Vol.2010-BIO-22 No.6
2010/7/29

期待精度最大化に基づくRNAシュードノット予測

佐藤　健吾[*1,†1]　加藤　有己[*1,†2]
阿久津　達也[†2]　浅井　潔[†1,†3]

RNAに観測されるシュードノットと呼ばれる部分構造は，多くの場合3次元空間上での折り畳みを補助する役割を担うことが知られており，シュードノットを含めたRNA 2次構造予測はその立体構造決定への手がかりを与えるものと期待される．本稿では，期待精度最大化に基づくシュードノット構造予測法 IPknot を提案する．IPknot では，シュードノットを考慮した事後塩基対確率分布を，シュードノットを含まない塩基対の各集合に対する事後分布の積で近似する．この期待精度最大化問題は閾値カットを用いた整数計画法で解く．また，計算機実験により，IPknot は高い予測精度と高速な計算速度を達成することを示す．

RNA Pseudoknot Prediction Based on Maximizing Expected Accuracy

Kengo Sato[*1,†1] Yuki Kato[*1,†2] Tatsuya Akutsu[†2]
and Kiyoshi Asai[†1,†3]

Pseudoknots, substructures observed in RNA secondary structures, play a role in assisting the overall 3D folding in many cases, and thus prediction of RNA secondary structures including pseudoknots is expected to provide a clue to determine the 3D structures of RNA molecules. In this technical report, we propose IPknot, a computational method for predicting pseudoknotted structures based on maximizing expected accuracy. IPknot approximates a posterior base pairing probability distribution that considers pseudoknots by decomposing it into the product of respective posterior distributions over pseudoknot-free structures. We solve the problem of maximizing expected accuracy using integer programming with threshold cut. Experimental results show that IPknot achieves high prediction accuracy and fast computation time.

1. はじめに

生体分子の一種であるRNAは，遺伝情報のメッセンジャーとしての受動的な役割から，遺伝子発現量の調節や細胞のプロセスにおける触媒としての能動的な役割まで，幅広い役割を担う核酸であり，その機能の解明に注目が集まっている[9]．RNAの立体構造と機能の間には相関があると言われており，機能の解明のためには構造を予測することが重要となる．立体構造を2次元平面上に射影し，塩基間の水素結合の情報のみを表したものを2次構造と呼ぶ．特に，2次構造の中で構成する塩基対が1次配列上で互いに交差する関係にある部分構造はシュードノットと呼ばれている．シュードノットはリボソームRNAやグループIイントロン，ウイルスRNAなどの多くのRNA分子の2次構造で観測されている．また，シュードノットは翻訳やスプライシングの調節，リボソームの読み枠の移動などに関与していることが知られている．さらに，シュードノットは多くの場合，3次元空間上での折り畳みを補助する役割を担うとされており，立体構造解析のためにはシュードノットを考慮した2次構造を予測することが重要となる．

現在まで，多くのRNA 2次構造予測法が情報科学的アプローチに基づいて開発されている．これらは比較配列解析法と単一配列解析法に大別される．初期の2次構造解析法は，計算の複雑さの観点からシュードノットを考慮しない構造を扱うことが多く，以下に述べる2次構造解析法はシュードノットを考慮しないことに注意されたい．比較解析に基づく手法として代表的なものは，確率文脈自由文法（SCFG）の構文解析技術に基づく2次構造予測法が挙げられる[8),14)]．比較解析によるアプローチは進化的な保存情報を考慮するため予測精度が高いという利点があるが，事前に相同な配列群を必要とするため，常に適用可能とは限らない．一方，単一配列解析法では，自由エネルギー最小化に基づくアプローチが代表的である．自由エネルギー最小の2次構造は動的計画法により $O(n^3)$ 時間で計算することができ，mfold[24),25)]，RNAfold[11),12)] などで利用可能である．ここで，n は入力塩基配

†1 東京大学 大学院新領域創成科学研究科
Graduate School of Frontier Sciences, University of Tokyo
†2 京都大学 化学研究所 バイオインフォマティクスセンター
Bioinformatics Center, Institute for Chemical Research, Kyoto University
†3 産業技術総合研究所 生命情報工学研究センター
Computational Biology Research Center (CBRC), National Institute of Advanced Industrial Science and Technology (AIST)
∗1 The authors wish it to be known that in their opinion the first two authors should be regarded as joint First Authors.

列の長さを表す．なお，2次構造のエネルギーは各種ループ部分構造の持つエネルギーパラメータの和で計算され，実験的に決定されているものが利用可能である[16]．また，上記動的計画法に基づく手法は，RNA 2次構造の分配関数の計算に応用されており[17]，これにより塩基対確率などの事後確率の計算が可能になる．最近では，自由エネルギー最小の2次構造は実際の構造と異なる場合が多いと指摘されているため，可能な2次構造全体からなる空間上で，予測構造の期待精度が最大となるような2次構造を求める手法（CONTRAfold[7]，CentroidFold[10] など）が開発されており，高い予測精度を上げることに成功している．

一方，任意のシュードノットを考慮したとき，自由エネルギー最小の2次構造を予測する問題は NP 困難であることが証明されている[1),15)]．そのため，考慮するシュードノットのクラスを限定した2次構造予測に対する $O(n^4) \sim O(n^6)$ 時間の動的計画アルゴリズムがいくつか開発されている[1),2),6),15),19),21)]．それらの中で，PKNOTS[21] は最も広範囲なシュードノットのクラスを扱えるアルゴリズムであるが[5]，その時間計算量は $O(n^6)$ であり，数百塩基といった長い塩基配列への適用は難しい．pknotsRG[19] は，扱うシュードノットのクラスを PKNOTS からさらに狭めることで，その時間計算量を $O(n^4)$ に削減したアルゴリズムである．また，整数計画法を用いることで，最小自由エネルギー構造を計算するアプローチも提案されている[18]．シュードノットに対する厳密なエネルギーパラメータは現在のところ利用困難であるため，上記いずれの手法もシュードノットに対する近似エネルギーパラメータを用いており，予測性能はその数理モデルにより限定的であると考えられる．また，入力配列サイズが大きくなると実行不能となる可能性をはらんでいる．

動的計画法が扱うシュードノットの複雑さに起因する計算量の問題を克服するために，ヒューリスティクスに基づくいくつかのアプローチが提案されている[2),4),20),22)]．ILM[22]，HotKnots[20]，FlexStem[4] は，2次構造の構成要素の1つであるステムの候補を段階的に追加，削除する手法であり，数百塩基以上の比較的長い配列に対しても，高速に2次構造を予測することが可能となっている．これらヒューリスティクスに基づく手法は予測構造の最適性が保証されないものの，計算量を増加させることなく広範囲なシュードノットのクラスを扱うことができ，長い配列に対しても現実的な時間内で予測できる利点を持つ．

本稿では，期待精度最大化に基づくシュードノット構造予測法 IPknot を提案する．ここで，予測構造の期待精度を計算するために，シュードノットを考慮した2次構造全体からなる空間上の事後確率分布が必要になる．IPknot では，シュードノットを考慮した事後塩基対確率分布を，シュードノットを含まない塩基対の有限個の集合に対する事後分布の積で近似する．この分解法により，任意のシュードノットを表現することが可能となる．期待精度最大化問題は閾値カットを用いた整数計画法で解く．閾値カットにより，予測構造の期待精度の向上に寄与しない塩基対はあらかじめ問題の定式化から除外するため，整数計画問題のサイズが格段に小さくなり，最適化に要する時間が大幅に短縮される．計算機実験では，シュードノットを含むことが知られている配列データセットに対し，いくつかの RNA 2次構造予測法との比較を行い，IPknot が高い予測精度と高速な計算速度を達成することを示す．

2. 手　　法

RNA 配列 $x = x_1 x_2 \cdots x_n$ が取りうるシュードノットを含む2次構造全体の集合を $\mathcal{S}(x)$，シュードノットを含まない2次構造全体の集合を $\mathcal{S}'(x)$ とする．2次構造 $y \in \mathcal{S}(x)$ あるいは $y \in \mathcal{S}'(x)$ は2値の上三角行列 $y = (y_{ij})_{i<j}$ $(y_{ij} \in \{0,1\})$ で表され，y_{ij} の値は塩基 x_i と x_j が塩基対を形成しているかどうかを表す．

ここで，シュードノットを含む2次構造 $y \in \mathcal{S}(x)$ が，シュードノットを含まない m 個の互いに素な2次構造からなる集合 $\{y^{(1)}, y^{(2)}, \ldots, y^{(m)} \mid y^{(p)} \in \mathcal{S}'(x) \text{ for } 1 \le p \le m\}$ に分割できると仮定する．ただし，$y^{(p)}$ に含まれる任意の塩基対は $y^{(q)}$ $(\forall q < p)$ に含まれる塩基対のいずれかとシュードノットの関係にあるとする[⋆1]．この分割に基づき，シュードノットを考慮した2次構造に関する利益関数を次のように定義する:

$$G(y, \hat{y}) = \sum_{1 \le p \le m} \alpha^{(p)} G_{\gamma^{(p)}}(y^{(p)}, \hat{y}^{(p)}), \tag{1}$$

$$\begin{aligned} G_\gamma(y, \hat{y}) &= \gamma TP(y, \hat{y}) + TN(y, \hat{y}) \\ &= \sum_{i<j} \left[\gamma\, I(y_{ij} = 1)\, I(\hat{y}_{ij} = 1) + I(y_{ij} = 0)\, I(\hat{y}_{ij} = 0)\right]. \end{aligned}$$

ここで，y は参照2次構造，$\hat{y}$ は予測2次構造，$\sum_p \alpha^{(p)} = 1$，$\gamma > 0$ とする．また，$I(condition)$ は条件 $condition$ が真のとき 1，偽のとき 0 の値をとる指標関数である．

与えられた RNA 配列 x のシュードノットを含む2次構造の確率分布 $P(y \mid x)$ の下で，利益関数 (1) の期待値

$$\mathbb{E}_{y|x}[G(y, \hat{y})] = \sum_{y \in \mathcal{S}(x)} G(y, \hat{y})\, P(y \mid x) \tag{2}$$

が最大となるような2次構造 $\hat{y} \in \mathcal{S}(x)$ を求めたい．しかしながら，あらゆるシュードノッ

⋆1 この分割は一意には定まらないが，最終的には以下に説明する目的関数に関して最適なものを選ぶため，問題にはならない．

Vol.2010-BIO-22 No.6
2010/7/29

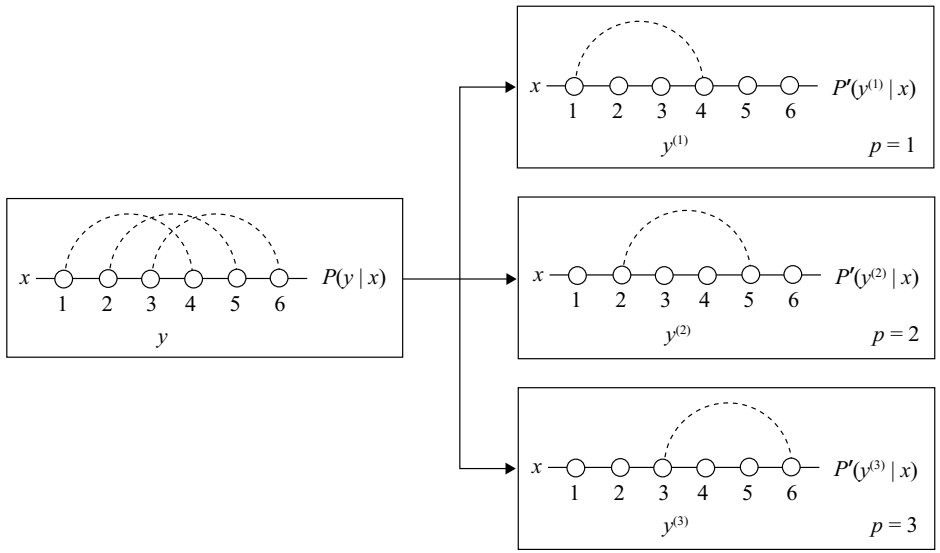

図 1 $m = 3$ のときの確率分布の分解の式 (3) の説明図．破線は塩基対を表す．

トを考慮した 2 次構造の確率分布の計算は NP 困難であることが知られており[2]，比較的単純なシュードノットに限定して動的計画法により計算したとしても $O(n^4) \sim O(n^6)$ 時間（n は配列長）もかかるため，式 (2) をそのまま計算することは実用的ではない．そこで本手法では上記の分割に基づき，シュードノットを含んだ 2 次構造の確率分布をシュードノットを含まない 2 次構造の確率分布の積で近似する（図 1 参照）：

$$P(y \mid x) \simeq \prod_{1 \le p \le m} P'(y^{(p)} \mid x). \tag{3}$$

ここで $P'(y \mid x)$ はシュードノットを含まない 2 次構造の確率分布であり，$O(n^3)$ で計算することができる．この近似により，期待利益関数 (2) は次のように置き換えることができる:

$$\begin{aligned}\mathbb{E}_{y|x}[G(y,\hat{y})] &\simeq \sum_{1 \le p \le m} \alpha^{(p)} \sum_{y^{(p)} \in \mathcal{S}'(x)} G_{\gamma^{(p)}}(y^{(p)}, \hat{y}^{(p)})\, P'(y^{(p)} \mid x) \\ &= \sum_{1 \le p \le m} \alpha^{(p)} \sum_{i<j} \left[(\gamma^{(p)}+1)p_{ij} - 1\right] \hat{y}_{ij}^{(p)} + C. \end{aligned} \tag{4}$$

ここで C は 予測 2 次構造 $\hat{y}$ に依存しない定数であり，$p_{ij} = \sum_{y \in \mathcal{S}'(x)} I(y_{ij}=1) P'(y \mid x)$ は塩基対確率（塩基 x_i と x_j が塩基対を形成する確率）を表している．式 (4) を最大化することは，各 p $(1 \le p \le m)$ の部分 2 次構造内の塩基対 $y_{ij}^{(p)}$ に対応する塩基対確率 p_{ij} が与えられた閾値 $\theta^{(p)} = 1/(\gamma^{(p)}+1)$ よりも大きくなるものだけを考えればよいことを意味する．

$m = 1$ の時は，シュードノットを考えない場合に相当する．この時，利益関数 (1) や期待利益関数 (4) は `CentroidFold`[10] の場合と等価になり，本手法は `CentroidFold` の自然な拡張となっていることがわかる．

最終的には，式 (4) を最大化する問題は次の整数計画問題として定式化できる:

$$\text{maximize} \quad \sum_{1 \le p \le m} \alpha^{(p)} \sum_{i<j} p_{ij} y_{ij}^{(p)} \tag{5}$$

$$\text{subject to} \quad \sum_{1 \le p \le m} \left\{ \sum_{j=1}^{i-1} y_{ji}^{(p)} + \sum_{j=i+1}^{n} y_{ij}^{(p)} \right\} \le 1 \quad (1 \le \forall i \le n), \tag{6}$$

$$y_{ij}^{(p)} + y_{kl}^{(p)} \le 1 \quad (1 \le \forall p \le m,\, 1 \le \forall i < \forall k < \forall j < \forall < l \le n), \tag{7}$$

$$\sum_{i<k<j<l} y_{ij}^{(q)} + \sum_{k<i'<l<j'} y_{i'j'}^{(q)} \ge y_{kl}^{(p)} \quad (1 \le \forall q < \forall p \le m;\ 1 \le \forall k < \forall l \le n). \tag{8}$$

式 (4) から，塩基対確率 p_{ij} が閾値 $\theta^{(p)}$ よりも大きい塩基対 $y_{ij}^{(p)}$ のみを考えればよいため，整数計画問題の規模を大幅に小さくすることができる．制約 (6) は，各々の塩基 x_i は他の塩基と塩基対を形成することができるのは高々一回のみであることを意味している．制約 (7) により，部分 2 次構造 $y^{(p)}$ の内部ではシュードノットを許さない．制約 (8) は，$y^{(p)}$ に含まれる任意の塩基対は $y^{(q)}$ $(\forall q < p)$ に含まれる塩基対のいずれかとシュードノットの関係にあることを表している．

一般に，RNA 2 次構造において塩基対が単独で孤立して存在することは稀であり，2 つ以上の塩基対が重なって現れることが多い（スタッキング）．本手法では，整数計画問題に以下のような制約を加えることにより，孤立塩基対が出現しないようにする:

$$\ell_{i-1}^{(p)} + (1 - \ell_i^{(p)}) + \ell_{i+1}^{(p)} \ge 1 \quad (1 < \forall i < n;\ 1 \le \forall p \le m), \tag{9}$$

$$r_{i-1}^{(p)} + (1 - r_i^{(p)}) + r_{i+1}^{(p)} \ge 1 \quad (1 < \forall i < n;\ 1 \le \forall p \le m). \tag{10}$$

ただし，$\ell_i^{(p)}, r_i^{(p)}$ は 2 値変数であり，下式で定義される:

$$\ell_i^{(p)} = \sum_{j=i+1}^{n} y_{ij}^{(p)} \quad (1 \le \forall i < n;\ 1 \le \forall p \le m),$$

$$r_i^{(p)} = \sum_{j=1}^{i-1} y_{ij}^{(p)} \quad (1 < \forall i \le n;\ 1 \le \forall p \le m).$$

3. 結　　果

2 節で述べた手法 `IPknot` を実装し，既存手法と比較するための計算機実験を行った．整数計画法のソルバーには GNU Linear Programming Toolkit (GLPK) を用いた．また，塩基対確率を計算するためのシュードノットを含まない 2 次構造の確率分布として CONTRAfold model[7] と McCaskill model[17] を使用し，パラメータには次の値を用いた: $m = 2$, $\alpha^{(1)} = 0.6$, $\alpha^{(2)} = 0.4$, $\gamma^{(1)}, \gamma^{(2)} \in \{2^n \mid n = -2, \ldots 4\}$．比較対象として，シュードノット予測が可能な手法 `ILM`[22], `pknotsRG`[19], `FlexStem`[4] と，シュードノット予測が不可能な手法 `RNAfold`[11),12)], `CentroidFold`[10] (CONTRAfold model, McCaskill model) を用いた．

実験に用いる配列は RNA STRAND データベース[3] から取得した．少なくとも 1 つはシュードノットを含み，配列長が 500 bp 以下の配列群をデータベースから選んだ後，blastclust による single linkage clustering を行い，配列相同性が 85 % 以上となる冗長な配列を除いた．その結果，399 本の RNA 配列が得られた．

2 次構造予測の精度は，次式で定義される塩基対の Sensitivity (Sen) と Positive Predictive Value (PPV) で評価した:

$$\text{sensitivity} = \frac{TP}{TP + FN}, \quad \text{PPV} = \frac{TP}{TP + FP}.$$

ここで，TP は正しい予測塩基対の数，FN は予測できなかった正解塩基対の数，FP は正しくなかった予測塩基対の数である．また，Sen と PPV はトレードオフの関係にあるため，両者のバランスを取った評価尺度として，次式の Matthews Correlation Coefficient (MCC) を用いた:

$$\text{MCC} = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}.$$

ここで，TN は塩基対でないと正しく予測した数である．

図 2 に，各手法の予測精度を横軸 PPV，縦軸 Sen でプロットしたグラフを示す．`IPknot` と `CentroidFold` は Sen と PPV を調節するためのパラメータ γ があるため，複数の点がプロットされている． このグラフから，本提案手法 `IPknot` は既存手法（シュードノット予測可，不可ともに）よりも精度良くシュードノットを考慮に入れた RNA 2 次構造が予測

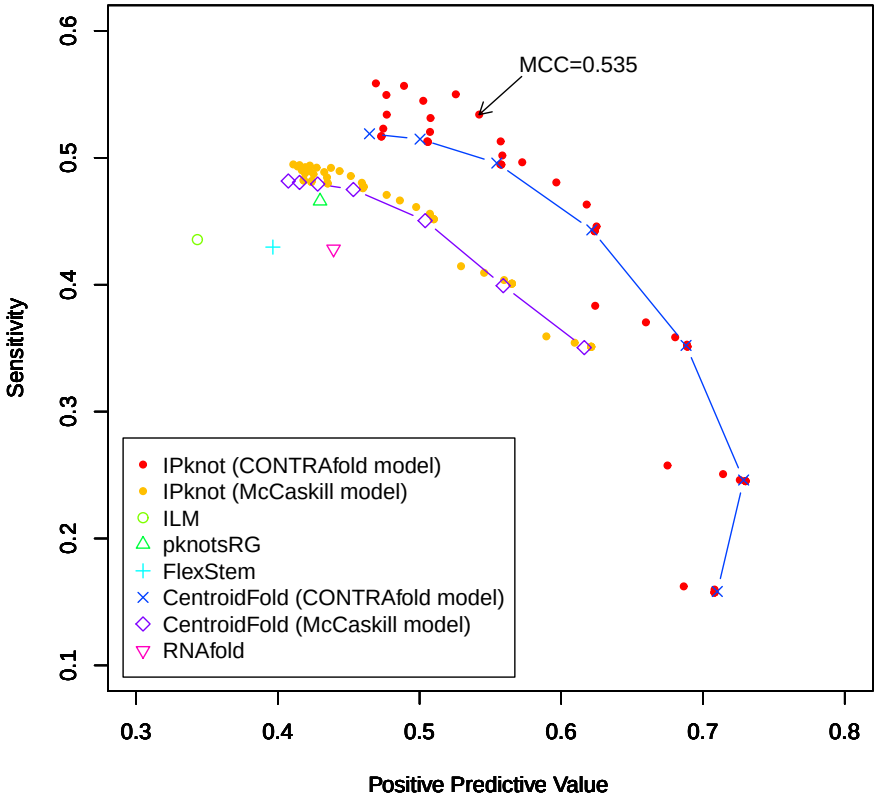

図 2 PPV–Sensitivity Plots. 矢印で示した点は MCC が最も高かったパラメータの組み合わせ $\gamma^{(1)} = 4$, $\gamma^{(2)} = 8$ である．

表 1 最も長い配列 CRW_00614 (497 bp) 上での 実行時間．CentroidFold は $\gamma = 4$, IPknot は $\gamma^{(1)} = 4$, $\gamma^{(2)} = 8$ の時の実行時間を表す．

Method	time (s)
IPknot (CONTRAfold model)	0.75
IPknot (McCaskill model)	0.44
pknotsRG	5.51
ILM	0.18
FlexStem	9.66
CentroidFold (CONTRAfold model)	0.73
CentroidFold (McCaskill model)	0.41
RNAfold	0.14

できることがわかる．

表 1 は，データセット中で最も長い配列 CRW_00614 (497 bp) の 2 次構造予測を実行するのに要した計算時間 (Linux OS, Intel Xeon E5540 2.53 GHz) である． ILM 以外のシュードノット予測ツールと比べて有意に高速であり，かつ精度の面でも上回っていることがわかる．

4. おわりに

本稿では，期待精度最大化に基づく RNA シュードノット予測法 IPknot を開発した．IPknot はシュードノットの平面的分解により，任意のシュードノットを扱うことが可能なモデルである．予測法の中核をなす期待精度最大化は閾値カットを用いた整数計画法により実現し，他の最新の予測手法と比較して高速かつ高精度な予測を達成した．なお，IPknot は，http://www.ncrna.org/software/ipknot/ から入手可能である．

今回の計算機実験による性能評価では，RNA STRAND[3] から独自に編纂した 399 本の RNA 配列からなるデータセットを用いた．一方，文献 4)，13) では，シュードノットを収集した信頼性の高いデータベース PseudoBase[23] から，単純シュードノット（H 型シュードノット）[2] を含むことが知られている冗長でない 168 本の配列セット pk168 をベンチマークとして用いている．IPknot が pk168 上でどの程度の予測性能を上げるかを検証することが，今後の課題として挙げられる．また，多重配列アライメントが与えられたとき，それらの共通 2 次構造を予測する問題も重要である．IPknot は共通 2 次構造予測問題にも適用可能と考えており，今後の課題として取り組む予定である．

謝辞 本研究は一部，文部科学省科学研究費補助金若手研究（B）[#22700305 to K.S., #22700313 to Y.K] からの助成金を受けている．

参 考 文 献

1) Akutsu,T. (2000) Dynamic programming algorithms for RNA secondary structure prediction with pseudoknots. *Discrete Appl. Math.*, **104**, 45–62.
2) Akutsu,T. (2006) Recent advances in RNA secondary structure prediction with pseudoknots. *Current Bioinformatics*, **1**, 115–129.
3) Andronescu,M., Bereg,V., Hoos,H.H. and Condon,A. (2008) RNA STRAND: the RNA secondary structure and statistical analysis database. *BMC Bioinform.*, **9**, 340.
4) Chen,X., He,S., Bu,D., Zhang,F., Wang,Z., Chen,R. and Gao,W. (2008) FlexStem: improving predictions of RNA secondary structures with pseudoknots by reducing the search space. *Bioinformatics*, **24**, 1994–2001.
5) Condon,A., Davy,B., Rastegari,B., Zhao,S. and Tarrant,F. (2004) Classifying RNA pseudoknotted structures. *Theor. Comput. Sci.*, **320**, 35–50.
6) Dirks,R.M. and Pierce,N.A. (2003) A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J. Comput. Chem.*, **24**, 1664–1677.
7) Do,C.B., Woods,D.A. and Batzoglou,S. (2006) CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics*, **22**, e90–e98.
8) Durbin,R., Eddy,S.R., Krogh,A. and Mitchison,G. (1998) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge, United Kingdom.
9) Eddy,S.R. (2001) Non-coding RNA genes and the modern RNA world. *Nat. Rev. Genet.*, **2**, 919–929.
10) Hamada,M., Kiryu,H., Sato,K., Mituyama,T. and Asai,K. (2009) Prediction of RNA secondary structure using generalized centroid estimators. *Bioinformatics*, **25**, 465–473.
11) Hofacker,I.L., Fontana,W., Stadler,P.F., Bonhöffer,S., Tacker,M. and Schuster,P. (1994) Fast folding and comparison of RNA secondary structures. *Monatsh. Chem.*, **125**, 167–188.
12) Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.
13) Huang,X. and Ali,H. (2007) High sensitivity RNA pseudoknot prediction. *Nucleic Acids Res.*, **35**, 656–663.
14) Knudsen,B. and Hein,J. (1999) RNA secondary structure prediction using stochastic context-free grammars and evolutionary history. *Bioinformatics*, **15**, 446–454.
15) Lyngsø,R.B. and Pedersen,C.N. (2000) RNA pseudoknot prediction in energy-based models. *J. Comput. Biol.*, **7**, 409–427.
16) Mathews,D.H., Sabina,J., Zuker,M. and Turner,D.H. (1999) Expanded sequence

Vol.2010-BIO-22 No.6
2010/7/29

dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J. Mol. Biol.*, **288**, 911–940.
17) McCaskill,J.S. (1990) The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, **29**, 1105–1119.
18) Poolsap,U., Kato,Y. and Akutsu,T. (2009) Prediction of RNA secondary structure with pseudoknots using integer programming. *BMC Bioinform.*, **10**(Suppl 1), S38.
19) Reeder,J. and Giegerich,R. (2004) Design, implementation and evaluation of a practical pseudoknot folding algorithm based on thermodynamics. *BMC Bioinform.*, **5**, 104.
20) Ren,J., Rastegari,B., Condon,A. and Hoos,H.H. (2005) HotKnots: heuristic prediction of RNA secondary structures including pseudoknots. *RNA*, **11**, 1494–1504.
21) Rivas,E. and Eddy,S.R. (1999) A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.*, **285**, 2053–2068.
22) Ruan,J., Stormo,G.D. and Zhang,W. (2004) An iterated loop matching approach to the prediction of RNA secondary structures with pseudoknots. *Bioinformatics*, **20**, 58–66.
23) van Batenburg,F.H.D., Gultyaev,A.P, Pleij,C.W.A, Ng,J. and Oliehoek,J. (2000) PseudoBase: a database with RNA pseudoknots. *Nucleic Acids Res.*, **28**, 201–204.
24) Zuker,M. and Stiegler,P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.*, **9**, 133–148.
25) Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.