

(19)



**Евразийское
патентное
ведомство**

(21) **201590195** (13) **A1**

(12) **ОПИСАНИЕ ИЗОБРЕТЕНИЯ К ЕВРАЗИЙСКОЙ ЗАЯВКЕ**

(43) Дата публикации заявки
2015.05.29

(51) Int. Cl. **C40B 40/06** (2006.01)

(22) Дата подачи заявки
2013.07.12

(54) **ДНК-КОДИРОВАННЫЕ БИБЛИОТЕКИ, СОДЕРЖАЩИЕ СВЯЗИ КОДИРУЮЩИХ
ОЛИГОНУКЛЕОТИДОВ, НЕ ДОСТУПНЫЕ ДЛЯ СЧИТЫВАНИЯ ПОЛИМЕРАЗАМИ**

(31) **61/671,406**

(32) **2012.07.13**

(33) **US**

(86) **PCT/US2013/050303**

(87) **WO 2014/012010 2014.01.16**

(71) Заявитель:
ИКС-ЧЕМ, ИНК. (US)

(72) Изобретатель:
**Кееф Энтони Д., Литовчик Александр,
Кларк Мэтью А. (US)**

(74) Представитель:
Лыу Т.Н., Угрюмов В.М. (RU)

(57) Настоящее изобретение относится к комплексам кодированных олигонуклеотидами библиотек и способов мечения и использования таких библиотек. В частности, олигонуклеотиды и способы могут включать в себя комплексы, содержащие по меньшей мере одну связь, для которой полимеразы обладает пониженной способностью считывать или перемещаться.

201590195
A1

201590195

A1

ДНК-КОДИРОВАННЫЕ БИБЛИОТЕКИ, СОДЕРЖАЩИЕ СВЯЗИ КОДИРУЮЩИХ ОЛИГОНУКЛЕОТИДОВ, НЕ ДОСТУПНЫЕ ДЛЯ СЧИТЫВАНИЯ ПОЛИМЕРАЗАМИ

ОПИСАНИЕ

Ссылка на родственные заявки

По настоящей заявке испрашивается приоритет в соответствии с предварительной заявкой на патент США № 61/671406, поданной 13 июля 2012 г., которая полностью включена в настоящий документ посредством ссылки.

Уровень техники

В общем, настоящее изобретение относится к библиотекам ДНК-кодированных соединений и способам применения и создания таких библиотек. Настоящее изобретение также относится к композициям для применения таких библиотек.

ДНК-кодированные комбинаторные библиотеки предоставляют много преимуществ для открытия новых лекарств. Эти библиотеки могут обеспечивать большое количество различных соединений, которые могут быть быстро подвергнуты скринингу и запрошены. Для дополнительного увеличения варибельности могут быть запрограммированы и автоматизированы различные стадии процесса обнаружения. Эти стадии включают в себя использование многоступенчатого синтеза разделением-и-объединением для добавления строительных блоков к атомным или многоатомным каркасным элементам и использования ферментного и/или химического лигирования для добавления олигонуклеотидных меток, которые кодируют как синтетические стадии, так и структурные элементы.

Несмотря на эти преимущества, может возникнуть множество вопросов, когда должны быть синтезированы и подвергнуты деконволюции очень большие или сложные библиотеки. Поскольку размер библиотеки увеличивается, могут быть необходимы более совершенные способы для обеспечения высокого выхода связывания кодирующих олигонуклеотидных меток с использованием надежных и быстрых техник. Для создания библиотек в разнообразных условиях реакции были бы полезны устойчивые олигонуклеотидные конструкции, например, конструкции, которые стабильны в условиях высоких значений рН и повышенной температуры. Для упрощения деконволюции меток последовательность меток может быть узнана ДНК-

или РНК-зависимыми полимеразми, так что демографические данные популяции меток могут быть определены с помощью зависимой от шаблона полимеризации и определения последовательности. Трудности могут возникнуть при создании библиотеки, характеризующейся всеми из этих полезных свойств. Соответственно, существует потребность в улучшенных, более надежных способах скрининга и идентификации небольших соединений в кодированных олигонуклеотидами библиотеках.

Сущность изобретения

К отличительным признакам настоящего изобретения относятся комплексы для использования в ДНК-кодированных библиотеках, где эти комплексы содержат одну или несколько таких связей, что полимеразы обладает пониженной способностью к считыванию или перемещению. Для того, чтобы обнаружить идентичность кодированных областей (например, головного фрагмента, одной или нескольких меток и/или хвостового фрагмента), связь либо обращают прежде чем она подвергается действию полимеразы, либо, альтернативно, обходят полимеразой. Эта связь может быть обойдена любым применимым образом, который может не захватывать всю информацию о последовательности меток (например, таких как в 5'- или 3'-коннекторах, как описано в настоящем документе), но захватывает информацию кодирующей последовательности меток. Такие связи могут привести к снижению числа случаев ошибочного мечения, расширить число и вид связей, которые будут использоваться в создании комплексов и библиотек, а также обеспечить неферментативные способы для создания и скрининга ДНК-кодированных комплексов. Дополнительные преимущества этих комплексов и способов описаны в настоящем документе.

Соответственно, согласно одному аспекту в настоящем изобретении предусмотрен комплекс, включающий в себя: химическое соединение, включающее в себя один или несколько каркасных элементов или один или несколько структурных элементов; первую олигонуклеотидную метку, кодирующую идентификацию по меньшей мере одного из одного или нескольких каркасных элементов или структурных элементов и головной фрагмент, содержащий первую функциональную группу и вторую функциональную группу, где первая функциональная группа функционально связывается с химическим соединением, и вторая функциональная группа

функционально связывается с первой меткой через первую связь, для которой полимеразы обладает пониженной способностью считывать или перемещаться.

Согласно другому аспекту в настоящем изобретении предусмотрен комплекс, включающий в себя: химическое соединение, включающее в себя один или несколько каркасных элементов или один или несколько структурных элементов; n количество олигонуклеотидных меток, содержащих $n-1$ связей, где n представляет собой целое число от 1 до приблизительно 10 и где каждая из связей между двумя соседними метками и каждая метка кодирует идентификацию по меньшей мере одного из одного или нескольких каркасных элементов или структурных элементов; и головной фрагмент, содержащий первую функциональную группу и вторую функциональную группу, где первая функциональная группа функционально связана с химическим соединением, и вторая функциональная группа функционально связана по меньшей мере с одним из n количества меток через первую связь, где полимеразы обладает пониженной способностью считывать или перемещаться через по меньшей мере одну из первой связи и $n-1$ связей.

Согласно некоторым вариантам осуществления n составляет от 1 до 2, от 1 до 3, от 1 до 4, от 1 до 5, от 1 до 6, от 1 до 7, от 1 до 8, от 1 до 9, от 1 до 10, от 1 до 12, от 1 до 15, от 1 до 18, от 1 до 20, от 2 до 3, от 2 до 4, от 2 до 5, от 2 до 6, от 2 до 7, от 2 до 8, от 2 до 9, от 2 до 10, от 2 до 12, от 2 до 15, от 2 до 18, от 2 до 20, от 3 до 4, от 3 до 5, от 3 до 6, от 3 до 7, от 3 до 8, от 3 до 9, от 3 до 10, от 3 до 12, от 3 до 15, от 3 до 18, от 3 до 20, от 4 до 5, от 4 до 6, от 4 до 7, от 4 до 8, от 4 до 9, от 4 до 10, от 4 до 12, от 4 до 15, от 4 до 18 или от 4 до 20.

Согласно некоторым вариантам осуществления полимеразы обладает пониженной способностью считывать или перемещаться через по меньшей мере приблизительно 10% (например, приблизительно 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или даже 100%, по сравнению с контролем) первой связи и $n-1$ связей. Согласно конкретным вариантам осуществления полимеразы обладает пониженной способностью считывать или перемещаться через приблизительно от 10% до приблизительно 100% первой связи и $n-1$ связей (например, от 20% до 100%, от 25% до 100%, от 50% до 100%, от 75% до 100%, от 90% до 100%, от 95% до 100%, от 10% до 95%, от 20% до 95%, от 25% до 95%, от 50% до 95%, от 75% до 95%, от 90% до 95%, от 10% до 90%, от 20% до 90%, от 25% до 90%, от 50% до 90% или от 75% до 90%, по сравнению с контролем (например, по сравнению с контрольным олигонуклеотидом без соединения)).

Согласно некоторым вариантам осуществления одна или несколько меток включают в себя 5'-коннектор на 5'-конце одной или нескольких меток и 3'-коннектор на 3'-конце одной или нескольких меток. Согласно конкретным вариантам осуществления каждый 5'-коннектор и/или каждый 3'-коннектор включает в себя ту же последовательность. Согласно другим вариантам осуществления каждый 5'-коннектор и/или каждый 3'-коннектор включает в себя другую последовательность.

Согласно некоторым вариантам осуществления первая связь и $n-1$ связей составляют по меньшей мере приблизительно 3 ангстрема в длину (например, по меньшей мере приблизительно 5, 8, 10, 15, 20, 30, 35, 40, 50, 60, 65 или 70 ангстрем). Согласно некоторым вариантам осуществления первая связь и $n-1$ связей составляют менее чем приблизительно 30 ангстрем в длину (например, менее чем приблизительно 25, 20, 15 или 10 ангстрем).

Согласно некоторым вариантам осуществления ДНК-полимераза и/или РНК-полимераза (например, описанная в настоящем документе) характеризуется пониженной способностью к считыванию или перемещению через первую связь и $n-1$ связей.

Согласно некоторым вариантам осуществления менее чем приблизительно 10% (например, приблизительно 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90% или 95%) первой связи и $n-1$ связей включают в себя ферментативную связь. Согласно некоторым вариантам осуществления первая связь и $n-1$ связей включают в себя от 0% до 90% ферментативных связей (например, приблизительно от 0% до 40%, от 0% до 45%, от 0% до 50%, от 0% до 55%, от 0% до 60%, от 0% до 65%, от 0% до 70%, от 0% до 75%, от 0% до 80%, от 0% до 85%, от 0% до 90%, от 0% до 95%, от 0% до 96%, от 0% до 97%, от 0% до 98%, от 0% до 99%, от 5% до 40%, от 5% до 45%, от 5% до 50%, от 5% до 55%, от 5% до 60%, от 5% до 65%, от 5% до 70%, от 5% до 75%, от 5% до 80%, от 5% до 85%, от 5% до 90%, от 5% до 95%, от 5% до 96%, от 5% до 97%, от 5% до 98%, от 5% до 99%, от 10% до 40%, от 10% до 45%, от 10% до 50%, от 10% до 55%, от 10% до 60%, от 10% до 65%, от 10% до 70%, от 10% до 75%, от 10% до 80%, от 10% до 85%, от 10% до 90%, от 10% до 95%, от 10% до 96%, от 10% до 97%, от 10% до 98%, от 10% до 99%, от 15% до 40%, от 15% до 45%, от 15% до 50%, от 15% до 55%, от 15% до 60%, от 15% до 65%, от 15% до 70%, от 15% до 75%, от 15% до 80%, от 15% до 85%, от 15% до 90%, от 15% до 95%, от 15% до 96%, от 15% до 97%, от 15% до 98%, от 15% до 99%, от 20% до 40%, от 20% до 45%, от 20% до 50%, от 20% до 55%, от 20% до 60%, от 20% до 65%, от 20% до 70%, от 20% до 75%, от

20% до 80%, от 20% до 85%, от 20% до 90%, от 20% до 95%, от 20% до 96%, от 20% до 97%, от 20% до 98% или от 20% до 99%).

Согласно некоторым вариантам осуществления первая связь и $n-1$ связей включают в себя химическую связь (например, химическую реакционноспособную группу, фотореакционноспособную группу, интеркалирующий фрагмент или сшивающий олигонуклеотид). Согласно конкретным вариантам осуществления по меньшей мере одна (например, две, три, четыре, пять или более) химическая реакционноспособная группа, фотореакционноспособная группа или интеркалирующий фрагмент присутствует в 5'-коннекторе на или в непосредственной близости от 5'-конца метки и/или в 3'-коннекторе на или в непосредственной близости от 3'-конца метки. Согласно другим вариантам осуществления последовательность по меньшей мере одного из 5'-коннектора комплементарна последовательности смежного 3'-коннектора или идентична или достаточно сходна, чтобы обеспечить гибридизацию с комплементарным олигонуклеотидом. Согласно некоторым вариантам осуществления по меньшей мере 10% (например, приблизительно 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или даже 100%) первой связи и/или $n-1$ связей представляют собой химические связи. Согласно другим вариантам осуществления приблизительно от 10% до приблизительно 100% первой связи и $n-1$ связей (например, от 20% до 100%, от 25% до 100%, от 50% до 100%, от 75% до 100%, от 90% до 100%, от 95% до 100%, от 10% до 95%, от 20% до 95%, от 25% до 95%, от 50% до 95%, от 75% до 95%, от 90% до 95%, от 10% до 90%, от 20% до 90%, от 25% до 90%, от 50% до 90% или от 75% до 90%) представляют собой химические связи.

Согласно некоторым вариантам осуществления химическую реакционноспособную группу выбирают из пары необязательно замещенной алкинильной группы и необязательно замещенной азидогруппы; пары необязательно замещенного диена, содержащего систему 4π -электронов, и необязательно замещенного диенофила или необязательно замещенного гетеродиенофила, содержащего систему 2π -электронов; пары нуклеофила и напряженного гетероциклического электрофила; пары необязательно замещенной аминогруппы и альдегида или кетонной группы; пары необязательно замещенной аминогруппы и карбоксильной группы; пары необязательно замещенного гидразина и альдегидной или кетонной группы; пары необязательно замещенного гидроксилamina и альдегидной или кетонной группы; пары нуклеофила и необязательно замещенного

алкилгалогенида; платинового комплекса; алкилирующего средства или фуран-модифицированного нуклеотида (например, любого описанного в настоящем документе).

Согласно некоторым вариантам осуществления фотореакционноспособная группа включает в себя интеркалирующий фрагмент, производное псоралена, необязательно замещенную циановинилкарбазольную группу (например, 3-циановинилкарбазольную группу, такую как 3-циановинилкарбазол-1'- β -деоксирибозид-5'-трифосфат), необязательно замещенную винилкарбазольную группу (например, амидовинилкарбазольную группу, карбоксивинилкарбазольную группу или C₂₋₇ алкоксикарбонилвинилкарбазольную группу, как описано в настоящем документе), необязательно замещенную циановинильную группу, необязательно замещенную акриламидную группу, необязательно замещенную диазириновую группу, необязательно замещенный бензофенон или необязательно замещенную азидную группу (например, любые описанные в настоящем документе).

Согласно некоторым вариантам осуществления интеркалирующий фрагмент представляет собой производное псоралена (например, псорален, 8-метоксипсорален или 4-гидроксиметил-4,5,8-триметилпсорален (НМТ-псорален)), производное алкалоида (например, берберин, палматин, коралин, сангвинарин (например, их иминиевые или алканоламиновые формы или аристололактam- β -D-глюкозид), катион этидия (например, этидия бромид), производное акридина (например, профлавин, акрифлавин или амсакрин), производное антрациклина (например, доксорубицин, эпирубицин, даунорубицин (дауномицин), идарубицин и акларубицина) или талидомид.

Согласно некоторым вариантам осуществления химическая связь включает в себя сшивающий олигонуклеотид, причем последовательность по меньшей мере пяти нуклеотидов на 5'-конце сшивающего олигонуклеотида комплементарна последовательности по меньшей мере пяти нуклеотидов на 3'-конце одной или нескольких меток или идентична или достаточно сходна, чтобы обеспечить гибридизацию с комплементарным олигонуклеотидом, и причем последовательность по меньшей мере пяти нуклеотидов на 3'-конце сшивающего олигонуклеотида комплементарна последовательности по меньшей мере пяти нуклеотидов на 5'-конце одной или нескольких меток или идентична или достаточно сходна, чтобы обеспечить гибридизацию с комплементарным олигонуклеотидом. Согласно конкретным вариантам осуществления 3'-конец одной или нескольких меток включает в себя 3'-

коннектор. Согласно конкретным вариантам осуществления 5'-конец одной или нескольких меток включает в себя 5'-коннектор.

Согласно некоторым вариантам осуществления 5'-конец и/или 3'-конец сшивающего олигонуклеотида включает в себя обратимую кореакционноспособную группу (например, циановинилкарбазольную группу, циановинильную группу, акриламидную группу, тиольную группу или винилсульфоновую группу, как описано в настоящем документе).

Согласно некоторым вариантам осуществления 3'-коннектор и/или 5'-коннектор включает в себя обратимую кореакционноспособную группу (например, циановинилкарбазольную группу, циановинильную группу, акриламидную группу, тиольную группу или винилсульфоновую группу, как описано в настоящем документе).

Согласно некоторым вариантам осуществления химическое соединение функционально связано с головным фрагментом с помощью бифункционального спейсера (например, любого описанного в настоящем документе). Согласно другим вариантам осуществления химическое соединение ковалентно прикреплено к головному фрагменту. Согласно конкретным вариантам осуществления головной фрагмент включает в себя олигонуклеотид, выбранный из группы, состоящей из двухцепочечного олигонуклеотида, одноцепочечного олигонуклеотида или олигонуклеотида-шпильки. Согласно некоторым вариантам осуществления головной фрагмент включает в себя связывающую праймер область.

Согласно любому из описанных в настоящем документе вариантов осуществления комплекс или способ дополнительно включает одну или несколько первых меток идентификации библиотек, метку(ки) использования и/или метку(ки) происхождения. Согласно некоторым вариантам осуществления комплекс или способ включает от 2 до 20 меток (например, от 2 до 10 меток структурных элементов или каркасных элементов, одну первую метку идентификации библиотеки, одну метку дополнительного использования и одну метку происхождения). Согласно некоторым вариантам осуществления комплекс или способ включает от 5 до 75 нуклеотидов (например, как описано в настоящем документе, например, приблизительно 40 или 50 нуклеотидов).

Согласно любому из описанных в настоящем документе вариантов осуществления каждая из меток в пределах одного набора меток характеризуется приблизительно той же массой.

Согласно любому из описанных в настоящем документе вариантов осуществления комплекс включает в себя РНК, ДНК, модифицированную ДНК и/или модифицированную РНК (например, ПНК, замкнутую нуклеиновую кислоту, гликольную нуклеиновую кислоту, треозную нуклеиновую кислоту или их смесь в пределах одного и того же олигонуклеотида, как описано в настоящем документе).

Согласно любому из описанных в настоящем документе вариантов осуществления комплекс или способ дополнительно включает хвостовой фрагмент.

Согласно другому аспекту настоящее изобретение включает библиотеку, включающую в себя один или несколько описанных в настоящем документе комплексов. Согласно некоторым вариантам осуществления библиотека включает в себя множество головных фрагментов. Согласно другим вариантам осуществления каждый головной фрагмент из множества головных фрагментов включает в себя область одинаковой последовательности (например, связывающую праймер область) и отличную кодирующую область (например, первую метку, которая кодирует использование библиотеки, происхождение библиотеки, идентификацию библиотеки, историю библиотеки, связь, спейсер или добавление первого компонента или олигонуклеотидной последовательности, которая облегчает технологии гибридизации, амплификации или секвенирования). Согласно конкретным вариантам осуществления библиотека включает в себя от приблизительно 10^2 до 10^{20} комплексов (например, приблизительно от 10^2 до 10^3 , от 10^2 до 10^4 , от 10^2 до 10^5 , от 10^2 до 10^6 , от 10^2 до 10^7 , от 10^2 до 10^8 , от 10^2 до 10^9 , от 10^2 до 10^{10} , от 10^2 до 10^{11} , от 10^2 до 10^{12} , от 10^2 до 10^{13} , от 10^2 до 10^{14} , от 10^2 до 10^{15} , от 10^2 до 10^{16} , от 10^2 до 10^{17} , от 10^2 до 10^{18} , от 10^2 до 10^{19} , от 10^4 до 10^5 , от 10^4 до 10^6 , от 10^4 до 10^7 , от 10^4 до 10^8 , от 10^4 до 10^9 , от 10^4 до 10^{10} , от 10^4 до 10^{11} , от 10^4 до 10^{12} , от 10^4 до 10^{13} , от 10^4 до 10^{14} , от 10^4 до 10^{15} , от 10^4 до 10^{16} , от 10^4 до 10^{17} , от 10^4 до 10^{18} , от 10^4 до 10^{19} , от 10^4 до 10^{20} , от 10^5 до 10^6 , от 10^5 до 10^7 , от 10^5 до 10^8 , от 10^5 до 10^9 , от 10^5 до 10^{10} , от 10^5 до 10^{11} , от 10^5 до 10^{12} , от 10^5 до 10^{13} , от 10^5 до 10^{14} , от 10^5 до 10^{15} , от 10^5 до 10^{16} , от 10^5 до 10^{17} , от 10^5 до 10^{18} , от 10^5 до 10^{19} или от 10^5 до 10^{20} комплексов). Согласно некоторым вариантам осуществления каждый комплекс отличается.

Согласно другому аспекту в настоящем изобретении предусмотрен способ мечения первой библиотеки, включающий в себя кодированные химические соединения, способ, включающий: (а) обеспечение головного фрагмента, содержащего первую функциональную группу и вторую функциональную группу (например, где головной фрагмент необязательно кодирует информацию); (б) связывание первой

функциональной группы головного фрагмента с первым компонентом химического соединения, где головной фрагмент непосредственно соединен с первым компонентом или головной фрагмент опосредованно соединен с первым компонентом с помощью бифункционального спейсера; и (с) связывание второй функциональной группы головного фрагмента с первой олигонуклеотидной меткой через первую связь с образованием комплекса, причем полимераза обладает пониженной способностью считывать или перемещаться через первую связь; причем стадии (b) и (с) могут быть выполнены в любом порядке, и причем первая метка кодирует реакцию связывания на стадии (b), тем самым обеспечивая меченую библиотеку.

Согласно еще одному аспекту в настоящем изобретении предусмотрен способ мечения первой библиотеки, включающий кодированное химическое соединение, способ, включающий: (a) обеспечение головного фрагмента, содержащего первую функциональную группу и вторую функциональную группу (например, где головной фрагмент необязательно кодирует информацию); (b) связывание первой функциональной группы головного фрагмента с первым компонентом химического соединения, причем головной фрагмент непосредственно соединен с первым компонентом или головной фрагмент опосредованно соединен с первым компонентом с помощью бифункционального спейсера; (с) связывание второй функциональной группы головного фрагмента с первой олигонуклеотидной меткой через первую связь; (d) связывание n_c количества дополнительных компонентов химического соединения, где n_c представляет собой целое число от 1 до 10; и (e) связывание n_t количества дополнительных меток олигонуклеотида, содержащих n_t связей с образованием комплекса, где n_t представляет собой целое число от 1 до 10, причем каждая из связей находится между двумя соседними метками и причем каждая метка кодирует идентификацию по меньшей мере одного из компонентов; причем полимераза обладает пониженной способностью считывать или перемещаться через по меньшей мере одну из первой связи и n_t связей; и причем стадии (b) и (с) могут быть выполнены в любом порядке, и причем первая метка кодирует реакцию связывания на стадии (b); причем стадии (d) и (e) могут быть выполнены в любом порядке и причем каждая дополнительная метка кодирует реакцию связывания каждого дополнительного компонента на стадии (d), тем самым обеспечивая меченую библиотеку.

Согласно некоторым вариантам осуществления n_c и n_t представляют собой каждый независимо целое число от 1 до 2, от 1 до 3, от 1 до 4, от 1 до 5, от 1 до 6, от 1 до 7, от 1 до 8, от 1 до 9, от 1 до 10, от 1 до 12, от 1 до 15, от 1 до 18, от 1 до 20, от 2 до

3, от 2 до 4, от 2 до 5, от 2 до 6, от 2 до 7, от 2 до 8, от 2 до 9, от 2 до 10, от 2 до 12, от 2 до 15, от 2 до 18, от 2 до 20, от 3 до 4, от 3 до 5, от 3 до 6, от 3 до 7, от 3 до 8, от 3 до 9, от 3 до 10, от 3 до 12, от 3 до 15, от 3 до 18, от 3 до 20, от 4 до 5, от 4 до 6, от 4 до 7, от 4 до 8, от 4 до 9, от 4 до 10, от 4 до 12, от 4 до 15, от 4 до 18 или от 4 до 20.

Согласно некоторым вариантам осуществления полимераза обладает пониженной способностью считывать или перемещаться через по меньшей мере приблизительно 10% первой связи и n_t связей (например, приблизительно 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или даже 100%, по сравнению с контролем первой связи и n_t связей). Согласно конкретным вариантам осуществления снижение способности считывать или перемещаться через первую связь и n_t связей составляет приблизительно от 10 до приблизительно 100% (например, от 20% до 100%, от 25% до 100%, от 50% до 100%, от 75% до 100%, от 90% до 100%, от 95% до 100%, от 10% до 95%, от 20% до 95%, от 25% до 95%, от 50% до 95%, от 75% до 95%, от 90% до 95%, от 10% до 90%, от 20% до 90%, от 25% до 90%, от 50% до 90% или от 75% до 90%, по сравнению с контролем (например, по сравнению с контрольным олигонуклеотидом без связи)).

Согласно некоторым вариантам осуществления первый компонент и/или дополнительные компоненты включают в себя каркасный элемент или структурный элемент.

Согласно некоторым вариантам осуществления стадия (b) включает в себя связывание головного фрагмента опосредованно с первым компонентом через бифункциональный спейсер.

Согласно некоторым вариантам осуществления способы дополнительно включают связывание метки использования, метки происхождения и/или метки идентификации первой библиотеки с комплексом. Согласно конкретным вариантам осуществления способ дополнительно включает обеспечение второй библиотеки и объединение первой библиотеки со второй библиотекой. Согласно некоторым вариантам осуществления способ дополнительно включает связывание хвостового фрагмента с комплексом.

Согласно некоторым вариантам осуществления стадии связывания включают в себя химическую реакционноспособную группу, фотореакционноспособную группу, интеркалирующий фрагмент или сшивающий олигонуклеотид (например, как описано в настоящем документе для любого из комплексов в настоящем документе). Согласно конкретным вариантам осуществления способ дополнительно включает выполнение

синтеза разделения-и-объединения в одной или нескольких стадиях, отделяя один или несколько комплексов и/или очищая один или несколько комплексов.

Согласно любому из описанных в настоящем документе способов способ приводит к любому из описанных в настоящем документе комплексов или любой из описанных в настоящем документе библиотек.

Согласно еще одному аспекту в настоящем изобретении предусмотрен способ скрининга множества химических соединений, способ, включающий: (а) контактирование мишени с любым описанным в настоящем документе комплексом и/или описанной в настоящем документе библиотекой; и (b) выбор одного или нескольких комплексов, характеризующихся заранее определенной характеристикой для мишени, по сравнению с контролем, таким образом, подвергая скринингу химическое соединение.

Согласно некоторым вариантам осуществления заранее определенная характеристика включает в себя увеличение связывания мишени (например, описанной в настоящем документе биологической мишени), по сравнению с контролем.

Согласно некоторым вариантам осуществления перед стадией (а) выполняют одну или несколько операций, выбранных из группы, состоящей из: отжига одного или нескольких релейных праймеров с указанным комплексом, причем указанный один или несколько релейных праймеров перекрывают указанную первую связь и/или $n-1$ связи; расширения указанных релейных праймеров с использованием полимеразы для получения олигонуклеотидных фрагментов и лигирования указанного олигонуклеотидного фрагмента для получения шаблона.

Согласно другому аспекту в настоящем изобретении предусмотрен способ определения нуклеотидной последовательности любого описанного в настоящем документе комплекса, способ, включающий: (а) отжиг одного или нескольких релейных праймеров с комплексом, причем один или несколько релейных праймеров охватывают первую связь и/или $n-1$ связей; (b) расширение релейных праймеров с использованием полимеразы для получения олигонуклеотидных фрагментов; (c) лигирование олигонуклеотидных фрагментов для получения шаблона; (d) дополнительную амплификацию шаблона с помощью полимеразной цепной реакции для получения амплифицированной смеси и (e) секвенирование дополнительно амплифицированной смеси для определения последовательности комплекса.

Согласно некоторым вариантам осуществления комплекс включает в себя 5'-коннектор на 5'-конце одной или нескольких меток и 3'-коннектор на 3'-конце одной

или нескольких меток, и причем каждый из одного или нескольких релейных праймеров гибридизуется с соседними 5'- и 3'-коннекторами. Согласно конкретным вариантам осуществления каждый 5'-коннектор включает в себя ту же последовательность и/или каждый 3'-коннектор включает в себя ту же последовательность. Согласно некоторым вариантам осуществления последовательность по меньшей мере одного из 5'-коннекторов комплементарна последовательности смежного 3'-коннектора (например, с образованием дуплекса между 5'- и 3'-коннекторами в условиях гибридизации) или идентична или достаточно близка, чтобы обеспечить гибридизацию с комплементарным олигонуклеотидом. Согласно дополнительным вариантам осуществления способ включает гибридизацию 5'-коннектора с соседним 3'-коннектором. Согласно некоторым вариантам осуществления каждый релейный праймер для специфической области соединения (например, таким образом, образуя трехспиральное соединение) между 5'- и 3'-коннекторами включает в себя одну ту же последовательность.

Согласно другому аспекту в настоящем изобретении предусмотрен способ определения нуклеотидной последовательности любого описанного в настоящем документе комплекса, способ, включающий: (а) получение комплекса, включающего в себя химическую связь, причем химическая связь представляет собой сшивающий олигонуклеотид, который охватывает область соединения между двумя соседними метками и который функционально связывается с двумя соседними метками через одну или несколько обратимых кореакционноспособных групп; (b) высвобождение сшивающего олигонуклеотида (например, любого описанного в настоящем документе) для получения шаблона; (с) дополнительную амплификацию шаблона с помощью полимеразной цепной реакции для получения амплифицированной смеси; и (d) секвенирование дополнительно амплифицированной смеси для определения последовательности комплекса. Согласно некоторым вариантам осуществления часть сшивающего олигонуклеотида гибридизуется с концами двух смежных меток, в результате чего получая одноцепочечный разрыв или пробел между двумя смежными метками.

Согласно дополнительным вариантам осуществления способ включает лигирование меток с использованием химического процесса или ферментативного процесса (например, 5'-фосфорилирования) перед стадией (b).

Согласно некоторым вариантам осуществления способ дополнительно включает репарацию шаблона перед стадией дополнительной амплификации или стадией

секвенирования. Согласно конкретным вариантам осуществления стадия репарации включает в себя модификацию по меньшей мере одной из первой восстанавливаемой связи и/или $n-1$ связей, способных к считыванию или перемещению полимеразой. Согласно некоторым вариантам осуществления полимеразы может считывать или перемещаться через по меньшей мере 50% (например, по меньшей мере 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или даже 100%) репарированных связей (например, по сравнению с контролем, таким как контрольный олигонуклеотид, содержащий первую связь и $n-1$ связей, которые не были исправлены). Согласно конкретным вариантам осуществления полимеразы может считывать или перемещаться через от приблизительно 10% до приблизительно 100% репарированных связей (например, от 20% до 100%, от 25% до 100%, от 50% до 100%, от 75% до 100%, от 90% до 100%, от 95% до 100%, от 10% до 95%, от 20% до 95%, от 25% до 95%, от 50% до 95%, от 75% до 95%, от 90% до 95%, от 10% до 90%, от 20% до 90%, от 25% до 90%, от 50% до 90%, от 75% до 90%, по сравнению с контролем (например, по сравнению с контрольным олигонуклеотидом, содержащим первую связь и $n-1$ связей, которые не подвергались репарации)). Согласно конкретным вариантам осуществления стадия репарации включает в себя обеспечение шаблона ферментом репарации (например, фотолиазой, гликозилазой, эндонуклеазой, флэп-эндонуклеазой, апуриновой/апириимидиновой (AP) эндонуклеазой, поли(АДФ-рибоза)полимеразой или метилтрансферазой, как описано в настоящем документе).

Согласно любому из приведенных выше вариантов осуществления способы или комплексы могут включать только одноцепочечные молекулы, где головной фрагмент, первая метка и/или одна или несколько дополнительных меток представляют собой одноцепочечные.

Согласно любому из приведенных выше вариантов осуществления способ дополнительно включает одну или несколько дополнительных стадий по обеспечению разнообразия библиотеки или перепроверке представителя библиотеки, как описано в настоящем документе. Согласно некоторым вариантам осуществления способ дополнительно включает идентификацию небольшого подобного лекарственному средству представителя библиотеки, который связывается или инактивирует представляющий терапевтический интерес белок. Согласно другим вариантам осуществления способ дополнительно включает контактирование представителя библиотеки с биологической мишенью в условиях, подходящих по меньшей мере одному представителю библиотеки для связывания с мишенью, удаление одного или

нескольких представителей библиотеки, которые не связываются с мишенью, и анализ одной или нескольких олигонуклеотидных меток, связанные с мишенью.

Как описано в настоящем документе, использование одноцепочечных молекул может характеризоваться множеством преимуществ. Таким образом, согласно любому из описанных в настоящем документе вариантов осуществления способы и комплексы включают головной фрагмент, одну или несколько меток, комплекс, химическое соединение, молекулу или любого представителя меченой библиотеки, характеризующегося сниженной массой, повышенной растворимостью (например, в органическом растворителе), сниженной стоимостью, повышенной реакционной способностью, повышенной доступностью мишени, сниженным гидродинамическим радиусом и/или повышенной точностью аналитических оценок, по сравнению со способом, включающим одну или несколько двухцепочечных молекул (например, двухцепочечный головной фрагмент или двухцепочечную метку). Согласно некоторым вариантам осуществления каждая из меток в пределах набора меток (например, первая метка или последующая метка, если присутствует) характеризуется приблизительно такой же массой (например, каждая метка характеризуется массой, которая составляет приблизительно +/- 10% от средней массы между двумя или более метками). Согласно конкретным вариантам осуществления метка характеризуется сниженной массой (например, менее чем приблизительно 15000 дальтон, приблизительно 14000 дальтон, приблизительно 13000 дальтон, приблизительно 12000 дальтон, приблизительно 11000 дальтон, приблизительно 10000 дальтон, приблизительно 9000 дальтон, приблизительно 8000 дальтон, приблизительно 7500 дальтон, приблизительно 7000 дальтон, приблизительно 6000 дальтон, приблизительно 6500 дальтон, приблизительно 5000 дальтон, приблизительно 5500 дальтон, приблизительно 4000 дальтон, приблизительно 4500 дальтон или приблизительно 3000 дальтон), по сравнению с двухцепочечной меткой (например, двухцепочечной меткой, характеризующейся массой, равной приблизительно 15000 дальтон, приблизительно 14000 дальтон, приблизительно 13000 дальтон или приблизительно 12000 дальтон). Согласно другим вариантам осуществления метка характеризуется уменьшенной длиной, по сравнению с двухцепочечной меткой (например, двухцепочечной меткой, характеризующейся длиной менее чем приблизительно 20 нуклеотидов, менее чем приблизительно 19 нуклеотидов, менее чем приблизительно 18 нуклеотидов, менее чем приблизительно 17 нуклеотидов, менее чем приблизительно 16 нуклеотидов, менее чем приблизительно 15 нуклеотидов, менее чем приблизительно 14 нуклеотидов, менее чем приблизительно 13

нуклеотидов, менее чем приблизительно 12 нуклеотидов, менее чем приблизительно 11 нуклеотидов, менее чем приблизительно 10 нуклеотидов, менее чем приблизительно 9 нуклеотидов, менее чем приблизительно 8 нуклеотидов или менее чем приблизительно 7 нуклеотидов). Согласно некоторым вариантам осуществления одна или несколько меток или представителей библиотеки не содержат область связывания праймера и/или константную область (например, во время стадии отбора, такой как отбор с использованием гель-проникающей хроматографии). Согласно некоторым вариантам осуществления одна или несколько меток или представителей библиотеки характеризуются наличием редуцированной константной области (например, длина составляет менее чем приблизительно 30 нуклеотидов, менее чем приблизительно 25 нуклеотидов, менее чем приблизительно 20 нуклеотидов, менее чем приблизительно 19 нуклеотидов, менее чем приблизительно 18 нуклеотидов, менее чем приблизительно 17 нуклеотидов, менее чем приблизительно 16 нуклеотидов, менее чем приблизительно 15 нуклеотидов, менее чем приблизительно 14 нуклеотидов, менее чем приблизительно 13 нуклеотидов, менее чем приблизительно 12 нуклеотидов, менее чем приблизительно 11 нуклеотидов, менее чем приблизительно 10 нуклеотидов, менее чем приблизительно 9 нуклеотидов, менее чем приблизительно 8 нуклеотидов или менее чем приблизительно 7 нуклеотидов). Согласно другим вариантам осуществления способы включают головной фрагмент, который кодирует молекулу, часть химического соединения, реакцию связывания (например, химического или ферментативного лигирования) стадии или идентичность библиотеки, причем кодирующий головной фрагмент устраняет необходимость в дополнительной метке для кодирования такой информации.

В любом из указанных выше вариантов осуществления олигонуклеотид (например, головной фрагмент, первая метка и/или одна или несколько дополнительных меток, если присутствуют) кодирует идентификацию библиотеки. Согласно некоторым вариантам осуществления олигонуклеотид (например, головной фрагмент, первая метка и/или одна или несколько дополнительных меток, если присутствуют) включает в себя последовательность идентификации первой библиотеки, причем последовательность кодирует идентификацию первой библиотеки. Согласно конкретным вариантам осуществления олигонуклеотид представляет собой метку идентификации первой библиотеки. Согласно некоторым вариантам осуществления способ включает обеспечение метки идентификации первой библиотеки, причем метка включает в себя последовательность, которая кодирует первую библиотеку, и/или связывание метки идентификации первой библиотеки с

комплексом. Согласно некоторым вариантам осуществления способ включает получение второй библиотеки и объединение первой библиотеки со второй библиотекой. Согласно другим вариантам осуществления способ включает получение метки идентификации второй библиотеки, причем метка включает в себя последовательность, которая кодирует вторую библиотеку. Согласно некоторым вариантам осуществления объединяют более двух библиотек (например, три, четыре, пять, шесть, семь, восемь, девять, десять или более библиотек).

Согласно любому из указанных выше вариантов осуществления закодированная информация предусмотрена в одной или нескольких метках или в комбинации более чем одной метки. Согласно некоторым вариантам осуществления закодированная информация представлена более чем одной меткой (например, двумя, тремя, четырьмя, пятью, шестью, семью, восемью, девятью, десятью или более метками). Согласно некоторым вариантам осуществления закодированная информация представлена более чем одной меткой, причем все кодирующие метки содержатся в пределах кодирующей последовательности (например, путем использования определенной комбинации меток для кодирования информации). Согласно некоторым вариантам осуществления закодированная информация представлена более чем одной меткой, причем менее чем все кодирующие метки содержатся в пределах кодирующей последовательности (например, путем использования одной метки из набора более чем одной отдельной метки для кодирования в пределах отдельной кодирующей последовательности). Согласно некоторым вариантам осуществления закодированная информация представлена ортогонально, причем закодированная информация представлена путем комбинации более чем одной метки менее чем всей кодирующей информации, содержащейся внутри отдельного представителя библиотеки, таким образом, что более чем один соответствующий представитель библиотеки нуждается в секвенировании для деконволюции закодированной информации. Согласно некоторым вариантам осуществления более чем один химический строительный блок представлен единственной меткой (например, для рацемического строительного блока, такого как два, три, четыре, пять, шесть, семь, восемь, девять, десять или более блоков, представленных одной меткой).

Согласно любому из указанных выше вариантов осуществления олигонуклеотид (например, головной фрагмент и/или один или несколько структурных элементов) кодирует применение представителя библиотеки (например, применение на стадии выбора или стадии связывания, как описано в настоящем документе). Согласно

некоторым вариантам осуществления олигонуклеотид (например, головной фрагмент, первая метка и/или одна или несколько дополнительных меток, если они присутствуют) включает в себя применение последовательности, причем последовательность кодирует применение подмножества представителей в библиотеке в одной или нескольких стадиях (например, стадии выбора и/или стадии связывания). Согласно конкретным вариантам осуществления олигонуклеотид представляет собой метку использования, включающую в себя последовательность использования. Согласно некоторым вариантам осуществления олигонуклеотид (например, головной фрагмент и/или одна или несколько олигонуклеотидных меток) кодирует происхождение представителя библиотеки (например, в определенной части библиотеки). Согласно некоторым вариантам осуществления олигонуклеотид (например, головной фрагмент, первая метка и/или одна или несколько дополнительных меток, если они присутствуют) включает в себя последовательность происхождения (например, случайную или вырожденную последовательность, характеризующуюся длиной, составляющей приблизительно 10, 9, 8, 7 или 6 нуклеотидов), причем последовательность кодирует происхождение представителя в библиотеке, по сравнению с другими представителями библиотеки. Согласно конкретным вариантам осуществления олигонуклеотид представляет собой метку происхождения, включающую в себя последовательность происхождения. Согласно некоторым вариантам осуществления способ дополнительно включает присоединение, связывание или функциональное связывание метки использования и/или метки происхождения с комплексом.

Согласно любому из вариантов осуществления в настоящем документе способы, композиции и комплексы необязательно включают в себя хвостовой фрагмент, причем хвостовой фрагмент включает в себя одно или несколько из последовательности идентификации библиотеки, последовательности применения или последовательности происхождения, как описано в настоящем документе. Согласно конкретным вариантам осуществления способы дополнительно включают присоединение, связывание или функциональное связывание хвостового фрагмента (например, включающего в себя одно или несколько из последовательности идентификации библиотеки, последовательности использования или последовательности происхождения) с комплексом.

Согласно любому из вышеприведенных вариантов осуществления способы, композиции и комплексы или их части (например, головной фрагмент, первая метка

и/или один или несколько дополнительных меток, если они присутствуют) включают в себя модификацию, которая поддерживает растворимость в эмульсиях, восстановленных или неводных (например, органических) условиях. Согласно некоторым вариантам осуществления бифункциональный спейсер, головной фрагмент или одна или несколько меток модифицируется для увеличения растворимости представителя указанной ДНК-кодированной химической библиотеки в органических условиях. Согласно некоторым вариантам осуществления модификация касается одного или нескольких из алкильной цепи, единицы полиэтиленгликоля, разветвленных соединений с положительными зарядами или гидрофобного кольцевой структуры. Согласно некоторым вариантам осуществления модификация включает в себя один или несколько модифицированных нуклеотидов, содержащих гидрофобный фрагмент (например, модифицированные в положениях С5 основания Т или С с алифатическими цепями, например, в 5'-диметокситритил-N4-диизобутиламинометилен-5-(1-пропинил)-2'-дезоксцитидин, 3'-[(2-цианоэтил)-(N,N-диизопропил)]фосфорамидите; 5'-диметокситритил-5-(1-пропинил)-2'-дезоксидеоксиуридин, 3'-[(2-цианоэтил)(N,N-диизопропил)]фосфорамидите; 5'-диметокситритил-5-фторо-2'-дезоксидеоксиуридин, 3'-[(2-цианоэтил)(N,N-диизопропил)]фосфорамидите и 5'-диметокситритил-5-(пирен-1-ил-этинил)-2'-дезоксидеоксиуридине или 3'-[(2-цианоэтил)-(N,N-диизопропил)]фосфорамидите) или вставку, содержащую гидрофобный фрагмент (например, азобензол). Согласно некоторым вариантам осуществления представитель библиотеки характеризуется коэффициентом октанол:вода, равным от приблизительно 1,0 до приблизительно 2,5 (например, от приблизительно 1,0 до приблизительно 1,5, от приблизительно 1,0 до приблизительно 2,0, от приблизительно 1,3 до приблизительно 1,5, от приблизительно 1,3 до приблизительно 2,0, от приблизительно 1,3 до приблизительно 2,5, от приблизительно 1,5 до приблизительно 2,0, от приблизительно 1,5 до приблизительно 2,5 или от приблизительно 2,0 до приблизительно 2,5).

Согласно любому из указанных выше вариантов осуществления головной фрагмент, хвостовой фрагмент, первая метка, одна или несколько дополнительных меток, метка идентификации библиотеки, метка использования и/или метка происхождения, если присутствуют, могут включать в себя от приблизительно 5 до приблизительно 75 нуклеотидов (например, от 5 до 7 нуклеотидов, от 5 до 8 нуклеотидов, от 5 до 9 нуклеотидов, от 5 до 10 нуклеотидов, от 5 до 11 нуклеотидов, от 5 до 12 нуклеотидов, от 5 до 13 нуклеотидов, от 5 до 14 нуклеотидов, от 5 до 15

нуклеотидов, от 16 до 17 нуклеотидов, от 16 до 18 нуклеотидов, от 16 до 19 нуклеотидов, от 16 до 20 нуклеотидов, от 17 до 18 нуклеотидов, от 17 до 19 нуклеотидов, от 17 до 20 нуклеотидов, от 18 до 19 нуклеотидов, от 18 до 20 нуклеотидов, от 19 до 20 нуклеотидов, от 20 до 30 нуклеотидов, от 20 до 40 нуклеотидов, от 20 до 50 нуклеотидов, от 20 до 60 нуклеотидов, от 20 до 70 нуклеотидов, от 20 до 75 нуклеотидов, от 30 до 40 нуклеотидов, от 30 до 50 нуклеотидов, от 30 до 60 нуклеотидов, от 30 до 70 нуклеотидов, от 30 до 75 нуклеотидов, от 40 до 50 нуклеотидов, от 40 до 60 нуклеотидов, от 40 до 70 нуклеотидов, от 40 до 75 нуклеотидов, от 50 до 60 нуклеотидов, от 50 до 70 нуклеотидов, от 50 до 75 нуклеотидов, от 60 до 70 нуклеотидов, от 60 до 75 нуклеотидов и от 70 до 75 нуклеотидов). Согласно конкретному варианту осуществления головной фрагмент, первая метка, вторая метка, одна или несколько дополнительных меток, метка идентификации библиотеки, метка использования и/или метка происхождения, если присутствуют, характеризуются длиной менее чем 20 нуклеотидов (например, менее чем 19 нуклеотидов, менее чем 18 нуклеотидов, менее чем 17 нуклеотидов, менее чем 16 нуклеотидов, менее чем 15 нуклеотидов, менее чем 14 нуклеотидов, менее чем 13 нуклеотидов, менее чем 12 нуклеотидов, менее чем 11 нуклеотидов, менее чем 10 нуклеотидов, менее чем 9 нуклеотидов, менее чем 8 нуклеотидов или менее чем 7 нуклеотидов).

Согласно любому из указанных выше вариантов осуществления кодирующая последовательность (например, головной фрагмент, хвостовой фрагмент, первая метка, одна или несколько дополнительных меток, метка идентификации библиотеки, метка использования и/или метка происхождения, если присутствуют) может включать в себя более 20 нуклеотидов (например, более 25, 30, 35, 40, 45, 50, 55, 60, 65, 70 или 75 нуклеотидов).

Определения

Под "2'-замещенным нуклеотидом" подразумевают нуклеотидное основание, содержащее замещение в 2'-положении рибозы.

Под "приблизительно" подразумевают +/- 10% от приведенного значения.

Под "бифункциональным" подразумевают наличие двух реакционноспособных групп, которые позволяют связывание двух химических фрагментов.

Под "бифункциональным спейсером" подразумевают пространственный фрагмент, содержащий два реакционноспособные группы, которые позволяют связывание химического соединения и кодирующей информации комплекса. В одном

неограничивающем примере предусмотрен бифункциональный спейсер, расположенный между химическим соединением и меткой. В другом неограничивающем примере предусмотрен бифункциональный спейсер, расположенный между химическим соединением и головным фрагментом. Иллюстративные бифункциональные спейсеры предусмотрены в настоящем документе.

Под "связыванием" подразумевают присоединение посредством ковалентной связи или нековалентной связи. Нековалентные связи включают в себя те, которые образуются ван-дер-ваальсовыми силами, водородными связями, ионными связями, захватом или физической инкапсуляцией, абсорбцией, адсорбцией и/или другими межмолекулярными силами. Связывание может быть осуществлено с помощью любых подходящих средств, например, путем ферментативного связывания (например, ферментативного лигирования, чтобы обеспечить ферментативную связь) или путем химического связывания (например, химическим лигированием, чтобы обеспечить химическую связь).

Под "структурным элементом" подразумевается структурная единица химического соединения, причем единица непосредственно связана с другими химическими структурными единицами или косвенно связана через каркасный элемент. Когда химическое соединение представляет собой полимерное или олигомерное, структурные элементы представляют собой мономерные звенья полимера или олигомера. Структурные элементы могут содержать один или несколько узлов разнообразия, которые позволяют добавление одного или нескольких других структурных элементов или каркасных элементов. В большинстве случаев каждый узел разнообразия представляет собой функциональную группу, способную взаимодействовать с одним или несколькими структурными элементами или каркасными элементами с образованием химического соединения. Как правило, структурные элементы содержат по меньшей мере два узла разнообразия (или реакционноспособные функциональные группы), но некоторые структурные элементы могут содержать один узел разнообразия (или реакционноспособную функциональную группу). Кроме того, закодированные химические или связывающие стадии могут включать в себя несколько химических компонентов (например, многокомпонентные реакции конденсации или многоступенчатые процессы). Реакционноспособные группы на двух различных структурных элементах должны быть комплементарны, т.е.,

способны вступать в реакцию друг с другом с образованием ковалентной или нековалентной связи.

Под "химическим соединением" подразумевают соединение, содержащее один или несколько структурных элементов и, необязательно, один или несколько каркасных элементов. Химическое соединение может представлять собой любую небольшую молекулу или пептидное лекарственное средство или кандидатное лекарственное средство, разработанное или созданное, чтобы обладать одной или несколькими желаемыми характеристиками, например, способностью связывать биологическую мишень, растворимостью, наличием доноров и акцепторов водородной связи, вращательными степенями свободы связей, положительным зарядом, отрицательным зарядом и тому подобным. Согласно некоторым вариантам осуществления химическое соединение может быть введено в реакцию дополнительно в качестве бифункционального или трифункционального (или более) соединения.

Под "химической реакционноспособной группой" понимают реакционноспособную группу, которая участвует в модульной реакции, таким образом, создавая связь. Иллюстративные реакции и реакционноспособные группы включают в себя те, которые выбраны из реакции Хьюсена 1,3-диполярного циклоприсоединения с парой необязательно замещенной алкинильной группы и необязательно замещенной азидогруппы; реакции Дильса-Альдера с парой необязательно замещенного диена, характеризующегося системой 4 π -электронов и необязательно замещенного диенофила или необязательно замещенного гетеродиенофила, характеризующегося системой 2 π -электронов; реакции раскрытия кольца с нуклеофилом и напряженным гетероциклическим электрофилом; реакции лигирования накладки с фосфоритоатной группой и группой йода и реакции восстановительного аминирования с альдегидной группой и аминогруппой, как описано в настоящем документе.

Под "комплементарным" подразумевают последовательность, способную к гибридизации, как определено в настоящем документе, с образованием вторичной структуры (дуплекса или двухцепочечной части молекулы нуклеиновой кислоты). Комплементарность не должна быть совершенной, но может включать в себя один или несколько несоответствий в одном, двух, трех или более нуклеотидах. Например, комплементарная последовательность может содержать нуклеотидные основания, которые могут образовывать водородные связи в соответствии с Уотсон-Криковскими правилами спаривания оснований (например, G с C, A с T или A с U) или другими мотивами водородных связей (например, диаминопурин с T, 5-метил-C с G, 2-

тиотимидин с А, инозин с С, псевдоизоцитозин с G). Последовательность и ее комплементарная последовательность могут присутствовать в одном и том же олигонуклеотиде или в различных олигонуклеотидах.

Под "комплексом" или "лигированным комплексом" подразумевают головной фрагмент, который функционально связан с химическим соединением и/или одной или несколькими олигонуклеотидными метками с помощью ковалентной связи или нековалентной связи. Комплекс может дополнительно включать в себя бифункциональный спейсер между химическим соединением и головным фрагментом.

Под "компонентом" химического соединения понимают либо каркасный элемент, либо структурный элемент.

Под "коннектором" олигонуклеотидной метки подразумевают часть метки на или в непосредственной близости от 5'-или 3'-конца, характеризующуюся фиксированной последовательностью. 5'-коннектор расположен на или в непосредственной близости от 5'-конца олигонуклеотида, а 3'-коннектор расположен на или в непосредственной близости от 3'-конца олигонуклеотида. Когда присутствует в комплексе, каждый 5'-коннектор может быть одинаковым или различным, и каждый 3'-коннектор может быть одинаковым или различным. В иллюстративном, не ограничивающем комплексе, содержащем более одной метки, каждая метка может включать в себя 5'-коннектор и 3'-коннектор, причем каждый 5'-коннектор характеризуется одной и той же последовательностью, и каждый 3'-коннектор характеризуется одной и той же последовательностью (например, где последовательность 5'-коннектора может быть такой же или отличаться от последовательности 3'-коннектора). В другом иллюстративном, не ограничивающем комплексе последовательность 5'-коннектора разрабатывается, чтобы быть комплементарной, как определено в настоящем документе, последовательности 3'-коннектора (например, для обеспечения гибридизации между 5'- и 3'-коннекторами). Коннектор может дополнительно включать в себя одну или несколько групп, способствующих образованию связи (например, связи, для которой полимеразы обладает пониженной способностью считывать или перемещаться, например, химической связи).

По "константной" или "фиксированной константной" последовательностью подразумевают последовательность олигонуклеотида, которая не кодирует информацию. Неограничивающие, иллюстративные части комплекса, характеризующегося константной последовательностью, включают в себя

связывающую праймер область, 5'-коннектор или 3'-коннектор. Головной фрагмент согласно настоящему изобретению может кодировать информацию (так, например, метка) или, альтернативно, не кодировать информацию (так, например, константная последовательность). Сходным образом, хвостовой фрагмент согласно настоящему изобретению может кодировать или не кодировать информацию.

Под "сшивающим олигонуклеотидом" подразумевают олигонуклеотид, который функционально связывается, как определено в настоящем документе, в конкретной области соединения между двумя соседними метками в комплексе. В качестве неограничивающего примера, один конец сшивающего олигонуклеотида гибридизуется с 3'-коннектором первой метки, а другой конец сшивающего олигонуклеотида гибридизуется с 5'-коннектором второй метки, который примыкает к первой метке. Иллюстративные, не ограничивающие варианты осуществления сшивающих олигонуклеотидов включают в себя те, которые содержат одну или несколько реакционно-способных групп (например, химическую реакционно-способную группу, фотореакционноспособную группу, интеркалирующий фрагмент или обратимую кореакционноспособную группу, или любую описанную в настоящем документе), которые функционально связываются с соседними метками или коннекторами соседних меток.

Под "узлом разнообразия" подразумевают функциональную группу в положении в каркасном элементе или структурном элементе, которая позволяет добавить другой структурный элемент.

Под "головным фрагментом" подразумевают химическую структуру для синтеза библиотеки, которая функционально связана с компонентом химического соединения и с меткой, например, начального олигонуклеотида. Дополнительно головной фрагмент может содержать мало или не содержать нуклеотиды, но может обеспечивать точку, в которой они могут быть функционально связаны. Необязательно, бифункциональный спейсер соединяет головной фрагмент с компонентом.

Под "гибридизоваться" подразумевают спаривание с образованием двухцепочечной молекулы между комплементарными олигонуклеотидами или их частями при различных условиях жесткости. (Смотрите, например, Wahl, GM и SL Berger (1987) *Methods Enzymol* 152: 399; Kimmel, AP (1987) *Methods Enzymol* 152: 507). Например, высокая жесткость гибридизации может быть получена с концентрацией солей, обычно составляющей менее чем приблизительно 750 мМ NaCl и 75 мМ тринатрийцитрата, менее чем приблизительно 500 мМ NaCl и 50 мМ тринатрийцитрата

или менее чем приблизительно 250 мМ NaCl и 25 мМ тринатрийцитрата. Низкая жесткость гибридизации может быть получена в отсутствие органического растворителя, например, формамида, в то время как высокая жесткость гибридизации может быть получена в присутствии по меньшей мере приблизительно 35% формамида или по меньшей мере приблизительно 50% формамида. Температурные условия гибридизации высокой жесткости будут, как правило, включать в себя температуру, составляющую по меньшей мере приблизительно 30°C, 37°C или 42°C. Изменение дополнительных параметров, таких как время гибридизации, концентрация детергента, например, додецилсульфата натрия (ДСН), и включение или исключение ДНК-носителя, хорошо известны специалистам в настоящей области техники. Различных уровней жесткости достигают путем сочетания этих различных условий, как это необходимо. Согласно одному варианту осуществления гибридизация будет происходить при 30°C в NaCl концентрацией 750 мМ, тринатрийцитрате концентрацией 75 мМ и 1% ДСН. Согласно альтернативному варианту осуществления гибридизация будет происходить при 37°C в NaCl концентрацией 500 мМ, тринатрийцитрате концентрацией 50 мМ, 1% ДСН, 35% формамиде и 100 мкг/мл денатурированной ДНК спермы лосося (оцДНК). Согласно другому альтернативному варианту осуществления гибридизация будет происходить при 42°C в NaCl концентрацией 250 мМ, тринатрийцитрате концентрацией 25 мМ, 1% ДСН, 50% формамиде и 200 мкг/мл оцДНК. Полезные вариации этих условиях будут очевидны специалистам в настоящей области техники.

Для большинства заявок стадии промывания, которые следуют после гибридизации, также будут варьировать в жесткости. Условия жесткости промывания могут быть определены концентрацией соли и температурой. Как указано выше, жесткость промывания может быть увеличена путем уменьшения концентрации соли или повышения температуры. Например, концентрации солей высокой жесткости для стадий промывания может составлять, например, менее чем приблизительно 30 мМ NaCl и 3 мМ тринатрийцитрат, или менее чем приблизительно 15 мМ NaCl и 1,5 мМ тринатрийцитрат. Температурные условия высокой жесткости для стадий промывания будут, как правило, включать в себя температуру, равную, например, по меньшей мере приблизительно 25°C, 42°C или 68°C. Согласно одному варианту осуществления стадии промывания будут происходить при температуре 25°C в 30 мМ NaCl, 3 мМ тринатрийцитрате и 0,1% ДСН. Согласно альтернативному варианту осуществления стадии промывания будут происходить при 42°C в 15 мМ NaCl, 1,5 мМ

тринатрийцитрате и 0,1% ДСН. Согласно дополнительному альтернативному варианту осуществления стадии промывания будут происходить при 68°C в 15 mM NaCl, 1,5 mM тринатрийцитрате и 0,1% ДСН. Дополнительные вариации этих условиях будут очевидны специалистам в настоящей области техники. Техники гибридизации хорошо известны специалистам в настоящей области техники и описаны, например, в Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel et al. (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); and Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

Под "интеркалирующим фрагментом" подразумевают реакционноспособную группу, которая приводит к включению фрагмента между двумя или более нуклеотидами. В качестве неограничивающего примера интеркалирующий фрагмент вступает в реакцию с одним или несколькими нуклеотидами с образованием меж- или внутринитевых сшивок между дуплексными или триплексными олигонуклеотидами. Иллюстративные, не ограничивающие интеркалирующие фрагменты описаны в настоящем документе.

Под "областью соединения" подразумевают однонитевый разрыв (отсутствие межнуклеотидной связи) или пробел (отсутствие одного или нескольких нуклеотидов) между двумя соседними метками в комплексе. Область соединения также может быть между двумя соседними коннекторами, присутствующими в двух смежных метках (например, между 3'-коннектором первой метки и 5'-коннектором второй метки, которая примыкает к первой метке).

Под "библиотекой" подразумевают коллекцию молекул или химических соединений. Дополнительно, молекулы или химические соединения связаны с одним или несколькими олигонуклеотидами, которые кодируют молекулы или части химического соединения.

Под "связью" подразумевают химическое соединяющее соединение, которое позволяет функциональное связывания двух или более химических структур, причем связь присутствует между головным фрагментом и одной или несколькими метками, между двумя метками или между меткой и хвостовым фрагментом. Химическое соединяющее соединение может представлять собой нековалентную связь (например, как описано в настоящем документе), ковалентную связь или продукт реакции между двумя функциональными группами. Под "химической связью" подразумевают связь,

образованную неферментативной химической реакцией между двумя функциональными группами. Иллюстративные, не ограничивающие функциональные группы включают в себя химическую реакционноспособную группу, фотореакционноспособную группу, интеркалирующий фрагмент или сшивающий олигонуклеотид (например, как описано в настоящем документе). Под "ферментативной связью" подразумевают межнуклеотидную или межнуклеозидную связь, образованную с помощью фермента. Иллюстративные, не ограничивающие ферменты включают в себя киназу, полимеразу, лигазу или их комбинации. Под связью, "для которой полимеразы обладает пониженной способностью считывать или перемещаться через", подразумевают связь, которая, когда присутствует в олигонуклеотидном шаблоне, обеспечивает уменьшенное количество удлинённых и/или амплифицированных продуктов с помощью полимеразы, по сравнению с контрольным олигонуклеотидом без связи. Иллюстративные, не ограничивающие способы определения такой связи включают в себя удлинение праймера, оцениваемое с помощью анализа ПЦР (например, количественной ПЦР), анализа ОТ-ПЦР, жидкостной хроматографии-масс-спектрометрии, демографических анализов последовательности или другими способами. Иллюстративные, не ограничивающие полимеразы включают в себя ДНК-полимеразы и РНК-полимеразы, такие как ДНК-полимераза I, ДНК-полимераза II, ДНК-полимераза III, ДНК-полимераза VI, Taq ДНК-полимераза, Deep VentR™ ДНК-полимераза (высокоточная термофильная ДНК-полимераза, доступная от New England Biolabs), T7 ДНК-полимераза, T4 ДНК-полимераза, РНК-полимераза I, РНК-полимераза II, РНК-полимераза III или РНК-полимераза T7.

Под "поливалентным катионом" подразумевают катион, способный к образованию более одной связи с более чем одним лигандом или анионом. Поливалентный катион может образовывать либо ионный комплекс, либо координационный комплекс. Иллюстративные поливалентные катионы включают в себя катионы из щелочно-земельных металлов (например, магния) и переходных металлов (например, марганца (II) или кобальта (III)), и те, которые необязательно связаны с одним или несколькими анионами и/или одним или несколькими одновалентными или полидентатными лигандами, такими как хлорид, амин и/или этилендиамин.

Под "олигонуклеотидом" подразумевают полимер нуклеотидов, содержащий 5'-конец, 3'-конец и один или несколько нуклеотидов во внутреннем положении между 5'-

и 3'-концами. Олигонуклеотид может включать в себя ДНК, РНК или любое производное, известное в настоящей области техники, которое может быть синтезировано и использовано для распознавания пар оснований. Олигонуклеотид не должен содержать смежные основания, но может содержать в промежутках линкерные фрагменты. Олигонуклеотидный полимер и нуклеотид (например, модифицированная ДНК или РНК) может включать в себя природные основания (например, аденозин, тимидин, гуанозин, цитидин, уридин, дезоксиаденозин, дезокситимидин, дезоксигуанозин, дезоксицитидин, инозин или диаминопурин), аналоги оснований (например, 2-аминоаденозин, 2-тиотимидин, инозин, пирролопиримидин, 3-метиладенозин, C5-пропинилцитидин, C5-пропинилуридин, C5-бромурин, C5-фторуридин, C5-йодуридин, C5-метилцитидин, 7-дезааденозин, 7-дезагуанозин, 8-оксоаденозин, 8-оксогуанозин, O-(6)-метилгуанин и 2-тиоцитидин), модифицированные основания (например, 2'-замещенные нуклеотиды, такие как 2'-O-метилированные основания и 2'-фтор основания), интеркалированные основания, модифицированные сахара (например, 2'-фторрибоза; рибоза; 2'-дезоксирибоза; арабиноза; гексоза; ангидрогекситол; альтрит; манит; циклогексанол; циклогексенол; морфолино, который также содержит фосфорамидатный скелет; замкнутые нуклеиновые кислоты (LNA, например, когда 2'-гидроксил рибозы соединен с помощью C₁₋₆ алкиленового или C₁₋₆ гетероалкиленового мостика с 4'-углеродом того же рибозного сахара, причем иллюстративные мосты включают в себя метилен, пропилен, эфир или аминокислотные мостики); гликольную нуклеиновую кислоту (GNA, например, R-GNA или S-GNA, где рибоза заменена гликольными единицами, присоединенными к фосфодиэфирным связям); треозную нуклеиновую кислоту (TNA, где рибоза заменена α-L-треофуранозил-(3'→2')); и/или замену кислорода в рибозе (например, на S, Se или алкилен, такой как метилен или этилен)), модифицированные остовы (например, пептидная нуклеиновая кислота (ПНК), где 2-аминоэтилглициновые связи заменяют рибозу и фосфодиэфирный остов), и/или модифицированные фосфатные группы (например, фосфоротиоаты, 5'-N-фосфорамидиты, селенаты фосфора, фосфаты борана, эфиры фосфатов борана, фосфонаты водорода, фосфорамидаты, фосфордиамидаты, алкильные или арильные фосфонаты, фосфотриэфиры, фосфорамидаты с перемычкой, фосфоротиоаты с перемычкой и метиленфосфонаты с перемычкой). Олигонуклеотид может быть одноцепочечным (например, шпилька), двухцепочечным или характеризоваться другой вторичной или

третичной структурой (например, структурой типа "стебель-петля", двойные спирали, триплексы, квадруплексы и т.д.).

Под "функционально связанный" или "функционально ассоциированный" подразумевают, что две или более химических структур прямо или опосредованно связаны друг с другом таким образом, что остаются связанными после различных манипуляций, которым они должны подвергнуться. Как правило, химическое соединение и головной фрагмент функционально связаны опосредованным образом (например, ковалентно с помощью соответствующего спейсера). Например, спейсер может представлять собой бифункциональный фрагмент с сайтом прикрепления для химического соединения и сайтом прикрепления к головному фрагменту.

Под "фотореакционноспособной группой" подразумевают реакционноспособную группу, которая участвует в реакции, вызванной поглощением ультрафиолетовой, видимой или инфракрасной радиации, тем самым производя связь. Иллюстративные, не ограничивающие фотореакционноспособные группы описаны в настоящем документе.

Под "защитной группой" подразумевается группа, предназначенная для защиты 3'-конца или 5'-конца олигонуклеотида или для защиты одной или нескольких функциональных групп химического соединения, каркасного элемента или структурного элемента в отношении нежелательных реакций во время одной или нескольких связывающих стадий получения, мечения или использования закодированной в олигонуклеотидах библиотеки. Обычно используемые защитные группы раскрыты в Greene, "Protective Groups in Organic Synthesis," 4th Edition (John Wiley & Sons, New York, 2007), включенной в настоящий документ посредством ссылки. Иллюстративные защитные группы для олигонуклеотидов включают в себя необратимые защитные группы, такие как дидезоксинуклеотиды и дидезоксинуклеозиды (ddNTP или ddN), и, более предпочтительно, обратимые защитные группы для гидроксильных групп, такие как эфирные группы (например, эфир O-(α -метоксиэтила), эфир O-изовалерила и эфир O-левулинила), тритиловые группы (например, диметокситритил и монометокситритил), ксантениловые группы (например, 9-фенилксантен-9-ил, и 9-(*p*-метоксифенил)ксантен-9-ил), ацильные группы (например, феноксиацетил и ацетил) и силильные группы (например, трет-бутилдиметилсилил). Иллюстративные, не ограничивающие защитные группы для химических соединений, каркасов и строительных блоков включают в себя N-защитные группы для защиты аминогруппы от нежелательных реакций во время процедуры синтеза (например, ацильные; арилоильные; карбамоильные группы, такие

как формил, ацетил, пропионил, пивалоил, трет-бутилацетил, 2-хлорацетил, 2-бромацетил, трифторацетил, трихлорацетил, фталил, о-нитрофеноксиацетил, α-хлорбутирил, бензоил, 4-хлорбензоил, 4-бромбензоил, 4-нитробензоил и хиральные вспомогательные средства, такие как защищенные или незащищенные D, L или D, L-аминокислоты, такие как аланин, лейцин, фенилаланин и тому подобные; содержащие сульфонил группы, такие как бензолсульфонил, п-толуолсульфонил и тому подобные; образующие карбамат группы, такие как бензилоксикарбонил, п-хлорбензилоксикарбонил, п-метоксибензилоксикарбонил, п-нитробензилоксикарбонил, 2-нитробензилоксикарбонил, п-бромбензилоксикарбонил, 3,4-диметоксибензилоксикарбонил, 3,5-диметоксибензилоксикарбонил, 2,4-диметоксибензилоксикарбонил, 4-метоксибензилоксикарбонил, 2-нитро-4,5-диметоксибензилоксикарбонил, 3,4,5-триметоксибензилоксикарбонил, 1-(п-бифенилил)-1-метилэтоксикарбонил, α,α-диметил-3,5-диметоксибензилоксикарбонил, бензгидрилоксикарбонил, трет-бутилоксикарбонил, диизопропилметоксикарбонил, изопропоксикарбонил, этоксикарбонил, метоксикарбонил, аллилоксикарбонил, 2,2,2-трихлорэтоксикарбонил, феноксикарбонил, 4-нитрофеноксикарбонил, флуоренил-9-метоксикарбонил, циклопентилоксикарбонил, адамантилоксикарбонил, циклогексиллоксикарбонил, фенилтиокарбонил и тому подобные; алкарильные группы, такие как бензил, трифенилметил, бензилоксиметил и тому подобные; и силильные группы, такие как триметилсилил и тому подобные; причем предпочтительные N-защитные группы представляют собой формил, ацетил, бензоил, пивалоил, трет-бутилацетил, аланил, фенилсульфонил, бензил, трет-бутилоксикарбонил (Boc) и бензилоксикарбонил (Cbz)); O-защитные группы для защиты гидроксильной группы от нежелательных реакций в процессе синтетической процедуры (например, алкилкарбонильные группы, такие как ацил, ацетил, пивалоил и тому подобные; необязательно замещенные арилкарбонильные группы, такие как бензоил; силильные группы, такие как триметилсилил (TMS), трет-бутилдиметилсилил (TBDMS), триизопропилсилилоксиметил (TOM), триизопропилсилил (TIPS) и тому подобные; образующих простые эфиры группы с гидроксилом, такие как метил, метоксиметил, тетрагидропиранил, бензил, п-метоксибензил, тритил и тому подобные; алкоксикарбонилы, такие как метоксикарбонил, этоксикарбонил, изопропоксикарбонил, н-изопропоксикарбонил, н-бутилоксикарбонил, изобутилоксикарбонил, втор-бутилоксикарбонил, трет-бутилоксикарбонил, 2-этилгексиллоксикарбонил, циклогексиллоксикарбонил, метилоксикарбонил и тому

подобные; алкоксиалкоксиалкильные группы, такие как метоксиметоксикарбонил, этоксиметоксикарбонил, 2-метоксиэтоксикарбонил, 2-этоксииэтоксикарбонил, 2-бутоксииэтоксикарбонил, 2-метоксиэтоксиметоксикарбонил, аллилоксикарбонил, пропаргилоксикарбонил, 2-бутеноксикарбонил, 3-метил-2-бутеноксикарбонил и тому подобные; галоалкоксикарбонилы, такие как 2-хлорэтоксикарбонил, 2-хлорэтоксикарбонил, 2,2,2-трихлорэтоксикарбонил и тому подобные; необязательно замещенные арилалкоксикарбонильные группы, такие как бензилоксикарбонил, п-метилбензилоксикарбонил, п-метоксибензилоксикарбонил, п-нитробензилоксикарбонил, 2,4-динитробензилоксикарбонил, 3,5-диметилоксикарбонил, п-хлорбензилоксикарбонил, п-бромбензилоксикарбонил и тому подобные; и необязательно замещенные арилоксикарбонильные группы, такие как феноксикарбонил, п-нитрофеноксикарбонил, о-нитрофеноксикарбонил, 2,4-динитрофеноксикарбонил, п-метилфеноксикарбонил, м-метилфеноксикарбонил, о-бромфеноксикарбонил, 3,5-диметилфеноксикарбонил, п-хлорфеноксикарбонил, 2-хлор-4-нитрофеноксикарбонил и тому подобные); карбонильные защитные группы (например, ацетальные и кетальные группы, такие как диметилформамид, 1,3-диоксолан и тому подобные; ацилальные группы, а также дитиановые группы, такие как 1,3-дитианы, 1,3-дитиолан и тому подобные); защитные группы – карбоксильные кислоты (например, сложноэфирные группы, такие как метиловый эфир, бензиловый эфир, трет-бутиловый эфир, ортоэфиры и тому подобные; силильные группы, такие как триметилсиллил, а также любые описанные в настоящем документе, и оксазолиновые группы); и фосфатные защитные группы (например, необязательно замещенные эфирные группы, такие как метиловый эфир, изопропиловый эфир, 2-цианоэтиловый эфир, аллиловый эфир, трет-бутиловый эфир, бензиловый эфир, флуоренилметиловый эфир, 2-(триметилсиллил)этиловый эфир, 2-(метилсульфонил)этиловый эфир, 2,2,2-трихлорэтиловый эфир и 3',5'-диметоксибензоиновый эфир, п-гидроксифенацилбромидный эфир и т.п.).

Под "близостью" или "в непосредственной близости" к концу олигонуклеотида подразумевают вблизи или ближе к указанному концу, по сравнению с другим оставшимся концом. Например, фрагмент или группа в непосредственной близости от 3'-конца олигонуклеотида находится рядом или ближе к 3'-концу, чем 5'-концу. Согласно конкретным вариантам осуществления фрагмент или группа в непосредственной близости от 3'-конца олигонуклеотида находится в одном, двух, трех, четырех, пяти, шести, семи, восьми, девяти, десяти, пятнадцати или более

нуклеотидах от 3'-конца. Согласно другим вариантам осуществления фрагмент или группа в непосредственной близости от 5'-конца олигонуклеотида находится в одном, двух, трех, четырех, пяти, шести, семи, восьми, девяти, десяти, пятнадцати или более нуклеотидах от 5'-конца.

Под "очисткой" подразумевают удаление любого непрореагировавшего продукта или любого средства, присутствующего в реакционной смеси, который может снизить активность химического или биологического средства, который будет использоваться на последующей стадии. Очистка может включать в себя одно или несколько из хроматографического разделения, электрофоретического разделения и осаждения непрореагировавшего продукта или реагента, который должен быть удален.

Под "обратимой кореакционноспособной группой" подразумевают реакционноспособную группу, которая участвует в обратимой реакции. Иллюстративные, не ограничивающие реакционноспособные группы включают в себя фотореакционноспособные группы, причем воздействие определенным абсорбционным излучением приводит к образованию связи между фотореакционноспособной группой и воздействию различным, конкретным абсорбционным излучением приводит к расщеплению образованной связи (например, циановинилкарбазольной группы, циановинильной группы и акриламидной группы). Другая иллюстративная, не ограничивающая реакционноспособная группа включает в себя окислительно-восстановительные реакционноспособные группы, причем такие группы могут быть обратимо восстановлены или окислены (например, тиольная группа).

Под "каркасным элементом" подразумевают химический фрагмент, который отображает один или несколько узлов разнообразия в определенной специальной геометрии. Узлы разнообразия, как правило, прикрепляются к каркасному элементу во время синтеза библиотеки, но в некоторых случаях один узел разнообразия может быть прикреплен к каркасному элементу до синтеза библиотеки (например, добавления одного или нескольких структурных элементов и/или одной или нескольких меток). Согласно некоторым вариантам осуществления каркасный элемент дериватизируется таким образом, что он может быть ортогонально незащищен при синтезе библиотеки, а затем подвергнут взаимодействию с различными узлами разнообразия.

Под "низкомолекулярным" лекарственным средством или "низкомолекулярным" кандидатным лекарственным средством подразумевают молекулу, которая характеризуется молекулярной массой ниже приблизительно 1000 дальтон. Небольшие молекулы могут быть органическими или неорганическими, выделенными (например,

из библиотек соединений или природных источников) или получены путем дериватизации известных соединений.

Под "существенной идентичностью" или "по существу идентичными" подразумевают полипептидную или полинуклеотидную последовательность, которая содержит ту же полипептидную или полинуклеотидную последовательность, соответственно, что и эталонная последовательности, или содержит определенный процент аминокислотных остатков или нуклеотидов, соответственно, которые такие же в соответствующем месте в эталонной последовательности, когда эти две последовательности оптимально выровнены. Например, аминокислотная последовательность, которая "по существу идентична" эталонной последовательности характеризуется по меньшей мере 50%, 60%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% или 100% идентичности по отношению к эталонной аминокислотной последовательности. Для полипептидов длина сравниваемых последовательностей будет, как правило, составлять по меньшей мере 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 или 20 примыкающих аминокислот, более предпочтительно, по меньшей мере 25, 50, 75, 90, 100, 150, 200, 250, 300 или 350 примыкающих аминокислот и, наиболее предпочтительно, полноразмерную аминокислотную последовательность. Для нуклеиновых кислот длина сравниваемых последовательностей будет, как правило, составлять по меньшей мере 5 примыкающих нуклеотидов, предпочтительно, по меньшей мере 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24 или 25 примыкающих нуклеотидов, а наиболее предпочтительно полноразмерную нуклеотидную последовательность. Идентичность последовательности может быть измерена с использованием программного обеспечения для анализа последовательностей на настройках по умолчанию (например, Пакет программного обеспечения анализа последовательностей от Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, WI 53705). Такое программное обеспечение может находить соответствия сходных последовательностей путем присвоения степени гомологии различным заменам, делециям и другим модификациям.

Под "меткой" или "олигонуклеотидной меткой" подразумевается олигонуклеотидная часть библиотеки, по меньшей мере часть из которой кодирует информацию. Не ограничивающие примеры такой информации включают в себя добавление (например, путем реакции связывания) компонента (например, каркасного элемента или структурного элемента, как в метке каркасного элемента или метке

структурного элемента, соответственно), головной фрагмент в библиотеке, идентификацию библиотеки (т.е., в качестве метки идентификации), использования библиотеки (т.е., как в метке использования) и/или происхождения представителя библиотеки (т.е., как в метке происхождения).

Под "хвостовым фрагментом" подразумевают олигонуклеотидную часть библиотеки, которая присоединяется к комплексу после добавления всех предыдущих меток и кодирует идентификацию библиотеки, использование библиотеки и/или происхождение представителя библиотеки.

Под "праймером" подразумевают олигонуклеотид, который способен к отжигу олигонуклеотидного шаблона, а затем к расширению с помощью полимеразы зависимым от шаблона образом.

Под "релейным праймером" подразумевают олигонуклеотид, способный к отжигу олигонуклеотидного шаблона, который содержит, в области шаблона, с которым праймер гибридизуют, по меньшей мере одну межнуклеотидную связь, которая снижает способность полимеразы считывать или перемещаться. После гибридизации один или несколько релейных праймеров позволяют расширение с помощью полимеразы зависимым от шаблона образом.

Используемый в настоящем документе термин "рекомбинация" означает создание полимеразного продукта в результате по меньшей мере двух различных событий гибридизации.

Другие особенности и преимущества настоящего изобретения будут очевидны из следующего подробного описания и формулы изобретения.

Краткое описание чертежей

На фигурах 1A-1B показаны иллюстративные комплексы. На фиг. 1A показан комплекс (в направлении от 5' к 3'), содержащий химическое соединение (звезда), фиксированную константную последовательность (например, головной фрагмент), затем три переменные кодирующие последовательности (например, три метки) и другую фиксированную константную последовательность (например, хвостовой фрагмент), причем связь между фиксированными константными последовательностями и переменными кодирующими последовательностями читаема. Комплекс может быть дополнительно расширен с использованием полимеразы с праймером (гибридизованным с 3'-концом комплекса), дополнительно амплифицирован с ПЦР и секвенирован. На фиг. 1B показан комплекс, содержащий нечитаемые связи. Комплекс

(в направлении от 5' к 3') содержит химическое соединение (звезда), фиксированную константную последовательность (например, головной фрагмент), затем три переменные кодирующие последовательности (например, три метки), и другую фиксированную константную последовательность (например, хвостовой фрагмент), причем связь между фиксированными константными последовательностями и переменными кодирующими последовательностями нечитаема. Каждая из переменных кодирующих последовательностей включает в себя 5'-коннектор и 3'-коннектор, которые представляют собой фиксированные константные последовательности. Поскольку связи представляют собой нечитаемые, используют релейные праймеры, чтобы охватить связи и позволить расширение с помощью полимеразы, таким образом, образуя олигонуклеотидные фрагменты. Фрагменты затем лигируют с использованием лигазы, дополнительно амплифицируют с ПЦР и секвенируют. Процесс считывания (например, дополнительное расширение, лигирование, дополнительная амплификация и секвенирование) может быть дополнительно полностью выполняться после отбора. Альтернативно, дополнительная часть расширения и лигирования процесса считывания может быть выполнена до отбора.

На фигуре 2 показан иллюстративный комплекс, содержащий нечитаемую связь, образованную псораленом. На схеме представлена (в направлении от 5' к 3') переменная кодирующая последовательность, 3'-коннектор, 5'-коннектор и другая переменная кодирующая последовательность. 3'-коннектор и 5'-коннектор гибридизуются с образованием дуплекса, и псорален (интеркалирующий, фотореакционноспособный фрагмент) образует нечитаемую связь между 5'- и 3'-коннекторами. Релейный праймер гибридизуется с 5'- и 3'-коннекторами и охватывает нечитаемую связь. Считываемые части комплекса (например, переменная кодирующая последовательность) расширяются с помощью полимеразы с образованием олигонуклеотидных фрагментов. Фрагменты и релейные праймеры лигируют с использованием лигазы, дополнительно амплифицируют с помощью ПЦР и секвенируют. Процесс считывания (например, расширение, лигирование, дополнительная амплификация и секвенирование) может быть дополнительно полностью выполняться после отбора. Альтернативно, часть расширения и лигирования процесса считывания может быть выполнена до отбора.

На фигуре 3 показан иллюстративный комплекс, содержащий нечитаемые связи, образованные сшивающим олигонуклеотидом. В верхней части схемы комплекс

включает в себя (в направлении от 5' к 3') химическое соединение (звезда), фиксированную константную последовательность (например, головной фрагмент), сшивающий олигонуклеотид, который сшит с помощью обратимой кореакционноспособной группы (например, циановинилкарбазольные группы сшиты приблизительно при 366 нм, поперечные связи обозначены X), а затем переменную кодирующую последовательность, содержащую фиксированные константные последовательности на 5'- и 3'-концах (например, метка, содержащая 5'- и 3'-коннекторы). Сочетание сшивающего олигонуклеотида и переменной кодирующей последовательности повторяется еще три раза. Затем продолжается конечный сшивающий олигонуклеотид и другая фиксированная константная последовательность (например, хвостовая часть). Поперечные шивки отмечены X. Далее, комплекс подвергают химическому процессу и/или ферментативному процессу (например, действию лигазы и, возможно, киназы) для лигирования меток. Далее, сшивающие олигонуклеотиды высвобождают, чтобы образовать шаблон (например, с использованием абсорбции при приблизительно 312 нм). И, наконец, шаблон дополнительно амплифицируют с помощью ПЦР и секвенируют. Процесс считывания (например, дополнительная стадия киназы, лигирование, дополнительная амплификация и секвенирование) может быть дополнительно полностью выполнен после отбора. Кроме того, дополнительная стадия киназы и стадия лигирования процесса считывания может быть выполнена до отбора.

На фигуре 4 показана иллюстративная обратимая реакция с обратимыми кореакционноспособными группами (парой циановинилкарбазольной группы и тимидина), и использование этих групп делает возможным образование шивков между сшивающим олигонуклеотидом и фиксированной константной последовательностью олигонуклеотидной метки. Обратимые кореакционноспособные группы (обозначенные X и X') присутствуют в 3'-коннекторе, 5'-коннекторе и сшивающем олигонуклеотиде.

На фигуре 5 показан олигонуклеотид CNVKJ, содержащий две модификации циановинилкарбазола CNVK2_P_TagB, CNVK_2_NP_TagB и CNVK2_TagA.

На фигуре 6 показаны результаты электрофореза продуктов фотохимического сшивания олигонуклеотидов CNVKJ, CNVK2_P_TagB и CNVK2_TagA.

На фигуре 7 показана ЖХМС очищенного CNVK2_P_TagB и CNVK2_TagA.

На фигуре 8 показана ЖХМС восстановления исходного Double_CNVK_template.

На фигуре 9 показана ЖХМС воздействия комплекса CNVK2_P_TagB, CNVK2_TagA и CNVKJ с T4 ДНК-лигазой.

На фигуре 10 показаны результаты ЖХМС реакции фосфорилирования-лигирования фотохимически сшитой нефосфорилированной пары модельных меток CNVK2_NP_TagB с CNVK2_TagA и CNVKJ.

На фигуре 11 показан олигонуклеотид 5PSO2_A9_TA, модифицированный с C2-псораленом на 5'-конце, и олигонуклеотид PSO_HP_A9_TCT.

На фигуре 12 показана ЖХМС фотохимического олигонуклеотидного сшивания с использованием псоралена в стволе с модельным одним событием мечения олигонуклеотида 5PSO2_A9_TA, модифицированного с C2-псораленом на 5'-конце, и олигонуклеотида PSO_HP_A9_TCT.

На фигуре 13 показаны олигонуклеотиды Tag1_PsoCVU и SplintC_PsoC2, и олигонуклеотид 5PSOC2_A9_GA (модифицированный на 5'-конце с C2 псораленом).

На фигуре 14 показана схема фотолигирования псоралена.

На фигуре 15 показан гель зависимости от времени фотолигирования псоралена.

На фигуре 16 показана ЖХМС фотолигированного конъюгата 5PsoC2_A9_GA и Tag1_PsoCVU.

На фигуре 17 показан 5'-биотинилированный олигонуклеотид 5Bio_Tag_PsoCVU и олигонуклеотид с 5'-псораленом 5PsoC2_A9_GA вместе с наложенным олигонуклеотидом SplintC_PsoC2.

На фигуре 18 показана ЖХМС из псоралена-тимидина, фотохимически лигированного олигонуклеотидного конъюгата (Bio_Pso).

На фигуре 19 показана ЖХМС кДНК, полученная от Bio_Pso с использованием удлинения и лигирования терминального и нетерминального праймера.

На фигуре 20 показана ЖХМС контрольной кДНК без лигирования, полученная из Bio_Pso с использованием удлинения терминального и нетерминального праймера.

На фигуре 21 показаны олигонуклеотиды CVU_G и CVU_A и олигонуклеотиды Tag1_PsoCVU, SplintC_CVU и SplintA_CVU.

На фигуре 22 показана структура области соединения фотохимического лигирования, образованная между облученным 5'-5-(карбоксивинил)-2'-дезоксисуридином и 3'-тимидином.

На фигуре 23 показан гель зависимости от времени фотолигирования CVU при 365 нм.

На фигуре 24 показана ЖХМС продукта фотохимического лигирования CVU_G с Tag1_PsoCVU.

На фигуре 25 показана ЖХМС фотохимически лигированного олигонуклеотидного конъюгата карбоксивинилирудин-тимидина (Bio_CVU).

На фигуре 26 показана ЖХМС шаблона Bio_CVU минус один нуклеотид dA.

На фигуре 27 показаны олигонуклеотиды S_IDTN3 и S_IDTalkyne.

На фигуре 28 показан анализ ЖХМС геля очищенного конъюгата S_IDTN3 и S_IDTalkyne.

На фигуре 29 показаны олигонуклеотиды TKR_Central и S_IDTN3 и модифицированный 5'-дибензоциклооктином (DBCO) олигонуклеотид TKR_DBCO_S.

На фигуре 30 показана ЖХМС зависимости от времени не содержащей меди клик-реакции между TKR_DBCO_S и TKR_Central.

На фигуре 31 показана ЖХМС конъюгата 2cl_s.

На фигуре 32 показаны Conjugate_Click_S и Conjugate_Click_L.

На фигуре 33 показано опосредованное зависимой от накладки рекомбинацией создание кДНК/ПЦР-амплификация Conjugate_Click_L и Conjugate_Click_S с помощью Deep Vent (exo-) полимеразы в присутствии и в отсутствие ePsplint. 24 цикла, начальная концентрация конъюгата 200 пМ. 1 - маркер; 2 – без шаблона; 3 - без шаблона + ePsplint; 4- Conjugate_Click_S; 5 - Conjugate_Click_S + ePsplint; 6 - Conjugate_Click_L.

На фигуре 34 показан предложенный механизм опосредованного зависимой от накладки рекомбинацией создания кДНК/ПЦР-амплификации.

На фигуре 35 показано опосредованное зависимой от накладки рекомбинацией создание кДНК/ПЦР-амплификация Conjugate_Click_S и Conjugate_Click_L с помощью Deep Vent (exo-) полимеразы как отдельно, так и смешанных вместе в свободном растворе. Все реакции с добавлением ePsplint. 1-маркер; 2 – без шаблона; 3- Conjugate_Click_S; 4 - Conjugate_Click_L; 5 - Conjugate_Click_S и Conjugate_Click_L.

На фигуре 36 показан возможный механизм рекомбинации продуктов опосредованного зависимой от накладки рекомбинацией создания кДНК/ПЦР-амплификации в свободном растворе с последующей перестановкой (потерей) ассоциаций меток (закодированной информации).

На фигуре 37 показано опосредованное рекомбинацией создание кДНК/ПЦР-амплификация смеси Conjugate_Click_S и Conjugate_Click_L с помощью Deep Vent (exo-) полимеразы в свободном растворе и эмульгированных водных компартментах со следующими условиями: 30 секунд при 94°C, 30 секунд при 52°C и 60 секунд при 72°C

в течение 35 циклов, и начальной концентрацией конъюгата 5 пМ; 1-маркер; 2- свободный раствор, 3- эмульгированные водные компартменты.

На фигуре 38 показаны конъюгаты TKR_S и TKR_L и их ЖХМС следы.

На фигуре 39 показано зависимое от повтора Deep Vent (экзо-) полимеразы опосредованное рекомбинацией создание кДНК/ПЦР-амплификация TKR_S и TKR_L. 1- маркер, 2- без шаблона, 3- TKR_S, 4- TKR_L, 5- ТКР L и TKR_S. Условия: 5 мкл 10x буфера Thermopol; 2,5 мкл 10 мкМ раствора каждого прямого и обратного праймера; 2,5 мкл 10 мМ смеси dNTP (NEB); конечная концентрация 40 пМ либо конъюгата, либо их смеси 1:1, 2,5 мкл Deep Vent (экзо-) полимеразы и вода до 50 мкл. Реакционная смесь циркулировала 21 раз следующим образом: 30 секунд при 94°C, 30 секунд при 52°C и 60 секунд при 72°C.

На фигуре 40 показаны данные последовательности, полученные из клонированного продукта зависимого от повтора опосредованного рекомбинацией создания кДНК/ПЦР-амплификация смешанных TKR_S и TKR_L в свободном растворе.

На фигуре 41 показаны конъюгаты TKR_2_click_S и TKR_2_click_L и их ЖХМС следы.

На фигуре 42 показано зависимое от повтора Deep Vent (экзо-) полимеразы опосредованное рекомбинацией создание кДНК/ПЦР-амплификация 2_click_S и 2_click_L в свободном растворе. 1 - маркер; 2 - контроль без шаблона; 3- 2_click_S; 4- 2_click_L; 5- смесь 1:1 2_click_S и 2_click_L. Условия: 5 мкл 10x буфера Thermopol; 2,5 мкл 10 мкМ раствора каждого прямого и обратного праймера; 2,5 мкл 10 мМ смеси dNTP (NEB); конечная концентрация 40 пМ либо конъюгата, либо их смеси 1:1, 2,5 мкл Deep Vent (экзо-) полимеразы и вода до 50 мкл. Реакционная смесь циркулировала 22 раза следующим образом: 30 секунд при 94°C, 30 секунд при 52°C и 60 секунд при 72°C.

На фигуре 43 показаны данные последовательности, полученные из клонированного продукта амплификации зависимого от повтора опосредованного рекомбинацией создания кДНК/амплификации ПЦР смеси TKR_2_click_S и TKR_2_click_L смеси в свободном растворе.

Подробное описание изобретения

Авторы настоящего изобретения разработали комплексы, содержащие по меньшей мере одну связь (например, химическую связь), для которой полимеразы

обладает пониженной способностью считывать или перемещаться. Возможность помечать библиотеки, не ограничиваясь способностью к считыванию полимеразы, значительно расширяет спектр способов мечения, которые могут быть использованы. Возможно, не ограничивающие преимущества включают в себя сниженную массу меток и/или представителей библиотеки; повышенную растворимость представителей библиотеки в различных условиях реакции (например, в водных и/или органических условиях, подходящих как для химического соединения, так и синтеза метки); повышенную стабильность (например, в водных и/или органических условиях реакции, когда такие олигонуклеотидные метки и связи могут быть более устойчивы к конкретным условиям реакции, по сравнению с меткой без таких связей); сниженную стоимость (например, сокращение использования дорогостоящих ферментативных реагентов); повышенную простоту использования (например, использование циановинилкарбазольного сшивания происходит в широком интервале pH 5,5-9,5, что позволяет сшивание в условиях реакции, подходящих для образования химического соединения-части комплекса); уменьшенное число стадий реакции и уменьшенное использование реагента(ов) (например, уменьшенное использование стадии обмена буфера, например, во время комбинаторных реакций разделения, в которых необходимо обработать независимо тысячи отдельных, небольших аликуот; стадии осаждения или стадии модификации pH); повышенную точность (например, когда используются опосредованные гибридизацией способы, чтобы уменьшить частоту возникновения событий неправильного мечения); и увеличенную совместимость с условиями реакции для образования химического соединения (например, позволяет использовать ортогональных функциональности, где могут быть использованы уникальные, помеченные химические вещества (например, использование УФ излучения), которые не будут происходить в процессе после стадии(й) разделения/смешивания, которые также снизят частоту событий неправильного мечения).

Поскольку такие связи могут характеризоваться уменьшенной способностью к считыванию с помощью полимеразы, авторы настоящего изобретения разработали способы, чтобы обеспечить секвенирование таких комплексов. Эти способы включают в себя использование релейных праймеров, чтобы охватывать связи и/или использование способных к высвобождению сшивающих олигонуклеотидов. Эти комплексы и способы могут быть использованы для создания различных библиотек выбираемых химических соединений путем создания кодированного отношения между

определенными метками и определенными химическими реакциями или структурными элементами. Чтобы идентифицировать одно или несколько химических соединений, олигонуклеотидные метки могут быть амплифицированы, клонированы, секвенированы и коррелированы с использованием установленного отношения. Способы создания и мечения библиотек этих комплексов подробно описаны ниже.

Комплексы

В настоящем изобретении предусмотрен комплекс, включающий в себя химическое соединение, одну или несколько меток, а также головной фрагмент, функционально связанный с химическим соединением и одной или несколькими метками. По меньшей мере одна из связей между головным фрагментом и меткой или между двумя метками представляет собой связь, для которых полимеразы обладает пониженной способностью считывать или перемещаться. Химические соединения, головные фрагменты, метки, связи и бифункциональные спейсеры дополнительно описаны ниже.

Химические соединения

Химические соединения или представители (например, небольшие молекулы или пептиды) согласно настоящему изобретению могут включать в себя один или несколько структурных элементов и, необязательно, включать в себя один или несколько каркасных элементов.

Каркасный элемент S может представлять собой одноатомный или молекулярный каркасный элемент. Иллюстративные одноатомные каркасные элементы включают в себя атом углерода, атом бора, атом азота или атом фосфора и т.д. Иллюстративные многоатомные каркасные элементы включают в себя циклоалкильную группу, циклоалкенильную группу, гетероциклоалкильную группу, гетероциклоалкенильную группу, арильную группу или гетероарильную группу. Конкретные варианты осуществления гетероарильного каркасного элемента включают в себя триазин, такой как 1,3,5-триазин, 1,2,3-триазин или 1,2,4-триазин; пиримидин; пиазин; пиридазин; фуран; пиррол; пирролин; пирролидин; оксазол; пиазол; изоксазол; пиран; пиридин; индол; индазол или пурин.

Каркасный элемент S может быть функционально связан с меткой с помощью любого применимого способа. В одном примере S представляет собой триазин, который непосредственно связан с головным фрагментом. Чтобы получить этот иллюстративный каркасный элемент, трихлортриазин (т.е., хлорированный предшественник триазина, содержащий три хлора) подвергают взаимодействию с

нуклеофильной группой головного фрагмента. Используя этот способ, S характеризуется тремя положениями, содержащими хлор, которые доступны для замены, причем два положения представляют собой узлы разнообразия и одно положение прикрепляется к головному фрагменту. Далее, структурный элемент A_n добавляют к узлу разнообразия каркасного элемента, и метку A_n , кодирующую структурный элемент A_n ("метка A_n "), лигируют с головным фрагментом, причем эти две стадии могут быть выполнены в любом порядке. Затем, структурный элемент B_n добавляют к оставшемуся узлу разнообразия, а метку B_n , кодирующую структурный элемент B_n , лигируют с концом метки A_n . В другом примере S представляет собой триазин, который функционально связан с меткой, причем трихлортриазин подвергают взаимодействию с нуклеофильной группой (например, аминогруппой) ПЭГа, алифатическим или ароматическим линкером метки. Структурные элементы и связанные с ними метки могут быть добавлены, как описано выше.

В еще одном примере S представляет собой триазин, который функционально связан со структурным элементом A_n . Чтобы получить этот каркасный элемент, структурный элемент A_n , содержащий два узла разнообразия (например, электрофильную группу и нуклеофильную группу, такую как Fmoc-аминокислоту), подвергают взаимодействию с нуклеофильной группой линкера (например, терминальной группой ПЭГ, алифатического или ароматического линкера, который прикреплен к головному фрагменту). Затем, трихлортриазин подвергают взаимодействию с нуклеофильной группой структурного элемента A_n . С использованием этого способа все три позиции хлора S используются в качестве узлов для структурных элементов. Как описано в настоящем документе, могут быть добавлены дополнительные структурные элементы и метки, а также могут быть добавлены дополнительные каркасные элементы S_n .

Иллюстративные структурные элементы A_n включают в себя, например, аминокислоты (например, альфа-, бета-, гамма-, дельта- и эpsilon-аминокислоты, а также производные природных и неприродных аминокислот), химические реакционноспособные реагенты (например, азидные или алкиновые цепи) с амином или тиольный реагент, или их комбинации. Выбор структурного элемента A_n зависит, например, от природы реакционноспособной группы, используемой в линкере, природы фрагмента каркасного элемента и растворителя, используемого для химического синтеза.

Иллюстративные структурные элементы V_n и C_n включают в себя любую применимую структурную единицу химического соединения, например, необязательно замещенные ароматические группы (например, необязательно замещенный фенил или бензил), необязательно замещенные гетероциклические группы (например, необязательно замещенный хинолинил, изохинолинил, индолил, изоиндолил, азаиндолил, бензимидазолил, азабензимидазолил, бензизоксазолил, пиридинил, пиперидинил или пирролидинил), необязательно замещенные алкильные группы (например, необязательно замещенные линейные или разветвленные C_{1-6} алкильные группы или необязательно замещенные C_{1-6} аминоалкильные группы) или необязательно замещенные карбоциклические группы (например, необязательно замещенный циклопропил, циклогексил или циклогексенил). Особенно применимые структурные элементы V_n и C_n включают в себя структурные элементы с одной или несколькими реакционноспособными группами, такими как необязательно замещенная группа (например, любая описанная в настоящем документе), содержащая один или несколько необязательных заместителей, которые представляют собой реакционноспособные группы, или могут быть химически модифицированы с образованием реакционноспособных групп. Иллюстративные реакционноспособные группы включают в себя одно или несколько из амина ($-NR_2$, где каждый R независимо представляет собой H или необязательно замещенный C_{1-6} алкил), гидроксильной, алкокси ($-OR$, где R представляет собой необязательно замещенный C_{1-6} алкил, такой как метокси), карбокси ($-COOH$), амид или химические реакционноспособные заместители. Сайт рестрикции может быть введен, например, в метку V_n или C_n , причем комплекс может быть идентифицирован посредством выполнения ПЦР и рестрикционного расщепления с одним из соответствующих ферментов рестрикции.

Головной фрагмент

В библиотеке головной фрагмент функционально связывает каждое химическое соединение с его кодирующей олигонуклеотидной меткой. Как правило, головной фрагмент представляет собой начальный олигонуклеотид, содержащий две функциональные группы, которые могут быть дополнительно дериватизированы, причем первая функциональная группа функционально связывает химическое соединение (или его компонент) с головным фрагментом и вторая функциональная группа функционально связывает одну или несколько меток с головным фрагментом. Бифункциональный спейсер необязательно может быть использован в качестве промежуточного фрагмента между головным фрагментом и химическим соединением.

Функциональные группы головного фрагмента могут быть использованы для образования ковалентной связи с компонентом химического соединения и другой ковалентной связи с меткой. Компонент может представлять собой любую часть небольшой молекулы, такую как каркасный элемент, содержащий узлы разнообразия, или структурный элемент. Альтернативно, головной фрагмент может быть дериватизирован, чтобы обеспечить спейсер (т.е., промежуточный фрагмент, отделяющий головной фрагмент от небольшой молекулы, которая будет образована в библиотеке), заканчивающийся в функциональной группе (например, гидроксильной, аминной, карбоксильной, сульфгидрильной, алкинильной, азидо или фосфатной группе), которая используется для образования ковалентной связи с компонентом химического соединения. Спейсер может быть прикреплен к 5'-концу в одном из внутренних положений или 3'-концу головного фрагмента. Когда спейсер прикреплен к одному из внутренних положений, спейсер может быть функционально связан с дериватизированным основанием (например, в положении C5 уридина) или помещен в пределах олигонуклеотида с использованием стандартных техник, известных в настоящей области техники. Иллюстративные спейсеры описаны в настоящем документе.

Головной фрагмент может характеризоваться любой применимой структурой. Головной фрагмент может составлять, например, от 1 до 100 нуклеотидов в длину, предпочтительно от 5 до 20 нуклеотидов в длину и, наиболее предпочтительно, от 5 до 15 нуклеотидов в длину. Головной фрагмент может быть одноцепочечным или двухцепочечным и может состоять из природных или модифицированных нуклеотидов, как описано в настоящем документе. Например, химический фрагмент может быть функционально связан с 3'-концом или 5'-концом головного фрагмента. Согласно конкретным вариантам осуществления головной фрагмент включает в себя структуру шпильки, образованную комплементарными основаниями в последовательности. Например, химический фрагмент может быть функционально связан с внутренним положением, 3'-концом или 5'-концом головного фрагмента.

Как правило, головной фрагмент включает в себя несамокомплементарную последовательность на 5'- или 3'-конце, что позволяет связывание олигонуклеотидной метки путем полимеризации, ферментативного лигирования или химической реакции. Головной фрагмент может позволить лигирование олигонуклеотидных меток и дополнительные стадии очистки и фосфорилирования. После добавления последней метки может быть добавлена дополнительная адаптерная последовательность к 5'-

концу последней метки. Иллюстративная адаптерная последовательность включает в себя связывающую праймер последовательность или последовательность, содержащую метку (например, биотин). В тех случаях, когда используется множество структурных элементов и соответствующих меток (например, 100), может быть использована стратегия смешивания и разделения во время стадии олигонуклеотидного синтеза, чтобы создать необходимое количество меток. Такие стратегии смешивания и разделения для синтеза ДНК известны в настоящей области техники. Полученные представители библиотеки могут быть амплифицированы с помощью ПЦР после отбора связывающих объектов против представляющей интерес мишени(ей).

Головной фрагмент или комплекс может дополнительно включать в себя один или несколько связывающих праймер последовательностей. Например, головной фрагмент содержит последовательность в области петли шпильки, которая служит в качестве связывающей праймер области для амплификации, причем связывающая праймер область характеризуется более высокой температурой плавления для его комплементарного праймера (например, который может включать фланкирующие участки идентификатора), чем для последовательности в головном фрагменте. Согласно другим вариантам осуществления комплекс включает в себя две связывающие праймер последовательности (например, для того, чтобы сделать возможной реакцию ПЦР) по обе стороны от одной или нескольких меток, которые кодируют один или несколько структурных элементов. Альтернативно, головной фрагмент может содержать одну связывающую праймер последовательность на 5'- или 3'-конце. Согласно другим вариантам осуществления головной фрагмент представляет собой шпильку, и область петли образует сайт связывания праймера, или сайт связывания праймера вводят путем гибридизации олигонуклеотида в головной фрагмент на 3'-стороне петли. Праймерный олигонуклеотид, содержащий область, гомологичную 3'-концу головного фрагмента и несущий связывающую праймер область на 5'-конце (например, для того, чтобы сделать возможной реакцию ПЦР), может быть гибридизован с головным фрагментом и может содержать метку, которая кодирует структурный элемент или добавление структурного элемента. Праймерный олигонуклеотид может содержать дополнительную информацию, такую как область рандомизированных нуклеотидов, например, от 2 до 16 нуклеотидов в длину, которая включена для анализа биоинформатики.

Головной фрагмент может дополнительно включать в себя структуру шпильки, причем эта структура может быть достигнута любым применимым способом.

Например, головной фрагмент может включать в себя комплементарные основания, которые образуют партнеров межмолекулярного спаривания оснований, например, путем Уотсон-Криковского спаривания оснований ДНК (например, аденин-тимин и гуанин-цитозин) и/или неоднозначного спаривания оснований (например, гуанин-урацил, инозин-урацил, инозин-аденин и инозин-цитозин). В другом примере головной фрагмент может включать в себя модифицированные или замещенные нуклеотиды, которые могут образовывать высокоаффинные дуплексные образования, по сравнению с немодифицированными нуклеотидами, такие модифицированные или замещенные нуклеотиды известны в настоящей области техники. В еще одном примере головной фрагмент содержит один или несколько сшитых оснований для образования структуры шпильки. Например, основания в пределах одной нити или основания в разных двойных нитях могут быть сшиты, например, с использованием псоралена.

Головной фрагмент или комплекс может дополнительно включать в себя одну или несколько меток, которые делают возможным обнаружение. Например, головной фрагмент, одна или несколько олигонуклеотидных метки и/или одна или несколько последовательностей праймеров могут включать в себя изотоп, средство радиовизуализации, маркер, меченое вещество, флуоресцентную метку (например, родамин или флуоресцеин), хемилюминесцентную метку, квантовую точку и репортерную молекулу (например, биотин или его his-метку).

Согласно другим вариантам осуществления головной фрагмент или метка могут быть модифицированы, чтобы поддерживать растворимость в эмульсиях, восстановленных или неводных (например, органических) условиях. Нуклеотидные основания головного фрагмента или метки могут быть сделаны более гидрофобными путем модификации, например, позиции С5 оснований Т или С с алифатическими цепями без значительного нарушения их способности к образованию водородной связи с их комплементарными основаниями. Иллюстративные модифицированные или замещенные нуклеотиды представляют собой 5'-диметокситритил-N4-диизобутиламинометилен-5-(1-пропинил)-2'-дезоксцитидин, 3'-[(2-цианоэтил)-(N,N-диизопропил)]фосфорамидит; 5'-диметокситритил-5-(1-пропинил)-2'-дезоксуридин, 3'-[(2-цианоэтил)-(N,N-диизопропил)]фосфорамидит; 5'-диметокситритил-5-фтор-2'-дезоксуридин, 3'-[(2-цианоэтил)-(N,N-диизопропил)]фосфорамидит и 5'-диметокситритил-5-(пирен-1-ил-этинил)-2'-дезоксуридин или 3'-[(2-цианоэтил)-(N, N-диизопропил)]фосфорамидит.

Кроме того, олигонуклеотид-головной фрагмент может содержать вкрапления модификаций, которые способствуют растворимости в органических растворителях. Например, азобензолфосфорамидит может вводить гидрофобный фрагмент в конструкцию головного фрагмента. Такие вставки гидрофобных амидитов в головной фрагмент могут происходить где угодно в молекуле. Тем не менее, вставка не может мешать последующему мечению, используя дополнительные метки ДНК во время синтеза библиотеки или последующей ПЦР, раз отбор завершен, или анализу микрочипов, если он используются для деконволюции меток. Такие описанные в настоящем документе дополнения к конструкции головного фрагмента будут вызывать растворимость головного фрагмента, например, в 15%, 25%, 30%, 50%, 75%, 90%, 95%, 98%, 99% или 100% органическом растворителе. Таким образом, добавление гидрофобных остатков в конструкцию головного фрагмента позволяет улучшить растворимость в эмульсиях или неводных (например, органических) условиях, при этом оставляя головной фрагмент компетентным для олигонуклеотидного мечения. Кроме того, метки ДНК, которые затем вводят в библиотеку, также могут быть модифицированы в положении С5 оснований Т или С, таким образом, что они также приводят библиотеку к более гидрофобному состоянию и растворимости в органических растворителях для последующих стадий синтеза библиотеки.

Согласно конкретным вариантам осуществления головной фрагмент и первая метка могут представлять собой одно и то же соединение, т.е., может быть построено множество соединений головной фрагмент-метка, которые все характеризуются наличием общих частей (например, связывающей праймер области) и все различаются другими частями (например, кодирующей областью). Они могут быть использованы в стадии "разделения" и объединены после того, как произошло событие, которое они кодируют.

Согласно конкретным вариантам осуществления головной фрагмент может кодировать информацию, например, путем включения последовательности, которая кодирует первую стадию разделения, или последовательности, которая кодирует идентификатор библиотеки, например, с использованием определенной последовательности, относящейся к конкретной библиотеке.

Олигонуклеотидные метки

Описанные в настоящем документе олигонуклеотидные метки (например, метка или часть головного фрагмента или часть хвостового фрагмента) могут быть использованы для кодирования любой применимой информации, такой как молекула,

часть химического соединения при добавлении компонента (например, каркасного элемента или структурного элемента), головного фрагмента в библиотеке, идентификации библиотеки, использования одного или нескольких представителей библиотеки (например, использования представителей в аликвоте библиотеки) и/или происхождения представителя библиотеки (например, путем использования последовательности происхождения).

Любая последовательность в олигонуклеотиде может быть использована для кодирования любой информации. Таким образом, одна последовательность олигонуклеотида может служить более чем одной цели, например, чтобы кодировать два или более типа информации или для обеспечения стартового олигонуклеотида, который также кодирует один или несколько типов информации. Например, первая метка может кодировать добавление первого структурного элемента, а также идентификацию библиотеки. В другом примере головной фрагмент может быть использован в качестве стартового олигонуклеотида, который функционально связывает химическое соединение с меткой, причем головной фрагмент дополнительно включает в себя последовательность, которая кодирует идентификацию библиотеки (т.е. идентифицирующую библиотеку последовательность). Соответственно, любая описанная в настоящем документе информация может быть закодированы в отдельных олигонуклеотидных метках или может быть объединена и закодирована в той же олигонуклеотидной последовательности (например, олигонуклеотидной метке, такой как метка, или головном фрагменте).

Последовательность структурного элемента кодирует идентификацию структурного элемента и/или тип реакции связывания, проведенной со структурным элементом. Эта последовательность структурного элемента включена в метку, причем метка может дополнительно включать в себя один или несколько типов описанных ниже последовательностей (например, идентифицирующую библиотеку последовательность, последовательность использования и/или последовательность происхождения).

Идентифицирующая библиотеку последовательность кодирует идентификацию конкретной библиотеки. Чтобы обеспечить смешивание двух или более библиотек, представитель библиотеки может содержать одну или несколько идентифицирующих библиотеку последовательностей, таких, как идентифицирующая библиотеку метка (т.е., олигонуклеотид, включающий в себя идентифицирующую библиотеку последовательность), в лигированной метке, в части последовательности головного

фрагмента или в последовательности хвостового фрагмента. Эти идентифицирующие библиотеку последовательности могут быть использованы для выведения отношений кодирования, причем последовательность метки транслируется и коррелирует с информацией химической (синтеза) истории. Соответственно, эти идентифицирующие библиотеку последовательности позволяют смешивание двух или более библиотек вместе для отбора, амплификации, очистки, секвенирования и т.д.

Последовательность использования кодирует историю (т.е., использование) одной или нескольких представителей библиотеки в отдельной аликвоте библиотеки. Например, отдельные аликвоты могут быть подвергнуты воздействию различных реакционных условий, структурных элементов и/или стадий отбора. В частности, эта последовательность может быть использована для идентификации таких аликвот и вывести свою историю (использование) и тем самым позволять смешивание вместе аликвот той же библиотеки с разными историями (использованиями) (например, различными экспериментами отбора) для целей смешивания вместе образцов для отбора, амплификации, очистки, секвенирования и т.д. Эти последовательности использования могут быть включены в головной фрагмент, хвостовой фрагмент, метку, метку использования (т.е. олигонуклеотид, включающий в себя последовательность использования) или любую другую описанную в настоящем документе метку (например, идентифицирующую библиотеку метку или метку происхождения).

Последовательность происхождения представляет собой вырожденную (случайную, сгенерированную стохастически) олигонуклеотидную последовательность любой применимой длины (например, приблизительно шесть олигонуклеотидов), которая кодирует происхождение представителя библиотеки. Эта последовательность служит для стохастического подразделения представителей библиотеки, которые в противном случае идентичны во всех отношениях в сущности, различимых по последовательности информации, например, эти наблюдения продуктов амплификации, полученных из уникальных шаблонов-предшественников (например, отдельных представителей библиотеки), можно отличить от наблюдений нескольких продуктов амплификации, полученных из того же шаблона-предшественника (например, выбранного элемента библиотеки). Например, после образования библиотеки и перед стадией отбора, каждый представитель библиотеки может включать в себя другую последовательность происхождения, такую как в метке происхождения. После отбора выбранные представители библиотеки могут быть амплифицированы для производства продуктов амплификации, и часть представителя

библиотеки, в которой ожидается нахождение последовательности происхождения (например, в метке происхождения), может наблюдаться и быть сравнена с последовательностью происхождения в каждом из других представителей библиотеки. Поскольку последовательности происхождения представляют собой вырожденные, каждый продукт амплификации каждого представителя библиотеки должен содержать другую последовательность происхождения. Тем не менее, наблюдение одной и той же последовательности происхождения в продукте амплификации может указывать на несколько ампликонов, полученных из одной и той же молекулы-шаблона. Когда желательно определение статистики и демографических характеристик популяции кодирующих меток до амплификации, в отличие от постаплификационного определения, может быть использована метка происхождения. Эти последовательности происхождения могут быть включены в головной фрагмент, хвостовой фрагмент, метку, метку происхождения (т.е., олигонуклеотид, включающий в себя последовательность происхождения) или любую другую описанную в настоящем документе метку (например, идентифицирующую библиотеку метку или метку использования).

Любой из описанных в настоящем документе типов последовательностей может быть включен в головной фрагмент. Например, головной фрагмент может включать в себя одну или несколько из последовательности структурного элемента, идентифицирующей библиотеку последовательности, последовательности использования или последовательности происхождения.

Любая из описанных в настоящем документе этих последовательностей может быть включена в хвостовой фрагмент. Например, хвостовой фрагмент может включать в себя одну или несколько из идентифицирующей библиотеку последовательности, последовательности использования или последовательности происхождения.

Любая из описанных в настоящем документе меток может включать в себя коннектор на или в непосредственной близости от 5'- или 3'-конца, характеризующийся фиксированной последовательностью. Коннекторы способствуют образованию связей (например, химических связей), обеспечивая реакционноспособную группу (например, химическую реакционноспособную группу или фотореакционноспособную группу) или путем предоставления сайта для средства, которое позволяет образование связи (например, средства интеркалирующего фрагмента или обратимой реакционноспособной группы в коннектор(ах) или сшивки олигонуклеотида). Каждый 5'-коннектор может быть одинаковым или различным и каждый 3'-коннектор может

быть одинаковым или различным. В иллюстративном, не ограничивающем комплексе, содержащем более одной метки, каждая метка может включать в себя 5'-коннектор и 3'-коннектор, причем каждый 5'-коннектор характеризуется той же последовательностью, и каждый 3'-коннектор характеризуется той же последовательностью (например, последовательность 5'-коннектора может быть такой же или отличаться от последовательности 3'-коннектора). Коннектор обеспечивает последовательность, которая может использоваться для одной или нескольких связей. Для обеспечения связывания релейного праймера или гибридизации сшивающего олигонуклеотида, коннектор может включать в себя одну или несколько функциональных групп, способствующих образованию связи (например, связи, для которой полимеразы обладает пониженной способностью считывать или перемещаться, например, химической связи).

Эти последовательности могут включать в себя любые описанные в настоящем документе модификации для олигонуклеотидов, такие как одна или несколько модификаций, которые способствуют растворимости в органических растворителях (например, любые описанные в настоящем документе, такие как для головного фрагмента), которые обеспечивают аналог природной фосфодиэфирной связи (например, фосфоротиоатный аналог) или которые обеспечивают один или несколько неприродных олигонуклеотидов (например, 2'-замещенные нуклеотиды, такие как 2'-О-метилированные нуклеотиды и 2'-фтор нуклеотиды или любые описанные в настоящем документе).

Эти последовательности могут включать в себя любые описанные в настоящем документе характеристики для олигонуклеотидов. Например, эти последовательности могут быть включены в метку, которая составляет менее чем 20 нуклеотидов (например, как описано в настоящем документе). В других примерах метки, включающие в себя одну или несколько из этих последовательностей, характеризуются приблизительно такой же массой (например, каждая метка характеризуется массой, которая составляет приблизительно +/- 10% от средней массы между представителями в пределах определенного набора меток, которые кодируют конкретную переменную область); не содержат связывающую праймер (например, константную) область; не содержат константную область или содержат константную область сокращенной длины (например, длина составляет менее чем 30 нуклеотидов, менее чем 25 нуклеотидов, менее чем 20 нуклеотидов, менее чем 19 нуклеотидов, менее чем 18 нуклеотидов, менее чем 17 нуклеотидов, менее чем 16 нуклеотидов, менее чем 15 нуклеотидов, менее чем

14 нуклеотидов, менее чем 13 нуклеотидов, менее чем 12 нуклеотидов, менее чем 11 нуклеотидов, менее чем 10 нуклеотидов, менее чем 9 нуклеотидов, менее чем 8 нуклеотидов или менее чем 7 нуклеотидов).

Стратегии секвенирования для библиотек и олигонуклеотидов этой длины могут необязательно включать в себя стратегии конкатенации или катенации по увеличению верности считывания или глубины секвенирования, соответственно. В частности, отбор кодированных библиотек, которые не содержат связывающие праймер области, был описан в литературе для SELEX, например, как описано в Jarosch et al., *Nucleic Acids Res.* 34: e86 (2006), которая включена в настоящий документ посредством ссылки. Например, представитель библиотеки может быть изменен (например, после стадии отбора), чтобы включать в себя первую адаптерную последовательность на 5'-конце комплекса, и вторую адаптерную последовательность на 3'-конце комплекса, причем первая последовательность, по существу, комплементарна второй последовательности, и приводит к образованию дуплекса. Для дополнительного повышения выхода два фиксированных оборванных нуклеотида (например, CC) добавляют к 5'-концу. Согласно конкретным вариантам осуществления первая адаптерная последовательность представляет собой 5'-GTGCTGC-3' (SEQ ID NO: 1), а вторая адаптерная последовательность представляет собой 5'-GCAGCACCC-3' (SEQ ID NO: 2).

Связи

Связи согласно настоящему изобретению находятся между олигонуклеотидами, которые кодируют информацию (например, такие как между головным фрагментом и меткой, между двумя метками или между меткой и хвостовым фрагментом), причем такие связи включают в себя любую связь, для которой полимеразы обладают пониженной способностью считывать или перемещаться. Иллюстративные связи включают в себя химические связи, включающие в себя одну или несколько из химической реакционноспособной группы, фотореакционноспособной группы, фрагмента интеркалирования, сшивающего олигонуклеотида или обратимой кореакционноспособной группы.

Связь может быть исследована, чтобы определить, обладает ли полимеразы пониженной способностью считывать или перемещаться через эту связь. Эта способность может быть исследована любым применимым способом, таким как жидкостная хроматография-масс-спектрометрия, анализ ОТ-ПЦР, демографические характеристики последовательности и/или анализ ПЦР.

Согласно конкретным вариантам осуществления, химическое лигирование включает в себя использование одной или нескольких химических реакционноспособных пар для обеспечения связи. Иллюстративные химические реакционноспособные пары представляют собой пару, включающую в себя необязательно замещенную алкинильную группу и необязательно замещенную азидогруппу для образования триазола через реакцию 1,3-диполярного циклоприсоединения Хьюсена; необязательно замещенный диен, содержащий систему 4π -электронов (например, необязательно замещенное 1,3-ненасыщенное соединение, такое как необязательно замещенный 1,3-бутадиен, 1-метокси-3-триметилсилилокси-1,3-бутадиен, циклопентадиен, циклогексадиен или фуран) и необязательно замещенный диенофил или необязательно замещенный гетеродиенофил, содержащий систему 2π -электронов (например, необязательно замещенная алкинильная группа или необязательно замещенная алкинильная группа) с образованием циклоалкенила посредством реакции Дильса-Альдера; нуклеофил (например, необязательно замещенный амин или необязательно замещенный тиол) с напряженным гетероциклическим электрофилом (например, необязательно замещенным эпоксидом, азиридином, ионом азиридиния или ионом эписульфония) с образованием гетероалкила через реакцию с раскрытием кольца; фосфортиоатную группу с группой йода, такой как в закрывающем лигировании олигонуклеотида, содержащем 5'-йод dT с 3'-фосфортиоатным олигонуклеотидом; необязательно замещенную аминогруппу с альдегидной группой или кетонной группой, такой как реакция 3'-альдегид-модифицированного олигонуклеотида, который может дополнительно быть получен окислением коммерчески доступного 3'-глицерил-модифицированного олигонуклеотида с 5'-амино олигонуклеотидом (т.е. в реакции восстановительного аминирования) или 5'-гидразида олигонуклеотидом; пару необязательно замещенной аминогруппы и группы карбоновой кислоты или тиольной группы (например, с использованием или без использования сукцинимидил-транс-4-(maleimidylmethyl)cyclohexane-1-carboxylate (SMCC) или 1-этил-3-(3-диметиламинопропил)карбодиимида (EDAC); пару необязательно замещенного гидразина и альдегида или кетонной группы; пару необязательно замещенного гидроксилamina и альдегидной или кетонной группы или пару нуклеофила и необязательно замещенного алкилгалогенида.

Платиновые комплексы, алкилирующие средства или модифицированные фураном нуклеотиды могут быть также использованы в качестве химической

реакционноспособной группы с образованием меж- или внутринитевых связей. Такие средства могут быть использованы между двумя олигонуклеотидами и могут необязательно присутствовать в сшивающем олигонуклеотиде.

Иллюстративные, не ограничивающие комплексы платины включают в себя цисплатин (цис-диаминдихлорплатина(II), например, для образования внутринитевой связи GG), трансплатин (транс-диаминдихлорплатина(II), например, для образования межнитевых связей GXG, где X может представлять собой любой нуклеотид), карбоплатин, пиколатин (ZD0473), ормаплатин или оксалиплатин, чтобы образовать, например, связи GC, CG, AG или GG. Любая из этих связей может представлять собой меж- или внутринитевую связь.

Иллюстративные, не ограничивающие алкилирующие средства включают в себя азотистый иприт (мехлорэтамин, например, с образованием GG связей), хлорамбуцил, мелфалан, циклофосфамид, пролекарственные формы циклофосфамида (например, 4-гидропероксициклофосфамид и ифосфамид), 1,3-бис-(2-хлорэтил)-1-нитрозомочевина (BCNU, кармустин), азиридин (например, митомицин С, триэтиленмеламин или триэтилтиофосфоридамид (тио-тепа) для образования связей GG или AG), гексаметилмеламин, алкилсульфонат (например, бусульфат для образования связей GG) или нитрозомочевина (например, 2-хлорэтилнитрозомочевина с образованием связей GG или CG, таких как кармустин (BCNU), хлорозотоцин, ломустин (CCNU) и семустин (метил-CCNU)). Любая из этих связей может представлять собой меж- или внутринитевую связь.

Модифицированные фураном нуклеотиды могут быть также использованы для образования связей. После окисления *in situ* (например, с N-бромсукцинимидом (NBS)), фрагмент фурана образует реакционноспособное оксо-еналь производное, которое реагирует с комплементарным основанием с образованием межнитевой связи. Согласно некоторым вариантам осуществления модифицированный фураном нуклеотид образует связи с комплементарным нуклеотидом А или С. Иллюстративные, не ограничивающие модифицированные фураном нуклеотиды включают в себя любой модифицированный 2'-(фуран-2-ил)пропаноиламином нуклеотид или ациклические, модифицированные нуклеотиды 2-(фуран-2-ил)этилгликолевой нуклеиновой кислоты.

Фотореакционноспособные группы также могут быть использованы в качестве реакционноспособной группы. Иллюстративные неограничивающие фотореакционноспособные группы включают в себя интеркалирующий фрагмент, производное псоралена (например, псорален, НМТ-псорален или 8-метоксипсорален),

необязательно замещенную циановинилкарбазольную группу, необязательно замещенную винилкарбазольную группу, необязательно замещенную циановинильную группу, необязательно замещенную акриламидную группу, необязательно замещенную диазириновую группу, необязательно замещенный бензофенон (например, сукцинимидиловый эфир 4-бензоилбензойной кислоты или бензофенонизотиоцианат), необязательно замещенную 5-(карбоксивинилуридиновую группу (например, 5-(карбоксивинил-2'-дезоксуридин) или необязательно замещенную азидную группу, (например, арилазид или галогенированный арилазид, такой как сукцинимидиловый эфир 4-азидо-2,3,5,6-тетрафторбензойной кислоты (ATFB)).

Интеркалирующие фрагменты могут быть также использованы в качестве реакционноспособной группы. Иллюстративные, не ограничивающие интеркалирующие фрагменты включают в себя производное псоралена, производное алкалоида (например, берберин, пальматин, коралин, сангвинарин (например, их иминиевые или алканоламиновые формы) или аристололактан- β -D-глюкозид), этидий катионов (например, этидий бромид), производное акридина (например, профлавин, акрифлавин или амсакрин), производное антрациклина (например, доксорубицин, эпирубицин, даунорубицин (дауномицин), идарубицин и акларубицина) или талидомид.

Для сшивающего олигонуклеотида может быть использована любая применимая реакционноспособная группа (например, описанная в настоящем документе) для образования меж- или внутринитевых связей. Иллюстративные реакционноспособные группы включают в себя химическую реакционноспособную группу, фотореакционноспособную группу, интеркалирующий фрагмент и обратимую кореакционноспособную группу. Сшивающие средства для использования с сшивающими олигонуклеотидами включают в себя без ограничения алкилирующие средства (например, описанные в настоящем документе), цисплатин (цис-диаминдихлорплатина(II)), транс-диаминдихлорплатина (II), псорален, НМТ-псорален, 8-метоксипсорален, модифицированные фураном нуклеотиды, 2-фтордезоксиинозин (2-F-dI), 5-бромдезоксцитозин (5-Br-dC), 5-бромдезоксуридин (5-Br-dU), 5-иоддезоксцитозин (5-I-dC), 5-иоддезоксуридин (5-I-dU), сукцинимидиловый транс-4-(малеимидилметил)циклогексан-1-карбоксилат, SMCC, EDAC или сукцинимидиловый ацетилтиоацетат (SATA).

Олигонуклеотиды также могут быть модифицированы, чтобы содержать тиоловые фрагменты, которые могут быть подвергнуты взаимодействию с различными

тиольными реакционноспособными группами, такими как малеимиды, галогены, йодацетамиды и, таким образом, могут быть использованы для сшивания двух олигонуклеотидов. Тиольные группы могут быть связаны с 5'- или 3'-концом олигонуклеотида.

Для межнитевого перекрестного сшивания между дуплексными олигонуклеотидами в положении пиримидина (например, тимидина) может быть выбран интеркалирующий, фотореакционноспособный фрагмент псорален. Псорален встраивается в дуплекс и образует ковалентные межнитевые поперечные сшивки с пиримидинами, преимущественно в сайтах 5'-ТрА, при облучении ультрафиолетовым светом (приблизительно 254 нм). Фрагмент псоралена может быть ковалентно присоединен к модифицированному олигонуклеотиду (например, с помощью алкановой цепи, такой как C_{1-10} алкил, или полиэтиленгликолевой группы, такой как $-(CH_2CH_2O)_nCH_2CH_2-$, где n равно целому числу от 1 до 50). Также могут быть использованы иллюстративные производные псоралена, причем не ограничивающие производные включают в себя 4'-(гидроксиэтоксиметил)-4,5',8'-триметилпсорален (НМТ-псорален) и 8-метоксипсорален.

Различные участки сшивающего олигонуклеотида могут быть изменены, чтобы ввести связь. Например, терминальные фосфоротиоаты в олигонуклеотидах также могут быть использованы для соединения двух соседних олигонуклеотидов. Галогенированные урацилы/цитозины могут также быть использованы в качестве сшивающих модификаций в олигонуклеотиде. Например, модифицированные 2-фтор-дезоксидеоксирибозой (2-F-dI) олигонуклеотиды могут быть подвергнуты взаимодействию с дисульфидсодержащими диаминами или тиопропиламинами с образованием дисульфидных связей.

Как описано ниже, обратимые кореакционноспособные группы включают в себя те, которые выбраны из циановинилкарбазольной группы, циановинильной группы, акриламидной группы, тиольной группы или сульфонилэтиловых тиоэфиров. Необязательно замещенная циановинилкарбазольная (CNV) группа может быть также использована в олигонуклеотидах для сшивания с пиримидиновым основанием (например, цитозином, тиминном и урацилом, а также их модифицированным основанием) в комплементарных нитях. CNV группы содействуют [2+2] циклодобавлению к соседним пиримидиновым основаниям при облучении при 366 нм, что приводит к межнитевой поперечной сшивке. Облучение при 312 нм изменяет поперечную сшивку и, таким образом, обеспечивает способ обратимого поперечного

сшивания олигонуклеотидных нитей. Неограничивающая группа CNV представляет собой 3-циановинилкарбазол, который может быть включен в качестве карбоксивинилкарбазольного нуклеотида (например, как 3-карбоксивинилкарбазол-1'-β-дезоксирибозид-5'-трифосфат).

Группа CNV может быть модифицирована, чтобы заменить реакционноспособную цианогруппу на другую реакционноспособную группу, чтобы обеспечить необязательно замещенную винилкарбазольную группу. Иллюстративные неограничивающие реакционноспособные группы для винилкарбазольной группы включают в себя амидную группу $-\text{CONR}_{N1}\text{R}_{N2}$, где каждый R_{N1} и R_{N2} может быть одинаковым или различным, и независимо представляет собой H и C_{1-6} алкил, например, $-\text{CONH}_2$; карбоксильную группу $-\text{CO}_2\text{H}$; или C_{2-7} алкоксикарбонильную группу (например, метоксикарбонил). Кроме того, реакционноспособная группа может быть расположена на альфа- или бета-углероде винильной группы. Иллюстративные винилкарбазольные группы включают в себя описанную в настоящем документе циановинилкарбазольную группу; амидовинилкарбазольную группу (например, амидовинилкарбазольный нуклеотид, такой как 3-амидовинилкарбазол-1'-β-дезоксирибозид-5'-трифосфат); карбоксивинилкарбазольную группу (например, карбоксивинилкарбазольный нуклеотид, такой как 3-карбоксивинилкарбазол-1'-β-дезоксирибозид-5'-трифосфат); и C_{2-7} алкоксикарбонилвинилкарбазольную группу (например, алкоксикарбонилвинилкарбазольный нуклеотид, такой как 3-метоксикарбонилвинилкарбазол-1'-β-дезоксирибозид-5'-трифосфат). Дополнительные необязательно замещенные винилкарбазольные группы и нуклеотиды, содержащие такие группы, представлены в химических формулах патента США №7972792 и в Yoshimura and Fujimoto, *Org. Lett.* 10:3227-3230 (2008), которые оба полностью включены в настоящий документ посредством ссылки.

Другие обратимые реакционноспособные группы включают в себя тиольную группу и другую тиольную группу для образования дисульфида, а также тиольную группу и винилсульфовую группу для образования сульфониэтиловых тиоэфиров. Группы тиол-тиольные могут дополнительно включать в себя связь, образованную в результате реакции с бис-((N-иодацетил)пиперазинил)сульфородаминол. Другие обратимые реакционноспособные группы (например, такие как некоторые фотореакционноспособные группы) включают в себя необязательно замещенные группы бензофенона. Не ограничивающий пример представляет собой бензофенонурацил (BPU), который может быть использован для избирательного по

сайту и последовательности образования межнитевой поперечной сшивки из ВРУ-содержащих олигонуклеотидных дуплексов. Эта поперечная сшивка может быть обратимой при нагревании, обеспечивая способ обратимого поперечного сшивания двух олигонуклеотидных нитей.

Согласно другим вариантам осуществления химическое лигирование включает в себя введение аналога фосфодиэфирной связи, например, для после отборного анализа ПЦР и секвенирования. Иллюстративные аналоги фосфодиэфирной связи включают в себя фосфоротиоатную связь (например, такую как введенная с использованием фосфоротиоатной группы и уходящей группы, такой как группа иода), фосфорамидную связь или фосфородитиоатную связь (например, такую как введенная с использованием фосфородитиоатной группы и уходящей группы, такой как группа иода).

Для любой из описанных в настоящем документе групп (например, химической реакционноспособной группы, фотореакционноспособной группы, интеркалирующего фрагмента, сшивающего олигонуклеотида или обратимой кореакционноспособной группы), группа может быть включена в или в непосредственной близости от конца олигонуклеотида или между 5'- и 3'-концами. Кроме того, одна или несколько групп могут присутствовать в каждом олигонуклеотиде. Когда необходимы пары реакционноспособных групп, тогда могут быть разработаны олигонуклеотиды для облегчения реакции между парой групп. В неограничивающем примере циановинилкарбазольной группы, которая кореагирует с пиримидиновым основанием, первый олигонуклеотид может быть разработан для включения циановинилкарбазольной группы на или в непосредственной близости от 5'-конца. В этом примере второй олигонуклеотид может быть разработан, чтобы быть комплементарным первому олигонуклеотиду и для включения кореакционноспособного пиримидинового основания в положение, которое выравнивается с циановинилкарбазольной группой, когда первый и второй олигонуклеотид гибридизуется. Любой из этих групп в настоящем документе и любой из олигонуклеотидов, содержащих одну или несколько групп, могут быть разработаны для облегчения реакции между группами с образованием одной или нескольких связей.

Бифункциональные спейсеры

Бифункциональный спейсер между головным фрагментом и химическим соединением может быть изменен, чтобы обеспечить соответствующий промежуточный фрагмент и/или для увеличения растворимости головного фрагмента в органическом растворителе. Коммерчески доступно широкое разнообразие спейсеров,

которые могут образовывать пару с головным фрагментом с библиотекой небольших молекул. Спейсер, как правило, состоит из линейных или разветвленных цепей и может включать в себя C_{1-10} алкил, гетероалкил, содержащий от 1 до 10 атомов, C_{2-10} алкенил, C_{2-10} алкинил, C_{5-10} арил, циклическую или полициклическую систему из 3-20 атомов, фосфодиэфира, пептида, олигосахарида, олигонуклеотида, олигомера, полимера или полиалкилгликоля (например, полиэтиленгликоля, такого как $-(CH_2CH_2O)_nCH_2CH_2-$, где n равно целому числу от 1 до 50) или их комбинации.

Бифункциональный спейсер может обеспечить соответствующий промежуточный фрагмент между головным фрагментом и химическим соединением библиотеки. Согласно некоторым вариантам осуществления бифункциональный спейсер состоит из трех частей. Часть 1 может представлять собой реакционноспособную группу, которая образует ковалентную связь с ДНК, такую как, например, карбоновая кислота, предпочтительно, активированная с помощью N-гидроксисукцинимидного (NHS) эфира для реагирования с аминогруппой на ДНК (например, амино-модифицированной dT), амидит для модификации 5' или 3'-конца одноцепочечного головного фрагмента (достигается с помощью стандартной химии олигонуклеотидов), химические реакционноспособные пары (например, азидо-алкиновое циклоприсоединение в присутствии катализатора Cu(I) или любого описанного в настоящем документе) или тиольные реакционноспособные группы. Часть 2 также может представлять собой реакционноспособную группу, которая образует ковалентную связь с химическим соединением, либо представлять собой структурный элемент A_n или каркасный элемент. Такая реакционноспособная группа может представлять собой, например, амин, тиол, азид или алкин. Часть 3 может представлять собой химически инертный промежуточный фрагмент переменной длины, введенный между частью 1 и 2. Такой промежуточный фрагмент может представлять собой цепь звеньев этиленгликоля (например, ПЭГ различной длины), алканов, алкенов, полиеновую цепь или пептидную цепь. Спейсер может содержать разветвления или вставки с гидрофобными фрагментами (такими как, например, бензольные кольца) для улучшения растворимости головного фрагмента в органических растворителях, а также люминесцентные фрагменты (например, флуоресцеин или Cy-3), используемые для целей обнаружения библиотеки. Гидрофобные остатки в конструкции головного фрагмента могут варьировать с конструкцией спейсера, чтобы облегчить синтез библиотеки в органических растворителях. Например, комбинация головного фрагмента и спейсера разработана,

чтобы содержать соответствующие остатки, причем коэффициент октанол:вода (P_{oct}) составляет, например, от 1,0 до 2,5.

Спейсеры могут быть эмпирически выбраны для данной разработки библиотеки небольших молекул, такой, чтобы библиотека могла быть синтезирована в органическом растворителе, например, в 15%, 25%, 30%, 50%, 75%, 90%, 95%, 98 %, 99% или 100% органическом растворителе. Спейсер может быть изменен с использованием модельных реакций до синтеза библиотеки для выбора соответствующей длины цепи, которая солюбилизирует головной фрагмент в органическом растворителе. Иллюстративные спейсеры включают в себя те, которые характеризуются увеличенной длиной алкильной цепи, увеличением единиц полиэтиленгликоля, разветвленных соединений с положительными зарядами (чтобы нейтрализовать отрицательные заряды фосфата на головном фрагменте) или увеличением количества гидрофобности (например, добавлением структур ароматических колец).

Примеры коммерчески доступных спейсеров включают в себя аминокислотные спейсеры, такие как пептиды (например, Z-Gly-Gly-Gly-Osu (N-альфа-бензилоксикарбонил-(глицин)₃-N-сукцинимидиловый эфир) или Z-Gly-Gly-Gly-Gly-Gly-Gly-Osu (N-альфа-бензилоксикарбонил-(глицин)₆-N-сукцинимидиловый эфир, SEQ ID NO: 3)), ПЭГ (например, Fmoc-aminoPEG2000-NHS или аминокислот-ПЭГ (12-24)-NHS) или цепи алкановой кислоты (например, Вос-ε-аминокапроновая кислота-Osu); химические реакционноспособные парные спейсеры, такие как описанные в настоящем документе химические реакционноспособные пары в комбинации с пептидным фрагментом (например, азидогомоаланин-Gly-Gly-Gly-OSu (SEQ ID NO: 4) или пропаргилглицин-Gly-Gly-Gly-OSu (SEQ ID NO: 5)), ПЭГ (например, азидо-ПЭГ-NHS) или фрагмент цепи алкановой кислоты (например, 5-азидопентановая кислота, (S)-2-(азидометил)-1-Вос-пирролидин, 4-азидоанилин или N-гидроксисукцинимидиловый эфир 4-азидобутан-1-овой кислоты); тиольные реакционноспособные спейсеры, такие как ПЭГ (например, SM(ПЭГ)_n NHS-ПЭГ-малеимид), алкановые цепи (например, 3-(пиридин-2-илдисульфанил)пропионовая кислота-Osu или сульфосукцинимидил-6-(3'-[2-пиридилдитио]пропионамидо)гексаноат)) и амидиты для синтеза олигонуклеотидов, такие как модификаторы аминокислот (например, 6-(трифторацетиламино)гексил-(2-цианоэтил)-(N,N-диизопропил)фосфораמידит), тиольные модификаторы (например, S-третил-6-меркаптогексил-1-[(2-цианоэтил)-(N,N-диизопропил)]фосфораמידит или химические реакционноспособные парные модификаторы (например, 6-гексин-1-ил-(2-

цианоэтил)-(N,N-диизопропил)фосфoramидит, 3-диметокситритилокси-2-(3-(3-пропаргиллоксипропанамидо)пропанамидо)пропил-1-О-сукциноил, длинноцепочечные алкиламино СРG или N-гидроксисукцинимидный эфир 4-азидобутан-1-овой кислоты)). Дополнительные спейсеры известны в настоящей области техники, и те, которые могут быть использованы при синтезе библиотеки, включают в себя без ограничения 5'-О-диметокситритил-1',2'-дидезоксирибоза-3'-[(2-цианоэтил)-(N,N-диизопропил)]фосфoramидит; 9-О-диметокситритилтриэтиленгликоль, 1-[(2-цианоэтил)-(N,N-диизопропил)]фосфoramидит; 3-(4,4-диметокситритилокси)пропил-1-[(2-цианоэтил)-(N,N-диизопропил)]фосфoramидит и 18-О-диметокситритилгексаэтиленгликоль, 1-[(2-цианоэтил)-(N,N-диизопропил)]фосфoramидит. Любой из спейсеров в настоящем документе может быть добавлен в тандеме друг с другом в различных комбинациях для создания спейсеров различных требуемых длин.

Спейсеры также могут быть разветвленными, причем разветвленные спейсеры, хорошо известны в настоящей области техники, и примеры могут включать симметричные или асимметричные дублиеры или симметричный утроитель. Смотрите, например, Newcome et al., *Dendritic Molecules: Concepts, Synthesis, Perspectives*, VCH Publishers (1996); Boussif et al., *Proc. Natl. Acad. Sci. USA* 92:7297-7301 (1995) и Jansen et al., *Science* 266:1226 (1994).

Техники ферментативного лигирования и химического лигирования

Могут быть использованы различные техники лигирования для добавления каркасных элементов, структурных элементов, спейсеров, связей, меток и/или головного фрагмента, чтобы произвести комплекс. Соответственно, любая из описанных в настоящем документе стадий построения может включать в себя любые применимые способы лигирования, такие как ферментативное лигирование и/или химическое лигирование. Эти стадии построения могут включать в себя добавление одной или нескольких меток к головному фрагменту или комплексу; добавление спейсера к головному фрагменту и добавление одного или нескольких каркасных элементов или структурных элементов к головному фрагменту или комплексу. Согласно конкретным вариантам осуществления используемые для любого олигонуклеотида техники лигирования обеспечивают полученный в результате продукт, который может быть транскрибирован и/или подвергнут обратной транскрипции, чтобы обеспечить декодирование библиотеки или для зависимой от шаблона полимеризации с одним или несколькими ДНК- или РНК-полимеразами.

Вообще, ферментативное лигирование производит олигонуклеотид, содержащий нативную фосфодиэфирную связь, которая может быть подвергнута транскрипции и/или обратной транскрипции. Иллюстративные способы ферментативного лигирования приведены в настоящем документе и включают в себя использование одной или нескольких РНК- или ДНК-лигаз, таких как РНК-лигазы Т4, ДНК-лигазы Т4, CircLigase™ оцДНК-лигазы, CircLigase™ II оцДНК-лигазы и ThermoPhage™ оцДНК-лигазы (Prokazyne Ltd., Reykjavik, Iceland).

Химическое лигирование также может быть использовано для получения олигонуклеотидов, способных к транскрипции или обратной транскрипции. Эффективность техники химического лигирования для обеспечения олигонуклеотидов, способных к транскрипции или обратной транскрипции, может потребовать исследования. Эта эффективность может быть исследована любым применимым способом, таким как жидкостная хроматография-масс-спектрометрия, анализ ОТ-ПЦР, и/или анализ ПЦР. Согласно конкретным вариантам осуществления химическое лигирование включает в себя использование одной или нескольких химических реакционноспособных пар, чтобы обеспечить промежуточный фрагмент, который может быть подвергнут транскрипции или обратной транскрипции. В частности, реакции, подходящие для химических реакционноспособных пар, представляют собой предпочтительных кандидатов для процесса лигирования (Kolb et al., *Angew. Chem. Int. Ed.*, 40:2004-2021 (2001); Van der Eycken et al., *QSAR Comb. Sci.*, 26:1115-1326 (2007)). Согласно одному варианту осуществления лигированные олигонуклеотиды содержат связь, для которой полимеразы характеризуются пониженной способностью считывать или перемещаться, например, "нечитаемую" связь.

Условия реакции для обеспечения ферментативного лигирования или химического лигирования

Описанные в настоящем документе способы могут включать одно или несколько условий реакции, которые способствуют ферментативному или химическому лигированию между головным фрагментом и меткой или между двумя метками. Эти условия реакции включают в себя использование описанных в настоящем документе модифицированных нуклеотидов внутри метки; использование донорных меток и акцепторных меток, характеризующихся различной длиной и различной концентрацией меток; использование различных видов лигазы, а также их комбинации (например, CircLigase™ ДНК-лигазы и/или РНК-лигазы Т4) и варьирования их концентрации; использование полиэтиленгликолей (ПЭГ), характеризующихся различными

молекулярными массами и варьирование их концентрации; использование не-ПЭГ средств краудинга (например, бетаина или бычьего сывороточного альбумина); варьирование температуры и продолжительности лигирования; варьирование концентрации различных средств, включающих в себя АТФ, $\text{Co}(\text{NH}_3)_6\text{Cl}_3$ и дрожжевой неорганический пирофосфат; использование ферментативно или химически фосфорилированных олигонуклеотидных меток; использование 3'-защищенных меток и использование преаденилированных меток. Эти условия реакции также включают в себя химические связи.

Головной фрагмент и/или метки могут включать в себя один или несколько модифицированных или замещенных нуклеотидов. Согласно предпочтительным вариантам осуществления головной фрагмент и/или метки включают в себя один или несколько модифицированных или замещенных нуклеотидов, которые способствуют ферментативному лигированию, такие как 2'-О-метилнуклеотиды (например, 2'-О-метилгуанин или 2'-О-метилурацил), 2'-фторнуклеотиды или любые другие модифицированные нуклеотиды, которые используются в качестве субстрата для лигирования. Кроме того, головной фрагмент и/или метки модифицируют, чтобы включать в себя один или несколько химических реакционноспособных групп для поддержания химического лигирования (например, необязательно замещенную алкильную группу и необязательно замещенную азидогруппу). Дополнительно, олигонуклеотиды-метки функциональны на обоих концах с химическими реакционноспособными группами, и, при необходимости, один из этих концов защищен таким образом, что группы могут быть направлены независимо и побочные реакции могут быть уменьшены (например, снижены побочные реакции полимеризации).

Ферментативное лигирование может включать в себя одну или несколько лигаз. Иллюстративные лигазы включают в себя CircLigaseTM оцДНК-лигазу (EPICENTRE Biotechnologies, Madison, WI), CircLigaseTM II оцДНК-лигазу (также из EPICENTRE Biotechnologies), ThermoPhageTM оцДНК-лигазу (Prokaryote Ltd., Reykjavik, Iceland), РНК-лигазу Т4 и ДНК-лигазу Т4. Согласно предпочтительным вариантам осуществления лигирование включает в себя использование РНК-лигазы или комбинации РНК-лигазы и ДНК-лигазы. Лигирование может дополнительно включать в себя один или несколько растворимых многовалентных катионов, таких как $\text{Co}(\text{NH}_3)_6\text{Cl}_3$ в сочетании с одной или несколькими лигазами.

Перед или после стадии лигирования комплекс может быть очищен по трем причинам. Во-первых, комплекс может быть очищен для удаления непрореагировавшего головного фрагмента или метки, которые могут приводить к перекрестным реакциям и вводить "шум" в процессе кодирования. Во-вторых, комплекс может быть очищен, чтобы удалить любые реагенты или непрореагировавший исходный материал, который может ингибировать или снижать активность лигирования лигазы. Например, фосфат может привести к пониженной активности лигирования. В-третьих, соединения, которые вводят на химической стадии или стадии лигирования, возможно, должны быть удалены, чтобы позволить последующую химическую стадию или стадию лигирования. Способы очистки комплекса описаны в настоящем документе.

Ферментативное и химическое лигирование может включать в себя полиэтиленгликоль, характеризующийся средней молекулярной массой более чем 300 дальтон (например, более чем 600 дальтон, 3000 дальтон, 4000 дальтон или 4500 дальтон). Согласно конкретным вариантам осуществления полиэтиленгликоль характеризуется средней молекулярной массой от приблизительно 3000 дальтон до 9000 дальтон (например, от 3000 дальтон до 8000 дальтон, от 3000 дальтон до 7000 дальтон, от 3000 дальтон до 6000 дальтон и от 3000 дальтон до 5000 дальтон). Согласно предпочтительным вариантам осуществления полиэтиленгликоль характеризуется средней молекулярной массой от приблизительно 3000 дальтон до приблизительно 6000 дальтон (например, от 3300 дальтон до 4500 дальтон, от 3300 дальтон до 5000 дальтон, от 3300 дальтон до 5500 дальтон, от 3300 дальтон до 6000 дальтон, от 3500 дальтон до 4500 дальтон, от 3500 дальтон до 5000 дальтон, от 3500 дальтон до 5500 дальтон и от 3500 дальтон до 6000 дальтон, такой как 4600 дальтон). Полиэтиленгликоль может присутствовать в любом применимом количестве, например, от приблизительно 25% (вес/объем) до приблизительно 35% (вес/объем), таком как 30% (вес/объем).

Согласно предпочтительному варианту осуществления настоящего изобретения метки устанавливают путем лигирования одноцепочечного олигонуклеотида с одноцепочечным олигонуклеотидом с использованием описанного ниже протокола лигирования: головной фрагмент: 25 мкМ (5'-конец: 5'-монофосфо/2'-ОМе G, промежуточные нуклеотиды: 2'-дезоксидезокси, и 3'-конец: 2'-блокированный/3'-блокированный); метка: 25 мкМ (5'-конец: 2'-ОМе/5'-ОН G, промежуточные нуклеотиды: 2'-дезоксидезокси и 3'-конец: 3'-ОН/2'-ОМе); $\text{Co}(\text{NH}_3)_6\text{Cl}_3$: 1 мМ; ПЭГ 4600: 30%

(вес/объем); РНК-лигаза T4 (Promega): 1,5 единиц/мкл; дрожжевая неорганическая пирофосфатаза: 0,0025 единиц/мкл; Трис: 50 mM; MgCl₂: 10 mM; АТФ: 1 mM; pH: 7,5 и вода: баланс.

Согласно дополнительным вариантам осуществления протокол включает в себя инкубацию при 37°C в течение 20 часов. Для целей актуального создания библиотеки могут быть использованы более высокие концентрации головного фрагмента, меток и/или лигазы, и такие модификации этих концентраций будут очевидны специалистам в настоящей области техники. Методология химического лигирования может включать в себя любой химический способ, который делает возможной функциональную связь двух кодирующих меток, не зависимо от того, способна ли полимераза считывать или перемещаться через функционально-ассоциированную связь.

Способы определения нуклеотидной последовательности комплекса

В настоящем изобретении предусмотрен способ определения нуклеотидной последовательности комплекса, таким образом, что могут быть установлены кодирующие отношения между последовательностью метки и структурными единицами (или структурными элементами) химического соединения. В частности, идентификация и/или история химического соединения может быть выведена из последовательности оснований в олигонуклеотиде. Используя этот способ, к библиотеке, включающей в себя разнообразные химические соединения или представителей (например, небольшие молекулы или пептиды), можно обращаться с помощью определенной последовательности метки.

На фигурах 1-4 предусмотрены различные иллюстративные способы определения нуклеотидной последовательности комплекса. На этих фигурах предусмотрены иллюстративные способы для охвата нечитаемой связи и/или высвобождения сшивающих олигонуклеотидов, которые приводят к образованию олигонуклеотидного шаблона.

На фигуре 1b предусмотрен иллюстративный способ использования релейных праймеров. В частности, релейные праймеры гибридизуют с фиксированными константными частями последовательности (например, 5'- и 3'-коннекторами), причем каждая последовательность для релейных праймеров может быть одинаковой или разной. Релейные праймеры охватывают нечитаемую связь, и расширение продолжается вдоль читаемых частей комплекса (например, кодирующей последовательности, например, одной или нескольких меток) для получения олигонуклеотидных фрагментов. Затем релейные праймеры и фрагменты могут быть

лигированы с образованием шаблона, который может быть дополнительно амплифицирован и секвенирован.

На фигурах 2-4 представлены иллюстративные нечитаемые связи. На фигуре 2 представлена связь, образованная псораленом, который реагирует в пределах дуплекса, образованного смежными 5'- и 3'-коннекторами. На фигуре 3 представлена связь, образованная сшитым олигонуклеотидом, который гибридизовал с соседними 5'- и 3'-коннекторами.

Любая из описанных в настоящем документе связей может быть обратимой или необратимой. Обратимые связи включают в себя фотореакционноспособные связи (например, циановинилкарбазольная группа и тимидин, как показано на фиг. 4) и окислительно-восстановительные связи. Дополнительные связи описаны в настоящем документе.

Согласно альтернативному варианту осуществления "нечитаемая" связь может быть ферментативно восстановлена, чтобы образовать читаемую или по меньшей мере способную к перемещению связь. Ферментативные процессы репарации хорошо известны специалистам в настоящей области техники и включают в себя без ограничения механизмы репарации димера пиримидина (например, тимидина) (например, с использованием фотолиазы или гликозилазы (например, T4 гликозилаза пиримидинового димера (PDG))), механизмы эксцизионной репарации оснований (например, с использованием гликозилазы, апуриновой/апиримидиновой (AP) эндонуклеазы, флэп-эндонуклеазы или поли(АДФ-рибоза)полимеразы (например, человеческой апуриновой/апиримидиновой (AP) эндонуклеазы, APE 1; эндонуклеазы III (Nth); эндонуклеазы IV; эндонуклеазы V; формамидопиримидин [fapy]-ДНК гликозилазы (Fpg); человеческой 8-оксогуанингликозилазы 1 (α изоформы) (hOGG1); человеческой подобной VIII эндонуклеазы 1 (hNEIL1); урацил-ДНК-гликозилазы (UDG); человеческой одонитевой селективной монофункциональной урацил-ДНК-гликозилазы (SMUG1) и человеческой алкиладенин-ДНК-гликозилазы (Haag)), которые могут быть необязательно в сочетании с одной или несколькими эндонуклеазами, ДНК или РНК-полимеразами и/или лигазами для репарации), механизмы репарации метилирования (например, с использованием метилгуанинметилтрансферазы), механизмы репарации AP (например, с использованием апуриновой/апиримидиновой (AP) эндонуклеазы (например, APE 1; эндонуклеазы III; эндонуклеазы IV; эндонуклеазы V; Fpg; hOGG1 и hNEIL1), которые могут быть необязательно в сочетании с одной или несколькими эндонуклеазами, ДНК или РНК-полимеразами

и/или лигазами для репарации), механизмы эксцизионной репарации нуклеотидов (например, с использованием перекрестно-комплементарных белков эксцизионной репарации или эксцизионных нуклеаз, которые могут быть необязательно в сочетании с одной или несколькими эндонуклеазами, ДНК или РНК-полимеразами и/или лигазами для репарации), а также механизмы репарации ошибочно спаренных оснований (например, с использованием эндонуклеазы (например, T7 эндонуклеазы I; MutS, MutH и/или MutL), которые могут быть необязательно в сочетании с одной или несколькими эндонуклеазами, экзонуклеазами, геликазами, ДНК или РНК-полимеразами и/или лигазами для репарации). Доступны коммерческие смеси ферментов для обеспечения этих видов механизмов репарации, например, репарационная смесь PreCR® (New England Biolabs Inc., Ipswich MA), которая включает в себя Taq ДНК-лигазу, эндонуклеазу IV, BST ДНК-полимеразу, Fpg, урацил-ДНК-гликозилазу (УДГ), T4 PDG (T4 эндонуклеазу V) и эндонуклеазу VIII.

Способы мечения кодированных библиотек

В настоящем изобретении предусмотрен способ функционального связывания олигонуклеотидных меток с химическими соединениями, такими, что отношения кодирования могут быть установлены между последовательностью метки и структурными единицами (или структурными элементами) химического соединения. В частности, идентификация и/или история химического соединения может быть выведена из последовательности оснований в олигонуклеотиде. Используя этот способ, библиотека, включающая в себя разнообразные химические соединения или представителей (например, небольшие молекулы или пептиды), может быть закодирована с помощью определенной последовательности метки.

Как правило, эти способы включают использование головного фрагмента, который содержит по меньшей мере одну функциональную группу, которая может быть разработана химически, и по меньшей мере одну функциональную группу, с которой одноцепочечный олигонуклеотид может быть связан (или лигирован). Связывание может быть осуществлено с помощью любых подходящих средств, например, путем ферментативного связывания (например, лигирования с одной или несколькими из РНК-лигазы и/или ДНК-лигазы) или путем химического связывания (например, с помощью реакции замещения между двумя функциональными группами, такими как нуклеофил и уходящая группа).

Для создания многочисленных химических соединений в библиотеке, раствор, содержащий головной фрагмент, может быть разделен на несколько аликвот, а затем

помещен во множество физически отделенных компартментов, такие как лунки многолуночного планшета. Как правило, это стадия "разделения". В каждом компартменте или лунке выполняются стадии последовательной химической реакции и лигирования с одноцепочечной меткой в каждой аликвоте. Отношение между условиями химической реакции и последовательностью одноцепочечной метки записывается. Реакционные стадии и стадии лигирования могут быть выполнены в любом порядке. Затем прореагировавшие и лигировавшие аликвоты сочетают или "объединяют", и может быть выполнена дополнительная очистка в этой точке. Эти стадии разделения и объединения могут необязательно повторяться.

Далее, библиотека может быть исследована и/или отобрана на конкретную описанную в настоящем документе характеристику или функцию. Например, смесь меченых химических соединений может быть разделена по меньшей мере на две популяции, причем первая популяция связывается с конкретной биологической мишенью, а вторая популяция - нет (например, путем негативной селекции или позитивной селекции). Первая популяция затем может быть избирательно захвачена (например, путем элюирования на колонке, обеспечивая представляющую интерес мишень, или путем инкубации аликвоты с представляющей интерес мишенью) и, необязательно, дополнительно проанализирована или исследована, например, с дополнительными стадиями промывания, очистки, негативной селекции, позитивной селекции или разделения.

Наконец, химические истории одного или нескольких представителей (или химических соединений) в пределах выбранной популяции могут быть определены с помощью последовательности функционально связанного олигонуклеотида. После коррелирования последовательности с конкретным структурным элементом этот способ может идентифицировать отдельных представителей библиотеки с выбранной характеристикой (например, повышенной склонностью к связыванию с белком-мишенью и, таким образом, к вызыванию терапевтического эффекта). Для дальнейшего исследования и оптимизации, кандидатные терапевтические соединения могут быть затем получены путем синтеза идентифицированных представителей библиотеки с или без связанных с ними олигонуклеотидных меток.

Описанные в настоящем документе способы могут включать любое количество дополнительных стадий для диверсификации библиотеки или для запроса представителей библиотеки. Для любого описанного в настоящем документе способа мечения последовательное количество "n" меток может быть добавлено с

дополнительным количеством "n" стадий лигирования, разделения и/или фосфорилирования. Иллюстративные дополнительные стадии включают в себя рестрикцию ассоциированных с представителями библиотеки кодирующих олигонуклеотидов с использованием одной или нескольких рестрикционных эндонуклеаз; репарацию ассоциированных кодирующих олигонуклеотидов, например, с любым репарационным ферментом, таким как те, которые описаны в настоящем документе; лигирование одной или нескольких адапторных последовательностей с одним или обоими концами ассоциированных с представителями библиотеки кодирующих олигонуклеотидов, например, такими, как одна или несколько адапторных последовательностей для обеспечения примированной последовательности для амплификации и секвенирования, или обеспечения метки, такой как биотин, для иммобилизации последовательности; обратную транскрипцию или транскрипцию, необязательно с последующей обратной транскрипцией собранных меток в комплексе с использованием обратной транскриптазы, транскриптазы или другой зависимой от шаблона полимеразы; амплификацию собранных меток в комплексе с использованием, например, ПЦР; создание клональных изолятов одной или нескольких популяций собранных меток в комплексе, например, путем использования бактериальной трансформации, образования эмульсии, разбавления, техник поверхностного захвата и т.д.; амплификацию клональных изолятов одной или нескольких популяций собранной метки в комплексе, например, путем использования клональных изолятов в качестве шаблонов для зависимой от шаблона полимеризации нуклеотидов; и определение последовательности клональных изолятов одной или нескольких популяций, собранных меток в комплексе, например, путем использования клональных изолятов в качестве шаблонов для зависимой от шаблона полимеризации с флуоресцентно мечеными нуклеотидами. Дополнительные способы амплификации и секвенирования олигонуклеотидных меток описаны в настоящем документе.

Эти способы могут быть использованы для идентификации и открытия любого количества химических соединений с конкретной характеристикой или функцией, например, на стадии отбора. Желаемая характеристика или функция может быть использована в качестве основы для разделения библиотеки по меньшей мере на две части с одновременным обогащением по меньшей мере одного из представителей или родственных представителей в библиотеке с желаемой функцией. Согласно конкретным вариантам осуществления способ включает идентификацию подобного лекарственному средству представителя библиотеки, который связывается или

инактивирует представляющий терапевтический интерес белок. Согласно другому варианту осуществления разрабатывают последовательность химических реакций, и набор структурных элементов выбирают таким образом, что реакция выбранных структурных элементов при определенных химических условиях будет создавать комбинаторное множество молекул (или библиотеку молекул), в котором одна или несколько молекул могут быть полезны в качестве терапевтического средства для конкретного белка. Например, химические реакции и структурные элементы выбираются так, чтобы создать библиотеку, содержащую структурные группы, обычно присутствующие в киназных ингибиторах. В любом из этих случаев олигонуклеотидные метки кодируют химическую историю представителя библиотеки, и в каждом случае набор химических возможностей может быть представлен любой конкретной комбинацией меток.

Согласно одному варианту осуществления библиотеку химических соединений или ее часть приводят в контакт с биологической мишенью в условиях, подходящих по меньшей мере для одного представителя библиотеки для связывания с мишенью, с последующим удалением представителей библиотеки, которые не связываются с мишенью, и анализом одной или нескольких олигонуклеотидных меток, связанных с мишенью. Этот способ может дополнительно включать амплификации меток с помощью известных в настоящей области техники способов. Иллюстративные биологические мишени включают в себя ферменты (например, киназы, фосфатазы, метилазы, деметилазы, протеазы и ферменты репарации ДНК), белки, участвующие в белок:белковых взаимодействиях (например, лиганды для рецепторов), рецепторные мишени (например, GPCR и RTK), ионные каналы, бактерии, вирусы, паразиты, ДНК, РНК, прионы и углеводы.

Согласно другому варианту осуществления химические соединения, которые связываются с мишенью, не подвергают амплификации, но анализируют непосредственно. Иллюстративные способы анализа включают микроматричный анализ, включающий в себя анализ исчезающего резонансного фотонного кристалла; основанные на гранулах способы для деконволиции меток (например, с использованием his-меток); анализ немеченых фотонных кристаллических биосенсоров (например, BIND® Reader из SRU Biosystems, Inc., Woburn, MA) или основанные на гибридизации подходы (например, с использованием матриц иммобилизованных олигонуклеотидов, комплементарных последовательностям, присутствующих в библиотеке меток).

Кроме того, химические реакционноспособные пары (или функциональные группы) могут быть легко включены в схемы твердофазного синтеза олигонуклеотидов и будет поддерживать эффективное химическое лигирование олигонуклеотидов. Кроме того, полученные лигированные олигонуклеотиды могут действовать в качестве шаблонов для зависимой от шаблонов полимеризации с одним или несколькими полимеразми. Соответственно, любая из описанных в настоящем документе стадий связывания для мечения закодированных библиотек может быть модифицирована, чтобы включать в себя одну или несколько техник ферментативного лигирования и/или химического лигирования. Иллюстративные техники лигирования включают в себя ферментативное лигирование, такое как использование одной или нескольких РНК лигаз и/или ДНК-лигаз; и химическое лигирование, такое как использование химических реакционноспособных пар (например, пары, включающей в себя необязательно замещенные алкильные и азидные функциональные группы).

Кроме того, одна или несколько библиотек могут быть объединены на стадии разделения и смешивания. Чтобы обеспечить смешивание двух или более библиотек, представитель библиотеки может содержать одну или несколько идентифицирующих библиотеку последовательностей, таких как в идентифицирующей библиотеку метке, в лигированной метке или как часть последовательности головного фрагмента, как описано в настоящем документе.

Способы снижения массы

Большая часть мотивации для одноцепочечных кодирующих стратегий возникает из-за сниженной массы одноцепочечной метки, по сравнению с двухцепочечными метками. Сниженная масса потенциально дает несколько преимуществ, включающих в себя повышенную растворимость, сниженную стоимость, увеличенную реактивность, повышенную доступность меток, сниженный гидродинамический радиус, повышенную точность аналитических оценок и т.д. В дополнение к использованию методологии одноцепочечного мечения, дальнейшее снижение массы может быть достигнуто, в том числе с использованием одного или нескольких из следующего: одной или нескольких меток, характеризующихся укороченной длиной, наборов меток с постоянной массой, кодирующего головного фрагмента, одного или нескольких представителей библиотеки с отсутствующей связывающей праймер областью и/или константной областью, одного или нескольких представителей библиотеки, содержащих сокращенную константную область или любые другие описанные в настоящем документе методологии.

Чтобы свести к минимуму массу представителей в библиотеке, может быть уменьшена длина одной или нескольких меток, например, до такой длины, которая настолько короткая, насколько возможно для кодирования каждого размера разделения. В частности, метки могут составлять менее чем 20 нуклеотидов (например, менее чем 19 нуклеотидов, менее чем 18 нуклеотидов, менее чем 17 нуклеотидов, менее чем 16 нуклеотидов, менее чем 15 нуклеотидов, менее чем 14 нуклеотидов, менее чем 13 нуклеотидов, менее чем 12 нуклеотидов, менее чем 11 нуклеотидов, менее чем 10 нуклеотидов, менее чем 9 нуклеотидов, менее чем 8 нуклеотидов или менее чем 7 нуклеотидов). Как описано ниже в примерах, могут быть использованы более короткие метки (например, приблизительно 10 нуклеотидов или меньше) для лигирования меток.

Также могут быть использованы стратегии постоянной массы, которые могут помочь в анализе во время синтеза библиотеки. Кроме того, наборы меток с постоянной массой могут позволить распознавание всех случаев отдельных ошибок (например, ошибок, связанных с неправильным считыванием последовательности или химическим или ферментативным лигированием метки) и большинство случаев множественных ошибок. Связь между длиной набора одноцепочечных меток с постоянной массой и кодирующей способностью (например, минимальные длины для поддержания конкретных размеров разделения структурных элементов или идентичности библиотеки и т.д.) приводится ниже в таблице 1. Таким образом, могут быть использованы наборы меток с постоянной массой, чтобы обеспечить преимущественную кодирующую способность при сохранении распознавания ошибок при образовании библиотеки.

Таблица 1.

Длина	Основание 1	Основание 2	Основание 3	Основание 4	Комбинации
1	1	0	0	0	1
2	1	1	0	0	2
3	1	1	1	0	6
4	1	1	1	1	24
5	2	1	1	1	60
6	2	2	1	1	180
7	2	2	2	1	630
8	2	2	2	2	2520
9	3	2	2	2	7560
10	3	3	2	2	25200
11	3	3	3	2	92400
12	3	3	3	3	369600
13	4	3	3	3	1201200
14	4	4	3	3	4204200

15	4	4	4	3	15765750
16	4	4	4	4	63063000
17	5	4	4	4	214414200
18	5	5	4	4	771891120
19	5	5	5	4	2933186256
20	5	5	5	5	11732745024

Чтобы свести к минимуму массу в библиотеке, головной фрагмент может быть использован не только для связи химического фрагмента и метки, но также для кодирования идентификации конкретной библиотеки или для конкретной стадии. Например, головной фрагмент может кодировать информацию, например, множество головных фрагментов, которые кодируют первое разделение(я) или идентификацию библиотеки, например, с использованием определенной последовательности, связанной с конкретной библиотекой.

Кроме того, связывающие праймер (например, константные) области из библиотеки ДНК-кодированных химических соединений могут быть исключены во время стадии(й) отбора. Затем эти области могут быть добавлены после отбора, например, путем одноцепочечного лигирования. Одна иллюстративная стратегия будет включать в себя предоставление химического соединения на 5'-конце кодирующего олигонуклеотида, выбор конкретного химического соединения на основе любой полезной конкретной характеристики или функции, и лигирование хвостового олигонуклеотида с 3'-концом кодирующего олигонуклеотида, который включает в себя связывающую праймер последовательность и необязательно может содержать одну или несколько меток, например, описанные в настоящем документе метку "использования", метку "происхождения" и т.д. Эта связывающая праймер последовательность затем может быть использована для инициирования зависимой от шаблона полимеризации для создания кДНК (или кРНК), комплементарной выбранному представителю библиотеки. кДНК или кРНК затем лигируют 3'-концом с олигонуклеотидом, который содержит связывающую праймер последовательность и теперь, когда кодирующая информация расположена по бокам с обеих сторон связывающих праймер последовательностей, олигонуклеотид может быть секвенирован и/или амплифицирован с использованием установленных подходов, таких, как любой из описанных в настоящем документе.

Масса может дополнительно быть минимизирована путем исключения или уменьшения размера одной или нескольких константных последовательностей,

которые отделяют метки кодирования. Одноцепочечное лигирование не требует комплементарной связи между подвергаемыми лигированию концами или между этими концами и закрывающей последовательностью. Таким образом, никакая фиксированная последовательность не требуется для поддержки ферментативного лигирования. Короткие фиксированные области между метками могут быть применимы для информатического разбора меток или других процессов деконволюции *in silico*.

Способы кодирования химических соединений в библиотеке

Способы согласно настоящему изобретению могут быть использованы для синтеза библиотеки, содержащей широкий ряд химических соединений, которые кодируются с помощью олигонуклеотидных меток. Примеры структурных элементов и кодирующих ДНК-меток можно найти в заявке на патент США № 2007/0224607, включенной в настоящий документ посредством ссылки.

Каждое химическое соединение образуется из одной или нескольких структурных элементов и, необязательно, каркасного элемента. Каркасный элемент служит для обеспечения одного или нескольких узлов разнообразия в конкретной геометрии (например, триазина, чтобы обеспечить три узла, пространственно расположенных вокруг гетероарильного кольца, или линейной геометрии).

Структурные элементы и их кодирующие метки могут быть добавлены непосредственно или опосредованно (например, через спейсер) к головному фрагменту с образованием комплекса. Когда головной фрагмент включает в себя спейсер, структурный элемент или каркасный элемент добавляют в конце спейсера. Когда спейсер отсутствует, структурный элемент может быть добавлен непосредственно к головному фрагменту, или сам структурный элемент может включать в себя спейсер, который реагирует с функциональной группой головного фрагмента. Иллюстративные спейсеры и головные фрагменты описаны в настоящем документе.

Каркасный элемент может быть добавлен любым применимым способом. Например, каркасный элемент может быть добавлен к концу спейсера или головному фрагменту, и последовательные структурные элементы могут быть добавлены к доступным узлам разнообразия каркасного элемента. В другом примере структурный элемент A_n сначала добавляют к спейсеру или головному фрагменту, а затем узел разнообразия каркасного элемента S взаимодействует с функциональной группой в структурном элементе A_n . Олигонуклеотидные метки, кодирующие конкретный каркасный элемент, необязательно могут быть добавлены к головному фрагменту или комплексу. Например, S_n добавляют к комплексу в n реакционных сосудов, где n

представляет собой целое число больше, чем один, и метка S_{n-} (т.е., метка $S_1, S_2, \dots, S_{n-1}, S_n$) связана с функциональной группой комплекса.

Структурные элементы могут быть добавлены на множественных синтетических стадиях. Например, аликвоту головного фрагмента, необязательно содержащего присоединенный спейсер, разделяют на n реакционных сосудов, где n представляет собой целое число два или более. На первой стадии структурный элемент A_n добавляют в каждый реакционный сосуд n (например, структурный элемент $A_1, A_2, \dots, A_{n-1}, A_n$ добавляют в реакционный сосуд $1, 2, \dots, n-1, n$), где n представляет собой целое число, а каждый структурный элемент A_n уникален. На второй стадии каркасный элемент S добавляют в каждый реакционный сосуд для образования комплекса A_n-S . Необязательно, каркасный элемент S_n может быть добавлен в каждый реакционный сосуд, чтобы образовать комплекс A_n-S_n , где n представляет собой целое число больше двух, и каждый каркасный элемент S_n может быть уникальным. На третьей стадии структурный элемент B_n добавляют в каждый реакционный сосуд n , содержащий комплекс A_n-S (т.е. структурный элемент $B_1, B_2, \dots, B_{n-1}, B_n$ добавляют в реакционный сосуд $1, 2, \dots, n-1, n$, содержащий комплекс $A_1-S, A_2-S, \dots, A_{n-1}-S, A_n-S$), где каждый структурный элемент B_n уникален. На дополнительных стадиях структурный элемент C_n может быть добавлен в каждый реакционный сосуд n , содержащий комплекс B_n-A_n-S (т.е., структурный элемент $C_1, C_2, \dots, C_{n-1}, C_n$ добавляют в реакционный сосуд $1, 2, \dots, n-1, n$, содержащий комплекс $B_1-A_1-S \dots B_n-A_n-S$), где каждый структурный элемент C_n уникален. Полученная библиотека будет содержать n^3 число комплексов, содержащих n^3 меток. Таким образом, дополнительные стадии синтеза могут быть использованы для связывания дополнительных структурных элементов для дальнейшей диверсификации библиотеки.

После образования библиотеки полученные в результате комплексы могут быть необязательно очищены и подвергнуты полимеризации или реакции лигирования, например, к хвостовому фрагменту. Это общая стратегия может быть расширена для включения дополнительных узлов разнообразия и структурных элементов (например, D, E, F и т.д.). Например, первый узел разнообразия подвергают взаимодействию со структурными элементами и/или S и кодируют с помощью олигонуклеотидной метки. Затем дополнительные структурные элементы подвергают взаимодействию с полученным комплексом и последующий узел разнообразия дериватизируют с помощью дополнительных структурных элементов, который кодируется праймером, используемым для реакции полимеризации или лигирования.

Для образования кодированной библиотеки олигонуклеотидные метки добавляют к комплексу до или после каждой стадии синтеза. Например, до или после добавления структурного элемента A_n в каждый реакционный сосуд, метка A_n связывается с функциональной группой головного фрагмента (т.е., метку $A_1, A_2, \dots A_{n-1}, A_n$, добавляют в реакционный сосуд 1, 2, ... n-1, n, содержащий головной фрагмент). Каждая метка A_n характеризуется определенной последовательностью, которая коррелирует с каждым уникальным структурным элементом A_n , и определение последовательности метки A_n обеспечивает химическую структуру структурного элемента A_n . Таким образом, дополнительные метки используются для кодирования дополнительных структурных элементов или дополнительных каркасных элементов.

Кроме того, последняя добавленная к комплексу метка может либо включать в себя связывающую праймер последовательность, либо обеспечивать функциональную группу, позволяющую связывание (например, путем лигирования) связывающей праймер последовательности. Связывающая праймер последовательность может быть использована для амплификации и/или секвенирования олигонуклеотидных меток комплекса. Иллюстративные способы амплификации и секвенирования включают полимеразную цепную реакцию (ПЦР), линейную амплификацию (LCR), амплификацию по типу катящегося кольца (RCA) или любой другой способ, известный в настоящей области техники, для амплификации или определения последовательности нуклеиновых кислот.

С использованием этих способов могут быть образованы большие библиотеки, содержащие большое количество закодированных химических соединений. Например, головной фрагмент подвергают взаимодействию со спейсером и структурным элементом A_n , который включает в себя 1000 различных вариантов (т.е., $n = 1000$). Для каждого структурного элемента A_n лигируют ДНК-метку A_n или праймер расширяют на головной фрагмент. Эти реакции могут быть выполнены в 1000-луночной планшете или 10 x 100-луночных планшетах. Все реакции могут быть объединены, необязательно, очищены и разделены на второй набор планшетов. Далее, та же процедура может быть выполнена со структурным элементом B_n , который также включает в себя 1000 различных вариантов. ДНК-метка B_n может быть лигирована к комплексу A_n -головной фрагмент, и все реакции могут быть объединены. Полученная в результате библиотека включает в себя 1000 x 1000 комбинаций A_n x B_n (т.е., 1000000 соединений), помеченных 1000000 различных комбинаций меток. Такой же подход может быть расширен, чтобы добавить структурные элементы C_n, D_n, E_n и т.д.

Созданная библиотека может быть затем использована для идентификации соединений, которые связываются с мишенью. Структура химических соединений, которые связываются с библиотекой, может необязательно быть оценена с помощью ПЦР и секвенирования ДНК-меток, чтобы идентифицировать соединения, которые были обогащены.

Этот способ может быть модифицирован, чтобы избежать мечения после добавления каждого структурного элемента или чтобы избежать объединения (или смешивания). Например, способ может быть модифицирован путем добавления структурного элемента A_n в n реакционных сосудов, где n представляет собой целое число, больше чем один, и добавления идентичного структурного элемента B_1 в каждую реакционную лунку. В настоящем документе B_1 идентичен для каждого химического соединения, и, следовательно, кодирующая этот структурный элемент олигонуклеотидная метка не требуется. После добавления структурного элемента комплексы могут быть объединены или не объединены. Например, библиотеку не объединяют после заключительной стадии добавления структурного элемента, и пулы подвергают скринингу индивидуально для идентификации соединения(ий), которые связываются с мишенью. Чтобы избежать объединения всех реакций после синтеза, может быть использован BIND[®] Reader (от SRU Biosystems, Inc.), например, для мониторинга связывания на сенсорной поверхности в формате высокой пропускной способности (например, 384-луночные планшеты и 1536-луночные планшеты). Например, структурный элемент A_n может быть закодирован с ДНК-меткой A_n , и структурный элемент B_n может быть закодирован по его положению внутри луночного планшета. Соединения-кандидаты могут быть идентифицированы с использованием анализа связывания (например, с использованием биосенсора BIND[®], который также доступен от SRU Biosystems, Inc., или с использованием анализа ELISA) и путем анализа меток A_n с помощью секвенирования, микроматричного анализа и/или анализа расщепления рестриктазами. Этот анализ позволяет идентифицировать комбинации структурных элементов A_n и B_n , которые производят необходимые молекулы.

Способ амплификации может дополнительно включать в себя образование эмульсии вода-в-масле, чтобы создать множество водных микрореакторов. Условия реакции (например, концентрация комплекса и размер микрореакторов) можно регулировать, чтобы обеспечить, в среднем, микрореактор, содержащий, по меньшей мере, одного представителя из библиотеки соединений. Каждый микрореактор может также содержать мишень, единственную гранулу, способную связываться с

комплексом или частью комплекса (например, одной или несколькими метками) и/или связывать мишень, и раствор для реакции амплификации, содержащий один или несколько необходимых реагентов для выполнения амплификации нуклеиновых кислот. После амплификации метки в микрореакторах, амплифицированные копии метки будут связываться с гранулами в микрореакторах, и покрытые гранулы могут быть идентифицированы с помощью любого применимого способа.

После того, как были идентифицированы структурные элементы из первой библиотеки, которые связываются с представляющей интерес мишенью, вторая библиотека может быть получена в итеративном режиме. Например, могут быть добавлены один или два дополнительных узла разнообразия, и вторую библиотеку создают и испытывают, как описано в настоящем документе. Этот процесс может быть повторен столько раз, сколько необходимо, чтобы создать молекулы с желаемыми молекулярными и фармацевтическими свойствами.

Могут быть использованы различные способы лигирования для добавления каркасного элемента, структурных элементов, спейсеров, связей и меток. Соответственно, любая из описанных в настоящем документе связывающих стадий может включать в себя любую применимую технику или техники лигирования. иллюстративные способы лигирования включают в себя ферментативное лигирование, такое как использование одной или нескольких описанных в настоящем документе РНК-лигаз и/или ДНК-лигаз; и химическое лигирование, такое как использование описанных в настоящем документе химических реакционноспособных пар.

Пример 1

Общая стратегия комплексов

Комплексы согласно настоящему изобретению могут быть образованы, чтобы включать в себя одну или несколько связей, для которых полимераза обладает пониженной способностью считывать или перемещаться. Например, комплексы могут быть образованы, чтобы включать в себя (в направлении от 5' к 3') химическое соединение (звезда), фиксированную константную последовательность (например, головной фрагмент), а затем три переменные кодирующие последовательности (например, три метки), и другую фиксированную константную последовательность (например, хвостовой фрагмент), причем связь между фиксированными константными последовательностями и переменными кодирующими последовательностями может быть нечитаемой (фиг. 1B).

Нечитаемые связи могут быть образованы с помощью любого применимого фрагмента или функциональной группы. Не ограничивающий пример представляет собой псорален (фиг. 2), который может образовывать нечитаемую связь между 5'- и 3'-коннекторами.

Нечитаемые связи могут быть образованы с помощью сшивающего олигонуклеотида с обратимой кореакционноспособной группой циановинилкарбазольной группы. Комплекс может включать в себя (в направлении от 5' к 3') химическое соединение (звезда), фиксированную константную последовательность (например, головной фрагмент), сшивающий олигонуклеотид, который сшивает через обратимую кореакционноспособную группу циановинилкарбазольной группы (сшитый приблизительно в 366 нм, отмеченный X), а затем переменную кодирующую последовательность, содержащую фиксированные константные последовательности на 5'-и 3'-концах (например, метку, содержащую 5'- и 3'-коннекторы) (фиг. 3). Иллюстративная обратимая реакция с обратимыми кореакционноспособными группами (парой циановинилкарбазольной группы и тимидина), которые могут присутствовать в 3'-коннекторе, 5'-коннекторе и сшивающем олигонуклеотиде (фиг. 4). Другие нечитаемые связи могут приводить к способности к деконволюции с использованием фермента(ов) репарации.

В качестве части процесса считывания или деконволюции комплекс может быть подвергнут реакции с химическим процессом и/или ферментативным процессом (например, лигазой и необязательно киназой) для лигирования меток. Когда область соединения между смежными метками представляет собой однонитевый разрыв, то лигаза может быть использована для лигирования меток. При желании, киназа может быть использована для преобразования концевых 5'-гидроксильных групп в фосфатные группы. Когда область соединения между смежными метками представляет собой пробел, требующий расширения одним или несколькими нуклеотидами, то полимеразы могут быть использованы для расширения через пробел перед лигированием меток (необязательно включая в себя киназу). Необязательно полимеразы обладают пониженной активностью замещения цепи. Далее, сшивающие олигонуклеотиды высвобождаются, чтобы образовать шаблон (например, с использованием излучения при приблизительно 312 нм). И, наконец, шаблон при желании амплифицируется с помощью ПЦР и секвенируется.

Для любого из этих комплексов каждая из переменных кодирующих последовательностей может дополнительно включать в себя 5'-коннектор и 3'-

коннектор, которые представляют собой фиксированные константные последовательности. Поскольку связи нечитаемы, могут быть использованы релейные праймеры, чтобы охватить связи и обеспечить расширения с полимеразой, таким образом, образуя олигонуклеотидные фрагменты. Фрагменты могут быть затем лигированы с использованием лигазы, при желании амплифицированы с ПЦР и секвенированы.

Альтернативно, кодирующие последовательности по обе стороны от нечитаемой связи могут быть скопированы в одну последовательность кДНК с использованием опосредованного рекомбинацией процесса, в котором события рекомбинации представляют собой предпочтительные для осуществления для полученных из одной и той же предшествующей шаблонной последовательности последовательностей. Этот механизм сохраняет информацию ассоциации метки и делает эти ассоциации выводимыми из полученных данных последовательностей. Один возможный способ создания нечитаемых связей, которые могут стать шаблонами для подверженной рекомбинации кДНК, представляет собой создание повторных гомологичных областей по обе стороны от связи.

Пример 2

Фотохимическое олигонуклеотидное сшивание с использованием циановинилкарбазола для моделирования одного события мечения

Олигонуклеотид CNVKJ, содержащий две циановинилкарбазольные модификации, получали от Nihon Techno Service (Japan) (фиг. 5). Олигонуклеотиды CNVK2_P_TagB, CNVK2_NP_TagB и CNVK2_TagA получали от Integrated DNA Technologies (IDT) (IA). Каждый олигонуклеотид растворяли в воде до концентрации 1 мМ.

Фотохимические сшивающие условия были следующими: олигонуклеотиды CNVKJ, CNVK2_P_TagB и CNVK2_TagA смешивали в эквимолярном соотношении 100 мкМ каждого в 500 мМ фосфатном буфере при pH 7,0 в 20 мкл или 50 мкл аликвоты в микроцентрифужных пробирках из полипропилена натурального цвета объемом 1,5 мл (Fisher Scientific, 02-682-550). Каждую реакционную смесь затем облучали с использованием лампы UVL-21 (4 Вт) (95-0018-02) (Ultra Violet Products, CA) при 365 нм в течение 2 часов при 4°C. Продукты реакции затем анализировали с помощью электрофореза на 10% денатурирующем ПААГ, а также с помощью ЖХМС (Thermo Fisher Scientific) (фиг. 6). Более 70% исходного материала конвертировали в фотохимически сшитый конъюгат олигонуклеотида с наблюдаемой молекулярной

массой 16894 Да (рассчитанной 16890 Да). Когда ту же сшивающую реакцию проводили в 500 мМ фосфатном буфере при pH 5,5 или 500 мМ боратном буфере при pH 9,5 наблюдались аналогичные результаты и выходы.

Фотохимически сшитый олигонуклеотидный конъюгат CNVK2_P_TagB, CNVK2_TagA и CNVKJ очищали на 10% денатурирующем ПААГ и расшивали при нагревании при 80°C в 50 мкл аликвоты в микроцентрифужных пробирках из полипропилена натурального цвета объемом 1,5 мл (Fisher Scientific, 02- 682-550) при облучении (при 80°C) с использованием лампы Spectroline E-Series EB-160C (6 Вт) (Spectronics Corp., NY) в 312 нм в течение 1 часа. Продукты анализировали с помощью ЖХ-МС. Реакция дает количественный выход неизменных стартовых олигонуклеотидных модельных меток CNVK2_P_TagB с наблюдаемой молекулярной массой 6578 Да (рассчитанной 6579 Да) и CNVK2_TagA с наблюдаемой молекулярной массой 5558 Да (рассчитанной 5558 Да) (фиг. 7).

С целью изучения диссоциации множественных содержащих CNVK сшитых олигонуклеотидов, авторы настоящего изобретения адаптировали вышеизложенный протокол, чтобы подготовить комплекс с двумя олигонуклеотидами CNVK, CNVKJ и Splint_14_CNVK (Splint_14_CNVK представляет собой CGAXCGTGTCAAXCG, где X представляет собой CNVK, синтезированный Biosearch Technologies, CA), фотохимически сшитый с одним олигонуклеотидом, Double_CNVK_Template (AAAAAAGTCGTGACACGTCCGAAAAAACAACGGTGACACGGTCGAAAAA, IDT (IA)). Этот комплекс анализировали с помощью ЖХ-МС: наблюдаемая молекулярная масса 25079 Да (рассчитанная 25055 Да). Затем его растворяли в воде до концентрации 100 мкМ и облучали с помощью УФ 312 нм в течение 1 часа при 80°C, снова, как описано выше. Продукты анализировали с помощью ЖХ-МС, чтобы найти восстановление исходного Double_CNVK_template: наблюдаемая молекулярная масса 16492 Да (рассчитанная 16492 Да) (фиг. 8).

Пример 3

Лигирование и фотохимическое расщивание циановинилкарбазольных фотохимически сшитых олигонуклеотидов для моделирования восстановления информации о последовательности из фотохимически сшитых меток

Получали комплекс CNVK2_P_TagB, CNVK2_TagA и CNVKJ, фотохимически сшивали и очищали на геле, как описано в примере 2.

Фотохимически сшитую модельную пару меток растворяли в 1-кратном буфере T4 ДНК-лигазы (NEB, MA) и инкубировали с T4 ДНК-лигазой (NEB, MA) в течение 1

часа при 37°C. Продукт этой реакции затем расшивали, как описано в примере 2, и анализировали с помощью денатурирующего гель-электрофореза и ЖХМС (фиг. 9). Продукт реакции лигирования наблюдался с высоким выходом, наблюдаемая молекулярная масса 12119 Да (рассчитанная 12119 Да).

Соответствующую фотохимически сшитую нефосфорилированную модельную пару меток CNVK2_NP_TagB с CNVK2_TagA и CNVKJ также получали, как описано выше. Анализ ЖХМС этого комплекса показал наблюдаемую молекулярную массу 16503 Да (рассчитанную 16499 Да, фиг. 10). Фотохимически сшитый комплекс растворяли в 1-кратном буфере T4 лигазы (NEB) и инкубировали со смесью T4 полинуклеотидной киназой и T4 ДНК-лигазы (NEB) в течение 1 часа при 37°C. Продукт этой спаренной реакции фосфорилирования-лигирования затем фотохимически расшивали облучением УФ при 312 нм в течение 1 часа при 80°C, как описано в примере 2. Продукт этой реакции фосфорилирования-лигирования наблюдали с помощью денатурирующего гель-электрофореза и ЖХ-МС, обнаруживая наблюдаемую молекулярную массу 11894 Да (рассчитанная молекулярная масса составляла 11896 Да (дополнительное фосфорилирование на 5'-конце лигированного продукта)) (фиг. 10).

Пример 4

Фотохимическое олигонуклеотидное сшивание с использованием псоралена в стволе для моделирования одного события мечения

Олигонуклеотид 5PSO2_A9_TA, модифицированный C2-псораленом на 5'-конце, получали из Biosearch, CA (фиг. 11). Олигонуклеотид PSO_HP_A9_TCT получали из IDT (IA) (фиг. 11). Оба олигонуклеотида растворяли до 0,5 мМ концентрации каждого в 500 мМ фосфатном буфере при pH 7,0 и отжигали при нагревании до 95°C, а затем охлаждали до комнатной температуры. Затем реакционную смесь облучали ультрафиолетовым светом при 365 нм при 4°C с использованием компактной УФ-лампы UVL-21 (UVP) в течение 30 минут. Продукты анализировали с помощью ЖХ-МС. Авторы настоящего изобретения наблюдали быстрое образование фотохимически сшитого продукта с наблюдаемой молекулярной массой 17643 Да (рассчитанной молекулярной массой 17643,4 Да, фиг. 12).

Пример 5

Фотохимическое олигонуклеотидное лигирование с использованием псоралена для моделирования одного события мечения

Олигонуклеотиды Tag1_PsoCVU и SplintC_PsoC2 получали из IDT (IA) (фиг. 13). Олигонуклеотид 5PSOC2_A9_GA (модифицированный на 5'-конце с C2 псораленом) получали из Biosearch (CA) (фиг. 13).

Фотохимически лигированный олигонуклеотидный конъюгат Tag1_PsoCVU и 5PSOC2_A9_GA получали следующим образом. Фотолигирование достигалось путем фотохимической реакции 5'-псораленпиринового фрагмента одного олигонуклеотида с 3'-тимидином второго олигонуклеотида в присутствии третьего олигонуклеотида, сконструированного для гибридизации с обоими первым и вторым олигонуклеотидом, и совмещали их реакционноспособные концы (фиг. 14).

Фотохимические условия лигирования были следующими: олигонуклеотиды 5PsoC2_A9_GA (конечная концентрация 1 мМ) и Tag1_PsoCVU (конечная концентрация 1,1 мМ) и SplintC_PsoC2 (конечная концентрация 1,1 мМ) объединяли в 10 мкл реакции фотолигирования в 500 мМ фосфатном буфере при pH 7,0 в полипропиленовых микроцентрифужных пробирках объемом 1,5 мл натурального цвета (Fisher Scientific, 02-682-550). Реакционную смесь сначала нагревали до 95°C с последующим медленным охлаждением до комнатной температуры. Затем реакционную смесь облучали ультрафиолетовым светом при 365 нм при 4°C с использованием компактной УФ-лампы UVL-21 (UVP). Аликвоты объемом 1 мкл брали в течение времени цикла и анализировали на 15% денатурирующем полиакриламидном геле с TBE/8M мочевиной (фиг. 15). Гель визуализировали и фотографировали с использованием УФ-затенения флуоресцентного TLC планшета. Фотолигированный конъюгат 5PsoC2_A9_GA и Tag1_PsoCVU образовывался с высоким выходом. Через 24 часа выход составлял приблизительно 90% (фиг. 15).

24-часовой образец анализировали с помощью ЖХ-МС (фиг. 16). Основной пик ЖХ соответствовал фотолигированному конъюгату 5PsoC2_A9_GA и Tag1_PsoCVU с наблюдаемой молекулярной массой 15252,4 Да (рассчитанная молекулярная масса 15243,1 Да).

Пример 6

Получение кДНК из нечитаемых полимеразой фотохимически лигированных сопряженных с псораленом модельных меток с использованием удлинения терминального праймера, удлинения нетерминального праймера и лигирования для моделирования чтения одного события мечения

Псорален-тимидиновый фотохимически лигированный олигонуклеотидный конъюгат (Bio_Pso) получали с использованием способа, аналогичного тому, который

описан в примере 5, со следующими двумя модельными метками-олигонуклеотидами: 5'-биотинилированным олигонуклеотидом 5Bio_Tag_PsoCVU и 5'-псорален олигонуклеотидом 5PsoC2_A9_GA вместе с олигонуклеотидом-накладкой SplintC_PsoC2, все три олигонуклеотида получали от IDT (IA) и они показаны на фиг. 17. Через 24 часа реакции Bio_Pso очищали на 10% денатурирующем ПААГ с ТВЕ-мочевинной и анализировали с помощью ЖХ-МС, обнаруживая продукт с наблюдаемой молекулярной массой 15834,0 Да (рассчитанная молекулярная масса составляла 15812,7 Да) (фиг. 18).

Псорален-тимидиновый фотохимически лигированный олигонуклеотидный конъюгат Bio_Pso использовали в качестве шаблона для создания кДНК с помощью ферментативного расширения как терминального олигонуклеотидного праймера (FAMprimer), так и нетерминального олигонуклеотидного праймера (Phos-SplintC_PsoC2) с Т4 ДНК-полимеразой (NEB, MA). Phos-SplintC_PsoC2 получали с помощью 5'-фосфорилирования SplintC_PsoC2 с использованием Т4 PNK (NEB, MA). Получение кДНК проводили в реакционной смеси объемом 100 мкл, содержащей 10 мкМ каждого из FAMprimer, Phos-SplintC_PsoC2 и Bio_Pso в 1-кратном буфере Т4 ДНК лигазы (NEB, MA), дополненном 1 мМ каждого dNTP и 10 мкл Т4 ДНК-полимеразы (NEB, MA). Реакционную смесь инкубировали при 37°C в течение 1 часа и дополняли АТФ концентрацией 0,5 мМ и Т4 ДНК-лигазой концентрацией 5 мкл (2000 ед/мкл, NEB, MA), а затем инкубировали в течение еще 1 часа при 37°C.

Реакционную смесь затем инкубировали с 200 мкл покрытыми стрептавидином Dynabead M280 (Invitrogen, предварительно промытыми PBS в течение 1 часа при комнатной температуре), гранулы промывали PBS объемом 1 мл, и продукт элюировали с 35 мкл NaOH концентрацией 100 мМ. Элюент затем нейтрализовали добавлением 5 мкл трис-HCl объемом 1 М рН 7,0 и анализировали с помощью ЖХ-МС.

Анализ ЖХМС полученного продукта с детекцией при 495нм показал, что приблизительно 50% FAMprimer были расширены до полноразмерной комплементарной последовательности к шаблону Bio_Pso минус один нуклеотид dA (наблюдаемая молекулярная масса 15170 Да, ожидаемая полноразмерная молекулярная масса 15458 Да). Большая часть остальной части FAMprimer была продлена до области соединения фотохимического лигирования (наблюдаемая молекулярная масса 11154 Да, рассчитанная молекулярная масса 11154 Да) (фиг. 19).

Когда описанное выше получение кДНК проводили без добавления фермента лигазы, полноразмерная комплементарная последовательность (минус один dA) не наблюдалась (фиг. 20).

Пример 7

Фотохимическое олигонуклеотидное лигирование с использованием карбоксивинилуридина для моделирования одного события мечения

Олигонуклеотиды CVU_G и CVU_A синтезировали с помощью Trilink (CA) (фиг. 21). Оба олигонуклеотида представляют собой 5'-модифицированные с 5-(карбоксивинил)-2'-дезоксинуридином (фиг. 21). Олигонуклеотиды Tag1_PsoCVU, SplintC_CVU и SplintA_CVU получали из IDT DNA (IA) (фиг. 21).

Структура области соединения фотохимического лигирования, образованной между облученным 5'-5-(карбоксивинил)-2'-дезоксинуридином и 3'-тимидином, показана на фигуре 22.

10 мкл реакционной смеси фотохимического лигирования создавали с CVU_G концентрацией 1 мМ, Tag1_PsoCVU концентрацией 1,1 мМ и SplintC_CVU концентрацией 1,1 мМ, растворенными в натрий-фосфатном буфере концентрацией 500 мМ, pH 7,0, в полипропиленовых микроцентрифужных пробирках натурального цвета объемом 1,5 мл (Fisher Scientific, 02-682-550). Реакционную смесь нагревали до 95°C с последующим медленным охлаждением до комнатной температуры. Затем реакционную смесь облучали ультрафиолетовым светом при 365 нм в течение 24 часов при 4°C с использованием компактной УФ-лампы UVL-21 (UVP). Аликвоты объемом 1 мкл брали в течение времени цикла и подвергали электрофорезу на 15% денатурирующем полиакриламидном геле с КЭ/8 М мочевиной (фиг. 23). Гель визуализировали и фотографировали с помощью УФ-затенения флуоресцентного TLC планшета. После 24 часов облучения более 90% олигонуклеотида CVU_G было фотохимически лигировано.

24-часовой образец анализировали с помощью ЖХ-МС (фиг. 24). Основной пик соответствует продукту фотохимического лигирования CVU_G с Tag1_PsoCVU. Наблюдаемая молекулярная масса составляла 15260,6 Да (рассчитанная молекулярная масса составляла 15238,7 Да).

Пример 8

Получение кДНК из нечитаемых полимеразой фотохимически лигированных сопряженных с карбоксивинилуридином модельных меток с использованием удлинения терминального праймера, удлинения

нетерминального праймера и лигирования для моделирования чтения одного события мечения

Карбоксивинилуридин-тимидиновый фотохимически лигированный олигонуклеотидный конъюгат (Bio_CVU) получали с использованием способа, аналогичного тому, который описан в примере 7, со следующими двумя модельными метками-олигонуклеотидами: 5'-5-карбоксивинилуридиновым олигонуклеотидом 5'-CVU_A и 5'-биотинилированным олигонуклеотидом 5Bio_Tag_PsoCVU вместе с олигонуклеотидом-накладкой SplintA_CVU, все три олигонуклеотида получали от IDT (IA) и они показаны на фиг. 21 и 25. Продукт реакции очищали на геле, как описано в примере 6, и анализ ЖХ-МС обнаруживал продукт с наблюдаемой молекулярной массой 15811,8 Да (рассчитанная молекулярная масса составляла 15792,3 Да) (фиг. 25).

Карбоксивинилуридин-тимидиновый фотохимически лигированный олигонуклеотидный конъюгат Bio_CVU использовали в качестве шаблона для создания кДНК с помощью ферментативного расширения как терминального олигонуклеотидного праймера (FAMprimer), так и нетерминального олигонуклеотидного праймера (Phos-SplintA_CVU) с Т4 ДНК-полимеразой (NEB, MA). Phos-SplintA_CVU получали с помощью 5'-фосфорилирования SplintA_CVU с использованием Т4 PNK (NEB, MA). Получение кДНК проводили в реакционной смеси объемом 100 мкл, содержащей 10 мкМ каждого из FAMprimer, Phos-SplintA_CVU и Bio_CVU в 1-кратном буфере Т4 ДНК лигазы (NEB, MA), дополненном 1 мМ каждого dNTP и 10 мкл Т4 ДНК-полимеразы (NEB, MA). Реакционную смесь инкубировали при 37°C в течение 1 часа и дополняли АТФ в концентрации 0,5 мМ и 5 мкл Т4 ДНК-лигазы (2000 ед/мкл, NEB, MA), а затем инкубировали в течение еще 1 часа при 37°C.

Реакционную смесь затем инкубировали с 200 мкл покрытыми стрептавидином Dynabead M280 (Invitrogen, предварительно промытыми PBS в течение 1 часа при комнатной температуре), гранулы промывали с 1 мл PBS, и продукт элюировали с 35 мкл NaOH концентрацией 100 мМ. Элюент затем нейтрализовали добавлением 5 мкл трис-HCl в концентрации 1 М pH 7,0 и анализировали с помощью ЖХ-МС.

ЖХМС анализ полученного продукта с детекцией при 495нм показал, что приблизительно 50% FAMprimer были продлены до полноразмерной комплементарной последовательности к шаблону Bio_CVU минус один нуклеотид dA (наблюдаемая молекулярная масса 15176 Да, ожидаемая полноразмерная молекулярная масса 15458 Да). Большая часть остальной части FAMprimer была продлена до области соединения

фотохимического лигирования (наблюдаемая молекулярная масса 10560 Да, рассчитанная молекулярная масса 10559 Да) (фиг. 26).

Когда описанное выше кДНК получение проводили без добавления фермента лигазы, полноразмерная комплементарная последовательность (минус один dA) не наблюдалась.

Пример 9

Олигонуклеотидное лигирование с использованием нечитаемой полимеразой клик-конъюгации для моделирования одного события мечения

Олигонуклеотиды S_IDTN₃ и S_IDTalkyne получали из IDT (IA) (фиг. 27).

3'-азидная модификация олигонуклеотида S_IDTN₃ представляет собой собственную модификацию IDT (3' IDT азидная модификация/3AzideN/(MM 350)).

Содержащие модификацию /3AzideN/ олигонуклеотиды очищали ВЭЖХ.

Олигонуклеотиды, содержащие гексинильную модификацию, обессоливали и использовали без дополнительной очистки.

Клик-конъюгация этих модельных меток достигалась с помощью Cu (I), катализируемой 3 + 1 азидоалкиновым циклоприсоединением следующим образом:

Маточные растворы готовили следующим образом: Cu(OAc)₂·H₂O (молекулярная масса по формуле 181,6), 10 мМ в ДМФ; аскорбат натрия (молекулярная масса по формуле 198,1), H₂O в концентрации 20 мМ и трис-(бензилтриазолилметил)амин (ТВТА) (молекулярная масса по формуле 530,6), DMF в концентрации 10 мМ. В пробирку Эппендорфа добавляли 10 мкл pH 7,0 фосфатного буфера (500 мМ водный раствор), 10 мкл S_IDTalkyne (1 мМ водный раствор, 10 нмоль, 1 эквивалент) и 10 мкл S-IDTN₃ (1 мМ водный раствор, 10 нмоль, 1 молярный эквивалент). К этому раствору добавляли 5 мкл Cu_Pre-смеси (2 молярных эквивалента Cu(OAc)₂, 4 молярных эквивалента аскорбата натрия, 1 молярный эквивалент ТВТА). Реакционную смесь инкубировали при комнатной температуре в течение ночи. Продукт анализировали с помощью ЖХ-МС (Thermo) (фиг. 28) и установленная молекулярная масса составляла 26304,8 Да (рассчитанная 26311,2 Да). Конъюгат подвергали дальнейшей очистке с помощью электрофореза с использованием 10% ПААГ/8М мочевины. Этот продукт конъюгации назвали Conjugate_Click_S.

Пример 10

Олигонуклеотидное лигирование с использованием пары ортогональных нечитаемых полимеразой клик-конъюгаций для моделирования пары ортогональных событий мечения

Олигонуклеотиды TKR_Central и S_IDTN₃ получали из IDT (фиг. 29). 5'-добензоциклооктин (DBCO)-модифицированный олигонуклеотид TKR_DBCO_S получали из Biosearch Technologies (CA) (фиг. 29).

Во-первых, клик-конъюгацию проводили с TKR_Central и TKR_DBCO_S с использованием основанной на циклооктине безмедной реакции. Равные молярные количества каждого олигонуклеотида (10 мкл 1 мМ водных растворов) смешивали в 0,2 М натрий-фосфатном буфере при pH 7,0. Ход реакции контролировали с помощью ЖХ-МС. Реакцию завершали через 30-60 мин с получением более чем 90% чистого конъюгата. Масс-спектрометрия указывала на молекулярную массу 25336 Да (рассчитанная 25341,8 Да) (фиг. 30).

Во-вторых, клик-конъюгацию проводили между продуктами конъюгации TKR_Central и TKR_DBCO_S с S_IDTN₃ с использованием катализируемой Cu (I) клик-процедуры, описанной в примере 9. Полученный окончательный конъюгат 2cl_S очищали на полиакриламидном геле 10% КЭ/8М мочевины и анализировали с помощью ЖХМС (Thermo), выявляя ожидаемую молекулярную массу для 2cl_S 38578,1 Да (рассчитанная 38560,4 Да) (фиг. 31).

Пример 11

ПЦР-амплификация кДНК, полученной из нечитаемых полимеразой клик-конъюгированных модельных меток, используя опосредованную нетерминальным праймером рекомбинацию в свободном растворе и в эмульгированных водных каплях для моделирования чтения информации о последовательности из химически конъюгированных меток

Conjugate_Click_S (фиг. 32) готовили путем конъюгации S_IDTN₃ и S_IDTalkyne, как описано в примере 9.

Conjugate_Click_L (фиг. 32) готовили путем конъюгации L_IDTN₃: TGCGGTCTAACTGTCTAGGCACTTGTTCGTTTGCCAGTGTGAGGAATGAACAGG/3 AzideN/ и L_IDTalkyne: /5hexynyl/TACCGAATTGCCTTCCTCGTACAGTTCTAAGGCGCTTGGACACCACTTC AATCGGGTGCTATGCTT (IDT, IA) по аналогии с протоколом конъюгации, описанным в примере 10. ЖХМС после очистки гелем подтвердил молекулярную массу как 37582,6 Да (рассчитанная 37584,4 Да).

Олигодезоксинуклеотидные праймеры, включающие в себя прямой праймер: TGCGGTCTAACTGTCTA, обратный праймер: AAG CATAGCACCCGATT и ePsplit: GGCAATTCGGTACCTGTTCATTCC получали из (IDT, IA).

Конъюгаты Conjugate_Click_L и Conjugate_Click_S каждый разводили до 1 нМ и использовали в качестве шаблонов для опосредованного зависимой от накладки рекомбинацией получения кДНК/ПЦР-амплификации с использованием Deep Vent (exo) ДНК-полимеразы (NEB, MA). 50 мкл реакционной смеси Deep Vent (exo-) получали следующим образом: 5 мкл 10-кратного буфера Thermopol; 2,5 мкл 10 мкМ раствора каждого прямого и обратного праймера, 1 мкл праймера ePsplint в концентрации 100 нМ; 2,5 мкл смеси dNTP (NEB) в концентрации 10 мМ; 5 мкл 2 нМ разбавления либо конъюгата, либо их смеси 1:1 2,5 мкл Deep Vent (exo-) и воды до 50 мкл. Реакционную смесь циркулировали 24 раза следующим образом: 30 секунд при 94°C, 30 секунд при 52°C и 60 секунд при 72°C.

Когда конъюгаты Conjugate_Click_L или Conjugate_Click_S амплифицируют отдельно в свободном растворе в присутствии обоих терминальных праймеров и праймера-накладки "ePsplint" и затем оценивали с помощью электрофореза, каждый приводит к росту одного продукта с подвижностью, которая коррелирует с длиной шаблона (полосы 5 и 7 на фиг. 33). Похожие амплификации при отсутствии праймера-накладки "ePsplint" не давали продукта (полосы 4 и 6 на фиг. 33). Это говорит о том, что опосредованное зависимой от накладки рекомбинацией получение кДНК/ПЦР-амплификация обоих конъюгатов с помощью Deep Vent (exo-) полимеразы зависит от наличия праймера ePsplint (фиг. 33). Возможный механизм этой рекомбинации и амплификации представлен на фиг. 34.

Когда оба конъюгата Conjugate_Click_L и Conjugate_Click_S смешивают и амплифицируют вместе в свободном растворе в присутствии обоих терминальных праймеров и праймера-накладки "ePsplint" и затем оценивают с помощью электрофореза, лестница продуктов с разной подвижностью создается с подвижностями, которые коррелируют с длинами каждого из шаблонов, а также с подвижностями, которые коррелируют с длинами, промежуточными между каждым из шаблонов (полоса 5 на фиг. 35). Возможный механизм этой рекомбинации и амплификации представлен на фигуре 36.

Когда оба конъюгата Conjugate_Click_L и Conjugate_Click_S смешивают и амплифицируют вместе в эмульсии вода-в-масле в присутствии как терминальных праймеров, так и праймера-накладки "ePsplint" при низкой концентрации конъюгатов по отношению к каплям эмульсии, а затем оценивают с помощью электрофореза, при этом наблюдается пара продуктов с подвижностями, которые коррелируют с длиной каждого из шаблонов (полоса 3 на фиг. 37). Капли эмульсии выступают в качестве

отдельных микрореакторов, в пределах которых отдельные выделенные конъюгаты амплифицируют. Фазу масло/поверхностно-активное вещество готовили смешиванием 4,5% (об/об) Span 80 (Fluka, номер в каталоге 85548), 0,4% (об/об) Твин 80 (Sigma, номер в каталоге S-8074) и 0,05% (об/об) Тритон X-100 (Sigma, номер в каталоге T-9284) в светлом минеральном масле (Sigma, номер в каталоге M-3516). Водную фазу (50 мкл) получали путем смешивания 5 мкл 10-кратного буфера Thermopol; 2,5 мкл раствора каждого праймера с концентрацией 10 мкМ, 1 мкл праймера ePsplit в концентрации 100 нМ; 2,5 мкл 10 мМ смеси dNTP (NEB); 2,5 мкл DVexo- (NEB), 2,5 мкл BSA (NEB) и воды до 50 мкл. Смесь 1:1 конъюгатов Conjugate_Click_L и Conjugate_Click_S добавляли к реакции до конечной концентрации 5 пМ. К 50 мкл водной фазы затем добавляли 150 мкл неводной фазы и эмульгирование проводили встряхиванием в течение 5 минут. Процесс рекомбинации/амплификации проводили следующим образом: 30 секунд при 94°C, 30 секунд при 52°C и 60 секунд при 72°C в течение 32 циклов.

После опосредованного зависимой от накладки рекомбинацией получения кДНК/ПЦР-амплификации эмульсию нарушали микроцентрифугированием при 21000g в течение 20 минут, а затем безводную фазу удаляли и водную фазу и межфазу экстрагировали в хлороформе. Водную фазу затем оценивали с помощью нативного электрофореза (4% агарозные минигели, Invitrogen, CA).

Результаты эмульсионной компартиментализации опосредованного зависимой от накладки рекомбинацией получения кДНК/амплификации ПЦР показывают, что степень образования перемешанных продуктов ПЦР значительно снижается или отсутствует (полоса 3, фиг. 37).

Пример 12

ПЦР-амплификация кДНК, полученной из нечитаемых полимеразой клик-конъюгированных модельных меток с использованием опосредованной повторной гомологичной рекомбинации в свободном растворе для моделирования чтения одного события мечения

Два конъюгата, TKR_S и TKR_L (фиг. 38), каждый из которых содержит одну нечитаемую полимеразой клик-конъюгированную связь, образованную между 3'-азидомодифицированным олигонуклеотидом и 5'-гексинил-модифицированным олигонуклеотидом с использованием катализируемого Cu (I) циклоприсоединения и протокола, описанного в примере 9. TKR_S и TKR-L каждый содержат дублированную 12-мерную последовательность GGAATGAACAGG, которая фланкирует нечитаемую

полимеразой связать. TKR_S получали путем конъюгирования следующих двух олигонуклеотидов IDT ДНК, TKR_S_N₃: TGCGGTCTAACTGTCTAGTTTCACSTTCTCCGGAATGAACAGG/3AzidoN/ (также упоминаемого как S_IDTN₃ в примере 9) и TKR_S_Hex:/5'-Hexynyl/GGAATGAACAGGGCCTGATTGAAGCAATCGGGTGCTATGCTT, в то время как TKR_L получали путем конъюгирования двух олигонуклеотидов IDT ДНК TKR_L_N₃: TGCGGTCTAACTGTCTAGGCACTTGTTTCGTTTGCCAGTGTGATGGACACCACTTGG AATGAACAGG/3AzideN/ и TKR_L_Hex: /5'Hexynyl/GGAATGAACAGGGGAATGAACAGGTTCCCTCGTACAGTTCTAAGGCGCTCAATCGGGTGCTATGCTT. Оба конъюгата очищали в ПААГ в денатурирующих условиях и анализировали с помощью ЖХ-МС для выявления молекулярных масс для TKR_S: 26447,5 Да (26449,2 Да рассчитанная), так для TKR_L 37582,6 Да (37584,4 рассчитанная).

Каждый из двух конъюгатов разводили до 1 нМ и использовали для опосредованного зависимой от повтора рекомбинации получения кДНК/ПЦР-амплификации с Deep Vent (exo-) ДНК-полимеразой (NEB) в свободном растворе с прямым праймером TGCGGTCTAACTGTCTA и обратным праймером AAGCATAGCACCCGATT (IDT).

Реакцию опосредованного зависимой от повтора рекомбинацией получения кДНК/ПЦР-амплификации проводили с использованием Deep Vent (exo-) ДНК-полимеразы (NEB). 5 мкл 10-кратного буфера Thermopol; 2,5 мкл 10 мкМ раствора каждого прямого и обратного праймера; 2,5 мкл 10 мМ смеси dNTP (NEB); 40 пМ в конечной концентрации либо конъюгата, либо их смеси 1: 1, 2,5 мкл Deep Vent (exo-) полимеразы и воду до конечного объема 50 мкл. Реакционную смесь подвергали термическим циклам 22 раза следующим образом: 30 секунд при 94°C, 30 секунд при 52°C и 60 секунд при 72°C.

Продукты визуализировали на 4% агарозном минигеле (Invitrogen) (фиг. 39).

Реакция опосредованного зависимой от повтора рекомбинацией получения кДНК/ПЦР-амплификации Deep Vent (exo-) полимеразы производила большое количество ампликона, полученного от каждого отдельно амплифицированного конъюгата (полосы 3 и 4, фиг. 39). Амплификация смеси конъюгатов в тех же условиях не приводила к образованию каких-либо продуктов амплификации промежуточной длины (полоса 5, фиг. 39). Для дальнейшего подтверждения отсутствия перестановки

(потери) ассоциаций меток (закодированной информации) в этих продуктах амплификации, авторы настоящего изобретения клонировали и секвенировали продукты амплификации, полученные из коамплифицированной смеси TKR_S и TKR_L. Эти данные последовательностей подтверждали полное отсутствие перестановки пар модельных меток в продуктах амплификации, полученных из коамплифицированных конъюгатов (фиг. 40).

Пример 13

ПЦР-амплификация кДНК, полученной из нечитаемых дважды клик-конъюгированных меток с использованием опосредованной повторной гомологией рекомбинации в свободном растворе для моделирования чтения пары помеченных событий

Два конъюгата, 2_cl_S и 2_cl_L (фиг. 41), каждый содержащий две нечитаемые связи клик-конъюгации, каждая фланкированная различными 12-мономерными повторяющимися последовательностями (GAAGGCGTTATG-клик-GAAGGCGTTATG) и (GGAATGAACAGG-клик-GGAATGAACAGG), соответственно, были подготовлены в процессе двухступенчатой процедуры, аналогичной той, что описана в примере 10.

Во-первых, конъюгацию проводили с использованием основанной на циклооктине свободной от меди клик-реакции между TKR_central:/5'-Hexynyl/GGAATGAACAGGGTAAGCTGGAGTGAAGGCGTTATG/3azideN/(IDT) и либо TKR_DBCO_S: (DBCO) GAAGGCGTTATGTCCGTA CTCTTGCAATCGGGTGCTATGCTT, либо TKR_DBCO_L (DBCO)

GAAGGCGTTATGTGATATCCGTGGTGTCTGAGTTCCAATCGGGTGCTATGCTT.

Оба содержащие DBCO олигонуклеотида получали из Biosearch (CA), DBCO представляет собой модификацию 5'-добензоциклооктина, для которого фосфорамидитный реагент можно получить из Glen Research (VA)). Продукты этих конъюгаций были названы TKR_S_N₃ и TKR_L_N₃, соответственно.

Во-вторых, конъюгацию проводили между продуктами первых конъюгаций, TKR_S_N₃ и TKR_L_N₃ с использованием катализируемой Cu (I) клик-реакции. Полученные конъюгаты, TKR_2_click_S и TKR_2_click_L (фиг. 41) очищали с использованием электрофореза на полиакриламидном геле с 10% КЭ/8 М мочевиной и анализировали с помощью ЖХМС (Thermo), выявляя ожидаемую молекулярную массу для TKR_2_click_S 38578,1 Да (рассчитанная 38560,4 Да) и для TKR_2_click_L 49913,3 Да (рассчитанная 49916,7 Да) (фиг. 41).

Каждый из двух конъюгатов разводили до 1 нМ и использовали для опосредованного зависимой от повтора рекомбинацией получения кДНК/ПЦР-амплификации с Deep Vent (exo-) полимеразой (NEB). Для амплификации прямой праймер TGCGGTCTAACTGTCTA и обратный праймер AAGCATAGCACCCGATT (IDT ДНК) использовали с 5 мкл 10-кратного буфера Thermopol; 2,5 мкл 10 мкМ раствора каждого прямого и обратного праймера; 2,5 мкл 10 мМ смеси dNTP (NEB); 40 пМ конечной концентрации либо конъюгата, либо их смеси 1:1, 2,5 мкл Deep Vent (exo-) полимеразы и воду до 50 мкл. Реакционную смесь подвергали термическим циклам 22 раза следующим образом: 30 секунд при 94°C, 30 секунд при 52°C и 60 секунд при 72°C. Продукты амплификации визуализировали на 4% агарозных минигелях (Invitrogen).

Реакция опосредованного зависимой от повтора рекомбинацией получения кДНК/ПЦР-амплификации Deep Vent (exo-) полимеразы производила большие количества ампликона, полученного от каждого амплифицированного отдельно конъюгата (полосы 3 и 4, фиг. 42). Амплификация смеси конъюгатов в тех же условиях не приводила к образованию каких-либо продуктов амплификации промежуточной длины (полоса 5, фиг. 42). Для дальнейшего подтверждения отсутствия перестановки (потери) ассоциаций меток (закодированной информации) в этих продуктах амплификации, авторы настоящего изобретения клонировали и секвенировали продукты амплификации, полученные из коамплифицированной смеси TKR_2_click_S и TKR_2_click_L. Эти данные последовательностей подтверждали полное отсутствие перестановки пар модельных меток в продуктах амплификации, полученных из коамплифицированных конъюгатов (фиг. 43).

Другие варианты осуществления

Все упоминаемые в настоящем документе публикации, патентные заявки и патенты включены в настоящий документ посредством ссылки.

Различные модификации и вариации описанного способа и системы согласно настоящему изобретению будут очевидны специалистам в настоящей области техники без отхождения от объема и сущности настоящего изобретения. Несмотря на то, что настоящее изобретение было описано в связи с конкретными желаемыми вариантами осуществления, следует понимать, что настоящее изобретение не должно быть неправомерно ограничено такими конкретными вариантами осуществления. В самом деле, различные модификации описанных способов осуществления настоящего

изобретения, которые очевидны специалистам в области медицины, фармакологии или смежных областях, предназначены, чтобы быть в пределах объема настоящего изобретения.

ПЕРЕЧЕНЬ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

<110> ИКС-ЧЕМ, ИНК.

<120> ДНК-КОДИРОВАННЫЕ БИБЛИОТЕКИ, СОДЕРЖАЩИЕ СВЯЗИ КОДИРУЮЩИХ
ОЛИГОНУКЛЕОТИДОВ, НЕ ДОСТУПНЫЕ ДЛЯ СЧИТЫВАНИЯ ПОЛИМЕРАЗАМИ

<130> 50719-021WO2

<140> PCT/US2013/050303

<141> 2013-07-12

<150> 61/671,406

<151> 2012-07-13

<160> 41

<170> PatentIn version 3.5

<210> 1

<211> 7

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic Construct

<400> 1
gtgctgc 7

<210> 2

<211> 9

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic Construct

<400> 2
gcagcaccc 9

<210> 3

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Synthetic Construct

<220>

<221> MISC_FEATURE

<222> (1)..(1)

<223> Xaa is N-alpha-benzyloxycarbonyl-(Glycine)3-N-succinimidyl ester)

<220>

<221> MISC_FEATURE

<222> (8)..(8)

<223> Xaa is OSU (N-succinimidyl ester)

<400> 3

Xaa Gly Gly Gly Gly Gly Gly Xaa
1 5

<210> 4
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> Xaa is azidohomoalanine

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa is OSu (N-succinimidyl ester)

<400> 4

Xaa Gly Gly Gly Xaa
1 5

<210> 5
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> MISC_FEATURE
<222> (1)..(1)
<223> Xaa is propargylglycine

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa is Osu (N-succinimidyl ester)

<400> 5

Xaa Gly Gly Gly Xaa
1 5

<210> 6
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>

<221> MISC_FEATURE
<222> (1)..(1)
<223> Xaa is an (N-alpha-benzyloxycarbonyl-(Glycine)3-N-succinimidyl ester

<220>
<221> MISC_FEATURE
<222> (5)..(5)
<223> Xaa is OSu (N-succinimidyl ester)

<400> 6

Xaa Gly Gly Gly Xaa
1 5

<210> 7
<211> 14
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> misc_feature
<222> (4)..(4)
<223> n is CNVK

<220>
<221> misc_feature
<222> (12)..(12)
<223> n is CNVK

<400> 7
cgancgtgtc ancg

14

<210> 8
<211> 53
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 8
aaaaaagtcg tgacacgtcg gaaaaaaaaa aaacggtgac acggtcgaaa aaa

53

<210> 9
<211> 55
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> misc_feature
<222> (55)..(55)
<223> n is a 3'-azide (WHS ester)

<400> 9
tgcggtctaa ctgtctaggg acttggttcgt ttgccagtgt gaggaatgaa caggn 55

<210> 10
<211> 67
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a 5'-hexynyl

<400> 10
ntaccgaatt gccttcctcg tacagttcta aggcgcttgg acaccacttc aatcgggtgc 60

tatgctt 67

<210> 11
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 11
tgcggtctaa ctgtcta 17

<210> 12
<211> 17
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 12
aagcatagca cccgatt 17

<210> 13
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 13
ggcaattcgg tacctgttca ttcc 24

<210> 14
<211> 12
<212> DNA

<213> Artificial Sequence
 <220>
 <223> Synthetic Construct
 <400> 14
 ggaatgaaca gg 12

<210> 15
 <211> 43
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct

<220>
 <221> misc_feature
 <222> (43)..(43)
 <223> n is a 3'-azide (WHS Esther)
 <400> 15
 tgcggtctaa ctgtctagtt caccttctcc ggaatgaaca ggn 43

<210> 16
 <211> 43
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a 5'-Hexynyl
 <400> 16
 nggaatgaac agggcctgat tgaagcaatc gggtgctatg ctt 43

<210> 17
 <211> 67
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct

<220>
 <221> misc_feature
 <222> (67)..(67)
 <223> n is a 3'-azide (WHS Ester)
 <400> 17
 tgcggtctaa ctgtctaggc acttggtcgt ttgccagtgt gatggacacc acttggaatg 60
 aacaggn 67

<210> 18
 <211> 67
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Construct

 <220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a 5'-Hexynyl

 <400> 18
 nggaatgaac aggggaatga acaggttcct cgtacagttc taaggcgctc aatcgggtgc 60
 tatgctt 67

 <210> 19
 <211> 17
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Construct

 <400> 19
 tgcggtctaa ctgtcta 17

 <210> 20
 <211> 38
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Construct

 <220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a 5'-Hexynyl

 <220>
 <221> misc_feature
 <222> (38)..(38)
 <223> n is a 3'-azide (WHS Ester)

 <400> 20
 nggaatgaac agggtaagct ggagtgaagg cgttatgn 38

 <210> 21
 <211> 43
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Construct

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a 5'-dibenzo-cyclooctyne

 <400> 21
 ngaaggcggt atgtccgtac tcttgcaatc gggtgctatg ctt 43

<210> 22
 <211> 55
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Construct

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a 5'-dibenzo-cyclooctyne

 <400> 22
 ngaaggcggt atgtgatatc cgtgggtgctg tgagttccaa tcgggtgcta tgctt 55

<210> 23
 <211> 18
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Construct

<400> 23
 aaaaaaaaaa gtcgtgac 18

<210> 24
 <211> 19
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Construct

<400> 24
 acgtcggtaa aaaaaaaaaa 19

<210> 25
 <211> 15
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic Construct

<220>
 <221> misc_feature
 <222> (1)..(1)

<223> Cyanovinylcarbozole (CNVK)
 <220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is Cyanovinylcarbozole (CNVK)
 <220>
 <221> misc_feature
 <222> (15)..(15)
 <223> Cyanovinylcarbozole (CNVK)
 <400> 25
 ncgacgtgac acgan 15
 <210> 26
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct
 <220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a 5' Psoralen (C2-linked)
 <220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a 5' Psoralen (C2)
 <400> 26
 ntacgagatg caaaaaaaaaa ggcgagcttg cgtactg 37
 <210> 27
 <211> 20
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct
 <400> 27
 aaaaaaaaaag catccgctct 20
 <210> 28
 <211> 37
 <212> DNA
 <213> Artificial Sequence
 <220>
 <223> Synthetic Construct
 <220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a 5' Psoralen (C2-linked)

<400> 28
ngagcggatg caaaaaaaaaa ggcgagcttg cgtactg 37

<210> 29
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 29
tccgctcgac ctgag 15

<210> 30
<211> 12
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 30
aaaaactcag gt 12

<210> 31
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a 5' biotin

<400> 31
naaaaactca ggt 13

<210> 32
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<400> 32
tccgctcgac ctgag 15

<210> 33
<211> 16
<212> DNA
<213> Artificial Sequence

<220>
 <223> Synthetic Construct

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a 56-FAM-Phosphoramidite

<400> 33
 nagtacgcaa gctcgc 16

<210> 34
 <211> 16
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<220>
 <221> misc_feature
 <222> (1)..(1)
 <223> n is a monophosphate

<400> 34
 ntccgctcga cctgag 16

<210> 35
 <211> 15
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<400> 35
 tccgctcaac ctgag 15

<210> 36
 <211> 15
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<400> 36
 tccgcttaac ctgag 15

<210> 37
 <211> 37
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic Construct

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a 5-(Carboxy)vinyl-2'-deoxyuridine

<400> 37
ngagcggatg caaaaaaaaaa ggcgagcttg cgtactg

37

<210> 38
<211> 37
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a 5-(Carboxy)vinyl-2'-deoxyuridine

<400> 38
naagcggatg caaaaaaaaaa ggcgagcttg cgtactg

37

<210> 39
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> misc_feature
<222> (43)..(43)
<223> n is a, c, g, or t

<220>
<221> misc_feature
<222> (47)..(47)
<223> n is a 3'-azide (WHS Ester)

<400> 39
tgcggtctaa ctgtctagtt caccttctcc ggaatgaaca ggn

43

<210> 40
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> misc_feature
<222> (1)..(1)
<223> n is a 5'-Hexyynyl

<400> 40
ntaccgaatt gccgcctgat tgaagcaatc gggtgctatg ctt

43

<210> 41
<211> 43
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic Construct

<220>
<221> misc_feature
<222> (43)..(43)
<223> n is a 3'-azide (WHS Ester)

<400> 41
tgcggtctaa ctgtctagtt caccttctcc ggaatgaaca ggn

43

ФОРМУЛА ИЗОБРЕТЕНИЯ

1. Комплекс, содержащий:

(i) химическое соединение, содержащее один или несколько каркасных элементов или один или несколько структурных элементов;

(ii) первую олигонуклеотидную метку, кодирующую идентификацию по меньшей мере одного из указанных одного или нескольких каркасных элементов или структурных элементов и

(iii) головной фрагмент, содержащий первую функциональную группу и вторую функциональную группу, причем указанная первая функциональная группа функционально связана с указанным химическим соединением и указанная вторая функциональная группа функционально связана с указанной первой меткой через первую связь, для которой полимераза обладает пониженной способностью к считыванию или перемещению.

2. Комплекс, содержащий:

(i) химическое соединение, содержащее один или несколько каркасных элементов или один или несколько структурных элементов;

(ii) n число олигонуклеотидных меток, содержащих $n-1$ связей, где n представляет собой целое число от 1 до 10 и где каждая из указанных связей находится между двумя соседними метками и каждая метка кодирует идентификацию по меньшей мере одного из указанных одного или нескольких каркасных элементов или структурных элементов и

(iii) головной фрагмент, содержащий первую функциональную группу, функционально связанную с указанным химическим соединением, и вторую функциональную группу, функционально связанную по меньшей мере с одной из указанного числа n меток через первую связь, причем полимераза обладает пониженной способностью считывать или перемещаться через по меньшей мере одну из указанных первой связи и $n-1$ связей.

3. Комплекс по п. 2, в котором полимераза обладает пониженной способностью считывать или перемещаться через по меньшей мере приблизительно 10% указанных первой связи и $n-1$ связей.

4. Комплекс по любому из пп. 1-3, в котором одна или несколько меток содержат 5'-коннектор на 5'-конце указанной одной или нескольких меток и 3'-коннектор на 3'-конце указанной одной или нескольких меток.

5. Комплекс по п. 4, в котором каждый 5'-коннектор содержит ту же последовательность.

6. Комплекс по п. 4 или 5, в котором каждый 3'-коннектор содержит ту же последовательность.

7. Комплекс по п. 4, в котором каждый 5'-коннектор и/или каждый 3'-коннектор содержит другую последовательность.

8. Комплекс по любому из пп. 1-7, в котором первая связь и/или $n-1$ связей составляют по меньшей мере приблизительно 3 ангстрема в длину.

9. Комплекс по любому из пп. 1-8, в котором первая связь и/или $n-1$ связей составляют по меньшей мере приблизительно 30 ангстрем в длину.

10. Комплекс по любому из пп. 1-9, в котором ДНК-полимераза и/или РНК-полимераза обладает пониженной способностью считывать или перемещаться через первую связь и/или $n-1$ связей.

11. Комплекс по любому из пп. 1-10, в котором менее чем приблизительно 50% от указанной первой связи и $n-1$ связей содержат ферментативную связь.

12. Комплекс по любому из пп. 1-11, в котором первая связь и/или $n-1$ связей содержат химическую связь.

13. Комплекс по п. 12, в котором указанная химическая связь содержит химическую реакционноспособную группу, фотореакционноспособную группу, интеркалирующий фрагмент или сшивающий олигонуклеотид.

14. Комплекс по п. 13, в котором по меньшей мере одна химическая реакционноспособная группа, фотореакционноспособная группа или интеркалирующий фрагмент присутствует в 5'-коннекторе на или в непосредственной близости от 5'-конца указанной метки и/или в 3'-коннекторе на или в непосредственной близости от 3'-конца указанной метки.

15. Комплекс по п. 14, в котором последовательность по меньшей мере одного указанного 5'-коннектора комплементарна последовательности смежного 3'-коннектора или идентична или достаточно схожа, чтобы обеспечить гибридизацию с тем же комплементарным олигонуклеотидом по обе стороны от области соединения в отдельных событиях гибридизации.

16. Комплекс по любому из пп. 13-15, в котором указанную химическую реакционноспособную группу выбирают из пары необязательно замещенной алкинильной группы и необязательно замещенной азидогруппы; пары необязательно замещенного диена, содержащего систему 4 π -электронов, и необязательно замещенного диенофила или необязательно замещенного гетеродиенофила, содержащего систему 2 π -электронов; пары нуклеофила и напряженного гетероциклического электрофила; пары необязательно замещенной аминогруппы и альдегидной или кетонной группы; пары необязательно замещенной аминогруппы и карбоксильной группы; пары необязательно замещенного гидразина и альдегидной или кетонной группы; пары необязательно замещенного гидроксилamina и альдегидной или кетонной группы; пары нуклеофила и необязательно замещенного алкилгалогенида; платинового комплекса; алкилирующего средства или модифицированного фураном нуклеотида.

17. Комплекс по любому из пп. 13-15, в котором фотореакционноспособная группа содержит интеркалирующий фрагмент, производное псоралена, необязательно замещенную циановинилкарбазольную группу, необязательно замещенную винилкарбазольную группу, необязательно замещенную циановинильную группу, необязательно замещенную акриламидную группу, необязательно замещенную диазириновую группу, необязательно замещенный бензофенон, необязательно замещенную 5-(карбоксивинилуридин)овую группу или необязательно замещенную азидную группу.

18. Комплекс по любому из пп. 13-15, в котором интеркалирующий фрагмент представляет собой производное псоралена, производное алкалоида, катион этидия, производное акридина, производное антрациклина или талидомид.

19. Комплекс по п. 18, в котором указанное производное псоралена представляет собой псорален, 8-метоксипсорален или 4-гидроксиметил-4,5,8-триметилпсорален.

20. Комплекс по п. 13, в котором указанная химическая связь содержит указанный сшивающий олигонуклеотид, причем последовательность по меньшей мере пяти нуклеотидов на 5'-конце указанного сшивающего олигонуклеотида комплементарна последовательности по меньшей мере пяти нуклеотидов на 3'-конце одной или нескольких меток и причем последовательность по меньшей мере пяти нуклеотидов на 3'-конце указанного сшивающего олигонуклеотида комплементарна последовательности по меньшей мере пяти нуклеотидов на 5'-конце одной или нескольких меток.

21. Комплекс по п. 20, в котором 3'-конец одной или нескольких меток содержит 3'-коннектор.

22. Комплекс по п. 20 или 21, в котором 5'-конец одной или нескольких меток содержит 5'-коннектор.

23. Комплекс по любому из пп. 20-22, в котором 5'-конец и/или 3'-конец указанного сшивающего олигонуклеотида содержит обратимую кореакционноспособную группу.

24. Комплекс по п. 21 или 22, в котором 3'-коннектор и/или 5'-коннектор содержит обратимую кореакционноспособную группу.

25. Комплекс по п. 23 или 24, в котором обратимая кореакционноспособная группа представляет собой циановинилкарбазольную группу, циановинильную группу, акриламидную группу, тиольную группу или винилсульфовую группу.

26. Комплекс по любому из пп. 1-25, в котором химическое соединение функционально связано с указанным головной фрагментом через бифункциональный спейсер.

27. Комплекс по любому из пп. 1-26, в котором химическое соединение ковалентно присоединено к указанному головному фрагменту.

28. Комплекс по любому из пп. 1-27, в котором головной фрагмент содержит олигонуклеотид, выбранный из группы, состоящей из двухцепочечного олигонуклеотида, одноцепочечного олигонуклеотида или олигонуклеотида-шпильки.

29. Комплекс по п. 28, в котором головной фрагмент содержит связывающую праймер область.

30. Комплекс по любому из пп. 1-29, дополнительно содержащий одну или несколько идентифицирующую первую библиотеку метку(и), метку(и) использования и/или метку(и) происхождения.

31. Комплекс по любому из пп. 1-30, в котором комплекс содержит от 2 до 20 меток.

32. Комплекс по любому из пп. 1-31, в котором каждая из указанных меток содержит от 5 до 75 нуклеотидов.

33. Комплекс по любому из пп. 1-32, в котором каждая из указанных меток в пределах отдельного набора меток характеризуется приблизительно той же массой.

34. Комплекс по любому из пп. 1-33, в котором комплекс содержит РНК, ДНК, модифицированную ДНК и/или модифицированную РНК.

35. Комплекс по п. 34, в котором модифицированная ДНК или модифицированная РНК представляют собой ПНК, LNA, GNA, TNA или их смесь в пределах одного и того же олигонуклеотида.

36. Комплекс по любому из пп. 1-35, дополнительно содержащий хвостовой фрагмент.

37. Библиотека, содержащая один или несколько комплексов по любому из пп. 1-36.

38. Библиотека по п. 37, причем указанная библиотека содержит множество головных фрагментов.

39. Библиотека по п. 38, в которой каждый головной фрагмент указанного множества головных фрагментов содержит идентичную область последовательности и отличную кодирующую область.

40. Библиотека по п. 39, в которой указанные идентичная область последовательности представляет собой связывающую праймер область.

41. Библиотека по п. 39 или 40, в которой указанная отличная кодирующая область представляет собой первую метку, которая кодирует использование библиотеки, происхождение библиотеки, идентификацию библиотеки, историю библиотеки, связь, спейсер или добавление первого компонента или олигонуклеотидную последовательность, которая облегчает технологии гибридизации, амплификации, клонирования или секвенирования.

42. Библиотека по любому из пп. 37-41, содержащая от приблизительно от 10^2 до 10^{20} комплексов.

43. Библиотека по любому из пп. 37-42, в которой каждый комплекс отличается.

44. Способ мечения первой библиотеки, включающий закодированное химическое соединение, причем указанный способ включает:

(а) получение головного фрагмента, содержащего первую функциональную группу и вторую функциональную группу;

(б) связывание указанной первой функциональной группы указанного головного фрагмента с первым компонентом указанного химического соединения, причем

головной фрагмент непосредственно соединен с указанным первым компонентом или указанный головной фрагмент опосредованно соединен с указанным первым компонентом с помощью бифункционального спейсера; и

(с) связывание указанной второй функциональной группы указанного головного фрагмента с первой олигонуклеотидной меткой через первую связь с образованием комплекса, в котором полимераза обладает пониженной способностью считывать или перемещаться через указанную первую связь;

причем указанные стадии (b) и (с) могут быть выполнены в любом порядке, и причем указанная первая метка кодирует реакцию связывания указанной стадии (b), тем самым обеспечивая меченую библиотеку.

45. Способ мечения первой библиотеки, включающий закодированное химическое соединение, причем указанный способ включает:

(a) получение головного фрагмента, содержащего первую функциональную группу и вторую функциональную группу;

(b) связывание указанной первой функциональной группой указанного головного фрагмента с первым компонентом указанного химического соединения, причем указанный головной фрагмент непосредственно соединен с указанным первым компонентом или указанный головной фрагмент опосредованно соединен с указанным первым компонентом с помощью бифункционального спейсера;

(с) связывание указанной второй функциональной группы указанного головного фрагмента с первой олигонуклеотидной меткой через первую связь;

(d) связывание n_c количества дополнительных компонентов указанного химического соединения, причем n_c представляет собой целое число от 1 до 10; и

(e) связывание n_t количества дополнительных олигонуклеотидных меток, содержащих n_t связей, с образованием комплекса, причем n_t представляет собой целое число от 1 до 10 и причем каждая из указанных связей находится между двумя соседними метками и каждая метка кодирует идентификацию по меньшей мере одного из указанных компонентов;

причем полимераза обладает пониженной способностью считывать или перемещаться через по меньшей мере одну из указанной первой связи и n_t связей; и

причем указанные стадии (b) и (с) могут быть выполнены в любом порядке, и причем указанная первая метка кодирует реакцию связывания указанной стадии (d); причем указанные стадии (d) и (e) могут быть выполнены в любом порядке, и причем

каждая дополнительная метка кодирует реакцию связывания каждого дополнительного компонента указанной стадии (d), тем самым обеспечивая меченую библиотеку.

46. Способ по п. 45, при котором полимеразы обладает пониженной способностью считывать или перемещаться через по меньшей мере приблизительно 10% от указанной первой связи и n , связей.

47. Способ по любому из пп. 44-46, при котором указанный первый компонент и/или дополнительные компоненты содержат каркасный элемент или структурный элемент.

48. Способ по любому из пп. 44-47, при котором стадия (b) включает связывание указанного головного фрагмента опосредованно с указанным первым компонентом с помощью бифункционального спейсера.

49. Способ по любому из пп. 44-48, дополнительно включающий связывание метки использования и/или метки происхождения с указанным комплексом.

50. Способ по любому из пп. 44-49, дополнительно включающий связывание идентифицирующей первую библиотеку метки с указанным комплексом.

51. Способ по п. 50, дополнительно включающий обеспечение второй библиотеки и объединение указанной первой библиотеки с указанной второй библиотекой.

52. Способ по любому из пп. 44-51, дополнительно включающий связывание хвостового фрагмента с указанным комплексом.

53. Способ по любому из пп. 44-52, при котором указанная стадия связывания содержит химическую реакционноспособную группу, фотореакционноспособную группу, интеркалирующий фрагмент или сшивающий олигонуклеотид.

54. Способ по любому из пп. 44-53, дополнительно включающий выполнение синтеза разделения-и-объединения в одной или нескольких стадиях, отделяя один или несколько комплексов, и/или очистку одного или нескольких комплексов.

55. Способ по любому из пп. 44-53, дополнительно включающий комплекс по любому из пп. 1-36 и/или библиотеку по любому из пп. 37-43.

56. Способ скрининга множества химических соединений, причем указанный способ включает:

(a) контактирование мишени с комплексом по любому из пп. 1-36 и/или библиотекой по любому из пп. 37-43 и

(b) выбор одного или нескольких комплексов, характеризующихся заранее определенной характеристикой для указанной мишени, по сравнению с контролем, таким образом, подвергая скринингу указанное химическое соединение.

57. Способ по п. 56, при котором указанная заранее определенная характеристика включает увеличенное связывание указанной мишени, по сравнению с контролем.

58. Способ по любому из пп. 56 или 57, при котором перед стадией (a) выполняется одна или несколько операций, указанные операции, выбирают из группы, состоящей из: отжига одного или нескольких релейных праймеров с указанным комплексом, причем указанный один или несколько релейных праймеров перекрывают указанную первую связь и/или $n-1$ связей; расширения указанных релейных праймеров с использованием полимеразы для получения олигонуклеотидных фрагментов и лигирования указанного олигонуклеотидного фрагмента для получения шаблона.

59. Способ определения нуклеотидной последовательности комплекса по любому из пп. 1-36, причем указанный способ включает:

(a) отжиг одного или нескольких релейных праймеров с указанным комплексом, причем указанный один или несколько релейных праймеров перекрывают указанную первую связь и/или $n-1$ связей;

(b) расширение указанных релейных праймеров с использованием полимеразы для получения олигонуклеотидных фрагментов;

(с) сшивание указанных олигонуклеотидных фрагментов для получения шаблона;

(d) дополнительная амплификация указанного шаблона с помощью полимеразной цепной реакции для получения амплифицированной смеси и

(е) секвенирование дополнительно амплифицированной смеси для определения последовательности указанного комплекса.

60. Способ по п. 59, при котором указанный комплекс содержит 5'-коннектор на 5'-конце одной или нескольких меток и 3'-коннектор на 3'-конце одной или нескольких меток и при котором каждый из указанных одного или нескольких релейных праймеров гибридизуется с соседними 5'- и 3'-коннекторами.

61. Способ по п. 60, при котором каждый 5'-коннектор содержит ту же последовательность.

62. Способ по п. 61, при котором каждый 3'-коннектор содержит ту же последовательность.

63. Способ по любому из пп. 60-62, при котором последовательность по меньшей мере одного из указанных 5'-коннекторов комплементарна последовательности смежного 3'-коннектора или идентична или достаточно схожа, чтобы позволить гибридизацию с тем же комплементарным олигонуклеотидом по обе стороны от области соединения в отдельных событиях гибридизации.

64. Способ по п. 63, дополнительно включающий гибридизацию указанного 5'-коннектора с указанным смежным 3'-коннектором.

65. Способ по любому из пп. 60-64, при котором каждый релейный праймер для конкретной области соединения между 5'- и 3'-коннекторами содержит ту же последовательность.

66. Способ определения нуклеотидной последовательности комплекса по любому из пп. 1-36, причем указанный способ включает:

(a) получение указанного комплекса, содержащего химическую связь, причем химическая связь представляет собой сшивающий олигонуклеотид, который перекрывает область соединения между двумя соседними метками и который функционально связан с указанными двумя соседними метками через одну или несколько обратимых кореакционноспособных групп;

(b) высвобождение указанного сшивающего олигонуклеотида для получения шаблона;

(c) дополнительную амплификацию указанного шаблона с помощью полимеразной цепной реакции для получения амплифицированной смеси и

(d) секвенирование указанной дополнительно амплифицированной смеси для определения последовательности указанного комплекса.

67. Способ по п. 66, при котором часть указанного сшивающего олигонуклеотида гибридизуется с концами двух смежных меток, таким образом, получая однонитевый разрыв или пробел между указанными двумя соседними метками.

68. Способ по п. 67, дополнительно включающий лигирование указанных меток с использованием химического процесса или ферментативного процесса перед стадией (b).

69. Способ по п. 68, при котором химический процесс или ферментативный процесс включает 5'-фосфорилирование.

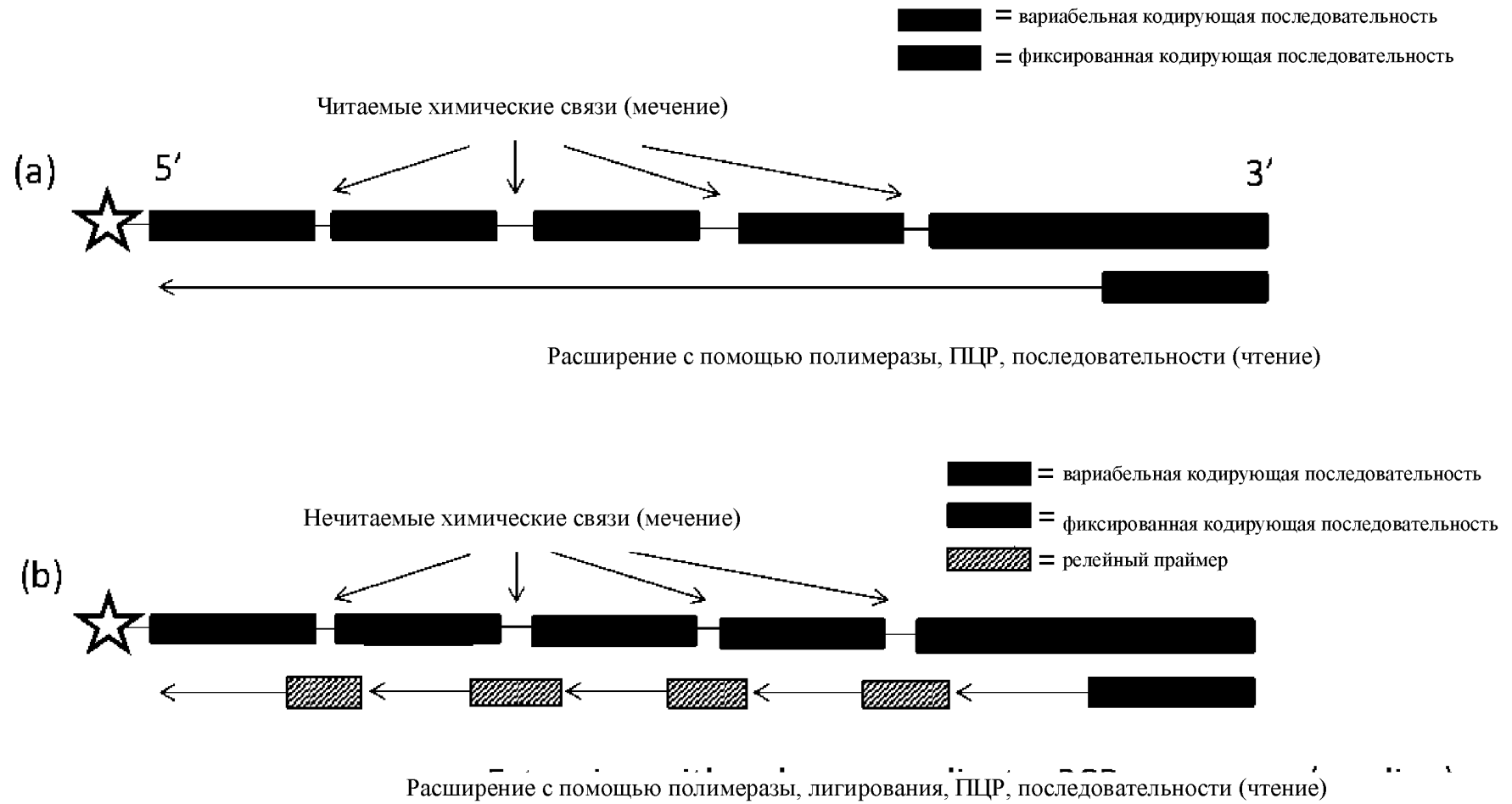
70. Способ по любому из пп. 59-69, дополнительно включающий стадию репарации указанного шаблона перед указанной дополнительной стадией амплификации или указанной стадией секвенирования.

71. Способ по п. 70, при котором стадия репарации включает модификацию по меньшей мере одной из указанных первой связи и/или $n-1$ связей для репарации связи для способности считывания или перемещения полимеразой.

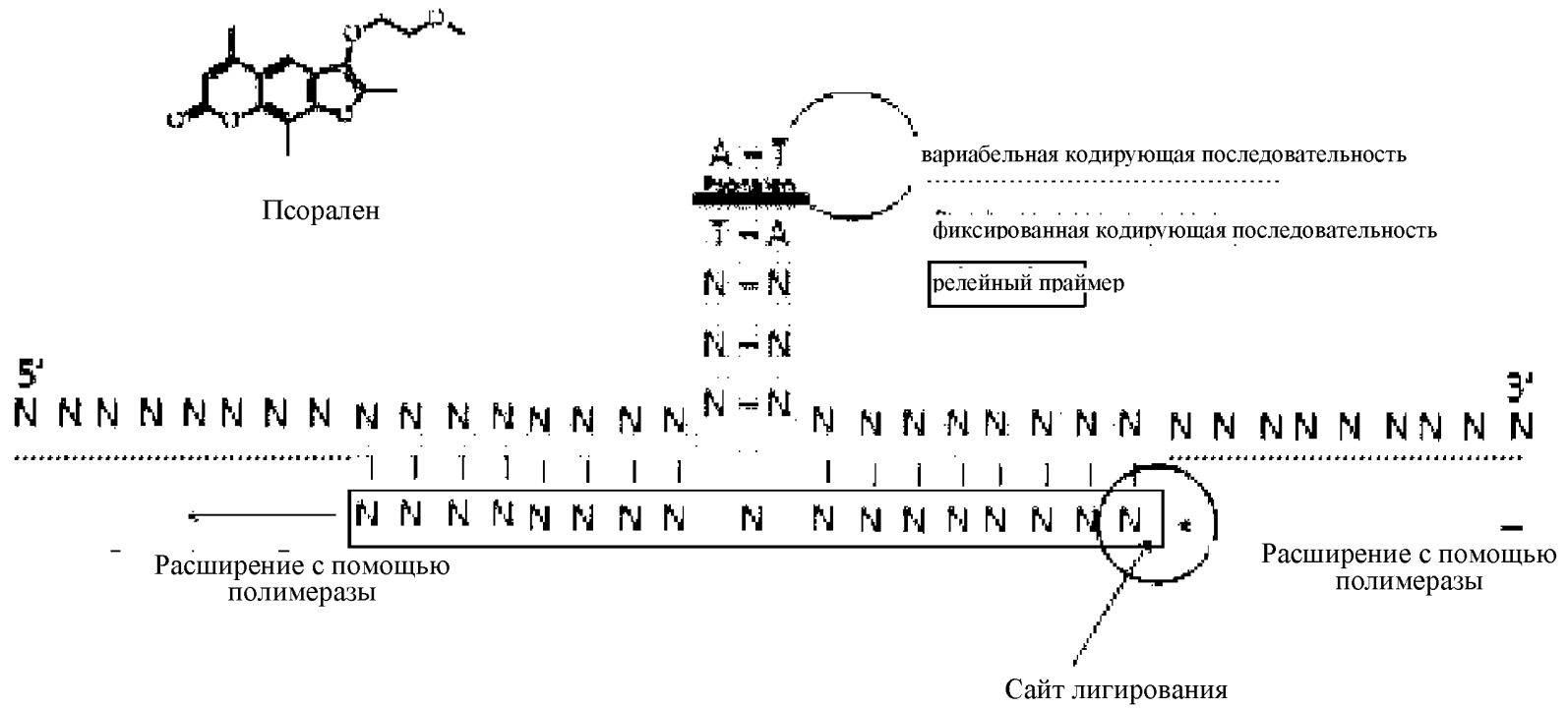
72. Способ по п. 71, при котором указанная полимераза может считывать или перемещаться через по меньшей мере 90% указанных репарированных связей.

73. Способ по любому из пп. 70-72, при котором указанная стадия репарации включает обеспечение указанного шаблона с ферментом репарации.

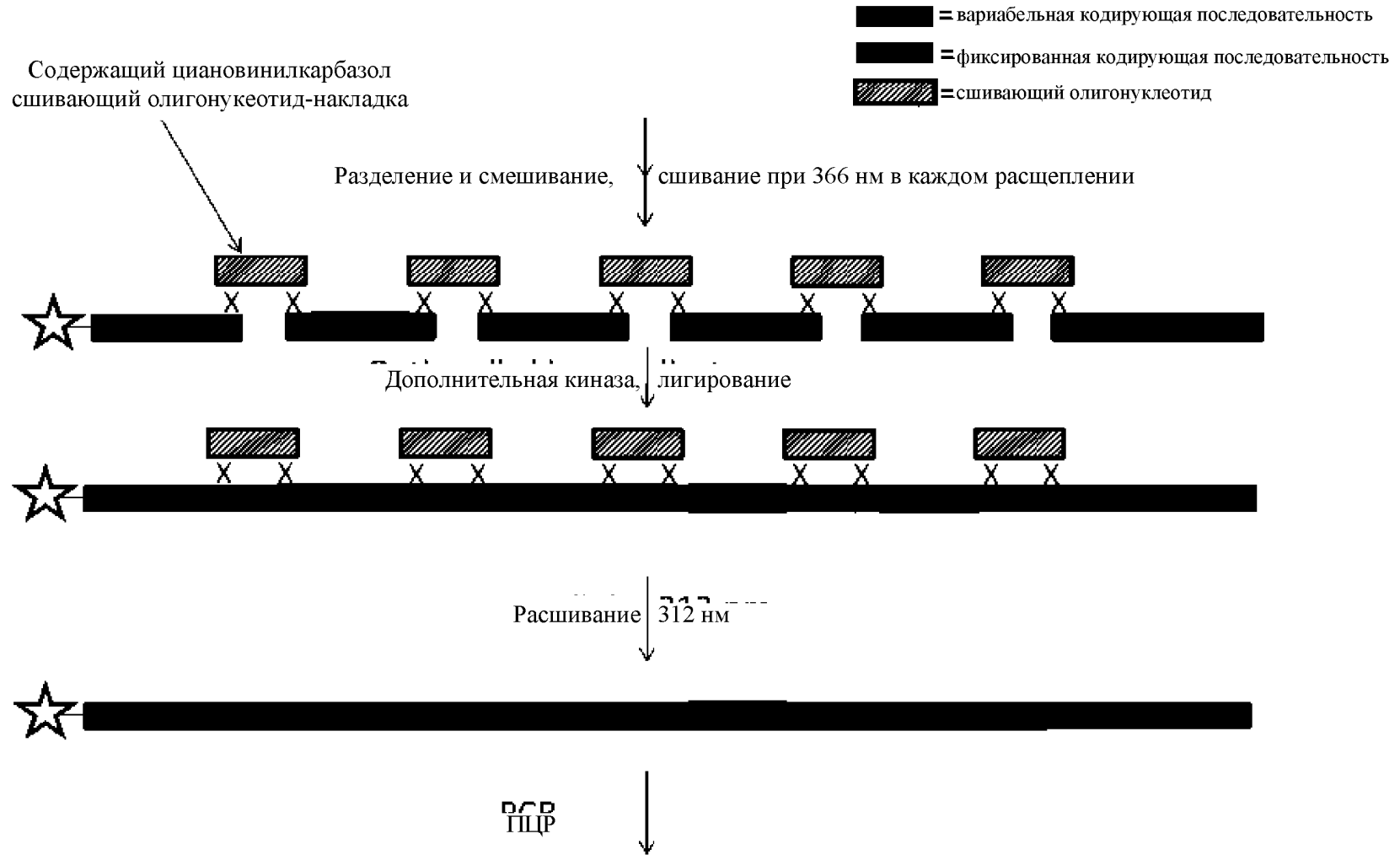
74. Способ по п. 73, при котором указанный фермент репарации представляет собой фотолиазу, гликозилазу, эндонуклеазу, флэп-эндонуклеазу, апуриную/апиримидиновую (AP) эндонуклеазу поли(АДФ-рибоза)полимеразу или метилтрансферазу.



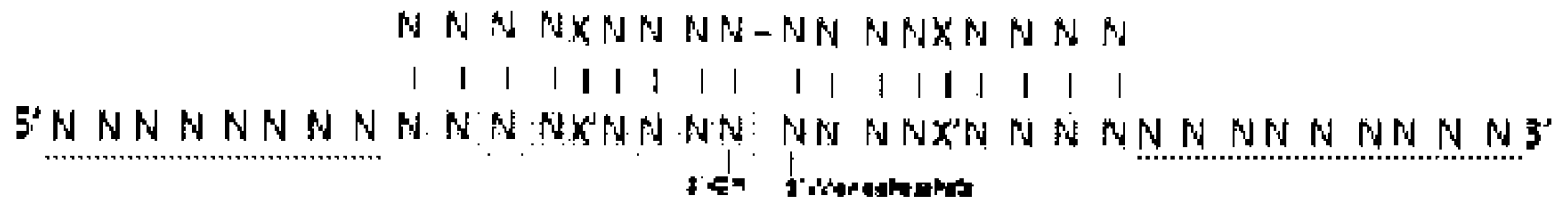
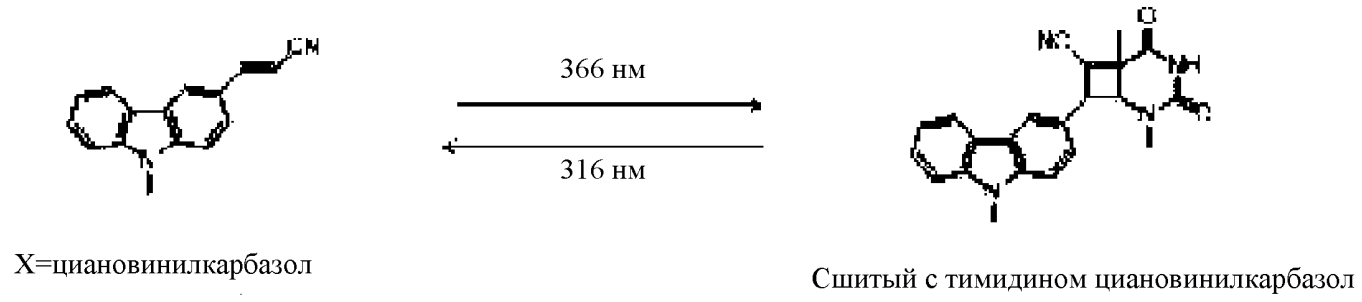
Фиг. 1



Фиг. 2



Фиг. 3



вариабельная кодирующая последовательность

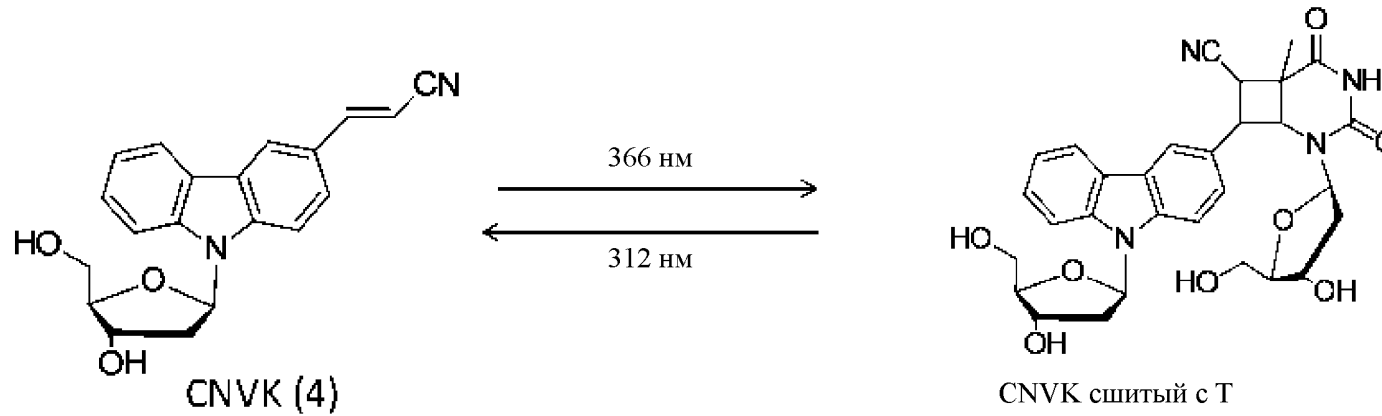
 фиксированная кодирующая последовательность
 сшивающий олигонуклеотид

Фиг. 4

Олигонуклеотид CNVKJ

5'-XCGACGTGTCACGAX-3'

X= циановинилкарбазол (CNVK)



Олигонуклеотид CNVK2_P_TagB (MM 6579)

5'-монофосфо-ACGTCGGTAAAAAAAAAAAA

Олигонуклеотид CNVK_2_NP_TagB (MM 6184)

ACGTCGGTAAAAAAAAAAAA

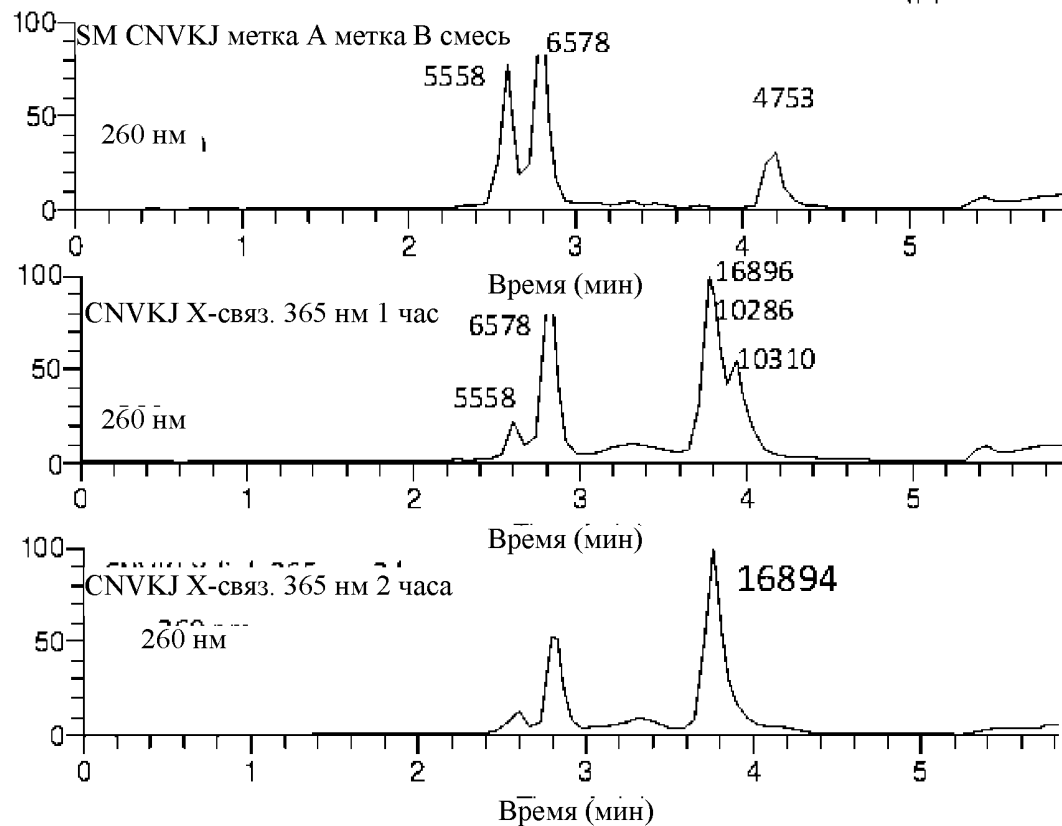
Олигонуклеотид CNVK2_TagA (MM 5558)

AAAAAAAAAAGTCGTGAC

Фиг. 5

5'-XCGACGT---- GTCACGAX-3'

3'-AAAAAAAAAATGGCTGCA(p) CAGTGCTGAAAAAAAA-5'



CNVKJ 4753 Да

AAAAAAAAAATGGCTGCA
CAGTGCTGAAAAAAAA-3'

CNVK2 TagB 6578 Да
CNVK2 Р TagB 6579 Да
AAAAAAAAAATGGCTGCA

A+B+CNVKJ = 16890

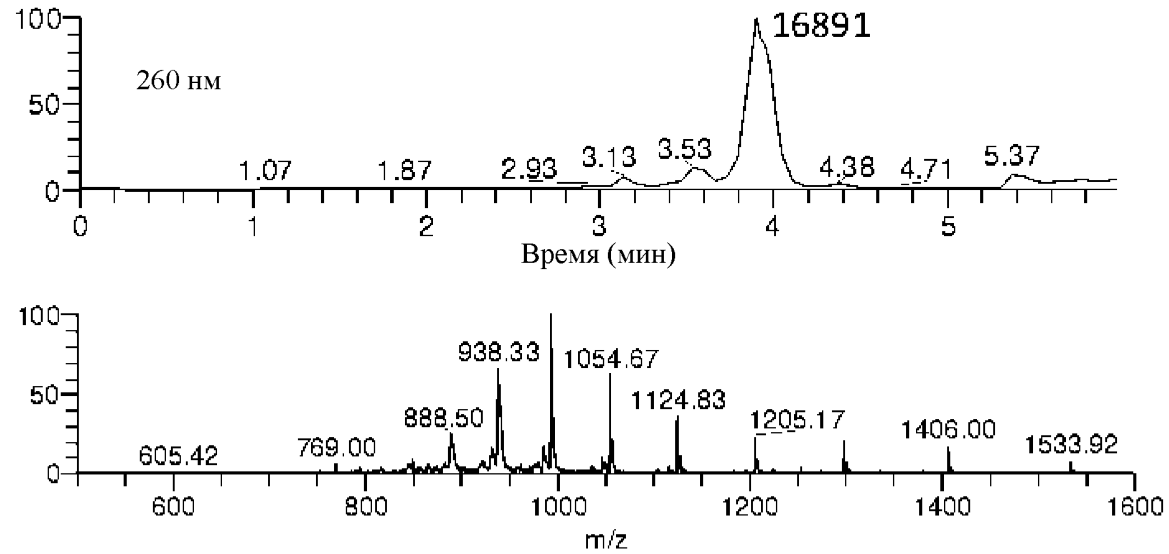
A+CNVKJ = 10311

B+CNVKJ = 11332

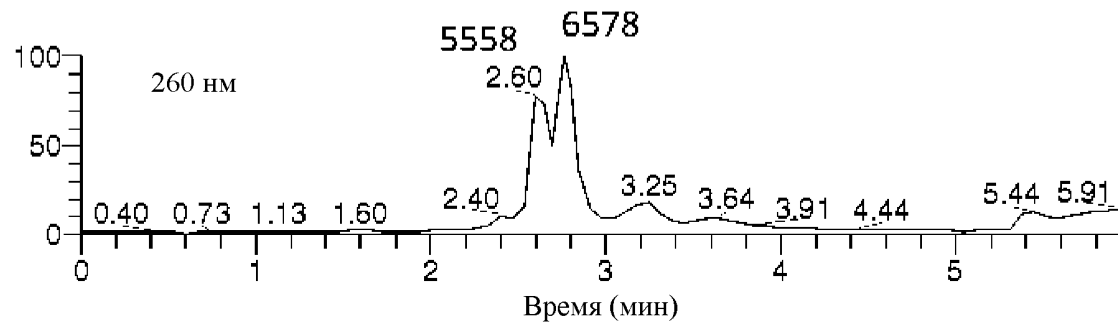
CNVK2 TagA 5558 Да
CNVK2 TagA 5558 Да
AAAAAAAAAATGGCTGCA

Фиг. 6

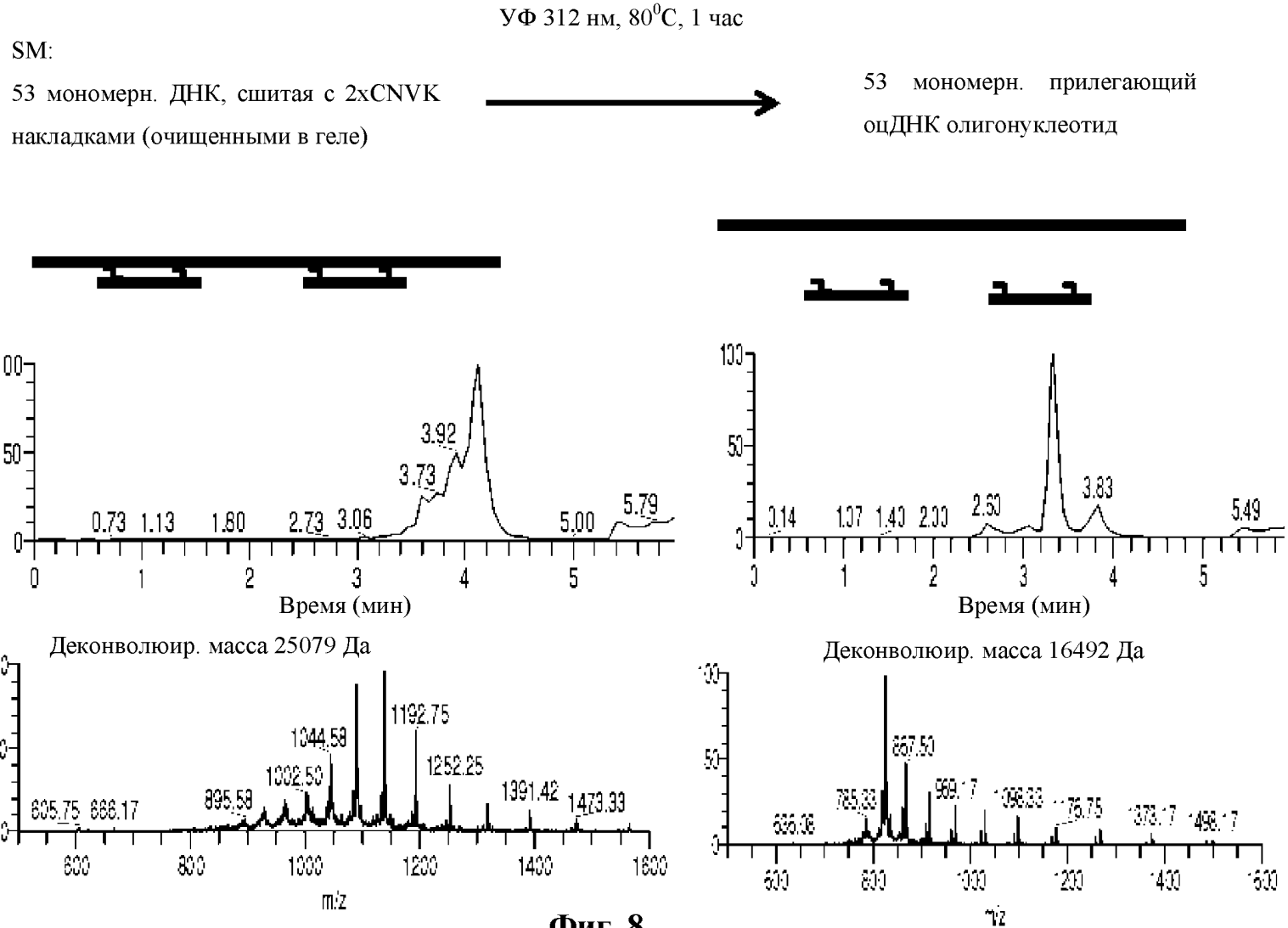
Очищенный в геле X-сшитый комплекс до лигирования



Очищенный в геле X-сшитый комплекс до лигирования, расшитый от x при 312 нм в течение 1,5 ч при комнатной темп.

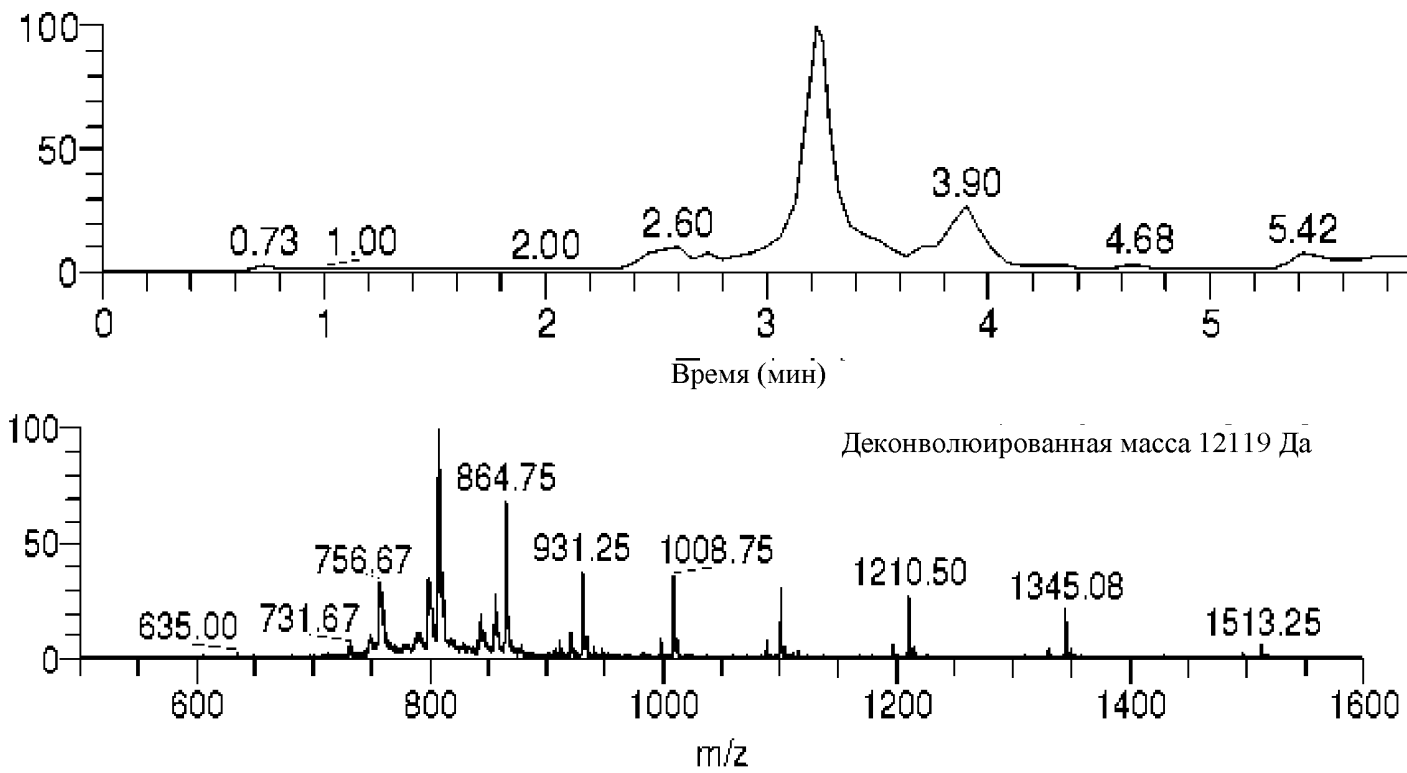


Фиг. 7



Фиг. 8

Т4 ДНК лигирование CNVK X-сшитого комплекса с последующим
расшиванием X в течение 1 часа при 312 нм при 80°C



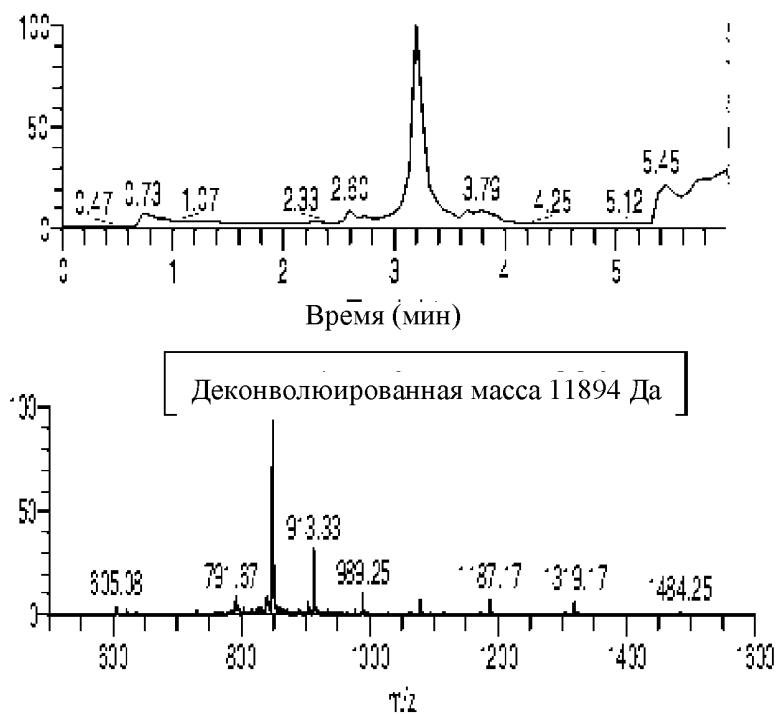
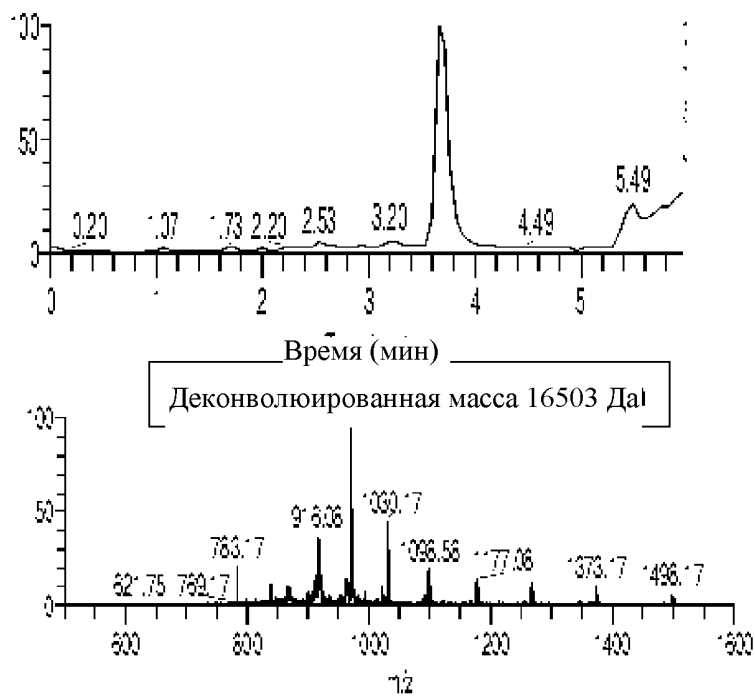
Фиг. 9

SM:

Метка А + метка В + комплекс-накладка

Продукт:

Метка А, лигированная с меткой В



Фиг. 10

5PSO2_A9_TA (MM 11556,3 Да)

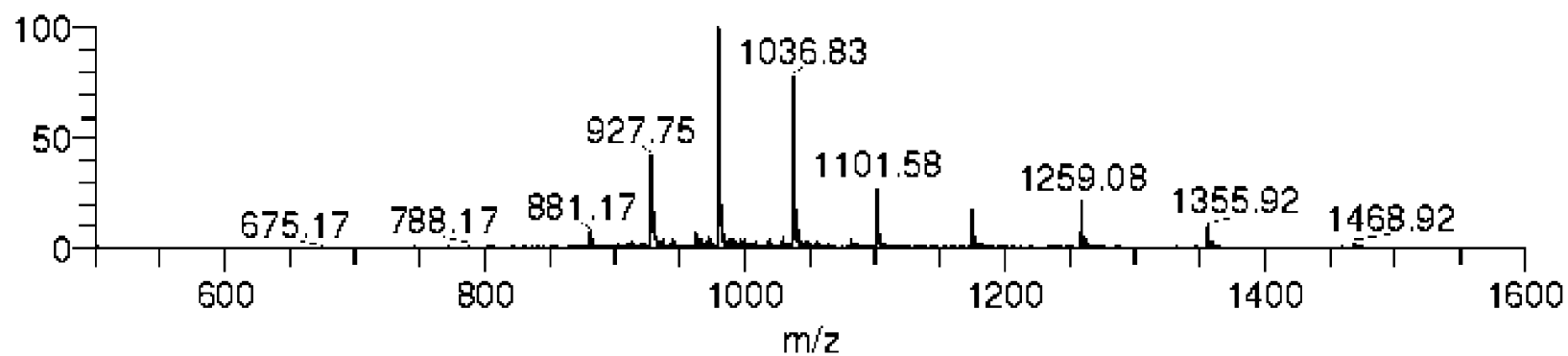
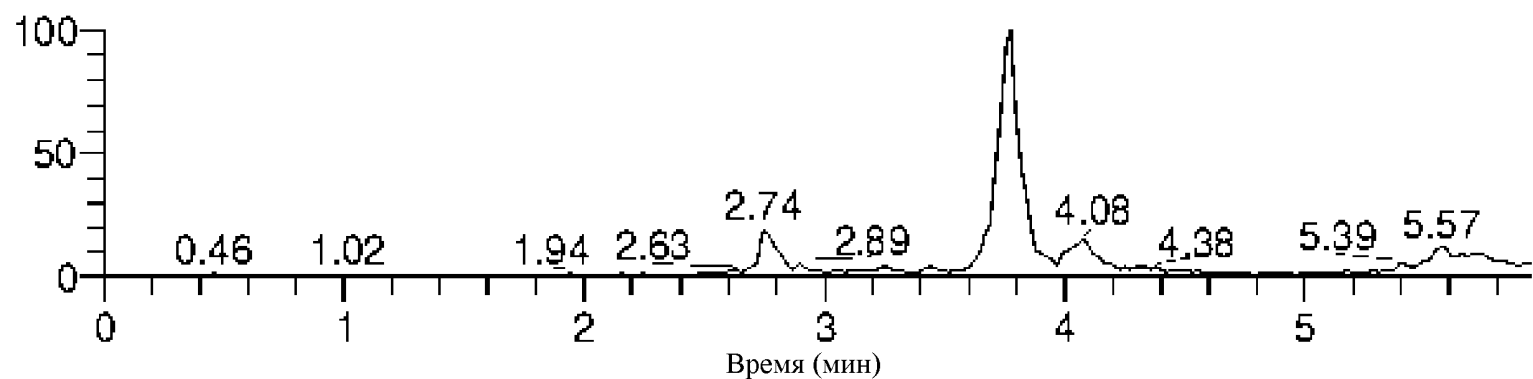
5'-псорален (С2-сшитый) -TAGCGGATGCAAAAAAAAAAGGCGAGCTTGCGTACTG-3'

PSO_HP_A9_TCT (MM 6087,1 Да)

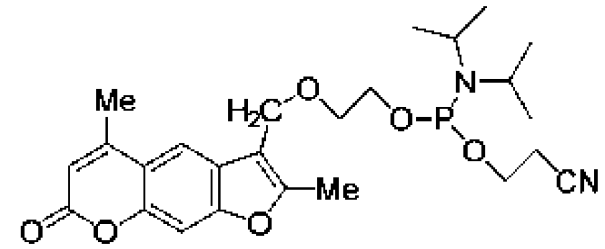
5'-AAAAAAAAAGCATCCGCTCT -3'

Фиг. 11

12/43



Фиг. 12



Псорален(С2)фосфорамидит

Tag1_PsoCVU (MM 3662,5 Да)

5'-AAAAACTCAGGT-3'

SplintC_PsoC2 (MM 4529,0 Да)

5'-TCCGCTCGACCTGAG-3'

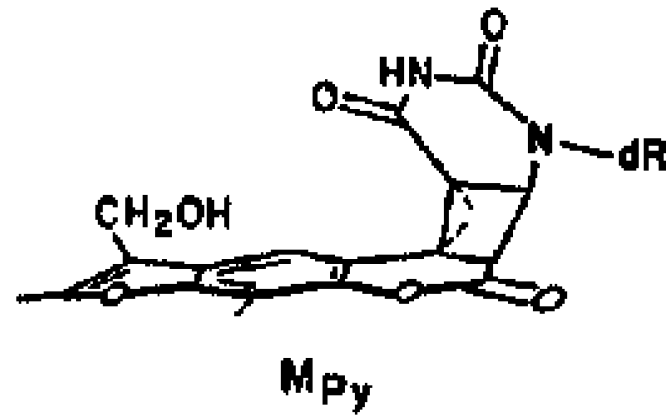
5PsoC2_A9_GA (MM 11580,6 Да)

5'-псорален (С2)-GAGCGGATGCAAAAAAAAAAGGCGAGCTTGCGTACTG-3'

Фиг. 13

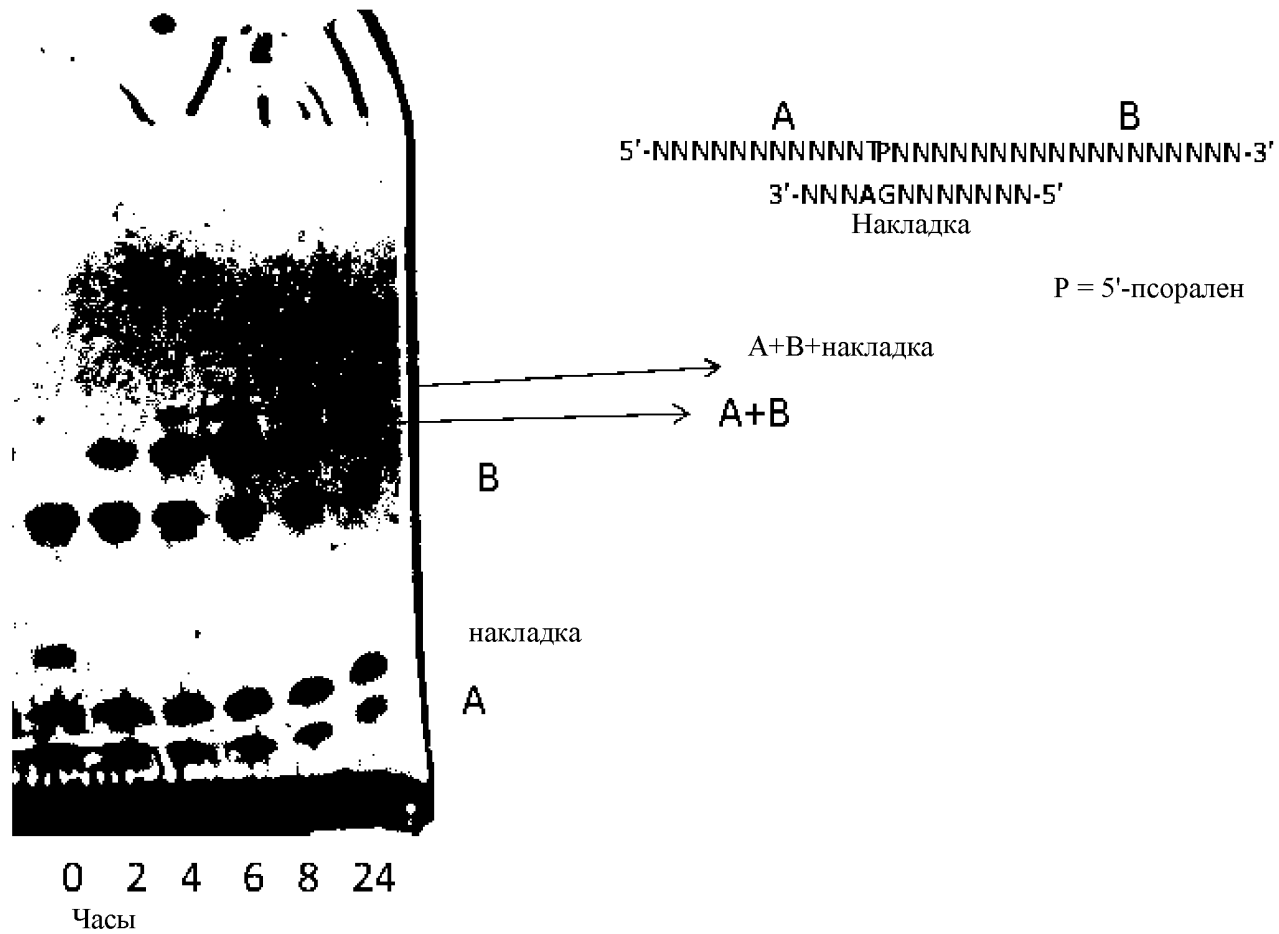


P = 5'-псорален



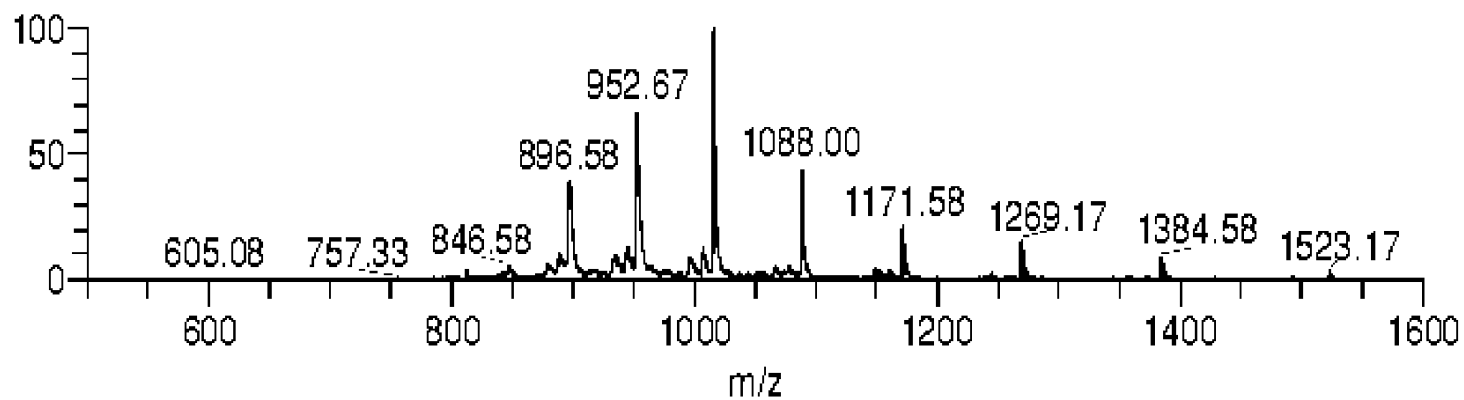
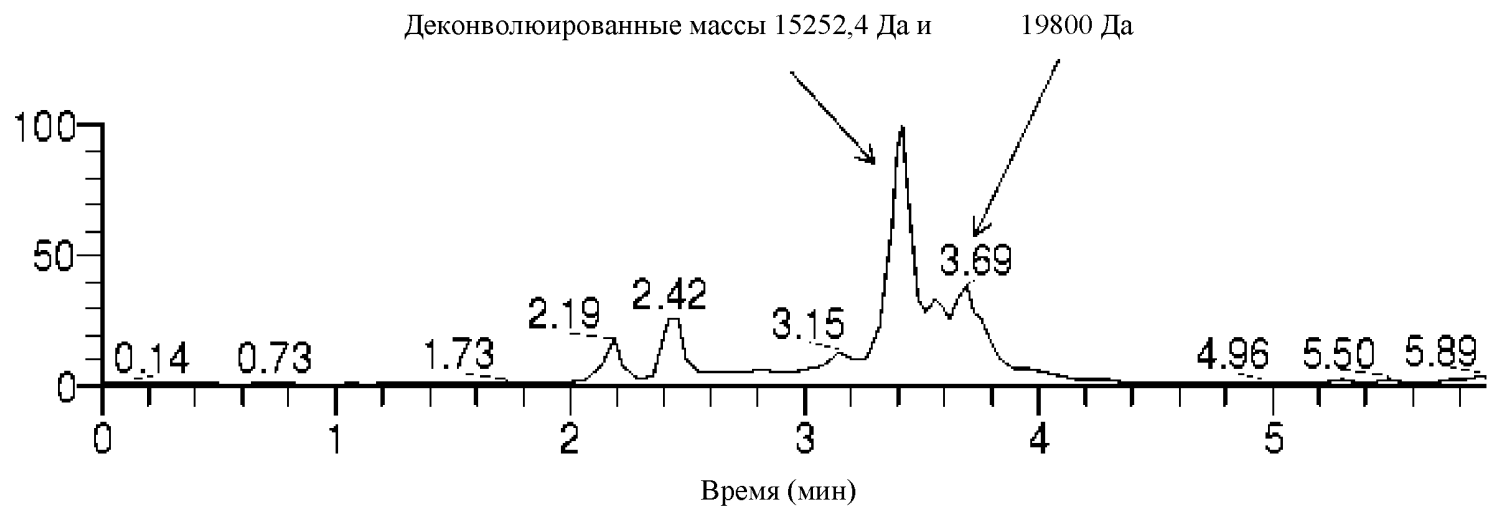
Фиг. 14

Зависимость от времени фотолигирования псоралена



Фиг. 15

16/43



Фиг. 16

5Bio_Tag_PsoCVU (MM 4232,1 Да)

5'-/5BioTEG/AAAAACTCAGGT-3'

/5BioTEG/ представляет собой
запатентованную модификацию биотина

5PsoC2_A9_GA (MW 11580,6 Да)

5'-псоларен (C2)-GAGCGGATGCAAAAAAAAAAGGCGAGCTTGCGTACTG-3'

SplintC_PsoC2 (MW 4529,0 Да)

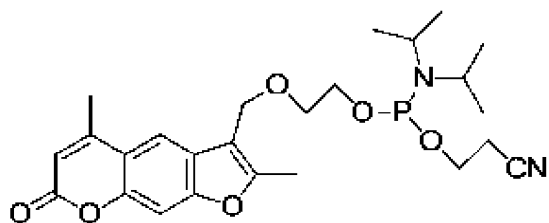
5'-TCCGCTCGACCTGAG-3'

FAMprimer (MM 5099,5 Да)

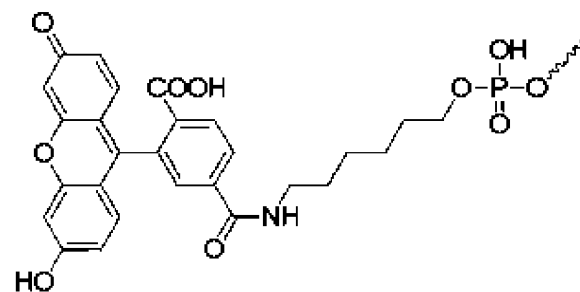
5'-/56-FAM/AGTACGCAAGCTCGC-3'

Phos-SplintC_PsoC2

5'-монофосфат-TCCGCTCGACCTGAG-3'



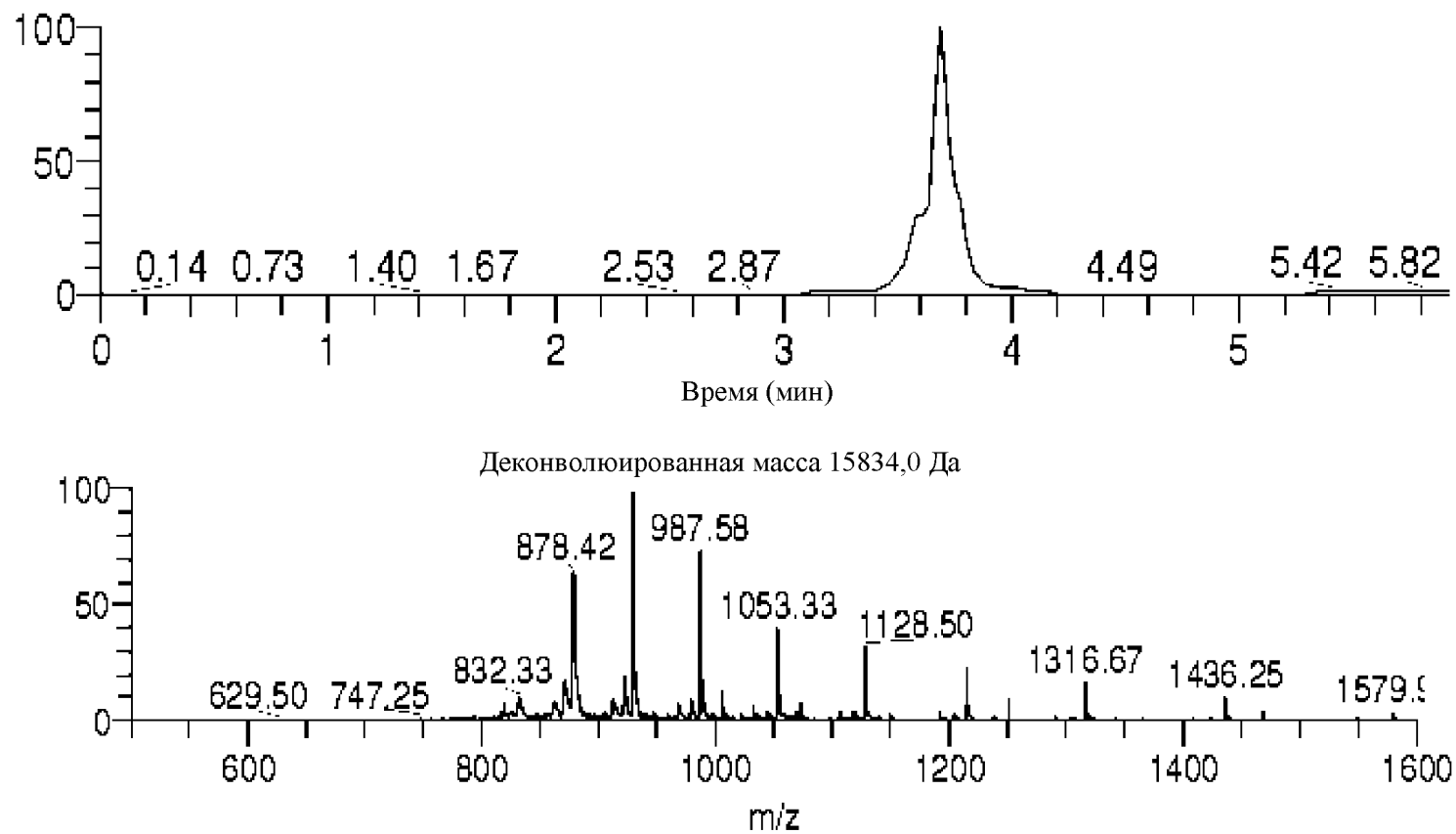
Псоларен(C2)фосфорамидит



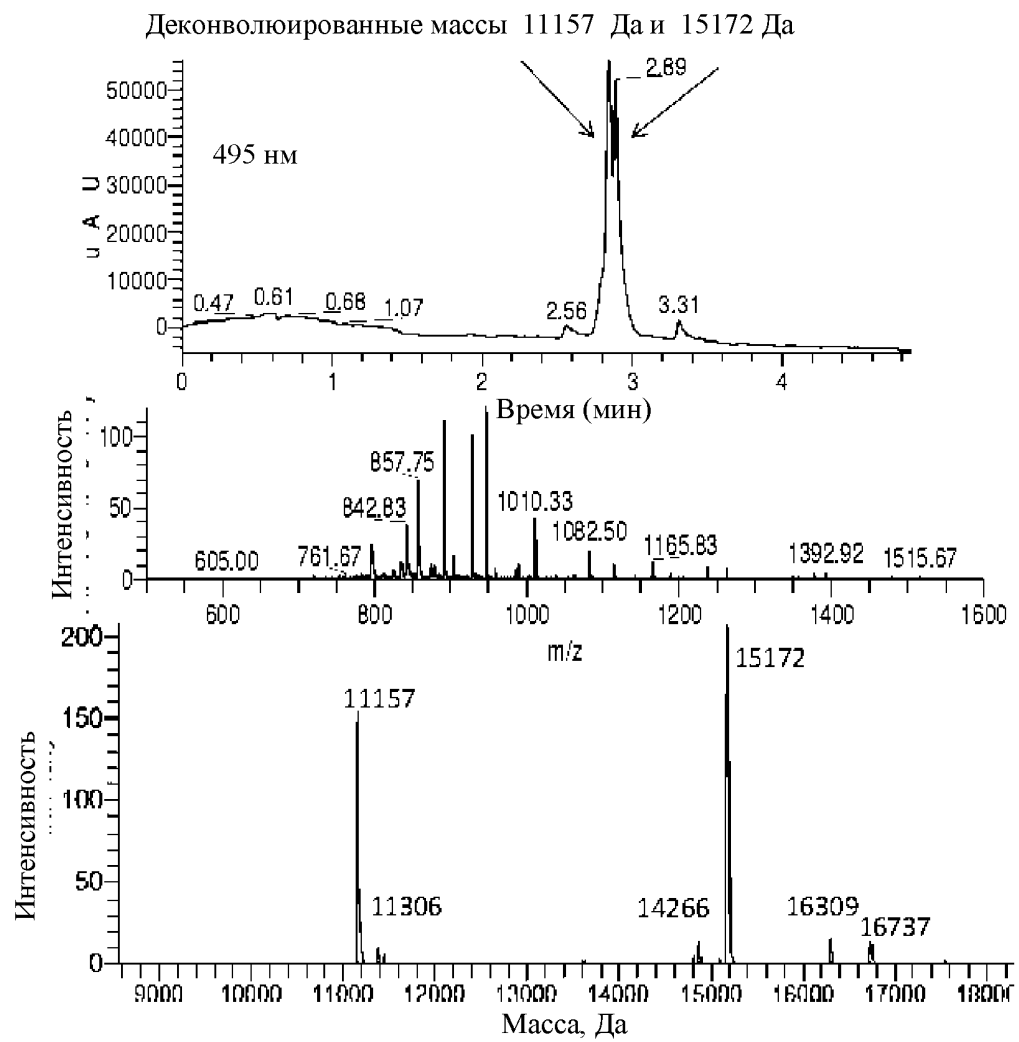
/56-FAM/ фосфорамидит

Фиг. 17

ЖХМС очищенного в геле фотохимического конъюгата Bio_Pso

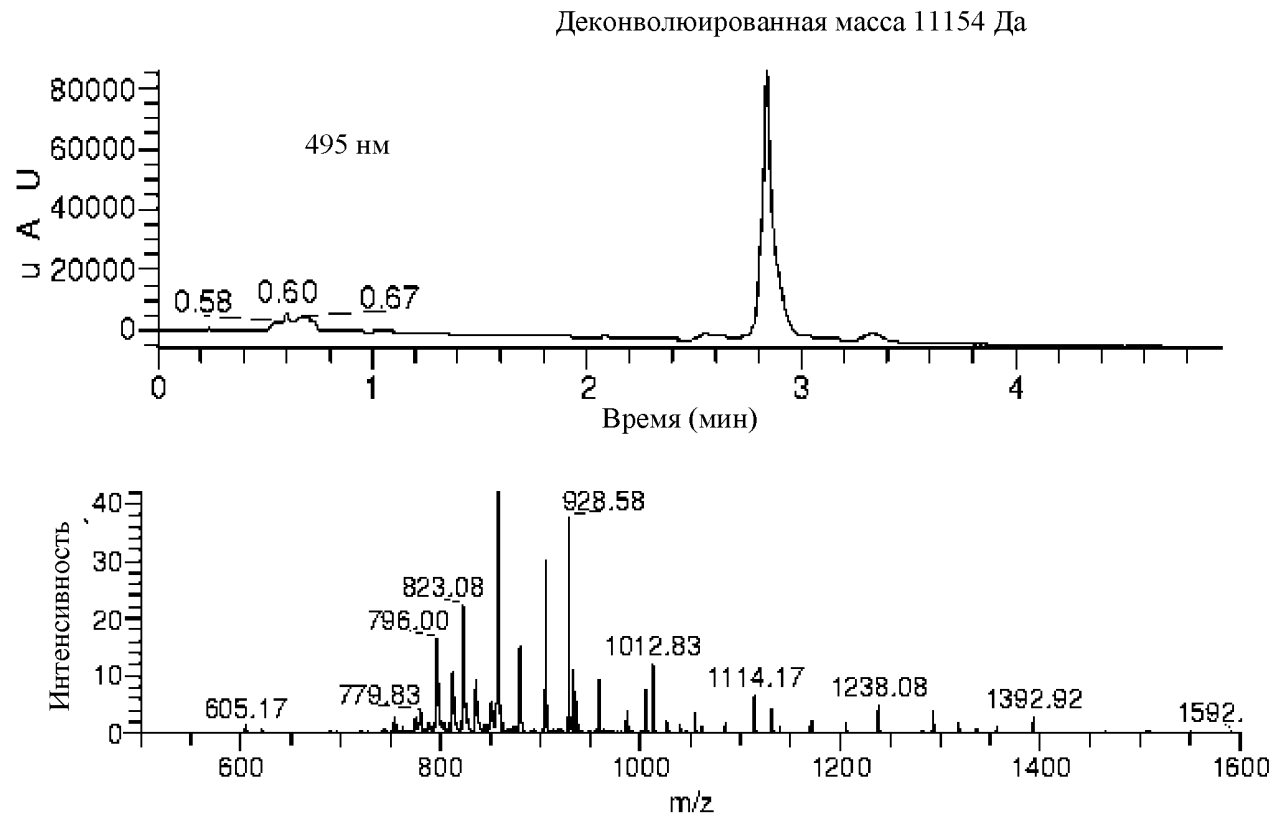


Фиг. 18

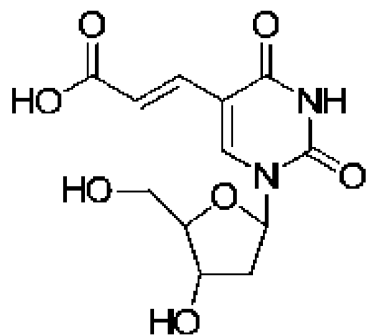


Фиг. 19

ЖХМС контрольной cDNA без лигирования, полученной от Bio_Pso
с использованием расширения терминального и нетерминального праймера



Фиг. 20



5-(карбоксивинил)-2'-дезоксинуридин

SplintC_CVU

5'-TCCGCTCAACCTGAG-3'

Tag1_PsoCVU (MW 3662,5 Да)

5'-AAAAACTCAGGT-3'

SplintA_CVU

5'-TCCGCTTAACCTGAG-3'

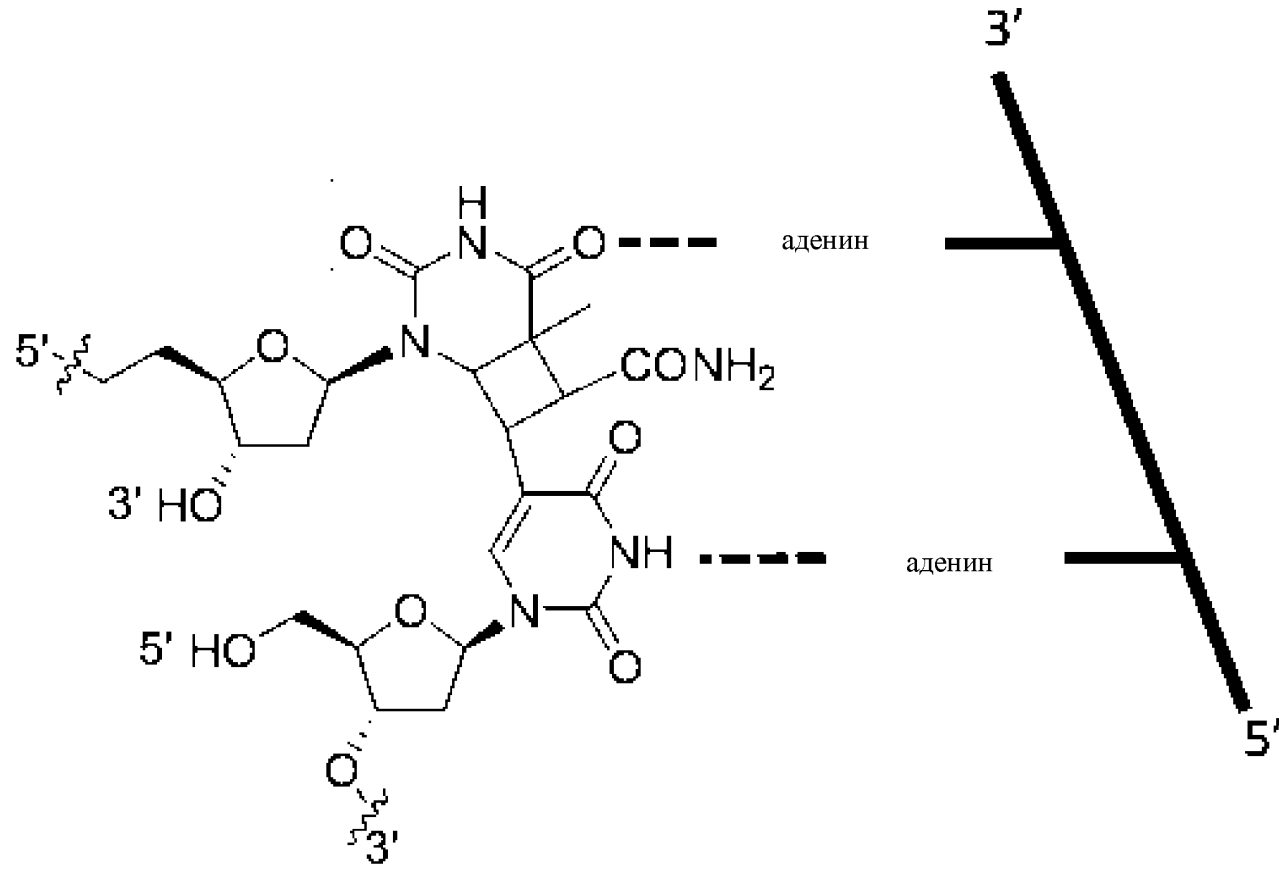
CVU_G (MW 11576,2 Да)

5'- (5-(карбоксивинил)-2'-дезоксинуридин)GAGCGGATGCAAAAAAAAAAAGGCGAGCTTGCGTACTG -3'

CVU_A (MW 11560,2 Да)

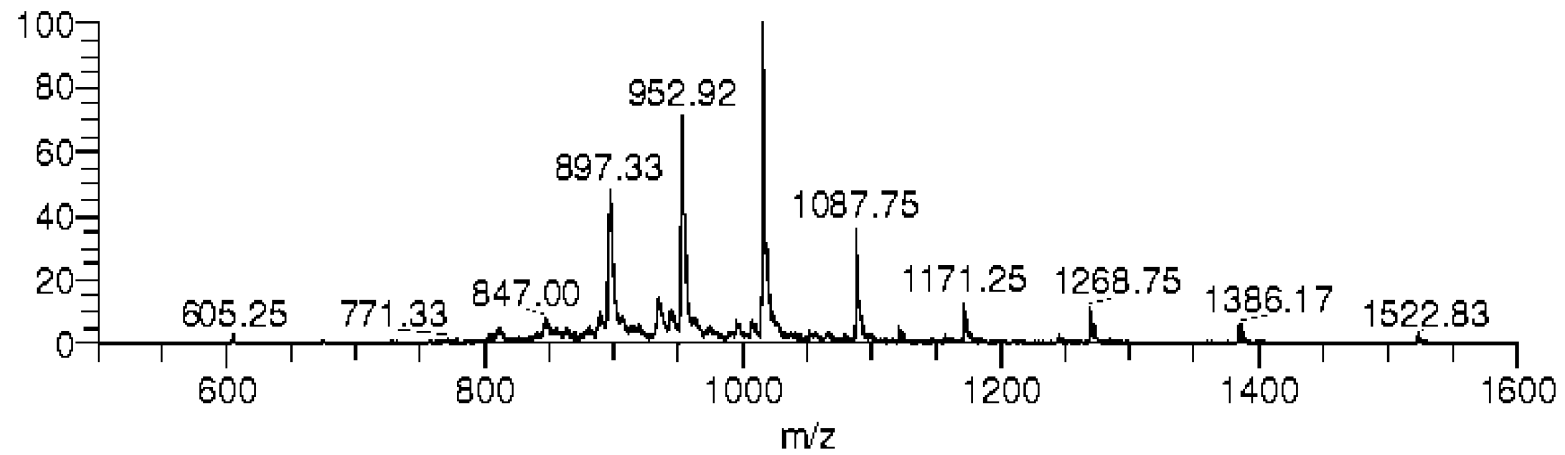
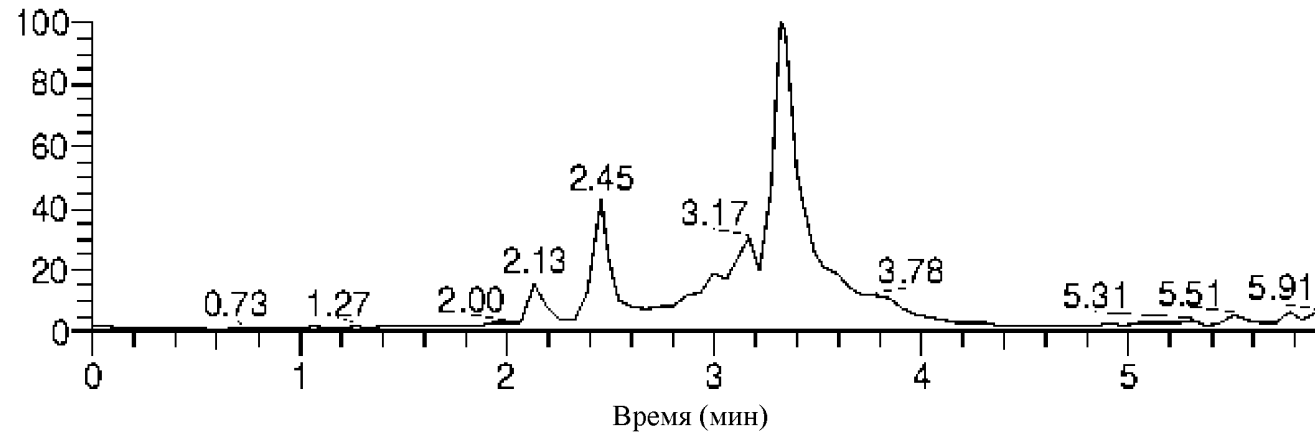
5'- (5-(карбоксивинил)-2'-дезоксинуридин)AAGCGGATGCAAAAAAAAAAAGGCGAGCTTGCGTACTG -3'

Фиг. 21



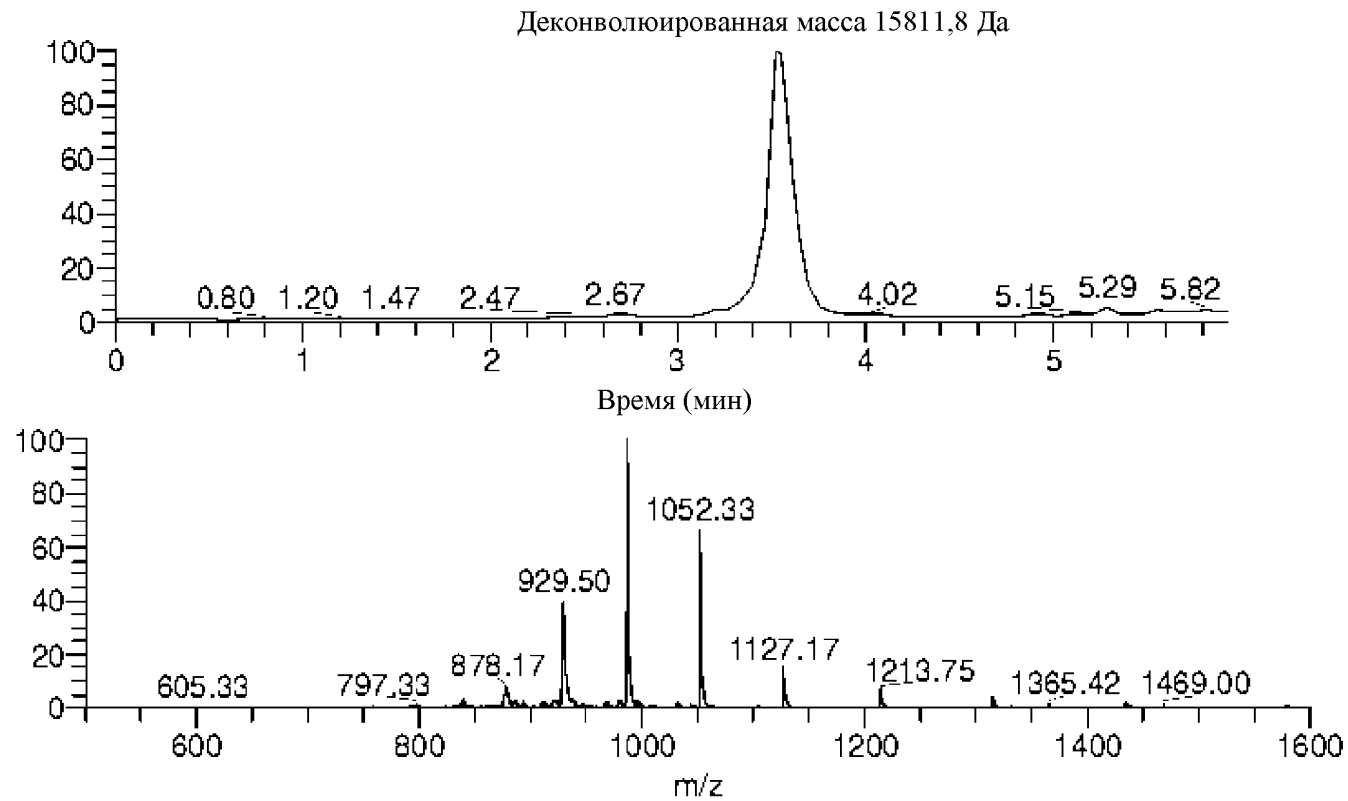
Фиг. 22

Деконволюированная масса 15260,6 Да



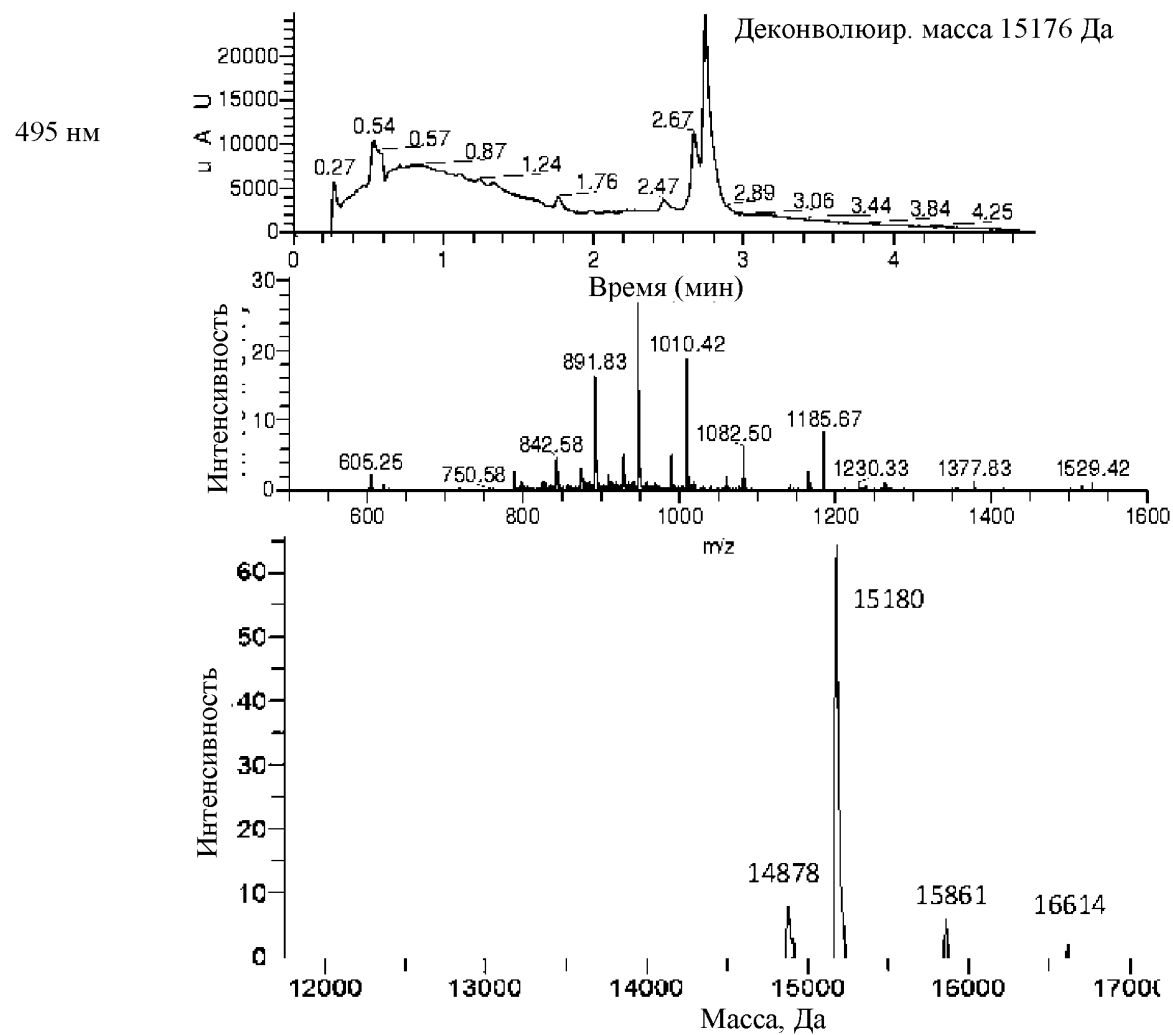
Фиг. 24

Иррадиация в течении 20 часов с последующей очисткой геля



5Bio_Tag_PsoCVU (MW 4232,1 Да)
/5BioTEG/AAAAA CTCAGGT

Фиг. 25



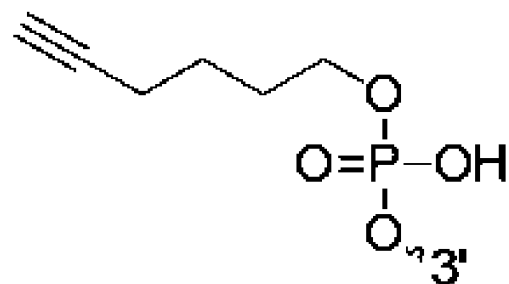
Фиг. 26

S_IDTN₃

TGCGGTCTAACTGTCTAG TTCACCTTCTCCGGAATGAACAGG/3азидN/

S_IDTalkyne

/5гексенил/TACCGAATTGCCGCCTGATTGAAGCAATCGGGTGCTATG

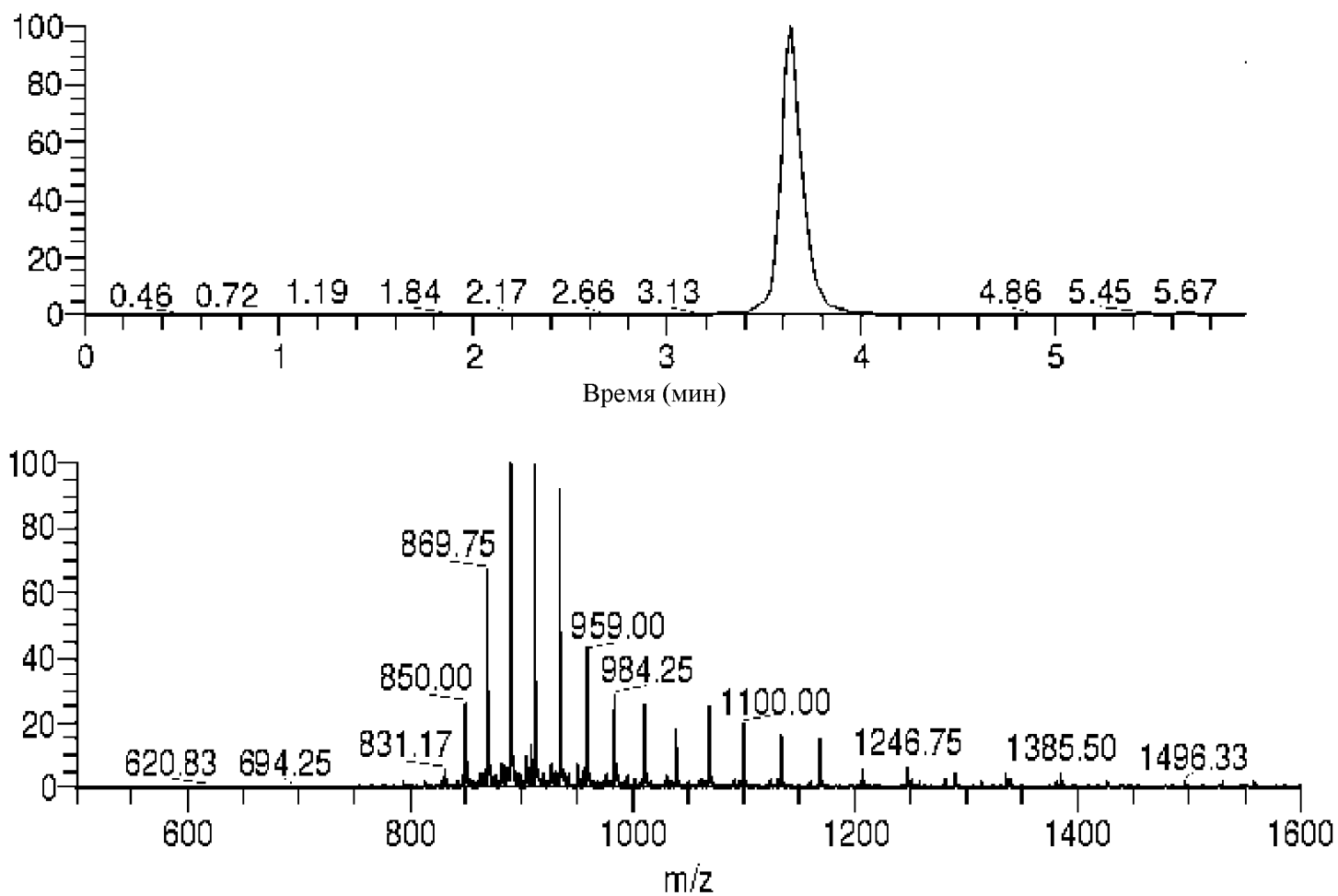


Модификация 5гексенила

Фиг. 27

Анализ ЖХ/МС очищенного в геле конъюгата S_IDTN₃ и S_IDTalkyne.

Деконволюированная наблюдаемая масса 26304,8 Да (рассчитанная 26311,2 Да)

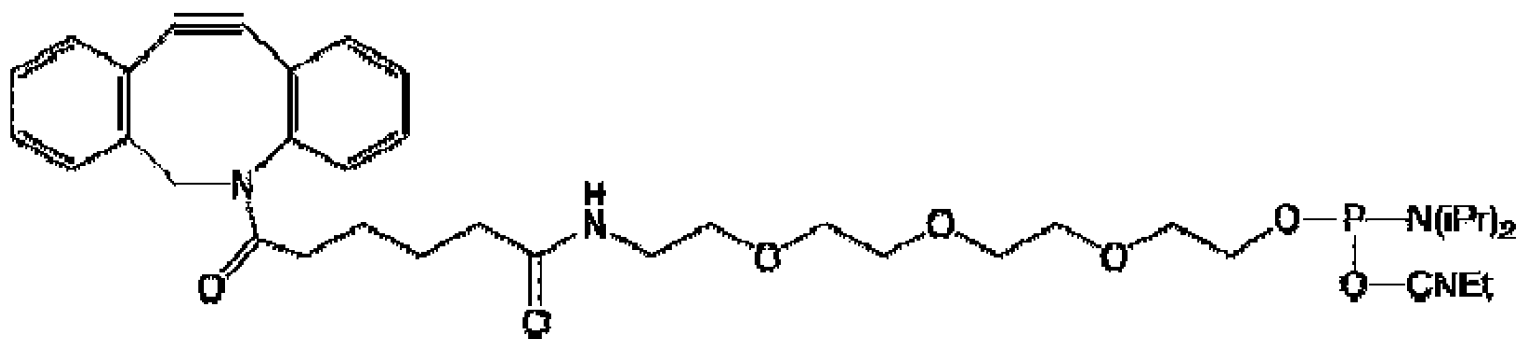


Фиг. 28

TKR_Central: /5Hexynyl/GGAATGAACAGGGTAAGCTGGAGTGAAGGCGTTATG/3азидN/

S_IDTN₃: TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAATGAACAGG/3азидN/

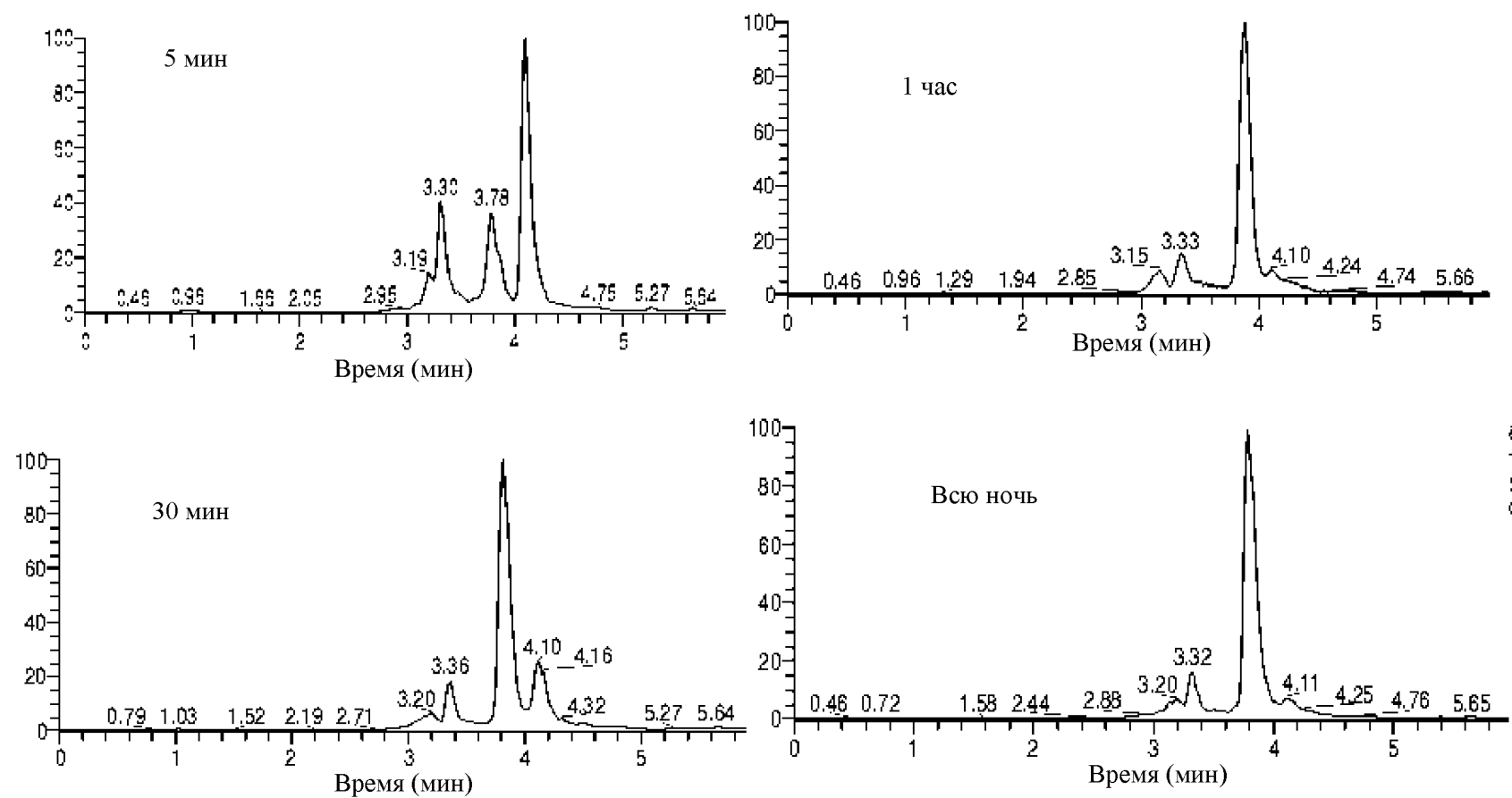
TKR_DBCO_S: /DBCO/GAAGGCGTTATGTCCGTA CTCTTGCAATCGGGTGCTATGCTT



DBCO фосфорамидит

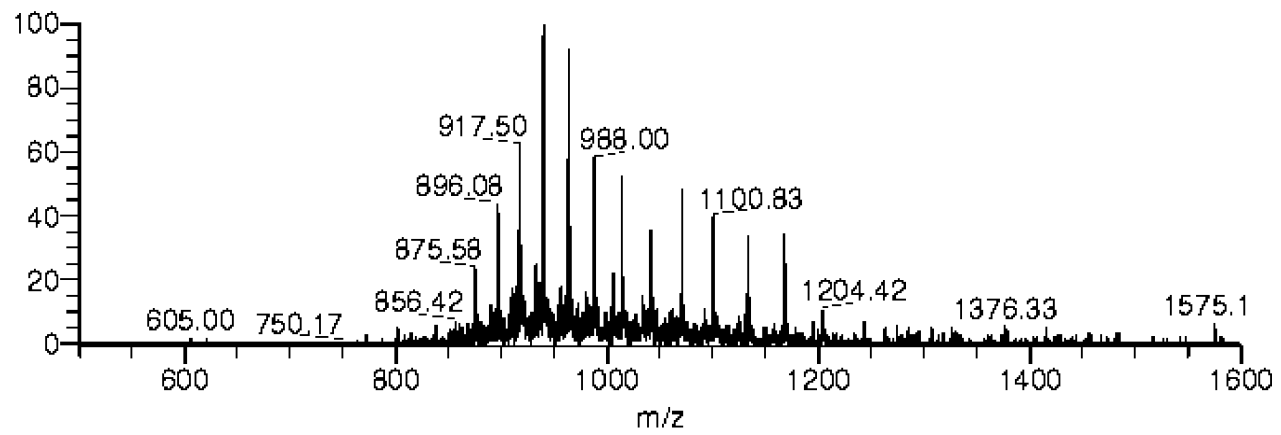
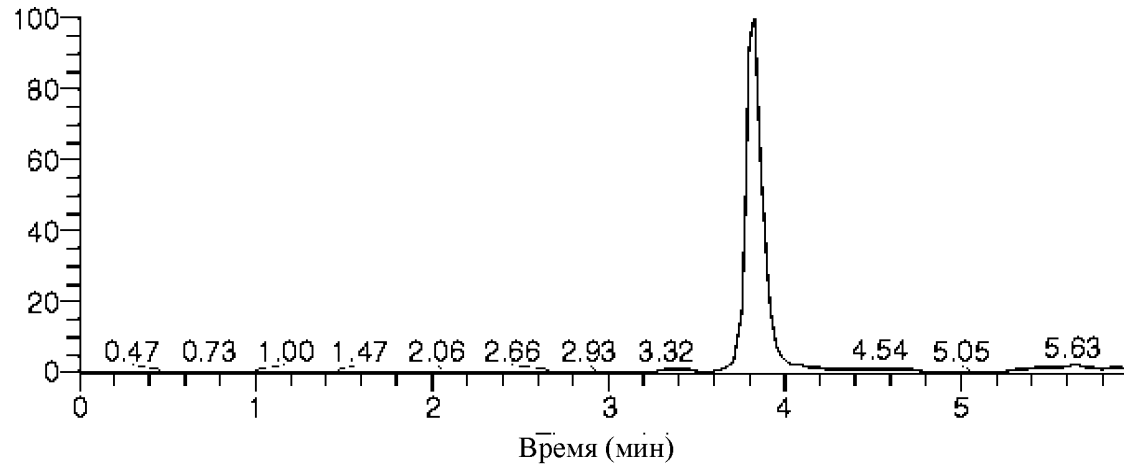
Фиг. 29

Наблюдаемая деконволюированная масса 25336 Да (рассчитанная 25341,8 Да)



Фиг. 30

Наблюдаемая деконволюированная масса 38578,1 Да (рассчитанная 38560,4 Да)



Фиг. 31

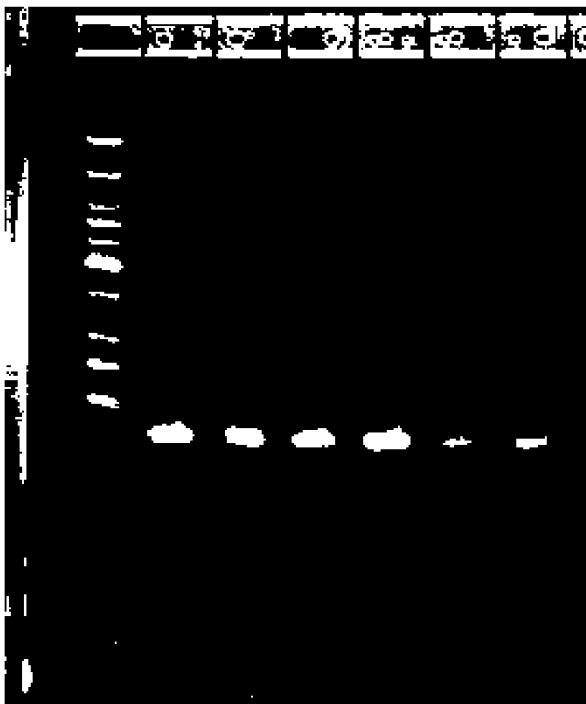
Conjugate_Click_S

TGCGGTCTAACTGTCTAGTTACCTTCTCCGGAATGAACAGG(click)TACCGAATTGCCGCCTGATTGAAG
-CAATCGGGTGCTATGCTT

Conjugate_Click_L

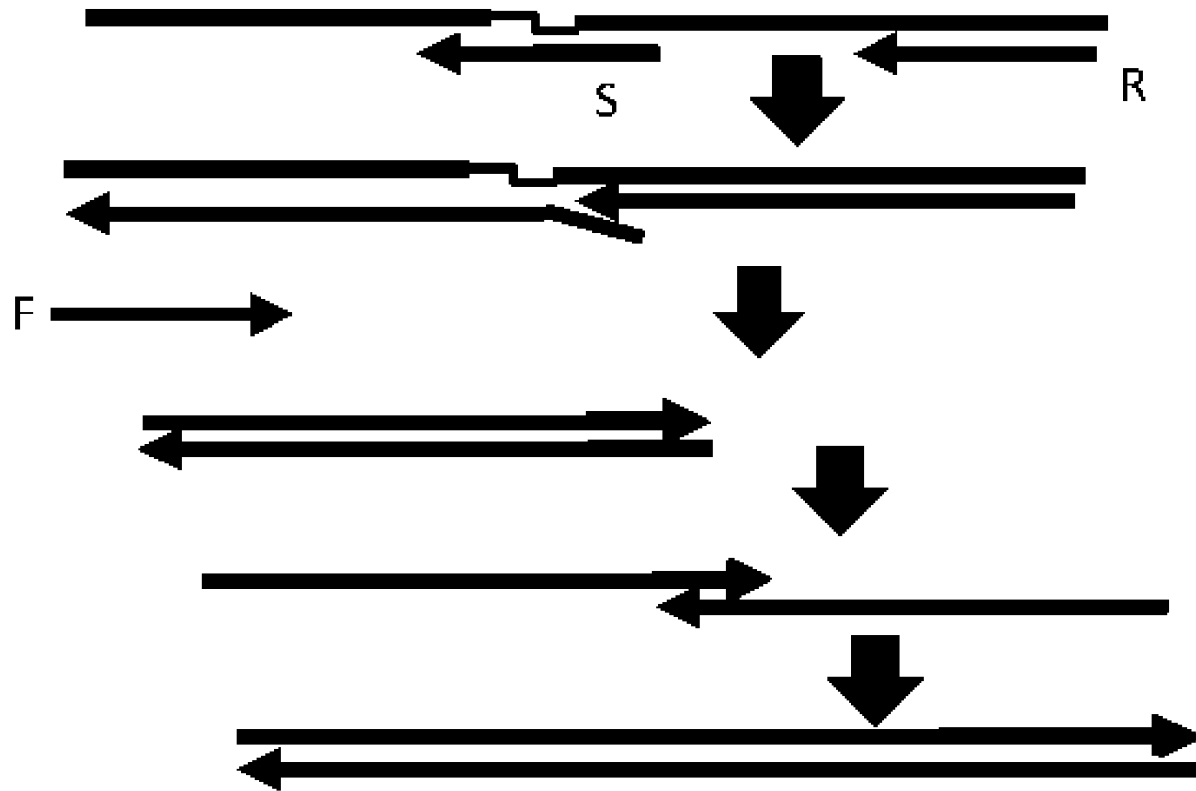
TGCGGTCTAACTGTCTAGGCACTTGTTCGTTTGCCAGTGTGAGGAATGAACAGG(click)TACCGAATTGC
CTTCCTCG-TACAGTTCTAAGGCGCTTGGACACCACTTCAATCGGGTGCTATGCTT

Фиг. 32

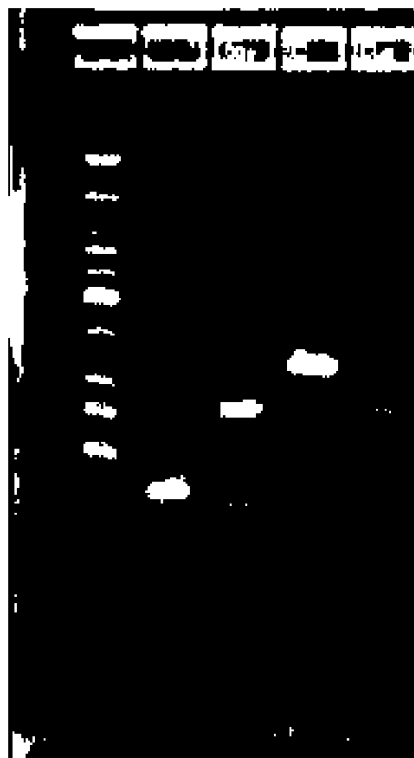


Полоса 1 2 3 4 5 6 7

Фиг. 33 ;

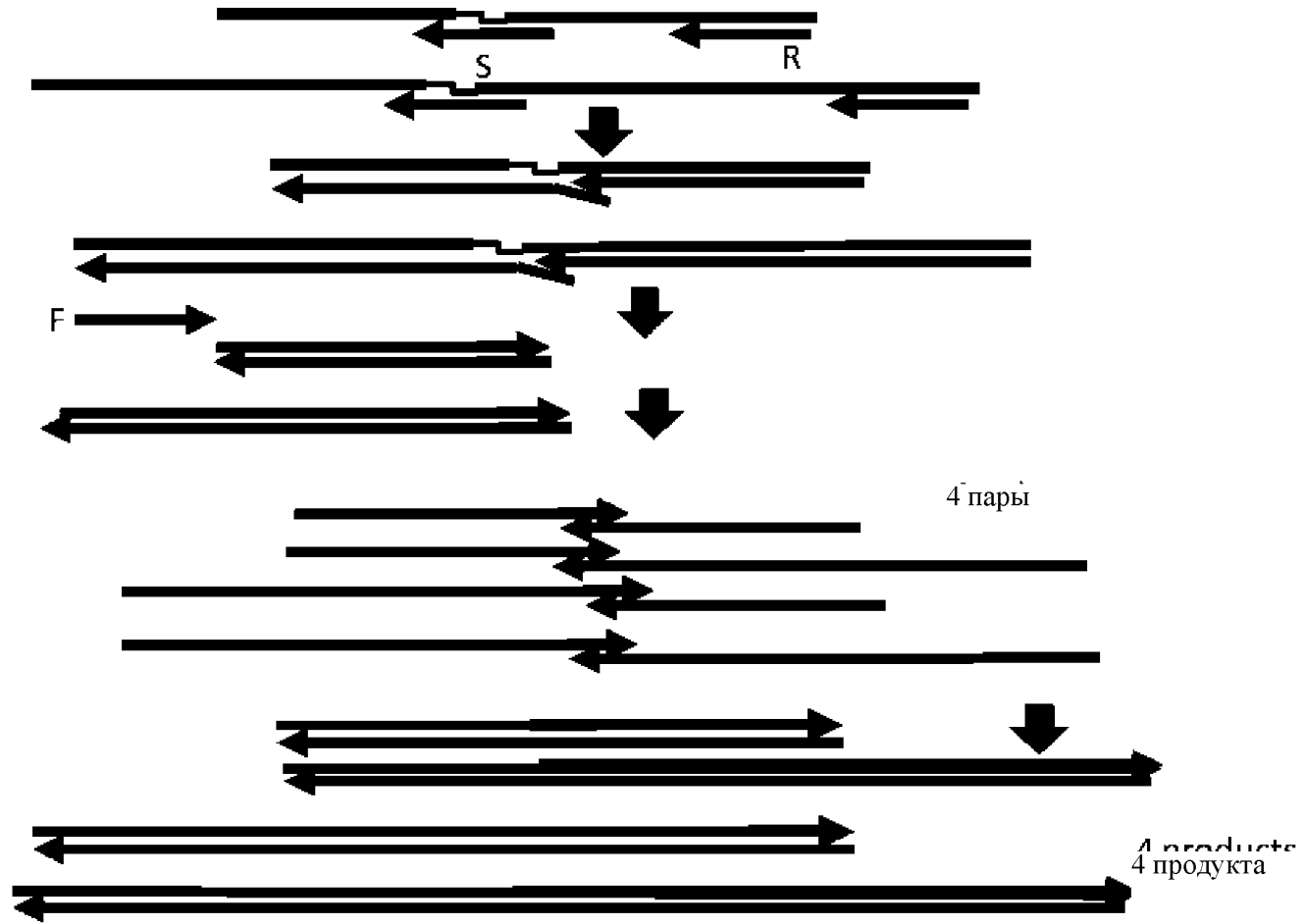


Фиг. 34



Полоса 1 2 3 4 5

Фиг. 35



Фиг. 36



Полоса 1 2 3 4

32 цикла, начальная концентрация 5 пМ

1-маркер

2- ПЦР в растворе

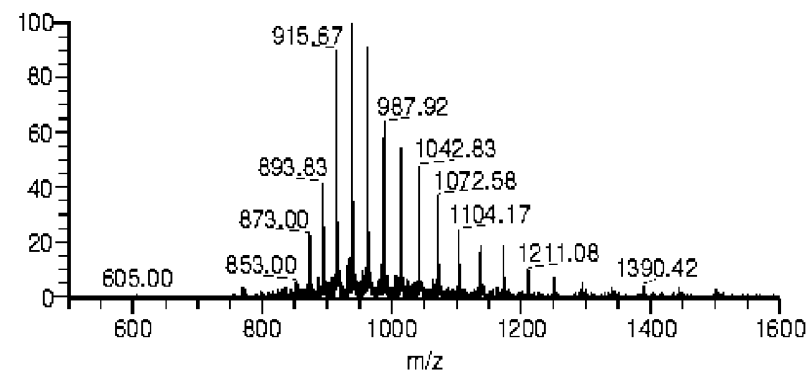
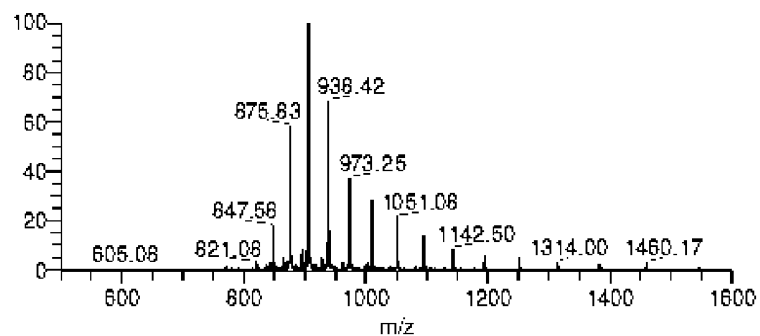
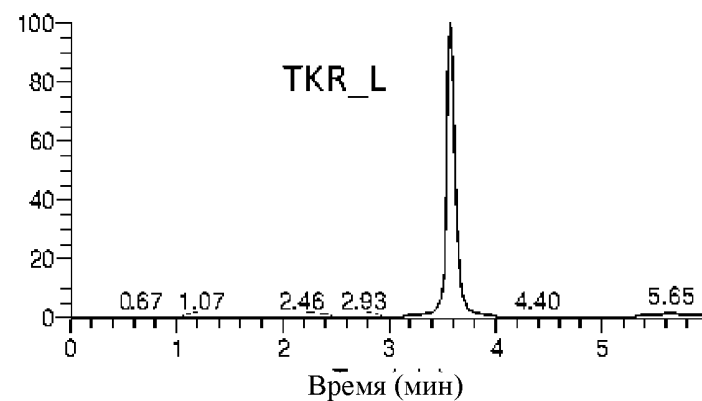
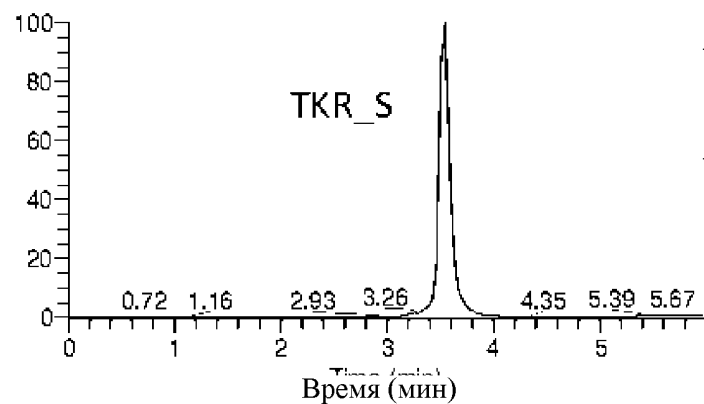
3-ПЦР в эмульсии

Фиг. 37

TKR_S:

TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAATGAACAGG(click)GGAATGAACAGGGCCTGATTGAA GCAATCGGGTGCTATGCTT

TKR_L:

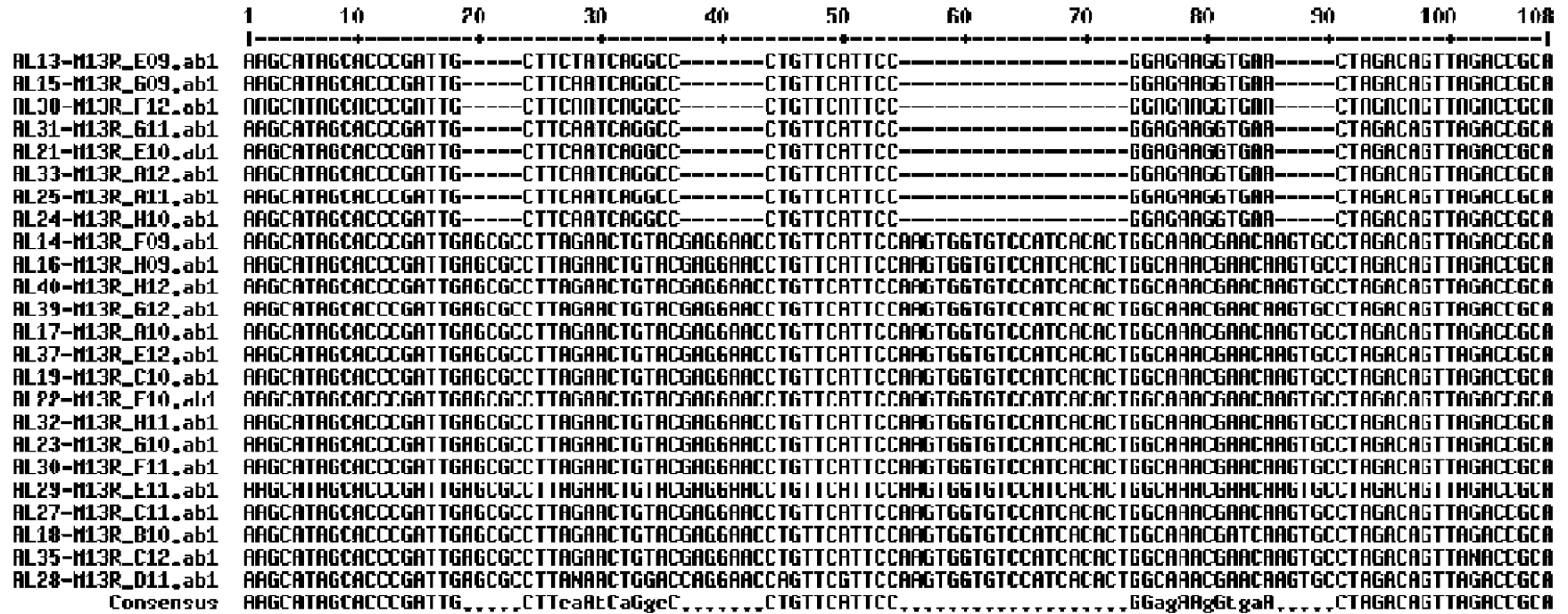
TGCGGTCTAACTGTCTAGGCACCTTGTTTCGTTTGCCAGTGTGATGGACACCACTTGGAATGAACAGG(click)GGAATGAACAGGGGAATGAACAG
GTTCTCGTACAGTTCCTAAGGCGCTCAATCGGGTGCTATGCTT

Фиг. 38



Полоса 1 2 3 4 5

Фиг. 39



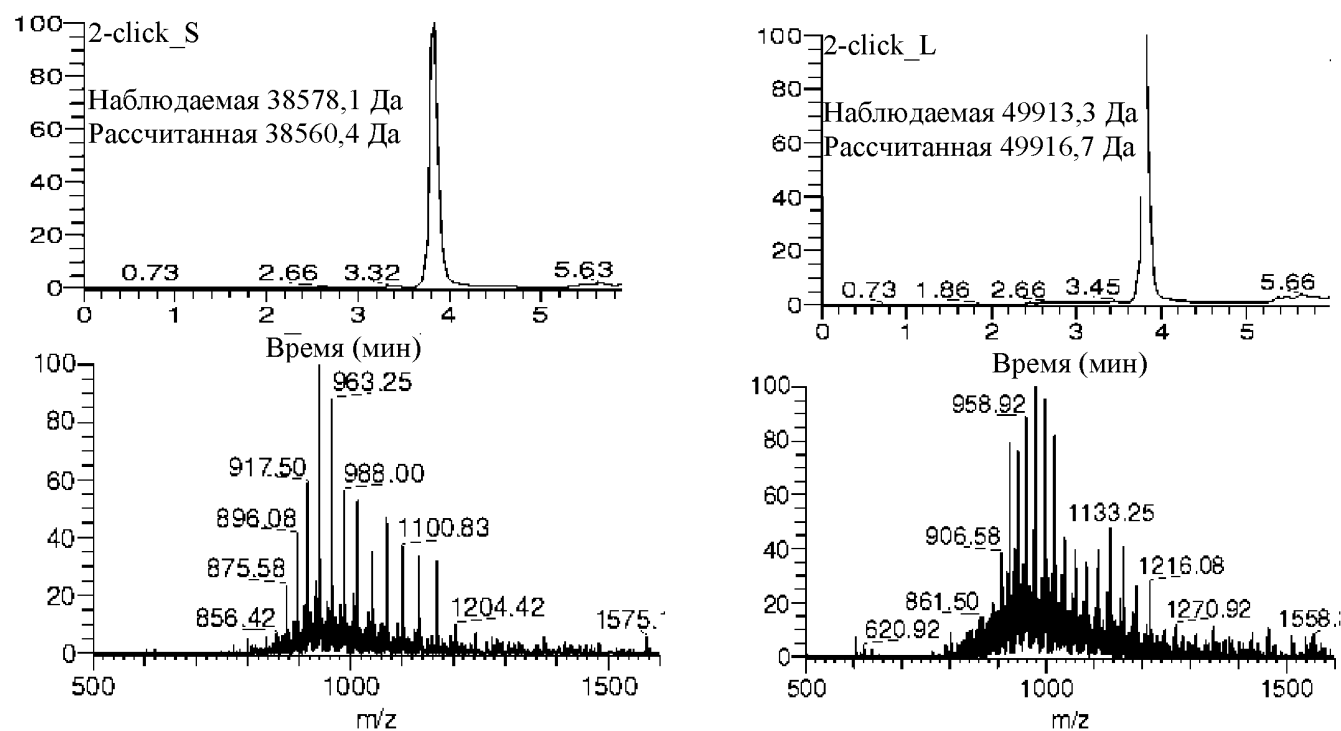
Фиг. 40

TKR 2 click S

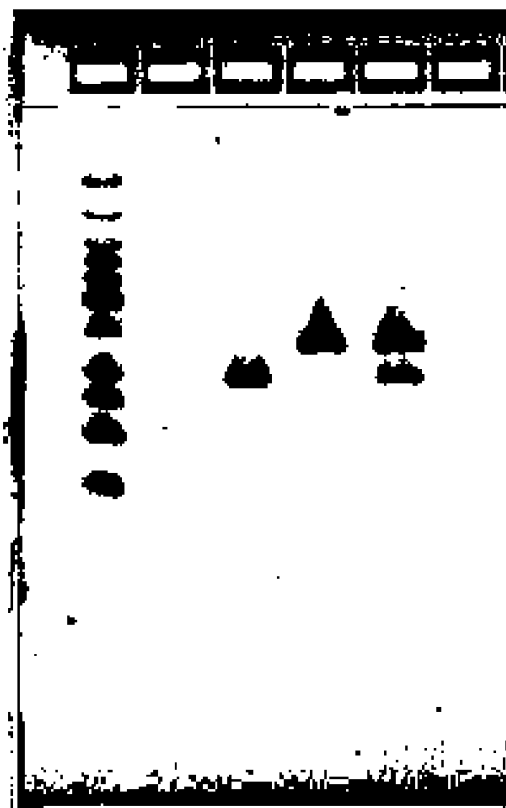
TGCGGTCTAACTGTCTAGTTCACCTTCTCCGGAATGAACAGG(CLICK)GGAATGAACAGGGTAAGCTGGAGTGAAGGCGTTATG(DBCO)
GAAGGCGTTATGTCCTACTCTTGCAATCGGGTGCTATGCTT

TKR 2 click L

TGCGGTCTAACTGTCTAGGCACTTGTTGTTGCCAGTGTGATGGACACCACTTGAATGAACAGG(CLICK)GGAATGAACAGGGTAAGC
TGGAGTGAAGGCGTTATG(DBCO)GAAGGCGTTATGTGATCCGTGGTGTGAGTTCGAATCGGGTGCTATGCTT



Фиг. 41



Полоса 1 2 3 4 5

Фиг. 42

