

様式 F - 7 - 1

科学研究費助成事業（学術研究助成基金助成金）実施状況報告書（研究実施状況報告書）（平成23年度）

1. 機関番号

1	4	6	0	3
---	---	---	---	---

 2. 研究機関名 奈良先端科学技術大学院大学
3. 研究種目名 挑戦的萌芽研究 4. 補助事業期間 平成23年度～平成24年度
5. 課題番号

2	3	6	5	0	1	5	3
---	---	---	---	---	---	---	---
6. 研究課題 機械学習と最適化に基づくRNAタンパク質相互作用予測
7. 研究代表者

研究者番号	研究代表者名	所属部局名	職名
8 0 1 9 6 9 4 8	セキ ヒロユキ 関 浩之	情報科学研究科	教授

8. 研究分担者

研究者番号	研究分担者名	所属研究機関名・部局名	職名
1 0 5 1 1 2 8 0	カトウ ユウキ 加藤 有己	情報科学研究科	助教

9. 研究実績の概要

生体内に存在する機能性RNAは多くの場合、タンパク質と結合することでその機能を発現することが知られている。また、両者とも折り畳み構造を持つことが多く、その構造が相互作用の形成に大きな影響を与えられている。RNAとタンパク質の相互作用予測は、その取り得る結合構造の複雑さやバリエーションの多さなどのため未だ確立された予測手法がなく、発展途上かつ挑戦的研究課題であると言える。本研究では、離散最適化法に基づくRNAおよびタンパク質の2次構造予測法を発展させ、RNA-タンパク質相互作用のモデル化を組織的に行い、精度の良い相互作用予測法の開発を目指す。

今年度では、基盤予測法として1本鎖RNA配列の構造解析に焦点を当て、昨年度分担者が開発したRNA間相互作用予測法RactIPの方法論を応用し、シュードノットと呼ばれる複雑な2次構造を考慮した高速な予測法を開発した。まず、シュードノットを含む2次構造の事後確率分布を、シュードノットを含まない2次構造の確率分布の積へ分解を行った。次に、最適化における目的関数の設定において、予測2次構造の期待精度の最大化に主眼を置き、期待精度最大化問題を閾値カット付き整数計画問題として実現した。さらに、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モデルの拡張を行った。提案手法（IPknot）の性能評価を構造既知の配列データセットを用いて行ったところ、複数の既存手法と比べて精度の点では同等以上、速度の点では桁違いの高速性を実現した。最後に、研究成果を広く世界に向けて発信するため、IPknotのWebサーバーを開発し、今年度補完したRactIPのサーバーと統合することで、世界最高レベルの高速性を実現するRNA構造解析ツールセットを公開した。

10. キーワード

(1) RNA-タンパク質相互作用	(2) RNA 2 次構造	(3) RNA間相互作用	(4) 機械学習
(5) 最適化	(6)	(7)	(8)

11. 現在までの達成度

(区分)(3) やや遅れている。

(理由)

本課題の主な目的はRNAとタンパク質の相互作用予測であり、単一RNA配列および多重配列に対する構造予測法の開発はあくまでその基盤を与えるものである。もちろん、シュドノットを含めた複雑な構造を高速かつ高精度に行うことは、相互作用予測を行う準備として十分貢献し得るものと確信する。また、分担者が昨年度開発したRNA間相互作用予測法は本課題遂行において重要な基盤技術を含むと考えている。今年度の進捗状況がやや芳しくない判断する理由は以下の通りである。相互作用予測に向けて詳細な調査を行って得られた知見は、RNAとタンパク質の相互作用を予測するためには、その両者の立体構造に近い情報までの予測を行わなければ、正確な予測が難しいということである。本課題は配列情報のみから解析を行うことに主眼を置いている。配列情報から立体構造などの高次構造を予測することはRNAに限って言及すれば容易ではなく、現在発展途上の課題となっている。特に、RNAのA-U、C-G、G-U以外の非正準塩基対、および立体構造上で観測される同軸スタックを予測することは立体構造予測への大きな足がかりとして期待されている一方で、計算機科学的な手法が必要となるスコア関数が十分に提案されていない状況でもある。現在、非正準塩基対と同軸スタックを予測するためのデコーディングモデルはほぼ完成しているが、最適化に用いるスコア関数については検討中である。一方、タンパク質の構造予測は過去に2次構造の一要素であるベータシートに限定して行った経緯があるが、それ以後の進展はないため、何らかの手法で得られた構造情報をうまくモデルに組み込むことが、今後の課題として行うべきものとなる。最後に、相互作用する生体分子のクラスをある程度限定する必要があると考えており、その上で機械学習を用いてスコア関数を定量化する必要があると思われる。

12. 今後の研究の推進方策 等

(今後の推進方策)

(1) 機械学習によるスコア関数の計算

まず、非正準塩基対確率分布の計算に用いるパラメータをマージン最大化問題などに定式化して学習する。各確率値は後に整数計画問題の目的関数に組み込まれる。また、タンパク質構造情報も同様に目的関数に統合する。相互作用の情報はRNAの1塩基とタンパク質の1アミノ酸残基の相互作用スコアによって表現する。そして、既知のRNA-タンパク質複合体を訓練データとして相互作用スコアを計算するアルゴリズムを開発する。ここでは、サポートベクトルマシンに基づく機械学習アルゴリズムを用いる予定である。また、1塩基-1残基モデルで十分な精度が得られない場合を想定し、1塩基-1残基ドメイン(短い残基列)モデルに基づくスコア計算アルゴリズムも設計する。

(2) 整数計画法によるRNA-タンパク質相互作用予測

RNA-タンパク質相互作用予測に対して計算コストの削減が可能となるような定式化を考える。ここでは、予測結合構造の期待精度を最大化することが目標となる。そのため、RNAとタンパク質の構造及びその相互作用に対応するスコアに、適切な確率を割り当てる必要がある。RNA 2次構造に対しては、非正準塩基対を考慮した塩基対確率を利用することを考える。タンパク質 2次構造に対しては、アミノ酸残基間のコンタクトポテンシャルを確率に変換することである程度対応できると思われる。また、相互作用に対しては、手順1のスコアを用いることにする。

(3) 計算機実験

上述の整数計画問題を解くツールをJava言語を用いて計算機に実装する。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-タンパク質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討を行う。

(次年度の研究費の使用計画)

未使用額が生じた要因は、研究の進捗状況に合わせ、予算執行計画を変更したことに伴うものである。また、次年度の請求額と合わせての執行計画は以下のとおりである。

物品費として、研究調査のための情報科学関連および生体生命情報学関連の書籍を購入するため書籍費を計上している。また、研究補助を行う人材が実装予定のシステム開発を円滑に行うためのノートパソコン1台の経費を計上している。

旅費として、研究成果発表(国内2回、海外1回)を行うための経費を計上している。

提案アルゴリズムの実装、計算機実験には多大な労力を要するため、高度なプログラミング能力を備えた人材に協力を要請する。そのための謝金(1人×6カ月/年)を計上している。

その他として、研究成果発表の場である国際会議での論文発表を計画している。そのための国際会議参加費を計上している。また、オープンアクセスジャーナルのための論文掲載費を計上している。

13.研究発表(平成23年度の研究成果)

〔雑誌論文〕計(2)件 うち査読付論文 計(2)件

著者名		論文標題			
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			
雑誌名	査読の有無	巻	発行年		最初と最後の頁
Bioinformatics	有	27(13)	2	0 1 1	i85-i93
掲載論文のDOI(デジタルオブジェクト識別子)					
10.1093/bioinformatics/btr215					

著者名		論文標題			
Unyane Poolsap, Yuki Kato, Kengo Sato and Tatsuya Akutsu		Using binding profiles to predict binding sites of target RNAs			
雑誌名	査読の有無	巻	発行年		最初と最後の頁
Journal of Bioinformatics and Computational Biology	有	9(6)	2	0 1 1	697-713
掲載論文のDOI(デジタルオブジェクト識別子)					
10.1142/S0219720011005628					

〔学会発表〕計(10)件 うち招待講演 計(1)件

発表者名		発表標題	
Hiroyuki Seki		A Comparative Approach to RNA Pseudoknotted Structure Prediction Based on Multiple Context-Free Grammar	
学会等名	発表年月日	発表場所	
8th Asian Workshop on Foundation of Software	2011年5月13日	中国 上海	

発表者名	発表標題	
Yuki Kato	RactIP: fast and accurate prediction of RNA-RNA interaction using integer programming	
学会等名	発表年月日	発表場所
Sixteenth Annual Meeting of the RNA Society (RNA2011)	2011年6月15日	京都府京都市

発表者名	発表標題	
Kengo Sato	IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming	
学会等名	発表年月日	発表場所
Sixteenth Annual Meeting of the RNA Society (RNA2011)	2011年6月17日	京都府京都市

発表者名	発表標題	
Kengo Sato	IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming	
学会等名	発表年月日	発表場所
19th Annual International Conference on Intelligent Systems for Molecular Biology and 10th European Conference on Computational Biology (ISMB/ECCB2011)	2011年7月17日	オーストリア ウィーン

発表者名	発表標題	
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	
学会等名	発表年月日	発表場所
次世代バイオインフォマティクス研究会(招待講演)	2011年8月2日	北海道札幌市

発表者名	発表標題	
Yuki Kato	Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming	
学会等名	発表年月日	発表場所
ICR Symposium to Celebrate the Bioinformatics Center's 10 Year Anniversary and New Restructuring	2011年8月29日	京都府宇治市

発表者名	発表標題	
Yuki Kato	Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming	
学会等名	発表年月日	発表場所
CBI/JSBi2011合同大会	2011年11月8日	兵庫県神戸市

発表者名	発表標題	
Kengo Sato	Simultaneous aligning and folding of RNA sequences by dual decomposition	
学会等名	発表年月日	発表場所
CBI/JSBi2011合同大会	2011年11月10日	兵庫県神戸市

発表者名	発表標題	
Kengo Sato	Rtips: fast and accurate tools for RNA 2D structure prediction using integer programming	
学会等名	発表年月日	発表場所
第34回日本分子生物学会年会	2011年12月14日	神奈川県横浜市

発表者名		発表標題	
佐藤 健吾		双対分解によるRNA構造アラインメント	
学会等名		発表年月日	発表場所
第28回情報処理学会バイオ情報学研究会		2012年3月28日	宮城県仙台市

〔図書〕計(1)件

著者名		出版社			
加藤 有己		羊土社			
書名				発行年	総ページ数
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」				2011	229-236

14. 研究成果による産業財産権の出願・取得状況

〔出願〕計(0)件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	出願年月日	国内・外国の別

〔取得〕計(0)件

産業財産権の名称	発明者	権利者	産業財産権の種類、番号	取得年月日	国内・外国の別
				出願年月日	

15. 備考

http://rna.naist.jp/
