

(12) 特許協力条約に基づいて公開された国際出願

(19) 世界知的所有権機関
国際事務局



(43) 国際公開日
2011 年 5 月 26 日(26.05.2011)

(10) 国際公開番号
WO 2011/062166 A1

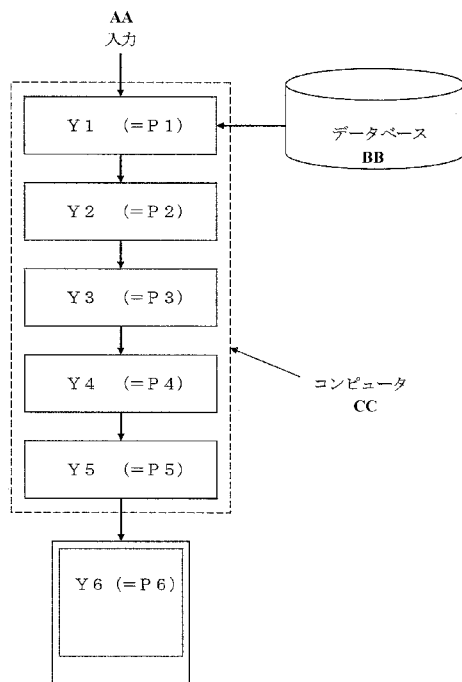
- (51) 国際特許分類:
G06F 19/16 (2011.01) *C12N 15/09* (2006.01)
- (21) 国際出願番号: PCT/JP2010/070400
- (22) 国際出願日: 2010 年 11 月 16 日(16.11.2010)
- (25) 国際出願の言語: 日本語
- (26) 国際公開の言語: 日本語
- (30) 優先権データ:
特願 2009-262258 2009 年 11 月 17 日(17.11.2009) JP
- (71) 出願人 (米国を除く全ての指定国について): 武田薬品工業株式会社(TAKEDA PHARMACEUTICAL COMPANY LIMITED) [JP/JP]; 〒5410045 大阪府大阪市中央区道修町四丁目 1 番 1 号 Osaka (JP).
- (72) 発明者; および
- (75) 発明者/出願人 (米国についてののみ): 中村 慎吾 (NAKAMURA, Shingo) [JP/JP]; 〒3004293 茨城県つくば市和台 1 0 番地 武田薬品工業株式会社内 Ibaraki (JP).
- (74) 代理人: 高島 一 (TAKASHIMA, Hajime); 〒5410044 大阪府大阪市中央区伏見町四丁目 1 番 1 号 明治安田生命大阪御堂筋ビル Osaka (JP).
- (81) 指定国 (表示のない限り、全ての種類の国内保護が可能): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.
- (84) 指定国 (表示のない限り、全ての種類の広域保護が可能): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), ユーラシア (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), ヨーロッパ (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI

[続葉有]

(54) Title: METHOD FOR ESTIMATING DYNAMIC STATE OF SECONDARY STRUCTURE OF RNA

(54) 発明の名称: RNAの二次構造の動態を予測する方法

【図1】



AA INPUT
BB DATABASE
CC COMPUTER

(57) Abstract: Provided are a method whereby the dynamic state of a secondary structure of an RNA in a non-equilibrium state can be estimated, a program for executing said method by a computer, and a device for carrying out said method. The method comprises identifying the secondary structures of an RNA in an equilibrium state by using a publicly known method for estimating secondary structure, representing inter-conversion among these secondary structures by simultaneous differential equations, and doing the simultaneous differential equations under appropriate initial conditions to thereby determine the probability at which each of the aforesaid secondary structure occurs at each point in the non-equilibrium state.

(57) 要約: 本発明は、非平衡状態でのRNAの二次構造の動態を予測することを可能とする方法、該方法をコンピュータに実行させるためのプログラム、および該方法を実施するための装置を提供することを課題とする。上記課題は、平衡状態でのRNAの二次構造を従来公知の二次構造予測方法を用いて決定し、それらの二次構造同士の相互変換を連立微分方程式で記述し、かつ該連立微分方程式を適当な初期条件の下で解くことにより、前記各二次構造が、非平衡状態にある各時点において何程の確率で存在し得るかを決定することを含む方法により解決される。

WO 2011/062166 A1



(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG). 添付公開書類:

— 国際調査報告 (条約第 21 条(3))

明 細 書

発明の名称 : R N Aの二次構造の動態を予測する方法

技術分野

[0001] 本発明は、R N Aの二次構造が平衡状態となる前の時点、すなわち、該二次構造がマクスウェル・ボルツマンの分布則に従った存在確率の分布を示すよりも前の時点における、該R N Aの二次構造の動態を予測する方法、ならびにそのためのプログラムおよび装置に関する。

[0002] (発明の背景)

近年、アンチセンスオリゴヌクレオチドやs i R N Aを用いて、m R N Aが創薬のターゲットになり得るという認識が広がり、ターゲット分子としてのm R N Aについて構造生物学的な研究が徐々に注目を集めつつある。

構造生物学は、生体内の分子の特定の構造とその構造に起因する機能を研究することで発展してきた。この分野では、タンパク質のみならず、機能を持つR N Aであるt R N Aやr R N Aの構造について、従来からよく研究が進んでいる。しかしながら、遺伝子であるD N Aの遺伝情報をタンパク質に翻訳するための「伝令」としての役割としての認識が強かったm R N Aについては、これまでその高次構造についてはほとんど注目されてこなかった。これは、m R N Aが遺伝情報を伝達する働き以外に、注目すべき機能を持たないと考えられていたからである（機能を持たないのであれば、それを引き起こす構造もないであろうと漠然と推察されていた）。

m R N Aは、細胞内のイベントに応じてその全体として形態（一次・二次・三次構造）を大きく変える。核内でm R N Aが合成されたとき、スプライシングを受けているとき、核外に運搬されたとき、タンパク質の翻訳が行われているとき、さらにその後など、m R N Aの全体的な構造は一意には論じることができない（本明細書では、核外のm R N Aの挙動について述べるので、以後m R N Aの一次構造の変化については注目しない）。

一方R N Aは自己相互作用しやすい高分子であるため、溶液中で何らかの高次構造を取り得ると推定される。しかし、長いR N Aは最安定構造と準最

安定構造群とがエネルギー的に差異が少ないので、結果として溶液中の長いRNAは、複数の高次構造の混合物として存在しているはずである（すなわち、細胞質内において一意の構造をとることが保証されていない）。

こうした二つの要因から、mRNAの全体の高次構造を研究するための方法が存在せず、ひいてはこのためmRNA構造が生命現象に影響を与えていると意識されることが少なかったのである。

[0003] ところが、昨今の研究から、mRNA上の局所的な二次構造が、RNAの編集、riboswitch、IRES、フレームシフトなどの、分子生物学的イベントに関与していることが次第にわかってきた。RNAiやアンチセンス医薬品の結合部位による効率の差、miRNAの結合部位にも、そうした局所的な二次構造の関与が示唆され得る。つまり、全体の構造を把握不可能なmRNAの高次構造であっても、対象をある細胞内イベントに対応したmRNAに限定すると、その局所的高次構造については生物学的に機能があることがわかってきた。このように機能を持つ構造体に対しては、医薬品の開発ができることが示唆され、事実、例えば、ピリチアミンはTPP riboswitchに対して作用することで抗生物質として働いていることが近年分かってきた。但し、その発見と解析の難しさから、このように「構造を持つ・機能を持つ」ことが報告されているmRNAはごく少数であり、したがって創薬ターゲットとなり得るmRNAの二次構造は非常に限定されていた。

[0004] 本発明者は以前、〔自己相互作用しやすいというRNA分子の特性に基づけば、あらゆるmRNAは確率的に何らかの高次構造をとることがあり、それが生体内で特別な機能を持たないかまたは機能を発揮し得るほど安定でないがために（或いは単純にその機能が未知であるがために）、これまで構造として認識されてこなかっただけである〕との発想の下で、mRNA分子内の局所的な二次構造を、コンピュータを用いてcDNAデータベースから網羅的に検索し得るソフトウェアを開発した（特許文献1）。さらに、該検索により抽出された局所的な二次構造が、生体内で比較的安定に存在し得る（すなわち、ある物質（例えば、低分子化合物）との特異的結合により容易に

安定化される) ことを、インビトロの実験系を用いて実証した。

本発明者はさらに、対象とするRNAが、抽出された局所的な二次構造を、実際に生体内でとるかどうかを精度よく予測し得る方法を開発した(特許文献2)。該方法は、mRNAのような長い配列の対象区間上のある位置に、ある短い長さのフレームを想定し、そのフレームで規定される小区間内の各特定位置において或る局所的な二次構造が何程の蓋然性で存在するか(第1の蓋然性)を求める、という操作を、フレームを微小長さ(フレーム長よりも十分短い長さ)だけ移動させながら各小区間で行った後、各小区間で判明した特定位置が、全ての小区間の間でどれだけ互いに関連しているかの指標、すなわち各特定位置自体がどの程度の信憑性を持つかを示す指標(第2の蓋然性)を求める、というものである。第1の蓋然性が高く、かつ、第2の蓋然性が高い特定位置は、その特定位置自体が該局所的な二次構造の位置である信憑性が高く、しかも、そこに該局所的な二次構造がより確かに存在するということになる。従って、該方法によって、目的とする局所的な二次構造が有意義な確実性で存在するかどうかを定量的に予測することが可能となった。

該方法は、mfold(GCG Software; 非特許文献1参照)などの従来の構造予測ソフトウェアが好ましい予測結果を与えない比較的長い区間(例、200塩基以上)に対して、対象区間がどれだけ長くても信頼性の高い予測結果を与えることができる。該方法によって、例えば、アンチセンス核酸、RNAi、低分子化合物などの、二次構造部分もしくはその近傍に結合して該構造を安定化することにより、mRNAからタンパク質への翻訳を抑制する(あるいはmRNAの分解安定性を改善して翻訳効率を増強する)化合物の開発・設計をより効率よく行うことが可能となった。

先行技術文献

特許文献

[0005] 特許文献1: 国際公開2006/054788号パンフレット

特許文献2: 国際公開2008/072713号パンフレット

非特許文献

[0006] 非特許文献1: Proc. Natl. Acad. Sci. USA, 86: 7706-10 (1989)

発明の概要

発明が解決しようとする課題

[0007] 上記特許文献2で開示された方法では、各小区間における局所的な二次構造の抽出および存在確率の算出は、各小区間について得られる予測構造群のうち、該局所的な二次構造が発見される予測構造にわたって、ギブス自由エネルギーから算出されるそれらの存在確率を総和することに基づいて行われる。これは、エネルギーのより低い二次構造の方が、その存在確率はより高い、ということを前提としている。ところでRNAは、生体内では、二次構造を形成するために十分な時間を与えられたときに、エネルギーの観点からより安定な二次構造として存在する可能性が高くなり、十分な時間経過後の各二次構造の存在確率の分布は、マクスウェル・ボルツマンの分布則に従うものとなると考えられる。それゆえ、上記前提はもつともであり、実際、上記特許文献2で開示された方法は、生物学的に妥当な結果を与え得ることが確認されてきた。さらに、局所的な二次構造の存在確率を論じるにあたり、それが発見される予測構造自体のエネルギーを用いることで、局所的な二次構造間の存在確率の大小を一般的に論じることができるようになった（それまで、長さや配列の違う局所的な二次構造間ではその存在確率の大小を論じることができなかった）。

[0008] しかしながら、本発明者によるその後の研究により、RNAの二次構造群の存在確率の分布を論じるにあたり、マクスウェル・ボルツマンの分布則に従うものとなるより前（すなわち、系が平衡状態に達する前）の時点の二次構造を予測することが重要であり得ることが明らかとなってきた。

すなわち、mRNAなどのRNAが構造を形成するために与えられた時間が非常に短かった場合に、いわゆる動力学的トラップ（kinetic trap）の状態（すなわち、熱力学的に安定な二次構造を取るまでの、速度論的に安定な構造を取る状態）となり、動力学的トラップに陥っていた構造は、さらに時

間を与えられたときに、カノニカルな塩基対相互作用だけを考慮した熱力学的に安定な状態（ここでは、「熱力学的トラップ（thermodynamic trap）の状態」と呼ぶ）となると考えられる。

このように、生体内でのRNAの高次構造は段階的に変遷し、それぞれの時間スケールとしては、シングルスランドのRNAの状態から開始して、動力学的トラップの状態については1秒以内、熱力学的トラップの状態については6～10秒程度の時間で達するものと非常におおまかに想定される。さらにその後、熱力学的トラップの構造間で平衡状態に達するであろう。上記特許文献2の方法は、熱力学的トラップの平衡状態での二次構造を暗に想定するものである。該方法が、上述したように、生物学的に妥当な結果を与え得ることは、生体内でのRNAの状態が、熱力学的トラップの平衡状態に少なくとも近似できることを示唆しているものと考えられる。

しかしながら、生体内でのRNA（特にmRNA）が構造を形成するために与えられた時間は、系が平衡状態に達するのに十分なものであると結論付ける理由がない。従って、生体内でのRNAの機能発現を検討する際に、平衡状態に達する前を想定し、とり得る二次構造として動力学的トラップの状態と熱力学的トラップの状態との間の二次構造を取り扱う方がより現実に近いと考えられる。ところが、従来、そのような時間的区間における二次構造を予測する方法はおろか、RNAの二次構造をそのような時間的区間において解析することの重要性の認識すら存在しなかった。

課題を解決するための手段

- [0009] 本発明者は、上記課題を解決するために鋭意検討した結果、構造が動力学的トラップから熱力学的トラップへ遷移する時間帯であっても、構造を組み始めるスタート時の影響を強く受ける時点を脱した時間帯を扱うのであれば、たとえ構造が遷移する時間帯においてであっても熱力学的トラップの状態のみを扱えばよいという考えに至った。その場合、動力学的トラップの状態から熱力学的トラップの状態への遷移はほぼ不可逆と見なすことができるから、熱力学的トラップの状態で存在する構造群が、動力学的トラップの状態

で存在する構造群を無視できるほど優勢に存在していると考えることができる。そのとき、RNAの二次構造を解析すべき時間的区間においては、熱力学的トラップの状態での二次構造のみが相互に変換していると考えてよい。それゆえ、熱力学的トラップのRNA二次構造を、従来公知の二次構造予測方法を用いて決定しかつそれらの存在確率の分布を適当な初期条件として与え、それらの二次構造同士の相互変換を連立微分方程式で記述し、かつ該連立微分方程式を上記初期条件の下で解くことによって、上記時間的区間内の任意の時点における各二次構造の存在確率を予測することが可能となることを見出し、本発明を完成させた。

[0010] すなわち、本発明は、次の特徴を有するものである。

[1] RNAの二次構造の動態を予測する方法であって、

(A) RNAの塩基配列から長さ $L-1$ を有する区間を選択し、該区間において、平衡状態における2種以上の特定二次構造を決定し、かつ、各特定二次構造のギブス自由エネルギーと平衡状態での存在確率とを算出するステップと、

(B) 前記特定二次構造同士の間の相互変換における遷移状態の二次構造を決定するとともに、各遷移状態の二次構造のギブス自由エネルギーを求めるステップと、

(C) 各遷移状態の二次構造およびそれらのギブス自由エネルギーから、各特定二次構造同士の間の相互変換の経路の活性化エネルギーを算出するステップと、

(D) 各活性化エネルギーを基にして、平衡状態に達するまでの時間的区間内の各時点における各特定二次構造の存在確率を決定するための連立微分方程式を形成するステップと、

(E) 下記(i)～(iv)のうちのいずれか1つの解法によって、各特定二次構造が、前記時間的区間内の各時点において何程の確率で存在し得るかを決定するステップと、

を有することを特徴とする、RNAの二次構造の動態を予測する方法。

(i) 前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とし、そこから導かれる該時点 t_0 での各特定二次構造の存在確率を初期値として前記連立微分方程式を解き、求める存在確率とする解法。

(ii) 前記時間的区間の或る時点 t_0 において各特定二次構造の存在比率が全て等しいものであるとして、各特定二次構造の該時点 t_0 での存在確率を決定し、これを初期値として、前記連立微分方程式を解き、求める存在確率とする解法。

(iii) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率を平均値とし、かつ、該平均値と、各特定二次構造が等しい比率で存在した場合の存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った 1 個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値として前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作を N 回繰り返した上で、各時点毎に、該 N 回の操作毎の存在確率の総和を N で除して平均し、求める存在確率とする解法。

(iv) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造が等しい比率で存在した場合の存在確率を平均値とし、かつ、該平均値と、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った 1 個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値とし、前記連立微分方

程式を解き、該初期値を用いた場合の存在確率を求める、という操作を N 回繰り返した上で、各時点毎に、該 N 回の操作毎の存在確率の総和を N で除して平均し、求める存在確率とする解法。

[2] 以下の (I) ~ (VI) のステップをさらに有するものである、上記 [1] に記載の方法。

(I) 前記長さ L_1 よりも大きい長さ L を有する対象区間 S を与えるステップ。

(II) 前記対象区間 S 内の位置 x の一端を位置 $x = x_1$ として、位置 x_1 と位置 $x_1 + L_1$ とを両端とする小区間を $S(x_1)$ とし、

上記のステップ (A) で選択される区間が前記小区間 $S(x_1)$ であるとして、上記のステップ (A) ~ (E) を行って、該小区間 $S(x_1)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定するステップ。

(III) 前記位置 x_1 を、所定の微小長さ r だけ他端側へ移動させた位置を $x_2 = x_1 + r$ とし、位置 x_2 と位置 $x_2 + L_1$ とを両端とする小区間を $S(x_2)$ とし、前記小区間 $S(x_1)$ の代わりに該小区間 $S(x_2)$ を用いる以外は前記 (II) のステップと同様にして、該小区間 $S(x_2)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する、という操作を、前記位置 $x = x_1, x_2$ に続いて、小区間の一端が対象区間 S の他端に達するまで、位置 $x = x_3, \dots, x_M$ について繰り返し行い、それによって、最初の小区間 $S(x_1)$ から最後の小区間 $S(x_M)$ までの各小区間毎の前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定するステップ。

(IV) 前記 (II) ~ (III) のステップの結果に基づいて、前記の小区間 $S(x_1)$ から $S(x_M)$ のうちの任意の小区間 $S(x_i)$ において、任意の時間 t における前記各特定二次構造の存在確率を決定し、

前記各特定二次構造において、位置および構造を表す所定の特徴パラメータ値を有する任意のモチーフ M_k が存在するか否かを調べ、その結果と上記で

得られた前記時間 t での各特定二次構造の存在確率とに基づいて、該小区間 $S(x_i)$ 内において、前記時間 t において該モチーフ M_k が何程の蓋然性で存在するかを決定し、これを、小区間 $S(x_i)$ における該モチーフ M_k に関する第 1 の蓋然性とする、

という操作を、前記の小区間 $S(x_i)$ から $S(x_M)$ の全てに対して行なうステップ。

(V) 前記 (IV) のステップにおいて各小区間内の各特定二次構造における前記モチーフ M_k の存在および非存在を調べた結果に基づいて、各小区間において、モチーフ M_k を有する特定二次構造が少なくとも 1 つ存在するか否かを決定する、

という操作を、前記の小区間 $S(x_i)$ から $S(x_M)$ の全てに行い、そのような特定二次構造が少なくとも 1 つ存在する小区間の頻度を、モチーフ M_k に関する第 2 の蓋然性とするステップ。

(VI) 前記 (IV) および (V) のステップで求められた、第 1 の蓋然性と、第 2 の蓋然性とに基づいて、前記時間 t における前記モチーフ M_k の存在を決定するステップ。

[3] 前記モチーフ M_k が、ステム・ループ構造を含むものである、上記 [2] に記載の方法。

[4] 前記モチーフ M_k が、2つのステム・ループ構造の間にあって他の塩基配列と相互作用していない間隙部分と、その両側にある前記ステム・ループ構造のそれぞれの内側のステム部分とを有してなる凹形構造を含むものである、上記 [2] または [3] に記載の方法。

[5] RNA の二次構造の動態を予測するためのコンピュータプログラムであって、

コンピュータを、

(P1) RNA の塩基配列から長さ L_1 を有する区間を選択し、該区間において、平衡状態における 2 種以上の特定二次構造を決定し、かつ各特定二次構造のギブス自由エネルギーおよび平衡状態での存在確率を算出手

段、

(P 2) 前記特定二次構造同士の間相互変換における遷移状態の二次構造を決定するとともに、各遷移状態の二次構造のギブス自由エネルギーを求める手段、

(P 3) 各遷移状態の二次構造およびそれらのギブス自由エネルギーから、各特定二次構造同士の間相互変換の経路の活性化エネルギーを算出する手段、

(P 4) 各活性化エネルギーを基にして、平衡状態に達するまでの時間的区間内の各時点における各特定二次構造の存在確率を決定するための連立微分方程式を形成する手段、

(P 5) 下記 (i) ~ (iv) のうちのいずれか 1 つの解法によって、各特定二次構造が、前記時間的区間内の各時点において何程の確率で存在し得るかを決定する手段、

として機能させるための前記コンピュータプログラム。

(i) 前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とし、そこから導かれる該時点 t_0 での各特定二次構造の存在確率を初期値として前記連立微分方程式を解き、求める存在確率とする解法。

(ii) 前記時間的区間の或る時点 t_0 において各特定二次構造の存在比率が全て等しいものであるとして、各特定二次構造の該時点 t_0 での存在確率を決定し、これを初期値として、前記連立微分方程式を解き、求める存在確率とする解法。

(iii) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率を平均値とし、かつ、該平均値と、各特定二次構造が等しい比率で存在した場合の存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った1個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値として前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作をN回繰り返した上で、各時点毎に、該N回の操作毎の存在確率の総和をNで除して平均し、求める存在確率とする解法。

(iv) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造が等しい比率で存在した場合の存在確率を平均値とし、かつ、該平均値と、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った1個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値とし、前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作をN回繰り返した上で、各時点毎に、該N回の操作毎の存在確率の総和をNで除して平均し、求める存在確率とする解法。

[6] コンピュータを、さらに以下の (p 1) ~ (p 6) の手段として機能させるためのものである、上記 [5] に記載のコンピュータプログラム。

(p 1) 前記長さ L_1 よりも大きい長さ L を有する対象区間 S を入力する手段。

(p 2) 前記対象区間 S 内の位置 x の一端を位置 $x = x_1$ として、位置 x_1 と位置 $x_1 + L_1$ とを両端とする小区間を $S(x_1)$ とし、

上記の手段 (P 1) で選択される区間が前記小区間 $S(x_1)$ であるとして、上記の手段 (P 1) ~ (P 5) を用いて、該小区間 $S(x_1)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する手段。

(p 3) 前記位置 x_1 を、所定の微小長さ r だけ他端側へ移動させた位置を

$x_2 = x_1 + r$ とし、位置 x_2 と位置 $x_2 + L_1$ とを両端とする小区間を $S(x_2)$ とし、前記小区間 $S(x_1)$ の代わりに該小区間 $S(x_2)$ を用いる以外は前記 (p 2) の手段と同様にして、該小区間 $S(x_2)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する、という操作を、前記位置 $x = x_1, x_2$ に続いて、小区間の一端が対象区間 S の他端に達するまで、位置 $x = x_3, \dots, x_M$ について繰り返し行い、それによって、最初の小区間 $S(x_1)$ から最後の小区間 $S(x_M)$ までの各小区間毎の前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する手段。

(p 4) 前記 (p 2) ~ (p 3) の手段を用いて得られた結果に基づいて、前記の小区間 $S(x_1)$ から $S(x_M)$ のうちの任意の小区間 $S(x_i)$ において、任意の時間 t における前記各特定二次構造の存在確率を決定し、

前記各特定二次構造において、位置および構造を表す所定の特徴パラメータ値を有する任意のモチーフ M_k が存在するか否かを調べ、その結果と上記で得られた前記時間 t での各特定二次構造の存在確率とに基づいて、該小区間 $S(x_i)$ 内において、前記時間 t において該モチーフ M_k が何程の蓋然性で存在するかを決定し、これを、小区間 $S(x_i)$ における該モチーフ M_k に関する第 1 の蓋然性とする、

という操作を、前記の小区間 $S(x_1)$ から $S(x_M)$ の全てに対して行なう手段。

(p 5) 前記 (p 4) の手段を用いて各小区間内の各特定二次構造における前記モチーフ M_k の存在および非存在を調べた結果に基づいて、各小区間において、モチーフ M_k を有する特定二次構造が少なくとも 1 つ存在するか否かを決定する、

という操作を、前記の小区間 $S(x_1)$ から $S(x_M)$ の全てに行い、そのような特定二次構造が少なくとも 1 つ存在する小区間の頻度を、モチーフ M_k に関する第 2 の蓋然性とする手段。

(p 6) 前記 (p 4) および (p 5) の手段で求められた、第 1 の蓋然性

と、第2の蓋然性とに基づいて、前記時間 t における前記モチーフ M_k の存在を決定するための出力を行う手段。

[7] 前記モチーフ M_k が、ステム・ループ構造を含むものである、上記 [6] に記載のコンピュータプログラム。

[8] 前記モチーフ M_k が、2つのステム・ループ構造の間において他の塩基配列と相互作用していない間隙部分と、その両側にある前記ステム・ループ構造のそれぞれの内側のステム部分とを有してなる凹形構造を含むものである、上記 [6] または [7] に記載のコンピュータプログラム。

[9] RNAの二次構造の動態を予測するための装置であって、

(Y1) RNAの塩基配列から長さ L_1 を有する区間を選択し、該区間において、平衡状態における2種以上の特定二次構造およびそれらの存在確率を算出する第一演算部と、

(Y2) 前記特定二次構造同士の間での相互変換における遷移状態の二次構造を決定するとともに、各遷移状態の二次構造のギブス自由エネルギーを求める第二演算部と、

(Y3) 各遷移状態の二次構造およびそれらのギブス自由エネルギーから、各特定二次構造同士の間での相互変換の経路の活性化エネルギーを算出する第三演算部と、

(Y4) 各活性化エネルギーを基にして、平衡状態に達するまでの時間的区間内の各時点における各特定二次構造の存在確率を決定するための連立微分方程式を形成する第四演算部と、

(Y5) 下記 (i) ~ (iv) のうちのいずれか1つの解法によって、各特定二次構造が、前記時間的区間内の各時点において何程の確率で存在し得るかを決定する第五演算部と、
を有する、前記装置。

(i) 前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とし、そこから導かれる該時点 t_0 での各特定二次構造の存在確率を初期値として前記連立微分

方程式を解き、求める存在確率とする解法。

(ii) 前記時間的区間の或る時点 t_0 において各特定二次構造の存在比率が全て等しいものであるとして、各特定二次構造の該時点 t_0 での存在確率を決定し、これを初期値として、前記連立微分方程式を解き、求める存在確率とする解法。

(iii) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率を平均値とし、かつ、該平均値と、各特定二次構造が等しい比率で存在した場合の存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った 1 個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値として前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作を N 回繰り返した上で、各時点毎に、該 N 回の操作毎の存在確率の総和を N で除して平均し、求める存在確率とする解法。

(iv) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造が等しい比率で存在した場合の存在確率を平均値とし、かつ、該平均値と、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った 1 個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値とし、前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作を N 回繰り返した上で、各時点毎に、該 N 回の操作毎の存在確率の総和を N で除して平均し、求める存在確率とする解法。

[10] さらに以下の (y 1) ~ (y 6) の手段を有するものである、上記 [9] に記載の装置。

(y 1) 前記長さ L_1 よりも大きい長さ L を有する対象区間 S を入力する手段。

(y 2) 前記対象区間 S 内の位置 x の一端を位置 $x = x_1$ として、位置 x_1 と位置 $x_1 + L_1$ とを両端とする小区間を $S(x_1)$ とし、

上記の手段 (Y 1) で選択される区間が前記小区間 $S(x_1)$ であるとして、上記の手段 (Y 1) ~ (Y 5) を用いて、該小区間 $S(x_1)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する手段。

(y 3) 前記位置 x_1 を、所定の微小長さ r だけ他端側へ移動させた位置を $x_2 = x_1 + r$ とし、位置 x_2 と位置 $x_2 + L_1$ とを両端とする小区間を $S(x_2)$ とし、前記小区間 $S(x_1)$ の代わりに該小区間 $S(x_2)$ を用いる以外は前記 (y 2) の手段と同様にして、該小区間 $S(x_2)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する、という操作を、前記位置 $x = x_1, x_2$ に続いて、小区間の一端が対象区間 S の他端に達するまで、位置 $x = x_3, \dots, x_M$ について繰り返し行い、それによって、最初の小区間 $S(x_1)$ から最後の小区間 $S(x_M)$ までの各小区間毎の前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する手段。

(y 4) 前記 (y 2) ~ (y 3) の手段によって決定された結果に基づいて、前記の小区間 $S(x_1)$ から $S(x_M)$ のうちの任意の小区間 $S(x_i)$ において、任意の時間 t における前記各特定二次構造の存在確率を決定し、

前記各特定二次構造において、位置および構造を表す所定の特徴パラメータ値を有する任意のモチーフ M_k が存在するか否かを調べ、その結果と上記で得られた前記時間 t での各特定二次構造の存在確率とに基づいて、該小区間 $S(x_i)$ 内において、前記時間 t において該モチーフ M_k が何程の蓋然性で存在するかを決定し、これを、小区間 $S(x_i)$ における該モチーフ M_k に関する第 1

の蓋然性とする、

という操作を、前記の小区間 $S(x_1)$ から $S(x_M)$ の全てに対して行なう手段。

(y 5) 前記 (y 4) の手段を用いて各小区間内の各特定二次構造における前記モチーフ M_k の存在および非存在を調べた結果に基づいて、各小区間において、モチーフ M_k を有する特定二次構造が少なくとも 1 つ存在するか否かを決定する、

という操作を、前記の小区間 $S(x_1)$ から $S(x_M)$ の全てに行い、そのような特定二次構造が少なくとも 1 つ存在する小区間の頻度を、モチーフ M_k に関する第 2 の蓋然性とする手段。

(y 6) 前記 (y 4) および (y 5) の手段で求められた、第 1 の蓋然性と、第 2 の蓋然性に基づいて、前記時間 t における前記モチーフ M_k の存在を決定するための出力を行う手段。

[1 1] 前記モチーフ M_k が、ステム・ループ構造を含むものである、上記 [1 0] に記載の装置。

[1 2] 前記モチーフ M_k が、2つのステム・ループ構造の間にあって他の塩基配列と相互作用していない間隙部分と、その両側にある前記ステム・ループ構造のそれぞれの内側のステム部分とを有してなる凹形構造を含むものである、上記 [1 0] または [1 1] に記載の装置。

発明の効果

[0011] 本発明の方法により、従来の方法では扱うことができなかった、RNA の二次構造がマクスウェル・ボルツマンの分布則に従う存在確率の分布を示すよりも前の時点における該 RNA の二次構造の動態を予測することが可能となる。それにより、生体内で RNA が実際に取っている二次構造をより反映する形で、RNA の二次構造を予測することができる。本発明はまた、該方法をコンピュータに実行させるためのプログラム、および該方法を実施するための装置を提供する。

図面の簡単な説明

[0012] [図1]本発明の一つの実施形態による、装置またはプログラムの構成の例を模式的に示すブロック図である。

[図2]本発明の一つの実施形態による、装置またはプログラムの構成の例を模式的に示すブロック図である。

[図3]本発明のプログラムにおける、パラメータの入力のためのパネルの一例を示す。

[図4]本発明のプログラムにおける、ユーザーに計算ステップの進捗状況を示すためのアイコンの一例を示す。

[図5]本発明のプログラムにおける、ユーザーに微分方程式の計算の進捗状況を示すためのアイコンの一例を示す。

[図6]実施例1からの3つのスペクトルを示す図である。(a)はスペクトル1、(b)はスペクトル2、(c)はスペクトル3に対応している。各スペクトルについて、横軸は塩基アドレス、縦軸は第1の蓋然性、ピークの色濃さは第2の蓋然性を表している。閾値を設定して、第3の蓋然性が90%以上のピークのみを表示させている。

[図7]実施例2からの3つのスペクトルを示す図である。(a)はスペクトル4、(b)はスペクトル5、(c)はスペクトル6に対応している。各スペクトルについて、横軸は塩基アドレス、縦軸は第1の蓋然性、ピークの色濃さは第2の蓋然性を表している。閾値を設定して、第3の蓋然性が90%以上のピークのみを表示させている。

[図8]実施例3における、 $T_{\text{fold}} = 3.0$ 秒以内で、ほぼ平衡状態に達する系の2つの例を示す図である。

[図9]実施例3における、平衡状態に向かって進みつつも、 $T_{\text{fold}} = 3.0$ 秒では平衡に達していない系の2つの例を示す図である。図中では二次構造の動態のおおよその様子を見るために、0.001秒の刻み幅を明記するとともに6秒までプロットしている。

[図10]実施例3における、0.001秒以内（最小刻み幅 Δt 内）に平衡状態に達してしまう系の例を示す図である。(a)は各特定二次構造の存在確率

に基づくプロットであり、(b)は構造間の活性化エネルギーと速度定数のマトリックスである。このマトリックスにおいては各特定二次構造を1～4とし、最上段にそれぞれの二次構造のその時点でのポピュレーションを表示するとともに、矢印“=>”の記載の後に、マクスウェル・ボルツマン分布した場合のポピュレーション（平衡状態でのポピュレーション）を記載している。さらに、下部のマトリックスにおいては、1=>2のように、どの構造からどの構造への遷移であるかの記載の下に、その遷移の活性化エネルギー（kcal/mol）と遷移の速度定数 k_{obs} と刻み幅 Δt をかけた実際の遷移の量が記載されている。また本計算では、ある数種の構造間で遷移の量が一定以上である場合には、その関係が刻み幅内の時間ですでに平衡に達しているとしているが、それはマトリックス中では対応する構造の遷移のマスと同じ色で塗りつぶすことによって明記している。

[図11]実施例3における、全く平衡状態に向かって進んで行かない系の例を示す図である。(a)は各特定二次構造の存在確率に基づくプロットであり、(b)は構造間の活性化エネルギーと速度定数のマトリックスである。このマトリックスにおいては各特定二次構造を1～6とし、最上段にそれぞれの二次構造のその時点でのポピュレーションを表示するとともに、矢印“=>”の記載の後に、マクスウェル・ボルツマン分布した場合のポピュレーション（平衡状態でのポピュレーション）を記載している。さらに、下部のマトリックスにおいては、1=>2のように、どの構造からどの構造への遷移であるかの記載の下に、その遷移の活性化エネルギー（kcal/mol）と遷移の速度定数 k_{obs} と刻み幅 Δt をかけた実際の遷移の量が記載されている。

[0013]（発明の詳細な説明）

以下、本発明の方法をより詳細に説明する。

なお、本発明においてRNA上でいう「長さ」とは、配列長であって、配列された塩基の数を意味する。長さの最小単位は塩基1個であり、「長さ」は「塩基数」と読み替えてよい。よって、本明細書では、〔長さが100個〕などと記載し、例えば〔長さ r だけ移動させる〕は〔塩基数 r だけ移動さ

せる]を意味する。

[0014] 上述した通り、本発明は、RNAの二次構造の動態を予測する方法を提供し、当該方法は、以下のステップ（A）～（E）を含む：

（A）RNAの塩基配列から長さ L_1 を有する区間を選択し、該区間において、平衡状態における2種以上の特定二次構造を決定し、かつ各特定二次構造のギブス自由エネルギーおよび平衡状態での存在確率を算出するステップ；

（B）前記特定二次構造同士の間の相互変換における遷移状態の二次構造を決定するとともに、各遷移状態の二次構造のギブス自由エネルギーを求めるステップ；

（C）各遷移状態の二次構造およびそれらのギブス自由エネルギーから、各特定二次構造同士の間の相互変換の経路の活性化エネルギーを算出するステップ；

（D）各活性化エネルギーを基にして、平衡状態に達するまでの時間的区間内の各時点における各特定二次構造の存在確率を決定するための連立微分方程式を形成するステップ；および、

（E）各特定二次構造が、前記時間的区間内の各時点において何程の確率で存在し得るかを決定するステップ。

[0015] 或る区間内のRNA配列は、二次構造を形成するために十分な時間を与えられたときに、その取り得る各二次構造の存在確率が時間的に実質的に変化しない状態に達し、そして該状態においては、各二次構造が、マクスウェル・ボルツマンの分布則に従う存在確率で分布すると考えられる。本明細書でいう「平衡状態」とは、二次構造の存在確率の分布が、実質的にマクスウェル・ボルツマンの分布則に従うものである状態をいう。ここで「実質的にマクスウェル・ボルツマンの分布則に従う」とは、各二次構造の存在確率が、例えば、マクスウェル・ボルツマンの分布則に従う存在確率に対して10%未満の差異、好ましくは1%未満の差異であることを意味する。

[0016] 本発明の方法は、上記のような平衡状態にある時点よりも前（すなわち、

「非平衡状態」)の任意の時点において、2種以上の二次構造が各々いかなる確率で存在し得るかを決定するための方法である。該任意の時点での各二次構造の存在確率を決定することにより、非平衡状態にある時間的区間内において、各二次構造の存在確率がどのように推移していくかを決定すること、すなわち、RNAの二次構造の動態を予測することが可能となる。従って、本明細書において、「RNAの二次構造の動態」とは、RNAが取り得る各二次構造の各時点における存在確率の経時的な変化の態様を意味している。なお、本発明の方法によって、非平衡状態にある時間的区間内でのRNAの二次構造の経時的な変化が予測され得るが、該時間的区間内の特定の時点指定したときには、当該特定の時点での各二次構造の存在確率が予測されたものと考えることができる。従って、別の観点からは、本発明により、平衡状態に達するまでの時間的区間内の或る時点におけるRNAの二次構造を予測するための方法が提供される。以下で詳細に説明するように、本発明の方法では、非平衡状態にある時間的区間内での存在確率を解析される2種以上の二次構造は、平衡状態において存在し得ると予測された二次構造である。

[0017] 本発明の方法による予測の対象とすべきRNAとしては、mRNA（本明細書では、特に矛盾のない限りはその前駆RNAをも包含する意味で用いられる）の他、miRNA（本明細書では、特に矛盾のない限りはその前駆RNAをも包含する意味で用いられる）、rRNA、tRNA、ncRNAなどが挙げられる。ncRNAは、タンパク質に翻訳されないRNAであって、機能性RNAなどとも呼ばれ、近年一段と注目されている。

これらのなかでも、mRNAは、従来より医薬品開発において重要な創薬ターゲットの1つとして位置付けられており、アンチセンス医薬品やsiRNAの効率は、mRNAの二次構造に左右されるという知見もある。また、そうした二次構造に関する知見は、miRNAなどの既知または未知の分子生物学現象に関与する構造の解明に寄与し得ると考えられる一方で、一部の特定のmRNAを除いては、生体内におけるその高次構造がこれまで全く明

らかにされていないことから、RNA中の有意義な二次構造の存在を予測、議論すべき対象としてきわめて重要である。

一方、miRNAはステム・ループ構造を有する前駆体（pre-miRNA）から酵素dicerで切断されて生成し、mRNAを切断分解することなく物理的にその翻訳を抑制するので、この前駆体RNAの構造を安定化することにより、成熟miRNAの生成を抑制して結果的に標的mRNAの翻訳を促進することができる。このような観点から、pre-miRNAも、好ましい予測対象の1つであるといえる。

また、ncRNAは、何らかの機能を生体内で果たしていると考えられている。まちがいなくそのような機能は、それらRNAが何らかの高次構造をとることで生体内での役割を担っていると考えられている。さらに、こうしたncRNAの中には、rRNAやtRNAのように（おそらく）一意の構造をとるであろうものも含まれる場合があるが、ほとんどのものはその構造的挙動がまったく未知である。このような観点から、ncRNAもまた、本発明にとって好ましい予測対象の1つであるといえる。

[0018] 上記ステップ（A）では、まず、RNAの塩基配列から、予測の対象とすべき長さL1の区間が選択される。区間の長さL1は、生物学的な条件をできるだけ模倣できるように設定することが好ましい。例えば、mRNAのような長いRNAが翻訳される過程においては、mRNA上の高次構造とタンパク質をリボソームが一本鎖に直しながら進行する。すなわち、リボソームが通り抜けた直後のmRNAは、いかような構造をもとり得る自由な状態にある。そうした自由な状態である部位は、次々と何らかの高次構造をとっていく。この状態は、現在の構造予測技術（例えば、本発明のプログラムにおける手段P1や、mfoldなどの公知の構造予測ソフトウェア）が想定している状況に近似していると考えられる。特に、mfoldにおいては、実験的に得られたパラメータを使用して構造予測を行っているが、そのパラメータは比較的短いRNAについての実験で得られたものである。つまりmfoldが正しい結果を与えるのは比較的短い（経験的に200塩基以下である）配列に対してだ

けである。そこで、リボソームの進行速度と、そうした現在の構造予測技術（後述するように、ステップ（A）での特定二次構造の決定は、従来公知の構造予測技術を用いて行われる）の特性とを考慮して、好ましい予測結果が得られるように、区間の長さ L_1 を設定すればよい。

また、区間の長さ L_1 は、生物学的な機能を有し得る局所的な二次構造（例、ステム・ループ、シュードノット構造、凹形構造（後述）など）を含み得る長さであるべきである。

以上のことから、具体的な区間の長さ L_1 としては、例えば、塩基数 50 ～ 400 個、好ましくは塩基数 100 ～ 300 個が挙げられる。

[0019] 次いで、上記のようにして選択された長さ L_1 の区間において、後のステップでの解析に供されるべき平衡状態における 2 種以上の二次構造が決定される。また、決定された各二次構造のギブス自由エネルギーおよび平衡状態での存在確率が算出される。ここで決定された平衡状態における二次構造を、本明細書では「特定二次構造」と呼ぶ。

特定二次構造には、最も安定な（すなわち、ギブス自由エネルギーが最も低い）二次構造が含まれることが好ましい。また、特定二次構造の種類は 2 種以上であれば特に制限されないが、典型的には 2 ～ 20 種である。

非平衡状態において殆ど存在しないと考えられる二次構造をこの段階で除外しておくことが好ましい。そのような除外されるべき二次構造としては、例えば、ギブス自由エネルギーがあまりに大きい（すなわち、平衡状態における存在確率があまりに低い）二次構造、および、実験的データなどから、非平衡状態において実質的に全く存在し得ないと予想される二次構造などが挙げられる。

特定二次構造の選択のために、例えば、特定二次構造の個数を予め設定しておき、平衡状態での存在確率が高いもの（つまりギブス自由エネルギーが低いもの）から優先的に選択すること、平衡状態での存在確率が所定の値以上の特定二次構造のみを選択すること、および予測ソフトウェアの望まれない特性のために算出された特定二次構造を排除することなどの基準を与え

することもできる。特定二次構造の選択は、当業者によって適宜為され得る。

[0020] 本発明の方法においては、長さ L の区間からの二次構造の抽出は、例えば、mfoldなどの従来公知のRNA構造予測ソフトウェアを利用して行えばよい。また、抽出された各二次構造のギブス自由エネルギーは、例えば、mfoldなどの構造予測ソフトウェアによって算出された値を用いることができる。あるいは、抽出された各二次構造のギブス自由エネルギーを、以下の参考文献：

1. Mathews et al, J. Mol. Biol. (1999) 288, 911-940.
2. Xia T et al, Biochemistry (1998) 37, 14719-14735.
3. Giese MR et al, Biochemistry (1998) 37, 1094-1100.
4. Draper DE et al, Methods in Enzymology (1995) 259, 281-242.
5. Serra MJ et al, Methods in Enzymology (1995) 259, 242-261.

に記載の方法およびパラメータをそのまま使用して算出してもよい。このエネルギー算出方法は、小さな差異はあるものの広く使われており、好ましい結果を与え得ると考えられる。各二次構造に対して算出されたギブス自由エネルギーを用いてマクスウェル・ボルツマンの分布則を適用することによって、各二次構造の平衡状態での存在確率を算出することができる。

また、mfoldなどの構造予測方法自体の条件は、目的に適合するよう適宜に変更してよい。例えば、通常 37°C の温度での構造予測が行われるが、好熱菌（高い温度の環境で生育する菌）などに見られるRNAの構造解析には、 37°C よりはるかに高い温度を指定して構造予測をする。また、実験的データなどが得られている場合などには、mfoldなどの構造予測結果に影響を与えるその他のパラメータを変更してよい（ステム・ループのループの最低長さや、強制的にある構造を組ませるように指定するなど）。

さらに、そのRNAのある特殊条件での構造解析をするため、推測に沿って、mfoldなどの構造予測結果に影響を与えるパラメータを変更してよい。例えば、ある領域にタンパク質またはアンチセンスなどが結合した条件での構造解析を行うのであれば、その領域を含むある一定領域を該RNA中での構

造形成に参加しないものとして、該領域をどことも相互作用しないと指定して構造予測方法を用いてもよい。

[0021] 次いで、上記ステップ（B）では、ステップ（A）で決定された複数の特定二次構造同士の間相互変換における遷移状態の二次構造が決定され、かつ各遷移状態の二次構造のギブス自由エネルギーが求められる。

さらに、上記ステップ（C）では、ステップ（B）で得られた各遷移状態の二次構造およびそれらのギブス自由エネルギーから、各特定二次構造同士の間相互変換の経路の活性化エネルギーが算出される。

[0022] ここで、ある2種の特定二次構造について、それらの間の相互変換における遷移状態の二次構造を、それらの特定二次構造の最大公約数的共通構造としている。すなわち、ある特定二次構造AからBへの（およびその逆の）遷移について、最低限ほどこねばならない部分二次構造すべてをシングルストランドにした二次構造を、それらの間の相互変換における遷移状態の二次構造とする。各遷移状態の二次構造のギブス自由エネルギーの算出は、上記したのと同様の方法で行うことができる。

[0023] 上述した通り、本発明の方法では、非平衡状態においても、ステップ（A）において平衡状態で存在し得ると決定された二次構造のみが存在すると想定されている。

上記ステップ（D）では、ステップ（A）で決定された各特定二次構造が、非平衡状態にある時間的区間内の各時点においていかなる確率で存在し得るかを決定するための連立微分方程式が形成される。該連立微分方程式の形成は、ステップ（C）で得られた各活性化エネルギーを用いて行われる。ここで、活性化エネルギーは、基準となる状態（そのギブス自由エネルギーを ΔG とする）から遷移状態（そのギブス自由エネルギーを ΔG^{TS} とする）へのエネルギー変化 $\Delta \Delta G = \Delta G^{TS} - \Delta G$ （正の値を取る）と定義される。該方法をより詳しく説明すると以下の通りである。

[0024] いま、ステップ（A）で決定された特定二次構造がN個であるとし、それらを二次構造 n （ $n = 1, 2, \dots, N$ ）とする。また、二次構造 n のギ

ツブス自由エネルギーを G_n 、時間 t での存在確率を $P_n(t)$ とする。平衡状態となる時点は $t = \infty$ であると考えることができるため、平衡状態での二次構造 n の存在確率を $P_n(\infty)$ と表す。ここで、以下の式 (1) が成り立つ。

[0025] [数1]

$$P_n(\infty) = \frac{\exp(-\frac{G_n}{RT})}{\sum_{x=1}^N \exp(-\frac{G_x}{RT})} \quad (1)$$

[0026] 上記式 (1) 中、 R は気体定数 ($R=8.31447 \text{ J K}^{-1}\text{mol}^{-1}$)、 T は絶対温度 (K) である。

さらに、二次構造 n と二次構造 m との間の相互変換における遷移状態 n, m の二次構造のギブス自由エネルギーを $G_{n,m}$ とする。二次構造 n から二次構造 m への変換の速度定数 ($k_{n \rightarrow m}$) (単位: 秒⁻¹) を示すアイリングの絶対反応速度式は以下の式 (2) の通りである。但し、 n と m が同じ値の場合は、速度定数はゼロとする。

[0027] [数2]

$$k_{n \rightarrow m} [\text{s}^{-1}] = \begin{cases} \exp(-\frac{G_{n,m} - G_n}{RT}) \times K \times \frac{k_B T}{h} & (n \neq m \text{ の場合}) \\ 0 & (n = m \text{ の場合}) \end{cases} \quad (2)$$

[0028] 上記式 (2) 中、 R は気体定数 ($R=8.31447 \text{ J K}^{-1}\text{mol}^{-1}$)、 T は絶対温度 (K)、 K はアイリング定数 ($K=1.00$)、 k_B はボルツマン定数 ($k_B=1.3806504 \times 10^{-23} \text{ J K}^{-1}$)、 h はプランク定数 ($h=6.62621 \times 10^{-34} \text{ J s}$) である。ここで式 (2) 中の ($G_{n,m} - G_n$) が、二次構造 n から二次構造 m への相互変換において遷移状態 n, m を通る経路の活性化エネルギーである。

すると、二次構造 n の存在確率の時間変化 ($d P_n(t) / d t$) は以下の式 (3) のように表される。

[0029]

[数3]

$$\frac{dP_n(t)}{dt} = \sum_{x=1}^N k_{x \rightarrow n} P_x(t) - \sum_{x=1}^N k_{n \rightarrow x} P_n(t) \quad (3)$$

[0030] 上記式(3)において $n = 1, 2, \dots, N$ とすることで得られる連立微分方程式を、適当な初期条件のもとで解くことによって、平衡状態に達するまでの時間的区間内の各時点における各特定二次構造の存在確率を決定することができる。

[0031] 現実の解析においては、通常、ステップ(D)で得られる連立微分方程式は数値的に解かれる。すなわち、該連立微分方程式を、存在確率の初期値を与える時点 $t = t_0$ から計算の終了時点 $t = T_{\text{fold}}$ まで適当な時間刻み幅 Δt 毎に数値的に解いていくことで、任意の時点 t における各特定二次構造の存在確率 $P_n(t)$ を得ることができる。このとき、1次近似を仮定すれば、存在確率 $P_n(t)$ の計算式は、以下の式(4)の通りとなる。

[0032] [数4]

$$P_n(t) = P_n(t - \Delta t) + \left(\sum_{x=1}^N k_{x \rightarrow n} P_x(t - \Delta t) - \sum_{x=1}^N k_{n \rightarrow x} P_n(t - \Delta t) \right) \Delta t \quad (4)$$

[0033] 次いで、上記ステップ(E)では、各特定二次構造が前記時間的区間内の各時点において何程の確率で存在し得るかを決定するために、ステップ(D)で得られた連立微分方程式が解かれる。連立微分方程式の初期値は、生物学的な条件をできるだけ模倣できるように設定することが好ましい。本発明の方法のステップ(E)では、該連立微分方程式は、以下に限定されないが、例えば、上記[1]のステップ(E)に示した(i)～(iv)のうちのいずれか1つの解法によって解かれ得る。

[0034] 上記(i)の解法は、〔平衡状態における各特定二次構造の存在確率は、形成される水素結合ペアの個数に比例しており、また水素結合ペアの個数が多い特定二次構造ほどその形成に時間がかかるはずであるため、非平衡状態の或る時点 t_0 においては、各特定二次構造は平衡状態におけるそれらの存在確

率の逆数に比例した確率で存在し得る]との考えに基づいている。そのようにして時点 t_0 における各特定二次構造の存在確率を初期条件として与え、例えば上記式 (4) をコンピュータを用いて数值的に解くことによって、各特定二次構造が非平衡状態にある時間的区間内の各時点において何程の確率で存在し得るかを決定することができる。

[0035] 上記 (ii) の解法は、非平衡状態の或る時点 t_0 において、全ての特定二次構造が等しい確率で存在し得るとする考えに基づいている。そのようにして時点 t_0 における各特定二次構造の存在確率を初期条件として与え、例えば上記式 (4) をコンピュータを用いて数值的に解くことによって、各特定二次構造が非平衡状態にある時間的区間内の各時点において何程の確率で存在し得るかを決定することができる。

[0036] 上記 (iii) の解法は、時点 t_0 における各特定二次構造の存在確率を、各々特定の数値として与えるのではなく、各特定二次構造毎に正規分布として与えるものである。或る特定二次構造のための正規分布の平均値は、上記 (i) の解法における初期時点 t_0 でのその存在確率と同様にして得られる。一方、或る特定二次構造のための正規分布の標準偏差は、上記のようにして得られた平均値と、上記 (ii) の解法と同様にして求めた存在確率との差の絶対値により与えられる。各特定二次構造毎に、各々の正規分布に従う確率で選択された1個の乱数を与え、さらに、与えられた乱数を全ての特定二次構造について総和することで得られる値で、各特定二次構造のために与えられた乱数を除することによって、各特定二次構造の時点 t_0 における存在確率が得られる。そのようにして得られた時点 t_0 における各特定二次構造の存在確率を初期条件として、例えば上記式 (4) をコンピュータを用いて数值的に解くことで、「該初期値を用いた場合の」、各特定二次構造の非平衡状態にある時間的区間内での任意の時点での存在確率が求まる。上記の〔正規分布に従った乱数を与えて、該乱数に基づいて時点 t_0 における各特定二次構造の存在確率を決定し、それを初期条件として連立微分方程式を解く〕という操作を合計で N 回繰り返す。各操作毎の、任意の時点 t における或る特定二次構造

の存在確率を算術平均する（すなわち、該存在確率の総和をNで除する）ことで得られる値が、当該特定二次構造の時点 t における存在確率となる。上記操作の繰り返し回数Nは、大きければ大きいほど乱数の選択に左右されない結果が得られるため好ましいが、実際の解析では、通常1～100の範囲から選択される適当な回数とすればよい。

[0037] 上記 (iv) の解法は、時点 t_0 における各特定二次構造の存在確率を、各々特定の数值として与えるのではなく、各特定二次構造毎に正規分布として与えるものである。或る特定二次構造のための正規分布の平均値は、上記 (ii) の解法における初期時点 t_0 でのその存在確率と同様にして得られる。一方、或る特定二次構造のための正規分布の標準偏差は、上記のようにして得られた平均値と、上記 (i) の解法と同様にして求めた存在確率との差の絶対値により与えられる。各特定二次構造毎に、各々の正規分布に従う確率で選択された1個の乱数を与え、さらに、与えられた乱数を全ての特定二次構造について総和することで得られる値で、各特定二次構造のために与えられた乱数を除することによって、各特定二次構造の時点 t_0 における存在確率が得られる。そのようにして得られた時点 t_0 における各特定二次構造の存在確率を初期条件として、例えば上記式 (4) をコンピュータを用いて数値的に解くことで、「該初期値を用いた場合の」、各特定二次構造の非平衡状態にある時間的区間内での任意の時点での存在確率が求まる。

上記の〔正規分布に従った乱数を与えて、該乱数に基づいて時点 t_0 における各特定二次構造の存在確率を決定し、それを初期条件として連立微分方程式を解く〕という操作を合計でN回繰り返す。各操作毎の、任意の時点 t における或る特定二次構造の存在確率を算術平均する（すなわち、該存在確率の総和をNで除する）ことで得られる値が、当該特定二次構造の時点 t における存在確率となる。上記操作の繰り返し回数Nは、大きければ大きいほど乱数の選択に左右されない結果が得られるため好ましいが、実際の解析では、通常1～100の範囲から選択される適当な回数とすればよい。

[0038] 上記 (i) ～ (iv) の解法において、コンピュータを用いた現実の数値解析

では、通常、存在確率の初期値を与える時点 t_0 を $t = 0$ として計算すればよい。また、上記時間刻み幅 Δt は、1 ピコ秒～1 ミリ秒であり得、例えば1 ミリ秒（0. 001 秒）である。

また、計算の終了時点 $t = T_{\text{fold}}$ については、生物学的な条件を考慮に入れて設定することが好ましい。あるいは、mRNA のある時点での構造を見たい場合には、任意の数値を入れることも間違いではない。そこで、 T_{fold} の値としては、例えば、0～30 秒であり得る。なお、後述する実施例3で示すように、各特定二次構造の存在確率の時間的変遷を解析したとき、おおまかには以下の4つのケースが観察される：

（c1）時点 $t = T_{\text{fold}}$ までに、各特定二次構造の存在確率が平衡状態での存在確率に収束するケース；

（c2）各特定二次構造の存在確率が平衡状態での存在確率に向かいつつも、時点 $t = T_{\text{fold}}$ までには収束に達しないケース；

（c3）初期時点 $t = 0$ の直後に、各特定二次構造の存在確率が平衡状態での存在確率に収束するケース；および、

（c4）時点 $t = T_{\text{fold}}$ までに、各特定二次構造の存在確率が、平衡状態での存在確率に向かって全く進まないケース。

上記（c2）のケースは、本発明の方法で想定される時間枠において、もっともドラスティックに各構造の存在確率が時間の影響を受けるケースと言える。

上記（c3）のケースは、各特定二次構造間の活性化エネルギーが極めて低く、二次構造間の変換が非常に速い（すなわち、速度定数が非常に大きい）ものである。このような場合、理論的に正しくは刻み幅を細かくすることで解決されるが、後述する実施例で行っているように、二次構造間の相互変換を数値的に解かずに、マクスウェル・ボルツマンの分布則によって存在比率を与えてもよい。

上記（c4）のケースは、各特定二次構造間の活性化エネルギーが極めて高く、二次構造間の変換が非常に遅い（すなわち、速度定数が非常に小さい）

）ものである。これは、事実上平衡状態に達しない系であり、各特定二次構造の存在確率は初期条件に大きく左右される。この場合、生体内でそれらの特定二次構造が相互変換し、やがて存在確率の分布が平衡状態に達するとする考え自体が正しくなく、むしろ、ある初期時点での二次構造が単純にそのまま持続すると考えるのが妥当であり得る。ここで、自然界の現象として、RNAの二次構造の初期値が毎回同じである保証はないため、その部分の二次構造の存在確率の分布は、毎回異なるものであり得る。二次構造の存在確率の分布が毎回異なるものであるRNA部分の構造が生物学的に意義がある可能性は低いと考えられる。このような考え方に基づいて、本発明の方法に従って、RNAの塩基配列上で生物学的機能を有する部分と有しない部分とを区別することが可能となり得る。

現実の解析では、単一の小区間であっても上記（c1）～（c4）のケースが混在する場合もある。

[0039] 本発明の方法ではまた、上記特許文献2で開示された概念に従って、上記長さL1を有する区間を該文献の方法における小区間と考えて該区間において上記ステップ（A）～（E）を行い、続いて該区間をRNA塩基配列上で数塩基ずらすことで得られる別の区間においても同様のステップを行う、という操作を繰り返した後、全てのそのような区間において得られた結果を集積する手法も用いることができる。すなわち、上述した本発明の方法は、一つの実施形態では、上記[2]に示した（I）～（VI）のステップをさらに有する。

[0040] 以下に、上記[2]に示した実施形態を詳細に説明する。

なお、以下では、配列中の塩基の位置（配列の順番）を特定し得るように1つ1つの塩基に付与された番号（本明細書では、5'側から3'側に向かって数える）を「塩基アドレス」と呼んで説明する。塩基配列上の位置xは塩基アドレスであり、各小区間の先頭（5'側）からどの位置にあるかを示す相対的な塩基アドレスを用いてもよいが、対象区間Sの先頭（5'側）からどの位置にあるかを示す塩基アドレスを用いても、両者を併用してもよい。

。

[0041] 上記 (I) のステップでは、二次構造の動態を解析すべき対象区間 S が選択される。本実施形態においても、該長さ L_1 は、上述したものと同様の大きさであってよい。

対象区間 S の長さ L は、長さ L_1 より大きい限り限定されず、用途に応じて任意に決定してよく、例えば RNA の全長としてもよい。例えば、所望の局所的な二次構造の存在をタンパク質への翻訳を抑制する目的で予測する場合、翻訳抑制への寄与が大きいと予測される位置、例えば、5' 非翻訳領域 (5' UTR)、5' UTR とコード領域 (CDS) とにまたがる領域、好ましくは翻訳開始点およびその近傍などが選択され得る。一方、所望の局所的な二次構造の存在をタンパク質への翻訳を増大させる目的で予測する場合、mRNA の安定性への寄与が大きいと予測される位置、例えば、3' 非翻訳領域 (3' UTR) などが選択され得る。あるいは、フレームシフトなどの生命現象を引き起こすシュードノット構造 (特殊なステム・ループ構造) を見つける目的ならばコード領域が主に選択され得る。

[0042] 上記 (II) のステップでは、対象区間 S 内の一端 (すなわち、最も 5' 側) である位置 $x = x_1$ と、位置 $x = x_1 + L_1$ とを両端とする小区間 $S(x_1)$ を、上記のステップ (A) で選択される長さ L_1 を有する区間であると想定して、該小区間 $S(x_1)$ において上記のステップ (A) ~ (E) を行うことにより、該小区間 $S(x_1)$ において決定された各特定二次構造が、非平衡状態にある時間的区間内の各時点では何程の確率で存在し得るかを決定する。上記のステップ (A) ~ (E) の実行自体は、上述した通りに行えばよい。なお、(II) のステップおよび以下で説明する (III) のステップにおいては、特定二次構造の個数が 1 つである小区間が存在してもよい。特定二次構造が 1 つである小区間においては、二次構造の動態を解析すべき前記時間的区間内の任意の時点において、当該 1 つの特定二次構造が 100% の存在確率で存在するとすればよく、従って、上記のステップ (B) ~ (E) を行わなくてよい。

[0043] 上記 (III) のステップでは、先ず、上記位置 x_1 を、所定の微小長さ r だけ

他端側（すなわち、 $5' \rightarrow 3'$ 方向）へ移動させた位置を $x_2 = x_1 + r$ とし、位置 x_2 と位置 $x_2 + L - 1$ とを両端とする小区間を $S(x_2)$ とする。

微小長さ r は、小区間同士が互いに重なり合うように小区間の長さ $L - 1$ よりも短い値であればよいが、数多く重なり合わせることによって、後に詳述する蓋然性の値自体の信頼性を高めることができる。微小長さ r は、塩基 1 個に近い小さい値が好ましいが、演算処理すべき小区間の数がそれに応じて増加するので、用いる演算装置の処理能力や、必要な信頼性の程度、計算する対象である RNA のおかれた生物学的環境などに応じて適当な数値を選択すればよい。これらの点からは、微小長さ r の塩基数は 1 ～ 10 個、特に、1 ～ 3 個が好ましい範囲である。なお、微小長さ r は、どの程度の精度で蓋然性を算出するかに影響する因子であり、分解能（resolution）と考えることもできる。

[0044] 上記（II）のステップでの操作と同様にして、小区間 $S(x_2)$ においても上記のステップ（A）～（E）を行うことにより、該小区間 $S(x_2)$ において決定された各特定二次構造が、非平衡状態にある時間的区間内の各時点で何程の確率で存在し得るかを決定する。

このように、小区間を微小長さ r だけ他端側へ移動させて、そのときの小区間でステップ（A）～（E）を行うという操作を、小区間の一端が対象区間 S の反対端に到達したときの最後の小区間 $S(x_M)$ まで繰り返し行う。このような繰り返し操作によって、各小区間において決定される各特定二次構造が、非平衡状態にある時間的区間内の各時点で何程の確率で存在し得るかが決定される。 x_i の添え字 i は、位置 x_i が位置 x_1 を $(i - 1)$ 回微小長さ r だけ（すなわち、長さ $\{r \times (i - 1)\}$ だけ）他端側へ移動させた位置であることを示している。また、小区間 $S(x_i)$ は、位置 x_i と位置 $x_i + L - 1$ とを両端とする小区間である。

[0045] 生体内の条件を適切に算出するために、上記（II）および（III）のステップにおいて、上記ステップ（E）における解法の選択、初期条件を与える時点 t_0 、計算の終了時点 T_{fold} を各小区間毎にそれぞれ独立に与えてもよいし、或

いは、簡単のためにそれらを全ての小区間で統一させてもよい。

[0046] また、上記 (II) および (III) のステップでは、対象区間 S の一端から小区間の位置を順次移動させて順番に解析しているように表現しているが、その演算や操作は、必ずしも端から順番通りに行なうことだけを意味するものではない。即ち、対象区間 S から、互いに一部がオーバーラップした小区間 $S(x_1) \sim S(x_m)$ を順不同に選び出し、上記のステップ (A) $\sim$ (E) を行ってもよい。すなわち、上記 (II) および (III) のステップは、結果として、対象区間 S にそのような互いに重なり合った小区間が設定されてそれぞれについて解析された状態を得ることを意味する。演算やプログラムの進行においては、端から順番の解析が好ましい手順である。

[0047] ステップ (IV) では、先ず、上記 (II) $\sim$ (III) のステップでの各小区間において、各小区間毎に決定された各特定二次構造が、任意に選択された時間 t において何程の確率で存在し得るかが決定される。その決定のためには、上記 (II) $\sim$ (III) のステップで得られた結果を参照すればよい。

[0048] 続いて各小区間における各特定二次構造毎に、所定の特徴パラメータ値を有する任意のモチーフ M_k が存在するか否かが調べられる。ここで、特徴パラメータ値は、モチーフの位置および構造を表すものであり、従って、本明細書におけるモチーフは、構造だけでなく、RNA 上での位置をも含む概念であることに留意されたい。求めるべきモチーフの塩基数は、長さ $L-1$ 以下である限り特に制限はなく、後述のモチーフを規定する各特徴パラメータ値のとり得る範囲の組み合わせによってその下限と上限が定まる。例えば、モチーフが、ステムの長さが $x_1 \sim x_2$ 塩基、ループの長さが $y_1 \sim y_2$ 塩基、ステム・ループ内の許容ミスマッチ数が i 個のステム・ループの場合、モチーフの長さは $[x_1 \times 2 + y_1] \sim [x_2 \times 2 + y_2 + i]$ 塩基となる。好ましいモチーフの塩基数としては、例えば、 $10 \sim 100$ 塩基程度が挙げられ、 $20 \sim 60$ 塩基程度がより好ましい塩基数として挙げられる。特定二次構造におけるモチーフの存在または非存在の決定は、該特定二次構造が、該モチーフの特徴パラメータ値を満たす局所的な二次構造を含むか否かを単純

にパターンマッチングすることによって行うことができる。また、明らかにように、当該モチーフの全配列（あるいは、当該モチーフを構成するために必要な全配列；以下同様であり、以後は特に断らない。）を含まない小区間については、小区間内で決定されたいずれの特定二次構造にも当該モチーフは存在し得ないため、パターンマッチングは、当該モチーフの全配列を含む小区間においてのみ行えばよい。

[0049] ここで、モチーフとはすなわち局所的な二次構造であり、その二次構造の特徴をパラメータ値で規定し得るものであればよい。また、モチーフを規定するパラメータ値が、特徴パラメータ値である。

具体的なモチーフとしては、ステム・ループ構造、シュードノット (pseudo knot) 構造（偽結び目構造）、凹形構造などが挙げられる。

凹形構造とは、本発明者が重要な局所的な二次構造として着目したものであって、互いに接近した2つのステム・ループ構造の間に存在し、他の塩基配列と相互作用していない間隙部分を有する構造である。その間隙部分自体は平坦ではあるが、その両側にステムに関与する塩基が存在する点から、モチーフとしている。凹形構造は、グローイン (groin) 構造と呼ぶこともできる。

[0050] ステム・ループ構造を規定する特徴パラメータ値としては、RNA上での位置、ステムの長さ、ループの長さ、ステム・ループ内の許容ミスマッチの数が挙げられる。

ステム・ループ構造は、必ず一対以上の塩基対(ステム部分)および折り返し部位(ループ部分)を持つので、必然的にパラメータの最小値は決定される(すなわち、ステムの長さ1以上、ループの長さ3以上、ミスマッチの数0以上)。しかし計算のスピードを上げるため、パラメータの最小値を上げても良い。

ステムの長さ(塩基対形成に寄与する片方の鎖の塩基数を意味する。G-Uのwobblingを含む)は、特に上限値は定めなくてもよいが、塩基数4～50、より好ましくは5～20である。

ループの長さは、塩基数 3～50、より好ましくは 4～10 である。

ステム・ループ内の許容ミスマッチの数は、塩基数 0～30、より好ましくは 0～10 である。

- [0051] シュードノット構造を規定する特徴パラメータ値としては、ステム・ループを規定するパラメータに加え、主たるステム・ループ（より 5' 端側に存在するステム・ループ）のループ部分のうち 3' 端側の配列と相互作用する連続したある塩基数と、その部分に相補的な配列の主たるステム・ループから見た距離（塩基数）があげられる。

シュードノット構造は、必ず一對以上の塩基対（ステム部分）と折り返し部位（ループ部分）、および一對以上の相互作用部位（ループ部分と相互作用する部位）を持つので、必然的に上記パラメータの最小値は決定される（すなわち、ステムの長さ 1 以上、ループの長さ 3 以上、ミスマッチの数 0 以上、もう一つの連続した塩基数 1 以上、距離 2 以上）。しかし計算のスピードを上げるため、パラメータの最小値を上げてても良い。

ループ部分のうち 3' 端側の配列と相互作用する連続したある塩基数は 3～50、より好ましくは、5～20 である。

その部分に相補的な配列の主たるステム・ループの 3' 端からの距離（塩基数）は 4～50、より好ましくは 5～20 である。

- [0052] 凹形構造では、中心となる間隙部分の両側にそれぞれステム a、b が存在する。両側のステム a、b は、それぞれ別個に塩基対をなしているが、凹形構造を規定するには、ステム a、b のそれぞれの塩基対のうちの内側の塩基（すなわち、それぞれの間隙部分側の塩基）の数を指定するだけで充分である。よって、本発明では、凹形構造を規定する好ましい特徴パラメータ値として、中心となる間隙部分の塩基数と、その両側にある 2 つのステム・ループ構造のそれぞれのステムの内側部分の塩基数の和、および該ステムの内側部分に許容されるミスマッチの数を採用する。

凹形構造は、必ず間隙部分をはさんで形成されている 2 個以上の塩基対に参加する塩基（間隙部分をはさむ 2 本のステムで、間隙部分側の塩基）を持つ

ので、必然的に上記パラメータの最小値は決定される（すなわち、間隙部分 0 以上、ステムの内側部分の塩基数 2 以上、ミスマッチの数 0 以上）。しかし計算のスピードを上げるため、パラメータの最小値を上げてても良い。

間隙部分の塩基数は 0 ~ 50、より好ましくは 0 ~ 10 である。

間隙部分の両側にある 2 つのステムのそれぞれの内側部分の塩基数の和は、4 ~ 50、より好ましくは 4 ~ 30 である。

ミスマッチは 0 ~ 10、より好ましくは 0 ~ 5 である。

[0053] なお、モチーフの「位置」そのものは、点（塩基 1 つ分の位置）ではなく、モチーフを直線状に展開したときの塩基長をもった区間である。演算などの取り扱い上、モチーフの位置を 1 点の塩基アドレスで表した方が便利の場合には、その位置が占める配列区間の先頭などの 1 点の塩基の塩基アドレスと、モチーフの塩基長とによって、モチーフの位置を表現してもよい。

[0054] 上記のようにして、各小区間毎に、時間 t での各特定二次構造の存在確率と、各特定二次構造内にモチーフ M_k が存在するか否かが決定され、続いて、これらを用いて各小区間内において、時間 t においてモチーフ M_k が何程の蓋然性で存在するかが決定される。第 1 の蓋然性 (V_S) は、例えば以下のようにして算出することができる。

一般に、或る小区間 $S(x_i)$ ($i = 1, 2, 3 \dots M$) について、時間 t における特定二次構造 j ($j = 1, 2, 3 \dots N$) の存在確率を $P_{i,j}(t)$ とし、或る小区間 i について得られた特定二次構造全ての存在確率の和を、

[0055] [数5]

$$P_{i,total}(t) = \sum_{j=1}^N P_{i,j}(t) \quad (5)$$

[0056] とする。

さらに、そのようにして求められた $P_{i,total}(t)$ を用いて、

[0057]

[数6]

$$VS_i(M_k)(t)[\%] = \frac{100}{P_{i,total}(t)} \sum_{j=1}^N \begin{cases} + P_{i,j}(t) & \text{(時間 } t \text{ において小区間 } S(x_i) \text{ の特定二次構造 } j \text{ に } M_k \text{ が発見された場合)} \\ + 0 & \text{(時間 } t \text{ において小区間 } S(x_i) \text{ の特定二次構造 } j \text{ に } M_k \text{ が発見されない場合)} \end{cases} \quad (6)$$

[0058] を計算する。このようにして求められる $VS_i(M_k)(t)$ が、時間 t における小区間 i でのモチーフ M_k についての第 1 の蓋然性となる。また上記の場合、特定二次構造 j に M_k が発見された場合であっても、その M_k が特定二次構造 j において最も 5' 側または最も 3' 側の構造であることを理由に、発見されなかったものとして扱うこともできる。

ところで、これが平衡状態にある場合には、以下の式：

[0059] [数7]

$$P_{i,j}(\infty) = \frac{\exp\left(\frac{-\Delta G_{i,j}}{RT}\right)}{\exp\left(\frac{-\Delta G_{i,min}}{RT}\right)} \quad (7)$$

[0060] が成り立つ。ここでの $\Delta G_{i,j}$ とは、或る小区間 i について、特定二次構造 j のギブス自由エネルギーであり、 $\Delta G_{i,min}$ は或る小区間 i について、最も安定な特定二次構造のギブス自由エネルギーである。

上記式 (7) は、各特定二次構造の平衡状態での存在確率を求めるに際して、マクスウェル・ボルツマンの分布則を適用したものであり、これと上記式 (5) および (6) を用いて求められる平衡状態での第 1 の蓋然性は、上記特許文献 2 に開示された方法で求められる第 1 の蓋然性に対応するものである。本発明者は、上記特許文献 2 で開示した発明の方法では、この方針を採らなかったが、それは、小区間内の特定二次構造を全て発見できていないだろうことを意識した上で、簡便性を優先したためである。本発明の方法においては、時間の概念を入れて厳密性を追及するため、上記式 (7) を利用することが好ましい。

ここで例えば、或る小区間 $S(x_i)$ において、上記 (II) または (III) のス

トップで4つの特定二次構造 s_1 、 s_2 、 s_3 および s_4 が得られており、時間 t におけるそれらの存在確率をそれぞれ、 $P_{s1}(t)$ 、 $P_{s2}(t)$ 、 $P_{s3}(t)$ および $P_{s4}(t)$ とし、かつ、得られた特定二次構造全ての存在確率の和を $P_{total}(t)$ とする。さらに、特定二次構造 s_1 、 s_2 および s_3 はモチーフ M_{k1} を有するが特定二次構造 s_4 はモチーフ M_{k1} を有さず、かつ特定二次構造 s_1 および s_4 はモチーフ M_{k2} を有するが特定二次構造 s_2 および s_3 はモチーフ M_{k2} を有さないものとする。この場合、時間 t において、モチーフ M_{k1} が存在する蓋然性は；

$$\{P_{s1}(t) + P_{s2}(t) + P_{s3}(t)\} / P_{total}(t),$$

モチーフ M_{k2} が存在する蓋然性は；

$$\{P_{s1}(t) + P_{s4}(t)\} / P_{total}(t)$$

となる。

決定された蓋然性の値が、小区間 $S(x_i)$ での時間 t における上記各モチーフに関する第1の蓋然性である。上記の例の場合、

$$VS_i(M_{k1})(t) = \{P_{s1}(t) + P_{s2}(t) + P_{s3}(t)\} / P_{total}(t),$$

$$VS_i(M_{k2})(t) = \{P_{s1}(t) + P_{s4}(t)\} / P_{total}(t)$$

となる。

- [0061] 次に、上記 (V) のステップでは、上記 (IV) のステップにおいて各小区間内の各特定二次構造におけるモチーフ M_k の存在および非存在を調べた結果に基づいて、各小区間を、(a) 当該小区間においてモチーフ M_k を有する特定二次構造が少なくとも1つ存在する、または (b) 当該小区間においてモチーフ M_k を有する特定二次構造が存在しない、のいずれであるかについて評価する。当該評価は、前記の小区間 $S(x_1)$ から $S(x_M)$ の全ての小区間に対して行うべきであるが、実際には、当該モチーフの全配列を含まない小区間にはモチーフ M_k を有する特定二次構造が存在し得ないため、当該モチーフの全配列を含む小区間のみ考慮すればよい。モチーフ M_k に対して、どれ程の小区間が上記 (a) に評価されるかは様々であって、1つの小区間のみの場合から、モチーフ M_k の全配列を含む全ての小区間の場合まで、様々である。ここで

、上記（a）に評価される小区間の頻度を求め、これをモチーフ M_k に関する第2の蓋然性とする。

この第2の蓋然性は、上記（a）に評価される小区間の個数で表してもよいし、当該個数を、モチーフ M_k の全配列を含む全ての小区間の個数で割って100を乗じた数値で表してもよい。

より好ましくは、第2の蓋然性は以下の通りに算出される。すなわち、注目しているモチーフ M_k に関して、該モチーフの全配列が含まれる全ての小区間 $S(x_i)$ ($i = 1, 2, 3 \dots M$)の各々に対して重み係数 $W(i, M_k)$ を与え、下記式：

[0062] [数8]

$$VD_i(M_k)[\%] = \frac{100 \times \begin{cases} +W(i, M_k) \\ +0 \end{cases}}{W(M_k)} \quad (8)$$

[0063] （式中、“” { “記号の表記は、小区間 $S(x_i)$ について、いずれかの特定二次構造にモチーフ M_k が発見された場合は $W(i, M_k)$ を、いずれの特定二次構造にもモチーフ M_k が発見されない場合は0を加えることを意味する。）を計算する。ここで、式（8）における $W(M_k)$ は：

[0064] [数9]

$$W(M_k) = \sum_{i=1}^M W(i, M_k) \quad (9)$$

[0065] で与えられる。重み係数 $W(i, M_k)$ は、通常全ての i に対して1とすればよいが、変数としてもよい。 $W(i, M_k)$ を定数1とした場合には、 $W(M_k)$ は、モチーフ M_k の全配列が含まれる小区間の総数に等しい。さらに、 $VD_i(M_k)$ を、モチーフ M_k の全配列が含まれる全ての小区間にわたって総和することで、

[0066] [数10]

$$VD(M_k)[\%] = \sum_{i=1}^M VD_i(M_k) \quad (10)$$

[0067] を計算する。このようにして得られる $VD_i(M_k)$ を、第2の蓋然性とすることができる。 $VD_i(M_k)$ の算出は一見冗長であるが、後述する第3の蓋然性の算出を容易にし得るため、好ましい。

第1の蓋然性 VS_i が、各小区間内で評価した場合のモチーフの存在の蓋然性であるのに対し、第2の蓋然性 VD_i は、各小区間を互いに関連させることで得られたモチーフの存在の蓋然性である。

[0068] 上記 (VI) のステップでは、第1の蓋然性と第2の蓋然性とに基づいて、時間 t におけるモチーフ M_k の存在を決定する。例えば、第1の蓋然性と第2の蓋然性とが共に高いものが、実際に存在する可能性の高いモチーフであると決定することができる。あるいは、第1の蓋然性と第2の蓋然性とが共に高い位置がどこにあるかをより効果的に示すために、上記特許文献2に記載したのと同様にして、これら2つの蓋然性を1つのグラフ内で関連づけて視覚的に表示してもよい。視覚的な表示の方法については、本発明のプログラムにおいて後述する。さらに代替的には、第1の蓋然性と第2の蓋然性とかから新たな値（これを「第3の蓋然性 VI 」とする）を算出し、第3の蓋然性に基づいてモチーフの存在を決定することもできる。第3の蓋然性としては、例えば、第1の蓋然性と第2の蓋然性とを掛け合わせた後に総和をとり、その結果を100で割ることで得られる値を用いることができ、そのようにして算出された第3の蓋然性の値が大きい場合に、モチーフが存在する蓋然性が高いと決定され得る。その場合、第3の蓋然性を、例えば、上記した $VS_i(M_k)(t)$ および $VD_i(M_k)$ を用いて、：

[0069] [数11]

$$VI(M_k)(t)[\%] = \frac{1}{100} \sum_{i=1}^M \{VS_i(M_k)(t) \times VD_i(M_k)\} \quad (11)$$

[0070] を計算することによって求めることができる。

[0071] 次に、本発明の方法を実施するための装置の構成を具体的に示す。

図1は当該装置の構成の一例を模式的に示すブロック図である。上記 [9] および図1に示すとおり、当該装置は、第一演算部（Y1）、第二演算部

(Y 2)、第三演算部(Y 3)、第四演算部(Y 4)および第五演算部(Y 5)を少なくとも有して構成される。

[0072] 第一演算部(Y 1)は、上記ステップ(A)を行うための演算部である。第一演算部(Y 1)には、予測の対象とすべき長さL 1のRNA塩基配列、温度など、予測に必要な初期値が入力される。RNA塩基配列は、オペレータによるデータ名の入力に基づいて、自体の記憶装置(HDD、CD-ROMなど)や外部データベースにアクセスしデータを取り込むように構成されていてもよい。これらの初期値については、本発明の方法において説明したとおりである。第一演算部(Y 1)は、入力されたこれらの初期値に基づいて、予測の対象とする区間において、平衡状態における複数の特定二次構造ならびにそれらのギブス自由エネルギーおよび存在確率を算出する演算を行う。その後の解析に利用すべき二次構造の選択を可能とするために、第一演算部(Y 1)は、初期値として二次構造の選択の基準の入力を可能とするか、あるいは、予測の対象とされた区間での特定二次構造の候補ならびにそれらのギブス自由エネルギーおよび／または存在確率を出力し、オペレータによる特定二次構造の選択の入力を可能とするものであることが好ましい。

第二演算部(Y 2)は、上記ステップ(B)を行うための演算部である。第二演算部(Y 2)では、第一演算部(Y 1)で算出された特定二次構造に対して、それらの間での相互変換における遷移状態の二次構造が決定され、さらに各遷移状態の二次構造のギブス自由エネルギーが求められる。

第三演算部(Y 3)は、上記ステップ(C)を行うための演算部である。第三演算部(Y 3)では、第二演算部(Y 2)で求められた各遷移状態の二次構造およびそれらのギブス自由エネルギーから、各特定二次構造同士の間の相互変換の経路の活性化エネルギーが算出される。

第四演算部(Y 4)は、上記ステップ(D)を行うための演算部である。第四演算部(Y 4)では、第三演算部(Y 3)で算出された各活性化エネルギーを基にして、平衡状態に達するまでの時間的区間内の各時点における、

第一演算部（Y 1）で決定された各特定二次構造の存在確率を決定するための連立微分方程式が形成される。

第五演算部（Y 5）は、上記ステップ（E）を行うための演算部である。第五演算部（Y 5）には、第四演算部（Y 4）で得られた連立微分方程式を数值的に解くために、初期条件、終了時点 T_{fold} 、時間刻み幅 Δt など、解析に必要な値が適宜入力される。本発明の方法において説明したように、本発明では、連立微分方程式は上記（i）～（iv）のうちのいずれか1つの解法によって解かれ得る。従って、第五演算部（Y 5）は、これら4つの演算プロセスを組み込んでおり、かつオペレータによるこれら4つの解法のうちのいずれを選択するかの入力を可能とするものであることが好ましい。上記（iii）または（iv）の解法が選択された場合には、本発明の方法によって説明したような、操作の繰り返し回数Nが入力される。

上述した第一から第五演算部でなされる演算のプロセスは、本発明の方法において説明したとおりである。

[0073] 図1に示すように、第一から第五演算部は、一体のまたはそれぞれ独立したコンピュータとして構成するのが装置としての好ましい態様である。その場合、前記コンピュータに付帯する入力装置および出力装置によって入出力がなされ得るが、入力装置をさらにクライアント用のコンピュータとしてもよい。あるいは、当該装置は、予測結果の出力のための出力部（Y 6）をさらに有してもよい。出力部（Y 6）は、液晶ディスプレイモニターなどの各種画像表示装置、プリンターなどの各種印刷出力装置などであってよく、それぞれの出力装置には専用のコンピュータが付帯していてもよい。

[0074] 本発明の装置は、上述した（I）～（VI）のステップをさらに有する実施形態で本発明の方法を行うために、上記[10]に示すように、さらに上記（y 1）～（y 6）の手段を有してもよい。その場合の装置の構成の一例を図2に示す。

[0075] （y 1）の手段は、上記（I）のステップを行うための手段である。（y 1）の手段には、解析の対象とすべきRNAの一次構造データが入力され、か

つ該一次構造データから選択される長さ L の対象区間 S が入力される。一次構造データは、必ずしもRNA全体のデータである必要はなく、予測の対象とすべき対象区間 S の一次構造データのみが入力されてもよい。また、必要なRNAの一次構造データは、自体の記憶装置に格納しておくだけでなく、オペレータの命令に応じて、CD-ROMや、外部データベース（外部コンピュータ）へのアクセスによって供給されるように構成してもよい。そのような一次構造データの供給源としては、公知の遺伝子データベース（例えば、米国のNational Center for Biotechnology Information (NCBI) が提供するGenBank）などが挙げられる。

(y_2) の手段は、上記 (II) のステップを行うための手段である。(y_2) の手段は、(y_1) の手段で決定された対象区間 S 内の位置 x の一端を位置 $x = x_1$ として、位置 x_1 と位置 $x_1 + L_1$ とを両端とする小区間を $S(x_1)$ とし、上記の手段(Y_1)で選択される区間が前記小区間 $S(x_1)$ であるとして、上記の手段(Y_1) ~ (Y_5)を用いて、該小区間 $S(x_1)$ での前記各特定二次構造が、非平衡状態にある時間的区間内の各時点で何程の確率で存在し得るかを決定する。

(y_3) の手段は、上記 (III) のステップを行うための手段である。(y_3) の手段は、前記位置 x_1 を、所定の微小長さ r だけ他端側へ移動させた位置を $x_2 = x_1 + r$ とし、位置 x_2 と位置 $x_2 + L_1$ とを両端とする小区間を $S(x_2)$ とし、前記小区間 $S(x_1)$ の代わりに該小区間 $S(x_2)$ を用いる以外は前記(y_2)の手段と同様にして、該小区間 $S(x_2)$ での前記各特定二次構造が、非平衡状態にある時間的区間内の各時点で何程の確率で存在し得るかを決定する、という操作を、前記位置 $x = x_1$ 、 x_2 に続いて、小区間の一端が対象区間の他端に達するまで、位置 $x = x_3$ 、...、 x_M について繰り返し行い、それによって、最初の小区間 $S(x_1)$ から最後の小区間 $S(x_M)$ までの各小区間毎の前記各特定二次構造が、非平衡状態にある時間的区間内の各時点で何程の確率で存在し得るかを決定する。

(y_4) の手段は、上記 (IV) のステップを行うための手段である。(y

4) の手段には、存在を解析すべきモチーフの特徴パラメータ値が入力される。モチーフの特徴パラメータ値については、本発明の方法で説明したとおりである。(y 4) の手段は、該特徴パラメータ値を有するモチーフ M_k が、任意の時間 t において、各小区間毎に決定された各特定二次構造中に存在するか否かを調べ、その結果と、時間 t での各特定二次構造の存在確率とに基づいて、各小区間において、時間 t において該モチーフが何程の蓋然性で存在するか（すなわち、上述した「第1の蓋然性」）を決定する。

(y 5) の手段は、上記(V)のステップを行うための手段である。(y 5) の手段は、前記(y 4)の手段を用いて調べた、各小区間における各特定二次構造における前記モチーフ M_k の存在および非存在の結果に基づいて、各小区間を、(a) 当該小区間においてモチーフ M_k を有する特定二次構造が少なくとも1つ存在する、または(b) 当該小区間においてモチーフ M_k を有する特定二次構造が存在しない、のいずれであるかについて評価する。そして、上記(a)に評価される小区間の頻度を求めることによって、モチーフ M_k に関する第2の蓋然性を算出する。その詳細については、本発明の方法において上述したとおりである。

(y 6) の手段は、上記(VI)のステップを行うための手段であり、前記(y 4)および(y 5)の手段で求められた、第1の蓋然性と、第2の蓋然性に基づいて、前記時間 t における前記モチーフ M_k の存在を決定するための出力を行う。(VI)のステップを行うに際して、複数のモチーフに対して蓋然性を決定することが有用であり得る。その場合、複数のモチーフについての蓋然性の算出結果を同時に視覚化する機能を(y 6)の手段が有することが有利である。そのため、(y 6)の手段は、例えば、各モチーフに対応する各位置毎に第1の蓋然性と第2の蓋然性とを共に出力する、これら2つの蓋然性を1つのグラフ内で関連づけて視覚的に表示する、または、各モチーフに対応する各位置毎に第1の蓋然性と第2の蓋然性から第3の蓋然性を算出してこれを出力する、などの機能を有し得る。これらの機能については、本発明のプログラムにおいて後述する。

[0076] 次に、本発明の方法を実施するためのプログラムの構成を具体的に示す。

当該プログラムは、上記した本発明の装置をコンピュータで構成した場合に、該コンピュータで本発明の方法を実行させるように構成されたプログラムである。当該プログラムは、1台のコンピュータ内で実行される態様だけではなく、手元のコンピュータ内にメインのプログラムが置かれ、他のコンピュータ内の演算プログラムへ初期値が送られ、演算結果がメインのプログラムに戻るというように、複数のコンピュータ内で実行される態様であってもよい。

[0077] 当該プログラムは、図1にそのフローを示すとおり、基本的には、上述した本発明の方法の流れと全く同様であって、コンピュータを、少なくとも、上記(A)のステップを行う手段(P1)、上記(B)のステップを行う手段(P2)、上記(C)のステップを行う手段(P3)、上記(D)のステップを行う手段(P4)、および上記(E)のステップを行う手段(P5)として機能させるように構成される。

[0078] 手段(P1)では、予測の対象とすべき長さL1のRNA塩基配列、温度など、予測に必要な初期値が入力される。RNA塩基配列は、オペレータによるデータ名の入力に基づいて、自体の記憶装置(HDD、CD-ROMなど)や外部データベースにアクセスしデータを取り込むプログラム構成であってもよい。これらの初期値については、本発明の方法において説明したとおりである。手段(P1)は、入力されたこれらの初期値に基づいて、予測の対象とする区間において、平衡状態における複数の特定二次構造ならびにそれらのギブス自由エネルギーおよび存在確率を算出する演算を行う。その後の解析に利用すべき二次構造の選択を可能とするために、手段(P1)は、初期値として二次構造の選択の基準の入力を可能とするか、あるいは、予測の対象とされた区間での特定二次構造の候補ならびにそれらのギブス自由エネルギーおよび/または存在確率を出力し、オペレータによる特定二次構造の選択の入力を可能とするものであることが好ましい。

手段(P2)では、手段(P1)を用いて算出された特定二次構造に対し

て、それらの間での相互変換における遷移状態の二次構造が決定され、さらに各遷移状態の二次構造のギブス自由エネルギーが求める。

手段（P 3）は、手段（P 2）を用いて求められた各遷移状態の二次構造およびそれらのギブス自由エネルギーから、各特定二次構造同士の間での相互変換の経路の活性化エネルギーを算出する。

手段（P 4）は、手段（P 3）を用いて算出された各活性化エネルギーを基にして、平衡状態に達するまでの時間的区間内の各時点における、手段（P 1）を用いて決定された各特定二次構造の存在確率を決定するための連立微分方程式を形成する。

手段（P 5）では、手段（P 4）を用いて得られた連立微分方程式を数値的に解くために、初期条件、終了時点 T_{fold} 、時間刻み幅 Δt など、解析に必要な値が適宜入力される。本発明の方法において説明したように、本発明では、連立微分方程式は上記（i）～（iv）のうちのいずれか1つの解法によって解かれ得る。従って、手段（P 5）は、これら4つの演算プロセスを実装しており、かつオペレータによるこれら4つの解法のうちのいずれを選択するかを入力を可能とするものであることが好ましい。上記（iii）または（iv）の解法が選択された場合には、本発明の方法によって説明したような、操作の繰り返し回数 N が入力される。

上述した手段（P 1）から（P 5）が行う演算のプロセスは、本発明の方法において説明したとおりである。

[0079] 当該プログラムは、コンピュータをさらに、上記手段（P 5）を用いて得られた各特定二次構造の存在確率の時間変化を出力するための手段（P 6）として機能させるように構成されることが好ましい。手段（P 6）は、各特定二次構造毎に、各時点における存在確率を数値として出力するものであってもよく、あるいは、横軸を時間とし、縦軸には、各時点毎に、各特定二次構造の存在確率を各々異なる色を有する存在確率に比例した長さの棒として、連続的かつ重ならないように一定の順序で並べる（その場合、全ての棒について合計した長さは常に一定である）などして、二次構造の挙動の視覚的

な理解を可能とするように工夫してもよい。後者に類似した方法の一例は、図8～11に示されている。

[0080] 本発明のプログラムは、上記[6]に示すように、コンピュータをさらに、上述した(I)～(VI)のステップをそれぞれ行うための(p1)～(p6)の手段として機能させるように構成することもできる。そのような構成によって、上述した(I)～(VI)のステップをさらに有する実施形態で本発明の方法を行うことが可能となる。その場合のプログラムの構成の一例を図2に示す。

[0081] (p1)の手段は、上記(I)のステップをコンピュータに行わせるための手段である。(p1)の手段では、解析の対象とすべきRNAの一次構造データが入力され、かつ該一次構造データから選択される長さLの対象区間Sが入力される。一次構造データは、必ずしもRNA全体のデータである必要はなく、予測の対象とすべき対象区間Sの一次構造データのみが入力されてもよい。また、必要なRNAの一次構造データは、コンピュータの記憶装置に格納しておいたものだけでなく、オペレータの命令に応じて、CD-ROMや、外部データベース(外部コンピュータ)へのアクセスによって供給されるように構成してもよい。そのような一次構造データの供給源としては、公知の遺伝子データベース(例えば、米国のNational Center for Biotechnology Information (NCBI)が提供するGenBank)などが挙げられる。

(p2)の手段は、上記(II)のステップをコンピュータに行わせるための手段である。(p2)の手段は、(p1)の手段を用いて選択された対象区間S内の位置xの一端を位置 $x = x_1$ として、位置 x_1 と位置 $x_1 + L - 1$ とを両端とする小区間を $S(x_1)$ とし、上記の手段(P1)で選択される区間が前記小区間 $S(x_1)$ であるとして、上記の手段(P1)～(P5)を用いて、該小区間 $S(x_1)$ での前記各特定二次構造が、非平衡状態にある時間的区間内の各時点で何程の確率で存在し得るかを決定する。

(p3)の手段は、上記(III)のステップをコンピュータに行わせるための手段である。(p3)の手段は、前記位置 x_1 を、所定の微小長さrだけ他

端側へ移動させた位置を $x_2 = x_1 + r$ とし、位置 x_2 と位置 $x_2 + L_1$ とを両端とする小区間を $S(x_2)$ とし、前記小区間 $S(x_1)$ の代わりに該小区間 $S(x_2)$ を用いる以外は前記 (p 2) の手段と同様にして、該小区間 $S(x_2)$ での前記各特定二次構造が、非平衡状態にある時間的区間内の各時点で何程の確率で存在し得るかを決定する、という操作を、前記位置 $x = x_1, x_2$ に続いて、小区間の一端が対象区間の他端に達するまで、位置 $x = x_3, \dots, x_M$ について繰り返し行い、それによって、最初の小区間 $S(x_1)$ から最後の小区間 $S(x_M)$ までの各小区間毎の前記各特定二次構造が、非平衡状態にある時間的区間内の各時点で何程の確率で存在し得るかを決定する。

(p 4) の手段は、上記 (IV) のステップをコンピュータに行わせるための手段である。(p 4) の手段では、存在を解析すべきモチーフの特徴パラメータ値が入力される。モチーフの特徴パラメータ値については、本発明の方法で説明したとおりである。(p 4) の手段は、該特徴パラメータ値を有するモチーフ M_k が、任意の時間 t において、各小区間毎に決定された各特定二次構造中に存在するか否かを調べ、その結果と、時間 t での各特定二次構造の存在確率とに基づいて、各小区間において、時間 t において該モチーフが何程の蓋然性で存在するか（すなわち、上述した「第 1 の蓋然性」）を決定する。

(p 5) の手段は、上記 (V) のステップをコンピュータに行わせるための手段である。(p 5) の手段は、前記 (p 4) の手段を用いて調べた、各小区間における各特定二次構造における前記モチーフ M_k の存在および非存在の結果に基づいて、各小区間を、(a) 当該小区間においてモチーフ M_k を有する特定二次構造が少なくとも 1 つ存在する、または (b) 当該小区間においてモチーフ M_k を有する特定二次構造が存在しない、のいずれであるかについて評価する。そして、上記 (a) に評価される小区間の頻度を求めることによって、モチーフ M_k に関する第 2 の蓋然性を算出する。その詳細については、本発明の方法において上述したとおりである。

(p 6) の手段は、上記 (VI) のステップをコンピュータに行わせるため

の手段であり、前記（p 4）および（p 5）の手段で求められた、第 1 の蓋然性と、第 2 の蓋然性とに基づいて、前記時間 t における前記モチーフ M_k の存在を決定するための出力をコンピュータに行わせる。（VI）のステップを行うに際して、複数のモチーフに対して蓋然性を決定することが有用であり得る。その場合、複数のモチーフについての蓋然性の算出結果を同時に視覚化する機能を（p 6）の手段が有することが有利である。そのため、（p 6）の手段は、コンピュータを、例えば、各モチーフに対応する各位置毎に第 1 の蓋然性と第 2 の蓋然性とを共に出力するように機能させる、これら 2 つの蓋然性を 1 つのグラフ内で関連づけて視覚的に表示するように機能させる、または、各モチーフに対応する各位置毎に第 1 の蓋然性と第 2 の蓋然性から第 3 の蓋然性を算出してこれを出力するように機能させる、などのプログラム構成であり得る。

第 1、第 2 の蓋然性の出力の仕方は、利用者が両方の値を考慮し易いように表示され得るものであればよい。例えば、第 1、第 2 の蓋然性のそれぞれの値を、対象区間 S に対応させて単純に表形式で並列的に出力し、それぞれの値について、規定値よりも高い値にマーキングを施して示すという態様や、第 1、第 2 の蓋然性のそれぞれに基準を設け、両方の基準を満たす位置をピックアップして示すという態様であってもよい。

第 3 の蓋然性の出力の仕方は、第 1、第 2 の蓋然性の出力をもって成しても良いし、利用者が第 3 の蓋然性の値を考慮し易いように表示され得るものであればよい。例えば、対象区間 S に対応させて単純に第 3 の蓋然性の値を順番にリスト出力する態様や、あるいは、対象区間 S 内の各位置を横軸、各位置に対応する第 3 の蓋然性を縦軸としたグラフで表示する態様であってもよい。

また、第 1 の蓋然性と第 2 の蓋然性とを 1 つのグラフ内で関連づけて視覚的に表示させる方法としては、上記特許文献 2 に記載されたものと同様の方法が挙げられる。該方法は、対象区間 S が一方の軸に対応したグラフ中に、第 1、第 2 の蓋然性を同時に描き、二次元、三次元のグラフィックなデザイ

ンにて出力するものである。

そのような表示例を次に示す。なお、以下の説明は、「特定の時点 t における」第 1、第 2 の蓋然性を、グラフ中に表示する例であることに留意されたい。適切な特定の時点を選択して該特定の時点 t でのモチーフの存在または非存在を決定することにより、生体内で RNA が実際に取っている可能性が高い二次構造であって、RNA の機能の発現に重要であると考えられる当該二次構造を予測することが可能となる。計算上は、通常、そのような特定の時点 t を T_{fold} とした上で、時点 $t = T_{\text{fold}}$ における第 1、第 2 の蓋然性をグラフ表示すればよい。

[0082] 本例は、対象区間 S が一方の軸に対応したグラフ中に、棒グラフの形式で第 1 の蓋然性を表し、同時に、第 2 の蓋然性をも示す例である。

例えば、(p 4) の手段の演算において、ある小区間内のある位置に第 1 の蓋然性が与えられる場合、グラフ内の対象区間 S 上の対応する位置に棒グラフの棒（グラフ棒）を表示し、そのグラフ棒の長さで個々の第 1 の蓋然性の大きさを表す。

小区間 $S(x_1) \sim S(x_M)$ においてそれぞれ得られた 1 以上の位置を、グラフの対象区間 S に対応させて、第 1 の蓋然性の値をグラフ棒として描いていくと、位置によっては、グラフ棒同士の重なり合いが多数生じることになる。

そこで、グラフ棒同士の重なり合いの程度が高いほど、すなわち、第 2 の蓋然性が高いほど、その重なり合った部分の色彩および／または模様を変えて表示する。

[0083] 色彩の変化は、色相（単色光の波長に相当するもの）、彩度（あざやかさ、すなわち白みを帯びていない度合）、明度（明るさ、すなわち光の強弱）から選ばれる任意の要素を変化させればよい。例えば、第 2 の蓋然性がより高いグラフ棒の色をより濃く示したり、グラフ棒の色を青色から赤色へ変化させて示す、モニター画面では最重要部分を点滅させるなど、見る者が識別し易いように種々の工夫を凝らして表示してよい。表示装置が白黒の場合に

は、ハッチングやあみがけなどの種々の模様が識別には有用となる場合もある。

このように表示することによって、第1の蓋然性については、グラフ棒の長さで直感でき、第2の蓋然性については、グラフ棒の色彩や模様の違いで直感できる。例えば、グラフ棒がより長くより濃い位置に、実際にモチーフが存在する蓋然性が高いと予想できる。

このような予測結果を表示するための手段は、演算手段とは別の独立した表示専用のプログラムとしてもよい。

[0084] 表示におけるオプションとしては、例えば、第2の蓋然性の値（グラフ棒同士の多重度）の閾値を変更できるようにし、その閾値以上の多重度を持っていないモチーフは表示しないようにしてもよい。それによって、色の薄い構造は表示されなくなる。

また、第1の蓋然性の閾値を変更できるようにし、その閾値よりも低い第1の蓋然性を示す構造については表示しないようにしてもよい。第1の蓋然性を各モチーフの ΔG を用いた関数 $f(\Delta G)$ で表した場合には、関数 $f(\Delta G)$ の値がより小さい方が、第1の蓋然性が高いことを意味する場合がある。その場合には、閾値よりも高い $f(\Delta G)$ の構造を表示させないようにする。これらの操作によって、存在する確からしさが低いモチーフを考慮に入れないようにすることができる。

上記のような閾値による表示変更は、第3の蓋然性について行ってもよい。その操作によって、総合的な観点から存在する確からしさが低いモチーフを考慮に入れないようにすることができる。

[0085] モチーフの予測結果をスペクトルとして表示する方法は、本発明の有用性が顕著となる特徴的な表示方法である。しかし、グラフ棒が長くかつ十分に重なって示された部分が、同じモチーフの重なり合いに由来するものなのか、あるいは、互いに異なるモチーフの抽出結果がたまたま重なって表示されているのかがわからない場合がある。

予測の信頼性を向上させるには、グラフ棒が多く重なって示された部分が

、同じモチーフが重なり合ったことに由来することが明示される必要がある。以下に、その明示のための方法を二つ提案する。

[0086] 1つの方法は、モチーフがステム・ループ構造や凹形構造などステムを持つ構造の場合に有用な方法であって、グラフに表示される全ての位置にあるモチーフについて塩基の状態を調べ、次の2種類の塩基 (i)、(ii) を、互いに識別可能なように (例えば、色彩などを変えて) グラフに重ねて表示するという方法である。

(i) 常に塩基対 (ステム) を形成しない塩基 (フリー塩基と呼ぶ)。

(ii) あるモチーフ中においては塩基対を形成しない塩基であるが、他のモチーフ中においては塩基対形成に参加している塩基 (偽フリー塩基と呼ぶ)。

例えば、モチーフが最も単純なステム・ループ構造であって、同じ構造のモチーフだけが重なり合って第2の蓋然性が高く表示された場合、その位置の中央には、フリー塩基 (ループ部分に対応する) だけが表示され、偽フリー塩基は表示されない可能性が高いはずである (注意すべき例外は後述する)。このように、フリー塩基と偽フリー塩基とを参照すれば、第2の蓋然性の高さが、同じモチーフの重なり合いに由来するかどうかを判断する指標になる。

[0087] もう1つの方法は、グラフに表示された位置の配列を、もう一度構造予測にかけるなどしてそのモチーフ自体の安定性値 ΔG 値を出し、それをグラフに重ねて表示する方法である。凹形構造についてはモチーフ自体が安定性値を持たないので、間隙部分をはさむステム・ループ構造二本の安定性値 ΔG の和をもってモチーフの安定性値としてもよい。

第2の蓋然性の高さが、同じモチーフの重なり合いに由来するならば、その重なり合っている構造の ΔG 値はすべて同一となり、逆に棒グラフが複数の構造に由来していれば重なり合っている構造の ΔG 値はそれぞれ互いに異なっている可能性が非常に高い。

また、この方法を用いて算出された各モチーフ M_k の ΔG について閾値を与

え、画面上で不安定な構造を表示させないようにすることもできる。

[0088] 次の二点に注意が必要である。

一点は、対象区間 S の両端では、第 2 の蓋然性が低くなる（重なり合いの度合いが少なくなる）ことである。これはただ単に、端部では、小区間の重なり合う数が少なくなることによることに起因する。そのため、対象区間 S の端部に近い二次構造を議論するときには、第 2 の蓋然性の低さが、小区間の重なり合う数が少ないことによるのか、その構造の第 2 の蓋然性が低いためなのかを詳細に調べればよい。或いは、上述したように、第 1 の蓋然性の算出にあたって、最初からこうした構造を除いても良い。

他の一点は、フリー塩基と偽フリー塩基とを参照する場合に、偽フリー塩基が確認されたからといって、その位置のモチーフが全て存在の蓋然性が低いとは限らないことである。例えば、一つの配列に対して十分な確からしさで存在するモチーフ M_1 が予測されていたとする。ところが、フレームの関係で該モチーフ M_1 を形成する配列の全長が含まれないフレームにおいては、該モチーフ M_1 を取れず、かわりに該配列が参加する形で別のモチーフ M_2 が予測される場合がある。こうした場合も最後の出力の結果としては、違う構造の重なりとして出力され、この二つの構造について互いに偽フリー塩基が出力される。しかしながら、この場合はモチーフ M_1 の蓋然性がより高いことは自明である。このような M_2 については上述の第一あるいは第 2 の蓋然性に対する閾値を設定することで除くことができる場合が多いが、より厳密には詳細を調べればよい。

[0089] また、RNA 上の各位置に対して、その位置を含む全ての小区間にわたって、各小区間を代表するギブス自由エネルギーを総和し、それをさらに、その位置を含む小区間の総数で割ることで得られた値を、その位置での RNA の一般的な構造化の度合いとして表示してもよい。ここで、小区間を代表するギブス自由エネルギーとは、その小区間において求められた各特定二次構造のギブス自由エネルギーに、各特定二次構造の時間 t での存在確率をかけ、該小区間での全ての特定二次構造にわたって総和したものである。

つまり、時間 t での或る小区間 i について、時間 t における特定二次構造 j の存在確率を $P_{i,j}(t)$ とし、その構造 j ($j = 1, 2, \dots, N$) のギブス自由エネルギーを $\Delta G_{i,j}$ とした場合、ある小区間 i を代表するギブス自由エネルギー $\Delta G_i(t)$ は、

[0090] [数12]

$$\Delta G_i(t) = \sum_{j=1}^N P_{i,j}(t) \times \Delta G_{i,j} \quad (12)$$

[0091] となる。さらに、或る位置 x に関して、位置 x を含む全ての小区間 i ($i = 1, 2, \dots, M$) にわたって、各小区間を代表するギブス自由エネルギーを総和し、それをさらに、その位置を含む小区間の総数 M で割ればよい。

[0092] [数13]

$$\Delta G(x)(t) = \frac{\sum_{i=1}^M \Delta G_i(t)}{M} \quad (13)$$

[0093] こうして求められた $\Delta G(x)(t)$ をスペクトル中に折れ線グラフとして表示してもよい。さらに、モチーフの存在する位置の $\Delta G(x)(t)$ について閾値を与え、それを満たさない位置にかかるモチーフは画面上で表示させないようにしてもよい。

[0094] 本プログラムの実施に際して、種々のパラメータの入力は、研究者が理解しやすい形で為されることが望ましい。さらに、その入力メソッドが、研究者の計算に対する理解を助けるように配置されていることが望ましい。そこで、たとえば計算の概要を示すようなアイコンと共に、それらの計算ステップで必要とされるパラメータを直観的に入力できるパネルが用意されていても良い。そのようなパネルの一例を図3に示す。

[0095] この計算は科学計算であるが、本プログラムはそうした科学計算に慣れていない実験系研究者をも対象にしている。そのため、計算の進捗にあたっては、上述したパラメータ設定のアイコンと対応するアイコンを表示し、計算がどのように行われているかを示して、研究者の計算に対する理解を助けて

もよい。各ステップの進捗がバーで示されると同時に、終了したステップにはチェックマークが入り、計算の進行を体感できるようにしても良い。その一例を図4に示す。

- [0096] また、本計算で行われる連立微分方程式の数値計算は、新たに示す「数値解析をイメージさせるアイコン」と共に進捗バーを表示することで、計算の意味と進捗を体感できるようにしても良い。その一例を図5に示す。

実施例

[0097] 実施例1：ADM mRNAの構造

アドレノメデュリン（ADM）の5' UTRには特徴的なステム・ループ構造があり、その構造がADMの翻訳を制御しているといわれている（Brenet et al., Oncogene 25(49): 6510-9 (2006)）。その構造は実験的に見つかり、そうした有意な構造をコンピュータを用いて予測することは非常に意味がある。

- [0098] そこで、本実施例1では、ADM mRNA配列（GenBank Accession No. NM_001124）について、3塩基（ $r=3$ ）ずつずらしながら200塩基（ $L=200$ ）の小区間で区切って構造予測を行い、各小区間について一つまたは複数の構造予測結果（すなわち、特定二次構造）を得た。

この構造予測結果から、妥当性が高いステム・ループを抽出した（抽出法の詳細については、上記Brenet et al., Oncogene 25(49): 6510-9 (2006)を参照）。そのとき、ステム・ループの評価には、各小区間での構造予測結果の存在確率（ポピュレーション）が必要になる。この「各小区間での構造予測結果の存在確率（ポピュレーション）」を以下の3つの計算方法で取得した。

第1の計算方法は、S. Nakamura, J. Biochem., Apr 22 (2009) [Epub ahead of print], PMID: 19386779 (<http://jb.oxfordjournals.org/cgi/reprint/mvp064?ijkey=UAuE0vg4VWCEPzL&keytype=ref>から入手可能)に記載されている既存の方法である。すなわち、構造予測に用いたソフトウェアGCG-mfoldで出力されるエネルギー値を用いて、それぞれの構造群の相互変換が平衡状

態にあるとしてポピュレーションを計算した。

第2の計算方法では、用いるエネルギー値の算出を、新たにRNA構造予測システムGeno Poemics™ (GP) 内部に搭載されているエネルギー計算ルーチンを用いて（計算手法は、以下の論文： Mathews et al, J. Mol. Biol. (1999) 288, 911-940. ; Xia T et al, Biochemistry (1998) 37, 14719-14735. ; Giese MR et al, Biochemistry (1998) 37, 1094-1100. ; Draper DE et al, Methods in Enzymology (1995) 259, 281-242. ; および、Serra MJ et al, Methods in Enzymology (1995) 259, 242-261. を参考にした）、構造群の相互変換が平衡状態にあるとしてポピュレーションを計算した。

第3の計算方法では、用いるエネルギー値の算出を、新たにGP内部に搭載されているエネルギー計算ルーチンを用いて、構造群の相互変換が平衡状態に達していない（非平衡状態）としてポピュレーションを計算した。

この非平衡状態の計算では、各構造のポピュレーションを計算するにあたり、初期値から微分方程式を数值的に解いた。初期値としては、各小区間において得られた構造予測結果が等しい比率で存在すると仮定することで得られる存在確率を用いた。さらに、時間刻み幅 $\Delta t=0.001$ 秒、計算の終了時点 $T_{\text{fold}}=3$ 秒、温度 37°C に設定の上、反応速度 k を算出するために、アイリングの絶対反応速度式を使用した（プランク定数 h (6.62621×10^{-34} J s)、ボルツマン定数 k_B ($1.3806504 \times 10^{-23}$ J K $^{-1}$)、気体定数 R (8.31447 J K $^{-1}$ mol $^{-1}$)、アイリング定数 K (1.00))。

微分方程式を数值的に解くにあたっては一次近似を採用した。さらに、ある構造間、例えば構造Aと構造Bとの間で $k_{A \rightarrow B} \times \Delta t$ と $k_{B \rightarrow A} \times \Delta t$ のどちらか一方が1.0以上であった場合には、その構造間での変換は Δt の時間内に既に平衡に達しているとした（その構造間では数值的に解かずにマクスウェル・ボルツマン分布によるポピュレーション配分を行った）。ここで、上記の関係が繋がって複数にわたる場合には、その複数の繋がった構造間で平衡に達していると考え（たとえばAとB、BとCに上記の関係があった場合には、A、BおよびCの3つの構造間で平衡に達していると考え。一方、A

とB、CとDのそれぞれに上記の関係があった場合には、A－B間とC－D間のそれぞれの間が独立して平衡状態に達していると考える）。

[0099] 上記3つの方法でのそれぞれの計算結果について、GPを用いてそれぞれスペクトル絵を作成した（第1から第3の計算方法が、それぞれスペクトル1～3に対応している）。簡単のため、VI=90%の閾値を適用した結果を図6に示す（すなわち、第3の蓋然性が90%以上のモチーフのみ表示している）。非平衡状態での計算から得られたスペクトル3については、時点 $t = T_{\text{fold}}$ における結果を示している。ここで、上記Brenet et al., Oncogene 25(49): 6510-9 (2006)では、40-80付近のステム・ループ構造が重要であると記載されている。スペクトルを見ると、スペクトル1～3のいずれにおいてもそのステム・ループ（該図中では、楕円で囲まれている）が抽出されていることが分かる。しかし、スペクトル3については、出力されているステム・ループ数がそもそも少ない。これは、スペクトル3はスペクトル1、2に比べてノイズが少ないことを意味している。

[0100] 以上より、本発明に従った非平衡状態での計算によって、mRNAの5' UTRを解析するにあたって、従来の方と比べて改善が得られることが分かった。また、本実施例では注目すべきステム・ループ以外をノイズと称したが、このように厳しい条件下でも残ってくるピークは、何らかの未知の生体内機能を担っている有意なものである可能性が考えられる。

[0101] 実施例2: CD83 mRNAの構造

CD83のCDSには特徴的なステム・ループ構造があり、その構造がHuRと相互作用することでCD83の翻訳が制御されているといわれている（Prechtel et al., J. Biol. Chem. 281(16): 10912-25 (2006)）。その構造は実験的に見つかり、そうした有意な構造をコンピュータを用いて予測することは非常に意味がある。

[0102] そこで、本実施例2では、CD83 mRNA配列（GenBank Accession No. NM_004233）についても同様に、3塩基（ $r=3$ ）ずつずらしながら200塩基（ $L1=200$ ）の小区間で区切って構造予測を行い、各小区間について一つまたは複数の構

造予測結果（すなわち、特定二次構造）を得た。この構造予測結果から、妥当性が高いステム・ループを抽出した（抽出法の詳細については上記Brenet et al., *Oncogene* 25(49): 6510-9 (2006)を参照）。そのとき、ステム・ループの評価には、各小区間での構造予測結果の存在確率（ポピュレーション）が必要になる。この「各小区間での構造予測結果の存在確率（ポピュレーション）」を次の3つの計算方法で取得した。

第1の計算方法は、S. Nakamura, *J. Biochem.*, Apr 22 (2009) [Epub ahead of print], PMID: 19386779 (<http://jb.oxfordjournals.org/cgi/reprint/mvp064?ijkey=UAuE0vg4VWCEPzL&keytype=ref>から入手可能)に記載されている既存の方法である。すなわち、構造予測に用いたソフトウェアGCG-mfoldで出力されるエネルギー値を用いて、それぞれの構造群の相互変換が平衡状態にあるとしてポピュレーションを計算した。

第2の計算方法では、用いるエネルギー値の算出を、新たにRNA構造予測システムGeno Poemics™ (GP) 内部に搭載されているエネルギー計算ルーチンを用いて（計算手法は、以下の論文：Mathews et al, *J. Mol. Biol.* (1999) 288, 911-940. ; Xia T et al, *Biochemistry* (1998) 37, 14719-14735. ; Giese MR et al, *Biochemistry* (1998) 37, 1094-1100. ; Draper DE et al, *Methods in Enzymology* (1995) 259, 281-242. ; および、Serra MJ et al, *Methods in Enzymology* (1995) 259, 242-261. を参考にした）、構造群の相互変換が平衡状態にあるとしてポピュレーションを計算した。

第3の計算方法では、用いるエネルギー値の算出を、新たにGP内部に搭載されているエネルギー計算ルーチンを用いて、構造群の相互変換が平衡状態に達していない（非平衡状態）としてポピュレーションを計算した。この非平衡状態での計算は、上記実施例1の方法と全く同じ方法を採用した。

[0103] 上記3つの計算方法でのそれぞれの計算結果について、GPを用いてそれぞれスペクトル絵を作成した（第1から第3の計算方法が、それぞれスペクトル4～6に対応している）。簡単のため、VI=90%の閾値を適用した結果を図7に示す（すなわち、第3の蓋然性が90%以上のモチーフのみ表示している）

。非平衡状態での計算から得られたスペクトル6については、時点 $t=T_{\text{fold}}$ における結果を示している。ここで、Prechtel AT et al., J Biol Chem. 2006; 281(16): 10912-25. では、540-580付近のステム・ループ構造がHuRとの相互作用に重要であると記載されている。また同文献では、440-510付近のステム・ループはHuRとの相互作用において重要ではないと記載されている。ここで、ノイズの減小と同時に、これら二つのステム・ループの差別化が目で見えれば、本発明に従った非平衡状態での計算の優位性が示される。

スペクトルを見ると、どのスペクトルでも540-580付近のステム・ループ（該図中では、楕円で囲まれている）が抽出されていることが分かる。しかし、スペクトル4では440-510付近のステム・ループ（該図中では、×印が付されている）が出力されているのに対して、スペクトル5、6では、明確に440-510付近のステム・ループの出力がない。一方、ノイズについて見ると、スペクトル4とスペクトル5ではピークの数等しく、ノイズの点では同程度である。しかし、スペクトル6においては、明確にピーク数が減少して4本となっている。これらのことから、本発明に従った非平衡状態での計算によって、重要な部分が差別化できると共に、ノイズの減少も達成できることが示された。

[0104] 以上より、本発明に従った非平衡状態での計算によって、mRNAのCDSを解析するにあたって、従来の方法と比べた改善が得られることが分かった。また、本実施例では注目すべきステム・ループ以外をノイズと称したが、このように厳しい条件下でも残ってくるピークは、何らかの未知の生体内機能を担っている有意義なものである可能性が考えられる。

[0105] 実施例3：ADM mRNAの構造解析からのmRNAの挙動についての検討

実施例1で示したように、ADM mRNAのモチーフについてスペクトルを得ることができた。本実施例3では、本発明に従った計算手法が、mRNAの構造的挙動を解析するにあたって、これまでにない視点を提供できることを示す。

[0106] 本手法は、mRNAの各小区間について予測される安定な二次構造の存在確率を、連立微分方程式を解いていくことで算出する。すなわち、これは存在確

率の時間変化を追っていることに他ならない。実施例 1 および 2 のように特定の時点でのスペクトル絵を生成するのではなく、本実施例では、個別の小区間毎の特定二次構造の存在確率の時間変化としてプロットした(その例を、図 8 ~ 11 に示す)。

[0107] このプロットの様子は下記の (1) ~ (4) のケースに大別された。計算にあたっては、 T_{fold} を 3.0 秒としている。

(1) $T_{\text{fold}} = 3.0$ 秒以内で、ほぼ平衡状態に達する系 (例: 図 8)。

(2) 平衡状態に向かって進みつつも、 $T_{\text{fold}} = 3.0$ 秒では平衡に達していない系 (例: 図 9)。図 9 に示した 2 つの例では、プロット線は、収束地点 (平衡状態) に向かって矢印の方に進んではいるが、収束までにはあと数十秒 ~ 数分かかると思われる。

(3) 0.001 秒以内 (最小刻み幅 Δt 内) に平衡状態に達してしまう系 (例: 図 10)。構造間の活性化エネルギーが極めて低く、構造間の変換がとても速い (速度定数 k が大きい)。図 10 に示した例では、活性化エネルギーは 2.57-5.72 kcal/mol である。

(4) 全く平衡状態に向かって進んで行かない系 (例: 図 11)。構造間の活性化エネルギーが高く、構造間の変換がとても遅い (速度定数 k が小さい)。図 11 に示した例では、活性化エネルギーは 23.8-51.18 kcal/mol である。事実上平衡状態に達しない系であり、ポピュレーションは初期値に大きく引っ張られる。この系のポピュレーションを考えるに、各特定二次構造が平衡状態にあると仮定することは本質的に正しくないと考えられる。

実際には、単一の小区間であってもこの 4 種類が混在している場合もある。例えば、ある小区間について 6 種の特定二次構造が予想されたとして、そのうち 5 種についてはそれぞれの構造間の変換が速くそれらの間で平衡に達するが、残りの 1 種がその他との構造変換を事実上起こさず全く平衡に向かって進まないという系もある。

[0108] ここで上記 (4) のケースについて検討する。

上述したように、この系を平衡状態にあると仮定してポピュレーション計

算することは本質的に正しくない。系が平衡状態にあると仮定してポピュレーション計算してよいかどうかをチェックするためには、二次構造のエネルギー量だけでなく、二次構造間の遷移状態の二次構造を決定し、各相互変換における活性化エネルギーを評価する必要がある。それは、本発明の手法を用いることで達成することができる。

この系でのポピュレーションは初期値（各構造の最初のポピュレーション）に大きく左右される。ここで、自然界の現象としてmRNAの挙動を見た場合に、毎回同じ初期値をとっている保証はない（というよりも同じ初期値をとっている理由がない）。とすると、毎回構造群が同じポピュレーションをとるわけではないということになる。さらに「毎回同じポピュレーションになるわけではない」部位の構造が生物学的に意味を持つ可能性は低いと考えられる。従って、このような構造は基本的に無視されるべきだろう。

実は、実施例1および2において、本発明の方法が有意にノイズが少ないのは、半ば自動的必然的に「毎回同じポピュレーションになるわけではない部位の構造を重視しない」という機能が働いているからである。つまり、実施例1および2においては、予測構造のポピュレーションに対して同一の初期値を与えている。平衡状態であれば予測構造のうち最安定なものが極めて優勢なポピュレーションをとるのであろうが（その場合、その最安定構造に含まれるモチーフは妥当性が高いものとして出力される）、平衡に達しない状態では、どの構造も優勢にならないのである。結果として、妥当な閾値（例えば、VI=90%）を設定すると、この部位の構造は完全に無視される。もちろん、平衡状態に達しない構造の中に共通モチーフがあって、それが有意なポピュレーションをとるのであれば、その共通モチーフは出力される。

GPシステムの考え方は、構造が一つに定まらないものを、複数の構造がどれくらいのポピュレーションで存在するかというプロファイルを定量的に算出し、そこから共通モチーフを抽出する（そしてそのモチーフの妥当性を算出する）ことにある。本実施例の手法は、そのプロファイル自体が定まるかどうかをチェックする方法として用いることができる。

[0109] 振り返るに、あるRNAの安定構造を論じるにあたって、従来、その安定構造のギブス自由エネルギー量だけを評価することが多かった。そこに、複数の安定構造が同時に存在し得るという事実を正面から捉え、複数の安定構造のポピュレーションのプロファイルから、共通モチーフの妥当性を定量的に算出および画像化するのを可能にしたのがGPシステムである。さらに、各安定構造間の遷移状態の構造を決定し、各相互変換の活性化エネルギーを評価することで、複数の安定構造のポピュレーションプロファイルが定まるかどうかを算出する（定まるのであればより確からしい値を算出する）のが本発明の方法である。

つまり、mRNAの構造を調べるということについては、その安定性だけを論じる時代、安定性だけから妥当性まで算出する時代、そして安定構造だけで論じるのではなく安定構造間の遷移状態の構造をも加味して妥当性を算出する時代、と順を追って進んできたとも言える。本発明の方法は、前記最後の、安定構造間の遷移状態の構造をも加味した妥当性の算出を実現できる現時点で唯一無二のものである。

産業上の利用可能性

[0110] 本発明の方法により、従来の方法では扱うことができなかった、RNAの二次構造がマクスウェル・ボルツマンの分布則に従う存在確率の分布を示すよりも前の時点における該RNAの二次構造の動態を予測することが可能となる。それにより、生体内でRNAが実際に取っている二次構造をより反映する形で、RNAの二次構造を予測することができる。本発明はまた、該方法をコンピュータに実行させるためのプログラム（該プログラムを記録した記録媒体を含む）、および該方法を実施するための装置を提供する。

本出願は日本で出願された特願2009-262258（出願日：2009年11月17日）を基礎としており、その内容は本明細書に全て包含されるものである。

請求の範囲

[請求項1]

RNAの二次構造の動態を予測する方法であって、

(A) RNAの塩基配列から長さ L を有する区間を選択し、該区間において、平衡状態における2種以上の特定二次構造を決定し、かつ、各特定二次構造のギブス自由エネルギーと平衡状態での存在確率とを算出するステップと、

(B) 前記特定二次構造同士の間相互変換における遷移状態の二次構造を決定するとともに、各遷移状態の二次構造のギブス自由エネルギーを求めるステップと、

(C) 各遷移状態の二次構造およびそれらのギブス自由エネルギーから、各特定二次構造同士の間相互変換の経路の活性化エネルギーを算出するステップと、

(D) 各活性化エネルギーを基にして、平衡状態に達するまでの時間的区間内の各時点における各特定二次構造の存在確率を決定するための連立微分方程式を形成するステップと、

(E) 下記(i)～(iv)のうちのいずれか1つの解法によって、各特定二次構造が、前記時間的区間内の各時点において何程の確率で存在し得るかを決定するステップと、
を有することを特徴とする、RNAの二次構造の動態を予測する方法。

(i) 前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とし、そこから導かれる該時点 t_0 での各特定二次構造の存在確率を初期値として前記連立微分方程式を解き、求める存在確率とする解法。

(ii) 前記時間的区間の或る時点 t_0 において各特定二次構造の存在比率が全て等しいものであるとして、各特定二次構造の該時点 t_0 での存在確率を決定し、これを初期値として、前記連立微分方程式

を解き、求める存在確率とする解法。

(iii) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率を平均値とし、かつ、該平均値と、各特定二次構造が等しい比率で存在した場合の存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った 1 個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値として前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作を N 回繰り返した上で、各時点毎に、該 N 回の操作毎の存在確率の総和を N で除して平均し、求める存在確率とする解法。

(iv) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造が等しい比率で存在した場合の存在確率を平均値とし、かつ、該平均値と、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った 1 個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値とし、前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作を N 回繰り返した上で、各時点毎に、該 N 回の操作毎の存在確率の総和を N で除して平均し、求める存在確率とする解法。

[請求項2]

以下の (I) ~ (VI) のステップをさらに有するものである、請求項 1 に記載の方法。

(I) 前記長さ L_1 よりも大きい長さ L を有する対象区間 S を与え

るステップ。

(II) 前記対象区間 S 内の位置 x の一端を位置 $x = x_1$ として、位置 x_1 と位置 $x_1 + L_1$ とを両端とする小区間を $S(x_1)$ とし、

上記のステップ (A) で選択される区間が前記小区間 $S(x_1)$ であるとして、上記のステップ (A) ~ (E) を行って、該小区間 $S(x_1)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定するステップ。

(III) 前記位置 x_1 を、所定の微小長さ r だけ他端側へ移動させた位置を $x_2 = x_1 + r$ とし、位置 x_2 と位置 $x_2 + L_1$ とを両端とする小区間を $S(x_2)$ とし、前記小区間 $S(x_1)$ の代わりに該小区間 $S(x_2)$ を用いる以外は前記 (II) のステップと同様にして、該小区間 $S(x_2)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する、という操作を、前記位置 $x = x_1, x_2$ に続いて、小区間の一端が対象区間 S の他端に達するまで、位置 $x = x_3, \dots, x_M$ について繰り返し行い、それによって、最初の小区間 $S(x_1)$ から最後の小区間 $S(x_M)$ までの各小区間毎の前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定するステップ。

(IV) 前記 (II) ~ (III) のステップの結果に基づいて、前記の小区間 $S(x_1)$ から $S(x_M)$ のうちの任意の小区間 $S(x_i)$ において、任意の時間 t における前記各特定二次構造の存在確率を決定し、

前記各特定二次構造において、位置および構造を表す所定の特徴パラメータ値を有する任意のモチーフ M_k が存在するか否かを調べ、その結果と上記で得られた前記時間 t での各特定二次構造の存在確率とに基づいて、該小区間 $S(x_i)$ 内において、前記時間 t において該モチーフ M_k が何程の蓋然性で存在するかを決定し、これを、小区間 $S(x_i)$ における該モチーフ M_k に関する第 1 の蓋然性とする、

という操作を、前記の小区間 $S(x_1)$ から $S(x_M)$ の全てに対して行

なうステップ。

(V) 前記 (IV) のステップにおいて各小区間内の各特定二次構造における前記モチーフ M_k の存在および非存在を調べた結果に基づいて、各小区間において、モチーフ M_k を有する特定二次構造が少なくとも 1 つ存在するか否かを決定する、

という操作を、前記の小区間 $S(x_i)$ から $S(x_m)$ の全てに対して行い、そのような特定二次構造が少なくとも 1 つ存在する小区間の頻度を、モチーフ M_k に関する第 2 の蓋然性とするステップ。

(VI) 前記 (IV) および (V) のステップで求められた、第 1 の蓋然性と、第 2 の蓋然性とに基づいて、前記時間 t における前記モチーフ M_k の存在を決定するステップ。

[請求項3] 前記モチーフ M_k が、ステム・ループ構造を含むものである、請求項 2 に記載の方法。

[請求項4] 前記モチーフ M_k が、2つのステム・ループ構造の間にあって他の塩基配列と相互作用していない間隙部分と、その両側にある前記ステム・ループ構造のそれぞれの内側のステム部分とを有してなる凹形構造を含むものである、請求項 2 または 3 に記載の方法。

[請求項5] RNA の二次構造の動態を予測するためのコンピュータプログラムであって、

コンピュータを、

(P 1) RNA の塩基配列から長さ L_1 を有する区間を選択し、該区間において、平衡状態における 2 種以上の特定二次構造を決定し、かつ各特定二次構造のギブス自由エネルギーおよび平衡状態での存在確率を算出する手段、

(P 2) 前記特定二次構造同士の間相互変換における遷移状態の二次構造を決定するとともに、各遷移状態の二次構造のギブス自由エネルギーを求める手段、

(P 3) 各遷移状態の二次構造およびそれらのギブス自由エネルギー

ギーから、各特定二次構造同士の間での相互変換の経路の活性化エネルギーを算出する手段、

(P 4) 各活性化エネルギーを基にして、平衡状態に達するまでの時間的区間内の各時点における各特定二次構造の存在確率を決定するための連立微分方程式を形成する手段、

(P 5) 下記 (i) ~ (iv) のうちのいずれか 1 つの解法によって、各特定二次構造が、前記時間的区間内の各時点において何程の確率で存在し得るかを決定する手段、

として機能させるための前記コンピュータプログラム。

(i) 前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とし、そこから導かれる該時点 t_0 での各特定二次構造の存在確率を初期値として前記連立微分方程式を解き、求める存在確率とする解法。

(ii) 前記時間的区間の或る時点 t_0 において各特定二次構造の存在比率が全て等しいものであるとして、各特定二次構造の該時点 t_0 での存在確率を決定し、これを初期値として、前記連立微分方程式を解き、求める存在確率とする解法。

(iii) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率を平均値とし、かつ、該平均値と、各特定二次構造が等しい比率で存在した場合の存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った 1 個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値として前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める

、という操作をN回繰り返した上で、各時点毎に、該N回の操作毎の存在確率の総和をNで除して平均し、求める存在確率とする解法。

(iv) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造が等しい比率で存在した場合の存在確率を平均値とし、かつ、該平均値と、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った1個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値とし、前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作をN回繰り返した上で、各時点毎に、該N回の操作毎の存在確率の総和をNで除して平均し、求める存在確率とする解法。

[請求項6]

コンピュータを、さらに以下の (p 1) ~ (p 6) の手段として機能させるためのものである、請求項5に記載のコンピュータプログラム。

(p 1) 前記長さ L_1 よりも大きい長さ L を有する対象区間 S を入力する手段。

(p 2) 前記対象区間 S 内の位置 x の一端を位置 $x = x_1$ として、位置 x_1 と位置 $x_1 + L_1$ とを両端とする小区間を $S(x_1)$ とし、

上記の手段 (P 1) で選択される区間が前記小区間 $S(x_1)$ であるとして、上記の手段 (P 1) ~ (P 5) を用いて、該小区間 $S(x_1)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する手段。

(p 3) 前記位置 x_1 を、所定の微小長さ r だけ他端側へ移動させた位置を $x_2 = x_1 + r$ とし、位置 x_2 と位置 $x_2 + L_1$ とを両端とする小区間を $S(x_2)$ とし、前記小区間 $S(x_1)$ の代わりに該小区間 $S(x_2)$

を用いる以外は前記 (p 2) の手段と同様にして、該小区間 $S(x_i)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する、という操作を、前記位置 $x = x_1, x_2$ に続いて、小区間の一端が対象区間 S の他端に達するまで、位置 $x = x_3, \dots, x_M$ について繰り返し行い、それによって、最初の小区間 $S(x_1)$ から最後の小区間 $S(x_M)$ までの各小区間毎の前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する手段。

(p 4) 前記 (p 2) ~ (p 3) の手段を用いて得られた結果に基づいて、前記の小区間 $S(x_i)$ から $S(x_M)$ のうちの任意の小区間 $S(x_i)$ において、任意の時間 t における前記各特定二次構造の存在確率を決定し、

前記各特定二次構造において、位置および構造を表す所定の特徴パラメータ値を有する任意のモチーフ M_k が存在するか否かを調べ、その結果と上記で得られた前記時間 t での各特定二次構造の存在確率とに基づいて、該小区間 $S(x_i)$ 内において、前記時間 t において該モチーフ M_k が何程の蓋然性で存在するかを決定し、これを、小区間 $S(x_i)$ における該モチーフ M_k に関する第 1 の蓋然性とする、

という操作を、前記の小区間 $S(x_i)$ から $S(x_M)$ の全てに対して行なう手段。

(p 5) 前記 (p 4) の手段を用いて各小区間内の各特定二次構造における前記モチーフ M_k の存在および非存在を調べた結果に基づいて、各小区間において、モチーフ M_k を有する特定二次構造が少なくとも 1 つ存在するか否かを決定する、

という操作を、前記の小区間 $S(x_i)$ から $S(x_M)$ の全てに対して行い、そのような特定二次構造が少なくとも 1 つ存在する小区間の頻度を、モチーフ M_k に関する第 2 の蓋然性とする手段。

(p 6) 前記 (p 4) および (p 5) の手段で求められた、第 1 の

蓋然性と、第2の蓋然性とに基づいて、前記時間 t における前記モチーフ M_k の存在を決定するための出力を行う手段。

[請求項7] 前記モチーフ M_k が、ステム・ループ構造を含むものである、請求項6に記載のコンピュータプログラム。

[請求項8] 前記モチーフ M_k が、2つのステム・ループ構造の間にあって他の塩基配列と相互作用していない間隙部分と、その両側にある前記ステム・ループ構造のそれぞれの内側のステム部分とを有してなる凹形構造を含むものである、請求項6または7に記載のコンピュータプログラム。

[請求項9] RNAの二次構造の動態を予測するための装置であって、

(Y1) RNAの塩基配列から長さ L_1 を有する区間を選択し、該区間において、平衡状態における2種以上の特定二次構造およびそれらの存在確率を算出する第一演算部と、

(Y2) 前記特定二次構造同士の間での相互変換における遷移状態の二次構造を決定するとともに、各遷移状態の二次構造のギブス自由エネルギーを求める第二演算部と、

(Y3) 各遷移状態の二次構造およびそれらのギブス自由エネルギーから、各特定二次構造同士の間での相互変換の経路の活性化エネルギーを算出する第三演算部と、

(Y4) 各活性化エネルギーを基にして、平衡状態に達するまでの時間的区間内の各時点における各特定二次構造の存在確率を決定するための連立微分方程式を形成する第四演算部と、

(Y5) 下記(i)～(iv)のうちのいずれか1つの解法によって、各特定二次構造が、前記時間的区間内の各時点において何程の確率で存在し得るかを決定する第五演算部と、
を有する、前記装置。

(i) 前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率

とし、そこから導かれる該時点 t_0 での各特定二次構造の存在確率を初期値として前記連立微分方程式を解き、求める存在確率とする解法。

(ii) 前記時間的区間の或る時点 t_0 において各特定二次構造の存在比率が全て等しいものであるとして、各特定二次構造の該時点 t_0 での存在確率を決定し、これを初期値として、前記連立微分方程式を解き、求める存在確率とする解法。

(iii) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率を平均値とし、かつ、該平均値と、各特定二次構造が等しい比率で存在した場合の存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った1個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値として前記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作をN回繰り返した上で、各時点毎に、該N回の操作毎の存在確率の総和をNで除して平均し、求める存在確率とする解法。

(iv) 前記時間的区間の或る時点 t_0 における各特定二次構造の存在確率を、それぞれ正規分布として与えるものとし、それぞれの正規分布は、前記平衡状態における各特定二次構造が等しい比率で存在した場合の存在確率を平均値とし、かつ、該平均値と、前記平衡状態における各特定二次構造の存在確率の逆数を、前記時間的区間の或る時点 t_0 における各特定二次構造の存在比率とすることで導かれる存在確率との差の絶対値を標準偏差とするものとし、

各特定二次構造毎に前記正規分布に従った1個の乱数を与えて対応する存在比率を求め、そこから導かれる各存在確率を初期値とし、前

記連立微分方程式を解き、該初期値を用いた場合の存在確率を求める、という操作をN回繰り返した上で、各時点毎に、該N回の操作毎の存在確率の総和をNで除して平均し、求める存在確率とする解法。

[請求項10]

さらに以下の(y 1)～(y 6)の手段を有するものである、請求項9に記載の装置。

(y 1) 前記長さL 1よりも大きい長さLを有する対象区間Sを入力する手段。

(y 2) 前記対象区間S内の位置xの一端を位置 $x = x_1$ として、位置 x_1 と位置 $x_1 + L_1$ とを両端とする小区間を $S(x_1)$ とし、

上記の手段(Y 1)で選択される区間が前記小区間 $S(x_1)$ であるとして、上記の手段(Y 1)～(Y 5)を用いて、該小区間 $S(x_1)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する手段。

(y 3) 前記位置 x_1 を、所定の微小長さrだけ他端側へ移動させた位置を $x_2 = x_1 + r$ とし、位置 x_2 と位置 $x_2 + L_1$ とを両端とする小区間を $S(x_2)$ とし、前記小区間 $S(x_1)$ の代わりに該小区間 $S(x_2)$ を用いる以外は前記(y 2)の手段と同様にして、該小区間 $S(x_2)$ での前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する、という操作を、前記位置 $x = x_1, x_2$ に続いて、小区間の一端が対象区間Sの他端に達するまで、位置 $x = x_3, \dots, x_M$ について繰り返し行い、それによって、最初の小区間 $S(x_1)$ から最後の小区間 $S(x_M)$ までの各小区間毎の前記各特定二次構造が、前記時間的区間内の各時点で何程の確率で存在し得るかを決定する手段。

(y 4) 前記(y 2)～(y 3)の手段によって決定された結果に基づいて、前記の小区間 $S(x_1)$ から $S(x_M)$ のうちの任意の小区間 $S(x_i)$ において、任意の時間tにおける前記各特定二次構造の存在確率を決定し、

前記各特定二次構造において、位置および構造を表す所定の特徴パラメータ値を有する任意のモチーフ M_k が存在するか否かを調べ、その結果と上記で得られた前記時間 t での各特定二次構造の存在確率とに基づいて、該小区間 $S(x_i)$ 内において、前記時間 t において該モチーフ M_k が何程の蓋然性で存在するかを決定し、これを、小区間 $S(x_i)$ における該モチーフ M_k に関する第1の蓋然性とする、

という操作を、前記の小区間 $S(x_i)$ から $S(x_m)$ の全てに対して行なう手段。

(y5) 前記(y4)の手段を用いて各小区間内の各特定二次構造における前記モチーフ M_k の存在および非存在を調べた結果に基づいて、各小区間において、モチーフ M_k を有する特定二次構造が少なくとも1つ存在するか否かを決定する、

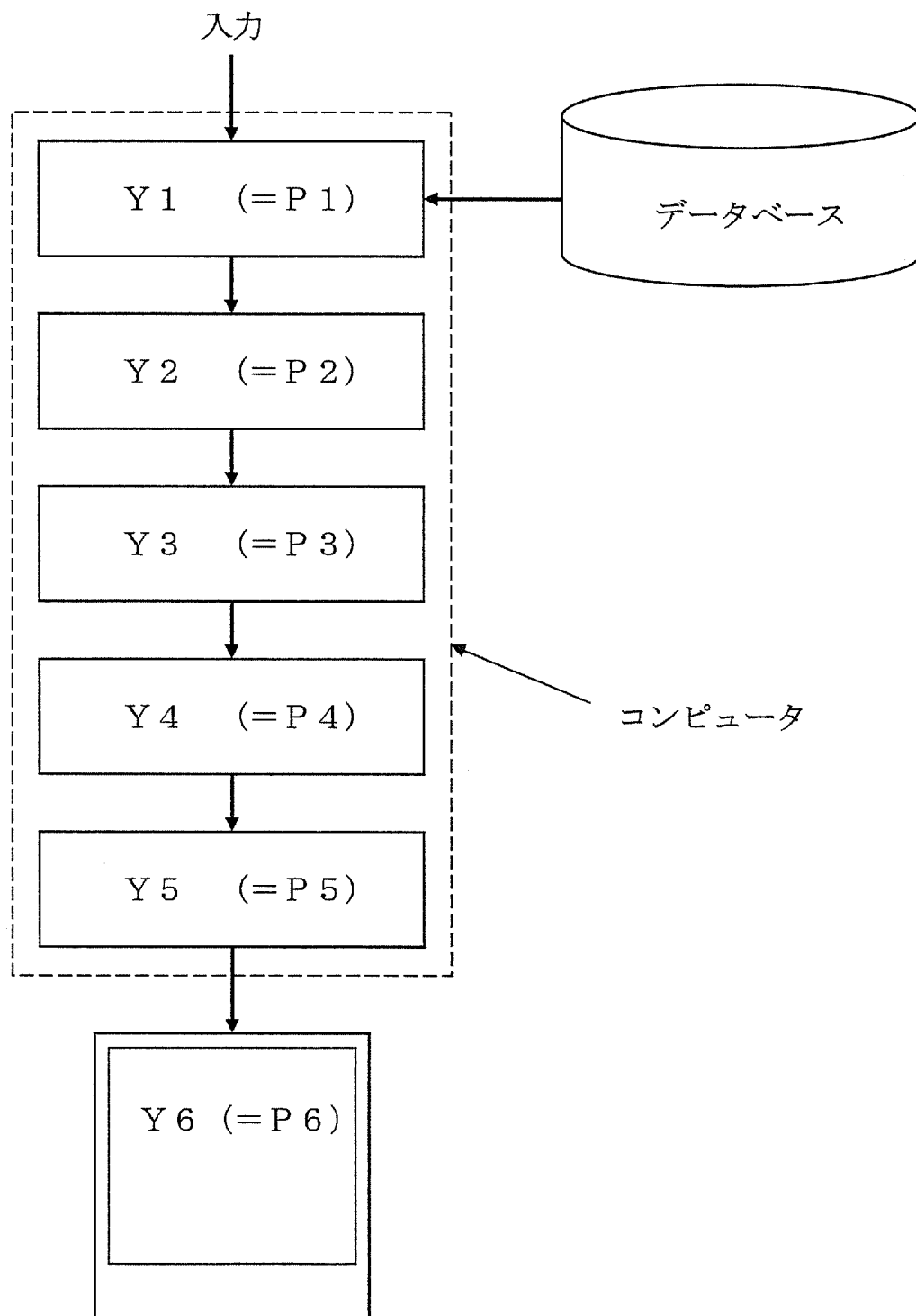
という操作を、前記の小区間 $S(x_i)$ から $S(x_m)$ の全てに対して行い、そのような特定二次構造が少なくとも1つ存在する小区間の頻度を、モチーフ M_k に関する第2の蓋然性とする手段。

(y6) 前記(y4)および(y5)の手段で求められた、第1の蓋然性と、第2の蓋然性とに基づいて、前記時間 t における前記モチーフ M_k の存在を決定するための出力を行う手段。

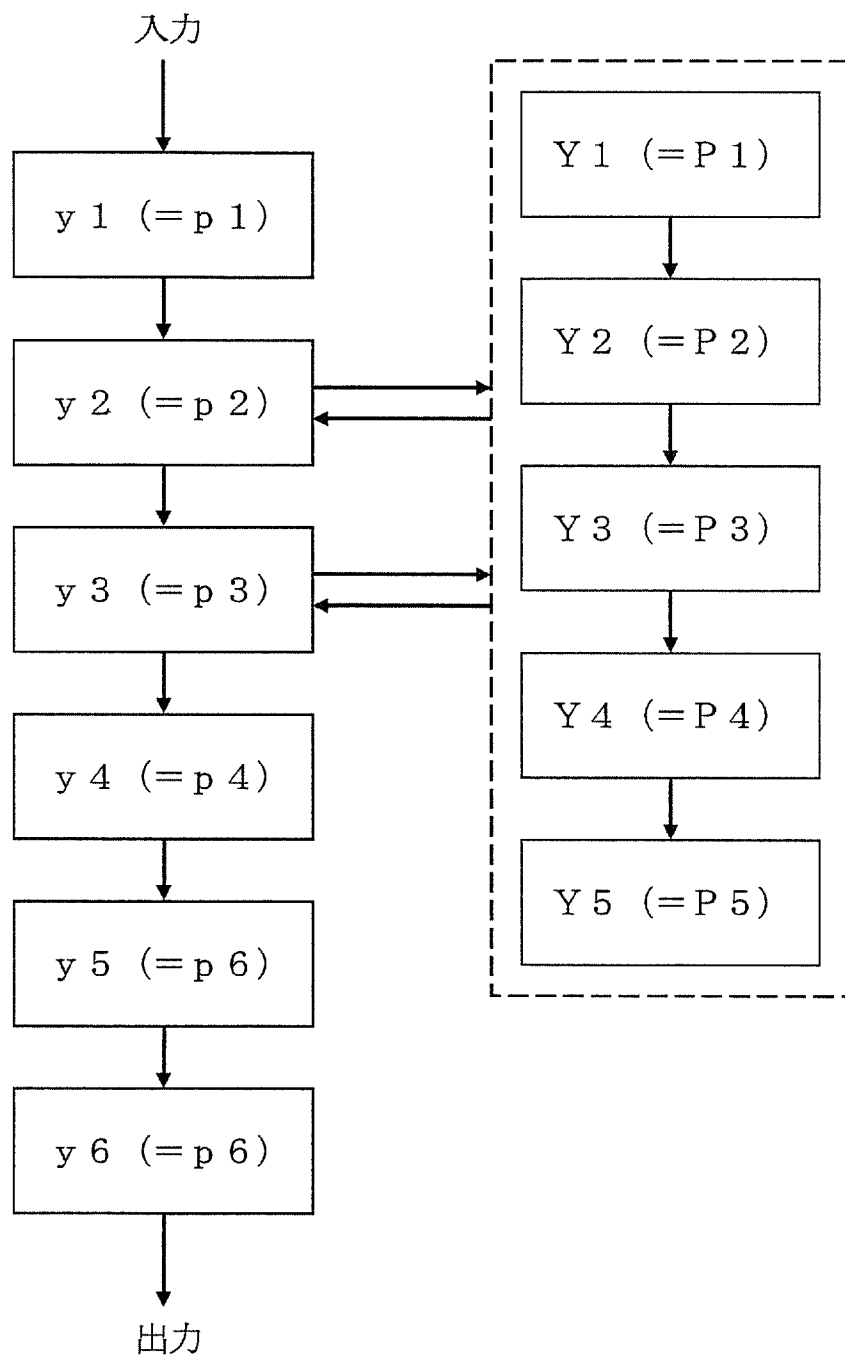
[請求項11] 前記モチーフ M_k が、ステム・ループ構造を含むものである、請求項10に記載の装置。

[請求項12] 前記モチーフ M_k が、2つのステム・ループ構造の間にあって他の塩基配列と相互作用していない間隙部分と、その両側にある前記ステム・ループ構造のそれぞれの内側のステム部分とを有してなる凹形構造を含むものである、請求項10または11に記載の装置。

[図1]



[図2]



[図3]

Option window

Parameters | Display settings | Advanced parameters

Calculation parameters

STEP 1 - baseAnalysis:

Damping window size:

STEP 2 - windowPreparation:

Peak sensitivity:

Window size:

Resolution:

STEP 3 - subOptPopulationAnalysis:

Maxwell-Boltzmann statistics:

Options:

STEP 4 - motifFind:

Search motifs: ☐ Extended

Minimum paired bases: 10

Maximum mismatch bases: 10

Minimum required center bases: 1

STEP 5 - motifBuild:

STEP 6 - spectrumBuild:

STEPS

Threshold values

STEP 4 - Threshold values: VS

☒ VS: 500 [K]

☐ VS+E: [Kcal/mol]

-1000 < ΔG motif < 0.0 [Kcal/mol]

-2000 < ΔG window < 900 [Kcal/mol]

Exclusion 5 [bits] 0 [bases]

STEP 5 - Threshold values: VS

☒ VS: 0.0 [K]

☐ Gc: 1 [K]

Exclusion 5 [bits] 0 [bases]

STEP 6 - Threshold values: VS

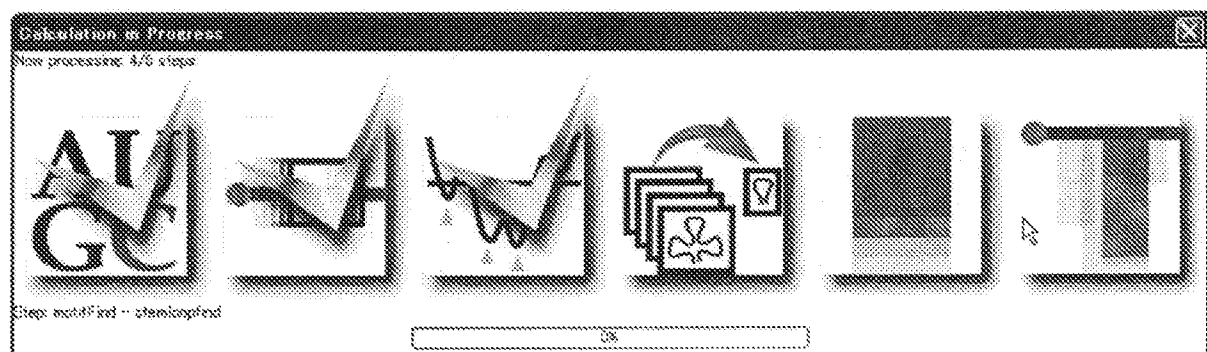
☒ VS: 0.0 [K]

☐ VS+E: [Kcal/mol]

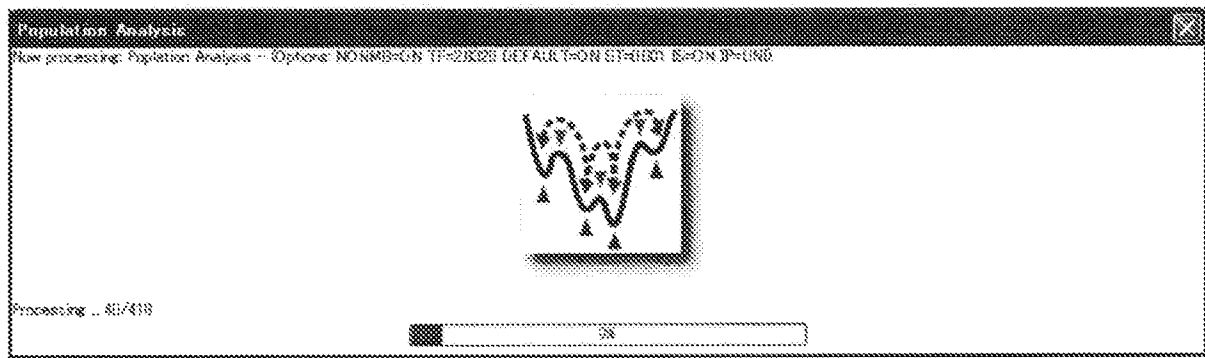
OK

OK Cancel/Close Remove default values Remove motif/center motifs Remove data/offset

[図4]

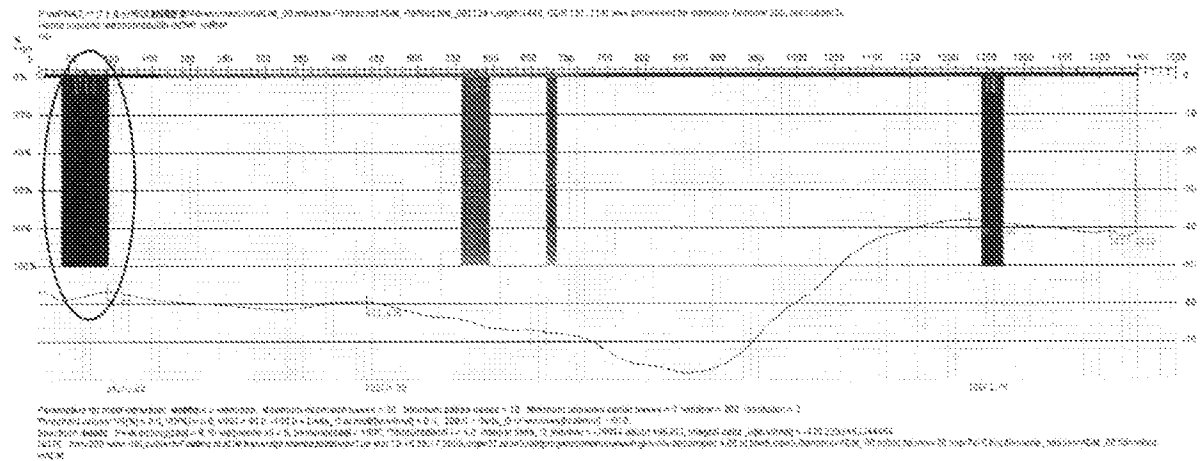


[図5]

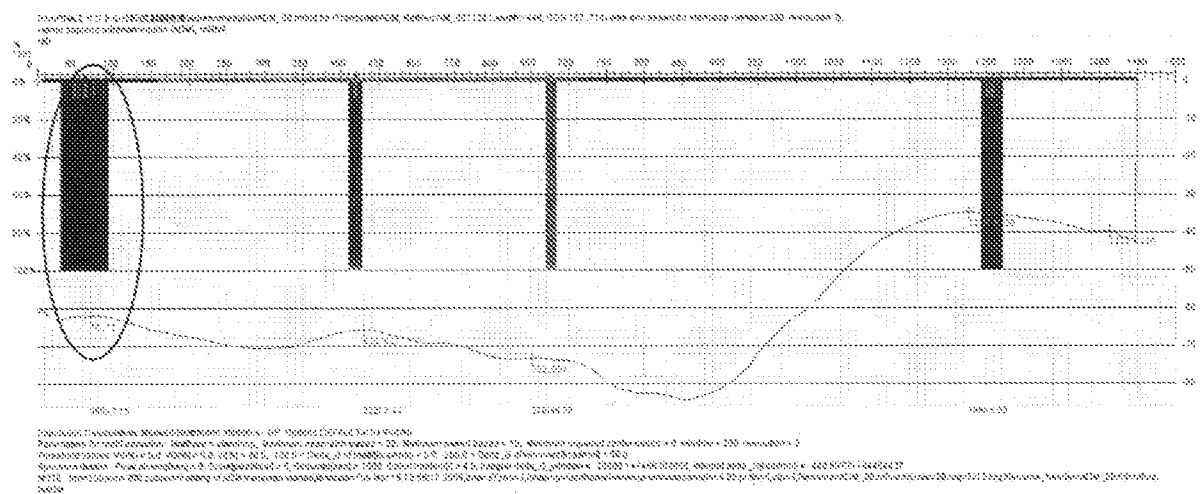


[図6]

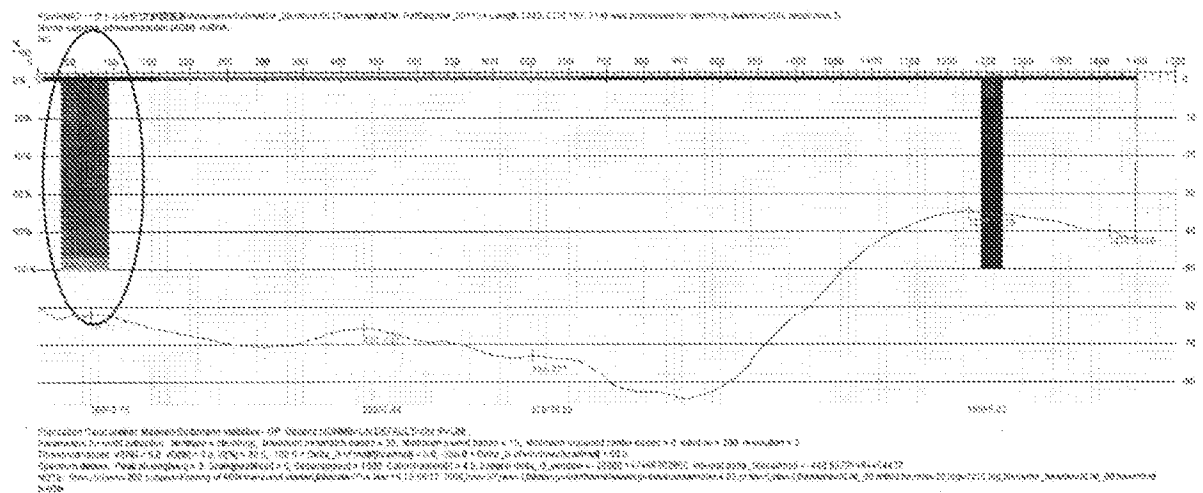
(a) スペクトル1



(b) スペクトル2

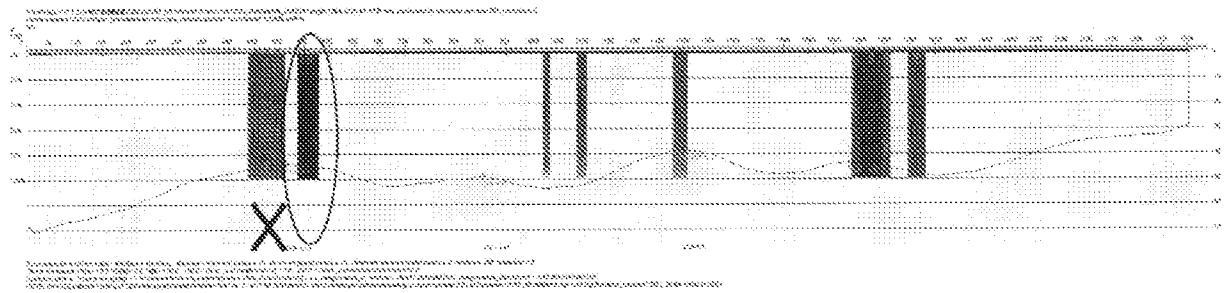


(c) スペクトル3

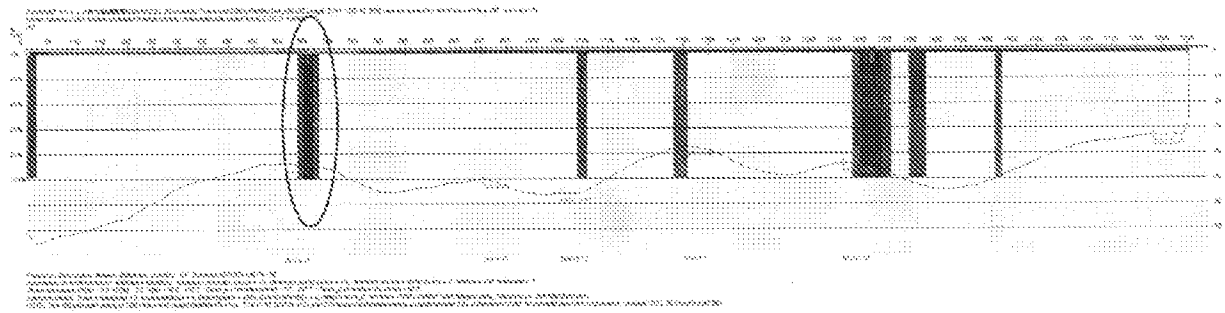


[図7]

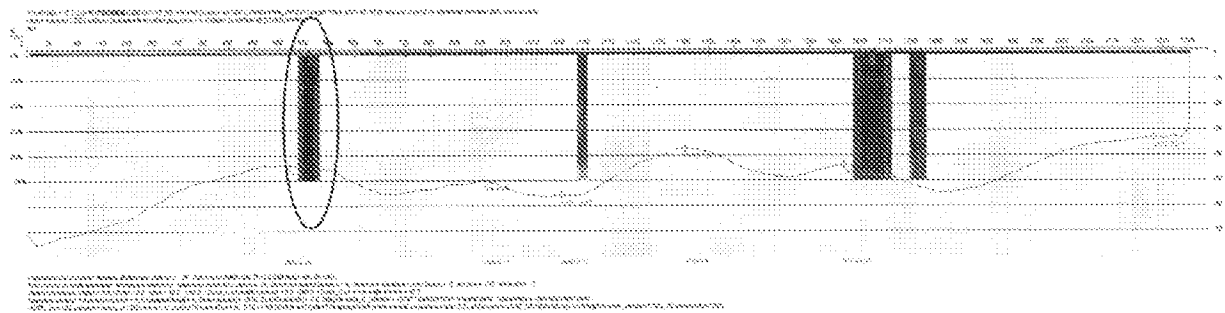
(a) スペクトル4



(b) スペクトル5

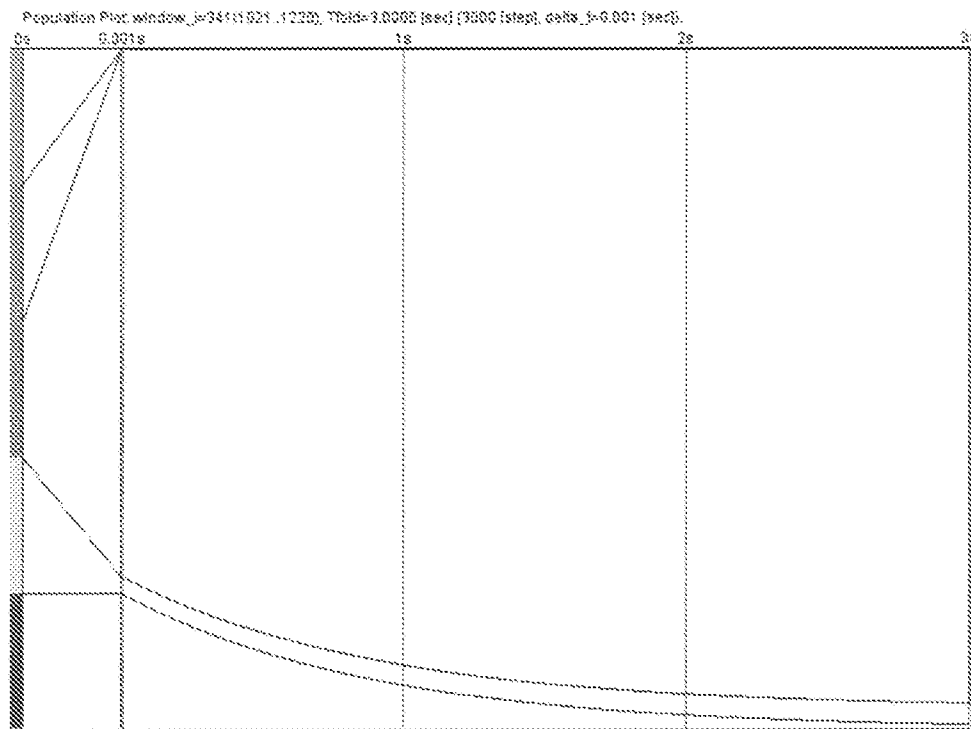


(c) スペクトル6

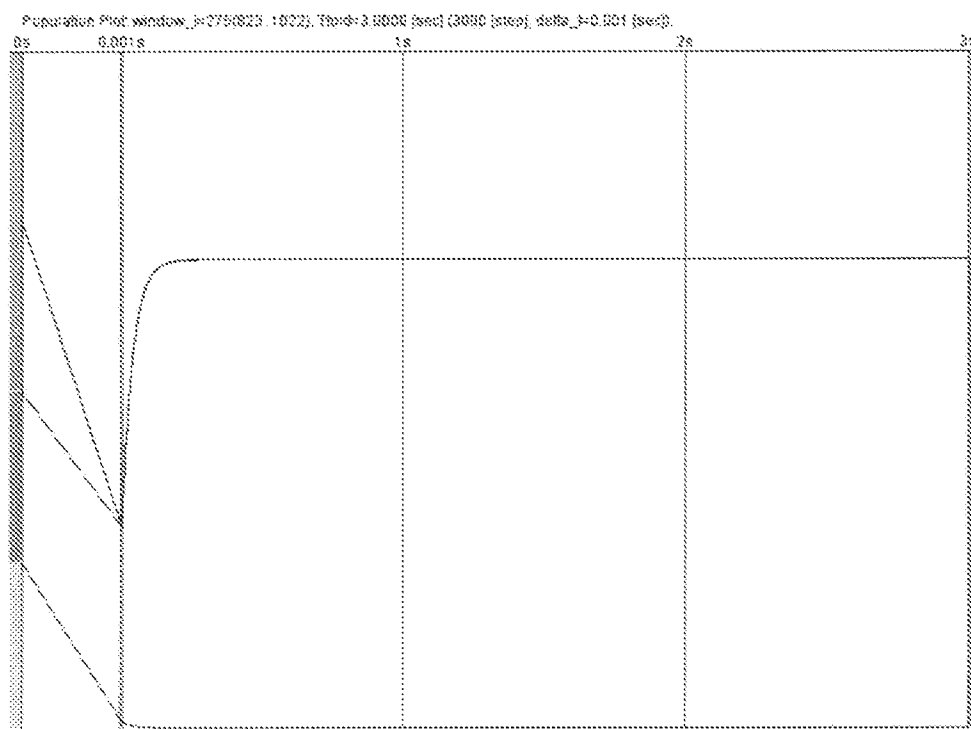


[図8]

(a)

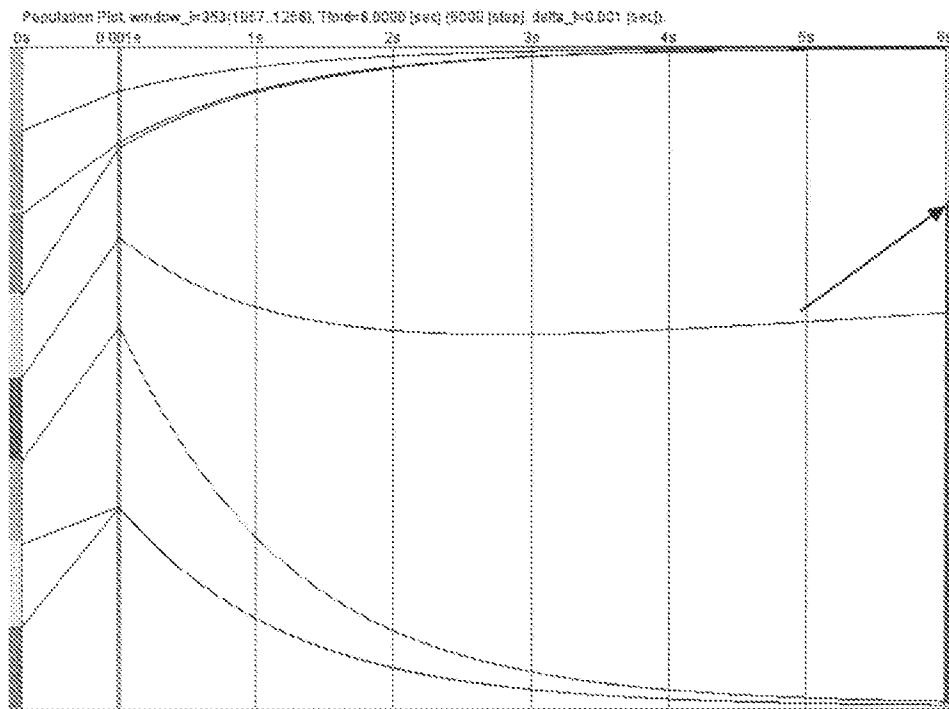


(b)

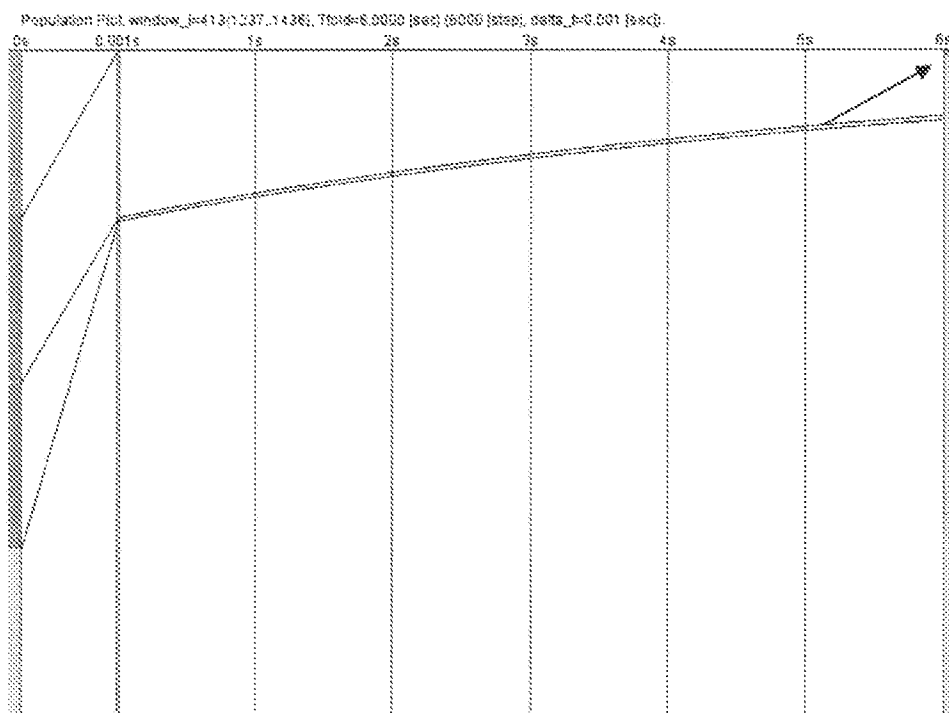


[図9]

(a)

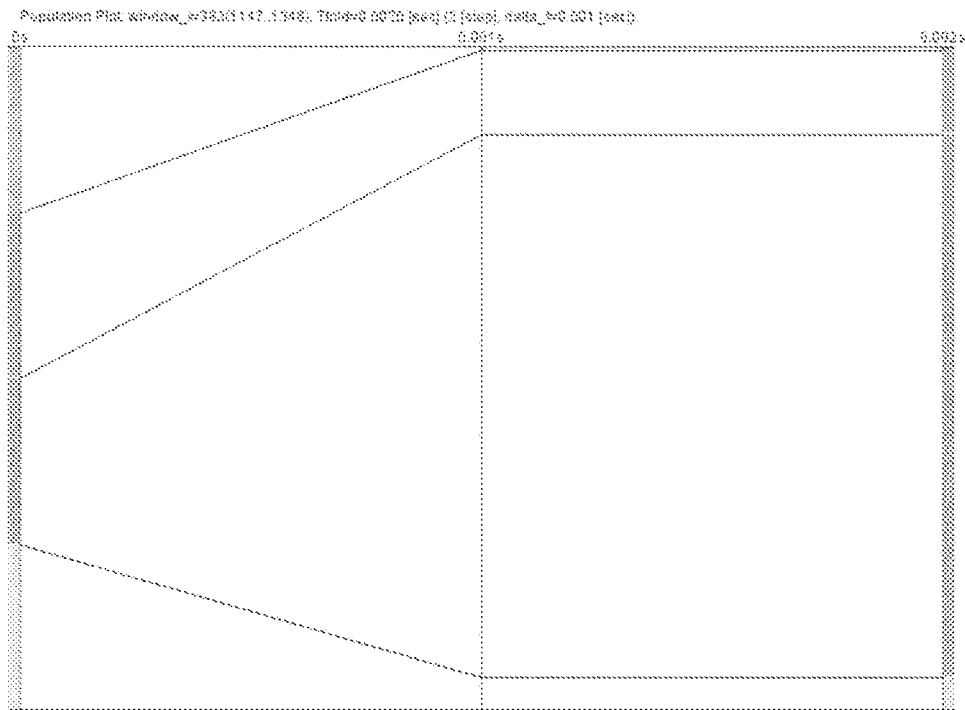


(b)



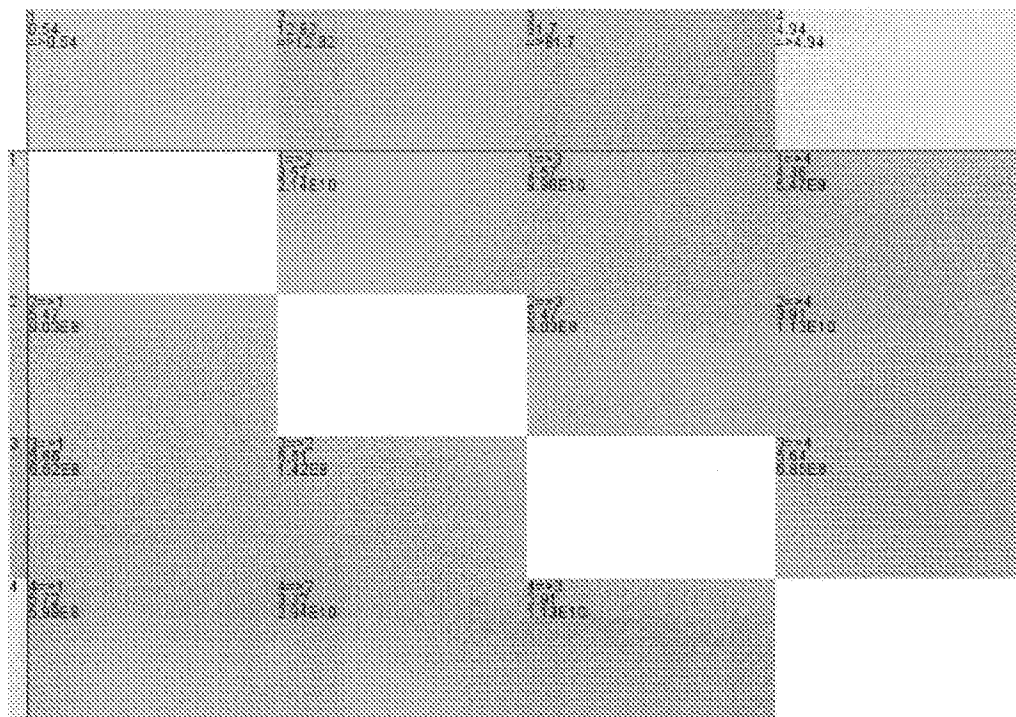
[図10]

(a)

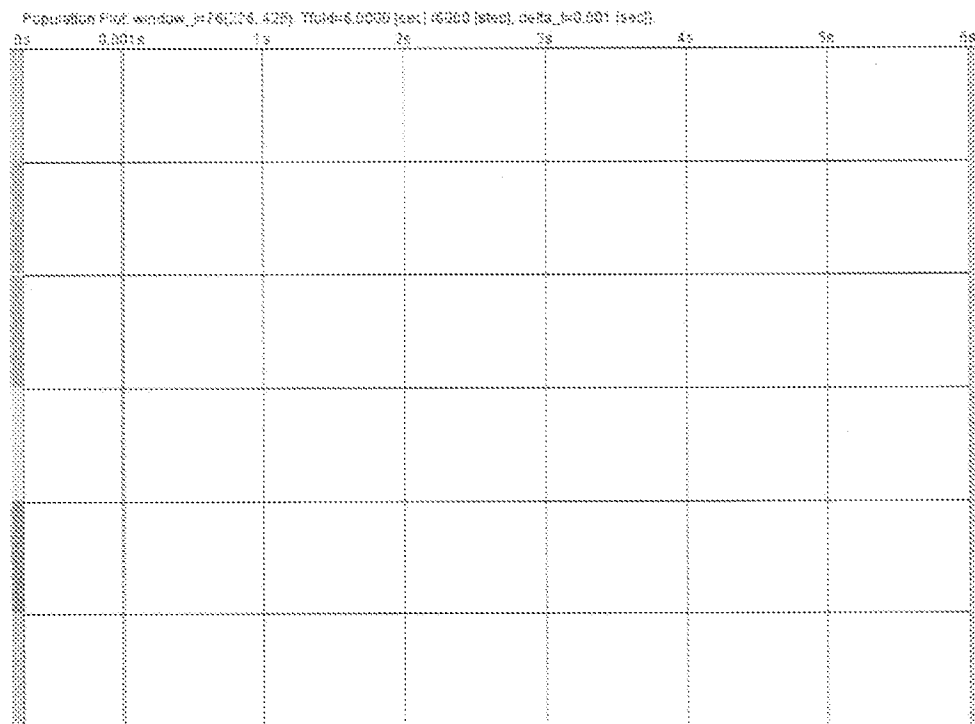


(b)

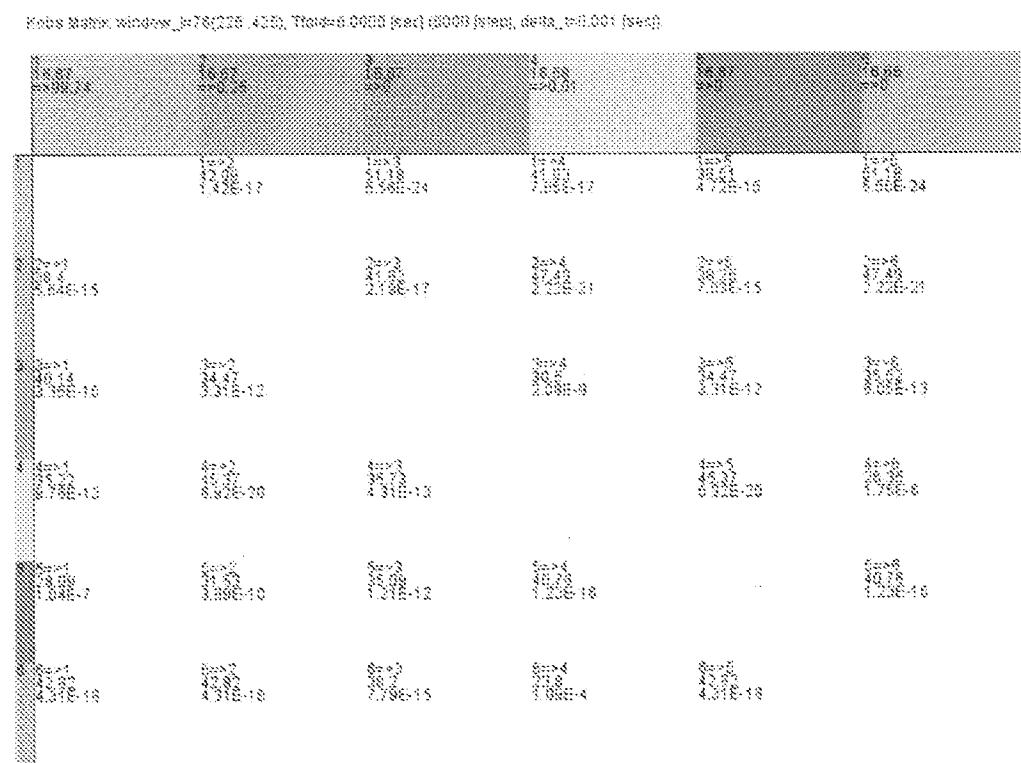
Kobs Matrix window_w=383(1147,1348), Time=0.0020 [sec] (2 steps), Cells_b=0.001 [sec]



(a)



(B)



INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/070400

A. CLASSIFICATION OF SUBJECT MATTER

G06F19/16(2011.01)i, C12N15/09(2006.01)n

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

G06F19/16, C12N15/09

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

Jitsuyo Shinan Koho	1922-1996	Jitsuyo Shinan Toroku Koho	1996-2011
Kokai Jitsuyo Shinan Koho	1971-2011	Toroku Jitsuyo Shinan Koho	1994-2011

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

JSTPlus/JMEDPlus/JST7580(JDREAMII), PubMed

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	ZHANG, W., Analyzing the biopolymer folding rates and pathways using kinetic cluster method, The Journal of chemical physics, 2003.10.22, Vol.119, No.16, p.8716-8729	1-12
A	DING, Y., A statistical sampling algorithm for RNA secondary structure prediction, Nucleic acids research, 2003.12.15, Vol.31, No.24, p.7280-7301	1-12
A	CHO, S.S., Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures, Proceedings of the National Academy of Sciences of the United States of America, 2009.10.13, Vol.106, No.41, p.17349-17354	1-12

☒ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

* Special categories of cited documents:

“A” document defining the general state of the art which is not considered to be of particular relevance

“E” earlier application or patent but published on or after the international filing date

“L” document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

“O” document referring to an oral disclosure, use, exhibition or other means

“P” document published prior to the international filing date but later than the priority date claimed

“T” later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

“X” document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

“Y” document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

“&” document member of the same patent family

Date of the actual completion of the international search
27 January, 2011 (27.01.11)

Date of mailing of the international search report
08 February, 2011 (08.02.11)

Name and mailing address of the ISA/
Japanese Patent Office

Authorized officer

Facsimile No.

Telephone No.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/JP2010/070400

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JP 2007-521813 A (Novartis AG.), 09 August 2007 (09.08.2007), entire text; all drawings & EP 1713913 A & WO 2005/075637 A & CA 2554218 A	1-12
A	WO 2008/072713 A1 (Takeda Chemical Industries, Ltd.), 19 June 2008 (19.06.2008), entire text; all drawings & US 2010/0063745 A & EP 2096567 A1 & CA 2672163 A	1-12
A	WO 2007/116787 A1 (NEC Soft, Ltd.), 18 October 2007 (18.10.2007), entire text; all drawings & US 2010/0235155 A & EP 2012246 A1	1-12

A. 発明の属する分野の分類（国際特許分類（I P C））
Int.Cl. G06F19/16(2011.01)i, C12N15/09(2006.01)n

B. 調査を行った分野

調査を行った最小限資料（国際特許分類（I P C））

Int.Cl. G06F19/16, C12N15/09

最小限資料以外の資料で調査を行った分野に含まれるもの

日本国実用新案公報	1 9 2 2 - 1 9 9 6 年
日本国公開実用新案公報	1 9 7 1 - 2 0 1 1 年
日本国実用新案登録公報	1 9 9 6 - 2 0 1 1 年
日本国登録実用新案公報	1 9 9 4 - 2 0 1 1 年

国際調査で使用した電子データベース（データベースの名称、調査に使用した用語）

JSTPlus/JMEDPlus/JST7580(JDreamII), PubMed

C. 関連すると認められる文献

引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	ZHANG, W., Analyzing the biopolymer folding rates and pathways using kinetic cluster method, The Journal of chemical physics, 2003.10.22, Vol.119, No.16, p.8716-8729	1 - 1 2
A	DING, Y., A statistical sampling algorithm for RNA secondary structure prediction, Nucleic acids research, 2003.12.15, Vol.31, No.24, p.7280-7301	1 - 1 2

☒ C 欄の続きにも文献が列挙されている。

☐ パテントファミリーに関する別紙を参照。

* 引用文献のカテゴリー

「A」特に関連のある文献ではなく、一般的技術水準を示すもの

「E」国際出願日前の出願または特許であるが、国際出願日以後に公表されたもの

「L」優先権主張に疑義を提起する文献又は他の文献の発行日若しくは他の特別な理由を確立するために引用する文献（理由を付す）

「O」口頭による開示、使用、展示等に言及する文献

「P」国際出願日前で、かつ優先権の主張の基礎となる出願

の日の後に公表された文献

「T」国際出願日又は優先日後に公表された文献であって出願と矛盾するものではなく、発明の原理又は理論の理解のために引用するもの

「X」特に関連のある文献であって、当該文献のみで発明の新規性又は進歩性がないと考えられるもの

「Y」特に関連のある文献であって、当該文献と他の1以上の文献との、当業者にとって自明である組合せによって進歩性がないと考えられるもの

「&」同一パテントファミリー文献

国際調査を完了した日

2 7 . 0 1 . 2 0 1 1

国際調査報告の発送日

0 8 . 0 2 . 2 0 1 1

国際調査機関の名称及びあて先

日本国特許庁（I S A / J P）

郵便番号 1 0 0 - 8 9 1 5

東京都千代田区霞が関三丁目4番3号

特許庁審査官（権限のある職員）

宮久保 博幸

電話番号 0 3 - 3 5 8 1 - 1 1 0 1 内線 3 5 6 2

5 L

3 1 3 6

C (続き) . 関連すると認められる文献		
引用文献の カテゴリー*	引用文献名 及び一部の箇所が関連するときは、その関連する箇所の表示	関連する 請求項の番号
A	CHO, S. S., Assembly mechanisms of RNA pseudoknots are determined by the stabilities of constituent secondary structures, Proceedings of the National Academy of Sciences of the United States of America, 2009.10.13, Vol.106, No.41, p.17349-17354	1 - 1 2
A	JP 2007-521813 A (ノバルティス アクチエンゲゼルシャフト) 2007.08.09, 全文, 全図 & EP 1713913 A & WO 2005/075637 A & CA 2554218 A	1 - 1 2
A	WO 2008/072713 A1 (武田薬品工業株式会社) 2008.06.19, 全文, 全図 & US 2010/0063745 A & EP 2096567 A1 & CA 2672163 A	1 - 1 2
A	WO 2007/116787 A1 (NECソフト株式会社) 2007.10.18, 全文, 全図 & US 2010/0235155 A & EP 2012246 A1	1 - 1 2