

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成24年5月31日現在

機関番号：14603

研究種目：若手研究（B）

研究期間：2010～2011

課題番号：22700313

研究課題名（和文） ハイブリッド型最適化による RNA 間相互作用予測

研究課題名（英文） RNA-RNA interaction prediction using hybrid optimization

研究代表者

加藤 有己 (KATO YUKI)

奈良先端科学技術大学院大学・情報科学研究科・助教

研究者番号：10511280

研究成果の概要（和文）：本研究では、整数計画法と期待精度最大化原理の融合により、RNA間の相互作用予測のための結合2次構造予測法を開発した。さらに、提案手法のアイデアに基づき、単一 RNA 配列に対してシュードノットと呼ばれる複雑な構造を考慮した2次構造予測法を発表した。いずれも実際のデータ上で性能評価を行った結果、既存の手法と比べて予測精度は同等以上で、実行時間では圧倒的な高速性を実現した。なお、提案手法は Web サーバーに実装して一般公開している。

研究成果の概要（英文）：In this work, we developed a novel method for RNA joint secondary structure prediction that aims at RNA-RNA interaction prediction by integrating the maximizing expected accuracy principle with integer programming. Furthermore, based on the idea of the proposed method, we presented an algorithm for predicting RNA secondary structures with pseudoknots, a class of complex substructures, when a single RNA sequence is given. Experimental results on real data sets indicate that both methods are at least comparable in prediction accuracy to existing methods, and even more importantly, are overwhelmingly faster than the early predictors. A web server that implements the proposed methods is free and open to all users.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,100,000	330,000	1,430,000
2011年度	800,000	240,000	1,040,000
年度			
年度			
年度			
総計	1,900,000	570,000	2,470,000

研究分野：バイオインフォマティクス

科研費の分科・細目：情報学・生体生命情報学

キーワード：RNA 間相互作用、RNA 2次構造、キッキングヘアピン、シュードノット、整数計画法

1. 研究開始当初の背景

近年、タンパク質に翻訳されない non-coding RNA (ncRNA) の機能に注目が集まっている。機能と構造の間には相関があると言われており、ncRNA の機能推定のためには構造を知ることが重要となる。生体生命情報学では、モデル化の容易さから、塩基の結合情報のみを表す 2 次構造を扱うことが多い。ncRNA のクラスの中で、標的となる mRNA に結合することでタンパク質への翻訳を抑制するものが多く発見されており、それらは調節 RNA と呼ばれている。調節 RNA は、標的 mRNA との間に主に Watson-Crick 相補性による塩基対を形成することで相互作用するが、この結合の仕方にバリエーションがあることが知られている。例えば、調節 RNA の全ての塩基が相互作用に関わる場合があり、主に miRNA や siRNA など小規模な RNA で観測されている。一方、比較的長い RNA 配列では、配列内での自分自身との結合 (内部結合) の間に外部への相互作用 (外部結合) が現れるような、キッキングヘアピンと呼ばれる構造も見られる。実用的相互作用予測という観点からは、これらの構造を組み合わせて扱えることが望ましい。計算機を用いた調節 RNA の新規標的の発見などの応用を考えると、RNA 間の結合部位を正確に予測することが重要であり、そのためには各配列自身の 2 次構造と相互作用部位を同時に考慮に入れることが必要不可欠である。

本研究の着想に至った経緯は次の通りである。研究代表者の加藤は数年前、整数計画法を用いる RNA 間相互作用予測問題に関する初等的な研究に取り組んだ経緯がある。整数計画問題とは、整数値をとる変数の線形不等式系で表された制約条件の下で、目的関数と呼ばれる線形関数を最大化 (あるいは最小

化) する問題である。整数計画問題は NP 困難と呼ばれる効率良く解くことが望めない計算量クラスに属することが証明されているが、そのモデル記述能力は極めて高く、生体生命情報学に現れる様々な最適化問題が整数計画問題として定式化できる。また、近年では最適化ソルバーの性能が目覚ましい進歩もあって、計算が困難であった大規模な問題も比較的高速に解けるようになってきている。研究代表者の過去の研究ではキッキングヘアピンを考慮した制約条件を考え、2 つの RNA の結合 2 次構造のスコアを表す目的関数が最大となる解を求める整数計画モデルを与えた。しかしながら、入力配列のサイズが大きくなると、最適化ソルバーが解を出力しないという状況に陥った。そのため、形式文法を用いるという異なる視点から、多項式時間の動的計画法に基づく相互作用予測法を開発し、実用的な時間内で良い予測精度をあげることに成功した。これまでのところ、整数計画法に基づく解析手法は実用上不向きかと思われたが、本研究課題の申請時に、整数計画問題の定式化を工夫することによって、RNA の構造アラインメントの問題を、より解析が容易な配列アラインメントの問題に帰着し、計算量を大幅に下げるという研究を目にする機会があった。この整数計画モデルでは、2 次構造に関する情報を制約条件に含めず目的関数に集約することで、整数計画問題の数学的な構造を簡略化している。この考え方を応用するならば、2 次構造の情報を反映した目的関数の精緻な設計が極めて重要となる。そこで、構造のスコアを表す目的関数の係数を計算効率の良い動的計画法により導出し、構造情報を省略することで制約条件を緩和した整数計画問題を解けば、高い性能と柔軟性をあわせもつ RNA 間相互作

用予測法が実現できるのではないかとの着想に至った。

2. 研究の目的

本研究課題の申請時における当初の研究目的は以下の通りである。

(1) 相互作用予測に要する計算コストの削減が可能な整数計画モデルを設計する。

(2) 構造のスコアを表す目的関数の各係数を効率良く求める動的計画アルゴリズムを開発する。

(3) 開発手法を計算機に実装し、既知の相互作用データを用いて提案手法の性能評価を行う。さらに、提案予測法を Web サーバーに実装し、商用目的以外には自由に利用できるようにすることを旨とする。

3. 研究の方法

(1) 研究代表者が以前行った整数計画問題によるモデル化を一から見直し、計算コストの削減が可能となるような新たな定式化を考える。具体的には、2本の RNA 配列の2次構造の形状に関する制約条件を、内部結合と外部結合の整合性を保ちながら可能な限り除去し、目的関数の係数に内部結合と外部結合の情報を集約させる。これにより、解空間の複雑さが緩和され、100塩基以上の長さをもつ RNA 配列の組に対しても比較的容易に解を求めることができるものと期待される。また同時に、モデルにフィードバックする情報があるかどうかを考察するため、得られた整数計画問題の数学的な構造についても調べる。

(2) 上記(1)で構造に関する情報を目的関数の

係数に集約させたため、配列からの構造スコア関数の計算は提案手法の予測精度を左右する重要な作業となる。ここでは、内部結合のスコアは外部結合のスコアに依存すると仮定する。まず、外部結合のスコアには、外部結合を構成するスタッキング塩基対を考慮した Markov モデルを設計し、2本の配列を入力とする動的計画法のフォワード・バックワードアルゴリズムを開発すれば、塩基間の相互作用に対応するスコアを計算できると考えている。なお、外部結合を先に扱う理由は、完全な配列相補性による相互作用を考慮するためである。一方、各配列の内部結合に対応するスコアには、その配列で任意の塩基対を形成する確率を表す塩基対確率行列を拡張して、外部結合のスコアを考慮した上で、内部結合に関する塩基対確率行列を求める動的計画法を新たに設計する。ただし、外部結合、内部結合とも全ての可能な塩基対を考えると効率が下がるので、ある閾値以下のスコアをもつ塩基対は予測の候補から外す。また、上記2つの動的計画アルゴリズムは、その時間計算量が配列長の3乗以下のオーダーになるように設計する。

(3) 上述の整数計画問題と動的計画法を、Java 言語を用いて計算機に実装する。なお、実際に整数計画問題を解くために、その性能の高さから世界中で広く使用されている（申請当時商用の）最適化ソルバー CPLEX のライブラリを用いる。その後、相互作用することが実験的に確かめられている RNA 複合体と提案手法が出力した予測構造を比較し、予測精度や計算速度などを評価することで、有効性の検証や問題点の検出を行う。

4. 研究成果

(1) 整数計画法を用いることで、最も一般的

と考えられている結合構造のクラスを扱える、高速で高精度の RNA 間相互作用予測法 RactIP を開発した。RactIP では RNA 配列の組が与えられたとき、以下の方法論を組み合わせることで構造予測を行う：

- ① 複雑な構造空間上で定義される確率分布を近似的に因数分解する (図 1 参照)；
- ② 予測構造の期待精度を最大化する目的で整数計画問題を解く。

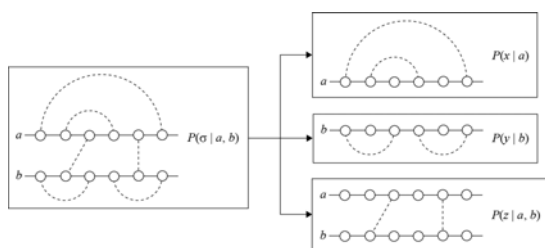


図 1： 結合 2 次構造の確率分布の分解。

ここでは、内部および外部塩基対確率を整数計画問題の目的関数に組み込むことで、結合 2 次構造のアンサンブルを考慮して期待精度を最適化するように設計している。実際に相互作用することが実験的に検証されている RNA 配列の組のデータセットを用いて計算機実験を行ったところ、RactIP は複数の既存手法と比べて同等以上の予測精度を達成し、さらに圧倒的な高速化 (最大で約 4 万倍) を実現した。一般的に、予測アルゴリズムが扱える構造の複雑度と計算効率の間にはトレードオフの関係があるが、RactIP はそのトレードオフの関係を打破したという意味で、非常にインパクトを与えた手法だと言える。

(2) 次に、(1) で得られた方法論を他の問題にも適用できないかと考え、RNA 配列に対してシュードノットと呼ばれる複雑な構造を考慮した 2 次構造予測問題に対し、IPknot と呼ばれる計算手法を開発した。シ

ュードノットは例えば図 2 のような構造を取り、シュードノットを考慮する場合は考慮しない場合に比べ、予測に要する計算量のオーダーが増加することが知られている。また、多くの RNA はシュードノットを含み、その機能解析のためには無視できない構造要素であると考えられている。

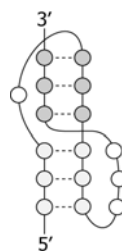


図 2： ヘアピン型シュードノット。

本研究では単一 RNA 配列解析に加え、多重配列アラインメントが与えられたとき、その共通 2 次構造を予測するように IPknot を拡張した。シュードノットを含むことが知られている多数の配列からなるデータセットを用いて提案手法を評価したところ、RactIP のときと同様、他の複数の予測手法と比べて同等以上の予測精度、圧倒的な高速性を達成した (図 3 参照)。シュードノットを考慮した RNA 2 次構造予測では、高速性と高精度を両立したツールは非常に数少なく、今後

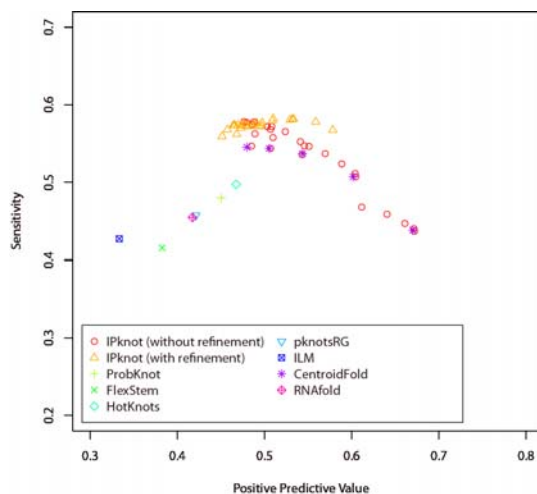


図 3： IPknot の予測精度。

IPknot がシュードノット構造予測ソフトウェアの決定版になることが期待される。

(3) 上記開発プログラム RactIP と IPknot を商用目的以外で自由に利用できるようにするため、Web サーバー Rtips を開発し、<http://rna.naist.jp/> に公開した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 4 件)

[1] Unyanee Poolsap, Yuki Kato, Kengo Sato and Tatsuya Akutsu, **Using binding profiles to predict binding sites of target RNAs**, *Journal of Bioinformatics and Computational Biology*, vol. 9, issue 6, pp. 697-713, Dec. 2011, 査読有り.

[2] Kengo Sato, Yuki Kato, Michiaki Hamada, Tatsuya Akutsu and Kiyoshi Asai, **IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming**, *Bioinformatics*, vol. 27, no. 13, pp. i85-i93, Jul. 2011, 査読有り.

[3] Yuki Kato, Kengo Sato, Michiaki Hamada, Yoshihide Watanabe, Kiyoshi Asai and Tatsuya Akutsu, **RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming**, *Bioinformatics*, vol. 26, no. 18, pp. i460-i466, Sep. 2010, 査読有り.

[4] Yusuke Ishida, Yuki Kato, Liang Zhao, Hiroshi Nagamochi and Tatsuya Akutsu, **Branch-and-bound algorithms for enumerating treelike chemical graphs with given path frequency using detachment-cut**, *Journal of Chemical Information and Modeling*, vol. 50, no. 5, pp. 934-946, May 2010, 査読有り.

[学会発表] (計 16 件)

[1] 佐藤 健吾, 加藤 有己, 阿久津 達也, 浅井 潔, 榊原 康文, **双対分解による RNA 構造アラインメント**, 第 28 回情報処理学会バイオ情報学研究会, 宮城県仙台市, 2012 年 3 月 28 日.

[2] Kengo Sato, Yuki Kato, Tatsuya Akutsu

and Kiyoshi Asai, **Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming**, 第 34 回日本分子生物学会年会, 神奈川県横浜市, 2011 年 12 月 14 日.

[3] Kengo Sato, Yuki Kato, Tatsuya Akutsu, Kiyoshi Asai and Yasubumi Sakakibara, **Simultaneous aligning and folding of RNA sequences by dual decomposition**, CBI/JSBi2011 合同大会, 兵庫県神戸市, 2011 年 11 月 10 日.

[4] Yuki Kato, Kengo Sato, Kiyoshi Asai and Tatsuya Akutsu, **Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming**, CBI/JSBi2011 合同大会, 兵庫県神戸市, 2011 年 11 月 8 日.

[5] Yuki Kato, Kengo Sato, Kiyoshi Asai and Tatsuya Akutsu, **Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming**, ICR Symposium to Celebrate the Bioinformatics Center's 10 Year Anniversary and New Restructuring, 京都府宇治市, 2011 年 8 月 29 日.

[6] 加藤 有己, **複雑な RNA 2 次構造予測のための高速計算ソフトウェアの開発と今後の展開**, 次世代バイオインフォマティクス研究会, 北海道札幌市, 2011 年 8 月 2 日 (招待講演).

[7] Kengo Sato, Yuki Kato, Michiaki Hamada, Tatsuya Akutsu and Kiyoshi Asai, **IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming**, 19th Annual International Conference on Intelligent Systems for Molecular Biology and 10th European Conference on Computational Biology (ISMB/ECCB2011), オーストリア ウィーン, 2011 年 7 月 17 日.

[8] Kengo Sato, Yuki Kato, Michiaki Hamada, Tatsuya Akutsu and Kiyoshi Asai, **IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming**, Sixteenth Annual Meeting of the RNA Society (RNA2011), 京都府京都市, 2011 年 6 月 17 日.

[9] Yuki Kato, Kengo Sato, Michiaki Hamada, Yoshihide Watanabe, Kiyoshi

Asai and Tatsuya Akutsu, **RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming**, Sixteenth Annual Meeting of the RNA Society (RNA2011), 京都府京都市, 2011年6月15日.

[10] Nobuyoshi Mizoguchi, Yuki Kato and Hiroyuki Seki, **A grammar-based approach to RNA pseudoknotted structure prediction for aligned sequences**, 1st IEEE International Conference on Computational Advances in Bio and medical Sciences (ICCABS2011), アメリカ合衆国 オーランド, 2011年2月3日.

[11] Yuki Kato, Kengo Sato, Michiaki Hamada, Yoshihide Watanabe, Kiyoshi Asai and Tatsuya Akutsu, **RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming**, 9th European Conference on Computational Biology (ECCB2010), ベルギー ゲント, 2010年9月27日.

[12] Kengo Sato, Yuki Kato, Tatsuya Akutsu and Kiyoshi Asai, **IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming**, 9th European Conference on Computational Biology (ECCB2010), ベルギー ゲント, 2010年9月27日.

[13] Yuki Kato, Kengo Sato, Michiaki Hamada, Yoshihide Watanabe, Kiyoshi Asai and Tatsuya Akutsu, **RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming**, Computational Biology Research Center Workshop 2010 (CBRC2010), 東京都江東区, 2010年7月29日.

[14] 佐藤 健吾, 加藤 有己, 阿久津 達也, 浅井 潔, **期待精度最大化に基づく RNA シュールドノット予測**, 第 22 回情報処理学会バイオ情報学研究会, 東京都江東区, 2010年7月29日.

[15] Yuki Kato, Kengo Sato, Michiaki Hamada, Yoshihide Watanabe, Kiyoshi Asai and Tatsuya Akutsu, **RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming**, 18th Annual International Conference on Intelligent Systems for Molecular Biology (ISMB2010), アメリカ合衆国 ボストン, 2010年7月11日.

[16] Yuki Kato, Kengo Sato, Michiaki Hamada, Yoshihide Watanabe, Kiyoshi Asai and Tatsuya Akutsu, **RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming**, 第 21 回情報処理学会バイオ情報学研究会, 沖縄県中頭郡西原町, 2010年6月19日.

〔図書〕(計1件)

[1] 加藤 有己, 実験医学増刊号, 使えるデータベース・ウェブツール, 第4章第8節「二次構造に基づく RNA 配列解析ソフトウェアの進展」, pp. 229-236, 羊土社, 2011年9月.

〔その他〕

ホームページ等

<http://rna.naist.jp/>

6. 研究組織

(1) 研究代表者

加藤 有己 (KATO YUKI)

奈良先端科学技術大学院大学・情報科学研究科・助教

研究者番号: 10511280

(2) 研究分担者

()

研究者番号:

(3) 連携研究者

()

研究者番号: