

# 論文審査の結果の要旨

氏名 傅 思縉

本論文は4章からなり、第1章はイントロダクション、第2章は研究に用いた手法について、第3章は研究の結果と考察、第4章は研究の結論が述べられている。真核生物に於ける細胞核は遺伝情報を格納する小器官であり、細胞の指令塔とも言える。その為、蛋白質の核への出入りは厳密に管理されており、核内移行シグナルや核外移行シグナルという「荷札」を持つものだけが運ばれることになっている。CRM1/Exportin 1は核内から核外への輸送を担う蛋白質であり、核内でロイシンに富んだ核外移行シグナル(leucine-rich Nuclear Export Signal, NES)を持つ「荷物」の蛋白質を認識し、核外へと運ぶ。

第1章では、NESに関するこれまでの研究の概要とその問題点が述べられている。

NESは重要な機能部位である。実験的に確かめられているものだけでも、NESを含む蛋白質は100個以上にのぼる。これらの蛋白質は信号伝達や細胞周期といった過程において重要な役割を担っている。また、人間の健康に直接関わるものも多く、HIV-1やインフルエンザといったウイルス蛋白質にNESが見られるほか、癌抑制遺伝子p53の核排出にも関与している。

NESを効率良く同定する為には、計算手法によるNES予測が不可欠である。アミノ酸配列上のどの部位がNESであるという仮説があれば、変異解析実験によりその確認はできる。が、こういった実験は手間がかかる為、数千残基にもおよぶ蛋白質のすべての候補部位を網羅的に調べることは非現実的である。しかし、実験を伴わない計算手法で候補部位が数ヶ所に絞れば、様々な蛋白質のNES部位決定が効率良く進むと期待できる。

残念ながら、NES部位を正確に予測する計算手法は未だ確立されていない。実際は、NESの特徴をある程度捉えたコンセンサス配列を用いることが多い。例えば、NES部位の7割は[LIVFM]-x(2,3)-[LIFVM]-x(2,3)-[LIVFM]-x-[LIVFM] ('x(2,3)'は、2塩基または3塩基の任意配列)に一致する。しかし、こういったコンセンサスは蛋白質の疎水性コアなど、NESとは関係のない部位にも一致することが多く、NESの候補部位を絞る効果は不十分である。

この状況を改善する為、2004年にla Courらは、隠れマルコフモデルとニューラル・ネットワークという機械学習技術に基づいたNES予測法、「NetNES」を開発した。しかし、NetNESは当時知られていた大部分のNES部位を学習に使った為、新規NES部位に対する予測精度(いわゆる汎化能力)は未知数であった。

コンセンサス配列に十分捉えられないNES部位の特徴として、決まった構造を取らない領域(ディスオーダー領域)の役割が挙げられる。疎水性のある残基を多く含むNESがCRM1の接触できる分子表面に現れる為、周辺領域の主鎖に柔軟性が必要であると考えられる。また、多くの立体配座が取れる部位ならば、CRM1と結合できる形のものを含む可能性は高くなるとも考えられる。しかし、ディスオーダー領域とNES部位の関係を量的に調べた研究はなく、こういった指摘は推測に過ぎなかったとも言える。

そこで本研究では、1) 新規NESデータの収集、2) ディスオーダー領域とNESの解析、3) NES予測法(NESsential)の開発、4) NESsentialとNetNESの予測精度比較、を行った。第2章では、これらの手法が説明されている。

第3章では、研究の具体的な結果と考察が述べられている。

新規NES部位のデータセットを構築する為、文献調査を行い、2004年以降に実験的に同定された85個のNES部位を収集した。データの選定に採用した確認実験条件と配列の重複度の詳細は3.1節に述べられている。

3.2節では2種類のディスオーダー領域予測法(POODLE-LとDISOPRED)を用い、NES部位とコンセンサス配列に一致する非NES部位(偽陽性)の予測されるディスオーダー傾向(ディスオーダー・スコア)の分布に顕著な差異があることを示した。特に、NES部位のディスオーダー・スコアは正規分布や一様分布ではなく、最小値に近いスコアと、比較的の高いスコアのいずれかとなる部位が多かった。それに対して、偽陽性のディスオーダー・スコアは低い値に集中していることを突き止めた。

本研究で開発したNES予測法「NESsential」は、NES部位を正例、[LIFVM]-x(2,3)-[LIVFM]-x-[LIVFM]のコンセンサスに一致する非NES部位を負例とした、2分類問題として捉えた。部位の特徴量として、疎水性、負電荷などといったアミノ酸の物性の他、予測されるディスオーダー傾向、二次構造と露出度(surface accessibility)といった、意味付けられる特徴量にした。特徴量の分類への寄与とディスオーダーとの関係は3.5節に詳細に検討されている。分類器として、標準的なSVM(Support Vector Machine)であるLIBSVMを採用した。

3.4節ではNESsentialと従来法のNetNESの予測比較について述べている。この比較は新規NESを含む蛋白質候補のスクリーニング(ふるい分け)問題を想定して行った。NESを含まない(未発見のNESがある可能性を除けば)、核または細胞質に局在する541個の酵母蛋白質と、NESを含む蛋白質60個の順位を付けた。各予測法の上位候補を調べた結果、NESsentialの精度はNetNESを大きく上回った。例えば、上位10個をみると、NetNESのリストにNESを含む蛋白質は1個もなかったが、NESsentialでは10個中6個のNESが含まれていた。この結果から見て、本研究は当初の研究目標を達成したと言える。

第4章では、上記の研究から導かれる結論について述べられている。まとめると、傅思縉は、新規NESのデータセットを収集し、NESとディスオーダーの関係を量的に示した上に、NESsentialという新しい予測法を開発した。さらに、スクリーニング問題においてNESsentialが従来法より遥かに有用であることを示した。

なお、本論文は、今井賢一郎、Paul Hortonとの共同研究であるが、論文提出者が主体となって開発、分析及び検証を行ったもので、論文提出者の寄与が十分であると判断する。

したがって、博士(科学)の学位を授与できると認める。