

氏 名 森 明弘

学位（専攻分野） 博士（理学）

学位記番号 総研大 1292 号

学位授与の日付 平成 21 年 9 月 30 日

学位授与の要件 生命科学研究科 遺伝学専攻
学位規則第 6 条第 1 項該当

学位論文題目 Computational and experimental studies on
cis-regulatory modules in *C.elegans* embryogenesis

論文審査委員 主 査 教授 大久保 公策
教授 荒木 弘之
教授 嶋本 伸雄
准教授 木村 暁
チームリーダー 杉本 亜砂子
(理化学研究所)

Adequate regulations of transcription are required for production of a wide array of cell types in development. These regulations are often triggered by binding of proteins (transcription factor or TF) to specific regions (*cis*-regulatory modules or CRM) of genes. Thus, it is important to reveal combinations of TF-CRM for understanding gene regulatory network, and I considered a crucial factor for understanding the network is the temporal/stage-specific transcriptional mechanisms regulated by CRMs. To this end, I developed a computer algorithm to predict CRMs candidates. Then, I verified them experimentally and extended the analysis to reveal gene regulatory mechanisms in *C.elegans* intestine cell lineage. *C. elegans* is one of the best animals to investigate gene regulatory network because of abundant information of gene expressions for prediction and availability of experimental methods for verification. Taking with the advantages, I address to elucidate that there exist independent mechanisms for stage-specific initiation of gene expression besides cell-specific transcriptional regulation. I think the combinatory regulations of the two regulatory mechanisms are essential for proper gene expressions.

First, I describe the development of a two-step computer algorithm named FEsystem (Filtering-Esteeming system) to identify CRMs. This system extracts short-sequence motifs specific to positive dataset (PD) (in this case, upstream region of genes that show the same expression patterns) by comparing with negative datasets (ND) (upstream regions of other genes). Previous *in silico* methods for CRM predictions often faced two problems known as scalability and generality problems. The scalability problem is that it becomes much harder to extract short sequence motifs when the search region becomes 1000bp or longer. The generality problem is that most of CRM discovery programs require parameter optimization such as window size of target CRMs for each CRM. Thus, those CRM discovery programs do not work efficiently on other datasets with a set of fixed parameters, and both of those obstacles caused difficulty to identify *ab initio* / *de novo* CRMs. I developed FEsystem to overcome these problems. I confirmed the performance of FEsystem prediction by the benchmark tests using the datasets of six known CRMs in *C. elegans* which had been verified experimentally. FEsystem identified all the CRMs from the datasets successfully. Then, I applied FEsystem to the sets of genes which show the same *in situ* hybridization patterns in NEXTDB (Nematode Expression Pattern Database). As a result, I identified many CRM candidates for tissue-/stage-specific transcriptional regulation in *C. elegans* embryogenesis.

Next, I show that I verified the functionality of the CRMs candidates by the GFP reporter assay. Here, I focused on three consecutive stages of the intestine E-lineage; E2, E4, and E8 (two, four, and eight E-cells) stages, to verify the CRM candidates *in vivo*. E-lineage forms only intestine and intestine is formed only by E-lineage, and it have been shown that many genes were activated by multiple GATA factors

stage-specifically in the E-lineage. I examined various deletion/substitution/insertion constructs for both the CRM candidates and GATA sites. These results showed that (i) there exist distinct CRMs at each stage, (ii) the candidate CRMs that are located near GATA sites may be required for regulating the initiation timing of proper gene expressions, and (iii) the initiation timing can be controlled by replacing the stage-specific CRMs. As a temporal transcriptional regulation, the affinity model is known in pharynx development of *C. elegans* embryogenesis. In the model, the initiation timing of gene expressions is determined by the binding affinity of TF and the numbers of transcriptional factor binding sites (TFBS). However, many lines of evidence have shown that the affinity model may be conflict with the transcriptional regulatory system in the intestinal lineage. For example, even though 5'-upstream region of genes contain multiple GATA sites, those genes are not activated at any stages in the lineage. Thus, I think my results support a temporal/stage-specific transcriptional regulation mechanism other than the affinity model.

Furthermore, I investigated TFs for the identified E-lineage CRMs to elucidate a mechanism of temporal transcriptional regulation. For TFBS prediction, I again applied FESystem to datasets of yeast-one hybrid, which contained more than 230 TFs expressed in digest trucks such as pharynx and intestine (Deplancke *et al.*, 2006). I confirmed approximately 70% of known TFBSs were predicted by this approach. As a result of comparing predicted TFBS with the E-lineage CRMs, I found DAF-12 is a good candidate as a temporal regulator at E8 stage. I evaluated the GFP expression patterns of E8-stage genes in *daf12* (m20) null mutant and found that the patterns were similar to those of the deletion/substitution constructs of E8-stage CRMs. Consistently with the results of previous studies, RNAi assay on GATA factors in the lineage showed that ELT-2 GATA factor also required for E8-gene expressions. These results strongly suggest that DAF-12 directly regulates the initiation timing of gene expressions with the GATA factor at E8 stage in intestinal lineage.

In this doctoral thesis, I demonstrated a strategy for elucidating a transcriptional mechanism of tissue-/stage-specific TF-CRM combination, which will lead to designing promoter regions of genes in the future.

博士論文の審査結果の要旨

申請者は、分化における時期特異的転写の機構の解明を目標として、同じ発現パターンを示す遺伝子には共通の転写因子が結合する同じモチーフ配列が存在するというモデルに基づいて

1. 同じ時期特異的発現を示す遺伝子の上流に存在するデジェネラシーの強いモチーフを検出するプログラムを開発した。
2. *C. elegans* 消化管形成期に器官特異的に発現する遺伝子を発現開始時期別に3セット選び、それぞれから当該発現の責任モチーフ候補を上記プログラムで抽出した。
3. GFPをレポーターとする実験系を用い、モチーフの欠損や塩基置換等の改変を対照に、プログラムから予言された候補の中に正解が含まれることを証明した。
4. さらに同じ実験系を用いて消化管細胞で早期発現と晚期発現をする遺伝子の間でそれぞれの発現に責任のあるモチーフを交換すると、早期と晚期が入れ替わることを示し、上記モチーフが組織特異性のみならず発現時期を直接コントロールしている可能性を補強した。
5. 発見したモチーフのうちE8発現開始に関わるひとつについて、同モチーフに結合する転写因子を推定した。モチーフの候補を求めたプログラムを、既存の酵母1ハイブリッド法による線虫転写因子とそれに結合するDNA断片とのペアをまとめたデータベースより、最も類似するものを選び、daf-12結合遺伝子群を得た。
6. daf-12たんぱく質がE8モチーフ結合因子であることを示す目的で、daf-12を欠損する個体にE8遺伝子群のプロモーターの活性を、GFPをレポーターとして測定したところ、daf-12たんぱく質をもつ個体における、E8責任モチーフを欠損したプロモーターと同程度に低下しており、求める転写因子はDaf-12であることが示された。

まとめると

1. 出願者が同定したいくつかのモチーフはすべて完全に新規なモチーフであり、モチーフ入れ替え実験は説得力のある証拠を与えている。
2. 同じ細胞系譜での遺伝子発現の順序が「時期を定めた独立のモチーフ」によって決定されていることが強く示唆された。対立する「同じモチーフの数の違いによる同じ転写因子に対する感度の差」であるとする仮説の存在との対比を考慮すると、遺伝子発現の順序を決定するメカニズムの解明に貢献したと言えるであろう。

研究成果の発表は水準に達しており、配列解析からデータベース、遺伝子導入実験、線虫実験分野の現状などの標準以上に広範な背景知識も備えている。加えて申請者は所属研究室のアドバンテージを生かしながら、独自の興味に基づいて「研究生活」を積極的に楽しんでいることがその表現からよく伝わり、強い研究動機が申請者の特筆すべき特徴であると断定できた。論文内容や発表内容において、バイオインフォマティクスや分子細胞生物学の独自の思考が多く認められた。

出願者が開発したプログラムは、複雑で十分説明出来ないパラメータ設定を伴うために、申請者が主張するようには既存のプログラムとの能力比較はできず、この点を、委員や委員会は長大な時間をかけたが解決できず将来の課題として残さざるを得なかったが、

少なくとも、現課題において機能していることは明白である。自らの興味と疑問に基づいて自ら調べ自ら試すという研究者として欠かせない特徴に加え、プログラミングから遺伝子導入実験までの広範囲な基本実験操作を習得しており、博士号授与の基準を満たしていると考えられる。

一方で自らのみで発展させる一連の研究スタイルは不完全で散漫になりやすいので、小さくとも確実なステップをつくり科学コミュニティに貢献することも意識するように、委員会は強く推薦した。具体的には不足実験やプログラムの整理を行い長大な博士論文を出版論文に変えてゆく過程に注力することで上記課題が達成されるであろう。

最後に、計算機とベンチの両方を駆使できる研究者が強く求められていた状況下で期待にこたえるべく双方に習熟せんとした出願者の高い動機に敬意を払う。

論文および発表、質疑はいずれも英語で行われたが、英語能力はコミュニケーションを含む研究遂行上必要な水準を満たしていると判定した。