

## 論文の内容の要旨

論文題目 RNA 転写過程の理論的研究

氏名 木立 尚孝

近年の生物データの爆発的増大により、それから生物学的知識を抽出するための技術の必要性は益々増大している。本研究では、RNA の転写に関する三つの話題について、数理的なモデルを構築し、それを実際のデータに対して検証するという理論的研究を行った。

まず、細胞の RNA 転写過程において、DNA の張力や捩じれなどの外力が、転写速度にどのような影響を与えるかをとりあげた。RNA ポリメラーゼが、DNA から RNA を合成する時に、DNA のらせん構造によって、RNA ポリメラーゼは DNA の周りを何度も回転しなければならない。この回転を妨げる障害があるとき、DNA は捩じれを起こすので、この時 RNA ポリメラーゼは、回転のトルクを受けながら転写を行うことになる。捩れた DNA 上の転写は、実験的には、ガラス板などに RNA ポリメラーゼを固定し、端にビーズをつけた DNA 断片をその RNA ポリメラーゼに転写させることで実現させることができる。この場合には、回転の抵抗は、ビーズの流体との粘性抵抗により生じることになる。また、抗癌剤などで、Topoisomerase 等の DNA の捩れを解消する酵素の活性が阻害されているがん細胞などでも、流体相互作用や周りのたんぱく質との相互作用により、DNA の捩れが生じていると考えられ、その様な場合の転写においても、RNA ポリメラーゼが受ける回転抵抗は重要になると考えられる。私は、まず最初に、たくさんの RNA ポリメラーゼの平均化された振る舞いを記述する粗視化された力学的・化学運動論的モデルを立て、その性質を調べた。その後、これまでに行われた実験とモデルを比較した。その結果 DNA にかかる張力と、ねじれの二つの外力は、系との結合の仕方が違っており、それによ

て、振れの抵抗は、NTP 濃度が十分大きい時の最大転写速度を著しく下げるが、張力は最大転写速度に影響を与えないことも明らかになった。

次に、原核生物において転写開始のシグナルとなるプロモータ配列の違いが、対応する転写発現量つまりプロモータ強度にどのように影響するかをマイクロアレイデータを用いることで調べた。プロモータ配列と強度の間関係は、1980 年代に数多くの研究がなされた。これらの研究は、配列と強度の間に強い相関があることを明らかにしたが、その骨の折れる実験方法のために、任意の生物種の何百ものプロモータに対して実験を行うことは不可能であった。ところが近年のマイクロアレイデータの増加のおかげで、数千もの遺伝子発現情報をその DNA 配列と比較出来るようになった。私はこれに動機づけられて、大腸菌マイクロアレイデータを使ったプロモータ配列・強度間関係の抽出を行った。具体的には、これらの関係を重み行列の方法を用いてモデル化し、その各成分をサポートベクトルマシンで最適化するという事を行った。それにより得られた結果は、コンセンサス配列に近いプロモータ配列は強度が大きいこと、プロモータの-35 領域は-10 領域に比べて強度への影響が大きいことなど従来の研究で知られている結果を再現するとともに、次のような新たな結果も得ることが出来た。プロモータ配列の各塩基位置からのプロモータ強度への寄与は、其々の場所の塩基保存度から予想されるよりも遥かに変化に富むこと、-35 領域の最初の 3 つの塩基位置と-10 領域の 1, 2, 5, 6 番目の塩基位置はプロモータ強度に強く影響すること、その一方で、-35 領域の 4, 6 番目と-10 領域の 2 番目の塩基位置は、塩基の保存度にも拘らずあまりプロモータ強度に影響しないこと。15bp-19bp の範囲内では、スペーサ長の変化は保存領域からの寄与に比べて、強度への影響は小さいこと、プロモータ領域のある塩基が、コンセンサスに一致しないときのプロモータ強度の現象の程度は、塩基の種類により大きく異なることなどである。私はまた、プロモータ配列からのプロモータ強度の予測と比べて、実際の遺伝子発現量が大きく異なっているデータを見ることで、転写制御を受けている可能性のある遺伝子をいくつか同定した。これらの結果は、マイクロアレイデータが手に入る以前に行われていた、塩基出現頻度の解析や、プロモータごとの塩基置換実験によって得られなかった結果である。私の方法は、プロモータ配列が知られていて、マイクロアレイデータが手に入る他の原核生物に対しても適用可能である。

最後に、近縁種ゲノムを比較する際に、遺伝子領域で適当な塩基置換パターンを用いてアライメントを行う方法について取り上げた。近年、様々な生物種のゲノムがシーケンシングされ、それによってゲノムを対象とした生物情報解析でも比較ゲノムが大きな話題となってきている。比較ゲノムを行う際に、それぞれの DNA 領域が、他方の DNA のどの領域に対応しているかを関連付けるアライメントの技術が非常に重要である。私は、たんぱく質をコードしている領域に特徴的な塩基置換パターンがどのような統計的

性質を持っているのかを研究した。通常 DNA 配列のアライメントをするためには、アライメントのよさを置換行列と呼ばれる行列を用いて評価する。このスコア行列は、DNA の塩基置換パターンから決定される。しかし DNA 配列がどこも同じ塩基置換パターンを持つわけではなく、DNA の各領域にどのような生物の情報が記録されているかによって DNA の置換パターンは異なったものとなってくる。例えば、たんぱく質がコードされている領域は、そうでない領域に比べて、たんぱく質の機能が保たれるような進化的圧力が加わっているので、この領域ではアミノ酸を変えるような置換が起こりにくくなる。一つのアミノ酸は、3 つの隣接した DNA によりコードされているので、この領域での塩基の置換行列は、隣接する 3 塩基の置換パターンを取り入れたものであるべきである。このたんぱく質コード領域で、一塩基ごとの置換パターンのみしか考慮に入れない置換行列でアライメントを行うと、アライメントのスコアの閾値を低くとった場合には、コード領域同士で、フレームがずれたようなアライメントが出来るという問題がおこる。反対に閾値を高く取ると、コドンの 3 文字目の塩基が頻繁に置換を起こしてスコアを押し下げるコード領域では、進化的に関連しているのにアラインされないという問題が出てくる。私は、コドン同士の置換パターンから置換行列をつくり、このコドンーコドンの置換行列を用いてアライメントを行えば、これらの問題をおこさない良いアライメントが得られることを示した。