

(19)日本国特許庁(JP)

(12)公表特許公報(A)

(11)公表番号

特表2023-538074

(P2023-538074A)

(43)公表日 令和5年9月6日(2023.9.6)

(51)国際特許分類	F I	テーマコード(参考)
C 1 2 Q 1/6844(2018.01)	C 1 2 Q 1/6844	Z Z N A 4 B 0 6 3
C 1 2 Q 1/6869(2018.01)	C 1 2 Q 1/6869	Z
C 1 2 N 9/10 (2006.01)	C 1 2 N 9/10	

審査請求 未請求 予備審査請求 未請求 (全76頁)

(21)出願番号	特願2023-511925(P2023-511925)	(71)出願人	518386818 エンコディア, インコーポレイテッド アメリカ合衆国 カリフォルニア 9 2 1 2 1, サンディエゴ, オーバーリン ドライブ 5 7 8 5
(86)(22)出願日	令和3年8月16日(2021.8.16)	(74)代理人	100078282 弁理士 山本 秀策
(85)翻訳文提出日	令和5年2月16日(2023.2.16)	(74)代理人	100113413 弁理士 森下 夏樹
(86)国際出願番号	PCT/US2021/046164	(74)代理人	100181674 弁理士 飯田 貴敏
(87)国際公開番号	WO2022/040098	(74)代理人	100181641 弁理士 石川 大輔
(87)国際公開日	令和4年2月24日(2022.2.24)	(74)代理人	230113332 弁護士 山本 健策
(31)優先権主張番号	63/067,744		
(32)優先日	令和2年8月19日(2020.8.19)		
(33)優先権主張国・地域又は機関	米国(US)		
(81)指定国・地域	AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA ,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA( AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,A T,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR ,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC, 最終頁に続く		最終頁に続く

(54)【発明の名称】 連続エンコーディング方法及び関連キット

(57)【要約】

本開示は、高分子を分析するための方法及びキットに関する。いくつかの実施形態において、本開示は、分子認識事象のバーコード化及び核酸エンコーディングを用いる高分子分析方法に関する。また、本明細書で提供されるのは、分析のための高分子と関連付けられた核酸分子とのライゲーション、伸長、及び切断反応を実行することを含む、複数の酵素を使用して情報を転送するための方法及び関連キットである。いくつかの実施形態において、分析のための高分子は、ペプチド、ポリペプチド、又はタンパク質を含む。

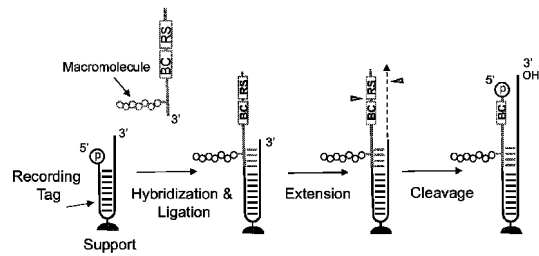


FIG. 1A

## 【特許請求の範囲】

## 【請求項 1】

高分子を分析するための方法であって、

- ( a ) 高分子及び支持体に接合された関連する記録タグを提供するステップと、
  - ( b ) 前記高分子を、前記高分子に結合することができる結合剤であって、前記結合剤に関する識別情報を有するコーディングタグを含む結合剤と接触させて、前記高分子と前記結合剤との間の結合を可能にするステップと、
  - ( c ) 核酸接合試薬によって、前記記録タグの 5 ' 末端を前記コーディングタグの 3 ' 末端に接合するステップと、
  - ( d ) ポリメラーゼによって、前記コーディングタグを鋳型として使用して前記記録タグを伸長し、二本鎖伸長記録タグを生成するステップと、
  - ( e ) 前記二本鎖伸長記録タグを二本鎖核酸切断試薬で切断して、前記伸長記録タグ内に 3 ' オーバーハングを生成するステップと、
- それによって、情報が、前記コーディングタグから前記記録タグに転送されて、前記伸長記録タグを生成するステップと、を含む、方法。

## 【請求項 2】

ステップ ( d ) において、前記二本鎖伸長記録タグが、前記二本鎖核酸切断試薬によって認識することができる認識配列を含む、請求項 1 に記載の方法。

## 【請求項 3】

ステップ ( e ) における前記切断が、前記高分子から前記結合剤を放出する、請求項 1 に記載の方法。

## 【請求項 4】

ステップ ( b )、( c )、( d )、及び ( e ) が、周期的な方法で連続 1 回以上繰り返される、請求項 1 に記載の方法。

## 【請求項 5】

前記高分子が、ポリペプチドである、請求項 1 に記載の方法。

## 【請求項 6】

前記ポリペプチドが、生物学的サンプルからタンパク質を断片化することによって得られる、請求項 5 に記載の方法。

## 【請求項 7】

分析のための前記高分子が、核酸ではない、請求項 1 に記載の方法。

## 【請求項 8】

ステップ ( b ) を繰り返す前に、前記高分子の一部を除去することを更に含む、請求項 4 に記載の方法。

## 【請求項 9】

ステップ ( b ) を繰り返す前に、前記ポリペプチドの N 末端アミノ酸 ( N T A A ) を除去して、前記ポリペプチドの新しい N T A A を露出させることを更に含む、請求項 5 に記載の方法。

## 【請求項 10】

ステップ ( e ) において、前記二本鎖核酸切断試薬によって生成された前記伸長記録タグの前記 3 ' オーバーハングが、ステップ ( b ) が繰り返されるときに、第 2 のコーディングタグとハイブリダイズするために利用可能である、請求項 4 に記載の方法。

## 【請求項 11】

前記方法が、ステップ ( b ) において、複数の高分子を、単一の結合剤又は複数の結合剤と接触させることを含む、請求項 1 に記載の方法。

## 【請求項 12】

ステップ ( b ) の前に、前記ポリペプチドの末端アミノ酸を修飾するための試薬で前記ポリペプチドを処理することを更に含む、請求項 5 に記載の方法。

## 【請求項 13】

前記高分子と関連付けられた前記記録タグが、核酸ヘアピンを含む、請求項 1 に記載の

方法。

【請求項 14】

ステップ(c)の前記核酸接合試薬、ステップ(d)の前記ポリメラーゼ、及びステップ(e)の前記二本鎖核酸切断試薬が同時に提供される、請求項1に記載の方法。

【請求項 15】

前記二本鎖核酸切断試薬が、IIS型制限酵素である、請求項1に記載の方法。

【請求項 16】

前記伸長記録タグのうちの1つ以上を分析することを更に含み、前記伸長記録タグを分析することが、核酸シーケンシング方法を含む、請求項1に記載の方法。

【請求項 17】

高分子分析のためのキットであって、  
 コーディングタグを含む結合剤であって、前記コーディングタグが、前記結合剤に関する識別情報を含む、結合剤と、  
 核酸接合試薬と、  
 ポリメラーゼと、  
 二本鎖核酸切断試薬と、を含み、  
 前記結合剤が、記録タグと関連付けられた高分子に結合するように構成され、前記コーディングタグからの前記識別情報が、前記コーディングタグから前記高分子と関連付けられた前記記録タグへの転送のために構成されている、キット。

10

【請求項 18】

前記高分子が、ポリペプチドである、請求項17に記載のキット。

20

【請求項 19】

前記ポリペプチドのN末端アミノ酸(NTAA)を除去するための試薬を更に含む、請求項18に記載のキット。

【請求項 20】

前記キットが、前記核酸接合試薬、前記ポリメラーゼ、及び前記二本鎖核酸切断試薬を含む混合物を含む、請求項17に記載のキット。

【請求項 21】

前記高分子及び/又は前記記録タグを固定化するための支持体を更に含む、請求項17に記載のキット。

30

【発明の詳細な説明】

【技術分野】

【0001】

関連出願

本出願は、2020年8月19日に提出された米国仮特許出願第63/067,744号の優先権を主張し、その開示は、あらゆる目的のためにその全体が参照により本明細書に組み込まれる。

【0002】

ASCIIテキストでの配列表

この特許又は出願ファイルには、コンピューターで読み取り可能なASCIIテキスト形式で提出された配列表が含まれている(ファイル名: 4614-2002540\_\_SeqList\_\_ST25.txt、記録日: 2021年7月27日、サイズ: 2,469バイト)。配列表ファイルの内容は、その全体が参照により本明細書に組み込まれる。

40

【0003】

本開示は、高分子を分析するための方法及びキットに関する。いくつかの実施形態において、本開示は、分子認識事象のバーコード化及び核酸エンコーディングを用いる高分子分析方法に関する。また、本明細書で提供されるのは、分析のための高分子と関連付けられた核酸分子とのライゲーション、伸長、及び切断反応を実行することを含む、複数の酵素を使用して情報を転送するための方法及び関連キットである。いくつかの実施形態において、分析のための高分子は、ペプチド、ポリペプチド、又はタンパク質を含む。

50

## 【背景技術】

## 【0004】

タンパク質などの高分子の高度に並列な特徴付け及び認識は、依然として課題である。プロテオミクスでは、1つの目標は、サンプル中の多数のタンパク質を識別しかつ定量化することであり、これは高スループットの方法で達成するのが難しいタスクである。イムノアッセイ及び質量分析に基づく方法などのアッセイが使用されているが、サンプル及び分析物レベルの両方に限定され、感度及びダイナミックレンジが制限され、交差反応性及びバックグラウンドシグナルに関する潜在的な問題がある。例えば、検出可能な標識を有する親和性剤を使用して、同族高分子の集合に対する親和性剤の集合の読み出しを多重化することは、依然として困難である。高分子分析に関連する改良された技法とともに、タンパク質シーケンシング及び/又は分析、並びにそれを達成するための製品、方法、及びキットへの適用の必要性が残っている。効率的で、高度に並列化され、正確で、感度が高く、高スループットである高分子分析を実行するためのプロテオミクス技術が必要である。本開示は、これら及び他の関連する必要性を満たす。

10

## 【0005】

本発明のこれら及び他の態様は、以下の発明を実施するための形態を参照することで明らかになるであろう。この目的のために、特定の背景情報、手順、化合物、及び/又は組成物をより詳細に説明する様々な参考文献が本明細書に記載されており、各々が、参照によりそれら全体が本明細書に組み込まれる。

20

## 【発明の概要】

## 【課題を解決するための手段】

## 【0006】

本発明の概要は、特許請求される主題の範囲を制限するために使用されることを意図しない。特許請求される主題の他の特徴、詳細、有用性、及び利点は、添付の図面及び添付の特許請求の範囲に開示された態様を含む、発明を実施するための形態から明らかになるであろう。

## 【0007】

本明細書で提供されるのは、高分子及び支持体に接合された関連する記録タグを提供するステップと、高分子を、高分子に結合することができる結合剤であって、結合剤に関する識別情報を有するコーディングタグを含む結合剤と接触させて、高分子と結合剤との間の結合を可能にするステップと、核酸接合試薬を使用して記録タグの5'末端をコーディングタグの3'末端に接合するステップと、ポリメラーゼによって、コーディングタグを鋳型として使用して記録タグを伸長し、二本鎖伸長記録タグを生成するステップと、二本鎖伸長記録タグを二本鎖核酸切断試薬で切断して、伸長記録タグ内に3'オーバーハングを生成するステップと、それによって、情報が、コーディングタグから記録タグに転送されて、伸長記録タグを生成するステップと、を含む高分子を分析するための方法である。

30

## 【0008】

本明細書で提供されるのは、コーディングタグを含む結合剤であって、コーディングタグが、結合剤に関する識別情報を含む、結合剤と、核酸接合試薬と、ポリメラーゼと、二本鎖核酸切断試薬と、を含み、結合剤が、記録タグと関連付けられた高分子に結合するように構成され、コーディングタグからの識別情報が、コーディングタグから高分子と関連付けられた記録タグへの転送のために構成されている、キットである。いくつかの実施形態において、キットは、複数の結合剤を含む。いくつかの実施形態において、核酸接合試薬、ポリメラーゼ及び二本鎖核酸切断試薬は、混合物として提供される。

40

## 【図面の簡単な説明】

## 【0009】

添付の図面を参照して、本発明の非限定的な実施形態を例として説明する。図面は模式的であり、正確な縮尺は意図されない。例示目的上、全ての構成要素が全ての図面において標識されるとは限らず、また、当業者が本発明を理解することを可能にするために例示が必要とされない場合は、本発明の各実施形態の全ての構成要素が示されるとも限らない

50

## 【 0 0 1 0 】

【図 1 A】本明細書で提供される方法を使用した情報転送を含む例示的な高分子分析アッセイを示す。(図 1 A)において、分析される高分子(例えば、ペプチド)は、ライゲーション反応を使用して支持体上に固定化された記録タグヘアピンに接合され、構造は、後続のステップのために切断される。図 1 A の第 1 の(最左)パネルは、凹部 5'リン酸化末端を有する捕捉核酸ヘアピンを示す。図 1 A の第 2 のパネルは、(反応性カップリング部分を介して)支持体上に固定化された捕捉核酸ヘアピンにハイブリダイズする bait 核酸に、分析されるペプチドが付着していることを示す。bait 核酸を捕捉核酸にライゲーションする。記録タグをヘアピンにライゲーションする。ブロック標識バーコード(BC)は、高分子に付着し、記録タグに組み込むことができるいずれかの任意のバーコード、例えば、サンプル特異的バーコード及び/又は UMI を示す。ブロック標識制限酵素認識部位(RS)部位は、IIS 型制限酵素が認識及び切断するための組み込まれた配列を表す。図 1 A の第 3 のパネルは、ペプチドが付着した二本鎖 DNA(dsDNA)構築物を作製するためのポリメラーゼ伸長を示す。図 1 A の第 4 のパネルは、IIS 型 RE を用いて消化した後の dsDNA 構築物を示しており、凹部 5'リン酸化末端を有する 3'オーバーハング(2塩基対配列)を生成する。このようにして、1つ以上のバーコードを含む記録タグが調製され、コーディングタグからの情報転送のために利用可能である。

10

【図 1 B】図 1 A で生成された構造によるエンコーディングのサイクルを示す。図 1 B の左パネルは、ペプチドに結合した結合剤を示し、結合剤に付着したコーディングタグを記録タグと近接させる。いくつかの実施形態において、結合剤は、描写されていない場所(例えば、コーディングタグ又は他のループ領域)でコーディングタグに付着又は接合することができる。示される結合剤は、リンカーによってコーディングタグに付着される。コーディングタグは、結合剤特異的バーコード(BC)、2bp スペース、及び IIS 型制限酵素部位(RS)を含む。図 1 B の中央パネルは、最初の 2 つの酵素反応の生成物を示す。記録タグの 5'末端をコーディングタグの 3'末端にライゲーションすると、ポリメラーゼは、記録タグの 3'(ライゲーションされていない)末端を伸長して、それぞれの IIS 型 RE 部位に隣接する 2塩基対スペースを含有する dsDNA 分子を作製する。二本鎖に続いて、IIS 型 RE は、その認識部位に隣接して結合及び切断する。図 1 B の右パネルは、dsDNA が、結合剤特異的バーコード及びスペース配列として機能する 2nt の 3'オーバーハング(OH)を含有する、3つの酵素ステップ全ての後の最終生成物を示す。いくつかの実施形態において、情報転送のサイクルの後、分析のためのポリペプチドの一部を、ポリペプチドから除去することができる。図 1 B に示されるステップのサイクルを、追加の結合剤及びコーディングタグを用いて 1 回以上繰り返して、記録タグを更に伸長し得る。

20

30

【図 2】図 1 A ~ 図 1 B に示されるエンコーディング方法によって生成された例示的なエンコーディング結果。2つの試験ポリペプチド(F-ペプチド、配列番号 6 及び L-ペプチド、配列番号 7)を、固定化されたビーズ付着核酸記録タグに接合した。対応するコーディングタグに付着した 2 つの結合剤(F-結合剤及び L-結合剤)を使用して、試験ポリペプチドに接触させた。ライゲーションステップでは、3つの異なる濃度の T4 DNA リガーゼを試験した。ライゲーション後、伸長ステップのためにクレノウ断片を添加し、切断ステップのために BtsI-V2 酵素を添加した。エンコードされた記録タグの画分を NGS シーケンシングによって評価し、両方の結合剤について特定のエンコーディング結果を示した。各誤差バーは、平均から 1 標準誤差を使用して構成されている。

40

## 【発明を実施するための形態】

## 【 0 0 1 1 】

本明細書で提供されるのは、高分子を分析するための方法及びキットである。いくつかの実施形態において、この分析は、分子認識事象のバーコーディング及び核酸エンコーディングを用いる。提供される方法は、(a)高分子及び支持体に接合された関連する記録タ

50

グを提供することと、(b)高分子を、高分子に結合することができる結合剤であって、結合剤に関する識別情報を有するコーディングタグを含む結合剤と接触させて、高分子と結合剤との間の結合を可能にすることと、(c)核酸接合試薬を使用して記録タグの5'末端をコーディングタグの3'末端に接合することと、(d)ポリメラーゼによって、コーディングタグを鋳型として使用して記録タグを伸長し、二本鎖伸長記録タグを生成することと、(e)二本鎖伸長記録タグを二本鎖核酸切断試薬で切断して、伸長記録タグ内に3'オーバーハングを生成することと、を含む。これらのステップを実行すると、情報が、コーディングタグから記録タグに転送されて、伸長記録タグを生成する。また、高分子シーケンシング及び/又は分析のために提供された方法を実行するための成分及び/又は試薬を含むキットも提供される。いくつかの実施形態において、キットはまた、本明細書で提供される方法のうちのいずれかを実行するためにキットを使用するための説明書を含む。

10

#### 【0012】

タンパク質などの高分子の高度に並列な特徴付け及び認識は、依然として課題である。プロテオミクスでは、1つの目標は、サンプル中の多数のタンパク質を識別しかつ定量化することであり、これは高スループットの方法で達成するのが難しいタスクである。イムノアッセイ及び質量分析に基づく方法などのアッセイが使用されているが、サンプル及び分析物レベルの両方に限定され、感度及びダイナミックレンジ、並びに交差反応性及びバックグラウンドシグナルが制限される。例えば、検出可能な標識を有する親和性剤を使用して、同族高分子の集合に対する親和性剤の集合の読み出しを多重化することは、依然として困難である。高分子分析に関連する改良された技法とともに、タンパク質シーケンシング及び/又は分析、並びにそれを達成するための製品、方法、及びキットへの適用の必要性が残っている。効率的で、高度に並列化され、正確で、感度が高く、高スループットである高分子分析を実行するためのプロテオミクス技術が必要である。本開示は、これら及び他の関連する必要性を満たす。

20

#### 【0013】

いくつかの実施形態において、本開示は、部分的に、タンパク質及びペプチドの特徴付け、定量化、及び/又はシーケンシングへの直接適用を伴う、情報転送を含む高分子を分析するための方法を提供する。いくつかの例では、転送される情報は、高分子に結合するように構成されている結合剤に関する識別情報を含む。いくつかの実施形態において、サンプルから得られた複数の高分子が分析される。いくつかの実施形態において、サンプルは、対象から得られる。いくつかの実施形態において、高分子シーケンシング又は分析方法は、コーディングタグに関連付けられた複数の結合剤を使用して、分析される複数の高分子を検出することを含む。

30

#### 【0014】

本明細書で提供される情報転送方法は、複数の酵素を利用して、核酸分子とのライゲーション、伸長、及び切断反応を実行する。いくつかの実施形態において、提供される方法は、ヘアピン構造及び制限酵素部位(又はその一部)を含むオリゴヌクレオチドを含む。いくつかの実施形態において、方法は、混合酵素が反応に提供される反応システムの使用を含む。例えば、ポリメラーゼ、核酸接合試薬、及び二本鎖核酸切断試薬の活性は、コーディングタグからの情報を記録タグに転送して、伸長記録タグを生成する、好適な条件で提供される。提供される方法において、使用される記録タグは、少なくとも部分的に二本鎖DNA構造を含む。説明された方法を使用するいくつかの利点は、高い情報転送(エンコーディング)の成功、ステップワイズ反応のためのシンプルな設計、単一のステップで実行する/単一のポット反応として実行するオプション、スペーサーの必要性を低減すること、又はスペーサーの長さを短縮すること、及び/又はシステム内のDNA-DNA相互作用を最小限に抑えることを含む。

40

#### 【0015】

本開示の完全な理解を提供するために、以下の説明には多くの具体的な詳細が記載される。これらの詳細は、例示目的で提供されており、特許請求される主題は、これらの特定の

50

詳細の一部又は全部がなくても、特許請求の範囲に従って実施することができる。特許請求される主題の範囲から逸脱することなく、他の実施形態を使用することができ、構造変更を行うことができることを理解されたい。個々の実施形態のうちの一つ以上で説明される様々な特徴及び機能は、それらが説明される特定の実施形態へのそれらの適用可能性に関して限定されないことを理解されたい。代わりに、それらは、そのような実施形態が説明されるかどうか、及びそのような特徴が説明される実施形態の一部として提示されるかどうかにかかわらず、単独で、又はある組み合わせで、本開示の他の実施形態の一つ以上に適用することができる。明確にするために、特許請求される主題に関連する技術分野で既知の技術項目 (technical material) は、特許請求される主題が不必要に不明瞭とならない程詳細には説明されない。

10

## 【0016】

本出願で言及される特許文書、科学論文及びデータベースを含む全ての出版物は、各個々の出版物が参照により個別に組み込まれるのと同じ程度まで、あらゆる目的のためにそれらの全体が参照により組み込まれる。出版物又は文書の引用は、これらのいずれかが関連する先行技術であることを承認することを意図しておらず、これらの出版物又は文書の内容又は日付に関するいかなる承認を構成するものではない。

## 【0017】

全ての見出しは、読者の利便性のために提供されるものであり、特に明記しない限り、見出しに続く本文の意味を限定するために使用されるべきでない。

## 【0018】

20

## 定義

別段の定義がない限り、本明細書で使用される全ての技術用語及び科学用語は、本開示が属する技術分野の当業者により通常理解されるものと同じ意味を有する。本章に記載される定義が、参照により本明細書に組み込まれる特許、出願、公開出願、及び他の出版物に記載される定義に反するか、さもなければそれと矛盾する場合、本章に記載される定義は、参照により本明細書に組み込まれる定義よりも優先される。

## 【0019】

本明細書で使用される場合、単数形「a」、「an」及び「the」は、文脈が明確に別のことを指示する場合を除き、複数の指示対象を包含する。したがって、例えば、「ペプチド (peptide)」への言及は、一つ以上のペプチド、又はペプチドの混合物を包含する。また、本明細書で使用される場合、具体的に述べられているか、又は文脈から明らかである場合を除き、「又は (or)」という用語は、包括的であると理解され、「又は (or)」及び「及び (and)」の両方を包含する。

30

## 【0020】

本明細書で使用される場合、「約 (about)」という用語は、当業者に容易に知られているそれぞれの値の通常の誤差範囲を指す。本明細書における「約」値又はパラメータへの言及は、その値又はパラメータ自体を対象とする実施形態を包含 (及び説明) する。例えば、「約 X」について言及する記載は、「X」についての記載を包含する。

## 【0021】

本明細書における「抗体」という用語は、最も広義に使用され、無傷の抗体を含むポリクローナル抗体及びモノクローナル抗体、並びに断片抗原結合 (Fab) 断片、F(ab')<sub>2</sub> 断片、Fab' 断片、Fv 断片、組換え IgG (rIgG) 断片、一本鎖可変断片 (scFv) を含む一本鎖抗体断片、及び単一ドメイン抗体 (例えば、sdAb、sdFv、ナノボディ) 断片を含む機能的 (抗原結合) 抗体断片を含む。この用語は、イントラボディ、ペプチボディ、キメラ抗体、完全ヒト抗体、ヒト化抗体、及びヘテロ共役抗体、多重特異性、例えば、二重特異性、抗体、ダイアボディ、トリアボディ、及びテトラボディ、タンデム di-scFv、タンデム tri-scFv などの、遺伝子操作及び/又は他の方法で修飾された形態の免疫グロブリンを包含する。別段の定めがない限り、「抗体」という用語は、その機能的抗体断片を包含することを理解されたい。この用語はまた、IgG 及びそのサブクラス、IgM、IgE、IgA、及び IgD を含む、任意のクラス又

40

50

はサブクラスの抗体を含む、無傷の又は全長の抗体を包含する。

【0022】

「個体」又は「対象」は、哺乳動物を含む。哺乳動物としては、家畜（例えば、ウシ、ヒツジ、ネコ、イヌ、及びウマ）、霊長類（例えば、ヒト及びサルなどの非ヒト霊長類）、ウサギ、及びげっ歯類（例えば、マウス及びラット）が挙げられるが、これらに限定されない。「個体」又は「対象」には、ニワトリなどの鳥類、魚類などの脊椎動物、並びにマウス、ラット、ウサギ、ネコ、イヌ、ブタ、ウシ、雄ウシ、ヒツジ、ヤギ、ウマ、サル、及び他の非ヒト霊長類などの哺乳動物が含まれ得る。特定の実施形態において、個体又は対象は、ヒトである。

【0023】

本明細書で使用される場合、「サンプル」という用語は、分析物アッセイが所望される分析物を含有し得る任意のものを指す。本明細書で使用される場合、「サンプル」は、溶液、懸濁液、液体、粉末、ペースト、水性、非水性、又はそれらの任意の組み合わせであり得る。サンプルは、生物学的流体又は生物学的組織などの生物学的サンプルであり得る。生物学的流体の例としては、尿、血液、血漿、血清、唾液、精液、便、痰、脳脊髄液、涙、粘液、羊水などが挙げられる。生物学的組織は、結合組織、上皮組織、筋肉及び神経組織を含む、ヒト、動物、植物、細菌、真菌又はウイルス構造の構造材料のうちの1つを形成する細胞間物質とともに、通常特定の種類の細胞の凝集体である。生物学的組織の例としては、臓器、腫瘍、リンパ節、動脈、及び個々の細胞も挙げられる。

【0024】

いくつかの実施形態において、サンプルは、生物学的サンプルである。本開示の生物学的サンプルは、溶液、懸濁液、液体、粉末、ペースト、水性サンプル、又は非水性サンプルの形態のサンプルを包含する。本明細書で使用される場合、「生物学的サンプル」は、生体若しくはウイルス（又はプリオン）供給源又は高分子及び生体分子の他の供給源から得られる任意のサンプルを含み、核酸、タンパク質及び/又は他の高分子を得ることができる対象の任意の細胞型又は組織を含む。生物学的サンプルは、生物学的供給源から直接得られるサンプル、又は処理されるサンプルであり得る。例えば、増幅された単離された核酸は、生物学的サンプルを構成する。生物学的サンプルとしては、血液、血漿、血清、脳脊髄液、滑液、尿及び汗などの体液、動物及び植物由来の組織及び臓器サンプル、並びにそこから得られる処理されたサンプルが挙げられるが、これらに限定されない。いくつかの実施形態において、サンプルは、組織若しくは体液、例えば、結合組織、上皮組織、筋肉組織、若しくは神経組織、脳、肺、肝臓、脾臓、骨髄、胸腺、心臓、リンパ、血液、骨、軟骨、膵臓、腎臓、胆嚢、胃、腸、睾丸、卵巣、子宮、直腸、神経系、腺、及び内部血管からなる群から選択される組織、又は血液、尿、唾液、骨髄、精子、腹水、及びそれらの細画分、例えば血清又は血漿、からなる群から選択される体液から誘導され得る。

【0025】

「レベル（level）」又は「レベル（levels）」という用語は、標的、例えば、疾患若しくは障害の病因の一部であり、定性的若しくは定量的に決定することができる物質又は生物の存在及び/又は量を指すために使用される。標的レベルにおける「定性的な」変化は、検出不可能であるか、若しくは正常対照から得られたサンプル中に存在する標的の出現又は消失を指す。1つ以上の標的のレベルにおける「定量的な」変化は、健全な対照と比較した場合の標的レベルの測定可能な増加又は減少を指す。

【0026】

本明細書で使用される場合、「高分子」という用語は、より小さなサブユニットから構成される大きな分子を包含する。高分子の例としては、ペプチド、ポリペプチド、タンパク質、核酸、炭水化物、脂質、大環状分子、又はそれらの組み合わせ若しくは複合体が挙げられるが、これらに限定されない。高分子はまた、一緒に共有結合された2つ以上のタイプの高分子の組み合わせから構成されるキメラ高分子（例えば、核酸に結合されたペプチド）を含む。高分子はまた、2つ以上の高分子の非共有結合性複合体から構成される「高分子集合体」を含み得る。高分子集合体は、同じタイプの高分子（例えば、タンパク質 -

10

20

30

40

50

タンパク質)又は2つ以上の異なるタイプの高分子(例えば、タンパク質 - DNA)から構成され得る。

【0027】

本明細書で使用される場合、「ポリペプチド」という用語は、ペプチド及びタンパク質を包含し、ペプチド結合によって接合された2つ以上のアミノ酸の鎖を含む分子を指す。いくつかの実施形態において、ポリペプチドは、2～50個のアミノ酸を含み、例えば、20～30個を超えるアミノ酸を有する。いくつかの実施形態において、ペプチドは、二次、三次、又はより高次の構造を含まない。いくつかの実施形態において、ポリペプチドは、タンパク質である。いくつかの実施形態において、タンパク質は、30個以上のアミノ酸を含み、例えば、50個を超えるアミノ酸を有する。いくつかの実施形態において、タンパク質は、一次構造に加えて、二次、三次、又はより高次の構造を含む。ポリペプチドのアミノ酸は、最も典型的にはL-アミノ酸であるが、D-アミノ酸、修飾アミノ酸、アミノ酸類似体、アミノ酸模倣物、又はそれらの任意の組み合わせでもあり得る。ポリペプチドは、天然に存在するか、合成的に産生されるか、又は組換え的に発現され得る。ポリペプチドは、合成的に産生されるか、単離されるか、組換え的に発現されるか、又は上記のような方法論の組み合わせによって産生され得る。ポリペプチドはまた、アミノ酸鎖を修飾する追加の基、例えば、翻訳後修飾を介して追加される官能基を含み得る。ポリマーは、直鎖状又は分岐鎖状であってもよく、修飾アミノ酸を含んでもよく、非アミノ酸によって中断されてもよい。この用語はまた、天然に又は介入によって修飾されたアミノ酸ポリマーを包含し、例えば、ジスルフィド結合の形成、グリコシル化、脂質化、アセチル化、リン酸化、又は標識成分との共役などの任意の他の操作又は修飾である。

【0028】

本明細書で使用される場合、「アミノ酸」という用語は、ペプチドの単量体サブユニットとして機能する、アミン基、カルボン酸基、及び各アミノ酸に特異的な側鎖を含む有機化合物を指す。アミノ酸としては、20個の標準的、天然又は正準アミノ酸、及び非標準的なアミノ酸が挙げられる。標準的な天然アミノ酸としては、アラニン(A又はAla)、システイン(C又はCys)、アスパラギン酸(D又はAsp)、グルタミン酸(E又はGlu)、フェニルアラニン(F又はPhe)、グリシン(G又はGly)、ヒスチジン(H又はHis)、イソロイシン(I又はIle)、リジン(K又はLys)、ロイシン(L又はLeu)、メチオニン(M又はMet)、アスパラギン(N又はAsn)、プロリン(P又はPro)、グルタミン(Q又はGln)、アルギニン(R又はArg)、セリン(S又はSer)、スレオニン(T又はThr)、バリン(V又はVal)、トリプトファン(W又はTrp)、及びチロシン(Y又はTyr)が挙げられる。アミノ酸は、L-アミノ酸又はD-アミノ酸であり得る。非標準的なアミノ酸は、天然に存在するか、又は化学的に合成される修飾アミノ酸、アミノ酸類似体、アミノ酸模倣物、非標準的なタンパク質原性アミノ酸、又は非タンパク質原性アミノ酸であり得る。非標準的なアミノ酸の例としては、セレノシステイン、ピロリシン、及びN-ホルミルメチオニン、 $\omega$ -アミノ酸、ホモアミノ酸、プロリン及びピルビン酸誘導体、3-置換アラニン誘導体、グリシン誘導体、環-置換フェニルアラニン及びチロシン誘導体、線状コアアミノ酸、N-メチルアミノ酸が挙げられるが、これらに限定されない。

【0029】

本明細書で使用される場合、「翻訳後修飾」という用語は、その翻訳、例えば、リボソームによる翻訳が完了した後にペプチド上で生じる修飾を指す。翻訳後修飾は、共有結合による化学修飾又は酵素修飾であり得る。翻訳後修飾の例としては、アシル化、アセチル化、アルキル化(メチル化を含む)、ピオチン化、プチリル化、カルバミル化、カルボニル化、脱アミド化、脱イミノ化(deimination)、ジフタミド形成、ジスルフィド架橋形成、脱離(eliminylation)、フラビン付着、ホルミル化、ガンマカルボキシル化、グルタミル化、グリシル化、グリコシル化、グリピエーション、ヘムC付着、ヒドロキシル化、ヒプシン形成、ヨウ素化、イソプレニル化、脂質化、リポイル化、マロニル化、メチル化、ミリストリル化(myristoylation)、酸化

、パルミトイル化、ペグ化、ホスホパンテテイン化、リン酸化、プレニル化、プロピオニル化、レチニリデンシフ塩基形成、S-グルタチオン付加、S-ニトロシル化、S-スルフェニル化、セレン化、スクシニル化、スルフィン化、ユビキチン化、及びC末端アミド化が挙げられるが、これらに限定されない。翻訳後修飾としては、ペプチドのアミノ末端及び/又はカルボキシル末端の修飾が挙げられる。末端アミノ基の修飾としては、デスアミノ、N-低級アルキル、N-ジ-低級アルキル、及びN-アシル修飾が挙げられるが、これらに限定されない。末端カルボキシ基の修飾としては、アミド、低級アルキルアミド、ジアルキルアミド、及び低級アルキルエステル修飾（例えば、低級アルキルはC<sub>1</sub>~C<sub>4</sub>アルキルである）が挙げられるが、これらに限定されない。翻訳後修飾としては、アミノ末端とカルボキシ末端との間にあるアミノ酸の修飾（例えば、上記のものであるが、これらに限定されない）も挙げられる。翻訳後修飾という用語はまた、1つ以上の検出可能な標識を含むペプチド修飾を含み得る。

10

#### 【0030】

本明細書で使用される場合、「結合剤」という用語は、結合標的、例えば、ポリペプチド又はポリペプチドの成分若しくは機構に結合する、それと会合する、それと一体化する、それを認識する、又はそれと組み合わせられる核酸分子、ペプチド、ポリペプチド、タンパク質、炭水化物、又は小分子を指す。結合剤は、ポリペプチド又はポリペプチドの成分若しくは機構と共有結合性会合又は非共有結合性会合を形成し得る。結合剤はまた、核酸分子-ペプチドキメラ結合剤又は炭水化物-ペプチドキメラ結合剤などの2タイプ以上の分子から構成されるキメラ結合剤であり得る。結合剤は、天然に存在するか、合成的に産生されるか、又は組換え的に発現される分子であり得る。結合剤は、ポリペプチドの単一の単量体若しくはサブユニット（例えば、ポリペプチドの単一アミノ酸）に結合し得るか、又はポリペプチドの複数の連結されたサブユニット（例えば、より長いペプチド、ポリペプチド、又はタンパク質分子のジペプチド、トリペプチド、又はより高次のペプチド）に結合し得る。結合剤は、線状分子又は三次元構造（立体配座とも称される）を有する分子に結合し得る。例えば、抗体結合剤は、線状ペプチド、ポリペプチド、若しくはタンパク質に結合するか、又は立体配座ペプチド、ポリペプチド、若しくはタンパク質に結合し得る。結合剤は、ペプチド、ポリペプチド、又はタンパク質分子のN末端ペプチド、C末端ペプチド、又は介在ペプチドに結合し得る。結合剤は、ペプチド分子のN末端アミノ酸、C末端アミノ酸、又は介在アミノ酸に結合し得る。結合剤は、好ましくは、修飾又は標識されていないアミノ酸よりも、化学的に修飾又は標識されたアミノ酸（例えば、化学試薬によって標識されたアミノ酸）に結合し得る。例えば、結合剤は、好ましくは、標識されていない又は修飾されていないアミノ酸よりも標識又は修飾されたアミノ酸に結合し得る。結合剤は、ペプチド分子の翻訳後修飾に結合し得る。結合剤は、ポリペプチドの成分又は機構への選択的結合を示し得る（例えば、結合剤は、20個の可能な天然アミノ酸残基のうちの1つに選択的に結合し、他の19個の天然アミノ酸残基には非常に低い親和性で結合するか、又は全く結合しない可能性がある）。結合剤が、ポリペプチドの複数の成分又は機構に結合することができるか又はそれらに結合するように構成される場合、結合剤は、低い選択的結合を示し得る（例えば、結合剤は、2つ以上の異なるアミノ酸残基に同様の親和性で結合し得る）。結合剤は、リンカーによって結合剤に接合され得るコーディングタグを含み得る。

20

30

40

#### 【0031】

本明細書で使用される場合、「リンカー」という用語は、2つの分子を接合するために使用されるヌクレオチド、ヌクレオチド類似体、アミノ酸、ペプチド、ポリペプチド、ポリマー、又は非ヌクレオチド化学部分のうちの1つ以上を指す。リンカーを使用して、例えば、結合剤をコーディングタグと接合し、記録タグをポリペプチドと接合し、ポリペプチドを支持体と接合し、記録タグを固体支持体と接合する。特定の実施形態において、リンカーは、酵素反応又は化学反応（例えば、クリック化学）によって、2つの分子を接合する。

#### 【0032】

50

本明細書で使用される場合、「リガンド」という用語は、本明細書に記載の化合物に接続された任意の分子又は部分を指す。「リガンド」は、化合物に付着した1つ以上のリガンドを指す場合がある。いくつかの実施形態において、リガンドは、ペンダント基又は結合部位（例えば、結合剤が結合する部位）である。

#### 【0033】

本明細書で使用される場合、「プロテオーム」という用語は、任意の生物の、特定の時間にゲノム、細胞、組織、又は生物によって発現されるタンパク質、ポリペプチド又はペプチド（それらの共役体又は複合体を含む）の全セットを含み得る。一態様において、プロテオームは、規定された条件下で、所与の時間に、所与のタイプの細胞又は生物において発現されるタンパク質のセットである。プロテオミクスは、プロテオームの研究である。例えば、「細胞プロテオーム」は、ホルモン刺激への曝露など、特定の一連の環境条件下で特定の細胞型において見られるタンパク質の集合を含み得る。生物の完全なプロテオームには、様々な細胞プロテオームの全てからのタンパク質の完全なセットが含まれ得る。プロテオームには、特定の細胞レベル下の生体系におけるタンパク質の集合も含まれ得る。例えば、ウイルス内の全てのタンパク質は、ウイルスプロテオームと呼ばれる可能性がある。本明細書で使用される場合、「プロテオーム」という用語は、キノーム；セクレトーム；レセプトーム（例えば、GPCRome）；免疫プロテオーム；ニュートリプロテオーム；翻訳後修飾（例えば、リン酸化、ユビキチン化、メチル化、アセチル化、グリコシル化、酸化、脂質化、及び/若しくはニトロシル化）によって規定されるプロテオームサブセット、例えば、ホスホプロテオーム（例えば、ホスホチロシン-プロテオーム、チロシン-キノーム、及びチロシン-ホスファトーム）、グリコプロテオームなど；組織若しくは器官、発達段階、又は生理学的若しくは病理学的状態に関連するプロテオームサブセット；細胞周期、分化（若しくは脱分化）、細胞死、老化、細胞移動、形質転換、又は転移などの細胞プロセスに関連するプロテオームサブセット；あるいはそれらの任意の組み合わせを含むが、これらに限定されないプロテオームのサブセットを含む。本明細書で使用される場合、「プロテオミクス」という用語は、細胞、組織、及び体液内のプロテオームの定量的分析、並びに細胞内及び組織内のプロテオームの対応する空間分布を指す。更に、プロテオミクス研究には、生物学及び規定された生物学的又は化学的刺激の関数として経時的に継続変化するプロテオームの動的状態が含まれる。

#### 【0034】

遊離アミノ基を有するペプチド又はポリペプチド鎖の一端の末端アミノ酸は、本明細書では「N末端アミノ酸」（NTAA）と称される。遊離カルボキシル基を有する鎖のもう一方の末端の末端アミノ酸は、本明細書では「C末端アミノ酸」（CTAA）と称される。ペプチドの長さが「n」個のアミノ酸である場合、ペプチドを構成するアミノ酸には、順に番号を付けることができる。本明細書で使用される場合、NTAAは、n目番目のアミノ酸（本明細書では「nNTAA」とも称される）とみなされる。この命名法を使用すると、次のアミノ酸は、n-1個のアミノ酸、次いでn-2個のアミノ酸というように、ペプチドのN末端からC末端までの長さを下っていく。特定の実施形態において、NTAA、CTAA、又はその両方は、部分又は化学部分で修飾又は標識され得る。

#### 【0035】

本明細書で使用される場合、「バーコード」という用語は、ポリペプチド、結合剤、結合サイクルからの結合剤のセット、サンプルポリペプチド、サンプルのセット、コンパートメント（例えば、液滴、ビーズ、又は分離された位置）内のポリペプチド、コンパートメントのセット内のポリペプチド、ポリペプチドの画分、ポリペプチド画分のセット、空間領域若しくは空間領域のセット、ポリペプチドのライブラリー、又は結合剤のライブラリーについての固有の識別子タグ又は起源情報を提供する、約2～約30塩基（例えば、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29又は30塩基）の核酸分子を指す。バーコードは、人工的な配列又は天然に存在する配列であり得る。特定の実施形態において、バーコードの集団内の各バーコードは異なる。他の実施形態に

10

20

30

40

50

において、バーコードの集団におけるバーコードの一部は異なり、例えば、バーコードの集団におけるバーコードの少なくとも約10%、15%、20%、25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、97%、又は99%異なる。バーコードの集団は、ランダムに生成されるか、又はランダムではなく生成される可能性がある。特定の実施形態において、バーコードの集団は、エラー訂正又はエラー許容バーコードである。バーコードを使用して、多重化されたシーケンシングデータを計算によりデコンポリューションし、個々のポリペプチド、サンプル、ライブラリーなどに由来する配列リードを識別することができる。また、バーコードを、マッピングを強化するために小さなコンパートメントに分配されたポリペプチド集合のデコンポリューションに使用することもできる。例えば、ペプチドを

10

#### 【0036】

本明細書で使用される場合、「コーディングタグ」という用語は、その関連する結合剤に関する識別情報を含む、任意の好適な長さを有するポリヌクレオチド、例えば、2~100(2及び100を含む)の任意の整数を含む、約2塩基~約100塩基の核酸分子を指す。「コーディングタグ」は、「シーケンシング可能なポリマー」から作製することもできる(例えば、Niu et al., 2013, Nat. Chem. 5: 282-292、Roy et al., 2015, Nat. Commun. 6: 7237、Lutz 2015, Macromolecules 48: 4759-4767を参照。これらの各々はその全体が参照により組み込まれる)。コーディングタグは、エンコーダー配列を含み得、このエンコーダー配列には、任意選択的に、片側に1つのスペーサーが隣接するか、又は任意選択的に、両側にスペーサーが隣接する。コーディングタグはまた、任意選択的なUMI及び/又は任意の結合サイクル特異的バーコードで構成され得る。コーディングタグは、一本鎖又は二本鎖であり得る。二本鎖コーディングタグは、平滑末端、突出末端、又はその両方を含み得る。コーディングタグは、結合剤に直接付着しているコーディングタグ、結合剤に直接付着しているコーディングタグにハイブリダイズした相補配列(例えば、二本鎖コーディングタグの場合)、又は伸長記録タグに存在するコーディングタグ情報を指す場合がある。特定の実施形態において、コーディングタグは、結合サイクル特異的スペーサー若しくはバーコード、固有の分子識別子、ユニバーサルプライミング

20

30

#### 【0037】

本明細書で使用される場合、「スペーサー」(Sp)という用語は、記録タグ又はコーディングタグの末端上に存在する、長さが約1塩基~約20個の塩基(例えば、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、又は20個の塩基)の核酸分子を指す。特定の実施形態において、スペーサー配列は、一端又は両端でコーディングタグのエンコーダー配列に隣接する。結合剤のポリペプチドへの結合に続いて、それらの関連するコーディングタグ及び記録タグ上の相補的スペーサー配列間のアニーリングは、それぞれ、プライマー伸長反応又は記録タグ、コーディングタグ、若しくはジタグ構築物へのライゲーションを介した結合情報の伝達を可能にする。Sp'は、Spに相補的なスペーサー配列を指す。好ましくは、結合剤ライブラリー内のスペーサー配列は、同じ数の塩基を有する。共通の(共有又は同一の)スペーサーを、結合剤ライブラリーで使用することができる。スペーサー配列は、特定の結合サイクルにおいて使用される結合剤を追跡するために、「サイクル特異的」配列を有し得る。スペーサー配列(Sp)は、全ての結合サイクルにわたって一定であるか、特定のクラスのポリペプチドに特異的であるか、又は結合サイクル数に特異的である可能性がある。ポリペプチドクラス特異的スペーサーにより、完了した結合/伸長サイクルからの伸長記録タグに存在する同種結合剤のコーディングタグ情報が、後続の結合サイクルにおける同じクラスのポリペプチドを認識する別の結合剤のコーディングタグに、クラス特異的スペーサーを介してアニーリングすることができる。正確な同種の対の逐次的な結合によってのみ、

40

50

相互作用するスペーサー要素及び効果的なプライマー伸長がもたらされる。スペーサー配列は、記録タグ内の相補的スペーサー配列にアニーリングしてプライマー伸長（ポリメラーゼ伸長とも称される）反応を開始させるか、又はライゲーション反応用の「副木」を提供するか、又は「粘着末端」ライゲーション反応を媒介するのに十分な数の塩基を含み得る。スペーサー配列は、コーディングタグ内のエンコーダー配列よりも少ない数の塩基を含み得る。

#### 【0038】

本明細書で使用される場合、「記録タグ」という用語は、部分、例えば、化学カップリング部分、核酸分子、又はシーケンシング可能なポリマー分子を指し（例えば、Niuet al., 2013, Nat. Chem. 5: 282-292、Roy et al., 2015, Nat. Commun. 6: 7237、Lutz, 2015, Macromolecules 48: 4759-4767を参照。これらの各々は、その全体が参照により組み込まれる）、そこにコーディングタグの識別情報が転送され得るか、又はそこから記録タグと関連付けられた高分子に関する識別情報（例えば、UMI情報）がコーディングタグに転送され得る。識別情報は、サンプル、画分、パーティション、空間位置、相互作用する隣接分子、サイクル数などに関連する情報など、分子を特徴付ける任意の情報を含むことができる。更に、UMI情報の存在も識別情報として分類することができる。特定の実施形態において、結合剤がポリペプチドに結合した後、結合剤に連結したコーディングタグからの情報を、結合剤がポリペプチドに結合している間に、ポリペプチドと関連付けられた記録タグに転送することができる。他の実施形態において、結合剤がポリペプチドに結合した後、ポリペプチドと関連付けられた記録タグからの情報を、結合剤がポリペプチドに結合している間に、結合剤に連結されたコーディングタグに転送することができる。記録タグは、ポリペプチドに直接連結され得るか、多官能性リンカーを介してポリペプチドに連結され得るか、又は支持体上でのその近接性（若しくは共局在化）によってポリペプチドと関連付けられ得る。記録タグは、その連結が、コーディングタグ情報を記録タグに、又はその逆に転送するために使用される方法と適合性がある限り、その5'末端若しくは3'末端を介して、又は内部部位で連結され得る。記録タグは、他の機能的成分、例えば、ユニバーサルプライミング部位、固有の分子識別子、バーコード（例えば、サンプルバーコード、画分バーコード、空間バーコード、コンパートメントタグなど）、コーディングタグのスペーサー配列に相補的であるスペーサー配列、又はそれらの任意の組み合わせを更に含み得る。記録タグのスペーサー配列は、好ましくは、ポリメラーゼ伸長を使用して、コーディングタグ情報を記録タグに転送する実施形態において、記録タグの3'末端にある。

#### 【0039】

本明細書で使用される場合、「ポリメラーゼ伸長」とも称される「プライマー伸長」という用語は、核酸ポリメラーゼ（例えば、DNAポリメラーゼ）によって触媒され、それにより、相補鎖にアニーリングする核酸分子（例えば、オリゴヌクレオチドプライマー、スペーサー配列）が、相補鎖を鋳型として使用してポリメラーゼによって伸長される反応を指す。

#### 【0040】

本明細書で使用される場合、「固有の分子識別子」又は「UMI」という用語は、UMIが連結される各高分子、ポリペプチド又は結合剤に固有の識別子タグを提供する、約3～約40個の塩基（3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、又は40個の塩基）の長さの核酸分子を指す。ポリペプチドUMIを使用して、複数の伸長記録タグからのシーケンシングデータを計算によりデコンボリューションして、個々のポリペプチドに由来する伸長記録タグを識別することができる。ポリペプチドUMIを使用して、NGSリードを固有のUMIに折り畳むことにより、元のポリペプチド分子を正確に計数することができる。結合剤UMIを使用して、特定のポリペプチドに結合する各個々の分子結合剤

10

20

30

40

50

を識別することができる。例えば、UMIを使用して、特定のペプチド分子に対して発生する単一のアミノ酸に特異的な結合剤の個々の結合事象の数を特定することができる。UMI及びバーコードを両方とも結合剤又はポリペプチドに関連して参照する場合、バーコードは、個々の結合剤又はポリペプチドに関するUMI以外の識別情報を指す（例えば、サンプルバーコード、コンパートメントバーコード、結合サイクルバーコード）ことが理解される。

#### 【0041】

本明細書で使用される場合、「ユニバーサルプライミング部位」又は「ユニバーサルプライマー」又は「ユニバーサルプライミング配列」という用語は、ライブラリーの増幅及び/又はシーケンシング反応に使用され得る核酸分子を指す。ユニバーサルプライミング部位には、PCR増幅用のプライミング部位（プライマー配列）、いくつかの次世代シーケンシングプラットフォームにおいてブリッジ増幅を可能にするフローセル表面上の相補的オリゴヌクレオチドにアニーリングするフローセルアダプター配列、シーケンシングプライミング部位、又はそれらの組み合わせが含まれ得るが、これらに限定されない。ユニバーサルプライミング部位は、次世代デジタルシーケンシングと組み合わせて一般的に使用されるものを含む、他のタイプの増幅に使用することができる。例えば、伸長記録タグ分子を環状化し、ユニバーサルプライミング部位をローリングサークル増幅に使用して、シーケンシング鑄型として使用することができるDNAナノボールを形成することができる（Drmanac et al., 2009, Science 327: 78-81）。あるいは、記録タグ分子を環状化し、ユニバーサルプライミング部位からのポリメラーゼ伸長によって直接シーケンシングすることができる（Korlach et al., 2008, Proc. Natl. Acad. Sci. 105: 1176-1181）。「ユニバーサルプライミング部位」又は「ユニバーサルプライマー」に関連して使用される場合の「フォワード」という用語は、「5」又は「センス」と称される場合もある。「ユニバーサルプライミング部位」又は「ユニバーサルプライマー」に関連して使用される場合の「リバース」という用語は、「3」又は「アンチセンス」と称される場合もある。

#### 【0042】

本明細書で使用される場合、「伸長記録タグ」という用語は、結合剤のポリペプチドへの結合に続いて、少なくとも1つの結合剤のコーディングタグ（又はその相補的配列）の情報が転送された記録タグを指す。コーディングタグの情報は、直接的に（例えば、ライゲーション）又は間接的に（例えば、プライマー伸長）記録タグに転送され得る。コーディングタグの情報は、酵素的又は化学的に記録タグに転送され得る。伸長記録タグは、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、20、21、22、23、24、25、26、27、28、29、30、31、32、33、34、35、36、37、38、39、40、45、50、55、60、65、70、75、80、85、90、95、100、125、150、175、200以上のコーディングタグの結合剤情報を含み得る。伸長記録タグの塩基配列は、それらのコーディングタグによって識別される結合剤の結合の時間的及び連続的な順序を反映し得るか、コーディングタグによって識別される結合剤の結合の一部の連続的順序を反映し得るか、又はコーディングタグによって識別される結合剤の結合のいかなる順序も反映しない可能性がある。特定の実施形態において、伸長記録タグに存在するコーディングタグ情報は、分析されているポリペプチド配列を、少なくとも25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、又は100%の同一性で表す。伸長記録タグが分析されているポリペプチド配列を100%の同一性で表さない特定の実施形態において、エラーは、結合剤によるオフターゲットの結合、若しくは結合サイクルの「失敗」（例えば、結合サイクル中に結合剤がポリペプチドに結合できないことが原因で、プライマー伸長反応の失敗が原因で）、又はその両方に起因する可能性がある。

#### 【0043】

10

20

30

40

50

本明細書で使用される場合、「固体支持体」、「固体表面」、又は「固体基質」、又は「シーケンシング基質」、又は「基質」という用語は、共有結合性の相互作用及び非共有結合性の相互作用、又はそれらの任意の組み合わせを含む、当技術分野で既知の任意の手段によって、直接的又は間接的にポリペプチドを関連付けることができる、多孔質材料及び非多孔質材料を含めた任意の固体材料を指す。固体支持体は、二次元（例えば、平面）又は三次元（例えば、ゲルマトリックス又はビーズ）であり得る。固体支持体は、ビーズ、マイクロビーズ、アレイ、ガラス表面、シリコン表面、プラスチック表面、フィルター、膜、PTFE膜、PTFE膜、ニトロセルロース膜、ニトロセルロースベースのポリマー表面、ナイロン、シリコンウェーハチップ、フロースルーチップ、フローセル、シグナル変換電子機器を含めたバイオチップ、チャンネル、マイクロタイターウェル、ELISAプレート、スピン干渉ディスク、ニトロセルロース膜、ニトロセルロースベースのポリマー表面、ポリマーマトリックス、ナノ粒子、又はマイクロスフェアを含むが、これらに限定されない任意の支持体表面であり得る。固体支持体の材料としては、アクリルアミド、アガロース、セルロース、デキストラン、ニトロセルロース、ガラス、金、石英、ポリスチレン、ポリエチレン酢酸ビニル、ポリプロピレン、ポリエステル、ポリメタクリレート、ポリアクリレート、ポリエチレン、ポリエチレンオキシド、ポリシリケート、ポリカーボネート、ポリビニルアルコール（PVA）、テフロン（登録商標）、フルオロカーボン、ナイロン、シリコンゴム、ポリ無水物、ポリグリコール酸、ポリ塩化ビニル、ポリ乳酸、ポリオルトエステル、官能化シラン、ポリプロピルフェレート、コラーゲン、グリコサミノグリカン、ポリアミノ酸、デキストラン、又はそれらの任意の組み合わせが挙げられるが、これらに限定されない。固体支持体は更に、薄膜、膜、ボトル、皿、繊維、織織維、チューブなどの成形ポリマー、粒子、ビーズ、マイクロスフェア、微粒子、又はそれらの任意の組み合わせを含む。例えば、固体表面がビーズである場合、ビーズは、セラミックビーズ、ポリスチレンビーズ、ポリマービーズ、ポリアクリレートビーズ、メチルスチレンビーズ、アガロースビーズ、セルロースビーズ、デキストランビーズ、アクリルアミドビーズ、固体コアビーズ、多孔質ビーズ、常磁性ビーズ、ガラスビーズ、制御多孔質ビーズ、シリカベースのビーズ、又はそれらの任意の組み合わせを含み得るが、これらに限定されない。ビーズは、球形であっても不規則な形状であってもよい。ビーズ又は支持体は、多孔質であり得る。ビーズのサイズは、ナノメートル、例えば100nm～数ミリメートル、例えば1mmまでの範囲であり得る。特定の実施形態において、ビーズのサイズは、約0.2ミクロン～約200ミクロン、又は約0.5ミクロン～約5ミクロンの範囲である。いくつかの実施形態において、ビーズは、直径約1、1.5、2、2.5、2.8、3、3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9、9.5、10、10.5、15、又は20μmであり得る。特定の実施形態において、「ビーズ」固体支持体は、個々のビーズ又は複数のビーズを指す場合がある。いくつかの実施形態において、固体表面はナノ粒子である。特定の実施形態において、ナノ粒子のサイズは、直径が約1nm～約500nm、例えば、直径が約1nm～約20nmの間、約1nm～約50nmの間、約1nm～約100nmの間、約10nm～約50nmの間、約10nm～約100nmの間、約10nm～約200nmの間、約50nm～約100nmの間、約50nm～約150の間、約50nm～約200nmの間、約100nm～約200nmの間、又は約200nm～約500nmの間の範囲である。いくつかの実施形態において、ナノ粒子は、直径が約10nm、約50nm、約100nm、約150nm、約200nm、約300nm、又は約500nmであり得る。いくつかの実施形態において、ナノ粒子は、直径が約200nm未満である。

#### 【0044】

本明細書で使用される場合、「核酸分子」又は「ポリヌクレオチド」という用語は、3' - 5' ホスホジエステル結合によって連結されたデオキシリボヌクレオチド又はリボヌクレオチドを含む一本鎖又は二本鎖ポリヌクレオチド、及びポリヌクレオチド類似体を指す。核酸分子としては、DNA、RNA、及びcDNAが挙げられるが、これらに限定されない。ポリヌクレオチド類似体は、天然ポリヌクレオチドに見られる標準的なホスホジエ

ステル結合以外の骨格、及び、任意選択的に、リボース又はデオキシリボース以外の修飾された糖部分を有し得る。ポリヌクレオチド類似体は、ワトソククリック塩基対形成による、標準的なポリヌクレオチド塩基への水素結合が可能な塩基を含み、類似体骨格は、オリゴヌクレオチド類似体分子と標準的なポリヌクレオチドの塩基との間の配列特異的な様式でのそのような水素結合を可能にする方式で塩基を提示する。ポリヌクレオチド類似体の例としては、異種核酸(XNA)、架橋核酸(BNA)、グリコール核酸(GNA)、ペプチド核酸(PNA)、PNA、モルフォリノポリヌクレオチド、ロックされた核酸(LNA)、トレオース核酸(TNA)、2'-O-メチルポリヌクレオチド、2'-O-アルキルリボシル置換ポリヌクレオチド、ホスホロチオエートポリヌクレオチド、及びポロノホスフェートポリヌクレオチドが挙げられるが、これらに限定されない。ポリヌクレオチド類似体は、例えば、7-デアザプリン類似体、8-ハロプリン類似体、5-ハロピリミジン類似体、又はヒポキサンチン、ニトロアゾール、イソカルボスチリル類似体、アゾールカルボキサミド、及び芳香族トリアゾール類似体、又は親和性結合のためのビオチン部分などの追加の官能基(functionality)を有する塩基類似体を含む任意の塩基と対になることができるユニバーサル塩基類似体を含む、プリン又はピリミジン類似体を有し得る。いくつかの実施形態において、核酸分子又はオリゴヌクレオチドは、修飾されたオリゴヌクレオチドである。いくつかの実施形態において、核酸分子又はオリゴヌクレオチドは、偽相補的塩基を有するDNA、保護された塩基を有するDNA、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、PNA分子、若しくはモルフォリノDNA、又はそれらの組み合わせである。いくつかの実施形態において、核酸分子又はオリゴヌクレオチドは、骨格修飾、糖修飾、又は核酸塩基修飾される。いくつかの実施形態において、核酸分子又はオリゴヌクレオチドは、Allocなどの核酸塩基保護基、チランなどの求電子性保護基、アセチル保護基、ニトロベンジル保護基、スルホネート保護基、又は従来塩基不安定性保護基を有する。

#### 【0045】

本明細書で使用される場合、「核酸シーケンシング」は、核酸分子又は核酸分子のサンプル中のヌクレオチドの順序の決定を意味する。

#### 【0046】

本明細書で使用される場合、「次世代シーケンシング」は、数百万～数十億の分子の並行したシーケンシングを可能にするハイスループットシーケンシング方法を指す。次世代シーケンシング方法の例としては、合成によるシーケンシング、ライゲーションによるシーケンシング、ハイブリダイゼーションによるシーケンシング、コロニーシーケンシング、イオン半導体シーケンシング、及びパイロシーケンシングが挙げられる。プライマーを固体基質に付着させ、相補配列を核酸分子に付着させることにより、プライマーを介して核酸分子を固体基質にハイブリダイズさせることができ、次いで、ポリメラーゼを使用して増幅することにより、固体基質上の別個の領域に複数のコピーを生成することができる(これらのグループ分けは、ポリメラーゼコロニー又はコロニーと称される場合もある)。その結果、シーケンシングプロセス中に、特定の位置のヌクレオチドを複数回(例えば、数百回又は数千回)シーケンシングすることができる。このカバレッジの深度は、「ディープシーケンシング」と称される。ハイスループット核酸シーケンシング技術の例としては、Illumina、BGI、Qiagen、Thermo-Fisher、及びRocheによって提供されるプラットフォームを含み、パラレルビーズアレイ、合成によるシーケンシング、ライゲーションによるシーケンシング、キャピラリー電気泳動、電子マイクロチップ、「バイオチップ」、マイクロアレイ、並列マイクロチップ、及び単一分子アレイなどの形式が挙げられる(例えば、Service, Science(2006) 311:1544-1546を参照)。

#### 【0047】

本明細書で使用される場合、「単一分子シーケンシング」又は「第3世代シーケンシング」は、単一分子シーケンシング機器からの読み取りがDNAの単一分子のシーケンシングによって生成される次世代シーケンシング方法を指す。段階的アプローチでのシーケンシ

ングのために多数のDNA分子を並行してクローン化するために増幅に依存する次世代シーケンシング方法とは異なり、単一分子シーケンシングはDNAの単一分子にインターロゲートし (interrogates)、増幅又は同期を必要としない。単一分子シーケンシングには、各塩基の取り込み後にシーケンシング反応を一時停止する必要がある方法 (「ウォッシュアンドスキャン」サイクル) と、読み取りステップ間で停止する必要がない方法が含まれる。単一分子シーケンシング方法の例としては、単一分子リアルタイムシーケンシング (Pacific Biosciences)、ナノポアベースのシーケンシング (Oxford Nanopore)、デュプレックスインタラプテッドナノポアシーケンシング、及び高度な顕微鏡を使用するDNAの直接イメージングが挙げられる。

10

#### 【0048】

本明細書で使用される場合、ポリペプチドを「分析する」とは、ポリペプチドの成分の全部又は一部分を識別、検出、定量化、特徴付け、区別すること、又はそれらの組み合わせを意味する。例えば、ペプチド、ポリペプチド又はタンパク質の分析には、ペプチドのアミノ酸配列 (連続又は非連続) の全部又は一部分を決定することが含まれる。ポリペプチドの分析には、ポリペプチドの成分の部分的な識別も含まれる。例えば、ポリペプチドタンパク質配列中のアミノ酸の部分的な識別は、可能なアミノ酸のサブセットに属するものとしてタンパク質中のアミノ酸を識別することができる。分析は通常、nNTAAの分析から始まり、ペプチドの次のアミノ酸 (すなわち、n-1、n-2、n-3など) に進む。これは、nNTAAを脱離させ、それによってペプチドのn-1個のアミノ酸をN末端アミノ酸 (本明細書では「n-1NTAA」と称する) に変換することによって達成される。ペプチドの分析には、ペプチドの翻訳後修飾の存在及び頻度決定も含まれ得る。これには、ペプチドの翻訳後修飾の順序に関する情報が含まれても、含まれなくてもよい。ペプチドの分析には、ペプチド中のエピトープの存在及び頻度の決定も含まれ得る。これには、ペプチド内のエピトープの順序又は位置に関する情報が含まれても、含まれなくてもよい。ペプチドの分析は、異なるタイプの分析の組み合わせ、例えば、エピトープ情報、アミノ酸配列情報、翻訳後修飾情報、又はそれらの任意の組み合わせの取得を含み得る。

20

#### 【0049】

本明細書に記載される本発明の態様及び実施形態は、態様及び実施形態「からなる (consisting of)」、並びに/又は態様及び実施形態「から本質的になる (consisting essentially of)」を包含することが理解される。

30

#### 【0050】

本開示を通じて、本発明の様々な態様が、範囲形式で提示される。範囲形式での記載は、単に便宜上及び簡潔にするためのものであり、本発明の範囲に対する不動の限定として解釈されるべきではないことを理解されたい。したがって、範囲の記載は、全ての可能な部分範囲、及びその範囲内の個々の数値を具体的に開示するとみなすべきである。例えば、1~6などの範囲の記載は、1~3、1~4、1~5、2~4、2~6、3~6などの部分範囲、及びその範囲内の個々の数 (例えば、1、2、3、4、5、及び6) を具体的に開示するとみなすべきである。これは、範囲の幅に関係なく適用される。

40

#### 【0051】

本発明の他の目的、利点及び特徴は、添付の図面と結び付けられた以下の明細書から明らかになるであろう。

#### 【0052】

##### I. 連続エンコーディング

本明細書で提供されるのは、高分子、例えば、ペプチド、ポリペプチド、及びタンパク質の分析のための方法及びキットであり、これは、情報を記録タグに転送するステップを含む。この分析は、分子認識事象の核酸エンコーディングを用いる。いくつかの態様において、転送される情報は、高分子に結合するように構成されている結合剤に関する識別情報を含む。情報転送のために提供される方法は、(a) 高分子及び支持体に接合された関連

50

する記録タグを提供することと、(b)高分子を、高分子に結合することができる結合剤であって、結合剤に関する識別情報を有するコーディングタグを含む結合剤と接触させて、高分子と結合剤との間の結合を可能にすることと、(c)核酸接合試薬によって記録タグの5'末端をコーディングタグの3'末端に接合することと、(d)ポリメラーゼによって、コーディングタグを鋳型として使用して記録タグを伸長し、二本鎖伸長記録タグを生成することと、(e)二本鎖伸長記録タグを二本鎖核酸切断試薬で切断して、伸長記録タグ内に3'オーバーハングを生成することと、を含む。これらのステップを実行すると、情報が、コーディングタグから記録タグに転送されて、伸長記録タグを生成する。いくつかの実施形態において、ポリメラーゼ、核酸接合試薬、及び二本鎖核酸切断試薬は、混合物中で、又は同時に提供される。いくつかの態様において、ステップ(c)、(d)、及び(e)は、ワンポット反応として実行される。いくつかの他の態様において、ステップ(c)、(d)、及び(e)は、連続的かつ別々に実行される。場合によっては、各試薬を別々に提供するか、又は3つの酵素試薬のうち2つを同時に提供することができる。場合によっては、提供された方法の記録タグ及びコーディングタグは、核酸を含む。

10

20

30

40

50

#### 【0053】

ステップ(b)~(e)のうち1つ以上のサイクルを繰り返して、複数のコーディングタグから転送された情報を含む伸長記録タグを生成することができる。例えば、ステップ(b)、(c)、(d)、及び(e)は、周期的な方法で、連続1回以上繰り返すことができる。いくつかの実施形態において、生成された伸長記録タグは、核酸ヘアピンを含む。本明細書で提供される方法は、複数の結合剤及び複数の高分子を提供し、結合剤及び高分子が相互作用することを可能にすることを含み得る。いくつかの実施形態において、複数の結合剤が、混合物として各サイクルにおいて提供される。いくつかの実施形態において、本方法は、単一の高分子を単一の結合剤と接触させること、複数の高分子を単一の結合剤と接触させること、又は複数の高分子を複数の結合剤と接触させることを含む。

#### 【0054】

いくつかの実施形態において、本開示は、部分的に、タンパク質及びペプチドの特徴付け、定量化、及び/又はシーケンシングへの直接適用を伴う、情報転送を含む高分子を分析するための方法を提供する。いくつかの特定の実施形態において、分析のための高分子は、核酸ではない。いくつかの特定の実施形態において、結合剤は、核酸ではない。本明細書で提供されるのは、結合剤に関連付けられるか又は接合したコーディングタグから、分析される高分子(例えば、ポリペプチド)と関連付けられた記録タグに情報を転送する方法である。情報の転送は、二本鎖核酸切断試薬(例えば、制限酵素)によるライゲーション、伸長、及び切断を含む3つの部分からなる反応を使用して実行される。コーディングタグから転送される情報は、結合剤の同一性に関する識別情報を含み、それによって、結合剤によって結合された高分子又はその一部に関する情報を提供する。例えば、タンパク質/ポリペプチド/ペプチド高分子が結合剤によって結合されている場合、識別情報は、結合剤によって結合された1つ以上のアミノ酸の同一性に関する情報を含み得る。いくつかの実施形態において、結合剤によって結合された高分子(又はその一部)の同一性に関する情報は、該結合剤と関連付けられたコーディングタグからのものであり、記録タグに転送される。高分子分析アッセイは、結合剤の識別情報をコーディングタグから分析される高分子と関連付けられた記録タグに転送する1つ以上のサイクルを含み得る。分析のための高分子と関連付けられた最終伸長記録タグは、1つ以上のコーディングタグからの情報を含むことができる。複数のサイクルが実行される場合、結果として得られる伸長記録タグは、一連の結合事象及びコーディングタグからの複数の情報転送事象から構築された情報を含む。一般に、核酸接合試薬、ポリメラーゼ、及び二本鎖核酸切断試薬の活性を含む、記載された方法を使用した情報転送のための改善は、高分子分析アッセイに特定の利点を提供し得る。

#### 【0055】

いくつかの態様において、高分子を分析するために提供される方法で使用される連続エンコーディングシステムは、アッセイの全体的な設計にある特定の利点を提供する。特に、

核酸接合試薬、ポリメラーゼ、及び二本鎖核酸切断試薬によって実行される連続的なステップによっていくつかの利点が提供される。いくつかの実施形態において、方法は、混合された酵素が反応に提供され、その結果、それが1つのステップで実行され、かつ/又は反応がワンポット反応で実行される、反応システムを含むか、又は使用する。本システムの設計により、ステップ(c)、(d)、及び(e)のライゲーション、伸長、及び切断が、それぞれ、段階的又は連続的な方法で起こる。いくつかの他の実施形態において、ステップの段階的な性質は、ブロック基の使用によって、又はライゲーション、伸長、及び/又は切断が、前のステップの完了を条件とする方法で実行されるように、追加の要件を導入することによって導入され得る。ポリメラーゼ、核酸接合試薬、及び二本鎖核酸切断試薬の活性は、コーディングタグからの情報を記録タグに転送して、伸長記録タグを生成するための好適な条件で提供される。説明された方法を使用するいくつかの利点は、高い情報転送(エンコーディング)の成功、ステップワイズ反応のためのシンプルな設計、単一のステップで実行する/単一のポット反応として実行するオプション、スペーサーの必要性を低減すること、又はスペーサーの長さを短縮すること、システム内のDNA-DNA質相互作用を最小限に抑えること及び/又はシステム内のDNA-タンパク質相互作用を最小限に抑えることを含む。例えば、可塑性であり、相互作用のために露出させた塩基を提供し得る一本鎖核酸分子と比較して、本明細書で提供される二本鎖記録タグは、DNAがシステム内の他の成分(例えば、核酸、ペプチド、結合剤など)と相互作用するための相互作用が少ないことを示し得る。

10

#### 【0056】

いくつかの実施形態において、連続エンコーディングを使用して情報を転送するために提供される方法は、使用される核酸成分のフォーマットに関連する特定の利点を可能にする。例えば、ポリメラーゼプライミング部位を形成するためにスペーサー要素に依存する伸長ベースの情報転送方法において、スペーサーのサイズは、少なくとも6塩基対などの特定の長さであることが必要とされ得る。短いスペーサーは、アッセイにおける非特異的相互作用を回避するなどの利点を有し得る。場合によっては、配列濃縮は、繰り返しの要素が最小化又は排除された場合、よりまっすぐである。場合によっては、提供される方法は、伸長ベースの方法と比較すると、特異性、バイアス、安定性、及び効率性に関する特定の問題を回避する。情報転送のためのライゲーションベースの方法では、核酸の好適な末端を接合する効率及び情報を転送するための他の必要なステップとの複雑さが問題となり得る。いくつかの実施形態において、ポリメラーゼ、核酸接合試薬、及び二本鎖核酸切断試薬を使用する提供される方法は、単に伸長ベース又はライゲーションベースのシステムよりも利点を提供する。ポリメラーゼステップの後に生じる切断ステップを用いることにより、ポリメラーゼが「A」オーバーハングを残すAテーリング問題を回避することができる。二本鎖DNAの使用は、一本鎖DNA成分を利用する他の系において生じ得る可能性のあるDNA-DNA相互作用を最小限に抑える。場合によっては、提供された方法で使用することができるより短いスペーサーは、2bpのスペーサーなどの長さを縮小することができる。場合によっては、スペーサーの使用を完全に排除することができる。いくつかの態様において、連続エンコーディング方法は、互換性のないオーバーハングのために、1サイクルの情報転送の後に自己終了する。

20

30

40

#### 【0057】

いくつかの実施形態において、記録タグ及びコーディングタグの両方は、スペーサー配列を含む。例えば、スペーサーは、10塩基以下、9塩基以下、8塩基以下、7塩基以下、6塩基以下、5塩基以下、4塩基以下、3塩基以下、又は2塩基以下の核酸分子である。スペーサーは、サイクル特異的スペーサー又はサイクル交互スペーサーであり得る。情報を転送するためのシステムの特定の設計は、前のコーディングタグによって付加されたスペーサーが第2のコーディングタグのスペーサーの少なくとも一部と一致する場合にのみ、情報が第2のコーディングタグから伸長記録タグに転送されるように、スペーサーを利用することができる。このようにして、後続のエンコーディング事象は、前のエンコーディング事象に依存し、条件付きである。いくつかの他の実施形態において、記録タグも

50

コーディングもスペース配列を含まない。

【0058】

いくつかの実施形態において、サイクル内のいくつかのステップは、任意選択的に繰り返すことができる。例えば、結合剤を提供し、コーディングタグから情報を転送する1サイクル後、結合剤を除去することができ、結合剤を再び提供して、結合剤の再結合を可能にし、例えば、第1の情報転送が発生しなかった場合に、情報転送のための別の機会が発生することを可能にすることができる。いくつかの態様において、交互スペースを使用して提供される方法は、情報転送のための第2の機会を可能にするために、この繰り返しの適している。第1の結合及び情報転送事象が成功した場合、成功したコーディングタグ情報とともに転送されたスペースは、繰り返される情報転送事象が発生することを可能にし、第1の結合及び情報転送事象が成功しなかった場合、再結合は、既存の記録タグ上のスペースに一致するスペースを有するコーディングタグをもたらす、第2の結合は、再結合ステップの後に情報転送事象が発生することを可能にし、それによって、第1の逃した事象を補うためのメカニズムを提供する。

10

【0059】

提供される方法は、ポリメラーゼ、核酸接合試薬、及び二本鎖核酸切断試薬の使用を使用する。二本鎖核酸の伸長、ライゲーション、及び切断のための任意の好適な酵素を使用することができる（例えば、特許公報第CN104212791B号、同第CN101560538A号及び同第CN100510069C号を参照）。

20

【0060】

提供される方法において、コーディングタグから記録タグへの情報の転送は、ライゲーションステップから始まる。記録タグの5'末端は、核酸接合試薬によってコーディングタグの3'末端に接合される（例えば、ライゲーションされる）。いくつかの実施形態において、記録タグが核酸接合試薬によってコーディングタグに接合された後、結合剤は、結合したままではないか、又は高分子に結合したままである必要はない。

【0061】

いくつかの例では、核酸接合試薬は、化学ライゲーション試薬又は酵素ライゲーション試薬である。ライゲーションステップは、化学ライゲーション又は酵素ライゲーション（例えば、粘着末端ライゲーション、ssDNAライゲーションなどの一本鎖(ss)ライゲーション、又はそれらの任意の組み合わせ）によって実行されてもよい。いくつかの設計では、ライゲーションは、平滑末端ライゲーション及び/又は粘着末端ライゲーションであり得る。提供される連続エンコーディング方法は、コーディングタグから転送された情報を含む二本鎖構造を生成するように設計されており、この目的を達成する任意の接合方法を適用することができる。リガーゼの例としては、CV DNAリガーゼ、Circularリガーゼ、CircularリガーゼII、T4 DNAリガーゼ、T7 DNAリガーゼ、T3 DNAリガーゼ、Taq DNAリガーゼ、E. coli DNAリガーゼ、9°N DNAリガーゼが挙げられるが、これらに限定されない（例えば、米国特許公開第2014/0378315A1号又は米国特許第10,494,671B2号に参照）。いくつかの好ましい実施形態において、核酸接合試薬は、T4 DNAリガーゼなどの2つのDNAセグメントを接合する酵素（例えば、リガーゼ）を指す。

30

40

【0062】

接合又はライゲーションステップの後に、ポリメラーゼ媒介反応（例えば、一本鎖核酸又は二本鎖核酸のプライマー伸長）を含み得る伸長ステップが続く。記録タグの3'末端の伸長（又は既に伸長された記録タグ）は、ライゲーションされたコーディングタグを鋳型として使用して発生する。伸長ステップにより、コーディングタグから導入され、ライゲーションされた制限酵素/エンドヌクレアーゼ部位は、伸長記録タグ上の二本鎖となる。いくつかの実施形態において、プライマー伸長のために使用されるDNAポリメラーゼは、鎖置換活性を有し、3'-5'エキソヌクレアーゼ活性が制限されたか、又は欠如する。このようなポリメラーゼの多くの例のいくつかには、クレノウエキソ-(DNA Po

50

11のクレノウ断片)、T4 DNAポリメラーゼエキソ-、T7 DNAポリメラーゼエキソ(Sequenase 2.0)、Pfuエキソ-、Ventエキソ-、Deep Ventエキソ-、Bst DNAポリメラーゼラージ断片エキソ-、Bca Pol、9°N Pol、及びPhi29 Polエキソ-が含まれる。好ましい実施形態において、DNAポリメラーゼは、室温及び最大45 で活性である。別の実施形態において、好熱性ポリメラーゼの「ウォームスタート」バージョンは、ポリメラーゼが活性化され、約40 ~ 50 で使用されるように用いられる。例示的なウォームスタートポリメラーゼは、Bst 2.0ウォームスタートDNAポリメラーゼ(New England Biolabs)である。反応のための任意の添加剤及び緩衝液を含む、伸長反応のための好適な条件も提供され得る。

10

**【0063】**

伸長ステップに続いて、二本鎖核酸切断試薬によって認識可能な認識配列を含む二本鎖伸長記録タグが生成される。二本鎖伸長記録タグを切断すると、記録タグ上に3'オーバーハングが生成される。いくつかの態様において、二本鎖核酸切断試薬によって生成された伸長記録タグの3'オーバーハングを、ステップ(b)を繰り返したときに、第2のコーディングタグとハイブリダイズするために利用可能である。いくつかの態様において、二本鎖伸長記録タグを切断することは、伸長記録タグから結合剤を取り外すか、又は切り離す。いくつかの特定の場合において、二本鎖記録タグを切断すると、伸長記録タグから結合剤が放出される。いくつかの事例において、切断は、結合剤を、分析される高分子から解放可能にする。二本鎖核酸切断試薬は、制限酵素又は制限エンドヌクレアーゼであり得る。いくつかの実施形態において、制限酵素は、IIS型制限酵素である。IIS型制限酵素を使用する利点は、図1B及び実施例2(2ntスパーサーのみ)に示されるように、古典的な回文制限酵素と比較して、切断反応後に残される一本鎖ヌクレオチドが少なくなることを含む。いくつかの好ましい実施形態において、切断試薬による切断後の配列は、3'オーバーハングを残す。例えば、切断試薬は、酵素Bts I、Mva1269 I、Bsa I、BsmBIなどであり得る。IIS型制限酵素は、5'...GCAGTGN...3' / 3'...CGTCACNN...5'を含む塩基の配列を認識し得る。いくつかの特定の事例において、制限酵素は、Nb.Bts I又はBts I-v2又はその誘導体である。

20

**【0064】**

いくつかの他の実施形態において、二本鎖核酸切断試薬は、回文特異性を有し、その認識配列内で切断するIIP型制限酵素である。コーディングタグを記録タグにライゲーションした後、ライゲーション部位における制限酵素切断部位の潜在的な再現を回避するために(例えば、図1B、第1のステップを参照)、ライゲーション部位に隣接するコーディングタグ配列は、酵素切断部位とは異なるべきである。いくつかの好ましい実施形態において、IIP型制限酵素は、図1Bに示すように、切断後に記録タグ上に3'オーバーハングを生成する。いくつかの他の実施形態において、平滑末端を生成するIIP型制限酵素は、開示される方法において使用するために採用され得る。これらの実施形態において、コーディングタグの3'末端は、制限酵素切断後に生成された記録タグの平滑末端にライゲーションされるべきであり、ライゲーション部位に隣接するコーディングタグ配列は、制限酵素切断部位とは異なるべきである。

30

40

**【0065】**

いくつかの他の実施形態において、酵素又はエンドヌクレアーゼの組み合わせを使用して、伸長記録タグの両方の鎖の切断を達成することができる。例えば、ニッキング酵素又はエンドヌクレアーゼを使用して、伸長再コーディングタグ(recoding tag)の第1の鎖を切断する場合、第2の鎖を切断するための二次メカニズムを使用することができる。いくつかの特定の実施形態において、ニッキング酵素又はニッキングエンドヌクレアーゼは、使用されない。適切な制限酵素は、好ましい切断部位、消化位置と認識位置との間の距離、消化位置の精度、及び方法の他のステップのために必要な核酸末端を生成する能力のための考慮に基づいて選択することができる。

**【0066】**

50

いくつかの実施形態において、核酸接合試薬及びポリメラーゼは、同時に提供される。いくつかの実施形態において、ポリメラーゼ及び二本鎖核酸切断試薬が同時に提供される。

【0067】

提供される実施形態のいくつかにおいて、ステップ(a)、(b)、(c)、(d)、及び(e)が、連続的に実行される。特定の実施形態において、高分子(例えば、ペプチド)への結合剤の結合事象情報は、コーディングタグから固定化された高分子と関連付けられた記録タグに周期的に転送される。いくつかの実施形態において、1回以上繰り返されるステップは、(b)高分子を、高分子に結合することができる結合剤であって、結合剤に関する識別情報を有するコーディングタグを含む結合剤と接触させて、高分子と結合剤との間の結合を可能にすることと、(c)核酸接合試薬によって記録タグの5'末端をコーディングタグの3'末端に接合することと、(d)ポリメラーゼによって、コーディングタグを鋳型として使用して記録タグを伸長し、二本鎖伸長記録タグを生成することと、(e)二本鎖伸長記録タグを二本鎖核酸切断試薬で切断して、伸長記録タグ内に3'オーバーハングを生成することと、を含む。

10

【0068】

いくつかの実施形態において、この方法は、結合剤を除去することを更に含む。例えば、結合剤は、コーディングタグの情報を記録タグに転送した後に除去することができる。エンコーディングのサイクルが実行されると、ステップ(b)を繰り返す前に、結合剤を除去することができる。結合剤は、洗浄中に適切な条件(例えば、試薬及び温度)を提供することによって除去又は放出され得る。

20

【0069】

いくつかの実施形態において、この方法は、ステップ(b)を繰り返す前に、高分子の一部を除去することを更に含む。例えば、ポリペプチドが分析されている場合、この方法は、ステップ(b)を繰り返す前に、ポリペプチドのうちの一つ以上のアミノ酸(例えば、末端から)を除去することを含むことができる。場合によっては、ステップ(b)を繰り返す前に、N末端アミノ酸(NTAA)ポリペプチドをポリペプチドから除去して、ポリペプチドの新しいNTAAを露出させる。いくつかの実施形態において、高分子の除去された部分は、化学剤又は酵素剤で処理されるか、又はそれらによって修飾されている。場合によっては、ポリペプチドは、ポリペプチドの末端アミノ酸を修飾するための試薬で処理される。NTAAなどのポリペプチドの修飾は、ステップ(b)の前に実行することができる。場合によっては、NTAAなどのポリペプチドの修飾は、ステップ(e)の後に実行することができる。

30

【0070】

場合によっては、この方法は、ステップのうちの一つ以上の前、その間、又はその後の一つ以上の洗浄ステップを更に含む。例えば、結合剤がステップ(b)で提供された後、洗浄ステップは、酵素試薬(例えば、核酸接合試薬、ポリメラーゼ、二本鎖核酸切断試薬)のいずれかが導入される前に実行され得る。そのような洗浄ステップは、非特異的に結合した結合剤を除去し得る。洗浄ステップのストリンジェンシーは、結合剤の親和性に応じて調整され得る。場合によっては、ステップ(c)、(d)、及び(e)の間に洗浄ステップが必要とされないことが好ましい。いくつかの他の態様において、洗浄ステップは、ステップ(c)のライゲーションの後に実行される。ライゲーション及び洗浄後、ポリメラーゼ及び二本鎖核酸切断試薬を含む後続のステップを、一つのステップで提供することができる。いくつかの他の態様において、洗浄ステップは、ステップ(d)の伸長の後に実行される。場合によっては、洗浄ステップが、ステップ(c)、ステップ(d)、及び/又はステップ(e)の前に実行される。一つのステップでライゲーション及び伸長が生じた後、二本鎖核酸切断試薬が提供される前に、洗浄ステップが実行され得る。場合によっては、この方法は、ステップ(e)及び/又はステップ(b)が後続のサイクルで繰り返される前など、情報が再コーディングタグに転送された後に結合剤を除去することを更に含む。

40

【0071】

50

いくつかの実施形態において、最終情報転送サイクルは、伸長記録タグにキャッピング配列を提供するために実行され、任意選択的に、キャッピング配列は、増幅、シーケンシング、又はその両方のためのユニバーサルプライミング部位を含む。キャッピング配列は、高分子の普遍的な特徴に結合するように構成されている結合剤によって、伸長記録タグに提供され得る。このようにして、キャッピング配列を送達するための結合剤は、サンプル内の全て又は多くの高分子によって含有される標的部分に結合するように構成されている。いくつかの例では、普遍的な特徴は、ポリペプチドの化学修飾である。キャッピング配列は、ユニバーサルリバースプライミング部位など、伸長記録タグの分析に有用な配列を含み得る。

【0072】

10

#### A. 記録タグ

分析のための高分子（例えば、タンパク質又はポリペプチド）は、核酸分子又はオリゴヌクレオチドを含む記録タグで標識され得る。いくつかの態様において、サンプル内の複数の高分子には、記録タグが提供される。記録タグは、任意の好適な手段を使用して高分子に直接的又は間接的に関連付けられ得るか、又は付着され得る。いくつかの実施形態において、高分子は、1つ以上の記録タグと関連付けられ得る。いくつかの態様において、記録タグは、結合剤と接触する前に、高分子に直接的又は間接的に関連付けられ得るか、又は付着され得る。

【0073】

いくつかの実施形態において、少なくとも1つの記録タグは、高分子（例えば、ポリペプチド）と直接的又は間接的に関連付けられるか、又は共局在化する。ステップ（a）で高分子及び関連付けられた記録タグを提供することは、記録タグ及び任意の関連付けられた核酸を処理して、アッセイのための記録タグを接合、切断、又は他の方法で調製することを含み得る。いくつかの実施形態において、ステップ（a）は、ライゲーション及び/又は伸長を使用して、バーコード及び/又はUMIを記録タグに提供することを含む。いくつかの態様において、ステップ（a）は、制限酵素を使用して記録タグを切断し、3'オーバーハングを生成することを含む。例えば、記録タグの3'オーバーハングは、ポリメラーゼを使用した伸長及び/又は二本鎖核酸切断試薬による切断によって生成される。記録タグを調製及び提供するための例示的なワークフローが、図1Aに示される。

20

【0074】

特定の実施形態において、単一の記録タグは、N末端又はC末端アミノ酸への付着を介してなど、ポリペプチドに付着される。別の実施形態において、複数の記録タグは、ポリペプチドに、例えば、リジン残基又はペプチド骨格に付着される。いくつかの実施形態において、複数の記録タグで標識されたポリペプチドは、より小さなペプチドに断片化又は消化され、各ペプチドは、平均して1つの記録タグで標識される。

30

【0075】

記録タグは、DNA、RNA、若しくはPNA、gPNA、GNA、HNA、BNA、XNA、TNAを含むポリヌクレオチド類似体、又はそれらの組み合わせを含み得る。記録タグは一本鎖、又は部分的又は完全に二本鎖であり得る。いくつかの特定の実施形態において、ある特定の利点は、二本鎖領域を含む記録タグに付随する。例えば、利点は、システムの他の核酸成分とのDNA-DNA相互作用の低減であり得る。場合によっては、記録タグは、核酸ヘアピンを含む。記録タグは、平滑末端又は突出末端を有する場合がある。いくつかの特定の実施形態において、高分子と関連付けられた記録タグは、それが3'オーバーハングを有するように処理又は処理される（例えば、消化を介して）。特定の実施形態において、サンプル内の高分子の全ての又は実質的な量（例えば、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、又は100%）は、記録タグで標識される。他の実施形態において、サンプル内の高分子のサブセットは、記録タグで標識される。特定の実施形態において、サンプルからの高分子のサブセットは、記録タグでの標的化（分析物特異的）標識を受ける。例えば、タンパク質の標的化記録タグ標識は、標的タンパク質特異的結合

40

50

剤（例えば、抗体、アプタマーなど）を使用して達成され得る。いくつかの実施形態において、記録タグ（又はその一部）は、支持体上にサンプルを提供する前に高分子に付着される。いくつかの実施形態において、記録タグは、支持体上にサンプルを提供した後、高分子に付着される。

**【0076】**

いくつかの実施形態において、記録タグは、他の核酸成分を含み得る。いくつかの実施形態において、記録タグは、固有の分子識別子、コンパートメントタグ、パーティションバーコード、サンプルバーコード、画分バーコード、スペーサー配列、ユニバーサルプライミング部位、又はそれらの任意の組み合わせを含み得る。いくつかの実施形態において、記録タグは、記録タグの3'末端などのブロック基を含み得る。場合によっては、記録タグの3'末端がブロックされて、ポリメラーゼによる記録タグの伸長を防ぐ。

10

**【0077】**

いくつかの実施形態において、記録タグは、バーコードを識別するサンプルを含むことができる。サンプルバーコードは、単一の反応容器内のサンプルのセットの多重分析に有用であるか、又は単一の固体基質若しくは固体基質の集合（例えば、平面スライド、単一のチューブ若しくは容器に含まれるビーズの集団など）に固定化される。例えば、多くの異なるサンプルからの高分子を、サンプル固有のバーコードが付いた記録タグで標識し、次いで、全てのサンプルと一緒にプールした後、支持体への固定化、結合剤の周期的結合、及び記録タグ分析を行うことができる。あるいは、DNAコードライブラリーの作成後までサンプルを分離しておき、サンプルバーコードをDNAコードライブラリーのPCR増幅中に付着させ、次いで、ともに混合した後、シーケンシングを行うこともできる。このアプローチは、異なる豊度クラス（*abundance classes*）の分析物（例えば、タンパク質）をアッセイする際に役立ち得る。

20

**【0078】**

特定の実施形態において、記録タグは、固有の分子識別子（UMI）が関連付けられた各高分子（例えば、ポリペプチド）に固有の識別子タグを提供する、任意選択的なUMIを含む。UMIは、約3～約40個の塩基、約3～約30個の塩基、約3～約20個の塩基、又は約3～約10個の塩基、又は約3～約8個の塩基であり得る。いくつかの実施形態において、UMIは、約3の塩基、4個の塩基、5個の塩基、6個の塩基、7個の塩基、8個の塩基、9個の塩基、10個の塩基、11個の塩基、12個の塩基、13個の塩基、14個の塩基、15個の塩基、16個の塩基、17個の塩基、18個の塩基、19個の塩基、20個の塩基、25個の塩基、30個の塩基、35個の塩基、又は40個の塩基の長さである。UMIを使用して、複数の伸長記録タグからのシーケンシングデータをデコンボリューションして、個々の高分子からの配列リードを識別することができる。いくつかの実施形態において、高分子のライブラリー内で、各高分子は、単一の記録タグと関連付けられ、各記録タグは、固有のUMIを含む。他の実施形態において、記録タグの複数のコピーは、単一の高分子と関連付けられ、記録タグの各コピーは、同じUMIを含む。いくつかの実施形態において、UMIは、配列分析中にこれらの成分を区別することを容易にするために、スペーサー又はコーディングタグとは異なる塩基配列を有する。いくつかの実施形態において、UMIは、位置識別子としての機能を提供し得、また、高分子分析アッセイにおいて情報を提供し得る。例えば、UMIを使用して、同一系統であり、したがって同じ初期分子に由来する分子を識別することができる。いくつかの態様において、この情報を使用して、増幅の変動を補正する、並びに分析中のシーケンシングエラーを検出及び補正することができる。

30

40

**【0079】**

いくつかの実施形態において、記録タグは、スペーサーポリマーを含む。特定の実施形態において、記録タグは、その末端、例えば3'末端にスペーサーを含む。本明細書で使用される場合、記録タグとの関連におけるスペーサー配列への言及は、その同族結合剤と関連付けられたスペーサー配列と同一のスペーサー配列、又はその同族結合剤と関連付けられたスペーサー配列に相補的なスペーサー配列を含む。記録タグ上の末端、例えば3'ス

50

ペーサーは、（例えば、プライマー伸長又は粘着末端ライゲーションのための相補的スペーサー配列のアニーリングにより）最初の結合サイクル中に同族結合剤の識別情報をコーディングタグから記録タグに転送することを可能にする。一実施形態において、スペーサー配列は、約1～20個の塩基の長さ、約2～12個の塩基の長さ、又は5～10個の塩基の長さである。いくつかの事例において、スペーサー配列は、約2～5個の塩基の長さである。スペーサーの長さは、コーディングタグ情報を記録タグに転送するための温度及び反応条件などの要因に応じて変化し得る。

**【0080】**

スペーサー配列を使用するいくつかの実施形態において、ポリペプチドのライブラリーと関連付けられた記録タグは、共通のスペーサー配列を共有する。他の実施形態において、ポリペプチドのライブラリーと関連付けられた記録タグは、アダプター分子の結合サイクル特異的スペーサー配列に相補的な結合サイクル特異的スペーサー配列を有する。いくつかの態様において、記録タグ内のスペーサー配列は、記録タグ内の他の領域に対して最小限の相補性を有するように設計される。場合によっては、記録タグのスペーサー配列は、記録タグ又はコーディングタグに存在する、固有分子識別子、バーコード（例えば、コンパートメント、パーティション、サンプル、空間位置）、ユニバーサルプライマー配列、コーディングタグ配列、サイクル特異的配列などの成分に対して最小限の配列相補性を有する必要がある。いくつかの実施形態において、スペーサーは、使用のために選択された二本鎖核酸切断試薬（例えば、制限酵素）に基づいて設計される。

**【0081】**

特定の実施形態において、記録タグは、ユニバーサルプライミング部位、例えば、フォワード又は5'ユニバーサルプライミング部位を含む。ユニバーサルプライミング部位は、ライブラリー増幅反応のプライミング及び/又はシーケンシングに使用することができる核酸配列である。ユニバーサルプライミング部位には、PCR増幅用のプライミング部位、フローセル表面上の相補的オリゴヌクレオチドにアニーリングするフローセルアダプター配列（例えば、Illuminaの次世代シーケンシング）、シーケンシングプライミング部位、又はそれらの組み合わせが含まれ得るが、これらに限定されない。ユニバーサルプライミング部位は、約10塩基～約60塩基であり得る。いくつかの実施形態において、ユニバーサルプライミング部位は、Illumina P5プライマー（AATGATACGGCGACCAACCGA、配列番号1）又はIllumina P7プライマー（CAAGCAGAAGACGGCATACGAGAT、配列番号2）を含む。

**【0082】**

特定の実施形態において、記録タグは、コンパートメントタグを含む。いくつかの実施形態において、コンパートメントタグは、記録タグ内の成分である。いくつかの実施形態において、記録タグはまた、液滴、マイクロウェル、支持体上の物理領域などのコンパートメントに固有のバーコードが割り当てられているコンパートメントタグを表すバーコードを含むことができる。コンパートメントと特定のバーコードとの会合は、バーコード化されたビーズをコンパートメントにカプセル化することによって、例えば、バーコード化された液滴をコンパートメントに直接マージ又は追加することによって、バーコード化された試薬をコンパートメントに直接印刷又は注入することによってなど、任意の数の方法で達成することができる。コンパートメント内のバーコード試薬を使用して、コンパートメント内の高分子又はその断片にコンパートメント固有のバーコードを追加する。コンパートメントへのタンパク質分画に適用され、バーコードを使用して、分析されたペプチドをコンパートメント内のそれらの起源のタンパク質分子にマッピングすることができる。これは、タンパク質の識別を非常に容易にすることができる。コンパートメントバーコードを使用して、タンパク質複合体を識別することもできる。他の実施形態において、コンパートメントの集団のサブセットを表す複数のコンパートメントは、サブセットを表す固有のバーコードを割り当てられ得る。いくつかの実施形態において、記録タグは、分画内の高分子の識別情報を含む分画バーコードを含む。

**【0083】**

いくつかの実施形態において、タグ（例えば、コンパートメントタグ、パーティションバーコード、サンプルバーコード、画分バーコードなど）のうちの一つ以上は、複数のタンパク質複合体、タンパク質、又はポリペプチド上の内部アミノ酸、ペプチド骨格、又はN末端アミノ酸と反応することができる官能性部分を更に含む。いくつかの実施形態において、官能性部分は、クリック化学部分、アルデヒド、アジド/アルキン、又はマレイミド/チオール、又はエポキシド/求核試薬、逆電子要求ディールス・アルダー（iE D D A）基、又はシュタウディング反応のための部分である。いくつかの特定の実施形態において、複数のコンパートメントタグは、コンパートメントタグをコンパートメントに印刷、スポットティング、インクジェット、又はそれらの組み合わせによって形成される。いくつかの実施形態において、タグは、ポリペプチドに付着されて、ポリペプチド - ポリペプチド連結を介してタグを高分子に連結する。いくつかの実施形態において、タグ付着ポリペプチドは、タンパク質リガーゼ認識配列を含む。

10

## 【0084】

特定の実施形態において、ペプチド又はポリペプチド高分子は、親和性捕捉試薬（及び任意選択的に共有結合された）によって支持体に固定化され得、記録タグは、親和性捕捉試薬と直接関連付けられ、あるいは、高分子は、記録タグで支持体に直接固定化され得る。

## 【0085】

いくつかの実施形態において、高分子は、b a i t 核酸に付着されて、核酸 - 高分子キメラを形成する。固定化方法は、b a i t 核酸を支持体に付着した捕捉核酸にハイブリダイズすることによって、核酸 - 高分子キメラを支持体に近接させること、及び核酸 - 高分子キメラを固体支持体に共有結合的にカップリングすることを含み得る。場合によっては、核酸 - 高分子キメラは、リンカーを介してなど、固体支持体に間接的にカップリングされる。いくつかの実施形態において、複数の核酸 - 高分子キメラは、固体支持体上でカップリングされ、任意の隣接してカップリングされた核酸 - 高分子キメラは、約50nm以上の平均距離で互いに離間している。場合によっては、b a i t 核酸は、ユニバーサルプライミング部位又はその一部を含む。

20

## 【0086】

図1Aの例示的なフォーマットに示されるように、分析されるペプチドは、支持体上に固定化された捕捉核酸ヘアピンにハイブリダイズするb a i t 核酸に付着される。b a i t 核酸は、支持体に付着させるための反応性カップリング部分を含む捕捉核酸にライゲーションされる。いくつかの例では、b a i t 又は捕捉核酸、又は接合されたb a i t 及び捕捉核酸は、ポリペプチドに関する情報をコーディングタグから転送することができる記録タグとして機能し得る。

30

## 【0087】

図1Aは、凹部5'リン酸化末端を有する核酸ヘアピンを含むか、又はそれである捕捉核酸を使用する記録タグを調製するための例示的なステップを示す。図1Aの第2パネルにおいて、付着したペプチドとのb a i t 核酸は、捕捉核酸ヘアピンとハイブリダイズし、捕捉核酸ヘアピンにライゲーションされる。b a i t 核酸は、示されるような一つ以上のバーコードを含み得、また、制限酵素部位（又は部分的制限酵素部位）を含み得る。図1Aの第3のパネルにおいて、ポリメラーゼ反応を使用して、捕捉核酸ヘアピンの3'末端を伸長し、ペプチドが付着した二本鎖記録タグ構築物を生成する。b a i t 核酸を鋳型として使用して伸長が生じると、消化部位は二本鎖であり、切断のためのIIS型制限酵素によって認識することができ、切断は、凹部5'リン酸化末端を有する3'オーバーハング（2塩基対配列）を有する記録タグを生成する。一つ以上のバーコードを含むこの記録タグは、コーディングタグからの情報転送のために利用可能である。いくつかの実施形態において、支持体上の記録タグ及び固定化を用いた分析のための高分子の調製は、国際特許出願第PCT/US2020/27840号に記載されている方法を使用して実行することができる。高分子の調製及び/又は固定化は、情報転送ステップの前に、及び/又は情報転送ステップとは別に実行することができる。

40

50

## 【0088】

いくつかの実施形態において、記録タグを備えた高分子の密度又は数は、制御又は滴定される。いくつかの例では、サンプル内の記録タグの所望の間隔、密度、及び/又は量は、希釈された、又は制御された数の記録タグを提供することによって滴定され得る。いくつかの例では、記録タグを提供する、関連付ける、及び/又は付着させるときに、競合又は「ダミー」競合分子をスパイクすることによって、記録タグの所望の間隔、密度、及び/又は量を達成することができる。場合によっては、「ダミー」競合分子は、サンプル内の高分子に関連付けられている、又は付着されている記録タグと同じように反応するが、競合分子は、記録タグとして機能しない。いくつかの特定の例では、所望の密度が、サンプル内の付着に利用可能な1,000部位当たり1つの機能的記録タグである場合、1,000の「ダミー」競合分子ごとに1つの機能的記録タグにおけるスパイクを使用して、所望の間隔を達成する。いくつかの例では、機能的記録タグの比率は、競合分子の反応速度と比較した機能的記録タグの反応速度に基づいて調整される。

10

## 【0089】

いくつかの例では、記録タグによる高分子の標識は、標準的なアミンカップリング化学を使用して実行される。例えば、e-アミノ基（例えば、リジン残基）及びN末端アミノ基は、反応のpHに応じて、アミン反応性カップリング剤での標識に感受性であり得る（Mendoza et al., Mass Spectrom Rev (2009) 28(5): 785-815）。特定の実施形態において、記録タグは、反応性部分（例えば、固体表面、多機能性リンカー、又は高分子への共役のため）、リンカー、ユニバーサルプライミング配列、バーコード（例えば、コンパートメントタグ、パーティションバーコード、サンプルバーコード、画分バーコード、又はそれらの任意の組み合わせ）、任意選択的なUMI、及び情報転送を容易にするためのスペーサー（Sp）配列を含む。別の実施形態において、タンパク質は、最初にユニバーサルDNAタグで標識することができ、バーコード-Sp配列（サンプル、コンパートメント、スライド上の物理的位置などを表す）は、後で酵素的又は化学的カップリングステップを通じてタンパク質に付着される。ユニバーサルDNAタグは、タンパク質又はポリペプチド高分子を標識するために使用されるヌクレオチドの短い配列を含み、バーコードの結合点（例えば、コンパートメントタグ、記録タグなど）として使用することができる。例えば、記録タグは、その末端に、ユニバーサルDNAタグに相補的な配列を含み得る。特定の実施形態において、ユニバーサルDNAタグは、ユニバーサルプライミング配列である。標識タンパク質上のユニバーサルDNAタグが（例えば、ビーズに結合された）記録タグ内の相補的な配列にハイブリダイゼーションすると、アニーリングされたユニバーサルDNAタグは、プライマー伸長を介して伸長され得、記録タグ情報をDNAタグ付けタンパク質に転送する。特定の実施形態において、タンパク質は、ペプチドへのプロテアーゼ消化の前にユニバーサルDNAタグで標識される。次いで、消化物からの標識されたペプチドのユニバーサルDNAタグを、有益かつ効果的な記録タグに変換することができる。

20

30

## 【0090】

記録タグは、高分子、例えば、タンパク質上に存在する同族の反応性部分に対する反応性部分を含み得る（例えば、クリック化学標識、光親和性標識）。例えば、記録タグは、アルキン誘導体化タンパク質と相互作用するためのアジド部分を含み得るか、又は記録タグは、天然タンパク質などと相互作用するためのベンゾフェノンを含み得る。標的タンパク質特異的結合剤による標的タンパク質の結合時に、記録タグ及び標的タンパク質は、それらの対応する反応性を介してカップリングされる。標的タンパク質を記録タグで標識した後、標的タンパク質特異的結合剤を、標的タンパク質特異的結合剤に連結されたDNA捕捉プローブの消化によって除去してもよい。例えば、DNA捕捉プローブは、ウラシル塩基を含むように設計され得、次に、ウラシル特異的切除試薬（例えば、USER（商標））による消化の標的とされる。標的タンパク質特異的結合剤は、標的タンパク質から解離され得る。いくつかの実施形態において、ハイブリダイゼーション以外の他のタイプの連結を使用して、記録タグを高分子に連結することができる。好適なリンカーは、内部位置

40

50

における 3' 末端などの記録タグの様々な位置に、又は記録タグの 5' 末端に付着したリンカー内に付着させることができる。

【0091】

1つ以上のコーディングタグからの情報が記録タグに転送されて、伸長記録タグが生成される。いくつかの実施形態において、伸長記録タグは、ユニバーサルフォワード（又は 5'）プライミング配列、1つ以上のコーディングタグから転送された情報、及びスペース配列を含む。いくつかの実施形態において、伸長記録タグは、ユニバーサルフォワード（又は 5'）プライミング配列、いずれかの任意のバーコード及び/又は UMI（例えば、サンプルバーコード、パーティションバーコード、コンパートメントバーコード、又は任意のそれらの組み合わせ）、1つ以上のコーディングタグ、スペース配列、及びユニ

10

【0092】

B. 結合剤

本明細書に記載の方法は、分析する高分子（例えば、ポリペプチド、ペプチド、タンパク質）と相互作用するように構成された結合剤を使用する。アッセイは、複数の結合剤を複数の高分子に接触させることを含むことができる。いくつかの実施形態において、本方法は、単一の高分子を単一の結合剤と接触させること、複数の高分子を単一の結合剤と接触させること、又は複数の高分子を複数の結合剤と接触させることを含む。いくつかの実施形態において、複数の結合剤は、異なる標的部分に結合するように構成された結合剤の混合物を含む。

20

【0093】

結合剤は、ポリペプチドの成分又は機構に結合することができる任意の分子（例えば、ペプチド、ポリペプチド、タンパク質、核酸、炭水化物、小分子など）であり得る。結合剤は、天然に存在するか、合成的に産生されるか、又は組換え的に発現される分子であり得る。いくつかの実施形態において、結合剤を操作するために使用される足場は、任意の種類、例えば、ヒト、非ヒト、トランスジェニックに由来することができる。結合剤は、標的高分子又はモチーフの一部に結合し得る。結合剤は、ポリペプチドの単一の単量体若しくはサブユニット（例えば、単一アミノ酸）に結合し得るか、又はポリペプチドの複数の連結されたサブユニット（例えば、より長いポリペプチド分子のジペプチド、トリペプチド、又はより高次のペプチド）に結合し得る。

30

【0094】

いくつかの例では、結合剤は、抗体、抗原結合抗体断片、単ドメイン抗体（s d A b）、組換え重鎖のみの抗体（V H H）、一本鎖抗体（s c F v）、サメ由来可変ドメイン（v N A R）、F v、F a b、F a b'、F（a b'）<sub>2</sub>、線状抗体、ダイアボディ、アブタマー、ペプチド模倣分子、融合タンパク質、反応性若しくは非反応性小分子、又は合成分子を含む。

【0095】

特定の実施形態において、結合剤は、共有結合するように設計され得る。共有結合は、条件付きであるか、又は正しい部分に結合すると有利になるように設計され得る。例えば、標的及びその同族の結合剤は各々、標的特異的結合剤が標的に結合されると、カップリング反応が実行されて、2つの間の共有結合を作成するように、反応性基で修飾され得る。同族の反応性基を欠く他の場所への結合剤の非特異的結合は、共有結合をもたらさないであろう。いくつかの実施形態において、標的は、結合剤への共有結合を形成することができるリガンドを含む。いくつかの実施形態において、標的は、結合剤に共有結合することができるリガンド基を含む。結合剤とその標的との間の共有結合により、より厳密な洗浄を使用して、非特異的に結合している結合剤を除去することが可能になり得、したがって、アッセイの特異性が高まる。いくつかの実施形態において、この方法は、結合剤を高分子に接触させて非特異的に結合した結合剤を除去した後の洗浄ステップを含む。洗浄ステップのストリンジェンシーは、標的に対する結合剤の親和性及び/又は形成された複合体の強度及び安定性に応じて調整され得る。

40

50

## 【0096】

いくつかの実施形態において、結合剤は、高分子への結合剤の特異性を提供するように構成されている。特定の実施形態において、結合剤は、選択的結合剤であり得る。本明細書で使用される場合、選択的結合は、異なるリガンド（例えば、アミノ酸又はアミノ酸のクラス）への結合と比較して、特定のリガンド（例えば、アミノ酸又はアミノ酸のクラス）に優先的に結合する結合剤の能力を指す。選択性は一般に、結合剤との複合体におけるあるリガンドの別のリガンドによる置換の反応の平衡定数と称される。典型的には、そのような選択性は、リガンドの空間幾何学、及び/又はリガンドが、例えば、水素結合、疎水性結合、及びファンデルワールス力（非共有結合性相互作用）によって、又は結合剤への可逆的若しくは不可逆的な共有結合によって結合剤に結合する方法及び程度と関連付けられる。選択性は絶対的ではなく相対的である可能性があり、リガンド濃度を含む異なる要因が同じものに影響を与える可能性があることも理解する必要がある。したがって、一例では、結合剤は、20個の標準的なアミノ酸のうちの一つに選択的に結合する。いくつかの例では、結合剤は、N末端アミノ酸残基、C末端アミノ酸残基、又は内部アミノ酸残基に結合する。

10

## 【0097】

いくつかの実施形態において、結合剤は、部分的に特異的又は選択的である。いくつかの態様において、結合剤は、1つ以上のアミノ酸に優先的に結合する。いくつかの例では、結合剤は、20個の標準的なアミノ酸のうち2つ以上に結合し得るか、又は結合することができる。例えば、結合剤は、他のアミノ酸よりもアミノ酸A、C、及びGに優先的に結合し得る。いくつかの他の例では、結合剤は、複数のアミノ酸に選択的又は特異的に結合し得る。いくつかの態様において、結合剤はまた、末端アミノ酸からの第2、第3、第4、第5などの位置にある1つ以上のアミノ酸を優先し得る。場合によっては、結合剤は、特定の末端アミノ酸及び最後から2番目のアミノ酸に優先的に結合する。例えば、結合剤は、AA、AC、及びAGに優先的に結合することができるか、又は結合剤は、AA、CA、及びGAに優先的に結合することができる。いくつかの実施形態において、結合剤は、標的の位置のいくつか又は全てにおいて、標的結合の優先性の柔軟性及び可変性を示し得る。いくつかの例では、結合剤は、1つ以上の特定の標的末端アミノ酸を優先し得、最後から2番目の標的を柔軟に優先し得る。いくつかの他の例では、結合剤は、最後から2番目のアミノ酸位置にある1つ以上の特定の標的アミノ酸を優先し得、末端アミノ酸位置にある標的を柔軟に優先し得る。いくつかの実施形態において、結合剤は、末端アミノ酸及び高分子の他の成分を含む標的に対して選択的である。いくつかの例では、結合剤は、末端アミノ酸及びペプチド骨格の少なくとも一部を含む標的に対して選択的である。いくつかの特定の例では、結合剤は、末端アミノ酸及びアミドペプチド骨格を含む標的に対して選択的である。場合によっては、ペプチド骨格は、天然のペプチド骨格又は翻訳後修飾を含む。いくつかの実施形態において、結合剤は、アロステリック結合を示す。

20

30

## 【0098】

いくつかの実施形態において、この方法は、結合剤の混合物を高分子の混合物と接触させることを含み、選択性は、標的が露出される他の結合剤に対してのみ必要である。結合剤の選択性は、特定の分子に対して絶対的である必要はないが、分子の一部に対して絶対的である可能性があることも理解されたい。いくつかの例では、結合剤の選択性は、特定のアミノ酸に対して絶対的である必要はないが、極性若しくは非極性側鎖を有する、あるいは電気的に（正に又は負に）帯電した側鎖を有する、あるいは芳香族側鎖、又はいくつかの特定のクラス若しくはサイズの側鎖などを有するアミノ酸などのアミノ酸のクラスに対して選択的であり得る。いくつかの実施形態において、高分子の特徴又は成分に選択的に結合する結合剤の能力は、結合剤の結合能力を比較することによって特徴付けられる。例えば、結合剤の標的への結合能力は、異なる標的に結合する結合剤の結合能力と比較することができる。例えば、アミノ酸のクラスに対して選択的な結合剤を、アミノ酸の異なるクラスに対して選択的な結合剤と比較することができる。いくつかの例では、非極性側鎖に対して選択的な結合剤を、極性側鎖に対して選択的な結合剤と比較する。いくつかの実

40

50

施形態において、特徴、ペプチドの成分、又は1つ以上のアミノ酸に対して選択的な結合剤は、異なる特徴、ペプチドの成分、又は1つ以上のアミノ酸に対して選択的な結合剤と比較して、少なくとも1倍、少なくとも2倍、少なくとも5倍、少なくとも10倍、少なくとも50倍、少なくとも100倍、又は少なくとも500倍高い結合を示す。

#### 【0099】

特定の実施形態において、結合剤は、目的の高分子、例えば、ポリペプチドに対して高い親和性及び高い選択性を有する。特に、低いオフレートでの高い結合親和性は、アダプター分子のコーディングタグへのハイブリダイゼーションに有効である可能性がある。特定の実施形態において、結合剤は、約500 nM未満、200 nM未満、100 nM未満、50 nM未満、10 nM未満、5 nM未満、1 nM未満、0.5 nM未満、又は0.1 nM未満のK<sub>d</sub>を有する。特定の実施形態において、結合剤は、そのK<sub>d</sub>の1倍超、5倍超、10倍超、100倍超、又は1000倍超の濃度でポリペプチドに添加されて、結合を完了まで駆動する。例えば、単一のタンパク質分子への抗体の結合速度論は、Chang et al., J Immunol Methods (2012) 378 (1-2): 102-115に記載されている。

10

#### 【0100】

特定の実施形態において、結合剤は、ペプチド、介在アミノ酸、ジペプチド(2つのアミノ酸の配列)、トリペプチド(3つのアミノ酸の配列)、又はペプチド分子のより高次のペプチドの末端アミノ酸に結合し得る。いくつかの実施形態において、結合剤のライブラリー中の各結合剤は、特定のアミノ酸、例えば、20個の標準的な天然に存在するアミノ酸のうちの一つに選択的に結合する。標準的な天然アミノ酸としては、アラニン(A又はAla)、システイン(C又はCys)、アスパラギン酸(D又はAsp)、グルタミン酸(E又はGlu)、フェニルアラニン(F又はPhe)、グリシン(G又はGly)、ヒスチジン(H又はHis)、イソロイシン(I又はIle)、リジン(K又はLys)、ロイシン(L又はLeu)、メチオニン(M又はMet)、アスパラギン(N又はAsn)、プロリン(P又はPro)、グルタミン(Q又はGln)、アルギニン(R又はArg)、セリン(S又はSer)、スレオニン(T又はThr)、バリン(V又はVal)、トリプトファン(W又はTrp)、及びチロシン(Y又はTyr)が挙げられる。いくつかの実施形態において、結合剤は、未修飾又は天然の(例えば、天然の)アミノ酸に結合する。いくつかの例では、結合剤は、未修飾又は天然のジペプチド(2つのアミノ酸の配列)、トリペプチド(3つのアミノ酸の配列)、又はペプチド分子のより高次のペプチドに結合する。結合剤は、天然若しくは未修飾のN末端アミノ酸(NTAA)に対する高い親和性、天然若しくは未修飾のNTAAに対する高い特異性、又はその両方のために設計され得る。いくつかの実施形態において、結合剤は、ファージディスプレイを使用する有望な親和性足場の指向性進化を通じて開発することができる。

20

30

#### 【0101】

特定の実施形態において、結合剤は、アミノ酸の翻訳後修飾に結合し得る。いくつかの実施形態において、ペプチドは、1つ以上の翻訳後修飾を含み、これは、同じであっても異なってもよい。ペプチドのNTAA、CTAA、介在アミノ酸、又はそれらの組み合わせは、翻訳後修飾され得る。アミノ酸に対する翻訳後修飾の例としては、アシル化、アセチル化、アルキル化(メチル化を含む)、ビオチン化、ブチリル化、カルバミル化、カルボニル化、脱アミド化、脱イミノ化(deiminiation)、ジフタミド形成、ジスルフィド架橋形成、脱離(eliminylation)、フラビン付着、ホルミル化、カルボキシル化、グルタミル化、グリシル化、グリコシル化、グリピエーション、ヘムC付着、ヒドロキシル化、ヒプシン形成、ヨウ素化、イソプレニル化、脂質化、リポイル化、マロニル化、メチル化、ミリストリル化(myristoylation)、酸化、パルミトイル化、ペグ化、ホスホパンテテイン化、リン酸化、プレニル化、プロピオニル化、レチニリデンシフ塩基形成、S-グルタチオン付加、S-ニトロシル化、S-スルフェニル化、セレン化、スクシニル化、スルフィン化、ユビキチン化、及びC末端アミド化が挙げられるが、これらに限定されない(Seo and Lee, 2004, J

40

50

. Biochem. Mol. Biol. 37: 35 - 44 も参照)。

【0102】

特定の実施形態において、レクチンは、タンパク質、ポリペプチド、又はペプチドのグリコシル化状態を検出するための結合剤として使用される。レクチンは、遊離炭水化物又は糖タンパク質のグリカンエピトープを選択的に認識することができる、炭水化物結合タンパク質である。様々なグリコシル化状態を認識するレクチン(例えば、コアフコース、シアル酸、N-アセチル-D-ラクトサミン、マンノース、N-アセチル-グルコサミン)の一覧には、A、AAA、AAL、ABA、ACA、ACG、ACL、AOL、ASA、BanLec、BC2L-A、BC2LCN、BPA、BPL、Calsepa、CGL2、CNL、Con、ConA、DBA、Discoidin、DSA、ECA、EEL、F17AG、Gal1、Gal1-S、Gal2、Gal3、Gal3C-S、Gal7-S、Gal9、GNA、GRFT、GS-I、GS-II、GSL-I、GSL-II、HHL、HIHA、HPA、I、II、Jacalin、LBA、LCA、LEA、LEL、Lentil、Lotus、LSL-N、LTL、MAA、MAH、MAL\_I、Malectin、MOA、MPA、MPL、NPA、Oryzata、PA-IIL、PA-IL、PALa、PHA-E、PHA-L、PHA-P、PHAE、PHAL、PNA、PPL、PSA、PSL1a、PTL、PTL-I、PWM、RCA120、RS-Fuc、SAMB、SBA、SJA、SNA、SNA-I、SNA-II、SSA、STL、TJA-I、TJA-II、TxLCI、UDA、UEA-I、UEA-II、VFA、VVA、WFA、WGAが含まれる(Zhang et al., 2016, MABS 8: 524 - 535を参照)。

【0103】

いくつかの実施形態において、結合剤は、天然又は非修飾又は非標識又は天然の末端アミノ酸に結合し得る。いくつかの例では、結合剤は、未修飾又は天然のジペプチド(2つのアミノ酸の配列)、トリペプチド(3つのアミノ酸の配列)、又はペプチド分子のより高次のペプチドに結合する。結合剤は、修飾されたNTAAに対する高い親和性、修飾されたNTAAに対する高い特異性、又はその両方のために操作され得る。いくつかの実施形態において、結合剤は、ファージディスプレイを使用する有望な親和性足場の指向性進化を通じて開発することができる。

【0104】

タンパク質/酵素足場の指向性進化を使用して、N末端標識の文脈でN末端アミノ酸を認識する、より高い親和性、より高い特異性の結合剤を生成することができる。例では、Havranakら(米国特許公開第2014/0273004号)は、アミノアシルtRNAシンテターゼ(aARS)を特定のNTAA結合剤として操作することについて説明している。aARSのアミノ酸結合ポケットは、同族のアミノ酸に結合する固有の能力を有しているが、一般的に低い結合親和性及び特異性を示す。更に、これらの天然アミノ酸結合剤は、N末端標識を認識しない。aARS足場の指向性進化を使用して、N末端標識の文脈でN末端アミノ酸を認識する、より高い親和性、より高い特異性の結合剤を生成することができる。

【0105】

特定の実施形態において、結合剤は、修飾又は標識された末端アミノ酸(例えば、官能化又は修飾されたNTAA)に結合し得る。いくつかの実施形態において、結合剤は、化学的又は酵素的に修飾された末端アミノ酸に結合し得る。修飾又は標識されたNTAAは、フェニルイソチオシアネート、PITC、1-フルオロ-2,4-ジニトロベンゼン(サンガー試薬、DNFB)、塩化ベンジルオキシカルボニル又は塩化カルボベンゾキシ(Cbz-Cl)、N-(ベンジルオキシカルボニルオキシ)スクシンイミド(Cbz-OSu若しくはCbz-O-NHS)、ダンシルクロリド(DNS-Cl、若しくは1-ジメチルアミノナフタレン-5-スルホニルクロリド)、4-スルホニル-2-ニトロフルオロベンゼン(SNFB)、N-アセチル-イサト酸無水物、イサト酸無水物、2-ピリジンカルボキサルデヒド、2-ホルミルフェニルボロン酸、2-アセチルフェニルボロン酸

、1-フルオロ-2,4-ジニトロベンゼン、無水コハク酸、4-クロロ-7-ニトロベンゾフラザン、ペンタフルオロフェニルイソチオシアネート、4-(トリフルオロメトキシ)-フェニルイソチオシアネート、4-(トリフルオロメチル)-フェニルイソチオシアネート、3-(カルボン酸)-フェニルイソチオシアネート、3-(トリフルオロメチル)-フェニルイソチオシアネート、1-ナフチルイソチオシアネート、N-ニトロイミダゾール-1-カルボキシミドアミド、N,N,A-ビス(ピバロイル)-1H-ピラゾール-1-カルボキサミジン、N,N,A-ビス(ベンジルオキシカルボニル)-1H-ピラゾール-1-カルボキサミジン、アセチル化試薬、グアニジニル化試薬、チオアシル化試薬、チオアセチル化試薬、又はチオベンジル化試薬、又は二複素環式メタンイミン試薬で官能化されるものであり得る。いくつかの例では、結合剤は、試薬と接触させることによって、又は国際特許公開第2019/089846号に記載されている方法を使用することによって標識されたアミノ酸に結合する。場合によっては、結合剤は、アミン修飾試薬によって標識されたアミノ酸に結合する。

#### 【0106】

結合剤は、ペプチド、ポリペプチド、又はタンパク質分子のN末端ペプチド、C末端ペプチド、又は介在ペプチドに結合し得る。結合剤は、ペプチド分子のN末端アミノ酸、C末端アミノ酸、又は介在アミノ酸に結合し得る。結合剤は、N末端又はC末端のジアミノ酸部分に結合し得る。N末端ジアミノ酸は、N末端アミノ酸及び最後から2番目のN末端アミノ酸で構成されている。C末端ジアミノ酸は、C末端についても同様に定義されている。いくつかの実施形態において、結合剤は、化学的に修飾されたN末端アミノ酸残基又は化学的に修飾されたC末端アミノ酸残基に結合する。ペプチドの小さなN末端アミノ酸(NTAA)に対する結合剤の親和性を高めるために、NTAAをジニトロフェノール(DNP)などの「免疫原性」ハプテンで修飾することができる。これは、サンガーの試薬であるジニトロフルオロベンゼン(DNFB)を使用したサイクルシーケンシングアプローチで実装することができ、これは、DNP基をNTAAのアミン基に付着させる。市販の抗DNP抗体は、低nM範囲(約8nM、LO-DNP-2)の親和性を有し(Bilgicer et al., J Am Chem Soc (2009) 131(26): 9361-9367)、したがって、(DNFBを介して)DNPで修飾された多数のNTAAに対して高親和性NTAA結合剤を操作し、同時に特定のNTAAに対して良好な結合選択性を達成することが可能であるのは当然のことである。別の例では、NTAAは、4-スルホニル-2-ニトロフルオロベンゼン(SNFB)を使用して、スルホニルニトロフェノール(SNP)で修飾され得る。同様の親和性の増強は、アセチル基又はアミジニル(グアニジニル)基などの代替のNTAA修飾剤でも達成され得る。

#### 【0107】

特定の実施形態において、結合剤は、アプタマー(例えば、ペプチドアプタマー、DNAアプタマー、又はRNAアプタマー)、ペプトイド、抗体若しくはその特異的結合断片、アミノ酸結合タンパク質若しくは酵素、抗体結合断片、抗体模倣物、ペプチド、ペプチド模倣物、タンパク質、あるいはポリヌクレオチド(例えば、DNA、RNA、ペプチド核酸(PNA)、gPNA、架橋核酸(BNA)、ゼノ核酸(XNA)、グリセロール核酸(GNA)、若しくはトレース核酸(TNA)、又はそのバリエーション)であり得る。

#### 【0108】

本明細書で使用される場合、抗体及び抗体(複数)という用語は、無傷の抗体分子、例えば限定されないが、免疫グロブリンA、免疫グロブリンG、免疫グロブリンD、免疫グロブリンE、及び免疫グロブリンMだけでなく、少なくとも1つのエピトープに免疫特異的に結合する抗体分子又はその一部の任意の免疫反応性成分を含むように、広い意味で使用される。抗体は、天然に存在し得るか、合成的に産生され得るか、又は組換え的に発現され得る。抗体は、融合タンパク質であり得る。抗体は、抗体模倣物であり得る。抗体の例としては、Fab断片、Fab'断片、F(ab')<sub>2</sub>断片、一本鎖抗体断片(scFv)、ミニ抗体、ダイアボディ、架橋抗体断片、Affibody(商標)、ナノボディ、単一ドメイン抗体、DVD-Ig分子、アルファボディ、アフィマー、アフィチン、サイク

ロチド、分子などが挙げられるが、これらに限定されない。抗体工学又はタンパク質工学技法を使用して誘導された免疫反応性産物もまた、明示的に抗体という用語の意味の範囲内にある。関連するプロトコルを含む、抗体及び/又はタンパク質工学の詳細な説明は、とりわけ、J. Maynard and G. Georgiou, 2000, Ann. Rev. Biomed. Eng. 2: 339 - 76、Antibody Engineering, R. Kontermann and S. Dubel, eds., Springer Lab Manual, Springer Verlag (2001)、米国特許第5,831,012号、及びS. Paul, Antibody Engineering Protocols, Humana Press (1995)において見出すことができる。

10

## 【0109】

抗体と同様に、高分子、例えば、ペプチド又はポリペプチドを特異的に認識する核酸及びペプチドアプタマーは、既知の方法を使用して産生することができる。アプタマーは、非常に特異的で立体配座依存的な方法で、典型的には非常に高い親和性で標的分子に結合するが、必要に応じて、結合親和性の低いアプタマーを選択することもできる。アプタマーは、メチル基又はヒドロキシル基の有無などの非常に小さな構造の違いに基づいて標的を区別することが示されており、特定のアプタマーは、D-エナンチオマーとL-エナンチオマーとを区別することができる。薬物、金属イオン、及び有機色素、ペプチド、ビオチン、及びタンパク質(ストレプトアビジン、VEGF、及びウイルスタンパク質を含むが、これらに限定されない)を含む、小分子標的に結合するアプタマーが得られている。アプタマーは、ビオチン化、フルオレセイン標識後、並びにガラス表面及びマイクロスフェアに付着したときに、機能的活性を保持することが示されている(例えば、Jayasena, 1999, Clin Chem 45: 1628 - 50、Kusser 2000, J. Biotechnol. 74: 27 - 39、Colas, 2000, Curr Opin Chem Biol 4: 54 - 9を参照)。アルギニン及びAMPに特異的に結合するアプタマーも同様に記載されている(Patel and Suri, 2000, J. Biotech. 74: 39 - 60を参照)。特定のアミノ酸に結合するオリゴヌクレオチドアプタマーは、Gold et al. (1995, Ann. Rev. Biochem. 64: 763 - 97)に開示されている。アミノ酸に結合するRNAアプタマーも記載されている(Ames and Breaker, 2011, RNA Biol. 8; 82 - 89、Mannironi et al., 2000, RNA 6: 520 - 27、Famulok, 1994, J. Am. Chem. Soc. 116: 1698 - 1706)。

20

30

## 【0110】

結合剤は、遺伝子工学によって天然に存在するか、又は合成的に産生されたタンパク質を修飾して、1つ以上の突然変異をアミノ酸配列に導入し、ポリペプチドの特定の成分又は機構に結合する操作されたタンパク質を産生することによって作製することができる(例えば、NTAA、CTAA、又は翻訳後修飾されたアミノ酸若しくはペプチド)。例えば、エキソペプチダーゼ(例えば、アミノペプチダーゼ、カルボキシペプチダーゼ、ジペプチジルペプチダーゼ、ジペプチジルアミノペプチダーゼ)、エキソプロテアーゼ、突然変異型エキソプロテアーゼ、突然変異型アンチカリン、突然変異型ClpS、抗体、又はtRNAシンテターゼを修飾して、特定のNTAAに選択的に結合する結合剤を作成することができる。別の例では、カルボキシペプチダーゼを修飾して、特定のCTAAに選択的に結合する結合剤を作成することができる。結合剤はまた、修飾されたNTAA又は修飾されたCTAA、例えば、翻訳後修飾を有するもの(例えば、リン酸化NTAA若しくはリン酸化CTAA)又は標識(例えば、PTC、1-フルオロ-2,4-ジニトロベンゼン(サンガー試薬、DNFBを使用)、塩化ダンシルク(DNS-Cl、若しくは1-ジメチルアミノナフタレン-5-スルホニルクロリドを使用)、又はチオアシル化試薬、チオアセチル化試薬、アセチル化試薬、アミド化(グアニジニル化)試薬、又はチオベンジル化試薬を使用する)で修飾されたものに特異的に結合するように設計又は修飾され、利

40

50

用され得る。タンパク質の定向進化の戦略は、当技術分野で知られており（例えば、Yuan et al., 2005, Microbiol. Mol. Biol. Rev. 69: 373 - 392）、ファージディスプレイ、リボソームディスプレイ、mRNAディスプレイ、CISディスプレイ、CADディスプレイ、エマルジョン、細胞表面ディスプレイ法、酵母表面ディスプレイ、細菌表面ディスプレイなどが含まれる。

【0111】

別の例では、高度に選択的に操作されたClpSもまた、文献に記載されている。Emiliらは、アスパラギン酸、アルギニン、トリプトファン、及びロイシン残基のNTAAに選択的に結合する能力を持つ4つの異なるバリエーションをもたらし、ファージディスプレイを介したE. coli ClpSタンパク質の定向進化について説明している（米国特許第9,566,335号、その全体が参照により組み込まれる）。一実施形態において、結合剤の結合部分は、天然のN末端タンパク質の認識及び結合に關与するアダプタータンパク質の進化的に保存されたClpSファミリーのメンバー又はそのバリエーションを含む。（例えば、Schuenemann et al., (2009) EMBO Reports 10(5)、Roman-Hernandez et al., (2009) PNAS 106(22): 8888 - 93、Guo et al., (2002) JBC 277(48): 46753 - 62、Wang et al., (2008) Molecular Cell 32: 406 - 414を参照）。いくつかの実施形態において、Schuenemannらで識別されたClpS疎水性結合ポケットに対応するアミノ酸残基は、所望の選択性を有する結合部分を生成するために修飾される。

10

20

【0112】

一実施形態において、結合部分は、UBRボックス認識配列ファミリーのメンバー、又はUBRボックス認識配列ファミリーのバリエーションを含む。UBR認識ボックスは、Tasaki et al., (2009), JBC 284(3): 1884 - 95に記載されている。例えば、結合部分は、UBR1、UBR2、又はそれらの変異体、バリエーション、又は相同体を含み得る。

【0113】

特定の実施形態において、結合剤は、結合部分に加えて、蛍光標識などの1つ以上の検出可能な標識を更に含む。いくつかの実施形態において、結合剤は、コーディングタグなどのポリヌクレオチドを含まない。任意選択的に、結合剤は、合成抗体又は天然抗体を含む。いくつかの実施形態において、結合剤は、アダプターを含む。一実施形態において、結合剤は、E. coli ClpS結合ポリペプチドのバリエーションなどのアダプタータンパク質のClpSファミリーの修飾されたメンバーなどのポリペプチド、及び検出可能な標識を含む。一実施形態において、検出可能な標識は、光学的に検出可能である。いくつかの実施形態において、検出可能な標識は、蛍光部分、色分けされたナノ粒子、量子ドット、又はそれらの任意の組み合わせを含む。一実施形態において、標識は、Fluosphere（商標）、ナイルレッド、フルオレセイン、ローダミンなどのコア色素、TAMRAなどの誘導体化ローダミン色素、蛍光体、ポリメタジン色素、蛍光ホスホルアミダイト、TEXAS RED、緑色蛍光タンパク質、アクリジン、シアニン、シアニン5色素、シアニン3色素、5-(2'-アミノエチル)-アミノナフタレン-1-スルホン酸(EDANS)、BODIPY、120 ALEXA、又は前述のうちのいずれかの誘導体又は修飾を含む、ポリスチレン色素を含む。一実施形態において、検出可能な標識は、高いシグナル対雑音比で、固有かつ容易に検出可能な波長で大量のシグナル（光子など）を産生しながら、光退色に対して耐性がある。

30

40

【0114】

特定の実施形態において、アンチカリンは、標識されたNTAA（例えば、PTC、修飾PTC、Cbz、DNP、SNP、アセチル、グアニジニル、アミノグアニジニル、二複素環メタニミンなど）に対する高い親和性及び高い特異性の両方のために操作される。特定の種類のアンチカリン足場は、バレル構造により、単一アミノ酸を結合するのに好適な形状を有する。N末端アミノ酸（修飾の有無にかかわらず）は、この「バレル」バケ

50

ットに収まり、認識される可能性がある。操作された新規結合活性を有する高親和性アンチカリンが記載されている (Skerra, 2008, F E B S J . 2 7 5 : 2 6 7 7 - 2 6 8 3 によるレビュー)。例えば、フルオレセイン及びジゴキシゲニンへの高親和性結合 (低 n M) を持つアンチカリンが設計されている (Gebauer et al., 2012, Methods Enzymol 503: 157 - 188.)。新しい結合機能のための代替足場のエンジニアリングもまた、Bantaら (2013, Annu. Rev. Biomed. Eng. 15: 93 - 113) によってレビューされている。

#### 【0115】

いくつかの実施形態において、結合剤は、直接的又は間接的に、多量体化ドメインに連結される。したがって、1つ以上の結合剤を含む単量体、二量体、及び高次 (例えば、3、4、5、又はそれ以上) の多量体ポリペプチドが本明細書で提供される。いくつかの特定の実施形態において、結合剤は、二量体である。いくつかの例では、本発明の2つのポリペプチドは、互いに共有結合又は非共有結合で付着して、二量体を形成することができる。

10

#### 【0116】

いくつかの実施形態において、結合剤は、生物学的、天然に存在する、非天然に存在する、又は合成供給源に由来する。いくつかの例では、結合剤は、新規タンパク質設計 (Huang et al., (2016) 537 (7620): 320 - 327) に由来する。いくつかの例では、結合剤は、第1の原理から設計された構造、配列、及び/又は活性を有する。

20

#### 【0117】

いくつかの実施形態において、修飾されたC末端アミノ酸 (CTAA) に選択的に結合する結合剤を利用することができる。カルボキシペプチダーゼは、遊離カルボキシル基を含む末端アミノ酸を切断/脱離するプロテアーゼである。多くのカルボキシペプチダーゼは、アミノ酸の優先性を示し、例えば、カルボキシペプチダーゼBは、アルギニン及びリジンなどの塩基性アミノ酸で優先的に切断する。カルボキシペプチダーゼを修飾して、特定のアミノ酸に選択的に結合する結合剤を作成することができる。いくつかの実施形態において、カルボキシペプチダーゼは、CTAAの修飾部分及び炭素R基の両方に選択的に結合するように操作され得る。したがって、操作されたカルボキシペプチダーゼは、C末端標識の文脈で標準的なアミノ酸を表す20の異なるCTAAを特異的に認識することができる。ペプチドのC末端からの段階的分解の制御は、標識の存在下でのみ活性 (例えば、結合活性又は触媒活性) である操作されたカルボキシペプチダーゼを使用することによって達成される。一例では、CTAAは、パラニトロアニリド又は7-アミノ-4-メチルクマリニル基によって修飾され得る。

30

#### 【0118】

本明細書に記載される方法で使用するための結合剤を生成するために操作され得る他の潜在的足場としては、アンチカリン、リポカリン、アミノ酸tRNAシンテターゼ (aaRS)、ClpS、Affilin (登録商標)、Adnectin (商標)、T細胞受容体、ジnkフィンガータンパク質、チオレドキシン、GST A1-1、DARPin、アフィマー、アフィチン、アルファボディ、アビマー、モノボディ、抗体、シングルドメイン抗体、ナノボディ、EETI-II、HPSTI、イントラボディ、PHD-フィンガー、V(NAR)LDTI、エピボディ、Ig(NAR)、ノッティン、マキシボディ、マイクロボディ、ネオカルジノスタチン、pVIIII、テンダミスタット、VLR、タンパク質A足場、MTI-II、エコチン、GCN4、Im9、クニツドメイン、PB P、トランスボディ、テトラネクチン、WWドメイン、CBM4-2、DX-88、GF P、iMab、Ldl受容体ドメインA、Min-23、PDZドメイン、鳥類臍臓ポリペプチド、カリプトトキシソ/10Fn3、ドメイン抗体 (Dab)、a2p8アンキリンリピート、昆虫防御Aペプチド、設計されたARタンパク質、C型レクチンドメイン、ブドウ球菌ヌクレアーゼ、Src相同性ドメイン3 (SH3)、又はSrc相同性ドメイン2 (SH2) が挙げられる。例えば、El-Gebali et al., (2019

40

50

) Nucleic Acids Research 47: D427 - D432 and Finn et al., (2013) Nucleic Acids Res. 42 (Database issue): D222 - D230を参照。いくつかの実施形態において、結合剤は、1つ以上のアミノ酸（例えば、アミノペプチダーゼ）に結合する酵素に由来する。特定の実施形態において、結合剤は、アンチカリン又はClpプロテアーゼアダプタータンパク質（ClpS）に由来することができる。

#### 【0119】

場合によっては、結合剤は、翻訳後修飾されたアミノ酸に結合し得る。いくつかの実施形態において、内部の翻訳後修飾アミノ酸（例えば、リン酸化、グリコシル化、スクシニル化、ユビキチン化、S-ニトロシル化、メチル化、N-アセチル化、脂質化など）の検出は、末端アミノ酸（例えば、NTAA又はCTAA）の検出及び脱離の前に達成される。一例では、ペプチドは、PTM修飾のために結合剤と接触され、対応するコーディングタグからの情報は、固定化されたペプチドと関連付けられた記録タグに転送される。アミノ酸修飾に関連する情報の検出及び転送が完了すると、N末端又はC末端分解法を使用して一次アミノ酸配列のコーディングタグ情報を検出及び転送する前に、PTM修飾基を除去することができる。したがって、結果として得られる伸長核酸は、一次アミノ酸配列情報とともに、ペプチド配列における翻訳後修飾の存在を示すが、順番ではない。

10

#### 【0120】

いくつかの実施形態において、内部の翻訳後修飾アミノ酸の検出は、一次アミノ酸配列の検出と同時に行為得る。一例では、NTAA（又はCTAA）を、単独で、又は結合剤のライブラリー（例えば、20個の標準アミノ酸及び選択された翻訳後修飾アミノ酸のための結合剤から構成されるライブラリー）の一部として、翻訳後修飾アミノ酸に特異的な結合剤と接触させる。末端アミノ酸の脱離及び結合剤（又は結合剤のライブラリー）との接触の連続サイクルが続く。したがって、固定化されたペプチドと関連付けられた記録タグ上の結果として得られる伸長核酸は、一次アミノ酸配列の文脈で、翻訳後修飾の存在及び順序を示す。

20

#### 【0121】

特定の実施形態において、高分子、例えば、ポリペプチドはまた、非同族の結合剤と接触させられる。本明細書で使用される場合、非同族の結合剤は、考慮されている特定の標的とは異なる標的（例えば、ポリペプチドの特徴又は成分）に対して選択的である結合剤を指す。例えば、nNTAAがフェニルアラニンであり、ペプチドがそれぞれフェニルアラニン、チロシン、及びアスパラギンに選択的な3つの結合剤と接触する場合、フェニルアラニンに選択的な結合剤は、n番目のNTAA（すなわち、フェニルアラニン）に選択的に結合することができる第1の結合剤になるが、他の2つの結合剤は、（それらがフェニルアラニン以外のNTAAに選択的であるため）そのペプチドの非同族の結合剤になる。しかしながら、チロシン及びアスパラギン結合剤は、サンプル内の他のペプチドの同族の結合剤であり得る。次いで、nNTAA（フェニルアラニン）がペプチドから切断され、それにより、ペプチドのn-1アミノ酸が、n-1 NTAA（例えば、チロシン）に変換され、次いで、ペプチドが同じ3つの結合剤と接触された場合、チロシンに選択的な結合剤は、n-1 NTAA（すなわち、チロシン）に選択的に結合することができる第2の結合剤となるが、他の2つの結合剤は、（それらがチロシン以外のNTAAに選択的であるため）非同族の結合剤となる。

30

40

#### 【0122】

したがって、薬剤が結合剤であるか、又は非同族の結合剤であるかは、結合に現在利用可能な特定のポリペプチドの特徴又は成分の性質に依存することを理解されたい。また、複数のポリペプチドが多重化反応で分析される場合、あるポリペプチドの結合剤は、別のポリペプチドの非同族の結合剤である可能性があり、逆もまた同様である。したがって、結合剤に関する以下の説明は、本明細書に記載の任意のタイプの結合剤（すなわち、同族及び非同族の両方の結合剤）に適用可能であることを理解されたい。

#### 【0123】

50

特定の実施形態において、溶液中の結合剤の濃度は、アッセイのバックグラウンド及び/又は偽陽性の結果を低減するように制御される。いくつかの実施形態において、結合剤の濃度は、任意の好適な濃度、例えば、約0.0001 nM、約0.001 nM、約0.01 nM、約0.1 nM、約1 nM、約2 nM、約5 nM、約10 nM、約20 nM、約50 nM、約100 nM、約200 nM、約500 nM、又は約1000 nMであり得る。他の実施形態において、アッセイで使用される可溶性共役体の濃度は、約0.0001 nM ~ 約0.001 nMの間、約0.001 nM ~ 約0.01 nMの間、約0.01 nM ~ 約0.1 nMの間、約0.1 nM ~ 約1 nMの間、約1 nM ~ 約2 nMの間、約2 nM ~ 約5 nMの間、約5 nM ~ 約10 nMの間、約10 nM ~ 約20 nMの間、約20 nM ~ 約50 nMの間、約50 nM ~ 約100 nMの間、約100 nM ~ 約200 nMの間、約200 nM ~ 約500 nMの間、約500 nM ~ 約1000 nMの間、又は約1000 nM超である。

#### 【0124】

いくつかの実施形態において、可溶性結合剤分子と固定化されたポリペプチド、例えば、ポリペプチドとの比率は、任意の好適な範囲、例えば、約0.00001 : 1、約0.0001 : 1、約0.001 : 1、約0.01 : 1、約0.1 : 1、約1 : 1、約2 : 1、約5 : 1、約10 : 1、約15 : 1、約20 : 1、約25 : 1、約30 : 1、約35 : 1、約40 : 1、約45 : 1、約50 : 1、約55 : 1、約60 : 1、約65 : 1、約70 : 1、約75 : 1、約80 : 1、約85 : 1、約90 : 1、約95 : 1、約100 : 1、約 $10^4$  : 1、約 $10^5$  : 1、約 $10^6$  : 1、若しくはそれ以上、又は上記の比率の間の任意の比率であり得る。可溶性結合剤分子と固定化されたポリペプチド及び/又は核酸とのより高い比率を使用して、結合及び/又はコーディングタグ情報の転送を完了させることができる。これは、サンプル中の少量のポリペプチドを検出及び/又は分析するのに特に有用である可能性がある。

#### 【0125】

##### C. コーディングタグ

記載される結合剤は、結合剤に関する識別情報（例えば、結合剤を表す、又は結合剤に関連する）を含有するコーディングタグを含むか、又はそれに関連付けられる。いくつかの実施形態において、コーディングタグからの識別情報は、結合剤によって結合された標的の同一性に関する情報を含む。いくつかの実施形態において、コーディングタグからの識別情報は、結合剤によって結合されたペプチド上の1つ以上のアミノ酸の同一性に関する情報を含むか、又はそれに関連付けられる。場合によっては、コーディングタグは、切断ステップに使用される部分的な制限酵素認識配列を含む。例えば、コーディングタグは、二本鎖核酸切断試薬（例えば、制限酵素）によって認識され得る配列（又はその一部）を含み得る。場合によっては、コーディングタグは、二本鎖核酸切断試薬（例えば、制限酵素）が二本鎖になるとそれによって認識され得る一本鎖配列を含み得る。

#### 【0126】

結合剤と関連付けられたコーディングタグは、その関連付けられた結合剤に関する識別情報を含む、任意の好適な長さを有するポリヌクレオチド、例えば、2 ~ 100（2及び100を含む）の任意の整数を含む、約2塩基 ~ 約100塩基の核酸分子であるか、又はそれを含む。「コーディングタグ」は、「シーケンシング可能なポリマー」から作製することもできる（例えば、Niu et al., 2013, Nat. Chem. 5 : 282 - 292 ; Roy et al., 2015, Nat. Commun. 6 : 7237, Lutz 2015, Macromolecules 48 : 4759 - 4767を参照。これらの各々はその全体が参照により組み込まれる）。コーディングタグは、エンコーダー配列又は識別情報を有する配列を含み得、これには、任意選択的に片側に1つのスペーサーが隣接するか、又は任意選択的に両側にスペーサーが隣接する。コーディングタグはまた、任意選択的なUMI及び/又は任意の結合サイクル特異的バーコードで構成され得る。コーディングタグは、結合剤に直接付着しているコーディングタグ、結合剤に直接付着しているコーディングタグにハイブリダイズした相補配列（例えば、二本鎖コーディン

グタグの場合)、又は記録タグ上の伸長核酸に存在するコーディングタグ情報を指す場合がある。特定の実施形態において、コーディングタグは、結合サイクル特異的スペースー若しくはバーコード、固有の分子識別子、ユニバーサルプライミング部位、又はそれらの任意の組み合わせを更に含み得る。

【0127】

コーディングタグは、一本鎖分子、二本鎖分子、又は部分的に二本鎖であり得る。コーディングタグは、平滑末端、突出末端、又は各々の1つを含み得る。いくつかの実施形態において、コーディングタグは、部分的に二本鎖であり、これは、成長する伸長記録タグ内の内部エンコーダー及びスペースー配列へのコーディングタグのアニールリングを防ぐ。いくつかの実施形態において、コーディングタグは、ヘアピンを含み得る。特定の実施形態において、ヘアピンは、相互に相補的な核酸領域を含み、核酸鎖を介して接続される。いくつかの実施形態において、核酸ヘアピンはまた、二本鎖ステムセグメントから延びる3'及び/又は5'一本鎖領域を更に含み得る。いくつかの例では、ヘアピンは、一本鎖の核酸を含む。

10

【0128】

いくつかの実施形態において、記載される結合剤は、結合剤に関する識別情報を含むコーディングタグを含む。いくつかの実施形態において、コーディングタグからの識別情報は、結合剤によって結合された標的の同一性に関する情報を含む。いくつかの実施形態において、コーディングタグからの識別情報は、結合剤によって結合されたペプチド上の1つ以上のアミノ酸の同一性に関する情報を含む。

20

【0129】

コーディングタグは、約3塩基～約100塩基の核酸分子であり、その関連付けられた結合剤に固有の識別情報を提供する。コーディングタグは、約3～約90塩基、約3～約80塩基、約3～約70塩基、約3～約60塩基、約3塩基～約50塩基、約3塩基～約40塩基、約3塩基～約30塩基、約3塩基～約20塩基、約3塩基～約10塩基、又は約3塩基～約8塩基を含み得る。いくつかの実施形態において、コーディングタグは、約3個の塩基、4個の塩基、5個の塩基、6個の塩基、7個の塩基、8個の塩基、9個の塩基、10個の塩基、11個の塩基、12個の塩基、13個の塩基、14個の塩基、15個の塩基、16個の塩基、17個の塩基、18個の塩基、19個の塩基、20個の塩基、25個の塩基、30個の塩基、35個の塩基、40個の塩基、55個の塩基、60個の塩基、65個の塩基、70個の塩基、75個の塩基、80個の塩基、85個の塩基、90個の塩基、95個の塩基、又は100個の塩基の長さである。コーディングタグは、DNA、RNA、ポリヌクレオチド類似体、又はそれらの組み合わせで構成され得る。ポリヌクレオチド類似体としては、PNA、gPNA、BNA、GNA、TNA、LNA、モルフォリノポリヌクレオチド、2'-O-メチルポリヌクレオチド、アルキルリボシル置換ポリヌクレオチド、ホスホロチオエートポリヌクレオチド、及び7-デアザプリン類似体が挙げられる。

30

【0130】

結合剤は、提供された方法の酵素反応(例えば、核酸接合試薬、ポリメラーゼ、及び二本鎖核酸切断試薬の機能)を破壊しないように、任意の好適な方法及び位置でコーディングタグに付着させることができる。例えば、コーディングタグが利用可能な3'末端を有するような方法でコーディングタグが付着されるのと同様に、結合剤は、任意の好適な方法でコーディングタグに付着され得る。いくつかの特定の実施形態において、コーディングタグは、3'オーバーハングであるスペースーを有する。いくつかの実施形態において、結合剤は、コーディングタグの5'末端においてコーディングタグに付着される。いくつかの実施形態において、結合剤は、ループ領域においてコーディングタグに付着される。いくつかの実施形態において、結合剤は、再コーディングタグの5'末端にループの近くにある位置においてコーディングタグに付着される。コーディングタグは、共有結合性及び非共有結合性相互作用を含む、当技術分野で知られている任意の手段によって、直接的又は間接的に(例えば、リンカーを介して)結合剤に接合され得る。いくつかの実施形態

40

50

において、コーディングタグは、酵素的又は化学的に結合剤に接合され得る。いくつかの実施形態において、コーディングタグは、ライゲーションを介して結合剤に接合され得る。他の実施形態において、コーディングタグは、親和性結合対（例えば、ビオチン及びストレプトアビジン）を介して結合剤に接合される。場合によっては、コーディングタグは、非天然アミノ酸との共有結合的相互作用を介してなど、非天然アミノ酸への結合剤に接合され得る。

**【0131】**

いくつかの実施形態において、結合剤は、SpyCatcher - SpyTag相互作用を介してコーディングタグに接合される。SpyTagペプチドは、自発的なイソペプチド連結を介してSpyCatcherタンパク質に対する不可逆的な共有結合を形成し、それにより、力及び過酷な条件に抵抗するペプチド相互作用をもたらす遺伝的にエンコードされた方法を提供する（Zakeri et al., 2012, Proc. Natl. Acad. Sci. 109: E690 - 697、Li et al., 2014, J. Mol. Biol. 426: 309 - 317）。結合剤は、SpyCatcherタンパク質を含む融合タンパク質として発現され得る。いくつかの実施形態において、SpyCatcherタンパク質は、結合剤のN末端又はC末端に付加される。SpyTagペプチドは、標準的な共役化学（Hermanson, Bioconjugate Techniques, (2013) Academic Press）を使用してコーディングタグにカップリングされ得る。

10

**【0132】**

いくつかの実施形態において、酵素ベースの戦略は、結合剤をコーディングタグに接合するために使用される。例えば、結合剤は、ホルミルグリシン（FGly）生成酵素（FGE）を使用してコーディングタグに接合され得る。一例では、タンパク質、例えば、SpyLigaseを使用して、結合剤をコーディングタグに接合する（Fierer et al., Proc Natl Acad Sci USA. 2014; 111(13): E1176 - E1181）。

20

**【0133】**

他の実施形態において、結合剤は、SnoopTag - SnoopCatcherペプチド - タンパク質相互作用を介してコーディングタグに接合される。SnoopTagペプチドは、SnoopCatcherタンパク質とイソペプチド結合を形成する（Veggiari et al., Proc. Natl. Acad. Sci. USA, 2016, 113: 1202 - 1207）。結合剤は、SnoopCatcherタンパク質を含む融合タンパク質として発現され得る。いくつかの実施形態において、SnoopCatcherタンパク質は、結合剤のN末端又はC末端に付加される。SnoopTagペプチドは、標準的な共役化学を使用してコーディングタグにカップリングされ得る。

30

**【0134】**

更に他の実施形態において、結合剤は、HaloTag（登録商標）タンパク質融合タグ及びその化学的リガンドを介してコーディングタグに接合される。HaloTagは、合成リガンド（HaloTagリガンド）に共有結合するように設計された修飾ハロアルカンデハロゲナーゼである（Los et al., 2008, ACS Chem. Biol. 3: 373 - 382）。合成リガンドは、様々な有用な分子に付着したクロロアルカンリンカーを含む。HaloTagとクロロアルカンリンカーとの間に共有結合が形成され、これは非常に特異的であり、生理学的条件下で急速に発生し、本質的に不可逆的である。

40

**【0135】**

場合によっては、結合剤は、酵素、例えば、ソルターゼ媒介標識を使用して付着（共役）することによってコーディングタグに接合される（例えば、Antos et al., Curr Protoc Protein Sci. (2009) CHAPTER 15: Unit - 15.3、国際特許公開第2013/003555号を参照）。ソルターゼ酵素は、トランスペプチド反応を触媒する（例えば、Falck et al., Anti

50

bodies (2018) 7(4): 1-19を参照)。いくつかの態様において、結合剤は、1つ以上のN末端又はC末端グリシン残基で修飾されるか、又はそれらに接合される。

【0136】

いくつかの実施形態において、結合剤は、システインバイオコンジュゲーション方法を使用してコーディングタグに接合される。いくつかの実施形態において、結合剤は、 $\kappa$ -クランプ媒介性システイン生体抱合を使用してコーディングタグに接合される(例えば、Zhang et al., Nat Chem. (2016) 8(2): 120-128を参照)。場合によっては、結合剤は、3-アリアルプロピオロニトリル(APN)媒介性タグ付けを使用してコーディングタグに接合される(例えば、Koniev et al., Bioconj Chem. 2014; 25(2): 202-206)。

10

【0137】

いくつかの実施形態において、結合及び情報転送のサイクルで使用されるコーディングタグのセットは、サイクル特異的な配列を使用するなどのサイクル情報を含み得る。一実施形態において、コーディングタグは、結合サイクル特異的な配列を含む。いくつかの実施形態において、結合剤のコレクション内のコーディングタグは、アッセイで使用される共通のスペーサー配列を共有する(例えば、複数の結合サイクル法で使用される結合剤のライブラリー全体は、それらのコーディングタグに共通のスペーサーを有する)。別の実施形態において、コーディングタグは、特定の結合サイクルを識別する結合サイクルタグから構成されている。他の実施形態において、結合剤のライブラリー内のコーディングタグは、結合サイクル特異的なスペーサー配列を有する。いくつかの実施形態において、コーディングタグは、1つの結合サイクル特異的なスペーサー配列を含む。例えば、第1の結合サイクルで使用される結合剤のコーディングタグは、「サイクル1」特異的なスペーサー配列を含み、第2の結合サイクルで使用される結合剤のコーディングタグは、「サイクル2」特異的なスペーサー配列を含むなど、最大「n」個の結合サイクルである。更なる実施形態において、第1の結合サイクルで使用される結合剤のコーディングタグは、「サイクル1」特異的なスペーサー配列及び「サイクル2」特異的なスペーサー配列を含み、第2の結合サイクルで使用される結合剤のコーディングタグは、「サイクル2」特異的なスペーサー配列及び「サイクル3」特異的なスペーサー配列など、最大「n」個の結合サイクルを含む。この実施形態は、結合サイクルが完了した後の非連結伸長記録タグの後続のPCRアセンブリに有用である。いくつかの実施形態において、スペーサー配列は、プライマー伸長反応又は粘着末端ライゲーション反応を開始するために、記録タグ又は伸長記録タグ内の相補的なスペーサー配列にアニーリングするのに十分な数の塩基を含む。

20

30

【0138】

サイクル特異的なスペーサー配列は、情報転送が条件付きで行われるように設計することもできる。一例では、第1の結合サイクルは、コーディングタグからの情報を記録タグに転送し、後続の結合サイクルは、以前に追加されたスペーサー配列に依存する方法で転送することができる。より具体的には、第1の結合サイクルで使用される結合剤のコーディングタグは、「サイクル1」特異的なスペーサー配列及び「サイクル2」特異的なスペーサー配列を含み、第2の結合サイクルで使用される結合剤のコーディングタグは、「サイクル2」特異的なスペーサー配列及び「サイクル3」特異的なスペーサー配列など、最大「n」個の結合サイクルを含む。第1の結合サイクルからの結合剤のコーディングタグは、相補的なサイクル1特異的なスペーサー配列を介して記録タグにアニーリングすることができる。コーディングタグ情報を記録タグに転送すると、サイクル2特異的なスペーサー配列は、結合サイクル1の終わりに伸長記録タグの3'末端に位置付けられる。第2の結合サイクルからの結合剤のコーディングタグは、相補的なサイクル2特異的なスペーサー配列を介して伸長記録タグにアニーリングすることができる。コーディングタグ情報が伸長記録タグに転送されると、サイクル3特異的なスペーサー配列は、「n」個の結合サイクルを通して、結合サイクル2の終わりに伸長記録タグの3'末端に位置付けられる。この実施形態は、複数の結合サイクル間の特定の結合サイクルにおける結合情報の転送が、前の結合サイクル

40

50

を経験した（伸長された）記録タグ上でのみ発生することを提供する。場合によっては、前のコーディングタグによって追加されたスペーサーが、第2のコーディングタグのスペーサーの少なくとも一部と一致する場合、情報は、第2の（又はより高次の）コーディングタグから伸長記録タグに転送される。

#### 【0139】

いくつかの実施形態において、結合剤は、同族高分子に結合できない場合がある。「追跡」ステップとして各結合サイクルの後に結合サイクル特異的スペーサーを含むオリゴヌクレオチドを使用して、結合サイクルが失敗した場合でも結合サイクルを同期させ続けることができる。例えば、同族の結合剤が結合サイクル1中に高分子に結合できない場合、サイクル1特異的スペーサー、サイクル2特異的スペーサーの両方、及び「ヌル」エンコーダー配列を含むオリゴヌクレオチドを使用して、結合サイクル1の後に追跡ステップを追加する。「ヌル」エンコーダー配列は、エンコーダー配列、又は好ましくは「ヌル」結合サイクルを明確に識別する特定のバーコードがないことであり得る。「ヌル」オリゴヌクレオチドは、サイクル1特異的スペーサーを介して記録タグにアニーリングすることができ、サイクル2特異的スペーサーは、記録タグに移される。したがって、結合サイクル2からの結合剤は、結合サイクル1事象の失敗にもかかわらず、サイクル2特異的スペーサーを介して伸長記録タグにアニーリングすることができる。「ヌル」オリゴヌクレオチドは、結合サイクル1を、伸長記録タグ内の結合事象の失敗としてマークする。

10

#### 【0140】

好ましい実施形態において、結合サイクル特異的エンコーダー配列は、コーディングタグに使用される。結合サイクル特異的エンコーダー配列は、完全に固有の分析物（例えば、NTAA）結合サイクルエンコーダーバーコードを使用するか、サイクル特異的バーコードに接合された分析物（例えば、NTAA）エンコーダー配列を組み合わせることで達成され得る。組み合わせアプローチを使用する利点は、設計する必要のある合計バーコードが少なくなることである。10サイクルにわたって使用される20個の分析物結合剤のセットの場合、20個の分析物エンコーダー配列バーコード及び10個の結合サイクル特異的バーコードのみを設計する必要がある。対照的に、結合サイクルが結合剤エンコーダー配列に直接埋め込まれている場合は、合計200の独立したエンコーダーバーコードを設計する必要があり得る。結合サイクル情報をエンコーダー配列に直接埋め込むことの利点は、ナノポア読み出しにエラー訂正バーコードを用いるときに、コーディングタグの全長を最小限に抑えることができることである。エラー耐性のあるバーコードを使用すると、エラーが発生しやすいが、分析の速度が速く、コストが低く、かつ/又は機器のポータブル性が高いなどの他の利点を有するシーケンシングプラットフォーム及びアプローチを使用して、高精度のバーコードを識別することができる。そのような例の1つは、ナノポアに基づくシーケンスリードである。

20

30

#### 【0141】

##### II. 高分子分析アッセイ

記録タグに記載される連続エンコーディング方法を使用して情報を転送するステップを含む、高分子、例えば、ペプチド、ポリペプチド、及びタンパク質の分析のために提供される方法は、追加のステップ、処理、及び反応を含み得る。いくつかの実施形態において、高分子はポリペプチドであり、ポリペプチド分析アッセイが実行される。いくつかの実施形態において、配列（若しくはその配列の一部）及び/又はタンパク質の同一性は、ポリペプチド分析アッセイを使用して決定される。いくつかの例では、ポリペプチド分析アッセイは、好適な技法又は手順を使用して、ポリペプチドの少なくとも部分的な配列又は同一性を評価することを含む。例えば、ポリペプチドの少なくとも部分的な配列は、N末端アミノ酸分析又はC末端アミノ酸分析によって評価することができる。いくつかの実施形態において、ポリペプチドの少なくとも部分的な配列は、ProteoCodeアッセイを使用して評価することができる。いくつかの例では、ポリペプチドの少なくとも部分的な配列は、米国特許公開第2019/0145982A1号、同第2020/0348308A1号、同第2020/0348307A1号、同第2021/0208150A

40

50

1号に開示及び/又は特許請求される技法又は手順のいくつかを適用することによって評価することができる。

【0142】

ペプチド又はポリペプチドを分析する方法に関する実施形態において、この方法は、一般に、結合剤を、ポリペプチド、タンパク質又はペプチドの少なくとも末端アミノ酸（例えば、NTAA又はCTAA）に接触させ、接触した際に、情報をコーディングタグからポリペプチド、タンパク質又はペプチドと関連付けられた記録タグに転送し、それにより、一次伸長記録タグ（エンコーディングのサイクル）を生成することを含む。本明細書に記載される方法によるエンコーディングの例示的なサイクルを図1Bに示す。図1Bに示される結合剤は、リンカーによって、結合剤（結合剤バーコード、BBC）に関する識別情報を含む、コーディングタグに付着される。いくつかの実施形態において、結合剤は、描写されていない場所（例えば、コーディングタグ又は他のループ領域）でコーディングタグに付着又は接合することができる。BBC情報を転送するために、ポリメラーゼ、核酸接合試薬（DNAリガーゼなど）、及び二本鎖核酸切断試薬（制限酵素など）が提供される。これらの試薬は、その後（一度に1つの酵素、又は2つの酵素（ポリメラーゼ及びDNAリガーゼ）に続いて制限酵素）、又は3つの酵素の混合物としてのいずれかで添加することができる。記録タグの5'末端をコーディングタグの3'末端にライゲーションすると、ポリメラーゼは、記録タグの3'（ライゲーションされていない）末端を伸長して、それぞれの制限酵素部位に隣接する2塩基対スペーサーを含有するdsDNA分子を作製する。二本鎖生成後、制限酵素は、その認識部位に隣接して結合及び切断する。切断後、記録タグ上のdsDNAは、結合剤特異的バーコード（BBC）及び次のラウンドのエンコーディングでスペーサー配列として機能する2nt3'オーバーハングを含有する。いくつかの実施形態において、情報転送（エンコーディング）のサイクルの後、分析のためのポリペプチドの一部は、ポリペプチド（末端アミノ酸又は末端ジペプチドなど）から除去され得る。図1Bに示されるステップのサイクルは、追加の結合剤及び対応するコーディングタグを用いて1回以上繰り返されて、記録タグを更に伸長し得る。

10

20

30

40

【0143】

いくつかの更なる実施形態において、この方法は、高分子が結合剤と接触する前又は後に、高分子（例えば、ペプチド）を標識又は修飾することを含む。例えば、結合剤によって結合されるポリペプチド、タンパク質、又はペプチドの末端アミノ酸は、化学的に標識された、又は修飾された末端アミノ酸であり得る。いくつかの更なる実施形態において、この方法は、情報転送ステップ後に、ポリペプチド、タンパク質又はペプチドから末端アミノ酸（例えば、NTAA又はCTAA）を除去又は除去することを更に含む。脱離される末端アミノ酸は、化学的に標識された、又は修飾された末端アミノ酸であり得る。酵素又は化学試薬と接触させることによるNTAAの除去は、ポリペプチド、タンパク質又はペプチドの最後から2番目のアミノ酸を末端アミノ酸に変換する。ポリペプチド分析は、追加の結合剤を末端アミノ酸に結合すること、及びコーディングタグからの情報を伸長核酸に転送し、それにより2つ以上の結合剤に関する情報を含むより高次の伸長記録タグを生成すること、及び末端アミノ酸を周期的な方法で脱離することの1つ以上のサイクルを含み得る。追加の結合、情報の転送、及び除去は、上記のように最大n個のアミノ酸で起こり、集合的にポリペプチド、タンパク質、又はペプチドを表すn次伸長核酸を生成する。記録タグは、1つ以上の結合剤と分析される高分子との間の1つ以上の結合事象から収集された情報を記録するために使用される。

【0144】

任意の提供された実施形態のいくつかにおいて、記載された例示的なアプローチにおけるNTAAを含むステップは、代わりに、C末端アミノ酸（CTAA）を用いて実行され得る。

【0145】

いくつかの実施形態において、分解によるペプチド又はポリペプチドシーケンシングアッセイのプロセスにおけるステップの順序は、逆にするか、又は様々な順序で実行すること

50

ができる。例えば、いくつかの実施形態において、末端アミノ酸標識は、ポリペプチドを結合剤に結合する前及び/又は後に実施することができる。

【0146】

本明細書で提供されるのは、高分子を分析するための方法であり、(a)高分子及び支持体に接合された関連する記録タグを提供するステップと、(b)高分子を、高分子に結合することができる結合剤であって、結合剤に関する識別情報を有するコーディングタグを含む結合剤と接触させて、高分子と結合剤との間の結合を可能にするステップと、(c)核酸接合試薬によって記録タグの5'末端をコーディングタグの3'末端に接合するステップと、(d)ポリメラーゼによって、コーディングタグを鋳型として使用して記録タグを伸長し、二本鎖伸長記録タグを生成するステップと、(e)二本鎖伸長記録タグを二本鎖核酸切断試薬で切断して、伸長記録タグ内に3'オーバーハングを生成するステップと、を含む。場合によっては、結合剤は、ステップ(e)の後に除去される。いくつかの実施形態において、この方法は、伸長再コーディングタグを分析することを更に含む。いくつかの態様において、この方法は、伸長記録タグを分析する前に、ユニバーサルプライミング部位を伸長記録タグに追加することを更に含む。

10

【0147】

いくつかの例では、ステップ(a)は、ステップ(b)、(c)、(d)、及び(e)の前に実行される。ステップ(c)、(d)、及び(e)は、設計の性質上、段階的な方法で実行することができるが、試薬は混合物中で提供することができる。いくつかの実施形態において、ステップ(b)は、例えば、最初に核酸接合試薬を提供し、次いで、ポリメ

20

【0148】

いくつかの実施形態において、この方法は、ポリペプチド、タンパク質、又はペプチドの末端アミノ酸(例えば、N末端アミノ酸(NTAA))などのポリペプチドの一部を除去して、ポリペプチド、タンパク質、又はペプチドの新しい末端アミノ酸を露出するステップを更に含む。場合によっては、ステップ(b)、(c)、(d)、(e)の第2のサイクルが、ステップ(b)、(c)、(d)、(e)の第1のサイクルの後に、ポリペプチドの少なくとも一部分を除去した後に繰り返される。

30

【0149】

いくつかの実施形態において、この方法は、標的ポリペプチド、タンパク質、又はペプチドを、ポリペプチド、タンパク質、又はペプチドの末端アミノ酸を修飾するための試薬で処理することを含む。いくつかの態様において、ポリペプチドの末端アミノ酸を修飾するための試薬は、化学剤又は酵素剤を含む。いくつかの実施形態において、標的ポリペプチド、タンパク質、又はペプチドは、ステップ(b)の前に、末端アミノ酸を修飾するための試薬と接触される。いくつかの実施形態において、標的ポリペプチド、タンパク質、又はペプチドは、末端アミノ酸を除去する前に、末端アミノ酸を修飾するために試薬と接触させる。

40

【0150】

いくつかの実施形態において、この方法は、コーディングAGから記録タグに情報を転送した後に結合剤を除去することを更に含む。いくつかの態様において、結合剤は、ステップ(e)の後に除去される。いくつかの態様において、結合剤は、ステップ(b)を繰り返す前に除去される。いくつかの態様において、結合剤の除去は、コーディングタグからの情報を分析のために高分子と関連付けられた記録タグに転送した後に実行される。

【0151】

A. サンプル及び高分子

いくつかの実施形態において、分析アッセイは、サンプルから得られた未知の同一性のう

50

ちの1つ以上の高分子に対して実行される。場合によっては、高分子は、サンプルから得られた分子の混合物からのものである。高分子は、より小さなサブユニットから構成される大きな分子であり得る。特定の実施形態において、高分子は、タンパク質、タンパク質複合体、ポリペプチド、ペプチド、核酸分子、炭水化物、脂質、大環状分子、又はキメラ高分子である。本明細書に開示される方法における高分子（例えば、タンパク質、ポリペプチド、ペプチド）は、任意の好適な供給源又はサンプルから入手することができる。

#### 【0152】

本明細書に開示される方法は、複数の高分子の同時（多重化）の検出、識別、定量化、及び/又はシーケンシングを含む分析に使用することができる。本明細書で使用される多重化は、同じアッセイにおける複数の高分子（例えば、ポリペプチド）の分析を指す。複数の高分子は、同じサンプル又は異なるサンプルに由来し得る。複数の高分子は、同じ対象又は異なる対象に由来し得る。分析される複数の高分子は、異なる高分子、又は異なるサンプルに由来する同じ高分子であり得る。複数の高分子は、2個以上の高分子、5個以上の高分子、10個以上の高分子、50個以上の高分子、100個以上の高分子、500個以上の高分子、1000個以上の高分子、5,000個以上の高分子、10,000個以上の高分子、50,000個以上の高分子、100,000個以上の高分子、又は1,000,000個以上の高分子を含む。

10

#### 【0153】

いくつかの実施形態において、高分子（例えば、タンパク質、ポリペプチド、又はペプチド）は、生物学的サンプルであるサンプルから取得される。いくつかの実施形態において、サンプルは、哺乳動物又はヒトの細胞、酵母細胞、及び/又は細菌細胞を含むが、これらに限定されない。いくつかの実施形態において、サンプルは、多細胞生物から得られたサンプルに由来する細胞を含む。例えば、サンプルは、個体から単離され得る。いくつかの実施形態において、サンプルは、単一の細胞型又は複数の細胞型を含み得る。いくつかの実施形態において、サンプルは、例えば、穿刺、又は他の収集若しくはサンプリング手順によって、哺乳類生物又はヒトから取得され得る。いくつかの実施形態において、サンプルは、2つ以上の細胞を含む。

20

#### 【0154】

いくつかの実施形態において、生物学的サンプルは、全細胞及び/又は生細胞及び/又は細胞破片を含み得る。いくつかの例では、好適な供給源又はサンプルとしては、事実上全ての生物の生物学的サンプル、例えば、生検サンプル、細胞培養物、細胞（初代細胞及び培養細胞株の両方）、細胞小器官又は小胞を含むサンプル、組織及び組織抽出物を挙げることができるが、これらに限定されない。例えば、好適な供給源又はサンプルとしては、生検；糞便物質；体液（血液、全血、血清、血漿、尿、リンパ液、胆汁、房水、母乳、耳垢（耳垢（ear wax））、乳び、粥状液、内リンパ、外リンパ、滲出液、脳髄液、間質液、水性若しくは硝子体液、初乳、痰、羊水、唾液、肛門及び膣の分泌物、胃酸、胃液、リンパ液、粘液（鼻汁及び痰を含む）、心膜液、腹水、胸水、膿、粘膜分泌物、唾液、皮脂（皮膚の油）、痰（sputum）、滑液、汗及び精液、漏出液、嘔吐物及びそれらの1つ以上の混合物、滲出液（例えば、膿瘍から得られた流体又は感染症若しくは炎症の任意の他の部位）あるいは好ましくはマイクロバイーム含有サンプルを含む哺乳動物由来のサンプル、及び特に好ましくはマイクロバイーム含有サンプルを含むヒト由来のサンプルを用いて、事実上全ての生物の関節から得られた流体（正常な関節又はリウマチ性関節炎、骨関節炎、痛風、若しくは化膿性関節炎などの疾患に罹患した関節）；環境サンプル（空気、農業、水、及び土壌サンプル）；微生物バイオフィーム及び/又は微生物群、並びに細菌孢子由来のサンプルを含む微生物サンプル；組織切片を含む組織サンプル、細胞外液を含む研究サンプル、細胞培養物からの細胞外上清、細菌内の封入体、ミトコンドリア及び細胞ペリプラズムを含む細胞成分を挙げることができるが、これらに限定されない。いくつかの実施形態において、生物学的サンプルは、体液を含むか、又は体液に由来し、体液は、哺乳動物又はヒトから得られる。いくつかの実施形態において、サンプルは、体液、又は体液からの細胞培養物を含む。

30

40

50

## 【0155】

いくつかの実施形態において、高分子（例えば、ポリペプチド及びタンパク質）は、単一の細胞型又は複数の細胞型から取得及び調製され得る。いくつかの実施形態において、サンプルは、細胞の集団を含む。いくつかの実施形態において、高分子（例えば、タンパク質、ポリペプチド、又はペプチド）は、細胞若しくは細胞内成分、細胞外小胞、細胞小器官、又はそれらの組織化された副成分に由来する。高分子（例えば、タンパク質、ポリペプチド、又はペプチド）は、細胞小器官、例えば、ミトコンドリア、核、又は細胞小胞に由来し得る。一実施形態において、1つ以上の特定のタイプの単一細胞又はそのサブタイプを単離することができる。いくつかの実施形態において、サンプルは、細胞小器官（例えば、核、ゴルジ装置、リボソーム、ミトコンドリア、小胞体、葉緑体、細胞膜、小胞など）を含み得るが、これらに限定されない。

10

## 【0156】

特定の実施形態において、高分子は、タンパク質、タンパク質複合体、ポリペプチド、又はペプチドであるか、又はそれらを含む。アミノ酸配列情報及びペプチド、ポリペプチド、又はタンパク質の翻訳後修飾は、次世代のシーケンシング法を介して分析することができる核酸でエンコードされたライブラリーに形質導入される。ペプチドは、L-アミノ酸、D-アミノ酸、又はその両方を含み得る。ペプチド、ポリペプチド、タンパク質、又はタンパク質複合体は、標準的な天然に存在するアミノ酸、修飾されたアミノ酸（例えば、翻訳後修飾）、アミノ酸類似体、アミノ酸模倣物、又はそれらの任意の組み合わせを含み得る。いくつかの実施形態において、ペプチド、ポリペプチド、又はタンパク質は、天然に存在するか、合成的に産生されるか、又は組換え的に発現される。前述のペプチド実施形態のうちいずれかにおいて、ペプチド、ポリペプチド、タンパク質、又はタンパク質複合体は、翻訳後修飾を更に含み得る。標準的な天然アミノ酸としては、アラニン（A又はAla）、システイン（C又はCys）、アスパラギン酸（D又はAsp）、グルタミン酸（E又はGlu）、フェニルアラニン（F又はPhe）、グリシン（G又はGly）、ヒスチジン（H又はHis）、イソロイシン（I又はIle）、リジン（K又はLys）、ロイシン（L又はLeu）、メチオニン（M又はMet）、アスパラギン（N又はAsn）、プロリン（P又はPro）、グルタミン（Q又はGln）、アルギニン（R又はArg）、セリン（S又はSer）、スレオニン（T又はThr）、バリン（V又はVal）、トリプトファン（W又はTrp）、及びチロシン（Y又はTyr）が挙げられる。非標準的なアミノ酸としては、セレノシステイン、ピロリシン、及びN-ホルミルメチオニン、 $\alpha$ -アミノ酸、ホモアミノ酸、プロリン及びピルビン酸誘導体、3-置換アラニン誘導体、グリシン誘導体、環-置換フェニルアラニン及びチロシン誘導体、線状コアアミノ酸、並びにN-メチルアミノ酸が挙げられる。

20

30

## 【0157】

ペプチド、ポリペプチド、又はタンパク質の翻訳後修飾（PTM）は、共有結合修飾又は酵素修飾であり得る。翻訳後修飾の例としては、アシル化、アセチル化、アルキル化（メチル化を含む）、ビオチン化、ブチリル化、カルバミル化、カルボニル化、脱アミド化、脱イミノ化（deimination）、ジフタミド形成、ジスルフィド架橋形成、脱離（eliminylation）、フラビン付着、ホルミル化、ガンマカルボキシル化、グルタミル化、グリシル化、グリコシル化（例えば、N-連結型、O-連結型、C-連結型ホスホグリコシル化）、グリピエーション、ヘムC付着、ヒドロキシル化、ヒプシン形成、ヨウ素化、イソプレニル化、脂質化、リポイル化、マロニル化、メチル化、ミリストリル化（myristoylation）、酸化、パルミトイル化、ペグ化、ホスホパンテテイン化、リン酸化、プロピオニル化、レチニリデンシフ塩基形成、S-グルタチオン付加（glutathionylation）、S-ニトロシル化、S-スルフェニル化、セレン化、スクシニル化、スルフィン化、ユビキチン化、及びC末端アミド化が挙げられるが、これらに限定されない。翻訳後修飾としては、ペプチド、ポリペプチド若しくはタンパク質のアミノ末端及び/又はカルボキシル末端の修飾が挙げられる。末端アミノ基の修飾としては、デスアミノ、N-低級アルキル、N-ジ-低級アルキル、及びN-

40

50

アシル修飾が挙げられるが、これらに限定されない。末端カルボキシ基の修飾としては、アミド、低級アルキルアミド、ジアルキルアミド、及び低級アルキルエステル修飾（例えば、低級アルキルはC<sub>1</sub>～C<sub>4</sub>アルキルである）が挙げられるが、これらに限定されない。翻訳後修飾としては、ペプチド、ポリペプチド又はタンパク質のアミノ末端とカルボキシ末端との間にあるアミノ酸の修飾（例えば、上記のものであるが、これらに限定されない）も挙げられる。翻訳後修飾は、細胞内のタンパク質の「生物学」、例えば、その活性、構造、安定性、又は局在化を調節することができる。例えば、リン酸化はタンパク質の調節、特に細胞シグナル伝達において重要な役割を果たす（Prabakaran et al., 2012, Wiley Interdiscip Rev Syst Biol Med 4: 565 - 583）。別の例では、グリコシル化などのタンパク質への糖の付加は、タンパク質の折り畳みを促進し、安定性を改善し、調節機能を変更し、脂質がタンパク質に付着することにより、細胞膜を標的とすることができることが示されている。翻訳後修飾はまた、1つ以上の検出可能な標識を含むようにペプチド、ポリペプチド、又はタンパク質修飾を含み得る。

10

20

30

40

50

#### 【0158】

特定の実施形態において、ペプチド、ポリペプチド、又はタンパク質は、断片化され得る。ペプチド、ポリペプチド、又はタンパク質は、プロテアーゼ又はエンドペプチダーゼによる断片化を含む、当技術分野において既知の任意の手段によって断片化され得る。いくつかの実施形態において、ペプチド、ポリペプチド、又はタンパク質の断片化は、特定のプロテアーゼ又はエンドペプチダーゼの使用によって標的とされる。特定のプロテアーゼ又はエンドペプチダーゼは、特定のコンセンサス配列（例えば、TEVプロテアーゼ）に結合して切断する。他の実施形態において、ペプチド、ポリペプチド又はタンパク質の断片化は、非特異的プロテアーゼ又はエンドペプチダーゼの使用による標的とされないか、又はランダムである。非特異的プロテアーゼは、コンセンサス配列ではなく、特定のアミノ酸残基に結合して切断する場合がある（例えば、プロテイナーゼKは、非特異的セリンプロテアーゼである）。いくつかの実施形態において、当技術分野で知られており、タンパク質又はポリペプチドをより小さなペプチド断片に切断するために使用され得るなどのプロテイナーゼ及びエンドペプチダーゼは、プロテイナーゼK、トリプシン、キモトリプシン、ペプシン、サーモリシン、トロンピン、第Xa因子、フリリン、エンドペプチダーゼ、パパイン、ペプシン、サブチリシン、エラスターゼ、エンテロキナーゼ、Genenase（商標）I、エンドプロテイナーゼLysC、エンドプロテイナーゼAspN、エンドプロテイナーゼGluCなどが挙げられる（Granvogel et al., 2007 Anal Bioanal Chem 389: 991 - 1002）。特定の実施形態において、ペプチド、ポリペプチド又はタンパク質は、プロテイナーゼK、又は任意選択的に、プロテイナーゼKの熱不安定性バージョンによって断片化されて、迅速な不活性化を可能にする。場合によっては、プロテイナーゼKは、尿素及びSDSなどの変性試薬で安定しており、完全に変性したタンパク質の消化を可能にする。ペプチドへのタンパク質及びポリペプチドの断片化は、DNAタグ又はDNA記録タグの付着の前又は後に実行することができる。

#### 【0159】

化学試薬を使用して、タンパク質をペプチド断片に消化することもできる。化学試薬は、特定のアミノ酸残基で切断する可能性がある（例えば、臭化シアンは、メチオニン残基のC末端のペプチド結合を加水分解する）。ポリペプチド又はタンパク質をより小さなペプチドに断片化するための化学試薬としては、臭化シアン（CNBr）、ヒドロキシルアミン、ヒドラジン、ギ酸、BNPS-スカトール[2-(2-ニトロフェニルスルフェニル)-3-メチルインドール]、ヨードソ安息香酸、 $\cdot$ NTCB + Ni（2-ニトロ-5-チオシロ安息香酸）などが挙げられる。

#### 【0160】

特定の実施形態において、酵素的又は化学的切断に続いて、得られるペプチド断片は、ほぼ同じ所望の長さ、例えば、約10個のアミノ酸～約70個のアミノ酸、約10個のアミ

ノ酸～約60個のアミノ酸、約10個のアミノ酸～約50個のアミノ酸、約10個～約40個のアミノ酸、約10個～約30個のアミノ酸、約20個のアミノ酸～約70個のアミノ酸、約20個のアミノ酸～約60個のアミノ酸、約20個のアミノ酸～約50個のアミノ酸、約20個～約40個のアミノ酸、約20個～約30個のアミノ酸、約30個のアミノ酸～約70個のアミノ酸、約30個のアミノ酸～約60個のアミノ酸、約30個のアミノ酸～約50個のアミノ酸、又は約30個のアミノ酸～約40個のアミノ酸である。切断反応は、好ましくは、タンパク質又はポリペプチドサンプルを、プロテイナーゼ又はエンドペプチダーゼ切断部位を含むペプチド配列を含む短い試験FRET（蛍光共鳴エネルギー転送）ペプチドでスパイクすることによって、リアルタイムで監視することができる。無傷のFRETペプチドでは、切断部位を含むペプチド配列のいずれかの末端に蛍光基及び消光基が付着しており、消光剤とフルオロフォアとの間の蛍光共鳴エネルギー転送により、蛍光が低くなる。プロテアーゼ又はエンドペプチダーゼによって試験ペプチドが切断されると、消光剤及びフルオロフォアが分離され、蛍光が大幅に増加する。特定の蛍光強度が達成されたときに切断反応を停止することができ、再現性のある切断終点を達成することができる。

10

#### 【0161】

いくつかの態様において、高分子（例えば、ペプチド、ポリペプチド、又はタンパク質）のサンプルは、タンパク質又はペプチドが、細胞位置、分子量、疎水性、等電点、又はタンパク質濃縮法などの1つ以上の特性によって分離されるタンパク質分画法を受けることができる。いくつかの実施形態において、サンプル内の高分子のサブセット（例えば、タンパク質）は、高分子のサブセットがサンプルの残りの部分からソートされるように分画される。例えば、サンプルは、支持体に付着する前に分画法を受けることができる。あるいは、又は更に、タンパク質濃縮法を使用して、特定のタンパク質又はペプチドを選択する（例えば、Whiteaker et al., (2007) Anal. Biochem. 362: 44-54を参照、その全体が参照により組み込まれる）、又は特定の翻訳後修飾を選択することができる（例えば、Huang et al., 2014. J. Chromatogr. A 1372: 1-17を参照、その全体が参照により組み込まれる）。あるいは、免疫グロブリンなどの1つ以上の特定クラスのタンパク質、又はIgGなどの免疫グロブリン（Ig）アイソタイプを、分析のためにアフィニティー濃縮又は選択することができる。免疫グロブリン分子の場合、親和性結合に関与する超可変配列の配列及び存在量又は頻度の分析は、特にそれらが疾患の進行にตอบสนองして変化するか、又は健康、免疫、及び/若しくは疾患の表現型と関連するため、特に興味深い。標準的な免疫親和性法を使用して、過剰に豊富なタンパク質をサンプルから減算することもできる。豊富なタンパク質の除去は、タンパク質成分の80%超がアルブミン及び免疫グロブリンである血漿サンプルに役立つ可能性がある。上位2～20の血漿タンパク質（Pierce, Agilent）を除去する枯渴スピンカラム、又はPROTIA及びPROT20（Sigma-Aldrich）を含む、過剰に豊富なタンパク質の血漿サンプルを除去するためのいくつかの市販製品が入手可能である。

20

30

#### 【0162】

特定の実施形態において、タンパク質サンプルのダイナミックレンジは、電気泳動及び液体クロマトグラフィー（Zhou et al., 2012, Anal Chem 84(2): 720-734）を含む標準的な分画法を使用してタンパク質サンプルを分画するか、又は分画を、限定された容量のタンパク質結合ビーズ/樹脂（例えば、ヒドロキシル化シリカ粒子）（McCormick, 1989, Anal Biochem 181(1): 66-74）を充填し、結合タンパク質を溶出するコンパートメント（例えば、液滴）に分配することによって調節することができる。各区画化画分の過剰なタンパク質を洗い流す。電気泳動方法の例としては、キャピラリー電気泳動（CE）、キャピラリー等電集束（CIEF）、キャピラリー等電泳動（CITP）、自由流電気泳動、ゲル溶出液分画嵌入電気泳動（GELFREE）が挙げられる。液体クロマトグラフィータンパク質分離方法の例としては、逆相（RP）、イオン交換（IE）、サイズ排除（SE）、親

40

50

水性相互作用などが挙げられる。コンパートメントパーティションの例としては、エマルション、液滴、マイクロウェル、平坦な基板上の物理的に分離された領域などが挙げられる。例示的なタンパク質結合ビーズ/樹脂としては、フェノール基又はヒドロキシル基（例えば、StrataClean Resin from Agilent Technologies, RapidClean from LabTech, etc.）で誘導されたシリカナノ粒子が挙げられる。ビーズ/樹脂の結合能力を制限することによって、所与の画分で溶出する非常に豊富なタンパク質は、ビーズに部分的にのみ結合され、過剰なタンパク質が除去される。

#### 【0163】

いくつかの実施形態において、サンプル内の高分子の集団から的高分子のサブサンプリングへの固有のバーコードの割り当てを含むパーティションバーコードが使用される。このパーティションバーコードは、同じバーコードで標識されたコンパートメント内の高分子のパーティション化から生じる同一のバーコードから構成され得る（例えば、複数のビーズが同じバーコードを共有するバーコード化ビーズ集団）。物理コンパートメントの使用は、パーティションバーコードの割り当てを提供するために、元のサンプルを効果的にサブサンプリングする。例えば、10,000の異なる区画バーコードでラベル付けされたビーズのセットが提供される。更に、所与のアッセイにおいて、100万個のビーズの集団がアッセイに使用されると仮定する。平均して、コンパートメントバーコード（ポアソン分布）ごとに100個のビーズが存在する。更に、ビーズが1000万個の高分子の凝集体を捕捉すると仮定する。平均して、ビーズごとに10個の高分子があり、コンパートメントバーコードごとに100個のコンパートメントがあり、パーティションバーコードごとに効果的に1,000個の高分子がある（100個の異なる物理コンパートメントに対して100個のコンパートメントバーコードで構成されている）。

#### 【0164】

別の実施形態において、ポリペプチドの単一分子分画及び分画バーコード化は、ポリペプチドを（化学的又は酵素的に）N末端若しくはC末端、又はその両方で増幅可能なDNAUMITAG（例えば、記録タグ）で標識することによって達成される。DNAタグは、非特異的なフォトラベリング又はリジンなどの反応性アミノ酸への特異的な化学付着を介して、ポリペプチド（内部アミノ酸）の本体に付着される。ペプチドの末端に付着した記録タグからの情報は、酵素エマルションPCR（Williams et al., Nat Methods, (2006) 3(7): 545-550; Schutze et al., Anal Biochem. (2011) 410(1): 155-157）又はエマルションインビトロ転写/逆転写（IVT/RT）ステップを介してDNAタグに転送される。好ましい実施形態において、平均して、50nm~1000nmのサイズのエマルション液滴当たりのポリペプチドが1つ未満であるようなナノエマルションが用いられる（Nishikawa et al., J Nucleic Acids. (2012) 2012: 923214, Gupta et al., Soft Matter. (2016) 12(11): 2826-41, Sole et al., Langmuir (2006), 22(20): 8326-8332）。更に、PCRの全ての成分は、プライマー、dNTP、Mg<sup>2+</sup>、ポリメラーゼ、及びPCR緩衝液を含む水性エマルション混合物に含まれる。IVT/RTを使用する場合、記録タグは、ポリペプチドの本体に付着されたDNAタグにハイブリダイズする転写物を生成するように、T7/SP6 RNAポリメラーゼプロモーター配列を用いて設計される（Ryckelynck et al., RNA. (2015) 21(3): 458-469）。逆転写酵素（RT）は、ハイブリダイズしたRNA分子からの情報をDNAタグにコピーする。このように、エマルションPCR又はIVT/RTを使用して、末端記録タグからの情報を、ポリペプチドの本体に付着した複数のDNAタグに効果的に転送することができる。

#### 【0165】

いくつかの実施形態において、高分子標的（例えば、ペプチド、ポリペプチド、又はタンパク質）のサンプルは、物理的領域又は体積、例えば、コンパートメントに処理され得る

。サンプルに対して様々な処理及び/又は標識ステップを実行することができる。いくつかの実施形態において、コンパートメントは、高分子のサンプルから高分子のサブセットを分離又は単離する。いくつかの例では、コンパートメントは、水性コンパートメント（例えば、マイクロ流体液滴）、固体コンパートメント（例えば、プレート、チューブ、バイアル、ビーズ上のピコタイターウェル又はマイクロタイターウェル）、又は表面上の分離された領域であり得る。場合によっては、コンパートメントは、高分子を固定化することができる1つ以上のビーズを含み得る。いくつかの実施形態において、コンパートメント内の高分子は、バーコードを含むコンパートメントタグで標識される。例えば、1つのコンパートメント内の高分子は、同じバーコードで標識することができ、又は複数のコンパートメント内の高分子は、同じバーコードで標識することができる。例えば、Valihrach et al., Int J Mol Sci. 2018 Mar 11; 19(3). pii: E807. Encapsulation of cellular contents via gelation in beads is a useful approach to single cell analysis (Tamminger et al., Front Microbiol (2015) 6: 195, Spencer et al., ISME J (2016) 10(2): 427-436)を参照されたい。単一細胞液滴をバーコード化することにより、単一細胞からの全ての成分を同じ識別子で標識することができる (Klein et al., Cell (2015) 161(5): 1187-1201, Zilionis et al., Nat Protoc (2017) 12(1): 44-73、国際特許公開第2016/130704号) 20。コンパートメントのバーコード化は、液滴接合 (Bio-Rad Laboratories)、液滴へのバーコード化されたビーズの導入 (10x Genomics)、又はGundersenら (国際特許公開第2016/130704号を参照、その全体が参照により組み込まれる) によって記載されるように、カプセル化及びゲル化後の液滴の成分の組み合わせバーコード化、並びにスプリットプール組み合わせバーコード化によって、各液滴への固有のバーコードの直接組み込みを含む様々な方法で達成することができる。同様の組み合わせ標識スキームを核に適用することもできる (Vitak et al., Nat Methods (2017) 14(3): 302-308)。

#### 【0166】

いくつかの実施形態において、高分子は、結合剤と接触する前に支持体に接合される。場合によっては、サンプル内の多数の高分子を固定するために、大きな担持能力を有する支持体を使用することが望ましい。いくつかの実施形態において、高分子を三次元支持体（例えば、多孔質マトリックス又はビーズ）を使用して固定化することが好ましい。いくつかの例では、調製物は、高分子を固定化する前又は後に、高分子を核酸分子又はオリゴヌクレオチドに接合することを含む。いくつかの実施形態において、結合剤と接触する前に、複数の高分子が支持体に付着される。 30

#### 【0167】

支持体は、ビーズ、マイクロビーズ、アレイ、ガラス表面、シリコン表面、プラスチック表面、フィルター、膜、PTFE膜、ナイロン、マイクロタイターウェル、ELISAプレート、スピン干渉ディスク、ニトロセルロース膜、ニトロセルロースベースのポリマー表面、ナノ粒子、又はミク로스フェアを含むが、これらに限定されない任意の固体又は多孔性支持体であり得る。支持体の材料としては、アクリルアミド、アガロース、セルロース、デキストラン、ニトロセルロース、ガラス、金、石英、ポリスチレン、ポリエチレン酢酸ビニル、ポリプロピレン、ポリエステル、ポリメタクリレート、ポリアクリレート、ポリエチレン、ポリエチレンオキシド、ポリシリケート、ポリカーボネート、ポリビニルアルコール (PVA)、テフロン (登録商標)、フルオロカーボン、ナイロン、シリコンゴム、シリカ、ポリ無水物、ポリグリコール酸、ポリ塩化ビニル、ポリ乳酸、ポリオルトエステル、官能化シラン、ポリプロピルフェレート、コラーゲン、グリコサミノグリカン、ポリアミノ酸、又はそれらの任意の組み合わせが挙げられるが、これらに限定されない。特定の実施形態において支持体は、ビーズ、例えば、ポリスチレンビーズ、ポリマービ 40 50

ーズ、ポリアクリレートビーズ、アガロースビーズ、セルロースビーズ、デキストランビーズ、アクリルアミドビーズ、固体コアビーズ、多孔質ビーズ、常磁性ビーズ、ガラスビーズ、シリカベースのビーズ、若しくは制御された多孔質ビーズ、又はそれらの任意の組み合わせである。いくつかの特定の実施形態において、支持体は、多孔質アガロースビーズである。

【0168】

いくつかの実施形態において、支持体は、共有結合性の相互作用及び非共有結合性の相互作用、又はそれらの任意の組み合わせを含む、当技術分野で既知の任意の手段によって、直接的又は間接的に高分子、例えば、ポリペプチドを関連付けることができる、多孔質材料及び非多孔質材料を含めた任意の好適な固体材料を含む得る。支持体は、二次元（例えば、平面）又は三次元（例えば、ゲルマトリックス又はビーズ）であり得る。支持体は、ビーズ、マイクロビーズ、アレイ、ガラス表面、シリコン表面、プラスチック表面、フィルター、膜、PTFE膜、PTFE膜、ニトロセルロース膜、ニトロセルロースベースのポリマー表面、ナイロン、マイクロタイターウェル、ELISAプレート、スピン干渉ディスク、ポリマーマトリックス、ナノ粒子、又はミクروسフェアを含むが、これらに限定されない任意の支持体表面であり得る。支持体の材料としては、アクリルアミド、アガロース、セルロース、デキストラン、ニトロセルロース、ガラス、金、石英、ポリスチレン、ポリエチレン酢酸ビニル、ポリプロピレン、ポリエステル、ポリメタクリレート、ポリアクリレート、ポリエチレン、ポリエチレンオキシド、ポリシリケート、ポリカーボネート、ポリビニルアルコール（PVA）、テフロン（登録商標）、フルオロカーボン、ナイロン、シリコンゴム、ポリ無水物、ポリグリコール酸、ポリ塩化ビニル、ポリ乳酸、ポリオルトエステル、官能化シラン、ポリプロピルフェレート、コラーゲン、グリコサミノグリカン、ポリアミノ酸、デキストラン、又はそれらの任意の組み合わせが挙げられるが、これらに限定されない。支持体は更に、薄膜、膜、ボトル、皿、繊維、織物、チューブなどの成形ポリマー、粒子、ビーズ、ミクروسフェア、微粒子、又はそれらの任意の組み合わせを含む。例えば、固体表面がビーズである場合、ビーズは、セラミックビーズ、ポリスチレンビーズ、ポリマービーズ、ポリアクリレートビーズ、メチルスチレンビーズ、アガロースビーズ、セルロースビーズ、デキストランビーズ、アクリルアミドビーズ、固体コアビーズ、多孔質ビーズ、常磁性ビーズ、ガラスビーズ、又は制御多孔質ビーズ、シリカベースのビーズ、又はそれらの任意の組み合わせを含み得るが、これらに限定されない。ビーズは、球形であっても不規則な形状であってもよい。ビーズ又は支持体は、多孔質であり得る。ビーズのサイズは、ナノメートル、例えば100nm～数ミリメートル、例えば1mmまでの範囲であり得る。特定の実施形態において、ビーズのサイズは、約0.2ミクロン～約200ミクロン、又は約0.5ミクロン～約5ミクロンの範囲である。いくつかの実施形態において、ビーズは、直径約1、1.5、2、2.5、2.8、3、3.5、4、4.5、5、5.5、6、6.5、7、7.5、8、8.5、9、9.5、10、10.5、15、又は20µmであり得る。特定の実施形態において、「ビーズ」支持体は、個々のビーズ又は複数のビーズを指す場合がある。いくつかの実施形態において、固体表面はナノ粒子である。特定の実施形態において、ナノ粒子のサイズは、直径が約1nm～約500nm、例えば、直径が約1nm～約20nmの間、約1nm～約50nmの間、約1nm～約100nmの間、約10nm～約50nmの間、約10nm～約100nmの間、約50nm～約100nmの間、約50nm～約150の間、約50nm～約200nmの間、約100nm～約200nmの間、又は約200nm～約500nmの間の範囲である。いくつかの実施形態において、ナノ粒子は、直径が約10nm、約50nm、約100nm、約150nm、約200nm、約300nm、又は約500nmであり得る。いくつかの実施形態において、ナノ粒子は、直径が約200nm未満である。

【0169】

高分子を支持体（例えば、固体支持体又は多孔性支持体）に付着させるために、様々な反応を使用することができる。高分子は、支持体に直接的又は間接的に付着させることがで

10

20

30

40

50

きる。場合によっては、高分子は、核酸を介して支持体に付着される。例示的な反応としては、トリアゾールを形成するためのアジド及びアルキンの銅触媒反応 (Huisgen 1, 3 - 双極子環付加)、株促進型アジドアルキン環付加 (SPAAC)、ジエン及びジエノフィルの反応 (ディールス・アルダー)、株促進型アルキン - ニترون環付加、歪みアルケンとアジド、テトラジン、又はテトラゾールとの反応、アルケン及びアジド [3 + 2] 環付加、アルケン及びテトラジン逆電子要求ディールス・アルダー (IEDDA) 反応 (例えば、m - テトラジン (mTet) 又はフェニルテトラジン (pTet) 及びトランス - シクロオクテン (TCO)、若しくは pTet 及びアルケン)、アルケン及びテトラゾール光化学反応、アジド及びホスフィンのシュタウディング - ライゲーション、並びに求電子性原子に対する求核攻撃による脱離基の置換などの様々な置換反応 (Horisawa 2014, Knall, Hollauf et al., 2014) が挙げられる。例示的な置換反応には、アミンと、活性化エステル、N - ヒドロキシスクシンイミドエステル、イソシアネート、イソチオシアネート、アルデヒド、エポキシドなどとの反応が含まれる。いくつかの実施形態において、IEDDA クリック化学は、迅速であり、かつ、低入力濃度で高収率を実現するという理由で、高分子 (例えば、ポリペプチド) を支持体に固定化するために使用される。別の実施形態において、m - テトラジンは結合安定性が改善されているので、テトラジンではなく m - テトラジンが IEDDA クリック化学反応において使用される。別の実施形態において、フェニルテトラジン (pTet) は、IEDDA クリック化学反応において使用される。1つの場合において、ポリペプチドは、アルキン - NHS エステル (アセチレン - PEG - NHS エステル) 試薬又はアルキン - ベンゾフェノンなどの二官能性クリック化学試薬で標識され、アルキン標識ポリペプチドを生成する。いくつかの実施形態において、アルキンはまた、ジベンゾシクロオクチル (DBCO) などを含むシクロオクチンなどの歪みアルキンであり得る。

10

20

30

40

50

#### 【0170】

複数の高分子が同じ支持体上に固定化される特定の実施形態において、高分子は、標的を評価するために使用される分析ステップに対応するために適切に離間させることができる。例えば、タンパク質を評価及びシーケンシングするための核酸ベースの方法が実行されることを可能にするために最適に高分子を離間させることが有利である可能性がある。場合によっては、支持体上の高分子の間隔は、1つの固定化された高分子に結合した結合剤のコーティングタグにハイブリダイズされたアダプター分子からの情報転送が、隣接する高分子に到達することがあることを考慮して決定される。

#### 【0171】

いくつかの実施形態において、支持体の表面は、不動態化 (ブロック) されている。「不動態化された」表面とは、材料の外層で処理された表面を指す。表面を不動態化する方法としては、蛍光単一分子分析の文献からの標準方法が挙げられ、ポリマー様ポリエチレングリコール (PEG) (Pan et al., 2015, Phys. Biol. 12: 045006)、ポリシロキサン (例えば、Pluronic F - 127)、星型ポリマー (例えば、星型 PEG) (Groll et al., 2010, Methods Enzymol. 472: 1 - 18)、疎水性ジクロロジメチルシラン (DDS) + 自己組織化 Tween (登録商標) - 20 (Hua et al., 2014, Nat. Methods 11: 1233 - 1236)、ダイヤモンド様炭素 (DLC)、DLC + PEG (Stavis et al., 2011, Proc. Natl. Acad. Sci. USA 108: 983 - 988)、及び両性イオン部分 (例えば、米国特許出願公開第 2006 / 0183863 号) で表面を不動態化することを含む。共有結合性表面修飾に加えて、Tween (登録商標) - 20 のような界面活性剤、溶液中のポリシロキサン (Pluronic シリーズ)、ポリビニルアルコール (PVA)、並びに BSA 及びカゼインのようなタンパク質を含む、多くの不動態化剤も同様に使用することができる。あるいは、高分子 (例えば、タンパク質、ポリペプチド、若しくはペプチド) の密度は、タンパク質、ポリペプチド、又はペプチドを固体基質に固定化するとき、競合物質又は「ダミー」反応性分子をスパイクすることによって、固体基質の表面上又はその体積内で滴

定することができる。

【0172】

支持体上の固定化標的の間隔を制御するために、標的を付着させるための官能性カップリング基（例えば、TCO又はカルボキシル基（COOH））の密度を、基質表面上で滴定することができる。いくつかの実施形態において、複数の標的分子（例えば、高分子）は、隣接した分子が、約50nm～約500nm、又は約50nm～約400nm、又は約50nm～約300nm、又は約50nm～約200nm、又は約50nm～約100nmの距離で離間するように、支持体の表面上又は体積内（例えば、多孔質支持体）で離間される。いくつかの実施形態において、複数の分子は、少なくとも50nm、少なくとも60nm、少なくとも70nm、少なくとも80nm、少なくとも90nm、少なくとも100nm、少なくとも150nm、少なくとも200nm、少なくとも250nm、少なくとも300nm、少なくとも350nm、少なくとも400nm、少なくとも450nm、又は少なくとも500nmの平均距離で、支持体の表面上で離間される。いくつかの実施形態において、複数の分子は、少なくとも50nmの平均距離で、支持体の表面上で離間される。いくつかの実施形態において、分子は、経験的に、分子間事象と分子内事象（例えば、情報の転送）との相対頻度が、 $< 1 : 10$ 、 $< 1 : 100$ 、 $< 1 : 1,000$ 、又は $< 1 : 10,000$ であるように、支持体の表面上又は体積内で離間される。

10

【0173】

いくつかの実施形態において、複数の高分子は、約50～100nm、約50～250nm、約50～500nm、約50～750nm、約50～1,000nm、約50～1,500nm、約50～2,000nm、約100～250nm、約100～500nm、約200～500nm、約300～500nm、約100～1000nm、約500～600nm、約500～700nm、約500～800nm、約500～900nm、約500～1000nm、約500～2,000nm、約500～5,000nm、約1000～5,000nm、又は約3,000～5,000nmの範囲の2つの隣接した分子間の平均距離で離間された支持体上でカップリングされる。

20

【0174】

いくつかの実施形態において、支持体上の高分子の適切な間隔は、基質表面上の利用可能な付着分子の比率を滴定することによって達成される。いくつかの例では、基質表面（例えば、ビーズ表面）は、活性化剤（例えば、活性化剤はEDC及びスルホ-NHSである）で処理されるカルボキシル基（COOH）で官能化される。いくつかの例では、基質表面（例えば、ビーズ表面）は、NHS部分を含む。いくつかの実施形態において、 $mPEG_n-NH_2$ 及び $NH_2-PEG_n-mTet$ の混合物が、活性化されたビーズに添加される（式中、 $n$ は、1～100などの任意の数値である）。 $mPEG_3-NH_2$ （カップリングには利用不可）と $NH_2-PEG_{24}-mTet$ （カップリングに利用可能）との間の比率を滴定して、ポリペプチドを基質表面上に付着させるために利用可能な官能性部分の適切な密度を生成する。特定の実施形態において、固体表面上のカップリング部分（例えば、 $NH_2-PEG_4-mTet$ ）間の平均間隔は、少なくとも50nm、少なくとも100nm、少なくとも250nm、又は少なくとも500nmである。いくつかの特定の実施形態において、 $NH_2-PEG_n-mTet$ 対 $mPEG_3-NH_2$ の比は、約1

30

40

【0175】

B. 切断

いくつかの実施形態において、提供される方法は、情報転送後に高分子の一部を除去することを更に含む。高分子の一部の除去は、分析のための高分子の新しい部分を露出させ、例えば、分析の次のサイクルで提供される結合剤によって結合するために利用可能である

50

。ポリペプチドが分析されている場合、方法は、1つ以上のアミノ酸（例えば、末端アミノ酸）を含むポリペプチドの一部の除去を含み得る。分解によるアプローチを使用してペプチド又はポリペプチドを分析する方法に関する実施形態において、第1の結合剤をn個のアミノ酸のペプチドのnNTAAに接触させて結合し、コーディングタグ情報をペプチドと関連付けられた核酸に転送し、それにより、一次伸長核酸を（例えば、記録タグ上に）生成し、nNTAAは脱離される。酵素又は化学試薬と接触させることによるn個の標識されたNTAAの除去は、ペプチドのn-1個のアミノ酸をN末端アミノ酸に変換し、これは本明細書ではn-1NTAAと称される。第2の結合剤をペプチド又はポリペプチドと接触させ、n-1NTAAに結合し、第2の結合剤の情報をコーディングタグから一次伸長核酸に転送し、それにより、（例えば、ペプチドを表す連結されたn次伸長核酸を生成するために）二次伸長核酸を生成する。n-1個の標識されたNTAAの脱離は、ペプチド又はポリペプチドのn-2個のアミノ酸をN末端アミノ酸に変換し、これは本明細書ではn-2NTAAと称される。追加の結合、情報の転送、及び除去は、上記のように最大n個のアミノ酸で起こり、集合的にペプチド又はポリペプチドを表すn次伸長核酸又はn個の別個の伸長核酸を生成する。本明細書で使用される場合、結合剤、コーディングタグ、又は伸長核酸に関して使用される場合のn「次（order）」は、結合剤及びその関連付けられたコーディングタグが使用されるn回目の結合サイクル、又は伸長核酸が（例えば、記録タグ上で）作成されるn回目の結合サイクルを指す。いくつかの実施形態において、記載された例示的なアプローチにおけるNTAAを含むステップは、代わりに、C末端アミノ酸（CTAA）を用いて実行され得る。

10

20

【0176】

ペプチド又はポリペプチドの分析に関連する特定の実施形態において、末端アミノ酸をペプチド又はポリペプチドから除去又は切断して、新しい末端アミノ酸を露出する。除去のためのポリペプチドの一部は、ポリペプチドの標識された（例えば、化学的に標識された、又は修飾された）部分であり得る。いくつかの実施形態において、末端アミノ酸は、NTAAである。他の実施形態において、末端アミノ酸は、CTAAである。末端アミノ酸の切断は、化学的切断及び酵素的切断を含む、任意の数の既知の技法によって達成することができる。いくつかの実施形態において、標識された（例えば、化学的に標識又は修飾された）N末端アミノ酸の除去を触媒する操作された酵素又はそれを促進する試薬が使用される。いくつかの実施形態において、末端アミノ酸は、国際特許公開第2019/089846号及び国際特許公開第PCT/US2020/29969号及び同第PCT/US2020/24521号に記載されている方法のうちのいずれかを使用して除去又は除去される。

30

【0177】

いくつかの実施形態において、末端アミノ酸を除去するための試薬は、カルボキシペプチダーゼ又はアミノペプチダーゼ又はそれらのバリエーション、変異体、若しくは修飾タンパク質；加水分解酵素又はそれらのバリエーション、変異体、若しくは修飾タンパク質；軽度のエドマン分解試薬；エドマナーゼ酵素；無水TFA、塩基；又はそれらの任意の組み合わせを含む。いくつかの実施形態において、軽度のエドマン分解は、ジクロロ酸若しくはモノクロロ酸を使用するか、軽度のエドマン分解は、TFA、TCA、若しくはDCAを使用するか、又は軽度のエドマン分解は、トリエチルアミン、トリエタノールアミン、若しくは酢酸トリエチルアンモニウム（Et<sub>3</sub>NHOAc）を使用する。

40

【0178】

場合によっては、アミノ酸を除去するための試薬は、塩基を含む。いくつかの実施形態において、塩基は、水酸化物、アルキル化アミン、環状アミン、炭酸塩緩衝液、リン酸三ナトリウム緩衝液、又は金属塩である。いくつかの例では、水酸化物は、水酸化ナトリウムである。アルキル化アミンは、メチルアミン、エチルアミン、プロピルアミン、ジメチルアミン、ジエチルアミン、ジプロピルアミン、トリメチルアミン、トリエチルアミン、トリプロピルアミン、シクロヘキシルアミン、ベンジルアミン、アニリン、ジフェニルアミン、N,N-ジイソプロピルエチルアミン（DIPEA）、及びリチウムジイソプロピル

50

アミド ( L D A ) から選択される。環状アミンは、ピリジン、ピリミジン、イミダゾール、ピロール、インドール、ピペリジン、プロリジン、1, 8 - ジアザビシクロ [ 5 . 4 . 0 ] ウンデカ - 7 - エン ( D B U )、及び 1, 5 - ジアザビシクロ [ 4 . 3 . 0 ] ノン - 5 - エン ( D B N ) から選択される。炭酸塩緩衝液は、炭酸ナトリウム、炭酸カリウム、炭酸カルシウム、重炭酸ナトリウム、重炭酸カリウム、又は重炭酸カルシウムを含む。金属塩は、銀を含む。又は金属塩は、 $A g C l O_4$  である。

**【 0 1 7 9 】**

いくつかの実施形態において、提供される方法は、P T C、修飾された P T C、C b z、D N P、S N P、アセチル、グアニジニル、二複素環メタンイミンなどの基によって修飾されている場合の N 末端アミノ酸の切断を含む。N T A A の酵素的切断は、アミノペプチダーゼ又は他のペプチダーゼによって達成され得る。アミノペプチダーゼは、単量体及び多量体酵素として天然に存在し、金属又は A T P 依存性である可能性がある。天然のアミノペプチダーゼは、特異性が非常に限られており、一般的に N 末端アミノ酸を進行的に切断し、次々とアミノ酸を切断する。本明細書に記載される方法では、アミノペプチダーゼ (例えば、金属酵素アミノペプチダーゼ) は、N 末端標識で修飾された場合にのみ N T A A への特異的結合又は触媒活性を有するように操作することができる。このように、アミノペプチダーゼは、N 末端から一度に単一のアミノ酸のみを切断し、分解サイクルの制御を可能にする。いくつかの実施形態において、修飾されたアミノペプチダーゼは、N 末端標識に対して選択的である一方で、アミノ酸残基の同一性に関して非選択的である。他の実施形態において、修飾されたアミノペプチダーゼは、アミノ酸残基の同一性及び N 末端標識の両方に対して選択的である。標識された (ビオチン化された) N T A A の個々のグループ又は小グループに結合して切断する操作されたアミノペプチダーゼ変異体が記載されている (P C T 公開第 2 0 1 0 / 0 6 5 3 2 2 号を参照)。

10

20

**【 0 1 8 0 】**

標識された (ビオチン化された) N T A A の個々のグループ又は小グループに結合して切断する操作されたアミノペプチダーゼ変異体が記載されている (P C T 公開第 2 0 1 0 / 0 6 5 3 2 2 号、その全体が参照により組み込まれる)。アミノペプチダーゼは、タンパク質又はペプチドの N 末端からアミノ酸を切断する酵素である。天然のアミノペプチダーゼは、特異性が非常に限られており、一般的に N 末端アミノ酸を進行的に脱離し、次々とアミノ酸を切断する (K i s h o r e t a l . , 2 0 1 5 , A n a l . B i o c h e m . 4 8 8 : 6 - 8 )。しかしながら、残基特異的アミノペプチダーゼが識別されている (E r i q u e z e t a l . , J . C l i n . M i c r o b i o l . 1 9 8 0 , 1 2 : 6 6 7 - 7 1 , W i l c e e t a l . , 1 9 9 8 , P r o c . N a t l . A c a d . S c i . U S A 9 5 : 3 4 7 2 - 3 4 7 7 , L i a o e t a l . , 2 0 0 4 , P r o t . S c i . 1 3 : 1 8 0 2 - 1 0 )。ペプチドの N 末端からの段階的分解の制御は、標識の存在下でのみ活性 (例えば、結合活性又は触媒活性) である操作されたアミノペプチダーゼを使用することによって達成され得る。特定の実施形態において、アミノペプチダーゼは、非特異的であるように操作され得、その結果、ある特定のアミノ酸を別のアミノ酸よりも選択的に認識するのではなく、標識された N 末端を単に認識する。更に別の実施形態において、環状切断は、操作されたアシルペプチド加水分解酵素 (A P H) を使用してアセチル化 N T A A を切断することによって達成される。更に別の実施形態において、N T A A のアミド化 (グアニジニル化) を用いて、N a O H を使用して標識された N T A A の穏やかな切断を可能にする (H a m a d a , ( 2 0 1 6 ) B i o o r g M e d C h e m L e t t 2 6 ( 7 ) : 1 6 9 0 - 1 6 9 5 )。

30

40

**【 0 1 8 1 】**

いくつかの実施形態において、この方法は、ステップ ( b ) の前に N 末端プロリンを切断するのに好適な条件下で、ポリペプチドをプロリンアミノペプチダーゼと接触させることを更に含む。いくつかの例では、プロリンアミノペプチダーゼ (P A P) は、ポリペプチドから N 末端プロリンを特異的に切断することができる酵素である。N 末端プロリンを切断する P A P 酵素は、プロリンイミノペプチダーゼ (P I P) とも称される。既知の単量

50

体PAPとしては、*B. coagulans*、*L. delbrueckii*、*N. gonorrhoeae*、*F. meningosepticum*、*S. marcescens*、*T. acidophilum*、*L. plantarum* (MEROPS S33.001)からのファミリーメンバーが挙げられる (Nakajima et al., *J Bacteriol.* (2006) 188(4):1599-606、Kitazono et al., *Bacteriol* (1992) 174(24):7919-7925)。既知の多量体PAPとしては、*D. hansenii* (Bolumar et al., (2003) 86(1-2):141-151)及び他の種からの同様の相同体 (Basten et al., *Mol Genet Genomics* (2005) 272(6):673-679)が挙げられる。PAPの天然の又は操作されたバリエーション/突然変異体のいずれかを用いることができる。 10

#### 【0182】

CTAA結合剤に関する実施形態については、ペプチド又はポリペプチドからCTAAを切断する方法もまた、当技術分野で知られている。例えば、米国特許第6,046,053号は、ペプチド又はタンパク質をアルキル酸無水物と反応させて、カルボキシ末端をオキサゾロンに変換し、酸及びアルコール又はエステルとの反応によってC末端アミノ酸を遊離させる方法を開示している。CTAAの酵素的切断は、カルボキシペプチダーゼによって達成することもできる。いくつかのカルボキシペプチダーゼは、アミノ酸の優先性を示し、例えば、カルボキシペプチダーゼBは、アルギニン及びリジンなどの塩基性アミノ酸で優先的に切断する。上記のように、カルボキシペプチダーゼはまた、アミノペプチダーゼと同じ方法で修飾されて、C末端標識を有するCTAAに特異的に結合するカルボキシペプチダーゼを操作することができる。このように、カルボキシペプチダーゼは、C末端から一度に単一のアミノ酸のみを切断し、分解サイクルの制御を可能にする。いくつかの実施形態において、修飾されたカルボキシペプチダーゼは、C末端標識に対して選択的である一方で、アミノ酸残基の同一性に関して非選択的である。他の実施形態において、修飾されたカルボキシペプチダーゼは、アミノ酸残基の同一性及びC末端標識の両方に対して選択的である。 20

#### 【0183】

##### C. 分析

いくつかの実施形態において、提供された方法を実行することから生成された伸長記録タグは、1つ以上のコーディングタグから転送された情報を含む。いくつかの実施形態において、伸長記録タグは、結合剤による結合の順序を示す複数のコーディングタグから転送された一連の情報を含む。この伸長記録タグは、任意の好適な方法によって分析することができる。いくつかの実施形態において、伸長記録タグ(又はその一部)は、少なくとも伸長記録タグ内のコーディングタグの配列を決定する前に増幅される。いくつかの実施形態において、伸長記録タグ(又はその一部)は、少なくとも伸長記録タグ内のコーディングタグの配列を決定する前に放出される。 30

#### 【0184】

本明細書に記載の方法によって生成される最終的な伸長記録タグの長さは、コーディングタグ(例えば、バーコード及びスペーサー)の長さ、核酸の長さ(例えば、任意選択的に、任意選択的に、任意の固有の分子識別子、スペーサー、ユニバーサルプライミング部位、バーコード、又はそれらの組み合わせを含む)を含む複数の要因に依存する。最終タグ情報を(例えば、任意のコーディングタグから)伸長核酸に転送した後、ライゲーション、プライマー伸長、又は当技術分野で知られている他の方法を介してユニバーサルリバースプライミング部位を付加することによって、タグをキャップすることができる。いくつかの実施形態において、(例えば、記録タグ上の)核酸内のユニバーサルフォワードプライミング部位は、最終的な伸長核酸に付加されるユニバーサルリバースプライミング部位と適合性がある。いくつかの実施形態において、ユニバーサルリバースプライミング部位は、Illumina P7プライマー又はIllumina P5プライマーである。コーディングタグからの識別情報が転送される核酸の鎖センスに応じて、センス又はアンチセ 40 50

ンス P7 を付加することができる。伸長核酸ライブラリーは、支持体（例えば、ビーズ）から直接切断又は増幅することができ、従来の次世代シーケンシングアッセイ及びプロトコルで使用することができる。

【0185】

いくつかの実施形態において、プライマー伸長反応は、（例えば、記録タグ上で伸長された）一本鎖伸長核酸のライブラリー上で実行され、その相補鎖をコピーする。いくつかの実施形態において、ペプチドシーケンシングアッセイ（例えば、ProteoCodeアッセイ）は、周期的進行におけるいくつかの化学的及び酵素的ステップを含む。

【0186】

伸長核酸記録タグは、様々な核酸シーケンシング方法を使用して処理及び分析することができる。いくつかの実施形態において、1つ以上のコーディングタグ及び任意の他の核酸成分からの情報を含む伸長記録タグが、処理及び分析される。いくつかの実施形態において、伸長記録タグの集合を連結することができる。いくつかの実施形態において、伸長記録タグは、配列を決定する前に増幅することができる。

10

【0187】

シーケンシング方法の例としては、チェーンターミネーションシーケンシング（サンガーシーケンシング）；合成によるシーケンシング、ライゲーションによるシーケンシング、ハイブリダイゼーションによるシーケンシング、ポロニーシーケンシング、イオン半導体シーケンシング、及びパイロシーケンシングなどの次世代シーケンシング方法；並びに単一分子リアルタイムシーケンシング、ナノポアベースのシーケンシング、デュプレックス

20

【0188】

本発明で使用するのに好適なシーケンシング方法としては、ハイブリダイゼーションによるシーケンシング、合成技術（例えば、HiSeq（商標）及びSolexa（商標）、Illumina）によるシーケンシング、SMRT（商標）（単一分子リアルタイム）技術（Pacific Biosciences）、真の単一分子シーケンシング（例えば、Heliscope（商標）、Helicos Biosciences）、大規模並列次世代シーケンシング（例えば、SOLID（商標）、Applied Biosciences、Solexa及びHiSeq（商標）、Illumina）、超並列半導体

30

【0189】

核酸のライブラリー（例えば、伸長核酸）は、様々な方法で増幅され得る。核酸のライブラリー（例えば、1つ以上のコーディングタグからの情報を含む記録タグ）は、例えば、PCR又はエマルジョンPCRを介して、指数関数的増幅を受ける。エマルジョンPCRは、より均一な増幅をもたらすことが知られている（Hori, Fukano et al., Biochem Biophys Res Commun (2007) 352 (2) : 323 - 328）。あるいは、核酸（例えば、伸長核酸）のライブラリーは、例えば、T7 RNAポリメラーゼを使用する鋳型DNAのインビトロ転写を介して、線形増幅を受けることができる。核酸（例えば、伸長核酸）のライブラリーは、そこに含まれるユニバーサルフォワードプライミング部位及びユニバーサルリバースプライミング部位と適合性のあるプライマーを使用して増幅させることができる。核酸のライブラリー（例えば、記録タグ）を、テールドプライマーを使用して増幅させて、伸長核酸の5'末端、3'末端、又はその両端のいずれかに配列を付加することもできる。伸長核酸の末端に付加することができる配列には、ライブラリー特異的索引配列が含まれ、1回のシーケンシング実行での複数のライブラリー、アダプター配列、リードプライマー配列、又は他の配列を多重化して、伸長核酸のライブラリーをシーケンシングプラットフォームと適合性のあるも

40

50

のにすることができる。次世代シーケンシングのための調製におけるライブラリー増幅の例は、以下のとおりである。約1 mgのビーズ(約10 ng)、200 μMのdNTP、1 μMの各フォワード及びリバース増幅プライマー、0.5 μl(1 U)のPhusion Hot Start酵素(New England Biolabs)から溶出された伸長核酸ライブラリーを使用して、20 μlのPCR反応容量を設定し、98 で30秒間、続いて98 で10秒間、60 で30秒間、72 で30秒間、72 で7分間を20サイクルのサイクル条件に供し、次いで4 で保持する。

#### 【0190】

特定の実施形態において、増幅前、増幅中、又は増幅後のいずれかで、核酸(例えば、伸長核酸)のライブラリーは、標的濃縮を受けることができる。いくつかの実施形態において、標的濃縮を使用して、シーケンシングの前に、伸長核酸のライブラリーから目的の高分子(例えば、ポリペプチド)を表す伸長核酸を選択的に捕捉又は増幅させることができる。いくつかの態様において、標的タンパク質に対して高度に特異的な結合剤を産生する際の費用が高く、かつ困難であるために、タンパク質シーケンシングのための標的濃縮は難解である。場合によっては、抗体は非特異的であり、何千ものタンパク質にまたがって生産を拡大することが難しいことで有名である。いくつかの実施形態において、本開示の方法は、タンパク質コードを核酸コードに変換することによってこの問題を回避し、核酸コードは、次に、DNAライブラリーに利用可能な広範囲の標的DNA濃縮戦略を利用することができる。場合によっては、対応する伸長核酸を濃縮することによって、目的のペプチドをサンプル内で濃縮することができる。標的濃縮の方法は、当技術分野で知られており、ハイブリッド捕捉アッセイ、TruSeqカスタムアンプリコン(Illumina)などのPCRベースのアッセイ、パドロックプローブ(分子反転プローブとも称される)などが含まれる(Mamanova et al., (2010) Nature Methods 7:111-118、Bodi et al., J. Biomol. Tech. (2013) 24:73-86、Ballester et al., (2016) Expert Review of Molecular Diagnostics 357-372、Mertes et al., (2011) Brief Funct. Genomics 10:374-386、Nilsson et al., (1994) Science 265:2085-8を参照、これらの各々は、参照によりそれら全体が本明細書に組み込まれる)。

#### 【0191】

一実施形態において、核酸(例えば、伸長記録タグ)のライブラリーは、ハイブリッド捕捉ベースのアッセイを介して濃縮される。ハイブリッド捕捉ベースのアッセイでは、伸長核酸のライブラリーは、親和性タグ(例えば、ビオチン)で標識された標的特異的オリゴヌクレオチドにハイブリダイズされる。標的特異的オリゴヌクレオチドにハイブリダイズした伸長核酸を、親和性リガンド(例えば、ストレプトアビジンでコーティングされたビーズ)を使用して、親和性タグを介して「プルダウン」し、バックグラウンド(非特異的)伸長核酸を洗い流す。次いで、濃縮された伸長核酸(例えば、伸長核酸)が、陽性濃縮(例えば、ビーズから溶出される)のために得られる。いくつかの実施形態において、目的のペプチドの対応する伸長核酸ライブラリー表現に相補的なオリゴヌクレオチドを、ハイブリッド捕捉アッセイにおいて使用することができる。いくつかの実施形態において、同じ又は異なるbaitセットを用いて、連続的なラウンド又は濃縮を実施することもできる。

#### 【0192】

別の実施形態において、プライマー伸長及びライゲーションベースの媒介増幅濃縮(AmpliSeq、PCR、TruSeq TSCAなど)を使用して、ポリペプチドのサブセットを表すライブラリー要素の濃縮画分を選択及びモジュール化することができる。競合するオリゴヌクレオチドを使用して、プライマーの伸長、ライゲーション、又は増幅の程度を調整することもできる。最も単純な実施において、これは、ユニバーサルプライマーテールを含む標的特異的プライマーと5'ユニバーサルプライマーテールを欠く競合プ

ライマーとの混合物を有することによって達成することができる。最初のプライマー伸長後、5'ユニバーサルプライマー配列を持つプライマーのみを増幅することができる。ユニバーサルプライマー配列がある場合とない場合のプライマーの比率は、増幅される標的の割合を制御する。他の実施形態において、ハイブリダイズするが非伸長性であるプライマーを含めることを使用して、プライマー伸長、ライゲーション、又は増幅を受けるライブラリー要素の割合を調節することができる。

#### 【0193】

標的濃縮法を負の選択モードで使用して、シーケンシングの前にライブラリーから伸長核酸を選択的に除去することもできる。除去することができる望ましくない伸長核酸の例は、例えば、タンパク質、アルブミン、免疫グロブリンなどのために、過剰に豊富なポリペプチド種を表すものである。

10

#### 【0194】

標的にハイブリダイズするがビオチン部分を欠く競合オリゴヌクレオチド *bait* をハイブリッド捕捉ステップで使用して、濃縮された特定の遺伝子座の画分を調節することもできる。競合オリゴヌクレオチド *bait* は、濃縮中にブルダウンされたターゲットの割合を効果的に調節する標準的なビオチン化 *bait* と、ターゲットへのハイブリダイゼーションをめぐって競合する。タンパク質発現の10桁のダイナミックレンジは、特にアルブミンなどの過剰に豊富な種の場合、この競合抑制アプローチを使用して数桁圧縮することができる。したがって、標準的なハイブリッド捕捉と比較して、特定の遺伝子座に対して捕捉されたライブラリー要素の割合は、100%~0%の濃縮に下方調整することができる。

20

#### 【0195】

更に、ライブラリー正準化技法を使用して、伸長核酸ライブラリーから過剰に豊富な種を除去することができる。このアプローチは、トリプシン、LysC、GluCなどの部位特異的プロテアーゼ消化によって生成されたペプチドに由来する定義された長さのライブラリーに最適である。一例では、正準化は、二本鎖ライブラリーを変性させ、ライブラリー要素を再アニーリングさせることによって達成することができる。二分子ハイブリダイゼーション速度論の二次速度定数により、豊富なライブラリー要素は、豊富でない要素よりも迅速に再アニーリングする (Bochman, Paeschke et al., 2012)。ssDNAライブラリー要素は、ヒドロキシアパタイトカラムでのクロマトグラフィ (Vander Noot, et al., 2012, Biotechniques 53:373-380)、又はdsDNAライブラリー要素を破壊するタラバガニ (Shagin et al., (2002) Genome Res. 12:1935-42) からの二重特異性ヌクレアーゼ (DSN) によるライブラリーの処理などの当技術分野で既知の方法を使用して、豊富なdsDNAライブラリー要素から分離することができる。

30

#### 【0196】

支持体に付着する前のポリペプチド及び/又は結果として得られる伸長核酸ライブラリーの画分化、濃縮、及び減算法の任意の組み合わせにより、シーケンシングリードを節約し、豊富でない種の測定を改善することができる。

#### 【0197】

いくつかの実施形態において、核酸 (例えば、伸長核酸) のライブラリーは、ライゲーション又は末端相補的PCRによって連結されて、それぞれ複数の異なる伸長レコーダタグ、伸長コーディングタグ、又はジタグを含む長いDNA分子を作成する (Duet al., (2003) BioTechniques 35:66-72、Muecke et al., (2008) Structure 16:837-841、米国特許第5,834,252号、これらの各々は、その全体が参照により組み込まれる)。この実施形態は、DNAの長い鎖がナノポアシーケンシングデバイスによって分析されるナノポアシーケンシングにとって好ましい。

40

#### 【0198】

いくつかの実施形態において、直接的な単一分子分析は、核酸 (例えば、伸長核酸) に対

50

して実行される（例えば、Harris et al., (2008) Science 320:106-109を参照）。核酸（例えば、伸長核酸）は、フローセル又はフローセル表面への負荷に適合性のあるビーズ（任意選択的に、マイクロセルパターン化される）などの支持体上で直接分析することができ、フローセル又はビーズは、単一分子シーケンサー又は単一分子デコーディング機器と統合され得る。単一分子デコーディングの場合、デコーディングオリゴヌクレオチドのプールされた蛍光標識された数ラウンドのハイブリダイゼーション（Gunderson et al., (2004) Genome Res. 14:970-7）を使用して、（例えば、記録タグ上の）伸長核酸内のコーディングタグの同一性及び順序の両方を確認することができる。いくつかの実施形態において、結合剤は、上記のようにサイクル特異的コーディングタグで標識することができる（Gunderson et al., (2004) Genome Res. 14:970-7も参照）。

10

20

30

40

50

#### 【0199】

核酸ライブラリーの（例えば、伸長核酸の）シーケンシングに続いて、結果として得られる配列は、使用される場合、UMIによって折り畳まれた後、それらの対応するポリペプチドと関連付けられ、プロテオーム全体に整列される。結果として得られる配列はまた、それらのコンパートメントタグによって折り畳まれ、それらの対応するコンパートメントプロテオームと関連付けられることができ、これは、特定の実施形態において、単一又は非常に限られた数のタンパク質分子のみを含む。タンパク質の識別及び定量化はいずれも、このデジタルペプチド情報から簡単に導き出すことができる。

#### 【0200】

本明細書に開示される方法は、複数の高分子の同時（多重化）の検出、定量化、及び/又はシーケンシングを含む分析に使用することができる。本明細書で使用される多重化は、同じアッセイにおける複数の高分子（例えば、ポリペプチド）の分析を指す。複数の高分子は、同じサンプル又は異なるサンプルに由来し得る。複数の高分子は、同じ対象又は異なる対象に由来し得る。分析される複数の高分子は、異なる高分子、又は異なるサンプルに由来する同じ高分子であり得る。複数の高分子は、2個以上の高分子、5個以上の高分子、10個以上の高分子、50個以上の高分子、100個以上の高分子、500個以上の高分子、1000個以上の高分子、5,000個以上の高分子、10,000個以上の高分子、50,000個以上の高分子、100,000個以上の高分子、500,000個以上の高分子、又は1,000,000個以上の高分子を含む。

#### 【0201】

##### III. キット及び製造品

本明細書で提供されるのは、高分子分析アッセイを予備形成するための構成要素を含むキット及び製造品である。いくつかの実施形態において、キットは、結合剤に関する識別情報を含むコーディングタグを含む結合剤を含む。いくつかの態様において、キットは、各々がコーディングタグを含む、複数の又は一組の結合剤を含む。いくつかの実施形態において、結合剤は、記録タグと関連付けられた高分子に結合するように構成され、コーディングタグからの情報を記録タグに転送するための試薬も提供される。キットには、核酸接合試薬、ポリメラーゼ、及び二本鎖核酸切断試薬も提供される。いくつかの態様において、核酸接合試薬、ポリメラーゼ、及び二本鎖核酸切断試薬は、混合物として提供される。例えば、核酸接合試薬は、化学ライゲーション試薬又は酵素ライゲーション試薬である。場合によっては、二本鎖核酸切断試薬は、制限酵素である。

#### 【0202】

いくつかの実施形態において、キットは、標的高分子（例えば、タンパク質、ポリペプチド、又はペプチド）を処理及び分析するための他の試薬を更に含有する。キット及び製造品は、セクションI及びIIに記載される方法で使用される試薬及び成分のうちのいずれかのうちの1つ以上を含み得る。いくつかの実施形態において、キットは、サンプルから高分子を調製し、支持体に接合するためのような、サンプルを調製するための試薬を含む。いくつかの実施形態において、キットは、任意選択的に、高分子分析アッセイを実行す

るための説明書を含む。いくつかの実施形態において、キットは、以下の構成要素のうち  
の1つ以上を含む：結合剤、核酸接合試薬、ポリメラーゼ、二本鎖核酸切断試薬、固体支  
持体、記録タグ、情報を転送するための試薬、シーケンシング試薬、及び/又は任意の必  
要な緩衝液など。記録タグ、結合剤、及びコーディングタグは、セクション I A、I B、  
及び I C にそれぞれ記載されるような特定の構造及び特徴を有し得る。

【0203】

一態様において、本明細書で提供されるのは、反応混合物を調製するために使用される成  
分である。好ましい実施形態において、反応混合物は、溶液である。好ましい実施形態に  
おいて、反応混合物は、以下のうちの1つ以上を含む：結合剤及び関連するコーディング  
タグ、固体支持体、記録タグ、核酸接合試薬、ポリメラーゼ、二本鎖核酸切断試薬、シー  
ケンシング試薬、緩衝液及び/又は他の添加剤を含む情報を転送するための試薬。

10

【0204】

別の態様において、本明細書に開示されるのは、結合剤のライブラリーを含む高分子分析  
アッセイを実行するためのキットであり、各結合剤は、コーディングタグを含むか、又は  
それに関連する。いくつかの態様において、コーディングタグは、結合剤の結合部分に関  
する識別情報を含む。いくつかの例では、結合部分は、標的ペプチド若しくはポリペプ  
チドのうち1つ以上のN末端、内部、若しくはC末端アミノ酸に結合することができるか  
、又は官能化/修飾試薬によって修飾されたペプチドの1つ以上のN末端、内部、若しく  
はC末端アミノ酸に結合することができる。

【0205】

いくつかの実施形態において、キット及び製造品は、複数の核酸分子又はオリゴヌクレオ  
チドを更に含む。いくつかの実施形態において、キットは、複数のバーコードを含む。バ  
ーコードは、コンパートメントバーコード、パーティションバーコード、サンプルバー  
コード、画分バーコード、又はそれらの任意の組み合わせを含み得る。場合によっては、バ  
ーコードは、固有の分子識別子 (UMI) を含む。いくつかの例では、バーコードは、D  
NA分子、偽相補的塩基を有するDNA、RNA分子、BNA分子、XNA分子、LNA  
分子、PNA分子、PNA分子、非核酸シーケンシング可能なポリマー、例えば、多糖  
類、ポリペプチド、ペプチド、若しくはポリアミド、又はそれらの組み合わせを含む。い  
くつかの実施形態において、バーコードは、サンプル内の分析のための高分子、例えば、  
タンパク質に付着するように、又は高分子と関連付けられた核酸成分に付着するように構  
成されている。

20

30

【0206】

いくつかの実施形態において、キットは、高分子、例えば、タンパク質を処理するための  
試薬を更に含む。タンパク質の画分化、濃縮、及び減算法の任意の組み合わせを実行す  
ることができる。例えば、試薬を使用して、タンパク質を断片化又は消化することができる  
。場合によっては、キットは、タンパク質を画分化、単離、減算、濃縮するための試薬及  
び成分を含む。いくつかの例では、キットは、トリプシン、Lys N、又はLys Cなど  
のプロテアーゼを更に含む。いくつかの実施形態において、キットは、1つ以上の標的を  
固定化するための支持体、及び支持体上で標的を固定化するための試薬を含む。いくつ  
かの実施形態において、キットは、化学剤又は酵素剤などの、ポリペプチドのN末端アミノ  
酸 (NTAA) を除去するための試薬を更に含む。いくつかの実施形態において、キット  
は、化学剤又は酵素剤などの、ポリペプチドの末端アミノ酸を修飾するための試薬を更  
に含む。

40

【0207】

いくつかの実施形態において、キットはまた、高分子分析アッセイのステップのいずれか  
を実行するのに必要な1つ以上の緩衝液又は反応流体を含む。洗浄緩衝液、反応緩衝液、  
及び結合緩衝液、溶出緩衝液などを含む緩衝液は、当業者に知られている。いくつ  
かの実施形態において、キットは、本明細書に記載の他の試薬に付随する緩衝液及び他の成分を  
更に含む。試薬、緩衝液、及び他の成分は、バイアル (密封されたバイアルなど)、容器  
、アンプル、ボトル、ジャー、柔軟な包装 (例えば、密封されたマイラー (Mylar))

50

又はビニール袋)などで提供され得る。キットの構成要素はいずれも滅菌及び/又は密封することができる。

【0208】

いくつかの実施形態において、キットは、核酸シーケンシング分析のための1つ以上の試薬を含む。いくつかの例では、配列分析用の試薬は、合成によるシーケンシング、ライゲーションによるシーケンシング、単一分子シーケンシング、単一分子蛍光シーケンシング、ハイブリダイゼーションによるシーケンシング、ポロニーシーケンシング、イオン半導体シーケンシング、パイロシーケンシング、単一分子リアルタイムシーケンシング、ナノポアベースのシーケンシング、若しくは高度な顕微鏡を使用するDNAの直接イメージング、又はそれらの任意の組み合わせにおいて使用するためのものである。

10

【0209】

上記の構成要素に加えて、対象キットは、対象方法を実施するためにキットの構成要素を使用するための説明書、すなわち、サンプル調製、処理及び/又は分析のための説明書を更に含み得る。本明細書に記載のキットはまた、他の緩衝液、希釈剤、フィルター、注射器、及び本明細書に記載の任意の方法を実行するための説明書を伴う添付文書を含む、商業的及びユーザの観点から望ましい他の材料を含み得る。

【0210】

上述のキット構成要素のうちのいずれか、及び任意の分子、分子複合体又は共役体、試薬(例えば、化学的若しくは生物学的試薬)、薬剤、構造(例えば、支持体、表面、粒子、若しくはビーズ)、反応中間体、反応産生物、結合複合体、又は例示的なキット及び方法で開示及び/又は使用される任意の他の製造品は、キットを形成するために別個に又は任意の好適な組み合わせで提供され得る。

20

【0211】

IV. 例示的な実施形態

提供される実施形態の中には、以下のものがある。

1. 高分子を分析するための方法であって、

(a) 高分子及び支持体に接合された関連する記録タグを提供するステップと、

(b) 高分子を、高分子に結合することができる結合剤であって、結合剤に関する識別情報を有するコーディングタグを含む結合剤と接触させて、高分子と結合剤との間の結合を可能にするステップと、

30

(c) 核酸接合試薬によって、記録タグの5'末端をコーディングタグの3'末端に接合するステップと、

(d) ポリメラーゼによって、コーディングタグを鋳型として使用して記録タグを伸長し、二本鎖伸長記録タグを生成するステップと、

(e) 二本鎖伸長記録タグを二本鎖核酸切断試薬で切断して、伸長記録タグ内に3'オーバーハングを生成するステップと、

それによって、情報が、コーディングタグから記録タグに転送されて、伸長記録タグを生成するステップと、を含む、方法。

2. ステップ(d)において、二本鎖伸長記録タグが、二本鎖核酸切断試薬によって認識することができる認識配列を含む、実施形態1に記載の方法。

40

3. ステップ(e)における切断が、高分子から結合剤を放出する、実施形態1又は実施形態2に記載の方法。

4. 記録タグ及びコーディングタグが、核酸を含む、実施形態1~3のいずれか1つに記載の方法。

5. 伸長記録タグが、核酸ヘアピンを含む、実施形態1~4のいずれか1つに記載の方法。

6. ステップ(a)、(b)、(c)、(d)及び(e)が、連続的に実行される、実施形態1~5のいずれか1つに記載の方法。

7. ステップ(b)、(c)、(d)、及び(e)が、周期的な方法で連続1回以上繰り返される、実施形態1~6のいずれか1つに記載の方法。

50

- 8．高分子が、脂質、炭水化物、タンパク質、ポリペプチド、又はペプチドである、実施形態 1 ~ 7 のいずれか 1 つに記載の方法。
- 9．ペプチドが、生物学的サンプルからタンパク質を断片化することによって得られる、実施形態 8 に記載の方法。
- 10．分析のための高分子が、核酸ではない、実施形態 1 ~ 9 のいずれか 1 つに記載の方法。
- 11．ステップ (b) を繰り返す前に、高分子の一部を除去することを更に含む、実施形態 7 ~ 10 のいずれか 1 つに記載の方法。
- 12．ステップ (b) を繰り返す前に、ポリペプチドの N 末端アミノ酸 (NTAA) を除去して、ポリペプチドの新しい NTAA を露出させることを更に含む、実施形態 8 ~ 11 のいずれか 1 つに記載の方法。
- 13．ステップ (e) において、二本鎖核酸切断試薬によって生成された伸長記録タグの 3' オーバーハングが、ステップ (b) が繰り返されるときに、第 2 のコーディングタグとハイブリダイズするために利用可能である、実施形態 7 ~ 12 のいずれか 1 つに記載の方法。
- 14．結合剤を除去することを更に含む、実施形態 1 ~ 13 のいずれか 1 つに記載の方法。
- 15．結合剤が、コーディングタグの情報を記録タグに転送した後に除去される、実施形態 14 に記載の方法。
- 16．結合剤が、ステップ (b) を繰り返す前に除去される、実施形態 14 又は実施形態 15 に記載の方法。
- 17．方法が、ステップ (b) において、複数の高分子を、単一の結合剤又は複数の結合剤と接触させることを含む、実施形態 1 ~ 16 のいずれか 1 つに記載の方法。
- 18．ポリペプチドの末端アミノ酸を修飾するための試薬でポリペプチドを処理することを更に含む、実施形態 8 ~ 17 のいずれか 1 つに記載の方法。
- 19．ポリペプチドの末端アミノ酸を修飾するための試薬が、化学剤又は酵素剤を含む、実施形態 18 に記載の方法。
- 20．ポリペプチドが、ステップ (b) の前に、ポリペプチドの末端アミノ酸を修飾するための試薬で処理される、実施形態 18 又は実施形態 19 に記載の方法。
- 21．高分子に関連付けられた記録タグが、二本鎖領域を含む、実施形態 1 ~ 20 のいずれか 1 つに記載の方法。
- 22．高分子に関連付けられた記録タグが、核酸ヘアピンを含む、実施形態 21 に記載の方法。
- 23．高分子に関連付けられた記録タグが、3' オーバーハングを有する、実施形態 1 ~ 22 のいずれか 1 つに記載の方法。
- 24．記録タグが、バーコードを含む、実施形態 1 ~ 23 のいずれか 1 つに記載の方法。
- 25．記録タグが、固有の分子識別子 (UMI) を含む、実施形態 1 ~ 24 のいずれか 1 つに記載の方法。
- 26．ステップ (a) が、バーコード及び / 又は UMI を記録タグに提供することを更に含む、実施形態 24 又は実施形態 25 に記載の方法。
- 27．バーコードが、サンプルバーコード、画分バーコード、空間バーコード、及び / 又はコンパートメントタグである、実施形態 24 又は実施形態 26 に記載の方法。
- 28．ステップ (a) が、ライゲーション及び / 又は伸長を使用して、バーコード及び / 又は UMI を記録タグに提供することを更に含む、実施形態 24 又は実施形態 25 に記載の方法。
- 29．ステップ (a) が、記録タグを切断して、3' オーバーハングを生成することを更に含む、実施形態 23 ~ 28 のいずれか 1 つに記載の方法。
- 30．記録タグの 3' オーバーハングが、二本鎖核酸切断試薬による伸長及び / 又は切断によって生成される、実施形態 23 ~ 29 のいずれか 1 つに記載の方法。
- 31．記録タグが、ユニバーサルプライミング部位を含む、実施形態 1 ~ 30 のいずれか

1 つに記載の方法。

32. ユニバーサルプライミング部位が、増幅、シーケンシング、又はその両方のためのプライミング部位を含む、実施形態 31 に記載の方法。

33. ステップ (c)、(d)、及び (e) が、ワンポット反応として実行される、実施形態 1 ~ 32 のいずれか 1 つに記載の方法。

34. ステップ (c)、(d)、及び (e) の核酸接合試薬、ポリメラーゼ、及び二本鎖核酸切断試薬が、それぞれ、混合物として提供される、実施形態 33 に記載の方法。

35. ステップ (c)、(d)、及び (e) が、連続的かつ別々に実行される、実施形態 1 ~ 32 のいずれか 1 つに記載の方法。

36. ステップ (c) の核酸接合試薬及びステップ (d) のポリメラーゼが、同時に提供される、実施形態 1 ~ 32 のいずれか 1 つに記載の方法。 10

37. ステップ (d) のポリメラーゼ及びステップ (e) の二本鎖核酸切断試薬が、同時に提供される、実施形態 1 ~ 32 のいずれか 1 つに記載の方法。

38. コーディングタグが、部分的な制限酵素認識配列を含む、実施形態 1 ~ 37 のいずれか 1 つに記載の方法。

39. コーディングタグの部分的制限酵素認識配列が、一本鎖である、実施形態 38 に記載の方法。

40. コーディングタグが、バーコード及び / 又は固有の分子識別子 (UMI) を含む、実施形態 1 ~ 39 のいずれか 1 つに記載の方法。

41. 記録タグ及びコーディングタグが、各々、スペーサーを含む、実施形態 1 ~ 40 のいずれか 1 つに記載の方法。 20

42. スペーサーが、10 塩基以下、9 塩基以下、8 塩基以下、7 塩基以下、6 塩基以下、5 塩基以下、4 塩基以下、3 塩基以下、又は 2 塩基以下の核酸分子である、実施形態 41 に記載の方法。

43. 記録タグ又は伸長記録タグが、二本鎖核酸切断試薬によって生成される切断部位オーバーハングであるスペーサーを含む、実施形態 41 又は実施形態 42 に記載の方法。

44. スペーサーが、サイクル特異的スペーサー又はサイクル交互スペーサーである、実施形態 41 ~ 43 のいずれか 1 つに記載の方法。

45. 方法が、互換性のないオーバーハングに起因する 1 サイクルの情報転送の後に自己終了する、実施形態 44 に記載の方法。 30

46. 前のコーディングタグによって追加されたスペーサーが、第 2 のコーディングタグのスペーサーの少なくとも一部と一致する場合、情報が、第 2 のコーディングタグから伸長記録タグに転送される、実施形態 41 ~ 45 のいずれか 1 つに記載の方法。

47. 記録タグ及びコーディングタグが、スペーサーを含まない、実施形態 1 ~ 47 のいずれか 1 つに記載の方法。

48. コーディングタグが、反応に利用可能な 3' 末端を有する、実施形態 1 ~ 47 のいずれか 1 つに記載の方法。

49. 結合剤が、コーディングタグの 5' 末端に付着される、実施形態 48 に記載の方法。

50. 方法が、1 つ以上の洗浄ステップを含む、実施形態 1 ~ 49 のいずれか 1 つに記載の方法。 40

51. 洗浄ステップが、ステップ (c) の前に実行される、実施形態 50 に記載の方法。

52. 洗浄ステップが、ステップ (d) の前に実行される、実施形態 50 に記載の方法。

53. 洗浄ステップが、ステップ (e) の前に実行される、実施形態 50 に記載の方法。

54. 核酸接合試薬が、化学ライゲーション試薬又は酵素ライゲーション試薬である、実施形態 1 ~ 53 のいずれか 1 つに記載の方法。

55. 酵素核酸接合試薬が、リガーゼである、実施形態 54 に記載の方法。

56. 二本鎖核酸切断試薬が、制限酵素である、実施形態 1 ~ 55 のいずれか 1 つに記載の方法。

57. 制限酵素が、IIS 型制限酵素である、実施形態 55 に記載の方法。

58. IIS 型制限酵素が、

5 ' ... G C A G T G N N ... 3 '

3 ' ... C G T C A C N N ... 5 ' を含む一連の塩基を認識する、実施形態 5 7 に記載の方法。

5 9 . 制限酵素が、N b . B t s I 又は B t s I - v 2 又はその誘導体である、実施形態 5 7 又は実施形態 5 8 に記載の方法。

6 0 . ステップ ( c )、( d )、及び ( e ) のライゲーシオン、伸長、及び切断が、それぞれ、段階的な方法で起こる、実施形態 1 ~ 5 9 のいずれか 1 つに記載の方法。

6 1 . 記録タグが、核酸接合試薬によってコーディングタグに接合された後に、結合剤が、高分子に結合したままではない、実施形態 1 ~ 6 0 のいずれか 1 つに記載の方法。

6 2 . 最終情報転送サイクルが、伸長記録タグにキャッピング配列を提供するために実行され、任意選択的に、キャッピング配列が、増幅、シーケンシング、又はその両方のためのユニバーサルプライミング部位を含む、実施形態 1 ~ 6 1 のいずれか 1 つに記載の方法。

6 3 . キャッピング配列が、高分子の普遍的な特徴に結合することができる結合剤と関連付けられたコーディングタグで提供される、実施形態 6 2 に記載の方法。

6 4 . 普遍的特徴が、ポリペプチドの N 末端アミノ酸の化学修飾である、実施形態 6 3 に記載の方法。

6 5 . 伸長再コーディングタグが、複数のコーディングタグから転送される一連の情報を含む、実施形態 8 ~ 6 4 のいずれか 1 つに記載の方法。

6 6 . 伸長記録タグのうちの 1 つ以上を分析することを更に含む、実施形態 1 ~ 6 5 のいずれか 1 つに記載の方法。

6 7 . 伸長記録タグのうちの 1 つ以上が、分析の前に増幅される、実施形態 6 6 に記載の方法。

6 8 . 伸長記録タグを分析することが、核酸シーケンシング方法を含む、実施形態 6 6 又は実施形態 6 7 に記載の方法。

6 9 . 核酸シーケンシング方法が、合成によるシーケンシング、ライゲーシオンによるシーケンシング、ハイブリダイゼーションによるシーケンシング、ポロニーシーケンシング、イオン半導体シーケンシング、パイロシーケンシング、単一分子リアルタイムシーケンシング、ナノポアベースのシーケンシング、又は高度な顕微鏡を使用する D N A の直接イメージングである、実施形態 6 8 に記載の方法。

7 0 . 結合剤が、ポリペプチド又はタンパク質である、実施形態 1 ~ 6 9 のいずれか 1 つに記載の方法。

7 1 . 結合剤が、アミノペプチダーゼ又はそれらのバリエーション、変異体、若しくは修飾タンパク質、アミノアシル t R N A シンターゼ又はそれらのバリエーション、変異体、若しくは修飾タンパク質、アンチリン又はそれらのバリエーション、変異体、若しくは修飾タンパク質、C l p S、C l p S 2 又はそれらのバリエーション、変異体、若しくは修飾タンパク質、U B R ボックスタンパク質又はそれらのバリエーション、変異体、若しくは修飾タンパク質、又はアミノ酸に結合する修飾された小分子、すなわちバンコマイシン若しくはそのバリエーション、変異体、若しくは修飾された分子、又は抗体若しくはその結合断片、又はそれらの任意の組み合わせである、実施形態 7 0 に記載の方法。

7 2 . 結合剤が、ポリペプチド高分子の単一のアミノ酸残基、ジペプチド、トリペプチド、又は翻訳後修飾に結合する、実施形態 8 ~ 7 1 のいずれか 1 つに記載の方法。

7 3 . 結合剤が、ポリペプチドの N 末端アミノ酸残基に結合するように構成されている、実施形態 7 2 に記載の方法。

7 4 . 結合剤が、ポリペプチドの化学的に修飾又は標識された N 末端アミノ酸残基に結合するように構成されている、実施形態 7 3 に記載の方法。

7 5 . 高分子分析のためのキットであって、

コーディングタグを含む結合剤であって、コーディングタグが、結合剤に関する識別情報を含む、結合剤と、

核酸接合試薬と、

ポリメラーゼと、

10

20

30

40

50

二本鎖核酸切断試薬と、を含み、

結合剤が、記録タグと関連付けられた高分子に結合するように構成され、コーディングタグからの識別情報が、コーディングタグから高分子と関連付けられた記録タグへの転送のために構成されている、キット。

76．高分子が、脂質又は炭水化物である、実施形態75に記載のキット。

77．高分子が、タンパク質、ポリペプチド、又はペプチドである、実施形態75に記載のキット。

78．ポリペプチドのN末端アミノ酸（NTAA）を除去するための試薬を更に含む、実施形態77に記載のキット。

79．ポリペプチドの末端アミノ酸を除去するための試薬が、化学剤又は酵素剤を含む、実施形態78に記載のキット。 10

80．ポリペプチドの末端アミノ酸を修飾するための試薬を更に含む、実施形態77～79のいずれか1つに記載のキット。

81．ポリペプチドの末端アミノ酸を修飾するための試薬が、化学剤又は酵素剤を含む、実施形態80に記載のキット。

82．高分子に関連付けられた記録タグが、二本鎖領域を含む、実施形態75～81のいずれか1つに記載のキット。

83．高分子と関連付けられた記録タグが、核酸ヘアピンを含む、実施形態82に記載のキット。

84．高分子に関連付けられた記録タグが、3'オーバーハングを有する、実施形態75～83のいずれか1つに記載のキット。 20

85．記録タグが、バーコードを含む、実施形態75～84のいずれか1つに記載のキット。

86．バーコードが、サンプルバーコード、画分バーコード、空間バーコード、及び/又はコンパートメントタグである、実施形態85に記載のキット。

87．記録タグが、固有の分子識別子（UMI）を含む、実施形態75～86のいずれか1つに記載のキット。

88．記録タグが、ユニバーサルプライミング部位を含む、実施形態75～87のいずれか1つに記載のキット。

89．ユニバーサルプライミング部位が、増幅、シーケンシング、又はその両方のためのプライミング部位を含む、実施形態88に記載のキット。 30

90．キットが、核酸接合試薬、ポリメラーゼ、及び二本鎖核酸切断試薬を含む混合物を含む、実施形態75～89のいずれか1つに記載のキット。

91．コーディングタグが、部分的な制限酵素認識配列を含む、実施形態75～90のいずれか1つに記載のキット。

92．コーディングタグの部分的制限酵素認識配列が、一本鎖である、実施形態91に記載のキット。

93．コーディングタグが、バーコード及び/又は固有の分子識別子（UMI）を含む、実施形態75～92のいずれか1つに記載のキット。

94．記録タグ及びコーディングタグが、各々、スペーサーを含む、実施形態75～93のいずれか1つに記載のキット。 40

95．スペーサーが、10塩基以下、9塩基以下、8塩基以下、7塩基以下、6塩基以下、5塩基以下、4塩基以下、3塩基以下、又は2塩基以下の核酸分子である、実施形態94に記載のキット。

96．スペーサーが、サイクル特異的スペーサー又はサイクル交互スペーサーである、実施形態94又は実施形態95に記載のキット。

97．記録タグ及びコーディングタグが、スペーサーを含まない、実施形態75～93のいずれか1つに記載のキット。

98．コーディングタグが、反応に利用可能な3'末端を有する、実施形態75～97のいずれか1つに記載のキット。 50

99. 酵素核酸接合試薬が、リガーゼである、実施形態75~98のいずれか1つに記載のキット。
100. 二本鎖核酸切断試薬が、制限酵素である、実施形態75~99のいずれか1つに記載のキット。
101. 制限酵素が、IIS型制限酵素である、実施形態100に記載のキット。
102. IIS型制限酵素が、  
5' ... G C A G T G N N ... 3'  
3' ... C G T C A C N N ... 5' を含む一連の塩基を認識する、実施形態101に記載のキット。
103. 制限酵素が、Nb.BtsI又はBtsI-v2又はその誘導体である、実施形態101又は実施形態102に記載のキット。 10
104. 結合剤が、キャッピング配列を含む、高分子の普遍的特徴に結合することができる結合剤を更に含む、実施形態75~103のいずれか1つに記載のキット。
105. キットが、複数の結合剤を含む混合物を含む、実施形態75~104のいずれか1つに記載のキット。
106. 結合剤が、ポリペプチド又はタンパク質である、実施形態75~105のいずれか1つに記載のキット。
107. 結合剤が、アミノペプチダーゼ又はそれらのバリエーション、変異体、若しくは修飾タンパク質、アミノアシルtRNAシントナーゼ又はそれらのバリエーション、変異体、若しくは修飾タンパク質、アンチカリン又はそれらのバリエーション、変異体、若しくは修飾タンパク質、C1pS、C1pS2又はそれらのバリエーション、変異体、若しくは修飾タンパク質、UBRボックスタンパク質又はそれらのバリエーション、変異体、若しくは修飾タンパク質、又はアミノ酸に結合する修飾された小分子、すなわちバンコマイシン若しくはそのバリエーション、変異体、若しくは修飾された分子、又は抗体若しくはその結合断片、又はそれらの任意の組み合わせである、実施形態106に記載のキット。 20
108. 結合剤が、ポリペプチド高分子の単一のアミノ酸残基、ジペプチド、トリペプチド、又は翻訳後修飾に結合する、実施形態77~107のいずれか1つに記載のキット。
109. 結合剤が、ポリペプチドのN末端アミノ酸残基に結合するように構成されている、実施形態108に記載のキット。
110. 結合剤が、ポリペプチドの末端アミノ酸を修飾するために試薬によって修飾される、ポリペプチドのN末端アミノ酸残基に結合するように構成されている、実施形態80~109のいずれか1つに記載のキット。 30
111. 1つ以上の洗浄緩衝液を更に含む、実施形態75~110のいずれか1つに記載のキット。
112. 高分子及び/又は記録タグを固定化するための支持体を更に含む、実施形態75~111のいずれか1つに記載のキット。
113. 支持体が、三次元支持体(例えば、多孔質マトリックス又はビーズ)である、実施形態112に記載のキット。
114. 支持体が、ポリスチレンビーズ、ポリアクリレートビーズ、ポリマービーズ、アガロースビーズ、セルロースビーズ、デキストランビーズ、アクリルアミドビーズ、固体コアビーズ、多孔質ビーズ、常磁性ビーズ、ガラスビーズ、制御された多孔質ビーズ、シリカベースのビーズ、又はそれらの任意の組み合わせを含む、実施形態113に記載のキット。 40

#### 【実施例】

#### 【0212】

#### V. 実施例

以下の実施例は、本明細書で提供される方法、組成物、及び使用を説明するために提供されるが、これらを限定するものではない。Proteocode(商標)ポリペプチドシークエンシングアッセイの実施形態、コーディングタグと記録タグとの間の情報転送、ヌクレオチド-ポリペプチド共役体の作製方法、ヌクレオチド-ポリペプチド共役体の支持体 50

への付着方法、バーコードの生成方法、ポリペプチドのN末端アミノ酸を認識する特異的結合剤の生成方法、試薬、並びにポリペプチドからN末端アミノ酸を修飾及び/又は除去する方法、伸長記録タグを分析する方法を含むが、これらに限定されない本発明の特定の態様は、先に公開された出願US 2019/0145982 A1、US 2020/0348308 A1、US 2020/0348307 A1、US 2021/0208150 A1、WO 2020/223000に開示され、それらの内容全体が参照により本明細書に組み込まれる。

#### 【0213】

実施例1．核酸ハイブリダイゼーション及び固体支持体への接合を使用した分析物固定化の評価。

10

この実施例は、核酸-ポリペプチド共役体を固体支持体に接合(固定化)するための例示的な方法を説明する。

#### 【0214】

ハイブリダイゼーションベースの固定化方法では、核酸-ポリペプチド複合体をハイブリダイズし、磁気ビーズ上に化学的に固定化されたヘアピン捕捉DNAにライゲーションした。捕捉核酸を、トランス-シクロオクテン(TCO)及びメチルテトラジン(mTet)ベースのクリック化学を使用してビーズに共役した。TCOで修飾された短いヘアピン捕捉核酸(16塩基対ステム、5塩基ループ、24塩基5'オーバーハング)を、mTetでコーティングされた磁気ビーズと反応させた。リン酸化された核酸-ポリペプチド複合体(10nM)を、ビーズに付着したヘアピンDNAに5xSSC、0.02%SDSでアニーリングし、37°Cで30分間インキュベートした。ビーズをPBSTで1回洗浄し、T4 DNAリガーゼを含む1xQuick ligation溶液(New England Biolabs, USA)に再懸濁した。25°Cで30分間インキュベートした後、ビーズをPBSTで2回洗浄し、50µLのPBSTに再懸濁した。アミノFA末端ペプチド(FAAGVAMPGAEDDVVGS GSK、配列番号3)、アミノAFA末端ペプチド(AFAAGVAMPGAEDDVVGS GSK、配列番号4)、及びアミノAA末端ペプチド(AAGVAMPGAEDDVVGS GSK、配列番号5)を含む固定化された全核酸-ペプチド共役体を、特定のプライマーセットを使用したqPCRによって定量化した。比較のために、ライゲーションステップを伴わない非ハイブリダイゼーションベースの方法を使用して、ペプチドをビーズに固定化した。非ハイブリダイゼーションベースの方法は、アミノFA末端ペプチド、アミノAFA末端ペプチド、及びアミノAA末端ペプチドを含む30µMのTCOで修飾されたDNAタグ付きペプチドを、mTetでコーティングされた磁気ビーズとともに25°Cで一晩インキュベートすることによって実行した。

20

30

#### 【0215】

表1に示されるように、1:100,000のグラフト密度の非ハイブリダイゼーション調製法、及び1:10,000のグラフト密度のハイブリダイゼーションベースの調製法で同様のCt値が観察された。ハイブリダイゼーションベースの調製法のDNAタグ付きペプチドのローディング量は、非ハイブリダイゼーション調製法の場合と比較して、1/3000であった。一般に、ハイブリダイゼーションベースの固定化法に必要な出発物質が少ないことが観察された。

40

【表 1】

表1:ローディングハイブリダイゼーションと非ハイブリダイゼーション固定化法との比較		
グラフト:不動態化	非ハイブリダイゼーションベースの固定化法 (-ライゲーション)	ハイブリダイゼーションベースの固定化法(+ライゲーション)
1:100,000	19.4	25.4
1:10,000	-	21.1

10

## 【0216】

実施例 2 . 例示的な連続エンコーディングアッセイ。

この実施例において、固体支持体上に固定化された核酸 - ポリペプチド共役体に特異的に結合するために、2つの例示的な結合剤を使用した。一方の結合剤は、ポリペプチドのN末端フェニルアラニン残基 (F - 結合剤、31 - F) に結合し、他方の結合剤は、ポリペプチドのN末端ロイシン残基 (L - 結合剤、44 - L) に結合する。両方の結合剤は、US 2021 / 0208150 A 1の実施例 1 に具体的に記載されているように、指向性進化によってリポカリン足場から操作された。結合剤を、結合剤に関する識別情報を有するバーコードを含む対応する核酸コーディングタグに共役した。各結合剤に特異的なコーディングタグを、PEGリンカーを介して Spy Tag に付着させ、得られた融合物を、本質的にUS 2021 / 0208150 A 1に記載されているように、Spy Tag - Spy Catcher 相互作用を介して結合剤 - Spy Catcher 融合タンパク質と反応させた。

20

## 【0217】

エンコーディングアッセイのために、2つの試験高分子 (F S G V A R G D V R G G K (アジド)、以下、F - ペプチド、配列番号 6 及び L A E S A F S G V A R G D V R G G K (アジド)、以下、L - ペプチド、配列番号 7) を、アルキン - アジド反応 (本質的に実施例 1 に記載されるように) を介して固定化ビーズ付着捕捉 DNA (配列番号 8) に接合した。DNA - ポリペプチド共役体 (20 nM) を、ビーズに付着した捕捉 DNA に 5 x SSC、0.02% SDS でアニーリングし、37 °C で 30 分間インキュベートした。ビーズを PBST で 1 回洗浄し、T4 DNA リガーゼを含む 1 x Quick Ligation 溶液 (New England Biolabs, USA) に再懸濁した。25 °C で 30 分間インキュベートし、ビーズを、PBST、2 回の 0.1 M の NaOH + 0.1% の Tween (登録商標) 20 及び 2 回の PBST で洗浄した。共役体中の記録タグは、高分子、2 nt のオーバーハング相補領域、II 型制限酵素結合領域、及び隣接領域のバーコードを含む。ハイブリダイゼーション及びライゲーション後、サンプルを、クレノウ断片 (3' -> 5' エキソ-) (MCLAB, USA)、Bts I - V2 (0.5 単位 / ul、New England Biolabs, USA)、dNTP (各 125 uM) 、及び Cut Smart 緩衝液 (50 mM 酢酸カリウム、20 mM 酢酸トリス、10 mM 酢酸マグネシウム、100 µg / ml の BSA、pH 7.9、New England Biolabs, USA) と 25 °C で 30 分間インキュベートし、PBST、2 回の 0.1 M の NaOH + 0.1% Tween (登録商標) 20 及び 2 回の PBST で洗浄した。その結果、3' で 2 nt オーバーハングを生成した (図 1 A 参照)。

30

40

## 【0218】

ビーズを、DMA 中の 18 mM の PMI 試薬 (ピラゾールメタニミン、4 - トリフルオロメチル - 1H - ピラゾール) 及び MOPS 混合物 (60% DMA 及び 40% MOPS、pH 7.6) で 25 °C で 30 分間処理して、固定化されたペプチドの N 末端を修飾し、PBST で 3 回洗浄した。F - 結合剤及び L - 結合剤に付着されたコーディングタグは、各々、3' で 8 bp の二本鎖及び 2 nt のオーバーハングを有するループを形成し、これは、ビーズ上の記録タグの 3' オーバーハングに相補的である。コーディングタグは、結合剤

50

を識別するための固有のバーコードを含み、次の結合サイクルのための B t s I - V 2 結合配列及び 2 n t の相補的オーバーハング領域も有する。2つの結合剤（各 1 0 0 n M）を、q u i c k L i g a s e 緩衝液（6 6 m M の T r i s - H C l、1 0 m M の M g C l<sub>2</sub>、1 m M のジチオスレイトール、1 m M の A T P、7 . 5 % のポリエチレングリコール（P E G 6 0 0 0）、7 . 6 の p H）中の 0 . 1 2 5 単位 / u L のクレノウ断片（3 ' - > 5 ' エキソ - ）（M C L A B , U S A）、d N T P 混合物（各 1 2 5 u M）、0 . 5 単位 / u L の B t s I - V 2、並びに異なる濃度の T 4 D N A リガーゼ（N e w E n g l a n d B i o l a b s , U S A）の存在下で、2 5 ° C で 1 5 分間、ビーズとともにインキュベートし、続いて、1 回の P B S T、2 回の 0 . 1 M の N a O H + 0 . 1 % T w e e n（登録商標）2 0、及び 2 回の P B S T で洗浄した。結合中、記録タグの伸長が発生し、コーディングタグからのバーコード情報が記録タグに転送され、伸長記録タグが形成された（図 1 B を参照）。

10

#### 【 0 2 1 9 】

結合サイクルの後、キャッピングを実行して、分析のために伸長記録タグを増幅する下流 P C R 用のプライマー部位を導入した。伸長又は非伸長の記録タグの 3 ' オーバーハングに相補的な 2 n t の 3 ' オーバーハングを有するループ DNA を含有するキャッピングオリゴを導入した。別々のキャッピングステップを実施する代わりに、相補的プライマー配列を含有するより長いコーディングタグを使用して、伸長反応中に下流 P C R のプライマー部位を導入することができる。4 0 0 n M のキャッピングオリゴを、1 2 . 5 単位 / u L の T 4 D N A リガーゼの存在下で、q u i c k L i g a s e 緩衝液（6 6 m M の T r i s - H C l、1 0 m M の M g C l<sub>2</sub>、1 m M のジチオスレイトール、1 m M の A T P、7 . 5 % のポリエチレングリコール（P E G 6 0 0 0）、7 . 6 の p H）中のビーズとともに 2 5 ° C で 1 5 分間インキュベートした。サンプルを、1 回の P B S T、2 回の 0 . 1 M の N a O H + 0 . 1 % の T w e e n（登録商標）2 0、及び 2 回の P B S T で洗浄した。アッセイの伸長記録タグを P C R 増幅にかけ、次世代シーケンシング（N G S）によって分析し、高分子と相互作用した結合剤に関するバーコード情報を明らかにした。

20

#### 【 0 2 2 0 】

記載されたエンコーディング方法によって生成された例示的なエンコーディング結果を図 2 に示す。ライゲーションステップでは、3つの異なる濃度の T 4 D N A リガーゼ（0 . 1 2 5、1 . 2 5、1 2 . 5 単位 / u l）を試験した。ライゲーション後、伸長ステップのためにクレノウ断片を添加し、切断ステップのために B t s I - V 2 酵素を添加した。エンコードされた記録タグの画分（ビーズ上の記録タグの総量に対する伸長記録タグの割合（伸長及び非伸長））を N G S シーケンシングによって評価し、両方の結合剤についての特定のエンコーディング結果を示した。

30

#### 【 0 2 2 1 】

実施例 3 . 他のタイプの高分子のためのエンコーディングアッセイ。

図 2 に示されるデータは、伸長記録タグの分析を表し、F - 結合剤又は L - 結合剤がペプチド高分子に結合した後、コーディングタグからのバーコード情報の記録タグへの転送が成功したことを確認する。記載された技術は、脂質、炭水化物、又は大環状分子などの他のタイプの高分子に対して採用することができる。エンコーディングアッセイを実行するために、そのような高分子は、固体支持体（ビーズなど）上に固定化され、核酸記録タグと関連付けられる必要がある。エンコーディングステップは、固定化された高分子の種類に関係なく同じままである。記録タグとの会合は、直接的（共有結合性結合など）又は間接的（固体支持体を介した会合など）であり得る。後者の場合、記録タグは、エンコーディングアッセイ中に高分子と共局在化するか、又は近接している必要がある。結合剤は、高分子の成分に特異的に結合するように選択することができる。各結合剤は、結合剤に関する識別情報を有するバーコードを含有する対応する核酸コーディングタグに共役体する必要がある。エンコーディング中に、バーコード情報が高分子と関連付けられた記録タグに転送され、伸長記録タグが生成され、高分子の結合履歴が伸長記録タグ内に記録される。高分子と相互作用する異なる結合剤を使用して、結合サイクルを、別々に、又は混合物

40

50

中のいずれかで、複数回繰り返すことができる。以下では、炭水化物、脂質、又は大環状分子などの異なるタイプの高分子のための開示されたエンコーディングアッセイの適合に利用することができる、当技術分野で知られている代表的な方法が開示されている。

#### 【0222】

まず、炭水化物、脂質、又は大環状分子の成分に特異的に結合することができる例示的な結合剤が知られている。例えば、レクチンは、遊離炭水化物又は糖タンパク質のグリカンエピトープを選択的に認識することができ、炭水化物を含有する高分子のための特異的結合剤として利用することができる、炭水化物結合タンパク質である。重要なことに、マンノース結合レクチン、ガラクトース/N-アセチルグルコサミン結合レクチン、シアル酸/N-アセチルグルコサミン結合レクチン、フコース結合レクチン（例えば、WO2012/049285A1に開示されている）などの炭水化物の異なる成分を認識する既知のレクチンが存在する。また、脂質結合タンパク質は周知であり、結合剤として利用することができる（例えば、Bernlohr DA, et al., *Intracellular lipid-binding proteins and their genes*. *Annu Rev Nutr*. 1997; 17: 277-303を参照）。脂質結合抗体は、一般的に知られており、脂質を含有する高分子の結合剤として利用することができる（例えば、Alving CR. *Antibodies to lipids and liposomes: immunology and safety*. *J Liposome Res*. 2006; 16(3): 157-66を参照）。更に、大環状分子に特異的に結合するタンパク質も知られている（例えば、Villar EA, et al., *How proteins bind macrocycles*. *Nat Chem Biol*. 2014 Sep; 10(9): 723-31; Hunter TM, et al., *Protein recognition of macrocycles: binding of anti-HIV metallocyclams to lysozyme*. *Proc Natl Acad Sci U S A*. 2005 Feb 15; 102(7): 2288-92を参照）。

#### 【0223】

第二に、例示的な炭水化物検出エンコーディングアッセイは、当技術分野で知られている方法を利用して、以下のように実行することができる。

#### 【0224】

アプローチI. 還元アミノ化 Yang SJ, Zhang H. *Glycan analysis by reversible reaction to hydrazide beads and mass spectrometry*. *Anal Chem*. 2012; 84(5): 2232-2238に基づく。

(a) 固定化された記録タグ付着炭水化物共役体を生成する。

炭水化物を過ヨウ素酸ナトリウムで酸化してアルデヒドを生成する。アミンで終端したDNA記録タグを共役し、シアノ水素化ホウ素ナトリウムを使用して、結果として得られるイミンを還元して、炭水化物記録タグ共役体を生成する。好ましくは、ヒドラジド、アルコキシアミン、又は同様の反応性DNAを使用して、還元剤を必要としないより安定な反応生成物（例えば、ヒドラゾン）を生成し得る。実施例2に記載されるように、DNA記録タグを介して、DNAカップリングされた炭水化物を固体支持体に固定する。

(b) 前述のSpy Catcher-コンカナバリンA (ConA) 融合を利用することによって、レクチン-DNAコーディングタグ（結合剤コーディングタグ）共役体を生成する。コーディングタグは、ConAに関する識別情報を含むバーコードを含有する。

(c) 実施例2に記載されるように、レクチン関連コーディングタグから記録タグにバーコード情報を転送することにより、炭水化物がConAに結合する成分を含有するかどうかを分析する。

#### 【0225】

アプローチII. ジアゾカップリング (Matsuura K, et al., *Facile synthesis of stable and lectin-recog*

nizable DNA-carbohydrate conjugates via diazo coupling. *Bioconjug Chem.* 2000 Mar-Apr; 11(2): 202-11に基づく)。アプローチ I Iにおいて、ステップ (a) (記録タグ付着炭水化物共役体の固定化)は、以下のように実行することができる。1)炭水化物を水中で炭酸水素アンモニウムでアミノ化し、 $\alpha$ -グリコシルアミンを生成する、2)アミンをニトロフェニル官能基を有するカルボキシレート誘導體とアミドに変換する。パラジウム触媒上にニトロ基を水素化し、 $\text{NaNO}_2$ 及び $\text{HCl}$ で処理してジアゾ化合物を提供する。

ステップ (b)及び (c)は、アプローチ Iと同じである。

【0226】

第三に、例示的な脂質検出エンコーディングアッセイは、当技術分野で知られている方法を利用して、以下のように実行することができる。

【0227】

アプローチ I . 脂肪酸 (Hiroshi Miwa, High-performance liquid chromatographic determination of free fatty acids and esterified fatty acids in biological materials as their 2-nitrophenylhydrazides, *Analytica Chimica Acta*, Volume 465, Issues 1-2, 2002, Pages 237-255, ISSN 0003-2670に基づく)。

(a)生物学的供給源から脂肪酸を抽出し、EDC/CDI化学でカルボン酸を活性化する。アミン末端又はヒドラジド末端のDNA記録タグをカップリングして、記録タグ付着脂質共役体を生成する。実施例2に記載されるように、DNA記録タグを介して、DNAカップリングされた脂質を固体支持体に固定する。

【0228】

アプローチ I I . 反応性脂質 (X. Wei & H. Yin (2015)に基づく)脂質過酸化由来の、不飽和アルデヒドによるDNAの共有修飾: Recent progress and challenges, *Free Radical Research*, 49:7, 905-917)。

(a)マロンジアルデヒド(MDA)又は4-ヒドロキシノネナル(HNE)などの反応性脂質基質を得、ヒドラジド末端のDNA記録タグを反応性脂質種にカップリングする。あるいは、アミン末端DNA記録タグを反応性脂質上のアルデヒドにカップリングし、結果として得られるイミンをシアノ水素化ホウ素ナトリウムで還元する。

両方のアプローチの次のステップでは、前述のSpyCatcher結合剤融合を利用することによって、結合剤-DNAコーディングタグ共役体を生成する。コーディングタグは、結合剤に関する識別情報を有するバーコードを含む。脂肪酸結合タンパク質(FABP)、他の脂質結合タンパク質又は脂質結合抗体を、結合剤として使用することができる。最後に、本出願の実施例2に記載されるように、結合剤関連コーディングタグから記録タグにバーコード情報を転送し、したがって、脂質が結合剤に結合する成分を含有するかどうかを分析する。

【0229】

第四に、McElhiney J, et al., Rapid isolation of a single-chain antibody against the cyanobacterial toxin microcystin-LR by phage display and its use in the immunoaffinity concentration of microcystins from water. *Appl Environ Microbiol.* 2002 Nov; 68(11): 5288-95に基づき、当業者に既知の方法を利用して、以下のように例示的マクロサイクル(マイクロシスチン)検出エンコーディングアッセイを実行することができる。

10

20

30

40

50

( a ) マイクロシスチンのデヒドロアラニン を 2 - メルカプトエチルアミン と反応させて一次アミンを生成し、続いて、アミン反応性 DNA 記録タグ ( 例えば、NHS - DNA 誘導体 ) を使用して、DNA 記録タグを一次アミンにカップリングすることによって、DNA 記録タグ結合マイクロシスチンを生成する。

( b ) マイクロシスチンを認識する一本鎖抗体 - SpyCatcher 結合剤を生成する。一本鎖抗体産生は、McElhiney J, et al. 2002 に記載されている。SpyTag ( コーディングタグは、一本鎖抗体に関する識別情報を有するバーコードを含有する ) へのカップル DNA コーディングタグ、続いて、SpyCatcher と反応して、前述のように、結合剤コーディングタグ共役体を生成する。

( c ) バーコード情報を、実施例 2 に記載されるように、一本鎖抗体関連コーディングタグから記録タグに転送し、したがって、高分子がマイクロシスチンを含有するかどうかを分析する。

10

## 【表 2】

配列表

配列番号	配列(5'-3')	説明
1	AATGATACGGCGACCACCGA	P5プライマー
2	CAAGCAGAAGACGGCATACGAGAT	P7プライマー
3	FAGVAMPGAEDDVVGSISK	試験ペプチド
4	AFAGVAMPGAEDDVVGSISK	試験ペプチド
5	AAGVAMPGAEDDVVGSISK	試験ペプチド
6	FSGVARGDVRGGK(アジド)	Fペプチド
7	LAESAFSGVARGDVRGGK(アジド)	Lペプチド
8	/5Phos/TGT AGG GAA AGA GTG TTT /iAmMC6T/T/iSpC3/A CAC TCT TTC CCT ACA CGA CGC TCTTCC GAT CT	捕捉DNA

20

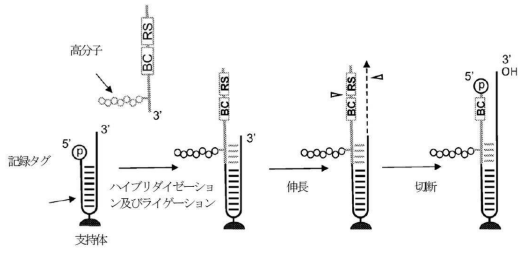
30

40

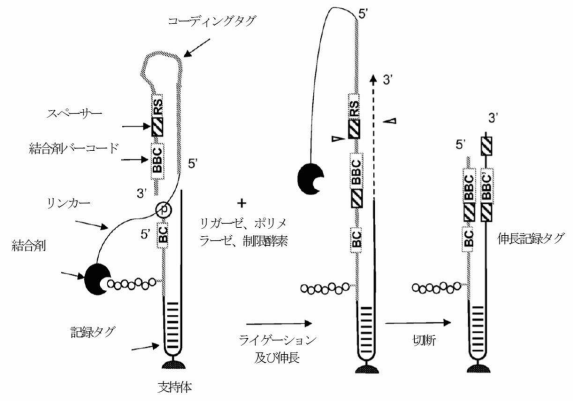
50

【図面】

【図 1 A】

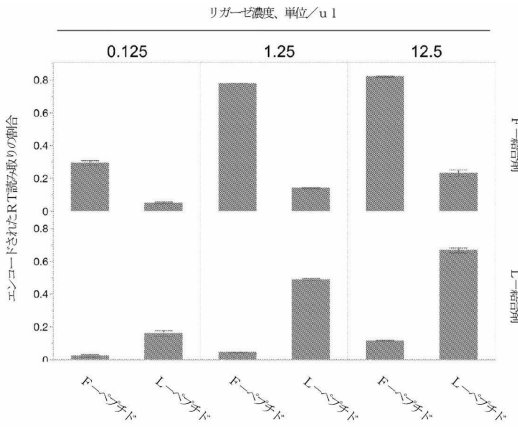


【図 1 B】



10

【図 2】



20

【配列表】

202353807400001.app

30

40

50

【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US2021/046164

**A. CLASSIFICATION OF SUBJECT MATTER**  
 IPC(B) - C12N 15/10; C40B 20/04; C40B 40/04; C40B 70/00; G06K 19/06 (2021.01)  
 CPC - C12N 15/10; C40B 20/04; C40B 40/04; C40B 70/00; G01N 33/68; G01N 33/6845; G01N 2570/00; G06K 19/06 (2021.08)

According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)  
 see Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched  
 see Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
 see Search History document

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2019/0145982 A1 (ENCODIA INC et al) 16 May 2019 (16.05.2019) entire document	1, 2, 4-12, 14, 16 --- 13, 15
X --- Y	WO 2019/089835 A1 (ENCODIA INC) 09 May 2019 (09.05.2019) entire document	17-19, 21 --- 13, 15
A	US 2018/0127744 A1 (CELLULAR RESEARCH INC) 10 May 2018 (10.05.2018) entire document	1-21
A	US 2018/0267036 A1 (CELLULAR RESEARCH INC) 20 September 2018 (20.09.2018) entire document	1-21
A	WO 2020/014586 A1 (BOARD OF REGENTS THE UNIVERSITY OF TEXAS SYSTEM) 16 January 2020 (16.01.2020) entire document	1-21

Further documents are listed in the continuation of Box C.       See patent family annex.

\* Special categories of cited documents:  
 "A" document defining the general state of the art which is not considered to be of particular relevance  
 "D" document cited by the applicant in the international application  
 "E" earlier application or patent but published on or after the international filing date  
 "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)  
 "O" document referring to an oral disclosure, use, exhibition or other means  
 "P" document published prior to the international filing date but later than the priority date claimed  
 "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention  
 "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone  
 "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art  
 "&" document member of the same patent family

Date of the actual completion of the international search  
29 October 2021

Date of mailing of the international search report  
**DEC 06 2021**

Name and mailing address of the ISA/US  
Mail Stop PCT, Attn: ISA/US, Commissioner for Patents  
P.O. Box 1450, Alexandria, VA 22313-1450  
Facsimile No. 571-273-8300

Authorized officer  
Harry Kim  
Telephone No. PCT Helpdesk: 571-272-4300

10

20

30

40

50

**INTERNATIONAL SEARCH REPORT**

International application No.

PCT/US2021/046164

**Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)**

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

a.  forming part of the international application as filed:

in the form of an Annex C/ST.25 text file.

on paper or in the form of an image file.

b.  furnished together with the international application under PCT Rule 13ter.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c.  furnished subsequent to the international filing date for the purposes of international search only:

in the form of an Annex C/ST.25 text file (Rule 13ter.1(a)).

on paper or in the form of an image file (Rule 13ter.1(b) and Administrative Instructions, Section 713).

2.  In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

10

20

30

40

50

フロントページの続き

MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,SM,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,KM,ML,MR,N  
E,SN,TD,TG),AE,AG,AL,AM,AO,AT,AU,AZ,BA,BB,BG,BH,BN,BR,BW,BY,BZ,CA,CH,CL,CN,CO,CR,CU,  
CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IR,IS,IT,JO,JP,K  
E,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,ME,MG,MK,MN,MW,MX,MY,MZ,NA,N  
G,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,  
TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

(特許庁注：以下のものは登録商標)

1 . P L U R O N I C

(72)発明者 チー , マーク エス .

アメリカ合衆国 カリフォルニア 9 2 1 2 1 , サンディエゴ , オーバーリン ドライブ 5 7 8 5

Fターム(参考) 4B063 QA01 QA13 QA18 QQ42 QQ52 QQ79 QR08 QR32 QR35 QS32