

平成21年5月31日現在

研究種目：基盤研究(G)

研究期間：2007～2008

課題番号：19570157

研究課題名（和文）生体高分子間相互作用構造推定のため情報抽出法の開発

研究課題名（英文）Development of a method to obtain information for prediction of interacting structures of biomolecules

研究代表者

由良 敬 (YURA KEI)

お茶の水女子大学・大学院人間文化創成科学研究科・教授

研究者番号：50252226

研究成果の概要：細胞内の生体高分子（タンパク質やRNAなど）は、お互いに相互作用して働いている。生体高分子単体がどのような形をしているのかは、ずいぶん明らかにされてきたが、生体高分子のどの部分で相互作用するのか、また相互作用するとどのような形になるのかは、数少ない場合でしかわかっていない。本補助金研究では、相互作用部位と相互作用構造をコンピュータを用いて推定する方法の基礎研究を行い、それぞれにおいて新しい方法を開発することができた。

交付額

(金額単位：円)

	直接経費	間接経費	合計
2007年度	2,800,000	840,000	3,640,000
2008年度	700,000	210,000	910,000
年度			
年度			
年度			
総計	3,500,000	1,050,000	4,550,000

研究分野：生物学

科研費の分科・細目：生物科学・生物物理学

キーワード：タンパク質間相互作用、タンパク質-RNA相互作用、シングレット、ダブレット、相互作用推定、タンパク質立体構造、計算生物学

1. 研究開始当初の背景

多くの生物種におけるゲノム塩基配列が決定し、ある生物がもつ全タンパク質のアミノ酸配列がわかるようになった。次に明らかにすべきことは、ゲノムにコードされているタンパク質やRNAがどのような相互作用をおこなって、細胞を形成しているかである。タンパク質間相互作用及びタンパク質と核酸高分子の相互作用の同定は、様々な手法で実験的に明らかにされてきている。タンパク質間及びタンパク質と核酸高分子(まとめて生体高分子とよぶ)の相互作用測定で明らかにな

ってきていることは、どの生体高分子がどの生体高分子と相互作用する場合があるかであり、その情報を積み上げることで、タンパク質などのネットワークがわかってきた。しかし、この手法ではそれぞれのタンパク質がどのように相互作用しているかを知ることはできない。例えば、生体高分子のネットワークを阻害することで、新しい医薬農薬を開発することを考えるならば、原子分解能で相互作用がわからなければならない。生体高分子の相互作用を原子分解能で解明するために、生体高分子の複合体立体構造解析の必

要性が主張されている。しかし、複合体構造の結晶化は容易なことではなく、産出されるデータ量は、タンパク質ネットワークの情報量と比較すると圧倒的に少ない。ゲノム塩基配列決定プロジェクトと構造ゲノミクスプロジェクトのおかげで、様々なタンパク質のアミノ酸配列や、DNA/RNAのタンパク質相互作用部位、およびタンパク質単体の原子分解能立体構造は多くわかってきている。そこで、(1)大量の核酸配列にもとづく生体高分子相互作用部位の高精度推定と、(2)生体高分子単体の構造から複合体構造の高精度推定ができれば、生体高分子ネットワークの情報を原子分解能まで高めることができ、創薬などの一助となることのできるはずである。

2. 研究の目的

立体構造が複合体で判明している生体高分子から得られる統計的相互作用情報とゲノム塩基配列から得られる進化的情報にもとづき、本申請研究は以下の2点を目的とする。

(1)RNAのどの部分にタンパク質が相互作用するかを予測する方法を開発する。

(2)タンパク質が相互作用する際の界面がどこにあるかを推定する手法を開発する。

3. 研究の方法

(1)RNAのどの部分にタンパク質が相互作用するか:陸上植物の葉緑体のゲノムから転写される多くのメッセンジャーRNAは翻訳される前に編集を受けることが知られている(RNA編集)。シトシン(C)がウラシル(U)に変化する化学反応が、核にコードされている酵素によって触媒される。この反応はRNAのすべてのCで起こるわけではなく、ある特定の定まったCのみがUに特異的に変化する。酵素が特定のCを認識することで、編集が達成されている。酵素がどのようにして特別なCを認識するのかは、明らかにされていない。またすべてを実験的に明らかにするには膨大な労力を要する。よってこの酵素がRNAのどの部分を認識するかをコンピュータを用いて推定することが重要になる。そこで実験的に確かめられている編集部位前後の塩基配列情報を用いて、酵素によって認識される部位を推定する方法を構築する。共同研究者が収集したナンジャモンジャゴケ葉緑体のRNA編集部位のデータを整理し、編集部位前後の塩基配列の特徴を抽出する。編集部位から特定の距離離れた部位にどのような核酸が統計的に現れやすいかを計算する。また編集部位から特定の距離離れている2カ所の部分でどのような核酸ペアが統計的に現れやすいかを計算する。その結果得られる傾向値をもとにして、ナンジャモンジャゴケ葉緑体のゲノム塩基配列からRNA編集

部位未同定部分のRNA編集部位を予測する。その予測部位が、実際にRNA編集を受けているかどうかを実験的に検証することで、予測法の精度を確認する。

(2)タンパク質が相互作用する際の界面がどこにあるか:タンパク質の複合体立体構造が非常に多くわかってきている(1000組程度)。しかし相互作用構造を明らかにしたいタンパク質群の数は、構造既知相互作用よりもはるかに多い。そこで、1000組程度の複合体構造から、タンパク質相互作用界面にどのようなアミノ酸残基が統計的に出現しやすいかを測定する。アミノ酸残基単体(シングレット)の出現頻度と、アミノ酸残基ペア(ダブルット)の出現頻度、およびタンパク質表面のアミノ酸残基単体とペアの出現頻度を測定し、それらの値から、タンパク質界面におけるアミノ酸残基の出現スコアを構築する。そのスコアにもとづきタンパク質相互作用面を推定する。

4. 研究成果

(1)RNA相互作用部位の予測:共同研究者が同定したナンジャモンジャゴケ葉緑体のRNA編集部位302カ所の前後40残基の配列を、その核酸配列類似性にもとづいて8個のグループに分類することができた(図1)。

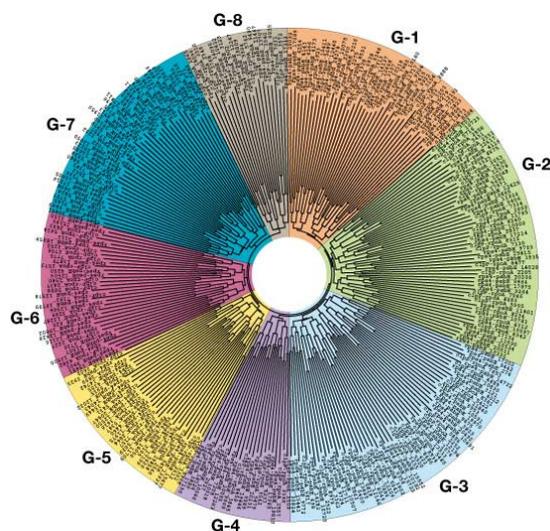


図1:RNA配列の分類結果

各分類の中で、ある部位にどのような塩基が現れやすいか(シングレット傾向)を調べたところ、各分類ごとに異なった傾向があることがわかった(図2)。注目すべきことは、ある部位に好まれる塩基があることだけでなく、ある部位で出現が低く押さえられている塩基種があることである。例えば、第2番目のグループ(G-2)においては、CからUに編集される塩基の2つ5'側の塩基が

Tである傾向が非常に強い。さらに3塩基5'側の部位では、Cが極端に避けられていることがわかった。

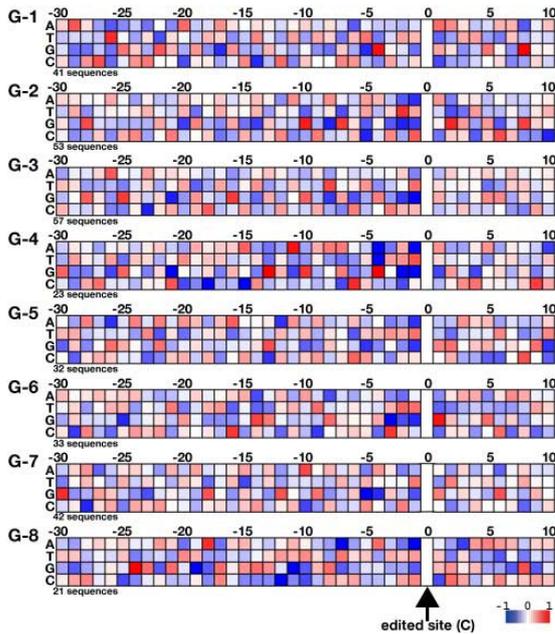


図2: シングレット傾向

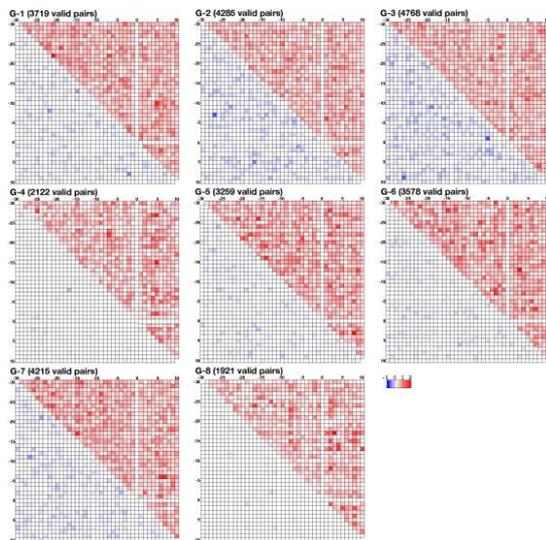


図3: ダブルレット傾向

RNAは塩基のペアリングによりステム構造を形成することが知られている。RNA編集部位の前後にステム構造が形成され、この構造を酵素が認識しているとすれば、編集部位前後で塩基ペアに特異な出現頻度が見えるはずである。そこで、分類ごとにペアの出現頻度(ダブルレット傾向)を調べた(図3)。核酸配列の量が少ないためあまり明確な傾向は見えていないが、それでも好まれて出現するペアがあることがわかった。ところが多くのペアは、一般的には水素結合をするペアではなく CC や GG のペアが頻繁に見

られることがわかった。よってここで見えたダブルレット傾向は、ステム構造の形成とは関係がないと考えられる。

シングレット傾向とダブルレット傾向を用いることで、31番目がCである40塩基から形成されるRNA塩基配列が、RNA編集酵素に認識されるかどうかを推定する方法を作ることができた。その方法を用いて、傾向値を導出したデータに対して編集部位を予測すると、図4のように実際に編集されることが明らかになっている配列はすべて高いスコアになることがわかった。スコアを導出したデータに対する予測なので、このことは当然の帰結である。

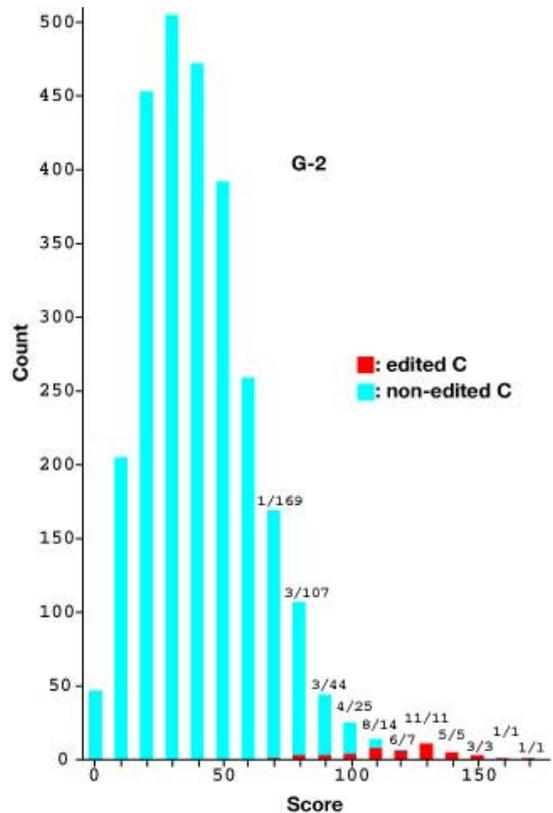


図4: スコアを形成したデータに対する予測結果

そこで、ナンジャモンジャゴケ葉緑体のRNA編集を同定していないRNAに対して、CがUに編集されるかどうかを予測した(表1)。その後RNAの配列を実験的に決定し、RNA編集が起こっているか否かを確認した。その結果、約60%の部位でRNA編集部位を正しく推定することに成功した。ここで示した60%とは、予測した部位のうち60%が本当に編集を受けていたことを意味する。実際にはもっと多くの場所が編集を受けていることがわかっているため、本予測方法には多くの取りこぼしがあり、この取りこぼしをどのようにして少なくしていくかが、今後の課題である。しかし、本予測による方法は実験家との今後の共同研究には有効であると考えている。

表1:新規予測結果と確認実験の結果

Predicted edited site with -30 to +10 nucleotides	Gene	Position*	Score	Codon*	Unedited - edited
TTTGTGATTTCTGATCTACAGAACTGCTTCTTCTTCCA	<i>clpP</i>	-6243	120.00	199	S(Ca) → L(Ts)
AAGTCTGATCTACTGACAGGCTTCTGCTTCTTCCA	<i>pefL</i>	1897	112.01	1	T(Gs) → M(Tg)
TATTTCTGCTTCTGATTTTCTGACTTCTTCTTCTTCTG	<i>pefL</i>	1957	109.97	21	S(Ca) → L(Ts)
AAGATTCACGACGCTGCTGCTTCTTCTTCTGACGAGAG	<i>clpP</i>	-6440	108.17		false positive
AAGTTFATGCGATTTGATTTCTGACTGACAAAGATGCT	<i>clpP</i>	-6255	107.40	195	T(Ga) → (GtA)
CTCCCGTGGAGGATTTACGAGGATATGTTTATGAT	<i>clpP</i>	-7143	106.91		false positive
GGATTCACGAGGCTGATTTGAACTATTTGACTCTTAA	<i>rip20</i>	-4681	106.11	33	R(Cga) → T(Ga)
CAATGCAAGATGATAAAGATCTTCTTCTGACGACTCTG	<i>psuJ</i>	3317	106.02		false positive
AAATGTTGAAAAAATAAATGATGCTTCTTCTTCTTCT	<i>rip33</i>	3834	105.03		false positive
AATTTCTGCTGACAGGCTTCTTCTTCTGACTCTG	<i>clpP</i>	-6237	103.90		false positive
TTTGGCTTCTGAGAGAGATGCTTCTGCTGATGCT	<i>clpP</i>	-8036	103.15	27	L(Ct) → L(Ct)
AGAGCAAAATTTATGATGCTTCTTCTTCTTCTGACG	<i>clpP</i>	-6263	101.19	192	F(Hc) → F(tT)

本研究で開発した方法は、RNA塩基配列のみからRNA編集部位を予測する方法であり、これはまったく新規の方法である。RNAにコードされているタンパク質のアミノ酸配列や、類縁タンパク質のアミノ酸配列の比較情報を用いてRNA編集部位を推定する方法は、従来からある。しかし、これらの方法は、RNAが実際に編集される現場には存在しない情報を用いて予測をしているため、予測方法をいくら開発しても、RNA編集の機構にせまることができない。また今回開発した方法では、アミノ酸残基が変化しないRNA編集部位も推定することに成功している。これは今までの方法では予測できなかった部位である(雑誌論文①、学会発表①⑤など)。今回はある特別な生物種にのみこの方法を適応して成功したが、他の生物種のRNA編集酵素についてもうまいくと考えており、平成21年度以降に対象生物の拡張を予定している。

(2)タンパク質相互作用部位の予測:タンパク質-タンパク質複合体のデータをタンパク質立体構造データベースPDBから抽出したところ、お互いに独立な複合体構造を1007個得ることができた。それらの溶媒接触表面積と生体高分子単体の溶媒接触表面積を残基単位と原子単位で計算し、接触表面積の差から生体高分子の相互作用面を同定することができた。同定された全相互作用面におけるアミノ酸残基単体の出現頻度を数え、各アミノ酸残基が相互作用面に現れる傾向を測定した。さらに、相互作用面を形成する残基において、同じタンパク質に属し空間的に近傍に位置するアミノ酸残基種ペアの出現頻度を数え上げ、ダブレット頻度を測定することができた(図5)。

この頻度から残基種ペアの出現に特徴を見いだすことができた。シングレットとダブレットの両パラメータを用いて、タンパク質間相互作用面の推定を行った。その結果約60%の精度で相互作用に関与するアミノ酸残基を推定することができた。このパラメータを用いて、共同研究者が立体構造を決定した運動タンパク質の相互作用面を推定したところ、NMRによって測定された相互作用界面と整合性のある結果を得ることができた(論文投稿中)。

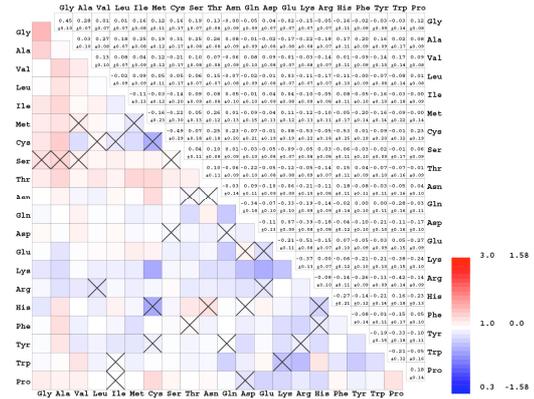


図5:ダブレット傾向

また Signal Recognition Particle 54 M (SRP54M) ドメインのシグナルペプチド結合部位の推定もおこなった(図6)。SRP54M はリボソームから出てようとしているペプチドが特別な配列を持っているときに、その部分に結合することがわかっているが、SRP54M のどの部分でペプチドに結合するかは明らかにされていない。そこで、今回開発した方法で、ペプチド結合部位を推定したところ、図6の赤色になっている Met 残基を中心とする表面が結合部位候補として予測された。この予測の実験的検証はまだ行われていない(雑誌論文②、学会発表⑥⑦など)。

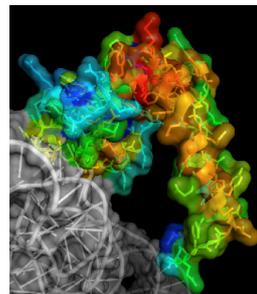


図6:Signal Recognition Particle 54 M Domain (SRP54M)の推定シグナルペプチド結合部位(赤色部分)

さらにタンパク質複合体推定国際コンテスト(CAPRI)に共同研究者とともに参加し、ここで構築した予測方法を試した。その結果、ある程度の成果を得ることができた。出題されたタンパク質複合体構造予測問題では、単体の構造がすでにX線結晶解析で明らかになっており、複合体構造もすでに判明しているが公開はされていない。単体の構造から複合体を形成する界面を、本研究で開発した方法で予測し、その情報をもとにして複合体構造の推定を試みた。実験的に明らかになった構造予測結果と全出場者の予測構造とを出題者が比較したところ、本方法で正確な位置を推定することはできていないことが明らかになったが、本予測がドッキング構造を構築するための束縛条件にはなり得ることが明らかになった。界面傾向値の導出法は平成21

年度に論文として発表する予定である。

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3件)

- ① Yura, K., Miyata, Y., Arikawa, T., Higuchi, M., Sugita, M. "Characteristics and prediction of RNA editing sites in transcripts of the moss *Takakia lepidozoides* chloroplast" *DNA Research*, **15**(5), 309-321, (2008). 査読あり.
- ② Yura, K. "Prediction of Protein-RNA, Protein-Protein and Protein-Ligand Interfaces Based on Protein Sequences and 3D-Structures" *Proceedings of International Symposium on Frontiers of Computational Science 2008*, 77-86 (2008). 査読あり.
- ③ 由良 敬 "タンパク質の立体構造に基づく相互作用構造の推定"、*薬学雑誌*, **128**(11), 1547-1555, (2008). 査読なし.

[学会発表] (計7件)

- ① 由良敬 「ナンジャモンジャゴケ葉緑体ゲノムデータに基づく RNA エディティング部位の予測」 *プロテイン・インフォマティクス: 蛋白質科学のためのバイオインフォマティクス*, 第9回日本蛋白質科学会年会ワークショップ, 2009年5月20日～22日, 熊本.
- ② Yura, K. "Prediction of Protein-RNA, Protein-Protein And Protein-Ligand Interfaces Based on Protein Sequences and 3D-Structures", *ATI International Forum 2009 "Protein Structure Determination and Applications"*, 2009年3月9日～10日, 日本原子力研究開発機構.
- ③ 由良敬 「たくさんのデータ観察からわかるタンパク質が低分子を認識するようす」 *第3回お茶の水女子大学糖鎖科学研究教育センター公開シンポジウム「糖鎖が語る生命と病気」* 2009年3月5日, お茶の水女子大学.
- ④ Yura, K. "Prediction of interface residues on proteins for RNA and protein molecules", *Korea-Japan Seminars on Biomolecular Sciences - Experiments and Simulations*, 2009年2月27日～3月2日, 韓国ソウル KIAS.
- ⑤ Yura, K., Miyata, Y., Arikawa, T., Higuchi, M., Sugita, M. " Prediction of RNA editing sites solely from the newly characterized DNA sequences of the moss *Takakia lepidozoides* chloroplast", 第46回日本生物物理学会年会(福岡大会), 2008年12月3日～5日, 福岡国際会議場.
- ⑥ Yura, K. "Prediction of Protein-RNA, Protein-Protein and Protein-Ligand Interfaces Based on Protein Sequences and 3D-Structures", *Frontiers of Computational*

Biology on Protein Structures, 2008年12月2日, お茶の水女子大学.

- ⑦ Yura, K. "Prediction of Protein-RNA, Protein-Protein and Protein-Ligand Interfaces Based on Protein Sequences and 3D-Structures", *International Symposium on Frontiers of Computational Science 2008*, 2008年11月27日～29日, 名古屋大学.

[図書] (計1件)

- ① Kinoshita, K., Kono, H., Yura, K. "Prediction of molecular interactions from 3D-structures: from small ligands to large protein complexes", in *Prediction of Protein Structures, Functions, and Interactions* (ed. Bujnicki, J.M.), Wiley and Sons, 159-186 (2009).

[その他]

ホームページなど

<http://cib.cf.ocha.ac.jp/~yura/present0.html>

<http://cib.cf.ocha.ac.jp/~yura/RNAE/>

6. 研究組織

(1) 研究代表者

由良 敬 (YURA KEI)

お茶の水女子大学・

大学院人間文化創成科学研究科・教授

研究者番号: 50252226

(2) 研究分担者

該当なし

(3) 連携研究者

該当なし