

(19) 日本国特許庁(JP)

## (12) 公表特許公報(A)

(11) 特許出願公表番号

特表2004-509645

(P2004-509645A)

(43) 公表日 平成16年4月2日(2004.4.2)

(51) Int.Cl.<sup>7</sup>

**C12N 15/09**  
**B01J 20/24**  
**C07K 14/00**  
**C07K 19/00**  
**C12P 21/02**

F 1

C12N 15/00  
B01J 20/24  
C07K 14/00  
C07K 19/00  
C12P 21/02

Z N A A  
C  
C  
C  
C

テーマコード(参考)

2 G O 4 5  
4 B O 2 4  
4 B O 6 3  
4 B O 6 4  
4 G O 6 6

審査請求 未請求 予備審査請求 有 (全 84 頁) 最終頁に続く

(21) 出願番号 特願2002-530855 (P2002-530855)  
(86) (22) 出願日 平成13年8月17日 (2001.8.17)  
(85) 翻訳文提出日 平成15年2月17日 (2003.2.17)  
(86) 國際出願番号 PCT/GB2001/003693  
(87) 國際公開番号 WO2002/027327  
(87) 國際公開日 平成14年4月4日 (2002.4.4)  
(31) 優先権主張番号 0020357.0  
(32) 優先日 平成12年8月17日 (2000.8.17)  
(33) 優先権主張国 イギリス(GB)  
(31) 優先権主張番号 60/247,995  
(32) 優先日 平成12年11月14日 (2000.11.14)  
(33) 優先権主張国 米国(US)

(71) 出願人 501385112  
センス・プロテオミック・リミテッド  
イギリス国、シービー2・4エーティー、  
ケンブリッジ、バプラハム、バプラハム・  
ホール、ザ・バプラハム・バイオインキュ  
ベイター(番地なし)  
(74) 代理人 100058479  
弁理士 鈴江 武彦  
(74) 代理人 100091351  
弁理士 河野 哲  
(74) 代理人 100088683  
弁理士 中村 誠  
(74) 代理人 100108855  
弁理士 蔵田 昌俊

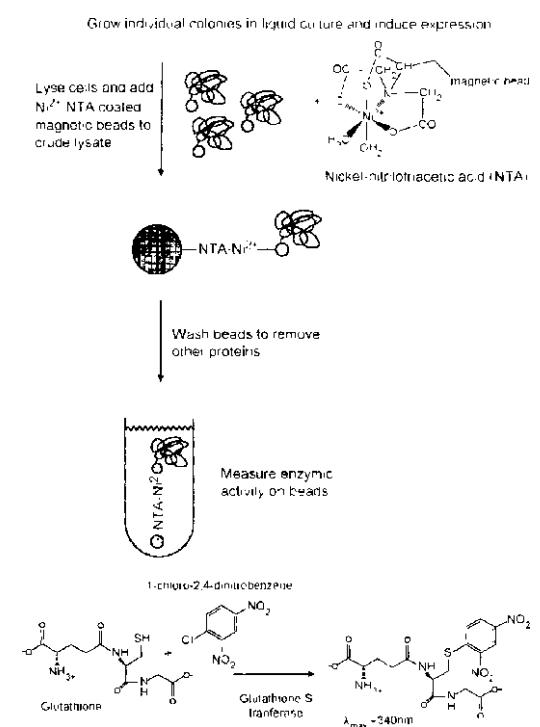
最終頁に続く

## (54) 【発明の名称】方法

## (57) 【要約】

## 【課題】

【解決手段】本発明は、一以上のドメインが完全長で正しく折り畳まれてあり、且つN末端又はC末端の何れかに一以上のマーカー部分が各々タグ標識されたタンパク質を作製する新規方法、このようなタンパク質を含有するアレイ、並びに迅速なスクリーニング用アレイにおけるこのようなタンパク質の使用に関する。



**【特許請求の範囲】****【請求項 1】**

一以上のドメインが完全長で正しく折り畳まれてあり且つN末端又はC末端の何れかに一以上のマーク部分が各々タグ標識された一以上のタンパク質を作製する方法であって、(a)前記タンパク質をコードするオープンリーディングフレームを、5'及び/又は3'非翻訳領域とともに有する一以上のDNA分子を準備することと、

(b)dNTP及び-S-dNTPが統計学的に娘DNA分子中に取り込まれる条件下で、前記DNA分子を増幅させることと、

(c)前記DNA分子の5'又は3'末端をヌクレアーゼ消化から特異的に保護することと、

(d)まず5'、3'ヌクレアーゼ又は3'、5'ヌクレアーゼを用いて前記DNA分子を処理して一群の入れ子型欠失物を得た後、前記オープンリーディングフレームの開始又は停止コドンを含む前記5'又は3'非翻訳領域を除去せしめる条件下で、一本鎖ヌクレアーゼで処理することと、

(e)一以上の5'又は3'マーク部分のコード配列を含有する発現ベクターの中に、工程(d)で得た前記断片をクローニングすることと、

(f)前記コードされているタンパク質を発現させることとを備えた方法。

**【請求項 2】**

前記DNA分子の前記増幅が統計学的に单一の-S-dNTPを取り込む請求項1に記載の方法。

**【請求項 3】**

前記単一の-S-dNTPが-S-dTTP又は-S-dATPの何れかである請求項2に記載の方法。

**【請求項 4】**

前記ヌクレアーゼがエキソヌクレアーゼIII又はエキソヌクレアーゼである請求項1~3の何れか1項に記載の方法。

**【請求項 5】**

前記一本鎖ヌクレアーゼがリヨクトウヌクレアーゼ又はT4DNAポリメラーゼである請求項1~4の何れか1項に記載の方法。

**【請求項 6】**

前記マーク部分によって前記オープンリーディングフレームの発現を確認することが可能となる請求項1~5の何れか1項に記載の方法。

**【請求項 7】**

前記マーク部分によって前記オープンリーディングフレームの折り畳みを確認することが可能となる請求項1~5の何れか1項に記載の方法。

**【請求項 8】**

前記マーク部分が緑色蛍光タンパク質をコードする請求項1~7の何れか1項に記載の方法。

**【請求項 9】**

前記マーク部分が、ペプチド配列(例えば、ヘキサヒスチジンタグ)、完全なタンパク質、又はタンパク質ドメイン(例えば、マルトース結合タンパク質ドメイン)である請求項1~7の何れか1項に記載の方法。

**【請求項 10】**

前記タグが前記アレイ中の各タンパク質の精製を可能とする請求項9に記載の方法。

**【請求項 11】**

前記各タンパク質の開始又は停止コドンが置換されるように、前記タグが挿入される請求項1~10の何れか1項に記載の方法。

**【請求項 12】**

前記各タンパク質の末端に近接する折り畳みと機能にとって重要でない領域内に、前記タグがインフレームに挿入される請求項1~10の何れか1項に記載の方法。

10

20

30

40

50

**【請求項 13】**

オープンリーディングフレームの中であるが、特定のドメイン境界の外側に位置する折り畳みと機能にとって重要でない領域の中に、前記タグがインフレームに挿入される請求項1～10の何れか1項に記載の方法。

**【請求項 14】**

工程(a)における前記DNA分子の増幅が、非プルーフリーディングポリメラーゼ(例えば、Taqポリメラーゼ又はDNAポリメラーゼIのクレノー断片)によって行われる請求項1～13の何れか1項に記載の方法。

**【請求項 15】**

dNTPに対する前記S-dNTPの比率が1:1乃至1:3の間にある請求項1～10の何れか1項に記載の方法。

**【請求項 16】**

前記5'、3'スクレアーゼ又は3'、5'スクレアーゼが-S-ホスホジエステル結合を加水分解できない請求項1～15の何れか1項に記載の方法。

**【請求項 17】**

前記DNA分子がmRNA配列からの逆転写によって得られたcDNAである請求項1～16の何れか1項に記載の方法。

**【請求項 18】**

前記方法が複数のDNA分子に対して平行して行われる請求項1～17の何れか1項に記載の方法。

**【請求項 19】**

前記方法が単一の容器中に存在するDNA分子の集団に対して行われる請求項1～18の何れか1項に記載の方法。

**【請求項 20】**

請求項1～19の何れか1項に記載の方法によって得られたタグ標識タンパク質のライブラリー。

**【請求項 21】**

タンパク質アレイを製造する方法であって、

- (a) 請求項20の前記ライブラリーの各要素をクローン毎に分離すること、
- (b) 空間的に離隔されたフォーマットで、前記タグ標識された各タンパク質を発現させること、
- (c) 前記マーカー部分によって、タグ標識された各タンパク質を精製すること、
- (d) 空間的に規定されたアレイに各タンパク質を配置することとを備えた方法。

**【請求項 22】**

請求項1～19の何れか1項に記載の方法によって調製された、又は請求項21に記載の方法によって作製されたタンパク質を備えたアレイ。

**【請求項 23】**

前記アレイの成分が例えば固相表面に固定化される請求項22に記載のアレイ。

**【請求項 24】**

前記各タンパク質がタグ部分によって固定化される請求項22又は23に記載のアレイ。

**【請求項 25】**

一以上の化合物の生物活性をスクリーニングする方法であって、前記一以上の化合物を請求項22～24の何れか1項に記載のタンパク質アレイと接触させる工程と、前記アレイ中の前記タンパク質への前記一以上の化合物の結合を測定する工程とを備えた方法。

**【請求項 26】**

一以上のタンパク質の特異的なタンパク質-タンパク質相互作用をスクリーニングする方法であって、前記一以上のタンパク質(例えば、細胞表面受容体)を請求項22～24の何れか1項に記載のアレイと接触させる工程と、前記アレイの前記タンパク質への前記一以上の特異的タンパク質の結合を測定する工程とを備えた方法。

**【請求項 27】**

20

40

50

ー以上のタンパク質の特異的なタンパク質 - 核酸相互作用をスクリーニングする方法であって、前記一以上の核酸プローブを請求項 22～24 の何れか 1 項に記載のタンパク質アレイと接触させる工程と、前記アレイ中の前記タンパク質への前記プローブの結合を測定する工程とを備えた方法。

【請求項 28】

化合物、タンパク質、又は核酸の迅速なスクリーニングにおける、請求項 22～24 の何れか 1 項に記載のアレイの使用。

【請求項 29】

前記アレイ中の各タンパク質を認識する分子のスクリーニングにおける請求項 22～24 の何れか 1 項に記載のアレイの使用であって、前記分子が好ましくは抗体である使用。

【請求項 30】

抗体アレイを作製する方法であって、タンパク質アレイ中の一以上のタンパク質が抗体ライブラリー中の少なくとも一つの抗体に結合するように、請求項 22～24 の何れか 1 項に記載のタンパク質アレイを前記抗体ライブラリーに接触させることと、未結合の抗体を全て除去することと、前記タンパク質アレイ中のタンパク質に結合した抗体を固定化することとを備えた方法。

【請求項 31】

タンパク質の機能又は量をスクリーニングする方法であって、請求項 30 に記載の抗体アレイを一以上のタンパク質の混合物に接触させる工程を備えた方法。

【請求項 32】

請求項 25～27 又は 30 の何れか 1 項に記載の方法であって、まず請求項 22～24 の何れか 1 項に記載のタンパク質アレイを準備する工程も備えた方法。

【請求項 33】

前記アレイ中の前記タンパク質が、単一の工程で精製され固定化される請求項 21 に記載の方法。

【請求項 34】

本明細書に記載されている方法によって得られたタグ標識されたタンパク質の使用。

【請求項 35】

発現されたタンパク質と他のタンパク質との相互作用を分析するための、本明細書に記載された方法によって得られたタグ標識されたタンパク質の使用。

【請求項 36】

例えば a ) 相互作用するタンパク質、 b ) DNA、又は c ) 化学的化合物のアフィニティークロマトグラフィーによる精製を可能とするためにアフィニティーカラム / 担体上に固定化するための、本明細書に記載された方法によって得られるタグ標識タンパク質の使用。

【請求項 37】

診断ツールとして抗体による検査を行うためのアフィニティー精製による固定化 (ELISA アッセイ) における、本明細書に記載の方法によって得られるタグ標識されたタンパク質の使用。

【請求項 38】

cDNA マイクロアレイ用のプローブとしての、本明細書に記載された方法によって得られるタグ標識されたタンパク質の使用。

【請求項 39】

DNA 結合タンパク質を同定するための請求項 38 に記載の使用。

【請求項 40】

「プロテオーム」中のタンパク質の種類を明らかにするための、本明細書に記載された方法によって得られるタグ標識されたタンパク質の使用。

【請求項 41】

本発明の方法によって改変された原ライブラリーの被発現タンパク質成分又は初発物質に質量分析法を行う請求項 40 に記載の使用。

10

20

30

40

50

**【発明の詳細な説明】****【0001】****【発明の属する技術分野】**

本発明は、一以上のドメインが完全長で正しく折り畳まれておる且つN末端又はC末端の何れかに一以上のマーク部分が各々タグ標識されたタンパク質を作製する新規方法、及びこのようないわゆるアレイ、並びに迅速なスクリーニングにおけるこのようなアレイの使用に関する。

**【0002】****【従来の技術】**

ゲノムマッピングプロジェクトは、治療における標的の発見過程及びそれを用いた薬物探索過程に大きな変革をもたらしている。新たな治療の標的が同定されるにつれて、既存の化学ライブラリー及びコンビナトリアル化学ライブラリーに対する高速大量処理スクリーニングによって、これらの標的に対して活性を有するリード化合物の候補が数多く提案されるであろう。初期段階の臨床試験を経た場合でさえ、全てのリード化合物を追求することは明らかに非経済的である。しかしながら、このようないわゆるリード化合物が示すと思われる活性のプロファイルをある生物中の全てのタンパク質につき迅速に評価する方法は現在のところ存在しない。もしこのような方法が利用できれば、全リード化合物の潜在的毒性プロファイルを初期の段階で評価することが可能となり、この情報によって、何れのリード化合物を追求し、何れのリード化合物を排除すべきかを決定する過程が著しく促進されるであろう。

10

20

**【0003】**

他方、製薬産業では、既存の薬物（既に市場に出ているか、又は開発途中である）の標的を全て同定することによって、それらの作用機序を明らかにする必要がある。現在、新薬の規制当局が作用機序の解明を殊に重要視していることがさらに明瞭となっているので、そのような情報が利用できれば、新薬の規制認可を得る過程が大幅に短縮されるであろう。さらに、この種の情報により、改良された第二世代の薬物デザインが可能となるであろう。薬物の大多数は、薬物又はその代謝物が望ましくない標的に結合することによって生じると思われる軽微な副作用を少なくとも有しており、改良された薬物をデザインするために必要な基準を明らかにするために、これらの標的タンパク質を全て同定する必要があるからである。しかし、今のところ、この情報を得る簡易な方法は存在せず、何百万ドルもの規模となり得る数多くの薬物が、作用の標的が不明であるというだけで脱落している。

30

**【0004】**

タンパク質 - タンパク質相互作用は、内的及び外的ストレスの両者に対する細胞内反応を支配する上で、決定的に重要であることが益々認識されつつある。従って、特異的なタンパク質 - タンパク質相互作用は、感染及び他の疾患への薬物による介入に対する標的となる可能性がある。現在のところ、タンパク質 - タンパク質相互作用を評価するための信頼するに足る唯一の方法は、酵母のツーハイブリッドアッセイであるが、この種のインビボアッセイは、高速大量処理形式でない場合でも、タンパク質 - タンパク質相互作用の特異的アゴニスト又はアンタゴニストを同定するのには余り適していないであろう。機能的プロテオーム発現アレイ、又は「プロテオームチップ」を用いれば、タンパク質 - タンパク質相互作用の特異性及び薬物によって媒介される任意の効果の特異性をインビトロ形式で決定することが可能となる。機能的プロテオーム発現アレイは、実際にこの研究分野に多大な変革をもたらすので、非常に大きな発展性を有すると考えられる。

40

**【0005】**

機能的プロテオームアレイを作製できる一つの方法は、特定のプロテオーム中で発現させた全タンパク質を個別にクローニングし、発現させ、精製し、固定化するというものである。しかし、ここで考慮すべき最も重要な問題は、ゲノム全体の配列データが利用可能であるかということと、同時に、所定のゲノムの絶対的大きさに関することがある。

**【0006】**

50

ゲノムの大きさの例を挙げると、典型的な細菌ゲノムは、~5 Mbpであり、既に完全に配列決定されたものも少数存在する（例えば、*Helicobacter pylori*、*Escherichia coli*及び*Mycobacterium tuberculosis*）。真菌ゲノムは、典型的には~40 Mbp、哺乳類ゲノムは~3 Gbp、植物ゲノムの場合は~10 Gbpである。かかる情報のうちどの程度が公有されるかということは大いに疑問であるが、現時点での推測によれば、ヒトゲノム配列は2003年頃に完了する予定である。代表的なモデル生物以外の生物のゲノムが現実的な時間の枠内で入手できるようになると予想することが全く非現実的であることは明らかであり、さらに、機能的プロテオミクスの観点から、モデル生物が有する価値は限定的なものにすぎない。それゆえ、原理的には、今から4年のうちに、プライマーをデザインし、合成して、cDNAライブラリーからヒトゲノム中に存在する約100,000の各遺伝子をクローニングすることが可能かもしれないが、必要な配列データが利用可能であるとしても、実際、これは、多大な費用を要し（プライマーの費用だけで数百万ドルに達するであろう）、且つ非常に手間のかかる過程となろう。

10

20

30

40

50

#### 【0007】

しかし、完全な配列データが利用できない薬物開発に適した生物はどうなるのであろうか。これらの機能的プロテオミクスを単に無視することはできないが、では、いかなる代案を探り得るのであろうか。原理的には、非特異的固定化と共に発現cDNAライブラリーを使用してタンパク質アレイを作製できるであろうが、非特異的固定化の過程で、通常、タンパク質の折り畳みが破壊されることにより機能が失われる所以、この技術には大きな制約がある。さらに、宿主細胞のタンパク質も全て固定化されることになって、最良の場合でもシグナル-ノイズ比が大きく減少し、最悪の場合には、ポジティブな結果が不明瞭になるであろう。従って、機能に影響を及ぼさず且つ全ゲノム配列を知る必要もなしに、共通のモチーフ又はタグを介して個々のタンパク質が特異的に固定化、精製される機能的プロテオームアレイを作製する技術は、機能的プロテオミクスの分野に大きな進歩をもたらすであろう。

#### 【0008】

##### 【発明の概要】

本発明者らは、そのタンパク質に対応する遺伝子のDNA配列を予め知る必要なくして、プロテオーム中の各タンパク質内の所定の位置に共通のマーカーをタグ標識する方法を提供することによって、上記問題を解決する新規アプローチを開発した。次いで、この「タグ」を使用して、その後の固定化と精製手段に共通性及び特異性を与えることがあり、続いて、あるプロテオームに由来する何千ものタンパク質が並べられた、空間的に規定されたアレイの作製が可能となる。

#### 【0009】

ここで考慮すべき重要な事柄は、「タグ」の正確な位置決定に関するものである。タグがある遺伝子の特定しないランダムな位置にインフレームで挿入された場合、得られたタグ付加タンパク質は不特定に末端切断され、殆どの場合、正しい折り畳みが破壊され、このため機能が破壊されることになる可能性がある。完全長のタンパク質には、折り畳み又は機能に影響を与えずに末端切断することができる短いポリペプチドの伸長部分がN-末端又はC-末端の一方（又は双方）に存在することが多い。しかし、末端切断が何れかのN-末端又はC-末端伸長部分を除去し、末端切断がドメイン境界を横切れば、通常、タンパク質の折り畳み及び機能に影響が及ぶ。本明細書に記載した方法によれば、アレイに特異的に固定化された時に、各タグ標識タンパク質が正しく折り畳まれて機能を保持するように、正確に各タンパク質のN又はC末端の何れかに、又はタンパク質の折り畳み及び機能に重要でない何れかの末端に近接した領域内に、タグを正確なリーディングフレームで挿入することが可能となる。個々のドメインが異なる機能を有する複数ドメインタンパク質の場合、本明細書に記載した方法によって、アレイに特異的に固定化された時に個々のタグ挿入ドメインが正しく折り畳まれ、そのタンパク質が機能を保持するように、コード配列全体の内側であるが特定のドメイン境界の外側にタグを挿入することも可能となる

。

### 【0010】

アレイ中の各タンパク質は完全に機能的であると考えられるので、そのアレイを直接スクリーニングして、薬物の標的及び他の生物学的に関連する分子を同定することが可能である。アレイを空間的に仕切ることにより、各タンパク質の表現型を直接的にその遺伝子型と関連させることができるとなり、「ヒット」を同定できるであろう。

### 【0011】

#### 【発明の実施の形態】

このように、第一の側面において、本発明は、一以上のドメインが完全長で正しく折り畳まれてあり、且つN末端又はC末端の何れかに一以上のマーカー部分が各々タグ標識された一以上のタンパク質を作製する方法であって、  
10

(a) 前記タンパク質をコードするオープンリーディングフレームを5'及び/又は3'非翻訳領域とともに有する一以上のDNA分子を準備することと、

(b) dNTP及び-S-dNTPが統計学的に娘DNA分子中に取り込まれる条件下で、前記DNA分子を増幅させることと、

(c) 前記DNA分子の5'又は3'末端をヌクレアーゼ消化から特異的に保護することと、

(d) まず5'、3'ヌクレアーゼ又は3'、5'ヌクレアーゼを用いて前記DNA分子を処理して一群の入れ子型欠失物を得た後、前記オープンリーディングフレームの開始又は停止コドンを含む前記5'又は3'非翻訳領域を除去せしめる条件下で、一本鎖ヌクレアーゼを用いて処理することと、  
20

(e) 一以上の5'又は3'マーカー部分のコード配列を含有する発現ベクターの中に、工程(d)で得た前記断片をクローニングすることと、

(f) 前記コードされているタンパク質を発現させることとを備えた方法を提供する。

### 【0012】

好ましくは、前記一又は複数のDNA分子の前記増幅は統計学的に単一の-S-dNTP、より好ましくは-S-dTTP又は-S-dATPの何れかを取り込む。

### 【0013】

前記マーカー部分(marker moiety)は、ペプチド配列(例えば、ヘキサヒスチジンタグ)、抗体エピトープ又はビオチン模倣物(biotin mimetic)、又は完全なタンパク質、又はタンパク質ドメイン(例えば、マルトース結合タンパク質ドメイン)の何れかであり得る。例えば、ビオチン又は脂質分子を加えることによって前記マーカー部分自体を翻訳後修飾することも可能である。好ましい実施態様では、前記マーカー部分によって、「タグ標識された(tagged)」タンパク質の精製も可能となるであろう。  
30

### 【0014】

このように、本発明の方法は、個々の遺伝子の配列に関する知識に全く依拠せずに、一つの容器内で、cDNAライブラリーの全ての要素を特異的に修飾することを可能とする。本発明の方法は、その代わりに、各cDNAの5'又は3'非翻訳領域の何れかが除去されるように、ヌクレアーゼによって各cDNAを非進行的(non-processive)に末端切断することに依拠している。次いで、得られた各cDNAの入れ子型欠失物(nested deletion)群に、既知のマーカー部分をコードする既知のDNA配列をさらに付け足す。このため、前記マーカー部分が個々のcDNAと同じリーディングフレームの中に存在し、インフレームな停止コドンが先行していなければ、本発明の方法によって作製した遺伝的修飾が施された得られた各cDNAは、共通の部分(例えば、そのN末端又はC末端の何れかに融合されたポリペプチド「タグ」)を有するに至った各タンパク質をコードしているであろう。次いで、正しく折り畳まれたタグ標識タンパク質をスクリーニングすることによって、ドメインの境界を横切って個々のタンパク質の折り畳みに(従って機能に)影響を与える全ての末端切断と前記タグへのフレーム外融合物(out-of-frame fusion)とを除去することが可能となる。cDNA  
40  
50

ライブラリーの全ての要素は同様に修飾されると思われる所以、最終的には、cDNAライブラリーによってコードされる全てのタンパク質のN末端又はC末端の何れかに共通の部分がタグ標識されるに至るであろう。

【0015】

一般的に、前記cDNAライブラリーから発現される前記タンパク質は「タグ標識」されているものと思われ、即座に同定し単離することができる。一旦精製されると、例えば、マイクロアレイに前記タンパク質を付着させることができる。付着はタグ自体によって行ってもよいし、あるいは、まず前記タンパク質に付着される別の部分によって行ってもよい。

【0016】

本明細書に記載されている方法によって形成されたアレイは、本発明の第二の側面を成す。このようなアレイは、通常固相支持体上に固定化された「タグ標識」タンパク質発現ライブラリーを備えている。当業者であれば、アレイの分野で、固相支持体となり得る広範な支持体が一般に使用されており、本発明のアレイを製造する際に、これらの「担体(substrate)」の何れも使用できることが理解できるであろう。

【0017】

本明細書で使用する「タンパク質アレイ」という用語は、ある表面上に、あるパターンで一以上のタンパク質部分を空間的に規定して配置したものを意味する。好ましくは、前記タンパク質部分は、直接又は間接的に前記表面に付着されるであろう。前記付着は、(例えば、前記表面上への物理的吸収による、又は非特異的な共有結合性相互作用の形成による)非特異的なものであってもよい。好ましい実施態様では、前記タンパク質部分は、本明細書に記載された方法を用いて、各タンパク質に連結された共通のマーカー部分を介して前記表面に付着されるであろう。

【0018】

別の実施態様では、前記表面に繋留された(tethered)小胞又はリポソーム中に前記タンパク質部分を取り込ませてもよい。

【0019】

このように、前記パターン中の各位置は、例えば、一コピー以上の  
 a) 単一のタンパク種(单量体、二量体、三量体、四量体、又はこれ以上の多量体の形態)のサンプル、  
 b) 相互作用する分子(例えば、DNA、抗体、他のタンパク質)に結合された单一のタンパク種のサンプル、  
 c) 合成分子(例えば、ペプチド、化学的化合物)に結合された单一のタンパク種のサンプル、又は  
 d) 前記アレイの前記パターンの各位置に位置する2~100の異なるタグ標識タンパク質部分の混合物

を含有し得る。

【0020】

前記アレイを支える表面は、例えば化学的な処理によってコート/誘導体化してもよい。適切な表面の例には、スライドグラス、ポリプロピレン若しくはポリスチレン、シリカ、金若しくは金属の支持体、又は例えば、ニトロセルロース、PVDF、ナイロン、若しくはホスホセルロース製の膜が含まれる。

【0021】

本明細書に論述されているように、本発明の方法によれば、あるプロテオーム中の全てのタンパク質のN末端又はC末端の何れかに特異的にタギングすることが可能となる。タンパク質の中にはN末端に伸長部を備えられないものや、C末端に伸長部を備えられないものもあるが、大多数のタンパク質は、このような伸長部のうちの何れかを備え得るであろう。しかしながら、既存のライブラリークローニング法は、完全長の未修飾cDNAとして、又は何らかのタンパク質対へのランダム且つほぼ不可避的に末端切断された融合物として遺伝子をクローニングする所以この課題には全く応えられない。後者と比べて、本發

10

20

30

40

50

明の方法は、例えば所望のペプチド対への融合によって c D N A 産物の折り畳み又は機能が影響を受けないように、c D N A 産物の N 末端又は C 末端残基に又はその近傍に位置する前記配列へ、前記タグの位置を誘導することを可能とする。前者と比べて、本明細書に記載されているアレイにタンパク質を固定化する方法は、非特異的な相互作用ではなく特異的な相互作用を介しており、これらの特異的な相互作用は各 c D N A の末端に付加されたタグの機能である。加えて、本明細書に記載されている方法は、非細菌性宿主生物中で発現された精製固定化タンパク質をスクリーニングして、正確な折り畳みと翻訳後修飾を通じて機能の維持を補助するために使用することができるのに対して、ファージディスプレー又は c D N A 発現ライブラリーのような既存の方法は、不適切な折り畳み又は不正確な翻訳後修飾のために、その中では真核細胞のタンパク質の多くが非機能的形態で合成されることが明らかとなっている細菌性宿主に限定される。

10

#### 【 0 0 2 2 】

本発明の方法には、広くインビトロでの用途が考えられ、これは大きく主な三つの分野に分けることができる。これらは、タンパク質 - リガンド相互作用の研究、タンパク質 - タンパク質相互作用の研究、及びタンパク質 - D N A 相互作用の研究である。

#### 【 0 0 2 3 】

##### タンパク質 - リガンド相互作用

本明細書に記載されている方法によれば、ある新規化学物質とあるプロテオーム中に存在する全てのタンパク質との相互作用を迅速に解明することが可能となるであろう。いわば逆高処理量スクリーニング ( r e v e r s e h i g h t h r o u g h p u t s c r e e n ) において、様々なストリングエンシーで、新規化学物質 ( N C E ) を用いて適切なプロテオームをプロープするだけで、これを達することができる。このようなスクリーニングから得られた計測値は、多くの場面で直接役に立つものと思われ、そのうちの幾つかは以下に示されている。

20

#### 【 0 0 2 4 】

細胞又は生物全体に対して化合物のライブラリーを調べる高処理量スクリーニングプログラムによって、しばしば、リード化合物が同定され、これは、スクリーニングに先立って標的が知られていなくても表現型に変化をもたらす。しかしながら、これに引き続いて行われる一次標的の同定は極めて骨の折れる作業であり得る。本発明の方法は、対象としている種の機能的なプロテオームアレイを創作した後、リード化合物を用いて該アレイをスクリーニングして、プロテオーム中の何れのタンパク質を標的としているかを明らかにすることが可能と思われる所以、この種の課題に直接応用できる。タンパク質 - リガンド相互作用を同定するためのこのような大規模な同時並行的アプローチによって、新規化学物質の一次標的を決定することが大幅に加速且つ簡易化され、同じく重要であるかもしれないさらに弱い二次的相互作用を同定することも可能となろう。さらに、本方法は、種間の交差反応性の問題に直接適用することができ、例えば、ヒトのプロテオーム中に存在する全てのタンパク質との抗真菌候補化合物の相互作用を迅速に評価することが可能となる。その後リード化合物に何らかの最適化を行う上で、この種の情報は極めて有用であることが明らかとなるであろう。

30

#### 【 0 0 2 5 】

現在、高処理量スクリーニング法によれば、それ自体治療の標的となる可能性があることが以前に明らかとされたあるタンパク質に結合する小分子を迅速に同定することが可能である。しかしながら、これらの方法は、何らかの所定の相互作用がどのくらい選択的であるかを知ることが所定のリード化合物の検査を継続するかどうかを決定する上で極めて重要なかも知れないにもかかわらず、相互作用がどのくらい選択的であるかという問題に応えるものではない。十分な理解を有する者であれば、数多くの関連するタンパク質又は無関連なタンパク質にもヒットする化合物と比べて、単一のタンパク質を標的とする化合物の方が副作用が少ない可能性があると主張するであろう。

40

#### 【 0 0 2 6 】

臨床試験の第三相に進むことができたにもかかわらず、主たる作用機序が不明であるため

50

に、当局の承認を得ることができない化合物の例が数多く存在している。抗うつ剤ミアンセリンとトラザドン及びファイザーの抗関節炎薬テニダップがその例であり、それぞれ何億ドルもの投資が全く回収されていない。このような薬物の主たる標的を発見した後、作用機序を明らかにすることさえできれば、極めて多大なコストを費やした臨床試験のデータが規制当局の認可を得ることができるので、本明細書に記載されている方法は、このような失敗に終わった薬物を復活させるために利用できる可能性がある。既存の薬物は全て、多かれ少なかれ副作用を有しており、それ余の点では有益な抗統合失调症薬であるクロザピンがその例である。このような副作用の分子的な起始点を決定することができれば、最適な主要効果に加えて副作用が最小である次世代の薬物を設計することが極めて容易となるであろう。ある化合物とプロテオーム中の全タンパク質との相互作用のプロファイルを作製する際に、異常な二次的相互作用が同定され、これらの二次的相互作用は、引き続き、既知の副作用と関連しているかどうかという見地から評価することができると思われる。10  
ので、本明細書に記載の方法はこのような課題にも直接適用することができる。

#### 【0027】

一般的な阻害剤を有するプロテオームアレイをスクリーニングすることによって、本発明の方法は、セリンプロテアーゼ等のタンパク質のファミリーを同定するために使用することができる。次いで、この方法を用いれば、より焦点を絞ったリード化合物のスクリーニングを行うために、例えば、全てのヒトセリンプロテアーゼ又は全てのキナーゼ又は全てのp450酵素を提示するバイオチップを統一して開発することが可能となるであろう。p450によって媒介される水酸化は、代謝過程の第一段階であることが多く、患者毎の薬物反応が異なる主な原因の一つと考えられるので、例えば、p450バイオチップは、あるリード化合物が代謝される可能性があるか否かを評価する際に有用であろう。現に、現時の薬物設計の最終目標の一つは、最初から代謝されない化合物を作製することであり、この場合にも、p450チップは極めて有用性を有する可能性がある。  
20

#### 【0028】

##### タンパク質 - タンパク質相互作用

タンパク質 - タンパク質相互作用及び多タンパク質複合体は、細胞生物学において特に重要な。30  
ある。例えば、シグナル伝達経路は、一般に細胞表面受容体と外部リガンドとの相互作用によって開始され、この後にタンパク質 - タンパク質相互作用のカスケードが続き、最終的に特定の遺伝子を活性化させる。各タンパク質 - タンパク質相互作用は、特異的なリガンドの存在に依存しているか、あるいは特異的なリガンドによって遮断されることがあるが、多タンパク質複合体の中には、リガンド依存的にしか形成されないものもある。

#### 【0029】

ツーハイブリッド技術を用いて、何千もの新しいタンパク質 - タンパク質相互作用が同定されてきた。本明細書に記載されている方法は、このような方法の限界を克服するものであり、各標識タンパク質を用いて、相互作用の相手方のみならず、各相互作用の相対的な強度をも明らかにするためにプロテオームアレイをスクリーニングするために使用することができる。本方法は、多タンパク質複合体の会合がリガンド依存的である場合さえ、多タンパク質複合体の成分を同定するために使用することもできる。  
40

#### 【0030】

新規タンパク質 - タンパク質相互作用を明らかにする際に、本発明の方法をこのように使用する例には、疾病状態への関与が示唆されている細胞表面受容体の細胞質ドメインのシグナル伝達の相手方を同定することが挙げられるであろう。このようなタンパク質 - タンパク質相互作用から治療の標的となり得る標的が直ちに明らかとなる場合があるので、このようなシグナル伝達の相手方を同定することは、薬学的見地から全く適切であろう。

#### 【0031】

##### タンパク質 - DNA相互作用

ヒトゲノム中に存在する全ての遺伝子の概ね10%が転写因子をコードしていると推定されているが、これらのうち現時点で同定されているものは僅かである。DNAエンハンサ  
50

—要素に特異的転写因子が（多くの場合外的な刺激に応じて）結合することは、エンハンセオソーム複合体の形成に必須であり、次いで該複合体が遺伝子発現のスイッチを入れる。原理的に、遺伝子発現が薬物投与によって影響を受け得る様々なポイントが存在する。薬物によっては、タンパク質又は小分子が細胞表面受容体に結合するのをブロックし、これによりシグナル伝達カスケードを元から遮断する場合もあるし、タンパク質・タンパク質相互作用を遮断し、又はシグナル伝達カスケード内の酵素活性を阻害する場合もある。あるいは、薬物によっては、エンハンセオソーム複合体内の特異的なタンパク質-DNA相互作用又はタンパク質・タンパク質相互作用の形成を遮断する場合もある。このような例として、転写因子NF-Bは、免疫及び炎症反応、四肢の発育、敗血性ショック、喘息、並びにHIVプロペプチドの产生など多岐にわたる細胞のプロセスに関与している。NF-Bの活性化における細胞内シグナル伝達カスケードの多くはこれらのプロセス全てに共通しているので、治療的介入を行うための有望な標的とはなり得ない。従って、これらの反応の差異は、最初のリガンド・受容体相互作用に存するか、又は特異的なエンハンセオソーム複合体の形成に存するかの何れかである。NF-Bは、少なくとも14の異なるエンハンサー要素に結合することが知られており、エンハンセオソーム複合体は、従って、治療の標的になり得る可能性がある。しかしながら、それぞれのエンハンセオソーム複合体の概要を把握するためには、関与する各DNA結合タンパク質の数およびお互いのタンパク質・タンパク質相互作用をともに知ることが必要となる。本発明の方法は、これらの両課題を直接解決するために使用することができる。プロテオームアレイは、新規DNA結合タンパク質を同定するために、特異的なDNAプローブを用いてスクリーニングすることができる。あるいは、転写因子のトランス活性化ドメインが相互作用する他のタンパク質を同定するために、転写因子のトランス活性化ドメインを用いて、プロテオームアレイをスクリーニングすることができる。このようなスクリーニングの相互関連によって、特定のエンハンセオソーム複合体の新規成分の同定が可能となるはずである。

#### 【0032】

本発明の方法によって得られるタンパク質アレイを用いれば、アレイ中に提示された各タンパク質を認識する分子を選別することも可能となるであろう。好ましい実施態様において、選別される分子は抗体又は抗体様タンパク質が考えられ、ファージ若しくはリボソーム上にディスプレイされるか、又はコードするmRNAに共有結合されるであろう。

#### 【0033】

このように、アレイ中の固定化された各タンパク質に、ファージディスプレイされた抗体ライブラリーを与えることができ、未結合の抗体は洗浄によって除去される。続いて、選別されたファージを回収し、通常の操作に従って細菌を感染するために使用することができる。次いで、さらに選別を進めるために、ファージを感染させた細菌には、選別された前記抗体をディスプレイするファージ粒子を産生させるか、又はそのまま使用するために可溶性の抗体断片を産生させることができる。「抗体」又は「抗体断片」という用語は、本明細書において、マウス、ヒト、ラクダ、又はその他の生物から得られる一本鎖Fvs、FAB断片、各軽鎖又は重鎖断片を意味する。

#### 【0034】

好ましい実施態様において、選択工程後に、各ウェルに適切な細菌細胞を加えることによって（細菌は選択されたファージ粒子にウェルの中で感染させる）ファージ粒子を回収できるように、前記タンパク質アレイはマイクロフェルフォーマットであり得る。続いて、アレイ中に固定された各タンパク質に対して選別された抗体断片を物理的に隔離し続けながら、各ウェルに増殖培地を加え、感染させた前記細菌を増殖させて、抗体断片を発現させることができる。所望であれば、その後の選別の際に、感染細菌によって産生された新規ファージ粒子を使用することができる。このような操作は、現在では、ポリクローナル又はモノクローナル抗体断片を单一の精製及び固定化タンパク質に選別するための慣用的操作である。実際、その後、元のタンパク質アレイは、ここで、それ以外は標準的なインビトロ抗体選別法を用いつつ、極めて同時並行的に、ポリクローナル又はモノクローナル抗体断片を何千もの正しく折り畳まれたタンパク質にすることができるであろう。

10

20

30

40

50

**【 0 0 3 5 】**

単一の所定位置に固定化された元のアレイ中のタンパク質に対して新しいアレイの各位置に存在する抗体断片自体が選択されるように、空間的に規定された新しいアレイに、元のアレイの各ウェルに由来する選別され可溶状態で発現されている前記抗体断片自体を固定化することができる。このようにして作製された抗体アレイは、可溶性抗体断片の固定化に先立って実施された選択の回数に応じて、各位置に、ポリクローナル又はモノクローナル抗体断片を含有することになるであろう。

**【 0 0 3 6 】**

このような抗体アレイには、関連するプロテオームの発現差をモニタリングするために、未精製細胞又は組織可溶化液から得られる各タンパク質を捕捉することを含む数多くの用途が存在し得るであろう。あるいは、抗体補足されたタンパク質のリガンド結合機能を直接スクリーニングしてもよい。一般に、標的タンパク質に結合してその機能を阻害するモノクローナル抗体もあるし、標的タンパク質に結合するが機能を阻害しないモノクローナル抗体もある。大規模な同時並行的アプローチにおいては、全てのモノクローナル抗体について、機能に影響を与えずにプロテオーム中の全タンパク質に結合する能力を個別に評価することが非現実的であることは明らかである。しかしながら、プロテオーム中の全てのタンパク質に対する一群のポリクローナル抗体の中には、機能に影響を与えずに結合するという所望の能力を有する個別の抗体が含有されている可能性があり、さらに、あるタンパク質の全ての翻訳後修飾を認識する個別の抗体が含有されているものと思われる。このため、一般的には、既述のごとく作製されたモノクローナル抗体アレイよりポリクローナル抗体アレイの方が、捕捉されたタンパク質の機能を直接スクリーニングする上では有用であろう。

10

20

30

40

50

**【 0 0 3 7 】**

元のタンパク質アレイに比べて、本明細書に記載された方法によって作られる抗体アレイは、アレイ上に固定化された全てのタンパク質が同様の条件下で安定であるという利点を有するであろう。未精製細胞又は組織可溶化液から捕捉されたタンパク質は組換えタンパク質ではなく、天然に発現されたものであろう。さらに、捕捉されたタンパク質は、未精製細胞又は組織可溶化液から捕捉した後に、機能又はリガンド結合等につき直接スクリーニングすることもでき、これによって、機能の維持を助けるはずである。

**【 0 0 3 8 】**

このように、さらなる側面において、本発明は、

( 1 ) 一以上の化合物の生物活性をスクリーニングする方法であって、前記一以上の化合物を本明細書に記載のタンパク質アレイと接触させる工程と、前記アレイ中の前記タンパク質への前記一以上の化合物の結合を測定する工程とを備えた方法、

( 2 ) 一以上のタンパク質の特異的なタンパク質 - タンパク質相互作用をスクリーニングする方法であって、前記一以上のタンパク質(例えば、細胞表面受容体)を本明細書に記載のタンパク質アレイに接触させる工程と、前記アレイ中の前記タンパク質との前記一以上のタンパク質の結合を測定する工程とを備えた方法、

( 3 ) 一以上のタンパク質の特異的なタンパク質 - 核酸相互作用をスクリーニングする方法であって、前記一以上の核酸プローブを本明細書に記載のタンパク質アレイと接触させる工程と、前記アレイ中の前記タンパク質への前記プローブの結合を測定する工程とを備えた方法、

( 4 ) 化合物、タンパク質、又は核酸の迅速なスクリーニングにおける、本明細書に記載のアレイの使用、

( 5 ) 前記アレイ中の各タンパク質を認識する分子のスクリーニングにおける本明細書に記載のアレイの使用であって、前記分子が好ましくは抗体である使用、

( 6 ) 抗体を作製する方法であって、タンパク質アレイ中の一以上のタンパク質が抗体ライブラリー中の少なくとも一つの抗体に結合するように、本明細書に記載のタンパク質アレイを前記抗体ライブラリーに接触させることと、未結合の抗体を全て除去することと、前記タンパク質アレイ中のタンパク質に結合した抗体を固定化することとを備えた方法、

及び

(7) タンパク質の機能又は量をスクリーニングする方法であって、本明細書に記載の抗体アレイを一以上のタンパク質の混合物に接触させる工程を備えた方法を提供する。

【0039】

前記方法(1)、(2)、(3)、及び(4)は、まず、本発明の一以上的方法によって前記アレイを準備する工程を含んでもよい。

【0040】

本明細書に記載された方法から得られたタンパク質の使用は、本発明の別の側面を構成する。当業者であれば、修飾タンパク質が利用され得る広範な用途が本分野において公知であることが理解できるであろう。

10

【0041】

従って、さらなる側面において、本発明は、

(1) 多数の発現宿主(すなわち、細菌、酵母、哺乳類細胞)における本発明の方法によつて作製されたタグ標識タンパク質の発現(例えば、Walker EA, Clark AM, Hewison M, Ride JP, Stewart PM. Functional expression, characterization, and purification of the catalytic domain of human 11-beta-hydroxysteroid dehydrogenase type 1. *J Biol Chem* 2001 Jun 15; 276(24): 21343-50; Cai J, Daoud R, Georges E, Gros P. Functional Expression of Multidrug Resistance Protein 1 in *Pichia pastoris*. *Biochemistry* 2001 Jul 17; 40(28): 8307-16, and Hara H, Yoshimura H, Uchida S, Toyoda Y, Aoki M, Sakai Y, Morimoto S, Shiokawa K. Molecular cloning and functional expression analysis of a cDNA for human hepassocin, a liver-specific protein with hepatocyte mitogenic activity(1). *Biochim Biophys Acta* 2001 Jul 30; 1520(1): 45-53参照)、  
20

20

(2) 本明細書に記載されている方法によって得られたタグ標識されたタンパク質の使用

、

(3) 前記修飾されたDNA分子を酵母ツーハイブリッド発現ベクター中にクローニングすることを介して酵母ツーハイブリッド系内での発現されたタンパク質と他のタンパク質との相互作用を分析するための、本明細書に記載された方法によつて作製されたタグ標識タンパク質の使用(例えば、Staudinger J, Perry M, Elledge SJ, Olson EN. Interactions among vertebrate helix-loop-helix proteins in yeast using the two-hybrid system. *J Biol Chem* 1993 Mar 5; 268(7): 4608-11, and Vojtek AB, Hollenberg SM, Cooper JA. Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* 1993 Jul 16; 74(1): 205-14参照)、  
30

30

(4) アフィニティーカラム/担体上に固定化して例えばa)相互作用するタンパク質、b)DNA、又はc)化学的化合物をアフィニティークロマトグラフィーによって精製できるようにするための、本明細書に記載された方法によつて作製されたタグ標識タンパク質の使用(例えば、Rhodes N, Gilmer TM, Lansing TJ. Expression and purification of active recombinant atm protein from transiently t  
40

40

50

ransfected mammalian cells. Protein Expr Purif 2001 Aug; 22(3): 462-6; Zwicker N, Adelhelm K, Thiericke R, Grabley S, Hanel F. Strep-tag II for one-step affinity purification of active bHLHzip domain of human c-Myc. Biotechniques 1999 Aug; 27(2): 368-75, and Giuliani CD, Iemma MR, Bondioli AC, Souza DH, Ferreira LL, Amaral AC, Salvini TF, Selistre-de-Araujo HS. Expression of an active recombinant lysine 49 phospholipase A(2) myotoxin as a fusion protein in bacteria. Toxicon 2001 Oct; 39(10): 1595-600 参照)。 10

#### 【0042】

(5) 診断ツールとして抗体による精査(interrogation)を行うためのアフィニティー精製による固定化(ELISAアッセイ)における、本明細書に記載の方法によって作製されたタグ標識タンパク質の使用(例えば、Doellgast GJ, Triscott MX, Beard GA, Bottoms JD, Cheng T, Roh BH, Roman MG, Hall PA, Brown JE. Sensitive enzyme-linked immunosorbent assay for detection of Clostridium botulinum neurotoxins A, B, and E using signal amplification via enzyme-linked coagulation assay. J Clin Microbiol 1993 Sep; 31(9): 2402-9 参照)。 20

(6) DNA結合タンパク質を同定するためのcDNAマイクロアレイ用のプローブとしての、本明細書に記載された方法によって得られるタグ標識されたタンパク質の使用(例えば、Derisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 1996 Dec; 14(4): 457-60 参照)。 30

#### 【0043】

(7) 本発明の方法によって改変された原ライブラリー又は初発物質の被発現タンパク質成分の質量分析によって「プロテオーム」中のタンパク質の種類(identity)を明らかにするための本明細書に記載された方法によって作製されたタグ標識タンパク質の使用(例えば、Bordini E, Hamdan M. Investigation of some covalent and noncovalent complexes by matrix-assisted laser desorption/ionization time-of-flight and electrospray mass spectrometry. Rapid Commun Mass Spectrom 1999; 13(12): 1143-51 参照)。 40

#### 【0044】

本発明の各側面の好ましい特徴は、その他の各側面に対しても互いに当てはまるものである。

#### 【0045】

##### 【実施例】

ここで、次に挙げる例を参照して本発明を説明するが、これらの例は、いかなる意味においても、本発明の範囲を限定するものとして解釈されるべきではない。

## 【0046】

例1

## (a) ベクターの構築(図1a参照)

本発明者らは、強力なハイブリッドプロモーター(P trc)を含有し、該プロモーター配列のすぐ下流に存在するNcoIサイトにクローニングした遺伝子発現を誘導するpUC19由来のベクターpMM106Hを構築した。本発明者らは、676bpのナンセンスDNA配列を詰め込み断片としてNcoIサイトと下流のHpaIサイトとの間に挿入した。HpaIは、平滑末端カッターであり、リーディングフレームが平滑末端の第一の塩基にある場合、下流のDNAがポリアスパラギン、ヘキサヒスチジンペプチドをコードするように、ベクターを切斷するために置かれている。ヘキサヒスチジンタグに続いて、アンバー停止コドン(TAG)があり、それに続いてクラゲ、Aequorea victoriaの緑色蛍光タンパク質(GFP)をコードする遺伝子が存在する。配列決定によりpMM106Hの構築を確認した。

## 【0047】

NcoI / 平滑末端断片としてpMM106Hにクローニングした遺伝子は、クローニングの際に、HpaIサイトに正確なリーディングフレームが作出された場合にしか、His-タグ及びGFPへの融合物を生じない。ここでは、リポーター遺伝子としてGFPを使用し、Hisタグを発現しているクローンの可視的スクリーニングを容易にし、同時に、GFPは正確なコンフメーションに折り畳まれた時にのみ活性を有するので、GFPは融合タンパク質の正確な折り畳みの指標となる。アンバー停止コドンにより、緑色を呈するコロニーを可視化する全長融合タンパク質が少量得られるが、殆どの融合タンパク質はHisタグの直後で終わり、それら融合タンパク質を次の固定化及び酵素分析に使用することができる。タンパク質の多くの様々なペプチドのうち、その可溶性発現物がある観察可能な表現型を細胞に付与するペプチドを、タグ標識タンパク質の発現及び折り畳みのマークとしてGFPの代わりに使用できるであろうことが理解されるべきである。これらには、クロラムフェニコールアセチルトランスフェラーゼ、-ガラクトシダーゼ、-ガラクトシダーゼのlacZ断片、及び-CIリプレッサーのような転写を抑制し得るタンパク質が含まれるが、これらに限定されない。

## 【0048】

下記に概説されている方法において使用する鋳型は、pGSTNであった。標準的条件下で、プライマー「GST fwd 2」(5' - ATG CTG CAG ACG TCA ACA GTA TCC ATG GCC CCT ATA CTA GG - 3')及び「GST Hind III」(5' - GCG AGG AAG CTT GTC AAT CAG TCA CGA TGA ATT CCC G - 3')を用いて、pGEX-2T(Pharmacia)からSchistosoma japonicumのグルタチオンSトランスフェラーゼ(GST)遺伝子に第一のPCR増幅を行うことによって、このプラスミドを構築した。これらのプライマーは、GSTの開始コドンにNcoI制限サイトを導入し、GSTの第二の残基をセリンからアラニンに変異させ、HindIII制限サイトが後置されたGST遺伝子のマルチクローニングサイト3' - に停止コドンを導入する。次に、標準的条件下で、NcoI / HindIII断片として、前もってNcoI / HindIIIで消化したpTrcHisA(Invitrogen)にPCR産物をクローニングし、pGSTNを作製した。

## 【0049】

## (b) タギング前の遺伝子のPCR増幅及びエキソヌクレアーゼ消化(図1b参照)

本発明者らは、ベクターの開始コドンから156bp上流及び停止コドンの84bp下流にそれぞれ結合する特注設計したベクター特異的プライマー「ST forward」(5' - ATG CTG ACG TCA TGA GGC CCA TGG GGC CCC G GAT AAC AAT TTC ACA CAG G - 3')及び「ST reverse」(5' - GCG GAT CCT TGC GGC CGC CAG GCA AAT TCT GTT T - 3')を用いたポリメラーゼ連鎖反応を利用してコンスト

10

20

40

50

ラクトpGSTNからGST遺伝子を増幅した。4個の別々の100μl反応系の中で、30サイクル(94で1分、57で1分、72で2分)のPCRを行った。各PCR反応系には、~20ngの錆型DNA、50pmolの各プライマー、及び2.5ユニットのPwoポリメラーゼが含有されていた。標準的緩衝液中(10mM Tris·HCl pH 8.8、25mM KCl、5mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>、2mM MgSO<sub>4</sub>、10%DMSO)中で各PCR反応を行った。4つの前記PCR反応は、それぞれ、以下の非標準的デオキシヌクレオチドトリホスフェート混合液も含有していた。

## 【0050】

反応1) 200μM dATP、200μM dTTP、200μM dCTP、150μM dGTP、50μM -S-dGTP、

10

反応2) 200μM dATP、200μM dTTP、200μM dGTP、150μM dCTP、50μM -S-dCTP、

反応3) 200μM dATP、200μM dGTP、200μM dCTP、150μM dTTP、50μM -S-dTTP、

反応4) 200μM dGTP、200μM dTTP、200μM dCTP、150μM dATP、50μM -S-dATP

言うまでもなく、3'・5'エキソヌクレアーゼ活性を欠失したTapポリメラーゼのような温度安定性ポリメラーゼ及び、T4 DNAポリメラーゼ又はDNAポリメラーゼIのクレノー断片などの非温度安定性ポリメラーゼを含む多くの異なる種類のDNAポリメラーゼを用いたプライマー伸長反応により、-S-dNTP存在下での錆型DNAの増幅を行うことができる。

## 【0051】

各特異的PCR混合液に単一の-チオデオキシヌクレオチドトリホスフェートを添加することにより、適切な-S-dNTPが、ランダムであるが統計学的に、特定の最終PCR産物中に取り込まれる。これらの改変ヌクレオチドは、エキソヌクレアーゼIIの基質ではなく、酵素によるヌクレオチド除去の進行を停止させるために使用される。4種類の個々のPCR混合液を集め、標準的条件下で、QIAquick PCR clean up kit (Qiaagen)を用いて精製し、制限酵素AatIIで完全に消化した。次に、得られた~1000bpのPCR産物をゲル精製した。AatIIを用いた制限酵素切断により、エキソヌクレアーゼII活性で消化されない3'-突出部が得られるため、前記PCR産物の5'末端は分解から保護される。

## 【0052】

エキソヌクレアーゼ消化から特異的にPCR産物の片方の末端を保護するこれ以外の方法は以下のものを含め容易に挙げることができるが、これらに限定されない。必要なサイトがPCRプライマーのデザイン中に取り込まれる限り、4塩基以上の3'-突出部を生ぜしめる任意の制限酵素をAatIIの替わりに使用することが可能であろう。5'-突出部を形成する任意の制限酵素も、必要なサイトがPCRプライマーのデザインに取り込まれる限り、AatIIの替わりに使用することが可能であろう。この場合、PCR産物の新しい3'-末端がエキソヌクレアーゼ消化から保護されるように、dNTPの替わりに適切な-チオdNTPを使用するDNA-ポリメラーゼ媒介性の埋め込み反応(filin reaction)が5'-突出部の形成に続いて行われるであろう。

## 【0053】

次に、150μlの反応系で、消化した10~15μgの前記PCR産物を75unitのエキソヌクレアーゼII/μgDNAと共に、37で30分間インキュベーションした。標準的反応緩衝液(66mM Tris·HCl pH 8.0、6.6mM MgCl<sub>2</sub>、5mM DTT、50μg/mlウシ血清アルブミン)でExoIII消化を行った。これらの条件により、確実に、ExoIIIによる消化が完全に行われる。次に、75まで15分間加熱して酵素を不活性化した。ExoIII消化の産物は、前記PCR産物の3'から得られる一群の入れ子型欠失物である。

## 【0054】

50

エキソヌクレアーゼ I I I は - チオ - 含有ヌクレオチドを加水分解することができない非進行的な 3' 5' エキソヌクレアーゼなので、現在のプロトコールでは、E x o I I I が - チオ - デオキシヌクレオチド塩基に到達するごとに、P C R 産物の 3' 陥凹末端における末端切断の進行が停止する。従って、最終的に得られるのは、これより前の段階で各 - S - d N T P をランダムに取り込んだ結果生じる一群の入れ子型欠失物である。最初のP C R 増幅で使用するd N T P に対する - S - d N T P の比率は、入れ子型欠失物のエンベロープが、元の完全長P C R 産物より約1 0 0 b p 短く中心に位置したサイズ4 0 0 b p のウインドウにわたるように経験的に決定した。

#### 【 0 0 5 5 】

通常のd N T P に対する - S - d N T P の比率を変化させることにより、得られる末端切断の大きさの範囲を調節し得る。真核c D N A は最も一般的な3' - U T R 長が2 0 0 ~ 3 0 0 b p の範囲に入るような様々な長さの3' 非翻訳領域を有するので、これは、本方法を真核c D N A に適用する際に重要である。ポリメラーゼの種類によって4 つのが - S - d N T P の相対的取り込み効率が変化するので、4 つの各塩基及び特定のポリメラーゼに最適化された - S - d N T P / 通常d N T P 比率を使用することが望ましい。典型的には、使用される通常のd N T P に対するラセミ体 - S - d N T P のモル比は、1 : 1 乃至 1 : 3 の範囲内に入るであろう。

#### 【 0 0 5 6 】

##### ( c ) 一本鎖領域の除去及びクローニングの準備( 図1 c を参照 )

エタノール沈殿により、前の段階でエキソヌクレアーゼ I I I 消化により作製した一群の入れ子型欠失物を精製し、1 × リヨクトウヌクレアーゼ緩衝液( 5 0 m M 酢酸ナトリウム pH 5 . 0 、 3 0 m M N a C l 、 1 m M Z n S O \_ 4 ) で再懸濁した。 1 0 0 μ l の反応系中で、 3 0 にて 3 0 分間( 2 u n i t s / μ g ) 3 0 U n i t s のリヨクトウヌクレアーゼを用いて、消化した前記D N A を処理した。この工程により、 5 ' - 及び 3 ' - 突出末端は除去され、平滑末端産物を得た。最終濃度が 5 m M になるようにE D T A を添加して反応を停止した。Q I A q u i c k P C R p u r i f i c a t i o n k i t ( Q u i a g e n ) を用いて、消化した前記産物を精製し、N c o I で消化し、標準物質として 1 0 0 b p D N A ラダーを用いて 1 % のアガロース / T B E ゲルで分離した。Q I A q u i c k g e l e x t r a c t i o n k i t ( Q u i a g e n ) を用いてアガロースゲルから大きさが 8 0 0 ~ 1 0 0 0 b p の範囲の産物を抽出した。明らかに、S 1 ヌクレアーゼのような他の一本鎖ヌクレアーゼを使用して、3' 5' - エキソヌクレアーゼにより生成した一群の3' - 入れ子型欠失物から 5 ' - 突出部を除去することもできる。

#### 【 0 0 5 7 】

##### ( d ) 入れ子型欠失物の調製の変法

一群の入れ子型欠失物を作製するための多くの他の標準的分子生物学的手法が本来の手技の変法となることが明らかである。これらには、任意の 3' 5' エキソヌクレアーゼ、任意の 5' 3' エキソヌクレアーゼ、又は線状D N A 断片の末端から漸進的に末端切断する任意のエンドヌクレアーゼの使用が含まれるが、これらに限定されない。例えば、G S T 遺伝子の開始位置からおよそ 2 k b 上流に結合するフォワードプライマーとともに、上記と同じリバースプライマーを使用して、最初のP C R 増幅を行うことができる。これにより、G S T 遺伝子の 5 ' 末端に > 2 k b が、 3 ' 端には僅かに 8 4 b p が隣接した断片が生じるであろう。次に、精製したP C R 断片をB a l 3 1 ヌクレアーゼで処理することが可能であり、それにより 5 ' - 末端及び 3 ' - 末端の両方から漸進的に線状の二本鎖D N A が分解される。この酵素は非進行的であり、D N A 分解速度は、D N A の塩基組成だけでなく反応時間及び温度に依存する。P C R 産物中のG S T 遺伝子の 3 ' 末端のフランкиング領域が 5 ' 末端のフランкиング領域よりも著しく短いので、他方の末端から開始コドンに到達するかなり前に、停止コドンまで、及び停止コドンを越えて分解が進んでいるであろう。時間的経過を調べる実験により、P C R 産物の 3 ' - 端から 4 0 0 b p までの長さを除去する至適反応条件を決定できるようになる。続いて、得られた一群の入れ

10

20

30

40

50

子型欠失物を平滑末端にし、残存する一本鎖領域を全て除去し、元のベクターによって前記遺伝子の5'末端にコードされたユニークな制限酵素で消化し、タグベクターに方向性を持ってクローニングした。或いは、ラムダエキソヌクレアーゼを使用して一群の入れ子型5' - 欠失物を生ぜしめることが可能である。この酵素に対する好ましい基質は5'リノ酸化二本鎖DNAであるので、5'ヒドロキシ末端を有することによってDNA基質の片方の末端を容易に保護することができる。T4 DNAポリメラーゼ又は、RNase T又はエキソヌクレアーゼT又はリヨクトウヌクレアーゼなどの一本鎖DNA特異的ヌクレアーゼを含む多く様々な酵素により、5' 3' - エキソヌクレアーゼにより生じた一群の入れ子型5' - 欠失物の一本鎖3'突出部を除去し得る。

## 【0058】

10

(e) 改変産物のクローニング及び分析(図1c参照)

ベクターpMM106H(3μg)を制限酵素NcoI及びHpaIを用いて完全に消化し、2870bpのバックボーン断片をゲル精製した。標準的条件下でベクターDNA及び上記のように調製した制限産物をライゲーションし、ライゲーション混合液を用いて、E.coli DH5細胞を形質転換し、次にこれを回収して100μg/mlのカルベニシリンを含有するLBプレートに播種した。

## 【0059】

20

前の段階で得られた一群の欠失物全てに対してこのクローニング操作を遂行した。しかし、GST遺伝子の停止コドンを切斷しインフレームコドンの直後で終わる欠失物のみが、この操作によるクローニング段階後にヘキサヒスチジンタグ及びGFPに対するインフレーム融合物を生じさせ得るはずであり、このようにしてクローニングした他の欠失産物は全て、ヘキサヒスチジンタグ及びGFPにへの読み替からはずれた融合物か、GST停止コドンでの翻訳の終結に起因する非融合GSTタンパク質のみを生じせしめるはずである。これは、ベクターの平滑末端に欠失産物の平滑末端がライゲーションすることによって、下流のベクターDNAの翻訳リーディングフレームがGSTコード領域の元のリーディングフレームにより支配される遺伝子融合が起こるからである。前記欠失産物がGST停止コドンを保持していれば、ヘキサヒスチジンタグへのGSTの翻訳可能な融合は起こらないと思われるのに対して、前記欠失産物が不完全なコドンで停止すれば、新しく付加されたヘキサヒスチジン - コード配列は、GST遺伝子に関して読み替からはずれたものとなろう。従って、上記工程全体から生じ得るヘキサヒスチジン - (及びGFP-)タグ標識タンパク質のみが必然的にポリアスパラギン、ヘキサヒスチジンタグに対するGST融合物となるであろう。これらは、必ずしも完全長のクローナーではないであろうが、さらなる工程において、それらが正確に折り畳みを行う能力及び酵素活性の保持についてのスクリーニングが行われることになろう。

30

## 【0060】

40

形質転換を行ったコロニーをUV光(365nm)下で可視化し、さらに分析を行うために、緑色蛍光を発する30コロニー(全体のおよそ10%)を肉眼で選択した。これらのコロニーのレプリカ - プレート作製し、抗-His-tag及び抗-GST抗体を用いて標準的条件下でコロニーウエスタンプロットによって分析した。抗-His-tag抗体は、ヘキサヒスチジン - タグ標識されたタンパク質を発現するコロニーのみと結合するので、ウエスタンプロットによりヘキサヒスチジン - タグに対するインフレーム融合物を発現するコロニー数に関する情報を直接得ることができる。一方、抗-GST抗体は、GSTタンパク質のC - 末端近傍に結合するため、全長又はほぼ全長のGSTタンパク質を発現するコロニーのみを認識する。本発明者らは、抗-His-tag及び抗-GST抗体の両者によってポジティブであると認識されたタンパク質を含有する19個のコロニー(緑色蛍光を発するコロニーのうち63%)を同定した。これらコロニーのうち、12個のコロニー由来のDNAを増幅し、精製して配列決定を行った。配列決定データにより、全長GSTに対する完璧な2つのインフレーム融合物2及び、GST遺伝子中に短い末端切断を有するが、ヘキサヒスチジンタグに関してはなおインフレームである10個のクローナーの存在が確認された。従って、この操作全体によって本発明者らが得た全長GSTクロ

50

ーンの単離頻度は、(緑色蛍光を発するコロニーの総数の)およそ17%であるのに対して、全長又はほぼ全長で、活性を保持すると予想されるGSTクローンの単離頻度は、(緑色蛍光を発するコロニーの総数の)およそ63%である。

#### 【0061】

(f) タグ付加タンパク質の固定化及び機能分析(図1d参照)

E.coli DH5 細胞に、上記の方法により作製した全長のヘキサヒスチジンタグ付加GSTプラスミドのうちの1つを形質転換した。10mlの液体培地の中で、単一のカルベニシリン-抵抗性コロニーを中央対数増殖期まで増殖させ、次に100μMのIPTGを補充してヘキサヒスチジン-タグ付加GSTの発現を誘導した。さらに4時間培養を行った後、細胞を回収し凍結融解/ライソザイムにより溶菌させた。未精製溶菌液のSDS-PAGEによって、予想したサイズ(27kDa)にタンパク質が過剰発現しており(総可溶性タンパク質のうちおよそ20%を占めた)、アンバー抑制によって生じた54kDaのGST-ヘキサヒスチジン-GFP融合タンパク質も少量存在することが示された。次に、未精製溶菌液(500μl、100μg)をニッケル-NTAマグネチックビーズ(50μl、結合容量 15μg ヘキサヒスチジン-タグ標識タンパク質)と混合し、磁場をかけてビーズを沈降させて回収した。上清を捨て、ビーズを洗浄した後、グルタチオン及び1-クロロ-2,4-ジニトロベンゼンをそれぞれ1mMずつ含有するグルタチオンSトランスフェラーゼアッセイ用緩衝液の中に再懸濁した。室温で30分後に、340nmでの吸光度を測定してエンドポイントアッセイのデータを集めた。この340nmという波長は、GST-触媒反応産物のmaxに相当する。

10

20

20

#### 【0062】

対照として、親ベクター(pMM106H)又は無関係なHis-タグ付加タンパク質(アラニンラセマーゼ)をコードするプラスミドの何れかを含有するDH5を培養、誘導、回収、溶菌し、平行してアッセイを行った。His-タグ付加GSTを含有する未精製溶菌液と混合しておいたビーズのみにGST活性が検出され、そ観察されたGST活性が専ら固定化したHis-タグ付加GSTによるものであり、さらにそのタンパク質が特異的固定化において活性を保持することが明らかに示している。

#### 【0063】

酵素アッセイ完了後、100mM イミダゾールを含有する緩衝液を添加してマグネチックビーズからタンパク質を溶出し、SDS-PAGEにより分析を行った。これにより、ポジティブな活性アッセイの結果を与えた試料がグルタチオンSトランスフェラーゼ(27kDa)に対して予想される正確な分子量を有する単一の固定化タンパク質を含有することが示され、従って、ビーズに対して観察された活性がこの組換えHis-タグ付加タンパク質のみによるものであると確認された。

30

#### 【0064】

例2

##### (a) ベクターの構築

本発明者らは、第二のベクターであるpMM111を構築した。このpMM111は、676bpのNcoI/HpaIナンセンスDNA詰め込み断片がEscherichia coli gdhA遺伝子由来の300bpのNcoI/HpaI断片と置換され、HpaIクローニングサイトがSmaIサイトで置換され、下流のヘキサヒスチジンタグがgdhA遺伝子により2ヌクレオチド分フレーム外に移動するように位置しており、GFP遺伝子のATG開始コドンがアラニン(GCG)のコドンに置換されていること以外は、pMM106H(例1参照)と基本的に同様である。SmaIサイトにクローニングされた挿入断片が、その3'末端においてコドンの第一のヌクレオチドを必ず含有して、その挿入断片がヘキサヒスチジンタグ及びGFPと共にフレーム内に置かれるように前記ベクターをデザインしている。pMM111の構築を配列決定により確認した。

40

#### 【0065】

##### (b) タグを導入するための改変方法

次に、本発明者らは、以下の改変を除き、例1にて記述したのと同様の操作を行った。ま

50

ず、オリジナルの P C R 増幅、即ちセクション( b )の反応番号 3 だけに - S - d T T P を取り込ませた。次に、ベクター p M M 1 1 1 の N c o I 及び S m a I サイトに最終産物をクローニングした。

#### 【 0 0 6 6 】

この操作には、例 1 で記述した操作よりも理論的に有利な点がいくつか存在する。これらは、主にオリジナルの P C R 産物に単一の - チオ d N T P が取り込まれる際の統計に起因するものである。従って、徹底的なエキソヌクレアーゼ I I I 消化を行うと、全ての入れ子型 3 ' 陥凹欠失物群は、4 個のうち何れかのヌクレオチドで終結するのではなく、3 ' - チミジン塩基で終結するであろう。3 ' - T が、G S T 遺伝子の第一のインフレーム停止コドン中の塩基であるか、又はこれより前に位置し、且つ第一のインフレーム停止コドンの「 T 」と同じリーディングフレーム中にある場合には、p M M 1 1 1 の S m a I サイトへこれらの断片をクローニングすることによって、ヘキサヒスチジンタグ及び G F P に対するインフレーム融合のみが起こるであろう(これは、S m a I による消化によって、タグのコード配列の前に 2 ヌクレオチドのギャップが生じるからである。)。

#### 【 0 0 6 7 】

統計的には、この改変操作におけるエキソヌクレアーゼ加水分解によって得られる入れ子型欠失物は例 1 の 1 / 4 であろう。しかし、停止コドンとなり得る 3 つのコドンは全てそれらの最初の塩基として「 T 」を含有するので、それらは全て前記欠失物群中に含まれることとなり、それ故、完全な欠失物群の 4 倍高い割合を占めるであろう。任意のある「 T 」が、停止コドンの第一の「 T 」と同じリーディングフレームである可能性があるならば、この改変操作から得られる全クローンのうち 3 3 % が、ベクターによりコードされる H i s タグへのインフレーム融合物となるはずであるが、折り畳み(その結果として機能に)に影響を及ぼすそのような欠失物からは(例 1 で述べた理由により)「白い」コロニーが生じる。本発明者らは、この改変操作に続いて、「緑色」のコロニー集団中の正確な完全長のクローンの割合が、例 1 の操作によって見いただされた割合より有意に高いということを見出した。既知の開始コドン( A T G , G T G , T T G , A T T , C T G ) 全てが第二番目に「 T 」を含有するので、言うまでもなく、同様の議論が「開始」コドンに対しても当てはまる。

#### 【 0 0 6 8 】

この改変操作の更なる利点は、最初の P C R 増幅時に使用するフォワードプライマーの 5 ' 末端にポリ A テイルを取り込みうることである(例えば、フォワード A 5 ' - A A A A A A A A A A G A T C G A T C T C A T G A C G G A T A A C A A T T T C A C A C A G G - 3 ' )。3 : 1 の d T T P : - S - d T T P 比で増幅を行う間に、最低 1 個の - S - d T T P 残基が、P C R 産物の 5 ' - 末端で、相補鎖の末端に取り込まれる可能性が高いであろう。取り込まれたこれらのヌクレオチドは、エキソヌクレアーゼ I I I 消化に対して抵抗性を有するので、酵素反応段階において P C R 産物の 5 ' - 末端を分解から特に保護する必要がないであろう。

#### 【 0 0 6 9 】

例 3

##### ( a ) ヘキサヒスチジンタグを用いた第 2 のタンパク質の改変

例 1 で記述したグルタチオン - S - トランスフェラーゼに対する操作に続いて、本発明者らは、この操作が、操作対象の遺伝子配列に依存しないことを示した。

#### 【 0 0 7 0 】

従って、ヒト転写因子である N F - B p 5 0 をコードするプラスミドを用いて開始し、特に断っていない限り例 1 で記述した操作に正確に従うことによって、本発明者らは、第一のインフレーム停止コドンが切断され、ポリアスパラギン、ヘキサヒスチジンタグ及び G F P をコードする D N A へのインフレーム融合によって置換されるように、N F - B p 5 0 の改変を行うことができた(アンバー停止コドンが抑制される場合)。遠紫外光( 3 6 5 n m )で励起させると緑色蛍光を発するクローンの性質をさらに調べた。抗 - H i s - T a g 抗体を用いたコロニーウエスタンプロットにより、ヘキサヘキスチジン -

10

20

30

40

50

タグ標識タンパク質を発現するクローンの同定を行うことができた。これらのクローンの可溶性タンパク質溶菌液を SDS - ポリアクリルアミドゲル電気泳動にて分離して、抗-His タグ抗体で調べた。およそ 65 kDa Mr (ヘキサヒスチジンタグ及び GFP に翻訳可能に融合した NF - B p50 に相当する)、及びおよそ 38 kDa Mr (NF - B p50 - His タグ) で免疫反応シグナルを観察した。さらに、27 kDa Mr 近辺にシグナルが観察されたが、これはおそらく His タグ標識 GFP タンパク質に対応する分解産物である。配列決定データから、ヘキサヒスチジンタグに対して全長又はほぼ全長の NF - B の完全なインフレーム融合物をコードするクローンがいくつか存在することが確認された。一回の実験で、緑色蛍光について 190 個のコロニーをスクリーニングした。遠紫外線光 (365 nm) で励起したところ、全部で 38 個 (スクリーニングを行った全クローン数の 20 %) が緑色蛍光を発した。抗-His tag 抗体を用いたコロニー-ウェスタンプロットティングを行ったところ、38 個のクローンのうち 29 個がヘキサヒスチジンタグを発現することが明らかになった。配列決定データにより、これらのクローンのうち 18 個が、ヘキサヒスチジンタグに対する NF - B p50 の全長又はほぼ全長のインフレーム融合物であり、これらのクローンのうち 7 個が完全に全長のヒストグ付加 NF - B p50 遺伝子であり、残りの 11 個の His - タグ標識クローンは 4 ~ 1 アミノ酸残基の短い末端切断を有していることが確認された。

10

## 【0071】

この実験から、タンパク質の発現及び正確な折り畳みを表すリポーター系を有することによる利点が明らかに示される。緑色蛍光を発するクローンのうち大体 50 % が His - タグへの完全長インフレーム融合物であるか、又はドメイン境界にまたがらない微小な末端切断を有し、His - タグに対してインフレームに融合するものであった。

20

## 【0072】

(b) ヘキサヒスチジンタグ付加 NF - B p50 の固定化及び機能分析

上記方法により作製した全長ヘキサヒスチジンタグ付加 NF - B プラスミドの一つを E. coli DH5 細胞に形質転換した。10 ml の液体培養系で 1 個のカルベニシリソ抵抗性コロニーを中心対数増殖期まで増殖させ、次に 100 μM の IPTG を添加してヘキサヒスチジンタグ付加 NF - B p50 の発現を誘導した。さらに 4 時間培養した後、細胞を回収し超音波破碎により溶菌した。未精製溶菌液の SDS - PAGE により、予想した分子量 (38 kDa) でタンパク質が過剰発現しており、これは全可溶性タンパク質のおよそ 5 % であることが示された。

30

## 【0073】

B モチーフ

5' - C G T	A T G	T T G	T G G	G G A	A T T	C C C	A G C	G G A	T
A A	C - 3'								
3' - G C A	T A C	A A C	A C C	<u>C C T</u>	<u>T A A</u>	<u>G G G</u>	T C G	C C T	A
T T	G - 5'								

NF - B p50 結合サイト

NF - B p50 に対するパリンドローム結合サイトを含有する二重鎖オリゴヌクレオチド「B モチーフ」を、標準的条件下で 3 - ターミナルトランスフェラーゼを用いて 3' - 塩基にジゴキシゲニンで標識した。

40

## 【0074】

5 mM - メルカプトエタノールを含有する PBS (リン酸塩類緩衝液、pH 7.5) 中でライソザイム / 凍結 - 融解法を用いてタンパク質溶菌液を調製した。各クローン由來の 200 μl の可溶性タンパク質溶菌液を Ni - NTA 被覆マイクロウェルに添加して、室温にて 45 分間インキュベーションを行った。インキュベーション終了時に、PBST (0.02% Triton X-100 を含有する PBS) でウェルを 3 回洗浄して未結合タンパク質を全て除去した。DNA 結合緩衝液 (5 mM - メルカプトエタノールを含有する 10 mM Tris . HCl、pH 7.4, 75 mM KCl) を添加して、1 分間浸漬し、そのウェルを 3 回洗浄した。1 μg のポリ(dI - dC) 非 - 特異的 DNA を

50

含有する 200 μl の DNA 結合緩衝液が入ったウェルに 3' ジゴキシゲニン標識した B モチーフ (2 pmol) を添加した。さらに 30 分のインキュベーションを行った後に、0.02% の Triton X-100 を含有する 10 mM Tris·HCl、pH 7.4、25 mM KCl でウェルを 3 回洗浄して未結合 DNA を除去した。150 mU/ml になるように、0.2% ウシ血清アルブミンを補充した「抗体希釀緩衝液」(10 mM Tris·HCl、pH 7.4、25 mM KCl) 中にアルカリホスファターゼ標識抗 - ジゴキシゲニン抗体を、希釀した。次に、希釀した抗体 (200 μl) をマイクロウェルに添加した。室温で 30 分間置いた後、0.02% Triton X-100 を補充した「抗体希釀緩衝液」でマイクロウェルを洗浄して (3 × 350 μl) 未結合抗体を除去した。次に、250 μM の p - ニトロフェニルホスフェート (pNPP) (アルカリホスファターゼの基質) を含有する 200 μl の緩衝液 (100 mM Tris·HCl、pH 9.5, 100 mM NaCl、50 mM MgCl<sub>2</sub>) をウェルに添加し、室温で一晩反応を進行させた後、405 nm で各ウェルでの黄色の呈色を定量した (反応物である p - ニトロフェノールの形成に対応する)。基質 pNPP の加水分解に対するバックグラウンドの割合は低かったため、従って、ポジティブなアッセイ結果がウェル中における黄色の呈色由来のものであることは明らかであった。このアッセイの対照として、本発明者らは、未精製ライセート又は標識したオリゴヌクレオチド又は抗体の何れかを除去するか、又は 20 倍過剰の非標識二重鎖オリゴを添加するか、又はヘキサヒスチジンタグ付加 NF-B p50 を含有する未精製ライセートを、同じベクターバックグラウンド中にヘキサヒスチジン - タグ付加 GST を発現する DH5 細胞由来の等量の未精製細胞ライセートと置換した。  
10  
20

## 【0075】

この分析において、NF-B p50 はまず特異的結合サイトにおいて、標識したオリゴヌクレオチドと結合する。次に、ヘキサヒスチジンタグを介してマイクロウェル中で、タンパク質 - DNA 複合体を固定化し、全ての未結合標識オリゴと共に他の全タンパク質 (標識したオリゴと未精製ライセート中に存在する他の DNA 結合タンパク質との複合体を含む) を洗浄除去する。本標識抗体はオリゴ上の標識を認識し、ヘキサヒスチジンタグ標識タンパク質を認識しないので、本アッセイにおけるポジティブなシグナルは、NF-B p50 - DNA 相互作用がタグを介した NF-B p50 の固定化に維持される場合にのみ観察され得る。この相互作用が維持されない場合は、このオリゴは洗浄段階で除去されるので呈色変化は見られないであろう。  
30

## 【0076】

本発明者らは、ヘキサヒスチジン - タグ付加 NF-B p50 未精製溶菌液とジゴキシゲニン標識オリゴヌクレオチドとを含有し、アルカリホスファターゼ標識抗ジゴキシゲニン抗体が添加されたマイクロウェル中で黄色を呈する産物のみが検出されることを見出した。このことは、観察される色の変化が、専ら固定化された NF-B p50 - オリゴヌクレオチド複合体によるものであることと、さらに、特異的固定化の際に NF-B p50 が活性を保持することを示していた。

## 【0077】

例 4

40

## (a) 10 個の遺伝子のプールからの 1 個のタンパク質の同定

本発明者らは、以下の表に挙げた 10 個の種々の遺伝子を含む遺伝子プールに対して、明示した場合を除いて、例 1 で記述したように正確に前記操作を遂行した。本発明者らは、この操作を用いた結果、アレイにおける各位置が、標識したタグを介して固定化された一個の組換えタンパク質に対応するように、得られた特異的改変タンパク質のアレイを作製した。次に本発明者らは、機能的アッセイによって本アレイをスクリーニングし、前記プールの個々のタンパク質成分を首尾よく同定した。

## 【0078】

## 【表 1】

## 遺伝子プール内の、10個の遺伝子の大きさ及び機能

遺伝子	大きさ	由来及び機能
グルタチオンSトランスフェラーゼ	950bp	細菌、解毒
NF- $\kappa$ B p50	1165bp	ヒト、転写因子
マルトース結合タンパク質	1325bp	細菌、糖質トランスポーター
アラニンラセマーゼ	1342bp	細菌、細胞壁生合成
T細胞活性化因子 (NFAT)	1087bp	マウス、転写因子
インドールグリセロール ホスフェートシンターゼ	1528bp	細菌、アミノ酸生合成
ホスホリボシルアントラニル酸 イソメラーゼ	920bp	細菌、アミノ酸生合成
トリプトファンシンターゼ ( $\alpha$ -サブユニット)	1122bp	細菌、アミノ酸生合成
キモトリプシンインヒビター2	389bp	オオムギ、セリンプロテアーゼインヒビター
$\beta$ -ラクタマーゼ	1040bp	細菌、抗生物質抵抗性

10

まず、同じ p T r c H i s A ベクターバックボーンに 10 個の遺伝子全てをサブクローニングしたが、これはとりわけ、cDNA ライブラリーにより起こる状況を模倣するからである。p T r c H i s A ベクターバックボーン内にコードされる遺伝子の増幅に対して、例 1 で記述したプライマー「 S T f o r w a r d 」及び「 S T r e v e r s e 」がユニバーサルプライマーになるように設計した。

## 【 0 0 7 9 】

次のように多数の制限酵素サイトをコードするように、プライマー「 S T f o r w a r d 」を設計した。

## 【 0 0 8 0 】

5' - A T G C T G A C G T C A T G A G G C C C A T G G G G C C C G G A  
T A A C A A T T T C A C A C A G G - 3'  
A a t I I B s p H I S f i I

30

このように、制限酵素 A a t I I 又は S f i I の何れかを使用して、エキソヌクレアーゼからの保護を目的として 3' - 突出部を形成させ得る。本改変操作の最終段階で、方向が決まったクローニングを行うために、この例において、本発明者らは B s p H I を選択したが、これは統計的には B s p H I はライブラリー中で切断する頻度が高いが、ここで使用するタグベクター p M M 1 0 6 H の N c o I クローニングサイトと合致する粘着末端を生ぜしめ、本プール中に存在する 11 個の何れの遺伝子中でも切断しないからである。明らかに、原理的には、タグベクターがプロモーター下流に同等のクローニングサイトを含有していれば、制限サイトをコードする任意のプライマーを使用することができるであろう。S f i I の認識配列は 8 bp であり、そのため、ある遺伝子内での S f i I サイトのランダムな発生頻度は、B s p H I のように認識部位が 6 bp であるもの ( 4 , 0 9 6 のうち 1 ) よりもかなり低い ( 6 . 5 X 1 0 <sup>4</sup> のうち 1 ) と思われる所以、この点で、より大きいライブラリー形式の場合 S f i I は非常に大きな利点を有することになるだろう。タグベクター p M M 1 0 6 H は「 A T G ベクター」である、つまり 5' - クローニングサイト ( N c o I ) が、ネイティブタンパク質を発現するためのリボソーム結合サイト ( R B S ) の下流に位置する A T G 開始コドンと重なり合っている。しかしながら、本明細書に記述した操作において、本発明者らは、開始コドンに共通の制限サイトを有するクローニング遺伝子に依存してはいない。それよりも、本発明者らは、翻訳の開始がクローニングされた遺伝子そのものによってもたらされるために必要なシグナルにより転写を開始して m R N A を生成させる、ベクターにコードされたプロモーターのみに依存する。従つ

40

50

て、この例で、元のプールの全遺伝子は、A T G におけるクローニングサイトの有無に関わりなく、R B S が直前に位置している開始コドンを有する。初めの 11 個のクローン全てにおいてプライマー「S T f o r w a r d」が R B S の上流に結合するので、プライマーにコードされる任意の制限サイトを用いて行うその後の改変後クローニングによって、新しく改変された遺伝子が、その元の R B S 及び A T G と共にタグベクターに導入され、それにより翻訳開始が確実となるであろう。c D N A ライブライマー形式において、全ての全長 c D N A が、真核細胞翻訳開始シグナルを含有するそれ自身の 5' - 非翻訳領域 (U T R ) を有するという場合に同様の状況が当てはまる。正確な翻訳開始に必要であるのは、転写開始シグナルを与える真核ベクターに対して、各改変 c D N A をその 5' - U T R と共にクローニングすることのみであるので、配列非依存的に、単一容器内の c D N A ライブライマーの全要素を改変するに際して、この例で使用したものと同等の P C R プライマーのユニバーサルセットを使用できるだろう。

10

20

30

**【0081】**  
次に挙げる改変を加え、例 1 で記述したように本実験操作を遂行した。プライマー「S T f o r w a r d」、「S T r e v e r s e」を使用する最初の P C R 増幅に対する鋳型として 10 個の遺伝子全てを等モルずつ含有するプールを使用した後、5' 末端を保護するために A a t I I を用いて断片を消化した。例 1 のように正確にエキソヌクレアーゼ I I I 及びリヨクトウヌクレアーゼ - 処理断片を生ぜしめ、フォワード P C R プライマー結合サイト内でユニークに断片を制限消化し、ベクター p M M 1 0 6 H にクローニングするために粘着末端を生ぜしめる B s p H I で消化した。得られた断片をゲル精製し、該ベクターにライゲーションした。形質転換した細胞を U V 光 (365 nm) で可視化し、ウェスタンプロットによる分析を行うために、緑色蛍光発光コロニーを肉眼で選択した。形質転換した全コロニーのうちおよそ 2 % が緑色蛍光を発した。これらのうち、103 個 (42 %) が、抗 - H i s タグ抗体により認識されるタンパク質を発現した。これらのコロニーを、96 穴のディープウェルブロック中の 1 . 5 m l の液体培地に各別に植え込み、一晩培養した。遠心によって細胞を回収し、凍結 - 融解 / ライソザイムにより溶菌した。次に個々の未精製溶菌液を、ニッケル - N T A - 被覆した 96 ウェルプレートのそれぞれ別個のウェルに添加して、未結合タンパク質を洗浄により除去したが、各ウェルに固定化した個々の H i s タグ付加タンパク質は残存した。次に、例 1 及び 3 で記述したアッセイを用いて、固定化タンパク質の G S T 又は N F - B 活性の何れかをアッセイし、各ケースにおいて、それぞれ緑色又は黄色の何れかの呈色によって、ポジティブな「ヒット」を示したウェルを同定した。

30

**【0082】**

最初のアッセイでは、アレイ中の 3 個のタンパク質が G S T 活性を示し、およそ 3 % のヒット率であることが分かった。対応するプラスミドの配列から、3 個全てが G S T 遺伝子とヘキサヒスチジンタグと G F P 遺伝子との間のインフレーム融合をコードしており、3 つのうち 2 つは完全長 G S T であり、1 つは明らかに活性には影響を与えないごく短い末端切断が見られることが明らかになった。

40

**【0083】**

第二のアッセイでは、アレイ中の 3 個のタンパク質がポジティブな「B - モチーフ」 D N A 結合活性を示した。さらに、これらポジティブクローンの性状を調べたところ、前記クローンのうち 2 個が、N F - B p 5 0 遺伝子のヘキサヒスチジンタグへのインフレーム融合物であり、そのうち 1 個はほぼ全長（末端のアミノ酸が 1 個切断されていた）であったのに対して、もう一方は末端切断の程度がこれより大きかったものの N F - B p 5 0 の完全な D N A 結合ドメインを含有することが示された。興味深いことに、前記アッセイは、同系の D N A 配列への結合を調べるために設計されているので、改変されず、折り畳まれ、機能を有する D N A 結合ドメインをなお有する末端切断断片は本アッセイではポジティブとなるであろう。第 3 のクローンがマウス転写因子 N F A T の D N A 結合ドメインのヒスタグに対するインフレーム融合であることが明らかとなった。前記アッセイで用いる 3' - ジゴキシゲニン標識「B モチーフ」は、N F - B p 5 0 に対する特

50

異的、高親和性 (K<sub>d</sub> およそ pM) 結合サイトを含有するが、この同じ結合サイトは、およそ nM オーダーの親和性で、特異的に NFKB の DNA 結合ドメインによっても認識される。従って、この結果から、この操作によって作製したアレイで機能を精査することによって、特異的相互作用、及びこれより弱い作用ではあるが、なお特異的であり生物学的に有用な相互作用を何れも同定することができるということが示される。

#### 【 0 0 8 4 】

次の実験において、この例の方法に従って、約 340 の His - 標識タンパク質を含有するアレイを作製した。GST 活性アッセイによるアレイのアッセイにより、アレイ中の全タンパク質のうち 8 % が強い GST 活性を保持するということが示された。さらに、遺伝子特異的プライマーを使用して 340 個のコードプラスミド DNA プールに対して PCR 解析を行い、その結果、各遺伝子が His - タグ標識集団に相当するものであることが示された。従って、これらのデータにより、この例の方法が配列非依存的であり、様々な遺伝子の集団に適用しうるということが更に確認される。10

#### 【 0 0 8 5 】

要約すると、つまり、本発明者らは、これらの例で記述した方法を使用して、マイクロウェル形式で機能的なタンパク質のアレイを作製し、本発明者らはこれらのアレイを用いて特異的タンパク質 - リガンド相互作用 (GST 活性アッセイ) 又は特異的タンパク質 - DNA 相互作用 (NFKB 結合分析) の何れかに基づいて、プールから 3 つの異なるタンパク質を同定することに成功した。

#### 【 0 0 8 6 】

#### [ 配列表 ]

## SEQUENCE LISTING

<110> Sense Proteomic Limited  
Blackburn, Jonathan  
Mulder, Michelle  
Samaddar, Mitali

<120> Method

<130> P30762WO/PWC/NCT

<140> PCT/GB01/03693  
<141> 2001-08-17

10

<150> GB 0020357.0  
<151> 2000-08-17

<150> US 60/247,995  
<151> 2000-11-14

<160> 9

<170> PatentIn version 3.1

<210> 1  
<211> 41  
<212> DNA  
<213> Artificial Sequence

20

<220>  
<223> Primer - GSTfwd2

<220>  
<221> misc\_feature  
<222> (1)..(41)  
<223> Primer

30

<400> 1  
atgctgcaga cgtcaacagt atccatggcc cctatactag g  
41

<210> 2  
<211> 37  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer - GSTHindIII

40

<220>

<221> misc\_feature  
<222> (1)..(37)  
<223> Primer

<400> 2  
gcgaggaagc ttgtcaatca gtcacgatga attcccg  
37

<210> 3  
<211> 49  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer - STforward

<220>  
<221> misc\_feature  
<222> (1)..(49)  
<223> Primer

<400> 3  
atgctgacgt catgaggccc atggggcccg gataacaatt tcacacagg  
49

<210> 4  
<211> 34  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Primer - STreverse

<220>  
<221> misc\_feature  
<222> (1)..(34)  
<223> Primer

<400> 4  
gcggatcctt gcggccgcca ggcaaattct gttt  
34

<210> 5  
<211> 47  
<212> DNA  
<213> Artificial Sequence

10

20

30

40

<220>  
<223> Primer - Forward-A

<220>  
<221> misc\_feature  
<222> (1)..(47)  
<223> Primer

<400> 5  
aaaaaaaaaa aagatcgatc tcatgacgga taacaatttc acacagg  
47

10

<210> 6  
<211> 31  
<212> DNA  
<213> Homo sapiens

<220>  
<221> misc\_feature  
<222> (12)..(21)  
<223> NF-Kappa B P50 Binding site

<400> 6  
cgtatgttgt gggaaattcc cagcggataa c  
31

20

<210> 7  
<211> 51  
<212> DNA  
<213> Artificial Sequence

<220>  
<223> Polyasparagine hexahistidine / Green Fluorescent Protein T  
ag vect  
or cloning site

30

<220>  
<221> CDS  
<222> (1)..(36)  
<223> Polyasparagine Hexahistidine Tag

<220>  
<221> CDS  
<222> (40)..(51)  
<223> Green Fluorescent Protein N-terminal sequence

40

<220>  
<221> misc\_feature  
<222> (37)..(39)  
<223> Amber stop codon

<400> 7  
gtt aac aac aac aac cac cac cac cac tag ggc tct atg  
48

Val Asn Asn Asn Asn His His His His His Gly Ser Met

1 5 10 15

10

agt 51  
Ser

<210> 8  
<211> 12  
<212> PRT  
<213> Artificial Sequence

20

<220>  
<223> Polyasparagine hexahistidine / Green Fluorescent Protein T  
ag vect  
or cloning site

<220>  
<221> misc\_feature  
<222> (37)..(39)  
<223> Amber stop codon

<400> 8

Val Asn Asn Asn Asn His His His His His  
1 5 10

30

<210> 9  
<211> 4  
<212> PRT  
<213> Artificial Sequence

<220>  
<223> Polyasparagine hexahistidine / Green Fluorescent Protein T  
ag vect

40

or cloning site

```
<220>
<221> misc_feature
<222> (37)..(39)
<223> Amber stop codon
```

<400> 9

Gly Ser Met Ser  
1

10

## 【図面の簡単な説明】

【図 1 a】

図 1 a は、ベクタ - pMM106H の構造を示す。

【 図 1 b 】

図 1 b は、タグ挿入前の、ある遺伝子の例 (GST) での PCR 増幅及びエキソヌクレアーゼ消化の詳細を示す。

【 义 1 c 】

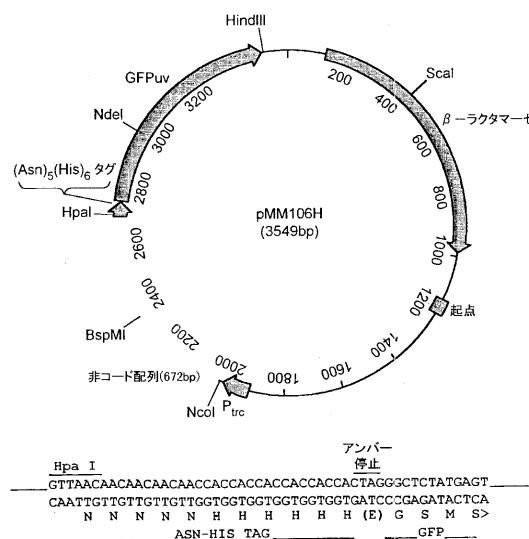
図 1 c は、タグを導入するための特異的なライゲーション及び P C R 増幅の詳細を示す。

【 1 d 】

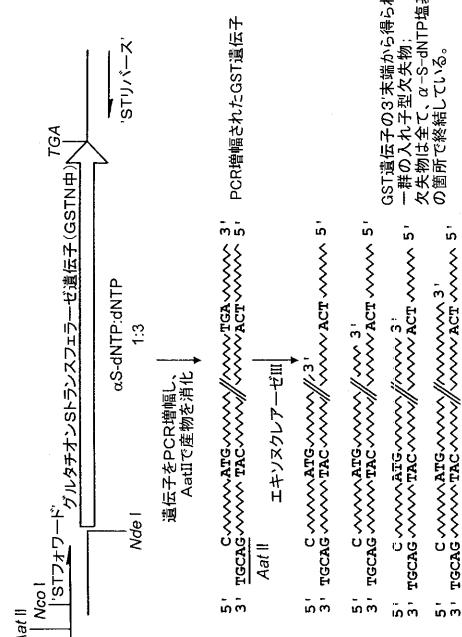
図 1 d は、GSTにより触媒されるグルタチオンと1-クロロ-2,4-ジニトロベンゼンとの反応を示す。

20

【 义 1 a 】



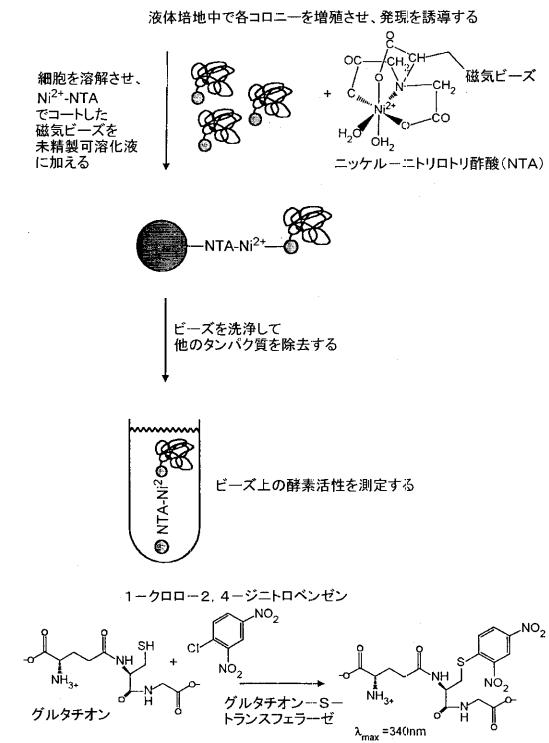
【 1 b 】



【図1c】



【図1d】



## 【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
4 April 2002 (04.04.2002)

PCT

(10) International Publication Number  
WO 02/27327 A2

(51) International Patent Classification: G01N 33/68 (74) Agents: CHAPMAN, P., W. et al.; Kilburn &amp; Strode, 20 Red Lion Street, London WC1R 4PJ (GB).

(21) International Application Number: PCT/GB01/03693

(22) International Filing Date: 17 August 2001 (17.08.2001)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

0020357.0 17 August 2000 (17.08.2000) GB  
60/247,995 14 November 2000 (14.11.2000) US

(71) Applicant (for all designated States except US): SENSE PROTEOMIC LIMITED [GB/GB]; The Babraham Bioincubator, Babraham Hall, Babraham, Cambridgeshire CB2 4AT (GB).

(72) Inventors: and

(75) Inventors/Applicants (for US only): BLACKBURN, Jonathan, Michael [GB/GB]; 36 Woodlark Road, Cambridge CB3 0HS (GB). MULDER, Michelle, Anne [ZA/GB]; 83 Beche Road, Cambridge CB5 8HU (GB). SAMADDAR, Mitoli [IN/GB]; 40 Chatsworth Avenue, Cambridge CB4 3LT (GB). KOZLOWSKI, Rolaund [GB/GB]; Sense Proteomic Limited, The Babraham Bioincubator, Babraham Hall, Babraham, Cambridge CB2 4AT (GB).

(81) Designated States (national): AE, AG, AL, AM, AT, AU,

AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EB, ES, FI, GB, GD, GE, GH,  
GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI,  
SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU,  
ZA, ZW.(84) Designated States (regional): ARIPO patent (GH, GM,  
KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD,  
TG).

Published:

— without international search report and to be republished  
upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



A2

WO

02/27327

Title: METHOD

(54) Title: METHOD

(57) Abstract: The present invention relates to novel methods of producing proteins in which one or more domains are full length and correctly folded and which are each tagged at either the N- or C-terminus with one or more marker moieties and arrays containing such proteins, as well as the use of such proteins in arrays for rapid screening.

**METHOD**

The present invention relates to novel methods of producing proteins in which one or more domains are full length and correctly folded and which are each tagged at either the N- or C-terminus with one or more marker moieties and arrays containing such 5 proteins, as well as the use of such arrays in rapid screening.

The genome mapping projects are revolutionising the therapeutic target discovery process and with it the drug discovery process. As new therapeutic targets are identified, high throughput screening of existing and combinatorial chemical libraries 10 will suggest many potential lead compounds which are active against these targets. It will clearly be uneconomic to pursue all lead compounds through even early phase clinical trials; currently however no rapid method exists for evaluating such lead compounds in terms of their likely activity profiles against all proteins in an organism. If available, such a method would allow the potential toxicology profiles of all the lead 15 compounds to be assessed at an early stage and this information would significantly enhance the process of deciding which lead compounds to pursue and which to set aside.

There is a complementary need in the pharmaceutical industry to identify all the 20 targets of existing drugs (either in the market already or still in development) and hence to define their mechanism of action. The availability of such information will greatly facilitate the process of gaining regulatory approval for new drugs since it is increasingly clear that the regulatory bodies now regard a knowledge of the mechanism of action to be of paramount importance. In addition, this type of 25 information would enable the design of improved second generation drugs. This follows because the majority of drugs have at least minor side effects, which probably result from binding of the drug or a metabolite thereof to undesirable targets; all of these target proteins need to be identified in order to define the criteria necessary for design of improved drugs. Currently however no simple method exists to generate this 30 information and a number of potential multi-million dollar drugs fall by the wayside

simply for lack of knowledge of the target of action.

Protein-protein interactions are being increasingly recognised as being of critical importance in governing cellular responses to both internal and external stresses.

- 5 Specific protein-protein interactions therefore represent potential targets for drug-mediated intervention in infections and other disease states. Currently the yeast two-hybrid assay is the only reliable method for assessing protein-protein interactions but *in vivo* assays of this type will not be readily compatible even in a non-high throughput format with the identification of specific agonists or antagonists of protein-protein interactions. Functional proteome expression arrays, or "proteome chips", will enable the specificity of protein-protein interactions and the specificity of any drug-mediated effect to be determined in an *in vitro* format. They will therefore have enormous potential because they will simply revolutionise this area of research.
- 10
- 15 One way in which functional proteome arrays could be generated is to individually clone, express, purify and immobilise all proteins expressed in the specific proteome. Here though, an important initial consideration concerns the absolute size of the genome of interest together with considerations about the availability of sequence data for the entire genome.
- 20
- By way of illustration of these points, a typical bacterial genome is ~5Mbp and a small number have now been completely sequenced (for example *Helicobacter pylori*, *Escherichia coli*, and *Mycobacterium tuberculosis*); fungal genomes are typically ~40Mbp, mammalian genomes at ~3Gbp and plant genomes at ~10Gbp. Current estimates are that the human genome sequence will be finished around 2003, although how much of this information will be in the public domain is very much open to question. Clearly it will be completely impractical to expect that the genomes of anything other than representative model organisms will become available in a realistic time frame, yet from the perspective of functional proteomics, model organisms are of only limited value. So, whilst in principle within the next four years
- 25
- 30

it may be possible to design and synthesise primers to clone each of the ~100,000 genes in the human genome from cDNA libraries, in practice this will be both enormously expensive (the cost of primers alone would run in to several millions of dollars) and a hugely laborious process, even if the necessary sequence data is

5 available.

But what about those pharmaceutically relevant organisms for which the complete sequence data will not be available? These cannot be simply ignored by functional proteomics so what are the alternatives? Expression cDNA libraries could in principle

10 be used together with non-specific immobilisation to create an array of proteins, but this technology is significantly limited by the fact that non-specific immobilisation is usually associated with loss of function because the fold of the protein is disrupted. In addition, all host cell proteins will also be immobilised which will at best markedly reduce signal-to-noise ratios and at worst result in obfuscation of positive results. The  
15 ability to create a functional proteome array in which individual proteins are specifically immobilised and purified *via* a common motif or tag without affecting function and without requiring knowledge of the entire genome sequence would therefore represent a huge advance in the field of functional proteomics.

20 The Inventors have now developed a novel approach which solves the problems described above by providing methodology which allows each protein in a proteome to be tagged with a common marker at a defined position within the protein without requiring any prior knowledge of the DNA sequence of the corresponding genes. This 'tag' can then be used to impart a commonality and specificity to downstream  
25 immobilisation and purification procedures, which in turn enables the creation of spatially defined arrays in which many thousands of proteins from a given proteome are displayed.

An important consideration here relates to the precise positioning of the 'tag'. If the  
30 tag is inserted in-frame in to any gene at an undefined, random position, the likelihood

is that the resultant tagged protein will be truncated in an undefined manner and in the majority of cases correct folding, and hence function, will be destroyed. It is often found that full-length proteins have short polypeptide extensions at either (or both) the N- and C-termini which can be truncated without affecting folding or function.

5 However, if the truncations remove any N- or C-terminal extensions and cross a domain boundary, folding and function of the protein are usually then compromised. The methodology described here allows the tag to be inserted in the correct reading frame either precisely at the N- or C-terminus of each protein, or within a region close to either terminus which is unimportant in the folding and function of the protein, such

10 that the individual tagged proteins fold correctly and hence retain function when specifically immobilised in the array. In the case of multidomain proteins where individual domains have discrete functions, the methodology described here also allows insertion of the tag within the overall coding sequence but outside specific domain boundaries such that the individual tagged domains fold correctly and hence

15 retain function when specifically immobilised in the array.

Since each protein in the array will be fully functional, the arrays can then be screened directly to identify the targets of drugs and other biologically relevant molecules. The spatial definition of the arrays will allow the phenotype of each protein to be related

20 directly to its genotype to allow the identification of 'hits'.

Thus, in a first aspect, the present invention provides a method producing one or more proteins in which one or more domains are full length and correctly folded and which are each tagged at either the N- or C-terminus with one or more marker moieties, said method comprising:

(a) providing one or more DNA molecules having an open reading frame encoding said proteins together with 5' and/or 3' untranslated regions;

(b) amplifying said DNA molecules under conditions that statistically incorporate  $\alpha$ -S-dNTPs as well as dNTPs into the daughter DNA molecules;

- (c) specifically protecting the 5' or 3' end of said DNA molecules from nuclease digestion;
- (d) treating said DNA molecules first with a 5' to 3'- or 3' to 5'-nuclease to generate a set of nested deletions followed by treating with a single-strand nuclease under conditions that allow removal of said 5' or 3' untranslated regions including the start or stop codons of said open reading frame;
- (e) cloning the fragments generated by step (d) into an expression vector containing a coding sequence for one or more 5' or 3' marker moieties;
- (f) expressing said encoded proteins.

10

Preferably the amplification of the DNA molecule or molecules statistically incorporates a single  $\alpha$ -S-dNTP, more preferably either  $\alpha$ -S-dTTP or  $\alpha$ -S-dATP.

15

The marker moiety can be either a peptide sequence, eg a hexa-histidine tag, an antibody epitope or a biotin mimic, or indeed a complete protein, or protein domain, eg the maltose binding protein domain. The marker moiety itself can be post-translationally modified, eg by addition of a biotin or lipid molecule. In a preferred embodiment, the marker moiety would also allow purification of "tagged" proteins.

20

Thus, the methods of the present invention allow the specific modification, in one pot, of every member of a cDNA library in a manner which does not rely on any knowledge of the sequence of individual genes. Instead, it relies on non-processive truncation of each cDNA by a nuclease such that either the 5'- or the 3'- untranslated region of each cDNA is removed. Additional known DNA sequence encoding a known marker moiety is then appended to the resultant set of nested deletions of each cDNA. If the marker moiety is in the same reading frame as the individual cDNA and is not preceded by any in-frame stop codon, each resultant genetically modified cDNA produced according to the methods of the present invention will thus encode an individual protein which now has a common moiety, eg. a polypeptide "tag" fused to

25

either its N- or C-terminus. A screen for correctly folded, tagged proteins then allows all truncations which cross a domain boundary and affect the folding (and hence function) of the individual protein and all out-of-frame fusions to the tag to be discarded.

5

Since every member of a cDNA library will be modified in the same manner, the net result will be that every protein encoded by the cDNA library will now be tagged with a common moiety at either their N- or C-termini.

10 In general, the proteins expressed from the cDNA library will be "tagged" and can be readily identified and isolated. Once purified they can be attached to microarrays, for example. Attachment can be effected by means of the tag itself, or alternatively, by means of another moiety which is first attached to the proteins.

15 Arrays formed by the methods described herein form a second aspect of the invention. Such arrays comprise the "tagged" protein expression library, immobilised, usually on a solid support. The skilled person will understand that a range of possible solid supports are in common usage in the area of arrays and any of these "substrates" can be utilised in the production of arrays of the present invention.

20

As discussed herein the term "protein array" relates to a spatially defined arrangement of one or more protein moieties in a pattern on a surface. Preferably the protein moieties will be attached to the surface either directly or indirectly. The attachment can be non-specific (e.g. by physical absorption onto the surface or by formation of a 25 non-specific covalent interaction). In a preferred embodiment the protein moieties will be attached to the surface through the common marker moiety linked to each protein using the methods described herein.

25

In another embodiment, the protein moieties may be incorporated into a vesicle or liposome which is tethered to the surface.

30

Thus, for example, each position in the pattern may contain one or more copies of:

- a) a sample of a single protein type (in the form of a monomer, dimer, trimer, tetramer or higher multimer);  
5
  - b) a sample of a single protein type bound to an interacting molecule (e.g. DNA, antibody, other protein);
  - c) a sample of a single protein type bound to a synthetic molecule (e.g. peptide, chemical compound); or  
10
  - d) mixtures of between 2 and 100 different tagged protein moieties at each position in the pattern of the array.
- 15 The surface which supports the array may be coated/derivatised by chemical treatment, for instance. Examples of suitable surfaces include glass slides, polypropylene or polystyrene, silica, gold or metal support or membranes made of, for example, nitrocellulose, PVDF, nylon or phosphocellulose.
- 20 As discussed herein, the methods of the present invention allow tagging of all proteins in a given proteome specifically at either the N- or C-terminus. Whilst some proteins may not tolerate N-terminal extensions and others might not tolerate C-terminal extensions, it is likely that the vast majority of proteins will tolerate one or other such extensions. Existing library cloning methods, however, simply cannot address this  
25 problem since they clone genes either as full-length, unmodified cDNAs or as random and almost inevitably truncated fusions to some protein partner. Compared to the latter, the present methods allow the position of the tag to be targeted to the sequences at or close to the N- or C-terminal residues of the cDNA products such that fusion to e.g. a desired peptide partner does not affect folding or function of the cDNA product  
30 Compared to the former, the method of immobilising proteins in an array as described

- herein is through specific rather than non-specific interactions, and these specific interactions are a function of the tag added to the termini of each cDNA. Additionally, the methods described herein can be used to screen purified, immobilised proteins which have been expressed in non-bacterial host organisms to aid maintenance of
- 5 function through correct folding and post-translational modification, whereas existing methods such as phage display or  $\lambda$ -cDNA expression libraries are restricted to bacterial hosts in which the majority of eukaryotic proteins are found to be synthesised in a non-functional form, either due to mis-folding or incorrect post-translational modification.
- 10 The methods of the present invention have a wide range of potential *in vitro* applications, which can be broadly divided into three main areas. These are the study of protein-ligand interactions, the study of protein-protein interactions, and the study of protein-DNA interactions.
- 15 **Protein-Ligand Interactions**
- The methods described herein will allow the rapid profiling of the interactions between a given new chemical entity and all proteins in a given proteome. This can be achieved simply through probing the appropriate proteome array with the NCE at
- 20 varying stringencies in what might be considered a reverse high throughput screen. The readout from such a screen will be directly useful in many situations, some of which are described below.
- High throughput screening programs in which libraries of compounds are tested
- 25 against cells or whole organisms often identifies leads, which give rise to a phenotypic change without the target being known prior to screening. Subsequent identification of the primary target can, however, be a very laborious process. The methods of the present invention can be applied directly to this type of problem since it will be possible to create a functional proteome array for the species concerned and then
- 30 screen this array with the lead compound to identify which proteins within the

- proteome it is targeting. This massively parallel approach to identifying protein-ligand interactions will greatly speed up and simplify the determination of primary targets of NCEs, and will also allow identification of weaker secondary interactions which may also be important. In addition, the methods can be applied directly to the question of species cross-reactivity, allowing a potential antifungal compound, for example, to be quickly assessed in terms of its interactions with, for example, all proteins in a human proteome; this type of information is likely to prove very useful in any subsequent optimisation of lead compounds.
- 10 High throughput screening methods now allow the rapid identification of small molecules which bind to a given protein which has itself previously been identified as a potential therapeutic target. However, these methods do not address the question of how selective any given interaction might be yet this knowledge is potentially crucial in deciding whether to pursue a given lead compound or not; perceived wisdom would argue that compounds which target single proteins are likely to show fewer side effects than those which also hit a large number of related or unrelated proteins.

20 There are a number of examples of compounds which have progressed successfully through third phase clinical trials yet have failed to win regulatory approval because their primary mechanism of action is not known. The antidepressant drugs mianserin and trazadone and the Pfizer anti-arthritis drug tenidap are examples here, each representing hundreds of millions of dollars investment for no return. The methods described herein can potentially be applied to the resurrection of such failed drugs since if the primary targets of such drugs can be discovered and subsequently verified 25 in terms of mechanism of action, the vastly expensive clinical trial data is already in place for regulatory approval.

30 All existing drugs have side effects, to a greater or lesser extent, an example here being the otherwise attractive anti-schizophrenia drug clozapine. If the molecular origin of such side effects could be determined, this would greatly facilitate the design of future generation drugs with optimised primary effects combined with minimised

side effects. Again the presently described methods can be applied directly to such problems since in creating a profile of the interactions between a compound and all proteins in a proteome, aberrant secondary interactions will be identified and these can subsequently be assessed in terms of whether they are linked to known side effects.

5

The methods of the present invention can also be used to identify families of proteins, such as serine proteases, through screening proteome arrays with generic inhibitors. This would then allow the subsequent development of biochips displaying, for 10 example, all human serine proteases or, alternately, all kinases or all p450 enzymes for more focused screening of lead compounds. A p450 biochip, for example, would have utility in assessing whether a given lead compound is likely to be metabolised or not, since p450-mediated hydroxylation is often the first step in this process and is thought to be one of the primary sources of patient-to-patient variability in drug response; 15 indeed one of the goals of drug design now is to generate compounds which are not metabolised in the first place and here again a p450 chip would have significant potential utility.

20

**Protein-protein interactions**

Protein-protein interactions and multiprotein complexes are of critical importance in cellular biology. Signalling pathways, for example, are commonly initiated by an interaction between a cell surface receptor and an external ligand, and this is followed by a cascade of protein-protein interactions which ultimately result in the activation of 25 a specific gene. Individual protein-protein interactions might be dependent on the presence of a specific ligand or alternatively might be blocked by a specific ligand, whilst some multiprotein complexes will only form in a ligand-dependent manner.

30

Thousands of new protein-protein interactions have been identified using two-hybrid technologies. The methods described herein overcome the limitations of such methods and can be used to screen proteome arrays with individual labelled proteins to identify

not only interacting partners but also the relative strengths of individual interactions. The methods can also be applied to the identification of the components of multiprotein complexes, even where their assembly is ligand dependent.

- 5 An example of the use of the methods in this way in defining novel protein-protein interactions would be the identification of the signalling partners of the cytosolic domain of a particular cell surface receptor which has been implicated in a disease state; identification of such signalling partners would be directly relevant from a pharmaceutical perspective since such protein-protein interactions might immediately  
10 represent possible therapeutic targets.

#### Protein-DNA Interactions

- It has been estimated that roughly 10% of all genes in the human genome encode transcription factors yet only a small percentage of these are at present identified. The  
15 binding of specific transcription factors to DNA enhancer elements, often in response to external stimuli, is a prerequisite for the formation of enhanceosome complexes which then switch on gene expression. There are various points at which gene expression can in principle be affected by drug administration: a drug might block the binding of a protein or small molecule to a cell surface receptor and hence block the  
20 signalling cascade at the beginning; a drug might block a protein-protein interaction or inhibit an enzymatic activity within the signalling cascade; or alternatively, a drug might block formation of specific protein-DNA or protein-protein interactions within the enhanceosome complex. As an example here, the transcription factor NF- $\kappa$ B is involved in cellular processes as diverse as immune and inflammation responses, limb  
25 development, septic shock, asthma, and HIV propeptide production. The majority of the intracellular signalling cascades in NF- $\kappa$ B activation are common to all these processes so do not represent viable targets for intervention. The differences between the responses therefore lie in either the original ligand-receptor interaction or in the  
30 formation of specific enhanceosome complexes. NF- $\kappa$ B is known to bind to at least 14 different enhancer elements and the enhanceosome complexes therefore represent

potential therapeutic targets. However, delineation of an individual enhanceosome complex requires knowledge of both the number of individual DNA-binding proteins involved and also their protein-protein interactions with each other. The present methods can be used to directly address both these questions. A proteome array can be screened with specific DNA probes to identify novel DNA binding proteins, Alternatively, the proteome array can be screened with the transactivation domain of a given transcription factor to identify other proteins with which it interacts. Cross correlation of such screens should allow identification of new components of specific enhanceosome complexes

10 The protein arrays generated by the methods of the present invention will also allow the selection of molecules which recognise each protein displayed in the arrays. In a preferred embodiment, the selected molecules will be antibodies or antibody-like proteins and will be displayed on phage or on ribosomes or will be covalently linked to the encoding mRNA.

20 Thus, a phage displayed antibody library can be applied to each immobilised protein in the array and non-binding antibodies removed by washing. The selected phage can then be recovered and used to infect bacteria according to normal procedures. The phage-infected bacteria can then produce either phage particles displaying the selected antibodies for further rounds of selection, or they can produce soluble antibody fragments for direct use. The terms 'antibody' or 'antibody fragments' here refer to single chain Fvs, FAB fragments, individual light or heavy chain fragments, derived from mouse, human, camel or other organisms.

25 In a preferred embodiment, the protein array will be in microwell format such that after the selection step, the phage particles can be recovered by addition of appropriate bacterial cells to each well where they will become infected by the selected phage particles. Growth media can then be added to each well and the infected bacteria allowed to grow and express the antibody fragments, whilst maintaining the physical

separation of the antibody fragments selected to each immobilised protein in the array.

If so desired, new phage particles produced by the infected bacteria can be used in subsequent rounds of selection. Such procedures are now routine for selecting polyclonal or monoclonal antibody fragments to a single purified and immobilised

5 protein. In effect then the original protein arrays here will allow the generation of polyclonal or monoclonal antibody fragments to thousands of correctly folded proteins in a massively parallel manner whilst otherwise using standard in vitro antibody selection methods.

- 10 The selected, solubly expressed antibody fragments from each well of the original array can themselves be immobilised in to a new spatially defined array such that the antibody fragments in each position of the new array were selected against the proteins immobilised in a single, defined position in the original array. The antibody arrays so-generated will contain at each position either polyclonal or monoclonal antibody
- 15 fragments, depending on the number of rounds of selection carried out prior to immobilisation of the soluble antibody fragments.

Such antibody arrays will have a number of potential uses including capture of individual proteins from a crude cell or tissue lysate for differential expression  
20 monitoring of the relevant proteome. Alternatively, the antibody-captured proteins might be screened directly for ligand-binding function. In general, any one monoclonal antibody might bind to the target protein so as to block its function, but another monoclonal antibody might bind but not block function. In a massively parallel approach, it is clearly impractical to assess all monoclonal antibodies to all  
25 proteins in a proteome individually for their ability to bind but not affect function. A polyclonal set of antibodies to all proteins in a proteome however is likely to contain individual antibodies which have the desired ability to bind but not affect function and will, in addition, contain individual antibodies which recognise all post-translational modifications of a given protein. Thus in general, polyclonal rather than monoclonal

antibody arrays generated as described will likely be advantageous for screening captured proteins directly for function.

Compared to the original protein arrays, the antibody arrays created by the methods described here will have the advantage that all proteins immobilised on the array will be stable under similar conditions. The proteins captured from the crude cell or tissue lysate will not be recombinant but will have been naturally expressed. Moreover, the captured proteins can be screened for function or ligand binding *etc* directly after capture from the crude cell or tissue lysate, which should aid maintenance of function.

10

Thus, in further aspects, the present invention provides:

- (i) a method of screening one or more compounds for biological activity which comprises the step of bringing said one or more compounds into contact with a protein array as defined herein and measuring binding of the one or more compounds to the proteins in the array;
- (ii) a method of screening one or more proteins for specific protein-protein interactions which comprises the step of bringing said one or more proteins, eg a cell surface receptor, into contact with an array as defined herein, and measuring binding of the one or more specific proteins with the proteins of the array;
- (iii) a method of screening one or more proteins for specific protein-nucleic acid interactions which comprises the step of bringing said one or more nucleic acid probes into contact with an array as defined herein and measuring binding of the probes to the proteins in the array;
- (iv) the use of an array as defined herein in the rapid screening of a compound, protein or nucleic acid;

25

20

25

30

- (v) the use of an array as defined herein in screening for molecules which recognise each protein in the array, wherein the molecules are preferably antibodies;
- 5 (vi) a method of generating an antibody array which comprises bringing a protein array, as defined herein, into contact with an antibody library, such that one or more proteins in the protein array bind to at least one antibody in the antibody library, removing any unbound antibodies and immobilisation of those antibodies bound to proteins in the protein array; and
- 10 (vii) a method for the screening of protein function or abundance which comprises the step of bringing an antibody array as defined herein into contact with a mixture of one or more proteins.

The methods (i), (ii), (iii) and (vi) may also include the step of first providing the array according to one or more of the methods of the present invention.

20 Use of the proteins derived from the methods described herein form additional aspects of the invention. The skilled person will understand that a range of applications are known in the art in which modified proteins may be employed.

Thus, in further aspects, the present invention provides:

25 (i) the expression of tagged proteins produced by the methods of the invention in numerous expression hosts i.e. bacteria, yeast, mammalian cells (for example see, Walker EA, Clark AM, Hewison M, Ride JP, Stewart PM. Functional expression, characterization, and purification of the catalytic domain of human 11-beta -hydroxysteroid dehydrogenase type 1. J Biol Chem

30

2001 Jun 15;276(24):21343-50; Cai J, Daoud R, Georges E, Gros P. Functional Expression of Multidrug Resistance Protein 1 in *Pichia pastoris*. *Biochemistry* 2001 Jul 17;40(28):8307-16, and Hara H, Yoshimura H, Uchida S, Toyoda Y, Aoki M, Sakai Y, Morimoto S, Shiokawa K. Molecular cloning and functional expression analysis of a cDNA for human hepassocin, a liver-specific protein with hepatocyte mitogenic activity(I). *Biochim Biophys Acta* 2001 Jul 30;1520(1):45-53)

(ii) the use of a tagged protein produced by the methods as defined herein.

10

(iii) the use of a tagged protein produced by the methods as defined herein for analysis of interaction between expressed protein and other proteins within a yeast two-hybrid system via the cloning of said modified DNA molecule into a yeast two-hybrid expression vector (for example see, Staudinger J, Perry M, Elledge SJ, Olson EN. Interactions among vertebrate helix-loop-helix proteins in yeast using the two-hybrid system. *J Biol Chem* 1993 Mar 5;268(7):4608-11, and Vojtek AB, Hollenberg SM, Cooper JA. Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* 1993 Jul 16;74(1):205-14).

20

(iv) the use of a tagged protein produced by the methods as defined herein for immobilization on an affinity column/substrate, for example to allow the purification by affinity chromatography of, for example, a) interacting proteins, b) DNA or c) chemical compounds. (for example see, Rhodes N, Gilmer TM, Lansing TJ. Expression and purification of active recombinant atm protein from transiently transfected mammalian cells. *Protein Expr Purif* 2001 Aug;22(3):462-6; Zwicker N, Adelhelm K, Thiericke R, Grabley S, Hanel F. Strep-tag II for one-step affinity purification of active bHLHzip domain of human c-Myc. *Biotechniques* 1999 Aug;27(2):368-75, and Giuliani CD, Iemma MR, Bondioli AC, Souza DH, Ferreira LL, Amaral AC, Salvini TF, Selistre-de-Araujo HS. Expression of an active recombinant lysine 49

25

30

phospholipase A(2) myotoxin as a fusion protein in bacteria. Toxicon 2001 Oct;39(10):1595-600)

- (v) the use of a tagged protein produced by  
5 the methods as defined herein in the immobilization by affinity purification for interrogation by antibodies (ELISA assay) as a diagnostic tool (for example see, Doellgast GJ, Triscott MX, Beard GA, Bottoms JD, Cheng T, Roh BH, Roman MG, Hall PA, Brown JE. Sensitive enzyme-linked immunosorbent assay for detection of Clostridium botulinum neurotoxins A, B, and E using signal amplification via enzyme-linked coagulation assay. J Clin Microbiol 1993 Sep;31(9):2402-9)
- (vi) the use of a tagged protein produced by the methods as defined herein  
15 as a probe for a cDNA microarray to identify DNA binding proteins (for example, see DeRisi J, Penland L, Brown PO, Bittner ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM. Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nat Genet 1996 Dec;14(4):457-60).
- (vii) the use of a tagged protein produced by the methods as defined herein  
20 for elucidating the identity of proteins in the 'proteome' by mass spectrometric analysis of expressed protein components of source library or start material modified by the methods of the invention (for example, see Bordini E, Hamdan M. Investigation of some covalent and noncovalent complexes by matrix-assisted laser desorption/ionization time-of-flight and electrospray mass spectrometry. Rapid Commun Mass Spectrom 1999;13(12):1143-51).

Preferred features of each aspect of the invention are applicable to each other aspect,  
*mutatis mutandis.*

The present invention will now be described with reference to the following examples, which should not in any way be construed as limiting the scope of the invention.

- 5           **FIGURE 1a:** shows the construction of the vector pMM106H;
- 10          **FIGURE 1b:** shows details of the PCR amplification and exonuclease digestion of an example gene (GST) prior to tagging;
- 15          **FIGURE 1c:** shows details of the specific ligation and PCR amplification to introduce the tag;
- FIGURE 1d:** shows the reaction between Glutathione and 1-chloro-2,4-dinitrobenzene catalysed by GST.

15          **Example 1**

*(a) Vector construction (see figure 1a)*

The Inventors constructed a vector pMM106H derived from pUC19 which contains a strong hybrid promoter (*P<sub>trc</sub>*) to drive the expression of genes cloned into an *Nco* I site immediately downstream of the promoter sequence. The Inventors inserted a 676 bp nonsense DNA sequence as a stuffer fragment between the *Nco* I site and a downstream *Hpa* I site. *Hpa* I is a blunt-end cutter and is positioned to cleave the vector such that the downstream DNA encodes a polyasparagine, hexahistidine peptide if the reading frame is on the first base of the blunt-end. Following the hexahistidine tag is an amber stop codon (TAG) followed by the gene encoding the green fluorescent protein (GFP) of the jellyfish *Aequorea victoria*. The construction of pMM106H was confirmed by sequencing.

Genes cloned into pMM106H as *Nco* I/blunt-end fragments result in fusions to the His-tag and GFP only if the correct reading frame is created at the *Hpa* I site during cloning. GFP is used here as a reporter gene to facilitate visual screening of clones expressing the His tag, while also providing an indication of the correct folding of the fusion protein, since GFP is only active when folded into the correct conformation.

- The amber stop codon will result in a small amount of the full length fusion protein for visualisation of green colonies, while most of the fusion protein will terminate immediately after the His tag and can be used for subsequent immobilisation and enzyme assays. It should be understood that a number of different peptides of 5 proteins, whose soluble expression confers some observable phenotype on the cells, could be used in place of GFP as markers for expression and folding of the tagged proteins. These include, but are not limited to, chloramphenicol acetyl transferase,  $\beta$ -galactosidase, the lacZ fragment of  $\beta$ -galactosidase, and proteins capable of repressing transcription, such as the  $\lambda$ -CI repressor.
- 10 The template used in the procedure outlined below was pGSTN. This plasmid was constructed by first PCR-amplifying the *Schistosoma japonicum* glutathione S transferase (GST) gene from pGEX-2T (Pharmacia) under standard conditions using primers 'GSTfwd2' (5'-ATG CTG CAG ACG TCA ACA GTA TCC ATG GCC CCT ATA CTA CG-3') and 'GSTHindIII' (5'-GCG AGG AAG CTT GTC AAT CAG TCA CGA TGA ATT CCC G-3'). These primers introduce an *Nco* I restriction site at the start codon of GST, mutate the second residue of GST from serine to alanine, and introduce a stop codon in the multiple cloning site 3'- of the GST gene followed by a *Hin* dIII restriction site. The PCR product was then cloned under standard conditions as an 15 *Nco* I/*Hin* dIII fragment into pTrcHisA (Invitrogen) previously digested with *Nco* I/*Hin* dIII to generate pGSTN.
- 20

*(b) PCR amplification and exonuclease digestion of genes prior to tagging (see figure 1b)*

- 25 The Inventors amplified the GST gene from the construct pGSTN using the polymerase chain reaction with custom-designed vector-specific primers 'STforward' (5'-ATG CTG ACG TCA TGA GGC CCA TGG GGC CCG GAT AAC AAT TTC ACA CAG G-3') and 'STreverse' (5'-GCG GAT CCT TGC GGC CGC CAG GCA AAT TCT GTT T-3') which bind to the vector 156 bp upstream of the start and 84 bp downstream of the stop codons respectively. 30 cycles of PCR (94°C 1min; 57°C 1min, 72°C 2min) were

carried out in four separate 100 $\mu$ l reactions. Each PCR reaction contained ~20ng template DNA, 50pmol each primer and 2.5 units *Pwo* polymerase. Each PCR reaction was carried out in a standard buffer (10mM Tris.HCl pH8.8, 25mM KCl, 5mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 2mM MgSO<sub>4</sub>, 10% DMSO). Each of the four PCR reactions then also contained a non-standard deoxynucleotide triphosphate mix, as follows:

Reaction 1) 200 $\mu$ M dATP, 200 $\mu$ M dTTP, 200 $\mu$ M dCTP, 150 $\mu$ M dGTP, 50 $\mu$ M  $\alpha$ -S-dGTP;

Reaction 2) 200 $\mu$ M dATP, 200 $\mu$ M dTTP, 200 $\mu$ M dGTP, 150 $\mu$ M dCTP, 50 $\mu$ M  $\alpha$ -S-dCTP;

Reaction 3) 200 $\mu$ M dATP, 200 $\mu$ M dGTP, 200 $\mu$ M dCTP, 150 $\mu$ M dTTP, 50 $\mu$ M  $\alpha$ -S-dTTP;

Reaction 4) 200 $\mu$ M dGTP, 200 $\mu$ M dTTP, 200 $\mu$ M dCTP, 150 $\mu$ M dATP, 50 $\mu$ M  $\alpha$ -S-dATP.

The amplification of the template DNA in the presence of  $\alpha$ -S-dNTPs can of course be carried out by primer extension reactions using many different DNA polymerase, including thermostable polymerases which lack a 3' to 5' exonuclease activity, such as Tap polymerase, and non-thermostable polymerases, such as T4 DNA polymerase or the Klenow fragment of DNA polymerase I.

The inclusion of a single  $\alpha$ -thio deoxynucleotide triphosphate in each specific PCR mix results in a random but statistical incorporation of the relevant  $\alpha$ -S-dNTP into the specific final PCR product. These modified nucleotides are not substrates for Exonuclease III, and are used to halt the progressive removal of nucleotides by the enzyme. The four individual PCR mixes were then pooled, and purified using a QIAquick PCR cleanup kit (Qiagen), under standard conditions, and digested to completion with the restriction enzyme *Aat* II. The resulting ~1000bp PCR products were then gel-purified. Restriction with *Aat* II results in a 3'-overhang which is

resistant to Exonuclease III activity and, therefore, protects the 5' end of the PCR product from degradation.

Alternative methods for specifically protecting one end of the PCR product from exonuclease digestion can be readily envisaged including, but not restricted to the following. Any restriction enzyme which generates a 3'-overhang of 4 bases or more could be used in place of *Aat* II providing the requisite site is incorporated into the design of the PCR primers. Any restriction enzyme which generates a 5'-overhang could also be used in place of *Aat* II providing the requisite site is incorporated into the design of the PCR primers; in this case, generation of the 5'-overhang would be followed by a DNA-polymerase-mediated fill-in reaction in which the relevant  $\alpha$ -thio dNTPs were used in place of the dNTPs such that the new 3'-end of the PCR product is now protected from exonuclease digestion.

15 10-15 $\mu$ g of the digested PCR product was then incubated with 75 units of Exonuclease III/ $\mu$ g of DNA for 30 minutes at 37°C in a 150 $\mu$ l reaction. The Exo III digestion was carried out in a standard reaction buffer (66mM Tris.HCl pH8.0, 6.6mM MgCl<sub>2</sub>, 5mM DTT, 50 $\mu$ g/ml bovine serum albumin). These conditions ensure that digestion by Exo III has reached completion. The enzyme was then inactivated by heating to 75°C for 20 15 minutes. The product of the Exo III digestion is a nested set of deletions from the 3'-end of the PCR product.

Exonuclease III is a non-processive 3'- to 5'- exonuclease which is unable to hydrolyse  $\alpha$ -thio-containing nucleotides so, in the present protocol, every time Exo III reaches an  $\alpha$ -thio-deoxynucleotide base, the progressive truncation of the recessed 3'-end of the PCR product is halted. The net result is thus a nested set of deletions as a consequence of the random incorporation of each  $\alpha$ -S-dNTP at the earlier stage. The 25 ratio of  $\alpha$ -S-dNTP to dNTP used in the original PCR amplifications was determined

empirically such that the envelope of nested deletions spanned a 400bp window of sizes centred approximately 100bp shorter than the original full length PCR product.

The size range of the truncations obtained can be controlled by altering the ratio of  $\alpha$ -S-dNTP to normal dNTP. This is important when the method is applied to eukaryotic cDNAs because such cDNAs have variable length 3' untranslated regions, with the most common 3'-UTR length falling in the range of 200-300 bp. Since the relative efficiency of incorporation of each of the four  $\alpha$ -S-dNTPs varies according to the identity of the polymerase, it is desirable to use  $\alpha$ -S-dNTP to normal dNTP ratios which are optimised for each of the four bases and for the particular polymerase. Typically, the molar ratio of racemic  $\alpha$ -S-dNTP to normal dNTP used will lie in the range 1:1 to 1:3.

(c) Removal of single-stranded regions and preparation for cloning (see figure

15 1e)

The nested set of deletions generated by Exonuclease III digestion in the previous step was purified by ethanol precipitation and resuspended in 1x mung bean nuclease buffer (50mM sodium acetate pH5.0, 30mM NaCl, 1mM ZnSO<sub>4</sub>). The digested DNA was treated with (2units/ $\mu$ g) 30 Units of mung bean nuclease in a 100 $\mu$ l reaction at 20 30°C for 30 minutes. This process removed the 5' and 3'-overhangs to yield blunt-end products. The reaction was stopped by the addition of EDTA to a final concentration of 5mM. The digested products were purified using a QIAquick PCR purification kit (Qiagen), digested with *Nco* I, and separated on a 1% agarose/TBE gel, using a 100bp DNA ladder as a standard. Products ranging in size from 800 to 1000bp 25 were extracted from the agarose using a QIAquick gel extraction kit (Qiagen).

Clearly, other single-strand nucleases such as S1 nuclease, can also be used to remove the 5'-overhang from the nested set of 3'-deletions generated by a 3'- 5'- exonuclease.

*(d) Variation on the preparation of nested deletions*

A number of other standard molecular biology methods for generating a nested set of deletions represent obvious variations on the original procedure. These include, but are not limited to, the use of any 3'- to 5'- exonuclease, any 5'- to 3'- exonuclease, or any endonuclease which truncates progressively from the termini of a linear DNA fragment. For example, the initial PCR amplification can be performed using the same reverse primer as above but with a forward primer which binds approximately 2 kb upstream of the start of the GST gene. This will generate a fragment in which the GST gene is flanked by >2 kb on the 5' end and only 84 bp on the 3' end. The purified PCR fragment can then be treated with Bal 31 nuclease, which progressively degrades linear duplex DNA from both the 5'- and 3'-ends. The enzyme is non-processive and the rate of degradation of the DNA depends on the time and temperature of the reaction, as well as the base composition of the DNA. Since the flanking region on the 3'-end of the GST gene in the PCR product is significantly shorter than that on the 5' end, degradation up to and beyond the stop codon will occur long before the start codon is reached from the other end. Time course experiments allow the optimum reaction conditions for removal of up to 400bp from the 3' end of the PCR product to be determined. The resulting nested set of deletions can then be blunt-ended to remove any remaining single-stranded regions, digested with a unique restriction enzyme encoded at the 5'-end of the gene by the original vector, and directionally cloned into the tag vector. Alternatively, Lambda exonuclease can be used to generate a nested set of 5'-deletions. The preferred substrate for this enzyme is 5' phosphorylated double stranded DNA so one end of the DNA substrate can be easily protected by having a 5' hydroxyl terminus.

The single-stranded 3' overhangs of the nested set of 5'-deletions generated by a 5'-3- exonuclease can be removed by a number of different enzymes, including T4 DNA polymerase or a single-stranded DNA specific nucleases such as RNase T or Exonuclease T or mung bean nuclease.

30 (e) Cloning and analysis of the modified products (see figure 1c)

The vector pMM106H (3 $\mu$ g) was digested to completion with the restriction enzymes *Nco* I and *Hpa* I and the 2870bp backbone fragment was gel purified. The vector DNA and the restricted products prepared as described above were then ligated together under standard conditions and the ligation mix was used to transform *E. coli*

5 DH5 $\alpha$  cells which were then recovered and plated onto LB plates containing 100 $\mu$ g/ml carbenicillin.

This cloning procedure was carried out on the full set of deletions obtained in the previous step. However, only those deletions which excise the stop codon of the GST gene and end immediately after an in-frame codon should be able to give rise to in-frame fusions to the hexahistidine tag and GFP after cloning steps via this procedure; all other deletion products cloned in this manner should only lead to out-of-frame fusions to the hexahistidine tag and GFP or to unfused GST proteins, due to translational termination at the GST stop codon. This follows because ligation of the blunt end of the deletion product to the blunt end of the vector results in a genetic fusion in which the translation reading frame of the downstream vector DNA is dictated by the original reading frame of the GST coding region. If the deletion product ends in an incomplete codon, the newly appended hexahistidine-coding sequence will be out-of-frame with respect to the GST gene, whilst if the deletion 15 product retains the GST stop codon, no translational fusion of GST to the hexahistidine tag will occur. The only hexahistidine-(and GFP-) tagged proteins which can arise from the overall process described above will therefore necessarily be GST fusions to the polyasparagine, hexahistidine tag. These will not necessarily be absolutely full-length clones, however their ability to fold correctly and retention of 20 enzyme activity will be screened for in further steps.

Transformed colonies were visualised under UV light (365nm) and 30 colonies (approximately 10% of the total) which fluoresced green were selected by eye for further analysis. These colonies were replica-plated and analysed by colony Western blot under standard conditions using anti-His-tag and anti-GST antibodies. The anti-

25

His-tag antibody only binds to colonies which express a hexahistidine-tagged protein so the Western blot gives direct information about the number of colonies expressing in-frame fusions to the hexahistidine-tag. The anti-GST antibody, on the other hand, binds close to the C-terminus of the GST protein and therefore only recognises 5 colonies expressing full- or nearly full-length GST proteins. The Inventors identified 19 colonies (63% of the green fluorescent colonies) containing protein which was positively recognised by both anti-His-tag and anti-GST antibodies. The DNA from 12 of these colonies was amplified, purified and sequenced. The sequencing data confirmed the presence of two perfect in-frame fusions to full length GST and 10 clones with short truncations in the GST gene, which were still in-frame with the hexahistidine tag. The frequency of isolation of full-length GST clones the Inventors obtained via this overall procedure is therefore approximately 17% (of the total 15 number of green fluorescent colonies), while the frequency of isolation of full- or almost full-length GST clones, which are expected to retain activity, is approximately 63% (of the total number of green fluorescent colonies).

*(f) Immobilisation and functional analysis of tagged proteins (see figure 1d)*

*E. coli* DH5 $\alpha$  cells were transformed with one of the full-length, hexahistidine-tagged GST plasmids created via the above methodology. A single carbenicillin-resistant 20 colony was grown to mid-log phase in 10ml liquid culture and then supplemented with 100 $\mu$ M IPTG to induce expression of the hexahistidine-tagged GST. After growth for a further 4 hours, cells were harvested and lysed by freeze-thaw/lysozyme. SDS-PAGE of the crude lysate showed an overexpressed protein at the expected size (27kDa), which represented roughly 20% of total soluble protein, as well as a small 25 amount of the 54 kDa GST-hexahistidine-GFP fusion, generated through amber suppression. The crude lysate (500 $\mu$ l; 100 $\mu$ g) was then mixed with Nickel-NTA magnetic beads (50 $\mu$ l; binding capacity 15 $\mu$ g hexahistidine-tagged protein) and the beads recovered by sedimentation under a magnetic field. The supernatant was discarded and the beads were washed and then resuspended in a glutathione S

transferase assay buffer containing 1mM each of glutathione and 1-chloro-2,4-dinitrobenzene. End point assay data was collected after 30 minutes at room temperature by measuring the absorbance at 340nm; this wavelength corresponds to the  $\lambda_{max}$  of the product of the GST-catalysed reaction.

5

As controls, cultures of DH5 $\alpha$  containing either the parent vector (pMM106H) or a plasmid encoding an unrelated His-tagged protein (alanine racemase) were grown, induced, harvested, lysed and assayed in parallel. GST activity was only detected on the beads which had been mixed with the crude lysate containing the His-tagged GST, clearly demonstrating that the observed GST activity was due specifically to the immobilised His-tagged GST and moreover that the protein retained activity on specific immobilisation.

10

After completion of the enzymatic assay, protein was eluted from the magnetic beads by addition of buffer containing 100mM imidazole and analysed by SDS-PAGE. This showed that the sample which gave the positive activity assay result contained a single immobilised protein of the exact size expected for glutathione S transferase (27kDa), thus confirming that the observed activity on the beads was due to this recombinant His-tagged protein alone.

15

#### Example 2

##### (a) Vector construction

20

The Inventors constructed a second vector, pMM111, which is essentially the same as pMM106H (see Example 1), except that the 676bp *Nco I/Hpa* I nonsense DNA stuffer fragment is replaced with a 300bp *Nco I/Hpa* I fragment derived from the *Escherichia coli gdhA* gene; the *Hpa* I cloning site is replaced with a *Sma* I site, positioned such that the downstream hexahistidine tag is out of frame with the *gdhA* gene by 2

nucleotides; and the ATG start codon of the GFP gene is replaced with the codon for alanine (GCG). The vector has been designed such that an insert cloned into the *Sma* I site must contain the first nucleotide of a codon at its 3' end to put it in frame with the hexahistidine tag and GFP. The construction of pMM111 was confirmed by sequencing.

5 **(b) Modification procedure to introduce tag**

The Inventors then carried out a procedure identical to that described in Example 1 except for the following modifications. Firstly, only  $\alpha$ -S-dTTP was incorporated in 10 the original PCR amplification, i.e. reaction number 3 in section (b). Secondly, the final products were cloned in to the *Neo* I and *Sma* I sites of the vector pMM111.

This procedure has several theoretical advantages over that described in Example 1. These arise principally from the statistics associated with incorporation of a single  $\alpha$ -thio dNTP in to the original PCR products. Thus, upon exhaustive exonuclease III 15 digestion, the nested set of 3'-recessed deletions will all now terminate with a 3'- thymidine base rather than with any of the four nucleotides. Cloning of these fragments into the *Sma* I site of pMM111 will only result in an in-frame fusion to the hexahistidine tag and GFP if the 3'-T is either that of the first in-frame stop codon of 20 the GST gene or precedes, and is in the same reading frame, as the 'T' of the first in-frame stop codon (this follows because digestion with *Sma* I results in a gap of 2 nucleotides before the coding sequence of the tag).

Statistically, 4-fold fewer nested deletions will be created by the exonuclease 25 hydrolysis in this modified procedure than in Example 1. However, since all three possible stop codons contain 'T' as their first base, they will all be represented in the set of deletions and will therefore constitute a four-fold higher fraction of the full set of deletions. Given the probability of any given 'T' being in the same reading frame as the first 'T' of the stop codon, 33% of all clones resulting from this modified procedure should be in-frame fusions to the His tag encoded by the vector but those

deletions which affect folding (and hence function) give rise to 'white' colonies (for the reasons set out in Example 1). The Inventors have found that following this modified procedure, the fraction of precise, full length clones within the 'green' population is significantly higher than that found following the procedure in Example

- 5 1. The same argument hold of course for 'start' codons since all known start codons (ATG, GTG, TTG, ATT, CTG) contain a 'T' in the second position.

A further advantage of this modified procedure is that a polyA tail can be incorporated into the 5'-end of the forward primer used in the initial PCR amplification (e.g.

- 10 Forward-A 5'-AAA AAA AAA AAA GAT CGA TCT C AT GAC GGA TAA CAA TTT CAC  
ACA GG-3'). During amplification with a dTTP: $\alpha$ -S-dTTP ratio of 3:1, there is a high probability then that at least one  $\alpha$ -S-dTTP residue will be incorporated at the end of the complementary strand, at the 5' end of the PCR product. These incorporated nucleotides will be resistant to Exonuclease III digestion and will therefore remove the requirement for enzymatic steps in specifically protecting that end of the PCR product
- 15 from degradation.

### Example 3

20

#### (a) Modification of a second protein using the hexahistidine tag

Following the procedure as described in Example 1 for glutathione-S-transferase, the Inventors have demonstrated that the procedure is independent of the sequence of the gene being manipulated.

- 25 Thus starting with a plasmid encoding human transcription factor NF- $\kappa$ B p50 and following exactly the procedure described in Example 1 unless otherwise specified, the Inventors have been able to demonstrate the modification of NF- $\kappa$ B p50 such that the first in-frame stop codon has been excised and replaced by an in-frame fusion to DNA encoding a polyasparagine, hexahistidine tag and GFP (when the amber stop

codon is suppressed). The clones that fluoresced green, when excited with far uv light (365 nm) were further characterised. Colony Western blots using an anti-His-tag antibody allowed identification of clones expressing hexahistidine-tagged protein. The soluble protein lysates of these clones were resolved by SDS-polyacrylamide gel electrophoresis and probed with anti-His tag antibody. Immunoreactive signals were observed at approx. 65 kDa  $M_r$  (corresponding to translationally fused NF- $\kappa$ B p50 to the hexahistidine tag and GFP) and at approx. 38 kDa  $M_r$  (NF- $\kappa$ B p50-histag). In addition there was a signal at around 27 kDa  $M_r$ , which is probably a degradation product corresponding to the his tagged GFP protein. The sequencing data confirmed that several clones encoded perfect in-frame fusions of full- or almost full-length NF- $\kappa$ B to the hexahistidine tag. In a single experiment, 190 colonies were screened for green fluorescence. A total of 38 clones (20% of the total number of clones screened) fluoresced green when excited with far uv light (365 nm). Colony western blotting using anti-His tag antibody revealed that 29 of the 38 clones expressed the hexahistidine tag. Sequencing data confirmed that 18 of these clones were full-length, or close to full length, in-frame fusions of NF- $\kappa$ B p50 to the hexahistidine tag; 7 of these clones were absolutely full length, His tagged NF- $\kappa$ B p50 genes and the remaining 11 His-tagged clones had short truncations of 4 to 1 amino acid residues. This experiment clearly demonstrates the advantage of having a reporter system indicative of expression and proper folding of the protein. Roughly 50% of the clones that fluoresced green were full length in-frame fusions to a His tag, or had minor truncations which did not cross a domain boundary, fused in-frame to a His tag.

**(b) Immobilisation and functional analysis of hexahistidine-tagged NF- $\kappa$ B p50**

*E. coli* DH5 $\alpha$  cells were transformed with one of the full-length, hexahistidine-tagged NF- $\kappa$ B plasmids created via the above methodology. A single carbenicillin-resistant colony was grown to mid-log phase in 10ml liquid culture and then supplemented with 100 $\mu$ M IPTG to induce expression of the hexahistidine-tagged NF- $\kappa$ B p50. After growth for a further 4 hours, cells were harvested and lysed by sonication. SDS-

PAGE of the crude lysate showed an overexpressed protein at the expected size (38kDa) which represented roughly 5% of total soluble protein.

5           **κB motif**        5'-CGT ATG TTG TGG GGA ATT CCC AGC GGA TAA C-3'  
5' -GCA TAC AAC ACC CCC TAA GGG TCG CCT ATT G-5'  
                                NF-κ B p50 binding site

A duplex oligonucleotide, 'κB motif', which contains a palindromic binding site for NF-κB p50, was labelled at the 3'-bases with digoxigenin using 3-terminal transferase  
10 under standard conditions.

The protein lysates were prepared using the lysozyme/freeze-thaw method in PBS (phosphate buffered saline pH 7.5) containing 5 mM β-mercaptoethanol. 200 µl of the soluble protein lysate from each clone, was applied to the Ni-NTA coated microwell  
15 and incubated at room temperature for 45 minutes. At the end of the incubation period, the wells were washed three times with PBST (PBS containing 0.02 % Triton X-100) to remove all the unbound proteins. The wells were washed three times with DNA binding buffer ( 10 mM Tris.HCl pH 7.4, 75 mM KCl containing 5 mM β- mercaptoethanol with a soak time of 1 minute. The 3' digoxigenin labelled κB motif  
20 (2 pmol) was added to the wells in 200 µl of the DNA binding buffer containing 1 µg of poly (dI-dC) non-specific DNA. After another 30 minutes incubation the unbound DNA was removed by washing the wells three times with 10 mM Tris.HCl pH 7.4, 25 mM KCl containing 0.02% Triton X-100. An anti-digoxigenin antibody-alkaline phosphatase conjugate was diluted to 150mU/ml in 'antibody dilution buffer' (10mM  
25 Tris.HCl pH7.4, 25mM potassium chloride) supplemented with 0.2% bovine serum albumin. The diluted antibody (200µl) was then applied to the microwells. After 30 minutes at room temperature, unbound antibody was removed by washing the microwells with 'antibody dilution buffer' (3x350µl) supplemented with 0.02% Triton X-100. 200µl of a buffer (100mM Tris.HCl pH9.5, 100mM NaCl, 50mM MgCl<sub>2</sub>)

containing 250µM *p*-nitrophenyl phosphate (pNPP), an alkaline phosphatase substrate, was then added to the wells and the reaction allowed to proceed overnight at room temperature, after which the yellow colouration in each well (corresponding to formation of the product, *p*-nitrophenol) was quantitated at 405nm. The background rate of hydrolysis of the substrate pNPP was low so a positive assay result was therefore immediately clear from the appearance of yellow colour in the wells.

As controls in this assay the Inventors omitted either the crude lysate, or the labelled oligonucleotide, or the antibody, or added a 20-fold excess of unlabelled duplex oligo or replaced the hexahistidine-tagged NF-κB p50 containing crude lysates with equivalent amounts of a crude cell lysate from DH5 $\alpha$  cells expressing hexahistidine-tagged GST in the same vector background.

In this assay, NF-κB p50 first binds to the labelled oligonucleotide *via* the specific binding site. The protein-DNA complex is then immobilised in the microwells *via* the hexahistidine tag and all other proteins (including complexes between the labelled oligo and other DNA binding proteins present in the crude lysate) together with any unbound, labelled oligo, are then washed away. Since the antibody-conjugate recognises the label on the oligo, not the hexahistidine-tagged protein, a positive signal in the assay can only be observed if the NF-κB p50-DNA interaction is maintained on immobilisation of NF-κB p50 *via* the tag; if this interaction is not maintained, the oligo will be lost during the washing steps so no colour change will be observed.

The Inventors found that the yellow product was only detected in the microwells which had contained the hexahistidine-tagged NF-κB p50 crude lysate and the digoxigenin-labelled oligonucleotide and to which the anti-digoxigenin antibody-alkaline phosphatase conjugate had been added. This demonstrated that the observed colour change was due specifically to the immobilised NF-κB p50-oligonucleotide complex and moreover that NF-κB p50 retained activity on specific immobilisation.

**Example 4****(a) Identification of one protein from a pool of 10 genes**

The Inventors have applied the procedure exactly as described in Example 1 except where specified to the pool of 10 different genes listed in the table below. The Inventors have generated arrays of the resultant specifically modified proteins such that each position in the array corresponds to a single recombinant protein immobilised through the tag appended as a result of this procedure. The Inventors have then screened the array by functional assay and have successfully identified individual protein components of the pool.

**Table 1. Size and function of the ten genes in the pool**

Gene	Size	Source and Function
glutathione S transferase	950bp	bacterial; detoxification
NF- $\kappa$ B p50	1165bp	human; transcription factor
maltose binding protein	1325bp	bacterial; carbohydrate transport
alanine racemase	1342bp	bacterial; cell wall biosynthesis
nuclear factor of activated T cells (NFAT)	1087bp	murine; transcription factor
indoleglycerolphosphate synthase	1523bp	bacterial; amino acid biosynthesis
phosphoribosylanthranilate isomerase	920bp	bacterial; amino acid biosynthesis
tryptophan synthase ( $\alpha$ -subunit)	1122bp	bacterial; amino acid biosynthesis
chymotrypsin inhibitor 2	389bp	barley; serine protease inhibitor
$\beta$ -lactamase	1040bp	bacterial; antibiotic resistance

15

Initially, all ten genes were subcloned into the same pTrcHisA vector backbone since amongst other things this mimics the situation encountered with a cDNA library. The primers 'STforward' and 'STreverse' described in Example 1 were designed to be universal primers for the amplification of genes encoded within a pTrcHisA vector backbone.

20

The primer 'STforward' was designed such that it encodes a number of restriction sites as follows:

5' -ATG CTG ACG TCA TGA GGC CCA TGG GGC CCG GAT AAC AAT TTC ACA CAG  
5      G-3'  
      *Aat II*    *Bsp HI*        *Sfi I*

Thus, either of the restriction enzymes *Aat* II or *Sfi* I can be used to generate 3'-overhangs for exonuclease protection purposes. For directional cloning purposes at 10 the end of the modification procedure, in this Example the Inventors chose to use *Bsp* HI since although statistically it will cut more frequently within a library, it generates cohesive ends which are compatible with the *Nco* I cloning site in the tag vector pMM106H used here and does not cut within any of the 11 genes in the present pool. Clearly, in principle any of the primer encoded restriction sites could be used 15 providing that the tag vector contains an equivalent cloning site downstream of the promoter; *Sfi* I would have significant advantages in this regard in a larger library format because it has an 8bp recognition sequence so the frequency of random occurrence of an *Sfi* I site within a given gene will be much lower (1 in 6.5x10<sup>4</sup>) than that for a 6bp recognition sequence such as that of *Bsp* HI (1 in 4,096). 20 The tag vector pMM106H is an 'ATG' vector, i.e. the 5'-cloning site (*Nco* I) overlaps the ATG start codon positioned downstream of a ribosome binding site (RBS) for expression of native proteins. However, in the procedure described here the Inventors are not reliant on the cloned genes having a common restriction site at their start 25 codons. Instead, the Inventors simply rely on the vector-encoded promoter initiating transcription to produce mRNA, with the requisite signals for translational initiation being provided by the cloned genes themselves. Thus in this Example, all the genes in the original pool have a start codon immediately preceded by an RBS, irrespective of the presence or absence of a cloning site at the ATG. Since the primer 'STforward' binds upstream of the RBS in all eleven initial clones, subsequent post-modification 30 cloning using any of the primer encoded restriction sites will introduce the newly

modified genes in to the tag vector together with their original RBS and ATG so translation initiation will be ensured. In a cDNA library format, a similar situation applies in that all full-length cDNAs will have their own 5'-untranslated regions (UTR) which contain the eukaryotic translational initiation signals. All that is required to obtain proper translational initiation then is to clone each modified cDNA together with its 5'-UTR in to a eukaryotic vector which provides transcriptional initiation signals so an equivalent universal set of PCR primers to those used in this Example could therefore be used in the modification of every member of a cDNA library in a single pot in a sequence-independent manner.

10

The experimental procedure was carried out as described in Example 1 with the following modifications. An equimolar pool of all ten genes was used as the template for initial PCR amplification using primers 'STforward' and 'STreverse', after which fragments were digested with *Aat II* to protect the 5' end. Exonuclease III and mung bean nuclease-treated fragments were generated exactly as in Example 1 and were then digested with *Bsp HI*, which restricts the fragments uniquely within the forward PCR primer binding site and generates a cohesive end for cloning into the vector pMM106H. The resulting fragments were gel purified and ligated in to the vector. Transformed cells were visualised under UV light (365nm) and colonies which fluoresced green were selected by eye for analysis by Western blot. Approximately 2% of the total number of transformed colonies fluoresced green. Of these, 103 (42%) expressed proteins which are recognised by anti-His tag antibodies. These colonies were inoculated individually into 1.5 ml of liquid medium in 96-deep-well blocks and grown overnight. Cells were harvested by centrifugation and lysed by freeze-thaw/lysozyme. The individual crude lysates were then applied to individual wells of a Nickel-NTA-coated 96-well plate and unbound proteins were removed by washing, leaving a discrete His tagged recombinant protein immobilised in each well. The immobilised proteins were then assayed for either GST or NF- $\kappa$ B activities using the assays described in Examples 1 and 3 and wells containing positive 'hits' were

identified in each case by the appearance of either green or yellow colouration respectively.

In the first assay, three proteins in the array were found to exhibit GST activity, giving a hit rate of approximately 3%. Sequencing of the corresponding plasmids revealed 5 that all three encoded in-frame fusions between a GST gene and the hexahistidine tag and GFP gene; of these three, two were absolutely full-length GST and one was a slight truncation which clearly did not affect activity.

In the second assay, three proteins in the array showed positive 'kB-motif' DNA binding activity. Further characterisation of the positive clones, showed two of the 10 clones were in-frame fusions of the NF- $\kappa$ B p50 gene to the hexahistidine tag, one of which was almost full length (truncated by one amino acid) whilst the other was more severely truncated but contained the entire DNA binding domain of NF- $\kappa$ B p50.

Interestingly, since the assay is designed for binding to the cognate DNA sequence, 15 truncations that still have the DNA binding domain intact, folded, and functional will be positive in the assay. The third clone was found to be an in-frame fusion of the DNA binding domain of the murine transcription factor NFAT to the His tag. The 3'-digoxigenin labelled 'kB motif' used in the assay contains a specific, high affinity ( $K_d$  approximately pM) binding site for NF- $\kappa$ B p50 but this same binding site is also specifically recognised by the DNA binding domain of NFAT with approximately nM 20 affinity. This result therefore demonstrates that functional interrogation of arrays generated by this procedure can identify both specific interactions and also weaker interactions which are nonetheless specific and biologically relevant.

In a subsequent experiment, an array containing ca. 340 of His-tagged proteins was 25 prepared according to the method of this Example. Analysis of the array by GST activity assay showed that 8% of all proteins in the array possessed strong GST activity. In addition, PCR analysis was carried out on a pool of the 340 encoding plasmid DNAs using gene-specific primers and this showed that each gene was represented in the His-tagged collection. These data therefore provide further

confirmation that the method of this example is sequence-independent and can be applied to a collection of different genes.

In summary therefore, the Inventors have used the procedures described in these Examples to create arrays of functional proteins in a microwell format and using these

- 5 arrays the Inventors have successfully identified three different proteins from a pool based on either specific protein-ligand interactions (GST activity assay) or specific protein-DNA interactions (NF- $\kappa$ B binding assay).

**CLAIMS**

1. A method producing one or more proteins in which one or more domains are full length and correctly folded and which are each tagged at either the N- or C- terminus with one or more marker moieties, said method comprising:
  - (a) providing one or more DNA molecules having an open reading frame encoding said proteins together with 5' and/or 3' untranslated regions;
  - (b) amplifying said DNA molecules under conditions that statistically incorporate  $\alpha$ -S-dNTPs as well as dNTPs into the daughter DNA molecules;
  - (c) specifically protecting the 5' or 3' end of said DNA molecules from nuclease digestion;
  - (d) treating said DNA molecules first with a 5' to 3'- or 3' to 5'-nuclease to generate a set of nested deletions followed by treating with a single-strand nuclease under conditions that allow removal of said 5' or 3' untranslated regions including the start or stop codons of said open reading frame;
  - (e) cloning the fragments generated by step (d) into an expression vector containing a coding sequence for one or more 5' or 3' marker moieties;
  - (f) expressing said encoded proteins.
2. A method as claimed in claim 1, wherein said amplification of said DNA molecule statistically incorporates a single  $\alpha$ -S-dNTP.
3. A method as claimed in claim 2, wherein the single  $\alpha$ -S-dNTP is either  $\alpha$ -S-dTTP or  $\alpha$ -S-dATP.
4. A method as claimed in any one of claims 1 to 3, wherein said nuclease is exonuclease III or  $\lambda$  exonuclease.
5. A method as claimed in any one of claims 1 to 4, wherein said single-strand

nuclease is mung bean nuclease or T4 DNA polymerase.

6. A method as claimed in any one of claims 1 to 5, wherein the marker moiety allows confirmation of expression of said open reading frame.

5

7. A method as claimed in any one of claims 1 to 5, wherein the marker moiety allows confirmation of folding of said open reading frame.

8. A method as claimed in any one of claims 1 to 7, wherein the marker moiety 10 encodes the green fluorescent protein.

9. A method as claimed in any one of claims 1 to 7 wherein the marker moiety is a peptide sequence, eg a hexa-histidine tag, a complete protein or protein domain, eg the maltose binding protein domain.

15

10. A method as claimed in claim 9 wherein the tag allows for purification of the individual proteins in the array.

11. A method as claimed in any one of claims 1 to 10 wherein the tag is inserted 20 such that the start or stop codon for each of the proteins is replaced.

12. A method as claimed in any one of claims 1 to 10 wherein the tag is inserted in-frame in a region close to the terminus of each of the proteins which is unimportant for folding and function.

25

13. A method as claimed in any one of claims 1 to 10 wherein the tag is inserted in-frame within the open reading frame but in a region outside specific domain boundaries which is unimportant for folding and function.

30 14. A method as claimed in any one of claims 1 to 13 wherein amplification of said

DNA molecule in step (a) is by a non-proof-reading polymerase, e.g. Taq polymerase or the Klenow fragment of DNA polymerase I.

5 15. A method as claimed in any one of claims 1 to 14 wherein the ratio of said  $\alpha$ -S-dNTPs to dNTPs is between 1:1 and 1:3

16. A method as claimed in any one of claims 1 to 15 wherein said 5' to 3' or 3'-to 5'- nuclease is unable to hydrolyse  $\alpha$ -S-phosphodiester linkages.

10 17. A method as claimed in any one of claims 1 to 16 wherein said DNA molecule is a cDNA produced by reverse transcription from a mRNA sequence.

18. A method as claimed in any one of claims 1 to 17 wherein said method is carried out on multiple DNA molecules in parallel.

15 19. A method as claimed in any one of claims 1 to 18 wherein said method is carried out on a population of DNA molecules in a single pot.

20 20. A library of tagged proteins produced by the method of any one of claims 1 to 19.

25 21. A method for producing a protein array, said method comprising:  
(a) clonally separating each member of the library of claim 20;  
(b) expressing the individual tagged proteins in a spatially separated format;  
(c) purifying each tagged protein by means of the marker moiety;  
(d) depositing each protein in to a spatially defined array.

22. An array comprising proteins prepared by a method as defined in any one of claims 1 to 19 or produced by the method defined in claim 21.

23. An array as claimed in claim 22 wherein the components of the array are immobilised, eg to a solid surface.
24. An array as claimed in claim 22 or 23 wherein the individual proteins are immobilised by means of the tag moiety.
25. A method of screening one or more compounds for biological activity which comprises the step of bringing said one or more compounds into contact with a protein array as defined in any one of claims 22 to 24 and measuring binding of the one or 10 more compounds to the proteins in the array.
26. A method of screening one or more proteins for specific protein-protein interactions which comprises the step of bringing said one or more proteins, eg a cell surface receptor, into contact with an array as defined in any one of claims 22 to 24, 15 and measuring binding of the one or more specific proteins with the proteins of the array.
27. A method of screening one or more proteins for specific protein-nucleic acid interactions which comprises the step of bringing said one or more nucleic acid probes 20 into contact with an array as defined in any one of claims 22 to 24, and measuring binding and measuring binding of the probes to the proteins in the array.
28. The use of an array as defined in any one of claims 22 to 24 in the rapid screening of a compound, protein or nucleic acid.
- 25 29. The use of an array as defined in any one of claims 22 to 24 in screening for molecules which recognise each protein in the array, wherein the molecules are preferably antibodies.
- 30 30. A method of generating an antibody array which comprises bringing a protein

array as defined in any one of claims 22 to 24 into contact with an antibody library, such that one or more proteins in the protein array bind to at least one antibody in the antibody library, removing any unbound antibodies and immobilisation of those antibodies bound to proteins in the protein array.

5

31. A method for the screening of protein function or abundance which comprises the step of bringing an antibody array as defined in claim 30 into contact with a mixture of one or more proteins.

10

32. A method as claimed in any one of claims 25 to 27 or 30 which also comprises the step of first providing the protein array as defined in any one of claims 22 to 24

33. The method as defined in claim 21, wherein the proteins in the array are purified and immobilised in a single step.

15

34. The use of a tagged protein produced by the methods as defined herein.

35. The use of a tagged protein produced by the methods as defined herein for analysis of interaction between expressed protein and other proteins

20

36. The use of a tagged protein produced by the methods as defined herein for immobilization on an affinity column/substrate, for example to allow the purification by affinity chromatography of, a) interacting proteins, b) DNA or c) chemical compounds.

25

37. The use of a tagged protein produced by the methods as defined herein in the immobilization by affinity purification for interrogation by antibodies (ELISA assay) as a diagnostic tool.

30

38. The use of a tagged protein produced by the methods as defined herein as a

probe for a cDNA microarray

39. The use as defined in claim 38 for the identification of DNA binding proteins.

5 40. The use of a tagged protein produced by the methods as defined herein for  
elucidating the identity of proteins in the 'proteome'

41. The use as defined in claim 40 wherein mass spectrometric analysis of  
expressed protein components of source library or start material modified by the  
10 methods of the invention is performed.

WO 02/27327

PCT/GB01/03693

1/4

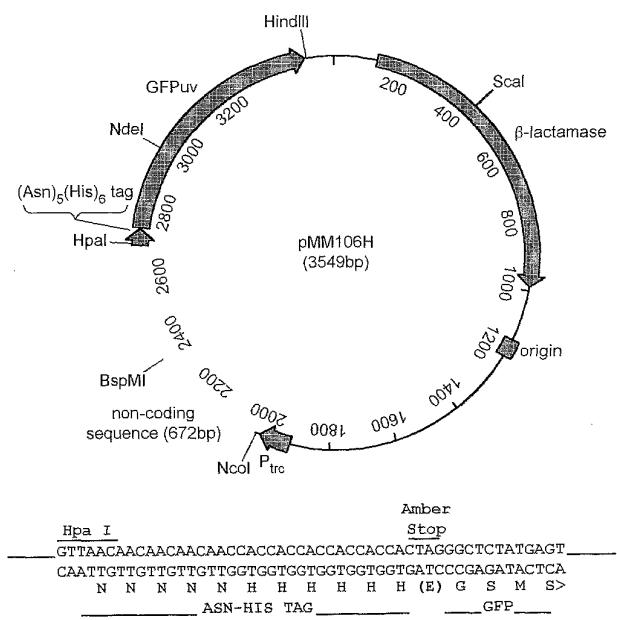


FIG. 1a

**SUBSTITUTE SHEET (RULE 26)**

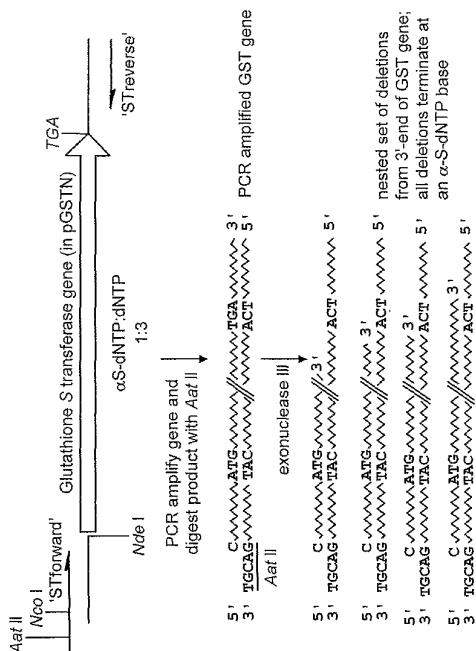


FIG. 1b

SUBSTITUTE SHEET (RULE 26)

3 / 4

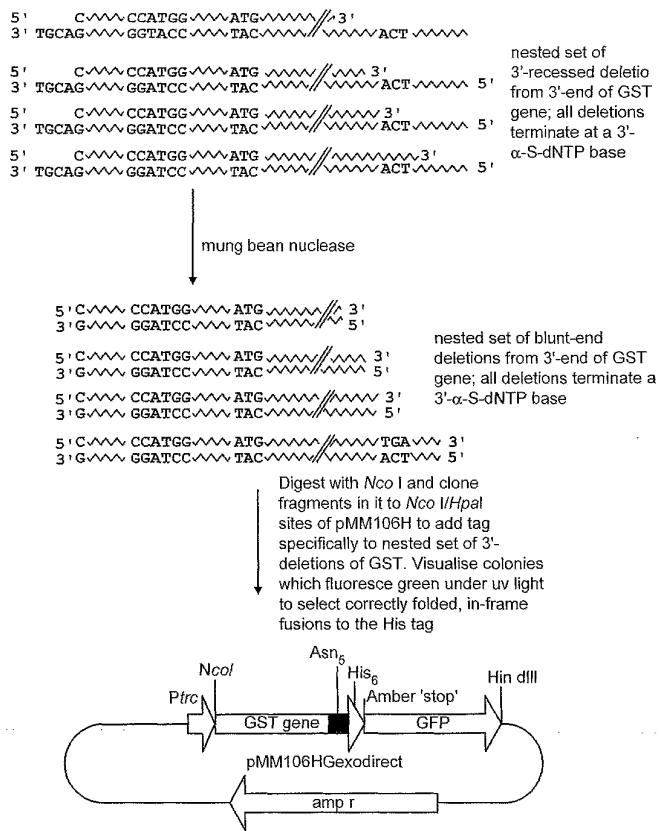


FIG. 1C

SUBSTITUTE SHEET (RULE 26)

WO 02/27327

PCT/GB01/03693

4 / 4

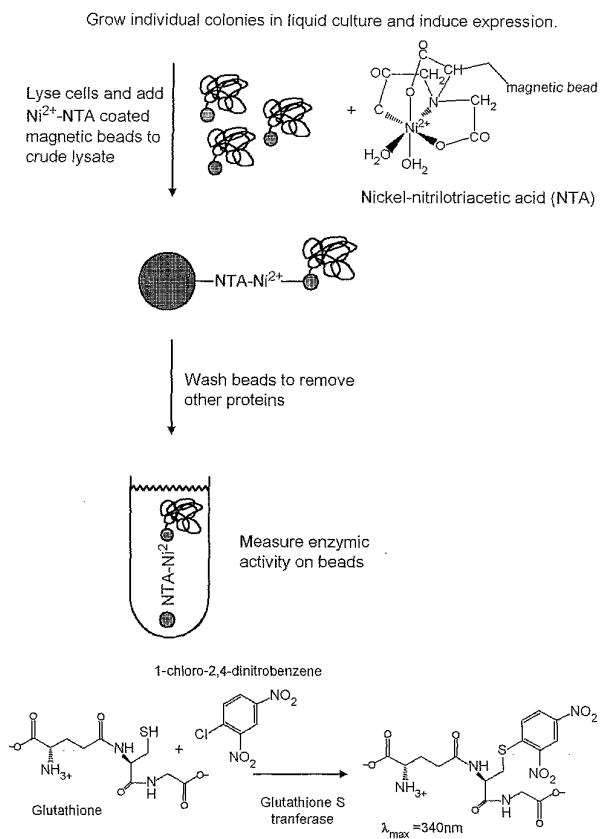


FIG. 1d

SUBSTITUTE SHEET (RULE 26)

## 【国際公開パンフレット（コレクトバージョン）】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization  
International Bureau(43) International Publication Date  
4 April 2002 (04.04.2002)

PCT

(10) International Publication Number  
WO 02/027327 A3(51) International Patent Classification: C12N 15/10,  
15/66, 15/62, C07K 1/13, 16/00, C12Q 1/68, G01N 33/68

(21) International Application Number: PCT/GB01/03693

(22) International Filing Date: 17 August 2001 (17.08.2001)

(25) Filing Language: English

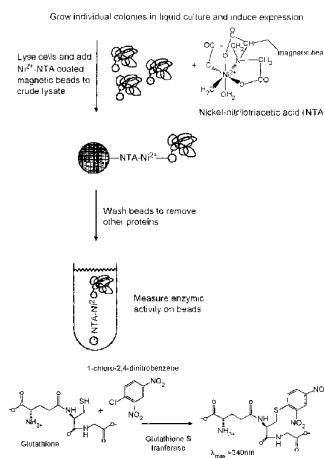
(26) Publication Language: English

(30) Priority Data:  
0020357.0 17 August 2000 (17.08.2000) GB  
60/247,995 14 November 2000 (14.11.2000) US(71) Applicant (for all designated States except US): SENSE  
PROTEOMIC LIMITED [GB/GB]: The Babraham  
Biocentre, Babraham Hall, Babraham, Cambridgeshire  
CB2 4AT (GB).(72) Inventors; and  
(73) Inventors/Applicants (for US only): BLACKBURN,  
Jonathan, Michael [GB/GB]; 36 Woodlark Road, Cam-  
bridge CB3 0IS (GB), MULDER, Michelle, Anne  
[ZA/GB]; 83 Beech Road, Cambridge CB5 8IU (GB).SAMADIDAR, Mitai [IN/GB]; 40 Chatsworth Avenue,  
Cambridge CB4 3LT (GB), KOZLOWSKI, Roland  
[GB/GB], Sense Proteomic Limited, The Babraham Bioin-  
cubator, Babraham Hall, Babraham, Cambridge CB2 4AT  
(GB).(74) Agents: CHAPMAN, P. W. et al.; Kilburn & Strode, 20  
Red Lion Street, London WC1R 4PJ (GB).(81) Designated States (national): AE, AG, AL, AM, AT, AU,  
AZ, BA, BB, BG, BR, BY, BZ, CA, CI, CN, CO, CR, CU,  
CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GI,  
GM, IIR, IIU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC,  
LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW,  
MX, MZ, NO, NZ, PH, PL, PT, RO, RU, SD, SE, SG, SI

{Continued on next page}

(54) Title: RAPID PROFILING OF THE INTERACTIONS BETWEEN A CHEMICAL ENTITY AND PROTEINS IN A GIVEN  
PROTEOME

WO 02/027327 A3

(57) Abstract: The present invention relates to novel meth-  
ods of producing proteins in which one or more domains are  
full length and correctly folded and which are each tagged  
at either the N- or C-terminus with one or more marker mo-  
ieties and arrays containing such proteins, as well as the use  
of such proteins in arrays for rapid screening.

WO 02/027327

A3

**WO 02/027327 A3**SK, SL, TJ, TM, TR, TI, TZ, UA, UG, US, UZ, VN, YU,  
ZA, ZW.**Published:**  
— with international search report**(84) Designated States (regional):** ARIPO patent (GH, GM,  
KI, LS, MW, MZ, SD, SI, SZ, TZ, UG, ZW), Eurasian  
patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European  
patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE,  
IT, LU, MC, NL, PL, SE, TR), OAPI patent (BF, BJ, CF,  
CG, CI, CM, GA, GN, GQ, GW, MI, MR, NE, SN, TD,  
TG).**(88) Date of publication of the international search report:**  
17 April 2003*For two-letter codes and other abbreviations, refer to the "Guide-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*

## 【国際調査報告】

INTERNATIONAL SEARCH REPORT		International Application No PCT/GB 01/03693															
<b>A. CLASSIFICATION OF SUBJECT MATTER</b> IPC 7 C12N15/10 C12N15/66 C12N15/62 C07K1/13 C07K16/00 C12Q1/68 G01N33/68																	
According to International Patent Classification (IPC) or to both national classification and IPC																	
<b>B. FIELDS SEARCHED</b> Minimum documentation searched (classification system followed by classification symbols) IPC 7 C12N C07K C12Q G01N																	
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched																	
Electronic data base consulted during the international search (name of data base and, where practical, search terms used) WPI Data, PAJ, CAB Data, SEQUENCE SEARCH, BIOSIS, EPO-Internal, MEDLINE																	
<b>C. DOCUMENTS CONSIDERED TO BE RELEVANT</b> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="text-align: left; width: 15%;">Category *</th> <th style="text-align: left; width: 60%;">Citation of document, with indication, where appropriate, of the relevant passages</th> <th style="text-align: left; width: 25%;">Relevant to claim No.</th> </tr> </thead> <tbody> <tr> <td>P, X</td> <td>WO 01 57198 A (SENSE PROTEOMIC LTD ; SUTHERLAND JOHN DAVID (GB); BLACKBURN JONATHAN) 9 August 2001 (2001-08-09) the whole document ---</td> <td>1-41</td> </tr> <tr> <td>A</td> <td>WO 00 04382 A (ZYOMYX INC) 27 January 2000 (2000-01-27) the whole document ---</td> <td></td> </tr> <tr> <td>A</td> <td>WO 00 09654 A (HEDLEY MARY LYNNE ; CHICZ ROMAN M (US); HSU CHARLES (US); ZYCOS INC) 24 February 2000 (2000-02-24) the whole document ---</td> <td></td> </tr> <tr> <td>A</td> <td>WO 99 39210 A (MILLER SAMUEL ; HUMPHREY SMITH IAN (AU)) 5 August 1999 (1999-08-05) the whole document ---</td> <td>-/-</td> </tr> </tbody> </table>			Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.	P, X	WO 01 57198 A (SENSE PROTEOMIC LTD ; SUTHERLAND JOHN DAVID (GB); BLACKBURN JONATHAN) 9 August 2001 (2001-08-09) the whole document ---	1-41	A	WO 00 04382 A (ZYOMYX INC) 27 January 2000 (2000-01-27) the whole document ---		A	WO 00 09654 A (HEDLEY MARY LYNNE ; CHICZ ROMAN M (US); HSU CHARLES (US); ZYCOS INC) 24 February 2000 (2000-02-24) the whole document ---		A	WO 99 39210 A (MILLER SAMUEL ; HUMPHREY SMITH IAN (AU)) 5 August 1999 (1999-08-05) the whole document ---	-/-
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.															
P, X	WO 01 57198 A (SENSE PROTEOMIC LTD ; SUTHERLAND JOHN DAVID (GB); BLACKBURN JONATHAN) 9 August 2001 (2001-08-09) the whole document ---	1-41															
A	WO 00 04382 A (ZYOMYX INC) 27 January 2000 (2000-01-27) the whole document ---																
A	WO 00 09654 A (HEDLEY MARY LYNNE ; CHICZ ROMAN M (US); HSU CHARLES (US); ZYCOS INC) 24 February 2000 (2000-02-24) the whole document ---																
A	WO 99 39210 A (MILLER SAMUEL ; HUMPHREY SMITH IAN (AU)) 5 August 1999 (1999-08-05) the whole document ---	-/-															
<input checked="" type="checkbox"/> Further documents are listed in the continuation of box C.		<input checked="" type="checkbox"/> Patent family members are listed in annex.															
* Special categories of cited documents: *A* document defining the general state of the art which is not considered to be of "particular relevance" *E* earlier document but published on or after the international filing date *L* document which may throw doubt on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *C* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date but not in conflict with the application but of interest to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. *Z* document member of the same patent family																	
Date of the actual completion of the international search	Date of mailing of the international search report																
21 October 2002	07/11/2002																
Name and mailing address of the ISA European Patent Office, P.B. 5518 Patenttaan 2 NL 2280 RD The Hague Tel: (+31-70) 340-2040, Fax: 31 651 epo nl, Fax: (+31-70) 340-3016	Authorized officer  Hornig, H																

Form PCT/ISA/210 (second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT		tional Application No PCT/GB 01/03693
C(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 99 51773 A (PHYLOS INC) 14 October 1999 (1999-10-14) the whole document	
A	WO 99 11777 A (CARR FRANCIS JOSEPH ; ADAIR FIONA (GB); BIOVATION LTD (GB); CARTER) 11 March 1999 (1999-03-11) the whole document	
P,A	WO 01 04265 A (PHYLOS INC) 18 January 2001 (2001-01-18) the whole document	
A	PANDEY AKHILESH ET AL: "Proteomics to study genes and genomes" NATURE, MACMILLAN JOURNALS LTD. LONDON, GB, vol. 405, no. 6788, 15 June 2000 (2000-06-15), pages 837-846, XP002172041 ISSN: 0028-0836 the whole document	
A	ROBERT L STRAUSBERG AND M J FINLEY AUSTIN: "Functional genomics: technological challenges and opportunities" PHYSIOLOGICAL GENOMICS, AMERICAN PHYSIOLOGICAL SOCIETY, BETHESDA, MD, US, vol. 1, no. 1, 15 July 1999 (1999-07-15), pages 25-32, XP002172042 ISSN: 1094-8341 the whole document	
T	SPEIGHT R E ET AL: "A new plasmid display technology for the in vitro selection of functional phenotype-genotype linked proteins." CHEMISTRY & BIOLOGY. ENGLAND OCT 2001, vol. 8, no. 10, October 2001 (2001-10), pages 951-965, XP002217504 ISSN: 1074-5521 the whole document	

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

## INTERNATIONAL SEARCH REPORT

Information on patent family members

National Application No  
PCT/GB 01/03693

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0157198	A 09-08-2001	AU 3037501 A EP 1203238 A2 WO 0157198 A2 GB 2361698 A	14-08-2001 08-05-2002 09-08-2001 31-10-2001
WO 0004382	A 27-01-2000	US 6406921 B1 AU 5102399 A AU 5102599 A CA 2337075 A1 CA 2337654 A1 EP 1097377 A2 EP 1097380 A1 JP 2002520618 T JP 2002520620 T US 2002132272 A1 WO 0004389 A2 WO 0004382 A1 US 2002106702 A1 US 2002110933 A1 US 6329209 B1 US 6365418 B1 US 2002119579 A1	18-06-2002 07-02-2000 07-02-2000 27-01-2000 27-01-2000 09-05-2001 09-05-2001 09-07-2002 09-07-2002 19-09-2002 27-01-2000 27-01-2000 08-08-2002 15-08-2002 11-12-2001 02-04-2002 29-08-2002
WO 0009654	A 24-02-2000	AU 5336799 A CA 2339817 A1 EP 1105731 A2 JP 2002522779 T WO 0009654 A2	06-03-2000 24-02-2000 13-06-2001 23-07-2002 24-02-2000
WO 9939210	A 05-08-1999	AU 740830 B2 AU 2259799 A WO 9939210 A1 CA 2319828 A1 EP 1051624 A1 JP 2002502038 T	15-11-2001 16-08-1999 05-08-1999 05-08-1999 15-11-2000 22-01-2002
WO 9951773	A 14-10-1999	AU 3463699 A CA 2323638 A1 EP 1068356 A1 JP 2002510505 T WO 9951773 A1	25-10-1999 14-10-1999 17-01-2001 09-04-2002 14-10-1999
WO 9911777	A 11-03-1999	AU 8879298 A CA 2302147 A1 EP 1003853 A1 WO 9911777 A1 JP 2001514850 T	22-03-1999 11-03-1999 31-05-2000 11-03-1999 18-09-2001
WO 0104265	A 18-01-2001	AU 7133800 A EP 1194594 A2 WO 0104265 A2	30-01-2001 10-04-2002 18-01-2001

Form PCT/I(SA)210 (patent family annex) (July 1992)

## フロントページの続き

(51) Int.Cl. <sup>7</sup>	F I	テーマコード(参考)
C 1 2 Q 1/68	C 1 2 Q 1/68	A 4 H 0 4 5
G 0 1 N 27/62	G 0 1 N 27/62	V
G 0 1 N 33/15	G 0 1 N 33/15	Z
G 0 1 N 33/50	G 0 1 N 33/50	Z
G 0 1 N 33/53	G 0 1 N 33/53	D
G 0 1 N 33/566	G 0 1 N 33/53	M
	G 0 1 N 33/566	

(81)指定国 AP(GH,GM,KE,LS,MW,MZ,SD,SL,SZ,TZ,UG,ZW),EA(AM,AZ,BY,KG,KZ,MD,RU,TJ,TM),EP(AT,BE,CH,CY,DE,DK,ES,FI,FR,GB,GR,IE,IT,LU,MC,NL,PT,SE,TR),OA(BF,BJ,CF,CG,CI,CM,GA,GN,GQ,GW,ML,MR,NE,SN,TD,TG),AE,AG,AL,AM,AT,AU,AZ,BA,BB,BG,BR,BY,BZ,CA,CH,CN,CO,CR,CU,CZ,DE,DK,DM,DZ,EC,EE,ES,FI,GB,GD,GE,GH,GM,HR,HU,ID,IL,IN,IS,JP,KE,KG,KP,KR,KZ,LC,LK,LR,LS,LT,LU,LV,MA,MD,MG,MK,MN,MW,MX,MZ,NO,NZ,PH,PL,PT,R0,RU,SD,SE,SG,SI,SK,SL,TJ,TM,TR,TT,TZ,UA,UG,US,UZ,VN,YU,ZA,ZW

(74)代理人 100075672

弁理士 峰 隆司

(74)代理人 100109830

弁理士 福原 淑弘

(74)代理人 100084618

弁理士 村松 貞男

(74)代理人 100092196

弁理士 橋本 良郎

(72)発明者 ブラックバーン、ジョナサン・マイケル

イギリス国、シービー3・0エイチエス、ケンブリッジ、ウッドラーク・ロード 3 6

(72)発明者 マルダー、ミッチャエル・アン

南アフリカ共和国 7 9 4 5 ケープタウン、トーカイ、パース・ロード 1 0

(72)発明者 サマッダー、ミタリ

インド国 5 0 0 0 1 6 ハイデラバード、アメールペット 7 - 1 - 2 7、ドクター・レディース・ラボラトリーズ・リミテッド内、バイオテクノロジー・ディビジョン

(72)発明者 コズロースキー、ローランド

イギリス国、オーエックス20・1ディーエル、オックスフォードシャー、ウッドストック、ウートン、ウエスト・エンド、ザ・コーナー・ハウス(番地なし)

F ターム(参考) 2G045 AA34 AA35 BB50 BB51 DA13 DA36 FA27 FA29 FA40 FB01  
FB02 FB03 FB05 FB06 FB12 GC15 JA01  
4B024 AA11 BA80 CA04 DA02 DA05 DA11 EA02 EA03 EA04 FA01  
GA11 HA03 HA11 HA12  
4B063 QA01 QA18 QQ42 QQ79 QR48 QR55 QR59 QR80 QR82 QS05  
QS15 QS25 QS32 QS36 QX02  
4B064 AG01 CA01 CA19 CC24 DA13  
4G066 AB26B AE19 AE20 CA54 DA07 EA02 FA01 FA15 FA21  
4H045 AA20 AA30 BA10 BA41 EA50 FA72 FA74