総説

天然変性, ハブ性, 細胞内局在を整理する

太田元規¹, 福地佐斗志², 小池亮太郎¹

¹名古屋大学大学院情報科学研究科

²前橋工科大学工学部

Protein-protein interactions are fundamental for all biological phenomena. The hub proteins interacting with a number of partner proteins play the vital role in the protein-protein interaction network. We investigated the subcellular localization of proteins in the network, and found that the proteins localized in the multiple subcellular compartments, especially the nucleus and cytoplasm, tend to be hub proteins. Examination on keywords associated with the proteins suggested that those related to post-translational modifications (PTMs) and transcriptions contributed to numerous interactions. Triggered by PTMs in the intrinsically disordered regions, they change interaction partners in the protein complex, and are translocated from cytoplasm to nucleus.

Intrinsic disorder / hub proteins / subcellular localization / nucleus and cytoplasm proteins / post-translational modification / transcription

1. はじめに

今世紀に入るとプロテオームワイドにタンパク質の相互作用情報が得られるようになり, 適当にデータをダウンロードしてツールに放り込むと, 非常に複雑なネットワーク図を手軽に描くことができるようになった. こういう図を見た時に着目すべきは, 多くのタンパク質と相互作用しているタンパク質で, それらはハブタンパク質と言われている. ハブタンパク質は相互作用ネットワークの要に位置するわけだから, それらがおかしくなってしまうと重篤な疾患を引き起こす. つまり, 紛れもなく機能的に重要なタンパク質である¹⁾. 一方, 多くの生物種でゲノム配列が明らかになるとプロテオーム単位でタンパク質の構造予測が可能となり, その結果, 真核生物ゲノムは相当量の**天然変性タンパク質**(生理的条件下にあっても立体構造をとらない領域(天然変性領域)を持つタンパク質:用語解説参照)をコードしていることが明らかとなった²⁾. 特に核に局在するタンパク質には天然変性タンパク質が多いことが判明し²⁾, 真核生物が行う転写, スプライシング, シグナル伝達やエピゲノム関連の細胞機能を実現するには天然変性タンパク質が不可欠であると認識されるようになった^{2), 3)}. 更に, 相互作用ネットワークのハブタンパク質を調べてみると, それらは天

然変性であることが多いことも知られるようになってきた⁴⁾. つまりこれまでに得られた知見を統合すると, ハブタンパク質は天然変性の傾向があり, 天然変性タンパク質は核に局在しがちとなる. この言説を非常に単純に解釈し, ハブタンパク質, 天然変性タンパク質, 核内タンパク質はほぼ同じだと捉えると, 三段論法的にハブタンパク質は核内タンパク質, となる. だが, これは本当だろうか. タンパク質の細胞内局在, つまりそのタンパク質が働く場所(オルガネラ)は, タンパク質の分子機能を考える上で非常に重要な情報である⁵⁾. それが細胞質であればハウスキーピングかな, 酵素かな, と推察するし, 核であれば転写因子かな, と推察する⁶⁾. 当然ながら, ハブタンパク質についても細胞内局在は調査されているのだが, 核内タンパクの傾向があるという報告はついぞ聞いたことがない^{4), 7)}. だから, 天然変性, ハブタンパク質, 細胞内局在についての先の言説は, 大事な但し書きが欠落した言葉足らずの文言のように感じられる. 何かがしっくりこないのだ. この総説では天然変性を示すハブタンパク質を特徴付けるために私たちが行った研究を概説する⁸⁾. 研究の結果, 一群のタンパク質の構造, 機能, 機能発現機構が浮き彫りとなった.

Relationships between Subcellular Localization, Hub Proteins and Intrinsic Disorder

Motonori OTA¹, Satoshi FUKUCHI² and Ryotaro KOIKE¹

¹Graduate School of Information Science, Nagoya University

²Faculty of Engineering, Maebashi Institute of Technology

2. 多重局在性を露わに取り扱う

私たちはまず, 細胞内局在とハブタンパク質の関係から調べることにした. その過程で, 多重局在性をどう扱うかで研究方針が分岐することに気がついた. SwissProt⁹⁾ などのタンパク質のデータベースを見ると細胞内局在の情報を得ることができる. 例えば, 図 1a のような記述であれば, このタンパク質は「核のみ」にしか局在しないので, これは誰がどう解釈しても核内タンパク質としか考えようがない. では, 図 1b のようであれば, どうすれば良いか. 図 1b のタンパク質は核と細胞質に多重 (二重) 局在するタンパク質である. これは「核にも細胞質にも」局在する特別なタンパク質と考えるべきか, それとも, 核内タンパク質でもあるが細胞質タンパク質でもあると, 二重登録処理をすべきなのか. SwissProt では, 局在情報のあるヒトタンパク質のおよそ 1/3 は多重局在するものなので⁸⁾, 多重局在性をタンパク質のグループ分けにどう反映させるのかという問題は看過できない¹⁰⁾. ハブタンパク質の細胞内局在を調べた先行研究を見ると, 後者のように重複を許してグループ分けをしている

図 1 タンパク質の多重局在性を考慮した分類. a) 核のみに局在するタンパク質の SwissProt での記述. b) 核と細胞質のみに局在するタンパク質の記述. c) 少なくとも核に局在するタンパク質の分割. 括弧内の数字は解析に利用したヒトタンパク質の数. d) タンパク質の細胞内局在別にとった相互作用数の平均.

るものがほとんどで, 核内タンパク質とは「少なくとも核に」局在するものと考えられていた^{4),7)}. 図 1c に, 「少なくとも核に」局在するタンパク質を更に分類した結果を示す. それらは, 「核のみ」にしか局在しないものと「核にも他の場所にも」多重に局在できるものに分解でき, 後者は更に「核と細胞質のみ」に局在するものや「核と染色体のみ」に局在するものなどに細かく分解される. 私たちは図 1c の右端のように多重局在性を露わに扱い, タンパク質を排他的に細かく分類して相互作用数の統計を取ることにした⁸⁾. 図 1d に結果を示す. 先行研究にならい, 「少なくとも核に」局在するものや「少なくとも細胞質に」局在するものを集めて HPRD データベース¹¹⁾ を参照して相互作用数の平均を求めると, 7 から 8 程度であり差がみられない. 一方, 排他的な分類に従い「核のみ」(赤)や「細胞質のみ」(青)に局在するタンパク質について統計をとると, 相互作用数の平均は 4 程度とかなり少なくなってしまう. これらに対し「核と細胞質のみ」(緑)という二重局在性を持つタンパク質の相互作用数の平均は 10 程度で, 圧倒的に多くのタンパク質と相互作用していることがわかった. つまり, 細胞内局在の観点では, 「核と細胞質」の両方に局在するタンパク質が強いハブ性を有するグループであり⁸⁾, 先行研究ではこれらが「核のみ」や「細胞質のみ」に局在するタンパク質に混じって統計をとられていたために, その特性が隠蔽されていたのだ (この他統計的には, 「細胞質と細胞膜のみ」, 「核, 細胞質, 細胞膜のみ」というグループのタンパク質も強いハブ性を持つ)⁸⁾.

3. 核と細胞質に局在するハブタンパク質の機能と構造

核と細胞質に局在するハブタンパク質はどういう分子機能を有しているのか, そしてその機能とはハブタンパク質であることとどういう関係性があるのか. これを調べるために核と細胞質のみに局在するタンパク質に付加されている SwissProt のキーワードを統計的に処理し, どういう性質がハブ性と強く関係しているのかを解析した. その結果, 翻訳後修飾と転写の両方に関係しているタンパク質 (具体的には, (“phosphoprotein” or “acetylation” or “Ubl conjugation”) and (“transcription” or “DNA binding” or “activation” or “repressor”) というキーワードを持つもの) が, ハブ性の源となっていると結論づけることができた⁸⁾. これらは核と細胞質のみに局在するタンパク質の 3 割弱を占めてお

り, 相互作用数の平均は19である. 次にこれらについて天然変性の観点から構造を調べた. 図2aに各細胞内局在のグループごとに計算した, 天然変性領域の割合の累積密度分布 (DICHOT¹²) による予測値) を示す. これまでに言われている傾向通り, 天然変性領域は核のみに局在するタンパク質では多いが, 細胞質のみに局在するタンパク質では少ない. 核と細胞質のみに局在するタンパク質はその中間くらいの天然変性具合である. しかし, このうち翻訳後修飾と転写の両方に関与するものは, 核のみに局在するタンパク質に匹敵するほどの天然変性を示した⁸⁾. つまり, 天然変性となるハブタンパク質は, 「核と細胞質を行き来し, 翻訳後修飾を受けて転写に関係するもの」だったのだ. これらがいろいろな観点で仕分けられたグループに顔を出した結果, 冒頭で述べたようなハブタンパク質や細胞内局在に関する言説が生まれたのだろう. これら一連の解析により, 抜け落ちていた但し書きが明示的になり, 気分はすっきりした. 「核と細胞質への局在」「翻訳後修飾と転写」の観点で

図2 細胞内局在と天然変性. a) 天然変性領域の全長に対する割合の累積密度分布. 立ち上がり早い曲線は天然変性領域が少ないタンパク質が多いこと, 遅い曲線は天然変性領域を多く含むタンパク質が多いことを表している. 各線の示すグループは図中に記した. b) ハブタンパク質, 天然変性タンパク質, 核タンパク質の関係性を表したベン図. 赤, 緑, 緑点線の囲みの位置は, そのタンパク質を一言で言うところのどこに対応するのかを示している.

加えることで, 天然変性, ハブタンパク質, 細胞内局在の関連性がきれいに整理された (図2b). より詳細な分子機能と作用機序の関連については次節で掘り下げる.

4. 天然変性ハブによる細胞内情報処理

「核と細胞質を行き来し, 翻訳後修飾を受けて転写に関係する天然変性ハブタンパク質」(これを仮に天然変性ハブと言う) を具体的に調べてみると Smad¹³⁾ や NF- κ B¹⁴⁾ などが該当する⁸⁾. これらはいわゆる細胞内シグナリング経路で働く転写因子であり, その役割を大雑把に述べると, 細胞外からの刺激を受けてそれを核内に伝達し, 最終的には転写を制御する, というものだ. その過程でリン酸化やユビキチン化などの翻訳後修飾を受ける. キナーゼやユビキチンリガーがターゲットとする翻訳後修飾部位は, それらの酵素が容易に接近できるよう, 天然変性領域に多く存在することが統計的に知られている¹⁵⁾. 分子機能を実現する機構として翻訳後修飾を積極的に利用することと天然変性であることには明確な因果関係がある. 翻訳後修飾はタンパク質の物性を変えることで, 相互作用パートナーとの親和性を制御し, 複合体構造の組換えを誘起する. その結果, 多くのタンパク質との相互作用を実現する. これらの細胞内シグナリング経路では, 過渡的複合体構造が動的に遷移し, その遷移の連鎖が手順通りに的確に起きることで, 細胞外からの刺激に対する情報処理が行われている. また, 核と細胞質の行き来を制御する場合にも, 翻訳後修飾と複合体の組換えが利用されている. インボーチンが認識する核移行シグナルは天然変性領域にある¹⁶⁾. 図3に天然変性ハブが行う細胞内情報処理を一般化して表した. Smadの系でも NF- κ Bの系でも図3に描かれている過渡的複合体に相当するものは, 立体構造情報が圧倒的に不足している. ここに構造生物学の「新大陸」があると予見するのは, 私たちだけではないだろう.

実際の細胞内には多くのタンパク質が高密度で存在しているため, 複合体の組換えを行うのは簡単ではなさそうだ. 立体構造のある構造ドメインからなるタンパク質が相互作用の相手を交換する場合には, 相互作用面の位置や配向からくる制約のために, 相手を正確な場所に配置する必要があるだろう. これを混み合い環境下で実行するためには相手に見合う空隙を特定の場所に用意しないといけないので, エントロピー的に大きなコストがかかるはずだ. それに対し天然変性で

図 3

天然変性ハブが行う細胞内情報処理の概念図。翻訳後修飾を契機に、複合体の組換えを行いながら多くのタンパク質と相互作用し、転写を調節する。

あれば相互作用部位は空間的に可動だから、適当な空隙は必要だがその位置や方向性には自由度がある。よって構造ドメインからなるタンパク質に比べエントロピー的には相当優位性があるはずだ。天然変性タンパク質の研究をしていると、しばしば「どうして生命現象に天然変性タンパク質が必要なのか」と問われることがある。これに対し私たちは「真核生物が行う高度な細胞内情報処理の本質が翻訳後修飾を利用した過渡的複合体の組換えにあるのだとすれば、それは天然変性でなければなし得ないのだ」と応じることができるのではと考えている。

5. おわりに

この研究では HPRD と SwissProt の記述を信頼し、統計的な解析を行った⁸⁾。しかしタンパク質の相互作用、細胞内局在や分子機能については全てがわかっているわけではない。これらは不完全なデータであり、これから追加や訂正がされていくだろう。また歴史的な理由から研究が進んでいるタンパク質もあり、それらの情報は他に比べて相対的に多いと考えるのが自然

だ。そう考えると多重局在性とハブタンパク質の相関は科学的真実ではなく、人間の研究意欲を反映した情報量の多寡によるアーティファクトということになる。そうではない、ということを実証する方法がないものかと自問自答したのだが妙案はなく、最終的にはデータベースを変えて同様のことを行い、結果の安定性を主張した。バイオインフォマティクスの研究をしていると似たような局面に何度も遭遇する。不完全で、間違いが混入していて、歴史のバイアスから逃れられないデータから、科学的真実を導くことが求められる。これを成し遂げる有効な一般論があると、本当にありがたい。

方法論とは少し異なるが、決め手がない時に私たちが依る検証法がある。それは結果をサポートする理由付けがしっかりできるかどうかだ。多重局在性とハブ性の相関は極めてリーズナブルで(活動場所が増えればつきあう相手も増える)、多重局在性や場所の移動性はタンパク質以外の「ハブ」についても当てはまる重要で一般的な概念だと思う。また、天然変性と細胞内情報処理との関係も、全て妥当に説明することができている。だから私たちはここで述べたことはこの先データが変わっても、定性的には真実を捉えていると信じることができた。逆に言うと、統計はとても重宝で多様かつ大量のデータから多くの法則を導くことができるが、説明と解釈が伴わない言説は幻なのかもしれないのだ。

文 献

- 1) Han, D. J. *et al.* (2004) *Nature* **430**, 88-93. DOI: 10.1038/nature02555.
- 2) Ward, J. J. *et al.* (2004) *J. Mol. Biol.* **337**, 635-645. DOI: 10.1016/j.jmb.2004.02.002.
- 3) Iakoucheva, L. M. *et al.* (2002) *J. Mol. Biol.* **323**, 573-584. DOI: 10.1016/S0022-2836(02)00969-5.
- 4) Haynes, C. *et al.* (2006) *PLoS Comp. Biol.* **8**, e100. DOI: 10.1371/journal.pcbi.0020100.
- 5) Alberts, B. *et al.*, eds. (2014) *Molecular Biology of the Cell*, Garland Science.
- 6) Yang, L. *et al.* (2014) *J. Theor. Biol.* **358**, 61-73. DOI: 10.1016/j.jtbi.2014.05.008.
- 7) Patil, A. *et al.* (2010) *Protein Sci.* **19**, 1461-1468. DOI: 10.1002/pro.425.
- 8) Ota, M. *et al.* (2016) *PLoS ONE* **11**, e0156455. DOI: 10.1371/journal.pone.0156455.
- 9) UniProt-Consortium (2011) *Nucleic Acids Res.* **39**, D214-219. DOI: 10.1093/nar/gkq1020.
- 10) Zhang, S. *et al.* (2008) *BMC Bioinf.* **9**, 127. DOI: 10.1186/1471-2105-9-127.
- 11) Keshava Prasad, T. S. *et al.* (2009) *Nucleic Acids Res.* **37**, D767-772. DOI: 10.1093/nar/gkn892.

- 12) Fukuchi, S. *et al.* (2011) BMC Struct. Biol. **11**, 29. DOI: 10.1186/1472-6807-11-29.
- 13) Qin, B. Y. *et al.* (2002) Genes Dev. **16**, 1950-1963. DOI: 10.1101/gad.1002002.
- 14) Huang, D. B. *et al.* (1997) Structure **5**, 1427-1436. DOI: 10.1016/S0969-2126(97)00293-1.
- 15) Collins, M. O. *et al.* (2008) Mol. Cell. Proteomics **7**, 1331-1348. DOI: 10.1074/mcp.M700564-MCP200.
- 16) Lee, B. J. *et al.* (2006) Cell **126**, 543-558. DOI 10.1016/j.cell.2006.05.049.

太田元規

太田元規 (おおた もとのり)

名古屋大学大学院情報科学研究科教授
1996 年早稲田大学大学院理工学研究科修了。
2008 年より現職。
研究内容: 生命情報学
連絡先: 〒464-8601 愛知県名古屋市千種区不老町
E-mail: mota@is.nagoya-u.ac.jp
URL: <http://www.force.cs.is.nagoya-u.ac.jp>

福地佐斗志

福地佐斗志 (ふくち さとし)

前橋工科大学工学部准教授
1995 年東京理科大学大学院理工学研究科修了。
2011 年より現職。
研究内容: タンパク質情報学
連絡先: 〒371-0816 群馬県前橋市上佐鳥町 460 番地 1
E-mail: sfukuchi@maebashi-it.ac.jp
URL: <http://www.maebashi-it.ac.jp/~sfukuchi/>

小池亮太郎

小池亮太郎 (こいけ りょうたろう)

名古屋大学大学院情報科学研究科助教
2003 年京都大学大学院理学研究科修了。12 年より現職。
研究内容: 構造バイオインフォマティクス。
連絡先: 〒464-8601 愛知県名古屋市千種区不老町
E-mail: rkoike@is.nagoya-u.ac.jp