

# 論文審査の結果の要旨

氏名 加藤 護

真核生物の遺伝子発現は、転写因子とプロモータ上に存在する転写因子結合モチーフとの組み合わせによって制御されており、これらさまざまな組み合わせが複雑な生命現象を生み出す第一歩であると考えられている。したがって、転写因子とその結合モチーフとの組み合わせによる転写制御の機構を解明することは、真核生物における遺伝子発現とそれに関わる生命現象を理解する上で重要である。しかしながら、ゲノムレベルでこの組み合わせ的転写制御を予測する計算機的手法は存在していなかった。本論文では、DNA マイクロアレイデータ、ChIP (Chromatin Immunoprecipitation) マイクロアレイデータ、全プロモータ配列データなどの複数のゲノム・データの統合的処理により、上記の組み合わせ的転写制御を予測する計算機的手法を開発し、これを酵母細胞周期に関わる遺伝子発現制御の解析に適用している。

本論文は、第1章序論、第2章方法、第3章結果、第4章考察の4章より成り、第1章序論では、転写制御研究に用いられる DNA マイクロアレイなどのゲノム・テクノロジー、DNA マイクロアレイ解析のための計算機的手法、最近の ChIP アレイデータの出現が紹介されている。従来の計算機的手法では、DNA マイクロアレイデータから得られる遺伝子の mRNA 発現パターンの類似性に従って遺伝子をグループ分けし、分けられた遺伝子群の上流プロモータ配列の中から、配列上統計的有意に高頻出している短い DNA パターンを抽出して、それが転写因子結合モチーフであると推測していた。しかし、この方法には以下の2つの問題点がある。(1)たとえ有意な DNA パターンを抽出したとしても、そのモチーフにどの転写因子が結合するのかわからない。(2)真核生物では転写因子と結合モチーフによる組み合わせ的転写制御が重要であると考えられるが、従来の方法では種類の DNA パターンしか探索していない。

第2章方法においては、データの前処理、高頻出単一モチーフ探索、高頻出モチーフ組み合わせの探索、発現パターンのコヒーレンスの調査などに関する方法が述べられている。上記(1)の問題点を克服するために、どのプロモータが特定の転写因子と結合しているのかをゲノムレベルで調べることができる、ChIP マイクロアレイデータが用いられている。ChIP マイクロアレイデータ自体はプロモータ内のどの DNA 配列に転写因子が結合するかに関する情報を直接与えるわけではないが、本論文の手法を用いると、プロモータ・セットに対して統計的有意に高頻出する DNA 配列を探索することによって、転写因子が結合する短い DNA 配列を推測することができる。

さらに上記(2)の問題点を克服するため、与えられた遺伝子セットに対して、従来法のような種類のモチーフ探索ではなく、モチーフ組み合わせの探索を行っている。すなわち、まず、ChIP データを利用して得られたモチーフのあらゆる組合せに対し、統計的有意に現れる共頻出モチーフを選別する。次に、これら共頻出モチーフを持つ下流遺伝子の発現パターンを DNA マイクロアレイデータから得て、発現パターンがコヒーレントであるモチーフ組み合わせだけを選別することにより、転写因子と結合モチーフによる組み合わせ的転写制御に関する知見を得ることができる。

第3章結果の部分では、まず、ChIP マイクロアレイデータを利用したモチーフ抽出の結果が述べられている。ここでは従来知られていたモチーフに加え、多くの新しい転写因子結合モチーフが予測されている。例えば、Mth1 転写因子に対する CAGCAG モチーフ、Ash1 転写因子に対する GCGGCA 等である。これらのうち Mcm1 転写因子に対する ATAATTA に関しては、予測と合う実験結果が後に報告され、本方法の有効性が実証されている。

結果の後半部分では、酵母細胞周期の各相(Phase)に特異的な転写因子と結合モチーフの組み合わせが予測されている。ここでは、従来知られていた組み合わせに加え、多くの新しい組み合わせが予測されている。例えば G1 相における Ste12、Swi6、Swi4 転写因子と、TGAAAC、CGCGTC モチーフの組み合わせ等である。この様に転写因子と結合モチーフの組み合わせを網羅的に予測することは、従来の方法ではなし得なかったことである。

第4章考察においては、上記の結果に基づき、酵母細胞周期の組み合わせ的転写制御において見出される、幾つかの特徴が論じられている。これらの特徴のうち、接合相組み合わせ(joint-phase combination)という特徴は、一つの相で主要な役割を果たす転写因子と、次の相で主要な役割を果たす転写因子が、その中間の相で組合せを取っているものであり、一細胞周期に渡って三つの組合せが存在し、全体として環を形成している。本研究の予測結果が今まで欠けていた組み合わせ部分を埋めたために、このような環の存在が初めて明らかになったのである。

以上のように、本研究によって初めて、転写因子とその結合モチーフの組み合わせをゲノムレベルで推測する方法が確立された。また、数多くの新しいモチーフや、転写因子と結合モチーフの組み合わせが明らかになった。これらの成果は、生物物理学の一分野、すなわち、計算機的手法による全遺伝子規模の生命現象解析の分野において、多大の寄与をなすものである。

なお、本論文はミカエル・ザン(Michael Zhang)氏との共同研究であるが、論文提出者が主体となって研究を行ったものであり、審査員一同は同提出者が博士(理学)の学位を授与するのに十分であると判断した。