

論文審査の結果の要旨

氏名 VANDENBON ALEXIS

本論文が扱っている分野はバイオインフォマティクスと呼ばれる、ゲノム情報のコンピュータ解析である。具体的には、多細胞生物における遺伝子の組織特異的発現をコントロールする制御領域（プロモーター）がどのような原理で塩基配列上に記述されているのかという基本的な問題にアプローチしている。本論文は以下に述べるように、3部構成をとる。

第一部は、マルコフ連鎖を用いたプロモーター構造のモデリングを扱っている。組織特異的プロモーターにはしばしば頻出するシスエレメント（モチーフ）の組み合わせが存在するが、それらエレメントの一次構造上の出現順序や相対距離は一定せず、またエレメントの出現自体も完全とは言えない。このような特徴をモデル化するのに、確率的な状態遷移を扱うマルコフ連鎖を用いるのは自然である。具体例としては、線虫とカタユレイボヤにおける筋肉特異的遺伝子のプロモーター群をそれぞれ用いた。カタユレイボヤについては、既知プロモーターから構築したモデルをゲノム中の全予測遺伝子上流配列に適用して、既知プロモーター以外に存在する可能性のある未知の筋肉特異的プロモーターの発見を試みた。実験的な制約の中で、4つの候補遺伝子を選び、共同研究者の *in situ* ハイブリダイゼーション実験によって発現を確認したところ、完全に筋肉でのみに発現が抑制されているわけではないが、3つの遺伝子については筋肉での発現が確認できた。この実験による確認を含め、本研究は兵庫県立大学（当時）の日下部博士らのグループとの共同研究であるが、論文の主要な内容は論文提出者によるものであり、論文提出者の寄与は十分であると判断する。

第二部は、第一部とは別の手法、すなわち単純なルールの組み合わせによって、プロモーター構造の特徴を記述する試みについて述べている。第一部で導入したマルコフ連鎖によるモデル化は数学的に単純で見通しがよいことが魅力であるが、プロモーターの配列上の特徴を記述する能力にはやはり限界がある。また、従来様々な研究者がいわゆるシス制御モジュールの記述を試みてきたが、それらの多くは厳しい制約条件のもとに行われている。ここでは単純な配列モチーフの存在の有無、モチーフ対の相対距離範囲、翻訳開始点からのモチーフの距離範囲などをパラメーターとした様々なモチーフを生成し、それらの組み合わせを遺伝的アルゴリズムによって最適化している。具体例として、第一部でも用いた線虫の筋肉特異的発現遺伝子のプロモーターを用い、それらを少数のルールの組み合わせで記述した。得られたルールセットをゲノム中の全遺伝子上流配列に適用すると、高いスコアをもち、学習データに含まれていない遺伝子の多くは、公共のマイクロアレイデータによって、筋肉での発現が認められることが確認された。

第三部では、第二部で開発したルールベースのプロモーターモデル化技法をヒトとマウスの様々な組織特異的プロモーター群に適用し、組織によってモデル化のし易さにどんな違いがあるのかを検討している。まず一般公開されているマイクロアレイデータを用いて、ヒトでは 26 組織、マウスでは 34 組織に対する特異的発現遺伝子群を定義し、それらについて各々モデル化を行ったところ、60 中 35 の組織については統計的に有意に識別力のあるモデルを構築できたとしている。比較的モデル化が容易だった組織は、肝臓、骨格筋、腎臓、舌などであった。これらについて得られたモデルは、ヒトで得たモデルがマウスの同じ組織で有用であるなどの意味で、進化的に保存されていた。また、全部のモデルを網羅的に全組織のプロモーター領域に適用してみたところ、比較的類縁性のある他の組織においても、モデルの有効性が確認された。これらの結果は、少なくとも半数以上の組織については、妥当なモデル化が行えていることを強く示唆しているのと同時に、組織特異的プロモーターのモデル化には、組織によってかなり難易度の差があることを意味しているものと結論された。

これらの研究を通して、多細胞生物における転写制御領域（プロモーター領域）の基本的構造を網羅的かつ理論的に明らかにする研究の先鞭がつけられたと言える。審査委員の間でも、論文の体裁を修正する要求がでた以外には、めだった批判もせず、博士課程の間の研究として、十分な成果をあげているという点で意見が一致した。

従って、博士（生命科学）の学位を授与できると認める。