(19)日本国特許庁(JP)　　(12)公表特許公報(A)　　(11)公表番号
特表2025-515452
(P2025-515452A)
(43)公表日　令和7年5月15日(2025.5.15)

(51)国際特許分類　　ＦＩ
Ｇ１６Ｂ　30/00　(2019.01)　　Ｇ１６Ｂ　30/00

審査請求　未請求　予備審査請求　未請求　（全130頁）

(21)出願番号　特願2024-562211(P2024-562211)
(86)(22)出願日　令和5年4月20日(2023.4.20)
(85)翻訳文提出日　令和6年12月20日(2024.12.20)
(86)国際出願番号　PCT/US2023/019283
(87)国際公開番号　WO2023/205345
(87)国際公開日　令和5年10月26日(2023.10.26)
(31)優先権主張番号　63/333,305
(32)優先日　令和4年4月21日(2022.4.21)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　63/338,760
(32)優先日　令和4年5月5日(2022.5.5)
(33)優先権主張国・地域又は機関
米国(US)
(31)優先権主張番号　63/481,873
最終頁に続く

(71)出願人　516040017
ツイスト　バイオサイエンス　コーポレーション
アメリカ合衆国　９４０８０　カリフォルニア州　サウス・サンフランシスコ　ゲートウェイ・ブールバード　６８１
(74)代理人　110003797
弁理士法人清原国際特許事務所
(72)発明者　トッパニ，ドミニク
アメリカ合衆国　９４０８０　カリフォルニア州　サウス・サンフランシスコ　ゲートウェイ・ブールバード　６８１
(72)発明者　ピッチ，ステファン
アメリカ合衆国　９４０８０　カリフォルニア州　サウス・サンフランシスコ　ゲー
最終頁に続く

(54)【発明の名称】　ＤＮＡデータ・ストレージのためのコーデック

(57)【要約】
本明細書では、デジタル・データをオリゴヌクレオチドにエンコードし、オリゴヌクレオチドを再びデジタル・データにデコードするためのシステムおよび方法が説明される。エンコーディング方式およびデコーディング方式は、デジタル・データを塩基に変換するため、および塩基をデジタル・データに変換するための内部コーデックを含む。エンコーディング方式およびデコーディング方式は、エラー訂正方式を含んでいる外部コーデックも含む。
【選択図】図７

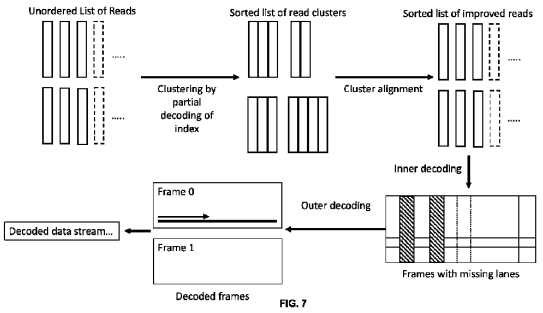


FIG. 7

【特許請求の範囲】
【請求項１】
複数のポリヌクレオチド配列内の複数のシンボルによって表されたデータをエンコードするための方法であって、
（ａ）データを複数のフレームに分割することであって、前記複数のフレームのうちの各フレームがフレーム・インデックスを含む、分割することと、
（ｂ）外部コーデックを前記複数のフレームうちの各フレームに適用することであって、前記外部コーデックがエラー訂正方式を含む、適用することと、
（ｃ）各フレームを複数のレーンに分割することであって、前記複数のレーンのうちの各レーンがレーン・インデックスを含む、分割することと、
（ｄ）前記レーン・インデックスに少なくとも部分的に基づいて各レーンをシャッフルすることと、
（ｅ）内部コーデックを適用して、前記複数のポリヌクレオチド配列のうちの１つのポリヌクレオチド配列内の各レーンをエンコードすることとを含む、方法。
【請求項２】
前記データがバイナリ・データを含み、前記バイナリ・データがバイト・ストリームまたはバイト配列を含む、請求項１に記載の方法。
【請求項３】
（ｄ）における前記シャッフルすることが、各レーン内の回転方式を含む、請求項１に記載の方法。
【請求項４】
（ｄ）における前記シャッフルすることが、各レーン内の擬似ランダム・プロセスを含む、請求項１に記載の方法。
【請求項５】
（ｄ）における前記シャッフルすることが、エラーに対する耐性をもたらす、請求項１に記載の方法。
【請求項６】
前記エラーが、ヌクレオチド合成エラーまたはシーケンシング・エラーである、請求項５に記載の方法。
【請求項７】
前記エラーが、欠失、挿入、または置換を含む、請求項５に記載の方法。
【請求項８】
前記エラー訂正方式が、リード・ソロモン（ＲＳ）コード、低密度パリティ・チェック（ＬＤＰＣ）コード、ターボコード、極コード、またはこれらの任意の組み合わせを含む、請求項１に記載の方法。
【請求項９】
前記データが少なくとも約１ＧＢ～約１ＴＢのデータを含む、請求項１に記載の方法。
【請求項１０】
前記複数のフレームが、約１００～約１０，０００フレームを含む、請求項１に記載の方法。
【請求項１１】
各フレームが、最大約５０００レーンを含む、請求項１に記載の方法。
【請求項１２】
各レーンが、約１００～約３００ビットを含む、請求項１に記載の方法。
【請求項１３】
前記フレーム・インデックスが、約１６～約２０ビットを含む、請求項１に記載の方法。
【請求項１４】
前記レーン・インデックスが、約１２ビットまたは約１６ビットを含む、請求項１に記載の方法。

【請求項１５】
前記ポリヌクレオチド配列が、約１００～約３００塩基の長さである、請求項１に記載の方法。
【請求項１６】
（ｄ）の前に、前記フレーム・インデックスおよび／または前記レーン・インデックスが各レーンの先頭に追加される、請求項１に記載の方法。
【請求項１７】
前記内部コーデックを適用することが、前記複数のポリヌクレオチド配列にわたって冗長性を追加することを含む、請求項１に記載の方法。
【請求項１８】
前記冗長性が約５％～約１０％である、請求項１７に記載の方法。
【請求項１９】
一部分において前記複数のポリヌクレオチド配列にわたる前記冗長性のため、エラーが存在する場合に、前記複数のポリヌクレオチド配列がデコードされ得る、請求項１７に記載の方法。
【請求項２０】
前記エラーが、挿入、欠失、置換、またはこれらの任意の組み合わせを含む、請求項１９に記載の方法。
【請求項２１】
前記内部コーデックを適用することが、
（ａ）レーンからのシンボル、シンボル履歴、およびシンボル位置を結合することと、
（ｂ）ルックアップ・テーブル、ハッシュ、またはその両方を使用して塩基候補を生成することとを含む、請求項１に記載の方法。
【請求項２２】
塩基反復チェックを実行することをさらに含む、請求項２１に記載の方法。
【請求項２３】
前記シンボル履歴を更新すること、前記レーン・インデックスをインクリメントすること、前記フレーム・インデックスをインクリメントすること、またはこれらの任意の組み合わせをさらに含む、請求項２１に記載の方法。
【請求項２４】
更新された前記シンボル履歴、インクリメントされた前記レーン・インデックス、インクリメントされた前記フレーム・インデックス、またはこれらの組み合わせが、後続のレーンのシンボルと結合される、請求項２３に記載の方法。
【請求項２５】
前記複数のポリヌクレオチド配列を合成する前にＧＣフィルタリングを実行することをさらに含む、請求項２１に記載の方法。
【請求項２６】
前記ＧＣフィルタリングが、前記複数のレーンのうちの約５％～約１０％のレーンを除去することを含む、請求項２５に記載の方法。
【請求項２７】
前記複数のポリヌクレオチド配列が、約４５％～約５５％のＧＣ含量を含む、請求項１に記載の方法。
【請求項２８】
前記複数のポリヌクレオチド配列の少なくとも９０％が、約４５％～約５５％のＧＣ含量を含む、請求項１に記載の方法。
【請求項２９】
前記内部コーデックを適用することが、
（ａ）ルックアップ・テーブルを使用して、レーン内のシンボルごとに塩基候補を生成することと、
（ｂ）以前にエンコードされたシンボルに少なくとも部分的に基づいて、次のルックア

ップ・テーブルを選択することとを含む、請求項１に記載の方法。
【請求項３０】
複数のポリヌクレオチド配列をデコードして、複数のシンボルによって表されたデータを含む出力を生成するための方法であって、
（ａ）前記複数のポリヌクレオチド配列を決定することと、
（ｂ）内部コーデックを前記複数のポリヌクレオチド配列に適用することであって、前記内部コーデックが、前記複数のポリヌクレオチド配列の各々を、複数のシンボルを含むレーンに変換し、前記内部コーデックが、貪欲アルゴリズムおよび最尤（ＭＬ）アルゴリズムを含んでいる混合デコーディング・アルゴリズムを含む、適用することと、
（ｃ）データのレーンを、レーン・インデックスおよび各レーンのフレーム・インデックスに基づいてフレームに配置することと、
（ｄ）外部コーデックを前記フレームに適用することであって、前記外部コーデックがエラー訂正方式を含み、
前記外部コーデックからの前記フレームがマージされて、前記データを含む出力を生成する、適用することとを含む、方法。
【請求項３１】
（ｂ）の前に前記ポリヌクレオチド配列をクラスタ化することをさらに含む、請求項３０に記載の方法。
【請求項３２】
前記クラスタ化することがインデックスに基づく、請求項３１に記載の方法。
【請求項３３】
クラスタ化することが、前記フレーム・インデックス、前記レーン・インデックス、またはその両方をデコードすることを部分的に含む、請求項３２に記載の方法。
【請求項３４】
前記クラスタ化することが、ハッシュ関数を使用して実行される、請求項３１に記載の方法。
【請求項３５】
（ｂ）の前に前記ポリヌクレオチド配列をアラインメントすることをさらに含む、請求項３０に記載の方法。
【請求項３６】
アラインメントが、アラインメント・アルゴリズムを使用して前記ヌクレオチドのコンセンサスを分析することを含む、請求項３５に記載の方法。
【請求項３７】
前記アライメント・アルゴリズムが、ペアワイズ・アライメント・アルゴリズム、マルチシーケンス・アライメント・アルゴリズム、またはこれらの組み合わせを含む、請求項３６に記載の方法。
【請求項３８】
前記アライメント・アルゴリズムが、
（ａ）複数の読み取りのうちの読み取りごとに位置を初期化することであって、初期化が、ポリヌクレオチド配列を位置０にアライメントすることを含む、初期化することと、
（ｂ）各読み取り間に次の１つまたは複数の塩基のコンセンサスを分析することと、
（ｃ）読み取りごとに、前記次の１つまたは複数の塩基の各々が正しいか、またはエラーを有するかを含む判定を決定することと、
（ｄ）読み取りごとに、前記判定を前提として前記位置をインクリメントすることと、
（ｅ）ステップ（ｂ）～（ｄ）を繰り返すこととを含む、請求項３６に記載の方法。
【請求項３９】
前記複数の読み取りが、約３～約１０回の読み取りを含む、請求項３８に記載の方法。
【請求項４０】
各読み取りの長さが、約１００～約３００塩基である、請求項３８に記載の方法。
【請求項４１】

前記次の１つまたは複数の塩基が、約２、３、４、または５塩基である、請求項３８に記載の方法。
【請求項４２】
前記混合デコーディング・アルゴリズムが、１つまたは複数の状態からの遷移確率に基づいてデコードすることを含む、請求項３０に記載の方法。
【請求項４３】
前記１つまたは複数の状態が、約１００～約１０００個の最も可能性の高い状態を含む、請求項４２に記載の方法。
【請求項４４】
前記内部コーデックがドリフト項をさらに含む、請求項４２に記載の方法。
【請求項４５】
前記ドリフト項が整数を含む、請求項４４に記載の方法。
【請求項４６】
前記整数が、ポリヌクレオチド配列内の挿入または欠失の総数に関連付けられる、請求項４５に記載の方法。
【請求項４７】
前記整数が、挿入または欠失の前記総数のうちの、挿入の値および／または欠失の値を合計することによって計算される、請求項４６に記載の方法。
【請求項４８】
前記挿入の前記値が＋１を含み、前記欠失の前記値が－１を含む、請求項４７に記載の方法。
【請求項４９】
（ｃ）が、前記レーン・インデックスに基づいて前記レーンをシャッフル解除し、前記フレーム・インデックスに基づいて前記レーンをフレームにグループ化することを含む、請求項３０に記載の方法。
【請求項５０】
前記エラー訂正方式が、リード・ソロモン（ＲＳ）コード、低密度パリティ・チェック（ＬＤＰＣ）コード、ターボコード、極コード、またはこれらの任意の組み合わせを含む、請求項３０に記載の方法。
【請求項５１】
前記複数のポリヌクレオチド配列のうちの少なくとも１つのポリヌクレオチド配列がエラーを含む、請求項３０に記載の方法。
【請求項５２】
前記エラーが、挿入、欠失、置換、またはこれらの任意の組み合わせを含む、請求項５１に記載の方法。
【請求項５３】
（ａ）メモリと、
（ｂ）前記メモリに動作可能に結合された処理デバイスとを備えている装置であって、前記処理デバイスが、
（ｉ）データを複数のフレームに分割することであって、前記複数のフレームのうちの各フレームがフレーム・インデックスを含む、分割することと、
（ｉｉ）外部コーデックを前記複数のフレームのうちの各フレームに適用することであって、前記外部コーデックがエラー訂正方式を含む、適用することと、
（ｉｉｉ）各フレームを複数のレーンに分割することであって、前記複数のレーンのうちの各レーンがレーン・インデックスを含む、分割することと、
（ｉｖ）前記レーン・インデックスに少なくとも部分的に基づいて各レーンをシャッフルすることと、
（ｖ）内部コーデックを適用して、ポリヌクレオチド配列内の各レーンをエンコードすることとを実行するように構成される、装置。
【請求項５４】

（ａ）メモリと、
（ｂ）複数のヌクレオチドの配列を決定するように構成されたシーケンシング・デバイスと、
（ｃ）前記メモリおよび前記シーケンシング・デバイスに動作可能に結合された処理デバイスとを備えている装置であって、前記処理デバイスが、
（ｉ）内部コーデックを前記配列に適用することであって、前記内部コーデックが、前記配列の各々を、複数のシンボルを含むレーンに変換し、前記内部コーデックが、貪欲アルゴリズムおよび最尤（ＭＬ）アルゴリズムを含んでいる混合デコーディング・アルゴリズムを含む、適用することと、
（ｉｉ）前記レーンを、レーン・インデックスおよび各レーン内のフレーム・インデックスに基づいてフレームに配置することと、
（ｉｉｉ）外部コーデックを前記フレームに適用することであって、前記外部コーデックがエラー訂正方式を含み、
前記外部コーデックからの前記フレームがマージされて、前記データを含む出力を生成する、適用することとを実行するように構成される、装置。
【請求項５５】
ポリヌクレオチド配列内のデータをエンコードするための方法であって、
（ａ）コードブックを含む内部コーデックを生成することであって、前記コードブックが１つまたは複数の制約に対して最適化される、生成することと、
（ｂ）前記内部コーデックを適用して、前記データを複数のポリヌクレオチド配列としてエンコードすることとを含む、方法。
【請求項５６】
前記１つまたは複数の制約が、核酸合成、後処理、ストレージ、シーケンシング、またはこれらの任意の組み合わせに関連する、請求項５５に記載の方法。
【請求項５７】
前記核酸合成が、電気化学的合成、酵素合成、ホスホラミダイト合成、インクジェット印刷、またはこれらの任意の組み合わせを含む、請求項５６に記載の方法。
【請求項５８】
前記核酸合成に関連する前記１つまたは複数の制約が合成エラーを含む、請求項５６に記載の方法。
【請求項５９】
前記合成エラーが、挿入、欠失、または突然変異を含む、請求項５８に記載の方法。
【請求項６０】
前記後処理が、ライゲーション、切断、ハイブリダイゼーション、変性、固体支持体への固定、伸長、エラー訂正、濃縮、分離、精製、または増幅のうちの１つまたは複数を含む、請求項５６に記載の方法。
【請求項６１】
前記ストレージがコールド・データ・ストレージを含む、請求項５６に記載の方法。
【請求項６２】
前記ストレージが、液相または固相での核酸ストレージを含む、請求項５６に記載の方法。
【請求項６３】
前記ストレージに関連する前記１つまたは複数の制約が、温度、湿度、圧力、塩分濃度、ｐＨ、濃度、時間、光、ＵＶ、O_2、またはこれらの任意の組み合わせを含む、請求項５６に記載の方法。
【請求項６４】
前記温度が室温を含む、請求項６３に記載の方法。
【請求項６５】
前記シーケンシングが、次世代シーケンシング、並列シーケンシング、単一分子リアルタイム・シーケンシング、ナノポア・シーケンシング、合成によるシーケンシング、サン

ガー・シーケンシング、またはこれらの任意の組み合わせを含む、請求項５６に記載の方法。
【請求項６６】
（ｃ）前記複数のポリヌクレオチド配列を含んでいる複数のポリヌクレオチドを合成することをさらに含む、請求項５５に記載の方法。
【請求項６７】
前記コードブックが、塩基順序に部分的に基づいて生成されるコードワードを含む、請求項５５に記載の方法。
【請求項６８】
前記塩基順序が既定の塩基転移を含む、請求項６７に記載の方法。
【請求項６９】
前記内部コーデックが２つ以上のコードブックを含む、請求項５５に記載の方法。
【請求項７０】
前記２つ以上のコードブックの各々が、複数のポリヌクレオチドの合成中に層をエンコードする、請求項６９に記載の方法。
【請求項７１】
前記層が、少なくとも１つの塩基による前記複数のポリヌクレオチドのうちの各ポリヌクレオチドの伸長を含む、請求項７０に記載の方法。
【請求項７２】
前記層の合成が１つまたは複数のサイクルを含み、前記１つまたは複数のサイクルの各々が、前記コードブックの１つまたは複数の塩基転移に従って塩基を流すことを含む、請求項７１に記載の方法。
【請求項７３】
前記１つまたは複数のサイクルのうちの１つのサイクルが、Ａ、Ｔ、Ｃ、またはＧのうちの１つまたは複数の追加を含む、請求項７２に記載の方法。
【請求項７４】
前記２つ以上のコードブックの各々が異なる塩基順序を含む、請求項６９に記載の方法。
【請求項７５】
前記コードブックが約１２コードワードを含む、請求項５５に記載の方法。
【請求項７６】
（ｂ）が、前記コードブックに基づいて前記データを前記複数のポリヌクレオチド配列にマッピングすることを含む、請求項５５に記載の方法。
【請求項７７】
前記内部コーデックが、前記複数のポリヌクレオチド配列の長さ、ＧＣ含量、反復、エラー、またはこれらの任意の組み合わせを含む１つまたは複数の制約に対してさらに最適化される、請求項５５に記載の方法。
【請求項７８】
前記複数のポリヌクレオチド配列の４０％～６０％が、冗長性のためにエンコードする、請求項５５に記載の方法。
【請求項７９】
合成が複数の合成サイクルを含む、請求項５５に記載の方法。
【請求項８０】
前記合成サイクルの数が、前記内部コーデックを使用してエンコードされないポリヌクレオチド配列を合成するために必要とされる前記合成サイクルの数と比較して減少する、請求項７９に記載の方法。
【請求項８１】
前記合成サイクルの数の減少が、フロー順序に部分的に基づく、請求項８０に記載の方法。
【請求項８２】

前記合成サイクルの数が少なくとも３０％減少する、請求項８０に記載の方法。
【請求項８３】
前記合成サイクルの数が５０％減少する、請求項８０に記載の方法。
【請求項８４】
前記合成サイクルの数が、１００塩基を含むポリヌクレオチド配列の場合、３００未満である、請求項８０に記載の方法。
【請求項８５】
前記合成サイクルの数が、１００塩基を含むポリヌクレオチド配列の場合、約１５５である、請求項８０に記載の方法。
【請求項８６】
前記ポリヌクレオチド配列が、Ａ、Ｔ、Ｃ、またはＧのうちの１つまたは複数を含む、請求項８４に記載の方法。
【請求項８７】
（ｃ）が、固体支持体上で前記複数のポリヌクレオチドを合成することを含む、請求項６６に記載の方法。
【請求項８８】
前記固体支持体が複数の特徴を備える、請求項８７に記載の方法。
【請求項８９】
合成サイクルごとに、前記複数の特徴の２５％より多くがデブロッキングされる、請求項８８に記載の方法。
【請求項９０】
合成サイクルごとに、前記複数の特徴の少なくとも５０％がデブロッキングされる、請求項８８に記載の方法。
【請求項９１】
前記複数のポリヌクレオチド配列の各々が同じ長さを有する、請求項５５に記載の方法。
【請求項９２】
前記複数のポリヌクレオチド配列の８０％～１００％が同じ長さを有する、請求項５５に記載の方法。
【請求項９３】
複数のポリヌクレオチドをシーケンシングして、複数の出力配列を生成することをさらに含む、請求項５５に記載の方法。
【請求項９４】
複数の出力配列が、貪欲アルゴリズム、最尤（ＭＬ）アルゴリズム、または混合貪欲ＭＬアルゴリズムを使用してデコードされる、請求項５５に記載の方法。
【請求項９５】
エラーの確率を計算することに少なくとも部分的に基づいて、複数の出力配列がデコードされる、請求項５５に記載の方法。
【請求項９６】
前記エラーが、欠失、挿入、突然変異、またはこれらの任意の組み合わせを含む、請求項９５に記載の方法。
【請求項９７】
データをエンコードするためのハイブリッド有機インシリコ・プラットフォームであって、
（ａ）少なくとも１つのプロセッサ、および動作を実行するために前記少なくとも１つのプロセッサによって実行可能な命令を備えるコンピューティング・システムを備え、前記動作が、
（ｉ）コードブックを含む内部コーデックを生成することであって、前記コードブックが１つまたは複数の制約に対して最適化される、生成することと、
（ｉｉ）前記内部コーデックを適用して、前記データを複数のポリヌクレオチド配列

としてエンコードすることとを含み、
前記ハイブリッド有機インシリコ・プラットフォームがさらに、
（ｂ）前記複数のポリヌクレオチド配列を含む複数のポリヌクレオチドを生成するためのシンセサイザを備える、ハイブリッド有機インシリコ・プラットフォーム。
【請求項９８】
前記１つまたは複数の制約が、核酸合成、後処理、ストレージ、シーケンシング、またはこれらの任意の組み合わせに関連する、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項９９】
前記核酸合成が、電気化学的合成、酵素合成、ホスホラミダイト合成、インクジェット印刷、またはこれらの任意の組み合わせを含む、請求項９８に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１００】
前記核酸合成に関連する前記１つまたは複数の制約が合成エラーを含む、請求項９８に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１０１】
前記合成エラーが、挿入、欠失、または突然変異を含む、請求項１００に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１０２】
前記後処理が、ライゲーション、切断、ハイブリダイゼーション、変性、固体支持体への固定、伸長、エラー訂正、濃縮、分離、精製、および増幅のうちの１つまたは複数を含む、請求項９８に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１０３】
ストレージがコールド・データ・ストレージを含む、請求項９８に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１０４】
前記ストレージが、液相または固相での核酸ストレージを含む、請求項９８に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１０５】
前記ストレージに関連する前記１つまたは複数の制約が、温度、湿度、圧力、塩分濃度、ｐＨ、濃度、時間、光、ＵＶ、O_2、またはこれらの任意の組み合わせを含む、請求項９８に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１０６】
前記温度が室温を含む、請求項１０５に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１０７】
前記シーケンシングが、次世代シーケンシング、並列シーケンシング、単一分子リアルタイム・シーケンシング、ナノポア・シーケンシング、合成によるシーケンシング、サンガー・シーケンシング、またはこれらの任意の組み合わせを含む、請求項９８に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１０８】
前記コンピューティング・システムがクラウド・コンピューティング・システムを含む、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１０９】
前記クラウド・コンピューティング・システムが、プライベート・クラウド、パブリック・クラウド、ハイブリッド・クラウド、マルチクラウド、またはこれらの任意の組み合わせを含む、請求項１０８に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１０】
前記クラウド・コンピューティング・システムが、サービスとしてのインフラストラクチャ（ＩａａＳ）、サービスとしてのプラットフォーム（ＰａａＳ）、サービスとしての

ソフトウェア（ＳａａＳ）、またはこれらの任意の組み合わせを含む、請求項１０８に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１１】
前記コードブックが、塩基順序に部分的に基づいて生成されるコードワードを含む、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１２】
塩基順序が既定の塩基転移を含む、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１３】
前記内部コーデックが２つ以上のコードブックを含む、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１４】
前記２つ以上のコードブックの各々が、前記複数のポリヌクレオチドの合成中に層をエンコードする、請求項１１３に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１５】
前記層が、少なくとも１つの塩基による前記複数のポリヌクレオチドのうちの各ポリヌクレオチドの伸長を含む、請求項１１４に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１６】
前記層の合成が１つまたは複数のサイクルを含み、前記１つまたは複数のサイクルの各々が、前記コードブックの１つまたは複数の塩基転移に従って塩基を流すことを含む、請求項１１５に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１７】
前記１つまたは複数のサイクルのうちの１つのサイクルが、Ａ、Ｔ、Ｃ、またはＧのうちの１つまたは複数の追加を含む、請求項１１６に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１８】
前記２つ以上のコードブックの各々が異なる塩基順序を含む、請求項１１３に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１１９】
前記命令が、前記シンセサイザに前記複数のポリヌクレオチドをさらに生成させる、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１２０】
前記複数のポリヌクレオチドをシーケンシングして、複数の出力配列を生成するためのシーケンサをさらに備える、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１２１】
前記命令が、前記コンピューティング・システムに前記複数の出力配列をさらに受信させる、請求項１２０に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１２２】
前記コンピューティング・システムが、（ｉｉｉ）前記複数の出力配列をデコードすることを含む動作をさらに実行する、請求項１２０に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１２３】
前記複数の出力配列が、貪欲アルゴリズム、最尤（ＭＬ）アルゴリズム、または混合貪欲ＭＬアルゴリズムを使用してデコードされる、請求項１２２に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１２４】
欠失、挿入、突然変異、またはこれらの任意の組み合わせの確率を計算することに少なくとも部分的に基づいて、前記複数の出力配列がデコードされる、請求項１２２に記載の

ハイブリッド有機インシリコ・プラットフォーム。
【請求項１２５】
前記複数のポリヌクレオチドを格納するためのストレージ・ユニットをさらに備える、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１２６】
前記動作が、前記シンセサイザ、シーケンサ、前記ストレージ・ユニット、またはこれらの任意の組み合わせの間で、前記複数のポリヌクレオチドを転送することをさらに含む、請求項１２５に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１２７】
特定の塩基転移が、フロー順序に従って合成を可能にする、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１２８】
前記コードブックが１２コードワードを含む、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１２９】
（ａ）（ｉｉ）が、前記コードブックに基づいてバイナリ・データを複数のポリヌクレオチド配列にマッピングすることを含む、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３０】
前記内部コーデックが、前記複数のポリヌクレオチド配列の長さ、ＧＣ含量、反復、エラー、またはこれらの任意の組み合わせを含む制約に対してさらに最適化される、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３１】
前記複数のポリヌクレオチド配列の４０％～６０％が、冗長性のためにエンコードする、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３２】
前記複数のポリヌクレオチドを生成することが、複数の合成サイクルを含む、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３３】
前記合成サイクルの数が、前記内部コーデックを使用してエンコードされないポリヌクレオチド配列を合成するために必要とされる前記合成サイクルの数と比較して減少する、請求項１３２に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３４】
前記合成サイクルの数の減少が、フロー順序に部分的に基づく、請求項１３３に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３５】
前記合成サイクルの数が少なくとも３０％減少する、請求項１３３に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３６】
前記合成サイクルの数が５０％減少する、請求項１３３に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３７】
前記合成サイクルの数が、１００塩基を含むポリヌクレオチド配列の場合、３００未満である、請求項１３３に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３８】
前記合成サイクルの数が、１００塩基を含むポリヌクレオチド配列の場合、１５５である、請求項１３３に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１３９】
前記ポリヌクレオチド配列が、１つまたは複数のＡ、Ｔ、Ｃ、またはＧを含む、請求項１３８に記載のハイブリッド有機インシリコ・プラットフォーム。

【請求項１４０】
前記複数のポリヌクレオチドを生成することが、塩基ごとの合成を含む、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１４１】
前記シンセサイザが、複数の特徴を備えている固体支持体を備える、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１４２】
前記複数の特徴の各々が、前記固体支持体の１つまたは複数の電極を介して独立してアドレス指定可能である、請求項１４１に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１４３】
前記複数の特徴の各々が、マスキングを介してアドレス指定可能である、請求項１４１に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１４４】
前記マスキングが物理的バリアを含む、請求項１４３に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１４５】
前記マスキングが、前記複数の特徴のうちの１つまたは複数で反応性を制御することを含む、請求項１４３に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１４６】
反応性を制御することが、前記複数の特徴のうちの１つまたは複数での脱保護を含む、請求項１４５に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１４７】
前記脱保護が酸生成を含む、請求項１４６に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１４８】
前記脱保護が電気化学的脱保護である、請求項１４６に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１４９】
合成サイクルごとに、前記複数の特徴の２５％より多くがデブロッキングされる、請求項１４１に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１５０】
合成サイクルごとに、前記複数の特徴の少なくとも５０％がデブロッキングされる、請求項１４１に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１５１】
前記複数のポリヌクレオチド配列の各々が同じ長さを有する、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１５２】
前記複数のポリヌクレオチド配列の８０％～１００％が同じ長さを有する、請求項９７に記載のハイブリッド有機インシリコ・プラットフォーム。
【請求項１５３】
データをＤＮＡに格納するためのシステムであって、
１つまたは複数の処理ユニットと、
前記１つまたは複数の処理ユニットと通信するメモリと、
命令であって、前記メモリに格納され、前記１つまたは複数の処理ユニット上で実行されて、前記システムに、
複数のプールを生成することであって、前記複数のプールの各々がプール記述子、前記データのペイロードを含むプール・アイテム、および終了記述子を含む、生成すること、
プール・アイテムごとの前記ペイロードの第１の１つまたは複数のハッシュを決定す

ること、および
エンコーディング方式を適用して、前記複数のプールを複数のポリヌクレオチドの配列としてエンコードすること
を実行させる、命令と
を備える、システム。
【請求項１５４】
前記データが１つまたは複数のオブジェクトを含む、請求項１５３に記載のシステム。
【請求項１５５】
命令が前記メモリに格納され、前記１つまたは複数の処理ユニット上で実行されて、前記システムに、前記１つまたは複数のオブジェクトの各々の第２の１つまたは複数のハッシュを決定させる、請求項１５４に記載のシステム。
【請求項１５６】
前記１つまたは複数のオブジェクトが、ファイルまたは前記ファイルに関連付けられたメタデータを含む、請求項１５３に記載のシステム。
【請求項１５７】
前記プール記述子が、バージョン、プールＩＤ、プール・アイテム記述子のリスト、またはこれらの任意の組み合わせを含む、請求項１５３に記載のシステム。
【請求項１５８】
前記プールＩＤが一意のＩＤを含む、請求項１５７に記載のシステム。
【請求項１５９】
前記一意のＩＤが、汎用一意識別子（ＵＵＩＤ）またはコンテンツＩＤを含む、請求項１５８に記載のシステム。
【請求項１６０】
前記プール・アイテム記述子のリストが、オブジェクトのパス、オブジェクトのサイズ、オブジェクト内の前記プール・アイテムの範囲、プール内の前記プール・アイテムのオフセット、またはこれらの任意の組み合わせを含む、請求項１５７に記載のシステム。
【請求項１６１】
前記１つまたは複数のプール・アイテムの各々が、前記第１の１つまたは複数のハッシュからの前記プール・アイテムのハッシュ、またはこれらの組み合わせをさらに含む、請求項１５３に記載のシステム。
【請求項１６２】
終了プール記述子が、オブジェクト記述子のリストを含む、請求項１５３に記載のシステム。
【請求項１６３】
前記オブジェクト記述子のリストが、オブジェクトのパス、前記第１の１つまたは複数のハッシュからのオブジェクトのハッシュ、またはこれらの組み合わせを含む、請求項１６２に記載のシステム。
【請求項１６４】
前記複数のプールの各々が、約１ＧＢ～約１ＴＢである、請求項１５３に記載のシステム。
【請求項１６５】
前記複数のプールが冗長なプールを含む、請求項１５３に記載のシステム。
【請求項１６６】
前記第１の１つまたは複数のハッシュ、前記第２の１つまたは複数のハッシュ、またはその両方が、ハッシング・モジュールを使用して決定される、請求項１５５に記載のシステム。
【請求項１６７】
前記ハッシング・モジュールが、前記１つまたは複数の処理ユニット上で実行される、請求項１６５に記載のシステム。
【請求項１６８】

前記第１の１つまたは複数のハッシュが、前記１つまたは複数のオブジェクトよりも少ないメモリを必要とする、請求項１５３に記載のシステム。
【請求項１６９】
前記第２の１つまたは複数のハッシュが、前記１つまたは複数のプール・アイテムよりも少ないメモリを必要とする、請求項１５３に記載のシステム。
【請求項１７０】
前記ハッシング・モジュールがハッシュ関数を含む、請求項１６５に記載のシステム。
【請求項１７１】
前記ハッシュ関数が、ＳＨＡ－２２４、ＳＨＡ－２５６、ＳＨＡ－３８４、ＳＨＡ－５１２、ＳＨＡ－５１２／２２４、またはＳＨＡ－５１２／２５６を含む、請求項１７０に記載のシステム。
【請求項１７２】
前記命令が、前記システムに１つまたは複数のインデックス・プールをさらに生成させる、請求項１５３に記載のシステム。
【請求項１７３】
前記１つまたは複数のインデックス・プールが、インデックス・プール記述子およびオブジェクト・インデックス作成のリストを含む、請求項１７２に記載のシステム。
【請求項１７４】
前記インデックス・プール記述子が、バージョン、プールＩＤ、プールのサイズ、およびタイムスタンプを含む、請求項１７３に記載のシステム。
【請求項１７５】
前記プールＩＤが一意のＩＤを含む、請求項１７４に記載のシステム。
【請求項１７６】
前記一意のＩＤが、ＵＵＩＤまたはコンテンツＩＤを含む、請求項１７５に記載のシステム。
【請求項１７７】
前記オブジェクト・インデックス作成のリストが、オブジェクトのパス、オブジェクトのハッシュ、オブジェクト・フラグメントのリスト、オブジェクト・メタデータのリスト、またはこれらの任意の組み合わせを含む、請求項１７３に記載のシステム。
【請求項１７８】
前記オブジェクト・フラグメントのリストが、フラグメントを含むプールのプールＩＤ、フラグメントの範囲、またはこれらの組み合わせを含む、請求項１７７に記載のシステム。
【請求項１７９】
前記オブジェクト・メタデータのリストが、メタデータの種類、メタデータ・ペイロード、またはこれらの組み合わせを含む、請求項１７７に記載のシステム。
【請求項１８０】
前記メタデータの種類が、オブジェクトに添付されたキーワードのリスト、サムネイル画像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはこれらの任意の組み合わせを含む、請求項１７９に記載のシステム。
【請求項１８１】
前記１つまたは複数のインデックス・プールの各々が約１ＧＢ～約１ＴＢである、請求項１７２に記載のシステム。
【請求項１８２】
前記命令が前記メモリに格納され、前記１つまたは複数の処理ユニット上で実行されて、前記システムに、前記ＤＮＡに格納された前記データを取り出させる、請求項１５３に記載のシステム。
【請求項１８３】
前記命令が、
デコーディング方式を適用して、前記複数のプールの各々における前記複数のポリヌク

レオチドの前記配列をデコードすることと、
前記第１の１つまたは複数のハッシュを使用して、少なくとも各プール・アイテムの前記ペイロードを検証することとを含む、請求項１８２に記載のシステム。
【請求項１８４】
１つまたは複数のコンパートメントを備えている、情報をＤＮＡに格納するためのデバイスであって、各コンパートメントが、
（ａ）複数のポリヌクレオチドを含むライブラリであって、１つまたは複数のオブジェクトに対応する情報を含むプールをエンコードする、ライブラリと、
（ｂ）前記複数のポリヌクレオチドを格納するための媒体とを備える、デバイス。
【請求項１８５】
前記１つまたは複数のコンパートメントが通信する、請求項１８４に記載のデバイス。
【請求項１８６】
前記１つまたは複数のコンパートメントが通信しない、請求項１８４に記載のデバイス。
【請求項１８７】
前記媒体が、固体、液体、気体、またはこれらの任意の組み合わせを含む、請求項１８４に記載のデバイス。
【請求項１８８】
媒体が、前記ＤＮＡ内のリン酸基に対する塩の陽イオンのモル比が２０：１未満の塩溶液を含む、請求項１８４に記載のデバイス。
【請求項１８９】
前記塩溶液が乾燥されて、乾燥生成物を作り出す、請求項１８８に記載のデバイス。
【請求項１９０】
表面を備えている固体支持体をさらに備える、請求項１８４に記載のデバイス。
【請求項１９１】
前記表面上に位置する複数の構造体をさらに備え、前記複数のポリヌクレオチドが前記複数の構造体から伸ばされる、請求項１８４に記載のデバイス。
【請求項１９２】
前記１つまたは複数のオブジェクトが、ファイルまたは前記ファイルに関連付けられたメタデータを含む、請求項１８４に記載のデバイス。
【請求項１９３】
前記プールが、プール記述子、１つまたは複数のプール・アイテム、および終了プール記述子を含む、請求項１８４に記載のデバイス。
【請求項１９４】
前記プール記述子が、バージョン、プールＩＤ、プール・アイテム記述子のリスト、またはこれらの任意の組み合わせを含む、請求項１９３に記載のデバイス。
【請求項１９５】
前記プールＩＤが一意のＩＤを含む、請求項１９４に記載のデバイス。
【請求項１９６】
前記一意のＩＤが、汎用一意識別子（ＵＵＩＤ）またはコンテンツＩＤを含む、請求項１９５に記載のデバイス。
【請求項１９７】
前記プール・アイテム記述子のリストが、オブジェクトのパス、オブジェクトのサイズ、オブジェクト内のプール・アイテムの範囲、プール内の前記プール・アイテムのオフセット、またはこれらの任意の組み合わせを含む、請求項１９４に記載のデバイス。
【請求項１９８】
前記１つまたは複数のプール・アイテムの各々が、データ・ペイロード、前記プール・アイテムのハッシュ、またはこれらの組み合わせを含む、請求項１８４に記載のデバイス。
【請求項１９９】

前記終了プール記述子が、オブジェクト記述子のリストを含む、請求項１８４に記載のデバイス。
【請求項２００】
前記オブジェクト記述子のリストが、オブジェクトのパス、オブジェクトのハッシュ、またはこれらの組み合わせを含む、請求項１９９に記載のデバイス。
【請求項２０１】
前記プールが約１ＧＢ～約１ＴＢのデジタル情報を含む、請求項１８４に記載のデバイス。
【請求項２０２】
１つまたは複数の第２のコンパートメントをさらに備え、前記１つまたは複数の第２のコンパートメントの各々が、インデックス・プールをエンコードする第２のライブラリを備える、請求項１８４に記載のデバイス。
【請求項２０３】
前記インデックス・プールが、インデックス・プール記述子およびオブジェクト・インデックス作成のリストを含む、請求項２０２に記載のデバイス。
【請求項２０４】
前記インデックス・プール記述子が、バージョン、プールＩＤ、プールのサイズ、およびタイムスタンプを含む、請求項２０３に記載のデバイス。
【請求項２０５】
前記プールＩＤが一意のＩＤを含む、請求項２０４に記載のデバイス。
【請求項２０６】
前記一意のＩＤが、ＵＵＩＤまたはコンテンツＩＤを含む、請求項２０５に記載のデバイス。
【請求項２０７】
前記オブジェクト・インデックス作成のリストが、オブジェクトのパス、オブジェクトのハッシュ、オブジェクト・フラグメントのリスト、オブジェクト・メタデータのリスト、またはこれらの任意の組み合わせを含む、請求項２０３に記載のデバイス。
【請求項２０８】
前記オブジェクト・フラグメントのリストが、フラグメントを含むプールのプールＩＤ、フラグメントの範囲、またはこれらの組み合わせを含む、請求項２０７に記載のデバイス。
【請求項２０９】
前記オブジェクト・メタデータのリストが、メタデータの種類、メタデータ・ペイロード、またはこれらの組み合わせを含む、請求項２０７に記載のデバイス。
【請求項２１０】
前記メタデータの種類が、オブジェクトに添付されたキーワードのリスト、サムネイル画像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはこれらの任意の組み合わせを含む、請求項２０９に記載のデバイス。
【請求項２１１】
前記インデックス・プールが約１ＧＢ～約１ＴＢである、請求項２０３に記載のデバイス。
【請求項２１２】
データを複数のポリヌクレオチドに格納するための方法であって、
複数のプールを生成することであって、前記複数のプールの各々がプール記述子、前記データのペイロードを含むプール・アイテム、および終了記述子を含む、生成することと、
プール・アイテムごとの前記ペイロードの第１の１つまたは複数のハッシュを決定することと、
エンコーディング方式を適用して、前記複数のプールを複数のヌクレオチドの配列としてエンコードすることとを含む、方法。

【請求項２１３】
前記データが１つまたは複数のオブジェクトを含む、請求項２１２に記載の方法。
【請求項２１４】
前記方法が、前記１つまたは複数のオブジェクトの各々の第２の１つまたは複数のハッシュを決定することをさらに含む、請求項２１３に記載の方法。
【請求項２１５】
前記複数のポリヌクレオチドを格納することをさらに含む、請求項２１２に記載の方法。
【請求項２１６】
前記複数のプールのうちの各プールに対応する前記複数のポリヌクレオチドのうちのポリヌクレオチドが、データ・ストレージ・システムの別々のコンテナに格納される、請求項２１３に記載の方法。
【請求項２１７】
前記複数のポリヌクレオチドを生成することをさらに含む、請求項２１２に記載の方法。
【請求項２１８】
前記複数のポリヌクレオチドを生成することが、デオキシリボ核酸（ＤＮＡ）のホスホラミダイトベースの合成を含む、請求項２１７に記載の方法。
【請求項２１９】
前記ホスホラミダイトベースの合成のための試薬が、ヌクレオシド・ホスホラミダイト、酸化剤、活性化剤、またはデブロッカーを含み、あるいは前記溶媒がアセトニトリルを含む、請求項２１７に記載の方法。
【請求項２２０】
前記複数のポリヌクレオチドを生成することが、酵素的ＤＮＡ合成を含む、請求項２１７に記載の方法。
【請求項２２１】
酵素的ＤＮＡ合成のための試薬が、末端デオキシヌクレオチジル・トランスフェラーゼ（ＴｄＴ）またはデブロッカーを含み、あるいは前記溶媒が水を含む、請求項２２０に記載の方法。
【請求項２２２】
前記１つまたは複数のオブジェクトが、ファイルまたは前記ファイルに関連付けられたメタデータを含む、請求項２１２に記載の方法。
【請求項２２３】
前記プール記述子が、バージョン、プールＩＤ、プール・アイテム記述子のリスト、またはこれらの任意の組み合わせを含む、請求項２１２に記載の方法。
【請求項２２４】
前記プールＩＤが一意のＩＤを含む、請求項２２３に記載の方法。
【請求項２２５】
前記一意のＩＤが、汎用一意識別子（ＵＵＩＤ）またはコンテンツＩＤを含む、請求項２２４に記載の方法。
【請求項２２６】
前記プール・アイテム記述子のリストが、オブジェクトのパス、オブジェクトのサイズ、オブジェクト内の前記プール・アイテムの範囲、プール内の前記プール・アイテムのオフセット、またはこれらの任意の組み合わせを含む、請求項２２３に記載の方法。
【請求項２２７】
前記１つまたは複数のプール・アイテムの各々が、前記第１の１つまたは複数のハッシュからの前記プール・アイテムのハッシュをさらに含む、請求項２１２に記載の方法。
【請求項２２８】
前記終了プール記述子が、オブジェクト記述子のリストを含む、請求項２２７に記載の方法。

【請求項２２９】
前記オブジェクト記述子のリストが、オブジェクトのパス、前記第１の１つまたは複数のハッシュからのオブジェクトのハッシュ、またはこれらの組み合わせを含む、請求項２２８に記載の方法。
【請求項２３０】
前記複数のプールの各々が、約１ＧＢ～約１ＴＢである、請求項２１２に記載の方法。
【請求項２３１】
前記複数のプールが冗長なプールを含む、請求項２１２に記載の方法。
【請求項２３２】
前記第１の１つまたは複数のハッシュ、前記第２の１つまたは複数のハッシュ、またはその両方が、ハッシング・モジュールを使用して決定される、請求項２１３に記載の方法。
【請求項２３３】
前記第２の１つまたは複数のハッシュが、前記１つまたは複数のオブジェクトよりも少ないメモリを必要とする、請求項２１３に記載の方法。
【請求項２３４】
前記第１の１つまたは複数のハッシュが、前記１つまたは複数のプール・アイテムよりも少ないメモリを必要とする、請求項２１２に記載の方法。
【請求項２３５】
前記ハッシング・モジュールがハッシュ関数を含む、請求項２３２に記載の方法。
【請求項２３６】
前記ハッシュ関数が、ＳＨＡ－２２４、ＳＨＡ－２５６、ＳＨＡ－３８４、ＳＨＡ－５１２、ＳＨＡ－５１２／２２４、またはＳＨＡ－５１２／２５６を含む、請求項２３５に記載の方法。
【請求項２３７】
１つまたは複数のインデックス・プールを作成することをさらに含む、請求項２１２に記載の方法。
【請求項２３８】
前記１つまたは複数のインデックス・プールが、インデックス・プール記述子およびオブジェクト・インデックス作成のリストを含む、請求項２３７に記載の方法。
【請求項２３９】
前記インデックス・プール記述子が、バージョン、プールＩＤ、プールのサイズ、およびタイムスタンプを含む、請求項２３８に記載の方法。
【請求項２４０】
前記プールＩＤが一意のＩＤを含む、請求項２３９に記載の方法。
【請求項２４１】
前記一意のＩＤが、ＵＵＩＤまたはコンテンツＩＤを含む、請求項２４０に記載の方法。
【請求項２４２】
前記オブジェクト・インデックス作成のリストが、オブジェクトのパス、オブジェクトのハッシュ、オブジェクト・フラグメントのリスト、オブジェクト・メタデータのリスト、またはこれらの任意の組み合わせを含む、請求項２３８に記載の方法。
【請求項２４３】
前記オブジェクト・フラグメントのリストが、フラグメントを含むプールのプールＩＤ、フラグメントの範囲、またはこれらの組み合わせを含む、請求項２４２に記載の方法。
【請求項２４４】
前記オブジェクト・メタデータのリストが、前記メタデータの種類、前記メタデータ・ペイロード、またはこれらの組み合わせを含む、請求項２４２に記載の方法。
【請求項２４５】
前記メタデータの種類が、オブジェクトに添付されたキーワードのリスト、サムネイル

画像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはこれらの任意の組み合わせを含む、請求項２４４に記載の方法。
【請求項２４６】
前記インデックス・プールのうちの１つまたは複数の各々が約１ＧＢ～約１ＴＢである、請求項２３７に記載の方法。
【請求項２４７】
複数のポリヌクレオチドに格納されたデータを取り出すための方法であって、
前記複数のポリヌクレオチドの配列を決定することであって、前記複数のポリヌクレオチドが複数のプール内にある、決定することと、
デコーディング方式を適用して、前記複数のプールの各々における前記複数のポリヌクレオチドの前記配列をデコードすることであって、各プールが、プール記述子、前記データのペイロードを含むプール・アイテム、および終了記述子を含む、デコードすることと、
第１の１つまたは複数のハッシュを使用して少なくとも各プール・アイテムの前記ペイロードを検証することとを含む、方法。
【請求項２４８】
前記データが１つまたは複数のオブジェクトを含む、請求項２４７に記載の方法。
【請求項２４９】
前記１つまたは複数のオブジェクトが、ファイルまたは前記ファイルに関連付けられたメタデータを含む、請求項２４８に記載の方法。
【請求項２５０】
前記方法が、第２の１つまたは複数のハッシュを使用して前記１つまたは複数のオブジェクトを検証することをさらに含む、請求項２４８に記載の方法。
【請求項２５１】
少なくとも前記ペイロードを検証することが、ハッシュ関数を使用して前記第１の１つまたは複数のハッシュを検証することを含む、請求項２４７に記載の方法。
【請求項２５２】
前記データを取り出すために、各プール・アイテムからの前記ペイロードを結合することをさらに含む、請求項２４７に記載の方法。
【請求項２５３】
前記データをメモリに格納することをさらに含む、請求項２４７に記載の方法。
【請求項２５４】
前記複数のプールの各々が、約１ＧＢ～約１ＴＢである、請求項２４７に記載の方法。
【請求項２５５】
前記１つまたは複数のオブジェクトを検証することが、ハッシュ関数を使用して前記第２の１つまたは複数のハッシュを検証することを含む、請求項２５０に記載の方法。
【請求項２５６】
前記ハッシュ関数が、ＳＨＡ－２２４、ＳＨＡ－２５６、ＳＨＡ－３８４、ＳＨＡ－５１２、ＳＨＡ－５１２／２２４、またはＳＨＡ－５１２／２５６を含む、請求項２５１に記載の方法。
【請求項２５７】
前記配列を決定することが、前記複数のポリヌクレオチドをシーケンシングすることを含む、請求項２４７に記載の方法。
【請求項２５８】
シーケンシングが、次世代シーケンシング、並列シーケンシング、単一分子リアルタイム・シーケンシング、ナノポア・シーケンシング、合成によるシーケンシング、サンガー・シーケンシング、またはこれらの任意の組み合わせを含む、請求項２５７に記載の方法。
【請求項２５９】
前記１つまたは複数のオブジェクトを含む複数のプールを決定するために、１つまたは

複数のインデックス・プールのうちの１つのインデックス・プールにアクセスすることをさらに含む、請求項２４８に記載の方法。
【請求項２６０】
前記インデックス・プールが、インデックス・プール記述子およびオブジェクト・インデックス作成のリストを含む、請求項２５９に記載の方法。
【請求項２６１】
前記インデックス・プール記述子が、バージョン、プールＩＤ、プールのサイズ、およびタイムスタンプを含む、請求項２６０に記載の方法。
【請求項２６２】
前記プールＩＤが一意のＩＤを含む、請求項２６１に記載の方法。
【請求項２６３】
前記一意のＩＤが、ＵＵＩＤまたはコンテンツＩＤを含む、請求項２６２に記載の方法。
【請求項２６４】
前記オブジェクト・インデックス作成のリストが、オブジェクトのパス、オブジェクトのハッシュ、オブジェクト・フラグメントのリスト、オブジェクト・メタデータのリスト、またはこれらの任意の組み合わせを含む、請求項２６０に記載の方法。
【請求項２６５】
前記オブジェクト・フラグメントのリストが、フラグメントを含むプールのプールＩＤ、フラグメントの範囲、またはこれらの組み合わせを含む、請求項２６４に記載の方法。
【請求項２６６】
前記オブジェクト・メタデータのリストが、前記メタデータの種類、前記メタデータ・ペイロード、またはこれらの組み合わせを含む、請求項２６４に記載の方法。
【請求項２６７】
前記メタデータの種類が、オブジェクトに添付されたキーワードのリスト、サムネイル画像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはこれらの任意の組み合わせを含む、請求項２６６に記載の方法。
【請求項２６８】
前記インデックス・プールのうちの１つまたは複数の各々が約１ＧＢ～約１ＴＢである、請求項２５９に記載の方法。
【発明の詳細な説明】
【技術分野】
【０００１】
相互参照
本出願は、２０２２年４月２１日に出願された米国特許仮出願第６３／３３３，３０５号、２０２２年５月５日に出願された米国特許仮出願第６３／３３８，７６０号、および２０２３年１月２７日に出願された米国特許仮出願第６３／４８１，８７３号の利益を主張し、これらは、全体として参照によって組み込まれている。
【０００２】
参照による組み込み
本明細書において言及されたすべての公開文献、特許、および特許出願は、本明細書では、個々の公開文献、特許、または特許出願が具体的かつ個別に参照によって組み込まれるように示された場合と同じ程度まで、参照によって組み込まれている。
【背景技術】
【０００３】
ＤＮＡは、現在使用されている電子媒体と比較して、その優れた密度、安定性、エネルギー効率、および寿命を前提として、魅力的なデータ・ストレージ媒体である。しかし、シーケンシングならびにシーケンシングに関連する動作およびプロセスのさまざまな段階で、またはさまざまな段階の間に、エラーおよび曖昧さが導入されるか、または発生する可能性がある。したがって、そのようなエラーが存在する場合に、ＤＮＡを効率的にエン

コードおよびデコードする方法を開発する必要がある。

【発明の概要】

【０００４】

本明細書では、デジタル・データ（例えば、バイナリ・データ）をオリゴ・プールにエンコードし、プールをデジタル・データに再びデコードするさまざまなコーデックの設計および実装が提供される。コーデックは、デジタル・データを塩基に変換するための内部コーデックを備えることができる。コーデックは、格納されるデータを多くのオリゴに分散させるための外部コーデックを備え、消失を修正するための冗長性を構築することもできる。本明細書に記載されたコーデックは、合成、格納、および／またはシーケンシング中のオリゴの損失、ならびに欠失、突然変異、および挿入の高い割合に耐えることができる。一部の実施形態では、本明細書に記載されたコーデックは、低いシーケンシング・カバレッジ用に設計されている。一部の実施形態では、本明細書に記載されたコーデックは、複数のポリヌクレオチドの合成を最適化するために設計されている。

【０００５】

さらに、本明細書では、複数のポリヌクレオチドからデジタル情報を取り出す方法が提供される。コーデックは、１つまたは複数のプール内の、デジタル情報を含んでいる１つまたは複数のオブジェクトの格納をサポートする、バケットのようなストレージ・システムを備えてよい。コーデックは、効率的なデータ・ストレージのために、インデックス作成（例えば、インデックス・プール）およびハッシュ処理（例えば、ハッシング・モジュール）などのストレージ戦略をさらに含んでよい。コーデックは、デジタル情報の格納または取り出し中に発生する可能性のある消失またはエラーを修正するために、１つまたは複数のプールに冗長性を構築することもできる。

【０００６】

１つの態様では、本明細書において複数のポリヌクレオチド配列内のデータをエンコードするための方法が提供され、この方法は、（ａ）データを複数のフレームに分割することであって、複数のフレームのうちの各フレームがフレーム・インデックスを含む、分割することと、（ｂ）外部コーデックを複数のフレームのうちの各フレームに適用することであって、外部コーデックがエラー訂正方式を含む、適用することと、（ｃ）各フレームを複数のレーンに分割することであって、複数のレーンのうちの各レーンがレーン・インデックスを含む、分割することと、（ｄ）レーン・インデックスに少なくとも部分的に基づいて各レーンをシャッフルすることと、（ｅ）内部コーデックを適用して、複数のポリヌクレオチド配列のうちの１つのポリヌクレオチド配列内の各レーンをエンコードすることとを含む。場合によっては、データは複数のシンボルを含む。場合によっては、データはバイナリ・データを含む。場合によっては、バイナリ・データは、バイト・ストリームまたはバイト配列を含む。場合によっては、（ｄ）におけるシャッフルすることは、各レーン内の回転方式を含む。場合によっては、（ｄ）におけるシャッフルすることは、各レーン内の擬似ランダム・プロセスを含む。場合によっては、（ｄ）におけるシャッフルすることは、エラーに対する耐性をもたらす。場合によっては、エラーは、ヌクレオチド合成エラーまたはシーケンシング・エラーである。場合によっては、エラーは、欠失、挿入、または置換を含む。場合によっては、エラー訂正方式は、リード・ソロモン（ＲＳ：Ｒｅｅｄ－Ｓｏｌｏｍｏｎ）コード、低密度パリティ・チェック（ＬＤＰＣ：ｌｏｗ－ｄｅｎｓｉｔｙ　ｐａｒｉｔｙ－ｃｈｅｃｋ）コード、ターボコード、極コード、またはこれらの任意の組み合わせを含む。場合によっては、データは、少なくとも約１ＧＢ～約１ＴＢを含む。場合によっては、複数のフレームは、約１００～約１０，０００フレームを含む。場合によっては、各フレームは、最大で約５０００レーンを含む。場合によっては、各レーンは、約１００～約３００ビットを含む。場合によっては、フレーム・インデックスは、約１６ビット～約２０ビットを含む。場合によっては、レーン・インデックスは、約１２ビットまたは約１６ビットを含む。場合によっては、ポリヌクレオチド配列の長さは、約１００～約３００塩基である。場合によっては、（ｄ）の前に、フレーム・インデックスおよび／またはレーン・インデックスが各レーンの先頭に追加される。場合によっ

ては、内部コーデックを適用することは、複数のポリヌクレオチド配列にわたって冗長性を追加することを含む。場合によっては、冗長性は約５％～約１０％である。場合によっては、一部分において複数のポリヌクレオチド配列にわたる冗長性のため、エラーが存在する場合に、複数のポリヌクレオチド配列がデコードされ得る。場合によっては、エラーは、挿入、欠失、置換、またはこれらの任意の組み合わせを含む。場合によっては、内部コーデックを適用することは、（ａ）レーンからのシンボル、シンボル履歴、およびシンボル位置を組み合わせることと、（ｂ）ルックアップ・テーブル、ハッシュ、またはその両方を使用して塩基候補を生成することとを含む。場合によっては、この方法は、塩基反復チェックを実行することをさらに含む。場合によっては、シンボルはビットである。場合によっては、この方法は、シンボル履歴を更新すること、レーン・インデックスをインクリメントすること、フレーム・インデックスをインクリメントすること、またはこれらの任意の組み合わせをさらに含む。場合によっては、更新されたシンボル履歴、インクリメントされたレーン・インデックス、インクリメントされたフレーム・インデックス、またはこれらの任意の組み合わせが、後続のレーンのシンボルと結合される。場合によっては、この方法は、複数のポリヌクレオチド配列を合成する前にＧＣフィルタリングを実行することをさらに含む。場合によっては、ＧＣフィルタリングは、複数のレーンのうちの約５％～約１０％のレーンを除去することを含む。場合によっては、複数のポリヌクレオチド配列は、約４５％～約５５％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列の少なくとも９０％が、約４５％～約５５％のＧＣ含量を含む。場合によっては、内部コーデックを適用することは、（ａ）ルックアップ・テーブルを使用して、レーン内のシンボルごとに塩基候補を生成することと、（ｂ）以前にエンコードされたシンボルに少なくとも部分的に基づいて、次のルックアップ・テーブルを選択することとを含む。場合によっては、内部コーデックを適用することは、エンコーディング方式を適用することを含む。

【０００７】

別の態様では、本明細書において複数のポリヌクレオチド配列をデコードして、データを含む出力を生成するための方法が提供され、この方法は、（ａ）複数のポリヌクレオチド配列を決定することと、（ｂ）内部コーデックを複数のポリヌクレオチド配列に適用することであって、内部コーデックが、複数のポリヌクレオチド配列の各々を、複数のシンボルを含むレーンに変換し、内部コーデックが、貪欲アルゴリズムおよび最尤（ＭＬ：ｍａｘｉｍｕｍ　ｌｉｋｅｌｉｈｏｏｄ）アルゴリズムを含んでいる混合デコーディング・アルゴリズム（ｍｉｘｅｄ　ｄｅｃｏｄｉｎｇ　ａｌｇｏｒｉｔｈｍ）を含む、適用することと、（ｃ）データのレーンを、レーン・インデックスおよび各レーンのフレーム・インデックスに基づいてフレームに配置することと、（ｄ）外部コーデックをフレームに適用することであって、外部コーデックがエラー訂正方式を含み、外部コーデックからのフレームがマージされて、データを含む出力を生成する、適用することとを含む。場合によっては、データは複数のシンボルを含む。場合によっては、データはバイナリ・データを含む。場合によっては、バイナリ・データは、バイト・ストリームまたはバイト配列を含む。場合によっては、内部コーデックはデコーディング方式を含む。場合によっては、この方法は、（ｂ）の前にポリヌクレオチド配列をクラスタ化することをさらに含む。場合によっては、このクラスタ化することはインデックスに基づく。場合によっては、このクラスタ化することは、フレーム・インデックス、レーン・インデックス、またはその両方をデコードすることを部分的に含む。場合によっては、このクラスタ化することは、ハッシュ関数を使用して実行される。場合によっては、この方法は、（ｂ）の前にポリヌクレオチド配列をアラインメントすることをさらに含む。場合によっては、アラインメントは、アラインメント・アルゴリズムを使用してヌクレオチドのコンセンサスを分析することを含む。場合によっては、アライメント・アルゴリズムは、ペアワイズ・アライメント・アルゴリズム、マルチシーケンス・アライメント・アルゴリズム、またはこれらの組み合わせを含む。場合によっては、アライメント・アルゴリズムは、（ａ）複数の読み取りのうちの読み取りごとに位置を初期化することであって、初期化が、ポリヌクレオチド配列

を位置０にアライメントすることを含む、初期化することと、（ｂ）各読み取り間に次の１つまたは複数の塩基のコンセンサスを分析することと、（ｃ）読み取りごとに、次の１つまたは複数の塩基の各々が正しいか、またはエラーを有するかを含む判定を決定することと、（ｄ）読み取りごとに、この判定を前提として位置をインクリメントすることと、（ｅ）ステップ（ｂ）～（ｄ）を繰り返すこととを含む。場合によっては、エラーは欠失、置換、または挿入である。場合によっては、複数の読み取りは、約３～約１０回の読み取りを含む。場合によっては、各読み取りの長さは、約１００～約３００塩基である。場合によっては、次の１つまたは複数の塩基は、約２、３、４、または５塩基である。場合によっては、混合デコーディング・アルゴリズムは、１つまたは複数の状態からの遷移確率に基づいてデコードすることを含む。場合によっては、１つまたは複数の状態は、約１００～約１０００個の最も可能性の高い状態を含む。場合によっては、内部コーデックは、ドリフト項をさらに含む。場合によっては、ドリフト項は整数を含む。場合によっては、整数は、ポリヌクレオチド配列内の挿入または欠失の総数に関連付けられる。場合によっては、整数は、挿入、欠失、またはその両方の総数のうちの、１つもしくは複数の挿入の値または１つもしくは複数の欠失の値を合計することによって計算される。場合によっては、１つまたは複数の挿入の各々の値は＋１を含み、１つまたは複数の欠失の各々の値は－１を含む。場合によっては、（ｃ）は、レーン・インデックスに基づいてレーンをシャッフル解除し、フレーム・インデックスに基づいてレーンをフレームにグループ化することを含む。場合によっては、エラー訂正方式は、リード・ソロモン（ＲＳ）コード、低密度パリティ・チェック（ＬＤＰＣ）コード、ターボコード、極コード、またはこれらの任意の組み合わせを含む。場合によっては、複数のポリヌクレオチド配列のうちの少なくとも１つのポリヌクレオチド配列がエラーを含む。場合によっては、エラーは、挿入、欠失、置換、またはこれらの任意の組み合わせを含む。

【０００８】

別の態様では、本明細書において、（ａ）メモリ、および（ｂ）メモリに動作可能に結合された処理デバイスを備えている装置が提供され、処理デバイスは、（ｉ）データを複数のフレームに分割することであって、複数のフレームのうちの各フレームがフレーム・インデックスを含む、分割することと、（ｉｉ）外部コーデックを複数のフレームのうちの各フレームに適用することであって、外部コーデックがエラー訂正方式を含む、適用することと、（ｉｉｉ）各フレームを複数のレーンに分割することであって、複数のレーンのうちの各レーンがレーン・インデックスを含む、分割することと、（ｉｖ）レーン・インデックスに少なくとも部分的に基づいて各レーンをシャッフルすることと、（ｖ）内部コーデックを適用して、ポリヌクレオチド配列内の各レーンをエンコードすることとを実行するように構成される。場合によっては、ポリヌクレオチド配列にエラーが存在する場合にデジタル・データがデコードされ得るように、内部コーデックは冗長性を追加する。場合によっては、内部コーデックはエンコーディング方式を含む。場合によっては、データは複数のシンボルを含む。場合によっては、データはデジタル・データを含む。場合によっては、装置は、ポリヌクレオチド配列を生成するためのシンセサイザをさらに備える。場合によっては、メモリ、処理デバイス、またはその両方は、コンピューティング・システムの一部である。場合によっては、コンピューティング・システムはクラウド・コンピューティング・システムを含む。場合によっては、クラウド・コンピューティング・システムは、プライベート・クラウド、パブリック・クラウド、ハイブリッド・クラウド、マルチクラウド、またはこれらの任意の組み合わせを含む。場合によっては、クラウド・コンピューティング・システムは、サービスとしてのインフラストラクチャ（ＩａａＳ：ｉｎｆｒａｓｔｒｕｃｔｕｒｅ ａｓ ａ ｓｅｒｖｉｃｅ）、サービスとしてのプラットフォーム（ＰａａＳ：ｐｌａｔｆｏｒｍ ａｓ ａ ｓｅｒｖｉｃｅ）、サービスとしてのソフトウェア（ＳａａＳ：ｓｏｆｔｗａｒｅ ａｓ ａ ｓｅｒｖｉｃｅ）、またはこれらの任意の組み合わせを含む。

【０００９】

別の態様では、本明細書において、（ａ）メモリと、（ｂ）複数のポリヌクレオチドの

配列を決定するように構成されたシーケンシング・デバイスと、（ｃ）メモリおよびシーケンシング・デバイスに動作可能に結合された処理デバイスとを備えている装置が提供され、処理デバイスは、（ｉ）内部コーデックを配列に適用することであって、内部コーデックが、配列の各々を、複数のシンボルを含むレーンに変換し、内部コーデックが、貪欲アルゴリズムおよび最尤（ＭＬ）アルゴリズムを含んでいる混合デコーディング・アルゴリズムを含む、適用することと、（ｉｉ）レーンを、レーン・インデックスおよび各レーン内のフレーム・インデックスに基づいてフレームに配置することと、（ｉｉｉ）外部コーデックをフレームに適用することであって、外部コーデックがエラー訂正方式を含み、外部コーデックからのフレームがマージされて、データを含む出力を生成する、適用することとを実行するように構成される。場合によっては、内部コーデックはデコーディング方式を含む。場合によっては、データは複数のシンボルを含む。場合によっては、データはデジタル・データを含む。場合によっては、メモリ、処理デバイス、またはその両方は、コンピューティング・システムの一部である。場合によっては、コンピューティング・システムは、クラウド・コンピューティング・システムを含む。場合によっては、クラウド・コンピューティング・システムは、プライベート・クラウド、パブリック・クラウド、ハイブリッド・クラウド、マルチクラウド、またはこれらの任意の組み合わせを含む。場合によっては、クラウド・コンピューティング・システムは、サービスとしてのインフラストラクチャ（ＩａａＳ）、サービスとしてのプラットフォーム（ＰａａＳ）、サービスとしてのソフトウェア（ＳａａＳ）、またはこれらの任意の組み合わせを含む。

【００１０】

別の態様では、本明細書において、ポリヌクレオチド配列内のデータをエンコードするための方法が提供され、この方法は、（ａ）コードブックを含む内部コーデックを生成することであって、コードブックが１つまたは複数の制約に対して最適化される、生成することと、（ｂ）内部コーデックを適用して、データを複数のポリヌクレオチド配列としてエンコードすることとを含む。場合によっては、データは複数のシンボルを含む。場合によっては、データはバイナリ・データを含む。場合によっては、１つまたは複数の制約は、核酸合成、後処理、ストレージ、シーケンシング、またはこれらの任意の組み合わせに関連する。場合によっては、核酸合成は、電気化学的合成、酵素合成、ホスホラミダイト合成、インクジェット印刷、またはこれらの任意の組み合わせを含む。場合によっては、核酸合成に関連する１つまたは複数の制約は合成エラーを含む。場合によっては、合成エラーは、挿入、欠失、または突然変異を含む。場合によっては、後処理は、ライゲーション、切断、ハイブリダイゼーション、変性、固体支持体への固定、伸長、エラー訂正、濃縮、分離、精製、または増幅のうちの１つまたは複数を含む。場合によっては、ストレージはコールド・データ・ストレージを含む。場合によっては、ストレージは、液相または固相での核酸ストレージ（ｎｕｃｌｅｉｃ ａｃｉｄ ｓｔｏｒａｇｅ）を含む。場合によっては、ストレージに関連する１つまたは複数の制約は、温度、湿度、圧力、塩分濃度、ｐＨ、濃度、時間、光、ＵＶ、O_2、またはこれらの任意の組み合わせを含む。場合によっては、温度は室温を含む。場合によっては、シーケンシングは、次世代シーケンシング、並列シーケンシング、単一分子リアルタイム・シーケンシング、ナノポア・シーケンシング、合成によるシーケンシング、サンガー・シーケンシング、またはこれらの任意の組み合わせを含む。場合によっては、（ｃ）複数のポリヌクレオチド配列を含んでいる複数のポリヌクレオチドを合成することをさらに含む。場合によっては、コードブックは、塩基順序に部分的に基づいて生成されるコードワードを含む。場合によっては、塩基順序は、既定の塩基転移を含む。場合によっては、内部コーデックは２つ以上のコードブックを含む。場合によっては、２つ以上のコードブックの各々は、複数のポリヌクレオチドの合成中に層をエンコードする。場合によっては、層は、少なくとも１つの塩基による複数のポリヌクレオチドのうちの各ポリヌクレオチドの伸長を含む。場合によっては、層の合成は１つまたは複数のサイクルを含み、１つまたは複数のサイクルの各々は、コードブックの１つまたは複数の塩基転移に従って塩基を流すことを含む。場合によっては、１つまたは複数のサイクルのうちの１つのサイクルは、Ａ、Ｔ、Ｃ、またはＧのうちの１つまた

は複数の追加を含む。場合によっては、２つ以上のコードブックの各々は、異なる塩基順序を含む。場合によっては、コードブックは、約１２コードワードを含む。場合によっては、（ｂ）は、コードブックに基づいてデータを複数のポリヌクレオチド配列にマッピングすることを含む。場合によっては、内部コーデックは、複数のポリヌクレオチド配列の長さ、ＧＣ含量、反復、エラー、またはこれらの任意の組み合わせを含む１つまたは複数の制約に対して、さらに最適化される。場合によっては、複数のポリヌクレオチド配列の４０％～６０％が、冗長性のためにエンコードする。場合によっては、合成は複数の合成サイクルを含む。場合によっては、この合成サイクルの数は、内部コーデックを使用してエンコードされないポリヌクレオチド配列を合成するために必要とされる合成サイクルの数と比較して、減少する。場合によっては、合成サイクルの数の減少は、フロー順序に部分的に基づく。場合によっては、合成サイクルの数は、少なくとも３０％減少する。場合によっては、合成サイクルの数は５０％減少する。場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、３００未満である。場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、約１５５である。場合によっては、ポリヌクレオチド配列は、Ａ、Ｔ、Ｃ、またはＧのうちの１つまたは複数を含む。場合によっては、（ｃ）は、固体支持体上で複数のポリヌクレオチドを合成することを含む。場合によっては、固体支持体は、複数の特徴を備える。場合によっては、合成サイクルごとに、複数の特徴の２５％より多くがデブロッキングされる。場合によっては、合成サイクルごとに、複数の特徴の少なくとも５０％がデブロッキングされる。場合によっては、複数のポリヌクレオチド配列の各々は、同じ長さを有する。場合によっては、複数のポリヌクレオチド配列の８０％～１００％は、同じ長さを有する。場合によっては、複数のポリヌクレオチドをシーケンシングして、複数の出力配列を生成することをさらに含む。場合によっては、複数の出力配列は、貪欲アルゴリズム、最尤（ＭＬ）アルゴリズム、または混合貪欲ＭＬアルゴリズムを使用してデコードされる。場合によっては、エラーの確率を計算することに少なくとも部分的に基づいて、複数の出力配列がデコードされる。場合によっては、エラーは、欠失、挿入、突然変異、またはこれらの任意の組み合わせを含む。

【００１１】

別の態様では、本明細書において、データをエンコードするためのハイブリッド有機インシリコ・プラットフォーム（ｈｙｂｒｉｄ　ｏｒｇａｎｉｃ－ｉｎ　ｓｉｌｉｃｏ　ｐｌａｔｆｏｒｍ）が提供され、このプラットフォームは、（ａ）少なくとも１つのプロセッサ、および動作を実行するために少なくとも１つのプロセッサによって実行可能な命令を備えるコンピューティング・システムであって、これらの動作が、（ｉ）コードブックを含む内部コーデックを生成することであって、コードブックが１つまたは複数の制約に対して最適化される、生成することと、（ｉｉ）内部コーデックを適用して、データを複数のポリヌクレオチド配列としてエンコードすることとを含む、コンピューティング・システムと、（ｂ）複数のポリヌクレオチド配列を含む複数のポリヌクレオチドを生成するためのシンセサイザとを構成する。場合によっては、データは複数のシンボルを含む。場合によっては、１つまたは複数の制約は、核酸合成、後処理、ストレージ、シーケンシング、またはこれらの任意の組み合わせに関連する。場合によっては、核酸合成は、電気化学的合成、酵素合成、ホスホラミダイト合成、インクジェット印刷、またはこれらの任意の組み合わせを含む。場合によっては、核酸合成に関連する１つまたは複数の制約は合成エラーを含む。場合によっては、合成エラーは、挿入、欠失、または突然変異を含む。場合によっては、後処理は、ライゲーション、切断、ハイブリダイゼーション、変性、固体支持体への固定、伸長、エラー訂正、濃縮、分離、精製、および増幅のうちの１つまたは複数を含む。場合によっては、ストレージはコールド・データ・ストレージを含む。場合によっては、ストレージは、液相または固相での核酸ストレージを含む。場合によっては、ストレージに関連する１つまたは複数の制約は、温度、湿度、圧力、塩分濃度、ｐＨ、濃度、時間、光、ＵＶ、O_2、またはこれらの任意の組み合わせを含む。場合によっては、温度は室温を含む。場合によっては、シーケンシングは、次世代シーケンシング、並列

シーケンシング、単一分子リアルタイム・シーケンシング、ナノポア・シーケンシング、合成によるシーケンシング、サンガー・シーケンシング、またはこれらの任意の組み合わせを含む。場合によっては、コンピューティング・システムは、クラウド・コンピューティング・システムを含む。場合によっては、クラウド・コンピューティング・システムは、プライベート・クラウド、パブリック・クラウド、ハイブリッド・クラウド、マルチクラウド、またはこれらの任意の組み合わせを含む。場合によっては、クラウド・コンピューティング・システムは、サービスとしてのインフラストラクチャ（ＩａａＳ）、サービスとしてのプラットフォーム（ＰａａＳ）、サービスとしてのソフトウェア（ＳａａＳ）、またはこれらの任意の組み合わせを含む。場合によっては、コードブックは、塩基順序に部分的に基づいて生成されるコードワードを含む。場合によっては、塩基順序は、既定の塩基転移を含む。場合によっては、内部コーデックは２つ以上のコードブックを含む。場合によっては、２つ以上のコードブックの各々は、複数のポリヌクレオチドの合成中に層をエンコードする。場合によっては、層は、少なくとも１つの塩基による複数のポリヌクレオチドのうちの各ポリヌクレオチドの伸長を含む。場合によっては、層の合成は１つまたは複数のサイクルを含み、１つまたは複数のサイクルの各々は、コードブックの１つまたは複数の塩基転移に従って塩基を流すことを含む。場合によっては、１つまたは複数のサイクルのうちの１つのサイクルは、Ａ、Ｔ、Ｃ、またはＧのうちの１つまたは複数の追加を含む。場合によっては、２つ以上のコードブックの各々は、異なる塩基順序を含む。場合によっては、命令は、シンセサイザに複数のポリヌクレオチドをさらに生成させる。場合によっては、複数のポリヌクレオチドをシーケンシングして、複数の出力配列を生成するためのシーケンサをさらに備える。場合によっては、命令は、コンピューティング・システムに複数の出力配列をさらに受信させる。場合によっては、コンピューティング・システムは、（ｉｉｉ）複数の出力配列をデコードすることを含む動作をさらに実行する。場合によっては、複数の出力配列は、貪欲アルゴリズム、最尤（ＭＬ）アルゴリズム、または混合貪欲ＭＬアルゴリズムを使用してデコードされる。場合によっては、欠失、挿入、突然変異、またはこれらの任意の組み合わせの確率を計算することに少なくとも部分的に基づいて、複数の出力配列がデコードされる。場合によっては、複数のポリヌクレオチドを格納するためのストレージ・ユニットをさらに備える。場合によっては、動作は、シンセサイザ、シーケンサ、ストレージ・ユニット、またはこれらの任意の組み合わせの間で、複数のポリヌクレオチドを転送することをさらに含む。場合によっては、特定の塩基転移は、フロー順序に従って合成を可能にする。場合によっては、コードブックは、約１２コードワードを含む。場合によっては、（ａ）（ｉｉ）は、コードブックに基づいてデータを複数のポリヌクレオチド配列にマッピングすることを含む。場合によっては、内部コーデックは、複数のポリヌクレオチド配列の長さ、ＧＣ含量、反復、またはこれらの任意の組み合わせを含む制約に対して、さらに最適化される。場合によっては、複数のポリヌクレオチド配列の４０％～６０％が、冗長性のためにエンコードする。場合によっては、複数のポリヌクレオチドを生成することは、複数の合成サイクルを含む。場合によっては、この合成サイクルの数は、内部コーデックを使用してエンコードされないポリヌクレオチド配列を合成するために必要とされる合成サイクルの数と比較して、減少する。場合によっては、合成サイクルの数の減少は、フロー順序に部分的に基づく。場合によっては、合成サイクルの数は、少なくとも３０％減少する。場合によっては、合成サイクルの数は５０％減少する。場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、３００未満である。場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、約１５５である。場合によっては、ポリヌクレオチド配列は、１つまたは複数のＡ、Ｔ、Ｃ、またはＧを含む。場合によっては、複数のポリヌクレオチドを生成することは、塩基ごとの合成を含む。場合によっては、シンセサイザは、複数の特徴を備えている固体支持体を備える。場合によっては、複数の特徴の各々は、固体支持体の１つまたは複数の電極を介して独立してアドレス指定可能である。場合によっては、複数の特徴の各々は、マスキングを介してアドレス指定可能である。場合によっては、マスキングは物理的バリアを含む。場合によっては、マスキングは、複数

の特徴のうちの１つまたは複数で反応性を制御することを含む。場合によっては、反応性を制御することは、複数の特徴のうちの１つまたは複数での脱保護を含む。場合によっては、脱保護は酸生成を含む。場合によっては、脱保護は電気化学的脱保護である。場合によっては、合成サイクルごとに、複数の特徴の２５％より多くがデブロッキングされる。場合によっては、合成サイクルごとに、複数の特徴の少なくとも５０％がデブロッキングされる。場合によっては、複数のポリヌクレオチド配列の各々は、同じ長さを有する。場合によっては、複数のポリヌクレオチド配列の８０％～１００％は、同じ長さを有する。

【００１２】

１つの態様では、本明細書において、１つまたは複数の処理ユニットを備えている、データをＤＮＡに格納するためのシステムが提供され、メモリが１つまたは複数の処理ユニットと通信し、命令がメモリに格納され、１つまたは複数の処理ユニット上で実行されて、システムに、複数のプールを生成することであって、複数のプールの各々がプール記述子、データのペイロードを含むプール・アイテム、および終了記述子を含む、生成することと、プール・アイテムごとのペイロードの第１の１つまたは複数のハッシュを決定することと、エンコーディング方式を適用して、複数のプールを複数のポリヌクレオチドの配列としてエンコードすることとを実行させる。一部の実施形態では、エンコーディング方式は、本明細書に記載されている内部コーデック、外部コーデック、またはその両方を含む。一部の実施形態では、データは、本明細書に記載されている情報のアイテムまたはデジタル情報を含む。一部の実施形態では、データは、１つまたは複数のオブジェクトを含む。一部の実施形態では、１つもしくは複数の処理ユニット、メモリ、またはその両方は、コンピューティング・システムの一部である。場合によっては、コンピューティング・システムは、クラウド・コンピューティング・システムを含む。場合によっては、クラウド・コンピューティング・システムは、プライベート・クラウド、パブリック・クラウド、ハイブリッド・クラウド、マルチクラウド、またはこれらの任意の組み合わせを含む。場合によっては、クラウド・コンピューティング・システムは、サービスとしてのインフラストラクチャ（ＩａａＳ）、サービスとしてのプラットフォーム（ＰａａＳ）、サービスとしてのソフトウェア（ＳａａＳ）、またはこれらの任意の組み合わせを含む。一部の実施形態では、命令がメモリに格納され、１つまたは複数の処理ユニット上で実行されて、システムに、１つまたは複数のオブジェクトの各々の第２の１つまたは複数のハッシュを決定させる。一部の実施形態では、１つまたは複数のオブジェクトは、ファイルまたはファイルに関連付けられたメタデータを含む。一部の実施形態では、プール記述子は、バージョン、プールＩＤ、プール・アイテム記述子のリスト、またはこれらの任意の組み合わせを含む。一部の実施形態では、プールＩＤは一意のＩＤを含む。一部の実施形態では、一意のＩＤは、汎用一意識別子（ＵＵＩＤ：ｕｎｉｖｅｒｓａｌ ｕｎｉｑｕｅ ｉｄｅｎｔｉｆｉｅｒ）またはコンテンツＩＤを含む。一部の実施形態では、プール・アイテム記述子のリストは、オブジェクトのパス、オブジェクトのサイズ、オブジェクト内のプール・アイテムの範囲、プール内のプール・アイテムのオフセット、またはこれらの任意の組み合わせを含む。一部の実施形態では、１つまたは複数のプール・アイテムの各々は、第１の１つまたは複数のハッシュからのプール・アイテムのハッシュをさらに含む。一部の実施形態では、終了プール記述子（ｅｎｄ ｐｏｏｌ ｄｅｓｃｒｉｐｔｏｒ）は、オブジェクト記述子のリストを含む。一部の実施形態では、オブジェクト記述子のリストは、オブジェクトのパス、第１の１つもしくは複数のハッシュからのオブジェクトのハッシュ、またはこれらの組み合わせを含む。一部の実施形態では、複数のプールの各々は、約１ＧＢ～約１ＴＢである。一部の実施形態では、複数のプールは、冗長なプールを含む。一部の実施形態では、第１の１つもしくは複数のハッシュ、第２の１つもしくは複数のハッシュ、またはその両方は、ハッシング・モジュールを使用して決定される。一部の実施形態では、ハッシング・モジュールは、１つまたは複数の処理ユニット上で実行される。一部の実施形態では、第１の１つまたは複数のハッシュは、１つまたは複数のオブジェクトよりも少ないメモリを必要とする。一部の実施形態では、第２の１つまたは複数のハッシュは、１つまたは複数のプール・アイテムよりも少ないメモリを必要とする。一部の

実施形態では、ハッシング・モジュールはハッシュ関数を含む。一部の実施形態では、ハッシュ関数は、ＳＨＡ－２２４、ＳＨＡ－２５６、ＳＨＡ－３８４、ＳＨＡ－５１２、ＳＨＡ－５１２／２２４、またはＳＨＡ－５１２／２５６を含む。一部の実施形態では、命令は、システムに１つまたは複数のインデックス・プールをさらに生成させる。一部の実施形態では、１つまたは複数のインデックス・プールは、インデックス・プール記述子およびオブジェクト・インデックス作成のリストを含む。一部の実施形態では、インデックス・プール記述子は、バージョン、プールＩＤ、プールのサイズ、およびタイムスタンプを含む。一部の実施形態では、プールＩＤは、一意のＩＤを含む。一部の実施形態では、一意のＩＤは、ＵＵＩＤまたはコンテンツＩＤを含む。一部の実施形態では、オブジェクト・インデックス作成のリストは、オブジェクトのパス、オブジェクトのハッシュ、オブジェクト・フラグメントのリスト、オブジェクト・メタデータのリスト、またはこれらの任意の組み合わせを含む。一部の実施形態では、オブジェクト・フラグメントのリストは、フラグメントを含むプールのプールＩＤ、フラグメントの範囲、またはこれらの組み合わせを含む。一部の実施形態では、オブジェクト・メタデータのリストは、メタデータの種類、メタデータ・ペイロード、またはこれらの組み合わせを含む。一部の実施形態では、メタデータの種類は、オブジェクトに添付されたキーワードのリスト、サムネイル画像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはこれらの任意の組み合わせを含む。一部の実施形態では、１つまたは複数のインデックス・プールの各々は、約１ＧＢ～約１ＴＢである。一部の実施形態では、命令がメモリに格納され、１つまたは複数の処理ユニット上で実行されて、システムに、ＤＮＡに格納されたデータを取り出させる。一部の実施形態では、命令は、デコーディング方式を適用して、複数のプールの各々における複数のポリヌクレオチドの配列をデコードすることと、第１の１つまたは複数のハッシュを使用して、少なくとも各プール・アイテムのペイロードを検証することとを含む。

【００１３】

１つの態様では、本明細書において、１つまたは複数のコンパートメントを備えている、情報をＤＮＡに格納するためのデバイスが提供され、各コンパートメントは、（ａ）複数のポリヌクレオチドを含むライブラリであって、１つまたは複数のオブジェクトに対応する情報を含むプールをエンコードする、ライブラリと、（ｂ）複数のポリヌクレオチドを格納するための媒体とを備える。一部の実施形態では、情報は、本明細書に記載されている情報のアイテムまたはデジタル情報を含む。一部の実施形態では、情報は複数のシンボルを含む。一部の実施形態では、１つまたは複数のコンパートメントは通信する。一部の実施形態では、１つまたは複数のコンパートメントは通信しない。一部の実施形態では、媒体は、固体、液体、気体、またはこれらの任意の組み合わせを含む。一部の実施形態では、媒体は、ＤＮＡ内のリン酸基に対する塩の陽イオンのモル比が２０：１未満の塩溶液を含む。一部の実施形態では、塩溶液が乾燥されて、乾燥生成物を作り出す。一部の実施形態では、このデバイスは、表面を備えている固体支持体をさらに備える。一部の実施形態では、このデバイスは、表面上に位置する複数の構造体をさらに備え、複数のポリヌクレオチドが複数の構造体から伸ばされる。一部の実施形態では、１つまたは複数のオブジェクトは、ファイルまたはファイルに関連付けられたメタデータを含む。一部の実施形態では、プールは、プール記述子、１つまたは複数のプール・アイテム、および終了プール記述子を含む。一部の実施形態では、プール記述子は、バージョン、プールＩＤ、プール・アイテム記述子のリスト、またはこれらの任意の組み合わせを含む。一部の実施形態では、プールＩＤは一意のＩＤを含む。一部の実施形態では、一意のＩＤは、汎用一意識別子（ＵＵＩＤ）またはコンテンツＩＤを含む。一部の実施形態では、プール・アイテム記述子のリストは、オブジェクトのパス、オブジェクトのサイズ、オブジェクト内のプール・アイテムの範囲、プール内のプール・アイテムのオフセット、またはこれらの任意の組み合わせを含む。一部の実施形態では、１つまたは複数のプール・アイテムの各々は、データ・ペイロード、プール・アイテムのハッシュ、またはこれらの組み合わせを含む。一部の実施形態では、終了プール記述子は、オブジェクト記述子のリストを含む。一部の

実施形態では、オブジェクト記述子のリストは、オブジェクトのパス、オブジェクトのハッシュ、またはこれらの組み合わせを含む。一部の実施形態では、プールは、約１ＧＢ～約１ＴＢのデジタル情報を含む。一部の実施形態では、このデバイスは、１つまたは複数の第２のコンパートメントをさらに備え、１つまたは複数の第２のコンパートメントの各々は、インデックス・プールをエンコードする第２のライブラリを備える。一部の実施形態では、１つまたは複数のインデックス・プールは、インデックス・プール記述子およびオブジェクト・インデックス作成のリストを含む。一部の実施形態では、インデックス・プール記述子は、バージョン、プールＩＤ、プールのサイズ、およびタイムスタンプを含む。一部の実施形態では、プールＩＤは、一意のＩＤを含む。一部の実施形態では、一意のＩＤは、ＵＵＩＤまたはコンテンツＩＤを含む。一部の実施形態では、オブジェクト・インデックス作成のリストは、オブジェクトのパス、オブジェクトのハッシュ、オブジェクト・フラグメントのリスト、オブジェクト・メタデータのリスト、またはこれらの任意の組み合わせを含む。一部の実施形態では、オブジェクト・フラグメントのリストは、フラグメントを含むプールのプールＩＤ、フラグメントの範囲、またはこれらの組み合わせを含む。一部の実施形態では、オブジェクト・メタデータのリストは、メタデータの種類、メタデータ・ペイロード、またはこれらの組み合わせを含む。一部の実施形態では、メタデータの種類は、オブジェクトに添付されたキーワードのリスト、サムネイル画像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはこれらの任意の組み合わせを含む。一部の実施形態では、１つまたは複数のインデックス・プールの各々は、約１ＧＢ～約１ＴＢである。

【００１４】

別の態様では、本明細書において、データを複数のポリヌクレオチドに格納するための方法が提供され、この方法は、複数のプールを生成することであって、複数のプールの各々がプール記述子、データのペイロードを含むプール・アイテム、および終了記述子を含む、生成することと、プール・アイテムごとのペイロードの第１の１つまたは複数のハッシュを決定することと、エンコーディング方式を適用して、複数のプールを複数のヌクレオチドの配列としてエンコードすることとを含む。一部の実施形態では、エンコーディング方式は、本明細書に記載されている内部コーデック、外部コーデック、またはその両方を含む。一部の実施形態では、データは、本明細書に記載されている情報のアイテムまたはデジタル情報を含む。場合によっては、データは複数のシンボルを含む。一部の実施形態では、データは、１つまたは複数のオブジェクトを含む。一部の実施形態では、この方法は、１つまたは複数のオブジェクトの各々の第２の１つまたは複数のハッシュを決定することをさらに含む。一部の実施形態では、複数のポリヌクレオチドを格納することをさらに含む。一部の実施形態では、複数のプールのうちの各プールに対応する複数のポリヌクレオチドのうちのポリヌクレオチドは、データ・ストレージ・システムの別々のコンテナに格納される。一部の実施形態では、複数のポリヌクレオチドを生成することをさらに含む。一部の実施形態では、複数のポリヌクレオチドを生成することは、デオキシリボ核酸（ＤＮＡ：ｄｅｏｘｙｒｉｂｏｎｕｃｌｅｉｃ ａｃｉｄ）のホスホラミダイトベースの合成を含む。一部の実施形態では、ホスホラミダイトベースの合成のための試薬は、ヌクレオシド・ホスホラミダイト、酸化剤、活性化剤、またはデブロッカーを含み、あるいは溶媒はアセトニトリルを含む。一部の実施形態では、複数のポリヌクレオチドを生成することは、酵素的ＤＮＡ合成を含む。一部の実施形態では、酵素的ＤＮＡ合成のための試薬は、末端デオキシヌクレオチジル・トランスフェラーゼ（ＴｄＴ：ｔｅｒｍｉｎａｌ ｄｅｏｘｙｎｕｃｌｅｏｔｉｄｙｌ ｔｒａｎｓｆｅｒａｓｅ）またはデブロッカーを含み、あるいは溶媒は水を含む。一部の実施形態では、１つまたは複数のオブジェクトは、ファイルまたはファイルに関連付けられたメタデータを含む。一部の実施形態では、プール記述子は、バージョン、プールＩＤ、プール・アイテム記述子のリスト、またはこれらの任意の組み合わせを含む。一部の実施形態では、プールＩＤは一意のＩＤを含む。一部の実施形態では、一意のＩＤは、汎用一意識別子（ＵＵＩＤ）またはコンテンツＩＤを含む。一部の実施形態では、プール・アイテム記述子のリストは、オブジェクトのパス、オ

ブジェクトのサイズ、オブジェクト内のプール・アイテムの範囲、プール内のプール・アイテムのオフセット、またはこれらの任意の組み合わせを含む。一部の実施形態では、１つまたは複数のプール・アイテムの各々は、第１の１つまたは複数のハッシュからのプール・アイテムのハッシュをさらに含む。一部の実施形態では、終了プール記述子は、オブジェクト記述子のリストを含む。一部の実施形態では、オブジェクト記述子のリストは、オブジェクトのパス、第１の１つもしくは複数のハッシュからのオブジェクトのハッシュ、またはこれらの組み合わせを含む。一部の実施形態では、複数のプールの各々は、約１ＧＢ～約１ＴＢである。一部の実施形態では、複数のプールは、冗長なプールを含む。一部の実施形態では、第１の１つまたは複数のハッシュ、第２の１つまたは複数のハッシュ、またはその両方は、ハッシング・モジュールを使用して決定される。一部の実施形態では、第２の１つまたは複数のハッシュは、１つまたは複数のオブジェクトよりも少ないメモリを必要とする。一部の実施形態では、第１の１つまたは複数のハッシュは、１つまたは複数のプール・アイテムよりも少ないメモリを必要とする。一部の実施形態では、ハッシング・モジュールはハッシュ関数を含む。一部の実施形態では、ハッシュ関数は、ＳＨＡ－２２４、ＳＨＡ－２５６、ＳＨＡ－３８４、ＳＨＡ－５１２、ＳＨＡ－５１２／２２４、またはＳＨＡ－５１２／２５６を含む。一部の実施形態では、１つまたは複数のインデックス・プールを作成することをさらに含む。一部の実施形態では、１つまたは複数のインデックス・プールは、インデックス・プール記述子およびオブジェクト・インデックス作成のリストを含む。一部の実施形態では、インデックス・プール記述子は、バージョン、プールＩＤ、プールのサイズ、およびタイムスタンプを含む。一部の実施形態では、プールＩＤは、一意のＩＤを含む。一部の実施形態では、一意のＩＤは、ＵＵＩＤまたはコンテンツＩＤを含む。一部の実施形態では、オブジェクト・インデックス作成のリストは、オブジェクトのパス、オブジェクトのハッシュ、オブジェクト・フラグメントのリスト、オブジェクト・メタデータのリスト、またはこれらの任意の組み合わせを含む。一部の実施形態では、オブジェクト・フラグメントのリストは、フラグメントを含むプールのプールＩＤ、フラグメントの範囲、またはこれらの組み合わせを含む。一部の実施形態では、オブジェクト・メタデータのリストは、メタデータの種類、メタデータ・ペイロード、またはこれらの組み合わせを含む。一部の実施形態では、メタデータの種類は、オブジェクトに添付されたキーワードのリスト、サムネイル画像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはこれらの任意の組み合わせを含む。一部の実施形態では、インデックス・プールのうちの１つまたは複数の各々は、約１ＧＢ～約１ＴＢである。

【００１５】

別の態様では、本明細書において、複数のポリヌクレオチドに格納されたデータを取り出すための方法が提供され、この方法は、複数のポリヌクレオチドの配列を決定することであって、複数のポリヌクレオチドが複数のプール内にある、決定することと、デコーディング方式を適用して、複数のプールの各々における複数のポリヌクレオチドの配列をデコードすることであって、各プールが、プール記述子、データのペイロードを含むプール・アイテム、および終了記述子を含む、デコードすることと、第１の１つまたは複数のハッシュを使用して少なくとも各プール・アイテムのペイロードを検証することとを含む。一部の実施形態では、デコーディング方式は、本明細書に記載されている内部コーデック、外部コーデック、またはその両方を含む。一部の実施形態では、データは、本明細書に記載されている情報のアイテムまたはデジタル情報を含む。一部の実施形態では、データは、１つまたは複数のオブジェクトを含む。一部の実施形態では、１つまたは複数のオブジェクトは、ファイルまたはファイルに関連付けられたメタデータを含む。一部の実施形態では、この方法は、第２の１つまたは複数のハッシュを使用して１つまたは複数のオブジェクトを検証することをさらに含む。一部の実施形態では、少なくともペイロードを検証することは、ハッシュ関数を使用して第１の１つまたは複数のハッシュを検証することを含む。一部の実施形態では、この方法は、データを取り出すために、各プール・アイテムからのペイロードを結合することをさらに含む。一部の実施形態では、この方法は、デ

ータをメモリに格納することをさらに含む。一部の実施形態では、複数のプールの各々は、約１ＧＢ～約１ＴＢである。一部の実施形態では、１つまたは複数のオブジェクトを検証することは、ハッシュ関数を使用して第２の１つまたは複数のハッシュを検証することを含む。一部の実施形態では、ハッシュ関数は、ＳＨＡ－２２４、ＳＨＡ－２５６、ＳＨＡ－３８４、ＳＨＡ－５１２、ＳＨＡ－５１２／２２４、またはＳＨＡ－５１２／２５６を含む。一部の実施形態では、配列を決定することは、複数のポリヌクレオチドをシーケンシングすることを含む。一部の実施形態では、シーケンシングは、次世代シーケンシング、並列シーケンシング、単一分子リアルタイム・シーケンシング、ナノポア・シーケンシング、合成によるシーケンシング、サンガー・シーケンシング、またはこれらの任意の組み合わせを含む。一部の実施形態では、この方法は、１つまたは複数のオブジェクトを含む複数のプールを決定するために、１つまたは複数のインデックス・プールのうちの１つのインデックス・プールにアクセスすることをさらに含む。一部の実施形態では、インデックス・プールは、インデックス・プール記述子およびオブジェクト・インデックス作成のリストを含む。一部の実施形態では、インデックス・プール記述子は、バージョン、プールＩＤ、プールのサイズ、およびタイムスタンプを含む。一部の実施形態では、プールＩＤは、一意のＩＤを含む。一部の実施形態では、一意のＩＤは、ＵＵＩＤまたはコンテンツＩＤを含む。一部の実施形態では、オブジェクト・インデックス作成のリストは、オブジェクトのパス、オブジェクトのハッシュ、オブジェクト・フラグメントのリスト、オブジェクト・メタデータのリスト、またはこれらの任意の組み合わせを含む。一部の実施形態では、オブジェクト・フラグメントのリストは、フラグメントを含むプールのプールＩＤ、フラグメントの範囲、またはこれらの組み合わせを含む。一部の実施形態では、オブジェクト・メタデータのリストは、メタデータの種類、メタデータ・ペイロード、またはこれらの組み合わせを含む。一部の実施形態では、メタデータの種類は、オブジェクトに添付されたキーワードのリスト、サムネイル画像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはこれらの任意の組み合わせを含む。一部の実施形態では、インデックス・プールのうちの１つまたは複数の各々は、約１ＧＢ～約１ＴＢである。

【００１６】

本主題の特徴および優位性のより良い理解は、実施形態例を示す以下の詳細な説明および添付の図面を参照することによって得られる。

【図面の簡単な説明】

【００１７】

【図１】一部の実施形態による、低レベル・コーデックのエンコーディング方式の非限定的な例を示す図である。

【図２】一部の実施形態による、低レベル・コーデックのデコーディング方式の非限定的な例を示す図である。

【図３】一部の実施形態による、外部コーデックを含むエンコーディング方式の非限定的な例を示す図である。

【図４】一部の実施形態による、データのレーンをシャッフルすることを含むエンコーディング方式の非限定的な例を示す図である。

【図５】一部の実施形態による、第１の内部コーデックを含むエンコーディング方式の非限定的な例を示す図である。

【図６】一部の実施形態による、第２の内部コーデックを含むエンコーディング方式の非限定的な例を示す図である。

【図７】一部の実施形態による、内部コーデックおよび外部コーデックを含むデコーディング方式の非限定的な例を示す図である。

【図８】一部の実施形態による、デコードするための貪欲アルゴリズムの非限定的な例を示す図である。

【図９】一部の実施形態による、デコードするための最尤（ＭＬ）アルゴリズムの非限定的な例を示す図である。

【図１０】コンピューティング・デバイス、この場合、１つまたは複数のプロセッサ、メモリ、ストレージ、およびネットワーク・インターフェイスを備えるデバイスの、非限定的な例を示す図である。
【図１１Ａ】一部の実施形態による、ポリヌクレオチド合成表面の製造のための「リフトオフ」プロセスの非限定的な例を示す図である。
【図１１Ｂ】一部の実施形態による、ポリヌクレオチド合成表面の製造のためのウェット・エッチング・プロセスの非限定的な例を示す図である。このプロセスは、ドライ・エッチング・プロセスに適応されてもよい。
【図１２】一部の実施形態による、高レベル・コーデックのエンコーディング方式の非限定的な例を示す図である。
【図１３】一部の実施形態による、高レベル・コーデックのデコーディング方式の非限定的な例を示す図である。
【図１４】一部の実施形態による、デジタル情報ストレージの非限定的な例を示す図である。
【図１５】一部の実施形態による、ハッシュを生成する非限定的な例を示す図である。
【図１６】一部の実施形態による、複数のポリヌクレオチドを合成、格納、およびシーケンシングするためのシステムの非限定的な例を示す図である。
【図１７Ａ】一部の実施形態による、複数のポリヌクレオチドを格納するための構造体またはコンパートメントの非限定的な例を示す図である。図１７Ａは、実質的に管状である構造体を示している。
【図１７Ｂ】一部の実施形態による、複数のポリヌクレオチドを格納するための構造体またはコンパートメントの非限定的な例を示す図である。図１７Ｂは、一緒に面一溶接された（ｆｌｕｓｈ－ｗｅｌｄｅｄ）キャップおよび本体を備えている構造体を示している。
【図１７Ｃ】一部の実施形態による、複数のポリヌクレオチドを格納するための構造体またはコンパートメントの非限定的な例を示す図である。図１７Ｃは、取り外し可能なスクリューキャップを備えている構造体を示している。
【図１７Ｄ】一部の実施形態による、複数のポリヌクレオチドを格納するための構造体またはコンパートメントの非限定的な例を示す図である。図１７Ｄは、隔膜を備えている構造を示している。
【図１７Ｅ】一部の実施形態による、複数のポリヌクレオチドを格納するための構造体またはコンパートメントの非限定的な例を示す図である。図１７Ｅは、一方の半分が他方に挿入されたときにシールを形成する、２つの丸みを帯びた錠剤状の半分を備えている構造体を示している。
【図１７Ｆ】一部の実施形態による、複数のポリヌクレオチドを格納するための構造体またはコンパートメントの非限定的な例を示す図である。図１７Ｆは、密閉可能な蓋を有する実質的に平らなディスク・コンテナを備えている構造体を示している。
【図１７Ｇ】一部の実施形態による、複数のポリヌクレオチドを格納するための構造体またはコンパートメントの非限定的な例を示す図である。図１７Ｇは、任意選択的に取り付けられた蓋を有するボックスを備えている構造体を示している。
【発明を実施するための形態】

【００１８】
本明細書では、デジタル情報を核酸に格納するための方法およびシステムが提供される。多くのストレージ媒体と同様に、合成ＤＮＡは、欠失、挿入、突然変異、断片化などの固有のエラーを有することがあり、これらのエラーは、完全なオリゴヌクレオチドの消去につながる可能性がある。経年劣化またはサンプル処理に起因して、一部のオリゴヌクレオチドの損失があることもある。コンピュータ・サイエンスおよび電気通信において使用される標準的な技術は、消失および／または突然変異のみに対処し、オリゴヌクレオチド・プールの特定の挙動には対処しない。例えば、オリゴヌクレオチド・プールのシーケンシングは、オリゴをランダムな順序で提供し、一方、ハード・ドライブのような標準的なストレージ媒体は、書き込み中に作成される既知の予想される順序でデータのストリーム

を提供する。さらに、デジタル情報を格納するための多くのコーデックは、デジタル情報を核酸にエンコードすることに重点を置いているが、構造化されたファイルのリストを格納して取り出す方法を提供しないことがある。そのため、本明細書では、複数の「オブジェクト」を受け取り、それらを１つまたは複数のプールとして効率的に格納するか、または１つもしくは複数のプールから取り出すことができるコーデックおよび実装が提供される。オブジェクトは、ファイルまたはファイルに関連付けられたメタデータを含んでよい。そのようなコーデックの実装は、例えば、本明細書に記載されたエラー訂正コードなどの、ただしこれに限定されないエラー訂正コードを含む、核酸および／または外部コーデックにおけるデジタル情報をエンコードするための低レベル・コーデックと組み合わせられ得る。

【００１９】

場合によっては、この方法は、データを複数のポリヌクレオチド配列にエンコードする。データは、複数のシンボルとして表されてよい。場合によっては、方法は、データを複数のフレームに分割するステップと、外部コーデックを複数のフレームのうちの各フレームに適用するステップと、各フレームを複数のレーンに分割するステップと、レーン・インデックスに少なくとも部分的に基づいて各レーンをシャッフルするステップと、内部コーデック（例えば、エンコーディング方式）を適用して、複数のポリヌクレオチド配列のうち１つのポリヌクレオチド配列内の各レーンをエンコードするステップと、のうちの１つまたは複数のステップを含む。場合によっては、複数のフレームのうちの各フレームはフレーム・インデックスを含む。場合によっては、外部コーデックはエラー訂正方式を含む。場合によっては、複数のレーンのうちの各レーンはレーン・インデックスを含む。

【００２０】

場合によっては、方法は、複数のポリヌクレオチド配列をデコードして、データを含む出力を生成する。データは、複数のシンボルとして表されてよい。場合によっては、方法は、複数のポリヌクレオチド配列を決定するステップと、内部コーデック（例えば、デコーディング方式）を複数のポリヌクレオチド配列に適用するステップと、データのレーンの各々におけるレーン・インデックスおよびフレーム・インデックスに基づいて、データのレーンをフレームに配置するステップと、外部コーデックをフレームに適用するステップと、のうちの１つまたは複数を含む。場合によっては、内部コーデックは、複数のポリヌクレオチド配列の各々を、複数のシンボルを含むレーンに変換する。場合によっては、内部コーデックは、貪欲アルゴリズムおよび最尤（ＭＬ）アルゴリズムを含んでいる混合デコーディング・アルゴリズムを含む。場合によっては、外部コーデックはエラー訂正方式を含む。場合によっては、外部コーデックからのフレームがマージされて、データを含む出力を生成する。

【００２１】

場合によっては、システムは、データを複数のポリヌクレオチド配列にエンコードする。場合によっては、システムは、メモリおよびメモリに動作可能に結合された処理デバイスのうちの１つまたは複数を備えている装置を備える。場合によっては、処理デバイスは、データを複数のフレームに分割するステップと、外部コーデックを複数のフレームのうちの各フレームに適用するステップと、各フレームを複数のレーンに分割するステップと、レーン・インデックスに少なくとも部分的に基づいて各レーンをシャッフルするステップと、内部コーデックを適用して、ポリヌクレオチド配列内の各レーンをエンコードするステップと、を含むステップのうちの１つまたは複数を実行するように構成される。場合によっては、複数のフレームのうちの各フレームはフレーム・インデックスを含む。場合によっては、複数のレーンのうちの各レーンはレーン・インデックスを含む。場合によっては、外部コーデックはエラー訂正方式を含む。場合によっては、ポリヌクレオチド配列にエラーが存在する場合にデータがデコードされ得るように、内部コーデックは冗長性を追加する。場合によっては、内部コーデックはエンコーディング方式を含む。

【００２２】

場合によっては、システムは、複数のポリヌクレオチド配列をデコードして、データを

含む出力を生成する。場合によっては、システムは、メモリと、複数のポリヌクレオチド配列を決定するように構成されたシーケンシング・デバイスと、メモリに動作可能に結合された処理デバイスと、のうちの１つまたは複数を備えている装置を備える。場合によっては、処理デバイスは、内部コーデックを複数のポリヌクレオチド配列に適用するステップと、データのレーンの各々におけるレーン・インデックスおよびフレーム・インデックスに基づいて、データのレーンをフレームに配置するステップと、外部コーデックをフレームに適用するステップと、を含むステップのうちの１つまたは複数を実行するように構成される。場合によっては、内部コーデックは、配列の各々を複数のシンボルを含むレーンに変換する。場合によっては、内部コーデックは、貪欲アルゴリズムおよび最尤（ＭＬ）アルゴリズムを含んでいる混合デコーディング・アルゴリズムを含む。場合によっては、外部コーデックはエラー訂正方式を含む。場合によっては、外部コーデックからのフレームがマージされて、データを含む出力を生成する。場合によっては、内部コーデックはデコーディング方式を含む。

【００２３】

場合によっては、方法は、データをポリヌクレオチド配列にエンコードする。場合によっては、方法は、（ａ）コードブックを含む内部コーデックを生成するステップであって、コードブックが１つまたは複数の制約に対して最適化される、ステップと、（ｂ）内部コーデックを適用して、データを複数のポリヌクレオチド配列としてエンコードするステップと、のうちの１つまたは複数を含む。場合によっては、この方法は、複数のポリヌクレオチド配列を含んでいる複数のポリヌクレオチドを生成することをさらに含む。

【００２４】

場合によっては、本明細書では、データをエンコードするためのハイブリッド有機インシリコ・プラットフォームが提供される。プラットフォームは、少なくとも１つのプロセッサ、および動作を実行するために少なくとも１つのプロセッサによって実行可能な命令を備えるコンピューティング・システムと、複数のポリヌクレオチド配列を含む複数のポリヌクレオチドを生成するためのシンセサイザと、のうちの１つまたは複数を備える。場合によっては、動作は、コードブックを含む内部コーデックを生成することであって、コードブックが１つまたは複数の制約に対して最適化される、生成することと、内部コーデックを適用して、データを複数のポリヌクレオチド配列としてエンコードすることと、のうちの１つまたは複数を含む。

【００２５】

場合によっては、システムは情報をＤＮＡに格納する。場合によっては、システムは、１つまたは複数の処理ユニットと、１つまたは複数の処理ユニットと通信するメモリと、メモリに格納され、１つまたは複数の処理ユニット上で実行される命令と、のうちの任意の１つまたは組み合わせを備える。場合によっては、命令は、システムに、１つまたは複数のオブジェクトのデジタル情報を複数のプールに分割することと、複数のプールの各々においてプール記述子、１つまたは複数のプール・アイテム、および終了プール記述子を生成することと、１つまたは複数のプール・アイテムの各々のデータ・ペイロードの第１の１つまたは複数のハッシュ、および１つまたは複数のオブジェクトの各々の第２の１つまたは複数のハッシュを決定することと、エンコーディング方式を適用して、デジタル情報を複数のポリヌクレオチドとして複数のプールにエンコードすることと、のうちの任意の１つまたは組み合わせを実行させる。

【００２６】

場合によっては、デバイスは、情報をＤＮＡに格納するためのものである。場合によっては、デバイスは１つまたは複数のコンパートメントを備える。場合によっては、各コンパートメントは、複数のポリヌクレオチドを含むライブラリと、複数のポリヌクレオチドを格納するための媒体と、のうちの任意の１つまたは組み合わせを備える。場合によっては、ライブラリは、１つまたは複数のオブジェクトに対応する情報を含むプールをエンコードする。

【００２７】

場合によっては、この方法は、データを複数のポリヌクレオチドに格納する。場合によっては、この方法は、複数のプールを生成することと、プール・アイテムごとにペイロードの第１の１つまたは複数のハッシュを決定することと、エンコーディング方式を適用して、複数のプールを複数のヌクレオチドの配列としてエンコードすることと、のうちの任意の１つまたは組み合わせを含む。場合によっては、複数のプールの各々は、プール記述子、データのペイロードを含むプール・アイテム、および終了記述子を含む。

【００２８】

場合によっては、この方法は、複数のポリヌクレオチドに格納されたデータを取り出す。場合によっては、この方法は、複数のポリヌクレオチドの配列を決定することと、デコーディング方式を適用して、複数のプールの各々における複数のポリヌクレオチドの配列をデコードすることと、第１の１つまたは複数のハッシュを使用して、少なくとも各プール・アイテムのペイロードを検証することと、のうちの任意の１つまたは組み合わせを含む。場合によっては、複数のポリヌクレオチドは複数のプール内にある。場合によっては、各プールは、プール記述子、データのペイロードを含むプール・アイテム、および終了記述子を含む。

【００２９】

本明細書では、ポリヌクレオチドの合成を最適化するための方法およびシステムがさらに提供される。場合によっては、合成は、本明細書において提供される合成最適化コーデック（ｓｙｎｔｈｅｓｉｓ　ｏｐｔｉｍｉｚｅｄ　ｃｏｄｅｃ）などの合成最適化コーデックを使用して最適化される。ポリヌクレオチドは、本明細書において提供されるデバイスに従って合成されてよい。電子合成は、通常、特定の部位（例えば、ポリヌクレオチド合成のための表面上の特徴または遺伝子座）をデブロッキングすること、および特定の塩基（例えば、核酸モノマー）を流すことを含み、これらが塩基ごとに繰り返される。これは、特に数百万のポリヌクレオチドを一緒に合成する場合、ポリヌクレオチドのセクションが合成順序において一致する可能性が非常に低いため、特定の塩基順序を伴わないポリヌクレオチドが、層ごとに４サイクル（例えば、Ａ、Ｔ、Ｃ、Ｇ）を必要とする可能性があるということを意味する。例えば、表面が、塩基付加から特定の部位を保護するためにマスクされ（各部位は、固有のポリヌクレオチドを含み、独立してアドレス指定可能である）、塩基が保護されていない部位に結合され、次に、異なる部位での塩基の結合を可能にするようにマスクが変更される。層は、通常、少なくとも１つの塩基による各ポリヌクレオチドの伸長を含む。例えば、ポリヌクレオチドの長さがＭ塩基である場合、合成は、ポリヌクレオチドへの単一核酸の付加ごとに４サイクルを仮定して、４×Ｍサイクルを必要とする可能性がある。この手法は、より多くの時間、より多くの試薬、またはその両方を必要とすることがあるため、よりコストがかかる可能性がある。この手法は、各サイクルが、より高いエラー率をもたらす可能性がある酸化ステップおよびデブロッキング・ステップを必要とするため、ＤＮＡ損傷の可能性を高めることもある。

【００３０】

合成を最適化するための方法、システム、およびプラットフォームは、塩基合成の特定の順序に従ってポリヌクレオチドを生成するように最適化された内部コーデックを含むことができる。これによって、４×Ｍサイクル未満でのポリヌクレオチドの合成を可能にすることができ、Ｍはポリヌクレオチドの塩基数である。この手法は、外部コーデックまたはエラー訂正コード（ＥＣＣ：ｅｒｒｏｒ　ｃｏｒｒｅｃｔｉｏｎ　ｃｏｄｅ）などを使用して、エラー訂正のための冗長性を提供することもできる。この手法は、合成されるポリヌクレオチドが同じ量のデータをエンコードする場合、最適化されていない（例えば、４×Ｍサイクルを必要とする）合成手法と比較して、ポリヌクレオチドの合成を加速することもできる。場合によっては、塩基の混合物（例えば、２つまたは３つ）が、単一のサイクルで表面全体に流される。場合によっては、本明細書で提供される１つまたは複数のコードブックと共に使用するために、合成法が構成される。本明細書に記載されるような最適化されていない合成手法は、通常、塩基順序を伴わないポリヌクレオチドの合成のことを指してよい。場合によっては、最適化されていない合成手法と比較して、合成速度が

約１．５倍、２倍、２．５倍、３倍、３．５倍、または４倍加速される。場合によっては、最適化されていない合成手法と比較して、合成速度が最大で２倍、２．５倍、３倍、３．５倍、または４倍まで加速される。場合によっては、最適化されていない合成手法と比較して、合成速度が最大で約１．５倍、２倍、２．５倍、３倍、または３．５倍加速される。場合によっては、より少ない酸化ステップが必要とされるため、ＤＮＡの品質を改善しながら合成速度が加速される。場合によっては、エラーを減らしながら合成速度が加速される。

【００３１】

場合によっては、本明細書において提供される方法は、データをエンコードする。データは、デジタル情報または情報のアイテムであってよい。データは、１つまたは複数のシンボルとして表されてよい。一部の例では、１つまたは複数のシンボルは、バイナリ・データなどの数値を含む。場合によっては、シンボルのセットとして表されたデータが、コーデックを使用してシンボルの異なるセットとしてエンコードされる。場合によっては、そのようなコーデックは、内部コーデックと呼ばれる。場合によっては、シンボルの異なるセットは、ポリヌクレオチド配列などのシンボルの配列を含む。

【００３２】

本明細書に記載された方法は、内部コーデックの使用または生成を含んでよい。場合によっては、この方法は、コードブックを含んでいる内部コーデックを生成することを含む。場合によっては、コードブックは、データの集合（例えば、核酸にエンコードされたデジタル情報）のコンテンツ、構造、およびレイアウトを含む。場合によっては、内部コーデックは２つ以上のコードブックを含む。場合によっては、２つ以上のコードブックの各々は、ポリヌクレオチドの合成中に層をエンコードする。場合によっては、コードブックは、１つまたは複数の制約に対して最適化される。場合によっては、１つまたは複数の制約は、核酸合成、後処理、ストレージ、シーケンシング、またはこれらの任意の組み合わせに関連する。

【００３３】

場合によっては、コードブックは、塩基順序を使用して生成される。場合によっては、コードブックは、１つまたは複数の塩基転移に対して最適化される。場合によっては、塩基順序は、１つまたは複数の塩基転移を生成する。そのような１つまたは複数の塩基転移は、特定の塩基転移または既定の塩基転移と呼ばれてよい。場合によっては、２つ以上のコードブックの各々は、異なる塩基順序を含む。場合によっては、２つ以上のコードブックの各々は、異なる１つまたは複数の塩基転移を含む。場合によっては、コードブックは、特定の層、サイクル・インデックス、履歴、またはこれらの任意の組み合わせでの特定の塩基転移に対して最適化される。一部の例では、履歴は、前の層、前の１つまたは複数の層をエンコードする１つまたは複数のコードブック、１つまたは複数の前の層のサイクル・インデックス、またはこれらの任意の組み合わせのうちの１つまたは複数を含む。一部の例では、この方法は、内部コーデックを適用して、データを複数のポリヌクレオチド配列としてエンコードすることを含む。

【００３４】

本明細書において提供される方法は、プラットフォーム上で実行されてよい。場合によっては、プラットフォームは、ハイブリッド有機インシリコ・プラットフォームを含む。場合によっては、プラットフォームはデータ（例えば、バイナリ・データ）をエンコードする。場合によっては、プラットフォームは、少なくとも１つのプロセッサと、動作を実行するために少なくとも１つのプロセッサによって実行可能な命令とを備えているコンピューティング・システムを備える。場合によっては、これらの動作は、コードブックを含んでいる内部コーデックを生成することを含む。場合によっては、コードブックは、塩基順序を使用して生成される。場合によっては、塩基順序は、１つまたは複数の塩基転移を含むコードワードを生成する。場合によっては、これらの動作は、内部コーデックを適用して、データを複数のポリヌクレオチド配列としてエンコードすることを含む。場合によっては、プラットフォームはシンセサイザを備える。場合によっては、プラットフォーム

は、複数のポリヌクレオチド配列を含む複数のポリヌクレオチドを生成するためのシンセサイザを備える。場合によっては、シンセサイザは、合成、ライゲーション、アセンブリ、またはこれらの任意の組み合わせによって、複数のポリヌクレオチド配列を生成する。場合によっては、プラットフォームは、従来の磁気ストレージ・デバイスまたはテープ・ストレージ・デバイスなどの、１つまたは複数の追加のシステムに統合される。

【００３５】

核酸に基づく情報格納

【００３６】

本明細書では、核酸に基づく情報（データ）格納のためのデバイス、構成、システム、および方法が提供される。ＤＮＡ分子などの生体分子は、従来のバイナリ情報符号化とは対照的に、長期にわたる安定性および強化された情報符号化の能力に部分的に起因して、情報格納に適したホストを提供する。第１のステップでは、第１の複数のシンボルを含むデータ、例えば、情報のアイテム（すなわち、コンピュータによって処理するためのバイナリ・コードでのデジタル情報）をエンコードするデジタル配列が受信される。暗号化方式が適用されて、第１の複数のシンボルを第２の複数のシンボルに変換する。第２の複数のシンボルは、核酸配列を含むことができる。例えば、暗号化方式が適用されて、デジタル配列をバイナリ・コードからポリヌクレオチド配列に変換する。核酸伸長のための表面材料、核酸伸長のための遺伝子座の設計（アレンジメント・スポット（ａｒｒａｎｇｅｍｅｎｔ　ｓｐｏｔｓ）としても知られている）、および／または核酸合成のための試薬が選択される。構造体の表面は、核酸合成のために準備されている。その後、デノボ・ポリヌクレオチド合成が実行される。合成されたポリヌクレオチドは格納され、全体的または部分的に、その後の放出に利用可能である。放出されると、ポリヌクレオチドは、全体的または部分的にシーケンシングされ、核酸配列をデジタル配列に再び変換するための復号を受ける。次に、デジタル配列が組み立てられて、元の情報のアイテムのアライメント・エンコーディング（ａｌｉｇｎｍｅｎｔ　ｅｎｃｏｄｉｎｇ）を取得する。

【００３７】

情報のアイテム

【００３８】

任意選択的に、本明細書において開示されたデータ格納プロセスの初期ステップは、１つまたは複数の情報のアイテムを含むデータを初期コードの形態で取得または受信することを含む。情報のアイテムは、テキスト情報、音声情報、および視覚情報を含むが、これらに限定されない。情報のアイテムのソースの例としては、書籍、定期刊行物、電子データベース、医療記録、手紙、フォーム、音声録音、動物の録音、生物学的プロファイル、放送、映画、短いビデオ、電子メール、経理の電話ログ、インターネット活動ログ、図面、絵画、印刷物、写真、ピクセル化されたグラフィックス、およびソフトウェア・コードが挙げられるが、これらに限定されない。情報のアイテムの生物学的プロファイルのソースの例としては、遺伝子ライブラリ、ゲノム、遺伝子発現データ、およびタンパク質活性データが挙げられるが、これらに限定されない。情報のアイテムの形式の例としては、．ｔｘｔ、．ＰＤＦ、．ｄｏｃ、．ｄｏｃｘ、．ｐｐｔ、．ｐｐｔｘ、．ｘｌｓ、．ｘｌｓｘ、．ｒｔｆ、．ｊｐｇ、．ｇｉｆ、．ｐｓｄ、．ｂｍｐ、．ｔｉｆｆ、．ｐｎｇ、および．ｍｐｅｇが挙げられるが、これらに限定されない。デジタル形式での、情報のアイテムをエンコードする個々のファイル・サイズ、または情報のアイテムをエンコードする複数のファイルの量は、最大１０２４バイト（１ＫＢに等しい）、１０２４ＫＢ（１ＭＢに等しい）、１０２４ＭＢ（１ＧＢに等しい）、１０２４ＧＢ（１ＴＢに等しい）、１０２４ＴＢ（１ＰＢに等しい）、１エクサバイト、１ゼタバイト、１ヨタバイト、１ゼノタバイト、またはそれ以上を含むが、これらに限定されない。場合によっては、デジタル情報の量は、少なくとも１ギガバイト（ＧＢ）である。場合によっては、デジタル情報の量は、少なくとも１、２、３、４、５、６、７、８、９、１０、２０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００ギガバイトであるか、または１０００ギガバイトを超える。場合によっては、デジタル情報の量は、少な

くとも１テラバイト（ＴＢ）である。場合によっては、デジタル情報の量は、少なくとも１、２、３、４、５、６、７、８、９、１０、２０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００テラバイトであるか、または１０００テラバイトを超える。場合によっては、デジタル情報の量は、少なくとも１ペタバイト（ＰＢ）である。場合によっては、デジタル情報の量は、少なくとも１、２、３、４、５、６、７、８、９、１０、２０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００ペタバイトであるか、または１０００ペタバイトを超える。場合によっては、デジタル情報は、生物から取得されたゲノム・データを含まない。場合によっては、情報のアイテムはエンコードされる。非限定的なエンコーディング方法の例としては、１ビット／塩基、２ビット／塩基、４ビット／塩基、または他のエンコーディング方法が挙げられる。

【００３９】

情報格納のための方法およびシステム

本明細書では、情報（例えば、デジタル情報）を格納するための方法およびシステムが提供される。場合によっては、本明細書では、エンコードするための方法およびシステムが提供される。場合によっては、情報は、１つまたは複数のオブジェクトを含む。場合によっては、１つまたは複数のオブジェクトは、本明細書に記載された情報のアイテムなどの情報のアイテムを含むが、これに限定されない。場合によっては、１つまたは複数のオブジェクトは、ファイルまたはファイルに関連付けられたメタデータを含む。場合によっては、方法およびシステムは、バイナリ・データなどのデジタル・データをエンコードする。場合によっては、方法およびシステムは、内部コーデック、外部コーデック、またはこれらの組み合わせを含む。場合によっては、バイナリ・データは、バイト・ストリームまたはバイト配列を含む。場合によっては、データまたは１つもしくは複数のオブジェクトは、約１ＧＢ～約１ＴＢである。場合によっては、データは、約１ＧＢ～約１ＴＢである。場合によっては、データまたは１つもしくは複数のオブジェクトは、約１ＧＢ～約１０ＧＢ、約１ＧＢ～約５０ＧＢ、約１ＧＢ～約１００ＧＢ、約１ＧＢ～約５００ＧＢ、約１ＧＢ～約１ＴＢ、約１０ＧＢ～約５０ＧＢ、約１０ＧＢ～約１００ＧＢ、約１０ＧＢ～約５００ＧＢ、約１０ＧＢ～約１ＴＢ、約５０ＧＢ～約１００ＧＢ、約５０ＧＢ～約５００ＧＢ、約５０ＧＢ～約１ＴＢ、約１００ＧＢ～約５００ＧＢ、約１００ＧＢ～約１ＴＢ、または約５００ＧＢ～約１ＴＢである。場合によっては、データは、約１ＧＢ、約１０ＧＢ、約５０ＧＢ、約１００ＧＢ、約５００ＧＢ、または約１ＴＢである。場合によっては、データまたは１つもしくは複数のオブジェクトは、少なくとも約１ＧＢ、約１０ＧＢ、約５０ＧＢ、約１００ＧＢ、または約５００ＧＢである。場合によっては、データまたは１つもしくは複数のオブジェクトは、最大で約１０ＧＢ、約５０ＧＢ、約１００ＧＢ、約５００ＧＢ、または約１ＴＢである。

【００４０】

デジタル情報を格納するシステムは、１つもしくは複数の処理ユニット、１つもしくは複数の処理ユニットと通信するメモリ、メモリに格納され、１つもしくは複数の処理ユニット上で実行される命令、またはこれらの任意の組み合わせを備えることができる。場合によっては、１つまたは複数の処理ユニットおよびメモリは、１つまたは複数の物理的位置または論理的位置にわたって分散される。場合によっては、１つまたは複数の処理ユニットは、中央処理ユニット（ＣＰＵ：ｃｅｎｔｒａｌ ｐｒｏｃｅｓｓｉｎｇ ｕｎｉｔｓ）、グラフィカル処理ユニット（ＧＰＵ：ｇｒａｐｈｉｃａｌ ｐｒｏｃｅｓｓｉｎｇ ｕｎｉｔｓ）、シングル・コア・プロセッサ、マルチコア・プロセッサ、プロセッサ・クラスタ、特定用途向け集積回路（ＡＳＩＣ：ａｐｐｌｉｃａｔｉｏｎ－ｓｐｅｃｉｆｉｃ ｉｎｔｅｇｒａｔｅｄ ｃｉｒｃｕｉｔｓ）、フィールド・プログラマブル・ゲート・アレイ（ＦＰＧＡ：Ｆｉｅｌｄ Ｐｒｏｇｒａｍｍａｂｌｅ Ｇａｔｅ Ａｒｒａｙｓ）などのプログラマブル回路、ＡＩアクセラレータ、およびこれらの変形の任意の組み合わせを含む。場合によっては、処理ユニットのうちの１つまたは複数は、単一命令複数データ（ＳＩＭＤ：Ｓｉｎｇｌｅ Ｉｎｓｔｒｕｃｔｉｏｎ Ｍｕｌｔｉｐｌｅ Ｄａｔａ）ま

たは単一プログラム複数データ（ＳＰＭＤ：Ｓｉｎｇｌｅ　Ｐｒｏｇｒａｍ　Ｍｕｌｔｉｐｌｅ　Ｄａｔａ）並列アーキテクチャを備える。１つまたは複数の処理ユニットの例としては、ＳＩＭＤまたはＳＰＭＤを実装する１つまたは複数のＧＰＵまたはＣＰＵが挙げられる。場合によっては、ＡＩアクセラレータは、Ｇｏｏｇｌｅ－ＴＰＵ、Ｇｒａｐｈｃｏｒｅ、Ｃｅｒｅｂｒａｓ、ＳａｍｂａＮｏｖａ、またはこれらの組み合わせを含む。一部の実施形態では、処理ユニットのうちの１つまたは複数は、ハードウェア実装に加えて、ソフトウェアおよび／またはファームウェアに実装される。処理ユニットのソフトウェアまたはファームウェアの実装は、本明細書に記載されたさまざまな機能を実行するために任意の適切なプログラミング言語で記述された、コンピュータ実行可能命令または機械実行可能命令を含むことができる。１つまたは複数の処理ユニットのソフトウェア実装は、全体的または部分的にメモリに格納され得る。代替的または追加的に、システムは、１つまたは複数のハードウェア・ロジック・コンポーネントを備えることができる。例えば、使用され得るハードウェア・ロジック・コンポーネントの種類の例としては、フィールドプログラマブル・ゲート・アレイ（ＦＰＧＡ）、特定用途向け集積回路（ＡＳＩＣ）、特定用途向け標準製品（ＡＳＳＰ：Ａｐｐｌｉｃａｔｉｏｎ－ｓｐｅｃｉｆｉｃ　Ｓｔａｎｄａｒｄ　Ｐｒｏｄｕｃｔｓ）、システムオンチップ・システム（ＳＯＣ：Ｓｙｓｔｅｍ－ｏｎ－ａ－ｃｈｉｐ　ｓｙｓｔｅｍｓ）、複合プログラマブル・ロジック・デバイス（ＣＰＬＤ：Ｃｏｍｐｌｅｘ　Ｐｒｏｇｒａｍｍａｂｌｅ　Ｌｏｇｉｃ　Ｄｅｖｉｃｅｓ）などが挙げられるが、これらに限定されない。場合によっては、メモリは、リムーバブル・ストレージ、非リムーバブル・ストレージ、ローカル・ストレージ、および／またはリモート・ストレージを含み、命令、データ構造、プログラム・モジュール（例えば、ハッシング・モジュール）、および本明細書に記載された任意の他のデータのストレージを提供する。場合によっては、メモリは、本明細書に記載されたアルゴリズムに関連する情報（例えば、ソフトウェア・コード、パラメータ、実行可能命令など）を格納するために使用される。

【００４１】

メモリに格納された命令は、デジタル情報を格納するための１つまたは複数のステップを含むことができる。デジタル情報を記憶するための１つまたは複数の動作が、図１２に例示されている。点線の動作は、一部の実施形態では実行されてよいが、他の実施形態では実行されなくてよい。場合によっては、１つまたは複数のステップは、１つまたは複数のオブジェクトのデジタル情報を複数のプールに分割すること（１２０５）を含む。場合によっては、１つまたは複数のオブジェクトのうちの１つのオブジェクトは、２つ以上のプールにわたって分割される。場合によっては、複数のプールの各々は、約１ＧＢ～約１ＴＢである。場合によっては、複数のプールの各々は、約１ＧＢ～約１ＴＢである。場合によっては、複数のプールの各々は、約１ＧＢ～約１０ＧＢ、約１ＧＢ～約５０ＧＢ、約１ＧＢ～約１００ＧＢ、約１ＧＢ～約５００ＧＢ、約１ＧＢ～約１ＴＢ、約１０ＧＢ～約５０ＧＢ、約１０ＧＢ～約１００ＧＢ、約１０ＧＢ～約５００ＧＢ、約１０ＧＢ～約１ＴＢ、約５０ＧＢ～約１００ＧＢ、約５０ＧＢ～約５００ＧＢ、約５０ＧＢ～約１ＴＢ、約１００ＧＢ～約５００ＧＢ、約１００ＧＢ～約１ＴＢ、または約５００ＧＢ～約１ＴＢである。場合によっては、複数のプールの各々は、約１ＧＢ、約１０ＧＢ、約５０ＧＢ、約１００ＧＢ、約５００ＧＢ、または約１ＴＢである。場合によっては、複数のプールの各々は、少なくとも約１ＧＢ、約１０ＧＢ、約５０ＧＢ、約１００ＧＢ、または約５００ＧＢである。場合によっては、複数のプールの各々は、最大で約１０ＧＢ、約５０ＧＢ、約１００ＧＢ、約５００ＧＢ、または約１ＴＢである。

【００４２】

場合によっては、１つまたは複数のオブジェクトは、本明細書において前述したように、ファイルなどの情報のアイテムを含む。場合によっては、１つまたは複数のオブジェクトは、情報のアイテムに関連付けられたメタデータ（例えば、ファイルに関連付けられたメタデータ）を含む。オブジェクトに関連付けられたメタデータの非限定的な例としては、オブジェクトに添付されたキーワードのリスト、オブジェクト・サイズ、サムネイル画

像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはオブジェクトの１つまたは複数の側面に関する情報を提供する任意の他のデータ、あるいはこれらの任意の組み合わせが挙げられる。一部の例では、メタデータはカスタマイズ可能である。一部の例では、メタデータは、複数のプール内のオブジェクトを検索するために使用される。

【００４３】

デジタル情報ストレージの例示的な図が、図１４に示されている。図に示されているように、１つまたは複数のオブジェクト１４０５は、複数のプール１４１０に分割され得る。場合によっては、１つのオブジェクトが複数のプールに分割される。場合によっては、サイズに部分的に基づいて、１つのオブジェクトが複数のプールに分割される。場合によっては、１つのオブジェクトが２、３、４、５、６、７、８、９、または１０個のプールに分割される。場合によっては、２つ以上のオブジェクトが複数のプールに分割される。場合によっては、１つまたは複数のオブジェクトがプール内にある。場合によっては、１、２、３、４、５、６、７、８、９、または１０個のオブジェクトがプール内にある。場合によっては、複数のプールが重複している。場合によっては、複数のプールが冗長なプールを含み、２つ以上のプールが同じ１つまたは複数のオブジェクトを含む。場合によっては、２、３、４、５、６、７、８、９、または１０個のプールが同じ１つまたは複数のオブジェクトを含む。

【００４４】

複数のプールのうちの各プールは、プール記述子、プール・アイテム、または終了記述子のうちの任意の１つまたは組み合わせを含むことができる。場合によっては、プールは少なくとも１つのプール・アイテムを含む。場合によっては、プールは２つ以上のプール・アイテムを含む。場合によっては、プールは少なくとも１つのプール記述子を含む。場合によっては、プールは２つ以上のプール記述子を含む。場合によっては、プールは少なくとも１つの終了記述子を含む。場合によっては、プールは２つ以上の終了記述子を含む。例として、各プールは、プール記述子１４１５、１つまたは複数のプール・アイテム１４２０、および終了記述子１４２５を含む。場合によっては、プールは、冗長なプール・アイテム、プール記述子、終了プール記述子、またはこれらの組み合わせを含む。そのような場合、２つ以上のプール・アイテム、プール記述子、終了プール記述子、またはこれらの組み合わせは同一である。場合によっては、２、３、４、５、６、７、８、９、または１０個のプール記述子、終了プール記述子、またはこれらの組み合わせは同一である。

【００４５】

図１２を参照すると、場合によっては、命令における１つまたは複数の動作は、プール記述子、プール・アイテム、および終了記述子を含んでいる複数のプールを生成すること（１２１０）を含む。場合によっては、データがプールに分割され、命令は、プール記述子、プール・アイテム、終了記述子、またはこれらの任意の組み合わせを、複数のプールのうちの各プールに生成することを含む。そのような場合、生成されたプール記述子、プール・アイテム、終了記述子は、プールの各々に追加される。場合によっては、プール記述子は、バージョン、プールＩＤ、プール・アイテム記述子のリスト、またはこれらの任意の組み合わせを含む。場合によっては、バージョンは、（例えば、情報が更新される場合）情報のバージョンを含む。場合によっては、バージョンは、プールの構造のバージョンである。場合によっては、バージョンは、さまざまなファイル・システムの全体的なプール構造を変更することを可能にする。

【００４６】

場合によっては、プールＩＤはプールの一意のＩＤを含む。一部の例では、一意のＩＤは汎用一意識別子（ＵＵＩＤ）を含む。一部の例では、一意のＩＤはコンテンツＩＤを含む。一部の例では、コンテンツＩＤは、コンテンツの著作権または所有権を識別および／または管理するために使用され得るデジタル・フィンガープリント・システムを含む。場合によっては、プール・アイテム記述子のリストは、オブジェクトのパス、オブジェクトのサイズ（例えば、オブジェクトの合計サイズ）、オブジェクト内のプール・アイテムの

範囲、プール内のプール・アイテムのオフセット、またはこれらの任意の組み合わせを含む。一部の例では、オブジェクト内のプール・アイテムの範囲は、オブジェクト内のプール・アイテム内のペイロードの１つまたは複数の位置を含む。一部の例では、１つまたは複数の位置は、プール・アイテム内のペイロードの開始範囲および／または終了範囲（例えば、プール内のプール・アイテム１内の１～６行目、プール・アイテム２内の７～１３行目など）を含む。一部の例では、プール・アイテムのオフセットは、プールのペイロード内の１つまたは複数のプール・アイテムの各々の最初のバイトのペイロード位置を含む。例えば、最初のプール・アイテムのオフセットは０バイトである。最初のプール・アイテムの範囲が１０００～２０００である場合、そのサイズは１０００バイトである。そのような例では、次のプール・アイテムのオフセットは１０００バイトである。場合によっては、プール・アイテムは、データ・ペイロードおよび／またはプール・アイテムのハッシュを含む。場合によっては、データ・ペイロードは、格納されているオブジェクトまたはオブジェクトの一部を含む。場合によっては、プール・アイテムのハッシュは、格納されているオブジェクトまたはオブジェクトの一部のハッシュ値を含む。場合によっては、終了プール記述子は、オブジェクト記述子のリストを含む。場合によっては、オブジェクト記述子のリストは、オブジェクトのパスおよび／またはオブジェクトのハッシュを含む。一部の例では、オブジェクトのパスは一意のパスを含む。一部の例では、オブジェクトのパスは階層（例えば、ディレクトリ階層）を含む。一部の例では、オブジェクトのパスは階層を含まない。

【００４７】

デジタル情報を格納するためのシステムおよび方法は、１つまたは複数のハッシュを含むことができる。場合によっては、ハッシング・モジュールを使用して１つまたは複数のハッシュが決定される。場合によっては、ハッシング・モジュールは、本明細書に記載された処理ユニットなどの、１つまたは複数の処理ユニット上で実行される。場合によっては、ハッシング・モジュールは、１つまたは複数のハッシュを決定するための命令（例えば、ハッシュ関数）を含む。場合によっては、命令（例えば、ハッシュ関数）は、本明細書に記載されたメモリなどのメモリに格納される。場合によっては、オブジェクト、オブジェクトの一部、またはプール・アイテムを含む情報は、ハッシュを使用して格納される。場合によっては、１つまたは複数のプール・アイテムの各々のデータ・ペイロードの第１の１つもしくは複数のハッシュが決定され、および／または１つもしくは複数のオブジェクトの各々の第２の１つもしくは複数のハッシュが決定される（１２１５）。場合によっては、データ・ペイロードは、オブジェクトまたはオブジェクトの一部を含む。場合によっては、プール・アイテムのハッシュは、データ・ペイロードに追加される。場合によっては、オブジェクトのハッシュは、終了プール記述子に追加される。

【００４８】

ハッシュは、ハッシュ関数によって決定されてよい（図１５）。ハッシュ関数は、通常、任意の長さの入力を、固定長（例えば、２２４、２５６、３８４、５１２ビットまたは文字）を有する出力に変換する関数を含む。場合によっては、ハッシュ関数は暗号ハッシュ関数を含む。場合によっては、ハッシュ関数は、ＭＤ－５、ＳＨＡ－１、ＳＨＡ－２、ＳＨＡ－３、ＲＩＰＥＭＤ－１６０、Ｗｈｉｒｌｐｏｏｌ、ＢＬＡＫＥ、ＢＬＡＫＥ２、ＢＬＡＫＥ３、またはこれらの変形を含む。場合によっては、ハッシュ関数はＳＨＡ－２を含む。一部の例では、ＳＨＡ－２は、ＳＨＡ－２２４、ＳＨＡ－２５６、ＳＨＡ－３８４、ＳＨＡ－５１２、ＳＨＡ－５１２／２２４、またはＳＨＡ－５１２／２５６を含む。ハッシュ関数の出力は決定的であり、リバースエンジニアリングの実行を不可能にすることができる。さらに、ハッシュの復号に関与するどの関係者も入力の長さを見分けることができないため、固定長の出力の生成はセキュリティを向上させることができる。一部の例では、識別コード、暗号化キー、パスワード、またはこれらの任意の変形を入力するときに、ハッシュが生成される。一部の例では、ハッシュは、デコーディング中にコンテンツ（例えば、プールに格納された情報のアイテムまたはデジタル情報）の検証を可能にする。

【００４９】
場合によっては、入力１５０５は、オブジェクトを含む。一部の例では、ハッシュ関数１５１０は、ハッシュされた出力（またはハッシュ）１５１５を決定するために使用される。場合によっては、入力１５２０は、オブジェクトを含む。一部の例では、ハッシュ関数１５２５は、ハッシュされた出力（またはハッシュ）１５３０を決定するために使用される。一部の例では、ハッシュ関数１５１０およびハッシュ関数１５２５は、同じハッシュ関数である。一部の例では、ハッシュ関数１５１０およびハッシュ関数１５２５は、両方ともＳＨＡ－２５６である。一部の例では、ハッシュ関数１５１０およびハッシュ関数１５２５は、異なるハッシュ関数である。一部の例では、出力１５１５および出力１５３０は、同じ長さである。一部の例では、出力１５１５および出力１５３０は、両方とも２５６ビットである。一部の例では、出力１５１５および出力１５３０は、異なる長さである。

【００５０】
ハッシュ関数は、ハッシュを生成するための１つまたは複数の動作を含むことができる。場合によっては、ハッシュ関数における１つまたは複数のステップは、ビットを埋め込むことを含む。場合によっては、余分なビットが、ハッシュされるデジタル情報（またはメッセージ）に追加される。一部の例では、デジタル・メッセージの長さが、総ビット数よりも小さいモジュラス値になるように、余分なビットがメッセージに追加される。一部の例では、モジュラス値は６４ビットである。一部の例では、ビット数は５１２ビットであり、デジタル情報の長さは４４８ビットである（例えば、ＳＨＡ－２５６の場合）。一部の例では、最初の余分なビットは、１の２進数を含む。一部の例では、その後に追加される余分なビットは、０の２進数を含む。

【００５１】
場合によっては、ハッシュ関数における１つまたは複数のステップは、ある長さを埋め込むことを含む。場合によっては、この長さを埋め込むことは、モジュラス値をデジタル情報に追加することを含む（例えば、バイエンディアン（ＢＥ：ｂｉ－ｅｎｄｉａｎ）整数とも呼ばれる）。モジュラス値またはＢＥ整数は、通常、元のデジタル情報を２進数で含んでいる元の入力の長さを表す。一部の例では、モジュラス値は６４ビットである。一部の例では、６４ビットが４４８ビットのデジタル・メッセージに追加され、総ビット数は５１２ビットである（例えば、ＳＨＡ－２５６の場合）。場合によっては、モジュラス値は、モジュラスを元のデジタル情報に適用することによって計算される。例として、元のデジタル情報が２進数での「ｈｅｌｌｏ　ｗｏｒｌｄ」である場合、元の入力の長さは８８ビットであり、２進数では「１０１１０００」である。そのため、総ビット数が５１２になるように、複数の０の後に「１０１１０００」が続く値が、４４８ビットのデジタル情報の末尾に追加される。

【００５２】
場合によっては、ハッシュ関数における１つまたは複数のステップは、１つまたは複数のハッシュ値またはバッファを初期化することを含む。場合によっては、８つのハッシュ値またはバッファが初期化される。場合によっては、初期化されたハッシュ値はハードコードされる（例えば、定数である）。場合によっては、初期化されたハッシュ値は、最初の８つの素数（例えば、２、３、５、７、１１、１３、１７、１９）の平方根の小数部の最初の３２ビットを表す。場合によっては、ハッシュ関数における１つまたは複数のステップは、ラウンド定数（またはキー）を初期化することをさらに含む。場合によっては、６４個のラウンド定数が初期化される。一部の例では、６４個のラウンド定数の各々は、最初の６４個の素数（例えば、２～３１１）の立方根の小数部の最初の３２ビットを表す。場合によっては、６４個の異なるラウンド定数は、配列に格納される。

【００５３】
場合によっては、ハッシュ関数における１つまたは複数のステップは、圧縮を含む。場合によっては、（例えば、５１２ビットごとの）情報の各ブロックが圧縮される。圧縮中に、情報の各ブロックは、固定回数のラウンドを経る。場合によっては、ラウンド数は６

４である。場合によっては、圧縮は、一方向圧縮関数によって実行される。場合によっては、一方向圧縮関数は、単一ブロック長圧縮関数である。一部の例では、圧縮関数は、Ｄａｖｉｅｓ－Ｍｅｙｅｒ、Ｍａｔｙａｓ－Ｍｅｙｅｒ－Ｏｓｅａｓ、またはＭｉｙａｇｕｃｈｉ－Ｐｒｅｎｅｅｌ圧縮関数である。場合によっては、一方向圧縮関数は、倍ブロック長圧縮関数である。一部の例では、圧縮関数は、ＭＤＣ－２／Ｍｅｙｅｒ－Ｓｃｈｉｌｌｉｎｇ、ＭＤＣ－４、またはＨｉｒｏｓｅ圧縮関数である。場合によっては、圧縮関数からの出力は、情報のブロックよりも小さい。一部の例では、この出力は２５６ビットの長さを有する。

【００５４】

場合によっては、ハッシュ（例えば、プール・アイテムのハッシュ、オブジェクトのハッシュ）のうちの１つまたは複数が、情報の格納中に計算される。場合によっては、ハッシュ（例えば、プール・アイテムのハッシュ、オブジェクトのハッシュ）のすべてが、情報の格納中に計算される。一部の例では、これによって、オブジェクトのサイズに関係なく、安定した少ないメモリ使用量を可能にする。場合によっては、各プール・アイテムのデータ・ペイロードの第１の１つまたは複数のハッシュは、１つまたは複数のオブジェクトよりも少ないメモリを必要とする。場合によっては、１つまたは複数のオブジェクトの各々の第２の１つまたは複数のハッシュは、１つまたは複数のプール・アイテムよりも少ないメモリを必要とする。場合によっては、ソース・データ（例えば、情報のアイテム）は、一度だけ読み取られる。場合によっては、プールの各々は、シークなしで一度書き込まれる。一部の例では、これによって、データ転送および待ち時間を最小限に抑える。

【００５５】

場合によっては、本明細書に記載されたハッシュは、１つまたは複数の目的を果たすことができる。１つまたは複数の目的は、非限定的な例として、１つまたは複数の情報のアイテム（例えば、オブジェクト）の完全性の検証、署名の生成および検証（例えば、デジタル署名の場合）、パスワード検証、プルーフ・オブ・ワーク、または情報のアイテムの識別子のうちの１つまたは複数を含むことができる。

【００５６】

場合によっては、暗号化および／または圧縮が、さらに追加され得る。一部の例では、暗号化および／または圧縮は、ストリーミング・アプリケーション・プログラマブル・インターフェイス（ＡＰＩ：ａｐｐｌｉｃａｔｉｏｎ ｐｒｏｇｒａｍｍａｂｌｅ ｉｎｔｅｒｆａｃｅ）を使用して実装され得る。一部の例では、これによって、中間結果を格納する必要性を回避する。場合によっては、格納されるデジタル情報は、例えばデータ転送コストを削減するために、すでに圧縮されている。場合によっては、例えばセキュリティ上の理由から、格納されるデジタル情報はすでに暗号化されている。

【００５７】

メモリに格納された命令における１つまたは複数の動作は、複数のインデックス・プールを作成することをさらに含むことができる。場合によっては、複数のインデックス・プールはインデックスのみを含む。場合によっては、インデックス・プールは、複数のポリヌクレオチドにエンコードされた複数のプールに格納されたオブジェクトを取り出すときに使用される。場合によっては、インデックス・プールは、順序付けられ、オブジェクトを検索するためにデジタル・ストレージ・システム（例えば、フラッシュ・ドライブ）に一時的に格納される。一部の例では、プールが識別された後に、プールをエンコードする複数のポリヌクレオチドがシーケンシングされる。

【００５８】

場合によっては、１つまたは複数のインデックス・プールは、インデックス・プール記述子および／またはオブジェクト・インデックス作成のリストを含む。場合によっては、インデックス・プール記述子は、バージョン、プールＩＤ、プールのサイズ、タイムスタンプ、またはこれらの組み合わせを含む。一部の例では、プールＩＤはプールの一意のＩＤを含む。一部の例では、一意のＩＤは汎用一意識別子（ＵＵＩＤ）を含む。一部の例では、一意のＩＤはコンテンツＩＤを含む。一部の例では、コンテンツＩＤは、コンテンツ

の著作権または所有権を識別および／または管理するために使用され得るデジタル・フィンガープリント・システムを含む。一部の例では、複数のインデックス・プールの各々のサイズは、約１ＧＢ～約１ＴＢである。場合によっては、オブジェクト・インデックス作成のリストは、オブジェクトのパス、オブジェクトのハッシュ、オブジェクト・フラグメントのリスト、オブジェクト・メタデータのリスト、またはこれらの任意の組み合わせを含む。一部の例では、オブジェクトのパスは一意のパスを含む。一部の例では、オブジェクトのパスは階層（例えば、ディレクトリ階層）を含む。一部の例では、オブジェクトのパスは階層を含まない。一部の例では、オブジェクトのハッシュは、本明細書において前述したようなハッシュである（例えば、ＳＨＡ－２５６）。一部の例では、オブジェクト・フラグメントのリストは、フラグメントを含むプールのプールＩＤ、フラグメントの範囲、またはこれらの組み合わせを含む。一部の例では、オブジェクト・メタデータのリストは、メタデータの種類、メタデータ・ペイロード、またはこれらの組み合わせを含む。一部の例では、メタデータの種類は、オブジェクトに添付されたキーワードのリスト、サムネイル画像、テキスト要約、ソートされたキー値データベースのＩＤ範囲、タイムスタンプ、バージョン、またはこれらの任意の組み合わせを含む。一部の例では、メタデータはカスタマイズ可能である。一部の例では、メタデータは、複数のプール内のオブジェクトを検索するために使用される。

【００５９】

場合によっては、インデックス・プールは、約１～約１００万個のプールの情報を格納することができる。場合によっては、インデックス・プールは、約１個のプール～約１０個のプール、約１個のプール～約１００個のプール、約１個のプール～約１，０００個のプール、約１個のプール～約５，０００個のプール、約１個のプール～約１０，０００個のプール、約１個のプール～約５０，０００個のプール、約１個のプール～約１００，０００個のプール、約１個のプール～約５００，０００個のプール、約１個のプール～約１００万個のプール、約１０個のプール～約１００個のプール、約１０個のプール～約１，０００個のプール、約１０個のプール～約５，０００個のプール、約１０個のプール～約１０，０００個のプール、約１０個のプール～約５０，０００個のプール、約１０個のプール～約１００，０００個のプール、約１０個のプール～約５００，０００個のプール、約１０個のプール～約１００万個のプール、約１００個のプール～約１，０００個のプール、約１００個のプール～約５，０００個のプール、約１００個のプール～約１０，０００個のプール、約１００個のプール～約５０，０００個のプール、約１００個のプール～約１００，０００個のプール、約１００個のプール～約５００，０００個のプール、約１００個のプール～約１００万個のプール、約１，０００個のプール～約５，０００個のプール、約１，０００個のプール～約１０，０００個のプール、約１，０００個のプール～約５０，０００個のプール、約１，０００個のプール～約１００，０００個のプール、約１，０００個のプール～約５００，０００個のプール、約１，０００個のプール～約１００万個のプール、約５，０００個のプール～約１０，０００個のプール、約５，０００個のプール～約５０，０００個のプール、約５，０００個のプール～約１００，０００個のプール、約５，０００個のプール～約５００，０００個のプール、約５，０００個のプール～約１００万個のプール、約１０，０００個のプール～約５０，０００個のプール、約１０，０００個のプール～約１００，０００個のプール、約１０，０００個のプール～約５００，０００個のプール、約１０，０００個のプール～約１００万個のプール、約５０，０００個のプール～約１００，０００個のプール、約５０，０００個のプール～約５００，０００個のプール、約５０，０００個のプール～約１００万個のプール、約１００，０００個のプール～約５００，０００個のプール、約１００，０００個のプール～約１００万個のプール、または約５００，０００個のプール～約１００万個のプールの情報を格納することができる。場合によっては、インデックス・プールは、約１個のプール、約１０個のプール、約１００個のプール、約１，０００個のプール、約５，０００個のプール、約１０，０００個のプール、約５０，０００個のプール、約１００，０００個のプール、約５００，０００個のプール、または約１００万個のプールの情報を格納することがで

きる。場合によっては、インデックス・プールは、少なくとも約１個のプール、約１０個のプール、約１００個のプール、約１，０００個のプール、約５，０００個のプール、約１０，０００個のプール、約５０，０００個のプール、約１００，０００個のプール、または約５００，０００個のプールの情報を格納することができる。場合によっては、インデックス・プールは、最大で約１０個のプール、約１００個のプール、約１，０００個のプール、約５，０００個のプール、約１０，０００個のプール、約５０，０００個のプール、約１００，０００個のプール、約５００，０００個のプール、または約１００万個のプールの情報を格納することができる。

【００６０】

場合によっては、１つまたは複数のインデックス・プールの各々は、約１ＧＢ～約１ＴＢである。場合によっては、複数のプールの各々は、約１ＧＢ～約１ＴＢである。場合によっては、１つまたは複数のインデックス・プールの各々は、約１ＧＢ～約１０ＧＢ、約１ＧＢ～約５０ＧＢ、約１ＧＢ～約１００ＧＢ、約１ＧＢ～約５００ＧＢ、約１ＧＢ～約１ＴＢ、約１０ＧＢ～約５０ＧＢ、約１０ＧＢ～約１００ＧＢ、約１０ＧＢ～約５００ＧＢ、約１０ＧＢ～約１ＴＢ、約５０ＧＢ～約１００ＧＢ、約５０ＧＢ～約５００ＧＢ、約５０ＧＢ～約１ＴＢ、約１００ＧＢ～約５００ＧＢ、約１００ＧＢ～約１ＴＢ、または約５００ＧＢ～約１ＴＢである。場合によっては、１つまたは複数のインデックス・プールの各々は、約１ＧＢ、約１０ＧＢ、約５０ＧＢ、約１００ＧＢ、約５００ＧＢ、または約１ＴＢである。場合によっては、１つまたは複数のインデックス・プール各々は、少なくとも約１ＧＢ、約１０ＧＢ、約５０ＧＢ、約１００ＧＢ、または約５００ＧＢである。場合によっては、１つまたは複数のインデックス・プールの各々は、最大で約１０ＧＢ、約５０ＧＢ、約１００ＧＢ、約５００ＧＢ、または約１ＴＢである。

【００６１】

エンコーディング方式は、複数のプールおよび／またはインデックス・プールの各々に適用され得る。場合によっては、エンコーディング方式は、複数のプール内のデジタル情報を複数のポリヌクレオチドとしてエンコードする（１２２０）。場合によっては、エンコーディング方式は、インデックス・プール内のデジタル情報を複数のポリヌクレオチドとしてエンコードする。場合によっては、エンコーディング方式は、バイナリ・データをポリヌクレオチド配列としてエンコードするためのコーデック（例えば、内部コーデック）を含む。場合によっては、エンコード方式はエラー訂正コード（ＥＣＣ）を含む。場合によっては、エンコーディング方式（例えば、内部コーデックまたは低レベル・コーデック）は、ストリーミング読み取りおよび書き込みＡＰＩアクセスを可能にするようにも設計および実装される。場合によっては、エンコーディング方式（例えば、内部コーデックまたは低レベル・コーデック）は、本明細書に記載されたデジタル・ストレージ（例えば、高レベル・コーデック）のためのシステムおよび方法のストリーミングに一致するようにも設計および実装される。

【００６２】

エンコーディング方式は、通常、１つまたは複数の動作を含むことができる。１つまたは複数の動作は、データ（例えば、デジタル情報）を操作または変換するための１つまたは複数の動作を含むことができる。１つまたは複数の動作は、非限定的な例として、データまたはデータの一部を分割すること、シャッフルすること、連結すること、転置すること、変換すること、複製すること、（例えば、インデックスを使用して）ラベル付けすること、あるいはこれらの任意の組み合わせを含むことができる。

【００６３】

デジタル情報（例えば、バイナリ・データ）を複数のポリヌクレオチド配列にエンコードする方法が、図１で概略的に説明されている。場合によっては、デジタルまたはデータを複数のポリヌクレオチド配列にエンコードするための方法は、データを分割することを含む。場合によっては、データは複数のフレームに分割される（１０５）。場合によっては、複数のフレームは、約１００～約１０，０００フレームを含む。場合によっては、複数のフレームは、約１００フレーム～約２５０フレーム、約１００フレーム～約５００フ

レーム、約１００フレーム～約７５０フレーム、約１００フレーム～約１，０００フレーム、約１００フレーム～約２，５００フレーム、約１００フレーム～約５，０００フレーム、約１００フレーム～約７，５００フレーム、約１００フレーム～約１０，０００フレーム、約２５０フレーム～約５００フレーム、約２５０フレーム～約７５０フレーム、約２５０フレーム～約１，０００フレーム、約２５０フレーム～約２，５００フレーム、約２５０フレーム～約５，０００フレーム、約２５０フレーム～約７，５００フレーム、約２５０フレーム～約１０，０００フレーム、約５００フレーム～約７５０フレーム、約５００フレーム～約１，０００フレーム、約５００フレーム～約２，５００フレーム、約５００フレーム～約５，０００フレーム、約５００フレーム～約７，５００フレーム、約５００フレーム～約１０，０００フレーム、約７５０フレーム～約１，０００フレーム、約７５０フレーム～約２，５００フレーム、約７５０フレーム～約５，０００フレーム、約７５０フレーム～約７，５００フレーム、約７５０フレーム～約１０，０００フレーム、約１，０００フレーム～約２，５００フレーム、約１，０００フレーム～約５，０００フレーム、約１，０００フレーム～約７，５００フレーム、約１，０００フレーム～約１０，０００フレーム、約２，５００フレーム～約５，０００フレーム、約２，５００フレーム～約７，５００フレーム、約２，５００フレーム～約１０，０００フレーム、約５，０００フレーム～約７，５００フレーム、約５，０００フレーム～約１０，０００フレーム、または約７，５００フレーム～約１０，０００フレームを含む。場合によっては、複数のフレームは、約１００フレーム、約２５０フレーム、約５００フレーム、約７５０フレーム、約１，０００フレーム、約２，５００フレーム、約５，０００フレーム、約７，５００フレーム、または約１０，０００フレームを含む。場合によっては、複数のフレームは、少なくとも約１００フレーム、約２５０フレーム、約５００フレーム、約７５０フレーム、約１，０００フレーム、約２，５００フレーム、約５，０００フレーム、または約７，５００フレームを含む。場合によっては、複数のフレームは、最大で約２５０フレーム、約５００フレーム、約７５０フレーム、約１，０００フレーム、約２，５００フレーム、約５，０００フレーム、約７，５００フレーム、または約１０，０００フレームを含む。場合によっては、フレームは、それぞれ同じ量のデータを含む。別の事例では、フレームは、それぞれ異なる量のデータを含む。場合によっては、各フレームにフレーム・インデックスが割り当てられる。一部の例では、フレーム・インデックスは、フレーム・インデックスごとに増加する（例えば、０、１、２、３、４、５、．．．など）。一部の例では、フレーム・インデックスは、フレーム・インデックスごとに単調に増加する。

【００６４】

デジタル・データまたはバイナリ・データをエンコードするための方法は、外部コーデックを含む。場合によっては、デジタル・データまたはバイナリ・データを複数のポリヌクレオチド配列にエンコードするための方法は、外部コーデックを含む。場合によっては、外部コーデックは、データ（例えば、バイナリ・データ）に適用される。場合によっては、データが複数のフレームに分割された後に、外部コーデックがデータに適用される（１１０）。そのような場合、外部コーデックは、複数のフレームの各々に適用される。データ・ストリームをフレームに分割し、外部コーデックを適用する例示的な図が、図３に例示されている。

【００６５】

場合によっては、外部コーデックは、エラー訂正方式、またはリード・ソロモン（ＲＳ）コードなどのエラー訂正コードを含む。この外部コーデックは、格納されるデジタル・データまたはバイナリ・データを多くのオリゴヌクレオチドにわたって分散するために使用される。場合によっては、データを分散することによって、消失（例えば、失われたオリゴ）を修正するために使用され得る冗長性を形成する。さらに場合によっては、データを分散することによって、内部コーデックからのエラーを修正するための冗長性も形成する。

【００６６】

場合によっては、エラー訂正方式はリード・ソロモン（ＲＳ）コードを含む。そのよう

な場合、バイナリ・データまたはバイナリ・データを含む複数のフレームをエンコードするために、ＲＳエンコーダが使用される。通常、ＲＳコードは、有限体要素のセットとして扱われるデータのブロックに対して動作する。場合によっては、ＲＳコードは、データ、例えば、$x=(x_1,...,x_k)\in F^k$を多項式p_xにマッピングすることを含み、ここで、

【００６７】

【数１】

$$p_x(a)=\sum_{i=1}^{k} x_i\, a^{i-1}$$

である。エンコードされたデータＣ（ｘ）は、体Ｆ内のさまざまな異なるｎ個の点$a_1,...,a_n$でp_xを評価することによって取得される（例えば、$C(x)=p_x(a_1),...,p_x(a_n)$）。

【００６８】

一部のさらなる実施形態では、ＲＳコードは、各コードワードがメッセージを接頭辞として含み、エラー訂正シンボルが接尾辞として追加されるエンコーディング方式を含む。場合によっては、ＲＳコードは、ｍビットのシンボルを使用してＲＳ（ｎ，ｋ）として指定される。そのような場合、エンコーダは、それぞれｍビットのｋ個のデータ・シンボルを受け取り、パリティ・シンボル（エラー訂正シンボルまたはチェック・シンボル）を追加して、ｎ個のシンボル・コードワードを作成する。ここで、それぞれｍビットの$n-k$個のパリティ・シンボル（またはチェック・シンボル、ｔ）が存在する。場合によっては、ＲＳデコーダは、コードワードにエラーを含む最大でｔ個までのシンボルを修正し、$2t=n-k$である。コードワードＣ（ｘ）は、メッセージ情報Ｍ（ｘ）に体系的に追加されるパリティ・チェック情報ＣＫ（ｘ）を含む。コードワードＣ（ｘ）は、$C(x)=x^{n-k}M(x)+CK(x)=x^{n-k}M(x)+x^{n-k}M(x) \bmod g(x)$として計算され得る。ここで、ｋはメッセージ長（例えば、シンボル）を指し、ｔは修正されるエラーの数を指し、ｎはブロック長（例えば、メッセージ長ｎに修正長さｔを加算した値）を指し、ｍはシンボル幅を指し、シンボルサイズｍを前提として、ＲＳコードの最大コードワード長ｎは$n=2^m-1$である。さらに、x^{n-k}はメッセージ内の変位シフトを指し、ｇ（Ｘ）は生成器多項式を指し、生成器多項式は、ガロア体（ＧＦ：Ｇａｌｏｉｓ　ｆｉｅｌｄ）原始αの逐次累乗である根を有する多項式として定義される（例えば、$g(x)=(x-\alpha^i)(x-\alpha^{i+1})\cdots(x-\alpha^{i+n-k-1})=g_0+g_1x+\cdots+g_{n-k-1}x^{n-1-1}+x^{n-k}$）。

【００６９】

例えば、８ビットのシンボルを使用するＲＳ（２５５，２２３）では、ブロック長ｎは２５５コードワード・バイトであり、メッセージ長ｋは２２３バイトであり、パリティ２ｔは３２バイトである。そのような例では、ＲＳデコーダは、コードワード内の最大で１６個までのシンボル・エラーを修正し、つまり、デコーダによって最大で１６バイトまでが修正され得る。ＲＳコードは、ガロア体に基づいてＧＦ（2^m）として示されることも可能である。例えば、図３に示されているように、ＲＳ　ＧＦ（2^{12}）エンコーディング方式では、ｎは４０９５である（例えば、$n=2^{12}-1=4096-1=4095$）。ｋが例えば２４９９である場合、$2t=4095-2499=1596$であり、ｔは７９８である。

【００７０】

場合によっては、エラー訂正方式は、低密度パリティ・チェック（ＬＤＰＣ）コードなどの線形エラー訂正コード（または線形ブロック・コード）を含む。場合によっては、エラー訂正方式は、極コードなどの線形ブロック・エラー訂正コードを含む。一部のさらなる実施形態では、エラー訂正方式は、ターボコードなどの高性能な前方エラー訂正（ＦＥＣ：ｆｏｒｗａｒｄ　ｅｒｒｏｒ　ｃｏｒｒｅｃｔｉｏｎ）を含む。場合によっては、エラー訂正方式は、ＲＳコード、ＬＤＰＣコード、ターボコード、極コード、またはこれら

の任意の組み合わせ（例えば、ＲＳベースのＬＤＰＣコード）を含む。

【００７１】

場合によっては、エラー訂正方式は、低密度パリティ・チェック（ＬＤＰＣ）コードを含む。そのような場合、バイナリ・データまたはバイナリ・データを含む複数のフレームをエンコードするために、ＬＤＰＣコードが使用される。通常、ＬＤＰＣコードの構造は、ほとんどのエントリで０を含み、それ以外で１を含むパリティ・チェック行列によって定義される。例えば、Ｋ個の情報ビットの（Ｎ，Ｋ）ＬＤＰＣコードは、Ｎのブロック・サイズを有する線形ブロック・コードであり、１以外のすべての要素が０であるスパースな（Ｎ－Ｋ）×Ｎパリティ・チェック行列によって定義される。行または列内の１の数は、行または列の次数と呼ばれる。場合によっては、長さＮのコードワードはベクトルＣとして表され、長さＫの情報ビットの場合、２Ｋ個のコードワードを含む（Ｎ，Ｋ）コードが使用される。場合によっては、（Ｎ，Ｋ）ＬＤＰＣコードは、$HC^{T}=0$という条件を満たす（Ｎ－Ｋ）×Ｎパリティ・チェック行列Ｈによって定義される。

【００７２】

場合によっては、ＬＤＰＣコードは、パリティ・チェック行列の各行および各列が一定の次数を有する場合、規則的であり、そうでない場合、不規則である。場合によっては、不規則なＬＤＰＣコードは、規則的なＬＤＰＣコードよりも性能が優れている。場合によっては、行間および列間で次数が異なるため、不規則なＬＤＰＣコードは、行の次数および列の次数が適切に調整されている場合にのみ、改善された性能を期待させる。

【００７３】

場合によっては、エラー訂正方式は極コードを含む。場合によっては、極コードは、理論的証明によってシャノン容量を達成できる。場合によっては、極コードは、低いエンコーディングおよびデコーディングの複雑さを含む。極コードは、通常、生成器行列G_Nを含み、情報は、$x_1^N=u_1^N G^N$に従ってエンコードされることが可能であり、x_1^Nはエンコーディング後の出力ビットであり、u_1^Nはエンコーディング前の入力ビットであり、生成器行列は

【００７４】

【数２】

$$G^N=B_N F^{\otimes n}$$

として定義される。コード長ＮはＮ＝２ｎとして定義され、ｎ≧０である。B_Nは、例えば、ビット反転行列などの転置行列を含む。

【００７５】

【数３】

$$F^{\otimes n}$$

はＦのクロネッカー累乗を含み、

【００７６】

【数４】

$$F^{\otimes n}=F^{\otimes}\ F^{\otimes n-1}$$

として定義され、

【００７７】

【数５】

$$F=\begin{bmatrix}1 & 0\\ 1 & 1\end{bmatrix}$$

である。
【００７８】
場合によっては、極コードは、コセカント・コード（ｃｏｓｅｃ　ｃｏｄｅ）を使用して（Ｎ、Ｋ、Ａ、$u_A{}^c$）として表され、エンコーディング・プロセスは、
【００７９】
【数６】

$$x_1^N = u_A G_N(A) \otimes u_A{}^c G_N(A^c)$$

として定義され、Ａは情報ビット・インデックス・セットであり、G_N（Ａ）は、G_N内のセットＡにおけるインデックスに対応する、行から取得された部分行列である。さらに、G_N（A^c）は、G_N内のセットA^cにおけるインデックスに対応する、行から取得された部分行列であり、$u_A{}^c$は、その数が（Ｎ－Ｋ）である凍結ビットであり、Ｎはコード長であり、Ｋは情報ビットの長さである。場合によっては、凍結ビットが０に設定され、上記のエンコーディング・プロセスは$x_1{}^N = u_A G_N$（Ａ）として表される。
【００８０】
場合によっては、エラー訂正方式はターボコードを含む。ターボコードは、通常、同じ情報配列の異なるインターリーブされたバージョンに適用される、２つ以上のコンポーネント・コードの並列連結を含む。通常、再帰的系統畳み込み（ＲＳＣ：ｒｅｃｕｒｓｉｖｅ　ｓｙｓｔｅｍａｔｉｃ　ｃｏｎｖｏｌｕｔｉｏｎａｌ）コードが、コンポーネント・コードとして使用される。例えば、ターボコードの構造は、並列に連結された２つのＲＳＣエンコーダ（例えば、Ｍ＝２）を含み、（近似的に）Ｒ＝１／（Ｍ＋１）であるため、コード・レートＲはＲ＝１／３である。第１のＲＳＣエンコーダへの入力は、元の情報配列である。元の情報配列ｄは、インターリーブされたバージョンｄ’を生成するために、インターリーバにも適用される。情報配列のインターリーブされたバージョンｄ’は、第２のＲＳＣエンコーダへの入力である。ターボ・エンコーダからの出力は、ｕならび冗長な部分に$x_{(1)}$（第１のＲＳＣエンコーダからの出力）および$x_{(2)}$（第２のエンコーダからの出力）の系統的配列を含む。したがって、エンコーダの出力はu_1、$x_{1(1)}$、$x_{1(2)}$、u_2、$x_{2(1)}$、$x_{2(2)}$を含み、u_kはｋ番目の系統的ビット（すなわち、データ・ビット）であり、$x_{k(1)}$はｋ番目の系統的ビットu_kに関連付けられた第１のＲＳＣエンコーダからのパリティ出力であり、$x_{k(2)}$はｋ番目の系統的ビットu_kに関連付けられた第２のＲＳＣエンコーダからのパリティ出力である。ターボコードのデコーディング手順は、通常、反復的デコーディングを含む。ターボコードのデコーディング手順は、（２つのＲＳＣエンコーダに対応する）２つのコンポーネント・デコーダ、インターリーバ、およびデインターリーバを含むことができる。場合によっては、２つのコンポーネント・デコーダは、ソフト入力およびソフト出力（ＳＩＳＯ：ｓｏｆｔ－ｉｎｐｕｔ　ａｎｄ　ｓｏｆｔ－ｏｕｔｐｕｔ）デコーダである。場合によっては、２つのコンポーネント・デコーダの出力は、符号化されたデータ配列に関する尤度情報を含む。
【００８１】
場合によっては、外部コーデックが適用されると、データのサイズが増大する。場合によっては、外部コーデックがデータを含んでいるフレームの各々に適用されると、フレーム・サイズが増大する。場合によっては、フレームが複数のレーンに分割される（１１５）。場合によっては、各レーンはレーン・インデックスを含む。場合によっては、各フレームは、約１０００～約１０，０００レーンを含む。場合によっては、各フレームは約５０００レーンを含む。場合によっては、各フレームは、約１，０００レーン～約２，５００レーン、約１，０００レーン～約５，０００レーン、約１，０００レーン～約７，５００レーン、約１，０００レーン～約１０，０００レーン、約２，５００レーン～約５，０００レーン、約２，５００レーン～約７，５００レーン、約２，５００レーン～約１０，０００レーン、約５，０００レーン～約７，５００レーン、約５，０００レーン～約１０

，０００レーン、または約７，５００レーン～約１０，０００レーンを含む。場合によっては、各フレームは、約１，０００レーン、約２，５００レーン、約５，０００レーン、約７，５００レーン、または約１０，０００レーンを含む。場合によっては、各フレームは、少なくとも約１，０００レーン、約２，５００レーン、約５，０００レーン、または約７，５００レーンを含む。場合によっては、各フレームは、最大で約２，５００レーン、約５，０００レーン、約７，５００レーン、または約１０，０００レーンを含む。各レーンは、約１００～約３００ビットをさらに含む。場合によっては、各レーンは、約１００ビット～約１５０ビット、約１００ビット～約２００ビット、約１００ビット～約２５０ビット、約１００ビット～約３００ビット、約１５０ビット～約２００ビット、約１５０ビット～約２５０ビット、約１５０ビット～約３００ビット、約２００ビット～約２５０ビット、約２００ビット～約３００ビット、または約２５０ビット～約３００を含む。場合によっては、各レーンは、約１００、１１０、１２０、１３０、１４０、１５０、１６０、１７０、１８０、１９０、２００、２１０、２２０、２３０、２４０、２５０、２６０、２７０、２８０、２９０、３００ビットを含む。場合によっては、各レーンは、少なくとも約１００、１１０、１２０、１３０、１４０、１５０、１６０、１７０、１８０、１９０、２００、２１０、２２０、２３０、２４０、２５０、２６０、２７０、２８０、２９０、３００ビットを含む。場合によっては、各レーンは、最大で約１００、１１０、１２０、１３０、１４０、１５０、１６０、１７０、１８０、１９０、２００、２１０、２２０、２３０、２４０、２５０、２６０、２７０、２８０、２９０、３００ビットを含む。本明細書において提供されるエンコードするための方法は、例えばバイナリ・データを使用して示されているが、場合によっては、方法は、一般に、複数のシンボルを含んでいるデータに適用されてよい。

【００８２】

場合によっては、データを複数のポリヌクレオチド配列にエンコードするための方法は、データをシャッフルすることを含む。場合によっては、各レーンは、レーン・インデックスに少なくとも部分的に基づいてシャッフルされる（１２０）。場合によっては、外部コーデックをバイナリ・データに適用した後に、各レーンがシャッフルされる。場合によっては、各レーンをシャッフルすることによって、オリゴヌクレオチド・プール全体に影響を与えるエラーなどの、合成またはシーケンシング中に発生する可能性があるエラーに対する耐性を可能にする。エラーは、挿入、欠失、置換、またはこれらの組み合わせを含むことができる。場合によっては、シャッフルすることは、各レーン・インデックスに部分的に基づく各レーン内の回転方式を含む。例えば、レーン内の各ビットは、各レーン・インデックスによってシフトされてよい（例えば、レーン０ではシャッフルなし、レーン１では１ビットのシフト、レーン２では２ビットのシフトなど）。

【００８３】

さらに場合によっては、シャッフルすることは、各レーン内の擬似ランダム・プロセスを含む。この擬似ランダム・シャッフル・プロセスでは、ランダム・シードを使用して疑似乱数生成器を初期化する。場合によっては、疑似乱数生成器によって生成される数値は、ランダム・シードによって決定される。したがって、同じシードを使用して、疑似乱数生成器によって数値の同じ配列が生成される。例として、シャッフルを使用することは擬似ランダム・プロセスを含み、レーン内の各ビットは、疑似乱数生成器によって生成された数値に従ってシフトされる。

【００８４】

さらに場合によっては、レーン・インデックスは、そのレーンの一部またはすべてのビットの順列を作成するためのシードとして使用される。場合によっては、一部またはすべてのビットの順列は、乱数生成器からのサンプリングによって作成される。場合によっては、順列は、コンパイル済みの形態で格納される。場合によっては、疑似乱数生成器の使用は、より小さい実装のソース・コードを可能にする。

【００８５】

場合によっては、フレーム・インデックスおよびレーン・インデックスが先頭に追加さ

れる。場合によっては、各レーンがシャッフルされた後に、フレーム・インデックスおよびレーン・インデックスが各レーンの先頭に追加される。レーンをシャッフルし、フレーム・インデックスおよびレーン・インデックスを先頭に追加する例示的な図が、図４に示されている。場合によっては、フレーム・インデックスは、約１２ビット～約２０ビットを含む。場合によっては、フレーム・インデックスは、約１２ビット～約１４ビット、約１２ビット～約１６ビット、約１２ビット～約１８ビット、約１２ビット～約２０ビット、約１４ビット～約１６ビット、約１４ビット～約１８ビット、約１４ビット～約２０ビット、約１６ビット～約１８ビット、約１６ビット～約２０ビット、または約１８ビット～約２０ビットを含む。場合によっては、フレーム・インデックスは、約１２ビット、約１４ビット、約１６ビット、約１８ビット、または約２０ビットを含む。場合によっては、フレーム・インデックスは、少なくとも約１２ビット、約１４ビット、約１６ビット、または約１８ビットを含む。場合によっては、フレーム・インデックスは、最大で約１４ビット、約１６ビット、約１８ビット、または約２０ビットを含む。場合によっては、レーン・インデックスは、約１２ビット～約１６ビットを含む。場合によっては、レーン・インデックスは、約１２ビット～約１４ビット、約１２ビット～約１６ビット、または約１４ビット～約１６ビットを含む。場合によっては、レーン・インデックスは、約１２ビット、約１４ビット、または約１６ビットを含む。場合によっては、レーン・インデックスは、少なくとも約１２ビットまたは約１４ビットを含む。場合によっては、レーン・インデックスは、最大で約１４ビットまたは約１６ビットを含む。図４に示されているように、場合によっては、レーン・インデックスは１２ビットであり、フレーム・インデックスは２０ビットである。場合によっては、レーン・インデックスは、ＲＳコードからのシンボル幅ｍである。

【００８６】

場合によっては、データを複数のポリヌクレオチド配列にエンコードするための方法は、内部コーデックを含む。場合によっては、内部コーデックは、データ（例えば、バイナリ・データ）に適用される。場合によっては、内部コーデックは、外部コーデックからのデータに適用される。場合によっては、内部コーデックは、データのレーンに適用される。場合によっては、内部コーデックは、レーンがシャッフルされた後にデータのレーンに適用される。

【００８７】

場合によっては、内部コーデックはエンコーディング方式を含む。場合によっては、エンコーディング方式含んでいる内部コーデックが各レーンに適用され、データをポリヌクレオチド配列としてエンコードする（１２５）。内部コーデックは、データ（例えば、デジタル・データまたはバイナリ・データ）をヌクレオチド塩基に変換するために使用される。場合によっては、内部コーデックは、欠失エラー、置換エラー、または挿入エラー、あるいはこれらの任意の組み合わせなどのエラーを修正することができる。一部のさらなる実施形態では、内部コーデックは、オリゴの妥当性を確認し、疑わしいオリゴを破棄して外部デコーディングを汚染することを防ぐために使用される。内部コーデックは、インデックス（フレーム・インデックスおよびレーン・インデックス）をさらにエンコードし、デコード中の効率的なクラスタ化を可能にすることができる。

【００８８】

場合によっては、エンコーディング方式は、複数のポリヌクレオチド配列にわたって冗長性を追加する。場合によっては、冗長性は約５％～約１０％である。場合によっては、冗長性は、約５％～約６％、約５％～約７％、約５％～約８％、約５％～約９％、約５％～約１０％、約６％～約７％、約６％～約８％、約６％～約９％、約６％～約１０％、約７％～約８％、約７％～約９％、約７％～約１０％、約８％～約９％、約８％～約１０％、または約９％～約１０％である。場合によっては、冗長性は、約５％、約６％、約７％、約８％、約９％、または約１０％である。場合によっては、冗長性は、少なくとも約５％、約６％、約７％、約８％、または約９％である。場合によっては、冗長性は、最大で約６％、約７％、約８％、約９％、または約１０％である。場合によっては、この冗長性

は、個々のオリゴに挿入、欠失、置換、またはこれらの任意の組み合わせなどのエラーが存在する場合に、オリゴのプールがデコードされることを可能にする。

【００８９】

エンコーディング方式の例示的な図が、図５に示されている。この例示的な図では、内部コーデックにおけるエンコーディング方式は、各レーンからのビット、ビット履歴、およびビット位置のうちの２つ以上を組み合わせる。場合によっては、既知のビットをコンテキストに分割するためにモデル（例えば、適応モデル）が使用され、各コンテキストがビット履歴にマッピングされる。場合によっては、ビット履歴は、８ビットの状態によって表される。場合によっては、例えば、ルックアップ・テーブルの使用によってコンテキストが検出されるたびに、ビット履歴が更新される。ビット位置は、固定数の最下位ビット（ＬＳＢ：ｌｅａｓｔ　ｓｉｇｎｉｆｉｃａｎｔ　ｂｉｔｓ）を含む。場合によっては、ＬＳＢは、エンコードするビットのビット・インデックスを含む。例えば、１００ビットが１００ｍｅｒオリゴヌクレオチドをエンコードする場合、「ビット・インデックス」は、エンコードするビットにおける０～９９のインデックスを指す。例えば、１００ビットが１００ｍｅｒオリゴヌクレオチドをエンコードする場合、「ビット・インデックス」は、エンコードするビットにおける０～９９のインデックスを指す。ＬＳＢは、整数の２進数１の場所を表す２進整数でのビット位置を含む。場合によっては、ＬＳＢインデックスは任意の長さである。場合によっては、ＬＳＢインデックスは、２ビットの状態、３ビットの状態、または４ビットの状態によって表される。例として、インデックス０、１、２、３、４、５、６、７、．．．は、それぞれ２ビットの状態００、０１、１０、１１、００、０１、１０、１１、．．．として表され得る。図５に示されたエンコーディング方式はバイナリ・データを含むが、場合によっては、エンコーディング方式は、一般に、複数のシンボルを含んでいるデータに適用されてよい。

【００９０】

場合によっては、内部コーデックは、バイナリ・データのビットの塩基候補を生成することを含む。バイナリ・データの塩基候補は、ルックアップ・テーブル、ハッシュ、またはこれらの組み合わせを使用して生成される。場合によっては、ハッシュは、本明細書において前述した方法を使用して決定される。場合によっては、バイナリ・データは、各レーンからのビット、ビット履歴、およびビット位置のうちの２つ以上を含む。場合によっては、エンコーディングのビット・レートは、１塩基当たり約１ビット～１塩基当たり約２ビットである。場合によっては、エンコーディングのためのビット・レートは、１塩基当たり約１ビット～１塩基当たり約１．１ビット、１塩基当たり約１ビット～１塩基当たり約１．２ビット、１塩基当たり約１ビット～１塩基当たり約１．３ビット、１塩基当たり約１ビット～１塩基当たりあたり約１．４ビット、１塩基当たり約１ビット～１塩基当たり約１．５ビット、１塩基当たり約１ビット～１塩基当たり約１．６ビット、１塩基当たり約１ビット～１塩基当たり約１．７ビット、１塩基当たり約１ビット～１塩基当たり約１．８ビット、１塩基当たり約１ビット～１塩基当たり約１．９ビット、１塩基当たり約１ビット～１塩基当たり約２ビット、１塩基当たり約１．１ビット～１塩基当たり約１．２ビット、１塩基当たり約１．１ビット～１塩基当たり約１．３ビット、１塩基当たり約１．１ビット～１塩基当たり約１．４ビット、１塩基当たり約１．１ビット～１塩基当たり約１．５ビット、１塩基当たり約１．１ビット～１塩基当たり約１．６ビット、１塩基当たり約１．１ビット～１塩基当たり約１．７ビット、１塩基当たり約１．１ビット～１塩基当たり約１．８ビット、１塩基当たり約１．１ビット～１塩基当たり約１．９ビット、１塩基当たり約１．１ビット～１塩基当たり約２ビット、１塩基当たり約１．２ビット～１塩基当たり約１．３ビット、１塩基当たり約１．２ビット～１塩基当たり約１．４ビット、１塩基当たり約１．２ビット～１塩基当たり約１．５ビット、１塩基当たり約１．２ビット～１塩基当たり約１．６ビット、１塩基当たり約１．２ビット～１塩基当たり約１．７ビット、１塩基当たり約１．２ビット～１塩基当たり約１．８ビット、１塩基当たり約１．２ビット～１塩基当たり約１．９ビット、１塩基当たり約１．２ビット～１塩基当たり約２ビット、１塩基当たり約１．３ビット～１塩基当たり約１．４ビット、１塩

基当たり約１．３ビット～１塩基当たり約１．５ビット、１塩基あたり約１．３ビット～１塩基当たり約１．６ビット、１塩基当たり約１．３ビット～１塩基当たり約１．７ビット、１塩基当たり約１．３ビット～１塩基当たり約１．８ビット、１塩基当たり約１．３ビット～１塩基当たり約１．９ビット、１塩基当たり約１．３ビット～１塩基当たり約２ビット、１塩基当たり約１．４ビット～１塩基当たり約１．５ビット、１塩基当たり約１．４ビット～１塩基当たり約１．６ビット、１塩基当たり約１．４ビット～１塩基当たり約１．７ビット、１塩基当たり約１．４ビット～１塩基当たり約１．８ビット、１塩基当たり約１．４ビット～１塩基当たり約１．９ビット、１塩基当たり約１．４ビット～１塩基当たり約２ビット、１塩基当たり約１．５ビット～１塩基当たり約１．６ビット、１塩基当たり約１．５ビット～１塩基当たり約１．７ビット、１塩基当たり約１．５ビット～１塩基当たり約１．８ビット、１塩基当たり約１．５ビット～１塩基当たり約１．９ビット、１塩基当たり約１．５ビット～１塩基当たり約２ビット、１塩基当たり約１．６ビット～１塩基当たり約１．７ビット、１塩基当たり約１．６ビット～１塩基当たり約１．８ビット、１塩基当たり約１．６ビット～１塩基当たり約１．９ビット、１塩基当たり約１．６ビット～１塩基当たり約２ビット、１塩基当たり約１．７ビット～１塩基当たり約１．８ビット、１塩基当たり約１．７ビット～１塩基当たり約１．９ビット、１塩基当たり約１．７ビット～１塩基当たり約２ビット、１塩基当たり約１．８ビット～１塩基当たり約１．９ビット、１塩基当たり約１．８ビット～１塩基当たり約２ビット、または１塩基当たり約１．９ビット～１塩基当たり約２ビットである。場合によっては、エンコードするためのビット・レートは、１塩基当たりあたり約１ビット、１塩基当たり約１．１ビット、１塩基当たり約１．２ビット、１塩基当たり約１．３ビット、１塩基当たり約１．４ビット、１塩基当たり約１．５ビット、１塩基当たり約１．６ビット、１塩基当たり約１．７ビット、１塩基当たり約１．８ビット、１塩基当たり約１．９ビット、または１塩基当たり約２ビットである。場合によっては、エンコードするためのビット・レートは、少なくとも１塩基当たり約１ビット、１塩基当たり約１．１ビット、１塩基当たり約１．２ビット、１塩基当たり約１．３ビット、１塩基当たり約１．４ビット、１塩基当たり約１．５ビット、１塩基当たり約１．６ビット、１塩基当たり約１．７ビット、１塩基当たり約１．８ビット、または１塩基当たり約１．９ビットである。場合によっては、エンコードするためのビット・レートは、最大で１塩基当たり約１．１ビット、１塩基当たり約１．２ビット、１塩基当たり約１．３ビット、１塩基当たり約１．４ビット、１塩基当たり約１．５ビット、１塩基当たり約１．６ビット、１塩基当たり約１．７ビット、１塩基当たり約１．８ビット、１塩基当たり約１．９ビット、または１塩基当たり約２ビットである。場合によっては、ビットをヌクレオチドにマッピングするためにルックアップ・テーブルが使用される（例えば、Ａ＝００、Ｔ＝１０、Ｃ＝０１、Ｇ＝１１）。場合によっては、ハッシュは、任意のサイズ（例えば、任意の数のビット）のデータを固定サイズ値（例えば、ヌクレオチドまたはハッシュ値）にマッピングするために使用され得る関数を含む。一部の例では、ハッシュされた値がポリヌクレオチド配列にマッピングされる。

【００９１】

場合によっては、内部コーデックは、塩基反復チェックを含む。場合によっては、塩基候補が選択された後に、塩基反復チェックが実行される。場合によっては、塩基反復チェックは、２つ以上の連続的な塩基における反復をチェックする。場合によっては、塩基反復チェックは、２つ以上の連続的な塩基に反復がある場合に、１つの塩基を別の塩基に置換する。場合によっては、ルックアップ・テーブルまたはハッシュは、塩基反復チェック中に更新された塩基に基づいて更新される。さらに、塩基反復チェック後に、ビット履歴が更新される。場合によっては、フレーム・インデックスおよび／またはレーン・インデックスがインクリメントされる。場合によっては、このプロセスは、複数のポリヌクレオチド配列のすべての配列が決定されるまで繰り返される。

【００９２】

場合によっては、内部コーデックは、複数のポリヌクレオチド配列を合成する前にＧＣフィルタリングを実行することをさらに含む。場合によっては、ＧＣフィルタリングは、

複数のレーンのうちの約１％～約１０％のレーンを除去する。場合によっては、ＧＣフィルタリングは、複数のレーンのうちの約５％～約１０％のレーンを除去する。場合によっては、ＧＣフィルタリングは、複数のレーンうちのレーンを除去しない。場合によっては、ＧＣフィルタリングは、約１％、約２％、約３％、約４％、約５％、約６％、約７％、約８％、約９％、または約１０％を除去する。場合によっては、ＧＣフィルタリングは、少なくとも約１％、約２％、約３％、約４％、約５％、約６％、約７％、約８％、または約９％を除去する。場合によっては、ＧＣフィルタリングは、最大で約２％、約３％、約４％、約５％、約６％、約７％、約８％、約９％、または約１０％を除去する。場合によっては、複数のポリヌクレオチド配列は、約４０％～約６０％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列は、約４０％～約４５％、約４０％～約５０％、約４０％～約５５％、約４０％～約６０％、約４５％～約５０％、約４５％～約５５％、約４５％～約６０％、約５０％～約５５％、約５０％～約６０％、または約５５％～約６０％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列は、約４０％、約４５％、約５０％、約５５％、または約６０％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列は、少なくとも約４０％、約４５％、約５０％、または約５５％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列は、最大で約４５％、約５０％、約５５％、または約６０％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列の少なくとも９０％が、約４０％～約６０％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列の少なくとも９０％が、約４０％～約４５％、約４０％～約５０％、約４０％～約５５％、約４０％～約６０％、約４５％～約５０％、約４５％～約５５％、約４５％～約５５％、約４５％～約６０％、約５０％～約５５％、約５０％～約６０％、または約５５％～約６０％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列の少なくとも９０％が、約４０％、約４５％、約５０％、約５５％、または約６０％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列の少なくとも９０％が、少なくとも約４０％、約４５％、約５０％、または約５５％のＧＣ含量を含む。場合によっては、複数のポリヌクレオチド配列の少なくとも９０％が、最大で約４５％、約５０％、約５５％、または約６０％のＧＣ含量を含む。場合によっては、内部コーデックからの出力は、最終的なオリゴヌクレオチド・プールを含む。

【００９３】

代替のエンコーディング方式の例示的な図が、図６に示されている。場合によっては、内部コーデックにおけるエンコーディング方式は、デフォルトのルックアップ・テーブルから開始することを含む。デフォルトのルックアップ・テーブルは、各レーン内でエンコードするワードを選択するために使用される。場合によっては、ワードは複数のシンボルを含む。一部の例では、ワードは、８ビットのワードまたはバイトである。ルックアップ・テーブルは、各レーン内のワードまたはバイトごとに塩基候補を生成するために適用される。以前にエンコードされたワードまたはバイトに基づいて、次のルックアップ・テーブルが選択される。場合によっては、エンコーディング方式は、本明細書において前述したように、塩基反復チェック、ＧＣフィルタリング、またはこれらの組み合わせを実行することをさらに含む。場合によっては、このプロセスは、複数のポリヌクレオチド配列のすべての配列が決定され得るまで繰り返される。場合によっては、内部コーデックからの出力は、最終的なオリゴヌクレオチド・プールまたは最終的なオリゴヌクレオチド・ライブラリを含む。

【００９４】

場合によっては、ライブラリ内のオリゴヌクレオチド（またはポリヌクレオチド）の各々の長さは、約２０～約５００塩基である。場合によっては、ライブラリ内のオリゴヌクレオチド（またはポリヌクレオチド）の各々の長さは、約２０塩基～約５０塩基、約２０塩基～約１００塩基、約２０塩基～約２００塩基、約２０塩基～約３００塩基、約２０塩基～約４００塩基、約２０塩基～約５００塩基、約５０塩基～約１００塩基、約５０塩基～約２００塩基、約５０塩基～約３００塩基、約５０塩基～約４００塩基、約５０塩基～約５００塩基、約１００塩基～約２００塩基、約１００塩基～約３００塩基、約１００塩

基～約４００塩基、約１００塩基～約５００塩基、約２００塩基～約３００塩基、約２００塩基～約４００塩基、約２００塩基～約５００塩基、約３００塩基～約４００塩基、約３００塩基～約５００塩基、または約４００塩基～約５００塩基である。場合によっては、ライブラリ内のオリゴヌクレオチド（またはポリヌクレオチド）の各々の長さは、約２０塩基、約５０塩基、約１００塩基、約２００塩基、約３００塩基、約４００塩基、または約５００塩基である。場合によっては、ライブラリ内のオリゴヌクレオチド（またはポリヌクレオチド）の各々の長さは、少なくとも約２０塩基、約５０塩基、約１００塩基、約２００塩基、約３００塩基、または約４００塩基である。場合によっては、ライブラリ内のオリゴヌクレオチド（またはポリヌクレオチド）の各々の長さは、最大で約５０塩基、約１００塩基、約２００塩基、約３００塩基、約４００塩基、または約５００塩基である。

【００９５】

場合によっては、本明細書に記載されているように、複数のポリヌクレオチド配列内のデータをエンコードするための方法は、システム上で実行される。場合によっては、そのようなシステムは、メモリ、メモリに動作可能に結合された処理デバイス、またはこれらの組み合わせを備えている装置を備える。場合によっては、メモリは、バイナリ・データ、ポリヌクレオチド配列、またはこれらの組み合わせの情報を格納するために使用される。場合によっては、データ（例えば、バイナリ・データ）、ポリヌクレオチド配列、またはこれらの組み合わせの情報は、本明細書に記載されたエンコーディング方法における１つまたは複数のステップからのものである。場合によっては、メモリは、本明細書に記載されたアルゴリズムに関連する情報（例えば、ソフトウェア・コード、パラメータ、実行可能命令など）を格納するために使用される。一部の例では、メモリは、本明細書に記載された任意の適切なメモリを含むことができる。一部の例では、メモリは、本明細書に記載された実施形態に従って構成され得る。

【００９６】

場合によっては、処理デバイスは、１つまたは複数のエンコーディング・ステップを実行するように構成される。場合によっては、処理デバイスは、データを複数のフレームに分割することと、外部コーデックを複数のフレームのうちの各フレームに適用することと、各フレームを複数のレーンに分割することと、レーン・インデックスに少なくとも部分的に基づいて各レーンをシャッフルすることと、エンコーディング方式を含む内部コーデックを適用して、ポリヌクレオチド配列内の各レーンをエンコードすることと、を含む１つまたは複数の動作を実行するように構成される。場合によっては、複数のフレームのうちの各フレームはフレーム・インデックスを含む。場合によっては、複数のレーンのうちの各レーンはレーン・インデックスを含む。場合によっては、外部コーデックはエラー訂正方式を含む。場合によっては、ポリヌクレオチド配列にエラーが存在する場合にバイナリ・データがデコードされ得るように、エンコーディング方式は冗長性を追加する。

【００９７】

データをエンコードするための方法、システム、およびプラットフォームは、１つまたは複数の制約に対して最適化された内部コーデックを含むことができる。１つまたは複数の制約は、非限定的な例として、核酸合成、後処理、ストレージ、またはシーケンシングに関連することができる。場合によっては、核酸合成は、電気化学的合成、酵素合成、ホスホラミダイト合成、インクジェット印刷、またはこれらの任意の組み合わせを含む。場合によっては、核酸合成に関連する１つまたは複数の制約は、挿入、欠失、または突然変異などの合成エラーを含む。場合によっては、後処理は、ライゲーション、切断、ハイブリダイゼーション、変性、固体支持体への固定、伸長、エラー訂正、濃縮、分離、精製、および増幅のうちの１つまたは複数を含む。場合によっては、ストレージはコールド・データ・ストレージを含む。コールド・データ・ストレージは、一般に、例えば、めったにアクセスされない核酸内のデータのストレージを指すことができる。コールド・データ・ストレージは、頻繁にアクセスされるデータを指す「ホット・ストレージ」の反対であることができる。一部の例では、ストレージはホット・ストレージを含み、ホット・ストレ

ージでは、核酸に格納されたデータが頻繁にアクセスされる。場合によっては、ストレージは、液相または固相での核酸ストレージを含む。一部の例では、ストレージに関連する１つまたは複数の制約は、温度（例えば、室温）、湿度、圧力、塩分濃度、ｐＨ、濃度、時間、光、ＵＶ、O_2、またはこれらの任意の組み合わせを含む。場合によっては、シーケンシングは、次世代シーケンシング、並列シーケンシング、単一分子リアルタイム・シーケンシング、ナノポア・シーケンシング、合成によるシーケンシング、サンガー・シーケンシング、またはこれらの任意の組み合わせを含む。

【００９８】

データをエンコードするための方法、システム、およびプラットフォームは、ポリヌクレオチドの生成に対して最適化された内部コーデックを含むことができる。場合によっては、ポリヌクレオチドの生成はポリヌクレオチドのアセンブリを含む。場合によっては、ポリヌクレオチドの生成はポリヌクレオチドの合成を含む。合成は、本明細書に記載された方法およびシステム、または当技術分野において知られた任意の適切な方法およびシステムを含んでよい。場合によっては、データは１つまたは複数のシンボルを含む。場合によっては、データはシンボルの列またはシンボルの配列を含む。場合によっては、１つまたは複数のシンボルは、バイナリ・データを含む。場合によっては、内部コーデックは、データに適用される。場合によっては、内部コーデックは、本明細書において提供される外部コーデックなどの外部コーデック（例えば、エラー訂正方式）からのデータに適用される。場合によっては、内部コーデックは、暗号化されていないデータに適用される。場合によっては、内部コーデックは、暗号化されたデータに適用される。内部コーデックは、塩基の特定の順序に従ってポリヌクレオチドを生成するように最適化されてよい。場合によっては、これによって、本明細書において提供される内部コーデックを使用してエンコードされない配列を有するポリヌクレオチドを合成するために必要とされる合成サイクルの数（例えば、最適化されていない合成手法）と比較して、合成サイクルの総数が減らされるため、ポリヌクレオチドのより効率的な合成を可能にする。場合によっては、これによって、合成中の酸化ステップおよび脱保護ステップの数が減らされるため、より低いエラー率を可能にする。

【００９９】

本明細書では、データをエンコードするための方法が提供される。場合によっては、この方法は、コードブックを含んでいる内部コーデックを生成することを含む。コードブックは、データをエンコードする核酸の適用、操作、動作、または使用に基づいて最適化されてよい。コードブックは、本明細書において説明されるように、（例えば、核酸合成、後処理、ストレージ、シーケンシングなどに関連する）１つまたは複数の制約に基づいて最適化されてよい。コードブックは、塩基順序を使用して生成されてよい。場合によっては、コードブックは、塩基順序に部分的に基づいて生成されるコードワードを含む。場合によっては、塩基順序は、既定の塩基転移を含む。場合によっては、コードブックは、コードワードを使用して、１つまたは複数のシンボル（例えば、２進数の「０」および「１」）によって表されたデータを、核酸（例えば、Ａ、Ｔ、Ｃ、Ｇ）などの別の１つまたは複数のシンボルにマッピングすることによって、ポリヌクレオチド配列を生成する。場合によっては、特定の、または既定の塩基転移は、塩基順序に従って合成を可能にする。一部の例では、各層で合成順序を変えることによって、パターンの繰り返しが減らされる。特定の層での合成順序の非限定的な例は、［Ａ、Ｇ、Ｃ、Ｔ］、［Ｃ、Ａ、Ｔ、Ｇ］、［Ｔ、Ｇ、Ａ、Ｃ］、または塩基Ａ、Ｔ、Ｇ、Ｃの任意の他の組み合わせを含むことができる。そのような例では、コードブックは、層ごとに変化する。一部の例では、２つの連続する層は同じコードブックを有さない。一部の例では、各層は一意のコードブックを含む。一部の例では、２つ以上の層が同じコードブックを含む。

【０１００】

一部の例では、各塩基で特定の塩基転移のみを許可することによって、パターンの繰り返しが減らされる。例えば、アデニン（Ａ：ａｄｅｎｉｎｅ）の後に、グアニン（Ｇ：ｇｕａｎｉｎｅ）、シトシン（Ｃ：ｃｙｔｏｓｉｎｅ）、またはチミン（Ｔ：ｔｈｙｍｉｎ

ｅ）のみが、配列内の次の塩基として選択され得る。代替として、Ａの後に、塩基が選択されない。一部の例では、Ｇが選択される場合、ＣまたはＴのみが選択され得るか、または代替として、塩基が選択されない。一部の例では、Ｃが選択される場合、Ｔのみが選択され得るか、または代替として、塩基が選択されない。

【０１０１】

場合によっては、コードブックは、１つ、２つ、３つ、４つ、または５つのヌクレオチドを含む。場合によっては、コードブックは、少なくとも１つ、２つ、３つ、または４つのヌクレオチドを含む。場合によっては、コードブックは、最大で２つ、３つ、４つ、または５つのヌクレオチドを含む。場合によっては、コードブックは、４つのヌクレオチド（例えば、アデニン（Ａ）、チミン（Ｔ）、シトシン（Ｃ）、グアニン（Ｇ））を含む。例えば、１つまたは複数の層に対する特定の塩基転移は、（ａ）［Ａ，Ｔ，Ｃ，Ｇ］、（ｂ）［Ａ，Ｔ，Ｇ，Ｃ］、（ｃ）［Ａ，Ｇ，Ｔ，Ｃ］、（ｄ）［Ａ，Ｇ，Ｃ，Ｔ］、（ｅ）［Ａ，Ｃ，Ｇ，Ｔ］、（ｆ）［Ａ，Ｃ，Ｔ，Ｇ］、（ｇ）［Ｔ，Ｃ，Ｇ，Ａ］、（ｈ）［Ｔ，Ｃ，Ａ，Ｇ］、（ｉ）［Ｔ，Ｇ，Ａ，Ｃ］、（ｊ）［Ｔ，Ｇ，Ｃ，Ａ］、（ｋ）［Ｔ，Ａ，Ｇ，Ｃ］、（ｌ）［Ｔ，Ａ，Ｃ，Ｇ］、（ｍ）［Ｃ，Ｇ，Ａ，Ｔ］、（ｎ）［Ｃ，Ｇ，Ｔ，Ａ］、（ｏ）［Ｃ，Ａ，Ｇ，Ｔ］、（ｐ）［Ｃ，Ａ，Ｔ，Ｇ］、（ｑ）［Ｃ，Ｔ，Ａ，Ｇ］、（ｒ）［Ｃ，Ｔ，Ｇ，Ａ］、（ｓ）［Ｇ，Ａ，Ｔ，Ｃ］、（ｔ）［Ｇ，Ａ，Ｃ，Ｔ］、（ｕ）［Ｇ，Ｔ，Ａ，Ｃ］、（ｖ）［Ｇ，Ｔ，Ｃ，Ａ］、（ｗ）［Ｇ，Ｃ，Ｔ，Ａ］、（ｘ）［Ｇ，Ｃ，Ａ，Ｔ］、または（ｙ）これらの任意の組み合わせ、のうちの任意の１つを含む。場合によっては、１つまたは複数の層に対する特定の塩基転移は、天然塩基または標準塩基を含む。場合によっては、１つまたは複数の層に対する特定の塩基転移は、天然塩基または標準塩基を有するヌクレオチドおよび非天然塩基または非標準塩基を有する１つまたは複数のヌクレオチドを含む。例として、コードブックは、［Ａ，Ｇ，Ｃ，Ｔ］（例えば、Ａ，Ｇ，Ｃ，Ｔ，Ａ，Ｇ，Ｃ，Ｔ，．．．）の繰り返しに従う合成順序を含むことができる。そのような例では、コードブックは、Ａ、Ｇ、Ｃ、Ｔ、ＡＧ、ＡＣ、ＡＴ、ＧＣ、ＧＴ、ＡＧＣ、ＡＣＴ、およびＡＧＣＴというコードワードを含むことができる。場合によっては、コードブック内のコードワードは、コードブック内のヌクレオチド数と同等のサイクル数で合成され得る。場合によっては、コードブック内のコードワードは、１、２、３、４、または５サイクルの合成で合成され得る。場合によっては、コードブック内のコードワードは、少なくとも１、２、３、４、または５サイクルの合成で合成され得る。場合によっては、コードブック内のコードワードは、最大で１、２、３、４、または５サイクルの合成で合成され得る。場合によっては、コードブックに関連付けられた転移は、非ランダムまたは擬似非ランダムである。場合によっては、コードブックに関連付けられた転移は、事前に定義された数学的アルゴリズムまたは統計アルゴリズムによって定義される。

【０１０２】

場合によっては、合成順序は、１つまたは複数の層に関して変化することができる。層は、通常、特定の、または既定の順序で、各塩基のフローを含むことができる。例えば、塩基転移が［Ａ，Ｔ，Ｃ，Ｇ］である場合、層は、合成中にＡのフロー、それに続くＴ、Ｃ、およびその後のＧのフローを含む。場合によっては、１つまたは複数の層は、（ａ）［Ａ，Ｔ，Ｃ，Ｇ］、（ｂ）［Ａ，Ｔ，Ｇ，Ｃ］、（ｃ）［Ａ，Ｇ，Ｔ，Ｃ］、（ｄ）［Ａ，Ｇ，Ｃ，Ｔ］、（ｅ）［Ａ，Ｃ，Ｇ，Ｔ］、（ｆ）［Ａ，Ｃ，Ｔ，Ｇ］、（ｇ）［Ｔ，Ｃ，Ｇ，Ａ］、（ｈ）［Ｔ，Ｃ，Ａ，Ｇ］、（ｉ）［Ｔ，Ｇ，Ａ，Ｃ］、（ｊ）［Ｔ，Ｇ，Ｃ，Ａ］、（ｋ）［Ｔ，Ａ，Ｇ，Ｃ］、（ｌ）［Ｔ，Ａ，Ｃ，Ｇ］、（ｍ）［Ｃ，Ｇ，Ａ，Ｔ］、（ｎ）［Ｃ，Ｇ，Ｔ，Ａ］、（ｏ）［Ｃ，Ａ，Ｇ，Ｔ］、（ｐ）［Ｃ，Ａ，Ｔ，Ｇ］、（ｑ）［Ｃ，Ｔ，Ａ，Ｇ］、（ｒ）［Ｃ，Ｔ，Ｇ，Ａ］、（ｓ）［Ｇ，Ａ，Ｔ，Ｃ］、（ｔ）［Ｇ，Ａ，Ｃ，Ｔ］、（ｕ）［Ｇ，Ｔ，Ａ，Ｃ］、（ｖ）［Ｇ，Ｔ，Ｃ，Ａ］、（ｗ）［Ｇ，Ｃ，Ｔ，Ａ］、（ｘ）［Ｇ，Ｃ，Ａ，Ｔ］、または（ｙ）これらの任意の組み合わせ、のうちの任意の１つを含むことができる。場合によっては、層の特定の塩基転移のうちの１つまたは複数は、２回以上繰り返され得る。例として、合成順序は、

［Ａ，Ｇ，Ｃ，Ｔ］、［Ｃ，Ａ，Ｔ，Ｇ］、［Ｔ，Ｇ，Ａ，Ｃ］．．．を含むことができ、配列は、ＡＧＣＴＡＧＣＴＣＡＴＧＴＧＡＣ．．．を含むことができ、最初の層が２回繰り返される。場合によっては、１つまたは複数の層を変化させることによって、配列（例えば、反復的な塩基、高ＧＣ／ＡＴ、または二次構造）内のパターンの繰り返しを減らす。

【０１０３】

場合によっては、内部コーデックは１つまたは複数のコードブックを含む。場合によっては、内部コーデックは、１、２、３、４、５、６、７、８、９、または１０個のコードブックを含む。場合によっては、内部コーデックは、少なくとも１、２、３、４、５、６、７、８、９、または１０個のコードブックを含む。場合によっては、内部コーデックは、最大で１、２、３、４、５、６、７、８、９、または１０個のコードブックを含む。場合によっては、各コードブックは、ポリヌクレオチドの合成中に層をエンコードする。場合によっては、各コードブックは、一意の塩基順序を使用して生成される。場合によっては、各コードブックは、１つまたは複数の塩基転移に対して最適化される。場合によっては、一意の塩基順序は、１つまたは複数の一意の塩基転移を生成する。場合によっては、各コードブックは、特定の層、サイクル・インデックス、履歴、またはこれらの任意の組み合わせでの特定の塩基転移に対して最適化される。一部の例では、履歴は、前の層、前の１つもしくは複数の層をエンコードする１つもしくは複数のコードブック、１つもしくは複数の前の層のサイクル・インデックス、またはこれらの任意の組み合わせのうちの１つもしくは複数を含む。場合によっては、各コードブックは、事前に定義された数学的アルゴリズムまたは統計アルゴリズムによって生成される。

【０１０４】

場合によっては、コードブックは、１つまたは複数のヌクレオチド類似体または非天然／非標準ヌクレオチドを含む。ヌクレオチド類似体または非天然ヌクレオチドは、何らかの種類の修飾を含んでいるヌクレオチドを含む。ヌクレオチド類似体または非天然ヌクレオチドは、塩基、糖、またはリン酸部分のいずれかに対する何らかの種類の修飾を含んでいるヌクレオチドを含む。修飾は、化学的修飾を含むことができる。修飾は、例えば、３’ＯＨまたは５’ＯＨ基、骨格、糖成分、またはヌクレオチド塩基の修飾であってよい。修飾は、非自然発生のリンカー分子の追加、および／または鎖間または鎖内架橋の追加を含んでよい。１つの態様では、修飾核酸は、３’Ｈまたは５’ＯＨ基、骨格、糖成分、またはヌクレオチド塩基のうちの１つまたは複数の修飾、ならびに／あるいは非自然発生のリンカー分子の追加を含む。１つの態様では、修飾骨格は、ホスホジエステル骨格以外の骨格を含む。１つの態様では、修飾糖は、デオキシリボース（修飾ＤＮＡ内）以外またはリボース（修飾ＲＮＡ）以外の糖を含む。１つの態様では、修飾塩基は、修飾塩基は、アデニン、グアニン、シトシン、またはチミン（修飾ＤＮＡ内）以外の塩基、あるいはアデニン、グアニン、シトシン、またはウラシル（修飾ＲＮＡ内）以外の塩基を含む。

【０１０５】

核酸は、少なくとも１つの修飾塩基を含んでよい。塩基部分に対する修飾は、Ａ、Ｃ、Ｇ、およびＴ／Ｕ、ならびに異なるプリン塩基またはピリミジン塩基の、天然および合成の修飾を含む。一部の実施形態では、修飾は、アデニン、グアニン、シトシン、もしくはチミン（修飾ＤＮＡ内）の修飾形態、またはアデニン、グアニン、シトシン、もしくはウラシル（修飾ＲＮＡ）の修飾形態へのものである。修飾塩基のさらなる例は、例えば、ＷＯ２０１９／０１４２６７およびＵＳ２０２２／０２４３２４４において見られることが可能であり、これらは、その全体が参照によって本明細書に組み込まれている。

【０１０６】

一部の実施形態では、コードブックは、１つまたは複数の標準ヌクレオチドおよび１つまたは複数の非標準ヌクレオチドを含む。場合によっては、標準ヌクレオチドは、Ａ、Ｔ、Ｃ、Ｇ、またはＵのうちの１つまたは複数を含む。場合によっては、非標準ヌクレオチドは、本明細書において提供される１つまたは複数のヌクレオチド類似体または非天然ヌクレオチドを含む。場合によっては、非標準ヌクレオチドは、修飾を有する１つまたは複

数の標準ヌクレオチドを含む。場合によっては、コードブックは、約１つ、２つ、３つ、４つ、または５つの標準ヌクレオチドを含む。場合によっては、コードブックは、約１つ、２つ、３つ、４つ、または５つの非標準ヌクレオチドを含む。場合によっては、コードブックは、少なくとも約１つ、２つ、３つ、４つ、または５つの標準ヌクレオチドを含む。場合によっては、コードブックは、少なくとも約１つ、２つ、３つ、４つ、または５つの非標準ヌクレオチドを含む。場合によっては、コードブックは、最大で約１つ、２つ、３つ、４つ、または５つの標準ヌクレオチドを含む。場合によっては、コードブックは、最大で約１つ、２つ、３つ、４つ、または５つの非標準ヌクレオチドを含む。場合によっては、コードブックは、本明細書において提供されるような、標準ヌクレオチドおよび非標準ヌクレオチドの任意の組み合わせを含む。

【０１０７】

場合によっては、コードブックは、約１～約３０個のコードワードを含む。場合によっては、コードブックは、約１～約５、約１～約１０、約１～約１２、約１～約１５、約１～約１８、約１～約２０、約１～約２２、約１～約２５、約１～約２８、約１～約３０、約５～約１０、約５～約１２、約５～約１５、約５～約１８、約５～約２０、約５～約２２、約５～約２５、約５～約２８、約５～約３０、約１０～約１２、約１０～約１５、約１０～約１８、約１０～約２０、約１０～約２２、約１０～約２５、約１０～約２８、約１０～約３０、約１２～約１５、約１２～約１８、約１２～約２０、約１２～約２２、約１２～約２５、約１２～約２８、約１２～約３０、約１５～約１８、約１５～約２０、約１５～約２２、約１５～約２５、約１５～約２８、約１５～約３０、約１８～約２０、約１８～約２２、約１８～約２５、約１８～約２８、約１８～約３０、約２０～約２２、約２０～約２５、約２０～約２８、約２０～約３０、約２２～約２５、約２２～約２８、約２２～約３０、約２５～約２８、約２５～約３０、または約２８～約３０個のコードワードを含む。場合によっては、コードブックは、約１、約５、約１０、約１２、約１５、約１８、約２０、約２２、約２５、約２８、または約３０個のコードワードを含む。場合によっては、コードブックは、少なくとも約１、約５、約１０、約１２、約１５、約１８、約２０、約２２、約２５、または約２８個のコードワードを含む。場合によっては、コードブックは、最大で約５、約１０、約１２、約１５、約１８、約２０、約２２、約２５、約２８、または約３０個のコードワードを含む。

【０１０８】

コードブックを含んでいる内部コーデックは、データを複数のポリヌクレオチド配列としてエンコードするために適用され得る。場合によっては、データはデジタル・データを含む。場合によっては、データは１つまたは複数のシンボルを含む。場合によっては、コードブックに基づいて、１つまたは複数のシンボルが複数のポリヌクレオチド配列にマッピングされる。例えば、２進数（０または１の配列）などの数値が、コードブック内のコードワードにマッピングされ得る。場合によっては、内部コーデックは、１つまたは複数の制約に対してさらに最適化される。１つまたは複数の制約は、複数のポリヌクレオチド配列に関連する制約を含むことができる。一部の例では、１つまたは複数の制約は、複数のポリヌクレオチド配列の長さを含む。一部の例では、１つまたは複数の制約は、複数のポリヌクレオチド配列のＧＣ含量を含む。一部の例では、１つまたは複数の制約は、複数のポリヌクレオチド配列の塩基の繰り返しを含む。一部の例では、１つまたは複数の制約は、挿入、突然変異、または欠失などの、１つまたは複数のエラーを含む。場合によっては、バイナリ・データをコードワードにマッピングすることによって、転移のグラフを作成する。転移は、バイナリ・データの値および位置（例えば、インデックス）に基づくコードワードおよびコードブック間の１つまたは複数の転移を含むことができる。場合によっては、デコーディング中に推定された欠失率、挿入率、および／または突然変異率に基づいて、１つまたは複数の確率が計算される。場合によっては、デコーディング・アルゴリズムは、本明細書（例えば、図８および図９）において提供されるように、転移確率を最大化するための１つまたは複数の解決策を見つける。

【０１０９】

場合によっては、複数のポリヌクレオチド配列の一部が、冗長性のためにエンコードする。場合によっては、冗長性のためにエンコードする複数のポリヌクレオチド配列の一部は、約２０％～約８０％である。場合によっては、冗長性のためにエンコードする複数のポリヌクレオチド配列の一部は、約２０％～約３０％、約２０％～約４０％、約２０％～約５０％、約２０％～約６０％、約２０％～約７０％、約２０％～約８０％、約３０％～約４０％、約３０％～約５０％、約３０％～約６０％、約３０％～約７０％、約３０％～約８０％、約４０％～約５０％、約４０％～約６０％、約４０％～約７０％、約４０％～約８０％、約５０％～約６０％、約５０％～約７０％、約５０％～約８０％、約６０％～約７０％、約６０％～約８０％、または約７０％～約８０％である。場合によっては、冗長性のためにエンコードする複数のポリヌクレオチド配列の一部は、約２０％、約３０％、約４０％、約５０％、約６０％、約７０％、または約８０％である。場合によっては、冗長性のためにエンコードする複数のポリヌクレオチド配列の一部は、少なくとも約２０％、約３０％、約４０％、約５０％、約６０％、または約７０％である。場合によっては、冗長性のためにエンコードする複数のポリヌクレオチド配列の一部は、最大で約約３０％、約４０％、約５０％、約６０％、約７０％、または約８０％である。

【０１１０】

場合によっては、複数のポリヌクレオチド配列は、同じ長さである。場合によっては、複数のポリヌクレオチド配列の約７０％～約１００％は、同じ長さを有する。場合によっては、複数のポリヌクレオチド配列の約７０％～約７５％、約７０％～約８０％、約７０％～約８５％、約７０％～約９０％、約７０％～約９５％、約７０％～約１００％、約７５％～約８０％、約７５％～約８５％、約７５％～約９０％、約７５％～約９５％、約７５％～約１００％、約８０％～約８５％、約８０％～約９０％、約８０％～約９５％、約８０％～約１００％、約８５％～約９０％、約８５％～約９５％、約８５％～約１００％、約９０％～約９５％、約９０％～約１００％、または約９５％～約１００％は、同じ長さを有する。場合によっては、複数のポリヌクレオチド配列の約７０％、約７５％、約８０％、約８５％、約９０％、約９５％、または約１００％は、同じ長さを有する。場合によっては、複数のポリヌクレオチド配列の少なくとも約７０％、約７５％、約８０％、約８５％、約９０％、または約９５％は、同じ長さを有する。場合によっては、複数のポリヌクレオチド配列の最大で約７５％、約８０％、約８５％、約９０％、約９５％、または約１００％は、同じ長さを有する。場合によっては、複数のポリヌクレオチド配列は、異なる長さである。場合によっては、複数のポリヌクレオチド配列は、１％～約３０％異なる。場合によっては、複数のポリヌクレオチド配列は、約１％～約５％、約１％～約１０％、約１％～約１５％、約１％～約２０％、約１％～約２５％、約１％～約３０％、約５％～約１０％、約５％～約１５％、約５％～約２０％、約５％～約２５％、約５～約３０％、約１０％～約１５％、約１０％～約２０％、約１０％～約２５％、約１０％～約３０％、約１５％～約２０％、約１５％～約２５％、約１５％～約３０％、約２０％～約２５％、約２０％～約３０％、または約２５％～約３０％異なる。場合によっては、複数のポリヌクレオチド配列は、約１％、約５％、約１０％、約１５％、約２０％、約２５％、または約３０％異なる。場合によっては、複数のポリヌクレオチド配列は、少なくとも約１％、約５％、約１０％、約１５％、約２０％、または約２５％異なる。場合によっては、複数のポリヌクレオチド配列は、最大で約５％、約１０％、約１５％、約２０％、約２５％、または約３０％異なる。

【０１１１】

複数のポリヌクレオチド配列を含む複数のポリヌクレオチドが生成され得る。場合によっては、複数のポリヌクレオチドが合成される。場合によっては、合成は塩基ごとの合成を含む。場合によっては、合成は合成サイクルを含む。合成サイクルは、一般に、ヌクレオチド結合を実現するために実行される１つまたは複数のステップのことを指す。合成サイクルは、デブロッキング（または脱保護）、結合、酸化、およびキャッピングのうちの１つまたは複数を含むことができる。場合によっては、合成は複数の合成サイクルを含む。内部コーデックは、必要とされる合成サイクルの数を減らすことによって、より効率的

な合成を可能にすることができる。場合によっては、内部コーデックによってエンコードされた複数のポリヌクレオチド配列を含む複数のポリヌクレオチドを合成するために必要とされる合成サイクルの数は、内部コーデックによってエンコードされていない配列を有する複数のポリヌクレオチドを合成するために必要とされる合成サイクルの数と比較して、減少する。場合によっては、合成サイクルの数は約５～約８０％減少する。場合によっては、合成サイクルの数は、約５％～約１０％、約５％～約２０％、約５％～約３０％、約５％～約４０％、約５％～約５０％、約５％～約６０％、約５％～約７０％、約５％～約８０％、約１０％～約２０％、約１０％～約３０％、約１０％～約４０％、約１０％～約５０％、約１０％～約６０％、約１０％～約７０％、約１０％～約８０％、約２０％～約３０％、約２０％～約４０％、約２０％～約５０％、約２０％～約６０％、約２０％～約７０％、約２０％～約８０％、約３０％～約４０％、約３０％～約５０％、約３０％～約６０％、約３０％～約７０％、約３０％～約８０％、約４０％～約５０％、約４０％～約６０％、約４０％～約７０％、約４０％～約８０％、約５０％～約６０％、約５０％～約７０％、約５０％～約８０％、約６０％～約７０％、約６０％～約８０％、または約７０％～約８０％減少する。場合によっては、合成サイクルの数は、約５％、約１０％、約２０％、約３０％、約４０％、約５０％、約６０％、または約７０％、または約８０％減少する。場合によっては、合成サイクルの数は、少なくとも約５％、約１０％、約２０％、約３０％、約４０％、約５０％、約６０％、または約７０％減少する。場合によっては、合成サイクルの数は、最大で約１０％、約２０％、約３０％、約４０％、約５０％、約６０％、約７０％、または約８０％減少する。例として、４つのヌクレオチドを使用する１２コードワードを有する内部コーデックは、約５０％の冗長性で複数のポリヌクレオチド配列をエンコードする。したがって、この例では、バイナリ・データの６つの値が、$\log_2(12)=3.58$ビットの情報に相当する１２コードワードにマッピングされる。しかし、冗長性、例えば２倍の冗長性を考慮すると、この情報は、１コードワード当たり３．５８／２＝１．７９ビットの情報に対応する。複数のポリヌクレオチド配列の各々におけるペイロードが約１００ビットである場合、約１００ビット／１コードワード当たり１．７９ビット＝５５．８コードワードが必要である。本明細書において説明されるように、最適化された内部コーデックおよびサイクルの順序付けを使用すると、コードワードは４サイクルを必要とし、その結果、約５５．８コードワード×１コードワード当たり４サイクル＝約２２４サイクルの合成が得られる。しかし、内部コーデックを使用しない場合、合成は４００サイクル（例えば、４×１００）を必要とする。さらなる例として、複数のポリヌクレオチド配列の各々におけるペイロードが約３００ビットである場合、約４４７サイクル（例えば、２００／１．７９×４）の合成が必要である。しかし、内部コーデックを使用しない場合、合成は８００サイクル（例えば、４×２００）を必要とする。追加の例として、複数のポリヌクレオチド配列の各々におけるペイロードが約３００ビットである場合、約６７０サイクル（例えば、３００／１．７９×４）の合成が必要である。しかし、内部コーデックを使用しない場合、合成は１２００サイクル（例えば、４×３００）を必要とする。

【０１１２】

場合によっては、複数のポリヌクレオチドは、同じ長さである。場合によっては、複数のポリヌクレオチドの約７０％～約１００％は、同じ長さを有する。場合によっては、複数のポリヌクレオチドの約７０％～約７５％、約７０％～約８０％、約７０％～約８５％、約７０％～約９０％、約７０％～約９５％、約７０％～約１００％、約７５％～約８０％、約７５％～約８５％、約７５％～約９０％、約７５％～約９５％、約７５％～約１００％、約８０％～約８５％、約８０％～約９０％、約８０～約９５％、約８０％～約１００％、約８５％～約９０％、約８５％～約９５％、約８５％～約１００％、約９０％～約９５％、約９０％～約１００％、または約９５％～約１００％は、同じ長さを有する。場合によっては、複数のポリヌクレオチドの約７０％、約７５％、約８０％、約８５％、約９０％、約９５％、または約１００％は、同じ長さを有する。場合によっては、複数のポリヌクレオチドの少なくとも約７０％、約７５％、約８０％、約８５％、約９０％、また

は約９５％は、同じ長さを有する。場合によっては、複数のポリヌクレオチドの最大で約７５％、約８０％、約８５％、約９０％、約９５％、または約１００％は、同じ長さを有する。場合によっては、複数のポリヌクレオチドは、異なる長さである。場合によっては、複数のポリヌクレオチドは、１％～約３０％異なる。場合によっては、複数のポリヌクレオチドは、約１％～約５％、約１％～約１０％、約１％～約１５％、約１％～約２０％、約１％～約２５％、約１％～約３０％、約５％～約１０％、約５％～約１５％、約５％～約２０％、約５％～約２５％、約５％～約３０％、約１０％～約１５％、約１０％～約２０％、約１０％～約２５％、約１０％～約３０％、約１５％～約２０％、約１５％～約２５％、約１５％～約３０％、約２０％～約２５％、約２０％～約３０％、または約２５％～約３０％異なる。場合によっては、複数のポリヌクレオチドは、約１％、約５％、約１０％、約１５％、約２０％、約２５％、または約３０％異なる。場合によっては、複数のポリヌクレオチドは、少なくとも約１％、約５％、約１０％、約１５％、約２０％、または約２５％異なる。場合によっては、複数のポリヌクレオチドは、最大で約５％、約１０％、約１５％、約２０％、約２５％、または約３０％異なる。場合によっては、ＰＣＲの効率は、同じ長さを有するポリヌクレオチドの量に関連する。場合によっては、同じ長さを有する複数のポリヌクレオチドは、ＰＣＲがポリヌクレオチドの分布を変更しないことを保証する。場合によっては、同じ長さを有する複数のポリヌクレオチドの９０％以上は、ＰＣＲがポリヌクレオチドの分布を変更しないことを保証する。

【０１１３】

場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、４００未満である。場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、３００未満である。場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、２００未満である。場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、約３００である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、約２００である。場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、約２２４である。場合によっては、合成サイクルの数は、１００塩基を含むポリヌクレオチド配列の場合、約１００である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、８００未満である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、６００未満である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、５００未満である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、４００未満である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、３００未満である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、約５００である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、約４００である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、約３００である。場合によっては、合成サイクルの数は、２００塩基を含むポリヌクレオチド配列の場合、約２００である。場合によっては、合成サイクルの数は、３００塩基を含むポリヌクレオチド配列の場合、１２００未満である。場合によっては、合成サイクルの数は、３００塩基を含むポリヌクレオチド配列の場合、１０００未満である。場合によっては、合成サイクルの数は、３００塩基を含むポリヌクレオチド配列の場合、８００未満である。場合によっては、合成サイクルの数は、３００塩基を含むポリヌクレオチド配列の場合、６００未満である。場合によっては、合成サイクルの数は、３００塩基を含むポリヌクレオチド配列の場合、４００未満である。場合によっては、合成サイクルの数は、３００塩基を含むポリヌクレオチド配列の場合、約６００である。場合によっては、合成サイクルの数は、３００塩基を含むポリヌクレオチド配列の場合、約５００である。場合によっては、合成サイクルの数は、３００塩基を含むポリヌクレオチド配列の場合、約４５０である。場合によっては、合成サイクルの数は、３００塩基を含むポリヌクレオチド配列の場合、約４５０である。場合によっては、ポリヌクレオチド配列は、４つのヌ

クレオチドを含む。場合によっては、ポリヌクレオチド配列は、Ａ、Ｔ、Ｃ、およびＧのうちの１つまたは複数を含む。場合によっては、ポリヌクレオチド配列は、１つ、２つ、３つ、４つ、または５つのヌクレオチドを含む。場合によっては、ポリヌクレオチド配列は、少なくとも１つ、２つ、３つ、４つ、または５つのヌクレオチドを含む。場合によっては、ポリヌクレオチド配列は、最大で１つ、２つ、３つ、４つ、または５つのヌクレオチドを含む。場合によっては、ポリヌクレオチド配列の約１０％、２０％、２５％、３０％、３３％、４０％、５０％、６０％、６６％、７０％、７５％、８０％、または９０％は、冗長性のためにエンコードする。場合によっては、ポリヌクレオチド配列の約１０％、２０％、２５％、３０％、３３％、４０％、５０％、６０％、６６％、７０％、７５％、８０％、または９０％までは、冗長性のためにエンコードする。場合によっては、ポリヌクレオチド配列の最大で約１０％、２０％、２５％、３０％、３３％、４０％、５０％、６０％、６６％、７０％、７５％、８０％、または９０％は、冗長性のためにエンコードする。一部の例では、ポリヌクレオチド配列は、約１．５倍、２倍、２．５倍、３倍、３．５倍、または４倍の冗長性を含む。

【０１１４】

場合によっては、複数のポリヌクレオチドは、本明細書において提供される固体支持体などの固体支持体上で合成される。固体支持体は、本明細書において提供されるような基板であることができる。場合によっては、固体支持体は、複数の特徴（または遺伝子座）を備える。複数のポリヌクレオチドは、複数の特徴の上で合成され得る。場合によっては、合成サイクルごとに、複数の特徴の約２５％～約８０％がデブロッキングされる。場合によっては、合成サイクルごとに、複数の特徴の約２５％～約３０％、約２５％～約３５％、約２５％～約４０％、約２５％～約４５％、約２５％～約５０％、約２５％～約５５％、約２５％～約６０％、約２５％～約６５％、約２５％～約７０％、約２５％～約７５％、約２５～約８０％、約３０％～約３５％、約３０％～約４０％、約３０％～約４５％、約３０％～約５０％、約３０％～約５５％、約３０％～約６０％、約３０％～約６５％、約３０％～約７０％、約３０％～約７５％、約３０％～約８０％、約３５％～約４０％、約３５％～約４５％、約３５％～約５０％、約３５％～約５５％、約３５％～約６０％、約３５～約６５％、約３５～約７０％、約３５～約７５％、約３５％～約８０％、約４０％～約４５％、約４０％～約５０％、約４０％～約５５％、約４０％～約６０％、約４０％～約６５％、約４０％～約７０％、約４０％～約７５％、約４０％～約８０％、約４５％～約５０％、約４５％～約５５％、約４５％～約６０％、約４５％～約６５％、約４５％～約７０％、約４５％～約７５％、約４５％～約８０％、約５０％～約５５％、約５０％～約６０％、約５０％～約６５％、約５０％～約７０％、約５０％～約７５％、約５０％～約８０％、約５５％～約６０％、約５５％～約６５％、約５５％～約７０％、約５５％～約７５％、約５５％～８０％、約６０％～約６５％、約６０％～約７０％、約６０％～約７５％、約６０％～約８０％、約６５％～約７０％、約６５％～約７５％、約６５％～約８０％、約７０％～約７５％、約７０％～約８０％、または約７５％～約８０％がデブロッキングされる。場合によっては、合成サイクルごとに、複数の特徴の約２５％、約３０％、約３５％、約４０％、約４５％、約５０％、約５５％、約６０％、約６５％、約７０％、約７５％、または約８０％がデブロッキングされる。場合によっては、合成サイクルごとに、複数の特徴の少なくとも約２５％、約３０％、約３５％、約４０％、約４５％、約５０％、約５５％、約６０％、約６５％、約７０％、または約７５％がデブロッキングされる。場合によっては、合成サイクルごとに、複数の特徴の最大で約３０％、約３５％、約４０％、約４５％、約５０％、約５５％、約６０％、約６５％、約７０％、約７５％、または約８０％がデブロッキングされる。

【０１１５】

固体支持体上の複数の特徴は、独立してアドレス指定可能である。場合によっては、複数の特徴は、特定のセクションへの試薬のアクセスを制御することによって、独立してアドレス指定可能である。場合によっては、複数の特徴は、複数の特徴のうちの各特徴でのポリヌクレオチドの反応性を制御することによって、独立してアドレス指定可能である。

場合によっては、複数の特徴は、固体支持体の１つまたは複数の電極を介して独立してアドレス指定可能である。アドレス指定可能な遺伝子座（例えば、特徴）を含んでいる固体支持体を備えるデバイスの例は、米国特許第１０９３６９５３号または米国特許第９２６７２１３号に記載されており、これらはその全体が参照によって本明細書に組み込まれている。場合によっては、複数の特徴は、特定の領域をマスクすることによってアドレス指定可能である。一部の例では、特定の領域は、例えば、疎水性または親水性の化学基で表面を修飾することなどによって、化学的に官能化される。例として、複数の特徴は、米国特許第１０８９４２４２号、米国特許第１０１９５５８０号、またはＷＯ２０２２／０４７０７６に記載されている方法を使用してマスクされてよく、これらはその全体が参照によって本明細書に組み込まれている。場合によっては、複数の特徴は、電気化学的デブロッキングによってアドレス指定可能である。場合によっては、複数の特徴は、酸生成によってアドレス指定可能である。場合によっては、１つまたは複数の化学反応（例えば、ヌクレオチド脱保護のための電気化学的に生成された酸（ＥＧＡ：ｅｌｅｃｔｒｏｃｈｅｍｉｃａｌｌｙ　ｇｅｎｅｒａｔｅｄ　ａｃｉｄ））を生成するために、１つまたは複数の電極が使用され得る。場合によっては、電気化学的デブロッキングは、さまざまなオリゴマー（例えば、オリゴヌクレオチド）のいずれかの合成中にデブロッキングするための有機溶媒ベースの溶液を含む。そのような場合、酸ベースの化学的デブロッキングは、分子上のブロッキング部分の除去を含み、次のヌクレオチドの共有結合を可能にすることができる。電気化学的デブロッキングは、酸または塩基（電極が陽極であるか、または陰極によって異なる）を局所的に生成するための、固体支持体（例えば、電極マイクロアレイ）上の１つまたは複数の電極を介した１つまたは複数の特徴への電圧または電流の印加を含み、化学種に結合された、酸または塩基に不安定な保護基（部分）の除去に影響を与える可能性がある。場合によっては、光生成酸を使用してアドレス指定可能なマスキング技術が、デブロッキングのために光増感剤と組み合わせて使用される。場合によっては、複数の特徴は、金属触媒による脱保護（例えば、パラジウム触媒による脱保護）によってアドレス指定可能である。

【０１１６】

場合によっては、複数の特徴は、マスキング方法によってアドレス指定可能である。場合によっては、リフトオフ製造方法が使用され得る（図１１Ａ）。リフトオフ法は、場合によっては、酸化物層でコーティングされたベース層への犠牲層（例えば、フォトレジストまたは「ＰＲ」）の追加、導電層の追加、および犠牲層の除去を含む。場合によっては、ドライエッチング製造方法が使用され得る（図１１Ｂ）。ドライエッチング法は、場合によっては、ベース層への、酸化物層、第１の中間層（例えば、ＴｉＮまたは他の材料）、導電層（例えば、白金）、第２の中間層（例えば、ＴｉＮまたは他の材料）、および犠牲層（例えば、フォトレジスト）などの１つまたは複数の層の追加、導電層を露出するための第２の中間層の部分的除去、第１の中間層を露出するための導電層の部分的除去、第１の中間層を露出するための第１の導電層の部分的除去、ならびに酸化物層を露出するための第１の中間層の部分的除去を含む。例として、シリコンのベース層および酸化物を含んでいる最上層を含む表面が、フォトレジストなどの除去できるマスキング材料を使用してパターン形成され得る（図１１Ａ）。マスクを含む全面が、白金を使用してめっきされることが可能であり、次に、マスク層が除去され得る。その後、前にマスクされていた領域は、露出された酸化物になり、マスクされていない領域は、酸化物層の上に白金を含む。さらなる例として、シリコンのベース層、酸化物を含んでいる第１の層、窒化チタンの第２の層、白金を含んでいる第３の層、窒化チタンを含んでいる第４の層を（下から上へ）含む表面が、フォトレジストなどの除去できるマスキング材料を使用してパターン形成され得る（図１１Ｂ）。マスクされていない第４の層が除去されて、第３の層を露出することができ、フォトレジストが除去されて、マスクされた第４層を露出することができる。残りのすべての第２の層および第４の層の除去によって、シリコンのベース層および酸化物の最上層、ならびに窒化チタンの上にパターン形成された白金の「島」を含む表面を生成することができる。

【０１１７】
場合によっては、電気化学試薬を生成するための１つまたは複数の電極は、非限定的な例として、イリジウムおよび／または白金などの金属、ならびにパラジウム、金、銀、銅、水銀、ニッケル、亜鉛、チタン、タングステン、アルミニウムなどの他の金属に加えて、さまざまな金属の合金、およびガラス状炭素、網状ガラス質炭素、基底面グラファイト、エッジ・プレーン・グラファイト、またはグラファイトを含む炭素などの、他の導電性材料を含んでよい。場合によっては、インジウム・スズ酸化物などのドープ酸化物、ならびに酸化ケイ素およびガリウムヒ素などの半導体が使用されてもよい。さらに、電極は、導電性ポリマー、金属ドープ・ポリマー、導電性セラミックス、および導電性粘土で構成されてよい。場合によっては、白金およびパラジウムは、水素を吸収する能力（例えば、使用される前に水素を「あらかじめ投入される」能力）に関連する有利な特性を含む。場合によっては、１つまたは複数の電極が電源に接続されてよい。場合によっては、電極は、ＣＭＯＳ（相補型金属酸化物半導体：ｃｏｍｐｌｅｍｅｎｔａｒｙ ｍｅｔａｌ ｏｘｉｄｅ ｓｅｍｉｃｏｎｄｕｃｔｏｒ）スイッチング回路、無線およびマイクロ波周波数アドレス指定可能スイッチ、光アドレス指定可能スイッチ、電極から半導体チップの周囲のボンド・パッドへの直接接続、またはこれらの任意の組み合わせによって、電源に接続される。ＣＭＯＳスイッチング回路は、ＣＭＯＳトランジスタ・スイッチへの電極の各々の接続を含むことができる。スイッチは、電子アドレス信号を、共通バスを下って各電極に関連付けられたＳＲＡＭ（スタティック・ランダム・アクセス・メモリ：ｓｔａｔｉｃ ｒａｎｄｏｍ ａｃｃｅｓｓ ｍｅｍｏｒｙ）回路に送信することによって、アクセスされてよい。スイッチが「オン」であるとき、電極は電源に接続され得る。無線およびマイクロ波周波数アドレス指定可能スイッチは、ＲＦ信号またはマイクロ波信号によって電極がスイッチングされることに関与することができる。これによって、スイッチング論理を使用する場合、および／またはスイッチング論理を使用しない場合の両方で、スイッチが入れられることを可能にすることができる。スイッチは、特定の周波数または変調周波数を受信するように調整され、スイッチング論理なしでスイッチングすることができる。光アドレス指定可能スイッチは、光によってスイッチングされ得る。場合によっては、スイッチング論理を使用して、および／またはスイッチング論理を使用せずに、１つまたは複数の電極がスイッチングされることも可能である。一部の例では、光信号は、スイッチング論理なしでスイッチングを可能にするために、例えば、電極アレイ上上でレーザー・ビームを走査することによって、空間的に局在することができ、レーザーが電極を照らすたびに、電極がスイッチングされる。

【０１１８】
複数のポリヌクレオチドの配列が決定されてよい。場合によっては、本明細書において提供されるシステムおよび方法に従って、複数のポリヌクレオチドがシーケンシングされてよい。シーケンシングは、非限定的な例として、次世代シーケンシング、並列シーケンシング、単一分子リアルタイム・シーケンシング、ナノポア・シーケンシング、合成によるシーケンシング、サンガー・シーケンシング、またはこれらの任意の組み合わせを含んでよい。場合によっては、複数のポリヌクレオチドは、シーケンサによってシーケンシングされてよい。場合によっては、複数のポリヌクレオチドをシーケンシングすることは、複数の出力配列を生成する。場合によっては、複数の出力配列は、複数のポリヌクレオチド配列と重複する。場合によっては、この重複は約５０％～約１００％である。場合によっては、この重複は、約５０％～約６０％、約５０％～約７０％、約５０％～約８０％、約５０％～約９０％、約５０％～約１００％、約６０％～約７０％、約６０％～約８０％、約６０％～約９０％、約６０％～約１００％、約７０％～約８０％、約７０％～約９０％、約７０％～約１００％、約８０％～約９０％、約８０％～約１００％、または約９０％～約１００％である。場合によっては、この重複は、約５０％、約６０％、約７０％、約８０％、約９０％、または約１００％である。場合によっては、この重複は、少なくとも約５０％、約６０％、約７０％、約８０％、または約９０％である。場合によっては、この重複は、最大で約約６０％、約７０％、約８０％、約９０％、または約１００％であ

る。場合によっては、複数の出力配列は、本明細書に記載された方法を使用してデコードされる。例えば、複数の出力配列は、貪欲アルゴリズム、最尤（ＭＬ）アルゴリズム、または混合貪欲ＭＬアルゴリズムを使用してデコードされる。場合によっては、欠失、挿入、突然変異、またはこれらの任意の組み合わせの確率を計算することに少なくとも部分的に基づいて、複数の出力配列がデコードされる。

【０１１９】

本明細書では、データをエンコードするためのプラットフォームがさらに提供される。場合によっては、プラットフォームは、ハイブリッド有機インシリコ・プラットフォームを含む。場合によっては、プラットフォームは、コンピューティング・システム、シンセサイザ、またはこれらの組み合わせを含む。場合によっては、コンピューティング・システムは、少なくとも１つのプロセッサと、動作を実行するために少なくとも１つのプロセッサによって実行可能な命令とを備える。コンピューティング・システムまたは少なくとも１つのプロセッサは、本明細書で提供されるものであってよい。場合によっては、コンピューティング・システムは、分散コンピューティング・システムを含む。場合によっては、コンピューティング・システムは、クラウド・コンピューティング・システムを含む。クラウド・コンピューティング・システムは、プライベート・クラウド、パブリック・クラウド、ハイブリッド・クラウド、マルチクラウド、またはこれらの任意の組み合わせを含むことができる。クラウド・コンピューティング・システムは、サービスとしてのインフラストラクチャ（ＩａａＳ）、サービスとしてのプラットフォーム（ＰａａＳ）、サービスとしてのソフトウェア（ＳａａＳ）、またはこれらの任意の組み合わせを含むことができる。場合によっては、これらの動作は、本明細書において提供されるコードブックなどのコードブックを含んでいる内部コーデックを生成することを含む。場合によっては、コードブックは、核酸合成、後処理、ストレージ、またはシーケンシングに関連する１つまたは複数の制約などの、１つまたは複数の制約に対して最適化される。場合によっては、核酸合成は、電気化学的合成、酵素合成、ホスホラミダイト合成、インクジェット印刷、またはこれらの任意の組み合わせを含む。場合によっては、核酸合成に関連する１つまたは複数の制約は、挿入、欠失、または突然変異などの合成エラーを含む。場合によっては、後処理は、ライゲーション、切断、ハイブリダイゼーション、変性、固体支持体への固定、伸長、エラー訂正、濃縮、分離、精製、および増幅のうちの１つまたは複数を含む。場合によっては、ストレージはコールド・データ・ストレージを含む。コールド・データ・ストレージは、一般に、例えば、めったにアクセスされない核酸内のデータのストレージを指すことができる。コールド・データ・ストレージは、頻繁にアクセスされるデータを指す「ホット・ストレージ」の反対であることができる。一部の例では、ストレージはホット・ストレージを含み、ホット・ストレージでは、核酸に格納されたデータが頻繁にアクセスされる。場合によっては、ストレージは、液相または固相での核酸ストレージを含む。一部の例では、ストレージに関連する１つまたは複数の制約は、温度（例えば、室温）、湿度、圧力、塩分濃度、ｐＨ、濃度、時間、光、ＵＶ、O_2、またはこれらの任意の組み合わせを含む。場合によっては、シーケンシングは、次世代シーケンシング、並列シーケンシング、単一分子リアルタイム・シーケンシング、ナノポア・シーケンシング、合成によるシーケンシング、サンガー・シーケンシング、またはこれらの任意の組み合わせを含む。

【０１２０】

場合によっては、コードブックは、塩基順序（例えば、［Ａ，Ｔ，Ｃ，Ｇ］など）を使用して生成される。場合によっては、コードブックは、塩基順序に基づいて生成されたコードワードを含む。場合によっては、塩基順序は、既定の塩基転移を含む。場合によっては、これらの動作は、本明細書において提供される方法を使用し、内部コーデックを適用して、バイナリ・データを複数のポリヌクレオチド配列としてエンコードすることを含む。

【０１２１】

場合によっては、シンセサイザは、複数のポリヌクレオチド配列を含む複数のポリヌク

レオチドを生成する。場合によっては、シンセサイザは、合成、ライゲーション、アセンブリ、またはこれらの任意の組み合わせによって、複数のポリヌクレオチド配列を生成する。合成の方法は、本明細書において提供される方法（例えば、ホスホラミダイト、酵素など）であってよい。場合によっては、コンピューティング・システムからの命令は、シンセサイザに複数のポリヌクレオチドをさらに生成させる。場合によっては、シンセサイザは、ポリヌクレオチドの合成に使用される。場合によっては、シンセサイザは、ポリヌクレオチドのアセンブリに使用される。場合によっては、代替のアセンブリ・モジュールが、ポリヌクレオチドのアセンブリに使用される。一部の例では、アセンブリは、オーバーラップ伸長ポリメラーゼ連鎖反応（ＰＣＲ：ｐｏｌｙｍｅｒａｓｅ　ｃｈａｉｎ　ｒｅａｃｔｉｏｎ）、ポリメラーゼ・サイクリング・アセンブリ（ＰＣＡ：ｐｏｌｙｍｅｒａｓｅ　ｃｙｃｌｉｎｇ　ａｓｓｅｍｂｌｙ）、粘着末端ライゲーション、バイオブリック・アセンブリ、ゴールデン・ゲート・アセンブリ、ギブソン・アセンブリ、リコンビナーゼ・アセンブリ、リガーゼ・サイクリング反応、テンプレート指向ライゲーション、またはこれらの任意の組み合わせを含む。場合によっては、シンセサイザおよびアセンブリ・モジュールは、流体連通するか、電子通信するか、またはこれらの組み合わせである。

【０１２２】

プラットフォームは、シーケンサをさらに備えることができる。シーケンサは、本明細書において提供されるシーケンシング方法を実行するためのシステムおよびデバイス、または当技術分野において知られたシステムおよびデバイスを含んでよい。場合によっては、シーケンサは、複数のポリヌクレオチドをシーケンシングして、複数の出力配列を生成する。シーケンシングの方法は、本明細書において提供される方法であってよい。場合によっては、命令は、コンピューティング・システムに複数の出力配列をさらに受信させる。場合によっては、コンピューティング・システムは、複数の出力配列をデコードすることを含む動作をさらに実行する。コンピューティング・システムは、本明細書において提供されるデコーディング方式を使用して、複数の出力配列または任意の他のポリヌクレオチド配列をデコードすることができる。場合によっては、複数の出力配列は、貪欲アルゴリズム、最尤（ＭＬ）アルゴリズム、または混合貪欲ＭＬアルゴリズムを使用してデコードされる。場合によっては、欠失、挿入、突然変異、またはこれらの任意の組み合わせの確率を計算することに少なくとも部分的に基づいて、複数の出力配列がデコードされる。プラットフォームは、さらにストレージ・ユニットを備えることができる。場合によっては、ストレージ・ユニットは、複数のポリヌクレオチドを格納する。ポリヌクレオチドは、液体として溶液に格納されるか、または固体として乾燥されてよい。ポリヌクレオチドは、本明細書において提供される基板などの基板に格納されてよい。場合によっては、コンピューティング・システムの命令は、シンセサイザ、シーケンサ、ストレージ・ユニット、またはこれらの任意の組み合わせの間で、複数のポリヌクレオチドの転送を引き起こす。

【０１２３】

デノボ・ポリヌクレオチド合成

【０１２４】

本明細書では、基板上のポリヌクレオチドの合成のためのシステムおよび方法が提供される。場合によっては、内部コーデックからの最終的なオリゴヌクレオチド・プールが合成される。場合によっては、（図１２に示されているように）エンコーディング方式からの複数のポリヌクレオチドを含むライブラリが合成される（１２２５）。一部の例では、エンコード方式からの複数のポリヌクレオチドを含むライブラリは、複数のプールのうちのプールをエンコードする。一部の例では、エンコード方式からの複数のポリヌクレオチドを含むライブラリは、インデックス・プールをエンコードする。場合によっては、方法は電気化学的脱保護の使用を含む。場合によっては、基板は柔軟な基板である。場合によっては、少なくとも10^{10}、10^{11}、10^{12}、10^{13}、10^{14}、または10^{15}個の塩基が１日で合成される。場合によっては、１日で少なくとも10×10^{8}、10×10^{9}、10×10^{10}、10×10^{11}、または10×10^{12}個のポリヌクレオチド

が合成される。場合によっては、合成される各ポリヌクレオチドは、少なくとも２０、５０、１００、２００、３００、４００、または５００個の核酸塩基を含む。場合によっては、これらの塩基は、１００、２００、３００、４００、５００、１０００、２０００、５０００、１００００、１５０００、２００００塩基に約１つ未満の合計平均エラー率で合成される。場合によっては、これらのエラー率は、合成されたポリヌクレオチドの少なくとも５０％、６０％、７０％、８０％、９０％、９５％、９８％、９９％、９９．５％、またはそれ以上に関する。場合によっては、これらの、合成されたポリヌクレオチドの少なくとも９０％、９５％、９８％、９９％、９９．５％、またはそれ以上は、これらのポリヌクレオチドがエンコードする既定の配列と異ならない。場合によっては、本明細書に記載された方法およびシステムを使用して基板上で合成されたポリヌクレオチドのエラー率は、約２００分の１未満、約１，０００分の１未満、約２，０００分の１未満、約３，０００分の１未満、または約５，０００分の１未満である。エラー率の個々の種類は、基板上で合成されたポリヌクレオチドに対する不一致、欠失、挿入、および／または置換を含む。「エラー率」という用語は、既定のポリヌクレオチド配列の集計に対する合成されたポリヌクレオチドの総量の比較を指す。場合によっては、本明細書において提供される合成方法（例えば、インクジェットベースの合成方法）は、約０．１％の欠失率、０．１％の突然変異率（または置換率）、０．０５％の挿入率、またはこれらの任意の組み合わせよりも良好な結果を有する。例えば、合成されたポリヌクレオチドは、約０．００１％、０．００５％、０．０１％、０．０５％、０．１％以下の欠失率、約０．００１％、０．００５％、０．０１％、０．０５％、もしくは０．１％以下の突然変異率、約０．００１％、０．００５％、０．０１％、もしくは０．０５％以下の挿入率、またはこれらの任意の組み合わせを有してよい。場合によっては、本明細書において開示される合成されたポリヌクレオチドは、１２～２５塩基の鎖を含む。場合によっては、この鎖は、１０、１１、１２、１３、１４、１５、１６、１７、１８、１９、２０、２１、２２、２３、２４、２５、２６、２７、２８、２９、３０、３１、３２、３３、３４、３５、３６、３７、３８、３９、４０、４１、４２、４３、４４、４５、４６、４７、４８、４９、５０塩基以上を含む。

【０１２５】

本明細書では、ポリヌクレオチド合成において使用される化学反応が、電気化学を使用して制御される、方法、システム、デバイス、および構成が説明される。電気化学反応は、場合によっては、光、熱、放射線、または電気などの、任意のエネルギー源によって制御される。例えば、表面上の離散的な遺伝子座のすべてまたは一部として化学反応を制御するために、電極が使用される。場合によっては、電極に電位を印加して、ポリヌクレオチド合成における１つまたは複数の化学ステップを制御することによって、電極が帯電される。場合によっては、これらの電極はアドレス指定可能である。本明細書に記載された任意の数の化学ステップは、場合によっては、１つまたは複数の電極を使用して制御される。電気化学反応は、酸化、還元、酸／塩基化学反応、または電極によって制御される他の反応を含んでよい。場合によっては、電極は、化学変換の試薬として使用される電子または陽子を生成する。電極は、場合によっては、酸などの試薬を直接生成する。場合によっては、酸は陽子である。電極は、場合によっては、塩基などの試薬を直接生成する。酸または塩基は、多くの場合、例えば反応溶液のｐＨを調整することによって、保護基を切断するか、またはさまざまなポリヌクレオチド合成反応の動態に影響を与えるために使用される。電気化学的に制御されたポリヌクレオチド合成反応は、場合によっては、酸化還元活性金属または他の酸化還元活性有機材料を含む。場合によっては、これらの電気化学反応では、金属触媒または有機触媒が使用される。場合によっては、キノンの酸化から酸が生成される。

【０１２６】

化学反応の制御は、試薬の電気化学的生成に限定されず、化学反応性は、電極によって生成される電場（または勾配）を介する基板または試薬に対する生物物理学的変化を通じて、間接的に影響を受けてよい。場合によっては、基板は核酸を含むが、これに限定され

ない。場合によっては、特定の試薬または基板を電極または表面から遠ざけるか、または電極または表面に引き付ける電場が生成される。そのような電場は、場合によっては、１つまたは複数の電極への電位の印加によって生成される。例えば、負に帯電した核酸は、負に帯電した電極表面から遠ざけられる。局所電場によって引き起こされるポリヌクレオチドまたは他の試薬のそのような反発力または引力は、場合によっては、合成デバイスまたは構造体の領域内外へのポリヌクレオチドまたは他の試薬の移動をもたらす。場合によっては、電極は電場を生成し、ポリヌクレオチドを合成表面、構造体、またはデバイスから遠ざける。場合によっては、電極は電場を生成し、ポリヌクレオチドを合成表面、構造体、またはデバイスに引き付ける。場合によっては、陽子が正に帯電した表面から遠ざけられて、基板およびその一部との陽子の接触を制限する。場合によっては、反発力または引力は、試薬または基板が合成表面の特定の領域に入るのを許可またはブロックするために使用される。場合によっては、一方または両方の成分の近傍に電場を印加することによって、ヌクレオシド・モノマーがポリヌクレオチド鎖に接触するのが妨げられる。そのような配置は、特定の試薬のゲート制御を可能にし、試薬および／または基板の間の接触の濃度または速度が制御されている場合に、基を保護する必要性をなくすことができる。場合によっては、保護されていないヌクレオシド・モノマーが、ポリヌクレオチド合成に使用される。代替として、一方または両方の成分の近傍における電場の印加は、ポリヌクレオチド鎖とのヌクレオシド・モノマーの接触を促進する。さらに、基板への電場の印加は、基板の反応性および配座を変更することができる。例示的な応用では、電極によって生成された電場は、隣接する遺伝子座でのポリヌクレオチドが相互作用するのを防ぐために使用される。場合によっては、基板はポリヌクレオチドであり、任意選択的に表面に取り付けられる。電場の印加は、場合によっては、ポリヌクレオチドの３次元構造を変更する。そのような変更は、らせん、ヘアピン、ループ、または他の３次元核酸構造などのさまざまな構造の折り畳みまたは展開を含む。そのような変更は、ウェル、チャネル、または他の構造内の核酸を操作するのに役立つ。場合によっては、二次構造を防ぐために核酸基板に電場が印加される。場合によっては、電場は、ポリヌクレオチド合成中の固体支持体へのリンカーまたは取り付けの必要性をなくす。

【０１２７】

本開示の基板上のポリヌクレオチド合成に適した方法は、ＤＮＡのホスホラミダイトベースの合成である。場合によっては、ホスホラミダイトベースの合成のための試薬は、ヌクレオシド・ホスホラミダイト、酸化剤、活性化剤、またはデブロッカーのうちの任意の１つまたは組み合わせを含み、あるいは溶媒はアセトニトリルを含む。場合によっては、ホスホラミダイトベースの合成法は、ホスホラミダイト構成要素と基板に結合されたヌクレオシドとの間のホスファイト・トリエステル結合を形成する結合ステップにおける、成長するポリヌクレオチド鎖へのホスホラミダイト構成要素、すなわちヌクレオシド・ホスホラミダイトの制御された追加を含む。場合によっては、ヌクレオシド・ホスホラミダイトは、活性化された基板に提供される。場合によっては、ヌクレオシド・ホスホラミダイトは、活性化剤と共に基板に提供される。場合によっては、ヌクレオシド・ホスホラミダイトは、基板に結合されたヌクレオシドよりも１．５、２、３、４、５、６、７、８、９、１０、１１、１２、１３、１４、１５、１６、１７、１８、１９、２０、２５、３０、３５、４０、５０、６０、７０、８０、９０、１００倍、またはそれ以上多く、基板に提供される。場合によっては、ヌクレオシド・ホスホラミダイトの追加は、無水環境内、例えば無水アセトニトリル中で実行される。結合ステップでのヌクレオシド・ホスホラミダイトの追加および結合の後に、基板が任意選択的に洗浄される。場合によっては、結合ステップは、任意選択的に基板へのヌクレオシド・ホスホラミダイトの追加間の洗浄ステップを伴って、さらに１回または複数回繰り返される。場合によっては、本明細書において使用されるポリヌクレオチド合成法は、１つ、２つ、３つ、またはそれ以上の逐次的結合ステップを含む。結合の前に、多くの場合、基板に結合されたヌクレオシドは、保護基の除去によって脱保護され、保護基は、重合を防ぐように機能する。保護基は、ポリヌクレオチド鎖の伸長を防ぐ任意の化学基を含んでよい。場合によっては、酸が存在する場合に

保護基が切断（または除去）される。場合によっては、塩基が存在する場合に保護基が切断される。場合によっては、保護基は、光、熱、または他のエネルギー源などの電磁放射を使用して除去される。場合によっては、保護基は、酸化反応または還元反応によって除去される。場合によっては、保護基はトリアリールメチル基を含む。場合によっては、保護基はアリール・エーテルを含む。場合によっては、保護はジスルフィドを含む。場合によっては、保護基は酸に不安定なシランを含む。場合によっては、保護基はアセタールを含む。場合によっては、保護基はケタールを含む。場合によっては、保護基はエノール・エーテルを含む。場合によっては、保護基はメトキシベンジル基を含む。場合によっては、保護基はアジドを含む。場合によっては、保護基は４，４’－ジメトキシトリチル（ＤＭＴ）である。場合によっては、保護基は炭酸ｔｅｒｔ－ブチルである。場合によっては、保護基はｔｅｒｔ－ブチル・エステルである。場合によっては、保護基は塩基に不安定な基を含む。

【０１２８】

結合の後に、ホスホラミダイト・ポリヌクレオチド合成法は、任意選択的にキャッピング・ステップを含む。キャッピング・ステップでは、成長するポリヌクレオチドが、キャッピング剤を使用して処理される。キャッピング・ステップは、通常、結合の後の未反応の基板に結合された５’－ＯＨ基をさらなる鎖伸長からブロックする役割を果たし、内部塩基欠失を伴うポリヌクレオチドの形成を防ぐ。さらに、１Ｈ－テトラゾールで活性化されたホスホラミダイトは、多くの場合、グアノシンのＯ６位とわずかに反応する。理論に縛られることなく、Ｉ２／水による酸化時に、この副生成物は、Ｏ６－Ｎ７の移動を介して脱プリン化される可能性がある。アプリン部位は、ポリヌクレオチドの最終的な脱保護の過程で切断される可能性があり、したがって、完全長生成物の収率を低下させる。Ｏ６修飾は、Ｉ２／水による酸化の前にキャッピング試薬を使用して処理することによって除去され得る。場合によっては、ポリヌクレオチド合成中にキャッピング・ステップを含めることによって、キャッピングのない合成と比較してエラー率が減少する。例として、キャッピング・ステップは、無水酢酸および１－メチルイミダゾールの混合物を使用して、基板に結合されたポリヌクレオチドを処理することを含む。キャッピング・ステップの後に、基板が任意選択的に洗浄される。

【０１２９】

ヌクレオシド・ホスホラミダイトの追加の後に、および任意選択的にキャッピングおよび１つまたは複数の洗浄ステップの後に、本明細書に記載された基板は、酸化され得る結合された成長する核酸を含む。酸化ステップは、亜リン酸トリエステルを酸化して、自然発生のリン酸塩ジエステル・インターヌクレオシド結合の保護された前駆体である四配位リン酸トリエステルにする。場合によっては、亜リン酸トリエステルは電気化学的に酸化される。場合によっては、成長するポリヌクレオチドの酸化は、ヨウ素および水を使用して処理することによって、任意選択的に、ピリジン、ルチジン、またはコリジンなどの弱い塩基が存在する場合に、実現される。酸化は、ｔｅｒｔ－ブチル・ヒドロペルオキシドまたは（１Ｓ）－（＋）－（１０－カンファースルホニル）－オキサジリジン（ＣＳＯ）を使用して無水条件下で行われることがある。一部の方法では、酸化後にキャッピング・ステップが実行される。持続することがある酸化からの残留水が、その後の結合を抑制する可能性があるため、第２のキャッピング・ステップは基板の乾燥を可能にする。酸化後に、基板および成長するポリヌクレオチドが、任意選択的に洗浄される。場合によっては、酸化のステップは、ポリヌクレオチド・ホスホロチオエートを得るための硫化ステップに置き換えられ、硫化後に任意のキャッピング・ステップが実行され得る。多くの試薬は、３－（ジメチルアミノメチリデン）アミノ）－３Ｈ－１，２，４－ジチアゾール－３－チオン、ＤＤＴＴ、３Ｈ－１，２－ベンゾジチオール－３－オン１，１－ジオキシド（ボカージュ試薬としても知られている）、およびＮ，Ｎ，Ｎ’，Ｎ’－テトラエチルチウラム・ジスルフィド（ＴＥＴＤ）を含むが、これらに限定されない、効率的な硫黄移動の能力がある。

【０１３０】

結合を通じてヌクレオシド組み込みのその後のサイクルが発生するために、基板に結合された成長するポリヌクレオチドの保護された５’末端（または合成が５’から３’の方向に行われる場合は、３’末端）が除去され、一次水酸基が次のヌクレオシド・ホスホラミダイトと反応できるようにする。場合によっては、保護基はＤＭＴであり、ジクロロメタン中のトリクロロ酢酸によってデブロッキングが発生する。場合によっては、保護基はＤＭＴであり、電気化学的に生成された陽子によってデブロッキングが発生する。脱トリチル化を長時間行うか、または推奨された酸溶液よりも強い酸溶液を使用して行うことは、固体支持体に結合されたポリヌクレオチドの脱プリン化の増加につながることがあり、したがって、所望の完全長生成物の収率を低下させる。本明細書に記載された方法および構成は、望ましくない脱プリン化反応を制限する制御されたデブロッキング条件を提供する。場合によっては、デブロッキング後に、基板に結合されたポリヌクレオチドが洗浄される。場合によっては、デブロッキング後の効率的な洗浄は、合成されたポリヌクレオチドが低いエラー率を有することに寄与する。

【０１３１】

本明細書に記載された基板上のポリヌクレオチドの合成のための方法は、表面、リンカー、またはすでに脱保護されたモノマーのいずれかと連結するための、基板の特徴の表面への保護されたモノマーの適用、その後に適用される保護されたモノマーと反応できるようにするための、適用されたモノマーの脱保護、および連結のための別の保護されたモノマーの適用という、一連のステップの反復を含んでよい。１つまたは複数の中間ステップは、酸化および／または硫化を含む。場合によっては、１つまたは複数の洗浄ステップが、ステップのうちの１つまたはすべてに先行するか、またはそれらの後に続く。

【０１３２】

本明細書に記載された基板上のポリヌクレオチドの合成のための方法は、酸化ステップを含んでよい。例えば、方法は、表面、リンカー、またはすでに脱保護されたモノマーのいずれかと連結するための、基板の特徴の表面への保護されたモノマーの適用、その後に適用される保護されたモノマーと反応できるようにするための、適用されたモノマーの脱保護、連結のための別の保護されたモノマーの適用、ならびに酸化および／または硫化という、一連のステップの反復を含む。場合によっては、１つまたは複数の洗浄ステップが、ステップのうちの１つまたはすべてに先行するか、またはそれらの後に続く。

【０１３３】

本明細書に記載された基板上のポリヌクレオチドの合成のための方法は、表面、リンカー、またはすでに脱保護されたモノマーのいずれかと連結するための、基板の特徴の表面への保護されたモノマーの適用、その後に適用される保護されたモノマーと反応できるようにするための、適用されたモノマーの脱保護、ならびに酸化および／または硫化という、一連のステップの反復をさらに含んでよい。場合によっては、１つまたは複数の洗浄ステップが、ステップのうちの１つまたはすべてに先行するか、またはそれらの後に続く。

【０１３４】

本明細書に記載された基板上のポリヌクレオチドの合成のための方法は、表面、リンカー、またはすでに脱保護されたモノマーのいずれかと連結するための、基板の特徴の表面への保護されたモノマーの適用、ならびに酸化および／または硫化という、一連のステップの反復をさらに含んでよい。場合によっては、１つまたは複数の洗浄ステップが、ステップのうちの１つまたはすべてに先行するか、またはそれらの後に続く。

【０１３５】

本明細書に記載された基板上のポリヌクレオチドの合成のための方法は、表面、リンカー、またはすでに脱保護されたモノマーのいずれかと連結するための、基板の特徴の表面への保護されたモノマーの適用、その後に適用される保護されたモノマーと反応できるようにするための、適用されたモノマーの脱保護、ならびに酸化および／または硫化という、一連のステップの反復をさらに含んでよい。場合によっては、１つまたは複数の洗浄ステップが、ステップのうちの１つまたはすべてに先行するか、またはそれらの後に続く。

【０１３６】

場合によっては、ポリヌクレオチドは、光不安定性保護基を使用して合成され、表面に生成されたヒドロキシル基が、光不安定性保護基によってブロックされる。フォトリソグラフィ・マスクなどを介して表面が紫外線にさらされたときに、遊離ヒドロキシル基のパターンが表面に生成され得る。これらのヒドロキシル基は、ホスホラミダイトの化学的性質に従って、光保護されたヌクレオシド・ホスホラミダイトと反応することができる。第２のフォトリソグラフィ・マスクが適用されることが可能であり、表面が紫外線にさらされて、ヒドロキシル基の第２のパターンを生成し、それに続いて、５’光保護されたヌクレオシド・ホスホラミダイトと結合することができる。同様に、パターンが生成されることが可能であり、オリゴマー鎖が伸ばされ得る。理論に縛られることなく、光切断可能な基の不安定性は、使用される溶媒の波長および極性に依存し、光切断の速度は、露光の持続時間および光の強度による影響を受け得る。この方法は、マスクのアライメントの精度、光保護基の除去の効率、およびホスホラミダイト結合ステップの収率などの、複数の要因を活用することができる。さらに、意図しない隣接部位への光の漏れが、最小限に抑えられ得る。スポットごとの合成されたオリゴマーの密度は、合成の表面上のリーダー・ヌクレオシドの投入を調整することによって、監視され得る。

【０１３７】

ポリヌクレオチド合成のための支持体を提供する本明細書に記載された基板の表面は、合成されたポリヌクレオチド鎖が表面から切断されることを可能にするように、化学的に修飾されてよい。場合によっては、ポリヌクレオチドが脱保護されるのと同時に、ポリヌクレオチド鎖が切断される。場合によっては、ポリヌクレオチドが脱保護された後に、ポリヌクレオチド鎖が切断される。例示的な方式では、$(CH_3CH_2O)_3Si-(CH_2)_2-NH_2$などのトリアルコキシシリル・アミンが、基板の表面のＳｉＯＨ基と反応され、その後に、核酸鎖の成長が支援されるアミド結合および遊離ＯＨを作成するための、アミンを有する無水コハク酸との反応が続く。切断は、アンモニアまたはメチルアミンを使用するガス切断を含む。場合によっては、切断は、酸または塩基などの電気的に生成された試薬を使用するリンカー切断を含む。場合によっては、ポリヌクレオチドは、表面から解放された後に、格納された情報を抽出するためにシーケンシングされてデコードされる、より大きい核酸に組み立てられる。

【０１３８】

場合によっては、合成は酵素合成を含む。酵素合成は、本明細書に記載された表面で実行されてよい。場合によっては、酵素合成は鎖伸長酵素を含む。場合によっては、鎖伸長酵素はポリメラーゼである。場合によっては、ポリメラーゼは、テンプレートに依存しないポリメラーゼである。場合によっては、ポリメラーゼは、ＲＮＡポリメラーゼまたはＤＮＡポリメラーゼである。場合によっては、ポリメラーゼはＤＮＡポリメラーゼである。場合によっては、酵素ＤＮＡ合成は溶媒として水を使用し、試薬は酵素末端デオキシヌクレオチジル・トランスフェラーゼ（ＴｄＴ）またはデブロッカーである。場合によっては、ＤＮＡの酵素合成は、テンプレートに依存しないＤＮＡポリメラーゼ、すなわち自然発生のｄＮＴＰの結合を迅速に触媒するように進化したタンパク質である末端デオキシヌクレオチジル・トランスフェラーゼ（ＴｄＴ）を使用する。ＴｄＴはヌクレオチドを無差別に追加するため、ＴｄＴをつなぐこと、変種酵素を作成すること、および鎖の伸長を防ぐための可逆終端を含むヌクレオチドを使用することなどのさまざまな技術によって、制御されていない合成の継続が停止される。ＴｄＴ活性は約３７°Ｃで最大化され、水性環境内で酵素反応を行う。ＤＮＡポリメラーゼの例としては、ｐｏｌＡ、ｐｏｌＢ、ｐｏｌＣ、ｐｏｌＤ、ｐｏｌＹ、ｐｏｌＸ、逆転写酵素（ＲＴ：ｒｅｖｅｒｓｅ ｔｒａｎｓｃｒｉｐｔａｓｅｓ）、および高忠実度ポリメラーゼが挙げられるが、これらに限定されない。場合によっては、ポリメラーゼは修飾ポリメラーゼである。一部の実施形態では、ポリメラーゼは、Φ２９、Ｂ１０３、ＧＡ－１、ＰＺＡ、Φ１５、ＢＳ３２、Ｍ２Ｙ、Ｎｆ、Ｇ１、Ｃｐ－１、ＰＲＤ１、ＰＺＥ、ＳＦ５、Ｃｐ－５、Ｃｐ－７、ＰＲ４、ＰＲ５、ＰＲ７２２、Ｌ１７、ＴｈｅｒｍｏＳｅｑｕｅｎａｓｅ（登録商標）、９°Ｎｍ（商標）、Ｔｈｅｒｍｉｎａｔｏｒ（商標）ＤＮＡポリメラーゼ、Ｔｎｅ、Ｔｍａ、ＴｆＩ、Ｔｔｈ

、ＴＩｉ、Ｓｔｏｆｆｅｌフラグメント、Ｖｅｎｔ（商標）およびＤｅｅｐ　Ｖｅｎｔ（商標）ＤＮＡポリメラーゼ、ＫＯＤ　ＤＮＡポリメラーゼ、Ｔｇｏ、ＪＤＦ－３、Ｐｆｕ、Ｔａｑ、Ｔ７　ＤＮＡポリメラーゼ、Ｔ７　ＲＮＡポリメラーゼ、ＰＧＢ－Ｄ、ＵｌＴｍａ　ＤＮＡポリメラーゼ、大腸菌ＤＮＡポリメラーゼＩ、大腸菌ＤＮＡポリメラーゼＩＩＩ、古細菌ＤＰ１Ｉ／ＤＰ２　ＤＮＡポリメラーゼＩＩ、９°Ｎ　ＤＮＡポリメラーゼ、Ｔａｑ　ＤＮＡポリメラーゼ、Ｐｈｕｓｉｏｎ（登録商標）ＤＮＡポリメラーゼ、Ｐｆｕ　ＤＮＡポリメラーゼ、ＳＰ６　ＲＮＡポリメラーゼ、ＲＢ６９　ＤＮＡポリメラーゼ、鳥骨髄芽球症ウイルス（ＡＭＶ：Ａｖｉａｎ　Ｍｙｅｌｏｂｌａｓｔｏｓｉｓ　Ｖｉｒｕｓ）逆転写酵素、モロニー・マウス白血病ウイルス（ＭＭＬＶ：Ｍｏｌｏｎｅｙ　Ｍｕｒｉｎｅ　Ｌｅｕｋｅｍｉａ　Ｖｉｒｕｓ）逆転写酵素、ＳｕｐｅｒＳｃｒｉｐｔ（登録商標）ＩＩ逆転写酵素、ならびにＳｕｐｅｒＳｃｒｉｐｔ（登録商標）ＩＩＩ逆転写酵素を含む。一部の実施形態では、ポリメラーゼは、ＤＮＡポリメラーゼ１－Ｋｌｅｎｏｗフラグメント、Ｖｅｎｔポリメラーゼ、Ｐｈｕｓｉｏｎ（登録商標）ＤＮＡポリメラーゼ、ＫＯＤ　ＤＮＡポリメラーゼ、Ｔａｑポリメラーゼ、Ｔ７　ＤＮＡポリメラーゼ、Ｔ７　ＲＮＡポリメラーゼ、Ｔｈｅｒｍｉｎａｔｏｒ（商標）ＤＮＡポリメラーゼ、ＰＯＬＢポリメラーゼ、ＳＰ６　ＲＮＡポリメラーゼ、大腸菌ＤＮＡポリメラーゼＩ、大腸菌ＤＮＡポリメラーゼＩＩＩ、鳥骨髄芽球症ウイルス（ＡＭＶ）逆転写酵素、モロニー・マウス白血病ウイルス（ＭＭＬＶ）逆転写酵素、ＳｕｐｅｒＳｃｒｉｐｔ（登録商標）ＩＩ逆転写酵素、またはＳｕｐｅｒＳｃｒｉｐｔ（登録商標）ＩＩＩ逆転写酵素である。本明細書に記載された方法において使用されるポリメラーゼ分子は、ポリメラーゼ・シータ、ＤＮＡポリメラーゼ、またはヌクレオチド鎖を伸ばすことができる任意の酵素であることができる。一部の実施形態では、ポリメラーゼはｔｒｉ２９である。一部の実施形態では、ポリメラーゼは、末端リン酸基、例えば、三リン酸基の周囲で機能するポケットを有するタンパク質である。

【０１３９】

一部の実施形態では、酵素合成は、１、２、３、４、５、６、７、８、９、または１０のアミノ酸変異を有するＴｄＴを使用して、定義されたポリヌクレオチドを合成する。一部の実施形態では、記載された方法は、表面アクセス可能なアミノ酸残基に対する１、２、３、４、５、６、７、８、９、または１０のアミノ酸変異を有するＴｄＴを使用する。一部の実施形態では、ＴｄＴはＴｄＴの変形である。一部の実施形態では、ＴｄＴの変形は、システイン変異（例えば、ＮＴＴ－１）を含む。一部の実施形態では、ＴｄＴの変形は、ＮＴＴ－１、ＮＴＴ－２、またはＮＴＴ－３である。場合によっては、ＴｄＴの変形は、野生型ＴｄＴに対する少なくとも７０％、８０％、９０％、または９５％の配列同一性を含む。一部の実施形態では、酵素合成は、１、２、３、４、５、６、７、８、９、または１０のアミノ酸変異を有するポリメラーゼ・シータを使用して、定義されたポリヌクレオチドを合成することができる。一部の実施形態では、酵素合成は、表面アクセス可能なアミノ酸残基に対する１、２、３、４、５、６、７、８、９、または１０のアミノ酸変異を有するポリメラーゼ・シータを使用することができる。一部の実施形態では、ポリメラーゼ・シータは、ポリメラーゼ・シータの変形である。場合によっては、ポリメラーゼ・シータの変形は、野生型ポリメラーゼ・シータに対する少なくとも７０％、８０％、９０％、または９５％の配列同一性を含む。一部の実施形態では、ポリメラーゼ・シータはＰＯＬＱによってエンコードされる。

【０１４０】

本明細書に記載された酵素（例えば、ＴｄＴ）は、一部の実施形態では、１つまたは複数の非天然アミノ酸を含む。場合によっては、非天然アミノ酸は、リジン類似体、芳香族側鎖、アジド基、アルキン基、またはアルデヒドもしくはケトン基を含む。場合によっては、非天然アミノ酸は芳香族側鎖を含まない。一部の実施形態では、非天然アミノ酸は、Ｎ６－アジドエトキシ－カルボニル－Ｌ－リジン（ＡｚＫ）、Ｎ６－プロパルギルエトキシ－カルボニル－Ｌ－リジン（ＰｒａＫ）、Ｎ６－（プロパルギルオキシ）－カルボニル－Ｌ－リジン（ＰｒＫ）、ｐ－アジド－フェニルアラニン（ｐＡｚＦ）、ＢＣＮ－Ｌ－リ

ジン、ノルボルネン・リジン、ＴＣＯ－リジン、メチルテトラジン・リジン、アリロキシカルボニルリジン、２－アミノ－８－オキソノナン酸、２－アミノ－８－オキソオクタン酸、ｐ－アセチル－Ｌ－フェニルアラニン、ｐ－アジドメチル－Ｌ－フェニルアラニン（ｐＡＭＦ）、ｐ－ヨード－Ｌ－フェニルアラニン、ｍ－アセチルフェニルアラニン、２－アミノ－８－オキソノナン酸、ｐ－プロパルギルオキシフェニルアラニン、ｐ－プロパルギル－フェニルアラニン、３－メチル－フェニルアラニン、Ｌ－ドーパ、フッ素化フェニルアラニン、イソプロピル－Ｌ－フェニルアラニン、ｐ－アジド－Ｌ－フェニルアラニン、ｐ－アシル－Ｌ－フェニルアラニン、ｐ－ベンゾイル－Ｌ－フェニルアラニン、ｐ－ブロモフェニルアラニン、ｐ－アミノ－Ｌ－フェニルアラニン、イソプロピル－Ｌ－フェニルアラニン、Ｏ－アリルチロシン、Ｏ－メチル－Ｌ－チロシン、Ｏ－４－アリル－Ｌ－チロシン、４－プロピル－Ｌ－チロシン、ホスホノチロシン、トリ－Ｏ－アセチル－ＧｌｃＮＡｃｐ－セリン、Ｌ－ホスホセリン、ホスホノセリン、Ｌ－３－（２－ナフチル）アラニン、２－アミノ－３－（（２－（（３－（ベンジルオキシ）－３－オキソプロピル）アミノ）エチル）セラニル）プロパン酸、２－アミノ－３－（フェニルセラニル）プロパン酸、セレノシステイン、Ｎ６－（（（２－アジドベンジル）オキシ）カルボニル）－Ｌ－リジン、Ｎ６－（（（３－アジドベンジル）オキシ）カルボニル）－Ｌ－リジン、およびＮ６－（（（４－アジドベンジル）オキシ）カルボニル）－Ｌ－リジンから選択される。一部の実施形態では、本明細書に記載された酵素は、１つの他の酵素に融合される。例えば、ＴｄＴは、ヘリカーゼなどの他の酵素に融合される。

【０１４１】

ポリヌクレオチドの酵素合成中に、酵素または他の核酸（例えば、ポリメラーゼ）結合部分を１つまたは複数の塩基対部分、例えば修飾ヌクレオチドに結合させるために、さまざまなリンカーが使用されてよい。リンカーへのヌクレオチドまたは他の塩基対部分の結合は、化学結合法の当技術分野において知られた任意の手段によって実現されてよい。例えば、遊離アミン基を追加する塩基修飾を含むヌクレオチドは、本明細書において説明されるように、リンカーへの結合における使用を企図される。例えば、第１級アミンは、ヘテロ二官能性ポリエチレン・グリコール（ＰＥＧ：ｐｏｌｙｅｔｈｙｌｅｎｅ ｇｌｙｃｏｌ）リンカーと反応して、酵素活性部位に依然として適切に結合する可変長ＰＥＧリンカーを含むヌクレオチドを作成できるような方法で、塩基に連結されてよい。そのようなアミン含有ヌクレオチドの例としては、５－プロパルギルアミノ－ｄＮＴＰ、５－プロパルギルアミノ－ＮＴＰ、アミノ・アリル－ｄＮＴＰ、およびアミノ・アリル－ＮＴＰが挙げられる。一部の実施形態では、アミン含有ヌクレオチドは、ＰＥＧベースのリンカーとの結合に適している。ＰＥＧリンカーの長さは、例えば、１～１０００、１～５００、１～１１、１～１００、１～５０、または１～１０サブユニットの範囲で変化してよい。他の適切なリンカーの非限定的な例は、ポリ－Ｔおよびポリ－Ａオリゴヌクレオチド鎖（例えば、約１塩基～約１，０００塩基の長さの範囲にわたる）、ペプチド・リンカー（例えば、約１残基～約１，０００残基の長さの範囲にわたるポリグリシンまたはポリアラニン）、または炭素鎖リンカー（例えば、Ｃ６、Ｃ１２、Ｃ１８、Ｃ２４など）を含んでよいが、これらに限定されない。一部の実施形態では、リンカーは、Ｎ－ヒドロキシスクシンイミド・エステル（ＮＨＳ）基を含む。一部の実施形態では、リンカーはマレイミド基を含む。ヌクレオチドの接続は、ジスルフィドの形成（容易に切断可能な接続の形成）、アミドの形成、エステルの形成、タンパク質－リガンド結合（例えば、ビオチン－ストレプトアビジン結合）、アルキル化（例えば、置換ヨードアセトアミド試薬の使用）、あるいはアルデヒドおよびアミンまたはヒドラジンを使用する付加物の形成によって、実現され得る。一部の実施形態では、リンカーは、例えば、マルトース基、ビオチン基、Ｏ２－ベンジルシトシン基またはＯ２－ベンジルシトシン誘導体、Ｏ６－ベンジルグアニン基、またはＯ６－ベンジルグアニン誘導体を含む。リンカーの長さは、ヌクレオチド（または他の塩基対部分）および酵素（または他の核酸結合部分）の種類に応じて変化してよい。場合によっては、ヌクレオチドを酵素に接続するためのリンカーは、約０．１～１，０００ｎｍ、０．５～５００ｎｍ、０．５～４００ｎｍ、０．５～３００ｎｍ、０．５～２００

ｎｍ、０．５～１００ｎｍ、０．５～５０ｎｍ、０．６～５００ｎｍ、０．６～４００ｎｍ、０．６～３００ｎｍ、０．６～２００ｎｍ、０．６～１００ｎｍ、０．６～５０ｎｍ、１～５００ｎｍ、１～４００ｎｍ、１～３００ｎｍ、１～２００ｎｍ、１～１００ｎｍ、１．５～５００ｎｍ、１．５～４００ｎｍ、１．５～３００ｎｍ、１．５～２００ｎｍ、１．５～１００ｎｍ、１．５～５０ｎｍ、１～５０ｎｍ、５～５００ｎｍ、５～４００ｎｍ、５～３００ｎｍ、５～２００ｎｍ、５～１００ｎｍ、または５～５０ｎｍの持続長を有することができる。一部の実施形態では、化学リンカーは、酸切断可能なリンカーである。一部の実施形態では、化学リンカーは、酸切断可能なリンカーである。一部の実施形態では、化学リンカーは、光切断可能なリンカーである。一部の実施形態では、化学リンカーは、シリル・リンカー、アルキル・リンカー、ポリエーテル・リンカー、ポリスルホニル・リンカー、ポリスルホキシド・リンカー、およびこれらの任意の組み合わせからなる群から選択される。一部の実施形態では、リンカーは酵素によって切断される。一部の実施形態では、酵素は、プロテアーゼ、エステラーゼ、グリコシラーゼ、またはペプチダーゼである。一部の実施形態では、切断酵素は、ポリメラーゼにおける結合を切断する。一部の実施形態では、切断酵素は、連結されたヌクレオシドを直接切断する。

【０１４２】

本明細書に記載された表面は、ポリヌクレオチドの切断後に、ポリヌクレオチド合成の追加サイクルを支援するために再利用され得る。例えば、リンカーは、追加の処理／化学修飾なしで再利用され得る。場合によっては、リンカーは、基板表面またはポリヌクレオチドに非共有結合される。一部の実施形態では、リンカーは、表面からの切断後に、ポリヌクレオチドに取り付けられたままである。リンカーは、一部の実施形態では、エステル、アミド、ケタール、ベータ置換ケトン、複素環、または可逆的に切断され得る他の基などの可逆的な共有結合を含む。そのような可逆的な切断反応は、場合によっては、試薬の添加または除去によって、または電極によって制御される電気化学的プロセスによって、制御される。任意選択的に、化学リンカーまたは表面に結合された化学基は、反応性を回復し、そのようなリンカーまたは表面に結合された化学基上の不要な副生成物形成を除去するために、複数のサイクル後に再生成される。

【０１４３】

ポリヌクレオチド・ストレージのためのデバイス

【０１４４】

ポリヌクレオチドの合成されたライブラリは、デバイスに格納され得る。場合によっては、このデバイスは、ポリヌクレオチド・データ・ストレージ・システムを備える。場合によっては、ライブラリ・エンコーディング・プール（ｌｉｂｒａｒｉｅｓ　ｅｎｃｏｄｉｎｇ　ｐｏｏｌｓ）（例えば、複数のプールまたはインデックス・プール）が、コンパートメントに格納される。場合によっては、コンパートメントは、非限定的な例として、活性表面（例えば、遺伝子座）、チューブ、または任意の他の物理的格納の解決策を含む。一部の例では、コンパートメントはラベルでマークされる。一部の例では、ラベルは、バーコード、名前（例えば、顧客名、サンプルの種類など）、タイムスタンプ、格納されたオブジェクトのリスト、またはこれらの任意の組み合わせを含む。

【０１４５】

場合によっては、デジタル情報をＤＮＡに格納するためのデバイスは、１つまたは複数のコンパートメントを備える。場合によっては、１つまたは複数のコンパートメントの各々は、複数のポリヌクレオチドを含んでいるライブラリを備える。一部の実施例では、ライブラリは、１つまたは複数のオブジェクトに対応するデジタル情報を含むプール（例えば、本明細書に記載された複数のプールのうちのプール）をエンコードする。一部の実施例では、プールは、本明細書に記載されたような、プール記述子、１つまたは複数のプール・アイテム、および終了プール記述子を含む。一部の例では、プールは、本明細書において前述したように、約１ＧＢ～約１ＴＢのデジタル情報を含む。

【０１４６】

複数のポリヌクレオチドを格納するためのコンパートメントまたは構造体は、任意の形

状またはサイズであってよい。場合によっては、コンパートメントは、実質的に球形、チューブ状（図１７Ａ）、卵形、円錐形、立方体、直方体、円筒形、くさび形、六角柱、底面が正方形の角錐、底面が三角形の角錐、三角柱、トロイド、半球、らせん形、ハート型、または他の形状である。場合によっては、形状は、構造体が外部環境に対して開かれるか、または閉じられることを可能にするように構成される。場合によっては、そのような閉鎖は、溶接、シール、隔膜、または構造体の内外への気体もしくは他の物質の移動を制限するための他のメカニズムによって、容易にされる。場合によっては、コンパートメントは、核酸、流体、気体、または他の材料を構造体の内外に追加または除去するための穴、スロット、隔膜、バルブ、またはポートを含む。場合によっては、複数のポリヌクレオチドを格納するための構造体は、一緒に面一溶接されたキャップおよび本体を備える（図１７Ｂ）。場合によっては、複数のポリヌクレオチドを格納するためのコンパートメントは、取り外し可能なスクリューキャップを備える（図１７Ｃ）。場合によっては、構造体は隔膜を備える（図１７Ｄ）。場合によっては、構造体は、一方の半分が他方に挿入されたときにシールを形成する、２つの丸みを帯びた錠剤状の半分を備える（図１７Ｅ）。場合によっては、構造体は、密閉可能な蓋を有する実質的に平らなディスク・コンテナを備える（図１７Ｆ）。場合によっては、コンパートメントは、任意選択的に取り付けられた蓋を有するボックスを備える（図１７Ｇ）。一部の例では、形状は円筒または円盤である。一部の例では、円筒形または円盤形は、コンパートメントの自動処理および／または充填のために好ましい。

【０１４７】

場合によっては、１つまたは複数のコンパートメントの各々は、複数のポリヌクレオチドを格納するための媒体を備える。一部の例では、媒体は、固体、液体、気体、またはこれらの任意の組み合わせを含む。一部の例では、媒体は塩溶液を含む。一部の例では、塩とＤＮＡのモル比は、約２０：１～約２：１の範囲であってよい。一部の例では、モル比は、使用される塩の分子量、および組み合わせられる塩とＤＮＡの相対量に依存する。一部の例では、モル比は、塩の陽イオンとＤＮＡの負に帯電したリン酸基との間で計算される。一部の例では、塩溶液は、ＤＮＡ内のリン酸基に対するモル比が２０：１未満の塩の陽イオンを含む。一部の例では、塩溶液が乾燥されて、乾燥生成物を作り出す。場合によっては、塩溶液は、非限定的な例として、塩化カルシウム、硝酸カルシウム、炭酸カルシウム、リン酸カルシウム、塩化マグネシウム、硫酸マグネシウム、硝酸マグネシウム、炭酸マグネシウム、塩化ランタン、硝酸ランタン、炭酸ランタン、臭化ランタン、またはこれらの混合物を含む。場合によっては、塩溶液は、塩化バリウム（ＩＩ）二水和物、塩化カルシウム二水和物、無水塩化銅（ＩＩ）、三塩化ランタン、二塩化マグネシウム六水和物、塩化ナトリウム、または塩化ストロンチウム六水和物を含む。場合によっては、塩溶液の濃度は約０．０１ｎＭ～約０．１ｎＭである。

【０１４８】

場合によっては、１つまたは複数のコンパートメントの各々は通信する。場合によっては、１つまたは複数のコンパートメントの各々は、媒体を介して通信する。場合によっては、１つまたは複数のコンパートメントの各々は通信しない。場合によっては、１つまたは複数のコンパートメントの各々は、媒体を介して通信しない。

【０１４９】

場合によっては、デバイスは１つまたは複数の第２のコンパートメントをさらに備える。場合によっては、１つまたは複数の第２のコンパートメントの各々は、第２のライブラリを備える。一部の例では、第２のライブラリは、本明細書に記載されたインデックス・プールなどのインデックス・プールをエンコードする。場合によっては、１つまたは複数の第２のコンパートメントは、本明細書において前述したような媒体を備える。場合によっては、１つまたは複数の第２のコンパートメントは、１つまたは複数のコンパートメントと同じ媒体を備える。場合によっては、１つまたは複数の第２のコンパートメントは、１つまたは複数のコンパートメントと異なる媒体を含む。場合によっては、１つまたは複数の第２のコンパートメントの各々は、互いに、および／または１つもしくは複数のコン

パートメントと、（例えば、媒体を介して）通信する。場合によっては、１つまたは複数の第２のコンパートメントの各々は、互いに、および／または１つもしくは複数のコンパートメントと、通信しない。

【０１５０】

場合によっては、このデバイスは、表面を備えている固体支持体をさらに備える。したがって、本明細書では、そのような固体支持体に基づく核酸合成および格納のためのデバイスが説明され、固体支持体はさまざまな寸法を有する。場合によっては、固体支持体のサイズは、約４０～１２０ｍｍ×約２５～１００ｍｍである。場合によっては、固体支持体のサイズは、約８０ｍｍ×約５０ｍｍである。場合によっては、固体支持体の幅は、少なくともまたは約１０ｍｍ、２０ｍｍ、４０ｍｍ、６０ｍｍ、８０ｍｍ、１００ｍｍ、１５０ｍｍ、２００ｍｍ、３００ｍｍ、４００ｍｍ、５００ｍｍであるか、または５００ｍｍを超える。場合によっては、固体支持体の高さは、少なくともまたは約１０ｍｍ、２０ｍｍ、４０ｍｍ、６０ｍｍ、８０ｍｍ、１００ｍｍ、１５０ｍｍ、２００ｍｍ、３００ｍｍ、４００ｍｍ、５００ｍｍであるか、または５００ｍｍを超える。場合によっては、固体支持体は、少なくともまたは約１００ｍｍ^2、２００ｍｍ^2、５００ｍｍ^2、１，０００ｍｍ^2、２，０００ｍｍ^2、４，５００ｍｍ^2、５，０００ｍｍ^2、１０，０００ｍｍ^2、１２，０００ｍｍ^2、１５，０００ｍｍ^2、２０，０００ｍｍ^2、３０，０００ｍｍ^2、４０，０００ｍｍ^2、５０，０００ｍｍ^2、またはそれ以上の平面表面積を有する。場合によっては、固体支持体の厚さは、約５０ｍｍ～約２０００ｍｍ、約５０ｍｍ～約１０００ｍｍ、約１００ｍｍ～約１０００ｍｍ、約２００ｍｍ～約１０００ｍｍ、または約２５０ｍｍ～約１０００ｍｍである。固体支持体の非限定的な例示的厚さは、２７５ｍｍ、３７５ｍｍ、５２５ｍｍ、６２５ｍｍ、６７５ｍｍ、７２５ｍｍ、７７５ｍｍ、および９２５ｍｍを含む。場合によっては、固体支持体の厚さは、少なくともまたは約０．５ｍｍ、１．０ｍｍ、１．５ｍｍ、２．０ｍｍ、２．５ｍｍ、３．０ｍｍ、３．５ｍｍ、４．０ｍｍであるか、または４．０ｍｍを超える。

【０１５１】

本明細書では、２つ以上の固体支持体が組み立てられるデバイスが説明される。場合によっては、固体支持体は、より大きいユニット上で一緒にインターフェイスされる。インターフェイスは、固体支持体間の流体、電気信号、または交換の他の媒体の交換を含んでよい。このユニットは、任意の数のサーバ、コンピュータ、またはネットワーク・デバイスとのインターフェイスの能力がある。例えば、複数の固体支持体がラック・ユニットに統合され、ラック・ユニットは、便利にサーバ・ラックに挿入されるか、またはサーバ・ラックから取り外される。ラック・ユニットには、任意の数の固体支持体を備えてよい。場合によっては、ラック・ユニットは、少なくとも１、２、５、１０、２０、５０、１００、２００、５００、１０００、２０００、５０００、１０，０００、２０，０００、５０，０００、１００，０００個の、または１００，０００個を超える、固体支持体を備える。場合によっては、２つ以上の固体支持体が互いにインターフェイスされない。固体支持体上に存在する核酸（およびそれらの核酸に格納された情報）は、ラック・ユニットからアクセスされ得る。アクセスは、固体支持体からのポリヌクレオチドの除去、固体支持体上のポリヌクレオチドの直接分析、または核酸に格納された情報が操作または識別されることを可能にする任意の他の方法を含む。情報は、場合によっては、情報は、複数のラック、単一のラック、ラック内の単一の固体支持体、固体支持体の一部、または固体支持体上の単一の遺伝子座からアクセスされる。さまざまな場合において、アクセスは、核酸を、質量分析計、ＨＰＬＣ、シーケンシング機器、ＰＣＲサーモサイクラー、または核酸を操作するための他のデバイスなどの追加のデバイスとインターフェイスすることを含む。核酸情報へのアクセスは、場合によっては、固体支持体のすべてまたは一部からのポリヌクレオチドの切断によって実現される。切断は、場合によっては、化学試薬（アンモニアまたは他の試薬）、電位、放射線、熱、光、音響、または化学結合を操作できる他の形態のエネルギーへの曝露を含む。場合によっては、ポリヌクレオチドの近傍で１つまたは複数の電極を帯電させることによって、切断が発生する。場合によっては、紫外線の形態

での電磁放射が、ポリヌクレオチドの切断に使用される。場合によっては、ランプがポリヌクレオチドの切断に使用され、マスクが、表面への紫外線の露光位置を仲介する。場合によっては、レーザーがポリヌクレオチドの切断に使用され、シャッターの開かれた／閉じられた状態が、表面への紫外線の露光を制御する。場合によっては、核酸情報へのアクセス（ラック、固体支持体、試薬、核酸、または他のコンポーネントの除去／追加を含む）は、完全に自動化される。

【０１５２】

本明細書において説明されたような固体支持体は、活性化領域を含む。場合によっては、活性化領域は、核酸合成のための領域または遺伝子座を含む。場合によっては、活性化領域は、核酸ストレージのための領域または遺伝子座を含む。一部の例では、領域または遺伝子座は、１つまたは複数のコンパートメントを含む。一部の例では、領域または遺伝子座は、第２の１つまたは複数のコンパートメントを含む。場合によっては、領域はアドレス指定可能である。一部の例では、領域は、電極を介してアドレス指定可能である。

【０１５３】

活性化領域は、さまざまな寸法を含む。例えば、活性化領域の寸法は、約１ｍｍ～約５０ｍｍ×約１ｍｍ～約５０ｍｍである。場合によっては、活性化領域は、少なくともまたは約０．５、１、１．５、２、２．５、３、５、５、１０、１２、１４、１６、１８、２０、２５、３０、３５、４０、４５、５０、６０、７０、８０ｍｍの、または８０ｍｍを超える、幅を含む。場合によっては、活性化領域は、少なくともまたは約０．５、１、１．５、２、２．５、３、５、５、１０、１２、１４、１６、１８、２０、２５、３０、３５、４０、４５、５０、６０、７０、８０ｍｍの、または８０ｍｍを超える、高さを含む。

【０１５４】

本明細書では、固体支持体に基づく核酸合成および格納のためのデバイスが説明され、固体支持体は、合成または格納のための複数の部位（例えば、スポット）または位置を有する。場合によっては、固体支持体は、領域内の最大でまたは約１０，０００×１０，０００個の位置を含む。場合によっては、固体支持体は、領域内の約１０００～２０，０００×約１０００～２０，０００個の位置を含む。場合によっては、固体支持体は、領域内の少なくともまたは約１０、３０、５０、７５、１００、２００、３００、４００、５００、１０００、２０００、３０００、４０００、５０００、６０００、７０００、８０００、９０００、１０，０００、１２，０００、１４，０００、１６，０００、１８，０００、２０，０００個の位置×少なくともまたは約１０、３０、５０、７５、１００、２００、３００、４００、５００、１０００、２０００、３０００、４０００、５０００、６０００、７０００、８０００、９０００、１０，０００、１２，０００、１４，０００、１６，０００、１８，０００、２０，０００の個の位置を含む。場合によっては、面積は、最大で０．２５、０．５、０．７５、１．０、１．２５、１．５、または２．０平方インチである。場合によっては、固体支持体は、少なくともまたは約０．１、０．２、０．２５、０．３、０．４、０．５、１．０、１．５、２．０、２．５、３．０、３．５、４．０、４．５、５、６、７、８、９、１０ｕｍの、または１０ｕｍを超える、ピッチを有する遺伝子座を含む。場合によっては、固体支持体は、約５ｕｍのピッチを有する遺伝子座を含む。場合によっては、固体支持体は、約２ｕｍのピッチを有する遺伝子座を含む。場合によっては、固体支持体は、約１ｕｍのピッチを有する遺伝子座を含む。場合によっては、固体支持体は、約０．２ｕｍのピッチを有する遺伝子座を含む。場合によっては、固体支持体は、約０．２ｕｍ～約１０ｕｍ、約０．２～約８ｕｍ、約０．５～約１０ｕｍ、約１ｕｍ～約１０ｕｍ、約２ｕｍ～約８ｕｍ、約３ｕｍ～約５ｕｍ、約１ｕｍ～約３ｕｍ、または約０．５ｕｍ～約３ｕｍのピッチを有する遺伝子座を含む。場合によっては、固体支持体は、約０．１ｕｍ～約３ｕｍのピッチを有する遺伝子座を含む。

【０１５５】

本明細書において説明されるような核酸合成または格納のための固体支持体は、データの格納のための大きい容量を含む。例えば、固体支持体の容量は、少なくともまたは約１

、２、３、４、５、６、７、８、９、１０、２０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００メガバイトであるか、または１０００メガバイトを超える。場合によっては、固体支持体の容量は、約１～１０、１～５０、１～１００、１～５００、１～１０００、１０～５０、１０～１００、１０～５００、１０～１０００、５０～１００、５０～５００、５０～１０００、１００～５００、１００～１０００、２００～５００、２００～１０００、５００～１０００、または約８００～１０００メガバイトである。例えば、固体支持体の容量は、少なくともまたは約１、２、３、４、５、６、７、８、９、１０、２０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００ギガバイトであるか、または１０００ギガバイトを超える。場合によっては、固体支持体の容量は、約１～１０、１～５０、１～１００、１～５００、１～１０００、１０～５０、１０～１００、１０～５００、１０～１０００、５０～１００、５０～５００、５０～１０００、１００～５００、１００～１０００、２００～５００、２００～１０００、５００～１０００、または約８００～１０００ギガバイトである。例えば、固体支持体の容量は、少なくともまたは約１、２、３、４、５、６、７、８、９、１０、２０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００テラバイトであるか、または１０００テラバイトを超える。場合によっては、固体支持体の容量は、約１～１０、１～５０、１～１００、１～５００、１～１０００、１０～５０、１０～１００、１０～５００、１０～１０００、５０～１００、５０～５００、５０～１０００、１００～５００、１００～１０００、２００～５００、２００～１０００、５００～１０００、または約８００～１０００テラバイトである。例えば、固体支持体の容量は、少なくとも、または約１、２、３、４、５、６、７、８、９、１０、２０、５０、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００ペタバイトであるか、または１０００ペタバイトを超える。場合によっては、固体支持体の容量は、約１～１０、１～５０、１～１００、１～５００、１～１０００、１０～５０、１０～１００、１０～５００、１０～１０００、５０～１００、５０～５００、５０～１０００、１００～５００、１００～１０００、２００～５００、２００～１０００、５００～１０００、または約８００～１０００ペタバイトである。場合によっては、固体支持体の容量は、約１００ペタバイトである。

【０１５６】

場合によっては、データは、液滴としてのパケットの配列として格納される。一部の例では、パケットの配列は、アドレス指定可能なパケットである。一部の例では、パケットは、電極を使用してアドレス指定可能である。場合によっては、データは、スポット上の液滴としてのパケットの配列として格納される。場合によっては、データは、乾燥したウェルとしてのパケットの配列として格納される。場合によっては、配列は、少なくともまたは約１、２、３、４、５、６、７、８、９、１０、２０、５０、１００、２００ギガバイトの、または２００ギガバイトを超える、データを含む。場合によっては、配列は、少なくともまたは約１、２、３、４、５、６、７、８、９、１０、２０、５０、１００、２００テラバイトの、または２００テラバイトを超える、データを含む。場合によっては、情報のアイテムが、データのバックグラウンドに格納される。例えば、情報のアイテムは、約１０～約１００メガバイトのデータに関してエンコードし、１ペタバイトのバックグラウンド・データに格納される。場合によっては、情報のアイテムは、少なくともまたは約１、１０、２０、３０、４０、５０、６０、７０、８０、９０、１００、１５０、２００、３００、４００、５００メガバイトの、または５００メガバイトを超える、データに関してエンコードし、１、１０、２０、３０、４０、５０、６０、７０、８０、９０、１００、１５０、２００、３００、４００、５００ペタバイトの、または５００ペタバイトを超える、バックグラウンド・データに格納される。

【０１５７】

本明細書では、固体支持体に基づく核酸合成および格納のためのデバイスが提供され、合成後に、ポリヌクレオチドが１つまたは複数の液滴としてパケットに収集される。場合

によっては、ポリヌクレオチドは、１つまたは複数の液滴としてパケットに収集されて格納される。場合によっては、液滴の数は、少なくともまたは約１、１０、２０、５０、１００、２００、３００、５００、１０００、２５００、５０００、７５０００、１０，０００、２５，０００、５０，０００、７５，０００、１００，０００、１００万、５００万、１０００万、２５００万、５，０００万、７，５００万、１億、２億５，０００万、５億、７億５，０００万個であるか、または７億５，０００万個を超える。場合によっては、液滴の量は、５、１０、１５、２０、２５、３０、３５、４０、４５、５０、５５、６０、６５、７０、７５、８０、８５、９０、９５、１００ｕｍの直径、または１００ｕｍ（マイクロメートル）を超える直径を含む。場合によっては、液滴の量は、１～１００ｕｍ、１０～９０ｕｍ、２０～８０ｕｍ、３０～７０ｕｍ、または４０～５０ｕｍの直径を含む。

【０１５８】

場合によっては、パケットに収集されるポリヌクレオチドは、類似する配列を含む。場合によっては、ポリヌクレオチドは、タグまたはバーコードとして使用される同一でない配列をさらに含む。例えば、同一でない配列は、固体支持体に格納されたポリヌクレオチドにインデックスを付けるため、および同一でない配列に基づいて後で特定のポリヌクレオチドを検索するために使用される。例示的なタグまたはバーコードの長さは、約１、２、３、４、５、６、７、８、９、１０、１５、２０、２５、またはそれ以上の塩基の長さを含むが、これらに限定されない、バーコード配列を含む。場合によっては、タグまたはバーコードは、少なくともまたは約１０、５０、７５、１００、２００、３００、４００塩基対、または４００を超える塩基対の長さを含む。

【０１５９】

本明細書では、固体支持体に基づく核酸合成および格納のためのデバイスが提供され、ポリヌクレオチドが、冗長性を含むパケットに収集される。例えば、パケットは、各ポリヌクレオチドの約１００～約１０００個のコピーを含む。場合によっては、パケットは、各ポリヌクレオチドの少なくともまたは約５０、７５、１００、２００、３００、４００、５００、６００、７００、８００、９００、１０００、１２００、１４００、１６００、１８００、２０００個のコピー、または２０００個を超えるコピーを含む。場合によっては、パケットは、約１０００倍～約５０００倍の合成の冗長性を含む。合成の冗長性は、場合によっては、少なくともまたは約５００倍、１０００倍、１５００倍、２０００倍、２５００倍、３０００倍、３５００倍、４０００倍、５０００倍、６０００倍、７０００倍、８０００倍であるか、または８０００倍を超える。本明細書において説明されるような固体支持体に基づく方法を使用して合成されるポリヌクレオチドは、さまざまな長さを含む。場合によっては、ポリヌクレオチドは、合成されて固体支持体にさらに格納される。場合によっては、ポリヌクレオチドの長さは、約１００～約１０００塩基である。場合によっては、ポリヌクレオチドは、少なくともまたは約１０、２０、３０、４０、５０、６０、７０、８０、９０、１００、１２５、１５０、１７５、２００、２２５、２５０、２７５、３００、３２５、３５０、３７５、４００、４２５、４５０、４７５、５００、６００、７００、８００、９００、１０００、１１００、１２００、１３００、１４００、１５００、１６００、１７００、１８００、１９００、２０００塩基、または２０００を超える塩基の長さを含む。

【０１６０】

シーケンシング

【０１６１】

ポリヌクレオチドは、それらが合成または格納される表面から抽出および／または増幅される。構造体の表面からのポリヌクレオチドの抽出および／または増幅の後に、適切なシーケンシング技術が使用されて、ポリヌクレオチドをシーケンシングしてよい。場合によっては、ＤＮＡ配列は、基板上または構造体の特徴内で読み取られる。場合によっては、基板に格納されたポリヌクレオチドは、抽出され、任意選択的により長い核酸に組み立てられ、その後、シーケンシングされる。ポリヌクレオチドは、本明細書に記載されたシ

ステムおよび方法を使用して基板から抽出されてよい。

【０１６２】

合成されて本明細書に記載された構造体に格納されたポリヌクレオチドは、データをエンコードし、このデータは、合成されたポリヌクレオチドの配列を読み取り、その配列をコンピュータによって読み取り可能な表現（例えば、バイナリ・コードなどのシンボルの列）に変換することによって、取り出されるか、または解釈され得る。場合によっては、配列は組み立てを必要とし、組み立てステップは、ポリヌクレオチド配列段階またはデジタル配列段階である必要があることがある。

【０１６３】

本明細書では、構造体上で直接、および／または主要な構造体（例えば、合成構造体、格納構造体など）からの除去の後に、格納されたポリヌクレオチドをシーケンシングすることができるデバイスを備える検出システムが提供される。構造体が可撓性材料のオープンリールテープである場合、検出システムは、構造体を保持し、検出位置を通って前進させるためのデバイスと、テープのセクションが検出位置にあるときにそのセクションからから生じた信号を検出するために検出位置の近くに配置される検出器とを備える。場合によっては、信号はポリヌクレオチドの存在を示す。場合によっては、信号はポリヌクレオチドの配列（例えば、蛍光信号）を示す。場合によっては、連続的なテープ上のポリヌクレオチド内にエンコードされた情報は、テープが、コンピュータに動作可能に接続された検出器を通って連続的に運ばれるときに、コンピュータによって読み取られる。場合によっては、検出システムは、ポリヌクレオチド・シーケンシング・デバイス、ポリヌクレオチド配列に関連するデータの格納および取り出しのためのデータベース、ポリヌクレオチド配列のＤＮＡコードをバイナリ・コードなどのシンボルの列に変換するためのソフトウェア、バイナリ・コードを読み取るためのコンピュータ、またはこれらの任意の組み合わせを備えているコンピュータ・システムを含む。

【０１６４】

本明細書では、本明細書に記載されたデバイスに統合され得るシーケンシング・システムが提供される。シーケンシングのさまざまな方法は、当技術分野において周知であり、対象のポリヌクレオチド中の塩基の同一性が識別される「塩基呼び出し」を含む。場合によっては、本明細書に記載された方法、デバイス、構成、およびシステムを使用して合成されたポリヌクレオチドは、合成表面からの切断後にシーケンシングされる。場合によっては、シーケンシングは、ポリヌクレオチド合成中またはポリヌクレオチド合成と同時に発生し、塩基呼び出しは、成長するポリヌクレオチド鎖へのヌクレオシド・モノマーの伸長の直後または直前に発生する。塩基呼び出しの方法は、ポリメラーゼ触媒によるテンプレート鎖へ塩基の追加によって生成された電流／電圧の測定を含む。場合によっては、合成表面は、ポリメラーゼなどの酵素を含む。場合によっては、そのような酵素は、電極または合成表面につながれる。場合によっては、酵素は、末端デオキシヌクレオチジル・トランスフェラーゼまたはその変形を含む。

【０１６５】

場合によっては、基板表面から切断されたポリヌクレオチドまたは増幅されたポリヌクレオチドは、従来のシーケンシングまたは超並列シーケンシングなどの技術によって処理され得る。シーケンシングは、本分野において利用可能なさまざまな方法、例えば、１つまたは複数の鎖末端ヌクレオチドを組み込むことを含む方法、例えば、ＡｐｐｌｉｅｄＢｉｏｓｙｓｔｅｍｓのＳｅｑＳｔｕｄｉｏ（登録商標）Ｇｅｎｅｔｉｃ Ａｎａｌｙｚｅｒによって実行され得るサンガー・シーケンシング法によって、実行され得る。他の実施形態では、シーケンシングは、次世代シーケンシング（ＮＧＳ：Ｎｅｘｔ Ｇｅｎｅｒａｔｉｏｎ Ｓｅｑｕｅｎｃｉｎｇ）法、例えば、プライマー伸長とそれに続く半導体ベースの検出（例えば、Ｔｈｅｒｍｏ Ｆｉｓｈｅｒ ＳｃｉｅｎｔｉｆｉｃのＩｏｎ Ｔｏｒｒｅｎｔ（商標）システム）または蛍光検出（例えば、Ｉｌｌｕｍｉｎａシステム）を実行することを含むことができる。

【０１６６】

情報の取り出しのための方法およびシステム

【０１６７】

本明細書では、情報（例えば、デジタル情報）を取り出すための方法およびシステムが提供される。場合によっては、本明細書では、デコードするための方法およびシステムが提供される。場合によっては、方法およびシステムは、ポリヌクレオチド配列（例えば、ポリヌクレオチド、オリゴヌクレオチド、複数のポリヌクレオチドなど）をデコードする。場合によっては、ポリヌクレオチド配列は、本明細書に記載された方法を使用してエンコードされる。場合によっては、方法およびシステムは、内部コーデック、外部コーデック、またはこれらの組み合わせを含む。場合によっては、情報は、本明細書において前述したように、１つまたは複数のオブジェクトを含む。場合によっては、１つまたは複数のオブジェクトの各々は、本明細書において前述したように、約１ＧＢ～約１ＴＢである。場合によっては、１つまたは複数のオブジェクトは、本明細書に記載された情報のアイテムなどの情報のアイテムを含むが、これに限定されない。

【０１６８】

場合によっては、システムおよび方法は、ポリヌクレオチド配列（例えば、ポリヌクレオチド、オリゴヌクレオチド、複数のポリヌクレオチドなど）をデコードする。複数のポリヌクレオチドに格納されたデジタル情報を取り出すための例示的な方法が、図１３に示されている。そのような場合、複数のポリヌクレオチドは、図１２に示されている一般的な動作に従って複数のプールに分割されていてよい。場合によっては、複数のポリヌクレオチドに格納されたデジタル情報を取り出す方法は、図１３に示されている１つまたは複数の動作を含む。

【０１６９】

場合によっては、複数のポリヌクレオチドに格納されたデジタル情報を取り出すことは、インデックス・プールにアクセスすることを含む（１３００）。場合によっては、インデックス・プールにアクセスすることは、インデックス・プールをエンコードするライブラリを完全に、または部分的にシーケンシングすることを含む。一部の例では、インデックス・プールは、本明細書に記載されたシステムおよび方法を使用してライブラリ内でエンコードされる。一部の例では、インデックス・プールをエンコードするライブラリ内のポリヌクレオチドは、本明細書に記載されたシステムおよび方法を使用してシーケンシングされる。場合によっては、２つ以上のインデックス・プールがアクセスされる。場合によっては、２つ以上のライブラリ内のポリヌクレオチドがシーケンシングされる。場合によっては、シーケンシングされたライブラリは、メモリ・ストレージ・システム（例えば、フラッシュ・ドライブ）に一時的に格納される。場合によっては、シーケンシングされたライブラリは、インデックス・プールを取り出すためにデジタル情報に変換される。場合によっては、インデックス・プールは、メモリ・ストレージ・システム（例えば、フラッシュ・ドライブ）に一時的に格納される。場合によっては、インデックス・プール内のデジタル情報は、対象の１つまたは複数のオブジェクトを検索するために使用される。一部の例では、対象の１つまたは複数のオブジェクトは、１つまたは複数のオブジェクトをエンコードする複数のポリヌクレオチドを含むライブラリに格納される。一部の例では、対象の１つまたは複数のオブジェクトは、１つまたは複数のオブジェクトに関連付けられたメタデータを使用して検索される。場合によっては、インデックス・プールにアクセスすることによって、１つまたは複数のオブジェクトに対応する複数のプールを決定する。しかし、場合によっては、複数のプールのうちの１つまたは複数のプール内の１つまたは複数のオブジェクトは既知であってよく、インデックス・プールへのアクセスが必要とされなくてよい。

【０１７０】

場合によっては、対象の１つまたは複数のオブジェクトが、例えば、ストレージ・デバイス内のコンパートメントから取り出される。場合によっては、複数のポリヌクレオチドに格納されたデジタル情報を取り出すことは、複数のプール内の１つまたは複数のオブジェクトに対応する複数のポリヌクレオチドをシーケンシングすることを含む（１３０５）

。場合によっては、複数のポリヌクレオチドはライブラリ内にある。場合によっては、ライブラリは、本明細書において前述したように、デバイスのコンパートメント内にある。場合によっては、プールをエンコードするライブラリ内の複数のポリヌクレオチドは、本明細書に記載されたシステムおよび方法を使用してシーケンシングされる。場合によっては、プールは、本明細書に記載されたシステムおよび方法を使用してライブラリ内でエンコードされる。場合によっては、１つまたは複数のオブジェクトを取り出すために、２つ以上のコンパートメント内の複数のポリヌクレオチドがシーケンシングされる。

【０１７１】

場合によっては、複数のポリヌクレオチドに格納されたデジタル情報を取り出すことは、デコーディング方式を適用することをさらに含む（２１０）。場合によっては、デコーディング方式は、複数のプール内のデジタル情報をデコードする。場合によっては、デコーディング方式は、複数のポリヌクレオチドを含むシーケンシングされたライブラリに適用される。場合によっては、デコーディング方式は、内部コーデック、ＥＣＣ、またはこれらの組み合わせを含む。場合によっては、デコーディング方式は、複数のポリヌクレオチド配列をデコードして、デジタル情報（例えば、オブジェクト）を含む出力を生成する。場合によっては、デコーディング方式は、エンコーディング方式における動作を取り消すことを含む。一部の例では、動作は、データまたはデータの一部を分割すること、シャッフルすること、連結すること、転置すること、変換すること、複製すること、（例えば、インデックスを使用して）ラベル付けすること、あるいはこれらの任意の組み合わせを含む。

【０１７２】

複数のポリヌクレオチド配列をデコードして、データ（例えば、バイナリ・データ）を含む出力を生成する方法は、例えば、図２に概略的に示されている。場合によっては、複数のポリヌクレオチド配列をデコードする方法は、複数のポリヌクレオチド配列を決定することを含んでよい（２０５）。場合によっては、複数のポリヌクレオチド配列を決定することは、ヌクレオチドをシーケンシングすることを含む。場合によっては、ヌクレオチドは、本明細書に記載された方法を使用してシーケンシングされる。

【０１７３】

複数のヌクレオチドをシーケンシングした後に、エンコードされたデータ（例えば、１つまたは複数のオブジェクト）がデコードされる。場合によっては、複数のヌクレオチドは、非限定的な例として、図７に示されている概略図を使用してデコードされる。シーケンシングからの出力は、図７に示されているように、読み取りの順序付けられていないリスト（例えば、ポリヌクレオチド配列）を含む。

【０１７４】

場合によっては、シーケンシング後に、シーケンシングされた読み取りおよび／または順序付けられていない読み取りがクラスタ化される。場合によっては、内部コーデックを適用する前にクラスタ化が実行される。場合によっては、読み取りは、フレーム・インデックス、レーン・インデックス、またはこれらの組み合わせなどのインデックスに基づいてクラスタ化される。そのような場合、読み取りは、フレーム・インデックス、レーン・インデックス、またはこれらの組み合わせを取得するために部分的にデコードされる。場合によっては、クラスタ化は、本明細書において前述したように、ハッシュ関数を使用して実行される。場合によっては、本明細書において前述したように、エンコーディング方式においてハッシュを使用してポリヌクレオチド配列内の塩基が決定された場合に、ハッシュ関数が使用される。

【０１７５】

場合によっては、シーケンシングされた読み取りがアライメントされる。場合によっては、シーケンシングされポリヌクレオチドは、クラスタ化された後にアライメントされる。場合によっては、クラスタ化された読み取りがアライメントされる。場合によっては、内部コーデックを適用する前に読み取りがアライメントされる。場合によっては、アラインメントは、アラインメント・アルゴリズムを使用して読み取り（例えば、核酸またはポ

リヌクレオチド配列）のコンセンサスを分析することを含む。一部の例では、アライメント・アルゴリズムは、ペアワイズ・アライメント・アルゴリズム、マルチシーケンス・アライメント・アルゴリズム、またはこれらの組み合わせを含む。

【０１７６】

場合によっては、ペアワイズ・アライメント・アルゴリズムは、読み取りごとに位置を初期化することを含む。初期化は、ポリヌクレオチド配列を位置０にアラインメントすることを含む。読み取り間で、次の１つまたは複数の塩基のコンセンサスが分析される。場合によっては、約３～１０回の読み取りがコンセンサスのために分析される。場合によっては、約３～約４、約３～約５、約３～約６、約３～約７、約３～約８、約３～約９、約３～約１０、約４～約５、約４～約６、約４～約７、約４～約８、約４～約９、約４～約１０、約５～約６、約５～約７、約５～約８、約５～約９、約５～約１０、約６～約７、約６～約８、約６～約９、約６～約１０、約７～約８、約７～約９、約７～約１０、約８～約９、約８～約１０、または約９～約１０回の読み取りが、コンセンサスのために分析される。場合によっては、約３、約４、約５、約６、約７、約８、約９、または約１０回の読み取りが、コンセンサスのために分析される。場合によっては、少なくとも約３、約４、約５、約６、約７、約８、または約９回の読み取りが、コンセンサスのために分析される。場合によっては、最大で約４、約５、約６、約７、約８、約９、または約１０回の読み取りが、コンセンサスのために分析される。場合によっては、次の１つまたは複数の塩基は、次の２～１０塩基を含む。場合によっては、次の１つまたは複数の塩基は、約２、３、４、５、６、７、８、９、または１０塩基である。場合によっては、次の１つまたは複数の塩基は、少なくとも約２、３、４、５、６、７、８、または９塩基である。場合によっては、次の１つまたは複数の塩基は、最大で約３、４、５、６、７、８、９、または１０塩基である。場合によっては、次の１つまたは複数の塩基は、約２、３、４、または５塩基である。コンセンサスが読み取り間で分析され、次の１つまたは複数の基が正しいかどうかが判定される。すべての読み取り間で、ある位置、例えばｘでの塩基間にコンセンサスが存在する場合、後続の塩基、例えばｘ＋１が次に分析されてよい。読み取り間で、ある位置、例えばｘでの塩基に不一致が存在する場合、その不一致を含んでいる読み取りにエラーがあるかどうかが判定される。場合によっては、エラーは挿入、欠失、または置換である。その後、読み取りごとに決定（例えば、読み取りが正しいか、またはエラーを有するか）が与えられて、位置がインクリメントされ、例えばｘ＋１になる。場合によっては、これらのステップは、読み取りの終了に達するまで繰り返される。

【０１７７】

場合によっては、複数のポリヌクレオチド配列をデコードするための方法（またはデコーディング方式）は、内部コーデックを含む。場合によっては、内部コーデックは、複数の核酸（またはポリヌクレオチド）配列に適用される。場合によっては、内部コーデックはデコーディング方式を含む。内部コーデックは、ポリヌクレオチド配列をデータ（例えば、デジタル・データまたはバイナリ・データ）に変換するために使用される。場合によっては、内部コーデックは、欠失エラー、置換エラー、または挿入エラー、あるいはこれらの任意の組み合わせを修正することができる。例えば、本明細書において提供される内部コーデックは、最大で１２％の欠失、６％の突然変異、または２％の挿入、あるいはこれらの任意の組み合わせのエラーを修正してよい。一部の例では、内部コーデックは、最大で６％の欠失、３％の突然変異、または１％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、内部コーデックは、約１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、または１２％の欠失、約１％、２％、３％、４％、５％、または６％の突然変異、あるいは約１％または２％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、内部コーデックは、最大で約１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、または１２％の欠失、約１％、２％、３％、４％、５％、または６％の突然変異、あるいは約１％または２％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部のさらなる実施形態では、内部コーデックは、オリゴ

の妥当性を確認し、疑わしいオリゴを破棄して外部デコーディングを汚染することを防ぐために使用される。場合によっては、内部コーデックは、インデックス（フレーム・インデックスおよびレーン・インデックス）を使用して効率的なデコーディングを可能にする。

【０１７８】

デコーディング方式を含んでいる内部コーデックが、複数のポリヌクレオチド配列に適用される（２１０）。内部コーデックのデコーディング方式は、複数のポリヌクレオチド配列の各々をデータのレーンに変換してよい。場合によっては、内部コーデックは、シーケンシングされた複数のヌクレオチドに適用される。場合によっては、内部コーデックは、順序付けられていない読み取りに適用される。場合によっては、内部コーデックは、本明細書において説明されるように、読み取りまたは複数のヌクレオチドがクラスタ化された後に、それらに適用される。場合によっては、内部コーデックは、本明細書において説明されるように、読み取りまたは複数のヌクレオチドがアライメントされた後に、それらに適用される。

【０１７９】

場合によっては、デコーディング方式は、ソフトウェアが８コア処理チップ、例えば８コアＩｎｔｅｌ（登録商標）９ｉ上で動作しているという条件で、少なくとも毎秒約５０，０００、１００，０００、１５０，０００、または２００，０００回の読み取りの速度で、読み取りをデコードしてよい。一部の例では、デコーディング方式は、ソフトウェアが８コア処理チップ（例えば、８コアＩｎｔｅｌ（登録商標）９ｉ）上で動作しているという条件で、少なくとも毎秒約１００，０００回の読み取り（例えば、毎秒約１３８，０００回の読み取りに対応する１時間当たり約５億回の読み取り）の速度で読み取りをデコードする。しかし、当業者は、１つまたは複数のハードウェア・パラメータ、１つまたは複数のソフトウェア手法、またはその両方を変更することによって、デコーディングの速度が高速化され得るということを理解するであろう。場合によっては、デコーディング方式は、水平方向または垂直方向にスケーリングされてよい。１つまたは複数のハードウェア・パラメータは、非限定的な例として、クロック速度、コア、キャッシュ・サイズ、ＲＡＭサイズ、ＣＰＵ、図１０のコンポーネント、または当技術分野において知られた任意の他のハードウェア・パラメータ、あるいはパラメータの組み合わせを含んでよい。１つまたは複数のソフトウェア手法は、非限定的な例として、同時実行、並列性、分散手法、または当技術分野において知られた任意の他の手法を使用する実装を含んでよい。

【０１８０】

場合によっては、内部コーデックは、貪欲アルゴリズムを含んでいるデコーディング方式を含む。場合によっては、内部コーデックは、最尤（ＭＬ）アルゴリズムを含んでいるデコーディング方式を含む。場合によっては、内部コーデックは、混合貪欲ＭＬアルゴリズムを含んでいるデコーディング方式を含む。

【０１８１】

貪欲アルゴリズム（例えば、貪欲デコーダ）を含むデコーディング方式が、図８に例示的に示されている。図に示されているように、貪欲アルゴリズムは、配列内の各ビット位置をデコードするときに、最も可能性の高い状態からの遷移のみを考慮する。場合によっては、貪欲アルゴリズムを使用して各ビットが１つずつ推測される。場合によっては、特定の時間に貪欲アルゴリズムを使用して２つ以上のビットが推測される。場合によっては、ｘ軸はビット位置を含み、ｙ軸は状態を含む。場合によっては、状態は、各ビット位置で分析される１つまたは複数の有効なエンコーディング状態Ｓを含む。場合によっては、各状態Ｓに確率が割り当てられる。場合によっては、状態Ｓは、各レーンからのエンコードされたビット、ビット履歴、およびビット位置として定義される。場合によっては、状態Ｓは、ビット履歴およびビット・ワードとして定義される。貪欲アルゴリズムは、最も可能性の高い終了状態に達するまで、各位置で最も可能性の高い状態を繰り返し見つける。場合によっては、デコードされたビットは、各ビット位置で最も可能性の高い状態をたどることによってバックトラックされる。場合によっては、これによって、完全にデコー

ドされたビットが得られる。場合によっては、貪欲デコーダは、局所最適解を見つける。場合によっては、局所最適解は、グローバル最適解の近似値である。貪欲デコーダは、本明細書に記載されたデコーディング方式などの他のデコーディング方式と比較して、妥当な時間量で解（または最終状態）を提供する。

【０１８２】

一部の例では、貪欲デコーダは、最大で６％の欠失、４％の突然変異、または１％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、貪欲デコーダは、最大で３％の欠失、２％の突然変異、または０．５％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、貪欲デコーダは、最大で３％の欠失、２％の突然変異、または０．５％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、貪欲デコーダは、約１％、２％、３％、４％、５％、または６％の欠失、約１％、２％、３％、または４％の突然変異、あるいは約０．５％または１％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、貪欲デコーダは、約１％、２％、３％、４％、５％、または６％の欠失、約１％、２％、３％、または４％の突然変異、あるいは約０．５％または１％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。

【０１８３】

場合によっては、ポリヌクレオチド配列が終了する位置を知ることによって、デコーディング方式の性能が改善される。場合によっては、オリゴヌクレオチドの長さは、例えばペアエンド・シーケンシングによって、シーケンシング中に決定される。場合によっては、ドリフト項が貪欲アルゴリズムに導入される。ドリフト項は、挿入および欠失の総数に関連付けられた整数を含む。各挿入は＋１の値として表され、各欠失は－１の値として表される。例えば、挿入がなく、２つの欠失がある場合、合計ドリフトは－２になる。そのような例では、貪欲アルゴリズムは、オリゴの長さと一致しないすべての最終デコーディング状態を無効として破棄する。したがって、ドリフト項は、貪欲アルゴリズムが、どの最終デコーディング状態が有効であるかを知ることを可能にし、性能をさらに改善することができる。そのため、場合によっては、図８および図９に示されているように、デコーディング方式は、ドリフトに対応するｚ軸をさらに含む。

【０１８４】

ＭＬアルゴリズム（またはＭＬデコーダ）を含むデコーディング方式（または内部コーデック）が、図９に例示的に示されている。図に示されているように、ＭＬアルゴリズムは、配列内の各ビット位置をデコードするときに、すべての状態からの遷移を考慮する。状態は、本明細書において前述したように定義される。場合によっては、ＭＬアルゴリズムを使用して各ビットが１つずつ推測される。場合によっては、特定の時間にＭＬアルゴリズムを使用して２つ以上のビットが推測される。場合によっては、ＭＬアルゴリズムは、最終候補状態が決定されるまで、各位置ですべての遷移状態を繰り返し検出する。場合によっては、本明細書において前述したように、ｘ軸はビット位置を含み、ｙ軸は状態を含む。場合によっては、本明細書において前述したように、ドリフト項は、最終候補状態をフィルタリングするために使用される。場合によっては、ＭＬアルゴリズムは、すべての状態遷移を追跡することよってグローバル最適解を提供する。場合によっては、ＭＬアルゴリズムは、本明細書に記載されたデコーディング方式などの他のデコーディング方式と比較して、計算負荷が高い。

【０１８５】

一部の例では、ＭＬデコーダは、最大で１２％の欠失、６％の突然変異、または２％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、ＭＬデコーダは、最大で６％の欠失、３％の突然変異、または１％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、ＭＬデコーダは、約１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、または１２％の欠失、約１％、２％、３％、４％、５％、または６％の突然変異、あるいは

約０．５％、１％、１．５％、または２％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、ＭＬデコーダは、最大で約１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、または１２％の欠失、約１％、２％、３％、４％、５％、または６％の突然変異、あるいは約約０．５％、１％、１．５％、または２％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。

【０１８６】

場合によっては、内部コーデックのデコーディング方式は、混合貪欲ＭＬアルゴリズムを含む。場合によっては、混合貪欲ＭＬアルゴリズムは、貪欲アルゴリズムおよびＭＬアルゴリズムを含む。混合貪欲ＭＬアルゴリズムは、配列内の各ビット位置をデコードするときに、複数の状態からの遷移を考慮する。場合によっては、複数の状態は、配列内の各ビット位置をデコードするときに、約１００～約１０００個の状態である。場合によっては、複数の状態は、約１００～約２００、約１００～約３００、約１００～約４００、約１００～約５００、約１００～約６００、約１００～約７００、約１００～約８００、約１００～約９００、約１００～約１，０００、約２００～約３００、約２００～約４００、約２００～約５００、約２００～約６００、約２００～約７００、約２００～約８００、約２００～約９００、約２００～約１，０００、約３００～約４００、約３００～約５００、約３００～約６００、約３００～約７００、約３００～約８００、約３００～約９００、約３００～約１，０００、約４００～約５００、約４００～約６００、約４００～約７００、約４００～約８００、約４００～約９００、約４００～約１，０００、約５００～約６００、約５００～約７００、約５００～約８００、約５００～約９００、約５００～約１，０００、約６００～ら約７００、約６００～約８００、約６００～約９００、約６００～約１，０００、約７００～約８００、約７００～約９００、約７００～約１，０００、約８００～約９００、約８００～約１，０００、または約９００～約１，０００個の状態である。場合によっては、複数の状態は、約１００、約２００、約３００、約４００、約５００、約６００、約７００、約８００、約９００、または約１，０００個の状態である。場合によっては、複数の状態は、少なくとも約１００、約２００、約３００、約４００、約５００、約６００、約７００、約８００、または約９００個の状態である。場合によっては、複数の状態は、最大で約２００、約３００、約４００、約５００、約６００、約７００、約８００、約９００、または約１，０００個の状態である。状態は、本明細書において前述したように定義される。場合によっては、混合貪欲ＭＬアルゴリズムを使用して各ビットが１つずつ推測される。場合によっては、特定の時間に混合貪欲ＭＬアルゴリズムを使用して２つ以上のビットが推測される。場合によっては、混合貪欲ＭＬアルゴリズムは、最終候補状態が決定されるまで、各位置で約１００～約１０００個の遷移状態を繰り返し検出する。場合によっては、本明細書において前述したように、ドリフト項は、最終候補状態をフィルタリングするために使用される。場合によっては、混合貪欲ＭＬアルゴリズムは、グローバル最適解を提供するが、本明細書に記載されたＭＬアルゴリズムなどの他のデコーディング方式と比較して、計算コストが低い。

【０１８７】

一部の例では、混合貪欲ＭＬデコーダは、最大で１５％の欠失、１０％の突然変異、または５％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、混合貪欲ＭＬデコーダは、最大で１２％の欠失、６％の突然変異、または２％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、混合貪欲ＭＬデコーダは、最大で６％の欠失、３％の突然変異、または１％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、混合貪欲ＭＬデコーダは、約１％、２％、３％、４％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、または１５％の欠失、約１％、２％、３％、４％、５％、６％、７％、８％、９％、または１０％の突然変異、あるいは約１％、２％、３％、または４％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。一部の例では、混合貪欲ＭＬデコーダは、最大で約１％、２％、３％、４

％、５％、６％、７％、８％、９％、１０％、１１％、１２％、１３％、１４％、または１５％の欠失、約１％、２％、３％、４％、５％、６％、７％、８％、９％、または１０％の突然変異、あるいは約１％、２％、３％、または４％の挿入、あるいはこれらの任意の組み合わせのエラーを修正することができる。

【０１８８】

場合によっては、内部コーデックにおけるデコーディング方式は、ビーム探索デコーダまたはランダム・サンプリング・デコーダ（例えば、純粋なサンプリング・デコーダ、トップＫサンプリング・デコーダなど）を含む。場合によっては、ビーム探索デコーダまたはランダム・サンプリング・デコーダは、貪欲デコーダと比較して多様な候補状態を提供する。

【０１８９】

場合によっては、内部コーデックは、チェックサムをさらに含む。場合によっては、データ完全性を検証するため、エラーを検出するため、またはこれらの組み合わせのために、チェックサムが使用される。場合によっては、チェックサム関数またはチェックサム・アルゴリズム（例えば、パリティ・バイトまたはパリティ・ワーク（ｐａｒｉｔｙ　ｗｏｒｋ）（縦方向のパリティ・チェック）、合計補数、位置依存、ファジー・チェックサムなど）を使用して、チェックサムが生成される。チェックサム関数またはチェックサム・アルゴリズムの例としては、ＢＳＤチェックサム（Ｕｎｉｘ）、ＳＹＳＶチェックサム（Ｕｎｉｘ）、ｓｕｍ４、ｓｕｍ８、ｓｕｍ１６、ｓｕｍ３２、ｆｌｅｔｃｈｅｒ－４、ｆｌｅｔｃｈｅｒ－８、ｆｌｅｔｃｈｅｒ－１６、ｆｌｅｔｃｈｅｒ－３２、Ａｄｌｅｒ－３２、ｘｏｒ８、Ｌｕｈｎアルゴリズム、Ｖｅｒｈｏｅｆｆアルゴリズム、またはＤａｍｍアルゴリズムが挙げられるが、これらに限定されない。場合によっては、最も可能性の高いパスのみを選択する代わりに、最も可能性の高いパスのいくつかが考慮され、チェックサムに対してテストされる。場合によっては、チェックサムは、ＲＳコード（例えば、小さいＲＳコード）を含む。そのような場合、デコーダは、どの可能性が実際に存在するかをユーザが決定できると仮定して、可能性のリスト（例えば、「リスト・デコーディング」）を提供する。

【０１９０】

場合によっては、デコーディングのための方法およびシステムは、レーンをフレームに配置することを含む。場合によっては、内部コーデックからのデコードされたレーンは、レーン・インデックスおよびフレーム・インデックスに基づいてフレームに配置される（２１５）。場合によっては、図７に示されているように、１つまたは複数のレーンがフレームから欠落している。場合によっては、ヌクレオチドの合成またはシーケンシング中に発生したエラーに起因して、レーンが欠落している。場合によっては、レーンの約１％～約１０％がフレームから欠落している。場合によっては、レーンの約１～２％、約１～約４％、約１～約６％、約１～約８％、約１～約１０％、約２～約４％、約２～約６％、約２％～約８％、約２％～約１０％、約４～約６％、約４～約８％、約４％～約１０％、約６％～約８％、約６％～約１０％、または約８％～約１０％がフレームから欠落している。場合によっては、レーンの約１％、約２％、約４％、約６％、約８％、または約１０％がフレームから欠落している。場合によっては、レーンの少なくとも約１％、約２％、約４％、約６％、または約８％がフレームから欠落している。場合によっては、レーンの最大で約２％、約４％、約６％、約８％、または約１０％がフレームから欠落している。

【０１９１】

場合によっては、内部コーデックは「フォーマット」を含む。場合によっては、デコーディング中にデータ（例えば、バイナリ・データ）のサイズに関する先験的な情報が存在しない。したがって、場合によっては、フレーム・インデックス０がデータのサイズを含む。場合によっては、レーンをフレームに配置した後、および／またはフレームを順序付けた後に、最初にフレーム０がデコードされる。次に、フレーム０からデータが抽出され、期待されるデータ・サイズの範囲外の（例えば、誤ってデコードされたオリゴからの）フレームを拒否する。

【０１９２】
場合によっては、内部コーデックはハッシュ（例えば、ＳＨＡ－２５６）を含む。場合によっては、ハッシュは、正しくデコードされたことを検証する。場合によっては、（外部コーデックまたはＥＣＣの後に）末尾でハッシュを使用することによって、エンコーディングおよびデコーディングがストリームとして実行される。場合によっては、これによって、メモリの使用を一時バッファのみに制限することができる。
【０１９３】
複数のポリヌクレオチド配列をデコードするための方法は、外部コーデックまたはエラー訂正コード（ＥＣＣ）を含む。場合によっては、複数のポリヌクレオチド配列がデータ（例えば、バイナリ・データ）にデコードされる。場合によっては、外部コーデックまたはＥＣＣがフレームの各々に適用される（２２０）。場合によっては、外部コーデックまたはＥＣＣは、内部コーデックからのレーンに適用される。場合によっては、内部コーデックからのレーンがフレームに配置された後に、外部コーデックまたはＥＣＣが適用される。
【０１９４】
場合によっては、外部コーデックがエラー訂正方式を含むか、またはコードが、データ（例えば、バイナリ・データ）をエンコードするために使用されるエラー訂正方式に基づく。場合によっては、エラー訂正方式は、リード・ソロモン（ＲＳ）コード、ＬＤＰＣコード、極コード、ターボコード、またはこれらの任意の組み合わせを含む。
【０１９５】
場合によっては、外部コーデックのエラー訂正方式は、リード・ソロモン（ＲＳ）コードを含む。そのような場合、ＲＳデコーダは、元のコードワードｃ（ｘ）にエラーｅ（ｘ）を加えたコードワードｒ（ｘ）（例えば、ｒ（ｘ）＝ｃ（ｘ）＋ｅ（ｘ））を受信する。場合によっては、エラーｅ（ｘ）は０である。場合によっては、ＲＳデコーダは、最大でｔ個のエラー（または２ｔ個の消失）の位置および大きさを識別しようとする。次に、ＲＳコードは、これらの識別されたエラーおよび／または消失を訂正しようとする。
【０１９６】
場合によっては、ＲＳデコーダはシンドローム計算を含む。場合によっては、シンドローム計算は、本明細書において前述したように、入力シンボルを受け取り、それらを生成器多項式ｇ（ｘ）に分割することを含む。場合によっては、シンドロームは、生成器多項式ｇ（ｘ）の２ｔ個の根（またはＲＳコードワードｃ（ｘ）のシンドローム）をｒ（ｘ）に代入することによって計算される。場合によっては、生成器多項式ｇ（ｘ）は、デコーダの既知のパラメータである。場合によっては、ＲＳコードワードｃ（ｘ）は、エラーに依存する２ｔ個のシンドロームを有する。
【０１９７】
場合によっては、ＲＳデコーダは、シンボル・エラーの位置を検出することを含む。場合によっては、パリティまたはチェック・シンボルｔは、エラーがない場合にシンドロームの計算が０になることを引き起こす。場合によっては、パリティまたはチェック・シンボルｔは、ＲＳエンコーダにおける剰余を含む。エラーが存在する場合、結果として得られた多項式ｇ（ｘ）がユークリッド・アルゴリズムに渡される。場合によっては、剰余の因子は、ユークリッド・アルゴリズムを使用して求められる。場合によっては、入力シンボルの各々の反復にわたって結果が評価される。場合によっては、エラーが検出され、エラーが訂正される。場合によっては、訂正されたコードワードｃ（ｘ）がＲＳデコーダから出力される。場合によっては、ＲＳコードによって訂正され得るエラーよりも多くのエラーがコードワードに存在する（例えば、ｅ（ｘ）＞２ｔ）。そのような場合、受信されたコードワードｒ（ｘ）がＲＳデコーダから出力される。場合によっては、受信されたコードワードｒ（ｘ）が、エラー訂正が失敗したことの指示（例えば、フラグ）と共に出力される。場合によっては、受信されたコードワードｒ（ｘ）（例えば、本明細書に記載されているようなバイナリ・データを含むレーンまたはフレーム）が、破棄される。
【０１９８】

場合によっては、外部コーデック（またはＥＣＣ）からのフレームがマージされて、データを含む出力を生成する（２２５）。場合によっては、データは、本明細書において前述したように、バイト・ストリームまたはバイト配列であることができるバイナリ・データを含む。

【０１９９】

本明細書に記載されたデコーディング方法（例えば、内部コーデック、外部コーデック、またはその両方）は、格納された複数のポリヌクレオチド配列のうちの少なくとも１つのポリヌクレオチド配列にエラーが存在する場合にデータを回復するために使用され得る。場合によっては、エラーは、挿入、欠失、置換、またはこれらの任意の組み合わせを含む。場合によっては、データは、複数のヌクレオチドにおけるポリヌクレオチド配列の約０．００１％～約３０％にエラー（例えば、エラー率）が存在する場合に回復される。場合によっては、データは、約０．００１％～約０．０１％、約０．００１％～約０．１％、約０．００１％～約０．５％、約０．００１％～約１％、約０．００１％～約２％、約０．００１％～約５％、約０．００１％～約１０％、約０．００１％～約１５％、約０．００１％～約２０％、約０．００１～約２５％、約０．００１～約３０％、約０．０１～約０．１％、約０．０１～約０．５％、約０．０１～約１％、約０．０１～約２％、約０．０１％～約５％、約０．０１～約１０％、約０．０１～約１５％、約０．０１～約２０％、約０．０１～約２５％、約０．０１～約３０％、約０．１～０．５％、約０．１～約１％、約０．１～約２％、約０．１～約５％、約０．１～約１０％、約０．１～約１５％、約０．１～約２０％、約０．１～約２５％、約０．１～約３０％、約０．５～約１％、約０．５～約２％、約０．５～約５％、約０．５～１０％、約０．５～約１５％、約０．５～約２０％、約０．５～約２５％、約０．５～約３０％、約１％～約２％、約１％～約５％、約１％～約１０％、約１％～約１５％、約１％～約２０％、約１％～約２５％、約１～約３０％、約２～５％、約２～１０％、約２～１５％、約２～２０％、約２％～約２５％、約２％～約３０％、約５～約１０％、約５～約１５％、約５～約２０％、約５～約２５％、約５～約３０％、約１０～約１５％、約１０～約２０％、約１０～約２５％、約１０～約３０％、約１５～約２０％、約１５～約２５％、約１５～約３０％、約２０～約２５％、約２０～約３０％、または約２５～約３０％のエラー率が存在する場合に回復される。場合によっては、データは、約０．００１％、約０．０１％、約０．１％、約０．５％、約１％、約２％、約５％、約１０％、約１５％、約２０％、約２５％、または約３０％のエラー率が存在する場合に回復される。場合によっては、データは、少なくとも約０．００１％、約０．０１％、約０．１％、約０．５％、約１％、約２％、約５％、約１０％、約１５％、約２０％、または約２５％のエラー率が存在する場合に回復される。場合によっては、データは、最大で約約０．０１％、約０．１％、約０．５％、約１％、約２％、約５％、約１０％、約１５％、約２０％、約２５％、または約３０％のエラー率が存在する場合に回復される。

【０２００】

場合によっては、デコーディング方式（例えば、外部デコーディングおよび内部デコーディング）は、ソフト・デコーディングと共に使用される。ソフト・デコーディングは、一般に、可能な値の範囲を考慮する（例えば、確率推定値を使用する）ことによってデコードすることを指す。例として、シーケンシングは、塩基ごとに品質を保持することができ、この品質は、確率計算時に考慮され得る。そのような例では、各状態は最終確率を含み、この最終確率は、外部デコーダで、例えば、その外部デコーダがソフト・デコーディングをサポートしている場合に、対数尤度として使用され得る。さらに、クラスタ化およびアライメントは、アライメントの信頼性に関するソフト情報を提供することができる。さらなる例として、ＬＤＰＣ外部コーデックは、反復的デコーダを含む。これは、１回のパスではなく、内部デコーダと外部デコーダの間を反復的に行き来する可能性をもたらす。しかし、場合によっては、これは、より高いコンピューティング要件のコストを伴う。

【０２０１】

デコーディングは、少なくとも１つの論理要素、プログラマブル・ロジック、またはプ

ロセッサ上で実行され得る。少なくとも１つの論理要素、プログラマブル・ロジック、またはプロセッサの非限定的な例としては、プログラマブル・ロジック・コントローラ（ＰＬＣ：ｐｒｏｇｒａｍｍａｂｌｅ　ｌｏｇｉｃ　ｃｏｎｔｒｏｌｌｅｒ）、プログラマブル・ロジック・アレイ（ＰＬＡ：ｐｒｏｇｒａｍａｂｌｅ　ｌｏｇｉｃ　ａｒｒａｙ）、プログラマブル・アレイ・ロジック（ＰＡＬ：ｐｒｏｇｒａｍｍａｂｌｅ　ａｒｒａｙ　ｌｏｇｉｃ）、汎用ロジック・アレイ（ＧＬＡ：ｇｅｎｅｒｉｃ　ｌｏｇｉｃ　ａｒｒａｙ）、コンプレックス・プログラマブル・ロジック・デバイス（ＣＰＬＤ：ｃｏｍｐｌｅｘ　ｐｒｏｇｒａｍｍａｂｌｅ　ｌｏｇｉｃ　ｄｅｖｉｃｅ）、フィールド・プログラマブル・ゲート・アレイ（ＦＰＧＡ）、または特定用途向け集積回路（ＡＳＩＣ）、ＧＰＵ、ＣＰＵ、ＡＩアクセラレータ、またはこれらの任意の組み合わせが挙げられる。場合によっては、ＡＩアクセラレータは、Ｇｏｏｇｌｅ－ＴＰＵ、Ｇｒａｐｈｃｏｒｅ、Ｃｅｒｅｂｒａｓ、ＳａｍｂａＮｏｖａ、またはこれらの組み合わせを含む。場合によっては、デコーディングは、ＵｐＭｅｍなどの、ただしこれに限定されない、コンピュートオンメモリ（ｃｏｍｐｕｔｅ－ｏｎ－ｍｅｍｏｒｙ）技術で実行される。

【０２０２】

本開示のハッシュは、取り出し中のデジタル情報の検証を可能にすることができる。場合によっては、複数のポリヌクレオチドに格納されたデジタル情報を取り出すことは、少なくとも１つまたは複数のオブジェクトを検証することをさらに含む（１３１５）。場合によっては、複数のプール内の第１の１つまたは複数のハッシュを使用して１つまたは複数のオブジェクトが検証される。場合によっては、複数のポリヌクレオチドに格納されたデジタル情報を取り出すことは、１つまたは複数のプール・アイテムを検証することをさらに含む。場合によっては、複数のプール内の第２の１つまたは複数のハッシュを使用して１つまたは複数のプール・アイテムが検証される。一部の例では、オブジェクトが複数のプールのうちの２つ以上のプールにまたがって格納される場合、２つ以上のプール・アイテムがこのオブジェクトに組み立てられる。そのような例では、プール・アイテムの各々のデータ・ペイロードの第１の１つまたは複数のハッシュ、１つまたは複数のオブジェクトの第２の１つまたは複数のハッシュ、またはこれらの組み合わせが、適切なアセンブリ検証を可能にする。

【０２０３】

ハッシュを検証することは、一般に、ハッシュ（例えば、暗号ハッシュ）を生成することを含む。検証は、生成されたハッシュを以前に決定されたハッシュと比較することをさらに含むことができる。場合によっては、以前のハッシュおよび新しいハッシュは、同じハッシュ関数を使用して決定される。場合によっては、ハッシュ関数は暗号ハッシュ関数を含む。場合によっては、ハッシュ関数は、ＭＤ－５、ＳＨＡ－１、ＳＨＡ－２、ＳＨＡ－３、ＲＩＰＥＭＤ－１６０、Ｗｈｉｒｌｐｏｏｌ、ＢＬＡＫＥ、ＢＬＡＫＥ２、ＢＬＡＫＥ３、またはこれらの変形を含む。場合によっては、ハッシュ関数はＳＨＡ－２を含む。一部の例では、ＳＨＡ－２は、ＳＨＡ－２２４、ＳＨＡ－２５６、ＳＨＡ－３８４、ＳＨＡ－５１２、ＳＨＡ－５１２／２２４、またはＳＨＡ－５１２／２５６を含む。場合によっては、新しいハッシュと以前のハッシュが一致する場合、情報のアイテム（例えば、オブジェクト）の完全性が検証される。場合によっては、新しいハッシュと以前のハッシュが一致しない場合、検証が失敗する。場合によっては、検証が失敗する場合、情報のアイテムの完全性は検証されない。場合によっては、検証が失敗する場合、情報のアイテムは変更されている、および／または破損している。

【０２０４】

デジタル情報を取り出すことは、プール・アイテムおよび／または複数のプールにわたって格納された情報を結合することを含むことができる。場合によっては、複数のポリヌクレオチドに格納されたデジタル情報を取り出すことは、複数のプール内のデジタル情報を結合することをさらに含む（１３２０）。場合によっては、１つまたは複数のプール・アイテム内のデータ・ペイロードが結合される。場合によっては、複数のプールにわたる１つまたは複数のプール・アイテム内のデータ・ペイロードが結合される。場合によって

は、結合されたデータ・ペイロードは、デジタル情報を含む。場合によっては、取り出されたデータまたはデジタル情報は、メモリに格納される（１３２５）。

【０２０５】

場合によっては、取り出されたデジタル情報は、ユーザに提示される。場合によっては、情報は、インターフェイス上でユーザに提示される。場合によっては、インターフェイスは、電子デバイス（例えば、個人用電子デバイス）のインターフェイスである。場合によっては、電子デバイスは、情報にアクセスするために、コンピュータ・ネットワークを介して、本明細書に記載されたシステムと通信するように構成されたアプリケーションを含む。

【０２０６】

場合によっては、本明細書において説明されるように、複数のポリヌクレオチド配列をデコードして、デジタル・データ（例えば、バイナリデータ）を含む出力を生成するための方法は、システム上で実行される。場合によっては、システムは、図１、図２、またはその両方に概して示されている動作を実行する。場合によっては、そのようなシステムは、メモリ、シーケンシング・デバイス、メモリに動作可能に結合された処理デバイス、またはこれらの組み合わせを備えている装置を備える。場合によっては、シーケンシング・デバイスは、メモリ、処理デバイス、またはこれらの組み合わせに動作可能に結合される。場合によっては、メモリは、バイナリ・データ、ポリヌクレオチド配列、またはこれらの組み合わせの情報を格納するために使用される。場合によっては、バイナリ・データ、ポリヌクレオチド配列、またはこれらの組み合わせの情報は、本明細書に記載されたエンコーディング方法における１つまたは複数のステップからのものである。場合によっては、メモリは、本明細書に記載されたアルゴリズムに関連する情報（例えば、ソフトウェア・コード、パラメータ、実行可能命令など）を格納するために使用される。一部の例では、メモリは、本明細書に記載された任意の適切なメモリを含むことができる。一部の例では、メモリは、本明細書に記載された実施形態に従って構成され得る。一部の例では、シーケンシング・デバイスは、本明細書に記載された方法を使用して複数のポリヌクレオチド配列を決定するように構成される。

【０２０７】

場合によっては、処理デバイスは、１つまたは複数のデコーディング・ステップを実行するように構成される。場合によっては、処理デバイスは、デコーディング方式を含む内部コーデックを複数のポリヌクレオチド配列に適用するステップと、バイナリ・データのレーンの各々におけるレーン・インデックスおよびフレーム・インデックスに基づいて、バイナリ・データのレーンをフレームに配置するステップと、外部コーデックをフレームに適用するステップと、を含む１つまたは複数のステップを実行するように構成される。場合によっては、デコーディング方式は、複数のポリヌクレオチド配列の各々をバイナリ・データのレーンに変換する。場合によっては、デコーディング方式は、貪欲アルゴリズムおよび最尤（ＭＬ）アルゴリズムを含んでいる混合デコーディング・アルゴリズムを含む。場合によっては、外部コーデックはエラー訂正方式を含む。場合によっては、外部コーデックからのフレームがマージされて、バイナリ・データを含む出力を生成する。

【０２０８】

ＤＮＡ（またはポリヌクレオチド）内のデジタル情報を取り出すための方法は、システム上で実行され得る。場合によっては、システムは、図１２、図１３、またはその両方に概して示されている動作を実行する。場合によっては、そのようなシステムは、１つまたは複数の処理ユニット、メモリ、命令、シーケンシング・デバイス、またはこれらの組み合わせを備えている装置を備える。場合によっては、メモリは、１つまたは複数の処理ユニットと通信する。場合によっては、命令はメモリに格納される。場合によっては、シーケンシング・デバイスは、メモリ、１つまたは複数の処理ユニット、またはこれらの組み合わせと通信する。場合によっては、１つまたは複数の処理ユニットおよびメモリは、１つまたは複数の物理的位置または論理的位置にわたって分散される。

【０２０９】

場合によっては、メモリは、データまたはデジタル情報、ポリヌクレオチド配列（例えば、部分的または完全にデコードされた配列）、またはこれらの組み合わせを格納するために使用される。場合によっては、メモリは、本明細書に記載されたアルゴリズムに関連する情報（例えば、ソフトウェア・コード、パラメータ、実行可能命令など）を格納するために使用される。一部の例では、メモリは、本明細書に記載された任意の適切なメモリを含むことができる。一部の例では、メモリは、本明細書に記載された実施形態に従って構成され得る。一部の例では、シーケンシング・デバイスは、本明細書に記載された方法を使用して複数のポリヌクレオチド配列を決定するように構成される。

【０２１０】

場合によっては、１つまたは複数の処理ユニットは、中央処理ユニット（ＣＰＵ）、グラフィカル処理ユニット（ＧＰＵ）、シングル・コア・プロセッサ、マルチコア・プロセッサ、プロセッサ・クラスタ、特定用途向け集積回路（ＡＳＩＣ）、フィールド・プログラマブル・ゲート・アレイ（ＦＰＧＡ）などのプログラマブル回路、ＡＩアクセラレータ、およびこれらの変形の任意の組み合わせを含む。場合によっては、処理ユニットのうちの１つまたは複数は、単一命令複数データ（ＳＩＭＤ）または単一プログラム複数データ（ＳＰＭＤ）並列アーキテクチャを備える。１つまたは複数の処理ユニットの例としては、ＳＩＭＤまたはＳＰＭＤを実装する１つまたは複数のＧＰＵまたはＣＰＵが挙げられる。場合によっては、ＡＩアクセラレータは、Ｇｏｏｇｌｅ－ＴＰＵ、Ｇｒａｐｈｃｏｒｅ、Ｃｅｒｅｂｒａｓ、ＳａｍｂａＮｏｖａ、またはこれらの組み合わせを含む。一部の実施形態では、処理ユニットのうちの１つまたは複数は、ハードウェア実装に加えて、ソフトウェアおよび／またはファームウェアに実装される。処理ユニットのソフトウェアまたはファームウェアの実装は、本明細書に記載されたさまざまな機能を実行するために任意の適切なプログラミング言語で記述された、コンピュータ実行可能命令または機械実行可能命令を含むことができる。１つまたは複数の処理ユニットのソフトウェア実装は、全体的または部分的にメモリに格納され得る。代替的または追加的に、システムは、１つまたは複数のハードウェア・ロジック・コンポーネントを備えることができる。例えば、使用され得るハードウェア・ロジック・コンポーネントの種類の例としては、フィールドプログラマブル・ゲート・アレイ（ＦＰＧＡ）、特定用途向け集積回路（ＡＳＩＣ）、特定用途向け標準製品（ＡＳＳＰ）、システムオンチップ・システム（ＳＯＣ）、複合プログラマブル・ロジック・デバイス（ＣＰＬＤ）などが挙げられるが、これらに限定されない。場合によっては、デコーディングは、ＵｐＭｅｍなどの、ただしこれに限定されない、コンピュートオンメモリ技術で実行される。

【０２１１】

場合によっては、１つまたは複数の処理ユニットは、１つまたは複数のデコーディング・ステップを実行するように構成される。場合によっては、処理デバイスは、デコーディング方式を適用して、複数のプール内のデジタル情報をデコードするステップと、複数のプール内の第１の１つまたは複数のハッシュを使用して少なくともプール・アイテム内のデータ・ペイロードを検証するステップと、複数のプール内のデジタル情報を結合して、１つまたは複数のオブジェクトを取り出すステップと、デジタル情報をメモリに格納するステップと、を含む１つまたは複数のステップを実行するように構成される。場合によっては、１つまたは複数の処理ユニットは、内部コーデックを複数のポリヌクレオチドに適用するか、またはＥＣＣを複数のポリヌクレオチドに適用するステップを含む１つまたは複数のステップを実行するように構成される。場合によっては、内部コーデックは、複数のポリヌクレオチドの各々をデジタル情報に変換する。場合によっては、内部コーデックは、貪欲アルゴリズムおよび最尤（ＭＬ）アルゴリズムを含んでいる混合デコーディング・アルゴリズムを含む。場合によっては、ＥＣＣからの出力がマージされて、デジタル情報を含む出力を生成する。

【０２１２】

ＤＮＡデータ・ストレージ

【０２１３】

本明細書に記載された情報をエンコードするポリヌクレオチドは、データ・ストレージ・システムに格納されてよい。データ・ストレージのためのシステムは、１つまたは複数のモジュールを備えることができる。場合によっては、１つまたは複数のモジュールの一部またはすべては通信する。一部の例では、１つまたは複数のモジュールの一部またはすべては、それらのモジュール間のポリヌクレオチドの転送を可能にするために通信する。一部の例では、１つまたは複数のモジュールの一部またはすべては流体結合される。一部の例では、１つまたは複数のモジュールの一部またはすべては、１つまたは複数のチューブを使用して流体結合される。流体は、一般に、合成、増幅、シーケンシングの準備、およびシーケンシングを含むが、これらに限定されない、ポリヌクレオチドを処理することに関与するさまざまなプロセスにおいて使用される１つまたは複数の液体のことを指してよい。一部の例では、モジュールの一部またはすべては、システムのモジュール間の制御コマンドの転送を可能にするために通信する。一部の例では、モジュールの一部またはすべては、電子的に結合される。システム内のモジュールは、シンセサイザ・ユニット、増幅室、シーケンサ・ユニット、ストレージ・ユニット、コントローラ、ロボット・システム、またはこれらの任意の組み合わせを含むことができるが、これらに限定されない。一部の例では、モジュールは、流体源、データベースまたはファイル・システムあるいはその両方を、さらに備えることができる。一部の例では、データベースまたはファイル・システムは、システムのストレージ容量を追跡する。例えば、データベースまたはファイル・システムは、使用可能なラック（またはトレイ）、スロット（カプセル用）、またはその両方を追跡することができる。一部の例では、データベースまたはファイル・システムは、ストレージ・システム内のラックの配置を決定するために使用される。場合によっては、システムの１つまたは複数のモジュール間のポリヌクレオチドの移動は、１つまたは複数のチューブまたはロボット・システムによって実現される。一部の例では、データベースまたはファイル・システムは、ロボット・システムをストレージ・システム内の正しい位置に向かわせるために使用される。場合によっては、システムは自律的である。

【０２１４】

データ・ストレージのためのシステムの非限定的な例が、図１６に示されている。データ・ストレージのためのシステムは、シンセサイザ・ユニット１６１０を備えてよい。シンセサイザ・ユニットは、デジタル情報をエンコードする複数のポリヌクレオチドを合成するために使用され得る。場合によっては、システムは、２つ以上のシンセサイザ・ユニット１６１０を備える。ポリヌクレオチドは、本明細書において提供される方法または当技術分野において知られた任意の他の適切な合成方法を使用して合成されてよい。シンセサイザ・ユニット１６１０におけるポリヌクレオチド合成の流体制御および／または電子制御は、コントローラ１６３５によって実行されてよい。場合によっては、シンセサイザ・ユニット１６１０内の電子機器は、コントローラ１６３５と通信する。場合によっては、シンセサイザ・ユニット１６１０は、ＤＮＡ配列を受信するための入力を有する。場合によっては、シンセサイザ・ユニット１６１０は、ポリヌクレオチド合成のための流体を受け取るための入力を有する。場合によっては、シンセサイザ・ユニット１６１０は、合成されたポリヌクレオチドを溶出するための出力を有する。場合によっては、合成されたポリヌクレオチドは、非限定的な例として、ストレージ・ユニット、増幅室、またはシーケンシング・ユニットなどの、システムの別のコンポーネントに転送される。

【０２１５】

シンセサイザ・ユニットは、固体支持体を備えてよい。固体支持体は、本明細書に記載されたポリヌクレオチドの格納のためのデバイスを備えてよい。固体支持体は、ポリヌクレオチド合成のための表面を備えてよい。場合によっては、固体支持体、表面、またはその両方は、本明細書に記載された材料を含む。場合によっては、材料は、金属または有機ポリマーを含む。場合によっては、材料は、鋼（例えば、ステンレス鋼）または他の金属合金を含む。場合によっては、材料は、ポリエチレン、ポリプロピレン、または他のポリマーを含む。場合によっては、構造体は、本明細書において提供される材料などの柔軟な材料を含む。柔軟な材料の例としては、改質ナイロン、非改質ナイロン、ニトロセルロー

ス、およびポリプロピレンが挙げられるが、これらに限定されない。場合によっては、材料は、本明細書において提供される材料などの硬質材料を含む。硬質材料の例としては、ガラス、ヒューズド・シリカ、ケイ素、二酸化ケイ素、窒化ケイ素、プラスチック（例えば、ポリテトラフルオロエチレン、ポリプロピレン、ポリスチレン、ポリカーボネート、およびこれらの混合物）、ならびに金属（例えば、鋼、金、白金）が挙げられるが、これらに限定されない。場合によっては、本明細書において開示された材料は、ケイ素、ポリスチレン、アガロース、デキストラン、セルロース系ポリマー、ポリアクリルアミド、ポリジメチルシロキサン（ＰＤＭＳ）、ガラス、またはこれらの任意の組み合わせを含む材料から製造されてよい。一部の例では、本明細書において開示された材料は、本明細書においてリストされた材料または当技術分野において知られた任意の他の適切な材料の組み合わせで製造される。

【０２１６】

場合によっては、ポリヌクレオチドは、シンセサイザ・ユニット１６１０から脱保護され、切断され、および／または溶出され、システム内の別のモジュールに転送される。場合によっては、ロボット・システム１６３０または流体チューブは、ポリヌクレオチドをシステム内の別のモジュールに輸送するために使用される。ロボット・システム１６３０は、コントローラ１６３５によって制御されてよい。ロボット・システムは、一般に、複数のポリヌクレオチドの操作のためのシステムを含む。場合によっては、ロボット・システムは、本明細書に記載されたポリヌクレオチドなどの、複数のポリヌクレオチドを含む構造体を操作するために使用される。操作は、非限定的な例として、移動、格納、取り出し、処理、転送、またはこれらの任意の組み合わせを含むことができる。ロボット・システムは、処理デバイス間でウェハーおよびチップのトレイを移動するために半導体処理において使用されるロボット・システムに類似してよい。ロボット・システム１６３０は、システムのモジュール間でポリヌクレオチドを選択し、転送するために使用されてよい。例えば、ロボット・システム１６３５は、ストレージ・ユニット１６１５内の構造体を検証するためのタグ・リーダを含んでよい。場合によっては、ロボット・システム１６３５はタグ・リーダを備え、ストレージ・ユニット１６１５内の構造体は、タグ（例えば、バーコードまたはＲＦＩＤタグ）を備える。検証された後に、ロボット・システム１６３０は、構造体をシステムのコンポーネントに転送してよい。さらに、ロボット・システム１６３０は、構造体をシステムのコンポーネント内の正確な位置に転送してよい。場合によっては、ロボット・システムは、ポリヌクレオチドがデータ・ストレージ・システム内のモジュールに追加されること、および／またはモジュールから除去されることを可能にすることができる。場合によっては、ロボット・システムは、複数のポリヌクレオチドを含む構造体が、ストレージ・ユニット１６１５内の識別可能なレイアウトでの位置に配置されること、および／またはその位置から取り出されることを可能にする。ロボット・システム１６３０は、本明細書においてさらに説明されるように、コントローラ１６３５を使用して制御されてよい。

【０２１７】

場合によっては、ポリヌクレオチドを含む１つまたは複数の液滴が、シンセサイザ・ユニット１６１０からストレージ・ユニット１６１５に転送される。場合によっては、固体支持体上で合成されたポリヌクレオチドの一部またはすべてが、格納のために構造体に転送される。構造体またはコンパートメントは、さまざまな形状およびサイズを有してよい。構造体は、タグ（例えば、バーコードまたはＲＦＩＤタグ）をさらに備えてよい。場合によっては、複数のポリヌクレオチドが、シンセサイザ・ユニット１６１０内の構造体に転送される。場合によっては、複数のポリヌクレオチドが、ストレージ・ユニット１６１５内の構造体に転送される。ストレージ・ユニット１６１５におけるポリヌクレオチド合成の流体制御および／または電子制御は、コントローラ１６３５によって実行されてよい。場合によっては、ストレージ・ユニット１６１５内の電子機器は、コントローラ１６３５と通信する。場合によっては、ポリヌクレオチドは、ストレージ・ユニット１６１５に室温で格納される。場合によっては、システムは、ストレージ・ユニット１６１５内のス

トレージ容量を追跡するためのデータベースまたはファイル・システムを備える。一部の例では、データベースは制御アプリケーション・データベースを含む。場合によっては、データベースまたはファイル・システムは、コントローラ１６３５の一部である。

【０２１８】

複数のポリヌクレオチドを含む構造体は、識別可能なレイアウトでストレージ・ユニット１６１５に格納され得る。識別可能なレイアウトは、１つまたは複数のラックまたはその変形を含んでもよい。ラックは、複数のポリヌクレオチドを含む１つまたは複数の構造体を保持するために使用されてよい。場合によっては、各構造体は、識別可能なレイアウトで固定位置に格納される。場合によっては、タグは、識別可能なレイアウトでの構造体の位置に関する情報を含む。例として、タグは、識別可能なレイアウトでの構造体の位置を含むメタデータをエンコードすることができる。場合によっては、ラックはデータ・センターに位置してよい。場合によっては、ラックは、従来のコンピューティング・リソースおよびデータス・トレージ・リソースをラック・ユニットに取り付けるために一般に使用される機械構造を使用する。例えば、ラックは、ディスク・ドライブ、処理ブレード、および／または他のコンピュータ機器を支えるように適応された開口部を備えてよい。場合によっては、ラックはタグを備える。一部の例では、タグは、ラック内／ラック上に格納された構造体の情報を含む。一部の例では、タグは、ラック内／ラック上に格納された構造体のリストを含む。

【０２１９】

場合によっては、ストレージ・ユニット１６１５は、ロボット・システム１６３０を使用してアクセスされてよい。場合によっては、ストレージ・ユニット１６１５内の識別可能なレイアウトは、ロボット制御でアドレス指定可能なスロットを含む。各スロットは、複数のポリヌクレオチドを含む構造体を保持してよい。一部の例では、各スロットは、数のポリヌクレオチドを含む構造体を収容するための幅、深さ、長さ、またはこれらの任意の組み合わせを含む。場合によっては、ラックは複数のスロットを備え、各スロットは複数のポリヌクレオチドを含む構造体を保持する。

【０２２０】

ポリヌクレオチドを格納するためのシステムは、増幅室１６２０をさらに備えてよい。増幅ユニットは、複数のポリヌクレオチドを増幅するために使用されてよい。場合によっては、システムは、２つ以上の増幅室１６２０を備える。場合によっては、構造体がストレージ・ユニット１６１５から選択され、構造体からのポリヌクレオチドが増幅室１６２０に転送される。場合によっては、シンセサイザ・ユニット１６１０からのポリヌクレオチドは、サイズ選択、ＰＣＲ、または格納のための他の種類の増幅もしくは準備のために増幅室１６２０に転送される。サイズ選択は、通常、対象のサイズのＤＮＡを選択すること、および非常に短いか、または非常に長い鎖を拒否することを含む。場合によっては、特定のサイズ範囲のＤＮＡを捕捉するようにフィルタが調整される。場合によっては、他の方法は、ＰＣＲ、電気泳動、合成されたオリゴヌクレオチドの末端配列を補完する固相結合プライマーによる捕捉、または等温ポリメラーゼの使用を含む。増幅室１６２０におけるポリヌクレオチド合成の流体制御および／または電子制御は、コントローラ１６３５によって実行されてよい。場合によっては、増幅室１６２０内の電子機器は、コントローラ１６３５と通信する。

【０２２１】

ポリヌクレオチドを格納するためのシステムは、シーケンシング・ユニット１６２５をさらに備えてよい。シーケンシング・ユニット１６２５は、複数のポリヌクレオチドをシーケンシングするために使用されてよい。場合によっては、複数のポリヌクレオチドが、増幅室１６２０からシーケンシング・ユニット１６２５に転送される。場合によっては、システムは、追加のシーケンシング準備ステップを実行するための追加のモジュールを備えてよい。一部の例では、１つまたは複数のチューブまたはロボット・システム１６３０を使用して、複数のポリヌクレオチドが増幅室１６２０からシーケンシング・ユニット１６２５に転送される。場合によっては、増幅室１６２０およびシーケンシング・ユニット

１６２５は流体結合される。シーケンシング・ユニット１６２５におけるポリヌクレオチド合成の流体制御および／または電子制御は、コントローラ１６３５によって実行されてよい。場合によっては、シーケンシング・ユニット１６２５内の電子機器は、コントローラ１６３５と通信する。

【０２２２】

場合によっては、システムは、ポリヌクレオチドの大規模なシーケンシングを含む。場合によっては、大規模なシーケンシングは、高密度の高度に並列化されたシーケンサを含む。場合によっては、システムは、２つ以上のシーケンシング・ユニット１６２５を備える。場合によっては、シーケンシング・ユニット１６２５は、遠心力および／または真空／圧力を使用して、シーケンシング・ユニット１６２５に試薬を追加するか、またはシーケンシング・ユニット１６２５から試薬を排出する。場合によっては、シーケンシング・ユニット１６２５は、光ベース（例えば、光源およびセンサをチップ上に備える）、ナノポアベース（例えば、Ｏｘｆｏｒｄ　Ｎａｎｏｐｏｒｅ　Ｔｅｃｈｎｏｌｏｇｉｅｓ（ＯＮＴ））であるか、または他の動作（例えば、ＰａｃＢｉｏまたは他のシーケンシング技術などの光ベースの方法）を含む。場合によっては、シーケンシング・ユニット１６２５は、本明細書において提供されるシーケンシング方法を採用する。場合によっては、シーケンシング・ユニット１６２５は、半導体製造装置によって提供されるバルク流体工学から恩恵を受けるナノポアまたは他の電気的シーケンシング技術を使用する。場合によっては、本明細書に記載された１つまたは複数のモジュールは、カメラを備える。カメラは、モジュール内のポリヌクレオチドの１つまたは複数の光学的特徴を捕捉するために使用されてよい。例として、カメラは、本明細書に記載されているような固体支持体上の表面に取り付けられたポリヌクレオチドの光学的特徴を捕捉するために、シンセサイザ・ユニット、シーケンシング・ユニット、またはその両方において使用されてよい。

【０２２３】

ポリヌクレオチドを格納するためのシステムは、本明細書に記載されているようなロボット・システム１６３０を備えることができる。ロボット・システムは、通常、システム内のポリヌクレオチドを操作するために使用されてよい。操作は、移動、格納、取り出し、処理、転送、またはこれらの任意の組み合わせを含むことができるが、これらに限定されない。場合によっては、ロボット・システムは、システム内のモジュール間で複数のポリヌクレオチドを転送する。一部の例では、ロボット・システムは、本明細書において説明されるように、格納のために構造体内の複数のポリヌクレオチドを操作（例えば、転送）する。場合によっては、ロボット・システムは、ラック内の複数のポリヌクレオチドを操作（例えば、転送）する。一部の例では、ラックは、タグをそれぞれ備えている複数の構造体を備える。一部の例では、ラックは、合成および／またはシーケンシングのための複数の固体支持体を備える。場合によっては、ロボット・システムは、ロボット・ハンドまたはロボット・ピッカーを備える。場合によっては、ロボット・システム１６３０は、コントローラ１６３５内のストレージ・システム制御ソフトウェアおよび／またはファームウェアと完全に統合される。場合によっては、ロボット・システム１６３０は、外部ホスト・アプリケーションと完全に統合される。場合によっては、ロボット・システム１６３０は完全に自動化される。

【０２２４】

ポリヌクレオチドを格納するためのシステムは、コントローラ１６３５を備えることができる。コントローラは、通常、モジュール、コンポーネント、流体工学、ロボット、またはこれらの任意の組み合わせを制御するために使用されてよい。モジュール、コンポーネント、流体工学、電子機器、ロボット、またはこれらの任意の組み合わせは、ポリヌクレオチドの合成、格納、取り出し、シーケンシング、および／または増幅に使用されてよい。場合によっては、コントローラ１６３５は、ラック内にロードされた、アンロードされた、および／または格納されたすべてのストレージ構造体をカタログ化することができる。ポリヌクレオチドは、本明細書において説明されるように、デジタル情報をエンコードすることができる。モジュール、コンポーネント、流体工学、電子機器、ロボット、ま

たはこれらの任意の組み合わせは、ポリヌクレオチドをエンコードすること、またはデコードすることなどの、方法、モデル、またはアルゴリズムを実行するために使用されてよい。

【０２２５】

場合によっては、コントローラ１６３５は、複数のポリヌクレオチドの物理的位置を制御する。場合によっては、コントローラ１６３５は、コマンドをシステムの１つまたは複数のモジュールに提供する。一部の例では、コントローラ１６３５は、ロボット工学（例えば、ロボット・システム１６３０）、アクチュエータ、および流体バルブ、またはシステムの任意の他の機器を制御する。場合によっては、コントローラ１６３５は、ポリヌクレオチドを処理および／または転送するためのモジュールを同期させて制御することを可能にする。一部の例では、ポリヌクレオチドは、流体工学によって処理および／または転送される。一部の例では、ポリヌクレオチドは、電子機器によって処理および／または転送される。場合によっては、コントローラ１６３５は、圧力、真空、温度、体積（例えば、流体）、またはこれらの任意の組み合わせなどの、これらに限定されない、１つまたは複数のモジュール内の物理的パラメータを制御する。

【０２２６】

場合によっては、コントローラ１６３５は、エンコーダ・モジュールまたはデコーダ・モジュールを呼び出す。場合によっては、エンコーダ・モジュールは、デジタル情報を複数のポリヌクレオチドとしてエンコードする。場合によっては、エンコーダ・モジュールは、本明細書（例えば、図１、図３～６、図１２、図１３～１４）に記載されたコーデックなどの１つまたは複数のコーデックをデジタル情報に適用する。場合によっては、デコーダ・モジュールは、複数のポリヌクレオチドの配列をデコードしてデジタル情報を取り出す。場合によっては、デコーダ・モジュールは、本明細書（例えば、図２、図７～９、図１３）に記載されたコーデックなどの１つまたは複数のコーデックを、複数のポリヌクレオチドの配列に適用する。場合によっては、デコーダ・モジュールは、リアセンブリ、エラー訂正を実行し、デジタル情報（例えば、バイナリ・データ）を出力する。場合によっては、デジタル情報を含む出力は、オペレーティング・システムおよび／またはファイル・システムに転送される。出力は、デジタル情報を提供するために、グラフィカル・ユーザ・インターフェイス（ＧＵＩ：ｇｒａｐｈｉｃａｌ　ｕｓｅｒ　ｉｎｔｅｒｆａｃｅ）などのディスプレイ、または本明細書に記載されたディスプレイなどの任意の他の適切なディスプレイ上で提供されてよい。場合によっては、コントローラ１６３５は、本明細書に記載されたソフトウェア・モジュールなどの１つまたは複数のソフトウェア・モジュールに実装される。場合によっては、コントローラ１６３５は、本明細書に記載されたオペレーティング・システムなどのオペレーティング・システムからのコマンドに応答する。

【０２２７】

特定の定義

特に定義されない限り、本明細書において使用されるすべての技術用語は、本主題が属する技術の当業者によって一般的に理解されている意味と同じ意味を有する。

【０２２８】

本開示全体を通じて、数値的特徴は範囲形式で提示される。範囲形式における説明が、単に便宜および簡潔さのためであり、任意の実施形態の範囲に対する柔軟性のない制限であると解釈されるべきではないということが、理解されるべきである。したがって、範囲の説明は、文脈が特に明確に指示しない限り、下限の単位の１０分の１までの、その範囲内の個別の数値に加えて、すべての可能な部分範囲を特に開示していると見なされるべきである。例えば、１～６などの範囲の説明は、その範囲内の個別の値（例えば、１．１、２、２．３、５、および５．９）に加えて、１～３、１～４、１～５、２～４、２～６、３～６などの部分範囲を特に開示していると見なされるべきである。このことは、範囲の幅に関わらず当てはまる。これらの介在する範囲の上限および下限は、記載された範囲内で特に除外される任意の制限を条件として、より小さい範囲に独立して含まれてよく、本

発明にやはり包含される。記載された範囲が制限のうちの１つまたは両方を含む場合、文脈が特に明確に指示しない限り、それらの含まれる制限のいずれかまたは両方を除外する範囲も本発明に含まれる。

【０２２９】

本明細書において使用される用語は、特定の実施形態を説明することのみを目的としており、任意の実施形態を限定することを意図されていない。本明細書において使用されるとき、単数形「ａ」、「ａｎ」、および「ｔｈｅ」は、文脈が特に明確に示さない限り、複数形も含むよう意図されている。「備える」または「備えている」あるいはその両方の用語は、本明細書において使用されるとき、記載された特徴、整数、ステップ、動作、要素、および／またはコンポーネンの存在を示すが、１つまたは複数の他の特徴、整数、ステップ、動作、要素、コンポーネント、および／またはこれらのグループの存在または追加を除外していないということが、さらに理解されるであろう。本明細書において使用されるとき、「および／または」という用語は、関連するリストされた項目のうちの１つまたは複数のありとあらゆる組み合わせを含む。

【０２３０】

本明細書全体を通じて、「一部の事例」、「さらなる事例」、または「特定の事例」への言及は、事例に関連して説明される特定の特徴、構造、または特性が少なくとも１つの事例に含まれるということを意味する。したがって、本明細書全体を通じてさまざまな場所に現れる「場合によっては」、または「さらなる事例では」、または「特定の事例では」という語句は、必ずしもすべてが同じ事例を参照しているわけではない。さらに、特定の特徴、構造、または特性は、１つまたは複数の事例において、任意の適切な方法で組み合わせられてよい。

【０２３１】

特に述べられるか、または文脈から明白でない限り、本明細書において使用されるとき、数値または数値の範囲を参照する「約」という用語は、述べられた数値およびその数値の±１０％、または範囲に関してリストされた値の場合、リストされた下限より１０％低い数値、およびリストされた上限より１０％高い数値を意味すると理解される。

【０２３２】

本明細書において使用されるとき、「予め選択された配列」、「事前に定義された配列」、または「既定の配列」という用語は、同じ意味で使用される。この用語は、ポリマーの配列が既知であり、ポリマーの合成またはアセンブリの前に選択されることを意味する。特に、本発明のさまざまな態様は、主に核酸分子の準備に関して本明細書に記載され、ポリヌクレオチドの配列は既知であり、核酸分子の合成またはアセンブリの前に選択される。

【０２３３】

本明細書において使用されるとき、「ハッシュ」または「複数のハッシュ」という用語は、通常、ハッシュ関数から出力される固定長の列のことを指してよい。ハッシュ関数は、通常、任意の長さの入力を受け取り、固定長を有する出力に変換する関数を含んでよい。場合によっては、入力は、ハッシュ関数を通過してハッシュを生成することができる、１つまたは複数のビットであってよい。場合によっては、ハッシュ関数は決定的であってよく、ハッシュされた出力から入力をリバースエンジニアリングすることは、不可能であってよい。入力をハッシュ関数に供給する行為は、「ハッシュする」と呼ばれることがある。

【０２３４】

「シンボル」という用語は、本明細書において使用されるとき、通常、デジタル情報の単位の表現のことを指す。デジタル情報は、１つまたは複数のシンボルに分割されるか、または変換される。１つの例では、シンボルはビットであってよく、ビットは数値を有してよい。一部の例では、シンボルは「０」または「１」の値を有してよい。一部の例では、デジタル情報は、シンボルの配列またはシンボルの列として表されてよい。一部の例では、シンボルの配列またはシンボルの列は、バイナリ・データを含んでよい。

【０２３５】
本明細書に記載されたポリヌクレオチド配列は、特に述べられない限り、ＤＮＡもしくはＲＮＡ、またはそれらの類似体もしくは誘導体を含んでよい。本明細書において使用されるとき、核酸、ヌクレオチド、ポリヌクレオチド、オリゴヌクレオチド、オリゴ、オリゴ核酸という用語は、ヌクレオシド・モノマーのポリマーを表すために、全体を通して同義的に使用される。本明細書において使用されるとき、核酸配列、ポリヌクレオチド配列、ポリヌクレオチド配列、オリゴヌクレオチド配列、オリゴ配列、またはオリゴ核酸配列という用語も、ヌクレオシド・モノマーのポリマーの配列を表すために、全体を通して同義的に使用される。場合によっては、核酸は、リン酸塩または硫黄含有結合を介して接続される。核酸は、場合によっては、ＤＮＡ、ＲＮＡ、非標準核酸、非天然核酸、または他のヌクレオシドを含む。場合によっては、ヌクレオチドは、非標準塩基、糖、または他の部分を含む。場合によっては、ヌクレオチドは、伸長反応を防ぐように構成された終端を含む。場合によっては、そのような終端は、成長する鎖への後続のヌクレオチドの追加の前に除去される。

【０２３６】
コンピューティング・システム
図１０を参照すると、コンピュータ・システム１０００（例えば、処理システムまたはコンピューティング・システム）を含む例示的なマシンを示すブロック図が示されており、コンピュータ・システム１０００内で、命令のセットが実行されて、デバイスに、本開示の静的コード・スケジューリングのための態様および／または方法のうちの任意の１つまたは複数を実行させることができる。図１０のコンポーネントは、単なる例であり、特定の実施形態を実装する任意のハードウェア、ソフトウェア、組み込みロジック・コンポーネント、または２つ以上のそのようなコンポーネントの組み合わせの使用または機能の範囲を限定しない。

【０２３７】
図１０に示されているようなコンピュータ・システムを備えるプラットフォームは、シンボルのセットとして表されたデータを別のシンボルのセットにエンコードするために使用されてよい。場合によっては、コンピュータ・システムは、プログラムを使用して、シンボルの第１の列をシンボルの第２の列に変換する。場合によっては、コンピュータ・システムは、プログラムを実行して、データを複数のポリヌクレオチド配列に変換するか、複数のポリヌクレオチド配列をデータに変換するか、またはその両方を実行する。例として、図１０に概して示されているようなコンピューティング・システムは、１つまたは複数のソフトウェア・プログラムを実行して、（情報のアイテムを表す）シンボルの列をポリヌクレオチド配列にエンコードするため（例えば、図１、図３～６、図１２、または図１４～１５に示された方法のうちの１つまたは複数）、ポリヌクレオチド配列をデコードしてシンボルの列に戻すため（例えば、図２、図７～９、または図１３に示された方法のうちの１つまたは複数）、またはその両方のために使用されてよい。より詳細には、データは、「０」および「１」の２進値などの数値シンボルとして表されてよく、コンピュータ・システム１０００は、コンピュータ・プログラム（例えば、内部コーデック、外部コーデック、またはその両方）を実行して、データを複数のポリヌクレオチド配列に変換してよい。場合によっては、コンピュータ・システムは、プログラムを実行して、第１の１つまたは複数のポリヌクレオチド配列を第２の１つまたは複数のポリヌクレオチド配列に変換する。

【０２３８】
データをエンコードするためのプラットフォームは、シンセサイザ、シーケンサ、ストレージ・ユニット、またはこれらの任意の組み合わせなどの、１つまたは複数のコンポーネントをさらに含むことができる。そのようなプラットフォームでは、場合によっては、コンピュータ・システム１０００は、シンセサイザ、シーケンサ、ストレージ・ユニット、またはこれらの任意の組み合わせなどの１つまたは複数のコンポーネントのうちの任意の１つと、電子通信する。場合によっては、１つまたは複数のコンポーネントは、コンピ

ュータ・システムに動作可能にリンクされ、コンピュータを介してローカルまたはリモートのいずれかで任意選択的に自動化される。さまざまな場合において、本明細書に記載された方法およびシステムは、コンピュータ・システムのプラットフォームの１つまたは複数のコンポーネントの動作およびその使用のためのソフトウェア・プログラムをさらに含む。したがって、材料堆積デバイスの動き、分配動作、および真空作動を調整すること、および同期させることなどの、分配／真空／補充機能の同期のためのコンピュータ化された制御は、本明細書において提供される開示の範囲内である。場合によっては、コンピュータ・システムは、ユーザによって指定された塩基配列と材料堆積デバイスの位置の間をインターフェイスして、正しい構成要素および／または試薬を基板の指定された領域（例えば、特定の遺伝子座）に供給するようにプログラムされる。さらに、図１０に示されたシステムなどのコンピュータ・システムは、プラットフォーム内の１つまたは複数のコンポーネントを監視するために使用されてよい。例えば、コンピュータ・システムは、コンポーネントに統合されたか、またはコンポーネントに接続されたセンサからの１つまたは複数のセンサ・データを監視するために使用されてよい。場合によっては、コンピュータ・システムは、プログラムを使用して、圧力、体積、流量、温度、真空、方位角、湿度、または本明細書に記載されたシステムおよびプラットフォーム内で測定され得る任意の他の物理的パラメータなどの、１つまたは複数のパラメータにおける不規則性を監視および検出する。プログラムを含むコンピュータ・システムは、１つまたは複数のセンサ・データ内のパターンを分析し、不規則性が検出された場合、あるいはデータまたはデータの組み合わせがしきい値（例えば、既定のしきい値または動的しきい値）の範囲外にある場合に、ＨＭＩを介して任意選択的にユーザに警告してよい。

【０２３９】

プログラムは、本明細書において提供されたコンピュータ・システム上で実行されてよい。場合によっては、プログラムは、統計アルゴリズムまたは機械学習アルゴリズムを含む。場合によっては、機械学習（ＭＬ）を含むアルゴリズムは、本明細書に記載された機能または動作を実行するようにトレーニングされる。場合によっては、アルゴリズムは、分類および／またはクラスタ化のための古典的なＭＬアルゴリズム（例えば、Ｋ平均クラスタ化、平均シフト・クラスタ化、ノイズを伴うアプリケーションの密度ベースの空間クラスタ化（ＤＢＳＣＡＮ：ｄｅｎｓｉｔｙ－ｂａｓｅｄ ｓｐａｔｉａｌ ｃｌｕｓｔｅｒｉｎｇ ｏｆ ａｐｐｌｉｃａｔｉｏｎｓ ｗｉｔｈ ｎｏｉｓｅ）、期待値最大化（ＥＭ：ｅｘｐｅｃｔａｔｉｏｎ－ｍａｘｉｍｉｚａｔｉｏｎ）クラスタ化、凝集型階層クラスタ化、ロジスティック回帰、単純ベイズ、Ｋ最近傍法、ランダム・フォレストまたは決定木、勾配ブースティング、サポート・ベクター・マシン（ＳＶＭ：ｓｕｐｐｏｒｔ ｖｅｃｔｏｒ ｍａｃｈｉｎｅｓ）、あるいはこれらの組み合わせ）を含む。

【０２４０】

場合によっては、アルゴリズムは、１つまたは複数のニューラル・ネットワークなどの、層を含んでいる学習アルゴリズムを含む。ニューラル・ネットワークは、入力データの変換または翻訳などの機能を実行することができる、ネットワーク内の接続されたノードを含んでよい。一部の例では、特定のノードからの出力が別のノードへの入力として渡されてよい。一部の実施形態では、ネットワーク内のノードは、入力ユニット、隠れユニット、出力ユニット、またはこれらの組み合わせを含んでよい。場合によっては、入力ノードは、１つまたは複数の隠れユニットに接続されてよい。場合によっては、１つまたは複数の隠れユニットが出力ユニットに接続されてよい。ノードは入力を受け取ってよく、活性化関数に基づいて出力を生成してよい。一部の実施形態では、入力または出力は、テンソル、行列、ベクトル、配列、またはスカラーであってよい。一部の実施形態では、活性化関数は、整流線形ユニット（ＲｅＬＵ：Ｒｅｃｔｉｆｉｅｄ Ｌｉｎｅａｒ Ｕｎｉｔ）活性化関数、シグモイド活性化関数、または双曲線正接活性化関数であってよい。一部の実施形態では、活性化機能は、ソフトマックス活性化関数であってよい。ノード間の接続は、特定のノードへの入力データを調整する（例えば、入力データを有効化するか、または入力データを無効化する）ための重みをさらに含んでよい。一部の実施形態では、重

みは、ニューラル・ネットワークによって学習されてよい。一部の実施形態では、ニューラル・ネットワークは、勾配ベースの最適化を使用してトレーニングされてよい。場合によっては、勾配ベースの最適化は、１つまたは複数の損失関数を含んでよい。一部の例では、勾配ベースの最適化は、共役勾配降下法、確率的勾配降下法、またはこれらの変形（例えば、適応モーメント推定（Ａｄａｍ））であってよい。さらなる例では、勾配ベースの最適化における勾配は、逆伝搬を使用して計算されてよい。一部の実施形態では、ノードがグラフに編成されて、ネットワーク（例えば、グラフ・ニューラル・ネットワーク）を生成してよい。一部の実施形態では、ノードが１つまたは複数の層に編成されて、ネットワーク（例えば、フィード・フォワード・ニューラルネットワーク、畳み込みニューラル・ネットワーク（ＣＮＮ：ｃｏｎｖｏｌｕｔｉｏｎａｌ ｎｅｕｒａｌ ｎｅｔｗｏｒｋｓ）、リカレント・ニューラル・ネットワーク（ＲＮＮ：ｒｅｃｕｒｒｅｎｔ ｎｅｕｒａｌ ｎｅｔｗｏｒｋｓ）など）を生成してよい。場合によっては、ニューラル・ネットワークは、２つ以上の層からなる深層ニューラル・ネットワークであってよい。

【０２４１】

場合によっては、ニューラル・ネットワークは、１つまたは複数の回帰層を含んでよい。一部の例では、１つまたは複数の回帰層は、１つまたは複数の長短期記憶（ＬＳＴＭ：ｌｏｎｇ ｓｈｏｒｔ－ｔｅｒｍ ｍｅｍｏｒｙ）層またはゲート付き回帰型ユニット（ＧＲＵ：ｇａｔｅｄ ｒｅｃｕｒｒｅｎｔ ｕｎｉｔ）であってよく、逐次的データ分類およびクラスタ化を実行してよい。一部の実施形態では、ニューラル・ネットワークは、１つまたは複数の畳み込み層を含んでよい。入力および出力は、特徴マップ（または活性化マップ）と呼ばれることがある、データ・セット（例えば、特徴）内の変数または属性を表すテンソルであってよい。場合によっては、畳み込みは、１次元（１Ｄ：ｏｎｅ ｄｉｍｅｎｓｉｏｎａｌ）畳み込み、２次元（２Ｄ：ｔｗｏ ｄｉｍｅｎｓｉｏｎａｌ）畳み込み、３次元（３Ｄ：ｔｈｒｅｅ ｄｉｍｅｎｓｉｏｎａｌ）畳み込み、またはこれらの任意の組み合わせであってよい。さらなる場合において、畳み込みは、１Ｄ転置畳み込み、２Ｄ転置畳み込み、３Ｄ転置畳み込み、またはこれらの任意の組み合わせであってよい。一部の例では、１次元畳み込み層は、並列畳み込みによって時系列を分類することができるため、時系列データに適していることがある。一部の例では、畳み込み層は、本明細書に記載されたシステムの１つまたは複数のコンポーネントからの信号（例えば、センサ・データ）を分析するために使用されてよい。

【０２４２】

ニューラル・ネットワーク内の層は、畳み込み層の前または後に１つまたは複数のプーリング層をさらに含んでよい。１つまたは複数のプーリング層は、行列の領域を集約するフィルタを使用して特徴マップの次元を減らすことができる。これによって、出力の数をダウンサンプリングし、このようにして、ニューラル・ネットワークに必要とされるパラメータおよび計算リソースを減らすことができる。一部の実施形態では、１つまたは複数のプーリング層は、最大プーリング、最小プーリング、平均プーリング、グローバル・プーリング、ノルム・プーリング、またはこれらの組み合わせであってよい。最大プーリングは、行列の領域内の最大値のみを取得することによってデータの次元を減らすことができ、有意な特徴を捕捉するのに役立つ。一部の実施形態では、１つまたは複数のプーリング層は、１次元（１Ｄ）、２次元（２Ｄ）、３次元（３Ｄ）、またはこれらの任意の組み合わせであってよい。ニューラル・ネットワークは、次の層に渡される入力を平坦化することができる１つまたは複数の平坦化層をさらに含んでよい。場合によっては、入力は、１次元配列に縮小することによって平坦化されてよい。平坦化された入力は、オブジェクトの分類（例えば、本明細書に記載されたシステムにおける信号（例えば、センサ・データ）の分類）を出力するために使用されてよい。ニューラル・ネットワークは、１つまたは複数のドロップアウト層をさらに含んでよい。ドロップアウト層は、ニューラル・ネットワークのトレーニング中に（例えば、２値分類またはマルチクラス分類を実行するために）使用されてよい。１つまたは複数のドロップアウト層は、特定の重みをランダムに０に設定することができ、特徴マップ内の対応する要素を０に設定することができるため、

ニューラル・ネットワークは過学習を防ぐことができる。ニューラル・ネットワークは、完全に接続されたネットワークを含む１つまたは複数の緻密層をさらに含んでよい。緻密層では、完全に接続されたネットワークを介して情報が渡されてオブジェクトの予測分類を生成することができ、エラーが計算され得る。一部の実施形態では、エラーは、予測を改善するために逆伝播されてよい。１つまたは複数の緻密層は、数値のベクトルを確率のベクトルに変換することができるソフトマックス活性化関数を含んでよい。これらの確率は、その後、本明細書に記載されたシステムからの信号（例えば、センサ・データ）または（例えば、コーデックの一部として）デコード中の可能性の高い核酸塩基の分類などの、分類において使用されてよい。

【０２４３】

コンピュータ・システム１０００は、バス１０４０を介して互いに、および他のコンポーネントと通信する、１つまたは複数のプロセッサ１００１、メモリ１００３、およびストレージ１００８を含んでよい。バス１０４０は、ディスプレイ１０３２、１つまたは複数の入力デバイス１０３３（例えば、キーパッド、キーボード、マウス、スタイラスなどを含んでよい）、１つまたは複数の出力デバイス１３４、１つまたは複数のストレージ・デバイス１０３５、およびさまざまな有形のストレージ媒体１０３６をリンクしてもよい。これらの要素のすべては、直接、または１つまたは複数のインターフェイスまたはアダプタを介して、バス１０４０にインターフェイスしてよい。例えば、さまざまな有形のストレージ媒体１０３６は、ストレージ媒体インターフェイス１０２６を介してバス１０４０とインターフェイスすることができる。コンピュータ・システム１０００は、１つまたは複数の集積回路（ＩＣ：ｉｎｔｅｇｒａｔｅｄ　ｃｉｒｃｕｉｔｓ）、プリント回路基板（ＰＣＢ：ｐｒｉｎｔｅｄ　ｃｉｒｃｕｉｔ　ｂｏａｒｄｓ）、モバイル・ハンドヘルド・デバイス（携帯電話またはＰＤＡなど）、ラップトップまたはノートブック・コンピュータ、分散コンピュータ・システム、コンピューティング・グリッド、あるいはサーバを含むが、これらに限定されない、任意の適切な物理的形態を有してよい。

【０２４４】

コンピュータ・システム１０００は、機能を実行する１つまたは複数のプロセッサ１００１（例えば、中央処理装置（ＣＰＵ）、汎用グラフィックス処理ユニット（ＧＰＧＰＵ：ｇｅｎｅｒａｌ　ｐｕｒｐｏｓｅ　ｇｒａｐｈｉｃｓ　ｐｒｏｃｅｓｓｉｎｇ　ｕｎｉｔｓ）、または量子処理ユニット（ＱＰＵ：ｑｕａｎｔｕｍ　ｐｒｏｃｅｓｓｉｎｇ　ｕｎｉｔｓ））を含む。プロセッサ１００１は、命令、データ、またはコンピュータ・アドレスの一時的なローカル・ストレージのためのキャッシュ・メモリ・ユニット１００２を任意選択的に含む。プロセッサ１００１は、コンピュータ可読命令の実行を支援するように構成される。コンピュータ・システム１０００は、プロセッサ１００１が、１つまたは複数の有形のコンピュータ可読ストレージ媒体に具現化された非一時的なプロセッサ実行可能命令を実行する結果として、メモリ１００３、ストレージ１００８、ストレージ・デバイス１０３５、および／またはストレージ媒体１０３６などの、図１０に示されたコンポーネントの機能を提供してよい。コンピュータ可読媒体は、特定の実施形態を実装するソフトウェアを格納してよく、プロセッサ１００１は、ソフトウェアを実行してよい。メモリ１００３は、ネットワーク・インターフェイス１０２０などの適切なインターフェイスを介して、（マス・ストレージ・デバイス１０３５、１０３６などの）１つまたは複数の他のコンピュータ可読媒体から、または１つまたは複数の他のソースから、ソフトウェアを読み取ってよい。ソフトウェアは、プロセッサ１００１に、本明細書において説明されたか、または示された１つまたは複数のプロセスまたは１つまたは複数のプロセスの１つまたは複数のステップを実行させてよい。そのようなプロセスまたはステップを実行することは、メモリ１００３に格納されたデータ構造を定義すること、およびソフトウェアによって指示されたとおりにデータ構造を変更することを含んでよい。

【０２４５】

メモリ１００３は、ランダム・アクセス・メモリコン・ポーネント（例えば、ＲＡＭ（ｒａｎｄｏｍ　ａｃｃｅｓｓ　ｍｅｍｏｒｙ）１００４）（例えば、スタティックＲＡＭ

（ＳＲＡＭ）、ダイナミックＲＡＭ（ＤＲＡＭ：ｄｙｎａｍｉｃ　ＲＡＭ）、強誘電体ランダム・アクセス・メモリ（ＦＲＡＭ（登録商標）：ｆｅｒｒｏｅｌｅｃｔｒｉｃ　ｒａｎｄｏｍ　ａｃｃｅｓｓ　ｍｅｍｏｒｙ）、相変化ランダム・アクセス・メモリ（ＰＲＡＭ：ｐｈａｓｅ－ｃｈａｎｇｅ　ｒａｎｄｏｍ　ａｃｃｅｓｓ　ｍｅｍｏｒｙ）など）、読み取り専用メモリ・コンポーネント（例えば、ＲＯＭ１００５）、およびこ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、さまざまなコンポーネント（例えば、機械可読媒体）を含んでよい。ＲＯＭ１００５は、データおよび命令をプロセッサ１００１に一方向に伝達するように動作してよく、ＲＡＭ１００４は、データおよび命令をプロセッサ１００１と双方向に伝達するように動作してよい。ＲＯＭ１００５およびＲＡＭ１００４は、以下に説明される任意の適切な有形のコンピュータ可読媒体を含んでよい。１つの例では、起動中などにコンピュータ・システム１０００内の要素間で情報を転送するのに役立つ基本ルーチンを含んでいる基本入出力システム１００６（ＢＩＯＳ：ｂａｓｉｃ　ｉｎｐｕｔ／ｏｕｔｐｕｔ　ｓｙｓｔｅｍ）が、メモリ１００３に格納されてよい。

【０２４６】

固定ストレージ１００８は、任意選択的にストレージ制御ユニット１００７を介して、プロセッサ１００１に双方向に接続される。固定ストレージ１００８は、追加のストレージ容量を提供し、本明細書に記載された任意の適切な有形のコンピュータ可読媒体を含んでもよい。ストレージ１００８は、オペレーティング・システム１００９、実行ファイル１０１０、データ１０１１、アプリケーション１０１２（アプリケーション・プログラム）などを格納するために使用されてよい。ストレージ１００８は、光ディスク・ドライブ、半導体メモリ・デバイス（例えば、フラッシュベースのシステム）、またはこれらのいずれかの組み合わせを含むこともできる。ストレージ１００８内の情報は、適切な場合において、仮想メモリとしてメモリ１００３に組み込まれてよい。

【０２４７】

１つの例では、ストレージ・デバイス１０３５は、ストレージ・デバイス・インターフェイス１０２５を介して（例えば、外部ポート・コネクタ（図示されていない）を介して）コンピュータ・システム１０００と取り外し可能にインターフェイスされてよい。特に、ストレージ・デバイス１０３５および関連する機械可読媒体は、コンピュータ・システム１０００のための機械可読命令、データ構造、プログラム・モジュール、および／または他のデータの不揮発性および／または揮発性のストレージを提供してよい。１つの例では、ソフトウェアは、ストレージ・デバイス１０３５上の機械可読媒体内に完全にまたは部分的に存在してよい。別の例では、ソフトウェアは、プロセッサ１００１内に完全にまたは部分的に存在してよい。

【０２４８】

バス１０４０は、多種多様なサブシステムを接続する。本明細書では、バスへの言及は、適切な場合に共通の機能を果たす、１つまたは複数のデジタル信号線を包含してよい。バス１０４０は、さまざまなバス・アーキテクチャのいずれかを使用する、メモリ・バス、メモリ・コントローラ、周辺バス、ローカル・バス、およびこ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、複数の種類のバス構造のいずれかであってよい。制限としてではなく、そのようなアーキテクチャの例としては、Ｉｎｄｕｓｔｒｙ　Ｓｔａｎｄａｒｄ　Ａｒｃｈｉｔｅｃｔｕｒｅ（ＩＳＡ）バス、拡張ＩＳＡ（ＥＩＳＡ：Ｅｎｈａｎｃｅｄ　ＩＳＡ）バス、Ｍｉｃｒｏ　Ｃｈａｎｎｅｌ　Ａｒｃｈｉｔｅｃｔｕｒｅ（ＭＣＡ）バス、Ｖｉｄｅｏ　Ｅｌｅｃｔｒｏｎｉｃｓ　Ｓｔａｎｄａｒｄｓ　Ａｓｓｏｃｉａｔｉｏｎローカル・バス（ＶＬＢ）、Ｐｅｒｉｐｈｅｒａｌ　Ｃｏｍｐｏｎｅｎｔ　Ｉｎｔｅｒｃｏｎｎｅｃｔ（ＰＣＩ）バス、ＰＣＩ－Ｅｘｐｒｅｓｓ（ＰＣＩ－Ｘ）バス、Ａｃｃｅｌｅｒａｔｅｄ　Ｇｒａｐｈｉｃｓ　Ｐｏｒｔ（ＡＧＰ）バス、ＨｙｐｅｒＴｒａｎｓｐｏｒｔ（ＨＴＸ）バス、Ｓｅｒｉａｌ　Ａｄｖａｎｃｅｄ　Ｔｅｃｈｎｏｌｏｇｙ　Ａｔｔａｃｈｍｅｎｔ（ＳＡＴＡ）バス、およびこれらの任意の組み合わせが挙げられる。

【０２４９】

コンピュータ・システム１０００は、入力デバイス１０３３を含んでもよい。１つの例では、コンピュータ・システム１０００のユーザは、入力デバイス１０３３を介してコマンドおよび／または他の情報をコンピュータ・システム１０００に入力してよい。入力デバイス１０３３の例としては、英数字入力デバイス（例えば、キーボード）、ポインティング・デバイス（例えば、マウスまたはタッチパッド）、タッチパッド、タッチ・スクリーン、マルチタッチ・スクリーン、ジョイスティック、スタイラス、ゲームパッド、音声入力デバイス（例えば、マイクロホン、音声応答システムなど）、光学スキャナ、ビデオまたは静止画捕捉デバイス（例えば、カメラ）、およびこれらの任意の組み合わせが挙げられるが、これらに限定されない。一部の実施形態では、入力デバイスは、Ｋｉｎｅｃｔ、Ｌｅａｐ　Ｍｏｔｉｏｎなどである。入力デバイス１０３３は、シリアル、パラレル、ゲーム・ポート、ＵＳＢ、ＦＩＲＥＷＩＲＥ（登録商標）、ＴＨＵＮＤＥＲＢＯＬＴ（登録商標）、またはこ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、さまざまな入力インターフェイス１０２３（例えば、入力インターフェイス１０２３）のいずれかを介してバス１０４０にインターフェイルされてよい。

【０２５０】

特定の実施形態では、コンピュータ・システム１０００がネットワーク１０３０に接続されている場合、コンピュータ・システム１０００は、ネットワーク１０３０に接続されている他のデバイス、詳細には、モバイル・デバイスおよびエンタープライズ・システム、分散コンピューティング・システム、クラウド・ストレージ・システム、クラウド・コンピューティング・システムなどと通信してよい。クラウド・コンピューティング・システムは、プライベート・クラウド、パブリック・クラウド、ハイブリッド・クラウド、マルチクラウド、またはこれらの任意の組み合わせを含むことができる。クラウド・コンピューティング・システムは、サービスとしてのインフラストラクチャ（ＩａａＳ）、サービスとしてのプラットフォーム（ＰａａＳ）、サービスとしてのソフトウェア（ＳａａＳ）、またはこれらの任意の組み合わせを含むことができる。コンピュータ・システム１０００との間の通信は、ネットワーク・インターフェイス１０２０を介して送信されてよい。例えば、ネットワーク・インターフェイス１０２０は、ネットワーク１０３０から１つまたは複数のパケット（インターネット・プロトコル（ＩＰ：Ｉｎｔｅｒｎｅｔ　Ｐｒｏｔｏｃｏｌ）パケットなど）の形態で着信通信（他のデバイスからの要求または応答など）を受信してよく、コンピュータ・システム１０００は、処理のために着信通信をメモリ１００３に格納してよい。コンピュータ・システム１０００は、同様に、メモリ１００３内の１つまたは複数のパケットの形態で、ネットワーク・インターフェイス１０２０からネットワーク１０３０に伝達される発信通信（他のデバイスへの要求または応答など）を格納してよい。プロセッサ１００１は、メモリ１００３に格納されたこれらの通信パケットに、処理のためにアクセスしてよい。

【０２５１】

ネットワーク・インターフェイス１０２０の例としては、ネットワーク・インターフェイス・カード、モデム、およびこれらの任意の組み合わせが挙げられるが、これらに限定されない。ネットワーク１０３０またはネットワーク・セグメント１０３０の例としては、分散コンピューティング・システム、クラウド・コンピューティング・システム、広域ネットワーク（ＷＡＮ：ｗｉｄｅ　ａｒｅａ　ｎｅｔｗｏｒｋ）（例えば、インターネット、エンタープライズ・ネットワーク）、ローカル・エリア・ネットワーク（ＬＡＮ：ｌｏｃａｌ　ａｒｅａ　ｎｅｔｗｏｒｋ）（例えば、オフィス、ビル、キャンパス、または他の比較的小さい地理的空間に関連付けられたネットワーク）、電話網、２つのコンピューティング・デバイス間の直接接続、ピアツーピア・ネットワーク、およびこれらの任意の組み合わせが挙げられるが、これらに限定されない。ネットワーク１０３０などのネットワークは、有線通信モードおよび／またはワイヤレス通信モードを採用してよい。一般に、任意のネットワーク・トポロジが使用されてよい。

【０２５２】

情報およびデータは、ディスプレイ１０３２を介して表示され得る。ディスプレイ１０

３２の例としては、ブラウン管（ＣＲＴ：ｃａｔｈｏｄｅ　ｒａｙ　ｔｕｂｅ）、液晶ディスプレイ（ＬＣＤ：ｌｉｑｕｉｄ　ｃｒｙｓｔａｌ　ｄｉｓｐｌａｙ）、薄膜トランジスタ液晶ディスプレイ（ＴＦＴ－ＬＣＤ：ｔｈｉｎ　ｆｉｌｍ　ｔｒａｎｓｉｓｔｏｒ　ｌｉｑｕｉｄ　ｃｒｙｓｔａｌ　ｄｉｓｐｌａｙ）、パッシブマトリックスＯＬＥＤ（ＰＭＯＬＥＤ：ｐａｓｓｉｖｅ－ｍａｔｒｉｘ　ＯＬＥＤ）またはアクティブマトリックスＯＬＥＤ（ＡＭＯＬＥＤ：ａｃｔｉｖｅ－ｍａｔｒｉｘ　ＯＬＥＤ）ディスプレイなどの有機液晶ディスプレイ（ＯＬＥＤ：ｏｒｇａｎｉｃ　ｌｉｑｕｉｄ　ｃｒｙｓｔａｌ　ｄｉｓｐｌａｙ）、プラズマ・ディスプレイ、およびこれらの任意の組み合わせが挙げられるが、これらに限定されない。ディスプレイ１０３２は、バス１０４０を介して、プロセッサ１００１、メモリ１００３、および固定ストレージ１００８、ならびに入力デバイス１０３３などの他のデバイスに、インターフェイスすることができる。ディスプレイ１０３２は、ビデオ・インターフェイス１０２２を介してバス１０４０にリンクされ、ディスプレイ１０３２とバス１０４０の間のデータの輸送は、グラフィック制御１０２１を介して制御され得る。一部の実施形態では、ディスプレイはビデオ・プロジェクタである。一部の実施形態では、ディスプレイは、ＶＲヘッドセットなどのヘッドマウント・ディスプレイ（ＨＭＤ：ｈｅａｄ－ｍｏｕｎｔｅｄ　ｄｉｓｐｌａｙ）である。さらなる実施形態では、適切なＶＲヘッドセットは、非限定的な例として、ＨＴＣ　Ｖｉｖｅ、Ｏｃｕｌｕｓ　Ｒｉｆｔ、Ｓａｍｓｕｎｇ　Ｇｅａｒ　ＶＲ、Ｍｉｃｒｏｓｏｆｔ　ＨｏｌｏＬｅｎｓ、Ｒａｚｅｒ　ＯＳＶＲ、ＦＯＶＥ　ＶＲ、Ｚｅｉｓｓ　ＶＲ　Ｏｎｅ、Ａｖｅｇａｎｔ　Ｇｌｙｐｈ、Ｆｒｅｅｆｌｙ　ＶＲヘッドセットなどを含む。さらなる実施形態では、ディスプレイは、本明細書において開示されたデバイスなどのデバイスの組み合わせである。

【０２５３】

ディスプレイ１０３２に加えて、コンピュータ・システム１０００は、オーディオ・スピーカ、プリンタ、ストレージ・デバイス、およびこ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、１つまたは複数の他の周辺出力デバイス１０３４を含んでよい。そのような周辺出力デバイスは、出力インターフェイス１０２４を介してバス１０４０に接続されてよい。出力インターフェイス１０２４の例としては、シリアル・ポート、パラレル接続、ＵＳＢポート、ＦＩＲＥＷＩＲＥ（登録商標）ポート、ＴＨＵＮＤＥＲＢＯＬＴ（登録商標）ポート、およびこれらの任意の組み合わせが挙げられるが、これらに限定されない。

【０２５４】

加えて、または代替として、コンピュータ・システム１０００は、回路においてハード配線されたか、または他の方法で具現化された論理の結果として機能を提供してよく、この回路は、本明細書において説明されたか、または示された１つまたは複数のプロセスまたは１つまたは複数のプロセスの１つまたは複数のステップを実行するために、ソフトウェアの代わりに、またはソフトウェアと共に動作してよい。本開示におけるソフトウェアへの言及は論理を包含してよく、論理への言及はソフトウェアを包含してよい。さらに、コンピュータ可読媒体への言及は、適切な場合に、実行のためのソフトウェアを格納する回路（ＩＣなど）、実行のための論理を具体化する回路、またはその両方を包含してよい。本開示は、ハードウェア、ソフトウェア、またはその両方の任意の適切な組み合わせを包含する。

【０２５５】

当業者は、本明細書において開示された実施形態に関連して説明されたさまざまな例示的な論理ブロック、モジュール、回路、およびアルゴリズム・ステップが、電子ハードウェア、コンピュータ・ソフトウェア、またはその両方の組み合わせとして実装されてよいということを理解するであろう。ハードウェアおよびソフトウェアのこの互換性を明確に示すために、さまざまな例示的なコンポーネント、ブロック、モジュール、回路、およびステップは、それらの機能の観点から上記では一般的に説明された。

【０２５６】

本明細書において開示された実施形態に関連して説明されたさまざまな例示的な論理ブロック、モジュール、および回路は、汎用プロセッサ、デジタル信号プロセッサ（ＤＳＰ：ｄｉｇｉｔａｌ　ｓｉｇｎａｌ　ｐｒｏｃｅｓｓｏｒ）、特定用途向け集積回路（ＡＳＩＣ）、フィールド・プログラマブル・ゲート・アレイ（ＦＰＧＡ）または他のプログラマブル・ロジック・デバイス、個別のゲートまたはトランジスタ論理、個別のハードウェア・コンポーネント、あるいは本明細書に記載された機能を実行するように設計されたこれらの任意の組み合わせを使用して、実装または実行されてよい。汎用プロセッサはマイクロプロセッサであってよいが、代替として、プロセッサは、任意の従来のプロセッサ、コントローラ、マイクロコントローラ、またはステート・マシンであってよい。プロセッサは、コンピューティング・デバイスの組み合わせ、例えば、ＤＳＰとマイクロプロセッサの組み合わせ、複数のマイクロプロセッサ、ＤＳＰコアと組み合わせた１つまたは複数のマイクロプロセッサ、または任意の他のそのような構成として、実装されてもよい。

【０２５７】

本明細書において開示された実施形態に関連して説明された方法またはアルゴリズムのステップは、直接ハードウェア内で、１つまたは複数のプロセッサによって実行されるソフトウェア・モジュール内で、またはこれら２つの組み合わせで、具現化されてよい。ソフトウェア・モジュールは、ＲＡＭメモリ、フラッシュ・メモリ、ＲＯＭメモリ、ＥＰＲＯＭメモリ、ＥＥＰＲＯＭメモリ、レジスタ、ハード・ディスク、取り外し可能なディスク、ＣＤ－ＲＯＭ、または当技術分野において知られたストレージ媒体の任意の他の形態に存在してよい。例示的なストレージ媒体は、プロセッサがストレージ媒体から情報を読み取り、ストレージ媒体に情報を書き込むことができるように、プロセッサに結合される。代替として、ストレージ媒体はプロセッサと一体であってよい。プロセッサおよびストレージ媒体は、ＡＳＩＣに存在してよい。ＡＳＩＣは、ユーザ端末に存在してよい。代替として、プロセッサおよびストレージ媒体は、個別のコンポーネントとしてユーザ端末に存在してよい。

【０２５８】

本明細書における説明によれば、適切なコンピューティング・デバイスは、非限定的な例として、サーバ・コンピュータ、デスクトップ・コンピュータ、ラップトップ・コンピュータ、ノートブック・コンピュータ、サブノートブック・コンピュータ、ネットブック・コンピュータ、ネットパッド・コンピュータ、セットトップ・コンピュータ、メディア・ストリーミング・デバイス、ハンドヘルド・コンピュータ、インターネット・アプライアンス、モバイル・スマートフォン、タブレット・コンピュータ、パーソナル・デジタル・アシスタント、ビデオ・ゲーム・コンソール、および車両を含む。当業者は、任意選択的なコンピュータ・ネットワーク接続を有する選ばれたテレビ、ビデオ・プレーヤー、およびデジタル音楽プレーヤーが、本明細書に記載されたシステムにおける使用に適しているということも認識するであろう。適切なタブレット・コンピュータは、さまざまな実施形態では、当業者に知られている、ブックレット、スレート、および変換可能な構成を有するタブレット・コンピュータを含む。

【０２５９】

一部の実施形態では、コンピューティング・デバイスは、実行可能な命令を実行するように構成されたオペレーティング・システムを含む。オペレーティング・システムは、例えば、デバイスのハードウェアを管理し、アプリケーションの実行のためのサービスを提供する、プログラムおよびデータを含むソフトウェアである。当業者は、適切なサーバ・オペレーティング・システムが、非限定的な例として、ＦｒｅｅＢＳＤ、ＯｐｅｎＢＳＤ、ＮｅｔＢＳＤ（登録商標）、Ｌｉｎｕｘ（登録商標）、Ａｐｐｌｅ（登録商標）ＭａｃＯＳ　Ｘ（登録商標）サーバ、Ｏｒａｃｌｅ（登録商標）Ｓｏｌａｒｉｓ（登録商標）、Ｗｉｎｄｏｗｓ　Ｓｅｒｖｅｒ（登録商標）、およびＮｏｖｅｌｌ（登録商標）ＮｅｔＷａｒｅ（登録商標）を含むということを認識するであろう。当業者は、適切なパーソナル・コンピュータ・オペレーティング・システムが、非限定的な例として、Ｍｉｃｒｏｓｏｆｔ（登録商標）Ｗｉｎｄｏｗｓ（登録商標）、Ａｐｐｌｅ（登録商標）Ｍａｃ　ＯＳ

Ｘ（登録商標）、ＵＮＩＸ（登録商標）、およびＧＮＵ／Ｌｉｎｕｘ（登録商標）などのＵＮＩＸ（登録商標）のようなオペレーティング・システムを含むということを認識するであろう。一部の実施形態では、オペレーティング・システムは、クラウド・コンピューティングによって提供される。当業者は、適切なモバイル・スマートフォン・オペレーティング・システムが、非限定的な例として、Ｎｏｋｉａ（登録商標）Ｓｙｍｂｉａｎ（登録商標）ＯＳ、Ａｐｐｌｅ（登録商標）ｉＯＳ（登録商標）、Ｒｅｓｅａｒｃｈ Ｉｎ Ｍｏｔｉｏｎ（登録商標）ＢｌａｃｋＢｅｒｒｙ ＯＳ（登録商標）、Ｇｏｏｇｌｅ（登録商標）Ａｎｄｒｏｉｄ（登録商標）、Ｍｉｃｒｏｓｏｆｔ（登録商標）Ｗｉｎｄｏｗｓ Ｐｈｏｎｅ（登録商標）ＯＳ、Ｍｉｃｒｏｓｏｆｔ（登録商標）Ｗｉｎｄｏｗｓ Ｍｏｂｉｌｅ（登録商標）ＯＳ、Ｌｉｎｕｘ（登録商標）、およびＰａｌｍ（登録商標）ＷｅｂＯＳ（登録商標）を含むということも認識するであろう。当業者は、適切なメディア・ストリーミング・デバイス・オペレーティング・システムが、非限定的な例として、Ａｐｐｌｅ ＴＶ（登録商標）、Ｒｏｋｕ（登録商標）、Ｂｏｘｅｅ（登録商標）、Ｇｏｏｇｌｅ ＴＶ（登録商標）、Ｇｏｏｇｌｅ Ｃｈｒｏｍｅｃａｓｔ（登録商標）、Ａｍａｚｏｎ Ｆｉｒｅ（登録商標）、およびＳａｍｓｕｎｇ（登録商標）ＨｏｍｅＳｙｎｃ（登録商標）を含むということも認識するであろう。当業者は、適切なビデオ・ゲーム・コンソール・オペレーティング・システムが、非限定的な例として、Ｓｏｎｙ（登録商標）ＰＳ３（登録商標）、Ｓｏｎｙ（登録商標）ＰＳ４（登録商標）、Ｍｉｃｒｏｓｏｆｔ（登録商標）Ｘｂｏｘ ３６０（登録商標）、Ｍｉｃｒｏｓｏｆｔ Ｘｂｏｘ Ｏｎｅ、Ｎｉｎｔｅｎｄｏ（登録商標）Ｗｉｉ（登録商標）、Ｎｉｎｔｅｎｄｏ（登録商標）Ｗｉｉ Ｕ（登録商標）、およびＯｕｙａ（登録商標）を含むということも認識するであろう。

【０２６０】

非一時的コンピュータ可読ストレージ媒体

一部の実施形態では、本明細書において開示されるプラットフォーム、システム、媒体、および方法は、任意選択的にネットワーク化されたコンピューティング・デバイスのオペレーティング・システムによって実行可能な命令を含むプログラムを使用してエンコードされた、１つまたは複数の非一時的コンピュータ可読ストレージ媒体を含む。さらなる実施形態では、コンピュータ可読ストレージ媒体は、コンピューティング・デバイスの有形のコンポーネントである。さらなる実施形態では、コンピュータ可読ストレージ媒体は、任意選択的に、コンピューティング・デバイスから取り外し可能である。一部の実施形態では、コンピュータ可読ストレージ媒体は、非限定的な例として、ＣＤ－ＲＯＭ、ＤＶＤ、フラッシュ・メモリ・デバイス、ソリッド・ステート・メモリ、磁気ディスク・ドライブ、磁気テープ・ドライブ、光ディスク・ドライブ、クラウド・コンピューティング・システムおよびサービスを含む分散コンピューティング・システムなどを含む。場合によっては、プログラムおよび命令は、永続的、実質的に永続的、半永続的、または非一時的に、媒体にエンコードされる。

【０２６１】

コンピュータ・プログラム

一部の実施形態では、本明細書において開示されるプラットフォーム、システム、媒体、および方法は、少なくとも１つのコンピュータ・プログラムまたはコンピュータ・プログラムの使用を含む。コンピュータ・プログラムは、指定されたタスクを実行するために書き込まれた、コンピューティング・デバイスのＣＰＵの１つまたは複数のプロセッサによって実行可能な一連の命令を含む。場合によっては、コンピュータ・プログラムは、垂直方向または水平方向にスケーリングされる。例えば、コンピュータ・プログラムは、１つまたは複数のハードウェア・パラメータ（例えば、クロック速度、コア、キャッシュ・サイズ、ＲＡＭサイズ、ＣＰＵ、図１０のコンポーネントなど）、１つまたは複数のソフトウェア手法（例えば、同時実行、並列性、分散手法など）、またはその両方を使用して、スケールアップまたはスケールダウンされてよい。

【０２６２】

一部の実施形態では、１つまたは複数のプロセッサによって実行可能な命令は、本明細

書に記載されたエンコーディング方法またはデコーディング方法を含む。場合によっては、エンコーディング方法またはデコーディング方法は、図１～９または図１２～１５において提供される１つまたは複数の動作または一般的な手法を含む。例えば、１つまたは複数のプロセッサによって実行可能な命令は、コードブックを含んでいる内部コーデックを生成することを含んでよい。場合によっては、コードブックは、塩基順序を使用して生成される。場合によっては、塩基順序は、ユーザ、コンピュータ・プログラム、またはその両方によって選択される。場合によっては、１つまたは複数のプロセッサによって実行可能な命令は、内部コーデックを適用して、シンボルのセットとして表されたデータをシンボルの別のセットにエンコードすることを含む。例として、データは、「０」および「１」の２進値などの数値シンボルとして表されてよく、コンピュータ・プログラムは、内部コーデックを適用して、データを複数のポリヌクレオチド配列に変換してよい。

【０２６３】

コンピュータ・プログラムを含むコンピュータ・システムは、プラットフォームの１つまたは複数のコンポーネント、例えば、シンセサイザ、シーケンサ、またはストレージ・ユニットなどと、電子通信してよい。そのような場合、コンピュータ・プログラムは、システムに１つまたは複数の動作を実行させる命令をさらに実行してよい。この動作は、シンセサイザに複数のポリヌクレオチドを生成させること、シーケンサに複数のポリヌクレオチドをシーケンシングさせること、あるいはシンセサイザ、シーケンサ、ストレージ・ユニット、またはこれらの任意の組み合わせの間で複数のポリヌクレオチドを転送することを含むことができる。場合によっては、コンピュータ・システムは、複数の出力配列を受信する。そのような場合、１つまたは複数のプロセッサによって実行可能な命令は、複数の出力配列をデコードすることを含む。

【０２６４】

コンピュータ可読命令は、関数、オブジェクト、アプリケーション・プログラミング・インターフェイス（ＡＰＩ）、計算データ構造などの、特定のタスクを実行するか、または特定の抽象データ型を実装する、プログラム・モジュールとして実装されてよい。本明細書において提供される開示を踏まえて、当業者は、コンピュータ・プログラムがさまざまな言語のさまざまなバージョンで書かれてよいということを認識するであろう。

【０２６５】

コンピュータ可読命令の機能は、さまざまな環境内で、必要に応じて組み合わせられるか、または配布されてよい。一部の実施形態では、コンピュータ・プログラムは、命令の１つのシーケンスを含む。一部の実施形態では、コンピュータ・プログラムは、命令の複数のシーケンスを含む。一部の実施形態では、コンピュータ・プログラムは、１つの場所から提供される。他の実施形態では、コンピュータ・プログラムは、複数の場所から提供される。さまざまな実施形態では、コンピュータ・プログラムは、１つまたは複数のソフトウェア・モジュールを含む。さまざまな実施形態では、コンピュータ・プログラムは、１つまたは複数のウェブ・アプリケーション、１つまたは複数のモバイル・アプリケーション、１つまたは複数のスタンドアロン・アプリケーション、１つまたは複数のウェブ・ブラウザ・プラグイン、拡張機能、アドイン、またはアドオン、あるいはこれらの組み合わせを、部分的または全体的に含む。

【０２６６】

ウェブ・アプリケーション

一部の実施形態では、コンピュータ・プログラムは、ウェブ・アプリケーションを含む。本明細書において提供される開示を踏まえて、当業者は、ウェブ・アプリケーションが、さまざまな実施形態において、１つまたは複数のソフトウェア・フレームワークおよび１つまたは複数のデータベース・システムを利用するということを認識するであろう。一部の実施形態では、ウェブ・アプリケーションは、Ｍｉｃｒｏｓｏｆｔ（登録商標）．ＮＥＴまたはＲｕｂｙ　ｏｎ　Ｒａｉｌｓ（ＲｏＲ）などのソフトウェア・フレームワーク上に作成される。一部の実施形態では、ウェブ・アプリケーションは、非限定的な例として、リレーショナル、非リレーショナル、オブジェクト指向、連想、ＸＭＬ、およびドキ

ュメント指向のデータベース・システムを含む、１つまたは複数のデータベース・システムを利用する。さらなる実施形態では、適切なリレーショナル・データベース・システムは、非限定的な例として、Ｍｉｃｒｏｓｏｆｔ（登録商標）ＳＱＬ Ｓｅｒｖｅｒ、ｍｙＳＱＬ（登録商標）、およびＯｒａｃｌｅ（登録商標）を含む。当業者は、ウェブ・アプリケーションが、さまざまな実施形態において、１つまたは複数の言語の１つまたは複数のバージョンで書かれるということも認識するであろう。ウェブ・アプリケーションは、１つまたは複数のマークアップ言語、プレゼンテーション定義言語、クライアント側スクリプト言語、サーバ側コーディング言語、データベース・クエリ言語、またはこれらの組み合わせで書かれてよい。一部の実施形態では、ウェブ・アプリケーションは、ある程度、ハイパーテキスト・マークアップ言語（ＨＴＭＬ：Ｈｙｐｅｒｔｅｘｔ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ）、拡張可能ハイパーテキスト・マークアップ言語（ＸＨＴＭＬ：Ｅｘｔｅｎｓｉｂｌｅ Ｈｙｐｅｒｔｅｘｔ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ）、または、拡張可能マークアップ言語（ＸＭＬ：ｅＸｔｅｎｓｉｂｌｅ Ｍａｒｋｕｐ Ｌａｎｇｕａｇｅ）などのマークアップ言語で書かれる。一部の実施形態では、ウェブ・アプリケーションは、ある程度、カスケード・スタイル・シート（ＣＳＳ：Ｃａｓｃａｄｉｎｇ Ｓｔｙｌｅ Ｓｈｅｅｔｓ）などのプレゼンテーション定義言語で書かれる。一部の実施形態では、ウェブ・アプリケーションは、ある程度、Ａｓｙｎｃｈｒｏｎｏｕｓ ＪａｖａＳｃｒｉｐｔ ａｎｄ ＸＭＬ（ＡＪＡＸ）、Ｆｌａｓｈ（登録商標）ＡｃｔｉｏｎＳｃｒｉｐｔ、ＪａｖａＳｃｒｉｐｔ、またはＳｉｌｖｅｒｌｉｇｈｔ（登録商標）などの、クライアント側スクリプト言語で書かれる。一部の実施形態では、ウェブ・アプリケーションは、ある程度、Ａｃｔｉｖｅ Ｓｅｒｖｅｒ Ｐａｇｅｓ（ＡＳＰ）、ＣｏｌｄＦｕｓｉｏｎ（登録商標）、Ｐｅｒｌ、Ｊａｖａ（商標）、ＪａｖａＳｅｒｖｅｒ Ｐａｇｅｓ（ＪＳＰ）、ハイパーテキスト・プリプロセッサ（ＰＨＰ）、Ｐｙｔｈｏｎ（商標）、Ｒｕｂｙ、Ｔｃｌ、Ｓｍａｌｌｔａｌｋ、ＷｅｂＤＮＡ（登録商標）、またはＧｒｏｏｖｙなどの、サーバ側コーディング言語で書かれる。一部の実施形態では、ウェブ・アプリケーションは、ある程度、Ｓｔｒｕｃｔｕｒｅｄ Ｑｕｅｒｙ Ｌａｎｇｕａｇｅ（ＳＱＬ）などの、データベース・クエリ言語で書かれる。一部の実施形態では、ウェブ・アプリケーションは、ＩＢＭ（登録商標）Ｌｏｔｕｓ Ｄｏｍｉｎｏ（登録商標）などのエンタープライズ・サーバ製品を統合する。一部の実施形態では、ウェブ・アプリケーションは、メディア・プレーヤー要素を含む。さまざまなさらなる実施形態では、メディア・プレーヤー要素は、非限定的な例として、Ａｄｏｂｅ（登録商標）Ｆｌａｓｈ（登録商標）、ＨＴＭＬ ５、Ａｐｐｌｅ（登録商標）ＱｕｉｃｋＴｉｍｅ（登録商標）、Ｍｉｃｒｏｓｏｆｔ（登録商標）Ｓｉｌｖｅｒｌｉｇｈｔ（登録商標）、Ｊａｖａ（商標）、およびＵｎｉｔｙ（登録商標）を含む、多くの適切なマルチメディア技術のうちの１つまたは複数を利用する。

【０２６７】

モバイル・アプリケーション

一部の実施形態では、コンピュータ・プログラムは、モバイル・コンピューティング・デバイスに提供されるモバイル・アプリケーションを含む。一部の実施形態では、モバイル・アプリケーションは、モバイル・コンピューティング・デバイスが製造される時点で、モバイル・コンピューティング・デバイスに提供される。他の実施形態では、モバイル・アプリケーションは、本明細書に記載されたコンピュータ・ネットワークを介してモバイル・コンピューティング・デバイスに提供される。

【０２６８】

本明細書において提供される開示を考慮して、モバイル・アプリケーションは、当技術分野で知られているハードウェア、言語、および開発環境を使用して、当業者に知られている技術によって作成される。当業者は、モバイル・アプリケーションが複数の言語で書かれるということを認識するであろう。適切なプログラミング言語は、非限定的な例として、Ｃ、Ｃ＋＋、Ｃ＃、Ｏｂｊｅｃｔｉｖｅ－Ｃ、Ｊａｖａ（商標）、ＪａｖａＳｃｒｉｐｔ、Ｐａｓｃａｌ、Ｏｂｊｅｃｔ Ｐａｓｃａｌ、Ｐｙｔｈｏｎ（商標）、Ｒｕｂｙ、

ＶＢ．ＮＥＴ、ＷＭＬ、およびＣＳＳを含むか、または含まないＸＨＴＭＬ／ＨＴＭＬ、あるいはこれらの組み合わせを含む。

【０２６９】

適切なモバイル・アプリケーション開発環境は、複数のソースから入手可能である。市販されている開発環境は、非限定的な例として、ＡｉｒｐｌａｙＳＤＫ、ａｌｃｈｅＭｏ、Ａｐｐｃｅｌｅｒａｔｏｒ（登録商標）、Ｃｅｌｓｉｕｓ、Ｂｅｄｒｏｃｋ、Ｆｌａｓｈ Ｌｉｔｅ、．ＮＥＴ Ｃｏｍｐａｃｔ Ｆｒａｍｅｗｏｒｋ、Ｒｈｏｍｏｂｉｌｅ、ＷｏｒｋＬｉｇｈｔ Ｍｏｂｉｌｅ Ｐｌａｔｆｏｒｍを含む。非限定的な例として、Ｌａｚａｒｕｓ、ＭｏｂｉＦｌｅｘ、ＭｏＳｙｎｃ、およびＰｈｏｎｅｇａｐを含む他の開発環境が、無料で利用可能である。また、モバイル・デバイス製造業者は、非限定的な例として、ｉＰｈｏｎｅ（登録商標）およびｉＰａｄ（登録商標）（ｉＯＳ）ＳＤＫ、Ａｎｄｒｏｉｄ（商標）ＳＤＫ、ＢｌａｃｋＢｅｒｒｙ（登録商標）ＳＤＫ、ＢＲＥＷ ＳＤＫ、Ｐａｌｍ（登録商標）ＯＳ ＳＤＫ、Ｓｙｍｂｉａｎ ＳＤＫ、ｗｅｂＯＳ ＳＤＫ、ならびにＷｉｎｄｏｗｓ（登録商標）Ｍｏｂｉｌｅ ＳＤＫを含む、ソフトウェア開発キットを配布している。

【０２７０】

当業者は、非限定的な例として、Ａｐｐｌｅ（登録商標）Ａｐｐ Ｓｔｏｒｅ、Ｇｏｏｇｌｅ（登録商標）Ｐｌａｙ、Ｃｈｒｏｍｅ ＷｅｂＳｔｏｒｅ、ＢｌａｃｋＢｅｒｒｙ（登録商標）Ａｐｐ Ｗｏｒｌｄ、Ｐａｌｍデバイス用Ａｐｐ Ｓｔｏｒｅ、ｗｅｂＯＳ用Ａｐｐ Ｃａｔａｌｏｇ、モバイル用Ｗｉｎｄｏｗｓ（登録商標）Ｍａｒｋｅｔｐｌａｃｅ、Ｎｏｋｉａ（登録商標）デバイス用Ｏｖｉ Ｓｔｏｒｅ、Ｓａｍｓｕｎｇ（登録商標）Ａｐｐｓ、およびニンテンドー（登録商標）ＤＳｉショップを含む複数の商業フォーラムが、モバイル・アプリケーションの配布に利用可能であるということを認識するであろう。

【０２７１】

スタンドアロン・アプリケーション

一部の実施形態では、コンピュータ・プログラムは、既存のプロセスへのアドオンではなく、例えばプラグインではなく、独立したコンピュータ・プロセスとして実行されるプログラムである、スタンドアロン・アプリケーションを含む。当業者は、スタンドアロン・アプリケーションが、多くの場合、コンパイルされるということを認識するであろう。コンパイラは、プログラミング言語で記述されたソース・コードをアセンブリ言語または機械語などのバイナリ・オブジェクト・コードに変換するコンピュータ・プログラムである。適切なコンパイルされるプログラミング言語は、非限定的な例として、Ｃ、Ｃ＋＋、Ｏｂｊｅｃｔｉｖｅ－Ｃ、ＣＯＢＯＬ、Ｄｅｌｐｈｉ、Ｅｉｆｆｅｌ、Ｊａｖａ（商標）、Ｌｉｓｐ、Ｐｙｔｈｏｎ（商標）、Ｖｉｓｕａｌ Ｂａｓｉｃ、およびＶＢ．ＮＥＴ、またはこれらの組み合わせを含む。コンパイルは、多くの場合、実行可能プログラムを作成するために、少なくとも部分的に実行される。一部の実施形態では、コンピュータ・プログラムは、１つまたは複数の実行可能なコンパイルされたアプリケーションを含む。

【０２７２】

ウェブ・ブラウザ・プラグイン

一部の実施形態では、コンピュータ・プログラムは、ウェブ・ブラウザ・プラグイン（例えば、拡張機能など）を含む。コンピューティングでは、プラグインは、特定の機能をより大きいソフトウェア・アプリケーションに追加する１つまたは複数のソフトウェア・コンポーネントである。ソフトウェア・アプリケーションのメーカーは、サードパーティの開発者がアプリケーションを拡張する機能を作成できるようするため、新しい機能を簡単に追加することをサポートするため、およびアプリケーションのサイズを縮小するために、プラグインをサポートしている。プラグインは、サポートされている場合、ソフトウェア・アプリケーションの機能のカスタマイズを可能にする。例えば、プラグインは、ビデオの再生、対話性の生成、ウイルスのスキャン、および特定のファイル・タイプの表示のために、ウェブ・ブラウザで一般的に使用される。当業者は、Ａｄｏｂｅ（登録商標）

Ｆｌａｓｈ（登録商標）Ｐｌａｙｅｒ、Ｍｉｃｒｏｓｏｆｔ（登録商標）Ｓｉｌｖｅｒｌｉｇｈｔ（登録商標）、およびＡｐｐｌｅ（登録商標）ＱｕｉｃｋＴｉｍｅ（登録商標）を含む、複数のウェブ・ブラウザ・プラグインを熟知しているであろう。一部の実施形態では、ツールバーは、１つまたは複数のウェブ・ブラウザの拡張機能、アドイン、またはアドオンを含む。一部の実施形態では、ツールバーは、１つまたは複数のエクスプローラ・バー、ツール・バンド、またはデスク・バンドを含む。

【０２７３】

本明細書において提供される開示を考慮して、当業者は、非限定的な例として、Ｃ＋＋、Ｄｅｌｐｈｉ、Ｊａｖａ（商標）、ＰＨＰ、Ｐｙｔｈｏｎ（商標）、およびＶＢ．ＮＥＴ、またはこれらの組み合わせを含む、さまざまなプログラミング言語でのプラグインの開発を可能にする複数のプラグイン・フレームワークが利用可能であるということを、認識するであろう。

【０２７４】

ウェブ・ブラウザ（インターネット・ブラウザとも呼ばれる）は、ワールド・ワイド・ウェブ上の情報リソースを取り出し、提示し、トラバースするために、ネットワーク接続されたコンピューティング・デバイスで使用するために設計されたソフトウェア・アプリケーションである。適切なウェブ・ブラウザは、非限定的な例として、Ｍｉｃｒｏｓｏｆｔ（登録商標）Ｉｎｔｅｒｎｅｔ Ｅｘｐｌｏｒｅｒ（登録商標）、Ｍｏｚｉｌｌａ（登録商標）Ｆｉｒｅｆｏｘ（登録商標）、Ｇｏｏｇｌｅ（登録商標）Ｃｈｒｏｍｅ、Ａｐｐｌｅ（登録商標）Ｓａｆａｒｉ（登録商標）、Ｏｐｅｒａ Ｓｏｆｔｗａｒｅ（登録商標）Ｏｐｅｒａ（登録商標）、およびＫＤＥ Ｋｏｎｑｕｅｒｏｒを含む。一部の実施形態では、ウェブ・ブラウザは、モバイル・ウェブ・ブラウザである。モバイル・ウェブ・ブラウザ（マイクロブラウザ、ミニブラウザ、およびワイヤレス・ブラウザとも呼ばれる）は、非限定的な例として、ハンドヘルド・コンピュータ、タブレット・コンピュータ、ネットブック・コンピュータ、サブノートブック・コンピュータ、スマートフォン、音楽プレーヤー、パーソナル・デジタル・アシスタント（ＰＤＡ：ｐｅｒｓｏｎａｌ ｄｉｇｉｔａｌ ａｓｓｉｓｔａｎｔｓ）、ハンドヘルド・ビデオ・ゲーム・システムを含む、モバイル・コンピューティング・デバイスで使用するために設計される。適切なモバイル・ウェブ・ブラウザは、非限定的な例として、Ｇｏｏｇｌｅ（登録商標）Ａｎｄｒｏｉｄ（登録商標）ブラウザ、ＲＩＭ ＢｌａｃｋＢｅｒｒｙ（登録商標）ブラウザ、Ａｐｐｌｅ（登録商標）Ｓａｆａｒｉ（登録商標）、Ｐａｌｍ（登録商標）Ｂｌａｚｅｒ、Ｐａｌｍ（登録商標）ＷｅｂＯＳ（登録商標）ブラウザ、モバイル用Ｍｏｚｉｌｌａ（登録商標）Ｆｉｒｅｆｏｘ（登録商標）、Ｍｉｃｒｏｓｏｆｔ（登録商標）Ｉｎｔｅｒｎｅｔ Ｅｘｐｌｏｒｅｒ（登録商標）Ｍｏｂｉｌｅ、Ａｍａｚｏｎ（登録商標）Ｋｉｎｄｌｅ（登録商標）Ｂａｓｉｃ Ｗｅｂ、Ｎｏｋｉａ（登録商標）ブラウザ、Ｏｐｅｒａ Ｓｏｆｔｗａｒｅ（登録商標）Ｏｐｅｒａ（登録商標）Ｍｏｂｉｌｅ、およびＳｏｎｙ（登録商標）ＰＳＰ（商標）ブラウザを含む。

【０２７５】

ソフトウェア・モジュール

一部の実施形態では、本明細書において開示されるプラットフォーム、システム、媒体、および方法は、ソフトウェア、サーバ、および／またはデータベース・モジュール、またはそれらの使用を含む。本明細書において提供される開示を考慮して、ソフトウェア・モジュールは、当技術分野で知られているマシン、ソフトウェア、および言語を使用して、当業者に知られている技術によって作成される。本明細書において開示されるソフトウェア・モジュールは、多数の方法で実装される。さまざまな実施形態では、ソフトウェア・モジュールは、ファイル、コードのセクション、プログラミング・オブジェクト、プログラミング構造、分散コンピューティング・リソース、クラウド・コンピューティング・リソース、またはこれらの組み合わせを含む。さらにさまざまな実施形態では、ソフトウェア・モジュールは、複数のファイル、コードの複数のセクション、複数のプログラミング・オブジェクト、複数のプログラミング構造、複数の分散コンピューティング・リソー

ス、複数のクラウド・コンピューティング・リソース、またはこれらの組み合わせを含む。さまざまな実施形態では、１つまたは複数のソフトウェア・モジュールは、非限定的な例として、ウェブ・アプリケーション、モバイル・アプリケーション、スタンドアロン・アプリケーション、および分散またはクラウド・コンピューティング・アプリケーションを含む。一部の実施形態では、ソフトウェア・モジュールは、１つのコンピュータ・プログラムまたはアプリケーション内にある。他の実施形態では、ソフトウェア・モジュールは、２つ以上のコンピュータ・プログラムまたはアプリケーション内にある。一部の実施形態では、ソフトウェア・モジュールは、１つのマシン上でホストされる。他の実施形態では、ソフトウェア・モジュールは、２つ以上のマシン上でホストされる。さらなる実施形態では、ソフトウェア・モジュールは、クラウド・コンピューティング・システムなどの分散コンピューティング・プラットフォーム上でホストされる。一部の実施形態では、ソフトウェア・モジュールは、１つの場所にある１つまたは複数のマシン上でホストされる。他の実施形態では、ソフトウェア・モジュールは、２つ以上の場所にある１つまたは複数のマシン上でホストされる。

【０２７６】

データベース

一部の実施形態では、本明細書において開示されるプラットフォーム、システム、媒体、および方法は、１つまたは複数のデータベースまたはデータベースの使用を含む。本明細書において提供される開示を考慮して、当業者は、多くのデータベースが情報の格納および取り出しに適しているということを認識するであろう。さまざまな実施形態では、適切なデータベースは、非限定的な例として、リレーショナル・データベース、非リレーショナル・データベース、オブジェクト指向データベース、オブジェクト・データベース、エンティティリレーションシップ・モデル・データベース、連想データベース、ＸＭＬデータベース、ドキュメント指向データベース、およびグラフ・データベースを含む。さらなる非限定的な例としては、ＳＱＬ、ＰｏｓｔｇｒｅＳＱＬ、ＭｙＳＱＬ（登録商標）、Ｏｒａｃｌｅ、ＤＢ２、Ｓｙｂａｓｅ、およびＭｏｎｇｏＤＢが挙げられる。一部の実施形態では、データベースは、インターネットベースである。さらなる実施形態では、データベースはウェブベースである。さらなる実施形態では、データベースは、クラウド・コンピューティングベースである。特定の実施形態では、データベースは分散データベースである。他の実施形態では、データベースは、１つまたは複数のローカル・コンピュータ・ストレージ・デバイスに基づく。

【実施例】

【０２７７】

以下の実施例は、本明細書に記載されたソフトウェア・アプリケーション、システム、および方法の実施形態を代表しており、いかなる方法でも限定するよう意図されていない。

【実施例１】

【０２７８】

レーン・インデックス・シャッフルを使用するオリゴヌクレオチドにおけるバイナリ・データのエンコーディング

４．５ＧＢのデータ・ストリームが１００，０００フレームに分割され、各フレームは、１フレーム当たり４５ＫＢを有する。各フレームは２５００レーンに分割され、各レーンは１４４ビットを有する。外部リード・ソロモンＧＦ（2^{12}）コードが各フレームに適用されて、フレームのサイズを増やし、１フレーム当たり４０９５レーンを生成する。

【０２７９】

その後、各レーンは、回転方式を使用してシャッフルされる。回転方式は、レーン・インデックスに基づく。例えば、レーン・インデックスが０である場合、そのレーンの回転はない。レーン・インデックスが１である場合、レーン内のビットがそれぞれ１ずつシフトされる。レーン・インデックスが２である場合、レーン内のビットがそれぞれ２ずつシフトされる、などとなる。さらに、レーン・インデックスおよびフレーム・インデックス

が各レーンの先頭に追加される。レーン・インデックスは１２ビットであり、フレーム・インデックスは２０ビットである。

【０２８０】

レーンがシフトされると、内部エンコーディング方式が各レーンに適用され、バイナリ・データをポリヌクレオチド配列としてエンコードする。データからエンコードするビット（例えば、１４４ビットに加えて、レーン・インデックスおよびフレーム・インデックス）が、８ビットの履歴および４最下位ビット（ＬＳＢ：ｌｅａｓｔ　ｓｉｇｎｉｆｉｃａｎｔ　ｂｉｔ）のインデックスと結合される。ビットは、ルックアップ・テーブルを使用してエンコードされる。ビットは、１塩基当たり１ビットのレートでエンコードされる。したがって、生成される配列の各々の長さは、１８８塩基である。ルックアップ・テーブルを使用して塩基候補が生成される。同じ塩基が隣り合ってエンコードされるのを防ぐために、塩基反復チェックがさらに実行される。例えば、２つの塩基（例えば、「ＡＡ」）が隣り合っている場合、塩基のうちの１つが更新される（例えば、「ＡＴ」）。その後、ビット履歴が更新され、レーン・インデックスおよび／またはフレーム・インデックスがインクリメントされる。新しいビット履歴が、エンコードする後続のビットに追加される。

【０２８１】

その後、オリゴヌクレオチド配列の塩基に対してＧＣフィルタリングが実行される。オリゴヌクレオチドの約５％～約１０％が、ＧＣフィルタリング中に除去される。最終的なプール内の塩基含有は、オリゴヌクレオチドであり、約４５％～約５５％のＧＣ含量である。

【０２８２】

次に、最終的なオリゴヌクレオチド・プールが合成され、格納される。

【実施例２】

【０２８３】

混合貪欲－最尤（ＭＬ）アルゴリズムを使用するオリゴヌクレオチドからのバイナリ・データのデコーディング

実施例１の一般的な方法に従ってアレイに格納されているオリゴヌクレオチドが、ペアエンド・シーケンシングによってシーケンシングされる。シーケンシングされたオリゴヌクレオチドは、レーン・インデックスおよび／またはフレーム・インデックスを回復するために部分的にデコードされる。オリゴヌクレオチドは、それらのレーン・インデックスに従ってクラスタ化される。

【０２８４】

次に、クラスタ化されたオリゴヌクレオチドがアライメントされる。アライメント中に、アライメント・アルゴリズムを使用してオリゴヌクレオチドのコンセンサスが分析される。オリゴヌクレオチド配列の各々の最初の位置が０に初期化され、約５つのオリゴヌクレオチド配列の読み取り間で、次の２塩基または３塩基のコンセンサスが分析される。オリゴヌクレオチド配列ごとに、次の２塩基または３塩基が正しいかどうか、あるいは欠失、挿入、または置換が存在するかどうかの判定が行われる。この判定を前提として、読み取りごとに位置がインクリメントされる。これらのステップは、読み取りの終了に達するまで繰り返される。

【０２８５】

デコーディング方式を含んでいる内部コーデックが、クラスタ化されてアライメントされたオリゴヌクレオチド配列に適用される。デコーディング方式は、混合貪欲最尤（ＭＬ）アルゴリズムを含む。この混合貪欲ＭＬアルゴリズムでは、約１００個の最も可能性の高い状態からの遷移確率に基づいてデコーディングが実行される。デコーディングは、オリゴヌクレオチド配列内の塩基をバイナリ・データのレーンに変換する。実施例１では、１塩基当たり１ビットのレートでビットがエンコードされるため、各塩基は、貪欲ＭＬアルゴリズムにおいて１ビットとしてデコードされる。

【０２８６】

オリゴヌクレオチド配列が終了する位置を知ることによって、デコーディング方式の性能が改善される。オリゴヌクレオチドの長さは、ペアエンド・シーケンシングによって決定される。各配列は、１８８－ｍｅｒのオリゴヌクレオチドである。ドリフト項は、貪欲ＭＬアルゴリズムに導入され、挿入および欠失の総数に関連付けられた整数である。各挿入は＋１の値を表し、各欠失は－１の値を表す。例えば、挿入がなく、２つの欠失がある場合、合計ドリフトは－２になる。そのような例では、貪欲ＭＬアルゴリズムは、１８６－ｍｅｒの配列以外のすべての最終デコーディング状態を無効として破棄する。したがって、ドリフト項は、混合貪欲ＭＬアルゴリズムが、どの最終デコーディング状態が有効であるかを知ることを可能にし、性能をさらに改善する。

【０２８７】

混合貪欲ＭＬアルゴリズムからの出力は、バイナリ・データのレーンである。レーンは、データの１４４ビット、レーン・インデックスをエンコードする１２ビット、およびフレーム・インデックスをエンコードする２０ビットを含む。

【０２８８】

バイナリ・データのレーンが取得されると、レーンは、レーン・インデックスおよびフレーム・インデックスに基づいてフレームに配置される。レーンはフレーム・インデックスに従ってデシャッフルされ、グループ化される。合計１００，０００フレームが取得され、各フレームは４０９５レーンを含む。

【０２８９】

エラー訂正方式を含む外部コーデックが、各フレームに適用される。外部リード・ソロモンＧＦ（2^{12}）コードが各フレームに適用されて、フレームのサイズを減らし、１フレーム当たり２５００レーンを生成する。

【０２９０】

次に、フレームがマージされて、バイナリ・データを含む出力を生成する。各フレームは４５ＫＢのデータを含み、回復されるバイナリ・データの合計は、４．５ＧＢのデータ・ストリームである。

【実施例３】

【０２９１】

コードブックを使用するオリゴヌクレオチドとの間のバイナリ・データのエンコーディングおよびデコーディング

ＬＤＰＣ外部コーデックは、本明細書において概して説明された手順を使用して、２０，０００ビットを含むデータに適用される。ＬＤＰＣ外部コーデックからの出力は、３５，０００ビットを含む。

【０２９２】

（８ビットごとの）８つの塩基シンボルの４つの異なるコードブックを使用する内部コーデックが、外部コーデックからの出力に適用される。コードブックは、シンボル間の編集距離を最適に最大化するように作成される。最初に、すべての可能性のあるシンボルが、塩基のすべての組み合わせから生成され、次に、最大反復回数およびＧＣ含量によってフィルタリングされる。初期シンボルが選択され、その後、初期シンボルからの最大編集距離を有する次のシンボルが選択される。後続の各シンボルが、前のシンボルに基づいて選択される。後続の各シンボルは、以前に選択されたすべてのシンボルへの最大平均編集距離に関して最小化される。このプロセスが、コードブックがいっぱいになるまで繰り返される。

【０２９３】

次のコードブックは、異なるシンボルから開始し、前のコードブックからの共通のシンボルを最小化することによって生成される。内部エンコーディング中に、ｉ番目の８ビット・ワードをエンコードするときに、（ｉモジュロ４の）コードブックが使用されて、割り当てられる８つの塩基シンボルを見つける。

【０２９４】

デコーディング中に、シンボルごとに、（突然変異（または置換）、欠失、および挿入

の設定された制限を有する）可能性のある受信された塩基ごとの尤度確率が、事前に計算される。この事前計算は高価になる可能性があるため、可能性のある塩基ごとの尤度確率は、ファイルにも保存され、そのため、後で読み込まれ得る。次に、最尤（ＭＬ）を含むデコーディング方式が実行されて、ビットを回復する。その後、外部ＬＤＰＣコードが適用されて、データを回復する。

【実施例４】

【０２９５】

バイナリ・データを格納するための合成最適化コーデック

特定の塩基転移を可能にするために、特定の合成順序が選択される。例えば、合成順序はＡ、Ｇ、Ｃ、Ｔである。この合成順序が、コードブックを含む内部コーデックを生成するために使用され得る。この合成順序に従って得られるコードブックは、［Ａ，Ｇ，Ｃ，Ｔ，ＡＧ，ＡＣ，ＡＴ，ＧＣ，ＧＴ，ＡＧＣ，ＡＣＴ，ＡＧＣＴ］というコードワードを含む。これらの１２個のコードワードは、４サイクルで合成され得る。

【０２９６】

バイナリ・データは、実施例１の一般的な方法に従って、データを分割し、外部コーデックを適用し、データをシャッフルすることによって処理される。コードブックを含んでいる内部コーデックは、バイナリ・データを複数のポリヌクレオチド配列としてエンコードするために、バイナリ・データに適用される。詳細には、バイナリ・データが、コードブックに基づいて複数のポリヌクレオチド配列にマッピングされる。マッピングは、編集距離、塩基の繰り返し、またはその両方に基づいてさらに最適化され得る。

【０２９７】

ポリヌクレオチド配列の約５０％が冗長性（例えば、２倍の冗長性）のためにエンコードするように、コーデックが適用される。したがって、バイナリ・データの６つの値が、１２個のコードワードにマッピングされる。これらの６つの値は、$\log_2(12)=3.58$ビットの情報に相当する。したがって、コードワードごとに３．５８／２＝１．７９ビットの情報がエンコードされる。この例では、複数のポリヌクレオチド配列の各々におけるペイロードは、約１００ビットである。このペイロードは、１００ビット／１コードワード当たり１．７９ビット＝５５．８コードワードを必要とする。したがって、最適化されたコーデックを使用して必要とされる合成サイクルの数は、約５５．８コードワード×１コードワード当たり４サイクル＝２２４サイクルの合成である。コーデックを使用しない場合、必要とされる合成サイクル数は、（塩基の追加ごとに４サイクルを仮定して）約４００サイクルである。内部コーデックの実装は、各合成サイクル中に、ポリヌクレオチド合成のための表面上の特徴の少なくとも約半分がデブロッキングされることを可能にする。

【実施例５】

【０２９８】

コードブックを含む合成最適化コーデック

層ごとの特定の塩基転移を可能にするために、特定の合成順序が選択される。層は、本明細書に記載されたデバイスを使用する、少なくとも１つの塩基による各ポリヌクレオチドの伸長を含む。例えば、各特定の連続する層での合成順序は、（１）［Ａ，Ｇ，Ｃ，Ｔ］、（２）［Ｃ，Ａ，Ｔ，Ｇ］、および（３）［Ｔ，Ｇ，Ａ，Ｃ］の繰り返しを含んでよい。この合成順序は、３つのコードブックを含む内部コーデックを生成するために使用され得る。この合成順序に従って最初に得られるコードブックは、（１）［Ａ，Ｇ，Ｃ，Ｔ，ＡＧ，ＡＣ，ＡＴ，ＧＣ，ＧＴ，ＡＧＣ，ＡＣＴ，ＡＧＣＴ］というコードワードを含む。２番目のコードブックは、（２）［Ｃ，Ａ，Ｔ，Ｇ，ＣＡ，ＣＴ，ＣＧ，ＡＴ，ＡＧ，ＣＡＴ，ＣＴＧ，ＣＡＴＧ］というコードワードを含む。３番目のコードブックは、（３）［Ｔ，Ｇ，Ａ，Ｃ，ＴＧ，ＴＡ，ＴＣ，ＴＧＡ，ＴＡＣ，ＴＧＡＣ］というコードワードを含む。これらの１２個のコードワードの各々は、４サイクルで合成され得る。

【０２９９】

バイナリ・データは、実施例１の一般的な方法に従って、データを分割し、外部コーデ

ックを適用し、データをシャッフルすることによって処理される。コードブックを含んでいる内部コーデックは、バイナリ・データを複数のポリヌクレオチド配列としてエンコードするために、バイナリ・データに適用される。詳細には、バイナリ・データが、コードブックに基づいて複数のポリヌクレオチド配列にマッピングされる。マッピングは、編集距離、塩基の繰り返し、またはその両方に基づいてさらに最適化され得る。

【０３００】

ポリヌクレオチド配列の約５０％が冗長性（例えば、２倍の冗長性）のためにエンコードするように、コーデックが適用される。したがって、バイナリ・データの６つの値が、１２個のコードワードにマッピングされる。これらの６つの値は、$\log_2(12)=3.58$ビットの情報に相当する。したがって、コードワードごとに３．５８／２＝１．７９ビットの情報がエンコードされる。この例では、複数のポリヌクレオチド配列の各々におけるペイロードは、約１００ビットである。このペイロードは、１００ビット／１コードワード当たり１．７９ビット＝５５．８コードワードを必要とする。したがって、最適化されたコーデックを使用して必要とされる合成サイクルの数は、約５５．８コードワード×１コードワード当たり４サイクル＝２２４サイクルの合成である。コーデックを使用しない場合、必要とされる合成サイクル数は、（塩基の追加ごとに４サイクルを仮定して）約４００サイクルである。内部コーデックの実装は、各合成サイクル中に、ポリヌクレオチド合成のための表面上の特徴の少なくとも約半分がデブロッキングされることを可能にする。

【実施例６】

【０３０１】

ＤＮＡへのデジタル情報の格納

デジタル情報は、１つまたは複数の処理ユニット、１つまたは複数の処理ユニットと通信するメモリ、およびメモリに格納された命令を備えるシステムを使用して、ＤＮＡに格納される。命令は、デジタル情報をＤＮＡに格納するために、１つまたは複数の処理ユニットで実行される。

【０３０２】

デジタル情報は、それぞれ約１ＭＢである１０００個のオブジェクトを含む。オブジェクトは、テキスト情報、音声情報、および／または視覚的情報を含んでいるファイルを含む。１０００個のオブジェクトが２つのプールに分割され、各プールは、合計で約５００ＭＢである５００個のオブジェクトを含む。

【０３０３】

２つのプールの各々において、プール記述子、１つまたは複数のプール・アイテム、および終了プール記述子が生成される。さらに、プール・アイテムの各々のデータ・ペイロードの第１の１つまたは複数のハッシュおよび１つまたは複数のオブジェクトの各々の第２の１つまたは複数のハッシュが決定される。

【０３０４】

各プール内の５００個のオブジェクトは、１つまたは複数のプール・アイテムにデータ・ペイロードとして格納される。プール・アイテム内の各データ・ペイロードは約２００ビット～約４００ビットであり、オブジェクトごとに約2×10^4～約4×10^4個のプール・アイテムが生成される。データ・ペイロードの各々のハッシュは、ＳＨＡ－２５６を使用するハッシング・モジュールを使用して生成され、２５６ビットである。ハッシュは、各プール・アイテムに追加される。

【０３０５】

プール記述子は、バージョン、プールＩＤ、およびプール・アイテム記述子のリストを含む。バージョンは、プールの最初のバージョンとして保存される（例えば、「００１」）。プールＩＤは、プールに固有の１２８ビットのランダムなラベルである、ＵＵＩＤである。プール・アイテム記述子のリストは、オブジェクトのパス、オブジェクトのサイズ、オブジェクト内のプール・アイテムの範囲、およびプール内のプール・アイテムのオフセットを含むことができる。オブジェクトのパスは、複数のプール内のオブジェクトが位

置する場所を含む（例えば、／ｈｏｍｅ／ｐｏｏｌ１）。オブジェクト内のプール・アイテムの範囲は、各プール・アイテム内のオブジェクトのビットの範囲を含む（例えば、プール・アイテム内１の最初の０～２００ビット）。プール内のプール・アイテムのオフセットは、プールのペイロード内の１つまたは複数のプール・アイテムの各々の最初のバイトのペイロード位置を含む。例えば、プール１内の最初のプール・アイテムのオフセットは０バイトである。最初のプール・アイテムの範囲は最初の５０バイトであり、次のプール・アイテムのオフセットは５０バイトになる。

【０３０６】

終了プール記述子は、オブジェクト記述子のリストを含む。終了プール記述子は、オブジェクトのパスおよびオブジェクトのハッシュを含む。オブジェクトのハッシュはＳＨＡ－２５６を使用して生成され、１ＭＢのオブジェクトのハッシュは２５６ビットである。

【０３０７】

エンコーディング方式が適用されて、２つのプールを複数のポリヌクレオチドを含む２つのライブラリとしてエンコードする。プール内のプール記述子、１つまたは複数のプール・アイテム、および終了プール記述子の各々が、複数のポリヌクレオチドのうちの１つのポリヌクレオチドとしてエンコードされる。

【０３０８】

プール記述子、１つまたは複数のプール・アイテム、または終了プール記述子からエンコードするビットは、８ビットの履歴および４最下位ビット（ＬＳＢ）のインデックスと結合される。ビットは、ルックアップ・テーブルを使用してエンコードされる。ビットは、１塩基当たり１～２ビットのレートでエンコードされる。生成される配列の各々の長さは、約２００～約５００塩基の長さである。ルックアップ・テーブルを使用して塩基候補が生成される。同じ塩基が隣り合ってエンコードされるのを防ぐために、塩基反復チェックがさらに実行される。例えば、２つの塩基（例えば、「ＡＡ」）が隣り合っている場合、塩基のうちの１つが更新される（例えば、「ＡＴ」）。その後、ビット履歴が更新され、レーン・インデックスおよび／またはフレーム・インデックスがインクリメントされる。新しいビット履歴が、エンコードする後続のビットに追加される。

【０３０９】

その後、オリゴヌクレオチド配列の塩基に対してＧＣフィルタリングが実行される。オリゴヌクレオチドの約５％～約１０％が、ＧＣフィルタリング中に除去される。最終的なプール内の塩基含有は、オリゴヌクレオチドであり、約４５％～約５５％のＧＣ含量である。

【０３１０】

次に、最終的なオリゴヌクレオチド・プールが合成され、格納される。

【実施例７】

【０３１１】

ＤＮＡへのデジタル情報のエンコーディングおよびデコーディング

２つの１ＴＢのオブジェクトが１０００個のプールに格納され、各プールは、２ＧＢの最大ペイロード・サイズを有する。各オブジェクトを２ＧＢのプール・アイテムに分割することによって、両方のプール・レイアウトが最初に計算されて検証され、プール・アイテムの範囲およびオフセットが割り当てられる。各オブジェクトは、低レベル・コーデックによって２ＧＢのセグメントで処理される。それと同時に、各セグメントおよび各オブジェクトのハッシュが同時に計算され、プール終了記述子に追加される。

【０３１２】

デコーディング中に、低レベル・コーデックは、各セグメントをストリーミングする。プールの識別情報が必ずしも知られていないため、既にデコードされたオブジェクトのセクションを追跡しながら、セグメントが、各オブジェクトを表す送信先ファイルにストリーミングされる。高レベル・コーデックが、オブジェクトのすべてのセグメントがデコードされたことを検出すると、デコーディングの完全性を確認するために、オブジェクト全体のハッシュが格納されたハッシュと比較される。その後、オブジェクトに完了としてフ

ラグを立てることができ、エンド・ユーザのための準備ができる。
【実施例８】
【０３１３】
ＤＮＡへの画像の格納および取り出し
画像が収集され、ディレクトリに整理された。収集された画像の合計サイズは、約５０メガバイトだった。低レベル・コーデック（例えば、実施例１の一般的な手順に従う、図１および図３～図６に概して示されたようなエンコーディング方式）および高レベル・コーデック（例えば、実施例６の一般的な手順に従う、図１２および図１４～図１５に概して示されたようなエンコーディング方式）を含むエンコーディング・ソフトウェアが、２２－ｍｅｒのフォワード・プライマーおよび１８－ｍｅｒのリバース・プライマーを含む１９８－ｍｅｒのオリゴ長を対象として実行された。インクジェットベースの合成法を使用して、固体支持体を備えるチップ上で、オリゴが合成された。このチップは、合成のための１６０万個を超える固有の部位を含んでいた。したがって、１チップ当たりのデータ量は、選択されたコーデック・パラメータの場合、約１０メガバイトだった。コーデック・ソフトウェアは、データを自動的に５つのチップに分割し、合成される１チップ当たり最大１６０万個のオリゴを生成した。
【０３１４】
インクジェットベースのシンセサイザで、標準的なプロトコルを使用して５回の実行が実行され、５つのオリゴ・プールが得られた。各プールがＩｌｌｕｍｉｎａシーケンサでシーケンシングされ、結果として得られた読み取りがランダムに選定され、平均５倍のシーケンシング・カバレッジを得た。低レベル・コーデック（例えば、実施例２の一般的な手順に従うか、または図２および図８～図９に概して示されたようなデコーディング方式）および高レベル・コーデック（例えば、実施例７の一般的な手順に従うか、または図１３に概して示されたようなデコーディング方式）を含むデコーディング・ソフトウェアが、５回のシーケンシング実行のすべてで実行され、各プールを自動的にデコードし、データをすべての画像に再び縫い合わせた。すべての画像が、元のディレクトリ構造と共に正常に回復された。コーデックは、複数のレベルのＳＨＡ２５６ハッシュを使用して、回復されたデータおよびデータ構造が１００％正確であることを検証した。
【実施例９】
【０３１５】
ＤＮＡへのＰＤＦの格納および取り出し
約１メガバイトの合計サイズを有する、科学出版物の５つのＰＤＦが提供された。低レベル・コーデック（例えば、実施例１の一般的な手順に従う、図１および図３～図６に概して示されたようなエンコーディング方式）および高レベル・コーデック（例えば、実施例６の一般的な手順に従う、図１２および図１４～図１５に概して示されたようなエンコーディング方式）を含むエンコーディング・ソフトウェアが、提供されたそのままのファイルに対して、２２－ｍｅｒのフォワード・プライマーおよび１８－ｍｅｒのリバース・プライマーを含む１９８－ｍｅｒのオリゴ長を対象として実行された。インクジェットベースの合成法を使用して、固体支持体を備えるチップ上で、オリゴが合成され、約１０００００個のオリゴの１つのプールが得られた。
【０３１６】
このプールは、完全な品質管理（ＱＣ：ｑｕａｌｉｔｙ　ｃｏｎｔｒｏｌ）を行うために、低レベル・コーデック（例えば、実施例２の一般的な手順に従うか、または図２および図８～図９に概して示されたようなデコーディング方式）および高レベル・コーデック（例えば、実施例７の一般的な手順に従うか、または図１３に概して示されたようなデコーディング方式）を使用して、増幅され、シーケンシングされ、デコードされた。プールは、複数のコピーに分割され、個々のカプセルに格納された。カプセルは、シーケンシング命令およびコンテナ化された（ｄｏｃｋｅｒｉｚｅｄ）デコーダと共にユーザに送られた。ユーザによってカプセルが開かれ、シーケンシングが実行され、ユーザのコンピュータ上でデコーダが実行された。元のＰＤＦデータが正常に回復され、読み取りプロセスが

ユーザのコンピュータ上で実行され得ることを示した。

【実施例１０】

【０３１７】

ＤＮＡへのペイロードの格納および取り出し

代表的なコンテンツを作成するために、ランダムに選択されたさまざまなダウンロードされたファイルを使用して１ギガバイトのペイロードが構築され、ディレクトリに格納された。低レベル・コーデック（例えば、実施例１の一般的な手順に従う、図１および図３～図６に概して示されたようなエンコーディング方式）および高レベル・コーデック（例えば、実施例６の一般的な手順に従う、図１２および図１４～図１５に概して示されたようなエンコーディング方式）を含むエンコーディング・ソフトウェアが、ディレクトリ上で実行され、おおよそ１億個のオリゴが得られた。シミュレーション・ソフトウェアが実行され、２％の欠失率、１％の突然変異率、および０．５％の挿入率、ならびに5倍のシーケンシング・カバレッジを使用して、このプールの合成、格納、およびシーケンシングをシミュレートし、１０億個の読み取り結果が得られた。

【０３１８】

低レベル・コーデック（例えば、実施例２の一般的な手順に従うか、または図２および図８～図９に概して示されたようなデコーディング方式）および高レベル・コーデック（例えば、実施例７の一般的な手順に従うか、または図１３に概して示されたようなデコーディング方式）を含むデコーディング・ソフトウェアが実行され、元の１ギガバイトのペイロードを正常に回復した。このプロセスは、Ｕｂｕｎｔｕ　Ｌｉｎｕｘ（登録商標）を実行している８コアＩｎｔｅｌ（登録商標）９ｉ上で２時間かかった。

【実施例１１】

【０３１９】

エラーを含むペイロードの格納および取り出し

乱数生成器を使用して、１メガバイトのペイロードが構築された。低レベル・コーデック（例えば、実施例１の一般的な手順に従う、図１および図３～図６に概して示されたようなエンコーディング方式）を含むエンコーディング・ソフトウェアが、冗長性およびコードワード・テーブルの変化するパラメータを使用して、結果として得られたペイロードに対して繰り返し実行された。シミュレーション・ソフトウェアが繰り返し実行され、エラーなし～最大６％の欠失率、３％の突然変異率、および１％の挿入率、ならびにオリゴ損失なし～２０％のオリゴ損失の範囲にわたるパラメータについて、合成、格納、およびシーケンシングをシミュレートした。

【０３２０】

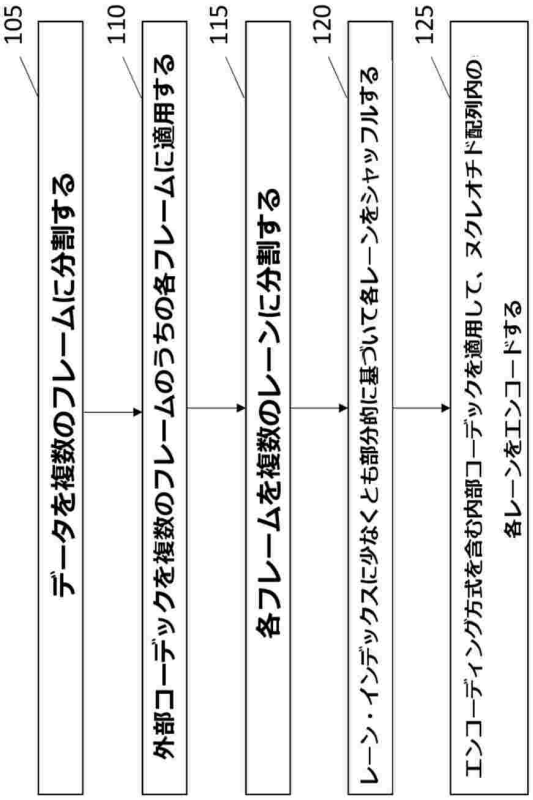
デコーディング・ソフトウェア（例えば、実施例２の一般的な手順に従う、または図２および図８～図９に概して示されたようなデコーディング方式）が、各シミュレーション結果に対して実行され、純粋な内部貪欲デコーダ（例えば、図８）から純粋な最尤デコーダ（例えば、図９）までの異なるデコード戦略を使用して、完全なペイロードの回復を検証した。貪欲デコーダは、３％の欠失率、２％の突然変異率、および０．５％の挿入率で失敗し始めた。最尤デコーダは、６％の欠失率、３％の突然変異率、および１％の挿入率まで正常に機能し、かなり低い合成の品質およびシーケンシングの品質を示した。参考のために、本明細書に記載された一般的な合成方法、例えば、実施例８および９で使用された合成方法は、０．１％の欠失率、０．１％の突然変異率、および０．０５％の挿入率よりも良い結果を有する。

【０３２１】

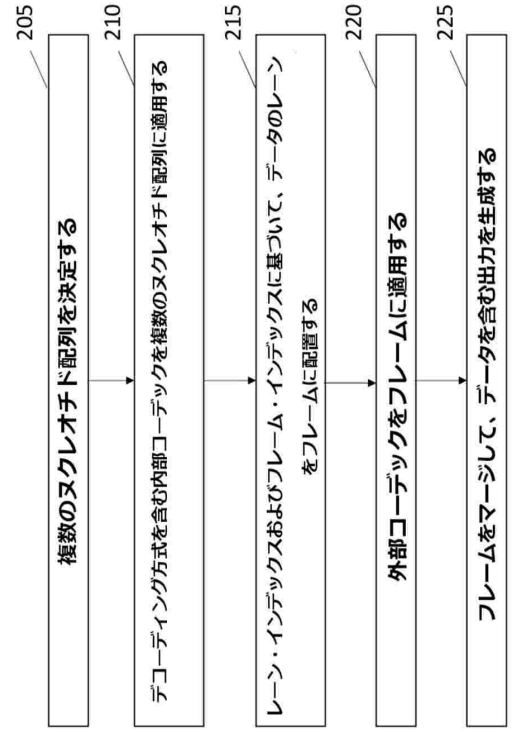
本明細書では、本主題の好ましい実施形態が示され、説明されたが、そのような実施形態が単に例として提供されているということは、当業者にとって明らかであろう。当業者は、本主題から逸脱することなく、多数の変形、変更、および置換に思い当たるであろう。本明細書に記載された本主題の実施形態のさまざまな代替手段が、本主題の実践において採用されてよいということが、理解されるべきである。

【図面】

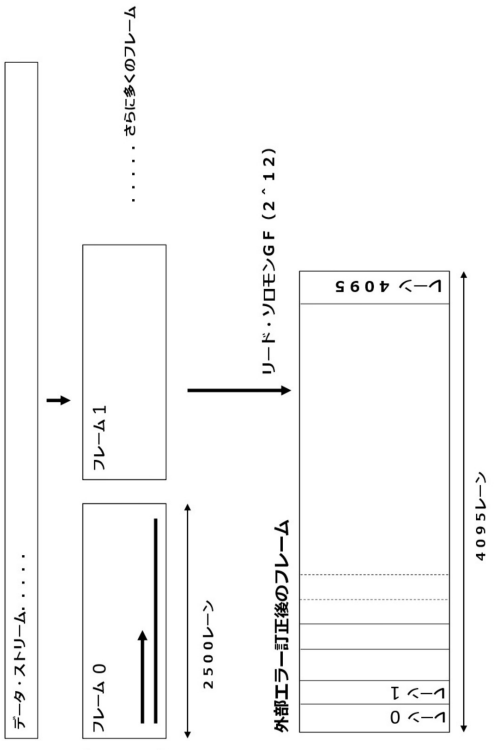
【図１】


105
データを複数のフレームに分割する
110
外部コーデックを複数のフレームのうちの各フレームに適用する
115
各フレームを複数のレーンに分割する
120
レーン・インデックスに少なくとも部分的に基づいて各レーンをシャッフルする
125
エンコーディング方式を含む内部コーデックを適用して、ヌクレオチド配列内の各レーンをエンコードする

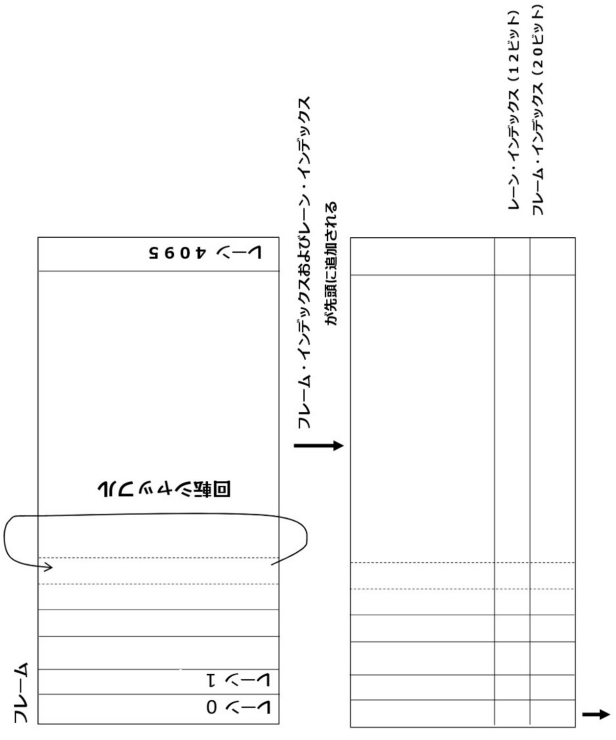
【図２】


205
複数のヌクレオチド配列を決定する
210
デコーディング方式を含む内部コーデックを複数のヌクレオチド配列に適用する
215
レーン・インデックスおよびフレーム・インデックスに基づいて、データのレーンをフレームに配置する
220
外部コーデックをフレームに適用する
225
フレームをマージして、データを含む出力を生成する

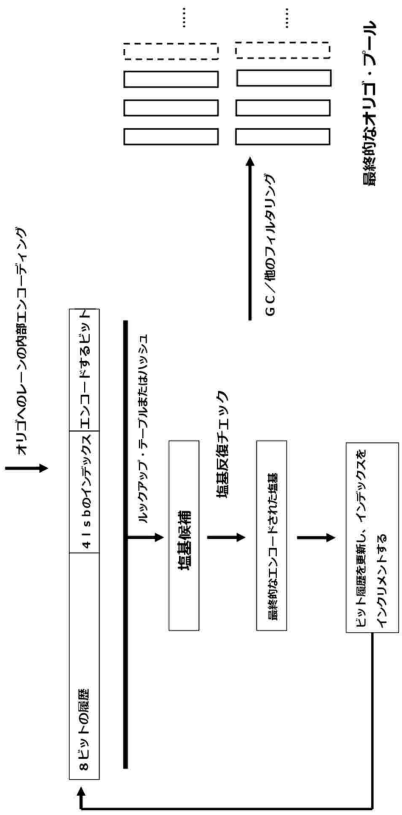
【図３】


データ・ストリーム・・・・・
フレーム 0
フレーム 1
・・・・・さらに多くのフレーム
144ビット
2500レーン
リード・ソロモンGF（2^12）
外部エラー訂正後のフレーム
レーン 0
レーン 1
レーン 4095
4095レーン

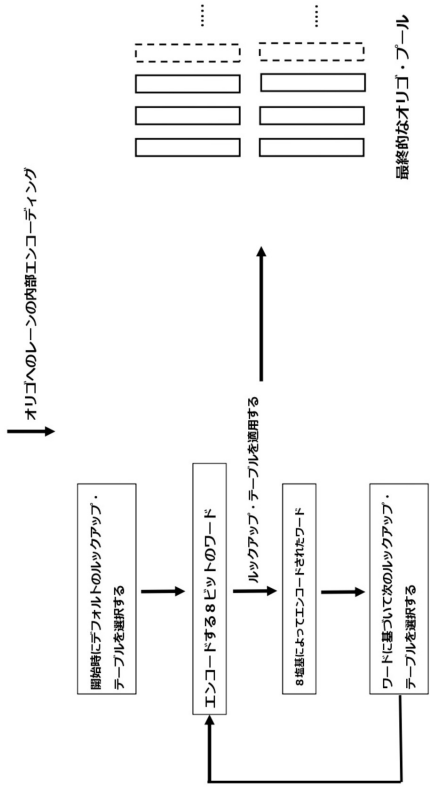
【図４】


フレーム
レーン 0
レーン 1
レーン 4095
回転シャッフル
フレーム・インデックスおよびレーン・インデックスが先頭に追加される
レーン・インデックス（12ビット）
フレーム・インデックス（20ビット）

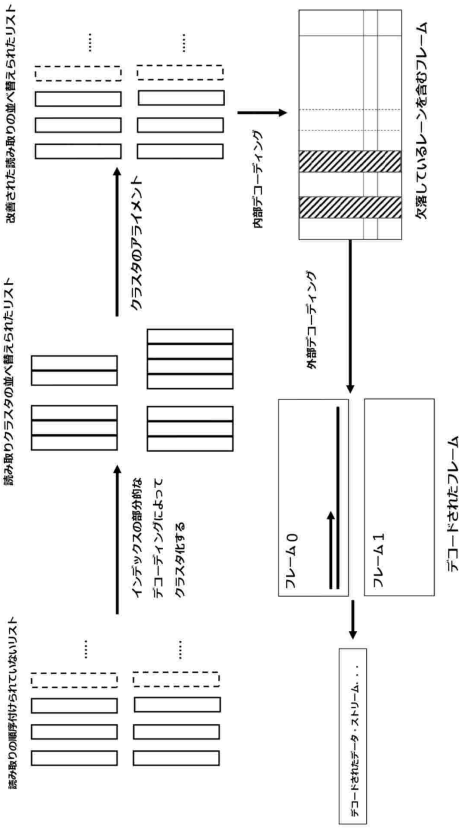
【図５】


オリゴへのレーンの内部エンコーディング
８ビットの履歴
４１ｓｂのインデックス
エンコードするビット
ルックアップ・テーブルまたはハッシュ
塩基候補
塩基反復チェック
最終的なエンコードされた塩基
ヒット履歴を更新し、インデックスをインクリメントする
ＧＣ／他のフィルタリング
最終的なオリゴ・プール

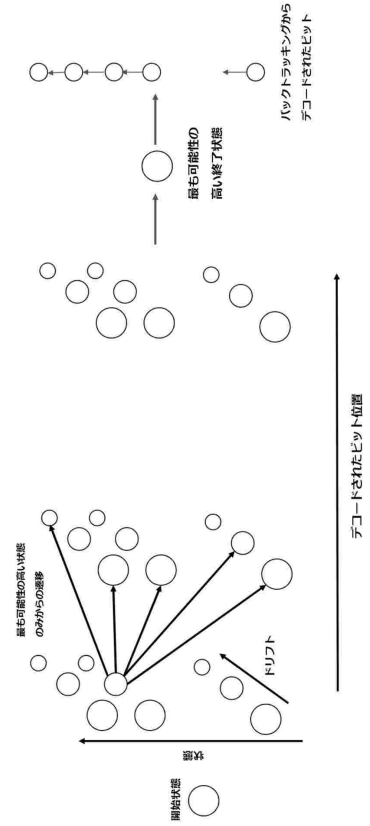
【図６】


オリゴへのレーンの内部エンコーディング
開始時にデフォルトのルックアップ・テーブルを選択する
エンコードする８ビットのワード
ルックアップ・テーブルを適用する
８塩基によってエンコードされたワード
ワードに基づいて次のルックアップ・テーブルを選択する
最終的なオリゴ・プール

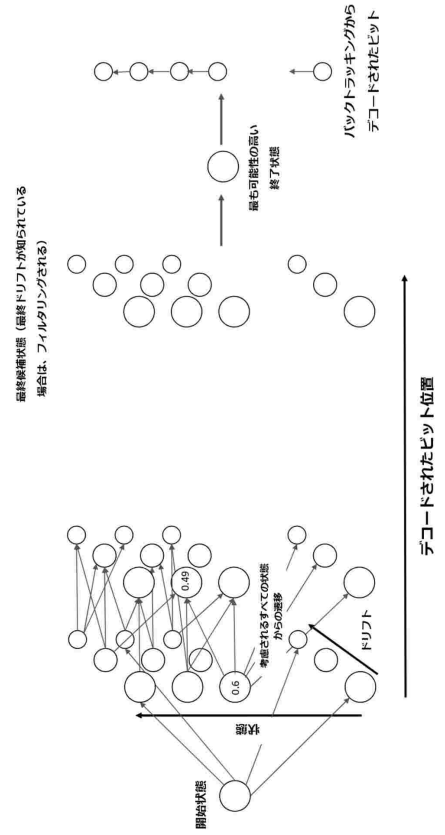
【図７】


読み取りの順序付けられていないリスト
インデックスの部分的なデコーディングによってクラスタ化する
読み取りクラスタの並べ替えられたリスト
クラスタのアライメント
改善された読み取りの並べ替えられたリスト
内部デコーディング
欠落しているレーンを含むフレーム
外部デコーディング
フレーム０
フレーム１
デコードされたフレーム
デコードされたデータ・ストリーム．．．

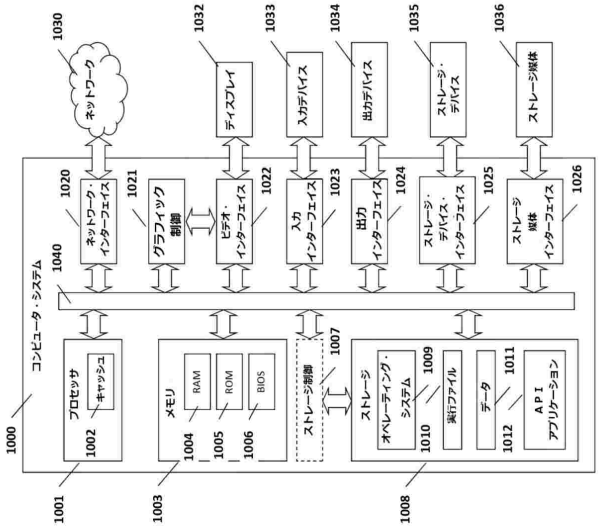
【図８】


バックトラッキングからデコードされたヒット
最も可能性の高い終了状態
最も可能性の高い状態のみからの遷移
ドリフト
状態
開始状態
デコードされたヒット位置

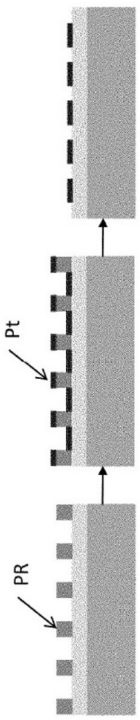
【図９】



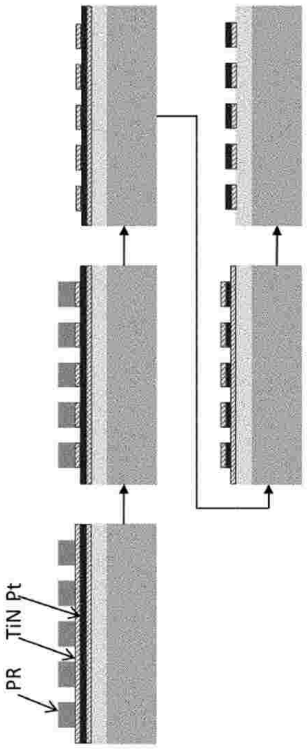
【図１０】



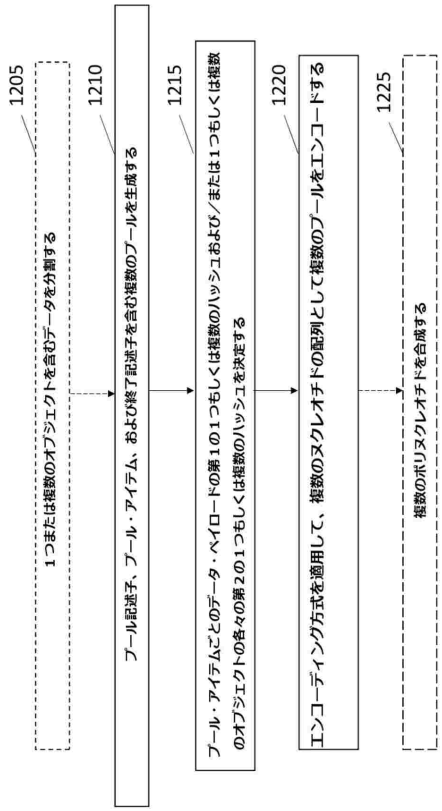
【図１１Ａ】



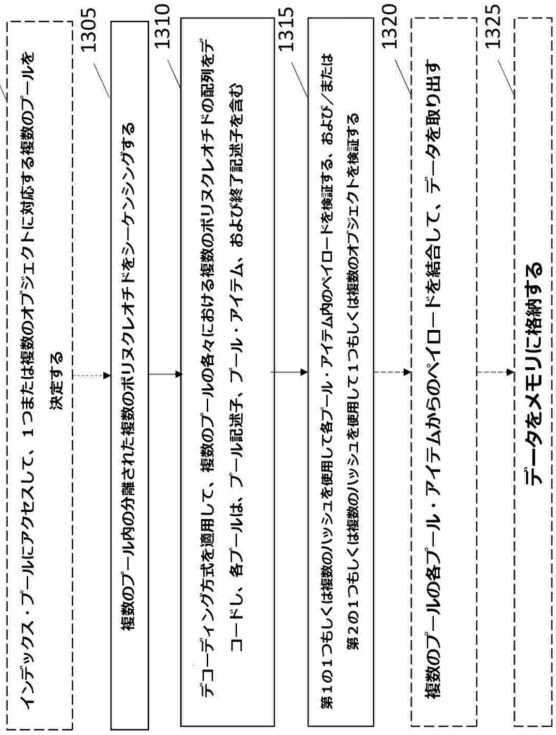
【図１１Ｂ】



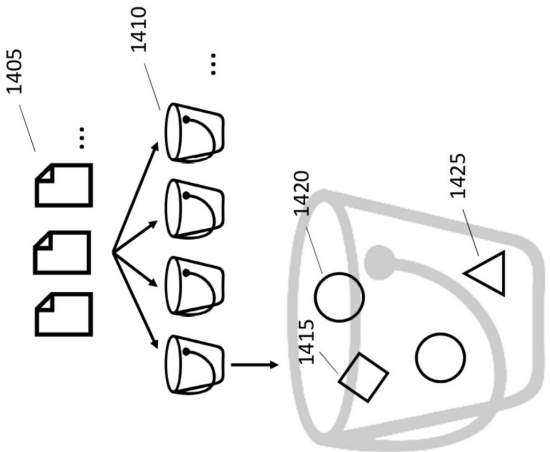
【図１２】



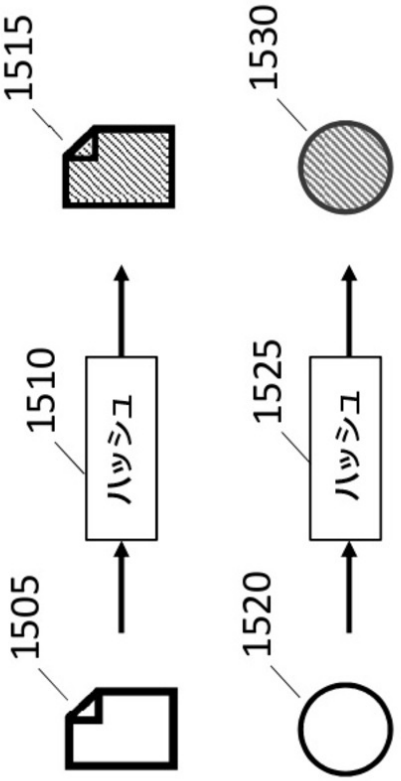
【図１３】



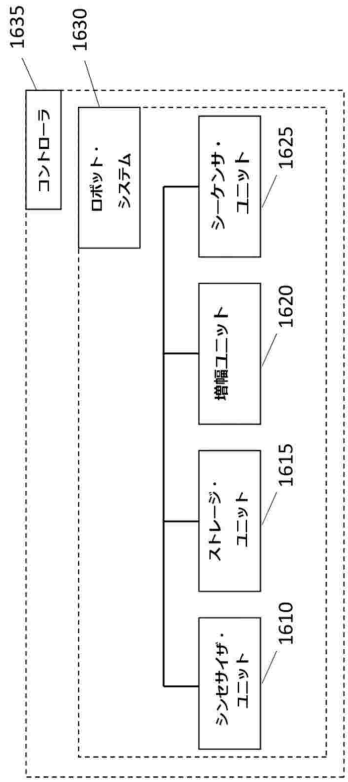
【図１４】



【図１５】



【図１６】



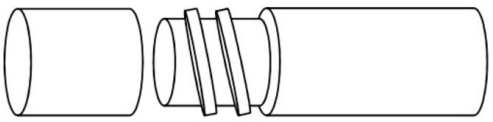
【図１７Ａ】



【図１７Ｂ】



【図１７Ｃ】



【図１７Ｄ】



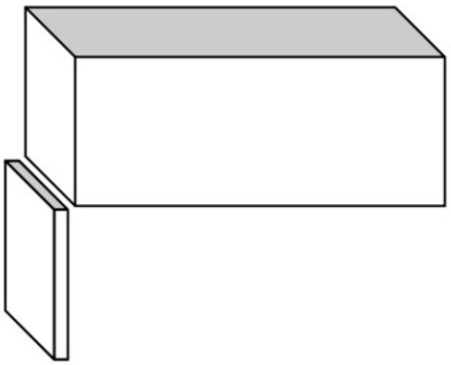
【図１７Ｅ】



【図１７Ｆ】



【図１７Ｇ】



【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 23/19283

A. CLASSIFICATION OF SUBJECT MATTER
IPC - INV. G16B 30/10, G06F 11/08, G06F 16/901, H03M 13/00 (2024.01)
ADD. H03M 7/00 (2024.01)

CPC - INV. G16B 30/10, G06F 11/08, G06F 16/901, H03M 13/00

ADD. H03M 7/00

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.
Y	US 2019/0050495 A1 (Intel Corporation), 14 February 2019 (14.02.2019), entire document	1-15, 17-20, 27-37, 42-54
A		- 16, 21-26, 38-41
Y	Organick et al. "Scaling up DNA data storage and random access retrieval." In: bioRxiv preprint, March 7, 2017, [online] [retrieved on 15 August 2023 (15.08.2023)] Retrieved from the Internet < URL: https://www.biorxiv.org/content/10.1101/114553v1>, entire document	1-15, 17-20, 27-37, 42-54
A		- 16, 21-26, 38-41
Y	US 2010/0058149 A1 (Lunglmayr et al.), 04 March 2010 (04.03.2010), entire document	
A		30-37, 42-52, 54
Y	Wang "Coding for DNA-Based Storage." In: Master's Thesis In Communication Engineering, Department of Electrical Engineering Communication Systems Group Chalmers University of Technology Gothenburg, Sweden 2021, [online] [retrieved on 15 August 2023 (15.08.2023)] Retrieved from the Internet < URL: https://odr.chalmers.se/server/api/core/bitstreams/3da1700d-ad8e-475d-b6af-a71a5aeca098/content >, entire document	- 16, 21-26, 38-41 42-48
A	US 2021/0074380 A1 (Microsoft Technology Licensing, LLC), 11 March 2021 (11.03.2021), entire document	
A	US 2021/0248318 A1 (AT & T Intellectual Property I, L.P.), 12 August 2021 (12.08.2021), entire document	1-54

☐ 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
16 August 2023	SEP 22 2023
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Kari Rodriquez Telephone No. PCT Helpdesk: 571-272-4300

Form PCT/ISA/210 (second sheet) (July 2022)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 23/19283

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.:
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:
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 searched, the appropriate additional search fees must be paid.

Group I - claims 1-54 are directed to a method for encoding data represented by a plurality of symbols in a plurality polynucleotide sequences.

Group II - claims 55-152 are directed to a hybrid organic-in silico platform and method for encoding data in polynucleotide sequences.

Group III - claims 153-268 are directed to a system for storing data in DNA.

--- (See Continuation in Supplemental Box) ---

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.:
1-54

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/US 23/19283

Continuation of Box No. III, Observations where unity of invention is lacking:

The inventions listed as 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:

Special Technical Features:

The invention of Group I included the features of splitting data into a plurality of frames, wherein each frame in the plurality of frames comprises a frame index; dividing each frame into a plurality of lanes, wherein each lane in the plurality of lanes comprises a lane index; shuffling each lane based at least in part on the lane index; and applying an inner codec to encode each lane in a polynucleotide sequence of the plurality of polynucleotide sequences, not required by any other group.

The invention of Group II included the features of generating an inner codec comprising a codebook, wherein the codebook is optimized for one or more constraints; a synthesizer for generating a plurality of polynucleotides comprising the plurality of polynucleotide sequences, not required by any other group.

The invention of Group III included the features of generate a plurality of pools, wherein each of the plurality of pools comprises a pool descriptor, a pool item comprising a payload of the data, and an end descriptor; determine a first one or more hashes of the payload for each pool item; verifying at least the payload of each pool item using a first one or more hashes, not required by any other group.

Common Technical Features

Groups I-III share the features of (a) a computing system comprising at least one processor and instructions executable by the at least one processor to perform operations comprising: applying the inner codec to encode the data as a plurality of polynucleotide sequences.

However, the shared technical features do not represent a contribution over prior art as being anticipated by US 2021/0248318 A1 (AT & T Intellectual Property I, L.P.), 12 August 2021 (12.08.2021).

AT & T Intellectual Property I, L.P teaches a computing system comprising at least one processor and instructions executable by the at least one processor to perform operations comprising (para [0018]- Server 102 further includes a processing device 108 communicatively coupled to a non-transitory memory device 110. Non-transitory memory device 110 is used to store computer program code instructions 112 for causing processing device 108); applying the inner codec to encode the data as a plurality of polynucleotide sequences (para [0018], [0024]-[0025], [0028] - causing processing device 108 to perform operations for DNA encoding of text; processing device 108 determines the modulus of the kth power of two of F for encoding the word in a k-mer DNA sequence; DNA sequence consists of four nucleotides represented by the letters A, T, C and G. Each nucleotide character, A, T, C and G, needs two bits for storage).

As the common features were known in the art at the time of the invention, this cannot be considered a common technical feature that would otherwise unify the groups. Therefore, Groups I-III lack unity under PCT Rule 13.

フロントページの続き

(32)優先日　令和5年1月27日(2023.1.27)
(33)優先権主張国・地域又は機関
米国(US)
(81)指定国・地域　AP(BW,CV,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,ME,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,CV,CZ,DE,DJ,DK,DM,DO,DZ,EC,EE,EG,ES,FI,GB,GD,GE,GH,GM,GT,HN,HR,HU,ID,IL,IN,IQ,IR,IS,IT,JM,JO,JP,KE,KG,KH,KN,KP,KR,KW,KZ,LA,LC,LK,LR,LS,LU,LY,MA,MD,MG,MK,MN,MU,MW,MX,MY,MZ,NA,NG,NI,NO,NZ,OM,PA,PE,PG,PH,PL,PT,QA,RO,RS,RU,RW,SA,SC,SC,SD,SE,SG,SK,SL,ST,SV,SY,TH,TJ,TM,TN,TR,TT,TZ,UA,UG,US,UZ,VC,VN,WS,ZA,ZM,ZW

（特許庁注：以下のものは登録商標）
１．ＶＩＳＵＡＬ ＢＡＳＩＣ
２．ＪＡＶＡＳＣＲＩＰＴ
３．ＤＢ２
４．ＭＯＮＧＯＤＢ
トウェイ・ブールバード ６８１