

(19)



**Евразийское
патентное
ведомство**

(21) **201992593** (13) **A1**

(12) ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ

(43) Дата публикации заявки
2020.03.05

(22) Дата подачи заявки
2018.05.07

(51) Int. Cl. *C12N 9/00* (2006.01)
C12N 15/09 (2006.01)
C12N 15/31 (2006.01)
C12N 15/52 (2006.01)

(54) ВАРИАНТЫ СКОНСТРУИРОВАННОЙ ЛИГАЗЫ

(31) 62/503,075; 62/540,734; 62/587,030

(32) 2017.05.08; 2017.08.03; 2017.11.16

(33) US

(86) PCT/US2018/031376

(87) WO 2018/208665 2018.11.15

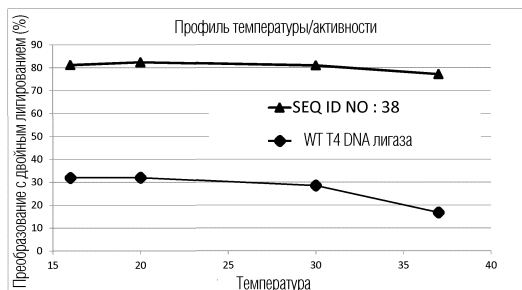
(71) Заявитель:
КОДЕКСИС, ИНК. (US)

(72) Изобретатель:

**Миллер Мэттью Дж., Врум Джонатан,
Деллас Никки, Баскервилл Доналд С.,
Гомес Сэнди М., Элгарт Дэвид, Видия
Джуди Виктория Антонио (US)**

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

(57) В настоящем изобретении представлены полипептиды сконструированной лигазы и их композиции, а также полинуклеотиды, кодирующие полипептиды сконструированной лигазы. В изобретении также представлены способы применения композиций, содержащих полипептиды сконструированной лигазы для диагностических и других целей.



A1

201992593

201992593

A1

ОПИСАНИЕ ИЗОБРЕТЕНИЯ

2420–559664EA/011

ВАРИАНТЫ СКОНСТРУИРОВАННОЙ ЛИГАЗЫ

Согласно настоящей заявке испрашивается приоритет в соответствии с предварительной заявкой на выдачу патента США № 62/503075, поданной 8 мая 2017 года, предварительной заявкой на выдачу патента США № 62/540734, поданной 3 августа 2017 года и предварительной заявкой на выдачу патента США № 62/587030, поданной 16 ноября 2017 года, каждая из которых полностью включена посредством ссылки и для всех целей.

Ссылка на список последовательностей, таблицу или компьютерную программу

Список последовательностей, одновременно представленный в данном документе в соответствии с §1,821 37 раздела свода законов США в машиночитаемом виде (CRF) посредством EFS–Web с названием файла CX9–160WO1_ST25.txt, включен в данный документ посредством ссылки. Электронная копия списка последовательностей создана 7 мая 2018 года с размером файла 100 кбит.

Область техники, к которой относится изобретение

В настоящем изобретении представлены полипептиды сконструированной лигазы и их композиции, а также полинуклеотиды, кодирующие полипептиды сконструированной лигазы. В изобретении также представлены способы применения композиций, содержащих полипептиды сконструированной лигазы для диагностических и других целей.

Уровень техники изобретения

ДНК–лигазы катализируют образование новых фосфодиэфирных связей в молекулах нуклеиновых кислот через конденсацию соседних 3'–гидроксильных и 5'–фосфатных концов. Фермент соединяет тупые и когезивные «липкие» терминальные концы, а также восстанавливает пробелы одной цепи в двухцепочечной ДНК и некоторых ДНК/РНК гибридах. Применение находят разные лигазы, одной из которых является ДНК–лигаза из бактериофага Т4. Т4 ДНК–лигаза представляет собой один из наиболее широко используемых ферментов в биотехнологии. Несмотря на то, что существуют разные ДНК–лигазы, которые находят применение, в данной области остается потребность в улучшенных лигазах для диагностических и исследовательских целей.

Сущность изобретения

В настоящем изобретении представлены полипептиды сконструированной лигазы и их композиции, а также полинуклеотиды, кодирующие полипептиды сконструированной лигазы. В изобретении также представлены способы применения композиций, содержащих полипептиды сконструированной лигазы для диагностических и других целей.

В настоящем изобретении представлены сконструированные лигазы, содержащие полипептидные последовательности, имеющие идентичность последовательности по

меньшей мере 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более с эталонной последовательностью SEQ ID NO: 2, 6, 32, 34 и/или 38 или ее функциональным фрагментом, при этом сконструированные лигазы содержат по меньшей мере одну замену или набор замен в их полипептидных последовательностях, и при этом положения аминокислот полипептидных последовательностей нумеруют со ссылкой на SEQ ID NO: 2, 6, 32, 34 или 38.

В настоящем изобретении также представлены сконструированные лигазы, содержащие по меньшей мере одну замену или набор замен, при этом по меньшей мере одну замену или набор замен выбирают из 52/56/404, 52/56/404/412, 127/207, 127/213, 127/213/276/339, 140/181/234, 165/181/299, 165/181/281/299, 238/241/404/412/462 и 462 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 2. В некоторых вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из 52E/56R/404K, 52E/56V/404K/412T, 127K/207R, 127K/213M, 127K/213M/276G/339V, 140A/181T/234M, 165A/181T/299P, 165A/181T/281A/299A, 238L/241L/404K/412T/462K и 462K, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 2. В некоторых дополнительных вариантах осуществления замены или наборы замен включают в себя замены или наборы замен, выбираемых из K52E/A56R/N404K, K52E/A56V/N404K/K412T, P127K/I207R, P127K/L213M, P127K/L213M/C276G/I339V, S140A/S181T/L234M, C165A/S181T/K299P, C165A/S181T/V281A/K299A, Y238L/N241L/N404K/K412T/I462K и I462K, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 2.

В настоящем изобретении дополнительно представлены сконструированные лигазы, содержащие по меньшей мере одну замену или набор замен, включающих в себя замену (замены) в положениях аминокислот, выбираемых из 52/127/140/181/462, 52/127/140/181/238, 52/127/181/462, 52/127/276/372/462, 52/127/404, 52/140/181/238/276/293/404, 52/140/181/276/299/404/462, 52/140/207/299/372/404/462, 52/140/238/276/299/372/404, 52/181, 52/181/238/276, 52/181/238/299/404, 52/181/293, 52/207/238/293/299/404/462, 52/276/299/404, 52/238/404/462, 52/293/299/404/462, 52/404/462, 58/63/89, 58/88/89/226/440, 58/88/199/225/226, 58/88/226/306, 58/88/306/470, 58/440/470, 58/451, 63/88/89, 63/88/451, 63/89/226/440/451, 63/89/451, 63/199/297/375, 88/225/440/451, 88/306/440/451, 88/470, 89, 127, 127/140/238, 127/140/276, 127/140/299/372/462, 127/181/207, 127/181/238/372, 127/181/276, 127/181/404, 127/207/238/372, 127/238/293/462, 127/238/293/299/372/404, 127/238/293/299/404, 127/238/372/462, 127/293, 127/293/372/462, 127/293/404/462, 127/462, 140/238/372/462, 140/276/293/404, 140/285/293/404, 140/299/372/404/462, 140/372, 140,181/207/238, 181/207/238/276/293/372/404, 181/207/238/372, 181/238/276, 181/238/299/404, 181,238/462, 181/276, 181/293, 181/462, 238/293/299/372/462, 238/293/372, 238/299/404, 238/404/462, 276/293/462, 276/404, 293/372, 299/372/462, 299/404/462, 372, 372/462, 404, 451 и 462 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6. В некоторых вариантах осуществления замена (замены) или наборы замен

включают в себя замены или наборы замен, выбираемых из 2E/127K/140A/181T/462K, 52E/127K/140A/181T/238L, 52E/127K/181T/462K, 52E/127K/276G/372I/462K, 52E/127K/404K, 52E/140A/181T/238L/276G/293E/404K, 52E/140A/181T/276G/299P/404K/462K, 52E/140A/207R/299P/372I/404K/462K, 52E/140A/238L/276G/299P/372I/404K, 52E/181T, 52E/181T/238L/276G, 52E/181T/238L/299P/404K, 52E/181T/293E, 52E/207R/238L/293E/299P/404K/462K, 52E/276G/299P/404K, 52E/238L/404K/462K, 52E/293E/299P/404K/462K, 52E/404K/462K, 58K/63R/89K, 58K/88R/89K/226E/440K, 58K/88R/199E/225A/226E, 58K/88R/226E/306A, 58K/88R/306A/470E, 58K/440K/470E, 58K/451K, 63R/88R/89K, 63R/88R/451K, 63R/89K/226E/440K/451K, 63R/89K/451K, 63R/199E/ 297G/375E, 88R/225A/440K/451K, 88R/306A/440K/451K, 88R/470E, 89K, 127K, 127K/140A/238L, 127K/140A/276G, 127K/140A/299P/372I/462K, 127K/181T/207R, 127K/181T/238L/372I, 127K/181T/276G, 127K/181T/404K, 127K/207R/238L/372I, 127K/238L/293E/462K, 127K/238L/293E/299P/372I/404K, 127K/238L/293E/299P/404K, 127K/238L/372I/462K, 127K/293E, 127K/293E/372I/462K, 127K/293E/404K/462K, 127K/462K, 140A/238L/372I/462K, 140A/276G/293E/404K, 140A/285A/293E/404K, 140A/299P/372I/404K/462K, 140A/372I, 140T, 181T/207R/238L, 181T/207R/238L/276G/293E/372I/404K, 181T/207R/238L/372I, 181T/238L/276G, 181T/238L/299P/404K, 181T/238L/462K, 181T/276G, 181T/293E, 181T/462K, 238L/293E/299P/372I/462K, 238L/293E/372I, 238L/299P/404K, 238L/404K/462K, 276G/293E/462K, 276G/404K, 293E/372I, 299P/372I/462K, 299P/404K/462K, 372I, 372I/462K, 404K, 451K и 462K, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6. В некоторых дополнительных вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из K52E/P127K/S140A/S181T/I462K, K52E/P127K/S140A/S181T/Y238L, K52E/P127K/S181T/I462K, K52E/P127K/C276G/V372I/I462K, K52E/P127K/N404K, K52E/S140A/S181T/Y238L/C276G/L293E/N404K, K52E/S140A/S181T/C276G/K299P/N404K/I462K, K52E/S140A/I207R/K299P/V372I/N404K/I462K, K52E/S140A/Y238L/C276G/K299P/V372I/N404K, K52E/S181T, K52E/S181T/Y238L/C276G, K52E/S181T/Y238L/K299P/N404K, K52E/S181T/L293E, K52E/I207R/Y238L/L293E/K299P/N404K/I462K, K52E/C276G/K299P/N404K, K52E/Y238L/N404K/I462K, K52E/L293E/K299P/N404K/I462K, K52E/N404K/I462K, Q58K/L63R/E89K, Q58K/E88R/E89K/K226E/E440K, Q58K/E88R/K199E/K225A/K226E, Q58K/E88R/K226E/K306A, Q58K/E88R/K306A/K470E, Q58K/E440K/K470E, Q58K/T451K, L63R/E88R/E89K, L63R/E88R/T451K, L63R/E89K/K226E/E440K/T451K, L63R/E89K/T451K, L63R/K199E/R297G/K375E, E88R/K225A/E440K/T451K, E88R/K306A/E440K/T451K, E88R/K470E, E89K, P127K, P127K/S140A/Y238L, P127K/S140A/C276G, P127K/S140A/K299P/V372I/I462K, P127K/S181T/I207R, P127K/S181T/Y238L/V372I, P127K/S181T/C276G, P127K/S181T/N404K,

P127K/I207R/Y238L/V372I, P127K/Y238L/L293E/I462K,
 P127K/Y238L/L293E/K299P/V372I/N404K, P127K/Y238L/L293E/K299P/N404K,
 P127K/Y238L/V372I/I462K, P127K/L293E, P127K/L293E/V372I/I462K,
 P127K/L293E/N404K/I462K, P127K/I462K, S140A/Y238L/V372I/I462K,
 S140A/C276G/L293E/N404K, S140A/V285A/L293E/N404K,
 S140A/K299P/V372I/N404K/I462K, S140A/V372I, S140T, S181T/I207R/Y238L,
 S181T/I207R/Y238L/C276G/L293E/V372I/N404K, S181T/I207R/Y238L/V372I,
 S181T/Y238L/C276G, S181T/Y238L/K299P/N404K, S181T/Y238L/I462K, S181T/C276G,
 S181T/L293E, S181T/I462K, Y238L/L293E/K299P/V372I/I462K, Y238L/L293E/V372I,
 Y238L/K299P/N404K, Y238L/N404K/I462K, C276G/L293E/I462K, C276G/N404K,
 L293E/V372I, K299P/V372I/I462K, K299P/N404K/I462K, V372I, V372I/I462K, N404K,
 T451K и I462K, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6.

В настоящем изобретении дополнительно представлены сконструированные лигазы, содержащие по меньшей мере одну замену или набор замен, включающих в себя замену (замены) в положениях аминокислот, выбираемых из 19, 19/127/199, 19/127/306, 19/238, 89, 89/127, 89/127/238/306, 127, 127/133/238/375, 127/177/238/293/306, 127/238, 127/306, 127/385, 176/244/247/373/438, 176/250/373/438/480, 238, 238/306/372, 244, 244/247, 244/247/250, 244/250/438, 244/438, 247/373/427/438, 297, 306, 372, 404 и 438 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32. В некоторых вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из 19K, 19K/127K/199S, 19K/127K/306A, 19K/238L, 89K, 89K/127K, 89K/127K/238L/306A, 127K, 127K/133H/238L/375R, 127K/177A/238L/293P/306A, 127K/238L, 127K/306A, 127K/385E, 176G/244S/247K/373A/438D, 176G/250S/373A/438D/480S, 238L, 238L/306A/372I, 244S, 244S/247K, 244S/247K/250S, 244S/250S/438D, 244S/438D, 247K/373A/427K/438D, 297S, 306A, 372I, 404K и 438D, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32. В некоторых дополнительных вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из Q19K, Q19K/P127K/K199S, Q19K/P127K/K306A, Q19K/Y238L, E89K, E89K/P127K, E89K/P127K/Y238L/K306A, P127K, P127K/Q133H/Y238L/K375R, P127K/V177A/Y238L/L293P/K306A, P127K/Y238L, P127K/K306A, P127K/D385E, D176G/A244S/F247K/D373A/E438D, D176G/V250S/D373A/E438D/D480S, Y238L, Y238L/K306A/V372I, A244S, A244S/F247K, A244S/F247K/V250S, A244S/V250S/E438D, A244S/E438D, F247K/D373A/E427K/E438D, R297S, K306A, V372I, N404K и E438D, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32.

В настоящем изобретении дополнительно представлены сконструированные лигазы, содержащие по меньшей мере одну замену или набор замен, включающих в себя замену (замены) в положениях аминокислот, выбираемых из 51, 56, 60, 63, 86, 149, 174, 184, 199, 207, 233, 237, 238, 240, 314, 329, 371, 373, 385, 427, 438, 439, 446, 448, 451, 452, 453, 454, 461, 466, 476 и 485 и/или любых их комбинаций, при этом положения

аминокислот нумеруют со ссылкой на SEQ ID NO: 32. В некоторых вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из 51R, 56S, 60G/V, 63T, 86R, 149R, 174P, 184A, 199T, 207Q/V, 233A/T, 237N/R, 238L, 240P, 314V, 329G/L, 371V/W, 373A/G, 385A/W, 427L/R, 438D/F/G, 439S, 446R, 448A/G/P, 451G, 452P/V, 453G/L/R/T, 454L, 461C, 466G/P, 476A и 485G/Y, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32. В некоторых дополнительных вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из P51R, A56S, F60G/V, L63T, A86R, N149R, L174P, G184A, K199T, I207Q/V, F233A/T, A237N/R, Y238L, E240P, Y314V, D329G/L, D371V/W, D373A/G, D385A/W, E427L/R, E438D/F/G, C439S, K446R, D448A/G/P, K451G, D452P/V, Y453G/L/R/T, V454L, A461C, E466G/P, D476A и T485G/Y, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32.

В настоящем изобретении дополнительно представлены сконструированные лигазы, содержащие по меньшей мере одну замену или набор замен, включающих в себя замену (замены) в положениях аминокислот, выбираемых из 7, 17, 52, 54, 59, 74, 85, 183, 199, 240, 241, 242, 280, 321, 235, 237, 371, 404, 405, 451, 452, 453, 454, 462 и 483 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32. В некоторых вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из 7L, 17R, 52G, 54E, 59M, 74G/T, 85T, 183N, 199G, 240P, 241G, 242H, 280L, 321A/R, 235R, 237G, 371G, 404S/G, 405G, 451G, 452P, 453L, 454A, 462Q и 483G/Q, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32. В некоторых дополнительных вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из N7L, Q17R, K52G, G54E, S59M, F74G/T, A85T, A183N, K199G, E240P, N241G, S242H, Q280L, E321A/R, F235R, A237G, D371G, N404S, A405G, K451G, D452P, Y453L, V454A, I462Q и E483G/Q, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32.

В настоящем изобретении дополнительно представлены сконструированные лигазы, содержащие по меньшей мере одну замену или набор замен, включающих в себя замену (замены) в положениях аминокислот, выбираемых из 7, 11, 13, 14, 54, 62, 89, 149, 183, 184, 185, 186, 231, 232, 233, 238, 239, 240, 385, 386, 413 и 453 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6. В некоторых вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из 7K, 11K, 13K, 14K, 54K, 62K, 89K, 149K, 183K, 184K, 185K, 186K, 231K, 232K, 233K, 238K, 239K, 240K, 385K, 386K, 413K и 453K, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6. В некоторых дополнительных вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из N7K, S11K, G13K, S14K, G54K, M62K, E89K, N149K, A183K, G184K, N185K, E186K, L231K, D232K, F233K, Y238K, P239K, E240K, D385K, P386K, A413K и Y453K, при этом положения аминокислот нумеруют со

ссылкой на SEQ ID NO: 6.

В настоящем изобретении дополнительно представлены сконструированные лигазы, содержащие по меньшей мере одну замену или набор замен, включающих в себя замену (замены) в положениях аминокислот, выбираемых из 19/63/233/237/371/452, 19/237/453, 63/89/448/452/453, 63/149/240/371/452, 63/233/240/452/454, 86/89/149/233/237/240, 86/89/149/233/237/314/452, 86/89/233/237/240/448, 89/233/237/240/448/453/454, 89/240/454, 149/233/237/454, 149/237/240, 149/237/240/329/404/453, 233/237/371/404/452/454 и 233/237/404 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 34. В некоторых вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из 19K/63T/233A/237R/371W/452P, 19K/237N/453G, 63T/89K/448A/452P/453G, 63T/149R/240P/371W/452P, 63T/233A/240P/452P/454L, 86R/89K/149R/233A/237N/240P, 86R/89K/149R/233A/237R/314V/452P, 86R/89K/233A/237N/240P/448A, 89K/233A/237R/240P/448A/453G/454L, 89K/240P/454L, 149R/233A/237N/454L, 149R/237N/240P, 149R/237N/240P/329G/404K/453G, 233A/237N/371W/404K/452P/454L и 233A/237R/404K, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 34. В некоторых дополнительных вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из Q19K/L63T/F233A/A237R/D371W/D452P, Q19K/A237N/Y453G, L63T/E89K/D448A/D452P/Y453G, L63T/N149R/E240P/D371W/D452P, L63T/F233A/E240P/D452P/V454L, A86R/E89K/N149R/F233A/A237N/E240P, A86R/E89K/N149R/F233A/A237R/Y314V/D452P, A86R/E89K/F233A/A237N/E240P/D448A, E89K/F233A/A237R/E240P/D448A/Y453G/V454L, E89K/E240P/V454L, N149R/F233A/A237N/V454L, N149R/A237N/E240P, N149R/A237N/E240P/D329G/N404K/Y453G, F233A/A237N/D371W/N404K/D452P/V454L и F233A/A237R/N404K, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 34.

В настоящем изобретении дополнительно представлены сконструированные лигазы, содержащие по меньшей мере одну замену или набор замен, включающих в себя замену (замены) в положениях аминокислот, выбираемых из 13/89/183/231, 13/89/183/232/386/451, 13/183/232/329/453/466, 13/183/232/386/451, 13/232/385/451, 89/183/329/451/453, 149/183, 183, 183/207/386, 183/207/386/427/453, 183/207/439, 183/231/373, 183/231/385/427, 183/231/427/466, 183/373/386, 183/385, 183/385/427, 183/413/427, 183/427/451 и 385/453/466 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 38. В некоторых вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из 13K/89K/183K/231K, 13K/89K/183K/232K/386K/451G, 13K/183K/232K/329L/453G/466G, 13K/183K/232K/386K/451G, 13K/232K/385K/451G, 89K/183K/329G/451G/453R, 149R/183K, 183K, 183K/207V/386K,

183K/207V/386K/427R/453G, 183K/207V/439S, 183K/231K/373G, 183K/231K/385K/427R, 183K/231K/427R/466G, 183K/373A/386K, 183K/385K, 183K/385K/427R, 183K/413K/427R, 183K/427R/451G и 385K/453R/466G, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 38. В некоторых дополнительных вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из G13K/E89K/A183K/L231K, G13K/E89K/A183K/D232K/P386K/K451G, G13K/A183K/D232K/D329L/Y453G/E466G, G13K/A183K/D232K/P386K/K451G, G13K/D232K/D385K/K451G, E89K/A183K/D329G/K451G/Y453R, N149R/A183K, A183K, A183K/I207V/P386K, A183K/I207V/P386K/E427R/Y453G, A183K/I207V/C439S, A183K/L231K/D373G, A183K/L231K/D385K/E427R, A183K/L231K/E427R/E466G, A183K/D373A/P386K, A183K/D385K, A183K/D385K/E427R, A183K/A413K/E427R, A183K/E427R/K451G и D385K/Y453R/E466G, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 38.

В настоящем изобретении дополнительно представлены сконструированные лигазы, содержащие по меньшей мере одну замену или набор замен, включающих в себя замену (замены) в положениях аминокислот, выбираемых из 13/19/63/88/127/183/225/232/233/237/329/371/440/451/452/453/466, 13/19/63/88/127/183/225/232/233/237/371/386/440/451/452, 19/63/88/127/183/225/231/233/237/371/427/440/451/452/466, 19/63/88/127/183/225/233/237/371/373/386/440/451/452, 19/63/88/127/225/233/237/371/385/440/451/452/453/466, 19/63/88/127/225/233/237/371/440/451/452, 63/88/127/149/225/240/371/440/451/452, 86/88/89/127/149/225/233/237/240/440/451, 88/89/127/225/233/237/240/440/448/451/453/454, 88/127/149/225/233/237/440/451/454, 88/127/225/440/451 и 88/225/440/451 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6. В некоторых вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из 13K/19K/63T/88R/127K/183K/225A/232K/233A/237R/329L/371W/440K/451K/452P/453G/466G, 13K/19K/63T/88R/127K/183K/225A/232K/233A/237R/371W/386K/440K/451G/452P, 19K/63T/88R/127K/183K/225A/231K/233A/237R/371W/427R/440K/451K/452P/466G, 19K/63T/88R/127K/183K/225A/233A/237R/371W/373A/386K/440K/451K/452P, 19K/63T/88R/127K/225A/233A/237R/371W/385K/440K/451K/452P/453R/466G, 19K/63T/88R/127K/225A/233A/237R/371W/440K/451K/452P, 63T/88R/127K/149R/225A/240P/371W/440K/451K/452P, 86R/88R/89K/127K/149R/225A/233A/237N/240P/440K/451K, 88R/89K/127K/225A/233A/237R/240P/440K/448A/451K/453G/454L, 88R/127K/149R/225A/233A/237N/440K/451K/454L, 88R/127K/225A/440K/451K и 88R/225A/440K/451K, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6. В некоторых дополнительных вариантах осуществления замена (замены) или наборы замен включают в себя замены или наборы замен, выбираемых из

G13K/Q19K/L63T/E88R/P127K/A183K/K225A/D232K/F233A/A237R/D329L/D371W/E440K/T451K/D452P/Y453G/E466G,
 G13K/Q19K/L63T/E88R/P127K/A183K/K225A/D232K/F233A/A237R/D371W/P386K/E440K/T451G/D452P,
 Q19K/L63T/E88R/P127K/A183K/K225A/L231K/F233A/A237R/D371W/E427R/E440K/T451K/D452P/E466G,
 Q19K/L63T/E88R/P127K/A183K/K225A/F233A/A237R/D371W/D373A/P386K/E440K/T451K/D452P,
 Q19K/L63T/E88R/P127K/K225A/F233A/A237R/D371W/D385K/E440K/T451K/D452P/Y453R/E466G, Q19K/L63T/E88R/P127K/K225A/F233A/A237R/D371W/E440K/T451K/D452P,
 L63T/E88R/P127K/N149R/K225A/E240P/D371W/E440K/T451K/D452P,
 A86R/E88R/E89K/P127K/N149R/K225A/F233A/A237N/E240P/E440K/T451K,
 E88R/E89K/P127K/K225A/F233A/A237R/E240P/E440K/D448A/T451K/Y453G/V454L,
 E88R/P127K/N149R/K225A/F233A/A237N/E440K/T451K/V454L,
 E88R/P127K/K225A/E440K/T451K и E88R/K225A/E440K/T451K, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6.

В настоящем изобретении также представлены сконструированные лигазы, содержащие полипептидные последовательности, которые по меньшей мере на 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более идентичны последовательности по меньшей мере одного варианта сконструированной лигазы, представленной в таблице 4.1, 4.2, 4.3, 5.1, 5.2, 5.3, 5.4, 5.5 и/или 6.1. В некоторых вариантах осуществления сконструированной лигазой является вариант сконструированной лигазы, представленной в таблице 4.1, 4.2, 4.3, 5.1, 5.2, 5.3, 5.4, 5.5 и/или 6.1. В некоторых дополнительных вариантах осуществления сконструированная лигаза обладает активностью ДНК-лигазы. В некоторых дополнительных вариантах осуществления сконструированная лигаза обладает по меньшей мере одним улучшенным свойством по сравнению с Т4 ДНК-лигазой дикого типа. В некоторых дополнительных вариантах осуществления сконструированные лигазы согласно настоящему изобретению обладают по меньшей мере одним улучшенным свойством по сравнению с Т4 ДНК-лигазой дикого типа, при этом улучшенное свойство выбирают из демонстрации большей активности при низких концентрациях ДНК-субстрата и получения меньшего количества димеров адаптеров. В некоторых дополнительных вариантах осуществления сконструированные лигазы согласно настоящему изобретению являются более термостабильными, чем Т4 ДНК-лигаза дикого типа. В некоторых дополнительных вариантах осуществления сконструированные лигазы согласно настоящему изобретению являются стабильными на протяжении более широкого диапазона pH чем Т4 ДНК-лигаза дикого типа. В некоторых дополнительных вариантах осуществления сконструированные лигазы согласно настоящему изобретению являются очищенными.

В настоящем изобретении также представлены полинуклеотидные последовательности, кодирующие по меньшей мере одну представленную в данном

документе сконструированную лигазу. В некоторых вариантах осуществления полинуклеотидная последовательность кодирует по меньшей мере одну сконструированную лигазу, имеющую идентичность последовательности по меньшей мере 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более с эталонной последовательностью SEQ ID NO: 2, 6, 32, 34 и/или 37 или ее функциональным фрагментом, при этом сконструированный полипептид содержит по меньшей мере одну замену в одной или более положений аминокислот. В некоторых вариантах осуществления полинуклеотидная последовательность содержит последовательность, имеющую идентичность последовательности по меньшей мере 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более с эталонной последовательностью SEQ ID NO: 1, 5, 31, 33 и/или 37. В некоторых дополнительных вариантах осуществления полинуклеотидная последовательность содержит SEQ ID NO: 1, 5, 31, 33 и/или 37. В некоторых дополнительных вариантах осуществления полинуклеотидная последовательность функционально связана с регуляторной последовательностью. В некоторых дополнительных вариантах осуществления полинуклеотидная последовательность является кодон-оптимизированной.

В настоящем изобретении также представлены векторы экспрессии, содержащие по меньшей мере одну полинуклеотидную последовательность, представленную в данном документе. В настоящем изобретении также представлены клетки-хозяева, трансформированные по меньшей мере вектором экспрессии, представленным в данном документе.

В настоящем изобретении дополнительно представлены способы получения полипептида сконструированной лигазы в клетке-хозяине, включающие культивирование клетки-хозяина, представленной в данном документе, в подходящих условиях культивирования, таким образом чтобы получить по меньшей мере одну сконструированную лигазу. В некоторых вариантах осуществления способы дополнительно включают выделение по меньшей мере одной сконструированной лигазы из культуры и/или клеток-хозяев. В некоторых дополнительных вариантах осуществления способы дополнительно включают стадию очистки по меньшей мере одной сконструированной лигазы.

В настоящем изобретении также представлены композиции, содержащие по меньшей мере одну сконструированную лигазу, представленную в данном документе. Не предполагается, чтобы композиция содержала любые сконструированные лигазы, представленные в данном документе, которые были получены с использованием какого-либо конкретного способа. Предполагается, что настоящее изобретение ограничено сконструированными лигазами, полученными с использованием какого-либо конкретного способа.

В настоящем изобретении также представлены способы получения по меньшей мере одного продукта лигирования, включающие предоставление по меньшей мере одной сконструированной лигазы, представленной в данном документе, и реакционной смеси,

содержащей по меньшей мере два фрагмента нуклеиновых кислот, совмещение по меньшей мере одного сконструированного полипептида и реакционной смеси в таких условиях, чтобы происходило лигирование фрагментов нуклеиновых кислот и получался по меньшей мере один продукт лигирования. В некоторых вариантах осуществления способов входящая двухцепочечная ДНК содержит фрагменты ДНК с тупыми концами.

В настоящем изобретении также представлены способы получения продукта лигирования, включающие предоставление по меньшей мере одной сконструированной лигазы, представленной в данном документе, и реакционной смеси, содержащей по меньшей мере два фрагмента нуклеиновых кислот, совмещение по меньшей мере одного сконструированного полипептида и реакционной смеси в таких условиях, чтобы происходило лигирование фрагментов нуклеиновых кислот и получался по меньшей мере один продукт лигирования. В некоторых вариантах осуществления способов входящая двухцепочечная ДНК содержит фрагменты ДНК с тупыми концами.

В настоящем изобретении также представлены способы получения продукта, содержащего библиотеку ДНК, включающие предоставление по меньшей мере одной сконструированной лигазы, представленной в данном документе, и композиции реакционной смеси, содержащей входящую двухцепочечную ДНК, по меньшей мере один Т-образный адапторный олигонуклеотид, аденозин и реакционный буфер; выдерживание по меньшей мере одной сконструированной лигазы и реакционной смеси в таких условиях, чтобы добавить аденозин к 3' концам обеих нитей ДНК, и лигирование Т-образного адаптера к концам входящей ДНК для получения библиотеки ДНК. В некоторых вариантах осуществления способов входящая двухцепочечная ДНК содержит фрагменты ДНК с тупыми концами.

В настоящем изобретении также представлены способы получения продукта, содержащего множество фрагментов ДНК, подходящих для секвенирования, включающие предоставление по меньшей мере одной сконструированной лигазы, представленной в данном документе, и реакционной смеси, содержащей входящую двухцепочечную ДНК, олигонуклеотид, содержащий 3' выступы с одним основанием дезоксиаденином и 5' концы с монофосфатом, адапторный олигонуклеотид, содержащий 5' дезокситимидиновый выступ и 5' фосфат на совместимом с лигированием конце, и; выдерживание по меньшей мере одной сконструированной лигазы и реакционной смеси в таких условиях, чтобы происходило лигирование олигонуклеотида, адапторного олигонуклеотида и входящей двухцепочечной ДНК, с получением продукта, содержащего множество фрагментов ДНК, подходящих для секвенирования.

В некоторых вариантах осуществления способов использования сконструированной лигазы согласно настоящему изобретению выдерживание проводят в присутствии краудинг-агента. В некоторых вариантах осуществления способов продукт трансформируют в *E. coli* после тепловой инактивации продукта. В некоторых дополнительных вариантах осуществления продукт используют для создания библиотеки молекул ДНК. В некоторых дополнительных вариантах осуществления библиотеку

молекул ДНК подвергают секвенированию. В некоторых дополнительных вариантах осуществления реакционная смесь содержит по меньшей мере одно соединение, которое ингибирует лигирование. В некоторых дополнительных вариантах осуществления способы включают присутствие по меньшей мере одного соединения, которое ингибирует лигирование, представляющего собой субоптимальный буфер для лигирования. В некоторых вариантах осуществления в способе получают больше продукта, чем в таких же способах, в которых используют Т4 ДНК-лигазу дикого типа. В некоторых дополнительных вариантах осуществления реакционная смесь содержит по меньшей мере один фермент. В некоторых дополнительных вариантах осуществления фермент выбирают из полимеразы, полинуклеотидкиназы, экзонуклеазы, эндонуклеазы и цитидиндезаминазы. В некоторых дополнительных вариантах осуществления концентрация входящей двухцепочечной ДНК составляет менее, чем 100 нМ, менее, чем 50 нМ, менее, чем 10 нМ, менее, чем 1 нМ или менее, чем 100 пМ. В некоторых дополнительных вариантах осуществления концентрация адаптеров составляет менее, чем 10-кратное, менее, чем 5-кратное, менее, чем 3-кратное или менее, чем 2-кратное превышение концентрации включений в реакционную смесь. В некоторых дополнительных вариантах осуществления способы дополнительно включают стадию амплификации продукта. В некоторых дополнительных вариантах осуществления способы дополнительно включают стадию секвенирования продукта. В некоторых дополнительных вариантах осуществления способы дополнительно включают стадии амплификации и секвенирования продукта. В некоторых дополнительных вариантах осуществления не предпринимают шаги для удаления димера адаптера из продукта. В некоторых дополнительных вариантах осуществления входящую двухцепочечную ДНК выбирают из выделенной внеклеточной ДНК, циркулирующей опухолевой ДНК, ДНК, выделенной из циркулирующих опухолевых клеток, циркулирующей фетальной ДНК и тонкоигольных аспиратов. В некоторых вариантах осуществления входящую двухцепочечную ДНК предоставляют в виде необработанного образца. В некоторых дополнительных вариантах осуществления входящую двухцепочечную ДНК очищают перед ее включением в реакционную смесь. В некоторых дополнительных вариантах осуществления входящую двухцепочечную ДНК доставляют из образца, содержащего нуклеиновую кислоту, включая одноцепочечную ДНК, одноцепочечную РНК, двухцепочечную ДНК, двухцепочечную РНК и/или любую другую синтетическую или существующую в природе последовательность нуклеиновой кислоты. На самом деле, не предполагается, чтобы настоящее изобретение было ограничено любым конкретным исходным образцом ДНК. В некоторых вариантах осуществления способ осуществляют в условиях, выбираемых из микроструйных устройств и капель. В некоторых дополнительных вариантах осуществления объем комбинации реакционной смеси и сконструированной лигазы составляет менее, чем 5000 пл, менее, чем 1000 пл, менее, чем 100 пл, менее, чем 10 пл или менее, чем 1 пл. В некоторых дополнительных вариантах осуществления двухцепочечную входящую ДНК иммобилизируют, тогда как в некоторых

альтернативных вариантах осуществления сконструированную лигазу иммобилизируют. В некоторых дополнительных вариантах осуществления иммобилизируют двухцепочечную входящую ДНК и сконструированную лигазу. В некоторых дополнительных вариантах осуществления в дополнение или вместо иммобилизированной ДНК и/или иммобилизированной сконструированной лигазы иммобилизируют по меньшей мере одно соединение в реакционной смеси. В некоторых дополнительных вариантах осуществления продукт используют для создания библиотек для секвенирования ДНК, высокопроизводительного скрининга, генетического отбора, фогового дисплея, дрожжевого дисплея, рибосомального дисплея, клеточных анализов, биохимических анализов, одновременного многопараметрического анализа на основе изображений или определения конформации хроматина (С3).

В некоторых вариантах осуществления способов согласно настоящему изобретению продолжительность времени выдерживания составляет менее, чем 30 минут. В некоторых дополнительных вариантах осуществления продолжительность времени выдерживания составляет менее, чем 15 минут. В некоторых дополнительных вариантах осуществления продолжительность времени выдерживания составляет менее, чем 10, 9, 8, 7, 6, 5, 4, 3 или 2 минуты. В некоторых дополнительных вариантах осуществления продолжительность времени выдерживания составляет менее, чем 5 минут.

В некоторых вариантах осуществления настоящего изобретения продукт содержит меньше димеров адаптеров, чем в способах лигирования, в которых продолжительность времени выдерживания составляет 15 минут или более. Таким образом, в некоторых вариантах осуществления в настоящем изобретении представлены более быстрые способы, в которых получают меньше димеров адаптеров, что приводит к более продуктивному считыванию на анализируемый образец по сравнению с Т4 ДНК–лигазой дикого типа, а в некоторых вариантах осуществления с другими лигазами. В некоторых дополнительных вариантах осуществления способы согласно настоящему изобретению являются внеклеточными.

В настоящем изобретении также представлены способы получения продукта лигирования, включающие предоставление по меньшей мере одной сконструированной лигазы, представленной в данном документе, субстрата, содержащего по меньшей мере два фрагмента нуклеиновых кислот, и реакционной смеси; совмещение по меньшей мере одной сконструированной лигазы, субстрата и реакционной смеси в таких условиях, чтобы происходило лигирование фрагментов нуклеиновых кислот и получался по меньшей мере один продукт лигирования.

В настоящем изобретении также представлены способы получения библиотеки ДНК, включающие предоставление по меньшей мере одной сконструированной лигазы, представленной в данном документе, субстрата, содержащего входящую двухцепочечную ДНК, и композиции реакционной смеси, содержащей по меньшей мере один Т–образный адапторный олигонуклеотид, аденозин и реакционный буфер; совмещение по меньшей мере одной сконструированной лигазы, субстрата и реакционной смеси в таких условиях,

чтобы добавить аденозин к 3' концам обеих нитей ДНК, и лигирование Т-образного адаптера к концам входящей ДНК, для получения продукта, содержащего библиотеку ДНК. В некоторых вариантах осуществления входящая двухцепочечная ДНК содержит фрагменты ДНК с тупыми концами.

В настоящем изобретении также представлены способы получения множества фрагментов ДНК, подходящих для секвенирования, включающие предоставление по меньшей мере одной сконструированной лигазы, представленной в данном документе, субстрата, содержащего входящую двухцепочечную ДНК, и реакционной смеси, содержащей олигонуклеотид, содержащий 3' выступы с одним основанием дезоксиаденином и 5' концы с монофосфатом, адапторный олигонуклеотид, содержащий 5' дезокситимидиновый выступ и 5' фосфат на совместимом с лигированием конце; и совмещение по меньшей мере одной сконструированной лигазы, субстрата и реакционной смеси в таких условиях, чтобы происходило лигирование олигонуклеотида, адапторного олигонуклеотида и входящей двухцепочечной ДНК, с получением продукта, содержащего множество фрагментов ДНК, подходящих для секвенирования.

В настоящем изобретении также представлены способы получения множества фрагментов ДНК, подходящих для секвенирования, включающие предоставление по меньшей мере одной сконструированной лигазы, представленной в данном документе, субстрата, содержащего входящую двухцепочечную ДНК, и реакционной смеси, содержащей олигонуклеотид, содержащий 3' выступы с одним основанием дезоксиаденином и 5' концы с монофосфатом, адапторный олигонуклеотид, содержащий 5' дезокситимидиновый выступ и 5' фосфат на совместимом с лигированием конце; и совмещение по меньшей мере одной сконструированной лигазы, субстрата и реакционной смеси в таких условиях, чтобы происходило лигирование олигонуклеотида, адапторного олигонуклеотида и входящей двухцепочечной ДНК, с получением продукта, содержащего множество фрагментов ДНК, подходящих для секвенирования, при этом концентрация адапторного олигонуклеотида в реакционной смеси составляет менее, чем 20-кратное молярное превышение концентрации субстрата.

В некоторых вариантах осуществления способов, представленных в данном документе, выдерживание проводят в присутствии краудинг-агента. В некоторых дополнительных вариантах осуществления продукт трансформируют в *E. coli* после тепловой инактивации продукта. В некоторых дополнительных вариантах осуществления продукт используют для создания библиотеки молекул ДНК. В некоторых дополнительных вариантах осуществления библиотеку молекул ДНК подвергают секвенированию. Кроме того, в некоторых дополнительных вариантах осуществления реакционная смесь содержит по меньшей мере одно соединение, которое ингибирует лигирование. В некоторых вариантах осуществления по меньшей мере одно соединение, которое ингибирует лигирование, содержит субоптимальный буфер для лигирования. В некоторых особенно предпочтительных вариантах осуществления в способах получают больше продукта, чем в таких же способах при наличии Т4 ДНК-лигазы дикого типа. В

некоторых дополнительных вариантах осуществления реакционная смесь содержит по меньшей мере один фермент. В некоторых вариантах осуществления фермент выбирают из полимеразы, полинуклеотидкиназы, экзонуклеазы, эндонуклеазы и цитидиндезаминазы. В некоторых вариантах осуществления фермент несовместим с лигазным буфером. В некоторых дополнительных вариантах осуществления при использовании входящей двухцепочечной ДНК концентрация входящей двухцепочечной ДНК составляет менее, чем 100 нМ, менее, чем 50 нМ, менее, чем 10 нМ, менее, чем 1 нМ или менее, чем 100 пМ. В некоторых вариантах осуществления при использовании адаптеров концентрация адаптеров составляет менее, чем 10-кратное, менее, чем 5-кратное, менее, чем 3-кратное или менее, чем 2-кратное превышение концентрации включений в реакционную смесь. В некоторых дополнительных вариантах осуществления способы дополнительно включают стадию амплификации продукта. Кроме того, в некоторых дополнительных вариантах осуществления способы дополнительно включают стадию секвенирования продукта. В некоторых дополнительных вариантах осуществления способы дополнительно включают стадии амплификации и секвенирования продукта. В некоторых вариантах осуществления не предпринимают шаги для удаления димера адаптера из продукта. В некоторых дополнительных вариантах осуществления субстрат выбирают из выделенной внеклеточной ДНК, циркулирующей опухолевой ДНК, ДНК, выделенной из лейкозных клеток, ДНК, выделенной из клеток лимфомы, ДНК, выделенной из циркулирующих опухолевых клеток, ДНК, выделенной из инфицированных вирусом клеток, циркулирующей фетальной ДНК и тонкоигольных аспиратов. В некоторых вариантах осуществления субстрат содержит входящую двухцепочечную ДНК, предоставленную в виде необработанного образца. В некоторых альтернативных вариантах осуществления субстрат содержит входящую двухцепочечную ДНК, которую очищают перед ее включением в реакционную смесь. В некоторых дополнительных вариантах осуществления способы осуществляют в условиях использования микроструйных устройств и/или капель. В некоторых вариантах осуществления объем комбинации реакционной смеси и сконструированной лигазы составляет менее, чем 5000 пл, менее, чем 1000 пл, менее, чем 100 пл, менее, чем 10 пл или менее, чем 1 пл. В некоторых вариантах осуществления субстрат содержит иммобилизованную двухцепочечную входящую ДНК. В некоторых альтернативных вариантах осуществления сконструированную лигазу иммобилизуют. В некоторых дополнительных вариантах осуществления иммобилизуют субстрат, содержащий двухцепочечную входящую ДНК и сконструированную лигазу. В некоторых дополнительных вариантах осуществления иммобилизуют по меньшей мере одно соединение в реакционной смеси. В некоторых дополнительных вариантах осуществления продукт способов используют для создания библиотек для секвенирования ДНК, высокопроизводительного скрининга, генетического отбора, фагового дисплея, дрожжевого дисплея, рибосомального дисплея, клеточных анализов, биохимических анализов, одновременного многопараметрического анализа на основе изображений или

определения конформации хроматина (С3). В некоторых вариантах осуществления продолжительность времени совмещения составляет менее, чем 30 минут. В некоторых дополнительных вариантах осуществления продолжительность времени совмещения составляет менее, чем 15 минут. В некоторых дополнительных вариантах осуществления продолжительность времени совмещения составляет менее, чем 10, 9, 8, 7, 6, 5, 4, 3 или 2 минуты. В некоторых вариантах осуществления продолжительность времени совмещения составляет менее, чем 5 минут. В некоторых дополнительных вариантах осуществления способов продукт содержит меньше димеров адаптеров, чем в способах лигирования, в которых продолжительность времени совмещения составляет 15 минут или более. В некоторых дополнительных вариантах осуществления способы являются внеклеточными. В некоторых вариантах осуществления субстратом является внеклеточная ДНК, экстрактированная из текучей среды, полученной у пациента. В некоторых дополнительных вариантах осуществления текучая среда содержит сыворотку или плазму. В некоторых дополнительных вариантах осуществления субстрат содержит нуклеиновые кислоты с отличающимися последовательностями на их 3' и 5' концах. В некоторых дополнительных вариантах осуществления в способах достигают низкого смещения при лигировании. В некоторых дополнительных вариантах осуществления совмещение осуществляют при температуре в диапазоне между приблизительно 10° и приблизительно 40°C. В некоторых вариантах осуществления диапазон температур составляет от приблизительно 16° до приблизительно 37°C. В некоторых вариантах осуществления использование температуры в диапазоне от 16° до 37° обеспечивает улучшенное преобразование в продукт по сравнению с другими лигазами (например, с Т4 ДНК-лигазой дикого типа или с другими лигазами, известными в данной области). В некоторых вариантах осуществления совмещение осуществляют при рН в диапазоне от приблизительно рН 7 до приблизительно рН 10. В некоторых вариантах осуществления совмещение осуществляют при рН между приблизительно 7,5 и приблизительно 9. В некоторых вариантах осуществления рН способов находится в диапазоне от 7,5 до 9, при этом извлечение возможно в буфере, который несовместим для использования с Т4 ДНК-лигазой дикого типа или с другими лигазами (например, с другими лигазами, известными в данной области). Не предполагается, чтобы настоящее изобретение было ограничено способами, в которых рН находится в диапазоне от 7,5 до 9, так как в настоящем изобретении находят применение другие рН. Также не предполагается, чтобы настоящее изобретение было ограничено каким-либо конкретным буфером (буферами), так как в настоящем изобретении находят применение разные буферы. В некоторых дополнительных вариантах осуществления использование концентраций адаптера менее, чем 20-кратное молярное превышение концентраций субстрата находит применение для достижения как эффективного преобразования библиотеки, так и/или предотвращения переноса молекул адаптера на последующие стадии, на которых используют продукты. Не предполагается, чтобы способы согласно настоящему изобретению были ограничены таким 20-кратным молярным превышением субстрата, так как в способах согласно

настоящему изобретению применение находят другие концентрации субстрата. В некоторых вариантах осуществления сконструированная лигаза находит применение для предотвращения скачкообразного изменения показателя.

Описание чертежей

На фиг. 1 представлен график, показывающий профиль температуры/активности для преобразования в продукты с двухсторонним лигированием для множества температурных условий с помощью полипептида SEQ ID NO: 38 и T4 ДНК–лигазы дикого типа, как описано в примере 9.

На фиг. 2 представлен график, показывающий преобразование в продукты с двухсторонним лигированием при разных pH с помощью полипептида SEQ ID NO: 38 и T4 ДНК–лигазы дикого типа, как описано в примере 10.

На фиг. 3 представлен график, показывающий преобразование в продукты с двухсторонним лигированием с помощью полипептида SEQ ID NO: 38 и T4 ДНК–лигазы дикого типа, как описано в примере 11.

На фиг. 4 представлен график, показывающий преобразование субстрата в продукты с двухсторонним лигированием в присутствии 200 нМ адаптера или 40 нМ адаптера с помощью полипептида SEQ ID NO: 38 и T4 ДНК–лигазы дикого типа, как описано в примере 12.

На фиг. 5 представлен график, показывающий процентное преобразование в продукты с двухсторонним лигированием при входящей ДНК 10 нг и 100 нг с помощью полипептида SEQ ID NO: 38 и T4 ДНК–лигазы дикого типа, как описано в примере 13.

На фиг. 6 представлен график, показывающий преобразование в продукты с двухсторонним лигированием во внеклеточной ДНК с помощью полипептида SEQ ID NO: 38 и T4 ДНК–лигазы дикого типа, как описано в примере 14.

Описание изобретения

В настоящем изобретении представлены полипептиды сконструированной лигазы и их композиции, а также полинуклеотиды, кодирующие полипептиды сконструированной лигазы. В изобретении также представлены способы применения композиций, содержащих полипептиды сконструированной лигазы для диагностических и других целей. В некоторых вариантах осуществления полипептиды сконструированной лигазы оптимизируют для обеспечения улучшенной активности лигирования, особенно в условиях, включающих низкие концентрации входящей ДНК и другие условия, неблагоприятные для получения продуктов лигированной ДНК, особенно ДНК, подходящей для высокоскоростного анализа и/или реакций секвенирования. В некоторых вариантах осуществления в настоящем изобретении представлены способы и композиции, содержащие сконструированные лигазы для диагностических и исследовательских целей. В настоящем изобретении также представлены полипептиды сконструированной лигазы, мутанты, биологически активные фрагменты и их аналоги, и содержащие их композиции.

ДНК–лигазы катализируют образование новых фосфодиэфирных связей в молекулах нуклеиновых кислот через конденсацию соседних 3'-гидроксильных и 5'-

фосфатных концов. Нативным субстратом для T4 ДНК–лигазы является интермедиат содержащей пробел двухцепочечной («ds») ДНК, генерируемый в процессе репликации ДНК. В практических вариантах применения *in vitro*, таких как молекулярное клонирование и получение библиотеки секвенирования ДНК, T4 ДНК–лигазу обычно используют по причине ее относительной эффективности по сравнению с другими существующими в природе лигазами для соединения фрагментов двухцепочечной ДНК с когезивными или тупыми концами.

На эффективность лигирования влияют концентрации субстрата и свойства соединяемых двухцепочечных ДНК–субстратов. T4 ДНК–лигаза обладает очень низкой абсолютной аффинностью ($K_m \sim 50$ мкМ) для субстратов в реакциях соединения двухцепочечных концов. Однако, ее часто используют в реакциях, имеющих концентрации ДНК до 1 нМ (Matsumura, biotechn., 95: IV–XIII [2015]). Тогда как T4 ДНК–лигаза будет соединять когезивные выступы с тупыми концами и одним основанием (например, T/A выступ), эффективность этих реакций значительно уменьшается относительно реакции между субстратами с более длинными когезивными концами, которые могут образовывать временно отожденные интермедиаты при 16–20°C. Технологические процессы получения библиотеки секвенирования следующего поколения (NGS) зависят от лигирования двухцепочечных адаптерных молекул к входящей ДНК, представляя возможно наиболее жесткие условия для лигирования: вставка с одним основанием или тупыми концами и адаптерные субстраты и низкие концентрации входящей ДНК (например, из внеклеточной ДНК, отдельных клеток, тонкоигольных аспиратов или других образцов с низким выходом ДНК.)

Было разработано несколько подходов улучшения скорости и эффективности лигирования ДНК. В настоящее время наиболее широко используемый подход включает добавление в реакционную смесь неспецифических полимеров (например, краудинг–агентов). Добавление полиэтиленгликоля (PEG; молекулярная масса 1000–8000) и Ficoll 70 может улучшить как скорость, так и общее преобразование субстрата (см. Pfeiffer и Zimmerman, Nucl. Acids Res., 11: 7853–7871 [1983]). Буферы для лигирования, содержащие PEG6000, широко используют в наборах для быстрого клонирования с лигированием и получения библиотек NGS (например, в наборах, поставляемых коммерческими поставщиками, такими как Promega [Promega, Inc., Madison, WI] и New England Biolabs [New England Biolabs, Ipswich, MA]). Однако, при более высоких концентрациях краудинг–агента увеличивается несколько нежелательных результатов, включая образование конкатамеров с более высокой молекулярной массой и лигирование неподходящих субстратов, таких как димеры адаптеров, образованные в реакциях NGS. Кроме того, краудинг–агенты могут быть несовместимы с другими ферментативными реакциями, выполняемыми в одноканальных или микроструйных технологических процессах. Например, перед трансформацией *E. coli* обычно выполняют тепловую инактивацию ДНК–лигазы, но тепловая инактивация в присутствии PEG значительно снижает эффективность трансформации. Буферы и мастер–миксы лигирования,

содержащие высокие концентрации краудинг–агентов, также могут быть очень вязкими, что осложняет автоматическую обработку жидкости и высокопроизводительную обработку образцов.

Для повышения эффективности лигирования в технологических процессах клонирования и NGS использовали энхансеры лигирования с более низкой молекулярной массой (например, 1,2–пропандиол), (см., например, публикацию заявки на патент США № 2014/0187447, включенную в данный документ посредством ссылки). Использование буфера для лигирования, 1,2–пропандиола, в 5 раз увеличивало преобразование NGS библиотеки относительно буфера PEG при использовании входящей ДНК с низкой концентрацией (5 нг ДНК *E. coli*). Тогда как 1,2–пропандиол и другие низкомолекулярные энхансеры совместимы с тепловой инактивацией, они могут быть несовместимы с другими ферментами или стадиями в одноканальных NGS технологических процессах или при микроструйном получении образцов.

Варианты сконструированной улучшенной T4 ДНК–лигазы разработали путем создания разных слияний пептидов с доменами связывания двухцепочечной ДНК (см. Wilson et al., *Prot. Engin. Des. Select.*, 7:471–478 [2013]). Несколько этих вариантов лигаз хорошо экспрессировались и демонстрировали улучшенную активность на субстратах с когезивными концами при концентрации ~40 нМ или на субстратах с тупыми концами при концентрации ~30 нМ. Однако, не было ни одного сообщения, что эти лигазы тестировали при низких концентрациях субстрата, обычно используемых для внеклеточных входящих ДНК (1–5 нМ) или других низко затратных вариантов применения NGS.

Чувствительность реакций секвенирования на основе NGS в настоящее время ограничена степенью преобразования фрагментированной входящей ДНК в двухсторонние лигированные с помощью адаптера фрагменты, которые могут составлять до 5%, с использованием входящих компонентов субстрата низкой концентрации. Несмотря на то, что разные лигазы, краудинг–агенты и энхансеры лигирования находят применение, чувствительность анализов секвенирования на основе NGS и устойчивость других молекулярных биологических технологических процессов ограничены текущими доступными способами. Сконструированные лигазы, композиции и способы согласно настоящему изобретению удовлетворяют потребности в данной области в улучшенных диагностических способах и методиках на основе NGS секвенирования и других.

В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению находят применение в диагностических и исследовательских вариантах применения с использованием небольших количеств ДНК из образцов пациентов, включая внеклеточную ДНК, циркулирующую опухолевую ДНК, ДНК, выделенную из циркулирующих опухолевых клеток, циркулирующую фетальную ДНК, ДНК, выделенную из инфицированных вирусами клеток, тонкоигольные аспираты или отдельные клетки, выделенные посредством FACS (сортировки флуоресцентно–активированных клеток), лазерной микроскопии или микроструйных устройств. Однако,

не предполагается, чтобы образец, используемый в настоящем изобретении, был ограничен каким-либо конкретным типом образца, так как в настоящем изобретении находит применение любой подходящий образец, включая образцы с низкими концентрациями ДНК.

В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению находят применение в создании библиотек секвенирования ДНК для интермедиата в образцы ДНК высокой концентрации. Сконструированным лигазам, представленным в данном документе, требуются более низкие концентрации адаптера для достижения эквивалента преобразования лигирования с двойным адаптером в WT ДНК-лигазы (например, T4 ДНК-лигазу дикого типа). В результате использования более низких концентраций адаптера минимизируют получение димеров адаптеров. В некоторых вариантах осуществления адаптеры используют в ограничивающих концентрациях таким образом, чтобы нужный продукт, лигированный с помощью двойного адаптера, находился с высоким молярным превышением относительно димера адаптера, и исключают стадии очистки, которые в противном случае требуются для удаления димера адаптера. Это упрощает множество стандартных технологических процессов секвенирования, таких как ресеквенирование генома. Более низкие концентрации адаптера также снижают количество адаптера, требуемое для лигирования, что может снижать влияние стоимости адаптера на общий технологический процесс. Это особенно полезно для дорогостоящих адаптеров, таких как метилированные адаптеры, используемые для бисульфитного секвенирования или секвенирования метилома.

В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению находят применение в вариантах использования с молекулярным клонированием, особенно в вариантах, где концентрация ДНК является низкой по сравнению с K_m существующих в природе ферментов. В некоторых вариантах осуществления это применимо к вариантам применения с высокопроизводительным клонированием, где образец получают в небольших объемах, или к любому образцу ДНК с низкой концентрацией, такому как образцы окружающей среды, образцы пациента или древняя ДНК.

В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению находят применение в упрощенных технологических процессах в молекулярной биологии, включая автоматизированные технологические процессы, в которых исключены стадии очистки между операциями. Так как сконструированные лигазы активны на субстратах низкой концентрации, в реакционную смесь лигирования можно добавлять меньший объем (или разведение) образца субстрата, содержащего ингибитор. Релевантные содержащие ингибитор образцы ДНК могут включать ДНК в ПЦР буфере, ДНК в электрофорезном буфере или ДНК в неочищенных экстрактах. Сконструированные лигазы согласно настоящему изобретению способны эффективно лигировать разбавленные образцы по сравнению с нативными лигазами. Альтернативно, в других вариантах осуществления сконструированные лигазы согласно настоящему

изобретению находят применение на неразбавленных образцах, содержащих ингибитор (ингибиторы). Благодаря их более высокой специфической активности эффективность лигирования этих сконструированных лигаз превышает эффективность лигаз дикого типа (например, T4 ДНК–лигазы дикого типа) в присутствии ингибиторов.

Тепловую инактивацию T4 ДНК–лигазы обычно выполняют при реакциях лигирования перед трансформацией *E. coli*. Данная стадия является удобной по сравнению с очисткой на основе колонки или гранул и существенно увеличивает эффективность трансформации плазмиды. Для увеличения скорости и эффективности T4 ДНК–лигазы (и других нативных лигаз) часто используют полиэтиленгликоль (PEG), но тепловая инактивация лигазы в присутствии PEG сильно ингибирует трансформацию. Таким образом, для PEG лигирования требуются более сложные стадии очистки, сводя на нет удобство и скорость стратегии быстрого PEG лигирования. В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению выполняют быстрое и эффективное лигирование в отсутствии PEG и могут быть инактивированы с помощью теплового преобразования перед трансформацией. Таким образом, использование этих сконструированных лигаз обеспечивает как быстрое лигирование, так и удобную очистку перед трансформацией.

В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению находят применение в микроструйных вариантах применения, включая варианты, которые несовместимы с очень вязкими энхансерами лигирования или краудинг–агентами (например, PEG, Ficoll или высокими концентрациями глицерина). Сконструированные лигазы согласно настоящему изобретению эффективно преобразуют субстраты низкой концентрации в отсутствие краудинг–агентов, которые требовались бы для достижения аналогичной эффективности с использованием нативной T4 ДНК–лигазы или других лигаз.

В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению находят применение в однокамерных многоферментных реакциях, выполняемых в микроструйных каплях или планшетах с лунками. Высокоспецифическая активность лигаз предусматривает буферные составы, выбираемые по эффективности других ферментов в реакции, которые достигают эффективности лигирования, которая не ограничивает общий технологический процесс.

В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению находят применение в создании библиотек ДНК. Эти библиотеки можно использовать для секвенирования ДНК, высокопроизводительного скрининга, генетического отбора, фагового дисплея, дрожжевого дисплея, рибосомального дисплея, клеточных анализов, биохимических анализов или одновременного многопараметрического анализа на основе изображений. В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению находят специальное применение, когда размер, разнообразие или точность библиотеки ограничены концентрацией субстрата для лигирования при использовании лигазы дикого

типа.

В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению находят применение при проведении анализов на основе определения конформации хроматина (C3), включая 3C, 4C, 5C и Hi-C. Эти анализы зависят от эффективного лигирования расщепленных субстратов в очень разбавленных условиях, которые требуются для стимулирования внутримолекулярного лигирования. В некоторых вариантах осуществления сконструированные лигазы согласно настоящему изобретению действуют более эффективно, чем T4 ДНК-лигаза дикого типа в этих условиях.

Сокращения и Определения:

Если не указано иное, все технические и научные термины, используемые в данном документе, обычно имеют такое же значение, которое обычно понятно рядовому специалисту в области, к которой относится данное изобретение. Обычно, терминология, используемая в данном документе, и лабораторные методики клеточной культуры, молекулярной генетики, микробиологии, органической химии, аналитической химии и химии нуклеиновых кислот, описанные ниже, хорошо известны и широко используются в данной области. Такие методики хорошо известны и описаны в многочисленных текстах и справочных работах, хорошо известных специалистам в данной области. Стандартные методики или их модификации используют для химического синтеза и химического анализа.

Все патенты, заявки на патенты, статьи и публикации, упоминаемые в данном документе, как выше, так и ниже, в явном виде включены настоящим в данный документ посредством ссылки.

Хотя при практическом использовании настоящего изобретения находят применение любые подходящие способы и материалы, аналогичные или эквивалентные способам и материалам, описанным в данном документе, в данном документе описаны некоторые способы и материалы. Должно быть понятно, что данное изобретение не ограничено конкретными описанными методикой, протоколами и реагентами, так как они могут варьировать в зависимости от контекста, в котором их используют специалисты в данной области. Соответственно, термины, приведенные непосредственно ниже, более полно описаны посредством ссылки на заявку в целом. Все патенты, заявки на патенты, статьи и публикации, упоминаемые в данном документе, как выше, так и ниже, в явном виде включены настоящим в данный документ посредством ссылки.

В рамках изобретения, если в контексте четко не указано иное, формы единственного числа включают ссылки на множественное число.

Числовые диапазоны включают числа, ограничивающие диапазон. Таким образом, каждый числовой диапазон, раскрытый в данном документе, предназначен для охвата каждого более узкого числового диапазона, который попадает в пределы такого более широкого числового диапазона, как если бы в данном документе были явно описаны все такие более узкие числовые диапазоны. Также предполагается, что каждое максимальное

(или минимальное) числовое ограничение, раскрытое в данном документе, включает каждое более низкое (или более высокое) числовое ограничение, как если бы такие более низкие (или более высокие) числовые ограничения были явно описаны в данном документе.

Термин «приблизительно» означает допустимую ошибку для конкретного значения. В некоторых случаях «приблизительно» означает в пределах 0,05%, 0,5%, 1,0% или 2,0%, заданного диапазона значений. В некоторых случаях «приблизительно» означает в пределах 1, 2, 3 или 4 стандартных отклонений от заданного значения.

Кроме того, представленные в данном документе заголовки, не являются ограничениями разных аспектов или вариантов осуществления изобретения, которые могут иметься посредством ссылки на заявку в целом. Соответственно, термины, приведенные непосредственно ниже, более полно определены посредством ссылки на заявку в целом. Тем не менее, чтобы облегчить понимание изобретения, ниже определен ряд терминов.

Если не указано иное, нуклеиновые кислоты описаны слева направо в ориентации от 5' к 3'; аминокислотные последовательности описаны слева направо в ориентации от amino к карбокси, соответственно.

В рамках изобретения термин «содержащий» и его однокоренные слова используют в их включительном смысле (т.е. как эквивалент термина «включающий» и его соответствующих однокоренных слов).

В рамках изобретения число «ЕС» относится к Номенклатуре Ферментов комитета по номенклатуре международного союза биохимии и молекулярной биологии (NC-IUBMB). Биохимическая классификация IUBMB представляет собой систему числовой классификации для ферментов на основе химической реакции, которую они катализируют.

В рамках изобретения «ATCC» относится к Американской коллекции типовых культур, собрание биологических материалов которой включает гены и штаммы.

В рамках изобретения «NCBI» относится к Национальному центру для биологической информации и имеющимся в нем базам данных последовательностей.

В рамках изобретения термин «ДНК» относится к дезоксирибонуклеиновой кислоте.

В рамках изобретения термин «РНК» относится к рибонуклеиновой кислоте.

В рамках изобретения термины «белок слияния» и «химерный белок» и «химера» относятся к гибридным белкам, созданным посредством соединения двух или более генов, которые первоначально кодировали отдельные белки. В некоторых вариантах осуществления белки слияния создают с помощью рекомбинантной технологии (например, методик молекулярной биологии, известных в данной области).

В рамках изобретения термин «лигаза» относится к классу ферментов, которые обычно используют для соединения полинуклеотидов вместе или для соединения концов одного полинуклеотида. Лигазы включают АТФ-зависимые двухцепочечные

полинуклеотидные лигазы, NAD^+ –зависимые двухцепочечная ДНК или РНК лигазы и одноцепочечные полинуклеотидные лигазы. В некоторых вариантах осуществления в настоящем изобретении представлены лигазы бактериофагов (например, Т3 ДНК–лигаза, Т4 ДНК–лигаза и Т7 ДНК–лигаза) и их варианты. В некоторых дополнительных вариантах осуществления в настоящем изобретении представлены лигазы слияния или химерные лигазы. ДНК–лигазы часто находят применение с рестрикционными ферментами для вставки фрагментов ДНК (например, генов) в плазмиды. Для лигирования фрагментов с когезивными концами при выполнении эффективной рекомбинации важно регулирование оптимальной температуры. Т4 ДНК–лигаза наиболее активна при 37°C, но для оптимальной эффективности лигирования с фрагментами с когезивными концами оптимальную температуру для фермента следует уравновесить с температурой плавления лигируемых концов; чем короче выступ, тем ниже температура плавления фрагментов. Реакции лигирования имеют тенденцию к наибольшей эффективности, когда когезивные концы уже стабильно отожджены. Для лигирования фрагментов ДНК с тупыми концами температура плавления не является фактором, который следует принимать во внимание, когда реакция происходит в пределах нормальных диапазонов температур, используемых для лигирования. В этих реакциях ограничивающим фактором является количество выравниваний между концами фрагментов ДНК, которые могут происходить, а не активность лигазы. Таким образом, наиболее эффективной температурой для лигирования фрагментов ДНК с тупыми концами является температура, при которой в реакции может происходить наибольшее количество выравниваний.

В рамках изобретения термин «адаптер» относится к одно или двухцепочечному олигонуклеотиду с совместимыми концами ДНК для лигирования. Концы адаптера могут быть одно или двухцепочечными и могут содержать выступы, совместимые с комплементарными выступами на обработанной ДНК вставки библиотеки. Адаптеры могут иметь как одноцепочечные, так и двухцепочечные области. В некоторых вариантах осуществления термин «адаптер» используют для ссылки на полноразмерные адаптеры, используемые в реакциях NGS (т.е. секвенирования следующего поколения), которые могут содержать участки связывания праймеров, штрихкоды и другие особенности, а также ссылку на адаптеры упрощенной модели, используемые в НТР анализах скрининга и лигирования, имеющие одинаковые совместимые с лигированием концы в качестве полноразмерных адаптеров, но не имеющие этих дополнительных особенностей. NGS адаптеры, разработанные для использования на платформе для секвенирования Illumina®, имеют дезокситимидиновые 3' выступы, совместимые с лигированием с дезоксиаденозиновыми 3' выступами, присутствующими на А–образных фрагментах вставки. Т–образные адаптеры не эффективно лигируют друг с другом вследствие избирательности Т4 ДНК–лигаз дикого типа к некомплементарным ДНК концам. Димеризация адаптера будет происходить в результате экстремальных условий лигирования, включая длительные периоды инкубации, высокие концентрации адаптера

или высокие концентрации краудинг-агента. Важно, нуклеазные контаминанты в реакции лигирования могут удалять выступы на концах адаптера, что приводит к субстратам с тупыми концами, которые совместимы с самолигированием.

В рамках изобретения термин «совместимые концы» относится к концам двух двойных фрагментов ДНК с 5' или 3' выступами, которые гибридизируют в 5'–3' антипараллельной ориентации таким образом, чтобы все основания на выступах были комплементарными. В контексте лигирования по меньшей мере один фрагмент ДНК должен иметь 5' фосфат на нуклеотиде, который расположен рядом с 3' гидроксилом нуклеотида из другой молекулы при гибридизации 3' или 5' выступа. Лигирование приводит к ковалентной связи двух молекул субстрата на совместимых концах. В некоторых вариантах осуществления, включая получение библиотеки для секвенирования ДНК, две молекулы ДНК, такие как адаптер и фрагмент вставки, должны иметь совместимые концы, и обе нити гибрида адаптер/вставка должны быть лигированы, чтобы обеспечить продуктивную амплификацию библиотеки посредством ПЦР или секвенирование посредством удлинения полимеразой праймера, гибридизированного к адаптеру.

В рамках изобретения термин «выступ» относится к области одного или более непарных полинуклеотидов, имеющихся на конце двухцепочечного фрагмента ДНК. В непарной области может иметься либо 5', либо 3' конец ДНК. Двухцепочечным фрагментом ДНК может быть дуплекс из двух комплементарных одноцепочечных полинуклеотидов, или это может быть один полинуклеотид с самокомплементарностью, который образует область двухцепочечной ДНК.

В рамках изобретения термины «дуплекс» и «ds» относятся к молекуле двухцепочечной нуклеиновой кислоты (например, ДНК), состоящей из двух одноцепочечных полинуклеотидов, которые являются комплементарными в их последовательности (А парная с Т, С парная с G), расположенной в антипараллельной 5'–3' ориентации и удерживаемой вместе с помощью водородных связей между нуклеиновыми основаниями (т.е. аденином [A], гуанином [G], цитозином [C] и тиминем [T]).

В рамках изобретения термин «тупые» относится к концу дуплекса ДНК или одноцепочечной («ss») ДНК с самокомплементарностью, который не имеет 5' или 3' выступа. Тупые концы могут иметь 5' фосфаты на одной или обеих нитях, которые делают их совместимыми с лигированием посредством лигазы, такой как T4 ДНК-лигаза.

В рамках изобретения термин «димер адаптера» относится к любому ковалентному продукту лигирования между двумя адаптерами. Димеры адаптеров могут образовываться в процессе реакций лигирования.

В рамках изобретения термины «библиотечная вставка» и «вставка» относятся к двухцепочечному фрагменту ДНК, который был обработан посредством концевого восстановления и/или А-наращивания, для предоставления совместимых концов для использования в лигировании адаптера.

В рамках изобретения термин «концевое восстановление» относится к способам восстановления ДНК (например, фрагментированной или поврежденной ДНК или молекул ДНК, которые несовместимы с другими молекулами ДНК). В некоторых вариантах осуществления способ включает две функции: 1) преобразование двухцепочечной ДНК с выступами в двухцепочечную ДНК без выступов с помощью такого фермента, как Т4 ДНК полимеразы и/или фрагмент Кленова; и 2) добавление фосфатной группы в 5' концам ДНК (одно- или двухцепочечной) с помощью такого фермента, как полинуклеотидкиназа.

В рамках изобретения термин «А-наращивание» относится к добавлению одного дезоксиаденозинового остатка к концу фрагмента двухцепочечной ДНК с тупыми концами для образования 3' выступа из одного основания дезоксиаденозина. А-концевые фрагменты не совместимы с самолигированием (т.е. самораспространения и соединения ДНК), но они совместимы с 3' выступами дезокситимидина, такими как выступы, имеющиеся на адаптерах.

В рамках изобретения термин «блокированный аминокислотой» относится к концу одно- или двухцепочечной ДНК, для которого 3' гидроксил был заменен 6-углеродным линкером, заканчивающимся аминокислотой. ДНК-лигазы не могут катализировать реакции лигирования в отсутствие 3' гидроксила.

«Белок», «полипептид» и «пептид» используют в данном документе взаимозаменяемо для обозначения полимера по меньшей мере двух аминокислот, ковалентно связанных амидной связью, независимо от длины или посттрансляционной модификации (например, гликозилирования или фосфорилирования).

«Аминокислоты» обозначены в данном документе либо с помощью их общеизвестных трехбуквенных символов, либо с помощью однобуквенных символов, рекомендованных комиссией по биохимической номенклатуре. Нуклеотиды также могут быть обозначены с помощью их общепринятых однобуквенных кодов.

Термины «сконструированный», «рекомбинантный», «несуществующий в природе» и «вариант» при использовании со ссылкой на клетку, полинуклеотид или полипептид относятся к материалу или материалу, соответствующему натуральной или нативной форме материала, который был модифицирован таким образом, чтобы он иначе не существовал в природе, или является идентичным ему, но получен или произошел из синтетических материалов и/или путем манипулирования с использованием рекомбинантных методов.

В рамках изобретения «дикий тип» и «существующий в природе» относятся к форме, обнаруженной в природе. Например, полипептидная или полинуклеотидная последовательность дикого типа представляет собой последовательность, присутствующую в организме, которую можно выделить из источника в природе и которая не была преднамеренно модифицирована человеческими манипуляциями.

В рамках изобретения «кодирующая последовательность» относится к той части нуклеиновой кислоты (например, гену), которая кодирует аминокислотную

последовательность белка.

В рамках изобретения термин «процент (%) идентичности последовательности» относится к сравнениям между полинуклеотидами и полипептидами, и его определяют путем сравнения двух оптимально выровненных последовательностей в окне сравнения, при этом часть полинуклеотидной или полипептидной последовательности в окне сравнения может содержать добавления или делеции (т.е. пробелы) по сравнению с эталонной последовательностью для оптимального выравнивания двух последовательностей. Процентное значение можно рассчитать путем определения количества положений, в которых идентичное основание нуклеиновой кислоты или аминокислотный остаток встречается в обеих последовательностях, для получения количества совпадающих положений, деления количества совпадающих положений на общее количество положений в окне сравнения и умножения результата на 100 для получения процентного значения идентичности последовательности. Альтернативно, процентное значение можно рассчитать путем определения количества положений, в которых либо идентичное основание нуклеиновой кислоты, либо аминокислотный остаток встречается в обеих последовательностях, или основание нуклеиновой кислоты или аминокислотный остаток выровнен с пробелом, для получения количества совпадающих положений, деления количества совпадающих положений на общее количество положений в окне сравнения и умножения результата на 100 для получения процентного значения идентичности последовательности. Специалистам в данной области понятно, что существует множество доступных установленных алгоритмов для выравнивания двух последовательностей. Оптимальное выравнивание последовательностей для сравнения можно проводить, например, с помощью алгоритма поиска локальной гомологии Смита–Уотермана (Smith и Waterman, *Adv. Appl. Math.*, 2:482 [1981]), с помощью алгоритма выравнивания областей гомологии Нидлмана–Вунша (Needleman и Wunsch, *J. Mol. Biol.*, 48:443 [1970]), с помощью способа поиска сходства по Пирсону–Липману (Pearson и Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 [1988]), с помощью компьютерного выполнения этих алгоритмов (например, GAP, BESTFIT, FASTA и TFASTA в программном пакете GCG Wisconsin) или с помощью визуального осмотра, как известно в данной области. Примеры алгоритмов, которые подходят для определения процента идентичности последовательности и сходства последовательности включают, но без ограничения, алгоритмы BLAST и BLAST 2,0 (см., например, Altschul et al., *J. Mol. Biol.*, 215: 403–410 [1990]; и Altschul et al., *Nucleic Acids Res.*, 3389–3402 [1977]). Программное обеспечение для проведения анализов BLAST общедоступно на веб-сайте Национального центра биотехнологической информации. Этот алгоритм включает в себя сначала идентификацию пар последовательностей с высокой оценкой (HSP) путем идентификации коротких слов с длиной «W» в запрашиваемой последовательности, которые либо совпадают, либо удовлетворяют некоторой пороговой оценке «Т» с положительным значением при выравнивании со словом такой же длины в последовательности базы данных. Т упоминается как пороговое значение оценки соседних слов (см. Altschul et al,

выше). Эти первоначальные совпадения соседних слов выступают в качестве затравки для начала поиска, чтобы найти более длинные HSP, содержащие их. Затем совпадения слов распространяют в обоих направлениях вдоль каждой последовательности насколько возможно увеличить совокупную оценку выравнивания. Совокупные оценки рассчитывают с использованием для нуклеотидных последовательностей параметров «M» (вознаграждающая оценка для пары совпадающих остатков; всегда >0) и «N» (штрафная оценка для несовпадающих остатков; всегда <0). Для аминокислотных последовательностей для расчета совокупной оценки используют матрицу замен. Нарастивание совпадений слов в каждом направлении останавливают, когда: совокупная оценка выравнивания падает на величину «X» от его максимально достигнутого значения; совокупная оценка падает до нуля или ниже вследствие накопления одного или нескольких выравниваний остатков с отрицательной оценкой; или достигнут конец любой последовательности. Параметры алгоритма BLAST W, T, и X определяют чувствительность и скорость выравнивания. В программе BLASTN (для нуклеотидных последовательностей) используют по умолчанию длину слова (W) 11, ожидание (E) 10, M=5, N=-4 и сравнение обеих нитей. Для аминокислотных последовательностей в программе BLASTN по умолчанию используют длину слова (W) 3, ожидание (E) 10 и матрицу замен BLOSUM62 (см., например, Henikoff и Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 [1989]). Для иллюстративного определения выравнивания последовательности и % идентичности последовательностей можно использовать программы BESTFIT или GAP в пакете программного обеспечения GCG Wisconsin (Accelrys, Madison WI) с использованием по умолчанию предоставленных параметров.

В рамках изобретения «эталонная последовательность» относится к определенной последовательности, используемой в качестве основы для сравнения последовательностей. Эталонной последовательностью может быть подмножество большей последовательности, например, сегмент полноразмерного гена или полипептидной последовательности. Обычно, эталонная последовательность составляет по меньшей мере 20 нуклеотидов или аминокислотных остатков в длину, по меньшей мере 25 остатков в длину, по меньшей мере 50 остатков в длину, по меньшей мере 100 остатков в длину или полную длину нуклеиновой кислоты или полипептида. Поскольку каждый из двух полинуклеотидов или полипептидов (1) может содержать последовательность (т.е. часть полной последовательности), которая похожа между двумя последовательностями, и (2) может дополнительно содержать последовательность, которая расходится между двумя последовательностями, сравнение последовательностей между двумя (или более) полинуклеотидами или полипептидами обычно выполняют путем сравнения последовательностей двух полинуклеотидов или полипептидов в «окне сравнения» для идентификации и сравнения локальных областей сходства последовательностей. В некоторых вариантах осуществления «эталонная последовательность» может быть основана на первичной аминокислотной последовательности, где эталонная последовательность представляет собой

последовательность, которая может иметь одно или более изменений в первичной последовательности. Например, фраза «эталонная последовательность на основе SEQ ID NO: 4, имеющая валин в остатке, соответствующем X39», относится к эталонной последовательности, в которой соответствующий остаток в положении X39 в SEQ ID NO: 4 (например, аланин), был заменен на валин.

В рамках изобретения «окно сравнения» относится к концептуальному сегменту из по меньшей мере приблизительно 20 смежных положений нуклеотидов или аминокислотных остатков, в котором последовательность можно сравнить с эталонной последовательностью по меньшей мере из 20 смежных нуклеотидов или аминокислот, и при этом часть последовательности в окне сравнения может содержать добавления или делеции (т.е. пробелы), составляющие 20 процентов или менее, по сравнению с эталонной последовательностью (которая не содержит добавлений или делеций) для оптимального выравнивания двух последовательностей. Окно сравнения может быть более длинным, чем 20 смежных остатков, и включает в себя, необязательно 30, 40, 50, 100 или более длинных окон.

В рамках изобретения «соответствующий», «ссылка на» и «относительно» при использовании в контексте нумерации заданной аминокислотной или полинуклеотидной последовательности относятся к нумерации остатков указанной эталонной последовательности, когда заданную аминокислотную или полинуклеотидную последовательность сравнивают с эталонной последовательностью. Другими словами, номер остатка или положение остатка заданного полимера обозначают относительно эталонной последовательности, а не фактическим числовым положением остатка в пределах заданной аминокислотной или полинуклеотидной последовательности. Например, заданную аминокислотную последовательность, такую как последовательность сконструированной лигазы, можно выравнивать с эталонной последовательностью путем введения пробелов для оптимизации совпадений остатков между двумя последовательностями. В этих случаях, несмотря на наличие пробелов, нумерация остатка в данной аминокислотной или полинуклеотидной последовательности производится относительно эталонной последовательности, с которой ее выравнивали. В некоторых вариантах осуществления последовательность метят (например, гистидиновой меткой).

В рамках изобретения «различие аминокислот» и «различие остатков» относится к различию в аминокислотном остатке в положении полипептидной последовательности относительно аминокислотного остатка в соответствующего положения в эталонной последовательности. Положения разных аминокислот обычно называются в данном документе «X_n», где n относится к соответствующему положению в эталонной последовательности, на котором основано различие остатков. Например, «различие остатков в положении X91 по сравнению с SEQ ID NO: 4» относится к различию аминокислотного остатка в положении полипептида, соответствующего положению 91 SEQ ID NO: 4. Таким образом, если эталонный полипептид SEQ ID NO: 4 имеет в положении 91 аланин, то «различие остатков в положении X91 по сравнению с SEQ ID

NO: 4» относится к аминокислотной замене любого остатка кроме аланина в положении полипептида, соответствующему положению 91 SEQ ID NO: 4. В большинстве случаев в данном документе конкретное различие аминокислотных остатков в положении обозначено «X_nY», где «X_n» указывает соответствующий остаток и положению эталонного полипептида (как описано выше), а «Y» представляет собой однобуквенный идентификатор аминокислоты, обнаруженной в сконструированном полипептиде (т.е. иной остаток, чем в эталонном полипептиде). В некоторых случаях (например, в таблицах в примерах) в настоящем раскрытии также представлено конкретное различие аминокислот, обозначенное общепринятым обозначением «A_nB», где A представляет собой однобуквенный идентификатор остатка в эталонной последовательности, «n» представляет собой номер положения остатка в эталонной последовательности, а B представляет собой однобуквенный идентификатор замены остатка в последовательности сконструированного полипептида. В некоторых случаях полипептид согласно настоящему раскрытию может содержать одно или более различий аминокислотных остатков относительно эталонной последовательности, что указано с помощью списка указанных положений, где различия остатков представлены относительно эталонной последовательности. В некоторых вариантах осуществления, где в конкретном положении остатка полипептида может быть использована более чем одна аминокислота, разные аминокислотные остатки, которые можно использовать, разделены «/» (например, X307G/X307Q или X307G/Q). В настоящем раскрытии включены сконструированные полипептидные последовательности, содержащие одну или более разных аминокислот, которые включают либо/или как консервативные, так и неконсервативные аминокислотные замены.

В рамках изобретения термины «набор аминокислотных замен» и «набор замен» относится к группе аминокислотных замен в пределах полипептидной последовательности. В некоторых вариантах осуществления наборы замен включают в себя 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 или более аминокислотных замен. В некоторых вариантах осуществления набор замен относится к набору аминокислотных замен, который присутствует в любом из вариантных полипептидов лигазы, перечисленных в любой из Таблиц в примерах. В этих наборах замен отдельные замены разделены точкой с запятой («;»; например, C165A;S181T;K299P) или косой чертой («/»; например, C165A/S181T/K299P).

В рамках изобретения «консервативная аминокислотная замена» относится к замене остатка другим остатком, имеющим похожую боковую цепь, и таким образом обычно включает замену аминокислоты в полипептиде аминокислотами в пределах того же или аналогичного определенного класса аминокислот. В качестве примера, а не ограничения, аминокислота с алифатической боковой цепью может быть заменена другой алифатической аминокислотой (например, аланин, валин, лейцин и изолейцин); аминокислоту с гидроксильной боковой цепью заменяют другой аминокислотой с гидроксильной боковой цепью (например, серин и треонин); аминокислоты, имеющие

ароматические боковые цепи, заменяют другой аминокислотой, имеющей ароматическую боковую цепь (например, фенилаланин, тирозин, триптофан и гистидин); аминокислоту с основной боковой цепью заменяют другой аминокислотой с основной боковой цепью (например, лизин и аргинин); аминокислоту с кислой боковой цепью заменяют другой аминокислотой с кислой боковой цепью (например, аспарагиновая кислота или глутаминовая кислота); а гидрофобную или гидрофильную аминокислоту заменяют другой гидрофобной или гидрофильной аминокислотой, соответственно.

В рамках изобретения «неконсервативные замена» относится к замене аминокислоты в полипептиде аминокислотой со значительно отличающимися свойствами боковой цепи. В неконсервативных заменах можно использовать аминокислоты между, а не в пределах, определенных групп, и они влияют на: (a) структуру остова пептида в области замены (например, пролин для глицин); (b) заряд или гидрофобность; и/или (c) большую часть боковой цепи. В качестве примера, а не ограничения, иллюстративные неконсервативные замены включают кислую аминокислоту, замененную основной или алифатической аминокислотой; ароматическую аминокислоту, замененную малой аминокислотой; и гидрофильную аминокислоту, замененную гидрофобной аминокислотой.

В рамках изобретения «делеция» относится к модификации полипептида за счет удаления из эталонного полипептида одной или более аминокислот. Делеции могут включать удаление 1 или более аминокислот, 2 или более аминокислот, 5 или более аминокислот, 10 или более аминокислот, 15 или более аминокислот или 20 или более аминокислот, до 10% общего количества аминокислот, или до 20% общего количества аминокислот, составляющих эталонный фермент, сохраняя при этом ферментативную активность и/или сохраняя улучшенные свойства сконструированного трансаминазного фермента. Делеции могут быть направлены на внутренние части и/или терминальные части полипептида. В разных вариантах осуществления делеция может содержать непрерывный сегмент или может быть прерывистой.

В рамках изобретения «вставка» относится к модификации полипептида путем добавления одной или более аминокислот из эталонного полипептида. Вставки могут быть во внутренних частях полипептида или на карбокси или амино концах. В рамках изобретения вставки включают белки слияния, которые известны в данной области. Вставка может представлять собой смежный сегмент аминокислот или может быть отделена одной или более аминокислотами в существующем в природе полипептиде.

В рамках изобретения «функциональный фрагмент» и «биологически активный фрагмент» используют в данном документе взаимозаменяемо для ссылки на полипептид, который имеет амино-терминальные и/или карбокси-терминальные делецию (делеции) и/или внутренние делеции, но где остальная аминокислотная последовательность идентична соответствующим положениям в последовательности, с которой ее сравнивают (например, полноразмерная сконструированная лигаза согласно настоящему изобретению) и которая сохраняет по существу всю активность полноразмерного

полипептида.

В рамках изобретения «выделенный полипептид» относится к полипептиду, который по существу отделен от других контаминантов, которые сопровождают его в природе (например, белок, липиды и полинуклеотиды). Термин охватывает полипептиды, которые были удалены или очищены из их естественной среды или системы экспрессии (например, клетка-хозяин или синтез *in vitro*). Полипептиды рекомбинантной лигазы могут присутствовать в клетке, присутствовать в клеточной среде или могут быть получены в различных формах, таких как лизаты или выделенные препараты. В связи с этим, в некоторых вариантах осуществления представленными в данном документе полипептидами рекомбинантной лигазы являются выделенные полипептиды.

В рамках изобретения «по существу чистый полипептид» относится к композиции, в которой полипептидные частицы являются преобладающими присутствующими частицами (т.е. в молярном или массовом отношении он является более распространенным, чем любые другие отдельные макромолекулярные частицы в композиции), и обычно представляет собой по существу очищенную композицию, когда на целевые частицы приходится по меньшей мере приблизительно 50 процентов макромолекулярных частиц, присутствующих в молях или в мас.%. Обычно, композиция по существу чистой лигазы будет содержать приблизительно 60% или более, приблизительно 70% или более, приблизительно 80% или более, приблизительно 90% или более, приблизительно 95% или более и приблизительно 98% или более всех макромолекулярных частиц в молях или в мас.%, присутствующих в композиции. В некоторых вариантах осуществления целевые частицы очищают до существенной гомогенности (т.е. контаминантные частицы нельзя обнаружить в композиции обычными способами обнаружения), при этом композиция состоит по существу из одной макромолекулярной частицы. Частицы растворителя, малые молекулы (<500 Дальтон) и элементарные ионные частицы не считаются макромолекулярными частицами. В некоторых вариантах осуществления выделенные полипептиды рекомбинантной лигазы представляют собой по существу чистые полипептидные композиции.

В рамках изобретения «улучшенное свойство фермента» относится к полипептиду сконструированной лигазы, который демонстрирует улучшение любого свойства фермента по сравнению с эталонным полипептидом лигазы, таким как полипептид лигазы дикого типа (например, T4 лигаза дикого типа SEQ ID NO: 2) или другой полипептид сконструированной лигазы. Улучшенные свойства включают, но без ограничения, такие свойства как повышенная экспрессия белка, повышенная термоактивность, повышенная термостабильность, повышенная стабильность, повышенная ферментативная активность, повышенная специфичность и/или аффинность к субстрату, повышенная специфическая активность, повышенная устойчивость к ингибированию субстрата и/или конечного продукта, повышенная химическая стабильность, улучшенное хемоселективность, улучшенная стабильность растворителя, повышенная устойчивость к кислому pH, повышенная устойчивость к протеолитической активности (т.е. пониженная

чувствительность к протеолизу), повышенная растворимость и измененный температурный профиль.

В рамках изобретения «повышенная ферментативная активность» и «повышенная каталитическая активность» относятся к улучшенному свойству полипептидов сконструированной лигазы, которое может быть представлено в виде увеличения специфической активности (например, вырабатываемый продукт/время/масса белка) и/или увеличения процентного преобразования субстрата в продукт (например, процентное преобразование исходного количества субстрата в продукт за указанный период времени при использовании указанного количества лигазы) по сравнению с эталонным ферментом лигазой (например, T4 лигазой дикого типа и/или другой сконструированной лигазой). Иллюстративные способы определения ферментативной активности приведены в примерах. Может быть затронуто любое свойство, относящееся к ферментативной активности, включая классические свойства фермента K_m , V_{max} или k_{cat} , изменения которых могут приводить к повышенной ферментативной активности. Улучшения ферментативной активности могут составлять от приблизительно 1,1 раза по сравнению с ферментативной активностью соответствующего фермента дикого типа, до 2-кратной, 5-кратной, 10-кратной, 20-кратной, 25-кратной, 50-кратной, 75-кратной, 100-кратной, 150-кратной, 200-кратной или более ферментативной активности чем у существующей в природе лигазы или другой сконструированной лигазы, из которой были получены полипептиды лигазы.

Термины «протеолитическая активность» и «протеолиз», используемые в данном документе взаимозаменяемо, относятся к расщеплению белков на меньшие полипептиды или аминокислоты. Расщепление белков обычно является результатом гидролиза пептидной связи ферментами протеазами (протеиназами). Ферменты протеазы включают, но без ограничения, пепсин, трипсин, химотрипсин, эластазу; карбоксипептидазу А и В и пептидазы (например, аминопептидазы, дипептидазу и энтеропептидазу).

Фразы «снижение чувствительности к протеолизу» и «снижение протеолитической чувствительности» используются в данном документе взаимозаменяемо, что означает, что полипептид сконструированной лигазы согласно изобретению будет иметь более высокую ферментативную активность по сравнению с эталонной лигазой в стандартном анализе (например, как описано в примерах) после обработки одной или несколькими протеазами.

В рамках изобретения «конверсия» относится к ферментативному преобразованию (или биотрансформации) субстрата (субстратов) в соответствующий продукт (продукты). «Процент конверсии» относится к проценту субстрата, который превращается в продукт в течение определенного периода времени при определенных условиях. Таким образом, «ферментативная активность» или «активность» полипептида лигазы может быть выражена как «процентное преобразование» субстрата в продукт в течение определенного периода времени.

В рамках изобретения «жесткость гибридизации» относится к условиям гибридизации, таким как условия промывания, при гибридизации нуклеиновых кислот.

Обычно реакции гибридизации проводят в условиях более низкой строгости с последующими промываниями с иной, но более высокой строгостью. Термин «умеренно строгая гибридизация» относится к условиям, которые позволяют ДНК–мишени связывать комплементарную нуклеиновую кислоту, которая имеет идентичность приблизительно 60%, идентичность предпочтительно приблизительно 75%, идентичность приблизительно 85% с целевой ДНК, с идентичностью более чем приблизительно 90% с целевым полинуклеотидом. Примерами умеренно строгих условий являются условия, эквивалентные гибридизации в 50% формамиде, 5–кратном растворе Денхарта, 5–кратном SSPE, 0,2% SDS при 42°C с последующей промывкой в $0,2 \times \text{SSPE}$, 0,2% SDS при 42°C. «Гибридизация высокой строгости» обычно относится к условиям, которые находятся примерно на 10°C или менее от температуры T_m термического плавления, определенной в условиях раствора для определенной полинуклеотидной последовательности. В некоторых вариантах осуществления условие высокой строгости относится к условиям, которые допускают гибридизацию только тех последовательностей нуклеиновых кислот, которые образуют стабильные гибриды в 0,018 M NaCl при 65°C (то есть, если гибрид не стабилен в 0,018 M NaCl при 65°C, он не будет стабильным в условиях высокой жесткости, как предусмотрено в настоящем документе). Условия высокой жесткости можно обеспечить, например, путем гибридизации в условиях, эквивалентных 50% формамиду, 5–кратному раствору Денхарта, 5–кратному SSPE, 0,2% SDS при 42°C, с последующей промывкой в $0,1 \times \text{SSPE}$ и 0,1% SDS при 65°C. Другим условием высокой жесткости является гибридизация в условиях, эквивалентных гибридизации в 5X SSC, содержащей 0,1% (м/о) SDS, при 65°C и промывка в $0,1 \times \text{SSC}$, содержащей 0,1% SDS, при 65°C. Другие условия гибридизации высокой строгости, а также умеренно строгие условия описаны в ссылках, приведенных выше.

В рамках изобретения «кодон оптимизированный» относится к изменениям в кодонах полинуклеотида, кодирующего белок, по сравнению с кодонами, которые преимущественно используются в конкретном организме, так что кодируемый белок более эффективно экспрессируется в этом организме. Хотя генетический код является вырожденным, так как большинство аминокислот представлено несколькими кодонами, называемыми «синонимами» или «синонимичными» кодонами, хорошо известно, что использование кодонов отдельными организмами неслучайно и смещено в сторону конкретных триплетов кодонов. Это смещение использования кодонов может быть выше в отношении данного гена, генов общей функции или наследственного происхождения, высокоэкспрессируемых белков по сравнению с белками с низким числом копий и областей, кодирующих агрегированный белок в геноме организма. В некоторых вариантах осуществления полинуклеотиды, кодирующие ферменты лигазы, оптимизированы по кодонам для оптимальной продукции из организма–хозяина, выбранного для экспрессии.

В рамках изобретения «регуляторная последовательность» относится в данном документе к включению всех компонентов, которые являются необходимыми или предпочтительными для экспрессии полинуклеотида и/или полипептида согласно

настоящему раскрытию. Каждая регуляторная последовательность может быть нативной или чужеродной для последовательности нуклеиновой кислоты, кодирующей полипептид. Такие регуляторные последовательности включают, но без ограничения, лидеры, последовательности аденилирования, пропептидные последовательности, промоторные последовательности, сигнальные пептидные последовательности, последовательности инициации и терминаторы транскрипции. Как минимум, регуляторные последовательности содержат промотор и сигналы остановки транскрипции и трансляции. В некоторых вариантах осуществления регуляторные последовательности снабжены линкерами с целью введения специфических сайтов рестрикции, облегчающих лигирование регуляторных последовательностей с кодирующей областью последовательности нуклеиновой кислоты, кодирующей полипептид.

«Функционально связанный» определен в данном документе в качестве конфигурации, в которой регуляторная последовательность надлежащим образом размещена (т.е. в функциональном отношении) в положении относительно интересующего полинуклеотида таким образом, чтобы регуляторная последовательность направляла или регулировала экспрессию полинуклеотида, кодирующего интересующий полипептид.

В рамках изобретения «промоторная последовательность» относится к последовательности нуклеиновой кислоты, которую распознает клетка-хозяин для экспрессии интересующего полинуклеотида, такой как кодирующая последовательность. Промоторная последовательность содержит регуляторные последовательности транскрипции, которые опосредуют экспрессию интересующего полинуклеотида. Промотором может быть любая последовательность нуклеиновой кислоты, которая демонстрирует транскрипционную активность в выбранной клетке-хозяине, включая мутантные, укороченные и гибридные промоторы, и он может быть получен из генов, кодирующих внеклеточные или внутриклеточные полипептиды либо гомологичные, либо гетерологичные для клетки-хозяина.

В рамках изобретения «подходящие условия реакции» относится к тем условиям в реакционном растворе для ферментативного превращения (например, диапазонам загрузки фермента, загрузки субстрата, температуры, pH, буферов, сорастворителей и т.д.), в которых полипептид лигазы согласно настоящему раскрытию способен превращать субстрат в нужное соединение продукта. В настоящем документе приведены иллюстративные «подходящие условия реакции» (см. примеры).

В рамках изобретения «загрузка», например, «загрузка соединения» или «загрузка фермента» относится к концентрации или количеству компонента в реакционной смеси в начале реакции. «Субстрат» в контексте процесса реакции ферментативного преобразования относится к соединению или молекуле, на которую воздействует полипептид лигазы.

В рамках изобретения «продукт» в контексте способа ферментативного преобразования относится к соединению или молекуле, возникающей в результате

воздействия полипептида лигазы на субстрат.

В рамках изобретения «культивирование» относится к росту популяции микробных клеток в подходящих условиях с использованием любой подходящей среды (например, жидкости, геля или твердого вещества).

Рекомбинантные полипептиды (например, варианты фермента лигазы) могут быть получены с использованием любых подходящих способов, известных в данной области. Например, существует множество различных методов мутагенеза, хорошо известных специалистам в данной области. Кроме того, наборы для мутагенеза также доступны от многих коммерческих поставщиков в молекулярной биологии. Доступны способы выполнения специфических замен определенных аминокислот (сайт-направленных), специфических или случайных мутаций в локализованной области гена (регио-специфических) или случайного мутагенеза по всему гену (например, насыщающего мутагенеза). Специалистам в данной области известно множество подходящих способов создания вариантов ферментов, включая, но без ограничения, сайт-направленный мутагенез одноцепочечной ДНК или двухцепочечной ДНК с использованием ПЦР, кассетный мутагенез, синтез генов, допускающую ошибки ПЦР, перетасовку и химический насыщающий мутагенез или любой другой подходящий метод, известный в данной области. Неограничивающие примеры методов, используемых для инженерии ДНК и белков, представлены в следующих патентах: Патент США 6117679; Патент США 6420175; Патент США № 6376246; Патент США № 6586182; Патент США № 7747391; Патент США № 7747393; Патент США № 7783428; и Патент США № 8383346. После получения вариантов их можно подвергать скринингу на любое желаемое свойство (например, высокую или повышенную активность или низкую или пониженную активность, повышенную термическую активность, повышенную термостабильность и/или стабильность кислого pH и т.д.). В некоторых вариантах осуществления «полипептиды рекомбинантной лигазы» (также называемые здесь «полипептидами рекомбинантной лигазы», «вариантными ферментами лигазы» и «вариантами лигазы») находят применение.

В рамках изобретения «вектор» представляет собой конструкцию ДНК для введения последовательности ДНК в клетку. В некоторых вариантах осуществления вектором является вектор экспрессии, который функционально связан с подходящей регуляторной последовательностью, способной влиять на экспрессию в подходящем хозяине полипептида, кодируемого в последовательности ДНК. В некоторых вариантах осуществления «вектор экспрессии» имеет промоторную последовательность, функционально связанную с последовательностью ДНК (например, трансгеном), для управления экспрессией в клетке-хозяине, и в некоторых вариантах осуществления также содержит последовательность терминатора транскрипции.

В рамках изобретения термин «экспрессия» включает любую стадию, вовлеченную в получение полипептида, включая, но без ограничения, транскрипцию, посттранскрипционную модификацию, трансляцию и посттрансляционную модификацию.

В некоторых вариантах осуществления термин также охватывает секрецию полипептида из клетки.

В рамках изобретения термин «продуцирует» относится к продуцированию белков и/или других соединений клетками. Предполагается, что термин охватывает любую стадию, вовлеченную в получение полипептидов, включая, но без ограничения, транскрипцию, посттранскрипционную модификацию, трансляцию и посттрансляционную модификацию. В некоторых вариантах осуществления термин также охватывает секрецию полипептида из клетки.

В рамках изобретения аминокислотная или нуклеотидная последовательность (например, промоторная последовательность, сигнальный пептид, терминаторная последовательность и т.д.) является «гетерологичной» по отношению к другой последовательности, с которой она функционально связана, если две последовательности по своей природе не связаны.

В рамках изобретения термины «клетка-хозяин» и «штамм-хозяин» относятся к подходящим хозяевам для векторов экспрессии, содержащих ДНК, представленную в данном документе (например, полинуклеотидные последовательности, кодирующие по меньшей мере один вариант лигазы). В некоторых вариантах осуществления клетками-хозяевами являются прокариотические или эукариотические клетки, которые были трансформированы или трансфицированы с помощью векторов, сконструированных с использованием методов рекомбинантные ДНК, известных в данной области.

В рамках изобретения термин «аналог» означает полипептид, имеющий более чем 70% идентичность последовательности, но менее, чем 100% идентичность последовательности (например, более чем 75%, 78%, 80%, 83%, 85%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% идентичность последовательности) с эталонным полипептидом. В некоторых вариантах осуществления аналоги содержат несуществующие в природе аминокислотные остатки, включая, но без ограничения, гомоаргинин, орнитин и норвалин, а также существующие в природе аминокислоты. В некоторых вариантах осуществления аналоги также включают один или более D-аминокислотных остатков и непептидных связей между двумя или более аминокислотными остатками.

В рамках изобретения термин «эффективное количество» означает количество, достаточное для получения нужного результата. Специалист в данной области может определить эффективное количество с использованием обычных экспериментов.

Термины «выделенный» и «очищенный» используют для обозначения молекулы (например, выделенной нуклеиновой кислоты, полипептида и т.д.) или другого компонента, который выделен из по меньшей мере одного другого компонента, с которым она связана в природе. Термин «очищенный» не требует абсолютной чистоты, вместо этого он предназначен в качестве относительного определения.

Термин «субъект» охватывает млекопитающих, таких как люди, нечеловеческие приматы, домашний скот, домашних животных и лабораторных животных (например,

грызунов и зайцеобразных). Предполагается, что термин охватывает самок, а также самцов.

В рамках изобретения термин «пациент» означает любого субъекта, которого оценивают на заболевание, лечат от заболевания или который испытывает заболевание.

В рамках изобретения «композиция» и «готовая форма» охватывает продукты, содержащие по меньшей мере одну

В рамках изобретения «внеклеточная ДНК» относится к ДНК, свободно циркулирующей в кровотоке и не содержащейся в клетках или не связанной с ними. В некоторых вариантах осуществления внеклеточная ДНК содержит ДНК, первоначально полученную и высвобожденную из клеток нормальных соматических или зародышевых линий, раковых клеток, фетальных клеток, клеток микробов или вирусов.

В рамках изобретения «скачкообразное изменение показателя» относится к способу, в результате которого данные секвенирования неправильно присваиваются библиотекам, меченым индексом из другой библиотеки. Скачкообразное изменение показателя может происходить, когда нелигированные адаптерные молекулы не эффективно удаляют в процессе очистки библиотеки, а затем включают во фрагменты библиотеки на более поздних стадиях в технологическом процессе секвенирования.

Полипептиды сконструированной лигазы:

Когда конкретный вариант лигазы (т.е. полипептид сконструированной лигазы) упоминается посредством ссылки на модификацию конкретных аминокислотных остатков в последовательности лигазы дикого типа или эталонной лигазы, должно быть понятно, что данный документ охватывает варианты другой лигазы, модифицированной в эквивалентном положении (положениях) (что определяют из необязательного выравнивания аминокислотных последовательностей между соответствующими аминокислотными последовательностями).

Варианты полипептида сконструированной лигазы согласно настоящему изобретению выполняют определение одноконцевого лигирования в высокопроизводительных анализах с планшетным определением со значительно более высокой скоростью, чем с T4 ДНК–лигазой дикого типа, как указано в примерах. Кроме того, эти варианты лигазы способны выполнять реакции лигирования в отсутствие таких средств, как PEG. Кроме того, эти варианты лигаз демонстрируют повышенную толерантность к ингибиторам.

Варианты сконструированной лигазы согласно настоящему изобретению эффективно создают библиотеки ДНК, подходящие для NGS и других диагностических методов. Эти варианты лигаз находят применение в растворе, а также в иммобилизованных вариантах осуществления.

В некоторых дополнительных вариантах осуществления полипептид сконструированной лигазы согласно настоящему изобретению содержит полипептид, имеющий идентичность последовательности по меньшей мере 70%, по меньшей мере 75%, по меньшей мере 80%, по меньшей мере 85%, по меньшей мере 90%, по меньшей

мере 91%, по меньшей мере 92%, по меньшей мере 93%, по меньшей мере 94%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98% или по меньшей мере 99% с SEQ ID NO: 2, 6, 32, 34 и/или 38.

В некоторых вариантах осуществления полипептиды сконструированной лигазы получают путем культивирования микроорганизмов, содержащих по меньшей мере одну полинуклеотидную последовательность, кодирующую по меньшей мере один полипептид сконструированной лигазы в условиях, которые способствуют получению полипептида сконструированной лигазы. В некоторых вариантах осуществления полипептид сконструированной лигазы впоследствии выделяют из полученной в результате культуральной среды и/или клеток.

В настоящем изобретении представлены иллюстративные полипептиды сконструированной лигазы, имеющие активность лигазы. В примерах представлены Таблицы, показывающие информацию о структуре последовательности, связывающую специфические особенности аминокислотной последовательности с функциональной активностью полипептидов сконструированной лигазы. Данная информация о корреляции функции структуры представлена в виде конкретных различий аминокислотных остатков относительно эталонного сконструированного полипептида SEQ ID NO: 2, 6, 32, 34 и/или 38, а также связанных экспериментально определенных данных об активности для иллюстративных полипептидов сконструированной лигазы.

В некоторых вариантах осуществления полипептиды сконструированной лигазы согласно настоящему изобретению, обладающие активностью лигазы, содержат аминокислотную последовательность, имеющую по меньшей мере 85% идентичность последовательности с эталонной последовательностью SEQ ID NO: 2, 6, 32, 34 и/или 38, и которая демонстрирует по меньшей мере одно улучшенное свойство по сравнению с эталонной последовательностью (например, T4 ДНК–лигазы дикого типа).

В некоторых вариантах осуществления полипептиды сконструированной лигазы, демонстрирующие по меньшей мере одно улучшенное свойство, имеют по меньшей мере 85%, по меньшей мере 88%, по меньшей мере 90%, по меньшей мере 91%, по меньшей мере 92%, по меньшей мере 93%, по меньшей мере 94%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98%, по меньшей мере 99% или большую идентичность аминокислотной последовательности с SEQ ID NO: 2, 6, 32, 34 и/или 38, и различие аминокислотных остатков в одном или более положениях аминокислот (например, в 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 20 или более положениях аминокислот) по сравнению с SEQ ID NO: 2, 6, 32, 34 и/или 38. В некоторых вариантах осуществления полипептид сконструированной лигазы представляет собой полипептид, приведенный в таблицах, предоставленных в примерах.

В некоторых вариантах осуществления в настоящем изобретении представлены функциональные фрагменты полипептидов сконструированной лигазы. В некоторых вариантах осуществления функциональные фрагменты содержат по меньшей мере приблизительно 90%, по меньшей мере приблизительно 95%, по меньшей мере

приблизительно 96%, по меньшей мере приблизительно 97%, по меньшей мере приблизительно 98% или по меньшей мере приблизительно 99% активности полипептида сконструированной лигазы, из которых он был получен (т.е. исходной сконструированной лигазы). В некоторых вариантах осуществления функциональные фрагменты содержат по меньшей мере приблизительно 90%, по меньшей мере приблизительно 91%, по меньшей мере приблизительно 92%, по меньшей мере приблизительно 93%, по меньшей мере приблизительно 94%, по меньшей мере приблизительно 95%, по меньшей мере приблизительно 96%, по меньшей мере приблизительно 97%, по меньшей мере приблизительно 98% или по меньшей мере приблизительно 99% исходной последовательности сконструированной лигазы. В некоторых вариантах осуществления функциональный фрагмент будет укорочен менее, чем на 5, менее, чем на 10, менее, чем на 15, менее, чем на 10, менее, чем на 25, менее, чем на 30, менее, чем на 35, менее, чем на 40, менее, чем на 45 и менее, чем на 50 аминокислот.

В некоторых вариантах осуществления в настоящем изобретении представлены функциональные фрагменты полипептидов сконструированной лигазы. В некоторых вариантах осуществления функциональные фрагменты имеют по меньшей мере приблизительно 95%, 96%, 97%, 98% или 99% активности полипептида сконструированной лигазы, из которого он был получен (т.е. исходной сконструированной лигазы). В некоторых вариантах осуществления функциональные фрагменты содержат по меньшей мере 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% или 99% исходной последовательности сконструированной лигазы. В некоторых вариантах осуществления функциональный фрагмент будет укорочен менее, чем на 5, менее, чем на 10, менее, чем на 15, менее, чем на 10, менее, чем на 25, менее, чем на 30, менее, чем на 35, менее, чем на 40, менее, чем на 45, менее, чем на 50, менее, чем на 55, менее, чем на 60, менее, чем на 65 или менее, чем на 70 аминокислот.

В некоторых вариантах осуществления полипептиды сконструированной лигазы, демонстрирующие по меньшей мере одно улучшенное свойство, имеют по меньшей мере 85%, по меньшей мере 88%, по меньшей мере 90%, по меньшей мере 91%, по меньшей мере 92%, по меньшей мере 93%, по меньшей мере 94%, по меньшей мере 95%, по меньшей мере 96%, по меньшей мере 97%, по меньшей мере 98%, по меньшей мере 99% или большую идентичность аминокислотной последовательности с SEQ ID NO: 2, 6, 32, 34 и/или 38, и различие аминокислотных остатков в одном или более положениях аминокислот (например, в 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15 или более положениях аминокислот) по сравнению с SEQ ID NO: 2, 6, 32, 34 и/или 38. В некоторых вариантах осуществления сконструированные лигазы имеют по меньшей мере 90% идентичность последовательности с SEQ ID NO: 2, 6, 32, 34 и/или 38 и имеют различие аминокислот в по меньшей мере 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более положениях аминокислот. В некоторых вариантах осуществления полипептид сконструированной лигазы состоит из последовательности SEQ ID NO: 6, 32, 34 и/или 38.

Полинуклеотиды, кодирующие Сконструированные Полипептиды, Векторы

экспрессии и Клетки-хозяева:

В настоящем изобретении представлены полинуклеотиды, кодирующие полипептиды сконструированной лигазы, описанные в данном документе. В некоторых вариантах осуществления полинуклеотиды функционально связаны с одной или более гетерологичными регуляторными последовательностями, которые регулируют экспрессию гена для создания рекомбинантного полинуклеотида, способного экспрессировать полипептид. В некоторых вариантах осуществления экспрессионные конструкции, содержащие по меньшей мере один гетерологичный полинуклеотид, кодирующий полипептид (полипептиды) сконструированной лигазы, вводят в подходящие клетки-хозяева для экспрессии соответствующего полипептида (полипептидов) лигазы.

Как будет очевидно специалисту в данной области, доступность белковой последовательности и знание кодонов, соответствующих различным аминокислотам, обеспечивают описание всех полинуклеотидов, способных кодировать рассматриваемые полипептиды. Вырожденность генетического кода, когда одни и те же аминокислоты кодируются альтернативными или синонимичными кодонами, позволяет получать чрезвычайно большое количество нуклеиновых кислот, которые все кодируют сконструированный полипептид лигазы. Таким образом, настоящее изобретение относится к способам и композициям для получения всех и каждого возможного варианта полинуклеотидов лигазы, которые могут быть получены, которые кодируют полипептиды лигазы, описанные в данном документе, путем выбора комбинаций на основе возможных вариантов кодонов, и все такие варианты должны рассматриваться специально раскрытыми для любого описанного в данном документе полипептида, включая аминокислотные последовательности, представленные в примерах (например, в различных таблицах).

В некоторых вариантах осуществления кодоны предпочтительно оптимизируют для использования путем выбора клетки-хозяина для получения белка. Например, предпочтительные кодоны, используемые в бактериях, обычно используют для экспрессии в бактериях. Следовательно, кодон-оптимизированные полинуклеотиды, кодирующие полипептиды сконструированной лигазы, содержат предпочтительные кодоны приблизительно в 40%, 50%, 60%, 70%, 80%, 90% или более чем в 90% положений кодонов в полноразмерной кодирующей области.

В некоторых вариантах осуществления полинуклеотид лигазы кодирует сконструированный полипептид, имеющий активность лигазы со свойствами, раскрытыми в данном документе, при этом полипептид содержит аминокислотную последовательность, имеющую идентичность по меньшей мере 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более с эталонной последовательностью, выбираемой из SEQ ID NO: 2, 6, 32, 34 и/или 38, или аминокислотную последовательность любого варианта (например, вариантов, предоставленных в примерах), и один или более разных остатков по сравнению с

эталонным полинуклеотидом SEQ ID NO: 2, 6, 32, 34 и/или 38, или аминокислотную последовательность любого варианта, который раскрыт в примерах (например, в 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более положениях аминокислотных остатков). В некоторых вариантах осуществления эталонную последовательность выбирают из SEQ ID NO: 2, 6, 32, 34 и/или 38. В некоторых вариантах осуществления варианты сконструированной лигазы содержат полипептидную последовательность, приведенную в SEQ ID NO: 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52 и/или 54. В некоторых вариантах осуществления варианты сконструированной лигазы содержат замену (замены) или набор (наборы) замен вариантных лигаз 1–261, которые предоставлены в примерах (например, Таблицы 4.1, 4.2, 4.3, 5.1, 5.2, 5.3, 5.4, 5.5 и 6.1).

В настоящем изобретении представлены полинуклеотиды, кодирующие варианты сконструированной лигазы, представленные в данном документе. В некоторых вариантах осуществления полинуклеотиды содержат нуклеотидную последовательность, имеющую идентичность по меньшей мере 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более с эталонной последовательностью, выбираемой из SEQ ID NO: 1, 5, 31, 33 и/или 37, или аминокислотную последовательность любого варианта (например, вариантов, предоставленных в примерах), и один или более разных остатков по сравнению с эталонным полинуклеотидом SEQ ID NO: 1, 5, 31, 37 и/или 38, или аминокислотную последовательность любого варианта, который раскрыт в примерах (например, в 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 или более положениях аминокислотных остатков). В некоторых вариантах осуществления эталонную последовательность выбирают из SEQ ID NO: 1, 5, 31, 33 и/или 37. В некоторых вариантах осуществления полинуклеотиды допускают гибридизацию в очень жестких условиях с эталонной полинуклеотидной последовательностью, выбираемой из SEQ ID NO: 1, 5, 31, 33 и/или 37, или ее комплементом, или с полинуклеотидной последовательностью, кодирующей любой из вариантных полипептидов лигазы, представленных в данном документе. В некоторых вариантах осуществления полинуклеотид, допускающий гибридизацию в очень жестких условиях, кодирует полипептид лигазы, содержащий аминокислотную последовательность, которая имеет одно или более различий остатков по сравнению с SEQ ID NO: 2, 6, 32, 34 и/или 38. В некоторых вариантах осуществления варианты сконструированной лигазы кодирует полинуклеотидная последовательность, приведенная в SEQ ID NO: 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51 и/или 53.

В некоторых вариантах осуществления изобретения с выделенным полинуклеотидом, кодирующим любой из сконструированных полипептидов лигазы в настоящем документе, манипулируют различными способами для облегчения экспрессии полипептида лигазы. В некоторых вариантах осуществления полинуклеотиды, кодирующие полипептиды лигазы, содержат векторы экспрессии, в которых присутствует одна или несколько регуляторных последовательностей, чтобы регулировать экспрессию полинуклеотидов и/или полипептидов лигазы. Манипуляции с выделенным полинуклеотидом перед его вставкой в вектор могут быть желательными или

необходимыми в зависимости от используемого вектора экспрессии. Методы модификации полинуклеотидов и последовательностей нуклеиновых кислот с использованием методов рекомбинантной ДНК хорошо известны в данной области. В некоторых вариантах осуществления регуляторные последовательности среди прочего содержат промоторы, лидерные последовательности, последовательности полиаденилирования, пропептидные последовательности, сигнальные пептидные последовательности и терминаторы транскрипции. В некоторых вариантах осуществления подходящие промоторы выбирают на основании выбора клеток-хозяев. Для бактериальных клеток-хозяев подходящие промоторы для направления транскрипции конструкций нуклеиновых кислот согласно настоящему изобретению включают, но не без ограничения, промоторы, полученные из *lac* оперона *E. coli*, гена агаказы *Streptomyces coelicolor* (*dagA*), гена левансахаразы *Bacillus subtilis* (*sacB*), гена альфа-амилазы *Bacillus licheniformis* (*amyL*), гена мальтогенной амилазы *Bacillus stearothermophilus* (*amyM*), гена альфа-амилазы *Bacillus amyloliquefaciens* (*amyQ*), гена пенициллиназы *Bacillus licheniformis* (*penP*), гена *xylA* и *xylB* *Bacillus subtilis* и гена прокариотической бета-лактамазы (см., например, Villa-Kamaroff et al., Proc. Natl Acad. Sci. USA 75: 3727–3731 [1978]), а также промотор *tac* (см., например, DeBoer et al., Proc. Natl. Acad. Sci. USA, 80: 21–25 [1983]). Иллюстративные промоторы для клеток-хозяев мицелиальных грибов включают, но без ограничения, промоторы, полученные из генов амилазы *ТАКА Aspergillus oryzae*, аспарагиновой протеиназы *Rhizomucor miehei*, нейтральной альфа-амилазы *Aspergillus niger*, кислотостабильной альфа-амилазы *Aspergillus niger*, глюкоамилазы *Aspergillus niger* или *Aspergillus awamori* (*glaA*), липазы *Rhizomucor miehei*, щелочной протеазы *Aspergillus oryzae*, триозофосфатизомеразы *Aspergillus oryzae*, ацетамидазы *Aspergillus nidulans* и трипсиноподобной протеазы *Fusarium oxysporum* (см., например, WO 96/00787), а также промотор *NA2-tri* (гибрид промоторов из генов нейтральной альфа-амилазы *Aspergillus niger* и триозофосфатизомеразы *Aspergillus oryzae*), и их мутантные, укороченный и гибридные промоторы. Иллюстративные промоторы дрожжевых клеток могут быть из генов енолазы *Saccharomyces cerevisiae* (*ENO-1*), галактокиназы *Saccharomyces cerevisiae* (*GAL1*), алкогольдегидрогеназы/глицеральдегид-3-фосфатдегидрогеназы *Saccharomyces cerevisiae* (*ADH2/GAP*) и 3-фосфоглицерат киназы *Saccharomyces cerevisiae*. В данной области известны другие полезные промоторы для дрожжевых клеток-хозяев (см., например, Romanos et al., Yeast 8: 423–488 [1992]).

В некоторых вариантах осуществления регуляторная последовательность также является подходящей терминаторной последовательностью транскрипции (то есть последовательностью, распознаваемой клеткой-хозяином для завершения транскрипции). В некоторых вариантах осуществления терминаторная последовательность функционально связана с 3'-концом последовательности нуклеиновой кислоты, кодирующей полипептид лигазы. В настоящем изобретении находит применение любой подходящий терминатор, который функционирует в выбранной клетке-хозяине.

Иллюстративные терминаторы транскрипции для клеток-хозяев мицелиальных грибов могут быть получены из генов для амилазы ТАКА *Aspergillus oryzae*, глюкоамилазы *Aspergillus niger*, антранилатсинтазы *Aspergillus nidulans*, альфа-глюкозидазы *Aspergillus niger* и трипсинподобной протеазы *Fusarium oxysporum*. Иллюстративные терминаторы для дрожжевых клеток-хозяев могут быть получены из генов енолазы *Saccharomyces cerevisiae*, цитохрома С (CYC1) *Saccharomyces cerevisiae* и глицеральдегид-3-фосфатдегидрогеназы *Saccharomyces cerevisiae*. В данной области известны другие полезные терминаторы для дрожжевых клеток-хозяев (см., например, Romanos et al., выше).

В некоторых вариантах осуществления регуляторная последовательность также является подходящей лидерной последовательностью (то есть нетранслируемой областью мРНК, которая важна для трансляции клеткой-хозяином). В некоторых вариантах осуществления лидерная последовательность функционально связана с 5'-концом последовательности нуклеиновой кислоты, кодирующей полипептид лигазы. Любая подходящая лидерная последовательность, которая является функциональной в выбранной клетке-хозяине, находит применение в настоящем изобретении. Иллюстративные лидеры для клеток-хозяев мицелиальных грибов получены из генов амилазы ТАКА *Aspergillus oryzae* и триозофосфатной изомеразы *Aspergillus nidulans*. Подходящие лидеры для дрожжевых клеток-хозяев получают из генов енолазы *Saccharomyces cerevisiae* (ENO-1), 3-фосфоглицераткиназы *Saccharomyces cerevisiae*, альфа-фактора *Saccharomyces cerevisiae* и алкогольдегидрогеназы/глицеральдегид-3-фосфатдегидрогеназы *Saccharomyces cerevisiae* (ADH2/GAP).

В некоторых вариантах осуществления регуляторная последовательность также представляет собой последовательность полиаденилирования (то есть последовательность, функционально связанную с 3'-концом последовательности нуклеиновой кислоты и которая при транскрибировании распознается клеткой-хозяином как сигнал для добавления остатков полиаденозина в транскрибируемую мРНК). Любая подходящая последовательность полиаденилирования, которая является функциональной в выбранной клетке-хозяине, находит применение в настоящем изобретении. Иллюстративные последовательности полиаденилирования для клеток-хозяев мицелиальных грибов включают, но без ограничения, гены для ТАКА-амилазы *Aspergillus oryzae*, глюкоамилазы *Aspergillus niger*, антранилатсинтазы *Aspergillus nidulans*, трипсиноподобной альфа-протеазы *Fusarium oxysporum* и альфа-глюкозидазы *Aspergillus niger*. Известны подходящие последовательности полиаденилирования для дрожжевых клеток-хозяев (см., например, Guo and Sherman, Mol. Cell. Bio., 15: 5983-5990 [1995]).

В некоторых вариантах осуществления регуляторная последовательность также представляет собой сигнальный пептид (то есть кодирующую область, которая кодирует аминокислотную последовательность, связанную с аминоконцом полипептида, и направляет закодированный полипептид в секреторный путь клетки). В некоторых вариантах осуществления 5'-конец кодирующей последовательности из

последовательности нуклеиновой кислоты по своей природе содержит кодирующую область сигнального пептида, естественно связанную в рамке считывания трансляции с сегментом кодирующей области, который кодирует секретируемый полипептид. Альтернативно, в некоторых вариантах осуществления 5'-конец кодирующей последовательности содержит кодирующую область сигнального пептида, которая является чужеродной для кодирующей последовательности. Любая подходящая кодирующая область сигнального пептида, которая направляет экспрессированный полипептид в секреторный путь выбранной клетки-хозяина, находит применение для экспрессии сконструированного полипептида (полипептидов). Эффективные области, кодирующие сигнальные пептиды для бактериальных клеток-хозяев, представляют собой области, кодирующие сигнальные пептиды, которые включают, но без ограничения, области, полученные из генов мальтогенной амилазы *Bacillus* NCIB 11837, альфа-амилазы *Bacillus stearothermophilus*, субтилизина *Bacillus licheniformis*, бета-лактамазы *Bacillus licheniformis*, нейтральной протеазы *Bacillus stearothermophilus* (nprT, nprS, nprM) и prsA *Bacillus subtilis*. Другие сигнальные пептиды известны в данной области (см., например, Simonen and Palva, *Microbiol. Rev.*, 57: 109–137 [1993]). В некоторых вариантах осуществления эффективные кодирующие области сигнального пептида для клеток-хозяев мицелиальных грибов включают, но без ограничения, кодирующие области сигнального пептида, полученные из генов амилазы TAKA *Aspergillus oryzae*, нейтральной амилазы *Aspergillus niger*, глюкоамилазы *Aspergillus niger*, аспарагиновой протеиназы *Rhizomucor miehei*, целлюлазы *Humicola insolens* и липазы *Humicola lanuginosa*. Полезные сигнальные пептиды для дрожжевых клеток-хозяев включают, но без ограничения, сигнальные пептиды из генов альфа-фактора *Saccharomyces cerevisiae* и инвертазы *Saccharomyces cerevisiae*.

В некоторых вариантах осуществления регуляторная последовательность также представляет собой область, кодирующую пропептид, которая кодирует аминокислотную последовательность, расположенную на аминоконце полипептида. Полученный полипептид называется «проэнзим», «прополипептид» или «зимоген». Пропептид может быть преобразован в зрелый активный полипептид путем каталитического или автокаталитического отщепления пропептида от прополипептида. Область, кодирующая пропептид, может быть получена из любого подходящего источника, включая, но без ограничения, гены щелочной протеазы *Bacillus subtilis* (aprE), нейтральной протеазы *Bacillus subtilis* (nprT), альфа-фактора *Saccharomyces cerevisiae*, аспарагиновой протеиназы *Rhizomucor miehei* и лактазы *Myceliophthora thermophila* (см., например, WO 95/33836). Когда на аминоконце полипептида присутствуют области как сигнального пептида, так и пропептида, область пропептида располагается рядом с аминоконцом полипептида, а область сигнального пептида располагается рядом с аминоконцом пропептидной области.

В некоторых вариантах осуществления также используют регуляторные последовательности. Эти последовательности облегчают регуляцию экспрессии

полипептида относительно роста клетки–хозяина. Примерами регуляторных систем являются те, которые вызывают включение или выключение экспрессии гена в ответ на химический или физический стимул, включая присутствие регуляторного соединения. В прокариотических клетках–хозяевах подходящие регуляторные последовательности включают, но без ограничения, операторные системы *lac*, *tac* и *trp*. В дрожжевых клетках–хозяевах подходящие регуляторные системы включают, но без ограничения, систему *ADH2* или систему *GAL1*. В мицелиальных грибах подходящие регуляторные последовательности включают, но без ограничения, промотор альфа–амилазы *TAKA*, промотор глюкоамилазы *Aspergillus niger* и промотор глюкоамилазы *Aspergillus oryzae*.

В другом аспекте настоящее изобретение направлено на рекомбинантный вектор экспрессии, содержащий полинуклеотид, кодирующий полипептид сконструированной лигазы, и одну или несколько областей, регулирующих экспрессию, таких как промотор и терминатор, источник репликации и т.д., в зависимости от типа хозяев, в которых они должны быть введены. В некоторых вариантах осуществления различные последовательности нуклеиновой кислоты и регуляторные последовательности, описанные в настоящем документе, объединяют, чтобы получить рекомбинантные векторы экспрессии, которые включают в себя один или несколько удобных сайтов рестрикции, чтобы обеспечить возможность вставки или замены последовательности нуклеиновой кислоты, кодирующей полипептид лигазы в таких сайтах. Альтернативно, в некоторых вариантах осуществления последовательность нуклеиновой кислоты согласно настоящему изобретению экспрессируют путем вставки последовательности нуклеиновой кислоты или конструкции нуклеиновой кислоты, содержащей последовательность, в подходящий вектор для экспрессии. В некоторых вариантах осуществления, включающих в себя создание вектора экспрессии, кодирующая последовательность расположена в векторе, так что кодирующая последовательность функционально связана с соответствующими регуляторными последовательностями для экспрессии.

Рекомбинантный вектор экспрессии может быть любым подходящим вектором (например, плазмидой или вирусом), который можно удобно подвергать процедурам рекомбинантной ДНК и вызывать экспрессию последовательности полинуклеотида лигазы. Выбор вектора обычно зависит от совместимости вектора с клеткой–хозяином, в которую этот вектор должен быть введен. Векторы могут быть линейными или замкнутыми кольцевыми плазмидами.

В некоторых вариантах осуществления вектор экспрессии представляет собой автономно реплицирующийся вектор (то есть вектор, который существует в виде внехромосомного объекта, репликация которого не зависит от репликации хромосомы, такой как плазида, внехромосомный элемент, минихромосома или искусственная хромосома). Вектор может содержать любые средства для обеспечения саморепликации. В некоторых альтернативных вариантах осуществления вектор представляет собой вектор, в котором при введении в клетку–хозяина он интегрируется в геном и реплицируется вместе с хромосомой (хромосомами), в которую он был интегрирован.

Кроме того, в некоторых вариантах осуществления используют один вектор или плазмиду или два или более векторов или плазмид, которые вместе содержат общую ДНК, которая должна быть введена в геном клетки–хозяина, и/или транспозон.

В некоторых вариантах осуществления вектор экспрессии содержит один или несколько селективируемых маркеров, которые позволяют легко выбирать трансформированные клетки. «Селективируемый маркер» представляет собой ген, продукт которого обеспечивает устойчивость к биоцидам или вирусам, устойчивость к тяжелым металлам, прототрофию к ауксотрофам и тому подобное. Примеры бактериальных селективируемых маркеров включают, но без ограничения, гены *dal* из *Bacillus subtilis* или *Bacillus licheniformis* или маркеры, которые придают устойчивость к антибиотикам, например, устойчивость к ампициллину, канамицину, хлорамфениколу или тетрациклину. Подходящие маркеры для дрожжевых клеток–хозяев включают, но без ограничения, ADE2, HIS3, LEU2, LYS2, MET3, TRP1 и URA3. Выбираемые маркеры для использования в клетках–хозяевах мицелиальных грибов включают, но без ограничения, *amdS* (ацетамидазу, например, из *A. nidulans* или *A. oryzae*), *argB* (орнитинкарбамоилтрансферазу), *bar* (фосфинотрицин–ацетилтрансферазу; например, из *S. hygrosopicus*), *hph* (гигромицинфосфотрансферазу), *niaD* (нитратредуктазу), *purG* (оротидин–5′–фосфатдекарбоксилазу; например, из *A. nidulans* или *A. oryzae*), *sC* (сульфат–аденилтрансферазу) и *trpC* (антранилатсинтазу), а также их эквиваленты. В другом аспекте настоящее изобретение относится к клетке–хозяину, содержащей по меньшей мере один полинуклеотид, кодирующий по меньшей мере один полипептид сконструированной лигазы согласно настоящему изобретению, причем полинуклеотид (полинуклеотиды) оперативно связан с одной или несколькими регуляторными последовательностями для экспрессии фермента (ферментов) сконструированной лигазы в клетке–хозяине. Клетки–хозяева, подходящие для использования в экспрессии полипептидов, кодируемых векторами экспрессии согласно настоящему изобретению, хорошо известны в данной области и включают, но без ограничения, бактериальные клетки, такие как клетки *E.coli*, *Vibrio fluvialis*, *Streptomyces* и *Salmonella typhimurium*; грибковые клетки, такие как дрожжевые клетки (например, *Saccharomyces cerevisiae* или *Pichia pastoris* (учетный номер ATCC 201178)); клетки насекомых, такие как клетки *Drosophila S2* и *Spodoptera Sf9*; клетки животных, такие как клетки меланомы CHO, COS, BHK, 293 и Bowes; и клетки растений. Иллюстративные клетки–хозяева также включают различные штаммы *Escherichia coli* (например, W3110 (Δ h_uA) и BL21).

Соответственно, в другом аспекте настоящее изобретение относится к способам получения полипептидов сконструированной лигазы, причем способы включают культивирование клетки–хозяина, способной экспрессировать полинуклеотид, кодирующий полипептид сконструированной лигазы, в условиях, подходящих для экспрессии полипептида. В некоторых вариантах осуществления способы дополнительно включают стадии выделения и/или очистки полипептидов лигаз, как описано в данном документе.

Подходящие культуральные среды и условия роста для клеток-хозяев хорошо известны в данной области. Предполагается, что любой подходящий способ введения полинуклеотидов для экспрессии полипептидов лигазы в клетки найдет применение в настоящем изобретении. Подходящие методы включают, но без ограничения, электропорацию, бомбардировку биолистическими частицами, опосредованную липосомами трансфекцию, трансфекцию хлоридом кальция и слияние протопластов.

Полипептиды сконструированной лигазы со свойствами, раскрытыми в данном описании, могут быть получены путем воздействия на полинуклеотид, кодирующий встречающийся в природе или сконструированный полипептид лигазы, любым подходящим методом мутагенеза и/или направленной эволюции, известным в данной области и/или описанным в данном документе. Иллюстративным методом направленной эволюции является мутагенез и/или перетасовка ДНК (см., например, Stemmer, *Proc. Natl. Acad. Sci. USA* 91: 10747–10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966); WO 98/27230; WO 00/42651; WO 01/75767 и патент США 6537746). Другие процедуры направленной эволюции, которые могут быть использованы, включают, среди прочего, ступенчатый процесс удлинения (StEP), рекомбинацию *in vitro* (см., например, Zhao et al., *Nat. Biotechnol.*, 16: 258–261 [1998]), мутагенную ПЦР (см., например, Caldwell et al., *PCR Methods Appl.*, 3: S136–S140 [1994]) и кассетный мутагенез (см., например, Black et al., *Proc. Natl. Acad. Sci. USA* 93: 3525–3529 [1996]).

Методы мутагенеза и направленной эволюции могут быть легко применены к полинуклеотидам, кодирующим лигазу, для создания библиотек вариантов, которые можно экспрессировать, подвергать скринингу и анализировать. Любые подходящие методы мутагенеза и направленной эволюции находят применение в настоящем изобретении и хорошо известны в данной области (см., например, патенты США №№ 5605793, 5811238, 5830721, 5834252, 5837458, 5928905, 6096548, 6117679, 6132970, 6165793, 6180406, 6251674, 6265201, 6277638, 6287861, 6287862, 6291242, 6297053, 6303344, 6309883, 6319713, 6319714, 6323030, 6326204, 6335160, 6335198, 6344356, 6352859, 6355484, 6358740, 6358742, 6365377, 6365408, 6368861, 6372497, 6337186, 6376246, 6379964, 6387702, 6391552, 6391640, 6395547, 6406855, 6406910, 6413745, 6413774, 6420175, 6423542, 6426224, 6436675, 6444468, 6455253, 6479652, 6482647, 6483011, 6484105, 6489146, 6500617, 6500639, 6506602, 6506603, 6518065, 6519065, 6521453, 6528311, 6537746, 6573098, 6576467, 6579678, 6586182, 6602986, 6605430, 6613514, 6653072, 6686515, 6703240, 6716631, 6825001, 6902922, 6917882, 6946296, 6961664, 6995017, 7024312, 7058515, 7105297, 7148054, 7220566, 7288375, 7384387, 7421347, 7430477, 7462469, 7534564, 7620500, 7620502, 7629170, 7702464, 7747391, 7747393, 7751986, 7776598, 7783428, 7795030, 7853410, 7868138, 7783428, 7873477, 7873499, 7904249, 7957912, 7981614, 8014961, 8029988, 8048674, 8058001, 8076138, 8108150, 8170806, 8224580, 8377681, 8383346, 8457903, 8504498, 8589085, 8762066, 8768871, 9593326 и все родственные PCT и неамериканские копии; Ling et al., *Anal. Biochem.*, 254(2):157–78 [1997]; Dale et al., *Meth. Mol. biol.*, 57:369–74 [1996]; Smith, *Ann.*

Rev. Genet., 19:423–462 [1985]; Botstein et al., Science, 229:1193–1201 [1985]; Carter, Biochem. J., 237:1–7 [1986]; Kramer et al., Cell, 38:879–887 [1984]; Wells et al., Gene, 34:315–323 [1985]; Minshull et al., Curr. Op. Chem. biol., 3:284–290 [1999]; Christians et al., Nat. Biotechnol., 17:259–264 [1999]; Cramer et al., Nature, 391:288–291 [1998]; Cramer et al., Nat. Biotechnol., 15:436–438 [1997]; Zhang et al., Proc. Nat. Acad. Sci. U.S.A., 94:4504–4509 [1997]; Cramer et al., Nat. Biotechnol., 14:315–319 [1996]; Stemmer, Nature, 370:389–391 [1994]; Stemmer, Proc. Nat. Acad. Sci. USA, 91:10747–10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; и WO 2009/152336, которые все включены в данный документ посредством ссылки).

В некоторых вариантах осуществления ферментные клоны, полученные после обработки мутагенезом, подвергают скринингу, подвергая ферментные препараты воздействию определенной температуры (или другим условиям анализа) и измеряя количество активности фермента, остающееся после термической обработки или других подходящих условий анализа. Клоны, содержащие полинуклеотид, кодирующий полипептид лигазы, затем выделяют из гена, секвенируют для идентификации изменений нуклеотидной последовательности (если таковые имеются) и используют для экспрессии фермента в клетке-хозяине. Измерение активности фермента из библиотек экспрессии может быть выполнено с использованием любого подходящего способа, известного в данной области (например, стандартных биохимических методов, таких как анализ HPLC).

Для сконструированных полипептидов известной последовательности полинуклеотиды, кодирующие фермент, могут быть получены стандартными твердофазными методами в соответствии с известными синтетическими методами. В некоторых вариантах осуществления фрагменты до примерно 100 оснований могут быть индивидуально синтезированы, а затем объединены (например, с помощью методов ферментативного или химического лигирования или методов, опосредованных полимеразой) с образованием любой нужной непрерывной последовательности. Например, полинуклеотиды и олигонуклеотиды, раскрытые в данном документе, могут быть получены с помощью химического синтеза с использованием классического фосфорамидитного метода (см., например, Beaucage et al., Tet. Lett., 22: 1859–69 [1981]; и Matthes et al., EMBO J., 3:801–05 [1984]), как это обычно практикуется в автоматизированных синтетических методах. Согласно фосфорамидитному методу олигонуклеотиды синтезируют (например, в автоматическом синтезаторе ДНК, очищают, отжигают, лигируют и клонируют в соответствующих векторах).

Соответственно, в некоторых вариантах осуществления способ получения полипептида сконструированной лигазы может включать: (а) синтез полинуклеотида, кодирующего полипептид, содержащий аминокислотную последовательность, выбираемую из аминокислотной последовательности любого варианта, который описан в данном документе, и (b) экспрессию полипептида лигазы, кодируемого полинуклеотидом. В некоторых вариантах осуществления способа аминокислотная последовательность,

кодируемая полинуклеотидом, необязательно может иметь один или несколько (например, до 3, 4, 5 или до 10) делеций, вставок и/или замен аминокислотных остатков. В некоторых вариантах осуществления аминокислотная последовательность необязательно имеет 1–2, 1–3, 1–4, 1–5, 1–6, 1–7, 1–8, 1–9, 1–10, 1–15, 1–20, 1–21, 1–22, 1–23, 1–24, 1–25, 1–30, 1–35, 1–40, 1–45 или 1–50 делеций, вставок и/или замен аминокислотных остатков. В некоторых вариантах осуществления аминокислотная последовательность необязательно имеет 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45 или 50 делеций, вставок и/или замен аминокислотных остатков. В некоторых вариантах осуществления аминокислотная последовательность необязательно имеет 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24 или 25 делеций, вставок и/или замен аминокислотных остатков. В некоторых вариантах осуществления замены представляют собой консервативные или неконсервативные замены.

Экспрессированный полипептид сконструированной лигазы можно оценить на любое нужное улучшенное свойство или комбинацию свойств (например, активность, избирательность, стабильность, кислотостойкость, чувствительность к протеазе и т.д.) с использованием любого подходящего анализа, известного в данной области, включая, но без ограничения, анализы и условия, описанные в данном документе.

В некоторых вариантах осуществления любой из полипептидов сконструированной лигазы, экспрессируемых в клетке–хозяине, выделяют из клеток и/или культуральной среды с использованием любого одной или нескольких хорошо известных методик очистки белков, включая среди прочего обработку лизоцимом, обработку ультразвуком, фильтрацию, высаливание, ультрацентрифугирование и хроматографию.

Хроматографические методы выделения полипептидов лигазы включают, среди прочего, хроматографию с обращенной фазой, высокоэффективную жидкостную хроматографию, ионообменную хроматографию, хроматографию с гидрофобным взаимодействием, эксклюзионную хроматографию, гель–электрофорез и аффинную хроматографию. Условия для очистки конкретного фермента частично зависят от таких факторов, как суммарный заряд, гидрофобность, гидрофильность, молекулярная масса, молекулярная форма и т.д., и будут очевидны для специалистов в данной области. В некоторых вариантах осуществления аффинные методы могут быть использованы для выделения улучшенных ферментов лигазы. Для очистки аффинной хроматографией может найти применение любое антитело, которое специфически связывает интересующий полипептид лигазы. Для получения антител различных животных–хозяев, включая, но без ограничения, кроликов, мышей, крыс и т.д., иммунизируют с помощью инъекции полипептида лигазы или его фрагмента. В некоторых вариантах осуществления полипептид или фрагмент лигазы присоединяют к подходящему носителю, такому как BSA, посредством функциональной группы боковой цепи или линкеров, связанных с функциональной группой боковой цепи.

В некоторых вариантах осуществления сконструированный полипептид лигазы получают в клетке–хозяине способом, включающим культивирование клетки–хозяина

(например, штамма *E.coli*), содержащей полинуклеотидную последовательность, кодирующую полипептид сконструированной лигазы, который описан в данном документе, в условиях, способствующих получению полипептида сконструированной лигазы, и выделение полипептида сконструированной лигазы из клеток и/или культуральной среды. В некоторых вариантах осуществления клетка-хозяин продуцирует более одного полипептида сконструированной лигазы.

В некоторых вариантах осуществления в настоящем изобретении представлен способ получения полипептида сконструированной лигазы, включающий культивирование рекомбинантной бактериальной клетки, содержащей полинуклеотидную последовательность, кодирующую полипептид сконструированной лигазы, имеющий по меньшей мере 85%, 90%, 95%, 96%, 97%, 98% или 99% идентичность последовательности с эталонными последовательностями SEQ ID NO: 2, 6, 32, 34 и/или 38, и одно или более различий аминокислотных остатков, в подходящих условиях культивирования, чтобы обеспечить получение полипептида сконструированной лигазы и необязательно выделение полипептида сконструированной лигазы из культуры и/или культивируемых бактериальных клеток. В некоторых вариантах осуществления клетка-хозяин продуцирует более чем один полипептид сконструированной лигазы.

В некоторых вариантах осуществления после извлечения полипептидов сконструированной лигазы из рекомбинантных клеток-хозяев и/или культуральной среды, их дополнительно очищали с помощью любого подходящего способа (способов), известных в данной области. В некоторых дополнительных вариантах осуществления очищенные полипептиды сконструированной лигазы объединяют с другими ингредиентами и соединениями для предоставления композиций и готовых форм, содержащих полипептид сконструированной лигазы, подходящий для разных вариантов применения и вариантов использования (например, диагностических способов и композиций).

Экспериментальная часть

Следующие примеры, включая эксперименты и достигнутые результаты, приведены только для иллюстративных целей и не должны рассматриваться как ограничивающие настоящее изобретение.

В приведенном ниже экспериментальном описании применяются следующие сокращения: ч/млн (частей на миллион); М (молярный); mM (миллимолярный), uM и μ M (микромолярный); nM (наномолярный); mol (моль); gm и g (грамм); мг (миллиграмм); ug и μ g (микрограмм); L и l (литр); ml и mL (миллилитр); cm (сантиметр); mM (миллиметр); μ m и μ m (микрометр); sec. (секунда); min(s) (минута (минуты)); h(s) и hr(s) (час (часы)); U (единицы); MW (молекулярная масса); об/мин (оборотов в минуту); gcf (относительное центробежное ускорение); psi и PSI (фунтов на квадратный дюйм); °C (градусы по Цельсию); RT и rt (комнатная температура); NGS (секвенирования следующего поколения); ds (двухцепочечный); ss (одноцепочечный); CDS (кодирующая последовательность); ДНК (дезоксирибонуклеиновая кислота); РНК (рибонуклеиновая

кислота); E. coli W3110 (широко используемый лабораторный штамм E. coli, поставляемый Coli Genetic Stock Center [CGSC], New Haven, CT); HTP (высокопроизводительный); HPLC (жидкостная хроматография высокого давления); CFSE (карбоксифлуоресцеин сукцинимидиловый эфир); IPTG (изопропил β -D-1-тиогалактопиранозид); PES (полиэфирсульфон); FIOPC (кратное улучшение по сравнению с положительным контролем); EB буфер (10 mM Tris); LB (Лурия–Бертани); SPRI (твердофазная обратимая иммобилизация); Sigma–Aldrich (Sigma–Aldrich, St. Louis, MO); Promega (Promega, Inc., Madison, WI); EMD Millipore (EMD Millipore или Millipore Sigma, part of Merck KGaA, Darmstadt, Germany); Perkin Elmer (Perkin Elmer, Inc, Waltham, MA); Millipore (Millipore, Corp., Billerica MA); MagBio Genomics (MagBio Genomics, Gaithersburg, MD); BioChain (BioChain Institute, Inc., Newark, CA); Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Molecular Devices (Molecular Devices, LLC, Sunnyvale, CA); Kuhner (Adolf Kuhner, AG, Basel, Switzerland); Microfluidics (Microfluidics, Corp., Westwood, MA); NEB (New England Biolabs, Inc., Ipswich, MA); Applied Biosystems (Applied Biosystems, part of Life Technologies, Corp., Grand Island, NY); Zymo (Zymo Research, Irvine, CA); Agilent (Agilent Technologies, Inc., Santa Clara, CA); Integrated DNA Technologies (Integrated DNA Technologies, Inc., Coralville, IA); Illumina (Illumina, Inc., San Diego, CA); Thermo Scientific (part of Thermo Fisher Scientific, Waltham, MA); GE Healthcare (GE Healthcare Bio–Sciences, Piscataway, NJ); Enzymatics (Enzymatics, Inc., Beverly, MA); и Bio–Rad (Bio–Rad Laboratories, Hercules, CA).

В настоящем изобретении находят применение следующие полинуклеотидные и полипептидные последовательности. В некоторых случаях (как показано ниже) за полинуклеотидной последовательностью следует кодированный полипептид. SEQ ID NO: 1 и 2 соответствуют ДНК–лигазы фага дикого типа T4 (без метки). SEQ ID NO: 3 и 4 соответствуют ДНК–лигазы фага дикого типа T4, которая была мечена гистидином. SEQ ID NO: 5 и 6 соответствуют оптимизированной синтетической T4 ДНК–лигазе (немеченной).

ATGATTCTTA	AAATTCTGAA	CGAAATAGCA	TCTATTGGTT	CAACTAAACA
GAAGCAAGCA	ATTCTTGAAA	AGAATAAAGA	TAATGAATTG	CTTAAACGAG
TATATCGTCT	GACTTATTCT	CGTGGGTTAC	AGTATTATAT	CAAGAAATGG
CCTAAACCTG	GTATTGCTAC	CCAGAGTTTT	GGAATGTTGA	CTCTTACCGA
TATGCTTGAC	TTCATTGAAT	TCACATTAGC	TACTCGGAAA	TTGACTGGAA
ATGCAGCAAT	TGAGGAATTA	ACTGGATATA	TCACCGATGG	TAAAAAAGAT
GATGTTGAAG	TTTTGCGTCG	AGTGATGATG	CGAGACCTTG	AATGTGGTGC
TTCAGTATCT	ATTGCAAACA	AAGTTTGGCC	AGGTTTAATT	CCTGAACAAC
CTCAAATGCT	CGCAAGTTCT	TATGATGAAA	AAGGCATTAA	TAAGAATATC
AAATTTCCAG	CCTTTGCTCA	GTAAAAGCT	GATGGAGCTC	GGTGTTTTGC
TGAAGTTAGA	GGTGATGAAT	TAGATGATGT	TCGTCTTTTA	TCACGAGCTG
GTAATGAATA	TCTAGGATTA	GATCTTCTTA	AGGAAGAGTT	AATTAAAATG
ACCGCTGAAG	CCCGCCAGAT	TCATCCAGAA	GGTGTGTTGA	TTGATGGCGA

ATTGGTATAC CATGAGCAAG TTAAAAAGGA GCCAGAAGGC CTAGATTTTC
 TTTTGTATGC TTATCCTGAA AACAGTAAAG CTAAAGAATT CGCCGAAGTA
 GCTGAATCAC GTACTGCTTC TAATGGAATC GCCAATAAAT CTTTAAAGGG
 AACCATTCTT GAAAAAGAAG CACAATGCAT GAAGTTTCAG GTCTGGGATT
 ATGTCCCGTT GGTAGAAATA TACAGTCTTC CTGCATTTTCG TTTGAAATAT
 GATGTACGTT TTTCTAAACT AGAACAAATG ACATCTGGAT ATGATAAAGT
 AATTTTAATT GAAAACCAGG TAGTAAATAA CCTAGATGAA GCTAAGGTAA
 TTTATAAAAA GTATATTGAC CAAGGTCTTG AAGGTATTAT TCTCAAAAAT
 ATCGATGGAT TATGGGAAAA TGCTCGTTCA AAAAATCTTT ATAAATTTAA
 AGAAGTAATT GATGTTGATT TAAAAATTGT AGGAATTTAT CCTCACCGTA
 AAGACCCTAC TAAAGCGGGT GGATTTATTC TTGAGTCAGA GTGTGGAAAA
 ATTAAGGTAA ATGCTGGTTC AGGCTTAAAA GATAAAGCCG GTGTAAAATC
 GCATGAACCTT GACCGTACTC GCATTATGGA AAACCAAAAT TATTATATTG
 GAAAAATTCT AGAGTGCGAA TGCAACGGTT GGTAAAAATC TGATGGCCGC
 ACTGATTACG TTAAATTATT TCTTCCGATT GCGATTCGTT TACGTGAAGA
 TAAAACATAA GCTAATACAT TCGAAGATGT ATTTGGTGAT TTTCATGAGG
 TAACTGGTCT ATAA (SEQ ID NO: 1)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Gln Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Leu Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Glu Glu
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Pro Gly Leu Ile Pro Glu Gln Pro
 Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Lys Lys Glu Pro Glu Gly Leu Asp Phe Leu Phe Asp Ala Tyr Pro Glu Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
 Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Asp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
 Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Glu Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Thr Asp Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 2)

ATGCATCACC ATCACCATCA CGGTGGCAGC GGTATGATTC TTAAAATTCT

GAACGAAATA	GCATCTATTG	GTTCAACTAA	ACAGAAGCAA	GCAATTCTTG
AAAAGAATAA	AGATAATGAA	TTGCTTAAAC	GAGTATATCG	TCTGACTTAT
TCTCGTGGGT	TACAGTATTA	TATCAAGAAA	TGGCCTAAAC	CTGGTATTGC
TACCCAGAGT	TTTGGGAATGT	TGACTCTTAC	CGATATGCTT	GACTTCATTG
AATTCACATT	AGCTACTCGG	AAATTGACTG	GAAATGCAGC	AATTGAGGAA
TTAACTGGAT	ATATCACCGA	TGGTAAAAAA	GATGATGTTG	AAGTTTTGCG
TCGAGTGATG	ATGCGAGACC	TTGAATGTGG	TGCTTCAGTA	TCTATTGCAA
ACAAAGTTTG	GCCAGGTTTA	ATTCCTGAAC	AACCTCAAAT	GCTCGCAAGT
TCTTATGATG	AAAAAGGCAT	TAATAAGAAT	ATCAAATTTT	CAGCCTTTGC
TCAGTTAAAA	GCTGATGGAG	CTCGGTGTTT	TGCTGAAGTT	AGAGGTGATG
AATTAGATGA	TGTTTCGTCTT	TTATCACGAG	CTGGTAATGA	ATATCTAGGA
TTAGATCTTC	TTAAGGAAGA	GTTAATTAAA	ATGACCGCTG	AAGCCCGCCA
GATTCATCCA	GAAGGTGTGT	TGATTGATGG	CGAATTGGTA	TACCATGAGC
AAGTTAAAAA	GGAGCCAGAA	GGCCTAGATT	TTCTTTTTGA	TGCTTATCCT
GAAAACAGTA	AAGCTAAAGA	ATTCGCCGAA	GTAGCTGAAT	CACGTACTGC
TTCTAATGGA	ATCGCCAATA	AATCTTTAAA	GGGAACCATT	TCTGAAAAAG
AAGCACAATG	CATGAAGTTT	CAGGTCTGGG	ATTATGTCCC	GTTGGTAGAA
ATATACAGTC	TTCCTGCATT	TCGTTTGAAA	TATGATGTAC	GTTTTTCTAA
ACTAGAACAA	ATGACATCTG	GATATGATAA	AGTAATTTTA	ATTGAAAACC
AGGTAGTAAA	TAACCTAGAT	GAAGCTAAGG	TAATTTATAA	AAAGTATATT
GACCAAGGTC	TTGAAGGTAT	TATTCTCAAA	AATATCGATG	GATTATGGGA
AAATGCTCGT	TCAAAAAATC	TTTATAAATT	TAAAGAAGTA	ATTGATGTTG
ATTTAAAAAT	TGTAGGAATT	TATCCTCACC	GTAAAGACCC	TACTAAAGCG
GGTGGATTTA	TTCTTGAGTC	AGAGTGTGGA	AAAATTAAGG	TAAATGCTGG
TTCAGGCTTA	AAAGATAAAG	CCGGTGTAAG	ATCGCATGAA	CTTGACCGTA
CTCGCATTAT	GGAAAACCAA	AATTATTATA	TTGGAAAAAT	TCTAGAGTGC
GAATGCAACG	GTTGGTTAAA	ATCTGATGGC	CGCACTGATT	ACGTAAATT
ATTTCTCCG	ATTGCGATTC	GTTTACGTGA	AGATAAAACT	AAAGCTAATA
CATTCGAAGA	TGTATTTGGT	GATTTTCATG	AGGTAAGTGG	TCTATAA (SEQ ID NO:

3)

Met His His His His His His Gly Gly Ser Gly Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Gln Ala Ile Leu Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Leu Thr Leu Thr Asp Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Glu Glu Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Pro Gly Leu Ile Pro Glu Gln Pro Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr Ala Glu Ala Arg Gln Ile His Pro

Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val Lys Lys Glu Pro Glu Gly Leu
 Asp Phe Leu Phe Asp Ala Tyr Pro Glu Asn Ser Lys Ala Lys Glu Phe Ala Glu Val Ala Glu Ser
 Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile Ser Glu Lys Glu Ala Gln Cys
 Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr Ser Leu Pro Ala Phe Arg Leu
 Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly Tyr Asp Lys Val Ile Leu Ile
 Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys Lys Tyr Ile Asp Gln Gly
 Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg Ser Lys Asn Leu Tyr
 Lys Phe Lys Glu Val Ile Asp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His Arg Lys Asp Pro Thr
 Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn Ala Gly Ser Gly Leu Lys
 Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met Glu Asn Gln Asn Tyr Tyr
 Ile Gly Lys Ile Leu Glu Cys Glu Cys Asn Gly Trp Leu Lys Ser Asp Gly Arg Thr Asp Tyr Val
 Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys Ala Asn Thr Phe Glu Asp
 Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 4)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA				
AAAACAAGCC	ATTCTGGAAA	AAAATAAAGA	CAATGAACTG	TTAAAGCGCG
TGTATCGCCT	GACCTATTCG	CGTGGCCTGC	AATACTATAT	TAAAAAATGG
CCCAAACCGG	GCATTGCGAC	GCAGAGCTTC	GGCATGCTGA	CCCTGACCGA
TATGCTGGAT	TTTATCGAGT	TACTCTTGC	GACGCGCAAA	CTTACCGGCA
ACGCTGCTAT	TGAAGAATTG	ACGGGCTATA	TTACCGATGG	CAAAAAAGAT
GATGTTGAGG	TGCTGCGTCG	CGTCATGATG	CGTGATCTGG	AGTGCGGTGC
GTCAGTGAGC	ATCGCTAACA	AAGTGTGGCC	GGGTTTGATC	CCGGAACAGC
CACAGATGCT	TGCAAGCAGC	TACGATGAAA	AGGGCATTAA	CAAAAATATC
AAATTTCCGG	CTTTCGCCCA	GCTGAAAGCG	GATGGCGCGC	GCTGCTTCGC
CGAGGTACGC	GGTGACGAAC	TGGACGACGT	TCGTCTGCTG	TCTCGCGCTG
GCAATGAATA	TCTGGGTCTG	GATCTGCTGA	AAGAAGAACT	GATTAAGATG
ACCGCCGAGG	CGCGCCAAAT	CCACCCCGAA	GGGGTGCTGA	TTGATGGCGA
ACTGGTGTAT	CACGAACAGG	TTAAAAAGGA	GCCGGAGGGG	TTGGATTTC
TGTTTGATGC	CTACCCGGAG	AATAGCAAAG	CGAAAGAATT	TGCGGAAGTG
GCGGAATCCC	GCACCGCAAG	CAATGGTATC	GCAAATAAAT	CGCTGAAGGG
CACCATTCT	GAGAAAGAAG	CACAGTGTAT	GAAGTTCCAG	GTGTGGGACT
ACGTGCCACT	GGTTGAGATC	TACTCCCTGC	CAGCGTTTCG	CCTGAAATAC
GATGTGCGTT	TTTCAAAACT	GGAACAGATG	ACGAGCGGAT	ATGATAAAGT
AATTTTAATC	GAGAACCAGG	TCGTGAACAA	CCTGGATGAA	GCAAAAAGTTA
TCTATAAAAA	ATATATTGAT	CAGGGCTTAG	AAGGCATTAT	CCTGAAGAAC
ATTGATGGCC	TTTGGGAAAA	TGCACGCAGC	AAAAACCTGT	ATAAATTTAA
AGAAGTGATT	GATGTAGATC	TGAAGATTGT	TGGTATTTAC	CCGCATCGCA
AAGACCCGAC	CAAAGCAGGT	GGTTTCATCC	TGGAATCTGA	ATGCGGTAAA
ATTAAAGTGA	ACGCAGGCAG	TGGTTTGAAA	GATAAAGCTG	GTGTTAAAAG
TCACGAAC TG	GATCGCACGC	GCATCATGGA	AAACCAGAAC	TATTATATCG
GCAAAATTCT	GGAATGCGAG	TGCAACGGGT	GGCTTAAGAG	CGACGGGCGC

ACGGACTATG TTAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGAAGA
 CAAAATAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTCATGAAG
 TCACGGGTCT GTAA (SEQ ID NO: 5)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Gln Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Leu Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Glu Glu
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Pro Gly Leu Ile Pro Glu Gln Pro
 Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Lys Lys Glu Pro Glu Gly Leu Asp Phe Leu Phe Asp Ala Tyr Pro Glu Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
 Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Asp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
 Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Glu Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Thr Asp Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 6)

Следующими последовательностями являются последовательности нуклеиновых
 кислот олигонуклеотидов, используемые в следующих примерах.

/5Phos/TGCTACTCATCCTAGTCCTGTTGCT/iCy3/GCCAAGCTATTTAATATCAT
 GCACA (SEQ ID NO: 7)

/5Phos/GTGCATGATATTAAATAGCTTGGCAGCAACAGGACTAGGATGAGTAG
 CAA (SEQ ID NO: 8)

ACACGACGCTCTTCCGATC*T (SEQ ID NO: 9)

/5Phos/GATCGGAAGAGCGTCGTGT/3BioTEG/ (SEQ ID NO: 10)

/5Phos/TGCTACTCATCCTAGTCCTGTTGCTGCCAAGCTATTTAATATCATGCAC
 A (SEQ ID NO: 11)

/5Phos/GTGCATGATATTAAATAGCTTGGCAGCAACAGGACTAGGATGAGTAG
 CAA (SEQ ID NO: 12)

ACACGACGCTCTTCCGATC*T (SEQ ID NO: 13)

/5Phos/GATCGGAAGAGCGTCGTGT (SEQ ID NO: 14)

/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTC/ideoxyU/ACACTCTTTCC

CTACACGACGCTCTTCCGATC*T (SEQ ID NO: 15)

/5Phos/GATCGGAAGAGCGTCGTGT/3A mMO/ (SEQ ID NO: 16)

AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTCTTCC
GATC*T (SEQ ID NO: 17)

/5Phos/GATCGGAAGAGCACACGTCTGAACTCCAGTCACATCACGATCTCGTAT
GCCGTCTTCTGCTTG (SEQ ID NO: 18)

AAGCCGGTGTA AAAATCGCATGA (SEQ ID NO: 19)

AGTAAACGAATCGCAATCGGAAGA (SEQ ID NO: 20)

CAGCCGGTGTA AAAATCGCATGA (SEQ ID NO: 21)

CGTAAACGAATCGCAATCGGAAGA (SEQ ID NO: 22)

GAGCCGGTGTA AAAATCGCATGA (SEQ ID NO: 23)

GGTAAACGAATCGCAATCGGAAGA (SEQ ID NO: 24)

TAGCCGGTGTA AAAATCGCATGA (SEQ ID NO: 25)

TGTAAACGAATCGCAATCGGAAGA (SEQ ID NO: 26)

AAGCCGGTGTA AAAATCGCATGAACTTGACCGTACTCGCATTATGGAAAACCA
AAATTATTATATTGGAAAAATTCTAGAGTGCGAATGCAACGGTTGGTTAAAATCTGA
TGGCCGCACTGATTACGTAAATTATTTCTTCCGATTGCGATTCGTTTACA (SEQ ID
NO: 27)

CAGCCGGTGTA AAAATCGCATGAACTTGACCGTACTCGCATTATGGAAAACCA
AAATTATTATATTGGAAAAATTCTAGAGTGCGAATGCAACGGTTGGTTAAAATCTGA
TGGCCGCACTGATTACGTAAATTATTTCTTCCGATTGCGATTCGTTTACC (SEQ ID
NO: 28)

GAGCCGGTGTA AAAATCGCATGAACTTGACCGTACTCGCATTATGGAAAACCA
AAATTATTATATTGGAAAAATTCTAGAGTGCGAATGCAACGGTTGGTTAAAATCTGA
TGGCCGCACTGATTACGTAAATTATTTCTTCCGATTGCGATTCGTTTACG (SEQ ID
NO: 29)

TAGCCGGTGTA AAAATCGCATGAACTTGACCGTACTCGCATTATGGAAAACCA
AAATTATTATATTGGAAAAATTCTAGAGTGCGAATGCAACGGTTGGTTAAAATCTGA
TGGCCGCACTGATTACGTAAATTATTTCTTCCGATTGCGATTCGTTTACT (SEQ ID
NO: 30)

Следующие последовательности соответствуют вариантам сконструированной
лигазы, предоставленным в настоящем изобретении.

ATGATTCTTA	AAATTCTGAA	CGAAATTGCA	AGCATTGGTT	CCACTAAACA
AAAACAAGCC	ATTCTGGAAA	AAAATAAAGA	CAATGAACTG	TTAAAGCGCG
TGTATCGCCT	GACCTATTCG	CGTGGCCTGC	AATACTATAT	TAAAAAATGG
CCCAAACCGG	GCATTGCGAC	GCAGAGCTTC	GGCATGCTGA	CCCTGACCGA
TATGCTGGAT	TTTATCGAGT	TTACTCTTGC	GACGCGCAA	CTTACCGGCA
ACGCTGCTAT	TAGAGAATTG	ACGGGCTATA	TTACCGATGG	CAAAAAAGAT
GATGTTGAGG	TGCTGCGTCG	CGTCATGATG	CGTGATCTGG	AGTGCGGTGC
GTCAGTGAGC	ATCGCTAACA	AAGTGTGGCC	GGGTTTGATC	CCGGAACAGC

CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTAA CAAAAATATC
AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCGCTG
GCAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG
ACCGCCGAGG CGCGCCAAAT CCACCCCGAA GGGGTGCTGA TTGATGGCGA
ACTGGTGTAT CACGAACAGG TTGCAAAAGA GCCGGAGGGG TTGGATTTC
TGTTTGATGC CTACCCGGAG AATAGCAAAG CGAAAGAATT TGCGGAAGTG
GCGGAATCCC GCACCGCAAG CAATGGTATC GCAAATAAAT CGCTGAAGGG
CACCATTCT GAGAAAGAAG CACAGTGTAT GAAGTTCCAG GTGTGGGACT
ACGTGCCACT GGTTGAGATC TACTCCCTGC CAGCGTTTCG CCTGAAATAC
GATGTGCGTT TTTCAAACT GGAACAGATG ACGAGCGGAT ATGATAAAGT
AATTTTAATC GAGAACCAGG TCGTGAACAA CCTGGATGAA GCAAAAGTTA
TCTATAAAAA ATATATTGAT CAGGGCTTAG AAGGCATTAT CCTGAAGAAC
ATTGATGGCC TTTGGGAAAA TGCACGCAGC AAAAACCTGT ATAAATTTAA
AGAAGTGATT GATGTAGATC TGAAGATTGT TGGTATTTAC CCGCATCGCA
AAGACCCGAC CAAAGCAGGT GGTTCATCC TGGAATCTGA ATGCGGTAAA
ATTAAAGTGA ACGCAGGCAG TGGTTTGAAA GATAAAGCTG GTGTTAAAG
TCACGAACTG GATCGCACGC GCATCATGGA AAACCAGAAC TATTATATCG
GCAAAATTCT GGAATGCAAA TGCAACGGGT GGCTTAAGAG CGATGGGCGC
AAAGACTATG TTAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGAAGA
CAAACTAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTCATGAAG
TCACGGGTCT GTAA (SEQ ID NO: 31)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Gln Ala Ile Leu
Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Leu Thr Leu Thr Asp
Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu
Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Pro Gly Leu Ile Pro Glu Gln Pro
Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
Ala Lys Glu Pro Glu Gly Leu Asp Phe Leu Phe Asp Ala Tyr Pro Glu Asn Ser Lys Ala Lys Glu
Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Asp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn

Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Lys Asp Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 32)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA
 AAAACAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
 TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
 CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGCTGA CCCTGACCGA
 TATGCTGGAT TTTATCGAGT TTA CTCTTGC GACGCGCAAA CTTACCGGCA
 ACGCTGCTAT TAGAGAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
 GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
 GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC
 CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTAA CAAAAATATC
 AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
 CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCGCTG
 GCAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG
 ACCGCCGAGG CGCGCCAAAT CCACCCCGAA GGGGTGCTGA TTGATGGCGA
 ACTGGTGTAT CACGAACAGG TTGCAAAAGA GCCGGAGGGG TTGGATTTC
 TGTTTGATGC CTACCCGGAG AATAGCAAAG CGAAAGAATT TGCGGAAGTG
 GCGGAATCCC GCACCGCAAG CAATGGTATC GCAAATAAAT CGCTGAAGGG
 CACCATTCT GAGAAAGAAG CACAGTGTAT GAAGTTCCAG GTGTGGGACT
 ACGTGCCACT GGTTGAGATC TACTCCCTGC CAGCGTTTCG CCTGAAATAC
 GATGTGCGTT TTTCAAACT GGAACAGATG ACGAGCGGAT ATGATAAAGT
 AATTTTAATC GAGAACCAGG TCGTGAACAA CCTGGATGAA GCAAAAGTTA
 TCTATAAAAA ATATATTGAT CAGGGCTTAG AAGGCATTAT CCTGAAGAAC
 ATTGATGGCC TTTGGGAAAA TGCACGCAGC AAAAACCTGT ATAAATTTAA
 AGAAGTGATT GATGTAGATC TGAAGATTGT TGGTATTTAC CCGCATCGCA
 AAGACCCGAC CAAAGCAGGT GGTTCATCC TGGAATCTGA ATGCGGTAAA
 ATTAAAGTGA ACGCAGGCAG TGGTTTGAAA GATAAAGCTG GTGTTAAAG
 TCACGAACTG GATCGCACGC GCATCATGGA AAACCAGAAC TATTATATCG
 GCAAAATTCT GGAATGCAAA TGCAACGGGT GGCTTAAGAG CGATGGGCGC
 AAAGACTATG TTAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGAAGA
 CAAAATAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTCATGAAG
 TCACGGGTCT GTAA (SEQ ID NO: 33)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Gln Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Leu Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro

Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Ala Lys Glu Pro Glu Gly Leu Asp Phe Leu Phe Asp Ala Tyr Pro Glu Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
 Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Asp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
 Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Lys Asp Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 34)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA
 AAAACAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
 TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
 CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGACTA CCCTGACCGA
 TATGCTGGAT TTTATCGAGT TTA CTCTTG C GACGCGCAA CTTACCGGCA
 ACGCTGCTAT TAGAGAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
 GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
 GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC
 CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTAA CAAAAGAATC
 AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
 CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCGCTG
 GCAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG
 ACCGCCGAGG CGCGCCAAAT CCACCCCGAA GGGGTGCTGA TTGATGGCGA
 ACTGGTGTAT CACGAACAGG TTGCAAAAGA GCCGGAGGGG TTGGATTTC
 TGTTTGATGC TTACCCGCCT AATAGCAAAG CGAAAGAATT TGCGGAAGTG
 GCGGAATCCC GCACCGCAAG CAATGGTATC GCAAATAAAT CGCTGAAGGG
 CACCATTTCT GAGAAAGAAG CACAGTGTAT GAAGTTCCAG GTGTGGGACT
 ACGTGCCACT GGTTGAGATC TACTCCCTGC CAGCGTTTCG CCTGAAATAC
 GATGTGCGTT TTTCAAAACT GGAACAGATG ACGAGCGGAT ATGATAAAGT
 AATTTTAATC GAGAACCAGG TCGTGAACAA CCTGGATGAA GCAAAAGTTA
 TCTATAAAAA ATATATTGAT CAGGGCTTAG AAGGCATTAT CCTGAAGAAC
 ATTGATGGCC TTTGGGAAAA TGCACGCAGC AAAAACCTGT ATAAATTTAA
 AGAAGTGATT TGGGTAGATC TGAAGATTGT TGGTATTTAC CCGCATCGCA
 AAGACCCGAC CAAAGCAGGT GGTTCATCC TGGAATCTGA ATGCGGTAAA

ATTAAAGTGA ACGCAGGCAG TGGTTTGAAA GATAAAGCTG GTGTTAAAAG
 TCACGAACTG GATCGCACGC GCATCATGGA AAACCAGAAC TATTATATCG
 GCAAAATTCT GGAATGCAAA TGCAACGGGT GGCTTAAGAG CGATGGGCGC
 AAACCCTATG TGAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGAAGA
 CAAAATAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTCATGAAG
 TCACTGGTCT GTAATAA (SEQ ID NO: 35)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Gln Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Thr Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro
 Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Arg Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Ala Lys Glu Pro Glu Gly Leu Asp Phe Leu Phe Asp Ala Tyr Pro Pro Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
 Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Trp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
 Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Lys Pro Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 36)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA
 AAAAAAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
 TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
 CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGACTA CCCTGACCGA
 TATGCTGGAT TTTATCGAGT TTA CTCTTGC GACGCGCAAA CTTACCGGCA
 ACGCTGCTAT TAGAGAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
 GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
 GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC
 CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTAA CAAAAATATC
 AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
 CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCGCTG
 GCAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG

ACCGCCGAGG	CGCGCCAAAT	CCACCCCGAA	GGGGTGCTGA	TTGATGGCGA
ACTGGTGTAT	CACGAACAGG	TTGCAAAAGA	GCCGGAGGGG	TTGGATGCTC
TGTTTGATCG	ATACCCGGAA	AATAGCAAAG	CGAAAGAATT	TGCGGAAGTG
GCGGAATCCC	GCACCGCAAG	CAATGGTATC	GCAAATAAAT	CGCTGAAGGG
CACCATTCT	GAGAAAGAAG	CACAGTGTAT	GAAGTTCCAG	GTGTGGGACT
ACGTGCCACT	GGTTGAGATC	TACTCCCTGC	CAGCGTTTCG	CCTGAAATAC
GATGTGCGTT	TTTCAAAACT	GGAACAGATG	ACGAGCGGAT	ATGATAAAGT
AATTTTAATC	GAGAACCAGG	TCGTGAACAA	CCTGGATGAA	GCAAAAGTTA
TCTATAAAAA	ATATATTGAT	CAGGGCTTAG	AAGGCATTAT	CCTGAAGAAC
ATTGATGGCC	TTTGGGAAAA	TGCACGCAGC	AAAAACCTGT	ATAAATTTAA
AGAAGTGATT	TGGGTAGATC	TGAAGATTGT	TGGTATTTAC	CCGCATCGCA
AAGACCCGAC	CAAAGCAGGT	GGTTTCATCC	TGGAATCTGA	ATGCGGTAAA
ATTAAAGTGA	ACGCAGGCAG	TGGTTTGAAA	GATAAAGCTG	GTGTTAAAAG
TCACGAACTG	GATCGCACGC	GCATCATGGA	AAACCAGAAC	TATTATATCG
GCAAAATTCT	GGAATGCAAA	TGCAACGGGT	GGCTTAAGAG	CGATGGGCGC
AAACCCTATG	TGAAATTGTT	CCTGCCGATT	GCGATTCGCC	TTCGCGAAGA
CAAAACTAAG	GCGAATACTT	TCGAAGATGT	GTTCGGTGAT	TTTCATGAAG

TCACGGGTCT GTAA (SEQ ID NO: 37)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Lys Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Thr Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro
 Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Ala Lys Glu Pro Glu Gly Leu Asp Ala Leu Phe Asp Arg Tyr Pro Glu Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
 Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Trp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
 Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Lys Pro Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 38)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA
 AAAACAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
 TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
 CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGCTTA CCCTGACCGA
 TATGCTGGAT TTTATCGAGT TTACTCTTGC GACGCGCAAA CTTACCGGCA
 ACGCTGCTAT TAGAAAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
 GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
 GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC
 CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTAA CAAAAATATC
 AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
 CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCGCTG
 GCAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG
 ACCGCCGAGG CGCGCCAAAT CCACCCCGAA GGGGTGCTGA TTGATGGCGA
 ACTGGTGTAT CACGAACAGG TTGCAAAAGA GCCGGAGGGG TTGGATGCTC
 TGTTTGATCG ATACCCGCCT AATAGCAAAG CGAAAGAATT TGCGGAAGTG
 GCGGAATCCC GCACCGCAAG CAATGGTATC GCAAATAAAT CGCTGAAGGG
 CACCATTCTT GAGAAAGAAG CACAGTGTAT GAAGTTCCAG GTGTGGGACT
 ACGTGCCACT GGTTGAGATC TACTCCCTGC CAGCGTTTCG CCTGAAATAC
 GATGTGCGTT TTTCAAAACT GGAACAGATG ACGAGCGGAT ATGATAAAGT
 AATTTTAATC GAGAACCAGG TCGTGAACAA CCTGGATGAA GCAAAAGTTA
 TCTATAAAAA ATATATTGAT CAGGGCTTAG AAGGCATTAT CCTGAAGAAC
 ATTGATGGCC TTTGGGAAAA TGCACGCAGC AAAAACCTGT ATAAATTTAA
 AGAAGTGATT GATGTAGATC TGAAGATTGT TGGTATTTAC CCGCATCGCA
 AAGACCCGAC CAAAGCAGGT GGTTCATCC TGGAATCTGA ATGCGGTAAA
 ATTAAAGTGA ACGCAGGCAG TGGTTTGAAA GATAAAGCTG GTGTTAAAAG
 TCACGAACTG GATCGCACGC GCATCATGGA AAACCAGAAC TATTATATCG
 GCAAAATTCT GGAGTGCAAA TGCAACGGGT GGCTTAAGAG CGCTGGGCGC
 AAAGACGGTC TGAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGAAGA
 CAAAATAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTTCATGAAG
 TCACGGGTCT GTAA (SEQ ID NO: 39)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Gln Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Leu Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Lys
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro
 Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val

Ala Lys Glu Pro Glu Gly Leu Asp Ala Leu Phe Asp Arg Tyr Pro Pro Asn Ser Lys Ala Lys Glu
Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Asp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Ala
Gly Arg Lys Asp Gly Leu Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 40)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA
AAAACAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGCTTA CCCTGACCGA
TATGCTGGAT TTTATCGAGT TTACTCTTGC GACGCGCAAA CTTACCGGCA
ACGCTGCTAT TAGAGAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC
CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTAA CAAAAGAATC
AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCGCTG
GCAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG
ACCGCCGAGG CGCGCCAAAT CCACCCCGAA GGGGTGCTGA TTGATGGCGA
ACTGGTGTAT CACGAACAGG TTGCAAAAGA GCCGGAGGGG TTGGATGCTC
TGTTTGATAA TTACCCGGAA AATAGCAAAG CGAAAGAATT TGCGGAAGTG
GCGGAATCCC GCACCGCAAG CAATGGTATC GCAAATAAAT CGCTGAAGGG
CACCATTCT GAGAAAGAAG CACAGTGTAT GAAGTTCCAG GTGTGGGACT
ACGTGCCACT GGTGAGATC TACTCCCTGC CAGCGTTTCG CCTGAAATAC
GATGTGCGTT TTTCAAACT GGAACAGATG ACGAGCGGAT ATGATAAAGT
AATTTTAATC GAGAACCAGG TCGTGAACAA CCTGGATGAA GCAAAAGTTA
TCTATAAAAA ATATATTGAT CAGGGCTTAG AAGGCATTAT CCTGAAGAAC
ATTGATGGCC TTTGGGAAAA TGCACGCAGC AAAAACCTGT ATAAATTTAA
AGAAGTGATT GATGTAGATC TGAAGATTGT TGGTATTTAC CCGCATCGCA
AAGACCCGAC CAAAGCAGGT GGTTCATCC TGAATCTGA ATGCGGTAAA
ATTAAAGTGA ACGCAGGCAG TGGTTTGAAA GATAAAGCTG GTGTTAAAAG
TCACGAACTG GATCGCACGC GCATCATGGA AAACCAGAAC TATTATATCG
GCAAAATTCT GGAGTGCAAA TGCAACGGGT GGCTTAAGAG CGATGGGCGC
AAAGACTATC TGAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGAAGA

CAAAACTAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTCATGAAG
TCACGGGTCT GTAA (SEQ ID NO: 41)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Gln Ala Ile Leu
Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Leu Thr Leu Thr Asp
Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu
Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro
Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Arg Ile Lys Phe Pro Ala Phe Ala Gln
Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
Ala Lys Glu Pro Glu Gly Leu Asp Ala Leu Phe Asp Asn Tyr Pro Glu Asn Ser Lys Ala Lys Glu
Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Asp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
Gly Arg Lys Asp Tyr Leu Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 42)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA
AAAACAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGCTGA CCCTGACCGA
TATGCTGGAT TTTATCGAGT TTA CTCTTGC GACGCGCAA CTTACCGGCA
ACGCTCGGAT TAGAAAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC
CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTA CAAAAGAATC
AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCGCTG
GCAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG
ACCGCCGAGG CGCGCCAAAT CCACCCCGAA GGGGTGCTGA TTGATGGCGA
ACTGGTGTAT CACGAACAGG TTGCAAAAGA GCCGGAGGGG TTGGATGCTC
TGTTTGATAA TTACCCGCCT AATAGCAAAG CGAAAGAATT TGCGGAAGTG
GCGGAATCCC GCACCGCAAG CAATGGTATC GCAAATAAAT CGCTGAAGGG

CACCATTCTCT GAGAAAGAAG CACAGTGTAT GAAGTTCCAG GTGTGGGACT
 ACGTGCCACT GGTTGAGATC TACTCCCTGC CAGCGTTTCG CCTGAAATAC
 GATGTGCGTT TTTCAAAACT GGAACAGATG ACGAGCGGAT ATGATAAAGT
 AATTTTAATC GAGAACCAGG TCGTGAACAA CCTGGATGAA GCAAAAAGTTA
 TCTATAAAAA ATATATTGAT CAGGGCTTAG AAGGCATTAT CCTGAAGAAC
 ATTGATGGCC TTTGGGAAAA TGCACGCAGC AAAAACCTGT ATAAATTTAA
 AGAAGTGATT GATGTAGATC TGAAGATTGT TGGTATTTAC CCGCATCGCA
 AAGACCCGAC CAAAGCAGGT GGTTCATCC TGGAATCTGA ATGCGGTAAA
 ATTAAAGTGA ACGCAGGCAG TGGTTTGAAA GATAAAGCTG GTGTTAAAG
 TCACGAACTG GATCGCACGC GCATCATGGA AAACCAGAAC TATTATATCG
 GCAAAATTCT GGAGTGCAAA TGCAACGGGT GGCTTAAGAG CGATGGGCGC
 AAAGACTATG TTAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGAAGA
 CAAAATAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTCATGAAG
 TCACGGGTCT GTAA (SEQ ID NO: 43)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Gln Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Leu Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Arg Ile Arg Lys
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro
 Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Arg Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Ala Lys Glu Pro Glu Gly Leu Asp Ala Leu Phe Asp Asn Tyr Pro Pro Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
 Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Asp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
 Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Lys Asp Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 44)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA
 AAAAAAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
 TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
 CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGACTA CCCTGACCGA

TATGCTGGAT	TTTATCGAGT	TTACTCTTGC	GACGCGCAAA	CTTACCGGCA
ACGCTGCTAT	TAGAGAATTG	ACGGGCTATA	TTACCGATGG	CAAAAAAGAT
GATGTTGAGG	TGCTGCGTCG	CGTCATGATG	CGTGATCTGG	AGTGCGGTGC
GTCAGTGAGC	ATCGCTAACA	AAGTGTGGAA	GGGTTTGATC	CCGGAACAGC
CACAGATGCT	TGCAAGCAGC	TACGATGAAA	AGGGCATTA	CAAAAACATC
AAATTTCCGG	CTTTCGCCCA	GCTGAAAGCG	GATGGCGCGC	GCTGCTTCGC
CGAGGTACGC	GGTGACGAAC	TGGACGACGT	TCGTCTGCTG	TCTCGCAAAG
GTAATGAATA	TCTGGGTCTG	GATCTGCTGA	AAGAAGAACT	GATTAAGATG
ACCGCCGAGG	CGCGCCAAAT	CCACCCCGAA	GGGGTGCTGA	TTGATGGCGA
ACTGGTGTAT	CACGAACAGG	TTGCAAAAGA	GCCGGAGGGG	AAAGACGCTC
TGTTTGATCG	ATACCCGGAA	AATAGCAAAG	CGAAAGAATT	TGCGGAAGTG
GCGGAATCCC	GCACCGCAAG	CAATGGTATC	GCAAATAAAT	CGCTGAAGGG
CACCATTTCT	GAGAAAGAAG	CACAGTGTAT	GAAGTTCCAG	GTGTGGGACT
ACGTGCCACT	GGTTGAGATC	TACTCCCTGC	CAGCGTTTCG	CCTGAAATAC
GATGTGCGTT	TTTCAAAACT	GGAACAGATG	ACGAGCGGAT	ATGATAAAGT
AATTTTAATC	GAGAACCAGG	TCGTGAACAA	CCTGGATGAA	GCAAAAGTTA
TCTATAAAAA	ATATATTGAT	CAGGGCTTAG	AAGGCATTAT	CCTGAAGAAC
ATTGATGGCC	TTTGGGAAAA	TGCACGCAGC	AAAAACCTGT	ATAAATTTAA
AGAAGTGATT	TGGGTAGATC	TGAAGATTGT	TGGTATTTAC	CCGCATCGCA
AAGACCCTAC	CAAAGCAGGT	GGTTTCATCC	TGGAATCTGA	ATGCGGTAAA
ATTAAAGTGA	ACGCAGGCAG	TGGTTTGAAA	GATAAAGCTG	GTGTTAAAG
TCACGAACTG	GATCGCACGC	GCATCATGCG	GAACCAGAAC	TATTATATCG
GCAAAATTCT	GGAATGCAAA	TGCAACGGGT	GGCTTAAGAG	CGATGGGCGC
AAGCCCTATG	TGAAATTGTT	CCTGCCGATT	GCGATTCGCC	TTCGCGGAGA
CAAAACTAAG	GCGAATACTT	TCGAAGATGT	GTTCGGTGAT	TTTCATGAAG

TCACGGGTCT GTAA (SEQ ID NO: 45)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Lys Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Thr Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro
 Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Lys Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Ala Lys Glu Pro Glu Gly Lys Asp Ala Leu Phe Asp Arg Tyr Pro Glu Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly

Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Trp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
 Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Arg Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Lys Pro Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Gly Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 46)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA
 AAAAAAGGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
 TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
 CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGACTA CCCTGACCGA
 TATGCTGGAT TTTATCGAGT TTA CTCTTGC GACGCGCAA CTTACCGGCA
 ACGCTGCTAT TAGAGAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
 GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
 GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC
 CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTAA CAAAAATATC
 AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
 CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCGCTG
 GCAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG
 ACCGCCGAGG CGCGCCAAAT CCACCCCGAA GGGGTGCTGA TTGATGGCGA
 ACTGGTGTAT CACGAACAGG TTGCAAAAGA GCCGGAGGGG TTGGATGCTC
 TGTTTGATCG ATACCCGGAA AATAGCAAAG CGAAAGAATT TGCGGAAGTG
 GCGGAATCCC GCACCGCAAG CAATGGTATC GCAAATAAAT CGCTGAAGGG
 CACCATTCT GAGAAAGAAG CACAGTGTAT GAAGTTCCAG GTGTGGGACT
 ACGTGCCACT GGTTGAGATC TACTCCCTGC CAGCGTTTCG CCTGAAATAC
 GATGTGCGTT TTTCAAACT GGAACAGATG ACGAGCGGAT ATGATAAAGT
 AATTTTAATC GAGAACCAGG TCGTGAACAA CCTGGATGAA GCAAAAGTTA
 TCTATAAAAA ATATATTGAT CAGGGCTTAG AAGGCATTAT CCTGAAGAAC
 ATTGATGGCC TTTGGGAAAA TGCACGCAGC AAAAACCTGT ATAAATTTAA
 AGAAGTGATT TGGGTAGATC TGAAGATTGT TGGTATTTAC CCGCATCGCA
 AAAAACCTAC CAAAGCAGGT GGTTCATCC TGGAATCTGA ATGCGGTAAA
 ATTAAAGTGA ACGCAGGCAG TGGTTTGAAA GATAAAGCAG GTGTTAAAAG
 TCACGAACTG GATCGCACGC GCATCATGGA GAACCAGAAC TATTATATCG
 GCAAAATTCT GGAATGCAAA TGCAACGGGT GGCTTAAGAG CGATGGGCGC
 AAGCCCCGTG TGAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGGAGA
 CAAAATAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTTCATGAAG
 TCACGGGTCT GTAA (SEQ ID NO: 47)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Lys Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln

Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Thr Leu Thr Asp Met
 Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu Leu
 Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg Asp
 Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro Gln
 Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Ala Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Ala Lys Glu Pro Glu Gly Leu Asp Ala Leu Phe Asp Arg Tyr Pro Glu Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
 Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Trp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Lys Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn Ala
 Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met Glu
 Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp Gly
 Arg Lys Pro Arg Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Gly Asp Lys Thr Lys Ala
 Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 48)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTAAGT CCACTAAACA
 AAAAAAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
 TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
 CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGACTA CCCTGACCGA
 TATGCTGGAT TTTATCGAGT TTA CTCTTGC GACGCGCAA CTTACCGGCA
 ACGCTGCTAT TAGAGAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
 GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
 GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC
 CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTA CAAAAACATC
 AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
 CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCAAAG
 GTAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG
 ACCGCCGAGG CGCGCCAAAT CCACCCCGAA GGGGTGCTGA TTGATGGCGA
 ACTGGTGTAT CACGAACAGG TTGCAAAAGA GCCGGAGGGG TTAAAAGCTC
 TGTTTGATCG ATACCCGGAA AATAGCAAAG CGAAAGAATT TGCGGAAGTG
 GCGGAATCCC GCACCGCAAG CAATGGTATC GCAAATAAAT CGCTGAAGGG
 CACCATTTCT GAGAAAGAAG CACAGTGTAT GAAGTTCCAG GTGTGGGACT
 ACGTGCCACT GGTTGAGATC TACTCCCTGC CAGCGTTTCG CCTGAAATAC
 GATGTGCGTT TTTCAAAACT GGAACAGATG ACGAGCGGAT ATGATAAAGT
 AATTTTAATC GAGAACCAGG TCGTGAACAA CCTGGATGAA GCAAAAGTTA

TCTATAAAAA ATATATTGAT CAGGGCTTAG AAGGCATTAT CCTGAAGAAC
 ATTGATGGCC TTTGGGAAAA TGCACGCAGC AAAAACCTGT ATAAATTTAA
 AGAAGTGATT TGGGTAGATC TGAAGATTGT TGGTATTTAC CCGCATCGCA
 AAGACAAAAC CAAAGCAGGT GGTTCATCC TGAATCTGA ATGCGGTAAA
 ATTAAGTGA ACGCAGGCAG TGGTTTGAAA GATAAAGCTG GTGTTAAAAG
 TCACGAACTG GATCGCACGC GCATCATGGA GAACCAGAAC TATTATATCG
 GCAAAATTCT GGAATGCAAA TGCAACGGGT GGCTTAAGAG CGATGGGCGC
 GGTCCCTATG TGAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGAAGA
 CAAAATAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTCATGAAG
 TCACGGGTCT GTAA (SEQ ID NO: 49)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Lys Ser Thr Lys Gln Lys Lys Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Thr Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro
 Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Lys Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Ala Lys Glu Pro Glu Gly Leu Lys Ala Leu Phe Asp Arg Tyr Pro Glu Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
 Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Trp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Asp Lys Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
 Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Gly Pro Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 50)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTGGTT CCACTAAACA
 AAAAAAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
 TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
 CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGACTA CCCTGACCGA
 TATGCTGGAT TTTATCGAGT TTAATCTTGC GACGCGCAAA CTTACCGGCA
 ACGCTGCTAT TAGAGAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
 GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
 GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC

CACAGATGCT	TGCAAGCAGC	TACGATGAAA	AGGGCATTAA	CAAAAATATC
AAATTTCCGG	CTTTCGCCCA	GCTGAAAGCG	GATGGCGCGC	GCTGCTTCGC
CGAGGTACGC	GGTGACGAAC	TGGACGACGT	TCGTCTGCTG	TCTCGCAAAG
GTAATGAATA	TCTGGGTCTG	GATCTGCTGA	AAGAAGAACT	GATTAAGATG
ACCGCCGAGG	CGCGCCAAAT	CCACCCCGAA	GGGGTGCTGA	TTGATGGCGA
ACTGGTGTAT	CACGAACAGG	TTGCAAAAGA	GCCGGAGGGG	TTGGATGCTC
TGTTTGATCG	ATACCCGGAA	AATAGCAAAG	CGAAAGAATT	TGCGGAAGTG
GCGGAATCCC	GCACCGCAAG	CAATGGTATC	GCAAATAAAT	CGCTGAAGGG
CACCATTTCT	GAGAAAGAAG	CACAGTGTAT	GAAGTTCCAG	GTGTGGGACT
ACGTGCCACT	GGTTGAGATC	TACTCCCTGC	CAGCGTTTCG	CCTGAAATAC
GATGTGCGTT	TTTCAAAACT	GGAACAGATG	ACGAGCGGAT	ATGATAAAGT
AATTTTAATC	GAGAACCAGG	TCGTGAACAA	CCTGGATGAA	GCAAAAGTTA
TCTATAAAAA	ATATATTGAT	CAGGGCTTAG	AAGGCATTAT	CCTGAAGAAC
ATTGATGGCC	TTTGGGAAAA	TGCACGCAGC	AAAAACCTGT	ATAAATTTAA
AGAAGTGATT	TGGGTAGCTC	TGAAGATTGT	TGGTATTTAC	CCGCATCGCA
AAGACAAAAC	CAAAGCAGGT	GGTTTCATCC	TGGAATCTGA	ATGCGGTAAA
ATTAAAGTGA	ACGCAGGCAG	TGGTTTGAAA	GATAAAGCAG	GTGTTAAAAG
TCACGAACTG	GATCGCACGC	GCATCATGGA	GAACCAGAAC	TATTATATCG
GCAAAATTCT	GGAATGCAAA	TGCAACGGGT	GGCTTAAGAG	CGATGGGCGC
AAACCCTATG	TGAAATTGTT	CCTGCCGATT	GCGATTCGCC	TTCGCGAAGA
CAAAACTAAG	GCGAATACTT	TCGAAGATGT	GTTCGGTGAT	TTTCATGAAG

TCACGGGTCT GTAA (SEQ ID NO: 51)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Gly Ser Thr Lys Gln Lys Lys Ala Ile Leu
Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Thr Thr Leu Thr Asp
Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu
Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro
Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
Leu Ser Arg Lys Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
Ala Lys Glu Pro Glu Gly Leu Asp Ala Leu Phe Asp Arg Tyr Pro Glu Asn Ser Lys Ala Lys Glu
Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Asp Glu Ala Lys Val Ile Tyr Lys
Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Trp Val Ala Leu Lys Ile Val Gly Ile Tyr Pro His
Arg Lys Asp Lys Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn

Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Lys Pro Tyr Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Glu Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 52)

ATGATTCTTA AAATTCTGAA CGAAATTGCA AGCATTAAGT CCACTAAACA
 AAAAAAAGCC ATTCTGGAAA AAAATAAAGA CAATGAACTG TTAAAGCGCG
 TGTATCGCCT GACCTATTCG CGTGGCCTGC AATACTATAT TAAAAAATGG
 CCCAAACCGG GCATTGCGAC GCAGAGCTTC GGCATGACTA CCCTGACCGA
 TATGCTGGAT TTTATCGAGT TTA CTCTTGC GACGCGCAAA CTTACCGGCA
 ACGCTGCTAT TAGAGAATTG ACGGGCTATA TTACCGATGG CAAAAAAGAT
 GATGTTGAGG TGCTGCGTCG CGTCATGATG CGTGATCTGG AGTGCGGTGC
 GTCAGTGAGC ATCGCTAACA AAGTGTGGAA GGGTTTGATC CCGGAACAGC
 CACAGATGCT TGCAAGCAGC TACGATGAAA AGGGCATTAA CAAAAACATC
 AAATTTCCGG CTTTCGCCCA GCTGAAAGCG GATGGCGCGC GCTGCTTCGC
 CGAGGTACGC GGTGACGAAC TGGACGACGT TCGTCTGCTG TCTCGCAAAG
 GTAATGAATA TCTGGGTCTG GATCTGCTGA AAGAAGAACT GATTAAGATG
 ACCGCCGAGG CGCGCCAAAT CCACCCCGAA GGGGTGCTGA TTGATGGCGA
 ACTGGTGTAT CACGAACAGG TTGCAAAAGA GCCGGAGGGG TTAAAAGCTC
 TGTTTGATCG ATACCCGGAA AATAGCAAAG CGAAAGAATT TGCGGAAGTG
 GCGGAATCCC GCACCGCAAG CAATGGTATC GCAAATAAAT CGCTGAAGGG
 CACCATTCT GAGAAAGAAG CACAGTGTAT GAAGTTCCAG GTGTGGGACT
 ACGTGCCACT GGTTGAGATC TACTCCCTGC CAGCGTTTCG CCTGAAATAC
 GATGTGCGTT TTTCAAAACT GGAACAGATG ACGAGCGGAT ATGATAAAGT
 AATTTTAATC GAGAACCAGG TCGTGAACAA CCTGCTGGAA GCAAAAGTTA
 TCTATAAAAA ATATATTGAT CAGGGCTTAG AAGGCATTAT CCTGAAGAAC
 ATTGATGGCC TTTGGGAAAA TGCACGCAGC AAAAACCTGT ATAAATTTAA
 AGAAGTGATT TGGGTAGATC TGAAGATTGT TGGTATTTAC CCGCATCGCA
 AAGACCCTAC CAAAGCAGGT GGTTCATCC TGGAATCTGA ATGCGGTAAA
 ATTAAAGTGA ACGCAGGCAG TGGTTTGAAA GATAAAGCAG GTGTTAAAAG
 TCACGAACTG GATCGCACGC GCATCATGGA GAACCAGAAC TATTATATCG
 GCAAAATTCT GGAATGCAAA TGCAACGGGT GGCTTAAGAG CGATGGGCGC
 AAGCCCGGTG TGAAATTGTT CCTGCCGATT GCGATTCGCC TTCGCGGAGA
 CAAAATAAG GCGAATACTT TCGAAGATGT GTTCGGTGAT TTTCATGAAG
 TCACGGGTCT GTAA (SEQ ID NO: 53)

Met Ile Leu Lys Ile Leu Asn Glu Ile Ala Ser Ile Lys Ser Thr Lys Gln Lys Lys Ala Ile Leu
 Glu Lys Asn Lys Asp Asn Glu Leu Leu Lys Arg Val Tyr Arg Leu Thr Tyr Ser Arg Gly Leu Gln
 Tyr Tyr Ile Lys Lys Trp Pro Lys Pro Gly Ile Ala Thr Gln Ser Phe Gly Met Thr Thr Leu Thr Asp
 Met Leu Asp Phe Ile Glu Phe Thr Leu Ala Thr Arg Lys Leu Thr Gly Asn Ala Ala Ile Arg Glu
 Leu Thr Gly Tyr Ile Thr Asp Gly Lys Lys Asp Asp Val Glu Val Leu Arg Arg Val Met Met Arg
 Asp Leu Glu Cys Gly Ala Ser Val Ser Ile Ala Asn Lys Val Trp Lys Gly Leu Ile Pro Glu Gln Pro

Gln Met Leu Ala Ser Ser Tyr Asp Glu Lys Gly Ile Asn Lys Asn Ile Lys Phe Pro Ala Phe Ala Gln
 Leu Lys Ala Asp Gly Ala Arg Cys Phe Ala Glu Val Arg Gly Asp Glu Leu Asp Asp Val Arg Leu
 Leu Ser Arg Lys Gly Asn Glu Tyr Leu Gly Leu Asp Leu Leu Lys Glu Glu Leu Ile Lys Met Thr
 Ala Glu Ala Arg Gln Ile His Pro Glu Gly Val Leu Ile Asp Gly Glu Leu Val Tyr His Glu Gln Val
 Ala Lys Glu Pro Glu Gly Leu Lys Ala Leu Phe Asp Arg Tyr Pro Glu Asn Ser Lys Ala Lys Glu
 Phe Ala Glu Val Ala Glu Ser Arg Thr Ala Ser Asn Gly Ile Ala Asn Lys Ser Leu Lys Gly Thr Ile
 Ser Glu Lys Glu Ala Gln Cys Met Lys Phe Gln Val Trp Asp Tyr Val Pro Leu Val Glu Ile Tyr
 Ser Leu Pro Ala Phe Arg Leu Lys Tyr Asp Val Arg Phe Ser Lys Leu Glu Gln Met Thr Ser Gly
 Tyr Asp Lys Val Ile Leu Ile Glu Asn Gln Val Val Asn Asn Leu Leu Glu Ala Lys Val Ile Tyr Lys
 Lys Tyr Ile Asp Gln Gly Leu Glu Gly Ile Ile Leu Lys Asn Ile Asp Gly Leu Trp Glu Asn Ala Arg
 Ser Lys Asn Leu Tyr Lys Phe Lys Glu Val Ile Trp Val Asp Leu Lys Ile Val Gly Ile Tyr Pro His
 Arg Lys Asp Pro Thr Lys Ala Gly Gly Phe Ile Leu Glu Ser Glu Cys Gly Lys Ile Lys Val Asn
 Ala Gly Ser Gly Leu Lys Asp Lys Ala Gly Val Lys Ser His Glu Leu Asp Arg Thr Arg Ile Met
 Glu Asn Gln Asn Tyr Tyr Ile Gly Lys Ile Leu Glu Cys Lys Cys Asn Gly Trp Leu Lys Ser Asp
 Gly Arg Lys Pro Gly Val Lys Leu Phe Leu Pro Ile Ala Ile Arg Leu Arg Gly Asp Lys Thr Lys
 Ala Asn Thr Phe Glu Asp Val Phe Gly Asp Phe His Glu Val Thr Gly Leu (SEQ ID NO: 54)

ПРИМЕР 1

Получение гена лигазы и конструирование экспрессионных векторов

Фермент Т4 ДНК–лигаза дикого типа (WT) (SEQ ID NO: 2) кодируется геном бактериофага Т4. Синтетический ген (SEQ ID NO: 3), кодирующий меченную 6–гистидином версию WT Т4 ДНК–лигазы (SEQ ID NO: 4), сконструировали и субклонировали в вектор экспрессии *Escherichia coli* pCK100900i (см., например, патент США № 7629157 и публикацию заявки на патент США 2016/0244787, которые оба включены в настоящий документ посредством ссылки). Второй синтетический ген (SEQ ID NO: 5), кодирующий меченную 6–гистидином Т4 ДНК–лигазу (SEQ ID NO: 6), сконструировали с оптимизацией кодонов для экспрессии в *E. coli*, синтезировали и клонировали в pCK100900i. Эти плазмидные конструкции трансформировали в штамм *E. coli*, полученный из W3110. Для создания библиотек вариантов генов из этих плазмид использовали методы направленной эволюции, общеизвестные специалистам в данной области (см., например, патент США № 8383346 и WO 2010/144103, которые оба включены в настоящий документ посредством ссылки). Замены в вариантах ферментов, описанных в данном документе, обозначены со ссылкой на немеченный фермент WT Т4 ДНК–лигаза (т.е. SEQ ID NO: 2) или его варианты, как указано.

ПРИМЕР 2

Экспрессия и очистка ДНК–лигазы Т4 с высокой пропускной способностью (НТР)

Высокопроизводительный (НТР) Рост Т4 ДНК–лигазы и Вариантов

Трансформированные клетки *E. coli* отбирали путем посева на чашки с агаром Лурия–Бертани, содержащие 1% глюкозу и 30 мкг/мл хлорамфеникола. После инкубации в течение ночи при 37°C колонии помещали в лунки 96–луночных неглубоких плоскодонных планшетов NUNC™ (Термо–Scientific), наполненные 180 мкл/лунку среды

Лурия–Бертани, дополненной 1% глюкозой и 30 мкг/мл хлорамфеникола. Культурам позволяли расти в течение ночи в течение 18–20 часов в шейкере (200 об/мин, 30°C и относительная влажность 85%; Kuhner). Растущие в течение ночи образцы (20 мкл) переносили в 96–луночные глубокие планшеты Costar, наполненные 380 мкл Terrific Broth, дополненной 30 мкг/мл хлорамфеникола. Планшеты инкубировали в течение 120 минут в шейкере (250 об/мин, 30°C и относительная влажность 85%; Kuhner) до достижения OD₆₀₀ между 0,4 и 0,8. Затем клетки индуцировали 40 мкл 10 mM IPTG в стерильной воде и инкубировали в течение ночи в течение 18–20 часов в шейкере (250 об/мин, 30°C и относительная влажность 85%; Kuhner). Клетки осаждали (4000 об/мин x 20 мин), супернатанты выбрасывали, и клетки замораживали при –80°C перед анализом.

Лизис НТР осадка

Клеточный осадок ресуспендировали в 25 mM Tris–HCl, pH 7,5 по 150 мкл/лунку, и в суспензии клеток добавляли 300 мкл лизирующего буфера (реагент В–Per (Термо Fisher), дополненного 1 мг/мл лизоцима и 0,1 mM сульфата магния). Смесь взбалтывали в течение 1,25 часов при комнатной температуре, осаждали (4000 об/мин x 20 мин), и супернатанты оставляли для очистки.

НТР Очистка Т4 Лигазы из неочищенных лизатов

Т4 ДНК–лигазу очищали из неочищенных экстрактов *E. coli* с помощью металл–аффинной хроматографии с использованием планшетов с никелевым покрытием большой емкости (НС) HIS–Select® (Sigma) в соответствии с инструкциями производителя. Планшеты HIS–Select уравнивали в общей сложности 800 мкл промывочного буфера (50 mM фосфата натрия pH 7,5, 300 mM NaCl, 25 mM имидазола, 0,1% о/о реагента TWEEN–20® (Sigma)) на лунку. Затем 200 мкл НТР лизата, содержащего Т4 лигазу и 200 мкл промывочного буфера перемешивали, загружали на планшет и центрифугировали в течение 1 мин при 2000 относительного центробежного ускорения (rcf) и 4°C. Планшет дважды промывали 600 мкл промывочного буфера/лунку с 3 мин центрифугированием при 3000 rcf и 4°C для каждого промывания. Образцы лигазы элюировали с добавлением 200 мкл элюирующего буфера (50 mM фосфата натрия pH 7,5, 300 mM NaCl, 25 mM имидазола, 0,1% о/о реагента TWEEN®–20) путем центрифугирования в течение 1 мин @ 3000 rcf при 4°C.

В элюатах заменяли буфер с использованием вращающихся планшетов для обессоливания Zeba™ (Термо Fisher). Кратко, планшеты дважды уравнивали 375 мкл буфера хранения 2x Т4 ДНК–лигазы (100 mM Tris.HCl pH 7,5, 200 mM NaCl, 2 mM DTT, 2 mM EDTA, 0,2% м/о Triton X–100) на лунку и центрифугировали в течение 2 мин @ 1100 xg при 4°C. На планшеты для обессоливания загружали 100 мкл элюата образца HIS–Select и центрифугировали в течение 2 мин @ 1100 x g при 4°C. Элюат из планшета для обессоливания оставляли и перемешивали с равным объемом глицерина для итоговой концентрации буфера хранения 50 mM Tris.HCl pH 7,5, 100 mM NaCl, 1 mM DTT, 1 mM EDTA, 0,1% м/о Triton X–100.

ПРИМЕР 3

Экспрессия и Очистка Т4 ДНК–лигазы Во Встряхиваемой колбе

Экспрессия Во Встряхиваемой колбе

Отобранные НТР культуры, выращенные как описано выше, помещали на чашки с агаром Лурия–Бертани с 1% глюкозой и 30 мкг/мл хлорамфеникола и выращивали в течение ночи при 37°C. Одну колонию из каждой культуры переносили в 6 мл бульона Лурия–Бертани с 1% глюкозой и 30 мкг/мл хлорамфеникола. Культуры выращивали в течение 18 ч при 30°C, 250 об/мин и пересеивали при разведении приблизительно 1:10 в 250 мл Terrific Broth с 30 мкг/мл хлорамфеникола до итоговой OD₆₀₀ 0,2. Культуры инкубировали приблизительно в течение 3 часов при 30°C, 250 об/мин до OD₆₀₀ 0,6–0,8, а затем индуцировали добавлением IPTG с итоговой концентрацией 1 mM. Индуцированные культуры инкубировали в течение 20 ч при 30°C, 250 об/мин. После этого периода инкубации культуры центрифугировали при 4000 об/мин x 10 мин. Супернатант культуры выбрасывали, и осадок ресуспендировали в 35 мл 25 mM триэтаноламина, pH 7,5. Эту клеточную суспензию охлаждали в ледяной ванне и лизировали с использованием устройства разрушения клеток Microfluidizer (Microfluidics M- 110L). Необработанный лизат осаждали путем центрифугирования (16000 об/мин в течение 60 мин при 4°C), а супернатант затем фильтровали через мембрану PES 0,2 мкм для дополнительной очистки лизата.

Очистка Т4 ДНК–лигазы из Лизатов из Встряхиваемой колбы

Лизаты Т4 лигазы дополняли 1/10^{ой} объема SF элюирующего буфера (50 mM фосфата натрия pH 7,5, 500 mM NaCl, 300 mM имидазола, 0,1% o/o Tween–20®) на лунку. Затем лизаты очищали с использованием системы очистки AKTA Start и 5 мл колонки HisTrap FF (GE Healthcare) с использованием настройки AC Step HiF (рабочие параметры приведены ниже). SF промывочный буфер содержал 50 mM фосфата натрия pH 7,5, 500 mM NaCl, 25 mM имидазола и 0,1% o/o TWEEN–20® (Sigma).

Таблица 3.1 Параметры очистки

Параметр	Объем
Объем Колонки	5 мл
Скорость потока	5 мл/мин
Предел давления	0,3 МПа
Объем образца	35 млс
Уравновешивающий объем	5 объемов колонки (CV) = 25 млс
Промывающий несвязанный объем	15 CV=75 млс
Элюирование	Изократная (стадия)
Объем элюирования	5 CV=25 млс
Объем фракций	3 млс
RE- уравновешивающий объем	5 CV=25 млс

Одну наиболее концентрированную 3 мл фракцию идентифицировали с помощью

УФ поглощения (A280) и диализировали в течение ночи в 2х буфере хранения лигазы (20 mM Tris-HCl pH 7,5, 100 mM KCl, 2 mM DTT, 0,2 mM EDTA) в течение ночи в кассете для диализа 10K Slide-A-Lyzer™ (Термо Fisher) для замены буфера, и в диализированный материал добавляли равный объем глицерина. Концентрации лигазы в препаратах измеряли с помощью анализа Бредфорда и поглощения при 280 нм.

ПРИМЕР 4

Анализ Лигирования С Планшетным Определением

Для использования в скрининге библиотек вариантов Т4 ДНК-лигазы при вводе ДНК с низкой концентрацией разработали высокопроизводительный анализ лигирования на планшете с лунками. В анализе обнаруживают лигирование двух меченых двухцепочечных ДНК-субстратов. Меченый Cy®3 50-мерный фрагмент двухцепочечной ДНК «50-мерная Cy®3 вставка» состоял из двух одноцепочечных очищенных с помощью HPLC синтетических олигонуклеотидов (Integrated DNA Technologies) (SEQ ID NO: 7, SEQ ID NO: 8), и его получали с помощью отжига этих двух олигонуклеотидов в 1х буфере для отжига (10 mM Tris pH 7,5, 50 mM NaCl, 1 mM eEDTA). Полученная двухцепочечная «50-мерная Cy®3 вставка» имеет 3' выступы с одним основанием дезоксиаденином и 5' концы с монофосфатом на обоих концах молекулы, а внутри мечена Cy®3 красителем, присоединенным к фосфатному остову. Двухцепочечную молекулу «20-мерного биотинового адаптера», содержащего два одноцепочечных очищенных с помощью HPLC олигонуклеотида (Integrated DNA Technologies) (SEQ ID NO: 9 и SEQ ID NO: 10), также получали путем отжига в 1х буфере для отжига. Полученный 20-мерный дуплекс биотин-адаптер имеет защищенный фосфоротиоатом 5' дезокситимидиновый выступ и 5' фосфат на совместимом с лигированием конце и биотинилирован на противоположном конце. Совместимые с лигированием концы этих коротких модельных вставок и адаптерные субстраты были идентичны совместимым Т-А когезивным концам, создаваемым в технологических процессах получения NGS образцов для технологического процесса Illumina® TruSeq. При лигировании этих двух субстратов образуются молекулы ковалентно связанной двухцепочечной ДНК, меченые как биотином, так и красителем Cy®3. Одно- или двух-лигированные продукты могут образоваться при лигировании 20-мерного биотинового адаптера к одному или обоим концам 50-мерной Cy®3 вставки.

Реакции лигирования проводили в объеме 80 мкл в 1х буфере для лигирования (66 mM Tris, pH 7,5, 10 mM MgCl₂, 1 mM DTT) и при низких концентрациях субстрата для лигирования (1 нМ 50-мерной Cy®3 вставки и 5 нМ или 10 нМ 20-мерного биотинового адаптера). В реакционную смесь добавляли НТР-очищенную лигазу (5 мкл), и реакции инкубировали в течение 4 часов при 20°C, с последующими 16 часами при 4°C. Реакции останавливали добавлением 40 мкл 3х останавливающего раствора (45 mM EDTA и 0,15% реагент TWEEN®-20).

Стрептавидиновые планшеты с большой связывающей емкостью (Pierce) предварительно промывали 3 раза 200 мкл буфера TBSTE (25 mM Tris pH 7,5, 150 mM

NaCl, 0,05% реагент TWEEN–20, 1 mM EDTA). Затем в предварительно промытые стрептавидиновые планшеты добавляли 100 мкл реакционной смеси с остановленным лигированием, планшеты встряхивали в течение 1 ч при комнатной температуре, чтобы обеспечить связывание, и выбрасывали несвязанные супернатанты. Затем стрептавидиновые планшеты 3 раза промывали 200 мкл буфера TBSTE, один раз буфером TBS (25 mM Tris, 150 mM NaCl), и наполняли 200 мкл буфера TBS. Промывка эффективно удаляла нелигированный 50–мерный Cy®3 адаптер и оставляла флуоресцентно меченые продукты с одно– и двойным лигированием на поверхности с покрытыми стрептавидином лунками. Интенсивность флуоресценции измеряли с использованием многорежимного планшет–ридера Paradigm® (Beckman Coulter) с использованием фильтров возбуждения 535/35 нМ и эмиссии 595/35 нМ. В Таблице 4.1 представлены данные об улучшении активности для разных вариантов лигаз относительно SEQ ID NO: 2., тогда как в Таблице 4.2 представлена активность для разных вариантов лигаз относительно SEQ ID NO: 6, а в Таблице 4.3 представлена активность для разных вариантов лигаз относительно SEQ ID NO: 32.

Таблица 4.1 Улучшение активности вариантов лигаз Относительно SEQ ID NO: 2 (с 1 нМ Вставкой и 10 нМ Адаптером)		
Вариант #	Улучшение активности	Замены аминокислот Относительно SEQ ID NO: 2
1	++	C165A/S181T/K299P
2	++	C165A/S181T/V281A/K299A
3	++	S140A/S181T/L234M
4	++	P127K/I207R
5	+	P127K/L213M/C276G/I339V
6	+	P127K/L213M
7	++	K52E/A56R/N404K
8	++	Y238L/N241L/N404K/K412T/I462K
9	++	K52E/A56V/N404K/K412T
10	+	K52E/A56R/N404K
11	+	I462K
Уровни активности определяли относительно эталонного полипептида SEQ ID NO: 2 Улучшения активности определяют следующим образом: ++ = 1,3–1,6 + = 1,2–1,3		
Таблица 4.2 Улучшения активности вариантов лигазы Относительно SEQ ID NO: 6		
Вариант #	Улучшение активности	Замены аминокислот Относительно

		SEQ ID NO: 6
12	+++	E88R/K225A/E440K/T451K
13	+++	E88R/K306A/E440K/T451K
14	++	Q58K/E88R/K226E/K306A
15	++	L63R/E89K/T451K
16	++	T451K
17	++	Q58K/T451K
18	++	L63R/E89K/K226E/E440K/T451K
19	++	E89K
20	++	L63R/E88R/E89K
21	++	L63R/E88R/T451K
22	++	E88R/K470E
23	+	Q58K/E88R/E89K/K226E/E440K
24	+	Q58K/E88R/K199E/K225A/K226E
25	+	Q58K/E88R/K306A/K470E
26	+	Q58K/E440K/K470E
27	+	Q58K/L63R/E89K
28	+	L63R/K199E/R297G/K375E
29	+++	K52E/P127K/S181T/I462K
30	+++	K52E/P127K/S140A/S181T/Y238L
31	+++	P127K/S181T/Y238L/V372I
32	+++	Y238L/N404K/I462K
33	+++	K52E/P127K/S140A/S181T/I462K
34	+++	P127K/S181T/I207R
35	+++	P127K/Y238L/V372I/I462K
36	+++	K52E/Y238L/N404K/I462K
37	+++	P127K/L293E/N404K/I462K
38	+++	S140A/Y238L/V372I/I462K
39	+++	N404K
40	+++	P127K/Y238L/L293E/I462K
41	+++	S181T/Y238L/K299P/N404K
42	+++	Y238L/K299P/N404K
43	+++	K52E/S140A/S181T/Y238L/C276G/L293E/N404K

44	+++	P127K/Y238L/L293E/K299P/N404K
45	+++	P127K
46	+++	V372I/I462K
47	+++	K52E/P127K/N404K
48	+++	K52E/N404K/I462K
49	+++	I462K
50	+++	S181T/Y238L/I462K
51	+++	S181T/I462K
52	++	P127K/S181T/C276G
53	++	S181T/I207R/Y238L
54	++	S140A/C276G/L293E/N404K
55	++	V372I
56	++	P127K/Y238L/L293E/K299P/V372I/ N404K
57	++	P127K/I462K
58	++	P127K/S140A/K299P/V372I/I462K
59	++	P127K/S140A/C276G
60	++	K52E/P127K/C276G/V372I/I462K
61	++	Y238L/L293E/V372I
62	++	C276G/L293E/I462K
63	++	S181T/I207R/Y238L/V372I
64	++	K52E/S181T
65	++	S140A/V285A/L293E/N404K
66	++	K52E/S140A/I207R/K299P/V372I/N4 04K/I462K
67	++	S140T
68	++	K52E/S140A/S181T/C276G/K299P/N 404K/I462K
69	++	P127K/L293E
70	++	P127K/S140A/Y238L
71	++	K52E/S181T/Y238L/K299P/N404K
72	++	S181T/Y238L/C276G
73	++	K52E/L293E/K299P/N404K/I462K
74	++	L293E/V372I

75	+	S140A/K299P/V372I/N404K/I462K
76	+	P127K/L293E/V372I/I462K
77	+	P127K/S181T/N404K
78	+	S181T/C276G
79	+	K52E/S181T/Y238L/C276G
80	+	K52E/S181T/L293E
81	+	C276G/N404K
82	+	K52E/I207R/Y238L/L293E/K299P/N404K/I462K
83	+	Y238L/L293E/K299P/V372I/I462K
84	+	K299P/V372I/I462K
85	+	P127K/I207R/Y238L/V372I
86	+	K52E/C276G/K299P/N404K
87	+	S181T/I207R/Y238L/C276G/L293E/V372I/N404K
88	+	K299P/N404K/I462K
89	+	S181T/L293E
90	+	S140A/V372I
91	+	K52E/S140A/Y238L/C276G/K299P/V372I/N404K
Уровни активности определяли относительно эталонного полипептида SEQ ID NO: 6 Улучшения активности определяют следующим образом: +++ = >2 ++ = 1,5–2 + = 1,25–1,5		
Таблица 4.3 Улучшения активности вариантов лигазы Относительно SEQ ID NO: 32		
Вариант #	Улучшение активности	Замены аминокислот Относительно SEQ ID NO: 32
92	+++	P127K/Y238L
93	+++	P127K
94	+++	P127K/D385E
95	+++	E89K/P127K
96	+++	Y238L/K306A/V372I
97	+++	E89K/P127K/Y238L/K306A

98	+++	N404K
99	++	V372I
100	++	P127K/K306A
101	++	P127K/V177A/Y238L/L293P/K306A
102	++	K306A
103	+	E89K
104	+++	Q19K
105	++	Y238L
106	++	P127K/Q133H/Y238L/K375R
107	++	Q19K/Y238L
108	+	R297S
109	+	Q19K/P127K/K306A
110	+	Q19K/P127K/K199S
111	+++	F247K/D373A/E427K/E438D
112	+++	D176G/V250S/D373A/E438D/D480S
113	+++	A244S/F247K
114	++	A244S/V250S/E438D
115	++	D176G/A244S/F247K/D373A/E438D
116	++	A244S/E438D
117	++	A244S/F247K/V250S
118	++	A244S
119	+	E438D
Уровни активности определяли относительно эталонного полипептида SEQ ID NO: 32 Улучшения активности определяют следующим образом: +++ = >2 ++ = 1,5–2 + = 1,25–1,5		

ПРИМЕР 5

Капиллярный электрофорез Анализ лигирования

Анализ лигирования с помощью капиллярного электрофореза разработали, чтобы обеспечить прямое измерение продуктов одно и двойного лигирования. Для данного анализа не требуется флуоресцентно меченых субстратов, и он обеспечивает гибкость для использования широко используемых адаптеров и вставок с соответствующей длиной.

Были разработаны и синтезированы короткие, немеченые фрагменты адаптера и вставки с последовательностью, идентичной тем, что используют в планшетном анализе

определения. Меченых ^{32}P фрагмент двухцепочечной 50-мерной ДНК «50-мерная вставка» содержал два одноцепочечных очищенных с помощью HPLC синтетических олигонуклеотида (Integrated DNA Technologies) (SEQ ID NO: 11 и SEQ ID NO: 12), и его получали с помощью отжига этих двух олигонуклеотидов в 1х буфере для отжига (10 mM Tris pH 7,5, 50 mM NaCl, 1 mM EDTA). Полученная двухцепочечная 50-мерная вставка имеет 3' выступы с одним основанием дезоксиаденином и 5' концы с монофосфатом на обоих концах молекулы, и внутри мечена ^{32}P , присоединенным к фосфатному остову. Двухцепочечную молекулу «20-мерный адаптер», содержащую два одноцепочечных очищенных с помощью HPLC олигонуклеотидов (Integrated DNA technologies) (SEQ ID NO: 13 и SEQ ID NO: 14), также получали путем отжига в 1х буфере для отжига (10 mM Tris pH 7,5, 50 mM NaCl, 1 mM EDTA). Полученный 20-мерный адаптерный дуплекс имел защищенный фосфоротиоатом 5' дезокситимиридиновый выступ и 5' фосфат на совместимом с лигированием конце.

С помощью ПЦР амплификации сконструировали набор более длинных определенных вставок по 160 п.о. Для амплификации фрагментов по 160 п.о. с использованием в качестве матрицы фаговой последовательности, кодирующей T4 ДНК-лигазу (SEQ ID NO: 1), в отдельных реакциях использовали четыре набора пар праймеров прямой/обратной ПЦР, оканчивающихся нуклеотидами 5' аденозином (праймеры SEQ ID NO: 15 и SEQ ID NO: 16; продукт SEQ ID NO: 17), 5' цитозином (SEQ ID NO: 18 и SEQ ID NO: 19; продукт SEQ ID NO: 20), 5' гуанозином (SEQ ID NO: 21 и SEQ ID NO: 22; продукт SEQ ID NO: 23) и 5' тимидином (SEQ ID NO: 24 и SEQ ID NO: 25; продукт SEQ ID NO: 26). Эти продукты имеют идентичные внутренние последовательности ДНК за исключением оснований на 5' концах, переданных праймером. Затем эквимольные количества этих включений объединяли и обрабатывали с помощью модуля NEB Next UltraTM II для концевого восстановления/dA-наращивания с использованием инструкций производителя для добавления 5' фосфатов и 3' дезоксиаденозиновых выступов для их подготовки к лигированию.

Продукты очищали с использованием набора для очистки Wizard® SV ПЦР (Promega) и элюировали в воде.

Реакции лигирования с короткими 50/20-мерными субстратами проводили в объеме 80 мкл в 1х буфере для лигирования (66 mM Tris, pH 7,5, 10 mM MgCl_2 , 1 mM DTT) и с низкими концентрациями субстрата для лигирования (2 нМ 50-мерной вставки, 10 нМ 20-мерного адаптера). Для реакций с набором более длинных 160-мерных субстратов использовали 2 нМ субстрата и 20 нМ шпилечного адаптера NEB (SEQ ID NO: 15). В реакционную смесь добавляли НТП-очищенную лигазу (10 мкл), и реакционные смеси инкубировали в течение 4 часов при 20°C с последующими 16 часами при 4°C.

Реакции останавливали путем добавления 40 мкл 3х останавливающего раствора (45 mM EDTA, 0,6 мг/мл протеиназы К). Остановленные реакционные смеси инкубировали при 50°C в течение 1 ч для протеолиза ДНК-лигазы, которая препятствует последующему электрофорезу из-за сдвига электрофоретической подвижности. Затем 100

мкл подвергнутой протеолизу реакционной смеси загружали на планшет для очистки Montage SEQ96 (EMD Millipore), и для концентрирования образца использовали разрежение. Затем образец промывали 100 мкл буфера Low-TE (2 mM Tris pH 7,5, 0,5 mM EDTA), и промокали поверхность фильтра. Для элюирования в лунки добавляли ДНК-субстраты и продукты лигирования, 25 мкл буфера Low-TE, планшет встряхивали в течение 10 минут при комнатной температуре, и для извлечения элюата в 384-луночный планшет BioRad Hardshell® использовали жидкостной манипулятор Biomek NX (Beckman Coulter) или 12-канальную пипетку. Реакционные смеси анализировали с использованием прибора для капиллярного электрофореза Caliper (теперь Perkin Elmer) Labchip GX с использованием высокочувствительного анализа ДНК согласно инструкциям производителя.

Преобразование рассчитывали путем сравнения молярных концентраций продуктов с нелигированными вставками, с одним лигированием и двойным лигированием. Для коротких вставок регистрировали долю продукта с двойным лигированием для родительского контроля и вариантов для использования при расчете улучшений активности. Для набора 160-мерных вставок и шпилечного адаптера NEB регистрировали долю общих преобразованных частиц (продуктов с одно и двойным лигированием) для использования при расчете улучшений активности.

Таблица 5.1 Улучшения активности вариантов лигазы Относительно SEQ ID NO: 32		
Вариант #	Улучшение активности	Замены аминокислот Относительно SEQ ID NO: 32
120	++	F233A
121	++	N149R
122	++	Y453R
123	++	D371W
124	++	L63T
125	++	Y453G
126	+	E240P
127	+	Y453G
128	+	V454L
129	+	A237N
130	+	E240P
131	+	F60G
132	+	D371W
133	+	D385A
134	+	F60V
135	+	A237R

136	+	Y238L
137	+	Y453T
138	+	G184A
139	+	A56S
140	+	A461C
141	+	P51R
142	+	D371V
143	+	Y453L
144	+	F233T
145	+	D385W
146	+	K446R
147	+++	D452P
148	+++	D448A
149	+++	D448P
150	+++	D452V
151	++	A86R
152	++	Y314V
153	++	K199T
154	++	E438F
155	++	D329G
156	++	T485G
157	++	D329G
158	++	K451G
159	++	D373A
160	++	E466G
161	++	E427R
162	++	D329L
163	++	I207V
164	++	C439S
165	++	D373G
166	++	D476A
167	++	E466P
168	+	E427L
169	+	E438G

170	+	L174P
171	+	I207Q
172	+	T485Y
173	+	E438D
174	+	D448G
Уровни активности определяли относительно эталонного полипептида SEQ ID NO: 32 Улучшения активности определяют следующим образом: +++ = >1,5 ++ = 1,3–1,5 + = 1,2–1,3		
Таблица 5.2 Улучшения активности вариантов лигазы Относительно SEQ ID NO: 32		
Вариант #	Улучшение активности	Замены аминокислот Относительно SEQ ID NO: 32
175	+	E321A
176	+	E321R
177	+	Q280L
178	+++	N7L
179	+++	K52G
180	+++	N404S
181	+++	F235R
182	+++	A237G
183	+++	I462Q
184	+++	E240P
185	+++	A405G
186	+++	G54E
187	+++	Y453L
188	+++	S242H
189	+++	D371G
190	++	V454A
191	++	S59M
192	++	Q17R
193	++	N241G
194	++	D371G
195	++	A183N

196	++	D452P
197	++	E483G
198	++	F74T
199	++	E483Q
200	++	K451G
201	++	A85T
202	+	F74G
203	+	K199G

Уровни активности определяли относительно эталонного полипептида SEQ ID NO: 32

Улучшения активности определяют следующим образом:

+++ = >2

++ = 1,5–2

+ = 1,2–1,5

Таблица 5.3 Улучшения активности вариантов лигазы Относительно SEQ ID NO: 6

Вариант #	Улучшение активности	Замены аминокислот Относительно SEQ ID NO: 6
204	+++	N149K
205	+++	D385K
206	+++	A413K
207	++	L231K
208	++	A183K
209	++	G184K
210	++	D232K
211	++	P386K
212	++	G13K
213	++	E186K
214	++	E240K
215	++	E89K
216	++	S11K
217	+	Y238K
218	+	F233K
219	+	S14K
220	+	Y453K

221	+	N185K
222	+	P239K
223	+	G54K
224	+	N7K
225	+	M62K
Уровни активности определяли относительно эталонного полипептида SEQ ID NO: 6 Улучшения активности определяют следующим образом: +++ = >2 ++ = 1,5–2 + = 1,25–1,5		
Таблица 5.4 Улучшения активности вариантов лигазы Относительно SEQ ID NO: 34		
Вариант #	Улучшение активности	Замены аминокислот Относительно SEQ ID NO: 34
226	+++	L63T/N149R/E240P/D371W/D452P
227	+++	Q19K/L63T/F233A/A237R/D371W/D452P
228	+++	E89K/F233A/A237R/E240P/D448A/Y453G/V454L
229	++	N149R/F233A/A237N/V454L
230	++	A86R/E89K/N149R/F233A/A237N/E240P
231	++	E89K/E240P/V454L
232	++	Q19K/A237N/Y453G
233	++	N149R/A237N/E240P
234	+	A86R/E89K/F233A/A237N/E240P/D448A
235	+	N149R/A237N/E240P/D329G/N404K/Y453G
236	+	F233A/A237N/D371W/N404K/D452P/V454L
237	+	F233A/A237R/N404K
238	+	L63T/E89K/D448A/D452P/Y453G
239	+	A86R/E89K/N149R/F233A/A237R/Y314V/D452P

240	+	L63T/F233A/E240P/D452P/V454L
Уровни активности определяли относительно эталонного полипептида SEQ ID NO: 34. Улучшения активности определяют следующим образом: +++ = >1,5 ++ = 1,3–1,5 + = 1,2–1,3		
Таблица 5.5 Улучшения активности вариантов лигазы Относительно SEQ ID NO: 38		
Вариант #	Улучшение активности	Замены аминокислот Относительно SEQ ID NO: 38
241	+++	G13K/A183K/D232K/D329L/Y453G/E466G
242	+++	G13K/A183K/D232K/P386K/K451G
243	+++	G13K/E89K/A183K/D232K/P386K/K451G
244	+++	D385K/Y453R/E466G
245	+++	A183K/D373A/P386K
246	++	A183K/L231K/E427R/E466G
247	++	A183K/I207V/P386K/E427R/Y453G
248	++	G13K/E89K/A183K/L231K
249	++	G13K/D232K/D385K/K451G
250	++	N149R/A183K
251	++	A183K
252	++	A183K/A413K/E427R
253	++	E89K/A183K/D329G/K451G/Y453R
254	++	N149R/A183K
255	++	A183K/D385K
256	++	A183K/D385K/E427R
257	++	A183K/L231K/D385K/E427R
258	++	A183K/E427R/K451G
259	++	A183K/I207V/P386K
260	++	A183K/I207V/C439S
261	++	A183K/L231K/D373G

Уровни активности определяли относительно эталонного полипептида SEQ ID NO: 38 и определяют в виде:

+++ = >1,5

++ = 1,3–1,5

ПРИМЕР 6

Анализ лигирования с помощью капиллярного электрофореза на препаратах в масштабе встряхиваемой колбы

Варианты лигаз экспрессировали и очищали в масштабе встряхиваемой колбы, как описано в примере 3. Анализ лигирования с помощью капиллярного электрофореза проводили в 1x буфере для лигирования (66 mM Tris, pH 7,5, 10 mM MgCl₂, 1 mM DTT), как описано в примере 5, с использованием 875 нМ SF-очищенной ДНК-лигазы, 1 нМ 160-мерной A-образной вставки (см. пример 5), и 200 нМ коммерчески доступного Y адаптера (Illumina). Данный Y адаптер состоит из двух очищенных с помощью HPLC олигонуклеотидов (SEQ ID NO: 17 и SEQ ID NO: 18), отождествленных как описано в примере 4. Реакции останавливали путем добавления EDTA до 15 mM и протеиназы K до 0,4 мг/мл (в итоге), и инкубировали при 50°C в течение 1 часа для протеолиза ДНК-лигазы, которая препятствует последующему электрофорезу. Подвергнутые протеолизу реакционные смеси лигирования затем очищали и заменяли буфер с использованием 96-луночного набора Zymo DNA Clean & Concentrator и элюировали в 25 мкл буфера с низким содержанием TE (2 mM Tris, 0,5 mM EDTA). Реакционные смеси анализировали с использованием прибора для капиллярного электрофореза Perkin Elmer Labchip GX с использованием анализа высокой чувствительности ДНК (Perkin Elmer) в соответствии с инструкциями производителя.

Преобразование вставки в продукты двойного лигирования рассчитывали путем сравнения молярных концентраций продуктов с нелигированными вставками, одним лигированием и двойным лигированием. долю вставок, преобразованных в продукт с двойным лигированием регистрировали для вариантов, как указано в таблице 6.1, ниже.

Таблица 6.1 Результаты Анализа лигирования (Замены аминокислот Относительно SEQ ID NO: 6)

SEQ ID NO:	Замены	Двухстороннее Лигирование (% Преобразование)
SEQ ID NO: 2	(Дикого типа)	+
SEQ ID NO: 32	E88R/K225A/E440K/T451K	+++
SEQ ID NO: 34	E88R/P127K/K225A/E440K/T451K	+++
SEQ ID NO: 36	L63T/E88R/P127K/N149R/K225A/E240P/D371 W/E440K/T451K/D452P	+++

SEQ ID NO: 38	Q19K/L63T/E88R/P127K/K225A/F233A/A237R/ D371W/E440K/T451K/D452P	+++
SEQ ID NO: 40	E88R/E89K/P127K/K225A/F233A/A237R/E240P /E440K/D448A/T451K/Y453G/V454L	+++
SEQ ID NO: 42	E88R/P127K/N149R/K225A/F233A/A237N/E44 0K/T451K/V454L/	+++
SEQ ID NO: 44	A86R/E88R/E89K/P127K/N149R/K225A/F233A/ A237N/E240P/E440K/T451K	++
SEQ ID NO: 46	Q19K/L63T/E88R/P127K/A183K/K225A/L231K /F233A/A237R/D371W/E427R/E440K/T451K/D 452P/E466G	++
SEQ ID NO: 48	Q19K/L63T/E88R/P127K/K225A/F233A/A237R/ D371W/D385K/E440K/T451K/D452P/Y453R/E4 66G	+++
SEQ ID NO: 50	G13K/Q19K/L63T/E88R/P127K/A183K/K225A/ D232K/F233A/A237R/D371W/P386K/E440K/T4 51G/D452P	++
SEQ ID NO: 52	Q19K/L63T/E88R/P127K/A183K/K225A/F233A/ A237R/D371W/D373A/P386K/E440K/T451K/D4 52P	++
SEQ ID NO: 54	G13K/Q19K/L63T/E88R/P127K/A183K/K225A/ D232K/F233A/A237R/D329L/D371W/E440K/T4 51K/D452P/Y453G/E466G	++
значения преобразования, указанные выше, соответствуют: +++ = >80% продукта с двойным лигированием ++ = >65% продукта с двойным лигированием + = ~50% продукта с двойным лигированием		

ПРИМЕР 7

Анализ димеризации адаптера на препаратах в масштабе встряхиваемой колбы

NGS адаптеры, разработанные для использования на платформе секвенирования Illumina®, имеют дезокситимидиновые 3' выступы, совместимые с лигированием с дезоксиаденозиновыми 3' выступами, присутствующими на А-образных фрагментах вставки. Т-образные адаптеры не эффективно лигируют друг с другом вследствие избирательности Т4 ДНК-лигазы дикого типа к некомплементарным концам ДНК. Димеризация адаптера будет происходить в результате экстремальных условий лигирования, включая длительные периоды инкубации, высокие концентрации адаптера

или высокие концентрации краудинг–агента. Важно, нуклеазные контаминанты в реакции лигирования могут удалять выступы на концах адаптера, что приводит к субстратам с тупыми концами, которые совместимы с самолигированием.

Для тестирования избирательности иллюстративной вариантной Т4 ДНК–лигазы разработали стратегию очистки для уменьшения нуклеазного загрязнения, которое могло бы помешать прямому наблюдению за любой присущей адаптеру димеризации активности.

Градиентная Очистка Т4 ДНК–Лигаза из Лизатов во Встряхиваемой колбе

Лигаза экспрессировали во встряхиваемых колбах в соответствии со способом в примере 3 и ресуспендировали в 50 мМ Tris HCl pH 7,5 перед лизисом. Данную суспензию клеток охлаждали в ледяной ванне и лизировали с использованием устройства разрушения клеток Microfluidizer (Microfluidics M- 110L). Неочищенные лизаты дополняли 500 мМ NaCl и 30 мМ имидазола перед очисткой путем центрифугирования (16000 об/мин в течение 60 мин при 4°C), а затем супернатанты фильтровали через 0,2 мкм PES мембрану для дополнительной очистки лизатов.

Лизаты очищали с использованием системы AKTA Start и 1 мл HisTrap FF колонки (GE Healthcare) с использованием протокола никелевого NTA и элюирования с градиентом имидазола (рабочие параметры приведены в таблице 7.1). SF промывочный буфер состоял из 50 мМ Tris HCl pH 7,5, 500 мМ NaCl, 30 мМ имидазола и 1 мМ DTT. SF элюирующий буфер состоял из 50 мМ Tris HCl pH 7,5, 500 мМ NaCl, 300 мМ имидазола и 1 мМ DTT.

Таблица 7.1 Параметры очистки	
Параметр	Объем
Объем Колонки	1 мл
Скорость потока	1 мл/мин
Предел давления	0,4 МПа
Объем образца	~35 млс
Уравновешивающий объем	5 объемов колонки (CV) = 25 млс
Промывающий несвязанный объем	30 CV=30 млс
Элюирование	Элюирующий буфер с градиентом 0–50%
Объем элюирования	20 CV=20 млс
Объем фракций	1 млс
RE- уравновешивающий объем	5 CV=25 млс

Четыре наиболее концентрированные фракции идентифицировали с помощью УФ поглощения (A280) и диализировали в течение ночи в 1х буфере хранения лигазы (10 мМ Tris HCl pH 7,5, 50 мМ KCl, 1 мМ DTT, 0,1 мМ EDTA) в течение ночи в кассете для диализа 10K Slide-A-Lyzer™ (Термо Fisher) для замены буфера. концентрации лигазы в препаратах измеряли с помощью анализа Бредфорда и поглощения при 280 нм.

Крупномасштабную очистку из ферментативного осадка проводили с использованием таких же параметров хроматографии для экспериментов в примерах 10–14. в этом случае наиболее концентрированные фракции из никель–сефарозного элюата объединяли и пропускали в проточном режиме через два последовательных нано (1 мл) фильтра с заряженной мембраной STIC–PA (Sartorius) для удаления загрязняющих нуклеиновых кислот.

Анализ димеризации адаптера проводили в 1х буфере для лигирования (66 mM Tris, pH 7,5, 10 mM MgCl₂, 1 mM DTT), как описано в примере 5, с использованием 875 nM SF–очищенной ДНК–лигазы и 1000 nM блокированного адаптера, состоящего из двух олигонуклеотидов (SEQ ID NO: 13 и SEQ ID NO: 16, отождествленных, как описано в примере 4). Этот блокированный дуплекс 20–мерных адаптеров имеет один пригодный для лигирования конец с 3' дезокситимидиновым выступом и 5' фосфатом и второй тупой непригодный для лигирования конец без 5' фосфата и аминокислотного 3' конца, который не может служить в качестве субстрата для лигирования. Лигирование проводили при 20°C в течение 16 часов, чтобы обеспечить образование димера. Реакции останавливали путем добавления EDTA до 15 mM и протеиназы K до 0,2 мг/мл (в итоге), и инкубировали при 50°C в течение 2 часов для протеолиза ДНК–лигазы. Реакционные смеси прямо анализировали с использованием прибора для капиллярного электрофореза Perkin Elmer Labchip GX с использованием анализа ДНК 1k (Perkin Elmer), согласно инструкциям производителя.

Процентное преобразование с димеризацией адаптера рассчитывали путем сравнения молярных концентраций продуктов с нелигированным адаптером и димером адаптера, измеренных посредством капиллярного электрофореза. доля блокированного 20–мерного адаптера, преобразованного в продукт с двойным лигированием, показана в таблице 7.2. Образец WT T4 ДНК–лигазы, полученный в соответствии с протоколом из примера 3, и содержащий, как известно, нуклеазный загрязнитель, включили в качестве положительного контроля для образования димера адаптера, а «ультра–чистый» коммерческий препарат T4 ДНК–лигазы (Enzymatics) использовали в качестве ссылки на активность димеризации WT лигазы.

Таблица 7.2 Процентные значения Преобразования Димеризации адаптера Лигаза		
образец Лигаза	Способ Получения	Димеризация адаптера (% Преобразование)
T4 ДНК–лигаза WT	коммерческий «ультра чистый» препарат	+
T4 ДНК–лигаза WT	Способ Примера 3	++
SEQ ID NO: 38	Способ Примера 3	+++
SEQ ID NO: 38	Способ Примера 7	+

указанные выше значения преобразования соответствуют:

+++ = >10% преобразование

++ = 2–10% преобразование

+ = <2% преобразование

ПРИМЕР 8

Динамика Лигирования ДНК

Анализ динамики реакции лигирования проводили в 1х буфере для лигирования, дополненном краудинг–агентом (66 mM Tris, pH 7,5, 10 mM MgCl₂, 1 mM DTT, 6% PEG₆₀₀₀ (м/о)), как описано в примере 5, с использованием либо итоговой концентрации 855 нМ вариантной лигазы (SEQ ID NO: 38), очищенной в соответствии со способом, описанным в примере 7, либо коммерческого препарата Т4 ДНК–лигазы дикого типа («Ультра–Чистая» Т4 ДНК–лигаза (Быстрая), Enzymatics). ДНК–субстраты содержали 1 нМ А–образного 160–мерного ПЦР продукта, описанного в примере 6, и 40 нМ адаптера, состоящего из двух олигонуклеотидов (SEQ ID NO: 17 и SEQ ID NO: 18, отождествленных, как описано в примере 4). Реакции инициировали путем добавления лигазы и останавливали спустя 2, 5, 15 и 30 минут путем добавления EDTA до итоговой концентрации 15 mM. Протеиназу К добавляли до итоговой концентрации 0,2 мг/мл, и образцы подвергали протеолизу в течение 2 ч при 50°C, с последующей очисткой образцов с использованием набора для очистки планшета с лунками Zymo ZR–96 DNA Clean & Concentrator–5 (Zymo Research). Образцы элюировали в 25 мкл EB буфера и анализировали с помощью капиллярного электрофореза в соответствии со способом, описанным в примере 6. Преобразование в продукты с двухсторонним лигированием регистрировали для каждого момента времени в таблице 8.1.

Таблица 8.1 Данные Динамики Лигирования

SEQ ID NO	Описание	Двухстороннее Преобразование Лигирования (%)			
		2 мин	5 мин	15 мин	30 мин
SEQ ID NO: 2	WT Т4 ДНК–лигаза коммерческий препарат	+	+	++	++
SEQ ID NO: 38	Т4 вариант лигазы	+++++	+++++	+++++	+++++
+++++	90,1–95% продукт с двойным лигированием				
++++	85,1–90%				
+++	50,1–85%				
++	30,1–50%				

+

0–30%

ПРИМЕР 9

Температурный Профиль Активности

Анализ динамики реакции лигирования проводили в 1x буфере для лигирования (66 mM Tris, pH 7,5, 10 mM MgCl₂, 1 mM DTT), как описано в примере 5, с использованием либо итоговой концентрации 855 нМ вариантной лигазы (SEQ ID NO: 38), очищенной в соответствии со способом, описанным в примере 7, либо коммерческого препарата T4 ДНК–лигазы дикого типа («Ультра–Чистая» T4 ДНК–лигаза (Быстрая); Enzymatics).

ДНК–субстраты содержали 1 нМ А–образного 160–мерного ПЦР продукта, описанного в примере 6, и 200 нМ адаптера, состоящего из двух олигонуклеотидов (SEQ ID NO: 13 и SEQ ID NO: 16, отождествленных, как описано в примере 4). Реакции инициировали с добавлением лигазы и инкубировали при 16°C, 20°C, 30°C или 37°C в течение 30 минут, затем останавливали путем добавления EDTA до итоговой концентрации 15 mM. Протеиназы К добавляли до итоговой концентрации 0,2 мг/мл, и образцы подвергали протеолизу в течение 1 часа при 50°C с последующей очисткой образцов с использованием набора для очистки планшета с лунками Zymo ZR–96 DNA Clean & Concentrator–5 (Zymo Research). Образцы элюировали в 25 мкл EB буфера и анализировали путем капиллярного электрофореза в соответствии со способом, описанным в примере 6. Преобразование в продукты с двухсторонним лигированием отображено на графике для каждого температурного условия на фиг. 1.

ПРИМЕР 10

pH Профиль Активности

Реакции лигирования проводили в 1x буфере для лигирования, дополненном краудинг–агентом (66 mM Tris, pH 7,5, 10 mM MgCl₂, 1 mM DTT, 6% PEG₆₀₀₀ (м/о)), как описано в примере 5, с использованием либо итоговой концентрации 855 нМ вариантной лигазы с), очищенной в соответствии со способом, описанным в примере 7, либо коммерческого препарата T4 ДНК–лигазы дикого типа («Ультра–Чистая» T4 ДНК–лигаза (Быстрая); Enzymatics). ДНК–субстраты содержали 1 нМ А–образного 160–мерного ПЦР продукта, описанного в примере 6, и 40 нМ адаптера, состоящего из двух олигонуклеотидов (SEQ ID NO: 17 и SEQ ID NO: 18, отождествленных как описано в примере 4). Реакции инициировали путем добавления лигазы и останавливали спустя 15 минут путем добавления шариков SPRI для ПЦР HiPrepTM (MagBio Genomics, 80 мкл в 100 мкл реакционной смеси). Образцы элюировали в 25 мкл EB буфера и анализировали путем капиллярного электрофореза в соответствии со способом, описанным в примере 6. Преобразование в продукты с двухсторонним лигированием отражено на графике на фиг. 2.

ПРИМЕР 11

Сдвиг последовательности лигирования

Анализ динамики реакции лигирования проводили в 1x буфере для лигирования, дополненном краудинг–агентом (66 mM Tris, pH 7,5, 10 mM MgCl₂, 1 mM DTT, 6% PEG₆₀₀₀

(м/о)), как описано в примере 5, с использованием либо итоговой концентрации 855 нМ вариантной лигазы (SEQ ID NO: 38), очищенной в соответствии со способом, описанным в примере 7, либо коммерческого препарата Т4 ДНК-лигазы дикого типа («Ультра-Чистая» Т4 ДНК-лигаза (Быстрая); Enzymatics). ДНК-субстраты, содержащие 50 нг А-образного 160-мерного продукта ПЦР, описанного в примере 6, за исключением отдельных продуктов ПЦР, не объединяли, но лигировали по отдельности.

С помощью ПЦР амплификации сконструировали набор определенных вставок по 160 п.о. с варьирующими последовательностями на их 5' концах. Для амплификации фрагментов по 160 п.о. с использованием в качестве матрицы фаговой последовательности, кодирующей Т4 ДНК-лигазу (SEQ ID NO: 1), в отдельных реакциях использовали четыре набора пар праймеров прямой/обратной ПЦР, оканчивающихся нуклеотидами 5' аденозином (праймеры SEQ ID NO: 15 и SEQ ID NO: 16; продукт SEQ ID NO: 17), 5' цитозином (SEQ ID NO: 18 и SEQ ID NO: 19; продукт SEQ ID NO: 20), 5' гуанозином (SEQ ID NO: 21 и SEQ ID NO: 22; продукт SEQ ID NO: 23) и 5' тимидином (SEQ ID NO: 24 и SEQ ID NO: 25; продукт SEQ ID NO: 26). Эти продукты имеют идентичные внутренние последовательности ДНК за исключением оснований на 5' концах, которые переданы праймером. Эти вставки затем отдельно обрабатывали с помощью модуля NEB Next Ultra™ II для концевой реставрации/дА-наращивания с использованием инструкций производителя для добавления 5' фосфатов и 3' дезоксиаденозиновых выступов для подготовки их для лигирования. Полученные субстраты отдельно очищали в соответствии с инструкциями производителя с использованием спин-колонок DNA Clean and Concentrator-5 (Zymo Research). Лигирование проводили с использованием 40 нМ адаптера, состоящего из двух олигонуклеотидов (SEQ ID NO: 17 и SEQ ID NO: 18, отождествленных, как описано в примере 4). Реакции инициировали путем добавления лигазы и останавливали спустя 15 минут путем добавления шариков SPRI для ПЦР HiPrep™ (MagBio Genomics; 80 мкл в 100 мкл реакционной смеси) и готовили в соответствии с инструкциями производителя. Образцы элюировали в 25 мкл не содержащего нуклеазу ТЕ буфера (10 мМ Tris, 1 мМ EDTA) и анализировали путем капиллярного электрофореза с использованием способа, описанного в примере 6. Преобразование в продукты с двухсторонним лигированием отражено на графике на фиг. 3.

ПРИМЕР 12

Преобразование в качестве Функции Концентрация лигазы

Реакции лигирования проводили в 1х буфере для лигирования, дополненном краудинг-агентом (66 мМ Tris, pH 7,5, 10 мМ MgCl₂, 1 мМ DTT, 6% PEG₆₀₀₀ (м/о)), как описано в примере 5, с использованием диапазона итоговых концентраций вариантной лигазы от 2 до 4000 нМ (SEQ ID NO: 38), очищенной в соответствии со способом, описанным в примере 7, или коммерческого препарата Т4 ДНК-лигазы дикого типа («Ультра-Чистая» Т4 ДНК-лигаза (Быстрая); Enzymatics). ДНК-субстраты содержали 10 нг А-образного 160-мерного ПЦР продукта, описанного в примере 6, и 40 нМ или 200 нМ

адаптера, состоящего из двух олигонуклеотидов (SEQ ID NO: 17 и SEQ ID NO: 18, отождествленных, как описано в примере 4). Реакции инициировали путем добавления лигазы и останавливали спустя 15 минут путем добавления шариков SPRI для ПЦР HiPrep™ (MagBio Genomics; 80 мкл в 100 мкл реакционной смеси). Образцы элюировали в 25 мкл EB буфера и анализировали путем капиллярного электрофореза с использованием способа, описанного в примере 6. Преобразование в продукты с двухсторонним лигированием отражено на графике на фиг. 4.

ПРИМЕР 13

Преобразование в качестве Функции концентрации адаптеров

Реакции лигирования проводили в 1x буфере для лигирования, дополненном краудинг-агентом (66 mM Tris, pH 7,5, 10 mM MgCl₂, 1 mM DTT, 6% PEG₆₀₀₀ (м/о)), как описано в примере 5, с использованием итоговой концентрации 855 нМ вариантной лигазы (SEQ ID NO: 38), очищенной в соответствии со способом, описанным в примере 7, либо коммерческого препарата T4 ДНК-лигазы дикого типа («Ультра-Чистая» T4 ДНК-лигаза (Быстрая); Enzymatics). ДНК-субстраты содержали 10 нг или 100 нг А-образного 160-мерного ПЦР продукта, описанного в примере 6, и 3–400 нМ титра адаптера, состоящего из двух олигонуклеотидов (SEQ ID NO: 17 и SEQ ID NO: 18, отождествленных как описано в примере 4). Реакции инициировали путем добавления лигазы и останавливали спустя 15 минут путем добавления шариков SPRI для ПЦР HiPrep™ (MagBio Genomics; 80 мкл в 100 мкл реакционной смеси), затем очищали в соответствии с протоколом производителя. Образцы элюировали в 25 мкл EB буфера и анализировали путем капиллярного электрофореза с использованием способа, описанного в примере 6. Процентное преобразование в продукты с двухсторонним лигированием отражено на графике на фиг. 5.

ПРИМЕР 14

Преобразование Внуклеточных ДНК-субстратов

Внуклеточные образцы ДНК, выделенной из человеческой сыворотки (BioChain), получали для лигирования с использованием модуля концевое-восстановления/А-наращивания NEB Next® Ultra II™ в объеме 60 мкл в соответствии с инструкциями производителя, с использованием 10 нг образца ДНК на реакционную смесь. Модуль лигирования добавляли прямо в продукт реакции концевое-восстановления/А-наращивания таким образом, чтобы итоговая концентрация дополнительных реакционных компонентов составляла 7 mM Tris pH 7,5, 9 mM MgCl₂, 1 mM DTT и 10% PEG 6000. Адаптер, состоящий из двух олигонуклеотидов (SEQ ID NO: 17 и SEQ ID NO: 18, отождествленных, как описано в примере 4), добавляли до итоговой концентрации 40 нМ или 200 нМ. Реакции проводили с итоговой концентрацией вариантной лигазы 855 нМ (SEQ ID NO: 38), очищенной в соответствии со способом, описанным в примере 7, либо коммерческого препарата T4 ДНК-лигазы дикого типа («Ультра-Чистая» T4 ДНК-лигаза (Быстрая); Enzymatics), для итогового общего объема 93,5 мкл. Реакции инициировали путем добавления лигазы и останавливали спустя 15 минут путем добавления шариков

SPRI для ПЦР HiPrep™ (MagBio Genomics; 75 мкл в реакционную смесь 93,5 мкл), затем очищали в соответствии с протоколом производителя. Образцы элюировали в 25 мкл ТЕ буфера (10 mM Tris, pH 7,5, 1 mM EDTA) и анализировали путем капиллярного электрофореза с использованием способа, описанного в примере 6. Процентное преобразование в продукты с двухсторонним лигированием показано на фиг. 6.

Несмотря на то, что изобретение было описано со ссылкой на конкретные варианты осуществления, могут быть сделаны разные изменения, и могут быть заменены эквиваленты для адаптации к конкретной ситуации, материалу, композиции, способу, стадии или стадиям способа, достигая тем самым преимуществ изобретения без отклонения от заявленного объема.

Для всех целей в США, все и каждая публикация и патентный документ, процитированный в этом раскрытии, включены в данный документ посредством ссылки, как если бы было конкретно и отдельно указано, что каждая из такой публикации или документа включена в данный документ посредством ссылки. Цитирование публикаций и патентных документов не предназначено для указания, что любой такой документ относится к предшествующему уровню техники и не является признанием его содержания или даты.

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Сконструированная лигаза, содержащая полипептидную последовательность, имеющую идентичность последовательности по меньшей мере 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более с эталонной последовательностью SEQ ID NO: 2, 6, 32, 34 и/или 38 или ее функциональным фрагментом, при этом сконструированная лигаза содержит по меньшей мере одну замену или набор замен в ее полипептидной последовательности, и при этом положения аминокислот полипептидной последовательности нумеруют со ссылкой на SEQ ID NO: 2, 6, 32, 34 или 38.

2. Сконструированная лигаза по п. 1, в которой по меньшей мере одну замену или набор замен выбирают из 52/56/404, 52/56/404/412, 127/207, 127/213, 127/213/276/339, 140/181/234, 165/181/299, 165/181/281/299, 238/241/404/412/462 и 462 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 2.

3. Сконструированная лигаза по п. 1, в которой по меньшей мере одна замена или набор замен содержит замены в положениях аминокислот, выбираемых из 52/127/140/181/462, 52/127/140/181/238, 52/127/181/462, 52/127/276/372/462, 52/127/404, 52/140/181/238/276/293/404, 52/140/181/276/299/404/462, 52/140/207/299/372/404/462, 52/140/238/276/299/372/404, 52/181, 52/181/238/276, 52/181/238/299/404, 52/181/293, 52/207/238/293/299/404/462, 52/276/299/404, 52/238/404/462, 52/293/299/404/462, 52/404/462, 58/63/89, 58/88/89/226/440, 58/88/199/225/226, 58/88/226/306, 58/88/306/470, 58/440/470, 58/451, 63/88/89, 63/88/451, 63/89/226/440/451, 63/89/451, 63/199/297/375, 88/225/440/451, 88/306/440/451, 88/470, 89, 127, 127/140/238, 127/140/276, 127/140/299/372/462, 127/181/207, 127/181/238/372, 127/181/276, 127/181/404, 127/207/238/372, 127/238/293/462, 127/238/293/299/372/404, 127/238/293/299/404, 127/238/372/462, 127/293, 127/293/372/462, 127/293/404/462, 127/462, 140/238/372/462, 140/276/293/404, 140/285/293/404, 140/299/372/404/462, 140/372, 140,181/207/238, 181/207/238/276/293/372/404, 181/207/238/372, 181/238/276, 181/238/299/404, 181,238/462, 181/276, 181/293, 181/462, 238/293/299/372/462, 238/293/372, 238/299/404, 238/404/462, 276/293/462, 276/404, 293/372, 299/372/462, 299/404/462, 372, 372/462, 404, 451 и 462 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6.

4. Сконструированная лигаза по п. 1, в которой по меньшей мере одно замена или набор замен содержит замены в положениях аминокислот, выбираемых из 19, 19/127/199, 19/127/306, 19/238, 89, 89/127, 89/127/238/306, 127, 127/133/238/375, 127/177/238/293/306, 127/238, 127/306, 127/385, 176/244/247/373/438, 176/250/373/438/480, 238, 238/306/372, 244, 244/247, 244/247/250, 244/250/438, 244/438, 247/373/427/438, 297, 306, 372, 404 и 438 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32.

5. Сконструированная лигаза по п. 1, в которой по меньшей мере одно замена или набор замен содержит замены в положениях аминокислот, выбираемых из 51, 56, 60, 63, 86, 149, 174, 184, 199, 207, 233, 237, 238, 240, 314, 329, 371, 373, 385, 427, 438, 439, 446,

448, 451, 452, 453, 454, 461, 466, 476 и 485 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32.

6. Сконструированная лигаза по п. 1, в которой по меньшей мере одна замена или набор замен содержит замены в положениях аминокислот, выбираемых из 7, 17, 52, 54, 59, 74, 85, 183, 199, 240, 241, 242, 280, 321, 235, 237, 371, 404, 405, 451, 452, 453, 454, 462 и 483 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 32.

7. Сконструированная лигаза по п. 1, в которой по меньшей мере одна замена или набор замен содержит замены в положениях аминокислот, выбираемых из 7, 11, 13, 14, 54, 62, 89, 149, 183, 184, 185, 186, 231, 232, 233, 238, 239, 240, 385, 386, 413 и 453 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6.

8. Сконструированная лигаза по п. 1, в которой по меньшей мере одна замена или набор замен содержит замены в положениях аминокислот, выбираемых из 19/63/233/237/371/452, 19/237/453, 63/89/448/452/453, 63/149/240/371/452, 63/233/240/452/454, 86/89/149/233/237/240, 86/89/149/233/237/314/452, 86/89/233/237/240/448, 89/233/237/240/448/453/454, 89/240/454, 149/233/237/454, 149/237/240, 149/237/240/329/404/453, 233/237/371/404/452/454 и 233/237/404 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 34.

9. Сконструированная лигаза по п. 1, в которой по меньшей мере одна замена или набор замен содержит замены в положениях аминокислот, выбираемых из 13/89/183/231, 13/89/183/232/386/451, 13/183/232/329/453/466, 13/183/232/386/451, 13/232/385/451, 89/183/329/451/453, 149/183, 183, 183/207/386, 183/207/386/427/453, 183/207/439, 183/231/373, 183/231/385/427, 183/231/427/466, 183/373/386, 183/385, 183/385/427, 183/413/427, 183/427/451 и 385/453/466 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 38.

10. Сконструированная лигаза по п. 1, в которой по меньшей мере одна замена или набор замен содержит замены в положениях аминокислот, выбираемых из 13/19/63/88/127/183/225/232/233/237/329/371/440/451/452/453/466, 13/19/63/88/127/183/225/232/233/237/371/386/440/451/452, 19/63/88/127/183/225/231/233/237/371/427/440/451/452/466, 19/63/88/127/183/225/233/237/371/373/386/440/451/452, 19/63/88/127/225/233/237/371/385/440/451/452/453/466, 19/63/88/127/225/233/237/371/440/451/452, 63/88/127/149/225/240/371/440/451/452, 86/88/89/127/149/225/233/237/240/440/451, 88/89/127/225/233/237/240/440/448/451/453/454, 88/127/149/225/233/237/440/451/454, 88/127/225/440/451 и 88/225/440/451 и/или любых их комбинаций, при этом положения аминокислот нумеруют со ссылкой на SEQ ID NO: 6.

11. Сконструированная лигаза по п. 1, причем указанная сконструированная лигаза содержит полипептидную последовательность, которая по меньшей мере на 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более идентична последовательности

по меньшей мере одного варианта сконструированной лигазы, представленной в таблице 4.1, 4.2, 4.3, 5.1, 5.2, 5.3, 5.4, 5.5 и/или 6.1.

12. Сконструированная лигаза по п. 1, причем указанной сконструированной лигазой является вариант сконструированной лигазы, представленной в таблице 4.1, 4.2, 4.3, 5.1, 5.2, 5.3, 5.4, 5.5 и/или 6.1.

13. Сконструированная лигаза по любому из пп. 1–12, при этом указанная лигаза обладает активностью ДНК–лигазы.

14. Сконструированная лигаза по любому из пп. 1–13, имеющая по меньшей мере одно улучшенное свойство по сравнению с Т4 ДНК–лигазой дикого типа.

15. Сконструированная лигаза по п. 14, имеющая по меньшей мере одно улучшенное свойство по сравнению с Т4 ДНК–лигазой дикого типа, при этом указанное улучшенное свойство выбирают из демонстрации большей активности при низких концентрациях ДНК–субстрата и получения меньшего количества димеров адаптеров.

16. Сконструированная лигаза по любому из пп. 1–15, при этом указанная сконструированная лигаза является более термостабильной, чем Т4 ДНК–лигаза дикого типа.

17. Сконструированная лигаза по любому из пп. 1–16, при этом указанная сконструированная лигаза демонстрирует более широкий диапазон стабильности pH, чем Т4 ДНК–лигаза дикого типа.

18. Сконструированная лигаза по любому из пп. 1–17, при этом указанный полипептид очищают.

19. Полинуклеотидная последовательность, кодирующая по меньшей мере одну сконструированную лигазу, по любому из пп. 1–18.

20. Полинуклеотидная последовательность, кодирующая по меньшей мере одну сконструированную лигазу, имеющая идентичность последовательности по меньшей мере 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более с эталонной последовательностью SEQ ID NO: 2, 6, 32, 34 и/или 38 или ее функциональным фрагментом, при этом указанный сконструированный полипептид содержит по меньшей мере одну замену в одном или более положениях аминокислот.

21. Полинуклеотидная последовательность по п. 19 и/или 20, при этом указанная полинуклеотидная последовательность содержит последовательность, имеющую идентичность последовательности по меньшей мере 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% или более с эталонной последовательностью SEQ ID NO: 1, 5, 31, 33 и/или 37.

22. Полинуклеотидная последовательность по п. 21, при этом указанная последовательность содержит SEQ ID NO: 1, 5, 31, 33 и/или 37.

23. Полинуклеотидная последовательность по любому из пп. 19–22, при этом указанная полинуклеотидная последовательность функционально связана с регуляторной последовательностью.

24. Полинуклеотидная последовательность по любому из пп. 19–22, при этом

указанная полинуклеотидная последовательность является кодон–оптимизированной.

25. Вектор экспрессии, содержащий по меньшей мере одну полинуклеотидную последовательность по любому из пп. 19–24.

26. Клетка–хозяин, трансформированная по меньшей мере вектором экспрессии по п. 25.

27. Способ получения полипептида сконструированной лигазы в клетке–хозяине, включающий культивирование клетки–хозяина по п. 26 в подходящих условиях культивирования таким образом, чтобы получить по меньшей мере одну сконструированную лигазу.

28. Способ по п. 27, дополнительно включающий выделение по меньшей мере одной сконструированной лигазы из культуры и/или клеток–хозяев.

29. Способ по п. 28, дополнительно включающий стадию очистки указанной по меньшей мере одной сконструированной лигазы.

30. Композиция, содержащая по меньшей мере одну сконструированную лигазу по любому из пп. 1–18.

31. Способ получения продукта лигирования, включающий предоставление по меньшей мере одной сконструированной лигазы по пп. 1–18, субстрата, содержащего по меньшей мере два фрагмента нуклеиновых кислот, и реакционной смеси; совмещение указанной по меньшей мере одной сконструированной лигазы, указанного субстрата и указанной реакционной смеси в таких условиях, чтобы происходило лигирование указанных фрагментов нуклеиновых кислот, и получался по меньшей мере один продукт лигирования.

32. Способ получения библиотеки ДНК, включающий предоставление по меньшей мере одной сконструированной лигазы по пп. 1–18, субстрата, содержащего входящую двухцепочечную ДНК, и композиции реакционной смеси, содержащей по меньшей мере один Т–образный адапторный олигонуклеотид, аденозин и реакционный буфер; совмещение указанной по меньшей мере одной сконструированной лигазы, указанного субстрата и указанной реакционной смеси в таких условиях, чтобы добавить аденозин к 3' концам обеих нитей указанных ДНК, и лигирование указанного Т–образного адаптера к концам входящей ДНК для получения продукта, содержащего библиотеку ДНК.

33. Способ по п. 32, в котором указанная входящая двухцепочечная ДНК содержит фрагменты ДНК с тупыми концами.

34. Способ получения множества фрагментов ДНК, подходящих для секвенирования, включающий предоставление по меньшей мере одной сконструированной лигазы по п. 1–18, субстрата, содержащего входящую двухцепочечную ДНК, и реакционной смеси содержащей олигонуклеотид, содержащий 3' выступы с одним основанием дезоксиаденином и 5' концы с монофосфатом, адапторный олигонуклеотид, содержащий 5' дезокситимиidinовый выступ и 5' фосфат на совместимом с лигированием конце; и совмещение указанной по меньшей мере одной сконструированной лигазы, указанного субстрата и указанной реакционной смеси в таких

условиях, чтобы происходило лигирование олигонуклеотида, адапторного олигонуклеотида и входящей двухцепочечной ДНК, с получением продукта, содержащего множество фрагментов ДНК, подходящих для секвенирования.

35. Способ получения множества фрагментов ДНК, подходящих для секвенирования, включающий предоставление по меньшей мере одной сконструированной лигазы по п. 1–18, субстрата, содержащего входящую двухцепочечную ДНК, и реакционной смеси содержащей олигонуклеотид, содержащий 3' выступы с одним основанием дезоксиаденином и 5' концы с монофосфатом, адапторный олигонуклеотид, содержащий 5' дезокситимиidinовый выступ и 5' фосфат на совместимом с лигированием конце; и совмещение указанной по меньшей мере одной сконструированной лигазы, указанного субстрата и указанной реакционной смеси в таких условиях, чтобы происходило лигирование олигонуклеотида, адапторного олигонуклеотида и входящей двухцепочечной ДНК, с получением продукта, содержащего множество фрагментов ДНК, подходящих для секвенирования, при этом концентрация указанного адапторного олигонуклеотида в указанной реакционной смеси составляет менее, чем 20-кратное молярное превышение концентрации субстрата.

36. Способ по любому из пп. 31–35, при этом указанное выдерживание проводят в присутствии краудинг-агента.

37. Способ по любому из пп. 31–36, в котором продукт трансформируют в *E. coli* после тепловой инактивации продукта.

38. Способ по любому из пп. 31–37, в котором продукт используют для создания библиотеки молекул ДНК.

39. Способ по п. 38, в котором указанную библиотеку молекул ДНК подвергают секвенированию.

40. Способ по любому из пп. 31–39, при этом указанная реакционная смесь содержит по меньшей мере одно соединение, которое ингибирует лигирование.

41. Способ по п. 40, при этом указанное по меньшей мере одно соединение, которое ингибирует лигирование, содержит субоптимальный буфер для лигирования.

42. Способ по любому из пп. 31–41, при этом в указанном способе получают больше продукта, чем в таких же способах при наличии T4 ДНК-лигазы дикого типа.

43. Способ по любому из пп. 31–42, при этом указанная реакционная смесь содержит по меньшей мере один фермент.

44. Способ по п. 43, при этом указанный фермент выбирают из полимеразы, полинуклеотидкиназы, экзонуклеазы, эндонуклеазы и цитидиндезаминазы.

45. Способ по любому из пп. 31–44, при этом указанная концентрация входящей двухцепочечной ДНК составляет менее, чем 100 нМ, менее, чем 50 нМ, менее, чем 10 нМ, менее, чем 1 нМ или менее, чем 100 пМ.

46. Способ по любому из пп. 31–45, при этом концентрация адаптеров составляет менее, чем 10-кратное, менее, чем 5-кратное, менее, чем 3-кратное или менее, чем 2-кратное превышение концентрации включений в реакционную смесь.

47. Способ по любому из пп. 31–46, дополнительно включающий стадию амплификации продукта.

48. Способ по любому из пп. 31–47, дополнительно включающий стадию секвенирования продукта.

49. Способ по любому из пп. 31–48, дополнительно включающий стадии амплификации и секвенирования продукта.

50. Способ по любому из пп. 31–49, в котором не предпринимают шаги для удаления димера адаптера из продукта.

51. Способ по любому из пп. 31–40, в котором указанный субстрат выбирают из выделенной внеклеточной ДНК, циркулирующей опухолевой ДНК, ДНК, выделенной из лейкозных клеток, ДНК, выделенной из клеток лимфомы, ДНК, выделенной из циркулирующих опухолевых клеток, ДНК, выделенной из инфицированных вирусом клеток, циркулирующей фетальной ДНК и тонкоигольных аспиратов.

52. Способ по любому из пп. 31–51, при этом указанный субстрат содержит входящую двухцепочечную ДНК, предоставленную в виде необработанного образца.

53. Способ по любому из пп. 31–51, при этом указанный субстрат содержит входящую двухцепочечную ДНК, которую очищают перед ее включением в реакционную смесь.

54. Способ по любому из пп. 31–53, при этом указанный способ осуществляют в условиях использования микроструйных устройств и/или капель.

55. Способ по п. 54, при этом объем комбинации указанной реакционной смеси и указанной сконструированной лигазы составляет менее, чем 5000 пл, менее, чем 1000 пл, менее, чем 100 пл, менее, чем 10 пл или менее, чем 1 пл.

56. Способ по любому из пп. 31–55, при этом указанный субстрат содержит иммобилизованную двухцепочечную входящую ДНК.

57. Способ по любому из пп. 31–56, в котором указанную сконструированную лигазу иммобилизируют.

58. Способ по любому из пп. 31–57, в котором иммобилизируют указанный субстрат, содержащий двухцепочечную входящую ДНК и указанную сконструированную лигазу.

59. Способ по любому из пп. 31–55, в котором иммобилизируют по меньшей мере одно соединение в указанной реакционной смеси.

60. Способ по любому из пп. 31–59, в котором продукт используют для создания библиотек для секвенирования ДНК, высокопроизводительного скрининга, генетического отбора, фагового дисплея, дрожжевого дисплея, рибосомального дисплея, клеточных анализов, биохимических анализов, одновременного многопараметрического анализа на основе изображений или определения конформации хроматина (C3).

61. Способ по любому из пп. 31–60, в котором продолжительность времени указанного совмещения составляет менее, чем 30 минут.

62. Способ по любому из пп. 31–61, в котором продолжительность времени

указанного совмещения составляет менее, чем 15 минут.

63. Способ по п. 62, в котором указанная продолжительность времени указанного совмещения составляет менее, чем 10, 9, 8, 7, 6, 5, 4, 3 или 2 минуты.

64. Способ по п. 63, в котором указанная продолжительность времени указанного совмещения составляет менее, чем 5 минут.

65. Способ по любому из пп. 31–64, в котором указанный продукт содержит меньше димеров адаптеров, чем в способах лигирования, в которых продолжительность времени совмещения составляет 15 минут или более.

66. Способ по любому из пп. 31–65, при этом указанный способ является внеклеточным.

67. Способ по любому из пп. 31–66, в котором субстратом является внеклеточная ДНК, экстрактированная из текучей среды, полученной у пациента.

68. Способ по п. 67, в котором указанная текучая среда представляет собой сыворотку или плазму.

69. Способ по любому из пп. 31–68, в котором указанный субстрат содержит нуклеиновые кислоты с отличающимися последовательностями на их 3' и 5' концах.

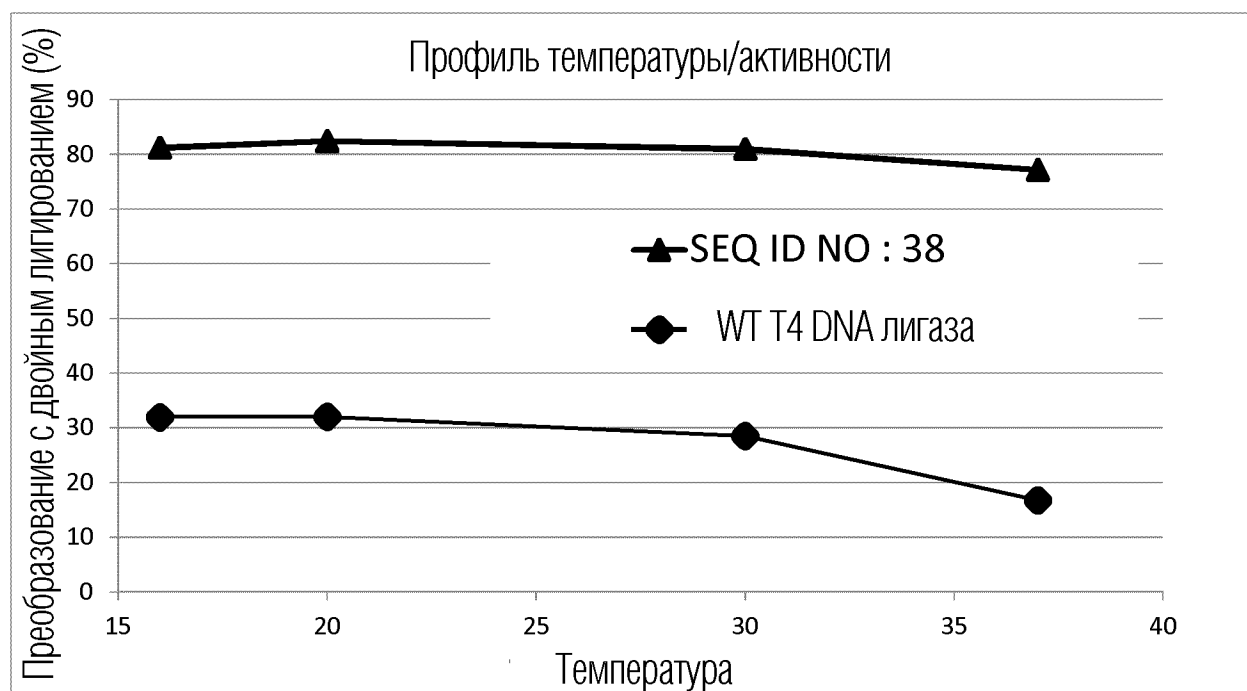
70. Способ по п. 69, при этом в способах достигают низкого смещения при лигировании.

71. Способы по любому из пп. 31–70, в котором указанное совмещение осуществляют при температуре в диапазоне между 10° и 40°C.

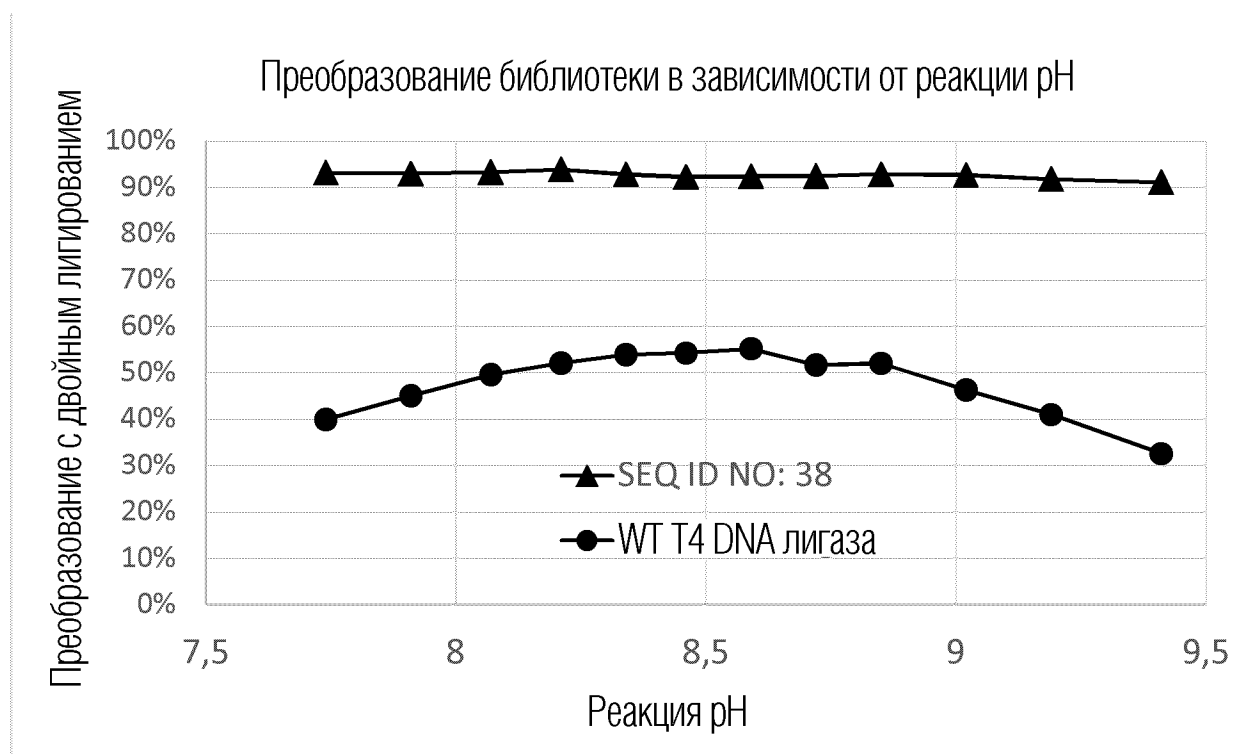
72. Способ по п. 71, в котором диапазон температур составляет 16°–37°C.

73. Способ по любому из пп. 31–72, в котором указанное совмещение осуществляют при pH в диапазоне от pH 7 до pH 10.

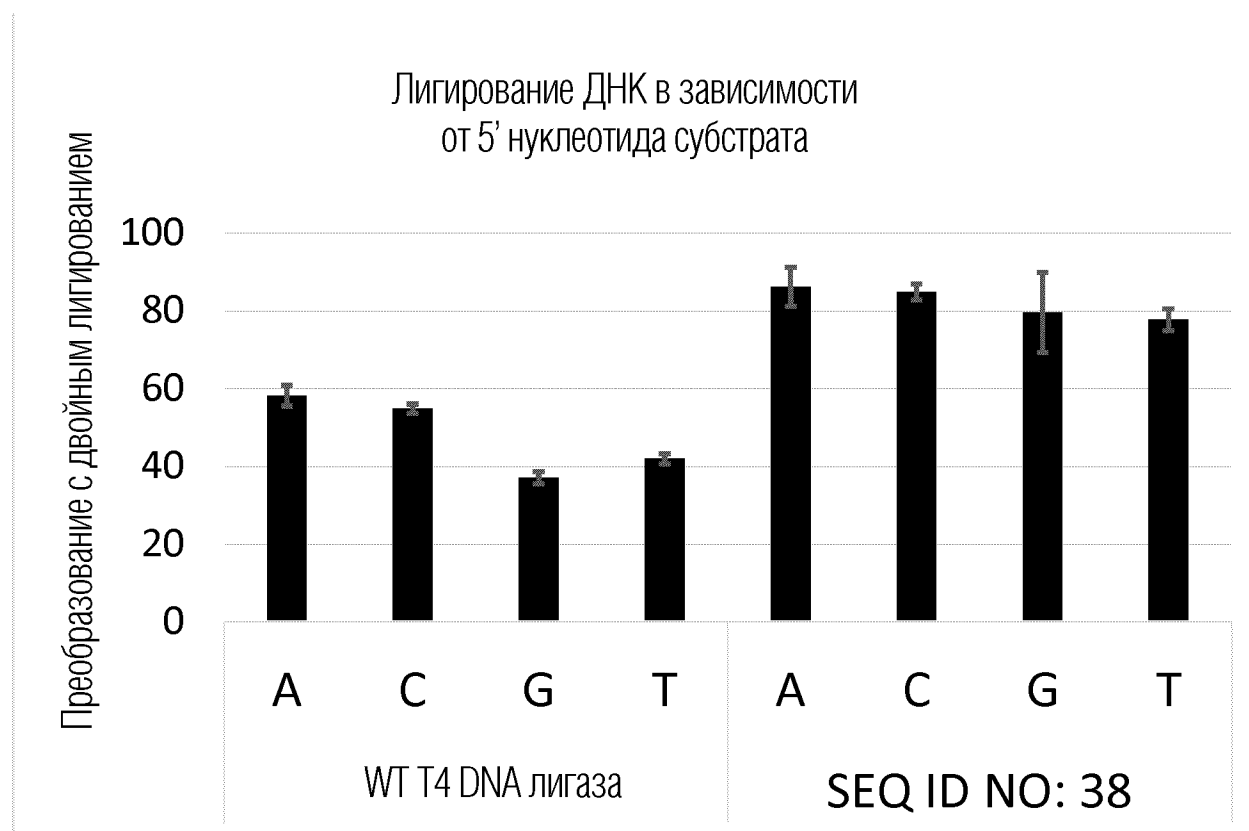
74. Способ по п. 73, в котором указанное совмещение осуществляют при pH между 7,5 и 9.



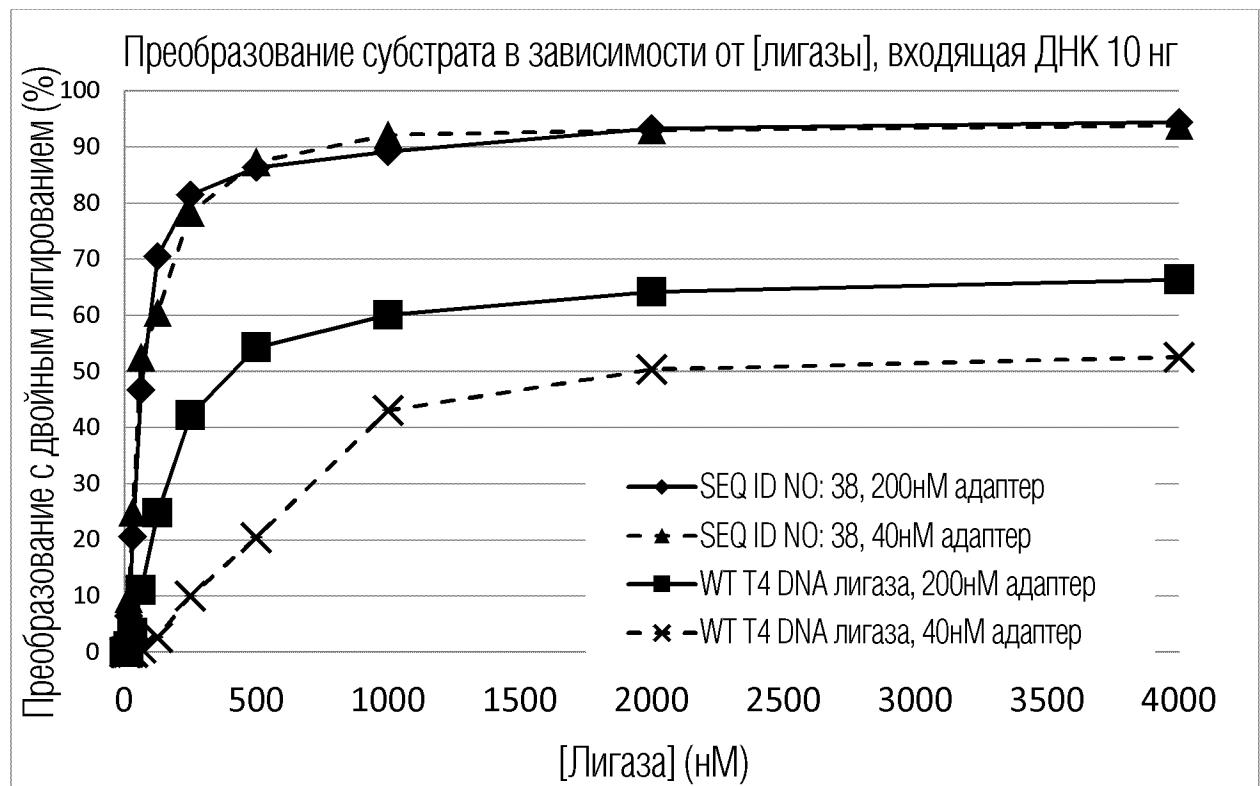
ФИГ. 1



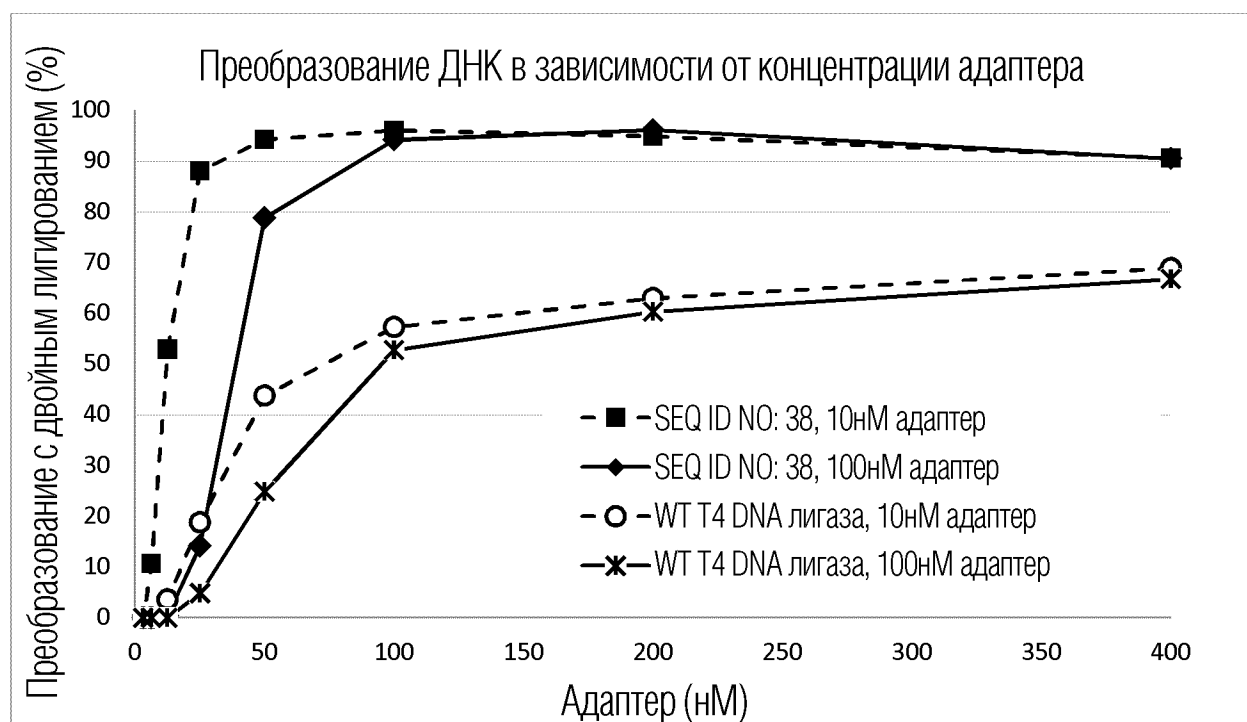
ФИГ. 2



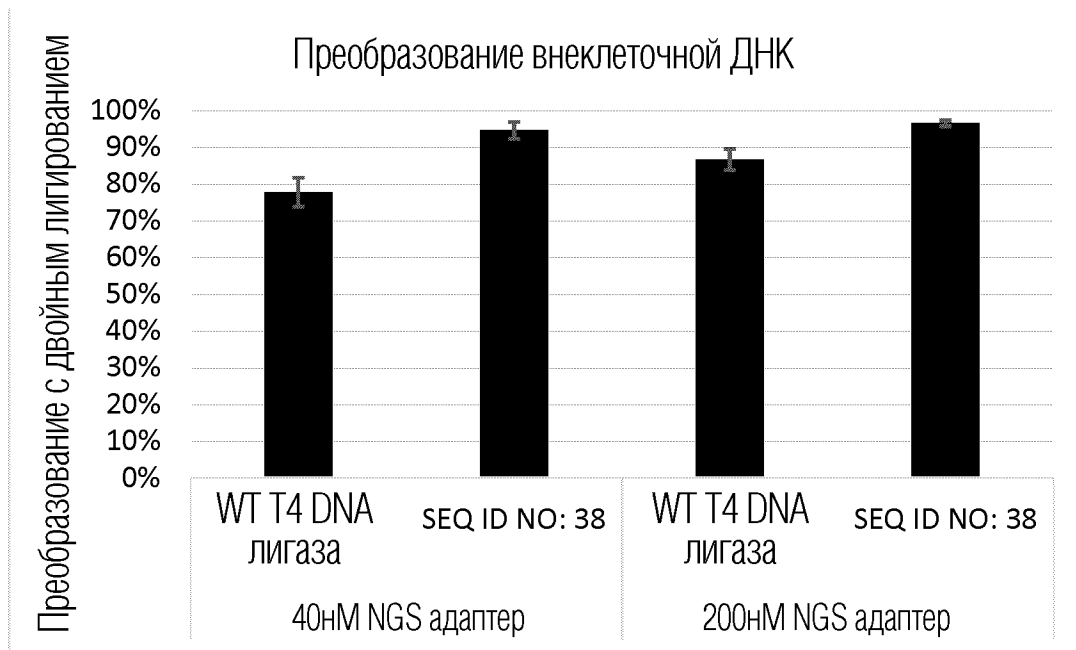
ФИГ. 3



ФИГ. 4



ФИГ. 5



ФИГ. 6