

# RNA とタンパク質の相互作用におけるタンパク質構造変化のデータ解析

理学専攻・情報科学コース 久保田千尋（指導教員：小林一郎，副指導教員：由良 敬）

## 1 研究背景

ゲノム塩基配列の読み取り完了後、数多くの RNA 分子が転写され何らかの機能を担っていることが明らかになってきた。これらの RNA 分子は、単独で機能しているわけではなく、他の様々な分子、特にタンパク質と相互作用して生体中で機能している[1]。この機能を明らかにする上では、タンパク質と RNA との複合体構造を知ることが不可欠だが、実験的に複合体構造を決定することは、容易ではない。また細胞中には数多くのタンパク質と RNA との複合体が存在していることから、バイオインフォマティクスの手法で複合体構造を推定することが不可欠になりつつある。

## 2 研究目的

すべてのタンパク質の立体構造をアミノ酸配列から推定することは、近い将来可能になることがわかっている [2,3]。立体構造が明らかになった上で、これらのタンパク質の機能を知るためには、タンパク質と他分子との相互作用構造を構築する必要がある。しかし、タンパク質と他分子が相互作用している状態の構造を推定する技術はまだ確立していない。たとえ、RNA 結合タンパク質の立体構造が明らかになったとしても、その機能様式を知るための技術がまだない。この技術を確立するためには、相互作用時のタンパク質の立体構造と RNA の立体構造、およびそれらがドッキングした構造を推定できるようになる必要がある。これらの必要技術の中で、一番開発が遅れているのは、相互作用時のタンパク質の立体構造推定技術である。多くの先行研究においては、タンパク質が単体で存在する時と、RNA と相互作用している時とでは、タンパク質の立体構造に変化はないことを仮定している[4]。そこで本研究では、タンパク質が RNA と結合した際に、タンパク質の立体構造がどの程度変化するかを明らかにすることを目的とした。仮定が正しいかを示すとともに、正しくない場合には、どのようにして相互作用時の立体構造を推定するかを明らかにする。

## 3 手法

バイオインフォマティクスの分野において、数多くのデータベースが存在する中、タンパク質と RNA との相互作用の様子を明確に整理し、その三次元構造を詳細に表示する Web データベースは存在しない。近年、タンパク質と RNA との複合体立体構造解析結果が増加している。それらを集積したデータベースがあれば、相互作用の詳細や構造変化をとらえることができると考え、Web データベースの開発に至った。

### 3.1 データ抽出方法

RNA 結合タンパク質が、RNA に結合している状態と結合していない状態の両方の状態で立体構造が明らかになっている場合を見だし、それらの状態をペアにすることで、構造変化の解析ができるようにする。Web データベースを作成するにあたって、原子座標、タンパク質の名称、タンパク質が由来する生物種名のデータを PDB より抽出した。両者の座標を得るにあたって、PDB から RNA に結合しているタンパク質と、そのタンパク質とアミノ酸配列が類似で RNA に結合していないタンパク質のエントリの抽出を試みた。

#### 3.1.1 PDB エントリのグループ分け

PDB のすべてのエントリ ID と鎖（ひとつのエントリには複数の分子（鎖）が含まれている）の情報からアミノ酸配列を抽出した。抽出したデータに対して、BLASTClust を用いてアミノ酸配列の類似性によるグループ分けを行った。グループ分けの基準は、アミノ酸配列の一致度が 90%以上、アミノ酸配列の相互包括度が 90%以上とした。

#### 3.1.2 RNA と結合しているタンパク質の抽出

RNA とタンパク質の両者の情報が入っているエントリを PDB の全エントリから抽出した。抽出した PDB エントリにおいて、どのタンパク質の鎖（分子）がどの RNA の鎖（分子）に結合しているかを計算した。タンパク質を構成する重原子と RNA を構成する重原子の距離が 4Å以下の場合にはそれらの原子は結合している（ファンデルワールス相互作用をしている）と見なした。

#### 3.1.3 RNA に結合するタンパク質の RNA と結合している状態と結合していない状態のペアデータの抽出

3.1.2 で抽出したタンパク質の鎖を 3.1.1 で分類したグループの中から探した。3.1.1 のグループの中で、RNA と結合しているタンパク質と RNA と結合していないタンパク質で可能なペアすべてを作った。このペアの中で、X 線結晶解析の解像度が一番よく、かつ座標が明らかになっている原子が最も多いペアを代表として選んだ。

### 3.2 Web データベースの概略

3.1 で述べた手法により得られたデータを、HTML と PHP を用いて、Web 上で検索閲覧できるようにした。図 1 に構築した Web データベースと Web ツールの概要を示す。

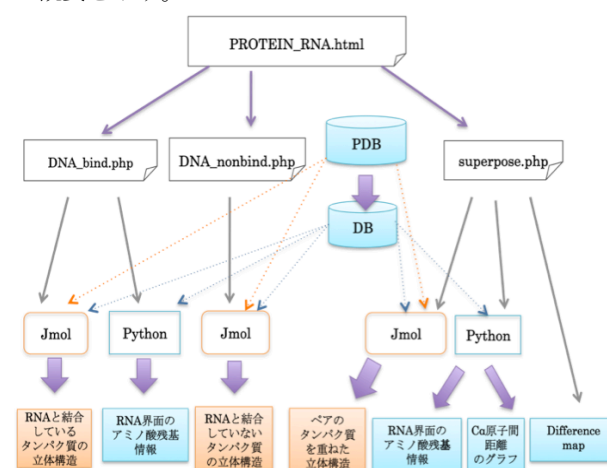


図 1 : Web データベースと Web ツールの流れ

各エントリ ID をクリックすることで、そのエントリにおけるタンパク質の立体構造を見ることができる。立体構造表示には、Jmol[5]を用いた。立体構造の表示では、RNA と相互作用している状態の座標、相互作用していない状態の座標、および両者の変化を視覚的に

理解するために両座標を重ね合わせ座標にもとづいて、立体構造を可視化した。立体構造比較には、Jmol での立体構造表示以外に、Difference Map (同一分子内の C $\alpha$  原子間距離の変化の度合いを可視化)、異なる状態の対応する C $\alpha$  原子間の距離のグラフを用いた。なお C $\alpha$  原子とは、アミノ残基においてアミノ基とカルボキシル基が共有結合している炭素のことであり、アミノ酸残基を代表する原子として用いられることが多い。

## 4 結果

### 4.1 データの抽出

PDB を検索した結果、RNA に結合した状態と結合していない状態の両方の立体構造が判明しているタンパク質を 108 組抽出することができた。

### 4.2 構造変化の程度によるグループ分け

108 組の RNA 結合タンパク質の立体構造がそれぞれの程度異なっているかを知るために、C $\alpha$  原子の RMSD を計算し、その値を用いてヒストグラムを作成した (図 2)。RMSD とは、2 つのタンパク質の各アミノ酸残基の C $\alpha$  原子の座標をずれが最も小さくなるように回転させた時のずれの二乗の平均値である。図 2 では、横軸が C $\alpha$  原子の RMSD の大きさ、縦軸がタンパク質のペアの個数をしめす。横軸の単位は Å である。横軸の値が大きいほど、タンパク質ペアの構造の差異が大きい。図 2 より、RNA 結合タンパク質において、構造変化の程度は多くの場合 RMSD が 2.0 Å 未満であり、2.0 Å 以上の変化をする場合はまれであることがわかった。RNA 結合タンパク質の構造変化を RMSD が 2.0 Å 未満の場合、2.0 Å 以上 10.0 Å 未満の場合、および 10.0 Å 以上の 3 つのグループに分けた場合、2.0 Å 未満のペアは 78 個 (約 72%)、2.0 Å 以上 10.0 Å 未満のペアは 23 個 (約 21%)、10.0 Å 以上のペアは 7 個 (約 7%) であった。

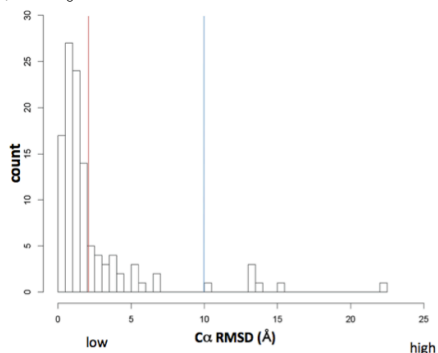


図 2 : C $\alpha$  原子の RMSD によるヒストグラム

### 4.3 RNA 結合タンパク質の構造変化

7 割以上の RNA 結合タンパク質において、RNA に結合する際の構造変化が、RMSD 2.0 Å 未満であることがわかった。この差異は測定誤差程度であり、RNA 結合タンパク質が RNA に結合する際に構造変化がほとんどないことを意味する。残り 3 割の構造変化について、詳細を以下に記す。

RMSD 10.0 Å 以上の変化の様子を図 3 に示す。図 3 左上にタンパク質の RNA に結合している状態 (赤色) (PDB ID: 1h3e) と RNA に結合していない状態 (黄色) (PDB ID: 1h3f) とを重ね合わせた結果を表示した。RNA と結合している状態のタンパク質と RNA と結合していない状態のタンパク質は互いに大きくねじれた構造をとっている。また、RNA と結合しているタンパク質は、RNA と結合していないタンパク質より RNA に近づくような構造変化を起こしていることがわかった。Difference Map では、同一分子内の C $\alpha$  原子間距

離が状態変化によって広がる場合を赤色で、狭まる場合を青色で、それらの程度に応じて白色からのグラデーションで描いている。図 3 右上の Difference Map の下部に青い領域があり、これはタンパク質が RNA と結合する際に、小型のドメインが RNA の方へ近づき、全体的に小さくなっている (慣性半径が小さくなる) ことを示している。この動きは図 3 左上の比較からも定性的に読み取ることができる。C $\alpha$  原子間距離のグラフでは、RNA と結合しているところ (図 3 下のグラフの上部に記した) において、C $\alpha$  原子間のずれが大きくなるとは限らないことがわかった。このデータでは、ドメイン構造そのものは大きな構造変化はなく、ドメインの位置関係が変化することで RNA と相互作用するようになり、その結果として RMSD が大きくなることがわかった。他のデータでも同様の傾向が見られた。

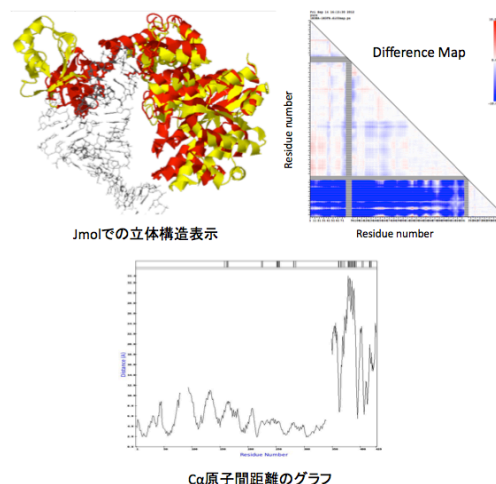


図 3 : 高度高熱菌 *Thermus thermophilus* 由来チロシルアミノアシル tRNA 合成酵素の構造変化

## 5 結論

今回の研究により、PDB から抽出した 108 組の RNA 結合タンパク質のうち、7 割以上の RNA 結合タンパク質において、RNA に結合する際に構造変化がほとんどないことがわかった。残り約 3 割の RNA 結合タンパク質において、作成した Web ツールを用いて構造変化を解析したところ、これらのタンパク質が RNA に結合する際には、タンパク質のドメイン内部の大きな構造変化はなく、構造変化はドメインの相対位置の変化に限られることがわかった。以上より、タンパク質のみの立体構造から RNA との複合体構造を推定するにあたり、ドメイン内部の構造変化は考慮する必要がないことがわかった。この事実は、タンパク質と RNA との複合体立体構造を予測するための重要な一歩となる。

### 参考文献

- [1] Keene J.D., 2001, *Proc. Natl. Acad. Sci. U.S.A.*, **98**, 7018-7024.
- [2] Yura K., Yamaguchi A., Go M. 2006, *J. Struct. Funct. Genomics*, **7**, 65-76.
- [3] Go M., Yura, K., Shionyu M. 2007 "Frontiers of Computational Science" eds. by Kaneda Y., Kawamura H., Sasai M., Springer-Verlag, Berlin, pp. 75-80.
- [4] Ellis J.J., Jones S. 2007, Evaluating conformational changes in protein structures binding RNA. *PROTEINS*, **70**, 1518-1526.
- [5] Jmol, <http://jmol.sourceforge.net/>