

(19)



Евразийское  
патентное  
ведомство

(21) 202091536 (13) A1

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки  
2020.09.23

(51) Int. Cl. C12N 15/113 (2010.01)  
C12N 15/70 (2006.01)  
C07H 3/06 (2006.01)

(22) Дата подачи заявки  
2018.12.19

(54) КОНСТРУКЦИЯ НУКЛЕИНОВОЙ КИСЛОТЫ ДЛЯ ЭКСПРЕССИИ ГЕНОВ IN VITRO И IN VIVO

(31) РА 2017 00737; РА 2018 00173; РА 2018 00231

(72) Изобретатель:  
Педерсен Маргит, Пападакис Манос  
(DK)

(32) 2017.12.21; 2018.04.18; 2018.05.24

(33) DK

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

(86) PCT/IB2018/060355

(87) WO 2019/123324 2019.06.27

(71) Заявитель:  
ГЛЮКОМ А/С (DK)

(57) Настоящее изобретение относится к области рекомбинантной продукции биологических молекул в клетках-хозяевах. Изобретение относится к нуклеотидным конструкциям, которые позволяют модифицировать экспрессию желаемого гена с использованием систем экспрессии генов in vitro и in vivo. Конструкции могут быть преимущественно использованы для рекомбинантного получения различных биологических молекул в промышленных масштабах, например олигосахаридов грудного молока (НМО).



A1

202091536

202091536

A1

## ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420-563708EA/042

### КОНСТРУКЦИЯ НУКЛЕИНОВОЙ КИСЛОТЫ ДЛЯ ЭКСПРЕССИИ ГЕНОВ IN VITRO И IN VIVO

#### Область техники, к которой относится изобретение

Настоящее изобретение относится к области рекомбинантной продукции биологических молекул в клетках-хозяевах. Изобретение относится к нуклеотидным конструкциям, которые позволяют модифицировать экспрессию желаемого гена с использованием систем экспрессии генов как *in vitro*, так и *in vivo*. Эти конструкции могут преимущественно использоваться для получения множества биологических молекул рекомбинантным способом в промышленных масштабах, например, олигосахаридов грудного молока (НМО).

#### Уровень техники

Коммерческая важность бактериальных клеток для получения рекомбинантных молекул возрастает. В настоящее время для продукции рекомбинантных белков в бактериальных хозяевах, в частности в *E. coli*, в основном используются плазмидные системы экспрессии. Поскольку эти системы обеспечивают высокое содержание генного материала и хорошо известны, они стали широко распространенными, в том числе благодаря простоте доступных протоколов клонирования. Однако использование систем экспрессии на основе плазмид, особенно в производственных масштабах, также имеет ряд недостатков.

Плазмидные прокариотические системы экспрессии обычно характеризуются высоким числом копий плазмиды, например, до нескольких сотен на клетку. Экспрессионные плазмиды обычно несут представляющий интерес ген под контролем промотора, точку инициации репликации (*ori*) и маркерный ген для отбора клонов, несущих плазмиду. Кроме того, в указанных плаزمидах (то есть векторах) часто присутствуют кодирующие или не кодирующие или нефункциональные каркасные последовательности. Присутствие плазмид и соответствующий механизм репликации изменяют метаболизм клетки-хозяина (Diaz-Rizzi and Hernandez, (2000) *Crit Rev Biotechnol.* ;20(2):79-108) и налагают высокую метаболическую нагрузку на клетки, тем самым ограничивая их ресурсы для производства рекомбинантного белка. Кроме того, использование сильных промоторов в сочетании с высоким содержанием генного материала дают скорость образования рекомбинантного белка, которая обычно оказывается слишком высокой для клетки-хозяина, и поэтому может привести к быстрому и необратимому нарушению клеточного метаболизма. Следовательно, потенциал клеток-хозяев не может полностью использоваться в системах на основе плазмид, что приводит к низкому выходу и качеству рекомбинантного белка. Таким образом, один из основных недостатков систем экспрессии на основе плазмид может быть объяснен повышенной потребностью в питательных веществах и энергии, которая требуется для репликации и поддержания плазмид.

Другим типичным явлением в системах на основе плазмид является изменение количества копий плазмиды в процессе культивирования. Продукция рекомбинантного белка при высоких скоростях экспрессии сопровождается голоданием и клеточным стрессом, которые приводят к увеличению пулов ненагруженных тРНК. Это приводит к вмешательству в механизм контроля количества копий плазмиды (PCN). В результате, PCN быстро увеличивается и вызывает нарушение процесса культивирования (так называемый «эффект убегания»).

Сегрегационная нестабильность (т.е. появление не содержащих плазмиду клеток-хозяев) и структурная нестабильность (т.е. мутации в плазмидной последовательности) представляют собой дополнительные проблемы, часто наблюдаемые в системах на основе плазмид. Во время клеточного деления клетки могут терять плазмиду и, следовательно, также и представляющий интерес ген. Такая потеря плазмиды зависит от нескольких внешних факторов и усиливается с увеличением количества клеточных делений (поколений). Это означает, что ферментация на основе плазмид ограничена количеством поколений или удвоений клеток.

В целом, эти свойства плазмидных экспрессионных систем лимитируют выход рекомбинантного белка и снижают управляемость и экономичность процесса.

В поиске эффективной альтернативы плазмидной экспрессии, в WO 1996/40722, посвященном геномной экспрессии, описан способ, в котором используется интеграция кольцевого вектора (так называемого «кольцевая ДНК-переносчик в хромосому», STD), включающего селективируемый маркер, в бактериальную хромосому (то есть в сайт attB *E. coli*). В этом способе с использованием дублицированных последовательностей ДНК, фланкирующих маркер селекции, была достигнута амплификация содержания хромосомного гена. Таким образом, полученное содержание хромосомного гена составляло приблизительно 15-40 копий на клетку, что было аналогично количеству, которое достигаются с традиционно используемыми плазмидными векторами. Культивирование клонов, содержащих ДНК-переносчик в хромосому, интегрированную в бактериальный геном, обеспечивало уровень рекомбинантных белков, аналогичный уровню, получаемому с помощью систем на основе плазмид (Olson et al., 1998). В этом способе требовалось *in vitro* лигирование STD, и он относительно встраивания был ограничен сайтом attB.

Геномные системы экспрессии, по-видимому, имеют большой потенциал для обеспечения стабильной экспрессии рекомбинантных генов без маркеров селекции. Однако часто экспрессия рекомбинантного гена в промышленном масштабе достижима только путем увеличения содержания копий гена в хромосоме до уровня копийности плазмиды, так как одна копия гена не способна обеспечить экспрессию в промышленном масштабе. Кроме того, выбор сайта интеграции является проблемой, а регулирование экспрессии часто является сложным и/или не подходит для промышленного получения. Таким образом, не существует простой и эффективной бактериальной системы экспрессии на основе генома для промышленного получения рекомбинантных полипептидов.

Одним из подходов к преодолению проблемы недостаточного уровня продукции и комплексной регуляции бактериальной геномной экспрессии гетерологичных полипептидов является использование сильных индуцибельных промоторов для контроля транскрипции встроенных рекомбинантных генов. Был описан ряд различных индуцибельных промоторов. Например, промоторы, индуцируемые высокой температурой, такие как  $\lambda P_R$  and  $\lambda P_L$ , истощением по триптофану, такой как *trp*, 1-арабинозой, такая как *araBAD*, маннитом, такой как *mtsE*, фосфатным голоданием, такой как *rhoA*, налидиксовой кислотой, такой как *recA*, осмолярностью, такой как *proU*, истощением глюкозы, такой как *cst-1*, тетрациклином, такой как *tetA*, pH, такой как *cadA*, анаэробными условиями, такой как *nar*, инфекцией T4, такой как ген 32 фага T4, алкил- или галогенбензоатами, такой как *Pm*, алкил- или галогентолуолами, такой как *Pu*, салицилатами, такой как *PsaI*, и кислородом, такой как *VHb*, все были исследованы в качестве альтернативы индуцируемому IPTG промоторам (см., например, Makrides, S. C. (1996) *Microbiol. Rev.* 60, 512-538; Hannig G. & Makrides, S. C. (1998) *TIBTECH* 16, 54-60; Stevens, R. C. (2000) *Structures* 8, R177-R185; Hoffmann J & Altenbuchner J (2015) *PLoS One*, 10(7) e0133248; J. Sanchez-Romero & V. De Lorenzo, Genetic Engineering of Nonpathogenic *Pseudomonas* strains as Biocatalysts for Industrial and Environmental Processes, in *Manual of Industrial Microbiology and Biotechnology* (A. Demain & J. Davies, eds.) pp. 460-74 (1999) (ASM Press, Washington, D.C.); H. Schweizer, Vectors to express foreign genes and techniques to monitor gene expression for *Pseudomonads*, *Current Opinion in Biotechnology*, 12:439-445 (2001); and R. Slater & R. Williams, The Expression of Foreign DNA in Bacteria, in *Molecular Biology and Biotechnology* (J. Walker & R. Rapley, eds.) pp. 125-54 (2000) (The Royal Society of Chemistry, Cambridge, UK). Однако у этих индуцибельных промоторов есть ряд проблем, такие как высокотемпературная индукция, которая вредна для клеток и со временем может оказаться непрактичной для крупномасштабной ферментации из-за ограничений оборудования; манипуляции с кислородом могут влиять на общую динамику аспектов плотности роста клеток в процессе ферментации, снижая идеальные условия; использование толуолов или других подобных типов потенциально токсичных химических веществ может потребовать дополнительной очистки, чтобы гарантировать, что эти соединения не присутствуют в конечном продукте; а pH может влиять на способность представляющего интерес пептида правильно сворачиваться или растворяться в клетке-хозяине, что делает очистку более дорогостоящей и трудоемкой, что делает плазмидные системы экспрессии по-прежнему предпочтительным выбором в случае промышленного производства.

Регуляция активности промотора источником углерода, вероятно, является наиболее привлекательным вариантом для контроля экспрессии целевого полипептида в промышленных условиях. Для этого есть несколько причин, например, более эффективное использование источника углерода и снижение длительных метаболических стрессов у клетки-хозяина. Однако в настоящее время выбор таких промоторов довольно ограничен, и большинство из них были адаптированы для плазмидной экспрессии (Terpe

К. Appl Microbiol Biotechnol (2006) 72:211-222). Однако геном бактериальной клетки, например *E. coli*, содержит тысячи промоторов, и многие из них регулируются изменениями в источнике углерода, что позволяет наличию углерода в окружающей среде влиять на характер экспрессии генов, находящихся под их контролем. Было высказано предположение, что глобальный регулятор транскрипции, cAMP-CRP, который формируется при снижении уровня глюкозы, регулирует как минимум 378 промоторов в бактериальной клетке (Shimada T. et al., PloS One 6(6): e20081, (2011)), однако нет данных, которые позволили бы предположить, какие из этих промоторов являются мощными, способными обеспечивать стабильную контролируемую геномную продукцию рекомбинантных полипептидов с высоким выходом в промышленных условиях.

Дыхательный метаболизм глицерина в *Escherichia coli* (*E. coli*) контролируется 12 генами, организованными в 5 *glp*-оперонов: *glpFKX*, *glpABC*, *glpTQ*, *glpD* и *PglpEGR*. Гены оперона *glpFKX* кодируют посредник диффузии глицерина, глицеринкиназу и фруктозо-1,6-бисфосфатазу. Ген оперона *glpABC* кодирует субъединицы анаэробной глицерин-3-фосфат дегидрогеназы - А, В и С. Гены оперона *glpTQ* кодируют носитель глицерин-3-фосфата и глицерофосфодиэстеразу. Ген *glpD* кодирует аэробную глицерин-3-фосфат дегидрогеназу. Ген *glpR* кодирует транскрипционный репрессор *GlpR*, *glpE* - тиосульфатсульфуртрансферазу, а *glpG* - сериновую протеазу. Транскрипция генов каждого из *glp*-оперонов контролируется промоторами *pglpFKX*, *pglpABC*, *pglpTQ*, *pglpD* и *pglpEGR*, соответственно, активность которых строго катаболически регулируется (Larson T J, J. et al, (1987); Biol. Chem. 262:15869-74 Zhao N et al (1994) J Bacteriol, 176: 2393-239). Недавно Selivano L. с савт. (Microb Cell Fact 15:28, (2016)) в поисках новых промоторов, подходящих для экспрессии рекомбинантных генов, протестировали *glp*-промотор, *pglpQ*, из *Streptomyces coelicolor* в бактериальной плазмидной системе экспрессии, однако этот промотор продемонстрировал довольно низкую активность и поэтому был исключен из перспективных кандидатов для промышленного применения.

### **Сущность изобретения**

Первый аспект изобретения относится к нуклеотидной конструкции, содержащей синтетическую некодирующую последовательность ДНК (i), которая содержит первый фрагмент ДНК и второй фрагмент ДНК, где первый фрагмент ДНК представляет собой последовательность ДНК, полученную из 5'-нетранслируемой области (5'-UTR) *glp*-гена из *Escherichia coli*, а второй фрагмент ДНК представляет собой ДНК-последовательность CAAGGAGGAAACAGCT (SEQ ID NO: 10) или вариант указанной последовательности, и где первый фрагмент расположен перед вторым фрагментом. В частности, изобретения относятся к нуклеотидной конструкции, содержащей три функционально связанных ДНК-последовательности: промоторную ДНК-последовательность (ii), синтетическую некодирующую ДНК-последовательность, содержащую сайт связывания рибосомы (RBS) (i), и кодирующую ДНК-последовательность (iii). Предпочтительно, первый фрагмент ДНК происходит из ДНК-последовательности 5'UTR генов *glpF*, *glpA* или *glpD* и содержал первые 5-65 последовательных нуклеотидов после сайта инициации

транскрипции в промоторах *glpF*, *glpA* или *glpD*. В одном предпочтительном варианте промоторная ДНК-последовательность (ii) соответствует ДНК-последовательности из промотора *glp*-оперона из *Escherichia coli* (*E. coli*), например из оперонов *glpFKX*, *glpABC*, *glpTQ* или *glpD*.

Во втором аспекте изобретение относится к рекомбинантной клетке, предпочтительно бактериальной рекомбинантной клетке, содержащей нуклеотидную конструкцию по изобретению.

В третьем аспекте изобретение относится к системе экспрессии, содержащей конструкцию по изобретению или рекомбинантную клетку по изобретению.

В четвертом аспекте изобретение относится к способу получения одной или более биологических молекул, например белка, нуклеиновой кислоты, олигосахарида и т.п., с использованием конструкции по изобретению и/или рекомбинантной клетки по изобретению.

Эти и другие аспекты изобретения подробно описаны ниже.

### **Краткое описание чертежей**

На фигуре 1 представлены:

(А) Уровни экспрессии репортерного гена (*lacZ*) с нуклеотидных конструкций, содержащих семнадцать различных промоторных элементов, объединенных с SEQ ID NO: 10. Каждая экспрессионная кассета, содержащая один промотор, также содержит нативную 5'UTR-последовательность (после промотора и перед SEQ ID NO: 10) из гена, который в природе транскрибируется с промотора конструкции. (В) Уровни экспрессии *lacZ* с экспрессионных кассет (А), содержащих выбранные промоторы - *PglpT*, *PglpA* и *PglpF* (заштрихованные столбики; b) по сравнению с уровнями экспрессии *lacZ* с экспрессионных кассет, содержащих промотор и ДНК-фрагмент исходной 5'UTR соответствующего гена, т.е. *glpT*, *glpA* и *glpF* (незаштрихованные столбики; a).

Данные показывают уровень активности экспрессированной  $\beta$ -галактозидазы в клетках-хозяевах. Активность измеряли в единицах Миллера (Ед./OD/мл/мин).

На фигуре 2 представлены:

(А) Схематическое изображение варианта осуществления нуклеотидной конструкции по изобретению;

(В) конструкция нуклеиновой кислоты по изобретению, содержащая ДНК-последовательность промотора *glpF* (SEQ ID NO: 54), синтетическую ДНК (i), содержащую SEQ ID NO: 36 и SEQ ID: 10. Указаны следующие структурные особенности:

- сайт инициации транскрипции в положении +1 выделен полужирным шрифтом;
- сайт связывания РНК-полимеразы, содержащий боксы -10 и -35, выделен полужирным шрифтом;
- четыре операторных сайта, OR1, OR2, OR3 и OR4, для связывания белка-репрессора транскрипции *GlpR* выделены серым цветом;
- два операторных сайта, CRP1 и CRP2, для связывания белка-активатора транскрипции CRP, выделены рамкой;

- синтетическая ДНК-последовательность, включающая 54-нуклеотидный фрагмент 5'UTR гена *glpF* (выделены пунктирной рамкой) (SEQ ID NO: 36) и SEQ ID NO: 10 (подчеркнуты).

На фиг.3 схематически представлена структура *glp*-промоторов:

(A) P*glpA*, (B) P*glpD*, (C) P*glpF* и (D) P*glpT*.

Относительное положение областей -35 и -10, распознаваемых РНК-полимеразой, показано черными прямоугольниками. Относительное положение сайтов связывания сAMP-CRP, *GlpR*, *FNR* и *FIS*, участвующих в транскрипционной регуляции промоторных элементов, указаны маленькими пустыми стрелками; относительное положение 16-нуклеотидного фрагмента 5'UTR, содержащего сайт связывания рибосомы (SD), указано маленькими закрашенными стрелками.

На фиг.4 представлены данные, демонстрирующие уровень экспрессии гена *lacZ*, экспрессируемого в *E. coli*, с одной копии интегрированной в геном экспрессионной кассеты, содержащей:

(a) фрагменты ДНК, соответствующие исходным промоторным последовательностям *glpF*, *glpA* или *glpT*, и ДНК-фрагментам 5'UTR из *glpF*, *glpA* или *glpT*, содержащим (нативный) сайт связывания рибосомы (RSB) (SEQ ID NO: 57, 56 и 55, соответственно) (незаштрихованные столбики);

(b) фрагменты ДНК, содержащие исходные промоторные последовательности *glpF*, *glpA* или *glpT*, и ДНК-фрагменты 5'UTR из исходных *glpF*, *glpA* или *glpT*, в которых отсутствует нативный RBS (SEQ ID NO: 1, 2 и 4, соответственно), но связанные с SEQ ID NO: 10 (заштрихованные столбики);

(c) фрагменты ДНК, содержащие исходные промоторные последовательности *glpF*, *glpA* или *glpT* (SEQ ID NO: 54, 48 и 49, соответственно), (каждая) функционально связанная с 54-нуклеотидным ДНК-фрагментом 5'UTR из *glpF* (SEQ ID NO: 36) и далее связанная с SEQ ID NO: 10 (заштрихованные крестиком столбики).

Данные показывают уровень активности экспрессированной  $\beta$ -галактозидазы в клетках-хозяевах. Активность измеряют в единицах Миллера (Ед./OD/мл/мин).

На фиг.5 представлены данные, демонстрирующие катаболическую репрессию P*glpF* (SEQ ID NO: 12), когда он функционально связан с репортерным геном *lacZ*, интегрированным (в единственной копии) в геном *E. coli*. Активность  $\beta$ -галактозидазы измеряли после роста в различных средах, таких как LB с глюкозой или без нее, минимальная среда, содержащая глицерин, сорбит, мальтозу или глюкозу. Активность  $\beta$ -галактозидазы измеряют в единицах Миллера (единицы ONPG, конвертированные на OD600 на миллилитр в минуту).

На фиг.6 показан уровень экспрессии *lacZ* из интегрированной в геном одной копии экспрессионной кассеты, содержащей либо промоторную последовательность *glpF* или *glpT* (SEQ ID NO: 54, либо SEQ ID NO: 50), 54-нуклеотидный фрагмент *glpF*-5'UTR (SEQ ID NO: 36) и SEQ ID NO: 10 или ее варианты: (A) варианты SEQ ID NO: 10; (B) измерения активности  $\beta$ -галактозидазы (репортерный ген экспрессируется с конструкций,

содержащих 10 вариантов SEQ ID NO: 10, которые имеют модифицированный RBS. Активность  $\beta$ -галактозидазы измеряется в единицах Миллера (единицы ONPG, конвертированные на OD600 на миллилитр в минуту).

На фиг.7 показано влияние на экспрессию *lacZ* (измеренную как уровень активности  $\beta$ -галактозидазы) после модификации области -10 промотора *glpF*. Семь различных конструкций, содержащих промотор *glpF* (SEQ ID NO: 12) и его варианты (P*glpF*\_19, P*glpF*\_20, P*glpF*\_17, P*glpF*\_11, P*glpF*\_13 или P*glpF*\_9), были функционально связаны с *lacZ* и встроены (в единственной копии) в геном *E. coli*, и активность репортерного гена, *lacZ*, оценивали как уровень активности  $\beta$ -галактозидазы, измеренной в единицах Миллера (единицы ONPG, конвертированные на OD600 на миллилитр в минуту). Все конструкции содержат SEQ ID NO: 10.

На фиг.8 показано влияние укорачивания 5'-конца последовательности *glpF* на экспрессию гена *lacZ*. Укороченные варианты промотора *glpF* (SEQ ID NO: 54) (последовательность промотора была укорочена на 15, 140, 165 или 180 пар оснований с 5'-конца) функционально связаны с *lacZ* и экспрессируются с одной копии, встроеной в геном *E. coli*. Все конструкции содержат 54-нуклеотидный фрагмент *glpF*-5'UTR (SEQ ID NO: 36) и SEQ ID NO: 10. Активность  $\beta$ -галактозидазы определяют в единицах Миллера (единицы ONPG, конвертированные на OD600 на миллилитр в минуту).

На фиг.9 показано влияние на уровень активности  $\beta$ -галактозидазы после разрушения гена *glpR* в клетке-хозяине с конструкции, содержащей *lacZ*, функционально связанный с промотором *glpF* (SEQ ID NO: 54, связанная с 54-нуклеотидным фрагментом *glpF*-5' UTR (SEQ ID NO: 36) и SEQ ID NO: 10. Активность  $\beta$ -галактозидазы измеряют в клетках, экспрессирующих белок-репрессор транскрипции, *GlpR* (то есть содержащих нативный ген *glpR*), и в клетках, где ген был разрушен введением *kanR* (т.е. в клетки с низкой экспрессией или без экспрессии *GlpR*). Активность  $\beta$ -галактозидазы определяют в единицах Миллера (единицы ONPG, конвертированные на OD600 на миллилитр в минуту).

На фиг.10 показаны результаты экспрессии гетерологичных генов в *E. coli*, экспрессируемых с ДНК-конструкций либо под *lac*-промотором (P*lac*; серые кружки), либо *glpF*-промотором (P*glpF*; черные кружки), вставленными в виде единичных копий в геном *E. coli*. Все конструкции содержат SEQ ID NO: 10 и нативные 5'UTR-фрагменты без RBS из соответствующих генов.

(А) Продукция 6'-сиалиллактозы (6'SL), оцененная в неочищенных экстрактах клеток, экспрессирующих  $\alpha$ -2,6-сиалилтрансферазу Pd2 (из *Photobacterium damsela* JT0160).

(В) Продукция 3'-сиаллилактозы (3'SL), оцененная в неочищенных экстрактах клеток, экспрессирующих  $\alpha$ -2,3-сиалилтрансферазу NST (из *Neisseria meningitidis* MC58).

Активность сиалилтрансферазы измеряется как продукция в мМ в час.

На фиг.11 показаны уровни экспрессии *lacZ* под контролем промоторов *lac*, *glpF*, *glpA* или *glpT* с мультикопийной плазмиды, содержащей одну копию соответствующей

экспрессионной кассеты. Каждая из экспрессионных кассет содержит SEQ ID NO: 10 и нативные 5'UTR-фрагменты без RBS из соответствующих генов (т.е. *glpF*, *glpA*, *glpT* или *lacZ*, соответственно). Активность  $\beta$ -галактозидазы измеряют после роста в различных средах, такие как LB с глюкозой или без нее (затемненные или окрашенные столбики, соответственно) в единицах Миллера (единицы ONPG, конвертированные на OD600 на миллилитр в минуту).

На фиг.12 показаны результаты продукции LNnT в рекомбинантных *E. coli*, экспрессирующих гетерологичные гены *galT* и *lgtA*, под контролем либо промотора *Plac*, либо *PglpF* (обе экспрессионные кассеты содержат SEQ ID NO: 10 и нативные 5'UTR-фрагменты без RBS из *lacZ* или *glpF*, соответственно). А) MDO1 экспрессирует *galT* и *lgtA* с *Plac* в плазмиде с высоким и средним числом копий соответственно. MP1497 экспрессирует *lgtA* с одной хромосомной копии гена с использованием *PglpF*, и *galT* в плазмиде с высоким числом копий с использованием *Plac*. MP1499 экспрессирует *galT* с одной копии хромосомного гена с использованием *PglpF*, и *lgtA* в плазмиде со средним числом копий с использованием *Plac*. В) MP2622 и MP166 экспрессируют *lgtA* и *galT* с одной или трех интегрированных в хромосому копий генов, соответственно, с использованием *Plac*; MP1825 экспрессирует *lgtA* и *lgtA* с одиночных хромосомных копий с использованием *PglpF*.

На фиг.13 показаны результаты продукции LNT в рекомбинантных *E. coli*, экспрессирующих гетерологичные гены *galTK* и *lgtA* под контролем либо *Plac*, либо *PglpF* (обе экспрессионные кассеты содержат SEQ ID NO: 10 и лишены RBS нативные 5'UTR-фрагменты из *lacZ* или *glpF*, соответственно). А) MD015 экспрессирует *galTK* и *lgtA* под контролем *Plac* в плазмидах с высоким и средним числом копий соответственно. MP1498 экспрессирует *lgtA* с одной встроенной в хромосому копии гена с использованием *PglpF*, и *galTK* в высококопийной плазмиде с использованием *Plac*. MP1655 экспрессирует *galTK* с двух хромосомных копий гена с использованием *PglpF*, и *lgtA* в плазмиде со средним числом копий с использованием *Plac*. В) MP245 экспрессирует 3 и 2 встроенные в хромосому копии генов *lgtA* и *galTK*, соответственно, с использованием *Plac*; MP1920 экспрессирует *lgtA* и *galTK* с одиночных встроенных в хромосому копий генов с использованием *PglpF*.

На фиг.14 показаны результаты продукции LNFP-I в рекомбинантных *E. coli*, экспрессирующих гетерологичные гены *galTK*, *lgtA* и *futC* под контролем *PglpF*. MP2239 и MP2374 экспрессируют *lgtA*, *galTK* и *futC* с одиночных встроенных в хромосому копий генов с использованием *PglpF*. Кроме того, MP2374 содержит дополнительную копию генов колановой кислоты *gmd*, *wcaJ* (*fcI*), *wcaH* (*gmm*), *wcaI*, *cpsB* (*manC*) и *cpsG* (*manB*), которые все экспрессируются под контролем *PglpF*. Экспрессионные кассеты содержат SEQ ID NO: 10 и нативный лишенный RBS 5'UTR-фрагмент из *glpF*.

На фиг.15 показаны результаты продукции 3'SL в рекомбинантных *E. coli*, экспрессирующих гетерологичные гены *nst*, *neuA*, *neuB* и *neuC*. MAP425 экспрессирует 2 хромосомно интегрированные копии *nst*, а также *neuA*, *neuB* и *neuC* с высококопийной

плазмиды с использованием Plac. MAP1214 экспрессирует *nst*, *neuA*, *neuB* и *neuC* с одной хромосомной копии гена с использованием P<sub>glpF</sub>. Экспрессионные кассеты содержат SEQ ID NO: 10 и нативные лишённые RBS 5'UTR-фрагменты из *lacZ* или *glpF*, соответственно.

На фиг.16 показаны результаты продукции 6'SL в модифицированных *E. coli*, экспрессирующих гетерологичные гены Pd2, *neuA*, *neuB* и *neuC*. MAP265 экспрессирует одну хромосомную копию Pd2, а также *neuA*, *neuB* и *neuC* с высококопийной плазмиды с использованием Plac. MAP1200 экспрессирует Pd2, *neuA*, *neuB* и *neuC* с одной хромосомной копии гена с использованием P<sub>glpF</sub>. Экспрессионные кассеты содержат SEQ ID NO: 10 и нативные лишённые RBS 5'UTR-фрагменты из *lacZ* или *glpF*, соответственно.

На фиг.17 показаны результаты продукции 2'FL в рекомбинантных *E. coli*, экспрессирующих гетерологичный ген *futC*: штамм FT18 содержит две плазмиды, экспрессирующие *futC*, и гены колановой кислоты *gmd*, *fcl*, *manC* и *manB* под контролем Plac, штамм MAP965 содержит одиночную копию *futC* и гены колановой кислоты: *gmd*, *wcaJ* (*fcl*), *wcaH* (*gmm*), *wcaI*, *cpsB* (*manC*) и *cpsG* (*manB*), которые экспрессируются под контролем P<sub>glpF</sub>. Экспрессионные кассеты содержат SEQ ID NO: 10 и нативные 5'UTR-фрагменты, лишённые RBS, из *lacZ* или *glpF*, соответственно.

### **Подробное описание изобретения**

Настоящее изобретение в целом относится к ДНК-конструкциям и экспрессионным системам, пригодным для рекомбинантного получения биологических молекул. В частности, настоящее изобретение относится к рекомбинантным бактериальным экспрессионным системам, способным обеспечить стабильную и исключительно высокую экспрессию гена, который функционально связан с промотором и синтетической некодирующей ДНК-последовательностью, расположенной выше гена, где указанная синтетическая ДНК-последовательность (взаимозаменяемо обозначаемая в настоящем описании как «синтетическая/искусственная/рекомбинантная ДНК-последовательность (i)»), содержит фрагмент 5'-нетранслируемой лидерной ДНК-последовательности (5'UTR-ДНК) гена *glp* из *Escherichia coli* (*E. coli*) и ДНК-последовательность CAAGGAGGAAACAGCT (SEQ ID NO: 10) или ее вариант. Последовательность CAAGGAGGAAACAGCT (SEQ ID NO: 10) представляет собой искусственную ДНК-последовательность, которая первоначально получена из 5'UTR *lacZ* из *E. coli* и была модифицирована в последовательности сайта связывания рибосомы (RBS). Ранее последовательность была описана в связи с ее способностью усиливать экспрессию репортерного гена (*lacZ*) примерно в 6 раз в модельной системе экспрессии генов с использованием нуклеотидной конструкции, где эта последовательность была функционально связана с искусственным промотором и с 30-нуклеотидной ДНК-последовательностью, способной стабилизировать мРНК. (Meunial-Salles I, et al (2005) *Appl Environ Microbiol* 71:2140-2144; WO 03/089605). Авторы изобретения неожиданно обнаружили, что SEQ ID NO: 10 или ее вариант, когда она не связана с описанной РНК-

стабилизирующей последовательностью ДНК, не способна усиливать экспрессию репортерного гена (*lacZ*) из всех случайно выбранных промоторов, а только из нескольких из них. Однако, если SEQ ID NO: 10 связана с фрагментом 5'UTR-ДНК гена *glp*, преимущественно, гена *glpF*, *glpA*, *glpT* или *glpD*, экспрессия репортерного гена значительно увеличивается, и уровень экспрессии гена не зависит или намного менее зависит от выбора промотора, т.е. сила промотора оказывает меньшее влияние на уровень экспрессии гена, когда синтетическая последовательность ДНК по изобретению вставлена между промотором и геном.

Соответственно, первый аспект изобретения относится к синтетической некодирующей ДНК-последовательности (i), содержащей первый фрагмент ДНК и второй фрагмент ДНК, где первый фрагмент представляет собой фрагмент 5'UTR-ДНК гена *glp*, преимущественно, гена *glpF*, *glpA*, *glpT* или *glpD*, а второй фрагмент представляет собой SEQ ID NO: 10 или ее вариант, и где второй фрагмент расположен ниже первого фрагмента (т.е. второй фрагмент связан с концом первого фрагмента). Предпочтительно, синтетическая некодирующая ДНК-последовательность (i) является частью нуклеотидной конструкции, где она функционально связана с последовательностью промоторной ДНК (ii) и, необязательно, с кодирующей последовательностью ДНК (iii), и где указанная синтетическая последовательность ДНК расположена ниже последовательности промоторной ДНК (ii) и, необязательно, выше последовательности кодирующей ДНК (iii). Термин «необязательно» в настоящем контексте означает, что в некоторых вариантах осуществления изобретение относится к нуклеотидным конструкциям, которые включают промоторную ДНК-последовательность (ii) и синтетическую ДНК-последовательность (i), но не кодирующую ДНК (iii). В других вариантах осуществления синтетическая ДНК-последовательность (i) может быть функционально связана с кодирующей последовательностью ДНК (iii), а промоторная ДНК не включена в конструкцию. Еще, в других вариантах осуществления конструкция может содержать синтетическую ДНК-последовательность (i) и не содержать ни промоторную ДНК, ни кодирующую ДНК-последовательность. Нуклеотидные конструкции, содержащие синтетическую последовательность ДНК (i), могут быть встроены в геном клетки-хозяина выше гена и ниже нативного промотора гена, например, заменяя существующую нативную ДНК-последовательность, или она может быть вставлена в геномную ДНК для замены любой из или обеих из промоторной и/или 5'UTR ДНК-последовательностей у представляющего интерес геномного гена. Конструкция по изобретению также может быть использована для модификации экспрессии представляющего интерес гена желаемым образом (т.е. для увеличения или уменьшения экспрессии гена) по сравнению с природной экспрессией гена, контролируемой нативными (не искусственно модифицированными) регуляторными последовательностями геномной ДНК. Как указано, в одном варианте осуществления конструкция может содержать только искусственную ДНК-последовательность (i), т.е. не иметь ни промоторной, ни кодирующей ДНК-последовательности, так как такая конструкция может быть вставлена в геном клетки-хозяина ниже последовательности

геномного промотора и выше последовательности гена/кодирующей последовательности (замена существующей нативной последовательности или добавление/удлинение существующей последовательности), и за счет этого изменяется экспрессия гена (увеличивается или уменьшается) по сравнению с его природной экспрессией с соответствующего геномного промотора. В других вариантах осуществления конструкция по изобретению может содержать функционально связанные промоторную ДНК-последовательность (ii), последовательность синтетической ДНК (i) и кодирующую ДНК-последовательность (iii), которая кодирует гетерологичную или гомологичную (по отношению к клетке-хозяину) биологическую молекулу, где синтетическая ДНК-последовательность расположена между промоторной ДНК и кодирующими последовательностями ДНК.

Если не указано иное, то все технические и научные термины, используемые в данном документе, имеют то же значение, которое обычно понимают специалисты в области техники, к которой относится это изобретение. Singleton et al. (1994) Dictionary of Microbiology and Molecular Biology, second edition, John Wiley and Sons (New York) дадут специалисту общий словарь многих терминов, используемых в данном изобретении. Хотя любые способы и материалы, подобные или эквивалентные тем, которые описаны в настоящем документе, могут использоваться при практическом применении или при проверке настоящего изобретения, описаны предпочтительные способы и материалы. Большинство номенклатуры и общих лабораторных методик, требуемых в этой заявке, можно найти в Sambrook et al., Molecular Cloning: A Laboratory Manual, Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York (2012); Wilson K. and Walker J., Principles and Techniques of Biochemistry and Molecular Biology (2010), Cambridge University Press; или в Maniatis et al., Molecular Cloning A laboratory Manual, Cold Spring Harbor Laboratory (2012); или в Ausubel et al., Current protocols in molecular biology, John Wiley and Sons (2010). В дальнейшем руководства называются «Sambrook et al.», «Wilson & Walker», «Maniatis et al», «Ausubel et al», соответственно.

Если не указано иное, то термины, определенные в описании, относятся ко всем аспектам и вариантам осуществления изобретения. Все варианты осуществления, описанные в описании и рабочих примерах, относятся ко всем любым аспектам изобретения.

Используемый в настоящем документе термин «нуклеиновая кислота» включает молекулы РНК, ДНК и кДНК. Понятно, что в результате вырожденности генетического кода может быть получено множество нуклеотидных последовательностей, кодирующих данный белок. Термин нуклеиновая кислота используется взаимозаменяемо с термином «полинуклеотид». «Олигонуклеотид» представляет собой молекулу нуклеиновой кислоты с короткой цепью. «Праймер» представляет собой олигонуклеотид, встречающийся в природе, например, в очищенном продукте рестрикции, или произведенный синтетическим путем, который способен выступать в качестве точки начала синтеза, когда находится в условиях, при которых происходит синтез продукта удлинения праймера,

который комплементарен цепи нуклеиновой кислоты (т.е. в присутствии нуклеотидов и индуцирующего агента, такого как ДНК-полимераза, и при подходящей температуре и рН). Праймер предпочтительно является одноцепочечным для максимальной эффективности амплификации, но в альтернативном варианте может быть двухцепочечным. Если праймер является двухцепочечным, то его сначала обрабатывают, чтобы разделить цепи, прежде чем использовать для получения продуктов удлинения цепи. Предпочтительно, чтобы праймер представлял собой дезоксирибонуклеотид. Праймер должен быть достаточно длинным, чтобы обеспечить синтез продуктов удлинения в присутствии индуцирующего агента. Точная длина праймеров будет зависеть от многих факторов, включая температуру, источник праймера и применение способа.

«Конструкция нуклеиновой кислоты» означает искусственно сконструированный сегмент нуклеиновой кислоты, в частности сегмент ДНК, который предназначен для «трансплантации» в клетку-мишень, например бактериальную клетку, для модификации экспрессии геномного гена или для экспрессии последовательности гена/кодирующей ДНК, которые могут быть включены в конструкцию. В контексте изобретения конструкция нуклеиновой кислоты содержит последовательность рекомбинантной ДНК, по существу состоящую из, необязательно, одной, двух или трех отдельных последовательностей ДНК: последовательности синтетической некодирующей ДНК (i), содержащей сайт связывания рибосомы (RBS), последовательности промоторной ДНК (ii) и последовательности кодирующей ДНК (iii). В вариантах осуществления, относящихся к конструкции, содержащей две или три из перечисленных последовательностей, последовательности в конструкции функционально связаны друг с другом. «Функционально связанный» определяется в настоящем документе как конфигурация, в которой регуляторная последовательность, т.е. промоторная последовательность и/или последовательность 5'UTR, надлежащим образом размещены относительно последовательности кодирующей ДНК, так что регуляторные последовательности направляют транскрипцию кодирующей последовательности и трансляцию мРНК в полипептидную последовательность, кодируемую кодирующей ДНК. В варианте осуществления, где конструкция содержит последовательность кодирующей ДНК, предпочтительно, чтобы кодирующая ДНК кодировала по меньшей мере одну молекулу белка или РНК, которая обладает активностью, которая прямо или косвенно участвует в продукции одного или более НМО в клетке-хозяине (т.е. активность является существенной или полезной для продуцирования одного или более НМО). Неограничивающими примерами такой активности могут быть: ферментативная активность, регулирующая экспрессию генов активность, шаперонная активность. Неограничивающие примеры последовательностей кодирующей ДНК (iii) описаны ниже и в рабочих примерах. Конструкция ДНК по изобретению в некоторых вариантах осуществления называется экспрессионной кассетой по изобретению. Конструкции ДНК/экспрессионные кассеты по изобретению в некоторых вариантах осуществления могут содержать более одной кодирующей ДНК-последовательности, которая может

кодировать разные биологические молекулы. Предпочтительно, чтобы конструкции (содержащие одну или более кодирующих последовательностей ДНК (iii)) содержали одну копию последовательности промоторной ДНК (ii) и одну копию последовательности синтетической ДНК (i). ДНК-конструкции по настоящему изобретению могут быть вставлены в плазмидную ДНК/вектор, перенесены в клетку-мишень/хозяина и экспрессированы с плазмид и/или с хромосом. Конструкции ДНК могут быть линейными или кольцевыми. Линейная или кольцевая конструкция ДНК, встроенная в бактериальный геном или экспрессионную плазмиду, взаимозаменяемо обозначается в данном документе как «экспрессионная кассета», «экспрессионный картридж» или «картридж». В одном варианте осуществления картридж представляет собой линейную конструкцию ДНК, содержащую три последовательности ДНК: промоторную (ДНК-последовательность (ii)), последовательность синтетической ДНК (i) ниже промотора, и кодирующую ДНК-последовательность (последовательность iii), кодирующую биологическую молекулу, представляющую интерес. Конструкция также может содержать дополнительные последовательности, например последовательность терминатора транскрипции, и две терминально фланкирующие области, которые гомологичны геномной области, и которые обеспечивают гомологичную рекомбинацию, и/или другие последовательности, описанные в настоящем документе. Картридж может быть получен способами, хорошо известными в данной области техники, например, с использованием стандартных способов, описанных в Wilson & Walker. Использование линейного экспрессионного картриджа может обеспечить преимущество, заключающееся в том, что сайт встраивания в геном может быть свободно выбран с помощью соответствующего подбора фланкирующих гомологичных областей картриджа. Таким образом, встраивание линейного экспрессионного картриджа обеспечивает большую вариабельность в отношении области генома. Линейные картриджи включены в предпочтительные варианты осуществления изобретения.

Под термином «сайт связывания рибосомы» (RBS) подразумевается короткая нуклеотидная последовательность, обычно включающая около 4-16 нуклеиновых оснований, которая функционирует путем позиционирования рибосомы на молекуле мРНК для трансляции кодируемого белка. Сайт «модифицированного связывания рибосомы» представляет собой сайт связывания рибосомы, в котором одна или несколько пар оснований были изменены. Согласно изобретению синтетическая некодирующая ДНК-последовательность (i) содержит RBS в SEQ ID NO: 10 или ее вариант. Варианты SEQ ID NO: 12 могут использоваться в другом варианте осуществления, например, с целью модификации экспрессии представляющей интерес последовательности геномной ДНК, как, например, авторы показывают в настоящем документе, что некоторые варианты могут усиливать экспрессию гена в еще большей степени по сравнению с достигнутой экспрессией с использованием SEQ ID NO: 10, тогда как другие варианты могут снижать экспрессию гена. Полезные, но не ограничивающие варианты осуществления последовательностей ДНК, которые включают RBS, можно найти в приведенной ниже

таблице 1 и описаны в спецификации.

Для целей данного изобретения «промотор» или «промоторная область» или «промоторный элемент» представляет собой нуклеотидную последовательность, которая распознается и связывается ДНК-зависимой РНК-полимеразой во время инициации транскрипции. Промотор вместе с другими транскрипционными и трансляционными регуляторными нуклеотидными последовательностями (также называемыми «регуляторными последовательностями») необходим для экспрессии данного гена или группы генов (оперона). В общем, транскрипционные и трансляционные регуляторные последовательности включают, но не ограничиваются ими, промоторные последовательности, сайты связывания рибосомы, последовательности инициации и терминации транскрипции, последовательности инициации и терминации трансляции и энхансерные или активаторные последовательности. «Сайт начала транскрипции» означает первый нуклеотид, который должен быть транскрибирован, и обозначается +1. Нуклеотиды вниз от сайта начала пронумерованы как +2, +3, +4 и т.д., а нуклеотиды в противоположном 5'-направлении (вверх) пронумерованы -1, -2, -3 и т.д. Промотором по изобретению является отдельную последовательность ДНК. Термин «отдельная ДНК-последовательность» означает, что последовательности представляют собой не интегрированный фрагмент геномной ДНК, а искусственный/клонированный фрагмент ДНК, который идентичен или гомологичен последовательности геномной ДНК; следуя этому определению, промоторная ДНК конструкции/экспрессионной кассеты, описанная в настоящем документе, рассматривается как «полученная» из последовательности геномной ДНК, содержащейся в промоторной области гена. Последовательность промоторной ДНК конструкции по изобретению может происходить из промоторной области любого гена из генома выбранного вида, предпочтительно, из промоторной области геномной ДНК *E. coli*. Согласно изобретению, любая последовательность промоторной ДНК, которая способна связываться с РНК-полимеразой и иницировать транскрипцию, является подходящей для осуществления изобретения. Как уже было указано, нуклеотидная последовательность промоторной ДНК конструкции может быть идентичной или иметь определенный процент идентичности, такой как примерно 65-70% идентичности, предпочтительно, по меньшей мере 80% идентичности, предпочтительно, от примерно 90% до примерно 99,9% идентичности с нуклеотидной последовательностью фрагмента последовательности геномной ДНК, предпочтительно, последовательности бактериальной геномной ДНК, которая рассматривается как промоторная область одного гена или оперона, например, *glp* или *lac*-оперонов *E. coli*. Термины «примерно», «около» и «приблизительно» используются взаимозаменяемо и означают отклонение от указанного значения на 1-10% или незначительное отклонение, которое не влияет на соответствующий признак. Под «опероном» подразумевается функционирующая единица геномной ДНК, содержащая кластер генов под контролем одного промотора. Под «*glp*-опероном» подразумевается группа генов, участвующих в дыхательном метаболизме глицерина бактерий. Изобретение в предпочтительных вариантах осуществления

относится к четырем *glp*-оперонам *E. coli*, в частности, *glpFKX*, *glpABC*, *glpTQ* и *glpD*. Промоторы указанных оперонов указаны в настоящем документе как промоторы *glpF*, *glpA*, *glpT* и *glpD* и сокращены в настоящем документе как *PglpF*, *PglpA*, *PglpT* и *PglpD* соответственно. В некоторых других вариантах осуществления изобретение относится к *lac*-оперону *E. coli*, содержащему гены *Z*, *Y* и *A* и их промотор *lac* (сокращенно обозначаемый в настоящем документе как *Plac*). Предпочтительно, чтобы промоторная последовательность оперона *glp*, содержащаяся в промоторной ДНК конструкции по изобретению, была идентична или имела по меньшей мере 80% идентичности, предпочтительно, 90-99,9% идентичности с нуклеотидной последовательностью фрагмента геномной ДНК *E. coli*, расположенной перед последовательностями имеющими идентификаторы (ID) Genetic Bank: EG10396 (*glpF*), EG10391 (*glpA*), EG10394 (*glpD*), EG10401 (*glpT*), EG10527 (*lacZ*). Геном *E. coli* в настоящем документе относится к полной последовательности геномной ДНК *E. coli* K-12 MG1655 (ID GenBank: U00096.3). Избранные, но не ограниченные варианты осуществления последовательностей промоторной ДНК по изобретению можно найти в приведенной ниже таблице 1 и описаны в заявке.

Последовательность промоторной ДНК (ii) может содержать несколько структурных признаков/элементов, таких как регуляторные области, способные влиять (облегчать или ингибировать) связывание РНК-полимеразы в клетке и инициировать транскрипцию расположенной ниже (в 3'-направлении) кодирующей последовательности, таких как, например, сайты связывания белков-регуляторов транскрипции, таких как, например, транскрипционный репрессор, белок *GlpR*, или активатор транскрипции - белок *CRP*. Регуляторная область включает белок-связывающие домены (консенсусные последовательности), ответственные за связывание РНК-полимеразы, такие как бокс -35 и бокс -10 (бокс Прибнова). Все упомянутые регуляторные последовательности промоторной ДНК конструкции могут иметь определенный процент идентичности с соответствующими геномными последовательностями промотора, т.е. изобретение предусматривает исходные ДНК-последовательности (нативного/дикого типа) или их варианты. Некоторые неограничивающие полезные варианты осуществления вариантов последовательностей промоторной ДНК (ii) можно найти в приведенной ниже Таблице 1 и описаны в спецификации.

Промоторная последовательность по изобретению, предпочтительно, содержит по меньшей мере 50 нуклеотидов, более предпочтительно, по меньшей мере 60 нуклеотидов, например, от примерно 65 до примерно 100, от примерно 75 до примерно 115 нуклеотидов, от примерно 85 до примерно 125, например от 90 до 115, 110-120, 120-130, 130-140, 140-150 или более 150 нуклеотидов, например, 155-165, 165-175, 175-185, 185-195, 195-205, 205-215, 215-225, 225-235, 235-245 245-255, 255-265, 250-350. В некотором варианте промоторная последовательность может иметь длину до 500-1000 нуклеотидов. В некоторых вариантах осуществления выбранная промоторная последовательность также может быть короче, то есть включать менее 50 нуклеотидов. Длина последовательности

промоторной ДНК не является общим ограничивающим фактором изобретения, так как изобретение в другом варианте осуществления предусматривает любую последовательность промоторной ДНК (i), которая способна связываться с РНК-полимеразой и инициировать транскрипцию гена в экспрессионной кассете или представляющего интерес гена в геноме. В одном предпочтительном варианте осуществления промоторная ДНК получена из геномной промоторной области оперона *glp*, например, из промоторной области *glpF* (*PglpF*), промоторной области *glpA* (*PglpA*), промоторной области *glpT* (*PglpT*), промоторной области *glpD* (*PglpD*). В некоторых предпочтительных вариантах осуществления промоторная ДНК-последовательность (i) может быть выбрана из любой из ДНК-последовательностей, указанных в настоящем документе как SEQ ID NO: 48, 49 или 54; в некоторых вариантах осуществления, вариант указанной последовательности может быть предпочтительным. Неограничивающие полезные примеры таких вариантов описаны в данном описании и приведены в таблице 1. В одном предпочтительном варианте осуществления последовательность промоторной ДНК имеет свое происхождение из геномной ДНК *E. coli* (GenBank ID EG10396), в частности из области геномной ДНК, содержащей промоторный элемент оперона *glpFХК* (в частности, 50-350-нуклеотидную ДНК-последовательность выше начала транскрипции оперона *glpFХК* или его вариант, который имеет по меньшей мере 90% идентичности по последовательности). Одним предпочтительным вариантом осуществления промоторной ДНК-последовательности (i) является SEQ ID NO: 54.

В различных вариантах осуществления изобретение может относиться к промотору, который является индуцибельным/регулируемым или постоянно активным. Термин «индуцибельный промотор» означает, что активность промотора, то есть способность промотора инициировать и поддерживать транскрипцию функционально связанного гена на определенном уровне, регулируется внешним фактором, например молекулой источника углерода. В некоторых вариантах осуществления активность промотора по изобретению можно контролировать с помощью присутствия или отсутствия молекулы источника углерода в среде, например глицерина, глюкозы, арабинозы и т.д. «Источник углерода» в целом относится к углеводу, который может поглощаться и метаболизироваться бактериальной клеткой. Предпочтительно, индуцибельный промотор по изобретению представляет собой индуцируемый источником углерода *glp*-промотор *E. coli*, предпочтительно промотор *glpF*, *glpA*, *glpD*, *glpT*, или его вариант, который индуцируется тем же источником углерода, что и соответствующий исходный промотор. Предпочтительно, промотор содержит по меньшей мере один сайт связывания с белком рецептора циклического АМФ (цАМФ-CRP). В других вариантах осуществления изобретение относится к промотору, который не индуцируется, то есть к активности промотора, которая не зависит от индукции, например, источником углерода. Предпочтительно, последний промотор представляет собой *glp*-промотор по изобретению, структура ДНК которого была модифицирована, чтобы сделать активность промотора независимой от источника углерода или сделать промотор конститутивно активным,

например, путем удаления/модификации сайтов для связывания GlpR в промоторной последовательности.

Как указано выше, изобретение также относится к вариантам последовательностей промоторной ДНК, которые могут быть использованы в конструкциях по изобретению. Под «вариантом» в настоящем контексте подразумевается последовательность искусственной нуклеиновой кислоты, которая предпочтительно имеет сходство примерно на 70-99% с нуклеотидной последовательностью рассматриваемой последовательности промоторной ДНК. Процент сходства сравниваемых последовательностей нуклеиновых кислот указывает на долю последовательностей, которые имеют идентичную структуру, то есть идентичный нуклеотидный состав. Процент сходства последовательностей для целей изобретения может быть определен с использованием любого метода, хорошо известного в данной области, например с помощью BLAST. Объем термина «вариант» включает нуклеотидные последовательности, комплементарные последовательностям ДНК, описанным в настоящем документе, последовательности мРНК и синтетические нуклеотидные последовательности, например, праймеры для ПЦР, и другие олигонуклеотиды, которые родственны нуклеотидным последовательностям конструкций по изобретению.

Согласно изобретению, конструкция/экспрессионная кассета содержит синтетическую некодирующую ДНК-последовательность (i), которая содержит сайт связывания рибосомы (RBS). Термин «синтетическая ДНК-последовательность» в настоящем контексте означает рукотворную ДНК-последовательность, т.е. ДНК-последовательность сконструирована искусственно и состоит из по меньшей мере двух фрагментов ДНК, причем по меньшей мере один из фрагментов получен из геномной ДНК (т.е. она соответствует ДНК-последовательности геномной ДНК), а другой фрагмент ДНК представляет собой искусственную ДНК-последовательность, содержащую около 16 нуклеиновых оснований, которая, предпочтительно, не соответствует последовательности природной бактериальной геномной ДНК, содержащей RBS. В частности, синтетическая ДНК-последовательность (i) состоит из двух фрагментов ДНК: первого фрагмента ДНК и второго фрагмента ДНК. Первый фрагмент ДНК имеет примерно 70-100% идентичности по последовательности с фрагментом последовательности геномной ДНК, полученной из нетранслируемой ДНК-последовательности, расположенной ниже точки инициации транскрипции и выше начала трансляции гена, например в гене *glp*. Предпочтительно, чтобы первый фрагмент содержал цепочку из по меньшей мере от 5 до 80 непрерывных нуклеотидов, например, 5-10 нуклеотидов, 10-15 нуклеотидов, 15-20 нуклеотидов, 20-30 нуклеотидов, 30-40 нуклеотидов, 40-50 нуклеотидов, 50-60 нуклеотидов после точки инициации транскрипции (начиная с нуклеотида +2) гена *glp*, предпочтительно, гена *glpF*, *glpA* или *glpD*. Последовательность ДНК первого фрагмента может быть гомологичной или гетерологичной по отношению к промотору ДНК (ii) и/или последовательности кодирующей ДНК (iii). «Гомологичный» в данном контексте означает, что ДНК-последовательность первого фрагмента получена из 5'UTR-области гена, который в

природе (в геноме исходного вида) расположен ниже промотора конструкции, или в природе она является встречающимся в природе фрагментом 5'UTR гена конструкции, или она в природе ассоциирована с обоими; «гетерологичный» означает, что первый фрагмент ДНК не соответствует фрагменту природной (геномной) 5'UTR-области, ассоциированной с геном или промотором конструкции. В одном предпочтительном варианте осуществления первый фрагмент ДНК имеет свое происхождение из геномной ДНК-последовательности 5'UTR гена *glp*, предпочтительно гена *glpF*. Предпочтительно, чтобы она была гетерологична по отношению к кодирующей ДНК-последовательности (iii). В некоторых вариантах осуществления может быть предпочтительным, чтобы ДНК-последовательность первого фрагмента ДНК была гомологичной относительно ДНК-последовательности промотора (ii); в других вариантах осуществления может быть предпочтительным, чтобы указанная первая последовательность была гетерологичной по отношению к ДНК-последовательности промотора (i). В одном предпочтительном варианте осуществления первый фрагмент ДНК представляет собой или содержит нуклеотидную последовательность, представленную в SEQ ID NO: 36.

Второй фрагмент последовательности искусственной ДНК (i) представляет собой нуклеотидную последовательность CAAGGAGGAAACAGCT (SEQ ID NO: 10) или вариант указанной последовательности. Некоторыми полезными неограничивающими вариантами осуществления варианта SEQ ID NO: 10 являются нуклеотидные последовательности SEQ ID NO: 38-47. В некоторых вариантах осуществления конструкции по изобретению, первый фрагмент ДНК (ДНК-последовательность 5'-UTR) расположен ниже ДНК-последовательности промотора (ДНК-последовательности (ii)) (и выше второго фрагмента ДНК); и второй фрагмент ДНК расположен выше кодирующей ДНК-последовательности (iii) (т.е. до точки инициации трансляции), то есть в искусственной ДНК-последовательности (i) первый фрагмент ДНК предшествует второму фрагменту ДНК. Одним предпочтительным вариантом осуществления последовательности синтетической ДНК (i) является SEQ ID NO: 37.

Неограничивающие варианты последовательностей ДНК (i) и (ii), фрагментов, вариантов и их комбинаций, применимых в различных аспектах изобретения, описаны в таблице 1.

Таблица 1

Название:	SEQ ID:	Последовательность (5'→3')	Описание
<i>PglpF_5'U</i> TR- <i>glpF</i> - d16nb	SEQ ID NO: 1	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA	284- нуклеотидный фрагмент ДНК, полученный из геномной ДНК <i>E.</i> <i>coli</i> (реф.послед.

		GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTTTCACGCATACAACAAACATTAAC	ID U00096.3), расположенный на 16 нуклеотидов выше кодона инициации трансляции гена glpF
<i>PglpA_5'U</i> TR-glpA- d16nb	SEQ ID NO: 2	GAAAACATTCATAAATTAAATGTGAATT GCCGCACACATTATTAATAAGATTTAC AAAATGTTCAAAATGACGCATGAAATCA CGTTTCACTTTCGAATTATGAGCGAATA TGCGCGAAATCAAACAATTCATGTTTTT ACTATGGCTAAATGGTAAAAAACGAA	166- нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный на 16 нуклеотидов выше кодона инициации трансляции гена glpA
<i>PglpD_5'U</i> TR-glpD- d16 nb	SEQ ID NO: 3	TGCGTCTCTCTTTCTTTACAAACAAGTGG GCAAATTTACCGCACAGTTTACGTCGAA GCGGCAGATAAACGCCATAATGTTATAC ATATCACTCTAAAATGTTTTTTCAATGTT ACCTAAAGCGCGATTCTTTGCTAATATG TTCGATAACGAACATTTATGAGCTTTAA CGAA	174- нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный на 16 нуклеотидов выше кодона инициации трансляции гена

			glpD
PglpT_ 5'UTR- glpT-d16nb	SEQ ID NO: 4	CCATTTAGCCATAGTAAAAACATGAATT GTTTGATTTTCGCGCATATTCGCTCATAAT TCGAAAGTGAAACGTGATTTTCATGCGTC ATTTTGAACATTTTGTAATCTTATTTAA TAATGTGTGCGGCAATTCACATTTAATTT ATGAATGTTTTCTTAACATCGCGGCAAC TCAAGAAACGGCAGGTTCTCTCACTGAA TCAGGCTGTTAATCATAAATAAGACCAC GG	229- нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный на 16 нуклеотидов выше кодона инициации трансляции гена glpT
16nb -glpF	SEQ ID NO: 5	TCTTCAGGATCCGATT	16-нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный непосредственно выше кодона инициации трансляции гена glpF
16nb -glpA	SEQ ID NO: 6	CTTCAGAGGGATAACA	16-нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный непосредственно

			выше кодона инициации трансляции гена glpA
16nb-glpD	SEQ ID NO: 7	AGTGAATGAGGGCAGC	16-нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный выше кодона инициации трансляции гена glpD
16nb -glpT	SEQ ID NO: 8	GCCACGGAGGCTATCA	16 -нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный непосредственно выше кодона инициации трансляции гена glpT
16bnb-lacZ	SEQ ID NO: 9	CACACAGGAAACAGCT	16-нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный выше lacZ

mut16bp - lacZ (recRBS)	SEQ ID NO:10	CAAGGAGGAAACAGCT	Вариант SEQ ID NO: 9) (CAC -> AGG )
Plac_org	SEQ ID NO:11	TGTGAGTTAGCTCACTCATTAGGCACCC CAGGCTTTACACTTTATGCTTCCGGCTCG TATGTTGTGTGGAATTGTGAGCGGATAA CAATTTACACACAGGAAACAGCT	107- нуклеотидный фрагмент ДНК, расположенный выше lacZ, полученного из геномной ДНК E. coli (реф.послед. ID U00096.3); элемент промотора lac- оперона
PglpF_54n b 5'URT- glpF_recRB S (PglpF_rec )	SEQ ID NO:12	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTTTCACGCATACAACAACATTA CCAAGGAGGAAACAGCT	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO: 10
PglpF_SD1	SEQ ID NO:13	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO: 38 (см. фигуру 6)

		TGACTTTCACGCATACAACAAACATTA CCAAATTCGAAACAGCT	
PglpF_SD2	SEQ ID NO:14	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTTTCACGCATACAACAAACATTA CCAAGCGCAAAACAGCT	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO: 39 (см. фигуру 6)
PglpF_SD3	SEQ ID NO:15	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTTTCACGCATACAACAAACATTA CCAAGAACAAAACAGCT	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO: 40 (см. фигуру 6)
PglpF_SD4	SEQ ID NO:16	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTTTCACGCATACAACAAACATTA CCAAGTAGGAAACAGCT	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO:41 (см. фигуру 6)

PglpF_SD5	SEQ ID NO:17	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTIONCACGCATACAACAACATTA CCAACCGAGAAACAGCT	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO:42 (см. фигуру 6)
PglpF_SD6	SEQ ID NO:18	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTIONCACGCATACAACAACATTA CCAAGAGCTAAACAGCT	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO:43 (см. фигуру 6)
PglpF_SD7	SEQ ID NO:19	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTIONCACGCATACAACAACATTA CCAAGAGCAAAACAGCT	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO:44 (см. фигуру 7)4
PglpF_SD8	SEQ ID NO:20	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT	300- нуклеотидный

		AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTTTCACGCATACAACAAACATTA CCAAGAGAAAAACAGCT	фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO:45 (см. фигуру 7)
PglpF_SD9	SEQ ID NO:21	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTTTCACGCATACAACAAACATTA CCAAAGGAAAAACAGCT	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO:46 (см. фигуру 7)
PglpF_SD1 0	SEQ ID NO:22	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTTTCACGCATACAACAAACATTA CCAAGGAAAAACAGCT	300- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 1 и SEQ ID NO: 47 (см. фигуру 7)
PglpF_9	SEQ ID NO:23	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAAACA	Вариант SEQ ID NO:12, содержащий модификацию

		TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATTTAATTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTIONCACGCATACAACAAACATTA CCAAGGAGGAAACAGCT	области -10 (см. фигуру 7)
PglpF_11	SEQ ID NO:24	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATCAGAATAC AGCATGCCTACAAGCATCGTGGAGGTCC GTGACTTTCACGCATACAACAAACATTA ACCAAGGAGGAAACAGCT	Вариант SEQ ID NO:12, содержащий модификацию области -10 (см. фигуру 7)
PglpF_13	SEQ ID NO:25	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATATCCTTCT ACAGCATGCCTACAAGCATCGTGGAGGT CCGTGACTTTCACGCATACAACAAACAT TAACCAAGGAGGAAACAGCT	Вариант SEQ ID NO:12, содержащий модификацию области -10 (см. фигуру 7)
PglpF_17	SEQ ID NO:26	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA	Вариант SEQ ID NO:12, содержащий модификацию области -10 (см. фигуру 7)

		GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAATGATAC AGCATGCCTACAAGCATCGTGGAGGTCC GTGACTTTCACGCATACAACAAACATTA ACCAAGGAGGAAACAGCT	
PglpF_19	SEQ ID NO:27	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATGAAGCTAC AGCATGCCTACAAGCATCGTGGAGGTCC GTGACTTTCACGCATACAACAAACATTA ACCAAGGAGGAAACAGCT	Вариант SEQ ID NO:12, содержащий модификацию области -10 (см. фигуру 7)
PglpF_20	SEQ ID NO:28	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGGCGCGATAACGCTCATTTTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATCAGTATACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTIONTTCACGCATACAACAAACATTA CCAAGGAGGAAACAGCT	Вариант SEQ ID NO:12, содержащий модификацию области -10 (см. фигуру 7)
D15PglpF	SEQ ID NO:29	GATTACGGTTTGCCACACTTTTCATCCTT CTCCTGGTGACATAATCCACATCAATCG AAAATGTTAATAAATTTGTTGCGCGAAT GATCTAACAACATGCATCATGTACAAT CAGATGGAATAAATGGCGCGATAACGCT CATTTTATGACGAGGCACACACATTTTA AGTTCGATATTTCTCGTTTTTGCTCGTTA ACGATAAGTTTACAGCATGCCTACAAGC	285- нуклеотидный фрагмент ДНК of SEQ ID NO: 12

		ATCGTGGAGGTCCGTGACTTTCACGCAT ACAACAAACATTAACCAAGGAGGAAAC AGCT	
D140PglpF	SEQ ID NO:30	ATGGCGCGATAACGCTCATTTTATGACG AGGCACACACATTTTAAGTTCGATATTT CTCGTTTTTGCTCGTTAACGATAAGTTTA CAGCATGCCTACAAGCATCGTGGAGGTC CGTGACTTTCACGCATACAACAAACATT AACCAAGGAGGAAACAGCT	160- нуклеотидный фрагмент ДНК of SEQ ID NO: 12
D165PglpF	SEQ ID NO:31	ACGAGGCACACACATTTTAAGTTCGATA TTTCTCGTTTTTGCTCGTTAACGATAAGT TTACAGCATGC CTACAAGCATCGTGGAGGTCCGTGACTT TCACGCATACAACAAACATTAACCAAGG AGGAAACAGCT	135- нуклеотидный фрагмент ДНК of SEQ ID NO: 12
D180PglpF	SEQ ID NO:32	TTTAAGTTCGATATTTCTCGTTTTTGCTC GTTAACGATAAGTTTACAGCATGCCTAC AAGCATCG TGGAGGTCCGTGACTTTCACGCATACAA CAAACATTAACCAAGGAGGAAACAGCT	120- нуклеотидный фрагмент ДНК of SEQ ID NO: 12
PglpA_ 5'UTR- glpA_recRB S (rec PglpA)	SEQ ID NO:33	GAAAACATTCATAAATTAATGTGAATT GCCGCACACATTATTAATAAGATTTAC AAAATGTTCAAAATGACGCATGAAATCA CGTTTCACTTTCGAATTATGAGCGAATA TGCGCGAAATCAAACAATTCATGTTTTT ACTATGGCTAAATGGTAAAAAACGAAC AAGGAGGAAACAGCT	182- нуклеотидный фрагмент ДНК comprising SEQ ID NO: 2 and SEQ ID NO: 10
PglpD_5'U TR- glpD_recR BS (rec PglpD)	SEQ ID NO:34	TGCGTCTCTCTTTCTTTACAAACAAGTGG GCAAATTTACCGCACAGTTTACGTCGAA GCGGCAGATAAACGCCATAATGTTATAC ATATCACTCTAAAATGTTTTTTCAATGTT ACCTAAAGCGCGATTCTTTGCTAATATG TTCGATAACGAACATTTATGAGCTTTAA CGAACAAGGAGGAAACAGCT	190- нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 3 и SEQ ID NO: 10
PglpT_5'U	SEQ ID	CCATTTAGCCATAGTAAAAACATGAATT	245-

TR- glpT_recRBS (rec PglpT)	NO:35	GTTTGATTTTCGCGCATATTCGCTCATAAT TCGAAAGTGAAACGTGATTTTCATGCGTC ATTTTGAACATTTTGTAATCTTATTTAA TAATGTGTGCGGCAATTCACATTTAATTT ATGAATGTTTTCTTAACATCGCGGCAAC TCAAGAAACGGCAGGTTCTCTCACTGAA TCAGGCTGTTAATCATAAATAAGACCAC GGCAAGGAGGAAACAGCT	нуклеотидный фрагмент ДНК, содержащий SEQ ID NO: 4 и SEQ ID NO: 10
54nb 5'UTR- glpF	SEQ ID NO:36	TGCCTACAAGCATCGTGGAGGTCCGTGA CTTTCACGCATACAACAACATTAAC	54-нуклеотидный фрагмент ДНК о 5'UTR-glpF, расположенный ниже сайта инициации транскрипции и на 16 нуклеотидов выше кодона инициации трансляции
synDNA(i) (70UTR)	SEQ ID NO:37	TGCCTACAAGCATCGTGGAGGTCCGTGA CTTTCACGCATACAACAACATTAACCA AGGAGGAAACAGCT	70-нуклеотидная синтетическая некодирующая последовательнос ть ДНК (ii), содержащая SEQ ID NO: 36 и SEQ ID NO: 10
recRBS_v1 (SD1)	SEQ ID NO:38	CAAATTCGAAACAGCT	Вариант SEQ ID NO: 10 (см. фиг.6, A)
recRBS_v2 (SD2)	SEQ ID NO:39	CAAGCGCAAAACAGCT	Вариант SEQ ID NO: 10 (см. фиг.6, A)
recRBS_v3	SEQ ID	CAAGAACAAAACAGCT	Вариант SEQ ID

(SD3)	NO:40		NO: 10 (см. фиг.6, A)
recRBS_v4 (SD4)	SEQ ID NO:41	CAACTAGGAAACAGCT	Вариант SEQ ID NO: 10 (см. фиг.6, A)
recRBS_v5 (SD5)	SEQ ID NO:42	CAACCGAGAAACAGCT	Вариант SEQ ID NO: 10 (см. фиг.6, A)
recRBS_v6 (SD6)	SEQ ID NO:43	CAAGAGCTAAACAGCT	Вариант SEQ ID NO: 10 (см. фиг.6, A)
recRBS_v7 (SD7)	SEQ ID NO:44	CAAGAGCAAAACAGCT	Вариант SEQ ID NO: 10 (см. фиг.6, A)
recRBS_v8 (SD8)	SEQ ID NO:45	CAAGAGAAAAACAGCT	Вариант SEQ ID NO: 10 (см. фиг.6, A)
recRBS_v9 (SD9)	SEQ ID NO:46	CAAAGGAAAAACAGCT	Вариант SEQ ID NO: 10 (см. фиг.6, A)
recRBS_v1 0 (SD10)	SEQ ID NO:47	CAACTGAGAAACAGCT	Вариант SEQ ID NO: 10 (см. фиг.6, A)
PglpA_org	SEQ ID NO:48	GAAAACATTCATAAATTAAATGTGAATT GCCGCACACATTATTAATAAGATTTAC AAAATGTTCAAAATGACGCATGAAATCA CGTTTCACTTTCGAATTATGAGCGAATA TGCGCGA	119- нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный выше точки инициации транскрипции

			glpA; элемент промотора glpA
PglpT_org	SEQ ID NO:49	CCATTTAGCCATAGTAAAAACATGAATT GTTTGATTTTCGCGCATATTCGCTCATAAT TCGAAAGTGAAACGTGATTTTCATGCGTC ATTTTGAACATTTTGTAATCTTATTTAA TAATGTGTGCGGCAATTCACATTTAATTT ATGAATGTTTTCTTAACATCGCGGCA	169-нуклеотидный фрагмент ДНК, полученный из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенный выше точки инициации транскрипции glpT; элемент промотора glpT
PglpA_70U TR	SEQ ID NO:50	GAAAACATTCATAAATTAATGTGAATT GCCGCACACATTATTAATAAGATTTAC AAAATGTTCAAAATGACGCATGAAATCA CGTTTCACTTTCGAATTATGAGCGAATA TGCGCGATGCCTACAAGCATCGTGGAGG TCCGTGACTTTCACGCATACAACAACA TTAACCAAGGAGGAAACAGCT	189-нуклеотидный фрагмент ДНК, полученный путем объединения SEQ ID NO: 49 с SEQ ID NO 37
PglpT_70U TR	SEQ ID NO: 51	CCATTTAGCCATAGTAAAAACATGAATT GTTTGATTTTCGCGCATATTCGCTCATAAT TCGAAAGTGAAACGTGATTTTCATGCGTC ATTTTGAACATTTTGTAATCTTATTTAA TAATGTGTGCGGCAATTCACATTTAATTT ATGAATGTTTTCTTAACATCGCGGCA TGCCTACAAGCATCGTGGAGGTCCGTGA CTTTCACGCATACAACAACATTAACCA AGGAGGAAACAGCT	239-нуклеотидный фрагмент ДНК, полученный путем объединения SEQ ID NO: 50 с синтетической некодирующей последовательностью ДНК (ii) (SEQ ID NO 37)

PglpT_70U TR_SD4	SEQ ID NO:52	CCATTTAGCCATAGTAAAAACATGAATT GTTTGATTTTCGCGCATATTCGCTCATAAT TCGAAAGTGAAACGTGATTTTCATGCGTC ATTTTGAACATTTTGTAATCTTATTTAA TAATGTGTGCGGCAATTCACATTTAATTT ATGAATGTTTTCTTAACATCGCGGCA TGCCTACAAGCATCGTGGAGGTCCGTGA CTTTCACGCATACAACAACATTAACCA ACTAGGAAACAGCT	239- нуклеотидный фрагмент ДНК SEQ ID NO: 50 где 16 п.о., расположенные выше сайта инициации трансляции, идентичны SEQ ID NO 41
PglpT_70U TR_SD9	SEQ ID NO:53	CCATTTAGCCATAGTAAAAACATGAATT GTTTGATTTTCGCGCATATTCGCTCATAAT TCGAAAGTGAAACGTGATTTTCATGCGTC ATTTTGAACATTTTGTAATCTTATTTAA TAATGTGTGCGGCAATTCACATTTAATTT ATGAATGTTTTCTTAACATCGCGGCA TGCCTACAAGCATCGTGGAGGTCCGTGA CTTTCACGCATACAACAACATTAACCA AAGGAAAAACAGCT	239- нуклеотидный фрагмент ДНК SEQ ID NO: 50, где 16 п.о., расположенные выше сайта инициации трансляции, идентичны SEQ ID NO 46
PglpF_org	SEQ ID NO:54	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGCGCGATAACGCTCATTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCA	230-нуклеотидная последовательнос ть ДНК, полученная из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенная выше точки инициации транскрипции glpF; элемент промотора glpF

PglpT- 5'UTR- glpT_org	SEQ ID NO:55	CCATTTAGCCATAGTAAAAACATGAATT GTTTGATTTTCGCGCATATTCGCTCATAAT TCGAAAGTGAAACGTGATTTTCATGCGTC ATTTTGAACATTTTGTAAATCTTATTTAA TAATGTGTGCGGCAATTCACATTTAATTT ATGAATGTTTTCTTAACATCGCGGCAAC TCAAGAAACGGCAGGTTCTCTCACTGAA TCAGGCTGTTAATCATAAATAAGACCAC GGGCCACGGAGGCTATCA	245-нуклеотидная последовательность ДНК, полученная из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенная выше сайта инициации трансляции glpT
PglpA- 5'UTR- glpA_org	SEQ ID NO:56	GAAAACATTCATAAATTAATGTGAATT GCCGCACACATTATTAATAAGATTTAC AAAATGTTCAAAATGACGCATGAAATCA CGTTTCACTTTCGAATTATGAGCGAATA TGCGCGAAATCAAACAATTCATGTTTTT ACTATGGCTAAATGGTAAAAAACGAA CTTCAGAGGGATAACA	182-нуклеотидная последовательность ДНК, полученная из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенная выше сайта инициации трансляции glpA
PglpF - 5'UTR- glpF_org	SEQ ID NO:57	GCGGCACGCCTTGCAGATTACGGTTTGC CACACTTTTCATCCTTCTCCTGGTGACAT AATCCACATCAATCGAAAATGTTAATAA ATTTGTTGCGCGAATGATCTAACAACA TGCATCATGTACAATCAGATGGAATAAA TGCGCGATAACGCTCATTATGACGA GGCACACACATTTTAAGTTCGATATTTCT CGTTTTTGCTCGTTAACGATAAGTTTACA GCATGCCTACAAGCATCGTGGAGGTCCG TGACTTTCACGCATACAACAACATTA CTCTTCAGGATCCGATT	300-нуклеотидная последовательность ДНК, полученная из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенная выше сайта инициации трансляции glpF
PglpD - 5'UTR-	SEQ ID NO:10	TGCGTCTCTCTTTCTTTACAACAAGTGG GCAAATTTACCGCACAGTTTACGTCGAA	190-нуклеотидная последовательность

glpD_org	4	GCGGCAGATAAACGCCATAATGTTATAC ATATCACTCTAAAATGTTTTTCAATGTT ACCTAAAGCGCGATTCTTTGCTAATATG TTCGATAACGAACATTTATGAGCTTTAA CGAA AGTGAATGAGGGCAGC	ть ДНК, полученная из геномной ДНК E. coli (реф.послед. ID U00096.3), расположенная выше точки инициации трансляции glpD
----------	---	---	--

Как указано выше, в некоторых предпочтительных вариантах осуществления конструкция включает три функционально связанных ДНК-последовательности: промоторную ДНК-последовательность (ii), последовательность синтетической некодирующей ДНК (i) и по меньшей мере одну последовательность кодирующей ДНК (iii) (ген). Кодирующая ДНК-последовательность (iii) представляет собой выделенную ДНК-последовательность, которая приблизительно на 70-100% идентична последовательности фрагмента геномной ДНК, содержащей ген, кодирующий биологическую молекулу, например белок или РНК. Кодирующая ДНК (iii) конструкции может быть гомологичной или гетерологичной промоторной ДНК-последовательности (ii). «Гетерологичный» в настоящем контексте означает, что экспрессия соответствующей геномной кодирующей ДНК-последовательности обычно контролируется другим промотором, а не промотором конструкции. Соответственно, «гомологичный» в настоящем контексте означает, что соответствующие геномные последовательности промоторной последовательности ДНК (ii) и кодирующей последовательности ДНК (iii) естественным образом связаны в геноме исходного вида.

Под термином «кодирующая нуклеотидная последовательность» подразумевается нуклеотидная последовательность, которая содержит набор последовательных неперекрывающихся триплетов (кодонов), которые транскрибируются в мРНК и транслируются в полипептид, когда находятся под контролем соответствующих контрольных последовательностей, т.е. промотора. Границы кодирующей последовательности обычно определяются сайтом связывания рибосомы, расположенным непосредственно перед открытой рамкой считывания на 5'-конце мРНК, стартовым кодоном транскрипции (AUG, GUG или UUG) и стоп-кодом трансляции (UAA, UGA или UAG). Кодирующая последовательность может включать, но не ограничивается ими, геномную ДНК, кДНК, синтетические и рекомбинантные нуклеотидные последовательности.

В предпочтительном варианте кодирующая нуклеотидная последовательность конструкции по изобретению является гетерологичной по отношению к промотору и, в некоторых вариантах, также по отношению к первому фрагменту ДНК некодирующей ДНК-последовательности (i) в конструкции. Тем не менее, в отношении клетки-хозяина, в

которой должна экспрессироваться кодирующая ДНК, указанная ДНК может быть либо гетерологичной (то есть полученной из другого биологического вида или рода), либо гомологичной (то есть полученной из клетки-хозяина). Например, в одном варианте кодирующая ДНК-последовательность конструкции может кодировать биологическую молекулу, например белок, который является чужеродным для хозяина, то есть нуклеотидная последовательность кодирующей ДНК является гетерологичной по отношению к виду-хозяину, поскольку она имеет свое происхождение из вида-донора, который отличается от организма хозяина, или нуклеотидная последовательность кодирующей ДНК содержит модификацию, которая приводит к экспрессии полипептида, который не идентичен полипептиду, экспрессирующемуся с соответствующей немодифицированной ДНК-последовательности хозяина, т.е. искусственно модифицированная кодирующая последовательность ДНК, исходно полученная из хозяина, рассматривается в настоящем контексте как гетерологичная. В случае, если хозяином является конкретный прокариотический вид, гетерологичная нуклеотидная последовательность может иметь свое происхождение из другого рода семейства, другого порядка или класса, другого типа (отдела), или другого домена (надцарства) организмов. Гетерологичная нуклеотидная последовательность, происходящая от донора, отличного от хозяина, может быть модифицирована перед ее введением в клетку-хозяина мутациями, вставками, делециями или заменами отдельных нуклеиновых кислот или части гетерологичной нуклеотидной последовательности до тех пор, пока такие модифицированные последовательности обладают той же функцией (функционально эквивалентные), что и эталонная последовательность. Гетерологичная нуклеотидная последовательность, как указано в настоящем документе, охватывает также нуклеотидные последовательности, происходящие из другого домена (надцарства) организмов, таких как эукариоты (эукариотическое происхождение), такие как, например, ферменты, участвующие в синтезе или деградации олигосахаридов грудного молока (НМО). Тем не менее, в других вариантах осуществления изобретения кодирующая нуклеиновая кислота может быть гомологичной по отношению к клетке-хозяину. Термин «гомологичная нуклеотидная последовательность» (синонимично используемый в настоящем документе как «нуклеотидная последовательность, нативная для хозяина» или «нуклеотидная последовательность, полученная из хозяина») в данном контексте означает, что нуклеотидная последовательность происходит (или получена) из того же самого организма, или того же рода семейства, или того же порядка или класса, того же типа (отдела) или того же домена (надцарства) организмов, что и организм-хозяин. В одном варианте осуществления кодирующая ДНК конструкции, описанной в настоящем документе, может кодировать фермент или белок-переносчик сахара, которые обычно экспрессируются бактериальной клеткой-хозяином, которая естественным образом содержит в своем геноме гены, кодирующие указанный белок-переносчик фермента или сахара.

В общем, изобретение предусматривает любую кодирующую ДНК, поскольку

любая кодирующая ДНК может быть включена в конструкцию по изобретению и транскрибирована с промотора, включенного в конструкцию. В некоторых предпочтительных вариантах кодирующая ДНК кодирует белок, например фермент, транспортный белок, регуляторный белок, шаперон и т.д. Термин «белок» взаимозаменяемо указывается в данном документе как «полипептид». В других предпочтительных вариантах кодирующая ДНК может кодировать регуляторную (некодирующую) молекулу РНК (нкРНК), например, такую как функционально важные типы некодирующих РНК, таких как трансферные РНК (тРНК) и рибосомные РНК (рРНК), а также малые РНК, такие как микроРНК, миРНК и длинные нкРНК. В предпочтительном варианте осуществления по меньшей мере одна кодирующая ДНК конструкции по изобретению кодирует белок или РНК, связанные с синтезом, деградацией или транспортом олигосахаридов человеческого молока, их предшественников или производных. «По меньшей мере, одна кодирующая ДНК-последовательность» означает, что конструкция в разных вариантах осуществления может содержать более одной кодирующей ДНК-последовательности, например две кодирующие последовательности, такие как первая и вторая кодирующая последовательность; три кодирующие последовательности, такие как первая, вторая и третья кодирующие последовательности и т.д. Предпочтительно, чтобы множественные кодирующие ДНК-последовательности в этих вариантах осуществления экспрессировались в виде тандема, и транскрипция контролировалась одной копией промоторной ДНК (ii) конструкции. Первая, вторая, третья и т.д. кодирующие ДНК-последовательности могут в разных вариантах осуществления кодировать различные ферменты или другие белки, функция которых является существенной или полезной для продукции НМО клеткой-хозяином, например ферменты, транспортерные белки, регуляторные белки, шапероны и т.д. Под «существенным» в данном контексте подразумевается, что белок непосредственно участвует в синтезе НМО, например, он представляет собой фермент, который помогает процессу получения НМО из предшественника НМО, например, фермент с активностью глюкозилтрансферазы. Под «полезным» в данном контексте подразумевается, что белок не участвует непосредственно в синтезе НМО, но он способствует процессу, который является полезным для продукции НМО клеткой-хозяином, например это белок, который способствует транспорту (внутри или наружу из клетки-хозяина) НМО или предшественника НМО. Некоторые неограничивающие варианты белков, которые рассматриваются в настоящем документе как существенные для получения одного или более НМО клеткой-хозяином, можно найти в таблице 2, и белки, которые рассматриваются как полезные для получения одного или более НМО клеткой-хозяином приведены в таблице 3 ниже.

Таблица 2

Ген	Последовательность, ID	Описание	Пример НМО
-----	------------------------	----------	------------

	(GenBank)		
lgtA	WP_002248149.1	$\beta$ -1,3-N- ацетилглюкозаминилтранс фераза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP- V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
galT	NP_207619.1	$\beta$ -1,4- галактозилтрансфераза	LNnT, LNFP-III, LNFP-VI, pLNH, F-pLNH I, pLNnH
cpsIBJ	AB050723	$\beta$ -1,3- галактозилтрансфераза	LNT, LNFP-I, LNFP-II, LNFP-V, LNDFH-I, LNDFH-II, pLNH, F-pLNH I
MAMA_R 764	AGC02224.1	$\alpha$ -1,3-фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP- VI, LNDFH-II, F-pLNH I
Mg791	AEQ33441.1	$\alpha$ -1,3-фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP- VI, LNDFH-II, F-pLNH I
Moumou_ 00703	AGC02224.1	$\alpha$ -1,3-фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP- VI, LNDFH-II, F-pLNH I
futA	NP_207177.1	$\alpha$ -1,3-фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP- VI, LNDFH-II, F-pLNH I
futC	CP003904	$\alpha$ -1,2-фукозилтрансфераза	2'FL, DFL, LNFP-I, LNDFH-I
fucT	AAB81031.1	$\alpha$ -1,3-фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP- VI, LNDFH-II, F-pLNH I
fucTIII	AY450598	$\alpha$ -1,4-фукозилтрансфераза	LNDFH-I, LNDFH-II
fucTa	AF194963	$\alpha$ -1,3/4- фукозилтрансфераза	LNFP-II, LNDFH-I, LNDFH-II
Pd2,6ST	BAA25316.1	$\alpha$ -2,6-сиалилтрансфераза	6'SL

PspST6	BAF92026.1	$\alpha$ -2,6-сиалилтрансфераза	6'SL
PiST6_14 5	BAF91416.1	$\alpha$ -2,6-сиалилтрансфераза	6'SL
PiST6_11 9	BAI49484.1	$\alpha$ -2,6-сиалилтрансфераза	6'SL
NST	AAC44541.1	$\alpha$ -2,3-сиалилтрансфераза	3'SL
neuA	AF400048	CMP-Neu5 Ас синтетаза	3'SL, 6'SL
neuB	AF400048	Синтаза сиаловой кислоты	3'SL, 6'SL, сиаловая кислота
neuC	AF400048	GlcNAc-6-фосфат-2-эпимераза	3'SL, 6'SL, сиаловая кислота

Таблица 3

Ген	Последовательность, ID (UniProt)	Описание	Продукты НМО, описание
gmd	P0AC88	GDP-манноза-4,6-дегидратаза	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
wcaG	P32055	GDP-фукозосинтаза	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
wcaH	P32056	GDP-манноза-маннозилгидролаза	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
cpsB	P24174	Манноза-1-фосфат гуанилилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-II,

			LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
cpsG	P24175	Фосфоманномутаза	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
glmS	P17169	L-глутамин-D-фруктоза-6-фосфат аминотрансфераза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
glmU	P0ACC7	Слитые N-ацетилглюкозамин-1-фосфатуридилтрансфераза и глюкозамин-1-фосфатацетилтрансфераза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
glmM	P31120	Фосфоглюкозамин-мутаза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
ampG	P0AE16	Муропептид: H <sup>+</sup> симпортер	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I,

			LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
nagA	P0AF18	N-ацетилглюкозамин-6-фосфат-деацетилаза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
nagK	P75959	N-ацетил-D-глюкозаминакиназа	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
nagZ	P75949	$\beta$ -N-ацетилгексозаминидаза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
phop	P23836	ДНК-связывающий транскрипционный двойной регулятор PhoP	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
glnA	P0A9C5	Глютаминсинтетаза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I,

			LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
ppk	P0A7B1	Полифосфаткиназа	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
pykA	P21599	пируваткиназа II	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
pgm	P36938	Фосфоглюкомутаза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
galU	P0AEP3	UTP - глюкозо-1-Фосфат-уридилтрансфераза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH
galE	P09147	UDP-глюкозо-4-эпимераза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH

nagC	P0AF20	ДНК-связывающий транскрипционный регулятор двойной NagC	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, 3'SL, 6'SL
glk	P0A6V8	Глюкокиназа	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
pfkB	P06999	6-фосфофруктокиназа II	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
gpt	P0A9M5	Ксантин-гуанин фосфорибозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
gmk	P60546	Гуанилаткиназа	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
ndk	P0A763	Нуклеозид-дифосфаткиназа	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-

			I, LNDFH-II, F-pLNH I
zwf	P0AC53	НАДФ <sup>+</sup> -зависимая глюкозо-6-фосфатдегидрогеназа	2'FL, 3'FL, DFL, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I
galF	P0AAB6	УТР:глюкозо-1-фосфат-уридилтрансфераза	LNT, LNnT, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH

Термин «олигосахарид грудного молока» или «НМО» в данном контексте означает сложный углевод, содержащийся в грудном молоке человека (см. Urashima et al.: *Milk Oligosaccharides*. Nova Science Publisher (2011); или Chen, *Adv. Carbohydr. Chem. Biochem.* **72**, 113 (2015)). НМО имеют базовую структуру, содержащую лактозную единицу на восстанавливаемом конце, которая может быть удлинена одним или несколькими  $\beta$ -N-ацетиллактозаминилами и/или одним или несколькими  $\beta$ -лакто-N-биозильными звеньями, и эта базовая структура может быть замещена  $\alpha$ -L-фукопиранозильной и/или  $\alpha$ -N-ацетилнейраминильной (сиалильной) группой. В этом отношении неокислотные (или нейтральные) НМО лишены сиалильного остатка, а кислотные НМО имеют по меньшей мере один сиалильный остаток в своей структуре. Неокислотные (или нейтральные) НМО могут быть фукозилированными или нефукозилированными. Примеры таких нейтральных нефукозилированных НМО включают лакто-N-тетраозу (LNT), лакто-N-неотетраозу (LNnT), лакто-N-неогексаозу (LNnH), пара-лакто-N-неогексаозу (pLNnH), пара-лакто-N-гексаозу (pLNH) и лакто-N-гексаозу (LNH). Примеры нейтральных фукозилированных НМО включают 2'-фукозиллактозу (2'-FL), лакто-N-фукопентаозу I (LNFP-I), лакто-N-дифукогексаозу I (LNDFH-I), 3-фукозиллактозу (3-FL), дифукозиллактозу (DFL), лакто-N-фукопентазу II (LNFP-II), лакто-N-фукопентаозу III (LNFP-II I), лакто-N-дифукогексаозу III (LNDFH-III), фукозил-лакто-N-гексаозу II (FLNH-II), лакто-N-фукопентазу V (LNFP-V), лакто-N-дифукогексаозу II (LNDFH-II), фукозил-лакто-N-гексаозу I (FLNH-I), фукозил-пара-лакто-N-гексаозу I (FpLNH-I), фукозил-пара-лакто-N-неогексаоза II (FpLNnH II) и фукозил-лакто-N-неогексаозу (FLNnH). Примеры кислых НМО включают 3'-сиалиллактозу (3'-SL), 6'-сиалиллактозу (6'-SL), 3-фукозил-3'-сиалиллактозу (FSL), 3'-О-сиалиллакто-N-тетраозу а (LST a), фукозил-LST a (FLST a), 6'-О-сиалиллакто-N-тетраозу b (LST b), фукозил-LST b (FLST b), 6'-О-сиалиллакто-N-неотетраозу (LST c), фукозил-

LST с (FLST с), 3'-О-сиалиллакто-N-неотетраозу (LST d), фукозил-LST d (FLST d), сиалил-лакто-N-гексаозу (SLNH), сиалил-лакто-N-неогексаозу I (SLNH-I), сиалил-лакто-N-неогексаозу II (SLNH-II) и дисиалил-лакто-N-тетраозу (DSLNT). В контексте настоящего изобретения лактозу рассматривают как представителя НМО.

Термин «предшественник НМО» в настоящем контексте относится к соединению, участвующему в пути биосинтеза одной или более НМО по изобретению, которые продуцируются и естественным образом присутствуют в клетке-хозяине или импортируются в клетку из внеклеточной среды. Некоторые неограничивающие примеры предшественников НМО перечислены ниже:

Предшественник:	Продукт:
UDP-GlcNAc	LNT, LN <sub>n</sub> T, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, 3'SL, 6'SL, pLNnH, (F)LSTa, (F)LSTb, (F)LSTc, (F)LSTd
UDP-Gal	LNT, LN <sub>n</sub> T, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, pLNH, F-pLNH I, pLNnH, LSTa, LSTb, LSTc, LSTd
GDP-фукоза	LNT, LN <sub>n</sub> T, LNFP-I, LNFP-II, LNFP-III, LNFP-V, LNFP-VI, LNDFH-I, LNDFH-II, F-pLNH I, 2'FL, 3'FL, DFL, FLSTa, FLSTb, FLSTc, FLSTd

Термин «транспортёр НМО» означает биологическую молекулу, например белок, которая облегчает транспортировку/экспорт НМО, синтезированного клеткой-хозяином, через клеточную мембрану, например, в клеточную среду, или транспортировку/импорт НМО из клеточной среды в клеточный цитозоль.

Термин «производное НМО» означает молекулу, которая получена из молекулы НМО или содержит группу НМО, например молекулу ганглиозида, структуру искусственного углевода/белка, содержащую группу НМО.

Экспрессионная кассета по настоящему изобретению может быть использована для рекомбинантной продукции одной или более НМО либо интегрированной в геном, либо в составе плазмиды, или, в некоторых вариантах, клетка-хозяин может содержать как интегрированную в геном, так и плазмидную экспрессирующую кассету, где по меньшей мере одна или обе экспрессионных кассеты содержат один или несколько генов, которые необходимы и/или полезны для продуцирования одного или более НМО, и где экспрессия по меньшей мере одного из указанных генов находится под контролем промотора *g<sub>l</sub>p* по изобретению (т.е. *P<sub>g<sub>l</sub>pF</sub>*, *P<sub>g<sub>l</sub>pA</sub>*, *P<sub>g<sub>l</sub>pD</sub>* или *P<sub>g<sub>l</sub>pT</sub>*, предпочтительно, *P<sub>g<sub>l</sub>pF</sub>*). Предпочтительно, чтобы интегрированная в геном кассета содержала по меньшей мере одну (или первый набор) кодирующих последовательностей ДНК, а плазмидная кассета содержала по меньшей мере одну вторую кодирующую ДНК (или второй набор кодирующих последовательностей ДНК), где по меньшей мере одна первая и/или по меньшей мере одна вторая кодирующие последовательности ДНК функционально

связаны с промотором *glp* по изобретению. В некоторых предпочтительных вариантах осуществления по меньшей мере одна из экспрессионных кассет экспрессируется под контролем *Pglp*, например, кодирующая последовательность интегрированной в геном кассеты функционально связана с промотором *glp* по изобретению, например, *PglpF*, а плазмидная кодирующая последовательность функционально связана с другим промотором, например *lac*-промотором, или другим промотором. В некоторых вариантах осуществления, как интегрированные в геном, так и плазмидные кассеты, могут экспрессироваться под контролем одного и того же или другого промотора *glp* по изобретению, например, промотор интегрированной в геном кассеты представляет собой *PglpF*, а плазмидный промотор представляет собой *PglpA*. В других вариантах осуществления все экспрессионные кассеты, содержащиеся в клетке-хозяине, могут содержать один и тот же промотор *glp*. В одном предпочтительном варианте осуществления клетка-хозяин содержит по меньшей мере одну копию интегрированной в геном экспрессионной кассеты по изобретению, содержащей *PglpF*. Предпочтительно, чтобы геном клетки-хозяина содержал одну или небольшое количество копий интегрированной в геном экспрессионной кассеты, например, две или три копии. Тем не менее, в некоторых вариантах осуществления хозяин может содержать несколько копий экспрессионной плазмиды, где каждая плазида содержит одну копию экспрессионной кассеты по изобретению. В некоторых вариантах осуществления клетка-хозяин может содержать несколько различных нуклеотидных конструкций по изобретению, как интегрированных в геном, так и/или плазмидных. Каждая из нескольких различных нуклеотидных конструкций может быть встроена в геном клетки-хозяина или в плазмиду в единственной копии или во множестве копий. В некоторых вариантах осуществления предпочтительно, чтобы конструкции встраивались в одной копии или с низким числом копий.

В соответствии с изобретением одна копия экспрессионной кассеты по изобретению, содержащаяся в клетке-хозяине, либо интегрированная в геном, либо в плазмиде, может обеспечить количество биологических молекул, кодируемых кодирующей последовательностью ДНК (ii) (предпочтительно, под контролем промотора *glp*, например, *PglpF*), которое достаточно для обеспечения высокого уровня продукции одного или более НМО клеткой-хозяином. Удивительно, что одна интегрированная в геном копия экспрессионной кассеты по настоящему изобретению может обеспечить уровни продукции НМО, которые сопоставимы или выше (например, в 2-10 раз выше), чем уровни продукции, достигнутые с использованием экспрессии с той же кассеты в высококопийной плазмиде (100-500 копий). В некоторых вариантах осуществления может быть выгодно экспрессировать два или несколько генов, связанных с продукцией НМО, в клетке-хозяине. Гены, связанные с НМО, могут быть включены в одну конструкцию и экспрессированы в виде тандема с одной (или множества) копий в геноме или с плазмиды; или гены могут быть включены в различные конструкции по изобретению, и один ген экспрессируется с интегрированной в геном кассеты, а другой ген - с плазмиды. В других

вариантах осуществления может быть предусмотрен другой способ экспрессии, состав или количество копий экспрессионных кассет. Предпочтительно, чтобы по меньшей мере один ген, включенный в вышеуказанные экспрессионные кассеты, кодирует белок с ферментативной активностью, которая необходима для синтеза НМО в клетке-хозяине. Неограничивающие варианты генов, которые могут быть преимущественно экспрессированы под контролем промотора *glp*, описаны в таблицах 2 и 3 и в рабочих примерах.

В соответствии с вышеизложенным второй аспект изобретения относится к рекомбинантной клетке, содержащей нуклеотидную конструкцию по изобретению. Рекомбинантная клетка в настоящем документе взаимозаменяемо обозначается как клетка-хозяин. Предпочтительно, клетка-хозяин представляет собой бактериальную клетку. Термины «вид бактерий-хозяев», «бактериальная клетка-хозяин» используются взаимозаменяемо для обозначения бактериальной клетки, которая была трансформирована для содержания ДНК-конструкции по изобретению и способна экспрессировать гетерологичный полипептид, кодируемый соответствующей гетерологичной кодирующей последовательностью ДНК конструкции. Термины «трансформация», «трансформированный» и «трансплантированный» являются синонимами и обозначают процесс, в котором внеклеточная нуклеиновая кислота, такая как вектор, содержащий конструкцию по изобретению, с сопутствующим материалом или без него, проникает в клетку-хозяина. Трансформация подходящих клеток-хозяев, например, вектором экспрессии, может быть осуществлена хорошо известными способами, такими как электропорация, конъюгация, или химическими способами, такими как трансформация, опосредованная фосфатом кальция, и системами природной трансформации, описанными, например, в Maniatis et al. или в Ausubel et al.

Что касается бактериальных клеток-хозяев, в принципе нет никаких ограничений; они могут быть зубактериями (грамположительными или грамотрицательными) или архебактериями, если они позволяют генетические манипуляции для вставки гена, представляющего интерес, и могут выращиваться в производственных масштабах. Предпочтительно, чтобы клетка-хозяин позволяла культивирование до высокой плотности клеток. Неограничивающими примерами бактериальных клеток-хозяев, которые подходят для рекомбинантного промышленного получения НМО по изобретению, могут быть *Erwinia herbicola* (*Pantoea agglomerans*), *Citrobacter freundii*, *Pantoea citrea*, *Pectobacterium carotovorum* или *Xanthomonas campestris*. Бактерии рода *Bacillus* также могут быть использованы, включая *Bacillus subtilis*, *Bacillus licheniformis*, *Bacillus coagulans*, *Bacillus thermophilus*, *Bacillus laterosporus*, *Bacillus megaterium*, *Bacillus mycoides*, *Bacillus pumilus*, *Bacillus lentus*, *Bacillus cereus* и *Bacillus circulans*. Так же, бактерии родов *Lactobacillus* и *Lactococcus* могут быть модифицированы с использованием способов по настоящему изобретению, включая, но не ограничиваясь этим, *Lactobacillus acidophilus*, *Lactobacillus salivarius*, *Lactobacillus plantarum*, *Lactobacillus helveticus*, *Lactobacillus delbrueckii*, *Lactobacillus rhamnosus*, *Lactobacillus bulgaricus*, *Lactobacillus crispatus*, *Lactobacillus*

*gasseri*, *Lactobacillus casei*, *Lactobacillus reuteri*, *Lactobacillus jensenii* и *Lactococcus lactis*. *Streptococcus thermophiles* и *Propionibacterium freudenreichii* также являются подходящими бактериальными видами для изобретения, описанного в настоящем документе. Также включены в качестве части этого изобретения штаммы, модифицированные, как описано в настоящем документе, из родов *Enterococcus* (например, *Enterococcus faecium* и *Enterococcus thermophiles*), *Bifidobacterium* (например, *Bifidobacterium longum*, *Bifidobacterium infantis* и *Bifidobacterium bifidum*), *Sporolactobacillus* spp., *Micromonospora* spp., *Micrococcus* spp., *Rhodococcus* spp., и *Pseudomonas* (например, *Pseudomonas fluorescens* и *Pseudomonas aeruginosa*). Бактерии, обладающие характеристиками, описанными в настоящем документе, культивируют в присутствии лактозы, и НМО, продуцируемые клеткой, извлекают либо из самой бактерии, либо из культурального супернатанта бактерий. НМО очищают с использованием подходящей процедуры, доступной в данной области (например, такой, как описанная в WO2015188834, WO2017182965 или WO2017152918).

В предпочтительном варианте осуществления клеткой-хозяином является *E. coli*. Однако, как уже упоминалось, для целей изобретения могут быть использованы различные клетки-хозяева.

Одно требование к клетке-хозяину состоит в том, что она содержит функциональную ДНК-зависимую РНК-полимеразу, которая может связываться с промотором и инициировать транскрипцию ДНК конструкции. РНК-полимераза может быть эндогенной (нативной), гомологичной (рекомбинантной) или чужеродной/гетерологичной (рекомбинантной) для клетки-хозяина.

Конструкция по изобретению, трансформированная в отобранного бактериального хозяина, может быть экспрессирована как интегрированная в геном экспрессионная кассета или клонирована в подходящий вектор экспрессии и экспрессирована в плазмиде. В различных вариантах осуществления может быть предпочтительным использовать систему экспрессии на основе генома, в других вариантах осуществления предпочтительной может быть экспрессия с плазмиды. Однако преимуществом является использование конструкции по изобретению в системе экспрессии на основе генома, поскольку, как ни удивительно, единственная копия конструкции, интегрированной в геном и экспрессируемой из него, может обеспечить высокий и стабильный уровень экспрессии встроенного генного продукта. Дополнительным преимуществом является то, что геномная экспрессия устойчива в течение длительных периодов времени. Для целей изобретения могут быть использованы стандартные способы для встраивания конструкций по изобретению в геном клетки-хозяина или в экспрессионные плазмиды, которые, например, описаны в Sambrook et al., Wilson & Walker, Maniatis et al. и Ausubel et al.

Термины «трансформация», «трансформированный» и «трансплантированный» являются синонимами и обозначают процесс, в котором внеклеточная нуклеиновая кислота, такая как вектор, содержащий конструкцию по изобретению, с сопутствующим

материалом или без него, проникает в клетку-хозяина. Трансформация подходящих клеток-хозяев, например, экспрессионным вектором, может быть осуществлена хорошо известными способами, такими как электропорация, конъюгация, или химическими способами, такими как трансформация, опосредованная фосфатом кальция, и системами природной трансформации, описанными, например, в Maniatis et al. или в Ausubel et al.

Для экспрессии на основе генома существует требование к клетке-хозяину - клетка должна быть способна осуществлять гомологичную рекомбинацию (что важно для встраивания экспрессионного картриджа в геном). Следовательно, предпочтительно, чтобы клетка-хозяин имела функциональный рекомбинационный белок RecA. Однако, поскольку RecA может вызывать нежелательные события рекомбинации во время культивирования, предпочтительно, чтобы клетка-хозяин имела геномную мутацию в своем геномном сайте *recA* (что делает его дисфункциональным), но вместо этого имела функциональный RecA, обеспечиваемый последовательностью *recA*, присутствующей на хелперной плазмиде, которая может быть удалена (элиминирована) после рекомбинации с использованием термочувствительного репликона хелперной плазмиды (Datsenko K.A. and Wanner B.L., (2000) *Proc Natl Acad Sci U S A.* 97(12):6640-5). С учетом рекомбинации, в дополнение к RecA, предпочтительно, чтобы клетка-хозяин содержала ДНК-последовательности, кодирующие рекомбинационные белки (например, *Exo*, *Beta* и *Gam*). В этом случае может быть выбрана клетка-хозяин, которая уже имеет эту функцию, или клетка-хозяин создается *de novo* с помощью генной инженерии для вставки этих последовательностей.

Что касается локуса интеграции, то система экспрессии, используемая в изобретении, допускает широкую вариабельность. В принципе, может быть выбран любой локус с известной последовательностью, при условии, что функция последовательности либо необязательна, либо, если необходимо, может быть восполнена (как, например, в случае ауксотрофии). Многие локусы интеграции, подходящие для целей изобретения, описаны в прототипах (см., например, Francia VM & Lobo JMG (1996), *J. Bacteriol* v178 p. 894-898; Juhas M et al (2014) [doi.org/10.1371/journal.pone.0111451](https://doi.org/10.1371/journal.pone.0111451); Juhas M & Aijoka FW (2015) *Microbial Biothechnol* v.8:617-748; Sabi A et al (2013) *Microbial Cell Factories* **12**:60).

ДНК-конструкция также может быть вставлена сайт-специфично. Принимая во внимание сайт-специфичную вставку гена, другое требование к клетке-хозяину состоит в том, чтобы она содержала по меньшей мере одну геномную область (либо кодирующую, либо не кодирующую функциональную или нефункциональную область, либо область с неизвестной функцией), у которой известна ее последовательность, и структура которой может быть нарушена или иным образом модифицирована, чтобы позволить вставку гетерологичной последовательности, не нанося вреда клетке.

В некоторых вариантах осуществления клетка-хозяин несет в своем геноме маркерный ген для проведения отбора.

При выборе локуса интеграции необходимо учитывать, что частота мутаций ДНК, вызванная так называемой «адаптивной эволюцией», варьирует по геному *E. coli*, и что

метаболическая нагрузка, запускаемая хромосомно кодируемой экспрессией рекомбинантного гена, может вызывать повышенную частоту мутаций в сайте интеграции. Чтобы получить экспрессирующую и стабильную клетку-хозяина, в качестве сайта интеграции предпочтительно выбрать высококонсервативную область генома, что обеспечивает сниженную частоту мутаций. Такими высоко консервативными областями генома *E. coli* являются, например, гены, кодирующие компоненты рибосомы, или гены, участвующие в биосинтезе пептидогликана, и эти области предпочтительно могут быть выбраны для интеграции картриджа экспрессии.

Геномная область с известной последовательностью, которая может быть выбрана для интеграции картриджа, может быть выбрана из кодирующей области несущественного гена или его части; из необязательной некодирующей функциональной области (т.е. промотора, транспозона и т. д.), из генов, удаление которых может иметь полезные эффекты с точки зрения продукции конкретного представляющего интерес белка, например определенных протеаз, белков наружных мембран, потенциальных загрязнителей продукта, генов, кодирующих метаболитические белки (например, важных для метаболизма сахарных молекул, который нежелателен или необязателен для данного штамма-хозяина и/или процесса ферментации) или белки путей передачи стрессовых сигналов, например, тех, которые возникают при ответе на жесткие условия культуры, механизма трансляционного контроля прокариот, который подавляет синтез тРНК и рРНК во время аминокислотного голодания. В альтернативном варианте, сайт интеграции может представлять собой маркерный ген, который позволяет проводить селекцию по исчезновению указанного маркерного фенотипа после интеграции. В альтернативном варианте, сайт, который можно выбрать для интеграции, представляет собой функцию, которая при удалении обеспечивает ауксотрофию, то есть неспособность организма синтезировать конкретное органическое соединение, необходимое для его роста. В этом случае сайт интеграции может представлять собой фермент, участвующий в биосинтезе или метаболических путях, удаление такого фермента приводит к ауксотрофному штамму. Положительные клоны, то есть те, которые несут экспрессионную кассету, могут быть отобраны по ауксотрофии по субстрату или молекулам-предшественникам указанных ферментов. В альтернативном варианте, сайт интеграции может представлять собой ауксотрофный маркер (нефункциональный, т.е. дефектный ген), который заменен/дополнен соответствующим прототрофным маркером (т.е. последовательностью, которая дополняет или заменяет дефектную последовательность), присутствующим в экспрессионной кассете, таким образом позволяя прототрофный отбор.

В одном аспекте область является несущественным геном. Согласно одному аспекту, это может быть ген, который сам по себе несущественен для клетки. Незаменимые бактериальные гены известны из литературы, например, из базы данных PEC (профилирование хромосомы *E. coli*) (<http://www.shigen.nig.ac.jp/ecoli/pec/genes.jsp>) или из так называемой «Коллекции Кейо» (Baba et al., *Molecular Systems Biology* (2006) 2, 2006.0008). Одним из примеров несущественного гена является *RecA*. Интеграция

экспрессионной кассеты в этом сайте обеспечивает геномную мутацию, описанную выше в контексте требований к клеткам-хозяевам.

Подходящие сайты интеграции, например сайты, которые легко доступны и/или, как ожидается, обеспечивают более высокий уровень экспрессии, могут быть определены в предварительном скрининге. Такой скрининг может быть выполнен путем создания серии одиночных делеционных мутантов в соответствии с коллекцией Keio (Baba et al., 2006), в результате чего интеграционный картридж содержит в качестве переменных элементов различные рекомбинационные последовательности, которые были предварительно выбраны с учетом конкретных сайтов встраивания, и в качестве константных элементов - базовые последовательности для интеграции и отбора, включая в качестве суррогатного «гена, представляющего интерес» ДНК-последовательность, кодирующую легко обнаруживаемый белок под контролем индуцибельного промотора, например, зеленый флуоресцентный белок. Уровень экспрессии созданных таким образом мутантов с одним нокаутом можно легко определить количественно с помощью измерения флуоресценции. На основании результатов этой процедуры может быть достигнут индивидуальный уровень экспрессии желаемого целевого белка путем изменения сайта интеграции и/или количества встраиваемых картриджей.

В вариантах осуществления, в которых клетка-хозяин содержит ДНК-последовательности, кодирующие рекомбинационные белки (например, Eho, Beta и Gam - либо в качестве признака исходной клетки, либо полученные генной инженерией), интеграция может происходить в геномном сайте, где расположены эти последовательности рекомбинационных белков. При интеграции экспрессионного картриджа последовательности, кодирующие рекомбинационные белки, разрушаются или удаляются и, следовательно, нет необходимости, как в случае кодируемых плазмидой хелперных белков, удалять их на отдельном этапе.

Интеграция представляющего интерес гена в бактериальный геном может быть достигнута обычными способами, например, с использованием линейных картриджей, которые содержат фланкирующие последовательности, гомологичные определенному сайту в хромосоме, как описано для сайта attTn7 (Waddell C.S. and Craig N.L., *Genes Dev.* (1988) Feb;2(2):137-49.); способами геномной интеграции последовательностей нуклеиновых кислот, в которых рекомбинация опосредуется функцией Red-рекомбиназы фага  $\lambda$  или функцией RecE/RecT-рекомбиназы профага  $\lambda$  (Murphy, *J Bacteriol.* (1998);180(8):2063-7; Zhang et al., *Nature Genetics* (1998) **20**: 123-128 Muyrers et al., *EMBO Rep.* (2000) 1(3): 239-243); способами, основанными на рекомбинации Red/ET (Wenzel et al., *Chem Biol.* (2005), 12(3):349-56.; Vetcher et al., *Appl Environ Microbiol.* (2005);71(4):1829-35); или

Положительные клоны, то есть клоны, которые несут экспрессионную кассету, могут быть отобраны, например, с помощью маркерного гена или потери или усиления функции гена.

В некоторых вариантах осуществления используются клетки-хозяева, которые уже

содержат маркерный ген, интегрированный в их геном, например ген устойчивости к антибиотику или ген, кодирующий флуоресцентный белок, например GFP. В этом случае экспрессионный картридж, который не содержит селектируемого маркера, интегрируется в локус хромосомного маркерного гена, и положительные клоны отбираются по потере/исчезновению соответствующего фенотипа, например, они выбираются по чувствительности к антибиотикам или исчезновению флуоресценции, которая может быть непосредственно визуализирована на культуральных чашках. Преимущество этих вариантов осуществления состоит в том, что маркер либо разрывается, либо полностью заменяется кассетой экспрессии, и, таким образом, функциональная маркерная последовательность отсутствует после интеграции и не требует удаления, если это нежелательно, как в случае генов устойчивости к антибиотику.

В альтернативном варианте, маркерный ген является частью экспрессионного картриджа. В случае, когда маркер, используемый для селекции, является геном, обеспечивающим устойчивость к антибиотикам (например, канамицину или хлорамфениколу), положительные клоны отбираются по устойчивости к антибиотикам (то есть по росту в присутствии соответствующего антибиотика). Маркерный ген (независимо от того, присутствует ли он в геноме клетки-хозяина или был введен посредством экспрессионного картриджа), может быть удален при интеграции кассеты.

В некоторых вариантах осуществления экспрессирующая клетка может быть сконструирована так, чтобы она несла дефектный селектируемый маркерный ген, например ген устойчивости к антибиотику, такому как хлорамфеникол или канамицин, флуоресцентный маркер или ген, участвующий в метаболическом пути сахара или аминокислоты. В этом случае картридж с геном, представляющим интерес, несет недостающую часть маркерного гена, и путем интеграции маркерный ген восстанавливает свою функциональность. Например, картридж несет недостающую часть маркерного гена на одном из своих концов и интегрируется непосредственно рядом с дефектным маркерным геном, интегрированным в геном, так что слияние двух фрагментов делает маркерный ген полным и позволяет его функциональную экспрессию. В случае гена устойчивости к антибиотику, клетки, несущие экспрессионную кассету, устойчивы к определенному антибиотику, в случае флуоресцентного маркера клетки можно визуализировать по флуоресценции, а в случае гена метаболического пути клетки приобретают способность метаболизировать соответствующий компонент. Преимущество этого варианта осуществления заключается в том, что необходимо синтезировать только короткий участок маркерного гена картриджа, что позволяет использовать более короткие или меньшие по размеру картриджи для вставки по сравнению с предшествующим уровнем техники.

В некоторых вариантах осуществления отбор положительных клонов (т.е. клонов, которые несут экспрессионную кассету) может быть осуществлен путем коррекции (т.е. комплементации) ауксотрофии клетки-хозяина. В таких вариантах осуществления используется клетка-хозяин, которая имеет мутацию, которая была выбрана для простого

отбора положительных трансформированных колоний, например штамма, который имеет делецию или мутацию, делающую его неспособным синтезировать соединение, которое необходимо для его роста (такую мутацию называют ауксотрофным маркером). Например, бактериальный мутант, в котором инактивирован ген пути синтеза пролина, является ауксотрофом по пролину. Такой штамм не способен синтезировать пролин и поэтому может расти только в том случае, если пролин может быть извлечен из окружающей среды, в отличие от пролинового прототрофа, который может расти в отсутствие пролина.

Можно использовать любую клетку-хозяина, имеющую ауксотрофный маркер. Предпочтительно использовать в качестве ауксотрофных маркеров мутации в генах, необходимые для синтеза аминокислот, например мутации в генах, относящихся к синтезу пролина, лейцина или треонина или ко-факторов, таких как тиамин. Согласно изобретению ауксотрофию клеток-хозяев корректируют путем интеграции в геном отсутствующего/дефектного гена в качестве компонента экспрессионного картриджа вместе с интеграцией представляющего интерес гена. Полученные таким образом прототрофные клетки можно легко отобрать, выращивая их на так называемой «минимальной среде» (прототрофный отбор), которая не содержит соединения, по которому исходная клетка-хозяин является ауксотрофной, что позволяет расти только положительным клонам.

Прототрофный отбор не зависит от локуса интеграции. Локусом интеграции для прототрофного отбора может быть любой ген в геноме или в локусе, несущем ауксотрофный маркер. Особое преимущество прототрофного отбора заключается в том, что после успешной интеграции в геноме не остается ни маркера устойчивости к антибиотикам, ни какого-либо другого чужеродного для хозяина маркера. Следовательно, нет необходимости удалять указанные маркерные гены, что обеспечивает быструю и простую процедуру клонирования и селекции. Другое преимущество состоит в том, что восстановление функции гена полезно для клетки и обеспечивает более высокую стабильность системы.

В альтернативном варианте, маркерный ген, который вставляется в геном вместе с экспрессионным картриджем, может представлять собой метаболический ген, который допускает определенный режим отбора. Такой метаболический ген может позволить клетке расти на определенном (необычном) сахаре или других источниках углерода, и отбор положительных клонов может быть достигнут путем выращивания клеток на указанном сахаре в качестве единственного источника углерода.

Как описано выше, во время длительного культивирования бактерий адаптивная эволюция может вызывать повышенную частоту мутаций в месте интеграции во время экспрессии хромосомно кодируемого рекомбинантного белка. Использование мутантного штамма с ауксотрофным нокаутом в сочетании с экспрессионным картриджем, дополняющим отсутствующую функцию мутантного штамма (тем самым генерируя штамм-прототроф из мутанта-ауксотрофа), имеет дополнительное преимущество в том,

что восстановленный ген обеспечивает преимущества для клетки, благодаря которым клетка получает конкурентное преимущество, так что клетки, в которых произошла адаптивная эволюция, подавляются. Таким образом, обеспечивается способ отрицательного отбора для мутированных клонов.

В некоторых вариантах осуществления (в случае, когда представляющий интерес белок позволяет обнаружение на уровне одной клетки или одной колонии, например, с помощью анализа FACS или иммунологического анализа (ELISA)), маркерный ген не требуется, поскольку положительные клоны могут быть идентифицированы с помощью прямого обнаружения белка, представляющего интерес.

Способы интеграции для получения экспрессирующей клетки-хозяина не ограничиваются интеграцией одного представляющего интерес гена в одном сайте в геноме; они обеспечивают возможность варибельности как в отношении сайта интеграции, так и в отношении экспрессионных кассет. Например, может быть вставлено более одного представляющего интерес гена, т.е. две или несколько одинаковых или разных последовательностей под контролем идентичных или разных промоторов могут быть интегрированы в один или несколько разных локусов в геноме. Например, так можно экспрессировать два различных белка, которые образуют гетеродимерный комплекс. Гетеродимерные белки состоят из двух индивидуально экспрессированных белковых субъединиц. Одним примером такого белка является молекула антитела, например, тяжелая и легкая цепь моноклонального антитела или фрагмента антитела; другими примерами гетеродимерных белков являются CapZ, Ras, человеческая геликаза II и т.д. Эти две последовательности, кодирующие мономеры, могут присутствовать в одном экспрессионном картридже, который вставлен в один локус интеграции. В альтернативном варианте, эти две последовательности также могут присутствовать в двух разных экспрессионных кассетах, которые вставляются независимо друг от друга в два разных локуса интеграции. В любом случае, промоторы и режимы индукции могут быть одинаковыми или разными.

Хотя изобретение позволяет и может быть преимущественно использовано на практике для безплазмидного получения представляющих интерес биологических молекул, кодируемых геном конструкции по изобретению, оно не исключает того, что в системе экспрессии по изобретению содержится плазида, которая несет подлежащие экспрессии последовательности, кроме гена, представляющего интерес, например, хелперные белки и/или рекомбинационные белки, описанные выше. Естественно, следует позаботиться о том, чтобы в таких вариантах осуществления преимущества изобретения не отменялись присутствием плазмиды, то есть предпочтительно, чтобы такая плазида присутствовала в низком числе копий и не оказывала метаболической нагрузки на клетку.

Система экспрессии, используемая в способе по изобретению, может быть подобрана так, что она практически или полностью не содержит фаговых функций.

Суммируя вышеприведенные варианты осуществления, геномная экспрессия экспрессионных кассет по изобретению обеспечивает следующие основные

преимущества:

В отношении процедуры конструирования экспрессионного хозяина, преимуществами являются: (i) простой метод синтеза и амплификации линейного картриджа для вставки, (ii) высокая степень гибкости (т.е. без ограничений) по отношению к локусу интеграции, (iii) высокая степень гибкости в отношении селективируемого маркера и принципа селекции, (iv) возможность последующего удаления селективируемого маркера, (v) дискретное и определенное количество вставленных экспрессионных картриджей (обычно один или два).

Интеграция одного или более рекомбинантных генов в геном приводит к дискретному и заранее определенному числу генов, представляющих интерес, на клетку. В варианте осуществления изобретения, в котором встраивается одна копия гена, это число обычно равно единице (за исключением случая, когда клетка содержит более одного генома, как это происходит временно во время деления клетки), по сравнению с экспрессией на основе плазмид, которая сопровождается увеличением числа копий до нескольких сотен. В системе экспрессии, используемой в способе по настоящему изобретению, за счет освобождения метаболизма хозяина от репликации плазмиды, увеличенная доля синтетического клеточного аппарата используется для продукции рекомбинантного белка. Поэтому, может применяться элемент конструкции, позволяющий сильную экспрессию, например, Pglp, например, PglpF, PglpA, PglpD или PglpT, без побочных эффектов в отношении метаболизма хозяина за счет снижения содержания копий гена.

Как указано выше, системы экспрессии на основе плазмид имеют недостаток, заключающийся в том, что во время клеточного деления клетки могут терять плазмиду и, следовательно, представляющий интерес ген. Такая потеря плазмиды зависит от нескольких внешних факторов и повышается с увеличением количества клеточных делений (поколений). Это означает, что ферментация на основе плазмид ограничена числом поколений (в обычных ферментациях это число составляет приблизительно от 20 до 50). Напротив, основанная на геноме система экспрессии, используемая в способе по изобретению, обеспечивает стабильное, предварительно заданное содержание гена для практически бесконечного числа поколений и, таким образом, теоретически бесконечное время культивирования в контролируемых условиях (без недостатка появления клеток, которые не продуцируют представляющий интерес белок, и с единственным ограничением в виде потенциально встречающихся природных мутаций, поскольку они могут встречаться в любом гене).

В случае химически индуцируемых промоторов изобретение обеспечивает конкретное преимущество, заключающееся в том, что количество молекулы-индуктора, при добавлении, например, в непрерывном режиме, прямо пропорционально содержанию гена на клетку, либо постоянно в течение всего культивирования, либо изменяется в течение времени культивирования в заранее определенных значениях. Тем самым может быть достигнут контроль скорости рекомбинантной экспрессии, что представляет

большой интерес для регулирования скорости экспрессии гена.

Поскольку основанная на геноме система экспрессии позволяет осуществлять точный контроль экспрессии белка, она особенно предпочтительно в сочетании с экспрессией, мишенью которой являются сигнальные пути, которая зависит от хорошо контролируемой экспрессии или полагается на нее.

Как описано выше, изобретение позволяет создавать упрощенные процессы, улучшать предсказуемость процесса и высокую воспроизводимость от ферментации до ферментации. Способ по изобретению, использующий описанную выше систему экспрессии, может быть осуществлен в периодическом процессе с подпиткой, или полунепрерывном или непрерывном режиме, благодаря чему преимущества системы экспрессии, кодируемой геномом, используются оптимальным образом. Нет ограничений в отношении параметров процесса, таких как скорость роста, температура и компоненты культуральной среды, за исключением случаев, определенных требованиями клетки-хозяина и определенных выбранным промотором.

Другое преимущество относится к выбору молекулы-индуктора: большинство доступных систем для высокого уровня экспрессии рекомбинантных генов в *E. coli* представляют собой промоторно-операторные системы на основе *lac*, индуцируемые IPTG. Система экспрессии, используемая в изобретении, позволяет культивирование в среде с ограниченным углеродом при непрерывной или импульсной подаче источника углерода, например лактозы, и обеспечивает жесткий контроль скорости экспрессии с широким спектром недорогих индукторов источников углерода, таких как глицерин, фукоза, лактоза, глюкоза.

Важно, что система экспрессии, используемая в изобретении, имеет преимущество, заключающееся в обеспечении высокого выхода рекомбинантно продуцируемых биологических молекул, как в отношении концентрации молекулы на объем культуральной среды (т.е. титра), так и в отношении содержания молекулы в полученной биомассе. Эта особенность делает систему экспрессии, используемую в изобретении, превосходной по сравнению с системами экспрессии предшествующего уровня техники.

Кроме того, изобретение предлагает преимущество, заключающееся в том, что выбор клетки-хозяина для экспрессии и/или оптимальной конструкции картриджа экспрессии может быть легко достигнут в предварительных скрининговых тестах. В качестве примера, в таких предварительных скринингах сконструирована серия картриджей с линейной экспрессией, которые варьируют по меньшей мере в отношении одного элемента, который влияет на свойства экспрессии представляющего интерес белка (уровень экспрессии или качественные характеристики, такие как биологическая активность), а именно, регуляторных элементов (например, промотора и/или сайта связывания полимеразы) и/или последовательности представляющего интерес гена (т.е. различные варианты использования кодонов), и/или направляющих последовательностей для рекомбинации, и/или любых других элементов в картридже, таких как секреторные лидеры. Варианты картриджа встраивают в геном предварительно выбранной клетки-

хозяина, и полученные варианты экспрессионных хозяев культивируют, включая индукцию экспрессии белка, в контролируемых условиях. Сравнивая экспрессию белка, выбирают вариант клетки-хозяина, показывающий наиболее благоприятные результаты с точки зрения процесса промышленного производства. В варианте этого предварительного подхода вместо определения оптимального экспрессионного картриджа, может быть идентифицирован оптимальный бактериальный штамм путем интеграции идентичных картриджей в панель различных клеток-хозяев. Поскольку у стратегии интеграции есть преимущество, заключающееся в том, что она позволяет интегрировать дискретное число копий гена (например, только одну) в геном, предварительный скрининг различных параметров может быть выполнен без вмешательства со стороны процесса репликации плазмиды или изменения числа копий плазмиды.

Согласно изобретению, термин «культивировать» (или «культивирование», также называемое «ферментация») относится к размножению бактериальных экспрессирующих клеток в контролируемом биореакторе в соответствии со способами, известными в промышленности.

Производство рекомбинантных белков обычно осуществляется путем выращивания в больших объемах. Термины «производство» и «производственный масштаб» в значении изобретения определяют ферментацию с минимальным объемом 5 л культуральной среды. Обычно процесс «производственного масштаба» определяется тем, что он способен обрабатывать большие объемы препарата, содержащего представляющий интерес рекомбинантный белок, и получать количества представляющего интерес белка, которые соответствуют, например, в случае терапевтического белка, требованиям к клиническим испытаниям, а также для предложения на рынке. В дополнение к большому объему, метод производственного масштаба, в отличие от простых лабораторных методов, таких как культивирование в колбах, характеризуется использованием технической системы биореактора (ферментера), который оснащен устройствами для перемешивания, аэрации, подачи питательных веществ, мониторинга и контроля параметров процесса (рН, температуры, напряжения растворенного кислорода, противодавления и т. д.). Поведение системы экспрессии в лабораторном методе не позволяет прогнозировать поведение этой системы в сложной среде биореактора.

Системы экспрессии по настоящему изобретению могут преимущественно использоваться для производства рекомбинантных средств в промышленном масштабе (как в отношении объема, так и в отношении технической системы) в сочетании с режимом культивирования, который основан на подаче питательных веществ, в частности, в процессе с подпиткой, или в непрерывном или полунепрерывном процессе.

В некоторых вариантах осуществления способ по изобретению представляет собой периодический процесс с подпиткой.

Принимая во внимание, что периодический процесс представляет собой режим культивирования, в котором все питательные вещества, необходимые для культивирования клеток, содержатся в исходной культуральной среде без дополнительной

подачи дополнительных питательных веществ во время ферментации, в периодическом процессе с подпиткой, после периодической фазы наступает фаза подпитки, в которой одно или несколько питательных веществ поступают в культуру путем подпитки. Целью подпитки питательными веществами является увеличение количества биомассы (так называемый «процесс культивирования клеток при высокой плотности» или «HCDC»), чтобы также увеличить количество рекомбинантного белка. Хотя в большинстве процессов культивирования режим подпитки является критическим и важным, настоящее изобретение не ограничено в отношении определенного режима подпитки.

Подача питательных веществ (подпитка) может осуществляться в непрерывном или периодическом режиме в соответствии со способами, известными в данной области. Режим подпитки может быть задан заранее (то есть подпитка осуществляется независимо от фактических параметров процесса), например, линейная постоянная подпитка, линейное увеличение подпитки, ступенчатое увеличение подпитки или следование математической функции при подпитке, например экспоненциальная подпитка.

В предпочтительном варианте осуществления способ по изобретению представляет собой процесс с подпиткой, в котором режим подпитки предварительно задан в соответствии с экспоненциальной функцией. При использовании режима экспоненциального питания, удельная скорость роста  $\mu$  клеточной популяции может быть предварительно задана на постоянном уровне и оптимизирована с учетом максимальной экспрессии рекомбинантного белка. Контроль скорости подпитки основан на желаемой удельной скорости роста  $\mu$ . Когда используется определенная среда, как описано ниже, рост может быть точно предсказан и предварительно определен путем расчета аликвоты биомассы, которая должна быть образована на основе введенной единицы субстрата.

В другом предпочтительном варианте осуществления режим экспоненциальной подачи может сопровождаться на конечных стадиях культивирования линейной постоянной подачей.

В другом варианте осуществления процесса с подпиткой применяется линейная постоянная подача. Линейная постоянная подача характеризуется скоростью подпитки (объемом питательной среды в единицу времени), которая является постоянной (то есть неизменной) на протяжении определенных фаз культивирования.

В другом варианте осуществления процесса с подпиткой применяется линейное увеличение подпитки. Линейное увеличение подпитки характеризуется скоростью подачи питательной среды, следующей линейной функцией. Подпитка в соответствии с линейной возрастающей функцией характеризуется определенным увеличением скорости подпитки за определенный период времени.

В другом варианте осуществления способа с подпиткой в соответствии с настоящим изобретением для управления подпиткой применяется алгоритм управления с обратной связью (в отличие от предварительно определенного режима подпитки). В процессе подпитки с обратной связью скорость подпитки зависит от фактического уровня определенного параметра культивирования. Параметры культивирования, подходящие

для подпитки с обратной связью, представляют собой, например, биомассу (и полученные для нее химические или физические параметры), растворенный кислород, дыхательный коэффициент, рН или температуру. Другой пример режима питания с обратной связью основан на фактической концентрации глюкозы в биореакторе.

В другом варианте осуществления бактериальные клетки, несущие экспрессионную кассету на основе генома согласно настоящему изобретению, культивируют в непрерывном режиме. Процесс непрерывной ферментации характеризуется определенной, постоянной и непрерывной скоростью подачи свежей культуральной среды в биореактор, в результате чего культуральная среда одновременно удаляется из биореактора с той же определенной, постоянной и непрерывной скоростью удаления. Поддерживая питательную среду, скорость подпитки и скорость удаления на одном и том же постоянном уровне, параметры и условия культивирования в биореакторе остаются постоянными (так называемое «устойчивое состояние»). Удельная скорость роста  $\mu$  может быть задана заранее и является исключительно результатом скорости подачи и объема питательной среды в биореакторе. Поскольку клетки, имеющие одну или более экспрессионных кассет на основе генома, являются генетически очень стабильными (в отличие от структурно и сегрегационно нестабильных систем экспрессии на основе плазмид или систем экспрессии, в которых вставленная в геном кассета зависит от геномной амплификации), число поколений клеток (удвоений клеток) по изобретению теоретически не ограничено, а также, следовательно, время культивирования. Преимущество культивирования генетически стабильной системы экспрессии на основе генома в непрерывном режиме заключается в том, что можно получить более высокое общее количество рекомбинантного белка за период времени по сравнению с генетически нестабильными системами предшествующего уровня техники. Кроме того, из-за теоретически неограниченного времени выращивания, непрерывное культивирование клеток по изобретению может приводить к более высокому общему количеству белка за период времени, даже по сравнению с процессами культивирования с подпиткой. Неограничивающие рабочие примеры, приведенные ниже, показывают высокую стабильность и продуктивность экспрессирующей конструкции на основе генома.

Другой предпочтительный вариант осуществления относится к полунепрерывному культивированию клеток. Полунепрерывный процесс культивирования в контексте изобретения представляет собой процесс, который на первом этапе используется как процесс с подпиткой (т.е. этап периодической ферментации с последующим этапом подпитки). После получения определенного объема или биомассы (т.е. обычно при получении верхнего предела объема ферментера) значительную часть клеточной среды, содержащей представляющий интерес рекомбинантный белок, удаляют из биореактора. Впоследствии подпитка возобновляется до тех пор, пока биомасса или объем культуральной среды снова не достигнут определенного значения. Этот способ (удаление культуральной среды и повторная наполнение при подпитке) может проводиться как минимум один раз, и теоретически неопределенное время.

Что касается типа культуральной среды, используемой в процессе ферментации, ограничений нет. Культуральная среда может быть полуопределенной, т.е. может содержать сложные компоненты среды (например, дрожжевой экстракт, соевый пептон, казаминовые кислоты и т.д.), или она может являться химически определенной, без каких-либо комплексных соединений.

Предпочтительно используется «определенная среда». «Определенная» среда (также называемая «минимальной» или «синтетической» средой) состоит исключительно из химически определенных веществ, то есть источников углерода, таких как глюкоза или глицерин, солей, витаминов и, с учетом возможной ауксотрофии по штамму, специфических аминокислот или других веществ, такие как тиамин. Наиболее предпочтительно, чтобы в качестве источника углерода использовалась глюкоза. Обычно источником углерода в питательной среде служит ограничивающий рост компонент, который контролирует удельную скорость роста.

В способах по изобретению получают значительно более высокие выходы, потому что рост бактерий и высокая, но физиологически приемлемая скорость экспрессии рекомбинантных генов могут поддерживаться в течение всего производственного процесса.

Как описано выше, в наиболее предпочтительном варианте осуществления изобретения представляющий интерес белок находится под контролем «индуцибельного» или «контролируемого» промотора.

Нет ограничений в отношении способа, которым осуществляется индукция экспрессии белка. Например, индуктор может быть добавлен в виде единичного или многократного болюса или путем непрерывной подпитки, причем последняя также известна как «подпитка индуктором (индукция)». Нет никаких ограничений в отношении момента времени, в который происходит индукция. Индуктор может быть добавлен в начале культивирования или в точке начала непрерывной подпитки питательными веществами или после начала подпитки. Подпитка индуктором может быть достигнута либо с помощью индуктора, содержащегося в культуральной среде, либо путем отдельной подпитки.

Преимущество подпитки индуктором состоит в том, что оно позволяет контролировать дозировку индуктора, то есть позволяет поддерживать дозировку определенного или постоянного количества индуктора на постоянное число генов, представляющих интерес, в производственной системе. Например, подпитка индуктором позволяет дозировать индуктор, который пропорционален биомассе, что приводит к постоянному соотношению индуктора и биомассы. Единицами биомассы, на которых может основываться доза индуктора, могут быть, например, сухой вес клеток (CDW), вес влажных клеток (WCW), оптическая плотность, общее количество клеток (TCN; клетки на объем) или колониеобразующие единицы (КОЕ на объем), или он-лайн контролируемые сигналы, которые пропорциональны биомассе (например, флуоресценция, мутность, диэлектрическая проницаемость и т.д.). По существу, способ по изобретению позволяет

точно дозировать индуктор по любому параметру или сигналу, который пропорционален биомассе, независимо от того, измеряется ли сигнал в автономном режиме или в режиме онлайн. Поскольку количество генов, представляющих интерес, определяется и является постоянным единицы для биомассы (один или несколько генов на клетку), следствием этого режима индукции является постоянная доза индуктора на представляющий интерес ген. В качестве дополнительного преимущества точную и оптимальную дозировку количества индуктора относительно количества биомассы можно определить и оптимизировать экспериментально.

Возможно, нет необходимости определять фактический уровень биомассы аналитическими методами. Например, может быть достаточно добавить индуктор в количестве, которое основано на предыдущих культивированиях (прошлые данные по биомассе). В другом варианте осуществления может быть предпочтительным добавить количество индуктора на одну единицу биомассы, как теоретически рассчитано или предсказано. Например, для культивирования на основе подпитки (например, периодической или непрерывной) хорошо известно, что одна единица ограничивающего рост компонента в питательной среде, обычно источника углерода, приведет к определенному количеству биомассы. Например, 1 г глюкозы в качестве субстрата, ограничивающего рост, приведет к приблизительно 0,33 г сухой массы клеток (также выраженной коэффициентом выхода субстрата  $Y_{x/s}=0,33$ ). Следовательно, определенная дозировка индуктора на представляющий интерес ген также может быть достигнута определенной дозировкой индуктора на единицу, ограничивающую рост компонента, поскольку определенная единица, ограничивающая рост, приводит к определенной единице биомассы, а определенная единица биомассы содержит определенное количество молекул представляющих интерес белков согласно способу по изобретению.

Существенным преимуществом является то, что подпитка ограниченным количеством индуктора предотвращает метаболическую нагрузку и снижает стресс в пользу максимизации способности синтеза белка.

Соотношение индуктора и биомассы (или гена, или субстрата, ограничивающего рост) не обязательно может быть постоянным. Это также может быть линейное увеличение, линейное уменьшение, увеличение или уменьшение в соответствии с экспоненциальными или другими математическими функциями и т.д. Существенным признаком согласно изобретению является то, что определяется значение дозы индуктора на представляющий интерес ген.

В некоторых вариантах осуществления способ по изобретению представляет собой периодический процесс с подпиткой, в котором индуктор присутствует в загруженной среде с начала культивирования.

Способ индукции экспрессии также может быть конститутивным, что означает, что индукция не запускается химически или другими стимулами, но является постоянной с начала культивирования. Конститутивная индукция является предпочтительным индукционным режимом для непрерывного культивирования, но также полезна для

культивирования с подпиткой.

Рекомбинантные бактерии и способы получения НМО хорошо известны (см., например, Priem B et al, (2002) *Glycobiology*;12(4):235-40; Drouillard S et al, (2006) *Angew. Chem. Int. Ed.* 45:1778 -1780; Fierfort N & Samain E (2008) *J Biotechnol* 134:261-265; Drouillard S. et al. (2010) *Carbohydrate Research* 345 1394-1399; Gebus C et al (2012) *Carbohydrate Research* 363 83-90).

Следуя способам, описанным в данной области техники, и согласно изобретению, бактериальный хозяин может использовать эндогенный или экзогенный путь синтеза гуанозиндифосфата (GDP) -фукозы для получения фукозилированных НМО. Под «путем синтеза GDP-фукозы» подразумевается последовательность реакций, обычно контролируемых и катализируемых ферментами, которая приводит к синтезу GDP-фукозы. Примерный путь синтеза GDP-фукозы в *E. coli* изложен ниже. В этом пути синтеза ферменты для синтеза GDP-фукозы включают в себя: 1) *manA*=фосфоманноизомеразу (PMI), 2) *manB*=фосфоманномутазу (PMM), 3) *manC*=манноза-1-фосфат-гуанилитрансферазу (GMP), 4) *gmd*=GDR-манноза-4,6-дегидратаза (GMD), 5) *fcl*=GDP-фукозосинтазу (GFS) и 6) *ΔwcaJ*=мутированная трансфераза UDP-глюкозы-липидный-носитель.

Глюкоза → Glc-6-P → Fru-6-P → <sup>1</sup>Man-6-P → <sup>2</sup>Man-1-P → <sup>3</sup>GDP-Man → <sup>4,5</sup>GDP-Fuc<sup>6</sup>Колановая кислота.

Путь синтеза GDP-фукозы из фруктозо-8-фосфата, общего метаболического интермедиата у всех организмов, состоит из 5 ферментативных стадий: 1) PMI (фосфоманноизомеразы), 2) PMM (фосфоманномутаза), 3) GMP (манноза-1-фосфат-гуанилитрансфераза), 4) GMD (GDP-манноза-4,6-дегидратаза) и 5) GFS (GDP-фукозосинтаза). В контексте настоящего изобретения ферменты пути синтеза GDP, которые способствуют увеличению внутриклеточного пула GDP-фукозы, включены в группу полезных белков, которые косвенно участвуют в синтезе НМО. Отдельные виды бактерий имеют различные генетические возможности синтеза GDP-фукозы. Например, *E. coli* может синтезировать ферменты, которые способны выполнять все пять стадий, в то время как в *Bacillus licheniformis* отсутствуют ферменты, способные выполнять стадии 4 и 5 (то есть GMD и GFS). Любые ферменты в пути синтеза GDP, которые в природе отсутствуют в каких-либо конкретных бактериальных видах, могут быть введены хозяину с помощью молекулярной инженерии с использованием рекомбинантных конструкций ДНК по изобретению. Гены, кодирующие отсутствующие ферменты, могут быть представлены либо в плазмидном экспрессионном векторе, либо в виде экзогенных генов, интегрированных в хромосому хозяина (ферменты пути синтеза GDP в контексте изобретения представляют собой ферменты, которые косвенно участвуют в производстве НМО, т.е. являются ферментами, необходимыми для производства НМО).

Бактерия, пригодная для продукции НМО, например, *E. coli*, может содержать ген эндогенной β-галактозидазы или экзогенный ген β-галактозидазы, например, *E. coli* содержит эндогенный ген *lacZ* (например, с номером доступа в базе GenBank - V00296

(GI: 41901)). Для целей изобретения НМО-продуцирующую бактериальную клетку генетически модифицируют либо для включения любого гена  $\beta$ -галактозидазы, либо для включения инактивированного гена. Ген может быть инактивирован путем полной или частичной делеции соответствующей нуклеотидной последовательности из бактериального генома, или последовательность гена мутируют так, чтобы она совсем не транскрибировалась, или при транскрипции не транслировалась, или при трансляции в белок (т.е.  $\beta$ -галактозидазу), белок не обладал бы соответствующей ферментативной активностью. Таким образом, бактерия, продуцирующая НМО, накапливает увеличенный внутриклеточный пул лактозы, что полезно для производства НМО.

Функциональный ген лактозопермеазы предпочтительно присутствует в НМО-продуцирующей бактерии по изобретению. Ген лактозопермеазы представляет собой эндогенный ген лактозопермеазы или экзогенный ген лактозопермеазы. Например, ген лактозопермеазы содержит ген *lacY* из *E. coli* (например, номер доступа в GenBank - V00295 (GI: 41897)). Многие бактерии обладают генетической способностью транспортировать лактозу из питательной среды в клетку, используя транспортный белок, который является либо гомологом лактозопермеазы *E. coli* (например, обнаруженный в *Bacillus licheniformis*), или транспортер, который является представителем распространенного PTS-семейства переносчиков сахара (например, обнаруженный в *Lactobacillus casei* и *Lactobacillus rhamnosus*). Для бактерий, не обладающих генетической способностью транспортировать внеклеточную лактозу в цитоплазму клетки, эта способность обеспечивается экзогенным геном переносчика лактозы (например, *lacY E. coli*), предоставленным на рекомбинантных конструкциях ДНК и доставляемым или в плазмидном экспрессионном векторе, или в виде экзогенных генов, интегрированных в хромосому хозяина.

Для получения фукозилированного олигосахарида путем биосинтеза бактерия предпочтительно содержит мутацию в гене синтеза эндогенной колановой кислоты (фукозосодержащего экзополисахарида). Под «геном синтеза колановой кислоты» подразумевается ген, участвующий в последовательности реакций, обычно контролируемых и катализируемых ферментами, которые приводят к синтезу колановой кислоты. Типичные гены синтеза колановой кислоты включают ген *rcaA* (например, номер доступа в GenBank - M58003 (GI:1103316)), ген *rcaB*, (например, номер доступа в GenBank - E04821 (GI:2173017)), ген *wcaJ*, (например, номер доступа в GenBank - (аминокислотная последовательность) BAA15900 (GI:1736749)), ген *wzc*, (например, номер доступа в GenBank - (аминокислотная последовательность) BAA15899 (GI:1736748)), ген *wcaD*, (например, номер доступа в GenBank - (аминокислотная последовательность) BAE76573 (GI:85675202)), ген *wza*, (например, номер доступа в GenBank - (аминокислотная последовательность) BAE76576 (GI:85675205)), ген *wzb* (например, номер доступа в GenBank - (аминокислотная последовательность) BAE76575 (GI:85675204)), и ген *wzc* (например, номер доступа в GenBank - (аминокислотная последовательность) BAA15913 (GI:1736763)).

Это достигается с помощью ряда генетических модификаций эндогенных генов *E. coli*, участвующих либо непосредственно в биосинтезе предшественника колановой кислоты, либо в общем контроле синтетического регулона колановой кислоты. В частности, способность штамма-хозяина *E. coli* синтезировать внеклеточный капсульный полисахарид, колановую кислоту, устраняется удалением гена *wcaJ*, кодирующего трансферазу UDP-глюкозы на липидный носитель. На фоне полного отсутствия *wcaJ* GDP-фукоза накапливается в цитоплазме *E. coli*. Сверхэкспрессия положительного регуляторного белка, *RcsA*, в пути синтеза колановой кислоты приводит к увеличению внутриклеточного уровня GDP-фукозы. Сверхэкспрессия дополнительного положительного регулятора биосинтеза колановой кислоты, а именно *RcsB*, также используется вместо или в дополнение к сверхэкспрессии *RcsA*, чтобы повысить внутриклеточный уровень GDP-фукозы. В альтернативном варианте, биосинтез колановой кислоты увеличивается после введения нулевой мутации в ген *lon* из *E. coli* (например, номер доступа в GenBank - L20572 (GI: 304907), включенный в настоящий документ в качестве ссылки). *Lon* представляет собой аденозин-5'-трифосфат(АТФ)-зависимую внутриклеточную протеазу, которая ответственна за деградацию *RcsA*, который был упомянут выше в качестве положительного транскрипционного регулятора биосинтеза колановой кислоты в *E. coli*. На фоне нулевой экспрессии *lon*-протеазы *RcsA* стабилизируется, уровень *RcsA* повышается, повышается экспрессия генов, отвечающих за синтез GDP-фукозы в *E. coli* (т.е. они сверхэкспрессируются), и увеличивается внутриклеточная концентрация GDP-фукозы. Белки *RcsA* и *RcsB* рассматриваются как полезные для целей изобретения, и в некоторых вариантах их уровни в клетках-хозяевах повышаются с использованием конструкций по изобретению, где соответствующие гены функционально связаны с промотором *glp*, предпочтительно, *PglpF*.

Предпочтительно, чтобы бактерия, продуцирующая фукозилированный НМГ, содержала ген экзогенной фукозилтрансферазы. Например, экзогенный ген фукозилтрансферазы кодирует  $\alpha(1,2)$ -фукозилтрансферазу и/или  $\alpha(1,3)$ -фукозилтрансферазу. Примером гена  $\alpha(1,2)$ -фукозилтрансферазы является ген *wcfW* из *Bacteroides fragilis* NCTC 9343. Примером гена  $\alpha(1,3)$ -фукозилтрансферазы является ген *futA* из *Helicobacter pylori* 26695. Один пример гена *futA* из *Helicobacter pylori* представлен в GenBank под регистрационным номером HV532291 (GI: 365791177). В контексте изобретения ферменты с фукозилтрансферазной активностью (такие как, например, фукозилтрансферазы, кодируемые указанными выше генами) упоминаются в данном документе как белки, которые важны для получения фукозилированных НМО и непосредственно участвуют в продукции одного или более фукозилированных НМО.

Способ получения фукозилированных НМО путем биосинтеза в соответствии с изобретением может включать следующие стадии: предоставление бактерии, которая содержит ген дисфункциональной  $\beta$ -галактозидазы, ген экзогенной фукозилтрансферазы, где ген экзогенной фукозилтрансферазы является частью экспрессионной кассеты, где ген функционально связан с промотором *glp*, например *PglpF*, мутацию в кластере генов

колановой кислоты и функциональный ген лактозопермеазы; культивирование бактерии в присутствии источника углерода, например глицерина, глюкозы, сахарозы, лактозы и т.д.; и выделение фукозилированных НМО из бактерии или из культурального супернатанта бактерии. Используемые в настоящем документе бактерии-продуценты НМО генетически модифицированы для увеличения внутриклеточного пула лактозы (по сравнению с диким типом), для увеличения уровня активности фукозилтрансферазы и, необязательно, для увеличения пула внутриклеточного гуанозиндифосфат(GDP)-фукозы. Согласно изобретению, указанные бактерии содержат по меньшей мере одну нуклеотидную конструкцию, которая содержит нуклеотидную последовательность, кодирующую фермент, который прямо или косвенно участвует в продукции НМО, и промотор *glp* по изобретению, предпочтительно *PglpF*, который функционально связан с этой нуклеотидной последовательностью. Бактерия также может содержать мутацию в гене пути синтеза колановой кислоты (фукозосодержащего экзополисахарида), такого как ген *wcaJ*, что приводит к увеличению внутриклеточного пула GDP-фукозы. Эндогенный ген *lacZ* *E. coli* предпочтительно удален или функционально инактивирован, но таким образом, что экспрессия расположенного за ним гена лактозопермеазы (*lacY*) остается неизменной. Организм, модифицированный, как описано выше, сохраняет способность переносить лактозу из питательной среды и создает внутриклеточный пул лактозы для использования в качестве акцепторного сахара в синтезе олигосахаридов. Бактерия может дополнительно содержать экзогенный ген *rcsA* и/или *rcsB* (например, в эктопической нуклеотидной конструкции, такой как плазида), и бактерия необязательно дополнительно содержит мутацию в гене *lacA*.

Бактерии, обладающие фукозилтрансферазной активностью, могут включать один или оба из экзогенного гена фукозилтрансферазы, кодирующего  $\alpha(1,2)$ -фукозилтрансферазу, и экзогенного гена фукозилтрансферазы, кодирующего  $\alpha(1,3)$ -фукозилтрансферазу. Примером гена  $\alpha(1,2)$ -фукозилтрансферазы является ген *wcfW* из *Bacteroides fragilis* NCTC 9343. Другие гены  $\alpha(1,2)$ -фукозилтрансферазы, которые используют лактозу в качестве акцепторного сахара (например, ген *futC* из *Helicobacter pylori* 26695 или ген *wbsJ* из *E. coli* 0128: B12), могут быть легко заменены на ген *wcfW* из *Bacteroides fragilis*. Один пример гена *futC* из *Helicobacter pylori* представлен под номером доступа в GenBank - EF452503 (GI: 134142866). Примером гена  $\alpha(1,3)$ -фукозилтрансферазы является ген *futA* из *Helicobacter pylori* 26695, хотя другие гены  $\alpha(1,3)$ -фукозилтрансферазы, известные в данной области, могут быть замещены (например, гены  $\alpha(1,3)$ -фукозилтрансферазы из *Helicobacter hepaticus* Nh0072, *Helicobacter bills*, *Campylobacter jejuni* или из видов *Bacteroides*). Некоторые примеры  $\alpha(1,3)$ -фукозилтрансфераз и других ферментов, которые участвуют в продукции различных фукозилированных НМО, показаны в таблице 4 ниже.

Таблица 4

<i>Ген</i>	<i>Исходный биологический</i>	<i>Номер доступа</i>	<i>Фермент</i>	<i>Пример НМО</i>
------------	-------------------------------	----------------------	----------------	-------------------

	<i>вид</i>			
MAMA_R764	<i>Мимивирус</i> <i>Acanthamoeba</i> <i>polyphaga</i>	AGC02224.1	$\alpha$ —1,3- фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F- pLNH I
Mg791	Megavirus <i>chiliensis</i>	AEQ33441.1	$\alpha$ —1,3- фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F- pLNH I
Moumou_00703	<i>Мимивирус</i> <i>Acanthamoeba</i> <i>polyphaga</i> M10A	AGC02224.1	$\alpha$ —1,3- фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F- pLNH I
futA	<i>Helicobacter</i> <i>pylori</i> ATCC 26695	NP_207177.1	$\alpha$ —1,3- фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI, LNDFH-II, F- pLNH I
fucT	<i>Helicobacter</i> <i>pylori</i> NCTC 11639 (укороченный)	AAB81031.1	$\alpha$ —1,3- фукозилтрансфераза	2'FL, 3'FL, DFL, LNFP-I, LNFP-III, LNFP-V, LNFP-VI,

				LNDFH-II, F-pLNH I
fucTIII	<i>Helicobacter pylori</i> ATCC 43504	AY450598.1	$\alpha$ -1,4-фукозилтрансфераза	LNDFH-I, LNDFH-II
fucTa	<i>Helicobacter pylori</i> UA948	AF194963.2	$\alpha$ -1,3/4-фукозилтрансфераза	LNFP-II, LNDFH-I, LNDFH-II

Бактерия может содержать экспрессионную кассету по настоящему изобретению, обеспечивающую экспрессию или сверхэкспрессию одной или более из указанных выше фукозилтрансфераз и, соответственно, более высокую продукцию одного или более фукозилированных НМО, например 2'-FL, 3FL, DFL, LNFP-I, -II, -III, -V, VI, LNDFH-I, -II или -III.

Изобретение в дополнительных вариантах осуществления относится к НМО-продуцирующим клеткам-хозяевам, которые содержат одну или более нуклеотидных конструкций, включающих один, два, три или большее число любого из генов, описанных в настоящем документе (то есть генов, кодирующих важные или полезные белки для продукции НМО), где, предпочтительно, по меньшей мере одна из конструкций содержит промотор Pglp, функционально связанный с по меньшей мере одним из генов, где по меньшей мере одна из конструкций интегрирована в геном, где предпочтительно по меньшей мере одна из конструкций, интегрированных в геном, содержит промотор Pglp, функционально связанный по меньшей мере с одним из генов, кодирующих белок, который необходим для синтеза НМО, и указанная конструкция присутствует в геноме в низкой копийности. В некоторых вариантах осуществления клетки-хозяева по настоящему изобретению могут продуцировать фукозилированные НМО, в других вариантах осуществления клетки могут продуцировать сиалирированные НМО, в других вариантах осуществления клетки могут продуцировать нефукозилированные нейтральные НМО.

Для получения описанных в настоящем документе сиалирированных НМО могут быть использованы общие принципы и способы, ранее описанные в данной области техники (см., например, Drouillard S et al, (2010) *Carbohydrate Research* 345:1394-1399, or Fierfort N & Samain E (2008) *J Biotechnol* 134:261-265).

В общем, сконструированная бактериальная клетка, которая способна продуцировать сиалирированный олигосахарид человеческого молока, например 6'-SL (6'-сиалиллактозу), содержит экзогенный ген сиалилтрансферазы, кодирующий  $\alpha$ (2,6)-сиалилтрансферазу. Бактериальная клетка может быть *E. coli*. Экзогенный ген сиалилтрансферазы, используемый для получения 6'-SL, может быть получен из любых доступных источников, например источников, описанных для ряда организмов рода *Photobacterium*. Еще одна сиалирированная НМО, например 3'-SL (3'-сиалиллактоза), может быть получена с помощью сконструированных бактерий, содержащих экзогенную

молекулу нуклеиновой кислоты, кодирующую  $\alpha(2,3)$ -сиалилтрансферазу. Ген экзогенной сиалилтрансферазы, используемый для получения 3'-SL, может быть получен из любого доступного источника, например, из *Neisseria meningitidis* и *Neisseria gonorrhoeae*. Некоторые примеры подходящих сиалилтрансфераз перечислены в таблице 5 ниже.

Таблица 5

<i>Ген</i>	<i>Исходный биологический вид</i>	<i>Номер доступа</i>	<i>Фермент</i>	<i>Пример НМО</i>
Pd2,6ST	<i>Photobacterium damsela</i> JT0160	BAA25316.1	$\alpha$ -2,6-сиалилтрансфераза	6'SL
PspST6	<i>Photobacterium</i> sp. JT-ISH-224	BAF92026.1	$\alpha$ -2,6-сиалилтрансфераза	6'SL
PiST6_145	<i>Photobacterium leiognathi</i> JT-SHIZ-145	BAF91416.1	$\alpha$ -2,6-сиалилтрансфераза	6'SL
PiST6_119	<i>Photobacterium leiognathi</i> JT-SHIZ-119	BAI49484.1	$\alpha$ -2,6-сиалилтрансфераза	6'SL
NST	<i>Neisseria meningitidis</i> MC58	AAC44541.1	$\alpha$ -2,3-сиалилтрансфераза	3'SL
NGO_1081	<i>Neisseria gonorrhoeae</i> ( <i>штамм</i> ATCC 700825/FA 1090)	YP_208160.1	$\alpha$ -2,3-сиалилтрансфераза	3'SL

Предпочтительно, чтобы созданная бактерия содержала дефицитный (нарушенный) катаболический путь сиаловой кислоты. «Катаболический путь сиаловой кислоты» означает последовательность реакций, обычно контролируемых и катализируемых ферментами, которые приводят к деградации сиаловой кислоты. Типичным катаболическим путем сиаловой кислоты, описанным в настоящем документе, является путь *E. coli*. В этом пути сиаловая кислота (Neu5Ac; N-ацетилнейраминная кислота) разлагается ферментами NanA (лиаза N-ацетилнейраминной кислоты) и NanK (N-ацетилманнозаминакиназа) и NanE (N-ацетилманнозамин-6-фосфат-эпимераза), все они кодируются опероном *nanATEK-yhcH* и подавляются NanR (<http://ecocyc.org/ECOLI>). Дефицит катаболического пути сиаловой кислоты достигается в хозяине *E. coli* путем введения мутации в эндогенную *nanA* (N-ацетилнейраминат лиазу) (например, номер доступа в GenBank - D00067.1(GL216588)) и/или *nanK* (N-ацетилманнозаминакиназу) (например, номер доступа в GenBank (аминокислотная последовательность) BAE77265.1 (GL85676015)) и/или *nanE* (N-ацетилманнозамин-6-фосфат-эпимеразу, GI: 947745,

включенную в настоящий документ посредством ссылки). Необязательно, чтобы ген *panT* (N-ацетилнейраминатного транспортера) также был инактивирован или мутирован. Другие промежуточные продукты метаболизма сиаловой кислоты включают: (ManNAc-6-P) N-ацетилманнозамин-6-фосфат; (GlcNAc-6-P) N-ацетилглюкозамин-6-фосфат; (GlcN-6-P) Гикозамин-6-фосфат и (FruC-6-P) фруктоза-6-фосфат. В некоторых предпочтительных вариантах осуществления *panA* мутирован. В других предпочтительных вариантах осуществления *panA* и *panK* являются мутированными, тогда как *panE* остается функциональным. В другом предпочтительном варианте осуществления *panA* и *panE* мутированы, тогда как *panK* не был мутирован, инактивирован или удален. Мутация представляет собой одно или несколько изменений в последовательности нуклеиновой кислоты, кодирующей продукт гена *panA*, *panK*, *panE* и/или *panT*. Например, мутация может представлять собой 1, 2, до 5, до 10, до 25, до 50 или до 100 изменений в последовательности нуклеиновой кислоты. Например, гены *panA*, *panK*, *panE* и/или *panT* мутированы нулевой мутацией. Нулевые мутации, как описано в настоящем документе, охватывают аминокислотные замены, добавления, делеции или вставки, которые либо вызывают потерю функции фермента (т.е. снижение активности или ее отсутствие), либо потерю фермента (т.е. отсутствие продукта гена). Под «делецией» подразумевается, что кодирующая область удаляется полностью или частично таким образом, что (функциональный) продукт гена не образуется. Под инактивацией подразумевается, что кодирующая последовательность была изменена таким образом, что полученный продукт гена является функционально неактивным, или кодируемый продукт гена имеет менее 100%, например, 90%, 80%, 70%, 60%, 50%, 40%, 30% или 20% активности от нативного эндогенного продукта гена природного происхождения. «Немутированный» ген или белок не отличаются от нативной, встречающейся в природе или эндогенной кодирующей последовательности на 1, 2, до 5, до 10, до 20, до 50, до 100, до 200 или до 500 или более кодонов, или до длины соответствующей закодированной аминокислотной последовательности.

Кроме того, бактерия (например, *E. coli*) также обладает способностью к синтезу сиаловой кислоты. Например, бактерия обладает способностью к синтезу сиаловой кислоты за счет наличия экзогенной UDP-GlcNAc 2-эпимеразы (например, *neuC* из *Campylobacter jejuni* (GenBank AAK91727.1; GI: 15193223) или эквивалента (например, *neuC* из *E. coli* S88 (GenBank YP\_002392936.1; GI: 218560023), Neu5Ac-синтазы (например, *neuB* из *C. jejuni* (GenBank AAK91726.1; GI: 15193222) или эквивалента (например, синтаза сиаловой кислоты из *Flavobacterium limnosediminis*, GenBank GL559220424) и/или CMP-Neu5Ac-синтаза (например, *neuA* из *C. jejuni* (GenBank AAK91728.1; GI: 15193224) или эквивалента (например, синтаза CMP-сиаловой кислоты из *Vibrio brasiliensis*, GenBank GI: 493937153).

Бактерии, продуцирующие сиалилированные НМО, содержат одну или более экзогенных сиалилтрансфераз, которые кодируются кодирующей ДНК экспрессионной кассеты по изобретению, которая присутствует в клетках-хозяевах либо в виде плазмиды,

либо интегрирована в геном. Предпочтительно, чтобы по меньшей мере одна из одной или более последовательностей ДНК, кодирующих сиалилтрансферазу, была функционально связана с промотором *glp*, описанным в настоящем документе, предпочтительно *PglpF*. Неограниченные примеры подходящих сиалилтрансфераз приведены в таблице 5.

Бактерия, обладающая способностью к синтезу сиаловой кислоты, может быть предпочтительно сконструирована так, чтобы иметь повышенную продукцию UDP-GlcNAc. Типичным способом достижения этого является избыточная экспрессия положительного эндогенного регулятора синтеза UDP-GlcNAc, например одновременная сверхэкспрессия генов *nagC* и *glmS* *E. coli*. Эта сверхэкспрессия *nagC* и *glmS* предпочтительно достигается путем функционального связывания генов с промотором *glp* по изобретению и экспрессии кассеты в виде интегрированной в геном, или, в альтернативном варианте, это может быть достигнуто путем предоставления дополнительных копий генов *nagC* и *glmS*, связанных с *glp* или другим промотором в плазмидном векторе.

Производство нейтральных N-ацетилглюкозаминсодержащих НМО в сконструированных бактериях также известна в данной области (см., например, Gebus C et al (2012) *Carbohydrate Research* 363 83-90).

Для получения N-ацетилглюкозаминсодержащих НМО, таких как лакто-N-триоза 2 (LNT2), лакто-N-тетраоза (LNT), лакто-N-неотетраоза (LNnT), лакто-N-фукопентаза I (LNFP-I), лакто-N-фукопентаза II (LNFP-II), лакто-N-фукопентаза III (LNFP-III), лакто-N-фукопентаза V (LNFP-V), лакто-N-дифукогексаза I (LDFH-I), лакто-N-дифукогексаоза II (LDFH-II) и лакто-N-неодифукогексаоза II (LNDFH-111), бактерия содержит функциональный *lacY* и дисфункциональный ген *lacZ*, как описано выше, и она сконструирована таким образом, чтобы содержать экзогенный ген UDP-GlcNAc:Gal $\alpha$ / $\beta$ -R $\beta$ 3-N-ацетилглюкозаминилтрансферазы или его функциональный вариант или фрагмент. Этот экзогенный ген UDP-GlcNAc:Gal $\alpha$ / $\beta$ -R $\beta$ 3-N-ацетилглюкозаминилтрансферазы может быть получен из любого из ряда источников, например, из генов *IgtA*, описанных для *N. meningitidis* (Genbank, номер доступа белка AAF42258.1) или *N. gonorrhoeae* (Genbank, номер доступа белка ACF31229.1). Необязательно, дополнительный ген экзогенной гликозилтрансферазы может быть коэкспрессирован в бактерии, содержащей экзогенную UDP-GlcNAc:Gal $\alpha$ / $\beta$ -R $\beta$ 3-N-ацетилглюкозаминилтрансферазу. Например, ген  $\beta$ -1,4-галактозилтрансферазы коэкспрессируется с геном UDP-GlcNAc:Gal $\alpha$ / $\beta$ -R $\beta$ 3-N-ацетилглюкозаминилтрансферазы. Этот экзогенный ген  $\beta$ -1,4-галактозилтрансферазы может быть получен из любого из ряда источников, например, из описанного для *N. meningitidis* гена *IgtB* (Genbank, номер доступа белка AAF42257.1) или описанного для *N. pylogi* гена HP0826/galT (Genbank, номер доступа белка NP 207619.1). Необязательно, дополнительный ген экзогенной гликозилтрансферазы, коэкспрессируемый в бактериях, содержащих экзогенный ген UDP-GlcNAc:Gal $\alpha$ / $\beta$ -R $\beta$ 3-N-ацетилглюкозаминилтрансферазы, представляет собой ген  $\beta$ -1,3-галактозилтрансферазы, например, описанный для *E. coli* 055:H7, ген *wbgO* (Genbank, номер доступа белка YP

003500090.1), или для *H. pylori* ген jhp0563 (Genbank, номер доступа белка AEZ55696.1), или для *Streptococcus agalactiae* *mun* Ib O12 ген cps1BJ (Genbank, номер доступа белка AV050723). Функциональные варианты и фрагменты любого из ферментов, описанных выше, также охвачены настоящим изобретением.

Предпочтительно, чтобы по меньшей мере один ген, кодирующий фермент, как любой из вышеперечисленных, то есть оба/любой ген N-ацетилглюкозаминилтрансферазы и/или ген галактозилтрансферазы, был функционально связан с Pglp по изобретению и экспрессировался с соответствующей интегрированной в геном кассеты. В одном варианте осуществления ген, который интегрирован в геном, представляет собой ген, кодирующий галактозилтрансферазу, например ген HP0826, кодирующий фермент GalT из *H. pylori* (Genbank, номер доступа белка NP 207619.1); в другом варианте осуществления ген, который интегрирован в геном, представляет собой ген, кодирующий  $\beta$ -1,3-N-ацетилглюкозаминилтрансферазу, например ген IgtA из *N. meningitidis* (Genbank, номер доступа белка AAF42258.1). В этих вариантах осуществления второй ген, то есть ген  $\beta$ -1,3-N-ацетилглюкозаминилтрансферазы или галактозилтрансферазы, соответственно, может быть экспрессирован либо с интегрированной в геном, либо с плазмидной кассеты. Второй ген может быть необязательно экспрессирован либо под контролем промотора glp, либо под контролем любого другого промотора, подходящего для системы экспрессии, например, Plac.

Преимущественно, бактерии, продуцирующие N-ацетилглюкозаминсодержащие НМО, могут быть сконструированы таким образом, чтобы иметь увеличенный внутриклеточный пул UDP-GlcNAc. Типичным средством достижения этого признака является избыточная экспрессия положительного эндогенного регулятора синтеза UDP-GlcNAc, например одновременная сверхэкспрессия генов nagC и glmS *E. coli*. Эта сверхэкспрессия nagC и glmS, предпочтительно, достигается путем оперативного связывания генов с промотором glp по изобретению и интеграции кассеты в геном хозяина, или, в альтернативном варианте, это может быть достигнуто путем предоставления дополнительных копий генов nagC и glmS, связанных с glp или другим промотором в плазмидном векторе.

Для продукции НМО, продуцирующие НМО бактерии, описанные в настоящем документе, культивируют в соответствии с процедурами, известными в данной области техники, в присутствии подходящего источника углерода, например глюкозы, глицерина, лактозы и т.д., и полученные НМО собирают из культуральной среды и микробной биомассы, образовавшихся в процессе культивирования. После этого НМО очищают в соответствии с процедурами, известными в данной области, например, такими, как описанные в WO2015188834, WO2017182965 или WO2017152918, и очищенные НМО используют в качестве нутрицевтиков, фармацевтических препаратов или для любых других целей, например, для исследований.

Другие признаки и преимущества изобретения будут очевидны из приведенного ниже описания рабочих примеров и из формулы изобретения. Если не указано иное, то все

технические и научные термины, используемые в данном документе, имеют то же значение, которое обычно понимают специалисты в области техники, к которой относится это изобретение. Хотя способы и материалы, подобные или эквивалентные тем, которые описаны в настоящем документе, могут использоваться при практическом применении или испытании настоящего изобретения, подходящие способы и материалы описаны ниже. Кроме того, материалы, способы и примеры являются только иллюстративными и поэтому не ограничивают объем изобретения.

## ПРИМЕРЫ

### Материалы и способы

Если не указано иное, то для манипуляций с нуклеиновыми кислотами, их трансформации и экспрессии используются стандартные методы, векторы, элементы контрольной последовательности и другие элементы экспрессионных систем, известные в области молекулярной биологии. Такие стандартные методы, векторы и элементы можно найти, например, в: Ausubel et al. (eds.), *Current Protocols in Molecular Biology* (1995) (John Wiley & Sons); Sambrook, Fritsch, & Maniatis (eds.), *Molecular Cloning* (1989) (Cold Spring Harbor Laboratory Press, NY); Berger & Kimmel, *Methods in Enzymology 152: Guide to Molecular Cloning Techniques* (1987) (Academic Press); Bukhari et al. (eds.), *DNA Insertion Elements, Plasmids and Episomes* (1977) (Cold Spring Harbor Laboratory Press, NY); Miller, J.H. *Experiments in molecular genetics* (1972.) (Cold spring Harbor Laboratory Press, NY).

#### *Штаммы и плазмиды*

Используемый бактериальный штамм MDO был сконструирован из *Escherichia coli* K12 DH1. Генотип *E. coli* K12 DH1:  $F^-$ ,  $\lambda^-$ , *gyrA96*, *recA1*, *relA1*, *endA1*, *thi-1*, *hsdR17*, *supE44*. В дополнение к *E. coli* K12 DH1 генотип MDO имеет следующие модификации: *lacZ*: делеция 1,5 т.п.н., *lacA*: делеция 0,5 т.п.н., *nanKETA*: делеция 3,3 т.п.н. *mdoH*  $\Delta$ mc/oH: делеция 0,5 т.п.н., и вставка промотора Plac перед геном *gmd*.

Штаммы, используемые в настоящих примерах, описаны в таблице 6. Донорские и вспомогательные плазмиды, использованные для конструирования этих штаммов, включены в таблицу 7 вместе с высококопийными плазмидами, введенными в некоторые из сконструированных штаммов.

Таблица 6

Исходные штаммы		
Идентификатор штамма, ID	Описание генома	Описание плазмиды
DH1	$F^-$ $\lambda^-$ <i>endA1</i> <i>recA1</i> <i>relA1</i> <i>gyrA96</i> <i>thi-1</i> <i>glnV44</i> <i>hsdR17</i> ( $r_K^- m_K^-$ )	
MDO	<i>E. coli</i> DH1 $\Delta$ <i>lacZ</i> , $\Delta$ <i>lacA</i> , $\Delta$ <i>nanKETA</i> , $\Delta$ <i>melA</i> , $\Delta$ <i>wcaJ</i> , $\Delta$ <i>mdoH</i>	
<b>Штаммы, экспрессирующие репортерные гены</b>		

Идентификатор штамма, ID	Описание генома	Описание плазмиды
MAP808	MDO galK::PglpF-lacZ-T1	--
MAP1010-9	MDO galK::PglpF400-9-lacZ-T1	--
MAP1010-11	MDO galK::PglpF400-11-lacZ-T1	--
MAP1010-13	MDO galK::PglpF400-13-lacZ-T1	--
MAP1010-17	MDO galK::PglpF400-17-lacZ-T1	--
MAP1010-19	MDO galK::PglpF400-19-lacZ-T1	--
MAP1010-20	MDO galK::PglpF400-20-lacZ-T1	--
MAP1025	MDO galK::PglpA-lacZ-T1	--
MAP1026	MDO galK::PglpD-lacZ-T1	--
MAP1027	MDO galK::PglpT-lacZ-T1	--
MAP1086	MDO galK:: <i>Δ175PglpF</i> -lacZ-T1	--
MAP1176	MDO galK::PglpF_SD1-lacZ-T1	--
MAP1178	MDO galK::PglpF_SD3-lacZ-T1	--
MAP1179	MDO galK::PglpF_SD4-lacZ-T1	--
MAP1180	MDO galK::PglpF_SD5-lacZ-T1	--
MAP1181	MDO galK::PglpF_SD6-lacZ-T1	--
MAP1182	MDO galK::PglpF_SD7-lacZ-T1	--
MAP1183	MDO galK::PglpF_SD8-lacZ-T1	--
MAP1184	MDO galK::PglpF_SD9-lacZ-T1	--
MAP1185	MDO galK::PglpF_SD10-lacZ-T1	--
MAP1206	MDO galK::PglpF_SD2-lacZ-T1	--
MAP1209	MDO galK:: <i>Δ190PglpF</i> -lacZ-T1	--
MAP1210	MDO galK:: <i>Δ25PglpF</i> -lacZ-T1	--
MAP1211	MDO galK:: <i>Δ150PglpF</i> -lacZ-T1	--
MAP1356	MDO galK::PglpF-lacZ-T1-galK glpR::kanR	--
MAP1365	MDO galK::PglpA_org-lacZ-T1-galK	--
MAP1366	MDO galK::PglpD_org-lacZ-T1-galK	--
MAP1367	MDO galK::PglpT_org-lacZ-T1-galK	--

MAP1368	MDO galK::PglpF_org-lacZ-T1-galK	--
MAP1370	MDO galK::Plac_org-lacZ-T1-galK	--
<b>Штаммы, экспрессирующие репортерные гены</b>		
<b>Идентификатор штамма, ID</b>	<b>Описание генома</b>	<b>Описание плазмиды</b>
MAP219	MDO Plac-Pd2	--
MAP700	MDO Plac-nst	--
MAP710	MDO PglpF-nst	--
MAP986	MDO PglpF-Pd2	--
MDO1	MDO	pBBR3-Plac-lgtA-tet, pBS-Plac-galT-amp
MDO15	MDO	pBBR3-Plac-lgtA-tet, pBS-Plac-galTK-amp
MP166	MDO 3xPlac-lgtA 3xPlac-galT lacI::CP6-galK	-
MP245	MDO 3xPlac-lgtA 2xPlac-galTK $\Delta$ lacI	-
MP1497	MDO PglpF-lgtA	pBS-Plac-galT-amp
MP1498	MDO PglpF-lgtA	pBS-Plac-galTK-amp
MP1499	MDO PglpF-galT	pBBR3-Plac-lgtA-tet
MP1655	MDO 2xPglpF-galTK	pBBR3-Plac-lgtA-tet
MP1825	MDO PglpF-galT PglpF-lgtA lacI::CP6-galK	-
MP1920	MDO PglpF-galTK PglpF-lgtA lacI::CP6-galK	-
MP2239	MDO PglpF-galTK PglpF-lgtA PglpF-futC $\Delta$ lacI	-
MP2374	MDO PglpF-galTK PglpF-lgtA PglpF-futC PglpF-CA $\Delta$ lacI	-
MP2622	MDO Plac-lgtA Plac-galT	-
MAP265	MDO Plac-Pd2 nadC::galK	pBS-Plac- neuBCA-nadC
MAP425	MDO 2xPlac-nst Plac-neuBCA $\Delta$ 50lacI	pBS-Plac- neuBCA-nadC
MAP1200	MDO PglpF-neuA PglpF-neuB PglpF-neuC PglpF-Pd2	-
MAP1214	MDO PglpF-neuA PglpF-neuB PglpF-neuC	-

	PglpF-nst	
FT18	MDO	pBP-Plac-futC-kan pBBR3-Plac-gmd-fcI- cpsB-cpsG
MAP965	MDO PglpF-gmd-fcI-gmm-wcaI-cpsB-cpsG PglpF-futC_op	-

Таблица 7

<b>Идентификатор плазмиды, ID</b>	<b>Описание</b>
pACBSR	Para-I-SceI- $\lambda$ Red, p15A ori, cam*
pUC57	pMB1, bla
pUC57::gal	<i>pUC57::galTK' / T1-galKM'</i>
pMAP99	<i>pUC57-galTK'-Plac-Pd2_op-T1-galKM'</i>
pMAP205	<i>pUC57::galTK'-PglpF-lacZ-T1-galKM'</i>
pMAP216	<i>pUC57-galTK'-Plac-<math>\Delta</math>29nst_op-T1-galKM'</i>
pMAP228	<i>pUC57-galTK'-PglpF-<math>\Delta</math>29nst_op-T1-galKM'</i>
pMAP391	<i>pUC57-galTK'-PglpF-Pd2-T1-galKM'</i>
pMAP431	<i>pUC57-galTK'-PglpA-lacZ-T1-galKM'</i>
pMAP432	<i>pUC57-galTK'-PglpD-lacZ-T1-galKM'</i>
pMAP433	<i>pUC57-galTK'-PglpT-lacZ-T1-galKM'</i>
pMAP457	<i>pUC57-galTK'-D25PglpF-lacZ-T1-galKM'</i>
pMAP462	<i>pUC57-galTK'-D150PglpF-lacZ-T1-galKM'</i>
pMAP463	<i>pUC57-galTK'-D175PglpF-lacZ-T1-galKM'</i>
pMAP486	<i>pUC57-galTK'-PglpF_SD1-lacZ-T1-galKM'</i>
pMAP487	<i>pUC57-galTK'-PglpF_SD2-lacZ-T1-galKM'</i>
pMAP488	<i>pUC57-galTK'-PglpF_SD3-lacZ-T1-galKM'</i>
pMAP489	<i>pUC57-galTK'-PglpF_SD4-lacZ-T1-galKM'</i>
pMAP490	<i>pUC57-galTK'-PglpF_SD5-lacZ-T1-galKM'</i>
pMAP491	<i>pUC57-galTK'-PglpF_SD6-lacZ-T1-galKM'</i>
pMAP492	<i>pUC57-galTK'-PglpF_SD7-lacZ-T1-galKM'</i>
pMAP493	<i>pUC57-galTK'-PglpF_SD8-lacZ-T1-galKM'</i>
pMAP494	<i>pUC57-galTK'-PglpF_SD9-lacZ-T1-galKM'</i>
pMAP495	<i>pUC57-galTK'-PglpF_SD10-lacZ-T1-galKM'</i>
pMAP537	<i>pUC57-galTK'-D190PglpF-lacZ-T1-galKM'</i>

pMAP689	pUC57- <i>galTK'</i> -PglpA_org-lacZ-T1- <i>galKM'</i>
pMAP690	pUC57- <i>galTK'</i> -PglpD_org-lacZ-T1- <i>galKM'</i>
pMAP691	pUC57- <i>galTK'</i> -PglpT_org-lacZ-T1- <i>galKM'</i>
pMAP693	pUC57- <i>galTK'</i> -PglpF_org-lacZ-T1- <i>galKM'</i>
pMAP695	pUC57- <i>galTK'</i> -Plac_org-lacZ-T1- <i>galKM'</i>
MP55	pBBR3-Plac- <i>lgtA-tet</i>
MP46	pBS-Plac- <i>galT-amp</i>
MP139	pBS-Plac- <i>galTK-amp</i>
pMAP101	pBS-Plac- <i>neuBCA-nadC</i>
MP415	pBP-Plac- <i>futC-kan</i>
MP416	pBBR3-Plac- <i>gmd-fcI-cpsB-cpsG</i>

### *Среды*

Среду Luria Broth (LB) готовили с использованием LB Broth Powder, Millers (Fisher Scientific), и чашки с LB-агаром приготавливали с использованием LB Agar Powder, Millers (Fisher Scientific). Скрининг штаммов на чашках LB, содержащих 5-бром-4-хлор-3-индолил-pD-галактопиранозид (X-gal), проводили с использованием концентрации X-gal 40 мкг/мл. При добавлении подходящего антибиотика, ампициллина (100 мкг/мл) и/или хлорамфеникола (20 мкг/мл).

Базальная минимальная среда имела следующий состав: NaOH (1 г/л), KOH (2,5 г/л), K<sub>2</sub>HPO<sub>4</sub> (7 г/л), NH<sub>4</sub>H<sub>2</sub>PO<sub>4</sub> (7 г/л), лимонная кислота (0,5 г/л), микроэлементный раствор (5 мл/л). Концентрированный раствор микроэлементов содержал: ZnSO<sub>4</sub>\*7H<sub>2</sub>O 0,82 г/л, лимонную кислоту 20 г/л, MnSO<sub>4</sub>\*H<sub>2</sub>O 0,98г/л, FeSO<sub>4</sub>\*7H<sub>2</sub>O 3,925 г/л, CuSO<sub>4</sub>\*5H<sub>2</sub>O 0,2 г/л. pH базальной минимальной среды довели до 7,0 с помощью 5 N NaOH и автоклавировали. Перед инокуляцией в базовую минимальную среду вводили 1 mM MgSO<sub>4</sub>, 4 мкг/мл тиамин, 0,5% заданного источника углерода (глюкоза, глицерин, сорбит, ксилоза, лактоза, мальтоза (MgSO<sub>4</sub>)) и при необходимости добавляли изопропил-β-D-тиогалактозид (IPTG) (0,2 mM), ампициллин (100 мкг/мл) и/или хлорамфеникол (20 мкг/мл). Тиамин, антибиотики и IPTG стерилизовали фильтрацией. Все процентные концентрации для глицерина представлены как объем/объем, а для глюкозы, сорбита, ксилозы, лактозы и мальтозы - как масса/объем.

Чашки M9, содержащие 2-дезоксигалактозу, имели следующий состав: агар 15 г/л (Fisher Scientific), 2,26 г/л 5x минимальной соли M9 (Sigma-Aldrich), 2 mM MgSO<sub>4</sub>, 4 мкг/мл тиамин, 0,2% глицерина и 0,2% 2-деокси-D-галактозы (Carbosynth).

Индикаторные планшеты MacConkey, содержащие галактозу, имели следующий состав: 40 г/л агаризованной основы MacConkey (BD Difco™). После автоклавирования и охлаждения до 50°C добавляли D-галактозу (Carbosynth) до конечной концентрации 1%.

### *Культивирование*

Если не указано иное, штаммы *E. coli* культивировали в среде Luria-Bertani (LB),

содержащей 0,2% глюкозы, при 37°C при перемешивании.

Культуры, собираемые для анализа на  $\beta$ -галактозидазу, готовили следующим образом: одну колонию из планшета с LB предварительно культивировали в 1 мл минимальной базальной среды, содержащей глюкозу (0,5%), в 10 мл в 24-луночном планшете с глубокими лунками (Deep well plate, Axygen). Планшет герметизировали перед культивированием с помощью Hydrophobic Gas Permeable Adhesive Seal (Axygen) и инкубировали в течение 24 часов при 34°C со встряхиванием при 700 об/мин в орбитальном шейкере (Edmund Bühler GmbH). Плотность клеток культуры контролировали при 600 нм с использованием спектрофотометра S-20 (Воесо, Германия). 20 мкл ночной культуры использовали для инокуляции в 2 мл среды LB или базальной минимальной среды, содержащей глюкозу или другой источник углерода (0,5%) в 24-луночном планшете с глубокими лунками. Антибиотики и/или IPTG добавляли при необходимости. Глубоколуночные планшеты покрывали герметизирующей фольгой и инкубировали в течение 24 часов при 28°C на орбитальном шейкере при 700 об/мин. После инкубации измеряли OD600 и собирали 0,5 мл клеточной культуры центрифугированием для предварительного анализа на  $\beta$ -галактозидазу.

#### *Химические компетентные клетки и трансформации*

*E. coli* инокулировали из LB-планшетов в 5 мл LB, содержащей 0,2% глюкозы, и культивировали при 37°C и встряхивании до OD 600 приблизительно 0,4. 2 мл культуры собирали центрифугированием в течение 25 секунд при 13000 g. Супернатант удаляли, и клеточный осадок ресуспендировали в 600 мкл холодного раствора ТВ (10 mM PIPES, 15 mM CaCl<sub>2</sub>, 250 mM KCl). Клетки инкубировали на льду в течение 20 минут с последующим осаждением в течение 15 секунд при 13000 g. Супернатант удаляли, и осадок клеток ресуспендировали в 100 мкл холодного раствора ТВ. Трансформацию плазмид осуществляли с использованием 100 мкл компетентных клеток и 1-10 нг плазмидной ДНК. Клетки и ДНК инкубировали на льду в течение 20 минут, а затем подвергали тепловому шоку при 42°C в течение 45 секунд. После 2-минутной инкубации на льду добавляли 400 мл среды SOC (20 г/л триптона, 5 г/л дрожжевого экстракта, 0,5 г/л NaCl, 0,186 г/л KCl, 10 mM MgCl<sub>2</sub>, 10 mM MgSO<sub>4</sub> и 20 mM глюкозы), и клеточную культуру инкубировали при 37°C со встряхиванием в течение 1 часа перед посевом на селективные чашки.

Лигирования плазмид трансформировали в химически компетентные клетки TOP10 при условиях, рекомендованных поставщиком (ThermoFisher Scientific).

#### *Методы работы с ДНК*

Плазмидную ДНК из *E. coli* выделяли с использованием набора QIAprep Spin Miniprep (Qiagen). Хромосомную ДНК из *E. coli* выделяли с использованием набора QIAmp DNA Mini Kit (Qiagen). Продукты ПЦР очищали с использованием набора для очистки ПЦР QIAquick (Qiagen). DreamTaq PCR Master Mix (Thermofisher), Phusion U hot start PCR master mix (Thermofisher), USER Enzym (New England Biolab) были использованы в соответствии с рекомендациями поставщика. Праймеры были

предоставлены Eurofins Genomics, Германия. Фрагменты ПЦР и плазмиды были секвенированы в Eurofins Genomics.

ПЦР на колониях проводили с использованием DreamTaq PCR Master Mix в условиях, рекомендованных поставщиком (ThermoFisher) в термоциклере T100™ (Bio-Rad). Например, во время конструирования штаммов, экспрессирующих репортерный или рекомбинантный ген из локуса galK, в реакции ПЦР на колониях были использованы праймеры 048 (5'-CCCAGCGAGACCTGACCGCAGAAC-3') (SEQ ID NO: 58) и 049 (5'-CCCCAGTCCATCAGCGTGACTACC-3') (SEQ ID NO: 59) для подтверждения правильности предполагаемой модификации.

Таблица 8: Праймеры, использованные для конструирования каркаса, используемого для получения донорских плазмид.

Название	Последовательность олигонуклеотида, 5'-3'	Описание	SEQ ID NO
O40	ATTAACCCUCCAGGCATCAAATAAACGAAAGGC	Backbone.for	100
O79	ATTTGCGCAUCACCAATCAAATTCACGCGGCC	Backbone.rev	101
O261	ATGCGCAA AUGCGGCACGCCTTGCAGATTACG	PglpF.for	102
O262	AGCTGTTUCCTCCTTGGTTAATGTTTGTGTATGC G	PglpF.rev	103
O68	ATGCGCAAUTGTGAGTTAGCTCACTCATTAG	Plac.for	84
O113	AGCTGTTUCCTCCTTAGGTACCCAGCTTTTGTTC C	Plac.rev	117

Таблица 9: Гетерологичные гены, включенные в экспрессионные кассеты, содержащие последовательность промотора и последовательность искусственной ДНК (i) (SEQ ID NO: 70), чтобы обеспечить микробную продукцию НМО

Ген	Источник генов	Номер доступа	Функция
futC	Helicobacter pylori 26695	EF452503	$\alpha$ -1,2-фукозилтрансфераза
lgtA	Neisseria meningitidis 053442	CP000381	$\beta$ -1,3-N-ацетилглюкозаминилтрансфераза
galT	Helicobacter pylori 26695	AE000511	$\beta$ -1,4-галактозилтрансфераза
galT	Helicobacter pylori	Гомологичный	$\beta$ -1,3-галактозилтрансфераза

K	43504	BD182026	
neuA	Campylobacter jejuni ATCC43438	AF400048	СМР-Neu5A-синтетаза
neuB	Campylobacter jejuni ATCC43438	AF400048	Синтетаза сиаловой кислоты
neuC	Campylobacter jejuni ATCC43438	AF400048	GlcNAc-6-фосфат-2-эпимераза
nst	Neisseria meningitides L3 MC58	U60660	$\alpha$ -2,3-сиалилтрансфераза
Pd2	Photobacterim damsela JT0160	BAА25316.1	$\alpha$ -2,6-сиалилтрансфераза

Таблица 10: Праймеры, которые были использованы для амплификации представляющих интерес гетерологичных генов и обеспечения совместимости полученных продуктов ПЦР с плазмидными каркасами, чтобы обеспечить конструирование донора

Идентификатор олигонуклеотида ID	Последовательность олигонуклеотида	Описание	SEQ ID NO
MP452	AAACAGCUATGATCTCTGTCTACATCATC AGTCTG	galTK_opt for	105
MP453	AGGGTTAAUTGCGCGTTAGACTTCTTTTCG GGGTTTTCA	galTK_opt rev	106
O123	AAACAGCUATGGCGTTCAAAGTGGTCCAA ATC	futC_opt for	107
O124	AGGGTTAAUTGCGCGTTAGCCCAGCGCGT TATATTTCTG	futC_opt rev	108
O142	AAACAGCUATGCAACCGCTGGTCTCCGTG C	lgtA_opt for	109
O143	AGGGTTAAUTGCGCGTTAACGGTTTTTCA GCAGGCGG	lgtA_opt rev	110
O342	AAACAGCUATGTCAAAAAGTCGCTCTCATC ACCGG	CA for	111
O126	AGGGTTAAUTGCGCGTTACTCGTTCAGCA	CA rev	112

	ACGTCAGC		
O144	AAACAGCUATGCGTGTCTTCGCCATTTCT C	galT_opt for	113
O145	AGGGTTAAUTGCGCGTTAGACGAATTGCC AGTATTTTCAGG	galT_opt rev	114
O95	AAACAGCUATGGAACGTAACGCCGTGAG CCTGC	nst_opt for	115
O93	AGGGTTAAUTGCGGCTTAGTTTTTATCGTC AAAGGTCAG	nst_opt rev	116
O26	AAACAGCUATGTGCAATAGCGATAACACC	Pd2_opt for	98
O27	AGGGTTAAUTGCGCGTTAGGCCCAGAACA GAACATC	Pd2_opt rev	99

#### *Конструирование плазмид*

Была создана плаزمида, содержащая два сайта эндонуклеазы I-SceI, разделенных двумя фрагментами ДНК оперона gal (необходим для гомологичной рекомбинации в galK), и последовательность терминатора транскрипции T1 (pUC57::gal) (GeneScript). Последовательности ДНК, используемые для гомологичной рекомбинации в gal опероне, покрывали пары оснований 3.628.621-3.628.720 и 3.627.572-3.627.671 в последовательности Escherichia coli K12 MG155 полного генома GenBank: ID: CP014225.1. Вставка с помощью гомологичной рекомбинации приведет к удалению 949 пар оснований galK и фенотипа galK.

Стандартные методики, хорошо известные в области молекулярной биологии, были использованы для конструирования праймеров и амплификации специфических последовательностей ДНК хромосомной ДНК Escherichia coli K-12 DH1. Такие стандартные методы, векторы и элементы можно найти, например, в: Ausubel et al. (eds.), Current Protocols in Molecular Biology (1995) (John Wiley & Sons); Sambrook, Fritsch, & Maniatis (eds.), Molecular Cloning (1989) (Cold Spring Harbor Laboratory Press, NY); Berger & Kimmel, Methods in Enzymology 152: Guide to Molecular Cloning Techniques (1987) (Academic Press); Bukhari et al. (eds.)

Плазмидный каркас 3,5 т.п.н., содержащий pUC57-sceI-galTK-T1-galKM-sceI, амплифицировали с использованием праймеров 040 (SEQ ID: 7) и 079 (SEQ ID: 8) (таблица 8), и фрагмент ДНК размером 3,3 т.п.н., содержащий lacZ, амплифицировали с хромосомной ДНК, выделенной из E. coli K-12 DH1.

Хромосомную ДНК, полученную из E. coli DH1, использовали для амплификации фрагмента ДНК размером 300 п.н., содержащего промотор PglpF (SEQ ID NO: 57), с использованием олигонуклеотидов 0261 (SEQ ID NO: 102) и 0262 (SEQ ID NO: 103), или промотор Plac (SEQ ID NO: 11) - с использованием олигонуклеотидов 068 и 0113 (SEQ ID NO: 84 и 117) (таблица 8). Аналогично фрагменту ДНК размером 107 п.н.,

содержащему промотор Plac, Plac\_org, фрагмент ДНК размером 182 п.н., содержащий промотор PglpA, PglpA\_org, фрагмент ДНК длиной 190 п.н., содержащий промотор PglpD, PglpD\_org, фрагмент ДНК размером 245 п.н., содержащий промотор PglpT, PglpT\_org, фрагмент ДНК размером 300 п.н., содержащий промотор PglpF, PglpF\_org, амплифицировали из генома *E. coli* DH1.

Шестнадцатинуклеотидная последовательность, расположенная перед сайтом начала трансляции glp-промоторов (SEQ ID NO: 5-8 - glpF, A, D и T соответственно) (включающая сайт связывания рибосомы), была изменена с помощью ПЦР в исходных (org) промоторных фрагментах. Аналогичным образом, последовательность Шайна-Дальгарно в экспрессирующем элементе PglpF была модифицирована с использованием праймеров путем введения специфических модификаций в олигонуклеотиды, используемые для амплификации фрагментов ДНК, в результате чего образуются промоторные экспрессионные элементы PglpF\_SD1, PglpF\_SD2, PglpF\_SD3, PglpF\_SD4, PglpF\_SD5, PglpF\_SD6, PglpF\_SD7, PglpF\_SD8, PglpF\_SD9 и PglpF\_SD10 (SEQ ID NO: 13-22, соответственно).

Укорачивание 5'-конца экспрессирующего элемента PglpF проводили с использованием специфических праймеров, дающих промоторные экспрессионные элементы  $\Delta$ 15PglpF,  $\Delta$ 140PglpF,  $\Delta$ 165PglpF и  $\Delta$ 198PglpF (SEQ ID NOs:29-32).

Все ПЦР-фрагменты были очищены, и плазмидный каркас, промоторный элемент (Plac\_org, Plac, PglpF\_org, PglpF, PglpT\_org, PglpT, PglpA\_org, PglpA, PglpD\_org, PglpD, PglpF\_SD1, PglpF\_SD2, PglpF\_SD3, PglpF\_SD4, PglpF\_SD5, PglpF\_SD6, PglpF\_SD7, PglpF\_SD8, PglpF\_SD9, PglpF\_SD10,  $\Delta$ 15PglpF,  $\Delta$ 140PglpF,  $\Delta$ 165PglpF или  $\Delta$ 180PglpF) и lacZ были клонированы, трансформированы в клетки TOP10 и отобраны на планшетах с LB, содержащей 100 мг/мл ампициллина и 0,2% глюкозы. Сконструированные плазмиды (см. таблицу 10) были очищены. Последовательность промотора и 5'-конец гена lacZ подтверждали секвенированием ДНК (MWG Eurofins Genomics).

Плазмидные каркасы на основе pTOP0 (ThermoFisher Scientific) или любой другой плазмиды могут быть получены аналогично тому, как описано выше. Все сконструированные плазмидные каркасы содержали два специфических фрагмента ДНК, гомологичных *Escherichia coli* K-12 DH1, используемых для гомологичной рекомбинации. Таким образом, генетическая кассета, содержащая промотор PglpF или Plac, любой представляющий интерес ген и терминатор транскрипции T1, была вставлена специфично в геном *Escherichia coli*. Конструирование плазмид, используемых для рекомбинации, было выполнено с использованием стандартных методов клонирования.

Последовательности ДНК гетерологичных генов, кодирующих гликозилтрансферазы или другие представляющие интерес ферменты, были оптимизированы по кодонам и синтезированы с помощью Genescript. Любой представляющий интерес ген, например, гены хозяина или гетерологичные гены, lgtA, galT, galTK, futC, neuA, neuB, neuB, nst или Pd2 (таблица 9), могут быть амплифицированы с помощью ПЦР с использованием соответствующих праймеров, охватывающих

стартовый кодон, ATG, и стоп-кодон, TAA, гена (таблица 10). Например, ген Pd2 был амплифицирован с использованием праймеров 026 (SEQ ID NO: 98) и 027 (SEQ ID NO: 99), тогда как ген nst был амплифицирован с использованием праймеров 095 и 093 (SEQ ID NO: 115 и 116) (таблица 9). Для конструирования донорских плазмид с этими генами использовали следующую процедуру: плазмидный каркас 3,5 т.п.н., содержащий PglpF (SEQ ID NO: 12), амплифицировали с использованием pMAP205 в качестве матрицы. Кодированные последовательности гена Pd2 из *Photobacterium damsela* (JT0160) (см. Drouillard et al. 2010. Carbohydrate Research. 345: 1394-1399. *Efficient synthesis of 6'-sialyllactose, 6,6'-disialyllactose, and 6'-KDO-lactose by metabolically engineered E. coli expressing a multifunctional sialyltransferase from the Photobacterium sp. JT-ISH-224*) и гена nst из *Neisseria meningitidis* (MC58) (см. Fierfort and Samian. 2008. J. Biotech. 134: 261-265. Genetic engineering of *Escherichia coli* for the economical production of sialylated oligosaccharides) были оптимизированы по кодонам для экспрессии в *E. coli* и синтезированы с помощью Genescript. Гены Pd2 и nst, были клонированы в плазмидные каркасы, как описано выше, что позволило получить плазмиды pMAP216, pMAP228, pMAP99, pMAP391 (таблица 3). Плазмиды очищали, трансформировали в клетки TOP10 и отбирали на чашках с LB, содержащих 100 мкг/мл ампициллина и 0,2% глюкозы. Последовательность промотора и 5'-конец генов Pd2 nst были подтверждены секвенированием ДНК (MWG Eurofins Genomics).

В целом и для любого представляющего интерес гетерологичного гена все фрагменты ПЦР очищали, а плазмидный каркас, промоторный элемент glpF (SEQ ID NO: 54), синтетическую ДНК-последовательность (i) (70UTR/SEQ ID NO: 37/synDNA (i) - см. таблицу 1)) и представляющий интерес ген клонировали стандартным клонированием USER. Клонирование в соответствующую плазмиду может быть выполнено с использованием любой стандартной методики клонирования ДНК. После клонирования ДНК трансформировали в клетки TOP10 и отбирали на чашках с LB, содержащих 100 мкг/мл ампициллина (или 50 мг/мл канамицина в случае конструкций на основе pTOP0) и 0,2% глюкозы. Сконструированные плазмиды были очищены, и последовательность промотора и 5'-конец представляющего интерес гена были проверены секвенированием ДНК (MWG Eurofins Genomics).

#### *Конструирование штаммов*

Вставка промоторных экспрессионных элементов, слитых с репортерным геном или рекомбинантным геном, была выполнена методом наполнения генов (Gene Gorging) по существу, как описано в Herring et al (Herring, C.D., Glasner, J.D. and Blattner, F.R. (2003). Gene (311). 153-163). Коротко, донорскую плазмиду и вспомогательную плазмиду совместно трансформировали в MDO и отбирали на планшетах с LB, содержащих 0,2% глюкозы, ампициллина (100 мкг/мл) или канамицина (50 мг/мл) и хлорамфеникола (20 мкг/мл). Отдельную колонию инокулировали в 1 мл LB, содержащей хлорамфеникол (20 мкг/мл) и 10 мкл 20% L-арабинозы, и инкубировали при 37°C при встряхивании в течение 7-8 часов. Затем клетки высевали на чашки с M9-DOG и инкубировали при 37°C в течение

48 часов. Одиночные колонии, образовавшиеся на MM-DOG расштриховывали на планшетах с LB, содержащих 0,2% глюкозы, и инкубировали в течение 24 часов при 37°C.

Ожидалось, что для вставок в локусе galK колонии, которые казались белыми на чашках с агаром MacConkey-галактоза и чувствительными к ампициллину и хлорамфениколу, потеряют донорскую и вспомогательную плазмиду и содержат вставку в локусы galK. Вставки в galK-сайтах идентифицировали с помощью ПЦР на колониях с использованием праймеров 048 и 049, расположенных вне локусов galK. Хромосомную ДНК очищали, локус galK амплифицировали, используя праймеры 048 и 049, и вставленную ДНК проверяли секвенированием (Eurofins Genomics, Германия). Штаммы MAP1365, MAP1366, MAP1367, MAP1368 и MAP1370 были сконструированы с использованием донорских плазмид pMAP689, pMAP690, pMAP691, pMAP693 и pMAP695, в результате чего были вставлены pglpABC, pglpD, pglpTQ or pglpFKX или plac, слитые с lacZ, соответственно, в локус galK в E. coli MDO.

Штаммы MAP1025, MAP1026, MAP1027 и MAP808 были сконструированы с использованием донорских плазмид pMAP431, pMAP432, pMAP433 и pMAP205, соответственно, что дало вставки PglpA, PglpD, PglpT и PglpF, где 16 пар оснований выше сайта начала трансляции были изменены на 5'-CAAGGAGGAAACAGCT-3' (SEQ ID NO: 10), слитые с lacZ и вставленные в локус galK E. coli MDO.

Штаммы MAP1176 к MAP1185 были сконструированы с использованием донорских плазмид pMAP486-pMAP495, соответственно, в результате получая вставки PglpF\_SD1, PglpF\_SD2, PglpF\_SD3, PglpF\_SD4, PglpF\_SD5, PglpF\_SD6, PglpF\_SD7, PglpF\_SD8, PglpF\_SD9 and PglpF\_SD10, слитые с lacZ и вставленные в локусы galK E. coli MDO. Модификации, внесенные в последовательность Шайна-Дальгарно экспрессионного элемента PglpF, показаны на фиг. 6А и перечислены в таблице 1.

Штаммы MAP1010-9, MAP1010-11, MAP1010-13, MAP1010-17, MAP1010-19 и MAP1010-20 были сконструированы с использованием плазмидного препарата, созданного с использованием вырожденного праймера. Модифицированный экспрессирующий элемент PglpF, слитый с lacZ, идентифицировали секвенированием и слитые промотор-lacZ вставляли в локус galK E. coli MDO. Модификации, внесенные в область -10 экспрессионного элемента PglpF, показаны на фиг. 7А и перечислены в таблице 1.

Штаммы MAP1210, MAP1211, MAP1086 и MAP1209 были сконструированы с использованием донорских плазмид pMAP457, pMAP462, pMAP463 и pMAP537, что давало вставку экспрессирующего элемента PglpF с удаленными 25, 150, 175 или 190 парами оснований на 5'-конце. Укороченные версии экспрессирующих элементов PglpF были слиты с lacZ и вставлены в локусы galK E. coli MDO. Модификации экспрессионных элементов PglpF перечислены в таблице 1.

Штамм MAP1356 был сконструирован путем двойной рекомбинации, как описано Sharan et al (2009). Nat. Protoc. 4(29: 206-223). Ген glpR в MAP808 был заменен геном, устойчивым к канамицину, что давало MAP1356.

MAP700 и MAP710: штаммы были сконструированы с использованием хелперной плазмиды и донорской плазмиды, содержащей кассеты *nst* (на основе *PglpF*- или *Plac*-), что давало штаммы *E. coli* MDO, экспрессирующие *Nst*-гликозилтрансферазу.

MAP219 и MAP986: штаммы были сконструированы с использованием хелперной плазмиды и донорской плазмиды, содержащей кассеты *Pd2* (на основе *PglpF*- или *Plac*-), что давало штаммы *E. coli* MDO, экспрессирующие фермент *Pd2*.

MDO1 и MDO15: химически компетентные *Escherichia coli* K-12 (DH1) MDO трансформировали 2 плазмидами (таблица 10) для получения LNT и LNT соответственно. MP1497 - MP1499: *Escherichia coli* K-12 (DH1) MDO трансформировали хелперной плазмидой *pACBSR* и донорскими плазмидами, содержащими представляющие интерес генные кассеты (*PglpF*-*lgtA* or *PglpF*-*galT*), чтобы интегрировать данную гликозилтрансферазу в геном. Полученные в результате хозяева были впоследствии трансформированы соответствующей плазмидой, несущей маркер антибиотика (таблица 10).

MP166, MP245, MP1825, MP1920, MP2239, MP2374, MP2525: *Escherichia coli* K-12 (DH1) MDO последовательно трансформировали хелперной плазмидой *pACBSR* и донорскими плазмидами, содержащими представляющие интерес генные кассеты, чтобы обеспечить интеграцию соответствующих модификаций в его геном в экспериментах с использованием метода наполнения генов (таблица 10).

MAP265 и MAP425: *Escherichia coli* K-12 (DH1) MDO трансформировали хелперной плазмидой *pACBSR* и донорскими плазмидами, содержащими представляющие интерес генные кассеты, чтобы обеспечить возможность введения желаемых вставок в геноме хозяина методом наполнения генов. Ген *padC* был удален путем встраивания *galK* в локусы *padC*, а также методом наполнения генов. Наконец, штаммы трансформировали плазмидой *pMAP101* для создания штаммов MAP265 и MAP425.

MAP1200 и MAP1214: *Escherichia coli* K-12 (DH1) MDO трансформировали хелперной плазмидой *pACBSR* и донорскими плазмидами, содержащими представляющие интерес генные кассеты, чтобы обеспечить возможность введения желаемых вставок в геном хозяина методом наполнения генов (Herring, C.D., Glasner, J.D. and Blattner, F.R. (2003). *Gene* (311). 153-163) (таблица 10). Все вставки были проверены секвенированием (Eurofins Genomics, Германия).

#### *Ферментативный анализ: lacZ*

Активность  $\beta$ -галактозидазы оценивали, как описано ранее (см., например, Miller J.H. *Experiments in molecular genetics*, Cold spring Harbor Laboratory Press, NY, 1972). Кратко, клетки разводили в *Z*-буфере и пермеабилizировали додецилсульфатом натрия (0,1%) и хлороформом. Анализ проводили при 30°C. Образцы предварительно нагревали, анализ инициировали добавлением 200 мкл орто-нитрофенил- $\beta$ -галактозидазы (4 мг/мл) и останавливали добавлением 500 мкл 1 М  $\text{Na}_2\text{CO}_3$ , когда образец становился слегка желтым. Выход орто-нитрофенола впоследствии определяли как изменение оптической плотности при 420 нм. Удельная активность указана в единицах Миллера

[A420/(мин\*мл\*А600)]. Активности, перечисленные в таблице 7, представляют собой средние значения по меньшей мере из двух независимых экспериментов.

*Ферментативный анализ: экспрессия гетерологичных генов Pd2 и Δ29nst*

Штаммы MAP219, MAP986, MAP700 и MAP710 предварительно культивировали в 1 мл базальной среды (см. выше) с добавлением MgSO<sub>4</sub>, тиамин и глюкозы, при 34°C в течение 24 часов. 0,5 мл предварительной культуры инокулировали в 50 мл минимальной базальной среды с добавлением MgSO<sub>4</sub>, тиамин и 0,5% глицерина, и инкубировали при встряхивании при 28°C в течение 24 часов. 10 мл клеточных культур собирали в 50 мл центрифужные пробирки при 8000×g, при -10°C в течение 15 минут. Супернатант удаляли, а осадок клеток хранили при -80°C до использования.

Клеточные осадки ресуспендировали в 1 мл физиологически подобной среде 1-X (125 mM KCl, 25 mM K<sub>3</sub>PO<sub>4</sub>, 10 mM глутамата натрия, 0,001 mM CaCl<sub>2</sub>, 5 mM MgSO<sub>4</sub>, pH 7,5), содержащий 1x BugBuster (Merk Milipore). Образцы клеток лизировали ультразвуком 4 раза по 20 секунд при амплитуде 30%. Нерастворимый клеточный дебрис удаляли центрифугированием при 10 000×g в течение 10 минут при -10°C. В анализе *in vitro* 5 mM CMP-SA использовали в качестве донора, а 10 mM лактозу использовали в качестве акцептора. Донор, акцептор и лизат смешивали, а образцы отбирали по истечении 0, 5, 10, 20 и 30 минут реакции. Образцы кипятили в течение 10 минут и супернатант анализировали на 3'SL и 6'SL. Активность ферментов измеряли в mM продукции 6'SL или 3'SL в час и корректировали на OD клеток.

и вставка генетических кассет была сделана с помощью наполнения геннов, как описано ниже, но могла быть осуществлена любым другим методом с использованием гомологичной рекомбинации ДНК.

*Анализ в планшетах с глубокими лунками*

Одну колонию из планшета с LB предварительно культивировали в 1 мл минимальной базальной среды, содержащей глюкозу (0,5%), в 10 мл в 24-луночном планшете с глубокими лунками (Deep well plate, Axygen). Планшет герметизировали перед культивированием с помощью Hydrophobic Gas Permeable Adhesive Seal (Axygen) и инкубировали в течение 24 часов при 34°C со встряхиванием при 700 об/мин в орбитальном шейкере (Edmund Bühler GmbH). Плотность клеток культуры контролировали при 600 нм с использованием спектрофотометра S-20 (Воесо, Германия). 40 мкл ночной культуры использовали для инокуляции в 2 мл минимальной базальной среды, содержащей 0,01% глюкозы, 0,5% лактозы и 200 мкг SUN (Sigma). IPTG добавляли при необходимости. Глубоколуночные планшеты покрывали герметизирующей фольгой и инкубировали в течение 48 или 72 часов при 28°C в орбитальном шейкере при 700 об/мин. После инкубации измеряли OD<sub>600</sub> и планшет покрывали герметизирующей лентой для нагревания (Saveen Werner) и инкубировали в термомиксере в течение 1 часа при 100°C со встряхиванием при 400 об/мин. Клеточный лизат осаждали центрифугированием в течение 10 минут при 4000 об/мин. Концентрацию НМО в супернатанте определяли методами HPLC или HPAC.

**Пример 1. Клонирование 17 случайно выбранных регуляторных элементов, выделенных из E. coli.**

Для идентификации однокопийных экспрессионных кассет, эффективных для экспрессии в штамме-хозяине E. coli с хромосомной ДНК E. coli K-12 DH1, амплифицировали семнадцать фрагментов ДНК, содержащих промоторные элементы. Все промоторные элементы содержали сайты связывания регуляторов транскрипции, а также сайт связывания РНК-полимеразы (области -35, -10), сайты инициации транскрипции и 5'-концевую нетранслируемую последовательность, расположенную в 16 парах оснований выше кодона инициации трансляции. 16-нуклеотидную ДНК-последовательность (5'-CAAGGAGGAAACAGCT-3') (SEQ ID NO: 10), охватывающую сайт связывания рибосомы (включая сайт Шайна-Дальгарно), была введена на 3'-конце во все фрагменты ДНК с использованием специфичных для последовательности олигонуклеотидов. Названия промоторов, длины фрагментов промоторов и олигонуклеотиды, использованные для амплификации и введения SEQ ID NO: 10, перечислены в таблице 11 ниже:

Промоторный элемент	Длина фрагмента (п.о.), амплифицируемого с DH1	Название олигонуклеотида	Последовательность олигонуклеотида	SEQ ID NO
PacnB	334	O350	ATGCGCAAUCGGATCTCAAGGAA ATCGCAATGG	SEQ ID NO:60
		O351	AGCTGTTUCCTCCTTGCTCATTGTC ATAGTGCGGCAGG	SEQ ID NO:61
Pactp	134	O354	ATGCGCAAUGCTGAATCCGAACA CCAGCGTC	SEQ ID NO:62
		O355	AGCTGTTUCCTCCTTGGCAGGACTT CATTATTAAGACGG	SEQ ID NO:63
PdcuB	584	O358	ATGCGCAAUACTCACTACTGAA ACAATATTGCC	SEQ ID NO:64
		O359	AGCTGTTUCCTCCTTGTAATCCTAT TTAAATTTTGGCTGAATAG	SEQ ID NO:65
Pdps	182	O274	ATGCGCAAUCCGAAAATTCCTGG CGAGCAG	SEQ ID NO:66

		O275	AGCTGTTUCCTCCTTGGATGTTATG TCCCAGTAATTAAC	SEQ NO:67	ID
PgalP	414	O360	ATGCGCAAUAUGAAGTAATCTTTCT TCACCTGCGTTC	SEQ NO:68	ID
		O361	AGCTGTTUCCTCCTTGGTTATTTTT TATTGTGAATTAAGATAGG	SEQ NO:69	ID
PgapA	355	O265	ATGCGCAAUUCAGTTCTTCTGCCG AAGGTT	SEQ NO:70	ID
		O266	AGCTGTTUCCTCCTTGTGTTAGTG AATAAAAGGTTGCC	SEQ NO:71	ID
PglpA	166	O378	ATGCGCAAUGAAAACATTCATAA ATTAATGTG	SEQ NO:72	ID
		O379	AGCTGTTUCCTCCTTGTTCGTTTTTT ACCATTTAGCCATAG	SEQ NO:73	ID
PglpD	173	O376	ATGCGCAAUGCGTCTCTTTCTT TACAAAC	SEQ NO:74	ID
		O377	AGCTGTTUCCTCCTTGTTCGTTAAA GCTCATAAATGTTCG	SEQ NO:75	ID
PglpF	284	O261	ATGCGCAAUGCGGCACGCCTTGC AGATTACG	SEQ NO:76	ID
		O262	AGCTGTTUCCTCCTTGGTTAATGTT TGTTGTATGCG	SEQ NO:77	ID
PglpT	229	O380	ATGCGCAAUCCATTTAGCCATAG TAAAAACATG	SEQ NO:78	ID
		O381	AGCTGTTUCCTCCTTGCCGTGGTCT TATTTATGATTAAC	SEQ NO:79	ID
PkatE	254	O270	ATGCGCAAUGCGCGGGTCCGTG CGTGGG	SEQ NO:80	ID
		O271	AGCTGTTUCCTCCTTGATTTATTAC TGAAAGGGCCGC	SEQ NO:81	ID
PkatG	254	O272	ATGCGCAAUGTGATCACAAATTT TAAACAG	SEQ NO:82	ID
		O273	AGCTGTTUCCTCCTTGACAGTGTTA CCGTTACGATAC	SEQ NO:83	ID

Plac	91	O68	ATGCGCAA AUTGTGAGTTAGCTCA CTCATTAG	SEQ ID NO:84
		O268	AGCTGTTUCCTCCTTGAAATTGTTA TCCGCTCACAA	SEQ ID NO:85
Pmlc	134	O257	ATGCGCAA AUGAATGCTCTCAGGT GAGGG	SEQ ID NO:86
		O258	AGCTGTTUCCTCCTTGTTTCGCGCT CCGAAATAATC	SEQ ID NO:87
PpoxB	184	O366	ATGCGCAA AUCCGAAATCGCTGAA GGTTACGTAC	SEQ ID NO:88
		O367	AGCTGTTUCCTCCTTGAATGTGATA ACGGTAACAAGTTTAG	SEQ ID NO:89
PptsG	384	O255	ATGCGCAA AUGGCTGTGTTGAAAG GTGTTG	SEQ ID NO:90
		O256	AGCTGTTUCCTCCTTGAGTATGGGT GCTTTTTTTACG	SEQ ID NO:91
PptsH	382	O259	ATGCGCAA AUGAATTGCAACAGTA ATGCCAG	SEQ ID NO:92
		O260	AGCTGTTUCCTCCTTGATAGGTTTA GTGTTGTGGAAC	SEQ ID NO:93

Из клонированных семнадцати фрагментов ДНК (17-промоторов), все они были идентичны на 3'-конце (SEQ ID NO: 10), все были слиты с лишенным промотора геном *lacZ*, связанным с последовательностью терминатора транскрипции, T1. Единственная копия экспрессионной кассеты Promoter-SEQ ID NO:10-*lacZ*-T1, была интегрирована в хромосомную ДНК. Активность экспрессии гена *lacZ*, введенного в единственном экземпляре, измеряли как активность  $\beta$ -галактозидазы. Активность *lac*-промотора измеряли в присутствии IPTG. Результаты экспрессии репортерного гена из конструкций показаны на фиг.1А. Примечательно, что не все протестированные промоторы продемонстрировали увеличение активности после замены нативного 16-нуклеотидного фрагмента выше кодона инициации трансляции, содержащего сайт связывания рибосомы (RBS) для SEQ ID NO: 10. Замечательная активность наблюдалась для промотора *glpF*, активность для промоторов *glpA* и *D* была также значительной, но активность большинства протестированных промоторов либо не изменялась вообще, либо не увеличивалась в значительной степени (что можно было ожидать в соответствии с Meynial-Salles I, et al (2005) *Appl Environ Microbiol* 71:2140-2144; и WO 03/08960). На фиг.1В представлены данные об экспрессии репортерного гена из конструкций, включающих три репрезентативных промотора из фиг.1А (*PglpF*, *PglpA* и *PglpT*) и любую

нативную 16-нуклеотидную последовательность, полученную из 5'UTR ДНК выше соответствующего гена (т.е. *glpF*, *glpA* и *glpT*), или SEQ ID NO: 10.

На фиг.2 показано сравнение уровней экспрессии репортерного гена с промоторной последовательности, выделенной из оперонов *glp*: *glpFKX*, *glpABC*, *glpTQ* и *glpD*. Все промоторы негативно регулируются репрессором *GlpR*. Все клонированные фрагменты ДНК содержали сайты связывания ДНК для cAMP-CRP плюс ряд сайтов связывания других регуляторных элементов, таких как один или несколько сайтов связывания репрессора *GlpR*. Схематический вид клонированных промоторных элементов показан на фиг.2 и 3. Клонирование проводили, как описано выше.

### **Пример 2. Уровень экспрессии с единственной копии слитых конструкций P*glp-lacZ***

Плазмиду для изучения промоторов, содержащую ген *lacZ* без промотора, использовали в качестве системы клонирования для идентификации промоторных элементов *E. coli*, которые могли бы поддерживать высокую и регулируемую экспрессию белка. Уровни экспрессии *lacZ* определяли как с единственной копии P*glp-lacZ*, интегрированной в хромосомную ДНК, так и с плазмиды с большим количеством копий (как описано в примере 11). Делеция  $\Delta$ *lacZM15* в гене *lacZ* в *E. coli* MDO не может продуцировать активный фермент  $\beta$ -галактозидазу и поэтому использовалась в качестве фонового штамма при скрининге. В качестве положительного эталона для экспрессии *lacZ* был использован промоторный элемент P*lac*. Промоторные элементы, происходящие из оперонов *glp*, *glpTQ*, *glpACB*, *glpD* и *glpFKX*, или P*lac* были слиты с лишенным промотора геном *lacZ* и вставлены в геном *Escherichia coli* с помощью сайт-специфической рекомбинации, что привело к получению штаммов MAP1367, MAP1365, MAP1366, MAP1368 и MAP1370, соответственно. Все выделенные промоторные фрагменты ДНК, *pglpTQ*, *pglpACB*, *pglpD*, *pglpFKX* и *plac*, слитые с *lacZ*, могут экспрессировать фермент  $\beta$ -галактозидазу, и активность фермента была измерена, как показано на рисунке 4. Следует обратить внимание, что активность промотора P*lac* была измерена в присутствии IPTG.

В другой серии экспериментов нуклеотидная последовательность из 16 пар оснований, расположенная перед сайтами начала трансляции в генах *glpF*, *glpA*, *glpD*, *glpT* и *lacZ*, включенных в экспрессионные кассеты, содержащие *pglpFKX*, *pglpACB*, *pglpD* и *pglpTQ*, была заменена последовательностью 5'-CAAGGAGGAAACAGCT-3' (SEQ ID NO: 10), что дало штаммы MAP808, MAP1025, MAP1026 и MAP1027, соответственно. Модификация 16 пар оснований перед сайтом начала трансляции в конструкциях, содержащих промоторные элементы P*glpF*, P*glpA* и P*glpT* и оригинальный ДНК-фрагмент 5'UTR соответствующего гена *glp*, увеличивала экспрессию фермента  $\beta$ -галактозидазы приблизительно в 10000 раз для P*glpF*, в 2 раза для P*glpA* и P*glpT* (рис. 4). Неожиданно, но дополнительная замена исходной 54-нуклеотидной ДНК-последовательности 5'UTR, расположенной ниже точки инициации транскрипции в конструкциях P*glpA* и P*glpT* на SEQ ID NO: 36 из 5'UTR ДНК гена *glpF* (полученные конструкции ДНК P*glpA*\_70UTR (SEQ ID NO: 50) и P*glpT*\_70UTR (SEQ ID NO: 51)), давала значительное увеличение

экспрессии репортерного гена с этих конструкций (фиг.4), демонстрируя неожиданный синергетический эффект последовательности 70UTR на экспрессию гена.

Таблица 12: праймеры, используемые для конструирования PglpA\_70UTR и PglpT\_70UTR

PglpA_70UTR	O378	ATGCGCAAAUGAAAACATTCATAAATTAATGTG	SEQ ID NO:94
	O812	AGCTGTTUCCTCCTTGGTTAATGTTTGTGTATGCGT GAAAGTCACGGACCTCCACGATGCTTGTAGGCATCG CGCATATTCGCTCATAATTC	SEQ ID NO:95
PglpT_70UTR	O380	ATGCGCAAAUCCATTTAGCCATAGTAAAAACATG	SEQ ID NO:96
	O815	AGCTGTTUCCTCCTTGGTTAATGTTTGTGTATGCGT GAAAGTCACGGACCTCCACGATGCTTGTAGGCATGC CGCGATGTTAAGAAAAC	SEQ ID NO:97

**Пример 3. Уровень экспрессии с единственной копии слитых конструкций Pglp-lacZ является катаболически подавленным**

Поскольку известно, что экспрессия с промоторов glp является катаболически подавленной, измеряли уровень экспрессии со слитых конструкций Pglp-lacZ, содержащих SEQ ID NO: 10. Штаммы MAP808, MAP1025, MAP1026 и MAP1027 выращивали в присутствии или в отсутствии глюкозы, и определяли активность β-галактозидазы, кодируемой lacZ (таблица 13). Как в LB, так и в минимальной среде, уровень экспрессии значительно снижался в присутствии глюкозы.

Таблица 13: уровни экспрессии с единственных копий слитых конструкций Pglp-lacZ

Штамм	Промоторный элемент	LB (MU)	LB-глюкоза (MU)	Степень подавления (разы)	ММ-глицерин	ММ-глюкоза	Степень подавления (разы)
MAP808	PglpF	8,850±1,415	279±3	32	13,450±786	3,699±353	4
MAP1025	PglpA	3,732±234	72±70	51	5,677±154	1,874±332	3
MAP1026	PglpD	3,969±603	618±223	6	4,529±132	1,398±121	3

MAP102 7	PglpT	647±23	15±15	43	1,110±139	290±44	4
-------------	-------	--------	-------	----	-----------	--------	---

**Пример 4. Экспрессия с единственной копии PglpF-lacZ регулируется и зависит от источника углерода в носителе**

Чтобы определить, регулируется ли уровень экспрессии с PglpF-lacZ другими источниками углерода, кроме глюкозы, штамм MAP808, который содержит единственную копию слитой конструкции PglpF-lacZ, выращивали в среде LB с глюкозой и без нее, и в минимальной среде с использованием глицерина, сорбита, мальтозы или глюкозы в качестве единственного источника углерода. Высокий уровень экспрессии lacZ наблюдался в среде LB, однако добавление 0,2% глюкозы значительно снижало экспрессию lacZ (в 32 раза) в среде LB (таблица 13, фиг.5). Высокий уровень экспрессии lacZ в штамме MAP808 наблюдался в минимальной среде, когда глицерин или сорбит использовались в качестве единственного источника углерода, тогда как использование мальтозы или глюкозы в минимальной среде снижало экспрессию lacZ.

**Пример 5. Уровень экспрессии в конструкциях PglpF и T может быть значительно изменен при изменении сайта связывания рибосомы, содержащегося в SEQ ID NO: 10.**

Эффект модификаций рекомбинантной последовательности (SEQ ID NO: 10), содержащей сайт связывания рибосомы (последовательность Шайна-Дальгарно), в конструкциях, содержащих PglpF и PglpT (SEQ ID NO: 54 и 49), и последовательности 70UTR (SEQ ID NO: 37) на экспрессию генов определяли путем измерения уровня экспрессии репортерного гена (lacZ), экспрессируемого с различных конструкций PglpF и PglpT, содержащих варианты SEQ ID NO: 10, как показано на фиг.6. Уровень экспрессии lacZ изменялся значительно (снижался на 9%-95% или увеличивался на 60%) (фиг.6).

**Пример 6. Уровень экспрессии с PglpF можно изменить, изменив последовательность области -10.**

Нуклеотидные модификации вводили в область -10: 5'-ТААГТ-3' экспрессионного элемента PglpF длиной 310 п.н., что давало штаммы MAP1010-19, MAP1010-20, MAP1010-13, MAP1010-17, MAP1010-11 и MAP1010-9. Уровень экспрессии lacZ измеряли в каждом штамме, и результаты показали, что экспрессия lacZ снижалась на 16%-60%, показывая важность области -10 для изменения уровня экспрессии (фиг.7).

**Пример 7. Фрагмент в 120 п.н., содержащий PglpF, содержит функциональный промотор**

Активность PglpF исследовали путем укорачивания его на 15, 140, 165 или 180 пар оснований с 5'-конца, что привело к образованию штаммов MAP1210, MAP1211, MAP1086 и MAP1209, соответственно. Уровень экспрессии lacZ измеряли в каждом штамме, и результаты показали, что экспрессия с PglpF сохраняется, даже если длина последовательности PglpF была уменьшена с 300 пар оснований до 120 пар оснований (фиг.8).

**Пример 8. Удаление транскрипционного репрессора GlpR увеличивает**

### **экспрессию с элемента PglpF**

Известно, что промотор PglpF негативно регулируется белком-репрессором транскрипции GlpR. PglpF из 300 пар оснований содержит четыре сайта связывания для белка-репрессора GlpR (фиг.2). Для того, чтобы определить важность репрессора GlpR на транскрипцию с PglpF, ген репрессора glpR был отключен, что давало штамм MAP1356. Уровень экспрессии lacZ измеряли в отсутствие и в присутствии репрессора GlpR, и уровень экспрессии повышался почти на 20% в его отсутствии (фиг.9).

**Пример 9. Экспрессия рекомбинантных генов с использованием экспрессионного элемента PglpF зависит от источника углерода, используемого в питательной среде.**

Выращивание клеток в разных средах оказывает значительное влияние на уровень LacZ, экспрессируемого с PglpF (фиг.5). Чтобы проверить, могут ли гетерологичные гены экспрессироваться с промоторного элемента PglpF, связанного с последовательностью ДНК 70UTR, и влияет ли на экспрессию источник углерода, используемый в среде, были сконструированы штаммы MAP710 и MAP986, экспрессирующие транссиалидазу nst или Pd2 с PglpF в единственной копии, соответственно, и культивировались в минимальных средах, содержащих глюкозу или глицерин в качестве источника углерода. Когда глицерин использовали в качестве источника углерода, ферментативная активность была по меньшей мере в 7 раз выше, чем когда клетки выращивались на средах, содержащих глюкозу в качестве источника углерода (фиг.10), показывая, что рекомбинантные гены можно экспрессировать с использованием PglpF, и что эта экспрессия может регулироваться с помощью источник углерода, присутствующего в среде.

**Пример 10. Высокий уровень экспрессии рекомбинантных генов с использованием экспрессионного элемента PglpF по сравнению с использованием промоторного элемента Plac**

Экспрессию гетерологичных генов из слитых конструкций PglpF-70UTR-Pd2 и PglpF-70UTR-nst определяли для единичных копий, интегрированным в хромосому хозяина, с последующим измерением ферментативной активности экспрессированных белков. Ферментативную активность измеряли в клеточных лизатах, экспрессирующих Pd2 с промотора Plac (при индукции с IPTG) или с промотора PglpF (фиг.10). Продукция 6'SL была в 7 раз выше в клеточных лизатах, экспрессирующих Pd2 из PglpF, по сравнению с Plac. Аналогичным образом, ферментативная активность клеточного лизата, экспрессирующего сиаллилтрансферазу nst с промотора Plac (индуцируемую с помощью IPTG) или с промотора PglpF (фиг. 10), показала почти в 14 раз более высокую активность, когда nst экспрессировалась с PglpF, чем с Plac.

**Пример 11. Pglp-содержащие конструкции нуклеиновых кислот могут быть использованы для экспрессии в высококопийных плазидах.**

Все конструкции нуклеиновых кислот PglpF, PglpA и PglpT содержат фрагмент промоторной ДНК, выделенный из соответствующих оперонов glp, и синтетическую ДНК-последовательность, содержащую ДНК-последовательность 5'UTR, выделенную из

соответствующего промотора, которая была слита с SEQ ID NO: 10 и помещена выше сайта начала трансляции репортерного гена (*lacZ*). Конструкции были клонированы в высококопийную плазмиду. Уровни экспрессии *lacZ* измеряли как активность фермента  $\beta$ -галактозидазы в клетках, выращенных в среде LB в присутствии или в отсутствие глюкозы (таблица 14, фигура 11). Результаты в таблице 4 показывают, что Pglp поддерживает высокую экспрессию с плазмид с большим числом копий и что эта экспрессия регулируется катаболической репрессией в присутствии глюкозы (таблица 14, фиг. 11).

Таблица 14: Уровни экспрессии со слитых конструкций промотор-*lacZ* в плазмиде с большим числом копий

Плазида	Промоторный элемент	Исходный оперон	Активность в LB (MU)	Активность в LB+глюкоза (MU)
pMAP205	PglpF	glpFKX	51,245±4,560	6,664
pMAP431	PglpA	glpACB	27,377±549	457±469
pMAP433	PglpT	glpTQ	11,019±575	400±10

**Пример 12. Создание *Escherichia coli* для получения LNnT с использованием промотора PglpF.**

Штамм *Escherichia coli* K-12 (DH1) MDO можно изменять для эпизодической экспрессии представляющих интерес гетерологичных генов. Например, штамм MDO1 является 2-плазмидным штаммом с плазмидой со средним числом копий (30-40 копий на клетку), несущей ген *lgtA*, и плазмидой с высоким числом копий (300-500 копий на клетку) с геном *galT*. Промотор P<sub>lac</sub> контролирует экспрессию обоих гетерологичных генов в этих плаزمидах. В альтернативном варианте, гетерологичные гены могут быть интегрированы в геном штамма *Escherichia coli* K-12 (DH1) MDO для создания продуцируемых геномом производственных систем. Таким образом, среднекопийная плазида с геном *lgtA* заменяется одной геномной копией экспрессионной кассеты PglpF-*lgtA* в штамме MP1497, которая все еще несет высококопийную плазмиду с P<sub>lac</sub>-*galT*. В другом примере высококопийную плазмиду с геном *galT* заменяют одной геномной копией экспрессионной кассеты PglpF-*galT* в штамме MP1499, которая все еще несет среднекопийную плазмиду с P<sub>lac</sub>-*lgtA*. Как показано на фиг. 12А, сходные титры LNnT достигаются, когда экспрессия обоих гетерологичных генов, *lgtA* и *galT*, экспрессируется с помощью плазмидного промотора P<sub>lac</sub> (штамм MDO1) или когда один из генов экспрессируется с фрагмента с промотором PglpF, интегрированного в хромосомную ДНК, и второй ген экспрессируется с плазмидного промотора P<sub>lac</sub> (штаммы MP1497 и MP1499).

Штаммы MP2622 и MP1825 экспрессируют гены *lgtA* и *galT* с одной геномной копии под контролем промоторов P<sub>lac</sub> или PglpF, соответственно. Штамм MP166 имеет

три Plac-контролируемые геномные копии обоих генов. За исключением штамма MP2622, ген *lacI* удаляется из генома обсуждаемых в настоящем документе штаммов, а кассета CP6-galK вставляется в локус *lacI*. Из гистограмм на фиг. 12B очевидно, что экспрессия генов *lgtA* и *galT* с одной, контролируемой PglpF геномной копии (штамм MP1825) достаточна для достижения титров LNnT, которые значительно выше, чем титры, достигаемые при одиночных (штамм MP2622) или множественных (штамм MP166) Plac-контролируемых геномных копиях *lgtA* и *galT*, интегрированных в геном.

**Пример 13. Конструирование Escherichia coli для получения LNT с использованием промотора PglpF.**

Штамм Escherichia coli K-12 (DH1) MDO можно модифицировать для эпизодической экспрессии представляющих интерес гетерологичных генов. Штамм MD015 представляет собой 2-плазмидный штамм с плазмидой со средним количеством копий (30-40 копий на клетку), несущей ген *lgtA*, и плазмидой с большим количеством копий (300-500 копий на клетку) с геном *galTK*. Экспрессия обоих гетерологичных генов в этих плазмидах контролируется промотором Plac. В альтернативном варианте, гетерологичные гены могут быть интегрированы в геном штамма Escherichia coli K-12 (DH1) MDO для создания продуцируемых геномом продуцирующих систем. Таким образом, плазида со средним числом копий с геном *lgtA* заменяется одной копией экспрессионной кассеты PglpF-*lgtA* в штамме MP1498, которая все еще несет высококопийную плазмиду с Plac-*galTK* (фиг. 13A). В другом примере высококопийную плазмиду с геном *galTK* заменяют двумя геномными копиями экспрессионной кассеты PglpF-*galTK* в штамме MP1655, который все еще несет плазмиду со средним числом копий с Plac-*lgtA* (фиг. 13A). Как показано на фиг. 13A, сходные титры LNT могут быть достигнуты, когда экспрессия обоих гетерологичных генов, *lgtA* и *galTK*, контролируется плазмидным промотором Plac (штамм MD015), или когда экспрессия *lgtA* управляется кассетой PglpF-*lgtA*, вставленной в хромосомную ДНК хозяина, и плазмидный промотор Plac контролирует экспрессию гена *galTK* (штамм MP1498). Подобные титры продукта также могут быть достигнуты, когда две интегрированные в геном кассеты PglpF-*lgtA* управляют экспрессией гена *galTK*, а экспрессия *lgtA* находится под контролем плазмидного промотора Plac (штамм MP1655 на фиг. 13A).

Экспрессия трех геномных копий *lgtA* и двух *galTK* в штамме MP245 находится под контролем промотора Plac. Одна копия каждого гетерологичного гена экспрессируется с промотора PglpF в штамме MP1920. Ген *lacI* удаляется из генома этих штаммов путем введения кассеты CP6-galK в локус *lacI*. Из графиков на фиг. 13B очевидно, что экспрессия генов *lgtA* и *galTK* с одной, контролируемой PglpF геномной копии (штамм MP1920), достигает титров LNT, которые намного выше, чем достигаемые в варианте, когда множественные геномные копии *lgtA* и *galT*, контролируемые Plac (штамм MP245), интегрированы в хромосому.

**Пример 14. Конструирование Escherichia coli для получения LNFP-I с использованием промотора PglpF.**

Геномную систему продукции LNFP-I можно разработать путем интеграции соответствующих гетерологичных генов в геном штамма *Escherichia coli* K-12 (DH1) MDO. Например, штамм MP2239 экспрессирует гетерологичные гены *lgtA*, *galTK* и *futC* с одной геномной копии под контролем промотора *PglpF*, тогда как экспрессия генов колановой кислоты (CA) - *gmd*, *wcaJ* (*fcl*), *wcaH* (*gmm*), *weal*, *cpsB* (*manC*) и *cpsG* (*manB*) - контролируется с помощью *Plac*. Интеграция внегеномных копий генов пути синтеза колановой кислоты в штамме MP2239 под контролем промотора *PglpF* дает штамм MP2374. Ген *lacI* удаляется из генетического фона обоих штаммов. Заметное улучшение титра LNFP-I наблюдается, когда дополнительные копии генов пути синтеза колановой кислоты интегрируются в штамм MP2239 и экспрессируются с промотора *PglpF* (фиг. 14).

**Пример 15. Конструирование *Escherichia coli* для продукции 3'SL с использованием промотора *PglpF***

В штамме MAP425 промотор *Plac* был использован для экспрессии i) гетерологичного гена *nst* и кластера гетерологичных генов *neuBCA*, интегрированных в хромосому *Escherichia coli* K-12 (DH1) MDO, и ii) кластера гетерологичных генов *neuBCA* с высококопийной плазмиды (300-500 копий на клетку). В штамме MAP1214 промотор *PglpF* использовали для экспрессии гетерологичных генов *nst*, *neuA*, *neuB* и *neuC*, интегрированных в единичных копиях в хромосому *Escherichia coli* K-12 (DH1) MDO. Как показано на рисунке 15, сходные 3'SL-титры были достигнуты, когда *Plac* использовался для экспрессии кластера генов *neuBCA* с высококопийной плазмиды (штамм MAP425), и когда *PglpF* использовался для экспрессии *nst*, *neuA*, *neuB* и *neuC* с единичных интегрированных копий генов (штамм MAP1214) (фиг.15).

**Пример 16. Конструирование *Escherichia coli* для продукции 6'SL с использованием промотора *PglpF*.**

В штамме MAP265 промотор *Plac* использовали для экспрессии i) гетерологичного гена *Pd2*, введенного в хромосому *Escherichia coli* K-12 (DH1) MDO, и ii) кластера гетерологичных генов *neuBCA* с высококопийной плазмиды (300-500 копий). В штамме MAP1200 фрагмент промотора *PglpF* был использован для экспрессии гетерологичных генов *Pd2*, *neuA*, *neuB* и *neuC*, интегрированных в единичных копиях в хромосому *Escherichia coli* K-12 (DH1) MDO. Как показано на фиг.16, титр 6'SL был значительно выше, когда *PglpF* использовался для экспрессии *Pd2*, *neuA*, *neuB* и *neuC* с единичных интегрированных копий генов, чем когда *Plac* использовался для экспрессии кластера генов *neuBCA* с высококопийной плазмиды (штамм MAP265) (фиг.16).

**Пример 17. Конструирование *Escherichia coli* для продукции 2'FL с использованием промотора *PglpF*.**

Штамм FT18 представляет собой 2-плазмидный штамм с плазмидой со средним числом копий (30-40 копий на клетку), несущей гены пути биосинтеза колановой кислоты *gmd*, *wcaJ* (*fcl*), *cpsB* (*manC*) и *cpsG* (*manB*), выделенные из *Escherichia coli* K-12 DH1 и высококопийной плазмидой (300-500 копий на клетку) с геном *futC*. Промотор *Plac* контролирует экспрессию клонированных генов в этих плазмидах. Экспрессия генов *futC*

и генов биосинтеза колановой кислоты *gmd*, *wcaJ* (*fcl*), *cpsB* (*manC*) и *cpsG* (*manB*) в штамме MP965 находится под контролем промотора *PglpF*. Единственная копия генов *futC* и колановой кислоты экспрессируется с промотора *PglpF* в штамме MAP965. Результаты, представленные на фиг.17, показывают, что экспрессия генов *futC* и *CA* из генов биосинтеза колановой кислоты с одной, контролируемой *PglpF* геномной копии (штамм MAP965) позволяет получить почти такие же титры 2'FL, что и при экспрессии с множества копий генов, контролируемых *Plac* (фиг.17).

## ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Конструкция нуклеиновой кислоты, содержащая синтетическую некодирующую ДНК-последовательность (i), которая содержит первый фрагмент ДНК и второй фрагмент ДНК, где первый фрагмент ДНК представляет собой ДНК-последовательность, полученную из 5'-нетранслируемой области (5'-UTR) гена *glp* из *Escherichia coli*, и второй фрагмент ДНК представляет собой ДНК-последовательность CAAGGAGGAAACAGCT (SEQ ID NO: 10), или вариант указанной последовательности, и где первый фрагмент расположен выше второго фрагмента.

2. Конструкция нуклеиновой кислоты по п.1, дополнительно содержащая промоторную ДНК-последовательность (ii), где промоторная ДНК функционально связана с последовательностью синтетической некодирующей ДНК (i) и расположена перед первым фрагментом ДНК последовательности синтетической некодирующей ДНК (i).

3. Конструкция нуклеиновой кислоты по п.1 или 2, дополнительно содержащая кодирующую ДНК-последовательность (iii), где кодирующая ДНК-последовательность (iii) функционально связана с синтетической некодирующей последовательностью ДНК (i) и расположена ниже второго фрагмента синтетической некодирующей ДНК-последовательности (i).

4. Конструкция нуклеиновой кислоты по п.1, 2 или 3, где первый фрагмент ДНК синтетической некодирующей ДНК-последовательности (i) получен из ДНК-последовательности 5'UTR гена *glpF*, *glpA* или *glpD* *Escherichia coli*.

5. Конструкция нуклеиновой кислоты по п.1, 2 или 3, где первый фрагмент ДНК синтетической некодирующей ДНК-последовательности (i) содержит от примерно 5 до примерно 65 последовательных нуклеотидов после начала транскрипции гена *glp*.

6. Конструкция нуклеиновой кислоты по п.5, где ген представляет собой *glpF*.

7. Конструкция нуклеиновой кислоты по любому из пп. 1-6, где первый фрагмент ДНК представляет собой или содержит SEQ ID NO: 36, его фрагмент.

8. Конструкция нуклеиновой кислоты по любому из пп. 1-7, где указанная конструкция содержит вариант SEQ ID NO: 10, причем указанный вариант выбран из любой из SEQ ID NO: 13-22.

9. Конструкция нуклеиновой кислоты по любому из пп. 2-10, где последовательность промоторной ДНК (ii) происходит из последовательности промоторной ДНК, которая в природе инициирует транскрипцию оперона *glpFKX*, *glpABC*, *glpD* или *glpTQ* *Escherichia coli*.

10. Конструкция нуклеиновой кислоты по п.9, где промоторная ДНК-последовательность (ii) имеет последовательность, выбранную из любой из нуклеотидных последовательностей, указанных в SEQ ID NO: 12-35, 50-53, или она представляет собой фрагмент или вариант указанной последовательности.

11. Конструкция нуклеиновой кислоты по любому из пп. 1-10, где кодирующая ДНК-последовательность (iii) кодирует белок или РНК.

12. Конструкция нуклеиновой кислоты по п.11, где белок или РНК имеют

функцию, которая необходима или полезна для продукции одного или более олигосахаридов человеческого молока (НМО) в рекомбинантном организме.

13. Конструкция нуклеиновой кислоты по любому из пп. 1-12, где конструкция представляет собой экспрессионную кассету, которая интегрирована в геномную или плазмидную ДНК рекомбинантного организма.

14. Рекомбинантная клетка, содержащая одну или более нуклеотидных конструкций по любому из пп. 1-13.

15. Рекомбинантная клетка по п.14, где клетка представляет собой бактериальную клетку.

16. Рекомбинантная клетка, где бактериальная клетка представляет собой *E.coli*.

17. Рекомбинантная клетка по любому из пп. 14-16, где клетка содержит более одной копии одной и той же нуклеотидной конструкции.

18. Рекомбинантная клетка по любому из пп. 15-16, где клетка содержит по меньшей мере две разные нуклеотидных конструкции.

19. Рекомбинантная клетка по п.18, где нуклеотидные конструкции отличаются любой одной или обеими из последовательности промоторной ДНК (ii) и синтетической некодирующей последовательности ДНК (i) или/и кодирующей последовательности ДНК (iii).

20. Экспрессионная система, содержащая нуклеотидную конструкцию по любому из пп. 1-13 или рекомбинантную клетку по пп. 14-19.

21. Способ рекомбинантной продукции одной или более биологических молекул, включающий:

- получение нуклеотидной конструкции по любому из пп. 1-13;
- получение рекомбинантной клетки по любому из пп. 14-19;
- культивирование рекомбинантной клетки в условиях, позволяющих продуцировать одну или более биологических молекул;
- получение одной или более биологических молекул.

22. Способ по п.21, где по меньшей мере одна из биологических молекул представляет собой олигосахарид.

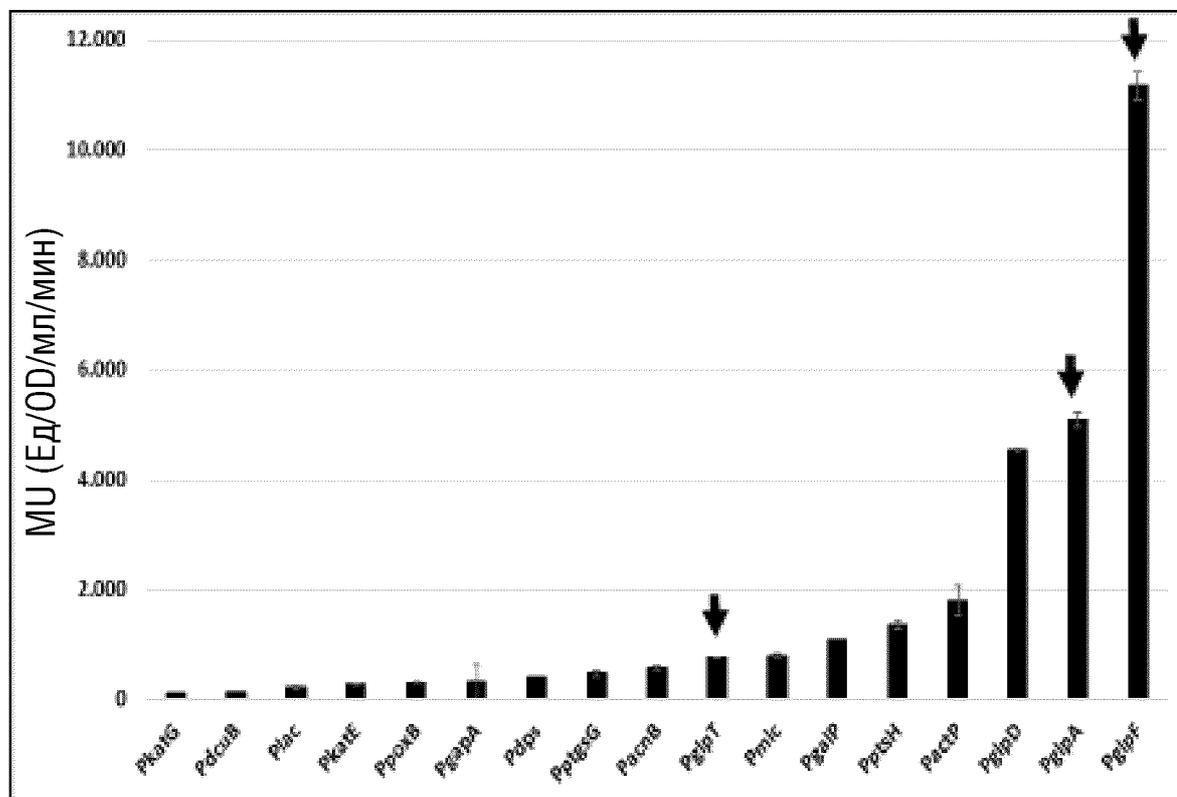
23. Способ по п.22, где по меньшей мере один из олигосахаридов представляет собой олигосахарид грудного молока (НМО).

24. Способ по п.23, где по меньшей мере одна НМО выбрана из следующей группы: лакто-N-тетраозы (LNT), лакто-N-неотетраозы (LNnT), лакто-N-неогексаозы (LNnH), пара-лакто-N-неогексаозы (pLNnH), пара-лакто-N-гексаозы (pLNH) и лакто-N-гексаозы (LNH), 2'-фукозиллактозы (2'-FL), лакто-N-фукопентазы I (LNFP-I), лакто-N-дифукогексаозы I (LNDFH-I), 3-фукозиллактозы (3-FL), дифукозиллактозы (DFL), лакто-N-фукопентазы II (LNFP-II), лакто-N-фукопентазы III (LNFP-III), лакто-N-дифукогексаозы III (LNDFH-III), фукозил-лакто-N-гексаозы II (FLNH-II), лакто-N-фукопентазы V (LNFP-V), лакто-N-дифукогексаозы II (LNDFH-II), фукозил-лакто-N-гексаозы I (FLNH-I), фукозил-пара-лакто-N-гексаозы I (FpLNH-1), фукозил-пара-лакто-N-неогексаозы II (F-

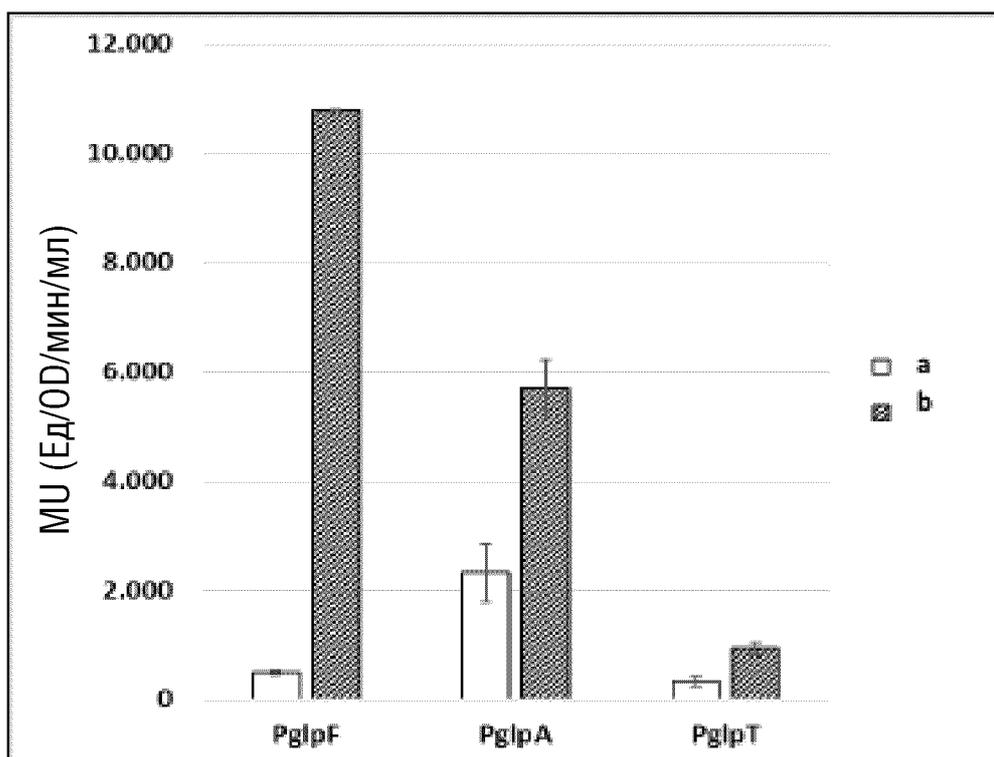
pLNnH II) и фукозил-лакто-N-неогексаозы (FLNnH), 3'-сиалиллактозы (3'-SL), 6'-сиалиллактозы (6'-SL), 3-фукозил-3'-сиалиллактозы (FSL), LST a, фукозил-LST a (FLST a), LST b, фукозил-LST b (FLST b), LST c, фукозил-LST c (FLST c), сиалил-лакто-N-гексаозы (SLNH), сиалил-лакто-N-неогексаозы I (SLNH-I), сиалил-лакто-N-неогексаозы II (SLNH-II) и дисиалил-лакто-N-тетраозы (DSLNT).

По доверенности

ФИГ.1А



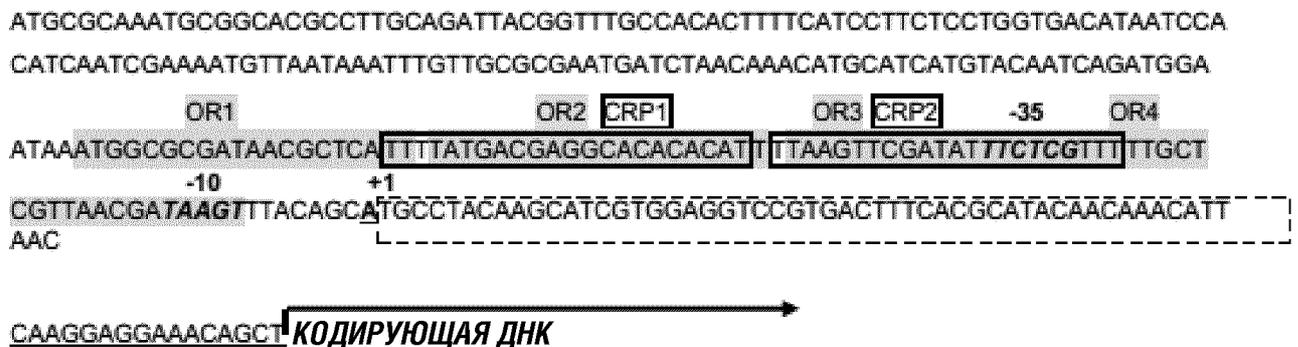
ФИГ.1В



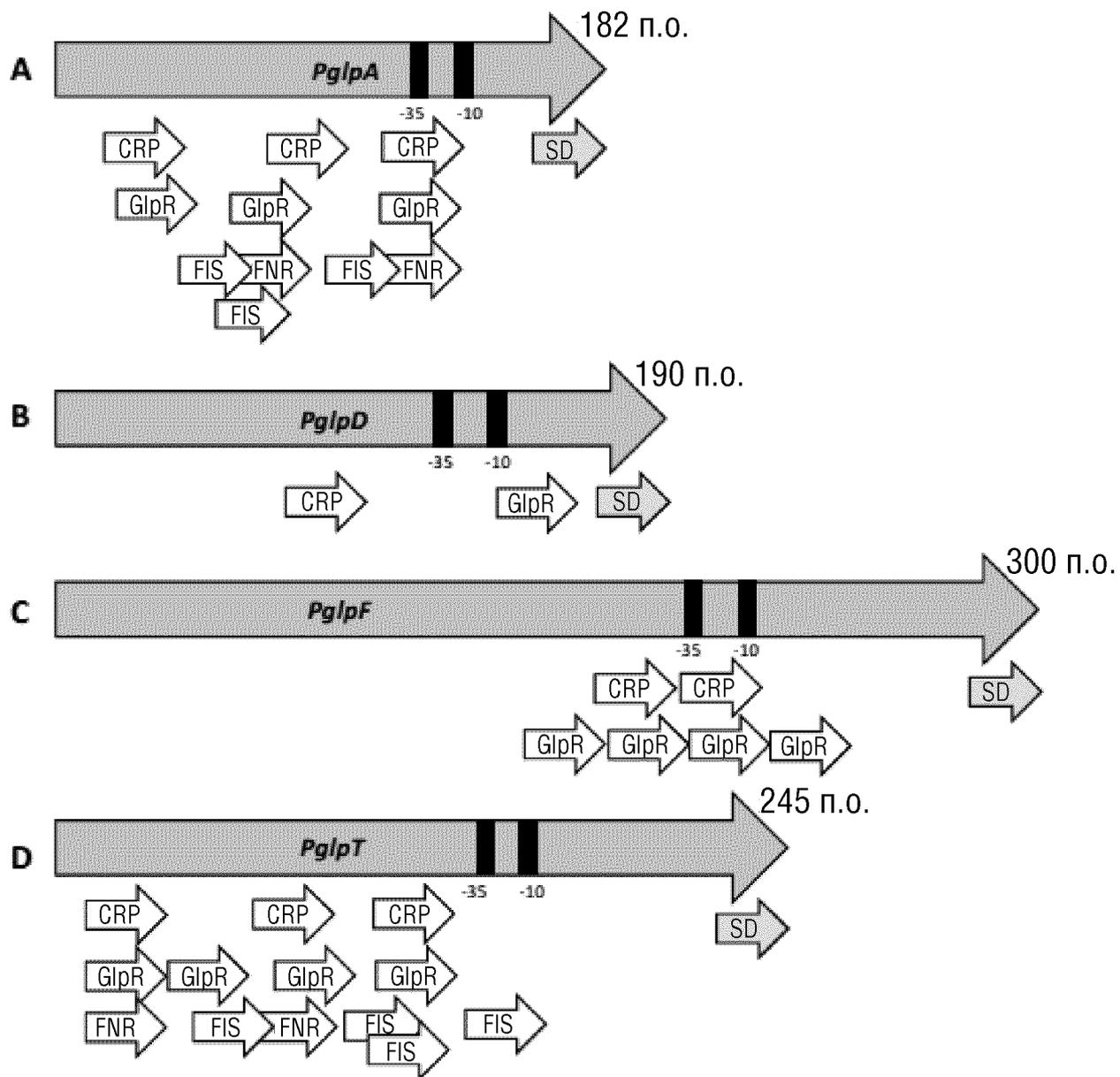
ФИГ.2А



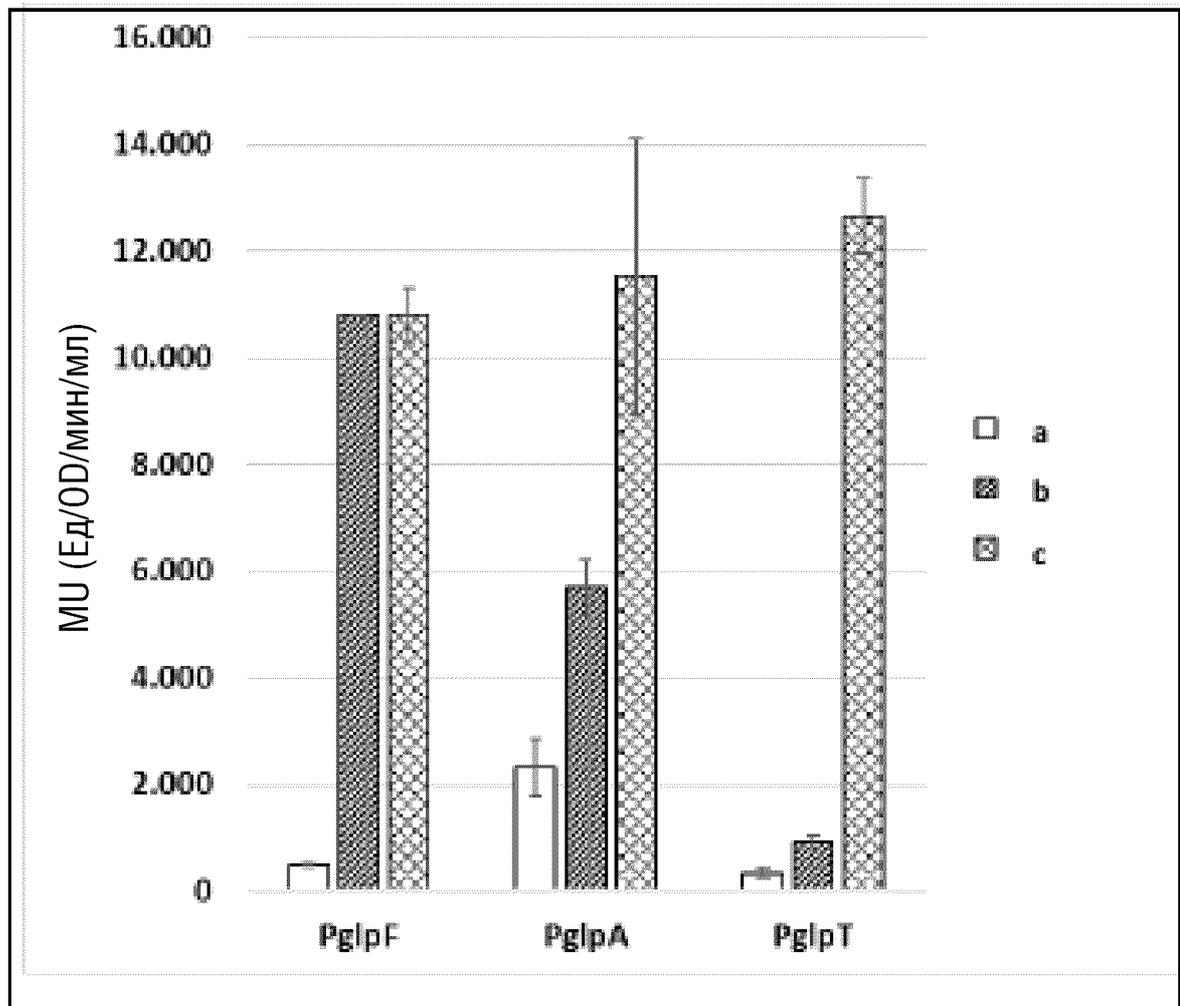
ФИГ.2В



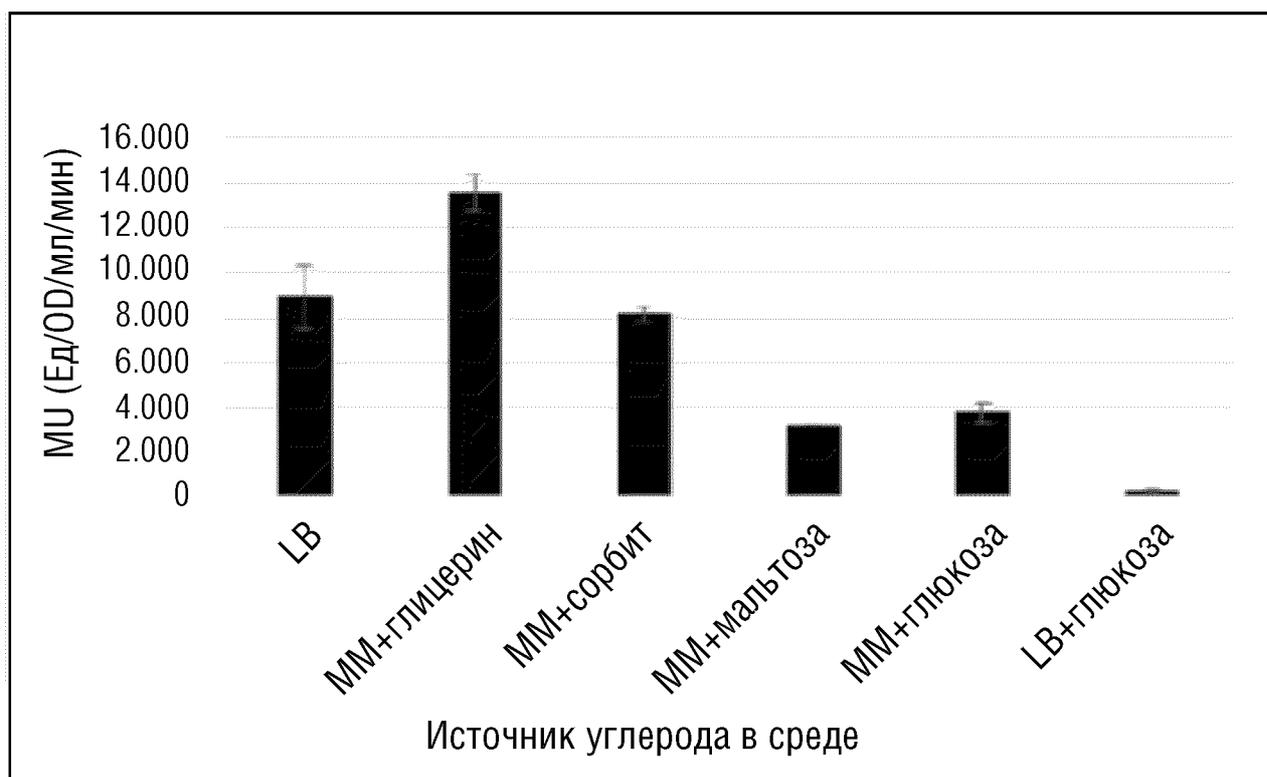
## ФИГ.3



ФИГ.4



ФИГ.5

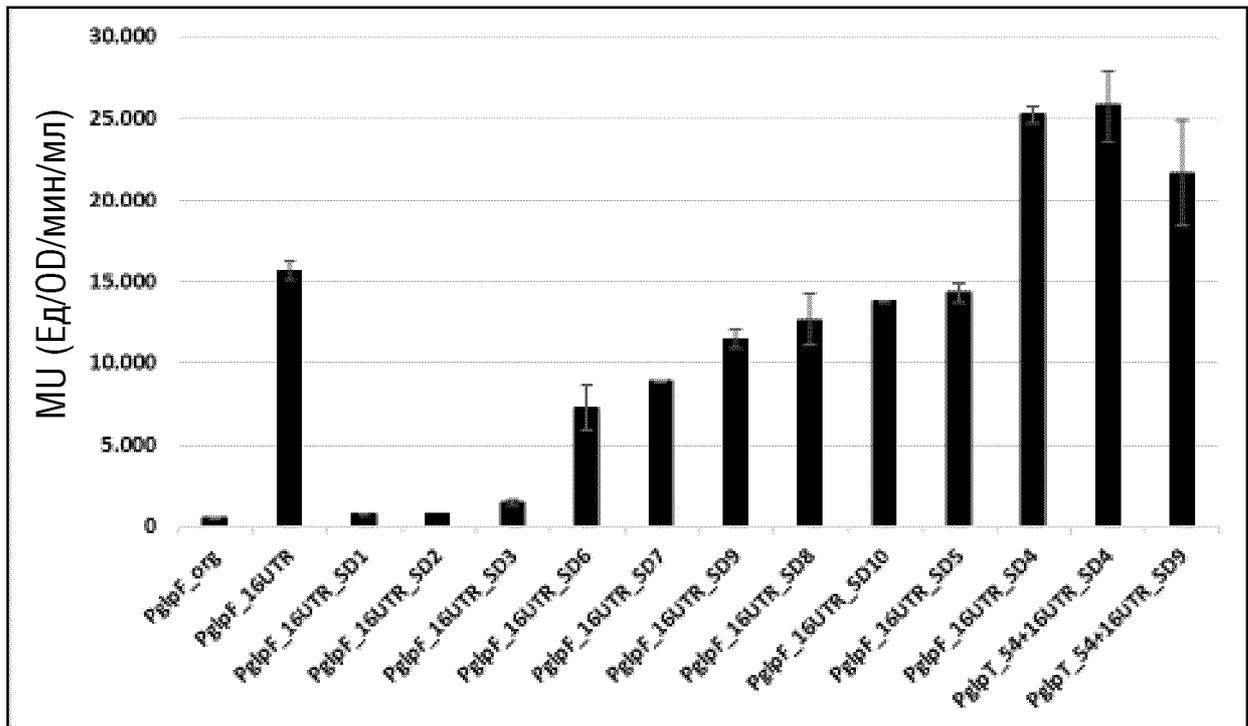


## ФИГ.6А

## Варианты SEQ ID NO:10

	5' - CAAGGAGGAAACAGCT- 3'	SEQ ID NO:10
SD1	5' ---- <u>ATTC</u> -----3'	SEQ ID NO:38
SD2	5' ---- <u>CGCA</u> -----3'	SEQ ID NO:39
SD3	5' ---- <u>A-CA</u> -----3'	SEQ ID NO:40
SD4	5' ---- <u>CT</u> -----3'	SEQ ID NO:41
SD5	5' ---- <u>CCGA</u> -----3'	SEQ ID NO:42
SD6	5' ---- <u>AGCT</u> -----3'	SEQ ID NO:43
SD7	5' ---- <u>AGCA</u> -----3'	SEQ ID NO:44
SD8	5' ---- <u>AGAA</u> -----3'	SEQ ID NO:45
SD9	5' ---- <u>A-GAA</u> -----3'	SEQ ID NO:46
SD10	5' ---- <u>CTGA</u> -----3'	SEQ ID NO:47

## ФИГ.6В

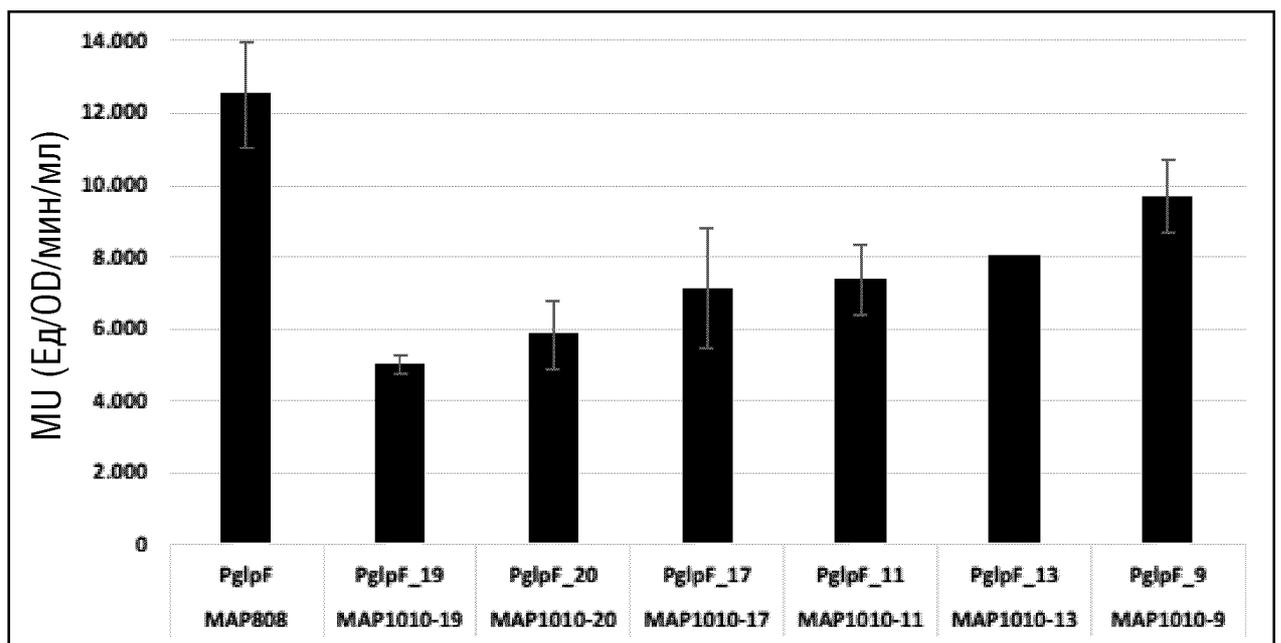


## ФИГ.7А

## Варианты SEQ ID NO:12, содержащие модификацию в боксе -10

Штамм	Промотор	Фрагмент ДНК-последовательности, содержащий модификацию		
		-35	-10	+1
MAP808	PglpF (SEQ ID NO: 12)	5' -		
		<u>TTCTCGTTTTTGCTCGTTAACGATAAGTTTACAGCA</u> -3'		
MAP1010-9	PglpF_9 (SEQ ID NO: 23)	5' -----	-TTAA-----	
		-3'		
MAP1010-11	PglpF_11 (SEQ ID NO: 24)	5' -----	-C--AA-----	
		-3'		
MAP1010-13	PglpF_13 (SEQ ID NO: 25)	5' -----	-TTCC-----	
		-3'		
MAP1010-17	PglpF_17 (SEQ ID NO: 26)	5' -----	-TGA-----	
		-3'		
MAP1010-19	PglpF_19(SEQ ID NO: 27)	5' -----	-G-AGC-----	
		-3'		
MAP1010-20	PglpF_20 (SEQ ID NO: 30)	5' -----	-C--A-----	
		-3'		

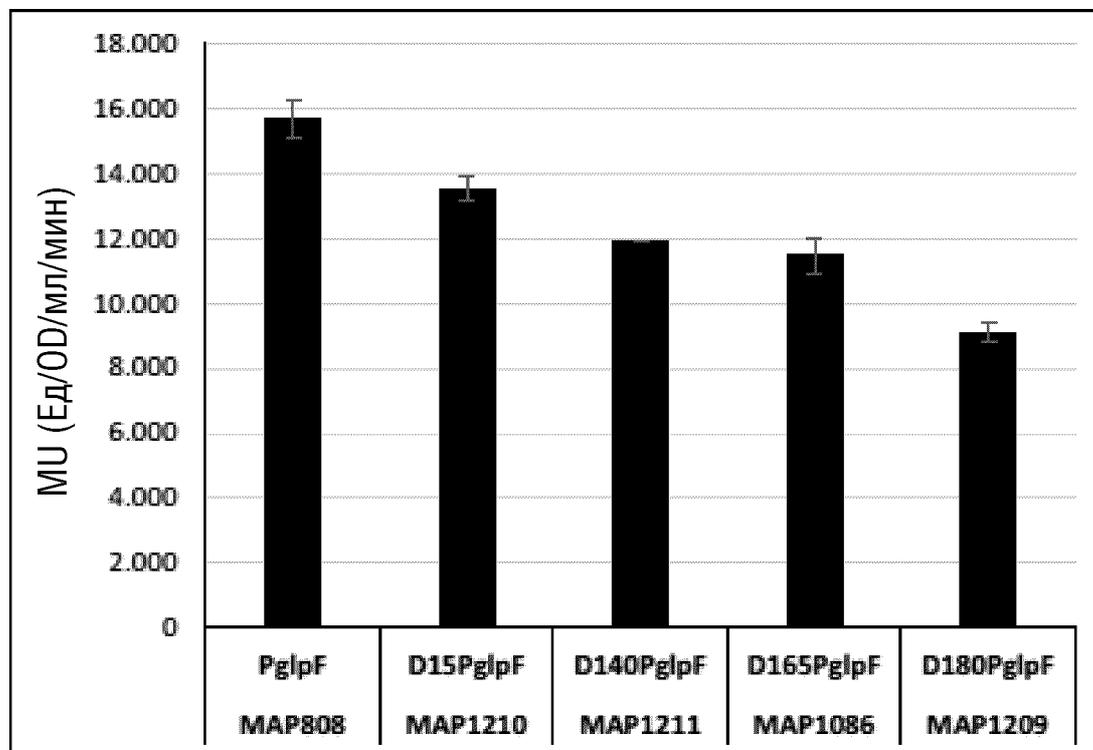
## ФИГ.7В



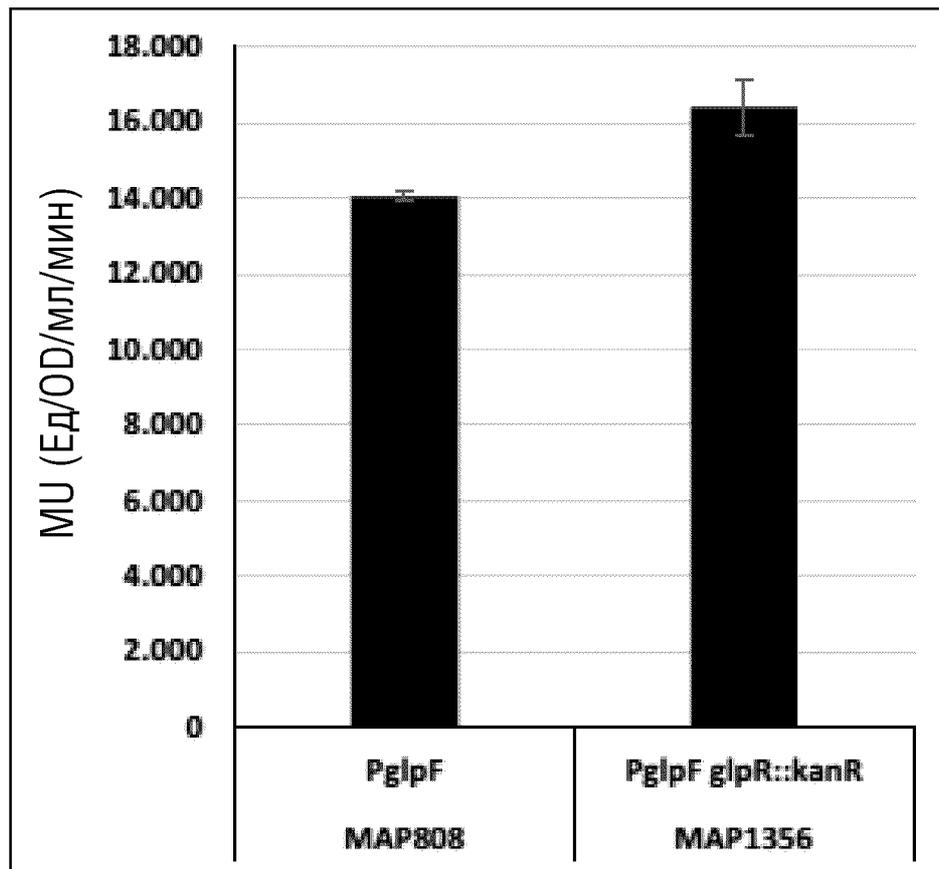
## ФИГ.8А

Штамм	Промотор	Длина промотора
MAP808	PglpF	300 nb (SEQ ID NO: 12)
MAP1210	D15PglpF	285 nb (SEQ ID NO: 29)
MAP1211	D140PglpF	160 nb (SEQ ID NO: 30)
MAP1086	D165PglpF	135 nb (SEQ ID NO: 31)
MAP1209	D180PglpF	120 nb (SEQ ID NO: 32)

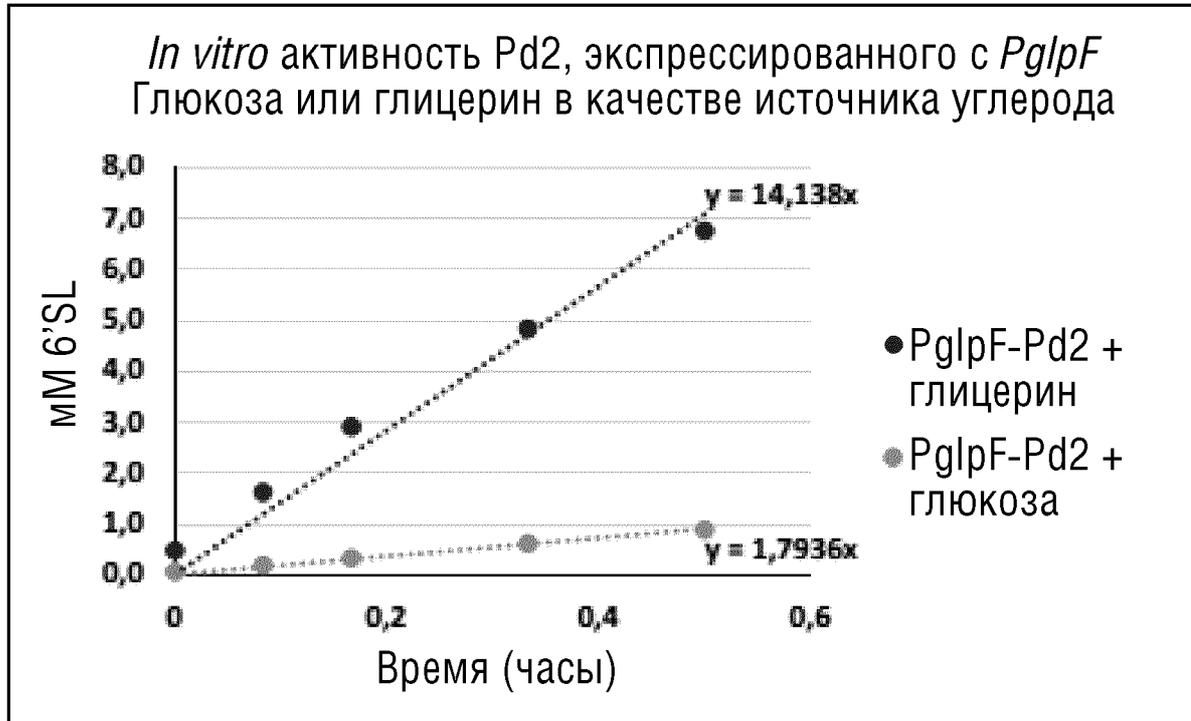
## ФИГ.8В



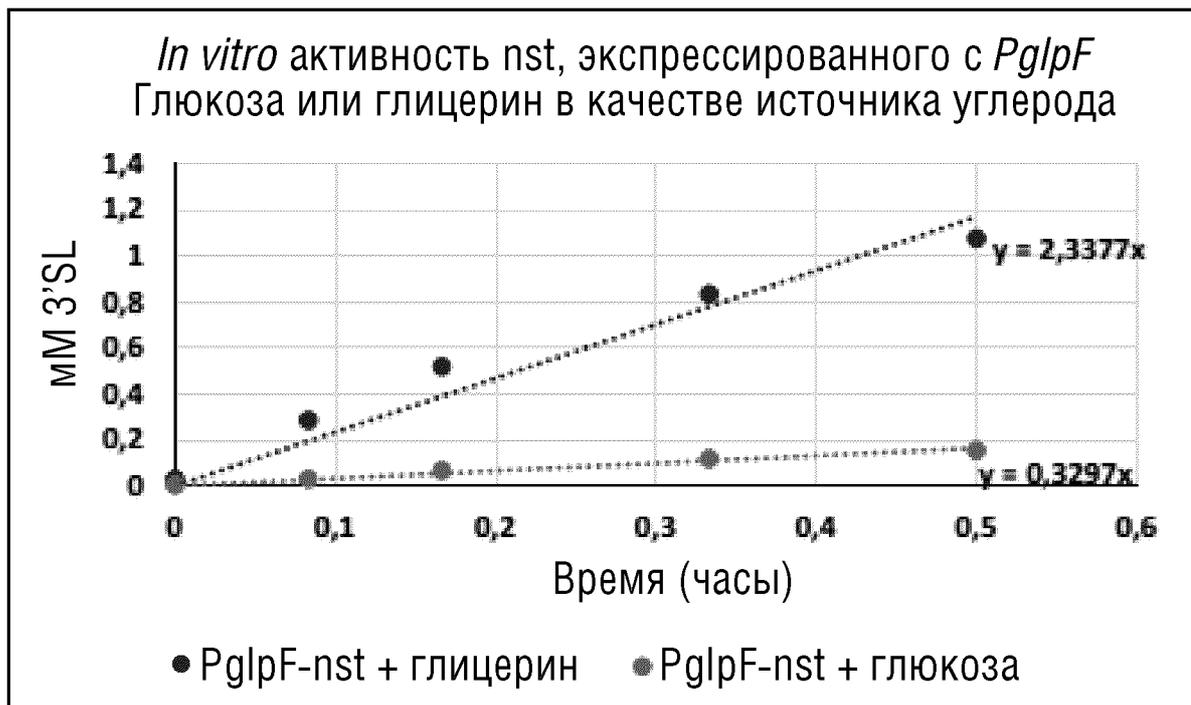
ФИГ.9



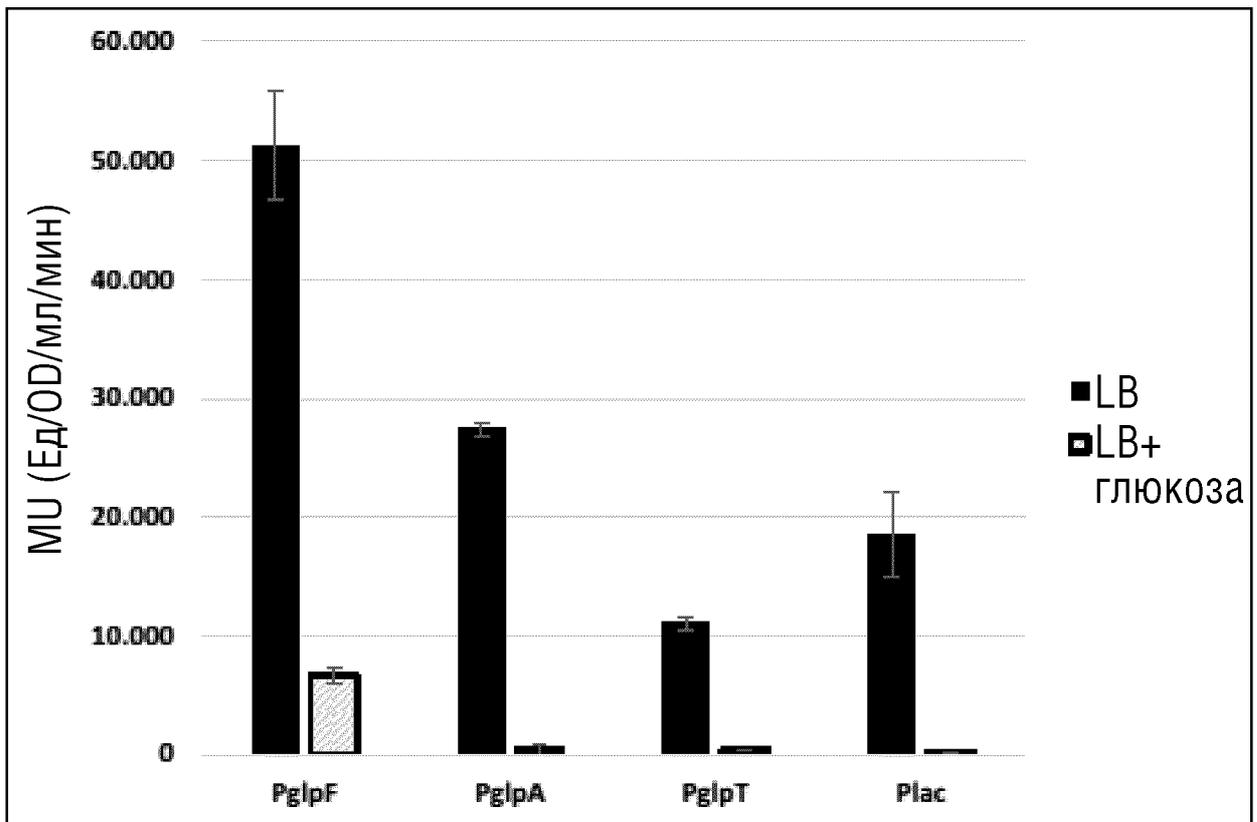
ФИГ.10А



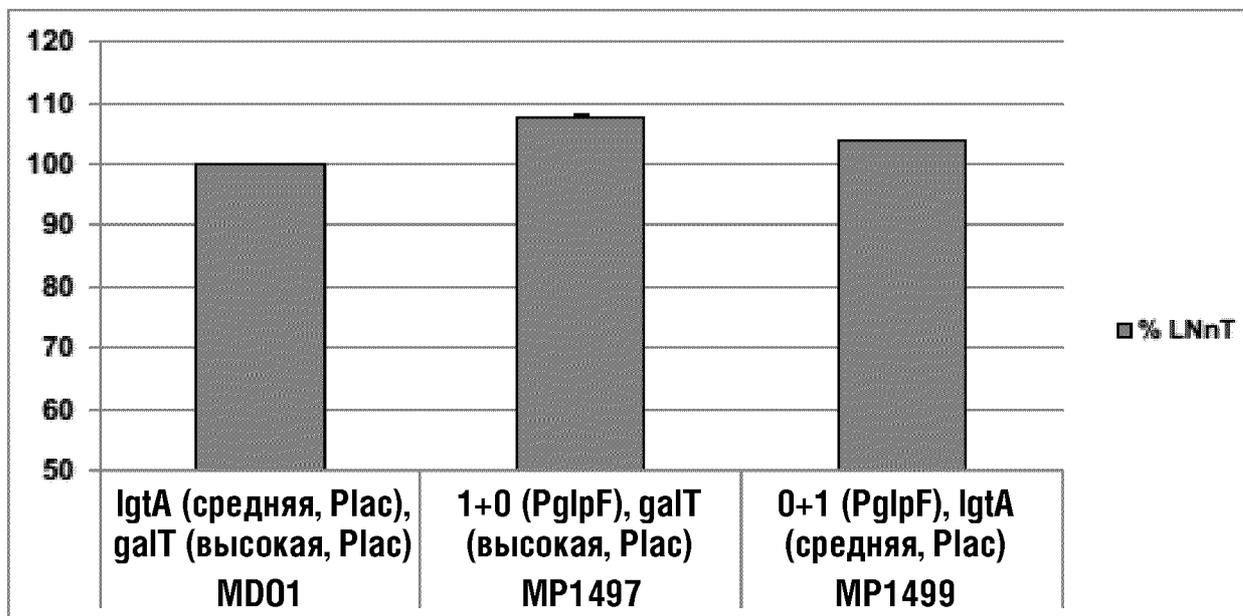
ФИГ.10В



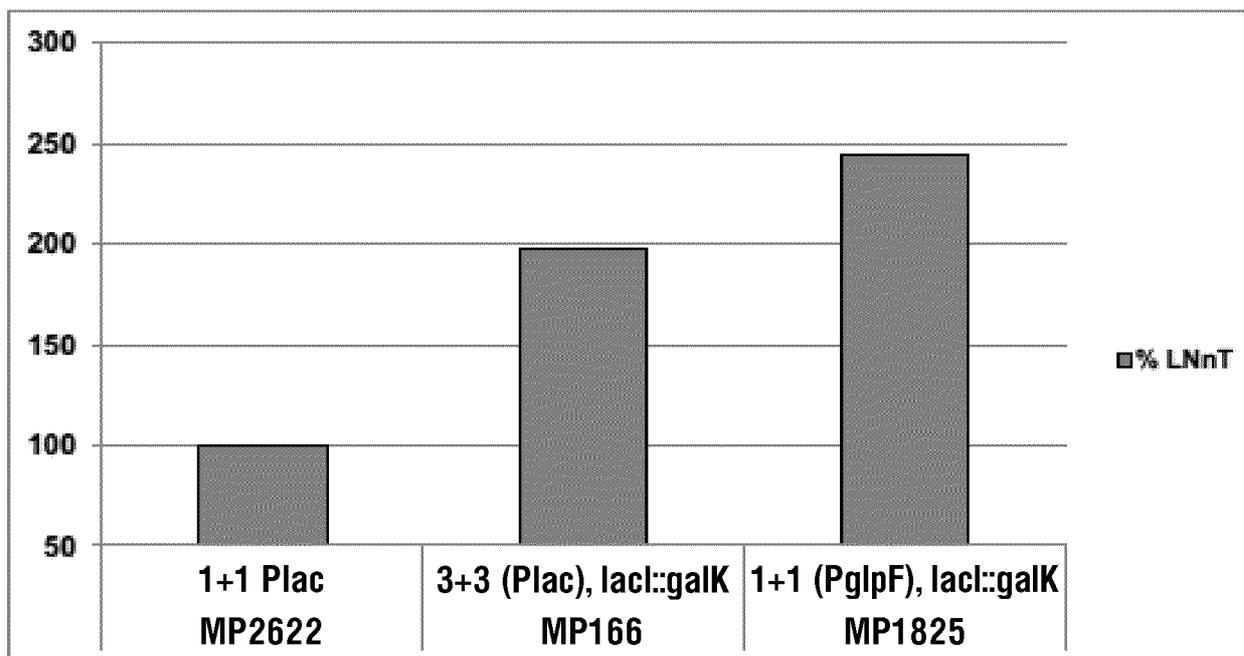
ФИГ.11



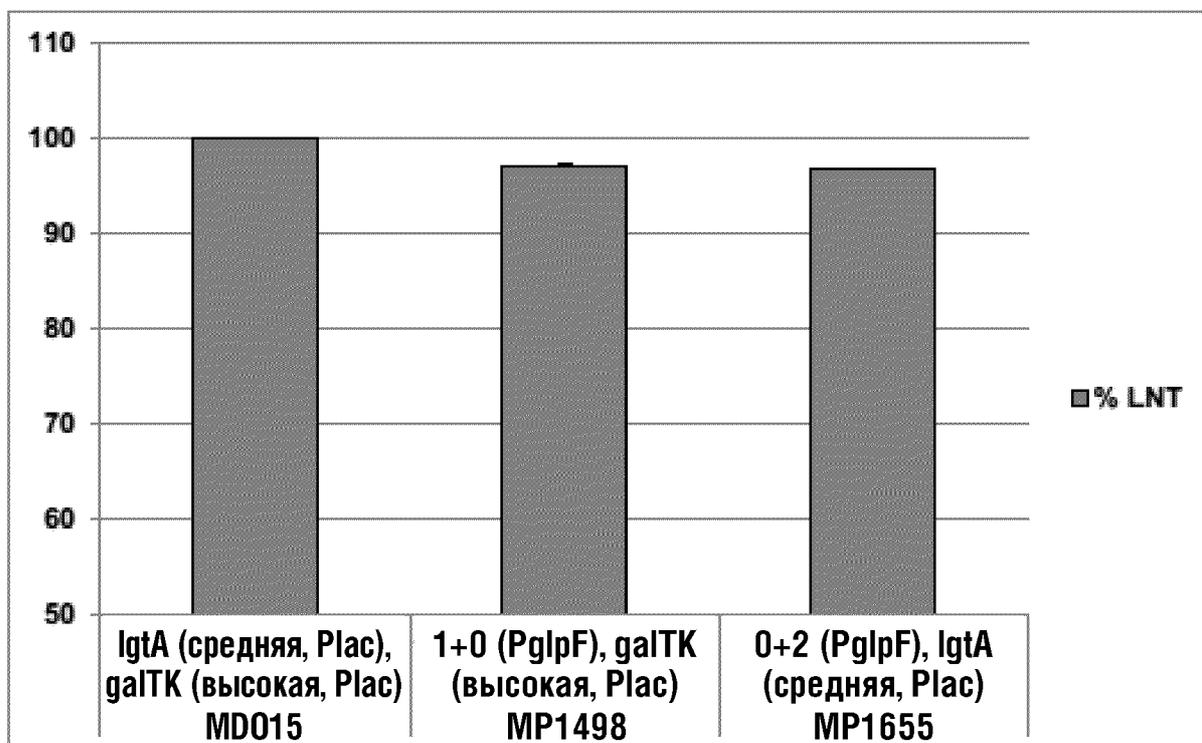
ФИГ.12А



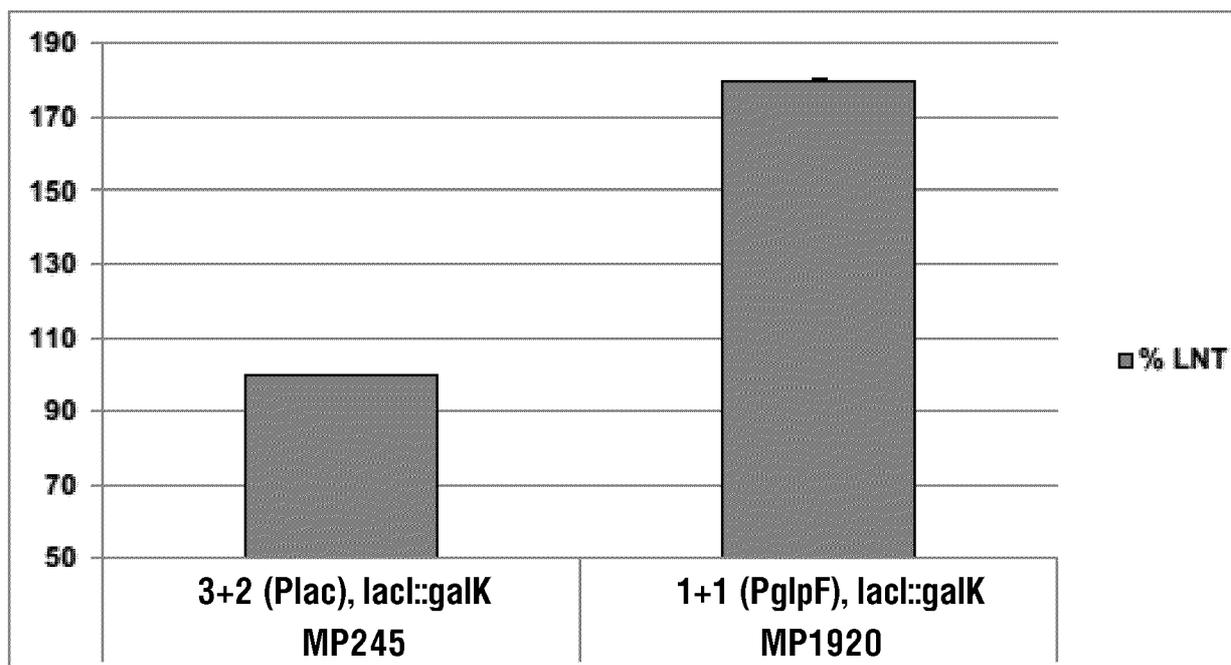
ФИГ.12В



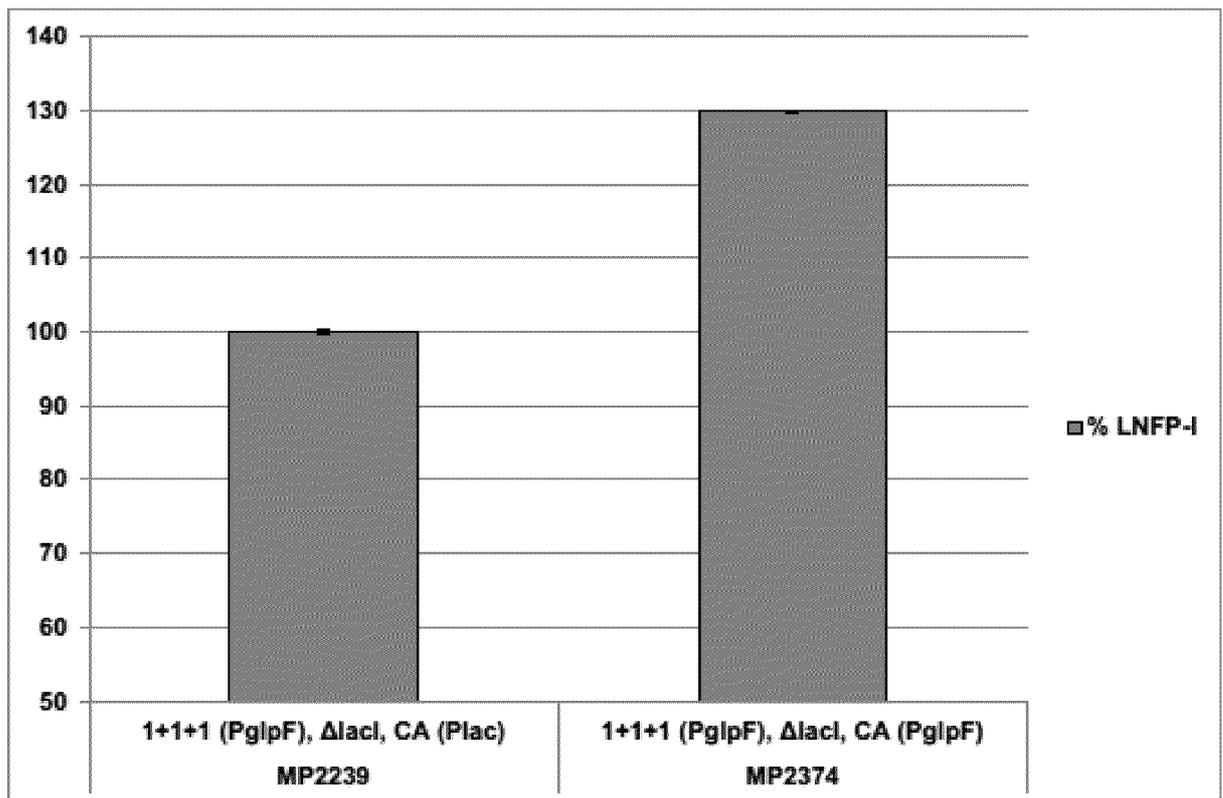
ФИГ.13А



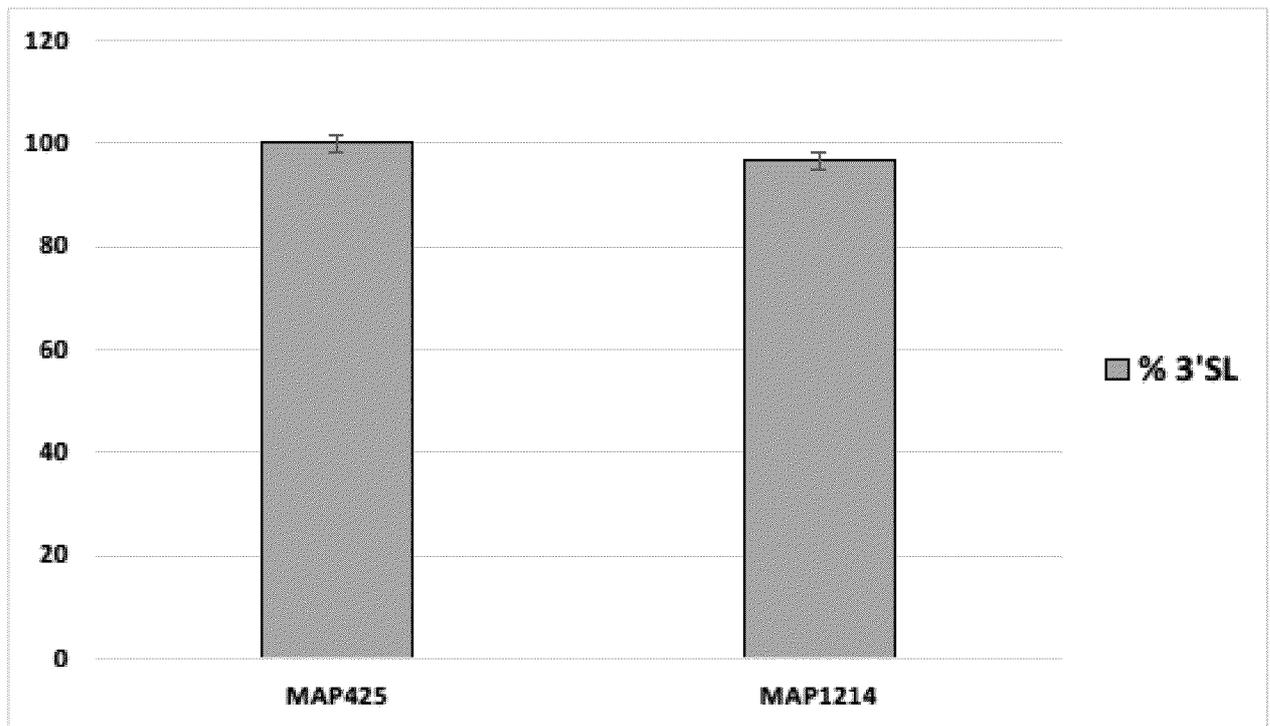
ФИГ.13В



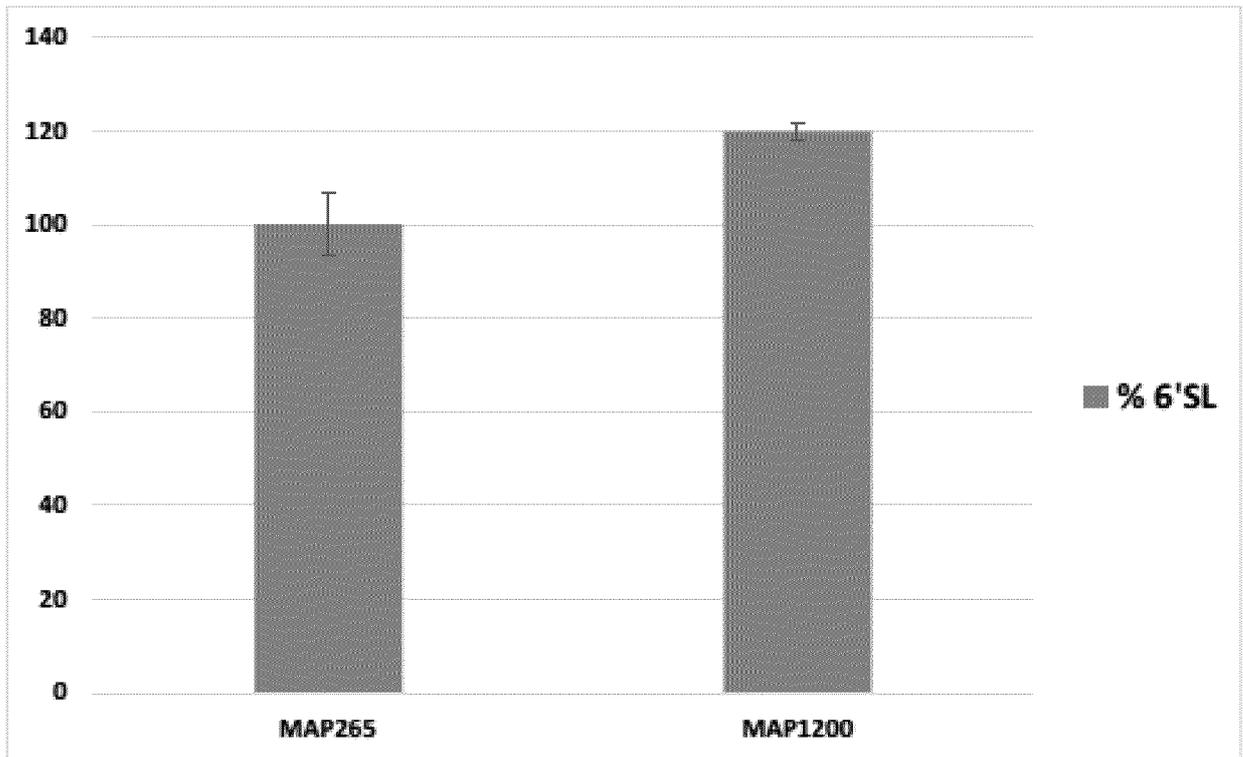
ФИГ.14



ФИГ.15



ФИГ.16



ФИГ.17

