

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6457811号
(P6457811)

(45) 発行日 平成31年1月23日(2019. 1. 23)

(24) 登録日 平成30年12月28日(2018. 12. 28)

(51) Int.Cl.

F I

C 1 2 Q 1/6869 (2018. 01)
C 1 2 M 1/00 (2006. 01)
C 1 2 M 1/34 (2006. 01)
G O 1 N 27/00 (2006. 01)
G O 1 N 27/02 (2006. 01)

C 1 2 Q 1/6869 Z N A Z
C 1 2 M 1/00 A
C 1 2 M 1/34 Z
G O 1 N 27/00 Z
G O 1 N 27/02 D

請求項の数 30 (全 63 頁) 最終頁に続く

(21) 出願番号 特願2014-531315 (P2014-531315)
(86) (22) 出願日 平成24年9月21日(2012. 9. 21)
(65) 公表番号 特表2014-531901 (P2014-531901A)
(43) 公表日 平成26年12月4日(2014. 12. 4)
(86) 国際出願番号 PCT/GB2012/052343
(87) 国際公開番号 W02013/041878
(87) 国際公開日 平成25年3月28日(2013. 3. 28)
審査請求日 平成27年9月24日(2015. 9. 24)
(31) 優先権主張番号 61/538, 721
(32) 優先日 平成23年9月23日(2011. 9. 23)
(33) 優先権主張国 米国 (US)
(31) 優先権主張番号 61/617, 880
(32) 優先日 平成24年3月30日(2012. 3. 30)
(33) 優先権主張国 米国 (US)

(73) 特許権者 511252899
オックスフォード ナノボール テクノロ
ジーズ リミテッド
イギリス国 オックスフォードシャー オ
ーエックス4 4 ジーエー, オックスフォ
ード, オックスフォード サイエンス パ
ーク, ロバート ロビンソン アベニュー
4, エドマンド カートライト ハウス
(74) 代理人 100092783
弁理士 小林 浩
(74) 代理人 100120134
弁理士 大森 規雄
(74) 代理人 100104282
弁理士 鈴木 康仁

最終頁に続く

(54) 【発明の名称】 ポリマー単位を含むポリマーの解析

(57) 【特許請求の範囲】

【請求項 1】

ナノポア中を通るポリマーの移行中に行われる測定からポリマー中のポリマー単位の配列を推定する方法であって、

それぞれの測定の値が、 k が 2 以上の正整数である k 個のポリマー単位のグループである k マーに依拠しており、

可能な k マーのセットについて、

起点 k マーから目的地 k マーまでの遷移の可能性を表す遷移重み付け、および

その k マーについての所与の測定値を観測する可能性を表すそれぞれの k マーに関する放出重み付けを含むモデルを提供するステップ、

グループにおける測定の数についての先験的知識なしで、複数の測定のグループが同じ k マーに依拠している測定の入力シリーズを含む少なくとも 1 つの入力シグナルを受けるステップ、

少なくとも 1 つの入力シグナルを処理して、連続するグループの測定を同定し、それぞれの同定されたグループに関して予め定められた数の測定を導き出し、前記予め定められた数は 1 以上であり、これにより入力シグナルまたはそれぞれの入力シグナルに関する測定のシリーズを導き出すステップ、ならびに

これにより導き出された測定のシリーズを前記モデルを参照する解析技法を使用して解析し、測定のシリーズが、ポリマー単位の配列により生み出されるという前記モデルにより予測される尤度に基づいて、ポリマー中のポリマー単位の少なくとも 1 つの推定された

配列を推定するステップ
を含む方法。

【請求項 2】

放出重み付けがあらゆる可能な測定を観測する非ゼロの可能性を表す、請求項 1 に記載の方法。

【請求項 3】

それぞれの k マーに関する放出重み付けが測定値にわたり単峰性または多峰性分布を有する、請求項 1 または 2 に記載の方法。

【請求項 4】

それぞれの k マーに関する放出重み付けが測定値にわたりガウス分布、ラプラス分布、四角分布または三角分布を有する、請求項 3 に記載の方法。

10

【請求項 5】

遷移重み付けが、起点 k マーから、最初の $(k-1)$ 個のポリマー単位が前記起点 k マーの最後の $(k-1)$ 個のポリマー単位である配列を有する目的地 k マーまでの遷移である好ましい遷移の非ゼロ可能性を表し、起点 k マーから前記起点 k マーとは異なる配列を有し、最初の $(k-1)$ 個のポリマー単位が前記起点 k マーの最後の $(k-1)$ 個のポリマー単位ではない目的地 k マーまでの遷移である好ましくない遷移のより低い可能性を表す、請求項 1 から 4 のいずれか一項に記載の方法。

【請求項 6】

遷移重み付けが前記好ましくない遷移のうちの少なくとも一部の非ゼロ可能性を表す、請求項 5 に記載の方法。

20

【請求項 7】

遷移重み付けが、起点 k マーから、最初の $(k-2)$ 個のポリマー単位が前記起点 k マーの最後の $(k-2)$ 個のポリマー単位である配列を有する目的地 k マーまでの好ましくない遷移の非ゼロ可能性を表す、請求項 6 に記載の方法。

【請求項 8】

解析技法が確率的技法である、請求項 1 から 7 のいずれか一項に記載の方法。

【請求項 9】

遷移重み付けが確率であり、および／または放出重み付けが確率である、請求項 1 から 8 のいずれか一項に記載の方法。

30

【請求項 10】

モデルが隠れマルコフモデルである、請求項 1 から 9 のいずれか一項に記載の方法。

【請求項 11】

解析のステップが、測定のシリーズがポリマー単位の推定された配列により生み出されるというモデルにより予測される尤度を表す推定された配列またはそれぞれの推定された配列に関してクオリティスコアを導き出すことをさらに含む、請求項 1 から 10 のいずれか一項に記載の方法。

【請求項 12】

解析のステップが、ポリマー単位の推定された配列に対応する個々の k マーに関するクオリティスコアであって、測定のシリーズが個々の k マーを含む配列により生み出されるというモデルにより予測される尤度を表すクオリティスコアを導き出すことをさらに含む、請求項 1 から 11 のいずれか一項に記載の方法。

40

【請求項 13】

解析のステップが、ポリマー単位の推定された配列に対応する k マーの配列に関するクオリティスコアであって、測定のシリーズが k マーの所与の配列により生み出されるというモデルにより予測される尤度を表すクオリティスコアを導き出すことをさらに含む、請求項 1 から 12 のいずれか一項に記載の方法。

【請求項 14】

解析のステップがポリマー中のポリマー単位の複数の推定された配列を導き出す、請求項 1 から 13 のいずれか一項に記載の方法。

50

【請求項 1 5】

ポリマー中のポリマー単位の少なくとも 1 つの推定された配列を推定するステップが、測定シリーズが個々の k マーにより生み出されるというモデルにより予測される尤度に基づいて k マーの配列を推定するステップ、および k マーの推定された配列からポリマー単位の配列を推定するステップを含む、請求項 1 から 1 4 のいずれか一項に記載の方法。

【請求項 1 6】

ポリマー中のポリマー単位の少なくとも 1 つの推定された配列を推定するステップが、測定シリーズが k マーの全体の配列により生み出されるというモデルにより予測される尤度に基づいて k マーの少なくとも 1 つの配列を推定するステップ、および k マーの推定された配列からポリマー単位の配列を推定するステップを含む、請求項 1 から 1 5 のいずれか一項に記載の方法。

10

【請求項 1 7】

前記測定は、一方向のみに動く様式でナノポア中を通るポリマーの移行中に行われる、請求項 1 から 1 6 のいずれか一項に記載の方法。

【請求項 1 8】

ポリマーがポリヌクレオチドであり、ポリマー単位がヌクレオチドである、請求項 1 から 1 7 のいずれか一項に記載の方法。

【請求項 1 9】

ナノポアが生物学的ポアである、請求項 1 から 1 8 のいずれか一項に記載の方法。

20

【請求項 2 0】

測定が、電流測定、インピーダンス測定、トンネリング測定、FET 測定および光学的測定の中の 1 つまたは複数を含む、請求項 1 から 1 9 のいずれか一項に記載の方法。

【請求項 2 1】

方法が、それぞれが前記ポリマーに関係している測定の複数の入力シリーズで実施され、それぞれの測定の値が k マーに依拠しており、解析技法が、複数のそれぞれの次元で配置されている測定の前記複数のシリーズを扱う、請求項 1 から 2 0 のいずれか一項に記載の方法。

【請求項 2 2】

測定のそれぞれの入力シリーズが同じポリマーの同じ領域の測定である、請求項 2 1 に記載の方法。

30

【請求項 2 3】

測定の複数の入力シリーズが測定の 2 つのシリーズを含み、測定の最初の入力シリーズがポリマーの第 1 の領域の測定であり、測定の第 2 の入力シリーズが前記第 1 の領域に関係しているポリマーの第 2 の領域の測定である、請求項 2 1 に記載の方法。

【請求項 2 4】

前記第 1 の領域と第 2 の領域が同じポリマーの関係する領域である、請求項 2 3 に記載の方法。

【請求項 2 5】

前記関係する領域が相補的である、請求項 2 3 または 2 4 に記載の方法。

40

【請求項 2 6】

モデルがメモリに記憶される、請求項 1 から 2 5 のいずれか一項に記載の方法。

【請求項 2 7】

モデルを提供し測定を解析するステップが、ハードウェア装置においてまたはコンピュータ装置において実行される、請求項 1 から 2 6 のいずれか一項に記載の方法。

【請求項 2 8】

ポリマーの前記測定を行うことをさらに含む、請求項 1 から 2 7 のいずれか一項に記載の方法。

【請求項 2 9】

ポリマー中のポリマー単位の配列を前記ポリマーに関係のある測定の少なくとも 1 つの

50

シリーズから推定するための解析デバイスであって、それぞれの測定の値が、 k が2以上の正整数である k 個のポリマー単位のグループである k マーに依拠しており、方法が

可能な k マーのセットについて、

起点 k マーから目的地 k マーまでの遷移の可能性を表す遷移重み付け、および

その k マーの所与の測定値を観測する可能性を表すそれぞれの k マーに関する放出重み付けを含むモデルを記憶するメモリ、ならびに

グループにおける測定の数についての先験的知識なしで、複数の測定のグループが同じ k マーに依拠している測定の入力シリーズを含む少なくとも 1 つの入力シグナルを受け、

少なくとも 1 つの入力シグナルを処理して、連続するグループの測定を同定し、それぞれの同定されたグループに関して予め定められた数の測定を導き出し、前記予め定められた数は 1 以上であり、これにより入力シグナルまたはそれぞれの入力シグナルに関する測定のシリーズを導き出し、

これにより導き出された測定のシリーズを前記モデルを参照する解析技法を使用して、測定のシリーズがポリマー単位の配列により生み出されるというモデルにより予測される尤度に基づいて、ポリマー中のポリマー単位の少なくとも 1 つの推定された配列を推定するように構成された解析ユニット

を含む、解析デバイス。

【請求項 30】

請求項 29 に記載の解析デバイス、およびポリマーの前記測定を行うように構成された測定デバイス

を備える塩基配列決定装置。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は、一般的には、ポリマー単位を含むポリマーを解析する、例えば、限定されるものではないが、ポリヌクレオチドをポリマーに関係する測定を行うことにより解析する分野に関する。本発明の第一の態様は、具体的にはポリマー中のポリマー単位の配列を推定することに関する。本発明の第二および第三の態様は、ポリマーの解析のためにポリマーの移行中にナノポア中を流れるイオン電流の測定に関する。

【背景技術】

【0002】

ポリマーを解析しおよび／またはポリマー単位の配列を決定するためのポリマーの測定を提供する多くの種類の測定システムが存在する。

【0003】

例えば、限定されるものではないが、1つの種類の測定システムは、ポリマーを移行させる際に通るナノポアを利用する。前記システムのある特性は、ナノポア中のポリマー単位に依拠しており、その特性の測定が行われる。例えば、測定システムは、絶縁膜にナノポアを置き分析物分子の存在下でナノポアを通る電圧駆動イオン輸送を測定することにより作製しうる。ナノポアの性質に応じて分析物の正体を、その特徴的なイオン電流サイン、特に、電流ブロックの持続時間および程度ならびに電流レベルの分散を通じて明らかにしうる。ナノポアを使用するそのような種類の測定システムは、特に、DNA または RNA などのポリヌクレオチドの配列を決定する分野ではかなり有望であり、つい最近の開発の主題であった。

【0004】

幅広い適用にわたり迅速で安価な核酸（例えば、DNA または RNA）配列決定技術に対する必要性が現在存在する。既存の技術は、主に増幅技法を利用して大量の核酸を産生し、シグナル検出には高品質の専門の蛍光化学物質を必要とするために、時間がかかり高価である。ナノポア検知には、必要なヌクレオチドおよび試薬の量を減少させることにより迅速で安価な核酸配列決定を提供できる潜在性がある。

【発明の概要】

【発明が解決しようとする課題】

【0005】

本発明は、それぞれの測定値が、 k が正整数である k 個のポリマー単位（すなわち、 k マー）のグループに依拠している状況に関する。

【0006】

さらに、それぞれの測定値が、 k が複数の整数である k マーに依拠していることは、大多数の現在公知の生物学的ナノポアを含む多くの種類の測定システムに典型的である。これは、1つよりも多いポリマー単位が、観察されるシグナルに寄与しており、概念的には、測定されているポリマー単位よりも大きな「鈍いリーダーヘッド」を有する測定システムと見なしてもよいからである。そのような状況では、分解される異なる k マーの数が k の冪乗まで増加する。例えば、 n 個の可能なポリマー単位が存在する場合では、分解される異なる k マーの数は n^k である。異なる k マーについての測定間は明確に分離していることが望ましいが、これらの測定の一部が重複していることはよく起こる。特に大きな数のポリマー単位が k マー中にある、すなわち、大きな値の k では、異なる k マーにより生み出される測定を分解するのは困難になり、ポリマーについての情報、例えば、ポリマー単位の根底にある配列の推定値を導き出すのに障害となる。

10

【0007】

したがって、開発研究の多くは、測定の分解を改善する測定システムの設計に向けられてきた。根底にある物理的または生物学的系の固有の変動からばらついて生じることがある測定の変動および／または測定されている特性が小規模である結果避けられない測定ノイズによって、これは実際の測定システムでは困難である。

20

【0008】

多くの研究が、単一のポリマー単位に依拠している分解可能な測定を提供する測定システムの設計を目指してきた。しかし、これは実際には困難であった。

【0009】

他の研究は、 k が複数の整数である k マーに依拠している測定を受け入れたが、異なる k マーからの測定が互いに分解可能である測定システムの設計を目指してきた。しかし、現実の限界はこれが極めて困難であることを再び意味している。いくつかの異なる k マーが発するシグナルの分布は重複することが多いことがある。

【0010】

30

原理的には、それぞれが同じポリマー単位に部分的に頼ってポリマー単位のレベルで分解される単一の値を得る、 k が複数の整数である k 個の測定からの情報を組み合わせることは可能であるだろう。しかし、これは実際には困難である。第一に、これは、 k 個の測定のセットを変換する適切な変換を同定する可能性にかかっている。しかし、多くの測定システムでは、根底にある物理的または生物学的系における相互作用の複雑さのために、そのような変換は存在しないまたは同定するのが非現実的であるのいずれかである。第二に、所与の測定システムについてそのような変換が原理的に存在するとしても、測定の変動のために変換を同定するのは困難であるおよび／または変換はそれでも互いに分解することができない値を与える可能性がある。第三に、そのような技法では、例えば、測定を行うことができない測定システムによって、またはそれに続くデータ処理のエラーによって、実際の測定システムで時折ありうるように、見逃された測定、すなわち、所与の k マーに依拠している測定がポリマー単位の配列のなかで失われる場合を考慮に入れることは困難であるまたは不可能である。

40

【課題を解決するための手段】

【0011】

本発明の第一の態様は、 k マーに依拠しているような測定からポリマー中のポリマー単位の配列を推定する正確度を改善する技法を提供することに関する。

【0012】

本発明の第一の態様によれば、ポリマーに関係する測定の少なくとも1つのシリーズからポリマー中のポリマー単位の配列を推定する方法であって、それぞれの測定値が、 k

50

が正整数である k 個のポリマー単位のグループである k マーに依拠しており、可能な k マーのセットについて、起点 k マーから目的地 k マーまでの遷移の可能性を表す遷移重み付け、およびその k マーについて所与の測定値を観測する可能性を表すそれぞれの k マーに関する放出重み付けを含むモデルを提供するステップ、ならびに前記モデルを参照する解析技法を使用して測定シリーズを解析し、測定シリーズがポリマー単位の配列により生み出されるというモデルにより予測される尤度に基づいて、ポリマー中のポリマー単位の少なくとも 1 つの推定された配列を推定するステップを含む方法が提供される。

10

【0013】

さらに本発明の第一の態様によれば、類似する方法を実行する解析装置が提供される。

【0014】

したがって、本発明の第一の態様は、測定を生み出す測定システムのモデルを使用する。測定のいかなるシリーズを考慮しても、モデルはその測定を生み出した k マーの異なる配列の可能性を表す。本発明の第一の態様は、それぞれの測定の値が、 k が複数の整数である k マーに依拠している状況に特に適している。

【0015】

前記モデルは可能な k マーを考慮する。例えば、それぞれのポリマー単位が 4 個のポリマー単位（またはさらに一般的には n 個のポリマー単位）のうちの 1 つでありうるポリマーでは、どれか特定の k マーが物理的に存在しないのでなければ、 4^k 個の可能な k マー（またはさらに一般的には n^k 個の可能な k マー）が存在する。存在しうるすべての k マーでは、放出重み付けは所与の測定値を観測する可能性を考慮に入れる。それぞれの k マーに関する放出重み付けは、その k マーについての所与の測定値を観測する可能性を表す。

20

【0016】

遷移重み付けは、起点 k マーから目的地 k マーまでの遷移の可能性を表し、したがって、測定が依拠している k マーの、異なる k マー間を遷移する可能性を考慮に入れる。したがって、遷移重み付けは、多少なりとも可能性がある遷移を考慮しうる。例として、 k が複数の整数である場合、所与の起点 k マーでは、これは、起点 k マーとは異なる配列を有し最初の $(k-1)$ 個のポリマー単位が起点 k マーの最後の $(k-1)$ 個のポリマー単位ではない目的地 k マーまでの遷移である好ましくない遷移よりも、最初の $(k-1)$ 個のポリマー単位が起点 k マーの最後の $(k-1)$ 個のポリマー単位である配列を有する目的地 k マーまでの遷移である好ましい遷移のより大きな可能性を表しうる。例えば、ポリマー単位が天然に存在する DNA 塩基である 3 マーでは、状態 C G T は、G T C、G T G、G T T および G T A への好ましい遷移を有する。限定のない例として、モデルは、遷移重み付けおよび放出重み付けが確率である隠れマルコフモデルであってもよい。

30

【0017】

これにより、測定シリーズは、モデルを参照する解析技法を使用して解析することが可能になる。ポリマー中のポリマー単位の少なくとも 1 つの推定された配列が、測定シリーズがポリマー単位の配列により生み出されるというモデルにより予測される尤度に基づいて推定される。例えば、限定されるものではないが、解析技法は確率的技法であってもよい。

40

【0018】

特に、個々の k マーからの測定は互いに分解可能である必要はなく、同じポリマー単位に依拠している k 個の測定のグループからその変換に関しての値までの変換が存在する必要はない、すなわち、観測される状態のセットはもっと少数のパラメータの関数である必要はない（がこれは排除されない）。代わりに、モデルの使用は、測定シリーズがポリマー単位の配列により生み出されるというモデルにより予測される尤度を検討する際に複数の測定を考慮に入れることにより正確な推定を提供する。概念的には、遷移重み付けは

50

、一部はそのポリマー単位に、および実際、配列中のより大きな距離からの測定にも依拠している少なくとも k 個の測定を、所与のポリマー単位を推定する際に、モデルが考慮に入れることを可能にすると見なしてもよい。モデルは、所与のポリマー単位を推定する際に多数の測定を効果的に考慮に入れて、より正確でありうる結果を出すことができる。

【0019】

同様に、そのようなモデルを使用すれば、解析技法は所与の k マーからの失われた測定を考慮に入れるおよび／または所与の k マーにより生じる測定の外れ値を考慮に入れることが可能になる。これは、遷移重み付けおよび／または放出重み付けにおいて説明しうる。例えば、遷移重み付けは好ましくない遷移の少なくとも一部の非ゼロの可能性を表しうるおよび／または放出重み付けはあらゆる可能な測定を観測する非ゼロの可能性を表しうる。

10

【0020】

本発明の第二および第三の態様は、ポリマーがナノポア中を通して移行している間にナノポア中を流れるイオン電流の測定を使用してポリマーの解析を支援する技法の提供に関する。

【0021】

本発明の第二の態様によれば、ポリマー単位を含むポリマーを解析する方法であって、ナノポアを横断して電圧が印加されている間にナノポア中を通るポリマーの移行中に、 k が正整数である前記ポリマーの k 個のポリマー単位であるナノポア中の k マーの正体に依拠している測定を行い、前記測定が個々の k マーに関して、ナノポアを横断して印加される前記電圧の異なるレベルで行われる別々の測定を含むステップ、および前記電圧の前記異なるレベルでの測定を解析してポリマーの少なくとも一部の正体を決定するステップを含む方法が提供される。

20

【0022】

前記方法は、 k が正整数であるポリマーの k 個のポリマー単位であるナノポア中の k マーの正体に依拠している測定を行うことを伴う。特に、測定は、個々の k マーに関して、ナノポアを横断して印加された前記電圧の異なるレベルで行われる別々の測定を含む。本発明者らは、ナノポアを横断して印加された前記電圧の異なるレベルでのそのような測定は単に重複しているというよりもむしろ追加の情報を提供すると認識し実証してきた。例えば、異なる電圧での測定は、異なる状態の分解を可能にする。例えば、所与の電圧で分解することができないいくつかの k マーは別の電圧では分解することができる。

30

【0023】

本発明の第三の態様は、ナノポアを横断する電圧の異なるレベルの印加下で行われる測定を行う方法であって、場合により本発明の第二の態様で用いてもよい方法を提供する。特に、本発明の第三の態様によれば、ポリマー単位を含むポリマーの測定を行う方法であって、

電圧がナノポアを横断して印加されている間に、ナノポア中を通る前記ポリマーの移行を実施するステップ、

ナノポア中を通るポリマーの前記移行中に、前記電圧の異なるレベルを周期的に印加するステップ、および

40

k が正整数である前記ポリマーの k 個のポリマー単位であるナノポア中の k マーの正体に依拠している測定であり、前記測定が個々の k マーに依拠している状態よりも短い繰返し周期を有する前記周期での前記電圧の前記異なるレベルでの前記個々の k マーに関する別々の測定を含む測定を行うステップ

を含む方法が提供される。

【0024】

したがって、本発明の第三の態様は、本発明の第二の態様と同じ利点であって、特に測定が単に重複しているというよりもむしろ追加の情報を提供するという利点を与える。異なる電圧での測定は、それに続く測定の解析において異なる状態の分解を可能にする。例

50

えば、所与の電圧で分解することができないいくつかの状態は別の電圧では分解することができる。

【0025】

これは、異なる電圧での測定がナノポア中を通るポリマーの1回の移行中に得られる技術革新に基づいている。これは、繰返し周期が測定される状態の持続時間よりも短くなるように選択される周期で前記電圧のレベルを変化させることにより達成される。

【0026】

しかし、本発明の第二の態様内でこの方法を使用することは不可欠ではない。代案として、電圧の異なる大きさでのイオン電流測定は、同方向への移行でもよいし、または反対方向への移行を含んでいてもよいナノポア中を通るポリマーの異なる移行中に行うことができる。

10

【0027】

したがって、本発明の第二の態様および第三の態様の方法は、ポリマーについての情報を導き出す測定のその後の解析を改善する追加の情報を提供することができる。導き出さる情報の種類のいくつかの例は以下の通りである。

【0028】

解析は、状態間の遷移の時機を導き出すことでありうる。このケースでは、異なる電位でのそれぞれの状態の測定により提供される追加の情報は、正確度を改善する。例えば、2つの状態間の遷移を1つの電圧では分解することができないケースでは、遷移は別の電圧でのイオン電流測定のレベルの変化により同定しうる。これが、1つの電圧だけでは明らかな作動にならないと考えられる遷移を同定すること、または遷移が実際には起こらなかったことをより高い信頼度で決定することを潜在的に可能にする。この同定は、測定のそれに続く解析において使用しうる。

20

【0029】

一般に、異なる電圧レベルで測定を実行することは、1つの電圧レベルで得られるよりも多くの情報を提供する。例えば、ナノポア中を通るイオン流の測定において、測定から得られる情報には、特定の状態についての電流レベルおよびシグナル分散（ノイズ）が含まれる。例えば、ナノポア中を通るDNAの移行では、ヌクレオチド塩基Gを含むkマーはシグナル分散が増大した状態を生じる傾向がある。例えば、類似の電流レベルを有するそれぞれの状態によって、またはそれぞれの状態のうちの1つもしくは両方が高いシグナル分散を有する場合には、状態の遷移が起こったかどうかを確定するのは困難であることもある。特定の状態についての電流レベルおよびシグナル分散は異なる電圧レベルでは異なることがあり、したがって、異なる電圧レベルでの測定は、高分散状態の決定を可能にするまたは状態を決定する信頼度のレベルを増加しうる。その結果、別の電圧レベルと比べて1つの電圧レベルでの状態間の遷移を決定する方が容易でありうる。

30

【0030】

解析は、ポリマーの正体を推定するまたはポリマー中のポリマー単位の配列を推定することでありうる。このケースでは、異なる電位でのそれぞれの状態の測定により与えられる追加の情報は推定の正確度を改善する。

【0031】

40

ポリマー単位の配列を推定するケースでは、解析は本発明の第一の態様に従った方法を使用しうる。したがって、本発明の第一の態様の特長は、本発明の第二の態様および／または第三の態様の特長と、いかなる組合せでも組み合わせることができる。

【0032】

さらに、本発明の第二および第三の態様によれば、類似の方法を実行する解析装置が提供される。

ポリマーに関係する少なくとも1つのシリーズの測定からポリマー中のポリマー単位の配列を推定する方法であって、それぞれの測定値が、kが正整数であるk個のポリマー単位のグループであるkマーに依拠しており、可能なkマーのセットについて、起点kマーから目的地kマーまでの遷移の可能性を表す遷移重み付け、およびそのkマーについての

50

所与の測定値を観測する可能性を表すそれぞれの k マーに関する放出重み付けを含むモデルを提供するステップ、ならびに前記モデルを参照する解析技法を使用して測定シリーズを解析し、測定シリーズがポリマー単位の配列により生み出されるというモデルにより予測される尤度に基づいて、ポリマー中のポリマー単位の少なくとも1つの推定された配列を推定するステップを含む方法が好ましい。

遷移重み付けおよび放出重み付けのうちの少なくとも1つが非2値変数の値を含んでもよい。

遷移重み付けと放出重み付けの両方が非2値変数の値を含んでもよい。

放出重み付けがあらゆる可能な測定を観測する非ゼロの可能性を表してもよい。

それぞれの k マーに関する放出重み付けが測定値にわたり単峰性または多峰性分布を有してもよい。

それぞれの k マーに関する放出重み付けが測定値にわたりガウス分布、ラプラス分布、四角分布または三角分布を有してもよい。

k が複数の整数であってもよい。

遷移重み付けが、起点 k マーから、最初の $(k-1)$ 個のポリマー単位が前記起点 k マーの最後の $(k-1)$ 個のポリマー単位である配列を有する目的地 k マーまでの遷移である好ましい遷移の非ゼロ可能性を表し、起点 k マーから前記起点 k マーとは異なる配列を有し、最初の $(k-1)$ 個のポリマー単位が前記起点 k マーの最後の $(k-1)$ 個のポリマー単位ではない目的地 k マーまでの遷移である好ましくない遷移のより低い可能性を表してもよい。

遷移重み付けが前記好ましくない遷移のうちの少なくとも一部の非ゼロ可能性を表してもよい。

遷移重み付けが、起点 k マーから、最初の $(k-2)$ 個のポリマー単位が前記起点 k マーの最後の $(k-2)$ 個のポリマー単位である配列を有する目的地 k マーまでの好ましくない遷移の非ゼロ可能性を表してもよい。

解析技法が確率的技法であってもよい。

遷移重み付けが確率であり、および／または放出重み付けが確率であってもよい。

モデルが隠れマルコフモデルであってもよい。

解析のステップが、測定シリーズがポリマー単位の推定された配列により生み出されるというモデルにより予測される尤度を表す推定された配列またはそれぞれの推定された配列に関してクオリティスコアを導き出すことをさらに含んでもよい。

解析のステップが、ポリマー単位の推定された配列に対応する個々の k マーに関するクオリティスコアであって、測定シリーズが個々の k マーを含む配列により生み出されるというモデルにより予測される尤度を表すクオリティスコアを導き出すことをさらに含んでもよい。

解析のステップが、ポリマー単位の推定された配列に対応する k マーの配列に関するクオリティスコアであって、測定シリーズが k マーの所与の配列により生み出されるというモデルにより予測される尤度を表すクオリティスコアを導き出すことをさらに含んでもよい。

解析のステップがポリマー中のポリマー単位の複数の推定された配列を導き出してもよい。

ポリマー中のポリマー単位の少なくとも1つの推定された配列を推定するステップが、測定シリーズが個々の k マーにより生み出されるというモデルにより予測される尤度に基づいて k マーの配列を推定するステップ、および k マーの推定された配列からポリマー単位の配列を推定するステップを含んでもよい。

ポリマー中のポリマー単位の少なくとも1つの推定された配列を推定するステップが、測定シリーズが k マーの全体の配列により生み出されるというモデルにより予測される尤度に基づいて k マーの少なくとも1つの配列を推定するステップ、および k マーの推定された配列からポリマー単位の配列を推定するステップを含んでもよい。

測定の少なくとも1つのシリーズにおいて、予め定められた数の測定がそれぞれの k マ

10

20

30

40

50

ーに依拠しており、予め定められた数が1以上であってもよい。

グループにおける測定の数についての先験的知識なしで、複数の測定のグループが同じkマーに依拠している測定の入力シリーズを含む少なくとも1つの入力シグナルを受けるステップ、および解析のステップ前に、少なくとも1つの入力シグナルを処理して、連続するグループの測定を同定し、それぞれの同定されたグループに関して前記予め定められた数の測定を導き出し、解析のステップがこのようにして導き出された測定のシリーズまたはそれぞれの測定のシリーズで実施されるステップを含んでもよい。

少なくとも1つのシリーズの測定において、複数の測定のグループが、前記グループにおける測定の数について先験的知識なしで同じkマーに依拠してもよい。

ポリマーの前記測定を行うことをさらに含んでもよい。

ポリマーの前記測定がナノポア中を通るポリマーの移行中に行われてもよい。

ポリマーの移行が、複数の測定のグループが同じkマーに依拠するように実施されてもよい。

ナノポア中を通るポリマーの移行は一方方向のみに動く様式で実施されてもよい。

ポリマーがポリヌクレオチドであり、ポリマー単位がヌクレオチドであってもよい。

測定のシリーズが、ナノポア中を通るポリマーの移行中に行われる測定であってもよい。

ナノポアが生物学的ポアであってもよい。

測定が、電流測定、インピーダンス測定、トンネリング測定、FET測定および光学的測定のうちの1つまたは複数を含んでもよい。

方法が、それぞれが前記ポリマーに関係している測定の複数のシリーズで実施され、それぞれの測定の値がkマーに依拠しており、解析技法が、複数のそれぞれの次元で配置されている測定の複数のシリーズを扱ってもよい。

測定のそれぞれのシリーズが同じポリマーの同じ領域の測定であってもよい。

測定の複数のシリーズが測定の2つのシリーズを含み、測定の最初のシリーズがポリマーの第1の領域の測定であり、測定の第2のシリーズが前記第1の領域に関係しているポリマーの第2の領域の測定であってもよい。

前記第1の領域と第2の領域が同じポリマーの関係する領域であってもよい。

前記関係する領域が相補的であってもよい。

モデルがメモリに記憶されてもよい。

モデルを提供し測定を解析するステップが、ハードウェア装置においてまたはコンピュータ装置において実行されてもよい。

前記方法を実施するように構成されたデバイスが好ましい。

ポリマー中のポリマー単位の配列を前記ポリマーに関係のある測定の少なくとも1つのシリーズから推定するための解析デバイスであって、それぞれの測定の値が、kが複数の整数であるk個のポリマー単位のグループであるkマーに依拠しており、方法が可能なkマーのセットについて、起点kマーから目的地kマーまでの遷移の可能性を表す遷移重み付け、およびそのkマーの所与の測定値を観測する可能性を表すそれぞれのkマーに関する放出重み付けを含むモデルを記憶するメモリ、ならびに前記モデルを参照する解析技法を使用して、測定のシリーズを解析し、測定のシリーズがポリマー単位の配列により生み出されるというモデルにより予測される尤度に基づいて、ポリマー中のポリマー単位の少なくとも1つの推定された配列を推定するように構成された解析ユニットを含む、解析デバイスが好ましい。

ポリマーの前記測定を行うように構成された測定デバイス、および請求項38または39に記載の解析デバイスを備える塩基配列決定装置が好ましい。

ポリマー単位を含むポリマーを解析する方法であって、ナノポアを横断して電圧が印加されている間にナノポア中を通るポリマーの移行中に、kが正整数である前記ポリマーのk個のポリマー単位であるナノポア中のkマーの正体に依拠している測定を行い、前記測定が個々のkマーに関して、ナノポアを横断して印加される前記電圧の異なるレベルで行われる別々の測定を含むステップ、および前記電圧の前記異なるレベルで測定を解析して

10

20

30

40

50

ポリマーの少なくとも一部の正体を決定するステップを含む方法が好ましい。

測定を行う前記ステップが、異なる移行において電圧がナノポアを横断して異なるレベルで印加されている間に、ナノポア中を通る前記ポリマーの複数の移行を実施するステップ、前記異なる移行中に、ナノポアを横断する前記電圧の前記異なるレベルでの前記kマーの測定を行うステップを含んでもよい。

前記複数の移行がナノポア中を通る第1の方向への移行およびナノポア中を通る前記第1の方向とは反対の方向への移行を含んでもよい。

測定を行う前記ステップが、電圧がナノポアを横断して印加されている間に、ナノポア中を通る前記ポリマーの移行を実施するステップ、ナノポア中を通るポリマーの前記移行中に、前記測定が前記個々のkマーに依拠している状態の持続期間よりも短い繰返し周期を有する周期で前記電圧の前記異なるレベルを印加するステップ、および前記周期において前記電圧の前記異なるレベルでの前記個々のkマーに関して前記別々の測定を行うステップを含んでもよい。

10

ポリマー単位を含むポリマーの測定を行う方法であって、電圧がナノポアを横断して印加されている間に、ナノポア中を通る前記ポリマーの移行を実施するステップ、ナノポア中を通るポリマーの前記移行中に、前記電圧の異なるレベルを周期的に印加するステップ、およびkが正整数である前記ポリマーのk個のポリマー単位であるナノポア中のkマーの正体に依拠している測定であり、前記測定が個々のkマーに依拠している状態よりも短い繰返し周期を有する前記周期で前記電圧の前記異なるレベルでの前記個々のkマーに関する別々の測定を含む測定を行うステップを含む方法が好ましい。

20

繰返し周期が最長で3秒であってもよい。

繰返し周期が少なくとも0.5msであってもよい。

前記電圧の異なるレベルがそれぞれ、前記周期の部分的期間連続して印加されてもよい。

前記周期における前記電圧の前記異なるレベル間の遷移が、電圧変化により引き起こされる測定の容量性遷移を減少するように形作られてもよい。

測定を解析してポリマーの正体を決定することをさらに含んでもよい。

測定を解析してポリマーの正体を推定するステップが、測定を解析してポリマー中のポリマー単位の配列を推定することを含んでもよい。

測定を解析してポリマー中のポリマー単位の配列を推定するステップが、可能なkマーのセットについて、起点kマーから目的地kマーまでの遷移の可能性を表す遷移重み付け、およびそのkマーについての所与の測定値を観測する可能性を表すそれぞれのkマーに関する放出重み付けを含むモデルを提供するステップ、ならびに前記モデルを参照し、ナノポアを横断する電圧の異なるレベルの印加下で行われる測定を扱う解析技法を使用して測定を複数の次元での測定として解析し、測定のシリーズがポリマー単位の配列により生み出されるというモデルにより予測される尤度に基づいて、ポリマー中のポリマー単位の少なくとも1つの推定された配列を推定するステップを含んでもよい。

30

測定を解析してポリマーの正体を決定するステップが、前記異なる電圧レベルで行われる別々の測定を比較して、前記測定が前記個々のkマーに依拠している状態間の遷移を決定することをさらに含んでもよい。

40

電圧の前記異なるレベル間の違いが10mVから1.5Vの範囲であってもよい。

前記異なるレベルが2つの異なるレベルからなってもよい。

電圧の異なるレベルが同じ極性であってもよい。

前記測定がナノポア中を通るイオン電流の流れの測定であってもよい。

ナノポア中を通るイオン電流の流れの前記測定が、ナノポア中を通るDCイオン電流の流れの測定であってもよい。

前記電圧の前記異なるレベルのそれぞれ1つでグループの複数の測定を行うステップ、および前記異なるレベルのそれぞれ1つでの複数の測定のそれぞれのグループから1つまたは複数のサマリー測定を導き出して、個々のkマーに関して前記別々の測定を構成する

50

ステップを含んでもよい。

前記電圧の異なるレベルがそれぞれ一定期間連続して印加され、それぞれ各自の期間中、それぞれの期間中に印加される前記電圧の前記異なるレベルのうちの1つでグループのうちの1つの複数の測定を行ってもよい。

ポリマーがポリヌクレオチドであり、ポリマー単位がヌクレオチドであってもよい。

ナノポアが生物学的ポアであってもよい。

ナノポア中を通るポリマーの前記移行が、連続するkマーがナノポアで登録される一方向のみに動く様式で実施されてもよい。

ポリマーの移行が分子歯止めにより制御されてもよい。

分子歯止めが酵素であってもよい。

ポリマー単位を含むポリマーを解析するための装置であって、ポリマーが中を通して移行しうるナノポア、ナノポア中を通るポリマーの移行中にナノポアを横断して電圧を印加するように配置された制御回路、およびkが正整数であるポリマーのk個のポリマー単位であるkマーの正体に依拠している測定をナノポア中で行うように配置された測定回路であって、前記制御回路がナノポアを横断して電圧の異なるレベルを印加するように配置され、前記測定回路がナノポアを横断して印加される前記電圧の異なるレベルで、個々のkマーに関して別々の測定を行うように配置されている、測定回路、および前記電圧の前記異なるレベルで測定を解析してポリマーの少なくとも一部の正体を決定するように配置されている解析ユニットを備える装置が好ましい。

制御回路が、ナノポア中を通る前記ポリマーの異なる移行中にナノポアを横断して電圧の異なるレベルを印加するように配置されており、測定回路が前記電圧の異なるレベルでの前記異なる移行中に、個々のkマーに関して別々の測定を行うように配置されている。

制御回路が、ナノポア中を通るポリマーの前記移行中に、前記測定が前記個々のkマーに依拠している状態の持続時間よりも短い繰返し周期を有する周期で前記電圧の前記異なるレベルを印加するように配置されており、測定回路が前記周期において前記電圧の前記異なるレベルで、個々のkマーに関して別々の測定を行うように配置されている。

ポリマー単位を含むポリマーを測定するための装置であって、ポリマーが中を通して移行しうるナノポア、ナノポア中を通るポリマーの移行中に、前記測定が前記個々のkマーに依拠している状態の持続時間よりも短い繰返し周期を有する周期で前記電圧の異なるレベルを印加するように配置されている制御回路、およびナノポアを横断して印加される前記電圧の異なるレベルで、個々のkマーに関して別々の測定を行うように配置されている測定回路を備える装置が好ましい。

前記電圧の前記異なるレベルで測定を解析して、ポリマーの少なくとも一部の正体を決定するように配置されている解析ユニットをさらに備えてもよい。

【0033】

さらに十分な理解を可能にするため、本発明の実施形態は添付図面を参照して非限定的例としてここで説明されることになる。

【図面の簡単な説明】

【0034】

【図1】は、ナノポアを含む測定システムの概略図である。

【図2】は、測定システムにより経時的に測定された事象のシグナルのプロットである。

【図3】は、ナノポアを含む測定システムにおける2つの異なるポリヌクレオチドの測定の度数分布のグラフである。

【図4】は、実験的に導き出される電流測定のセットに適用される一次線形モデルから予測される値に対する64の3マー係数のプロットである。

【図5】は、実験的に導き出される電流測定値のセットに適用される一次線形モデルから予測される値に対する1024の5マー係数のプロットである。

【図6】は、ポリマーの測定を含む入力シグナルを解析する方法のフローチャートである。

。

【図 7】は、図 6 の状態検出ステップのフローチャートである。

【図 8】は、図 6 の解析ステップのフローチャートである。

【図 9】は、状態検出ステップに供される入力シグナルのプロットである。

【図 10】は、測定の間隔の得られたシリーズのプロットである。

【図 11】は、遷移マトリックスの図形表示である。

【図 12】は、シミュレーションされた例における k マー状態に関する予測される測定のグラフである。

【図 13】は、図 12 において図示される予測される測定からシミュレーションされた入力シグナルを示している。

【図 14】は、図 13 の入力シグナルから導き出される測定のシリーズを示している。

10

【図 15】は、遷移重み付けの遷移マトリックスを示している。

【図 16】は、遷移重み付けの遷移マトリックスを示している。

【図 17】は、ガウスである可能な分布を有する放出重み付けのグラフである。

【図 18】は、三角である可能な分布を有する放出重み付けのグラフである。

【図 19】は、四角である可能な分布を有する放出重み付けのグラフである。

【図 20】は、図 12 に示される 1 セットのシミュレーションされた測定と予測される測定間の電流スペースアライメントのグラフである。

【図 21】は、実際の k マーと図 20 のシミュレーションされた測定から推定される k マー間の k マースペースアライメントのグラフである。

【図 22】は、シミュレーションされた測定の追加のセットと図 12 に示される予測される測定間の電流スペースアライメントのグラフである。

20

【図 23】は、図 15 の遷移マトリックスを用いた、実際の k マーと図 22 のシミュレーションされた測定から推定される k マー間の k マースペースアライメントのグラフである。

【図 24】は、図 16 の遷移マトリックスを用いた、実際の k マーと図 22 のシミュレーションされた測定から推定される k マー間の k マースペースアライメントのグラフである。

【図 25】は、分布が図 12 の予測される測定を中心とする小非ゼロバックグラウンドのある四角分布を有する放出重み付けのグラフである。

【図 26】は、図 15 の遷移マトリックスおよび図 25 の放出重み付けを用いた、実際の k マーと図 20 のシミュレーションされた測定から推定される k マー間の k マースペースアライメントのグラフである。

30

【図 27】は、分布が図 12 の予測される測定を中心とするゼロバックグラウンドのある四角分布を有する放出重み付けのグラフである。

【図 28】は、図 15 の遷移マトリックスおよび図 27 の放出重み付けを用いた、実際の k マーと図 20 のシミュレーションされた測定から推定される k マー間の k マースペースアライメントのグラフである。

【図 29】は、ストレプトアビジンを使用する MS - (B2) 8 ナノポア中に保持される DNA 鎖から得られる電流測定の散布図である。

【図 30】は、実例トレーニングプロセスについての遷移マトリックスである。

40

【図 31】は、図 30 の遷移マトリックスの拡大部分である。

【図 32】は、静的トレーニングプロセスから導き出される 64 の k マーのモデルについての放出重み付けのグラフである。

【図 33】は、図 32 のモデルのおよそ 400 の状態のモデルへの変換についての放出重み付けのグラフである。

【図 34】は、トレーニングプロセスのフローチャートである。

【図 35】は、図 34 のトレーニングプロセスにより決定される放出重み付けのグラフである。

【図 36】は、モデルから予測される測定を用いていくつかの実験にわたって集められた電流測定のグラフである。

50

【図 37】は、実際の k マーと推定された k マー間の k マースペースアライメントのグラフである。

【図 38】は、実際の配列と整列された推定された k マーの推定された配列を示している。

【図 39】は、ポリマーのセンス領域およびアンチセンス領域の別々の推定された配列を、2つのそれぞれの次元で配置されたセンス領域およびアンチセンス領域からの測定を処理することにより導かれる推定された配列と一緒に示している。

【図 40】は、最初の例における 3 つの異なる電圧でのナノポア中の DNA 鎖のセットについてのイオン電流測定のプロットのセットである。

【図 41】は、第二の例におけるナノポア中の一本鎖についての共通の時間期間にわたり付加された電位および得られたイオン電流の対になったグラフである。

10

【図 42】は、第二の例において電圧レベル + 60 mV で水平方向に表示された DNA 鎖ごとの測定された電流の散布図である。

【図 43】は、第二の例において電圧レベル + 100 mV で水平方向に表示された DNA 鎖ごとの測定された電流の散布図である。

【図 44】は、第二の例において電圧レベル + 140 mV で水平方向に表示された DNA 鎖ごとの測定された電流の散布図である。

【図 45】は、第二の例において電圧レベル + 180 mV で水平方向に表示された DNA 鎖ごとの測定された電流の散布図である。

【図 46】は、第二の例において印加された電圧に対するそれぞれの DNA 鎖の測定された電流のプロットである。

20

【図 47】は、印加された電圧に対する第二の例における DNA 鎖ごとの電流測定の標準偏差のプロットである。

【図 48】は、イオン電流測定を行う方法のフローチャートである。

【図 49】は、第三の例における共通の時間期間にわたって付加された電位および得られたイオン電流の対になったグラフである。

【図 50】は、第三の例における共通の時間期間にわたって付加された電位および得られたイオン電流の対になったグラフである。

【図 51】は、イオン電流測定を行う別の方法のフローチャートである。

【図 52】A 及び B は、ナノポアを横断して印加された形状電圧ステップおよび得られた電流の同じ時間尺度上のプロットである。本発明のすべての態様は、以下の通りの広範なポリマーに適用しうる。

30

【発明を実施するための形態】

【0035】

ポリマーは、ポリヌクレオチド（または核酸）、タンパク質などのポリペプチド、多糖類、または他の任意のポリマーであってよい。ポリマーは天然でも合成でもよい。

【0036】

ポリヌクレオチドまたは核酸のケースでは、ポリマー単位はヌクレオチドであってよい。核酸は典型的にはデオキシリボ核酸（DNA）、リボ核酸（RNA）、cDNA またはペプチド核酸（PNA）、グリセロール核酸（GNA）、トレオース核酸（TNA）、ロックド核酸（LNA）もしくはヌクレオチド側鎖のある他の合成ポリマーなどの当技術分野で公知の合成核酸である。核酸は一本鎖でも二本鎖でもまたは一本鎖領域と二本鎖領域の両方を含んでもよい。典型的には、cDNA、RNA、GNA、TNA または LNA は一本鎖である。本発明の方法を使用して、いかなるヌクレオチドでも同定しうる。ヌクレオチドは天然に存在するもので人工的なものでもよい。ヌクレオチドは典型的には、核酸塩基、糖および少なくとも 1 つのリン酸基を含有する。核酸塩基は典型的には複素環式である。適切な核酸塩基には、プリンおよびピリミジンが、さらに具体的にはアデニン、グアニン、チミン、ウラシルおよびシトシンが含まれる。糖は典型的には五炭糖である。適切な糖には、リボースおよびデオキシリボースが含まれるがこれらに限定されない。ヌクレオチドは典型的にはリボヌクレオチドまたはデオキシリボヌクレオチドである。ヌ

40

50

クレオチドは典型的には、一リン酸塩、二リン酸塩または三リン酸塩を含有する。

【0037】

ヌクレオチドは損傷した塩基でも後成的塩基でも可能である。ヌクレオチドは標識されてまたは修飾されて、はっきりしたシグナルを有するマーカーとして働くことができる。この技法を使用して、ポリヌクレオチド内の塩基の非存在、例えば、塩基脱落単位またはスパーサーを同定することが可能である。前記方法はどんな種類のポリマーにも適用することができるであろう。

【0038】

修飾されたまたは損傷したDNAの測定（または類似のシステム）を検討する際に特に有用なのが、補完的データが考慮される方法である。提供される追加の情報は、さらに多数の根底にある状態の区別を可能にする。

10

【0039】

ポリペプチドのケースでは、ポリマー単位は天然に存在するまたは合成のアミノ酸であってよい。

【0040】

多糖類のケースでは、ポリマー単位は単糖であってよい。

【0041】

本発明は、下でさらに考察されるように、広範囲の測定システムにより行われる測定に適用しうる。

【0042】

20

本発明のすべての態様に従って、測定システムはナノポアを含むナノポアシステムでありうる。このケースでは、測定はナノポア中を通るポリマーの移行中に行いうる。ナノポア中を通るポリマーの移行は、観測されうる、全体では「事象」と呼びうる測定される特性における特徴的シグナルを生じる。

【0043】

ナノポアは、典型的にはナノメートルオーダーのサイズを有し、その中を通してポリマーを通過させるポアである。ポリマー単位がポア中を通して移行することに依拠している特性は測定しうる。前記特性は、ポリマーとポア間の相互作用に関連していることがある。ポリマーの相互作用はポアの狭窄領域で起こりうる。測定システムは前記特性を測定し、ポリマーのポリマー単位に依拠している測定を生み出す。

30

【0044】

ナノポアは生物学的ポアでも固体状態ポアでもよい。

【0045】

ナノポアが生物学的ポアである場合、以下の特性を有していることがある。

【0046】

生物学的ポアは膜貫通タンパク質ポアでありうる。本発明に従って使用するための膜貫通タンパク質ポアは、βバレルポアまたはαヘリックス束状ポアに由来することが可能である。βバレルポアは、β鎖から形成されるバレルまたはチャネルを含む。適切なβバレルポアには、α溶血素、炭疽毒素およびロイコシジンなどのβ毒素、ならびにマイコバクテリウム・スメグマチス (*Mycobacterium smegmatis*) ポリン (Msp)、例えば、MspA、外膜ポリンF (OmpF)、外膜ポリンG (OmpG)、外膜ホスホリパーゼAおよびナイセリア (*Neisseria*) 自己輸送体リポタンパク質 (NalP) などの細菌の外膜タンパク質／ポリン、が含まれるがこれらに限定されない。αヘリックス束状ポアは、αヘリックスから形成されるバレルまたはチャネルを含む。適切なαヘリックス束状ポアには、内膜タンパク質ならびにWZAおよびClyA毒素などのα外膜タンパク質が含まれるがこれらに限定されない。膜貫通ポアは、Mspにまたはα溶血素 (α-HL) に由来しうる。

40

【0047】

膜貫通タンパク質ポアは典型的にはMspに、好ましくはMspAに由来する。そのようなポアは、オリゴマーであり、典型的にはMsp由来の7、8、9または10モノマー

50

を含む。ポアは、同一のモノマーを含むM s p由来のホモオリゴマーポアであってもよい。代わりに、ポアは、その他のモノマーとは異なる少なくとも1つのモノマーを含むM s p由来のヘテロオリゴマーポアであってもよい。ポアは、M s p由来の2つ以上の共有結合モノマーを含む1つまたは複数の構築物を含んでいてもよい。適切なポアは、米国特許仮出願第61／441,718号(2011年2月11日提出)に開示されている。好ましくは、ポアはM s p Aまたはそのホモログもしくはパラログ由来である。

【0048】

生物学的ポアは天然に存在するポアでもよいし、変異ポアでもよい。典型的なポアは、WO-2010/109197、Stoddart D et al., Proc Natl Acad Sci, 12; 106(19):7702-7、Stoddart D et al., Angew Chem Int Ed Engl. 2010; 49(3):556-9、Stoddart D et al., Nano Lett. 2010 Sep 8; 10(9):3633-7、Butler TZ et al., Proc Natl Acad Sci 2008; 105(52):20647-52および米国特許仮出願第61／441718号に記載されている。

【0049】

生物学的ポアはMS-(B1)8でありうる。B1をコードするヌクレオチド配列およびB1のアミノ酸配列は下に示されている(配列番号1および配列番号2)。

配列番号1: MS-(B1)8=MS-(D90N/D91N/D93N/D118R/D134R/E139K)8

ATGGGTCTGGATAATGAACCTGAGCCTGGTGGACGGTCAAG
ATCGTACCCCTGACGGTGCAACAATGGGATACCTTTCTGAA
TGGCGTTTTTCCGCTGGATCGTAATCGCCTGACCCGTGAA
TGGTTTTCAATTCGGGTCGCGCAAAATATATCGTCGCGAGGCC
CGGGTGCTGACGAATTCGAAGGCACGCTGGAACCTGGGGTTA
TCAGATTGGCTTTCCGTGGTCACTGGGCGTTGGTATCAAC
TTCTCTGTACACCAACGCCGAATATTTCTGATCAACAATGGTA
ACATTACCGCACCGCGCTTTGGCCTGAACAGCGTGATTAC
GCCGAACCTGTTTTCCGGGTGTTAGCATCTCTGCCCGTCTG
GGCAATGGTCCGGGCAATTCAGAAGTGGAACCTTTAGTG
TGCGCGTTTTCCGGCGCTAAAGGCGGTGTCGCGGTGTCCTAA
CGCCCAACGGTACCGTTACGGGCGCGGGCCGGCGGTGTCCTG
CTGCGTCCGTTCGCGCGCCTGATTGCCCTCTACCGGCGACA
GCGTTACGACCTATGGCGAACCGTGGAATATGAACCTAA

配列番号2: MS-(B1)8=MS-(D90N/D91N/D93N/D118R/D134R/E139K)8

GLDNELSLV D G Q D R T L T V Q Q W D T F L N G V F P L D R N R L T R E W
F H S G R A K Y I V A G P G A D E F E G T L E L G Y Q I G F P W S L G V G I N F
S Y T T P N I L I N N G N I T A P P F G L N S V I T P N L F P G V S I S A R L G
N G P G I Q E V A T F S V R V S G A K G G V A V S N A H G T V T G A A G G V L L
R P F A R L I A S T G D S V T T Y G E P W N M N

【0050】

生物学的ポアはより好ましくはMS-(B2)8である。B2のアミノ酸配列は、変異L88Nを除いてはB1のアミノ酸配列と同じである。B2をコードするヌクレオチド配列およびB2のアミノ酸配列は下に示されている(配列番号3および配列番号4)。

配列番号3: MS-(B2)8=MS-(L88N/D90N/D91N/D93N/D118R/D134R/E139K)8

ATGGGTCTGGATAATGAACCTGAGCCTGGTGGACGGTCAAG
ATCGTACCCCTGACGGTGCAACAATGGGATACCTTTCTGAA
TGGCGTTTTTCCGCTGGATCGTAATCGCCTGACCCGTGAA
TGGTTTTCAATTCGGGTCGCGCAAAATATATCGTCGCGAGGCC
CGGGTGCTGACGAATTCGAAGGCACGCTGGAACCTGGGGTTA

TCAGATTGGCTTTCCGTGGTCACTGGGCGTTGGTATCAAC
TTCTCGTACACCAACGCCGAATATTAAACATCAACAATGGTA
ACATTACCGCACCGCCGTTTGGCCGTGAACAGCGTGATTAC
GCCGAACCTGTTTCCGGGTGTTAGCATCTCTGCCCGTCTG
GGCAATGGTCCGGGCATTCAAGAAAGTGGCAACCTTTAGTG
TGCGCGTTTCCGGCGCTAAAGGCGGTGTCGCGGTGTCTAA
CGCCCCACGGTACCGTTACGGGCGCGGCGCGGTGTCTCTG
CTGCGTCCGTTTCGCGCGCGCTGATTGCCCTCTACCGGCGACA
GCGTTACGACCTATGGCGAACCGGTGGAATATGAACATAA

配列番号4: MS-(B2)8=MS-(L88N/D90N/D91N/D93N/D
118R/D134R/E139K)8

10

GLDNE LSLVDGQDRTLTVQQWDTFLNGVFPLDRNRRLTREW
FHSGRACYIVAGPGADEFEGTLELGYQIGFPWSLGVGINF
SYTTPNININNGNITAPPFGLNSVITPNLFPGVSI SARLG
NGPGIQEVATFSVRVSGAKGGVAVSNAHGTVTGAAGGVLL
RPFARLIASTGDSVTTYGEPWNMN

【0051】

生物学的ポアは、生体膜、例えば、脂質二重層などの両親媒性層に挿入しうる。両親媒性層は、親水性も親油性も有するリン脂質などの両親媒性分子から形成される層である。両親媒性層は単層でも二重層でもよい。両親媒性層は、(Gonzalez-Perez et al., Langmuir, 2009, 25, 10447-10450)により開示されているなどのコブロックポリマーであってもよい。代わりに、生物学的ポアは固体状態層に挿入されてもよい。

20

【0052】

代わりに、ナノポアは、固体状態層内で形成される開口部を含む固体状態ポアであってもよい。

【0053】

固体状態層は生体起源ではない。言い換えると、固体状態層は生物もしくは細胞などの生物学的環境に由来してもそこから単離されてもおらず、または生物学的に利用可能な構造体の合成的に製造されたものでもない。固体状態層は、マイクロエレクトロニクス材料、Si₃N₄、Al₂O₃およびSiO₂などの絶縁材料、ポリアミドなどの有機および無機ポリマー、Teflon（登録商標）などのプラスチックまたは二成分付加硬化型シリコーンゴムなどのエラストマーならびにガラスを含むがこれらに限定されない有機材料からでも無機材料からでも形成することが可能である。固体状態層はグラフェンから形成しうる。適切なグラフェン層は、WO 2009/035647およびWO-2011/046706に開示されている。

30

【0054】

固体状態ポアは典型的には固体状態層中の開口部である。開口部は化学的にまたは他の方法で改変して、ナノポアとしてのその特性を増強してもよい。固体状態ポアは、トンネル電極(Ivanov AP et al., Nano Lett. 2011 Jan 12;11(1):279-85)または電界効果トランジスター(FET)デバイス(国際出願WO2005/124888)などのポリマーの別のまたは追加の測定を提供する追加の成分と組み合わせて使用してもよい。固体状態ポアは、例えば、WO 00/79257に記載されている工程を含む公知の工程により形成してもよい。

40

【0055】

測定システムの一つでは、ナノポア中を流れるイオン電流の測定を使用しうる。これらのおよび他の電氣的測定は、Stoddart D et al., Proc Natl Acad Sci, 12; 106(19):7702-7, Lieberman KR et al., J Am Chem Soc. 2010; 132(50):17961-72および国際出願WO-2000/28312に記載されている標準単一チャンネル記録装置を使用して行いうる。代わりに、電氣的測定は、例えば、国際出願WO-2009/077734および国際出願WO-2011/067559に記載されているマルチチャンネルシステムを使用して

50

行いうる。

【0056】

ポリマーがナノポア中を通して移行する時に測定を行うことを可能にするために、移行速度はポリマー結合部分により制御することができる。典型的には、前記部分は、印加電界に合わせてまたは逆らってナノポア中を通してポリマーを移動させることができる。前記部分は、例えば、前記部分が酵素のケースでは酵素活性を使用して分子モーターになる、または分子ブレーキとなることができる。ポリマーがポリヌクレオチドである場合、ポリヌクレオチド結合酵素の使用を含む移行速度を制御するためのいくつかの方法が提唱されている。ポリヌクレオチドの移行速度を制御するのに適した酵素には、ポリメラーゼ、ヘリカーゼ、エキソヌクレアーゼ、一本鎖および二本鎖結合タンパク質、ならびにジャイレースなどのトポイソメラーゼが含まれるがこれらに限定されない。他のポリマー型では、そのポリマー型と相互作用する部分を使用することができる。ポリマー相互作用部分は、国際出願番号 P C T / G B 1 0 / 0 0 0 1 3 3 または U S 6 1 / 4 4 1 7 1 8 (Lieberman KR et al, J Am Chem Soc. 2010;132(50):17961-72) に開示されているおよび電位開口型スキーム (Luan B et al., Phys Rev Lett. 2010; 104(23):238103) についてのいかなる部分でもよい。

10

【0057】

ポリマー結合部分は、ポリマー運動を制御するいくつかの方法でを使用することが可能である。前記部分は、印加電界に合わせてまたは逆らってナノポア中を通してポリマーを移動させることができる。前記部分は、例えば、前記部分が酵素のケースでは酵素活性を使用して分子モーターとして、または分子ブレーキとして使用することができる。ポリマーの移行は、ポア中を通るポリマーの動きを制御する分子歯止めにより制御しうる。分子歯止めはポリマー結合タンパク質であってよい。ポリヌクレオチドでは、ポリヌクレオチド結合タンパク質は好ましくはポリヌクレオチドハンドリング酵素である。ポリヌクレオチドハンドリング酵素は、ポリヌクレオチドと相互作用をし、ポリヌクレオチドの少なくとも1つの特性を改変することができるポリペプチドである。前記酵素は、ポリヌクレオチドを切断して個々のヌクレオチドまたはジヌクレオチドもしくはトリヌクレオチドなどのヌクレオチドの比較的短い鎖を形成することによりポリヌクレオチドを改変することができる。前記酵素は、ポリヌクレオチドを特定の位置に方向付けるまたは移動させることによりポリヌクレオチドを改変することができる。ポリヌクレオチドハンドリング酵素は、標的ポリヌクレオチドと結合しポア中を通るその動きを制御することができさえすれば酵素活性を示す必要はない。例えば、前記酵素はその酵素活性を取り除くように改変されてもよいし、酵素として作用するのを妨げる条件下で使用してもよい。そのような条件は下でさらに詳細に考察される。

20

30

【0058】

ポリヌクレオチドハンドリング酵素は、核酸分解酵素に由来していてもよい。酵素の構築物中で使用されるポリヌクレオチドハンドリング酵素は、より好ましくは、酵素分類 (E C) 群、3. 1. 11、3. 1. 13、3. 1. 14、3. 1. 15、3. 1. 16、3. 1. 21、3. 1. 22、3. 1. 25、3. 1. 26、3. 1. 27、3. 1. 30 および 3. 1. 31 のうちのいずれかのメンバー由来である。酵素は、国際出願番号 P C T / G B 1 0 / 0 0 0 1 3 3 (W O 2 0 1 0 / 0 8 6 6 0 3 として公開された) に開示されている酵素のうちのいずれであってもよい。

40

【0059】

好ましい酵素は、ポリメラーゼ、エキソヌクレアーゼ、ヘリカーゼおよびジャイレースなどのトポイソメラーゼである。適切な酵素には、大腸菌 (E.coli) 由来のエキソヌクレアーゼ I (配列番号 8)、大腸菌 (E.coli) 由来のエキソヌクレアーゼ I I I 酵素 (配列番号 10)、サーマス・サーモフィルス (T. thermophilus) 由来の R e c J (配列番号 12) およびバクテリオファージラムダエキソヌクレアーゼ (配列番号 14) ならびにそのバリエーションが含まれるがこれらに限定されない。配列番号 14 に示される配列またはそのバリエーションを含む3つのサブユニットは相互作用してトリマーエキソヌクレアーゼを形

50

成する。前記酵素は好ましくはP h i 2 9 DNAポリメラーゼ由来である。P h i 2 9ポリメラーゼ由来の酵素は、配列番号6に示される配列またはそのバリエーションを含む。

【0060】

配列番号6、8、10、12または14のバリエーションは、配列番号6、8、10、12または14のアミノ酸配列とは異なるアミノ酸配列を有しポリヌクレオチド結合力は保持している酵素である。前記バリエーションは、ポリヌクレオチドの結合を促進しならびに／または高塩濃度および／もしくは室温でその活性を促進する改変を含みうる。

【0061】

配列番号6、8、10、12または14のアミノ酸配列の全長にわたり、バリエーションは好ましくは、アミノ酸同一性に基づいてその配列に少なくとも50%相同であることになる。さらに好ましくは、前記バリエーションポリペプチドは、アミノ酸同一性に基づいて、配列番号6、8、10、12または14のアミノ酸配列にその全配列にわたり少なくとも55%、少なくとも60%、少なくとも65%、少なくとも70%、少なくとも75%、少なくとも80%、少なくとも85%、少なくとも90%およびさらに好ましくは少なくとも95%、97%または99%相同であってよい。200以上の、例えば、230、250、270または280以上の連続するアミノ酸のストレッチにわたって少なくとも80%、例えば、少なくとも85%、90%または95%アミノ酸同一性（「ハード相同性」）が存在していてもよい。相同性は上に記載される通りに決定される。バリエーションは、配列番号2を参照して上で考察されるいかなる点においても野生型配列とは異なりうる。酵素は上で考察されたポアに共有結合していてもよい。

【0062】

一本鎖DNA塩基配列決定のための2つの戦略は、シスからトランスへとトランスからシスへの両方、付加電位に合わせてまたは逆らってのどちらかでナノポア中を通るDNAの移行である。鎖塩基配列決定のための最も有利な機構は、付加電位下でのナノポア中を通る一本鎖DNAの制御された移行である。二本鎖DNA上で前進的にまたは進行的に作用するエキソヌクレアーゼは、付加電位下で残りの一本鎖を中に送り込むためにポアのシス側上で、または逆電位下ではトランス側でを使用することができる。同様に、二本鎖DNAをほどくヘリカーゼも類似する様式でを使用することが可能である。付加電位に逆らった鎖移行を必要とする塩基配列決定適用の可能性もあるが、DNAはまず逆電位下でまたは電位なしで酵素により「捕捉」されなければならない。次に結合に続いて電位が切り替えられると、前記鎖はポア中をシスからトランスへ通過し、電流の流れにより伸ばされた立体構造に保持されることになる。一本鎖DNAエキソヌクレアーゼまたは一本鎖DNA依存性ポリメラーゼは分子モーターとして働いて、付加電位に逆らってトランスからシスに、制御された段階的な様式でポア中で移行したばかりの一本鎖を引き戻すことができる。代わりに、一本鎖DNA依存性ポリメラーゼは、ポア中を通るポリヌクレオチドの動きを遅くする分子ブレーキとして働くことができる。仮出願US 61/441718またはUS仮出願第61/402903号に記載されているいかなる部分、技法または酵素を使用してもポリマーの動きを制御することができるであろう。

【0063】

しかし、別の種類の測定システムおよび測定も可能である。

【0064】

別の種類の測定システムのいくつかの非限定的例は以下の通りである。

【0065】

測定システムは、走査型プローブ顕微鏡であってよい。走査型プローブ顕微鏡は、原子間力顕微鏡（AFM）、走査型トンネル顕微鏡（STM）または別の形態の走査型顕微鏡である。

【0066】

読取り装置がAFMであるケースでは、AFMチップの解像度は、個々のポリマー単位の寸法ほど微細ではないことがある。したがって、測定は複数のポリマー単位の関数になりうる。AFMチップは、チップが機能化されていない場合とは別の様式でポリマー単位

10

20

30

40

50

と相互作用するように機能化させうる。A F Mは、接触モード、非接触モード、タッピングモードまたは他のどんなモードでも作動させうる。

【0067】

読取り装置がS T Mであるケースでは、測定の解像度は個々のポリマー単位の寸法ほど微細ではないことがあり、そのため測定は複数のポリマー単位の関数になる。S T Mは、従来法でまたは分光学的測定（S T S）をするようにまたは他のどんなモードでも作動させうる。

【0068】

別の種類の測定のいくつかの例には、電氣的測定および光学的測定が限定なく含まれる。蛍光の測定を伴う適切な光学的方法は、J. Am. Chem. Soc. 2009、131 1652-1653により開示されている。可能な電氣的測定には、電流測定、インピーダンス測定、トンネリング測定（例えば、Ivanov AP et al., Nano Lett. 2011 Jan 12;11(1):279-85に開示されている）およびF E T測定（例えば、国際出願WO 2005/124888に開示されている）が含まれる。光学的測定は電氣的測定と組み合わせうる（Soni GV et al., Rev Sci Instrum. 2010 Jan;81(1):014301）。測定は、ナノポア中を通るイオン電流の流れの測定などの膜貫通電流測定でもよい。イオン電流は典型的には、直流イオン電流でよいが、原理的には代案は交流電流の流れ（すなわち、交流電圧の印加下で流れる交流電流の大きさ）を使用することである。

【0069】

本明細書では、用語「k マー」とは、k マーが単一ポリマー単位であるk が1であるケースを含む、k が正整数であるk 個のポリマー単位のグループのことである。いくつかの文脈では、一般にはk が1であるケースを除外してk マーのサブセットである、k が複数の整数であるk マーに言及する。

【0070】

理想的には測定は単一のポリマー単位に依拠すると考えられるが、多くの典型的な測定システムに関しては、測定は、k が複数の整数であるポリマーのk マーに依拠する。すなわち、それぞれの測定は、k が複数の整数であるk マー中のそれぞれのポリマー単位の配列に依拠する。典型的には、測定はポリマーと測定システム間の相互作用に関連する特性である。

【0071】

本発明のいくつかの実施形態では、小グループのポリマー単位、例えば、ダブレットまたはトリプレットのポリマー単位（すなわち、k = 2 または k = 3）に依拠している測定を使用するのが好ましい。他の実施形態では、さらに大きなグループのポリマー単位に依拠している測定を、すなわち、「広い」解像度で使用するのが好ましい。そのような広い解像度は、ホモポリマー領域を調べるのに特に有用になることがある。

【0072】

特に測定が、k が複数の整数であるk マーに依拠している場合、できる限り多くの可能なk マーについて分解可能である（すなわち、分離される）ことが望ましい。典型的には、これは異なるk マーにより生み出される測定が測定範囲にわたり十分に広がっているおよび／または狭い分布を有するならば、達成することが可能である。これは、異なる測定システムにより様々な程度に達成しうる。しかし、異なるk マーにより生み出される測定が分解可能であることが不可欠ではないことは本発明の特別な利点である。

【0073】

図1は、両親媒性層などの生体膜2に挿入された生物学的ポア1であるナノポアを含む測定システム8の例を模式的に示している。ポリマー単位のシリーズ4を含むポリマー3は、矢印で示されるように生物学的ポア1中を通して移行される。ポリマー3は、ポリマー単位4がヌクレオチドであるポリヌクレオチドでありうる。ポリマー3は、生物学的ポア1内部のk マーに依拠して変化する膜貫通電流などの電気特性を引き起こす生物学的ポア1の活性部分5と相互作用する。この例では、活性部分5は3つのポリマー単位4のk マーと相互作用しているものとして示されているが、これは限定的なものではない。

【0074】

生体膜2のそれぞれの側に配置された電極6は、制御回路71および測定回路72を含めて、電気回路7と接続されている。

【0075】

制御回路71は、生物学的ポア1を横断する印加のために電極6に電圧を供給するように配置されている。

【0076】

測定回路72は、電気特性を測定するように配置されている。したがって、測定は生物学的ポア1内部のkマーに依拠している。

【0077】

測定システムにより出力される典型的タイプのシグナルであり、本発明に従って解析されることになる入力シグナルであるシグナルは「ノイズステップ波」であるが、このシグナルタイプに限定されない。この形態を有する入力シグナルの例は、ナノポアを含む測定システムを使用して得られるイオン電流測定の場合について図2に示されている。

【0078】

このタイプの入力シグナルは、連続するグループの複数の測定が同じkマーに依拠している測定の入力シリーズを含む。それぞれのグループ内の複数の測定は、下で考察されるある分散を受ける一定値であり、したがって、測定システムの状態に対応してシグナルに「レベル」を形成する。シグナルは、大きなセットであることもあるレベルのセット間を移動する。計測手段のサンプリング速度およびシグナル上のノイズを考慮すると、レベル間の遷移は瞬間的だと見なすことが可能であり、したがって、シグナルは理想化されたステップトレースにより概算することが可能である。

【0079】

それぞれの状態に対応した測定は事象の時間尺度にわたり一定であるが、大半の測定システムでは短い時間尺度にわたり分散を受けることになる。分散は、例えば、電気回路およびシグナル処理から、特に電気生理学という特定のケースでは増幅器から生じる測定ノイズに起因することがある。そのような測定ノイズは、測定されている特性が小規模であるため避けられない。分散は、測定システムの根底にある物理的または生物学的系における固有の変動または拡散からも生じることがある。大半の測定システムは、そのような固有の変動を大小の程度の差はあっても経験することになる。所与の測定システムでは、両方の変動源が一因となることがあるまたはこれらのノイズ源のうちの1つが優勢であることもある。

【0080】

さらに、典型的にはグループにおける測定数に先験的知識はなく、この数は予測不能に変化する。

【0081】

分散および測定数についての知識の欠如というこれら2つの要因のため、例えば、グループが短いおよび／または2つの連続するグループの測定のレベルが互いに近い場合、グループのいくつかを区別するのが困難になることがある。

【0082】

シグナルは、測定システムにおいて起きている物理的または生物学的過程の結果として、このような形態をとる。したがって、測定のそれぞれのグループは「状態」と呼ぶうる。

【0083】

例えば、ナノポアを含むいくつかの測定システムでは、ナノポア中を通るポリマーの移行からなる事象は、一方向だけに動く様式で起こりうる。一方向だけの動きのそれぞれのステップ中、ナノポアを横断する所与の電圧でのナノポア中を流れるイオン電流は一定であり、上で考察された分散を受ける。したがって、測定のそれぞれのグループは、一方向だけの動きのステップに関連している。それぞれのステップは、ポリマーがナノポアに対してそれぞれの位置にある状態に対応する。状態の期間中の正確な位置にはある程度の変

10

20

30

40

50

動がありうるが、状態間にはポリマーの大規模な動きがある。測定システムの性質に応じて、状態はナノポア中での結合事象の結果として起こりうる。

【0084】

個々の状態の持続時間は、ポアを横断して印加される電位、ポリマーの歯止めをするのに使用される酵素の種類、ポリマーが酵素によりポア中を通して押されているのかまたは引っ張られているのか、pH、塩濃度および存在するヌクレオシド三リン酸の種類などのいくつかの要因に依拠していることがある。状態の持続時間は、測定システムに応じて0.5msから3秒まで変化することがあり、所与のナノポアシステムでは、状態間である程度の無作為な変動がある。持続時間の予測される分布は、所与の測定システムについて実験的に決定しうる。

10

【0085】

前記方法は、測定の複数の入力シリーズであって、それぞれがそれぞれのシリーズにおける複数の測定の連続するグループが同じkマーに依拠している上記の形態をとる入力シリーズを使用することがある。そのような複数のシリーズは登録されることがあるので、例えば、それぞれシリーズの測定が同時に行われる場合、それぞれのシリーズからのどの測定が対応し同じkマーに依拠しているのかが先験的に分かる。例えば、測定が、異なる測定システムにより同調して測定された異なる特性である場合、こうなる可能性がある。代わりに、そのような複数のシリーズは登録されないことがあるので、それぞれのシリーズからのどの測定が対応し同じkマーに依拠しているのかは先験的には分からない。例えば、測定のシリーズが異なる時間に行われる場合、こうなる可能性がある。

20

【0086】

ナノポアを横断する異なるレベルの電圧の印加下で測定が行われる、下で考察される第三の態様に従った方法は、それぞれのレベルの電圧に関する測定のシリーズを提供する。このケースでは、測定の繰返し周期は、問題の測定システムについての状態の繰返し周期を考慮して選択される。理想的には、繰返し周期はすべての状態の持続時間より短く、これは、測定システムについての最小の予想される繰返し周期よりも短い繰返し周期を選択することにより達成される。しかし、いくつかの状態のみの持続時間よりも短い、例えば、状態の持続時間の平均で60%、70%、80%、90%、95%、または99%よりも短い繰返し周期中に行われる測定から有用な情報を得ることができる。典型的には、繰返し周期は最長で3秒、さらに典型的には最長で2秒または最長で1秒であってよい。典型的には、繰返し周期は少なくとも0.5ms、さらに典型的には少なくとも1msまたは少なくとも2msであってよい。

30

【0087】

1つよりも多い、例えば、2から10までの数の電圧周期を状態の持続時間に適用してもよい。

【0088】

それぞれのkマーに関して複数の測定を1つの電圧レベルで（または、複数の電圧レベルのそれぞれで複数の測定を）行ってもよい。1つの可能なアプローチでは、異なるレベルの電圧をそれぞれ、例えば、電圧波形がステップ波である時に、一期間連続して印加してもよく、前記複数の期間のそれぞれの期間中に、その期間に印加される電圧のうちの1つで1グループの複数の測定が行われる。

40

【0089】

複数の測定は、それ自体それにく続く解析において使用しうる。代わりに、その（またはそれぞれの）電圧レベルでの1つまたは複数のサマリー測定を、複数の測定のそれぞれのグループから導き出しうる。前記1つまたは複数のサマリー測定は、所与のkマーに関して所与の電圧レベルでの複数の測定から、いかなる様式でも、例えば、平均もしくは中央値として、または統計的変動の尺度、例えば標準偏差として導き出しうる。次に、前記1つまたは複数のサマリー測定は、それにく続く解析において使用しうる。

【0090】

電圧周期はいくつかの異なる波形から選択しうる。波形は、非対称的、対照的、規則的

50

または不規則的でもよい。

【0091】

周期の一例では、異なるレベルの電圧をそれぞれ、それらの異なるレベル間の遷移と共に、一期間、すなわち、周期、例えば方形波またはステップ波の部分的期間連続して印加してもよい。電圧レベル間の遷移は、急なこともあれば、一期間にわたり傾斜していることもある。

【0092】

周期の別の例では、電圧レベルは連続して変化し、例えば、異なるレベル間、例えば、三角波またはのこぎり波で傾斜していてもよい。このケースでは、異なるレベルでの測定は、所望の電圧レベルに対応する周期内で時々測定を行うことにより行ってもよい。

10

【0093】

情報は、電圧平坦域での測定からまたは勾配の測定から導き出さう。さらに情報は、異なる電圧レベルで行われる測定に加えて、例えば、1つの電圧レベルと別の電圧レベル間の過渡電流の形状の測定により導き出さう。

【0094】

段階的電圧スキームでは、電圧レベル間の遷移は、どんな容量過渡電流をも最小にするように形作ることができる。ナノポアシステムを単純なRC回路と見なせば、電流の流れ、 I 、は式 $I = V / R + C \, dV / dt$ により与えられ、 V は付加電位、 R は抵抗（典型的にはポアの）、 t は時間および C は容量（典型的には二重層の）である。このモデル系では2つの電圧レベル間の遷移は、時定数、 $\tau = RC$ の指数関数（ $V = V_2 - (V_2 - V_1) * \exp(-t / \tau)$ ）に従うと考えられる。

20

【0095】

図52aおよび52bは、電圧レベル間の遷移の時定数 τ が、遷移速度が最適化されている、非常に速いおよび非常に遅くなるように選択されるケースを図示している。電圧遷移が非常に速い場合、測定される電流シグナルにスパイク（オーバーシュート）が見られ、非常に遅いと測定されるシグナルは急には平らにならない（アンダーシュート）。遷移速度が最適化されているケースでは、測定される電流が理想的な急な遷移から歪んでいる時間は最小化される。遷移の時定数 τ は、測定システムの電気的特性の測定から、または異なる遷移の試験から決定さう。

【0096】

30

2つ以上のうちのいかなる数のレベルの電圧でも測定は行いうる。電圧のレベルは、それぞれのレベルの電圧での測定が、前記測定が依拠している k マーの正体についての情報を提供するように選択される。したがって、レベルの選択は測定システムの性質に依拠している。ナノポアを横断して印加される電位差の程度は、両親媒性層の安定性、使用される酵素の種類および所望の移行速度などの要因に依拠することになる。典型的には、電圧のレベルのそれぞれは同じ極性になるが、一般には電圧のレベルのうちの1つまたは複数はそれ以外のレベルとは反対の極性であることも可能であろう。一般に、大半のナノポアシステムでは、それぞれのレベルの電圧は典型的には対地10mVから2Vであってもよい。したがって、電圧レベル間の電圧差は典型的には少なくとも10mV、さらに好ましくは少なくとも20mVであってよい。電圧レベル間の電圧差は典型的には最大で1.5V、さらに典型的には最大で400mVであってよい。電圧差が大きくなると電圧レベル間の電流差はさらに大きくなり、したがってそれぞれの状態間に潜在的にさらに大きな区別を生じる傾向がある。しかし、高電圧レベルは例えば、システムにより多くのノイズを生じるまたは酵素による移行の混乱を招くことがある。逆に、電圧差が小さくなると電流差は小さくなる傾向がある。最適電位差は、実験条件または酵素菌止めの種類に応じて選択さう。

40

【0097】

1つの電圧レベルで測定される k マーは、必ずしも異なる電圧レベルで測定されるのと同じ k マーではないことがある。 k の値は、異なる電位で測定される k マー間で異なることがある。しかし、これが本当であるならば、異なる電圧レベルで測定されるそれぞれの

50

k マーに共通であるポリマー単位が存在する可能性が高い。理論に縛られることなく、測定されている k マーのいかなる差も、ナノポアを横断して印加されるより高い電位差でのナノポア内のポリマーの立体構造の変化がリーダーヘッドにより測定されているポリマー単位の数の変化をもたらすことに起因している可能性があると考えられる。この立体構造の変化の程度は、1 つの値と別の値間の電位の差に依拠している可能性がある。

【0098】

測定の一部としてまたは登録情報を提供する追加の供給源からのいずれかで他の情報が入手可能になることがある。この他の情報により状態を同定できることがある。

【0099】

代わりに、シグナルが恣意的な形態をとることがある。これらのケースでは、k マーに 10
対応する測定は放出と遷移のセットの点からも記載されることがある。例えば、特定の k マーに依拠している測定は、これらの方法による記載を受け入れることができる形で起こる測定のシリーズを含むことがある。

【0100】

所与の測定システムが k マーおよび k マーのサイズに依拠している測定を提供する程度は実験的に調べることができる。例えば、公知のポリマーが合成され、測定システムに対して予め定められた位置に保持されて、得られた測定から、前記測定が測定システムと相互作用をする k マーの正体にどのように依拠しているのかを調べることができる。

【0101】

1 つの可能なアプローチは、同一配列を有するポリマーのセットを、そのセットのポリマーごとに 20
変化する予め定められた位置の k マーを除いて使用することである。前記 k マーのサイズおよび正体を変化させて、測定に対するその効果を調べることができる。

【0102】

別の可能なアプローチは、予め定められた位置で調査中の k マーの外側のポリマー単位がそのセットのポリマーごとに 30
変化するポリマーのセットを使用することである。そのようなアプローチの例として、図 3 は、ナノポアを含む測定システムにおける 2 つのポリヌクレオチドの電流測定の頻度分布である。前記ポリヌクレオチドのうちの 1 つ (poly T と呼ばれる) では、ナノポアの領域におけるすべての塩基が T (poly T と呼ばれる) であり、ポリヌクレオチドのもう一方 (N11-TATGAT-N8 と呼ばれる) では、特定の決められた 6 マー (配列 TATGAT を有する) から左側に 11 塩基および右側に 8 塩基を変化させる。図 3 の例は電流測定の点で前記 2 つの鎖の見事な分離を示している。N11-TATGAT-N8 鎖により見られる値の範囲も、poly T により見られる値の範囲よりもごくわずかに広い。このようにしておよび他の配列を有するポリマーも測定して、問題の特定の測定システムでは、測定が 6 マーに良好な近似で依拠していることを確かめることが可能である。

【0103】

このアプローチまたは類似のアプローチは、位置および最小 k マー記述を決定することを可能にするどんな測定システムについても一般化することが可能である。

【0104】

異なる条件下でまたは異なる検出方法によって複数の測定を適用する確率論的枠組み、 40
特に技法により、ポリマーのより低い k 記述を使用することが可能になることがある。例えば、下で考察されるセンス DNA およびアンチセンス DNA 測定のケースでは、それぞれの k マー測定のより正確な記述が 6 マーになると考えられる場合、3 マー記述で根底にあるポリマー k マーを決定するのに十分でありうる。同様に、複数の電位での測定のケースでは、それぞれの k マー測定のより正確な記述が k マーまたは k が比較的高い値を有する k マーになると考えられる場合、k が比較的低い値を有する k マー記述で根底にあるポリマー k マーを決定するのに十分でありうる。

【0105】

類似の方法を使用して、一般的測定システムにおける十分近似する k マーの位置および 50
幅を同定してもよい。図 3 の例では、これは、(例えば、N の数を前と後ろで変化させる

ことにより) ポアに対して6マーの位置を変えて、最もよく近似するkマーの位置を検出し、6から決められた塩基の数を増加するおよび減少することにより達成される。kの値は、十分に狭い値の拡散の影響を最小限受けることがある。kマーの位置は、ピーク幅を最小限にするように選択することができる。

【0106】

典型的な測定システムでは、異なるkマーに依拠している測定はすべてが独自に分解可能であるわけではないことは通常事実である。例えば、図3が関係する測定システムでは、決められた6マーを有するDNA鎖により生み出される測定の範囲は2 pAの桁であり、このシステムの近似の測定範囲は30 pAから70 pAであることが観測される。6マーでは、4096通りの可能なkマーが存在する。これらのそれぞれが2 pAの類似する変動を有することを考慮すると、40 pA測定範囲においてはこれらのシグナルは独自に分解可能ではないことは明らかである。いくつかのkマーの測定が分解可能である場合でさえ、多くの他のkマーの測定が分解可能ではないことが典型的に観測される。

10

【0107】

多くの実際の測定システムでは、それぞれが同じポリマー単位に一部依拠しているk個の測定を変換して、ポリマー単位のレベルで分解される単一の値を得る関数を同定することは可能ではなく、またはさらに一般的にはkマー測定がkマーの数よりも小さなパラメータのセットにより記述可能ではない。

【0108】

例として、ナノポアを含む特定の測定システムでは、ポリヌクレオチドの実験的に導かれたイオン電流測定は簡単な一次線形モデルにより正確に記述可能ではないことがここで実証されるであろう。これは、下でより詳細に説明される2つのトレーニングセットについて明らかにされる。この実証のために使用される簡単な一次線形モデルは、

20

$$\text{電流} = \text{Sum} [f_n (B_n)] + E$$

であり、 f_n は測定システムにおけるそれぞれの位置nで生じる塩基 B_n ごとの係数であり、Eは実験変動性に起因するランダム誤差を表す。データは最小二乗法によりこのモデルにフィットさせるが、当技術分野で公知の多くの方法のいずれか1つを代わりに使用することもできる。図4および5は、電流測定に対してフィットする最良モデルのプロットである。データがこのモデルで十分に記述されるのであれば、点は典型的な実験誤差(例えば、2 pA)内で対角線にきちんと従うはずである。これは、データがどちらのセットの係数でもこの線形モデルでは十分に記述されないことを示すケースではない。

30

【0109】

ノイズが多いステップ波である入力シグナルを解析する方法であって、本発明の第一の態様を具体化する特定の方法がここで説明されることになる。以下の方法は、測定が、kが2以上のkマーに依拠しているケースに関係するが、同じ方法はkが1であるkマーに依拠している測定に単純化された形態で適用しうる。

【0110】

前記方法は図6に図示されており、図6に模式的に図示されている解析ユニット10において実施しうる。解析ユニット10は測定回路72からの測定を含む入力シグナルを受けて解析する。したがって、解析ユニット10と測定システム8は接続され、合わせてポリマーを解析するための装置を構成する。解析ユニット10は、制御回路7に制御シグナルも提供して測定システム8において生物学的ポア1を横断して印加される電圧を選択し、印加された電圧に従って測定回路72からの測定を解析することができる。

40

【0111】

解析ユニット10と測定システム8を含む装置は、WO-2008/102210、WO-2009/07734、WO-2010/122293および/またはWO-2011/067559のいずれかに開示されている通りに配置しうる。

【0112】

解析ユニット10は、コンピュータ装置において実行されるコンピュータプログラムにより実施してもよいし、または専用のハードウェアデバイスまたはその任意の組合せによ

50

り実施してもよい。どちらのケースでも、前記方法により使用されるデータは解析ユニット10のメモリに記憶される。コンピュータ装置は、使用される場合、いかなる種類のコンピュータシステムでもよいが、典型的には従来の構造である。コンピュータプログラムは、いかなる適切なプログラム言語で書かれていてもよい。コンピュータプログラムは、コンピュータ可読記憶媒体上に記憶されることができ、前記媒体は、どんな種類でも、例えば、計算システムのドライブに差し込み可能であり、情報を磁氣的に、光学的にまたは光磁氣的に記憶しうる記憶媒体、ハードドライブなどのコンピュータシステムの固定記憶媒体、またはコンピュータメモリでもよい。

【0113】

前記方法は、どのグループでも測定の数についての先験的な知識なしで同じkマーに依拠している複数の測定の連続するグループを含む、上記の種類の測定のシリーズを（または、さらに下で説明されるように、より一般的にはどんな数のシリーズでも）を含む入力シグナル11で実施される。そのような入力シグナル11の例は、以前記載された図2に示されている。

10

【0114】

状態検出ステップS1では、入力シグナル11は処理されて、連続するグループの測定を同定し、それぞれの同定されたグループに関して予め定められた数（1または複数である）の測定からなる測定のシリーズ12を導き出す。解析ステップS2は、このようにして導き出された測定のシリーズ12で実施される。状態検出ステップS1の目的は、それぞれのkマー状態に関連する予め定められた数の測定まで入力シグナルを減らして、解析ステップS2を簡略化することである。例えば、図2に示されるように、ノイズの多いステップ波シグナルは、それぞれの状態に関連する単一の測定が平均電流でありうる状態まで減らしうる。この状態はレベルと呼ばれることもある。

20

【0115】

状態検出ステップS1は、以下の通りに入力シグナル11の導関数において短期増加を探す図7に示されている方法を使用して実施しうる。

【0116】

ステップS1-1では、入力シグナル11は微分されてその導関数を導き出す。

【0117】

ステップS1-2では、ステップS1-1からの導関数は、（微分が増幅させる傾向がある）高周波ノイズを抑制するための低域フィルタリングにかけられる。

30

【0118】

ステップS1-3では、ステップS1-2からフィルターにかけられた導関数は閾値処理され、測定のグループ間の遷移点を検出し、それによりデータのグループを同定する。

【0119】

ステップS1-4では、予め定められた数の測定は、ステップS1-3において同定されたそれぞれのグループにおける入力シグナル11から導き出される。最も簡単なアプローチでは、単一の測定は、それぞれの同定されたグループにおける測定値の、例えば平均、中央値または他の位置の尺度として導き出される。ステップS1-4からの測定出力は測定のシリーズ12を形成する。他のアプローチでは、それぞれのグループに関する複数の測定が導き出される。

40

【0120】

この技法の一般的簡略化は、データの2つの隣接する窓の平均を比較するスライディングウィンドウ解析を使用することである。次に、閾値は直接的に平均差に置くこともできるし、または前記2つの窓におけるデータ点の分散に基づいて設定することができる（例えば、ステューデントt統計量を計算することにより）。これらの方法の特定の利点は、これらの方法がデータに多くの仮定を負わせることなく適用することができることである。

【0121】

測定されたレベルに関連している他の情報は、解析において後で使用するために記憶さ

50

せることができる。そのような情報は、シグナルの分散、非対称情報、観察の信頼度、グループの長さのいずれでも制限なく含んでいてもよい。

【0122】

例として、図9は、移動するウィンドウ検定により減らされた実験的に決定された入力シグナル11を図示している。特に、図9は入力シグナル11を細線で示している。状態検出に続くレベルは暗線としてかぶせて示されている。図10は、遷移間の平均値からのそれぞれの状態のレベルを計算して、全トレースについて導き出された測定シリーズ12を示している。

【0123】

しかし、下でさらに詳細に説明されるように、状態検出ステップS1は任意選択であり、さらに下で説明される代案では、省いてもよい。このケースでは、図6中の点線により模式的に示されるように、解析ステップS2は、測定シリーズ12の代わりに、入力シグナル11自体で実施される。

【0124】

解析ステップS2はここで説明される。

【0125】

解析ステップS2は、解析ユニット10に記憶されたモデル13を参照する解析技法を使用する。解析ステップS2は、測定シリーズ12がポリマー単位の配列により生み出されるというモデル13により予測される尤度に基づいてポリマー中のポリマー単位の推定された配列16を推定する。最も単純なケースでは、推定された配列16は、ポリマー単位ごとに単一の推定された正体を提供する表示であることがある。さらに一般的には、推定された配列16は、ある最適性基準に従ってポリマー単位の配列の任意の表示であることがある。例えば、推定された配列16は、例えば、ポリマーの一部またはすべてにおける1つまたは複数のポリマー単位の複数の推定された正体を含めて、複数の配列を含むことがある。

【0126】

モデル13の数学的基礎はここで検討される。解析ステップS2は、さらに下で説明されるクオリティスコア17も提供する。

【0127】

電流をサンプリングする確率変数の配列 $\{X_1, X_2, \dots, X_n\}$ 間の関係は、変数

$$X_1 - X_2 - X_3 - \dots - X_n$$

間の条件付き独立関係を表す単純なグラフィックモデルAにより表される。

【0128】

それぞれの電流測定は読み取られているkマーに依拠しているので、確率変数の根底にあるセット $\{S_1, S_2, \dots, S_n\}$ がkマーの根底にある配列を表しており、対応するグラフィックモデルBでは、

【0129】

【化1】

$$X_1 X_2 X_3 \dots X_n$$

$$\begin{array}{ccccccc} | & | & | & & | & & \\ S_1 & - & S_2 & - & S_3 & - & \dots - S_n \end{array}$$

である。

【0130】

適用の電流領域に適用されるこれらのモデルはマルコフ性を利用する。モデルAでは、 $f(X_i)$ を確率変数 X_i の確率密度関数を表すようにとれば、マルコフ性は

$$f(X_m | X_{m-1}) = f(X_m | X_1, X_2, \dots, X_{m-1})$$

として表すことができる。

【0131】

モデルBでは、マルコフ性は

$$P(S_m | S_{m-1}) = P(S_m | S_1, S_2, \dots, S_{m-1})$$

として表すことができる。

【0132】

問題が正確にどのようにコード化されているかに応じて、解決のための自然の方法は、ベイジアンネットワーク、マルコフ確率場、隠れマルコフモデルを含むことがあり、これらのモデルの異形、例えば、そのようなモデルの条件付きまたは最大エントロピーフォーマクションも含む。これらのわずかに異なる枠組み内の解決法は多くの場合類似している。一般的に、モデル13は起点kマーから目的地kマーまでの遷移の可能性を表す遷移重み付け14、およびそれぞれのkマーに関してそのkマーについて測定の所与の値を観測する可能性を表す放出重み付け15を含んでいる。モデル13が隠れマルコフモデルであるケースではここで説明される。

10

【0133】

隠れマルコフモデル(HMM)は、グラフィックモデルBにおいてここで与えられる設定の自然表示である。HMMでは、個々の確率変数 S_m と S_{m+1} の間の関係は、このケースではそれぞれの確率変数を取ることで可能な状態間、すなわち、起点kマーから目的地kマーまでの遷移の確率を表す確率である遷移重み付け14の遷移マトリックスの点から定義されている。例えば、従来から、遷移マトリックスの (i, j) 番目エントリーは、 $S_m = S_{m,i}$ だと仮定すると $S_{m+1} = S_{m+1,j}$ である確率、すなわち、 S_m がその i 番目の可能な値を帯びると仮定すると S_{m+1} の j 番目の可能な値への遷移の確率を表す遷移重み付け14である。

20

【0134】

図11は、 S_m から S_{m+1} までの遷移マトリックスの図形表示である。ここでは、 S_m および S_{m+1} は、説明のために4つの値を示しているだけであるが、実際には異なるkマーが存在するのと同じ数の状態が存在すると考えられる。それぞれのエッジは遷移を表し、遷移確率を表す遷移マトリックスからのエントリーで標識してもよい。図11では、 S_m 層から S_{m+1} 層においてそれぞれの交点を接続する4つのエッジの遷移確率は古典的には合計で1になると考えられるが、非確率的重み付けを使用してもよい。

【0135】

一般に、遷移重み付け14は、非バイナリー変数(非バイナリー値)の値を含むことが望ましい。これにより、モデル13はkマー間の遷移の実際の確率を表すことが可能になる。

30

【0136】

モデル13がkマーを表すことを考慮すると、所与のkマーは、起点kマーから、最初の $(k-1)$ 個のポリマー単位が起点kマーの最後の $(k-1)$ 個のポリマー単位である配列を有する目的地kマーまでの遷移であるk個の好ましい遷移を有する。例えば、4種のヌクレオチドG、T、AおよびCからなるポリヌクレオチドのケースでは、起点3マーTACは3マーACA、ACC、ACTおよびACGへの好ましい遷移を有する。最初の近似まで、概念的には、4つの好ましい遷移の遷移確率は (0.25) に等しく、それ以外の好ましくない遷移の遷移確率はゼロであり、好ましくない遷移は、起点kマーから、起点kマーとは異なる配列を有し、最初の $(k-1)$ 個のポリマー単位が起点kマーの最後の $(k-1)$ 個のポリマー単位ではない目的地kマーまでの遷移であると考えてもよい。しかし、この近似は理解のためには有用であるが、遷移の実際の可能性は一般に、所与の測定システムではこの近似から変わることがある。これは、非バイナリー変数(非バイナリー値)の値をとる遷移重み付け14により反映されることがある。表されることがあるそのような変動のいくつかの例は以下の通りである。

40

【0137】

一例は、好ましい遷移の遷移確率は等しくなくてもよいことである。これによりモデル13は、ポリマーの配列間に相互関係性があるポリマーを表すことが可能になる。

【0138】

50

一例は、好ましくない遷移の少なくとも一部の遷移確率はゼロでなくてもよいことである。これにより、モデル13は見逃された測定、すなわち、実際のポリマー中のkマーのうちの1つ（または複数）に依拠している測定はないことを考慮することが可能になる。そのような見逃された測定は、測定システムの問題によって起こり、測定は物理的には行われない、または、例えば、所与のグループが短すぎるもしくは2つのグループが十分分離したレベルではないという理由で、測定値のグループのうちの1つを同定することができない状態検出ステップS1などのその後のデータ解析の問題によって起こることもある。

【0139】

遷移重み付け14が任意の値を持つことを可能にする一般性にもかかわらず、典型的には、遷移重み付け14が、起点kマーから、最初の $(k-1)$ 個のポリマー単位が起点kマーの最後の $(k-1)$ 個のポリマー単位である配列を有する目的地kマーまでの好ましい遷移の非ゼロの可能性を表し、好ましくない遷移のさらに低い可能性を表すことが事実となるであろう。典型的には、遷移重み付け14は、前記好ましくない遷移の少なくとも一部の非ゼロの可能性も表すが、その可能性はゼロに近いことがあり、または完全に除外される遷移の一部ではゼロであることもある。

【0140】

配列中の単一の見逃されたkマーを可能にするため、遷移重み付け14は、起点kマーから、最初の $(k-2)$ 個のポリマー単位が起点kマーの最後の $(k-2)$ 個のポリマー単位である配列を有する目的地kマーまでの好ましくない遷移の非ゼロの可能性を表すことがある。例えば、4種のヌクレオチドからなるポリヌクレオチドのケースで、起点3マーTACでは、これらはCから開始するすべての可能な3マーへの遷移である。これらの単一の見逃されたkマーに対応する遷移を「スキップ (skips)」と定義することができる。

【0141】

それぞれのkマーに関して単一測定を含む測定のシリーズ12を解析するケースでは、次に遷移重み付け14は、測定12ごとに遷移の高い可能性を表すことになる。測定の性質に応じて、起点kマーから起点kマーと同じである目的地kマーまでの遷移の可能性はゼロもしくはゼロに近いこともあれば、好ましくない遷移の可能性に類似することもある。

【0142】

同様に、それぞれのkマーに関して予め定められた数の測定を含む測定のシリーズ12を解析するケースでは、次に遷移重み付け14は、同じkマーに関する測定12間の遷移の低いまたはゼロの可能性を表すことがある。遷移重み付け14を変化させて起点kマーと目的地kマーが同じkマーになるようにすることは可能である。これは、例えば、誤検出された状態遷移を可能にする。これら繰り返される同じkマーに対応する遷移を「ステイ (stays)」と定義することができる。kマー中のポリマー単位すべてが同一である、すなわちホモポリマーであるケースでは、好ましい遷移はステイ遷移になると考えられることに我々は注目している。これらのケースでは、ポリマーは1位置移動しているがkマーは同じままである。

【0143】

同様に、それぞれのkマーに関して典型的に複数の測定がある測定のシリーズ12を解析するケースであるが未知の量であるケース（「スティッキング (sticking)」と呼ばれることがある）では、遷移重み付け14は、起点kマーと目的地kマーが同じkマーになる比較的高い確率を表すことがあり、物理系に応じて、遷移重み付け14は、いくつかのケースでは、上記の好ましい遷移が、起点kマーから最初の $(k-1)$ 個のポリマー単位が起点kマーの最後の $(k-1)$ 個のポリマー単位と同じである目的地kマーまでの遷移である確率よりも大きくなることがある。

【0144】

さらに、状態検出ステップS1を使用せずに、入力シグナル11を解析するケースでは

10

20

30

40

50

、次にこれは、単に、起点 k マーと目的地 k マーが同じ k マーになる比較的高い確率を表すように遷移重み付け 14 を適合させることにより達成しうる。これにより、根本的に、同じ解析ステップ S 2 を実施することが可能になり、モデル 13 の適合は状態検出を暗黙の裡に考慮に入れている。

【0145】

それぞれの k マーと関連して、その k マーについての測定の所与の値を観測する確率を表す放出重み付け 15 が存在する。したがって、図 11 における交点 $S_{m, i}$ により表される k マー状態では、放出重み付け 15 は、電流測定がサンプリングされる分布を記述する確率密度関数 $g(X_m | S_{m, i})$ として表されうる。放出重み付け 15 が、非バイナリー変数の値を含むことが望ましい。これにより、モデル 13 は、一般には簡単なバイナリー形をもたないことがある異なる電流測定の確率を表すことが可能になる。

10

【0146】

状態検出ステップ S 1 が、それぞれの同定されたグループに関して複数の測定からなる測定のシリーズ 12 を導き出すケースでは（例えば、平均および変動）、放出重み付け 15 は、その k マーについてのそれぞれの種類の測定の所与の値を観測する確率を表す。同様に、登録されているので、それぞれのシリーズ由来のどの測定が対応しており同じ k マーに依拠しているのかが先験的に分かっている測定の複数のシリーズ 12 で前記方法が実施されるさらに一般的なケースでは、放出重み付け 15 はその k マーについてのそれぞれのシリーズの測定の所与の値を観測する確率を再び表す。これらのケースでは、モデル 13 は、それぞれの k マー状態についての複数の測定の分布を記述する複数の次元における確率密度関数として放出重み付け 15 を使用して適用することができる。一般に、所与の k マーについての放出重み付け 15 は、測定の確率を反映するいかなる形態でも取りうる。異なる k マーは、単一モデル 13 内での同じ放出分布形態またはパラメータ付け (parameterisation) で放出重み付け 15 を有する必要はない。

20

【0147】

多くの測定システムでは、 k マーの測定は、測定される物理的または生物学的特性の拡散によりおよび／または測定誤差によりのいずれかで分散することがある特定の予測される値を有する。これは、適切な分布、例えば、単峰形 (unimodal) である分布を有する放出重み付け 15 を使用することによりモデル 13 にモデル化することができる。

【0148】

しかし、いくつかの測定システムでは、所与の k マーについての放出重み付け 15 は、多峰形 (multimodal) である、例えば、測定システムにおける 2 つの異なる種類の結合からおよび／または k マーが測定システム内で複数の立体構造をとることから物理的に生じることがある。

30

【0149】

有利なことに、放出重み付け 15 は、可能なすべての測定を観測する非ゼロの可能性を表すことがある。これにより、モデル 13 は、外れ値である、所与の k マーが生み出す思いがけない測定を考慮に入れることが可能になる。例えば、放出重み付け 15 確率密度関数を、非ゼロ確率を有する外れ値を可能にする広いサポート上で選択しうる。例えば、単峰形分布のケースでは、 k マーごとの放出重み付けは、すべての実数に対して非ゼロの重み付けを有するガウス分布またはラプラス分布を有することがある。

40

【0150】

放出重み付け 15 が恣意的に定義される分布であることを可能にして、外れ値測定の手際の良い取扱いおよび多価放出を有する単一状態のケースを取り扱うことが可能になることは有利でありうる。

【0151】

放出重み付け 15 を経験的に、例えば、下に説明されるトレーニング段階中に決定するのが望ましいことがある。

【0152】

放出重み付け 15 の分布は、測定空間を横断する任意の適切な数のビン (bin) を用い

50

て表すことができる。例えば、下記のケースでは、分布は、データ範囲全体で500個のビンにより定義される。外れ値測定は、すべてのビンにおいて非ゼロ確率（外れたビンでは低い）を、データが定義されたビンのうちの1つに収まらない場合は類似の確率を有することにより取り扱うことが可能である。十分な数のビンを定義すれば望ましい分布を近似することができる。

【0153】

したがって、前記好ましくない遷移のうちの少なくとも一部の非ゼロ可能性を表す遷移重み付け14の使用および／またはあらゆる可能な測定を観測する非ゼロ可能性を表す放出重み付け15の使用から特定の利点を導き出さう。特定の利点は、所与のkマーについての広い範囲の測定を観測する相対的な可能性に対応する放出重み付けの使用からも導き出さう。

10

【0154】

これらの利点を強調するために、配列を導き出すための簡単な非確率的方法は比較例と見なされる。この比較例では、観測される値の所与の範囲外の測定を生み出すkマーは無効とされ、見逃された測定（スキップ）に対応する遷移は、例えば、エッジおよび交点を削除することにより図11における遷移の数を減らして、無効とされる。比較例では、次に、 S_i ごとに正確に1つの交点を含み、ポリマー単位の根底にある配列に対応するkマー状態の独自の接続された配列が探索される。しかし、この比較例は恣意的な閾値に頼って無効とされた交点とエッジを同定するので、適切なエッジがグラフ内の存在しないため読み飛ばされた測定のケースではどんな経路も見つけることができない。同様に、外れた測定のケースでは、比較例は、図11において削除された対応する交点を生じることになり、再びグラフ中を通る適切な経路を確認するのは不可能になる。

20

【0155】

これとは対照的に、モデル13および解析ステップS2における確率的または重み付けの方法などの解析技法の使用の特定の利点は、このブレイクダウンのケースを回避することができる点である。別の利点は、複数の可能な経路が存在するケースでは、最も可能性の高い経路、または可能性のある経路のセットを決定することができる点である。

【0156】

この方法の別の特定の利点は、ホモポリマー、すなわち同一ポリマー単位の配列の検出に関係する。モデルをベースとする解析により、シグナルに寄与するポリマー単位の数に類似する長さまでホモポリマー領域の取扱いが可能になる。例えば、6マー測定は6ポリマー単位長までホモポリマー領域を同定することができるであろう。

30

【0157】

解析ステップS2の1つの可能な形態は図8に示されており、以下の通りに作動する。

【0158】

ステップS2-1では、kマーの推定された配列18は、測定のシリーズ12がkマーの配列により生み出されるというモデル13により予測される尤度に基づいてモデル13を参照して推定される。

【0159】

ステップS2-2では、ポリマー単位の推定された配列16は、ステップS2-1において推定されたkマーの推定された配列18から推定される。

40

【0160】

ステップS2-1でもS2-2でも、さらに下で考察されるように、それぞれkマーの推定された配列18とポリマー単位の推定された配列16の質を表すクオリティスコアも提供される。

【0161】

解析ステップS2において適用される解析技法は、モデル13に適している種々の形態をとり、測定のシリーズ12がポリマー単位の配列により生み出されるというモデル13により予測される尤度に基づいてポリマー中のポリマー単位の推定された配列16を提供しう。例えば、モデルがHMMであるケースでは、解析技法は、ステップS2-1にお

50

いて、いかなる公知のアルゴリズムでも、例えば、フォワードバックワードアルゴリズムまたはビタビ (Viterbi) アルゴリズムを使用してもよい。そのようなアルゴリズムは、一般に、状態の配列を通じたあらゆる可能な経路の尤度を力任せに風潰しに計算するのを回避し、代わりに、尤度に基づいて簡略化された方法を使用して状態配列を同定する。

【0162】

一代案では、ステップS2-1は、測定のシリーズが個々のkマーにより生み出されるというモデルにより予測される尤度に基づいて、配列の個々のkマーまたは配列中のkマーごとの複数のkマー推定値を推定することによりkマーの配列18を同定しうる。例として、解析技法がステップS2-1においてフォワードバックワードアルゴリズムを使用する場合、解析技法は、測定のシリーズが個々のkマーにより生み出されるというモデルにより予測される尤度に基づいて、kマーの配列18を推定する。フォワードバックワードアルゴリズムは当技術分野では周知である。フォワード部分では、所与のkマーで終わる全配列の全体の尤度は遷移および放出重み付けを使用して最初から最後の測定まで再帰的に前向きに計算される。バックワード部分は、類似する形態でしかし最後の測定からずっと最初まで働く。これらのフォワードおよびバックワード確率は組み合わせられ、データの全体の尤度と併せてそれぞれの測定が所与のkマー由来である確率を計算する。

10

【0163】

フォワード-バックワード確率から、配列18におけるそれぞれのkマーの推定値が導き出される。これは、それぞれ個々のkマーと関連する尤度に基づいている。1つの簡単なアプローチは、フォワード-バックワード確率がそれぞれの測定でのkマーの相対的尤度を示しているので、それぞれの測定で最も可能性の高いkマーを取ることである。

20

【0164】

ステップS2-1では、測定のシリーズ12が個々のkマーを含む配列により生み出されるというモデル13により予測される尤度を表すクオリティスコアも配列18中の個々のkマーに関して導き出される。これはステップS2-1において実施される解析から得られ、追加の有用な情報を提供する。

【0165】

もう1つの代案では、ステップS2-1は、測定のシリーズがkマーの全体配列により生み出されるというモデルにより予測される尤度に基づいて、全体配列または複数の全体配列を推定することによりkマーの配列18を同定しうる。別の例として、解析技法がステップS2-1においてビタビアルゴリズムを使用する場合、解析技法は、測定のシリーズがkマーの全体配列により生み出されるというモデルにより予測される尤度に基づいてkマーの配列18を推定する。ビタビアルゴリズムは当技術分野では周知である。

30

【0166】

ステップS2-1では、測定のシリーズ12がkマーの全体配列により生み出されるというモデル13により予測される尤度を表すクオリティスコアも配列18中の個々のkマーに関して導き出される。これはステップS2-1において実施される解析から得られ、追加の有用な情報を提供する。

【0167】

別の代案として、ステップS2-1は、測定のシリーズがkマーの全体配列により生み出されるというモデルにより予測される尤度に基づいてkマーの全体配列を同定する第1段階および第1段階の結果から前記配列の個々のkマーまたは前記配列中のkマーごとの複数のkマー推定値を推定することによりkマーの配列18を同定する第2段階を含む、2つの段階に分けてもよい。例として、この代案は、力任せの風潰し計算を使用しうる。

40

【0168】

ステップS2-2では、ポリマー単位の推定された配列16は、任意の適切な技法を使用してステップS2-1において推定されたkマーの推定された配列18から推定される。1つの単刀直入なアプローチは、kマーをポリマー単位と1対1の関係に関連付け、関連するkマーから単一のポリマー単位を取り出すだけである。さらに複雑なアプローチは、それぞれの所与のポリマー単位を含有する配列18中の推定されたkマーのグループか

50

らの情報の組合せを使用してそれぞれのポリマー単位を推定する。例えば、ポリマー単位はその推定されたkマーのうち最も確実なものから取り出してもよい。それぞれのポリマー単位は、ステップS 2-1における推定されたkマー配列に関して導き出されたクオリティスコア17を利用して推定することができる。

【0169】

ステップS 2-2では、測定シリーズ12がポリマー単位を含む配列により生み出されるというモデル13により予測される尤度を表すクオリティスコアも配列16中のそれぞれのポリマー単位に関して導き出される。これは、例えば、それぞれのkマーおよび関連するポリマー単位の相対的確率に基づいて、ステップS 2-2において実施される解析から得られ、追加の有用な情報を提供する。

10

【0170】

解析ステップS 2における上記技法は限定的ではない。確率的または他の解析技法を使用するモデルを利用する多くの方法が存在する。kマーの全体配列、個々のkマーまたは根底にあるポリマー単位を推定するプロセスは、特定の適用に適合させることが可能である。いかなる「硬い」kマー配列、kマーまたはポリマー単位コールも作る必要はない。あらゆるkマー配列、または可能性の高いkマー配列のサブセットを考慮することができる。kマー配列と関連するkマーもしくはkマーのセットまたは特定のkマー配列とは関係がないと見なされるkマーもしくはkマーのセット、例えば、すべてのkマー配列にわたる加重和を考慮することができる。kマーと関連するポリマー単位もしくはポリマー単位のセットまたは特定のkマーとは関係がないと見なされるポリマー単位もしくはポリマー単位のセット、例えば、すべてのkマーにわたる加重和、kマー配列もしくはkマー配列のセットに依拠しているkマーまたはkマー配列もしくはkマー配列のセットとは無関係なkマーを考慮することができる。

20

【0171】

例として、3マーポリヌクレオチド系を考慮しうる。可能性の高い塩基推定値のセットを導き出すいくつかの方法がある。第1の代案は、最も可能性の高い経路を考慮し（ビタビアルゴリズム）、その経路と関連する3マー状態のセットを導き出し、kマー由来の1つの塩基、例えば、中央の塩基を塩基コールとして使用することである。第2の代案は、すべての経路を考慮してそれぞれの地点で最も可能性の高いkマーを導き出すことである（フォワードバックワードアルゴリズム）。次に、最も可能性の高いkマー由来の1つの塩基（例えば、中央の塩基）を塩基推定値とすることもできるだろう。kマーから塩基推定値を導き出す別の方法は、塩基のうちの1つ（例えば、中央の塩基）の寄与を考慮し最も可能性の高い塩基を推定値として取りすべてのkマーにわたり合計することになると考えられる。kマーから塩基推定値を導き出す別の方法は、すべてのkマーにおけるすべての位置からの寄与を合計してそれぞれの位置で最も可能性の高い推定値を決定することになると考えられる。

30

【0172】

同様に、解析ステップS 2は、kマーの複数の配列18および／またはポリマー単位の複数の配列16を推定しうる。このケースでは、kマーの複数の配列18のそれぞれおよび／またはポリマー単位の複数の配列16のそれぞれに関してクオリティスコアを導き出しうる。このようにして、解析ステップS 2は、可能性がさらに低い配列に関する情報で、にもかかわらず一部の適用において有用であることもある情報を提供する。

40

【0173】

上記説明は、遷移重み付け14および放出重み付け15が確率であり、解析ステップS 2がモデル13を参照する確率的技法を使用するHMMであるモデル13に関して与えられている。しかし、遷移重み付け14および／または放出重み付け15が確率ではなくある他の方法で遷移または測定の可能性を表す枠組みをモデル13が使用することは代案としては可能である。このケースでは、解析ステップS 2は、測定シリーズがポリマー単位の配列により生み出されるというモデル13により予測される尤度に基づいている確率的技法以外の解析技法を使用しうる。解析ステップS 2により使用される解析技法は、尤

50

度関数を明確に使用しうるが、一般にはこれは不可欠ではない。したがって、本発明の文脈では、用語「尤度」は、計算または正式な尤度関数の使用を必要とせずに、測定シリーズがポリマー単位の配列により生み出される可能性を考慮に入れるという一般的意味で使用される。

【0174】

例えば、遷移重み付け14および／または放出重み付け15は、遷移または放出の可能性を表すが、確率ではなく、したがって、例えば、合計が1になるように制約されていないコスト（または距離）により表しうる。このケースでは、解析ステップS2は解析を、例えば、オペレーションズリサーチにおいて広く見られるように、最小コスト経路または最小経路問題として取り扱う解析技法を使用しうる。ダイクストラアルゴリズムなどの標準法（または他のさらに効率的なアルゴリズム）を解決のために使用することができる。

10

【0175】

モデル13が、鈍いリーダーヘッドシステムからのデータをモデル化し解析するのに使用されるHMMである特定の例がここで考察されることになる。ここでは、入力データ11は先ず既に記載された状態検出ステップS1により処理される。簡単にするために、しかし制限なく、この特定の例は、4種の可能な塩基を有し、そのため64の可能なkマーが存在するポリヌクレオチドについての3マーモデルに関係している。根底にあるモデル13および状態を参照して要点を説明することができるシミュレーションされたケースが提示される。

20

【0176】

このシミュレーションされたケースでは、3マー電流レベルが無作為に選択されるので、64のkマー状態の放出重み付け15を最も簡単に記述するのにも64の係数が必要である。測定からkマーの根底にある配列を決定するのは、記載された通りにモデルベースの解析により達成される。

【0177】

図12は、kマーごとに、測定の最も可能性の高い値を示している。したがって、これらの値は、それぞれのkマーの放出重み付け15の分布の中心値でもある。図12では、kマー状態インデックスは、G、T、A、Cの順に順次続き、すなわち、状態0＝「GGG」、状態1＝「GGT」、...、状態62＝「CCA」、状態63＝「CCC」である。kマー状態インデックスは解析中に使用され、最終ステップとして「塩基スペース（base space）」に再び変換される。

30

【0178】

所与の配列からの測定は、既に記載されている係数を使用してシミュレーションされる。例えば、配列ACTGTCAGは3マー、ACT、CTG、TGT、GTC、TCA、CAGで構成されている。これらは状態インデックス45、52、17、7、30、56に対応し、このインデックスは68.5、46.5、94.9、51.3、19.5、52.1の予測される測定を与える。シミュレーションされた測定は、入力シグナル12として図13に、状態検出ステップS1により生み出される測定のシリーズ12として図14に図示されている。

40

【0179】

実際には、行われたどんな測定も、その測定に付随するエラーがある。シミュレーションのケースでは、予測された測定値にノイズを加えることによりこれは考慮される。

【0180】

測定を見落とすまたは偽陽性の測定を挿入する可能性もある。これらは、ここで説明されることになる遷移マトリックスにおいて説明することができる。

【0181】

シミュレーションのケースについての遷移重み付け14の遷移マトリックスは、ここで考慮されることになる。

【0182】

50

測定のシリーズ1 2および放出重み付け1 5のセットを考慮して、解析ステップS 2は根底にある配列の推定値を決定する。概念的には、これは、観測された配列が比較される対象のすべての可能な遷移をモデル化する解析ステップS 2と見なしてもよい（が、実際、解析ステップS 2はこれを必要としないもっと効率的なアルゴリズムを使用することもある）。例えば、考慮中の3マーのケースでは、6 4の状態のそれぞれが他の4つの状態への好ましい遷移を有する。

【0 1 8 3】

図1 5は、好ましい遷移に対する遷移重み付け1 4がそれぞれ0. 2 5であり、好ましくない遷移に対する遷移重み付け1 4がそれぞれゼロであるシミュレーションされたモデルについての遷移重み付け1 4の遷移マトリックスを図示している。例えば、起点状態0（G G G）は、等しい確率で状態、0（G G G）、1（G G T）、2（G G A）または3（G G C）に遷移することが可能であると見ることができる。

10

【0 1 8 4】

図1 6は、見逃された測定を表す、すなわち、遷移が読み飛ばされている好ましくない遷移に対する非ゼロの遷移重み付け1 4を可能にすることにより、図1 5のシミュレーションされたモデルから修正されたシミュレーションされたモデルについての遷移重み付け1 4の遷移マトリックスのより複雑なケースを図示している。一般論として、遷移マトリックスは、根底にある測定システムをモデル化するのに必要であるように、恣意的に複雑になることがある。

20

【0 1 8 5】

測定のシリーズ1 2で作動するケースでは、我々が状態検出S 1を実施した場合、所与の起点k マーから離れる遷移確率は典型的に高く、要するに1に近づく。図1 5の第1の例では、遷移マトリックスは、好ましい「遷移」のうちの1つが同じk マーへの遷移である4つのホモポリマーのケースを除けば、遷移を必要とする。任意の状態からの4つの好ましい遷移のそれぞれの確率は0. 2 5である。このマトリックスは、他の適切な緩和が行われなければ、「実世界」のデータを取り扱うこと、例えば、放出重み付け1 5における外れ値取扱い、ができる可能性はない。

【0 1 8 6】

しかし、非ゼロの遷移は、それを扱う必要があるまたは起こる可能性があるいかなるケースにも可能にすることができる。図1 6の第2の例では、好ましい遷移の確率は0. 2 5未満であり、残りはステイおよびスキップ確率からなる。恣意的なレベルの複雑さまで、複数のスキップも類似する形態で許される。

30

【0 1 8 7】

遷移確率は、k マー間の遷移を測定することができる平易さを考慮に入れるように調整することが可能である。例えば、互いに非常に近い2つの連続するk マーからのシグナルのケースでは、状態検出ステップS 1がこの遷移を見逃すことは可能である。このケースでは、これら2つのk マー間の遷移マトリックスエレメントは、第2のk マーを読み飛ばす方向へ重み付けをしてもよい。

【0 1 8 8】

マトリックスは、所与の試料中の任意の配列バイアスを考慮に入れるように調整しうる。

40

【0 1 8 9】

上記の例では、放出および遷移重み付けは一定の値で固定されているが、これは不可欠ではない。代案として、放出重み付けおよび／または遷移重み付けは、おそらくプロセスについての追加の情報に導かれて、解析される測定値シリーズの異なる部分について変化させてもよい。例として、「ステイ」としての解釈を有する遷移重み付けのマトリックスのエレメントを、特定の事象（）がポリマーの実際の遷移を反映する信頼度に応じて調整することができるだろう。さらなる例として、放出重み付けを、測定デバイスのバックグラウンドノイズにおける系統的ドリフトまたは印加された電圧に加えられる変化を反映するように調整することができるだろう。重み付けに対する調整の範囲はこれらの例に限定さ

50

れない。

【0190】

上記の例では、それぞれのkマーが単回表示されているが、これは不可欠ではない。代案として、モデルはkマーの一部またはすべての複数回別個に表示されてもよく、したがって、所与のkマーに関して、遷移および／または放出重み付けのセットが複数あってもよい。ここでの遷移重み付けは別個の起点kマーと別個の目的地kマー間であってよく、したがって、それぞれの起点－目的地対は、それぞれのkマーの別個の表示の数に応じて、複数の重み付けがあってもよい。これら別個の表示の多くの可能な解釈の1つは、kマーが、直接観測することが可能ではないシステムのある挙動、例えば、ナノポア中を通る移行中にポリマーが取る可能性がある異なる立体構造または移行挙動の異なる動態を示すラベルをタグ付けされていることである。

10

【0191】

状態検出ステップS1を実施せずに生の入力シグナル11で作動するモデル13では、複数の測定のグループが、グループにおける測定の数についての先験的な知識なしで同じkマーに依拠している測定の入力シリーズに前記方法は直接適用される。このケースでは、非常に類似する技法であるが、所与の起点kマー状態から離れる遷移確率の合計がここでは1よりはるかに少ないという点で著しく調整した技法を、モデル13に適用することができる。例えば、平均してシステムが同じkマーで100の測定を費やすとすれば、遷移マトリックスにおける対角線上の確率（遷移を表さないまたは起点kマーと目的地kマーが同じkマーである遷移を表す）は0.99になり、0.01のスプリットがすべてのその他の好ましい遷移と好ましくない遷移の間にあることになる。好ましい遷移のセットは、状態検出ケースについての遷移に類似していることがある。

20

【0192】

放出重み付け15を考慮して、図17から19は、それぞれ、ガウス、三角および四角分布であるシミュレーションされた係数についての放出分布を示すが、どんな恣意的な分布（非パラメータ分布を含む）もこの様式で定義することができる。

【0193】

ノイズに対するこれらの方法の堅牢さを実証するため、ノイズ摂動がシミュレーションされた測定に加算される。この例では、標準偏差5 pAのガウス分布からサンプリングされたランダムノイズが図12に示される予測されるkマー測定に加算される。

30

【0194】

図20は、図12に示される予測される測定と比べたシミュレーションされた測定（測定のシリーズ12）を示しており、見ることができる加算されたノイズが厳密であることを図示している。

【0195】

遷移重み付けの適切な遷移マトリックス、例えば、図16に示されるマトリックスおよび放出重み付け15についての適切な分布、このケースでは、ガウス分布を用いて、モデル13が適用される。フォワードバックワードアルゴリズムが解析技法として使用されて、測定のシリーズにおけるそれぞれの点で最も可能性の高いkマーを推定する。推定されたkマーコールは、図21に示されているように、既知のkマー配列に対して比較される。この厳密なケースでも、大多数の状態が正しく推定されていることが分かる。

40

【0196】

配列中のkマーに関連する失われた測定に対する堅牢さはここで説明される。このケースでは、ノイズを予測されるkマー測定に加算することに加えて（この例では、1 pA標準偏差を有するノイズの厳密ではないケースを使用する）、このケースでは、0.1の削除の確率で、kマー測定がデータからも無作為に削除される測定のシリーズ12がシミュレーションされる。図22は、図12に示されている予測された測定と比べたシミュレーションされた測定（測定のシリーズ12）を示している。図22では、円で囲まれた失われたkマー状態を見ることができる。

【0197】

50

再び、遷移重み付けの適切な遷移マトリックス、このケースでは図 1 5 および 1 6 に示される両方のマトリックスおよび放出重み付け 1 5 についての適切な分布、このケースでは、ガウス分布を用いて、予測された k マー測定モデル 1 3 が適用される。フォワードバックワードアルゴリズムが解析技法として使用されて、測定のシリーズ 1 2 におけるそれぞれの点で最も可能性の高い k マーを推定する。

【0 1 9 8】

推定された k マーコールは、図 1 5 および 1 6 の遷移マトリックスについてそれぞれ図 2 3 および 2 4 に示されているように、既知の k マー配列に対して比較される。ここでは、図 2 3 と比べた場合、図 2 4 では、モデル遷移におけるスキップを可能にすることにより正確にコールされた k マーの数に改善がみられる。失われた k マー測定値が高信頼度の推定値に取り囲まれているケースでは、失われた k マーは周囲の k マーから推定することが可能である。これとは対照的に、スキップが許されないケースでは、失われたデータは、解析が k マーのシリーズの中を通る経路を見つけるためにゼロには達しない分布を有する放出重み付け 1 5 により収容される。放出分布における非ゼロのバックグラウンドは次のセクションでさらに考察される。

10

【0 1 9 9】

配列中の所与の k マーと関連がある外れた測定に対する堅牢さはここで説明される。遷移重み付け 1 4 が読み飛ばされた状態（すなわち、図 1 5 の遷移マトリックスを用いて）を許さない失われた測定に関する以前の説明では、解析が k マーの配列中を通る経路（非常に可能性が低い経路にもかかわらず）を見つけることができるようにゼロに達しない分布を有する放出重み付け 1 5 を使用する必要があった。すべての測定値について非ゼロの値を有する放出重み付け 1 5 の利点は、四角放出分布の単純なケースにおいて例証される。この例は、標準偏差 5 p A を有するノイズが加算される図 2 0 に示される測定のシミュレーションされたシリーズ 1 2 を使用する。

20

【0 2 0 0】

再び、図 1 5 に示されるように、好ましくない遷移が許されない遷移重み付け 1 4 の遷移マトリックスを用いて、および放出重み付け 1 5 についての 2 つの異なる分布を用いて、予測された k マー測定モデル 1 3 がこのケースでは適用される。フォワードバックワードアルゴリズムが解析技法として使用されて、測定のシリーズ 1 2 におけるそれぞれの点で最も可能性の高い k マーを推定する。

30

【0 2 0 1】

第 1 のケースでは、放出重み付け 1 5 は、図 2 5 に示される小非ゼロバックグラウンド（このケースでは 1×10^{-10} ）の四角分布を有し、これについては図 2 6 において、推定された k マーコールは既知の k マー配列に対して比較される。

【0 2 0 2】

第 2 のケースでは、放出重み付け 1 5 は図 2 7 に示されるゼロバックグラウンドの四角分布を有し、これについては図 2 8 において、推定された k マーコールが既知の k マー配列に対して比較される。

【0 2 0 3】

放出重み付け 1 5 の分布においてゼロバックグラウンドの第 2 のケースでは、それらの分布の幅が狭すぎる放出分布では k マー配列を通る経路は存在しない。この例では、図 2 7 において示されるように、解析が測定の中を通る経路を見つけることができるように幅 $+/- 14$ p A の放出分布を使用してきた。このケースでは、それぞれが大きな数の正確な状態を有する少数の経路が存在するよりは、多くの不正確にコールされた状態を含有する多数の経路が存在する。この例についての k マーコールのセットは図 2 8 に示されている。

40

【0 2 0 4】

図 2 5 に示されるように、バックグラウンドにおいて小非ゼロ放出が許される第 1 のケースでは、はるかに狭い分布を許容することが可能であり、図 2 8 よりも良好な結果を与える図 2 7 に示されるように、さらに大きな数の k マー状態を正確に推定することが可能に

50

なる。

【0205】

さらに、この例は、図27および28に示される四角分布の使用よりも良好な結果を提供する図20および21に示される例について使用されるガウス放出と四角分布ケースを比較することにより確率的方法の利点を示している。

【0206】

モデル13のトレーニング、すなわち、所与の測定システムについての放出重み付け15の誘導がここで考察されることになる。

【0207】

上記シミュレーションとは対照的に、実際の測定システムでは、それぞれのkマーからの個々の測定は前もって分かっているわけではないが、トレーニングセットから導き出すことができる。一般論として、これは既知のポリマーから測定を行い、それ自体がHMMにとって従来法であるトレーニング技法を使用することを含む。

10

【0208】

これらのトレーニング法では、特定種類の配列、すなわち、所与のkについてすべてのkマーを含有する最小長配列であるド・ブラン (deBruijn) 配列を利用する。ド・ブラン配列を使用するのは必要とされる実験数を最小限に抑えるための効率的な方法である。

【0209】

ポリヌクレオチドを測定するのに使用されるナノポアを含む測定システムについて2つのトレーニング法が記載されている。第1の方法は、ビオチン/ストレプトアビジン系によってナノポア内の特定の位置に保持された「静的」DNA鎖からの測定を使用する。第2の方法は、ナノポア中を通して移行されるDNA鎖からの測定を使用し、kマー推定のために記載された枠組みに類似する確率的枠組みを利用することにより係数を推定するまたは「訓練する」。

20

【0210】

第1の静的トレーニング法は以下の通りに実施される。

【0211】

これらの実験は、Stoddart D et al., Proc Natl Acad Sci, 12;106(19):7702-7に記載されている方法に類似するやり方でビオチン分子を使用してDNA鎖をストレプトアビジン「アンカー」に結合させた。このシステムでは、kの値は3である。DNA鎖は、400 mM KCl中でMS-(B2)8を使用するk=3ド・ブラン配列 (配列番号3) を表す。前記鎖は、付加電位下でナノポア中に捕捉され、その電流が記録される。下の表1に収載されているように、実験は配列を1ヌクレオチドずつ前進させている一連のDNA鎖を用いて繰り返すことが可能である。このようにして、下の表に収載されているように、移動している鎖から予測される電流レベルに対応する180 mVなどの特定の付加電位での電流レベルの測定が得られた。

30

配列番号3 (k3ド・ブラン) :

A T A A G A A C A T T A T G A T C A G T A G G A G C A C T A C G A C C T T T G T
T C T G G T G C T C G T C C G G G C G C C C A A A T

【0212】

40

【表 1 A】

表 1 :

鎖	配列	測定 (pA)
SD01	CTCTCTCTCTCCTCTCTCTCAAATAAGAACATTATGATCAGTAGG/3BioTEG/	63.3
SD02	CTCTCTCTCTCCTCTCTCTCAATAAGAACATTATGATCAGTAGGA/3BioTEG/	72.6
SD03	CTCTCTCTCTCCTCTCTCTCATAAGAACATTATGATCAGTAGGAG/3BioTEG/	68.2
SD04	CTCTCTCTCTCCTCTCTCTCTAAGAACATTATGATCAGTAGGAGC/3BioTEG/	56.7
SD05	CTCTCTCTCTCCTCTCTCTCAAGAACATTATGATCAGTAGGAGCA/3BioTEG/	55.3
SD06	CTCTCTCTCTCCTCTCTCTCAGAACATTATGATCAGTAGGAGCAC/3BioTEG/	75.6
SD07	CTCTCTCTCTCCTCTCTCTCGAACATTATGATCAGTAGGAGCACT/3BioTEG/	69.0
SD08	CTCTCTCTCTCCTCTCTCTCAACATTATGATCAGTAGGAGCACTA/3BioTEG/	64.5
SD09	CTCTCTCTCTCCTCTCTCTCACATTATGATCAGTAGGAGCACTAC/3BioTEG/	57.8
SD10	CTCTCTCTCTCCTCTCTCTCCATTATGATCAGTAGGAGCACTACG/3BioTEG/	64.3
SD11	CTCTCTCTCTCCTCTCTCTCATTATGATCAGTAGGAGCACTACGA/3BioTEG/	80.4
SD12	CTCTCTCTCTCCTCTCTCTCTTATGATCAGTAGGAGCACTACGAC/3BioTEG/	77.5
SD13	CTCTCTCTCTCCTCTCTCTCTATGATCAGTAGGAGCACTACGACC/3BioTEG/	65.3
SD14	CTCTCTCTCTCCTCTCTCTCATGATCAGTAGGAGCACTACGACCT/3BioTEG/	68.9
SD15	CTCTCTCTCTCCTCTCTCTCTGATCAGTAGGAGCACTACGACCTT/3BioTEG/	67.1
SD16	CTCTCTCTCTCCTCTCTCTCGATCAGTAGGAGCACTACGACCTTT/3BioTEG/	67.3
SD17	CTCTCTCTCTCCTCTCTCTCATCAGTAGGAGCACTACGACCTTTG/3BioTEG/	66.6
SD18	CTCTCTCTCTCCTCTCTCTCTCAGTAGGAGCACTACGACCTTTGT/3BioTEG/	77.7
SD19	CTCTCTCTCTCCTCTCTCTCCAGTAGGAGCACTACGACCTTTGTT/3BioTEG/	67.3
SD20	CTCTCTCTCTCCTCTCTCTCAGTAGGAGCACTACGACCTTTGTTC/3BioTEG/	71.6
SD21	CTCTCTCTCTCCTCTCTCTCGTAGGAGCACTACGACCTTTGTTCT/3BioTEG/	76.9
SD22	TTTTTTTTTTTTTTTTTTTTTAGGAGCACTACGACCTTTGTTCTG/3BioTEG/	58.2
SD23	TTTTTTTTTTTTTTTTTTTTTAGGAGCACTACGACCTTTGTTCTGG/3BioTEG/	68.8
SD24	CTCTCTCTCTCCTCTCTCTCGGAGCACTACGACCTTTGTTCTGGT/3BioTEG/	57.7
SD25	CTCTCTCTCTCCTCTCTCTCGAGCACTACGACCTTTGTTCTGGTG/3BioTEG/	49.1
SD26	CTCTCTCTCTCCTCTCTCTCAGCACTACGACCTTTGTTCTGGTGC/3BioTEG/	50.4
SD27	CTCTCTCTCTCCTCTCTCTCGCACTACGACCTTTGTTCTGGTGCT/3BioTEG/	65.8

10

20

30

【表 1 B】

SD28	TTTTTTTTTTTTTTTTTTTTTTTCACTACGACCTTTGTTCTGGTGCTC/3BioTEG/	50.3
SD29	TTTTTTTTTTTTTTTTTTTTTTTACTACGACCTTTGTTCTGGTGCTCG/3BioTEG/	53.0
SD30	CTCTCTCTCTCCTCTCTCTCCTACGACCTTTGTTCTGGTGCTCGT/3BioTEG/	52.6
SD31	CTCTCTCTCTCCTCTCTCTCTACGACCTTTGTTCTGGTGCTCGTC/3BioTEG/	60.4
SD32	CTCTCTCTCTCCTCTCTCTCACGACCTTTGTTCTGGTGCTCGTCC/3BioTEG/	69.9
SD33	CTCTCTCTCTCCTCTCTCTCCGACCTTTGTTCTGGTGCTCGTCCG/3BioTEG/	59.5
SD34	CTCTCTCTCTCCTCTCTCTCGACCTTTGTTCTGGTGCTCGTCCGG/3BioTEG/	50.7
SD35	CTCTCTCTCTCCTCTCTCTCACCTTTGTTCTGGTGCTCGTCCGGG/3BioTEG/	50.5
SD36	CTCTCTCTCTCCTCTCTCTCCCTTTGTTCTGGTGCTCGTCCGGGC/3BioTEG/	57.1
SD37	CTCTCTCTCTCCTCTCTCTCCTTTGTTCTGGTGCTCGTCCGGGCG/3BioTEG/	67.6
SD38	CTCTCTCTCTCCTCTCTCTCTTTGTTCTGGTGCTCGTCCGGGCGC/3BioTEG/	58.7
SD39	CTCTCTCTCTCCTCTCTCTCTTTGTTCTGGTGCTCGTCCGGGCGCC/3BioTEG/	66.8
SD40	CTCTCTCTCTCCTCTCTCTCTGTTCTGGTGCTCGTCCGGGCGCCC/3BioTEG/	49.6
SD41	CTCTCTCTCTCCTCTCTCTCGTTCTGGTGCTCGTCCGGGCGCCCA/3BioTEG/	58.7
SD42	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAA/3BioTEG/	57.3
SD43	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAAA/3BioTEG/	69.4
SD44	CTCTCTCTCTCCTCTCTCTCCTGGTGCTCGTCCGGGCGCCCAAAT/3BioTEG/	57.0
SD45	CTCTCTCTCTCCTCTCTCTCTGGTGCTCGTCCGGGCGCCCAAATA/3BioTEG/	54.0
SD46	CTCTCTCTCTCCTCTCTCTCGGTGCTCGTCCGGGCGCCCAAATAA/3BioTEG/	65.3
SD47	CTCTCTCTCTCCTCTCTCTCGTGCTCGTCCGGGCGCCCAAATAAG/3BioTEG/	66.2
SD48	CTCTCTCTCTCCTCTCTCTCTGCTCGTCCGGGCGCCCAAATAAGA/3BioTEG/	61.3
SD49	CTCTCTCTCTCCTCTCTCTCGCTCGTCCGGGCGCCCAAATAAGAA/3BioTEG/	75.5
SD50	CTCTCTCTCTCCTCTCTCTCCTCGTCCGGGCGCCCAAATAAGAAC/3BioTEG/	69.4
SD51	CTCTCTCTCTCCTCTCTCTCTCGTCCGGGCGCCCAAATAAGAACA/3BioTEG/	74.5
SD52	CTCTCTCTCTCCTCTCTCTCCGTCCGGGCGCCCAAATAAGAACAT/3BioTEG/	71.6
SD53	CTCTCTCTCTCCTCTCTCTCGTCCGGGCGCCCAAATAAGAACATT/3BioTEG/	79.2
SD54	CTCTCTCTCTCCTCTCTCTCTCGGGCGCCCAAATAAGAACATTA/3BioTEG/	58.5
SD55	CTCTCTCTCTCCTCTCTCTCCGGGCGCCCAAATAAGAACATTAT/3BioTEG/	78.2
SD56	CTCTCTCTCTCCTCTCTCTCCGGGCGCCCAAATAAGAACATTATG/3BioTEG/	81.5
SD57	CTCTCTCTCTCCTCTCTCTCGGGCGCCCAAATAAGAACATTATGA/3BioTEG/	84.7
SD58	CTCTCTCTCTCCTCTCTCTCGGCGCCCAAATAAGAACATTATGAT/3BioTEG/	71.7
SD59	CTCTCTCTCTCCTCTCTCTCGCGCCCAAATAAGAACATTATGATC/3BioTEG/	67.7
SD60	CTCTCTCTCTCCTCTCTCTCCGCCCAAATAAGAACATTATGATCA/3BioTEG/	59.7
SD61	CTCTCTCTCTCCTCTCTCTCGCCCAAATAAGAACATTATGATCAG/3BioTEG/	65.6
SD62	CTCTCTCTCTCCTCTCTCTCCCCAAATAAGAACATTATGATCAGT/3BioTEG/	66.5
SD63	CTCTCTCTCTCCTCTCTCTCCCAAATAAGAACATTATGATCAGTA/3BioTEG/	63.8
SD64	CTCTCTCTCTCCTCTCTCTCCAAATAAGAACATTATGATCAGTAG/3BioTEG/	70.6

【0 2 1 3】

それぞれ個々の鎖からのデータは順次プロットされて図 2 9 に示される電流状態の地図 (散布図) を生み出し、それぞれの点は S D 0 1 (左) から S D 6 4 (右) までの DNA 鎖を表す。データは、ポリ T 鎖からのふれとしてプロットされている。

【0 2 1 4】

これらの測定を使用して、図 2 9 に示される測定を中心とするそれぞれの k マーの分布として放出重み付け 1 5 を導き出すことができる。ガウス分布は、図 2 9 に示される測定から得られる標準偏差を用いて使用しうる。遷移重み付け 1 4 は手動で選択しうる。

【0 2 1 5】

第 2 の動的トレーニング技法は以下の通りに実施される。

【0 2 1 6】

静的鎖トレーニングは多くの利点を提供するが、骨が折れることがあり、一部では測定システムは完全な塩基配列決定システムを正確に反映しないこともある。解析ステップ S 2 において使用する枠組みに類似する枠組み (および、したがって、類似するアルゴリズム) を利用することにより、モデル 1 3 を別法で訓練することが可能である。このそのよ

10

20

30

40

50

うな1つの実行はここで説明されるが、多くのバリエーションを適用することができる。説明されるプロセスは反復性のプロセスなので、始めるのに用いるパラメータの合理的推定値（ベイズ理論用語で、プライア（prior））があるのは有用である。3マー静的係数は、より高度なkマーモデルを訓練するための合理的開始点を提供する。

【0217】

トレーニングが適用されるので、状態コーディングモデルよりもかなり柔軟性が少ないモデルが使用される。トレーニング鎖（複数可）の配列は既知であるので、大きな制約を適用することができる。すべてのkマー間の認められた遷移をモデル化するよりも、我々のトレーニング配列により認められた遷移のみがモデル化される。トレーニングをさらに制約するため、トレーニング鎖中のそれぞれの位置は独立してモデル化され、直後の状態への遷移のみが好ましい。したがって、これを「強制的経路（forced path）」モデルと呼ぶことができるであろう。

10

【0218】

例えば、およそ400単位のポリマーであれば、そのポリマー中の位置ごとの別々の状態インデックスを定義することができる。次に、図30および31に示されるように、ポリマー内での遷移を許す遷移マトリックスが構築され、図30が408のkマー状態についての遷移マトリックスを示し、図31が最初の10の遷移重み付けの大写しを示している。

【0219】

上記のモデル13における遷移重み付け14のkマー推定遷移マトリックスの場合と同じように、これが現実世界のシステムであるという事実を考慮するように柔軟性を加えることが可能である。この例では、遷移がないこと（または、起点状態インデックスと目的地状態インデックスが同じ状態である遷移）が認められ、状態を読み飛ばす好ましくない遷移について非ゼロの確率を使用することにより見落とされた測定が収容される。確率的（または、重み付けされた）枠組みの利点は、測定システムの既知の人為産物を遷移重み付けおよび／または放出重み付けにおいて具体的に扱うことができる点である。

20

【0220】

放出重み付けのトレーニングがここで説明される。放出重み付けの分布は、上記解析ステップS2のために使用された分布と類似していてもよい。しかし、この例では、ポリマー中のそれぞれの位置が別々に扱われるために、放出分布は位置ごとに定義される。図32は、上記の静的トレーニングプロセスから導かれる64kマーモデルの例を示している。図33は、およそ400の状態の配列に移し返られた図32の64kマーモデルの例を示している。既に記載されたように、外れ値データは、あらゆる可能な測定値について非ゼロの確率を有する放出重み付けの分布内に収容することができる。

30

【0221】

トレーニングプロセスは図34に示されており、ここで説明される。トレーニングプロセスは反復性であり、先ず、上記のモデル20の最初の推定値をモデル21の推定値として使用する。トレーニングプロセスは測定22も使用する。

【0222】

モデル21の推定値および測定22を考慮すれば、ステップS3において、広い範囲の既知のアルゴリズムのうちのいずれか1つを適用することにより、測定22がモデルにどのようなようにして適合するのかが計算される。HMMのケースでは、1つの適切なアルゴリズムはフォワードバックワードアルゴリズムである。

40

【0223】

次に、ステップS4では、ステップS3において計算されたモデルに適合するデータを使用して、どんな根底にある状態放出分布がその適合下にあるかを推定し、kマー状態中心を再推定し、それによってモデル21の推定値を更新する。

【0224】

ステップS5では、トレーニングプロセスが収束したかどうか、すなわち、ステップS4からのモデル21の更新された推定値が以前の繰り返しから著しく変化していないかど

50

うかが決定される。収束していなければ、前記プロセスはモデル 2 1 の更新された推定値を使用して繰り返される。

【0 2 2 5】

そのような繰り返しは、ステップ S 5 において収束が決定されるまで行われる。この時点で、モデル 2 1 の更新された推定値は測定 2 2 の記述に収束しており、出力モデル 2 3 として出力される。

【0 2 2 6】

これはトレーニングプロセスのための機械学習アルゴリズムの 1 つの可能な実行であるが、当技術分野で公知の他の機械学習方法を使用することもできるであろう。

【0 2 2 7】

図 6 の解析方法が図 9 の実験的に決定された入力シグナル 1 1 に適用される例がここで説明されることになる。上記のように、状態検出ステップ S 1 により導き出される測定のシリーズ 1 2 は図 1 0 に示されている。

【0 2 2 8】

ポリマーはポリヌクレオチドであり、測定を記述するのに使用される k マーモデルは 3 マーである。

【0 2 2 9】

モデル 1 3 は、図 1 6 に示され上に記載されている遷移重み付け 1 4 を含む。

【0 2 3 0】

前記モデルは、上記の図 3 4 のトレーニングプロセスを使用して決定された放出重み付け 1 5 を含む。図 3 5 は、小非ゼロバックグラウンドを有するガウス分布である結果として得られた放出重み付け 1 5 を示している。

【0 2 3 1】

図 3 6 は、モデル 1 3 から予測される測定値を用いて、いくつかの実験にわたり集計された状態データのセクションからの電流測定のオーバーレイを示している。

【0 2 3 2】

図 3 7 は、既知の配列（参照）と解析ステップ S 2 により推定された k マー状態の推定された配列（コール）の状態スペースアライメントを示している。正確に推定された k マー状態は大きな点として示されている。見ることにできるように、k マー状態の良好な推定が与えられている。

【0 2 3 3】

図 3 8 は、解析ステップ S 2 により推定され、実際の配列と整列させて示されるヌクレオチドの推定された配列 1 6 を示している。正確な k マー状態推定値は「#」として図示されている（k マー状態を直接塩基と関連付けてきたので、これを示すことができる）。正確な塩基推定値であるが不正確な k マー状態推定値は「*」として図示されている。

【0 2 3 4】

上記説明は、前記方法が単一入力シグナル 1 1 および測定の単一シリーズ 1 2 に基づいているケースに関係している。

【0 2 3 5】

代わりに、本発明の第一の態様は、それぞれが同じポリマーに関係している測定の複数のシリーズを使用してもよい。この文脈では、「同じ」ポリマーとは、同じ正体または組成を有するポリマーであり、物理的に同じポリマーまたは同じ正体を有する物理的に異なるポリマーである。測定の複数のシリーズは同じポリマーで行ってもよいし、関連する配列を有する異なるポリマーで行ってもよい。

【0 2 3 6】

測定の複数のシリーズはそれぞれが同じ技法で行ってもよいし、異なる技法で行ってもよい。測定の複数のシリーズは同じ測定システムで行ってもよいし、異なる測定システムで行ってもよい。

【0 2 3 7】

測定の複数のシリーズは、同じポリマーの同じ領域で同時に行われる異なる種類、例え

10

20

30

40

50

ば、同時に行われる膜貫通電流測定と F E T 測定であっても、または同時に行われる光学的測定と電氣的測定であってもよい (Heron AJ et al., J Am Chem Soc. 2009;131(5):1652-3)。所与のポリマーまたはその領域を 1 回よりも多くポア中を通して移行させることにより複数の測定を次々に行うことも可能である。これらの測定は同じ測定でもまたは異なる測定でもよく、同じ条件下でもまたは異なる条件下でも行うことができる。

【0238】

測定の複数のシリーズは関係のあるポリマーの領域で行いうる。このケースでは、測定のシリーズは、関係のある配列を有する別々のポリマーの測定でもよくまたは関係のある配列を有する同じポリマーの異なる領域の測定でもよい。後者の例として、関係が配列が相補的であるという場合に、ポリヌクレオチドについて提唱された技法が使用されることもある。このケースでは、センス鎖とアンチセンス鎖が、ポリヌクレオチド結合タンパク質を使用してまたはポリヌクレオチド試料調製を介して順次読み取られうる。特許仮出願第 61/511436 号または WO-2010/086622 に提示されているいかなる方法でも使用してセンス鎖とアンチセンス鎖を読み取らせてもよい。

10

【0239】

この例として、図 6 に図示されている方法は、状態検出ステップ S1 において処理される複数の入力シグナル 11 に適用して、測定の複数のシリーズ 12 を提供してもよい。このケースでは、上に詳細に記載されているように、それぞれの入力シグナル 11 および測定シリーズ 12 は、同じポリマーの同じ領域の測定であることにより、または同じポリマーもしくは異なるポリマーの異なっているが関係のある領域（例えば、DNA 鎖および相補的 DNA 鎖）の測定であることによりいずれかで前記ポリマーと関係している。

20

【0240】

このケースでは、解析方法は基本的に同じであるが、測定のそれぞれのシリーズ 12 からの測定は、複数のそれぞれの次元で配置されているステップ S2 における解析技法により扱われる。

【0241】

これは、解析ステップ S2 においてそれぞれの入力シグナル 11 と測定のシリーズ 12 を別々に処理するよりはかなり有利である。解析のこの初期段階で測定のシリーズ 12 からの情報を組み合わせることにより、根底にあるポリマー単位のより正確な推定を行うことが可能である。解析プロセス初期の情報を組み合わせると、解析プロセスの終了時に測定のシリーズ 12 と組合せを独立に処理するよりも正確な出力が可能になる。これは、根底にあるポリマー関係による以外に、測定のシリーズ 12 が関係があるといういかなる要件もなしで達成されうる。確率的技法または他の解析技法でも、解析が、測定の関係のあるシリーズ 12 の位置合わせまたはアライメントを推定することが可能になる。測定のどんなシリーズでも他の任意の測定のシリーズに位置合わせをすることは先験的に分かっていることもあれば分かっていることもあることに注目するのは重要である。位置合わせがないケースでは、シリーズ内のそれぞれの測定は別のシリーズからの測定と先験的に対になっていない。

30

【0242】

数学的に言えば、2 つのそれぞれの次元に配置されている測定のシリーズ 12 を扱う解析ステップ S2 の拡張は単刀直入である。放出重み付け 15 は複数の次元で起こり、測定のシリーズ 12 ごとに 1 つの次元である。方法が、登録されている測定の複数のシリーズ 12 で実施され、したがってそれぞれのシリーズからのどの測定が対応しており同じ k マーに依拠しているのかが先験的に分かっているケースでは、モデル 13 を、放出重み付け 15 を k マー状態ごとの複数の測定の分布を記述している確率密度関数として複数の次元で使用して適用しうる。

40

【0243】

これとは対照的に、方法が登録されていない複数のシリーズで実施され、したがってそれぞれのシリーズからのどの測定値が対応しており同じ k マーに依拠しているのかが先験的に分かっているケースでは、前記方法は、以下の通りに、測定の複数のシリーズを、

50

複数のそれぞれの次元で配置された全体として扱う。

【0244】

放出分布のそれぞれの次元は、スキップ状態で増大され、多次元重みはその発生の可能性を表す。個々のシリーズでスキップが起こる場合、放出分布は対応する次元での測定値よりはむしろ「スキップ」シグナル状態を放出するように取られる。これら「スキップ」状態は観測可能ではなく、これらの状態の未知の数および位置が位置合わせ問題を引き起こす。解析ステップS2は、測定の複数のシリーズ12がkマーおよびポリマー単位の異なる配列から導き出される尤度およびこれらの測定間の異なる位置合わせではそれぞれの位置合わせが放出分布において潜在している可能性に基づいて実施される。

【0245】

登録されたケースでも非登録のケースでも、測定の複数のシリーズ12が同じ特性（例えば、同じポリマーの繰返し測定について）の等しい測定である場合、それぞれのシリーズ12に関する放出重み付け15は同じであってよい。測定の複数のシリーズ12が異なる特性（例えば、同じポリマーの異なる測定について、またはポリマーの異なっているが関係のある領域の測定について）の測定である場合、それぞれのシリーズ12に関する放出重み付け15は異なってもよい。

【0246】

上記のグラフィックモデルBを考慮すると、概念的にはモデルは、 X_i がここでは単一の値というよりむしろ値のベクトルを表すこと以外は同じである。HMMのケースでは、1次元確率密度関数 $g()$ からの状態放出値というよりはむしろ、値は複数次元密度関数から放出され、例えば、センス鎖およびアンチセンス鎖の測定のケースでは、 X_i は電流ペア (X_{is}, X_{ia}) を放出し、 X_{is} はセンス鎖からの電流読取りであり X_{ia} は相補的kマーに対するアンチセンス鎖からの読みである。この放出された電流ペアは、観測されないスキップ状態ならびに実際の電流測定を含有していてもよい。基本的な1次元ケースと同じように、外れ値および失われたデータまたは読み飛ばされた状態をモデル化することができる。

【0247】

有利なことに、ポリマーのうちの1つでのスキップは関係するポリマーからの情報を使用して埋めることができる。例えば、センサーアンチセンスデータに関しては、2次元密度 $g()$ がスキップを非ゼロ確率で1次元に放出し、その間電流をもう一方の次元からサンプリングすることを可能にすることにより、スキップをアンチセンスではなくセンスにおいて（またはその逆）放出してもよく、したがって、 X_1 は形式 (X_{1s}, X_{1a}) 、 $(X_{1s}, -)$ または $(-, X_{1a})$ （-は観測されないスキップを表す）の電流ペアを放出しうる。さらに、両方のポリマーにおけるスキップもモデル化し、1Dケースにおけるのと同じように補正することができる。ここで、測定の1つのシリーズにおける「ステイ」も、その他のものについてのスキップ状態を放出することによりモデル化することができる。

【0248】

前記1次元HMMからの利点はすべてこの複数次元HMMに移動する。同様に、2つの別々の1次元HMMを実行し次にアライメント技法を通じて底空間で整列させることよりも利点がある。

【0249】

単なる例として、複数の次元で配置された測定にビタビアルゴリズムを適用することが考察される。ビタビアルゴリズムは当技術分野では周知である。1次元HMMでは、最も可能性の高い経路がそれぞれの可能なkマーKで終わる尤度 $L_i(k)$ は、最初の状態から最後の状態へ状態配列の中を前方に移動するそれぞれの状態 i ($i = 1 \dots n$) ごとに計算される。測定の複数のシリーズ間の位置合わせの欠如によって、そのような経路はすべて考慮しなければならない。値 $L_i(K)$ はすぐ前を先行する状態からの値 $L_{i-1}(\cdot)$ のみを遷移および放出確率と共に使用し、再帰を形成して計算することができる。m次元HMMでは、類似するスキームを使用しうる。スキップが組み込まれるためには、

m 個のインデックスがあり、したがって、 $L_{i_1, i_2, \dots, i_m}(K)$ は、次元1における状態 i_1 、次元2における状態 i_2 などを記述する最大尤度である。それは可能なあらゆる量 $L_{j_1, j_2, \dots, j_m}(K)$ を調べることにより再帰的に計算することができ、スキップが次元1で放出されるならば $j_1 = i_1$ または状態が次元1において放出されるならば $(i_1 - 1)$ であり、 j_2 、 j_3 、等についても同様である。

【0250】

この解析法は、それぞれの入力シグナル11および測定シリーズ12が同じポリマーの同じ領域の測定である場合に適用しうる。例えば、ポリマーまたはポリマーの領域が再読取りされるシステムでは、これらの読みは組み合わせることができ、位置合わせまたはアライメントを推定して根底にある k マー状態をさらに正確に決定することができる。前記方法は、異なる条件下でまたは組み合わせられる異なる方法により行われる測定も可能にする。

10

【0251】

上で考察されたように、例えば、測定の複数シリーズが複数の電氣的測定または電氣的および光学的測定を含む場合、複数の測定も同時に行いうる。これらの読みは組み合わせられるおよび／または位置合わせもしくはアラインメントが推定されて根底にあるポリマー配列をより正確に推定することができる。

【0252】

代わりに、測定の複数のシリーズ12は集合されて、1次元測定として解析ステップS2により使用される測定のサマリーシリーズを提供する。 m 個の異なる種類の複数の測定シリーズが存在する場合、同じ種類のすべてのシリーズに集合が適用され、 m 次元HMMがサマリー状態シリーズに用いることができる。代わりに、複数のシリーズが存在する場合、それぞれのシリーズまたはそれぞれのサマリー測定シリーズに1次元HMMを実行してよく、これらの解析からの出力に基づいてコンセンサスコールが行われる。

20

【0253】

この解析法は、入力シグナル11および測定の2つのシリーズを含む測定のシリーズ12にも適用することができ、測定の第1のシリーズはポリマーの第1の領域の測定であり、測定の第2のシリーズは前記第1の領域に関係しているポリマーの第2の領域、例えば、同じポリマーまたは異なるポリマーの相補的領域の測定である。

【0254】

この技法は、DNA配列の相補的対、すなわち、「センス」鎖およびその相補的「アンチセンス」鎖への特定の適用を有する。

30

【0255】

2つの別々の1次元HMMおよび次にアライメント技法を通じた底空間で整列させることにまさる2次元アプローチの利点はここで説明されることになる。

【0256】

極度に単純化した説明として、センス鎖上でのHMMから $Pr(AAACAAA) = 0.6$ 、 $Pr(AAAGAAA) = 0.39$ 、 $Pr(AAAAAAA) = 0.01$ およびアンチセンス鎖上でのHMMから $Pr(TTTTTTT) = 0.6$ 、 $Pr(TTTC TTT) = 0.39$ 、 $Pr(TTTG TTT) = 0.01$ が想定されている。センスおよびアンチセンスについて最も可能性の高い配列が取られ、センスーアンチセンスペアとして整列されるように企てられた場合、配列の中間の塩基でクラッシュ(clash)が得られる。2次元HMMは、配列の飛び抜けて最も可能性の高い一貫したペアは(AAAGAAA、TTTC TTT)であることを見つけ、配列ペア(AAACAAA、TTTG TTT)および(AAAAAAA、TTTTTTT)には低い確率を割り当てると考えられる。

40

【0257】

この極度に単純化した説明では、2番目に可能性が高い配列はそれぞれの1次元HMMにより問題を解決すると見なされるが、さらに長い配列についてのすべての必要なポリマー単位推定を一通り調べるのはすぐに非現実的になる。さらに、ポリマー単位を推定するためのいくつかの方法(例えば、ビタビ)は最も確実な経路を放出するだけで、ポリマー

50

単位を推定した後の可能性のより低い配列の組合せは不可能になる。

【0258】

ビタビアルゴリズムを使用するセンサーアンチセンスケースの特定の詳細な例は、改良を実証するためにここで説明される。

【0259】

センサーアンチセンスのケースでは、上記の m 次元ケースは、 $m = 2$ について使用され、 $L_{ij}(K)$ は、状態がセンスのみにより、アンチセンスのみにより、または両方により放出されるのかに応じて、値 $L_{i-1,j}(\cdot)$ 、 $L_{i,j-1}(\cdot)$ および $L_{i,j}(\cdot)$ を使用して計算される。

【0260】

図39は、最も可能性の高いセンスおよびアンチセンス配列の独立したコールが3マーモデルおよびHMMを使用して行われる例を図示している。ジョイントセンサーアンチセンスコールは上記の2次元ビタビアルゴリズムを使用して行われる。ジョイントコールは正確で例外は非常に少なく、特にセンスコールとアンチセンスコールの両方で不正確にコールされている塩基を正確にコールしている。正確な3マー状態推定値は「#」で示され、正確な塩基は「*」で示されている。この図解では、独立したセンス読取りとアンチセンス読取りの最良の領域を組み合わせても、センサーアンチセンス結果の正確なコールの数を占めていないことが見て取れる。解析プロセス初期のデータの組合せは、確率的アプローチと組み合わせると、「部分の総和超 (more than the sum of the parts)」の結果をもたらす。

【0261】

この多次元例は、加算された情報が1つの鎖は別の鎖に相補的であることであるセンサーアンチセンスDNAのケースのためであるが、ポリマーの領域間の他の関係は多次元アプローチにおいてコード化しうる。コード化できるであろう別の種類の情報の例は、ポリマー中の構造情報である。この情報は、機能的構造体を形成することがわかっているRNAに存在しうる。この情報は、ポリペプチド（タンパク質）にも存在しうる。タンパク質のケースでは、構造情報は、疎水性領域または親水性領域に関係していてもよい。前記情報は、アルファヘリックス、ベータシートまたは他の二次構造に関してでもよい。前記情報は、結合部位、触媒部位および他のモチーフなどの既知の機能的モチーフに関してでもよい。

【0262】

本発明の第二の態様および第三の態様に従ってポリマーの測定を行う方法がここで考察されることになる。下でさらに詳細に考察されるように、これは場合によっては、本発明の第一の態様に従った上記の方法と組み合わせてもよい。

【0263】

この方法では、測定はナノポア中を流れるイオン電流の測定である。この方法では、ポリマーは、ナノポアを横断して電圧が印加されている間にナノポアの中を移行する。測定はナノポア中の k マーの正体に依拠している。測定はナノポアを横断する異なるレベルの電圧の印加下で行われる。そのような測定により、単なる重複性ではなく追加の情報が与えられることは本発明者らにより認識されていた。この利点のいくつかの特定の実証がここで説明されることになる。

【0264】

最初の例は、付加電位下で測定システム中に静止して保持されたDNAの鎖であるポリマーのイオン電流測定の分解を図示している。この例では、電流が最初の通常の電圧レベルでは互いに類似しているDNA配列は、第2の電圧レベルで記録することにより分解された。

【0265】

DNA鎖は、Proc Natl Acad Sci U S A. 2009 May 12;106(19):7702-7に既に報告されている方法に類似するストレプトアビジンアンカーを使用してナノポア中に保持された。DNAの個々の鎖が当技術分野で公知の方法を使用してDPHP C二重層に包埋された単

10

20

30

40

50

一MS-(B1)₈ ナノポア中で測定される場合、ラン(run)が収集された。電圧はナノポアを横切って印加され、電流はナノポアのどちら側でも塩溶液中のイオンの動きから生み出された。

【0266】

ラン条件は、400mM KCl、10mM Hepes、pH8.0、+180mVであった。対照配列(TS01)はストレプトアビジンと一緒に2対1比でインキュベートされ、チャンバーに添加されて最終濃度200nM DNAを得た。分析物配列はストレプトアビジンと一緒に2対1比でチャンバーに添加され、最終分析物DNA濃度400nMを得た。両方のケースで、ビオチン化DNAとストレプトアビジンは、チャンバーへの添加に先立って5分間インキュベートされた。シングルチャネルレコーディングは、+180mV(2秒)から-180mV(0.2秒)まで付加電位を変化させる自動化手順を使用して実施された。正の付加電位を使用して捕捉してDNAレベルを読み取り、負電位を使用してナノポアからストレプトアビジン-DNA複合体を排出した。

【0267】

DNA結合事象(状態)ごとの平均電流は以下の通りに調べられた。

【0268】

TS01対照と分析物配列からの集団が記録された。分析物配列電流レベルは、以下の関係

$$I_{DNA \text{ Adjusted}} = I_{DNA \text{ Recorded}} - I_{TS01} + 32.2 \text{ pA}$$
を使用することにより調整された。このプロセスは、広範囲の様々なDNA配列について繰り返された。例として、表2は、+180mVの電圧で測定された場合、調整された電流レベルが類似する大きさ(54.5±0.5 pA)を示した選択された配列を提示している。

【0269】

【表2】

表2:

コード	配列	トリプレット	電流(pA)
TS01	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT/3BioTEG/	コントロール	28.5-33.3
SD90	CTCTCTCTCTCCTCTCTCTCGACGAGCACCAGAACAAGGTCGTA/3BioTEG/	ACA	54.6
SD85	TTTTTTTTTTTTTTTTTTTTTTTTTTGCCCGGACGAGCACCAGAACAAGG/3BioTEG/	CCA	55.0
SD81	CTCTCTCTCTCCTCTCTCTCGGGCGCCCGGACGAGCACCAGAACA/3BioTEG/	AGC	54.9
SD59	CTCTCTCTCTCCTCTCTCTCGCGCCCAATAAGAACATTATGATC/3BioTEG/	AAC	54.0
SD52	CTCTCTCTCTCCTCTCTCTCCGTCCGGGCGCCCAATAAGAACAT/3BioTEG/	AAA	54.7
SD18	CTCTCTCTCTCCTCTCTCTCTCAGTAGGAGCACTACGACCTTTGT/3BioTEG/	TAC	54.2
SD15	CTCTCTCTCTCCTCTCTCTCTGATCAGTAGGAGCACTACGACCTT/3BioTEG/	CAC	54.6
SD03	CTCTCTCTCTCCTCTCTCTCATAAGAACATTATGATCAGTAGGAG/3BioTEG/	GAT	54.3
S142	CTCTCTCTCTCCTCTCTCTCTTTGGGCGCCCGGACGAGCACCAGA/3BioTEG/	ACG	54.7
S117	CTCTCTCTCTCCTCTCTCTCGCTCCTACTGATCATAATGTTCTTA/3BioTEG/	ATA	54.1
S116	CTCTCTCTCTCCTCTCTCTCTGCTCCTACTGATCATAATGTTCTT/3BioTEG/	CAT	54.3

【0270】

それに続く実験では、DNAの同じ鎖はすべて、脂質膜に包埋された単一MS-(B1)₈ ナノポアを含有するチャンバー中に置かれた。条件は、上記条件の400mM KCl、10mM Hepes、pH8.0、+180mVに類似していた。分析物配列はすべてストレプトアビジンと一緒に2対1比でチャンバーに添加され、最終濃度は分析物D

NA ごとに 200 nM DNA であった。TSO1 はこの実験では添加されなかった。ビオチン化 DNA とストレプトアビジンは、チャンバーへの添加に先立って 5 分間インキュベートされた。

【0271】

DNA 識別に対する付加電位の効果を調べるため、この実験では電圧を変化させた。シングルチャネルレコーディングは、付加電位を +X (2 秒) から -X (0.2 秒) まで変化させる自動化手順を使用して実施され、X は 140 mV、180 mV および 220 mV である。シングルチャネルデータは、X の値ごとにおよそ 30 分間記録された。

【0272】

DNA 結合事象 (状態) ごとの平均電流レベルが記録され、それぞれ +140 mV、+180 mV および +220 mV の正電位に関して図 40 に示されているヒストグラムのセットにプロットされている。これらの結果を考慮すると、+180 mV でのデータは予想通りに振る舞っており、表 1.1 の 11 の鎖すべてが非常に類似する電流レベルを生じていることは明白である。+220 mV では、電流レベルヒストグラムが広がるまたは散らばっており、レベルが分離していることが示唆される。+140 mV でも、広がるまたは散らばっており、同様に電流レベルは明らかに多数のはっきり異なる集団に分解している。これらの結果から、+180 mV ではできなかった多くの DNA 鎖を +140 mV では互いに区別することができることが示唆される。実験の容易さのために、これはナノポア中で静止している鎖を用いて実施された例であるが、DNA 鎖が異なればナノポア中の関連のある位置で提供される k マーも異なり、イオン電流に影響を与えるので、DNA 鎖の異なる k マーにより生み出されるイオン電流間の類似の分離はポア中を動的に移行すると予測される。

【0273】

第 2 の例は、付加電位下で測定システム中で静止して保持される DNA の鎖であるポリマーのイオン電流測定の分離を図示している。この例では、異なる電圧レベルでのイオン電流の測定は異なる k マーを分解することが示されている。

【0274】

第 2 の例では、所与の鎖の電流レベルに対する付加電位の効果を決定するために、DNA 配列はすべての可能なトリプレットを含有するように選択された (ド・ブラン、GTA C、k3、配列番号 5)。

配列番号 5 (k3 ド・ブラン)

A T A A G A A C A T T A T G A T C A G T A G G A G C A C T A C G A C C T T T G T
T C T G G T G C T C G T C C G G G C G C C C A A A T

【0275】

鎖の動きから生じるどんな可能な複雑化もなく電流レベルの効果を評価するために、一連の異なる DNA 鎖が設計された。これらの鎖はそれぞれが、3' 末端にビオチン-TEG リンカー、k3 ド・ブラン配列の一部 (35 ヌクレオチド長)、および DNA をナノポア内に通すのを支援する低二次構造を有するセクション (10 ヌクレオチド長) を含有していた。k3 ド・ブランを含有するセクションの配列は、配列が鎖あたり 1 ヌクレオチド移動されるように変化させた。リーダーセクションは、ド・ブランセクションにハイブリダイズしないように選択された。これらのコードおよび対応する配列は表 3 に収載されている。

【0276】

【表 3 A】

表 3 :

鎖	配列	変化する付加電位での電流 (pA)			
		180 mV	140 mV	100 mV	60 mV
SD01	CTCTCTCTCTCCTCTCTCTCAAATAAGAACATTATGATCAGTAGG/3BioTEG/	63.3	36.2	17.2	7.8
SD02	CTCTCTCTCTCCTCTCTCTCAATAAGAACATTATGATCAGTAGGA/3BioTEG/	72.6	41.3	20.1	4.9
SD03	CTCTCTCTCTCCTCTCTCTCATAAGAACATTATGATCAGTAGGAG/3BioTEG/	68.2	37.3	18.2	7.0
SD04	CTCTCTCTCTCCTCTCTCTCTAAGAACATTATGATCAGTAGGAGC/3BioTEG/	56.7	34.6	18.0	8.5
SD05	CTCTCTCTCTCCTCTCTCTCAAGAACATTATGATCAGTAGGAGCA/3BioTEG/	55.3	30.4	15.8	6.4
SD06	CTCTCTCTCTCCTCTCTCTCAGAACATTATGATCAGTAGGAGCAC/3BioTEG/	75.6	40.5	18.7	7.5
SD07	CTCTCTCTCTCCTCTCTCTCGAACATTATGATCAGTAGGAGCACT/3BioTEG/	69.0	40.6	19.9	8.0
SD08	CTCTCTCTCTCCTCTCTCTCAACATTATGATCAGTAGGAGCACTA/3BioTEG/	64.5	40.5	21.2	7.5
SD09	CTCTCTCTCTCCTCTCTCTCACATTATGATCAGTAGGAGCACTAC/3BioTEG/	57.8	31.9	17.6	7.6
SD10	CTCTCTCTCTCCTCTCTCTCCATTATGATCAGTAGGAGCACTACG/3BioTEG/	64.3	35.7	17.0	7.6

【表 3 B】

SD11	CTCTCTCTCTCCTCTCTCTCATTATGATCAGTAGGAGCACTACGA/3BioTEG/	80.4	47.0	22.5	6.3
SD12	CTCTCTCTCTCCTCTCTCTCTTATGATCAGTAGGAGCACTACGAC/3BioTEG/	77.5	47.0	24.8	
SD13	CTCTCTCTCTCCTCTCTCTCTATGATCAGTAGGAGCACTACGACC/3BioTEG/	65.3	41.2	23.2	10.2
SD14	CTCTCTCTCTCCTCTCTCTCATGATCAGTAGGAGCACTACGACCT/3BioTEG/	68.9	40.0	21.6	8.8
SD15	CTCTCTCTCTCCTCTCTCTCTGATCAGTAGGAGCACTACGACCTT/3BioTEG/	67.1	39.8	21.4	10.4
SD16	CTCTCTCTCTCCTCTCTCTCGATCAGTAGGAGCACTACGACCTTT/3BioTEG/	67.3	38.8	20.9	10.5
SD17	CTCTCTCTCTCCTCTCTCTCATCAGTAGGAGCACTACGACCTTTG/3BioTEG/	66.6	39.3	21.0	10.0
SD18	CTCTCTCTCTCCTCTCTCTCAGTAGGAGCACTACGACCTTTGT/3BioTEG/	77.7	44.7	22.1	7.0
SD19	CTCTCTCTCTCCTCTCTCTCCAGTAGGAGCACTACGACCTTTGTT/3BioTEG/	67.3	37.7	19.0	8.5
SD20	CTCTCTCTCTCCTCTCTCTCAGTAGGAGCACTACGACCTTTGTTT/3BioTEG/	71.6	41.3	20.0	7.8
SD21	CTCTCTCTCTCCTCTCTCTCGTAGGAGCACTACGACCTTTGTTCT/3BioTEG/	76.9	47.3	24.6	7.9
SD22	TTTTTTTTTTTTTTTTTTTTTAGGAGCACTACGACCTTTGTTCTG/3BioTEG/	58.2	33.4	18.0	6.9
SD23	TTTTTTTTTTTTTTTTTTTTTAGGAGCACTACGACCTTTGTTCTGG/3BioTEG/	68.8	37.5	18.6	8.1
SD24	CTCTCTCTCTCCTCTCTCTCGGAGCACTACGACCTTTGTTCTGGT/3BioTEG/	57.7	34.4	17.1	7.4
SD25	CTCTCTCTCTCCTCTCTCTCGAGCACTACGACCTTTGTTCTGGTG/3BioTEG/	49.1	28.8	17.2	8.1
SD26	CTCTCTCTCTCCTCTCTCTCAGCACTACGACCTTTGTTCTGGTGC/3BioTEG/	50.4	25.8	13.5	7.9
SD27	CTCTCTCTCTCCTCTCTCTCGCACTACGACCTTTGTTCTGGTGCT/3BioTEG/	65.8	34.8	13.8	2.5
SD28	TTTTTTTTTTTTTTTTTTTTTCACTACGACCTTTGTTCTGGTGCTC/3BioTEG/	50.3	28.9	14.2	5.7
SD29	TTTTTTTTTTTTTTTTTTTTTACTACGACCTTTGTTCTGGTGCTCG/3BioTEG/	53.0	27.0	12.9	2.9
SD30	CTCTCTCTCTCCTCTCTCTCCTACGACCTTTGTTCTGGTGCTCGT/3BioTEG/	52.6	24.8	10.6	4.3
SD31	CTCTCTCTCTCCTCTCTCTCAGACCTTTGTTCTGGTGCTCGTC/3BioTEG/	60.4	30.4	11.9	5.0
SD32	CTCTCTCTCTCCTCTCTCTCAGACCTTTGTTCTGGTGCTCGTCC/3BioTEG/	69.9	39.8	17.0	2.2
SD33	CTCTCTCTCTCCTCTCTCTCCGACCTTTGTTCTGGTGCTCGTCCG/3BioTEG/	59.5	34.3	17.0	5.6
SD34	CTCTCTCTCTCCTCTCTCTCGACCTTTGTTCTGGTGCTCGTCCGG/3BioTEG/	50.7	30.2	16.6	6.5
SD35	CTCTCTCTCTCCTCTCTCTCACCTTTGTTCTGGTGCTCGTCCGGG/3BioTEG/	50.5	27.6	14.6	5.9
SD36	CTCTCTCTCTCCTCTCTCTCCCTTTGTTCTGGTGCTCGTCCGGGC/3BioTEG/	57.1	29.9	14.9	7.0
SD37	CTCTCTCTCTCCTCTCTCTCTTTGTTCTGGTGCTCGTCCGGGCG/3BioTEG/	67.6	37.4	17.2	
SD38	CTCTCTCTCTCCTCTCTCTTTGTTCTGGTGCTCGTCCGGGCGC/3BioTEG/	58.7	33.2	16.5	7.2
SD39	CTCTCTCTCTCCTCTCTCTTTGTTCTGGTGCTCGTCCGGGCGCC/3BioTEG/	66.8	37.6	17.1	5.0
SD40	CTCTCTCTCTCCTCTCTCTGTTCTGGTGCTCGTCCGGGCGCCC/3BioTEG/	49.6	30.8	18.5	
SD41	CTCTCTCTCTCCTCTCTCTCGTTCTGGTGCTCGTCCGGGCGCCCA/3BioTEG/	58.7	30.1	14.0	5.9
SD42	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAA/3BioTEG/	57.3	26.9	11.8	6.5
SD43	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAAA/3BioTEG/	69.4	37.1	14.6	5.4
SD44	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAAA/3BioTEG/	57.0	35.2	15.7	4.3
SD45	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAAA/3BioTEG/	54.0	32.0	19.5	7.1
SD46	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAAA/3BioTEG/	65.3	34.9	17.2	8.1
SD47	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAAA/3BioTEG/	66.2	38.7	19.4	8.7
SD48	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAAA/3BioTEG/	61.3	36.7	20.1	9.2
SD49	CTCTCTCTCTCCTCTCTCTCTTCTGGTGCTCGTCCGGGCGCCCAAA/3BioTEG/	75.5	43.4	21.1	6.8

10

20

30

40

【表 3 C】

SD50	CTCTCTCTCTCCTCTCTCTCTCGTCCGGGCGCCCAAATAAGAAC/3BioTEG/	69.4	39.0	19.6	8.9
SD51	CTCTCTCTCTCCTCTCTCTCTCGTCCGGGCGCCCAAATAAGAAC/3BioTEG/	74.5	44.2	21.6	8.8
SD52	CTCTCTCTCTCCTCTCTCTCTCGTCCGGGCGCCCAAATAAGAACAT/3BioTEG/	71.6	42.8	22.9	9.1
SD53	CTCTCTCTCTCCTCTCTCTCTCGTCCGGGCGCCCAAATAAGAACATT/3BioTEG/	79.2	45.9	23.3	7.8
SD54	CTCTCTCTCTCCTCTCTCTCTCGGCGCCCAAATAAGAACATTA/3BioTEG/	58.5	34.4	18.7	8.3
SD55	CTCTCTCTCTCCTCTCTCTCTCCGGGCGCCCAAATAAGAACATTAT/3BioTEG/	78.2	43.8	20.9	7.2
SD56	CTCTCTCTCTCCTCTCTCTCTCCGGGCGCCCAAATAAGAACATTATG/3BioTEG/	81.5	47.0	21.9	6.6
SD57	CTCTCTCTCTCCTCTCTCTCTCCGGGCGCCCAAATAAGAACATTATGA/3BioTEG/	84.7	50.2	25.0	7.6
SD58	CTCTCTCTCTCCTCTCTCTCTCGGCGCCCAAATAAGAACATTATGAT/3BioTEG/	71.7	42.1	21.7	
SD59	CTCTCTCTCTCCTCTCTCTCTCGCGCCCAAATAAGAACATTATGATC/3BioTEG/	67.7	42.0	22.9	9.5
SD60	CTCTCTCTCTCCTCTCTCTCTCGCCCAAATAAGAACATTATGATCA/3BioTEG/	59.7	34.2	19.1	8.6
SD61	CTCTCTCTCTCCTCTCTCTCTCGCCCAAATAAGAACATTATGATCAG/3BioTEG/	65.6	37.0	18.7	9.6
SD62	CTCTCTCTCTCCTCTCTCTCTCCCAAATAAGAACATTATGATCAGT/3BioTEG/	66.5	39.8	21.3	9.2
SD63	CTCTCTCTCTCCTCTCTCTCTCCCAAATAAGAACATTATGATCAGTA/3BioTEG/	63.8	36.7	19.3	6.5
SD64	CTCTCTCTCTCCTCTCTCTCTCCCAAATAAGAACATTATGATCAGTAG/3BioTEG/	70.6	38.0	17.4	6.1

10

【0277】

表3に示される鎖の電流レベルは、第1の例に記載されるアプローチに類似するアプローチを使用して得られた。TSO1鎖は内部対照としてチャンバーに添加され、電流レベルはこの対照に対して較正された。この実験で使用される方法と第1の例で使用された方法の間には2つの主な違いが存在していた。第1の違いはナノポアがMS-(B1-L88N)gミュータントに変えられていることであった。第2の違いは、適用された電圧スキームであった。これは、電流が4つの異なる付加電位で順次記録されるように選択された。ナノポアがDNAを捕捉する速度は付加電位に依拠しているので、最も大きな電位が最初に記録された。選択された電圧スキームは、+180mV(2.2秒)、+140mV(0.4秒)、+100mV(0.4秒)、+60mV(0.4秒)、-180mV(0.8秒)であった。

20

【0278】

図41は、下のトレースでは印加電圧の例を、上のトレースでは同じ時間スケールにわたるSDO1鎖について得られ測定されたイオン電流を示している。図41のこの例に見られるように、結合事象は+180mVの初期間中に起こり、イオン電流の降下を生じる。それに続く期間で電位が下がるに従って、観測されるイオン電流は減少する。最終期間は、逆になった電圧はDNA鎖を排出する。

30

【0279】

DNA鎖SDO1~SD54のすべてについて類似するパターンが観測され、それぞれの電圧での測定されたイオン電流レベルは表3に収載されている。

【0280】

このデータのグラフィック表示を提供するため、図42から45は、それぞれ4つのレベルの電圧で、順次水平方向に表示されたDNA鎖ごとの測定された電流の散布図である。見て取れるように、散布図の形状は電位が変わるに従って変化する。それは、異なる電圧での測定が、例えば、別の電圧では分解することができない2つの状態を分解する1つの電圧での測定により追加の情報を提供することになることを暗示している。

40

【0281】

同じデータの別の表示を与えるため、図46は印加電圧に対するそれぞれの鎖の測定された電流のグラフである。前記データは、それぞれの電圧での鎖ごとの点からなり、鎖ごとの点はグラフでは線で繋がれて鎖ごとの傾向を示している。図46におけるこの表示は変動の2つの主要な特長を図示している。

【0282】

50

第1の特長は、電圧が増大するに従って全体では異なる鎖についての測定された電流の広がりが増大することである。この全体の傾向は一般的に興味深い。これは、電圧の最適の選択に影響を与えると考えられるが、状態間の分離および個々の状態の測定の標準偏差にも依拠している状態間の分解の変化を示している可能性がある。しかし、全体の傾向は複数の電圧を使用する有益性を実証するものではない。

【0283】

第2の特長は、個々の鎖についての測定された電流が印加電圧への異なる依存度での挙動を示していることである。したがって、全体の傾向が電圧の増大に従って分岐することであるが、すべての鎖ごとの電流測定は同じ傾向を示してはいない。鎖についての測定は相互に分岐していないが、代わりに個々の鎖に変動がある。それどころか、一部の鎖は電圧と共に一般的線形変化を示しているが、他の鎖は非線形または振動性の変化を示し、いくつかのケースでは変曲点がある。全体的な分岐傾向に対して、一部の鎖に関する線は収束している。この観測の理由は重大ではないが、これは、異なる電圧の印加下での測定システムの物理的および／または生物学的変化により、おそらくナノポア中のDNAの立体構造変化により引き起こされると推測される。

【0284】

この第2の特長は、1つよりも多い電圧での測定が単に重複しているというよりはむしろ追加の情報を提供することである。異なる電圧でのイオン電流測定により異なる状態の分解が可能になる。例えば、1つの電圧では分解できないいくつかの状態を別の電圧では分解することができる。

【0285】

第2の例でのいくつかの追加の観測により、状態の標準偏差（または分散）に対する電圧を変えることの効果が調べられる。これらの状態の分散は、電流の分散がDNA鎖の制御された動き（例えば、酵素制御されたDNA移行）に類似する時間尺度であるときには問題を引き起こすことがある。このレジメでは、電流レベルの変化がそれぞれの状態内の分散またはDNAのネット運動（net movement）のためあるかどうかを決めるのは困難になる。この理由で、第2の例で収集されたデータは、移行を制御する酵素を使用するのではなく、ストレプトアビジンによりナノポア上に保持された鎖を使用して収集された。したがって、電流変化が鎖の動きから生じたのかまたはその電流状態の固有の特性から生じたのかどうかを描写するように電流レベルの分散を変化させることができるシステムがあるのが望ましい。

【0286】

状態分散に対する付加電位の効果を評価するため、第2の例の結果は解析されて、表3のDNA配列ごとに平均標準偏差を導き出した。図47は、印加電圧に対するそれぞれの鎖の標準偏差のグラフである。データは、それぞれの電圧での鎖ごとの点からなり、鎖ごとの点はグラフでは線で繋がれて鎖ごとの傾向を示している。電流レベルの分散は付加電位と共に確かに変化することは、図47から明らかである。大多数の鎖では、分散は付加電位の増大と共に増加するが+180 mVから+220 mVまで急上昇する。この変化は上記の電圧に合わせた電流の変動に類似する原因を有すると推測される。

【0287】

1つよりも多い電圧でイオン電流測定を行う方法であって、本発明の第二の態様および第三の態様を具体化する方法は、図48に図示されている。この方法では、付加電位はDNAがナノポア中を通して動いている間変調される。

【0288】

ステップS6では、ポリマーはナノポアを横断する電圧の印加下でナノポア内を通して移行される。

【0289】

ステップS7では、移行中、電圧のレベルは周期的に変化される。周期は2つ以上の電圧レベルを含みうる。電圧レベルは規則的にまたは不規則に繰り返してもよい。この期間を含めて、周期は個々の観測される状態、すなわち、測定される電流が異なるkマーに依

10

20

30

40

50

拠るようにポリマーが異なる位置である状態よりも短くなるように選択される。したがって、それぞれの状態中、電圧のレベルが同じである時、例えば、繰り返される周期で、ナノポア中を流れるイオン電流は同じであることが観察される。言い換えると、イオン電流は印加電圧と共に循環する。

【 0 2 9 0 】

ステップS8では、異なる電圧レベルの適用下でのナノポア中を流れるイオン電流はそれぞれの状態ごとに測定される。

【 0 2 9 1 】

第3の例は、この方法の例が以下の通りに実施された。分析物DNA鎖は、上記第2の例におけるストレプトアビジン系で特徴付けられていた配列を含有するように選択された。分析物DNA鎖は、ナノポア中に通させる5'オーバーハングでの低二次構造配列も含有していた。相補鎖は分析物鎖にハイブリダイズされた。前記相補鎖は、コレステロール-TEGリンカーを含有する短いオリゴがハイブリダイズされる短い5'オーバーハングも含有していた。コレステロールの組込みにより、DNAは二重層に繋ぎ止められ、必要なDNAの濃度が大幅に減少する。表4は、この例で使用される分析物DNA鎖の配列を収載している。

【 0 2 9 2 】

【表 4】

表 4 :

鎖	配列 (5'-3')
I198	TTTTTTTTTTTTTTTTTTTTTCCCCCCCCCCCCCAAATAAGAACATTATGATCAGT AGGAGCACTACGACCTTTGTTCTGGTGCTCGTCCGGGCGCCCAAAGTGAGCGA GTGCGAGAGGCGAGCGGTCAA
I305	GTATCTCCATCGCTGTTGACCGCTCGCCTCTCGCACTCGCTCCACTTTGGGCGCC CGGACGAGCACCAGAACAAAGGTCGTAGTGCTCCTACTGATCATAATGTTCTTA TTT
TE07	CAGCGATGGAGATAC-CholTEG

【 0 2 9 3 】

実験設定は上記に類似しており、溶液は、400mM KCl、10mM HEPES、pH8.0、1mM EDTA、1mM DTTを含有していた。バッファーはチャンバー内で予備混合溶液の一部として使用された。表4.1において使用されるDNAは1対1対1比でハイブリダイズされ、予備混合溶液に添加され、Phi DNA Pも添加され、予備混合溶液は室温で5分間混合させておいた。単一のMS-(B1-L88N)₈チャンネルが得られ、予備混合物を添加して、0.5nMの最終溶液DNA濃度および100nMの最終溶液Phi29 DNA P濃度が得られた。

【 0 2 9 4 】

印加電圧は、それぞれ10 msの長さの+180 mVと+140 mVの交互パルスを含む周期で適用された。

【 0 2 9 5 】

図4-9は、結果の図解部分を示しており、特に、下トレースでは印加電圧、および上トレースでは得られ測定されたイオン電流を示している。事象はP h i 2 9 DNA P-DNA複合体から見られた。図4-9では、付加電位の両方で、状態、例えば、標識された状態1から3を観測することができた。それぞれの状態中、連続する周期においてそれぞれの電圧レベルで流れるイオン電流は同じである。それぞれの状態で、+140 mVおよび+180 mVの付加電位での電流レベルは、鎖が一貫した位置にあり、ポア中の単一分子上2つの電圧で読みを与える間順次得られ、これは状態の期間よりも短い周期期間により達成される。容量性遷移は付加電位が変えられる直後に観察することができる。これは、

脂質二重層上に蓄えられた電荷が変化する時に起きる。この容量性遷移の持続時間は脂質膜のサイズに依拠し、さらに小さな膜サイズに進むことにより減らすことができる。この実験では、脂質膜は、直径50 μm の開口部にわたって浮遊させた。

【0296】

DNAは付加電位下Phi 29 DNA Pの中を通過して引っ張られるので、鎖が1つの位置から別の位置に動くときに起こる状態間の遷移を観察することも可能である。前記遷移により付加電位ごとに観察される電流が変化する。

【0297】

図49の例は、状態2と隣接する状態1および3における測定されるイオン電流間の違いは+140 mVの印加電圧よりも+180 mVの印加電圧のほうがはるかに大きいという点で、複数の電圧を使用する利点も図示している。これにより、+140 mVの印加電圧よりも+180 mVの印加電圧で状態1と3から状態2を分解するほうが容易になる。逆に、+180 mVの印加電圧よりも+140 mVの印加電圧で他の状態を分解するほうが容易である。

10

【0298】

図50は、図49と同じ種類のプロットにおいて、第3の例で記載された条件に類似する条件下で、しかし、MS-(B1-L88N)₈の代わりにMS-(B1)₈ポアを使用して得られた結果の別の図解部分を図示している。図50は図49に類似する全体的な形を有し、今回は、状態1から状態4と名付けられた4つの状態を含む。このケースでは、+140 mVの印加電圧での状態2と隣接する状態3の測定されたイオン電流間にはほとんど違いはないが、+140 mVの印加電圧では大きな違いが存在する。このケースでは、+140 mVでは状態2を状態3から分解するのは困難であるまたは不可能でさえあるが、これは+180 mVでは可能になる。再び、+180 mVの印加電圧よりも+140 mVの印加電圧で他の状態を分解するほうが容易である。

20

【0299】

上で実証され考察された複数レベルの印加電圧を使用して得られた追加の情報は、ポリマーについての情報を導き出すために測定されたイオン電流が解析される時には利点を提供する。

【0300】

測定を解析する1つの方法は、本発明の第一の態様に従って方法、例えば、第一の態様を（図6およびそれに続く図を参照して）具体化する上記の方法を適用することである。したがって、本明細書に記載される方法の様々な特長はいかなる組合せでも組み合わせうる。このケースでは、複数の電圧を使用することにより得られる追加の情報は推定の正確度を改善する。

30

【0301】

本発明の第一の態様に従った解析法は、ポリマーの少なくとも一部の配列を、したがって、正体を決定する。しかし、第二の態様および第三の態様に従った方法も、ポリマーの少なくとも一部の正体を決定する測定を解析する他の方法において利点を提供し、その利点のいくつかの非限定的例は以下の通りである。

【0302】

測定を解析して、本発明の第一の態様に従った技法以外の技法を使用してポリマーの少なくとも一部のポリマー単位の配列を推定しうる。

40

【0303】

測定を解析して、ポリマー単位の配列の完全な推定を提供せずにポリマーの少なくとも一部の正体を推定しうる。これらの種類の解析では、複数の電圧を使用することにより得られる追加の情報は推定の正確度を改善する。

【0304】

代わりに、測定を解析して、状態間の遷移のタイミングを導き出しうる。これらのタイミングはそれ自体が価値があり、またはさらなる解析において使用して、例えば、ポリマー単位の正体を決定しうる。この種類の解析では、追加の情報は遷移を検出する能力を改

50

善する。一部の遷移は1つの電位で観測するほうが容易であり、他の遷移はもう一方の電位で観測するほうが容易である。例として、図50の図解的結果では、状態2から状態3への遷移は+140 mVで観測するのは困難であるが、+180 mVでは容易に観測される。これとは対照的に、状態3から状態4への遷移は+180 mVでは弱い、+140 mVでは容易に観測される。したがって、1つよりも多い電位で記録するのには状態検出に対する利点が明らかに存在する。

【0305】

いくつかの解析方法では、異なるレベルでの測定は両方とも、例えば、ポリマーの少なくとも一部の正体の決定に両方とも同じように寄与する別々の測定として直接使用される。他の解析方法では、異なるレベルでの測定、例えば、正体を決定するのに使用される1つのレベルで行われる測定およびその結果を確認するのに使用される異なるレベルで行われる測定は異なるやり方で使用しうる。代わりに、1つのレベルでのノイズは、1つの電圧での特定の測定を使用することを決定するため、別のレベルでのノイズと比較されることがある。代わりに、解析方法は、それぞれのkマーについての異なるレベルでの測定間の選択とそれに続くポリマーの少なくとも一部の正体を決定するための選択された測定の使用を含んでいてもよい。

【0306】

異なるレベルでの2つの測定の使用により得られる追加の情報の程度はkマー間で変わることがある。そのケースでは、異なる数のレベルでの測定は異なるkマーについて使用され、例えば、いくつかのkマーについて、減少した数のレベル、おそらく、1つだけのレベルでの測定を使用し、一方、他のkマーについてさらに多くのレベルでの測定を使用することがある。この方法は、高分散状態にまたは類似の電流レベルを有するそれぞれの状態に特に有用でありうる。

【0307】

異なるレベルでの測定が使用される場合、異なる重み付けが異なる測定に付けられることがある。

【0308】

それでもなお、解析方法が様々な形で測定を使用しうるという事実にもかかわらず、いくつかのkマーに関する異なるレベルでの測定はある方法で使用される。

【0309】

本発明に従った2つの非限定的例がここで説明される。これらの例は両方とも、それぞれの電位での状態あたり典型的には少なくとも1つの測定があるケースに適用される。

【0310】

最初の例では、複数のレベルでの測定が使用されて、状態遷移を決定する。これは、状態遷移がある電位では観測可能であるが別の電位では可能ではないことがあるという事実を利用する。測定は、状態からの遷移の可能性が高い、状態検出ステップS1を含む上記の解析方法を受けることができる。図50では、例えば、状態についてのそれぞれの電位での全データの平均をとることにより、トレースはそれぞれ140および180 mVでの2つの測定に減らしてもよい。次に、これらの測定は2セットの放出分布から同時発生（すなわち、密に結合した次元）として扱われ、1Dケースに類似するセットの遷移で解析されてもよい。これは、我々が単一の電位で状態を1回よりも多く測定するケース、例えば、平均と分散に実行が類似している点に注目されたい。実際、我々は、例えば、それぞれの電位での平均と分散を考慮することによりこのアプローチを4つの密に結合された次元まで広げることがある。

【0311】

第2の例では、状態間の遷移は、ステップS1が省かれる上記ケースに似て、別々のステップとしてというよりもむしろ解析段階中に推定される。この例では、単純にするため、我々は、電位周期のそれぞれのステップでの測定のシリーズを単回測定まで減らしたケース、例えば、平均を考慮することになる。再び、図50を参照して、状態1は140から180 mVに交互に代わる28の測定からなる。したがって、測定ごとの放出確率は適切

な放出（140 mVまたは180 mV）およびこのデータに適している遷移に関して計算される。例えば、この状態からのおよそ0.05の全遷移確率が適切でありうる。このアプローチは、それぞれの周期からのサマリー測定またはそれぞれの周期からの複数のサマリー測定よりはむしろそれぞれの測定を考慮するように一般化されてもよい。

【0312】

本発明の第二の態様に従って異なる電圧で測定を行う方法では、ポリマーがナノポア中を通して移行される間付加電位が循環される本発明の第三の態様に従った方法を適用するのが有利であるが、代わりに他の方法を使用しうる。

【0313】

非限定的例として、本発明の第二の態様に従って1つよりも多い電圧でイオン電流測定を行う1つの別の方法は、図51に示されており以下の通りに実施される。

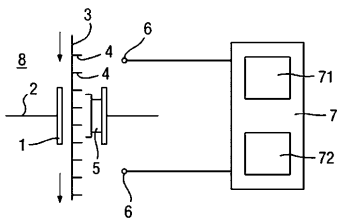
【0314】

ステップS9では、ポリマーはナノポア中を通して移行され、ステップS10では、移行中単一レベルの電圧がナノポアを横断して印加されそのレベルの電圧の印加下でナノポア中を流れるイオン電流が観測されるそれぞれの状態ごとに測定される。次に、前記方法は同じポリマーを移行させるステップS9およびステップS10を繰り返すが、異なるレベルの電圧を印加する。ステップS9およびS10は、いかなる数の回数でも繰り返して、いかなる数の電圧レベルでイオン電流測定を得てもよい。

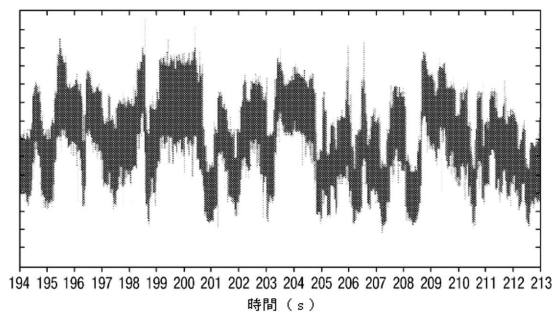
【0315】

望ましいのは、毎回同じポリヌクレオチドを読み取るために、ナノポアを離れるポリマーの能力は制限される。ポリヌクレオチドのケースでは、これは、鎖が離れていかないように電位を制御することにより、または鎖の移行を阻害する、ストレプトアビジンなどの化学的または生化学的ブロック剤を使用することにより実行しうる。

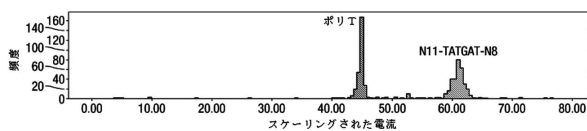
【図1】



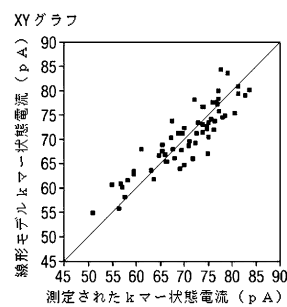
【図2】



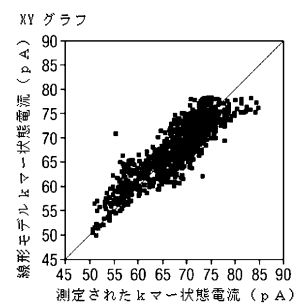
【図3】



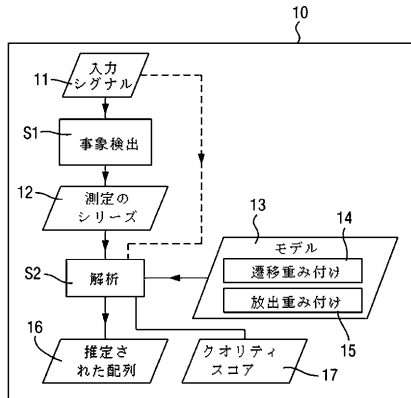
【図4】



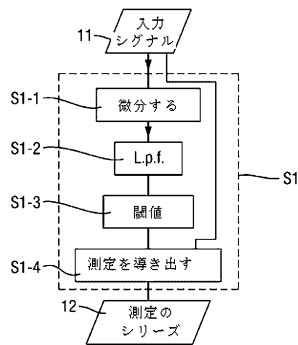
【図5】



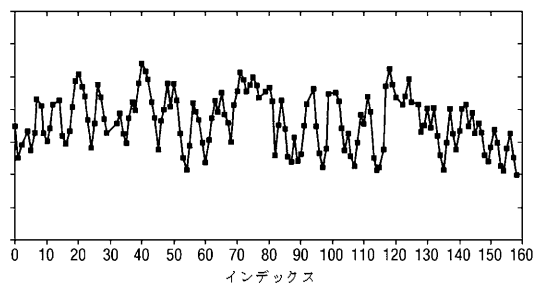
【図 6】



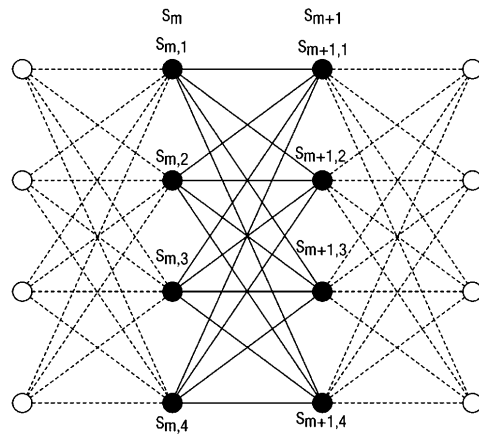
【图 7】



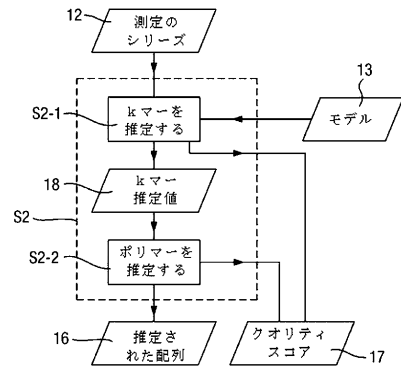
【図 10】



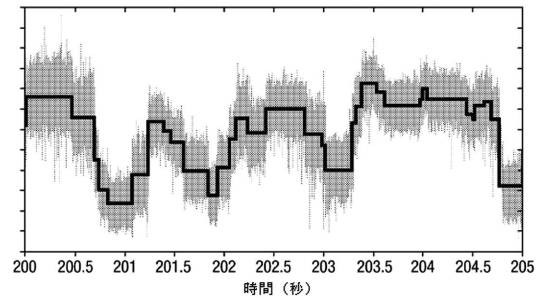
【図 1 1】



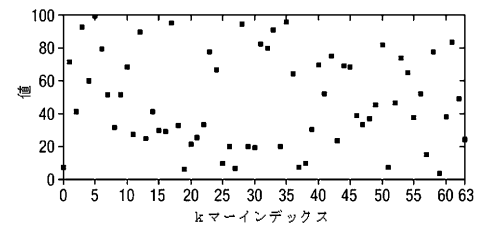
【图 8】



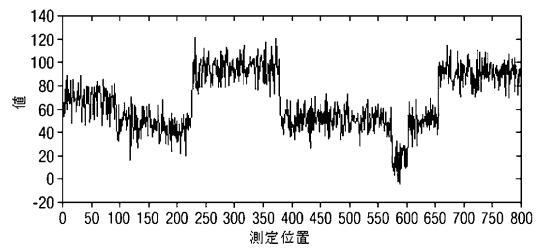
【图 9】



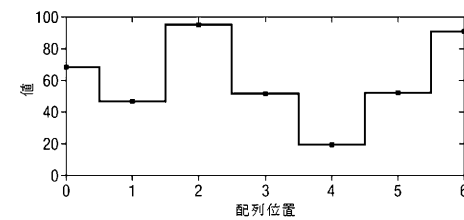
【图 1 2】



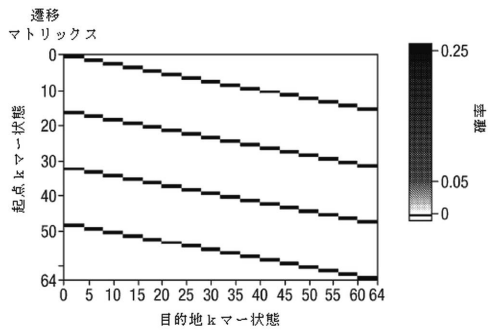
【図 1 3】



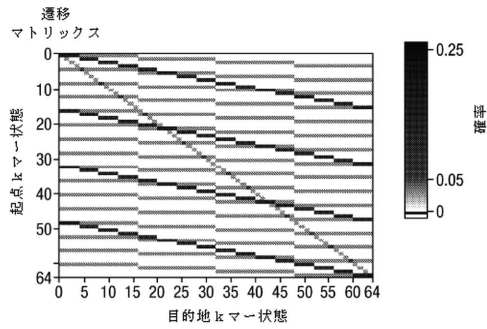
【图 14】



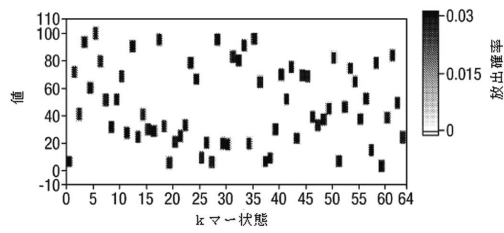
【図 15】



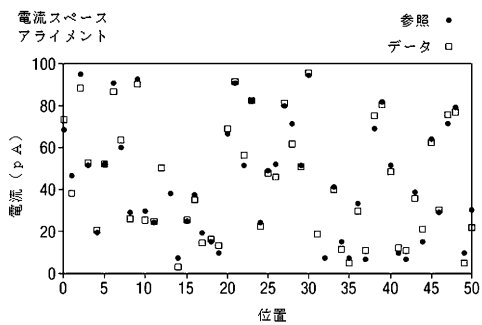
【図 16】



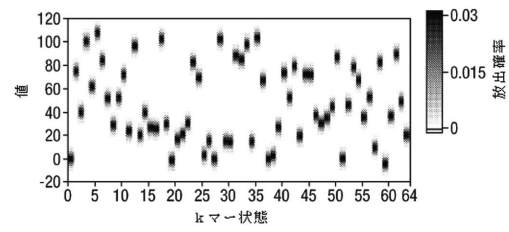
【図 19】



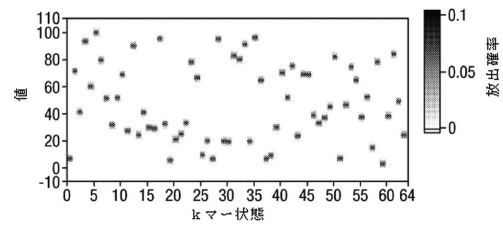
【図 20】



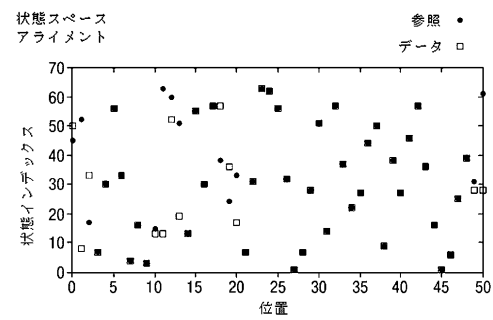
【図 17】



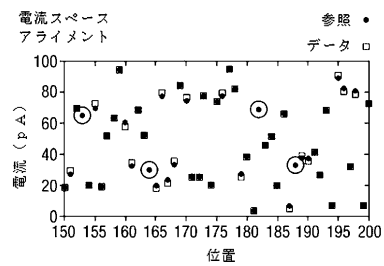
【図 18】



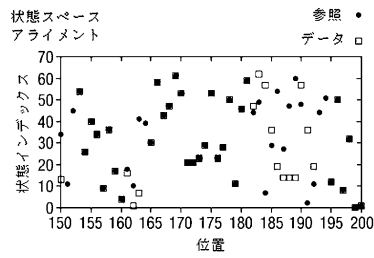
【図 21】



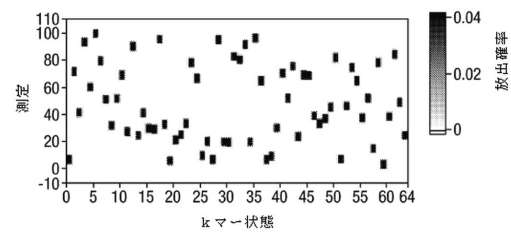
【図 22】



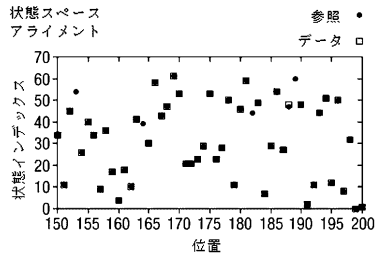
【図 2 3】



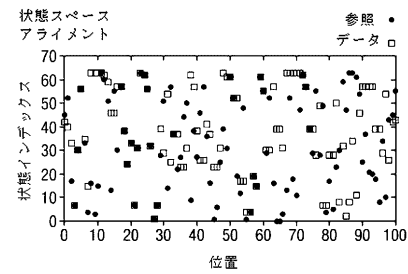
【図 2 5】



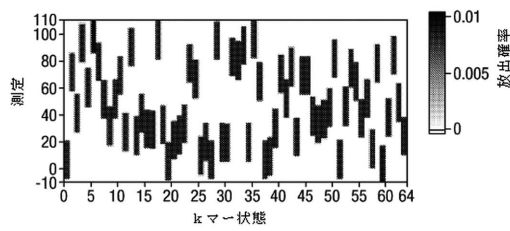
【図 2 4】



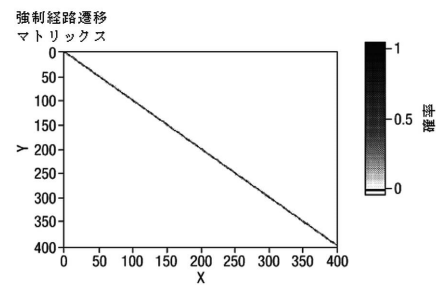
【図 2 6】



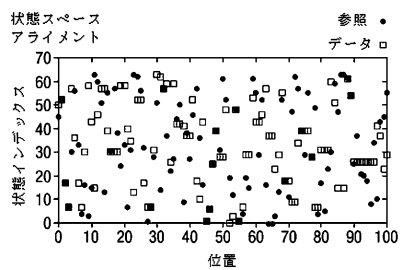
【図 2 7】



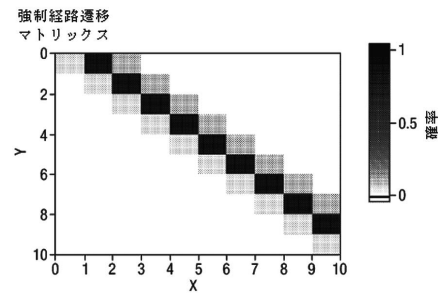
【図 3 0】



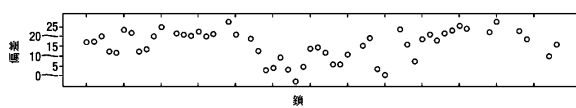
【図 2 8】



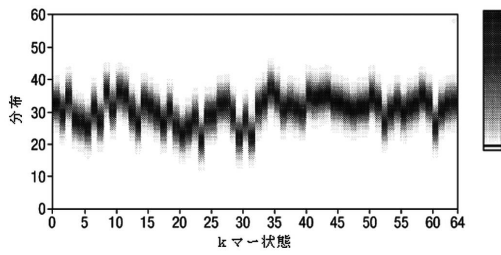
【図 3 1】



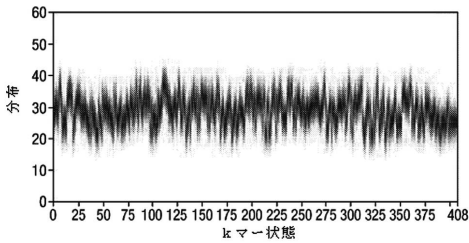
【図 2 9】



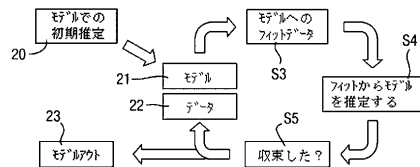
【図 3 2】



【図 3 3】



【図 3 4】



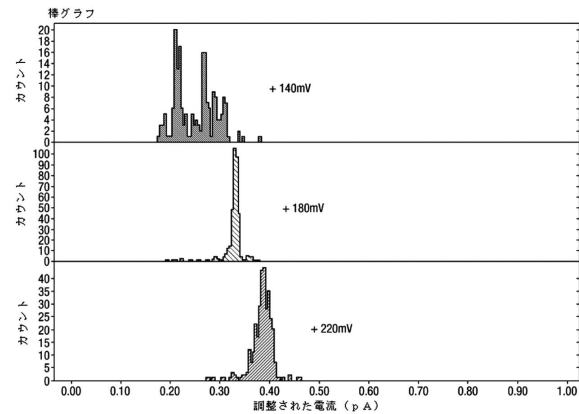
【図 3 8】

[illegible]

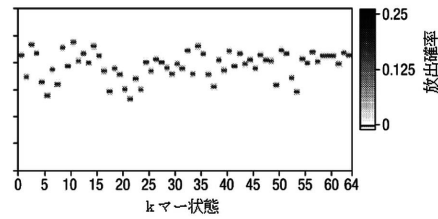
【図 39】

[illegible]

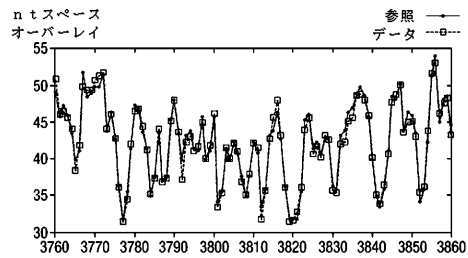
【図 40】



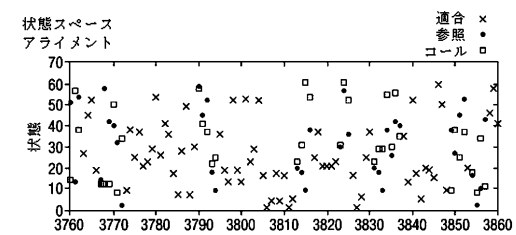
【図 3 5】



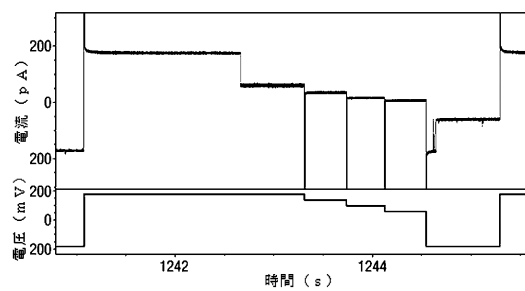
【図 3 6】



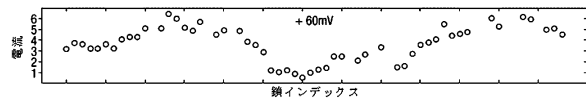
【図 3 7】



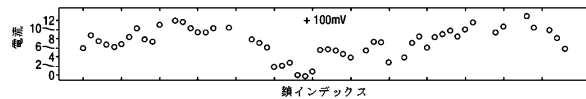
【図 4 1】



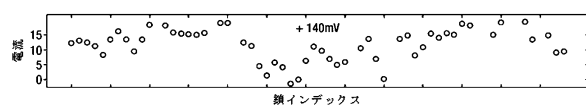
【図 4 2】



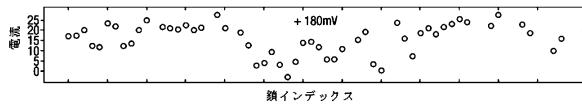
【図 4 3】



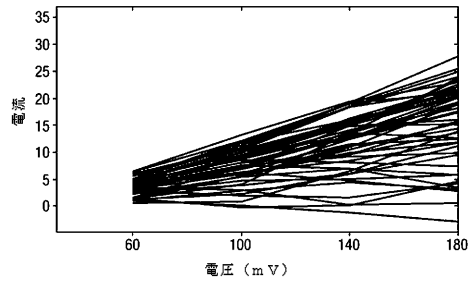
【図 4 4】



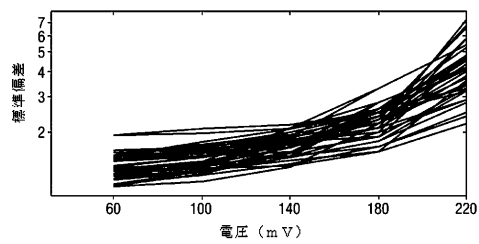
【図 4 5】



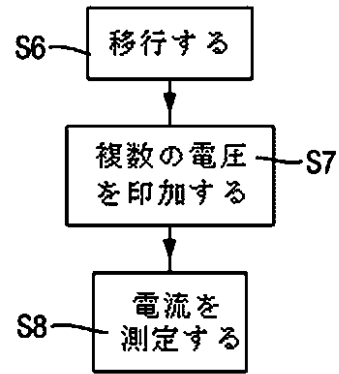
【図 4 6】



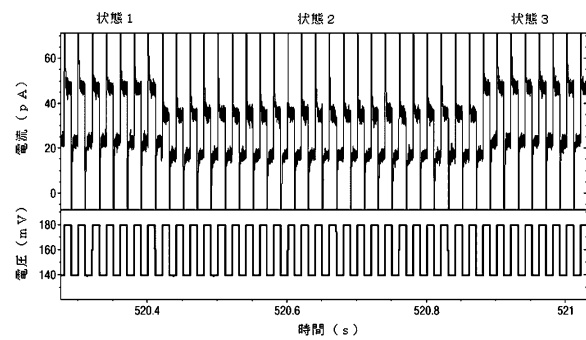
【図 4 7】



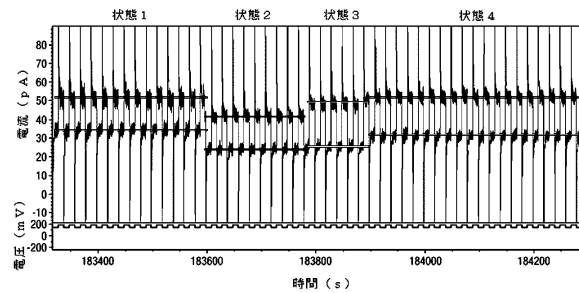
【図 4 8】



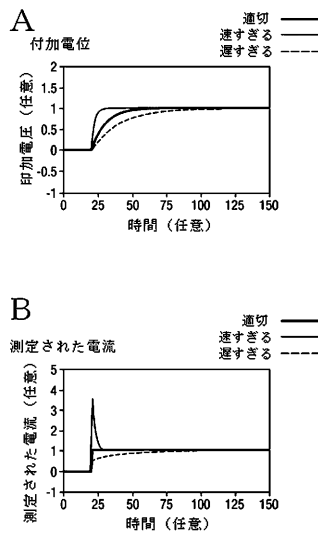
【図 4 9】



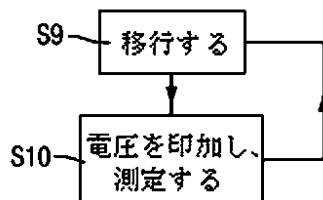
【図 5 0】



【図 5 2】



【図 5 1】



【配列表】

0006457811000001.app

フロントページの続き

(51)Int.Cl. F I
C 1 2 N 15/09 (2006.01) C 1 2 N 15/09 Z

- (72)発明者 レイド, スチュアート ウィリアム
イギリス国 オックスフォードシャー オーエックス4 4ジーエー, オックスフォード, オックス
フォード サイエンス パーク, ロバート ロビンソン アベニュー 4, エドモンド カート
ライト ハウス, オックスフォード ナノポール テクノロジーズ リミテッド
- (72)発明者 ハーパー, ギャヴィン
イギリス国 オックスフォードシャー オーエックス4 4ジーエー, オックスフォード, オックス
フォード サイエンス パーク, ロバート ロビンソン アベニュー 4, エドモンド カート
ライト ハウス, オックスフォード ナノポール テクノロジーズ リミテッド
- (72)発明者 ブラウン, クライヴ ギャヴィン
イギリス国 オックスフォードシャー オーエックス4 4ジーエー, オックスフォード, オックス
フォード サイエンス パーク, ロバート ロビンソン アベニュー 4, エドモンド カート
ライト ハウス, オックスフォード ナノポール テクノロジーズ リミテッド
- (72)発明者 クラーク, ジェームス アンソニー
イギリス国 オックスフォードシャー オーエックス4 4ジーエー, オックスフォード, オックス
フォード サイエンス パーク, ロバート ロビンソン アベニュー 4, エドモンド カート
ライト ハウス, オックスフォード ナノポール テクノロジーズ リミテッド
- (72)発明者 ヘロン, アンドリュー ジョン
イギリス国 オックスフォードシャー オーエックス4 4ジーエー, オックスフォード, オックス
フォード サイエンス パーク, ロバート ロビンソン アベニュー 4, エドモンド カート
ライト ハウス, オックスフォード ナノポール テクノロジーズ リミテッド

審査官 植原 克典

- (56)参考文献 米国特許出願公開第2011/0226623 (US, A1)
米国特許出願公開第2006/0019259 (US, A1)
国際公開第2008/092760 (WO, A1)
特表2015-509710 (JP, A)
John J. Kasianowicz, Nanoscopic Porous Sensors, Annual Review of Analytical Chemistry
, 2008年, Vol.1, No.1, Pages 737-766
Petros Boufounos, Basecalling using hidden Markov models, Journal of the Franklin Insti
tute, 2004年, Vol.341, Pages 23-36
LIANG KUO-CHIHG, BAYESIAN BASECALLING FOR DNA SEQUENCE ANALYSIS USING HIDDEN MARKOV MO
DELS, PROCEEDINGS OF 2006 IEEE CONFERENCE ON INFORMATION SCIENCES AND SYSTEMS, CISS 20
06, 2006年 3月22日, P1599-1604
Motoyuki Suzuki, A New HMnet Construction Algorithm Requiring No Contextual Factors, I
EICE Trans. Inf. & Syst., 1995年, Vol.E78-D, No.6, Pages 662-668
Jun-ichi Takami, Automatic Generation of Hidden Markov Networks by a Successive State
Splitting Algorithm, 電子情報通信学会論文誌, 1993年, Vol.J76-D-II, No.10, Pages 21
55-2164
Nanomedicine, 2007年, Vol.2, No.4, pp.459-481

(58)調査した分野(Int.Cl., DB名)

C 1 2 Q 1/6869

C 1 2 N 15/00-15/90

C A p l u s / M E D L I N E / E M B A S E / B I O S I S (S T N)