

連載講座

細胞の完全理解は可能か？—生物物理学的視点からの問い (第2回)

タンパク質に関するいろいろな数字

市川 一寿

Key words : 複雑さ, 分子数, 混雑度, 翻訳後修飾, ターンオーバー

1950 年から始まる 75 年で細胞生物学が分子細胞生物学へと進展し、それまでの悶々が払拭されたことを第 1 回に述べた。しかしそれと同時に、そして当然ながら新たな悶々が出現し、メカニズムの詳細が明らかになればなるほどその度合いは大きくなった。この最大の原因は細胞の複雑さである。75 年前は「たぶん複雑だろう」程度の漠然としたものであったが、解明が進むにつれ、タンパク質の種類の多さに直面した。しかしこれは入口にすぎず、その後、想像を超えるタンパク質動態とその複雑さが明らかになった。例えば、アミノ酸配列が同じでも翻訳後修飾で機能が変化し、複合体を形成し、細胞全体への分布も、局在も、移動もあることを知った。しかもこれらは、いずれも遺伝子による直接的制御が及ばない領域であった。細胞は「何だかわからない」から「どう理解したらいいかわからない」へと、悶々の内容が変化したのである。

対象がいかに複雑であろうとも、分類でき、一定の考え方の下に整理できれば理解の対象である。ところが現状ではそこまで到達していないし、到達できるかどうか不明である。しかしこれは進歩である。なぜなら、それまで記述できなかった複雑さの中身を、われわれはかなり具体的に記述できるようになったからである。そこで連載第 2 回では、新たな悶々からの脱出に役立つことを信じて、タンパク質に関わる数字を通して

複雑さを考える。

I 複雑さ

本題に入る前に、複雑さの概念を少し整理しておく。と言っても難しいことはなく、込み入っており、絡み合っていて、一言で「こうだ」と言えない説明困難な状況を複雑と呼ぶ。ただし、複雑さの概念には多数の要素と、要素間の相互作用の存在が含意されていることは意識する必要がある。日本語では対象を“複雑”として全体的に捉えがちだが、英語ではこれらを意識し、要素と相互作用の違いに応じて“complicated”と“complex”を使い分ける。

この使い分けは分野によって上記とは異なるようであるが、サイエンスでは同種の要素が多数存在していたり相互関係が線形的である場合に“complicated”を、異種の要素が多種存在していたり相互関係が非線形的である場合に“complex”を使う。細胞は complex かつ complicated である。この使い分けを意識することは、複雑な対象を考えるうえで重要である。

II タンパク質の分子数

1. 細胞内タンパク質の総数

1 細胞中には何分子のタンパク質があるのだろうか。ようやく 2013 年に、細胞 1 fL ($1 \mu\text{m}$ の立方体) 当たりのタンパク質総数が、 $2\text{--}4 \times 10^6$ 個と

Characteristic number of proteins

Ichikawa Kazuhisa : 株式会社 True Cell Simulations (〒107-0062 東京都港区南青山 2-2-15 WIN 青山 531)

0370-9531/04/¥500/論文/JCOPY

推定された¹⁾。これをあえて濃度換算すれば 5 mM 程度となり、非常に高い。この問題は次節で改めて考える。

その後、プロテオーム解析によるタンパク質の種類ごとの数が報告されるようになった。HeLa 細胞のミトコンドリアでは抗アポトーシスタンパク質 Bcl-2 が 1,300 個/mit, アポトーシス促進タンパク質 Bax が 220 個/mit であった²⁾。更に GFP でタグ付けした転写因子 NF- κ B (RelA) を 1 マクロファージ単位で計測し、32,400–188,000 個/細胞という大きなバラつきが報告された³⁾。Caspase3 も、98 個の Jurkat 細胞の計測から 629–12,171 個/細胞という、やはり大きなバラつきが報告された⁴⁾。こうして報告が蓄積されてメタ解析が行われ、タンパク質 1 種類当たり数個/細胞から 10^7 個/細胞に及ぶ広い範囲の値が報告された⁵⁾。このうちの 3 分の 2 は、1 種当たり 1,000–10,000 個/細胞であった^{*1)}。植物の葉肉細胞では、タンパク質総数が 2.5×10^{10} 個/細胞と推定された⁶⁾。

それにしても、なぜタンパク質の種類によって大きな差があるのだろうか？ 1 細胞当たり 10 個程度のタンパク質は細胞機能に関与できるのであろうか？ 特定のタンパク質の個数を 10 倍あるいは 10 分の 1 にできたとしたら、何がどのように変わってくるのだろうか？ このような設定は細胞シミュレーションでは簡単だが、実験で定量的に行うのは難しいであろう。

2. 細胞内タンパク質の混雑度

次にタンパク質の混雑度を考える。ここでは、自分の大きさに対して隣のタンパク質がどのくらい離れた距離にあるかで混雑度を測る。これは、混雑度についての日常的感覚に合致する。これに従えば、総タンパク質濃度が 5 mM 程度、すなわち概算で 1 fL (V) 中に 10^6 個 (N) のタンパク質が存在する場合は¹⁾、1 個に割り当てられた体積は V/N であり、タンパク質間平均距離の概算は $10 \text{ nm} [= (V/N)^{1/3}]$ である。この値は、大きさ 5 nm のタンパク質にとっては自分の大きさの 2

倍の距離に別の 1 個が存在することになり、非常に混雑した状況で、ブラウン運動 (拡散) や移動はかなり制限されるであろう^{*2)}。しかし、これは筆者が細胞シミュレーションを始めた当時、細胞内と試験管内での拡散定数の違いを調べた結果 (大きな差はなかったと記憶する) とは一致しない。悩ましい問題だが、細胞内では膜系と骨格系にタンパク質が高密度にパッキングされていれば⁷⁾、解決する可能性がある。

3. タンパク質の機能別分子数、細胞機能との関係

これは細胞理解に重要な数値である。HeLa 細胞では、10,255 種のタンパク質の約 40 種で、タンパク質全体の 25% を占めることが見いだされた⁷⁾。更に彼らは細胞コンパートメント別の解析を行い、膜系と骨格系に多いこと、機能別では転写、翻訳、フォールディングに関わるタンパク質数が多く、アポトーシスやヒストンアセチル化に関わるタンパク質は少ないことを報告した。

更に興味深くかつ重要なのは、細胞環境の変化によるタンパク質数の変化である。それによると、アポトーシス誘導刺激で Bax は 4.4 倍に増加したが Bcl-2 には変化がなかった²⁾。一方、環境ストレスによる変化は mRNA コピー数ではみられるが、タンパク質数ではみられないことが報告された⁵⁾。

4. タンパク質数と mRNA コピー数との相関

これも大変に興味深い数字である。まず Nagaraj らは、mRNA コピー数とタンパク質数の相関が低いことを報告した⁷⁾。Albayrak らも HEK の 1 細胞で CD147 タンパク質数と mRNA コピー数を同時計測して全 131 細胞の結果をまとめ、やはり相関が低いことを示した³⁾。どうやら mRNA コピー数と対応するタンパク質数の相関は低そうである。仮に単位時間当たりの翻訳量が mRNA コピー数と相関するなら、タンパク質の分解活性が数を決定することになり、しかもそれがタンパク質種ごとに異なる可能性がある。

* 1) 偶然だが、この値はシミュレーションを通して分子細胞生物学を行ってきた筆者の想定内である。

* 2) タンパク質のランダムウォークシミュレーションを行えば、混雑度の影響について何らかの見通しを得られるであろう。

化(3), メチル化(1), アセチル化・メチル化(4)が報告されている(カッコ内はサイト数)⁹⁾。

$$\text{NS}_{\text{PTM}} = 2^{10}(\text{P}) \times 2^3(\text{A}) \times 4^1(\text{M}) \times 5^4(\text{A/M}) \\ = 20,480,000$$

- ・ **p53**: がん抑制遺伝子の転写を活性化する転写因子である。4 量体として機能するが, 1 量体の p53 にはリン酸化(12)とアセチル化(11)が報告されている^{10,11)}。

$$\text{NS}_{\text{PTM}} = 2^{12}(\text{P}) \times 2^{11}(\text{A}) = 8,388,608$$

- ・ **エストロゲン受容体**: ホルモンの受容体で, リン酸化(4)が報告されている¹²⁾。

$$\text{NS}_{\text{PTM}} = 2^4(\text{A}) = 16$$

- ・ **SARS-CoV-2**: 2019 年の COVID-19 パンデミックのバリエーションで, そのスパイクタンパク質には N-グリコシル化(22), O-グリコシル化(18), パルミトイル化(11), グルタミン酸 3 サイトとアルギニン 1 サイトのメチル化が報告されており¹³⁾, 驚きの NS_{PTM} 数である。

$$\text{NS}_{\text{PTM}} = 2^{22}(\text{N-G}) \times 2^{18}(\text{O-G}) \times 2^{11}(\text{pal}) \times \\ 2^3(\text{M}) \times 3^1(\text{M}) = 54,043,195,528,445,952$$

このように, 現在わかっているだけでも, タンパク質 1 分子のとり得る状態数の多さは, 想像をはるかに超える。しかも, 例として挙げた範囲ではあるが, タンパク質によって NS_{PTM} の値が 15 桁も違う。これをどのように考えたらよいのであろうか?

2. 細胞の揺らぎの原因?

ここで深刻な問題を考える。 NS_{PTM} がそれぞれタンパク質の異なる機能に対応すると考えると, ヒストン H1, RelA, p53 の状態数 (= 機能数) は 10^7 のオーダーかそれ以上であり, これはそのタンパク質の細胞内分子数を超える(「**II タンパク質の分子数**」を参照)。ということは, アミノ酸配列が同じでも同じ機能のタンパク質は少数で, しかも時間変化する。これでは細胞の機能発現は安定しない。しかしこれは現実と異なる。なぜなら, 再現性ある機能発現を観測できるからである。この事実は, タンパク質状態数の理論的推定が現実から乖離している深刻な問題である。こ

れに対する 6 つの考え方を以下に示す。

- ①すべての PTM が等確率では生じず, 通常は限られた PTM のみ生じる。
- ②機能に無関係な PTM が存在する。
- ③ PTM サイトは独立ではなく, 相互に排他的あるいは協調的に PTM が生じる。
- ④注目の機能は 1 サイトの PTM では生じず, 複数の同時 PTM で単一機能に対応する協調性や, 異なる PTM 状態が同一機能に対応する重複性が存在する。
- ⑤タンパク質は(ということは細胞も)常に PTM 揺らぎのなかにあるが, 観測されるのはドミナントな PTM だけである。
- ⑥実際に, 細胞は PTM 揺らぎのなかにあり, それを積極的に利用している。

前述の 6 つの可能性のなかで, ②の場合は何のための PTM か, という疑問が生まれ, ⑤は観測されない PTM 揺らぎが実験結果や薬効/治療効果の揺らぎの原因となる悩ましい可能性につながる。一方, ⑥は細胞理解の新たな視点となる可能性がある。以前, 確率共鳴(stochastic resonance)が話題になった^{*4)}。タンパク質の活性化・非活性化でも, これに類することが生じる可能性がある¹⁴⁾。特に, 非線形性の強い系では起こり得る。

IV タンパク質のターンオーバー

タンパク質のターンオーバーは入れ替わりの概念であり, 既存タンパク質の分解と新規タンパク質合成の両方が含意される。ターンオーバーの計測には, 初期より放射性同位体で標識したアミノ酸が使われ¹⁵⁻¹⁷⁾, マウスのヒストンの寿命半値幅(τ_h)が脳で 54 日と 104 日, 肝臓で 18, 56, 93 日と報告された¹⁸⁾。その後 TAP タグや, SILAC ラベル法を使った解析が行われるようになった。ここでは, 分解・合成メカニズムには立ち入らない範囲で考える。

1. τ_h の分布

Belle らは出芽酵母の 3,751 種のタンパク質で τ_h を計測し, 分布が log-normal で中央値が約 43

* 4) 確率共鳴とは, 検出閾値以下の弱い信号にランダムなノイズが加わることで, 確率的に検出閾値以上の強度になり, その信号の存在が検知される現象であり, 神経細胞・神経系でその存在が知られる。

分であること、 τ_h が短いものはセリンが豊富で長いものはバリンが豊富であること、アミノ酸長と τ_h には負の相関があることを報告した¹⁹⁾。 τ_h の43分は驚くほど短い、出芽酵母の細胞周期約1.5時間と関連するのであろう。その後、約8,000種のHEK293Tのタンパク質を1細胞解像度で解析し、 τ_h が0.5時間の群と2時間の群から構成されること、アミノ酸長と τ_h には正の相関があること、短い τ_h のタンパク質はアミノ酸W, C, L, T, F, W, Y, Vが豊富で、長いものはE, D, K, N, R, Qが豊富であることが報告された²⁰⁾。アミノ酸長と τ_h の関係はBelleらとは逆の結果であるが、これが細胞機能に重要な議論かどうかには疑問がある。

2. 生物種、臓器、細胞種、機能、タンパク質種と τ_h の関係

生物種との関連については、HeLa細胞の4,106種のタンパク質とC2C12細胞の3,528種タンパク質を測定し、ヒトとマウスのオルソログではターンオーバーに高い相関があると報告された²¹⁾。これは、ターンオーバーが動物種より機能との関連が強いことを示唆する納得できる結果である。臓器との関連では、マウスの脳、肝臓、血液での τ_h が、それぞれ9.0日、3.0日、3.5日であることが報告された²²⁾。

細胞種との関連については、前述の報告に加え、哺乳類の分裂細胞(細胞周期24時間)では τ_h が0.5-35時間、非分裂細胞では約43時間、マウス(寿命約1.5年)では3-9日だが、脳細胞の核膜孔複合体(nuclear pore complex: NPC)の核膜埋込ユニットは数週間~2年、その周辺ユニットは秒のオーダーと報告された²³⁾。 τ_h は10-1,000時間以上の広い範囲に及ぶが、B細胞、NK細胞、単球では値が近く、肝細胞とマウス胚性神経細胞では大きく異なり、例外はあるもののヒストン複合体の τ_h は非常に長く、プロテアソームとNPCでは構成要素に依存すると報告された²⁴⁾。更に、3T3-L1脂肪細胞では τ_h が0-144時間で、比較的長いあるいは短い τ_h はPPAR γ パスウェイ、エネルギー代謝、ECM、ユビキチン分解、スプライシング、ゴルジ複合体、リソソームに関係していることが報告された²⁵⁾。

一方、培養海馬のタンパク質について細胞種、細胞部位、およびタンパク質機能別に解析され、全体的にはシナプスのタンパク質に比べて膜タンパク質の τ_h は短く、ミトコンドリアは長く、グリアは神経細胞より短く、グリアが神経細胞タンパク質の τ_h に影響することが報告された²⁶⁾。最近では、若い成体マウスの肝臓、骨格筋、軟骨、粘膜、血液の細胞の3,000種以上のタンパク質の τ_h を解析し、 τ_h は全体では1日以下~数百日の範囲に広がり、肝臓で最も短く(単峰性、中央値3.6日)、骨格筋で最も長い(二峰性、中央値10日)ことが報告された²⁷⁾。この報告で重要なのは、タンパク質の翻訳後修飾と τ_h との関連で、アセチル化、メチル化、リン酸化、ヒドロキシル化、カルボキシル化は、 τ_h に有意に影響した。これを補強するように、PTM解像度でタンパク質のターンオーバーが調べられ、PTMがターンオーバーを変化させることが最近報告された²⁸⁾。

以上のように、タンパク質ターンオーバーの研究には、全体→個別網羅的、臓器・動物種→細胞種・細胞局所、タンパク質バルク→タンパク質種→翻訳後修飾という流れがあり、細胞機能との関連で捉えようとする努力が積み重ねられてきたことがうかがえる。タンパク質の合成・分解は活性化・非活性化の究極的メカニズムであるが、PTMでも活性化・非活性化する。細胞機能と関連して、これがどのように使い分けられているのかは興味深い。

おわりに

PTMによる機能修飾は、環境変化に曝されている細胞が機能を変化させる手段として、有利であると考えられる。なぜなら、PTMがなければ、外界の変化に応じて遺伝子発現にまで遡るか、あるいは対応できるだけの種類と数のタンパク質をあらかじめ用意しておかなければならないからである。これらと比較すればPTMは、変化への対応の素早さとエネルギー効率の点ではるかに優れるであろう。

本稿の冒頭に、細胞の複雑さの原因として翻訳後修飾、複合体の形成、分布の多様性を挙げたが、切断(shedding)も機能修飾の一つとして挙

げなければならない。翻訳後のタンパク質はこのように多様な修飾や変化を受け、アミノ酸配列だけでは決まらない、現実に対処できる機能を生む。タンパク質はこれら多様な機能修飾のメカニズムをどのように使い分けているのか、改めて細胞の複雑さと手強さを認識せざるを得ない。セントラルドグマに比する指導的コンセプトがほしいところだが、はたしてそれは存在するのであろうか？

●文 献

- 1) Milo R : *Bioessays*. **35** : 1050-1055, 2013
- 2) Chen C, Zhang X, Zhang S et al : *Biosens Bioelectron*. **4** : 476-482, 2015
- 3) Albayrak C, Jordi CA, Zechner C et al : *Mol Cell*. **61** : 914-924, 2016
- 4) Shi M, Geng X, Wang C, Guan Y : *Anal Chem*. **91** : 11493-11496, 2019
- 5) Ho B, Baryshnikova A, Brown GW : *Cell Syst*. **6** : 192-205, 2018
- 6) Heinemann B, Künzler P, Eubel H et al : *Plant Physiol*. **185** : 385-404, 2021
- 7) Nagaraj N, Wisniewski JR, Geiger T et al : *Mol Syst Biol*. **7** : 548, 2011
- 8) Zhou T, Cheng X, He Y et al : *Front Immunol*. **13** : 981285, 2022
- 9) Huang B, Yang XD, Lamb A, Chen LF : *Cell Sigal*. **22** : 1282-1290, 2010
- 10) Reed SM, Quelle DE : *Cancers (Basel)*. **7** : 30-69, 2015
- 11) Appella E, Anderson CW : *Eur J Biochem*. **268** : 2764-2772, 2001
- 12) Dwyer AR, Truong TH, Ostrander JH, Lange CA : *J Mol Endocrinol*. **65** : T35-T48, 2020
- 13) Cheng N, Liu M, Li W et al : *Front Immunol*. **13** : 1068449, 2023
- 14) Song H, Smolen P, Av-Ron E et al : *Biophys J*. **92** : 3407-3424, 2007
- 15) Schoenheimer R, Rittenberg D : *J Biol Chem*. **127** : 285-290, 1939
- 16) Borsook H : *Physiol Rev*. **30** : 206-219, 1950
- 17) Lajtha A, Furst S, Gerstein A, Waelsch H : *J Neurochem*. **1** : 289-300, 1957
- 18) Piha RS, Cuénod M, Waelsch H : *J Biol Chem*. **241** : 2397-2404, 1966
- 19) Belle A, Tanay A, Bitincka L et al : *Proc Natl Acad Sci U S A*. **103** : 13004-13009, 2006
- 20) Yen HC, Xu Q, Chou DM et al : *Science*. **322** : 918-923, 2008
- 21) Cambridge SB, Gnad F, Nguyen C et al : *J Proteome Res*. **10** : 5275-5284, 2011
- 22) Price JC, Guan S, Burlingame A et al : *Proc Natl Acad Sci U S A*. **107** : 14508-14513, 2010
- 23) Toyama BH, Hetzer MW : *Nat Mol Cell Biol*. **14** : 55-61, 2013
- 24) Mathieson T, Franken H, Kosinski J et al : *Nat Commun*. **9** : 689, 2018
- 25) Choi S, Romero AR, Mashood F et al : *J Proteome Res*. **22** : 1723-1733, 2023
- 26) Dörrbaum AR, Kochen L, Langer JD, Schuman EM : *eLife*. **7** : e34202, 2018
- 27) Rolfs Z, Frey BL, Shi X et al : *Nat Commun*. **12** : 6778, 2021
- 28) Zecha J, Gabriel W, Spallek R et al : *Nat Commun*. **13** : 165, 2022