(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)公表番号
特表2023-549850
(P2023-549850A)
(43)公表日　令和5年11月29日(2023.11.29)

(51)国際特許分類		FI			テーマコード（参考）
Ｃ１２Ｎ	*15/10 (2006.01)*	Ｃ１２Ｎ	15/10	２１０Ｚ	４Ｂ０６３
Ｃ１２Ｑ	*1/6869(2018.01)*	Ｃ１２Ｑ	1/6869	Ｚ	４Ｂ０６５
Ｃ１２Ｎ	*1/20 (2006.01)*	Ｃ１２Ｎ	15/10	２２０Ｚ	
Ｃ１２Ｎ	*15/63 (2006.01)*	Ｃ１２Ｎ	1/20	Ａ	
		Ｃ１２Ｎ	15/63	Ｚ	

審査請求　未請求　予備審査請求　未請求　（全95頁）

(21)出願番号　特願2023-528634(P2023-528634)
(86)(22)出願日　令和3年11月13日(2021.11.13)
(85)翻訳文提出日　令和5年7月11日(2023.7.11)
(86)国際出願番号　PCT/US2021/059275
(87)国際公開番号　WO2022/104164
(87)国際公開日　令和4年5月19日(2022.5.19)
(31)優先権主張番号　63/113,571
(32)優先日　令和2年11月13日(2020.11.13)
(33)優先権主張国・地域又は機関
米国(US)
(81)指定国・地域　AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,
最終頁に続く

(71)出願人　523176152
トリプルバー バイオ， インコーポレイテッド
アメリカ合衆国 カリフォルニア ９４６０８， エメリービル， ４５ティーエイチ ストリート １２５０， スイート １５０
(74)代理人　100078282
弁理士　山本 秀策
(74)代理人　100113413
弁理士　森下 夏樹
(74)代理人　100181674
弁理士　飯田 貴敏
(74)代理人　100181641
弁理士　石川 大輔
最終頁に続く

(54)【発明の名称】　マルチパラメトリックな発見および最適化プラットフォーム

(57)【要約】
本明細書では、ハイスループット統合システムを使用して望ましい生物学的バリアントをスクリーニングするためのシステムおよび方法が提供される。統合システムは、生物学的バリアントの統合ライブラリと併せて、適用条件下で生物学的バリアントの機能研究から複数のパラメータを入力し、入力をフィルタリングして、入力した性能要件に基づいて望ましい生物学的バリアントを生成するように構成することができる。システムは、望ましい機能的特性を有すると予想される、最適化された株、分子、または新規な分子を出力することができる。したがって、本明細書に開示される方法およびシステムは、システム生物学における生物学的多様性および条件付き多様性のマルチパラメトリック研究を可能にする。

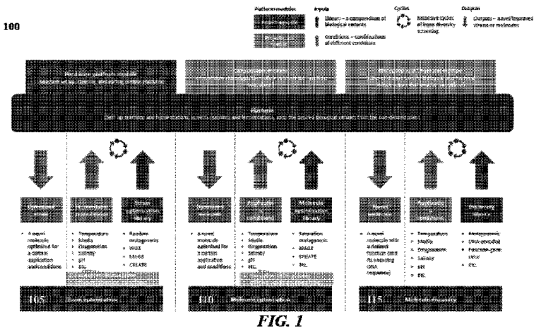


FIG. 1

【特許請求の範囲】
【請求項１】
改善されたポリマーバリアントを生成する方法であって、
（ａ）データセットを取得することであって、
ｉ）ポリマーバリアント配列のデータを含む配列データと、
ｉｉ）ポリマーバリアント機能に基づいて選別された試料中のポリマーバリアント配列の存在量に基づく濃縮データと、を含む、取得することと、
（ｂ）１またはそれを超える機械学習アルゴリズムを前記データセットに適用して、改善されたポリマーバリアントを設計することと、を含む、方法。
【請求項２】
前記ポリマーバリアント配列が、ｉ）選別されていない試料中およびｉｉ）前記ポリマーバリアント機能に基づいて選別された前記試料中の配列を含み、前記ポリマーバリアント機能に基づいて選別された前記試料および前記選別されていない試料が共通の試料に由来する、請求項１に記載の方法。
【請求項３】
前記濃縮データが、前記選別されていない試料中の前記ポリマーバリアント配列の存在量にさらに基づく、請求項２に記載の方法。
【請求項４】
前記ポリマーバリアント機能に基づいて選別された前記試料中の前記ポリマーバリアント配列の存在量がシーケンスカウントを含む、請求項３に記載の方法。
【請求項５】
前記選別されていない試料中の前記ポリマーバリアント配列の存在量がシーケンスカウントを含む、請求項３または４に記載の方法。
【請求項６】
前記濃縮データが、ｉ）前記選別されていない試料中の前記ポリマーバリアント配列中の前記第１のポリマーバリアント配列の存在量、およびｉｉ）前記ポリマーバリアント機能に基づいて選別された前記試料の前記サブセット中の前記ポリマーバリアント配列中の前記第１のポリマーバリアント配列の存在量に基づいて、第１のポリマーバリアント配列の濃縮係数を含む、請求項３から５のいずれか一項に記載の方法。
【請求項７】
前記濃縮係数が、前記第１のポリマーバリアント配列の機能的標識である、請求項６に記載の方法。
【請求項８】
前記方法が、（ａ）および（ｂ）の複数世代を実施することを含み、（ｂ）の前記改善されたポリマーバリアントを使用して、前記複数世代の後続世代において（ａ）の前記データセットを生成する、請求項１から７のいずれか一項に記載の方法。
【請求項９】
前記複数世代が、少なくとも３世代または４世代である、請求項８に記載の方法。
【請求項１０】
前記ポリマーバリアントがタンパク質を含む、請求項１から９のいずれか一項に記載の方法。
【請求項１１】
前記タンパク質が酵素を含む、請求項１０に記載の方法。
【請求項１２】
前記機能が前記酵素の活性を含む、請求項１１に記載の方法。
【請求項１３】
前記改善されたポリマーバリアントが、少なくとも１０００の改善されたポリマーバリアントを含む、請求項１から１２のいずれか一項に記載の方法。
【請求項１４】
前記ポリマーバリアント配列が、前記ポリマーバリアントをコードする核酸配列を含む

、請求項１から１３のいずれか一項に記載の方法。
【請求項１５】
前記データセットを取得する前に、前記ポリマーバリアントを含む細胞のライブラリを提供することをさらに含み、前記ポリマーバリアントが複数のポリマーバリアントを含む、請求項１に記載の方法。
【請求項１６】
前記ポリマーバリアントを含む細胞の前記ライブラリを生成することをさらに含む、請求項１５に記載の方法。
【請求項１７】
前記生成することが、各位置に２０個の天然に存在するアミノ酸のそれぞれを有するタンパク質をコードする部位飽和核酸分子ライブラリ（ＳＳＬ）を提供することを含む、請求項１６に記載の方法。
【請求項１８】
前記ポリマーバリアントをコードする核酸分子を生成するために、前記ＳＳＬに対して遺伝子シャッフリングおよびエラープローンポリメラーゼ連鎖反応（ｅｐＰＣＲ）を実施することをさらに含む、請求項１７に記載の方法。
【請求項１９】
前記ポリマーバリアントをコードする前記核酸分子を細胞に形質転換して、前記細胞のライブラリを生成することをさらに含む、請求項１８に記載の方法。
【請求項２０】
前記細胞が細菌細胞を含む、請求項１９に記載の方法。
【請求項２１】
複数のパーティションを生成するために前記細胞のライブラリを分配することをさらに含む、請求項１５に記載の方法。
【請求項２２】
前記複数のパーティションの各パーティションが、平均して、１つ以下の細胞を含む、請求項２１に記載の方法。
【請求項２３】
前記複数のパーティションにおいて前記細胞のライブラリを拡大することをさらに含む、請求項２１に記載の方法。
【請求項２４】
前記ポリマーバリアントを前記複数のパーティション内の細胞のライブラリにおいて発現させることをさらに含む、請求項２３に記載の方法。
【請求項２５】
前記ポリマーバリアントを含む細胞の前記ライブラリを使用して前記ポリマーバリアント機能を決定することをさらに含む、請求項１５に記載の方法。
【請求項２６】
前記ポリマーバリアントがタンパク質を含む、請求項２５に記載の方法。
【請求項２７】
前記タンパク質が酵素を含む、請求項２６に記載の方法。
【請求項２８】
前記ポリマーバリアント機能を決定することが、酵素基質および反応緩衝液を使用することを含む、請求項２７に記載の方法。
【請求項２９】
前記ポリマーバリアント機能が酵素活性を含む、請求項２８に記載の方法。
【請求項３０】
前記酵素活性が蛍光生成物を生成する、請求項２９に記載の方法。
【請求項３１】
前記ポリマーバリアントを含む前記細胞ライブラリのサブセットを前記ポリマーバリアント機能に基づいて選別して、前記ポリマーバリアント機能に基づいて選別された前記試

料および選別されていない試料を提供することをさらに含む、請求項１５に記載の方法。
【請求項３２】
前記機能が酵素活性の量である、請求項３１に記載の方法。
【請求項３３】
前記選別することが、前記細胞のライブラリの前記サブセットを、前記機能に基づいて複数のグループに選別することを含む、請求項３１に記載の方法。
【請求項３４】
前記複数のグループが少なくとも４つのグループである、請求項３２に記載の方法。
【請求項３５】
前記細胞のライブラリのサブセットが、前記細胞のライブラリに存在すると予想される数の約５０倍～約１００倍のユニークなポリマーバリアントを含む、請求項３１に記載の方法。
【請求項３６】
前記ポリマーバリアント機能に基づいて選別された前記試料を使用して前記ポリマーバリアント配列を決定することをさらに含む、請求項３１に記載の方法。
【請求項３７】
前記ポリマーバリアント機能に基づいて選別された前記試料が基準配列を含む、請求項３６に記載の方法。
【請求項３８】
前記ポリマーバリアント機能に基づいて選別された前記試料を使用して前記ポリマーバリアント配列を決定することが、前記ポリマーバリアント機能に基づいて選別された前記試料から核酸分子を精製することを含む、請求項３６に記載の方法。
【請求項３９】
前記選別されていない試料から核酸分子を精製することをさらに含む、請求項３１に記載の方法。
【請求項４０】
前記ポリマーバリアント機能に基づいて選別された前記試料から核酸分子を精製することが、前記選別されていない試料から核酸分子を精製することとは別個に生じる、請求項３９に記載の方法。
【請求項４１】
前記ポリマーバリアント機能に基づいて選別された前記試料からの前記核酸分子および前記選別されていない試料からの前記核酸分子をシーケンシングして、前記ポリマーバリアント配列を提供することをさらに含む、請求項３１に記載の方法。
【請求項４２】
前記シーケンシングすることが次世代シーケンシングすることを含む、請求項４１に記載の方法。
【請求項４３】
前記ポリマーバリアント配列を使用して、ｉ）前記ポリマーバリアント機能に基づいて選別された前記試料中およびｉｉ）前記選別されていない試料中の前記ポリマーバリアント配列の存在量を決定することをさらに含む、請求項４１に記載の方法。
【請求項４４】
ｉ）前記ポリマーバリアント機能に基づいて選別された前記試料中およびｉｉ）前記選別されていない試料中の前記ポリマーバリアント配列の存在量を使用して、前記濃縮データを決定することをさらに含む、請求項４３に記載の方法。
【請求項４５】
前記濃縮データが濃縮係数を含む、請求項４４に記載の方法。
【請求項４６】
前記改善されたポリマーバリアントを合成することをさらに含む、請求項１に記載の方法。
【請求項４７】

前記合成することが、オリゴヌクレオチドアレイ上での合成を含む、請求項４６に記載の方法。
【請求項４８】
前記改善されたポリマーバリアントをベクターにアセンブリすることをさらに含む、請求項４７に記載の方法。
【請求項４９】
前記ベクターがプラスミドである、請求項４８に記載の方法。
【請求項５０】
前記ベクターを宿主細胞に導入することをさらに含む、請求項４８に記載の方法。
【請求項５１】
前記複数世代のうちの第１世代の（ａ）が、飽和突然変異誘発およびｅｐＰＣＲを実施して細胞のライブラリを提供することを含み、前記複数世代のうちの第２世代の（ａ）が、ｅｐＰＣＲを実施することを含み、前記複数世代のうちの第３世代の（ａ）が、ｅｐＰＣＲを実施することを含む、請求項８に記載の方法。
【請求項５２】
前記複数世代の各世代が、複数回の前記選別を含む、請求項５１に記載の方法。
【請求項５３】
コンピュータプロセッサによって実行されると前記コンピュータプロセッサに請求項１から１４のいずれか一項に記載の方法を実行させる命令を含む、非一時的なコンピュータ可読媒体。
【請求項５４】
コンピュータプロセッサと、請求項５３に記載の非一時的なコンピュータ可読媒体とを含むシステム。
【請求項５５】
１またはそれを超えるプロセッサとコンピュータ可読媒体とを含む、細胞を処理するためのシステムであって、前記コンピュータ可読媒体が、前記１またはそれを超えるプロセッサによって実行されると前記システムに、
（ａ）細胞の集団を複数の反応体積に分配して、分配された細胞を得ることと、
（ｂ）前記分配された細胞を表現型についてアッセイして、分配された細胞に対するシグナルを生成することを実行させる命令を含み、
前記分配された細胞に対するシグナルの変動係数（ＣＶ）が、前記細胞の集団が均一な細胞の集団である場合、１１％またはそれ未満である、システム。
【請求項５６】
前記ＣＶが１０％またはそれ未満である、請求項５５に記載システム。
【請求項５７】
前記ＣＶが７％またはそれ未満である、請求項５５に記載システム。
【請求項５８】
前記ＣＶが約５％～約１０％である、請求項５５に記載システム。
【請求項５９】
前記均一な細胞の集団が少なくとも１００万個の細胞を含む、請求項５５に記載のシステム。
【請求項６０】
前記均一な細胞の集団が少なくとも１０，０００，０００個の細胞を含む、請求項５９に記載のシステム。
【請求項６１】
前記反応体積が液滴を含む、請求項５５に記載のシステム。
【請求項６２】
前記アッセイが、前記分配された細胞への検出試薬の添加を含む、請求項５５に記載のシステム。
【請求項６３】

前記アッセイが、前記分配された細胞への前記検出試薬の添加のタイミングの制御を含む、請求項６２に記載のシステム。
【請求項６４】
前記検出試薬が酵素基質を含む、請求項６３に記載のシステム。
【請求項６５】
前記アッセイが、前記分配された細胞に対する前記シグナルの検出を含む、請求項５５に記載のシステム。
【請求項６６】
前記検出が光学的検出を含む、請求項６５に記載のシステム。
【請求項６７】
前記アッセイが、前記分配された細胞に対する検出シグナルのタイミングの制御を含む、請求項５５に記載のシステム。
【請求項６８】
前記表現型が酵素活性を含む、請求項５５に記載のシステム。
【請求項６９】
（ａ）が、前記複数の反応体積の各反応体積の体積サイズの制御を含む、請求項５５に記載のシステム。
【請求項７０】
前記均一な細胞の集団が同じ遺伝物質を含む、請求項５５に記載のシステム。
【請求項７１】
前記均一な細胞の集団が、同じ配列を有する同じ目的の遺伝子を含む、請求項５５に記載のシステム。
【請求項７２】
改善された細胞を生成するための方法であって、
（ａ）突然変異誘発細胞のライブラリを生成することと、
（ｂ）前記突然変異誘発細胞のライブラリを複数のパーティションに分配して、分配された変異誘発細胞を得ることと、
（ｃ）前記分配された突然変異誘発細胞を表現型についてスクリーニングすることと、
（ｄ）前記スクリーニングに基づいて、閾値ゲートを使用して、前記分配された突然変異誘発細胞から改善された細胞を選択することと、
（ｅ）（ｂ）のフィードバックコントローラ（ｉ）を使用して前記複数のパーティションの生成頻度を維持すること、または（ｄ）のフィードバックコントローラ（ｉｉ）を使用して前記閾値ゲートを調節することと、
（ｆ）前記改善された細胞を使用して工程（ａ）～（ｄ）を繰り返すことと、を含む、方法。
【請求項７３】
前記方法が、前記フィードバックコントローラを使用して前記複数のパーティションの生成頻度を維持することを含む、請求項７２に記載の方法。
【請求項７４】
前記複数のパーティションが複数の液滴を含む、請求項７３に記載の方法。
【請求項７５】
前記フィードバックコントローラが、流量を調整して前記複数の液滴の生成頻度を維持する、請求項７４に記載の方法。
【請求項７６】
前記方法が、前記フィードバックコントローラを使用して前記閾値ゲートを調節することを含む、請求項７２に記載の方法。
【請求項７７】
前記フィードバックコントローラが、前記閾値ゲートを上昇または下降させる、請求項７６に記載の方法。
【請求項７８】

前記フィードバックコントローラが、前記閾値ゲートを調節して、前記分配された突然変異誘発細胞の標的パーセンテージを選択する、請求項７６に記載の方法。
【請求項７９】
コンピュータプロセッサによって実行されると、前記コンピュータプロセッサに請求項７２に記載の工程（ｂ）～（ｅ）を実行させる命令を含む、コンピュータ可読媒体。
【請求項８０】
コンピュータプロセッサと、請求項７９に記載のコンピュータ可読媒体と含むシステム。
【発明の詳細な説明】
【技術分野】
【０００１】
相互参照
本出願は、２０２０年１１月１３日に出願された米国仮出願第６３／１１３，５７１号に基づく利益を主張し、その全体が参照により本明細書に組み込まれる。
【背景技術】
【０００２】
背景
指向性進化は、種のライブラリをスクリーニングし、反復プロセスにおいて所望の機能性を選択するのに有用なツールである。そのようなプロセスは、機能的特性について多数の分子または株（例えば、遺伝子バリアント）のスクリーニングを可能にし、機能的目標に向けた分子（例えば、タンパク質）の操作を可能にする。
【０００３】
しかしながら、分子および／または株の発見および最適化のための現在のプロセスは、スループットが限られており、および／または結果を予測するために、または指向性進化プロセスを推進するために、様々なソースからの複数のデータ入力を利用しない。さらに、従来のハイスループットスクリーニング技術に関連する高いコストおよび長い時間は、典型的には、生成物を最適化することができる条件の多様性を制限する。その結果、分子および／または株の発見および最適化は、遅く、費用がかかり、最適ではない可能性がある。統合されたマルチパラメトリックでデータ駆動型のプロセスの欠如は、分子または株の発見をさらに妨げるか、または発見のために複数の入力を必要とする分子または株を発見できない可能性がある。
【発明の概要】
【０００４】
要旨
本明細書では、分子発見、分子最適化、および種または株の最適化（例えば、誘導された突然変異、指向性進化などを使用する）のためのモジュール式のハイスループットプラットフォームの必要性が認識されている。本開示は、分子（例えば、タンパク質などの生体分子）の発見、最適化、および機能的予測を推進するための同時機械学習プラットフォームへの大規模データ入力（オープンデータおよび公的ソースデータを含む）、経験的データ、および既知の条件を組み合わせて使用するためのシステムおよび方法を提供することによって、上記のニーズに対処する。
【０００５】
ある態様では、改善されたポリマーバリアントを生成する方法であって、
（ａ）データセットを取得することであって、（ｉ）ポリマーバリアント配列のデータを含む配列データと、（ｉｉ）ポリマーバリアント機能に基づいて選別された試料中のポリマーバリアント配列の存在量に基づく濃縮データと、を含む、取得することと、（ｂ）１またはそれを超える機械学習アルゴリズムをデータセットに適用して、改善されたポリマーバリアントを設計することと、を含む、方法が本明細書で提供される。
【０００６】
いくつかの実施形態では、ポリマーバリアント配列は、ｉ）選別されていない試料中お

よびｉｉ）ポリマーバリアント機能に基づいて選別された試料中の配列を含み、ポリマーバリアント機能に基づいて選別された試料および選別されていない試料が共通の試料に由来する。

【０００７】

いくつかの実施形態では、濃縮データは、選別されていない試料中のポリマーバリアント配列の存在量にさらに基づく。いくつかの実施形態では、ポリマーバリアント機能に基づいて選別された試料中のポリマーバリアント配列の存在量は、シーケンスカウントを含む。いくつかの実施形態では、選別されていない試料中のポリマーバリアント配列の存在量は、シーケンスカウントを含む。いくつかの実施形態では、濃縮データは、ｉ）選別されていない試料中のポリマーバリアント配列中の第１のポリマーバリアント配列の存在量、およびｉｉ）ポリマーバリアント機能に基づいて選別された試料のサブセット中のポリマーバリアント配列中の第１のポリマーバリアント配列の存在量に基づいて、第１のポリマーバリアント配列の濃縮係数を含む。いくつかの実施形態では、濃縮係数は、第１のポリマーバリアント配列の機能的標識である。いくつかの実施形態では、方法は、（ａ）および（ｂ）の複数世代を実施することを含み、（ｂ）の改善されたポリマーバリアントを使用して、複数世代の後続世代において（ａ）のデータセットを生成する。いくつかの実施形態では、複数世代は、少なくとも３世代または４世代である。

【０００８】

いくつかの実施形態では、ポリマーバリアントはタンパク質を含む。いくつかの実施形態では、タンパク質は酵素を含む。いくつかの実施形態では、機能は酵素の活性を含む。いくつかの実施形態では、改善されたポリマーバリアントは、少なくとも１０００の改善されたポリマーバリアントを含む。いくつかの実施形態では、ポリマーバリアント配列は、ポリマーバリアントをコードする核酸配列を含む。

【０００９】

いくつかの実施形態では、方法は、データセットを取得する前に、ポリマーバリアントを含む細胞のライブラリを提供することをさらに含み、ポリマーバリアントは複数のポリマーバリアントを含む。いくつかの実施形態では、方法は、ポリマーバリアントを含む細胞のライブラリを生成することをさらに含む。いくつかの実施形態では、方法は、生成することは、各位置に２０個の天然に存在するアミノ酸のそれぞれを有するタンパク質をコードする部位飽和核酸分子ライブラリ（ＳＳＬ）を提供することを含む。いくつかの実施形態では、方法は、ポリマーバリアントをコードする核酸分子を生成するために、ＳＳＬに対して遺伝子シャッフリングおよびエラープローンポリメラーゼ連鎖反応（ｅｐＰＣＲ）を実施することをさらに含む。いくつかの実施形態では、方法は、ポリマーバリアントをコードする核酸分子を細胞に形質転換して、細胞のライブラリを生成することをさらに含む。いくつかの実施形態では、細胞は細菌細胞を含む。いくつかの実施形態では、方法は、複数のパーティションを生成するために細胞のライブラリを分配することをさらに含む。いくつかの実施形態では、複数のパーティションの各パーティションは、平均して、１つ以下の細胞を含む。いくつかの実施形態では、方法は、複数のパーティションにおいて細胞のライブラリを拡大することをさらに含む。いくつかの実施形態では、方法は、ポリマーバリアントを複数のパーティション内の細胞のライブラリにおいて発現させることをさらに含む。いくつかの実施形態では、方法は、ポリマーバリアントを含む細胞のライブラリを使用してポリマーバリアント機能を決定することをさらに含む。いくつかの実施形態では、ポリマーバリアントはタンパク質を含む。いくつかの実施形態では、タンパク質は酵素を含む。いくつかの実施形態では、ポリマーバリアント機能を決定することは、酵素基質および反応緩衝液を使用することを含む。いくつかの実施形態では、ポリマーバリアント機能は、酵素活性を含む。いくつかの実施形態では、酵素活性は蛍光生成物を生成する。

【００１０】

いくつかの実施形態では、方法は、ポリマーバリアントを含む細胞ライブラリのサブセットをポリマーバリアント機能に基づいて選別して、ポリマーバリアント機能に基づいて

選別された試料および選別されていない試料を提供することをさらに含む。いくつかの実施形態では、機能が酵素活性の量である。いくつかの実施形態では、選別することは、細胞のライブラリのサブセットを、機能に基づいて複数のグループに選別することを含む。いくつかの実施形態では、複数のグループは、少なくとも４つのグループである。いくつかの実施形態では、細胞のライブラリのサブセットは、細胞のライブラリに存在すると予想される数の約５０倍～約１００倍のユニークなポリマーバリアントを含む。いくつかの実施形態では、方法は、ポリマーバリアント機能に基づいて選別された試料を使用してポリマーバリアント配列を決定することをさらに含む。いくつかの実施形態では、ポリマーバリアント機能に基づいて選別された試料が基準配列を含む。いくつかの実施形態では、ポリマーバリアント機能に基づいて選別された試料を使用してポリマーバリアント配列を決定することが、ポリマーバリアント機能に基づいて選別された試料から核酸分子を精製することを含む。いくつかの実施形態では、方法は、選別されていない試料から核酸分子を精製することをさらに含む。いくつかの実施形態では、ポリマーバリアント機能に基づいて選別された試料から核酸分子を精製することは、選別されていない試料から核酸分子を精製することとは別個に生じる。いくつかの実施形態では、方法は、ポリマーバリアント機能に基づいて選別された試料からの核酸分子および選別されていない試料からの核酸分子をシーケンシングして、ポリマーバリアント配列を提供することをさらに含む。いくつかの実施形態では、シーケンシングすることは、次世代シーケンシングすることを含む。いくつかの実施形態では、方法は、ポリマーバリアント配列を使用して、ｉ）ポリマーバリアント機能に基づいて選別された試料中およびｉｉ）選別されていない試料中のポリマーバリアント配列の存在量を決定することをさらに含む。いくつかの実施形態では、方法は、ｉ）ポリマーバリアント機能に基づいて選別された試料中およびｉｉ）選別されていない試料中のポリマーバリアント配列の存在量を使用して、濃縮データを決定することをさらに含む。いくつかの実施形態では、濃縮データは、濃縮係数を含む。いくつかの実施形態では、方法は、改善されたポリマーバリアントを合成することをさらに含む。いくつかの実施形態では、合成することは、オリゴヌクレオチドアレイ上での合成を含む。いくつかの実施形態では、方法は、改善されたポリマーバリアントをベクターにアセンブリすることをさらに含む。いくつかの実施形態では、ベクターはプラスミドである。いくつかの実施形態では、方法は、ベクターを宿主細胞に導入することをさらに含む。

【００１１】

いくつかの実施形態では、複数世代のうちの第１世代の（ａ）が、飽和突然変異誘発およびｅｐＰＣＲを実施して細胞のライブラリを提供することを含み、複数世代のうちの第２世代の（ａ）が、ｅｐＰＣＲを実施することを含み、複数世代のうちの第３世代の（ａ）が、ｅｐＰＣＲを実施することを含む。いくつかの実施形態では、複数世代の各世代は、複数回の選別を含む。

【００１２】

別の態様では、コンピュータプロセッサによって実行されるとコンピュータプロセッサに本明細書に記載の任意の実施形態に記載の方法を実行させる命令を含む、非一時的なコンピュータ可読媒体が本明細書に提供される。

【００１３】

別の態様では、コンピュータプロセッサと、非一時的なコンピュータ可読媒体とを含むシステムが本明細書に開示される。

【００１４】

さらに別の態様では、１またはそれを超えるプロセッサとコンピュータ可読媒体とを含む、細胞を処理するためのシステムであって、コンピュータ可読媒体が、１またはそれを超えるプロセッサによって実行されるとシステムに、（ａ）細胞の集団を複数の反応体積に分配して、分配された細胞を得ることと、（ｂ）分配された細胞を表現型についてアッセイして、分配された細胞に対するシグナルを生成することを実行させる命令を含み、分配された細胞に対するシグナルの変動係数（ＣＶ）が、細胞の集団が均一な細胞の集団である場合、１１％またはそれ未満である、システムが本明細書に開示される。

【００１５】
いくつかの実施形態では、ＣＶは１０％またはそれ未満である。いくつかの実施形態では、ＣＶは７％またはそれ未満である。いくつかの実施形態では、ＣＶは約５％～約１０％である。
【００１６】
いくつかの実施形態では、均一な細胞の集団は、少なくとも１００万個の細胞を含む。いくつかの実施形態では、均一な細胞の集団は、少なくとも１０，０００，０００個の細胞を含む。
【００１７】
いくつかの実施形態では、反応体積は、液滴を含む。
【００１８】
いくつかの実施形態では、アッセイは、分配された細胞への検出試薬の添加を含む。いくつかの実施形態では、アッセイは、分配された細胞への検出試薬の添加のタイミングの制御を含む。いくつかの実施形態では、検出試薬が酵素基質を含む。
【００１９】
いくつかの実施形態では、アッセイは、分配された細胞に対するシグナルの検出を含む。いくつかの実施形態では、検出は、光学的検出を含む。
【００２０】
いくつかの実施形態では、アッセイは、分配された細胞に対する検出シグナルのタイミングの制御を含む。
【００２１】
いくつかの実施形態では、表現型が酵素活性を含む。
【００２２】
いくつかの実施形態では、（ａ）は、複数の反応体積の各反応体積の体積サイズの制御を含む。
【００２３】
いくつかの実施形態では、均一な細胞の集団は、同じ遺伝物質を含む。
【００２４】
いくつかの実施形態では、均一な細胞の集団は、同じ配列を有する同じ目的の遺伝子を含む。
【００２５】
別の態様では、改善された細胞を生成するための方法であって、（ａ）突然変異誘発細胞のライブラリを生成することと、（ｂ）突然変異誘発細胞のライブラリを複数のパーティションに分配して、分配された変異誘発細胞を得ることと、（ｃ）分配された突然変異誘発細胞を表現型についてスクリーニングすることと、（ｄ）スクリーニングに基づいて、閾値ゲートを使用して、分配された突然変異誘発細胞から改善された細胞を選択することと、（ｅ）（ｂ）のフィードバックコントローラ（ｉ）を使用して複数のパーティションの生成頻度を維持すること、または（ｄ）のフィードバックコントローラ（ｉｉ）を使用して閾値ゲートを調節することと、（ｆ）改善された細胞を使用して工程（ａ）～（ｄ）を繰り返すことと、を含む、方法が本明細書に提供される。
【００２６】
いくつかの実施形態では、方法は、フィードバックコントローラを使用して複数のパーティションの生成頻度を維持することを含む。いくつかの実施形態では、複数のパーティションが複数の液滴を含む。
【００２７】
いくつかの実施形態では、フィードバックコントローラが、流量を調整して複数の液滴の生成頻度を維持する。いくつかの実施形態では、方法は、フィードバックコントローラを使用して閾値ゲートを調節することを含む。いくつかの実施形態では、フィードバックコントローラは、閾値ゲートを上昇または下降させる。フィードバックコントローラが、閾値ゲートを調節して、分配された突然変異誘発細胞の標的パーセンテージを選択する、

請求項７６に記載の方法。いくつかの実施形態では、フィードバックコントローラは、比例・積分・微分（ＰＩＤ）アルゴリズムである。いくつかの実施形態では、フィードバックコントローラは、機械的機器、液体粘度、または温度の偏差を考慮する（accounts for）。

【００２８】

いくつかの実施形態では、突然変異誘発細胞のライブラリを生成することは、細胞を変異原と接触させることを含む。いくつかの実施形態では、突然変異誘発細胞のライブラリを生成することは、トランスポゾン突然変異誘発、多重自動ゲノム工学法（ｍａｇｅ）、ゲノムシャッフリング、ランダム組換え、または非相同末端結合を含む。いくつかの実施形態では、突然変異誘発細胞のライブラリを生成することは、異種遺伝要素を細胞に導入することを含む。いくつかの実施形態では、異種遺伝要素は、プロモーター、５’非翻訳領域、リボザイム、ＲＮＡ安定性配列、分泌ペプチドをコードする核酸、シグナルペプチドをコードする核酸、融合タンパク質をコードする核酸、ＤＮＡ結合ドメインをコードする核酸、タンパク質－タンパク質結合領域をコードする核酸、コドン最適化遺伝子をコードする核酸、３’非翻訳領域、またはターミネーターを含む。

【００２９】

いくつかの実施形態では、突然変異誘発細胞は、少なくとも１０００個の細胞、少なくとも１０，０００個の細胞、または少なくとも１，０００，０００個の細胞を含む。

【００３０】

いくつかの実施形態では、本方法は、（ｂ）の前に突然変異誘発細胞のライブラリを拡大増殖することをさらに含む。いくつかの実施形態では、拡大増殖は液体培養で起こる。

【００３１】

いくつかの実施形態では、分配はコンピュータ可読媒体によって制御される。いくつかの実施形態では、複数のパーティションの各パーティションは、所定の範囲内の液体の体積を含む。いくつかの実施形態では、本方法は、（ｃ）の前に複数のパーティションの各パーティションを画像化することをさらに含む。いくつかの実施形態では、本方法は、画像化に基づいて、突然変異誘発細胞の所望の数の突然変異誘発細胞を含む複数のパーティションの分配を確実にすることをさらに含む。

【００３２】

いくつかの実施形態では、本方法は、複数のパーティションを（ｂ）と（ｃ）との間の細胞増殖条件に供することをさらに含む。

【００３３】

いくつかの実施形態では、上記供することは、複数のパーティションのサブセットのパーティションの温度、炭素源、ｐＨ、培養液成分、または通気を制御することを含む。

【００３４】

いくつかの実施形態では、本方法は、画像化を使用して細胞増殖条件に供された複数のパーティションにおける細胞増殖を監視することをさらに含む。

【００３５】

いくつかの実施形態では、スクリーニングは、複数のパーティションの各パーティションに検出試薬を添加することを含む。いくつかの実施形態では、検出試薬は、複数のパーティションの各パーティションの環境を変化させる。いくつかの実施形態では、検出試薬は、ｐＨ調整剤、イオン強度調整剤、イオン、基質、阻害剤および酵素、または競合剤を含む。

【００３６】

いくつかの実施形態では、突然変異誘発細胞のライブラリは、目的の分子をコードする。いくつかの実施形態では、スクリーニングは、目的の分子を検出すること、目的の分子の活性を測定することであって、必要に応じて活性が触媒活性であり、必要に応じて触媒活性が蛍光生成物を生成する、測定すること、目的の分子の量を決定すること、または目的の分子の機能を決定することを含む。

【００３７】

いくつかの実施形態では、スクリーニングは、複数のスクリーニングを実施することを含む。いくつかの実施形態では、複数のスクリーニングは、複数のスクリーニングの前のスクリーニングと比較して、複数のスクリーニングの後続のスクリーニングにおいて、目的の表現型に対してよりストリンジェントな閾値を適用することを含む。

【００３８】

いくつかの実施形態では、パーティションを選択することは、分光測定を実行することを含む。

【００３９】

いくつかの実施形態では、パーティションを選択することは、閾値を適用することを含む。

【００４０】

いくつかの実施形態では、上記（ｆ）は、選択されたパーティションから突然変異誘発細胞を除去して、改善された細胞を得ることを含む。いくつかの実施形態では、本方法は、突然変異誘発細胞の生存を維持することをさらに含む。いくつかの実施形態では、本方法は、突然変異誘発細胞を増殖させることをさらに含む。

【００４１】

いくつかの実施形態では、本方法は、目的の表現型について改善された細胞をスクリーニングすることをさらに含む。

【００４２】

いくつかの実施形態では、突然変異誘発細胞は細菌を含む。

【００４３】

別の態様では、コンピュータプロセッサによって実行されると、コンピュータプロセッサに工程（ｂ）～（ｅ）を実行させる命令を含む、コンピュータ可読媒体も本明細書で提供される。

【００４４】

別の態様では、コンピュータプロセッサとコンピュータ可読媒体とを含むシステムが本明細書で開示される。

【００４５】

別の態様では、株最適化ユニット、分子最適化ユニットおよび分子発見ユニットからなる群から選択される２またはそれを超える統合ユニットを含む所望の生物学的バリアントを選別するための統合システムが本明細書に開示され、ここで、（ａ）株最適化ユニットは、それぞれが異なる株最適化遺伝的バリアントまたは発酵条件を有する複数の別個の株最適化反応容器を含み、（ｂ）分子最適化ユニットは、それぞれが異なる標的タンパク質バリアントまたは適用条件を有する複数の別個の分子最適化反応容器を含み、（ｃ）分子発見ユニットは、それぞれが異なる発見分子または適用条件を有する複数の別個の分子発見反応容器を含む。

【００４６】

いくつかの実施形態では、２またはそれを超えるユニットは、以下の機能である（ｉ）データリポジトリ内の２またはそれを超える統合ユニットのそれぞれからのデータセットの記憶、（ｉｉ）データリポジトリに配信されるデータの閾値パラメータの設定、（ｉｉｉ）入力された多様性スクリーニングを繰り返すための２またはそれを超えるユニットのそれぞれの指示のうちの１またはそれを超える機能の実行ができるコンピュータ可読媒体を介して統合される。

【００４７】

いくつかの実施形態では、２またはそれを超えるユニットは、選別された所望の生物学的バリアントについて２またはそれを超えるユニットのそれぞれからのデータを統合する機械学習アルゴリズムを含むコンピュータ可読媒体を介して統合される。いくつかの実施形態では、機械学習アルゴリズムは、エラスティックネット回帰一般化線形モデル（ＧＬＭＮＥＴ）、サポートベクターマシン回帰（ＳＶＭ）、ランダムフォレスト（ＲＦ）、エクストリーム勾配ブースティング（ＸＧＢｏｏｓｔ）、多層パーセプトロン（ＭＬＰ）、

または畳み込みニューラルネットワーク（ＣＮＮ）を含む。
【００４８】
いくつかの実施形態では、株最適化反応容器、分子最適化反応容器、および／または分子発見反応容器のそれぞれは、平均で１個以下の細胞を含む。
【００４９】
いくつかの実施形態では、株最適化反応容器、分子最適化反応容器、および／または分子発見反応容器のそれぞれは、平均で１個を超える細胞を含む。いくつかの実施形態では、ウェルあたり１個を超える細胞の平均がクローン集団である。いくつかの実施形態では、ウェルあたり１個を超える細胞の平均は、多様な細胞集団を含む。
【００５０】
いくつかの実施形態では、株最適化反応容器、分子最適化反応容器、および／または分子発見反応容器のそれぞれは、平均で１０～１００個の細胞を含む。いくつかの実施形態では、株最適化ユニット、分子最適化ユニット、および分子発見ユニットは、単一細胞分析を実行するように構成される。いくつかの実施形態では、株最適化ユニット、分子最適化ユニット、または分子発見ユニットは、多細胞解析を実行するように構成される。
【００５１】
いくつかの実施形態では、株最適化ユニット、分子最適化ユニットおよび／または分子発見ユニットは、少なくとも毎分１００個の細胞をスクリーニングするように構成される。
【００５２】
いくつかの実施形態では、２またはそれを超える統合ユニットのそれぞれは、消費ユニットを含む。いくつかの実施形態では、消費ユニットはマイクロ流体デバイスを含む。いくつかの実施形態では、マイクロ流体デバイスは、液滴発生器、液滴凝縮器、緩衝液リザーバ、細胞インキュベータ、流体注入器、液滴スプリッター、液滴選別機、ディレイライン、および液滴シンギュレータを含む群から選択される１またはそれを超える機能要素を含む。
【００５３】
いくつかの実施形態では、株最適化反応容器、分子最適化反応容器、または分子発見反応容器の反応容器は、１マイクロリットル未満の体積を有する。いくつかの実施形態では、株最適化反応容器、分子最適化反応容器、および／または分子発見反応容器の反応容器は、ナノウェルアレイ内にある。いくつかの実施形態では、株最適化反応容器、分子最適化反応容器、および／または分子発見反応容器の反応容器は、液滴、プラグ、またはウェルである。
【００５４】
いくつかの実施形態では、統合システムは、１またはそれを超える当該ユニットに結合された検出器をさらに含み、検出器は、質量分析計、ラマン分光計、顕微鏡、セルカウンタ、蛍光顕微鏡、光学顕微鏡、フローサイトメータ、質量分析計、蛍光プレートリーダ、近赤外（ＮＩＲ）分光光度計、および圧電センサからなる群から選択される。
【００５５】
いくつかの実施形態では、統合システムは、ロボットマニピュレータ、フロージェネレータ、音響ドロップジェネレータ、光ピンセット、熱ドロップオンデマンド、圧電ドロップオンデマンドからなる群から選択される反応容器移動アクチュエータをさらに含む。
【００５６】
いくつかの実施形態では、異なる株最適化遺伝的バリアントのそれぞれ、標的タンパク質バリアントのそれぞれは異なる細胞内にあり、および／または標的発見分子のそれぞれは異なる細胞内にある。
【００５７】
いくつかの実施形態では、各株最適化遺伝的バリアントは、ランダム突然変異誘発（例えば、紫外線、Ｘ線照射、ガンマ線照射などの突然変異原、化学変異原、例えばエチルメチルスルホネート、活性酸素種、脱アミノ化剤、多環芳香族炭化水素、アルキル化剤（例

えば、エチルニトロソウレア、ニトロソアミン）芳香族アミン、アルカロイド、臭素、アジ化ナトリウム、ソラレン、ベンゼン、金属（例えば、ヒ素、カドミウム、クロム、ニッケル）または生物学的薬剤（例えば、トランスポゾン、ウイルス、細菌）などへの曝露）、多重自動ゲノム工学法（ＭＡＧＥ）、ＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、またはＣＲＩＳＰＲガイドＤＮＡポリメラーゼによって生成される。いくつかの実施形態では、ランダム突然変異誘発は、飽和突然変異誘発である。

【００５８】

いくつかの実施形態では、各標的タンパク質バリアントは、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、ＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、またはＣＲＩＳＰＲガイドＤＮＡポリメラーゼによって生成される。いくつかの実施形態では、ランダム突然変異誘発は、飽和突然変異誘発または循環置換である。

【００５９】

いくつかの実施形態では、各発見分子は、メタゲノムライブラリ、ＤＮＡコード化ライブラリ、１またはそれを超える微生物からのＤＮＡのライブラリ、またはファンクションジェネレータで生成されたライブラリに見られるＤＮＡコード化分子である。

【００６０】

いくつかの実施形態では、株最適化ユニットは、第１の株の複数の異なる遺伝的バリアントを含む株最適化ライブラリを含む。いくつかの実施形態では、複数の異なる遺伝的バリアントは、代謝経路によって分類される。いくつかの実施形態では、株最適化ライブラリは、表現型、配列および代謝経路データを含む。いくつかの実施形態では、株最適化ユニットは機械学習アルゴリズムを含む。いくつかの実施形態では、機械学習アルゴリズムは、複数の発酵条件にわたって最適化された遺伝的バリアントを同定する。

【００６１】

いくつかの実施形態では、２またはそれを超えるユニットのそれぞれは、生物学的多様性または条件付き多様性の追加の入力を含む再帰的サイクルを含む。

【００６２】

いくつかの実施形態では、複数の別々の反応体積は、少なくとも10^5個の別個の反応体積を含む。

【００６３】

いくつかの実施形態では、適用条件は、ｐＨ、基質、反応緩衝液、および温度のうちの１またはそれを超えるものを含む。

【００６４】

いくつかの実施形態では、発酵条件は、ｐＨ、酸素化、炭素源、緩衝液濃度、および温度のうちの１またはそれを超えるものを含む。いくつかの実施形態では、株最適化ライブラリは、第１の株の少なくとも１００個の異なるバリアントを含む。いくつかの実施形態では、株最適化ライブラリは、株のランダム突然変異誘発（例えば、紫外線、Ｘ線照射、ガンマ線照射などの突然変異原、化学変異原、例えばエチルメチルスルホネート、活性酸素種、脱アミノ化剤、多環芳香族炭化水素、アルキル化剤（例えば、エチルニトロソウレア、ニトロソアミン）、芳香族アミン、アルカロイド、臭素、アジ化ナトリウム、ソラレン、ベンゼン、金属（例えば、ヒ素、カドミウム、クロム、ニッケル）または生物学的薬剤（例えば、トランスポゾン、ウイルス、細菌）などへの曝露）によって生成される。いくつかの実施形態では、株最適化ライブラリは、全ゲノム突然変異誘発ライブラリ、ゲノムシャッフルライブラリ、標的ゲノムライブラリおよびトランスポゾンライブラリからなる群から選択される。いくつかの実施形態では、全ゲノム突然変異誘発ライブラリは、全ゲノムランダム突然変異誘発ライブラリである。いくつかの実施形態では、標的ゲノムライブラリは、プロモータースワップライブラリである。いくつかの実施形態では、標的ゲ

ノムライブラリは、プロモーター、５’非翻訳領域、３’非翻訳領域、リボザイム、ＲＮＡ安定性配列、分泌ペプチド、融合タンパク質、ＤＮＡ結合ドメイン、タンパク質－タンパク質結合領域、コドン最適化遺伝子、コドンランダム化遺伝子、ターミネーターおよび非相同末端結合からなる群から選択される１またはそれを超えるメンバーを含むか、またはそれらをコードする。いくつかの実施形態では、トランスポゾンライブラリは、プロモーター挿入ライブラリである。

【００６５】

いくつかの実施形態では、分子最適化ユニットは、第１の分子の複数の異なる遺伝的バリアントを含む分子最適化ライブラリを含む。いくつかの実施形態では、分子最適化ライブラリは、単一遺伝子ランダム突然変異ライブラリ、部位飽和ライブラリ、小型インサートメタゲノムライブラリ、および大型インサートメタゲノムライブラリからなる群から選択される。

【００６６】

別の態様では、所望の生物学的バリアントを選別するための統合システムであって、生物学的多様性データストリームと条件付き多様性データストリームとを含み、（ａ）データストリームのそれぞれは独立してプログラム可能であり、（ｂ）生物学的多様性データストリームは標的タンパク質または標的タンパク質を発現する株の異なるゲノムバリアントのスクリーニングからのデータを含み、（ｃ）条件付き多様性データストリームは株と標的タンパク質の異なる条件のスクリーニングからのデータを含み、（ｄ）生物学的多様性データストリームおよび条件付き多様性データストリームは標的パラメータを受け取るように構成された最適化ユニットに統合され、（ｉ）標的パラメータと一致する標的タンパク質または標的株が同定されるまで、追加の生物学的多様性または条件付き多様性を導入するように最適化ユニットに指示するか、またはｉｉ）標的パラメータと一致する標的タンパク質または標的株を同定する、統合システムが本明細書に開示される。

【００６７】

いくつかの実施形態では、最適化ユニットは、（ｉ）生物学的多様性データストリームおよび条件付き多様性データストリームのそれぞれからのデータセットをデータリポジトリに記憶する機能、（ｉｉ）データリポジトリに配信されるデータの閾値パラメータを設定する機能、（ｉｉｉ）入力された多様性スクリーニングを繰り返すように最適化ユニットに指示する機能のうちの１またはそれを超える機能を実行することができるコンピュータ可読媒体を含む。

【００６８】

いくつかの実施形態では、最適化ユニットは、生物学的バリアントの所望の選別のために生物学的多様性データストリームおよび条件付き多様性データストリームを統合する機械学習アルゴリズムを含むコンピュータ可読媒体を含む。いくつかの実施形態では、機械学習アルゴリズムは、エラスティックネット回帰一般化線形モデル（ＧＬＭＮＥＴ）、サポートベクターマシン回帰（ＳＶＭ）、ランダムフォレスト（ＲＦ）、エクストリーム勾配ブースティング（ＸＧＢｏｏｓｔ）、多層パーセプトロン（ＭＬＰ）、または畳み込みニューラルネットワーク（ＣＮＮ）を実行する。

【００６９】

いくつかの実施形態では、株最適化ユニットは、標的タンパク質または標的タンパク質を発現する株の１００個の異なるゲノムバリアントを毎分スクリーニングするように構成される。

【００７０】

いくつかの実施形態では、株最適化ユニットは、毎分１００個の細胞をスクリーニングするように構成される。

【００７１】

いくつかの実施形態では、株最適化ユニットのそれぞれは消費ユニットを備える。いくつかの実施形態では、消費ユニットはマイクロ流体デバイスを含む。いくつかの実施形態では、マイクロ流体デバイスは、液滴発生器、液滴凝縮器、緩衝液リザーバ、細胞インキ

ュベータ、流体注入器、液滴スプリッター、液滴選別機、ディレイライン、および液滴シンギュレータを含む群から選択される１またはそれを超える機能要素を含む。いくつかの実施形態では、マイクロ流体デバイスは１またはそれを超える反応容器を含む。いくつかの実施形態では、反応容器は、１マイクロリットル未満の体積を有する。いくつかの実施形態では、反応容器はナノウェルアレイ内にある。いくつかの実施形態では、反応容器は、液滴、プラグまたはウェルである。いくつかの実施形態では、液滴は油中水型液滴である。いくつかの実施形態では、液滴は表面上の水滴である。いくつかの実施形態では、反応容器は、２またはそれを超える反応容器ごとに平均約１個の細胞が存在するような細胞を含む。

【００７２】

いくつかの実施形態では、統合システムは、当該最適化ユニットに結合された検出器をさらに含み、検出器は、質量分析計、ラマン分光計、顕微鏡、セルカウンタ、蛍光顕微鏡、光学顕微鏡、フローサイトメータ、質量分析計、蛍光プレートリーダ、近赤外（ＮＩＲ）分光光度計、および圧電センサからなる群から選択される。

【００７３】

いくつかの実施形態では、統合システムは、ロボットマニピュレータ、フロージェネレータ、音響ドロップジェネレータ、光ピンセット、熱ドロップオンデマンド、圧電ドロップオンデマンドからなる群から選択される反応容器移動アクチュエータをさらに含む。

【００７４】

いくつかの実施形態では、標的タンパク質の異なるゲノムバリアントのそれぞれは、異なる細胞内にある。

【００７５】

いくつかの実施形態では、株の異なるゲノムバリアントのそれぞれは、ランダム突然変異誘発（例えば、紫外線、Ｘ線照射、ガンマ線照射などの突然変異原、化学変異原、例えばエチルメチルスルホネート、活性酸素種、脱アミノ化剤、多環芳香族炭化水素、アルキル化剤（例えば、エチルニトロソウレア、ニトロソアミン）芳香族アミン、アルカロイド、臭素、アジ化ナトリウム、ソラレン、ベンゼン、金属（例えば、ヒ素、カドミウム、クロム、ニッケル）または生物学的薬剤（例えば、トランスポゾン、ウイルス、細菌）などへの曝露）、多重自動ゲノム工学法（ＭＡＧＥ）、またはＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、またはＣＲＩＳＰＲガイドＤＮＡポリメラーゼによって生成される。いくつかの実施形態では、ランダム突然変異誘発は、飽和突然変異誘発または循環置換である。

【００７６】

いくつかの実施形態では、標的タンパク質バリアントの異なるゲノムバリアントのそれぞれは、ランダム突然変異誘発（例えば、紫外線、Ｘ線照射、ガンマ線照射などの突然変異原、化学変異原、例えばエチルメチルスルホネート、活性酸素種、脱アミノ化剤、多環芳香族炭化水素、アルキル化剤（例えば、エチルニトロソウレア、ニトロソアミン）芳香族アミン、アルカロイド、臭素、アジ化ナトリウム、ソラレン、ベンゼン、金属（例えば、ヒ素、カドミウム、クロム、ニッケル）または生物学的薬剤（例えば、トランスポゾン、ウイルス、細菌）などへの曝露）、多重自動ゲノム工学法（ＭＡＧＥ）、またはＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）によって生成される。いくつかの実施形態では、ランダム突然変異誘発は、飽和突然変異誘発、循環置換である。

【００７７】

別の態様では、標的タンパク質の遺伝要素および株の遺伝要素を含む生物学的に多様な遺伝要素の多様性スクリーニングからのデータを、標的タンパク質発現条件および標的タンパク質性能条件を含む多様な条件の条件付きスクリーニングからのデータと組み合わせることを含む、最終的な望ましい遺伝的バリアントを選別するための方法、組み合わせデータセットを使用して、最終的な望ましい遺伝的バリアントを同定するための方法が本明細書で提供される。

【００７８】
いくつかの実施形態では、条件付きスクリーニングは、多様性スクリーニングの一部である１またはそれを超える遺伝要素に対して行われる。
【００７９】
いくつかの実施形態では、各細胞が第１の遺伝要素のライブラリからの異なる遺伝要素を含む第１の細胞のライブラリを生成すること。いくつかの実施形態では、本方法は、各細胞が第２の遺伝要素のライブラリからの異なる遺伝要素を含む第２の細胞のライブラリを生成することをさらに含み、第１の遺伝要素のライブラリおよび第２の遺伝要素のライブラリは、複数の条件下で独立してスクリーニングされる。いくつかの実施形態では、第１の遺伝要素のライブラリおよび第２の遺伝要素のライブラリは、次世代シーケンシング（ＮＧＳ）、例えば、Ｉｌｌｕｍｉｎａ（登録商標）、Ｐａｃｉｆｉｃ Ｂｉｏｓｃｉｅｎｃｅｓ（ＰａｃＢｉｏ（登録商標））、Ｏｘｆｏｒｄ Ｎａｎｏｐｏｒｅ（登録商標）またはＬｉｆｅ Ｔｅｃｈｎｏｌｏｇｉｅｓ（Ｉｏｎ Ｔｏｒｒｅｎｔ（登録商標））によるシーケンシングシステムを使用してスクリーニングされ、複数の条件下で、濃縮されている突然変異が同定される。
【００８０】
いくつかの実施形態では、多様性スクリーニングは、株最適化に関連するバリアントに対して実行される。いくつかの実施形態では、株最適化は、収量の増加、発酵条件の改善、生存率の増加、イオン濃度に対する耐性の増加をもたらす。いくつかの実施形態では、収量の増加は、タンパク質発現の増加および／または細胞増殖の増加である。
【００８１】
いくつかの実施形態では、最終的な望ましい遺伝的バリアントを同定するために組み合わせデータセットを使用することは、（ａ）望ましい特性を有する１またはそれを超える中間的な望ましい遺伝的バリアントを同定すること、（ｂ）望ましい特性を有する１またはそれを超える中間的な望ましい遺伝的バリアントの遺伝要素に基づいて組換えライブラリを得ること、（ｃ）多様な条件の条件付きスクリーニングにより組換えライブラリをスクリーニングすること、および（ｄ）最終的な望ましい遺伝的バリアントを同定すること、または（ａ）望ましい特性を有する１またはそれを超える中間的な望ましい遺伝的バリアントを同定することの再帰的サイクルをさらに含む。
【００８２】
いくつかの実施形態では、多様性スクリーニングは、標的タンパク質発見に関連するバリアントに対して実行される。いくつかの実施形態では、多様性スクリーニングは、標的タンパク質最適化に関連するバリアントに対して実行される。いくつかの実施形態では、生物学的に多様な遺伝要素のそれぞれが並行してスクリーニングされる。いくつかの実施形態では、生物学的に多様な遺伝要素のそれぞれが連続してスクリーニングされる。
【００８３】
いくつかの実施形態では、組み合わせデータセットは、蛍光発生アッセイ、比色アッセイ、およびレポーター細胞ベースのアッセイの１またはそれを超えるアッセイからのデータを含む。いくつかの実施形態では、蛍光発生アッセイは、直接アッセイ、間接アッセイ、結合アッセイ、またはインビボレポーターアッセイである。
【００８４】
いくつかの実施形態では、条件付きスクリーニングは、温度、培養液、酸素化、塩分、ｐＨ、混合、反応時間、金属イオン濃度、添加剤濃度、供給速度、またはそれらの組み合わせを調整することを含む。いくつかの実施形態では、条件付きスクリーニングは、第１の遺伝要素に対して２またはそれを超える異なる条件をタンデムで実行することを含む。いくつかの実施形態では、多様な条件のそれぞれが並行して実行される。いくつかの実施形態では、多様な条件のそれぞれが連続して実行される。
【００８５】
いくつかの実施形態では、本方法は、多様な遺伝要素のライブラリを生成することをさらに含む。いくつかの実施形態では、生成することは、ランダム突然変異誘発、部位特異

的突然変異誘発、部位飽和またはゲノムシャッフリングによって行われる。いくつかの実施形態では、本方法は、多様な遺伝要素のライブラリからの各遺伝要素を細胞または無細胞系に導入することをさらに含む。いくつかの実施形態では、本方法は、多様な遺伝要素のライブラリからの遺伝的多様性を有する細胞を、各反応容器が１個以下の細胞を有するように、異なる反応容器に分離することをさらに含む。いくつかの実施形態では、反応容器は、液滴、プラグまたはウェルである。

【００８６】

いくつかの実施形態では、所望の遺伝的バリアントは、株最適化、標的分子最適化、または環境ＤＮＡもしくはＲＮＡライブラリからの分子発見に関連する。

【００８７】

いくつかの実施形態では、生物学的多様性は、株最適化のものである。いくつかの実施形態では、最終的な望ましい遺伝的バリアントを同定することは、発現レベル、細胞生存率、細胞増殖、および二価カチオン濃度に対する感受性を含む株の１またはそれを超える特性を測定することを含む。

【００８８】

いくつかの実施形態では、本方法は、部位特異的突然変異誘発、ランダム突然変異誘発（例えば、紫外線、Ｘ線照射、ガンマ線照射などの突然変異原、化学変異原、例えばエチルメチルスルホネート、活性酸素種、脱アミノ化剤、多環芳香族炭化水素、アルキル化剤（例えば、エチルニトロソウレア、ニトロソアミン）芳香族アミン、アルカロイド、臭素、アジ化ナトリウム、ソラレン、ベンゼン、金属（例えば、ヒ素、カドミウム、クロム、ニッケル）または生物学的薬剤（例えば、トランスポゾン、ウイルス、細菌）などへの曝露）、多重自動ゲノム工学法（ＭＡＧＥ）、ＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、またはＣＲＩＳＰＲガイドＤＮＡポリメラーゼによって改変株のライブラリを生成することをさらに含む。いくつかの実施形態では、ランダム突然変異誘発は、飽和突然変異誘発である。

【００８９】

いくつかの実施形態では、生物学的多様性は、標的タンパク質最適化のものである。いくつかの実施形態では、望ましい遺伝的バリアントを同定することは、基質特異性、基質状態、二価カチオン濃度に対する感受性、温度耐性、触媒性能、プロテアーゼ分解に対する耐性、ｐＨ耐性、阻害剤に対する感受性、補因子優先性、酵素速度、半減期、支持体に固定化された場合の活性、またはガスに対する感受性を含む標的タンパク質の１またはそれを超える特性を測定することを含む。いくつかの実施形態では、基質状態は可溶性または表面吸着性である。いくつかの実施形態では、多様性スクリーニングは、標的タンパク質の１またはそれを超える特性に基づいて遺伝要素を濃縮することをさらに含む。

【００９０】

いくつかの実施形態では、標的タンパク質の遺伝要素は、遺伝子領域、配列モチーフ、およびタンパク質構造クラスを含む。いくつかの実施形態では、標的タンパク質をコードする遺伝子は、ＣＯＬ１Ａ１（ヒト１型コラーゲン）である。いくつかの実施形態では、標的タンパク質は、構造タンパク質、酵素、表面受容体タンパク質、ペプチドホルモン、免疫系成分、または生物活性ペプチドである。いくつかの実施形態では、標的タンパク質は産業用酵素である。いくつかの実施形態では、標的タンパク質は治療用タンパク質である。いくつかの実施形態では、ペプチドホルモンはインスリンまたはシグナル伝達因子である。いくつかの実施形態では、構造タンパク質はコラーゲンまたは食品ベースのタンパク質である。いくつかの実施形態では、酵素は、リパーゼまたはポリ（エチレンテレフタレート）ヒドロラーゼ（ＰＥＴａｓｅ）である。

【００９１】

いくつかの実施形態では、生物学的多様性は発見ライブラリのものである。いくつかの実施形態では、発見ライブラリは、環境ＤＮＡ（ｅＤＮＡ）ライブラリである。

【００９２】

いくつかの実施形態では、条件付き多様性は、１またはそれを超える発酵条件のものである。いくつかの実施形態では、１またはそれを超える発酵条件は、温度、培養液、酸素化、塩分、ｐＨ、混合、供給スケジュール、炭素源、またはそれらのいずれか１つもしくは組み合わせの変更を含む。いくつかの実施形態では、本方法は、１００ミリリットルを超える体積で望ましい遺伝的バリアントを産生するために産業的発酵条件を同定することをさらに含む。

【００９３】

いくつかの実施形態では、条件付き多様性は、１またはそれを超える適用条件のものである。いくつかの実施形態では、適用条件は、ｐＨ、温度、基質、反応緩衝液および支持体への付着を含む。

【００９４】

いくつかの実施形態では、本方法は、シーケンシングを含む望ましい遺伝的バリアントを同定することをさらに含む。いくつかの実施形態では、シーケンシングは、ＰＣＲを行うことを含む。いくつかの実施形態では、シーケンシングは、次世代シーケンシング（ＮＧＳ）を行うことを含む。いくつかの実施形態では、同定することは、望ましい遺伝的バリアントをコードするオープンリーディングフレームを同定することを含む。いくつかの実施形態では、本方法は、１またはそれを超える核酸分子を精製することをさらに含み得る。

【００９５】

本開示のさらに別の態様では、性能基準を満たす最終最適化タンパク質または最終最適化株を同定するための方法であって、（ａ）（ｉ）ａ）株の異なる遺伝的バリアント、ｂ）タンパク質の異なる遺伝的バリアント、またはｃ）別個のタンパク質をコードする異なる遺伝子のうちの２またはそれを超えるものから選択される複数の異なる遺伝要素のデータ、および（ｉｉ）複数の異なる遺伝要素のそれぞれについて、複数の異なる条件のデータを含むデータセットを得ること、（ｂ）性能基準を満たす最終的な最適化されたタンパク質または最終的な最適化された株を同定または予測するためにデータセットを統合すること、または（ａ）に導入する追加データを同定することを含む方法が本明細書で提供される。

【００９６】

いくつかの実施形態では、性能基準は、標的タンパク質の発現レベル、標的タンパク質の安定性、標的タンパク質の基質特異性、標的タンパク質活性、株の増殖、発酵条件に対する株の耐性、標的タンパク質のフォールディング、標的タンパク質対副産物比、または標的タンパク質修飾を含む。

【００９７】

いくつかの実施形態では、複数の条件のデータは、標的タンパク質の安定性、二価カチオン濃度に対する株の耐性、標的タンパク質の基質特異性、温度耐性、ｐＨ耐性、発現レベル、または株の生存率を含む。

【００９８】

いくつかの実施形態では、データセットは機械学習アルゴリズムに適用され、（ｂ）の追加データは機械学習アルゴリズムによって同定される。いくつかの実施形態では、追加データは、遺伝要素または追加条件のライブラリである。

【００９９】

いくつかの実施形態では、データセットは、蛍光発生アッセイまたは比色アッセイのうちの１またはそれを超えるものからのデータを含む。いくつかの実施形態では、蛍光発生アッセイは、直接アッセイ、間接アッセイ、結合アッセイ、またはインビボレポーターアッセイである。

【０１００】

いくつかの実施形態では、（ｂ）の追加データを同定するためにデータセットを統合することは、（ｉ）性能基準のサブセットを満たす目的の中間タンパク質または目的の中間株を同定すること、（ｉｉ）（ａ）に導入する追加データとして、目的の中間タンパク質

または目的の中間株に対して少なくとも７０％、８０％、または９０％の相同性を有する追加の複数のバリアントを同定すること、および（ｉｉｉ）追加の複数のバリアントのそれぞれについて、（ａ）に導入する複数の異なる条件の追加データを取得することをさらに含む。いくつかの実施形態では、本方法は、ａ）追加の複数のバリアントが１００個未満のメンバーを含む場合に個々に、またはｂ）追加の複数のバリアントが１００個超のメンバーを含む場合に集団レベルで、複数の異なる条件の追加データを得ることをさらに含む。

【０１０１】

いくつかの実施形態では、株の遺伝的バリアントは、株の遺伝的バリアントのライブラリに由来する。いくつかの実施形態では、株の遺伝的バリアントのライブラリは、全ゲノム突然変異誘発ライブラリ、ゲノムシャッフルライブラリ、標的ゲノムライブラリおよびトランスポゾンライブラリからなる群から選択される。いくつかの実施形態では、全ゲノム突然変異誘発ライブラリは、全ゲノムランダム突然変異誘発ライブラリである。いくつかの実施形態では、標的ゲノムライブラリは、プロモーター、５’非翻訳領域、３’非翻訳領域、リボザイム、ＲＮＡ安定性配列、分泌ペプチド、融合タンパク質、ＤＮＡ結合ドメイン、タンパク質－タンパク質結合領域、コドン最適化遺伝子、コドンランダム化遺伝子、ターミネーターおよび非相同末端結合からなる群から選択される１またはそれを超えるメンバーを含むか、またはそれらをコードする。いくつかの実施形態では、標的ゲノムライブラリは、プロモータースワップライブラリである。いくつかの実施形態では、トランスポゾンライブラリは、プロモータースワップライブラリである。いくつかの実施形態では、株の遺伝的バリアントのライブラリは、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、またはＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）によって生成される。

【０１０２】

いくつかの実施形態では、タンパク質の異なる遺伝的バリアントは、標的タンパク質をコードする遺伝的バリアントのライブラリである。いくつかの実施形態では、データセットは、異なるタンパク質をコードする遺伝的バリアントを含み、追加データは、残基をコードするバリアントの部分配列ライブラリを含む。いくつかの実施形態では、部分配列ライブラリは遺伝子シャッフリングによって生成される。いくつかの実施形態では、標的タンパク質をコードする遺伝的バリアントのライブラリは、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、またはＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）によって生成される。いくつかの実施形態では、標的タンパク質をコードする遺伝的バリアントのライブラリは、単一遺伝子ランダム突然変異ライブラリ、部位飽和ライブラリ、小型インサートメタゲノムライブラリ、および大型インサートメタゲノムライブラリからなる群から選択される。

【０１０３】

いくつかの実施形態では、異なるタンパク質をコードする別個の遺伝子は、環境ＤＮＡ（ｅＤＮＡ）ライブラリである。

【０１０４】

本開示の別の態様は、１またはそれを超えるコンピュータプロセッサによって実行されると、上記または本明細書の他の箇所の方法のいずれかを実施する、マシン実行可能コードを含む非一時的なコンピュータ可読媒体を提供する。

【０１０５】

本開示の別の態様は、１またはそれを超えるコンピュータプロセッサと、それに結合されたコンピュータメモリとを含むシステムを提供する。コンピュータメモリは、１またはそれを超えるコンピュータプロセッサによって実行されると、上記または本明細書の他の箇所の方法のいずれかを実施する、マシン実行可能コードを含む。

【０１０６】

本開示のさらなる態様および利点は、本開示の例示的な実施形態のみが示され説明される以下の詳細な説明から当業者には容易に明らかになるであろう。理解されるように、本

開示は、他の異なる実施形態が可能であり、そのいくつかの詳細は、すべて本開示から逸脱することなく、様々な明白な点で修正が可能である。したがって、図面および説明は、本質的に例示と見なされるべきであり、限定と見なされるべきではない。

【０１０７】

参照による組み込み

本明細書で言及されるすべての刊行物、特許、および特許出願は、あたかも各個々の刊行物、特許、または特許出願が参照により組み込まれることが具体的かつ個別に示されているのと同程度に、参照により本明細書に組み込まれる。参照により組み込まれる刊行物および特許または特許出願が本明細書に含まれる開示と矛盾する限り、本明細書は、そのような矛盾する材料に取って代わるおよび／または優先することを意図している。

【図面の簡単な説明】

【０１０８】

本発明の新規な特徴は、添付の特許請求の範囲に詳細に記載されている。本発明の特徴および利点のより良い理解は、本発明の原理が利用される例示的な実施形態を説明する以下の詳細な説明、および添付の図面（本明細書ではまた、「図（Ｆｉｇｕｒｅ）」および「図（ＦＩＧ．）」を参照することによって得られるであろう。

【０１０９】

【図１】図１は、望ましい生物学的バリアントを選別するための株最適化ユニット、分子最適化ユニット、および分子発見ユニットを含む例示的な統合システムを概略的に示す。

【０１１０】

【図２】図２は、複数のパラメータを入力し、最適化された株、最適化された分子、または所望の機能的特性を有する新しい分子に関する予測を出力するように構成された例示的な統合システムを概略的に示す。

【０１１１】

【図３】図３は、本明細書で提供される方法を実施するようにプログラムまたは他の方法で構成されたコンピュータシステムを示す。

【０１１２】

【図４】図４は、株最適化のための例示的なワークフローを示す。

【０１１３】

【図５】図５は、複数回の濃縮を含む株最適化のための例示的なワークフローを示す。

【０１１４】

【図６Ａ－６Ｃ】図６Ａ～図６Ｃは、株最適化から得られた例示的なデータを示す。図６Ａは、第１世代における複数回の濃縮にわたる細胞のライブラリの進化を示す。図６Ｂは、第２世代における複数回の濃縮にわたる細胞のライブラリの進化を示す。図６Ｃは、第３世代における複数回の濃縮にわたる細胞のライブラリの進化を示す。

【０１１５】

【図７】図７は、世代にわたる目的の分子の量のプロットを示す。

【０１１６】

【図８】図８は、世代内の複数回の濃縮にわたる株最適化から得られた例示的なデータを示す。

【０１１７】

【図９】図９は、複数回の濃縮にわたる株最適化から得られた例示的なデータを示す。

【０１１８】

【図１０】図１０は、分子発見のための例示的なワークフローを示す。

【０１１９】

【図１１】図１１は、世代内の複数回の濃縮にわたる分子発見から得られた例示的なデータを示す。

【０１２０】

【図１２】図１２は、分子発見から発見された突然変異のデータ例を示す。

【０１２１】

【図１３－１】図１３は、分子発見プロセスを使用することによって分子の機能的性能を改善する例示的なデータを示す。
【図１３－２】図１３は、分子発見プロセスを使用することによって分子の機能的性能を改善する例示的なデータを示す。
【０１２２】
【図１４】図１４は、分子最適化のための例示的なワークフローを示す。
【０１２３】
【図１５】図１５は、世代内の複数回の濃縮にわたる分子最適化から得られた例示的なデータを示す。
【０１２４】
【図１６】図１６は、世代間にわたる分子最適化プロセスから得られた例示的なデータを示す。
【０１２５】
【図１７】図１７は、本明細書に記載のシステムにおけるドリフトを防止または考慮するためのアルゴリズムの例示的なデータを示す。
【発明を実施するための形態】
【０１２６】
詳細な説明
本明細書では、ハイスループットプラットフォームにおいて所望の生物学的バリアントを選別および同定するための方法およびシステムが開示される。一般に、本明細書に記載のシステムおよび方法は、株最適化ユニット、分子最適化ユニット、および分子発見ユニットからなる群から選択される１またはそれを超える統合ユニットを使用することを含む。本明細書で使用される方法およびシステムは、統合ユニット（例えば、株最適化ユニット、分子最適化ユニット、および／または分子発見ユニット）のうちの１またはそれを超えるものに入力するか、またはそれから出力される大きなデータセットを利用して、そのような所望の特性を示す所望の特性、最適化された分子（例えば、所望の特性の向上を有する）、または株（例えば、天然に存在するまたは操作された原核生物または真核生物）を有する新規な分子に関する予測または出力を生成することができる。本発明のいくつかの態様では、本明細書に記載のプラットフォーム、システムおよび方法は、所望の特性を最適化するための指向性進化、分子発見、および分子工学に適用することができる。
【０１２７】
以下の用語は、当業者によるこれらの用語の理解に加えて、本明細書で使用される用語の意味を説明するために説明される。「少なくとも」、「を超える」、または「およびそれを超える」という用語が２またはそれを超える一連の数値の第１の数値に先行するときはいつでも、「少なくとも」、「を超える」、または「およびそれを超える」という用語は、その一連の数値の各数値に適用される。例えば、１、２、または３もしくはそれを超えるは、１もしくはそれを超える、２もしくはそれを超える、または３もしくはそれを超えるに等しい。
【０１２８】
「以下」、「未満」、または「およびそれ未満」という用語が２またはそれを超える一連の数値の第１の数値に先行するときはいつでも、「以下」、「未満」、または「およびそれ未満」という用語は、その一連の数値の各数値に適用される。例えば、３、２、または１もしくはそれ未満は、３もしくはそれ未満、２もしくはそれ未満、または１もしくはそれ未満に等しい。
【０１２９】
本明細書で使用される場合、「生物学的多様性」という用語は、一般に、１またはそれを超える生物学的分子間の変動または差異を指す。生物学的分子は、例えば、イオン、小分子、高分子（例えば、ペプチド、タンパク質、脂質、炭水化物、核酸分子）、代謝産物、細胞、組織、および／またはそれらの組み合わせを含むがこれらに限定されない、生物学において存在するかまたは合成され得る分子であってもよい。１またはそれを超える生

物学的分子間の変動または差異は、機能的、構造的、またはその両方であってもよい。１またはそれを超える生物学的分子間の変動または差異は、生物学的、化学的、物理的、機械的、光学的、熱的、または他の活性、機能、構造、または特性の差異または変動を含み得る。生物学的多様性は、天然に存在するか、病的症状に見られるか、または天然に存在しない、例えば合成または操作された生物学的分子の差異を指すことができる。生物学的多様性は、遺伝子型の差異または表現型の差異を指すことができる。

【０１３０】

本明細書で使用される場合、「生物学的バリアント」という用語は、一般に、バリエーションまたは多様性を示す生物学的分子の形態またはバージョンを指す。生物学的バリアントは、細胞、細胞小器官、核酸分子、ペプチドもしくはタンパク質、脂質、炭水化物、代謝産物、またはそれらの組み合わせなどの異なる生物学的分子を含む群（例えば、ライブラリ）のメンバーであってもよい。異なる生物学的分子は、同じタイプまたは異なるタイプの分子であってもよい。例えば、２つの生物学的バリアントは、同じタンパク質またはペプチドをコードする２つのＤＮＡ分子を含むことができる。あるいは、２つの生物学的バリアントは、異なるタンパク質またはペプチドをコードする２つのＤＮＡ分子を含むことができる。いくつかの例では、２またはそれを超える生物学的バリアントは、配列相同性を示し得る。例えば、２またはそれを超える生物学的バリアントは、少なくとも１０％、少なくとも２０％、少なくとも３０％、少なくとも４０％、少なくとも５０％、少なくとも６０％、少なくとも７０％、少なくとも８０％、少なくとも９０％、少なくとも９５％またはそれを超える配列相同性または構造相同性を示し得る。いくつかの例では、２またはそれを超える生物学的バリアントは、組成、構造または機能においてわずかな差を示し得る。例えば、２またはそれを超える生物学的バリアントは、組成、構造、機能において、１％未満、０．１％未満、０．０１％未満、０．００１％未満、０．０００１％未満、またはそれ未満異なっていてもよい。いくつかの例では、生物学的バリアントは、遺伝的バリアントまたは分子バリアントであってもよい。

【０１３１】

本明細書で使用される場合、「遺伝的バリアント」は、一般に、ＲＮＡ、ＤＮＡ、または他の核酸分子の配列などの異なる遺伝要素を含む群（例えば、ライブラリ）のメンバーを指し、これらは、同じまたは異なるタンパク質、ペプチド、モチーフ、アミノ酸、または他のタンパク質構造要素をコードし得る。遺伝的バリアントは、異なる核酸配列、構造、修飾（例えば、メチル化）、塩基置換または遺伝要素の他の違いを有し得る。遺伝的バリアントは、突然変異誘発細胞を含むことができる。遺伝的バリアントは、タンパク質またはペプチドをコードし得るか、または非コード核酸分子（例えば、イントロン、トランスファーＲＮＡ、リボソームＲＮＡ、調節ＲＮＡなど）であってもよい。いくつかの例では、２またはそれを超える遺伝的バリアントが配列相同性を示し得る。例えば、２またはそれを超える遺伝的バリアントは、少なくとも１０％、少なくとも２０％、少なくとも３０％、少なくとも４０％、少なくとも５０％、少なくとも６０％、少なくとも７０％、少なくとも８０％、少なくとも９０％、少なくとも９５％またはそれを超える配列相同性を示し得る。

【０１３２】

本明細書で使用される場合、「分子バリアント」という用語は、一般に、小分子、治療用分子、ポリマーもしくはポリマー分子、生物学的分子、例えばＲＮＡ、ＤＮＡもしくは他の核酸分子、タンパク質、ペプチド、モチーフ、アミノ酸もしくは他のタンパク質構造要素、脂質、炭水化物、糖タンパク質、リポタンパク質、またはそれらの組み合わせなどの異なる分子を含む群（例えば、ライブラリ）のメンバーを指す。分子バリアント（例えば、ポリマーバリアント）は、異なる化学組成、構造、修飾、電荷、質量、または他の物理的もしくは化学的特性を有し得る。例えば、ポリマーバリアントは、核酸分子またはペプチドもしくはタンパク質であってもよい。核酸分子は、ＲＮＡまたはＤＮＡであってもよい。ＲＮＡは、コードまたは非コードであってもよく、例えば、メッセンジャーＲＮＡ（ｍＲＮＡ）、リボソームＲＮＡ（ｒＲＮＡ）またはトランスファーＲＮＡ（ｔＲＮＡ）

を含むことができる。ＲＮＡは転写物であってもよい。ＲＮＡは、２００核酸塩基長未満の小型ＲＮＡ、または２００核酸塩基長を超える大型ＲＮＡであってもよい。低分子ＲＮＡとしては、５．８Ｓリボソーム ＲＮＡ（ｒＲＮＡ）、５ＳｒＲＮＡ、トランスファーＲＮＡ（ｔＲＮＡ）、マイクロＲＮＡ（ｍｉＲＮＡ）、低分子干渉ＲＮＡ（ｓｉＲＮＡ）、低分子核小体ＲＮＡ（ｓｎｏＲＮＡ）、Ｐｉｗｉ結合ＲＮＡ（ｐｉＲＮＡ）、ｔＲＮＡ由来低分子ＲＮＡ（ｔｓＲＮＡ）および低分子ｒＲＮＡ由来ＲＮＡ（ｓｒＲＮＡ）を挙げることができる。ＲＮＡは、二本鎖ＲＮＡまたは一本鎖ＲＮＡであってもよい。ＲＮＡは環状ＲＮＡであってもよい。いくつかの例では、２またはそれを超える分子バリアントは、配列相同性または組成相同性を示し得る。例えば、２またはそれを超える分子バリアントは、少なくとも１０％、少なくとも２０％、少なくとも３０％、少なくとも４０％、少なくとも５０％、少なくとも６０％、少なくとも７０％、少なくとも８０％、少なくとも９０％、少なくとも９５％またはそれを超える組成相同性を示し得る。いくつかの例では、生物学的バリアントまたは遺伝的バリアントは、分子バリアントであるか、または分子バリアントを含む。分子バリアントは、生物学的バリアントまたは遺伝的バリアントであってもよいか、または逆に、生物学的バリアントは、遺伝的バリアントまたは分子バリアントであってもよい。

【０１３３】

株最適化、分子最適化および分子発見

本明細書では、所望の生物学的バリアントを選別および同定するためのシステムおよび方法が開示される。本明細書に開示されるシステムおよび方法は、一般に、統合システムにデータを入力し、分子、所望の特性を有する新規な分子、所望の特性を有する最適化された（例えば、操作された、または進化した）分子、株の情報（例えば、所望の特性を有する分子を産生する生物の核酸またはタンパク質情報）および／または所望の性能を有する最適化された株（例えば、細胞、核酸配列またはタンパク質配列）の構造－機能関係に関するデータおよび／または予測を出力することを含み得る。システムは、株最適化ユニット、分子最適化ユニット、および分子発見ユニットからなる群から選択される１またはそれを超える統合ユニットを含む統合システムであってもよい。統合ユニットは、本明細書で説明されるように、実験データを取得するか、または実験データを他の入力と共に入力として使用するように、個別にまたは集合的に構成することができる。代替的または追加的に、統合ユニットは、大きなデータセット（例えば、データベース、公開されているソースなどから）を入力として取得または使用するように個別にまたは集合的に構成することができる。全体として、本明細書に記載のシステムおよび方法は、分子および株の最適化および発見のために、経験的データならびに大きなデータセット（例えば、公開されているデータベースから）の両方を使用して、プラットフォーム技術における生物学的バリアントのハイスループットスクリーニングに有用である。

【０１３４】

いくつかの例では、１またはそれを超える統合ユニットのそれぞれは、データセット（例えば、データベース、経験的データ、他の統合ユニットから生成されたデータなどから）を入力として利用し、（ｉ）最適な株（例えば、所望の機能的特性（例えば、タンパク質もしくは酵素活性、分子結合もしくは親和性、または他の代謝もしくは表現型形質）のためのコード化分子（例えば、ＤＮＡまたはＲＮＡなどの核酸分子）を有する少なくとも１つの生物に関する情報を出力することができる株最適化ユニットを使用する）、（ｉｉ）最適な分子（例えば、小分子、ペプチド、またはタンパク質を出力することができる分子最適化ユニットを使用する）、または（ｉｉｉ）所望の機能的特性を有する新規な分子（例えば、小分子、ペプチド、またはタンパク質を出力することができる分子発見ユニットを使用する）に関するデータおよび／または１またはそれを超える予測を出力する。場合によっては、統合ユニットの少なくとも１つのユニットは、実験データを得るように構成され、実験データは、いくつかの例では、最適な株、最適な分子、または新規な分子に関する予測を生成するための同じまたは異なる統合ユニットへの入力として使用される。

【０１３５】

（ｉ）株最適化ユニット

いくつかの例では、本明細書に開示されるシステムおよび方法は、株最適化ユニットを利用し、これは、高度に予測可能なハイスループット形式で、機能的特性について株（例えば、細胞型、遺伝的バリアントを有する細胞など）をスクリーニングおよび最適化するのに有用であり得る。そのような株最適化ユニットは、１またはそれを超えるフィードバックループまたはコントローラの使用を含むことができ、これは、株最適化（例えば、生成された生成物の量、生成物の活性レベル、またはさらに改善される細胞の生成の一貫性などの予測可能な機能的特性）の予測可能性および一貫性を高めるのに役立ち得る。

【０１３６】

株最適化ユニットは、所望の機能的特性について株（例えば、遺伝的多様性を有する真核生物または原核生物）を反復的にスクリーニングするのに有用であり得る。例えば、株最適化のための方法は、（ａ）遺伝的バリアントのライブラリを複数の細胞に導入すること、（ｂ）遺伝的バリアントを有する複数の細胞を所望の機能的特性（例えば、表現型形質（例えば、タンパク質の産生または分泌の増加）または代謝形質）について複数の細胞をスクリーニングするのに十分な条件に供すること、（ｃ）所望の特性を有する細胞を単離すること、（ｄ）遺伝的バリアントの追加のライブラリを生成することであって、遺伝的バリアントの追加のライブラリが所望の特性をコードするバリアントを含む、生成すること、および（ｅ）性能要件が満たされるまで（ａ）～（ｄ）を繰り返すことを含み得る。別の例では、特性が改善された細胞を生成するために株最適化が実行されてもよい。そのような例では、改善された細胞または改善された特性を有する細胞を生成するための方法は、（ａ）突然変異誘発細胞のライブラリを生成すること、（ｂ）突然変異誘発細胞のライブラリを複数のパーティション（例えば、反応容器）に分配して、分配された突然変異誘発細胞を得ること、（ｃ）分配された変異誘発細胞を表現型についてスクリーニングすること、および（ｄ）スクリーニングに基づいて、分配された突然変異誘発細胞から改善された細胞を（例えば、閾値ゲートを使用して）選択すること、（ｅ）（ｉ）複数のパーティションの生成頻度を（ｂ）において維持するために、または（ｉｉ）閾値ゲートを（ｄ）において調節するためにフィードバックコントローラを使用すること、（ｆ）改善された細胞を使用して（ａ）～（ｄ）を繰り返すことを含み得る。一般に、株最適化ユニットは、遺伝的バリアントのライブラリを生成すること、遺伝的バリアントを細胞に導入すること、必要に応じて、様々な増殖、培養、または発酵条件下で細胞を培養すること、所望の形質について細胞をスクリーニングすること、所望の遺伝的バリアントの一部を組み合わせること、または形質を濃縮して（例えば、所望の形質を含む組換え遺伝子または導入遺伝子を生成することによって）遺伝的バリアントの新しいライブラリを生成すること、および性能基準または要件が満たされるまでプロセスを繰り返すことにおいて有用であってもよい（例えば、実施例１、４および５を参照されたい）。

【０１３７】

株最適化ユニットは、実験モジュールを含み得、これは、それぞれが異なる株最適化遺伝的バリアントまたは発酵条件を有する複数の別個の株最適化反応容器を含み得る。例として、株最適化ユニットは、所望の機能的特性を有するように、指向性進化手法および細胞または細胞集団の操作に使用することができる。そのような例では、遺伝的多様性を有する複数の細胞（これは、天然に存在し得るか、または例えば、導入遺伝子もしくは組換え遺伝子の突然変異誘発もしくは形質転換、形質導入などを介して誘導され得る）または突然変異誘発細胞のライブラリなどの生物学的バリアントを、株最適化ユニットの反応容器内で、必要に応じて複数の発酵条件下でスクリーニングすることができ、所望の機能的特性を有する細胞を濃縮（例えば、単離、収集）し、分析（例えば、シーケンシングを介して）して遺伝的バリアントを（例えば、核酸または遺伝子配列を使用して）同定することができる。所望の特性（例えば、目的の遺伝子、指定された発酵条件下でのタンパク質の発現、または他の表現型もしくは代謝形質）を有する遺伝的バリアントを単離し、複数の組換え遺伝子に組み換え、例えば、細胞への組換え遺伝子の形質転換、トランスフェクション、または形質導入、必要に応じて細胞の培養、および所望の機能的特性を有する、

必要に応じてよりストリンジェントな性能要件を有する細胞の濃縮または単離によって再スクリーニングすることができる。各サイクル（本明細書では「多様性スクリーニング」とも呼ばれる）がよりストリンジェントな性能要件（例えば、表現型形質のより高いもしくはより低い代謝活性、酵素活性、またはより高いもしくはより低い発現、一連の発酵条件下での収量の増加、頑強性、例えばイオン濃度、温度、ｐＨなどの差に対する耐性、生存率の増加、細胞増殖または増殖の増加）を有するように、プロセスを再帰的サイクルで反復することができる。代替的または追加的に、複数のプロセスを並列および／または連続して実行することができる。再帰的サイクルおよびそれに含まれる操作は、所望の性能要件を満たすために必要な数のサイクルにわたって実行されてもよい。いくつかの例では、生物学的多様性（例えば、遺伝子の組換えおよび細胞への組換え遺伝子の導入を介して、または突然変異誘発、例えば紫外線、エチルメチルスルホネートなどの突然変異誘発物質への曝露を介して）の追加の入力が再帰的サイクル中に導入され得る。代替的または追加的に、条件の多様性（例えば、発酵または培養条件の変更）が再帰的サイクル中に導入され得る。

【０１３８】

いくつかの実施形態では、株最適化は、複数回のスクリーニングおよび濃縮を実行し、生物学的多様性を（例えば、突然変異誘発を介して）追加して新しい株の世代を生成し、追加の複数回のスクリーニングおよび濃縮を行うことによって達成され得る。一例では、突然変異誘発細胞のライブラリを生成し、必要に応じて細胞を増殖させるのに十分な条件下で培養することができる。次いで、突然変異誘発細胞のライブラリを反応容器に分配し、必要に応じて細胞を増殖または拡大増殖させるのに十分な条件下で培養することができる（例えば、液体培養中、発酵条件の規定のセット（複数可）において）。いくつかの例では、個々の細胞を個々の反応容器に分配することができる。そのような分配は、本明細書の他の箇所で説明するように、コンピュータ可読媒体およびフィードバック制御によって支援され得る。続いて、分配された細胞をスクリーニングに供し、スクリーニングに基づいて、所望の機能的特性（例えば、所望の表現型）を有する細胞を濃縮（例えば、単離、収集、または選択）することができる。これらの濃縮された細胞は、追加の複数回の濃縮（例えば、再分配、培養、スクリーニングおよび濃縮）に供することができる。規定の（例えば、ユーザ決定またはコンピュータ誘導の決定）回数の濃縮後、濃縮された細胞をさらに（例えば、シーケンシングを介して）分析して、遺伝的バリアントを同定することができる。最も好ましい遺伝的バリアント（例えば、最も高い機能的性能をもたらすもの）を単離し、複数の組換え遺伝子に組み換えて、新しい世代の生物学的バリアント（例えば、遺伝的バリアント、突然変異誘発細胞など）を提供することができる。次いで、例えば、組換え遺伝子の細胞への形質転換、トランスフェクション、または形質導入によって、必要に応じて細胞を培養し、所望の機能的特性を有する、必要に応じてよりストリンジェントな性能要件を有する細胞を濃縮または単離することによって、新しい世代の生物学的バリアントを再スクリーニングすることができる。例えば、図４～図５を参照されたい。

【０１３９】

所望の機能的性能を達成するために、任意の数のラウンドまたは世代を反復することができる。例えば、一世代内で、少なくとも１回、少なくとも２回、少なくとも３回、少なくとも４回、少なくとも５回、少なくとも６回、少なくとも７回、少なくとも８回、少なくとも９回、少なくとも１０回、少なくとも５０回、少なくとも１００回、少なくとも５００回、少なくとも１，０００回またはそれを超える回数の濃縮または選別が行われ得る。あるいは、一世代内で、最大で約１，０００回、最大で約５００回、最大で約１００回、最大で約５０回、最大で約１０回、最大で約９回、最大で約８回、最大で約７回、最大で約６回、最大で約５回またはそれ未満の濃縮または選別が行われ得る。同様に、任意の数の世代（例えば、突然変異誘発または生物学的多様性の付加）を実行することができ、例えば、最適化のために１、２、３、４、５、６、７、８、９、１０、５０、１００またはそれを超える世代を生成することができる。あるいは、最大で約１００世代、最大で約

５０世代、最大で約１０世代、最大で約９世代、最大で約８世代、最大で約７世代、最大で約６世代、最大で約５世代、最大で約４世代、最大で約３世代、最大で約２世代またはそれ未満の世代（例えば、複数回の突然変異誘発）が行われ得る。

【０１４０】

いくつかの例では、株最適化ユニットを使用して様々な発酵条件をスクリーニングすることが有用であり得る。そのような場合、必要に応じて遺伝的多様性（これは、本明細書に記載するように、天然に存在し得るか、または誘導され得る）を含む複数の細胞を、異なる発酵条件で培養することができる。発酵条件の異なる操作可能なパラメータの非限定的な例としては、温度、供給速度、増殖または栄養培地組成、酸素化、塩分、ｐＨ、炭素源、二酸化炭素濃度、緩衝液濃度、イオン濃度、培養期間、灌流または混合、通気、反応時間、金属イオン濃度、添加剤濃度、供給スケジュール、それらの組み合わせ、またはそれらのいずれか１つもしくは組み合わせの変更が挙げられる。発酵条件の１またはそれを超える操作可能なパラメータを、所与の一連の遺伝的バリアントに対して調整することができる。複数の細胞は、細胞を反応容器に分配する前に規定の発酵条件下で培養することができ、または細胞は、反応容器内で規定の発酵条件で培養することができる。発酵条件は、反応容器のサブセットまたはすべてにわたって同じであっても異なっていてもよく、独立して、連続でまたは並列にスクリーニングされてもよい。様々な発酵条件下でのそのようなバリアントのスクリーニングは、規定の発酵条件下での機能的特性、例えば、低酸素、低温または高温、非正常ｐＨ、低グルコースなどの環境での収率、増殖またはタンパク質発現レベルの増加をもたらす遺伝的バリアントの同定に有用であり得る。いくつかの例では、様々な発酵条件下でのバリアントのスクリーニングは、所望の機能的特性をもたらし得る工業的（例えば、１００ミリリットル（ｍＬ）超、１リットル（Ｌ）、２Ｌ、５Ｌ、またはそれを超える体積で）発酵条件を同定するのに有用であり得る。

【０１４１】

次いで、培養細胞を株最適化ユニットの反応容器でスクリーニングし、規定の発酵条件下で所望の機能的特性を有する細胞を単離し、収集し、（例えば、シーケンシングを介して）分析して、遺伝的バリアントを（例えば、核酸または遺伝子配列を使用して）同定することができる。いくつかの例では、シーケンシングを使用して、規定の発酵または培養条件下で濃縮された突然変異を同定することができる。所望の特性（例えば、目的の遺伝子、指定された発酵条件下でのタンパク質の発現、または他の表現型もしくは代謝形質）を有する遺伝的バリアントを単離し、必要に応じてさらに培養し（例えば、生物の生存率または増殖を最大化するのに十分な条件下で）、複数の組換え遺伝子に組み換え、例えば、細胞への組換え遺伝子の形質転換、トランスフェクション、または形質導入、発酵条件下での細胞の培養、および所望の機能的特性を有する、必要に応じてよりストリンジェントな性能要件（例えば、よりストリンジェントなまたは指定された発酵条件下での増殖）を有する細胞の単離によって再スクリーニングすることができる。各サイクルがよりストリンジェントな性能要件（例えば、よりストリンジェントな発酵条件、例えば、低酸素条件、高温または低温、低グルコースなど）を有するように、プロセスを再帰的サイクルで反復することができる。再帰的サイクルおよびそれに含まれる操作は、所望の性能要件を満たすために必要な数のサイクルにわたって実行されてもよい。いくつかの例では、複数のプロセスを並行して実行することができる。いくつかの例では、生物学的多様性（例えば、遺伝子の組換えおよび細胞への組換え遺伝子の導入を介して、または突然変異誘発、例えば紫外線、エチルメチルスルホネートなどの突然変異誘発物質への曝露を介して）の追加の入力を再帰的サイクル中に導入し、それによって新しい世代の株または突然変異誘発細胞を生成することができる。代替的または追加的に、条件の多様性（例えば、発酵または培養条件の変更）が再帰的サイクル中に導入され得る。

【０１４２】

細胞は、任意の有用なまたは使いやすい工程で培養することができる。例えば、生物学的多様性（例えば、標的化または非標的化突然変異誘発を介して）を導入した後、細胞を培養することができる。細胞は、分配、スクリーニング、濃縮後、または生物学的多様性

を付加する追加のサイクルの後に培養することができる。細胞は、プロセスに応じて、異なる段階で異なる条件下で培養され得ることが理解されよう。例えば、細胞は、スクリーニング前に準最適（例えば、低酸素、低栄養素など）条件で培養することができるが、濃縮後、細胞をより最適な発酵条件で培養し、増殖を促進することができる。

【０１４３】

本明細書に記載のシステムおよび方法を使用して、複数の実験条件を同時にスクリーニングすることができる。例えば、株最適化ユニットは、同じまたは異なる発酵条件下で増殖させた数百万の反応体積または生物学的バリアント（例えば、数百から数百万またはそれを超える遺伝的バリアントを有する細胞）を同時にスクリーニングまたは分析するように構成することができる。スクリーニングは、例えば、反応容器の画像化ならびに画像化の結果に基づく細胞および／または反応容器の選別を介して、本明細書に記載の１またはそれを超えるアッセイを実施することによって実施され得る。有益なことに、株最適化ユニットは、遺伝的バリアント、および生物学的バリアントの環境（または発酵）条件に関する多重化されたマルチパラメトリック情報を提供することができる。さらに、株最適化ユニットは、所望の性能基準（例えば、閾値酵素活性レベルまたは他の代謝もしくは表現型形質などの所望の特性を有する）または発酵条件に基づいて株を選択するのに有益であり得る。

【０１４４】

反応容器：株最適化ユニットの反応容器は、例えば、液滴またはプラグ、ウェル（例えば、マイクロウェル、ナノウェル、ピコウェルなど）、バイアル、チューブ、フラスコ、または他の容器であってもよい。いくつかの例では、反応容器はマイクロ流体デバイスに含まれる。いくつかの例では、反応容器は、例えばマイクロウェルアレイまたはプレート（または他の基材）、ナノウェルアレイまたはプレート（または他の基材）などのアレイのメンバーである。本明細書に記載するプロセスのいずれかのために、同時または連続のいずれかで、複数のタイプの反応容器が使用され得る（例えば、液滴およびウェル）。反応容器は、物理的に別個のものであってもまたは別々のものであってもよく、または反応容器は、合流または合体することが可能であってもよい。反応容器は、１またはそれを超える試薬が吸着し得る基質を含むことができる、または試薬は溶液中にあってもよい。

【０１４５】

各反応容器は、スクリーニングのための異なる株最適化遺伝的バリアントを含むことができる。例えば、所望の機能的特性（例えば、生体分子（例えば、酵素）活性、生体分子相互作用、表現型または代謝形質）について遺伝的バリアントのライブラリをスクリーニングすることが有用であってもよく、したがって、各反応容器は、スクリーニングのために異なる遺伝的バリアントを含有するように構成することができる。場合によっては、各反応容器は、遺伝的バリアントを有する１またはそれを超える細胞を含む。例えば、各反応容器は、他の反応容器中の他の細胞の遺伝的バリアントとは異なる単一の遺伝的バリアントを有する単一の細胞または複数の細胞を含み得る。代替的または追加的に、各反応容器は、単一の遺伝的バリアントを有する細胞のクローン集団を含み得る。他の例では、単一の反応容器内で異なる遺伝的バリアントをスクリーニングすることが有用であってもよく、したがって、反応容器は、異なる遺伝的バリアントを有する細胞の混合物を含み得る。そのような例では、各反応容器は、細胞と遺伝的バリアントとの組み合わせを含み得る。液滴の場合などのいくつかの例では、１またはそれを超える液滴を合体または合流させて、細胞または遺伝的バリアントを組み合わせることができる。

【０１４６】

反応容器は、単一の細胞または複数の細胞を含み得る。例えば、単一の反応容器は、０個の細胞、１個の細胞、２個の細胞、３個の細胞、４個の細胞、５個の細胞、６個の細胞、７個の細胞、８個の細胞、９個の細胞、１０個の細胞、２０個の細胞、５０個の細胞、１００個の細胞、２００個の細胞、５００個の細胞、１，０００個の細胞またはそれを超える細胞を含み得る。１またはそれを超える反応容器は、異なる数の細胞を含み得、例えば、反応容器は、平均１０～１００個の細胞を含み得る。１またはそれを超える細胞は、

同じ起源を有し得るか、または異なる起源を有し得る。１またはそれを超える細胞は、同じタイプであってもよく、または異なるタイプであってもよい。１またはそれを超える細胞は、任意の有用な種または起源、例えば細菌、酵母、哺乳動物（例えば、ヒト、霊長類、ラット、マウス、ウサギ、ヤギ、ブタ、ウシ、ヒツジ、ウマ）、爬虫類、鳥類などに由来し得る。１またはそれを超える細胞は、原核生物、真核生物またはそれらの組み合わせであってもよい。細胞は、単細胞生物または多細胞生物に由来し得る。細胞は、対象から得られた細胞試料に由来し得る。対象は、哺乳動物（例えば、ヒト、霊長類、ラット、マウス、ウサギ、ヤギ、ブタ、ウシ、ヒツジ、ウマ）、爬虫類、鳥類などであってもよい。細胞試料は、生検、スワブ、灌注、ブラッシング、収集などによって得ることができる。各反応容器は、単一の細胞または複数の細胞の分析を実施するように構成することができる。

【０１４７】

いくつかの例では、反応容器は、細胞を含まなくてもよく、細胞プロセスを行うのに十分な反応条件（例えば、転写、翻訳、酵素反応など）を含んでもよい。そのような場合、反応容器は、酵素（例えば、ＤＮＡポリメラーゼ、ＲＮＡポリメラーゼ、逆転写酵素、ヘリカーゼ、リガーゼ）、核酸分子（例えば、ヌクレオチド、トランスファーＲＮＡ）、タンパク質（例えば、リボソーム）、脂質、炭水化物、代謝産物、またはそれらの任意の組み合わせを含むがこれらに限定されない、そのような細胞プロセスを行うための任意の有用な試薬を含む無細胞系を含み得る。

【０１４８】

反応容器は、ハイスループット方式で評価または分析されるように構成することができる。例えば、株最適化ユニット（例えば、実験モジュール）は、少なくとも１００細胞／分、少なくとも５００細胞／分、少なくとも１，０００細胞／分、少なくとも５，０００細胞／分、少なくとも１０，０００細胞／分、少なくとも５０，０００細胞／分、少なくとも１００，０００細胞／分またはそれを超える細胞をスクリーニングするように構成することができる。代替的または追加的に、反応容器をハイスループット方式でスクリーニングすることができる。例えば、株最適化ユニットは、少なくとも１０反応容器／分、少なくとも１００反応容器／分、少なくとも５００反応容器／分、少なくとも１，０００反応容器／分、少なくとも５，０００反応容器／分、少なくとも１０，０００反応容器／分、少なくとも５０，０００反応容器／分、少なくとも１００，０００反応容器／分をスクリーニングするように構成することができる。いくつかの例では、株最適化ユニットは、少なくとも１０反応体積／分、少なくとも１００反応体積／分、少なくとも５００反応体積／分、少なくとも１，０００反応体積／分、少なくとも５，０００反応体積／分、少なくとも１０，０００反応体積／分、少なくとも５０，０００反応体積／分、少なくとも１００，０００反応体積／分をスクリーニングするように構成することができる。

【０１４９】

所与の期間内に任意の数の反応容器を生成または使用することができる。いくつかの例では、異なる数の反応容器（または体積）をスクリーニングすることが有益であり得る。したがって、使用または生成される反応容器（または体積）の数は、適切な用途または実験目的に従って調整することができる。例えば、多数の別々の反応体積が、例えばスクリーニング目的（例えば、株最適化のために）に有用である場合、少なくとも１０，０００反応体積、少なくとも１００，０００反応体積、少なくとも１，０００，０００反応体積、少なくとも１０，０００，０００反応体積、少なくとも１００，０００，０００反応体積またはそれを超える反応体積が使用され得る。他の例では、より少ない数の別々の反応体積が有用または必要とされ得る（例えば、標的スクリーニングのために、またはスクリーニングのための限られた数の種を用いた実験の合理的な設計のために）。そのような場合、最大で１００，０００，０００反応体積、最大で１０，０００，０００反応体積、最大で１，０００，０００反応体積、最大で１００，０００反応体積、最大で１０，０００反応体積、最大で１，０００反応体積またはそれ未満の反応体積が使用または生成され得る。

【０１５０】
いくつかの例では、反応容器はマイクロ流体デバイス内に含まれる。マイクロ流体デバイスは、株最適化ユニットの消費または使い捨てユニットの一部であってもよい。場合によっては、反応容器は、ナノウェルまたは液滴（またはプラグ）を含み得る。そのような場合、マイクロ流体デバイスは、液滴発生器、液滴凝縮器、液滴合流器、緩衝液リザーバ、細胞インキュベータ、流体注入器、液滴スプリッター、液滴選別機、ディレイライン、液滴シンギュレータなどの１またはそれを超える機能要素を含み得る。
【０１５１】
反応容器は、任意の有用な体積を収容することができる。反応容器は、１００マイクロリットル（μＬ）未満、１０μＬ未満、１μＬ未満、５００ナノリットル（ｎＬ）未満、２００ｎＬ未満、１００ｎＬ未満、５０ｎＬ未満、２０ｎＬ未満、１０ｎＬ未満、５ｎＬ未満、２ｎＬ未満、１ｎＬ未満、５００ピコリットル（ｐＬ）未満、２００ｐＬ未満、１００ｐＬ未満、５０ｐＬ未満、２０ｐＬ未満、１０ｐＬ未満、５ｐＬ未満、２ｐＬ未満、１ｐＬ未満、５００フェムトリットル（ｆＬ）未満、またはそれ未満の体積を有し得る。いくつかの実施形態では、反応容器は、５００ｆＬを超える、１ｐＬを超える、２ｐＬを超える、５ｐＬを超える、１０ｐＬを超える、２０ｐＬを超える、５０ｐＬを超える、１００ｐＬを超える、２００ｐＬを超える、５００ｐＬを超える、１ｎＬを超える、２ｎＬを超える、５ｎＬを超える、１０ｎＬを超える、２０ｎＬを超える、５０ｎＬを超える、１００ｎＬを超える、２００ｎＬを超える、５００ｎＬを超える、１μＬを超える、２μＬを超える、５μＬを超える、１０μＬを超える、１００μＬを超える、またはそれを超える体積を有し得る。反応体積は、例えば、約１ｐＬ～約２０ｐＬ、約１０ｐＬ～１００ｎＬなどの体積の範囲内にあってもよいことが理解されよう。複数の反応容器を使用してもよく、複数の反応容器は、同じまたは異なる体積を含み得る。本明細書の他の箇所でさらに説明するように、複数の反応容器の容積は、コンピュータ実装方法を使用して制御することができる。
【０１５２】
液滴の場合、液滴は任意の有用な組成物を含み得る。例えば、液滴は、油中水型液滴、水中油型液滴、水－油－水の液滴、油－水－油の液滴などを含み得る。いくつかの例では、液滴は、表面上または表面に隣接する空気中の液滴（例えば、油、水）を含み、反応容器（例えば、ウェル）に包含され得る。液滴は、複数のエマルジョン、例えば二重エマルジョン、三重エマルジョンなどを含み得る。
【０１５３】
アッセイ：いくつかの例では、生物学的バリアントの選別は、１またはそれを超えるスクリーニング操作の使用を含み得る。スクリーニング操作は、アッセイ、例えば、直接アッセイ、間接アッセイ、結合アッセイ、レポーターアッセイ、蛍光発生アッセイ、レポーター細胞に基づくアッセイ、発色アッセイまたは比色アッセイなどを含み得る。酵素結合免疫吸着アッセイ（ＥＬＩＳＡ）、サンドイッチＥＬＩＳＡ、近接ライゲーションアッセイなどを含むがこれらに限定されない１またはそれを超える免疫学的アッセイを使用することができる。いくつかの例では、レポーター分子または選択分子がスクリーニングのために使用される。アッセイ分子またはレポーター分子または選択分子は、生物学的バリアントの遺伝的バリアントまたは機能的特性（例えば、表現型または代謝形質）の存在の検出に有用であり得る。例えば、所望の生物学的バリアントが酵素活性を含む細胞である場合、反応容器内の細胞は、所望の酵素活性によって基質が生成物に変換されると色が変化するかまたは蛍光を発する発色性または蛍光発生性基質の存在下でスクリーニングされ得る。そのような例では、色または蛍光の変化を示す反応容器は、細胞が所望の酵素活性を含み、続いて、その細胞を単離し、さらに分析して（例えば、次世代シーケンシング（ＮＧＳ）またはＰＣＲベースのシーケンシングなどのシーケンシング、オープンリーディングフレーム同定または分析、タンパク質分析などを介して）、生物学的バリアントに関する情報（例えば、ＤＮＡまたはＲＮＡ配列、タンパク質配列または組成物など）を決定することができることを示し得る。他の例では、所望の生物学的バリアントは、選択分子ま

たはレポーター分子を含み得、例えば、選択分子の場合、所望の形質を有する細胞のみが所与の培養条件で生存し得る（例えば、抗生物質耐性を包含する導入遺伝子または組換え遺伝子を有する細胞のみが、抗生物質を含む増殖培地中で生存する）。生存細胞は、所望の機能的特性についてスクリーニングされ得、所望の機能的特性を有するものは、さらなる分析に供することができる。さらに別の例では、所望の形質を有する細胞のみが、導入遺伝子または組換え遺伝子が存在することを示すレポーター分子（例えば、蛍光性または検出可能なタンパク質）を発現し得る。そのような例では、望ましい形質または機能的特性を示す細胞は、本明細書に記載するように、反応容器から単離され、さらなる分析に供することができる。他の例では、アッセイ（例えば、イムノアッセイ、結合アッセイなど）を使用して、各遺伝的バリアント中の目的の分子（例えば、タンパク質、核酸分子、脂質、炭水化物、代謝産物、またはそれらの組み合わせ）の存在を決定または測定することができ、目的の分子を含む遺伝的バリアントを単離し、さらに分析することができる。

【０１５４】

選択された生物学的バリアント（例えば、所望の特性を示す選択された細胞）に対してさらなる分析を行うことができる。そのような分析としては、本明細書に記載するように、タンパク質分析（例えば、質量分析、タンパク質アッセイ）、核酸分析（例えば、次世代シーケンシング（ＮＧＳ）などのシーケンシング）、脂質分析（例えば、ラマン散乱または分光法）、画像化（および画像分析）、分光法、または他の分析方法もしくはスクリーニングを挙げることができるが、これらに限定されない。そのような分析方法は、例えば、ＤＮＡ配列または遺伝子、アミノ酸配列、タンパク質特性（例えば、翻訳後修飾）に関する情報を得ることによって、遺伝的バリアントの同一性を決定するのに有用であり得る。次いで、そのような情報を収集してデータベースに保存するか、または必要に応じて所望の生物学的活性、表現型もしくは代謝形質などの任意の機能的特性を有する株最適化ライブラリと再統合することができる。そのようなさらなる分析は、任意の有用または使いやすい工程で行われ得る。例えば、いくつかの例では、生物学的バリアントの機能的性能を決定する前に、（例えば、シーケンシングまたは他のタンパク質分析を介して）一連の生物学的バリアントを同定または特性評価することが有益であり得る。そのような例では、機能的性能について生物学的バリアントをスクリーニングする前に、そのような特性評価、例えば（例えば、細胞のクローン集団の一部の）シーケンシングおよび任意のインデックス付けを行うことができる。機能的性能のスクリーニングに続いて、配列は、機能的性能に対して（例えば、コンピュータ可読媒体を使用して）インデックス付けまたはペアリングされ得、それにより、構造（例えば、配列）情報と機能的性能との間の相関に関する情報が提供される。

【０１５５】

いくつかの例では、株最適化ユニットは、検出器を備えるか、または検出器に結合されてもよい。検出器は、細胞または反応容器における目的の分子または機能的特性の有無を決定するために使用され得、所望の生物学的バリアントのスクリーニングおよび／または選択を可能にし得る。検出器は、１またはそれを超える分析機器、例えば質量分析計、ラマン分光計、近赤外（ＮＩＲ）分光光度計、フーリエ変換赤外（ＦＴＩＲ）分光計、分光光度計（例えば、蛍光または吸光度プレートまたはキュベットリーダ）、圧電センサなどを含み得る。検出器は、細胞を計数する（例えば、セルカウンタ、フローサイトメータ）ように構成することができる。検出器は、１またはそれを超える光学構成要素を備えてもよく、いくつかの例では、顕微鏡（例えば、立体鏡、複合顕微鏡、蛍光顕微鏡、偏光顕微鏡、共焦点顕微鏡、微分干渉顕微鏡、超解像顕微鏡など）、またはそれらの構成要素もしくは組み合わせを備えてもよい。検出器は、磁気、インピーダンス、トポグラフィまたは幾何学的形状、散乱、形態、吸光度、蛍光、色、細胞のクラスター化、細胞またはコロニーサイズなどを含むがこれらに限定されない、細胞または反応容器の物理的特性を検出するために使用され得る。検出器は、点検出器またはアレイもしくは画像検出器であってもよい。

【０１５６】

移動アクチュエータ：いくつかの例では、反応容器移動アクチュエータを使用して（例えば、検出器に向かって）反応容器を移動させる。容器移動アクチュエータは、ロボットマニピュレータ、フロージェネレータ、音響ドロップジェネレータ、光ピンセット、熱ドロップオンデマンド、圧電ドロップオンデマンド、リソグラフィユニット、デジタルマイクロ流体デバイス、誘電泳動、またはそれらの組み合わせなどの流体ハンドリングユニットであってもよい。移動アクチュエータは、ステージと、例えば、ねじ、歯車、ねじ山、モータ、ポンプ、圧縮機、抵抗器、トランジスタ、ばね、滑車などの任意の必要な移動部品とを備えることができる。いくつかの例では、反応容器は、移動アクチュエータに接続された基材上または基材に隣接して配置される。

【０１５７】

ライブラリ：株最適化ユニットは、株最適化ライブラリを含み得、これは、本明細書に記載のプロセスで使用され得る。株最適化ライブラリは、生物学的多様性のライブラリを含むことができ、例えば、株最適化ライブラリは、遺伝的バリアントなどの複数の異なる生物学的バリアントを含み得る。複数の遺伝的バリアントは、第１の株に由来し得る。例えば、第１の細胞株（例えば、細菌、酵母、動物、植物など）の細胞は、遺伝的バリアントの各々が異なるように、複数の異なる遺伝的バリアントを生成するのに十分な条件に供することができる。複数の遺伝的バリアントの各遺伝的バリアントは、例えば、複数の遺伝的バリアントの他の遺伝的バリアントとは異なる配列を有するＤＮＡまたはＲＮＡを含み得る。複数の異なる遺伝的バリアントは、同じ生体分子（例えば、ＤＮＡ、ＲＮＡ、ペプチド、タンパク質など）の変異をコードし得る。例えば、複数の異なる遺伝的バリアントは、後にタンパク質に翻訳されるＲＮＡ配列の変異をコードする異なるＤＮＡ配列を含み得る。ライブラリ中の異なるＤＮＡ配列の変異は、同じまたは異なるタンパク質をコードし得る、異なるＲＮＡ配列をもたらすことができる。いくつかの例では、遺伝的バリアント（ＤＮＡおよび／またはＲＮＡの変異を有し得る）によってコードされるタンパク質は酵素である。そのような場合、酵素は、代謝経路を形成し得るか、またはその一部であってもよい。酵素または代謝活性は、株最適化プロセス中のスクリーニングパラメータまたは性能要件として使用され得る。

【０１５８】

場合によっては、株最適化ユニットは、遺伝的バリアントをスクリーニングし、表現型または代謝測定に基づいて所望の生物学的バリアント（例えば、遺伝的バリアント）を選択および／または分類するように構成される。表現型または代謝測定は、非限定的な例では、タンパク質または酵素活性、タンパク質発現レベル、副産物または代謝産物の数量などを含むことができる。例えば、所望の生物学的バリアントは、規定のまたは最小閾値の酵素活性を有するものであってもよい。そのような場合、株最適化ライブラリの遺伝的バリアントを株最適化ユニットでスクリーニングすることができ、規定のまたは最小閾値の酵素活性を有する遺伝的バリアントのみをさらなる分析のために選択する。遺伝的バリアントをさらなる分析（例えば、シーケンシング、タンパク質分析）に供して、目的の酵素の核酸またはタンパク質配列を決定することができる。決定された配列情報（例えば、核酸またはタンパク質配列）および対応する機能的性能は、１またはそれを超えるコンピュータ可読媒体（例えば、シミュレータまたはモデル）に入力されてもよく、これは、本明細書の他の箇所に記載されているように、その後、構造要素（例えば、核酸またはタンパク質配列）を機能的性能と相関させるために使用されてもよい。場合によっては、核酸配列は、望ましい特性（例えば、同じまたは異なる酵素活性）を示す他の核酸配列と組み換えられ、それによって遺伝的バリアントの新しいライブラリが生成され得る。このプロセスは、酵素活性の所望の閾値に達するまで繰り返すことができる。

【０１５９】

場合によっては、株最適化ライブラリは、生成または合成された遺伝的バリアントを含む。複数の遺伝的バリアントは、１またはそれを超える技術によって生成され得る。例えば、遺伝的バリアントは、突然変異誘発（ランダム突然変異誘発（例えば、紫外線、Ｘ線照射、ガンマ線照射などの突然変異原、化学変異原、例えばエチルメチルスルホネート、

活性酸素種、脱アミノ化剤、多環芳香族炭化水素、アルキル化剤（例えば、エチルニトロソウレア、ニトロソアミン）芳香族アミン、アルカロイド、臭素、アジ化ナトリウム、ソラレン、ベンゼン、金属（例えば、ヒ素、カドミウム、クロム、ニッケル）または生物学的薬剤（例えば、トランスポゾン、ウイルス、細菌）などへの曝露）、飽和突然変異誘発、循環置換）、多重自動ゲノム工学法（ＭＡＧＥ）、ＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、ＣＲＩＳＰＲガイドＤＮＡポリメラーゼ、突然変異原との接触、エラープローン（ｅｒｒｏｒ－ｐｒｏｎｅ）ポリメラーゼ連鎖反応（ｅｐＰＣＲ）の実施、または他のゲノム工学的アプローチによって生成され得る。いくつかの例では、株最適化ライブラリは、全ゲノム突然変異誘発ライブラリ、ゲノムシャッフルライブラリ、プロモータースワップライブラリなどの標的ゲノムライブラリ、およびトランスポゾンライブラリ（例えば、プロモーター挿入ライブラリ）であるか、またはそれらを含む。いくつかの実施形態では、標的化ゲノムライブラリは、１またはそれを超える異種遺伝要素を含む。異種遺伝要素は、プロモーター、５’非翻訳領域、３’非翻訳領域、リボザイム、ＲＮＡ安定性配列、分泌ペプチド、融合タンパク質、ＤＮＡ結合ドメイン、タンパク質－タンパク質結合領域、コドン最適化遺伝子、コドンランダム化遺伝子、ターミネーターおよび非相同末端結合からなる群から選択される１またはそれを超えるメンバーを含むか、またはそれらをコードし得る。株最適化ライブラリは、標的突然変異誘発または非標的（例えば、ランダム）突然変異誘発を使用して生成された遺伝的バリアントを含み得る。いくつかの例では、株最適化ライブラリは、突然変異誘発細胞のライブラリを含む。

【０１６０】

株最適化ライブラリは、所望の生物学的バリアントの株をスクリーニングまたは最適化するのに有用な数の遺伝的バリアントを含み得る。株最適化ライブラリは、少なくとも１００の遺伝的バリアント、少なくとも１，０００の遺伝的バリアント、少なくとも１０，０００の遺伝的バリアント、少なくとも１００，０００の遺伝的バリアント、少なくとも１，０００，０００の遺伝的バリアント、少なくとも１０，０００，０００の遺伝的バリアント、少なくとも１００，０００，０００の遺伝的バリアントまたはそれを超える遺伝的バリアントを含み得る。本明細書で説明するように、株最適化プロセスは、再帰的または反復的であってもよい。そのような場合、反復ごとにスクリーニングされる遺伝的バリアントの数は様々であってもよく、例えば、最初のサイクルは１０，０００，０００の遺伝的バリアントをスクリーニングし、２番目のサイクルは１００，０００の遺伝的バリアントをスクリーニングすることができる。逆に、遺伝的バリアントの数は、例えば、新しい生物学的多様性またはバリアント（例えば、新しい株、新しい遺伝的バリアントなど）および／または条件付き多様性（例えば、新たに導入されたまたはスクリーニングされた生物学的バリアントの一部または全部に対する様々な発酵条件）の導入によって、サイクル間で増加し得る。

【０１６１】

株最適化のためのシステムおよび方法のさらなる例は、例えば、参照によりその全体が本明細書に組み込まれる、Ｊ．Ａｇｒｅｓｔｉら、２００９、「Ｕｌｔｒａｈｉｇｈ－ｔｈｒｏｕｇｈｐｕｔ ｓｃｒｅｅｎｉｎｇ ｉｎ ｄｒｏｐ－ｂａｓｅｄ ｍｉｃｒｏｆｌｕｉｄｉｃｓ ｆｏｒ ｄｉｒｅｃｔｅｄ ｅｖｏｌｕｔｉｏｎ．」、ＰＮＡＳに見出すことができる。以下の実施例１、４および５も参照されたい。

【０１６２】

（ｉｉ）分子発見ユニット

いくつかの例では、本明細書に開示されるシステムおよび方法は、機能的特性を有する新規な分子（例えば、タンパク質、核酸などの生体分子）の発見に有用であり得る、分子発見ユニットを利用する。本明細書に記載のシステムおよび方法は、バリアントのライブラリをハイスループットスクリーニング、バリアントの情報または特性評価を取得すること、次いで、取得した情報または特性評価を使用して、ハイスループットで一貫した手法

で、機能的特性を有すると予測される新しい分子を設計することを含み得る。

【０１６３】

分子発見ユニットは、所望の機能的特性を有する新規な分子を発見するのに有用であり得る。例えば、分子発見ユニットは、発見分子（例えば、試料もしくはライブラリからのＤＮＡ、またはライブラリからのＤＮＡ情報）を取得または収集し、表現型または代謝形質（例えば、酵素活性）などの所望の機能的特性について収集された発見分子をスクリーニングするように構成することができる。所望の機能的特性を示す分子を単離し、（例えば、シーケンシングを使用して）特性評価することにより、発見分子および機能的特性に関するデータ（例えば、酵素活性レベル、標的タンパク質の発現など）を生成することができる。次いで、データは、データベースを含み得るコンピュータ可読媒体（例えば、シミュレータまたはモデル）に入力されてもよく、これは、発見分子の構造－機能関係、所与の一連の適用条件に対する所与の構造もしくは配列の予想される性能、またはモデル入力に基づく他の予測を予測するように構成することができる。いくつかの実施形態では、コンピュータ可読媒体は機械学習アルゴリズムを含む。

【０１６４】

分子発見ユニットは、実験モジュールを含み得、この実験モジュールは、それぞれが異なる発見分子または適用条件を有する複数の別個の分子発見反応容器を含み得る。一例では、分子発見のための方法は、（ａ）ＤＮＡ配列を収集または採取すること、（ｂ）収集または採取されたＤＮＡ配列を複数の細胞に導入すること、（ｃ）収集または採取されたＤＮＡ配列を有する複数の細胞を、所望の機能的特性（例えば、表現型または代謝形質）について複数の細胞をスクリーニングするのに十分な条件に供すること、（ｄ）所望の特性を有する細胞を単離すること、および（ｅ）単離された細胞の配列を決定することを含み得る。得られたデータ（例えば、シーケンシングの読み取りまたはデータ）は、コンピュータ可読媒体（例えば、シミュレータまたはモデル）に入力されてもよく、コンピュータ可読媒体は、どの配列、領域、またはモチーフが所望の機能的特性を有するかを予測するために使用され得る。必要に応じて、コンピュータ可読媒体は、得られるデータ（例えば、シーケンシングの読み取りまたはデータ）を同じタイプのデータ（例えば、所望の特性を有するタンパク質の既知の配列）のデータベースと比較するように構成することができる。分子発見ユニットにおける１またはそれを超えるプロセスは、再帰的または反復的であってもよい。代替的または追加的に、複数のプロセスを並列および／または連続して実行することができる。

【０１６５】

別の例では、分子発見のための方法は、表現型の改善された機能的性能をもたらす生物学的バリアント（例えば、遺伝的変異体）の発見によって改善された細胞を生成することを含み得る。株最適化と同様に、そのような方法は、例えば、（ａ）突然変異誘発細胞のライブラリを生成すること、（ｂ）突然変異誘発細胞のライブラリを複数のパーティションに分配して、分配された変異誘発細胞を得ること、（ｃ）分配された突然変異誘発細胞を表現型についてスクリーニングすること、および（ｄ）スクリーニングに基づいて、分配された突然変異誘発細胞から改善された細胞を選択すること、および（ｅ）改善された細胞を使用して（ａ）～（ｄ）を繰り返すことを含み得る。分子発見プロセスは、表現型の所望の機能的特性について分子バリアント（例えば、突然変異誘発細胞のライブラリ）をスクリーニングし、分子バリアントを分析し、観察された表現型および得られた分析に基づいて構造－機能関係を決定するのに有用であり得る。

【０１６６】

一例では、分子発見ユニットは、表現型形質、例えばタンパク質の産生または酵素の触媒活性に対する遺伝子配列の構造的機能的関係を提供し得る。そのような発見は、複数回のスクリーニングおよび濃縮生物学的多様性の任意の追加（例えば、突然変異誘発を介して）を実行し、新しい世代の生物学的バリアントを生成し、追加の複数回のスクリーニングおよび濃縮を行うことによって達成され得る。例えば、突然変異誘発細胞のライブラリを生成し、必要に応じて細胞を増殖させるのに十分な条件下で培養することができる。次

いで、突然変異誘発細胞のライブラリを反応容器に分配し、必要に応じて細胞を増殖または拡大増殖させるのに十分な条件下で培養することができる（例えば、液体培養中、発酵条件の規定のセット（複数可）において）。いくつかの例では、個々の細胞を個々の反応容器に分配することができる。そのような分配は、本明細書の他の箇所で説明するように、コンピュータ可読媒体およびフィードバック制御によって支援され得る。続いて、分配された細胞をスクリーニングに供し、スクリーニングに基づいて、所望の機能的特性（例えば、所望の表現型）を有する細胞を濃縮（例えば、単離、収集、または選択）することができる。これらの濃縮された細胞は、追加の複数回の濃縮（例えば、再分配、培養、スクリーニングおよび濃縮）に供することができる。規定の（例えば、ユーザ決定またはコンピュータ誘導の決定）回数の濃縮後、濃縮された細胞をさらに（例えば、シーケンシングを介して）分析して、（例えば、核酸または遺伝子配列を使用して）遺伝的バリアントを同定することができる。

【０１６７】

いくつかの例では、構造－機能関係は、例えばコンピュータ可読媒体を使用して濃縮係数を計算することによって決定され得る。濃縮係数は、例えば、別の細胞、例えば、選別されていないまたは濃縮されていない細胞、野生型細胞、親細胞などと比較した、選別され濃縮された細胞から産生された生成物の量の差または比などの比較であってもよいか、またはそれを含み得る。いくつかの例では、濃縮係数は、選別されていない集団と比較した、選別された細胞の集団におけるバリアント（例えば、遺伝的バリアント）の出現の比を含み得る。濃縮係数は、機能的性能（例えば、機能的性能）の代理として機能し得る。並行して、濃縮された細胞（例えば、最高の機能的性能をもたらすもの）から配列情報を得て、遺伝子配列および／またはゲノム突然変異を決定することができる。次いで、同定された遺伝子配列（またはゲノム突然変異）を濃縮係数に割り当てるかまたはペアリングし、それにより、機能的特性に対する遺伝子配列の構造－機能関係を生成することができる。

【０１６８】

いくつかの例では、濃縮係数は、選別された試料と比較した、選別されていない試料中のバリアント（例えば、遺伝的バリアント）の出現数の比較であるか、またはそれを含み得る。例えば、各遺伝的バリアント（例えば、シーケンシングを使用して同定される場合）の存在量は、選別された濃縮細胞と同様に選別されていない細胞集団においても決定され得、濃縮係数は、選別されていない細胞または濃縮されていない細胞における遺伝的バリアントの存在量に対する濃縮された細胞における遺伝的バリアントの存在量の比または差をとることによって計算され得る。そのような例では、選別されていない試料またはその一部をシーケンシングに供して、選別されていない試料内の遺伝的バリアントを同定することができ、例えば、選別されていない試料中のすべての遺伝的バリアントの中の特定の遺伝的バリアントの出現数を決定することによって、特定の遺伝的バリアントの存在量を計算することができる。同様に、選別された試料またはその一部をシーケンシングに供して、選別された試料内の遺伝的バリアントを同定することができ、例えば、選別された試料中のすべての遺伝的バリアントの中の特定の遺伝的バリアントの出現数を決定することによって、特定の遺伝的バリアントの存在量を計算することができる。選別された試料と比較した、選別されていない試料中の特定の遺伝的バリアントの存在量は、濃縮プロセスからのその特定の遺伝的バリアントの濃縮を示し得る。例えば、選別されていない試料中の１００，０００配列のうち１０回出現し（または１：１０，０００の頻度）、選別された試料中の１００配列のうち１回出現する特定の遺伝的バリアントＸの場合、濃縮係数は（１／１００）／（１／１０，０００）または１００倍である。濃縮係数は、機能的性能（例えば、生成された生成物の量またはタンパク質もしくは酵素の活性レベル）の変化の代理または推定値として有用であり得る。あるいは、濃縮係数は、選別されていない例と比較して、選別された試料中の遺伝的バリアントの出現数によって計算され得る。濃縮係数は、（ｉ）選別されていない試料または（ｉｉ）前回の濃縮から得られた試料と比較して、選別または濃縮された試料中のバリアントの出現数を比較し得ることが理解されよ

う。

【０１６９】

濃縮係数は、いくつかの例では、約１、約２、約３、約４、約５、約６、約７、約８、約９、約１０、約２０、約３０、約４０、約５０、約１００、約５００、約１０００、約１０，０００、約１００，０００、約１，０００，０００、約１０，０００，０００倍またはそれを超える倍数であってもよい。濃縮係数は、例えば、約１０倍～約１００倍、約１００倍～約１，０００倍などの範囲であってもよい。

【０１７０】

いくつかの例では、分子発見プロセスは、細胞を含まなくてもよく、異なる発見分子を収集または採取すること、それらを１またはそれを超える適用条件に供すること、および所望の特性を有する分子バリアントを単離することを含み得る。異なる発見分子のさらなる特性評価（例えば、化学分析または生化学分析）を行うことができる。例えば、所望の特性（例えば、所望の物理的または化学的特性を示す代謝活性）を有する発見分子を、画像化（および画像解析）、分光法、質量分析法、重量測定、原子分光法、核磁気共鳴法（ＮＭＲ）、クロマトグラフィー、プラズマ質量分析法、Ｘ線光電子分光法（ＸＰＳ）、オージェ電子分光法などの化学分析または元素分析に供することができる。得られたデータ（例えば、化学組成、物理的特性など）を使用して、どの構造要素（例えば、元素、化学組成など）が所望の機能的特性を有するかを予測することができる。

【０１７１】

分子発見ユニットは、分子のライブラリ（例えば、異なる標的タンパク質をコードする複数の遺伝的バリアントもしくはＤＮＡ配列を含む遺伝的ライブラリ、または異なるタンパク質をコードするもしくは機能的特性をもたらす複数の分子バリアントを含む分子ライブラリ）を細胞または細胞集団に導入し、必要に応じて１またはそれを超える適用条件下で細胞を培養し、所望の機能的特性について細胞をスクリーニングすることによって、新しい分子（例えば、タンパク質またはペプチドなどの生体分子）を発見するために使用され得る。所望の機能的特性（例えば、指定された適用条件下での特定の標的タンパク質の発現、または他の表現型もしくは代謝形質）を示す細胞または細胞集団を単離し、さらに処理または分析することができる（例えば、シーケンシングを使用する）。そのようなプロセスの反復的サイクルまたは再帰的サイクルは、所与の一連の適用条件に対する分子の化学的または構造的組成（例えば、タンパク質モチーフ、ＤＮＡまたはアミノ酸配列）および機能または性能（例えば、所望の特性、表現型、代謝形質などを有する）に関する大量のデータを生成するのに有用であり得る。いくつかの例では、生物学的多様性（例えば、遺伝子の組換えおよび細胞への組換え遺伝子の導入を介して、または突然変異誘発を介して）の追加の入力を再帰的サイクル中に導入することができる。代替的または追加的に、条件の多様性（例えば、適用または培養条件の変更）が再帰的サイクル中に導入され得る。

【０１７２】

いくつかの例では、分子発見ユニットを使用して様々な適用条件をスクリーニングすることが有用であり得る。そのような場合、必要に応じて分子変異（これは、本明細書に記載するように、天然に存在し得るか、または誘導され得る）を含む反応容器を、異なる適用条件に供することができる。異なる適用条件の非限定的な例としては、温度、増殖または栄養培地組成、反応緩衝液組成、基質濃度、酸素化、塩分、ｐＨ、炭素源、緩衝液濃度、培養期間、マトリックス組成、タンパク質濃度、支持体（例えば、ビーズまたは他の固体支持体）への分子またはバリアントの付着または固定、および通気が挙げられる。例えば、複数の細胞におけるまたは複数の細胞を使用した分子発見の場合、複数の細胞は、細胞を反応容器に分配する前に規定の発酵条件のセットの下で培養することができ、または細胞は、反応容器内で規定の発酵条件で培養することができる。発酵条件は、反応容器のサブセットまたはすべてにわたって同じであっても異なっていてもよい。別の例では、複数の細胞は、複数の細胞の性能または所望の特性（例えば、標的タンパク質発現レベル、酵素活性レベル）に影響を及ぼし得る処理（例えば、薬物または小分子処理、基質の存在

、様々な反応条件など）を含む適用条件に供することができる。適用条件を変化させ、複数の細胞の性能への影響を観察することは、環境の変動が構造または機能にどのように影響するかを理解するのに有用である可能性があり、および／または所与の適用条件下で所望の特性を有する新規または最適化された分子の発見に使用され得る。

【０１７３】

細胞内でのまたは細胞を使用した生体分子発見の場合、培養細胞を分子発見ユニットの反応容器内でスクリーニングすることができ、規定の適用条件下で所望の機能的特性（例えば、標的タンパク質バリアントの発現、酵素活性、代謝または表現型形質）を有する細胞を単離し、収集し、分析（例えば、シーケンシングを介して）して、遺伝的バリアントまたはタンパク質バリアントを（例えば、核酸またはアミノ酸配列を使用して）同定することができる。所望の特性（例えば、指定された適用条件下での標的タンパク質の発現、または他の表現型もしくは代謝形質）を有する遺伝的バリアントまたはタンパク質バリアントを単離することができる。遺伝的バリアントまたはタンパク質バリアント（例えば、核酸またはアミノ酸配列）および対応する性能（例えば、標的タンパク質発現レベル）に関する情報を生成することができる。各サイクルが構造または配列要素および対応する性能に関するデータを生成するように、プロセスを再帰的サイクルで反復することができる。いくつかの例では、生物学的多様性（例えば、遺伝子の組換えおよび細胞への組換え遺伝子の導入を介して、または突然変異誘発を介して）の追加の入力を再帰的サイクル中に導入し、それによって構造または配列要素および性能に対するそのような構造または配列要素の効果に関する追加データを生成することができる。代替的または追加的に、条件の多様性（例えば、適用条件の変更）が再帰的サイクル中に導入され得る。

【０１７４】

異なる発見分子は、ライブラリ（例えば、タンパク質ライブラリ、ペプチドライブラリ、生体分子ライブラリ）に含まれてもよく、または異なる発見分子は、ゲノムバリアントライブラリにおいてコードされてもよく、これは、天然に存在するかまたは生成もしくは合成されたものからの遺伝的バリアントを含み得る。例えば、天然に存在する遺伝子ライブラリは、試料（例えば、生物学的試料、土壌試料など）から集められたＤＮＡ配列を含み得、環境ＤＮＡ（ｅＤＮＡ）ライブラリを構成または含み得る。他の例では、ライブラリは、合成される遺伝的バリアントを含み得る。例えば、遺伝的バリアントは、突然変異誘発（ランダム突然変異誘発（例えば、紫外線、Ｘ線照射、ガンマ線照射などの突然変異原、化学変異原、例えばエチルメチルスルホネート、活性酸素種、脱アミノ化剤、多環芳香族炭化水素、アルキル化剤（例えば、エチルニトロソウレア、ニトロソアミン）芳香族アミン、アルカロイド、臭素、アジ化ナトリウム、ソラレン、ベンゼン、金属（例えば、ヒ素、カドミウム、クロム、ニッケル）または生物学的薬剤（例えば、トランスポゾン、ウイルス、細菌）などへの曝露）、飽和突然変異誘発、循環置換）、多重自動ゲノム工学法（ＭＡＧＥ）、ＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、ＣＲＩＳＰＲガイドＤＮＡポリメラーゼ、エラープローン（ｅｒｒｏｒ－ｐｒｏｎｅ）ポリメラーゼ連鎖反応（ｅｐＰＣＲ）、または他のゲノム工学的アプローチによって生成され得る。いくつかの例では、ゲノムバリアントライブラリは、全ゲノム突然変異誘発ライブラリ、ゲノムシャッフルライブラリ、プロモータースワップライブラリなどの標的ゲノムライブラリ、およびトランスポゾンライブラリ（例えば、プロモーター挿入ライブラリ）であるか、またはそれらを含む。いくつかの例では、標的ゲノムライブラリは、プロモーター、５’非翻訳領域、３’非翻訳領域、リボザイム、ＲＮＡ安定性配列、分泌ペプチド、融合タンパク質、ＤＮＡ結合ドメイン、タンパク質－タンパク質結合領域、コドン最適化遺伝子、コドンランダム化遺伝子、ターミネーターおよび非相同末端結合からなる群から選択される１またはそれを超えるメンバーを含むか、またはそれらをコードする。いくつかの例では、ゲノムバリアントライブラリは、単一遺伝子ランダム突然変異ライブラリ、部位飽和ライブラリ（例えば、すべてのアミノ酸が、各部位において、２０個のアミノ酸のそれぞれに変異しているタンパク

質またはペプチドライブラリ）、小型インサートメタゲノムライブラリ、大型インサートメタゲノムライブラリ、またはアプローチの組み合わせであってもよい。いくつかの例では、異なる発見分子はＤＮＡコード化分子である。ＤＮＡコード化分子は、メタゲノムライブラリ、ＤＮＡコード化ライブラリ、１またはそれを超える生物からのライブラリ、またはファンクションジェネレータによって作成されたライブラリなどのライブラリに含まれ得る。ＤＮＡコード化分子ライブラリは、ランダムに生成されたＤＮＡコード化分子、天然に存在するＤＮＡコード化分子、またはそれらの組み合わせを含み得る。

【０１７５】

株最適化ユニットと同様に、本明細書に記載のシステムおよび方法を使用して複数の実験条件を同時にスクリーニングすることができる。例えば、分子発見ユニットは、同じまたは異なる適用条件に供された異なる発見分子または分子バリアントを含む数百万の反応体積を同時にスクリーニングまたは分析するように構成することができる。有益なことに、分子発見ユニットは、分子バリアントおよび分子バリアント（本明細書では「発見分子」とも呼ばれる）の環境（または適用）条件に関する多重化されたマルチパラメトリック情報を提供し得る。さらに、分子発見ユニットは、所望の性能基準を有する新しい分子（例えば、ＤＮＡによってコードされるタンパク質）、例えば閾値酵素活性レベルまたは他の代謝もしくは表現型形質または特定の適用条件下での標的タンパク質の産生などの所望の特性を有する分子を発見するのに有益であり得る。

【０１７６】

反応容器：分子発見ユニットの反応容器は、例えば、液滴またはプラグ、ウェル（例えば、マイクロウェル、ナノウェル、ピコウェルなど）、バイアル、チューブ、フラスコ、または他の容器であってもよい。いくつかの例では、反応容器は、マイクロ流体デバイスに含まれ、このマイクロ流体デバイスは、消費ユニットを含むかまたは消費ユニットの一部であってもよい。いくつかの例では、反応容器は、例えばマイクロウェルアレイまたはプレート、ナノウェルアレイまたはプレートなどのアレイのメンバーである。本明細書に記載するプロセスのいずれかのために、同時または連続のいずれかで、複数のタイプの反応容器が使用され得る（例えば、液滴およびウェル）。反応容器は、物理的に別個のものであってもまたは別々のものであってもよく、または反応容器は、合流または合体することが可能であってもよい。

【０１７７】

各反応容器は、スクリーニングのための異なる発見分子を含むことができる。例えば、株最適化ユニットと同様に、所望の機能的特性（例えば、生体分子（例えば、酵素）活性、生体分子相互作用、表現型または代謝形質）について異なるタンパク質をコードする遺伝的バリアントのライブラリをスクリーニングすることが有用であってもよく、したがって、各反応容器は、スクリーニングのために異なる遺伝的バリアントを含有するように構成することができる。場合によっては、各反応容器は、遺伝的バリアントを有する１またはそれを超える細胞を含む。例えば、各反応容器は、他の反応容器中の他の細胞の遺伝的バリアントとは異なる単一の遺伝的バリアントを有する単一の細胞または複数の細胞を含み得る。代替的または追加的に、各反応容器は、単一の遺伝的バリアントを有する細胞のクローン集団を含み得る。他の例では、単一の反応容器内で異なる遺伝的バリアントをスクリーニングすることが有用であってもよく、したがって、反応容器は、異なる遺伝的バリアントを有する細胞の混合物を含み得る。そのような例では、各反応容器は、細胞と遺伝的バリアントとの組み合わせを含み得る。

【０１７８】

上記のように、反応容器は、単一の細胞または複数の細胞を含み得、ハイスループット方式で評価されるように構成することができる。例えば、分子発見ユニットは、少なくとも１００細胞／分、少なくとも５００細胞／分、少なくとも１，０００細胞／分、少なくとも５，０００細胞／分、少なくとも１０，０００細胞／分、少なくとも５０，０００細胞／分、少なくとも１００，０００細胞／分以上をスクリーニングするように構成することができる。代替的または追加的に、反応容器をハイスループット方式でスクリーニング

することができる。例えば、分子発見ユニットは、少なくとも１０反応容器／分、少なくとも１００反応容器／分、少なくとも５００反応容器／分、少なくとも１，０００反応容器／分、少なくとも５，０００反応容器／分、少なくとも１０，０００反応容器／分、少なくとも５０，０００反応容器／分、少なくとも１００，０００反応容器／分をスクリーニングするように構成することができる。いくつかの例では、分子発見ユニットは、少なくとも１０反応体積／分、少なくとも１００反応体積／分、少なくとも５００反応体積／分、少なくとも１，０００反応体積／分、少なくとも５，０００反応体積／分、少なくとも１０，０００反応体積／分、少なくとも５０，０００反応体積／分、少なくとも１００，０００反応体積／分をスクリーニングするように構成することができる。

【０１７９】

所与の期間内に任意の数の反応容器を生成または使用することができる。いくつかの例では、異なる数の反応容器（または体積）をスクリーニングすることが有益であり得る。したがって、使用または生成される反応容器（または体積）の数は、適切な用途または実験目的に従って調整することができる。例えば、多数の別々の反応体積が、例えばスクリーニング目的（例えば、分子発見のために）に有用である場合、少なくとも１０，０００反応体積、少なくとも１００，０００反応体積、少なくとも１，０００，０００反応体積、少なくとも１０，０００，０００反応体積、少なくとも１００，０００，０００反応体積またはそれを超える反応体積が使用され得る。他の例では、より少ない数の別々の反応体積が有用または必要とされ得る（例えば、標的スクリーニングのために、またはスクリーニングのための限られた数の種を用いた実験の合理的な設計のために）。そのような場合、最大で１００，０００，０００反応体積、最大で１０，０００，０００反応体積、最大で１，０００，０００反応体積、最大で１００，０００反応体積、最大で１０，０００反応体積、最大で１，０００反応体積またはそれ未満の反応体積が使用または生成され得る。

【０１８０】

株最適化ユニットと同様に、分子発見ユニットの反応容器は、マイクロ流体デバイス内に含まれてもよく、任意の有用な体積を収容してもよい。液滴の場合、上述のように、液滴は任意の有用な組成物を含み得る。

【０１８１】

アッセイ：いくつかの例では、生物学的バリアントの選別は、１またはそれを超えるスクリーニング操作の使用を含み得る。スクリーニング操作は、アッセイ、例えば、直接アッセイ、間接アッセイ、結合アッセイ、レポーターアッセイ（例えば、レポーター細胞に基づくアッセイ）、蛍光発生アッセイ、発色アッセイまたは比色アッセイなどを含み得る。酵素結合免疫吸着アッセイ（ＥＬＩＳＡ）、サンドイッチＥＬＩＳＡ、近接ライゲーションアッセイなどを含むがこれらに限定されない１またはそれを超える免疫学的アッセイを使用することができる。いくつかの例では、レポーター分子または選択分子がスクリーニングのために使用される。アッセイ分子またはレポーター分子または選択分子は、生物学的バリアントの遺伝的バリアントまたは機能的特性（例えば、表現型または代謝形質）の存在の検出に有用であり得る。例えば、所望の生物学的バリアントが酵素活性を含む細胞である場合、反応容器内の細胞は、検出試薬、例えば所望の酵素活性によって基質が生成物に変換されると色が変化するか蛍光を発するか、またはそうでなければ周囲環境を変化させる発色性または蛍光発生性基質の存在下でスクリーニングされ得る。そのような例では、色または蛍光の変化を示す反応容器は、細胞が所望の酵素活性を含み、続いて、その細胞を単離し、さらに分析して（例えば、シーケンシング、タンパク質分析などを介して）、生物学的バリアントに関する情報（例えば、ＤＮＡまたはＲＮＡ配列、タンパク質配列または組成物など）を決定することができることを示し得る。他の例では、所望の生物学的バリアントは、選択分子またはレポーター分子を含み得、例えば、選択分子の場合、所望の形質を有する細胞のみが所与の培養条件で生存し得る（例えば、抗生物質耐性を包含する導入遺伝子または組換え遺伝子を有する細胞のみが、抗生物質を含む増殖培地中で生存する）。生存細胞は、所望の機能的特性についてスクリーニングされ得、所望の機

能的特性を有するものは、さらなる分析に供することができる。さらに別の例では、所望の形質を有する細胞のみが、導入遺伝子または組換え遺伝子が存在することを示すレポーター分子（例えば、蛍光性または検出可能なタンパク質）を発現し得る。そのような例では、望ましい形質または機能的特性を示す細胞は、本明細書に記載するように、反応容器から単離され、さらなる分析に供することができる。他の例では、アッセイ（例えば、イムノアッセイ、結合アッセイなど）を使用して、各遺伝的バリアント中の目的の分子（例えば、タンパク質、核酸分子、脂質、炭水化物、代謝産物、またはそれらの組み合わせ）の存在を決定することができ、目的の分子を含む遺伝的バリアントを単離し、さらに分析することができる。例えば、アッセイは、１またはそれを超える検出試薬、例えばｐＨ調整剤、イオン強度もしくはイオン濃度調整剤、イオン、基質、阻害剤、酵素または競合剤の添加を含み得る。

【０１８２】

１またはそれを超えるスクリーニング操作は、反応容器内のシグナルの画像化または検出を含み得る。例えば、所望の生物学的バリアントが酵素活性を含む細胞である場合、反応容器内の細胞は、検出試薬、例えば所望の酵素活性によって基質が生成物に変換されると色が変化するか蛍光を発するか、またはそうでなければ周囲環境を変化させる発色性または蛍光発生性基質の存在下で画像化され得る。取得された画像は、例えば、シグナル強度値を取得するために、コンピュータ可読媒体によって分析することができる。画像化の結果に基づいて、株最適化ユニットの構成要素は、例えば、細胞または反応容器を異なる位置に移動させることによって、細胞または反応容器を自動的に選別するように構成することができる。画像化は、例えば、分配、培養（適用可能な場合）、スクリーニング、選別などの前、間、または後に、任意の有用なまたは使いやすい工程行われ得る。

【０１８３】

選択された生物学的バリアント（例えば、所望の特性を示す選択された細胞）に対してさらなる分析を行うことができる。そのような分析としては、本明細書に記載するように、タンパク質分析（例えば、質量分析、タンパク質アッセイ）、核酸分析（例えば、シーケンシング）、脂質分析（例えば、ラマン散乱または分光法）、画像化（および画像分析）、分光法、化学組成分析（例えば、ＮＭＲ、ＸＰＳ、クロマトグラフィーなど）または他の分析方法もしくはスクリーニングを挙げることができるが、これらに限定されない。そのような分析方法は、例えば、ＤＮＡ配列または遺伝子、アミノ酸配列、タンパク質特性（例えば、翻訳後修飾）に関する情報を得ることによって、遺伝的バリアントの同一性を決定するのに有用であり得る。次いで、そのような情報を収集してデータベースに保存するか、または必要に応じて所望の生物学的活性、表現型もしくは代謝形質などの任意の機能的特性を有する分子発見ライブラリと再統合することができる。

【０１８４】

いくつかの例では、分子発見ユニットは、検出器を備えるか、または検出器に結合されてもよい。検出器は、細胞または反応容器における目的の分子または機能的特性の有無を決定するために使用され得、所望の生物学的バリアントのスクリーニングおよび／または選択を可能にし得る。検出器は、１またはそれを超える分析機器、例えば質量分析計、ラマン分光計、近赤外（ＮＩＲ）分光光度計、フーリエ変換赤外（ＦＴＩＲ）分光計、分光光度計（例えば、蛍光または吸光度プレートまたはキュベットリーダ）、圧電センサなどを含み得る。検出器は、細胞を計数する（例えば、セルカウンタ、フローサイトメータ）ように構成することができる。検出器は、１またはそれを超える光学構成要素を備えてもよく、いくつかの例では、顕微鏡（例えば、立体鏡、複合顕微鏡、蛍光顕微鏡、偏光顕微鏡、共焦点顕微鏡、微分干渉顕微鏡、超解像顕微鏡など）、またはそれらの構成要素もしくは組み合わせを備えてもよい。検出器は、磁気、インピーダンス、トポグラフィまたは幾何学的形状、散乱、形態、吸光度、蛍光、色、細胞のクラスター化、細胞またはコロニーサイズなどを含むがこれらに限定されない、細胞または反応容器の物理的特性を検出するために使用され得る。検出器は、点検出器またはアレイもしくは画像検出器であってもよい。検出器は、点のシグナル、アレイを収集し、点の測定値または画像を生成するよう

に構成することができる。

【０１８５】

移動アクチュエータ：いくつかの例では、反応容器移動アクチュエータを使用して（例えば、検出器に向かって）反応容器を移動させる。容器移動アクチュエータは、ロボットマニピュレータ、フロージェネレータ、音響ドロップジェネレータ、光ピンセット、熱ドロップオンデマンド、圧電ドロップオンデマンド、リソグラフィユニット、デジタルマイクロ流体デバイス、またはそれらの組み合わせであってもよい。移動アクチュエータは、ステージと、例えば、ねじ、歯車、ねじ山、モータ、ポンプ、圧縮機、抵抗器、トランジスタ、ばね、滑車などの任意の必要な移動部品とを備えることができる。いくつかの例では、反応容器は、移動アクチュエータに接続された基材上または基材に隣接して配置される。

【０１８６】

分子ライブラリ：分子発見ユニットは、発見分子ライブラリを含み得、これは、本明細書に記載のプロセスで使用され得る。分子発見ライブラリは、本明細書に記載するように、複数の異なる遺伝的バリアントを含むことができる。複数の遺伝的バリアントは、第１の株に由来し得る。例えば、第１の細胞株（例えば、細菌、酵母、動物、植物など）の細胞は、遺伝的バリアントの各々が異なるように、（例えば、標的化または非標的化突然変異誘発を介して）複数の異なる遺伝的バリアントを生成するのに十分な条件に供することができる。複数の遺伝的バリアントの各遺伝的バリアントは、例えば、複数の遺伝的バリアントの他の遺伝的バリアントとは異なる配列を有するＤＮＡまたはＲＮＡを含み得る。複数の異なる遺伝的バリアントは、同じ生体分子（例えば、ＤＮＡ、ＲＮＡ、ペプチド、タンパク質など）の変異をコードし得る。例えば、複数の異なる遺伝的バリアントは、後にタンパク質に翻訳されるＲＮＡ配列の変異をコードする異なるＤＮＡ配列を含み得る。ライブラリ中の異なるＤＮＡ配列の変異は、異なるＲＮＡ配列、したがって翻訳されたタンパク質またはペプチドをもたらすことができる。いくつかの例では、遺伝的バリアント（ＤＮＡおよび／またはＲＮＡの変異を有し得る）によってコードされるタンパク質またはペプチドは酵素である。そのような場合、酵素は、代謝経路を形成し得るか、またはその一部であってもよい。酵素または代謝活性は、分子発見プロセス中のスクリーニングパラメータまたは性能要件として使用され得る。

【０１８７】

場合によっては、分子発見ユニットは、分子バリアントをスクリーニングし、機能的特性（例えば、表現型または代謝形質または活性）または測定に基づいて所望の生物学的バリアント（例えば、分子バリアント）を選択および／または分類するように構成される。表現型または代謝測定は、非限定的な例では、タンパク質または酵素活性、タンパク質発現レベル、副産物または代謝産物の数量または量などを含むことができる。例えば、所望の生物学的バリアントは、規定のまたは最小閾値の酵素活性を有するものであってもよい。そのような場合、分子発見ライブラリの分子バリアントを分子発見ユニットでスクリーニングすることができ、規定のまたは最小閾値の酵素活性を有する分子バリアントのみをさらなる分析のために選択する。分子バリアントをさらなる分析（例えば、シーケンシング、タンパク質分析、化学分析）に供して、目的の酵素の核酸もしくはタンパク質配列または化学組成を決定することができる。いくつかの例では、分子バリアントの配列または組成をデータベースに追加し、および／または機能的特性または性能を有する分子発見ライブラリに再統合する。したがって、分子発見ユニットは、配列または構造情報を機能的特性または性能と相関させることにおいて、または非標準的な構造－機能関係を明らかにすること、例えば、所与の一連の条件下で類似の機能的性能を有するより高い多様性またはより少ない相同性を有する配列を発見することにおいて有用であり得る。

【０１８８】

分子発見ライブラリは、所望の機能的特性を有するかまたはもたらす分子をスクリーニングまたは発見するのに有用な数の分子バリアントを含み得る。分子発見ライブラリは、少なくとも１００個の異なる発見分子、少なくとも１，０００個の異なる発見分子、少な

くとも１０，０００個の異なる発見分子、少なくとも１００，０００個の異なる発見分子、少なくとも１，０００，０００個の異なる発見分子、少なくとも１０，０００，０００個の異なる発見分子、少なくとも１００，０００，０００個の異なる発見分子またはそれを超える個数の異なる発見分子を含み得る。本明細書に記載するように、分子発見プロセスまたは統合システムは、再帰的または反復的であってもよい。そのような場合、反復ごとにスクリーニングされる異なる発見分子の数は様々であってもよく、例えば、最初のサイクルは１０，０００，０００個の異なる発見分子をスクリーニングし、２番目のサイクルは１００，０００個の異なる発見分子をスクリーニングすることができる。逆に、異なる発見分子の数は、例えば、新しい生物学的多様性またはバリアント（例えば、新しい株、新しい異なる発見分子など）および／または条件付き多様性（例えば、新たに導入されたまたはスクリーニングされた生物学的バリアントの一部または全部に対する様々な発酵条件）の導入によって、サイクル間で増加し得る。

【０１８９】

（ｉｉｉ）分子最適化ユニット

いくつかの例では、本明細書に開示されるシステムおよび方法は、分子最適化ユニットを利用し、これは、機能的特性について分子（例えば、タンパク質、ペプチド、または他の生体分子）をスクリーニングおよび最適化するのに有用であり得る。本明細書に記載のシステムおよび方法は、バリアントのライブラリに情報またはデータを入力し、改善された方法で機能を実行するように構成された設計されたバリアントを出力する１またはそれを超える機械学習アルゴリズムの使用を用いてもよい。

【０１９０】

分子最適化ユニットは、分子の構造を操作することによって所望の機能的特性を最適化するだけでなく、所望の機能的特性について分子バリアントをスクリーニングし、分子の構造－機能関係（例えば、生体分子またはポリマー、例えば、タンパク質、ペプチド、核酸配列などのバイオポリマー）を予測するのに有用であり得る。一例では、分子最適化のための方法は、（ａ）分子（例えば、タンパク質、酵素）の構造的性能（例えば、配列情報）および機能的性能データを得ること、ならびに（ｂ）コンピュータ可読媒体を使用して、閾値性能基準の機能的性能または閾値性能基準を超える機能的性能を有する最適化された分子の構造に関する予測を生成することを含み得る。いくつかの例では、本方法は、（ｃ）予測された構造をバリアントのライブラリに導入すること（例えば、所望の機能的特性に関連する、または標的タンパク質をコードする遺伝子配列を含む組換え遺伝子または導入遺伝子を含み得る遺伝的バリアント）、および（ｄ）いくつかの例では反復的であってもよいバリアントのライブラリをスクリーニングすることをさらに含み得る。いくつかの例では、本方法は、コンピュータ可読媒体（例えば、モデル、シミュレータ）を使用して、性能基準閾値を超える所望の機能的特性を有する最適化された分子の構造（例えば、遺伝子配列またはアミノ酸配列、化学組成など）を予測することを含む。いくつかの例では、コンピュータ可読媒体は機械学習アルゴリズムを含む。

【０１９１】

別の例では、分子最適化のための方法は、（ａ）分子または株バリアントのライブラリを提供すること（例えば、１またはそれを超えるペプチドまたはタンパク質をコードするＤＮＡコード化ライブラリ、機能的性能を有する核酸分子（例えば、アプタマー）、分子（例えば、ポリマーまたはペプチド）バリアントを含む細胞のライブラリ）および（ｂ）いくつかの例では反復的であってもよいバリアントのライブラリをスクリーニングすることを含み得る。スクリーニングは、所望の機能的特性を有するか、または機能的性能基準を満たすバリアントを選択することを含み得る。例えば、タンパク質またはペプチドの場合、酵素活性または結合特性を監視することができ、所望の酵素活性または結合特性を有するものを選択し、必要に応じてさらに最適化することができる。

【０１９２】

いくつかの例では、分子最適化ユニットは、１またはそれを超えるデータセットを入力し、所望の機能的特性をもたらすと予測される分子または構造（例えば、配列、モチーフ

、化学組成など）の構造－機能関係の予測を出力するコンピュータを使用した構成要素（例えば、コンピュータ可読媒体）を含む。データセットは、他の統合ユニット（例えば、分子発見ユニット、株最適化ユニット）からの出力を含むことができる。例えば、どの発酵条件および遺伝的バリアントが所望の機能的特性をもたらすかなどの株最適化ユニットからのデータを分子最適化ユニットで使用して予測を生成することができる。同様に、分子最適化ユニットでは、所望の機能的特性を有する分子の配列または構造情報、材料または生物学的起源、適用条件などの分子発見ユニットからのデータを使用することができる。逆に、分子最適化ユニットからのデータは、それぞれ株最適化プロセスおよび分子発見プロセスのために株最適化ユニットまたは分子発見ユニットで使用することができる。

【０１９３】

いくつかの例では、分子最適化ユニットに入力されるデータセットは、公開されているデータベースなどのデータベースからのデータを含む。例えば、タンパク質またはペプチドの最適化の場合、ＤＤＢＪ、ＧｅｎｅＣａｒｄｓ、ＩｎｔｅｒＰｒｏ、ＮＣＢＩ、ＰＯＷＬ、ＰＩＲ、ＰＲＦ、ＳＹＳＴＥＲｓ、３Ｄｅｅ、ＡｒｃｈＤＢ、ＡＳＴＲＡＬ、ＢｉｏＭａｇＲｅｓＢａｎｋ、ＣＤＤＢ、ＥＭＤａｔａＢａｎｋ、Ｅｎｚｙｍｅ Ｓｔｒｕｃｔｕｒｅｓ、ｆＰＯＰ、Ｊｅｎａ Ｌｉｂ、ＰＭＰ、Ｐｒｏｔｅｏｐｅｄｉａ、ＭｏｌＭｏｖＤＢ、ＰＤＢ、ＰＩＳＡ、ＰＭＰ、ＳＢＫＢ、ＴＯＰＳＡＮ、ＣＯＧ、Ｐｆａｍ、ＰＲＩＮＴＳ、ｉＰｒｏＣｌａｓｓ、ＰｒｏＤｏｍ、ＰＲＯＳＩＴＥ、ＳＭＡＲＴ、ＡｕｔｏＰＳＩ、ＣＡＴＨ、ＣＤＤ、ＣＥ、ＣＬ、ＦＳＳＰ、ＨＯＭＳＴＲＡＤ、ＭＯＤＢＡＳＥ、ＰａｒｔｓＬｉｓｔ、ＰＤＢｅ、ＳＣＯＰ、ＶＡＳＴ、ＡＡＲＳＤＢ、ＡＳＤ、ＡＳＰＤ、ＢＲＥＮＤＡ、ＤＡＶＩＤ、Ｃａｔａｌｙｔｉｃ Ｓｉｔｅ Ａｔｌａｓ、ＥＦ－ｈａｎｄ ＣａＢＰ、ＥｃｏＣｙｃ、ＥＮＺＹＭＥ、Ｇｅｎｅ Ｏｎｔｏｌｏｇｙ、ＧＰＲＣＤ、Ｈｏｍｅｏｂｏｘ、ＫＥＧＧ、ＭａｔｒｉｘＤＢ、ＭＥＲＯＰＳ、ＭｅｔａＣｙｃ、Ｐ２ＣＳ、ＰＲＥＸ、Ｐｒｏｔｅｉｎ Ｋｉｎａｓｅ Ｒｅｓｏｕｒｃｅ、Ｐｒｏｔｅｉｎ Ｏｎｔｏｌｏｇｙ、ＲＮａｓｅ Ｐ、ＳｕｐｅｒＣＹＰ、ＴｒａｎｓｐｏｒｔＤＢ、ＣＰＬＡ、Ｐ３ＤＢ、Ｐｈｏｓｐｈｏ３Ｄ、ＰＨＯＳＩＤＡ、ＬＯＣＡＴＥ、ｅＳＬＤＢ、ＬＯＣＡＴＥ、ＮＰＤ、ＮＵＲＳＡ、ＯＲＧｅ、ＰＳＯＲＴｄｂ、ＳＵＢＡ、３ｄＩＤ、ＵｎｉＰｒｏｔ、ＧｅｎＢａｎｋ、ＲｅｆＳｅｑ、ＳｗｉｓｓＰｒｏｔ、Ｐｒｏｔｅｉｎ、ＭＥＤＬＩＮＥ、ＢｉｏＳｙｓｔｅｍｓ、Ｅｎｔｒｅｚ、ＳＲＳ、ＳＣＯＰＰＩ、ＲｅＬｉＢａｓｅ、Ｐｒｏｔｅｉｎ３ｄＨｏｍｅ、ＭＩＰＳ、ＩＢＩＳ、ＤＯＭＭＩＮＯ、ＤＯＭＩＮＥ、ＤＩＭＡ、ＤＩＰ、ＢｉｏＧＲＩＤなどのデータベースからのタンパク質情報を使用することができる。

【０１９４】

いくつかの例では、実験データとデータベースからのデータの両方が分子最適化ユニットで使用される。データベースからのデータ（例えば、配列情報、構造－機能関係、タンパク質－タンパク質相互作用）を経験的データ（例えば、他の統合ユニットから、または分子最適化ユニットから）と併せて使用して、所望の機能的特性またはある閾値レベルの機能的性能を有すると予想される分子バリアントを予測することができる。次いで、予測された分子バリアントを、分子最適化ユニットまたは他の統合ユニット（例えば、株最適化ユニットまたは分子発見ユニット）でスクリーニングすることができる。

【０１９５】

１つの非限定的な例では、構造または配列データを含む分子発見ユニットまたは株最適化ユニットからのデータ、例えば、目的の酵素をコードするＤＮＡ配列または所望の生物学的バリアント中の酵素のアミノ酸配列、ならびに酵素の性能（１またはそれを超える再帰的サイクルから）、および／または発酵条件もしくは適用条件を、分子最適化ユニット（例えば、分子発見ユニットのコンピュータを使用した構成要素またはコンピュータ可読媒体）に入力することができる。１またはそれを超えるタンパク質または酵素データベース（例えば、ＵｎｉＰｒｏｔ、ＤＤＢＪなど）からのデータも分子最適化ユニット（例えば、コンピュータを使用した構成要素またはコンピュータ可読媒体）に追加することができる。分子最適化ユニットのコンピュータ可読媒体は、入力を使用し、パラメータ（例え

ば、構造、配列、適用条件、発酵条件など）が性能（例えば、酵素活性）に影響を及ぼすモデルまたは予測を出力することができる１またはそれを超えるアルゴリズムを含むことができる。コンピュータ可読媒体はまた、パラメータ（例えば、遺伝子配列またはバリアント、発酵条件）が閾値性能基準（例えば、ベースライン活性レベル）の性能または閾値性能基準を超える性能をもたらす可能性が高い予測を出力することができる。

【０１９６】

いくつかの例では、予測された遺伝子バリアントは、例えば、制限消化などの組換え戦略、合成時解読法（ｓｅｑｕｅｎｃｉｎｇ－ｂｙ－ｓｙｎｔｈｅｓｉｓ）などを介して、または本明細書の他の箇所に記載されている突然変異誘発（例えば、紫外線、エチルメチルスルホネートなどの変異原への曝露）によって、プラスミドライブラリなどの遺伝子ライブラリに組み込むことができる。次いで、遺伝子ライブラリを、例えばトランスフェクション、形質転換または形質導入によって１またはそれを超える細胞に導入し、それによって遺伝的バリアントを含む複数の細胞を作成することができる。複数の細胞は、必要に応じて、１またはそれを超える適用または発酵条件下で培養することができる。本明細書の他の箇所に記載されているように、複数の細胞を、ハイスループット形式で所望の機能的特性または性能（例えば、酵素活性）についてスクリーニングすることができる。所望の機能的特性または性能を示す株を単離し、さらなる分析（例えば、シーケンシング）に供して、遺伝的バリアントを同定することができる。次いで、この情報またはデータ（例えば、機能的性能および遺伝子配列）をコンピュータ可読媒体（例えば、シミュレーションまたはモデル）に再入力して、予測品質をさらに改良することができる。そのようなプロセスは、株最適化で提示されたプロセスと同様に、所望の性能基準が満たされるまで反復的または再帰的に実行することができる。

【０１９７】

分子最適化ユニットは、実験モジュールを含み得、これは、それぞれが異なる標的タンパク質（または分子バリアント）または適用条件を有する複数の別個の分子最適化反応容器を含み得る。本明細書に記載するように、分子最適化のための方法は、分子（例えば、タンパク質、酵素）の構造（例えば、配列情報）および機能的性能データを得ること、閾値性能基準の機能的性能または閾値性能基準を超える機能的性能を有する最適化された分子の構造に関する予測を生成すること、予測された構造をバリアント（例えば、所望の機能的特性に関連する遺伝子配列を含む組換え遺伝子または導入遺伝子を含み得る遺伝的バリアント）のライブラリに導入すること、およびいくつかの例では反復的であってもよいバリアントのライブラリをスクリーニングすることを含み得る。いくつかの例では、分子最適化のための方法は、バリアント（例えば、ランダムに変異している遺伝子またはオペロンから作成された遺伝子バリアント）のライブラリを得ること、および所望の機能的性能または特性についてバリアントのライブラリをスクリーニングすることを含み得る。いくつかの例では、１またはそれを超える操作は反復的であってもよい。したがって、分子最適化ユニットは、多数の反応容器を処理するためのシステムを含み得る。

【０１９８】

例として、分子最適化ユニットは、株最適化ユニットと同様に、所望の機能的特性を有するように細胞または細胞集団の指向性進化手法および操作に使用され得る。そのような例では、１またはそれを超えるポリマーバリアントを含む細胞のライブラリまたは遺伝的多様性を有する複数の細胞（これは、天然に存在し得るか、または例えば、導入遺伝子もしくは組換え遺伝子の突然変異誘発もしくは形質転換、形質導入などを介して誘導され得る）などの生物学的バリアントを、分子最適化ユニットの反応容器内で、必要に応じて複数の適用条件下でスクリーニングすることができ、所望の機能的特性を有する細胞を選別、単離、収集、および／または分析（例えば、シーケンシングを介して）することができる。機能的特性（例えば、目的の分子の産生、目的の分子の活性レベル）は、本明細書に記載するように、画像化またはアッセイを使用して測定することができ、スクリーニングの前、スクリーニング中、またはスクリーニング後に実施することができる。いくつかの例では、生物学的バリアントのライブラリの一部は、選別されなくてもよく、または好ま

しい機能的特性を示す生物学的バリアントの位置とは異なる位置に選別されてもよい。例えば、試料は生物学的バリアントのライブラリを含むことができ、試料の一部は選別されてもよく、一部は選別されていないままである。選別されていない（または異なるように選別された）生物学的バリアントは、分子最適化プロセスからの機能的特性の増加または増強を比較または測定するための対照集団または対照群として役立ち得る。いくつかの例では、選別された細胞は、複数の細胞群が得られるように、別の数の階級に向けられてもよい。例えば、ゲーティングが実行されてもよく、最も高い機能的性能を示す細胞は第１の群に向けられてもよく、２番目に高い性能を示す細胞は第２の群に向けられてもよいなどである。

【０１９９】

選別された細胞を、さらなる特性評価または分析に供することができる。生物学的バリアントの分析は、遺伝的バリアントの同定（例えば、必要に応じて、核酸または遺伝子配列を決定するための精製およびシーケンシング）を含み得る。同様に、選別されていない細胞または異なるように選別された細胞を、例えば、配列情報またはデータを得るために、さらなる特性評価に供することができる。いくつかの例では、配列情報は、機能的特性とペアリングされてもよく、または相関されてもよい。一例では、配列情報は、例えばコンピュータ可読媒体を使用して、選別された細胞および選別されていない細胞における配列の存在量（例えば、シーケンスカウント）に基づく機能的特性とペアリングされてもよい。代替的または追加的に、各ポリマーバリアントの存在量を、選別された細胞および選別されていない細胞の両方において得ることができる。配列および存在量データから、例えば、選別された細胞における各ポリマーバリアントの存在量を選別されていない細胞における各ポリマーバリアントの存在量と比較することによって、濃縮データ（例えば、濃縮係数）を得ることができる。機能データ、配列データ、および濃縮データを、コンピュータ可読媒体（例えば、機械学習アルゴリズム）に個別にまたは集合的に入力して、次いで、改善された機能的特性を有すると予想される予測されたバリアントを設計および出力することができる。

【０２００】

いくつかの例では、バリアントのライブラリは、選別に使用され得る１またはそれを超える基準マーカーまたは配列を含み得る。例えば、試料またはバリアントのライブラリは、ポリマーバリアントの機能に基づいて選別されてもよく、ポリマーバリアントの少なくとも１つは基準配列を含む。基準配列を使用して、測定値を較正することができ、例えば、基準配列は、既知の配列および既知の機能活性を含むことができ、これを使用してスクリーニングプロセスを既知の機能活性に較正または正規化することができる。基準配列は、同じまたは異なるポリマーをコードし得る１またはそれを超える生物学的バリアントを含み得る。例えば、２つの基準配列は、同じポリマー（例えば、ペプチドまたはタンパク質）をコードする異なるＤＮＡ配列を含み得る。

【０２０１】

いくつかの例では、所望の特性（例えば、目的の遺伝子、指定された発酵条件下でのタンパク質の発現、または他の表現型もしくは代謝形質）を有する生物学的バリアントを単離し、複数の組換え遺伝子に組み換え、例えば、細胞への組換え遺伝子の形質転換、トランスフェクション、または形質導入、必要に応じて細胞の培養、および所望の機能的特性を有する、必要に応じてよりストリンジェントな性能要件を有する細胞の単離によって再スクリーニングすることができる。各サイクルがよりストリンジェントな性能要件（例えば、より高いもしくはより低い代謝活性、酵素活性、または表現型形質のより高いもしくはより低い発現）を有するように、プロセスは再帰的サイクルで反復されてもよい。再帰的サイクルおよびそれに含まれる操作は、所望の性能要件を満たすために必要な数のサイクルにわたって実行されてもよい。いくつかの例では、生物学的多様性（例えば、遺伝子の組換えおよび細胞への組換え遺伝子の導入を介して、または突然変異誘発を介して）の追加の入力を再帰的サイクル中に導入することができる。代替的または追加的に、条件の多様性（例えば、適用条件の変更）が再帰的サイクル中に導入され得る。

【０２０２】
いくつかの例では、条件付き多様性スクリーニング、例えば、分子最適化ユニットを使用して様々な適用条件をスクリーニングすることが有用であり得る。そのような場合、反応容器または分子バリアントを異なる適用条件に供することができ、必要に応じて、分子変異（これは、本明細書に記載するように、天然に存在し得るか、または誘導され得る）を含むことを異なる適用条件に供することができる。異なる適用条件の非限定的な例としては、温度、増殖または栄養培地組成、反応緩衝液組成、基質濃度、酸素化、塩分、ｐＨ、炭素源、緩衝液濃度、培養期間、供給スケジュール、および通気が挙げられる。例えば、複数の細胞におけるまたは複数の細胞を使用した分子最適化の場合、複数の細胞は、細胞を反応容器に分配する前に規定の発酵条件のセットの下で培養することができ、または細胞は、反応容器内で規定の発酵条件で培養することができる。発酵条件は、反応容器のサブセットまたはすべてにわたって同じであっても異なっていてもよい。別の例では、複数の反応容器（必要に応じて細胞を含む）は、複数の反応容器の性能または所望の特性（例えば、標的タンパク質発現レベル、酵素活性レベル）に影響を及ぼし得る処理（例えば、薬物または小分子処理、基質の存在、様々な反応条件など）を含む適用条件に供することができる。適用条件を変化させ、分子の性能への影響を観察することは、環境の変動が構造または機能にどのように影響するかを理解するのに有用である可能性があり、および／または所与の適用条件下で所望の特性を有する新規または最適化された分子（例えば、標的タンパク質または他の生体分子）の発見にさらに使用され得る。

【０２０３】
いくつかの例では、分子最適化プロセスは、細胞を含まず、異なる分子をスクリーニングし、それらを１またはそれを超える適用条件に供すること、および所望の特性を有する分子バリアントを単離することを含み得る。例えば、所望の分子バリアントは、特定のタンパク質－タンパク質相互作用または代謝活性を示すものであってもよい。本明細書に記載のプロセスを使用して多数の分子をスクリーニングすることができ、機能的特性を示す所望の分子バリアントを単離し、さらなる特性評価に供することができる。異なる発見分子のさらなる特性評価（例えば、化学分析または生化学分析）としては、非限定的な例では、質量分析法、重量測定、原子分光法、核磁気共鳴法（ＮＭＲ）、クロマトグラフィー、プラズマ質量分析法、Ｘ線光電子分光法（ＸＰＳ）、オージェ電子分光法などの化学分析または元素分析を挙げることができる。得られたデータ（例えば、化学組成、物理的特性など）を使用して、どの構造要素（例えば、元素、化学組成など）が所望の機能的特性を有するか、または改善された機能的性能（例えば、タンパク質－タンパク質相互作用または代謝活性）を有し得るかを（例えば、コンピュータ可読媒体を使用して）予測することができる。

【０２０４】
分子最適化ユニットの異なる分子バリアント（例えば、標的タンパク質バリアント）は、ライブラリ（例えば、タンパク質ライブラリ、ペプチドライブラリ、生体分子ライブラリ）に含まれてもよく、または異なる発見分子は、ゲノムバリアントライブラリにおいてコードされてもよく、これは、いくつかの例では生成または合成される遺伝的バリアントを含む。例えば、生成または合成された遺伝的バリアントの場合、遺伝的バリアントは、突然変異誘発（ランダム突然変異誘発（例えば、紫外線、Ｘ線照射、ガンマ線照射などの突然変異原、化学変異原、例えばエチルメチルスルホネート、活性酸素種、脱アミノ化剤、多環芳香族炭化水素、アルキル化剤（例えば、エチルニトロソウレア、ニトロソアミン）芳香族アミン、アルカロイド、臭素、アジ化ナトリウム、ソラレン、ベンゼン、金属（例えば、ヒ素、カドミウム、クロム、ニッケル）または生物学的薬剤（例えば、トランスポゾン、ウイルス、細菌）などへの曝露）、飽和突然変異誘発、循環置換）、多重自動ゲノム工学法（ＭＡＧＥ）、ＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、ＣＲＩＳＰＲガイドＤＮＡポリメラーゼ、エラープローン（ｅｒｒｏｒ－ｐｒｏｎｅ）ポリメラーゼ連鎖反応（ｅｐＰ

ＣＲ）、または他のゲノム工学的アプローチによって生成され得る。いくつかの例では、ゲノムバリアントライブラリは、全ゲノム突然変異誘発ライブラリ、ゲノムシャッフルライブラリ、プロモータースワップライブラリなどの標的ゲノムライブラリ、およびトランスポゾンライブラリ（例えば、プロモーター挿入ライブラリ）であるか、またはそれらを含む。いくつかの例では、標的ゲノムライブラリは、プロモーター、５’非翻訳領域、３’非翻訳領域、リボザイム、ＲＮＡ安定性配列、分泌ペプチド、融合タンパク質、ＤＮＡ結合ドメイン、タンパク質－タンパク質結合領域、コドン最適化遺伝子、コドンランダム化遺伝子、ターミネーターおよび非相同末端結合からなる群から選択される１またはそれを超えるメンバーを含むか、またはそれらをコードする。いくつかの例では、ゲノムバリアントライブラリは、単一遺伝子ランダム突然変異ライブラリ、部位飽和ライブラリ、小型インサートメタゲノムライブラリ、大型インサートメタゲノムライブであってもよい。いくつかの例では、異なる発見分子はＤＮＡコード化分子である。ＤＮＡコード化分子は、メタゲノムライブラリ、ＤＮＡコード化ライブラリ、１またはそれを超える生物からのライブラリ、またはファンクションジェネレータによって作成されたライブラリなどのライブラリに含まれ得る。ＤＮＡコード化分子ライブラリは、ランダムに生成されたＤＮＡコード化分子、天然に存在するＤＮＡコード化分子、またはそれらの組み合わせを含み得る。

【０２０５】

株最適化ユニットと同様に、本明細書に記載のシステムおよび方法を使用して複数の実験条件を同時にスクリーニングすることができる。例えば、分子最適化ユニットは、同じまたは異なる適用条件に供された異なる分子バリアント（例えば、標的タンパク質バリアント）を含む数百万の反応体積を同時にスクリーニングまたは分析するように構成することができる。有益なことに、分子最適化ユニットは、分子バリアントおよび分子バリアントの環境（または適用）条件に関する多重化されたマルチパラメトリック情報を提供し得る。さらに、分子最適化ユニットは、所望の性能基準を有する分子（例えば、ＤＮＡによってコードされるタンパク質標的）、例えば閾値酵素活性レベルまたは他の代謝もしくは表現型形質または特定の適用条件下での標的タンパク質の産生などの所望の特性を有する分子をスクリーニングするのに有益であり得る。

【０２０６】

反応容器：分子最適化ユニットの反応容器は、株最適化ユニットおよび分子発見ユニットの反応容器と同じであっても異なっていてもよい。分子発見ユニットの反応容器は、例えば、液滴またはプラグ、ウェル（例えば、マイクロウェル、ナノウェル、ピコウェルなど）、バイアル、チューブ、フラスコ、または他の容器であってもよい。いくつかの例では、反応容器は、マイクロ流体デバイスに含まれ、このマイクロ流体デバイスは、消費ユニットを含むかまたは消費ユニットの一部である。いくつかの例では、反応容器は、例えばマイクロウェルアレイまたはプレート、ナノウェルアレイまたはプレートなどのアレイのメンバーである。本明細書に記載するプロセスのいずれかのために、同時または連続のいずれかで、複数のタイプの反応容器が使用され得る（例えば、液滴およびウェル）。反応容器は、物理的に別個のものであってもまたは別々のものであってもよく、または反応容器は、合流または合体することが可能であってもよい。

【０２０７】

各反応容器は、スクリーニングのための異なる分子（例えば、標的タンパク質バリアント）を含むことができる。例えば、株最適化ユニットと同様に、所望の機能的特性（例えば、生体分子（例えば、酵素）活性、生体分子相互作用、表現型または代謝形質）について異なるタンパク質をコードする遺伝的バリアントのライブラリをスクリーニングすることが有用であってもよく、したがって、各反応容器は、スクリーニングのために異なる遺伝的バリアントを含有するように構成することができる。場合によっては、各反応容器は、遺伝的バリアントを有する１またはそれを超える細胞を含む。例えば、各反応容器は、他の反応容器中の他の細胞の遺伝的バリアントとは異なる単一の遺伝的バリアントを有する単一の細胞または複数の細胞を含み得る。代替的または追加的に、各反応容器は、単一

の遺伝的バリアントを有する細胞のクローン集団を含み得る。他の例では、単一の反応容器内で異なる遺伝的バリアントをスクリーニングすることが有用であってもよく、したがって、反応容器は、異なる遺伝的バリアントを有する細胞の混合物を含み得る。そのような例では、各反応容器は、細胞と遺伝的バリアントとの組み合わせを含み得る。

【０２０８】

本明細書に記載するように、反応容器は、単一の細胞または複数の細胞を含み得、ハイスループット方式で評価されるように構成することができる。所与の期間内に任意の数の反応容器を生成または使用することができる。いくつかの例では、異なる数の反応容器（または体積）をスクリーニングすることが有益であり得る。したがって、使用または生成される反応容器（または体積）の数は、適切な用途または実験目的に従って調整することができる。例えば、多数の別々の反応体積が、例えばスクリーニング目的（例えば、分子発見のために）に有用である場合、少なくとも１０，０００反応体積、少なくとも１００，０００反応体積、少なくとも１，０００，０００反応体積、少なくとも１０，０００，０００反応体積、少なくとも１００，０００，０００反応体積またはそれを超える反応体積が使用され得る。他の例では、より少ない数の別々の反応体積が有用または必要とされ得る（例えば、標的スクリーニングのために、またはスクリーニングのための限られた数の種を用いた実験の合理的な設計のために）。そのような場合、最大で１００，０００，０００反応体積、最大で１０，０００，０００反応体積、最大で１，０００，０００反応体積、最大で１００，０００反応体積、最大で１０，０００反応体積、最大で１，０００反応体積またはそれ未満の反応体積が使用または生成され得る。

【０２０９】

株最適化ユニットと同様に、分子発見ユニットの反応容器は、マイクロ流体デバイス内に含まれてもよく、任意の有用な体積を収容してもよい。液滴の場合、上述のように、液滴は任意の有用な組成物を含み得る。

【０２１０】

アッセイ：いくつかの例では、生物学的バリアントの選別は、１またはそれを超えるスクリーニング操作の使用を含み得る。スクリーニング操作は、アッセイ、例えば、直接アッセイ、間接アッセイ、結合アッセイ、レポーターアッセイ（インビトロまたはインビボであってもよい）、蛍光発生アッセイ、発色アッセイまたは比色アッセイなどを含み得る。酵素結合免疫吸着アッセイ（ＥＬＩＳＡ）、サンドイッチＥＬＩＳＡ、近接ライゲーションアッセイなどを含むがこれらに限定されない１またはそれを超える免疫学的アッセイを使用することができる。いくつかの例では、レポーター分子または選択分子がスクリーニングのために使用される。アッセイ分子またはレポーター分子または選択分子は、生物学的バリアントの遺伝的バリアントまたは機能的特性（例えば、表現型または代謝形質）の存在の検出に有用であり得る。例えば、所望の生物学的バリアントが酵素活性を含む細胞である場合、反応容器内の細胞は、所望の酵素活性によって基質が生成物に変換されると色が変化するかまたは蛍光を発する発色性または蛍光発生性基質の存在下でスクリーニングされ得る。そのような例では、色または蛍光の変化を示す反応容器は、細胞が所望の酵素活性を含み、続いて、その細胞を単離し、さらに分析して（例えば、シーケンシング、タンパク質分析などを介して）、生物学的バリアントに関する情報（例えば、ＤＮＡまたはＲＮＡ配列、タンパク質配列または組成物など）を決定することができることを示し得る。他の例では、所望の生物学的バリアントは、選択分子またはレポーター分子を含み得、例えば、選択分子の場合、所望の形質を有する細胞のみが所与の培養条件で生存し得る（例えば、抗生物質耐性を包含する導入遺伝子または組換え遺伝子を有する細胞のみが、抗生物質を含む増殖培地中で生存する）。生存細胞は、所望の機能的特性についてスクリーニングされ得、所望の機能的特性を有するものは、さらなる分析に供することができる。さらに別の例では、所望の形質を有する細胞のみが、導入遺伝子または組換え遺伝子が存在することを示すレポーター分子（例えば、蛍光性または検出可能なタンパク質）を発現し得る。そのような例では、望ましい形質または機能的特性を示す細胞は、本明細書に記載するように、反応容器から単離され、さらなる分析に供することができる。他の例

では、アッセイ（例えば、イムノアッセイ、結合アッセイなど）を使用して、各遺伝的バリアント中の目的の分子（例えば、タンパク質、核酸分子、脂質、炭水化物、代謝産物、またはそれらの組み合わせ）の存在を決定することができ、目的の分子を含む遺伝的バリアントを単離し、さらに分析することができる。

【０２１１】

選択された生物学的バリアント（例えば、所望の特性を示す選択された細胞）に対してさらなる分析を行うことができる。そのような分析としては、本明細書に記載するように、タンパク質分析（例えば、質量分析、タンパク質アッセイ）、核酸分析（例えば、シーケンシング）、脂質分析（例えば、ラマン散乱または分光法）、化学組成分析（例えば、ＮＭＲ、ＸＰＳ、クロマトグラフィーなど）、画像化（および画像分析）、分光法、または他の分析方法もしくはスクリーニングを挙げることができるが、これらに限定されない。そのような分析方法は、例えば、ＤＮＡ配列または遺伝子、アミノ酸配列、タンパク質特性（例えば、翻訳後修飾）に関する情報を得ることによって、遺伝的バリアントの同一性を決定するのに有用であり得る。次いで、そのような情報を収集してデータベースに保存するか、または必要に応じて所望の生物学的活性、表現型もしくは代謝形質などの任意の機能的特性を有する分子発見ライブラリと再統合することができる。

【０２１２】

いくつかの例では、分子最適化ユニットは、検出器を備えるか、または検出器に結合されてもよい。検出器は、細胞または反応容器における目的の分子または機能的特性の有無を決定するために使用され得、所望の生物学的バリアントのスクリーニングおよび／または選択を可能にし得る。検出器は、１またはそれを超える分析機器、例えば質量分析計、ラマン分光計、近赤外（ＮＩＲ）分光光度計、フーリエ変換赤外（ＦＴＩＲ）分光計、分光光度計（例えば、蛍光または吸光度プレートまたはキュベットリーダ）、圧電センサなどを含み得る。検出器は、細胞を計数する（例えば、セルカウンタ、フローサイトメータ）ように構成することができる。検出器は、１またはそれを超える光学構成要素を備えてもよく、いくつかの例では、顕微鏡（例えば、立体鏡、複合顕微鏡、蛍光顕微鏡、偏光顕微鏡、共焦点顕微鏡、微分干渉顕微鏡、超解像顕微鏡など）、またはそれらの構成要素もしくは組み合わせを備えてもよい。検出器は、磁気、インピーダンス、トポグラフィまたは幾何学的形状、散乱、形態、吸光度、蛍光、色、細胞のクラスター化、細胞またはコロニーサイズなどを含むがこれらに限定されない、細胞または反応容器の物理的特性を検出するために使用され得る。検出器は、点検出器またはアレイもしくは画像検出器であってもよい。

【０２１３】

移動アクチュエータ：いくつかの例では、反応容器移動アクチュエータを使用して（例えば、検出器に向かって）反応容器を移動させる。容器移動アクチュエータは、ロボットマニピュレータ、フロージェネレータ、音響ドロップジェネレータ、光ピンセット、熱ドロップオンデマンド、圧電ドロップオンデマンド、リソグラフィユニット、デジタルマイクロ流体デバイス、またはそれらの組み合わせであってもよい。移動アクチュエータは、ステージと、例えば、ねじ、歯車、ねじ山、モータ、ポンプ、圧縮機、抵抗器、トランジスタ、ばね、滑車などの任意の必要な移動部品とを備えることができる。いくつかの例では、反応容器は、移動アクチュエータに接続された基材上または基材に隣接して配置される。

【０２１４】

ライブラリ：分子最適化ユニットは、分子バリアント（例えば、標的タンパク質バリアント）ライブラリを含み得、これは、本明細書に記載のプロセスで使用され得る。分子最適化ライブラリは、本明細書に記載するように、複数の異なる遺伝的バリアントまたはタンパク質バリアントを含むことができる。例えば、ライブラリは、異なる遺伝的バリアントによってコードされる複数の標的タンパク質バリアントを含み得る。複数の遺伝的バリアントは、第１の株に由来し得る。例えば、第１の細胞株（例えば、細菌、酵母、動物、植物など）の細胞は、遺伝的バリアントの各々が異なるように、複数の異なる遺伝的バリ

アントを生成するのに十分な条件に供することができる。複数の遺伝的バリアントの各遺伝的バリアントは、例えば、複数の遺伝的バリアントの他の遺伝的バリアントとは異なる配列を有するＤＮＡまたはＲＮＡを含み得る。複数の異なる遺伝的バリアントは、同じ生体分子（例えば、ＤＮＡ、ＲＮＡ、ペプチド、タンパク質など）の変異をコードし得る。例えば、複数の異なる遺伝的バリアントは、後にタンパク質に翻訳されるＲＮＡ配列の変異をコードする異なるＤＮＡ配列を含み得る。ライブラリ中の異なるＤＮＡ配列の変異は、異なるＲＮＡ配列、したがってタンパク質もたらすことができる。いくつかの例では、遺伝的バリアント（ＤＮＡおよび／またはＲＮＡの変異を有し得る）によってコードされるタンパク質は酵素である。そのような場合、酵素は、代謝経路を形成し得るか、またはその一部であってもよい。酵素または代謝活性は、分子最適化プロセス中のスクリーニングパラメータまたは性能要件として使用され得る。

【０２１５】

場合によっては、分子最適化ユニットは、分子バリアントをスクリーニングし、機能的特性（例えば、表現型または代謝形質または活性）または測定に基づいて所望の生物学的バリアント（例えば、分子バリアント）を選択および／または分類するように構成される。表現型または代謝測定は、非限定的な例では、タンパク質または酵素活性、タンパク質発現レベル、副産物または代謝産物の数量などを含むことができる。例えば、所望の生物学的バリアントは、規定のまたは最小閾値の酵素活性を有するものであってもよい。そのような場合、分子ライブラリの分子バリアント（例えば、標的タンパク質バリアントライブラリ）を分子発見ユニットでスクリーニングすることができ、規定のまたは最小閾値の酵素活性を有する分子バリアントのみをさらなる分析のために選択する。分子バリアントをさらなる分析（例えば、シーケンシング、タンパク質分析、化学分析）に供して、目的の酵素の核酸もしくはタンパク質配列または化学組成を決定することができる。いくつかの例では、分子バリアントの配列または組成をデータベースに追加し、および／または機能的特性または性能を有する分子最適化ライブラリに再統合する。いくつかの例では、分子バリアントの配列または組成をコンピュータ可読媒体で使用して、所与の一連の条件（例えば、適用条件）および構造（例えば、配列、化学組成など）に基づいて分子（例えば、標的タンパク質バリアント）の性能に関する予測を生成する。

【０２１６】

分子最適化ライブラリは、所望の機能的特性を有するかまたはそれに影響する（例えば、タンパク質－タンパク質相互作用またはシグナル伝達経路を介して）分子、例えば標的タンパク質をスクリーニングするのに有用な任意の数の分子バリアントを含み得る。分子最適化ライブラリは、少なくとも１００個の異なる分子、少なくとも１，０００個の異なる分子、少なくとも１０，０００個の異なる分子、少なくとも１００，０００個の異なる分子、少なくとも１，０００，０００個の異なる分子、少なくとも１０，０００，０００個の異なる分子、少なくとも１００，０００，０００個の異なる分子またはそれを超える個数の異なる分子を含み得る。本明細書で説明するように、分子最適化プロセスは、再帰的または反復的であってもよい。そのような場合、反復ごとにスクリーニングされる異なる発見分子の数は様々であってもよく、例えば、最初のサイクルは１０，０００，０００個の異なる分子をスクリーニングし、２番目のサイクルは１００，０００個の異なる分子をスクリーニングしてもよい。逆に、異なる発見分子の数は、例えば、新しい生物学的多様性またはバリアント（例えば、新しい株、新しい異なる発見分子など）および／または条件付き多様性（例えば、新たに導入されたまたはスクリーニングされた生物学的バリアントの一部または全部に対する様々な発酵条件または適用条件）の導入によって、サイクル間で増加し得る。

【０２１７】

標的タンパク質：分子バリアントは、標的タンパク質をコードする遺伝的バリアントを含み得る。標的タンパク質は、構造タンパク質、酵素、表面受容体タンパク質、ペプチドホルモン、治療用ペプチド、免疫系成分（例えば、抗体、抗体フラグメント、抗体コンジュゲート）、生物活性ペプチド（例えば、ランチペプチド、ラッソペプチド、非リボソー

ム合成ペプチド）、または他の有用なタンパク質もしくはペプチドなどの任意の有用なタンパク質またはペプチドであってもよい。標的タンパク質は、ポリマー分解酵素（例えば、ポリエチレンテレフタレートヒドロラーゼ）、リポザイム、パラターゼ、セルラーゼ、アミラーゼ、レジナーゼ、アミダーゼ、イソメラーゼなどの工業用酵素であってもよい。標的タンパク質は、リパーゼ、ＤＮＡｓｅ、Ｌ－アスパラギナーゼ、Ｌ－グルタミナーゼ、ウロキナーゼ、ウリカーゼ、コラゲナーゼ、グルコシダーゼ、ガラクトシダーゼ、チロシナーゼなどの治療用酵素であってもよい。標的タンパク質は、抗体、抗体フラグメント、または他の結合タンパク質などの治療用タンパク質であってもよい。標的タンパク質は、アミリン、アディポネクチン、副腎皮質刺激ホルモン、アンジオテンシン、抗利尿ホルモン、ナトリウム利尿ペプチド、カルシトニン、コレシストキニン、コルチコトロピン放出ホルモン、コルチスタチン、エンケファリン、エンドセリン、エリスロポエチン、卵胞刺激ホルモン、ガラニン、胃抑制ポリペプチド、ガストリン、グレリン、グルカゴン、放出ホルモン、ヘプシジン、ヒト絨毛性ゴナドトロピン、ヒト胎盤ラクトゲン、成長ホルモン、インヒビン、インスリン、レプチン、リポトロピン、黄体形成ホルモン、刺激ホルモン、モチリン、オレキシン、オステオカルシン、オキシトシン、膵臓ポリペプチド、副甲状腺ホルモン、下垂体アデニル酸シクラーゼ活性化ペプチド、プロラクチン、リラキシン、レニン、セクレチン、ソマトスタチン、トロンボポエチン、甲状腺刺激ホルモン、甲状腺刺激ホルモン放出ホルモン、血管作用性腸ペプチド、グアニリン、ウログアニリン、または他のペプチドホルモンなどのペプチドホルモンであってもよい。標的タンパク質は、アクチン、アクチニン、アグリカン、ビグリカン、カドヘリン、クラスリン、コラーゲン、デコリン、弾性、フィブリン、フィブリノゲン、フィブロネクチン、ヘパリン、ケラチン、ラミニン、ムチン、ミエリン、ミオシン、スペクトリン、トロポミオシン、トロポニン、チューブリン、ビメンチン、ビトロネクチンなどの構造タンパク質であってもよい。標的タンパク質は、ポリメラーゼ（例えば、ＲＮＡポリメラーゼ）、リガーゼ、ヘリカーゼ、リボザイム、リボソーム、スプライセオソームタンパク質、シャペロンタンパク質などのＤＮＡ複製、転写または翻訳に関与するタンパク質または酵素であってもよい。標的タンパク質は、シグナル伝達因子またはサイトカイン（例えば、インターロイキン、インターフェロン、腫瘍壊死因子、トランスフォーミング増殖因子、免疫調節因子など）であってもよい。標的タンパク質は、食品ベースまたは栄養タンパク質、例えばオボアルブミン、ヘモグロビンであってもよい。

【０２１８】

データセット：分子最適化ユニットの１またはそれを超える構成要素は、以下にさらに詳細に記載するように、コンピュータ可読媒体を含み得る。いくつかの例では、コンピュータ可読媒体は、データを入力し、構造的態様（例えば、配列、ペプチドまたはタンパク質モチーフ、化学組成など）の機能的性能に関する予測を生成するように構成されたコンピュータモデルである。コンピュータ可読媒体は、複数のデータセットを入力するように構成することができ、このデータセットは、データベースまたは実験モジュールから、例えば統合ユニットからのものであってもよい。例えば、コンピュータ可読媒体は、機能データ、配列データ、および／または濃縮データ（例えば、ポリマーバリアント）を入力し、改善された機能を有すると予想される設計分子を出力する機械学習アルゴリズムを含み得る。コンピュータ可読媒体はまた、データセットを記憶すること、記憶されるデータセットの閾値を設定すること、自動処理のために１またはそれを超える統合ユニットと通信することなど、様々な他の機能を実行するように構成することができる。

【０２１９】

いくつかの例では、機械学習アルゴリズムが使用される場合、機械アルゴリズムは、教師あり学習アルゴリズム、教師なし学習アルゴリズム、または強化学習アルゴリズムを含むことができる。アルゴリズムは、線形回帰、ロジスティック回帰、決定樹、サポートベクターマシン（ＳＶＭ）、ナイーブベイズ、ｋ近傍法、ｋ平均法、ランダムフォレスト、次元削減アルゴリズム、勾配ブースティングアルゴリズム、ＸＧＢｏｏｓｔ、ライト勾配ブースティング（ｌｉｇｈｔ　ｇｒａｄｉｅｎｔ　ｂｏｏｓｔｉｎｇ）アルゴリズム、Ｃ

ａｔｂｏｏｓｔなど、またはそれらの組み合わせを含み得る。機械学習アルゴリズムは、自然言語モデルと、１またはそれを超えるニューラルネットワークとを含むことができる。自然言語モデルは、構文解析または意味解析などのコンピュータ実装のための１またはそれを超える技法を適用することができる。例えば、構文解析を使用することができ、トークン化、品詞タグ付け、見出語認定およびステミング、ストップワード除去などのサブタスクを含むことができる。意味解析は、語義の曖昧性解消、関係抽出などを含むことができる。

【０２２０】

いくつかの例では、分子最適化ユニットの実験系は、他の統合ユニットと統合されてもよい。例えば、分子最適化ユニットは、最適な分子または遺伝的バリアントに関する予測を生成するコンピュータ可読媒体を含み得、最適な分子または遺伝的バリアントは、株最適化ユニット内の株最適化ライブラリに導入され得る。したがって、以下に説明するように、相互作用する統合ユニットによって異なる機能を実行することができる。

【０２２１】

（ｉｖ）統合システム

コンピュータ可読媒体：いくつかの例では、１またはそれを超える統合ユニットは、コンピュータ可読媒体を介して統合されるか、またはコンピュータ可読媒体に結合される。コンピュータ可読媒体は、非限定的な例では、統合ユニットのいずれかからのデータセットをデータリポジトリに格納すること、データリポジトリに配信されるデータの閾値パラメータを設定すること、および入力多様性スクリーニングを繰り返すようにユニットのいずれかに命令することを含む、１またはそれを超える機能を実行することが可能であり得る。例えば、株最適化ユニットからの配列情報および機能的性能などのデータをコンピュータ可読媒体に統合することができ、これを使用して分子最適化の予測（例えば、閾値基準を超えて機能すると予測される遺伝的バリアントを提供する）を生成することができる。別の例では、分子発見ユニットからのデータをコンピュータ可読媒体に統合することができ、これを使用して、所与の発酵または適用条件下で機能的性能を向上させると予測される分子バリアントに関する予測を生成することができる。

【０２２２】

いくつかの例では、コンピュータ可読媒体は品質管理機能を含む。例えば、コンピュータ可読媒体は、品質閾値を超えるデータを（例えば、データベースまたはデータリポジトリ内に）格納するように構成することができる。したがって、コンピュータ可読媒体は、データの閾値パラメータを設定し、入力データを閾値にかけることができる。閾値を守るデータ（例えば、閾値を超えている）は、データリポジトリに配信される。閾値パラメータの非限定的な例としては、シグナル対ノイズ比、統計的有意性、標的性能基準、例えば酵素活性レベル、タンパク質または分子相互作用、表現型または代謝形質の存在などが挙げられる。

【０２２３】

システム自動化：いくつかの例では、コンピュータ可読媒体は、統合ユニットのうちの１またはそれを超えるユニットにおける１またはそれを超えるプロセスを自動化するように構成される。そのような場合、１またはそれを超える統合ユニットは、コンピュータ可読媒体の通信インターフェースと通信する通信インターフェース（例えば、無線または有線の通信インターフェース）を備えることができる。コンピュータ可読媒体は、（例えば、通信インターフェースを介した）コマンドまたは命令を統合ユニット（またはその構成要素）に指示することができる。例えば、コンピュータ可読媒体は、１またはそれを超えるコマンドまたは命令を中継することによって統合ユニットの１つまたはすべての反応運動アクチュエータ（例えば、流体ハンドリングユニット、フロージェネレータなど）を制御し、それによって反応容器の動きを制御するように構成することができる。同様に、コンピュータ可読媒体は、例えば、流体ハンドリングユニット、フロージェネレータ、移動アクチュエータなどを使用して、反応容器への試薬（例えば、培養培地、反応試薬または検出試薬）の供給を制御するように構成することができる。反応運動アクチュエータ、コ

ンピュータ可読媒体、および検出器はすべて統合されてもよく、いくつかの例では、フィードバックループを介して接続されてもよい。

【０２２４】

一例では、フィードバックループを使用して各反応容器内の体積（例えば、試薬体積）を制御することができ、これは各反応容器内の細胞間の一貫した増殖を確実にするのに有益であり得る。フィードバックループを使用して、各反応容器の条件を監視し、一連の所定の条件、例えば温度、ｐＨ、培養液成分などに基づいて条件を自動的に調整（例えば、フィードバックコントローラを介して）することができる。同様に、フィードバックループは、検出または画像認識ソフトウェアを含むことができ、このソフトウェアは、所与の反応容器が所望の数の細胞を含むかどうか、正しい体積、流量などを指定するために使用することができる。液滴が使用される場合、フィードバックループは、液滴生成の頻度を維持するための流量調整器を含むことができる。所望の数の細胞を有しない反応容器は、さらなるスクリーニングまたは分析から除外され得る。したがって、コンピュータ可読媒体は、そのようなフィードバックに基づいて、細胞または反応容器がさらにスクリーニングまたは分析されるかどうかについて決定を下すことができる。

【０２２５】

いくつかの例では、フィードバックループを本明細書に記載のスクリーニングプロセスで使用することができる。反応容器移動アクチュエータは、所望の生物学的バリアントを選別し、それらを単離するように構成することができる（例えば、蛍光または比色活性化細胞選別、磁気選別、誘電泳動などを使用したマイクロ流体デバイスにおける液滴の選別）。そのような例では、検出器は各反応容器からシグナルまたは画像を収集することができ、シグナルまたは画像は（例えば、コンピュータ可読媒体を介して）処理されてもよく、機能的性能を示す反応容器（例えば、閾値を超える蛍光または比色シグナル）は自動的に（例えば、別のチャンバ、容器、または装置の一部に）ゲーティングおよび／または選別することができる。したがって、反応運動アクチュエータ（または選別機）、検出器（シグナルを検出する）、およびコンピュータ可読媒体を一体化することができる。いくつかの例では、反応容器および／または細胞の量または割合、例えば固定割合（例えば、ユーザ規定のパーセンテージの上位）を選別することができる。いくつかの例では、細胞または反応容器の上位パーセンテージ（例えば、最も高い機能的性能）のみを収集およびさらなる分析のために選別することができる。例えば、細胞または反応器の上位５０％、４０％、３０％、３０％、２５％、１５％、１０％、５％、１％、０．５％、０．１％、０．０５％、０．０２５％、０．０１％などのみを収集のために選別することができる。いくつかの例では、反応運動アクチュエータは、フィードバックループを使用してコンピュータ可読媒体の出力に基づいて反応容器の運動を一時停止、遅延、または他の方法で変更するように構成することができる。例えば、エラーが検出器によって検出された場合（例えば、検出不能なシグナル、妨害の疑いなど）、コンピュータ可読媒体は、エラーを自動的に検出し、反応容器の移動を一時停止、遅延、または変更するように移動アクチュエータに指示することができる。

【０２２６】

同様に、コンピュータ可読媒体は、追加の試薬（例えば、検出試薬、または他の反応試薬、例えばｐＨ調整剤、イオン強度調整剤、イオン、基質、阻害剤、酵素、競合剤、酸素、二酸化炭素など）を反応容器に自動的に導入するように１またはそれを超えるシステム構成要素（例えば、流体ハンドリングユニット、フロージェネレータなど）に指示することができる。複数のシステム構成要素のそのような統合は、指定された時間枠内での検出、分析、選別などと共に、正確な規定の反応時間を可能にし得る。例えば、検出試薬（例えば、酵素活性レベルを測定するための酵素アッセイのための基質）の添加は、各反応容器が検出前に同じ期間、ある量の検出試薬と共にインキュベートされるように、コンピュータ可読媒体によって指示することができる。インキュベーション時間は読み出しに影響を与える可能性があるので（例えば、より長いインキュベーション時間は、より多くの基質が生成物に変換されるにつれてシグナル強度を増加させる）、添加される検出試薬の量

および／またはインキュベーション時間を正確に制御することにより、例えば、より長期間インキュベートされ、必ずしもより高い機能的性能を示さない反応容器からの誤った読み出しを防止することができる。

【０２２７】

したがって、本明細書に記載の方法およびシステムは、反応容器全体での変動が少ない正確なシステム制御を可能にすることができる。例えば、コンピュータ可読媒体は、１またはそれを超えるプロセッサによって実行されると、システムに、細胞の集団を複数の反応容器または反応体積に分配させて分配された細胞を生成させ、分配された細胞の表現型または機能的性能についてアッセイして、分配された細胞についてのシグナルを生成させる命令を含むことができる。生成されたシグナルは、反応容器にわたって低い変動係数を有し得る。例えば、所与の生物学的バリアント（例えば、同じＤＮＡ配列を含む株、均一な細胞集団など）について、変動係数（例えば、シグナル出力によって測定されるように、生成物の数量または活動レベルを表すことができる）は、約５００％未満、約４００％未満、約３００％未満、約２００％未満、約１００％未満、約９０％未満、約８０％未満、約７０％未満、約６０％未満、約５０％未満、約４０％未満、約３０％未満、約２０％未満、約１０％未満、約９％未満、約８％未満、約７％未満、約６％未満、約５％未満、約４％未満、約３％未満、約２％未満、または約１％未満またはそれ未満の変動係数であってもよい。いくつかの例では、変動係数は、例えば、１％～約１０％、または約２％～約８％、または約１０％～約１００％の値の範囲内にあってもよい。そのような変動係数は、システム内の異なるノイズ特性、例えば、生物学的変動（例えば、同一の株にわたって生物学において天然に生じる性能差）、試薬添加およびインキュベーション時間、光学／検出ノイズなどを説明することができる。いくつかの例では、本明細書に記載の株最適化、分子最適化、または分子発見ユニットは、約１０％未満の変動係数を含み得、本明細書で使用されるシステムおよび方法が、技術的ノイズまたは変動が低いバリアントのライブラリを正確に特徴付けるのに有利であり得ることを示している。

【０２２８】

いくつかの例では、コンピュータ可読媒体は、システム内のドリフトを補正することができる比例・積分・微分（ＰＩＤ）アルゴリズムを含むことができる。例えば、ＰＩＤアルゴリズムは、機械的機器、液体粘度、または温度の偏差に起因するドリフトを補正することができる。ドリフトの補正は、ドリフトによって引き起こされるシステムによる誤った選別決定を防止するのに有利であり得る。そのような補正は、例えば、反応容器または細胞を選別するために使用される閾値ゲートの変更、または選別された細胞の標的パーセンテージ（例えば、最も高い機能の細胞の上位パーセンテージ）の調節を含むことができる。いくつかの例では、コンピュータ可読媒体は、比例コントローラまたは比例積分コントローラを備えてもよい。

【０２２９】

いくつかの例では、コンピュータ可読媒体は、株最適化プロセス、分子発見プロセス、または分子最適化プロセスの実行に関連する１またはそれを超えるコマンドを実行するように構成することができる。コンピュータ可読媒体は、本明細書に記載の操作の１またはそれを超える操作を実行するように、ユニットのそれぞれに個別にまたは集合的に命令することができる。例えば、コンピュータ可読ユニットは、株最適化ユニットに、所与の実験条件下で生物学的条件のスクリーニングを実行するように（例えば、一連の生物学的または遺伝的バリアント、適用または発酵条件などを使用して）、実験条件を再実行するように、または多様性スクリーニングを繰り返すように（例えば、遺伝的バリアント、適用もしくは発酵条件、またはそれらの組み合わせを追加した実験の再実行）指示することができる。

【０２３０】

いくつかの例では、コンピュータ可読媒体は、１またはそれを超える標的パラメータを入力するように構成することができる。標的パラメータは、例えば、酵素活性、分子またはタンパク質相互作用、結合親和性などの性能閾値または所望の機能的性能を含むことが

できる。そのような例では、コンピュータ可読媒体は、目標パラメータに達するまで、株最適化プロセス、分子発見プロセス、または分子最適化プロセスの実行に関連する１またはそれを超えるコマンドを実行するように構成することができる。１つの非限定的な例では、所望のレベルの酵素活性の目標パラメータを株最適化ユニットに入力することができる。次いで、コンピュータ可読媒体は、酵素活性の目標閾値が満たされるまで、必要に応じて（例えば、新しい特異的または非特異的な遺伝的バリアントまたは発酵条件を導入することによって）追加の生物学的多様性または条件付き多様性を入力し、（例えば、蛍光アッセイを使用して）酵素活性をスクリーニングして、株最適化プロセスを反復的および／または自動的に実行するように構成することができる。例えば、コンピュータ可読媒体は、例えば、既知または経験的に導出された構造的（例えば、配列）－機能的関係間の相関を使用して、既知の遺伝子座（複数可）で細胞に１またはそれを超える点ベースまたは特異的突然変異を自動的に導入するように構成することができる。別の例では、コンピュータ可読媒体は、規定の遺伝子座または未規定の遺伝子座のいずれかにおいて、細胞にランダム突然変異を自動的に導入するように構成することができる。さらなるスクリーニングを行うことができ、標的閾値を満たす株を単離し、その株またはタンパク質を分析することができる。

【０２３１】

標的性能パラメータ：生物学的バリアントまたは分子バリアントをスクリーニングするために、１またはそれを超える標的性能パラメータをシステム（例えば、統合ユニットのうちの１またはそれを超える統合ユニット）に入力することができる。標的性能パラメータは、生物学的バリアントまたは分子バリアントの所望の機能的特性に関連し得る。例えば、所望の機能的特性は、所望の分子（例えば、タンパク質またはペプチド）の収率、細胞の成長または増殖速度、条件（例えば、ｐＨ、温度、例えばカルシウム、マグネシウムなどの二価カチオンなどのイオン濃度などの適用または発酵条件）に対する細胞の感受性、細胞生存率、力価、分子相互作用（例えば、タンパク質－タンパク質相互作用）、安定性（例えば、熱安定性、プロテアーゼ分解に対する耐性など）、結合特性（例えば、別の分子に対する親和性または結合活性、基質特異性）、代謝または表現型形質（例えば、酵素活性、触媒性能）、タンパク質の発現レベル、副産物または代謝産物の数量、標的タンパク質のフォールディングの存在または程度、標的タンパク質対副産物比、標的タンパク質の修飾（例えば、グリコシル化、リン酸化、ＳＵＭＯ化、アシル化、ユビキチン化、アセチル化、脂質化、メチル化、ヒドロキシル化、または他の翻訳後修飾）、または他の機能的特性であってもよい。いくつかの例では、標的性能パラメータは、条件パラメータを条件として選択されてもよい。例えば、所望の生物学的バリアントは、指定された発酵条件下（例えば、低酸素、低温、イオン濃度の変化など）で標的タンパク質の収率が増加した細胞であってもよい。別の例では、所望の生物学的バリアントは、指定された発酵条件下で増殖する細胞であってもよい。

【０２３２】

統合ユニットからのデータの統合：いくつかの例では、統合ユニットのそれぞれからのデータをコンピュータ可読媒体（例えば、データリポジトリ内）に追加することができる。例えば、本明細書に記載されるように、統合ユニットのうちの１またはそれを超える統合ユニットは、コンピュータ可読媒体を含むか、またはコンピュータ可読媒体と相互作用することができる。コンピュータ可読媒体は、１またはそれを超えるデータセットを入力し、分子の構造－機能関係、所望の機能的特性に対する最適動作条件（例えば、適用または発酵条件）などの予測を出力するように構成することができる。コンピュータ可読媒体に入力されるデータセットは、統合ユニットのいずれか１つからのものであってもよい。例えば、データセットは、統合ユニット（例えば、株最適化ユニット）のいずれか１つからの生物学的バリアントまたは分子バリアントのスクリーニングの１つから生成することができ、所望の生物学的バリアントおよびその特性に関する情報（例えば、構造、配列、組成など）ならびに条件付き情報（例えば、培養、発酵、適用または反応条件）を含み得る。いくつかの例では、統合ユニットから生成されないデータセットは、コンピュータ可

読媒体に含まれるか、またはコンピュータ可読媒体に追加されてもよい。例えば、データベースからのデータセットをコンピュータ可読媒体に入力することができる。

【０２３３】

生物学的多様性および条件付き多様性スクリーニング：いくつかの態様では、統合システムは、生物学的多様性スクリーニングよび／または条件付き多様性スクリーニングに関する情報を含むデータベース、データセット、またはデータストリームを入力することができる。生物学的多様性データは、生物学的多様性スクリーニング（例えば、望ましい生物学的バリアントを同定するためのスクリーニング）の入力用の生物学的バリアント（例えば、異なる遺伝要素を有する遺伝的バリアント、分子バリアントなど）に関する情報を含むことができ、入力は、非限定的な例では、株情報、ＤＮＡ配列、ＲＮＡ配列、アミノ酸配列、タンパク質修飾（例えば、翻訳後修飾）、機能的性能（例えば、酵素活性、収量、力価、増殖速度、タンパク質間相互作用、結合特性）、またはそれらの組み合わせを含むことができる。例えば、生物学的多様性データストリームは、株最適化ユニットからのスクリーニングから生成されたデータを含むことができ、株最適化ユニットは、所望の標的タンパク質を産生するかまたは所望の機能的特性を有する異なる遺伝子（またはゲノム）バリアントに関するデータを生成することができる。そのようなデータは、（ｉ）各株の遺伝子（またはゲノム）バリアントの同一性、例えば、遺伝子（またはゲノム）バリアントのそれぞれから産生されるＤＮＡもしくはプラスミド配列および／またはペプチドもしくはタンパク質、ならびに（ｉｉ）遺伝的バリアントまたはそれに由来するタンパク質もしくはペプチドのそれぞれの性能（例えば、酵素活性、増殖速度）特性を含み得る。公的なソースまたはデータベースからの他の生物学的多様性データ（例えば、タンパク質の構造および配列、タンパク質の性能）もまた、生物学的多様性データセット内に含まれ得る。

【０２３４】

条件付き多様性データは、所望の生物学的バリアントを生成するために使用される１またはそれを超える条件に関するデータを含み得る。例えば、株最適化中に、１またはそれを超える発酵条件（例えば、ｐＨ、栄養レベル、二酸化炭素レベル、温度など）を変更して、一連の遺伝的バリアントにおいて所望の機能的性能を達成することができる。これらの発酵条件、およびいくつかの例では機能的性能は、１またはそれを超える条件付き多様性データセットを構成することができる。同様に、分子最適化（例えば、標的タンパク質の産生または最適化のため）または分子発見で使用される適用条件は、条件付き多様性データに含まれ得る。公的なソースまたはデータベースからの他の条件付き多様性データ（例えば、適用または発酵条件）もまた、条件付き多様性データセット内に含まれ得る。

【０２３５】

いくつかの例では、生物学的多様性データおよび条件付き多様性データは、独立してプログラム可能なデータセットである。例えば、生物学的多様性データおよび条件付き多様性データは、単一の実験（例えば、株最適化スクリーニング）から生成することができるが、生成されたデータは、実験の入力パラメータまたは変数に離散化され得る。例えば、実験の入力パラメータまたは変数（例えば、遺伝的バリアントおよび発酵条件）は、遺伝的バリアントと発酵条件との各組み合わせの機能的性能を含む一連のデータをもたらし得る。しかしながら、プログラム可能なデータセットは、各パラメータまたは変数が独立して操作可能であるように構成することができる。一例では、コンピュータ可読媒体は、パラメータ化（例えば、マルチパラメトリック）モデルを生成することができ、このモデルは、株の性能に対する変数（例えば、発酵条件および遺伝的バリアント）のそれぞれの関連する重みを分析することができる。いくつかの例では、生物学的多様性データまたは条件付き多様性データは、実験データおよびデータベース（例えば、公開されているデータセット）からのデータの両方を含む。

【０２３６】

全体として、生物学的多様性スクリーニングおよび条件付き多様性スクリーニングからのデータを使用して、システムに入力される標的パラメータ（例えば、標的性能パラメー

タ）に基づいて、望ましい生物学的バリアント（例えば、遺伝的バリアント）を同定することができる。

【０２３７】

機械学習アルゴリズム：いくつかの例では、コンピュータ可読媒体は機械学習アルゴリズムを含む。機械学習アルゴリズムは、統合ユニットまたは外部ソース（例えば、公開されているデータベース）から生成されたデータを反復的または連続的に統合および分析し、機能的特性または標的性能を有すると予測される新しい特徴（例えば、タンパク質をコードする遺伝的バリアント）または分子を発見するために使用することができる１またはそれを超えるニューラルネットワークを備えることができる。機械学習アルゴリズムは、例えば、機能的特性を示すことが明らかにされている分子、例えばＤＮＡまたはアミノ酸配列の既知の構造－機能関係の公開されているまたは公的に提供されたデータベースなどの大型ライブラリからのデータセットを使用して訓練することができる。代替的または追加的に、機械学習アルゴリズムは、統合ユニットのうちの１またはそれを超える統合ユニットから生成されたデータセットを使用して訓練することができる。

【０２３８】

機械学習アルゴリズムは、任意の数またはタイプの有用なアルゴリズムを含むことができる。例えば、機械学習アルゴリズムは、エラスティックネット回帰一般化線形モデル（ＧＬＭＮＥＴ）、サポートベクターマシン回帰（ＳＶＭ）、ランダムフォレスト（ＲＦ）、エクストリーム勾配ブースティング（ＸＧＢｏｏｓｔ）、多層パーセプトロン（ＭＬＰ）、畳み込みニューラルネットワーク（ＣＮＮ）、ディープニューラルネットワーク（ＤＮＮ）、またはそれらの組み合わせを含むことができる。機械学習アルゴリズムは、教師あり学習、教師なし学習、サポートベクターマシン、または他のアルゴリズムを含み得る。機械学習アルゴリズムを使用して、所望の機能的特性を達成し得る生物学的バリアントを予測または生成することができる。例えば、機械学習アルゴリズムは、予測量または閾値量の機能的特性を有し得る少なくとも１、少なくとも１０、少なくとも１００、少なくとも１，０００、少なくとも１０，０００、少なくとも１００，０００、少なくとも１，０００，０００またはそれを超える生物学的バリアントを出力するように構成することができる。

【０２３９】

一例では、機械学習アルゴリズムを使用して、まだ経験的に試験または決定されていない所望のまたは改善された機能的性能（例えば、酵素活性）を有する生物学的バリアント（例えば、遺伝的バリアント）またはその構造的構成要素（例えば、遺伝子配列、アミノ酸配列、モチーフなどの分子バリアント）を予測することができる。予測される生物学的バリアントまたは分子バリアントは、新規な分子もしくは配列、既知の分子もしくは配列、またはそれらの組み合わせであってもよい。機械学習アルゴリズムは、統合ユニットのそれぞれからのデータならびに公開されているデータセットを入力して、そのような予測を生成することができる。いくつかの例では、入力データは、処理、例えば、クリーニング、変換、正規化、標準化などされてもよい。その後、予測されたバリアントは、統合ユニットのうちの１つにおいて性能について経験的に試験することができる。機能的性能は記録され、機械学習アルゴリズムに再統合され、予測アルゴリズムをさらに改良することができる。例えば、予測が正しいことをサポートするデータおよび／または予測の誤りを証明するデータを使用して、予測アルゴリズムを反復的に調整または訓練することができる。このプロセスは、機械学習性能が得られるまで、例えば機械学習アルゴリズムが指定された精度または性能基準で予測分子を出力するまで、例えば任意の数の世代にわたって、何度も繰り返すことができる。

【０２４０】

一例では、訓練データは、株または分子最適化ユニットからのデータを含み得る。例えば、変異体のライブラリ、例えば、ポリマー（例えば、タンパク質）をコードする遺伝的バリアントを含む細胞は、例えば、ペプチドまたはタンパク質などのポリマーの各位置が２０個すべてのアミノ酸に変異することができる飽和突然変異誘発を使用して生成するこ

とができる。必要に応じて、さらなるランダム突然変異誘発および組換えが行われ得る。変異体のライブラリは、本明細書で提供される方法を使用してスクリーニングすることができ、データ（例えば、配列データ、濃縮データ、機能的性能データ（例えば、ポリマー、例えばタンパク質もしくはペプチドをコードする所与の核酸配列について、またはタンパク質もしくはペプチド配列について、産生された産物の数量もしくは量、酵素活性など）、スクリーニングまたはアッセイ条件など）を入力して機械学習アルゴリズムを訓練し、改善された機能的特性を有すると予想される新しいバリアントを予測することができる。

【０２４１】

いくつかの例では、機械学習アルゴリズムは、統合ユニットのそれぞれからのデータを統合する。例えば、分子発見ユニットは、１またはそれを超える分子および適用条件を入力し、分子を所望の機能活性についてスクリーニングし、分子の同一性（例えば、ＤＮＡまたはアミノ酸配列）および分子のそれぞれの機能的性能を出力することができる。これらの出力は、分子最適化ユニットにおける入力として提供され得る。その後、分子最適化ユニットは、必要に応じてデータベース（例えば、ＵｎｉＰｒｏｔなどのＤＮＡまたはタンパク質データベース）からの生物学的多様性および／または条件付きデータを使用して、生物学的バリアント（例えば、標的タンパク質をコードする株の遺伝的バリアント）または分子バリアント（例えば、タンパク質、モチーフ、構造、組成物）が閾値レベルの性能を有すると予想される予測を生成することができる。

【０２４２】

いくつかの例では、出力予測は、閾値レベルの性能を有すると予想されるバリアントと配列相同性を有する追加のバリアントを含み得る。例えば、予測は、主な予測されたバリアントに対して少なくとも７０％の配列相同性、少なくとも８０％の配列相同性、９０％の配列相同性またはそれを超える配列相同性を有するバリアントを含み得る。相同性は、例えば、２またはそれを超える核酸分子またはタンパク質の配列を整列させ、重複の割合を決定することによって決定することができる。そのようなバリアントの性能に関する追加データを得るために、（例えば、株最適化ユニットを使用して）追加のバリアントをスクリーニングすることもできる。追加データを使用して、分子（例えば、標的タンパク質）の機能的性能に重要であると考えられる重要な残基に関する情報を含む配列または一部の配列（部分配列）の追加のライブラリを生成することができる。いくつかの実施形態では、部分配列ライブラリは、重要な残基をコードする遺伝子または配列の既存のライブラリを（例えば、遺伝子シャッフリングまたは他の組換え戦略を介して）操作することによって生成することができる。

【０２４３】

生物学的バリアントまたは分子バリアントは、ライブラリ（例えば、生物学的遺伝的バリアントの場合、遺伝子ライブラリの構築）に組み込まれてもよい。いくつかの例では、ライブラリは、複数の細胞に導入される。次いで、ライブラリをさらにスクリーニングすることができる。高い性能を有するバリアントを単離し、さらに解析して（例えば、シーケンシングまたは化学的に分析して）バリアントを同定することができる。異なるバリアントからのデータを機械学習アルゴリズムに入力して、予測能力をさらに改良することができる。機械学習アルゴリズムはまた、高い性能を有すると予想されるバリアントおよび／または適用条件の新規な組み合わせを生成することができる。次いで、これらのバリアントおよび条件を入力し、さらなるスクリーニングのために株最適化ユニットで使用することができる。

【０２４４】

図１は、本明細書に記載の統合システムを概略的に示す。統合システム１００は、ハードウェアユニット、消費ユニット、および生物学的または化学的ユニットを備えるプラットフォームであってもよく、ユニットのそれぞれは、本明細書に記載の１またはそれを超える統合ユニットを構成する一部分であってもよい。統合システム１００は、株最適化ユニット１０５、分子最適化ユニット１１０、および分子発見ユニット１１５などの統合ユ

ニットを含むことができる。統合ユニットのそれぞれは、（例えば、遺伝的バリアントまたは分子バリアントの）ライブラリおよび一連の条件（例えば、発酵条件、適用条件）を入力し、（例えば、最適化された株、最適化された分子、または新規な分子の）予測を出力するように構成することができる。本明細書に記載するように、統合ユニットのそれぞれ内で実行されるプロセスは、再帰的であってもよい。

【０２４５】

株最適化ユニット１０５は、例えばランダム突然変異誘発、ＭＡＧＥ、ＣＲＥＡＴＥ、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、ＣＲＩＳＰＲガイドＤＮＡポリメラーゼ、エラープローンポリメラーゼ連鎖反応（ｅｐＰＣＲ）を使用して生成することができる、またはデータベースからのライブラリ、例えば全ゲノムシーケンシング（ＷＧＳ）ライブラリを含むことができる、複数の遺伝的バリアントを含むことができる株最適化ライブラリを入力するように構成することができる。株最適化ユニット１０５はまた、遺伝的バリアントを含む細胞が増殖される一連の発酵条件を入力することができる。出力として、株最適化ユニット１０５は、一連の条件下で特定の用途（例えば、酵素活性）に対して閾値性能を有する最適化された株（またはその情報）を生成することができる。

【０２４６】

分子最適化ユニット１１０は、分子最適化ライブラリを入力するように構成することができ、分子最適化ライブラリは、１またはそれを超える標的タンパク質をコードする（または標的タンパク質をコードする遺伝子と配列相同性を有する）複数の遺伝的バリアントを含み得る。遺伝的バリアントは、飽和突然変異誘発、循環置換、ＭＡＧＥ、ＣＲＥＡＴＥ、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、ＣＲＩＳＰＲガイドＤＮＡポリメラーゼ、エラープローンポリメラーゼ連鎖反応（ｅｐＰＣＲ）などの技術を使用して生成することができる。いくつかの例では（図示せず）、分子最適化ライブラリは、機能的特性について分子最適化プロセス中にスクリーニングされる複数の分子バリアントを含み得る。分子最適化ユニット１１０はまた、温度、培養液組成、酸素化、塩分、ｐＨなどを含むがこれらに限定されない一連の適用条件を入力することができる。出力として、分子最適化ユニット１１０は、一連の条件下で特定の用途（例えば、触媒活性、分子相互作用など）に対して閾値性能を有する最適化された分子（またはその情報、例えばＤＮＡまたはタンパク質配列）を生成することができる。

【０２４７】

分子発見ユニット１１５は、分子発見ライブラリを入力するように構成することができ、分子発見ライブラリは、複数の分子バリアントを含み得る。いくつかの例では、分子バリアントは、既知および未知の配列を有し得る複数の遺伝的バリアントを含む。分子バリアントは、試料から収集されたライブラリ（例えば、生物学的試料、土壌試料など）から生じ得るか、またはライブラリは、例えばメタゲノムライブラリ、ＤＮＡコード化ライブラリ、ファンクションジェネレータライブラリなどから生成することができる。遺伝的バリアントは、１またはそれを超えるタンパク質をコードしていてもよく、かつ／または望ましい機能的特性を有していてもよい。分子発見ユニット１１５はまた、温度、培養液組成、酸素化、塩分、ｐＨなどを含むがこれらに限定されない一連の適用条件を入力することができる。出力として、分子発見ユニット１１５は、一連の条件下で特定の用途（例えば、触媒活性、分子相互作用など）に対して閾値性能を有する新規な分子（またはその情報）を生成することができる。

【０２４８】

図２は、望ましい生物学的バリアントを選別するための統合システムを概略的に示す。統合システム２００は、ライブラリ入力２１１（例えば、株ライブラリまたはＤＮＡライブラリ）、アッセイ入力２１３（例えば、ライブラリのメンバーの機能的特性に関する実験データ）、および例えば産生条件および性能形質（例えば、産生または発現レベル、基質変換、一連の環境条件下での産生、安定性など）を含む性能要件入力２１５を含む複数

の入力２１０を受け入れるように構成することができる。統合システム２００は、（例えば、再帰的サイクルのうちの１またはそれを超えるサイクル間に）１またはそれを超える中間の出力２１７を出力することができ、この出力は、配列データ、濃縮された株、または濃縮されたタンパク質などの統合ユニット（例えば、実験モジュール）からのデータを含み得る。中間の出力２１７はまた、一連のパラメータがより大規模にどのように機能することができるかについての予測、特定の条件下での入力の機能的性能を含むデータの出力、および性能の可能なスケールアップ予測を含むことができる。中間の出力２１７は、入力として統合システム２００に再統合することができる。統合システムの最終出力２１９は、予測され最適化された株、最適化された分子、または機能的性能が向上すると予想される新しい分子を含むことができる。

【０２４９】

コンピュータシステム

本開示は、本開示の方法を実施するようにプログラムされたコンピュータシステムを提供する。図３は、１またはそれを超えるデータストリームを入力し、１またはそれを超えるデータセットをデータリポジトリに格納し、データの閾値パラメータを設定し、かつ／または、統合ユニットのうちの１またはそれを超える統合ユニットに対して、多様性スクリーニングを実行または反復するように指示するようにプログラムまたはその他の方法で構成されたコンピュータシステム３０１を示す。コンピュータシステム３０１は、例えば、株最適化ユニット、分子最適化ユニット、分子発見ユニットの１またはそれを超える構成要素を制御すること、データを入力すること、分子の構造－機能関係に関する１またはそれを超える予測を出力することなど、本開示の株または分子最適化または発見の様々な態様を調節することができる。コンピュータシステム３０１は、ユーザの電子デバイス、または電子デバイスに対して遠隔に配置されたコンピュータシステムとすることができる。電子デバイスは、モバイル電子デバイスとすることができる。

【０２５０】

コンピュータシステム３０１は、中央処理装置（ＣＰＵ、本明細書では「プロセッサ」および「コンピュータプロセッサ」）３０５を含み、この中央処理装置は、シングルコアもしくはマルチコアプロセッサ、または並列処理のための複数のプロセッサとすることができる。コンピュータシステム３０１はまた、メモリまたはメモリロケーション３１０（例えば、ランダムアクセスメモリ、読み出し専用メモリ、フラッシュメモリ）と、電子記憶ユニット３１５（例えば、ハードディスク）と、１またはそれを超える他のシステムと通信するための通信インターフェース３２０（例えば、ネットワークアダプタ）と、キャッシュ、他のメモリ、データストレージ、および／または電子ディスプレイアダプタなどの周辺装置３２５とを含む。メモリ３１０、記憶ユニット３１５、インターフェース３２０および周辺装置３２５は、マザーボードなどの通信バス（実線）を介してＣＰＵ３０５と通信する。記憶ユニット３１５は、データを記憶するためのデータ記憶ユニット（またはデータリポジトリ）とすることができる。コンピュータシステム３０１は、通信インターフェース３２０の助けを借りてコンピュータネットワーク（「ネットワーク」）３３０に動作可能に結合することができる。ネットワーク３３０は、インターネット、インターネットおよび／もしくはエクストラネット、またはインターネットと通信するイントラネットおよび／もしくはエクストラネットとすることができる。ネットワーク３３０は、場合によっては、電気通信および／またはデータネットワークである。ネットワーク３３０は、クラウドコンピューティングなどの分散コンピューティングを可能にすることができる１またはそれを超えるコンピュータサーバを含むことができる。ネットワーク３３０は、場合によっては、コンピュータシステム３０１の助けを借りて、ピアツーピアネットワークを実装することができ、これによってコンピュータシステム３０１に結合されたデバイスがクライアントまたはサーバとして動作することを可能にすることができる。

【０２５１】

ＣＰＵ３０５は、プログラムまたはソフトウェアで具現化することができる、機械可読命令のシーケンスを実行することができる。命令は、メモリ３１０などのメモリロケーシ

ョンに格納することができる。命令は、ＣＰＵ３０５を対象とすることができ、ＣＰＵ３０５は、その後、本開示の方法を実施するようにＣＰＵをプログラムまたは構成することができる。ＣＰＵ３０５によって実行される動作の例は、フェッチ、デコード、実行、およびライトバックを含むことができる。

【０２５２】

ＣＰＵ３０５は、集積回路などの回路の一部とすることができる。システム３０１の１またはそれを超える他の構成要素を回路に含めることができる。場合によっては、回路は特定用途向け集積回路（ＡＳＩＣ）である。

【０２５３】

記憶ユニット３１５は、ドライバ、ライブラリおよび保存されたプログラムなどのファイルを記憶することができる。記憶ユニット３１５は、ユーザデータ、例えば、ユーザプレファレンスおよびユーザプログラムを記憶することができる。コンピュータシステム３０１は、場合によっては、イントラネットまたはインターネットを介してコンピュータシステム３０１と通信するリモートサーバ上に位置するなど、コンピュータシステム３０１の外部にある１またはそれを超える追加データ記憶ユニットを含むことができる。

【０２５４】

コンピュータシステム３０１は、ネットワーク３３０を介して１またはそれを超えるリモートコンピュータシステムと通信することができる。例えば、コンピュータシステム３０１は、ユーザのリモートコンピュータシステムと通信することができる。リモートコンピュータシステムの例としては、パーソナルコンピュータ（例えば、ポータブルＰＣ）、スレートもしくはタブレットＰＣ（例えば、Ａｐｐｌｅ（登録商標）ｉＰａｄ（登録商標）、Ｓａｍｓｕｎｇ（登録商標）Ｇａｌａｘｙ　Ｔａｂ）、電話、スマートフォン（例えば、Ａｐｐｌｅ（登録商標）ｉＰｈｏｎｅ（登録商標）、Ａｎｄｒｏｉｄ（登録商標）対応デバイス、Ｂｌａｃｋｂｅｒｒｙ（登録商標））、または携帯情報端末が挙げられる。ユーザは、ネットワーク３３０を介してコンピュータシステム３０１にアクセスすることができる。

【０２５５】

本明細書に記載の方法は、例えばメモリ３１０または電子記憶ユニット３１５などのコンピュータシステム３０１の電子記憶ロケーションに記憶された機械（例えば、コンピュータプロセッサ）実行可能コードによって実施することができる。マシン実行可能コードまたはマシン可読コードは、ソフトウェアの形態で提供することができる。使用中、コードはプロセッサ３０５によって実行することができる。場合によっては、コードは、記憶ユニット３１５から取得され、プロセッサ３０５による容易なアクセスのためにメモリ３１０に記憶することができる。いくつかの状況では、電子記憶ユニット３１５を除外することができ、マシン実行可能命令がメモリ３１０に記憶される。

【０２５６】

コードは、コードを実行するように適合されたプロセッサを有する機械で使用するためにプリコンパイルおよび構成することができ、またはランタイム中にコンパイルすることができる。コードは、コードがプリコンパイルされた方法またはコンパイルされた方法で実行することを可能にするように選択することができるプログラミング言語で供給することができる。

【０２５７】

コンピュータシステム３０１など、本明細書で提供されるシステムおよび方法の態様は、プログラミングにおいて具現化することができる。本技術の様々な態様は、典型的にはマシン（またはプロセッサ）実行可能コードおよび／またはある種のマシン可読媒体上に搭載されるかまたはそれに具体化される関連データの形態の「製品」または「製造品」と考えることができる。マシン実行可能コードは、メモリ（例えば、読み出し専用メモリ、ランダムアクセスメモリ、フラッシュメモリ）またはハードディスクなどの電子記憶ユニットに記憶することができる。「記憶」タイプの媒体は、コンピュータ、プロセッサなどの有形メモリ、または様々な半導体メモリ、テープドライブ、ディスクドライブなどの関

連モジュールのいずれかまたはすべてを含むことができ、ソフトウェアプログラミングのためにいつでも非一時的記憶を提供することができる。ソフトウェアのすべてまたは一部は、インターネットまたは様々な他の電気通信ネットワークを介して通信されることがある。そのような通信は、例えば、あるコンピュータまたはプロセッサから別のコンピュータまたはプロセッサへの、例えば管理サーバまたはホストコンピュータからアプリケーションサーバのコンピュータプラットフォームへのソフトウェアのロードを可能にすることができる。したがって、ソフトウェア要素を搭載することができる別のタイプの媒体は、ローカルデバイス間の物理インターフェースにわたって、有線および光の地上ネットワークを介して、および様々なエアリンクを介して使用されるような、光波、電気波、および電磁波を含む。有線または無線リンク、光リンクなど、そのような波を搬送する物理的要素もまた、ソフトウェアを搭載する媒体と考えることができる。本明細書で使用される場合、非一時的で有形の「記憶」媒体に限定されない限り、コンピュータまたは機械「可読媒体」などの用語は、実行のためにプロセッサに命令を提供することに関与する任意の媒体を指す。

【０２５８】

したがって、コンピュータ実行可能コードなどのマシン可読媒体は、有形記憶媒体、搬送波媒体、または物理伝送媒体を含むがこれらに限定されない多くの形態をとることができる。不揮発性記憶媒体は、例えば、図面に示されるデータベースなどを実装するために使用することができる任意のコンピュータなどの記憶装置のいずれかなどの光学または磁気ディスクを含む。揮発性記憶媒体は、そのようなコンピュータプラットフォームのメインメモリなどのダイナミックメモリを含む。有形伝送媒体は同軸ケーブル、コンピュータシステム内にバスを備えるワイヤを含む銅ワイヤおよび光ファイバを含む。搬送波伝送媒体は、電気信号もしくは電磁信号、または無線周波数（ＲＦ）および赤外線（ＩＲ）データ通信中に生成されるような音波もしくは光波の形態をとることができる。したがって、コンピュータ可読媒体の一般的な形態としては、例えば、フロッピー（登録商標）ディスク、フレキシブルディスク、ハードディスク、磁気テープ、任意の他の磁気媒体、ＣＤ－ＲＯＭ、ＤＶＤもしくはＤＶＤ－ＲＯＭ、任意の他の光学媒体、パンチカード、紙テープ、穴のパターンを有する任意の他の物理的記憶媒体、ＲＡＭ、ＲＯＭ、ＰＲＯＭおよびＥＰＲＯＭ、ＦＬＡＳＨ－ＥＰＲＯＭ、任意の他のメモリチップもしくはカートリッジ、データもしくは命令を搬送する搬送波、そのような搬送波を搬送するケーブルもしくはリンク、またはコンピュータがプログラミングコードおよび／もしくはデータを読み取ることができる任意の他の媒体が挙げられる。これらの形態のコンピュータ可読媒体の多くは、実行のために１またはそれを超える命令の１またはそれを超えるシーケンスをプロセッサに搬送することに関与することができる。

【０２５９】

コンピュータシステム３０１は、例えば、所望の生物学的バリアントを選別するための入力閾値パラメータを提供するためのユーザインターフェース（ＵＩ）３４０を含む電子ディスプレイ３３５を含むか、またはそれと通信することができる。ＵＩの例としては、グラフィカルユーザインターフェース（ＧＵＩ）およびウェブベースのユーザインターフェースが挙げられるが、これらに限定されない。

【０２６０】

本開示の方法およびシステムは、１またはそれを超えるアルゴリズムによって実施することができる。アルゴリズムは、中央処理装置３０５による実行時にソフトウェアによって実施することができる。アルゴリズムは、例えば、生物学的多様性または条件付き多様性に関するデータストリームを反復的に入力し、分子、最適化された株、最適化された分子、または発見分子の構造－機能関係に関する予測を出力する機械学習アルゴリズムであってもよい。

【実施例】

【０２６１】

実施例１：株最適化

マイクロスケールスクリーニング条件をより大規模で経験するスクリーニング条件と正確に一致させることは困難である可能性がある。さらに、大規模容器の混合の物理的制限は、高度に不均一な環境での株をもたらす可能性があり、細胞は、酸素レベル、ｐＨ、炭素源濃度および温度などの条件で勾配にさらされる。従来のスクリーニングプログラムでは、変異体のスクリーニングは多くの場合、時間およびコストのボトルネックとなる可能性があり、したがって、大規模産生条件に最も近い特定のスクリーニング条件を選択することができる。これらの条件を一致させることは困難である可能性があり、大規模での条件の不均一性により、特定の小規模スクリーニングアッセイから改善されるように選択された株が、実際により大きな規模で機能する可能性が低くなる可能性がある。

【０２６２】

アッセイの小型化は、様々な条件下での大規模な集団をスクリーニングする際に大きな支障がない程度まで、スクリーニングの時間およびコストを削減することができる。この実施例では、ライブラリを複数の条件下で独立してスクリーニングし、次世代シーケンシング（ＮＧＳ）を使用して、これらの異なる条件下で濃縮された突然変異を同定する。突然変異は、それらが影響を及ぼす代謝経路に分類される。次いで、この表現型、配列および代謝経路ネットワークデータを機械学習アルゴリズムに供給して、異なる環境条件下での性能にとって、どの突然変異または経路または組み合わせが最も重要であるか、またどの突然変異が複数の条件にわたってロバストな改善を提供し、したがって一般に有用であるかの予測を可能にする。

【０２６３】

株の構築

Ｔｈｅｒｍｏ－Ｆｉｓｈｅｒからのプロテアーゼ欠損市販株ｐｉｃｈｉａ ｐａｓｔｏｒｉｓ ＳＭＤ１１６８株から出発して、ヒトＩ型コラーゲン遺伝子ＣＯＬ１Ａ１のＰＣＲ増幅およびｐＰＩＣ９Ｋプラスミドへのクローニングを行い、続いて、得られたプラスミドでｐｉｃｈｉａ ｐａｓｔｏｒｉｓ ＳＭＤ１１６８株を形質転換する。ｐＰＩＣ９Ｋプラスミドは、Ｈｉｓ４選択下でのマルチコピープラスミドの生成を可能にし、タンパク質の細胞外産生を可能にする分泌タグを含有する。

【０２６４】

株最適化ライブラリの構築

ライブラリは、改変株の全ゲノム突然変異誘発を行うことによって作成される。次いで、対数期細胞をＮ－メチル－Ｎ’－ニトロ－Ｎ－ニトログアニジン（ＮＴＧ）に曝露し、適切な濃度および時間でチオ硫酸ナトリウムを用いてクエンチして、０．５～１％の生存率を得る。

【０２６５】

コラーゲン産生のためのマイクロスケールアッセイ

ピキア細胞からのコラーゲンの産生を検出するために、ＰＤＭＳナノウェルアレイを作成する（Ｔｏｒｒｅｓ ＡＪ、Ｈｉｌｌ ＡＳ、Ｌｏｖｅ ＪＣ．Ｎａｎｏｗｅｌｌ－ｂａｓｅｄ ｉｍｍｕｎｏａｓｓａｙｓ ｆｏｒ ｍｅａｓｕｒｉｎｇ ｓｉｎｇｌｅ－ｃｅｌｌ ｓｅｃｒｅｔｉｏｎ：ｃｈａｒａｃｔｅｒｉｚａｔｉｏｎ ｏｆ ｔｒａｎｓｐｏｒｔ ａｎｄ ｓｕｒｆａｃｅ ｂｉｎｄｉｎｇ．Ａｎａｌ．Ｃｈｅｍ．２０１４；８６（２３）：１１５２－１１５６９．ｄｏｉ：１０．１０２１／ａｃ４０３０２９７）。０．５％メタノールを含有する増殖培地にライブラリから細胞を懸濁させ、タンパク質発現を誘導する。単一細胞測定を確実にするために１０ウェルに約１個の細胞が存在するように細胞をＰＤＭＳナノウェルに分配する。抗ＣＯＬ１Ａ１抗体クローン５Ｄ８－Ｇ９（Ｍｉｌｌｉｐｏｒｅ－Ｓｉｇｍａ）でコーティングしたスライドガラスでウェルを覆う。細胞によって分泌されたコラーゲンは、ガラス表面上の抗体によって捕捉される。次いで、スライドガラスを取り出し、表面をＦＩＴＣコンジュゲート化ヤギ抗マウス抗体で染色し、蛍光顕微鏡で画像化する。ＦＩＴＣ強度は、細胞によって分泌されるコラーゲンの量に比例する。高発現クローンを含むナノウェルを特定し、ロボットマイクロマニピュレータを用いてナノウェルアレイから取り出し、次いでさらなる研究のために非誘導条件下で

増殖させる。
【０２６６】
株の改善
細胞が産生において経験する可能性が高い条件の範囲をカバーする複数のスクリーニング実験を並行して設定する。各条件について、最も高い産生細胞を選択し、再増殖させて、９６ウェルプレートにおいて１ｍｌスケールでの産生の改善を確認する。性能を確認した後、次にＤＮＡを抽出し、各条件からのプールされたＤＮＡまたは個々のクローンのいずれかで、ＩｌｌｕｍｉｎａのＭｉＳｅｑを使用するなどして次世代シーケンシング（ＮＧＳ）を行う。異なる条件下での産生に関連する突然変異を同定する。
【０２６７】
ライブラリから見出される株は、完全に最適化される可能性は低い。有益でない突然変異を除外し、有益な組み合わせを発見するために、プロトプラスト融合および「ゲノムシャッフリング」を使用して、高産生株のゲノム間の組換えを可能にする。組換えライブラリを、最も高い産生物を選別するために使用されるのと同じ初期条件下でスクリーニングする。
【０２６８】
ライブラリ作成、スクリーニングおよび分析の再帰的サイクルを、性能要件を満たす株が同定されるまで繰り返す。
【０２６９】
実施例２：分子発見
メタゲノムライブラリからの新規リパーゼ酵素の発見。
環境ＤＮＡ（ｅＤＮＡ）を土壌または他の供給源から収集および精製する。ｅＤＮＡを切断し、典型的な遺伝子およびプロモーターサイズを包含するように２～５ｋｂの範囲にサイズ選択する。断片をこれをプラスミドにクローニングし、次いで高形質転換効率の大腸菌に形質転換する。
【０２７０】
リパーゼアッセイ
大腸菌細胞を、市販の蛍光発生性基質ＥｎｚＣｈｅｃｋ　Ｌｉｐａｓｅ　Ｓｕｂｓｔｒａｔｅ（商標）（ＴｈｅｒｍｏＦｉｓｃｈｅｒ）および細胞溶解試薬Ｂ－ＰＥＲ（商標）（ＴｈｅｒｍｏＦｉｓｈｅｒ）と組み合わせる。混合物を、組織溶解剤ホモジナイザ（ＴｉｓｓｕｅＬｙｓｅｒ　ＩＩ、Ｑｉａｇｅｎ）で、２％「００８」フルオロ界面活性剤（ＲＡＮ　Ｂｉｏｔｅｃｈｎｏｌｏｇｉｅｓ）を含む界面活性剤と油の混合物ＨＦＥ－７５００（３Ｍ）を用いて震盪して乳化する。穿刺可能な光透過性膜を有する浅いフローセルにエマルジョンを移す。１０×５０×０．２５ｍｍのフローセルでは、約１２５ｕＬのエマルジョンを保存することができ、約１０８ピコリットル容量のエマルジョン区画に相当する。溶解した細胞から酵素が放出されると、活性リパーゼを発現する細胞を含む液滴は蛍光性になる。蛍光顕微鏡を用いてフローセルを画像化し、被覆膜を突き刺し、マイクロマニピュレーターニードルで吸引することによって蛍光の領域を採取する。
【０２７１】
活性リパーゼ酵素をコードする遺伝子が濃縮されたプラスミドを採取し、大腸菌に再形質転換する。形質転換、アッセイおよび回収のサイクルは、集団が活性リパーゼ遺伝子について実質的に濃縮されるまで繰り返される。
【０２７２】
次いで、個々のクローンを選択から単離し、別々に増殖させる。リパーゼ活性を確認し、インサートのシーケンシングを行って、リパーゼをコードするオープンリーディングフレームを同定する。配列を既知のリパーゼ遺伝子ファミリーとの相同性についてチェックし、新しいクラスを同定する。濃縮後にライブラリに残っているユニークなクローンの数に応じて、２つの経路がある。数が１～約１００である場合、酵素を個々に活性について評価する。対照的に、ユニークなクローンの数が１００を超える場合、さらなる特性評価が集団レベルで行われる。

【０２７３】
特性評価
集団レベルで配列を特性評価するために、数百から数千のユニークな活性酵素を含有する上記の活性濃縮からの出力が使用される。この同じプールは、いくつかの異なる条件下で並行して複数回、上記のようなハイスループットエマルジョンベースの濃縮に供される。この多条件特性評価では、基質状態（可溶性と表面吸着）、低温および高温での安定性および触媒性能、プロテアーゼ分解に対する耐性、二価カチオン濃度に対する感受性などのいくつかの重要なパラメータにわたって性能が評価される。これらのパラメータのそれぞれについての低条件および高条件での濃縮が行われ、各条件下での最良の性能について集団を濃縮する。次いで、ロングリードＮＧＳ（ＬｏｏｐＳｅｑ（商標）、ＬＯＯＰ　Ｇｅｎｏｍｉｃｓ）を使用して、これらのパラメータにわたる配列の差次的濃縮を測定する。濃縮度のこれらの差は、アッセイにおけるこれらのバリアントの性能と相関する。
【０２７４】
配列データ、構造推論、および性能データは、異なる条件下でどの領域、モチーフ、および構造クラスが良好に機能するかを予測するために使用できる機械学習訓練セットを構築するために使用される。このモデルは、分子最適化を推進するための入力として使用される。
【０２７５】
実施例３：分子最適化－低温での活性のためのリパーゼ酵素の最適化
分子発見および特性評価ユニットの出力は、配列同一性および性能に関する情報を含む。酵素のファミリー、酵素内の領域、および特定のアミノ酸残基が、より高い低温性能と関連しているかについてのデータを取得する。アルゴリズムに基づいて良好な性能を有すると予測される天然バリアントについて既存の配列データベースを検索する。この情報を使用して、これらの領域を標的とするリパーゼ遺伝子ライブラリを構築する。ライブラリを適切な宿主に形質転換し、１５℃の目標温度で最高の性能を有するものについてスクリーニングする。ライブラリ多様性が１，０００未満の場合、スクリーニングは従来のマイクロタイター形式で行われ、多様性が大きい場合、スクリーニングはマイクロ流体プラグまたは液滴で行われる。
【０２７６】
いずれの場合も、タンパク質バリアントが細菌中で発現され、次いで、これらの細菌の溶解物が蛍光発生性基質に曝露される。集団は、低温で最も高い活性を有するバリアントが濃縮されている。次いで、濃縮された集団をシーケンシングし、出発集団と比較する。陽性濃縮配列および陰性濃縮配列の両方が、本発明者らのシステムにおける最良の性能の酵素の予測を可能にするモデルをもたらす。このデータを使用して、性能に影響を及ぼす重要な残基が飽和している部分配列ライブラリを設計する。最も高い性能の配列はまた、遺伝子シャッフリングまたは有望な組み合わせを合成することによっても組み換えられる。濃縮およびライブラリの改良のサイクルは、性能目標に達するまで繰り返される。
【０２７７】
実施例４：ランダム突然変異誘発による株最適化
本明細書の方法、組成物、およびシステムは、高い予測可能性を有するハイスループット形式での株最適化を可能にすることができる。一例では、目的の分子を発現する株は、非標的突然変異誘発、例えばランダム突然変異誘発に供することができる。ランダム突然変異誘発は、目的の分子を発現する株を突然変異原および／またはランダムな遺伝子変化（例えば、トランスポゾン突然変異誘発、ＭＡＧＥ、ゲノムシャッフリング、ランダム組換え、非相同末端結合など）に供し、それによって突然変異誘発細胞のライブラリを生成することを含み得る。そのようなライブラリは、例えば、１０，０００を超えるバリアントまたは１，０００，０００を超えるバリアントを含み得る。突然変異誘発細胞のライブラリ（一世代と考えられる）は、液体培養液中での増殖によって拡大させることができ、細胞は、集団が単一細胞からなることを確実にするために処理することができる。
【０２７８】

図４は、株最適化を実行するための例示的なワークフローを概略的に示す。プロセス４０１では、上記のように、変異体のライブラリ（または本明細書では遺伝的バリアントとも呼ばれる突然変異誘発細胞）を生成することができる。プロセス４０３では、変異体のライブラリを個々の反応容器（例えば、マイクロリアクタ）に分配することができる。いくつかの例では、個々の細胞または単一の細胞が個々の容器に分配される。分配は、すべての個々の容器が所定の体積（範囲内）を含むことを保証するフィードバックループを有するコンピュータ実装方法によって支援されてもよく、これは個々の容器内の個々の細胞間の一貫した増殖を保証するのに有用であり得る。いくつかの例では、画像認識を使用して、所望の数の細胞が個々の容器に充填されることを確実にすることができる。

【０２７９】

個々の細胞を含む個々の容器を、発酵条件を模倣する条件（例えば、温度、炭素源、ｐＨ、培養液成分、通気など）に供することができる。個々の容器は、同じ組の発酵条件を含んでもよく、または異なる発酵条件を含んでもよく、または個々の容器のサブセットは、同じ組の発酵条件を含んでもよく、個々の容器の別のサブセットは、異なる組の発酵条件を含んでもよい。個々の容器内の細胞は、細胞増殖および／または目的の分子の産生を可能にする時間枠にわたって個々の容器内でインキュベートすることができる。マイクロバイオリアクタ内の増殖は、画像化およびコンピュータ画像認識によって監視することができる。

【０２８０】

プロセス４０５では、個々の容器をスクリーニングすることができ、スクリーニングに基づいて、（一連の分配された突然変異誘発細胞の）１またはそれを超える細胞を濃縮することができる（例えば、改善された細胞または所望の表現型を発現する細胞を選択することができる）。一例では、個々の容器またはその中の内容物を、試薬の添加および反応時間の制御を可能にする装置内に配置することができる。目的の分子の検出を行うことができ、検出に基づいて、１またはそれを超える細胞を選択することができる。検出および／または選択は、自動化された様式で行うことができる。例えば、本明細書に記載のシステムは、検出試薬の添加を介して分子生成および分子活性について調べることによって、どの個々の容器が目的の表現型を示す細胞を含有するかを決定することができるコンピュータ実施方法を含み得る。検出試薬の添加を使用して、目的の分子の環境（例えば、ｐＨ、イオン強度、イオンタイプ、基質、阻害剤、酵素、競合剤）を変化させることができる。このような環境の変化は、生成される分子の量および／または生成される分子の機能を反映する条件を生成するのに有用であり得る。試薬の添加後の規定の制御可能な時間に、分光シグナルを介して目的の分子について個々の容器を評価することができる。機器設計および機器のコンピュータ制御フィードバックは、タイミング、試薬添加、分子生成の評価、およびどのマイクロバイオリアクタが選択基準を満たすかに関する意思決定を制御することができる。いくつかの例では、ＰＩＤアルゴリズムを使用して、株最適化の過程での機器のドリフト（例えば、機械的機器、液体粘度、温度などのわずかな偏差に起因する）に起因する誤った選別決定を防止することができる。

【０２８１】

必要に応じて、選別後、例えば、特定のライブラリメンバー（例えば、突然変異誘発細胞）を保持するための機器の決定時、選択された細胞は、システムによってその容器から取り出され、生物の生存率を最大化する環境（例えば、十分な栄養素条件、塩条件、温度、通気下など）に保持することができる。細胞は、細胞増殖を介して拡大させることができ、必要に応じて、別の過程（例えば、プロセス４０３、および４０５、いくつかの例では４０７）のスクリーニングおよび濃縮を実施することができる。必要に応じて、行われるスクリーニングおよび濃縮の各回について、よりストリンジェントな条件を適用することができる（例えば、より低い酸素、より低い栄養素、ｐＨの変化など）。代替的または追加的に、濃縮中によりストリンジェントな性能基準（例えば、発現の増加、触媒活性など）を適用することができる。その後、濃縮された細胞を、プロセス４０７において特性評価することができる。濃縮された細胞は、例えば、シーケンシングまたは表現型スクリ

ーニングを介して分析することができる。例えば、濃縮された細胞は、さらなる試験条件下で所望の表現型について確認するために分析することができる。次いで、所望の表現型を保持する細胞は、遺伝的変化およびスクリーニング（例えば、プロセス４０１、４０３および４０５）の別の反復（「世代」）に導入され、プロセスは、性能基準が満たされるまで再帰的サイクルで反復することができる。

【０２８２】

図５は、本明細書に記載の株最適化プロセスの例示的なワークフローを示す。親株または細胞株５０１をランダム突然変異誘発（９回の紫外線曝露）に供して、変異体の第１のライブラリ、例えば突然変異誘発細胞５０３を提供する。突然変異誘発細胞５０３の第１のライブラリは、第１の過程の濃縮に供され、第１の過程は、分配、任意の増殖、および選別を含み得る。閾値を超える機能的性能を有する変異体を収集し、次いで別の過程の濃縮に供して、突然変異誘発細胞の第２のライブラリ５０５を生成することができる。濃縮プロセスは、突然変異誘発細胞の後続のライブラリ（例えば、５０７および５０９）を得るために、任意の数の有用な過程について実施することができる。次いで、突然変異誘発細胞の最終ライブラリを収集し、（例えば、シーケンシングを介して）特性評価して、追加の突然変異誘発を実施して、例えば本明細書に記載されるように、新しい世代を開始することができる。新しい世代の親株または細胞株５１１を突然変異誘発（エチルメチルスルホネート（ＥＭＳ）への曝露）にさらに供して、１またはそれを超える濃縮を行い、最適または所望の性能基準が満たされるまでプロセスを繰り返すことができる。

【０２８３】

図６Ａ～図６Ｃは、３世代にわたって株最適化を実行することによる例示的なデータを示す（例えば、３サイクルの突然変異誘発、分配、濃縮および特性評価）。図６Ａは、突然変異誘発されていない可能性のある株最適化６０１の前、ならびに第１の世代内の濃縮の第１の過程６０３後および濃縮の第２の過程６０５後の個々の細胞の、目的の分子（例えば、酵素）の量または数量を表し得る蛍光強度（ｘ軸）のヒストグラムを示す。次いで、第１の世代の濃縮の最後の過程の最も性能の高いバリアント（例えば、目的の分子の最大数量を産生するバリアント）をさらに突然変異誘発し、それによって第２の世代の変異体の生成を開始する。図６Ｂは、第２の世代中の株最適化６０７の前、第１の過程６０９後、第２の過程６１１後、および第３の過程６１３後の個々の細胞の蛍光強度（ｘ軸）のヒストグラムを示す。図６Ｃは、第３の世代中の株最適化６１５の前、第１の過程６１７後、第２の過程６１９後、および第３の過程６１３後の個々の細胞の蛍光強度（ｘ軸）のヒストグラムを示す。図６Ａ～図６Ｃに示すように、濃縮の追加の過程は、強度の増加（右シフト）をもたらし、性能の向上を示している。

【０２８４】

図７は、２世代の突然変異誘発および複数回の選別および濃縮にわたって得られた変異体の生成された生成物の濃度の例示的なプロットを、反応容器（例えば、ハイスループットスクリーニングおよび選別に使用される場合）で培養した場合と、（ｘ軸）バルク供給バッチ反応器での培養（ｙ軸）とを比較して示す。各点は、マイクロリアクタおよびバッチ増殖条件の両方で増殖させた場合の各世代からの単一の変異体（最も発現する突然変異体）から生成された生成物の量（濃度）を表す。図７に示すように、マイクロリアクタにおける目的の分子（例えば、タンパク質）の生成は、バッチスケールでの目的の分子（例えば、タンパク質）の生成とほぼ線形に比例する関係であり、したがって、株最適化プロセスは、バッチ増殖条件を代表する結果をマイクロスケールで得ることができることを示している。全体として、これらのデータは、株最適化がハイスループット方式で株の機能的性能を改善するための効果的なツールとなり得る方法を示している。

【０２８５】

実施例５：自然多様性ライブラリによる株最適化

本明細書の方法、組成物、およびシステムは、高い予測可能性を有するハイスループット形式での株最適化を可能にすることができる。一例では、目的の分子を発現する株を、大きな天然および／または操作された多様性セットに基づく遺伝子変化に供することがで

きる。多様性セットは、プロモーター、５’非翻訳領域、リボザイム、ＲＮＡ安定性配列、分泌ペプチド、シグナルペプチド、融合タンパク質、ＤＮＡ結合ドメイン、タンパク質－タンパク質結合領域、コドン最適化遺伝子、コドンランダム化遺伝子、３’非翻訳領域、ターミネーター、非相同末端結合などを含むことができ、これらは突然変異誘発細胞のライブラリを生成するために複数の細胞に導入することができる。突然変異誘発細胞のライブラリは、例えば、１，０００を超えるバリアント、１０，０００を超えるバリアント、１００，０００を超えるバリアントまたは１，０００，０００を超えるバリアントを含み得る。突然変異誘発細胞のライブラリ（一世代と考えられる）は、液体培養液中での増殖によって拡大させることができ、細胞は、集団が単一細胞からなることを確実にするために処理することができる。

【０２８６】

世代内で過程を開始するために、突然変異誘発細胞のライブラリの個々の細胞を個々の容器に分配し、発酵条件に供し、必要に応じてモニターし、上記のように（実施例４を参照）および図４～５に示すようにスクリーニングすることができる。

【０２８７】

図８は、単一の世代（例えば、突然変異誘発の１サイクル）にわたって株最適化を実行することによる例示的なデータを示す。図８では、各パネルは、産生された目的の分子の数量を表す強度のヒストグラムプロットを示す。非突然変異誘発細胞８０３（例えば、突然変異の導入前の細胞）は、予想されるように比較的類似したピークプロファイルを維持するが、各過程から得られた突然変異誘発細胞８０１（例えば、分配され、培養され、スクリーニングされる突然変異誘発細胞）の全集団の平均は、おそらく突然変異誘発の有害な影響により、最初は目的の分子の産生がより低いことを示している。その後のスクリーニング過程（例えば、分配、濃縮、および任意の特性評価）にわたって、突然変異誘発細胞８０１の全集団の平均は右へのピークシフトを示し、これは、株最適化およびスクリーニングプロセスの結果として、強度、したがって目的の生成された分子の数量の増加を示している。

【０２８８】

図９は、単一世代内の各スクリーニング過程にわたって目的の分子を測定するためのプレートベースのアッセイの結果を示す。各バーは、非突然変異誘発細胞生成９０３に対して正規化された、株最適化の過程にわたる強度の倍数変化（産生された目的の分子の数量を表す）の測定値を表す。各過程（バーの各セット）から得られた突然変異誘発細胞９０１（例えば、分配され、培養され、スクリーニングされる突然変異誘発細胞）は、１回あたりの強度が増加し、したがって、株最適化プロセスの結果として、産生された目的の分子の数量がより多いことを示している。

【０２８９】

実施例６：標的化または非標的化突然変異誘発による分子発見

本明細書の方法、組成物およびシステムは、分子発見、例えば、表現型の改善をもたらす遺伝的バリアントの発見を可能にすることができる。一例では、細胞を突然変異誘発（例えば、標的化または非標的化突然変異誘発）に供して、突然変異誘発細胞のライブラリを生成することができる。そのようなライブラリは、例えば、１０，０００を超えるバリアントまたは１，０００，０００を超えるバリアントを含み得る。突然変異誘発細胞のライブラリは、液体培養液中での増殖によって拡大させることができ、細胞は、集団が単一細胞からなることを確実にするために処理することができる。

【０２９０】

図１０は、分子発見を実行するための例示的なワークフローを概略的に示す。プロセスは、図４に示すプロセスと概ね同様であってもよい。プロセス１００１では、例えば、標的化または非標的化突然変異誘発を使用して、上記のように、変異体のライブラリ（または本明細書では遺伝的バリアントとも呼ばれる突然変異誘発細胞）を生成することができる。プロセス１００３では、変異体のライブラリを個々の容器（例えば、マイクロリアクタ）に分配することができる。いくつかの例では、個々の細胞または単一の細胞が個々の

容器に分配される。分配は、すべての個々の容器が所定の体積（範囲内）を含むことを保証するフィードバックループを有するコンピュータ実装方法によって支援されてもよく、これは個々の容器内の個々の細胞間の一貫した増殖を保証するのに有用であり得る。いくつかの例では、画像認識を使用して、所望の数の細胞が個々の容器に充填されることを確実にすることができる。

【０２９１】

個々の細胞を含む個々の容器を、発酵条件を模倣する条件（例えば、温度、炭素源、ｐＨ、培養液成分、通気など）に供することができる。個々の容器は、同じ組の発酵条件を含んでもよく、または異なる発酵条件を含んでもよく、または個々の容器のサブセットは、同じ組の発酵条件を含んでもよく、個々の容器の別のサブセットは、異なる組の発酵条件を含んでもよい。個々の容器内の細胞は、細胞増殖および／または目的の分子の産生を可能にする時間枠にわたって個々の容器内でインキュベートすることができる。マイクロバイオリアクタ内の増殖は、画像化およびコンピュータ画像認識によって監視することができる。

【０２９２】

プロセス１００５では、個々の容器をスクリーニングすることができ、スクリーニングに基づいて、（一連の分配された突然変異誘発細胞の）１またはそれを超える細胞を濃縮することができる（例えば、改善された細胞または所望の表現型を発現する細胞を選択することができる）。一例では、個々の容器またはその中の内容物を、試薬の添加および反応時間の制御を可能にする装置内に配置することができる。目的の分子の検出を行うことができ、検出に基づいて、１またはそれを超える細胞を選択することができる。検出および／または選択は、自動化された様式で行うことができる。例えば、本明細書に記載のシステムは、検出試薬の添加を介して分子生成および分子活性について調べることによって、どの個々の容器が目的の表現型を示す細胞を含有するかを決定することができるコンピュータ実施方法を含み得る。検出試薬の添加を使用して、目的の分子の環境（例えば、ｐＨ、イオン強度、イオンタイプ、基質、阻害剤、酵素、競合剤）を変化させることができる。このような環境の変化は、生成される分子の量および／または生成される分子の機能を反映する条件を生成するのに有用であり得る。試薬の添加後の規定の制御可能な時間に、分光シグナルを介して目的の分子について個々の容器を評価することができる。機器設計および機器のコンピュータ制御フィードバックは、タイミング、試薬添加、分子生成の評価、およびどのマイクロバイオリアクタが選択基準を満たすかに関する意思決定を制御することができる。いくつかの例では、ＰＩＤアルゴリズムを使用して、株最適化の過程での機器のドリフト（例えば、機械的機器、液体粘度、温度などのわずかな偏差に起因する）に起因する誤った選別決定を防止することができる。

【０２９３】

必要に応じて、選別後、例えば、特定のライブラリメンバー（例えば、突然変異誘発細胞）を保持するための機器の決定時、選択された細胞は、システムによってその容器から取り出され、生物の生存率を最大化する環境（例えば、十分な栄養素条件、塩条件、温度、通気下など）に保持することができる。細胞は、細胞増殖を介して拡大させることができ、必要に応じて、別の過程（例えば、プロセス１００３および１００５）のスクリーニングおよび濃縮を実施することができる。必要に応じて、行われるスクリーニングおよび濃縮の各回について、よりストリンジェントな条件を適用することができる（例えば、より低い酸素、より低い栄養素、ｐＨの変化など）。代替的または追加的に、濃縮中によりストリンジェントな性能基準（例えば、発現の増加、触媒活性など）を適用することができる。その後、濃縮された細胞を、プロセス１００７において特性評価することができる。濃縮された細胞は、例えば、シーケンシングまたは表現型スクリーニングを介して分析することができる。例えば、濃縮された細胞は、さらなる試験条件下で所望の表現型について確認するために分析することができる。

【０２９４】

いくつかの例では、プロセス１００７の特性評価は、遺伝子および／またはタンパク質

のシーケンシングを含む。例えば、濃縮された細胞をＤＮＡまたはＲＮＡシーケンシングに供して、１またはそれを超える突然変異の存在を決定もしくは同定するか、または突然変異を特性評価することができる。いくつかの例では、ＤＮＡまたはＲＮＡ配列に基づいて、ＤＮＡまたはＲＮＡ配列の得られた転写または翻訳ペプチドのアミノ酸配列を決定することができる。アミノ酸配列を親（例えば、非突然変異誘発）細胞のアミノ酸配列と比較して、目的のタンパク質のアミノ酸配列の特異的変化（例えば、ペプチドのどの残基が異なるか）を決定することができる。さらなる検証研究は、例えば、親（例えば、非突然変異誘発）細胞または娘細胞（例えば、親細胞に由来する突然変異誘発細胞）を濃縮されたバリアントの遺伝子配列で形質転換または形質導入することによって行うことができる。例えば、バリアントを濃縮し、それらの核酸またはタンパク質配列を同定し、次いで、濃縮されたバリアントの同定された核酸またはタンパク質配列を細胞が含むように、それらの変異体を（例えば、形質転換または形質導入を介して）細胞に移入するプロセス１００９は、改善された表現型または機能的性能を有する細胞をもたらすことができる。

【０２９５】

図１１は、上記のような、および図１０に示すようなワークフローを使用して分子発見を実行することからの例示的なデータを示す。分子発見は、２回の分配、濃縮、および任意の特性評価にわたって、単一世代（例えば、１サイクルの突然変異誘発）に対して行われる。図１１では、各パネルは、産生された目的の分子（例えば、分泌タンパク質）の数量を表す強度のヒストグラムを示す。一例では、産生された目的の分子は、ＳｉｇＡプロモーターによって発現が駆動され得るレポータータンパク質（例えば、緑色蛍光タンパク質）であってもよい。ＳｉｇＡの活性が高いほど、より多くのレポータータンパク質が発現され、シグナル強度が高い。非突然変異誘発細胞１１０３（例えば、突然変異の導入前の細胞）は、比較的類似したピークプロファイルを維持するが、各過程から得られた突然変異誘発細胞１１０１（例えば、分配され、培養され、スクリーニングされる変異誘発細胞）は右へのシフトを示し、分子発見プロセスの結果として、強度、したがって産生された目的の分子の数量の増加を示している。下のパネルは、２回目のゲーティングおよび選別後の突然変異誘発細胞１１０１を示しており、選別および濃縮後に、目的の表現型の増加（例えば、産生された目的の分子の数量がより多いことを表すより高い強度）が観察されることを示している。

【０２９６】

スクリーニングに続いて、またはスクリーニングと同時に、突然変異誘発細胞の集団を選別することができる。例えば、本明細書に記載のシステムは、細胞が選別される閾値機能的性能パラメータ（例えば、目的の分子の最小強度、最小濃度、最小数量など）を設定するコンピュータ可読媒体を含み得る。閾値を上回る細胞は、さらなる分析のために第１のビンで選別または濃縮され得るが、閾値を下回る細胞は、第２のビンで選別され得る。第１のビン内の細胞は、突然変異についてさらに分析して、例えば核酸またはタンパク質配列を決定することができ、これはどの突然変異が改善された機能的性能をもたらすかを決定するのに役立ち得る。

【０２９７】

図１２は、分子発見プロセス、例えば、図１０に概説されるプロセスから得られた例示的な突然変異の表を示す。各変異はＸＮＮＮＹの形式であり、Ｘは元のアミノ酸残基（例えば、非変異細胞の残基）を示し、ＮＮＮは突然変異のアミノ酸位置を示し、Ｙは変異した元のアミノ酸残基を示す。例えば、「Ｔ３１Ｒ」は、３１番目のアミノ酸残基上のトレオニンからアルギニンへの突然変異を指す。図１２の各突然変異は、ＳｉｇＡ転写因子に対するものであり、このＳｉｇＡ転写因子は、目的のタンパク質またはレポータータンパク質（例えば、ＧＦＰ）の発現を改変することができる。図１２に示す変異体は、目的のタンパク質またはレポータータンパク質のより高い発現をもたらす。突然変異は、改善された機能的性能（例えば、ＳｉｇＡ転写因子の分泌または活性の増加、したがってレポータータンパク質の発現および強度の増加）をもたらすことが実証されており、分子発見プロセス中の濃縮の結果である。

【０２９８】
図１３は、図１２の誘導された突然変異を細胞に導入する例示的なデータを示す。そのような例では、細胞は、非突然変異誘発細胞、親細胞、または突然変異誘発細胞であってもよい。細胞は、図１２で同定されたタンパク質バリアント（例えば、目的のタンパク質またはＧＦＰなどのレポータータンパク質の発現を促進し得るＳｉｇＡ変異体）を発現する遺伝子バリアントで形質転換することができる。タンパク質変異体（例えば、ＳｉｇＡ変異体）の発現は、誘導性プロモーターによって制御することができる。図１３のパネルＡは、親（例えば、非突然変異誘発細胞）を誘導可能なプロモーターを有するプラスミドベクター対照で形質転換することができる対照群を示す。発現の誘導（例えば、この例ではイソプロピルβ－ｄ－１－チオガラクトピラノシド、ＩＰＴＧの添加として示されている誘導性プロモーターをオンにすることによって）は、レポータータンパク質の蛍光シグナルの顕著な差をもたらさず、予想されるように、転写因子（例えば、ＳｉｇＡ）の発現または活性の変化がほとんどまたは全くないことを示している。パネルＢでは、転写因子（例えば、ＳｉｇＡ）を天然に産生するレポータータンパク質（例えば、発現がＳｉｇＡによって駆動されるＧＦＰ）を発現する細胞を、転写因子（例えば、ＳｉｇＡ）をコードする誘導可能な遺伝子を有するプラスミドで形質転換する。細胞によって既に天然に産生されている野生型転写因子（例えば、ＳｉｇＡ）の強制発現は、レポータータンパク質シグナルの増加をもたらし、転写因子（例えば、ＳｉｇＡ）の数量または活性のいくらかの増加を示している。パネルＣ～Ｆでは、導入された誘導ＳｉｇＡ変異体のそれぞれについて、レポータータンパク質（例えば、ＧＦＰ）の発現が増加している。パネルＣ～Ｆの各パネルでは、図１２に列挙された濃縮ＳｉｇＡ変異体のうちの１またはそれを超える変異体を含むプラスミドで細胞を形質転換する。変異体群（パネルＣ～Ｆ）のそれぞれは、誘導性プロモーターの誘導時にレポータータンパク質の発現をそれぞれ２．５倍、３．７倍、２．４倍および３．７倍増加させる。全体として、これらの結果は、本明細書に記載の分子発見システムおよび方法が、改善された機能的性能、例えばタンパク質発現または活性の増加をもたらす遺伝的バリアントの同定に有用であり得ることを示している。

【０２９９】
実施例７：機械学習を使用した分子の改善
本明細書の方法、組成物、およびシステムは、分子の改善または最適化、例えば改善された分子バリアントの生成を可能にすることができる。一例では、本明細書の方法は、ペプチドまたはタンパク質などの改善されたポリマーバリアントを産生するために使用することができる。

【０３００】
図１４は、改善されたポリマーバリアント（例えば、タンパク質）を産生するために分子の改善を実行するための例示的なワークフローを概略的に示す。プロセス１４０１では、ポリマーバリアントのライブラリが生成される。一例では、ポリマーバリアントのライブラリは、酵素をコードする遺伝子の変異体のライブラリを含み得、そのような変異体のライブラリは、突然変異誘発によって生成され得る。例えば、所与の遺伝子について、突然変異誘発を実施して、タンパク質中のすべてのアミノ酸位置が２０個すべてのアミノ酸に変更されている部位飽和ライブラリ（ＳＳＬ）を得ることができる。さらなる突然変異誘発は、例えば、変異体のライブラリを得るために遺伝子シャッフリングおよびエラープローンＰＣＲを介して行うことができる。次いで、変異体のライブラリを、細胞が単一の変異体（例えば、ポリマー、ペプチドまたはタンパク質バリアントをコードする配列）を含むように複数の細胞（例えば、適切な細菌発現宿主）に形質転換することができる。

【０３０１】
プロセス１４０３では、変異体のライブラリを個別化された容器に分配することができる。必要に応じて、容器は、細胞が増殖し、目的のポリマー（例えば、タンパク質または酵素）を産生するのに十分な条件を含み得る。数時間の増殖および産生の後、次いで、細胞をスクリーニングし、プロセス１４０５において濃縮することができる。スクリーニング中、反応緩衝液中の酵素基質を各個別化された容器に添加することができる。個々の容

器を正確な時間インキュベートし、次いで、ポリマーバリアントの機能を決定して機能データを得ることができる。例えば、酵素活性を測定してもよい（例えば、蛍光または発光生成物の量を測定することによって）。測定と同時にまたは測定に続いて、変異体のライブラリを、異なる性能（例えば、酵素活性）を表すいくつかの機能ゲート（例えば、４つのゲート）に選別することができる。いくつかの例では、変異体のライブラリのサブセットは選別されない場合があり、これは対照集団または比較（例えば、前濃縮）群として除外され得る。

【０３０２】

いくつかの例では、測定された細胞の総数は、ユニークなポリマーバリアントの数より多くてもよい。例えば、バリアント（例えば、遺伝的バリアントを含む突然変異誘発細胞）の増殖および培養後、各バリアントはバリアントのクローン集団を含み得る。単一細胞を個々の反応容器に分配した後、同じバリアントの複数のコピーが個々の反応容器内に存在し得る。したがって、アッセイされる細胞の総数は、ライブラリにおいて予想されるユニークなバリアントの数の５０～１００倍であってもよい。

【０３０３】

プロセス１４０７では、例えば、選別されたゲートのうちの１またはそれを超えるゲートからの選別された細胞をさらに分析することができる。例えば、各細胞（または細胞集団）を容器から単離し、例えば次世代シーケンシングを使用してＤＮＡをシーケンシングして配列データを得ることができる。選別されていない細胞をシーケンシングすることもできる。配列データに基づいて、濃縮前（選別されていない細胞から）および濃縮後（選別された細胞から）の両方の各変異体の存在量を計算することができる。各配列についての濃縮データ、例えば濃縮係数は、選別された細胞における各変異体の存在量を選別されていない細胞における各変異体の存在量と比較することによって計算することができる。濃縮係数は変異体の機能的性能（例えば、触媒活性）に比例するので、濃縮係数はライブラリ内の各配列の機能的標識として使用することができる。次いで、この配列／機能ラベル付きデータセットは、将来の設計のための推奨を行うために機械学習（ＭＬ）手法を使用してプロセス１４０９において入力および／または分析することができる。ＭＬアルゴリズムは、高い性能を有すると予測される何千もの設計（例えば、遺伝的バリアントまたはポリマーバリアント）を生成することができる。次いで、これらの予測されたバリアントを（例えば、オリゴヌクレオチドアレイを使用して）合成し、発現宿主への形質転換のためにプラスミドにアセンブルすることができる。次いで、分子改善または最適化プロセス（例えば、プロセス１４０１、１４０３、１４０５、１４０７、および１４０９）を数回（例えば、２～４世代）繰り返して、高度に活性な変異体を得ることができる。

【０３０４】

図１５は、分子最適化プロセスのデータ例を示す。各パネルは、その後の濃縮過程にわたる（上から下へ）生成された目的の分子の性能（例えば、酵素活性レベル）を表す強度のヒストグラムプロットを示す。非突然変異誘発細胞１５０３（例えば、突然変異の導入前の細胞）は、予想されるように比較的類似したピークプロファイルを維持するが、各過程から得られた突然変異誘発細胞１５０１（例えば、分配され、培養され、スクリーニングされる突然変異誘発細胞）は、おそらく突然変異誘発の有害な影響により、最初は目的の分子の性能（例えば、活性レベル）がより低いことを示している。その後のスクリーニング過程（例えば、分配、濃縮、および任意の特性評価）にわたって、突然変異誘発細胞１５０１は右へのピークシフトを示し、これは、分子最適化プロセスの結果として、強度、したがって目的の生成された分子の活性レベルの増加を示している。

【０３０５】

図１６は、分子最適化の世代にわたるバリアント活性の散布図を示す。図１６から分かるように、バリアント活性（ｘ軸）は数世代の過程内で増加する。

【０３０６】

実施例８：機器制御

本明細書では、システム内でまたは本明細書に記載の方法の間に発生するドリフト、ノ

イズ、および／またはクリープを補正するのに有用であり得る機器制御アルゴリズムも記載される。

【０３０７】

いくつかの例では、コンピュータ可読媒体は、分配頻度を制御する（例えば、液滴の頻度を制御する）ための１またはそれを超えるフィードバックアルゴリズムを含み得、これは、例えば、流量、圧力、温度などを自動的に調整することを含み得る。別の例では、選別されたゲート、例えば閾値パラメータは、ノイズ、クリープなどを考慮するように調整することができる。例えば、選別または濃縮中に、閾値ゲート（例えば、ライブラリの最小強度閾値または最大パーセンテージ）を設定することができる。フィードバックアルゴリズムは、収集比を維持するためにゲーティングパラメータを自動的に変更または変調することができる。

【０３０８】

図１７は、選別ゲートフィードバックシステムの例示的なデータを示す。上のプロットは、測定されたバリアントのシグナル強度を時間の積として示す。上の線１７０１および下の線１７０３はゲートを示す。シグナルが時間と共により高く発散する（時間とともにドリフトする）と、２つのゲート１７０１および１７０３の調整を行うフィードバックアルゴリズムにより、ゲート間で濃縮されたバリアントのパーセンテージが一定に維持される。下のプロットは、収集範囲内のバリアントの割合を時間の関数として示し、これは比較的一定を維持する。

【０３０９】

実施形態

場合によっては、本開示は、以下の実施形態による方法を提供する。

実施形態１．　改善されたポリマーバリアントを生成する方法であって、

（ａ）データセットを取得することであって、

ｉ）ポリマーバリアント配列のデータを含む配列データと、

ｉｉ）ポリマーバリアント機能に基づいて選別された試料中のポリマーバリアント配列の存在量に基づく濃縮データと、を含む、取得することと、

（ｂ）１またはそれを超える機械学習アルゴリズムを前記データセットに適用して、改善されたポリマーバリアントを設計することと、を含む、方法。

実施形態２．　前記ポリマーバリアント配列が、ｉ）選別されていない試料中およびｉｉ）前記ポリマーバリアント機能に基づいて選別された前記試料中の配列を含み、前記ポリマーバリアント機能に基づいて選別された前記試料および前記選別されていない試料が共通の試料に由来する、実施形態１に記載の方法。

実施形態３．　前記濃縮データが、前記選別されていない試料中の前記ポリマーバリアント配列の存在量にさらに基づく、実施形態２に記載の方法。

実施形態４．　前記ポリマーバリアント機能に基づいて選別された前記試料中の前記ポリマーバリアント配列の存在量がシーケンスカウントを含む、実施形態３に記載の方法。

実施形態５．　前記選別されていない試料中の前記ポリマーバリアント配列の存在量がシーケンスカウントを含む、実施形態３または４に記載の方法。

実施形態６．　前記濃縮データが、ｉ）前記選別されていない試料中の前記ポリマーバリアント配列中の前記第１のポリマーバリアント配列の存在量、およびｉｉ）前記ポリマーバリアント機能に基づいて選別された前記試料の前記サブセット中の前記ポリマーバリアント配列中の前記第１のポリマーバリアント配列の存在量に基づいて、第１のポリマーバリアント配列の濃縮係数を含む、実施形態３から５のいずれか一つに記載の方法。

実施形態７．　前記濃縮係数が、前記第１のポリマーバリアント配列の機能的標識である、実施形態６に記載の方法。

実施形態８．　前記方法が、（ａ）および（ｂ）の複数世代を実施することを含み、（ｂ）の前記改善されたポリマーバリアントを使用して、前記複数世代の後続世代において（ａ）の前記データセットを生成する、実施形態１から７のいずれか一つに記載の方法。

実施形態９．　前記複数世代が、少なくとも３世代または４世代である、実施形態８に

記載の方法。
実施形態１０．前記ポリマーバリアントがタンパク質を含む、実施形態１から９のいずれか一つに記載の方法。
実施形態１１．前記タンパク質が酵素を含む、実施形態１０に記載の方法。
実施形態１２．前記機能が前記酵素の活性を含む、実施形態１１に記載の方法。
実施形態１３．前記改善されたポリマーバリアントが、少なくとも１０００の改善されたポリマーバリアントを含む、実施形態１から１２のいずれか一つに記載の方法。
実施形態１４．前記ポリマーバリアント配列が、前記ポリマーバリアントをコードする核酸配列を含む、実施形態１から１３のいずれか一つに記載の方法。
実施形態１５．前記データセットを取得する前に、前記ポリマーバリアントを含む細胞のライブラリを提供することをさらに含み、前記ポリマーバリアントが複数のポリマーバリアントを含む、実施形態１に記載の方法。
実施形態１６．前記ポリマーバリアントを含む細胞の前記ライブラリを生成することをさらに含む、実施形態１５に記載の方法。
実施形態１７．前記生成することが、各位置に２０個の天然に存在するアミノ酸のそれぞれを有するタンパク質をコードする部位飽和核酸分子ライブラリ（ＳＳＬ）を提供することを含む、実施形態１６に記載の方法。
実施形態１８．前記ポリマーバリアントをコードする核酸分子を生成するために、前記ＳＳＬに対して遺伝子シャッフリングおよびエラープローンポリメラーゼ連鎖反応（ｅｐＰＣＲ）を実施することをさらに含む、実施形態１７に記載の方法。
実施形態１９．前記ポリマーバリアントをコードする前記核酸分子を細胞に形質転換して、前記細胞のライブラリを生成することをさらに含む、実施形態１８に記載の方法。
実施形態２０．前記細胞が細菌細胞を含む、実施形態１９に記載の方法。
実施形態２１．複数のパーティションを生成するために前記細胞のライブラリを分配することをさらに含む、実施形態１５に記載の方法。
実施形態２２．前記複数のパーティションの各パーティションが、平均して、１つ以下の細胞を含む、実施形態２１に記載の方法。
実施形態２３．前記複数のパーティションにおいて前記細胞のライブラリを拡大することをさらに含む、実施形態２１に記載の方法。
実施形態２４．前記ポリマーバリアントを前記複数のパーティション内の細胞のライブラリにおいて発現させることをさらに含む、実施形態２３に記載の方法。
実施形態２５．前記ポリマーバリアントを含む細胞の前記ライブラリを使用して前記ポリマーバリアント機能を決定することをさらに含む、実施形態１５に記載の方法。
実施形態２６．前記ポリマーバリアントがタンパク質を含む、実施形態２５に記載の方法。
実施形態２７．前記タンパク質が酵素を含む、実施形態２６に記載の方法。
実施形態２８．前記ポリマーバリアント機能を決定することが、酵素基質および反応緩衝液を使用することを含む、実施形態２７に記載の方法。
実施形態２９．前記ポリマーバリアント機能が酵素活性を含む、実施形態２８に記載の方法。
実施形態３０．前記酵素活性が蛍光生成物を生成する、実施形態２９に記載の方法。
実施形態３１．前記ポリマーバリアントを含む前記細胞ライブラリのサブセットを前記ポリマーバリアント機能に基づいて選別して、前記ポリマーバリアント機能に基づいて選別された前記試料および選別されていない試料を提供することをさらに含む、実施形態１５に記載の方法。
実施形態３２．前記機能が酵素活性の量である、実施形態３１に記載の方法。
実施形態３３．前記選別することが、前記細胞のライブラリの前記サブセットを、前記機能に基づいて複数のグループに選別することを含む、実施形態３１に記載の方法。
実施形態３４．前記複数のグループが少なくとも４つのグループである、実施形態３２に記載の方法。

実施形態３５．　前記細胞のライブラリのサブセットが、前記細胞のライブラリに存在すると予想される数の約５０倍～約１００倍のユニークなポリマーバリアントを含む、実施形態３１に記載の方法。
実施形態３６．　前記ポリマーバリアント機能に基づいて選別された前記試料を使用して前記ポリマーバリアント配列を決定することをさらに含む、実施形態３１に記載の方法。
実施形態３７．　前記ポリマーバリアント機能に基づいて選別された前記試料が基準配列を含む、実施形態３６に記載の方法。
実施形態３８．　前記ポリマーバリアント機能に基づいて選別された前記試料を使用して前記ポリマーバリアント配列を決定することが、前記ポリマーバリアント機能に基づいて選別された前記試料から核酸分子を精製することを含む、実施形態３６に記載の方法。
実施形態３９．　前記選別されていない試料から核酸分子を精製することをさらに含む、実施形態３１に記載の方法。
実施形態４０．　前記ポリマーバリアント機能に基づいて選別された前記試料から核酸分子を精製することが、前記選別されていない試料から核酸分子を精製することとは別個に生じる、実施形態３９に記載の方法。
実施形態４１．　前記ポリマーバリアント機能に基づいて選別された前記試料からの前記核酸分子および前記選別されていない試料からの前記核酸分子をシーケンシングして、前記ポリマーバリアント配列を提供することをさらに含む、実施形態３１に記載の方法。
実施形態４２．　前記シーケンシングすることが次世代シーケンシングすることを含む、実施形態４１に記載の方法。
実施形態４３．　前記ポリマーバリアント配列を使用して、ｉ）前記ポリマーバリアント機能に基づいて選別された前記試料中およびｉｉ）前記選別されていない試料中の前記ポリマーバリアント配列の存在量を決定することをさらに含む、実施形態４１に記載の方法。
実施形態４４．　ｉ）前記ポリマーバリアント機能に基づいて選別された前記試料中およびｉｉ）前記選別されていない試料中の前記ポリマーバリアント配列の存在量を使用して、前記濃縮データを決定することをさらに含む、実施形態４３に記載の方法。
実施形態４５．　前記濃縮データが濃縮係数を含む、実施形態４４に記載の方法。
実施形態４６．　前記改善されたポリマーバリアントを合成することをさらに含む、実施形態１に記載の方法。
実施形態４７．　前記合成することが、オリゴヌクレオチドアレイ上での合成を含む、実施形態４６に記載の方法。
実施形態４８．　前記改善されたポリマーバリアントをベクターにアセンブリすることをさらに含む、実施形態４７に記載の方法。
実施形態４９．　前記ベクターがプラスミドである、実施形態４８に記載の方法。
実施形態５０．　前記ベクターを宿主細胞に導入することをさらに含む、実施形態４８に記載の方法。
実施形態５１．　前記複数世代のうちの第１世代の（ａ）が、飽和突然変異誘発およびｅｐＰＣＲを実施して細胞のライブラリを提供することを含み、前記複数世代のうちの第２世代の（ａ）が、ｅｐＰＣＲを実施することを含み、前記複数世代のうちの第３世代の（ａ）が、ｅｐＰＣＲを実施することを含む、実施形態８に記載の方法。
実施形態５２．　前記複数世代の各世代が、複数回の前記選別を含む、実施形態５１に記載の方法。
実施形態５３．　コンピュータプロセッサによって実行されると前記コンピュータプロセッサに実施形態１から１４のいずれか一つに記載の方法を実行させる命令を含む、非一時的なコンピュータ可読媒体。
実施形態５４．　コンピュータプロセッサと、実施形態５３に記載の非一時的なコンピュータ可読媒体とを含むシステム。
実施形態５５．　改善された細胞を生成するための方法であって、

（ａ）突然変異誘発細胞のライブラリを生成することと、
（ｂ）前記突然変異誘発細胞のライブラリを複数のパーティションに分配して、分配された変異誘発細胞を得ることと、
（ｃ）前記分配された突然変異誘発細胞を表現型についてスクリーニングすることと、
（ｄ）前記スクリーニングに基づいて、閾値ゲートを使用して、前記分配された突然変異誘発細胞から改善された細胞を選択することと、
（ｅ）（ｂ）のフィードバックコントローラ（ｉ）を使用して前記複数のパーティションの生成頻度を維持すること、または（ｄ）のフィードバックコントローラ（ｉｉ）を使用して前記閾値ゲートを調節することと、
（ｆ）前記改善された細胞を使用して工程（ａ）～（ｄ）を繰り返すことと、を含む、方法。

実施形態５６．　前記方法が、前記フィードバックコントローラを使用して前記複数のパーティションの生成頻度を維持することを含む、実施形態５５に記載の方法。

実施形態５７．　前記複数のパーティションが複数の液滴を含む、実施形態５６に記載の方法。

実施形態５８．　前記フィードバックコントローラが、流量を調整して前記複数の液滴の生成頻度を維持する、実施形態５７に記載の方法。

実施形態５９．　前記方法が、前記フィードバックコントローラを使用して前記閾値ゲートを調節することを含む、実施形態５５に記載の方法。

実施形態６０．　前記フィードバックコントローラが、前記閾値ゲートを上昇または下降させる、実施形態５９に記載の方法。

実施形態６１．　前記フィードバックコントローラが、前記閾値ゲートを調節して、前記分配された突然変異誘発細胞の標的パーセンテージを選択する、実施形態５９に記載の方法。

実施形態６２．　前記フィードバックコントローラが、比例・積分・微分（ＰＩＤ）アルゴリズムである、実施形態５５に記載の方法。

実施形態６３．　前記フィードバックコントローラが、機械的機器、液体粘度、または温度の偏差を考慮する、実施形態５９に記載の方法。

実施形態６４．　突然変異誘発細胞の前記ライブラリを生成することが、前記細胞を変異原と接触させることを含む、実施形態５５に記載の方法。

実施形態６５．　突然変異誘発細胞の前記ライブラリを生成することが、トランスポゾン突然変異誘発、多重自動ゲノム工学法（ｍａｇｅ）、ゲノムシャッフリング、ランダム組換え、または非相同末端結合を含む、実施形態５５に記載の方法。

実施形態６６．　突然変異誘発細胞の前記ライブラリを生成することが、異種遺伝要素を細胞に導入することを含む、実施形態５５に記載の方法。

実施形態６７．　前記異種遺伝要素が、プロモーター、５’非翻訳領域、リボザイム、ＲＮＡ安定性配列、分泌ペプチドをコードする核酸、シグナルペプチドをコードする核酸、融合タンパク質をコードする核酸、ＤＮＡ結合ドメインをコードする核酸、タンパク質－タンパク質結合領域をコードする核酸、コドン最適化遺伝子をコードする核酸、３’非翻訳領域、またはターミネーターを含む、実施形態５５に記載の方法。

実施形態６８．　前記突然変異誘発細胞が、少なくとも１０００個の細胞、少なくとも１０，０００個の細胞、または少なくとも１，０００，０００個の細胞を含む、実施形態５５に記載の方法。

実施形態６９．　（ｂ）の前に突然変異誘発細胞のライブラリを拡大増殖することをさらに含む、実施形態５５に記載の方法。

実施形態７０．　前記拡大増殖が液体培養で起こる、実施形態６７に記載の方法。

実施形態７１．　前記分配がコンピュータ可読媒体によって制御される、実施形態５５に記載の方法。

実施形態７２．　前記複数のパーティションの各パーティションが、所定の範囲内の液体の体積を含む、実施形態５５に記載の方法。

実施形態７３．（ｃ）の前に複数のパーティションの各パーティションを画像化することをさらに含む、実施形態５５に記載の方法。
実施形態７４．前記画像化に基づいて、前記突然変異誘発細胞の所望の数の突然変異誘発細胞を含む前記複数のパーティションの分配を確実にすることをさらに含む、実施形態７１に記載の方法。
実施形態７５．前記複数のパーティションを（ｂ）と（ｃ）との間の細胞増殖条件に供することをさらに含む、実施形態５５に記載の方法。
実施形態７６．前記供することが、前記複数のパーティションのサブセットのパーティションの温度、炭素源、ｐＨ、培養液成分、または通気を制御することを含む、実施形態５５に記載の方法。
実施形態７７．画像化を使用して前記細胞増殖条件に供された前記複数のパーティションにおける細胞増殖を監視することをさらに含む、実施形態５５に記載の方法。
実施形態７８．前記スクリーニングが、前記複数のパーティションの各パーティションに検出試薬を添加することを含む、実施形態５５に記載の方法。
実施形態７９．前記検出試薬が、前記複数のパーティションの各パーティションの環境を変化させる、実施形態７６に記載の方法。
実施形態８０．前記検出試薬が、ｐＨ調整剤、イオン強度調整剤、イオン、基質、阻害剤および酵素、または競合剤を含む、実施形態７６に記載の方法。
実施形態８１．前記突然変異誘発細胞のライブラリが、目的の分子をコードする、実施形態５５に記載の方法。
実施形態８２．前記スクリーニングが、目的の分子を検出すること、目的の分子の活性を測定することであって、必要に応じて活性が触媒活性であり、必要に応じて触媒活性が蛍光生成物を生成する、測定すること、目的の分子の量を決定すること、または目的の分子の機能を決定することを含む、実施形態７９に記載の方法。
実施形態８３．前記スクリーニングが、複数のスクリーニングを実施することを含む、実施形態５５に記載の方法。
実施形態８４．前記複数のスクリーニングが、複数のスクリーニングの前のスクリーニングと比較して、複数のスクリーニングの後続のスクリーニングにおいて、目的の表現型に対してよりストリンジェントな閾値を適用することを含む、実施形態５５に記載の方法。
実施形態８５．前記パーティションを選択することが、分光測定を実行することを含む、実施形態５５に記載の方法。
実施形態８６．前記パーティションを選択することが、閾値を適用することを含む、実施形態５５に記載の方法。
実施形態８７．（ｆ）が、選択されたパーティションから突然変異誘発細胞を取り出して、改善された細胞を得ることを含む、実施形態５５に記載の方法。
実施形態８８．前記突然変異誘発細胞の生存を維持することをさらに含む、実施形態８７に記載の方法。
実施形態８９．前記突然変異誘発細胞を増殖させることをさらに含む、実施形態８７に記載の方法。
実施形態９０．目的の表現型について改善された細胞をスクリーニングすることをさらに含む、実施形態５５に記載の方法。
実施形態９１．前記突然変異誘発細胞は細菌を含む、実施形態５５に記載の方法。
実施形態９２．コンピュータプロセッサによって実行されると、前記コンピュータプロセッサに実施形態５５に記載の工程（ｂ）～（ｅ）を実行させる命令を含む、コンピュータ可読媒体。
実施形態９３．コンピュータプロセッサと、実施形態９２に記載のコンピュータ可読媒体と含むシステム。
実施形態９４．株最適化ユニット、分子最適化ユニットおよび分子発見ユニットからなる群から選択される２またはそれを超える統合ユニットを含む所望の生物学的バリアン

トを選別するための統合システムであって、ここで、（ａ）前記株最適化ユニットが、それぞれが異なる株最適化遺伝的バリアントまたは発酵条件を有する複数の別個の株最適化反応容器を含み、（ｂ）前記分子最適化ユニットが、それぞれが異なる標的タンパク質バリアントまたは適用条件を有する複数の別個の分子最適化反応容器を含み、（ｃ）前記分子発見ユニットが、それぞれが異なる発見分子または適用条件を有する複数の別個の分子発見反応容器を含む、統合システム。

実施形態９５。前記２またはそれを超えるユニットが、以下の機能である（ｉ）データリポジトリ内の前記２またはそれを超える統合ユニットのそれぞれからのデータセットの記憶、（ｉｉ）前記データリポジトリに配信されるデータの閾値パラメータの設定、（ｉｉｉ）入力された多様性スクリーニングを繰り返すための前記２またはそれを超えるユニットのそれぞれの指示のうちの１またはそれを超える機能の実行ができるコンピュータ可読媒体を介して統合される、実施形態９４に記載の統合システム。

実施形態９６。前記２またはそれを超えるユニットが、前記選別された所望の生物学的バリアントについて前記２またはそれを超えるユニットのそれぞれからのデータを統合する機械学習アルゴリズムを含むコンピュータ可読媒体を介して統合される、実施形態９４に記載の統合システム。

実施形態９７。前記機械学習アルゴリズムが、エラスティックネット回帰一般化線形モデル（ＧＬＭＮＥＴ）、サポートベクターマシン回帰（ＳＶＭ）、ランダムフォレスト（ＲＦ）、エクストリーム勾配ブースティング（ＸＧＢｏｏｓｔ）、多層パーセプトロン（ＭＬＰ）、または畳み込みニューラルネットワーク（ＣＮＮ）を含む、実施形態９６に記載の統合システム。

実施形態９８。前記株最適化反応容器、前記分子最適化反応容器、および／または前記分子発見反応容器のそれぞれは、平均で１個以下の細胞を含む、実施形態９４に記載の統合システム。

実施形態９９。前記株最適化反応容器、前記分子最適化反応容器、および／または前記分子発見反応容器のそれぞれが、平均で１個を超える細胞を含む、実施形態９４に記載の統合システム。

実施形態１００。ウェルあたり１個を超える細胞の前記平均がクローン集団である、実施形態９９に記載の統合システム。

実施形態１０１。ウェルあたり１個を超える細胞の前記平均が多様な細胞集団を含む、実施形態９９に記載の統合システム。

実施形態１０２。前記株最適化反応容器、前記分子最適化反応容器、および／または前記分子発見反応容器のそれぞれが、平均で１０～１００個の細胞を含む、実施形態９４に記載の統合システム。

実施形態１０３。前記株最適化ユニット、前記分子最適化ユニット、および前記分子発見ユニットが、単一細胞分析を実行するように構成される、実施形態９７に記載の統合システム。

実施形態１０４。前記株最適化ユニット、前記分子最適化ユニット、または前記分子発見ユニットが、多細胞解析を実行するように構成される、実施形態９９に記載の統合システム。

実施形態１０５。前記株最適化ユニット、前記分子最適化ユニットおよび／または前記分子発見ユニットが、少なくとも毎分１００個の細胞をスクリーニングするように構成される、実施形態９４に記載の統合システム。

実施形態１０６。前記２またはそれを超える統合ユニットのそれぞれは、消費ユニットを含む、実施形態９４に記載の統合システム。

実施形態１０７。前記消費ユニットがマイクロ流体デバイスを含む、実施形態１０６に記載の統合システム。

実施形態１０８。前記マイクロ流体デバイスが、液滴発生器、液滴凝縮器、緩衝液リザーバ、細胞インキュベータ、流体注入器、液滴スプリッター、液滴選別機、ディレイライン、および液滴シンギュレータを含む群から選択される１またはそれを超える機能要素

を含む、実施形態１０７に記載の統合システム。
実施形態１０９。 前記株最適化反応容器、前記分子最適化反応容器、または前記分子発見反応容器の反応容器が、１マイクロリットル未満の体積を有する、実施形態９４に記載の統合システム。
実施形態１１０。 前記株最適化反応容器、前記分子最適化反応容器、および／または前記分子発見反応容器の反応容器が、ナノウェルアレイ内にある、実施形態９４に記載の統合システム。
実施形態１１１。 前記株最適化反応容器、前記分子最適化反応容器、および／または前記分子発見反応容器の反応容器が、液滴、プラグ、またはウェルである、実施形態９４に記載の統合システム。
実施形態１１２。 前記１またはそれを超える当該ユニットに結合された検出器をさらに含み、前記検出器が、質量分析計、ラマン分光計、顕微鏡、セルカウンタ、蛍光顕微鏡、光学顕微鏡、フローサイトメータ、質量分析計、蛍光プレートリーダ、近赤外（ＮＩＲ）分光光度計、および圧電センサからなる群から選択される、実施形態９４に記載の統合システム。
実施形態１１３。 ロボットマニピュレータ、フロージェネレータ、音響ドロップジェネレータ、光ピンセット、熱ドロップオンデマンド、圧電ドロップオンデマンドからなる群から選択される反応容器移動アクチュエータをさらに含む、実施形態９４に記載の統合システム。
実施形態１１４。 前記異なる株最適化遺伝的バリアントのそれぞれ、前記標的タンパク質バリアントのそれぞれが異なる細胞内にあり、および／または前記標的発見分子のそれぞれが異なる細胞内にある、実施形態９４に記載の統合システム。
実施形態１１５。 各株最適化遺伝的バリアントが、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、ＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、またはＣＲＩＳＰＲガイドＤＮＡポリメラーゼによって生成される、実施形態９４に記載の統合システム。
実施形態１１６。 前記ランダム突然変異誘発が、飽和突然変異誘発である、実施形態１１５に記載の統合システム。
実施形態１１７。 各標的タンパク質バリアントが、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、ＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、またはＣＲＩＳＰＲガイドＤＮＡポリメラーゼによって生成される、実施形態９４に記載の統合システム。
実施形態１１８。前記ランダム突然変異誘発が、飽和突然変異誘発または循環置換である、実施形態１１７に記載の統合システム。
実施形態１１９。 各発見分子が、メタゲノムライブラリ、ＤＮＡコード化ライブラリ、１またはそれを超える微生物からのＤＮＡのライブラリ、またはファンクションジェネレータで生成されたライブラリに見られるＤＮＡコード化分子である、実施形態９４に記載の統合システム。
実施形態１２０。 前記株最適化ユニットが、第１の株の複数の異なる遺伝的バリアントを含む株最適化ライブラリを含む、実施形態９４に記載の統合システム。
実施形態１２１。 前記複数の異なる遺伝的バリアントが、代謝経路によって分類される、実施形態１２０に記載の統合システム。
実施形態１２２。 前記株最適化ライブラリが、表現型、配列および代謝経路データを含む、実施形態１２１に記載の統合システム。
実施形態１２３。 前記株最適化ユニットが、機械学習アルゴリズムを含む、実施形態１２２に記載の統合システム。
実施形態１２４。 前記機械学習アルゴリズムが、複数の発酵条件にわたって最適化された遺伝的バリアントを同定する、実施形態１２３に記載の統合システム。

実施形態１２５。 前記２またはそれを超えるユニットのそれぞれが、生物学的多様性または条件付き多様性の追加の入力を含む再帰的サイクルを含む、実施形態９４に記載の統合システム。

実施形態１２６。 前記複数の別々の反応体積が、少なくとも10^5個の別個の反応体積を含む、実施形態９４に記載の統合システム。

実施形態１２７。 前記適用条件が、ｐＨ、基質、反応緩衝液、および温度のうちの１またはそれを超えるものを含む、実施形態９４に記載の統合システム。

実施形態１２８。 前記発酵条件が、ｐＨ、酸素化、炭素源、緩衝液濃度、および温度のうちの１またはそれを超えるものを含む、実施形態９４に記載の統合システム。

実施形態１２９。 前記株最適化ライブラリが、前記第１の株の少なくとも１００個の異なるバリアントを含む、実施形態１２０に記載の統合システム。

実施形態１３０。 前記株最適化ライブラリが、株のランダム突然変異誘発によって生成される、実施形態１２０に記載の統合システム。

実施形態１３１。 前記株最適化ライブラリが、全突然変異誘発ライブラリ、ゲノムシャッフルライブラリ、標的ゲノムライブラリおよびトランスポゾンライブラリからなる群から選択される、実施形態１２０に記載の統合システム。

実施形態１３２。 前記全ゲノム突然変異誘発ライブラリが、全ゲノムランダム突然変異誘発ライブラリである、実施形態１３１に記載の統合システム。

実施形態１３３。 前記標的ゲノムライブラリが、プロモータースワップライブラリである、実施形態１３１に記載の統合システム。

実施形態１３４。 前記トランスポゾンライブラリが、プロモーター挿入ライブラリである、実施形態１３１に記載の統合システム。

実施形態１３５。 前記分子最適化ユニットが、第１の分子の複数の異なる遺伝的バリアントを含む分子最適化ライブラリを含む、実施形態９４に記載の統合システム。

実施形態１３６。 前記分子最適化ライブラリが、単一遺伝子ランダム突然変異ライブラリ、部位飽和ライブラリ、小型インサートメタゲノムライブラリ、および大型インサートメタゲノムライブラリからなる群から選択される、実施形態１３５に記載の統合システム。

実施形態１３７。 所望の生物学的バリアントを選別するための統合システムであって、生物学的多様性データストリームと条件付き多様性データストリームとを含み、（ａ）前記データストリームのそれぞれは独立してプログラム可能であり、（ｂ）前記生物学的多様性データストリームは標的タンパク質または前記標的タンパク質を発現する株の異なるゲノムバリアントのスクリーニングからのデータを含み、（ｃ）前記条件付き多様性データストリームは前記株と前記標的タンパク質の異なる条件のスクリーニングからのデータを含み、（ｄ）前記生物学的多様性データストリームおよび前記条件付き多様性データストリームは標的パラメータを受け取るように構成された最適化ユニットに統合され、ｉ）前記標的パラメータと一致する標的タンパク質または標的株が同定されるまで、追加の生物学的多様性または条件付き多様性を導入するように前記最適化ユニットに指示するか、またはｉｉ）前記標的パラメータと一致する標的タンパク質または標的株を同定する、統合システム。

実施形態１３８。 前記最適化ユニットが、（ｉ）前記生物学的多様性データストリームおよび前記条件付き多様性データストリームのそれぞれからのデータセットをデータリポジトリに記憶する機能、（ｉｉ）前記データリポジトリに配信されるデータの閾値パラメータを設定する機能、（ｉｉｉ）入力された多様性スクリーニングを繰り返すように前記最適化ユニットに指示する機能のうちの１またはそれを超える機能を実行することができるコンピュータ可読媒体を含む、実施形態１３７に記載の統合システム。

実施形態１３９。 前記最適化ユニットが、生物学的バリアントの前記所望の選別のために前記生物学的多様性データストリームおよび前記条件付き多様性データストリームを統合する機械学習アルゴリズムを含むコンピュータ可読媒体を含む、実施形態１３７に記載の統合システム。

実施形態１４０。前記機械学習アルゴリズムが、エラスティックネット回帰一般化線形モデル（ＧＬＭＮＥＴ）、サポートベクターマシン回帰（ＳＶＭ）、ランダムフォレスト（ＲＦ）、エクストリーム勾配ブースティング（ＸＧＢｏｏｓｔ）、多層パーセプトロン（ＭＬＰ）、または畳み込みニューラルネットワーク（ＣＮＮ）を実行する、実施形態１３９に記載の統合システム。

実施形態１４１。前記株最適化ユニットが、前記標的タンパク質または前記標的タンパク質を発現する前記株の１００個の異なるゲノムバリアントを毎分スクリーニングするように構成される、実施形態１３７に記載の統合システム。

実施形態１４２。前記株最適化ユニットが、毎分１００個の細胞をスクリーニングするように構成される、実施形態１３７に記載の統合システム。

実施形態１４３。前記株最適化ユニットのそれぞれが消費ユニットを備える、実施形態１３７に記載の統合システム。

実施形態１４４。前記消費ユニットが、マイクロ流体デバイスを含む、実施形態１４３に記載の統合システム。

実施形態１４５。前記マイクロ流体デバイスが、液滴発生器、液滴凝縮器、緩衝液リザーバ、細胞インキュベータ、流体注入器、液滴スプリッター、液滴選別機、ディレイライン、および液滴シンギュレータを含む群から選択される１またはそれを超える機能要素を含む、実施形態１４４に記載の統合システム。

実施形態１４６。前記マイクロ流体デバイスが、１またはそれを超える反応容器を含む、実施形態１４４に記載の統合システム。

実施形態１４７。前記反応容器が、１マイクロリットル未満の体積を有する、実施形態１４６に記載の統合システム。

実施形態１４８。前記反応容器がナノウェルアレイ内にある、実施形態１４６に記載の統合システム。

実施形態１４９。前記反応容器が、液滴、プラグまたはウェルである、実施形態１４６に記載の統合システム。

実施形態１５０。前記液滴が油中水型液滴である、実施形態１４９に記載の統合システム。

実施形態１５１。前記液滴が表面上の水滴である、実施形態１４９に記載の統合システム。

実施形態１５２。前記反応容器が、２またはそれを超える反応容器ごとに平均約１個の細胞が存在するような細胞を含む、実施形態１４６に記載の統合システム。

実施形態１５３。前記最適化ユニットに結合された検出器をさらに含み、前記検出器は、質量分析計、ラマン分光計、顕微鏡、セルカウンタ、蛍光顕微鏡、光学顕微鏡、フローサイトメータ、質量分析計、蛍光プレートリーダ、近赤外（ＮＩＲ）分光光度計、および圧電センサからなる群から選択される、実施形態１３７に記載の統合システム。

実施形態１５４。ロボットマニピュレータ、フロージェネレータ、音響ドロップジェネレータ、光ピンセット、熱ドロップオンデマンド、圧電ドロップオンデマンドからなる群から選択される反応容器移動アクチュエータをさらに含む、実施形態１３７に記載の統合システム。

実施形態１５５。前記標的タンパク質の異なるゲノムバリアントのそれぞれが、異なる細胞内にある、実施形態１５５に記載の統合システム。

実施形態１５６。前記株の異なるゲノムバリアントのそれぞれが、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、またはＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、またはＣＲＩＳＰＲガイドＤＮＡポリメラーゼによって生成される、実施形態１３７に記載の統合システム。

実施形態１５７。前記ランダム突然変異誘発が、飽和突然変異誘発または循環置換である、実施形態１５６に記載の統合システム。

実施形態１５８。　前記標的タンパク質バリアントの異なるゲノムバリアントのそれぞれが、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、またはＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）によって生成される、実施形態１３７に記載の統合システム。

実施形態１５９。　前記ランダム突然変異誘発が、飽和突然変異誘発、循環置換である、実施形態１５８に記載の統合システム。

実施形態１６０。　最終的な望ましい遺伝的バリアントを選別するための方法であって、標的タンパク質の遺伝要素および株の遺伝要素を含む生物学的に多様な遺伝要素の多様性スクリーニングからのデータを、標的タンパク質発現条件および標的タンパク質性能条件を含む多様な条件の条件付きスクリーニングからのデータと組み合わせること、前記組み合わせデータセットを使用して、前記最終的な望ましい遺伝的バリアントを同定することを含む、方法。

実施形態１６１。　前記条件付きスクリーニングが、多様性スクリーニングの一部である１またはそれを超える遺伝要素に対して行われる、実施形態１６０に記載の方法。

実施形態１６２。　各細胞が第１の遺伝要素のライブラリからの異なる遺伝要素を含む第１の細胞のライブラリを生成することをさらに含む、実施形態１６０に記載の方法。

実施形態１６３。　各細胞が第２の遺伝要素のライブラリからの異なる遺伝要素を含む第２の細胞のライブラリを生成することをさらに含み、第１の遺伝要素のライブラリおよび第２の遺伝要素のライブラリが、複数の条件下で独立してスクリーニングされる、実施形態１６２に記載の方法。

実施形態１６４。　前記第１の遺伝要素のライブラリおよび第２の遺伝要素のライブラリが、次世代シーケンシング（ＮＧＳ）によるシーケンシングシステムを使用してスクリーニングされ、前記複数の条件下で、濃縮されている突然変異が同定される、実施形態１６３に記載の方法。

実施形態１６５。　前記多様性スクリーニングが、株最適化に関連するバリアントに対して実行される、実施形態１６０に記載の方法。

実施形態１６６。　前記株最適化が、収量の増加、発酵条件の改善、生存率の増加、イオン濃度に対する耐性の増加をもたらす、実施形態１６５に記載の方法。

実施形態１６７。　前記収量の増加が、タンパク質発現の増加および／または細胞増殖の増加である、実施形態１６６に記載の方法。

実施形態１６８。　前記最終的な望ましい遺伝的バリアントを同定するために前記組み合わせデータセットを使用することが、（ａ）望ましい特性を有する１またはそれを超える中間的な望ましい遺伝的バリアントを同定すること、（ｂ）前記望ましい特性を有する前記１またはそれを超える中間的な望ましい遺伝的バリアントの前記遺伝要素に基づいて組換えライブラリを得ること、（ｃ）多様な条件の前記条件付きスクリーニングにより前記組換えライブラリをスクリーニングすること、および（ｄ）前記最終的な望ましい遺伝的バリアントを同定すること、または（ａ）望ましい特性を有する１またはそれを超える中間的な望ましい遺伝的バリアントを同定すること、の再帰的サイクルをさらに含む、実施形態１６０に記載の方法。

実施形態１６９。　前記多様性スクリーニングが、標的タンパク質発見に関連するバリアントに対して実行される、実施形態１６０に記載の方法。

実施形態１７０。　前記多様性スクリーニングが、標的タンパク質最適化に関連するバリアントに対して実行される、実施形態１６０に記載の方法。

実施形態１７１。　前記生物学的に多様な遺伝要素のそれぞれが並行してスクリーニングされる、実施形態１６０に記載の方法。

実施形態１７２。　前記生物学的に多様な遺伝要素のそれぞれが連続してスクリーニングされる、実施形態１６０に記載の方法。

実施形態１７３。　前記組み合わせデータセットが、蛍光発生アッセイ、比色アッセイ、およびレポーター細胞ベースのアッセイの１またはそれを超えるアッセイからのデータを含む、実施形態１６０に記載の方法。

実施形態１７４。　前記蛍光発生アッセイが、直接アッセイ、間接アッセイ、結合アッセイ、またはインビボレポーターアッセイである、実施形態１７３に記載の方法。
実施形態１７５。　前記条件付きスクリーニングが、温度、培養液、酸素化、塩分、ｐＨ、混合、反応時間、金属イオン濃度、添加剤濃度、供給速度、またはそれらの組み合わせを調整することを含む、実施形態１６０に記載の方法。
実施形態１７６。　前記条件付きスクリーニングが、第１の遺伝要素に対して２またはそれを超える異なる条件をタンデムで実行することを含む、実施形態１６０に記載の方法。
実施形態１７７。　前記多様な条件のそれぞれが並行して実行される、実施形態１６０に記載の方法。
実施形態１７８。　前記多様な条件のそれぞれが直列で（in series）実行される、実施形態１６０に記載の方法。
実施形態１７９。　多様な遺伝要素のライブラリを生成することをさらに含む、実施形態１６０に記載の方法。
実施形態１８０。　前記生成することが、ランダム突然変異誘発、部位特異的突然変異誘発、部位飽和またはゲノムシャッフリングによって行われる、実施形態１７９に記載の方法。
実施形態１８１。　前記多様な遺伝要素のライブラリからの各遺伝要素を細胞または無細胞系に導入することをさらに含む、実施形態１７９に記載の方法。
実施形態１８２。　前記多様な遺伝要素のライブラリからの遺伝的多様性を有する細胞を、各反応容器が１個以下の細胞を有するように、異なる反応容器に分離することをさらに含む、実施形態１７９に記載の方法。
実施形態１８３。　前記反応容器が、液滴、プラグまたはウェルである、実施形態１８２に記載の方法。
実施形態１８４。　前記所望の遺伝的バリアントが、株最適化、標的分子最適化、または環境ＤＮＡもしくはＲＮＡライブラリからの分子発見に関連する、実施形態１６０に記載の方法。
実施形態１８５。　前記生物学的多様性が、株最適化のものである、実施形態１６０に記載の方法。
実施形態１８６。　前記最終的な望ましい遺伝的バリアントを同定することが、発現レベル、細胞生存率、細胞増殖、および二価カチオン濃度に対する感受性を含む株の１またはそれを超える特性を測定することを含む、実施形態１６０に記載の方法。
実施形態１８７。　部位特異的突然変異誘発、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、ＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）、トランスポゾンベースの突然変異誘発、ゲノムシャッフリング、ランダム組換え、非相同末端結合、追跡可能な多重組み換え（ＴＲＭＲ）、またはＣＲＩＳＰＲガイドＤＮＡポリメラーゼによって改変株のライブラリを生成することをさらに含む、実施形態１６２に記載の方法。
実施形態１８８。　前記ランダム突然変異誘発が、飽和突然変異誘発である、実施形態１８７に記載の方法。
実施形態１８９。　前記生物学的多様性が、標的タンパク質最適化のものである、実施形態１６０に記載の方法。
実施形態１９０。　前記望ましい遺伝的バリアントを同定することが、基質特異性、基質状態、二価カチオン濃度に対する感受性、温度耐性、触媒性能、プロテアーゼ分解に対する耐性、ｐＨ耐性、阻害剤に対する感受性、補因子優先性、酵素速度、半減期、支持体に固定化された場合の活性、またはガスに対する感受性を含む標的タンパク質の１またはそれを超える特性を測定することを含む、実施形態１８９に記載の方法。
実施形態１９１。　基質状態が可溶性または表面吸着性である、実施形態１９０に記載の方法。
実施形態１９２。　前記多様性スクリーニングが、前記標的タンパク質の前記１または

それを超える特性に基づいて前記遺伝要素を濃縮することをさらに含む、実施形態１９０に記載の方法。

実施形態１９３。 前記標的タンパク質の遺伝要素が、遺伝子領域、配列モチーフ、およびタンパク質構造クラスを含む、実施形態１６０に記載の方法。

実施形態１９４。 前記標的タンパク質をコードする遺伝子は、ＣＯＬ１Ａ１（ヒト１型コラーゲン）である、実施形態１９０に記載の方法。

実施形態１９５。 前記標的タンパク質が、構造タンパク質、酵素、表面受容体タンパク質、ペプチドホルモン、免疫系成分、または生物活性ペプチドである、実施形態１９０に記載の方法。

実施形態１９６。 前記標的タンパク質が産業用酵素である、実施形態１９０に記載の方法。

実施形態１９７。 前記標的タンパク質が治療用タンパク質である、実施形態１９０に記載の方法。

実施形態１９８。 前記ペプチドホルモンがインスリンまたはシグナル伝達因子である、実施形態１９５に記載の方法。

実施形態１９９。 前記構造タンパク質がコラーゲンまたは食品ベースのタンパク質である、実施形態１９５に記載の方法。

実施形態２００。 前記酵素が、リパーゼまたはポリ（エチレンテレフタレート）ヒドロラーゼ（ＰＥＴａｓｅ）である、実施形態１９５に記載の方法。

実施形態２０１。 前記生物学的多様性が発見ライブラリのものである、実施形態１６０に記載の方法。

実施形態２０２。 前記発見ライブラリが環境ＤＮＡ（ｅＤＮＡ）ライブラリである、実施形態２０１に記載の方法。

実施形態２０３。 前記条件付き多様性が、１またはそれを超える発酵条件のものである、実施形態１６０に記載の方法。

実施形態２０４。 前記１またはそれを超える発酵条件が、温度、培養液、酸素化、塩分、ｐＨ、混合、供給スケジュール、炭素源、またはそれらのいずれか１つもしくは組み合わせの変更を含む、実施形態２０３に記載の方法。

実施形態２０５。 １００ミリリットルを超える体積で前記望ましい遺伝的バリアントを産生するために産業的発酵条件を同定することをさらに含む、実施形態２０３に記載の方法。

実施形態２０６。 前記条件付き多様性が、１またはそれを超える適用条件のものである、実施形態１６０に記載の方法。

実施形態２０７。 前記適用条件が、ｐＨ、温度、基質、反応緩衝液および支持体への付着を含む、実施形態２０６に記載の方法。

実施形態２０８。 前記望ましい遺伝的バリアントを同定することが、シーケンシングを含む、実施形態１６０に記載の方法。

実施形態２０９。 シーケンシングがＰＣＲを行うことを含む、実施形態２０８に記載の方法。

実施形態２１０。 シーケンシングが、次世代シーケンシング（ＮＧＳ）を行うことを含む、実施形態２０８に記載の方法。

実施形態２１１。 同定することが、望ましい遺伝的バリアントをコードするオープンリーディングフレームを同定することを含む、実施形態２０８に記載の方法。

実施形態２１２。 性能基準を満たす最終最適化タンパク質または最終最適化株を同定するための方法であって、（ａ）ｉ）ａ）株の異なる遺伝的バリアント、ｂ）タンパク質の異なる遺伝的バリアント、またはｃ）別個のタンパク質をコードする異なる遺伝子のうちの２またはそれを超えるものから選択される複数の異なる遺伝要素のデータ、およびｉｉ）前記複数の異なる遺伝要素のそれぞれについて、複数の異なる条件のデータを含むデータセットを得ること、（ｂ）前記性能基準を満たす最終的な最適化されたタンパク質または最終的な最適化された株を同定または予測するために前記データセットを統合するこ

と、または（ａ）に導入する追加データを同定することを含む方法。
実施形態２１３。前記性能基準が、標的タンパク質の発現レベル、標的タンパク質の安定性、標的タンパク質の基質特異性、標的タンパク質活性、株の増殖、発酵条件に対する株の耐性、標的タンパク質のフォールディング、標的タンパク質対副産物比、または標的タンパク質修飾を含む、実施形態２１２に記載の方法。
実施形態２１４。前記複数の条件のデータが、標的タンパク質の安定性、二価カチオン濃度に対する株の耐性、標的タンパク質の基質特異性、温度耐性、ｐＨ耐性、発現レベル、または株の生存率を含む、実施形態２１２に記載の方法。
実施形態２１５。前記データセットが機械学習アルゴリズムに適用され、（ｂ）の前記追加データは前記機械学習アルゴリズムによって同定される、実施形態２１２に記載の方法。
実施形態２１６。前記追加データが、遺伝要素または追加条件のライブラリである、実施形態２１５に記載の方法。
実施形態２１７。前記データセットが、蛍光発生アッセイまたは比色アッセイのうちの１またはそれを超えるものからのデータを含む、実施形態２１２に記載の方法。
実施形態２１８。前記蛍光発生アッセイが、直接アッセイ、間接アッセイ、結合アッセイ、またはインビボレポーターアッセイである、実施形態２１７に記載の方法。
実施形態２１９。（ｂ）の追加データを同定するために前記データセットを統合することが、ｉ）性能基準のサブセットを満たす目的の中間タンパク質または目的の中間株を同定すること、ｉｉ）（ａ）に導入する追加データとして、前記目的の中間タンパク質または前記目的の中間株に対して少なくとも７０％、８０％、または９０％の相同性を有する追加の複数のバリアントを同定すること、およびｉｉｉ）前記追加の複数のバリアントのそれぞれについて、（ａ）に導入する前記複数の異なる条件の追加データを取得することをさらに含む、実施形態２１２に記載の方法。
実施形態２２０。ａ）追加の複数のバリアントが１００個未満のメンバーを含む場合に個々に、またはｂ）前記追加の複数のバリアントが１００個超のメンバーを含む場合に集団レベルで、前記複数の異なる条件の追加データを得ることをさらに含む、実施形態２１９に記載の方法。
実施形態２２１。前記株の遺伝的バリアントが、株の遺伝的バリアントのライブラリに由来する、実施形態２１２に記載の方法。
実施形態２２２。前記株の遺伝的バリアントのライブラリが、全ゲノム突然変異誘発ライブラリ、ゲノムシャッフルライブラリ、標的ゲノムライブラリおよびトランスポゾンライブラリからなる群から選択される、実施形態２１２に記載の方法。
実施形態２２３。前記全ゲノム突然変異誘発ライブラリが、全ゲノムランダム突然変異誘発ライブラリである、実施形態２２２に記載の方法。
実施形態２２４。前記標的ゲノムライブラリが、プロモータースワップライブラリである、実施形態２２２に記載の方法。
実施形態２２５。前記トランスポゾンライブラリが、プロモータースワップライブラリである、実施形態２２２に記載の方法。
実施形態２２６。前記株の遺伝的バリアントのライブラリが、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、またはＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）によって生成される、実施形態２２１に記載の方法。
実施形態２２７。前記タンパク質の前記異なる遺伝的バリアントが、標的タンパク質をコードする遺伝的バリアントのライブラリである、実施形態２１２に記載の方法。
実施形態２２８。前記データセットが、異なるタンパク質をコードする遺伝的バリアントを含み、前記追加データが、残基をコードするバリアントの部分配列ライブラリを含む、実施形態２１２に記載の方法。
実施形態２２９。前記部分配列ライブラリが遺伝子シャッフリングによって生成される、実施形態２２８に記載の方法。
実施形態２３０。前記標的タンパク質をコードする遺伝的バリアントのライブラリが

、ランダム突然変異誘発、多重自動ゲノム工学法（ＭＡＧＥ）、またはＣＲＩＳＰＲ技術を介した追跡可能なゲノム工学法（ＣＲＥＡＴＥ）によって生成される、実施形態２２７に記載の方法。

実施形態２３１。　前記標的タンパク質をコードする遺伝的バリアントのライブラリが、単一遺伝子ランダム突然変異ライブラリ、部位飽和ライブラリ、小型インサートメタゲノムライブラリ、および大型インサートメタゲノムライブラリからなる群から選択される、実施形態２２７に記載の方法。

実施形態２３２。　前記異なるタンパク質をコードする別個の遺伝子が、環境ＤＮＡ（ｅＤＮＡ）ライブラリである、実施形態２１２に記載の方法。

実施形態２３３。　１またはそれを超えるプロセッサとコンピュータ可読媒体とを含む、細胞を処理するためのシステムであって、前記コンピュータ可読媒体が、前記１またはそれを超えるプロセッサによって実行されると前記システムに、

（ａ）細胞の集団を複数の反応体積に分配して、分配された細胞を得ることと、

（ｂ）前記分配された細胞を表現型についてアッセイして、前記分配された細胞に対するシグナルを生成することを実行させる命令を含み、

前記分配された細胞に対する前記シグナルの変動係数（ＣＶ）が、細胞の前記集団が均一な細胞の集団である場合、１１％またはそれ未満である、システム。

実施形態２３４。　前記ＣＶが１０％またはそれ未満である、実施形態２３３に記載システム。

実施形態２３５。　前記ＣＶが７％またはそれ未満である、実施形態２３３に記載システム。

実施形態２３６。　前記ＣＶが約５％～約１０％である、実施形態２３３に記載システム。

実施形態２３７。　前記均一な細胞の集団が少なくとも１００万個の細胞を含む、実施形態２３３に記載システム。

実施形態２３８。　前記均一な細胞の集団が少なくとも１０，０００，０００個の細胞を含む、実施形態２３３に記載システム。

実施形態２３９。　前記反応体積が液滴を含む、実施形態２３３に記載システム。

実施形態２４０。　前記アッセイが、前記分配された細胞への検出試薬の添加を含む、実施形態２３３に記載システム。

実施形態２４１。　前記アッセイが、前記分配された細胞への前記検出試薬の前記添加のタイミングの制御を含む、実施形態２４０に記載システム。

実施形態２４２。　前記検出試薬が酵素基質を含む、実施形態２４０に記載システム。

実施形態２４３。　前記アッセイが、前記分配された細胞に対する前記シグナルの検出を含む、実施形態２３３に記載システム。

実施形態２４４。　前記検出が光学的検出を含む、実施形態２４３に記載システム。

実施形態２４５。　前記アッセイが、前記分配された細胞に対する検出シグナルのタイミングの制御を含む、実施形態２４３に記載システム。

実施形態２４６。　前記表現型が酵素活性を含む、実施形態２３３に記載システム。

実施形態２４７。　（ａ）が、複数の反応体積の各反応体積の体積サイズの制御を含む、実施形態２３３に記載システム。

実施形態２４８。　前記均一な細胞の集団が同じ遺伝物質を含む、実施形態２３３に記載システム。

実施形態２４９。　前記均一な細胞の集団が、同じ配列を有する同じ目的の遺伝子を含む、実施形態２３３に記載システム。

【０３１０】

本発明の好ましい実施形態を本明細書に示し説明してきたが、そのような実施形態が例としてのみ提供されることは当業者には明らかであろう。本発明は、本明細書内で提供される特定の例によって限定されることを意図しない。本発明を前述の明細書を参照して説明してきたが、本明細書の実施形態の説明および例示は、限定的な意味で解釈されること

を意味しない。本発明から逸脱することなく、当業者には多数の変形、変更、および置換が思い浮かぶであろう。さらに、本発明のすべての態様は、様々な条件および変数に依存する本明細書に記載の特定の描写、構成または相対的な割合に限定されないことを理解されたい。本明細書に記載の本発明の実施形態に対する様々な代替形態が、本発明を実施する際に使用され得ることを理解されたい。したがって、本発明は、任意のそのような代替形態、修正形態、変形形態または均等物も包含すると考えられる。以下の特許請求の範囲が本発明の範囲を定義し、これらの特許請求の範囲内の方法および構造ならびにそれらの均等物がそれによって包含されることが意図される。

【図面】

【図１】

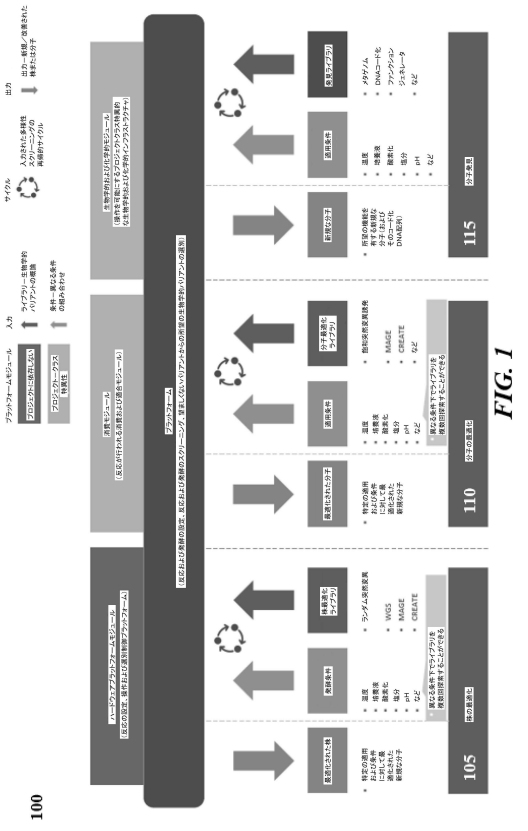


FIG. 1

【図２】

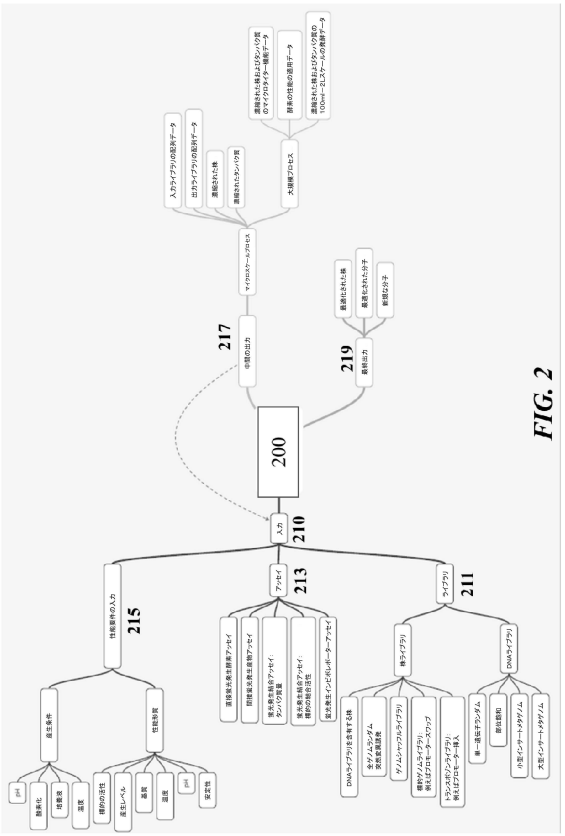


FIG. 2

【図３】

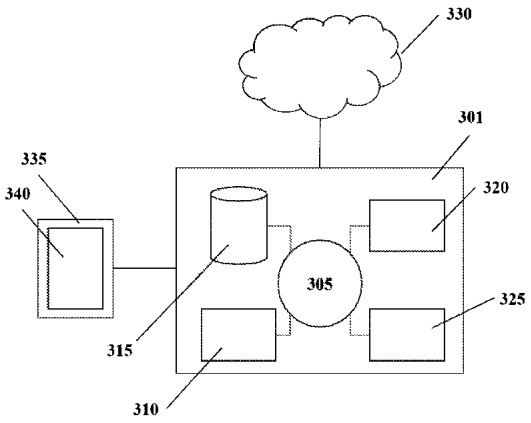


FIG. 3

【図４】

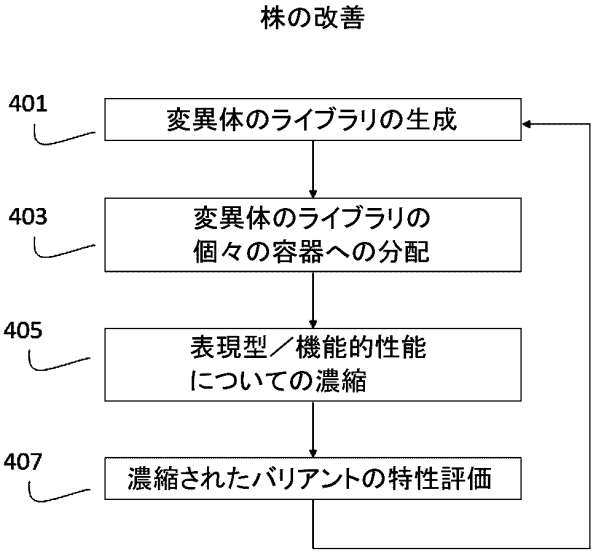


FIG. 4

【図５】

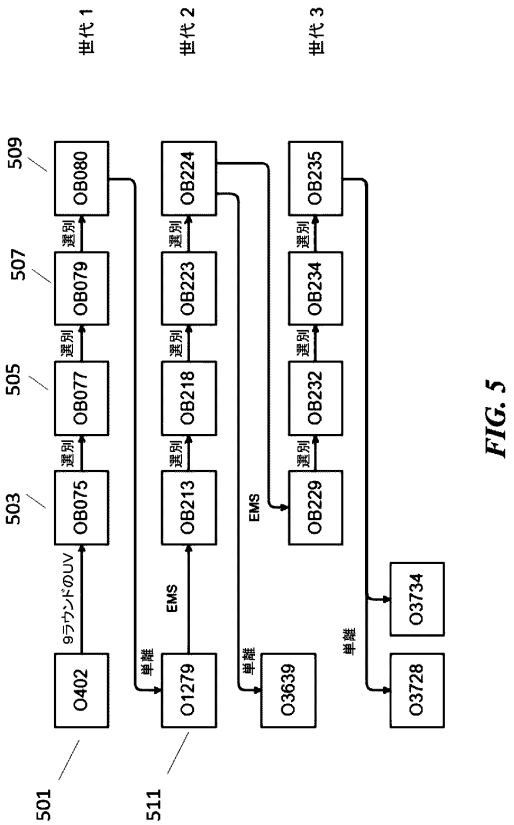


FIG. 5

【図６Ａ－６Ｃ】

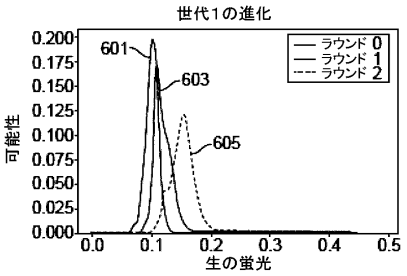


FIG. 6A

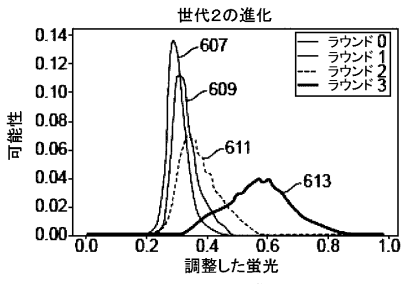


FIG. 6B

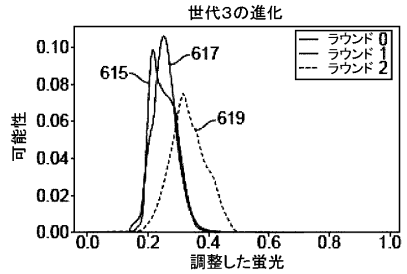


FIG. 6C

【図７】

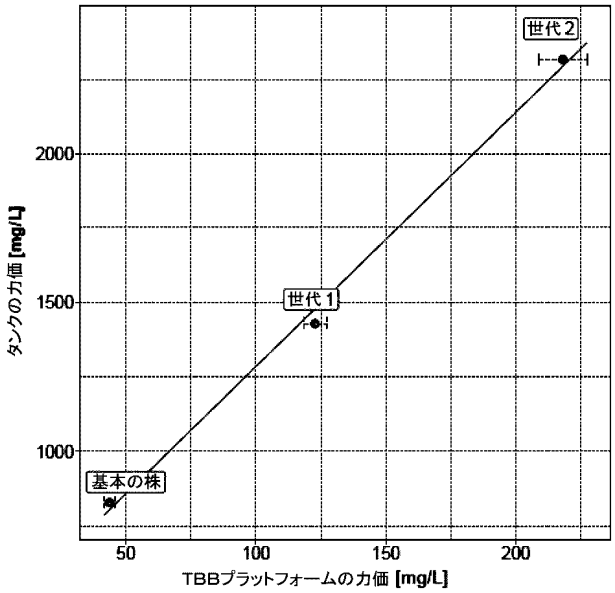


FIG. 7

【図８】

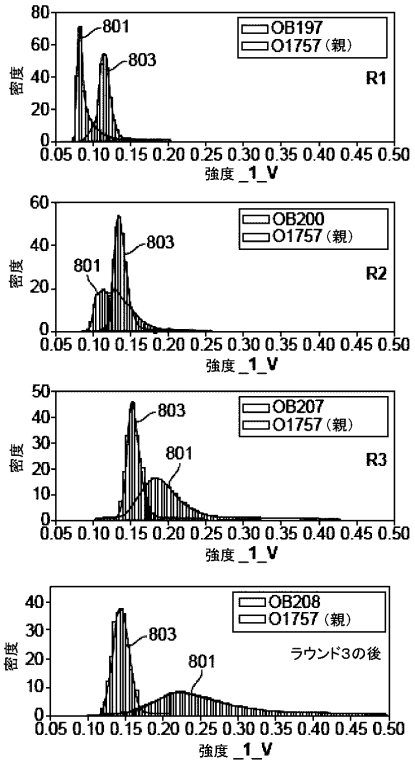


FIG. 8

【図９】

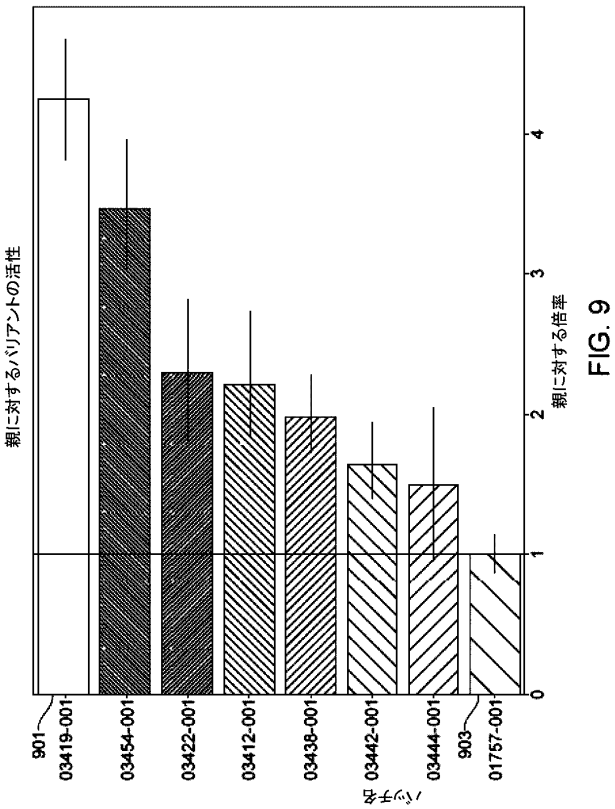


FIG. 9

【図１０】

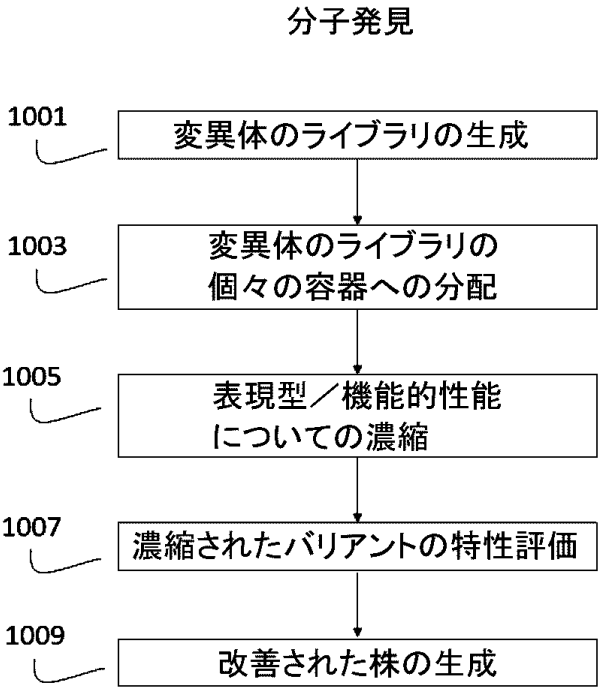


FIG. 10

【図１１】

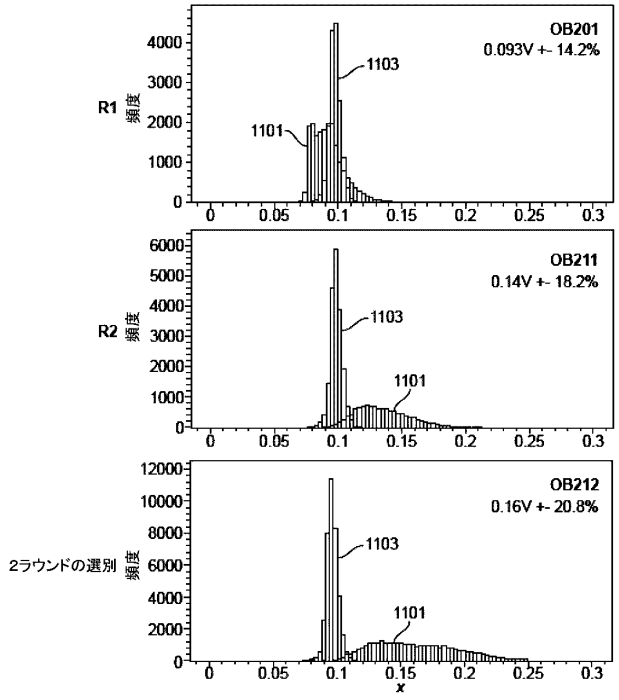


FIG. 11

【図１２】

元のヒットした株	変異体	元のO3297 nucB宿主における改善	O402 nucB シャッフリング単離株における改善	GFP発現 O707宿主における改善
O3549	sigA T31R, M39L, L56F, R210P, P237S, D275E	4.0		2.5
O3553	sigA V102D, R207S, I242F, A287S	2.9		3.7
O3565	sigA A2S, L30S, M158I, I209F, D305Y, L318M, A351V	4.3		2.4
O3606	sigA K4I, S23Y, K25N, Q48R, I109T, A122T, A197T, R210Q, T238S, G279S, F339L	4.2	1.6	3.7

FIG. 12

【図１３－１】

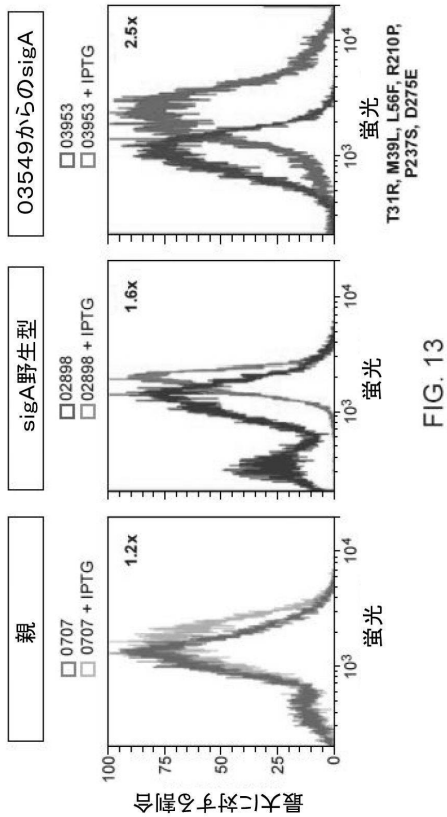


FIG. 13

【図１３－２】

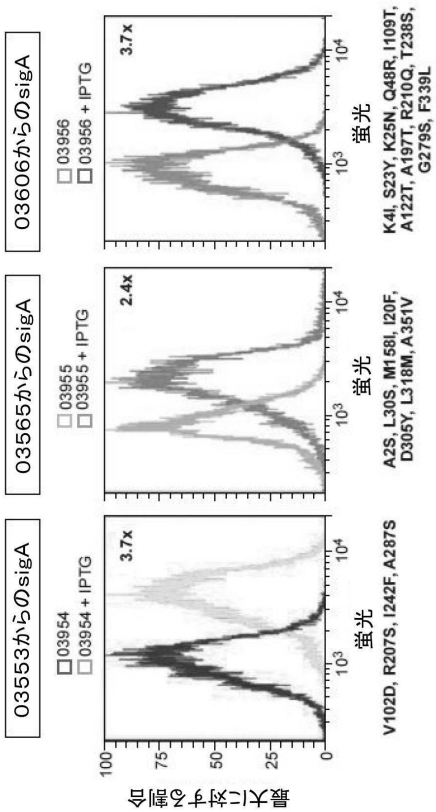


FIG. 13（続き）

【図１４】

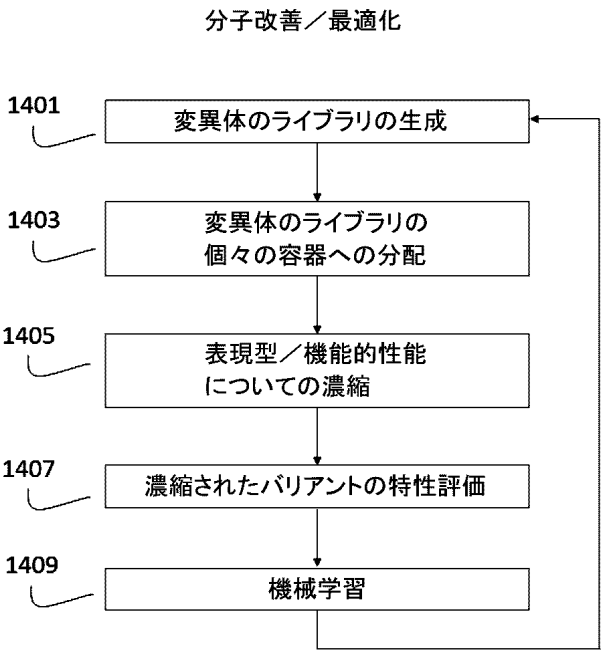


FIG. 14

【図１５】

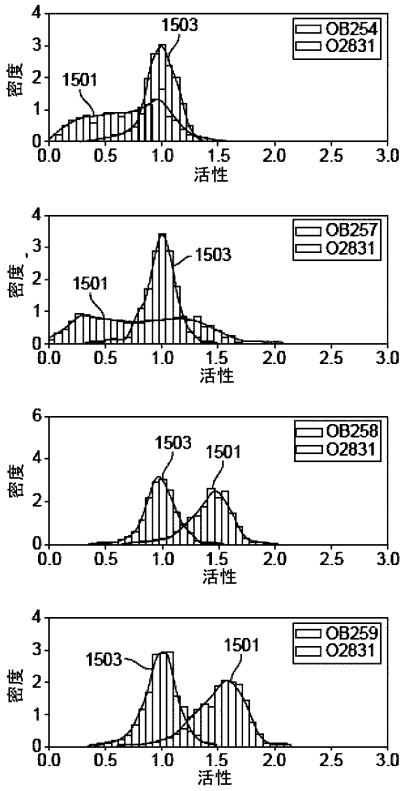


FIG. 15

【図１６】

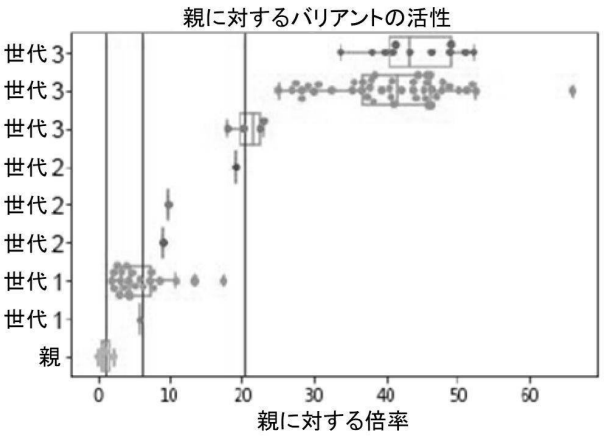


FIG. 16

【図１７】

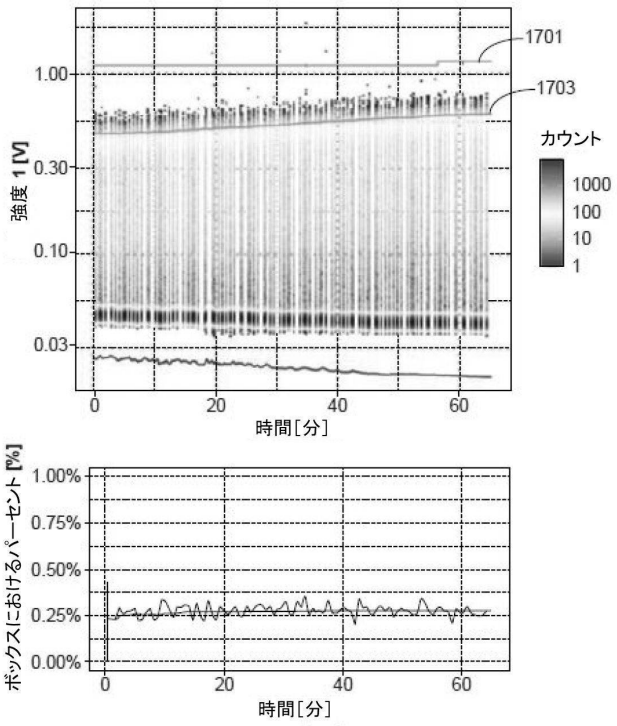


FIG. 17

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/059275

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C12N 15/10; C12Q 1/6886; G16B 20/20 (2022.01)
CPC - C12N 15/1037; C12N 15/1079; C12Q 2600/156; C12Q 2600/158; G16B 20/20; G16B 20/50; G16B 25/10 (2022.02)

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

B. FIELDS SEARCHED

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

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2010/0100331 A1 (GUSTAFSSON et al) 22 April 2010 (22.04.2010) entire document	1, 15, 16, 21-30, 46-48, 50
— Y		— 2-5, 17-20, 31-45, 49
Y	WO 2019/006022 A1 (THE BROAD INSTITUTE INC. et al) 03 January 2019 (03.01.2019) entire document	2-5, 31-45
Y	US 2009/0130718 A1 (SHORT) 21 May 2009 (21.05.2009) entire document	17-20, 49
A	US 2016/0210403 A1 (THE REGENTS OF THE UNIVERSITY OF CALIFORNIA et al) 21 July 2016 (21.07.2016) entire document	1-5, 15-50
A	US 2014/0256557 A1 (CODEXIS INC.) 11 September 2014 (11.09.2014) entire document	1-5, 15-50

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

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

Date of the actual completion of the international search	Date of mailing of the international search report
03 March 2022	MAR 16 2022
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, VA 22313-1450 Facsimile No. 571-273-8300	Authorized officer Harry Kim Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/210 (second sheet) (July 2019)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/059275

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 6-14, 51-54
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

See extra sheet(s).

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-5, 15-50

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2021/059275

Continued from Box No. III Observations where unity of invention is lacking

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees need to be paid.

Group I: claims 1-5 and 15-50 are drawn to method for generating improved polymer variants.

Group II: claims 55-71 are drawn to systems for processing cells.

Group III: claims 72-80 are drawn to methods for generating improved cells.

The inventions listed in Groups I-III do not relate to a single general inventive concept under PCT Rule 13.1, because under PCT Rule 13.2 they lack the same or corresponding special technical features for the following reasons:

The special technical features of Group I, a method for generating an improved polymer variant, are not present in Groups II and III; the special technical features of Group II, one or more processors and computer readable medium, the computer readable medium comprising instructions thereon, are not present in groups I or III, and the special technical features of Group III, based on the screening, selecting an improved cell from the partitioned mutagenized cells using a threshold gate, are not present in Groups I, and II.

Additionally, Groups I-III share the technical features of generating an improved, obtaining a dataset, a population of cells, partitioned cells, and a phenotype. However, these shared technical features do not represent a contribution over the prior art.

Specifically, disclose generating an improved (cancer regression (Improvement), Para. [0267]; generating a CpG methylation profile database, Para. [0005]), obtaining a dataset (the generating the pair-wise methylation difference dataset, Para. [0018]), a population of cells (stem cell population of interest, Para. [0285]), partitioned cells (The samples are randomly partitioned into fractional test sets, Para. [0235]), and a phenotype (factors such as the …phenotype …can be utilized, Para. [0237]).

The inventions listed in Groups I-III therefore lack unity under Rule 13 because they do not share a same or corresponding special technical features.

フロントページの続き

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

(74)代理人　230113332
弁護士　山本 健策
(72)発明者　アグレスティ， ジェレミー
アメリカ合衆国 カリフォルニア ９４６０８， エメリービル， ４５ティーエイチ ストリート １２５０， スイート １５０， トリプルバー バイオ， インコーポレイテッド 気付
(72)発明者　オルネラス バルガス， アンドレス
アメリカ合衆国 カリフォルニア ９４６０８， エメリービル， ４５ティーエイチ ストリート １２５０， スイート １５０， トリプルバー バイオ， インコーポレイテッド 気付
(72)発明者　ホフ， ケビン グレゴリー
アメリカ合衆国 カリフォルニア ９４６０８， エメリービル， ４５ティーエイチ ストリート １２５０， スイート １５０， トリプルバー バイオ， インコーポレイテッド 気付

Ｆターム（参考）　4B063　QA13 QQ44 QR08 QR62 QS28 QX02
4B065　AA01X AC20 BA16 BA30 BC50 CA60