



フィールド・レビュー

FIELD REVIEWS

生命情報科学

中村 周吾（大学院情報学環・准教授）

1. はじめに

生物の生命活動は、非常に微小で複雑なシステムの上に成り立っている。ヒトの遺伝子の数は2万以上といわれており、多数・多種類の生体分子が互いに精緻に相互作用を及ぼし合う。そしてそれらの協同作業によって多種多様な化合物が作り出されたり分解されたりすることで生命活動が維持される。また生体分子のサイズは概ね10ナノメートル（10万分の1ミリ）程度で、1分子を原子レベルで観察することは最新技術によっても困難である。このため自分の体内で実際に起きていても、詳細な仕組みがまだわかっていないことは本当にたくさんある。

これらの解明の手掛かりとなる生命科学関連データは、1990年代に入って生物の遺伝暗号をまるごと解読してしまおうという「ゲノムプロジェクト」の進展により、爆発的に増加した。コンピュータと分子生物学双方の技術革新を背景に、ゲノム塩基配列を断片化して自動DNAシーケンサと呼ばれる装置で読み取り、コンピュータでそれらをつなぎあわせて遺伝情報を得るもので、現在までにヒトを含むおよそ70種の脊椎動物や植物、そして多数の微生物の全遺伝情報が解読され公開されている。さら

に、ある時点の細胞内での各遺伝子の発現量を一気に解析できるDNAマイクロアレイも開発された。これらの膨大なデータを解析するための情報科学的アプローチは現在の生命科学研究に不可欠なものとなっている。

遺伝子の多くは、ゲノム塩基配列に書かれた順にアミノ酸が鎖状に数百～数千個並んだタンパク質である。各タンパク質は固有のアミノ酸配列をもち、固有の立体構造に折れたたまる。それにより特定の相手分子と相互作用したり特定の化学反応を触媒したりできるようになる（図1）。つまりアミノ酸配列がタンパク質の立体構造さらには機能を規定しており、アミノ酸配列情報の中には立体構造情報がすべて含まれていることになる。

タンパク質のアミノ酸配列と立体構造の関係を図2に示す。アミノ酸には、電荷や極性をもち水と親和性の高い「親水性アミノ酸」と、逆の性質の「疎水性アミノ酸」がある。細胞質内は水やイオンが多いため、水と仲の悪い疎水性アミノ酸は内側に集まり、それを親水性アミノ酸が取り囲むことで安定になる。アミノ酸の並び順が変われば安定な構造も変わる（図

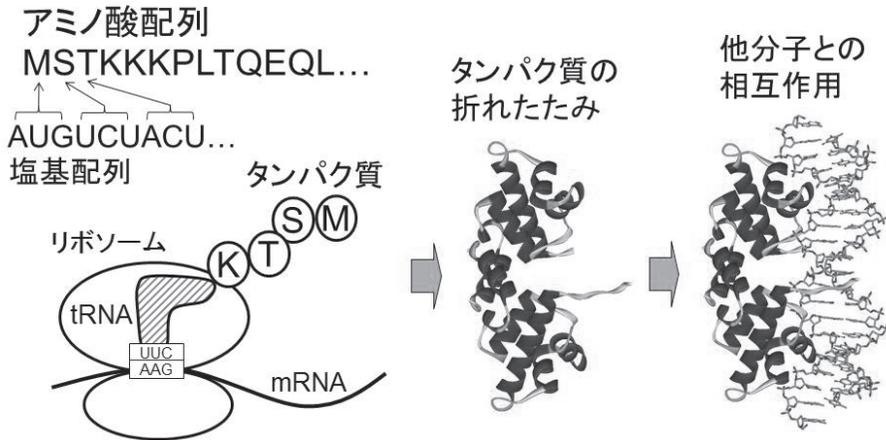


図1 : DNAが複製されてできたmRNA上の塩基配列にしたがってアミノ酸が並んでタンパク質となる。タンパク質は固有の立体構造に折れたたまり、他分子(図の場合はDNA)と結合することで機能を果たす。

2a,b)。鎖が折れたたまと、配列の上で遠いアミノ酸同士が立体構造上で近づいて特徴ある分子表面が形成され、そのタンパク質の機能が決まる。自然はこの折れたたみを「自由エネルギー最小化」という物理化学法則によっていと

も簡単にやっつけてのける。しかし自由エネルギーの精確な計算法はまだ知られておらず、アミノ酸配列情報からの立体構造情報の完全な再現は誰も成功していない。このため日々試行錯誤が続けられている。

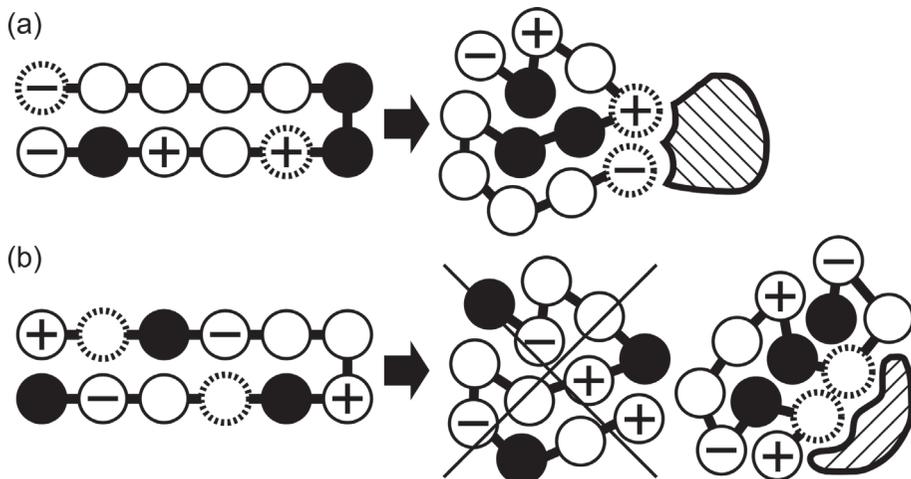


図2 : タンパク質のアミノ酸配列と立体構造の関係の模式図。黒丸が疎水性アミノ酸、白丸が親水性アミノ酸、+と-は荷電を表す。(a)と(b)のタンパク質は構成アミノ酸は同じだが並び順が異なるため、異なる立体構造に折れたたまる。これにより配列上離れたアミノ酸同士(点線)が近づいて特徴ある表面を構成し、特定の他分子(斜線)と相互作用する。

アミノ酸配列や立体構造の解析により、タンパク質の機能発現メカニズムを原子レベルで理解できれば、一部を別のアミノ酸に変えることで自然界にない反応を触媒できる変異タンパク質を創り出したり、タンパク質にうまく結合する薬剤分子を合理的に設計することで薬剤開発のコストを削減できることが期待される。リレンザやタミフルはそのような戦略で開発された薬剤の例である。

2. 機械学習法によるタンパク質の機能予測

機械学習法は、すでに答えがわかっている「学習データ」をもとにパラメータを最適化して予測器を構築する仕組みであり、さまざまな分野で幅広く応用されている。たとえばある機能をもつタンパク質ともたないタンパク質のデータを多数集め、着目する性質をもとに数値化する。これを機械学習法に与えると、その性質にしたがって両者をもっともうまく分けるパラメータが求まる。これにより、未知のタンパク質を与えたときに、そのタンパク質がその機能をもつかどうかを予測できる。

機械学習法の主な方法には、人工ニューラル

筆者は、情報学環兼任の清水謙多郎教授とともに、農学生命科学研究科・応用生命工学専攻・生物情報工学研究室において、以下に紹介するタンパク質をはじめとする生体分子の配列、立体構造、相互作用情報の解析と予測ツールの開発や、その他DNAマイクロアレイデータ解析法、大規模並列計算アルゴリズムの開発など、多岐にわたる研究を展開している。

ネットワーク、隠れマルコフモデル、サポートベクタマシンなどがある。我々はこれらを用いて、着目する性質やカーネルと呼ばれる類似度計算モデルを工夫することにより、タンパク質間相互作用部位、翻訳後修飾の有無、糖結合可能性など種々の予測のための、高精度の予測器を開発してきた¹⁻³⁾。また構築したGPIアンカー型タンパク質予測器を用いて、麹菌として知られる*Aspergillus oryzae*を含む3種の菌類の全遺伝子を検索し、同タンパク質の数や種間での相違などを見積もることに成功した⁴⁾。

3. 生体分子の立体構造・4次構造・複合体構造予測

タンパク質は微小な分子であるが、結晶化して強力なX線を当てたり溶液に強い磁場変化をかけたときの応答を解析することで、構成原子の座標すなわち立体構造を決定することができる。しかし結晶化が困難なために立体構造が決定できないタンパク質もあり、また構造決定に

は大きな手間と時間がかかる。そこでアミノ酸に潜む立体構造情報をコンピュータを用いて抽出することで立体構造を予測する方法の開発が盛んに行われている。我々は、既知構造データベースの統計解析により近似的な自由エネルギー関数を構築し、さらに局所的に配列が似た

構造既知タンパク質の情報を組み合わせるフラグメントアセンブリ法をベースにした立体構造予測ツールを開発した。また、予測構造が複数集まって複合体を形成したときの構造（4次構造）も予測できるように拡張した（図3）⁵⁾。我々の4次構造予測は、CASP9と呼ばれる立体構造予測コンテストにおいて、もっとも精度

が高いという評価を受けた⁶⁾。

さらに立体構造を入力とした薬剤など低分子リガンドの結合部位予測、タンパク質構造のペアを与えたときの複合体構造の高速予測など、さまざまな相互作用関連の予測ツールを開発している^{7,8)}。

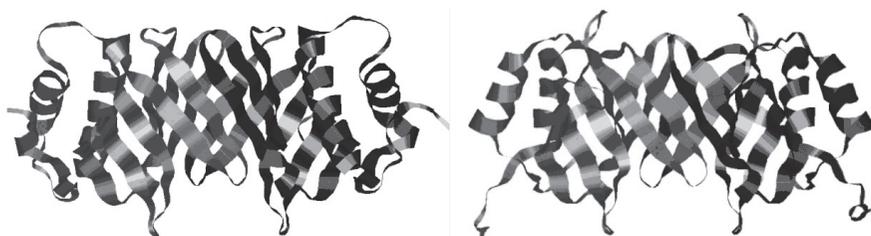


図3 : CASP9における我々の4次構造（複合体構造）予測の成功例。（右）タンパク質「3NA2」の予測構造。（左）後に実験的手法で明らかにされた天然構造。

4. 生体分子のシミュレーション解析

これまで述べてきた「立体構造」は固いものではなく、実際にはタンパク質は安定構造のまわりで熱ゆらぎによってゆらゆらと形を変えている。タンパク質を構成する原子同士あるいは周辺のイオンや水分子などの間の相互作用を2体間力の和としてモデル化し、ミクロなタンパク質を一種の構造物のようにとらえることで、周波数分析をしたり（基準振動モード解析）、刻一刻の形の変化を追いかけてたりすることができる（分子動力学法）。

我々は、精密な分子認識を行う転移RNA-タンパク質複合体に対する基準振動モード解析

を行い、両者が一体となって大きな低周波運動を行っていることを明らかにした⁹⁾。また、分子のとりうる構造を広範にサンプリングして自由エネルギー曲面を見積もることができる拡張アンサンブル法を分子動力学法と組み合わせ、10個のアミノ酸から成る小さなタンパク質などの安定性とアミノ酸配列の関係を明らかにした¹⁰⁾。また天然のものよりも強く相互作用する転移RNA-タンパク質複合体に対するシミュレーション解析により、その強い相互作用の理由を明らかにすることに成功した¹¹⁾。

5. 終わりに

ゲノム科学の分野では個人の全ゲノム情報を短時間で読める時代が近づいており、ますます大量データのマイニング技術が重要になってくる。また、細胞の中で起きている分子間の制御関係を表現するシステム生物学と、本稿で述べたような立体構造を直接扱う相互作用解析の間には、まだ大きな隔りがある。スパコンやアルゴリズムの発達により、タンパク質分子のシ

ミュレーション時間はこの20年で約1万倍に伸びたが、さらにまだ何桁も高速化を進める必要がある。遺伝子間の網羅的な相互作用ネットワークの分子レベル、原子レベルでの理解に向けて、計算技術にも分子生物学技術にもさらなるブレイクスルーが望まれており、我々もそれに向けて努力していく所存である。

参考文献

- 1) M. Kakuta, S. Nakamura, K. Shimizu, *IPJSJ Trans. Bioinform.*, 49, SIG 5 (TBIO 4) , 25-35, 2008.
- 2) W. Cao, K. Sumikoshi, S. Nakamura, T. Terada, K. Shimizu, *Bioinformation*, 6 (5) , 204-206, 2011.
- 3) S. Someya, M. Kakuta, M. Morita, K. Sumikoshi, W. Cao, Z. Ge, O. Hirose, S. Nakamura, T. Terada, K. Shimizu. *Adv. Bioinform.*, 2010, Article ID 289301, 2010.
- 4) W. Cao, J.-I. Maruyama, K. Kitamoto, K. Sumikoshi, T. Terada, S. Nakamura, K. Shimizu, *J. Gen. Appl. Microbiol.*, 55 (5) , 381-393, 2009.
- 5) M. Morita, M. Kakuta, K. Shimizu, S. Nakamura, *J. Proteome Sci. Comput. Biol.*, 1, 1, 2012.
- 6) V. Mariani, F. Kiefer, T. Schmidt, J. Haas, T. Schwede, *Proteins*, 79 (S10) , 37-58, 2011.
- 7) M. Morita, S. Nakamura, K. Shimizu, *Proteins*, 73 (2) , 468-479, 2008.
- 8) K. Sumikoshi, T. Terada, S. Nakamura, K. Shimizu, *Int. Proc. Chem., Biol. Environ. Eng.*, 5, 342-347, 2011.
- 9) S. Nakamura, M. Ikeguchi, K. Shimizu, *Chem. Phys. Lett.*, 372 (3-4) , 423-431, 2003.
- 10) D. Satoh, K. Shimizu, S. Nakamura, T. Terada, *FEBS Lett.*, 580 (14) , 3422-3426, 2006.
- 11) S. Yamasaki, S. Nakamura, T. Terada, K. Shimizu, *Biophys. J.*, 92 (1) , 192-200, 2007.



中村 周吾 (なかむら しゅうご)

1968年7月26日生まれ

[専門領域]

生命情報科学

[著書・論文]

- 中村周吾, 清水謙多郎, タンパク質立体構造予測 - de novo モデリング -, 遺伝子医学 MOOK 14 「次世代創薬テクノロジー 実践: インシリコ創薬の最前線」, 竹田-志鷹, 梅山編, メディカルドゥ, 49-54, 2009.
- M. Morita, M. Kakuta, K. Shimizu, S. Nakamura, Blind prediction of quaternary structures of homo-oligomeric proteins from amino acid sequences based on templates, *J. Proteome Sci. Comput. Biol.*, 1, 1, 2012.
- S. Nakamura, K. Shimizu, Comprehensive analysis of sequence-structure relationships in the loop regions of proteins, *Genome Inform.*, 23(1), 106-116, 2009.

[現在の所属]

大学院情報学環

(兼務) 大学院農学生命科学研究科

[所属学会] 日本生物物理学会、日本蛋白質科学会、日本バイオインフォマティクス学会

『東京大学大学院情報学環紀要』 投稿規定

- (1) 東京大学大学院情報学環教員等（教授、准教授、助教、客員教授・准教授、研究員等）は、本紀要および英文紀要に論文を日本語または英語で執筆することができる。
- (2) 東京大学大学院学際情報学府博士課程在籍者および東京大学大学院人文社会系研究科博士課程在籍者で大学院情報学環教員を指導教員としている者は、論文を日本語または英語で投稿することができる。大学院博士課程学生の投稿論文の採否は、図書・出版委員会が指名した情報学環教員と外部の委託された研究者による査読を経て、図書・出版委員会において決定される。
- (3) 執筆及び投稿される論文は未刊行のものに限る。定期刊行物（学術雑誌、商業雑誌、大学・研究所紀要など）や単行本として既刊、あるいは、これらに投稿中の論文は本誌に投稿できない。但し、学会発表抄録や科研費などの研究報告書はその限りではない。
- (4) 投稿する者は、指定された期日までに、執筆要項の諸規定にそって作成した原稿をプリントアウトしたもの2部およびそのデータファイルのフロッピーディスクやCD等を、東京大学大学院情報学環・学際情報学府図書室に提出しなければならない。
- (5) 本紀要に掲載された論文は、大学院情報学環のホームページで公開される。

『東京大学大学院情報学環紀要』 執筆要項

執筆・投稿

- (1) 執筆・投稿に際しては、東京大学大学院情報学環・学際情報学府図書室のホームページ（<http://www.lib.isics.u-tokyo.ac.jp/index.html>）に本投稿規定と執筆要項に関連する最新の情報が掲載されているので必ず参照すること。特にテンプレートに記載された細則に注意すること。
- (2) 原稿はA4版、横書きを原則とする。1頁は40字×34行。パソコンで作成する。
- (3) 分量は原則としてA4版で打ち出し10～30頁とする。大学院生の投稿の場合はA4版で打ち出し、表紙・英文要旨を除き本文14頁以内とする（注・参考文献・図表を含む）。枚数の上限は厳守すること。
- (4) 執筆要項に適した書式のテンプレートを東京大学大学院情報学環・学際情報学府図書室のホームページからダウンロードできるように準備してあるので、これらの雛形を用いて執

筆・提出を行うこと。

ファイル形式

- (5) ファイルは「Word」または「一太郎」の文書ファイルで提出すること。併せてプレーンテキストファイル形式で保存したものを提出する。その際、改行コードは1パラグラフに対して1つ入れること。表示行に対して改行コードの入ることのないように注意する。論理行に対して改行コードが1つとする。
- (6) Macを用いて執筆した場合は、提出するフロッピーディスクはDOS/Vフォーマットを使用すること。
- (7) フロッピーディスクのラベルには、論文名・執筆者名（複数の場合は代表者）・使用したワープロソフト名を明記すること。

全体の構成

- (8) 論文は、「表紙」「英文要旨」「本文」からなり、この順番で構成される。図・表は本文中に組み込む。
- (9) 右上ヘッダ部分に、通しのページ数をふること。
- (10) 1頁の余白は、上25mm 下30mm 右23mm 左23mmに設定する。
- (11) フォントはMS明朝10.5ポイントを標準とする。
- (12) 字句・叙述は簡潔・明確にして常用漢字、現代仮名遣い、算用数字を原則として用いる。

表紙書式

- (13) 表紙には、日本語の標題、著者名、著者の所属を、和文および英文で記載する。また主要著者の連絡先、研究助成に関する記述、謝辞、共同執筆の場合の執筆分担なども表紙に記す。
- (14) 日本語の標題は30字以内とする。副題がある場合は、「-」（ハイフン）の後に主題と明確に区別する形で記載する。その下に著者名と著者所属を日本語で記す。1頁目の日本語標題はMS明朝12ポイントで記す。
- (15) 日本語の標題、著者名、著者所属の下に、英語での標題（主題・副題）、著者名、著者所属を記す。英語標題は、筆頭語と主要語の頭文字を大文字で表記する。また英語の主題と副題は「:」で区切る。
- (16) 著者名の英語表記は原則としてFirst name を先とし、頭文字を大文字にする。日本名のローマ字使用法は執筆者の慣行を尊重し、統一しない。
- (17) 執筆者の所属に、教授・准教授・助教その他の別を記す必要はない。共同執筆の場合の記

載方法詳細はテンプレートを参照すること。

- (18) 標題、著者、著者所属に続けて、主要著者の連絡先、研究助成に関する記述、謝辞、共同執筆の場合の執筆分担などを記す。

英文要旨

- (19) 英文要旨の頭に「Abstract」（ゴシック体）と記す。
- (20) 英文要旨はA4版で1～2頁とする。英文に関しては、特に記述に注意し、執筆者の責任において英語を母語とする人の校閲を経ること。
- (21) 英文要旨の下に、キーワードを日本語と英語で記す。日本語キーワードは「キーワード：」に続けて6つ前後記す。日本語キーワードに続けて、英語キーワードを「Key Words:」（ゴシック体）に続けて記す。キーワードの筆頭語および主要語の頭文字は大文字とする。各キーワードはコンマで区切り、最後のキーワードの末尾にピリオドを付ける。

本文書式

- (22) 本文の開始ページの頭に、日本語および英語の標題を記す。
- (23) 本文中には、数字・記号を用いて章・節を設ける。章にあたるものは「1. , 2. , …」（全角数字及びドット）とし、節にあたるものは「1.1 …, 1.2 …,」（半角数字及びドット）とする。以下これに準ずる。章題・節題、強調部分は、太字ではなく、MSゴシック 10.5ポイントを用いること。

例) 章題の例	2. 携帯電話利用実態（全角の数字とドット）
節題の例	2.1 利用頻度・利用料金（半角の数字に全角スペース）
節以下の例	2.1.1 男性の利用頻度（上に同じ）
	2.1.1.a 男性の利用頻度の詳細（上に同じ）

- (24) 目次は、原則として各論文毎には付けない。但し、学位論文の一括掲載や長編の調査研究論文などの場合には付けることができる。
- (25) 本文中における外国人名などの固有名詞は、原綴りあるいは英語綴りを原則とするが、公式の名称として著名なものはカタカナでもよい。
- (26) 本文中での参考文献の引用は著者姓と発行年をつけて次の例のようにする。著者が3人以上の場合には初出の際には全著者の姓を書き、2度目以降は第一著者の姓を書き、和文献では「他」、欧文文献では「et al.」を書き添える。

例) Rumelhart, Hinton, & Williams(1980)は…
…と主張している（丸山・田中・谷口, 1998）。

- (27) 査読にあたっての匿名性を確保するため、自己の既発表論文等の引用にあたっては、「拙

- (39) 各論文執筆者には別刷30部と掲載誌3部を配布する。
- (40) 本投稿規定及び執筆要項の改正は図書・出版委員会の決議を経なければならない。

著者紹介の執筆

- (41) 論文の掲載が決まった著者は、著者紹介と自分の写真一葉を提出する。著者紹介には、生年月や出身大学などの履歴、専門、主たる著書・論文、所属、所属学会などを書くことができる。

附則 この規定・要項は、平成21年1月16日から施行する。
東京大学大学院情報学環 図書・出版委員会

東京大学大学院情報学環紀要 情報学研究 No.83

印 刷	平成24年10月1日
発 行	平成24年10月1日
編集・発行	東京大学大学院情報学環
郵便番号	113-0033
住 所	東京都文京区本郷7-3-1
電話番号	03-5841-5905
ファクシミリ	03-5841-5916
E-mail :	tosyo@iii.u-tokyo.ac.jp
装 丁	木 下 弥
印刷・製本	株式会社創志企画