

(19)日本国特許庁(JP)

## (12)特許公報(B2)

(11)特許番号

特許第7120630号  
(P7120630)

(45)発行日 令和4年8月17日(2022.8.17)

(24)登録日 令和4年8月8日(2022.8.8)

(51)国際特許分類

F I

C 1 2 Q 1/6869(2018.01)

C 1 2 Q 1/6869

Z Z N A

G 0 1 N 33/53 (2006.01)

G 0 1 N 33/53

D

G 0 1 N 33/543(2006.01)

G 0 1 N 33/543

5 0 1 D

G 0 1 N 33/566(2006.01)

G 0 1 N 33/566

C 1 2 N 15/10 (2006.01)

C 1 2 N 15/10

1 1 4 Z

請求項の数 21 (全180頁) 最終頁に続く

(21)出願番号 特願2018-558133(P2018-558133)

(86)(22)出願日 平成29年5月2日(2017.5.2)

(65)公表番号 特表2019-523635(P2019-523635  
A)

(43)公表日 令和1年8月29日(2019.8.29)

(86)国際出願番号 PCT/US2017/030702

(87)国際公開番号 WO2017/192633

(87)国際公開日 平成29年11月9日(2017.11.9)

審査請求日 令和2年5月1日(2020.5.1)

(31)優先権主張番号 62/330,841

(32)優先日 平成28年5月2日(2016.5.2)

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/339,071

(32)優先日 平成28年5月19日(2016.5.19)

最終頁に続く

(73)特許権者 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, ローゼル スト  
リート 1 1 0 4 5, スイート 1 2 0

(74)代理人 100078282

弁理士 山本 秀策

(74)代理人 100113413

弁理士 森下 夏樹

(74)代理人 100181674

弁理士 飯田 貴敏

(74)代理人 100181641

弁理士 石川 大輔

(74)代理人 230113332

弁護士 山本 健策

最終頁に続く

(54)【発明の名称】 核酸エンコーディングを使用した巨大分子解析

## (57)【特許請求の範囲】

## 【請求項1】

分子を解析するための方法であって、前記分子は、ペプチド、ポリペプチド、タンパク質、核酸、炭水化物、脂質、または大環状分子を含み、前記方法は、

(a) 固体支持体に接合した分子および付随する核酸記録タグを用意するステップと；

(b) 前記分子を、前記分子に結合することが可能な第1の結合性物質であって、前記第1の結合性物質に関する識別情報を有する第1の核酸コーディングタグを含む第1の結合性物質と接触させるステップと；

(c) 前記第1の核酸コーディングタグの情報を前記核酸記録タグに移行させて、一次伸長核酸記録タグを生成するステップであって、前記移行は、プライマー伸長またはライゲーションを含む、ステップと；

(d) 前記分子を、前記分子に結合することが可能な第2の結合性物質であって、前記第2の結合性物質に関する識別情報を有する第2の核酸コーディングタグを含む第2の結合性物質と接触させるステップと；

(e) 前記第2の核酸コーディングタグの情報を前記一次伸長核酸記録タグに移行させて、二次伸長核酸記録タグを生成するステップであって、前記移行は、プライマー伸長またはライゲーションを含む、ステップと；

(f) 前記二次伸長核酸記録タグを解析するステップと  
を含み、前記分子の解析は、前記分子の全てまたは一部の特徴付け、識別または定量を含む、方法。

10

20

**【請求項 2】**

ステップ ( e ) と ( f ) の間に、

前記第 2 の結合性物質を、前記分子に結合することが可能な第 3 の結合性物質であって、前記第 3 の結合性物質に関する識別情報を有する第 3 の核酸コーディングタグを含む第 3 の結合性物質に置き換えることにより、ステップ ( d ) および ( e ) を 1 回または複数回繰り返すステップと；

前記第 3 の核酸コーディングタグの情報を前記二次伸長核酸記録タグに移行させて、三次伸長核酸記録タグを生成するステップであって、前記移行は、プライマー伸長またはライゲーションを含む、ステップと

をさらに含み、

ステップ ( f ) において前記三次伸長核酸記録タグを解析する、請求項 1 に記載の方法。

**【請求項 3】**

ステップ ( a ) が前記固体支持体に接合した複数の分子および付随する核酸記録タグを提供することを含み、

ステップ ( b ) および ( d ) のそれぞれが、前記複数の分子を、前記分子に結合することが可能な第 1 または第 2 の複数の結合性物質と接触させることを含み、前記第 1 または第 2 の複数の結合性物質が、前記第 1 または第 2 の結合性物質に関する識別情報を有する第 1 または第 2 の核酸コーディングタグを含む、

請求項 1 または請求項 2 に記載の方法。

**【請求項 4】**

前記分子が、タンパク質、ポリペプチドまたはペプチドである、請求項 1 から 3 までのいずれか一項に記載の方法。

**【請求項 5】**

前記ペプチドが、生体試料由来のタンパク質を断片化することによって得られる、請求項 4 に記載の方法。

**【請求項 6】**

前記タンパク質、ポリペプチドまたはペプチドの N 末端アミノ酸 ( N T A A ) を化学部分で修飾し、修飾された N T A A を産生することをさらに含む、請求項 4 または請求項 5 に記載の方法。

**【請求項 7】**

前記第 1 の結合性物質および / または前記第 2 の結合性物質が前記修飾された N T A A に結合することが可能である、請求項 6 に記載の方法。

**【請求項 8】**

前記修飾された N T A A を除去して、前記タンパク質、ポリペプチドまたはペプチドの新しい N T A A を露出させることをさらに含む、請求項 6 または 7 に記載の方法。

**【請求項 9】**

前記核酸記録タグが、DNA 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、またはこれらの組合せであり、および / または

前記核酸コーディングタグが、DNA 分子、RNA 分子、BNA 分子、XNA 分子、LNA 分子、PNA 分子、PNA 分子、またはこれらの組合せである、  
請求項 1 から 8 までのいずれか一項に記載の方法。

**【請求項 10】**

前記核酸記録タグが、  
ユニバーサルプライミング部位、  
一意の分子識別子 ( U M I ) 、

バーコード、  
その 3 ' 末端におけるスペーサー、および / または

3 ' ブロック基  
を含む、請求項 1 から 9 までのいずれか一項に記載の方法。

10

20

30

40

50

**【請求項 1 1】**

前記複数の分子の間に前記固体支持体上で平均距離  $> 50 \text{ nm}$  の間隔をあける、請求項 3 から 10 までのいずれか一項に記載の方法。

**【請求項 1 2】**

前記核酸コーディングタグの情報の前記核酸記録タグへの移行が、DNA リガーゼ、DNA ポリメラーゼまたは化学的ライゲーションによって媒介される、請求項 1 から 11 までのいずれか一項に記載の方法。

**【請求項 1 3】**

前記第 1 の結合性物質および / または前記第 2 の結合性物質が、前記分子に選択的に結合可能である、請求項 1 から 12 までのいずれか一項に記載の方法。

10

**【請求項 1 4】**

前記第 1 の結合性物質および / または前記第 2 の結合性物質が、単一のアミノ酸残基、ジペプチド、トリペプチドまたは前記ペプチドの翻訳後修飾に結合する、請求項 1 から 13 までのいずれか一項に記載の方法。

**【請求項 1 5】**

前記翻訳後修飾がアシル化、アセチル化、アルキル化、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、ジフタミド形成、ジスルフィド架橋形成、エリミニル化、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、グリコシルホスファチジルイノシトール付加、ヘム C 付着、ヒドロキシル化、ハイブシン形成、ヨウ素化、イソプレニル化、脂質付加、リポイル化、マロニル化、メチル化、ミリストイル化、酸化、パルミトイル化、ペグ化、ホスホパンテテイル化、リン酸化、プレニル化、プロピオニル化、レチニリデンシッフ塩基形成、S - グルタチオン化、S - ニトロシル化、S - スルフェニル化、セレン化、サクシニル化、スルフィン化、ユビキチン化、および C 末端アミド化からなる群から選択される、請求項 1 4 に記載の方法。

20

**【請求項 1 6】**

前記方法が前記翻訳後修飾を除去することをさらに含む、請求項 1 4 または請求項 1 5 に記載の方法。

**【請求項 1 7】**

前記核酸コーディングタグが、スペーサー、結合サイクル特異的配列、一意の分子識別子、ユニバーサルプライミング部位、ターミネーターヌクレオチドまたはこれらに任意の組み合わせをさらに含む、請求項 1 から 16 までのいずれか一項に記載の方法。

30

**【請求項 1 8】**

前記伸長核酸記録タグの解析が、核酸配列決定法を含む、請求項 1 から 17 までのいずれか一項に記載の方法。

**【請求項 1 9】**

前記伸長核酸記録タグが解析前に増幅される、請求項 1 から 18 までのいずれか一項に記載の方法。

**【請求項 2 0】**

前記固体支持体が、ビーズ、多孔質ビーズ、多孔質マトリックス、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、ナイロン、シリコンウェーハチップ、フロースルーチップ、信号変換電子機器を含むバイオチップ、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、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアである、請求項 1 から 19 までのいずれか一項に記載の方法。

40

**【請求項 2 1】**

前記固体支持体が、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、制御ポアビーズまたはそれらの組合せである、請求項 20 に記載の方法。

**【発明の詳細な説明】**

50

## 【技術分野】

## 【0001】

## 配列表に関する記述

本出願に付随する配列表は、紙コピーの代わりにテキスト形式で提供され、また、これによって参照により本明細書に組み込まれる。配列表を含むテキストファイルの名称は、760229\_\_401WO\_\_SEQUENCE\_\_LISTING.txtである。テキストファイルは、38.7KBであり、2017年5月2日に作成され、EFS-Webを介して電子的に提出される。

## 【0002】

本開示は、一般に、分子認識事象のバーコーディングおよび核酸エンコーディングを使用した、ペプチド、ポリペプチド、およびタンパク質を含めた巨大分子の解析に関する。

10

## 【背景技術】

## 【0003】

タンパク質は、多くの異なる生物学的機能を実行および促進し、細胞生物学および生理学において不可欠な役割を果たす。異なるタンパク質分子のレパートリーは広範囲にわたり、翻訳後修飾（PTM）によって導入される追加的な多様性に起因して、トランスクリプトームよりもはるかに複雑である。さらに、細胞内のタンパク質は、環境、生理的状況、および病態にตอบสนองして動的に変化する（発現レベルおよび修飾の状態）。したがって、タンパク質は、特にゲノム情報と比べて、莫大な量のほとんど明らかになっていない関連情報を含有する。一般に、プロテオミクス解析ではゲノミクス解析と比べて革新が遅れている。ゲノミクスの分野では、次世代シーケンシング（NGS）により、数十億のDNA配列を単一の計器における実行で解析することが可能になることによって当該分野が変容しているが、一方で、タンパク質解析およびペプチド配列決定では、スループットが今でも限られている。

20

## 【0004】

それにもかかわらず、このタンパク質情報は健康および疾患におけるプロテオームダイナミクスをよりよく理解するため、ならびに高精度の医療を可能にするのを補助するために直接必要である。そのように、このプロテオミクス情報の収集を小型化および高度に並行化するための「次世代」ツールの開発に大きな関心が寄せられている。

## 【0005】

30

タンパク質の高度に並行な高分子特徴付けおよび認識は、いくつかの理由で困難である。親和性に基づくアッセイの使用は、多くの場合、いくつかの重要な難題に起因して難しい。1つの重要な難題は、親和性作用物質の集合の読み取りを同類の巨大分子の集合に多重化することである；別の難題は、親和性作用物質とオフターゲットの巨大分子との交差反応性を最小化することである；第3の難題は、効率的なハイスループットの読み取りプラットフォームを開発することである。この問題の例は、試料中のタンパク質の大多数または全てを識別および定量化することが1つの目標であるプロテオミクスにおいて生じる。さらに、タンパク質の種々の翻訳後修飾（PTM）を単一分子レベルで特徴付けることが望ましい。現在、これは、ハイスループットなやり方で実現するのが大変な課題である。

## 【0006】

40

タンパク質またはペプチド巨大分子の分子認識および特徴付けは、一般には、イムノアッセイを使用して実施される。ELISA、マルチプレックスELISA（例えば、スポット抗体アレイ（spotted antibody array）、液体粒子ELISAアレイ）、デジタルELISA（例えば、Quanterix、Singulex）、逆相タンパク質アレイ（RPPA）、および多くの他のものを含めた、多くの異なるイムノアッセイ形式が存在する。これらの異なるイムノアッセイプラットフォームは全て、高親和性かつ高度に特異的な（または選択的な）抗体（結合性物質）の開発、試料レベルおよび分析物レベルのどちらにおいても多重化能力が限られていること、感度およびダイナミックレンジが限られていること、ならびに交差反応性およびバックグラウンドシグナルを含めた、同様の難題に直面する。ペプチド配列決定（エドマン分解または質量分析）

50



による直接タンパク質特徴付けなどの、結合性物質にとらわれない手法により有用な代替的手法がもたらされる。しかし、これらの手法はいずれも、極めて並行またはハイスループットなものではない。

【0007】

エドマン分解に基づくペプチド配列決定は、1950年にPehr Edmanによって最初に提唱されたものであり、言い換えると、ペプチドのN末端アミノ酸の、一連の化学修飾による段階的分解および下流のHPLC分析（後に質量分析による解析に置き換えられた）である。第1のステップにおいて、N末端アミノ酸を穏やかな塩基性条件下（NMP/メタノール/H<sub>2</sub>O）でフェニルイソチオシアネート（PITC）を用いて修飾してフェニルチオカルバモイル（PTC）誘導体を形成させる。第2のステップでは、PTCで修飾されたアミノ基を酸（無水TFA）で処理して、切断された環状ATZ（2-アニリノ-5（4）-チアゾリノン（thiozolinone）修飾アミノ酸を創出し、新しいN末端をペプチド上に残す。切断された環状ATZアミノ酸をPTHアミノ酸誘導体に変換し、逆相HPLCによって分析する。このプロセスを、ペプチド配列を構成するアミノ酸の全てまたは部分的な数がN末端から除去され識別されるまで反復的に継続する。一般に、エドマン分解ペプチド配列決定は、時間がかかり、1日当たりほんの数ペプチドとスループットが限られている。

10

【0008】

ここ10～15年で、MALDI、エレクトロスプレー質量分析（MS）、およびLC-MS/MSを使用したペプチド解析が大きくエドマン分解に取って代わっている。MS器械使用（Rileyら、2016年、Cell Syst、2巻：142～143頁）における最近の進歩にもかかわらず、MSにはなお、計器費用が高いこと、洗練された使用者が求められること、数量化能力が不十分であること、およびプロテオームのダイナミックレンジにわたって測定を行う能力が限られていることを含めたいくつかの欠点がある。例えば、タンパク質は異なる効率レベルでイオン化するので、試料間の絶対的な定量化およびさらには相対的な定量化も困難である。質量タグの実装が相対的な定量化の改善に役立っているが、プロテオームの標識が必要になる。試料中のタンパク質の濃度が非常に大きな範囲にわたって（血漿に関しては10桁にわたって）変動し得るダイナミックレンジがさらなる複雑化の要因である。一般には、MSではより豊富な種のみが分析され、豊富さが低いタンパク質の特徴付けは困難になる。最後に、試料スループットは、一般には、実行当たり数千ペプチドに限られ、データ非依存性解析（DIA）に関して、このスループットは真のボトムアップ式のハイスループットなプロテオーム解析には不十分である。さらに、各試料について記録された何千もの複雑なMSスペクトルをデコンボリューションするために著しいコンピュータ処理の必要性がある。

20

30

【先行技術文献】

【非特許文献】

【0009】

【文献】Rileyら、2016年、Cell Syst、2巻：142～143頁

【発明の概要】

【課題を解決するための手段】

40

【0010】

したがって、タンパク質配列決定および/または解析に適用される巨大分子配列決定および/または解析に関する改善された技法、ならびに、それを実現するための製品、方法およびキットが当技術分野において依然として必要とされている。高度に並行化された、正確な、感度の高い、かつハイスループットなプロテオミクス技術が必要とされている。本開示は、これらおよび他の必要性を満たすものである。

【0011】

本発明のこれらおよび他の態様は、以下の詳細な説明を参照すれば明らかになる。この目的のために、ある特定のバックグラウンド情報、手順、化合物および/または組成物がより詳細に記載されている種々の参考文献が本明細書に記載され、これらはそれぞれが

50

、これによって全体が参照により組み込まれる。

【 0 0 1 2 】

本開示の複数の実施形態は、概して、高度に並行な、ハイスループットなデジタル巨大分子解析、特にペプチド解析の方法に関する。

【 0 0 1 3 】

第 1 の実施形態は、巨大分子を解析するための方法であって、

- ( a ) 固体支持体に接合した巨大分子および付随する記録タグを用意するステップと；
  - ( b ) 前記巨大分子を、前記巨大分子に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；
  - ( c ) 前記第 1 のコーディングタグの情報を前記記録タグに移行させて、一次伸長記録タグを生成するステップと；
  - ( d ) 前記巨大分子を、前記巨大分子に結合することが可能な第 2 の結合性物質であって、前記第 2 の結合性物質に関する識別情報を有する第 2 のコーディングタグを含む第 2 の結合性物質と接触させるステップと；
  - ( e ) 前記第 2 のコーディングタグの情報を前記一次伸長記録タグに移行させて、二次伸長記録タグを生成するステップと；
  - ( f ) 前記二次伸長記録タグを解析するステップと
- を含む方法である。

10

【 0 0 1 4 】

第 2 の実施形態は、接触させるステップ ( b ) および ( d ) を逐次的に実施する、実施形態 1 に記載の方法である。

20

【 0 0 1 5 】

第 3 の実施形態は、接触させるステップ ( b ) および ( d ) を同時に実施する、実施形態 1 に記載の方法である。

【 0 0 1 6 】

第 4 の実施形態は、ステップ ( e ) と ( f ) の間に、

- ( x ) 前記第 2 の結合性物質を、前記巨大分子に結合することが可能な第 3 の ( またはより高次の ) 結合性物質であって、前記第 3 の ( またはより高次の ) 結合性物質に関する識別情報を有する第 3 の ( またはより高次の ) コーディングタグを含む第 3 の ( またはより高次の ) 結合性物質に置き換えることにより、ステップ ( d ) および ( e ) を 1 回または複数回繰り返すステップと；
  - ( y ) 前記第 3 の ( またはより高次の ) コーディングタグの情報を前記第 2 の ( またはより高次の ) 伸長記録タグに移行させて、第 3 の ( またはより高次の ) 伸長記録タグを生成するステップと
- をさらに含み、
- ステップ ( f ) において前記第 3 の ( またはより高次の ) 伸長記録タグを解析する、実施形態 1 に記載の方法である。

30

【 0 0 1 7 】

第 5 の実施形態は、巨大分子を解析するための方法であって、

- ( a ) 固体支持体に接合した巨大分子、付随する第 1 の記録タグおよび付随する第 2 の記録タグを用意するステップと；
- ( b ) 前記巨大分子を、前記巨大分子に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；
- ( c ) 前記第 1 のコーディングタグの情報を前記第 1 の記録タグに移行させて、第 1 の伸長記録タグを生成するステップと；
- ( d ) 前記巨大分子を、前記巨大分子に結合することが可能な第 2 の結合性物質であって、前記第 2 の結合性物質に関する識別情報を有する第 2 のコーディングタグを含む第 2 の結合性物質と接触させるステップと；

40

50

( e ) 前記第 2 のコーディングタグの情報を前記第 2 の記録タグに移行させて、第 2 の伸長記録タグを生成するステップと；

( f ) 前記第 1 の伸長記録タグおよび第 2 の伸長記録タグを解析するステップとを含む方法である。

【 0 0 1 8 】

第 6 の実施形態は、接触させるステップ ( b ) および ( d ) を逐次的に実施する、実施形態 5 に記載の方法である。

【 0 0 1 9 】

第 7 の実施形態は、接触させるステップ ( b ) および ( d ) を同時に実施する、実施形態 5 に記載の方法である。

【 0 0 2 0 】

第 8 の実施形態は、ステップ ( a ) が、前記固体支持体に接合した付随する第 3 の ( またはより高次の ) 記録タグを用意するステップをさらに含む、実施形態 5 に記載の方法である。

【 0 0 2 1 】

第 9 の実施形態は、ステップ ( e ) と ( f ) の間に、

( x ) 前記第 2 の結合性物質を、前記巨大分子に結合することが可能な第 3 の ( またはより高次の ) 結合性物質であって、前記第 3 の ( またはより高次の ) 結合性物質に関する識別情報を有する第 3 の ( またはより高次の ) コーディングタグを含む第 3 の ( またはより高次の ) 結合性物質に置き換えることにより、ステップ ( d ) および ( e ) を 1 回または複数回繰り返すステップと；

( y ) 前記第 3 の ( またはより高次の ) コーディングタグの情報を前記第 3 の ( またはより高次の ) 記録タグに移行させて、第 3 の ( またはより高次の ) 伸長記録タグを生成するステップと

をさらに含み、

ステップ ( f ) において前記第 1 の伸長記録タグ、前記第 2 の伸長記録タグおよび前記第 3 の ( またはより高次の ) 伸長記録タグを解析する、実施形態 8 に記載の方法である。

【 0 0 2 2 】

第 1 0 の実施形態は、前記第 1 のコーディングタグ、前記第 2 のコーディングタグ、および任意のより高次のコーディングタグが、結合サイクル特異的スペーサー配列を含む、実施形態 5 から 9 までのいずれか 1 つに記載の方法である。

【 0 0 2 3 】

第 1 1 の実施形態は、ペプチドを解析するための方法であって、

( a ) 固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；

( b ) 前記ペプチドの N 末端アミノ酸 ( N T A A ) を化学薬剤で修飾するステップと；

( c ) 前記ペプチドを、修飾された前記 N T A A に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；

( d ) 前記第 1 のコーディングタグの情報を前記記録タグに移行させて、伸長記録タグを生成するステップと；

( e ) 前記伸長記録タグを解析するステップとを含む方法である。

【 0 0 2 4 】

第 1 2 の実施形態は、ステップ ( c ) が、前記ペプチドを、前記第 2 の ( またはより高次の ) 結合性物質に関する識別情報を有する第 2 の ( またはより高次の ) コーディングタグを含む第 2 の ( またはより高次の ) 結合性物質であって、ステップ ( b ) の前記修飾された N T A A 以外の修飾された N T A A に結合することが可能である第 2 の ( またはより高次の ) 結合性物質と接触させることをさらに含む、実施形態 1 1 に記載の方法である。

【 0 0 2 5 】

第 1 3 の実施形態は、前記ペプチドの前記第 2 の ( またはより高次の ) 結合性物質との

10

20

30

40

50

接触を、前記ペプチドの前記第 1 の結合性物質との接触後に逐次的に行う、第 1 2 の実施形態に記載の方法である。

【 0 0 2 6 】

第 1 4 の実施形態は、前記ペプチドの前記第 2 の（またはより高次の）結合性物質との接触を、前記ペプチドの前記第 1 の結合性物質との接触と同時に行う、第 1 2 の実施形態に記載の方法である。

【 0 0 2 7 】

第 1 5 の実施形態は、前記化学薬剤が、イソチオシアネート誘導体、2, 4 - ジニトロベンゼンスルホン酸 ( d i n i t r o b e n z e n e s u l f o n i c ) ( D N B S )、4 - スルホニル - 2 - ニトロフルオロベンゼン ( S N F B ) 1 - フルオロ - 2, 4 - ジニトロベンゼン、ダンシルクロリド、7 - メトキシマリリン酢酸、チオアシル化試薬、チオアセチル化試薬、またはチオベンジル化試薬である、実施形態 1 1 から 1 4 までのいずれか 1 つに記載の方法である。

10

【 0 0 2 8 】

第 1 6 の実施形態は、ペプチドを解析するための方法であって、

- ( a ) 固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；
  - ( b ) 前記ペプチドの N 末端アミノ酸 ( N T A A ) を化学薬剤で修飾して、修飾された N T A A を得るステップと；
  - ( c ) 前記ペプチドを、前記修飾された N T A A に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；
  - ( d ) 前記第 1 のコーディングタグの情報を前記記録タグに移行させて、第 1 の伸長記録タグを生成するステップと；
  - ( e ) 前記修飾された N T A A を除去して、新しい N T A A を露出させるステップと；
  - ( f ) 前記ペプチドの前記新しい N T A A を化学薬剤で修飾して、新しく修飾された N T A A を得るステップと；
  - ( g ) 前記ペプチドを、前記新しく修飾された N T A A に結合することが可能な第 2 の結合性物質であって、前記第 2 の結合性物質に関する識別情報を有する第 2 のコーディングタグを含む第 2 の結合性物質と接触させるステップと；
  - ( h ) 前記第 2 のコーディングタグの情報を前記第 1 の伸長記録タグに移行させて、第 2 の伸長記録タグを生成するステップと；
  - ( i ) 前記第 2 の伸長記録タグを解析するステップと
- を含む方法である。

20

30

【 0 0 2 9 】

第 1 7 の実施形態は、ペプチドを解析するための方法であって、

- ( a ) 固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；
  - ( b ) 前記ペプチドを、前記ペプチドの N 末端アミノ酸 ( N T A A ) に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；
  - ( c ) 前記第 1 のコーディングタグの情報を前記記録タグに移行させて、伸長記録タグを生成するステップと；
  - ( d ) 前記伸長記録タグを解析するステップと
- を含む方法である。

40

【 0 0 3 0 】

第 1 8 の実施形態では、ステップ ( b ) が、前記ペプチドを、前記第 2 の（またはより高次の）結合性物質に関する識別情報を有する第 2 の（またはより高次の）コーディングタグを含む第 2 の（またはより高次の）結合性物質であって、前記ペプチドの前記 N T A A 以外の N T A A に結合することが可能な第 2 の（またはより高次の）結合性物質と接触させることをさらに含む、第 1 7 の実施形態に記載の方法である。

【 0 0 3 1 】

50

第 19 の実施形態は、前記ペプチドの前記第 2 の（またはより高次の）結合性物質との接触を、前記ペプチドの前記第 1 の結合性物質との接触後に逐次的に行う、第 18 の実施形態に記載の方法である。

【0032】

第 20 の実施形態は、前記ペプチドの前記第 2 の（またはより高次の）結合性物質との接触を、前記ペプチドの前記第 1 の結合性物質との接触と同時に行う、第 18 の実施形態に記載の方法である。

【0033】

第 21 の実施形態は、ペプチドを解析するための方法であって、

（a）固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；  
（b）前記ペプチドを、前記ペプチドの N 末端アミノ酸（NTAA）に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；

（c）前記第 1 のコーディングタグの情報を前記記録タグに移行させて、第 1 の伸長記録タグを生成するステップと；

（d）前記 NTAA を除去して、前記ペプチドの新しい NTAA を露出させるステップと；

（e）前記ペプチドを、前記新しい NTAA に結合することが可能な第 2 の結合性物質であって、前記第 2 の結合性物質に関する識別情報を有する第 2 のコーディングタグを含む第 2 の結合性物質と接触させるステップと；

（h）前記第 2 のコーディングタグの情報を前記第 1 の伸長記録タグに移行させて、第 2 の伸長記録タグを生成するステップと；

（i）前記第 2 の伸長記録タグを解析するステップとを含む方法である。

【0034】

第 22 の実施形態は、前記巨大分子が、タンパク質、ポリペプチドまたはペプチドである、実施形態 1 から 10 までのいずれか 1 つに記載の方法である。

【0035】

第 23 の実施形態は、前記巨大分子が、ペプチドである、実施形態 1 から 10 までのいずれか 1 つに記載の方法である。

【0036】

第 24 の実施形態は、前記ペプチドが、生体試料由来の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 11 から 23 までのいずれか 1 つに記載の方法である。

【0037】

第 25 の実施形態は、前記巨大分子が、脂質、炭水化物、または大環状分子である、実施形態 1 から 10 までのいずれか 1 つに記載の方法である。

【0038】

第 26 の実施形態は、前記記録タグが、DNA 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、またはこれらの組合せである、実施形態 1 から 25 までのいずれか 1 つに記載の方法である。

【0039】

第 27 の実施形態は、前記記録タグが、ユニバーサルプライミング部位を含む、実施形態 1 から 26 までのいずれか 1 つに記載の方法である。

【0040】

第 28 の実施形態は、前記ユニバーサルプライミング部位が、増幅、配列決定、またはその両方のためのプライミング部位を含む、実施形態 27 に記載の方法である。

【0041】

第 29 の実施形態は、前記記録タグが、一意の分子識別子（UMI）を含む、実施形態 1 から 28 までに記載の方法である。

【0042】

10

20

30

40

50

第 3 0 の実施形態は、記録タグがバーコードを含む、第 1 ~ 第 2 9 の実施形態のいずれか 1 つに記載の方法である。

【 0 0 4 3 】

第 3 1 の実施形態は、前記記録タグが、その 3 ' 末端にスペーサーを含む、実施形態 1 から 3 0 までのいずれか 1 つに記載の方法である。

【 0 0 4 4 】

第 3 2 の実施形態は、前記巨大分子および前記付随する記録タグを、前記固体支持体に共有結合により接合させる、実施形態 1 から 3 1 までのいずれか 1 つに記載の方法である。

【 0 0 4 5 】

第 3 3 の実施形態は、前記固体支持体が、ビーズ、多孔質ビーズ、多孔質マトリックス、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、ナイロン、シリコンウェーハチップ、フロースルーチップ、信号変換電子機器を含むバイオチップ、マイクロタイターウェル、E L I S A プレート、スピン干渉ディスク、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアである、実施形態 1 から 3 2 までのいずれか 1 つに記載の方法である。

【 0 0 4 6 】

第 3 4 の実施形態は、前記固体支持体が、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズである、実施形態 3 3 に記載の方法である。

【 0 0 4 7 】

第 3 5 の実施形態は、複数の巨大分子および付随する記録タグを固体支持体に接合する、実施形態 1 から 3 4 までのいずれか 1 つに記載の方法である。

【 0 0 4 8 】

第 3 6 の実施形態は、前記複数の巨大分子の間に前記固体支持体上で平均距離 > 5 0 n m の間隔をあける、実施形態 3 5 に記載の方法である。

【 0 0 4 9 】

第 3 7 の実施形態は、前記結合性物質が、ポリペプチドまたはタンパク質である、実施形態 1 から 3 6 までのいずれか 1 つに記載の方法である。

【 0 0 5 0 】

第 3 8 の実施形態は、前記結合性物質が、改変アミノペプチダーゼ、改変アミノアシル t R N A 合成酵素、改変アンチカリン、または改変 C l p S である、実施形態 3 7 に記載の方法である。

【 0 0 5 1 】

第 3 9 の実施形態は、前記結合性物質が、巨大分子に選択的に結合することが可能である、実施形態 1 から 3 8 までのいずれか 1 つに記載の方法である。

【 0 0 5 2 】

第 4 0 の実施形態は、前記コーディングタグが、D N A 分子、R N A 分子、B N A 分子、X N A 分子、L N A 分子、P N A 分子、P N A 分子、またはこれらの組合せである、実施形態 1 から 3 9 までのいずれか 1 つに記載の方法である。

【 0 0 5 3 】

第 4 1 の実施形態は、前記コーディングタグが、エンコーダー配列を含む、実施形態 1 から 4 0 までのいずれか 1 つに記載の方法である。

【 0 0 5 4 】

第 4 2 の実施形態は、前記コーディングタグが、スペーサー、結合サイクル特異的配列、一意の分子識別子、ユニバーサルプライミング部位、またはそれらの任意の組合せをさらに含む、実施形態 1 から 4 1 までのいずれか 1 つに記載の方法である。

【 0 0 5 5 】

第 4 3 の実施形態は、前記結合性物質と前記コーディングタグが、リンカーによって接合されている、実施形態 1 から 4 2 までのいずれか 1 つに記載の方法である。

【 0 0 5 6 】

10

20

30

40

50

第44の実施形態は、前記結合性物質と前記コーディングタグが、SpyTag/SpyCatcherまたはSnoopTag/SnoopCatcherペプチド-タンパク質対によって接合されている、実施形態1から42までに記載の方法である。

【0057】

第45の実施形態は、前記コーディングタグの情報の前記記録タグへの移行が、DNAリガーゼによって媒介される、実施形態1から44までのいずれか1つに記載の方法である。

【0058】

第46の実施形態は、前記コーディングタグの情報の前記記録タグへの移行が、DNAポリメラーゼによって媒介される、実施形態1から44までのいずれか1つに記載の方法である。

10

【0059】

第47の実施形態は、前記コーディングタグの情報の前記記録タグへの移行が、化学的ライゲーションによって媒介される、実施形態1から44までのいずれか1つに記載の方法である。

【0060】

第48の実施形態は、前記伸長記録タグの解析が、核酸配列決定法を含む、実施形態1から47までのいずれか1つに記載の方法である。

【0061】

第49の実施形態は、前記核酸配列決定法が、合成による配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、ポロニーシーケンシング、イオン半導体シーケンシング、またはパイロシーケンシングである、実施形態48に記載の方法である。

20

【0062】

第50の実施形態は、前記核酸配列決定法が、単一分子リアルタイムシーケンシング、ナノポアに基づく配列決定、または先端顕微鏡を使用したDNAのダイレクトイメージングである、実施形態48に記載の方法である。

【0063】

第51の実施形態は、前記伸長記録タグを解析前に増幅する、実施形態1から50までのいずれか1つに記載の方法である。

30

【0064】

第52の実施形態は、前記伸長記録タグに含有されるコーディングタグ情報の順序が、前記結合性物質による前記巨大分子への結合の順序に関する情報を提供する、実施形態1から51までに記載の方法である。

【0065】

第53の実施形態は、前記伸長記録タグに含有される前記コーディングタグ情報の頻度が、前記結合性物質による前記巨大分子への結合の頻度に関する情報を提供する、実施形態1から52までに記載の方法である。

【0066】

第54の実施形態は、複数の巨大分子を表す複数の伸長記録タグを並行して解析する、実施形態1から53までに記載の方法である。

40

【0067】

第55の実施形態は、前記複数の巨大分子を表す複数の伸長記録タグを多重化アッセイで解析する、実施形態54に記載の方法である。

【0068】

第56の実施形態は、前記複数の伸長記録タグが、解析前に標的濃縮アッセイを受ける、実施形態1から55までのいずれか1つに記載の方法である。

【0069】

第57の実施形態は、前記複数の伸長記録タグが、解析前にサブトラクションアッセイを受ける、実施形態1から56までのいずれか1つに記載の方法である。

50

## 【 0 0 7 0 】

第 5 8 の実施形態は、前記複数の伸長記録タグが、極めて豊富な種を減少させるために解析前に正規化アッセイを受ける、実施形態 1 から 5 7 までのいずれか 1 つに記載の方法である。

## 【 0 0 7 1 】

第 5 9 の実施形態は、前記 N T A A を、改変アミノペプチダーゼ、改変アミノ酸 t R N A 合成酵素、穏やかなエドマン分解、エドマナーゼ ( E d m a n a s e ) 酵素、または無水 T F A によって除去する、実施形態 1 から 5 8 までのいずれか 1 つに記載の方法である。

## 【 0 0 7 2 】

第 6 0 の実施形態は、少なくとも 1 つの結合性物質が末端アミノ酸残基に結合する、実施形態 1 から 5 9 までのいずれか 1 つに記載の方法である。

10

## 【 0 0 7 3 】

第 6 1 の実施形態は、少なくとも 1 つの結合性物質が翻訳後修飾されたアミノ酸に結合する、実施形態 1 から 6 0 までのいずれか 1 つに記載の方法である。

## 【 0 0 7 4 】

第 6 2 の実施形態は、複数のタンパク質複合体、タンパク質、またはポリペプチドを含む試料由来の 1 つまたは複数のペプチドを解析するための方法であって、

## 【 0 0 7 5 】

( a ) 前記試料中の前記複数のタンパク質複合体、タンパク質、またはポリペプチドを複数のコンパートメントに分配するステップであって、各コンパートメントが、任意選択で固体支持体と接合した複数のコンパートメントタグを含み、前記複数のコンパートメントタグが、個々のコンパートメント内では同じであり、他のコンパートメントのコンパートメントタグとは異なる、ステップと；

20

( b ) 前記複数のタンパク質複合体、タンパク質、および / またはポリペプチドを複数のペプチドに断片化するステップと；

( c ) 前記複数のペプチドと前記複数のコンパートメントタグを、前記複数のペプチドと前記複数のコンパートメント内の前記複数のコンパートメントタグとのアニーリングまたは接合を可能にするのに十分な条件下で接触させ、それにより、複数のコンパートメントタグ付きペプチドを生成するステップと；

( d ) 前記コンパートメントタグ付きペプチドを前記複数のコンパートメントから収集するステップと；

30

( e ) 1 つまたは複数のコンパートメントタグ付きペプチドを、実施形態 1 から 2 1 までおよび実施形態 2 6 から 6 1 までのいずれか 1 つに記載の方法に従って解析するステップを含む方法である。

## 【 0 0 7 6 】

第 6 3 の実施形態は、前記コンパートメントがマイクロ流液体液滴である、実施形態 6 2 に記載の方法である。

## 【 0 0 7 7 】

第 6 4 の実施形態は、前記コンパートメントがマイクロウェルである、実施形態 6 2 に記載の方法である。

40

## 【 0 0 7 8 】

第 6 5 の実施形態は、前記コンパートメントが、表面上の分離された領域である、実施形態 6 2 に記載の方法である。

## 【 0 0 7 9 】

第 6 6 の実施形態は、各コンパートメントが、平均して単一の細胞を含む、実施形態 6 2 から 6 5 までのいずれか 1 つに記載の方法である。

## 【 0 0 8 0 】

第 6 7 の実施形態は、複数のタンパク質複合体、タンパク質、またはポリペプチドを含む試料由来の 1 つまたは複数のペプチドを解析するための方法であって、

50



( a ) 前記複数のタンパク質複合体、タンパク質、またはポリペプチドを複数のユニバーサル DNA タグで標識するステップと；

( b ) 前記試料中の前記複数の標識されたタンパク質複合体、タンパク質、またはポリペプチドを複数のコンパートメントに分配するステップであって、各コンパートメントが、複数のコンパートメントタグを含み、前記複数のコンパートメントタグが、個々のコンパートメント内では同じであり、他のコンパートメントのコンパートメントタグとは異なる、ステップと；

( c ) 前記複数のタンパク質複合体、タンパク質、またはポリペプチドと前記複数のコンパートメントタグを、前記複数のタンパク質複合体、タンパク質、またはポリペプチドと前記複数のコンパートメント内の前記複数のコンパートメントタグとのアニーリングまたは接合を可能にするのに十分な条件下で接触させ、それにより、複数のコンパートメントタグ付きタンパク質複合体、タンパク質またはポリペプチドを生成するステップと；

( d ) 前記コンパートメントタグ付きタンパク質複合体、タンパク質、またはポリペプチドを前記複数のコンパートメントから収集するステップと；

( e ) 任意選択で前記コンパートメントタグ付きタンパク質複合体、タンパク質、またはポリペプチドをコンパートメントタグ付きペプチドに断片化するステップと；

( f ) 1 つまたは複数のコンパートメントタグ付きペプチドを、実施形態 1 から 2 1 までおよび実施形態 2 6 から 6 1 までのいずれか 1 つに記載の方法に従って解析するステップと

を含む方法である。

#### 【 0 0 8 1 】

第 6 8 の実施形態は、コンパートメントタグ情報を、ペプチドに付随する記録タグにプライマー伸長またはライゲーションによって移行させる、実施形態 6 2 から 6 7 までのいずれか 1 つに記載の方法である。

#### 【 0 0 8 2 】

第 6 9 の実施形態は、前記固体支持体がビーズを含む、実施形態 6 2 から 6 8 までのいずれか 1 つに記載の方法である。

#### 【 0 0 8 3 】

第 7 0 の実施形態は、前記ビーズが、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズである、実施形態 6 9 に記載の方法である。

#### 【 0 0 8 4 】

第 7 1 の実施形態は、前記コンパートメントタグが、一本鎖または二本鎖核酸分子を含む、実施形態 6 2 から 7 0 までのいずれか 1 つに記載の方法である。

#### 【 0 0 8 5 】

第 7 2 の実施形態は、前記コンパートメントタグが、バーコードおよび任意選択で U M I を含む、実施形態 6 2 から 7 1 までのいずれか 1 つに記載の方法である。

#### 【 0 0 8 6 】

第 7 3 の実施形態は、前記固体支持体がビーズであり、前記コンパートメントタグがバーコードを含み、さらに、前記複数のコンパートメントタグが接合したビーズを、スプリット・アンド・プール ( s p l i t - a n d - p o o l ) 合成によって形成する、実施形態 7 2 に記載の方法である。

#### 【 0 0 8 7 】

第 7 4 の実施形態は、前記固体支持体がビーズであり、前記コンパートメントタグがバーコードを含み、さらに、複数のコンパートメントタグが接合したビーズを、個々の合成または固定化によって形成する、実施形態 7 2 に記載の方法である。

#### 【 0 0 8 8 】

第 7 5 の実施形態は、前記コンパートメントタグが記録タグ内の成分であり、前記記録タグが任意選択でスペーサー、一意の分子識別子、ユニバーサルプライミング部位、またはそれらの任意の組合せをさらに含む、実施形態 6 2 から 7 4 までのいずれか 1 つに記載

10

20

30

40

50

の方法である。

【 0 0 8 9 】

第 7 6 の実施形態は、前記コンパートメントタグが、前記複数のタンパク質複合体、タンパク質、またはポリペプチドの内部アミノ酸または N 末端アミノ酸と反応することが可能な機能的部分をさらに含む、実施形態 6 2 から 7 5 までのいずれか 1 つに記載の方法である。

【 0 0 9 0 】

第 7 7 の実施形態は、前記機能的部分が N H S 基である、実施形態 7 6 に記載の方法である。

【 0 0 9 1 】

第 7 8 の実施形態は、前記機能的部分がアルデヒド基である、実施形態 7 6 に記載の方法である。

【 0 0 9 2 】

第 7 9 の実施形態は、前記複数のコンパートメントタグが、前記コンパートメントタグを前記コンパートメントに印刷、スポットティング、インク噴射すること、またはその組合せによって形成されたものである、実施形態 6 2 から 7 8 までのいずれか 1 つに記載の方法である。

【 0 0 9 3 】

第 8 0 の実施形態は、前記コンパートメントタグが、ペプチドをさらに含む、実施形態 6 2 から 7 9 までのいずれか一項に記載の方法である。

【 0 0 9 4 】

第 8 1 の実施形態は、前記コンパートメントタグペプチドが、タンパク質リガーゼ認識配列を含む、実施形態 8 0 に記載の方法である。

【 0 0 9 5 】

第 8 2 の実施形態は、前記タンパク質リガーゼが、ブテラーゼ I またはそのホモログである、実施形態 8 1 に記載の方法である。

【 0 0 9 6 】

第 8 3 の実施形態は、前記複数のポリペプチドをプロテアーゼで断片化する、実施形態 6 2 から 8 2 までのいずれか一項に記載の方法である。

【 0 0 9 7 】

第 8 4 の実施形態は、前記プロテアーゼがメタロプロテアーゼである、実施形態 8 3 に記載の方法。

【 0 0 9 8 】

第 8 5 の実施形態は、前記メタロプロテアーゼの活性が金属カチオンの光活性化放出によってモジュレートされる、実施形態 8 4 に記載の方法である。

【 0 0 9 9 】

第 8 6 の実施形態は、前記複数のポリペプチドを前記複数のコンパートメントに分配する前に 1 つまたは複数の豊富なタンパク質を前記試料からサブトラクションすることをさらに含む、実施形態 6 2 から 8 5 までのいずれか 1 つに記載の方法である。

【 0 1 0 0 】

第 8 7 の実施形態は、前記複数のペプチドと前記コンパートメントタグを接合する前に前記コンパートメントタグを前記固体支持体から遊離させることをさらに含む、実施形態 6 2 から 8 6 までのいずれか 1 つに記載の方法である。

【 0 1 0 1 】

第 8 8 の実施形態は、ステップ ( d ) の後に、前記コンパートメントタグ付きペプチドを固体支持体に記録タグを伴って接合させることをさらに含む、実施形態 6 2 に記載の方法である。

【 0 1 0 2 】

第 8 9 の実施形態は、前記コンパートメントタグ付きペプチド上の前記コンパートメントタグの情報を前記付随する記録タグに移行させることをさらに含む、実施形態 8 8 に記

10

20

30

40

50

載の方法である。

【 0 1 0 3 】

第 9 0 の実施形態は、ステップ ( e ) の前に前記コンパートメントタグを前記コンパートメントタグ付きペプチドから除去することをさらに含む、実施形態 8 9 に記載の方法である。

【 0 1 0 4 】

第 9 1 の実施形態は、解析されるペプチドが由来する前記単一の細胞の同一性を前記解析されるペプチドのコンパートメントタグ配列に基づいて決定することをさらに含む、実施形態 6 2 から 9 0 までのいずれか 1 つに記載の方法である。

【 0 1 0 5 】

第 9 2 の実施形態は、解析されるペプチドが由来する前記タンパク質またはタンパク質複合体の同一性を前記解析されるペプチドのコンパートメントタグ配列に基づいて決定することをさらに含む、実施形態 6 2 から 9 0 までのいずれか 1 つに記載の方法である。

【 0 1 0 6 】

第 9 3 の実施形態は、複数の巨大分子を解析するための方法であって、

( a ) 固体支持体に接合した複数の巨大分子および付随する記録タグを用意するステップと；

( b ) 前記複数の巨大分子を、前記複数の巨大分子に結合することが可能な複数の結合性物質であって、各結合性物質が前記結合性物質に関する識別情報を有するコーディングタグを含む複数の結合性物質と接触させるステップと；

( c ) ( i ) 前記巨大分子に付随する記録タグの情報を前記巨大分子に結合した前記結合性物質の前記コーディングタグに移行させて、伸長コーディングタグを生成するステップ；または ( i i ) 巨大分子に付随する記録タグおよび前記巨大分子に結合した前記結合性物質のコーディングタグの情報をジタグ構築物に移行するステップと；

( d ) 前記伸長コーディングタグまたはジタグ構築物を収集するステップと；

( e ) 任意選択でステップ ( b ) ~ ( d ) を 1 回または複数回の結合サイクルにわたって繰り返すステップと；

( f ) 伸長コーディングタグまたはジタグ構築物の収集物を解析するステップとを含む方法である。

【 0 1 0 7 】

第 9 4 の実施形態は、前記巨大分子がタンパク質である、実施形態 9 3 に記載の方法である。

【 0 1 0 8 】

第 9 5 の実施形態は、前記巨大分子が、ペプチドである、実施形態 9 3 に記載の方法である。

【 0 1 0 9 】

第 9 6 の実施形態は、前記ペプチドが、生体試料由来の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 9 5 に記載の方法である。

【 0 1 1 0 】

第 9 7 の実施形態は、前記記録タグが、DNA 分子、RNA 分子、PNA 分子、BNA 分子、XNA 分子、LNA 分子、PNA 分子、またはこれらの組合せである、実施形態 9 3 から 9 6 までのいずれか 1 つに記載の方法である。

【 0 1 1 1 】

第 9 8 の実施形態は、前記記録タグが、一意の分子識別子 ( UMI ) を含む、実施形態 9 3 から 9 7 までのいずれか 1 つに記載の方法である。

【 0 1 1 2 】

第 9 9 の実施形態は、前記記録タグが、コンパートメントタグを含む、実施形態 9 3 から 9 8 までに記載の方法である。

【 0 1 1 3 】

第 1 0 0 の実施形態は、前記記録タグが、ユニバーサルプライミング部位を含む、実施

10

20

30

40

50

形態 93 から 99 までのいずれか 1 つに記載の方法である。

【0114】

第 101 の実施形態は、前記記録タグが、その 3' 末端にスパーサーを含む、実施形態 93 から 100 までのいずれか 1 つに記載の方法である。

【0115】

第 102 の実施形態は、前記記録タグの 3' 末端をブロックしてポリメラーゼによる前記記録タグの伸長を防止し、巨大分子に付随する記録タグおよび前記巨大分子に結合している前記結合性物質のコーディングタグの情報をジタグ構築物に移行させる、実施形態 93 から 101 までのいずれか 1 つに記載の方法である。

【0116】

第 103 の実施形態は、前記コーディングタグが、エンコーダー配列を含む、実施形態 93 から 102 までのいずれか 1 つに記載の方法である。

【0117】

第 104 の実施形態は、前記コーディングタグが、UMI を含む、実施形態 93 から 103 までのいずれか 1 つに記載の方法である。

【0118】

第 105 の実施形態は、前記コーディングタグが、ユニバーサルプライミング部位を含む、実施形態 93 から 104 までのいずれか 1 つに記載の方法である。

【0119】

第 106 の実施形態は、前記コーディングタグが、その 3' 末端にスパーサーを含む、実施形態 93 から 105 までのいずれか 1 つに記載の方法である。

【0120】

第 107 の実施形態は、前記コーディングタグが、結合サイクル特異的配列を含む、実施形態 93 から 106 までのいずれか 1 つに記載の方法である。

【0121】

第 108 の実施形態は、前記結合性物質と前記コーディングタグが、リンカーによって接合されている、実施形態 93 から 107 までのいずれか 1 つに記載の方法である。

【0122】

第 109 の実施形態は、前記記録タグの情報の前記コーディングタグへの移行が、プライマー伸長によってもたらされる、実施形態 93 から 108 までのいずれか 1 つに記載の方法である。

【0123】

第 110 の実施形態は、前記記録タグの情報の前記コーディングタグへの移行が、ライゲーションによってもたらされる、実施形態 93 から 108 までのいずれか 1 つに記載の方法である。

【0124】

第 111 の実施形態は、前記ジタグ構築物が、ギャップ充填、プライマー伸長、またはその両方によって生成される、実施形態 93 から 108 までのいずれか 1 つに記載の方法である。

【0125】

第 112 の実施形態は、前記ジタグ分子が、前記記録タグに由来するユニバーサルプライミング部位、前記記録タグに由来するコンパートメントタグ、前記記録タグに由来する一意の分子識別子、前記記録タグに由来する任意選択のスパーサー、前記コーディングタグに由来するエンコーダー配列、前記コーディングタグに由来する一意の分子識別子、前記コーディングタグに由来する任意選択のスパーサー、および前記コーディングタグに由来するユニバーサルプライミング部位を含む、実施形態 93 から 97 まで、107、108、および 111 のいずれか 1 つに記載の方法である。

【0126】

第 113 の実施形態は、前記巨大分子および前記付随する記録タグを、前記固体支持体に共有結合により接合させる、実施形態 93 から 112 までのいずれか 1 つに記載の方法

10

20

30

40

50

である。

【 0 1 2 7 】

第 1 1 4 の実施形態は、前記固体支持体が、ビーズ、多孔質ビーズ、多孔質マトリックス、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、ナイロン、シリコンウェハチップ、フロースルーチップ、信号変換電子機器を含むバイオチップ、マイクロタイターウェル、E L I S A プレート、スピン干渉ディスク、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアである、実施形態 1 1 3 に記載の方法である。

【 0 1 2 8 】

第 1 1 5 の実施形態は、前記固体支持体が、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズである、実施形態 1 1 4 に記載の方法である。

【 0 1 2 9 】

第 1 1 6 の実施形態は、前記結合性物質が、ポリペプチドまたはタンパク質である、実施形態 9 3 から 1 1 5 までのいずれか 1 つに記載の方法である。

【 0 1 3 0 】

第 1 1 7 の実施形態は、前記結合性物質が、改変アミノペプチダーゼ、改変アミノアシル t R N A 合成酵素、改変アンチカリン、または抗体もしくはその結合性断片である、実施形態 1 1 6 に記載の方法である。

【 0 1 3 1 】

第 1 1 8 の実施形態は、前記結合性物質が、単一のアミノ酸残基、ジペプチド、トリペプチドまたは前記ペプチドの翻訳後修飾に結合する、実施形態 9 5 から 1 1 7 までのいずれか 1 つに記載の方法である。

【 0 1 3 2 】

第 1 1 9 の実施形態は、前記結合性物質が、N 末端アミノ酸残基、C 末端アミノ酸残基、または内部アミノ酸残基に結合する、実施形態 1 1 8 に記載の方法である。

【 0 1 3 3 】

第 1 2 0 の実施形態は、前記結合性物質が、N 末端ペプチド、C 末端ペプチド、または内部ペプチドに結合する、実施形態 1 1 8 に記載の方法である。

【 0 1 3 4 】

第 1 2 1 の実施形態は、前記結合性物質が N 末端アミノ酸残基に結合し、前記 N 末端アミノ酸残基が各結合サイクル後に切断される、実施形態 1 1 9 に記載の方法である。

【 0 1 3 5 】

第 1 2 2 の実施形態は、前記結合性物質が C 末端アミノ酸残基に結合し、前記 C 末端アミノ酸残基が各結合サイクル後に切断される、実施形態 1 1 9 に記載の方法である。

【 0 1 3 6 】

実施形態 1 2 3 . N 末端アミノ酸残基がエドマン分解によって切断される、実施形態 1 2 1 に記載の方法。

【 0 1 3 7 】

実施形態 1 2 4 . 前記結合性物質が、アミノ酸または翻訳後修飾の部位特異的な共有結合性標識である、実施形態 9 3 に記載の方法。

【 0 1 3 8 】

実施形態 1 2 5 . ステップ ( b ) の後に、前記巨大分子および付随する結合性物質を含む複合体を前記固体支持体から解離させ、液滴またはマイクロ流体液滴のエマルジョン中に分配する、実施形態 9 3 から 1 2 4 までのいずれか 1 つに記載の方法。

【 0 1 3 9 】

実施形態 1 2 6 . 各マイクロ流体液滴が、平均して、前記巨大分子および前記結合性物質を含む複合体を 1 つ含む、実施形態 1 2 5 に記載の方法。

【 0 1 4 0 】

実施形態 1 2 7 . 伸長コーディングタグまたはジタグ構築物を生成する前に前記記録タグ

10

20

30

40

50

を増幅する、実施形態 1 2 5 または 1 2 6 に記載の方法。

【 0 1 4 1 】

実施形態 1 2 8 . エマルジョン融合 P C R を使用して、前記記録タグ情報を前記コーディングタグに移行させる、またはジタグ構築物の集団を創出する、実施形態 1 2 5 から 1 2 7 までのいずれか 1 つに記載の方法。

【 0 1 4 2 】

実施形態 1 2 9 . 伸長コーディングタグまたはジタグ構築物の収集物を解析前に増幅させる、実施形態 9 3 から 1 2 8 のいずれか 1 つに記載の方法。

【 0 1 4 3 】

実施形態 1 3 0 . 伸長コーディングタグまたはジタグ構築物の収集物の解析が、核酸配列決定法を含む、実施形態 9 3 から 1 2 9 までのいずれか 1 つに記載の方法。

10

【 0 1 4 4 】

実施形態 1 3 1 . 前記核酸配列決定法が、合成による配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、ポロニーシーケンシング、イオン半導体シーケンシング、またはパイロシーケンシングである、実施形態 1 3 0 に記載の方法。

【 0 1 4 5 】

実施形態 1 3 2 . 前記核酸配列決定法が、単一分子リアルタイムシーケンシング、ナノポアに基づく配列決定、または先端顕微鏡を使用した D N A のダイレクトイメージングである、実施形態 1 3 0 に記載の方法。

【 0 1 4 6 】

20

実施形態 1 3 3 . 前記巨大分子の部分的組成を、一意のコンパートメントタグおよび任意選択で U M I を使用する複数の伸長コーディングタグまたはジタグ構築物の解析によって決定する、実施形態 1 3 0 に記載の方法。

【 0 1 4 7 】

実施形態 1 3 4 . 前記解析ステップを、塩基当たりのエラー率が > 5 %、> 1 0 %、> 1 5 %、> 2 0 %、> 2 5 %、または > 3 0 % である配列決定法を用いて実施する、実施形態 1 から 1 3 3 までのいずれか 1 つに記載の方法。

【 0 1 4 8 】

実施形態 1 3 5 . コーディングタグ、記録タグ、またはその両方の識別成分が、エラー訂正コードを含む、実施形態 1 から 1 3 4 までのいずれか 1 つに記載の方法。

30

【 0 1 4 9 】

実施形態 1 3 6 . 前記識別成分が、エンコーダー配列、バーコード、U M I、コンパートメントタグ、サイクル特異的配列、またはそれらの任意の組合せから選択される、実施形態 1 3 5 に記載の方法。

【 0 1 5 0 】

実施形態 1 3 7 . 前記エラー訂正コードが、H a m m i n g コード、L e e 距離コード、非対称 L e e 距離コード、R e e d - S o l o m o n コード、および L e v e n s h t e i n - T e n e n g o l t s コードから選択される、実施形態 1 3 5 または 1 3 6 に記載の方法。

【 0 1 5 1 】

40

実施形態 1 3 8 . コーディングタグ、記録タグ、またはその両方の識別成分が、一意の電流またはイオンフラックスまたは光学的シグネチャを生成することが可能であり、前記解析ステップが、前記識別成分を識別するために前記一意の電流またはイオンフラックスまたは光学的シグネチャを検出することを含む、実施形態 1 から 1 3 4 までのいずれか 1 つに記載の方法。

【 0 1 5 2 】

実施形態 1 3 9 . 前記識別成分が、エンコーダー配列、バーコード、U M I、コンパートメントタグ、サイクル特異的配列、またはそれらの任意の組合せから選択される、実施形態 1 3 8 に記載の方法。

【 0 1 5 3 】

50

実施形態 1 4 0 . 複数の巨大分子を解析するための方法であって、

( a ) 固体支持体に接合した複数の巨大分子および付随する記録タグを用意するステップと ;

( b ) 前記複数の巨大分子を、同類の巨大分子に結合することが可能な複数の結合性物質であって、各結合性物質が前記結合性物質に関する識別情報を有するコーディングタグを含む複数の結合性物質と接触させるステップと ;

( c ) 第 1 の結合性物質の第 1 のコーディングタグの情報を第 1 の巨大分子に付随する第 1 の記録タグに移行させて、一次伸長記録タグを生成するステップであって、前記第 1 の結合性物質が前記第 1 の巨大分子に結合するステップと ;

( d ) 前記複数の巨大分子を、同類の巨大分子に結合することが可能な複数の結合性物質と接触させるステップと ;

10

( e ) 第 2 の結合性物質の第 2 のコーディングタグの情報を前記一次伸長記録タグに移行させて、二次伸長記録タグを生成するステップであって、前記第 2 の結合性物質が前記第 1 の巨大分子に結合するステップと ;

( f ) 任意選択でステップ ( d ) ~ ( e ) を「 n 」回の結合サイクルにわたって繰り返すステップであって、前記第 1 の巨大分子に結合する各結合性物質の各コーディングタグの情報を前の結合サイクルで生成した伸長記録タグに移行させて、前記第 1 の巨大分子を表す n 次伸長記録タグを生成するステップと ;

( g ) 前記 n 次伸長記録タグを解析するステップとを含む方法。

20

【 0 1 5 4 】

実施形態 1 4 1 . 複数の巨大分子を表す複数の n 次伸長記録タグを生成し、解析する、実施形態 1 4 0 に記載の方法。

【 0 1 5 5 】

実施形態 1 4 2 . 前記巨大分子が、タンパク質である、実施形態 1 4 0 または 1 4 1 に記載の方法。

【 0 1 5 6 】

実施形態 1 4 3 . 前記巨大分子が、ペプチドである、実施形態 1 4 2 に記載の方法。

【 0 1 5 7 】

実施形態 1 4 4 . 前記ペプチドが、生体試料由来の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 1 4 3 に記載の方法。

30

【 0 1 5 8 】

実施形態 1 4 5 . 前記複数の巨大分子が、多数のプールされた試料由来の巨大分子を含む、実施形態 1 4 0 から 1 4 4 までのいずれか 1 つに記載の方法。

【 0 1 5 9 】

実施形態 1 4 6 . 前記記録タグが、DNA 分子、RNA 分子、BNA 分子、XNA 分子、LNA 分子、PNA 分子、PNNA 分子、またはこれらの組合せである、実施形態 1 4 0 から 1 4 5 までのいずれか 1 つに記載の方法。

【 0 1 6 0 】

実施形態 1 4 7 . 前記記録タグが、一意の分子識別子 ( UMI ) を含む、実施形態 1 4 0 から 1 4 6 までに記載の方法。

40

【 0 1 6 1 】

実施形態 1 4 8 . 前記記録タグが、コンパートメントタグを含む、実施形態 1 4 0 から 1 4 7 までに記載の方法。

【 0 1 6 2 】

実施形態 1 4 9 . 前記記録タグが、ユニバーサルプライミング部位を含む、実施形態 1 4 0 から 1 4 8 までのいずれか 1 つに記載の方法。

【 0 1 6 3 】

実施形態 1 5 0 . 前記記録タグが、その 3 ' 末端にスパーサーを含む、実施形態 1 4 0 から 1 4 9 までのいずれか 1 つに記載の方法。

50

## 【 0 1 6 4 】

実施形態 1 5 1 . 前記コーディングタグが、エンコーダー配列を含む、実施形態 1 4 0 から 1 5 0 までのいずれか 1 つに記載の方法。

## 【 0 1 6 5 】

実施形態 1 5 2 . 前記コーディングタグが、U M I を含む、実施形態 1 4 0 から 1 5 1 までのいずれか 1 つに記載の方法。

## 【 0 1 6 6 】

実施形態 1 5 3 . 前記コーディングタグが、ユニバーサルプライミング部位を含む、実施形態 1 4 0 から 1 5 2 までのいずれか 1 つに記載の方法。

## 【 0 1 6 7 】

実施形態 1 5 4 . 前記コーディングタグが、その 3 ' 末端にスペーサーを含む、実施形態 1 4 0 から 1 5 3 までのいずれか 1 つに記載の方法。

## 【 0 1 6 8 】

実施形態 1 5 5 . 前記コーディングタグが、結合サイクル特異的配列を含む、実施形態 1 4 0 から 1 5 4 までのいずれか 1 つに記載の方法。

## 【 0 1 6 9 】

実施形態 1 5 6 . 前記コーディングタグが、一意の分子識別子を含む、実施形態 1 4 0 から 1 5 5 までのいずれか 1 つに記載の方法。

## 【 0 1 7 0 】

実施形態 1 5 7 . 前記結合性物質と前記コーディングタグが、リンカーによって接合されている、実施形態 1 4 0 から 1 5 6 までのいずれか 1 つに記載の方法。

## 【 0 1 7 1 】

実施形態 1 5 8 . 前記記録タグの情報の前記コーディングタグへの移行が、プライマー伸長によって媒介される、実施形態 1 4 0 から 1 5 7 までのいずれか 1 つに記載の方法。

## 【 0 1 7 2 】

実施形態 1 5 9 . 前記記録タグの情報の前記コーディングタグへの移行が、ライゲーションによって媒介される、実施形態 1 4 0 から 1 5 8 までのいずれか 1 つに記載の方法。

## 【 0 1 7 3 】

実施形態 1 6 0 . 前記複数の巨大分子、前記付随する記録タグ、またはその両方が、前記固体支持体に共有結合により接合している、実施形態 1 4 0 から 1 5 9 までのいずれか 1 つに記載の方法。

## 【 0 1 7 4 】

実施形態 1 6 1 . 前記固体支持体が、ビーズ、多孔質ビーズ、多孔質マトリックス、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、ナイロン、シリコンウェーハチップ、フロースルーチップ、信号変換電子機器を含むバイオチップ、マイクロタイターウェル、E L I S A プレート、スピン干渉ディスク、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアである、実施形態 1 4 0 から 1 6 0 までのいずれか 1 つに記載の方法。

## 【 0 1 7 5 】

実施形態 1 6 2 . 前記固体支持体が、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズである、実施形態 1 6 1 に記載の方法。

## 【 0 1 7 6 】

実施形態 1 6 3 . 前記結合性物質が、ポリペプチドまたはタンパク質である、実施形態 1 4 0 から 1 6 2 までのいずれか 1 つに記載の方法。

## 【 0 1 7 7 】

実施形態 1 6 4 . 前記結合性物質が、改変アミノペプチダーゼ、改変アミノアシル t R N A 合成酵素、改変アンチカリン、または抗体もしくはその結合性断片である、実施形態 1 6 3 に記載の方法。

## 【 0 1 7 8 】

10

20

30

40

50



実施形態 165 . 前記結合性物質が、単一のアミノ酸残基、ジペプチド、トリペプチドまたは前記ペプチドの翻訳後修飾に結合する、実施形態 142 から 164 までのいずれか 1 つに記載の方法。

【0179】

実施形態 166 . 前記結合性物質が、N末端アミノ酸残基、C末端アミノ酸残基、または内部アミノ酸残基に結合する、実施形態 165 に記載の方法。

【0180】

実施形態 167 . 前記結合性物質が、N末端ペプチド、C末端ペプチド、または内部ペプチドに結合する、実施形態 165 に記載の方法。

【0181】

実施形態 168 . 前記結合性物質が、修飾されたN末端アミノ酸残基、修飾されたC末端アミノ酸残基、または修飾された内部アミノ酸残基の化学標識に結合する、実施形態 142 から 164 までのいずれか 1 つに記載の方法。

【0182】

実施形態 169 . 前記結合性物質がN末端アミノ酸残基または前記修飾されたN末端アミノ酸残基の化学標識に結合し、前記N末端アミノ酸残基が各結合サイクル後に切断される、実施形態 166 または 168 に記載の方法。

【0183】

実施形態 170 . 前記結合性物質がC末端アミノ酸残基または前記修飾されたC末端アミノ酸残基に化学標識に結合し、前記C末端アミノ酸残基が各結合サイクル後に切断される、実施形態 166 または 168 に記載の方法。

【0184】

実施形態 171 . 前記N末端アミノ酸残基が、エドマン分解、エドマナーゼ、改変アミノペプチダーゼ、または改変アシルペプチドヒドロラーゼによって切断される、実施形態 169 に記載の方法。

【0185】

実施形態 172 . 前記結合性物質が、アミノ酸または翻訳後修飾の部位特異的な共有結合性標識である、実施形態 163 に記載の方法。

【0186】

実施形態 173 . 前記複数のn次伸長記録タグを解析前に増幅させる、実施形態 140 から 172 までのいずれか 1 つに記載の方法。

【0187】

実施形態 174 . 前記n次伸長記録タグの解析が、核酸配列決定法を含む、実施形態 140 から 173 までのいずれか 1 つに記載の方法。

【0188】

実施形態 175 . 複数の巨大分子を表す複数のn次伸長記録タグを並行して解析する、実施形態 174 に記載の方法。

【0189】

実施形態 176 . 前記核酸配列決定法が、合成による配列決定、ライゲーシオンによる配列決定、ハイブリダイゼーションによる配列決定、ポロニーシーケンシング、イオン半導体シーケンシング、またはパイロシーケンシングである、実施形態 174 または 175 に記載の方法。

【0190】

実施形態 177 . 前記核酸配列決定法が、単一分子リアルタイムシーケンシング、ナノポアに基づく配列決定、または先端顕微鏡を使用したDNAのダイレクトイメージングである、実施形態 174 または 175 に記載の方法。

【0191】

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ていない。例示が目的であるため、すべての構成要素がすべて図面で標記されているとは限らず、また、当業者による本発明の理解に例示が必要

10

20

30

40

50

とされない場合は、本発明の各実施形態のすべての構成要素が示されているとも限らない。

【図面の簡単な説明】

【0192】

【図1 - 1】図1 Aは、図面に示されている機能的エレメントの凡例を示す。図1 Bは、タンパク質コードをDNAコードへと転換するための基本的概要を示す。この転換では、複数のタンパク質またはポリペプチドが複数のペプチドに断片化され、それらがその後複数のペプチドを表わす伸長記録タグのライブラリーへと変換される。伸長記録タグは、ペプチド配列を表わすDNAコード付きライブラリーを構成する。ライブラリーは、適切に修飾して、任意の次世代シーケンシング(NGS)プラットフォームで配列決定することができる。

10

【図1 - 2】同上。

【0193】

【図2 - 1】図2 A ~ 2 Dは、単一のまたは複数の記録タグと共局在化または共標識されている固定化タンパク質と相互作用するコーディングタグを含む結合性物質(例えば、抗体、アンチカリン、N-レコグニンタンパク質(N-recognins protein)(例えば、ATP依存性Clpプロテアーゼアダプタータンパク質(Cl pS)、アプタマーなど、およびそれらの変異体/ホモログ)の複数のサイクルを使用する、本明細書で開示されている方法によるタンパク質巨大分子解析の例を示す。記録タグは、ユニバーサルプライミング部位、バーコード(例えば、分配バーコード、コンパートメントバーコード、画分バーコード)、任意選択の一意の分子識別子(UMI)配列、およびコーディングタグの情報移行に使用されるスペーサー配列(Sp)で構成される。スペーサー配列(Sp)は、すべての結合サイクルにわたって一定であってもよく、結合性物質特異的であってもよく、または結合サイクル数特異的であってもよい。コーディングタグは、結合性物質の識別情報を提供するエンコーダー配列、任意選択のUMI、および記録タグの相補的スペーサー配列にハイブリダイズし、コーディングタグ情報の記録タグへの移行(例えば、プライマー伸長、本明細書ではポリメラーゼ伸長とも呼ばれる)を容易にするスペーサー配列で構成されている。図2 Aは、同種結合性物質とタンパク質とのサイクル結合により伸長記録タグを創出するプロセス、および結合性物質のコーディングタグからタンパク質の記録タグへの対応する情報移行を示す。一連の連続した結合およびコーディングタグ情報移行ステップの後、結合性物質(例えば、抗体1(Ab1)、抗体2(Ab2)、抗体3(Ab3)、...抗体「n」(Abn))の識別情報を提供する、「n」結合サイクルからのエンコーダー配列を含む結合性物質コーディングタグ情報、記録タグに由来するバーコード/任意選択のUMI、結合性物質のコーディングタグに由来する任意選択のUMI配列、ならびに増幅およびデジタル次世代シーケンシングによる解析を容易にするための、ライブラリー構築物の各末端にある隣接ユニバーサルプライミング配列を含む最終伸長記録タグが産生される。図2 Bは、DNAバーコード付き記録タグでタンパク質を標識するためのスキームの例を示す。上段パネルでは、N-ヒドロキシスクシンイミド(NHS)は、アミン反応性カップリング剤であり、ジベンゾシクロオクチル(DBCO)は、固形基材の表面への「クリック」カップリングに有用な歪みアルキンである。このスキームでは、記録タグは、タンパク質のリシン(K)残基(および任意選択でN末端アミノ酸)の アミンに、NHS部分を介してカップリングされる。下段パネルでは、ヘテロ二機能性リンカーであるNHSアルキンを使用して、リシン(K)残基の アミンを標識し、アルキン「クリック」部分を創出する。その後、アジド標識DNA記録タグを、標準的クリック化学によりこれら反応性アルキン基に容易に付着させることができる。さらに、DNA記録タグは、逆IEDDA反応によりTCO誘導体化配列決定基材と下流でカップリングするための直交性メチルテトラジン(mTet)部分を用いて設計することができる。図2 Cは、記録タグを使用したタンパク質解析法の2つの例を示す。上段パネルでは、タンパク質巨大分子は、捕捉剤により固体支持体に固定化されており、任意選択で架橋されている。タンパク質または捕捉剤はいずれも、記録タグで標識されていてもよい。下段パネルでは、付随する記録タグを有するタンパク質は、固体支持体に直接固定

20

30

40

50

化されている。図 2 D は、同種結合物質の DNA コード化、および得られた伸長記録タグの配列決定を使用したタンパク質イムノアッセイの全体的なワークフローの例を示す。タンパク質は、記録タグによりバーコード化（つまりインデックス化）し、サイクル結合解析前にプールのことができ、試料スループットおよび結合試薬節約が大幅に増加される試料であってもよい。この手法は、逆相タンパク質アッセイ（RPPA）を実施するための、有効なデジタル式のより単純でより大規模化可能な手法である。

【図 2 - 2】同上。

【図 2 - 3】同上。

【図 2 - 4】同上。

【0194】

【図 3】図 3 A ~ 3 D は、ペプチド配列を表わす DNA 伸長記録タグを構築することによる、分解に基づくペプチド配列決定のプロセスを示す。これは、N 末端アミノ酸（NTAA）結合、ペプチドに付着している記録タグへのコーディングタグ情報の移行、NTAA 切断のサイクルプロセスを使用し、このプロセスをすべて固体支持体でサイクル様式で繰り返すエドマン分解様手法により達成される。ペプチドの N 末端分解に由来する伸長記録タグの例示的な構築の概要が提供されている：（A）ペプチドの N 末端アミノ酸を標識する（例えば、フェニルチオカルバモイル（PTC）、ジニトロフェニル（DNP）、スルホニルニトロフェニル（SNP）、アセチル、またはグアニジン（guanidyl）部分で）；（B）は、標識 NTAA に結合された結合性物質および付随するコーディングタグを示す；（C）は、固体支持体（例えば、ビーズ）に結合され、記録タグに（例えば、三機能性リンカーを介して）付随するペプチドを示し、結合性物質がペプチドの NTAA に結合すると、コーディングタグの情報が記録タグに移行され（例えば、プライマー伸長により）、伸長記録タグが生成される；（D）標識 NTAA を、化学的または酵素的な手段により切断して、新しい NTAA を露出させる。矢印により示されているように、このサイクルを「n」回繰り返して、最終伸長記録タグを生成する。最終伸長記録タグは、任意選択で、下流の増幅および DNA 配列決定を容易にするために、ユニバーサルプライミング部位により隣接されている。フォワードユニバーサルプライミング部位（例えば、Illumina の P5 - S1 配列）は、元の記録タグ設計の一部であってもよく、リバーズユニバーサルプライミング部位（例えば、Illumina の P7 - S2' 配列）は、記録タグの伸長の最終ステップとして添加してもよい。この最終ステップは、結合性物質とは独立して実施してもよい。

【0195】

【図 4 - 1】図 4 A ~ 4 B は、本明細書で開示されている方法による例示的なタンパク質配列決定ワークフローを示す。図 4 A は、例示的なワークフローを示し、代替モードが明灰色の破線で概説されており、四角の中に示されている特定の実施形態が矢印で関連付けられている。ワークフローの各ステップの代替モードが、矢印下方の四角の中に示されている。図 4 B は、サイクル結合およびコーディングタグ情報移行ステップを実施する際の、情報移行の効率を向上させるための選択肢を示す。1 分子当たり複数の記録タグを用いることができる。さらに、所与の結合事象毎に、コーディングタグ情報の記録タグへの移行を複数回実施してもよく、またはその代わりに表面増幅ステップを用いて、伸長記録タグライブラリーなどのコピーを創出してもよい。

【図 4 - 2】同上。

【0196】

【図 5】図 5 A ~ 5 B は、プライマー伸長を使用して、結合性物質のコーディングタグの識別情報を、巨大分子（例えば、ペプチド）に付随する記録タグへと移行して伸長記録タグを生成するための伸長記録タグの例示的な構築の概要を示す。結合性物質に関する識別情報を有する一意のエンコーダー配列を含むコーディングタグは各末端が、任意選択で、共通スペーサー配列（Sp'）により隣接されている。図 5 A は、ビーズに連結されている記録タグ標識ペプチドの NTAA に結合するコーディングタグを含む NTAA 結合性物質を示す。記録タグは、相補的スペーサー配列（Sp）を介してコーディングタグにアニー

10

20

30

40

50

リングし、プライマー伸長反応は、スペーサー (S p) をプライミング部位として使用したコーディングタグ情報の記録タグへの移行を媒介する。コーディングタグは、結合性物質から遠位にある末端での一本鎖スペーサー (S p') 配列との二本鎖として示されている。この構成は、記録タグの内部部位へのコーディングタグのハイブリダイゼーションを最小限に抑え、記録タグの末端スペーサー (S p) 配列と、コーディングタグの一本鎖スペーサー突出 (S p') とのハイブリダイゼーションに有利に働く。さらに、伸長記録タグは、コーディングタグと内部記録タグ配列エレメントとのハイブリダイゼーションを阻止するために、オリゴヌクレオチド (エンコーダー、スペーサー配列に相補的な) と事前にアニーリングされていてもよい。図 5 B は、結合、およびコーディングタグ情報の移行、およびユニバーサルプライミング部位の 3' 末端への付加の「n」回のサイクル後に産生された最終伸長記録タグを示す (「\*\*\*」は、伸長記録タグに示されていないその間の結合サイクルを表わす)。

【0197】

【図 6】図 6 は、コーディングタグ情報が、酵素的ライゲーションにより伸長記録タグへと移行されることを示す。それぞれの記録タグを有する 2 つの異なる巨大分子が示されており、記録タグ伸長は並行して進行する。ライゲーションは、スペーサー配列 (S p) が、記録タグの相補的スペーサー (S p') とアニーリングする「粘着末端」突出を有するように、二本鎖コーディングタグを設計することにより容易になり得る。二本鎖コーディングタグの相補鎖は、情報を記録タグへと移行させる。ライゲーションを使用して記録タグを伸長させる場合、伸長の方向は、示されているように、5' から 3' へであってもよく、

【0198】

【図 7】図 7 は、スペーサー配列を伸長記録タグに挿入せずに、記録タグまたは伸長記録タグの 3' ヌクレオチドを、コーディングタグ (またはその相補体) の 5' ヌクレオチドと連結させる化学的ライゲーションにより、コーディングタグ情報を記録タグへと移行させる、「無スペーサー」手法を示す。また、伸長記録タグおよびコーディングタグの向きを逆にして、記録タグの 5' 末端が、コーディングタグ (または相補体) の 3' 末端にライゲーションされるようにしてもよい。示されている例では、記録タグの相補的「ヘルパー」オリゴヌクレオチド配列 (「記録ヘルパー」) とコーディングタグとのハイブリダイゼーションを使用して、複合体を安定させ、記録タグのコーディングタグ相補鎖への特異的な化学的ライゲーションを可能にさせる。得られた伸長記録タグは、スペーサー配列を欠いている。また、DNA、PNA、または類似の核酸ポリマーを用いることができる「クリック化学」型の化学的ライゲーション (例えば、アジドおよびアルキン部分 (3 本線記号として示されている) が使用される) が示されている。

【0199】

【図 8】図 8 A ~ 8 B は、N 末端アミノ酸分解の前に、ペプチドの翻訳後修飾 (PTM) 情報を伸長記録タグに書き込むための例示的な方法を示す。図 8 A : 結合性物質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性物質 (例えば、ホスホチロシン抗体の識別情報を有するコーディングタグを含むホスホチロシン抗体) を、ペプチドに結合することが可能である。ホスホチロシンが、図示されているように記録タグ標識ペプチドに存在する場合、ホスホチロシン抗体がホスホチロシンに結合すると、コーディングタグおよび記録タグが、相補的スペーサー配列を介してアニーリングし、コーディングタグ情報が記録タグに移行され、伸長記録タグが生成される。図 8 B : 伸長記録タグは、一次アミノ酸配列 (例えば、「aa<sub>1</sub>」、「aa<sub>2</sub>」、「aa<sub>3</sub>」、...、「aa<sub>N</sub>」) およびペプチドの翻訳後修飾 (例えば、「PTM<sub>1</sub>」、「PTM<sub>2</sub>」) の両方のコーディングタグ情報を含んでいてもよい。

【0200】

【図 9】図 9 A ~ 9 B は、結合性物質を巨大分子に結合させ、結合性物質に付着しているコーディングタグの情報を、固体支持体 (例えば、ビーズ) に付着している単一の巨大分子の部位に共局在化されている複数の記録タグ中の個々の記録タグへと移行させ、それに

10

20

30

40

50

より巨大分子を集散的に表す複数の伸長記録タグを生成する複数サイクルのプロセスを示す。この図では、例示のために過ぎないが、巨大分子は、ペプチドであり、各サイクルは、結合性物質を、N末端アミノ酸（NTAA）に結合させること、コーディングタグ情報を記録タグに移行させることにより結合事象を記録すること、その後そのNTAAを除去して、新しいNTAAを露出させることを含む。図9Aは、巨大分子と共に固体支持体に共局在化されている複数の記録タグ（ユニバーサルフォワードプライミング配列およびUMIを含む）を示す。個々の記録タグは、伸長反応をプライムして、コーディングタグ情報を記録タグへと移行するために使用することができる、結合性物質のコーディングタグ内の共通スパーサー配列に相補的な共通スパーサー配列（Sp）を有する。図9Bは、各連続サイクルの結合に使用されるサイクル特異的NTAA結合性物質の様々なプールを示し、各プールは、サイクル特異的スパーサー配列を有する。

10

#### 【0201】

【図10】図10A～10Cは、結合性物質に付着しているコーディングタグの情報を、固体支持体（例えば、ビーズ）に付着している単一の巨大分子の部位に共局在化されている複数の記録タグのうちの記録タグへと移行させ、それにより巨大分子を集散的に表す複数の伸長記録タグを生成する複数のサイクルを含む例示的モードを示す。この図では、例示のために過ぎないが、巨大分子は、ペプチドであり、プロセスの各ラウンドは、NTAAと結合させること、結合事象を記録すること、その後そのNTAAを除去して、新しいNTAAを露出させることを含む。図10Aは、好ましくは1ビーズ当たり単一分子の巨大分子と共に固体支持体に共局在化された複数の記録タグ（ユニバーサルフォワードプライミング配列およびUMIを含む）を示す。個々の記録タグは、異なる「サイクル特異的」配列（例えば、C<sub>1</sub>、C<sub>2</sub>、C<sub>3</sub>、...C<sub>n</sub>）を有する異なるスパーサー配列を3'末端に有する。好ましくは、各ビーズの記録タグは、同じUMI配列を共有する。第1のサイクルの結合（サイクル1）では、複数のNTAA結合性物質を巨大分子と接触させる。サイクル1で使用される結合性物質は、記録タグのサイクル1C<sub>1</sub>スパーサー配列に相補的な共通5'スパーサー配列（C'<sub>1</sub>）を有する。また、サイクル1で使用される結合性物質は、サイクル2スパーサーC<sub>2</sub>に相補的な3'スパーサー配列（C'<sub>2</sub>）を有する。結合サイクル1中、第1のNTAA結合性物質は、巨大分子の遊離N末端に結合し、第1のコーディングタグの情報が、相補的C'<sub>1</sub>スパーサー配列にハイブリダイズされたC<sub>1</sub>配列から、プライマー伸長により、同種記録タグに移行される。NTAAを除去して、新しいNTAAを露出させた後、結合サイクル2は、サイクル1結合性物質の3'スパーサー配列と同一であるサイクル25'スパーサー配列（C'<sub>2</sub>）および共通のサイクル33'スパーサー配列（C'<sub>3</sub>）を有する複数のNTAA結合性物質を、巨大分子と接触させる。第2のNTAA結合性物質は、巨大分子のNTAAと結合し、第2のコーディングタグの情報が、プライマー伸長により、相補的なC<sub>2</sub>およびC'<sub>2</sub>スパーサー配列から同種記録タグへと移行される。これらのサイクルを、最大「n」結合サイクル繰り返し、最後の伸長記録タグを、ユニバーサルリバースプライミング配列でキャッピングし、単一の巨大分子と共局在化されている複数の伸長記録タグを生成し、各伸長記録タグは、1結合サイクルに由来するコーディングタグ情報を有する。各連続結合サイクルで使用される各セットの結合性物質は、コーディングタグにサイクル特異的スパーサー配列を有するため、結合サイクル情報を、得られる伸長記録タグの結合性物質情報と関連付けることができる。図10Bは、結合の各連続サイクルに使用されるサイクル特異的結合性物質の様々なプールを示し、各プールは、サイクル特異的スパーサー配列を有する。図10Cは、サイクル特異的スパーサー配列を使用し、伸長記録タグのPCRアセンブリに基づいて、巨大分子の部位に共局在化されている伸長記録タグのコレクションを逐次的に組み立て、それにより巨大分子の順序付けられた配列を提供することができる方法を示す。好ましいモードでは、鎖状体形成前に、増幅により各伸長記録タグの複数のコピーを生成する。

20

30

40

#### 【0202】

【図11】図11A～11Bは、記録タグからコーディングタグまたはジタグ構築物への情報移行を示す。結合情報を記録するための2つの方法（A）および（B）が示されてい

50

る。結合性物質は、本明細書に記載の任意のタイプの結合性物質であってもよく、抗ホスホチロシン結合性物質が示されているが、それは例示のために過ぎない。伸長コーディングタグまたはジタグ構築の場合、結合情報をコーディングタグから記録タグへと移行させるのではなく、情報を、記録タグからコーディングタグへと移行させて、伸長コーディングタグを生成するか（A）、または情報を、記録タグおよびコーディングタグの両方から第3のジタグ形成構築物へと移行させるか（B）のいずれかである。ジタグおよび伸長コーディングタグは、記録タグ（バーコード、任意選択のUMI配列、および任意選択のコンパートメントタグ（CT）配列（図示せず））、およびコーディングタグの情報を含む。ジタグおよび伸長コーディングタグを、記録タグから溶出させ、収集し、任意選択で増幅し、次世代シーケンサーで読み取ることができる。

10

#### 【0203】

【図12】図12A～12Dは、PNAコンビナトリアルバーコード/UMI記録タグの設計、および結合事象のジタグ検出を示す。図12Aには、化学的ライゲーションによる、4つの基本PNAワード配列（A、A'-B、B'-C、およびC'）のコンビナトリアルPNAバーコード/UMIの構築が示されている。DNAアームのハイブリダイズは、PNAバーコード/UMIのコンビナトリアルアセンブリの無スペースコンビナトリアル鋳型を創出するために含まれている。化学的ライゲーションは、アニーリングされたPNA「ワード」を縫い合わせるために使用される。図12Bは、記録タグのPNA情報をDNA中間体へと移行させるための方法を示す。DNA中間体は、情報をコーディングタグへと移行させることが可能である。すなわち、相補的DNAワード配列を、PNAにアニーリングさせ、化学的にライゲーションする（任意選択で、PNA鋳型を使用するリガーゼが発見された場合は、酵素的にライゲーションする）。図12Cでは、DNA中間体は、スペース配列Spを介して、コーディングタグと相互作用するように設計されている。鎖置換プライマー伸長ステップは、ライゲーションされたDNAを置換し、記録タグ情報をDNA中間体からコーディングタグへと移行させ、伸長コーディングタグを生成する。ターミネーターヌクレオチドを、DNA中間体の末端に組み込んで、プライマー伸長によるコーディングタグ情報のDNA中間体への移行を防止してもよい。図12D：あるいは、情報を、コーディングタグからDNA中間体へと移行させて、ジタグ構築物を生成してもよい。ターミネーターヌクレオチドを、コーディングタグの末端に組み込んで、DNA中間体からコーディングタグへの記録タグ情報の移行を防止してもよい。

20

30

#### 【0204】

【図13】図13A～13Eは、コンパートメントバーコード付きビーズへのプロテオーム分配、その後のエマルジョン融合PCRでのジタグアセンブリによる、ペプチド配列組成を表わすエレメントのライブラリーの生成を示す。その後、ペプチドのアミノ酸の内容は、N末端配列決定法により、またはその代わりに、アミノ酸特異的化学的標識もしくはコーディングタグが付随されている結合性物質を付着（共有結合または非共有結合）させることにより特徴付けることができる。コーディングタグは、ユニバーサルプライミング配列、ならびにアミノ酸を識別するためのエンコーダー配列、コンパートメントタグ、およびアミノ酸UMIで構成されている。情報移行の後、ジタグを、記録タグUMIにより元の分子にマッピングし戻す。図13Aでは、プロテオームを、バーコード付きビーズを有する液滴にコンパートメント化する。付随する記録タグ（コンパートメントバーコード情報を含む）を有するペプチドを、ビーズ表面に付着させる。液滴エマルジョンを破壊して、ペプチドが分配されているバーコード付きビーズを放出させる。図13Bでは、ペプチドの特定のアミノ酸残基を、部位特異的標識部分にコンジュゲートされているDNAコーディングタグで化学的に標識する。DNAコーディングタグは、アミノ酸バーコード情報および任意選択でアミノ酸UMIを含む。図13C：標識ペプチド-記録タグ複合体を、ビーズから放出させる。図13D：標識ペプチド-記録タグ複合体を、1コンパートメント当たり平均で1つ未満のペプチド-記録タグ複合体が存在するように、ナノエマルジョンまたはマイクロエマルジョンへと乳化する。図13E：エマルジョン融合PCRは、記録タグ情報（例えば、コンパートメントバーコード）を、アミノ酸残基に付着している

40

50

DNAコーディングタグのすべてに移行させる。

#### 【0205】

【図14】図14は、乳化されたペプチド記録タグ-コーディングタグ複合体からの伸長コーディングタグの生成を示す。図13Cのペプチド複合体を、1液滴当たり平均で単一のペプチド複合体となるように、PCR試薬と共に液滴内へと共乳化する。3プライマー融合PCR(three-primer fusion PCR)手法を使用して、ペプチドに付随されている記録タグを増幅し、増幅した記録タグを、複数の結合性物質コーディングタグまたは共有結合で標識されたアミノ酸のコーディングタグと融合し、プライマー伸長によりコーディングタグを伸長させて、ペプチドUMIおよびコンパートメントタグ情報を、記録タグからコーディングタグへと移行させ、得られた伸長コーディングタグを増幅する。1液滴当たり複数の伸長コーディングタグ種が存在し、存在する各アミノ酸エンコーダー配列-UMIコーディングタグ毎に種が異なる。このようにして、ペプチド内のアミノ酸の同一性および計数を両方とも決定することができる。U1ユニバーサルプライマーおよびSpプライマーは、U2<sub>tr</sub>ユニバーサルプライマーよりも高い融解T<sub>m</sub>を有するように設計される。これにより、2段階PCRが可能になり、2段階PCRでは、最初の少数回のサイクルをより高いアニーリング温度で実施して記録タグを増幅し、その後、PCR中に記録タグおよびコーディングタグが互いにプライミングして伸長コーディングタグが産生されるように、より低いT<sub>m</sub>の段階へと進み、U1およびU2<sub>tr</sub>ユニバーサルプライマーを使用して、得られた伸長コーディングタグ産物の増幅にプライミングする。ある特定の実施形態では、U2<sub>tr</sub>プライマーからの時期尚早ポリメラーゼ伸長は、光解離性3'ブロック基を使用することにより防止することができる(Youngら、2008年、Chem. Commun. (Camb) 4巻: 462~464頁)。記録タグを増幅する第1のラウンドのPCR、およびコーディングタグS<sub>ptr</sub>が、記録タグの増幅されたS<sub>p</sub>'配列のコーディングタグの伸長にプライミングする第2のラウンドの融合PCRステップの後、U2<sub>tr</sub>の3'ブロック基を除去し、U1およびU2<sub>tr</sub>プライマーを有する伸長コーディングタグを増幅するために、より高い温度でのPCRを開始させる。

#### 【0206】

【図15】図15は、タンパク質のマッピング性およびフェージングの増強を容易にするプロテオーム分配およびバーコード付与の使用を示す。ペプチド配列決定では、典型的には、タンパク質を消化してペプチドにする。このプロセスでは、親タンパク質分子に由来する個々のペプチド間の関係性に関する情報、および親タンパク質分子に対するそれらの関係性が失われる。この情報を再構築するために、個々のペプチド配列を、それらが由来する可能性のあるタンパク質配列のコレクションにマッピングし戻す。そのようなセット内に一意の一致を見出すタスクは、短鎖および/または部分ペプチド配列では、コレクションのサイズおよび複雑性(例えば、プロテオーム配列複雑性)が増加すると共に困難となる。バーコード付き(例えば、コンパートメントタグ付き)コンパートメントまたは区分へとプロテオームを分配し、その後タンパク質をペプチドへと消化し、コンパートメントタグをペプチドに接合することにより、ペプチド配列をマッピングする必要のある「タンパク質」空間が低減され、複雑なタンパク質試料の場合のタスクが大幅に単純化される。ペプチドへと消化する前に、一意の分子識別子(UMI)を有するタンパク質を標識することにより、ペプチドを元のタンパク質分子にマッピングし戻すことが容易になり、同じタンパク質分子に由来する翻訳後修飾(PTM)変異体間のフェージング情報の注釈、および個々のプロテオフォームの識別が可能になる。図15Aは、分配バーコードを含む記録タグでタンパク質を標識し、その後記録タグ標識ペプチドへと断片化することを含む、プロテオーム分配の例を示す。図15B:部分ペプチド配列情報または組成情報のみの場合でさえ、このマッピングは、高度に縮重性である。しかしながら、同じタンパク質に由来する複数のペプチドからの情報と結び付けられた部分ペプチド配列または組成情報は、元のタンパク質分子の一意の識別を可能にする。

#### 【0207】

【図 16】図 16 は、コンパートメントタグ付きビーズ配列設計の例示的なモードを示す。コンパートメントタグは、個々のコンパートメントを識別するための X<sub>5</sub> ~ 20 のバーコード、およびコンパートメントタグが接合されているペプチドを識別するための N<sub>5</sub> ~ 10 の一意の分子識別子 (UMI) を含み、X および N は、縮重した核酸塩基または核酸塩基ワードを表わす。コンパートメントタグは、一本鎖であってもよく (上段に図示)、または二本鎖であってもよい (下段に図示)。任意選択で、コンパートメントタグは、目的のペプチドと接合するためのタンパク質リガーゼ (例えば、ブテラーゼ I (butelase I)) の認識配列を有するペプチド配列を含むキメラ分子であってもよい (左側に図示)。あるいは、目的のペプチドとのカップリングのために、化学的部分が、コンパートメントタグに含まれていてもよい (例えば、右側の図に示されるようなアジド)。

10

【0208】

【図 17】図 17 A ~ 17 B は、(A) 複数のペプチドを表す複数の伸長記録タグおよび (B) 標準的ハイブリッド捕捉技法による標的ペプチド濃縮の例示的な方法を示す。例えば、ハイブリッド捕捉濃縮では、ペプチドのライブラリーを表わす伸長記録タグのライブラリーから、1 つまたは複数の目的のペプチド (「標的ペプチド」) を表わす伸長記録タグとハイブリダイズする 1 つまたは複数のビオチン化「ベイト」オリゴヌクレオチドを使用してもよい。ベイトオリゴヌクレオチド：標的とする伸長記録タグハイブリダイゼーション対を、ハイブリダイゼーション後にビオチンタグにより溶液からプルダウンして、1 つまたは複数の目的のペプチドを表わす伸長記録タグの濃縮画分を生成する。伸長記録タグの分離 (「プルダウン」) は、例えば、ストレプトアビジンコーティング磁気ビーズを使用して達成することができる。ビオチン部分を、ビーズのストレプトアビジンと結合させ、磁石を使用してビーズを局在化させ、溶液を除去または交換することにより分離を達成する。望ましくないかまたは過剰に豊富なペプチドを表わす伸長記録タグに競合的にハイブリダイズする非ビオチン化競合物質濃縮オリゴヌクレオチドを、任意選択で、ハイブリッド捕捉アッセイのハイブリダイゼーションステップに含めて、濃縮された標的ペプチドの量をモジュレートしてもよい。非ビオチン化競合オリゴヌクレオチドは、標的ペプチドとのハイブリダイゼーションを競合するが、ハイブリダイゼーション二本鎖は、ビオチン部分が存在しないため捕捉ステップ中に捕捉されない。したがって、競合オリゴヌクレオチドのビオチン化「ベイト」オリゴヌクレオチドに対する比を調整することにより、濃縮された伸長記録タグ画分を、広いダイナミックレンジにわたってモジュレートすることができる。このステップは、試料内のタンパク質存在量のダイナミックレンジ問題に対処するために重要になるであろう。

20

30

【0209】

【図 18】図 18 A ~ 18 B は、単一細胞およびバルクプロテオームを個々の液滴内に分配するための例示的な方法を示し、各液滴は、ペプチドをそれらの元のタンパク質複合体と、または単一の細胞に由来するタンパク質と相関させるために、複数のコンパートメントタグが付着しているビーズを含む。コンパートメントタグは、バーコードを含む。液滴形成後の液滴構成成分の操作：(A) 単一細胞を個々の液滴内に分配し、その後細胞溶解して細胞プロテオームを放出させ、タンパク質分解により細胞プロテオームをペプチドへと消化し、十分なタンパク質分解後にプロテアーゼを不活化する；(B) バルクプロテオームを、複数の液滴内に分配し、個々の液滴は、タンパク質複合体を含み、その後タンパク質分解によりタンパク質複合体をペプチドへと消化し、十分なタンパク質分解後にプロテアーゼを不活化する。熱不安定性メタロ-プロテアーゼを使用し、光ケージ化 2 価カチオンを光放出させてプロテアーゼを活性化した後、封入されているタンパク質をペプチドへと消化することができる。プロテアーゼは、十分なタンパク質分解後に加熱不活化してもよく、または 2 価カチオンをキレートしてもよい。液滴は、ペプチドの N - または C - 末端アミノ酸のいずれかにライゲートすることが可能な核酸バーコード (記録タグとは別の) を含む、ハイブリダイズされたまたは放出可能なコンパートメントタグを含む。

40

【0210】

【図 19】図 19 A ~ 19 B は、単一細胞およびバルクプロテオームを個々の液滴内に分

50



配するための例示的な方法を示し、各液滴は、ペプチドをそれらの元のタンパク質またはタンパク質複合体と、またはタンパク質を元の単一細胞と関連させるために、コンパートメントタグが付着している複数の二機能性記録タグを有するビーズを含む。液滴形成後の液滴構成成分の操作：(A) 単一細胞を個々の液滴内に分配し、その後細胞溶解して細胞プロテオームを放出させ、タンパク質分解により細胞プロテオームをペプチドへと消化し、十分なタンパク質分解後にプロテアーゼを不活化する；(B) バルクプロテオームを、複数の液滴内に分配し、個々の液滴は、タンパク質複合体を含み、その後タンパク質分解によりタンパク質複合体をペプチドへと消化し、十分なタンパク質分解後にプロテアーゼを不活化する。熱不安定性メタロ-プロテアーゼを使用し、光ケージ化2価カチオン（例えば、 $Zn^{2+}$ ）を光放出させた後、封入されているタンパク質をペプチドへと消化することができる。プロテアーゼは、十分なタンパク質分解後に加熱不活化してもよく、または2価カチオンをキレートしてもよい。液滴は、ペプチドのN-またはC-末端アミノ酸のいずれかにライゲートすることが可能な核酸バーコード（記録タグとは別の）を含む、ハイブリダイズされたまたは放出可能なコンパートメントタグを含む。

【0211】

【図20-1】図20A~20Lは、ペプチドに付着したコンパートメントバーコード付き記録タグの生成を示す。コンパートメントバーコード付与技術（例えば、マイクロ流体液滴中のバーコード付きビーズなど）を使用して、コンパートメント特異的バーコードを、特定のコンパートメント内に封入されている分子内容物に移行させることができる。(A) 特定の実施形態では、タンパク質分子を変性させ、リシン残基(K)の-アミノ基を、活性化されたユニバーサルDNAタグ分子(5'末端にNHS部分を有することが示されているユニバーサルプライミング配列(U1)を含む)と化学的にコンジュゲートさせる。ユニバーサルDNAタグをポリペプチドにコンジュゲーションした後、過剰なユニバーサルDNAタグを除去する。(B) ユニバーサルDNAタグ付きポリペプチドを、ビーズに結合された核酸分子とハイブリダイズさせ、個々のビーズに結合された核酸分子は、コンパートメントタグ(バーコード)配列の一意の集団を含む。コンパートメント化は、液滴などの異なる物理的コンパートメント内(破線楕円により示されている)に試料を分離することにより生じ得る。あるいは、コンパートメント化は、例えば、ポリペプチドのユニバーサルDNAタグを、ビーズのコンパートメントDNAタグにアニーリングさせることによって、標識ポリペプチドをビーズ表面に固定化することにより、追加の物理的分離を必要とせずに、直接的に達成することができる。単一のポリペプチド分子は、単一のビーズとのみ相互作用する(例えば、単一のポリペプチドは、複数のビーズにまたがっていない)。しかしながら、複数のポリペプチドが、同じビーズと相互作用する場合がある。コンパートメントバーコード配列(BC)に加えて、ビーズに結合された核酸分子は、共通Sp(スペーサー)配列、一意の分子識別子(UMI)、およびポリペプチドDNAタグU1'に相補的な配列で構成されていてもよい。(C) ユニバーサルDNAタグ付きポリペプチドを、ビーズに結合されたコンパートメントタグとアニーリングさせた後、付着リンカーを切断することによりコンパートメントタグをビーズから放出する。(D) アニーリングしたU1 DNAタグプライマーを、ビーズに由来するコンパートメントタグ核酸分子を鋳型として使用する、ポリメラーゼに基づくプライマー伸長により伸長させる。プライマー伸長ステップは、(C)に示されているようにコンパートメントタグをビーズから放出させた後で、または任意選択でコンパートメントタグが依然としてビーズに付着している間(図示せず)に実施してもよい。これにより、ビーズのコンパートメントタグに由来するバーコード配列が、ポリペプチドのU1 DNAタグ配列に効果的に書き込まれる。この新しい配列が、記録タグを構成する。プライマー伸長後、プロテアーゼ、例えば、Lys-C(リシン残基のC末端側を切断する)、Glu-C(グルタミン酸残基のC末端側を、および程度は低いグルタミン酸残基を切断する)、またはプロテイナーゼKなどのランダムプロテアーゼを使用して、ポリペプチド断片をペプチド断片へと切断する。(E) 各ペプチド断片は、本明細書で開示されているように下流でペプチド配列決定するために、そのC末端リシンを、記録タグを構成する伸長DNAタグ配列で標識する。(F)

10

20

30

40

50

記録タグ付きペプチドを、歪みアルキン標識 D B C O を介してアジドビーズにカップリングする。また、アジドビーズは、任意選択で、D B C O - アジド固定化の効率を促進するために、記録タグに相補的な捕捉配列を含む。なお、元のビーズからペプチドを除去し、新しい固体支持体（例えば、ビーズ）に再固定化することにより、ペプチド間の最適な分子間離間が可能になり、本明細書で開示されているペプチド配列決定法が容易になることが留意されるべきである。図 2 0 G ~ 2 0 L は、アルキンで予め標識したポリペプチド（図 2 B に記載のような）への D N A タグのクリック化学的コンジュゲーションを使用すること以外は、図 2 0 A ~ 2 0 F に示されているものと同様の概念を示す。アジドおよび m T e t 化学は、直交性であり、D N A タグへのクリックコンジュゲーション、および配列決定基材へのクリック i E D D A コンジュゲーション（m T e t および T C O）を可能にする。

10

【図 2 0 - 2】同上。

【図 2 0 - 3】同上。

【図 2 0 - 4】同上。

【0 2 1 2】

【図 2 1】図 2 1 は、流動フォーカス T 字路を使用して、単一細胞を、コンパートメントタグ付き（例えば、バーコード）ビーズにコンパートメント化するための例示的な方法を示す。2 つの水流を用いると、液滴形成時に、細胞溶解およびプロテアーゼ活性化（Z n <sup>2+</sup> 混合）を容易に開始することができる。

【0 2 1 3】

20

【図 2 2】図 2 2 A ~ 2 2 B は、例示的なタグ化詳細を示す。（A）コンパートメントタグ（D N A - ペプチドキメラ）を、プテラーゼ I によるペプチドライゲーションを使用して、ペプチドに付着させる。（B）ペプチド配列決定の開始前に、コンパートメントタグ情報を、付随する記録タグへと移行させる。任意選択で、アスパラギン酸残基の N 末端でペプチド結合を選択的に切断するエンドペプチダーゼ A s p N を使用して、記録タグへの情報移行後、コンパートメントタグを切断することができる。

【0 2 1 4】

【図 2 3】図 2 3 A ~ 2 3 C は、組織切片の空間的プロテオミクスに基づく解析のためのアレイに基づくバーコードを示す。（A）空間的にコードされた D N A バーコードのアレイ（B C i j と表記されているバーコードを特徴とする）を、組織切片（F F P E または凍結）と組み合わせる。一実施形態では、組織切片は、固定および透過処理されている。好ましい実施形態では、アレイ特徴サイズは、細胞サイズ（ヒト細胞の場合、約 1 0 μ m）よりも小さい。（B）アレイにマウントした組織切片を試薬で処理して架橋を元に戻す（例えば、シトラコン酸無水物を用いた抗原回復プロトコール（N a m i m a t s u , G h a z i z a d e h ら、2 0 0 5 年）、その後その中のタンパク質を、タンパク質分子をすべて D N A 記録タグで効果的に標識する部位反応性 D N A 標識で標識する（例えば、抗原回復後に遊離されたりシンの標識）。標識および洗浄後、アレイに結合した D N A バーコード配列を切断し、マウントした組織切片への拡散を可能にし、その中にあるタンパク質に付着した D N A 記録タグとハイブリダイズさせる。（C）ここで、アレイにマウントした組織を、ポリメラーゼ伸長に供して、ハイブリダイズされたバーコードの情報を、タンパク質を標識している D N A 記録タグへと移行させる。バーコード情報の移行後、アレイにマウントした組織を、スライドからこすり落とし、任意選択でプロテアーゼにより消化し、タンパク質またはペプチドを溶液内に抽出する。

30

【0 2 1 5】

【図 2 4 - 1】図 2 4 A ~ 2 4 B は、ビーズに固定化されており、コーディングタグに付着した結合性物質によりアッセイされる 2 つの異なる例示的な D N A 標的巨分子（A B および C D）を示す。このモデル系は、結合された物質から近位記録タグへのコーディングタグ移行の単一分子挙動を例示する役目を果たす。好ましい実施形態では、コーディングタグは、プライマー伸長により伸長記録コーディングタグに組み込まれる。図 2 4 A は、A B 巨分子が、A 特異的結合性物質（「A'」、A B 巨分子の「A」成分に相補的オ

40

50

リゴヌクレオチド配列)と相互作用して、付随するコーディングタグの情報がプライマー伸長により記録タグへと移行され、B特異的結合性物質(「B'」、AB巨大分子の「B」成分に相補的なオリゴヌクレオチド配列)と相互作用して、付随するコーディングタグの情報がプライマー伸長により記録タグへと移行されることを示す。コーディングタグAおよびBは配列が異なり、この図では、容易に識別することができるように長さも異なっている。長さが異なることにより、ゲル電気泳動法によるコーディングタグ移行の解析が容易になるが、次世代シーケンシングによる解析では長さが異なる必要はない。A'およびB'結合性物質の結合は、単一結合サイクルの代替的な可能性として示されている。第2のサイクルが追加されれば、伸長記録タグはさらに伸長されることになる。第1および第2のサイクルにてA'またはB'結合性物質のいずれが添加されるかに応じて、伸長記録タグは、形態AA、AB、BA、およびBBのコーディングタグ情報を含むことができる。したがって、伸長記録タグは、結合物質の同一性だけでなく、結合事象の順序に関する情報も含む。同様に、図24Bは、CD巨大分子が、C特異的結合性物質(「C'」、CD巨大分子の「C」成分に相補的なオリゴヌクレオチド配列)と相互作用して、付随するコーディングタグの情報がプライマー伸長により記録タグへと移行され、D特異的結合性物質(「D'」、CD巨大分子の「D」成分に相補的なオリゴヌクレオチド配列)と相互作用して、付随するコーディングタグの情報がプライマー伸長により記録タグへと移行されることを示す。コーディングタグCおよびDは配列が異なり、この図では、容易に識別することができるように長さも異なっている。長さが異なることにより、ゲル電気泳動法によるコーディングタグ移行の解析が容易になるが、次世代シーケンシングによる解析には長さが異なる必要はない。C'およびD'結合性物質の結合は、単一結合サイクルの代替的な可能性として示されている。第2のサイクルが追加されれば、伸長記録タグはさらに伸長されることになる。第1および第2のサイクルにてC'またはD'結合性物質のいずれが添加されるかに応じて、伸長記録タグは、形態CC、CD、DC、およびDDのコーディングタグ情報を含むことができる。コーディングタグは、任意選択でUMIを含んでいてもよい。コーディングタグにUMIが含まれることにより、結合事象に関する追加情報を記録することが可能になり、それにより、結合事象を個々の結合性物質のレベルで区別することが可能になる。これは、個々の結合性物質が、1つよりも多くの結合事象に参加することができる場合(例えば、その結合親和性が、1つよりも多くの事象に参加するのに十分な頻度で解離および再結合することができるような程度である場合)、有用であり得る。また、エラー訂正に有用であり得る。例えば、いくつかの状況下では、コーディングタグは、同じ結合サイクルで2回またはそれよりも多くの頻度で、情報を記録タグに移行させる場合がある。UMIを使用すれば、これらが、すべて単一結合事象に関連する繰り返し情報移行事象である可能性が高いことが明らかになるであろう。

【図24-2】同上。

【0216】

【図25】図25は、ビーズに固定化されており、コーディングタグに付着した結合性物質によりアッセイされる例示的なDNA標的的巨大分子(AB)を示す。A特異的結合性物質(「A'」、AB巨大分子のA成分に相補的なオリゴヌクレオチド)は、AB巨大分子と相互作用し、付随するコーディングタグの情報が、ライゲーションにより記録タグへと移行される。B特異的結合性物質(「B'」、AB巨大分子のB成分に相補的なオリゴヌクレオチド)は、AB巨大分子と相互作用し、付随するコーディングタグの情報が、ライゲーションにより記録タグへと移行される。コーディングタグAおよびBは配列が異なり、この図では、容易に識別することができるように長さも異なっている。長さが異なることにより、ゲル電気泳動法によるコーディングタグ移行の解析が容易になるが、次世代シーケンシングによる解析には長さが異なる必要はない。

【0217】

【図26】図26A~26Bは、プライマー伸長による結合/コーディングタグ移行の例示的なDNA-ペプチド巨大分子を示す。図26Aは、ビーズに固定化されている例示的なオリゴヌクレオチド-ペプチド標的的巨大分子(「A」オリゴヌクレオチド-cMy cペ

プチド)を示す。cMy c - 特異的結合性物質(例えば、抗体)は、巨大分子のcMy c ペプチド部分と相互作用し、付随するコーディングタグの情報が記録タグに移行される。cMy c コーディングタグの情報の記録タグへの移行は、ゲル電気泳動法により解析することができる。図26Bは、ビーズに固定化されている例示的なオリゴヌクレオチド - ペプチド標的的巨大分子(「C」オリゴヌクレオチド - 赤血球凝集素(HA)ペプチド)を示す。HA - 特異的結合性物質(例えば、抗体)は、巨大分子のHAペプチド部分と相互作用し、付随するコーディングタグの情報が記録タグに移行される。コーディングタグの情報の記録タグへの移行は、ゲル電気泳動法により解析することができる。cMy c 抗体 - コーディングタグおよびHA抗体 - コーディングタグの結合は、単一結合サイクルの代替的な可能性として示されている。第2のサイクルが実施されれば、伸長記録タグはさらに伸長されることになる。第1および第2の結合サイクルにてcMy c 抗体 - コーディングタグまたはHA抗体 - コーディングタグのいずれを添加するかに応じて、伸長記録タグは、形態cMy c - HA、HA - cMy c、cMy c - cMy c、およびHA - HAのコーディングタグ情報を含むことができる。また、図示されていないが、追加の結合性物質を導入して、巨大分子のAおよびCオリゴヌクレオチド成分の検出を可能にすることができる。したがって、異なるタイプの骨格を含むハイブリッド巨大分子は、情報を記録タグへと移行させ、結合事象の順序ならびに結合性物質の同一性に関する情報を含む伸長記録タグを読み出すことにより解析することができる。

【0218】

【図27-1】図27A~27Dは、エラー訂正バーコードの生成を示す。(A)65個のエラー訂正バーコード(配列番号1~65)のサブセットを、Rソフトウェアパッケージ「DNABarcodes」(<https://bioconductor.riken.jp/packages/3.3/bioc/manuals/DNABarcodes/man/DNABarcodes.pdf>)から、コマンドパラメーター[create.dnabarcodes(n=15, dist=10)]を使用して導出した77個のバーコードのセットから選択した。このアルゴリズムは、4個の置換の距離まで置換エラーを訂正することができ、9個の置換までエラーを検出することができる15mer「Hamming」バーコードを生成する。様々なナノポア電流レベル(ナノポアに基づく配列決定の場合)を示さなかったか、またはこのセットの他のメンバーとも相関していたバーコードを濾過することにより、65個のバーコードのサブセットを生成した。(B)細孔を通り抜ける15merバーコードの予測ナノポア電流レベルのプロット。予測電流は、各15merバーコードワードを、11個のオーバーラップ5merワードの複合セットに分割し、5mer R9ナノポア電流レベル参照テーブル(template\_median68pA, 5mers.model (<https://github.com/jts/nanopolish/tree/master/etc/r9-models>))を使用して、バーコードがナノポアを一度に一塩基ずつ通り抜ける際の対応する電流レベルを予測することにより算出した。(B)から理解することができるように、65個のバーコードのこのセットは、そのメンバーの各々に一意の電流シグネチャを示す。(C)DTRおよびDTRプライマーのオーバーラップセットを使用した、ナノポアシーケンシング用のモデル伸長記録タグとしてのPCR産物の生成が示されている。その後、PCRアンプリコンをライゲーションして、鎖状の伸長記録タグモデルを形成する。(D)図27Cに示されているように生成された、例示的な「伸長記録タグ」モデルのナノポアシーケンシングリード(リード長734塩基)。MinIon R9.4リードは、7.2の品質スコアを有する(リード品質は不良)。しかしながら、バーコード配列は、リード品質が低い場合でさえ(Qscore=7.2)、lalignを使用して容易に識別することができる。15merスペーサーエレメントには下線が引かれている。バーコードは、BCまたはBC'記号と表記されているフォワード方向またはリバース方向のいずれでもアラインすることができる。

【図27-2】同上。

【図27-3】同上。

10

20

30

40

50

## 【 0 2 1 9 】

【図 2 8】図 2 8 A ~ 2 8 D は、記録タグによるタンパク質の分析物特異的標識を示す。

( A ) その天然コンフォメーションをとっている目的のタンパク質分析物を標的とする結合性物質は、DNA 記録タグの相補的分析物特異的バーコード ( B C A ) にハイブリダイズする分析物特異的バーコード ( B C A ' ) を含む。あるいは、切断可能なリンカーを介して DNA 記録タグを結合性物質に付着させることができ、DNA 記録タグを、タンパク質に直接「クリック」させ、その後結合性物質から切断する ( 切断可能なリンカーにより ) 。 DNA 記録タグは、反応性カップリング部分 ( 目的のタンパク質にカップリングするためのクリック化学試薬 ( 例えば、アジド、m T e t など ) 、および他の機能性成分 ( 例えば、ユニバーサルプライミング配列 ( P 1 ) 、試料バーコード ( B C S ) 、分析物特異的バーコード ( B C A ) 、およびスペーサー配列 ( S p ) など ) を含む。また、試料バーコード ( B C S ) を使用して、異なる試料に由来するタンパク質を標識および区別することができる。また、DNA 記録タグは、その後の基材表面へのカップリングのための直交性カップリング部分 ( 例えば、m T e t ) を含んでいてもよい。目的のタンパク質に記録タグをクリック化学カップリングする場合、タンパク質を、DNA 記録タグのクリック化学カップリング部分と同種のクリック化学カップリング部分で予め標識する ( 例えば、タンパク質のアルキン部分は、DNA 記録タグのアジド部分と同種である ) 。クリック化学カップリング用のカップリング部分で DNA 記録タグを標識するための試薬の例としては、リシン標識用のアルキン - N H S 試薬、光親和性標識用のアルキン - ベンゾフェノン試薬などが挙げられる。 ( B ) 結合性物質が近位標的タンパク質と結合した後、記録タグの反応性カップリング部分 ( 例えば、アジド ) は、近位タンパク質の同種クリック化学カップリング部分 ( 3 本線記号として示されている ) に共有結合で付着する。 ( C ) 標的タンパク質分析物を記録タグで標識した後、ウラシル特異的切除試薬 ( 例えば、U S E R ( 商標 ) ) を使用してウラシル ( U ) を消化することにより、付着している結合性物質を除去する。 ( D ) DNA 記録タグで標識した標的タンパク質分析物を、クリック化学 ( アルキン - アジド結合対、メチルテトラジン ( m T E T ) - t r a n s - シクロオクテン ( T C O ) 結合対など ) などの好適なバイオコンジュゲート化学反応を使用して基材表面に固定化する。ある特定の実施形態では、標的タンパク質 - 記録タグ標識アッセイ全体を、結合性物質のプールおよび記録タグのプールを使用して、多数の異なる標的タンパク質分析物を含む単一チューブ中で実施する。試料バーコード ( B C S ) を含む記録タグで試料内のタンパク質分析物を標的標識した後、複数のタンパク質分析物試料を、( D ) の固定化ステップ前にプールしてもよい。したがって、ある特定の実施形態では、数百個の試料にわたって最大数千個のタンパク質分析物を、単一チューブ次世代タンパク質アッセイ ( N G P A ) で標識および固定化することができ、高価な親和性試薬 ( 例えば、抗体 ) が大幅に節約される。

## 【 0 2 2 0 】

【図 2 9】図 2 9 A ~ 2 9 E は、DNA 記録タグとポリペプチドとのコンジュゲーションを示す。 ( A ) 変性ポリペプチドを、アルキン - N H S エステル ( アセチレン - P E G - N H S エステル ) 試薬またはアルキン - ベンゾフェノンなどの二機能性クリック化学試薬で標識して、アルキン標識 ( 3 本線記号 ) ポリペプチドを生成する。また、アルキンは、ジベンゾシクロオクチル ( D B C O ) などを含むシクロオクチンなどの歪みアルキンであってもよい。 ( B ) アルキン標識ポリペプチドに化学的にカップリングされる DNA 記録タグ設計の例が示されている。記録タグは、ユニバーサルプライミング配列 ( P 1 ) 、バーコード ( B C ) 、およびスペーサー配列 ( S p ) を含む。記録タグを、基材表面にカップリングするための m T e t 部分、および標識ポリペプチドのアルキン部分とカップリングするためのアジド部分で標識する。 ( C ) 変性アルキン標識タンパク質またはポリペプチドを、アルキンおよびアジド部分を介して記録タグで標識する。任意選択で、記録タグ標識ポリペプチドを、コンパートメントバーコードで、例えば、コンパートメントピースに付着した相補的配列とアニーリングさせ、プライマー伸長 ( ポリメラーゼ伸長とも呼ばれる ) させることにより、または図 2 0 H ~ 2 0 J に示されているように、さらに標識す

10

20

30

40

50

ることができる。(D)記録タグ標識ポリペプチドをプロテアーゼ消化することにより、記録タグ標識ペプチドの集団が創出される。一部の実施形態では、一部のペプチドは、いかなる記録タグにも標識されていないであろう。他の実施形態では、一部のペプチドには、1つまたは複数の記録タグが付着していてもよい。(E)記録タグ標識ペプチドを、T C O基で機能化された基材表面とペプチドに付着した記録タグのm T e t部分との間の逆電子要請型ディールス - アルダー ( i E D D A ) クリック化学反応を使用して、基材表面に固定化する。ある特定の実施形態では、図示されている異なる段階間で、クリーンアップステップを用いてもよい。直交性クリック化学 (例えば、アジド - アルキンおよびm T e t - T C O) を使用することにより、記録タグによるポリペプチドのクリック化学標識、および記録タグ標識ペプチドの基材表面へのクリック化学固定化が両方とも可能になる (その全体が参照により組み込まれる、M c K a y ら、2014年、C h e m . B i o l .、21巻: 1075 ~ 1101頁を参照されたい)。

10

#### 【0221】

【図30】図30A ~ 30Eは、ポリペプチドの初期DNA標識後の、試料バーコードの記録タグへの書き込みを示す。(A)変性ポリペプチドを、アルキン - N H S 試薬またはアルキン - ベンゾフェノンなどの二機能性クリック化学試薬で標識して、アルキン標識ポリペプチドを生成する。(B)ポリペプチドをアルキン (またはその代わりにクリック化学部分) で標識した後、ユニバーサルプライミング配列 (P 1) を含み、アジド部分およびm T e t 部分で標識されているDNAタグを、アジド - アルキン相互作用によりポリペプチドにカップリングする。他のクリック化学相互作用を用いてもよいことが理解される。(C)試料バーコード情報 (B C S ' ) および他の記録タグ機能性成分 (例えば、ユニバーサルプライミング配列 (P 1 ' )、スパーサー配列 (S p ' ) ) を含む記録タグDNA構築物は、相補的ユニバーサルプライミング配列を介してDNAタグ標識ポリペプチドとアニーリングする (P 1 - P 1 ' )。記録タグ情報は、ポリメラーゼ伸長によりDNAタグに移行される。(D)記録タグ標識ポリペプチドをプロテアーゼ消化することにより、記録タグ標識ペプチドの集団が創出される。(E)記録タグ標識ペプチドを、T C O基で機能化された表面とペプチドに付着した記録タグのm T e t 部分との間の逆電子要請型ディールス - アルダー ( i E D D A ) クリック化学反応を使用して、基材表面に固定化する。ある特定の実施形態では、図示されている異なる段階間で、クリーンアップステップを用いてもよい。直交性クリック化学 (例えば、アジド - アルキンおよびm T e t - T C O) を使用することにより、記録タグによるポリペプチドのクリック化学標識、および記録タグ標識ペプチドの基材表面へのクリック化学固定化が両方とも可能になる (その全体が参照により組み込まれる、M c K a y ら、2014年、C h e m . B i o l .、21巻: 1075 ~ 1101頁を参照されたい)。

20

30

#### 【0222】

【図31】図31A ~ 31Eは、ポリペプチドにバーコードを付与するためのビーズコンパートメント化を示す。(A)ポリペプチドを、ヘテロ二機能性クリック化学試薬による標準的バイオコンジュゲーションまたは光親和性標識技法を使用して、溶液中で標識する。考え得る標識部位としては、リシン残基の - アミン (例えば、図示されているようなN H S - アルキンと) またはペプチドの炭素骨格 (例えば、ベンゾフェノン - アルキンと) が挙げられる。(B)ユニバーサルプライミング配列 (P 1) を含むアジド標識DNAタグを、標識ポリペプチドのアルキン部分にカップリングする。(C)DNAタグ標識ポリペプチドを、DNA記録タグ標識ビーズに、相補的DNA配列 (P 1 およびP 1 ' ) を介してアニーリングさせる。ビーズのDNA記録タグは、スパーサー配列 (S p ' )、コンパートメントバーコード配列 (B C p ' )、任意選択の一意の分子識別子 (U M I)、およびユニバーサル配列 (P 1 ' ) を含む。DNA記録タグ情報は、ポリメラーゼ伸長により (あるいは、ライゲーションを用いてもよい)、ポリペプチドのDNAタグに移行される。情報移行後、得られたポリペプチドは、コンパートメントバーコードを含むいくつかの機能性エレメントを含有する複数の記録タグを含む。(D)記録タグ標識ポリペプチドをプロテアーゼ消化することにより、記録タグ標識ペプチドの集団が創出される。記録タグ標識

40

50

ペプチドをビーズから解離させ、(E)配列決定基材に再固定化する(例えば、図示されているようなmTet部分とTCO部分との間のiEDDAクリック化学を使用して)。

【0223】

【図32-1】図32A~32Hは、次世代タンパク質アッセイ(NGPA)のワークフローの例を示す。タンパク質試料を、いくつかの機能性単位、例えば、ユニバーサルプライミング配列(P1)、バーコード配列(BC)、任意選択のUMI配列、およびスパーサー配列(Sp)(結合性物質コーディングタグとの情報移行を可能にする)で構成されているDNA記録タグで標識する。(A)標識タンパク質を、基材(例えば、ビーズ、多孔性ビーズ、または多孔性マトリックス)に固定化する(受動的にまたは共有結合で)。(B)基材をタンパク質でブロッキングし、任意選択で、分析物記録タグ配列の非特異的相互作用を最小限に抑えるために、スパーサー配列に相補的な競合オリゴヌクレオチド(Sp')を添加する。(C)分析物特異的抗体(付随するコーディングタグを有する)を、基質に結合されたタンパク質と共にインキュベートする。コーディングタグは、その後のウラシル特異的切断のためのウラシル塩基を含んでいてもよい。(D)抗体結合後、該当する場合は、過剰な競合オリゴヌクレオチド(Sp')を洗い流す。コーディングタグは、相補的なスパーサー配列を介して記録タグと一時的にアニーリングし、コーディングタグ情報は、プライマー伸長反応で記録タグへと移行され、伸長記録タグが生成される。固定化されたタンパク質が変性されている場合、結合されている抗体およびアニーリングされているコーディングタグを、0.1N NaOHなどのアルカリ洗浄条件下で除去することができる。固定化されたタンパク質が、天然コンフォメーションをとっている場合、より穏やかな条件で、結合されている抗体およびコーディングタグを除去する必要がある場合がある。より穏やかな抗体除去条件の例は、パネルE~Hに概説されている。(E)コーディングタグから記録タグへの情報移行後、コーディングタグを、ウラシル特異的切除試薬(例えば、USER(商標))酵素ミックスを使用して、そのウラシル部位にニックを入れる(切断する)。(F)結合されている抗体を、高塩濃度および低/高pH洗浄を使用して、タンパク質から除去する。抗体に付着したままの切断型DNAコーディングタグは、短鎖であり、同様に迅速に溶出する。より長鎖のDNAコーディングタグ断片は、記録タグにアニーリングしたままであってもよく、またはアニーリングしたままでなくてもよい。(G)第2の結合サイクルが、ステップ(B)~(D)と同様に開始し、第2のプライマー伸長ステップでは、プライマー伸長により、コーディングタグ情報が第2の抗体から伸長記録タグへと移行される。(H)2つの結合サイクルの結果は、記録タグに付着した第1の抗体および第2の抗体からの結合情報の鎖状物である。

【図32-2】同上。

【0224】

【図33】図33A~33Dは、複数の結合性物質および酵素媒介性連続情報移行を使用した1段階次世代タンパク質アッセイ(NGPA)を示す。2つの同種結合性物質(例えば、抗体)が同時に結合されている固定化タンパク質分子を用いたNGPAアッセイ。複数の同種抗体結合事象の後、プライマー伸長およびDNAニックングの組合せステップを使用して、結合されている抗体のコーディングタグから記録タグへと情報を移行させる。コーディングタグのキャレット記号(^)は、二本鎖DNAニックングエンドヌクレアーゼ部位を表わす。(A)図示されている例では、タンパク質のエピトープ1(Epi#1)に結合されている抗体のコーディングタグは、相補的なスパーサー配列のハイブリダイゼーション後のプライマー伸長ステップにて、コーディングタグ情報(例えば、エンコーダー配列)を記録タグへと移行させる。(B)伸長記録タグとコーディングタグとの間に二本鎖DNAが形成されたら、37℃で活性であるNt.BsmAIなどの、二本鎖DNA基質の一方のDNAの鎖のみを切断するニックングエンドヌクレアーゼを使用して、コーディングタグを切断する。ニックングステップ後、切断型コーディングタグ結合性物質および伸長記録タグで形成された二本鎖は、熱力学的に不安定になり、解離する。より長鎖のコーディングタグ断片は、記録タグにアニーリングしたままであってもよく、またはアニーリングしたままでなくてもよい。(C)これにより、タンパク質のエピトープ#2(

E p i # 2 ) に結合されている抗体のコーディングタグが、相補的スペーサー配列を介して伸長記録タグにアニーリングし、プライマー伸長により E p i # 2 抗体のコーディングタグから伸長記録タグへと情報を移行させることにより、伸長記録タグをさらに伸長させることが可能になる。( D ) この場合も、伸長記録タグと E p i # 2 抗体のコーディングタグとの間に二本鎖 DNA が形成された後、N b . B s s S I などのニッキングエンドヌクレアーゼによりコーディングタグにニックを入れる。ある特定の実施形態では、プライマー伸長(ポリメラーゼ伸長とも呼ばれる)中は非鎖置換ポリメラーゼの使用が好ましい。非鎖置換ポリメラーゼは、単一塩基よりも多くが記録タグにアニーリングしたままである切断されたコーディングタグ残部の伸長を防止する。( A ) ~ ( D ) のプロセスは、近位の結合されている結合性物質のコーディングタグがすべて、ハイブリダイゼーション、伸長記録タグへの情報移行、ニッキングステップにより「消費」されるまで自発的に繰り返すことができる。コーディングタグは、所与の分析物(例えば、同種タンパク質)に特異的なすべての結合性物質(例えば、抗体)と同一であるエンコーダー配列を含んでいてもよく、エピトープ特異的なエンコーダー配列を含んでいてもよく、または異なる分子事象を区別するための一意の分子識別子( U M I )を含んでいてもよい。

10

#### 【 0 2 2 5 】

【図 3 4】図 3 4 A ~ 3 4 C は、基材表面の反応性部分のタイトレーションを使用した記録タグ - ペプチド固定化の密度制御を示す。( A ) 基材表面のペプチド密度は、基材の表面の機能性カップリング部分の密度を制御することによりタイトレーションすることができる。これは、活性カップリング分子対「ダミー」カップリング分子の適切な比で、基質の表面を誘導体化することにより達成することができる。図示されている例では、N H S - P E G - T C O 試薬(活性カップリング分子)を、N H S - m P E G (ダミー分子)を規定されている比で組み合わせて、アミン表面を T C O で誘導体化する。機能化 P E G は、3 0 0 から 4 0 , 0 0 0 を超える種々の分子量のものが入手可能である。( B ) 二機能性 5 ' アミン DNA 記録タグ( m T e t は他方の機能的部分である)を、スクシンイミジル 4 - ( N - マレイミドメチル)シクロヘキサン - 1 ( S M C C ) 二機能性架橋剤を使用して、ペプチドの N 末端 C y s 残基にカップリングする。記録タグの内部 m T e t - d T 基を、m テトラジン - アジドを使用して、アジド - d T 基から創出する。( C ) 記録タグ標識ペプチドを、m T e t と T C O との i E D D A クリック化学反応を使用して、( A ) の活性化基材表面に固定化する。m T e t - T C O i E D D A カップリング反応は、非常に迅速であり、効率的であり、安定的である( m T e t - T C O は、T e t - T C O よりも安定的である)。

20

30

#### 【 0 2 2 6 】

【図 3 5】図 3 5 A ~ 3 5 C は、次世代タンパク質シーケンシング( N G P S ) 結合サイクル特異的コーディングタグを示す。( A ) サイクル特異的 N 末端アミノ酸( N T A A ) 結合性物質コーディングタグを用いた N G P S アッセイの設計。N T A A 結合性物質(例えば、N 末端 D N P 標識チロシンに特異的な抗体)は、ユニバーサルプライミング配列( P 1 )、バーコード( B C )、およびスペーサー配列( S p )を含む記録タグに付随されているペプチドの D N P 標識 N T A A に結合する。結合性物質がペプチドの同種 N T A A に結合すると、N T A A 結合性物質に付随されているコーディングタグは、記録タグに接近し、相補的スペーサー配列を介して記録タグにアニーリングする。コーディングタグ情報は、ポリメラーゼ伸長により記録タグに移行される。コーディングタグがどの結合サイクルを表わすのかを記録するために、コーディングタグは、サイクル特異的バーコードを含んでいてもよい。ある特定の実施形態では、分析物に結合する結合性物質のコーディングタグは、一意の結合サイクル特異的バーコードと組み合わせられている、サイクル数に依存しない同じエンコーダーバーコードを有する。他の実施形態では、分析物に対する結合性物質のコーディングタグは、分析物 - 結合サイクルの組合せ情報についての一意のエンコーダーバーコードを含む。いずれの手法でも、共通スペーサー配列を、各結合サイクルでの結合性物質のコーディングタグに使用することができる。( B ) この例では、各結合サイクルの結合性物質は、結合サイクルを識別するための短い結合サイクル特異的バーコ

40

50



ードを有し、これにより、結合性物質を識別するエンコーダーバーコードと共に、特定の結合性物質結合サイクルの組合せを識別する一意の組合せバーコードが提供される。(C) 結合サイクルの完了後、伸長記録タグを、キャッピングサイクルステップを使用して増幅可能なライブラリーに変換することができ、キャッピングサイクルステップでは、例えば、ユニバーサルプライミング配列 P 2 およびスパーサー配列 S p ' に連結されたユニバーサルプライミング配列 P 1 ' を含むキャップが、まず相補的 P 1 および P 1 ' 配列を介して伸長記録タグとアニーリングして、キャップが伸長記録タグと接近する。伸長記録タグおよびキャップの相補的 S p および S p ' 配列がアニーリングし、プライマー伸長により、第 2 のユニバーサルプライマー配列 ( P 2 ) が伸長記録タグに付加される。

【 0 2 2 7 】

【図 3 6 - 1】図 3 6 A ~ 3 6 E は、コーディングタグから記録タグへの情報移行を実証するための DNA に基づくモデル系を示す。例示的な結合および分子内書き込みを、オリゴヌクレオチドモデル系により実証した。コーディングタグの標的指向性物質 A ' および B ' を、記録タグの標的結合領域 A および B とハイブリダイズするように設計した。2 つの記録タグ s a R T \_ A b c \_ v 2 ( A 標的 ) および s a R T \_ B b c \_ V 2 ( B 標的 ) を等濃度でプールすることにより、記録タグ ( R T ) ミックスを調製した。記録タグは、5 ' 末端がビオチン化されており、一意の標的結合領域、ユニバーサルフォワードプライマー配列、一意の DNA バーコード、および 8 塩基共通スパーサー配列 ( S p ) を含む。コーディングタグは、8 塩基共通スパーサー配列 ( S p ' ) に隣接されている一意のエンコーダーバーコード塩基を含み、8 塩基共通スパーサー配列の 1 つは、ポリエチレングリコールリンカーを介して A または B 標的物質と共有結合で連結されている。( A ) ビオチン化記録タグオリゴヌクレオチド ( s a R T \_ A b c \_ v 2 および s a R T \_ B b c \_ V 2 ) を、ビオチン化ダミー T 1 0 オリゴヌクレオチドと共に、ストレプトアビジンビーズに固定化した。A または B 捕捉配列 (それぞれ同種結合性物質 A ' および B ' により認識される) および結合標的を識別するための対応するバーコード ( r t A \_ B C および r t B \_ B C ) を有する記録タグを設計した。このモデル系のバーコードはすべて、6 5 個の 1 5 m e r バーコード (配列番号 1 ~ 6 5 ) のセットから選択した。一部の場合では、ゲル解析を容易にするために、1 5 m e r バーコードを組み合わせて、より長いバーコードを構成した。特に、r t A \_ B C = B C \_ 1 + B C \_ 2 ; r t B \_ B C = B C \_ 3。また、記録タグの A および B 配列と同種の結合性物質の 2 つのコーディングタグ、つまり C T \_ A ' - b c (エンコーダーバーコード = B C \_ 5 ) および C T \_ B ' - b c (エンコーダーバーコード = B C \_ 5 + B C \_ 6 ) を合成した。任意選択で、コーディングタグを、ビーズに固定化された記録タグにアニーリングさせる前に、コーディングタグ配列の部分に相補的なブロッキングオリゴ ( D u p C T \_ A ' B C および D u p C T \_ A B ' B C ) (一本鎖 S p ' 配列が後に残る) を、コーディングタグに予めアニーリングさせた。鎖置換ポリメラーゼは、ポリメラーゼ伸長中にブロッキングオリゴを除去する。バーコード凡例 (挿入図) には、記録タグおよびコーディングタグの機能的バーコードへの 1 5 m e r バーコードの帰属が示されている。( B ) 記録タグバーコード設計およびコーディングタグエンコーダーバーコード設計は、記録タグとコーディングタグとの「分子内」対「分子間」相互作用の容易なゲル解析を提供する。この設計では、望ましくない「分子間」相互作用 ( A 記録タグと B ' コーディングタグとの、および B 記録タグと A ' コーディングタグとの ) は、所望の「分子内」 ( A 記録タグと A ' コーディングタグとの ; B 記録タグと B ' コーディングタグとの ) 相互作用産物よりも 1 5 塩基だけ長いまたは短いかなのいずれかであるゲル産物を生成する。プライマー伸長ステップでは、A ' および B ' コーディングタグバーコード ( c t A ' \_ B C , c t B ' \_ B C ) が、リバース相補体バーコード ( c t A \_ B C および c t B \_ B C ) に変更される。( C ) プライマー伸長アッセイは、コーディングタグから記録タグへと情報が移行されたこと、およびアダプター配列が、プライマー伸長により、P C R 解析のためのアニーリングされた E n d C a p オリゴに付加されたことを実証した。( D ) ダミー T 2 0 オリゴを使用して記録タグの表面密度をタイトレーションすることによる「分子内」情報移行の最適化。ビオチン化記録タグオリゴを、1 : 0 から、1 : 1 0、1

10

20

30

40

50

：10000までの全域にわたる種々の比で、ビオチン化ダミーT20オリゴと混合した。低減された記録タグ密度（1：10<sup>3</sup>および1：10<sup>4</sup>）では、「分子内」相互作用が、「分子間」相互作用よりも優勢である。（F）DNAモデル系の単純な伸長として、Nano-Tag15ペプチド-ストレプトアビジン結合対を含む単純なタンパク質結合系が示されているが（K<sub>D</sub> 約4 nM）（Perbandtら、2007年、Proteins、67巻：1147～1153頁）、任意の数のペプチド-結合性物質モデル系を用いることができる。Nano-Tag15ペプチド配列は、（fM）DVEAWLGARVPLVET（配列番号131）（fM=ホルミル-Met）である。Nano-Tag15ペプチドは、短い可撓性リンカーペプチド（GGGGS）およびDNA記録タグにカップリングするためのシステイン残基をさらに含む。他の例示的なペプチドタグ-同種結合性物質対としては、カルモジュリン結合ペプチド（CBP）-カルモジュリン（K<sub>D</sub> 約2 pM）（Mukherjeeら、2015年、J. Mol. Biol.、427巻：2707～2725頁）、アミロイドベータ（A16～27）ペプチド-US7/Lcn2アンチカリン（0.2 nM）（Rauthら、2016年、Biochem. J.、473巻：1563～1578頁）、PAタグ/NZ-1抗体（K<sub>D</sub> 約400 pM）、FLAG-M2 Ab（28 nM）、HA-4B2 Ab（1.6 nM）、およびMyc-9E10 Ab（2.2 nM）（Fujiiら、2014年、Protein Expr. Purif.、95巻：240～247頁）が挙げられる。（E）プライマー伸長による結合性物質のコーディングタグから記録タグへの分子内情報移行の試験としては、相補的DNA配列「A」に結合するオリゴヌクレオチド「結合性物質」を、試験および開発に使用することができる。このハイブリダイゼーション事象は、本質的にfMよりも大きな親和性を示す。Nano-tag15ペプチドエピトープの試験結合性物質としては、ストレプトアビジンを使用してもよい。ペプチドタグ-結合性物質相互作用は高親和性であるが、酸性での洗浄および/または高塩濃度での洗浄により容易に妨害することができる（Perbandtら、上記）。

【図36-2】同上。

【図36-3】同上。

【0228】

【図37】図37A～37Bは、ペプチドのUMI標識NまたはC末端からDNAタグ標識体へと情報を移行させるための、ナノ-またはマイクロ-エマルジョンPCRの使用を示す。（A）ポリペプチドのN-またはC-末端を、一意の分子識別子（UMI）を含む核酸分子で標識する。UMIは、その後のPCRにプライミングするために使用される配列により隣接されていてもよい。その後、ポリペプチドの内部部位を、UMIを隣接するプライミング配列に相補的な配列を含む別々のDNAタグで「本体標識」する。（B）得られた標識ポリペプチドを乳化し、エマルジョンPCR（ePCR）（あるいは、エマルジョンin vitro転写-RT-PCR（IVT-RT-PCR）反応または他の好適な増幅反応を実施することができる）を実施して、N-またはC-末端UMIを増幅する。マイクロエマルジョンまたはナノエマルジョンは、平均液滴直径が50～1000 nmで、1液滴当たり平均で1つ未満のポリペプチドが存在するように形成される。PCR前およびPCR後の液滴内容物の概略が、それぞれ左パネルおよび右パネルに示されている。UMIアンプリコンを、内部ポリペプチド本体DNAと相補的プライミング配列を介してハイブリダイズさせ、UMI情報を、プライマー伸長によりアンプリコンから内部ポリペプチド本体DNAタグへと移行させる。

【0229】

【図38】図38は、単一細胞プロテオミクスを示す。細胞を、ポリマー形成性サブユニット（例えば、アクリルアミド）を含む液滴に封入し、溶解する。ポリマー形成性サブユニットを重合させ（例えば、ポリアクリルアミド）、タンパク質を、ポリマーマトリックスに架橋する。エマルジョン液滴を破壊し、透過性ポリマーマトリックスに付着した単一細胞タンパク質ライセートを含む重合ゲルビーズを放出させる。タンパク質は、天然コンフォメーションで、または溶解および封入緩衝液中に尿素などの変性剤を含めることによ

10

20

30

40

50

り変性状態でのいずれかでポリマーマトリックスに架橋する。コンパートメントバーコードおよび他の記録タグ成分（例えば、ユニバーサルプライミング配列（P1）、スペーサー配列（Sp）、任意選択の一意的分子識別子（UMI））を含む記録タグを、バーコード付きビーズとの乳化またはコンビナトリアルインデックス化を含む、当技術分野で公知のおよび本明細書に記載のいくつかの方法を使用してタンパク質に付着させる。また、単一細胞タンパク質を含む重合ゲルビーズを、記録タグ付加後にプロテイナーゼ消化にかけて、ペプチド配列決定に好適な記録タグ標識ペプチドを生成することができる。ある特定の実施形態では、ポリマーマトリックスは、tris（2-カルボキシエチル）ホスフィン（TCPEP）またはジチオトレイトール（DTT）などの還元剤に曝されると破壊されるジスルフィド架橋ポリマーなどの適切な添加剤に溶解するように設計することができる。

10

【0230】

【図39】図39A～39Eは、二機能性N末端アミノ酸（NTAA）修飾因子およびキメラ切断試薬を使用したアミノ酸切断反応の増強を示す。（A）および（B）固相基材に付着しているペプチドを、ピオチン-フェニルイソチオシアネート（PITC）などの二機能性NTAA修飾因子で修飾する。（C）低親和性エドマナーゼ（ $> \mu\text{M}$  K<sub>d</sub>）を、ストレプトアビジン-エドマナーゼキメラタンパク質を使用して、ピオチン-PITC標識NTAAに動員する。（D）エドマナーゼ切断の効率は、ピオチン-ストレプトアビジン（streptavidin）相互作用の結果として有効局所濃度が増加するため、大幅に向上する。（E）切断されたピオチン-PITC標識NTAAおよび付随するストレプトアビジン-エドマナーゼキメラタンパク質は、切断後に遠方に拡散する。いくつかの他のバイオコンジュゲーション動員戦略も用いることができる。アジド修飾PITCは、市販の（4-アジドフェニルイソチオシアネート、Sigma）であり、アルキン-ピオチンとのクリック化学反応によるピオチン-PITCなどの、PITCの他のバイオコンジュゲートへのアジド-PITCのいくつかの単純な変換を可能にする。

20

【0231】

【図40-1】図40A～40Iは、タンパク質ライセート（ゲルビーズに封入されていてもよい）に由来するC末端記録タグ標識ペプチドの生成を示す。（A）変性ポリペプチドを酸無水物と反応させて、リシン残基を標識する。一実施形態では、アルキン（mTet）置換シトラコン酸無水物+プロピオン酸無水物のミックスを使用して、リシンをmTetで標識する。（縞模様の長方形として示されている）（B）その結果は、一部のリシンがプロピオン基（ポリペプチド鎖にある正方形として示されている）でブロックされたアルキン（mTet）標識ポリペプチドである。アルキン（mTet）部分は、クリック化学に基づくDNA標識に有用である。（C）DNAタグ（黒抜き長方形として示されている）を、アジドまたはtrans-シクロオクテン（TCO）標識を使用して、それぞれアルキンまたはmTet部分にクリック化学により付着させる。（D）バーコードならびにスペーサー（Sp）配列およびユニバーサルプライミング配列などの機能的エレメントを、図31に示されているようなプライマー伸長ステップを使用してDNAタグに追加して、記録タグ標識ポリペプチドを産生する。バーコードは、試料バーコード、分配バーコード、コンパートメントバーコード、空間位置バーコードなど、またはそれらの任意の組合せであってもよい。（E）得られた記録タグ標識ポリペプチドを、プロテアーゼでまたは化学的に記録タグ標識ペプチドに断片化する。（F）例示のために、2つの記録タグで標識されているペプチド断片が示されている。（G）記録タグのユニバーサルプライミング配列に相補的なユニバーサルプライミング配列を含むDNAタグを、ペプチドのC末端にライゲーションさせる。また、C末端DNAタグは、ペプチドを表面にコンジュゲートするための部分を含む。（H）C末端DNAタグの相補的ユニバーサルプライミング配列および確率論的に選択された記録タグがアニーリングする。分子内プライマー伸長反応を使用して、記録タグからC末端DNAタグへと情報を移行させる。（I）ペプチドの内部記録タグを、無水マレイン酸を介してリシン残基にカップリングする。このカップリングは酸性pHで可逆的である。内部記録タグを、酸性pHにてペプチドのリシン残基から切断し、C末端記録タグはそのまま残る。任意選択で、新しく露出したリシン残基を

30

40

50

、プロピオン無水物などの非加水分解性無水物でブリッキングしてもよい。

【図 4 0 - 2】同上。

【 0 2 3 2 】

【図 4 1】図 4 1 は、N G P S アッセイの好ましい実施形態のワークフローを示す。

【 0 2 3 3 】

【図 4 2】図 4 2 A ~ 4 2 D は、N G P S シーケンシングアッセイの例示的なステップを示す。記録タグで標識した表面に結合されているペプチドの N 末端アミノ酸 ( N T A A ) アセチル化またはアミジン化ステップは、N T A A 結合性物質が、アセチル化 N T A A に結合するように遺伝子操作されているか、または天然 N T A A に結合するように遺伝子操作されているかに応じて、N T A A 結合性物質による結合の前に生じてもよく、または結合の後で生じてもよい。第 1 の場合、( A ) まず、ペプチドの N T A A を、無水酢酸を使用して化学的手段により、または N 末端アセチルトランスフェラーゼ ( N A T ) を用いて酵素的にアセチル化する。( B ) N T A A は、遺伝子操作されたアンチカリン、アミノアシル t R N A シンテターゼ ( a a R S )、C l p S などの N T A A 結合性物質により認識される。D N A コーディングタグは、結合性物質に付着しており、特定の N T A A 結合性物質を識別するバーコードエンコーダー配列を含む。( C ) アセチル化 N T A A が N T A A 結合性物質と結合した後に、D N A コーディングタグは、相補的配列を介して記録タグと一時的にアニーリングし、コーディングタグ情報が、ポリメラーゼ伸長により記録タグへと移行される。代替的な実施形態では、記録タグ情報は、ポリメラーゼ伸長によりコーディングタグへと移行される。( D ) アセチル化 N T A A は、アセチル化ペプチドの末端アセチル化アミノ酸の加水分解を触媒する遺伝子操作されたアシルペプチドヒドロラーゼ ( A P H ) によりペプチドから切断される。アセチル化 N T A A の切断後、このサイクルは、新たに露出した N T A A のアセチル化から開始して自発的に繰り返される。N 末端アセチル化が、N T A A 修飾 / 切断の例示的なモードとして使用されているが、その代わりに、グアニル部分などの他の N 末端部分を、それに応じて切断化学を変更して置換してもよい。グアニジン化を用いる場合、グアニル化 N T A A は、0 . 5 ~ 2 % N a O H 溶液を使用して、穏やかな条件下で切断することができる ( その全体が参照により組み込まれる、H a m a d a、2 0 1 6 年を参照されたい )。A P H は、ブロッキングされたペプチドの N - アセチル化アミノ酸の除去を触媒することができるセリンペプチダーゼであり、プロリルオリゴペプチダーゼ ( P O P ) ファミリー ( クラン S C、ファミリー S 9 ) に属する。A P H は、真核生物細胞、細菌細胞、および古細菌細胞の N - 末端アセチル化タンパク質の重要な制御因子である。

【 0 2 3 4 】

【図 4 3】図 4 3 A ~ 4 3 B は、例示的な記録タグ - コーディングタグ設計特徴を示す。

( A ) 例示的な記録タグ付随タンパク質 ( またはペプチド ) および付随するコーディングタグを有する結合されている結合性物質 ( 例えば、アンチカリン ) の構造。プライマー伸長反応での確率論的な非鋳型 3 ' 末端アデノシン ( A ) 付加を受け入れるために、コーディングタグのスペーサー ( S p ' ) とバーコード ( B C ' ) 配列との間にチミジン ( T ) 塩基が挿入されている。( B ) D N A コーディングタグは、S p y C a t c h e r - S p y T a g タンパク質 - ペプチド相互作用により結合性物質 ( 例えば、アンチカリン ) に付着している。

【 0 2 3 5 】

【図 4 4】図 4 4 A ~ 4 4 E は、記録タグへの切断性物質のハイブリダイゼーションを使用した N T A A 切断反応の増強を示す ( A ) および ( B )。固相基材 ( 例えば、ビーズ ) に付着している記録タグ標識ペプチドの N T A A を、例えば、P I T C、D N P、S N P、アセチル修飾因子、グアニジン化などで修飾または標識する ( M o d )。( C ) 切断酵素 ( 例えば、アシルペプチドヒドロラーゼ ( A P H )、アミノペプチダーゼ ( A P )、エドマナーゼなど ) を、記録タグのユニバーサルプライミング配列に相補的なユニバーサルプライミング配列を含む D N A タグに付着させる。切断酵素は、切断酵素の D N A タグおよび記録タグの相補的ユニバーサルプライミング配列のハイブリダイゼーションにより、

修飾 N T A A に動員される。(D) このハイブリダイゼーションステップは、N T A A に対する切断酵素の有効親和性を大幅に向上させる。(E) 切断された N T A A は、遠方に拡散し、付随する切断酵素は、ハイブリダイズされた D N A タグを剥離することにより除去することができる。

#### 【0236】

【図45】図45は、ペプチドリガーゼ+プロテアーゼ+ジアミノペプチダーゼを使用したサイクル分解ペプチド配列決定を示す。プテラーゼIにより、T E V - プテラーゼI ペプチド基質 (T E N L Y F Q N H V、配列番号132) が、クエリペプチドの N T A A にライゲーションされる。プテラーゼは、ペプチド基質のC末端に N H V モチーフを必要とする。ライゲーション後、タバコエッチウイルス (T E V) プロテアーゼを使用して、グルタミン (Q) 残基の後でキメラペプチド基質を切断して、クエリペプチドのN末端にアスパラギン (N) 残基が付着したキメラペプチドが得られる。2つのアミノ酸残基をN末端から切断するジアミノペプチダーゼ (D A P) またはジペプチジルペプチダーゼにより、N付加されたクエリペプチドが2つのアミノ酸だけ短くなり、クエリペプチドのアスパラギン残基 (N) および元の N T A A が効果的に除去される。新たに露出した N T A A を、本明細書で提供されているような結合性物質を使用して読み取り、その後、サイクル全体を「n」回繰り返して、「n」個のアミノ酸を配列決定する。ストレプトアビジン - D A P 金属酵素キメラタンパク質を使用することにより、およびビオチン部分をN末端アスパラギン残基に係留することにより、D A P 処理能力の制御を可能にすることができる。

#### 【発明を実施するための形態】

#### 【0237】

本明細書において具体的に定義されていない用語には、本開示および文脈を踏まえて当業者により与えられる意味であろう意味が与えられるべきである。しかし、本明細書で使用される場合、それに反する指定がなければ、用語は示された意味を有する。

#### I. 緒言

#### 【0238】

本開示は、一部において、タンパク質およびペプチドの特徴付けおよび配列決定に直接適用される、高度に並行な、ハイスループットなデジタル巨大分子特徴付けおよび定量化方法を提供する(図1B、図2Aを参照されたい)。本明細書に記載の方法では、識別情報を核酸分子または配列決定可能なポリマーの形態で有するコーディングタグを含む結合性物質を使用し、ここで、結合性物質は、目的の巨大分子と相互作用するものである。多数の連続的な結合サイクルを実施し、各サイクルは、プールされた試料を表すものであることが好ましい、固体支持体上に固定化した複数の巨大分子を複数の結合性物質に曝露させることを含む。各結合サイクル中、巨大分子に結合する各結合性物質の同一性、および任意選択で結合サイクル数を、結合性物質コーディングタグから巨大分子と共局在する記録タグに情報を移行させることによって記録する。代替の実施形態では、付随する巨大分子に関する識別情報を含む記録タグからの情報を、結合した結合性物質のコーディングタグ(例えば、伸長コーディングタグを形成するため)または第3の「ジタグ」構築物に移行させることができる。多数サイクルの結合事象により、巨大分子と共局在する記録タグに関する歴史的な結合情報が構築され、それにより、多数のコーディングタグを含む伸長記録タグが、所与の巨大分子についての時間的な結合履歴を表す共直線的な順序で生じる。さらに、サイクル特異的コーディングタグを使用して各サイクルからの情報を追跡することができる。したがって、あるサイクルが何らかの理由でスキップされた場合に、伸長記録タグがその後のサイクルで情報を収集し続け、情報が欠如したサイクルを識別することができる。

#### 【0239】

あるいは、コーディングタグから記録タグに情報を書き込むまたは移行させる代わりに、付随する巨大分子に関する識別情報を含む記録タグから、伸長コーディングタグを形成するコーディングタグまたは第3のジタグ構築物に情報を移行させることができる。得られた伸長コーディングタグまたはジタグを、その後の配列解析のために各結合サイクル後

に収集することができる。バーコード（例えば、分配タグ、コンパートメントタグ、試料タグ、画分タグ、U M I、またはそれらの任意の組合せ）を含む記録タグ上の識別情報を使用して、伸長コーディングタグまたはジタグ配列読み取りを元の巨大分子にマッピングし戻すことができる。このように、巨大分子の結合履歴の核酸コードライブラリー表示を生成する。この核酸コードライブラリーを、非常にハイスループットの次世代デジタルシーケンシング法を使用して増幅させ、解析し、それにより、実行当たり数百万～数十億の分子を解析することができる。結合情報に関する核酸コードライブラリーの創出は、ハイブリダイゼーションを使用するDNAに基づく技法による濃縮、サブトラクション、および正規化が可能になるという点で、別のように有用である。これらのDNAに基づく方法は、容易におよび迅速に大規模化可能かつカスタマイズ可能であり、タンパク質ライブラリーなどの他の種類の巨大分子ライブラリーの直接操作のために利用可能なものよりも費用効果が高い。したがって、結合情報に関する核酸コードライブラリーを配列決定前に1つまたは複数の技法によって処理して、配列の表示を濃縮および/またはサブトラクションおよび/または正規化することができる。これにより、最大の目的の情報が、個々のメンバーの豊富さが多数の桁にわたって最初に変動し得る非常に大きなライブラリーから、はるかに効率的に、迅速に、かつ大きな費用効果で抽出される。重要なことに、ライブラリー表示を操作するためのこれらの核酸に基づく技法は、より慣習的な方法と直交性のものであり、それらと組み合わせて使用することができる。例えば、アルブミンなどの、一般的な、極めて豊富なタンパク質を、望ましくないタンパク質の全てではないが大多数を除去することができるタンパク質に基づく方法を使用してサブトラクションすることができる。その後、伸長記録タグライブラリーのアルブミン特異的メンバーもサブトラクションし、したがって、より徹底的な全体的サブトラクションを実現することができる。

#### 【0240】

一態様では、本開示は、DNA記録タグで標識されたペプチドの大きな集団（例えば、数百万～数十億）からの配列決定を可能にする、エドマン様分解手法を使用した、ペプチド配列決定のための高度に並行化された手法を提供する。これらの記録タグで標識されたペプチドは、タンパク質試料のタンパク質分解による消化または限定された加水分解に由来するものであり、記録タグで標識されたペプチドは、配列決定基板（例えば、多孔質ビーズ）に、基板上の適切な分子間間隔でランダムに固定化される。NTAA切断反応を触媒するまたは動員する、フェニルチオカルバモイル（PTC）、ジニトロフェノール（DNP）、スルホニルニトロフェノール（SNP）、ダンシル、7-メトキシクマリン、アセチル、またはグアニジニルなどの小さな化学的部分を用いたペプチドのN末端アミノ酸（NTAA）残基の修飾により、エドマン様分解プロセスの周期的制御が可能になる。修飾用化学的部分により、同類のNTAA結合性物質に対する結合親和性の増強をもたらすこともできる。各固定化ペプチドの修飾されたNTAAを、コーディングタグを含む同類のNTAA結合性物質の結合、および、コーディングタグ情報（例えば、結合性物質に関する識別情報をもたらすエンコーダー配列）のコーディングタグからペプチドの記録タグへの移行（例えば、プライマー伸長またはライゲーション）によって識別する。その後、修飾されたNTAAを化学的方法または酵素的手段によって除去する。ある特定の実施形態では、修飾されたNTAAの除去を触媒させるために酵素（例えば、エドマナーゼ）を工学的に操作する。他の実施形態では、アミノペプチダーゼまたはアシルペプチドヒドロラーゼなどの天然に存在するエキソペプチダーゼを、適切な化学修飾の存在下でのみ末端アミノ酸を切断するように工学的に操作することができる。

#### II. 定義

#### 【0241】

以下の説明では、種々の実施形態の詳細な理解をもたらすために、ある特定の具体的詳細を記載する。しかし、これらの詳細を伴わずに本化合物を作製および使用できることが当業者には理解されよう。他の場合では、実施形態の説明が不必要に不明瞭になるのを回避するために、周知の構造は詳細に示されていないまたは記載されていない。文脈上異なる解釈を要する場合を除き、本明細書およびそれが従う特許請求の範囲全体を通して、「

10

20

30

40

50

含む (comprise)」という単語および「含む (comprises)」および「含む (comprising)」などのその変形は、制限のない、包括的な意味で、すなわち、「含むが、これだけに限定されない (including, but not limited to)」と解釈されるべきである。さらに、「含む (comprising)」という用語 (および「含む (comprise)」または「含む (comprises)」または「有する (having)」または「含む (including)」などの関連する用語) は、他のある特定の実施形態では、例えば、本明細書に記載の任意の組成物 (composition of matter)、組成物 (composition)、方法、またはプロセスなどのある実施形態が、記載されている特徴「からなる (consist of)」または「から本質的になる (consist essentially of)」ものであり得ることを排除するものではない。本明細書で提示される表題は単に便宜上のものであり、特許請求された実施形態の範囲または意味とは解釈されない。

10

**【0242】**

本明細書全体を通して、「一実施形態」または「ある実施形態」への言及は、当該実施形態に関連して記載される特定の特徵、構造または特性が少なくとも1つの実施形態に含まれることを意味する。したがって、本明細書全体を通して各所での「一実施形態では」または「ある実施形態では」という句の出現は、必ずしも全てが同じ実施形態について言及しているのではない。さらに、特定の特徵、構造、または特性は、任意の適切な様式で組み合わせて1つまたは複数の実施形態にすることができる。

**【0243】**

20

本明細書で使用される場合、単数形「1つの (a)」、「1つの (an)」および「その (the)」は、文脈により明確に別段の規定がなされない限り、複数の指示対象を包含する。したがって、例えば、「1つのペプチド (a peptide)」への言及は、1つもしくは複数のペプチド、またはペプチドの混合物を含む。また、特に明記されていないまたは文脈から明らかでない限り、本明細書で使用される場合、「または (or)」という用語は、包括的であり、「または (or)」と「および (and)」の両方を包含するものと理解される。

**【0244】**

本明細書で使用される場合、「巨大分子」という用語は、より小さなサブユニットで構成される大きな分子を包含する。巨大分子の例としては、これだけに限定されないが、ペプチド、ポリペプチド、タンパク質、核酸、炭水化物、脂質、大環状分子が挙げられる。巨大分子は、共有結合により連結した2つまたはそれよりも多くの型の巨大分子の組合せで構成されるキメラ巨大分子 (例えば、核酸と連結したペプチド) も含む。巨大分子は、2つまたはそれよりも多くの巨大分子の非共有結合性の複合体で構成される「巨大分子集合体」も含み得る。巨大分子集合体は、同じ型の巨大分子で構成されるもの (例えば、タンパク質 - タンパク質) であってもよく、2つまたはそれよりも多くの異なる型の巨大分子で構成されるもの (例えば、タンパク質 - DNA) であってもよい。

30

**【0245】**

本明細書で使用される場合、「ペプチド」という用語は、ペプチド、ポリペプチドおよびタンパク質を包含し、ペプチド結合によって接合した2つまたはそれよりも多くのアミノ酸の鎖を含む分子を指す。一般に述べると、20~30個よりも多くのアミノ酸を有するペプチドが一般にポリペプチドと称され、50個よりも多くのアミノ酸を有するペプチドが一般にタンパク質と称される。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、修飾されたアミノ酸、アミノ酸類似体、アミノ酸模倣物、またはそれらの任意の組合せであってもよい。ペプチドは、天然に存在するものであってもよく、合成的に作製されたものであってもよく、組換えによって発現させたものであってもよい。ペプチドは、アミノ酸の鎖を修飾する追加的な基、例えば、翻訳後修飾によって付加された官能基も含んでよい。

40

**【0246】**

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

50

トとして機能する、アミン基、カルボン酸基、および各アミノ酸に特異的な側鎖を有する有機化合物を指す。アミノ酸は、20種の標準の天然に存在するまたは正規のアミノ酸ならびに非標準アミノ酸を含む。標準の天然に存在するアミノ酸としては、アラニン（AまたはAla）、システイン（CまたはCys）、アスパラギン酸（DまたはAsp）、グルタミン酸（EまたはGlu）、フェニルアラニン（FまたはPhe）、グリシン（GまたはGly）、ヒスチジン（HまたはHis）、イソロイシン（IまたはIle）、リシン（KまたはLys）、ロイシン（LまたはLeu）、メチオニン（MまたはMet）、アスパラギン（NまたはAsn）、プロリン（PまたはPro）、グルタミン（QまたはGln）、アルギニン（RまたはArg）、セリン（SまたはSer）、トレオニン（TまたはThr）、バリン（VまたはVal）、トリプトファン（WまたはTrp）、およびチロシン（YまたはTyr）が挙げられる。アミノ酸は、L-アミノ酸であってもD-アミノ酸であってもよい。非標準アミノ酸は、天然に存在するまたは化学的に合成された修飾されたアミノ酸、アミノ酸類似体、アミノ酸模倣物、非標準タンパク質新生性アミノ酸、または非タンパク質新生性アミノ酸であり得る。非標準アミノ酸の例としては、これだけに限定されないが、セレノシステイン、ピロリジン、およびN-ホルミルメチオニン、-アミノ酸、ホモアミノ酸、プロリンおよ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およびチロシン誘導体、直鎖コアアミノ酸、N-メチルアミノ酸が挙げられる。

#### 【0247】

本明細書で使用される場合、「翻訳後修飾」という用語は、リボソームによるペプチドの翻訳が完了した後に当該ペプチド上で生じる修飾を指す。翻訳後修飾は、共有結合性修飾または酵素的修飾であり得る。翻訳後修飾の例としては、これだけに限定されないが、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、ジフタミド形成、ジスルフィド架橋形成、エリミニル化、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、グリコシルホスファチジルイノシトール付加（glypiation）、ヘムC付着、ヒドロキシル化、ハイプシン形成、ヨウ素化、イソプレニル化、脂質付加、リポイル化、マロニル化、メチル化、ミリストイル化、酸化、パルミトイル化、ベグ化、ホスホパンテテイル化、リン酸化、プレニル化、プロピオニル化、レチニリデンシッフ塩基形成、S-グルタチオン化、S-ニトロシル化、S-スルフェニル化、セレン化、サクシニル化、スルフィン化、ユビキチン化、およびC末端アミド化が挙げられる。翻訳後修飾は、ペプチドのアミノ末端および/またはカルボキシル末端の修飾を含む。末端アミノ基の修飾としては、これだけに限定されないが、デスアミノ、N-低級アルキル、N-ジ低級アルキル、およびN-アシル修飾が挙げられる。末端カルボキシ基の修飾としては、これだけに限定されないが、アミド、低級アルキルアミド、ジアルキルアミド、およ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。翻訳後修飾は、例えば、これだけに限定されないが、アミノ末端とカルボキシ末端の間にあるアミノ酸の、上記のものなどの修飾も含む。翻訳後修飾という用語は、1つまたは複数の検出可能な標識を含むペプチド修飾も含み得る。

#### 【0248】

本明細書で使用される場合、「結合性物質」という用語は、巨大分子または巨大分子の成分もしくは特徴に結合する、結び付く、それと合体する、それを認識する、またはそれと組み合わせる核酸分子、ペプチド、ポリペプチド、タンパク質、炭水化物、または小分子を指す。結合性物質は、巨大分子または巨大分子の成分もしくは特徴と共有結合性の結び付きまたは非共有結合性の結び付きを形成し得る。結合性物質はまた、核酸分子-ペプチドキメラ結合性物質または炭水化物-ペプチドキメラ結合性物質などの、2つまたはそれよりも多くの型の分子で構成されるキメラ結合性物質であってもよい。結合性物質は、天然に存在する分子であってもよく、合成的に作製された分子であってもよく、組換えによって発現させた分子であってもよい。結合性物質は、巨大分子の単一の単量体またはサブユニット（例えば、ペプチドの単一のアミノ酸）に結合し得る、または巨大分子の複数



の連結したサブユニット（例えば、より長いペプチド、ポリペプチド、もしくはタンパク質分子のジ - ペプチド、トリ - ペプチド、またはより高次のペプチド）に結合し得る。結合性物質は、直鎖状分子または三次元構造（コンフォメーションとも称される）を有する分子に結合し得る。例えば、抗体結合性物質は、直鎖ペプチド、ポリペプチド、もしくはタンパク質に結合し得る、または、コンフォメーションペプチド、ポリペプチド、もしくはタンパク質に結合し得る。結合性物質は、ペプチド、ポリペプチド、またはタンパク質分子の N 末端ペプチド、C 末端ペプチド、または介在するペプチドに結合し得る。結合性物質は、ペプチド分子の N 末端アミノ酸、C 末端アミノ酸、または介在するアミノ酸に結合し得る。結合性物質は、好ましくは、化学修飾されたまたは標識されたアミノ酸に、改変されていないまたは標識されていないアミノ酸よりも優先的に結合し得る。例えば、結合性物質は、好ましくは、アセチル部分、グアニル部分、ダンシル部分、PTC 部分、DNB 部分、SNB 部分などで修飾されたアミノ酸に、前記部分を保有しないアミノ酸よりも優先的に結合し得る。結合性物質は、ペプチド分子の翻訳後修飾に結合し得る。結合性物質は、巨大分子の成分もしくは特徴への選択的結合を示し得る（例えば、結合性物質は、20 種の可能性のある天然のアミノ酸残基のうちの 1 種に選択的に結合することができ、他の 19 種の天然のアミノ酸残基には非常に低い親和性で結合するまたは全く結合しない）。結合性物質は、結合性物質が複数の巨大分子の成分もしくは特徴に結合することが可能な場合、より低い選択的結合を示し得る（例えば、結合性物質は、2 つまたはそれよりも多くの異なるアミノ酸残基に同様の親和性で結合し得る）。結合性物質は、コーディングタグを含み、これは、リンカーによって結合性物質に接合されている。

10

20

#### 【0249】

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、または非ヌクレオチド化学的部分の 1 つまたは複数を指す。リンカーは、結合性物質とコーディングタグを接合するため、記録タグと巨大分子（例えば、ペプチド）を接合するため、巨大分子と固体支持体を接合するため、記録タグと固体支持体などを接合するために使用することができる。ある特定の実施形態では、リンカーにより、2 つの分子が酵素反応または化学反応（例えば、クリックケミストリー）を介して接合される。

#### 【0250】

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、および体液内のプロテオーム、ならびに対応する細胞内および組織内のプロテオームの空間的分布の定量的分析を指す。さらに、プロテオミクス試験は、生物学および定義された生物学的または化学的刺激に応じて継続的に時間変化するプロテオームの動的状態を含む。

30

#### 【0251】

本明細書で使用される場合、「非同類結合性物質」という用語は、特定の結合サイクル反応において調査される巨大分子の特徴、成分、またはサブユニットに、対応する巨大分子の特徴、成分、またはサブユニットに高親和性で結合する「同類結合性物質」と比較して、結合することができないまたは低親和性で結合する結合性物質を指す。例えば、ペプチド分子のチロシン残基を結合反応において調査する場合、非同類結合性物質は、チロシン残基に低親和性で結合するまたは全く結合しないものであり、したがって、非同類結合性物質では、コーディングタグ情報を同類結合性物質から記録タグに移行させるために適した条件下でコーディングタグ情報が記録タグに効率的に移行されない。あるいは、ペプチド分子のチロシン残基を結合反応において調査する場合、非同類結合性物質は、チロシン残基に低親和性で結合するまたは全く結合しないものであり、したがって、伸長記録タグではなく伸長コーディングタグを伴う実施形態に適した条件下で、記録タグ情報はコーディングタグに効率的に移行されない。

40

#### 【0252】

遊離のアミノ基を有する、ペプチド鎖の一方の末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。遊離のカルボキシル基を有する、鎖の他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。

50

ペプチドを構成するアミノ酸には、順番に番号を付すことができ、ペプチドは「 $n$ 」アミノ酸長になる。本明細書で使用される場合、NTAAは、 $n$  番目のアミノ酸と考えられる（本明細書では「 $n$  NTAA」とも称される）。この命名法を使用すると、N末端からC末端までのペプチドの長さを下方に次のアミノ酸は $n - 1$ アミノ酸、次いで $n - 2$ アミノ酸などである。ある特定の実施形態では、NTAA、CTAA、またはその両方を化学的部分で修飾または標識することができる。

#### 【0253】

本明細書で使用される場合、「バーコード」という用語は、巨大分子（例えば、タンパク質、ポリペプチド、ペプチド）、結合性物質、結合サイクルからの結合性物質のセット、試料巨大分子、試料のセット、コンパートメント（例えば、液滴、ビーズ、または分離された位置）内の巨大分子、コンパートメントのセット内の巨大分子、巨大分子の画分、巨大分子画分のセット、空間的領域または空間的領域のセット、巨大分子のライブラリー、または結合性物質のライブラリーについての一意の識別子タグまたは起源情報をもたらす約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または30塩基）の核酸分子を指す。バーコードは、人工的な配列であっても天然に存在する配列であってもよい。ある特定の実施形態では、バーコードの集団内の各バーコードは異なるものである。他の実施形態では、バーコードの集団のバーコードの一部が異なる、例えば、バーコードの集団の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%、または99%が異なる。バーコードの集団は、ランダムに生成することもでき、非ランダムに生成することもできる。ある特定の実施形態では、バーコードの集団は、エラー訂正バーコードである。バーコードは、多重化された配列決定データをコンピュータによりデコンボリューションし、個々の巨大分子、試料、ライブラリーなどに由来する配列読み取りを識別するために使用することができる。バーコードはまた、マッピングを増強するための小さなコンパートメント中に分布させた巨大分子の集合のデコンボリューションのために使用することもできる。例えば、ペプチドをプロテオームにマッピングし戻すのではなく、ペプチドをその起源であるタンパク質分子またはタンパク質複合体にマッピングし戻す。

#### 【0254】

「試料バーコード」は、「試料タグ」とも称され、巨大分子がいずれの試料に由来するかを識別するものである。

#### 【0255】

「空間バーコード」は、巨大分子が2Dまたは3D組織切片のいずれの領域に由来するかを識別するものである。空間バーコードは、組織切片に関する分子病理学のために使用することができる。空間バーコードにより、組織切片（複数可）由来の複数の試料またはライブラリーのマルチプレックス配列決定が可能になる。

#### 【0256】

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、それに付随する結合性物質に関する識別情報を含む、2および100ならびにその間のあらゆる整数を含めて約2塩基～約100塩基の核酸分子を指す。「コーディングタグ」は、「配列決定可能なポリマー」で作られたものであってよい（例えば、それぞれ、その全体が参照により組み込まれる、Niura、2013年、Nat. Chem.、5巻：282～292頁；Royら、2015年、Nat. Commun.、6巻：7237頁；Lutz、2015年、Macromolecules、48巻：4759～4767頁を参照されたい）。コーディングタグは、任意選択で片側に1つのスペーサーが隣接するまたは両側にスペーサーが隣接するエンコーダー配列を含む。コーディングタグはまた、任意選択のUMIおよび/または任意選択の結合サイクル特異的バーコードで構成されてもよい。コーディングタグは、一本鎖であっても二本鎖であってもよい。二本鎖コーディングタグは、平滑末端、突出末端、またはその両方を含んでよい。コーディングタグとは、結合性物質、結合

性物質に直接付着したコーディングタグと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に関して）、または、伸長記録タグに存在するコーディングタグ情報に直接付着したコーディングタグを指し得る。ある特定の実施形態では、コーディングタグは、結合サイクル特異的スペーサーまたはバーコード、一意の分子識別子、ユニバーサルプライミング部位、またはそれらの任意の組合せをさらに含んでよい。

#### 【0257】

本明細書で使用される場合、「エンコーダー配列」または「エンコーダーバーコード」という用語は、それが付随する結合性物質に関する識別情報をもたらす、約2塩基～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または30塩基）の長さの核酸分子を指す。エンコーダー配列は、それが付随する結合性物質を一意的に識別することができるものである。ある特定の実施形態では、エンコーダー配列により、それが付随する結合性物質および結合性物質が使用される結合サイクルに関する識別情報がもたらされる。他の実施形態では、エンコーダー配列をコーディングタグ内の別の結合サイクル特異的バーコードと組み合わせる。あるいは、エンコーダー配列により、それに付随する結合性物質を2種またはそれよりも多くの異なる結合性物質のセットのメンバーに属するものと識別することができる。一部の実施形態では、解析のためにはこの識別のレベルで十分である。例えば、アミノ酸に結合する結合性物質を伴う一部の実施形態では、ペプチドの特定の位置におけるアミノ酸残基を決定的に識別するのではなく、ペプチドがその位置において2つの可能性のあるアミノ酸のうちの1つを含むことを知ることで十分であり得る。別の例では、タンパク質標的の1種よりも多くのエピトープを認識し、様々な特異性を有する抗体の混合物を含むポリクローナル抗体に共通のエンコーダー配列を使用する。他の実施形態では、エンコーダー配列により可能性のある結合性物質のセットを識別する場合、逐次的な脱コーディング手法を使用して、各結合性物質の一意的識別をもたらすことができる。これは、繰り返される結合のサイクルにおいて所与の結合性物質に対するエンコーダー配列を変動させるによって実現される（Gundersenら、2004年、Genome Res.、14巻：870～7頁を参照されたい）。各結合サイクルからのコーディングタグ情報を部分的に識別することにより、他のサイクルからのコーディング情報と組み合わせると、結合性物質について一意の識別子がもたらされ、例えば、個々のコーディングタグ（またはエンコーダー配列）ではなく、コーディングタグの特定の組合せにより、結合性物質に関する一意の識別情報がもたらされる。結合性物質のライブラリー内のエンコーダー配列は同じまたは同様の数の塩基を有することが好ましい。

#### 【0258】

本明細書で使用される場合、「結合サイクル特異的タグ」、「結合サイクル特異的バーコード」、または「結合サイクル特異的配列」という用語は、特定の結合サイクル内で使用される結合性物質のライブラリーを識別するために使用される一意の配列を指す。結合サイクル特異的タグは、約2塩基～約8塩基（例えば、2、3、4、5、6、7、または8塩基）の長さを含み得る。結合サイクル特異的タグは、結合性物質のコーディングタグ内に、スペーサー配列の一部として、エンコーダー配列の一部として、UMIの一部として、またはコーディングタグ内の別の成分として組み入れることができる。

#### 【0259】

本明細書で使用される場合、「スペーサー」（Sp）という用語は、記録タグまたはコーディングタグの末端上に存在する約1塩基～約20塩基（例えば、1、2、3、4、5、6、7、8、9、10、11、12、13、14、15、16、17、18、19、または20塩基）の長さの核酸分子を指す。ある特定の実施形態では、スペーサー配列は、一方の末端または両方の末端におけるコーディングタグのエンコーダー配列に隣接する。結合性物質の巨大分子への結合後、それらに付随するコーディングタグおよび記録タグ上の相補的なスペーサー配列間のアニーリングにより、それぞれ、結合情報の、プライマー伸長反応またはライゲーションによる記録タグ、コーディングタグ、またはジタグ構築物

への移行が可能になる。S p ' は、S p に相補的なスペーサー配列を指す。結合性物質の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ことが好ましい。共通する（共有されるまたは同一の）スペーサーを結合性物質のライブラリーにおいて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性物質を追跡するために、「サイクル特異的」配列を有してよい。スペーサー配列（S p）は、結合サイクル全てにわたって一定のものであってもよく、特定のクラスの巨大分子に対して特異的なものであってもよく、結合サイクル数に特異的なものであってもよい。巨大分子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同類結合性物質の伸長記録タグに存在するコーディングタグ情報が、その後の結合サイクルにおいて同じクラスの巨大分子を認識する別の結合性物質のコーディングタグと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可能になる。正確な同類の対の逐次的な結合によってのみ、相互作用するスペーサーエレメントおよび有効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なスペーサー配列と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、またはライゲーション反応のための「副子」をもたらす、または「粘着末端」ライゲーション反応を媒介するため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。

#### 【0260】

本明細書で使用される場合、「記録タグ」という用語は、それが付随する巨大分子に関する識別情報を含む核酸分子または配列決定可能なポリマー分子を指す（例えば、それぞれ、その全体が参照により組み込まれる、Niuら、2013年、Nat. Chem.、5巻：282～292頁；Royら、2015年、Nat. Commun.、6巻：7237頁；Lutz、2015年、Macromolecules、48巻：4759～4767頁を参照されたい）。ある特定の実施形態では、結合性物質が巨大分子に結合した後、結合性物質と連結しているコーディングタグからの情報を、結合性物質が巨大分子に結合している間に巨大分子に付随する記録タグに移行させることができる。他の実施形態では、結合性物質が巨大分子に結合した後、巨大分子に付随する記録タグからの情報を、結合性物質が巨大分子に結合している間に結合性物質と連結しているコーディングタグに移行させることができる。記録タグは、巨大分子に直接連結していてもよく、巨大分子に多機能性リンカーを介して連結していてもよく、固体支持体で巨大分子の近傍にある（または共局在する）ことによって巨大分子に付随していてもよい。記録タグは、連結がコーディングタグ情報を記録タグに移行させるまたはその逆のために使用される方法に適合するものである限りは、5'末端または3'末端に連結していてもよく、内側の部位に連結していてもよい。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、一意の分子識別子、バーコード（例えば、試料バーコード、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と相補的なスペーサー配列、またはそれらの任意の組合せをさらに含んでよい。記録タグのスペーサー配列は、コーディングタグ情報を記録タグに移行させるためにポリメラーゼ伸長を使用する実施形態では記録タグの3'末端にあることが好ましい。

#### 【0261】

本明細書で使用される場合、「プライマー伸長」という用語は「ポリメラーゼ伸長」とも称され、核酸ポリメラーゼ（例えば、DNAポリメラーゼ）によって触媒される反応であって、相補鎖と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）を、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せる反応を指す。

#### 【0262】

本明細書で使用される場合、「一意の分子識別子」または「UMI」という用語は、UMIが連結された各巨大分子（例えば、ペプチド）または結合性物質についての一意の識別子タグをもたらす、約3～約40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

10

20

30

40

50

、 26、27、28、29、30、31、32、33、34、35、36、37、38、39、または40塩基の長さの核酸分子を指す。巨大分子UMIは、複数の伸長記録タグからの配列決定データをコンピュータによりデコンボリューションして個々の巨大分子を起源とする伸長記録タグを識別するために使用することができる。結合性物質UMIは、特定の巨大分子に結合する個々の結合性物質それぞれを識別するために使用することができる。例えば、UMIを使用して、特定のペプチド分子に存在する単一のアミノ酸に特異的な結合性物質についての個々の結合事象の数を識別することができる。結合性物質または巨大分子に関してUMIおよびバーコードの両方に言及される場合、バーコードは、個々の結合性物質または巨大分子についてのUMI以外の識別情報（例えば、試料バーコード、コンパートメントバーコード、結合サイクルバーコード）を指すことが理解される。

10

#### 【0263】

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」または「ユニバーサルプライマー」または「ユニバーサルプライミング配列」という用語は、ライブラリー増幅のためおよび/または配列決定反応のために使用することができる核酸分子を指す。ユニバーサルプライミング部位としては、これだけに限定されないが、PCR増幅のためのプライミング部位（プライマー配列）、一部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なオリゴヌクレオチドとアニーリングするフローセルアダプター配列、配列決定プライミング部位、またはこれらの組合せを挙げることができる。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併せて一般に使用されるものを含めた他の型の増幅のため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、ローリングサークル増幅にユニバーサルプライミング部位を使用して、配列決定鋳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ら、2009年、Science、327巻：78～81頁）。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配列決定することができる（Korlachら、2008年、Proc. Natl. Acad. Sci.、105巻：1176～1181頁）。「フォワード」という用語は、「ユニバーサルプライミング部位」または「ユニバーサルプライマー」に関連して使用される場合、「5'」または「センス」と称される場合もある。「リバーズ」という用語は、ユニバーサルプライミング部位または「ユニバーサルプライマー」に関連して使用される場合は、「3'」または「アンチセンス」と称される場合もある。

20

30

#### 【0264】

本明細書で使用される場合、「伸長記録タグ」という用語は、結合性物質が巨大分子に結合した後に少なくとも1つの結合性物質のコーディングタグ（またはその相補配列）の情報が移行された記録タグを指す。コーディングタグの情報は、記録タグに直接移行させることもでき（例えば、ライゲーション）、間接的に移行させることもできる（例えば、プライマー伸長）。コーディングタグの情報は、記録タグに酵素的に移行させることもでき、化学的に移行させることもできる。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またはそれよりも多くのコーディングタグの結合性物質情報を含み得る。伸長記録タグの塩基配列は、コーディングタグによって識別される結合性物質の結合の時間的および逐次的順序を反映する場合もあり、コーディングタグによって識別される結合性物質の結合の部分的な逐次的順序を反映する場合もあり、コーディングタグによって識別される結合性物質の結合のいかなる順序も反映しない場合もある。ある特定の実施形態では、伸長記録タグ中に存在するコーディングタグ情報は、解析される巨大分子配列を少なくとも25%、30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、91%、92%、93%、94%、95%、96%、97%、98%、99%、また

40

50

は 100% の同一性で表す。解析される巨大分子配列が伸長記録タグによって 100% の同一性で表されないある特定の実施形態では、エラーは、結合性物質によるオフターゲットの結合、または「飛ばされた」結合サイクル（例えば、結合サイクル中に結合性物質が巨大分子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）、またはその両方に起因する可能性がある。

#### 【0265】

本明細書で使用される場合、「伸長コーディングタグ」という用語は、コーディングタグが接合している結合性物質の、記録タグが付随している巨大分子への結合後に少なくとも 1 つの記録タグ（またはその相補配列）の情報が移行されたコーディングタグを指す。記録タグの情報は、コーディングタグに直接移行させることもでき（例えば、ライゲーシ  
10 ョン）、間接的に移行させることもできる（例えば、プライマー伸長）。記録タグの情報は、酵素的に移行させることもでき、化学的に移行させることもできる。ある特定の実施形態では、伸長コーディングタグは、1 つの結合事象を反映する 1 つの記録タグの情報を  
20 含む。本明細書で使用される場合、「ジタグ」または「ジタグ構築物」または「ジタグ分子」という用語は、コーディングタグが接合している結合性物質の、記録タグが付随する巨大分子への結合後に少なくとも 1 つの記録タグ（またはその相補配列）の情報および少  
なくとも 1 つのコーディングタグ（またはその相補配列）が移行された核酸分子を指す（  
図 11B を参照されたい）。記録タグの情報およびコーディングタグは、ジタグに間接的  
に移行させることができる（例えば、プライマー伸長）。記録タグの情報は、酵素的に移  
行させることもでき、化学的に移行させることもできる。ある特定の実施形態では、ジ  
20 タグは、記録タグの UMI、記録タグのコンパートメントタグ、記録タグのユニバーサル  
プライミング部位、コーディングタグの UMI、コーディングタグのエンコーダー配列、結  
合サイクル特異的バーコード、コーディングタグのユニバーサルプライミング部位、ま  
たはそれらの任意の組合せを含む。

#### 【0266】

本明細書で使用される場合、「固体支持体」、「固体表面」、または「固体基板」または「基板」という用語は、巨大分子（例えば、ペプチド）を、共有結合性の相互作用および非共有結合性の相互作用またはそれらの任意の組合せを含めた当技術分野で公知の任意  
30 の手段によって直接または間接的に結び付けることができる、多孔質材料およ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、2 次元（例えば、平面）であっても  
よく、3 次元（例えば、ゲルマトリックスまたはビーズ）であってもよい。固体支持体は、これだけに限定されないが、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、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、信号変換電子機器を含めたバイオチップ、チャンネル、マイ  
40 クロタイターウェル、ELISA プレート、スピン干渉ディスク、ニトロセルロースメンブ  
レン、ニトロセルロースに基づくポリマー表面、ポリマーマトリックス、ナノ粒子、または  
マイクロスフェアを含めた任意の支持体表面であってよい。固体支持体用の材料として  
は、これだけに限定されないが、アクリルアミド、アガロース、セルロース、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、  
40 ポリメタクリレート、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、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、ポリ酸  
無水物、ポリグリコール酸、ポリ乳酸（poly lactic acid）、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、またはそれらの任意の組合せが挙げられる。固体支持体  
は、薄膜、膜、ピン、ディッシュ、繊維、織られた繊維、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、微小粒子、またはそれらの任意の組合せをさらに含む。  
例えば、固体表面がビーズである場合、ビーズは、これだけに限定されないが、セラミッ  
クビーズ、ポリスチレンビーズ、ポリマービーズ、メチルスチレンビーズ、アガロースビ  
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、または制御ポアビーズを含み得る。ビーズは、球状であっても不規則な形状であ  
50

てもよい。ビーズのサイズは、ナノメートル、例えば100nmから、ミリメートル、例えば1mmまでにわたり得る。ある特定の実施形態では、ビーズのサイズは、約0.2ミクロンから約200ミクロンまで、または約0.5ミクロンから約5ミクロンまでにわたる。一部の実施形態では、ビーズは、直径約1μm、約1.5μm、約2μm、約2.5μm、約2.8μm、約3μm、約3.5μm、約4μm、約4.5μm、約5μm、約5.5μm、約6μm、約6.5μm、約7μm、約7.5μm、約8μm、約8.5μm、約9μm、約9.5μm、約10μm、約10.5μm、約15μm、または約20μmであり得る。ある特定の実施形態では、「ビーズ (a bead)」固体支持体とは、個々のビーズを指す場合もあり、複数のビーズを指す場合もある。

【0267】

本明細書で使用される場合、「核酸分子」または「ポリヌクレオチド」という用語は、3'-5'リン酸ジエステル結合で連結されたデオキシリボヌクレオチドまたはリボヌクレオチドを含有する一本鎖または二本鎖ポリヌクレオチド、ならびにポリヌクレオチド類似体を指す。核酸分子としては、これだけに限定されないが、DNA、RNA、およびcDNAが挙げられる。ポリヌクレオチド類似体は、天然のポリヌクレオチドにおいて見出される標準のリン酸ジエステル連結以外の骨格、および任意選択で、修飾された糖部分またはリボースもしくはデオキシリボース以外の部分を有し得る。ポリヌクレオチド類似体は、標準のポリヌクレオチド塩基とワトソン・クリック塩基対合によって水素結合することが可能な塩基を含有し、ここで、類似体骨格は、当該塩基を、そのような水素結合がオリゴヌクレオチド類似体分子と標準のポリヌクレオチド内の塩基との間で配列特異的に可能になるように提示する。ポリヌクレオチド類似体の例としては、これだけに限定されないが、異種核酸 (xeno nucleic acid) (XNA)、架橋核酸 (BNA)、グリコール核酸 (GNA)、ペプチド核酸 (PNA)、PNA、モルホリノポリヌクレオチド、ロックド核酸 (LNA)、トレース核酸 (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、およびボロノホスフェートポリヌクレオチドが挙げられる。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、または、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、および芳香族トリアゾール類似体を含めた、任意の塩基と対合することができるユニバーサル塩基類似体、または、親和性結合のためのビオチン部分などの追加的な機能性を有する塩基類似体を含めたプリンまたはピリミジン類似体を有し得る。

【0268】

本明細書で使用される場合、「核酸配列決定」とは、核酸分子内または核酸分子の試料中のヌクレオチドの順序を決定することを意味する。

【0269】

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～数十億の分子を並行して配列決定することを可能にするハイスループットな配列決定法を指す。次世代シーケンシング法の例としては、合成による配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、ボロニーシーケンシング、イオン半導体シーケンシング、およ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板および核酸分子に対する相補配列に付着させることにより、核酸分子を固体基板にプライマーを介してハイブリダイズさせることができ、次いで、固体基板上の別個の領域においてポリメラーゼを使用することによって多数のコピーを生成させて、増幅させることができる（これらの群分けは、時にはポリメラーゼコロニーまたはボロニーと称される）。したがって、配列決定プロセス中、特定の位置にあるヌクレオチドについて多数回配列決定することができる（例えば、数百回または数千回）- このカバレッジの深さは、「ディープシーケンシング」と称される。ハイスループットな核酸配列決定技術の例としては、Service (Science, 311巻: 1544~1546頁, 2006年) によって概説されている通り、並行ビーズアレイ、合成による配列決定、ライゲーションによる配列決定、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、並行マイ

10

20

30

40

50

クロチップ、および単一分子アレイなどの形式を含めた、Illumina、BGI、Qiagen、Thermo-Fisher、およびRocheにより提供されるプラットフォームが挙げられる。

【0270】

本明細書で使用される場合、「単一分子シーケンシング」または「第3世代シーケンシング」とは、単一分子シーケンシング装置からの読み取りがDNAの単一分子の配列決定によって生成される次世代シーケンシング法を指す。段階的な手法で配列決定するために増幅に依拠して多くのDNA分子を並行してクローニングする次世代シーケンシング法とは異なり、単一分子シーケンシングでは、DNAの単一分子を調査し、増幅または同期化の必要はない。単一分子シーケンシングは、各塩基の組み入れ後に配列決定反応を休止する（「wash-and-scan」サイクル）必要のある方法、および読み取りステップ間の停止を必要としない方法を含む。単一分子シーケンシング法の例としては、単一分子リアルタイムシーケンシング（Pacific Biosciences）、ナノポアに基づく配列決定（Oxford Nanopore）、DI（duplex interrupted）ナノポアシーケンシング、および先端顕微鏡を使用したDNAのダイレクトイメージングが挙げられる。

10

【0271】

本明細書で使用される場合、巨大分子の「解析」とは、巨大分子の成分の全部または一部を数量化すること、特徴付けること、区別すること、またはこれらの組合せを意味する。例えば、ペプチド、ポリペプチド、またはタンパク質の解析は、ペプチドのアミノ酸配列（連続したまたは連続していない）の全部または一部を決定することを含む。巨大分子の解析は、巨大分子の成分の部分的な識別も含む。例えば、巨大分子タンパク質配列内のアミノ酸の部分的な識別により、タンパク質のアミノ酸を、可能性のあるアミノ酸のサブセットに属するものと識別することができる。解析は、一般にはn NTAAの解析で開始され、次いで、次のペプチドのアミノ酸（すなわち、n-1、n-2、n-3など）に進行する。これは、n NTAAが切断され、それにより、n-1ペプチドのアミノ酸がN末端アミノ酸に変わることによって実現される（本明細書では、「n-1 NTAA」と称される）。ペプチドの解析は、ペプチド上の翻訳後修飾の存在および頻度を決定することを含み得、これは、ペプチド上の翻訳後修飾の逐次的順序に関する情報が含む場合もあり、含まない場合もある。ペプチドの解析は、ペプチド内のエピトープの逐次的順序または位置に関する情報を含む場合もあり、含まない場合もある、ペプチド内のエピトープの存在および頻度を決定することを含み得る。ペプチドの解析は、異なる型の解析を組み合わせること、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、またはそれらの任意の組合せを得ることを含み得る。

20

30

【0272】

本明細書で使用される場合、「コンパートメント」という用語は、巨大分子のサブセットを巨大分子の試料から分離または隔離する物理的な領域または容積を指す。例えば、コンパートメントにより、個々の細胞を他の細胞から分離すること、または試料のプロテオームのサブセットを試料の残りのプロテオームから分離することができる。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、プレート上のピコタイターウェルまたはマイクロタイターウェル、チューブ、バイアル、ゲルビーズ）、または表面上の分離された領域であってよい。コンパートメントは、巨大分子を固定化することができる1つまたは複数のビーズを含んでよい。

40

【0273】

本明細書で使用される場合、「コンパートメントタグ」または「コンパートメントバーコード」という用語は、1つまたは複数のコンパートメント（例えば、マイクロ流体液滴）内の、構成物（例えば、単一の細胞のプロテオーム）に関する識別情報を含む、約4塩基～約100塩基（4塩基、100塩基、およびその間の任意の整数を含む）の一本鎖または二本鎖核酸分子を指す。コンパートメントバーコードにより、試料中の巨大分子のサブセット、例えば、複数（例えば、数百万～数十億）のコンパートメントから同じ物理的

50



コンパートメントまたはコンパートメントの群に分離されたタンパク質試料のサブセットが識別される。したがって、コンパートメントタグを使用して、同じコンパートメントタグを有する1つまたは複数のコンパートメントに由来する構成物と、異なるコンパートメントタグを有する別のコンパートメント内の構成物を、これらの構成物が一緒にプールされた後であっても区別することができる。各コンパートメント内または2つもしくはそれよりも多くのコンパートメントの群内のタンパク質および/またはペプチドを一意的コンパートメントタグで標識することにより、個々のコンパートメントまたはコンパートメントの群内の同じタンパク質、タンパク質複合体、または細胞に由来するペプチドを識別することができる。コンパートメントタグは、任意選択で片側またはその両側にスパーサー配列が隣接するバーコード、および任意選択のユニバーサルプライマーを含む。スパーサー配列は、記録タグのスパーサー配列に相補的なものであってよく、それにより、コンパートメントタグ情報の記録タグへの移行が可能になる。コンパートメントタグは、特に、コンパートメントタグが、本明細書に記載の下流のペプチド解析方法に使用される記録タグを含む実施形態に関しては、ユニバーサルプライミング部位、一意的分子識別子（それに付着したペプチドに関する識別情報をもたらす）、またはその両方も含んでよい。コンパートメントタグは、ペプチドとのカップリングのための機能的部分（例えば、アルデヒド、NHS、mTet、アルキンなど）を含んでよい。あるいは、コンパートメントタグは、コンパートメントタグと目的のペプチドのライゲーションを可能にするためのタンパク質リガーゼの認識配列を含むペプチドを含んでよい。コンパートメントは、単一のコンパートメントタグ、任意選択のUMI配列を保存する複数の同一のコンパートメントタグ、または2種またはそれよりも多くの異なるコンパートメントタグを含んでよい。ある特定の実施形態では、各コンパートメントは、一意的コンパートメントタグを含む（1対1のマッピング）。他の実施形態では、より大きなコンパートメントの集団からの多数のコンパートメントは、同じコンパートメントタグを含む（多対1のマッピング）。コンパートメントタグは、コンパートメント内の固体支持体（例えば、ビーズ）に接合していてもよく、コンパートメント自体の表面（例えば、ピコタイターウェルの表面）に接合していてもよい。あるいは、コンパートメントタグは、コンパートメント内の溶液中に遊離していてもよい。

#### 【0274】

本明細書で使用される場合、「分配」という用語は、試料内の巨大分子の集団に由来する巨大分子の亜集団への一意的バーコードのランダムな割り当てを指す。ある特定の実施形態では、分配は、巨大分子をコンパートメントに区分することによって実現することができる。分配は、単一のコンパートメント内の巨大分子で構成されるものであってよく、コンパートメントの集団に由来する多数のコンパートメント内の巨大分子で構成されるものであってよい。

#### 【0275】

本明細書で使用される場合、「分配タグ」または「分配バーコード」とは、分配に関する識別情報を含む、約4塩基〜約100塩基（4塩基、100塩基、およびその間の任意の整数を含む）の一本鎖または二本鎖核酸分子を指す。ある特定の実施形態では、巨大分子に関する分配タグは、巨大分子を同じバーコードで標識されたコンパートメント（複数可）に分配することにより生じる同一のコンパートメントタグを指す。

#### 【0276】

本明細書で使用される場合、「画分」という用語は、サイズ、疎水性、等電点、親和性などによる分画などの物理的または化学的分離方法を使用して残りの試料または細胞小器官から選別された試料内の巨大分子のサブセット（例えば、タンパク質）を指す。分離方法としては、HPLC分離、ゲル分離、アフィニティー分離、細胞分画、細胞小器官分画、組織分画などが挙げられる。流体の流れ、磁性、電流、質量、密度などの物理特性も分離のために使用することができる。

#### 【0277】

本明細書で使用される場合、「画分バーコード」という用語は、画分内の巨大分子に関

10

20

30

40

50

する識別情報を含む、約 4 塩基～約 1 0 0 塩基（4 塩基、1 0 0 塩基、およびその間の任意の整数を含む）の一本鎖または二本鎖核酸分子を指す。

### ⅠⅠⅠ．巨大分子の解析方法

#### 【0278】

本明細書に記載の方法は、巨大分子解析のための、高度に並行化された手法を提供する。高度に多重化された巨大分子結合アッセイを、次世代シーケンシングによる読み取りのために核酸分子ライブラリーに変換する。本明細書において提示される方法は、タンパク質またはペプチド配列決定に特に有用である。

#### 【0279】

好ましい実施形態では、タンパク質試料を、バーコード（例えば、試料バーコード、コンパートメントバーコード）および任意選択の一意の分子識別子を含む少なくとも 1 つの核酸記録タグを用いて単一分子レベルで標識する。タンパク質試料にタンパク質分解による消化を行って記録タグで標識されたペプチドの集団（例えば、数百万～数十億）を生成する。これらの記録タグで標識されたペプチドをプールし、固体支持体（例えば、多孔質ビーズ）上にランダムに固定化する。プールされ、固定化された、記録タグで標識されたペプチドを多数の連続的な結合サイクルに供し、各結合サイクルは、付随する結合性物質を識別するエンコーダー配列を含むコーディングタグで標識された複数の結合性物質（例えば、天然に存在するアミノ酸 20 種全てに対する結合性物質）に曝露することを含む。各結合サイクル中、結合性物質のペプチドへの結合に関する情報を、結合性物質のコーディングタグ情報を記録タグに移行させること（または記録タグ情報をコーディングタグに移行させることもしくは記録タグ情報およびコーディングタグ情報の両方を別のジタグ構築物に移行させること）によって捕捉する。結合サイクルが完了したら、アッセイされたペプチドの結合履歴を表す伸長記録タグ（または伸長コーディングタグまたはジタグ構築物）のライブラリーを生成し、これを、非常にハイスループットの次世代デジタルシーケンシング法を使用して解析することができる。記録タグに核酸バーコードを使用することにより、例えば、ペプチド配列の起源である試料、細胞、プロテオームのサブセット、またはタンパク質を識別するために、大量のペプチド配列決定データをデコンボリューションすることが可能になる。

#### 【0280】

一態様では、巨大分子を解析するための方法であって、（a）巨大分子および固体支持体に接合した付随するまたは共局在する記録タグを用意するステップと；（b）巨大分子を、巨大分子に結合することが可能な第 1 の結合性物質であって、第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；（c）第 1 のコーディングタグの情報を記録タグに移行させて、一次伸長記録タグを生成するステップと；（d）巨大分子を、巨大分子に結合することが可能な第 2 の結合性物質であって、第 2 の結合性物質に関する識別情報を有する第 2 のコーディングタグを含む第 2 の結合性物質と接触させるステップと；（e）第 2 のコーディングタグの情報を一次伸長記録タグに移行させて、二次伸長記録タグを生成するステップと；（f）二次伸長タグを解析するステップと（例えば、図 2 A～D を参照されたい）を含む方法が提供される。

#### 【0281】

ある特定の実施形態では、接触させるステップ（b）および（d）を逐次的に実施する、例えば、第 1 の結合性物質と第 2 の結合性物質を、別々の結合サイクル反応で巨大分子と接触させる。他の実施形態では、接触させるステップ（b）および（d）を同時に、例えば、第 1 の結合性物質、第 2 の結合性物質、および任意選択で追加的な結合性物質を含む単一の結合サイクル反応などで実施する。好ましい実施形態では、接触させるステップ（b）および（d）は、それぞれ、巨大分子を複数の結合性物質と接触させることを含む。

#### 【0282】

ある特定の実施形態では、方法は、ステップ（e）と（f）の間に、（x）第 2 の結合性物質を巨大分子に結合することが可能な第 3 の（またはより高次の）結合性物質であっ

10

20

30

40

50

て、第3の（またはより高次の）結合性物質に関する識別情報を有する第3の（またはより高次の）コーディングタグを含む第3の（またはより高次の）結合性物質に置き換えることにより、ステップ（d）および（e）を1回または複数回繰り返すステップと；（y）第3の（またはより高次の）コーディングタグの情報を第2の（またはより高次の）伸長記録タグに移行させて、第3の（またはより高次の）伸長記録タグを生成するステップと；（z）第3の（またはより高次の）伸長記録タグを解析するステップとをさらに含む。  
【0283】

第3の（またはより高次の）結合性物質を第1の結合性物質および第2の結合性物質とは別の結合サイクル反応で巨大分子と接触させることができる。あるいは、第3の（またはより高次の）結合性物質を第1の結合性物質、および第2の結合性物質と一緒に単一の結合サイクル反応で巨大分子と接触させることができる。

10

【0284】

第2の態様では、巨大分子を解析するための方法であって、（a）固体支持体に接合した巨大分子、付随する第1の記録タグおよび付随する第2の記録タグを用意するステップと；（b）巨大分子を、巨大分子に結合することが可能な第1の結合性物質であって、第1の結合性物質に関する識別情報を有する第1のコーディングタグを含む第1の結合性物質と接触させるステップと；（c）第1のコーディングタグの情報を第1の記録タグに移行させて、第1の伸長記録タグを生成するステップと；（d）巨大分子を、巨大分子に結合することが可能な第2の結合性物質であって、第2の結合性物質に関する識別情報を有する第2のコーディングタグを含む第2の結合性物質と接触させるステップと；（e）第2のコーディングタグの情報を第2の記録タグに移行させて、第2の伸長記録タグを生成するステップと；（f）第1の伸長記録タグおよび第2の伸長記録タグを解析するステップとを含む方法が提供される。

20

【0285】

ある特定の実施形態では、接触させるステップ（b）および（d）を逐次的に実施する、例えば、第1の結合性物質および第2の結合性物質を別々の結合サイクル反応で巨大分子と接触させる。他の実施形態では、接触させるステップ（b）および（d）を同時に、例えば、第1の結合性物質、第2の結合性物質、および任意選択で追加的な結合性物質を含む単一の結合サイクル反応などで実施する。

【0286】

ある特定の実施形態では、ステップ（a）は、固体支持体に接合した付随する第3の（またはより高次の）記録タグを用意するステップをさらに含む。さらなる実施形態では、方法は、ステップ（e）と（f）の間に、（x）第2の結合性物質を巨大分子に結合することが可能な第3の（またはより高次の）結合性物質であって、第3の（またはより高次の）結合性物質に関する識別情報を有する第3の（またはより高次の）コーディングタグを含む第3の（またはより高次の）結合性物質に置き換えることにより、ステップ（d）および（e）を1回または複数回繰り返すステップと；（y）第3の（またはより高次の）コーディングタグの情報を第3の（またはより高次の）記録タグに移行させて、第3の（またはより高次の）伸長記録タグを生成するステップと；（z）第1の伸長記録タグ、第2の伸長記録タグおよび第3の（またはより高次の）伸長記録タグを解析するステップとをさらに含む。

30

40

【0287】

第3の（またはより高次の）結合性物質を第1の結合性物質および第2の結合性物質とは別の結合サイクル反応で巨大分子と接触させることができる。あるいは、第3の（またはより高次の）結合性物質を第1の結合性物質、および第2の結合性物質と一緒に単一の結合サイクル反応で巨大分子と接触させることができる。

【0288】

ある特定の実施形態では、第1のコーディングタグ、第2のコーディングタグ、および任意のより高次のコーディングタグはそれぞれ、結合サイクル特異的配列を有する。

【0289】

50

第3の態様では、ペプチドを解析する方法であって、(a) 固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；(b) ペプチドのN末端アミノ酸(N T A A)を化学的部分で修飾して、修飾されたN T A Aを生成するステップと；(c) ペプチドを、修飾されたN T A Aに結合することが可能な第1の結合性物質であって、第1の結合性物質に関する識別情報を有する第1のコーディングタグを含む第1の結合性物質と接触させるステップと；(d) 第1のコーディングタグの情報を記録タグに移行させて、伸長記録タグを生成するステップと；(e) 伸長記録タグを解析するステップと(例えば、図3を参照されたい)を含む方法が提供される。

#### 【0290】

ある特定の実施形態では、ステップ(c)が、前記ペプチドを、前記第2の(またはより高次の)結合性物質に関する識別情報を有する第2の(またはより高次の)コーディングタグを含む第2の(またはより高次の)結合性物質であって、ステップ(b)の前記修飾されたN T A A以外の修飾されたN T A Aに結合することが可能である第2の(またはより高次の)結合性物質と接触させることをさらに含む。さらなる実施形態では、ペプチドの第2の(またはより高次の)結合性物質との接触を、ペプチドの第1の結合性物質との接触後に逐次的に行う、例えば、第1の結合性物質および第2の(またはより高次の)結合性物質を別々の結合サイクル反応でペプチドと接触させる。他の実施形態では、ペプチドの第2の(またはより高次の)結合性物質との接触を、ペプチドの第1の結合性物質との接触と同時に、例えば、第1の結合性物質および第2の(またはより高次の)結合性物質を含む単一の結合サイクル反応などで行う。

#### 【0291】

ある特定の実施形態では、化学的部分をN T A Aに化学反応または酵素反応を介して付加する。

#### 【0292】

ある特定の実施形態では、N T A Aを修飾するために使用される化学的部分は、フェニルチオカルバモイル(P T C)、ジニトロフェノール(D N P)部分；スルホニルオキシニトロフェニル(S N P)部分、ダンシル部分；7-メトキシクマリン部分；チオアシル部分；チオアセチル部分；アセチル部分；グアニジニル部分；またはチオベンジル部分である。

#### 【0293】

化学的部分は、化学薬剤を使用してN T A Aに付加することができる。ある特定の実施形態では、N T A AをP T C部分で修飾するための化学薬剤は、フェニルイソチシアネートまたはその誘導体である；N T A AをD N P部分で修飾するための化学薬剤は、2, 4-ジニトロベンゼンスルホン酸(D N B S)または1-フルオロ-2, 4-ジニトロベンゼン(D N F B)などのハロゲン化アリールである；N T A Aをスルホニルオキシニトロフェニル(S N P)部分で修飾するための化学薬剤は、4-スルホニル-2-ニトロフルオロベンゼン(S N F B)である；N T A Aをダンシル基で修飾するための化学薬剤は、ダンシルクロリドなどのスルホニルクロリドである；N T A Aを7-メトキシクマリン部分で修飾するための化学薬剤は、7-メトキシクマリン酢酸(M C A)である；N T A Aをチオアシル部分で修飾するための化学薬剤は、チオアシル化試薬である；N T A Aをチオアセチル部分で修飾するための化学薬剤は、チオアセチル化試薬である；N T A Aをアセチル部分で修飾するための化学薬剤は、アセチル化試薬(例えば、無水酢酸)である；N T A Aをグアニジニル(アミジニル)部分で修飾するための化学薬剤は、グアニジニル化試薬である、または、N T A Aをチオベンジル部分で修飾するための化学薬剤は、チオベンジル化試薬である。

#### 【0294】

第4の態様では、本開示は、ペプチドを解析するための方法であって、(a) 固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；(b) ペプチドのN末端アミノ酸(N T A A)を化学的部分で修飾して、修飾されたN T A Aを生成するステップと；(c) ペプチドを、修飾されたN T A Aに結合することが可能な第1の結合性

物質であって、第1の結合性物質に関する識別情報を有する第1のコーディングタグを含む第1の結合性物質と接触させるステップと；(d)第1のコーディングタグの情報を記録タグに移行させて、第1の伸長記録タグを生成するステップと；(e)修飾されたNTAAを除去して、新しいNTAAを露出させるステップと；(f)ペプチドの新しいNTAAを化学的部分で修飾して、新しく修飾されたNTAAを生成するステップと；(g)ペプチドを、新しく修飾されたNTAAに結合することが可能な第2の結合性物質であって、第2の結合性物質に関する識別情報を有する第2のコーディングタグを含む第2の結合性物質と接触させるステップと；(h)第2のコーディングタグの情報を第1の伸長記録タグに移行させて、第2の伸長記録タグを生成するステップと；(i)第2の伸長記録タグを解析するステップとを含む方法を提供する。

10

## 【0295】

ある特定の実施形態では、接触させるステップ(c)および(g)を逐次的に実施する、例えば、第1の結合性物質および第2の結合性物質を別々の結合サイクル反応でペプチドと接触させる。

## 【0296】

ある特定の実施形態では、方法は、ステップ(h)と(i)の間に、(x)第2の結合性物質を修飾されたNTAAに結合することが可能な第3の(またはより高次の)結合性物質であって、第3の(またはより高次の)結合性物質に関する識別情報を有する第3の(またはより高次の)コーディングタグを含む第3の(またはより高次の)結合性物質に置き換えることにより、ステップ(e)、(f)および(g)を1回または複数回繰り返すステップと；(y)第3の(またはより高次の)コーディングタグの情報を第2の(またはより高次の)伸長記録タグに移行させて、第3の(またはより高次の)伸長記録タグを生成するステップと；(z)第3の(またはより高次の)伸長記録タグを解析するステップとをさらに含む。

20

## 【0297】

ある特定の実施形態では、化学的部分をNTAAに化学反応または酵素反応を介して付加する。

## 【0298】

ある特定の実施形態では、化学的部分は、フェニルチオカルバモイル(PTC)、ジニトロフェノール(DNP)部分；スルホニルオキシニトロフェニル(SNP)部分、ダンシル部分；7-メトキシクマリン部分；チオアシル部分；チオアセチル部分；アセチル部分；グアニル部分；またはチオベンジル部分である。

30

## 【0299】

化学的部分は、化学薬剤を使用してNTAAに付加することができる。ある特定の実施形態では、NTAAをPTC部分で修飾するための化学薬剤は、フェニルイソチオシアネートまたはその誘導体である；NTAAをDNP部分で修飾するための化学薬剤は、2,4-ジニトロベンゼンスルホン酸(DNBS)または1-フルオロ-2,4-ジニトロベンゼン(DNFB)などのハロゲン化アリールである；NTAAをスルホニルオキシニトロフェニル(SNP)部分で修飾するための化学薬剤は、4-スルホニル-2-ニトロフルオロベンゼン(SNFB)である；NTAAをダンシル基で修飾するための化学薬剤は、ダンシルクロリドなどのスルホニルクロリドである；NTAAを7-メトキシクマリン部分で修飾するための化学試薬は、7-メトキシクマリン酢酸(MCA)である；NTAAをチオアシル部分で修飾するための化学薬剤は、チオアシル化試薬である；NTAAをチオアセチル部分で修飾するための化学薬剤は、チオアセチル化試薬である；NTAAをアセチル部分で修飾するための化学薬剤は、アセチル化剤(例えば、無水酢酸)である；NTAAをグアニル部分で修飾するための化学薬剤は、グアニジニル化試薬である、またはNTAAをチオベンジル部分で修飾するための化学薬剤は、チオベンジル化試薬である。

40

## 【0300】

第5の態様では、ペプチドを解析するための方法であって、(a)固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；(b)ペプチドを、ペプチド

50

のN末端アミノ酸（NTAA）に結合することが可能な第1の結合性物質であって、第1の結合性物質に関する識別情報を有する第1のコーディングタグを含む第1の結合性物質と接触させるステップと；（c）第1のコーディングタグの情報を記録タグに移行させて、伸長記録タグを生成するステップと；（d）伸長記録タグを解析するステップとを含む方法が提供される。

#### 【0301】

ある特定の実施形態では、ステップ（b）が、前記ペプチドを、前記第2の（またはより高次の）結合性物質に関する識別情報を有する第2の（またはより高次の）コーディングタグを含む第2の（またはより高次の）結合性物質であって、前記ペプチドの前記NTAA以外のNTAAに結合することが可能な第2の（またはより高次の）結合性物質と接触させることをさらに含む。さらなる実施形態では、ペプチドの第2の（またはより高次の）結合性物質との接触を、ペプチドの第1の結合性物質との接触後に逐次的に行う、例えば、第1の結合性物質および第2の（またはより高次の）結合性物質を別々の結合サイクル反応でペプチドと接触させる。他の実施形態では、ペプチドの第2の（またはより高次の）結合性物質との接触を、ペプチドを第1の結合性物質と接触させるのと同時に、例えば、第1の結合性物質および第2の（またはより高次の）結合性物質を含む単一の結合サイクル反応などで行う。

#### 【0302】

第6の態様では、ペプチドを解析するための方法であって、（a）固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；（b）ペプチドを、ペプチドのN末端アミノ酸（NTAA）に結合することが可能な第1の結合性物質であって、第1の結合性物質に関する識別情報を有する第1のコーディングタグを含む第1の結合性物質と接触させるステップと；（c）第1のコーディングタグの情報を記録タグに移行させて、第1の伸長記録タグを生成するステップと；（d）NTAAを除去して、ペプチドの新しいNTAAを露出させるステップと；（e）ペプチドを、新しいNTAAに結合することが可能な第2の結合性物質であって、第2の結合性物質に関する識別情報を有する第2のコーディングタグを含む第2の結合性物質と接触させるステップと；（f）第2のコーディングタグの情報を第1の伸長記録タグに移行させて、第2の伸長記録タグを生成するステップと；（g）第2の伸長記録タグを解析するステップとを含む方法が提供される。

#### 【0303】

ある特定の実施形態では、方法は、ステップ（f）と（g）の間に、（x）第2の結合性物質を巨大分子に結合することが可能な第3の（またはより高次の）結合性物質であって、第3の（またはより高次の）結合性物質に関する識別情報を有する第3の（またはより高次の）コーディングタグを含む第3の（またはより高次の）結合性物質に置き換えることにより、ステップ（d）、（e）および（f）を1回または複数回繰り返すステップと；（y）第3の（またはより高次の）コーディングタグの情報を第2の（またはより高次の）伸長記録タグに移行させて第3の（またはより高次の）伸長記録タグを生成するステップとをさらに含み、ステップ（g）において第3の（またはより高次の）伸長記録タグを解析する。

#### 【0304】

ある特定の実施形態では、接触させるステップ（b）および（e）を逐次的に実施する、例えば、第1の結合性物質および第2の結合性物質を別々の結合サイクル反応でペプチドと接触させる。

#### 【0305】

本明細書において提示される実施形態のいずれかでは、方法は、複数の巨大分子を並行して解析することを含む。好ましい実施形態では、方法は、複数のペプチドを並行して解析することを含む。

#### 【0306】

本明細書において提示される実施形態のいずれかでは、巨大分子（またはペプチド）を結合性物質と接触させるステップは、巨大分子（またはペプチド）を複数の結合性物質と

10

20

30

40

50

接触させることを含む。

【0307】

本明細書において提示される実施形態のいずれかでは、巨大分子は、タンパク質、ポリペプチド、またはペプチドであってよい。さらなる実施形態では、ペプチドは、生体試料に由来するタンパク質またはポリペプチドを断片化することによって得ることができる。

【0308】

本明細書において提示される実施形態のいずれかでは、巨大分子は、炭水化物、脂質、核酸、もしくは大環状分子であるまたはそれを含んでよい。

【0309】

本明細書において提示される実施形態のいずれかでは、記録タグは、DNA分子、修飾塩基を有するDNA分子、RNA分子、BNA分子、XNA分子、LNA分子、PNA分子、PNA分子(Dragulescu-Andrasiら、2006年、J. Am. Chem. Soc.、128巻：10258～10267頁)、GNA分子、またはそれらの任意の組合せであってよい。

10

【0310】

本明細書において提示される実施形態のいずれかでは、記録タグは、ユニバーサルプライミング部位を含んでよい。さらなる実施形態では、ユニバーサルプライミング部位は、増幅、ライゲーション、配列決定、またはこれらの組合せのためのプライミング部位を含む。

【0311】

本明細書において提示される実施形態のいずれかでは、記録タグは、一意の分子識別子、コンパートメントタグ、分配バーコード、試料バーコード、画分バーコード、スペーサー配列、またはそれらの任意の組合せを含んでよい。

20

【0312】

本明細書において提示される実施形態のいずれかでは、コーディングタグは、一意の分子識別子(UMI)、エンコーダー配列、結合サイクル特異的配列、スペーサー配列、またはそれらの任意の組合せを含んでよい。

【0313】

本明細書において提示される実施形態のいずれかでは、コーディングタグ内の結合サイクル特異的配列は、結合サイクル特異的スペーサー配列であってよい。

30

【0314】

ある特定の実施形態では、結合サイクル特異的配列は、エンコーダー配列とは別のバーコードとしてコードされる。他の実施形態では、エンコーダー配列および結合サイクル特異的配列は、結合性物質に対しておよび各結合サイクルに対して一意である単一のバーコードに記載される。

【0315】

ある特定の実施形態では、スペーサー配列は、多数の結合サイクルからの結合性物質の間で共有される共通の結合サイクル配列を含む。他の実施形態では、スペーサー配列は、同じ結合サイクルからの結合性物質の間で共有される一意の結合サイクル配列を含む。

【0316】

本明細書において提示される実施形態のいずれかでは、記録タグは、バーコードを含んでよい。

40

【0317】

本明細書において提示される実施形態のいずれかでは、巨大分子および付随する記録タグ(複数可)は、固体支持体に共有結合により接合してよい。

【0318】

本明細書において提示される実施形態のいずれかでは、固体支持体は、ビーズ、多孔質ビーズ、多孔質マトリックス、拡張可能なゲルビーズまたはマトリックス、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、ナイロン、シリコンウェーハチップ、フロースルーチップ、信号変換電子機器を含むバイオチップ、マイクロタイタ

50

ーウェル、E L I S A プレート、スピン干渉ディスク、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアであってよい。

【0319】

本明細書において提示される実施形態のいずれかでは、固体支持体は、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズであってよい。

【0320】

本明細書において提示される実施形態のいずれかでは、複数の巨大分子および付随する記録タグを固体支持体に接合することができる。さらなる実施形態では、複数の巨大分子の間に固体支持体上で平均距離 50 nm、100 nm、または 200 nm の間隔をあける。

10

【0321】

本明細書において提示される実施形態のいずれかでは、結合性物質は、ポリペプチドまたはタンパク質であり得る。さらなる実施形態では、結合性物質は、改変もしくはバリエーションアミノペプチダーゼ、改変もしくはバリエーションアミノアシル tRNA 合成酵素、改変もしくはバリエーションアンチカリン、または改変もしくはバリエーション ClpS である。

【0322】

本明細書において提示される実施形態のいずれかでは、結合性物質は、巨大分子に選択的に結合することが可能なものであってよい。

【0323】

20

本明細書において提示される実施形態のいずれかでは、コーディングタグは、DNA 分子、修飾塩基を有する DNA 分子、RNA 分子、BNA 分子、XNA 分子、LNA 分子、GNA 分子、PNA 分子、PNA 分子、またはこれらの組合せであってよい。

【0324】

本明細書において提示される実施形態のいずれかでは、結合性物質とコーディングタグは、リンカーによって接合されていてよい。

【0325】

本明細書において提示される実施形態のいずれかでは、結合性物質とコーディングタグは、SpyTag/SpyCatcher または SnoopTag/SnoopCatcher ペプチド-タンパク質対（それぞれ、その全体が参照により組み込まれる、Zakeri ら、2012 年、Proc Natl Acad Sci U S A、109 巻（12 号）：E690～697 頁；Veggiari ら、2016 年、Proc. Natl. Acad. Sci. USA、113 巻：1202～1207 頁）によって接合されていてよい。

30

【0326】

本明細書において提示される実施形態のいずれかでは、コーディングタグの情報の記録タグへの移行は、DNA リガーゼによって媒介される。あるいは、コーディングタグの情報の記録タグへの移行は、DNA ポリメラーゼまたは化学的ライゲーションによって媒介される。

【0327】

40

本明細書において提示される実施形態のいずれかでは、伸長記録タグを解析するステップは、核酸配列決定を含んでよい。さらなる実施形態では、核酸配列決定は、合成による配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、ポロニーシーケンシング、イオン半導体シーケンシング、またはパイロシーケンシングである。他の実施形態では、核酸配列決定は、単一分子リアルタイムシーケンシング、ナノポアに基づく配列決定、ナノギャップトンネリングシーケンシング、または先端顕微鏡を使用した DNA のダイレクトイメージングである。

【0328】

本明細書において提示される実施形態のいずれかでは、伸長記録タグを解析前に増幅させることができる。

50



## 【0329】

本明細書において提示される実施形態のいずれかでは、伸長記録タグに含有されるコーディングタグ情報の順序が、結合性物質による巨大分子への結合の順序に関する情報、したがって、結合性物質によって検出される分析物の配列を提供することができる。

## 【0330】

本明細書において提示される実施形態のいずれかでは、伸長記録タグに含有される特定のコーディングタグ情報（例えば、エンコーダー配列）の頻度は、特定の結合性物質による巨大分子への結合の頻度、したがって、結合性物質によって検出される巨大分子内の分析物の頻度に関する情報を提供することができる。

## 【0331】

本明細書に開示されている実施形態のいずれかでは、多数の巨大分子（例えば、タンパク質）試料であって、各試料内の巨大分子の集団が試料特異的バーコードを含む記録タグで標識されている試料をプールすることができる。そのような巨大分子試料のプールを単一の反応チューブ内での結合サイクルに供することができる。

## 【0332】

本明細書において提示される実施形態のいずれかでは、複数の巨大分子を表す複数の伸長記録タグを並行して解析することができる。

## 【0333】

本明細書において提示される実施形態のいずれかでは、複数の巨大分子を表す複数の伸長記録タグを多重化アッセイで解析することができる。

## 【0334】

本明細書において提示される実施形態のいずれかでは、複数の伸長記録タグに対して、解析前に標的濃縮アッセイを行うことができる。

## 【0335】

本明細書において提示される実施形態のいずれかでは、複数の伸長記録タグに対して、解析前にサブトラクションアッセイを行うことができる。

## 【0336】

本明細書において提示される実施形態のいずれかでは、複数の伸長記録タグに対して、極めて豊富な種を減少させるために解析前に正規化アッセイを行うことができる。

## 【0337】

本明細書において提示される実施形態のいずれかでは、NTAAを、改変アミノペプチダーゼ、改変アミノ酸tRNA合成酵素、穏やかなエドマン分解、エドマナーゼ（Edmanase）酵素、または無水TFAによって除去することができる。

## 【0338】

本明細書において提示される実施形態のいずれかでは、少なくとも1つの結合性物質は、末端アミノ酸残基に結合し得る。ある特定の実施形態では、末端アミノ酸残基は、N末端アミノ酸またはC末端アミノ酸である。

## 【0339】

本明細書に記載の実施形態のいずれかでは、少なくとも1つの結合性物質は、翻訳後修飾されたアミノ酸に結合し得る。

## 【0340】

上述の実施形態の特徴を以下の節でさらに詳細に提示する。

## I V . 巨大分子

## 【0341】

一態様では、本開示は、巨大分子の解析に関する。巨大分子は、より小さなサブユニットで構成される大きな分子である。ある特定の実施形態では、巨大分子は、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、またはキメラ巨大分子である。

## 【0342】

本明細書に開示されている方法に従って解析される巨大分子（例えば、タンパク質、ポ

10

20

30

40

50

リペプチド、ペプチド)は、これだけに限定されないが、細胞(初代細胞および培養細胞株の両方)、細胞溶解物または抽出物、エキソソームを含めた細胞小器官または小胞、組織および組織抽出物などの生体試料;生検材料;排泄物;事実上あらゆる生物体の体液(例えば、血液、全血、血清、血漿、尿、リンパ液、胆汁、脳脊髄液、間質液、眼房水または硝子体液、初乳、痰、羊水、唾液、肛門および膣分泌物、汗および精液、漏出液、滲出液(例えば、膿瘍もしくは任意の他の感染もしくは炎症の部位から得られる流体)または関節(正常な関節もしくは関節リウマチ、変形性関節症、痛風もしくは化膿性関節炎などの疾患の影響を受けている関節)から得られる流体)(マイクロバイームを含有する試料を含めた哺乳動物由来試料であることが好ましく、マイクロバイームを含有する試料を含めたヒト由来試料であることが特に好ましい);環境試料(例えば、空気、農業、水および土壌試料など);微生物バイオフィームおよび/または微生物群、ならびに微生物の胞子に由来する試料を含めた微生物試料;細胞外液、細胞培養物からの細胞外上清、細菌内の封入体、ミトコンドリア区画を含めた細胞区画、および細胞ペリプラズムを含めた研究試料を含めた適切な供給源または試料から得ることができる。

10

#### 【0343】

ある特定の実施形態では、巨大分子は、タンパク質、タンパク質複合体、ポリペプチド、またはペプチドである。ペプチド、ポリペプチド、またはタンパク質のアミノ酸配列情報および翻訳後修飾を、次世代シーケンシング法によって解析することができる核酸コードライブラリーに変換する。ペプチドは、L-アミノ酸、D-アミノ酸、またはその両方を含んでよい。ペプチド、ポリペプチド、タンパク質、またはタンパク質複合体は、標準の、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、またはそれらの任意の組合せを含んでよい。一部の実施形態では、ペプチド、ポリペプチド、またはタンパク質は、天然に存在するもの、合成的に作製されたもの、または組換えによって発現させたものである。上述のペプチド実施形態のいずれかでは、ペプチド、ポリペプチド、タンパク質、またはタンパク質複合体は、翻訳後修飾をさらに含んでよい。

20

#### 【0344】

標準の、天然に存在するアミノ酸としては、アラニン(AまたはAla)、システイン(CまたはCys)、アスパラギン酸(DまたはAsp)、グルタミン酸(EまたはGlu)、フェニルアラニン(FまたはPhe)、グリシン(GまたはGly)、ヒスチジン(HまたはHis)、イソロイシン(IまたはIle)、リシン(KまたはLys)、ロイシン(LまたはLeu)、メチオニン(MまたはMet)、アスパラギン(NまたはAsn)、プロリン(PまたはPro)、グルタミン(QまたはGln)、アルギニン(RまたはArg)、セリン(SまたはSer)、トレオニン(TまたはThr)、バリン(VまたはVal)、トリプトファン(WまたはTrp)、およびチロシン(YまたはTyr)が挙げられる。非標準アミノ酸としては、セレノシステイン、ピロリジン、およびN-ホルミルメチオニン、 $\alpha$ -アミノ酸、ホモアミノ酸、プロリンおよ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およびチロシン誘導体、直鎖コアアミノ酸、およびN-メチルアミノ酸が挙げられる。

30

#### 【0345】

ペプチド、ポリペプチド、またはタンパク質の翻訳後修飾(PTM)は、共有結合性修飾であっても酵素的修飾であってもよい。翻訳後修飾の例としては、これだけに限定されないが、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、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、ジフタミド形成、ジスルフィド架橋形成、エリミニル化、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結合、O結合、C結合、ホスホグリコシル化)、グリコシルホスファチジルイノシトール付加(glypiation)、ヘムC付着、ヒドロキシル化、ハイプシン形成、ヨウ素化、イソプレニル化、脂質付加、リポイル化、マロニル化、メチル化、ミリストイル化、酸化、パルミトイル化、ペグ化、ホスホパンテテイル化、リン酸化、プレニル化、プロピオニル化、レチニリデンシッフ塩

40

50

基形成、S - グルタチオン化、S - ニトロシル化、S - スルフェニル化、セレン化、サクシニル化、スルフィン化、ユビキチン化、およびC末端アミド化が挙げられる。翻訳後修飾は、ペプチドのアミノ末端および/またはカルボキシル末端、ポリペプチド、またはタンパク質の修飾を含む。末端アミノ基の修飾としては、これだけに限定されないが、デスアミノ、N - 低級アルキル、N - ジ - 低級アルキル、およびN - アシル修飾が挙げられる。末端カルボキシ基の修飾としては、これだけに限定されないが、アミド、低級アルキルアミド、ジアルキルアミド、およ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。翻訳後修飾は、例えば、これだけに限定されないが、ペプチド、ポリペプチド、またはタンパク質のアミノ末端とカルボキシ末端の間にあるアミノ酸の、上記のものなどの修飾も含む。翻訳後修飾により、細胞内のタンパク質の「生物学的性質」、例えば、その活性、構造、安定性、または局在を調節することができる。リン酸化は最も一般的な翻訳後修飾であり、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ら、2012年、Wiley Interdiscip Rev Syst Biol Med、4巻：565 ~ 583頁）。グリコシル化などの、タンパク質への糖の付加により、タンパク質のフォールディングが促進されること、安定性が改善されること、および調節機能が修飾されることが示されている。タンパク質に脂質を付着させるにより、細胞膜へのターゲティングが可能になる。翻訳後修飾は、1つまたは複数の検出可能な標識を含めるための、ペプチド、ポリペプチド、またはタンパク質の修飾も含み得る。

10

#### 【0346】

20

ある特定の実施形態では、ペプチド、ポリペプチド、またはタンパク質を断片化することができる。例えば、生体試料などの試料に由来するタンパク質を断片化することにより、断片化されたペプチドを得ることができる。ペプチド、ポリペプチド、またはタンパク質を、プロテアーゼまたはエンドペプチダーゼによる断片化を含めた当技術分野で公知の任意の手段によって断片化することができる。一部の実施形態では、ペプチド、ポリペプチド、またはタンパク質の断片化を、特異的なプロテアーゼまたはエンドペプチダーゼを使用することによって標的化する。特異的なプロテアーゼまたはエンドペプチダーゼは、特異的なコンセンサス配列に結合し、そこで切断する（例えば、ENLYFQ\ Sコンセンサス配列に対して特異的なTEVプロテアーゼ）。他の実施形態では、ペプチド、ポリペプチド、またはタンパク質の断片化を、非特異的なプロテアーゼまたはエンドペプチダーゼを使用することにより、標的化しないまたはランダムなものにする。非特異的なプロテアーゼは、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、そこで切断することができる（例えば、プロテイナーゼKは、非特異的なセリンプロテアーゼである）。プロテイナーゼおよびエンドペプチダーゼは当技術分野で周知であり、タンパク質またはポリペプチドを切断してより小さなペプチド断片にするために使用することができるものの例としては、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、トロンビン、第Xa因子、フューリン、エンドペプチダーゼ、パパイン、ペプシン、スブチリシン、エラスターゼ、エンテロキナーゼ、Genenase（商標）I、エンドプロテアーゼLysC、エンドプロテアーゼAspN、エンドプロテアーゼGluCなどが挙げられる（Granvogelら、2007年、Anal Bioanal Chem、389巻：991 ~ 1002頁）。ある特定の実施形態では、ペプチド、ポリペプチド、またはタンパク質を、プロテイナーゼKによって、または、任意選択で、迅速な不活化を可能にするために熱不安定性型のプロテイナーゼKによって、断片化する。プロテイナーゼKは、尿素およびSDSなどの変性試薬中で非常に安定であり、それにより、完全に変性したタンパク質を消化することが可能になる。タンパク質およびポリペプチドのペプチドへの断片化は、DNAタグまたはDNA記録タグを付着させる前に実施することもでき、その後にも実施することもできる。

30

40

#### 【0347】

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。化学試薬により、特定のアミノ酸残基において切断することができる（例えば、臭化シアンにより、メ

50

チオニン残基のC末端においてペプチド結合が加水分解される)。ポリペプチドまた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、NTCB + Ni(2-ニトロ-5-チオシアノ安息香酸)などが挙げられる。

#### 【0348】

ある特定の実施形態では、酵素的切断または化学的切断の後、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アミノ酸から約70アミノ酸まで、約10アミノ酸から約60アミノ酸まで、約10アミノ酸から約50アミノ酸まで、約10アミノ酸から約40アミノ酸まで、約10アミノ酸から約30アミノ酸まで、約20アミノ酸から約70アミノ酸まで、約20アミノ酸から約60アミノ酸まで、約20アミノ酸から約50アミノ酸まで、約20アミノ酸から約40アミノ酸まで、約20アミノ酸から約30アミノ酸まで、約30アミノ酸から約70アミノ酸まで、約30アミノ酸から約60アミノ酸まで、約30アミノ酸から約50アミノ酸まで、または約30アミノ酸から約40アミノ酸までである。切断反応は、タンパク質またはポリペプチド試料に、プロテイナーゼまたはエンドペプチダーゼ切断部位を含有するペプチド配列を含む短い試験FRET(蛍光共鳴エネルギー移動)ペプチドをスパイクすることにより、好ましくはリアルタイムで、モニターすることができる。インタクトなFRETペプチドでは、蛍光基およびクエンチャー基が切断部位を含有するペプチド配列のいずれかの末端に付着しており、クエンチャーとフルオロフォアのための蛍光共鳴エネルギー移動により、低い蛍光が生じる。試験ペプチドがプロテアーゼまたはエンドペプチダーゼによって切断されると、クエンチャーおよびフルオロフォアが分離し、それにより、蛍光の大きな増大がもたらされる。切断反応は、ある特定の蛍光強度が実現された時に停止させることができ、これにより、再現性のある切断エンドポイントを実現することが可能になる。

#### 【0349】

巨大分子(例えば、ペプチド、ポリペプチド、またはタンパク質)の試料に対して、固体支持体に付着させる前にタンパク質分画法を行うことができ、ここで、タンパク質またはペプチドを細胞内の位置、分子量、疎水性、もしくは等電点などの1つもしくは複数の性質、またはタンパク質濃縮法によって分離する。その代わりにまたはそれに加えて、特定のタンパク質またはペプチドを選択するため(例えば、その全体が参照により組み込まれる、Whiteakerら、2007年、Anal. Biochem., 362巻: 44~54頁を参照されたい)または特定の翻訳後修飾を選択するため(例えば、その全体が参照により組み込まれる、Huangら、2014年、J. Chromatogr. A, 1372巻: 1~17頁を参照されたい)に、タンパク質濃縮法を使用することができる。あるいは、免疫グロブリンなどの特定のクラス(複数可)のタンパク質、またはIgGなどの免疫グロブリン(Ig)アイソタイプを、解析のために親和性により濃縮または選択することができる。免疫グロブリン分子の場合では、親和性結合に関与する超可変配列の配列および豊富さまたは頻度を解析することが特に興味深く、これは、特に、これらが疾患の進行にตอบสนองして変動するまたは健康、免疫、および/もしくは疾患表現型と関連するからである。過度に豊富なタンパク質を、標準の免疫親和性方法を使用して試料からサブトラクションすることもできる。豊富なタンパク質の枯渇は、タンパク質構成物の80%よりも多くがアルブミンおよび免疫グロブリンである血漿試料に関して有用であり得る。例えばPROTIAおよびPROT20(Sigma-Aldrich)など、血漿試料の過度に豊富なタンパク質の枯渇のために、いくつかの市販品が入手可能である。

#### 【0350】

ある特定の実施形態では、巨大分子は、タンパク質またはポリペプチドで構成される。一実施形態では、タンパク質またはポリペプチドを標準のアミンカップリング化学によってDNA記録タグで標識する(例えば、図2B、2C、28、29、31、40)を参照されたい。 - アミノ基(例えばリシン残基の)およびN末端アミノ基は、反応のpHに

10

20

30

40

50

応じて、アミン反応性カップリング剤で特に標識しやすい (M e n d o z a および V a c h e t、2009 年)。特定の実施形態では (例えば、図 2 B および図 2 9 を参照されたい)、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、または巨大分子へのコンジュゲーションのための部分)、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、分配バーコード、試料バーコード、画分バーコード、またはそれらの任意の組合せ)、任意選択の U M I、およびコーディングタグへの / からの情報移行を容易にするためのスペーサー (S p) 配列で構成される。別の実施形態では、タンパク質を、まず、ユニバーサル D N A タグで標識し、その後で、酵素的または化学的カップリングステップによってバーコード - S p 配列 (試料、コンパートメント、スライド上の物理的位置を表すなど) をタンパク質に付着させることができる (例えば、図 2 0、3 0、3 1、4 0 を参照されたい)。ユニバーサル D N A タグは、タンパク質またはポリペプチド巨大分子を標識するために使用され、バーコード (例えば、コンパートメントタグ、記録タグなど) の付着点として使用することができるヌクレオチドの短い配列を含む。例えば、記録タグは、その末端に、ユニバーサル D N A タグと相補的な配列を含んでよい。ある特定の実施形態では、ユニバーサル D N A タグは、ユニバーサルプライミング配列である。標識されたタンパク質上のユニバーサル D N A タグが記録タグ (例えば、ビーズに結合させたもの) 内の相補配列とハイブリダイズしたら、アニーリングされたユニバーサル D N A タグをプライマー伸長によって伸長させ、それにより、D N A タグが付されたタンパク質に記録タグ情報を移行させることができる。特定の実施形態では、タンパク質を、プロテイナーゼにより消化してペプチドにする前に、ユニバーサル D N A タグで標識する。次いで、消化物に由来する標識されたペプチド上のユニバーサル D N A タグを、情報価値があり、有効な記録タグに変換することができる。

#### 【0351】

ある特定の実施形態では、タンパク質巨大分子を親和性捕捉試薬によって固体支持体に固定化すること (および任意選択で共有結合により架橋結合させること) ができ、ここで、記録タグは、親和性捕捉試薬に直接付随している、あるいは、タンパク質を固体支持体に記録タグと共に直接固定化することができる (例えば、図 2 C を参照されたい)。

#### V. 固体支持体

#### 【0352】

本開示の巨大分子を固体支持体の表面 (「基板表面」とも称される) に接合する。固体支持体は、これだけに限定されないが、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、ナイロン、シリコンウェーハチップ、フローセル、フロースルーチップ、信号変換電子機器を含むバイオチップ、マイクロタイターウェル、E L I S A プレート、スピン干渉ディスク、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアを含めた任意の多孔質または非多孔質支持体表面であってよい。固体支持体用の材料としては、これだけに限定されないが、アクリルアミド、アガロース、セルロース、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、ポリメタクリレート、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、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、ポリ酸無水物、ポリグリコール酸、ポリ乳酸、ポリオルトエステル、官能化シラン、ポリプロピルフメレート、コラーゲン、グリコサミノグリカン、ポリアミノ酸、またはそれらの任意の組合せが挙げられる。固体支持体は、薄膜、膜、ピン、ディッシュ、繊維、織られた繊維、チューブなどの成形ポリマー、粒子、ビーズ、微小粒子、またはそれらの任意の組合せをさらに含む。例えば、固体表面がビーズである場合、ビーズとしては、これだけに限定されないが、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズを挙げることができる。

#### 【0353】

ある特定の実施形態では、固体支持体は、フローセルである。フローセルの形態は、異

10

20

30

40

50

なる次世代シーケンシングプラットフォームの間で変動し得る。例えば、Illumina フローセルは、その表面に一連のオリゴヌクレオチドアンカーが結合した、顕微鏡スライドと同様の平面の光学的に透明な表面である。鋳型DNAは、末端にフローセル表面上のオリゴヌクレオチドと相補的なアダプターがライゲーションしている。アダプター付加した一本鎖DNAをフローセルに結合させ、固相「ブリッジ」PCRによって増幅した後、配列決定する。454 フローセル (454 Life Sciences) は、75ピコリットルのウェルを約160万個有する光ファイバースライドである「ピコタイター」プレートを支持するものである。せん断された鋳型DNAの個々の分子をそれぞれ別々のビーズに捕捉し、各ビーズを油エマルジョン中の水性PCR反応混合物の個々の液滴中に区画化する。鋳型をビーズ表面上でPCRによってクローン的に増幅し、次いで、鋳型がローディングされたビーズを配列決定反応のためにピコタイタープレートのウェル中に、理想的には1ウェル当たり1個またはそれ未満のビーズに分布させる。Applied Biosystems からの SOLiD (Supported Oligonucleotide Ligation and Detection) 計器では、454 システムと同様に、鋳型分子をエマルジョンPCRによって増幅する。増幅した鋳型を含有しないビーズを選別するステップの後、ビーズに結合した鋳型がフローセル上に蓄積する。フローセルは、TWIST (商標) DNA 合成カラム (Glen Research) などの単純なフィルターフリットであってもよい。

#### 【0354】

ある特定の実施形態では、固体支持体はビーズであり、これは、個々のビーズを指す場合もあり複数のビーズを指す場合もある。一部の実施形態では、ビーズは、下流の解析のために使用される選択された次世代シーケンシングプラットフォーム (例えば、SOLiD または 454) に適合するものである。一部の実施形態では、固体支持体は、アガロースビーズ、常磁性ビーズ、ポリスチレンビーズ、ポリマービーズ、アクリルアミドビーズ、固体コアビーズ、多孔質ビーズ、ガラスビーズ、または制御ポアビーズである。さらなる実施形態では、ビーズは、巨大分子への結合を容易にするために、結合機能性 (例えば、アミン基、ビオチン標識された巨大分子、抗体との結合用のストレプトアビジンなどの親和性リガンド) でコーティングすることができる。

#### 【0355】

タンパク質、ポリペプチド、またはペプチドは、固体支持体に、直接または間接的に、共有結合性の相互作用および非共有結合性の相互作用またはそれらの任意の組合せを含めた当技術分野で公知の任意の手段によって接合させることができる (例えば、それぞれ、これによってその全体が参照により組み込まれる、Chandra, 2007年、PLoS One, 2巻: e1164頁; Cazalisら、Bioconj. Chem., 15巻: 1005~1009頁; Soellnerら、2003年、J. Am. Chem. Soc., 125巻: 11790~11791頁; Sunら、2006年、Bioconj. Chem., 17巻: 52~57頁; Decreauら、2007年、J. Org. Chem., 72巻: 2794~2802頁; Camareroら、2004年、J. Am. Chem. Soc., 126巻: 14730~14731頁; Girishら、2005年、Bioorg. Med. Chem. Lett., 15巻: 2447~2451頁; Kalialら、2007年、Bioconj. Chem., 18巻: 1064~1069頁; Watzkeら、2006年、Angew Chem. Int. Ed. Engl., 45巻: 1408~1412頁; Parthasarathyら、2007年、Bioconjugate Chem., 18巻: 469~476頁; および Bioconjugate Techniques, G. T. Hermanson, Academic Press (2013年) を参照されたい)。例えば、ペプチドを固体支持体にライゲーション反応によって接合させることができる。あるいは、固体支持体は、直接または間接的にペプチドを固体支持体に接合することを容易にするための作用物質またはコーティングを含んでよい。タンパク質、核酸、炭水化物および小分子を含めた任意の適切な分子または材料をこの目的のために使用することができる。例えば、

10

20

30

40

50

一実施形態では、作用物質は、親和性分子である。別の例では、作用物質は、別の分子のアルキニル基と反応して固体支持体と他の分子の間の結び付きまたは結合を容易にすることができる、アジド基である。

#### 【0356】

タンパク質、ポリペプチド、またはペプチドは、固体支持体に、「クリックケミストリー」と称される方法を使用して接合させることができる。この目的のために、迅速かつ実質的に不可逆的である任意の反応を使用してタンパク質、ポリペプチド、またはペプチドを固体支持体に付着させることができる。例示的な反応としては、トリアゾールが形成される、アジドとアルキンの銅により触媒される反応（ヒュスゲン1, 3 - 双極子付加環化）、歪み促進型アジド - アルキン付加環化（SPAAC）、ジエンと求ジエン体の反応（ディールス・アルダー）、歪み促進型アルキン - ニトロ付加環化、歪アルケンとアジド、テトラジンまたはテトラゾールの反応、アルケン - アジド[3 + 2]付加環化、アルケン - テトラジン逆電子要請型ディールス・アルダー（IEDDA）反応（例えば、m - テトラジン（mTet） - trans - シクロオクテン（TCO））、アルケン - テトラゾール光化学反応、アジド - ホスフィンのシュタウディンガーライゲーション、ならびに、求電子原子に対する求核攻撃による脱離基の置換（Horisawa 2014年、Knall、Hollaufら、2014年）などの種々の置換反応が挙げられる。例示的な置換反応としては、アミンと活性化エステル（例えば、N - ヒドロキシスクシンイミドエステル）の反応；アミンとN - ヒドロキシスクシンイミドエステルの反応；アミンとイソシアネートの反応；アミンとイソチオシアネートの反応などが挙げられる。

#### 【0357】

一部の実施形態では、巨大分子および固体支持体を、2つの相補的な反応性基の反応によって形成することができる官能基、例えば、前述の「クリック」反応のうちの1つの生成物である官能基によって接合する。種々の実施形態では、官能基は、アルデヒド、オキシム、ヒドラゾン、ヒドラジド、アルキン、アミン、アジド、アシルアジド、ハロゲン化アシル、ニトリル、ニトロ、スルフヒドリル、ジスルフィド、ハロゲン化スルホニル、イソチオシアネート、イミドエステル、活性化エステル（例えば、N - ヒドロキシスクシンイミドエステル、ペンチン酸STPエステル）、ケトン、不飽和カルボニル、アルケン、マレイミド、ハロイミド、エポキシド、アジリジン、テトラジン、テトラゾール、ホスフィン、ピオチンまたはチラン官能基と、相補的な反応性基との反応によって形成することができる。例示的な反応は、アミン（例えば、第一級アミン）とN - ヒドロキシスクシンイミドエステルまたはイソチオシアネートの反応である。

#### 【0358】

さらに他の実施形態では、官能基は、アルケン、エステル、アミド、チオエステル、ジスルフィド、炭素環式、複素環式またはヘテロアリール基を含む。さらなる実施形態では、官能基は、アルケン、エステル、アミド、チオエステル、チオ尿素、ジスルフィド、炭素環式、複素環式またはヘテロアリール基を含む。他の実施形態では、官能基は、アミドまたはチオ尿素を含む。一部のより具体的な実施形態では、官能基は、トリアゾリル官能基、アミド、またはチオ尿素官能基である。

#### 【0359】

好ましい実施形態では、迅速かつ低インプット濃度で高収率がもたらされるので、IEDDAクリックケミストリーを巨大分子（例えば、タンパク質、ポリペプチド、ペプチド）を固体支持体に固定化するために使用する。別の好ましい実施形態では、テトラジンではなくm - テトラジンをIEDDAクリックケミストリー反応に使用し、これは、m - テトラジンが改善された結合安定性を有するからである。

#### 【0360】

好ましい実施形態では、基板表面をTCOで官能化し、記録タグで標識したタンパク質、ポリペプチド、ペプチドを、TCOでコーティングした基板表面に、付着したm - テトラジン部分を介して固定化する（図34）。

#### 【0361】

タンパク質、ポリペプチド、またはペプチドは、そのC末端、N末端、または内部アミノ酸により、例えば、アミン基、カルボキシル基、またはスルフィドリル基を介して、固体支持体の表面に固定化することができる。アミン基へのカップリングに使用される標準の活性化された支持体としては、C N B rで活性化された支持体、N H Sで活性化された支持体、アルデヒドで活性化された支持体、アズラクトンで活性化された支持体、およびC D Iで活性化された支持体が挙げられる。カルボキシルカップリングに使用される標準の活性化された支持体としては、アミン支持体にカップリングするカルボジイミドで活性化されたカルボキシル部分が挙げられる。システインカップリングでは、マレイミド、ヨードアセチル、およびピリジルジスルフィドで活性化された支持体を使用することができる。ペプチドカルボキシ末端固定化の代替方式では、C末端にリシンまたはアルギニン残基を含有するペプチドに、それらを切断することなく結合する、触媒として不活性なトリプシンの誘導体であるアンヒドロトリプシンを使用する。

10

#### 【0362】

ある特定の実施形態では、タンパク質、ポリペプチド、またはペプチドを固体支持体に、固体表面に結合させたリンカーをタンパク質、ポリペプチド、またはペプチドのリシン基に共有結合により付着させることによって固定化する。

#### 【0363】

固体支持体への固定化前または固定化後に、タンパク質、ポリペプチド、またはペプチドに記録タグを付着させることができる。例えば、タンパク質、ポリペプチド、またはペプチドを、まず記録タグで標識し、次いで、カップリングのための2つの機能的部分を含む記録タグを介して固体表面に固定化することができる(図28を参照されたい)。記録タグの一方の機能的部分をタンパク質とカップリングさせ、他方の機能的部分により、記録タグで標識したタンパク質を固体支持体に固定化する。

20

#### 【0364】

あるいは、タンパク質、ポリペプチドまたはペプチドを記録タグで標識する前に、タンパク質、ポリペプチド、またはペプチドを固体支持体に固定化する。例えば、タンパク質をまずクリックケミストリー部分などの反応性基で誘導体化することができる。次いで、活性化されたタンパク質分子を適切な固体支持体に付着させ、次いで、相補的なクリックケミストリー部分を使用して記録タグで標識することができる。例として、アルキンおよびm T e t部分で誘導体化したタンパク質を、アジドおよびT C Oで誘導体化したビーズに固定化し、アジドおよびT C Oで標識された記録タグに付着させることができる。

30

#### 【0365】

巨大分子(例えば、タンパク質、ポリペプチド、またはペプチド)を固体支持体に付着させるための本明細書において提示される方法はまた、記録タグを固体支持体に付着させるまたは記録タグを巨大分子(例えば、タンパク質、ポリペプチド、またはペプチド)に付着させるためにも使用することができることが理解される。

#### 【0366】

ある特定の実施形態では、結合性物質への非特異的吸収を最小限にするために、固体支持体の表面を不動態化(ブロッキング)する。「不動態化」された表面とは、結合性物質の非特異的結合を最小限にするために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、表面を、ポリエチレングリコール(P E G)(P a nら、2015年、P h y s . B i o l .、12巻:045006頁)、ポリシロキサン(例えば、P l u r o n i c F - 127)、星型ポリマー(例えば、星型P E G)(G r o l lら、2010年、M e t h o d s E n z y m o l .、472巻:1~18頁)、疎水性ジクロロジメチルシラン(D D S)+自己組織化T w e e n - 20(H u aら、2014年、N a t . M e t h o d s、11巻:1233~1236頁)、およびダイヤモンド様炭素(D L C)、D L C+P E G(S t a v i sら、2011年、P r o c . N a t l . A c a d . S c i . U S A、108巻:983~988頁)のようなポリマーで不動態化することを含めた、蛍光単一分子解析の文献からの標準の方法が挙げられる。共有結合性の表面修飾に加えて、T w e e n - 20のような界面活性物質、溶液中ポリシロキサン

40

50



(Pluronicシリーズ)、ポリビニルアルコール(PVA)、ならびにBSAおよびカゼインのようなタンパク質を含めたいくつもの不動態化剤を同様に使用することができる。あるいは、固体基板の表面上または容積内のタンパク質、ポリペプチド、またはペプチドの密度を、タンパク質、ポリペプチドまたはペプチドを固体基板に固定化する際に競合剤または「ダミー」反応性分子をスパイクすることによって調整することができる(図36Aを参照されたい)。

#### 【0367】

ある特定の実施形態では、多数の巨大分子を同じ固体支持体上に固定化する場合、例えば、結合性物質が第1の巨大分子に結合し、そのコーディングタグ情報が第1の巨大分子に付随する記録タグではなく近隣の巨大分子に付随する記録タグに移行する交差結合または分子間事象の発生を低減するまたはそれを防止するために、巨大分子の間に適切に間隔をあけることができる。固体支持体上の巨大分子の間隔(例えば、タンパク質、ポリペプチド、またはペプチドの間隔)を制御するために、基板表面上の機能的カップリング基(例えば、TCO)の密度を調整することができる(図34を参照されたい)。一部の実施形態では、固体支持体(例えば、多孔質支持体)の表面上または容積内で多数の巨大分子の間に約50nm~約500nm、または約50nm~約400nm、または約50nm~約300nm、または約50nm~約200nm、または約50nm~約100nmの距離で間隔をあける。一部の実施形態では、固体支持体の表面上で多数の巨大分子の間に少なくとも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、または少なくとも500nmの平均距離で間隔をあける。一部の実施形態では、固体支持体の表面上で多数の巨大分子の間に少なくとも50nmの平均距離で間隔をあける。一部の実施形態では、固体支持体の表面上または容積内で巨大分子の間に、経験的に分子内事象に対する分子間事象の相対的な頻度が<1:10; <1:100; <1:1,000; または<1:10,000になるように間隔をあける。適切な間隔頻度は、機能アッセイを使用して経験的に決定することができる(実施例23を参照されたい)、希釈によっておよび/または基板表面上の部位への付着について競合する「ダミー」 Spacer分子をスパイクすることによって実現することができる。

#### 【0368】

例えば、図34に示されている通り、PEG-5000(MW約5000)を使用して、基板表面(例えば、ビーズ表面)上のペプチド間の間隙の空間をブロッキングする。さらに、ペプチドを同じくPEG-5000分子に付着させた機能的部分とカップリングさせる。好ましい実施形態では、これは、NHS-PEG-5000-TCO+NHS-PEG-5000-メチルの混合物をアミン誘導体化ビーズにカップリングすることによって実現される(図34を参照されたい)。2つのPEG間(TCO対メチル)の化学量論比を調整して、基板表面上の機能的カップリング部分(TCO基)の適切な密度を生じさせる;メチル-PEGはカップリングに対して不活性である。TCO基間の有効な間隔は、表面上のTCO基の密度を測定することによって算出することができる。ある特定の実施形態では、固体表面上のカップリング部分(例えば、TCO)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、または少なくとも500nmである。ビーズをPEG5000-TCO/メチルで誘導体化した後、表面上の過剰なNH<sub>2</sub>基を反応性無水物(例えば、無水酢酸または無水コハク酸)でクエンチする。

#### VI. 記録タグ

#### 【0369】

少なくとも1つの記録タグを巨大分子に直接または間接的に付随または共同在させ、固体支持体に接合させる(例えば、図5を参照されたい)。記録タグは、DNA、RNA、PNA、PNA、GNA、BNA、XNA、TNA、ポリヌクレオチド類似体、またはこれらの組合せを含んでよい。記録タグは、一本鎖であってもよく、部分的にまたは完全に二本鎖であってもよい。記録タグは、平滑末端を有してもよく突出末端を有してもよい

10

20

30

40

50

。ある特定の実施形態では、結合性物質が巨大分子に結合したら、結合性物質のコーディングタグの識別情報を記録タグに移行させて、伸長記録タグを生成する。その後の結合サイクルにおいて伸長記録タグに対するさらなる伸長を行うことができる。

#### 【0370】

記録タグは、固体支持体に、共有結合性の相互作用および非共有結合性の相互作用、またはそれらの任意の組合せを含めた当技術分野で公知の任意の手段によって直接または間接的に（例えば、リンカーを介して）接合させることができる。例えば、記録タグを固体支持体にライゲーショ反応によって接合させることができる。あるいは、固体支持体は、記録タグを固体支持体に直接または間接的に接合させることを容易にするための作用物質またはコーティングを含んでよい。核酸分子を固体支持体（例えば、ビーズ）に固定化するための戦略は、それぞれ、その全体が参照により本明細書に組み込まれる、米国特許第5,900,481号；Steinbergら（2004年、Biopolymers、73巻：597～605頁）；Lundら、1988年（Nucleic Acids Res.、16巻：10861～10880頁）；およびSteinbergら（2004年、Biopolymers、73巻：597～605頁）に記載されている。

10

#### 【0371】

ある特定の実施形態では、巨大分子（例えば、ペプチド）と付随する記録タグの共局在は、巨大分子および記録タグを、固体支持体表面に直接付着させた二機能性リンカーとコンジュゲートすることによって実現される。Steinbergら（2004年、Biopolymer、73巻：597～605頁）。さらなる実施形態では、三機能性部分を使用して固体支持体（例えば、ビーズ）を誘導体化し、得られた二機能性部分を巨大分子および記録タグの両方とカップリングさせる。

20

#### 【0372】

巨大分子および固体支持体の付着に関して記載されているものなどの方法および試薬（例えば、クリックケミストリー試薬および光親和性標識試薬）を記録タグの付着にも使用することができる。

#### 【0373】

特定の実施形態では、単一の記録タグを巨大分子（例えば、ペプチド）に、好ましくは脱プロッキングしたN末端アミノ酸またはC末端アミノ酸に付着させることによって付着させる。別の実施形態では、多数の記録タグを巨大分子（例えば、タンパク質、ポリペプチド、またはペプチド）に、好ましくはリシン残基またはペプチド骨格に付着させる。一部の実施形態では、多数の記録タグで標識した巨大分子（例えば、タンパク質またはポリペプチド）を、それぞれが平均して1つの記録タグで標識された、より小さなペプチドに断片化または消化する。

30

#### 【0374】

ある特定の実施形態では、記録タグは、任意選択の、UMIが付随する各巨大分子（例えば、タンパク質、ポリペプチド、ペプチド）についての一意の識別子タグをもたらす一意の分子識別子（UMI）を含む。UMIは、約3～約40塩基、約3～約30塩基、約3～約20塩基、または約3～約10塩基、または約3～約8塩基であってよい。一部の実施形態では、UMIは、約3塩基、4塩基、5塩基、6塩基、7塩基、8塩基、9塩基、10塩基、11塩基、12塩基、13塩基、14塩基、15塩基、16塩基、17塩基、18塩基、19塩基、20塩基、25塩基、30塩基、35塩基、または40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配列決定データをデコンボリューションして個々の巨大分子からの配列読み取りを識別することができる。一部の実施形態では、巨大分子のライブラリー内で、各巨大分子に単一の記録タグを付随させ、各記録タグが一意的UMIを含む。他の実施形態では、記録タグの多数のコピーを単一の巨大分子に付随させ、記録タグの各コピーは同じUMIを含む。一部の実施形態では、UMIは、結合性物質のコーディングタグ内のスペーサーまたはエンコーダー配列と、これらの成分を配列解析中に区別することを容易にするために、異なる塩基配列を有する。

40

#### 【0375】

50

ある特定の実施形態では、記録タグは、例えば、存在する場合、U M I 以外のバーコードを含む。バーコードは、約 3 ~ 約 3 0 塩基、約 3 ~ 約 2 5 塩基、約 3 ~ 約 2 0 塩基、約 3 ~ 約 1 0 塩基、約 3 ~ 約 1 0 塩基、約 3 ~ 約 8 塩基の長さの核酸分子である。一部の実施形態では、バーコードは、約 3 塩基、4 塩基、5 塩基、6 塩基、7 塩基、8 塩基、9 塩基、1 0 塩基、1 1 塩基、1 2 塩基、1 3 塩基、1 4 塩基、1 5 塩基、2 0 塩基、2 5 塩基、または 3 0 塩基の長さである。一実施形態では、バーコードにより、複数の試料またはライブラリーのマルチプレックス配列決定が可能になる。バーコードを使用して、分配、画分、コンパートメント、試料、空間的位置、または巨大分子（例えば、ペプチド）が由来するライブラリーを識別することができる。バーコードを使用して、多重化された配列データをデコンボリューションし、個々の試料またはライブラリーからの配列読み取りを識別する。例えば、バーコードが付されたビーズは、試料のエマルジョンおよび分配を伴う方法に、例えば、プロテオームを分配するために、有用である。

10

#### 【 0 3 7 6 】

バーコードは、例えば液滴、マイクロウェル、固体支持体上の物理的領域などのコンパートメントに一意的バーコードが割り当てられたコンパートメントタグを表し得る。コンパートメントと特異的なバーコードの結び付きは、例えば、単一のバーコードが付されたビーズをコンパートメントに封入することによって、例えば、バーコードが付された液滴をコンパートメントに直接混合させるまたは添加することによって、バーコード試薬をコンパートメントに直接プリントまたは注射することによってなどの任意の数のやり方で実現することができる。コンパートメント内のバーコード試薬を使用して、コンパートメント特異的バーコードをコンパートメント内の巨大分子またはその断片に付加する。コンパートメント内へのタンパク質の分配に適用すると、バーコードを使用して、解析されるペプチドをコンパートメント内のそれらの起源であるタンパク質分子にマッピングし戻すことができる。これにより、タンパク質識別が著しく容易になる。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。

20

#### 【 0 3 7 7 】

他の実施形態では、コンパートメントの集団のサブセットを表す多数のコンパートメントにそのサブセットを表す一意のバーコードを割り当てることができる。

#### 【 0 3 7 8 】

あるいは、バーコードは、試料識別バーコードであってよい。試料バーコードは、単一反応容器中のまたは単一の固体基板または固体基板の集合（例えば、平面スライド、単一のチューブまたは容器などに含有されるビーズの集団）に固定化された試料のセットの多重化解析に有用である。多くの異なる試料由来の巨大分子を、試料特異的バーコードを有する記録タグで標識することができ、次いで、全ての試料を、固体支持体への固定化、周期的結合、および記録タグ解析の前に一緒にプールのすることができる。あるいは、DNA によりコードされるライブラリー作成後まで試料を別々に保持し、DNA によりコードされるライブラリーの PCR 増幅中に試料バーコードを付着させ、次いで、配列決定前に混合することができる。この手法は、豊富さのクラスが異なる分析物（例えば、タンパク質）をアッセイする場合に有用であり得る。例えば、試料を分割し、バーコードを付し、一方の部分を豊富さが低い分析物に対する結合性物質を使用して処理し、他方の部分を豊富さがより高い分析物に対する結合性物質を使用して処理することができる。特定の実施形態では、この手法は、特定のタンパク質分析物アッセイのダイナミックレンジをタンパク質分析物の標準の発現レベルの「スイートスポット」に入るように調整するのに役立つ。

30

40

#### 【 0 3 7 9 】

ある特定の実施形態では、多数の異なる試料に由来するペプチド、ポリペプチド、またはタンパク質を、試料特異的バーコードを含有する記録タグで標識する。多試料バーコードが付されたペプチド、ポリペプチド、またはタンパク質を周期的結合反応前に混合することができる。このように、デジタル逆相タンパク質アレイ（RPPA）に対する高度に多重化された代替物が有効に創出される（Guo、Liuら、2012年、Assadi、Lamerzら、2013年、Akbari、Beckerら、2014年、Crei

50

g h t o n および H u a n g、2015 年)。デジタル R P P A 様アッセイの創出には、翻訳研究、バイオマーカー検証、薬物発見、臨床、および高精度の医療における多数の適用がある。

#### 【0380】

ある特定の実施形態では、記録タグは、ユニバーサルプライミング部位、例えば、フォワードまたは 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をプライミングするためおよび/または配列決定のため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としては、これだけに限定されないが、P C R 増幅のためのプライミング部位、フローセル表面上の相補的なオリゴヌクレオチドとアニーリングするフローセルアダプター配列（例えば、I l l u m i n a 次世代シーケンシング）、配列決定プライミング部位、またはこれらの組合せを挙げることができる。ユニバーサルプライミング部位は、約 10 塩基～約 60 塩基であってよい。一部の実施形態では、ユニバーサルプライミング部位は、I l l u m i n a P5 プライマー（5' - A A T G A T A C G G C G A C C A C C G A - 3' - 配列番号 133）または I l l u m i n a P7 プライマー（5' - C A A G C A G A A G A C G G C A T A C G A G A T - 3' - 配列番号 134）を含む。

10

#### 【0381】

ある特定の実施形態では、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、記録タグに関してスペーサー配列への言及は、その同類結合性物質に付随するスペーサー配列、またはその同類結合性物質に付随するスペーサー配列と相補的なスペーサー配列と同一のスペーサー配列を含む。記録タグ上の末端、例えば 3' スペーサーにより、第 1 の結合サイクル中に同類結合性物質の識別情報をそのコーディングタグから記録タグに移行させることが可能になる（例えば、プライマー伸長または粘着末端ライゲーションのための相補的なスペーサー配列のアニーリングによって）。

20

#### 【0382】

一実施形態では、スペーサー配列は、約 1～20 塩基の長さ、約 2～12 塩基の長さ、または 5～10 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移行させるためのプライマー伸長反応の温度および反応条件などの因子に依存し得る。

#### 【0383】

好ましい実施形態では、記録タグ内のスペーサー配列を、記録タグ内の他の領域に対して最小の相補性を有するように設計する；同様に、コーディングタグ内のスペーサー配列はコーディングタグ内の他の領域に対して最小の相補性を有するべきである。言い換えれば、記録タグおよびコーディングタグのスペーサー配列は、記録タグまたはコーディングタグ内に存在する一意の分子識別子、バーコード（例えば、コンパートメント、分配、試料、空間的位置）、ユニバーサルプライマー配列、エンコーダー配列、サイクル特異的配列などの成分に対して最小の配列相補性を有するべきである。

30

#### 【0384】

結合性物質スペーサーについて記載されている通り、一部の実施形態では、巨大分子のライブラリーに付随する記録タグは、共通のスペーサー配列を共有する。他の実施形態では、巨大分子のライブラリーに付随する記録タグは、それらの同類結合性物質の結合サイクル特異的スペーサー配列と相補的な結合サイクル特異的スペーサー配列を有し、これは、非連鎖状伸長記録タグを使用する場合に有用であり得る（図 10 を参照されたい）。

40

#### 【0385】

伸長記録タグの収集物は、事後に連鎖状にすることができる（例えば、図 10 を参照されたい）。結合サイクルが完了した後、ビーズ固体支持体であって、各ビーズが平均してビーズ当たり 1 つまたは 1 つ未満の巨大分子を有し、各巨大分子が巨大分子の部位に共局在する伸長記録タグの収集物を有するビーズ固体支持体をエマルジョン中に入れる。エマルジョンは、各液滴が平均して最大で 1 ビーズによって占められるように形成する。任意選択のアセンブリ P C R 反応をエマルジョン中で実施してビーズ上の巨大分子と共局在する伸長記録タグを増幅し、それらを別の伸長記録タグの異なるサイクル特異的配列間のブ

50

ライミングによって共直線的な順序でアセンブルさせる (Xiong、Pengら、2008年)。その後、エマルジョンを破壊し、アセンブルした伸長記録タグを配列決定する。

#### 【0386】

別の実施形態では、DNA記録タグは、第1の結合サイクルに特異的なユニバーサルプライミング配列 (U1)、1つまたは複数のバーコード配列 (BCs)、およびスペーサー配列 (Sp1) で構成される。第1の結合サイクルでは、結合性物質はSp1相補のスペーサー、エンコーダーバーコード、および任意選択のサイクルバーコード、および第2のスペーサーエレメント (Sp2) で構成されるDNAコーディングタグを使用する。少なくとも2つの異なるスペーサーエレメントを使用することの有用性は、第1の結合サイクルにより潜在的に複数のDNA記録タグのうちの1つが選択され、単一のDNA記録タグが伸長し、その結果、伸長DNA記録タグの最後に新しいSp2スペーサーエレメントがもたらされることである。第2の結合サイクルおよびその後の結合サイクルでは、結合性物質は、Sp1'ではなくSp2'スペーサーだけを含む。このように、第1のサイクルからの単一の伸長記録タグのみがその後のサイクルで伸長する。別の実施形態では、第2のサイクルおよびその後のサイクルに結合性物質特異的スペーサーを使用することができる。

10

#### 【0387】

一部の実施形態では、記録タグは、5'から3'の方向に、ユニバーサルフォワード (または5') プライミング配列、UMI、およびスペーサー配列を含む。一部の実施形態では、記録タグは、5'から3'の方向に、ユニバーサルフォワード (または5') プライミング配列、任意選択のUMI、バーコード (例えば、試料バーコード、分配バーコード、コンパートメントバーコード、空間バーコード、またはそれらの任意の組合せ)、およびスペーサー配列を含む。一部の他の実施形態では、記録タグは、5'から3'の方向に、ユニバーサルフォワード (または5') プライミング配列、バーコード (例えば、試料バーコード、分配バーコード、コンパートメントバーコード、空間バーコード、またはそれらの任意の組合せ)、任意選択のUMI、およびスペーサー配列を含む。

20

#### 【0388】

改変DNAおよびPNAからUMIを生成するためにコンビナトリアル手法を使用することができる。一実施例では、UMIは、互いに直交性になるように設計された短いワード配列 (4~15mer) のセットの「化学的ライゲーション」によって構築することができる (SpiropoulosおよびHeemstra、2012年)。「ワード」ポリマーの化学的ライゲーションを導くためにDNA鋳型を使用する。DNA鋳型は、単に溶液中で副成分と一緒に混合することによってコンビナトリアル鋳型構造をアセンブルすることを可能にするハイブリダイズアームを用いて構築する (図12Cを参照されたい)。ある特定の実施形態では、この設計には「スペーサー」配列は存在しない。ワードスペースのサイズは、10ワードから10,000またはそれよりも多くのワードまで変動し得る。ある特定の実施形態では、ワードは、交差ハイブリダイズはしないように互いとは異なるが、それでも比較的均一なハイブリダイゼーション条件が保たれるように選択する。一実施形態では、ワードの長さは、およそ10塩基であり、サブセット内に約1000ワードを有する (これは、全10merワードスペースのたった0.1%である、約4<sup>10</sup> = 百万ワード)。これらのワードのセット (サブセット内に1000) を連鎖状にして、複雑さ = 1000<sup>n</sup> パワーの最終的なコンビナトリアルUMIを生成することができる。連鎖状にした4つのワードに関しては、これにより、10<sup>12</sup>種の異なるエレメントのUMI多様性が生じる。これらのUMI配列を巨大分子 (ペプチド、タンパク質など) に単一分子レベルで付加する。一実施形態では、UMIの多様性は、UMIを付着させる巨大分子の分子数を超える。このように、UMIにより、目的の巨大分子が一意的に識別される。コンビナトリアルワードUMIの使用により、多数塩基長のワードを読み取るために一塩基分解能は必要ないので、エラー率が高いシーケンサー (例えば、ナノポアシーケンサー、ナノギャップトンネリングシーケンシングなど) での読み取りが容易になる。コンビナトリアルワード手法は、コンパートメントタグ、分配バーコード、空間バーコード、

30

40

50

試料バーコード、エンコーダー配列、サイクル特異的配列、およびバーコードなどの、他の同一性に関する情報価値のある記録タグまたはコーディングタグの成分を生成するためにも使用することができる。ナノポアシーケンシングに関する方法およびエラー許容性ワード(コード)を有する情報をコードするDNAは、当技術分野で公知である(例えば、それぞれ、その全体が参照により組み込まれる、Kiahら、2015年、Codes for DNA sequence profiles, IEEE International Symposium on Information Theory (ISIT); Gabrysら、2015年、Asymmetric Lee distance codes for DNA-based storage, IEEE Symposium on Information Theory (ISIT); Laureら、2016年、Coding in 2D: Using Intentional Dispersity to Enhance the Information Capacity of Sequence-Coded Polymer Barcodes, Angew. Chem. Int. Ed. doi:10.1002/anie.201605279; Yazdiら、2015年、IEEE Transactions on Molecular, Biological and Multi-Scale Communications, 1巻:230~248頁;およびYazdiら、2015年、Sci Rep, 5巻:14138頁を参照されたい)。したがって、ある特定の実施形態では、本明細書に記載の実施形態のいずれかにおける伸長記録タグ、伸長コーディングタグ、またはジタグ構築物は、エラー訂正コードである識別成分(例えば、UMI、エンコーダー配列、バーコード、コンパートメントタグ、サイクル特異的配列など)で構成される。一部の実施形態では、エラー訂正コードは、Hammingコード、Lee距離コード、非対称Lee距離コード、Reed-Solomonコード、およびLevenshtein-Tenengoltsコードから選択される。ナノポアシーケンシングに関しては、電流またはイオンフラックスプロファイルおよび非対称の塩基呼び出しエラーは使用されるナノポアの型および生化学的性質に内因するものであり、上述のエラー訂正手法を使用してよりロバストなDNAコードを設計するためにこの情報を使用することができる。ロバストなDNAナノポアシーケンシングバーコードの使用に対する代替として、バーコード配列の電流またはイオンフラックスシグネチャを直接使用し(その全体が参照により組み込まれる、米国特許第7,060,507号)、それによりDNA塩基呼び出しを完全に回避し、バーコード配列を、Laszloら(2014年、Nat. Biotechnol., 32巻:829~833頁、その全体が参照により組み込まれる)に記載されている通り、予測される電流/フラックスシグネチャにマッピングし戻すことによってすぐに識別することができる。この論文において、Laszloらは、生物学的ナノポア、MspAから、異なるワードのつながりをナノポアを通過させた時に生じる電流シグネチャ、ならびに、得られた電流シグネチャを配列のユニバースからの可能性のある電流シグネチャのin silico予測にマッピングし戻すことによってDNA鎖をマッピングおよび識別する能力について記載している(2014年、Nat. Biotechnol., 32巻:829~833頁)。同様の概念を、DNAコードおよびナノギャップトンネル電流に基づくDNA配列決定によって生じる電気信号に適用することができる(Ohshiroら、2012年、Sci Rep, 2巻:501頁)。

【0389】

したがって、ある特定の実施形態では、コーディングタグ、記録タグ、またはその両方の識別成分は、一意の電流またはイオンフラックスまたは光学的シグネチャを生成することが可能であり、本明細書において提示される方法のいずれかの解析ステップは、識別成分を識別するために一意の電流またはイオンフラックスまたは光学的シグネチャを検出することを含む。一部の実施形態では、識別成分は、エンコーダー配列、バーコード、UMI、コンパートメントタグ、サイクル特異的配列、またはそれらの任意の組合せから選択される。

【0390】

10

20

30

40

50

ある特定の実施形態では、試料中の巨大分子（例えば、タンパク質、ポリペプチド、またはペプチド）の全量または実質的な量（例えば、少なくとも50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、96%、97%、98%、99%、または100%）を記録タグで標識する。巨大分子の標識は、巨大分子を固体支持体に固定化する前に行うこともでき、その後に行うこともできる。

#### 【0391】

他の実施形態では、試料中の巨大分子のサブセット（例えば、タンパク質、ポリペプチド、またはペプチド）を記録タグで標識する。特定の実施形態では、試料由来の巨大分子のサブセットに記録タグを用いて標的化（分析物特異的）標識を行う。タンパク質の標的化記録タグ標識は、相補的な標的特異的ベイト配列、例えば記録タグ内の分析物特異的バーコードとアニーリングする短い標的特異的DNA捕捉用プローブ、例えば分析物特異的バーコードと連結した標的タンパク質特異的結合性物質（例えば、抗体、アプタマーなど）を使用して実現することができる（図28Aを参照されたい）。記録タグは、標的タンパク質上に存在する同類の反応性部分（例えば、クリックケミストリー標識、光親和性標識）に対する反応性部分を含む。例えば、記録タグは、アルキンで誘導体化されたタンパク質と相互作用させるためのアジド部分を含んでよい、または、記録タグは、ネイティブなタンパク質で相互作用させるためのベンゾフェノンを含んでよい、などである（図28A～Bを参照されたい）。標的タンパク質に標的タンパク質特異的結合性物質を結合させたら、記録タグおよび標的タンパク質をそれらの対応する反応性部分を介してカップリングさせる（図28B～Cを参照されたい）。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性物質を、標的タンパク質特異的結合性物質に連結したDNA捕捉用プローブの消化によって除去することができる。例えば、DNA捕捉用プローブをウラシル塩基が含有されるように設計することができ、次いでこれを、ウラシル特異的切除試薬（例えば、USER（商標））を用いた消化の標的とし、標的タンパク質特異的結合性物質を標的タンパク質から解離させることができる。

#### 【0392】

一実施例では、標的タンパク質のセットに特異的な抗体を、相補的なベイト配列（例えば、図28の分析物バーコードBCAの）を用いて設計した記録タグとハイブリダイズするDNA捕捉用プローブ（例えば、図28の分析物バーコードBCA）で標識することができる。タンパク質の試料特異的標識は、試料特異的バーコードを含む記録タグ上の相補的なベイト配列とハイブリダイズするDNA捕捉用プローブで標識された抗体を使用することによって実現することができる。

#### 【0393】

別の例では、標的タンパク質特異的アプタマーを試料中のタンパク質のサブセットの標的化記録タグ標識のために使用する。標的特異的アプタマーを、記録タグ内の相補的なベイト配列とアニーリングするDNA捕捉用プローブに連結する。記録タグは、対応する反応性部分を有する標的タンパク質とカップリングするための反応性化学プローブまたは光反応性化学プローブ（例えば、ベンゾフェノン（BP））を含む。アプタマーがその標的タンパク質分子に結合し、記録タグが標的タンパク質の極めて近傍になり、その結果、記録タグと標的タンパク質がカップリングする。

#### 【0394】

小分子タンパク質親和性リガンドに付着させた光反応性化学プローブを使用した光親和性（PA）タンパク質標識は以前記載されている（Park, Kohら、2016年）。典型的な光反応性化学プローブとしては、以前に記載されている照射波長の下で活性化されたベンゾフェノンに基づくプローブ（反応性ジラジカル、365nm）、フェニルジアジリンに基づくプローブ（反応性炭素、365nm）、およびアジ化フェニルに基づくプローブ（反応性ニトロフリーラジカル、260nm）が挙げられる（SmithおよびCollins、2015年）。好ましい実施形態では、タンパク質試料中の標的タンパク質を、Liらにより開示された、ベンゾフェノンで標識した記録タグ内のベイト配列を同類結合性物質（例えば、核酸アプタマー（図28を参照されたい））に付着させたD

N A 捕捉用プローブとハイブリダイズさせる方法 (Li、Liuら、2013年) を使用して試料バーコードを含む記録タグで標識する。光親和性標識されたタンパク質標的に関しては、光親和性部分により標的タンパク質ではなく抗体が自己標識される可能性があるため、抗体よりもDNA/RNA アプタマーを標的タンパク質特異的結合性物質として使用することが好ましい。対照的に、光親和性標識は、核酸に対してはタンパク質よりも効率が高く、それにより、DNA 指向性化学標識または光標識についてはアプタマーがより良好なビヒクルになる。光親和性標識と同様に、Rosenら (Rosen、Kodalら、2014年、Kodal、Rosenら、2016年) 記載されているものと同様に、アプタマー結合性部位の近傍にある反応性リシン (または他の部分) のDNA 指向性化学標識を使用することもできる。

10

#### 【0395】

上述の実施形態では、標的特異的結合性物質と記録タグを連結するために、ハイブリダイゼーションに加えて他の型の連結を使用することができる (図28Aを参照されたい)。例えば、図28Bに示されている通り、2つの部分を、捕捉された標的タンパク質 (または他の巨大分子) が記録タグと共有結合したら、切断され、結合性物質を放出するように設計されたリンカーを使用して共有結合により連結することができる。適切なリンカーを記録タグの様々な位置、例えば、3' 末端、または記録タグの5' 末端に付着させたリンカー内などに付着させることができる。

#### VII. 結合性物質およびコーディングタグ

#### 【0396】

20

本明細書に記載の方法では、巨大分子に結合することが可能な結合性物質を使用する。結合性物質は、巨大分子の成分もしくは特徴に結合することが可能な任意の分子であってよい (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)。結合性物質は、天然に存在する分子であってもよく、合成的に作製された分子であってもよく、組換えによって発現させた分子であってもよい。結合性物質は、巨大分子の単一の単量体またはサブユニット (例えば、ペプチドの単一のアミノ酸) に結合し得る、または巨大分子の多数の連結したサブユニット (例えば、より長いペプチド分子のジペプチド、トリペプチド、またはより高次のペプチド) に結合し得る。

#### 【0397】

ある特定の実施形態では、結合性物質を、共有結合により結合するように設計することができる。共有結合は、正確な部分への結合が条件付けられるまたはそれが有利になるように設計することができる。例えば、NTAA およびその同類のNTAA 特異的結合性物質はそれぞれ、NTAA 特異的結合性物質が同類のNTAA に結合したらカップリング反応が行われてこれら2つの間に共有結合性の連結が生じるように、反応性基を用いて修飾することができる。同類の反応性基を欠く他の位置への結合性物質の非特異的結合では共有結合による付着はもたらされない。結合性物質とその標的の間の共有結合により、非特異的に結合した結合性物質を除去するためによりストリンジェントな洗浄を使用することが可能になり、したがって、アッセイの特異度が増大する。

30

#### 【0398】

ある特定の実施形態では、結合性物質は、選択的結合性物質であってよい。本明細書で使用される場合、選択的結合とは、結合性物質が、特異的なリガンド (例えば、アミノ酸またはアミノ酸のクラス) に、異なるリガンド (例えば、アミノ酸またはアミノ酸のクラス) への結合と比べて優先的に結合できることを指す。選択性は、一般に、結合性物質との複合体において1つのリガンドが別のリガンドで置換される反応についての平衡定数とされる。一般には、そのような選択性は、リガンドの空間的な幾何学的形状ならびに/またはリガンドが結合性物質に結合する様式および程度、例えば、水素結合またはファンデルワールス力 (非共有結合性の相互作用) によるもしくは可逆的または非可逆的な共有結合による、結合性物質への付着などに関連付けられる。選択性は、絶対的なものとは対照的に相対的なものであってよいこと、および、リガンド濃度を含めた種々の因子が選択性に影響を及ぼし得ることも理解されるべきである。したがって、一実施例では、結合性物

40

50



質は、20種の標準のアミノ酸のうちの1種に選択的に結合する。非選択的結合の例では、結合性物質は、20種の標準のアミノ酸のうちの2種またはそれよりも多くに結合し得る。

#### 【0399】

本明細書に開示されている方法の実施において、結合性物質の、巨大分子の特徴または成分に選択的に結合する能力は、結合性物質のコーディングタグ情報の巨大分子に付随する記録タグへの移行、記録タグ情報のコーディングタグへの移行、またはコーディングタグ情報および記録タグ情報のジタグ分子への移行を可能にするために十分なものであればよい。したがって、選択的とは、巨大分子が暴露される他の結合性物質に対して相対的なものであればよい。結合性物質の選択性は、特定のアミノ酸に対して絶対的なものである必要はなく、非極性もしくは非極性側鎖を有するアミノ酸、または電氣的に（正にもしくは負に）荷電した側鎖を有するアミノ酸、または芳香族側鎖を有するアミノ酸などのアミノ酸のクラス、または一部の特異的なクラスまたはサイズの側鎖などに選択的なものであってよいことも理解されるべきである。

10

#### 【0400】

特定の実施形態では、結合性物質は、目的の巨大分子に対して高い親和性および高い選択性を有する。特に、低い解離速度で高い結合親和性を有することがコーディングタグと記録タグの間の情報移行に効果的である。ある特定の実施形態では、結合性物質の $K_d$ は、 $< 10 \text{ nM}$ 、 $< 5 \text{ nM}$ 、 $< 1 \text{ nM}$ 、 $< 0.5 \text{ nM}$ 、または $< 0.1 \text{ nM}$ である。特定の実施形態では、結合性物質を巨大分子に、結合が完了するように駆動するために、結合性物質の $K_d$ の $> 10 \times$ 、 $> 100 \times$ 、または $> 1000 \times$ の濃度で添加する。抗体の単一のタンパク質分子への結合カインेटクスに関する詳細な考察は、Changら(Chang, Rissinら、2012年)に記載されている。

20

#### 【0401】

結合性物質の、ペプチドの小さなN末端アミノ酸(NTAA)に対する親和性を増大させるために、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、DNP基をNTAAのアミン基に付着させるサンガー試薬であるジニトロフルオロベンゼン(DNFB)を使用して周期的な配列決定手法で実行することができる。商業的な抗DNP抗体は、低nM範囲(約8 nM、LO-DNP-2)で親和性を有する(Bilgicer, Thomasら、2009年)；そのように、DNPで(DNFBを介して)修飾されたいくつものNTAAに対して親和性が高いNTAA結合性物質を工学的に操作し、同時に特定のNTAAに対する良好な結合選択性を実現することが可能であるはずなのは当然である。別の例では、NTAAを、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)で修飾することができる。アセチル基またはアミジニル(グアニジニル)基などの代替的なNTAA修飾因子を用いて同様の親和性の増強を実現することもできる。

30

#### 【0402】

ある特定の実施形態では、結合性物質は、ペプチド分子のNTAA、CTAA、介在するアミノ酸、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)、またはより高次のペプチドに結合し得る。一部の実施形態では、結合性物質のライブラリー内の各結合性物質は、特定のアミノ酸、例えば、20種の標準の天然に存在するアミノ酸のうちの1種に選択的に結合する。標準の、天然に存在するアミノ酸としては、アラニン(AまたはAla)、システイン(CまたはCys)、アスパラギン酸(DまたはAsp)、グルタミン酸(EまたはGlu)、フェニルアラニン(FまたはPhe)、グリシン(GまたはGly)、ヒスチジン(HまたはHis)、イソロイシン(IまたはIle)、リシン(KまたはLys)、ロイシン(LまたはLeu)、メチオニン(MまたはMet)、アスパラギン(NまたはAsn)、プロリン(PまたはPro)、グルタミン(QまたはGln)、アルギニン(RまたはArg)、セリン(SまたはSer)、トレオニン(TまたはThr)、バリン(VまたはVal)、トリプトファン(WまたはTrp)、およびチロシン(YまたはTyr)が挙げられる。

40

50

## 【0403】

ある特定の実施形態では、結合性物質は、アミノ酸の翻訳後修飾に結合し得る。一部の  
実施形態では、ペプチドは、1つまたは複数の翻訳後修飾を含み、これは、同じものであ  
っても異なるものであってもよい。ペプチドのNTAA、CTAA、介在するアミノ酸、  
またはこれらの組合せを翻訳後に修飾することができる。アミノ酸に対する翻訳後修飾と  
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、ジフタミド形成、ジスルフ  
イド架橋形成、エリミニル化、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、グリコシルホスファチジルイノシトール付加（  
glypiation）、ヘムC付着、ヒドロキシル化、ハイブシン形成、ヨウ素化、イ  
ソプレニル化、脂質付加、リポイル化、マロニル化、メチル化、ミリストイル化、酸化、  
パルミトイル化、ペグ化、ホスホパンテテイル化、リン酸化、プレニル化、プロピオニ  
ル化、レチニリデンシッフ塩基形成、S - グルタチオン化、S - ニトロシル化、S - スル  
フェニル化、セレン化、サクシニル化、スルフィン化、ユビキチン化、およびC末端アミ  
ド化が挙げられる（SeoおよびLee、2004年、J. Biochem. Mol. Biol.、37巻：35～44頁も参照されたい）。

10

## 【0404】

ある特定の実施形態では、タンパク質、ポリペプチド、またはペプチドのグリコシル化  
の状態を検出するための結合性物質としてレクチンを使用する。レクチンは、遊離の炭水  
化物また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  
- 結合性タンパク質である。種々のグリコシル化の状態（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）を  
認識するレクチンの一覧は、A、AAA、AAL、ABA、ACA、ACG、ACL、A  
OL、ASA、BanLec、BC2L - A、BC2LCN、BPA、BPL、Cal s  
epa、CGL2、CNL、Con、ConA、DBA、Discoidin、DSA、  
ECA、EEL、F17AG、Gal1、Gal1 - S、Gal2、Gal3、Gal3  
C - S、Gal7 - S、Gal9、GNA、GRFT、GS - I、GS - II、GSL -  
I、GSL - II、HHL、HIHA、HPA、I、II、Jacalin、LBA、L  
CA、LEA、LEL、Lentil、Lotus、LSL - N、LTL、MAA、MA  
H、MAL\_\_I、Malectin、MOA、MPA、MPL、NPA、Orysa ta  
、PA - IIL、PA - IL、PALa、PHA - E、PHA - L、PHA - P、PHA  
E、PHAL、PNA、PPL、PSA、PSL1a、PTL、PTL - I、PWM、R  
CA120、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を含む（Zhangら、2016年、M  
ABS、8巻：524～535頁を参照されたい）。

20

30

## 【0405】

ある特定の実施形態では、結合性物質は、修飾または標識されたNTAAに結合し得る  
。修飾または標識されたNTAAは、PITC、1 - フルオロ - 2, 4 - ジニトロベンゼ  
ン（サンガー試薬、DNFB）、ダンシルクロリド（DNS - Cl、もしくは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）、アセチル化試薬、グアニジン化試薬、チオアシル化試薬、チオア  
セチル化試薬、またはチオベンジル化試薬で標識されているものであってよい。

40

## 【0406】

ある特定の実施形態では、結合性物質は、アプタマー（例えば、ペプチドアプタマー、  
DNAアプタマー、またはRNAアプタマー）、抗体、アンチカリン、ATP依存性Clp  
プロテアーゼアダプタータンパク質（ClpS）、抗体結合性断片、抗体模倣物、ペプ  
チド、ペプチド模倣物、タンパク質、またはポリヌクレオチド（例えば、DNA、RNA  
、ペプチド核酸（PNA）、PNA、架橋核酸（BNA）、異種核酸（XNA）、グリ  
セロール核酸（GNA）、またはトレオース核酸（TNA）、またはそのバリエーション）で

50

あってよい。

【0407】

本明細書で使用される場合、抗体 (antibody) および抗体 (antibodies) という用語は、広範な意味で使用され、インタクトな抗体分子、例えば、これだけに限定されないが、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、および免疫グロブリン 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の任意の免疫反応性成分 (複数可) も含む。抗体は、天然に存在するものであってもよく、合成的に作製されたものであってもよく、組換えによって発現させたものであってもよい。抗体は、融合タンパク質であってよい。抗体は、抗体模倣物であってよい。抗体の例としては、これだけに限定されないが、Fab 断片、Fab' 断片、F(ab')<sub>2</sub> 断片、単鎖抗体断片 (scFv)、ミニボディ (miniantibody)、ダイアボディ (diabody)、架橋結合した抗体断片、Affibody (商標)、ナノボディ、単ドメイン抗体、DVD-Ig 分子、アルファボディ、アフィマー (affimer)、アフィチン (affitin)、サイクロチド、分子などが挙げられる。抗体工学またはタンパク質工学技法を使用して得られる免疫反応性産物もまた、明白に抗体という用語の意味の範囲内に入る。関連するプロトコルを含めた抗体および/またはタンパク質工学の詳細な説明は、他の場所の中でも、J. Maynard および G. Georgiou、2000 年、Ann. Rev. Biomed. Eng.、2 巻: 339~76 頁; Antibody Engineering、R. Kontermann および S. Dubel 編、Springer Lab Manual、Springer Verlag (2001 年); 米国特許第 5,831,012 号; および S. Paul、Antibody Engineering Protocols、Humana Press (1995 年) において見いだすことができる。

【0408】

抗体と同様に、ペプチドを特異的に認識する核酸およびペプチドアプタマーは、公知の方法を使用して作製することができる。アプタマーは、標的分子に、高度に特異的な、コンフォメーション依存性様式で結合し、一般には、非常に高い親和性を有するが、より低い結合親和性を有するアプタマーも所望であれば選択することができる。アプタマーは、標的間を、メチル基またはヒドロキシル基が存在するかしないかなどの非常に小さな構造的差異に基づいて区別することが示されており、また、ある特定のアプタマーは、D-鏡像異性体と L-鏡像異性体を区別することができる。薬物、金属イオン、および有機色素、ペプチド、ビオチン、ならびに、これだけに限定されないが、ストレプトアビジン、VEGF、およびウイルスタンパク質を含めたタンパク質を含めた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、ビオチン化の後、フルオレセイン標識の後、ならびにガラス表面およびマイクロスフェアに付着した際に機能活性を保持することが示されている (Jayasena、1999 年、Clin Chem、45 巻: 1628~50 頁; Kusser、2000 年、J. Biotechnol.、74 巻: 27~39 頁; Colas、2000 年、Curr Opin Chem Biol、4 巻: 54~9 頁を参照されたい)。アルギニンおよび AMP に特異的に結合するアプタマーも記載されている (Patel および Suri、2000 年、J. Biotech.、74 巻: 39~60 頁を参照されたい)。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が Goldra (1995 年、Ann. Rev. Biochem.、64 巻: 763~97 頁) 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 および Breaker、2011 年、RNA Biol.、8 巻: 82~89 頁; Mannironi ら、2000 年、RNA、6 巻: 520~27 頁; Famulok、1994 年、J. Am. Chem. Soc.、116 巻: 1698~1706 頁)。

【0409】

結合性物質は、天然に存在するまたは合成的に作製されたタンパク質を、アミノ酸配列内に 1 つまたは複数の変異が導入されるように遺伝子工学によって修飾して、特定の巨大分子の成分もしくは特徴 (例えば、NTAA、CTAA、または翻訳後修飾されたアミノ

10

20

30

40

50



高い特異性、またはその両方のために工学的に操作することができる。一部の実施形態では、結合性物質を、ファージディスプレイを使用した有望な親和性足場の定向進化によって開発することができる。

#### 【0413】

個々のまたは小さな群の標識（ビオチン化）されたNTAAに結合し、それを切断する工学的に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（その全体が参照により組み込まれる、PCT公開第WO2010/065322号を参照されたい）。アミノペプチダーゼは、タンパク質またはペプチドの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、非常に限られた特異性を有し、一般的に、N末端アミノ酸を前進的に切断し、アミノ酸を次々に切断する（Kishorら、2015年、Anal. Biochem., 488巻：6～8頁）。しかし、残基特異的アミノペプチダーゼが同定されている（Eriquezら、J. Clin. Microbiol., 1980年、12巻：667～71頁；Wilceら、1998年、Proc. Natl. Acad. Sci. USA、95巻：3472～3477頁；Liaoら、2004年、Prot. Sci., 13巻：1802～10頁）。アミノペプチダーゼを、特定の部分（例えば、PTC、DNP、SNPなど）で標識された、標準のアミノ酸を表す20種の異なるNTAAに特異的に結合するように工学的に操作することができる。ペプチドのN末端の段階的分解の制御は、標識の存在下でのみ活性な（例えば、結合活性または触媒活性）、工学的に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実現される。別の例では、Havranakら（米国特許公開第2014/0273004号）が、アミノアシルtRNA合成酵素（aaRS）を特異的なNTAA結合物質として工学的に操作することに関して記載している。aaRSのアミノ酸結合ポケットは、同類のアミノ酸に結合する内因性の能力を有するが、一般に、不十分な結合親和性および特異性を示す。さらに、これらの天然のアミノ酸結合物質はN末端標識を認識しない。aaRS足場の定向進化を使用して、N末端標識に関してN末端アミノ酸を認識する、親和性がより高く、特異性がより高い結合性物質を生成することができる。

#### 【0414】

別の例では、高度に選択的な工学的に操作されたClpSも文献に記載されている。Emilliらは、E. coli ClpSタンパク質のファージディスプレイによる定向進化の結果、NTAAにアスパラギン酸、アルギニン、トリプトファン、およびロイシン残基に対して選択的に結合する能力を有する4つの異なるバリエーションがもたらされることを記載している（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）。

#### 【0415】

特定の実施形態では、アンチカリンを、標識されたNTAA（例えば、DNP、SNP、アセチル化など）に対する高い親和性および高い特異性の両方について工学的に操作する。アンチカリン足場のある特定の變形は、ベータバレル構造に起因して、単一のアミノ酸への結合に適した形状を有する。N末端アミノ酸（修飾を伴うまたは伴わない）は、この「ベータバレル」バケットに潜在的に適合し、認識され得る。工学的に操作された新規の結合活性を有する親和性が高いアンチカリンが記載されている（Skerra、2008年、FEBS J., 275巻：2677～2683頁によって概説されている）。例えば、フルオレセインおよびジゴキシゲニンに対して親和性が高い結合性（低nM）を有するアンチカリンが工学的に操作されている（GebauerおよびSkerra、2012年）。新しい結合機能のために代替的足場を工学的に操作することについては、Bantaら（2013年、Annu. Rev. Biomed. Eng., 15巻：93～113頁）によっても概説されている。

#### 【0416】

所与の一価の結合性物質の機能的親和性（結合活性）を、一価の結合性物質の二価またはより高次の多量体を使用することによって少なくとも1桁分だけ増大させることができる（VauckelinおよびCharlton、2013年）。結合活性とは、多数の同時に存在する非共有結合性の相互作用の蓄積された強度を指す。個々の結合相互作用は

容易に解離し得る。しかし、多数の結合相互作用が同時に存在する場合、単一の結合相互作用の一過性の解離では結合性タンパク質は発散せず、結合相互作用は回復する可能性がある。結合性物質の結合活性を増大させるための代替的方法是、結合性物質に付着させるコーディングタグと巨大分子に付随させる記録タグに相補配列を含めることである。

#### 【0417】

一部の実施形態では、修飾されたC末端アミノ酸（CTAA）に選択的に結合する結合性物質を利用することができる。カルボキシペプチダーゼは、遊離のカルボキシル基を含有する末端アミノ酸を切断するプロテアーゼである。いくつかのカルボキシペプチダーゼが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、アルギニンおよびリシンなどの塩基性アミノ酸において優先的に切断する。カルボキシペプチダーゼを改変して、特定のアミノ酸に選択的に結合する結合性物質を創出することができる。一部の実施形態では、カルボキシペプチダーゼを、修飾部分ならびにCTAAのアルファ炭素R基のどちらにも選択的に結合するように工学的に操作することができる。したがって、工学的に操作されたカルボキシペプチダーゼは、C末端標識に関連して標準のアミノ酸を表す20種の異なるCTAAを特異的に認識し得る。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な（例えば、結合活性または触媒活性）工学的に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実現される。一実施例では、CTAAをパラ-ニトロアニリド基または7-アミノ-4-メチルクマリニル基によって修飾することができる。

#### 【0418】

本明細書に記載の方法において使用するための結合物質を生成するために工学的に操作することができる他の潜在的な足場としては、アンチカリン、アミノ酸tRNA合成酵素（aaRS）、ClpS、Affilin（登録商標）、Adnectin（商標）、T細胞受容体、ジンクフィンガータンパク質、チオレドキシン、GSTA1-1、DARPin、アフィマー、アフィチン、アルファボディ（alphabody）、アビマー（avimer）、クニツドメインペプチド、モノボディ（monobody）、単ドメイン抗体、EETI-II、HPSTI、細胞内抗体、リボカリン、PHD-フィンガー、V(NAR)LDTI、エビボディ（evibody）、Ig(NAR)、ノッチン、マキシボディ（maxibody）、ネオカルジノスタチン、pVIIII、テンダミスタット、VLR、プロテインA足場、MTI-II、エコチン、GCN4、Im9、クニツドメイン、ミクロボディ、PBP、トランスボディ（trans-body）、テトラネクチン（tetranectin）、WWドメイン、CBM4-2、DX-88、GFP、iMab、Ldl受容体ドメインA、Min-23、PDZ-ドメイン、トリ隣臓ポリペプチド、カリブドトキシノ/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）、またはSrc相同性ドメイン2（SH2）が挙げられる。

#### 【0419】

結合性物質を、より高い温度および穏やかな変性条件（例えば、尿素、グアニジンチオシアネート、イオン溶液の存在など）に耐えるように工学的に操作することができる。変性剤の使用は、結合性物質の直鎖ペプチドエピトープへの結合に干渉する可能性がある、表面に結合したペプチドの二次構造、例えば、 $\alpha$ -ヘリックス構造、 $\beta$ -ヘアピン、 $\beta$ -鎖、および他のこのような構造などを低減するのに役立つ。一実施形態では、結合サイクル中にペプチド二次構造を低減するために1-エチル-3-(3-ジメチルイミダゾリウム)アセレート（[EMIM]+[ACE]）などのイオン性液体を使用する（Lesch、Heuerら、2015年）。

#### 【0420】

記載されている任意の結合性物質はまた、結合性物質に関する識別情報を含有するコーディングタグも含む。コーディングタグは、それが付随する結合性物質に関する一意の識別情報をもたらす約3塩基～約100塩基の核酸分子である。コーディングタグは、約3

10

20

30

40

50

～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、または約3塩基～約8塩基を含み得る。一部の実施形態では、コーディングタグは、約3塩基、4塩基、5塩基、6塩基、7塩基、8塩基、9塩基、10塩基、11塩基、12塩基、13塩基、14塩基、15塩基、16塩基、17塩基、18塩基、19塩基、20塩基、25塩基、30塩基、35塩基、40塩基、55塩基、60塩基、65塩基、70塩基、75塩基、80塩基、85塩基、90塩基、95塩基、または100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、またはこれらの組合せで構成されるものであってよい。ポリヌクレオチド類似体は、PNA、PNA、BNA、GNA、TNA、LNA、モルホ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、および7-デアザプリン類似体を含む。

#### 【0421】

コーディングタグは、付随する結合性物質に関する識別情報をもたらすエンコーダー配列を含む。エンコーダー配列は、約3塩基～約30塩基、約3塩基～約20塩基、約3塩基～約10塩基、または約3塩基～約8塩基である。一部の実施形態では、エンコーダー配列は、約3塩基、4塩基、5塩基、6塩基、7塩基、8塩基、9塩基、10塩基、11塩基、12塩基、13塩基、14塩基、15塩基、20塩基、25塩基、または30塩基の長さである。エンコーダー配列の長さにより、生成され得る一意のエンコーダー配列の数が決定される。コード配列が短いほど、生成される一意のコード配列の数が少なくなり、これは、少数の結合性物質を使用する場合に有用であり得る。巨大分子の集団を解析する場合にはより長いエンコーダー配列が望ましい可能性がある。例えば、5塩基のエンコーダー配列は式5'-NNNNN-3'（配列番号135）（式中、Nは任意の天然に存在するヌクレオチドまたは類似体であってよい）を有する。4種の天然に存在するヌクレオチドA、T、C、およびGを使用すると、5塩基の長さを有する一意のエンコーダー配列の総数は1,024になる。一部の実施形態では、例えば、塩基が全て同一であるか、少なくとも3つの連続した塩基が同一であるか、またはその両方であるエンコーダー配列を除くことにより、一意のエンコーダー配列の総数を減らすことができる。特定の実施形態では、50種の一意のエンコーダー配列のセットを結合性物質ライブラリーに使用する。

#### 【0422】

一部の実施形態では、コーディングタグまたは記録タグの識別成分、例えば、エンコーダー配列、バーコード、UMI、コンパートメントタグ、分配バーコード、試料バーコード、空間的領域バーコード、サイクル特異的配列またはそれらの任意の組合せを、Hamming距離、Lee距離、非対称Lee距離、Reed-Solomon、Levenshtein-Tenengolts、または同様のエラー訂正方法に供する。Hamming距離は、長さが等しい2つのつながりの間の異なる位置の数を指す。Hamming距離により、1つのつながりを他のつながりに変えるために必要な置換の最小数が測定される。Hamming距離は、妥当な距離があいたエンコーダー配列を選択することによってエラーを訂正するために使用することができる。したがって、エンコーダー配列が5塩基である例では、使用できるエンコーダー配列の数は256種の一意のエンコーダー配列に減少する（144エンコーダー配列のHamming距離=256種のエンコーダー配列）。別の実施形態では、エンコーダー配列、バーコード、UMI、コンパートメントタグ、サイクル特異的配列、またはそれらの任意の組合せを、周期的な脱コーディングプロセスによって容易に読み取られるように設計する（Gundersen、2004年、Genome Res.、14巻：870～7頁）。別の実施形態では、エンコーダー配列、バーコード、UMI、コンパートメントタグ、分配バーコード、空間バーコード、試料バーコード、サイクル特異的配列、またはそれらの任意の組合せを、一塩基分解能が必要なのではなく、多数の塩基のワード（約5～20塩基の長さ）を読み取る必要があるため、正確度の低いナノポアシーケンシングによって読み取られるように設計する。本開示の方法において使用することができる15merのエラー訂正Hammingバーコ

ードのサブセットは配列番号 1 ~ 65 に記載されており、それらの対応する逆相補的な配列は配列番号 66 ~ 130 に記載されている。

#### 【0423】

一部の実施形態では、結合性物質のライブラリー内の一意の結合性物質のそれぞれが一意のエンコーダー配列を有する。例えば、20種の一意のエンコーダー配列を20種の標準のアミノ酸に結合する20種の結合性物質のライブラリーに使用することができる。修飾されたアミノ酸（例えば、翻訳後修飾されたアミノ酸）を識別するために追加的なコーディングタグ配列を使用することができる。別の例では、30種の一意のエンコーダー配列を20種の標準のアミノ酸および10種の翻訳後修飾されたアミノ酸（例えば、リン酸化アミノ酸、アセチル化アミノ酸、メチル化アミノ酸）に結合する30種の結合性物質のライブラリーに使用することができる。他の実施形態では、2種またはそれよりも多くの異なる結合性物質が同じエンコーダー配列を共有してよい。例えば、それぞれが異なる標準のアミノ酸に結合する2種の結合性物質が同じエンコーダー配列を共有してよい。

10

#### 【0424】

ある特定の実施形態では、コーディングタグは、一方の末端または両方の末端にスペーサー配列をさらに含む。スペーサー配列は、約1塩基~約20塩基、約1塩基~約10塩基、約5塩基~約9塩基、または約4塩基~約8塩基である。一部の実施形態では、スペーサーは、約1塩基、2塩基、3塩基、4塩基、5塩基、6塩基、7塩基、8塩基、9塩基、10塩基、11塩基、12塩基、13塩基、14塩基、15塩基または20塩基の長さである。一部の実施形態では、コーディングタグ内のスペーサーは、エンコーダー配列よりも短い、例えば、エンコーダー配列よりも少なくとも1塩基、2塩基、3塩基、4塩基、5塩基、6塩基、7塩基、8塩基、9塩基、10塩基、11塩基、12塩基、13塩基、14塩基、15塩基、20塩基、または25塩基短い。他の実施形態では、コーディングタグ内のスペーサーは、エンコーダー配列と同じ長さである。ある特定の実施形態では、スペーサーは、結合性物質特異的であり、したがって、前の結合サイクルからのスペーサーは現行の結合サイクルにおいて適切な結合性物質からのスペーサーとのみ相互作用する。例は、両方の抗体が巨大分子に逐次的に結合する場合にのみ情報移行を可能にするスペーサー配列を含有する同類の抗体の対である。スペーサー配列は、プライマー伸長反応のためのプライマーアニーリング部位、またはライゲーション反応における副子もしくは粘着末端として使用することができる。コーディングタグ上の5'スペーサー（図5A、' \* S p ' を参照されたい）は、 $T_m$ を上昇させるために、記録タグ上の3'スペーサーに対する偽性相補的塩基を任意選択で含有してよい（Lehoucqら、2008年、Nucleic Acids Res.、36巻：3409~3419頁）。

20

30

#### 【0425】

一部の実施形態では、結合性物質の集合内のコーディングタグは、アッセイに使用される共通のスペーサー配列を共有する（例えば、多数の結合サイクル方法に使用される結合性物質のライブラリー全体が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）。別の実施形態では、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で構成される。他の実施形態では、結合性物質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スペーサー配列を有する。一部の実施形態では、コーディングタグは、結合サイクル特異的スペーサー配列を1つ含む。例えば、第1の結合サイクルで使用される結合性物質に対するコーディングタグは「サイクル1」特異的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性物質に対するコーディングタグは「サイクル2」特異的スペーサー配列を含み、「n」回の結合サイクルまで同様である。さらなる実施形態では、第1の結合サイクルで使用される結合性物質に対するコーディングタグは「サイクル1」特異的スペーサー配列および「サイクル2」特異的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性物質に対するコーディングタグは「サイクル2」特異的スペーサー配列および「サイクル3」特異的スペーサー配列を含み、「n」回の結合サイクルまで同様である。この実施形態は、結合サイクルが完了した後の非連鎖状伸長記録タグのPCRアセンブリに有用である（図10を参照されたい）。一部の実施形態で

40

50



は、スペーサー配列は、プライマー伸長反応また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または伸長記録タグ内の相補的なスペーサー配列とアニーリングするのに十分な数の塩基を含む。

#### 【0426】

記録タグの集団が巨大分子に付随する場合、サイクル特異的スペーサー配列を使用してコーディングタグの情報を単一の記録タグ上に連鎖状にすることもできる。第1の結合サイクルでコーディングタグからランダムに選択された記録タグに情報を移行させ、その後の結合サイクルではサイクル依存性スペーサー配列を使用して伸長記録タグのみをプライミングすることができる。より詳細には、第1の結合サイクルで使用される結合性物質に対するコーディングタグは「サイクル1」特異的スペーサー配列および「サイクル2」特異的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性物質に対するコーディングタグは「サイクル2」特異的スペーサー配列および「サイクル3」特異的スペーサー配列を含み、「n」回の結合サイクルまで同様である。第1の結合サイクルからの結合性物質のコーディングタグは、相補的なサイクル1特異的スペーサー配列を介して記録タグとアニーリングすることが可能である。コーディングタグ情報が記録タグに移行したら、結合サイクル1の最後に、サイクル2特異的スペーサー配列が伸長記録タグの3'末端に位置する。第2の結合サイクルからの結合性物質のコーディングタグは、相補的なサイクル2特異的スペーサー配列を介して伸長記録タグとアニーリングすることが可能である。コーディングタグ情報が伸長記録タグに移行したら、結合サイクル2の最後に、サイクル3特異的スペーサー配列が伸長記録タグの3'末端に位置し、「n」回の結合サイクルまで同様である。この実施形態は、多数の結合サイクルの中で特定の結合サイクルにおける結合情報の移行が前の結合サイクルを経た（伸長）記録タグ上でのみ起こるものとする。しかし、時には、結合性物質は同類の巨大分子に結合し損ねる。「追跡」ステップとして各結合サイクル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、結合サイクルの事象が失敗したとしても結合サイクルを同期させたままにすることができる。例えば、同類結合性物質が結合サイクル1の間に巨大分子に結合し損ねた場合、結合サイクル1後に、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、および「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する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、または、好ましくは、「ヌル」結合サイクルを正に識別する特異的なバーコードが存在しないことであってよい。「ヌル」オリゴヌクレオチドは、記録タグとサイクル1特異的スペーサーを介してアニーリングすることが可能であり、サイクル2特異的スペーサーが記録タグに移行される。したがって、結合サイクル1事象の失敗にもかかわらず、結合サイクル2からの結合性物質が伸長記録タグとサイクル2特異的スペーサーを介してアニーリングすることが可能である。「ヌル」オリゴヌクレオチドにより、伸長記録タグ内で結合サイクル1に結合事象失敗の印が付けられる。

#### 【0427】

好ましい実施形態では、結合サイクル特異的エンコーダー配列をコーディングタグに使用する。結合サイクル特異的エンコーダー配列は、完全に一意の分析物（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ことによって、または分析物（例えば、NTAA）エンコーダー配列とサイクル特異的バーコードを接合させた組合せ使用によってのいずれかで実現することができる（図35を参照されたい）。組合せ手法を使用することの利点は、設計する必要がある総バーコードがより少なくなることである。10サイクルにわたって使用する20種の分析物結合性物質のセットに対しては、20種の分析物エンコーダー配列バーコードおよび10種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性物質エンコーダー配列に直接埋め込まれる場合には、合計200種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、ナノポア読み取りにエラー訂正バーコードを使用する場合にコーディングタグの全長を最小化することができることである。エラー耐性バーコードの使用により、配列決定プラットフォーム

10

20

30

40

50

ームおよびよりエラーが生じやすい手法を使用して高度に正確なバーコード識別が可能になるが、迅速な解析スピード、低費用、および/またはよりポータブルな器械使用などの他の利点がある。そのような例の1つは、ナノポアに基づく配列決定読み取りである。

#### 【0428】

一部の実施形態では、コーディングタグは、結合性物質の近位にある第2の(3')スパーサー配列内の切断可能なまたはニッキング可能なDNA鎖を含む(図32を参照されたい)。例えば、3'スパーサーは、ウラシル特異的切除試薬(USER)によってニッキングすることができる1つまたは複数のウラシル塩基を有してよい。USERにより、ウラシルの位置に一ヌクレオチドギャップが生成される。別の例では、3'スパーサーは、2重鎖の一方の鎖のみを加水分解するニッキングエンドヌクレアーゼの認識配列を含んでよい。3'スパーサー配列を切断またはニッキングするために使用する酵素は、一方のDNA鎖のみ(コーディングタグの3'スパーサー)に作用し、したがって、(伸長)記録タグに属する2重鎖内の他方の鎖は intact なまま残されることが好ましい。これらの実施形態は、プライマー伸長が起こった後に結合性物質を(伸長)記録タグから変性によらずに除去し、その後の結合サイクルに利用可能な一本鎖DNAスパーサー配列を伸長記録タグ上に残すことが可能になるので、タンパク質をそれらのネイティブなコンフォメーションで解析するアッセイにおいて特に有用である。

#### 【0429】

コーディングタグは、パリンドローム配列を含有するように設計することもできる。コーディングタグにパリンドローム配列を含めることにより、新生の成長している伸長記録タグが、コーディングタグ情報が移行するに従ってそれ自体でフォールディングすることが可能になる。伸長記録タグはより緻密な構造にフォールディングし、望ましくない分子間結合およびプライマー伸長事象が有効に減少する。

#### 【0430】

一部の実施形態では、コーディングタグは、同じ分析物を認識する結合性物質を用いて以前に伸長した記録タグ上でのみプライミング伸長することが可能な分析物特異的スパーサーを含む。伸長記録タグは、分析物特異的スパーサーおよびエンコーダー配列を含むコーディングタグを使用して一連の結合事象から組み立てることができる。一実施形態では、第1の結合事象では、一般的な3'スパーサープライマー配列および次の結合サイクルで使用するための5'末端の分析物特異的スパーサー配列で構成されるコーディングタグを伴う結合性物質を使用する; 次いで、その後の結合サイクルでは、コードされる分析物特異的3'スパーサー配列を伴う結合性物質を使用する。この設計により、正確な一連の同類結合事象のみから創出される増幅可能なライブラリーエレメントがもたらされる。オフターゲットのおよび交差反応性結合相互作用により、増幅可能でない伸長記録タグが導かれる。一実施例では、同類結合性物質と特定の巨大分子分析物の対を2つの結合サイクルに使用して分析物を識別する。第1の同類結合性物質は、一般的な記録タグのスパーサー配列上でのプライミング伸長のための一般的なスパーサー3'配列および次の結合サイクルで使用する5'末端のコードされる分析物特異的スパーサーで構成されるコーディングタグを含有する。対応する同類結合性物質対に関しては、第2の結合性物質の3'分析物特異的スパーサーと第1の結合性物質の5'分析物特異的スパーサーを対応させる。このように、結合性物質の同類の対の正確な結合によってのみ、増幅可能な伸長記録タグがもたらされる。交差反応性結合性物質は記録タグ上でプライミング伸長することができず、増幅可能な伸長記録タグ産物は生成しない。この手法では、本明細書に開示されている方法の特異性が著しく増強される。同じ原理を、3つの結合サイクルを使用するトリプレット結合性物質セットに適用することができる。第1の結合サイクルでは、記録タグ上の一般的な3'Sp配列と結合性物質コーディングタグ上の一般的なスパーサーを相互作用させる。プライマー伸長により、分析物特異的5'スパーサーを含めたコーディングタグ情報を記録タグに移行させる。その後の結合サイクルでは、結合性物質のコーディングタグ上の分析物特異的スパーサーを使用する。

#### 【0431】

10

20

30

40

50

ある特定の実施形態では、コーディングタグは、コーディングタグが連結した結合性物質についての一意の分子識別子をさらに含んでよい。結合性物質についてのUMIは、配列決定読み取りのために伸長コーディングタグまたはジタグ分子を利用し、エンコーダー配列と組み合わせて、結合性物質の同一性および巨大分子に対する一意の結合事象の数に関する情報をもたらす実施形態において有用であり得る。

#### 【0432】

別の実施形態では、コーディングタグは、ランダム化配列（Nのセット、ここで、N = A、C、G、Tからのランダムな選択、またはワードのセットからのランダムな選択である）を含む。一連の「n」回の結合サイクルおよびコーディングタグ情報の（伸長）記録タグへの移行の後、最終的な伸長記録タグ産物は、一連のこれらのランダム化配列で構成され、これは、最終的な伸長記録タグについての「複合性の」一意の分子識別子（UMI）を集散的に形成する。例えば、各コーディングタグが（NN）配列（ $4 \times 4 = 16$ の可能性がある配列）を含有する場合、10回の配列決定サイクル後に、分布した2merの組合せセットが10種形成され、可能性のある伸長記録タグ産物についての複合性のUMI配列1610～1012種という総多様性が創出される。ペプチド配列決定実験で約 $10^9$ 個の分子を使用するとすれば、この多様性は、配列決定実験のための有効なUMIのセットを創出するために十分すぎるほどである。多様性の増大は、単にコーディングタグ内により長いランダム化領域（NNN、NNNNなど）を使用することによって実現することができる。

#### 【0433】

コーディングタグは、3'スパーサー配列の3'末端に組み入れられたターミネーターヌクレオチドを含んでよい。結合性物質が巨大分子およびそれらの対応するコーディングタグに結合し、記録タグが相補的なスパーサー配列を介してアニーリングした後、情報をコーディングタグから記録タグに移行させるため、または情報を記録タグからコーディングタグに移行させるためにプライマー伸長することが可能になる。コーディングタグの3'末端にターミネーターヌクレオチドを付加することにより、記録タグ情報のコーディングタグへの移行が防止される。伸長コーディングタグの生成を伴う本明細書に記載の実施形態に関しては、コーディングタグ情報の記録タグへの移行を防止するために、記録タグの3'末端にターミネーターヌクレオチドを含めることが好ましい場合があることが理解される。

#### 【0434】

コーディングタグは、一本鎖分子であってもよく、二本鎖分子であってもよく、部分的に二本鎖であってもよい。コーディングタグは、平滑末端、突出末端、またはその一方を含んでよい。一部の実施形態では、コーディングタグは、部分的に二本鎖であり、それにより、コーディングタグが成長している伸長記録タグの内部のエンコーダーおよびスパーサー配列にアニーリングすることが防止される。

#### 【0435】

コーディングタグは、結合性物質に、共有結合性の相互作用および非共有結合性の相互作用を含めた当技術分野で公知の任意の手段を使用して直接または間接的に接合される。一部の実施形態では、コーディングタグを結合性物質に酵素的にまたは化学的に接合することができる。一部の実施形態では、コーディングタグを結合性物質にライゲーションによって接合することができる。他の実施形態では、コーディングタグを結合性物質に親和性結合対（例えば、ビオチンおよびスト렙トアビジン）を介して接合する。

#### 【0436】

一部の実施形態では、結合性物質とコーディングタグをSpy Catcher - Spy Tag相互作用によって接合する（図43Bを参照されたい）。Spy Tagペプチドは、Spy Catcherタンパク質と自発的なイソペプチド連結によって不可逆的な共有結合を形成し、それにより、力および厳しい条件に対して抵抗性であるペプチド相互作用を創出するための遺伝子によりコードされたやり方がもたらされる（Zakeriら、2012年、Proc. Natl. Acad. Sci.、109巻：E690～697頁；Liら、2014年、J. Mol. Biol.、426巻：309～317頁）。

結合性物質を、SpyCatcherタンパク質を含む融合タンパク質として発現させることができる。一部の実施形態では、SpyCatcherタンパク質を結合性物質のN末端またはC末端に付加する。SpyTagペプチドとコーディングタグを標準のコンジュゲーション化学を使用してカップリングすることができる(Bioconjugate Techniques、G. T. Hermanson、Academic Press (2013年))。

#### 【0437】

他の実施形態では、結合性物質とコーディングタグをSnoopTag-SnoopCatcherペプチド-タンパク質相互作用によって接合する。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(Veggi 10  
aniら、Proc. Natl. Acad. Sci. USA、2016年、113巻：1202～1207頁)。結合性物質を、SnoopCatcherタンパク質を含む融合タンパク質として発現させることができる。一部の実施形態では、SnoopCatcherタンパク質を結合性物質のN末端またはC末端に付加する。SnoopTagペプチドとコーディングタグを標準のコンジュゲーション化学を使用してカップリングすることができる。

#### 【0438】

さらに他の実施形態では、結合性物質とコーディングタグをHaloTag(登録商標)タンパク質融合タグとその化学的リガンドによって接合する。HaloTagは、合成リガンド(HaloTagリガンド)と共有結合するように設計された改変ハロアルカン 20  
デハロゲナーゼである(Losら、2008年、ACS Chem. Biol.、3巻：373～382頁)。合成リガンドは、種々の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の間に、高度に特異的であり、生理的条件下で迅速に起こり、基本的に不可逆的である共有結合が形成される。

#### 【0439】

ある特定の実施形態では、巨大分子を非同類結合性物質とも接触させる。本明細書で 30  
使用される場合、非同類結合性物質とは、考察されている特定の巨大分子とは異なる巨大分子の特徴または成分に対して選択的である結合性物質を指す。例えば、n NTAAがフェニルアラニンであり、ペプチドをそれぞれフェニルアラニン、チロシン、およびアスパラギンに対して選択的な3種の結合性物質と接触させる場合、フェニルアラニンに対して選択的な結合性物質が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可能な第1の結合性物質ということになり、一方、他の2種の結合性物質はそのペプチドに対して非同類結合性物質ということになる(フェニルアラニン以外のNTAAに対して選択的であるので)。しかし、チロシン結合性物質およびアスパラギン結合性物質は、試料中の他のペプチドに対して同類結合性物質になり得る。次いで、n NTAA(フェニルアラニン)をペプチドから切断し、それにより、ペプチドのn-1アミノ酸をn-1 NTAA(例えば、チロシン)に変換し、次いで、ペプチドを同じ3種の結合性物質と接触させると、チロシンに対して選択的な結合性物質が、n-1 NTAA(す 40  
なわち、チロシン)に選択的に結合することが可能な第2の結合性物質ということになり、一方、他の2種の結合性物質は非同類結合性物質ということになる(チロシン以外のNTAAに対して選択的であるので)。

#### 【0440】

したがって、作用物質が結合性物質であるか非同類結合性物質であるかは、結合のために現在利用可能な特定の巨大分子の特徴または成分の性質に依存することが理解されるべきである。また、多数の巨大分子を多重化反応において解析する場合、1つの巨大分子に対する結合性物質は別の巨大分子に対しては非同類結合性物質であり得、逆もまた同じである。したがって、結合性物質に関する以下の説明は本明細書に記載されているあらゆる型の結合性物質(すなわち、同類結合性物質および非同類結合性物質のどちらにも)に適用可能であることが理解されるべきである。

VIIII. コーディングタグ情報の記録タグへの周期的移行

10

20

30

40

50

## 【0441】

本明細書に記載の方法において、結合性物質が巨大分子に結合したら、その連結したコーディングタグの識別情報を巨大分子に付随する記録タグに移行させ、それにより、「伸長記録タグ」を生成する。伸長記録タグは、実施された各結合サイクルを表す、結合性物質のコーディングタグからの情報を含み得る。しかし、伸長記録タグはまた、例えば、結合性物質が巨大分子に結合し損ねたことが原因で、コーディングタグが見落とされた、損傷を受けた、または欠陥があることが原因で、プライマー伸長反応が失敗したことが原因で、「飛ばされた」結合サイクルを経る可能性もある。結合事象が起こったとしても、例えば、コーディングタグが損傷を受けたまたは欠陥があることが原因で、プライマー伸長反応にエラーが導入されたことが原因で、コーディングタグから記録タグへの情報の移行が不完全であるまたは100%未満の正確さになる可能性がある。したがって、伸長記録タグは、その付随する巨大分子において起こった結合事象の100%、または最大で95%、90%、85%、80%、75%、70%、65%、60%、65%、55%、50%、45%、40%、35%、30%を表す可能性がある。さらに、伸長記録タグ中に存在するコーディングタグ情報は、対応するコーディングタグに対して少なくとも30%、35%、40%、45%、50%、55%、60%、65%、70%、75%、80%、85%、90%、95%、または100%の同一性を有する可能性がある。

10

## 【0442】

ある特定の実施形態では、伸長記録タグは、多数の連続的な結合事象を表す、多数のコーディングタグからの情報を含む可能性がある。これらの実施形態では、単一の連鎖状の伸長記録タグは単一の巨大分子を表す可能性がある（図2Aを参照されたい）。本明細書で言及される通り、コーディングタグ情報の記録タグへの移行は、多数の連続的な結合事象を伴う方法において起こると思われる伸長記録タグへの移行も含む。

20

## 【0443】

ある特定の実施形態では、結合事象情報をコーディングタグから記録タグに周期的に移行させる（図2Aおよび2Cを参照されたい）。交差反応性結合事象を、配列決定後に、少なくとも2つの異なるコーディングタグを要求し、2つまたはそれよりも多くの独立した結合事象を識別し、同じ結合性物質のクラス（特定のタンパク質と同類）にマッピングすることにより、情報科学的にフィルターにかけて除去することができる。任意選択の試料バーコードまたはコンパートメントバーコードを記録タグに含めることができ、同じく任意選択のUMI配列も含めることができる。コーディングタグは、任意選択のUMI配列をエンコーダーおよびスペーサー配列と一緒に含有してもよい。ユニバーサルプライミング配列（U1およびU2）も増幅およびNGS配列決定のために伸長記録タグに含めることもできる（図2Aを参照されたい）。

30

## 【0444】

特定の結合性物質に付随するコーディングタグ情報を記録タグに種々の方法を使用して移行させることができる。ある特定の実施形態では、コーディングタグの情報を記録タグに、プライマー伸長によって移行させる（Chan、McGregorら、2015年）。記録タグまたは伸長記録タグの3'末端のスペーサー配列をコーディングタグの3'末端の相補的なスペーサー配列とアニーリングさせ、ポリメラーゼ（例えば、鎖置換ポリメラーゼ）により、アニーリングしたコーディングタグを鋳型として使用して記録タグ配列を伸長させる（図5～7を参照されたい）。一部の実施形態では、コーディングタグと伸長記録タグに存在する内部のエンコーダーおよびスペーサー配列のハイブリダイゼーションを防止するために、コーディングタグエンコーダー配列および5'スペーサーと相補的なオリゴヌクレオチドをコーディングタグとプレアニーリングさせることができる。一本鎖のままのコーディングタグ上の3'末端スペーサーを記録タグ上の末端3'スペーサーと結合させることが好ましい。他の実施形態では、コーディングタグと内部の部位とのアニーリングを防止するために、新生記録タグを一本鎖結合性タンパク質でコーディングすることができる。あるいは、完全に二本鎖のコーディングタグへの3'末端の侵入を容易にするために、新生記録タグをRecA（またはUvsXなどの関連する相同体）でコーディング

40

50

することもできる (Bellら、2012年、Nature、491巻: 274~278頁)。この形態により、二本鎖のコーディングタグが内部の記録タグエレメントと相互作用することが防止されるが、それでも、伸長記録タグのRecAコーティングされた3'尾部による鎖の侵入は起こりやすい (Bellら、2015年、Elife、4巻: e08646頁)。一本鎖結合性タンパク質が存在することにより、鎖置換反応が容易になり得る。

#### 【0445】

好ましい実施形態では、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限られているまたはそれを欠く。そのようなポリメラーゼの多数の例のいくつかとして、クレノウexo- (DNA Pol 1のクレノウ断片)、T4 DNAポリメラーゼexo-、T7 DNAポリメラーゼexo- (シーケナーゼ2.0)、Pfu exo-、Vent exo-、Deep Vent exo-、Bst DNAポリメラーゼ大断片exo-、Bca Pol、9°N Pol、およびPhi29 Pol exo-が挙げられる。好ましい実施形態では、DNAポリメラーゼは、室温および最大45℃までで活性である。別の実施形態では、「ウォームスタート」バージョンの好熱性ポリメラーゼを使用し、したがって、ポリメラーゼを約40~50℃で活性化し、使用する。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。

#### 【0446】

鎖置換複製に有用な添加剤としては、E. coliのSSBタンパク質、ファージT4遺伝子32産物、ファージT7遺伝子2.5タンパク質、ファージPf3 SSB、複製タンパク質A RPA32およびRPA14サブユニット (World、1997年)などの、細菌起源、ウイルス起源、または真核生物起源のいくつもの一本鎖DNA結合性タンパク質 (SSBタンパク質) のいずれか; アデノウイルスDNA結合性タンパク質、単純ヘルペスタンパク質ICP8、BMRF1ポリメラーゼアクセサリサブユニット、ヘルペスウイルスUL29 SSB様タンパク質などの他のDNA結合性タンパク質; ファージT7ヘリカーゼ/プライマーゼ、ファージT4遺伝子41ヘリカーゼ、E. coli Repヘリカーゼ、E. coli recBCDヘリカーゼ、recA、E. coliおよび真核生物トポイソメラーゼ (Champoux、2001年)などの、DNA複製に

#### 【0447】

記録タグの末端スペーサー配列により伸長自己伸長がプライミングされる場合などのミスプライミングまたは自己プライミング事象は、一本鎖結合性タンパク質 (T4遺伝子32、E. coli SSBなど)、DMSO (1~10%)、ホルムアミド (1~10%)、BSA (10~100 µg/ml)、TMAc1 (1~5 mM)、硫酸アンモニウム (10~50 mM)、ベタイン (1~3 M)、グリセロール (5~40%)、またはエチレングリコール (5~40%) をプライマー伸長反応に含めることによって最小化することができる。

#### 【0448】

大多数のA型ポリメラーゼは、3'エキソヌクレアーゼ活性を欠き (内因的または工学的除去)、例えば、クレノウexo-、T7 DNAポリメラーゼexo- (シーケナーゼ2.0)、およびTaqポリメラーゼは、2重鎖増幅産物の3'平滑末端へのヌクレオチド、好ましくはアデノシン塩基 (より低い程度でG塩基、配列の状況に依存する) の非鋳型付加を触媒する。Taqポリメラーゼに関しては、3'ピリミジン (C>T) により、非鋳型アデノシン付加が最小限になり、一方、3'プリンヌクレオチド (G>A) では非鋳型アデノシン付加が有利になる。プライマー伸長のためにTaqポリメラーゼを使用する実施形態では、コーディングタグにおいて、結合性物質から遠位のスペーサー配列と隣接するバーコード配列 (例えば、エンコーダー配列またはサイクル特異的配列) との間にチミジン塩基を置くことにより、記録タグのスペーサー配列の3'末端上の非鋳型アデノシンヌク

レオチドの散在的包含に適應させる（図43A）。このように、伸長記録タグ（非鋳型アデノシン塩基を有するまたは有さない）は、コーディングタグとアニーリングし、プライマー伸長を受けることが可能である。

【0449】

あるいは、特にO-ヘリックス領域における1つまたは複数の点変異によって非鋳型ターミナルトランスフェラーゼ活性を著しく低下させた変異ポリメラーゼ（中温性または好熱性）を使用することにより、非鋳型塩基の付加を減少させることができる（米国特許第7,501,237号を参照されたい）（Yang、Astatkera、2002年）。3'エキソヌクレアーゼが欠損しており、鎖置換能を有するPfu exo-も非鋳型ターミナルトランスフェラーゼ活性を有さない。

10

【0450】

別の実施形態では、最適なポリメラーゼ伸長緩衝液は、40～120mMの、例えばTris-酢酸、Tris-HCl、HEPESなどの緩衝剤、pH6～9で構成される。

【0451】

伸長記録タグの末端スペーサー配列と伸長記録タグの内部の領域の自己アニーリングによって開始される自己プライミング/ミスプライミング事象は、記録/伸長記録タグに偽相補的塩基を含めることによって最小化することができる（Lahoud、Timoshchukら、2008年）、（Hoshika、Chenら、2010年）。偽相補的塩基は、化学修飾が存在することに起因する互いの2重鎖の形成に対する有意に低下したハイブリダイゼーション親和性を示す。しかし、多くの偽相補的修飾塩基は、天然のDNAまたはRNA配列と強い塩基対を形成し得る。ある特定の実施形態では、コーディングタグスペーサー配列は、多数のA塩基およびT塩基で構成され、ホスホラミダイトオリゴヌクレオチド合成を使用して市販の偽相補的塩基2-アミノアデニンおよび2-チオチミンを記録タグに組み入れる。追加的な偽相補的塩基を、プライマー伸長中に、偽相補的ヌクレオチドを反応に添加することによって伸長記録タグに組み入れることができる（Gammer、Ararら、2006年）。

20

【0452】

溶液中のコーディングタグで標識された結合性物質と固定化されたタンパク質の記録タグとの非特異的な相互作用を最小限にするために、記録タグスペーサー配列と相補的な競合剤オリゴヌクレオチド（遮断オリゴヌクレオチドとも称される）を結合反応に添加して、非特異的な相互作用を最小限にする（図32A～D）。遮断オリゴヌクレオチドは、比較的短いものである。プライマー伸長前に過剰な競合オリゴヌクレオチドを結合反応から洗い流し、これにより、特にわずかな温度の上昇（例えば、30～50）に曝露させると、アニーリングした競合オリゴヌクレオチドが記録タグから有効に解離する。遮断オリゴヌクレオチドは、プライマー伸長を防止するために3'末端にターミネーターヌクレオチドを含んでよい。

30

【0453】

ある特定の実施形態では、記録タグ上のスペーサー配列とコーディングタグ上の相補的なスペーサー配列のアニーリングは、プライマー伸長反応条件下（すなわち、アニーリングTmが反応温度と同様である）で準安定である。これにより、コーディングタグのスペーサー配列で記録タグのスペーサー配列とアニーリングした任意の遮断オリゴヌクレオチドを置き換えることが可能になる。

40

【0454】

特定の結合性物質に付随するコーディングタグ情報は、ライゲーションによって記録タグに移行させることもできる（例えば、図6および7を参照されたい）。ライゲーションは、平滑末端ライゲーションであってもよく、粘着末端ライゲーションであってもよい。ライゲーションは、酵素的ライゲーション反応であってもよい。リガーゼの例としては、これだけに限定されないが、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、Electroligase（登録商標）が挙げられる。あるいは、ライゲーション

50

は、化学的ライゲーション反応であってよい（図7を参照されたい）。図において、スパーサーを欠くライゲーションは、「記録ヘルパー」配列とコーディングタグ上のアームのハイブリダイゼーションを使用することによって実現される。アニーリングした相補配列を、標準の化学的ライゲーションまたは「クリックケミストリー」を使用して化学的にライゲーションする（Gundersen、Huangら、1998年、Peng、Liら、2010年、El-Sagheer、Cheongら、2011年、El-Sagheer、Sanzoneら、2011年、Sharma、Kentら、2012年、RoloffおよびSeitz、2013年、Litovchick、Clarkら、2014年、Roloff、Fichtら、2014年）。

#### 【0455】

別の実施形態では、PNAの移行は、公開された技法を使用した化学的ライゲーションで実現することができる。PNAの構造は、5' N末端アミン基および非反応性3' C末端アミドを有するようなものである。PNAの化学的ライゲーションには、末端を化学的に活性になるように修飾することが必要である。これは、一般には、5' N末端をシステニル部分で誘導体化し、3' C末端をチオエステル部分で誘導体化することによってなされる。そのような修飾されたPNAは、標準のネイティブな化学的ライゲーション条件を使用することで容易にカップリングする（Roloffら、2013年、Bioorgan. Med. Chem.、21巻：3458～3464頁）。

#### 【0456】

一部の実施形態では、コーディングタグ情報を、トポイソメラーゼを使用して移行させることができる。トポイソメラーゼを使用して、記録タグ上のトポ荷電3' リン酸とコーディングタグまたはその相補物の5' 末端とライゲーションすることができる（Shumanら、1994年、J. Biol. Chem.、269巻：32678～32684頁）。

#### 【0457】

本明細書に記載の通り、結合性物質は、翻訳後修飾されたアミノ酸に結合し得る。したがって、ペプチド巨大分子を伴うある特定の実施形態では、伸長記録タグは、アミノ酸配列および翻訳後修飾に関するコーディングタグ情報を含む。一部の実施形態では、内部の翻訳後修飾されたアミノ酸（例えば、リン酸化、グリコシル化、サクシニル化、ユビキチン化、S-ニトロシル化、メチル化、N-アセチル化、脂質付加など）の検出は、末端アミノ酸（例えば、NTAAまたはCTAA）の検出および切断の前に実現される。一実施例では、ペプチドをPTM修飾のために結合性物質と接触させ、付随するコーディングタグ情報を上記の通り記録タグに移行する（図8Aを参照されたい）。アミノ酸修飾に関するコーディングタグ情報の検出および移行が完了したら、一次アミノ酸配列に関するコーディングタグ情報の検出および移行の前に、N末端またはC末端分解法を使用してPTM修飾基を除去することができる。したがって、得られた伸長記録タグにより、ペプチド配列における翻訳後修飾の存在が、逐次的順序ではないが、一次アミノ酸配列情報と併せて示される（図8Bを参照されたい）。

#### 【0458】

一部の実施形態では、内部の翻訳後修飾されたアミノ酸の検出は、一次アミノ酸配列の検出と並行して行うことができる。一実施例では、NTAA（またはCTAA）を、単独でまたは結合性物質のライブラリー（例えば、20種の標準のアミノ酸および選択された翻訳後修飾されたアミノ酸に対する結合性物質で構成されるライブラリー）の一部としての、翻訳後修飾されたアミノ酸に対して特異的な結合性物質と接触させる。末端アミノ酸切断および結合性物質（または結合性物質のライブラリー）との接触の連続的なサイクルを後に続ける。したがって、得られた伸長記録タグにより、一次アミノ酸配列に関して、翻訳後修飾の存在および順序が示される。

#### 【0459】

ある特定の実施形態では、コーディングタグ情報移行の全体的な頑強性および効率を改善するために、巨大分子ごとに記録タグのアンサンブルを使用することができる（例えば、図9を参照されたい）。単一の記録タグではなく、所与の巨大分子に付随する記録タグ

10

20

30

40

50



のアンサンブルを使用することにより、コーディングタグと記録タグのカップリング収率が潜在的に高いこと、およびライブラリーの全収率が高いことに起因して、ライブラリー構築の効率が改善される。単一の連鎖状の伸長記録タグの収率は、連鎖の段階的収率に直接依存し、一方、コーディングタグ情報を受容することが可能な多数の記録タグの使用では、指数関数的な連鎖の喪失を受けない。

#### 【0460】

そのような実施形態の例が図9および10に示されている。図9Aおよび10Aでは、固体支持体上で単一の巨大分子に多数の記録タグが付随している（空間的共同在または単一の巨大分子の単一のビーズへの限局によって）。結合性物質を固体支持体に周期的に曝露させ、各サイクルにおいて、それらの対応するコーディングタグにより、共同在する多数の記録タグのうちの1つに情報が移行される。図9Aに示されている例では、結合サイクル情報は、コーディングタグ上に存在するスペーサーにコードされている。各結合サイクルについて、結合性物質のセットに設計されたサイクル特異的スペーサー配列で印をつける（図9Aおよび9B）。例えば、NTAA結合性物質の場合では、同じアミノ酸残基に対する結合性物質を異なるコーディングタグで標識する、またはサイクル特異的情報をスペーサー配列に含めて、両方の結合性物質の同一性およびサイクル数を示す。

#### 【0461】

図9Aにおいて例示されている通り、第1の結合サイクル（サイクル1）において、複数のNTAA結合性物質を巨大分子と接触させる。サイクル1で使用する結合性物質は、記録タグのスペーサー配列と相補的な共通のスペーサー配列を有する。サイクル1で使用する結合性物質は、サイクル1特異的配列を含む3'-スペーサー配列も有する。結合サイクル1の間に、第1のNTAA結合性物質が巨大分子の遊離末端に結合し、第1のコーディングタグおよび記録タグ内の共通のスペーサー配列の相補配列がアニーリングし、第1のコーディングタグの情報が同類の記録タグに共通のスペーサー配列からのプライマー伸長を介して移行する。NTAAを除去して、新しいNTAAを露出させた後、結合サイクル2により、記録タグのスペーサー配列と相補的な共通のスペーサー配列を有する複数のNTAA結合性物質を接触させる。サイクル2で使用する結合性物質は、サイクル2特異的配列を含む3'-スペーサー配列も有する。第2のNTAA結合性物質が巨大分子のNTAAに結合し、第2のコーディングタグの情報が記録タグにプライマー伸長を介して移行する。これらのサイクルを最大「n」回の結合サイクルまで繰り返し、単一の巨大分子と共同在する複数の伸長記録タグを生成し、ここで、各伸長記録タグは、1つの結合サイクルからのコーディングタグ情報を有する。連続的な結合サイクルのそれぞれで使用する結合性物質の各セットは、コーディングタグ内のサイクル特異的スペーサー配列を有するので、結合サイクル情報を、得られた伸長記録タグ内の結合性物質情報と関連づけることができる。

#### 【0462】

代替の実施形態では、図9Aと同様に、固体支持体（例えば、ビーズ）上で単一の巨大分子に多数の記録タグが付随しているが、この場合、特定の結合サイクルで使用する結合性物質は、今の結合サイクルに対するサイクル特異的スペーサーおよび次の結合サイクルに対するサイクル特異的スペーサーが隣接するコーディングタグを有する（図10Aおよび10B）。この設計は、伸長記録タグの集団を単一の共直線性の伸長記録タグに変換するための最終的なアセンブリPCRステップを支持するためのものである（図10C）。単一の共直線性の伸長記録タグのライブラリーを配列決定前に濃縮、サブトラクションおよび/または正規化方法に供することができる。第1の結合サイクル（サイクル1）では、第1の結合性物質が結合すると、サイクル1特異的スペーサー（C'1）を含むコーディングタグの情報が、末端に相補的なサイクル1特異的スペーサー（C1）を含む記録タグに移行する。第2の結合サイクル（サイクル2）では、第2の結合性物質が結合すると、サイクル2特異的スペーサー（C'2）を含むコーディングタグの情報が、末端に相補的なサイクル2特異的スペーサー（C2）を含む異なる記録タグに移行する。このプロセスを第nの結合サイクルまで続ける。一部の実施形態では、伸長記録タグ内の第nのコーディ

10

20

30

40

50

ングタグにユニバーサルリバースプライミング配列を用いてキャップ形成する、例えば、ユニバーサルリバースプライミング配列を第  $n$  のコーディングタグ設計の一部として組み入れることもでき、ユニバーサルリバースプライミング配列を第  $n$  の結合サイクルの後に続く反応、例えば尾部を有するプライマーを使用した増幅反応などに追加することもできる。一部の実施形態では、各結合サイクル時に、巨大分子を、それらの対応する結合性物質に関する識別情報および結合サイクル情報を含むコーディングタグと接合した結合性物質の集合に曝露させる（図 9 および図 10）。特定の実施形態では、第  $n$  の結合サイクルが完了した後、伸長記録タグでコーディングしたビーズ基板を油乳剤に平均して液滴当たり 1 ビーズ未満またはそれとほぼ同等になるように入れる。次いで、アセンブリ PCR を使用して、伸長記録タグをビーズから増幅し、多数の別々の記録タグを、別々の伸長記録タグ内のサイクル特異的スパーサー配列を介してプライミングすることによって共直線的な順序でアセンブルさせる（図 10 C）（Xiong ら、2008 年、FEMS Microbiol. Rev.、32 巻：522～540 頁）。あるいは、結合性物質のコーディングタグを有するサイクル特異的スパーサーを使用する代わりに、各結合サイクル中または各結合サイクル後に、サイクル特異的スパーサーを別々に伸長記録タグに付加することができる。単一の巨大分子を集合的に表すものである伸長記録タグの集団を使用することの、単一の巨大分子を表すものである単一の連鎖状の伸長記録タグに対する 1 つの利点は、より高濃度の記録タグにより、コーディングタグ情報の移行の効率が上昇し得ることである。さらに、結合サイクルを数回繰り返して同類結合事象の完了を確実にすることができる。さらに、伸長記録タグの表面増幅により、情報移行の重複性をもたらすことができる（図 4 B を参照されたい）。コーディングタグ情報が必ずしも移行されない場合、ほとんどの場合、それでも、コーディングタグ情報の不完全な集合を、タンパク質などの、情報含有量が非常に高い巨大分子を識別するために使用することが可能であるはずである。短いペプチドであっても、非常に多数の可能性のあるタンパク質配列を具体化し得る。例えば、10mer のペプチドは、20<sup>10</sup> 種の可能性のある配列を有する。したがって、欠失および/または多義性を含有する可能性がある部分的または不完全な配列を、それでも、多くの場合、一意的にマッピングすることができる。

#### 【0463】

タンパク質のネイティブなコンフォメーションが照会される一部の実施形態では、結合性物質の近位のスパーサーエレメント内の切断可能なまたはニッキング可能な DNA 鎖で構成されるコーディングタグを有する結合性物質を用いて周期的結合アッセイを実施する（図 32）。例えば、結合性物質の近位のスパーサーは、ウラシル特異的切除試薬（USER）によってニッキングすることができる 1 つまたは複数のウラシル塩基を有してよい。別の例では、結合性物質の近位のスパーサーは、2 重鎖の一方の鎖のみを加水分解するニッキングエンドヌクレアーゼの認識配列を含んでよい。この設計により、結合性物質を伸長記録タグから変性によらずに除去し、その後のイムノアッセイサイクルのための遊離の一本鎖 DNA スパーサーエレメントを創出することが可能になる。好ましい実施形態では、プライマー伸長ステップ後の結合性物質の酵素的 USER 除去を可能にするために、コーディングタグにウラシル塩基を組み入れる（図 32 E～F）。ウラシルによる USER 切除後、結合性物質および切り詰められたコーディングタグを、タンパク質 - 結合性物質相互作用を破壊するための、高塩濃度（4 M の NaCl、25% ホルムアミド）および穏やかな熱を含めた種々の穏やかな条件下で除去することができる。記録タグとアニリングしたままの他の切り詰められたコーディングタグ DNA の残り（図 32 F）は、わずかに温度を上昇させると容易に解離する。

#### 【0464】

結合性物質の近位のスパーサーエレメント内の切断可能なまたはニッキング可能な DNA 鎖で構成されるコーディングタグにより、多数の結合した結合性物質からコーディングタグ情報を移行させるための単一の均一なアッセイも可能になる（図 33 を参照されたい）。好ましい実施形態では、結合性物質の近位のコーディングタグはニッキングエンドヌクレアーゼ配列モチーフを含み、これは、dsDNA に関して規定された配列モチーフに

においてニッキングエンドヌクレアーゼにより認識され、ニッキングされる。多数の結合性物質の結合後、複合ポリメラーゼ伸長（鎖置換活性を欠く）+ニッキングエンドヌクレアーゼ試薬混合物を使用して、コーディングタグの近位の記録タグまたは伸長記録タグへの反復移行を生じさせる。各移行ステップ後、得られた伸長記録タグ-コーディングタグ2重鎖をニッキングエンドヌクレアーゼによってニッキングし、それにより、結合性物質に付着した切り詰められたスパーサーを放出させ、追加的な近位の結合した結合性物質のコーディングタグとアニーリングすることが可能な伸長記録タグ3'スパーサー配列に曝露させる（図33B~D）。コーディングタグスパーサー配列内へのニッキングモチーフの配置は、切断されていないコーディングタグスパーサー配列と容易に交換することができる、準安定ハイブリッドが創出されるように設計する。このように、2種またはそれよりも多くの結合性物質が同じタンパク質分子に同時に結合する場合、多重に結合した結合性物質からのコーディングタグ情報の記録タグ上への連鎖による結合情報は、単一反応混合物において、いかなる周期的な試薬交換も伴わずに生じる（図33C~D）。この実施形態は、次世代タンパク質アッセイ（NGPA）、特に、タンパク質上の多価エピトープに対するポリクローナル抗体（またはモノクローナル抗体の混合集団）を用いるものに特に有用である。

#### 【0465】

変性したタンパク質、ポリペプチド、およびペプチドの解析を伴う実施形態に関しては、結合した結合性物質およびアニーリングしたコーディングタグを、プライマー伸長後に、高度変性条件（例えば、0.1~0.2NのNaOH、6Mの尿素、2.4Mのグアニジニウムイソチオシアネート、95%ホルムアミドなど）を使用することによって除去することができる。

IX. 記録タグ情報のコーディングタグまたはジタグ構築物への周期的移行

#### 【0466】

別の態様では、結合性物質が巨大分子に結合した後にコーディングタグから記録タグに情報を書き込むのではなく、任意選択のUMI配列（例えば、特定のペプチドまたはタンパク質分子を識別する）および少なくとも1つのバーコード（例えば、コンパートメントタグ、分配バーコード、試料バーコード、空間的位置バーコードなど）を含む記録タグからコーディングタグに情報を移行させ、それにより、伸長コーディングタグを生成することができる（図11Aを参照されたい）。ある特定の実施形態では、結合性物質および付随する伸長コーディングタグを、各結合サイクル後、および任意選択でエドマン分解化学ステップの前に収集する。ある特定の実施形態では、コーディングタグは、結合サイクル特異的タグを含む。周期的なエドマン分解におけるNTAAの検出などの全ての結合サイクルが完了した後、伸長コーディングタグの完全な収集物を増幅し、配列決定し、ペプチド上の情報を、UMI（ペプチド同一性）、エンコーダー配列（NTAA結合性物質）、コンパートメントタグ（単一の細胞またはプロテオームのサブセット）、結合サイクル特異的配列（サイクル数）、またはそれらの任意の組合せの間の関連性から決定することができる。同じコンパートメントタグ/UMI配列を有するライブラリーエレメントを同じ細胞、プロテオームのサブセット、分子などにマッピングし戻し、ペプチド配列を再構築することができる。この実施形態は、記録タグがエドマン分解プロセス中に過度の損傷を保持する場合に有用であり得る。

#### 【0467】

複数の巨大分子を解析するための方法であって、（a）固体支持体に接合した複数の巨大分子および付随する記録タグを用意するステップと；（b）複数の巨大分子を複数の巨大分子に結合することが可能な複数の結合性物質であって、各結合性物質が結合性物質に関する識別情報を有するコーディングタグを含む複数の結合性物質と接触させるステップと；（c）（i）巨大分子に付随する記録タグの情報を巨大分子に結合した結合性物質のコーディングタグに移行させて、伸長コーディングタグを生成するステップ（図11Aを参照されたい）；または（ii）巨大分子に付随する記録タグおよび巨大分子に結合した結合性物質のコーディングタグの情報をジタグ構築物に移行するステップと（図11Bを

参照されたい) ; ( d ) 伸長コーディングタグまたはジタグ構築物を収集するステップと ; ( e ) 任意選択でステップ ( b ) ~ ( d ) を 1 回または複数回の結合サイクルにわたって繰り返すステップと ; ( f ) 伸長コーディングタグまたはジタグ構築物の収集物を解析するステップと

を含む方法が本明細書において提示される。

#### 【 0 4 6 8 】

ある特定の実施形態では、記録タグからコーディングタグへの情報移行は、記録タグのプライマー伸長を防止するために記録タグの 3 ' 末端を任意選択でブロッキングする ( 例えば、図 1 1 A を参照されたい ) プライマー伸長ステップを使用して実現することができる。得られた伸長コーディングタグおよび付随する結合性物質を各結合事象の後および情報移行の完了後に収集することができる。図 1 1 B に例示されている例では、記録タグは、ユニバーサルプライミング部位 ( U 2 ' ) 、バーコード ( 例えば、コンパートメントタグ「 C T 」 ) 、任意選択の U M I 配列、および共通のスペーサー配列 ( S p 1 ) で構成される。ある特定の実施形態では、バーコードは、個々のコンパートメントを表すコンパートメントタグであり、また、U M I を使用して、配列読み取りを照会されている特定のタンパク質またはペプチド分子にマッピングし戻すことができる。図 1 1 B の例において例示されている通り、コーディングタグは、共通のスペーサー配列 ( S p 2 ' ) 、結合性物質エンコーダー配列、およびユニバーサルプライミング部位 ( U 3 ) で構成される。コーディングタグで標識した結合性物質の導入前に、記録タグの U 2 ' ユニバーサルプライミング部位と相補的であり、ユニバーサルプライミング配列 U 1 およびサイクル特異的タグを含むオリゴヌクレオチド ( U 2 ) を記録タグ U 2 ' とアニーリングさせる。さらに、アダプター配列 S p 1 ' - S p 2 を記録タグ S p 1 とアニーリングさせる。このアダプター配列は、コーディングタグ上の S p 2 ' 配列とも相互作用することができ、それにより、記録タグとコーディングタグが互いと近傍になる。結合事象の前または後のいずれかにギャップ充填伸長ライゲーションアッセイを実施する。ギャップ充填を結合サイクルの前に実施する場合、結合サイクル後のプライマー伸長ステップを使用してジタグ形成を完了させる。いくつもの結合サイクルにわたってジタグを収集した後、ジタグの収集物を配列決定し、U M I 配列を介して元のペプチド分子にマッピングし戻す。有効性を最大にするために、U M I 配列の多様性は U M I でタグが付された単一分子の数の多様性を超えるものでなければならないことが理解される。

#### 【 0 4 6 9 】

ある特定の実施形態では、巨大分子は、タンパク質またはペプチドである。ペプチドは生体試料由来のタンパク質を断片化することによって得ることができる。

#### 【 0 4 7 0 】

記録タグは、DNA 分子、RNA 分子、PNA 分子、BNA 分子、XNA 分子、LNA 分子、PNA 分子、またはこれらの組合せであってよい。記録タグは、それが付随する巨大分子 ( 例えば、ペプチド ) を識別する U M I を含む。ある特定の実施形態では、記録タグは、コンパートメントタグをさらに含む。記録タグは、ユニバーサルプライミング部位も含んでよく、これを下流の増幅に使用することができる。ある特定の実施形態では、記録タグは、3 ' 末端にスペーサーを含む。スペーサーは、コーディングタグ内のスペーサーと相補的であってよい。記録タグの 3 ' 末端をブロッキングして ( 例えば、光不安定性 3 ' ブロッキング基 ) ポリメラーゼによる記録タグの伸長を防止し、それにより、巨大分子に付随する記録タグの情報のコーディングタグへの移行または巨大分子に付随する記録タグおよびコーディングタグの情報のジタグ構築物への移行を容易にすることができる。

#### 【 0 4 7 1 】

コーディングタグは、コーディング物質が連結した結合性物質を識別するエンコーダー配列を含む。ある特定の実施形態では、コーディングタグは、コーディングタグが連結した各結合性物質に対する一意の分子識別子 ( U M I ) をさらに含む。コーディングタグは、下流の増幅のために使用することができるユニバーサルプライミング部位を含んでよい。コーディングタグは、3 ' 末端にスペーサーを含んでよい。スペーサーは、記録タグ内の

スペーサーに相補的であってよく、記録タグ情報をコーディングタグに移行させるためのプライマー伸長反応を開始するために使用することができる。コーディングタグは、伸長コーディングタグまたはジタグの起源である結合サイクルを識別するための結合サイクル特異的配列も含んでよい。

#### 【0472】

記録タグの情報のコーディングタグへの移行は、プライマー伸長またはライゲーションによってもたすことができる。記録タグおよびコーディングタグの情報のジタグ構築物への移行は、ギャップ充填反応、プライマー伸長反応、またはその両方で生じさせることができる。

#### 【0473】

ジタグ分子は、伸長記録タグのものと同様の機能的成分を含む。ジタグ分子は、記録タグに由来するユニバーサルプライミング部位、記録タグに由来するバーコード（例えば、コンパートメントタグ）、記録タグに由来する任意選択の一意の分子識別子（UMI）、記録タグに由来する任意選択のスペーサー、コーディングタグに由来するエンコーダー配列、コーディングタグに由来する任意選択の一意の分子識別子、結合サイクル特異的配列、コーディングタグに由来する任意選択のスペーサー、およびコーディングタグに由来するユニバーサルプライミング部位を含んでよい。

#### 【0474】

ある特定の実施形態では、記録タグを、バーコードのコンビナトリアル連鎖をコードするワードを使用して生成することができる。コンビナトリアルをコードするワードの使用により、アニーリングおよび化学的ライゲーションを使用して情報をPNA記録タグからコーディングタグまたはジタグ構築物に移行させることができる方法がもたらされる（例えば、図12A～Dを参照されたい）。本明細書に開示されているペプチドを解析する方法がエドマン分解による末端アミノ酸の切断を伴うある特定の実施形態では、PNAなどの、エドマン分解の厳しい条件に対して抵抗性の記録タグを使用することが望ましい可能性がある。エドマン分解プロトコルにおける1つの厳しいステップは、N末端アミノ酸を切断するための無水TFA処理である。このステップにより、一般には、DNAが破壊される。PNAは、DNAとは対照的に、酸加水分解に対して高度に抵抗性である。PNAでの問題は、情報移行の酵素的 방법이より難しくなる、すなわち、好ましい様式が化学的ライゲーションによる情報移行になることである。図11Bにおいて、記録タグおよびコーディングタグ情報は酵素的ギャップ充填伸長ライゲーションステップによって書き込まれるが、これは、現在、PNA鋳型を用いると、PNAを使用するポリメラーゼが開発されなければ、都合がよくない。化学的ライゲーションが必要であり、その産物は容易には増幅されないため、PNA記録タグからコーディングタグへのバーコードおよびUMIの書き込みには問題がある。化学的ライゲーションの方法は、文献に広範囲にわたって記載されている（Gundersenら、1998年、Genome Res.、8巻：1142～1153頁；Pengら、2010年、Eur. J. Org. Chem.、4194～4197頁；El-Sagheerら、2011年、Org. Biomol. Chem.、9巻：232～235頁；El-Sagheerら、2011年、Proc. Natl. Acad. Sci. USA、108巻：11338～11343頁；Litovchickら、2014年、Artif. DNA PNA XNA、5巻：e27896頁；Roloffら、2014年、Methods Mol. Biol.、1050巻：131～141頁）。

#### 【0475】

コンビナトリアルPNAバーコードおよびUMI配列を創出するために、 $n$ -merのライブラリーからのPNAワードのセットをコンビナトリアルにライゲーションすることができる。各PNAワードが1,000ワードの空間に由来する場合、4つの組合せ配列により、 $1,000^4 = 10^{12}$ コードのコーディング空間が生じる。このように、4,000種の異なるDNA鋳型配列の出発セットから、 $10^{12}$ を超えるPNAコードを生成することができる（図12A）。連鎖状のワードの数を調整すること、または基本のワ

10

20

30

40

50

ードの数を調整することにより、より小さなまたはより大きなコーディング空間を生成することができる。そのように、PNA記録タグとハイブリダイズさせたDNA配列を使用した情報移行は、DNAワードアセンブリハイブリダイゼーションおよび化学的ライゲーションを使用して完了させることができる(図12Bを参照されたい)。PNA鋳型上のDNAワードのアセンブリおよびDNAワードの化学的ライゲーションの後、得られた中間体を使用して、情報をコーディングタグにノから移行させることができる(図12Cおよび図12Dを参照されたい)。

#### 【0476】

ある特定の実施形態では、巨大分子および付随する記録タグは、固体支持体に共有結合により接合している。固体支持体は、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、ナイロン、シリコンウェハチップ、フロースルーチップ、信号変換電子機器を含むバイオチップ、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、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアであってよい。固体支持体は、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズであってよい。

#### 【0477】

ある特定の実施形態では、結合性物質は、タンパク質またはポリペプチドである。一部の実施形態では、結合性物質は、改変またはバリエーションアミノペプチダーゼ、改変またはバリエーションアミノアシルtRNA合成酵素、改変またはバリエーションアンチカリン、改変またはバリエーションC1pS、または改変またはバリエーション抗体またはその結合性断片である。ある特定の実施形態では、結合性物質は、単一のアミノ酸残基、ジペプチド、トリペプチド、またはペプチドの翻訳後修飾に結合する。一部の実施形態では、結合性物質は、N末端アミノ酸残基、C末端アミノ酸残基、または内部アミノ酸残基に結合する。一部の実施形態では、結合性物質は、N末端ペプチド、C末端ペプチド、または内部ペプチドに結合する。一部の実施形態では、結合性物質は、ペプチドの翻訳後修飾のアミノ酸の部位特異的な共有結合性標識である。

#### 【0478】

ある特定の実施形態では、ステップ(b)において複数の巨大分子を複数の結合性物質と接触させた後、巨大分子および付随する結合性物質を含む複合体を固体支持体から解離させ、液滴またはマイクロ流体液滴のエマルジョン中に分配する。一部の実施形態では、各マイクロ流体液滴は、巨大分子および結合性物質を含む複合体を最大で1つ含む。

#### 【0479】

ある特定の実施形態では、伸長コーディングタグまたはジタグ構築物を生成する前に前記記録タグを増幅する。巨大分子および付随する結合性物質を含む複合体を液滴またはマイクロ流体液滴に液滴当たりの複合体が最大で1つになるように分配する実施形態では、記録タグの増幅により、情報をコーディングタグまたはジタグ構築物に移行させるための鋳型として追加的な記録タグがもたらされる(図13および図14を参照されたい)。エマルジョン融合PCRを使用して、記録タグ情報をコーディングタグに移行させるか、またはジタグ構築物の集団を創出することができる。

#### 【0480】

生成された伸長コーディングタグまたはジタグ構築物の収集物を解析前に増幅させることができる。伸長コーディングタグまたはジタグ構築物の収集物の解析は、核酸配列決定法を含んでよい。合成による配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、ポロニーシーケンシング、イオン半導体シーケンシング、またはパイロシーケンシング。核酸配列決定法は、単一分子リアルタイムシーケンシング、ナノポアに基づく配列決定、または先端顕微鏡を使用したDNAのダイレクトイメージングであってよい。

#### 【0481】

エドマン分解およびPITC、サンガー薬剤(DNFB)、SNFB、アセチル化試薬

10

20

30

40

50

、アミジン化（グアニジン化）試薬などのN末端アミンを化学的に標識する方法により、アデニン、グアニン、およびシトシンなどの標準の核酸またはPNA塩基の内部アミノ酸および環外アミンを修飾することもできる。ある特定の実施形態では、ペプチドのリシン残基の $\alpha$ -アミンを、配列決定の前に酸無水物、グアニジン化剤、または同様のブロッキング試薬でブロッキングする。DNA塩基の環外アミンはペプチドのN末端第一級アミンよりもはるかに反応性が低い、DNA塩基の内部アミノ酸および環外アミンに対する非標的活性を低下させる、アミン反応性作用物質のN末端アミンに対する反応性の制御が、配列決定アッセイにとって重要である。修飾反応の選択性は、pH、溶媒（水性対有機、非プロトン性、非極性、極性非プロトン性、イオン性液体など）、塩基および触媒、共溶媒、温度、および時間などの反応条件を調整することによってモジュレートすることができる。さらに、DNA塩基の環外アミンの反応性は、DNAがssDNAの形態であるかdsDNAの形態であるかによってモジュレートされる。修飾を最小限にするために、NTAA化学修飾の前に、記録タグを相補DNAプローブ： $P1'$ 、 $\{試料BCs\}'$ 、 $\{Sp-BC\}'$ などとハイブリダイズさせることができる。別の実施形態では、保護された環外アミンを有する核酸の使用も使用することもできる（Ohkubo、Kasuyara、2008年）。さらに別の実施形態では、SNFBなどの「反応性が低い」アミン標識化合物により、DNAの内部アミノ酸および環外アミンに対するオフターゲットの標識が軽減される（CartyおよびHirs、1968年）。SNFBは、パラスルホニル基がパラニトロ基よりも電子求引性であり、それにより、DNFBよりも活性が低いSNFBでのフッ素置換が導かれるという事実起因してDNFBよりも反応性が低い。

#### 【0482】

NTAA $\alpha$ -アミン修飾を最適化し、オフターゲットのアミノ酸修飾またはDNA修飾を最小限にするためのカップリング条件およびカップリング試薬の調整は、化学および反応条件（濃度、温度、時間、pH、溶媒の型など）を慎重に選択することによって可能である。例えば、DNFBは、二級アミンと、水中よりもアセトニトリルなどの非プロトン性溶媒中の方が容易に反応することが公知である。環外アミンが軽度修飾されてもなお相補的なプローブが配列とハイブリダイズすることが可能になるが、ポリメラーゼに基づくプライマー伸長を妨害する可能性がある。環外アミンを保護しながら、それでも水素結合を可能にすることも可能である。これは、保護された塩基がなお目的の標的とハイブリダイズすることが可能であるという最近の刊行物に記載された（Ohkubo、Kasuyara、2008年）。一実施形態では、工学的に操作されたポリメラーゼを使用して、DNAコーディングタグ鋳型上の記録タグの伸長中に、保護された塩基を有するヌクレオチドを組み入れる。別の実施形態では、工学的に操作されたポリメラーゼを使用して、PNA記録タグ鋳型上のコーディングタグの伸長中に記録タグPNA鋳型（w/またはw/o保護された塩基）上のヌクレオチドを組み入れる。別の実施形態では、情報を、外因性オリゴヌクレオチドをPNA記録タグとアニーリングさせることによって記録タグからコーディングタグに移行させることができる。ハイブリダイゼーションの特異性は、n-merのワードのアセンブリに基づく設計などの配列空間が別個であるUMIを選択することによって容易にすることができる（Gerry、Witowskiら、1999年）。

エドマン様N末端ペプチド分解配列決定を使用してペプチドの直鎖状アミノ酸配列を決定することができるが、代替的な実施形態を使用して、伸長記録タグ、伸長コーディングタグ、およびジタグを利用する方法を用いたペプチドの部分的な組成分析を実施することができる。結合性物質または化学標識を使用してペプチド上のN末端および内部の両方のアミノ酸またはアミノ酸修飾を識別することができる。化学薬剤により、アミノ酸（例えば、標識）を部位特異的に共有結合により修飾することができる（SlettenおよびBertozzi、2009年、Basle、Joubertら、2010年）（SpicerおよびDavis、2014年）。部位特異的に標識されたアミノ酸のコーディングおよびその後の識別を容易にするために、単一のアミノ酸を標的とする化学標識剤にコーディングタグを付着させることができる（図13を参照されたい）。

#### 【0483】

ペプチド組成分析にはペプチドの環状分解は必要なく、したがって、DNAを含有するタグを厳しいエドマン化学に曝露することの問題が回避される。環状結合様式では、組成情報（アミノ酸またはジペプチド/トリペプチド情報）、PTM情報、および一次アミノ酸配列をもたらすために伸長コーディングタグまたはジタグを使用することもできる。一実施形態では、この組成情報は、本明細書に記載の伸長コーディングタグまたはジタグ手法を使用して読み取ることができる。UMIおよびコンパートメントタグ情報と組み合わせると、伸長コーディングタグまたはジタグの収集物により、ペプチドおよびそれらの起源であるコンパートメントのタンパク質（複数可）に関する組成情報がもたらされる。同じコンパートメントタグ（および表面上起源であるタンパク質分子）にマッピングし戻された伸長コーディングタグまたはジタグの収集物は、部分的な組成情報を有するペプチドをマッピングするための強力なツールである。プロテオーム全体にマッピングし戻すのではなく、コンパートメントタグ付きペプチドの集合を限られたタンパク質分子のサブセットにマッピングし戻し、これにより、マッピングの一意性が著しく増大する。

#### 【0484】

本明細書で使用される結合性物質は、単一のアミノ酸、ジペプチド、トリペプチド、またはさらに長いペプチド配列モチーフを認識し得る。Tessler (2011年、Digital Protein Analysis: Technologies for Protein Diagnostics and Proteomics through Single Molecule Detection, Ph.D., Washington University in St. Louis) は、荷電したジペプチドエピトープのサブセットに対して比較的選択的なジペプチド抗体を生成することができることを実証した (Tessler, 2011年)。代替のタンパク質足場（例えば、aaRS、アンチカリン、ClpSなど）およびアプタマーへの定向進化の適用を使用して、ジペプチド/トリペプチド結合性物質のセットを増大させることができる。単一のタンパク質分子にマッピングし戻すことと併せたジペプチド/トリペプチド組成分析からの情報は、各タンパク質分子を一意的に識別し、定量化するために十分なものであり得る。最大で、合計400種の可能性のあるジペプチドの組合せがある。しかし、最も頻度が高く、かつ最も抗原性が高い（荷電、親水性、疎水性）ジペプチドのサブセットが、結合性物質を生成するために十分なものであるはずである。この数は、40～100種の異なる結合性物質のセットを構成し得る。40種の異なる結合性物質のセットに関して、平均10merのペプチドに少なくとも1種の結合性物質が結合する見込みは約80%である。この情報を同じタンパク質分子に由来する全てのペプチドと組み合わせることにより、タンパク質分子の識別が可能になり得る。ペプチドおよびその起源であるタンパク質に関するこの情報全てを組み合わせ、より正確かつ確かなタンパク質配列特徴付けをもたらすことができる。

#### 【0485】

部分的なペプチド配列情報を使用する最近のデジタルタンパク質特徴付けアッセイが提唱された (Swaminathanら、2015年、PLOS Comput. Biol., 11巻: e1004080頁) (Yao, Doctorら、2015年)。すなわち、当該手法では、システイン、リシン、アルギニン、チロシン、アスパラギン酸/グルタミン酸などの、標準の化学を使用して容易に標識されるアミノ酸の蛍光標識を使用する (Basle, Joubertら、2010年)。部分的なペプチド配列情報を用いることの問題は、プロテオームにマッピングし戻すことが一対多数の関連であり、一意のタンパク質が識別されないことである。この一対多数のマッピング問題は、プロテオーム全体空間を、ペプチドがマッピングし戻される限られたタンパク質分子のサブセットに減少させることによって解決することができる。本質的に、単一の部分的なペプチド配列を100種または1000種の異なるタンパク質配列にマッピングし戻すことができるが、いくつかのペプチド（例えば、単一のタンパク質分子の消化に由来する10ペプチド）のセットを全て、コンパートメント内のタンパク質分子のサブセットに含有される単一のタンパク質分子にマッピングし戻した場合、タンパク質分子の同一性を推定することが容易である

10

20

30

40

50



ことが公知である。例えば、同じ分子に由来する全てのペプチドに関するペプチドプロテオームマップの交差により、可能性のあるタンパク質同一性のセットが著しく制限される（図15を参照されたい）。

#### 【0486】

特に、部分的なペプチド配列または組成のマッピング可能性は、コンパートメントのタグおよびUMIの革新的使用を行うことによって有意に増強される。すなわち、プロテオームを最初にバーコードが付されたコンパートメントに分配し、ここで、コンパートメントバーコードはUMI配列にも付着している。コンパートメントバーコードはコンパートメントに一意的な配列であり、UMIはコンパートメント内のバーコードが付された分子それぞれに一意的な配列である（図16を参照されたい）。一実施形態では、この分配は、その全体が参照により組み込まれる、PCT公開第WO2016/061517号に開示されているものと同様の方法を使用して、ビーズに付着したDNAコンパートメントバーコードとのハイブリダイゼーションによる、DNAタグで標識されたポリペプチドとビーズの表面との直接相互作用によって実現される（図31を参照されたい）。プライマー伸長ステップにより、ビーズに連結されたコンパートメントバーコードからポリペプチド上のDNAタグに情報を移行させる（図20）。別の実施形態では、この分配は、UMIを含有する、バーコードが付されたビーズおよびタンパク質分子をエマルジョンの液滴中に共封入することによって実現される。さらに、液滴は、任意選択でタンパク質をペプチドに消化するプロテアーゼを含有する。いくつものプロテアーゼを、レポータータグが付されたポリペプチドを消化するために使用することができる（Switzerland、Gieraら、2013年）。プテラーゼIなどの酵素的リガーゼとプロテアーゼの共封入では、酵素をプロテアーゼ消化に対して抵抗性にするために、ペグ化などの酵素の修飾が必要になり得る（FrokjaerおよびOtzen、2005年、Kang、Wangら、2010年）。消化後、ペプチドをバーコード-UMIタグとライゲーションする。好ましい実施形態では、下流の生化学的操作を容易にするために、バーコード-UMIタグをビーズ上に保持する（図13を参照されたい）。

#### 【0487】

バーコード-UMIとペプチドのライゲーション後、エマルジョンを破壊し、ビーズを回収する。バーコードが付されたペプチドを、それらの一次アミノ酸配列、またはそれらのアミノ酸組成によって特徴付けることができる。ペプチドに関するどちらの型の情報も、プロテオームのサブセットにマッピングし戻すために使用することができる。一般に、配列情報は、組成情報よりもはるかに小さいプロテオームのサブセットにマッピングし戻される。それにもかかわらず、多数のペプチド（配列または組成）からの情報を同じコンパートメントバーコードと組み合わせることにより、ペプチドの起源であるタンパク質（複数可）を一意的に識別することが可能である。このように、プロテオーム全体を特徴付け、定量化することができる。ペプチド上の一次配列情報は、ペプチド配列を表すDNAにコードされるライブラリー（DNA Encoded Library）（DEL）の伸長記録タグ創出を伴うペプチド配列決定反応を実施することによって引き出すことができる。好ましい実施形態では、記録タグは、コンパートメントバーコードおよびUMI配列で構成される。この情報を、コーディングタグから移行された一次またはPTMアミノ酸情報と共に使用して、最終的なマッピングされたペプチド情報を生成する。

#### 【0488】

ペプチド配列情報に対する代替は、コンパートメントバーコードおよびUMIと連結したペプチドアミノ酸またはジペプチド/トリペプチド組成情報を生成することである。これは、UMI-バーコードが付されたペプチドを伴うビーズを、各ペプチド上の選択されたアミノ酸（内部）をアミノ酸コード情報および別のアミノ酸UMI（AA UMI）を含むDNAタグで部位特異的に標識するアミノ酸標識ステップに供することによって実現される（図13を参照されたい）。最も化学標識しやすいアミノ酸（AA）は、リシン、アルギニン、システイン、チロシン、トリプトファン、およびアスパラギン酸/グルタミン酸であるが、他のAAに対する標識スキームを同様に展開することも実行可能であり得

10

20

30

40

50

る (Mendoza および Vachet、2009 年)。所与のペプチドは、同じ型の A A をいくつか含有し得る。同じ型の多数のアミノ酸の存在は、付着した A A U M I 標識によって区別することができる。各標識分子は、DNA タグ内に異なる U M I を有し、それによりアミノ酸の計数が可能になる。化学標識に対する代替は、A A を結合性物質で「標識」することである。例えば、A A コード情報および A A U M I を含むコーディングタグで標識されたチロシン特異的抗体を使用して、ペプチドの全てのチロシンに印を付けることができる。この手法に伴う注意事項は、大きなかさのある抗体では立体的な障害が生じることであり、この目的のためには、より小さな s c F v、アンチカリン、または C l p S バリアントを使用することが理想的である。

#### 【0489】

一実施形態では、A A へのタグ付け後、情報を、記録タグと、ペプチド上の結合したまたは共有結合によりカップリングした結合性物質に付随する多数のコーディングタグとの間で、液滴当たり単一のペプチドが含有されるようにペプチド複合体を区画化し、エマルジョン融合 P C R を実施して、区画化されたペプチドのアミノ酸組成を特徴付ける伸長コーディングタグまたはジタグのセットを構築することによって移行させる。ジタグの配列決定後、同じバーコードを有するペプチド上の情報を単一のタンパク質分子にマッピングし戻すことができる。

#### 【0490】

特定の実施形態では、タグが付されたペプチド複合体をビーズから解離させ (図 13 を参照されたい)、小さなミニコンパートメント (例えば、マイクロエマルジョン) に、平均して単一の標識された / 結合した結合性物質ペプチド複合体のみが所与のコンパートメント内に存在するように分配する。特定の実施形態では、この区画化は、マイクロエマルジョン液滴の生成によって実現される (Shim、Ranasinghe ら、2013 年、Shembekar、Chaipan ら、2016 年)。ペプチド複合体に加えて、P C R 試薬も、液滴中に 3 種のプライマー (U 1、S p、および U 2 t r) と共封入することができる。液滴形成後、U 1 および S p のみがアニーリングし、記録タグ産物が増幅するように、高いアニーリング温度でエマルジョン P C R を数サイクル実施する (約 5 ~ 10 サイクル) (図 13 を参照されたい)。この最初の 5 ~ 10 サイクルの P C R の後、アミノ酸コードタグ上の U 2 t r および S p t r が増幅に加わるようにアニーリング温度を低下させ、さらに約 10 ラウンドを実施する。3 プライマーエマルジョン P C R により、ペプチド U M I - バーコードと A A コードタグの全てが有効に組み合わせられ、それにより、ペプチドおよびそのアミノ酸組成のジタグライブラリー表示が生成する。3 プライマー P C R およびタグの連鎖を実施する他のモダリティを使用することもできる。別の実施形態は、光デブロッキング、または不安定なブロッキングされた 3' ヌクレオチドの 3' デブロッキングを開始するための油可溶性還元体の添加によって活性化される、3' ブロッキングされた U 2 プライマーの使用である。エマルジョン P C R 後、N G S 配列決定のためのライブラリーエレメントをフォーマットするための一般的なプライマーを用いて別のラウンドの P C R を実施することができる。

#### 【0491】

このように、ライブラリーエレメントの異なる配列成分を計数および分類のために使用する。所与のペプチド (コンパートメントバーコード - U M I の組合せによって識別される) について、多くのライブラリーエレメントが存在し、それぞれが識別用 A A コードタグおよび A A U M I を有する (図 13 を参照されたい)。A A コードおよび付随する U M I を使用して、所与のペプチド内の所与のアミノ酸型の存在を計数する。したがって、ペプチド (おそらく G l u C、L y s C、または E n d o A s n N 消化物) を、そのアミノ酸組成 (例えば、C y s が 2 つ、L y s が 1 つ、A r g が 1 つ、T y r が 2 つなど) によって、空間的順序は考慮せず、特徴付ける。それにもかかわらず、これにより、ペプチドをプロテオームのサブセットにマッピングするため、また、同じタンパク質分子に由来する他のペプチドと組み合わせ使用した場合には、タンパク質を一意的に識別および定量化するための、十分なシグネチャがもたらされる。

10

20

30

40

50

# X. 末端アミノ酸 (NTAA) 標識方法

## 【0492】

ある特定の実施形態では、本明細書に記載の方法においてペプチドを結合性物質と接触させる前に、ペプチドの末端アミノ酸 (例えば、NTAAまたはCTAA) を修飾または標識する。

## 【0493】

一部の実施形態では、NTAAをフェニルイソチオシアネート (PITC) と反応させて、フェニルチオカルバモイル (PTC) - NTAA誘導体を生成する。エドマン分解では、一般には、フェニルイソチオシアネート (PITC) を使用してN末端を標識する。PITCは、本明細書に開示されている方法に十分に適する2つの性質を有する：(1) PITCはN末端アミン基を高い効率で標識する；および(2) 得られるPTC誘導体化NTAAは、酸処理されると自己異性化を受け、それにより当該アミノ酸が残りのペプチドから切断される。

## 【0494】

NTAAを標識するために使用することができる他の試薬としては、4 - スルホフェニルイソチオシアネート、3 - ピリジルイソチオシアネート (PYITC)、2 - ピペリジノエチルイソチオシアネート (PEITC)、3 - (4 - モルホリノ) プロピルイソチオシアネート (MPIITC)、3 - (ジエチルアミノ) プロピルイソチオシアネート (DEPTIC) (Wangら、2009年、Anal Chem、81巻：1893 ~ 1900頁)、(1 - フルオロ - 2, 4 - ジニトロベンゼン (サンガー試薬、DNFB)、ダンシルクロリド (DNS - Cl、または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)、アセチル化試薬、アミジン化 (グアニジン化) 試薬、2 - カルボキシ - 4, 6 - ジニトロクロロベンゼン、7 - メトキシクマリン酢酸、チオアシル化試薬、チオアセチル化試薬、およびチオベンジル化試薬が挙げられる。NTAAを標識のためにブロッキングする場合、例えば、N - アセチルブロッキングをアシルペプチドヒドロラーゼ (APH) で除去することなど、末端をアンブロッキングするための手法がいくつも存在する (Farries、Harrisら、1991年、Eur. J. Biochem.、196巻：679 ~ 685頁)。ペプチドのN末端をアンブロッキングする方法は、当技術分野で公知である (例えば、それぞれ、その全体が参照により本明細書に組み込まれる、Krishnaら、1991年、Anal. Biochem.、199巻：45 ~ 50頁；Leoneら、2011年、Curr. Protoc. Protein Sci.、第11章：ユニット11.7；Fowlerら、2001年、Curr. Protoc. Protein Sci.、第11章：ユニット11.7を参照されたい)。

## 【0495】

ダンシルクロリドは、ペプチドの遊離のアミン基と反応して、NTAAのダンシル誘導体をもたらす。DNFBおよびSNFBはペプチドの - アミン基と反応して、それぞれDNP - NTAA、およびSNP - NTAAを生成させる。さらに、DNFBおよびSNFBはどちらも、リシン残基の - アミンとも反応する。DNFBはまた、チロシンおよびヒスチジンアミノ酸残基とも反応する。SNFBは、DNFBよりも良好なアミン基に対する選択性を有し、NTAA修飾に好ましい (CartyおよびHirs、1968年)。ある特定の実施形態では、ポリペプチドをプロテアーゼ消化してペプチドにする前に、リシン - アミンを有機無水物でプレブロッキングする。

## 【0496】

別の有用なNTAA修飾因子はアセチル基であり、その理由は、アセチル化NTAAを除去する公知の酵素、すなわち、N末端アセチル化アミノ酸を切断し、それによりペプチドを単一アミノ酸だけ有効に短縮するアシルペプチドヒドロラーゼ (APH) が存在するからである {Chang、2015年、#373；Friedmann、2013年、#374}。NTAAは、無水酢酸を用いて化学的にアセチル化することもでき、N末端アセチルトランスフェラーゼ (NAT) を用いて酵素的にアセチル化することもできる {C

hang、2015年、#373; Friedmann、2013年、#374}。さらに別の有用なNTAA修飾因子はアミジニル(グアニジニル)部分であり、その理由は、アミジン化NTAAの証明された切断化学、すなわち、N末端アミジン化ペプチドを0.5~2% NaOHと一緒に穏やかにインキュベートすることにより、N末端アミノ酸の切断がもたらされることが文献で知られているからである{Hamada、2016年、#383}。これにより、穏やかなエドマン様化学的N末端分解ペプチド配列決定プロセスが有効にもたらされる。さらに、ある特定のアミジン化(グアニジン化)試薬および下流のNaOH切断は、DNAエンコーディングに非常に適合性である。

#### 【0497】

NTAAにDNP/SNP基、アセチル基、またはアミジニル(グアニジニル)基が存在することにより、工学的に操作された結合性物質との相互作用のより良好な取り扱いがもたらされ得る。低nMの親和性を有する商業的なDNP抗体がいくつも存在する。他のNTAA標識方法としては、トリプリガーゼ(trypsinase)を用いた標識(Liebscherら、2014年、Angew Chem Int Ed Engl、53巻:3024~3028頁)、およびアミノアシルトランスフェラーゼを用いた標識(Wagnerら、2011年、J Am Chem Soc、133巻:15139~15147頁)が挙げられる。

#### 【0498】

イソチオシアネートは、イオン性液体の存在下では、第一級アミンに対して増強された反応性を有することが示されている。イオン性液体は、有機化学反応における優れた溶媒であり(また、触媒として作用する)、チオ尿素を形成する、イソチオシアネートとアミンの反応を増強することができる。例は、フェニルイソチオシアネート(PITC)による芳香族および脂肪族アミンの迅速かつ効率的な標識のためにイオン性液体1-ブチル-3-メチル-イミダゾリウムテトラフルオロボレート[Bmim][BF4]を使用することである(Le、Chenら、2005年)。エドマン分解は、PITCなどのイソチオシアネートとペプチドのアミノN末端の反応を伴う。そのように、一実施形態では、より穏やかな標識および分解条件をもたらしことによってエドマン分解プロセスの効率を改善するためにイオン性液体を使用する。例えば、イオン性液体[Bmim][BF4]中5%(vol./vol.)PITCを25で10分間使用することは、ピリジン、エタノール、およびddH2Oを含有する溶液(1:1:1 vol./vol./vol.)中5%(vol./vol.)PITCを55で60分間使用する標準のエドマンPITC誘導体化条件下で標識することよりも効率的である(Wang、Fangら、2009年)。好ましい実施形態では、ポリペプチドの内部のリシン、チロシン、ヒスチジン、およびシステインアミノ酸を、ペプチドへの断片化の前にブロッキングする。このように、ペプチド配列決定反応中、ペプチドNTAAの - アミン基のみが修飾に利用可能になるようにする。これは、特に、DNFB(サンガー試薬)およびダンシルクロリドを使用する場合に適切である。

#### 【0499】

ある特定の実施形態では、NTAAは、NTAA標識ステップの前にブロッキングされている(特にタンパク質の元のN末端)。その場合、例えば、N-アセチルブロッキングをアシルペプチドヒドロラーゼ(APH)で除去することなど、N末端をアンブロッキングするための手法がいくつも存在する(Farrises、Harrisら、1991年)。ペプチドのN末端をアンブロッキングする他の方法がいくつも当技術分野で公知である(例えば、それぞれ、その全体が参照により本明細書に組み込まれる、Krishnaら、1991年、Anal. Biochem.、199巻:45~50頁; Leoneら、2011年、Curr. Protoc. Protein Sci.、第11章:ユニット11.7; Fowlerら、2001年、Curr. Protoc. Protein Sci.、第11章:ユニット11.7を参照されたい)。

#### 【0500】

CTAAは、Hermanson(Hermanson、2013年)に記載されてい

10

20

30

40

50

る通り、いくつもの異なるカルボキシル反応性試薬を用いて修飾することができる。別の例では、CTAAを混合無水物およびイソチオシアネートで修飾して、チオヒダントインを生成する（LiuおよびLiang、2001年）および米国特許第5,049,507号）。チオヒダントインで修飾されたペプチドは、基剤中上昇した温度で切断されて、最後から2番目のCTAAが露出し、それにより、C末端に基づくペプチド分解配列決定手法を有効に生成することができる（LiuおよびLiang、2001年）。CTAAに対して行うことができる他の修飾としては、パラ-ニトロアニリド基の付加および7-アミノ-4-メチルクマリニル基の付加が挙げられる。

#### XI. 末端アミノ酸切断方法

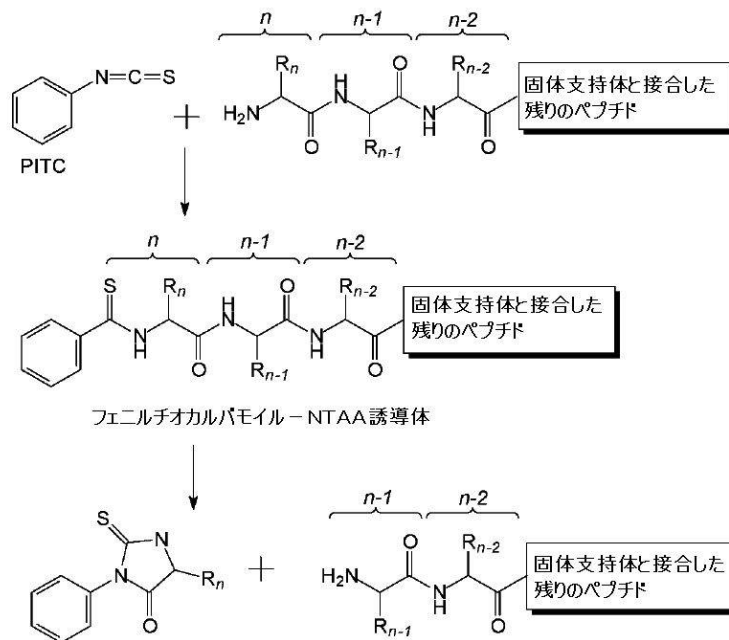
##### 【0501】

ペプチドの解析に関するある特定の実施形態では、末端アミノ酸（N末端またはC末端）への結合性物質の結合、およびコーディングタグ情報の記録タグへの移行、記録タグ情報のコーディングタグへの移行、記録タグ情報およびコーディングタグ情報のジタグ構築物への移行の後、末端アミノ酸をペプチドから除去または切断して、新しい末端アミノ酸を露出させる。一部の実施形態では、末端アミノ酸は、NTAAである。他の実施形態では、末端アミノ酸は、CTAAである。

##### 【0502】

末端アミノ酸の切断は、化学的切断および酵素的切断を含めた、任意の数の公知の技法によって実現することができる。化学的切断の例は、エドマン分解である。ペプチドのエドマン分解中、 $n$  NTAAをフェニルイソチオシアネート（PITC）と弱アルカリ性条件下で反応させて、フェニルチオカルバモイル-NTAA誘導体を形成する。次に、酸性条件下で、フェニルチオカルバモイル-NTAA誘導体を切断し、それにより、遊離のチアゾリノン誘導体を生成し、それにより、ペプチドの $n-1$ アミノ酸をN末端アミノ酸（ $n-1$  NTAA）に変換する。このプロセスのステップを以下に例示する：

##### 【化1】



##### 【0503】

典型的なエドマン分解では、上記の通り、長いインキュベーション時間にわたって厳しい高温の化学的条件（例えば、無水TFA）の発生が必要になる。これらの条件は、一般に、巨大分子の核酸エンコーディングには適合しない。

##### 【0504】

化学的エドマン分解を核酸エンコーディングに好都合の手法に変換するために、厳しい

化学的ステップを穏やかな化学的分解または効率的な酵素的ステップで置き換える。一実施形態では、化学的エドマン分解を、元々記載された条件よりも穏やかな条件を使用して利用することができる。アセトニトリル中無水TFAを酢酸トリエチルアミンで置き換えること（例えば、その全体が参照により組み込まれる、Barrett、1985年、Tetrahedron Lett.、26巻：4375～4378頁を参照されたい）を含め、エドマン分解のための穏やかな切断条件がいくつか文献に記載されている。NTAAの切断は、エドマン分解と比較して穏やかな切断条件を使用するチオアシル化分解を使用することによって実現することもできる（米国特許第4,863,870号を参照されたい）。

#### 【0505】

別の実施形態では、無水TFAによる切断を、穏やかな条件下で切れやすいペプチド結合のカルボニル基のチオ尿素硫黄原子の求核攻撃によるPITC誘導体化N末端アミノ酸の除去を触媒する、工学的に操作された酵素である「エドマナーゼ(Edmanase)」で置き換えることができる（その全体が参照により組み込まれる、米国特許公開第US2014/0273004号を参照されたい）。エドマナーゼは、Trypanosoma cruziに由来するシステインプロテアーゼであるクルザイン(cruzain)を改変することによって作出された(Borgo、2014年)。C25G変異により触媒性システイン残基が除去されたと同時に、エドマン試薬(PITC)のフェニル部分との立体的な適合を創出するための3つの変異(G65S、A138C、L160Y)が選択された。

#### 【0506】

NTAAの酵素的切断は、アミノペプチダーゼによって実現することもできる。アミノペプチダーゼは、単量体酵素および多量体酵素として天然に存在し、金属またはATP依存性であり得る。天然のアミノペプチダーゼは、非常に限られた特異性を有し、一般的に、N末端アミノ酸を前進的に切断し、それにより、アミノ酸を次々に切断する。本明細書に記載の方法に関しては、アミノペプチダーゼを、NTAAに対して、N末端標識で修飾されている場合にのみ特異的な結合活性または触媒活性を有するように工学的に操作することができる。例えば、アミノペプチダーゼを、N末端アミノ酸がDNP/SNP、PTC、ダンシルクロリド、アセチル、アミジニルなどの基によって修飾されている場合にN末端アミノ酸のみを切断するように工学的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、一度にN末端から単一のアミノ酸だけ切断し、分解サイクルの制御を可能にする。一部の実施形態では、改変アミノペプチダーゼは、アミノ酸残基同一性に関しては非選択的である一方で、N末端標識に関しては選択的である。他の実施形態では、改変アミノペプチダーゼは、アミノ酸残基同一性とN末端標識の両方に関して選択的である。酵素的NTAA分解の修飾特異性のモデルの例は、BorgoおよびHavranekにより例示されており、そこでは、構造-機能補助設計を通じて、メチオニンアミノペプチダーゼがロイシンアミノペプチダーゼに変換された(BorgoおよびHavranek、2014年)。DNP/SNPで修飾されたNTAAなどの修飾されたNTAAに同様の手法をとることができ、ここで、アミノペプチダーゼを、DNP/SNP基が存在するN末端アミノ酸のみを切断するように工学的に操作する（構造-機能に基づく設計および定向進化の両方を使用する）。個々のまたは小さな群の標識した（ビオチン化した）NTAAに結合し、それを切断する工学的に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（PCT公開第WO2010/065322号を参照されたい）。

#### 【0507】

ある特定の実施形態では、緻密な単量体金属酵素的アミノペプチダーゼを、DNPで標識されたNTAAを認識し、切断するように工学的に操作する。単量体メタロ-アミノペプチダーゼの使用には、重要な利点が2つある：1) 緻密な単量体タンパク質は、ファージディスプレイを使用してディスプレイおよびスクリーニングするのがはるかに容易である；2) メタロ-アミノペプチダーゼは、適切な金属カチオンを添加または除去することによってその活性をオン/オフにすることができるという点で、独特の利点を有する。例

10

20

30

40

50

示的なアミノペプチダーゼとしては、*Streptomyces* sp. KK506 (SKAP) (Yoo, Ahnら、2010年)、*Streptomyces griseus* (SGAP)、*Vibrio proteolyticus* (VPAP)、(SpunginおよびBlumberg、1989年、Ben-Meir、Spunginら、1993年)などの、アミノペプチダーゼのM28ファミリーが挙げられる。これらの酵素は、室温およびpH 8.0で安定であり、ロバストであり、かつ活性であり、したがって、ペプチド解析に好ましい穏やかな条件に適合する。

#### 【0508】

別の実施形態では、アミノペプチダーゼを、N末端アミノ酸標識の存在下でのみ活性になるように工学的に操作することにより、周期的切断が達成される。さらに、アミノペプチダーゼを、非特異的になるように、したがって、1つの特定のアミノ酸を別のアミノ酸に対して選択的に認識するのではなく、単に標識されたN末端を認識するように、工学的に操作することができる。好ましい実施形態では、メタロペプチダーゼ単量体アミノペプチダーゼ(例えば、*Vibro*ロイシンアミノペプチダーゼ)(Hernandez-Moreno、Villaseñorら、2014年)を、修飾されたNTAA(例えば、PTC、DNP、SNP、アセチル化された、アシル化されたなど)のみを切断するように工学的に操作する。

#### 【0509】

さらに別の実施形態では、アセチル化NTAAを切断するように工学的に操作されたアシルペプチドヒドロラーゼ(APH)を使用することによって周期的切断を達成する。APHは、ブロックされたペプチドからのN-アセチル化アミノ酸の除去を触媒することができるセリンペプチダーゼであり、真核細胞、細菌細胞および古細菌細胞において、N末端がアセチル化されたタンパク質の重要な調節因子である。ある特定の実施形態では、APHは、二量体であり、エキソペプチダーゼ活性のみを有する(Gogliettino、Balettrieriら、2012年、Gogliettino、Ricciolaら、2014年)。工学的に操作されたAPHは、内因性または野生型APHよりも高い親和性および低い選択性を有し得る。

#### 【0510】

さらに別の実施形態では、NTAAのアミジン化(グアニジニル化)を使用して、標識されたNTAAのNaOHを使用した穏やかな切断を可能にする(Hamada、2016年、その全体が参照により組み込まれる)。S-メチルイソチオ尿素、3,5-ジメチルピラゾール-1-カルボキサミジン、S-エチルチオウロニウムブロミド、S-エチルチオウロニウムクロリド、O-メチルイソ尿素、O-メチルイソウロニウム硫酸塩、O-メチルイソ尿素硫化水素塩、2-メチル-1-ニトロイソ尿素、アミノイミノメタンスルホン酸、シアナミド、シアノグアニド、ジシアンジアミド、3,5-ジメチル-1-グアニルピラゾール硝酸塩および3,5-ジメチルピラゾール、N,N'-ビス(オルト-クロロ-Cbz)-S-メチルイソチオ尿素およびN,N'-ビス(オルト-プロモ-Cbz)-S-メチルイソチオ尿素を含めたいくつものアミジン化(グアニジニル化)試薬が当技術分野で公知である(Katritzky、2005年、その全体が参照により組み込まれる)。

#### 【0511】

NTAAの標識、結合、および分解ワークフローの例は以下の通りである(図41および42を参照されたい): タンパク質分解による消化に由来する、記録タグで標識されたペプチドの大きな集団(例えば、5千万~10億)を単一分子シーケンシング基板(例えば、多孔質ビーズ)に適切な分子内間隔でランダムに固定化する。周期的に、各ペプチドのN末端アミノ酸(NTAA)を小さな化学的部分(例えば、DNP、SNP、アセチル)で修飾して、NTAA分解プロセスの周期的制御をもたらし、同類結合性物質による結合親和性を増強する。各固定化ペプチドの修飾されたN末端アミノ酸(例えば、DNP-NTAA、SNP-NTAA、アセチル-NTAA)に同類のNTAA結合性物質を結合させ、結合したNTAA結合性物質に付随するコーディングタグからの情報を固定化され

10

20

30

40

50

たペプチドに付随する記録タグに移行させる。NTAAの認識、結合、およびコーディングタグ情報の記録タグへの移行の後、標識されたNTAAを、標識の存在下でのみNTAAを切断することができる工学的に操作されたアミノペプチダーゼ（例えば、DNP-NTAAもしくはSNP-NTAAに対して）または工学的に操作されたAPH（例えば、アセチル-NTAAに対して）に曝露させることによって除去する。他のNTAA標識（例えば、PITC）も、適切に工学的に操作されたアミノペプチダーゼとともに使用することもできる。特定の実施形態では、単一の工学的に操作されたアミノペプチダーゼまたはAPHは、N末端アミノ酸標識を有する全ての可能性のあるNTAA（翻訳後修飾バリエーションを含む）を普遍的に切断する。別の特定の実施形態では、2種、3種、4種、またはそれよりも多くの工学的に操作されたアミノペプチダーゼまたはAPHを使用して、標識されたNTAAのレパートリーを切断する。

10

#### 【0512】

DNPまたはSNPで標識されたNTAAに対する活性を有するアミノペプチダーゼは、ベンジルペニシリンに対するメタロ-ベータ-ラクタマーゼ酵素の工学的操作におけるPonsardらにより記載されている手法（Ponsard、Galleniら、2001年、Fernandez-Gacio、Uguenら、2003年）のような、アポ酵素（金属補助因子の不在下では不活性）に対する密接結合の選択、その後の機能的触媒の選択ステップを組み合わせるスクリーニングを使用して選択することができる。この二段階選択は、 $Zn^{2+}$ イオンの添加によって活性化されたメタロ-APの使用を伴う。固定化されたペプチド基板への密接結合選択の後、 $Zn^{2+}$ を導入し、DNPまたはSNPで標識されたNTAAを加水分解することが可能な、触媒として活性なファージにより、結合したファージの上清中への放出が導かれる。DNPまたはSNPで標識されたNTAAを切断するための活性なAPを濃縮するために、選択ラウンドを繰り返し実施する。

20

#### 【0513】

本明細書において提示される実施形態のいずれかでは、NTAAの切断試薬のNTAAへの動員は、キメラ切断酵素およびキメラNTAA修飾因子によって増強することができる。ここで、キメラ切断酵素およびキメラNTAA修飾因子は、それぞれ、互いと密接結合反応することが可能な部分を含む（例えば、ピオチン-ストレプトアビジン）（図39を参照されたい）。例えば、NTAAをピオチン-PITCで修飾し、キメラ切断酵素（ストレプトアビジン-エドマナーゼ）を修飾されたNTAAにストレプトアビジン-ピオチン相互作用によって動員し、それにより切断酵素の親和性および効率を改善することができる。修飾されたNTAAが切断され、付随する切断酵素と共にペプチドから発散する。キメラエドマナーゼの例では、この手法により、親和性 $K_D$ が $\mu M$ からサブピコモルまで有効に増大する。記録タグと相互作用する切断剤上のDNAタグを使用した係留により、同様の切断の増強も実現することができる（図44を参照されたい）。

30

#### 【0514】

NTAAの切断のための代替として、ジペプチジルアミノペプチダーゼ（DAP）を使用して、最後の2つのN末端アミノ酸をペプチドから切断することができる。ある特定の実施形態では、単一のNTAAを切断することができる（図45を参照されたい）：図45は、プテラーゼIペプチド基板のN末端ライゲーションによりTEVエンドペプチダーゼ基板をペプチドのN末端に付着させる、N末端分解のための手法を示す。付着後、TEVエンドペプチダーゼにより、新しくライゲーションされたペプチドが照会ペプチド（配列決定を受けているペプチド）から切断され、NTAAに付着した単一のアスパラギン（N）が残る。N末端から2つのアミノ酸を切断するDAPと一緒にインキュベートすることにより、元のNTAAの最終的な除去がもたらされる。このプロセス全体をN末端分解プロセスとして周期的に繰り返すことができる。

40

#### 【0515】

CTAA結合性物質に関する実施形態に関しては、CTAAをペプチドから切断する方法も当技術分野で公知である。例えば、米国特許第6,046,053号には、ペプチドまた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

50



換させ、酸およびアルコールまたはエステルを用いた反応によってC末端アミノ酸を遊離させる方法が開示されている。CTAAの酵素的切断はまた、カルボキシペプチダーゼによっても実現することができる。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、アルギニンおよびリシンなどの塩基性アミノ酸において優先的に切断する。上記の通り、カルボキシペプチダーゼをアミノペプチダーゼと同じように改変して、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工学的に作製することもできる。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。一部の実施形態では、改変カルボキシペプチダーゼは、アミノ酸残基同一性に関しては非選択的であるが、C末端標識については選択的である。他の実施形態では、改変カルボキシペプチダーゼは、アミノ酸残基同一性およびC末端標識の両方に関して選択的である。

10

XII. 伸長記録タグ、伸長コーディングタグ、またはジタグの処理および解析

#### 【0516】

目的の巨大分子（複数可）を表す伸長記録タグ、伸長コーディングタグ、およびジタグライブラリーを、種々の核酸配列決定法を使用して処理および解析することができる。配列決定法の例としては、これだけに限定されないが、連鎖停止配列決定（サンガー配列決定）；合成による配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、ポロニーシーケンシング、イオン半導体シーケンシング、およびパイロシーケンシングなどの次世代シーケンシング法；および単一分子リアルタイムシーケンシング、ナノポアに基づく配列決定、DI (duplex interrupted) シーケンシング、および先端顕微鏡を使用したDNAのダイレクトイメージングなどの第3世代シーケンシング法が挙げられる。

20

#### 【0517】

伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーを様々なやり方で増幅することができる。伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーに、例えば、PCRまたはエマルジョンPCRによって指数関数的な増幅を行うことができる。エマルジョンPCRにより、より均一な増幅がもたらされることが公知である（Hori、Fukanoら、2007年）。あるいは、伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーに、例えば、T7 RNAポリメラーゼを使用した鋳型DNAのin vitro転写によって線形増幅を行うことができる。伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーは、それに含有されるユニバーサルフォワードプライミング部位およびユニバーサルリバースプライミング部位に適合するプライマーを使用して増幅することができる。伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーは、伸長記録タグ、伸長コーディングタグ、またはジタグの5'末端、3'末端または両末端のいずれかに配列を付加するための尾部を有するプライマーを使用して増幅することもできる。伸長記録タグ、伸長コーディングタグ、またはジタグの末端に付加することができる配列としては、多数のライブラリーを単一の配列決定の実行に多重化することを可能にするライブラリー特異的指数配列、アダプター配列、読み取りプライマー配列、または、配列決定プラットフォームに適合する、伸長記録タグ、伸長コーディングタグ、もしくはジタグのライブラリーを作出するための任意の他の配列が挙げられる。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通りである：ビーズ（約10 ng）約1 mgから溶出した伸長記録タグライブラリー、200 μMのdNTP、1 μMの各フォワードおよ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）を使用してPCR反応体積20 μlをセットアップし、以下のサイクル条件に供す：98 °Cで30秒、その後、98 °Cで10秒、60 °Cで30秒、72 °Cで30秒を20サイクル、その後、72 °Cで7分、次いで、4 °Cで保持。

30

40

#### 【0518】

ある特定の実施形態では、増幅前、増幅中または増幅後のいずれかにおいて、伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーに標的濃縮を行うことができ

50

る。標的濃縮は、配列決定前に、伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーから、目的の巨大分子を表す伸長記録タグを選択的に捕捉または増幅するために使用することができる。タンパク質配列に対する標的濃縮は、費用が高く、また、標的タンパク質に対して高度に特異的な結合性物質を作製することが難しいので、困難である。抗体は、非特異的であり、何千ものタンパク質にわたる規模生産が難しいことが周知である。本開示の方法では、タンパク質コードを核酸コードに変換し、次いで、DNAライブラリーに利用可能な広範囲の標的化DNA濃縮戦略を使用することによってこの問題を回避する。目的のペプチドを、それらの対応する伸長記録タグを濃縮することによって試料中で濃縮することができる。標的化濃縮の方法が当技術分野で公知であり、それらとして、ハイブリッド捕捉アッセイ、PCRに基づくアッセイ、例えば、TruSeq custom Amplicon (Illumina)、padlockプローブ（分子反転プローブとも称される）などが挙げられる（それぞれ、その全体が参照により本明細書に組み込まれる、Mamanovaら、2010年、Nature Methods、7巻：111～118頁；Bodiら、J. Biomol. Tech.、2013年、24巻：73～86頁；Ballesterら、2016年、Expert Review of Molecular Diagnostics、357～372頁；Mertesら、2011年、Brief Funct. Genomics、10巻：374～386頁；Nilssonら、1994年、Science、265巻：2085～8頁を参照されたい）。

10

#### 【0519】

20

一実施形態では、伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーをハイブリッド捕捉に基づくアッセイによって濃縮する（例えば、図17Aおよび図17Bを参照されたい）。ハイブリッド-捕捉に基づくアッセイでは、伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーを、アフィニティータグ（例えば、ビオチン）で標識した標的特異的オリゴヌクレオチドまたは「ベイトオリゴヌクレオチド」とハイブリダイズさせる。標的特異的オリゴヌクレオチドとハイブリダイズさせた伸長記録タグ、伸長コーディングタグ、またはジタグを、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したビーズ）を使用してそれらの親和性タグを介して「プルダウン」し、バックグラウンド（非特異的な）伸長記録タグを洗い流す（例えば、図17を参照されたい）。次いで、濃縮された伸長記録タグ、伸長コーディングタグ、またはジタグを正の濃縮のために得る（例えば、ビーズから溶出する）。

30

#### 【0520】

アレイに基づく「in situ」オリゴヌクレオチド合成およびその後のオリゴヌクレオチドプールの増幅によって合成されたベイトオリゴヌクレオチドに関しては、所与のオリゴヌクレオチドアレイ内でユニバーサルプライマーのいくつかのセットを使用することにより、競合するベイトを工学的に操作してプールにすることができる。それぞれの型のユニバーサルプライマーについて、ビオチン化プライマーと非ビオチン化プライマーの比により、濃縮比を制御する。いくつかのプライマー型の使用により、いくつかの濃縮比を設計して最終的なオリゴヌクレオチドベイトプールにすることができる。

#### 【0521】

40

ベイトオリゴヌクレオチドを目的の巨大分子を表す伸長記録タグ、伸長コーディングタグ、またはジタグと相補的になるように設計することができる。伸長記録タグ、伸長コーディングタグ、またはジタグ内のスペーサー配列に対するベイトオリゴヌクレオチドの相補性の程度は、0%から100%まで、およびその間の任意の整数にすることができる。このパラメーターは、少しの濃縮実験によって容易に最適化することができる。一部の実施形態では、コーディングタグ設計においてエンコーダー配列に対するスペーサーの長さを最小化する、または、スペーサーを、ベイト配列とのハイブリダイゼーションに利用できないように設計する。1つの手法は、補助因子の存在下で二次構造を形成するスペーサーを使用することである。そのような二次構造の例は、互いの上に積み重なった2つまたはそれよりも多くのグアニンカルテットによって形成される構造である、G-4重鎖であ

50

る (Bochman、Paeschkeら、2012年)。グアニンカルテットは、フーグスティーン水素結合によって結びついた4つのグアニン塩基によって形成される平面正方形構造である。G-4重鎖構造は、カチオン、例えば、K<sup>+</sup>イオン対Li<sup>+</sup>イオンの存在下で安定化される。

#### 【0522】

使用するベイトオリゴヌクレオチドの数を最小限にするために、各タンパク質に由来する比較的一意のペプチドのセットをバイオインフォマティクスにより識別することができ、目的のペプチドの対応する伸長記録タグライブラリー表示と相補的なベイトオリゴヌクレオチドのみをハイブリッド捕捉アッセイに使用する。同じまたは異なるベイトセットを用いて逐次的なラウンドまたは濃縮を行うこともできる。

10

#### 【0523】

その断片 (例えば、ペプチド) を表す伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリー中の巨大分子 (例えば、タンパク質またはポリペプチド) の全長を濃縮するために、「タイルド」ベイトオリゴヌクレオチドをタンパク質の核酸表示全体にわたって設計することができる。

#### 【0524】

別の実施形態では、プライマー伸長およびライゲーションに基づき媒介される増幅濃縮 (Ampliseq、PCR、TruSeq TSCAなど) を使用して、巨大分子のサブセットを表すライブラリーエレメントを選択し、モジュール画分を濃縮することができる。競合するオリゴを使用して、プライマー伸長、ライゲーション、または増幅の程度を調整することもできる。最も単純な実行では、これは、ユニバーサルプライマー尾部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尾部を欠く競合プライマーの混合物によって実現することができる。最初のプライマー伸長の後、5'ユニバーサルプライマー配列を有する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を有するプライマーとユニバーサルプライマー配列を有さないプライマーの比により、増幅する標的の分率を制御する。他の実施形態では、ハイブリダイズするが伸長しない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、または増幅を受けるライブラリーエレメントの分率をモジュレートすることができる。

20

#### 【0525】

配列決定前に伸長記録タグ、伸長コーディングタグ、またはジタグをライブラリーから選択的に除去するために標的化濃縮法を負の選択形式で使用することもできる。したがって、ビオチン化ベイトオリゴヌクレオチドおよびストレプトアビジンでコーティングしたビーズを使用する上記の例において、上清を配列決定のために保持する一方で、ビーズに結合したベイト-オリゴヌクレオチド：伸長記録タグ、伸長コーディングタグ、またはジタグハイブリッドは解析しない。除去することができる望ましくない伸長記録タグ、伸長コーディングタグ、またはジタグの例は、豊富な巨大分子種、例えば、タンパク質、アルブミン、免疫グロブリンなどを表すものである。

30

#### 【0526】

標的と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ベイトをハイブリッド捕捉ステップに使用して、濃縮しようとする任意の特定の遺伝子座の分率をモジュレートすることもできる。競合オリゴヌクレオチドベイトは、標的とのハイブリダイゼーションについて、濃縮中の標的ブルダウンの分率を有効にモジュレートする標準のビオチン化ベイトと競合する (図17)。特にアルブミンなどの過度に豊富な種に関して、この競合的な抑制手法を使用することで10桁のタンパク質発現のダイナミックレンジを何桁か圧縮することができる。したがって、所与の遺伝子座に関して、標準のハイブリッド捕捉に対する捕捉されるライブラリーエレメントの分率を100%から0%濃縮までモジュレートすることができる。

40

#### 【0527】

さらに、ライブラリー正規化技法を使用して、過度に豊富な種を伸長記録タグ、伸長コーディングタグ、またはジタグライブラリーから除去することができる。この手法は、ト

50

リブシン、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を起源とする定義された長さのライブラリーに対して最良に機能する。一実施例では、正規化は、二本鎖ライブラリーを変性させ、ライブラリーエレメントを再アニーリングさせることによって実現することができる。豊富なライブラリーエレメントは、2分子ハイブリダイゼーションカイネティクスの二次速度定数に起因して、豊富さがより低いエレメントよりも迅速に再アニーリングする（Bochman、Paeschkeら、2012年）。ヒドロキシアパタイトカラムでのクロマトグラフィー（VanderNootら、2012年、Biotechniques、53巻：373～380頁）またはライブラリーを、dsDNAライブラリーエレメントを破壊するタラバガニ由来の2重鎖特異的ヌクレアーゼ（DSN）で処理すること（Shaginら、2002年、Genome Res.、12巻：1935～42頁）などの当技術分野で公知の方法を使用して、ssDNAライブラリーエレメントを豊富なdsDNAライブラリーエレメントから分離することができる。

10

#### 【0528】

固体支持体に付着させる前の巨大分子および/または得られた伸長記録タグライブラリーの分画、濃縮、およびサブトラクション方法の任意の組合せにより、配列決定読み取りを節約し、豊富さが低い種の測定を改善することができる。

#### 【0529】

一部の実施形態では、伸長記録タグ、伸長コーディングタグ、またはジタグのライブラリーをライゲーションまたは末端相補的PCRによって連鎖状にして、それぞれ多数の異なる伸長記録タグ、伸長コーディングタグ、またはジタグを含む長いDNA分子を創出する（それぞれ、その全体が参照により組み込まれる、Duら、2003年、Biotechniques、35巻：66～72頁；Mueckeら、2008年、Structure、16巻：837～841頁；米国特許第5,834,252号）。この実施形態は、長鎖のDNAをナノポアシーケンシングデバイスによって解析するナノポアシーケンシングに好ましい。

20

#### 【0530】

一部の実施形態では、伸長記録タグ、伸長コーディングタグ、またはジタグに対して直接単一分子解析を実施する（例えば、Harrisら、2008年、Science、320巻：106～109頁を参照されたい）。伸長記録タグ、伸長コーディングタグ、またはジタグを、フローセル表面（任意選択でマイクロセルパターン）へのローディングに適合するフローセルまたはビーズなどの固体支持体上で直接解析することができ、ここで、フローセルまたはビーズは、単一分子シーケンサーまたは単一分子脱コーディング計器に組み込むことができる。単一分子脱コーディングに関しては、プールした蛍光標識した脱コーディングオリゴヌクレオチドの数ラウンドのハイブリダイゼーション（Gundersenら、2004年、Genome Res.、14巻：970～7頁）を使用して、伸長記録タグ内のコーディングタグの同一性および順序の両方を確認することができる。コーディングタグの結合の順序をデコンボリューションするために、結合性物質を上記の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（Gundersenら、2004年、Genome Res.、14巻：970～7頁も参照されたい）。サイクル特異的コーディングタグは、単一の巨大分子を表す単一の連鎖状の伸長記録タグ、または単一の巨大分子を表す伸長記録タグの集合のどちらに対しても機能する。

30

40

#### 【0531】

伸長レポータータグ、伸長コーディングタグ、またはジタグライブラリーの配列決定後、得られた配列をそれらのUMIにより崩壊させ、次いで、それらの対応する巨大分子（例えば、ペプチド、タンパク質、タンパク質複合体）に付随させ、細胞内の巨大分子型全体（例えば、ペプチド、ポリペプチド、タンパク質巨大分子についてのプロテオーム）とアラインメントすることができる。得られた配列をそれらのコンパートメントタグにより崩壊させ、特定の実施形態では、単一のタンパク質分子のみまたは非常に限られた数のタンパク質分子を含有する、それらの対応するコンパートメントのプロテオームに付随させ

50

ることにもできる。タンパク質の識別および数量化はどちらも、このデジタルペプチド情報から容易に引き出すことができる。

#### 【0532】

一部の実施形態では、コーディングタグ配列を特定の配列決定解析プラットフォームに対して最適化することができる。特定の実施形態では、配列決定プラットフォームは、ナノポアシーケンシングである。一部の実施形態では、配列決定プラットフォームの塩基当たりのエラー率は、 $> 5\%$ 、 $> 10\%$ 、 $> 15\%$ 、 $> 20\%$ 、 $> 25\%$ 、または $> 30\%$ である。例えば、ナノポアシーケンシング装置を使用して伸長記録タグを解析する場合、バーコード配列（例えば、エンコーダー配列）を、ナノポアを通過時に最適に電氣的に区別可能になるように設計することができる。本明細書に記載の方法に従ったペプチド配列決定は、ナノポアシーケンシングの一塩基正確度はまだ幾分低い（ $75\% \sim 85\%$ ）、「エンコーダー配列」の決定は、はるかにより正確であるはずである（ $> 99\%$ ）ことを考慮すると、ナノポアシーケンシングによく適し得る。さらに、DI (duplex interrupted) ナノポアシーケンシングと称される技法を、システム設計を著しく単純化する分子モーターを必要とせず、ナノポア鎖シーケンシングと共に使用することができる (Derrington、Butlerら、2010年)。DI ナノポアシーケンシングによる伸長記録タグの読み取りには、連鎖状の伸長記録タグライブラリー内のスペーサーエレメントを相補的なオリゴヌクレオチドとアニーリングする必要がある。本明細書で使用されるオリゴヌクレオチドは、得られる2重鎖の有効なTmを上昇させるために、LNA、または他の改変された核酸または類似体を含んでよい。これらの2重鎖スペーサー領域で装飾された一本鎖伸長記録タグがポアを通過するに従い、2本鎖領域が狭窄域で一過性にストールし、それにより、2重鎖領域に隣接する約3塩基の電流読み取りが可能になる。特定の実施形態では、DI ナノポアシーケンシングのために、エンコーダー配列を、スペーサーエレメントに隣接する3塩基により最大限に電氣的に区別可能なナノポアシグナルが生じるように設計する (Derringtonら、2010年、Proc. Natl. Acad. Sci. USA、107巻：16060～5頁)。モーターフリーDIシーケンシングの代替として、スペーサーエレメントを、ナノポアを通過するに従い、伸長記録タグ、伸長コーディングタグ、またはジタグを一過性にストールし、それにより、隣接するエンコーダー配列の読み取りを可能にするG-カルテットなどの二次構造をとるように設計することができる (Shim、Tanら、2009年、Zhang、Zhangら、2016年)。ストールを過ぎて進んだ後、次のスペーサーにより一過性のストールが創出され、それにより、次のエンコーダー配列の読み取りが可能になる、などである。

#### 【0533】

本明細書に開示されている方法は、複数の巨大分子（例えば、ペプチド）の同時（多重化）検出、定量化および/または配列決定を含めた解析のために使用することができる。多重化とは、本明細書で使用される場合、複数の巨大分子を同じアッセイで解析することを指す。複数の巨大分子は、同じ試料に由来するものであってもよく、異なる試料に由来するものであってもよい。複数の巨大分子は、同じ対象に由来するものであってもよく、異なる対象に由来するものであってもよい。解析される複数の巨大分子は、異なる巨大分子（例えば、ペプチド）であってもよく、異なる試料に由来する同じ巨大分子（例えば、ペプチド）であってもよい。複数の巨大分子は、2またはそれよりも多くの巨大分子、5またはそれよりも多くの巨大分子、10またはそれよりも多くの巨大分子、50またはそれよりも多くの巨大分子、100またはそれよりも多くの巨大分子、500またはそれよりも多くの巨大分子、1000またはそれよりも多くの巨大分子、5,000またはそれよりも多くの巨大分子、10,000またはそれよりも多くの巨大分子、50,000またはそれよりも多くの巨大分子、100,000またはそれよりも多くの巨大分子、または1,000,000またはそれよりも多くの巨大分子を含む。

#### 【0534】

試料多重化は、記録タグで標識された巨大分子試料を予めバーコーディングすることによって実現することができる。各バーコードは異なる試料を表し、試料を周期的結合アッセイまたは配列解析前にプールすることができる。このように、多くのバーコードで標識した試料を単一のチューブ内で同時に処理することができる。この手法は、逆相タンパク質アレイ (RPPA) で実施するイムノアッセイに対する有意な改善である (Akban i、Beckerら、2014年、CreightonおよびHuang、2015年、NishizukaおよびMills、2016年)。このように、本開示は、基本的に、単純なワークフローを用いるRPPAアッセイの代わりに、多重化された高度にデジタルの試料および分析物を提供する。

XIII. NTA Aの認識、記録タグ伸長、およびNTA Aの切断の周期的ラウンドによる巨大分子の特徴付け

10

#### 【0535】

ある特定の実施形態では、本開示において提示される巨大分子を解析するための方法は、多数の結合サイクルを有し、ここで、巨大分子を複数の結合性物質と接触させ、連続的な結合性物質の結合により、履歴的な結合情報が核酸に基づくコーディングタグの形態で少なくとも1つの巨大分子に付随する記録タグに移行する。このように、多数の結合事象に関する情報を含有する履歴的記録を核酸形式で生成する。

#### 【0536】

N末端分解に基づく手法を使用してペプチド巨大分子を解析する方法に関する実施形態では、(図3、図4、図41、および図42を参照されたい)、第1の結合性物質をnアミノ酸のペプチドのn NTA Aに接触させ、結合させ、第1の結合性物質のコーディングタグ情報をペプチドに付随する記録タグに移行させ、それにより、一次伸長記録タグを生成した後、n NTA Aを本明細書に記載の通り切断する。n NTA Aの切断により、ペプチドのn-1アミノ酸がN末端アミノ酸に変わり、これは、本明細書ではn-1 NTA Aと称される。本明細書に記載の通り、n NTA Aは、任意選択で、部分(例えば、PTC、DNP、SNP、アセチル、アミジニルなど)で標識することができ、これは、標識された形態のNTA Aに結合するように工学的に操作された切断酵素と併せて特に有用である。n NTA Aを標識した場合、次いで、n-1 NTA Aも同じ部分を用いて標識する。第2の結合性物質をペプチドと接触させ、n-1 NTA Aに結合させ、第2の結合性物質のコーディングタグ情報を一次伸長記録タグに移行させ、それにより、二次伸長記録タグ(例えば、ペプチドを表す連鎖状のn次伸長記録タグを生成するため)または異なる記録タグ(例えば、ペプチドを集合的に表す多数の伸長記録タグを生成するため)を生成する。n-1 NTA Aの切断により、ペプチドのn-2アミノ酸がN末端アミノ酸に変わり、これは、本明細書ではn-2 NTA Aと称される。追加的な結合、移行、切断、および任意選択でNTA A標識を、上記の通り、最大nアミノ酸まで行って、ペプチドを集合的に表す、n次伸長記録タグまたはn個の別々の伸長記録タグを生成することができる。本明細書で使用される場合、n「次」とは、結合性物質、コーディングタグ、または伸長記録タグに関して使用される場合、結合性物質およびそれに付随するコーディングタグを使用することができるn回の結合サイクル、または、伸長記録タグを創出するn回の結合サイクルを指す。

20

30

40

#### 【0537】

一部の実施形態では、第1の結合性物質および第2の結合性物質、ならびに任意選択で任意のさらなる結合性物質(例えば、第3の結合性物質、第4の結合性物質、第5の結合性物質など)の巨大分子への接触を同時に実施する。例えば、第1の結合性物質および第2の結合性物質、ならびに任意選択で任意のさらなる次数の結合性物質を、例えば結合性物質のライブラリーを形成するために、一緒にプールすることができる。別の例では、第1の結合性物質および第2の結合性物質、ならびに任意選択で任意のさらなる次数の結合性物質を、一緒にプールするのではなく、巨大分子に同時に添加する。一実施形態では、結合性物質のライブラリーは、20種の標準の天然に存在するアミノ酸に選択的に結合する少なくとも20種の結合性物質を含む。

50

## 【 0 5 3 8 】

他の実施形態では、第 1 の結合性物質および第 2 の結合性物質、ならびに任意選択で任意のさらなる次数の結合性物質をそれぞれ別々の結合サイクルで巨大分子をと接触させ、逐次的に添加する。ある特定の実施形態では、多数の結合性物質を同時に使用することが好ましく、その理由は、並行手法により時間が節約されるから、および、結合性物質が競合状態になり、それにより、同類結合性物質が結合する部位への非同類結合性物質による非特異的結合が低減するからである。

## 【 0 5 3 9 】

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、エンコーダー配列およびスパーサー）の長さ、記録タグ（例えば、一意の分子識別子、スパーサー、ユニバーサルプライミング部位、バーコード）の長さ、実施される結合サイクルの数、および各結合サイクルからのコーディングタグが同じ伸長記録タグに移行されるか多数の伸長記録タグに移行されるかを含めた多数の因子に依存する。ペプチドを表し、エドマン分解様切断方法によって作製される連鎖状の伸長記録タグの例では、コーディングタグが、両側に 5 塩基のスパーサーが隣接する 5 塩基のエンコーダー配列を有する場合、ペプチドの結合性物質履歴を示す最終的な伸長記録タグ上のコーディングタグ情報は、10 塩基 × エドマン分解サイクルの数になる。20 サイクルの実行に関しては、伸長記録は、少なくとも 200 塩基である（最初の記録タグ配列は含めない）。この長さは、標準の次世代シーケンシング計器に適合する。

## 【 0 5 4 0 】

最終的な結合サイクルおよび最終的な結合性物質のコーディングタグ情報の伸長記録タグへの移行の後、記録タグに、ユニバーサルリバースプライミング部位をライゲーション、プライマー伸長または当技術分野で公知の他の方法によって付加することによってキャップ形成することができる。一部の実施形態では、記録タグ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記録タグに付属するユニバーサルリバースプライミング部位に適合する。一部の実施形態では、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（5' - C A A G C A G A A G A C G G C A T A C G A G A T - 3' - 配列番号 134）または Illumina P5 プライマー（5' - A A T G A T A C G G C G A C C A C C G A - 3' - 配列番号 133）である。記録タグの鎖のセンスに応じてセンスまたはアンチセンス P7 を付属させることができる。伸長記録タグライブラリーを固体支持体（例えば、ビーズ）から直接切断または増幅し、従来の次世代シーケンシングアッセイおよびプロトコルに使用することができる。

## 【 0 5 4 1 】

一部の実施形態では、プライマー伸長反応を一本鎖伸長記録タグのライブラリーに対して実施して、その相補鎖をコピーする。

## 【 0 5 4 2 】

NGPS ペプチドシーケンシングアッセイは、いくつかの化学的ステップおよび酵素的ステップを周期的進行で含む。NGPS シーケンシングは単一分子であるという事実により、いくつかの重要な利点がプロセスに付与される。単一分子アッセイの第 1 の重要な利点は、種々の周期的化学的 / 酵素的ステップにおける非効率に対する頑強性である。これは、コーディングタグ配列内に存在するサイクル特異的バーコードを使用することによって可能になる。

## 【 0 5 4 3 】

サイクル特異的コーディングタグを使用して、各サイクルからの情報を追跡する。これは単一分子シーケンシング手法であるので、配列決定プロセスの各結合 / 移行サイクルにおける 70 % の効率でさえ、マッピング可能な配列情報を生成するために十分すぎるほどである。例として、10 塩基ペプチド配列「C P V Q L W V D S T」（配列番号 169）は、我々の配列プラットフォームでは「C P X Q X W X D X T」（配列番号 170）と読み取られる可能性がある（X = 任意のアミノ酸；アミノ酸の存在はサイクル数追跡によって推定される）。この部分的なアミノ酸配列読み取りは、BLASTP を使用してそれを

10

20

30

40

50

ヒト p 5 3 タンパク質に一意的にマッピングし戻すために十分すぎるほどである。そのように、我々のプロセスはいずれも完全にロバストにする必要はない。さらに、サイクル特異的バーコードを我々の分配概念と組み合わせれば、どのペプチドのセットが元のタンパク質分子にマッピングされるかが（コンパートメントバーコードによって）わかるので、10カ所の位置のうちの数アミノ酸を識別するだけでタンパク質の絶対的な識別を実現することができる。

#### 【0544】

XIV. 分画、区画化、および結合能が限定された樹脂によるタンパク質正規化

#### 【0545】

プロテオミクス解析の重要な問題の1つは、試料中のタンパク質の豊富さの大きなダイナミックレンジに取り組むことである。タンパク質は、血漿中で10桁を超えるダイナミックレンジにわたる（「上位20種」が枯渇した血漿でさえ）。ある特定の実施形態では、解析前に、試料からのある特定のタンパク質種（例えば、極めて豊富なタンパク質）のサブトラクションを実施する。これは、例えば、上位20種の血漿タンパク質を枯渇させるSigmaのPROT20免疫枯渇キットなどの市販のタンパク質枯渇試薬を使用して実現することができる。さらに、ダイナミックレンジを管理可能な3~4桁までさらに著しく低下させる手法をとることが有用であると思われる。ある特定の実施形態では、タンパク質試料のダイナミックレンジは、電気泳動および液体クロマトグラフィーを含めた標準的分画法を使用してタンパク質試料を分画するか（Zhou、Ningら、2012年）、または、画分を、限られた能力のタンパク質結合性ビーズ/樹脂（例えば、ヒドロキシル化シリカ粒子）をローディングしたコンパートメント（例えば、液滴）に分配し（McCormick、1989年）、結合したタンパク質を溶出することによってモジュレートすることができる。各区画化された画分中の過剰なタンパク質を洗い流す。

#### 【0546】

電気泳動による方法の例としては、キャピラリー電気泳動（CE）、キャピラリー等電点電気泳動（CIEF）、キャピラリー等速電気泳動（CITP）、フリーフロー電気泳動、ゲル溶出分画液相封入電気泳動（GELFREE）が挙げられる。液体クロマトグラフィータンパク質の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）、親水性相互作用などが挙げられる。コンパートメント分配の例としては、エマルジョン、液滴、マイクロウェル、平ら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性ビーズ/樹脂としては、フェノール基またはヒドロキシル基で誘導体化されたシリカナノ粒子（例えば、StrataClean Resin（Agilent Technologies製）、RapidClean（LabTech製）など）が挙げられる。ビーズ/樹脂の結合能を限定することにより、所与の画分中に溶出する高度に豊富なタンパク質は部分的にビーズに結合するだけになり、過剰なタンパク質は除去される。

XV. 単一の細胞のプロテオームの分配または分子サブサンプリング

#### 【0547】

別の態様では、本開示は、バーコーディングおよび分配技法を使用して試料中のタンパク質を大規模並列解析するための方法を提供する。タンパク質解析の現行の手法は、タンパク質巨大分子をペプチド配列決定に適した短いペプチド分子に断片化することを伴う。したがって、そのような手法を使用して得られる情報は、断片化ステップによって限定され、例えば、翻訳後修飾、各試料中に存在するタンパク質間相互作用、試料中に存在するタンパク質集団の組成、または特定の細胞または細胞の集団に由来するなどのタンパク質巨大分子の起源を含めたタンパク質の広範な連続した情報は排除される。タンパク質分子内の翻訳後修飾に関する広範な情報（例えば、プロテオフォーム特徴付け）により、より徹底的な生物学的像がもたらされ、どのペプチドがどのタンパク質分子に属するかに関する広範な情報により、ペプチド配列の、基礎をなすタンパク質配列へのよりロバストなマッピングがもたらされる（図15Aを参照されたい）。これは、ペプチド配列決定技術によりたった5種のアミノ酸型からの情報などの不完全なアミノ酸配列情報がもたらされる

10

20

30

40

50



場合に特に関連する。本明細書に開示されている分配方法を使用することにより、同じタンパク質分子に由来するいくつかのペプチドからの情報と組み合わせて、タンパク質分子の同一性（例えば、プロテオフォーム）をより正確に評価することができる。コンパートメントタグを同じコンパートメント（複数可）に由来するタンパク質およびペプチドに付随させることにより、分子および細胞情報の再構成が容易になる。典型的なプロテオーム解析では、細胞を溶解させ、タンパク質を短いペプチドに消化し、それにより、どのタンパク質がどの細胞または細胞型に由来するか、およびどのペプチドがどのタンパク質またはタンパク質複合体に由来するかに関する全体的な情報が破壊される。この全体的な情報は、細胞および組織内の生物学および生化学を理解するために重要である。

#### 【0548】

分配とは、試料内の巨大分子の集団に由来する巨大分子の亜集団への一意のバーコードのランダムな割り当てを指す。分配は、巨大分子をコンパートメントに区分することによって実現することができる。分配は、単一のコンパートメント内の巨大分子で構成されるものであってもよく、コンパートメントの集団に由来する多数のコンパートメント内の巨大分子で構成されるものであってもよい。

#### 【0549】

複数（例えば、数百万～数十億）のコンパートメントから同じ物理的コンパートメントまたはコンパートメントの群に、またはそれにおいて分離された巨大分子のサブセットまたはタンパク質試料のサブセットを一意的コンパートメントタグによって識別する。したがって、コンパートメントタグを使用して、同じコンパートメントタグを有する1つまたは複数のコンパートメントに由来する構成物と、異なるコンパートメントタグを有する別のコンパートメント（またはコンパートメントの群）内の構成物を、これらの構成物が一緒にプールされた後であっても区別することができる。

#### 【0550】

本開示は、複雑なプロテオーム試料（例えば、複数のタンパク質複合体、タンパク質、またはポリペプチド）または複雑な細胞試料を複数のコンパートメントに分配することによってタンパク質解析を増強する方法であって、各コンパートメントが、個々のコンパートメント内では同じであり（任意選択のUMI配列は除いて）、他のコンパートメントのコンパートメントタグとは異なる複数のコンパートメントタグを含む、方法を提供する（図18～20を参照されたい）。コンパートメントは、任意選択で、複数のコンパートメントタグが接合した固体支持体（例えば、ビーズ）を含む。複数のタンパク質複合体、タンパク質、またはポリペプチドを複数のペプチドに断片化し、次いで、それを複数のコンパートメントタグと、複数のコンパートメント内で複数のペプチドと複数のコンパートメントタグとのアニーリングまたは接合を可能にするのに十分な条件下で接触させ、それにより、複数のコンパートメントタグ付きペプチドを生成する。あるいは、複数のタンパク質複合体、タンパク質、またはポリペプチドに、複数のコンパートメントタグを、複数のタンパク質複合体、タンパク質またはポリペプチドと複数のコンパートメント内の複数のコンパートメントタグのアニーリングまたは接合を可能にするために十分な条件下で接合し、それにより、複数のコンパートメントタグ付きタンパク質複合体、タンパク質、ポリペプチドを生成する。次いで、コンパートメントタグ付きタンパク質複合体、タンパク質、またはポリペプチドを複数のコンパートメントから収集し、任意選択で複数のコンパートメントタグ付きペプチドに断片化する。1つまたは複数のコンパートメントタグ付きペプチドを本明細書に記載の方法のいずれかに従って解析する。

#### 【0551】

ある特定の実施形態では、コンパートメントタグ情報を巨大分子に付随する記録タグ（例えば、ペプチド）にプライマー伸長（図5）またはライゲーション（図6）によって移行させる。

#### 【0552】

一部の実施形態では、コンパートメントタグは、コンパートメント内で溶液中に遊離している。他の実施形態では、コンパートメントタグをコンパートメントの表面（例えば、

10

20

30

40

50

マイクロタイタープレートもしくはピコタイタープレートのウェルの底) またはビーズもしくはコンパートメント内のビーズに直接接合させる。

【0553】

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)であっても固体コンパートメントであってもよい。固体コンパートメントとしては、例えば、ナノ粒子、マイクロスフェア、マイクロタイターウェルもしくはピコタイターウェル、またはアレイ、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、ナイロン、シリコンウェーハチップ、フローセル、フロースルーチップ、信号変換電子機器を含むバイオチップ、ELISAプレート、スピン干渉ディスク、ニトロセルロースメンブレン、もしくはニトロセルロースに基づくポリマー表面上の分離された領域が挙げられる。ある特定の  
10 実施形態では、各コンパートメントは、平均して、単一の細胞を含有する。

【0554】

固体支持体は、これだけに限定されないが、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、ナイロン、シリコンウェーハチップ、フローセル、フロースルーチップ、信号変換電子機器を含むバイオチップ、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、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアを含めた任意の支持体表面であってもよい。固体支持体用の材料としては、これだけに限定されないが、アクリルアミド、アガロース、セルロース、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、ポリメタクリレート、  
20 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、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、ポリ酸無水物、ポリグリコール酸、ポリ乳酸、ポリオルトエステル、官能化シラン、ポリプロピルフメレート、コラーゲン、グリコサミノグリカン、ポリアミノ酸、またはそれらの任意の組合せが挙げられる。ある特定の実施形態では、固体支持体は、ビーズ、例えば、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズである。

【0555】

コンパートメントタグを付けたビーズを伴うコンパートメントに試料を分配する種々の方法は、Shembekarらにより概説されている(Shembekar、Chai  
30 panら、2016年)。一実施例では、プロテオームをエマルジョンにより液滴中に分配して、タンパク質分子およびタンパク質複合体に関する全体的な情報を本明細書に開示されている方法を使用して記録することを可能にする(例えば、図18および図19を参照されたい)。ある特定の実施形態では、プロテオームをコンパートメント(例えば、液滴)に、コンパートメントタグを付けたビーズ、活性化可能なプロテアーゼ(直接または間接的に熱、光などによって)、およびプロテアーゼ抵抗性になるように工学的に操作された(例えば、修飾リシン、ペグ化など)ペプチドリガーゼと一緒に分配する。ある特定の  
40 実施形態では、プロテオームを変性剤で処理して、タンパク質またはポリペプチドのペプチド構成物を評価することができる。タンパク質のネイティブな状態に関する情報が望まれる場合、相互作用するタンパク質複合体を、それに由来するペプチドのその後の分析のためにコンパートメントに分配することができる。

【0556】

コンパートメントタグは、任意選択で片側または両側にスペーサーまたはユニバーサルプライマー配列が隣接するバーコードを含む。プライマー配列は、記録タグの3'配列に対して相補的であってよく、それにより、プライマー伸長反応によるコンパートメントタグ情報の記録タグへの移行が可能になる(図22A~Bを参照されたい)。バーコードは、固体支持体もしくはコンパートメントに付着した一本鎖核酸分子または固体支持体もしくはコンパートメントとハイブリダイズしたその相補配列、または両方の鎖で構成されるものである  
50 であってよい(例えば、図16を参照されたい)。コンパートメントタグは、ペプチドとのカップリングのための、例えばスペーサーに付着した、機能的部分を含んでよい。一

実施例では、機能的部分（例えば、アルデヒド）は、複数のペプチド上のN末端アミノ酸残基と反応することが可能である。別の例では、機能的部分は、複数のペプチド上の内部アミノ酸残基（例えば、リシンまたは「クリック」反応性部分で標識されたリシン）と反応することが可能である。別の実施形態では、機能的部分は、単に、DNAタグで標識されたタンパク質とハイブリダイズすることが可能な相補DNA配列であってよい。あるいは、コンパートメントタグは、コンパートメントタグの目的のペプチドへのライゲーションを可能にするために、タンパク質リガーゼ（例えば、プテラーゼIまたはそのホモログ）の認識配列を含むペプチドをさらに含むキメラ分子であってよい（図22Aを参照されたい）。コンパートメントタグは、より大きな核酸分子内の成分であってよく、当該核酸分子は、それに接合したペプチドに関する識別情報をもたらすための一意の分子識別子、スパーサー配列、ユニバーサルプライミング部位、またはそれらの任意の組合せを任意選択でさらに含む。このUMI配列は、一般に、コンパートメント内のコンパートメントタグの集団の間で異なる。ある特定の実施形態では、コンパートメントタグは、記録タグ内の成分であり、したがって、個々のコンパートメント情報をもたらすために使用する同じタグを、それが付着したペプチドに関する個々のペプチド情報を記録するためにも使用する。

10

#### 【0557】

ある特定の実施形態では、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プリント、スポッティング、インク噴射することによって形成することができる。ある特定の実施形態では、複数のコンパートメントタグを付けたビーズであって、ビーズ当たり1つのバーコード型が存在するビーズを、Kleinら、2015年、Cell、161巻：1187～1201頁；Macoskoら、2015年、Cell、161巻：1202～1214頁；およびFanら、2015年、Science、347巻：1258367頁に記載されている通り、スプリット・アンド・プール（split-and-pool）オリゴヌクレオチドライゲーションまたは合成によって形成する。コンパートメントタグを付けたビーズは、個々の合成または固定化によって形成することもできる。ある特定の実施形態では、コンパートメントタグを付けたビーズは、一方の部分が記録タグを含むコンパートメントタグを含み、他方の部分が消化されたペプチドをカップリングすることができる機能的部分を含む二機能性記録タグをさらに含む（図19および図20）。

20

30

#### 【0558】

ある特定の実施形態では、複数のコンパートメント内の複数のタンパク質またはポリペプチドを、プロテアーゼを用いて複数のペプチドに断片化する。プロテアーゼは、メタロプロテアーゼであってよい。ある特定の実施形態では、メタロプロテアーゼの活性が金属カチオンの光活性化放出によってモジュレートされる。使用することができるエンドペプチダーゼの例としては、トリプシン、キモトリプシン、エラスターゼ、サーモリシン、ペプシン、クロストリパン、グルタミルエンドペプチダーゼ（GluC）、エンドペプチダーゼArgC、ペプチジル-aspmetaロ-エンドペプチダーゼ（AspN）、エンドペプチダーゼLysCおよびエンドペプチダーゼLysNが挙げられる。それらの活性化の形式は、緩衝液および二価カチオンの必要性に応じて変動する。任意選択で、タンパク質またはポリペプチドをペプチド断片に十分に消化した後、プロテアーゼを不活性化（例えば、熱、フルオロ油またはシリコン油可溶性阻害剤、例えば、二価カチオンキレート化剤など）。

40

#### 【0559】

コンパートメントタグを用いたペプチドバーコーディングのある特定の実施形態では、タンパク質分子（任意選択で、変性ポリペプチド）を、DNAタグを用いて、DNAタグをタンパク質のリシン基の - アミン部分とのコンジュゲーションによって、またはアルキンなどの反応性クリック部分で予め標識したタンパク質/ポリペプチドに付着させるクリックケミストリーによって間接的に標識する（図2Bおよび図20Aを参照されたい）。次いで、DNAタグで標識したポリペプチドを、コンパートメントタグ（例えば、液滴

50

内に含有されるビーズに結合したDNAバーコード)を含むコンパートメントに分配し(図20Bを参照されたい)、ここで、コンパートメントタグは、各コンパートメントを識別するバーコードを含有する。一実施形態では、ビーズに付随する単一のタンパク質/ポリペプチド分子と単一の種のDNAバーコードを共封入する(図20Bを参照されたい)。別の実施形態では、PCT公開第WO2016/061517号(その全体が参照により組み込まれる)に記載されているものと、DNAではなくタンパク質に適用されること以外は同様に、コンパートメントは、付着したコンパートメント(ビーズ)タグと共にビーズの表面を構成し得る。コンパートメントタグは、バーコード(BC)配列、ユニバーサルプライミング部位(U1')、UMI配列、およびスパーサー配列(Sp)を含んでよい。一実施形態では、分配と同時にまたはその後、コンパートメントタグをビーズから切断し、ポリペプチドに付着したDNAタグと、例えば、それぞれDNAタグおよびコンパートメントタグ上の相補的なU1配列およびU1'配列を介してハイブリダイズさせる。ビーズへの分配に関しては、DNAタグで標識したタンパク質をビーズ表面上のコンパートメントタグと直接ハイブリダイズさせることができる(図20Cを参照されたい)。このハイブリダイゼーションステップ後、DNAタグとハイブリダイズしたポリペプチドをコンパートメントから抽出し(例えば、エマルジョン「分解」、またはビーズからのコンパートメントタグの切断)、ポリメラーゼに基づくプライマー伸長ステップを使用して、バーコードおよびUMI情報をポリペプチド上のDNAタグに書き込んでコンパートメントバーコードが付された記録タグをもたらす(図20Dを参照されたい)。ポリペプチドを、C末端リシンにおいてユニバーサルプライミング配列を含有する記録タグ、コンパートメントタグ、およびUMIで標識されたペプチドに切断するために、LysCプロテアーゼ消化を使用することができる(図20Eを参照されたい)。一実施形態では、LysCプロテアーゼを、DNAタグが付されたリシン残基が許容されるように工学的に操作する。得られる記録タグで標識されたペプチドを、固体基板(例えば、ビーズ)に、記録タグが付されたペプチド間の分子間相互作用を最小限にするために適した密度で固定化する(図20Eおよび20Fを参照されたい)。

#### 【0560】

ペプチドのコンパートメントタグへの付着(または逆もまた同じ)は、固定化されたコンパートメントタグへの直接付着であってもよく、その相補配列(二本鎖の場合)への付着であってもよい。あるいは、コンパートメントタグを固体支持体またはコンパートメントの表面から引き離し、ペプチドおよび液相コンパートメントタグをコンパートメント内に接合させることができる。一実施形態では、コンパートメントタグ(例えば、オリゴヌクレオチドの末端)上の機能的部分は、ペプチドのN末端のアミンにシッフ塩基を通じて直接カップリングしたアルデヒドである(図16を参照されたい)。別の実施形態では、コンパートメントタグを、タンパク質リガーゼに対するペプチドモチーフ( $n-X\cdots XXCGSHV-c$ )を含む核酸-ペプチドキメラ分子として構築する。核酸-ペプチドコンパートメントタグ構築物を、消化されたペプチドと、プテラーゼIまたはそのホモログなどのペプチドリガーゼを使用してコンジュゲートする。プテラーゼI、および他のアスパラギニルエンドペプチダーゼ(AEP)相同体を使用して、オリゴヌクレオチド-ペプチドコンパートメントタグ構築物のC末端と消化されたペプチドのN末端をライゲーションすることができる(Nguyen、Wangら、2014年、Nguyen、Caoら、2015年)。この反応は、速く、高度に効率的である。得られたコンパートメントタグ付きペプチドを、その後、本明細書に記載の核酸ペプチド解析のために固体支持体に固定化することができる。

#### 【0561】

ある特定の実施形態では、コンパートメントタグと複数の断片化されたペプチドを接合させる前に、固体支持体またはコンパートメントの表面に接合しているコンパートメントタグを放出させる(図18を参照されたい)。一部の実施形態では、コンパートメントタグ付きペプチドを複数のコンパートメントから収集した後、コンパートメントタグ付きペプチドを、記録タグを伴って固体支持体に接合する。次いで、コンパートメントタグ情報

10

20

30

40

50

をコンパートメントタグ付きペプチド上のコンパートメントタグから付随する記録タグに移行させることができる（例えば、記録タグおよびコンパートメントタグ内の相補的なスペーサー配列からプライミングされるプライマー伸長反応によって）。一部の実施形態では、次いで、本明細書に記載の方法に従ったペプチド解析の前に、コンパートメントタグをコンパートメントタグ付きペプチドから除去する。さらなる実施形態では、最初に複数のタンパク質を消化するために使用した配列特異的プロテアーゼ（例えば、Endo A s p N）を、コンパートメントタグ情報を付随する記録タグに移行した後、ペプチドのN末端からのコンパートメントタグの除去にも使用する（図22Bを参照されたい）。

#### 【0562】

コンパートメントに基づく分配のための手法は、Tジャンクションおよびフローフォーカシング、攪拌または小さな穴を有する膜（例えば、エッチング飛跡膜）を通じた押出しを使用したエマルジョン生成などを使用するマイクロ流体デバイスによる液滴形成などを含む（図21を参照されたい）。区画化に伴う問題は、コンパートメントの内部の取扱いである。ある特定の実施形態では、コンパートメント内で一連の異なる生化学的ステップを行うことは、流体成分の交換が困難であるので、難しい場合がある。以前に記載されている通り、液滴内部の限られた特徴、例えば、pH、キレート化剤、還元剤などは、エマルジョンのフルオロ油に試薬を添加することによって改変することができる。しかし、水性相と有機相両方の溶解性を有する化合物の数は限られている。1つの手法は、コンパートメント内の反応を、基本的に目的の分子へのバーコードの移行に限定することである。

#### 【0563】

タンパク質/ペプチドをコンパートメントタグ（バーコード）で構成される記録タグで標識した後、タンパク質/ペプチドを、固体支持体に、結合した同類結合性物質のコーディングタグから結合したペプチドまたはタンパク質分子に付着した対応する記録タグ/タグへの情報の分子内移行を有利にするために適した密度で固定化する。分子間情報移行は、固体支持体の表面上の分子の分子間間隔を制御することによって最小化される。

#### 【0564】

ある特定の実施形態では、コンパートメントタグは、コンパートメントの集団内の各コンパートメントに対して一意である必要はない。コンパートメントの集団内のコンパートメントのサブセット（2つ、3つ、4つ、またはそれよりも多く）は、同じコンパートメントタグを共有してよい。例えば、各コンパートメントは、試料から巨大分子の亜集団を捕捉するように作用するビーズ表面の集団で構成されてよい（ビーズ当たり多くの分子が捕捉される）。さらに、ビーズは、捕捉される巨大分子に付着させることができるコンパートメントバーコードを含む。各ビーズは、単一のコンパートメントバーコード配列のみを含むが、このコンパートメントバーコードは、コンパートメント内の他のビーズ上で複製され得る（多くのビーズが同じバーコードにマッピングされる）。物理的コンパートメントとコンパートメントバーコードの間で多対1のマッピングをなすことができ（必要ではないが）、さらに、コンパートメント内の巨大分子間で多対1のマッピングをなすことができる（必要ではないが）。分配バーコードは、試料内の巨大分子の集団から巨大分子をサブサンプリングするための一意のバーコードの割り当てと定義される。この分配バーコードは、同じバーコードで標識されたコンパートメント内の巨大分子の分配から生じる同一のコンパートメントバーコードで構成されてよい。物理的コンパートメントの使用により、元の試料が有効にサブサンプリングされて、分配バーコードの割り当てがもたらされる。例えば、10,000種の異なるコンパートメントバーコードで標識されたビーズのセットがもたらされる。さらに、所与のアッセイにおいて、ビーズ百万個の集団をアッセイに使用されると仮定する。平均で、コンパートメントバーコード当たり100個のビーズが存在する（ポアソン分布）。さらに、ビーズにより巨大分子1千万個の凝集体が捕捉されると仮定する。平均で、ビーズ当たり10個の巨大分子が存在し、コンパートメントバーコード当たりのコンパートメントは100個であり、分配バーコード（100個の別個の物理的コンパートメント当たり100個のコンパートメントバーコードで構成される）当たり有効に1000個の巨大分子が存在する。

10

20

30

40

50

## 【0565】

別の実施形態では、ポリペプチドの単一分子分配および分配バーコーディングは、ポリペプチドを、N末端またはC末端またはその両方において、増幅可能なDNA UMIタグ（例えば、記録タグ）を用いて標識すること（化学的にまたは酵素的に）によって実現される（図37を参照されたい）。DNAタグをポリペプチドのボディ（内部アミノ酸）に非特異的な光標識または図2Bにおいて例示されている通りリシンなどの反応性アミノ酸への特異的な化学的付着によって付着させる。ペプチドの末端に付着させた記録タグからの情報をDNAタグに酵素的エマルジョンPCR（Williams、Peisajovichら、2006年、Schutze、Rubeltら、2011年）またはエマルジョンin vitro転写/逆転写（IVT/RT）ステップによって移行させる。好ましい実施形態では、ナノエマルジョンを使用し、その結果、平均して、50nm~1000nmのサイズのエマルジョン液滴当たり単一のポリペプチド未滴が存在するようにする（Nishikawa、Sunamiら、2012年、Gupta、Eralら、2016年）。さらに、プライマー、dNTP、Mg<sup>2+</sup>、ポリメラーゼ、およびPCR緩衝液を含めたPCRの全ての成分を水性エマルジョン混合物に含める。IVT/RTを使用する場合には、記録タグを、T7/SP6 RNAポリメラーゼプロモーター配列を用い、ポリペプチドのボディに付着したDNAタグにハイブリダイズする転写物が生成されるように設計する（Ryckelynck、Baudreyら、2015年）。逆転写酵素（RT）により、ハイブリダイズしたRNA分子からDNAタグに情報がコピーされる。このように、エマルジョンPCRまたはIVT/RTを使用して、末端記録タグからポリペプチドのボディに付着させた多数のDNAタグに情報を有効に移行させることができる。

10

20

## 【0566】

細胞内容物のビーズへのゲル化による封入は、単一の細胞を解析するために有用な手法である（TamminenおよびVirta 2015、Spencer、Tamminenら、2016年）。単一の細胞液滴のバーコーディングにより、単一の細胞に由来する成分を全て同じ識別子で標識することが可能になる（Klein、Mazutisら、2015年、Gundersen、Steemersら、2016年、Zilionis、Nainysら、2017年）。コンパートメントバーコーディングは、一意のバーコードを、各液滴に、液滴接合によって（Raindance）、バーコードが付されたビーズを液滴に導入することによって（10x Genomics）、または、その全体が参照により組み込まれる、Gundersenら（Gundersen、Steemersら、2016年）およびPCT公開第WO2016/130704号に記載されている通り、封入およびゲル化後にスプリット・プール（split-pool）コンビナトリアルバーコーディングを使用して液滴の成分をコンビナトリアルバーコーディングすることによって、直接組み入れることを含めた、いくつかのやり方で実現することができる。Adeyら（Vitak、Torkenczyら、2017年）に記載されている通り、同様のコンビナトリアル標識スキームを核にも適用することができる。

30

## 【0567】

上記の液滴バーコーディング手法は、DNA解析には使用されているが、タンパク質解析には使用されていない。上記の液滴バーコーディングプラットフォームをタンパク質を用いた研究に適合させるには、いくつかの革新的なステップが必要である。まず第1は、バーコードは主にDNA配列で構成され、このDNA配列情報をタンパク質分析物に付与する必要があることである。DNA分析物の場合では、DNA情報をDNA分析物に移行させることは比較的簡単である。対照的に、DNA情報をタンパク質に移行させることは、特に、下流の解析のためにタンパク質を変性させ、消化してペプチドにする場合には、より困難である。これにより、各ペプチドをコンパートメントバーコードで標識することが必要になる。問題は、細胞が液滴に封入されたら、タンパク質を変性させ、得られたポリペプチドをプロテアーゼ消化し、同時にペプチドをDNAバーコードで標識するのが難しいことである。細胞をポリマー形成液滴に封入し、それらを重合（ゲル化）して、水性緩衝液にすることができる多孔質ビーズにすることにより、液滴中の細胞とは異なり、多

40

50

数の異なる反応ステップを実施するためのビヒクルがもたらされる (Tamminen および Virta、2015 年、Spencer、Tamminen ら、2016 年) (Gundersen、Steemers ら、2016 年)。封入されたタンパク質がその後ゲルビーズから拡散することを防止するために、封入されたタンパク質をゲルマトリックスと架橋結合させることが好ましい。このゲルビーズ形式により、ゲル中に閉じ込められたタンパク質を化学的にまたは酵素的に変性させ、DNA タグで標識し、プロテアーゼ消化し、いくつもの他の介入に供することが可能になる。図 38 は、例示的な単一の細胞のゲルマトリックスへの封入および溶解を示す。

#### XVI. 組織および単一の細胞の空間的プロテオミクス

##### 【0568】

バーコードの別の使用は、空間的に分布した DNA バーコード配列のアレイ表面上での組織の空間的分割である。組織タンパク質を、アレイ表面に乗せた細胞組織内のタンパク質の空間的位置を反映するバーコードを含む DNA 記録タグで標識すれば、Stahl ら (2016 年、Science、353 巻 (6294 号) : 78 ~ 82 頁) および Crosetto ら (Crosetto、Bienko ら、2015 年) に記載されている空間的トランスクリプトームと同様に、組織スライス内のタンパク質分析物の空間的分布を配列解析後に再構築することができる。空間バーコードの付着は、アレイに結合させたバーコードをアレイから放出させ、それらを組織切片中に拡散させることによって実現することができる、あるいは、組織切片中のタンパク質を DNA 記録タグで標識し、次いで、タンパク質をプロテアーゼで消化して、拡散し、アレイ上の空間バーコードとハイブリダイズする標識されたペプチドを放出させることができる。次いで、バーコード情報をペプチドに付着させた記録タグに移行させることができる (酵素的にまたは化学的に)。

##### 【0569】

組織内のタンパク質の空間的バーコーディングは、DNA 記録タグで化学的に標識された、固定 / 透過処理した組織スライスを、空間的にコードされた DNA アレイであって、アレイ上の各特徴が空間的に識別可能なバーコードを有する、DNA アレイに置くことによって実現することができる (図 23 を参照されたい)。アレイバーコードを DNA タグに付着させるために、組織スライスをプロテアーゼで消化し、それにより、拡散し、組織スライスに隣接する近位のアレイ特徴にハイブリダイズすることができる DNA タグで標識されたペプチドを放出させることができる。アレイバーコード情報は、化学的 / 酵素的ライゲーションまたはポリメラーゼ伸長によって DNA タグに移行させることができる。あるいは、標識されたペプチドをアレイ表面に拡散させる代わりに、アレイ上のバーコード配列を切断し、組織スライス上の近位の領域に拡散させ、その中の DNA タグで標識されたタンパク質にハイブリダイズさせることができる。再度、バーコーディング情報は、化学的 / 酵素的ライゲーションまたはポリメラーゼ伸長によって移行させることができる。この第 2 の場合には、バーコード情報の移行後にプロテアーゼ消化を実施することができる。いずれの手法の結果も、記録タグで標識されたタンパク質またはペプチドの収集であり、ここで、記録タグは、元の組織内のタンパク質 / ペプチドの位置に関する 2 - D 空間的情報を有するバーコードを含む。さらに、翻訳後修飾の空間的分布を特徴付けることができる。この手法により、感度が高く、高度に多重化された in situ デジタル免疫組織化学的アッセイがもたらされ、また、現代の分子病理学の基礎が形成され、それによりはるかに正確な診断および予後判定が導かれるはずである。

##### 【0570】

別の実施形態では、細胞小器官および細胞区画内のタンパク質構成物 / PTM を識別するために、細胞内に空間的バーコーディングを使用することができる (Christoforou ら、2016 年、Nat. Commun.、7 巻 : 8992 頁、その全体が参照により組み込まれる)。近位のタンパク質に付着させることができる細胞内空間バーコードをもたらすために、いくつもの手法を使用することができる。一実施形態では、細胞または組織を構成物細胞小器官中に細胞内分画し、異なるタンパク質細胞小器官画分バーコードを付すことができる。空間的な細胞標識の他の方法は、その全体が参照により組み

10

20

30

40

50

込まれる、Marx、2015年、Nat Methods、12巻：815～819頁による総説に記載されている；同様の手法を本明細書で使用する事ができる。

【0571】

以下の実施例は例示のために提示されるものであり、限定するものではない。

【実施例】

【0572】

（実施例1）

プロテイナーゼKによるタンパク質試料の消化

ペプチドのライブラリーを、トリプシン、プロテイナーゼKなどのプロテアーゼで消化することによりタンパク質試料から調製する。トリプシンは、好ましくは、リシンおよびアルギニンのような正電荷のアミノ酸のC末端側を切断するが、プロテイナーゼKは、タンパク質の全体にわたって非選択的に切断する。そのため、プロテイナーゼK消化は、短鎖ペプチド（約30個のアミノ酸）を生成するのに十分なタンパク質分解を提供するが、試料が過剰消化されないような好ましい酵素対ポリペプチドの比を使用して、注意深くタイトレーションすることが必要である。一般的に、所与のプロテイナーゼKロット毎に、機能活性のタイトレーションを実施することが必要である。この例では、タンパク質試料を、1×PBS/1mM EDTA/0.5mM CaCl<sub>2</sub>/0.5% SDS（pH8.0）中で1:10～1:100（w/w）の酵素：タンパク質比にてプロテイナーゼKを用いて、1時間37℃にて消化する。インキュベーション後、PMSEを、終濃度が5mMになるように添加して、さらなる消化を阻害する。

【0573】

プロテイナーゼKの特異的活性は、「化学的基質」ベンゾイルアルギニン-p-ニトロアニリドをプロテイナーゼKと共にインキュベートし、約410nmで吸収を示す黄色p-ニトロアニリン産物の発色を測定することにより、測定することができる。酵素活性は、1単位が毎分1μモルのp-ニトロアニリドの産生と等しい単位で測定し、特異的活性は、酵素活性/mg総タンパク質の単位で測定する。その後、溶液中のタンパク質の総量で酵素活性を除算することにより、特異的活性を算出する。

【0574】

（実施例2）

SP3オンビーズプロテアーゼ消化および標識を使用した試料調製

Hughesら（2014年、Mol Syst Biol、10巻：757頁）により記載されているようなSP3試料調製プロトコールを使用して、タンパク質を抽出し、変性させる。抽出した後、タンパク質ミックス（およびビーズ）を、0.02% SDSで補完された1mM EDTAを有する50mMホウ酸緩衝液（pH8.0）中で1時間37℃で溶解した。タンパク質溶解後、ジスルフィド結合を、終濃度が5mMになるようにDTTを添加し、試料を50℃で10分間インキュベートすることにより還元する。終濃度が10mMになるようにヨードアセトアミドを添加することによりシステインをアルキル化し、室温で20分間、暗所でインキュベートする。反応を、50mMホウ酸緩衝液で2倍に希釈し、Glu-CまたはLys-Cを、1:50（w/w）の最終プロテイナーゼ：タンパク質比で添加する。試料を37℃で一晩（約16時間）インキュベートして、消化を完了した。Hughesら（上記）により記載されているように試料を消化した後、8分間のインキュベーション中、アセトニトリルの終濃度が95%になるように100%アセトニトリルを添加することによりペプチドをビーズに結合させ、アセトニトリルで洗浄する。洗浄した後、5分間のピペット混合ステップにより、10μlの2% DMSO中でビーズからペプチドを溶出する。

【0575】

（実施例3）

記録タグとペプチドとのカップリング

DNA記録タグを、いくつかの方法でペプチドにカップリングする（Aslamら、1998年、Bioconjugation: Protein Coupling Tec

10

20

30

40

50



h n i q u e s f o r t h e B i o m e d i c a l S c i e n c e s , M a c m i l l a n R e f e r e n c e L T D ; H e r m a n s o n G T 、 1 9 9 6 年 、 B i o c o n j u g a t e T e c h n i q u e s 、 A c a d e m i c P r e s s I n c . 、 1 9 9 6 年 を 参 照 さ れ たい ) 。 1 つ の 手 法 で は 、 カ ル ボ ジ イ ミ ド ( c a r b d i i m i d e ) 化 学 を 使 用 し て ペ プ チ ド の C 末 端 に カ ッ プ リ ン グ さ れ る 5 ' ア ミ ン 、 お よ び ク リ ッ ク 化 学 を 使 用 し て ア ジ ド ビ ー ズ に カ ッ プ リ ン グ さ れ る 内 部 歪 み ア ル キ ン D B C O - d T ( G l e n R e s e a r c h 、 V A ) を 用 い て 、 オ リ ゴ ヌ ク レ オ チ ド 記 録 タ グ を 構 築 す る 。 カ ル ボ ジ イ ミ ド カ ッ プ リ ン グ を 完 了 へ と 駆 動 し 、 ペ プ チ ド 間 カ ッ プ リ ン グ を 制 限 す る た め に 、 大 幅 に 過 剰 な モ ル 濃 度 の 記 録 タ グ を 使 用 し て 、 溶 液 中 で 記 録 タ グ を ペ プ チ ド に カ ッ プ リ ン グ す る 。 あ る い は 、 5 ' 歪 み ア ル キ ン ( D B C O - d T ) を 用 い て オ リ ゴ ヌ ク レ オ チ ド を 構 築 し 、 ア ジ ド 誘 導 体 化 ペ プ チ ド に カ ッ プ リ ン グ し ( ペ プ チ ド の C 末 端 に カ ッ プ リ ン グ す る ア ジ ド - P E G - ア ミ ン お よ び カ ル ボ ジ イ ミ ド に よ り ) 、 そ の 後 ア ル デ ヒ ド 反 応 性 H y N i c ヒ ド ラ ジ ン ビ ー ズ に カ ッ プ リ ン グ す る 。 こ の 目 的 の た め に は 、 記 録 タ グ オ リ ゴ ヌ ク レ オ チ ド を 、 内 部 ア ル デ ヒ ド ホ ル ミ ル イ ン ド ー ル ( T r i l i n k ) 基 で 容 易 に 標 識 す る こ と が で き る 。 あ る い は 、 C 末 端 ア ミ ン に カ ッ プ リ ン グ す る の で は な く 、 そ の 代 わ り に 、 記 録 タ グ を 、 内 部 リ シ ン 残 基 に カ ッ プ リ ン グ し て も よ い ( 好 ま し く は 、 L y s - C 消 化 後 、 ま た は そ の 代 わ り に G l u - C 消 化 後 ) 。 1 つ の 手 法 で は 、 こ れ は 、 N H S - ア ジ ド ( ま た は 、 N H S - P E G - ア ジ ド ) 基 を 用 い て リ シ ン ア ミ ン を 活 性 化 し 、 そ の 後 5 ' ア ミ ン 標 識 記 録 タ グ に カ ッ プ リ ン グ す る こ と に よ り 達 成 す る こ と が で き る 。 別 の 手 法 で は 、 5 ' ア ミ ン 標 識 記 録 タ グ を 、 D S S な ど の 過 剰 な N H S ホ モ 二 機 能 性 架 橋 試 薬 と 反 応 さ せ て 、 5 ' N H S 活 性 化 記 録 タ グ を 創 出 す る こ と が で き る 。 こ の 5 ' N H S 活 性 化 記 録 タ グ は 、 ペ プ チ ド の リ シ ン 残 基 の - ア ミ ノ 基 に 直 接 カ ッ プ リ ン グ す る こ と が で き る 。

【 0 5 7 6 】

( 実 施 例 4 )

ペ プ チ ド の ア ミ ノ 酸 の 部 位 特 異 的 標 識

活 性 化 D N A タ グ で 直 接 的 に ( ヘ テ ロ 二 機 能 性 ア ミ ノ 酸 部 位 特 異 的 試 薬 に よ る 活 性 化 を 使 用 し て ) 、 ま た は D N A タ グ の 同 種 ク リ ッ ク 部 分 に 付 着 さ せ る た め に 後 に 使 用 さ れ る ク リ ッ ク 部 分 を 有 す る ア ミ ノ 酸 を 部 位 特 異 的 に 標 識 す る ク リ ッ ク 化 学 ヘ テ ロ 二 機 能 性 試 薬 に よ り 間 接 的 に 修 飾 す る こ と が で き る ( L u n d b l a d 、 2 0 1 4 年 ) タ ン パ ク 質 ま た は ペ プ チ ド の ア ミ ノ 酸 の 異 な る 5 つ の 例 。 典 型 的 な タ ン パ ク 質 入 力 は 、 0 . 1 % R a p i G e s t ( 商 標 ) S F 界 面 活 性 剤 お よ び 5 m M T C E P を 含 む 5 0 μ l の 適 切 な 水 性 緩 衝 液 中 に 1 μ g の タ ン パ ク 質 を 含 む 。 R a p i G e s t ( 商 標 ) S D は 、 標 識 ま た は 消 化 を 向 上 さ せ る た め に タ ン パ ク 質 を ポ リ ペ プ チ ド へ と 変 性 さ せ る た め の 酸 分 解 性 界 面 活 性 剤 と し て 有 用 で あ る 。 以 下 の ア ミ ノ 酸 標 識 戦 略 を 使 用 す る こ と が で き る : マ レ イ ミ ド 化 学 を 使 用 し た シ ス テ イ ン - - - 2 0 0 μ M の ス ル ホ - S M C C 活 性 化 D N A タ グ を 使 用 し て 、 シ ス テ イ ン を 、 1 0 0 m M M E S 緩 衝 液 ( p H 6 . 5 ) + 1 % T X - 1 0 0 中 で 1 時 間 、 部 位 特 異 的 に 標 識 す る ; N H S 化 学 を 使 用 し た リ シ ン - - - 2 0 0 μ M の D S S ま た は B S <sup>3</sup> 活 性 化 D N A タ グ を 使 用 し て 、 溶 液 相 タ ン パ ク 質 ま た は ビ ー ズ 結 合 ペ プ チ ド の リ シ ン を 、 室 温 に て 1 時 間 、 ホ ウ 酸 緩 衝 液 ( 5 0 m M 、 p H 8 . 5 ) + 1 % T X - 1 0 0 中 で 部 位 特 異 的 に 標 識 す る ; チ ロ シ ン は 、 4 - フ ェ ニ ル - 3 H - 1 , 2 , 4 - ト リ ア ゾ リ ン - 3 , 5 ( 4 H ) - ジ オ ン ( P T A D ) で 修 飾 す る ; ま た は ジ ア ゾ ニ ウ ム 化 学 - - - ジ ア ゾ ニ ウ ム 化 学 の 場 合 、 D N A タ グ を 、 E D C お よ び 4 - カ ル ボ キ シ ル ベ ン ゼ ン ジ ア ゾ ニ ウ ム テ ト ラ フ ル オ ロ ボ レ ー ト ( A i k o n I n t e r n a t i o n a l 、 C h i n a ) で 活 性 化 す る 。 タ ン パ ク 質 ま た は ビ ー ズ 結 合 ペ プ チ ド を 、 ホ ウ 酸 緩 衝 液 ( 5 0 m M 、 p H 8 . 5 ) + 1 % T X - 1 0 0 中 で 、 2 0 0 μ M の ジ ア ゾ ニ ウ ム 誘 導 体 化 D N A タ グ と 共 に 1 時 間 氷 上 に て イ ン キ ュ ベ ー ト す る こ と に よ り 、 チ ロ シ ン と の ジ ア ゾ 連 結 を 創 出 す る ( N g u y e n 、 C a o ら 、 2 0 1 5 年 ) 。 E D C 化 学 を 使 用 し て ア ス パ ラ テ ー ト / グ ル タ メ ー ト を 修 飾 す る - - - ア ミ ン 標 識 D N A タ グ を 、 p H 6 . 5 の M E S 中 で ビ ー ズ 結 合 ペ プ チ ド お よ び 1 0 0 m M E D C / 5 0 m M イ ミ ダ ゴ ー ル と 共 に 、 室 温 で 1 時 間 に わ た っ て イ ン キ ュ ベ ー ト す る ( B a s l e ら 、 2 0 1 0 年 、 C h e m . B i o l . 、 1 7 巻 : 2 1 3 ~

227頁)。標識の後、過剰な活性化DNAタグを、C4レジンZip Tips (Millipore)からのタンパク結合溶出を使用して除去する。溶出したタンパク質を、1×PBS緩衝液で50 µlにする。

【0577】

(実施例5)

歪みアルキン記録タグ標識ペプチドのアジド活性化ビーズへの固定化

市販のアミンDynabeads (登録商標) M-270を、アジドPEG NHSエステルヘテロ二機能性リンカー (JenKem Technology, Tx)と反応させることにより、アジド誘導体化Dynabeads (登録商標) M-270ビーズを生成する。さらに、メトキシまたはヒドロキシルPEG NHSエステルと適切な比で混合することにより、アジドの表面密度をタイトレーションすることができる。所与のペプチド試料毎に、1~2 mgのアジド誘導体化Dynabeads (登録商標) M-270ビーズ (約 $1.3 \times 10^8$ 個のビーズ)を、100 µlのホウ酸緩衝液 (50 mMホウ酸ナトリウム、pH 8.5)で希釈し、1 ngの記録タグペプチドを添加し、23~37 で1時間インキュベートする。200 µlのホウ酸緩衝液で3回洗浄する。

【0578】

(実施例6)

ホルミルインドール反応性HyNicビーズの創出

アミンビーズのHyNic誘導体化により、ホルミルインドール反応性ビーズを創出する。20 mgのDynabeads (登録商標) M-270アミンビーズ (2.8 µm)のアリコート、200 µlのホウ酸緩衝液に懸濁する。短時間の超音波処理後、1~2 mgのスルホ-S-HyNic (スクシンイミジル6-ヒドラジノニコチネートアセトンヒドラゾン、S ANH) (カタログ# S-1002、Solulink, San Diego)を添加し、反応混合物を室温で1時間にわたって振とうする。その後、ビーズを、ホウ酸緩衝液で2回およびクエン酸緩衝液 (200 mMクエン酸ナトリウム)で1回洗浄する。ビーズを、終濃度が10 mg/mlになるようにクエン酸緩衝液に懸濁する。

【0579】

(実施例7)

記録タグホルミルインドール (formylindole) 標識ペプチドの活性化ビーズへの固定化

1~2 mgのHyNic活性化Dynabeads (登録商標) M-270ビーズのアリコート (約 $1.3 \times 10^8$ 個ビーズ)を、50 mMアニリンで補完された100 µlのクエン酸緩衝液で希釈し、約1 ngの記録タグペプチドコンジュゲートを添加し、37 で1時間にわたってインキュベートする。ビーズを、200 µlのクエン酸緩衝液で3回洗浄し、100 µlのホウ酸緩衝液に再懸濁した。

【0580】

(実施例8)

オリゴヌクレオチドモデル系 - コーディングタグの識別情報をサイクル様式で記録タグへと移行させることによる結合性物質履歴の記録

核酸コーディングタグおよび記録タグの場合、標準的核酸酵素学を使用したライゲーションまたはプライマー伸長により、結合されている結合性物質のコーディングタグから近位の記録タグへと、情報を移行させることができる。これは、結合性物質標的を表す5'部分および記録タグを表す3'部分を有するオリゴヌクレオチドで構成される単純なモデル系で実証することができる。オリゴヌクレオチドの内部部位を、dT-アルキン修飾 (DBCO-dT, Glen Research)によるクリック化学を使用して固定化することができる。図24Aに示されている例では、固定化されたオリゴヌクレオチド (AB標的)は、同種オリゴヌクレオチド「結合性物質」であるAオリゴおよびBオリゴが結合することができる、AおよびBと標識されている2つの標的結合領域を含む。AオリゴおよびBオリゴヌクレオチドは、共通スパーサー (Sp)を介して記録タグと相互作用して、プライマー伸長 (またはライゲーション)を開始するコーディングタグ (配列および長さ

10

20

30

40

50

が異なる)に連結されている。S pの長さは、結合性物質結合中の非特異的相互作用を最小限に抑えるために、短く(例えば、6~9塩基)しておくべきである。この特定の例では、コーディングタグの長さは、「A」オリゴ結合事象(10塩基エンコーダー配列)と「B」オリゴ結合事象(20塩基エンコーダー配列)がゲル解析で容易に区別されるように設計されている。

#### 【0581】

PAGEゲルを単に解析することにより、AまたはBコーディングタグ移行の効率を測定することが可能であり、実験パラメーターの容易な最適化が可能である。AB標的配列に加えて、CおよびDが、AおよびBと相互作用しない異なるハイブリダイゼーション配列であることを除いて、同様のオリゴヌクレオチドCD標的配列が用いられる(図24Bを参照されたい)。さらに、CおよびDは、それぞれ30塩基DNAコードおよび40塩基DNAコードを含む、異なる配列および長さのコーディングタグを含む。第2の標的配列CDの目的は、ABおよびCD標的分子間の交差相互作用を評価することである。特定のハイブリダイゼーションを考慮すると、AB標的に結合されたオリゴに接続されているAまたはBコーディングタグ間に分子間交差が生じない限り、CD標的の伸長記録タグが、AまたはBコーディングタグ情報を含むことはない。同様に、AB標的の伸長記録タグは、CまたはDコーディングタグ情報を含むはずがない。ABおよびCD標的が物理的近傍に接近している状況では(つまり<50nm)、掛け合い応答が起こる可能性が高い。したがって、表面の標的巨大分子を適切に離間させることが重要である。

#### 【0582】

このオリゴヌクレオチドモデル系は、結合性物質履歴の記録能力の十分な特徴付けを可能にする。図25は、プライマー伸長ではなくライゲーションによる情報移行を示す。まずゲルで最適化した後、種々の結合およびアッセイプロトコルを実施し、配列決定により評価する。一意の分子識別子(UMI)配列は、計数のために使用され、単一の巨大分子に由来するリードの識別を可能にし、元の試料中の総合的で全体的な巨大分子複雑性の尺度を提供する。例示的な履歴結合プロトコルとしては、A-B-C-B-A、A-B-A-A-B-A、A-B-C-D-A-Cなどが挙げられる：得られる最終産物は、それぞれ、UMI-Sp-A-Sp-B-Sp-B-Sp-A-Sp+UMI-Sp-C-Sp;UMI-Sp-A-Sp-B-Sp-A-Sp-A-Sp-B-Sp-A;UMI-A-Sp-B-Sp-A+UMI-Sp-C-Sp-D-Sp-C-Spと読み取られるはずである。この解析の結果は、さらなる最適化を可能にする。

#### 【0583】

##### (実施例9)

オリゴヌクレオチド-ペプチドモデル系-コーディングタグの識別情報をサイクル様式で記録タグへと移行させることによる結合性物質履歴の記録

オリゴヌクレオチドモデル系を検証した後、例示的な標的オリゴヌクレオチド配列の5'末端にペプチドエピトープタグをコンジュゲートすることにより、オリゴヌクレオチド系からペプチドモデル系を構築する(図26Aおよび26B)。例示的なペプチドエピトープタグとしては以下のものが挙げられる：FLAG(DYKDDDDK)(配列番号171)、V5(GKPIPNPLGLDST)(配列番号172)、c-My c(EQKLISEEDL)(配列番号173)、HA(YPYDVPDYA)(配列番号174)、V5(GKPIPNPLGLDST)(配列番号175)、Streptag II(NWSHPQFEK)(配列番号176)など。ペプチドエピトープタグをオリゴヌクレオチドにカップリングするための、任意選択のCys-Ser-Glyリンカーが含まれていてもよい。実施例7のABオリゴヌクレオチド鋳型を、A\_\_オリゴヌクレオチド-cMy cペプチド構築物に取り替え、実施例7のCDオリゴヌクレオチド鋳型を、C\_\_オリゴヌクレオチド-HAペプチド構築物に取り替える(図26を参照されたい)。また、A\_\_オリゴヌクレオチド-cMy cペプチド構築物は、CSGリンカーおよびN末端ホスホチロシンを含む。同様に、同種ペプチド結合性物質であるcMy c抗体およびHA抗体を、それぞれBオリゴヌクレオチドコーディングタグおよびDオリゴヌクレオチドコード

ィングタグでタグ化する。ホスホチロシン特異的抗体は、別の「E」コーディングタグでタグ化する。このように、ペプチドモデル系は、オリゴヌクレオチド系と対応しており、オリゴ結合および抗体結合の両方が、このモデル系で試験される。

#### 【0584】

抗c-myc抗体(2G8D5、マウスモノクローナル、GenScript)、抗HA抗体(5E11D8、マウスモノクローナル、GenScript)、strept-タグII抗体(5A9F9、マウスモノクローナル、GenScript)、または抗FLAG抗体(5AE85、マウスモノクローナル、GenScript)を使用した固定化DNAペプチド構築物の抗体染色を、0.1~1μg/mlの1×PBST(PBS+0.1%Tween20)を使用して実施する。インキュベーションを、典型的には室温で30分間実施する。また、1×PBST中1%のPVPを使用した標準的プレブロッキングおよび染色後洗浄を実施する。抗体脱染色は、高濃度塩(1M NaCl)および低pH(グリシン、pH2.5)または高pH(トリエチルアミン、pH11.5)のいずれかで洗浄することにより効果的に達成される。

#### 【0585】

標的オリゴヌクレオチドは、アジドビーズに付着させるための内部アルキン標識を含み、5'末端は、Williamsら(2010年、Curr Protoc Nucleic Acid Chem. 第4章:4.41項)に記載されているように、ペプチドのC末端システインにSMCC媒介性付着させるためのアミノ基を含む。あるいは、標準的カルボジイミドカップリングを、オリゴヌクレオチドおよびペプチドのコンジュゲーション反応に使用する(Luら、2010年、Bioconjug. Chem., 21巻:187~202頁)。この場合、過剰なオリゴを使用して、カルボジイミド反応を駆動し、ペプチド間カップリングを最小限に抑える。コンジュゲーション後、PAGEゲルから切除し、溶出することにより、最終産物を精製する。

#### 【0586】

##### (実施例10)

DNA/PNAコーディングタグ相補体を記録タグにライゲーションさせることによるコーディングタグ移行

コーディングタグを、ライゲーションにより直接的または間接的のいずれかで記録タグに移行させて、伸長記録タグを生成する。一実施例では、アニーリングされたコーディングタグの相補体を、記録タグにライゲーションさせる(図25)。このコーディングタグ相補体は、核酸(DNAまたはRNA)であってもよく、ペプチド核酸(PNA)であってもよく、または成長中の記録タグにライゲーションすることが可能ないくつかの他のコードディング分子であってもよい。ライゲーションは、DNAおよびRNAの場合、標準的なATP依存性およびNADH依存性リガーゼを使用して酵素的であってもよく、またはライゲーションは、DNA/RNAおよび特にペプチド核酸PNAの両方の場合、化学媒介性であってもよい。

#### 【0587】

DNAの酵素的ライゲーションの場合、アニーリングされたコーディングタグは、記録タグの3'ヒドロキシルとライゲーションするために、5'リン酸塩を必要とする。例示的な酵素的ライゲーション条件は、以下の通りである(Gunderson、Huangら、1998年)。標準的T4 DNAライゲーション反応は、50mM Tris-HCl(pH7.8)、10mM MgCl<sub>2</sub>、10mM DTT、1mM ATP、50μg/ml BSA、100mM NaCl、0.1%TX-100、および2.0U/μl T4 DNAリガーゼ(New England Biolabs)を含む。E. coli DNAリガーゼ反応は、40mM Tris-HCl(pH8.0)、10mM MgCl<sub>2</sub>、5mM DTT、0.5mM NADH、50μg/ml BSA、0.1%TX-100、および0.025U/μl E. coli DNAリガーゼ(Amersham)を含む。Taq DNAライゲーション反応は、20mM Tris-HCl(pH7.6)、25mM酢酸カリウム、10mM酢酸マグネシウム、10mM DTT、1mM NAD

H、50  $\mu$ g/ml BSA、0.1% Triton X-100、10% PEG、100 mM NaCl、および1.0 U/ $\mu$ l Taq DNAリガーゼ(New England Biolabs)を含む。T4およびE. coli DNAリガーゼ反応は、室温で1時間にわたって実施し、Taq DNAリガーゼ反応は、40℃で1時間にわたって実施する。

#### 【0588】

DNA/PNAコーディングタグ移行の場合、鋳型化DNA/PNAの化学的ライゲーションのいくつかの方法を用いることができる。そうした方法としては、標準的な化学的ライゲーションおよびクリック化学手法が挙げられる。鋳型DNAライゲーションの例示的な化学的ライゲーション条件は、以下の通りである(Gunderson、Huangら、1998年)：鋳型3'リン酸塩リポータータグと5'リン酸塩コーディングタグとのライゲーションは、50 mM 2-[N-モルホリノ]エタンスルホン酸(MES)(KOHでpH6.0)、10 mM MgCl<sub>2</sub>、0.001% SDS、新しく調製した200 mM EDC、50 mM イミダゾール(HClでpH6.0)または50 mM HOBt(HClでpH6.0)、および3.0~4.0 M TMAcI(Sigma)を含む反応中で、室温にて1時間以内に生じる。

#### 【0589】

PNAの鋳型依存性ライゲーションの例示的な条件としては、NH<sub>2</sub>-PNA-CHOポリマー(例えば、コーディングタグ相補体および伸長記録タグ)のライゲーションが挙げられ、Brudnoら(Brudno、Birnbaumら、2010年)により記載されている。PNAは、5'アミン等価物および3'アルデヒド等価物を有しており、化学的ライゲーションにより2つの部分がカップリングされてシッフ塩基が創出され、シッフ塩基は、その後シアノ水素化ホウ素ナトリウムで還元される。このカップリングの典型的な反応条件は、以下の通りである：100 mM TAPS(pH8.5)、80 mM NaCl、および80 mM シアノ水素化ホウ素ナトリウム、室温で60分間。5'アミノ末端1,2-アミノチオール修飾および3'C末端チオエステル修飾を含む機能化PNAを使用する未変性化学的ライゲーションの例示的な条件は、Roloffら(2014年、Methods Mol. Biol.、1050巻：131~141頁)により記載されている。また、他のN-およびC-末端PNA部分を、ライゲーションに使用することができる。別の例は、クリック化学を使用したPNAの化学的ライゲーションを含む。Pengら(2010年、European J. Org. Chem.、2010年：4194~4197頁)の手法を使用して、PNAを、5'アジドおよび3'アルキンで誘導体化し、クリック化学を使用してライゲーションすることができる。「クリック」化学ライゲーションの例示的な反応条件は、10 mM リン酸カリウム緩衝液、100 mM KCl、5 mM THPTA(tris-ヒドロキシプロピルトリゾリル(trizoly)アミン)、0.5 mM CuSO<sub>4</sub>、および2.5 mM Na-アスコルビン酸塩を含む100  $\mu$ lの反応ミックス中での、1~2 mg ビーズとの鋳型化PNA-PNAである。化学的ライゲーション反応は、室温で1時間インキュベートされる。PNAライゲーションの他の例示的な方法は、Sakuraiら(Sakurai、Snyderら、2005年)により記載されている。

#### 【0590】

##### (実施例11)

##### DNAへのPNA変換

PNA鋳型にアニーリングされたDNAオリゴヌクレオチドのクリック化学媒介性重合を使用して、PNAをDNAへと変換する。DNAオリゴは、DNAポリメラーゼにより複製することが可能なヌクレオチド間トリアゾール連結を創出するために、反応性5'アジドおよび3'アルキンを含む(El-Sagheerら、2011年、Proc. Natl. Acad. Sci. USA 108巻：11338~11343頁)。PNAのあらゆる考え得るコーディングタグに相補的なDNAオリゴの完全なセット(10 nM、1xハイブリダイゼーション緩衝液中：10 mM Na-ホウ酸塩(pH8.5)、0.

10

20

30

40

50

2 M NaCl) を、固相に結合されている PNA 分子と共に 30 分間インキュベートする (23 ~ 50)。アニーリングの後、固相に結合されている PNA-DNA 構築物を、アスコルビン酸ナトリウム緩衝液 (10 mM アスコルビン酸ナトリウム、200 mM NaCl) で 1 回洗浄する。「クリック化学」反応条件は、以下の通りである：ビーズ上の PNA-DNA を、新たなアスコルビン酸ナトリウム緩衝液中でインキュベートし、10 mM THPTA + 2 mM CuSO<sub>4</sub> のミックスと 1 : 1 で組み合わせて、室温で 1 時間インキュベートする。その後、ビーズを、ハイブリダイゼーション緩衝液で 1 回、および PCR 緩衝液で 2 回洗浄する。化学的ライゲーションの後、得られたライゲーション DNA 産物を、El-Sagheer ら (2011 年、Proc. Natl. Acad. Sci. USA 108 巻 : 11338 ~ 11343 頁) により記載されているような条件下で PCR により増幅する。

10

## 【0591】

## (実施例 12)

核酸記録およびコーディングタグと適合性の穏やかな N 末端エドマン分解

N 末端エドマン分解と DNA コーディングとの適合性は、この手法が、ペプチド配列決定で機能することを可能にする。無水 TFA が用いられる N 末端エドマン分解の標準的条件では、DNA は破壊される。しかしながら、この効果は、より穏やかな切断条件を開発し、より高い酸耐性を有する修飾 DNA を開発することにより緩和される。N 末端エドマン分解の穏やかな条件を、フェニルチオカルバモイル (PTC) - ペプチドの切断を最適化すること、および切断条件下で DNA/PNA コード付きライブラリーの安定性を測定することの組合せを使用して開発する。さらに、天然 DNA は、低 pH での脱プリンを低減する 7 - デアザプリンなどの塩基修飾、および脱ピリミジン化を低減する 5' メチル修飾シトシンを使用することにより、酸加水分解に対して安定化させることができる (Schneider および Chait、1995 年、Nucleic Acids Res.、23 巻 : 1570 ~ 1575 頁)。チミンが酸断片化に対して最も安定した塩基であることを考慮すると、T 豊富なコーディングタグも有用であり得る。穏やかな N 末端エドマン分解の条件は、無水 TFA 切断の代わりに、Barrett ら (その全体が参照により組み込まれる、1985 年、Tetrahedron Lett.、26 巻 : 4375 ~ 4378 頁) により記載されているように、アセトニトリル中のトリエチルアミンアセートを 60 で使用する穏やかな 10 分間の塩基切断を使用することである。こうした穏やかな条件は、ほとんどのタイプの DNA 記録およびコーディングタグと適合性である。代替選択肢として、PNA は完全に酸安定性であるため、PNA をコーディングタグに使用する (Ray および Norden、2000 年、FASEB J.、14 巻 : 1041 ~ 1060 頁)。

20

30

## 【0592】

NTAA 結合物質の同一性をコードするために DNA コーディングタグ/記録タグを使用することと、穏やかな N 末端エドマン分解反応を実施することとの適合性を、以下のアッセイを使用して実証する。抗ホスホチロシンおよび抗 cMy c 抗体の両方を使用して、モデルペプチドを読み取る。単一のエドマン分解ステップを使用した、C - My c および N 末端ホスホチロシン検出、コーディングタグ書き込み、および N 末端ホスホチロシンの除去。このステップの後、ペプチドを、抗ホスホチロシンおよび抗 cMy c 抗体で再び染色する。N 末端分解に対する記録タグの安定性を、qPCR により評価する。ホスホチロシンの効果的な除去は、配列決定、qPCR、またはゲル電気泳動により解析して、最終記録タグ配列に E - オリゴヌクレオチドコーディングタグ情報が存在しないことにより示される。

40

## 【0593】

## (実施例 13)

コンパートメントタグ付きビーズの調製。

## 【0594】

コンパートメントタグ付きビーズを調製するには、ホスホラミダイト合成またはスプリ

50

ット・アンド・プールライゲーションのいずれかが使用されるスプリット・アンド・プール合成手法を使用して、ビーズに固定化されているオリゴヌクレオチドにバーコードを組み込む。コンパートメントタグは、コンパートメントタグがそれに接合される各ペプチドまたはタンパク質分子を一義的に標識するための一意の分子識別子 (UMI) をさらに含む。例示的なコンパートメントタグ配列は、以下の通りである: 5' - NH<sub>2</sub> - GCGCAATCAG - XXXXXXXXXX - NNNNN - TGCAAGGAT - 3' (配列番号 177)。XXXXXXXXXXXX (配列番号 178) バーコード配列は、スプリット・プールオンビーズ合成により生成されるビーズ毎の核酸塩基配列の固定集団であり、この固定配列は、ビーズ毎に異なる。NNNNN (配列番号 179) 配列は、その後それに接合されるペプチド分子の一意的分子識別子 (UMI) としての役目を果たすように、ビーズ内で無作為化される。バーコード配列は、Macoskoら (その全体が参照により組み込まれる、2015年、Cell、161巻: 1202~1214頁) により記載されているようなスプリット・アンド・プール手法を使用して、ビーズ上で合成することができる。UMI配列は、縮重塩基混合物 (各カップリングステップに存在する4つのホスホラミダイト塩基すべての混合物) を使用して、オリゴヌクレオチドを合成することにより創出することができる。5' - NH<sub>2</sub> を、スクシンイミジル4 - (N - マレイミドメチル) シクロヘキサン - 1 - カルボキシレート (SMCC) およびN末端からC末端への配列「CGGSSGSNHV」 (配列番号 180) を有するシステイン含有プテラーゼIペプチド基質で活性化し、Williamsら (2010年、Curr Protoc Nucleic Acid Chem、第4章: 4.41項) により記載されているような修飾プロトコルを使用して、SMCC活性化コンパートメントタグ付きビーズにカップリングする。すなわち、200 µl の磁気ビーズ (10 mg/ml) を、1.5 ml の Eppendorf チューブに入れる。1 ml のカップリング緩衝液 (5 mM EDTA、0.01% Tween 20、pH 7.4 を有する 100 mM KH<sub>2</sub>PO<sub>4</sub> 緩衝液、pH 7.2) をチューブに添加し、短時間ボルテックスする。新たに調製した 40 µl のスルホ - SMCC (DMSO 中 50 mg/ml、ThermoFisher) を、磁気ビーズに添加し、混合する。反応を、室温にて1時間、ロータリーミキサーでインキュベートする。インキュベーション後、磁石でビーズを上清から分離し、500 µl のカップリング緩衝液で3回洗浄する。ビーズを、400 µl のカップリング緩衝液に再懸濁する。1 ml の CGGSSGSNHV (配列番号 180) ペプチドを磁気ビーズに添加する (TCEP還元 (5 mM) および氷冷アセトン沈殿後、カップリング緩衝液中 1 mg/mL)。反応を、室温にて2時間、ロータリーミキサーでインキュベートする。反応を、カップリング緩衝液で1回洗浄する。400 µl のクエンチング緩衝液 (10 mg/mL のメルカプトコハク酸、pH 7.4 を有する 100 mM KH<sub>2</sub>PO<sub>4</sub> 緩衝液、pH 7.2) を反応混合物に添加し、ロータリーミキサーで2時間インキュベートする。反応混合物を、カップリング緩衝液で3回洗浄する。得られたビーズを、保管緩衝液 (0.02% NaN<sub>3</sub>、0.01% Tween 20、pH 7.4 を含む 10 mM KH<sub>2</sub>PO<sub>4</sub> 緩衝液、pH 7.2) に再懸濁し、4 で保管する。

【0595】

(実施例 14)

封入ビーズおよびタンパク質の生成

コンパートメントタグ付きビーズおよびタンパク質を、エンドプロテアーゼ AspN (Endo AspN) などの亜鉛メタロ - エンドペプチダーゼ、任意選択の光ケージ化 Znキレート剤 (例えば、Zinc Cleav I)、および遺伝子操作された耐熱性プテラーゼIホモログ (Bandara、Kennedyら、2009年、Bandara、Walshら、2011年、Cao、Nguyenら、2015年) と組み合わせる。実施例 12 のコンパートメントタグ付きビーズをタンパク質と混合し、T字路型マイクロ流体または流動フォーカスデバイス (図 21 を参照されたい) で乳化する。二水流構成では、一方の流動中のタンパク質および Zn<sup>2+</sup> を、他方の流動からのメタロ - エンドペプチダーゼと組み合わせ、液滴形成時に直ちに消化を開始させることができる。一流動構成

では、すべての試薬を予め混合し、一緒に乳化する。これには、任意選択の光ケージ化 Zn キレート剤（例えば、Zinc Cleav I）を使用して、液滴形成後に、UV 光への曝露によりタンパク質消化を開始させることが必要である。濃度および流動条件は、1 液滴当たりのビーズが平均で 1 つ未満になるように調整する。最適化された実験では、 $10^8$  個のフェムト液滴を、液滴の約 10 % がビーズを含有する含有率で製作することができる（Shimら、2013 年、ACS Nano、7 巻：5955～5964 頁）。一流動手法では、液滴を形成した後、エマルジョンを UV - 365 nm の光に曝露して光ケージ化  $Zn^{2+}$  を放出させ、Endo AspN プロテアーゼを活性化することにより、プロテアーゼを活性化する。エマルジョンを、37 で 1 時間インキュベートして、タンパク質をペプチドへと消化する。消化した後、エマルジョンを 80 で 15 分間加熱することにより、Endo AspN を不活化する。二流動調合では、2 つの流動の組み合わせ中に、 $Zn^{2+}$  を液滴内に導入する。この場合、Endo AspN は、キレート剤が UV 光への曝露時に活性化される光活性化  $Zn^{2+}$  ケージ分子を使用することにより、または 2 - アルキルマロン酸もしくは EDTA - MO などの両親媒性  $Zn^{2+}$  キレート作用剤を、油相に添加することにより不活化することができる。両親媒性 EDTA 分子の例としては、EDTA - MO、EDTA - BO、EDTA - BP、DPTA - MO、DPTA - BO、DPTA - BP など（Ojha、Singhら 2010 年、Moghaddam、de Campoら 2012 年）が挙げられる。また、エマルジョン油に両親媒性の酸または塩基を添加することにより液滴の pH を変更することを含む、他のモダリティを使用して、液滴内部の反応を制御することができる。例えば、液滴 pH は、水 / 油に可溶性である酢酸を使用して低下させることができる。フルオロ - エマルジョンへの酢酸の添加は、酢酸分子が両親媒性の性質を持つため、液滴コンパートメント内の pH の低下に結びつく（Mashaghi および van Oijen、2015 年、Sci Rep、5 巻：11837 頁）。同様に、塩基であるプロピルアミンを添加すると、液滴内部はアルカリ化される。同様の手法を、油 / 水に可溶性の酸化還元試薬、還元剤、キレート剤、および触媒などの他のタイプの両親媒性分子に使用することができる。

#### 【0596】

コンパートメント化されたタンパク質をペプチドへと消化した後、ブテラーゼ I または化学的ライゲーション（例えば、アルデヒド - アミノなど）を使用して、ペプチドを、ビーズ上のコンパートメントタグにライゲーションする（オリゴヌクレオチドペプチドバーコードキメラ）（図 16 および図 22 A を参照されたい）。任意選択の手法では、オリゴ - チオデブシペプチド「化学的基質」を用いて、ブテラーゼ I ライゲーションを不可逆的にする（Nguyen、Caoら、2015 年）。ライゲーションの後、エマルジョンを「クラック」し、コンパートメントタグ付きペプチド構築物が固定化されているビーズをバルクで収集するか、またはコンパートメント付きペプチドをビーズから切断し、バルクで収集する。コンパートメントタグ付きペプチドが固定化されているビーズが記録タグを含む場合、これらビーズは、本明細書に記載されている核酸コーディングに基づくペプチド解析法に直接使用することができる。対照的に、コンパートメントタグ付きペプチドをビーズ基材から切断した場合、コンパートメントタグ付きペプチドを、コンパートメントタグ付きペプチドの C 末端へのコンジュゲーションにより記録タグに付随させ、その後、本明細書に記載のように、コーディングタグ付き結合性物質との結合サイクルおよび配列決定解析を行うために、固体支持体に固定化する。記録タグとコンパートメントタグ付きペプチドとの付随は、三機能性リンカー分子を使用して達成することができる。サイクルシーケンシング解析を行うために、付随する記録タグを有するコンパートメントタグ付きペプチドを固体支持体に固定化した後、コンパートメント情報を、プライマー伸長またはライゲーションを使用して、付随する記録タグへと移行させる（図 22 B を参照されたい）。コンパートメントタグ情報を記録タグに移行させた後、元のペプチド消化で使用されたものと同じ酵素を使用して、コンパートメントタグをペプチドから切断することができる（図 22 B を参照されたい）。これにより、ペプチドの元の N 末端が回復されるため、本明細書に記載のような N 末端分解ペプチド配列決定法が可能になる。

10

20

30

40

50



## 【0597】

## (実施例15)

3プライマー融合エマルジョンPCRにより、アミノ酸特異的コーディングタグで共有結合的に修飾されたペプチドの記録タグを付随させることによるジタグ生成

コンパートメントタグおよび分子UMIで構成される記録タグを有するペプチドを、コーディングタグ部位特異的の化学標識で化学的に修飾する。また、コーディングタグは、修飾ペプチド内の所与のタイプのアミノ酸の数を計数することが可能になるようにUMIを含む。TysonおよびArmor (TysonおよびArmour、2012年)の改変プロトコルを使用して、エマルジョンPCRを、1×PHUSION (商標) GC反応緩衝液 (Thermo Fisher Scientific)、200 μMの各dNTP (New England Biolabs)、1 μMプライマーU1、1 μMプライマーU2 tr、25 nMプライマーSp、14単位のPHUSION (商標) 高フィデリティDNAポリメラーゼ (Thermo Fisher Scientific) を含む100 μlの総水性容積中で調製する。10 μlの水相を、TurnerおよびHurles (2009年、Nat. Protoc.、4巻: 1771~1783頁) により以前に記載されているように、2 mlクライオバイアル中の軽油 (Sigma) に溶解した200 μlの油相 (4.5 % vol / vol) Span 80、0.4 % vol / volのTween 80、および0.05 % Triton X-100に、合計5分間にわたって1000 rpmで撹拌しながら、5~10秒ごとに添加する。得られたエマルジョンの平均液滴サイズは、約5ミクロンだった。T字路および流動フォーカスの使用などのエマルジョンを生成するための他の方法も用いることができる (Brouzes、Medkovaら、2009年)。エマルジョン生成後、100 μlの水/油混合物を0.5 mlのPCRチューブに移し、第1ラウンドの増幅を、以下の条件で実施する: 98 °Cで30秒間; 98 °Cで10秒間、70 °Cで30秒間、および72 °Cで30秒間を40サイクル; その後72 °Cで5分間の伸長。第2ラウンドの増幅反応を、以下の条件で実施する: 98 °Cで30秒間; 98 °Cで10秒間、55 °Cで30秒間、および72 °Cで30秒間を40サイクル; その後4 °Cで維持。PCRの最終サイクル後できるだけ早く、200 μlのヘキサン (Sigma) をPCRチューブに直接添加し、20秒間ボルテックスし、13,000 gで3分間遠心分離することにより、エマルジョンを崩壊させる。

## 【0598】

## (実施例16)

伸長記録タグ、伸長コーディングタグ、またはジタグ構築物の配列決定

記録タグまたはコーディングタグのスペーサー (Sp) またはユニバーサルプライミング部位は、配列の本体に3つの塩基のみ (例えばA、C、およびT) および配列の5'末端に第4の塩基 (例えば、G) を使用して設計することができる。合成による配列決定 (SBS) の場合、これにより、標準的非発光性 (dark) (未標識および非終端化) スクレオチド (dATP、dGTP、およびdTTP) および単一ffC色素標識可逆的ターミネーター (例えば、完全に機能的なシトシン三リン酸塩) のミックスを使用して、スペーサー配列全体にわたって非発光性塩基を迅速に組み込むことが可能になる。このようにすると、関連するエンコーダー配列、一意の分子識別子、コンパートメントタグ、伸長リポータータグの結合サイクル配列、伸長コーディングタグ、またはジタグのみがSBS配列決定され、無関連のスペーサーまたはユニバーサルプライミング配列は「スキップ」される。スペーサーの塩基および配列の5'末端の第4の塩基の同一性は、変更してもよく、上記の同一性は、例示のために提供されているに過ぎない。

## 【0599】

## (実施例17)

タンパク質ライセートの調製。

## 【0600】

種々の試料タイプからタンパク質ライセートを製作するための幅広く様々なプロトコルが、当技術分野で公知である。プロトコルの相違点の多くは、細胞タイプ、およびラ

10

20

30

40

50

イセート中の抽出されたタンパク質が未変性状態で解析されるかまたは変性状態で解析されるかに依存する。NGPAアッセイの場合、天然コンフォメーションのタンパク質または変性タンパク質のいずれも、固体基材に固定化することができる（図32を参照されたい）。さらに、未変性タンパク質を固定化した後、基材の表面に固定化されたタンパク質を変性させてもよい。変性タンパク質を用いる利点は2つある。第1に、多数の抗体試薬は、線形エピトープ（例えば、ウエスタンブロットAb）に結合し、変性タンパク質は、線形エピトープへのより良好な接近を提供する。第2に、NGPAアッセイワークフローは、変性タンパク質を使用すると単純化される。それは、固定化されているタンパク質が既に変性されているため、アルカリ性（例えば、0.1 NaOH）剥離条件を使用して、アニーリングされたコーディングタグを伸長記録タグから剥離することができるためである。これは、結合事象および情報移行後に、アニーリングされているコーディングタグを酵素的に除去することが必要とされる、天然コンフォメーションのタンパク質を含むアッセイを使用して、アニーリングされているコーディングタグを除去する場合と対照的である。

#### 【0601】

未変性タンパク質溶解緩衝液の例としては、50 mM HEPES (pH 7.4)、150 mM NaCl、1% Triton X-100、1.5 mM MgCl<sub>2</sub>、10%グリセロールで構成されるRPPA緩衝液；およびM-PER哺乳動物タンパク質抽出試薬（Thermo Fisher）などの市販の緩衝液が挙げられる。変性溶解緩衝液は、50 mM HEPES (pH 8.)、1% SDSを含む。尿素（1 M ~ 3 M）またはグアニジンHCl（1 ~ 8 M）の添加も、タンパク質試料の変性に使用することができる。溶解緩衝液の上記成分に加えて、一般的には、プロテアーゼおよびホスファターゼ阻害剤も含まれている。プロテアーゼ阻害剤および典型的な濃度の例としては、アプロチニン（aprotinin）（2 μg/ml）、ロイペプチン（5 ~ 10 μg/ml）、ベンズアミジン（15 μg/ml）、ペプスタチンA（1 μg/ml）、PMSF（1 mM）、EDTA（5 mM）、およびEGTA（1 mM）が挙げられる。ホスファターゼ阻害剤の例としては、Naピロリン酸塩（10 mM）、フッ化ナトリウム（5 ~ 100 mM）、およびオルトバナジン酸ナトリウム（1 mM）が挙げられる。追加の添加剤としては、タンパク質試料からDNAを除去するためのDNAase I、およびジスルフィド結合を還元するためのDTTなどの還元剤を挙げることができる。

#### 【0602】

組織培養細胞から調製される未変性タンパク質ライセートプロトコールの一例は、以下の通りである。接着細胞をトリプシン処理し（PBS中0.05%トリプシン-EDTA）、遠心分離（200 gで5分間）により収集し、氷冷PBSで2回洗浄する。プロテアーゼ/ホスファターゼ阻害剤および添加剤（例えば、EDTA非含有完全阻害剤（Roche）およびPhosStop（Roche）で補完された氷冷M-PER哺乳動物抽出試薬（10<sup>7</sup>細胞/100 mm皿または150 cm<sup>2</sup>フラスコ当たり約1 mL）を添加する。得られた細胞懸濁液を、4 にて20分間にわたって回転振とう器でインキュベートし、その後、4 にて20分間、約12,000 rpm（細胞タイプに依存する）で遠心分離して、タンパク質上清を単離する。BCAアッセイを使用してタンパク質を定量化し、PBSに1 mg/mlで再懸濁する。タンパク質ライセートは、直ちに使用してもよく、または液体窒素で瞬間凍結して、-80 で保管してもよい。

#### 【0603】

HughesらのSP3プロトコールに基づく、組織培養細胞から調製される変性タンパク質ライセートプロトコールの一例は、以下の通りである。接着細胞をトリプシン処理し（PBS中0.05%トリプシン-EDTA）、遠心分離（200 gで5分間）により収集し、氷冷PBSで2回洗浄する。プロテアーゼ/ホスファターゼ阻害剤および添加剤（例えば、1x Completeプロテアーゼ阻害剤カクテル（Roche））で補完された氷冷変性溶解緩衝液（10<sup>7</sup>細胞/100 mm皿または150 cm<sup>2</sup>フラスコ当たり約1 mL）を添加する。得られた細胞懸濁液を、95 で5分間インキュベートし、5分間

氷上に置く。ベンゾナーゼヌクレアーゼ (500 U/ml) をライセートに添加し、37 で30分間インキュベートしてDNAおよびRNAを除去する。

【0604】

ライセート100 uL当たり5 μLの200 mM DTTを添加することによりタンパク質を還元し、45 で30分間インキュベートする。タンパク質システイン基のアルキル化 (alkylation) は、ライセート100 uL当たり10 uLの400 mM ヨードアセトアミドを添加することにより達成し、24 で30分間、暗所でインキュベートする。ライセート100 uL当たり10 uLの200 mM DTTを添加することにより、反応をクエンチする。任意選択で、ライセート100 uL当たり2 uLの酸無水物および100 uLの1 M Na<sub>2</sub>CO<sub>3</sub> (pH 8.5) を添加することにより、タンパク質をアシル化する。室温で30分間インキュベートする。「in vivo」でアセチル化されたリシンと、アシル化によるリシン基の「in situ」ブロッキングとの区別が可能になるように、無水酢酸ではなく、吉草酸無水物、安息香酸無水物、およびプロピオン酸無水物が推奨される (Sidoli、Yuanら、2015年)。5 mgのTris (2-アミノエチル) アミン、ポリマー (Sigma) を添加し、室温で30分間インキュベーションすることにより反応をクエンチする。ポリマー樹脂は、ライセートを2000 gで1分間遠心分離して、0.45 μm酢酸セルロースSpin-Xチューブ (Corning) を通過させることにより除去する。BCAアッセイを使用してタンパク質を定量化し、PBSに1 mg/mlで再懸濁する。

【0605】

追加の例では、タンパク質閉じ込め、アルキル化、およびペプチダーゼ消化に、MWCO濾過デバイスが使用される、Erdeらに記載のフィルター支援試料調製 (FASP) プロトコルを使用して標識ペプチドを生成する (Erde、Looら 2014年、FeistおよびHummon 2015年)。

【0606】

(実施例18)

分配タグ付きペプチドの生成。

【0607】

DNAタグ (任意選択の試料バーコードおよび直交性付着部分を有する) を、標準的バイオコンジュゲーション法 (Hermanson、2013年) を使用して、変性ポリペプチドのリシンの - アミノ基を標識するために使用するか、またはその代わりにベンゾフェノンなどの光親和性標識 (PAL) 法を使用して、ポリペプチドに付着させる (Li、Liura、2013年)。ポリペプチドのリシン基をまたは無作為にCH基 (PALにより) をDNAタグで標識し、アシル無水物でのアシル化により未標識基をブロッキングした後、DNAタグ標識アシル化ポリペプチドを、ユニバーサルプライミング配列、コンパートメントバーコード、任意選択のUMI、およびポリペプチドに付着したDNAタグの部分に相補的なプライマー配列を含むDNAオリゴヌクレオチドが付着しているコンパートメントビーズにアニーリングさせる。複数のDNAハイブリダイゼーションタグには協同性があるため、単一のポリペプチド分子は、主に単一のビーズと相互作用し、同じコンパートメントバーコードを、ポリペプチド分子のすべてのDNAタグに書き込むことが可能である。アニーリング後、ポリペプチド結合DNAタグは、アニーリングされたビーズ結合DNA配列のポリメラーゼ伸長反応にプライミングする。このように、コンパートメントバーコードおよび他の機能的エレメントは、結合されているポリペプチドに付着したDNAタグに書き込まれる。このステップの完了時には、複数の記録タグがポリペプチドに付着しており、記録タグは、共通スペーサー配列、バーコード配列 (例えば、試料、画分、コンパートメント、空間など)、任意選択のUMI、および他の機能的エレメントを有する。この標識ポリペプチドは、トリプシン、GluC、プロテイナーゼKなどの標準的エンドプロテアーゼを使用して、ペプチド断片へと消化することができる。注: リシン標識ポリペプチドの消化にトリプシンを使用する場合、ポリペプチドは、Arg残基でのみ切断され、Lys残基では切断されない (Lys残基は標識されているため)。プロ

10

20

30

40

50

テアーゼ消化は、ビーズ上で直接実施してもよく、または標識ポリペプチドをバーコード付きビーズから除去した後で実施してもよい。

【0608】

(実施例19)

モデル系のDNA記録タグ-ペプチドコンジュゲートの調製

【0609】

5'NH<sub>2</sub>基および後にビーズにカップリングするための内部mTetrazine基を有する記録タグオリゴヌクレオチドを合成する(mTet-PEG-N<sub>3</sub>ヘテロ二機能性架橋剤により、アルキン-dTをmTetrazine-dTに変換する)。オリゴヌクレオチドの5'NH<sub>2</sub>を、Williamsら(WilliamsおよびChaput 2010年)により記載されているように、LC-SMCC(ThermoFisher Scientific)などのNHS/マレイミドヘテロ二機能性架橋剤を使用して、ペプチドの反応性システインにカップリングする。特に、20nmolの5'NH<sub>2</sub>標識オリゴヌクレオチドをエタノール沈殿し、シリコーン処理チューブ中の180ulリン酸カップリング緩衝液(0.1Mリン酸カリウム緩衝液、pH7.2)に再懸濁する。5mgのLC-SMCCを、1mLのDMF(5mg/ml)に再懸濁する(アリコートにして-20で保管)。20ulのLC-SMCC(5mg/ml)のアリコートを、180ulの再懸濁したオリゴヌクレオチドに添加し、混合し、室温で1時間インキュベートする。混合物を2回エタノール沈殿する。得られたマレイミド(maleimide)誘導体化オリゴヌクレオチドを、200ulのリン酸カップリング緩衝液に再懸濁する。システイン残基を含むペプチド(>95%純度、脱塩)を、1mg/mlのDMSO(約0.5mM)に再懸濁する。およそ50nmolのペプチド(100ul)を、反応ミックスに添加し、室温で一晩インキュベートする。得られたDNA記録タグ-ペプチドコンジュゲートを、Williamsら(WilliamsおよびChaput 2010年)により記載されているように、未変性PAGEを使用して精製する。コンジュゲートを、シリコーン処理チューブ中の100uM濃度のリン酸カップリング緩衝液に再懸濁する。

【0610】

(実施例20)

DNA-ペプチド固定化用の基材の開発。

【0611】

M-270アミン磁気Dynabeadsを、それぞれアルキンまたはメチルテトラジン標識オリゴ-ペプチドコンジュゲートにカップリングすることが可能なアジド誘導体化ビーズまたはTCO誘導体化ビーズのいずれかに変換することにより、クリック化学固定化に好適な磁気ビーズを創出する(例えば、図29D~29E;図30D~30Eを参照されたい)。すなわち、10mgのM-270ビーズを、500ulのホウ酸緩衝液(100mMホウ酸ナトリウム、pH8.5)で洗浄および再懸濁する。TCO-PEG(12-120)-NHS(Nanocs)およびメチル-PEG(12-120)-NHSの混合物を、1mMでDMSOに再懸濁し、M-270アミンビーズと共に室温で一晩インキュベートする。メチルのTCO-PEGに対する比をタイトレーションして、TCO部分が<100個/um<sup>2</sup>で存在するように、ビーズの最終TCO表面密度を調整する(例えば、図31E、図34を参照されたい)。未反応アミン基を、DMF中の0.1M無水酢酸および0.1M DIEAの混合物を用いて(10mgビーズ毎に500ul)、室温で2時間カップリングする。カップリングし、DMFで3回洗浄した後、ビーズを、10mg/mlでリン酸カップリング緩衝液に再懸濁する。

【0612】

(実施例21)

記録タグ標識ペプチドの基材への固定化。

【0613】

記録タグ標識ペプチドを、記録タグのmTet基および活性化ビーズまたは基材の表面のTCO基を使用してIEDDAクリック化学反応により、基材に固定化する。この反応

は、反応物の入力濃度が低い場合でさえ、迅速および効率的である。さらに、メチルテトラジンを使用することにより、より大きな安定性が結合に付与される (Selvaraj および Fox 2013 年、Knall、Hollaufら 2014 年、Wu および Devraj 2016 年)。200 ng の M-270 TCO ビーズを、100  $\mu$ l のリン酸カップリング緩衝液に再懸濁する。記録タグに mTet 部分を含む 5 pmol の DNA 記録タグ標識ペプチドを、終濃度が約 50 nM になるようにビーズに添加する。反応を、室温にて 1 時間インキュベートする。固定化した後、基材の未反応 TCO 基を、リン酸カップリング緩衝液中の 1 mM メチルテトラジン酸により、室温にて 1 時間クエンチする。

#### 【0614】

(実施例 22)

N 末端アミノ酸 (NTAA) 修飾

化学的 NTAA アセチル化:

ペプチドの NTAA を、有機または水性溶液 (スルホ - NHS - アセテート) 中で、無水酢酸または NHS - アセテートのいずれかを使用してアセチル化する。無水酢酸誘導体化の場合、DMF 中 10 mM の無水酢酸を、ペプチドと共に室温で 30 分間インキュベートする (Halpin、Leeら、2004 年)。あるいは、100 mM の 2 - (N - モルホリノ) エタンスルホネート (MES) 緩衝液 (pH 6.0) 中 50 mM の無水酢酸および 1 M の NaCl を使用して、室温で 30 分間、ペプチドを水溶液中でアセチル化する (Tse、Snyderら、2008 年)。NHS - アセテート誘導体化の場合、スルホ - NHS - アセテートのストック溶液 (DMSO 中 100 mM) を調製し、100 mM リン酸ナトリウム緩衝液 (pH 8.0) または 100 mM ホウ酸緩衝液 (pH 9.4) に、終濃度が 5 ~ 10 mM になるように添加し、室温で 10 ~ 30 分間インキュベートする (Goodnow、2014 年)。

#### 【0615】

酵素的 NTAA アセチル化:

以下の条件を使用して、N - アセチルトランスフェラーゼ (*Sulfolobus solfatarius* に由来する SsArd1) に曝露することにより、ペプチドの NTAA を酵素的にアセチル化する。ペプチドを、2  $\mu$ M の SsArd1 と共に、NAT 緩衝液 (20 mM Tris - HCl、pH 8.0、100 mM NaCl、1 mM EDTA、1 mM アセチル - CoA) 中で、10 分間 60 °C にてインキュベートする (Chang および Hsu、2015 年)。

#### 【0616】

化学的 NTAA アミジン化 (グアニジン化):

ペプチドを、10 mM N,N - ビス (tert - ブトキシカルボニル) チオ尿素、20 mM トリメチルアミン、および 12 mM 向山試薬 (2 - クロロ - 1 - メチルピリジニウムヨード) の DMF 溶液で、30 分間室温にてインキュベートする。あるいは、ペプチドを、10 mM 1H - ピラゾール - 1 - カルボキサミジン塩酸塩、10 mM DIEA の DMF 溶液で、30 分間室温にてインキュベートする。標準的脱プロテクト法を使用して、保護基を除去する。あるいは、ペプチドを、PBS 緩衝液 (pH 8.0) または 100 mM ホウ酸緩衝液 (pH 8.0) 中 10 mM S - メチルイソチオ尿素で、30 分間 100 °C にてインキュベートする (Tse、Snyderら、2008 年)。

#### 【0617】

PITC 標識:

ペプチドを、イオン性液体 [Bmim][BF<sub>4</sub>] 中 5% (vol/vol) の PITC で、5 分間室温にてインキュベートする。伸長 DNA 記録タグに存在するヌクレオチド塩基の環外アミンの異所標識を最小限に抑えつつ、NTAA が定量的に PITC で標識されるように、反応時間を最適化する。

#### 【0618】

DNFB 標識:

2,4 - ジニトロフルオロベンゼン (DNFB) を、メタノール中 5 mg/ml のスト

10

20

30

40

50

ックとして調製する。溶液は、光から保護し、毎日新しく調製する。10 mM ホウ酸緩衝液 (pH 8.0) 中 0.5 ~ 5.0  $\mu\text{g}/\text{ml}$  の DNFB で、5 ~ 30 分間 37 °C にてインキュベーションすることにより、ペプチドを標識する。

#### 【0619】

SNFB 標識：

4 - スルホニル - 2 - ニトロ - フルオロベンゼン (SNFB) を、メタノール中 5 mg / ml のストックとして調製する。溶液は、光から保護し、毎日新しく調製すべきである。10 mM ホウ酸緩衝液 (pH 8.0) 中 0.5 ~ 5.0  $\mu\text{g}/\text{ml}$  の DNFB で、5 ~ 30 分間 37 °C にてインキュベーションすることにより、ペプチドを標識する。

#### 【0620】

アセチル化 NTAA ペプチドの切断：

25 mM Tris - HCl (pH 7.5) 中で 10  $\mu\text{M}$  のアシルペプチドヒドロラーゼ (APH) 酵素 (Sulfolobus solfataricus 由来、SSO2693) と共に、10 分間 90 °C にてインキュベーションすることにより、アセチル化 NTAA をペプチドから切断する (Gogliettino、Balestrieri ら、2012 年)。

#### 【0621】

アミジン化 NTAA ペプチドの切断：

アミジン化 (グアニジン化) NTAA を、0.1 N NaOH 中で、10 分間 37 °C にてインキュベーションすることにより、ペプチドから切断する (Hamada、2016 年)。

#### 【0622】

(実施例 23)

モデル系によるコーディングタグ情報の記録タグへの分子内移行の実証

DNA モデル系を使用して、ビーズに固定化されている記録タグへのコーディングタグ情報の「分子内」移行を試験した (図 36A を参照されたい)。2 つの異なるタイプの記録タグオリゴヌクレオチドを使用した。saRT\_\_Abc\_\_v2 (配列番号 141) は、「A」DNA 捕捉配列 (配列番号 155) (「A'」結合性物質の模倣エピトープ) および対応する「A」バーコード (rtA\_\_BC) を含んでいた。saRT\_\_Bbc\_\_V2 (配列番号 142) は、「B」DNA 捕捉配列 (配列番号 156) (「B'」結合性物質の模倣エピトープ) および対応する「B」バーコード (rtB\_\_BC) を含んでいた。これらバーコードは、基本的な 65 セットの 15 mer バーコード (配列番号 1 ~ 65) およびそれらのリバース相補的配列 (配列番号 66 ~ 130) の組合せだった。rtA\_\_BC は、2 つのバーコード、BC\_\_1 および BC\_\_2 の同鎖上の組合せであり、rtB\_\_BC は、1 つのバーコード、BC\_\_3 のみである。同様に、コーディングタグのバーコード (エンコーダー配列) も、65 個の 15 mer バーコード (配列番号 1 ~ 65) の基本的なセットに由来するバーコードで構成されていた。CT\_\_A' - bc\_\_1PEG (配列番号 144) および CT\_\_B' - bc (配列番号 147) コーディングタグは、それぞれ相補的捕捉配列 A' および B' で構成され、それぞれ 15 mer バーコード BC\_\_5、および BC\_\_5 & BC\_\_6 と割り当てた。記録タグおよびコーディングタグのこの設計設定により、容易なゲル解析が可能になる。所望の「分子内」プライマー伸長は、類似サイズのオリゴヌクレオチド産物を生成するが、望ましくない「分子間」伸長は、「分子内」産物よりも 15 塩基大きな 1 つのオリゴ産物および 15 塩基短い別のオリゴ産物を生成する (図 36B)。

#### 【0623】

「分子内」対「分子間」情報移行に対する記録タグ密度の効果を評価した。正しい情報移行のためには、「分子間」情報移行 (A' コーディングタグは A 記録タグと結合するが、情報は B 記録タグへと移行されること、およびその逆) ではなく、「分子内」情報移行 (「A'」コーディングタグから A 記録タグ; B' コーディングタグから B 記録タグ) が観察されなければならない。ビーズ表面の記録タグ間隔の効果を試験するために、ビオチン化記録タグオリゴヌクレオチド saRT\_\_Abc\_\_v2 (配列番号 141) および saRT

10

20

30

40

50

— B b c — v 2 (配列番号 1 4 2) を 1 : 1 の比で混合し、その後、1 : 0、1 : 1 0、1 : 1 0 2、1 : 1 0 3、および 1 : 1 0 4 の比で、s a D u m m y - T 1 0 オリゴヌクレオチド (配列番号 1 4 3) に対してタイトレーションした。合計で 2 0 p m o l の記録タグオリゴヌクレオチドを、5 0 u l の固定化緩衝液 (5 m M T r i s - C l (p H 7 . 5)、0 . 5 m M E D T A、1 M N a C l) 中で 5 u l の M 2 7 0 ストレプトアビジンビーズ (T h e r m o) と共に、3 7 °C で 1 5 分間インキュベートした。ビーズを、1 0 0 u l の固定化緩衝液で 3 回、室温にて洗浄した。ほとんどのその後の洗浄ステップでは、1 0 0 u l の容積を使用した。ビーズを 2 5 u l の 5 × アニーリング緩衝液 (5 0 m M T r i s - C l (p H 7 . 5)、1 0 m M M g C l 2) に再懸濁し、コーディングタグミックスを添加することにより、コーディングタグ (後のサイクルには、D u p C T 配列との二本鎖アニーリングが必要である) を、ビーズに固定化されている記録タグとアニーリングさせる。6 5 °C で 1 分間加熱し、その後室温へと徐々に冷却する (0 . 2 °C / 秒) ことにより、コーディングタグを記録タグにアニーリングさせる。あるいは、コーディングタグを、3 7 °C にて P B S T 緩衝液中でアニーリングさせてもよい。ビーズを、室温にて P B S T (P B S + 0 . 1 % T w e e n - 2 0) で洗浄し、3 7 °C にて 5 分間 P B S T で 2 回洗浄し、室温にて P B S T で 1 回洗浄し、1 × アニーリング緩衝液で最終洗浄した。ビーズを、1 9 . 5 u l の伸長緩衝液 (5 0 m M T r i s - C l (p H 7 . 5)、2 m M M g S O 4、1 2 5 u M d N T P、5 0 m M N a C l、1 m M ジチオトレイトール、0 . 1 % T w e e n - 2 0、および 0 . 1 m g / m l B S A) に再懸濁し、3 7 °C で 1 5 分間インキュベートした。クレノウ e x o - D N A ポリメラーゼ (N E B、5 U / u l) を、0 . 1 2 5 U / u l の終濃度でビーズに添加し、3 7 °C で 5 分間インキュベートした。プライマー伸長後、ビーズを、P B S T で 2 回、5 0 u l の 0 . 1 N a O H で 1 回室温にて 5 分間、P B S T で 3 回、および P B S で 1 回洗浄した。下流 P C R アダプター配列 R 1 ' を付加するために、E n d C a p 2 T オリゴ (R 1 (配列番号 1 5 2) で構成される) を、コーディングタグオリゴヌクレオチドで実施したのと同様に、ハイブリダイズさせ、ビーズ上で伸長させた。アダプター配列を付加した後、最終伸長記録タグオリゴヌクレオチドを、9 5 % ホルムアミド / 1 0 m M E D T A 中で 5 分間 6 5 °C にてインキュベーションすることにより、ストレプトアビジンビーズから溶出した。溶出産物のおよそ 1 / 1 0 0 を、2 0 u l で 1 8 サイクルにて P C R 増幅し、P C R 産物の 1 u l を、1 0 % 変性 P A G E ゲルで解析した。得られたゲルは、コーディングタグ情報がポリメラーゼ伸長により記録タグに書き込まれるという原理の証明を実証し (図 3 6 C)、ビーズ表面の記録タグ密度を希釈すると、「分子間」伸長事象よりも「分子内」伸長事象を主に生成することができることを示した。

#### 【 0 6 2 4 】

このモデル系では、対応するエンコーダー配列およびユニバーサルリバープライマー部位を含む記録タグ R T \_ A B C および R T \_ B B C からの P C R 産物のサイズは、1 0 0 塩基対 (図 3 6 C) であるが、s a R T \_ A B C (配列番号 1 4 1) / C T \_ B ' B C (配列番号 1 4 7) および s a R T \_ B B C (配列番号 1 4 2) / C T \_ A ' B C (配列番号 1 4 4) の誤対合による産物は、それぞれ 1 1 5 塩基対および 8 5 塩基対である。図 3 6 D に示されているように、ビーズに s a R T \_ A B C (配列番号 1 4 1) および s a R T \_ B B C (配列番号 1 4 2) が高密度で存在する場合、3 つのバンドが観察された。高密度では、記録タグは、それ自体に結合する近位のコーディングタグ (分子内事象) または近隣の記録タグ (分子間事象) で伸長することが予想された。しかしながら、誤対合による産物のバンドは、ダミーオリゴヌクレオチドの記録タグを希釈することにより減少し、1 : 1 0 0 0 0 の比では消失した。この結果は、記録タグがビーズ表面に低密度で離間されていたため、分子間事象が減少したことを実証した。

10

20

30

40

【表 1】

表1.モデル系配列

名称	配列 (5'-3')	配列番号
saRT_Abc_v2	/5Biosg/TTTTTGCAAATGGCATTCTGACATCCCGTAGTCCGCGACACTAGATGTCTAGCATGCCGCCGTGTCTATGTGGAAACTGAGTG	141
saRT_Bbc_v2	/5Biosg/TTTTTTTTTTTGGACTGGTTCCAATTGACAAGCCGTAGTCCGCGACACTAGTAAGCCGGTATATCAACTGAGTG	142
saDummy-pT10	/5Biosg/TTTTTTTTTTT/3SpC3/	143
CT_A'-bc	GGATGTCAGAATGCCATTGCTTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGCACTCAGT/3SpC3/	144
CT_A'-bc_1PEG	GGATGTCAGAATGCCATTGCTTTTTTTTTTT/iSP18/CACTCAGTCCTAACGCGTATACGCACTCAGT/3SpC3/	145
CT_A'bc_5PEG	GGATGTCAGAATGCCATTGCTTTTTTTTTTT/iSP18//iSP18//iSP18//iSP18//iSP18/CACTCAGTCCTAACGCGTATACGTCAGT/3SpC3/	146
CT_B'bc	GCTTGTCGAATTGGAACCAGTCTTTT/iSP18/CACTCAGTCCTAACGCGTATACGGAATCTCGGCAGTTCAGT/3SpC3/	147
EndCap2T	CGATTTGCAAGGATCACTCGTCACTCAGTCCTAACGCGTATACG/3SpC3/	148
Sp	ACTGAGTG	149
Sp'	CACTCAGT	150
P1_f2	CGTAGTCCGCGACACTAG	151
R1	CGATTTGCAAGGATCACTCG	152
dupCT_A'BC	CGTATACGCGTTAGGACTGAGTG/3SpC3/	155
dupCT_B'BC	AACTGCCGAGATTCCTCGTATACGCGTTAGGACTGAGTG/3SpC3/	156

/ 3 S p C 3 / = 3 ' C 3 (炭素 3 個) スペース

/ 5 B i o s g / = 5 ' ビオチン

/ i S P 1 8 / = 1 8 原子ヘキサ - エチレングリコールスペース

【 0 6 2 5 】

( 実施例 2 4 )

ナノポアシーケンサーでの伸長記録タグ、伸長コーディングタグ、またはジタグ構築物の配列決定

DNA バーコードは、現在の塩基コールエラー率が 10 % 台またはそれよりも高いナノポアに基づくシーケンサーなどの、非常にエラーを起こしやすい NGS シーケンサーでの使用に耐えるように設計することができる。いくつかのエラー訂正コード系が文献に記載されている。こうしたエラー訂正コード系としては、Hamming コード、Reed-Solomon コード、Levenshtein コード、Lee コードなどが挙げられる。エラー耐性バーコードは、選択した設計パラメーターに応じて、挿入エラー、欠失エラー、および置換エラーを訂正することが可能な R Bioconductor パッケージ「DNA barcodes」を使用した、Hamming および Levenshtein コードに基づいていた (Buschmann および Bystrikh、2013 年)。65 個の異なる 15mer Hamming バーコードのセットが、図 27A に示されている (配列番号 1 ~ 65 に示されており、それらのリバース相補的配列はそれぞれ配列番号 66 ~ 130 に示されている)。これらのバーコードは、最小 Hamming 距離が 10 であり、4 つの置換エラーおよび 2 つのインデルエラーまでを自己訂正する。これは、1



0 %のエラー率でのナノポアシーケンサーの正確な読み出しに十分過ぎる程である。さらに、これらのバーコードは、予測ナノポア電流シグネチャを使用して、77個のオリジナルバーコードのセットから濾過されている（図27Bを参照されたい）。これらのバーコードは、バーコード全体にわたって大きな電流レベル差を示し、そのセットの他のバーコードと極力相関しないように濾過した。このようにすると、これらバーコードを使用したアッセイからの実際の生ナノポア電流レベルプロットを、塩基コールアルゴリズムを使用せずに、予測バーコードシグネチャに対して直接的にマッピングすることができる（Laszlo、Derringtonら、2014年）。

#### 【0626】

ナノポアシーケンシングを使用した伸長記録タグ、伸長コーディングタグ、またはジタグ構築物の解析を模倣するために、4つのフォワードプライマー（DTF1（配列番号157）、DTF2（配列番号158）、DTF3（配列番号159）、DTF4（配列番号160））および4つのリバースプライマー（DTR9（配列番号161）、DTR10（配列番号162）、DTR11（配列番号163）、DTR12（配列番号164））を使用した15merバーコードの小さなサブセットで構成されるPCR産物を生成した（図27C）。この8個プライマーのセットを、隣接するフォワードプライマーF1（配列番号165）およびリバースプライマーR1（配列番号166）と共にPCR反応に含めた。DTFおよびDTRプライマーは、相補的15merスペーサー配列（Sp15）（配列番号167）を介してアニーリングした。4つのDTFフォワードプライマーおよび4つのDTRリバースプライマーの組合せは、16個の考え得るPCR産物をもたらす。

#### 【表2】

PCR条件:

試薬	終濃度
F1 (5'リン酸化) (配列番号 165)	1 uM
R1 (5'リン酸化) (配列番号 166)	1 uM
DTF1-4 (配列番号 157-160); DTR9-12 (配列番号 161-164)	各 0.3 nM
VeraSeq 緩衝液 2	1X
dNTPs	200 uM
水	
VeraSeq 2.0 Ultra Pol	2 U/100 ul

PCRサイクル:

98 °C	30 秒
50 °C	2 分
98 °C	10 秒
55 °C	15 秒
72 °C	15 秒
最後の3ステップを19サイクル繰り返す	
72 °C	5 分

#### 【0627】

PCR後、アンプリコンを、平滑末端ライゲーション（図27C）により以下の通り鎖状化した。20ulのPCR産物を、20ulのQuick Ligase Mix（NEB）と直接混合し、室温で一晩インキュベートした。長さが約0.5~2kbの得られたライゲーション産物を、Zymo精製カラムを使用して精製し、20ulの水に溶出した。この精製ライゲーション産物の約7ulを、MinIon Library Rapid

Sequencing Prepキット (SQK-RAD002) に直接使用し、MinION Mk 1B (R9.4) デバイスで解析した。品質スコアが7.2 (正確性が約80%) の734bp ナノポアリードの一例が、図27Dに示されている。配列決定の正確性が不良であるにもかかわらず、MinION配列リードに対するバーコードの1alignに基づくアラインメントにより示されているように、配列中の多数のバーコードは容易に読み取り可能である (図27D)。

【0628】

(実施例25)

ゲルビーズへの単一細胞の封入

単一細胞を、標準的技法 (TamminenおよびVirta 2015年、Spencer、Tamminenら 2016年) を使用して、液滴 (約50μm) に封入する (図38を参照されたい)。ポリアクリルアミド (アクリルアミド: ビスアクリルアミド (29:1) (30% w/vol))、ベンゾフェノンメタクリルアミド (BM)、およびAPSを、細胞と共に不連続相に含有させて、連続油相にTEMEDを添加する (液滴内に拡散する) と重合を起こすことが可能な液滴を創出する。ベンゾフェノンを、ポリアクリルアミドゲル液滴のマトリックスに架橋させる。これは、その後のタンパク質とポリアクリルアミドマトリックスとの光親和性架橋を可能にする (Hughes、Spelkeら 2014年、Kang、Yamauchiら 2016年)。得られた単一細胞ゲルビーズ内に固定化されたタンパク質を、様々な方法を使用して単一細胞バーコード化することができる。一実施形態では、DNAタグを、以前に記載されているようなアミン反応性作用剤または光活性ベンゾフェノンDNAタグを使用して化学的にまたは光化学的に、単一細胞ゲルビーズ内の固定化されているタンパク質に付着させる。単一細胞ゲルビーズは、以前に記載されているようなバーコード付きビーズとの同時封入により、バーコードを含む液滴に封入することができ、タンパク質に移行されたDNAバーコードタグまたはその代わりに単一細胞ゲルビーズ内のタンパク質を、Amini、Cusanovich、およびGundersonら (Amini、Pushkarevら 2014年、Cusanovich、Dazaら 2015年) (Gunderson、Steemersら 2016年) により記載されているような一連のプール・アンド・スプリットステップによりコンビナトリアル的にインデックス化することができる。最も単純な実行例では、単一細胞ゲルビーズ内のタンパク質を、まず「クリック化学」部分で標識し (図40を参照されたい)、その後コンビナトリアルDNAバーコードを、プール・アンド・スプリット手法を使用して、タンパク質試料にクリックする。

参考文献:

【化2】

Harlow, Ed, and David Lane. Using Antibodies. Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press, 1999.

Hennessy BT, Lu Y, Gonzalez-Angulo AM, et al. A Technical Assessment of the Utility of Reverse Phase Protein Arrays for the Study of the Functional Proteome in Non-microdissected Human Breast Cancers. *Clinical proteomics*. 2010;6(4):129-151. Davidson, G. R., S. D. Armstrong and R. J. Beynon (2011). "Positional proteomics at the N-terminus as a means of proteome simplification." *Methods Mol Biol* 753: 229-242.

Zhang, L., Luo, S., and Zhang, B. (2016). The use of lectin microarray for assessing glycosylation of therapeutic proteins. *mAbs* 8, 524-535.

Akbani, R., K. F. Becker, N. Carragher, T. Goldstein, L. de Koning, U. Korf, L. Liotta, G. B. Mills, S. S. Nishizuka, M. Pawlak, E. F. Petricoin, 3rd, H. B. Pollard, B. Serrels and J. Zhu (2014). "Realizing the promise of reverse phase protein arrays for clinical, translational, and basic research: a workshop report: the RPPA (Reverse Phase Protein Array) society." *Mol Cell Proteomics* 13(7): 1625-1643.

10

20

30

40

50

## 【化 3】

- Amini, S., D. Pushkarev, L. Christiansen, E. Kostem, T. Royce, C. Turk, N. Pignatelli, A. Adey, J. O. Kitzman, K. Vijayan, M. Ronaghi, J. Shendure, K. L. Gunderson and F. J. Steemers (2014). "Haplotype-resolved whole-genome sequencing by contiguity-preserving transposition and combinatorial indexing." *Nat Genet* **46**(12): 1343-1349.
- Assadi, M., J. Lamerz, T. Jarutat, A. Farfsing, H. Paul, B. Gierke, E. Breiting, M. F. Templin, L. Essioux, S. Arbogast, M. Venturi, M. Pawlak, H. Langen and T. Schindler (2013). "Multiple protein analysis of formalin-fixed and paraffin-embedded tissue samples with reverse phase protein arrays." *Mol Cell Proteomics* **12**(9): 2615-2622.
- Bailey, J. M. and J. E. Shively (1990). "Carboxy-terminal sequencing: formation and hydrolysis of C-terminal peptidylthiohydantoins." *Biochemistry* **29**(12): 3145-3156.
- Bandara, H. M., D. P. Kennedy, E. Akin, C. D. Incavito and S. C. Burdette (2009). "Photoinduced release of Zn<sup>2+</sup> with ZinCleave-1: a nitrobenzyl-based caged complex." *Inorg Chem* **48**(17): 8445-8455.
- Bandara, H. M., T. P. Walsh and S. C. Burdette (2011). "A Second-generation photocage for Zn<sup>2+</sup> inspired by TPEN: characterization and insight into the uncaging quantum yields of ZinCleave chelators." *Chemistry* **17**(14): 3932-3941.
- Basle, E., N. Joubert and M. Pucheault (2010). "Protein chemical modification on endogenous amino acids." *Chem Biol* **17**(3): 213-227.
- Bilgicer, B., S. W. Thomas, 3rd, B. F. Shaw, G. K. Kaufman, V. M. Krishnamurthy, L. A. Estroff, J. Yang and G. M. Whitesides (2009). "A non-chromatographic method for the purification of a bivalently active monoclonal IgG antibody from biological fluids." *J Am Chem Soc* **131**(26): 9361-9367.
- Bochman, M. L., K. Paeschke and V. A. Zakian (2012). "DNA secondary structures: stability and function of G-quadruplex structures." *Nat Rev Genet* **13**(11): 770-780.
- Borgo, B. and J. J. Havranek (2014). "Motif-directed redesign of enzyme specificity." *Protein Sci* **23**(3): 312-320.
- Brouzes, E., M. Medkova, N. Savenelli, D. Marran, M. Twardowski, J. B. Hutchison, J. M. Rothberg, D. R. Link, N. Perrimon and M. L. Samuels (2009). "Droplet microfluidic technology for single-cell high-throughput screening." *Proc Natl Acad Sci U S A* **106**(34): 14195-14200.
- Brudno, Y., M. E. Birnbaum, R. E. Kleiner and D. R. Liu (2010). "An in vitro translation, selection and amplification system for peptide nucleic acids." *Nat Chem Biol* **6**(2): 148-155.
- Calcagno, S. and C. D. Klein (2016). "N-Terminal methionine processing by the zinc-activated *Plasmodium falciparum* methionine aminopeptidase 1b." *Appl Microbiol Biotechnol*.
- Cao, Y., G. K. Nguyen, J. P. Tam and C. F. Liu (2015). "Butelase-mediated synthesis of protein thioesters and its application for tandem chemoenzymatic ligation." *Chem Commun (Camb)* **51**(97): 17289-17292.
- Carty, R. P. and C. H. Hirs (1968). "Modification of bovine pancreatic ribonuclease A with 4-sulfonyloxy-2-nitrofluorobenzene. Isolation and identification of modified proteins." *J Biol Chem* **243**(20): 5244-5253.
- Chang, L., D. M. Rissin, D. R. Fournier, T. Piech, P. P. Patel, D. H. Wilson and D. C. Duffy (2012). "Single molecule enzyme-linked immunosorbent assays: theoretical considerations." *J Immunol Methods* **378**(1-2): 102-115.
- Chang, Y. Y. and C. H. Hsu (2015). "Structural basis for substrate-specific acetylation of Nalpa-acetyltransferase Ard1 from *Sulfolobus solfataricus*." *Sci Rep* **5**: 8673.

10

20

30

40

50

## 【化 4】

- Christoforou, A., C. M. Mulvey, L. M. Breckels, A. Geladaki, T. Hurrell, P. C. Hayward, T. Naake, L. Gatto, R. Viner, A. Martinez Arias and K. S. Lilley (2016). "A draft map of the mouse pluripotent stem cell spatial proteome." *Nat Commun* **7**: 8992.
- Creighton, C. J. and S. Huang (2015). "Reverse phase protein arrays in signaling pathways: a data integration perspective." *Drug Des Devel Ther* **9**: 3519-3527.
- Crosetto, N., M. Bienko and A. van Oudenaarden (2015). "Spatially resolved transcriptomics and beyond." *Nat Rev Genet* **16**(1): 57-66.
- Cusanovich, D. A., R. Daza, A. Adey, H. A. Pliner, L. Christiansen, K. L. Gunderson, F. J. Steemers, C. Trapnell and J. Shendure (2015). "Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing." *Science* **348**(6237): 910-914.
- Derrington, I. M., T. Z. Butler, M. D. Collins, E. Manrao, M. Pavlenok, M. Niederweis and J. H. Gundlach (2010). "Nanopore DNA sequencing with MspA." *Proc Natl Acad Sci U S A* **107**(37): 16060-16065.
- El-Sagheer, A. H., V. V. Cheong and T. Brown (2011). "Rapid chemical ligation of oligonucleotides by the Diels-Alder reaction." *Org Biomol Chem* **9**(1): 232-235.
- El-Sagheer, A. H., A. P. Sanzone, R. Gao, A. Tavassoli and T. Brown (2011). "Biocompatible artificial DNA linker that is read through by DNA polymerases and is functional in Escherichia coli." *Proc Natl Acad Sci U S A* **108**(28): 11338-11343.
- Emili, A., M. McLaughlin, K. Zagorovsky, J. B. Olsen, W. C. W. Chan and S. S. Sidhu (2017). Protein Sequencing Method and Reagents. USPTO. USA, The Governing Council of University of Toronto. **9,566,335 B1**.
- Erde, J., R. R. Loo and J. A. Loo (2014). "Enhanced FASP (eFASP) to increase proteome coverage and sample recovery for quantitative proteomic experiments." *J Proteome Res* **13**(4): 1885-1895.
- Farries, T. C., A. Harris, A. D. Auffret and A. Aitken (1991). "Removal of N-acetyl groups from blocked peptides with acylpeptide hydrolase. Stabilization of the enzyme and its application to protein sequencing." *Eur J Biochem* **196**(3): 679-685.
- Feist, P. and A. B. Hummon (2015). "Proteomic challenges: sample preparation techniques for microgram-quantity protein analysis from biological samples." *Int J Mol Sci* **16**(2): 3537-3563.
- Friedmann, D. R. and R. Marmorstein (2013). "Structure and mechanism of non-histone protein acetyltransferase enzymes." *FEBS J* **280**(22): 5570-5581.
- Frokjaer, S. and D. E. Otzen (2005). "Protein drug stability: a formulation challenge." *Nat Rev Drug Discov* **4**(4): 298-306.
- Fujii, Y., M. Kaneko, M. Neyazaki, T. Nogi, Y. Kato and J. Takagi (2014). "PA tag: a versatile protein tagging system using a super high affinity antibody against a dodecapeptide derived from human podoplanin." *Protein Expr Purif* **95**: 240-247.
- Gebauer, M. and A. Skerra (2012). "Anticalins small engineered binding proteins based on the lipocalin scaffold." *Methods Enzymol* **503**: 157-188.
- Gerry, N. P., N. E. Witowski, J. Day, R. P. Hammer, G. Barany and F. Barany (1999). "Universal DNA microarray method for multiplex detection of low abundance point mutations." *J Mol Biol* **292**(2): 251-262.
- Gogliettino, M., M. Balestrieri, E. Cocca, S. Mucerino, M. Rossi, M. Petrillo, E. Mazzella and G. Palmieri (2012). "Identification and characterisation of a novel acylpeptide hydrolase from *Sulfolobus solfataricus*: structural and functional insights." *PLoS One* **7**(5): e37921.

10

20

30

40

50

## 【化 5】

- Gogliettino, M., A. Riccio, M. Balestrieri, E. Cocca, A. Facchiano, T. M. D'Arco, C. Tesoro, M. Rossi and G. Palmieri (2014). "A novel class of bifunctional acylpeptide hydrolases--potential role in the antioxidant defense systems of the Antarctic fish *Trematomus bernacchii*." *FEBS J* **281**(1): 401-415.
- Granvogl, B., M. Ploscher and L. A. Eichacker (2007). "Sample preparation by in-gel digestion for mass spectrometry-based proteomics." *Anal Bioanal Chem* **389**(4): 991-1002.
- Gunderson, K. L., X. C. Huang, M. S. Morris, R. J. Lipshutz, D. J. Lockhart and M. S. Chee (1998). "Mutation detection by ligation to complete n-mer DNA arrays." *Genome Res* **8**(11): 1142-1153.
- Gunderson, K. L., F. J. Steemers, J. S. Fisher and R. Rigatti (2016). *Methods and Compositions for Analyzing Cellular Components*. WIPO, Illumina, Inc.
- Gunderson, K. L., F. J. Steemers, J. S. Fisher and R. Rigatti (2016). *Methods and compositions for analyzing cellular components*, Illumina, Inc.
- Guo, H., W. Liu, Z. Ju, P. Tamboli, E. Jonasch, G. B. Mills, Y. Lu, B. T. Hennessy and D. Tsavachidou (2012). "An efficient procedure for protein extraction from formalin-fixed, paraffin-embedded tissues for reverse phase protein arrays." *Proteome Sci* **10**(1): 56.
- Hamada, Y. (2016). "A novel N-terminal degradation reaction of peptides via N-amidation." *Bioorg Med Chem Lett* **26**(7): 1690-1695.
- Hermanson, G. (2013). *Bioconjugation Techniques*. Academic Press.
- Hernandez-Moreno, A. V., F. Villasenor, E. Medina-Rivero, N. O. Perez, L. F. Flores-Ortiz, G. Saab-Rincon and G. Luna-Barcenas (2014). "Kinetics and conformational stability studies of recombinant leucine aminopeptidase." *Int J Biol Macromol* **64**: 306-312.
- Hori, M., H. Fukano and Y. Suzuki (2007). "Uniform amplification of multiple DNAs by emulsion PCR." *Biochem Biophys Res Commun* **352**(2): 323-328.
- Horisawa, K. (2014). "Specific and quantitative labeling of biomolecules using click chemistry." *Front Physiol* **5**: 457.
- Hoshika, S., F. Chen, N. A. Leal and S. A. Benner (2010). "Artificial genetic systems: self-avoiding DNA in PCR and multiplexed PCR." *Angew Chem Int Ed Engl* **49**(32): 5554-5557.
- Hughes, A. J., D. P. Spelke, Z. Xu, C. C. Kang, D. V. Schaffer and A. E. Herr (2014). "Single-cell western blotting." *Nat Methods* **11**(7): 749-755.
- Hughes, C. S., S. Foehr, D. A. Garfield, E. E. Furlong, L. M. Steinmetz and J. Krijgsvelde (2014). "Ultrasensitive proteome analysis using paramagnetic bead technology." *Mol Syst Biol* **10**: 757.
- Kang, C. C., K. A. Yamauchi, J. Vlassakis, E. Sinkala, T. A. Duncombe and A. E. Herr (2016). "Single cell-resolution western blotting." *Nat Protoc* **11**(8): 1508-1530.
- Kang, T. S., L. Wang, C. N. Sarkissian, A. Gamez, C. R. Sriver and R. C. Stevens (2010). "Converting an injectable protein therapeutic into an oral form: phenylalanine ammonia lyase for phenylketonuria." *Mol Genet Metab* **99**(1): 4-9.
- Katritzky, A. R. and B. V. Rogovoy (2005). "Recent developments in guanylation agents." *ARKIVOC* **iv**(Issue in Honor of Prof. Nikolai Zefirov): 49-87.
- Klein, A. M., L. Mazutis, I. Akartuna, N. Tallapragada, A. Veres, V. Li, L. Peshkin, D. A. Weitz and M. W. Kirschner (2015). "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells." *Cell* **161**(5): 1187-1201.

10

20

30

40

50

## 【化 6】

- Knall, A. C., M. Hollauf and C. Slugovc (2014). "Kinetic studies of inverse electron demand Diels-Alder reactions (iEDDA) of norbornenes and 3,6-dipyridin-2-yl-1,2,4,5-tetrazine." *Tetrahedron Lett* **55**(34): 4763-4766.
- Le, Z. G., Z. C. Chen, Y. Hu and Q. G. Zheng (2005). "Organic Reactions in Ionic Liquids: Ionic Liquid-promoted Efficient Synthesis of Disubstituted and Trisubstituted Thioureas Derivatives." *Chinese Chemical Letters* **16**(2): 201-204.
- Lesch, V., A. Heuer, V. A. Tasis, C. Holm and J. Smiatek (2015). "Peptides in the presence of aqueous ionic liquids: tunable co-solutes as denaturants or protectants?" *Phys Chem Chem Phys* **17**(39): 26049-26053.
- Li, G., Y. Liu, Y. Liu, L. Chen, S. Wu, Y. Liu and X. Li (2013). "Photoaffinity labeling of small-molecule-binding proteins by DNA-templated chemistry." *Angew Chem Int Ed Engl* **52**(36): 9544-9549.
- Litovchick, A., M. A. Clark and A. D. Keefe (2014). "Universal strategies for the DNA-encoding of libraries of small molecules using the chemical ligation of oligonucleotide tags." *Artif DNA PNA XNA* **5**(1): e27896.
- Liu, Y. and S. Liang (2001). "Chemical carboxyl-terminal sequence analysis of peptides and proteins using tribenzylsilyl isothiocyanate." *J Protein Chem* **20**(7): 535-541.
- Lundblad, R. L. (2014). *Chemical reagents for protein modification*. Boca Raton, CRC Press, Taylor & Francis Group.
- Mashaghi, S. and A. M. van Oijen (2015). "External control of reactions in microdroplets." *Sci Rep* **5**: 11837.
- McCormick, R. M. (1989). "A solid-phase extraction procedure for DNA purification." *Anal Biochem* **181**(1): 66-74.
- Mendoza, V. L. and R. W. Vachet (2009). "Probing protein structure by amino acid-specific covalent labeling and mass spectrometry." *Mass Spectrom Rev* **28**(5): 785-815.
- Mikami, T., T. Takao, K. Yanagi and H. Nakazawa (2012). "N (alpha) Selective Acetylation of Peptides." *Mass Spectrom (Tokyo)* **1**(2): A0010.
- Moghaddam, M. J., L. de Campo, N. Kirby and C. J. Drummond (2012). "Chelating DTPA amphiphiles: ion-tunable self-assembly structures and gadolinium complexes." *Phys Chem Chem Phys* **14**(37): 12854-12862.
- Mukherjee, S., M. Ura, R. J. Hoey and A. A. Kossiakoff (2015). "A New Versatile Immobilization Tag Based on the Ultra High Affinity and Reversibility of the Calmodulin-Calmodulin Binding Peptide Interaction." *J Mol Biol* **427**(16): 2707-2725.
- Namimatsu, S., M. Ghazizadeh and Y. Sugisaki (2005). "Reversing the effects of formalin fixation with citraconic anhydride and heat: a universal antigen retrieval method." *J Histochem Cytochem* **53**(1): 3-11.
- Nguyen, G. K., Y. Cao, W. Wang, C. F. Liu and J. P. Tam (2015). "Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiopeptide." *Angew Chem Int Ed Engl* **54**(52): 15694-15698.
- Nguyen, G. K., S. Wang, Y. Qiu, X. Hemu, Y. Lian and J. P. Tam (2014). "Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis." *Nat Chem Biol* **10**(9): 732-738.
- Nishizuka, S. S. and G. B. Mills (2016). "New era of integrated cancer biomarker discovery using reverse-phase protein arrays." *Drug Metab Pharmacokinet* **31**(1): 35-45.
- Ohkubo, A., R. Kasuya, K. Sakamoto, K. Miyata, H. Taguchi, H. Nagasawa, T. Tsukahara, T. Watanobe, Y. Maki, K. Seio and M. Sekine (2008). "Protected DNA

10

20

30

40

50



## 【化 7】

Probes' capable of strong hybridization without removal of base protecting groups." *Nucleic Acids Res* **36**(6): 1952-1964.

Ojha, B., A. K. Singh, M. D. Adhikari, A. Ramesh and G. Das (2010). "2-Alkylmalonic acid: amphiphilic chelator and a potent inhibitor of metalloenzyme." *J Phys Chem B* **114**(33): 10835-10842.

Peng, X., H. Li and M. Seidman (2010). "A Template-Mediated Click-Click Reaction: PNA-DNA, PNA-PNA (or Peptide) Ligation, and Single Nucleotide Discrimination." *European J Org Chem* **2010**(22): 4194-4197.

Perbandt, M., O. Bruns, M. Vallazza, T. Lamla, C. Betzel and V. A. Erdmann (2007). "High resolution structure of streptavidin in complex with a novel high affinity peptide tag mimicking the biotin binding motif." *Proteins* **67**(4): 1147-1153.

Rauth, S., D. Hinz, M. Borger, M. Uhrig, M. Mayhaus, M. Riemenschneider and A. Skerra (2016). "High-affinity Anticalins with aggregation-blocking activity directed against the Alzheimer beta-amyloid peptide." *Biochem J* **473**(11): 1563-1578.

Ray, A. and B. Norden (2000). "Peptide nucleic acid (PNA): its medical and biotechnical applications and promise for the future." *FASEB J* **14**(9): 1041-1060.

Riley, N. M., A. S. Hebert and J. J. Coon (2016). "Proteomics Moves into the Fast Lane." *Cell Syst* **2**(3): 142-143.

Roloff, A., S. Ficht, C. Dose and O. Seitz (2014). "DNA-templated native chemical ligation of functionalized peptide nucleic acids: a versatile tool for single base-specific detection of nucleic acids." *Methods Mol Biol* **1050**: 131-141.

Roloff, A. and O. Seitz (2013). "The role of reactivity in DNA templated native chemical PNA ligation during PCR." *Bioorg Med Chem* **21**(12): 3458-3464.

Sakurai, K., T. M. Snyder and D. R. Liu (2005). "DNA-templated functional group transformations enable sequence-programmed synthesis using small-molecule reagents." *J Am Chem Soc* **127**(6): 1660-1661.

Schneider, K. and B. T. Chait (1995). "Increased stability of nucleic acids containing 7-deaza-guanosine and 7-deaza-adenosine may enable rapid DNA sequencing by matrix-assisted laser desorption mass spectrometry." *Nucleic Acids Res* **23**(9): 1570-1575.

Selvaraj, R. and J. M. Fox (2013). "trans-Cyclooctene--a stable, voracious dienophile for bioorthogonal labeling." *Curr Opin Chem Biol* **17**(5): 753-760.

Sharma, A. K., A. D. Kent and J. M. Heemstra (2012). "Enzyme-linked small-molecule detection using split aptamer ligation." *Anal Chem* **84**(14): 6104-6109.

Shembekar, N., C. Chaipan, R. Utharala and C. A. Merten (2016). "Droplet-based microfluidics in drug discovery, transcriptomics and high-throughput molecular genetics." *Lab Chip* **16**(8): 1314-1331.

Shenoy, N. R., J. E. Shively and J. M. Bailey (1993). "Studies in C-terminal sequencing: new reagents for the synthesis of peptidylthiohydantoins." *J Protein Chem* **12**(2): 195-205.

Shim, J. U., R. T. Ranasinghe, C. A. Smith, S. M. Ibrahim, F. Hollfelder, W. T. Huck, D. Klennerman and C. Abell (2013). "Ultrarapid generation of femtoliter microfluidic droplets for single-molecule-counting immunoassays." *ACS Nano* **7**(7): 5955-5964.

Shim, J. W., Q. Tan and L. Q. Gu (2009). "Single-molecule detection of folding and unfolding of the G-quadruplex aptamer in a nanopore nanocavity." *Nucleic Acids Res* **37**(3): 972-982.

Sidoli, S., Z. F. Yuan, S. Lin, K. Karch, X. Wang, N. Bhanu, A. M. Arnaudo, L. M. Britton, X. J. Cao, M. Gonzales-Cope, Y. Han, S. Liu, R. C. Molden, S. Wein, L.

10

20

30

40

50

## 【化 8】

- Afjehi-Sadat and B. A. Garcia (2015). "Drawbacks in the use of unconventional hydrophobic anhydrides for histone derivatization in bottom-up proteomics PTM analysis." *Proteomics* **15**(9): 1459-1469.
- Sletten, E. M. and C. R. Bertozzi (2009). "Bioorthogonal chemistry: fishing for selectivity in a sea of functionality." *Angew Chem Int Ed Engl* **48**(38): 6974-6998.
- Spencer, S. J., M. V. Tamminen, S. P. Preheim, M. T. Guo, A. W. Briggs, I. L. Brito, A. W. D., L. K. Pitkanen, F. Vigneault, M. P. Juhani Virta and E. J. Alm (2016). "Massively parallel sequencing of single cells by epicPCR links functional genes with phylogenetic markers." *ISME J* **10**(2): 427-436.
- Spicer, C. D. and B. G. Davis (2014). "Selective chemical protein modification." *Nat Commun* **5**: 4740. 10
- Spiropoulos, N. G. and J. M. Heemstra (2012). "Templating effect in DNA proximity ligation enables use of non-bioorthogonal chemistry in biological fluids." *Artif DNA PNA XNA* **3**(3): 123-128.
- Switzar, L., M. Giera and W. M. Niessen (2013). "Protein digestion: an overview of the available techniques and recent developments." *J Proteome Res* **12**(3): 1067-1077.
- Tamminen, M. V. and M. P. Virta (2015). "Single gene-based distinction of individual microbial genomes from a mixed population of microbial cells." *Front Microbiol* **6**: 195.
- Tessler, L. (2011). *Digital Protein Analysis: Technologies for Protein Diagnostics and Proteomics through Single-Molecule Detection*. Ph.D., WASHINGTON UNIVERSITY IN ST. LOUIS. 20
- Tyson, J. and J. A. Armour (2012). "Determination of haplotypes at structurally complex regions using emulsion haplotype fusion PCR." *BMC Genomics* **13**: 693.
- Vauquelin, G. and S. J. Charlton (2013). "Exploring avidity: understanding the potential gains in functional affinity and target residence time of bivalent and heterobivalent ligands." *Br J Pharmacol* **168**(8): 1771-1785.
- Veggiani, G., T. Nakamura, M. D. Brenner, R. V. Gayet, J. Yan, C. V. Robinson and M. Howarth (2016). "Programmable polypeptides built using twin peptide superglues." *Proc Natl Acad Sci U S A* **113**(5): 1202-1207.
- Wang, D., S. Fang and R. M. Wohlhueter (2009). "N-terminal derivatization of peptides with isothiocyanate analogues promoting Edman-type cleavage and enhancing sensitivity in electrospray ionization tandem mass spectrometry analysis." *Anal Chem* **81**(5): 1893-1900. 30
- Williams, B. A. and J. C. Chaput (2010). "Synthesis of peptide-oligonucleotide conjugates using a heterobifunctional crosslinker." *Curr Protoc Nucleic Acid Chem* **Chapter 4**: Unit4 41.
- Wu, H. and N. K. Devaraj (2016). "Inverse Electron-Demand Diels-Alder Bioorthogonal Reactions." *Top Curr Chem (J)* **374**(1): 3.
- Xiong, A. S., R. H. Peng, J. Zhuang, F. Gao, Y. Li, Z. M. Cheng and Q. H. Yao (2008). "Chemical gene synthesis: strategies, softwares, error corrections, and applications." *FEMS Microbiol Rev* **32**(3): 522-540.
- Yao, Y., M. Docter, J. van Ginkel, D. de Ridder and C. Joo (2015). "Single-molecule protein sequencing through fingerprinting: computational assessment." *Phys Biol* **12**(5): 055003. 40



## 【化 9】

Zakeri, B., J. O. Fierer, E. Celik, E. C. Chittock, U. Schwarz-Linek, V. T. Moy and M. Howarth (2012). "Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin." *Proc Natl Acad Sci U S A* **109**(12): E690-697.  
 Zhang, L., K. Zhang, S. Rauf, D. Dong, Y. Liu and J. Li (2016). "Single-Molecule Analysis of Human Telomere Sequence Interactions with G-quadruplex Ligand." *Anal Chem* **88**(8): 4533-4540.  
 Zhou, H., Z. Ning, A. E. Starr, M. Abu-Farha and D. Figeys (2012). "Advancements in top-down proteomics." *Anal Chem* **84**(2): 720-734.  
 Zilionis, R., J. Nainys, A. Veres, V. Savova, D. Zemmour, A. M. Klein and L. Mazutis (2017). "Single-cell barcoding and sequencing using droplet microfluidics." *Nat Protoc* **12**(1): 44-73.

10

## 【 0 6 2 9 】

上記の詳細な説明に照らして、これらおよび他の変更を実施形態に対して行うことができる。一般的に、以下の特許請求の範囲においては、使用されている用語は、特許請求の範囲を、本明細書および本特許請求の範囲で開示されている特定の実施形態に限定するものとは解釈されるべきでなく、考え得るすべての実施形態ならびにそのような特許請求の範囲が権利を有する等価物の完全な範囲を含むと解釈されるべきである。したがって、特許請求の範囲は、本開示により限定されない。

20

## 【 0 6 3 0 】

上記に記載されている種々の実施形態を組み合わせ、さらなる実施形態を提供することができる。米国特許仮出願第 6 2 / 3 3 0 , 8 4 1 号、米国特許仮出願第 6 2 / 3 3 9 , 0 7 1 号、および米国特許仮出願第 6 2 / 3 7 6 , 8 8 6 号を含む、本明細書で引用されている、および/または出願データシートに列挙されている米国特許、米国特許出願公開、米国特許出願、外国特許、外国特許出願、および非特許文献はすべて、それらの全体が参照により本明細書に組み込まれる。必要に応じて、種々の特許、出願、および刊行物の概念を用いるために、本実施形態の態様を改変して、またさらなる実施形態を提供することができる。

本発明は、例えば、以下の項目を提供する。

30

( 項目 1 )

巨大分子を解析するための方法であって、

( a ) 固体支持体に接合した巨大分子および付随する記録タグを用意するステップと；

( b ) 前記巨大分子を、前記巨大分子に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；

( c ) 前記第 1 のコーディングタグの情報を前記記録タグに移行させて、一次伸長記録タグを生成するステップと；

( d ) 前記巨大分子を、前記巨大分子に結合することが可能な第 2 の結合性物質であって、前記第 2 の結合性物質に関する識別情報を有する第 2 のコーディングタグを含む第 2 の結合性物質と接触させるステップと；

40

( e ) 前記第 2 のコーディングタグの情報を前記一次伸長記録タグに移行させて、二次伸長記録タグを生成するステップと；

( f ) 前記二次伸長記録タグを解析するステップとを含む方法。

( 項目 2 )

接触させるステップ ( b ) および ( d ) を逐次的に実施する、項目 1 に記載の方法。

( 項目 3 )

接触させるステップ ( b ) および ( d ) を同時に実施する、項目 1 に記載の方法。

( 項目 4 )

50

ステップ ( e ) と ( f ) の間に、

( x ) 前記第 2 の結合性物質を、前記巨大分子に結合することが可能な第 3 の ( またはより高次の ) 結合性物質であって、前記第 3 の ( またはより高次の ) 結合性物質に関する識別情報を有する第 3 の ( またはより高次の ) コーディングタグを含む第 3 の ( またはより高次の ) 結合性物質に置き換えることにより、ステップ ( d ) および ( e ) を 1 回または複数回繰り返すステップと；

( y ) 前記第 3 の ( またはより高次の ) コーディングタグの情報を前記第 2 の ( またはより高次の ) 伸長記録タグに移行させて、第 3 の ( またはより高次の ) 伸長記録タグを生成するステップと

をさらに含み、

ステップ ( f ) において前記第 3 の ( またはより高次の ) 伸長記録タグを解析する、項目 1 に記載の方法。

( 項目 5 )

巨大分子を解析するための方法であって、

( a ) 固体支持体に接合した巨大分子、付随する第 1 の記録タグおよび付随する第 2 の記録タグを用意するステップと；

( b ) 前記巨大分子を、前記巨大分子に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；

( c ) 前記第 1 のコーディングタグの情報を前記第 1 の記録タグに移行させて、第 1 の伸長記録タグを生成するステップと；

( d ) 前記巨大分子を、前記巨大分子に結合することが可能な第 2 の結合性物質であって、前記第 2 の結合性物質に関する識別情報を有する第 2 のコーディングタグを含む第 2 の結合性物質と接触させるステップと；

( e ) 前記第 2 のコーディングタグの情報を前記第 2 の記録タグに移行させて、第 2 の伸長記録タグを生成するステップと；

( f ) 前記第 1 の伸長記録タグおよび第 2 の伸長記録タグを解析するステップとを含む方法。

( 項目 6 )

接触させるステップ ( b ) および ( d ) を逐次的に実施する、項目 5 に記載の方法。

( 項目 7 )

接触させるステップ ( b ) および ( d ) を同時に実施する、項目 5 に記載の方法。

( 項目 8 )

ステップ ( a ) が、前記固体支持体に接合した付随する第 3 の ( またはより高次の ) 記録タグを用意するステップをさらに含む、項目 5 に記載の方法。

( 項目 9 )

ステップ ( e ) と ( f ) の間に、

( x ) 前記第 2 の結合性物質を、前記巨大分子に結合することが可能な第 3 の ( またはより高次の ) 結合性物質であって、前記第 3 の ( またはより高次の ) 結合性物質に関する識別情報を有する第 3 の ( またはより高次の ) コーディングタグを含む第 3 の ( またはより高次の ) 結合性物質に置き換えることにより、ステップ ( d ) および ( e ) を 1 回または複数回繰り返すステップと；

( y ) 前記第 3 の ( またはより高次の ) コーディングタグの情報を前記第 3 の ( またはより高次の ) 記録タグに移行させて、第 3 の ( またはより高次の ) 伸長記録タグを生成するステップと

をさらに含み、

ステップ ( f ) において前記第 1 の伸長記録タグ、前記第 2 の伸長記録タグおよび前記第 3 の ( またはより高次の ) 伸長記録タグを解析する、項目 8 に記載の方法。

( 項目 10 )

前記第 1 のコーディングタグ、前記第 2 のコーディングタグ、および任意のより高次の

10

20

30

40

50

コーディングタグが、結合サイクル特異的スパーサー配列を含む、項目 5 から 9 までのいずれか一項に記載の方法。

(項目 1 1)

ペプチドを解析するための方法であって、

- (a) 固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；
  - (b) 前記ペプチドの N 末端アミノ酸 (NTAA) を化学薬剤で修飾するステップと；
  - (c) 前記ペプチドを、修飾された前記 NTAA に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；
  - (d) 前記第 1 のコーディングタグの情報を前記記録タグに移行させて、伸長記録タグを生成するステップと；
  - (e) 前記伸長記録タグを解析するステップと
- を含む方法。

10

(項目 1 2)

ステップ (c) が、前記ペプチドを、前記第 2 の (またはより高次の) 結合性物質に関する識別情報を有する第 2 の (またはより高次の) コーディングタグを含む第 2 の (またはより高次の) 結合性物質であって、ステップ (b) の前記修飾された NTAA 以外の修飾された NTAA に結合することが可能である第 2 の (またはより高次の) 結合性物質と接触させることをさらに含む、項目 1 1 に記載の方法。

(項目 1 3)

20

前記ペプチドの前記第 2 の (またはより高次の) 結合性物質との接触を、前記ペプチドの前記第 1 の結合性物質との接触後に逐次的に行う、項目 1 2 に記載の方法。

(項目 1 4)

前記ペプチドの前記第 2 の (またはより高次の) 結合性物質との接触を、前記ペプチドの前記第 1 の結合性物質との接触と同時に行う、項目 1 2 に記載の方法。

(項目 1 5)

前記化学薬剤が、イソチオシアネート誘導体、2, 4 - ジニトロベンゼンスルホン酸 (dinitrobenzenesulfonic) (DNBS)、4 - スルホニル - 2 - ニトロフルオロベンゼン (SNFB) 1 - フルオロ - 2, 4 - ジニトロベンゼン、ダンシルクロリド、7 - メトキシクマリン酢酸、チオアシル化試薬、チオアセチル化試薬、またはチオベンジル化試薬である、項目 1 1 から 1 4 までのいずれか一項に記載の方法。

30

(項目 1 6)

ペプチドを解析するための方法であって、

- (a) 固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；
- (b) 前記ペプチドの N 末端アミノ酸 (NTAA) を化学薬剤で修飾して、修飾された NTAA を得るステップと；
- (c) 前記ペプチドを、前記修飾された NTAA に結合することが可能な第 1 の結合性物質であって、前記第 1 の結合性物質に関する識別情報を有する第 1 のコーディングタグを含む第 1 の結合性物質と接触させるステップと；
- (d) 前記第 1 のコーディングタグの情報を前記記録タグに移行させて、第 1 の伸長記録タグを生成するステップと；
- (e) 前記修飾された NTAA を除去して、新しい NTAA を露出させるステップと；
- (f) 前記ペプチドの前記新しい NTAA を化学薬剤で修飾して、新しく修飾された NTAA を得るステップと；
- (g) 前記ペプチドを、前記新しく修飾された NTAA に結合することが可能な第 2 の結合性物質であって、前記第 2 の結合性物質に関する識別情報を有する第 2 のコーディングタグを含む第 2 の結合性物質と接触させるステップと；
- (h) 前記第 2 のコーディングタグの情報を前記第 1 の伸長記録タグに移行させて、第 2 の伸長記録タグを生成するステップと；
- (i) 前記第 2 の伸長記録タグを解析するステップと

40

50

を含む方法。

(項目 17)

ペプチドを解析するための方法であって、

(a) 固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；

(b) 前記ペプチドを、前記ペプチドのN末端アミノ酸 (NTAA) に結合することが可能な第1の結合性物質であって、前記第1の結合性物質に関する識別情報を有する第1のコーディングタグを含む第1の結合性物質と接触させるステップと；

(c) 前記第1のコーディングタグの情報を前記記録タグに移行させて、伸長記録タグを生成するステップと；

(d) 前記伸長記録タグを解析するステップと

10

を含む方法。

(項目 18)

ステップ (b) が、前記ペプチドを、前記第2の (またはより高次の) 結合性物質に関する識別情報を有する第2の (またはより高次の) コーディングタグを含む第2の (またはより高次の) 結合性物質であって、前記ペプチドの前記NTAA以外のNTAAに結合することが可能な第2の (またはより高次の) 結合性物質と接触させることをさらに含む、項目 17 に記載の方法。

(項目 19)

前記ペプチドの前記第2の (またはより高次の) 結合性物質との接触を、前記ペプチドの前記第1の結合性物質との接触後に逐次的に行う、項目 18 に記載の方法。

20

(項目 20)

前記ペプチドの前記第2の (またはより高次の) 結合性物質との接触を、前記ペプチドの前記第1の結合性物質との接触と同時に行う、項目 18 に記載の方法。

(項目 21)

ペプチドを解析するための方法であって、

(a) 固体支持体に接合したペプチドおよび付随する記録タグを用意するステップと；

(b) 前記ペプチドを、前記ペプチドのN末端アミノ酸 (NTAA) に結合することが可能な第1の結合性物質であって、前記第1の結合性物質に関する識別情報を有する第1のコーディングタグを含む第1の結合性物質と接触させるステップと；

(c) 前記第1のコーディングタグの情報を前記記録タグに移行させて、第1の伸長記録タグを生成するステップと；

30

(d) 前記NTAAを除去して、前記ペプチドの新しいNTAAを露出させるステップと；

(e) 前記ペプチドを、前記新しいNTAAに結合することが可能な第2の結合性物質であって、前記第2の結合性物質に関する識別情報を有する第2のコーディングタグを含む第2の結合性物質と接触させるステップと；

(h) 前記第2のコーディングタグの情報を前記第1の伸長記録タグに移行させて、第2の伸長記録タグを生成するステップと；

(i) 前記第2の伸長記録タグを解析するステップと

を含む方法。

40

(項目 22)

前記巨大分子が、タンパク質、ポリペプチドまたはペプチドである、項目 1 から 10 までのいずれか一項に記載の方法。

(項目 23)

前記巨大分子が、ペプチドである、項目 1 から 10 までのいずれか一項に記載の方法。

(項目 24)

前記ペプチドが、生体試料由来のタンパク質を断片化することによって得られる、項目 11 から 23 までのいずれか一項に記載の方法。

(項目 25)

前記巨大分子が、脂質、炭水化物、または大環状分子である、項目 1 から 10 までのい

50

ずれか一項に記載の方法。

(項目 26)

前記記録タグが、DNA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、またはこれらの組合せである、項目1から25までのいずれか一項に記載の方法。

(項目 27)

前記記録タグが、ユニバーサルプライミング部位を含む、項目1から26までのいずれか一項に記載の方法。

(項目 28)

前記ユニバーサルプライミング部位が、増幅、配列決定、またはその両方のためのプライミング部位を含む、項目27に記載の方法。

(項目 29)

前記記録タグが、一意の分子識別子(UMI)を含む、項目1から28までに記載の方法。

(項目 30)

前記記録タグが、バーコードを含む、項目1から29までのいずれか一項に記載の方法。

(項目 31)

前記記録タグが、その3'末端にスペーサーを含む、項目1から30までのいずれか一項に記載の方法。

(項目 32)

前記巨大分子および前記付随する記録タグを、前記固体支持体に共有結合により接合させる、項目1から31までのいずれか一項に記載の方法。

(項目 33)

前記固体支持体が、ビーズ、多孔質ビーズ、多孔質マトリックス、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、ナイロン、シリコンウェーハチップ、フロースルーチップ、信号変換電子機器を含むバイオチップ、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、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアである、項目1から32までのいずれか一項に記載の方法。

(項目 34)

前記固体支持体が、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズである、項目33に記載の方法。

(項目 35)

複数の巨大分子および付随する記録タグを固体支持体に接合する、項目1から34までのいずれか一項に記載の方法。

(項目 36)

前記複数の巨大分子の間に前記固体支持体上で平均距離>50nmの間隔をあける、項目35に記載の方法。

(項目 37)

前記結合性物質が、ポリペプチドまたはタンパク質である、項目1から36までのいずれか一項に記載の方法。

(項目 38)

前記結合性物質が、改変アミノペプチダーゼ、改変アミノアシルtRNA合成酵素、改変アンチカリン、または改変C1pSである、項目37に記載の方法。

(項目 39)

前記結合性物質が、巨大分子に選択的に結合することが可能である、項目1から38までのいずれか一項に記載の方法。

(項目 40)

前記コーディングタグが、DNA分子、RNA分子、BNA分子、XNA分子、LNA

10

20

30

40

50

分子、PNA分子、PNA分子、またはこれらの組合せである、項目1から39までのいずれか一項に記載の方法。

(項目41)

前記コーディングタグが、エンコーダー配列を含む、項目1から40までのいずれか一項に記載の方法。

(項目42)

前記コーディングタグが、スパーサー、結合サイクル特異的配列、一意の分子識別子、ユニバーサルプライミング部位、またはそれらの任意の組合せをさらに含む、項目1から41までのいずれか一項に記載の方法。

(項目43)

前記結合性物質と前記コーディングタグが、リンカーによって接合されている、項目1から42までのいずれか一項に記載の方法。

(項目44)

前記結合性物質と前記コーディングタグが、SpyTag/SpyCatcherまたはSnoopTag/SnoopCatcherペプチド-タンパク質対によって接合されている、項目1から42までに記載の方法。

(項目45)

前記コーディングタグの情報の前記記録タグへの移行が、DNAリガーゼによって媒介される、項目1から44までのいずれか一項に記載の方法。

(項目46)

前記コーディングタグの情報の前記記録タグへの移行が、DNAポリメラーゼによって媒介される、項目1から44までのいずれか一項に記載の方法。

(項目47)

前記コーディングタグの情報の前記記録タグへの移行が、化学的ライゲーションによって媒介される、項目1から44までのいずれか一項に記載の方法。

(項目48)

前記伸長記録タグの解析が、核酸配列決定法を含む、項目1から47までのいずれか一項に記載の方法。

(項目49)

前記核酸配列決定法が、合成による配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、ポロニーシーケンシング、イオン半導体シーケンシング、またはパイロシーケンシングである、項目48に記載の方法。

(項目50)

前記核酸配列決定法が、単一分子リアルタイムシーケンシング、ナノポアに基づく配列決定、または先端顕微鏡を使用したDNAのダイレクトイメージングである、項目48に記載の方法。

(項目51)

前記伸長記録タグを解析前に増幅する、項目1から50までのいずれか一項に記載の方法。

(項目52)

前記伸長記録タグに含有されるコーディングタグ情報の順序が、前記結合性物質による前記巨大分子への結合の順序に関する情報を提供する、項目1から51までに記載の方法。

(項目53)

前記伸長記録タグに含有される前記コーディングタグ情報の頻度が、前記結合性物質による前記巨大分子への結合の頻度に関する情報を提供する、項目1から52までに記載の方法。

(項目54)

複数の巨大分子を表す複数の伸長記録タグを並行して解析する、項目1から53までに記載の方法。

(項目55)

10

20

30

40

50

前記複数の巨大分子を表す複数の伸長記録タグを多重化アッセイで解析する、項目 5 4 に記載の方法。

(項目 5 6)

前記複数の伸長記録タグが、解析前に標的濃縮アッセイを受ける、項目 1 から 5 5 までのいずれか一項に記載の方法。

(項目 5 7)

前記複数の伸長記録タグが、解析前にサブトラクシオンアッセイを受ける、項目 1 から 5 6 までのいずれか一項に記載の方法。

(項目 5 8)

前記複数の伸長記録タグが、極めて豊富な種を減少させるために解析前に正規化アッセイを受ける、項目 1 から 5 7 までのいずれか一項に記載の方法。

10

(項目 5 9)

前記 N T A A を、改変アミノペプチダーゼ、改変アミノ酸 t R N A 合成酵素、穏やかなエドマン分解、エドマナーゼ ( E d m a n a s e ) 酵素、または無水 T F A によって除去する、項目 1 から 5 8 までのいずれか一項に記載の方法。

(項目 6 0)

少なくとも 1 つの結合性物質が末端アミノ酸残基に結合する、項目 1 から 5 9 までのいずれか一項に記載の方法。

(項目 6 1)

少なくとも 1 つの結合性物質が翻訳後修飾されたアミノ酸に結合する、項目 1 から 6 0 までのいずれか一項に記載の方法。

20

(項目 6 2)

複数のタンパク質複合体、タンパク質、またはポリペプチドを含む試料由来の 1 つまたは複数のペプチドを解析するための方法であって、

( a ) 前記試料中の前記複数のタンパク質複合体、タンパク質、またはポリペプチドを複数のコンパートメントに分配するステップであって、各コンパートメントが、任意選択で固体支持体と接合した複数のコンパートメントタグを含み、前記複数のコンパートメントタグが、個々のコンパートメント内では同じであり、他のコンパートメントのコンパートメントタグとは異なる、ステップと；

( b ) 前記複数のタンパク質複合体、タンパク質、および / またはポリペプチドを複数のペプチドに断片化するステップと；

30

( c ) 前記複数のペプチドと前記複数のコンパートメントタグを、前記複数のペプチドと前記複数のコンパートメント内の前記複数のコンパートメントタグとのアニーリングまたは接合を可能にするのに十分な条件下で接触させ、それにより、複数のコンパートメントタグ付きペプチドを生成するステップと；

( d ) 前記コンパートメントタグ付きペプチドを前記複数のコンパートメントから収集するステップと；

( e ) 1 つまたは複数のコンパートメントタグ付きペプチドを、項目 1 から 2 1 までおよび項目 2 6 から 6 1 までのいずれか一項に記載の方法に従って解析するステップとを含む方法。

40

(項目 6 3)

前記コンパートメントがマイクロ流体液滴である、項目 6 2 に記載の方法。

(項目 6 4)

前記コンパートメントがマイクロウェルである、項目 6 2 に記載の方法。

(項目 6 5)

前記コンパートメントが、表面上の分離された領域である、項目 6 2 に記載の方法。

(項目 6 6)

各コンパートメントが、平均して単一の細胞を含む、項目 6 2 から 6 5 までのいずれか一項に記載の方法。

(項目 6 7)

50

複数のタンパク質複合体、タンパク質、またはポリペプチドを含む試料由来の１つまたは複数のペプチドを解析するための方法であって、

(a) 前記複数のタンパク質複合体、タンパク質、またはポリペプチドを複数のユニバーサルDNAタグで標識するステップと；

(b) 前記試料中の前記複数の標識されたタンパク質複合体、タンパク質、またはポリペプチドを複数のコンパートメントに分配するステップであって、各コンパートメントが、複数のコンパートメントタグを含み、前記複数のコンパートメントタグが、個々のコンパートメント内では同じであり、他のコンパートメントのコンパートメントタグとは異なる、ステップと；

(c) 前記複数のタンパク質複合体、タンパク質、またはポリペプチドと前記複数のコンパートメントタグを、前記複数のタンパク質複合体、タンパク質、またはポリペプチドと前記複数のコンパートメント内の前記複数のコンパートメントタグとのアニーリングまたは接合を可能にするのに十分な条件下で接触させ、それにより、複数のコンパートメントタグ付きタンパク質複合体、タンパク質またはポリペプチドを生成するステップと；

(d) 前記コンパートメントタグ付きタンパク質複合体、タンパク質、またはポリペプチドを前記複数のコンパートメントから収集するステップと；

(e) 任意選択で前記コンパートメントタグ付きタンパク質複合体、タンパク質、またはポリペプチドをコンパートメントタグ付きペプチドに断片化するステップと；

(f) １つまたは複数のコンパートメントタグ付きペプチドを、項目１から２１までおよび項目２６から６１までのいずれか一項に記載の方法に従って解析するステップとを含む方法。

(項目６８)

コンパートメントタグ情報を、ペプチドに付随する記録タグにプライマー伸長またはライゲーションによって移行させる、項目６２から６７までのいずれか一項に記載の方法。

(項目６９)

前記固体支持体がビーズを含む、項目６２から６８までのいずれか一項に記載の方法。

(項目７０)

前記ビーズが、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズである、項目６９に記載の方法。

(項目７１)

前記コンパートメントタグが、一本鎖または二本鎖核酸分子を含む、項目６２から７０までのいずれか一項に記載の方法。

(項目７２)

前記コンパートメントタグが、バーコードおよび任意選択でUMIを含む、項目６２から７１までのいずれか一項に記載の方法。

(項目７３)

前記固体支持体がビーズであり、前記コンパートメントタグがバーコードを含み、さらに、前記複数のコンパートメントタグが接合したビーズを、スプリット・アンド・プール(split-and-pool)合成によって形成する、項目７２に記載の方法。

(項目７４)

前記固体支持体がビーズであり、前記コンパートメントタグがバーコードを含み、さらに、複数のコンパートメントタグが接合したビーズを、個々の合成または固定化によって形成する、項目７２に記載の方法。

(項目７５)

前記コンパートメントタグが記録タグ内の成分であり、前記記録タグが任意選択でスペーサー、一意の分子識別子、ユニバーサルプライミング部位、またはそれらの任意の組合せをさらに含む、項目６２から７４までのいずれか一項に記載の方法。

(項目７６)

前記コンパートメントタグが、前記複数のタンパク質複合体、タンパク質、またはポリ

10

20

30

40

50



ペプチドの内部アミノ酸またはN末端アミノ酸と反応することが可能な機能的部分をさらに含む、項目62から75までのいずれか一項に記載の方法。

(項目77)

前記機能的部分がNH<sub>2</sub>基である、項目76に記載の方法。

(項目78)

前記機能的部分がアルデヒド基である、項目76に記載の方法。

(項目79)

前記複数のコンパートメントタグが、前記コンパートメントタグを前記コンパートメントに印刷、スポッティング、インク噴射すること、またはその組合せによって形成されたものである、項目62から78までのいずれか一項に記載の方法。

(項目80)

前記コンパートメントタグが、ペプチドをさらに含む、項目62から79までのいずれか一項に記載の方法。

(項目81)

前記コンパートメントタグペプチドが、タンパク質リガーゼ認識配列を含む、項目80に記載の方法。

(項目82)

前記タンパク質リガーゼが、プテラーゼIまたはそのホモログである、項目81に記載の方法。

(項目83)

前記複数のポリペプチドをプロテアーゼで断片化する、項目62から82までのいずれか一項に記載の方法。

(項目84)

前記プロテアーゼがメタロプロテアーゼである、項目83に記載の方法。

(項目85)

前記メタロプロテアーゼの活性が金属カチオンの光活性化放出によってモジュレートされる、項目84に記載の方法。

(項目86)

前記複数のポリペプチドを前記複数のコンパートメントに分配する前に1つまたは複数の豊富なタンパク質を前記試料からサブトラクションすることをさらに含む、項目62から85までのいずれか一項に記載の方法。

(項目87)

前記複数のペプチドと前記コンパートメントタグを接合する前に前記コンパートメントタグを前記固体支持体から遊離させることをさらに含む、項目62から86までに記載の方法。

(項目88)

ステップ(d)の後に、前記コンパートメントタグ付きペプチドを固体支持体に記録タグを伴って接合させることをさらに含む、項目62に記載の方法。

(項目89)

前記コンパートメントタグ付きペプチド上の前記コンパートメントタグの情報を前記付随する記録タグに移行させることをさらに含む、項目88に記載の方法。

(項目90)

ステップ(e)の前に前記コンパートメントタグを前記コンパートメントタグ付きペプチドから除去することをさらに含む、項目89に記載の方法。

(項目91)

解析されるペプチドが由来する前記単一の細胞の同一性を前記解析されるペプチドのコンパートメントタグ配列に基づいて決定することをさらに含む、項目62から90までのいずれか一項に記載の方法。

(項目92)

解析されるペプチドが由来する前記タンパク質またはタンパク質複合体の同一性を前記

10

20

30

40

50

解析されるペプチドのコンパートメントタグ配列に基づいて決定することをさらに含む、  
項目 6 2 から 9 0 までのいずれか一項に記載の方法。

( 項目 9 3 )

複数の巨大分子を解析するための方法であって、

( a ) 固体支持体に接合した複数の巨大分子および付随する記録タグを用意するステップと；

( b ) 前記複数の巨大分子を、前記複数の巨大分子に結合することが可能な複数の結合性物質であって、各結合性物質が前記結合性物質に関する識別情報を有するコーディングタグを含む複数の結合性物質と接触させるステップと；

( c ) ( i ) 前記巨大分子に付随する記録タグの情報を前記巨大分子に結合した前記結合性物質の前記コーディングタグに移行させて、伸長コーディングタグを生成するステップ；または ( i i ) 巨大分子に付随する記録タグおよび前記巨大分子に結合した前記結合性物質のコーディングタグの情報をジタグ構築物に移行するステップと；

( d ) 前記伸長コーディングタグまたはジタグ構築物を収集するステップと；

( e ) 任意選択でステップ ( b ) ~ ( d ) を 1 回または複数回の結合サイクルにわたって繰り返すステップと；

( f ) 伸長コーディングタグまたはジタグ構築物の収集物を解析するステップとを含む方法。

( 項目 9 4 )

前記巨大分子がタンパク質である、項目 9 3 に記載の方法。

( 項目 9 5 )

前記巨大分子が、ペプチドである、項目 9 3 に記載の方法。

( 項目 9 6 )

前記ペプチドが、生体試料由来のタンパク質を断片化することによって得られる、項目 9 5 に記載の方法。

( 項目 9 7 )

前記記録タグが、DNA 分子、RNA 分子、PNA 分子、BNA 分子、XNA 分子、LNA 分子、PNA 分子、またはこれらの組合せである、項目 9 3 から 9 6 までのいずれか一項に記載の方法。

( 項目 9 8 )

前記記録タグが、一意の分子識別子 ( UMI ) を含む、項目 9 3 から 9 7 までのいずれか一項に記載の方法。

( 項目 9 9 )

前記記録タグが、コンパートメントタグを含む、項目 9 3 から 9 8 までに記載の方法。

( 項目 1 0 0 )

前記記録タグが、ユニバーサルプライミング部位を含む、項目 9 3 から 9 9 までのいずれか一項に記載の方法。

( 項目 1 0 1 )

前記記録タグが、その 3 ' 末端にスペーサーを含む、項目 9 3 から 1 0 0 までのいずれか一項に記載の方法。

( 項目 1 0 2 )

前記記録タグの 3 ' 末端をブロックしてポリメラーゼによる前記記録タグの伸長を防止し、巨大分子に付随する記録タグおよび前記巨大分子に結合している前記結合性物質のコーディングタグの情報をジタグ構築物に移行させる、項目 9 3 から 1 0 1 までのいずれか一項に記載の方法。

( 項目 1 0 3 )

前記コーディングタグが、エンコーダー配列を含む、項目 9 3 から 1 0 2 までのいずれか一項に記載の方法。

( 項目 1 0 4 )

前記コーディングタグが、UMI を含む、項目 9 3 から 1 0 3 までのいずれか一項に記

10

20

30

40

50

載の方法。

(項目 1 0 5)

前記コーディングタグが、ユニバーサルプライミング部位を含む、項目 9 3 から 1 0 4 までのいずれか一項に記載の方法。

(項目 1 0 6)

前記コーディングタグが、その 3' 末端にスパーサーを含む、項目 9 3 から 1 0 5 までのいずれか一項に記載の方法。

(項目 1 0 7)

前記コーディングタグが、結合サイクル特異的配列を含む、項目 9 3 から 1 0 6 までのいずれか一項に記載の方法。

(項目 1 0 8)

前記結合性物質と前記コーディングタグが、リンカーによって接合されている、項目 9 3 から 1 0 7 までのいずれか一項に記載の方法。

(項目 1 0 9)

前記記録タグの情報の前記コーディングタグへの移行が、プライマー伸長によってもたらされる、項目 9 3 から 1 0 8 までのいずれか一項に記載の方法。

(項目 1 1 0)

前記記録タグの情報の前記コーディングタグへの移行が、ライゲーションによってもたらされる、項目 9 3 から 1 0 8 までのいずれか一項に記載の方法。

(項目 1 1 1)

前記ジタグ構築物が、ギャップ充填、プライマー伸長、またはその両方によって生成される、項目 9 3 から 1 0 8 までのいずれか一項に記載の方法。

(項目 1 1 2)

前記ジタグ分子が、前記記録タグに由来するユニバーサルプライミング部位、前記記録タグに由来するコンパートメントタグ、前記記録タグに由来する一意の分子識別子、前記記録タグに由来する任意選択のスパーサー、前記コーディングタグに由来するエンコーダー配列、前記コーディングタグに由来する一意の分子識別子、前記コーディングタグに由来する任意選択のスパーサー、および前記コーディングタグに由来するユニバーサルプライミング部位を含む、項目 9 3 から 9 7 まで、1 0 7、1 0 8、および 1 1 1 のいずれか一項に記載の方法。

(項目 1 1 3)

前記巨大分子および前記付随する記録タグを、前記固体支持体に共有結合により接合させる、項目 9 3 から 1 1 2 までのいずれか一項に記載の方法。

(項目 1 1 4)

前記固体支持体が、ビーズ、多孔質ビーズ、多孔質マトリックス、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、ナイロン、シリコンウェハチップ、フロースルーチップ、信号変換電子機器を含むバイオチップ、マイクロタイターウェル、E L I S A プレート、スピン干渉ディスク、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアである、項目 1 1 3 に記載の方法。

(項目 1 1 5)

前記固体支持体が、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズである、項目 1 1 4 に記載の方法。

(項目 1 1 6)

前記結合性物質が、ポリペプチドまたはタンパク質である、項目 9 3 から 1 1 5 までのいずれか一項に記載の方法。

(項目 1 1 7)

前記結合性物質が、改変アミノペプチダーゼ、改変アミノアシル t R N A 合成酵素、改変アンチカリン、または抗体もしくはその結合性断片である、項目 1 1 6 に記載の方法。

10

20

30

40

50

( 項目 1 1 8 )

前記結合性物質が、単一のアミノ酸残基、ジペプチド、トリペプチドまたは前記ペプチドの翻訳後修飾に結合する、項目 9 5 から 1 1 7 までのいずれか一項に記載の方法。

( 項目 1 1 9 )

前記結合性物質が、N末端アミノ酸残基、C末端アミノ酸残基、または内部アミノ酸残基に結合する、項目 1 1 8 に記載の方法。

( 項目 1 2 0 )

前記結合性物質が、N末端ペプチド、C末端ペプチド、または内部ペプチドに結合する、項目 1 1 8 に記載の方法。

( 項目 1 2 1 )

前記結合性物質がN末端アミノ酸残基に結合し、前記N末端アミノ酸残基が各結合サイクル後に切断される、項目 1 1 9 に記載の方法。

( 項目 1 2 2 )

前記結合性物質がC末端アミノ酸残基に結合し、前記C末端アミノ酸残基が各結合サイクル後に切断される、項目 1 1 9 に記載の方法。

( 項目 1 2 3 )

N末端アミノ酸残基がエドマン分解によって切断される、項目 1 2 1 に記載の方法。

( 項目 1 2 4 )

前記結合性物質が、アミノ酸または翻訳後修飾の部位特異的な共有結合性標識である、項目 9 3 に記載の方法。

( 項目 1 2 5 )

ステップ ( b ) の後に、前記巨大分子および付随する結合性物質を含む複合体を前記固体支持体から解離させ、液滴またはマイクロ流体液滴のエマルジョン中に分配する、項目 9 3 から 1 2 4 までのいずれか一項に記載の方法。

( 項目 1 2 6 )

各マイクロ流体液滴が、平均して、前記巨大分子および前記結合性物質を含む複合体を 1 つ含む、項目 1 2 5 に記載の方法。

( 項目 1 2 7 )

伸長コーディングタグまたはジタグ構築物を生成する前に前記記録タグを増幅する、項目 1 2 5 または 1 2 6 に記載の方法。

( 項目 1 2 8 )

エマルジョン融合 P C R を使用して、前記記録タグ情報を前記コーディングタグに移行させる、またはジタグ構築物の集団を創出する、項目 1 2 5 から 1 2 7 までに記載の方法。

( 項目 1 2 9 )

伸長コーディングタグまたはジタグ構築物の収集物を、解析前に増幅させる、項目 9 3 から 1 2 8 までのいずれか一項に記載の方法。

( 項目 1 3 0 )

伸長コーディングタグまたはジタグ構築物の収集物の解析が、核酸配列決定法を含む、項目 9 3 から 1 2 9 までのいずれか一項に記載の方法。

( 項目 1 3 1 )

前記核酸配列決定法が、合成による配列決定、ライゲーションによる配列決定、ハイブリダイゼーションによる配列決定、ポロニーシーケンシング、イオン半導体シーケンシング、またはパイロシーケンシングである、項目 1 3 0 に記載の方法。

( 項目 1 3 2 )

前記核酸配列決定法が、単一分子リアルタイムシーケンシング、ナノポアに基づく配列決定、または先端顕微鏡を使用した D N A のダイレクトイメージングである、項目 1 3 0 に記載の方法。

( 項目 1 3 3 )

前記巨大分子の部分的組成を、一意のコンパートメントタグおよび任意選択で U M I を使用する複数の伸長コーディングタグまたはジタグ構築物の解析によって決定する、項目

10

20

30

40

50

1 3 0 に記載の方法。

( 項目 1 3 4 )

前記解析ステップを、塩基当たりのエラー率が > 5 %、> 1 0 %、> 1 5 %、> 2 0 %、> 2 5 %、または > 3 0 % である配列決定法を用いて実施する、項目 1 から 1 3 3 までのいずれか一項に記載の方法。

( 項目 1 3 5 )

コーディングタグ、記録タグ、またはその両方の識別成分が、エラー訂正コードを含む、項目 1 から 1 3 4 までのいずれか一項に記載の方法。

( 項目 1 3 6 )

前記識別成分が、エンコーダー配列、バーコード、U M I、コンパートメントタグ、サイクル特異的配列、またはそれらの任意の組合せから選択される、項目 1 3 5 に記載の方法。

10

( 項目 1 3 7 )

前記エラー訂正コードが、Hammingコード、Lee 距離コード、非対称 Lee 距離コード、Reed - Solomonコード、およびLevenshtein - Tene ngol tsコードから選択される、項目 1 3 5 または 1 3 6 に記載の方法。

( 項目 1 3 8 )

コーディングタグ、記録タグ、またはその両方の識別成分が、一意の電流またはイオンフラックスまたは光学的シグネチャを生成することが可能であり、前記解析ステップが、前記識別成分を識別するために前記一意の電流またはイオンフラックスまたは光学的シグネチャを検出することを含む、項目 1 から 1 3 4 までのいずれか一項に記載の方法。

20

( 項目 1 3 9 )

前記識別成分が、エンコーダー配列、バーコード、U M I、コンパートメントタグ、サイクル特異的配列、またはそれらの任意の組合せから選択される、項目 1 3 8 に記載の方法。

( 項目 1 4 0 )

複数の巨大分子を解析するための方法であって、

( a ) 固体支持体に接合した複数の巨大分子および付随する記録タグを用意するステップと；

( b ) 前記複数の巨大分子を、同類の巨大分子に結合することが可能な複数の結合性物質であって、各結合性物質が前記結合性物質に関する識別情報を有するコーディングタグを含む複数の結合性物質と接触させるステップと；

30

( c ) 第 1 の結合性物質の第 1 のコーディングタグの情報を第 1 の巨大分子に付随する第 1 の記録タグに移行させて、一次伸長記録タグを生成するステップであって、前記第 1 の結合性物質が前記第 1 の巨大分子に結合するステップと；

( d ) 前記複数の巨大分子を、同類の巨大分子に結合することが可能な複数の結合性物質と接触させるステップと；

( e ) 第 2 の結合性物質の第 2 のコーディングタグの情報を前記一次伸長記録タグに移行させて、二次伸長記録タグを生成するステップであって、前記第 2 の結合性物質が前記第 1 の巨大分子に結合するステップと；

40

( f ) 任意選択でステップ ( d ) ~ ( e ) を「 n 」回の結合サイクルにわたって繰り返すステップであって、前記第 1 の巨大分子に結合する各結合性物質の各コーディングタグの情報を前の結合サイクルで生成した伸長記録タグに移行させて、前記第 1 の巨大分子を表す n 次伸長記録タグを生成するステップと；

( g ) 前記 n 次伸長記録タグを解析するステップとを含む方法。

( 項目 1 4 1 )

複数の巨大分子を表す複数の n 次伸長記録タグを生成し、解析する、項目 1 4 0 に記載の方法。

( 項目 1 4 2 )

50

前記巨大分子が、タンパク質である、項目 1 4 0 または 1 4 1 に記載の方法。

( 項目 1 4 3 )

前記巨大分子が、ペプチドである、項目 1 4 2 に記載の方法。

( 項目 1 4 4 )

前記ペプチドが、生体試料由来のタンパク質を断片化することによって得られる、項目 1 4 3 に記載の方法。

( 項目 1 4 5 )

前記複数の巨大分子が、多数のプールされた試料由来の巨大分子を含む、項目 1 4 0 から 1 4 4 までのいずれか一項に記載の方法。

( 項目 1 4 6 )

前記記録タグが、DNA 分子、RNA 分子、BNA 分子、XNA 分子、LNA 分子、PNA 分子、PNNA 分子、またはこれらの組合せである、項目 1 4 0 から 1 4 5 までのいずれか一項に記載の方法。

( 項目 1 4 7 )

前記記録タグが、一意の分子識別子 (UMI) を含む、項目 1 4 0 から 1 4 6 までに記載の方法。

( 項目 1 4 8 )

前記記録タグが、コンパートメントタグを含む、項目 1 4 0 から 1 4 7 までに記載の方法。

( 項目 1 4 9 )

前記記録タグが、ユニバーサルプライミング部位を含む、項目 1 4 0 から 1 4 8 までのいずれか一項に記載の方法。

( 項目 1 5 0 )

前記記録タグが、その 3' 末端にスペーサーを含む、項目 1 4 0 から 1 4 9 までのいずれか一項に記載の方法。

( 項目 1 5 1 )

前記コーディングタグが、エンコーダー配列を含む、項目 1 4 0 から 1 5 0 までのいずれか一項に記載の方法。

( 項目 1 5 2 )

前記コーディングタグが、UMI を含む、項目 1 4 0 から 1 5 1 までのいずれか一項に記載の方法。

( 項目 1 5 3 )

前記コーディングタグが、ユニバーサルプライミング部位を含む、項目 1 4 0 から 1 5 2 までのいずれか一項に記載の方法。

( 項目 1 5 4 )

前記コーディングタグが、その 3' 末端にスペーサーを含む、項目 1 4 0 から 1 5 3 までのいずれか一項に記載の方法。

( 項目 1 5 5 )

前記コーディングタグが、結合サイクル特異的配列を含む、項目 1 4 0 から 1 5 4 までのいずれか一項に記載の方法。

( 項目 1 5 6 )

前記コーディングタグが、一意の分子識別子を含む、項目 1 4 0 から 1 5 5 までのいずれか一項に記載の方法。

( 項目 1 5 7 )

前記結合性物質と前記コーディングタグが、リンカーによって接合されている、項目 1 4 0 から 1 5 6 までのいずれか一項に記載の方法。

( 項目 1 5 8 )

前記記録タグの情報の前記コーディングタグへの移行が、プライマー伸長によって媒介される、項目 1 4 0 から 1 5 7 までのいずれか一項に記載の方法。

( 項目 1 5 9 )

10

20

30

40

50

前記記録タグの情報の前記コーディングタグへの移行が、ライゲーションによって媒介される、項目140から158までのいずれか一項に記載の方法。

(項目160)

前記複数の巨大分子、前記付随する記録タグ、またはその両方が、前記固体支持体に共有結合により接合している、項目140から159までのいずれか一項に記載の方法。

(項目161)

前記固体支持体が、ビーズ、多孔質ビーズ、多孔質マトリックス、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、ナイロン、シリコンウェーハチップ、フロースルーチップ、信号変換電子機器を含むバイオチップ、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、ニトロセルロースメンブレン、ニトロセルロースに基づくポリマー表面、ナノ粒子、またはマイクロスフェアである、項目140から160までのいずれか一項に記載の方法。

(項目162)

前記固体支持体が、ポリスチレンビーズ、ポリマービーズ、アガロースビーズ、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、または制御ポアビーズである、項目161に記載の方法。

(項目163)

前記結合性物質が、ポリペプチドまたはタンパク質である、項目140から162までのいずれか一項に記載の方法。

(項目164)

前記結合性物質が、改変アミノペプチダーゼ、改変アミノアシルtRNA合成酵素、改変アンチカリン、または抗体もしくはその結合性断片である、項目163に記載の方法。

(項目165)

前記結合性物質が、単一のアミノ酸残基、ジペプチド、トリペプチドまたは前記ペプチドの翻訳後修飾に結合する、項目142から164までのいずれか一項に記載の方法。

(項目166)

前記結合性物質が、N末端アミノ酸残基、C末端アミノ酸残基、または内部アミノ酸残基に結合する、項目165に記載の方法。

(項目167)

前記結合性物質が、N末端ペプチド、C末端ペプチド、または内部ペプチドに結合する、項目165に記載の方法。

(項目168)

前記結合性物質が、修飾されたN末端アミノ酸残基、修飾されたC末端アミノ酸残基、または修飾された内部アミノ酸残基の化学標識に結合する、項目142から164までのいずれか一項に記載の方法。

(項目169)

前記結合性物質がN末端アミノ酸残基または前記修飾されたN末端アミノ酸残基の化学標識に結合し、前記N末端アミノ酸残基が各結合サイクル後に切断される、項目166または168に記載の方法。

(項目170)

前記結合性物質がC末端アミノ酸残基または前記修飾されたC末端アミノ酸残基に化学標識に結合し、前記C末端アミノ酸残基が各結合サイクル後に切断される、項目166または168に記載の方法。

(項目171)

前記N末端アミノ酸残基が、エドマン分解、エドマナーゼ、改変アミノペプチダーゼ、または改変アシルペプチドヒドロラーゼによって切断される、項目169に記載の方法

(項目172)

前記結合性物質が、アミノ酸または翻訳後修飾の部位特異的な共有結合性標識である、項目163に記載の方法。

(項目173)

10

20

30

40

50





【図 2 - 1】

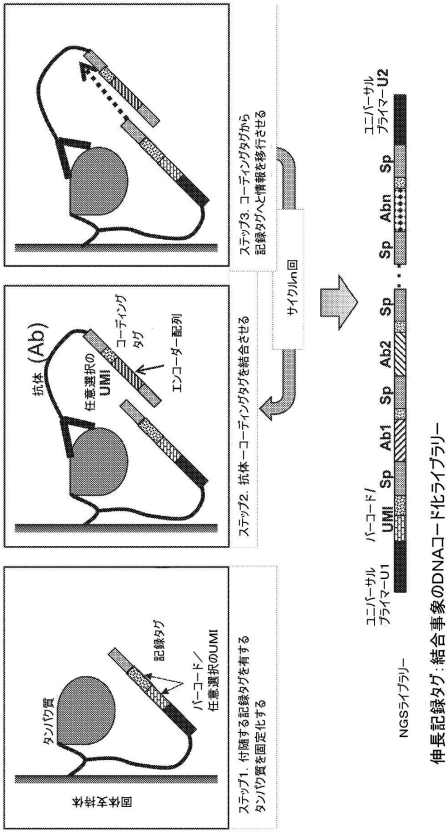


Fig. 2A

【図 2 - 2】

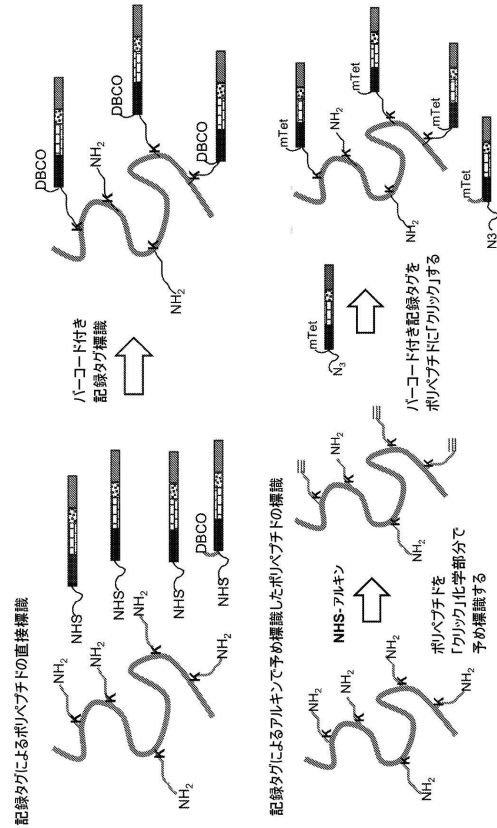


Fig. 2B

【図 2 - 3】

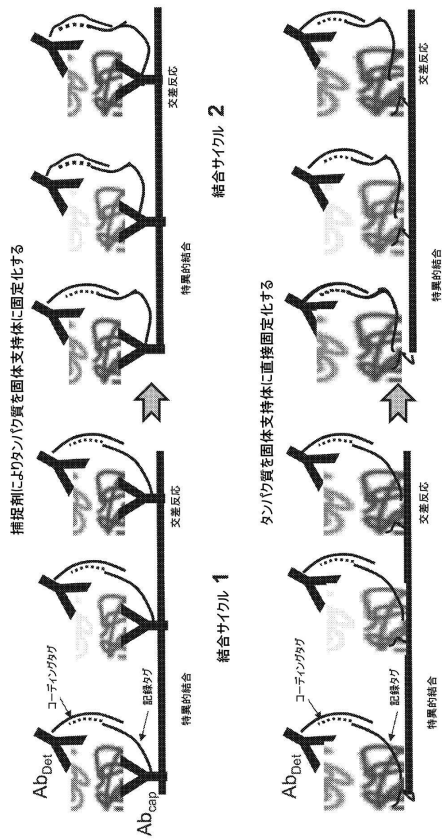


Fig. 2C

【図 2 - 4】

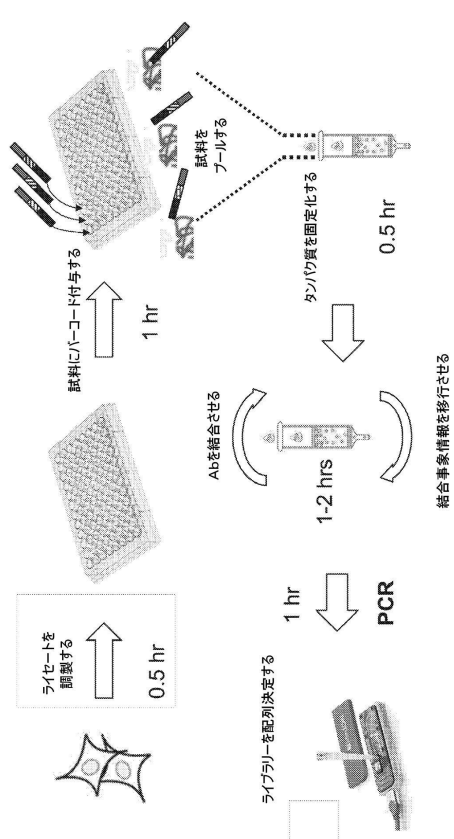
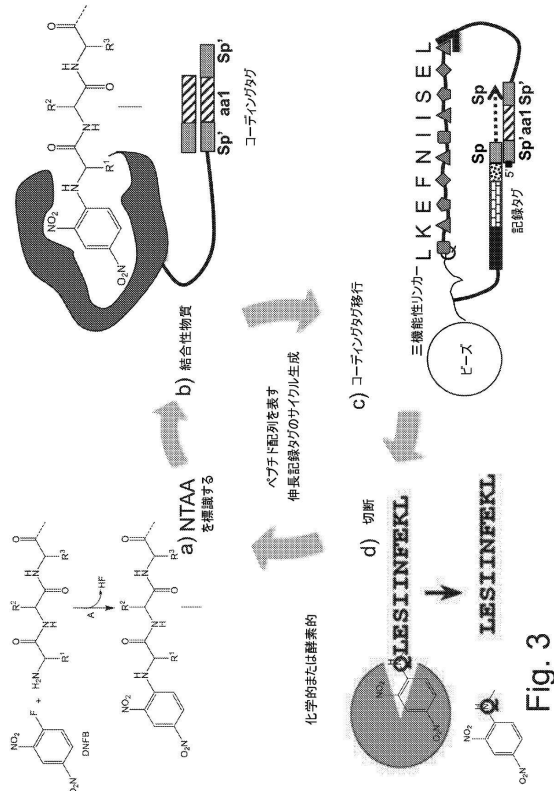
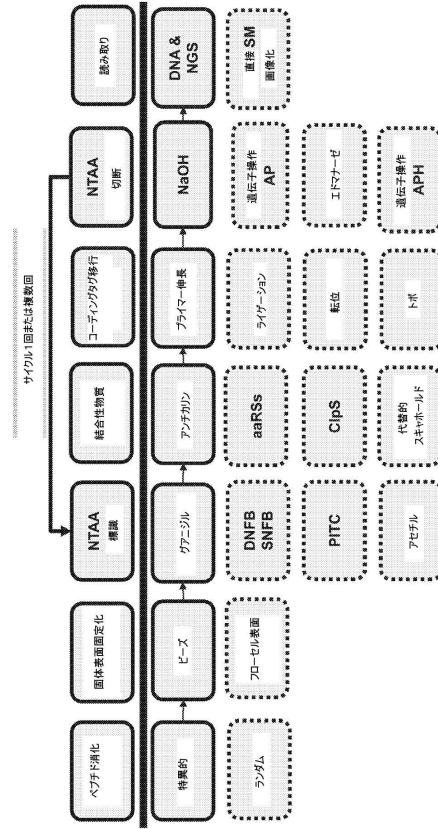


Fig. 2D

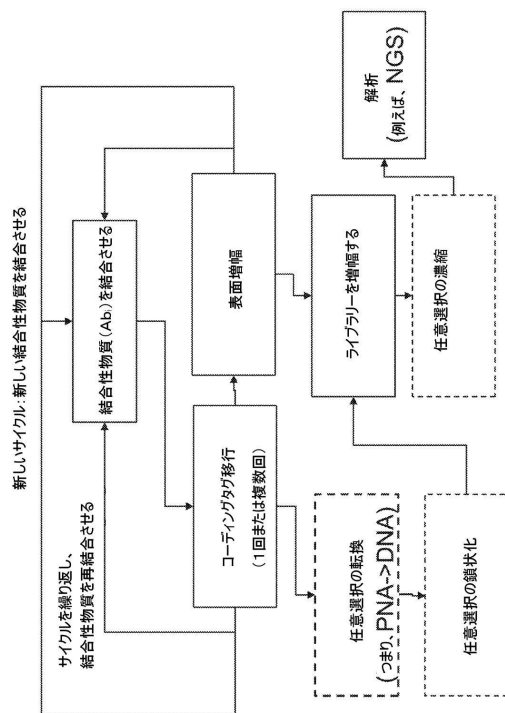
【図 3】



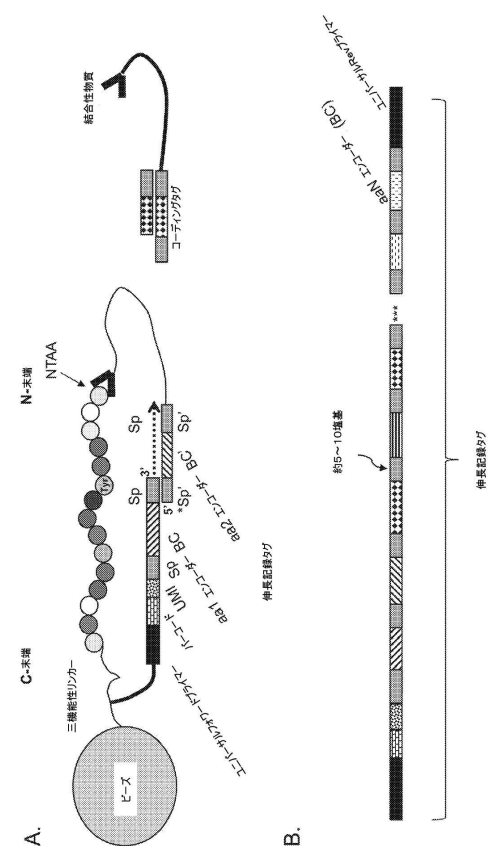
【図 4 - 1】



【図 4 - 2】



【図 5】



10

20

30

40

50

【図 6】

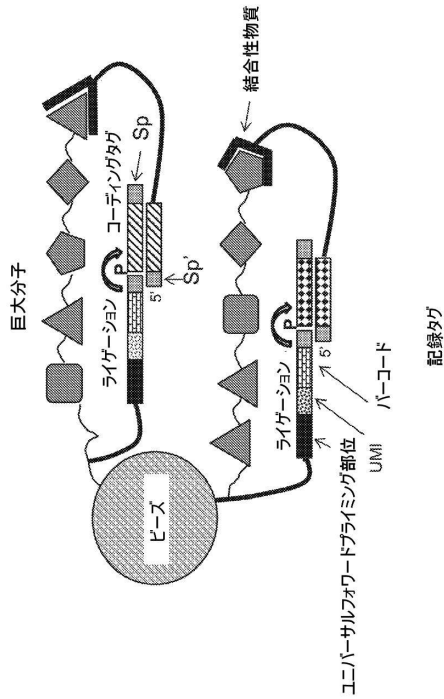


Fig. 6

【図 7】

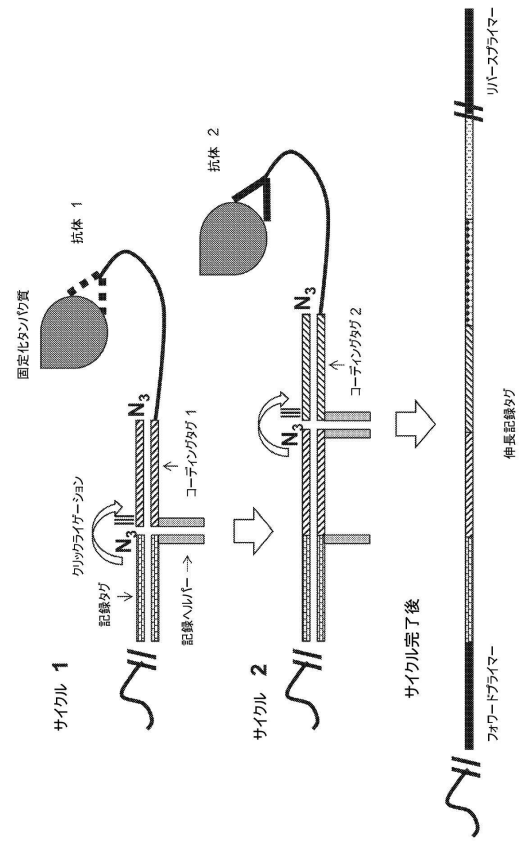


Fig. 7

【図 8】

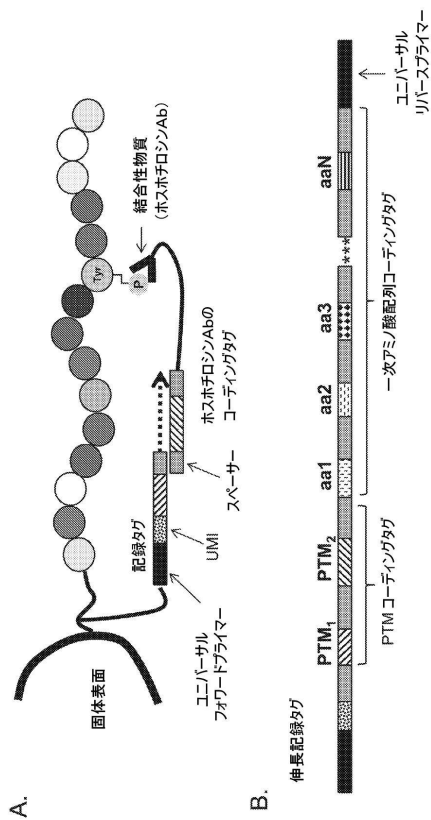


Fig. 8

【図 9】

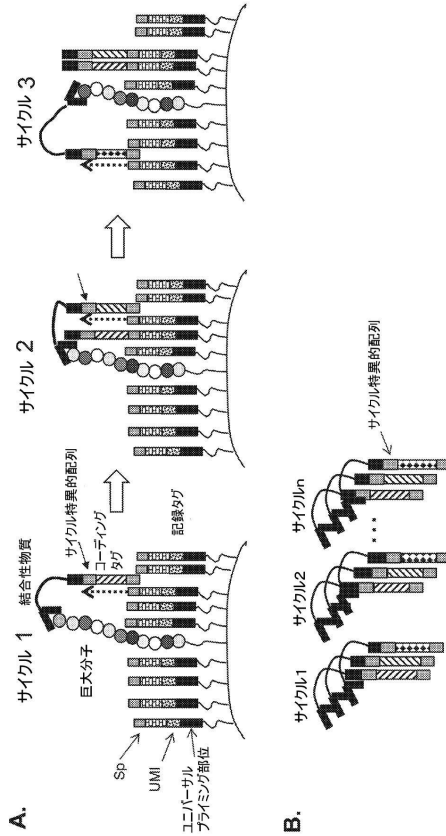


Fig. 9

【図 10】

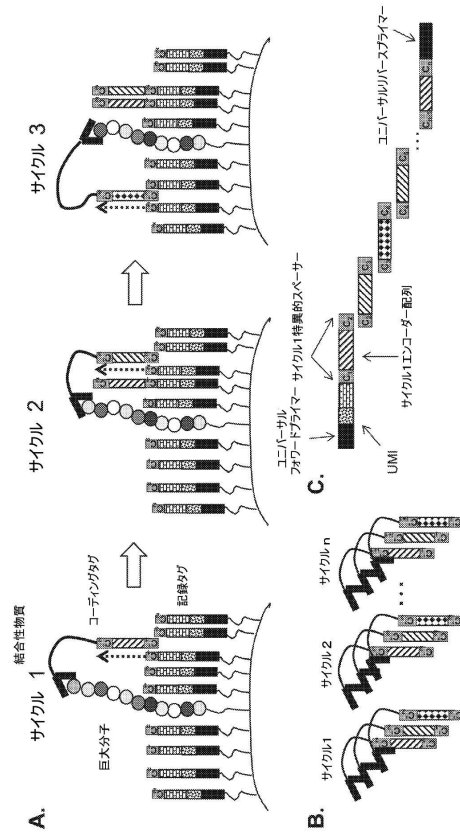


Fig. 10

【図 11】

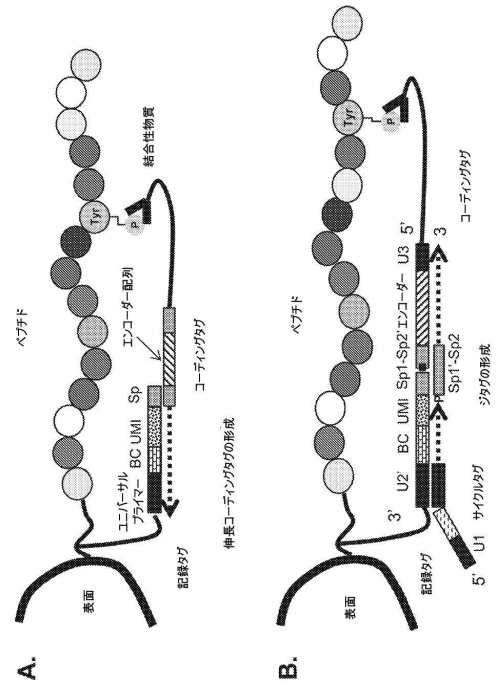


Fig. 11

【図 12】

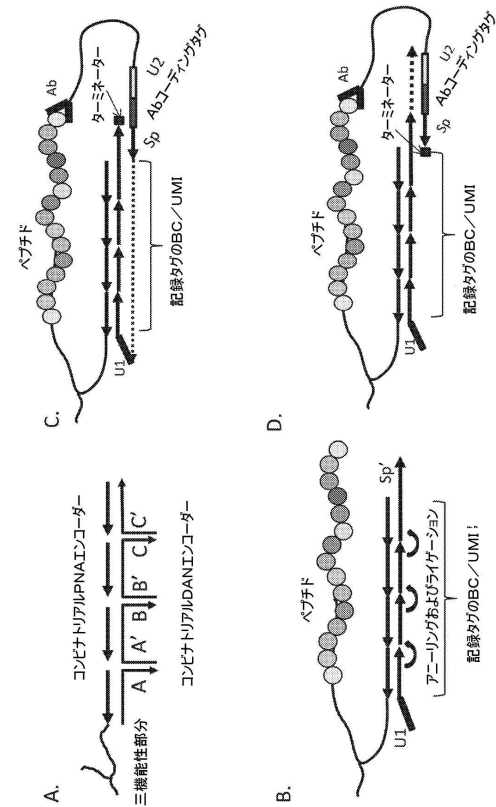


Fig. 12

【図 13】

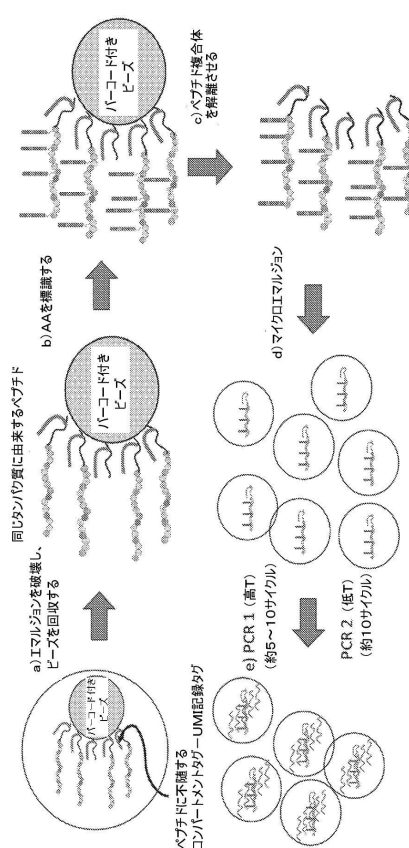


Fig. 13

10

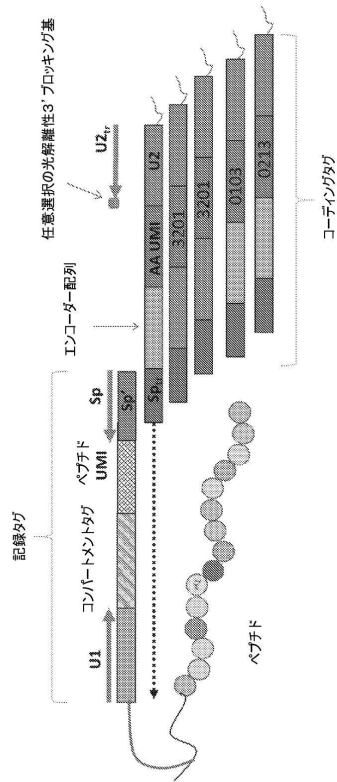
20

30

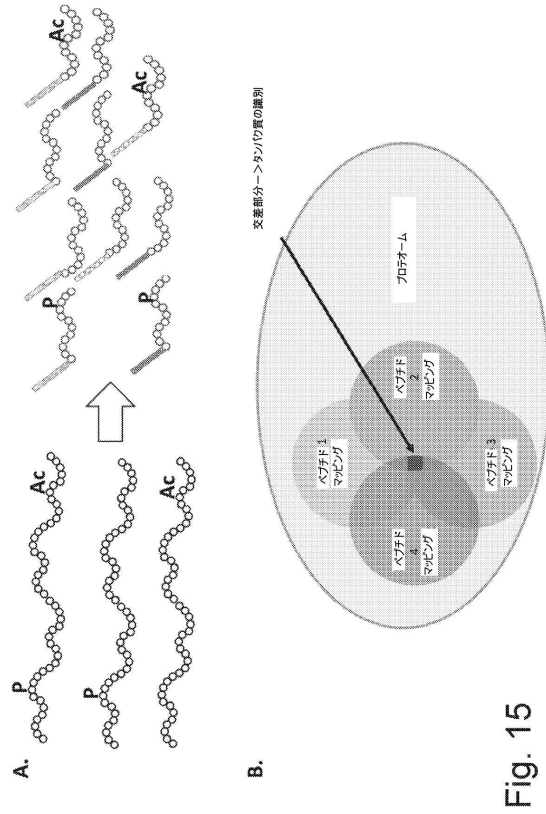
40

50

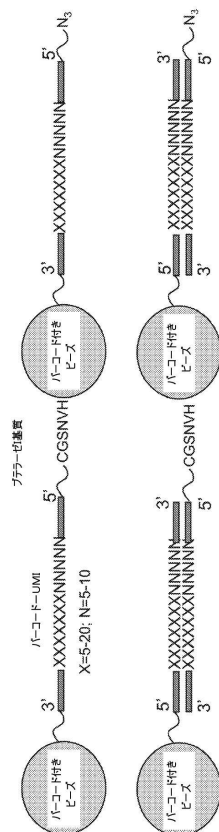
【 図 1 4 】



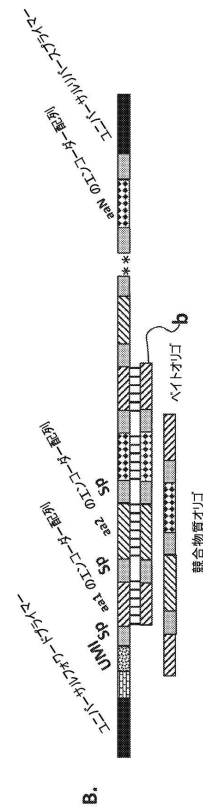
【 図 1 5 】



【 図 1 6 】



【 図 1 7 】



10

20

30

40

50

【図 18】

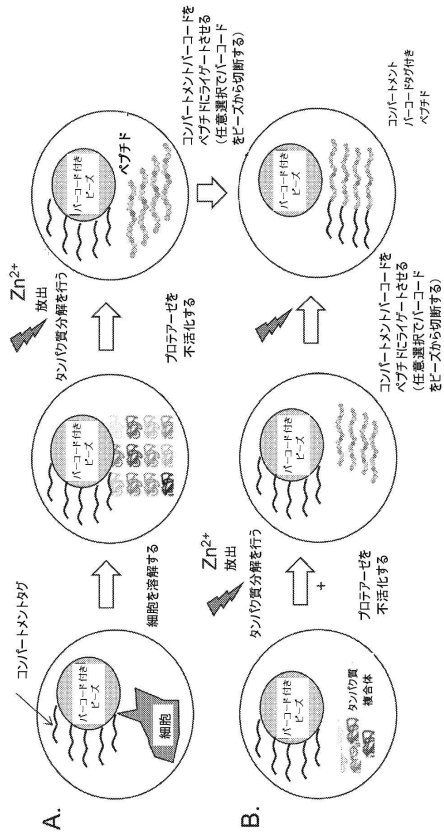


Fig. 18

【図 19】

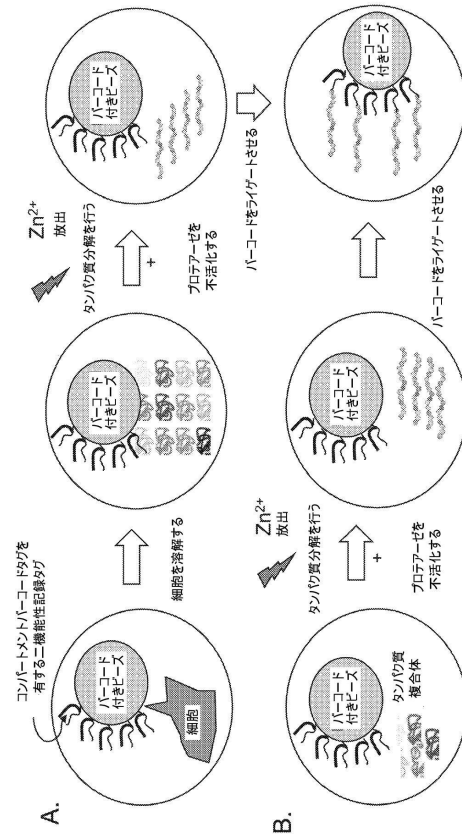


Fig. 19

【図 20 - 1】

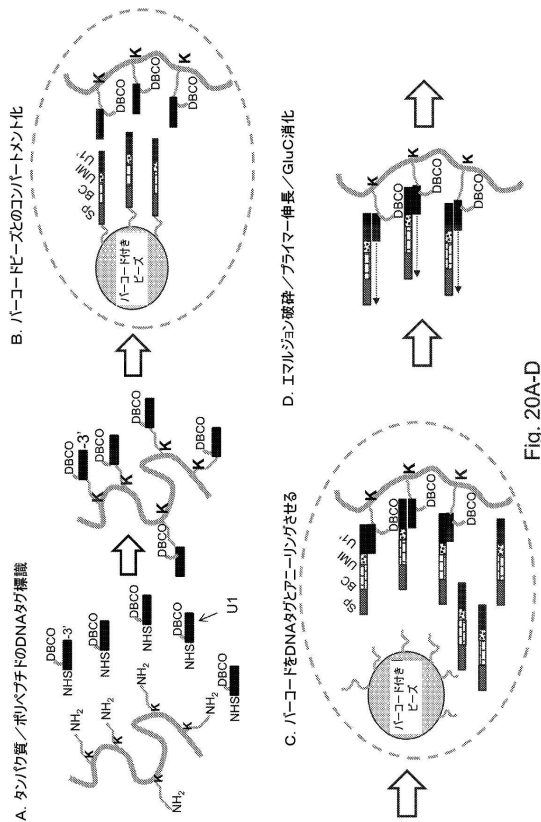


Fig. 20A-D

【図 20 - 2】

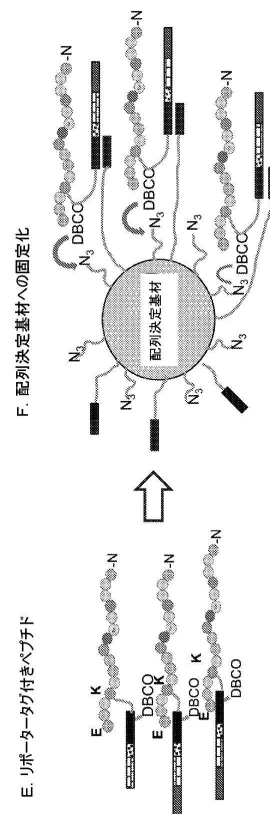


Fig. 20E-F

10

20

30

40

50

【図 20 - 3】

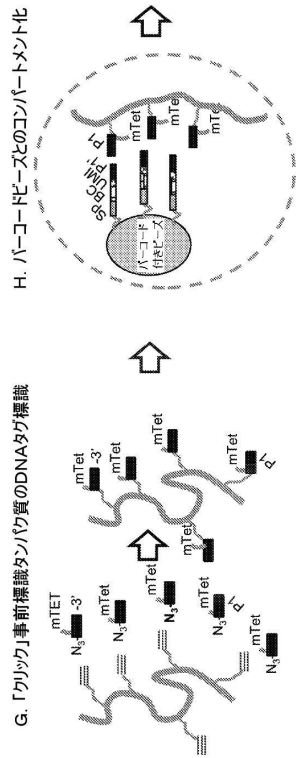


Fig. 20G-H

【図 20 - 4】

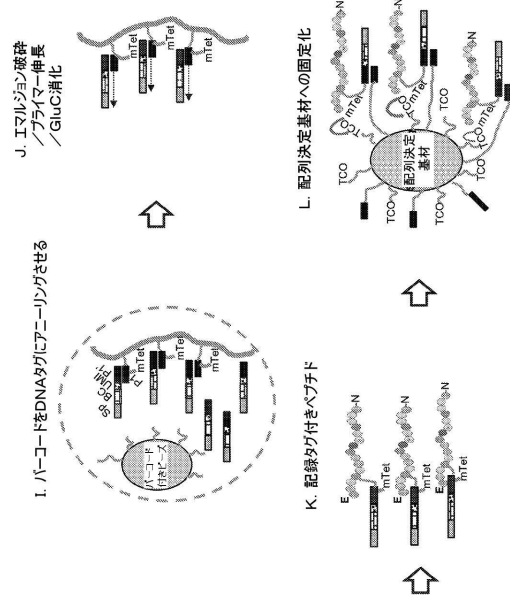


Fig. 20I-L

【図 21】

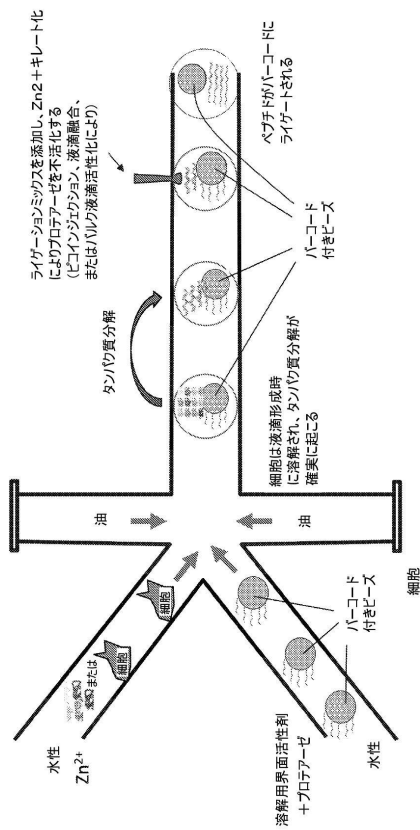


Fig. 21

【図 22】

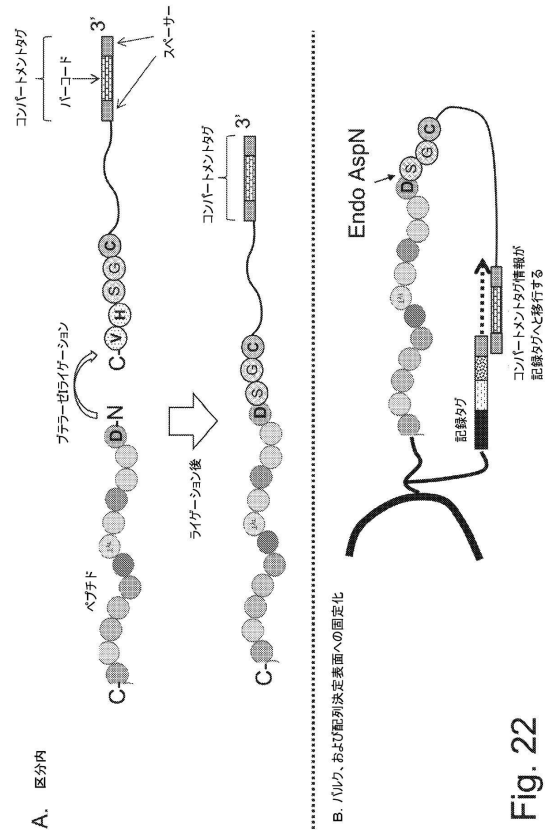


Fig. 22

10

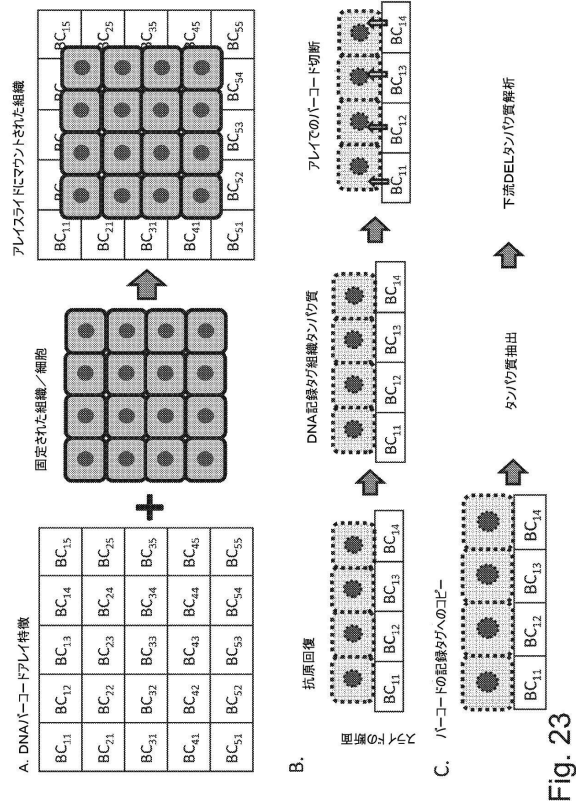
20

30

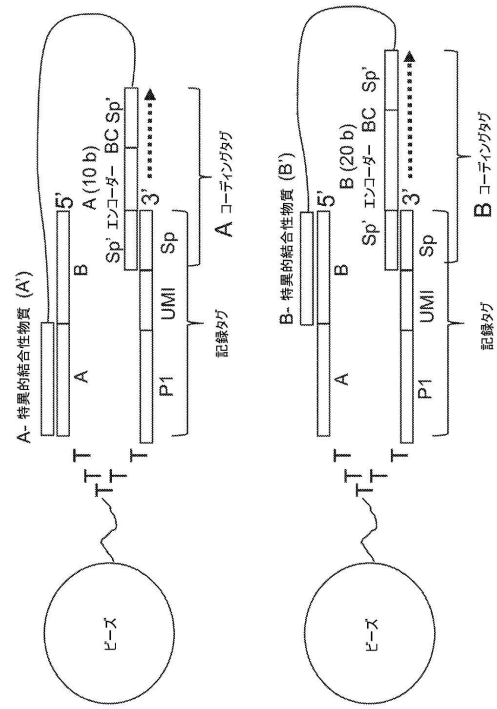
40

50

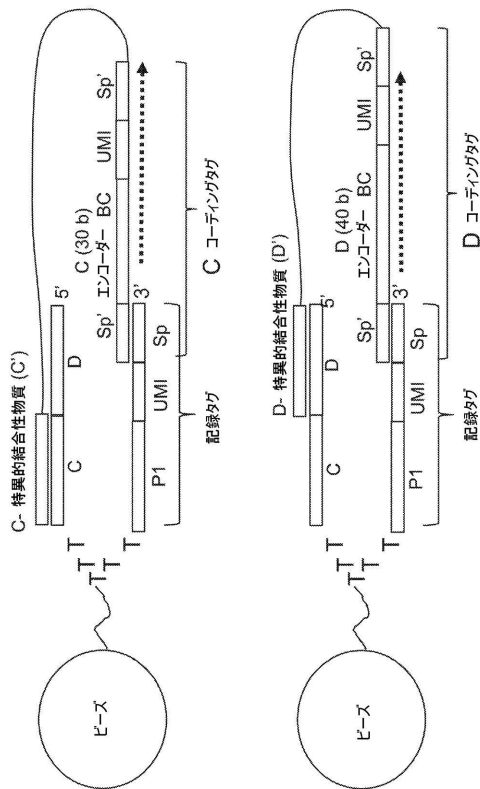
【図 2 3】



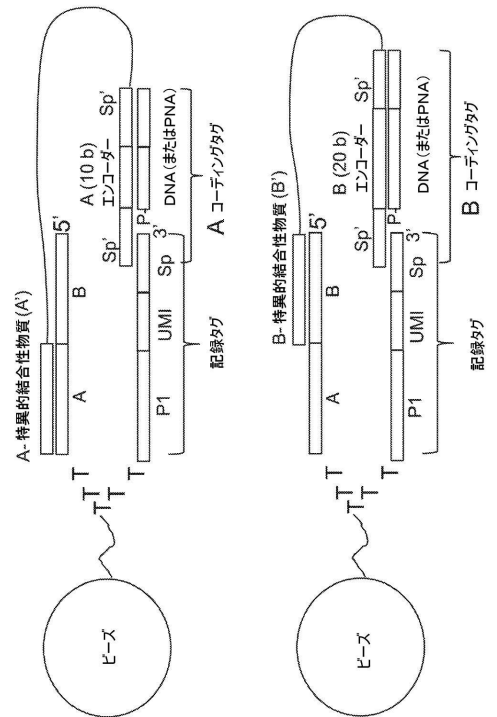
【図 2 4 - 1】



【図 2 4 - 2】



【図 2 5】



10

20

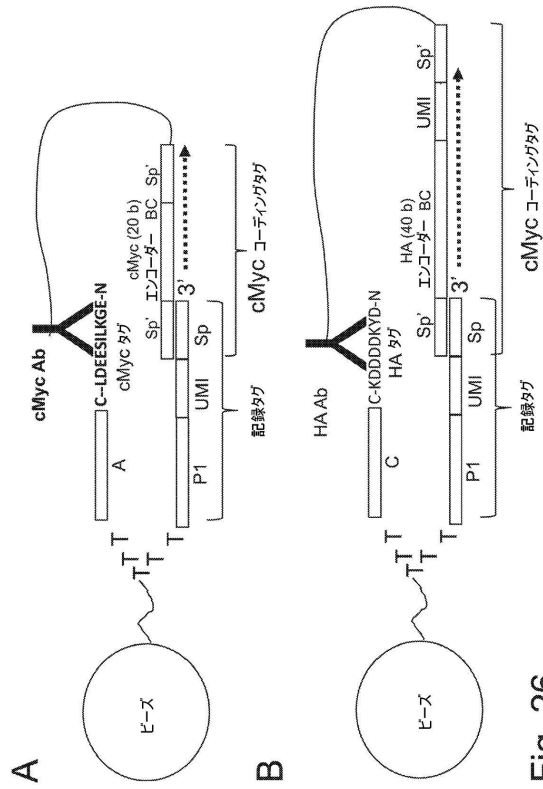
30

40

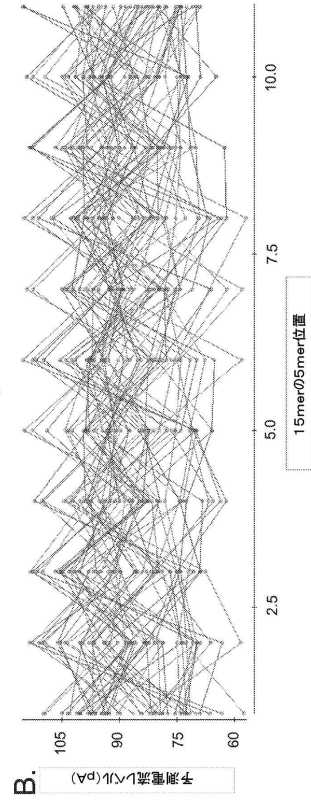
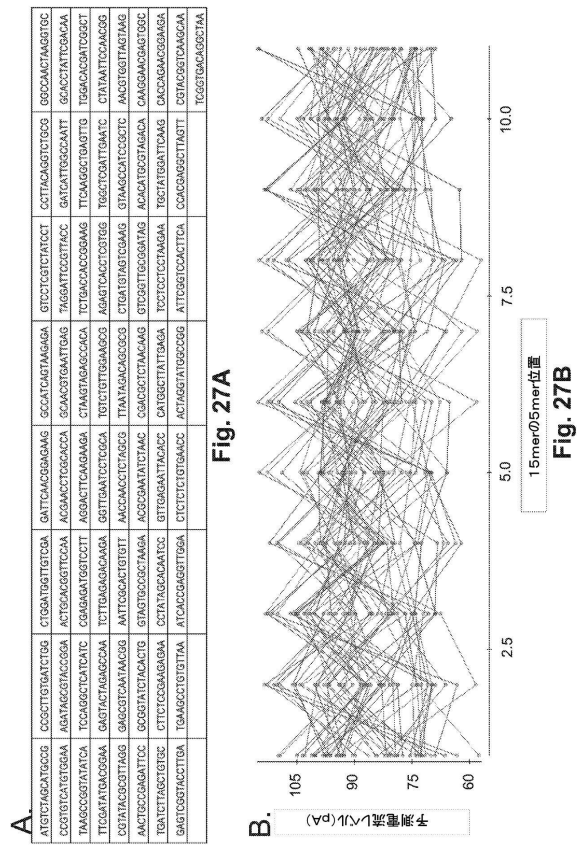
50



【 図 2 6 】

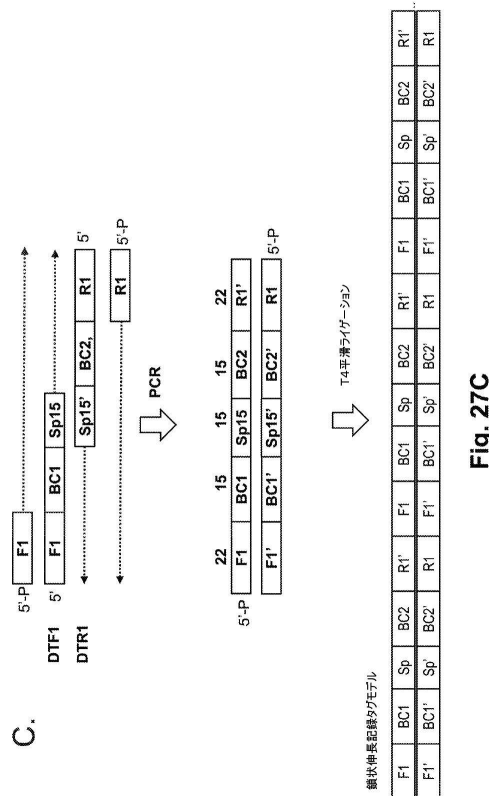


【 図 2 7 - 1 】



【 図 2 7 - 2 】

【 図 2 7 - 3 】

[illegible]

【図 28】

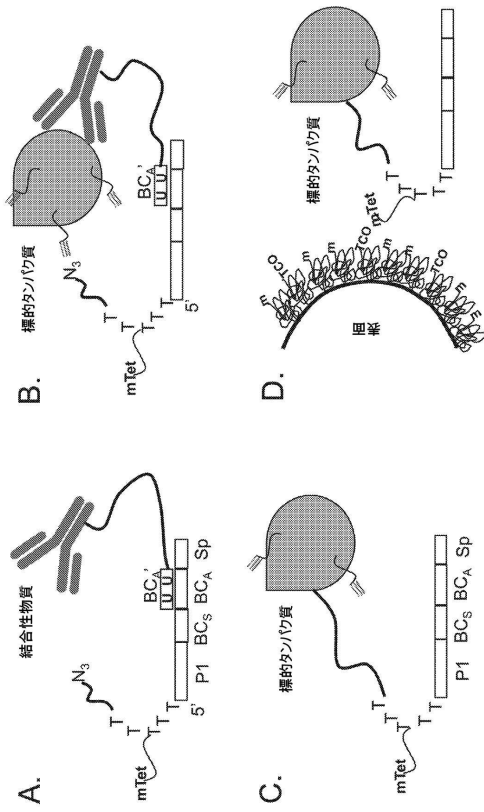


Fig. 28

【図 29】

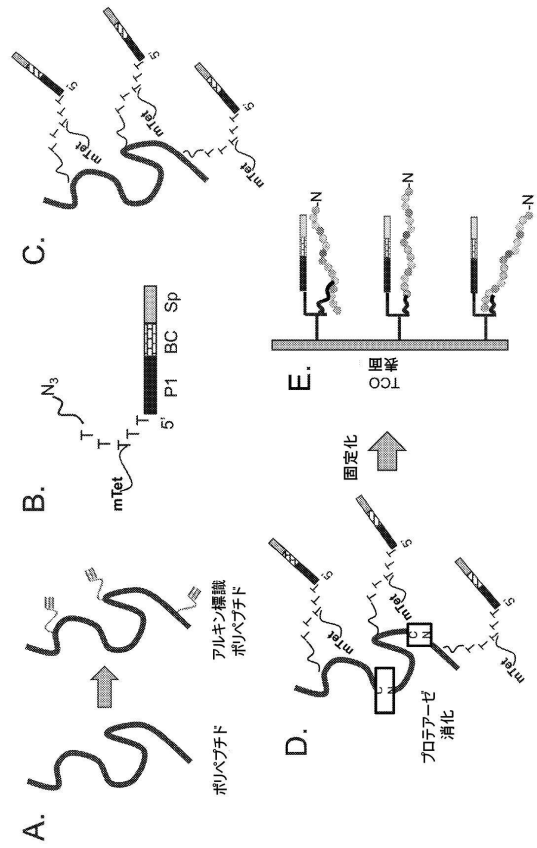


Fig. 29

【図 30】

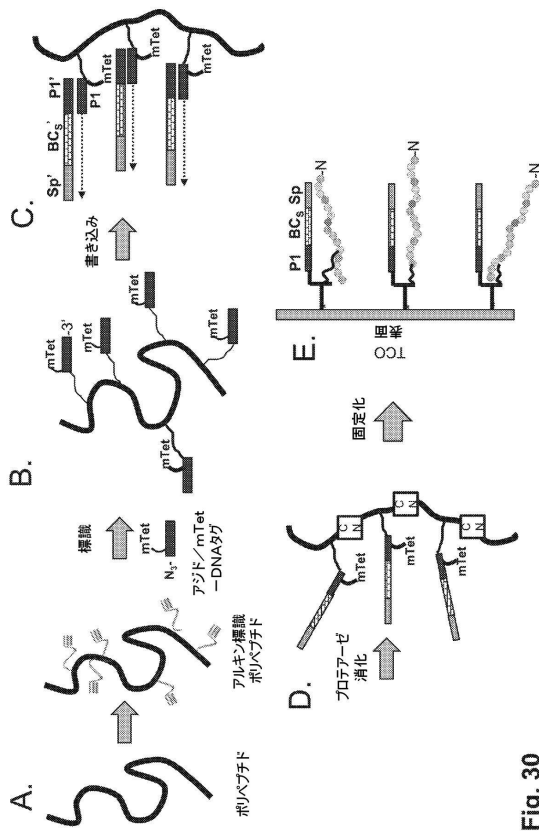


Fig. 30

【図 31】

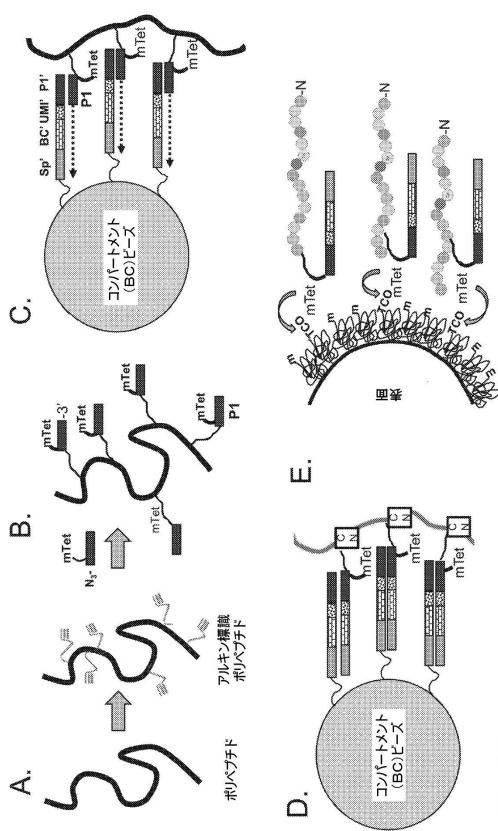


Fig. 31

10

20

30

40

50

【図 3 2 - 1】

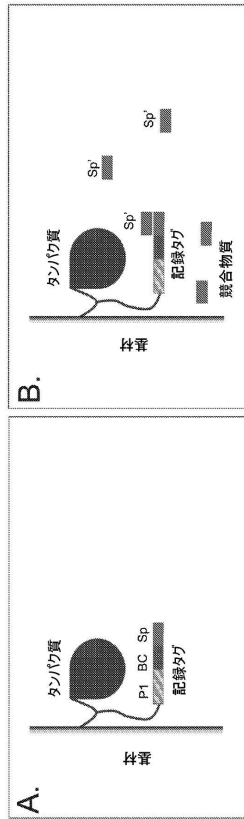


Fig. 32B

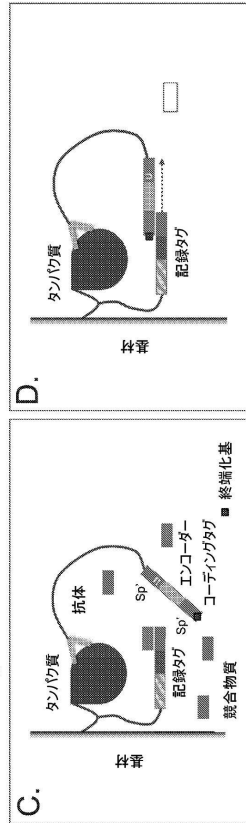


Fig. 32D

Fig. 32C

【図 3 2 - 2】

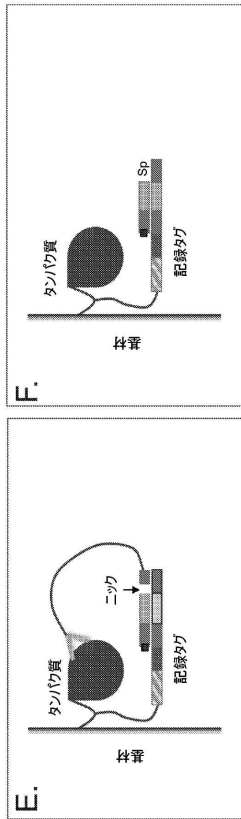


Fig. 32F

Fig. 32E

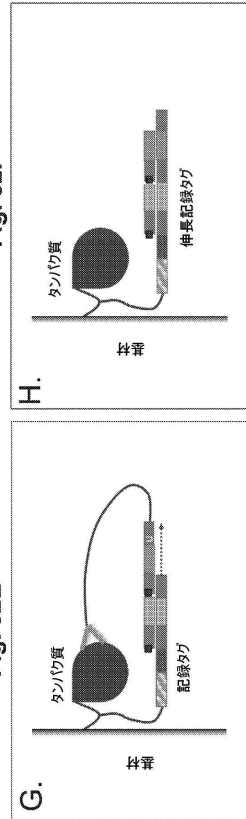


Fig. 32H

Fig. 32G

【図 3 3】

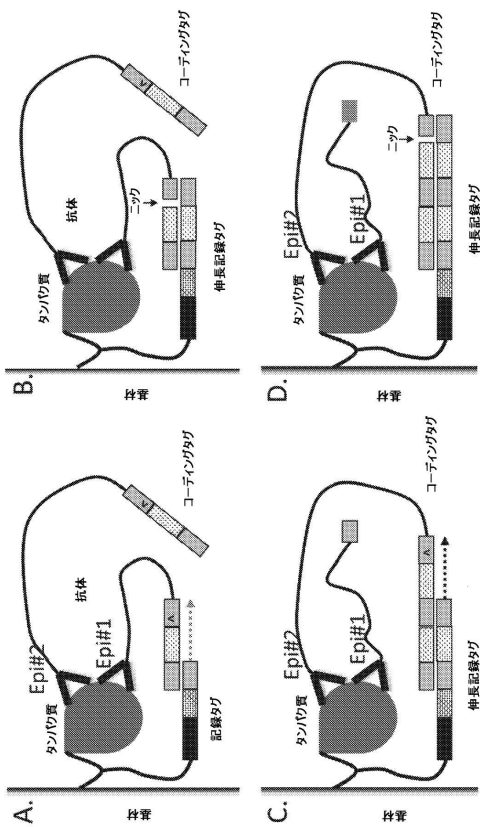


Fig. 33

【図 3 4】

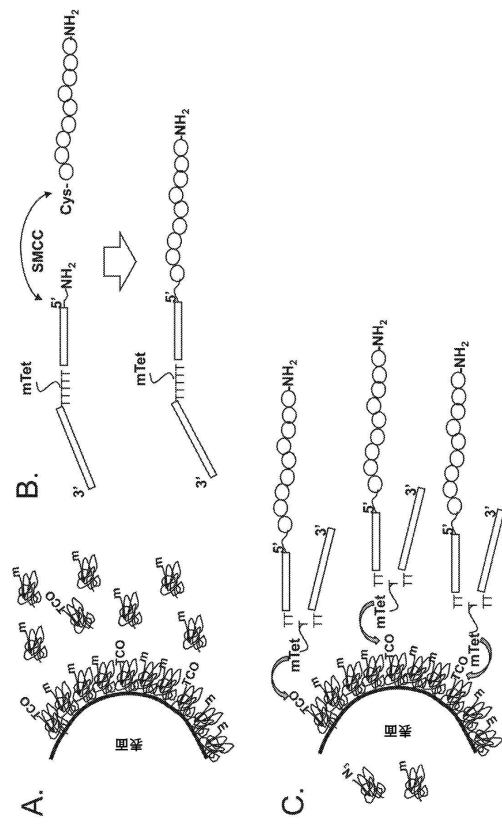
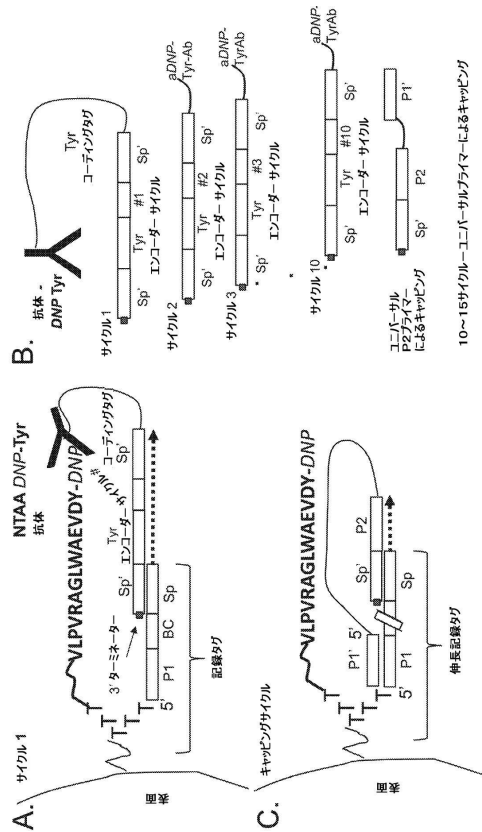
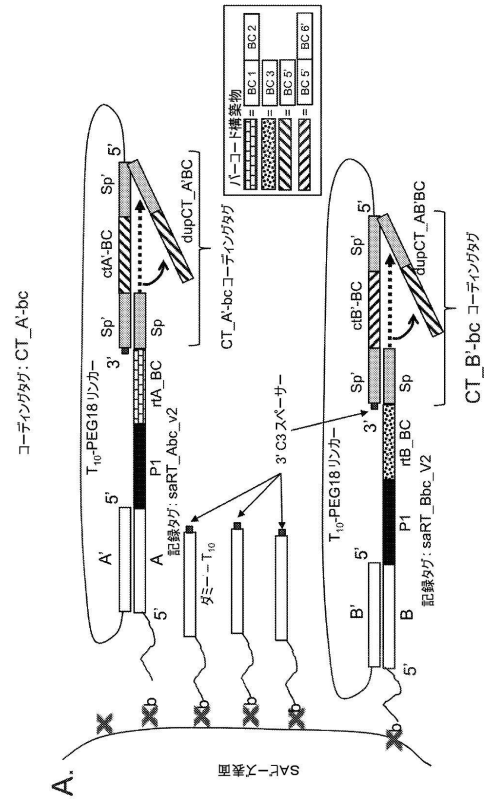


Fig. 34

【 図 3 5 】

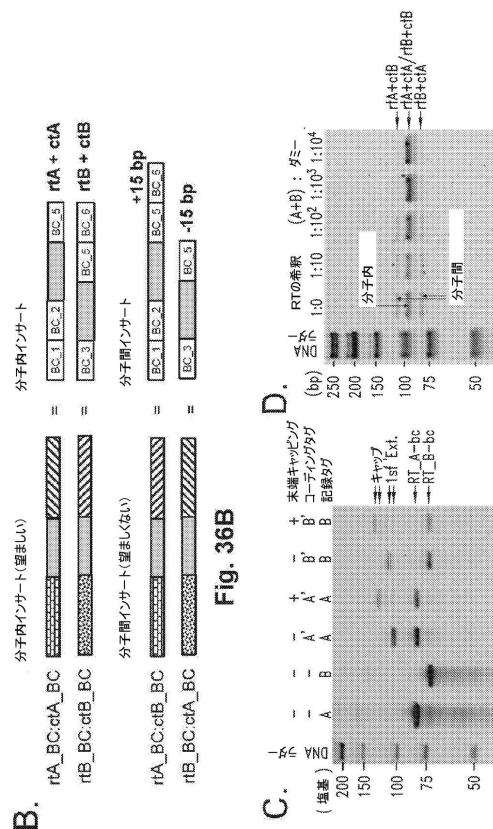


【 図 3 6 - 1 】

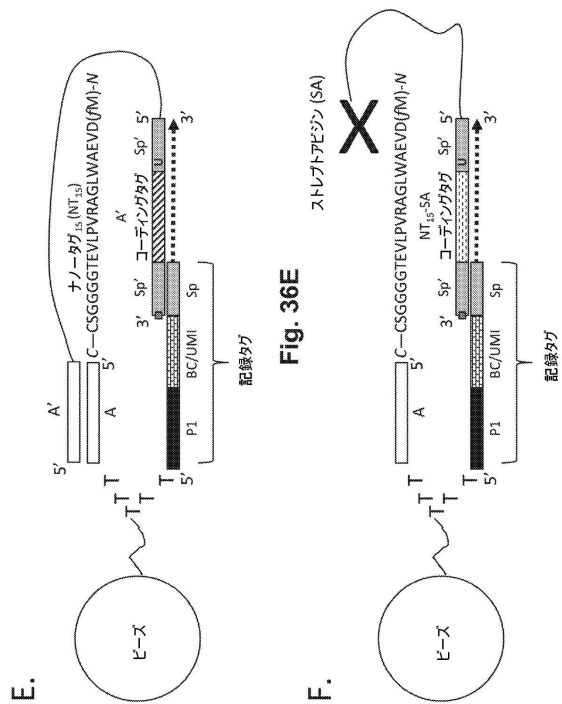


**Fig. 36A**

【 図 3 6 - 2 】



【 図 3 6 - 3 】



**Fig. 36F**

10

20

30

40

50

【図 37】

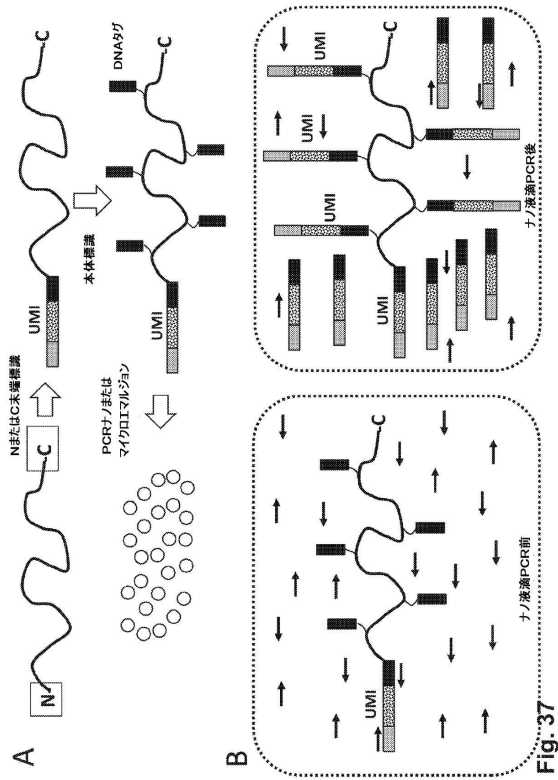


Fig. 37

【図 38】

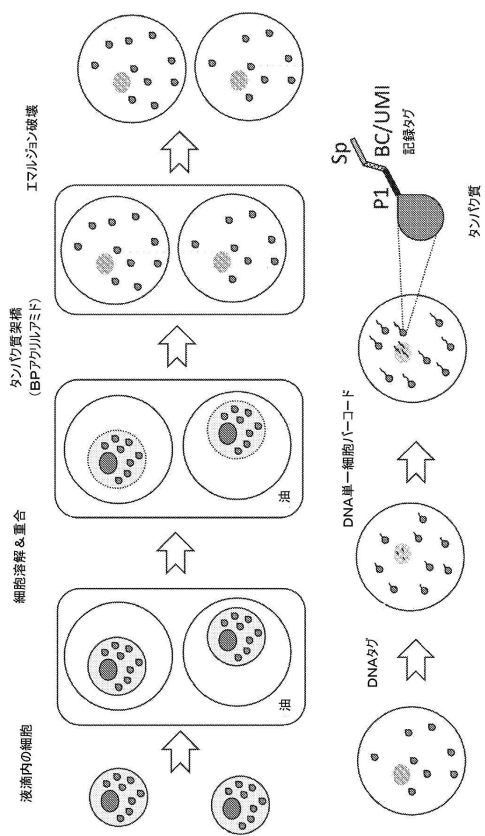


Fig. 38

【図 39】

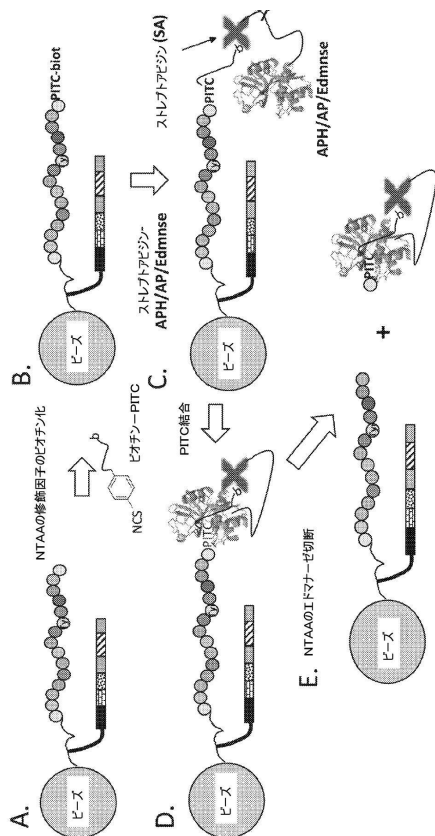
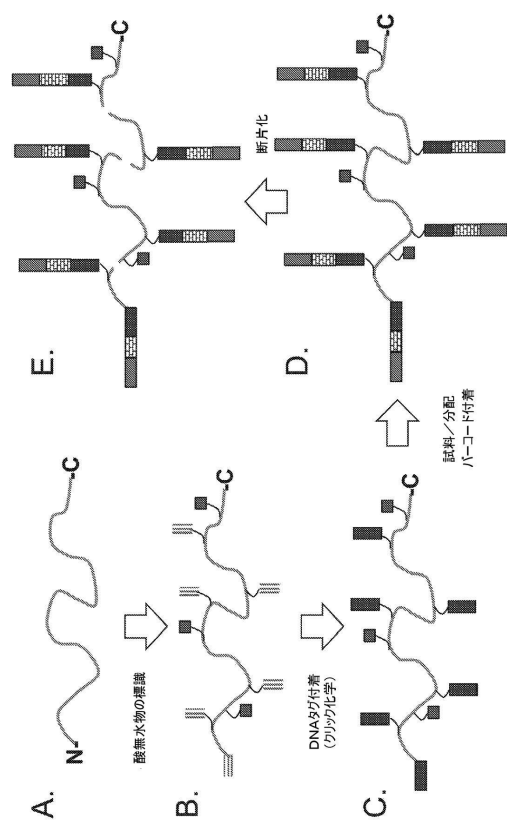


Fig. 39

【図 40 - 1】



Figs. 40A-D

10

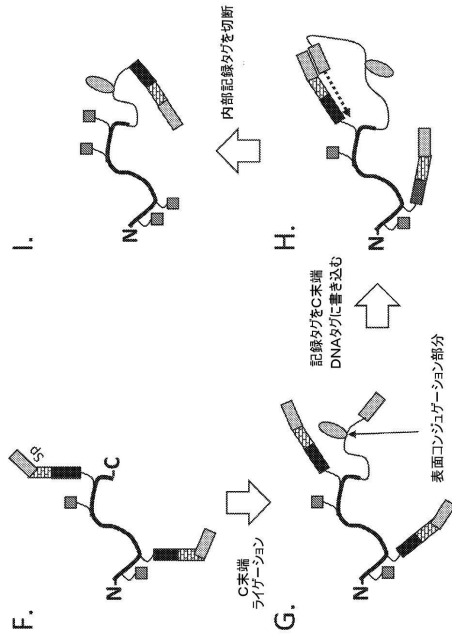
20

30

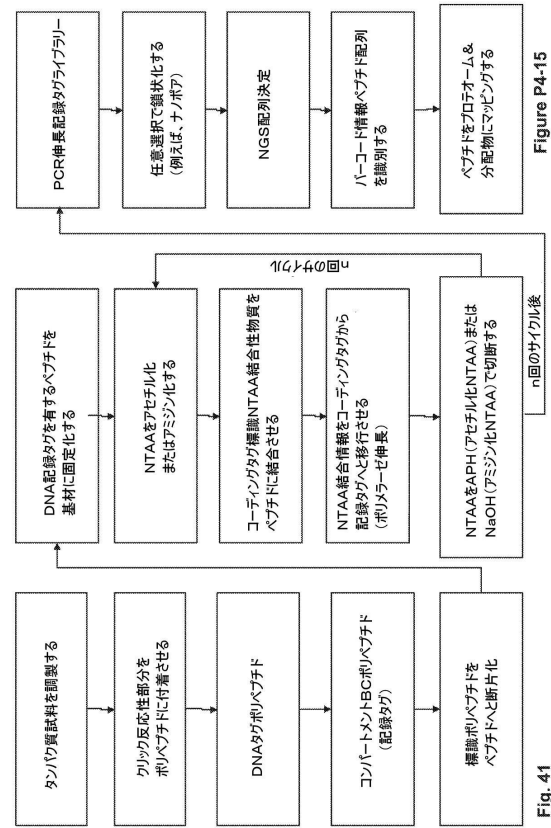
40

50

【図 40 - 2】



【図 41】



Figs. 40F-H

Fig. 41

【図 42】

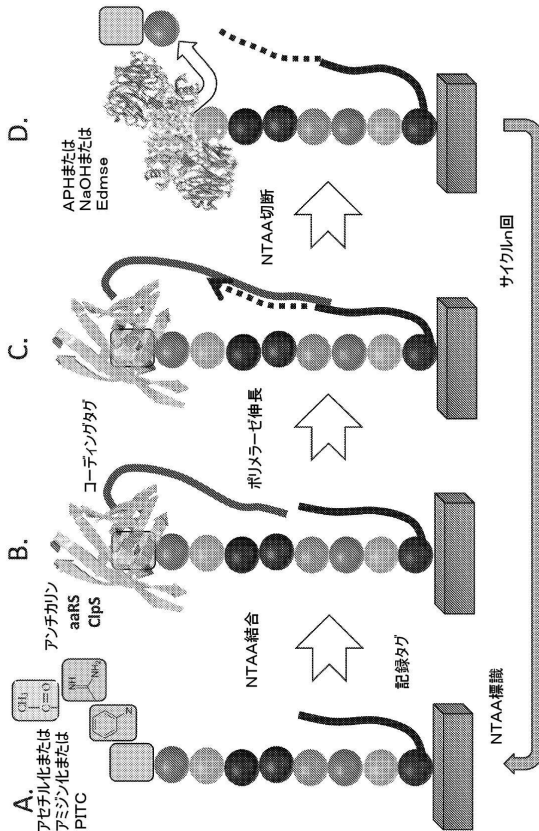


Fig. 42

【図 43】

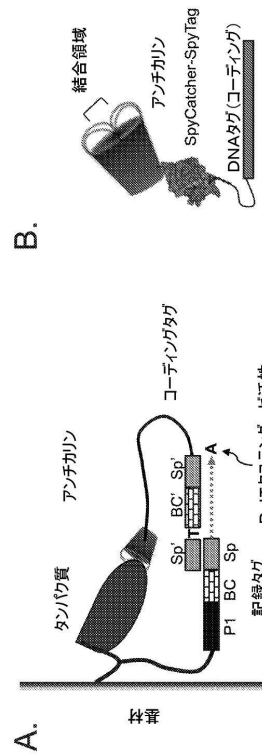


Fig. 43

10

20

30

40

50

【図 4 4】

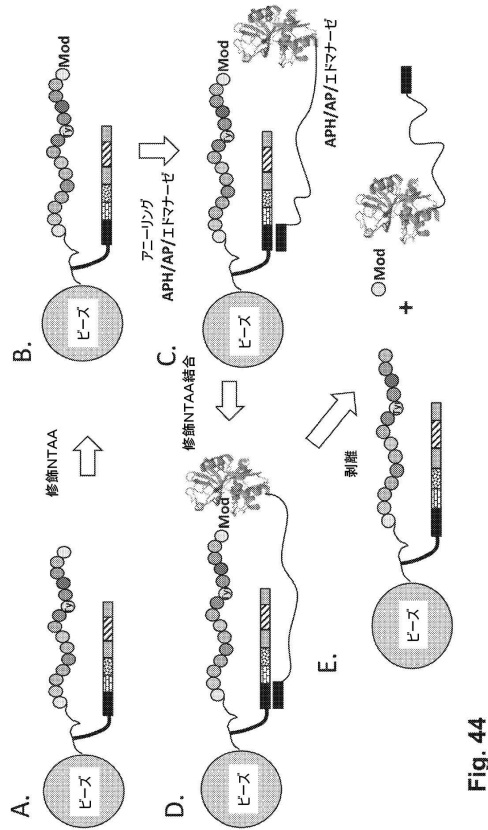


Fig. 44

【配列表】

0007120630000001.app

【図 4 5】

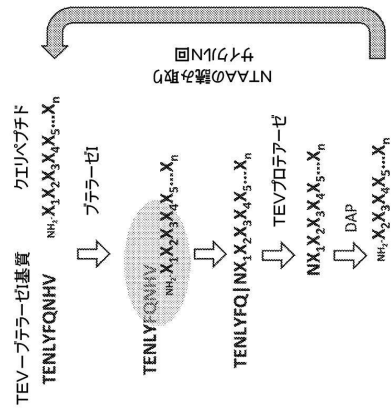


Fig. 45

10

20

30

40

50

## フロントページの続き

(51)国際特許分類

F I  
C 1 2 N 15/10 1 1 0 Z

(33)優先権主張国・地域又は機関

米国(US)

(31)優先権主張番号 62/376,886

(32)優先日 平成28年8月18日(2016.8.18)

(33)優先権主張国・地域又は機関

米国(US)

(72)発明者 チー, マーク

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, フリントコート アベニュー 1  
1 1 2 5, スイート ビー

(72)発明者 ガンダーソン, ケビン

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, フリントコート アベニュー 1  
1 1 2 5, スイート ビー

(72)発明者 ワイナー, マイケル フィリップ

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, フリントコート アベニュー 1  
1 1 2 5, スイート ビー

審査官 名和 大輔

(56)参考文献 国際公開第2007/062664(WO, A1)

国際公開第2009/050261(WO, A1)

特表2007-524105(JP, A)

(58)調査した分野 (Int.Cl., DB名)

C 1 2 N 15/00 - 15/90

C 1 2 Q 1/68 - 1/6897

G 0 1 N 33/50 - 33/98

C A p l u s / R e g i s t r y / M E D L I N E / E M B A S E / B I O S I S ( S T  
N )