимеразы необязательно проявляют специфическую активность в отношении природной нуклеиновой кислоты, которая по меньшей мере на 5% выше (например, 5%, 10%, 25%, 50%, 75%, 100% или выше), чем соответствующая 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), чем полимеразы дикого типа в присутствии природной нуклеиновой кислоты. Необязательно, полимеразы отображают  $k_{cat}/K_m$  или  $V_{max}/K_m$  в отношении встречающегося в природе нуклеотида, которые по меньшей мере приблизительно на 5% выше (например, приблизительно 5%, 10%, 25%, 50%, 75% или 100% или выше), чем полимеразы дикого типа.

Используемые в настоящем документе полимеразы, которые могут обладать способностью включать неприродную нуклеиновую кислоту определенной структуры, также могут быть получены с использованием подхода направленной эволюции. Анализ синтеза нуклеиновой кислоты может быть использован для скрининга вариантов полимеразы, обладающих специфичностью к любой из множества неприродных нуклеиновых кислот. Например, варианты полимеразы могут быть подвергнуты скринингу на способность включать неприродную нуклеиновую кислоту или UBP; например, dTPT3, аналог dNaM, или UBP dTPT3-dNaM в нуклеиновые кислоты. Согласно некоторым вариантам осуществления такой анализ представляет собой анализ *in vitro*, например, с использованием варианта рекомбинантной полимеразы. Такие способы направленной эволюции могут быть использованы для скрининга вариантов любой подходящей полимеразы на активность в отношении любой из неприродных нуклеиновых кислот, представленных в настоящем документе.

Модифицированные полимеразы описанных композиций могут необязательно представлять собой модифицированную и/или рекомбинантную ДНК-полимеразу типа Ф29. Необязательно, полимеразы могут представлять собой модифицированную и/или рекомбинантную полимеразу Ф29, В103, GA-1, PZA, Ф15, BS32, M2Y, Nf, G1, Cp-1, PRD1, PZE, SF5, Cp-5, Cp-7, PR4, PR5, PR722 или L17.

Полимеразы нуклеиновых кислот, как правило, используемые в настоящем изобретении, включают в себя ДНК-полимеразы, РНК-полимеразы, обратные транскриптазы и их мутантные или измененные формы. ДНК-полимеразы и их свойства подробно описаны, в частности, среди прочего в публикации *DNA Replication 2<sup>nd</sup> edition*, Kornberg and Baker, W.H. Freeman, New York, N. Y. (1991). Известные обычные ДНК-полимеразы, используемые в настоящем изобретении, включают в себя, без ограничения, 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. 475:32), ДНК-полимеразу *Thermococcus litoralis* (Tli) (также называемую ДНК-полимеразой Vent<sup>TM</sup>, Cariello et al., 1991, Polynucleotides Res, 19:4193, New England Biolabs), ДНК-полимеразу 9°Nm<sup>TM</sup> (New England Biolabs), фрагмент Stoffel, Thermo Sequenase® (Amersham Pharmacia Biotech UK), Therminator<sup>TM</sup> (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 *Pyrococcus kodakaraensis* KOD (Takagi et al., 1997, Appl. Environ. Microbiol. 63:4504), ДНК-полимеразу JDF-3 (от *thermococcus* sp. JDF-3, заявка на патент WO 0132887), ДНК-полимеразу *Pyrococcus GB-D* (PGB-D) (также называемую ДНК-полимеразой Deep Vent<sup>TM</sup>, Juncosa-Ginesta et al., 1994, Biotechniques, 16:820, New England Biolabs), ДНК-полимеразу UITma (из термофила *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 *E. coli* (Lecomte and Doubleday, 1983, Polynucleotides Res. 11:7505), ДНК-полимеразу T7 (Nordstrom et al., 1981, J Biol. Chem. 256:3112) и археальную ДНК-полимеразу II DP11/DP2 (Cann et al., 1998, Proc. Natl. Acad. Sci. USA 95: 14250). Предполагаются как мезофильные полимеразы, так и термофильные полимеразы. Термофильные ДНК-полимеразы включают в себя, без ограничения, ThermoSequenase®, 9°Nm<sup>TM</sup>, Therminator<sup>TM</sup>, Taq, Tne, Tma, Pfu, Tfl, Tth, Tli, фрагмент Stoffel, ДНК-полимеразу Vent<sup>TM</sup> и Deep Vent<sup>TM</sup>, ДНК-полимеразу KOD, Tgo, JDF-3 и мутанты, их варианты и производные. Полимераза, которая является 3-экзонуклеазодефицитным мутантом, также рассматривается. Обратные транскриптазы, применимые в настоящем изобретении, включают в себя, без ограничения, обратные транскриптазы от HIV, HTLV-I, HTLV-II, FeLV, FIV, SIV, AMV, MMTV, MoMuLV и дру-

гих 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 в себя, без ограничения, ДНК-полимеразу 9<sup>o</sup>N, 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 мезофильных организмов, могут быть инактивируемыми при нагревании. Примерами являются ДНК-полимеразы из фага. Понятно, что полимеразы из любого из множества источников можно модифицировать для увеличения или уменьшения их толерантности к условиям высокой температуры. Согласно некоторым вариантам осуществления полимеразы может быть термофильной. Согласно некоторым вариантам осуществления термофильная полимеразы может быть инактивируемой при нагревании. Термофильные полимеразы, как правило, применимы для условий высокой температуры или в условиях термоциклирования, таких как те, которые используются для техник полимеразной цепной реакции (ПЦР).

Согласно некоторым вариантам осуществления полимеразы содержит 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<sup>o</sup>Nm™, ДНК-полимеразу Therminator™, Tne, Tma, Tfl, Tth, Tti, фрагмент Stoffel, ДНК-полимеразу Vent™ и Deep Vent™, ДНК-полимеразу KOD, Tgo, JDF-3, Pfu, Taq, ДНК-полимеразу T7, РНК-полимеразу T7, PGB-D, ДНК-полимеразу UTMa, ДНК-полимеразу I E.coli, ДНК-полимеразу III E.coli, археальную ДНК-полимеразу DP11/DP2 II, ДНК-полимеразу 9-N, ДНК-полимеразу Taq, ДНК-полимеразу Phusion®, ДНК-полимеразу Pfu, РНК-полимеразу SP6, ДНК-полимеразу RB69, обратную транскриптазу вируса миелобластома птиц (AMV), обратную транскриптазу вируса мышинного лейкоза (MMLV), обратную транскриптазу II SuperScript® и обратную транскриптазу III SuperScript®.

Согласно некоторым вариантам осуществления полимеразы представляет собой ДНК-полимеразный фрагмент I-Кленова, полимеразу Vent, ДНК-полимеразу Phusion®, ДНК-полимеразу KOD, полимеразу Taq, ДНК-полимеразу T7, РНК-полимеразу T7, ДНК-полимеразу Therminator™, полимеразу POLB, РНК-полимеразу SP6, ДНК-полимеразу I E.coli, ДНК-полимеразу III E.coli, обратную транскриптазу вируса миелобластома птиц (AMV), обратную транскриптазу вируса мышинного лейкоза (MMLV), обратную транскриптазу II SuperScript® или обратную транскриптазу III SuperScript®.

Кроме того, такие полимеразы могут быть использованы для амплификации и/или секвенирования ДНК, в том числе в режиме реального времени, например, в контексте амплификации или секвенирования, которые включают в себя включение остатков неприродной нуклеиновой кислоты в ДНК с помощью полимеразы. Согласно другим вариантам осуществления неприродная нуклеиновая кислота, которая включена, может быть такой же, как природный остаток, например, когда метка или другой фрагмент неприродной нуклеиновой кислоты удаляется под действием полимеразы во время включения, или неприродная нуклеиновая кислота может иметь одну или несколько особенностей, которые отличают ее от природной нуклеиновой кислоты.

Начиная, по меньшей мере, с последнего общего предка всей жизни на земле, генетическая информация сохранялась в четырехбуквенном алфавите, который распространяется и извлекается путем формирования двух пар оснований. Основная цель синтетической биологии заключается в создании новых форм жизни и функций, и наиболее общий путь к этой цели заключается в создании полусинтетических организмов (SSO), ДНК которых содержит две дополнительные буквы, которые образуют третью неприродную пару оснований (UBP). Ранее усилия авторов настоящего изобретения по созданию таких SSO завершились созданием штамма *Escherichia coli*, который с помощью транспортера нуклеозидтрифосфата из *Phaeodactylum tricornutum* (PtNTT2) импортирует необходимые неприродные трифосфаты из среды, а затем использует их для репликации плазмиды, содержащей UBP dNaM-dTPT3 (фиг. 1A). Хотя SSO хранит увеличенную информацию, она не извлекает ее, что требует *in vivo* транскрипции UBP в mRNA и tRNA, аминокислотирования tRNA с неприродной аминокислотой и, наконец, эффективного участия UBP в декодировании на рибосоме. В настоящем документе авторы настоящего изобретения сообщают о транскрипции *in vivo* ДНК, содержащей dNaM и dTPT3, в mRNA с двумя различными неприродными кодонами и tRNA с родственными неприродными антикодонами, и об их эффективном декодировании на рибосоме, чтобы направлять сайт-специфическое включение природных или неканонических аминокислот (ncAA) в суперсворачивающийся зеленый флуоресцентный белок (sfGFP). Результаты показывают, что взаимодействия, отличные от водородных связей, могут способствовать каждой стадии хранения и поиска информации. В результате SSO кодирует и получает увеличенную информацию и должна служить платформой для создания новых форм жизни и функций.

Зеленый флуоресцентный белок и его варианты, такие как sfGFP, служили модельными системами для изучения включения ncAA с использованием системы амбер-супрессора, в том числе в положении Y151, которое, как было показано, переносит различные природные и ncAA (фиг. 4). Чтобы исследовать декодирование неприродных кодонов, авторы настоящего изобретения сначала сфокусировались на включении Ser в положении 151 sfGFP, так как серин-аминоацил-tRNA-синтетаза *E.coli* (SerRS) не пола-

гается на распознавание антикодонов для аминокислотирования tRNA, таким образом устраняя потенциальные осложнения неэффективной зарядки. Штамм SSO YZ3 трансформировали плазмидой, кодирующей sfGFP, и геном tRNA<sup>Ser</sup> E. coli (serT), с заменой кодона 151 sfGFP (TAC) на неприродный кодон AXC (sfGFP(AXC)<sup>151</sup>; X=NaM) и заменой антикодона serT на неприродный антикодон GYT (tRNA<sup>Ser</sup>(GYT); Y=TPTЗ) (фиг. 1B). Трансформанты выращивали в среде, дополненной dNaMTP и dTPTЗTP, затем дополнительно добавляли NaMTP и TPTЗTP, а также изопропил-β-D-тиогаалактозид (IPTG), чтобы индуцировать экспрессию РНК-полимеразы T7 (RNAP T7) и tRNA<sup>Ser</sup> (GYT). После короткого периода индукции tRNA добавляли ангидротетрациклин (aTc) для индукции экспрессии sfGFP(AXC)<sup>151</sup>.

После индукции клетки, трансформированные контрольной плазмидой, кодирующей sfGFP(AXC)<sup>151</sup>, но лишённые tRNA<sup>Ser</sup>(GYT), показали резко сниженную флуоресценцию по сравнению с клетками, трансформированными плазмидой, кодирующей sfGFP, с природным кодоном Ser в положении 151 (sfGFP(AGT)<sup>151</sup>; фиг. 1C). Кроме того, рост клеток начал плато после индукции sfGFP(AXC)<sup>151</sup> (фиг. 1D), вероятно, из-за остановки и секвестрации рибосом. Лизаты этих клеток подвергали вестерн-блоттингу с антителом к GFP, которое выявило значительное снижение экспрессии sfGFP и присутствие укороченного sfGFP в положении неприродного кодона (фиг. 1E). Напротив, клетки, трансформированные плазмидой, кодирующей как sfGFP(AXC)<sup>151</sup>, так и tRNA<sup>Ser</sup>(GYT), проявляли флуоресценцию, которая была почти равна флуоресценции контрольных клеток, экспрессирующих sfGFP(AGT)<sup>151</sup> (фиг. 1C), рост клеток не индуцировал плато при индукции sfGFP(AXC)<sup>151</sup> (фиг. 1D) и вестерн-блоты лизатов из этих клеток выявили только полноразмерный белок sfGFP (фиг. 1E). Кроме того, авторы настоящего изобретения оценили способность всех четырех природных почти родственных tRNA (tRNA<sup>Ser</sup>(GNT); N=G, C, A или T), экспрессируемых идентичным образом, декодировать кодон AXC. В каждом случае наблюдалась небольшая флуоресценция, и дефект роста оставался (фиг. 5A и 5B). Эти данные демонстрируют, что PtNTT2 способен импортировать как дезокси-, так и риботрифосфаты обоих неприродных нуклеотидов, что РНК-полимераза T7 способна транскрибировать mRNA и tRNA, содержащие неприродные нуклеотиды *in vivo*, и что рибосома только эффективно декодирует неприродный кодон с неприродным антикодоном.

Чтобы оценить правильность декодирования, авторы настоящего изобретения проанализировали белок, очищенный из клеток, экспрессирующих sfGFP(AXC)<sup>151</sup> и tRNA<sup>Ser</sup>(GYT), с помощью LC/MS-MS, и относительное количественное определение с помощью интенсивностей пиков, которые выявили 98,5±0,7% (95% CI, n=4) включение Ser в положение 151, где Ile/Leu является преобладающим загрязнителем (фиг. 1F, табл. 4). Учитывая, что удержание UBP в гене sfGFP(AXC)<sup>151</sup> составило 98±2% (95% CI, n=4) (табл. 5) и что X→T, как правило, является основной мутацией во время репликации (что для AXC приведет в результате к кодону Ile ATC), авторы настоящего изобретения получают, что большая часть белка, не содержащего Ser в положении 151, связана с потерей UBP во время репликации и делают вывод о высокой точности трансляции с неприродным кодоном.

Таблица 4

Образец	Относительные ионные интенсивности MS1 (%)									
	S	Y	PrK	I/L	N	V	K	G	C	M
sfGFP(AGT) <sup>151</sup>	99,80	0,03	0,06	0,00	0,04	0,03	0,00	0,02	0,02	0,00
sfGFP(AXC) <sup>151</sup> / tRNA <sup>Ser</sup> (GYT)	98,47	0,04	0,04	1,23	0,14	0,02	0,00	0,05	0,01	0,00
sfGFP(TAC) <sup>151</sup>	0,11	99,71	0,06	0,00	0,05	0,02	0,00	0,02	0,02	0,01
sfGFP(TAG) <sup>151</sup> / tRNA <sup>Pro</sup> (CTA)	0,06	0,04	99,53	0,00	0,04	0,01	0,29	0,01	0,01	0,00
sfGFP(AXC) <sup>151</sup> / tRNA <sup>Pro</sup> (GYT)	0,25	0,03	96,16	2,06	1,06	0,02	0,37	0,03	0,01	0,00
sfGFP(GXC) <sup>151</sup> / tRNA <sup>Pro</sup> (GYC)	0,06	0,04	97,50	0,00	0,01	1,26	0,74	0,37	0,01	0,00
95% CI (%)										
Образец	S	Y	PrK	I/L	N	V	K	G	C	M
sfGFP(AGT) <sup>151</sup>	0,31	0,04	0,09	0,00	0,06	0,05	0,01	0,03	0,03	0,00
sfGFP(AXC) <sup>151</sup> / tRNA <sup>Ser</sup> (GYT)	0,73	0,04	0,03	0,64	0,04	0,01	0,00	0,04	0,01	0,00
sfGFP(TAC) <sup>151</sup>	0,06	0,11	0,05	0,00	0,03	0,02	0,00	0,01	0,02	0,00
sfGFP(TAG) <sup>151</sup> / tRNA <sup>Pro</sup> (CTA)	0,03	0,02	0,11	0,00	0,02	0,02	0,03	0,01	0,01	0,00
sfGFP(AXC) <sup>151</sup> / tRNA <sup>Pro</sup> (GYT)	0,13	0,02	0,25	0,06	0,03	0,01	0,06	0,01	0,02	0,01
sfGFP(GXC) <sup>151</sup> / tRNA <sup>Pro</sup> (GYC)	0,05	0,04	0,70	0,00	0,01	0,24	0,28	0,22	0,01	0,00

Таблица 5

aaRS	tRNA	NaMTP	TPTЗTP	Кодон	% удерживания UBP <i>sfGFP</i>	Анти кодон	% удерживания UBP в гене tRNA
SerRS	-	+	+	AXC	98±0	-	нет данных
SerRS§	Ser	+	+	AXC	98±2	GYT	89±2
SerRS	Ser	+	+	AXC	94±8	GAT	нет данных
SerRS	Ser	+	+	AXC	94±2	GGT	нет данных
SerRS	Ser	+	+	AXC	95±0	GCT	нет данных
SerRS	Ser	+	+	AXC	95±1	GTT	нет данных
-	Pyl	+	+	AXC	97±1	GYT	89±2
PylRS	-	+	+	AXC	97±1	-	нет данных
PylRS	Pyl	+	+	TAC	нет данных	GYT	92±3
PylRS§	Pyl	+	+	AXC	96±1	GYT	90±2
PylRS*	Pyl	+	+	AXC	98±0	GYT	95±2
PylRS*	Pyl	+	-	AXC	98±1	GYT	96±1
PylRS*	Pyl	-	+	AXC	98±1	GYT	95±1
PylRS*	Pyl	-	-	AXC	97±1	GYT	94±4
-	Pyl	+	+	GXC	98±1	GYC	96±3
PylRS	-	+	+	GXC	97±3	-	нет данных
PylRS	Pyl	+	+	TAC	нет данных	GYC	96±1
PylRS§	Pyl	+	+	GXC	97±1	GYC	95±C
PylRS*	Pyl	+	+	GXC	96±3	GYC	97±1
PylRS*	Pyl	+	-	GXC	96±2	GYC	97±1
PylRS*	Pyl	-	+	GXC	97±2	GYC	97±0
PylRS*	Pyl	-	-	GXC	96±1	GYC	97±1
<i>pAzFRS</i> RS§	<i>pAzF</i>	+	+	AXC	98±0	GYT	90±1
<i>pAzFRS</i> RS	<i>pAzF</i>	+	+	TAC	нет данных	GYT	91±1

\* Соответствует культурам, проанализированным на фиг. 7А-7D.

Чтобы продемонстрировать кодирование *ncAA* с помощью UBP, авторы настоящего изобретения сконструировали плазмиды, аналогичные тем, которые использовались выше, но с заменой гена  $tRNA^{Ser}$  геном  $tRNA^{Pyl}(GYT)$  *Methanosarcina mazei*.  $tRNA^{Pyl}$  может быть селективно заряжен с помощью пирролизин-аминоацил-tRNA-синтетазы (*PylRS*) *Methanosarcina barkeri* с *ncAA* N<sup>6</sup>-[(2-пропилилокси)карбонил]-L-лизином (PrK). В дополнение к кодону AXC авторы настоящего изобретения также проанализировали кодон GXC и соответствующий  $tRNA^{Pyl}(GYC)$ . SSO, несущую отдельную плазмиду, кодирующую индуцируемый IPTG *PylRS*, трансформировали необходимыми плазмидами и выращивали с добавлением или без добавления PrK. В контрольных экспериментах с клетками, экспрессирующими *sfGFP*(AXC)<sup>151</sup> или *sfGFP*(GXC)<sup>151</sup> в отсутствие *PylRS*, родственного неприродного  $tRNA^{Pyl}$  или PrK, авторы настоящего изобретения наблюдали только низкую клеточную флуоресценцию (фиг. 2А), укорочение *sfGFP* (фиг. 6А и 6В) и плато в росте клеток (фиг. 6В). Напротив, для любой неприродной mRNA с ее родственной неприродной tRNA, когда присутствовал *PylRS* и добавляли PrK, авторы настоящего изобретения наблюдали высокую флуоресценцию (64 и 69% *sfGFP* (TAC)<sup>151</sup> для AXC и GXC, соответственно) (фиг. 2А и 2В), устойчивое производство полноразмерного *sfGFP* (фиг. 6А) и нормальный рост (фиг. 6В).

Чтобы проверить включение PrK, *sfGFP* подвергали аффинной очистке от клеточных лизатов с использованием С-концевого Strep-tag II и подвергали катализируемой медью клик-химии для присоединения красителя карбокситетраметилродамина (TAMRA) (TAMRA-PEG4-N3), который, как было установлено, смещает электрофоретическую подвижность *sfGFP* во время ДСН-ПААГ, что позволяет авторам настоящего изобретения оценить точность включения PrK с помощью вестерн-блоттинга (фиг. 2С). Авторы настоящего изобретения наблюдали сильный сигнал TAMRA и что практически весь *sfGFP* был сдвинут при очистке от клеток, экспрессирующих *sfGFP*(AXC)<sup>151</sup> и  $tRNA^{Pyl}(GYT)$  или *sfGFP*(GXC)<sup>151</sup> и  $tRNA^{Pyl}(GYC)$  и тех, которые культивировались в средах, дополненных PrK (фиг. 2С). Напротив, сигнал TAMRA или сдвиг *sfGFP* практически отсутствовали, когда NaMTP, TPTЗTP или оба отсутствовали (фиг. 7А и 7В). Наконец, не было обнаружено сигнала TAMRA или сдвинутого *sfGFP* в белке, очищенном от клеток, экспрессирующих *sfGFP*(TAC)<sup>151</sup>, с любой неприродной tRNA (фиг. 2С). Эти данные демонстрируют, что PrK специфически включается в *sfGFP* посредством декодирования неприродных кодонов посредством tRNA с неприродным антикодоном.

При оптимальных концентрациях PrK (фиг. 8А-8D) авторы настоящего изобретения очистили 54±4 и 55±6 мкг/мл *sfGFP* (s.d., n=4, ~40% от контроля *sfGFP*(TAC)<sup>151</sup> (табл. 6)) для кодонов AXC и GXC, соответственно. Кроме того, на основе масс-спектрометрического анализа чистота *sfGFP* с PrK составила 96,2±0,3% (95% CI, n=4) для кодона AXC и 97,5±0,7% (95% CI, n=4) для кодона GXC (фиг. 2D). Хотя выход очищенного белка *sfGFP* был несколько ниже, чем при амбер-супрессии (87±6 мкг/мл, s.d., n=4 (табл. 6)), что обусловлено умеренным снижением роста с добавлением неприродных риботрифосфатов (фиг. 7С и 7D), декодирование обоих неприродных кодонов приводило к более высокой флуоресценции, чем амбер-супрессия при нормировании к плотности клеток (фиг. 2А и 2В), подразумевая, что декодирование с помощью неприродных кодонов более эффективно, чем амбер-супрессия.

Чтобы исследовать кодирование других *ncAA* с помощью UBP, авторы настоящего изобретения исследовали кодирование *pAzF* с помощью кодона AXC и созданной пары *TyrRS/tRNA<sup>Tyr</sup>* (*pAzFRS/tRNA<sup>pAzF</sup>*) *Methanococcus jannaschii*. При индукции синтетазы и добавлении *pAzF* в питательную среду авторы настоящего изобретения наблюдали устойчивую флуоресценцию, эквивалентную флуоресценции клеток, экспрессирующих природный *sfGFP*(TAC)<sup>151</sup>, и нормальный рост с *sfGFP*(AXC)<sup>151</sup> и  $tRNA^{pAzF}(GYT)$  (фиг. 3А, фиг. 9). Полноразмерный *sfGFP* очищали (86±6 мкг/мл, s.d.,

n=4; 68% контроля sfGFP(TAC)<sup>151</sup>, табл. 6) и подвергали клик-химии без меди с использованием дибензоциклооктильной (DBCO) группы для прикрепления TAMRA (TAMRA-PEG4-DBCO). Авторы настоящего изобретения наблюдали надежную конъюгацию TAMRA с sfGFP, выделенным из клеток, экспрессирующих sfGFP(AXC)<sup>151</sup> и tRNA<sup>pAzF</sup>(GYT), и культивировали в присутствии pAzF (фиг. 3B). Хотя авторы настоящего изобретения не смогли точно оценить точность включения pAzF из-за разложения азидофрагмента, было сдвинуто ~93% белка sfGFP, что выгодно отличается от сдвига sfGFP на ~95%, полученного путем амбер-супрессии (фиг. 3B).

Таблица 6

Образец	aaRS	Выход (мкг/мл)	Относительно контроля (%)	Общая флуор. (а.е.)	Относительно контроля (%)
sfGFP(AGT) <sup>151</sup>	SerRS (эндогенный)	100±8	100	269	100
sfGFP(AXC) <sup>151</sup> /tRNA <sup>Ser</sup> (GYT)		97±9	96	259	96
sfGFP(TAC) <sup>151</sup>	PylRS	135±17	100	400	100
sfGFP(TAG) <sup>151</sup> /tRNA <sup>Pyl</sup> (CTA)		87±6	65	242	60
sfGFP(AXC) <sup>151</sup> /tRNA <sup>Pyl</sup> (GYT)		54±4	40	153	38
sfGFP(GXC) <sup>151</sup> /tRNA <sup>Pyl</sup> (GYC)		55±6	41	166	41
sfGFP(TAC) <sup>151</sup>	pAzFRS	127±15	100	405	100
sfGFP(TAG) <sup>151</sup> /tRNA <sup>pAzF</sup> (CTA)		75±9	59	287	71
sfGFP(AXC) <sup>151</sup> /tRNA <sup>pAzF</sup> (GYT)		86±6	68	333	82

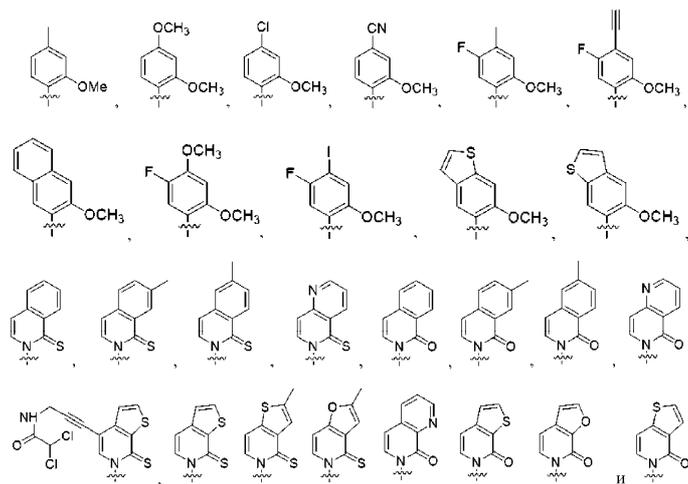
Начиная, по меньшей мере, с последнего общего предка всей жизни на земле, белки были получены путем декодирования кодонов, написанных исключительно с использованием четырехнуклеотидного генетического алфавита. Теперь авторы настоящего изобретения продемонстрировали расшифровку двух новых кодонов, написанных с расширенным генетическим алфавитом, и использовали новые кодоны для сайт-специфического включения псАА в белки. Авторы настоящего изобретения обнаруживают, что на каждой стадии хранения и поиска информации водородные связи, столь очевидно являющиеся центральными для природных пар оснований, могут, по меньшей мере, частично заменяться комплементарной упаковкой и гидрофобными силами. Несмотря на их новый механизм декодирования, неприродные кодоны могут быть декодированы так же эффективно, как и их полностью природные аналоги. Хотя авторы настоящего изобретения рассмотрели только декодирование двух неприродных кодонов, UBP вряд ли будет ограничен ими, и в сочетании с недавно сообщенной системой редактирования Cas9, которая усиливает удержание UBP, он, вероятно, сделает доступным больше кодонов, чем когда-либо может быть использовано. Таким образом, сообщаемая SSO может быть только первой из новой формы полусинтетической жизни, которая способна получить доступ к широкому спектру форм и функций, недоступных природным организмам.

Хотя предпочтительные варианты осуществления настоящего изобретения были показаны и описаны в настоящем документе, для специалиста в настоящей области техники будет очевидно, что такие варианты осуществления предоставлены только в качестве примера. Для специалистов в настоящей области техники будут происходить многочисленные вариации, изменения и замены без отступления от настоящего изобретения. Следует понимать, что различные альтернативы вариантам осуществления настоящего изобретения, описанным в настоящем документе, могут быть использованы при практическом применении настоящего изобретения. Предполагается, что нижеследующая формула изобретения определяет объем настоящего изобретения и что таким образом охватываются способы и структуры в пределах объема этой формулы изобретения и их эквивалентов.

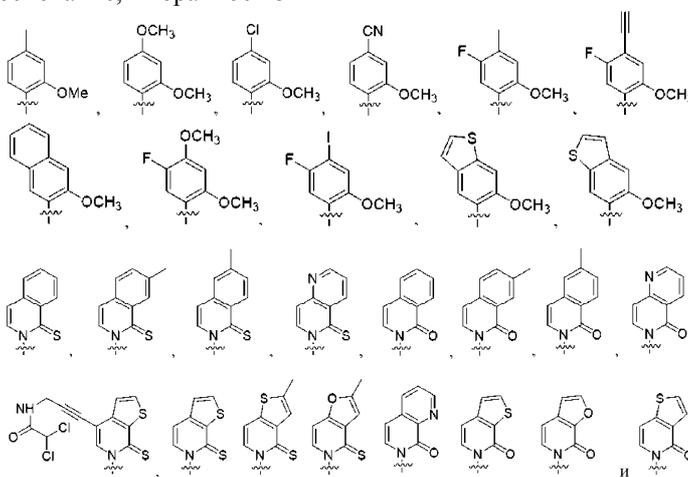
#### ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Способ получения белка, содержащего неприродную аминокислоту, причем способ предусматривает

получение мутантной tRNA, причем мутантная tRNA содержит последовательность мутантного антикодона, выбранную из табл. 1 или 2, при этом последовательность мутантного антикодона содержит неприродное нуклеотидное основание, выбранное из



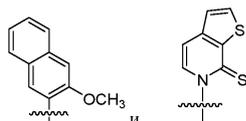
получение мутантной mRNA, причем мутантная mRNA содержит последовательность мутантного кодона, выбранную из табл. 1 или 2, при этом последовательность мутантного кодона содержит неприродное нуклеотидное основание, выбранное из



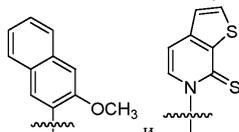
и синтез белка, содержащего неприродную аминокислоту, с использованием мутантной tRNA и мутантной mRNA;

причем белок синтезируют в клетке и указанная клетка содержит транспортер нуклеозидтрифосфата из *Phaeodactylum tricornutum* и аминоксил-tRNA-синтетазу из *Methanosarcina* или *Methanococcus jannaschii*.

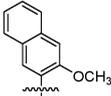
2. Способ по п.1, при котором клетка включает в себя микроорганизм.
3. Способ по п.1, при котором клетка включает в себя бактерию.
4. Способ по п.3, при котором клетка включает в себя *Escherichia coli*.
5. Способ по любому из пп.1-4, при котором мутантная tRNA заряжена неприродным аминокислотным остатком.
6. Способ по любому из пп.1-5, при котором клетка дополнительно содержит олигонуклеотид, кодирующий tRNA.
7. Способ по любому из пп.1-6, при котором клетка дополнительно содержит олигонуклеотид, кодирующий аминоксил-tRNA-синтетазу.
8. Способ по любому из пп.1-7, при котором клетка дополнительно содержит олигонуклеотид, кодирующий mRNA.
9. Способ по любому из пп.1-5, при котором клетка дополнительно содержит олигонуклеотид, который кодирует tRNA, mRNA и аминоксил-tRNA-синтетазу.
10. Способ по любому из пп.1-9, при котором аминоксил-tRNA-синтетаза происходит из *Methanosarcina*.
11. Способ по любому из пп.1-9, при котором аминоксил-tRNA-синтетаза происходит из *Methanosarcina barkeri*.
12. Способ по любому из пп.1-9, при котором аминоксил-tRNA-синтетаза происходит из *Methanococcus jannaschii*.
13. Способ по любому из пп.1-12, при котором tRNA содержит неприродное нуклеотидное основание, выбранное из



и mRNA содержит неприродное нуклеотидное основание, выбранное из

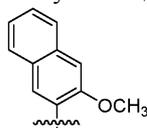


14. Способ по любому из пп.1-12, при котором mRNA содержит неприродное нуклеотидное осно-

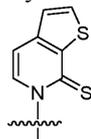
вание,  и tRNA содержит неприродное нуклеотидное основание .

15. Способ по любому из пп.1-14, при котором tRNA содержит один или несколько антикодонов, выбранных из GYT и GYC, причем Y представляет собой неприродное нуклеотидное основание.

16. Способ по п.15, при котором неприродное нуклеотидное основание Y представляет собой



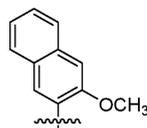
17. Способ по п.15, при котором неприродное нуклеотидное основание Y представляет собой



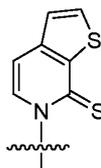
18. Способ по любому из пп.1-17, при котором mRNA содержит кодон GXC, причем X представляет собой неприродное нуклеотидное основание.

19. Способ по любому из пп.1-17, при котором mRNA содержит кодон AXC, причем X представляет собой неприродное нуклеотидное основание.

20. Способ по п.18 или 19, при котором неприродное нуклеотидное основание X представляет собой



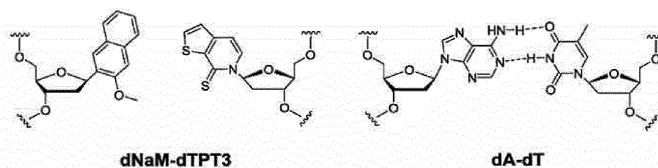
21. Способ по п.18 или 19, при котором неприродное нуклеотидное основание X представляет собой



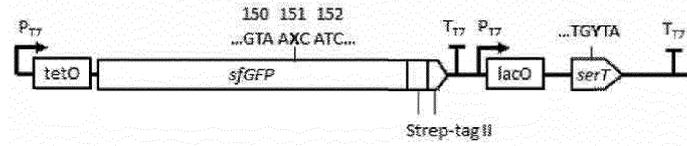
22. Способ по любому из пп.1-21, при котором мутантная tRNA представляет собой мутанта tRNA из *Methanosarcina mazei*.

23. Способ по любому из пп.1-22, при котором аминоксил-tRNA-синтетаза представляет собой пирролизин-аминоксил-tRNA-синтетазу.

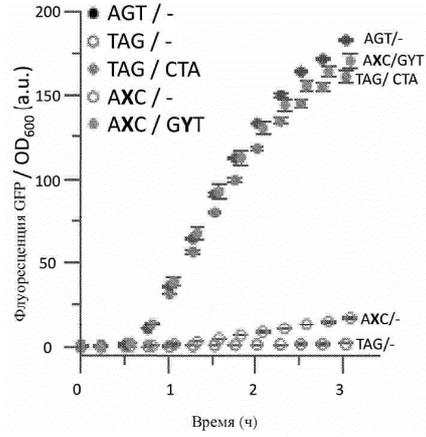
24. Способ по любому из пп.1-22, при котором аминоксил-tRNA-синтетаза представляет собой тирозин-аминоксил-tRNA-синтетазу.



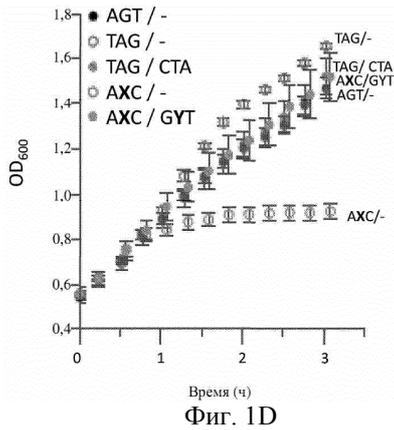
Фиг. 1А



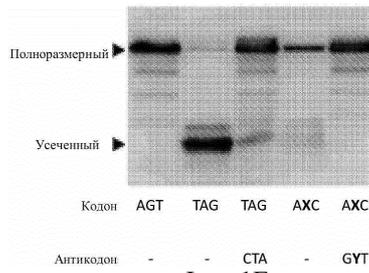
Фиг. 1В



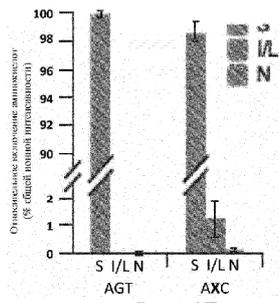
Фиг. 1С



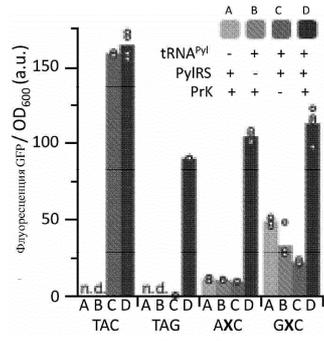
Фиг. 1D



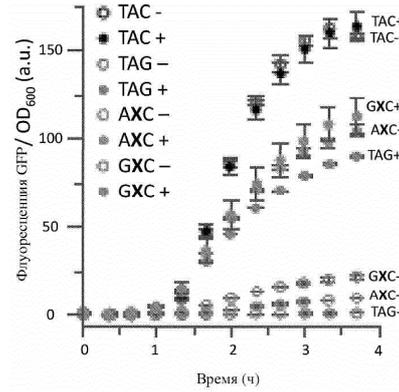
Фиг. 1E



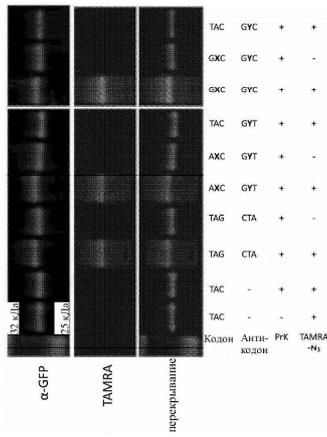
Фиг. 1F



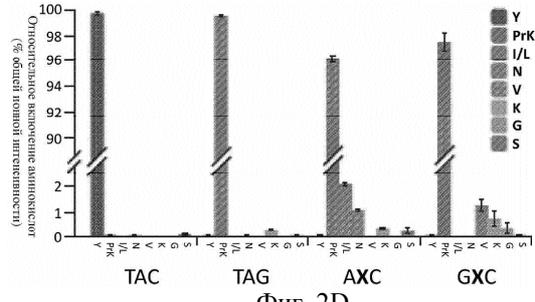
Фиг. 2А



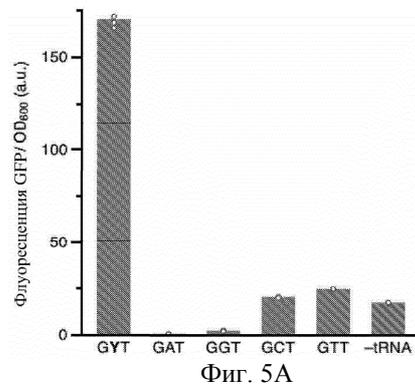
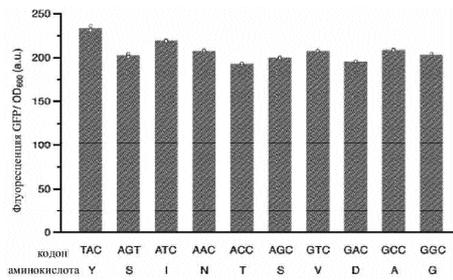
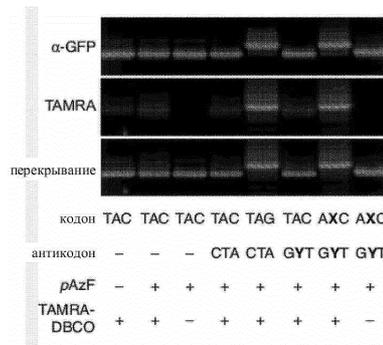
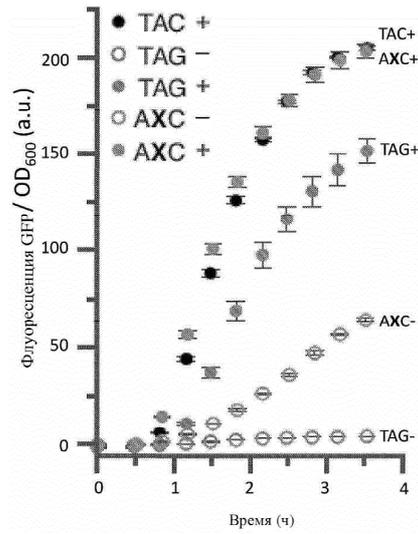
Фиг. 2В

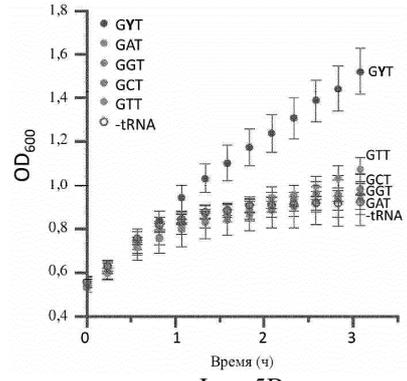


Фиг. 2С

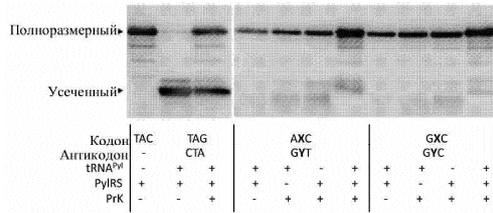


Фиг. 2D

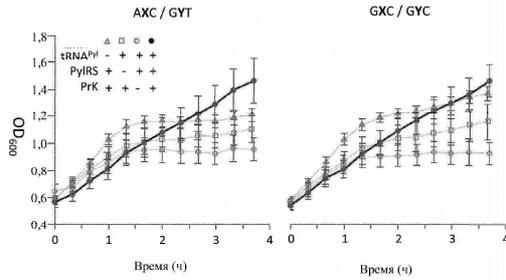




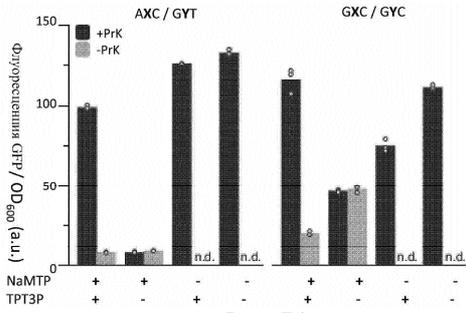
Фиг. 5В



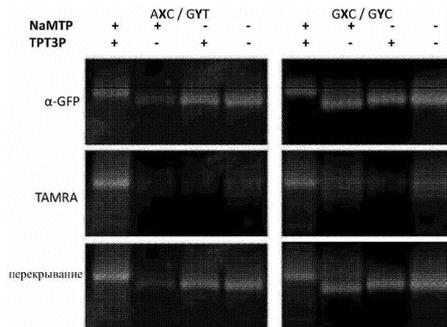
Фиг. 6А



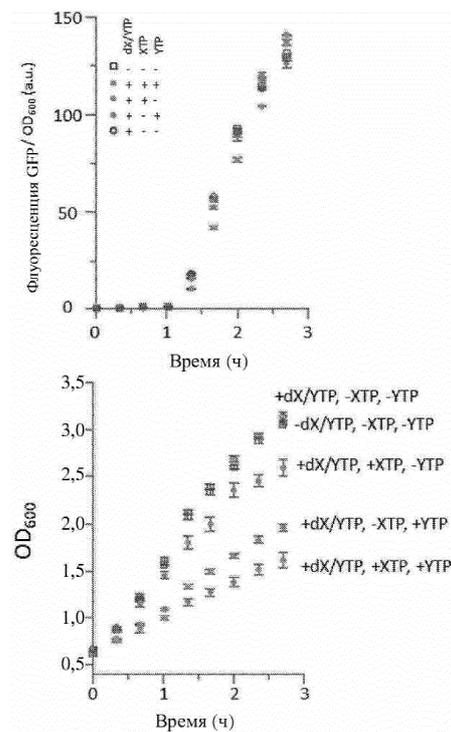
Фиг. 6В



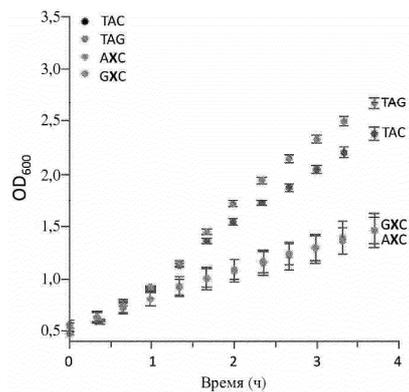
Фиг. 7А



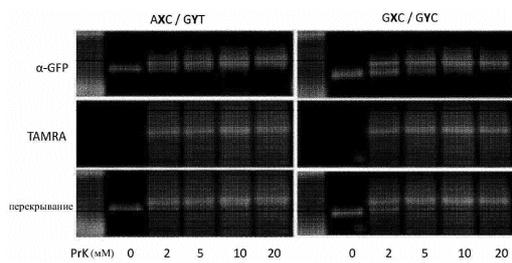
Фиг. 7В



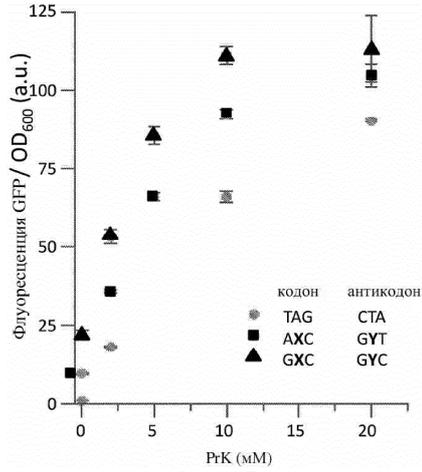
Фиг. 7С



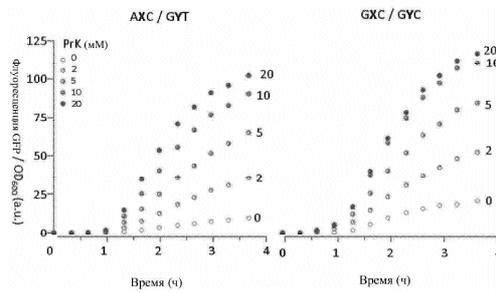
Фиг. 7D



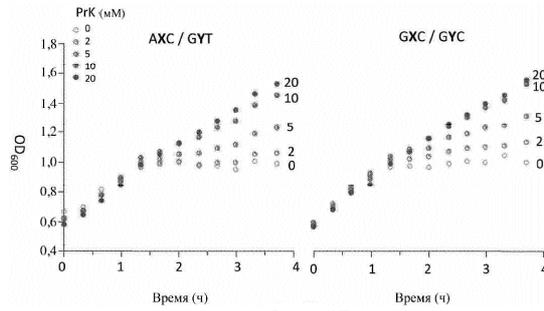
Фиг. 8А



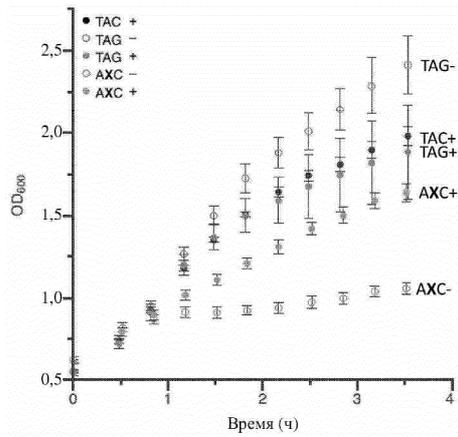
Фиг. 8В



Фиг. 8С



Фиг. 8D



Фиг. 9

