

# ミルクタンパク質の構造生物学 —実験データに基づく構造決定から AI とビッグデータを活用した構造予測まで

西尾俊亮\*

(福島大学食農学類附属発酵醸造研究所, 〒960-1296 福島県福島市金谷川1番地)

## Structural biology of milk proteins – From structure determination based on experimental data to AI and big data-aided structure prediction

Shunsuke NISHIO

(Institute of Fermentation Sciences (IFeS), Faculty of Food and Agricultural Sciences, Fukushima University, 1, Kanayagawa, Fukushima, 960-1296)

### 要旨

生命現象の理解には、その中で働くタンパク質の機能を知ることが不可欠であり、タンパク質の立体構造はその手助けとなりうる。しかし、実験的に構造を決定するのは多大な時間と労力が必要である。本稿ではまず、実験的にタンパク質の立体構造を決定する手法について紹介する。次に、2021年に発表され、瞬く間に全世界に広まった人工知能 (Artificial Intelligence, AI) とビッグデータを活用した構造予測について、その手法と、現時点で予測できることについて述べる。さらに、栄養成分として、また生体防御機構の調節に重要な役割を果たすものとして、ミルクタンパク質を例に、実験的に立体構造が決定されているものと予測構造との比較を試みる。最後に、未知のミルクタンパク質複合体を予測し、予測構造がこれまでに観察されている現象を理解するのに妥当であるか、また、これらの構造予測の現時点における限界についても検討する。

### 1. はじめに

すべてのタンパク質はアミノ酸が繋がっただけの紐のようなものであろうか？もちろんそのような訳はなく、折りたたまれて特定の立体構造を取るもの、部分的に折りたたまれるもの、あるいはまさに紐のように決まった構造を持たないものなど様々である。さらに、単独では機能を持たないが、2つのタンパク質あるいはそれ以上のタンパク質が複合体を形成したときに機能を発揮するものも存在する。この事実から言えることは、特定の構造をとらない

ということも含め、タンパク質の機能を理解するためにはその立体構造を知ることが重要であると言える。タンパク質の立体構造を知るために多くの実験技術が生まれ、今なお発展中である。さらに近年、AIを活用したタンパク質立体構造の予測技術が急速に進歩している。

### 2. 測定データに基づく構造解析法

最初にタンパク質の立体構造解析が成功した1958年から65年が経ち、タンパク質の構造解析はソフト、ハードの両面で著しい進歩を遂げている。今日、実験的にタンパク質の立体構造を決定

\* E-mail : snishio@agri.fukushima-u.ac.jp

する手段として用いられるのは、主に、X線結晶構造解析、核磁気共鳴 (Nuclear Magnetic Resonance, NMR) およびクライオ電子顕微鏡を用いた解析である。本稿ではこれらのうち、X線結晶構造解析およびクライオ電子顕微鏡を用いた解析について述べる。

## 2-1. X線結晶構造解析

X線結晶構造解析は、タンパク質の結晶にX線を照射し、タンパク質分子に当たって回折したX線のスポット (回折データ) を用いる。回折データと位相情報から電子密度を計算し、そこに分子モデルを当てはめていくことで立体構造を決定する。

タンパク質の結晶を作製するために重要なことは、精製した標的タンパク質の純度および均一性である。純度を上げる方法として、タンパク質の多段階精製が挙げられる。組換えタンパク質発現技術を用いた異種発現によって標的タンパク質を調製する場合は、数残基から数百残基のタンパク質タグを付加し、第1段階としてアフィニティークロマトグラフィー、第2段階以降にイオン交換クロマトグラフィーやサイズ排除クロマトグラフィーによってさらに精製するのが一般的である。

一方、均一性には異種発現に使用する宿主の翻訳後修飾が影響する。哺乳類細胞や昆虫細胞を宿主として分泌発現した場合、組換えタンパク質には糖鎖が付加する。糖鎖、特にN-結合型糖鎖の構造は多岐にわたり、局所的な不均一性の原因となる。これを回避するためには、糖鎖修飾配列の置換が考えられる。しかし、N-結合型糖鎖はタンパク質の折りたたみや安定化に寄与していることが報告されており、このような変異タンパク質は発現しないことも珍しくない<sup>1)</sup>。そこで、糖鎖の構造をある程度均一にするために、糖鎖合成遺伝子に変異が導入された細胞株 (例えば HEK293S GnTI<sup>-</sup>) を使用したり、培地へ糖鎖合成阻害剤 (キフネンシンなど) を添加したりする<sup>2)</sup>。これらによって合成されるハイマンノース型糖鎖は、その糖鎖構造を特異的に切断する酵素である Endoglycosidase H を用いて比較的温和

な条件で切断できる。

純度および均一性は、サイズ排除クロマトグラフィーにおけるピークの対称性や、多角度光散乱検出器 (Multi Angle Light Scattering, MALS) による分子量の確認で評価することが多い。

タンパク質結晶は、低分子化合物のそれと同じように、分子が3次元空間に規則正しく整列して形成される。ただ、低分子化合物と異なり、タンパク質表面に露出しているアミノ酸の側鎖がそれぞれ異なる電荷を持つため表面電荷が一樣ではない。これにより、緩衝液のpHや温度によって表面電荷が変化し、結晶中で異なる分子間相互作用を形成しうるため、一つのタンパク質でも異なる整列による結晶が形成される<sup>3)</sup>。

結晶形成には、タンパク質溶液、緩衝液および沈殿剤 (塩、ポリマー、アルコールなど) が必要となる。これらを混合しタンパク質溶液を過飽和の状態にすると、均一に分散していたタンパク質分子が会合し始め、小さな結晶の核を形成する。核形成によって周囲の溶液のタンパク質濃度が低下すると、結晶はタンパク質分子を取り込み成長する段階へ移行する<sup>4)</sup>。

それぞれのタンパク質は固有の形、大きさを持ち、外環境によって表面電荷などの性質が変化することから、結晶化の予測は困難である。そのため、キットを用いて結晶化スクリーニングを行う必要がある。小さくても結晶が出ればその条件を、回折実験に使用可能な結晶を得るために最適化する。X線結晶構造解析においてこの部分が律速になることは多く、筆者が高度好塩菌由来のタンパク質を結晶化した際、条件の最適化を試みても数ヶ月かけなければ結晶が成長しなかった<sup>5)</sup>。

十分に結晶が成長したら、放射光施設での回折実験に供する。結晶中のタンパク質分子に当たったX線は回折し、様々な方向に曲がった結果生じる多数のスポット (回折データ) が得られる。タンパク質分子は3次元的な構造を持つため、角度を変えながらX線を照射し、それぞれの角度における回折データを取得する。回折データには、それぞれのスポッ

トの位置およびシグナル強度の情報が含まれている。X線は電磁波の一種であるので振幅と位相というパラメータがあり、これらの情報を用いることで電子密度を計算できる。しかし、振幅はシグナル強度として得ることができるが、波の周期性を表す位相は回折データから得られない。

そこで、位相情報を取得するために様々な手法が開発された<sup>6)</sup>。ここでは3つの方法について述べる。

重原子同型置換法は、重原子を含む結晶と含まない結晶を用意し、これらの回折データから、重原子の結合により生じたシグナル強度の差を利用して位相情報を取得する方法である。この方法は、新規、既知を問わずどのようなタンパク質に対しても用いることができるが、同型の結晶が複数個必要なことや、重原子が結合する割合がわからないこと、重原子の浸漬で結晶そのものが割れてしまうなどの問題がある。

異常散乱法は、セレンメチオニンや硫黄含有アミノ酸（システインおよびメチオニン）の異常散乱を利用し、回折データの差を取得し、位相を決定する方法である。この方法は重原子同型置換法と異なり、結晶1つから回折データを取得すれば良いという利点がある。しかし、セレンメチオニンを含む特殊な培地があることや、タンパク質の発現量が大きく低下することなどが問題となる。さらに、1つの結晶に2つ以上の異なる波長のX線を使う必要があるため、それに対応したビームラインで実験を行わなければならない。硫黄の異常分散を利用する場合は1つの波長のX線を用いればよいが、異常分散シグナルがとても小さいため、結晶1つからのデータでは十分に構造決定に至らない場合が多く、複数個の結晶から得たデータを積算する必要がある。

分子置換法は、現在最も使われている方法で、1つの結晶から得られた回折データと、既知の類似構造や予測構造から取得した位相情報を用いる。特別な波長のX線や、複数の条件の結晶を用意する必要はないが、構造決定の成功は鑄型構造が取得できるかどうか大きく依存する。既知構造を鑄型とする場合は、アミノ酸配列の同一性が少なくとも

30%程度以上のものが必要となる。

回折データと位相情報が揃えば、電子密度を計算することができる。電子密度は、タンパク質を構成しているアミノ酸の主鎖および側鎖、結合している低分子化合物あるいは短い糖鎖などの形を反映している。実際のモデルの構築はPhenix AutoBuildなどのプログラムが自動で行うが、実験者が電子密度図に合わせながらモデルを適宜修正し、最終的な立体構造を決定する<sup>7)</sup>。結晶化や位相情報の取得に比べて、この段階は方法論がほぼ確立しているため、1日でモデルが構築できることも珍しくない。

## 2-2. クライオ電子顕微鏡を用いた構造解析

近年、タンパク質の構造決定において用いられる機会が増えてきたのが、クライオ電子顕微鏡を用いたものである。クライオ電子顕微鏡は、液体窒素による冷却下でタンパク質などの試料に対して電子線を照射し観察する透過型電子顕微鏡のひとつである<sup>8)</sup>。

クライオ電子顕微鏡を用いた構造決定法は単粒子解析とトモグラフィー、マイクロ電子回折の3つがあるが、ここでは単粒子解析とトモグラフィーについて述べる。

単粒子解析は、タンパク質粒子を非晶質の氷に包埋し、2次元の透過画像から3次元構造を再構築する方法である。単粒子解析の大きな特徴は、結晶化を必要としないため、より自然な状態の構造を解析できることである。また、局所的に構造が異なるタンパク質が含まれていても、モデルを構築する際にそれらを別々に解析することで、1つのサンプルで複数の構造を解くことも可能である。一方で、標的タンパク質の分子量が壁となっており、リボソームなどの超分子複合体や界面活性剤などで被覆された膜タンパク質複合体など、分子量が10万以上のもの構造解析が多い。ただし、可溶性の小分子タンパク質について、抗体などを基にしたスキャフォールドタンパク質と結合させて見かけの分子量を大きくするなどの改善策が提案されている<sup>9)</sup>。

標的タンパク質は、X線結晶構造解析と同様に培

養細胞などから精製する。精製したタンパク質は、直径3 mmほどの金属でできたグリッドに添加し、余分な液を除去したあとで液体エタンを用いて急速凍結する。グリッドには規則的に穴が開いたカーボン膜が貼られており、それぞれの穴にタンパク質粒子が氷包埋される。グリッドに添加するタンパク質は数 $\mu\text{g}$ 程度で、X線結晶構造解析に用いるそれより少ない。そのため、遺伝子改変動物や特異抗体を用いて生体内から微量のタンパク質およびタンパク質複合体を精製し、試料として用いることも増えている<sup>10)</sup>。

グリッドを作製後、クライオ電子顕微鏡を用いてグリッドの透過画像を数百枚から数千枚取得する。単粒子解析を行う前提として、グリッドのそれぞれの穴に包埋されているタンパク質粒子はランダムな方向を向いている必要がある。そのような状態で撮影した画像は、標的タンパク質を360°様々な角度から見たものが含まれている。また、染色をしていないため、観察される粒子のコントラストはとても低い。そこで、透過画像から粒子を抽出し、似た形、すなわち同じ向きと推定されるものを分類し整列させた後、重ね合わせることでノイズを軽減させる。ノイズを軽減させた画像に角度を割り当てて逆投影することで、最終的な3次元構造を再構築する<sup>11)</sup>。

トモグラフィー(クライオ電子線トモグラフィー)は、グリッドを連続的に傾斜させて1つのサンプルから多数の画像を撮影し3次元構造を取得する方法である。単粒子解析の場合と異なり、1つの構造体を様々な角度から撮影するため、細胞小器官や細胞といった、単粒子解析より大きい、より不均一なサンプルに適用できる<sup>12)</sup>。

クライオ電子顕微鏡を用いた構造解析は、結晶化を必要としない、必要なタンパク質量が少ないといった利点があり、サンプルを調製するのにX線結晶構造解析ほどは時間がかからない。しかし、画像データを取得するために一晩かかる場合があり、またデータ解析法も確立されていないため試行錯誤が必要となる。

### 3. ビッグデータとAIを活用した構造解析法

実験データを基にタンパク質の立体構造を決定する方法は、専用の実験機器や設備、さらにはデータを測定するための施設へのアクセスなど、超えるべき障壁が多い。それでは、これらの設備を持たないタンパク質研究者、生化学者はどのようにして標的タンパク質の構造にアクセスすればよいのか？また、標的タンパク質やそれに類似のタンパク質の構造がない場合はどうすればよいのか？これらの問いに答えるため、公共データベースの整備と立体構造予測技術の開発が進められた。

#### 3-1. 既知のタンパク質立体構造へのアクセス

PDB (Protein Data Bank) は、タンパク質や核酸などの立体構造を登録する公共データベースである<sup>13)</sup>。実験的に決定された構造は、論文と同時にPDBへの登録、公開が義務付けられている。PDBには21万を超える立体構造が登録されており(2023年11月)、これらの中にはタンパク質の全長構造だけでなく、特定のタンパク質ドメインの構造、基質や他のタンパク質との複合体構造など様々な状態のものが含まれている。PDBに登録されている標的タンパク質の構造を見るには、PDBから構造ファイルをダウンロードし、PyMOLやUCSF Chimera<sup>14)</sup>といった分子グラフィックツールを用いる。

#### 3-2. タンパク質の立体構造予測技術

##### ①ホモロジーモデリング法

ホモロジーモデリング法は、標的タンパク質のアミノ酸配列に相同性が高く、立体構造既知の鋳型タンパク質を出発点とする方法である。これは、アミノ酸配列が類似しているタンパク質同士は、進化的類縁関係にあり構造も類似であるということに基づいている。標的タンパク質と鋳型タンパク質との配列アライメント情報から、アミノ酸残基が挿入あるいは欠失している場合は主鎖の構造を再構成し、またアミノ酸置換が起こっている場合は側鎖の付け替えを行う。これにより標的タンパク質の立体構造を

予測するが、既知の構造がない場合は使用できないため、万能な手法ではない。

## ② AI 駆動型構造予測法

ホモロジーモデリング法に変わり近年急速に発達し、またその予測精度の高さが注目されているのが深層学習を利用した構造予測である。その代表が、2020年のCASP14<sup>1</sup>でDeepMind社が発表し、2021年に論文が公開されたAlphaFold (AlphaFold2) である<sup>15)</sup>。

AlphaFoldによる予測は、配列データベースから標的タンパク質の類縁配列を多数検索し、挿入や欠失に対応して作製されたマルチプルシーケンスアライメント (MSA) と、PDB データベースに登録されている類似の構造 (テンプレート) を用いる。MSA を利用する上で重要な考え方として、あるアミノ酸の変異がタンパク質の構造に影響を与える場合、近接するアミノ酸が安定化に寄与するように変異するというものがある。すなわち、MSAにおいてこのような共変異しているアミノ酸ペアを見つけることができれば、それらの位置情報を立体構造の予測に使うことが可能となる。AlphaFold は、MSA からアミノ酸ペアを探し出す処理に Evoformer という深層学習モデルを用いており、MSA の膨大な類縁配列から特徴的な部分や、特徴が見られない部分を、学習を繰り返しながら見出す。これにより、保存されたアミノ酸や共変異しているアミノ酸ペアを精度良く探し出すことが可能となる。

次に AlphaFold は、Evoformer によって得られた MSA に由来するアミノ酸ペアなどの特徴情報を基に標的タンパク質のアミノ酸配列を3次元空間に配置する。アミノ酸残基の座標情報は、再度 Evoformer へフィードバックされ、特徴情報を更新する。これらの処理を繰り返し行い、最終予測構造を出力する。

AlphaFold は、予測構造とともに構造を評価

するためのいくつかの指標を出力する。pLDDT (predicted Local Distance Difference Test) は、予測構造と既知構造とを比較した際のずれを、アミノ酸残基一つ一つに対して数値化したものである。AlphaFold は、予測構造と PDB に登録されている既知構造との誤差について学習済みであるため、実験構造がない場合でも、予測構造の pLDDT を算出する。pLDDT が予測構造における局所的な信頼性を表しているのに対して、pTM score (predicted Template Modeling score) は、予測構造と既知構造とを比較し、予測構造全体の信頼性を評価している。PAE (Predicted Aligned Error) は、予測構造と既知構造において、任意の2つの残基を選択し、一方を重ねたときにもう一方の残基の位置誤差を予測したものである。マルチドメインタンパク質において、ドメイン内の2つの残基の PAE は低くなるものの、ドメイン間の2つの残基では高く示される。これにより、マルチドメインタンパク質のドメイン間の境界を推測することができる。pTM score と PAE については関連があり、全体的に PAE が小さくなるほど、pTM score は大きくなる。なお、これらの仕組みについて詳しく知りたい場合は、森脇による総説<sup>16)</sup> にこれまでの背景なども含めてまとめられているので参照されたい。

## ③ AI 駆動型構造予測法の使い方

AlphaFold を用いて自分の興味があるタンパク質の立体構造を取得するにはどうしたらいいだろうか？ AlphaFold はソースコードが公開されているため、ローカルのコンピュータに実装することが可能である。しかしコンピュータの GPU によって、計算速度や計算できるタンパク質配列の長さが制限される。より簡単に予測したい場合は、DeepMind 社が Google Colaboratory で提供している AlphaFold の利用が可能である<sup>2</sup>。また、ハーバード大などのグループが AlphaFold の MSA 構築の部分を変更し計算速度を向上させた ColabFold<sup>17)</sup> も同じく Google

<sup>1</sup> Critical Assessment of protein Structure Prediction。2年に一回開催されるタンパク質構造予測技術の精度を評価する世界的なコンテスト。

<sup>2</sup> <https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb>

Colaboratory で利用できる<sup>3</sup>。Google Colaboratory での予測は、興味があるタンパク質のアミノ酸配列を入力するだけで行うことができる。また、DeepMind 社は、AlphaFold の論文が発表されたのと同時に、Uniprot に登録されているタンパク質の予測構造データベースを公開した<sup>18</sup>。2021 年の段階では 30 万程度の予測構造が公開されていたが、2023 年 11 月現在では 2 億を超えている。

AlphaFold による構造予測は単量体タンパク質にとどまらず、タンパク質複合体にも適応可能である。これは、複合体の相互作用界面においてそれぞれのタンパク質に由来するアミノ酸残基が共変異するという情報を利用している。また、複合体予測の精度を上昇するために、それぞれのポリペプチド鎖に由来する MSA から同じ生物種由来の配列を探して組み合わせている。これは同じ生物種由来のタンパク質同士のほうが、異なる種の組み合わせより相互作用が強いはずであるという性質を利用している。複合体予測は AlphaFold-Multimer として発表され<sup>19</sup>、Google Colaboratory では、複数の配列をインプットに用いるだけで簡単に予測できる。

#### ④ AI 駆動型の構造予測で何が変わったのか

AlphaFold の論文はすでに 1 万 6 千回以上引用されているが、AlphaFold の出現後、研究スタイルはどのように変わったのか。構造予測だけにとどまらない例を説明する。

1 つ目は、X 線結晶構造解析における構造決定法のひとつである分子置換法である。これまでの予測構造は分子置換法のモデルに用いられるほど正確ではなく、構造決定が困難であった。しかし、AlphaFold の正確性が高い予測構造は分子置換法への適応が十分可能である<sup>20</sup>。たとえば、筆者らが明らかにした構造は、ホモロジーモデリング法によって作製したモデルでは 1 年かかっても解けなかったが、AlphaFold によるモデルを用いたところ、新規の構造であったにも関わらず 1 日もかからずに解く

ことができた<sup>5</sup>。

2 つ目は、立体構造を利用したアミノ酸配列によらない類似構造のマイニング（採掘）である。類似の遺伝子やタンパク質を探す方法として、塩基配列やアミノ酸配列を用いて検索する BLAST が広く知られている<sup>21</sup>。構造マイニングはこれまでもあったが、データベースが PDB のみで、その利用は限定的であった。しかし、AlphaFold データベースの公開により予測構造が充実したため、Dali<sup>22</sup> や FoldSeek<sup>23</sup> を用いた大規模構造マイニングが可能になった。さらに、立体構造予測 AI のひとつである ESMFold<sup>24</sup> によるデータベースが公開されたことも大きい。ESMFold データベースには、メタゲノム配列から予測した約 7 億の予測構造が含まれている。これらを用いた検索は BLAST と同様で、インプット構造を入力するだけで類似構造が得られる。

3 つ目は、AlphaFold-Multimer を利用した新規相互作用ペアの探索である。AlphaFold-Multimer とオミクスデータとを組み合わせて網羅的な複合体予測を行い、*in silico* で新規タンパク質複合体を探すことが試みられている<sup>25</sup>。

4 つ目は、クライオ電子顕微鏡を用いた構造解析への応用である。前述の通り、クライオ電子顕微鏡を用いた単粒子解析およびトモグラフィーでは、2 次元画像から 3 次元構造を再構成する。このときに得られる「未知タンパク質の形状」という情報に対し、形状が合致する AlphaFold による予測構造を探索することにより、タンパク質複合体の新規構成成分を同定したことが報告されている<sup>26</sup>。

AlphaFold の公開当初に示唆されていた構造予測の制約・制限も解消されつつある。例えば、予測構造はタンパク質のある一状態のみを出力可能であるとされていた。しかし、MSA に含まれている配列を類似度でクラスターに分けると、同じタンパク質の異なるコンフォメーションを予測することが可能であると最近になって報告された<sup>27,28</sup>。また、様々な疾患で、その病態に関連するアミノ酸置換を伴う一塩基多型（Single Nucleotide Polymorphisms, SNPs）が報告されている。AlphaFold でこのような

<sup>3</sup> <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>

アミノ酸置換による構造変化を予測することは困難とされていた。それは、MSAで多くの配列を取得するため、アミノ酸置換という特徴がその他多数の配列によって消されてしまうからである。しかし、この問題もMSAをクラスター化することでアミノ酸置換による構造変化を予測することができたり<sup>27)</sup>、変異によるタンパク質機能への影響を予測するために改変されたAlphaFoldを含むAIであるAlphaMissense<sup>29)</sup>が発表されたりと、解決されつつある。

#### ⑤ AI駆動型構造予測の今後

タンパク質立体構造予測の次の段階として、タンパク質と核酸や金属イオン、低分子化合物との複合体や、翻訳後修飾されたタンパク質の構造予測が挙げられる。これらの予測に関しても近年成果が公開されている。

立体構造予測AIのひとつであるRoseTTAFold<sup>30)</sup>の改良版RoseTTAFoldNA (Nucleic Acid)は、タンパク質-核酸複合体を予測するAIである<sup>31)</sup>。RoseTTAFoldNAは、テストしたタンパク質-核酸複合体の約3割で相互作用界面を正しく予測できることを示した。

同じくRoseTTAFoldの改良版であるRoseTTAFoldAA (All-Atom)やUmolは、金属イオンや低分子化合物との複合体予測ができるAIである<sup>32,33)</sup>。タンパク質-低分子化合物複合体は、既存構造の受容体タンパク質に低分子化合物を当てはめるドッキングシミュレーションが主流であるが、これらのAIは、構造未知の受容体タンパク質の場合でもドッキングシミュレーションと同等の予測精度を達成している。この他に、標的タンパク質の予測構造から、既知の類似構造-低分子化合物複合体を探索し、低分子化合物を予測構造に埋め込むAlphaFillという手法も開発されている<sup>34)</sup>。

AlphaFoldの予測構造は、翻訳後修飾の一つであるN-結合型糖鎖の付加部位がタンパク質表面に正しく露出しており、そこにモデル糖鎖を埋め込むことができると報告されている<sup>35)</sup>。また、

RoseTTAFoldAAは既存の翻訳後修飾を含むタンパク質の立体構造のうち、5割弱のものを予測することに成功している<sup>32)</sup>。

タンパク質と非タンパク質の複合体予測については、AlphaFoldも今後のアップデートでそれらの予測について対応すると発表しており、先行で公開されたデータでは高精度での予測に成功している<sup>4)</sup>。

多くの場合、AlphaFoldなどのAIによる予測構造は、実験的に得られた構造を高精度で再現している。しかし、前述の指標により高い信頼性を示したモデルでも、モデル全体のねじれやドメインの配向が実験による構造と異なっていたり、局所的にはアミノ酸残基の主鎖および側鎖のコンフォメーションが異なっていたりすることもある<sup>36)</sup>。AlphaFoldに限らず、AIはもっともらしい嘘をつくこと(Hallucination)がある<sup>37)</sup>。そのため、今後はモデルの信頼性を予測精度の指標に頼るだけではなく、別の実験系でも正しく検証することがより重要になるとと思われる。

## 4. ミルク関連タンパク質の構造解析

ミルクは新生児に栄養を与えるために高機能化した、栄養学的に優れた食べ物である。新生児に対する母乳のみならず、牛乳や乳製品は広く日常的に摂取している。牛乳は、100 mLあたり3から4 gのタンパク質(ミルクタンパク質)を含む。ウシのミルクタンパク質は、約80%がカゼインで残りの約20%がホエイタンパク質で構成されるのに対し、ヒトのそれは約70%がホエイタンパク質である<sup>38,39)</sup>。

カゼインは、ウシで4種類( $\alpha_{s1}$ ,  $\alpha_{s2}$ ,  $\beta$ ,  $\kappa$ )、ヒトで3種類( $\alpha_{s1}$ ,  $\beta$ ,  $\kappa$ )のタンパク質が含まれており、乳中ではカゼイン同士が自己会合し、カゼインミセルと呼ばれる粒子として分散している。カゼインミセルは、水に溶けにくいリン酸カルシウムを内包し、新生児にカルシウムを送達する機能を担っている<sup>40)</sup>。

<sup>4)</sup> <https://deepmind.google/discover/blog/a-glimpse-of-the-next-generation-of-alphafold/>

ホエイはカゼインと異なり、多様な種類のタンパク質を含む。カゼインと比べて、生体内での消化・吸収速度が早いいため、新生児のアミノ酸源となる。さらに、その機能も多様で、IgA (Immunoglobulin A) などの免疫グロブリンは、新生児の未熟な免疫系を助ける役割を持つ。また、Lactoferrin は鉄結合能によって鉄吸収を調節したり、その分解物が抗菌作用を示したり、生体に有用な機能を示す<sup>41)</sup>。

これら以外にも、量的には僅かであるが重要な役割を果たすミルクタンパク質が存在する。乳は、100 mL あたり 3 から 5 g の脂肪 (乳脂肪) を含み、乳脂肪は三重膜で被覆された乳脂肪球となって分散している。この三重膜は、乳脂肪球被膜 (Milk Fat Globule Membrane, MFGM) と呼ばれ、乳腺上皮細胞内で合成され、脂質一重膜で被覆された脂肪滴が、離出分泌に伴い頂端側細胞膜によって被覆されて形成される<sup>42)</sup>。MFGM は、約 2-4% のミルクタンパク質を含み、タンパク質の種類も多岐にわたる。これらのタンパク質の中には、乳脂肪球の離出分泌 (Butyrophilin 1A1, BTN1A1 ; Xanthine Oxidoreductase, XOR) や、乳中での分散 (Milk fat globule-Epidermal growth factor (EGF) factor 8, MFG-E8) など、乳脂肪球の形態に寄与するものが

含まれる<sup>43-45)</sup>。

#### 4-1. ミルクタンパク質の立体構造

PDB に登録されているミルクタンパク質について、表 1 にまとめた。ミルクタンパク質のうち、ホエイタンパク質の立体構造がよく解かれている。これは、乳 (牛乳) が、大量に調製できるため、精製の出発物質として用いるのが容易であったこと、ホエイタンパク質が水溶性で扱いやすかったことが要因であると考えられる。さらに、免疫系の調節や抗菌など、生体防御に寄与するタンパク質が多く、その立体構造への関心が高かったことも理由であろう。また、それぞれのタンパク質について複数のエントリーがある。例えば Lactoferrin は、基質を含まないアポフォーム、鉄結合型、亜鉛結合型や糖結合型などの結晶構造や、Lactoferrin の分解物である抗菌ペプチド Lactoferricin の NMR 構造が報告されている<sup>46-50)</sup>。これらの構造は、新規基質結合サイトの同定や基質の結合に伴うタンパク質の構造変化を議論するために有用である。また、ホエイタンパク質の結晶構造はその由来がウシやヒトだけでなく、ヤギやヒツジ由来のものも登録されている。

MFGM タンパク質はホエイタンパク質と異なり、

表 1 PDB に登録されている乳タンパク質

ホエイタンパク質	PDB	生物種
Lactoferrin	1LFG など 100 以上	ヒト、ウシ
$\beta$ -Lactoglobulin	1BEB など 100 以上	ウシ、ヤギ
$\alpha$ -Lactalbumin	1HML など 30 以上	ヒト、ウシ、ヤギ
Lactoperoxidase	2IPS など 70 以上	ウシ、ヤギ、ヒツジ
カゼインタンパク質	PDB	生物種
$\alpha$ <sub>s2</sub> -Casein	6FS4 (Casocidin) 6FS5 (Casocidin)	ウシ
MFGM タンパク質	PDB	生物種
Mucin 1	2ACM (SEA domain)	ヒト
	6BSB (SEA domain)	
	6BSC (SEA domain)	
Xanthine oxidoreductase	1FO4 など 30 以上	ヒト、ウシ、ラット
Cluster of differentiation 36	5LGD	ヒト
Butyrophilin 1A1	4HH8	ウシ
	3BN6	
Milk fat globule- EGF factor 8	2PQS	ウシ、マウス
	2L9L	
Fatty acid binding protein 3	2HMB など 60 以上	ヒト、ウシ、マウス



完全長の構造が解かれているものは限定的である。XORは、ホエイタンパク質と同様に牛乳中から容易に精製できるため登録されている構造が多い。また、Fatty acid binding protein 3 (FABP3)の構造も多く登録されている。これは、FABP3が長鎖脂肪酸の細胞内輸送を担っており、その機構に関心が持たれた結果であると推定される。これら以外のMFGMタンパク質は、それらの機能に關与する特徴的なドメインの構造が登録されている。

一方で、ウシのミルクタンパク質の80%を占めるカゼインであるが、その構造は $\alpha_{s2}$ -カゼインの分解物であるCasocidinのものだけである<sup>51)</sup>。これは、カゼインが天然変性タンパク質 (Intrinsically Disordered Protein, IDP) であり、そのポリペプチド鎖の大部分が一定の立体構造をとらず、結晶化などの構造解析が困難なためである<sup>52)</sup>。

#### 4-2. ミルクタンパク質の構造予測

前述の通り、ミルクタンパク質の立体構造はホエイタンパク質を中心に明らかになっている。しかし、機能的に重要なタンパク質であるにも関わらず部分構造しか解かれていないものや、複合体で機能するがその立体構造は解明されていないものも多い。そこで、本稿ではいくつかのミルクタンパク質、複合体の構造を実際に予測し、実験データとの整合性や、予測構造から推察される機能などについて述べる。

##### ① MFG-E8

MFG-E8はMFGMタンパク質のひとつで、脂質二重膜の構成成分であるホスファチジルセリン (PS) に結合する<sup>53)</sup>。通常PSは脂質二重膜の内側に存在するが、アポトーシス細胞では外側に露出する。露出したPSにMFG-E8が結合すると、それを標的にしてマクロファージが貪食する<sup>54)</sup>。この機構は、乳腺退縮期に起こる乳腺上皮細胞のリモデリングにも利用される<sup>55)</sup>。また、MFG-E8を欠いた乳脂肪球は融合し大きくなることから、MFG-E8は乳中で乳脂肪球の大きさを制限し、乳脂肪球を分散させ

るために必要な分子である<sup>45)</sup>。

MFG-E8はマルチドメインタンパク質で、アミノ末端側にEGF-likeドメインを持ち、カルボキシ末端側にF5/8 type Cドメインを持つ (図1A)。EGF-likeドメインは生物種によって数が異なり、ウシでは2つ、ヒトでは1つである。F5/8 type Cドメイン (Cドメイン) は生物種問わず2つであり、2番目のドメインがPS結合能を持つ。図1Bに、AlphaFoldデータベースで公開されているウシMFG-E8の予測構造を示した。予測構造には、アミノ末端にpLDDT値が低い領域が見られた。AlphaFoldによる予測構造でアミノ末端側にpLDDT値が低い領域がある場合、その部分はシグナル配列であることが多い。2番目のEGF-likeドメインは、インテグリンとの結合モチーフであるRGD配列を持つ。予測構造から、RGD配列は大きいループ構造の先端にあり、タンパク質の外側に露出していることが示唆された。図1Cに、PAEプロットを示す。PAEプロットは縦軸、横軸ともにアミノ酸残基の番号に対応しており、任意の2つのアミノ酸の交点にそのアミノ酸ペアのPAE値が示されている。PAEプロットの濃くなっている部分は、アミノ酸残基同士の位置関係、配向性が適切であることを示しており、四角く囲まれている部分がタンパク質ドメインであると推定できる。また、隣接するドメイン間で相互作用しているかを見ると、2番目のEGF-likeドメインとC1ドメインとの間でPAEプロットが濃くなっており、これらのドメインの相対的な位置関係はある程度固定された状態であると推定される。他方、C1ドメインとC2ドメインの間ではPAEプロットが薄く、位置関係はある程度自由度があると考えられる。このことから、PSとの結合するC2ドメインと、インテグリンと結合するRGD配列を持つEGF-likeドメインとは柔軟に位置関係を変えられるため、マクロファージとアポトーシス細胞とを架橋するのに適していると推察される。さらに、結晶構造<sup>56)</sup>と予測構造とを重ねて比較したものを図1Dに示した。C2ドメインの結晶構造とAlphaFoldによるC2ドメインのアミ

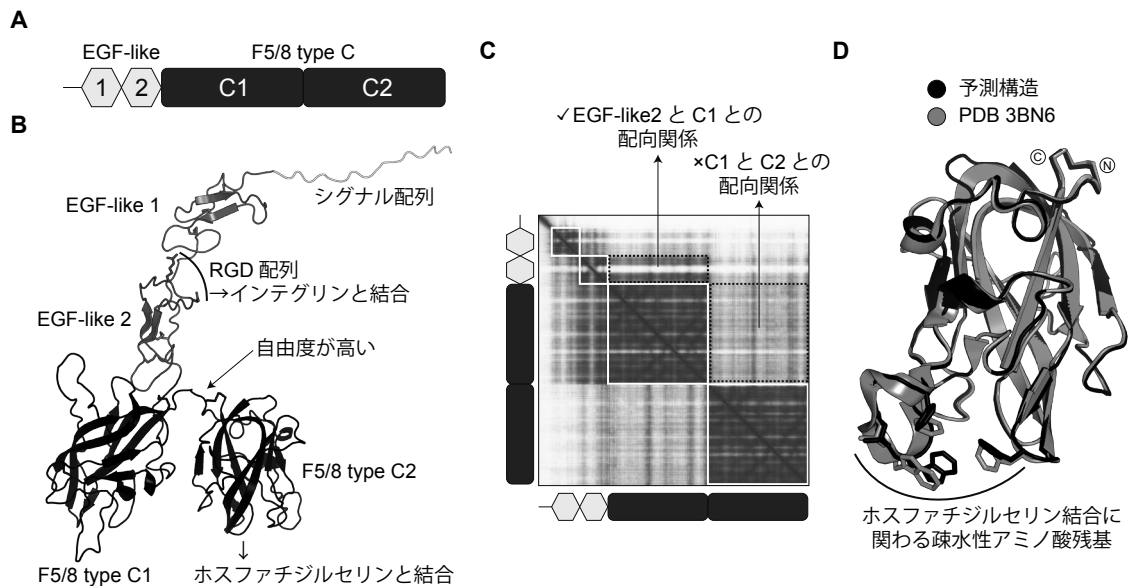


図1 ウシ MFG-E8 の予測構造

- (A) ウシ MFG-E8 の模式図。  
 (B) AlphaFold データベースで公開されているウシ MFG-E8 の予測構造。  
 (C) AlphaFold データベースで公開されているウシ MFG-E8 の PAE プロット。色が黒いほど PAE 値が低く、アミノ酸残基間の位置情報や配向性が信頼できる。  
 (D) ウシ MFG-E8 予測構造の C2 ドメインと、実験的に決定された C2 ドメイン (PDB 3BN6) との構造比較。

ノ酸主鎖はよく重なり、RMSD<sup>5</sup>は0.378であった。しかし、PS との結合に重要な疎水性アミノ酸残基の側鎖は、一部異なる向きであった。以上から、MFG-E8 の予測構造はこれまでの知見をうまく説明でき、また機能性ドメインの構造は、結晶構造を精度良く再現していた。

## ② BTN1A1-XOR 複合体

BTN1A1 および XOR は、MFGM に存在するタンパク質である。図 2A 左に AlphaFold データベースで公開されているウシ BTN1A1 の予測構造を示す。BTN1A1 は乳腺上皮細胞の細胞膜に局在する 1 回膜貫通型タンパク質で、細胞外領域に Immunoglobulin (Ig)-like ドメインを、細胞内領域に B30.2/SPRY ドメインを持つ。一方、XOR は細胞質に存在する全長 1000 アミノ酸残基を超える巨大な球状タンパク質である。これらのタンパク質をそれぞれ単独で欠損させたマウスは、乳腺上

皮細胞内に巨大な脂肪滴が蓄積し、乳脂肪球が乳中へ分泌されなくなる<sup>43,44</sup>。また、マウスにおいて *btn1a1* および *xor* 遺伝子は同一の mRNA 結合タンパク質 TDP-43 によって転写後制御を受けており、これらのタンパク質が協調して働くことを示唆している<sup>57</sup>。これらのタンパク質は、乳脂肪球形成時に乳腺上皮細胞由来の細胞膜と脂肪滴の膜をつなぐ役割を持つが、その相互作用は BTN1A1 の B30.2/SPRY ドメインと XOR との間で起こる。組換えタンパク質を用いたプルダウンアッセイによって、B30.2/SPRY ドメインのアミノ末端あるいはカルボキシ末端領域が XOR 認識に重要であると報告されている<sup>58,59</sup>。しかし、BTN1A1-XOR 複合体の構造に関する情報はない。そこで、これらの複合体を予測し、既報との整合性を検証した。なお、Google colabatory の AlphaFold ではアミノ酸残基数の制限から予測できなかったため、自前の計算システムを構築し予測した。

図 2A 右に、全長 BTN1A1 と XOR の複合体構造を、AlphaFold (ColabFold) を用いて予測した結果

<sup>5</sup> Root Mean Square Deviation. タンパク質の構造間の誤差を数値化したもの。小さいほど誤差が少ないと言える。

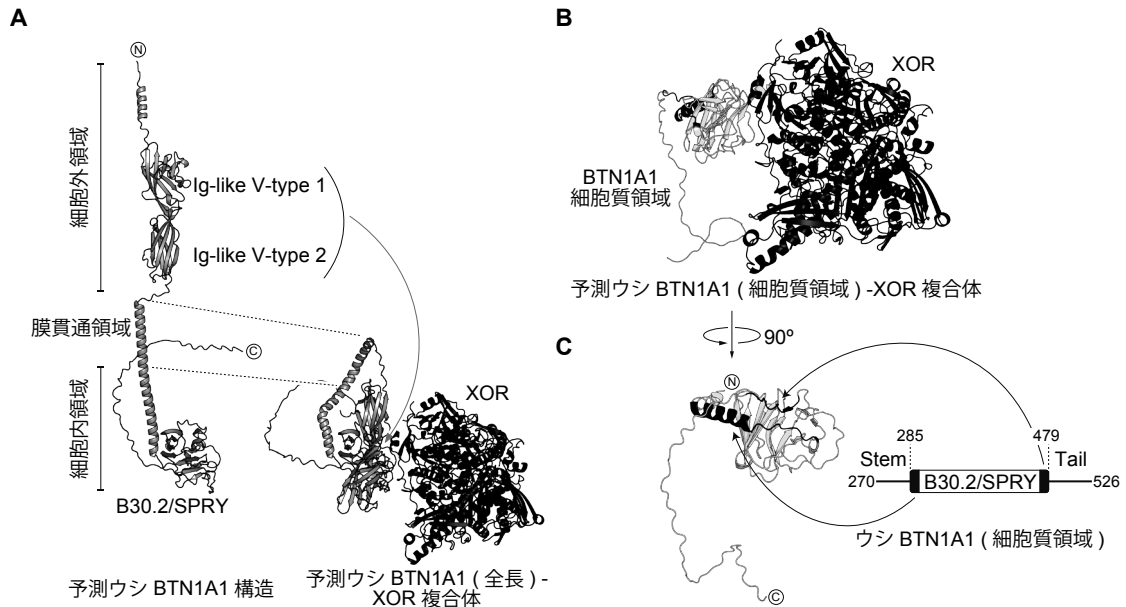


図2 ウシ BTN1A1 と XOR の複合体予測

(A) AlphaFold データベースで公開されているウシ BTN1A1 の予測構造 (左) および, ColabFold (AlphaFold-multimer) を用いて予測したウシ BTN1A1-XOR 複合体 (右)。  
 (B) ウシ BTN1A1 (細胞質領域) -XOR 複合体の予測構造。  
 (C) 予測ウシ BTN1A1 (細胞質領域) -XOR 複合体から, BTN1A1 (細胞質領域) のみを表示した (左)。また, BTN1A1 (細胞質領域) の模式図を右に示した。パネル B から 90° 回転している。既報で示された XOR 認識領域を黒で表した。

を示す。AlphaFold は, BTN1A1 の細胞外領域と細胞内領域のどちらもが XOR と相互作用するモデルを出力した。本来, 細胞膜によってこれらの領域は隔てられているため, 生体内でこのような相互作用はありえない。これは, AlphaFold が膜貫通領域および脂質二重膜に関する情報を持たないために起こったと考えられる。

そこで, 既報の実験結果に従い, BTN1A1 の細胞内領域である B30.2/SPRY ドメインと XOR との複合体を予測した。図 2B にその複合体の構造を示す。AlphaFold が出力したモデルは, プルダウンアッセイによって示された B30.2/SPRY ドメインの XOR 認識領域の反対側で結合するというものであった。なお, B30.2/SPRY ドメインに着目すると, プルダウンアッセイによって示された XOR 認識領域は, 一次構造では離れているものの立体構造上では近接していることが示唆された (図 2C)。

この結果の解釈として 2 つ考えられる。1 つ目は, 既報で証明された領域が結合界面として妥当で,

AlphaFold による予測構造は間違っているということである。AlphaFold の複合体予測について, 592 個の複合体 (ダイマー) を予測した場合の成功率は 60% ほどであり<sup>60)</sup>, 今回の場合も予測ができなかったことが考えられる。ただし, 生体内ではこの 2 者以外に相互作用するタンパク質の存在も示唆されており<sup>61)</sup>, 今回は計算システムの能力 (GPU) の制限で叶わなかったが, それらも含めた超分子複合体を予測できれば結果は変わるかもしれない。2 つ目は, 既報で用いた組換えタンパク質が 10 から 20 残基の比較的長い領域を欠損させたため, それにより全体的な立体構造が変化して結合能が失われたというものである。今回の予測構造が正しいかはまだ分からないが, この構造に基づき 1 アミノ酸レベルで変異を導入することで, この可能性が正しいか確認できると思われる。今後, 実験的に BTN1A1-XOR 複合体の構造が決定され, 予測構造との比較ができるようになり, 乳脂肪球の離出分泌機構が明らかになることを期待したい。

### ③ カゼイン

乳中のカゼインミセルの構造は長年議論されており、未だに統一見解は得られていない<sup>40)</sup>。カゼインミセル研究黎明期は、カゼインによる小さなミセル（サブミセル）がコロイド状リン酸カルシウム（Colloidal Calcium Phosphate, CCP）によって架橋され大きなミセルを形成するサブミセルモデルが提唱された。これらのサブミセルは、大きなミセルの表面に近いものでは親水性の $\kappa$ -カゼインを含み、内部のサブミセルは含まないとされた。しかし、現在ではサブミセルモデルは否定されており、かわりにナノクラスターモデルが提唱されている<sup>62)</sup>。ナノクラスターモデルは、リン酸カルシウム（CP）ナノクラスターの周りにカゼイン分子（ $\alpha$ -および $\beta$ -カゼイン）が、リン酸化セリンのリン酸基を介して結合したものを1つの単位とし、これらがカゼイン同士の相互作用で凝集し大きなミセルを形成すると

いうものである。こちらも、表面に親水性の $\kappa$ -カゼインを含むと考えられている。また、カゼインミセルは水を内包するという報告もされている<sup>63)</sup>。

これらのカゼインモデルに対して、複合体構造予測はなんらかのヒントを与えることができるだろうか？カゼインミセル一つには約  $10^4$  個のカゼイン分子が含まれているとされるが<sup>64)</sup>、それほどに巨大な複合体構造を予測することはできない。

そこで、ウシカゼイン分子それぞれの存在比に従って、 $\alpha_{s1}$ -カゼインを6分子、 $\beta$ -カゼインを6分子、 $\kappa$ -カゼインを2分子含む複合体を、AlphaFoldを用いて予測してみた。AlphaFoldに5つのモデルを出力してもらい、前述の指標を基にランク付けを行った。図3Aに5つのモデルのうち上位3つを示した。なお、これらのモデルについて、予測精度等の各指標は高くない点、また、AlphaFoldによる複合体予測では、同じ配列を持つ

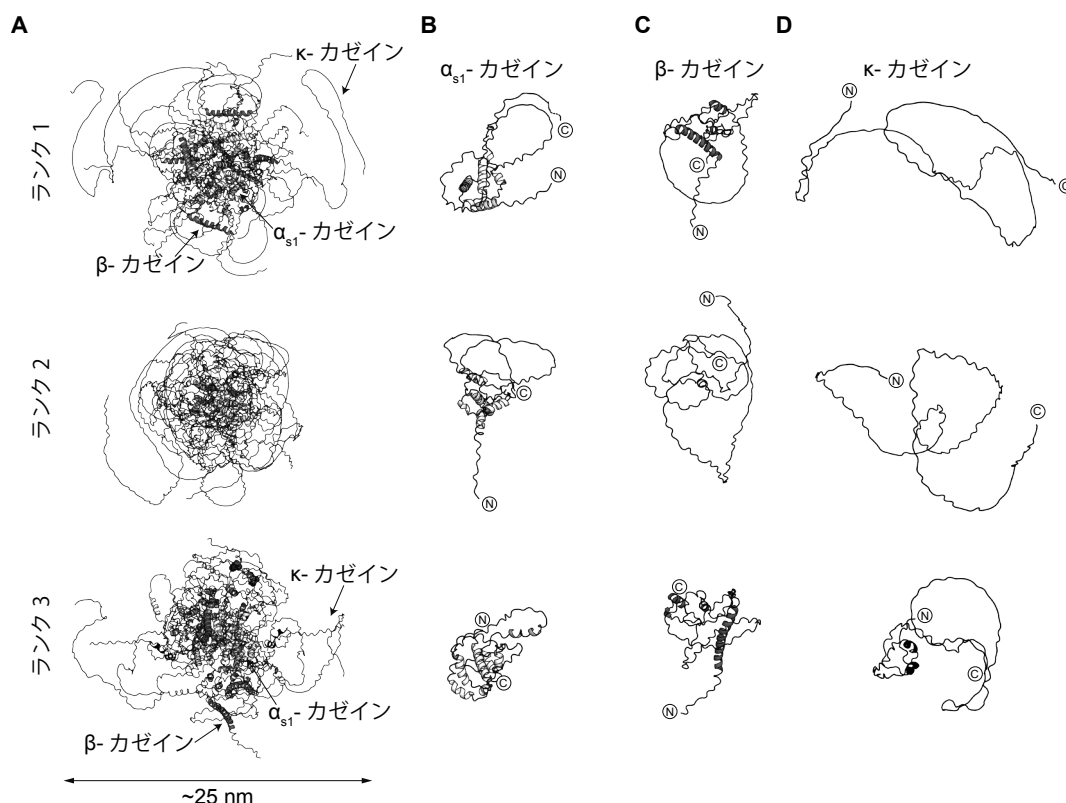


図3 カゼイン複合体の構造予測

(A) ColabFoldを用いて予測したカゼイン複合体。ウシ由来の $\alpha_{s1}$ -カゼイン6分子、 $\beta$ -カゼイン6分子、 $\kappa$ -カゼイン2分子からなる複合体を予測し、5つのモデルを出力した。そのうち上位3つを示した。  
 (B-D) それぞれのモデルを構成するカゼイン分子から1分子を抽出し、その予測構造を示した。(B)  $\alpha_{s1}$ -カゼイン、(C)  $\beta$ -カゼイン、(D)  $\kappa$ -カゼイン。

タンパク質が複数ある場合はそれらのタンパク質に  
対称性を持たせるようにモデルを形成するという点  
に留意していただきたい。二次構造を取り、構造が  
一つに決まるようなタンパク質およびタンパク質複  
合体の場合は5つのモデルがほとんど同じになる  
のに対して、IDPであるカゼイン分子による複合体  
はすべてのモデルが異なった構造を示した。これら  
のモデルは、直径約25 nm (250 Å) であった。全  
てのモデルで共通して見られる特徴として、 $\alpha_{s1}$ -カ  
ゼインが複合体の中心に集合し、その外側に $\beta$ -カ  
ゼイン、さらに外側に $\kappa$ -カゼインが結合している  
点が挙げられる。これは、サブミセルモデル、ナノ  
クラスターモデルの双方が提示している結合様式と  
一致している。次に、それぞれのサブユニットの  
予測構造を図3B-Dに示した。予測構造は、ほと  
んどの領域で二次構造を持たなかったが、一部で $\alpha$   
ヘリックス構造が見られた。 $\alpha_{s1}$ -カゼインは5モデル  
全てで、同じ領域の $\alpha$ ヘリックスが予測され(図  
3B)、ランク2のモデルではこの $\alpha$ ヘリックスが複  
合体の中心で集まっていた。一方、 $\beta$ -カゼインお  
よび $\kappa$ -カゼインはモデル間で構造が大きく異なっ  
ており、これら2つの分子は $\alpha_{s1}$ -カゼインより構造

の自由度が高いことが推定された(図3C, D)。そ  
の中で、 $\kappa$ -カゼインのランク3モデルで見られた $\alpha$   
ヘリックスは、他の分子との結合界面に見られた。  
これらの結果から、カゼインミセルを形成する際に  
それぞれのサブユニットは部分的に折りたたまれて  
 $\alpha$ ヘリックス構造をとり、この構造変化が凝集を引  
き起こすのに重要であることが推察された。

カゼインは種々の翻訳後修飾を受けたタンパク質  
であり、また部分分解によって機能性ペプチドを生  
じる<sup>65)</sup>。これらの領域が立体構造上どこに位置して  
いるかを、ランク1のモデルを用いて示した(図  
4)。カゼインの中で、 $\kappa$ -カゼインのみがO-結合  
型糖鎖を持つ糖タンパク質であり、6箇所のスレオ  
ニン残基に糖鎖が付加する<sup>66)</sup>。図4左の3つの球  
で示されている $\kappa$ -カゼインの糖鎖付加スレオニン  
は、複合体構造の最外部に位置していた。この修  
飾スレオニンを含む部分はグリコマクロペプチド  
(GMP)と呼ばれ、生体内で免疫系や消化器系の調  
節、腸内細菌叢の制御や抗菌作用など幅広い機能  
が報告されている<sup>67,68)</sup>。しかし、カゼインミセルに  
おける本質的な役割は、糖鎖付加により親水性を  
上昇し、乳中でミセルを安定化することである<sup>40)</sup>。

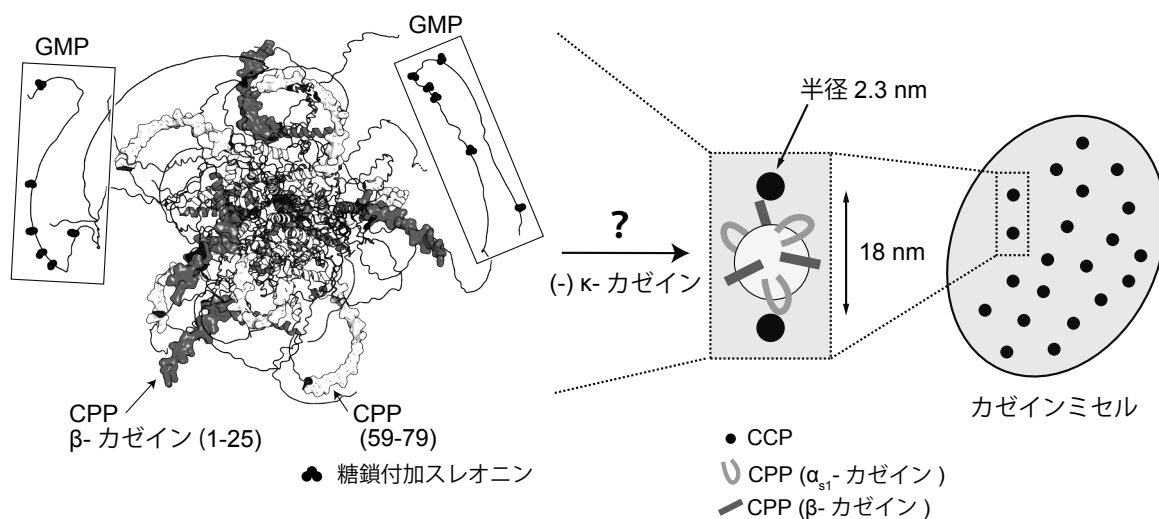


図4 予測カゼイン複合体(ランク1)の翻訳後修飾部位とカゼインミセルの構造

予測カゼイン複合体のランク1モデルに、トリプシンなどで消化されることで生じるリン酸化セリンに富んだペプチド(カゼインホスホペプチド, CPP)およびO-結合型糖鎖付加スレオニンに富んだペプチド(グリコマクロペプチド, GMP)の部位を示した。ナノクラスターモデルに基づいたカゼインミセル構造では、コロイド状リン酸カルシウム(CPP)が、18 nmの間隔で分散している。これは、ランク1モデルから $\kappa$ -カゼインを除いたもの(直径~20 nm)に近い。

AlphaFoldによる予測構造は、その機能をうまく説明しているように見えた。

別の重要な機能性ペプチドとして、カゼインホスホペプチド (CPP) がある。CPP はリン酸化セリン残基を含むペプチドで、生体内でカルシウムイオンと結合し小腸でのカルシウム吸収を促進する<sup>69)</sup>。主要な CPP として、 $\alpha_{s1}$ -カゼインおよび $\beta$ -カゼイン由来のものがある。図 4 左に、CPP の配列が複合体のどの部分に位置するかを白および灰色のサーフェス表示で示した。興味深いことに、現在の AlphaFold はセリンのリン酸化という情報を持たないにも関わらず、 $\alpha_{s1}$ -カゼインおよび $\beta$ -カゼインに由来する CPP 配列のどちらもが複合体の外側に露出していた。カゼインが持つリン酸化セリンは、カルシウムイオンを介して CCP と結合しカゼインミセルを形成する。カゼインミセル中の CCP 同士の間隔は、約 18 nm であると報告されている<sup>70)</sup>。 $\kappa$ -カゼインを含まない $\alpha_{s1}$ -カゼイン、 $\beta$ -カゼインそれぞれ 6 分子からなる複合体は直径が約 20 nm で、CPP 配列が中心から外向きに突出していたことから、 $\kappa$ -カゼインを除いた予測構造は、CCP 間を架橋する構造を再現しているのかもしれない (図 4)。

現在の AlphaFold を始めとした構造予測 AI における、カゼイン複合体予測の制限・限界についても述べる。最も重要な点として、翻訳後修飾やリガンド結合を考慮していないモデルであるということが挙げられる。特にカゼインミセルは、リガンドとして CP、翻訳後修飾としてリン酸化および糖鎖修飾が機能的にも構造的にも重要な役割を果たしている。今回提示したモデルはそれらを一切無視しているため、結果の解釈には注意が必要である。また、IDP は本質的に構造を取らないため、予測精度が低くなるという点も重要である。AlphaFold の予測した 5 つのモデルは、いずれも構造が全く異なっており、さらに予測精度を示す各指標がかなり低いものであった。今後は分子動力学シミュレーション (MD シミュレーション) などを組み合わせて、構造のスナップショットではない、動的なものとして

捉えていく必要がある。しかし、これらの問題点は今後の構造予測 AI の発展によって解消されることが十分予想される。その時に改めて構造を予測することで、カゼインミセルモデルの解明につなげていきたい。

## 5. おわりに

本稿ではいくつかのミルクタンパク質を例にして、タンパク質の立体構造を実験的に決定する手法と、AI を用いて *in silico* で構造を予測する手法について紹介した。未だに予測構造の精度が実験構造を完全に置き換えるには至っていないが、その差は確実に縮まっているように感じる。予測構造に容易にアクセスできるようになったことは、タンパク質の機能に対する理解を深化させる原動力となっている。今後も非常に速いスピードで発展していくであろう AI 駆動型の研究スタイルに置いていかれることなく、その利点を最大限に引き出すことが重要である。これらのツールを適切に活用することで、タンパク質の予想外の新たな機能がこれからも見つかっていくであろう。

## 参考文献

- 1) Molinari M. N-glycan structure dictates extension of protein folding or onset of disposal. *Nat. Chem. Biol.* **3**, 313–320. (2007)
- 2) Elbein A. D. et al. Kifunensine, a potent inhibitor of the glycoprotein processing mannosidase I. *J. Biol. Chem.* **265**(26), 15599–15605. (1990)
- 3) 沖英幸, 松浦良樹. タンパク質結晶におけるモルフォロジーと分子間接触の相関. *日本結晶成長学会誌*. **26**(4), 167–174. (1999)
- 4) Bijelic A. & Rompel A. Polyoxometalates: more than a phasing tool in protein crystallography. *ChemTexts*. **4**(3), 10. (2018)
- 5) Moi D. et al. Discovery of archaeal fusexins homologous to eukaryotic HAP2/GCS1 gamete fusion proteins. *Nat. Commun.* **13**(1), 3880. (2022)

- 6) Taylor G. The phase problem. *Acta Crystallogr. D Biol. Crystallogr.* **59(11)**, 1881–1890. (2003)
- 7) Terwilliger T. C. et al. Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallogr. D Biol. Crystallogr.* **64(1)**, 61–69. (2007)
- 8) Adrian M. et al. Cryo-electron microscopy of viruses. *Nature.* **308(5954)**, 32–36. (1984)
- 9) Wu X. & Rapoport T. Cryo-EM structure determination of small proteins by nanobody-binding scaffolds (Legobodies). *Proc. Natl. Acad. Sci. U. S. A.* **118(41)**, e2115001118. (2021)
- 10) Lin S. et al. Structure of a mammalian spermatid channel complex. *Nature.* **595(7869)**, 746–750. (2021)
- 11) Elmlund D. & Elmlund H. Cryogenic electron microscopy and single-particle analysis. *Annu. Rev. Biochem.* **84**, 499–517. (2015)
- 12) Turk M. & Baumeister W. The promise and the challenges of cryo-electron tomography. *FEBS Lett.* **594(20)**, 3243–3261. (2020)
- 13) Berman H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* **28(1)**, 235–242. (2000)
- 14) Pettersen E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25(13)**, 1605–1612. (2004)
- 15) Jumper J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* **596(7873)**, 583–589. (2021)
- 16) 森脇由隆. AlphaFold2 までのタンパク質立体構造予測の軌跡とこれから. *JSBi Bioinformatics Review.* **3(2)**, 47–60. (2022)
- 17) Mirdita M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods.* **19(6)**, 679–682. (2022)
- 18) Tunyasuvunakool K. et al. Highly accurate protein structure prediction for the human proteome. *Nature.* **596(7873)**, 590–596. (2021)
- 19) Evans R. et al. Protein complex prediction with AlphaFold-Multimer. *bioRxiv.* <https://doi.org/10.1101/2021.10.04.463034>
- 20) Terwilliger T. C. et al. Accelerating crystal structure determination with iterative AlphaFold prediction. *Acta Crystallogr. D Struct. Biol.* **79(Pt 3)**, 234–244. (2023)
- 21) Altschul S. F. et al. Basic local alignment search tool. *J. Mol. Biol.* **215(3)**, 403–410. (1990)
- 22) Holm L. et al. DALI shines a light on remote homologs: One hundred discoveries. *Protein Sci.* **32(1)**, e4519. (2023)
- 23) van Kempen M. et al. Fast and accurate protein structure search with Foldseek. *Nat. Biotechnol.* <https://doi.org/10.1038/s41587-023-01773-0> (2023)
- 24) Lin Z. et al. Evolutionary-scale prediction of atomic-level protein structure with a language model. *Science.* **379(6637)**, 1123–1130. (2023)
- 25) Homma F. et al. AlphaFold-Multimer predicts cross-kingdom interactions at the plant-pathogen interface. *Nat. Commun.* **14(1)**, 6040. (2023)
- 26) Chen Z. et al. De novo protein identification in mammalian sperm using in situ cryoelectron tomography and AlphaFold2 docking. *Cell.* **186(23)**, 5041–5053. (2023)
- 27) Wayment-Steele H. K. et al. Predicting multiple conformations via sequence clustering and AlphaFold2. *Nature.* <https://doi.org/10.1038/s41586-023-06832-9> (2023)
- 28) Porter L. L. et al. ColabFold predicts alternative protein structures from single sequences, coevolution unnecessary for AF-cluster. *bioRxiv.* <https://doi.org/10.1101/2023.11.21.567977>
- 29) Cheng J. et al. Accurate proteome-wide missense variant effect prediction with AlphaMissense. *Science.* **381(6664)**, eadg7492. (2023)
- 30) Baek M. et al. Accurate prediction of protein structures and interactions using a three-track

- neural network. *Science*. **373(6557)**, 871–876. (2021)
- 31) Baek M. et al. Accurate prediction of protein–nucleic acid complexes using RoseTTAFoldNA. *Nat. Methods*. <https://doi.org/10.1038/s41592-023-02086-5> (2023)
- 32) Krishna R. et al. Generalized Biomolecular Modeling and Design with RoseTTAFold All-Atom. *bioRxiv*. <https://doi.org/10.1101/2023.10.09.561603>
- 33) Bryant P. et al. Structure prediction of protein–ligand complexes from sequence information with Umol. *bioRxiv*. <https://doi.org/10.1101/2023.11.03.565471>
- 34) Hekkelman M. L. et al. AlphaFill: enriching AlphaFold models with ligands and cofactors. *Nat. Methods*. **20**, 205–213. (2023)
- 35) Bagdonas H. et al. The case for post-predictional modifications in the AlphaFold Protein Structure Database. *Nat. Struct. Mol. Biol.* **28**, 869–870. (2021)
- 36) Terwilliger T. C. et al. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat. Methods*. <https://doi.org/10.1038/s41592-023-02087-4> (2023)
- 37) Ji Z. et al. Survey of Hallucination in Natural Language Generation. arxiv. <https://doi.org/10.48550/arXiv.2202.03629>
- 38) 一般社団法人Jミルク. ファクトブック 乳たんぱく質のすべて. <https://www.j-milk.jp/report/study/hn0mvm00000a3dj.html>
- 39) Roy D. et al. Composition, Structure, and Digestive Dynamics of Milk From Different Species-A Review. *Front. Nutr.* **7**, 577759. (2020)
- 40) 青木孝良ら. カゼインミセルの構造モデルと乳の加工. *ミルクサイエンス*. **66(2)**, 125–143. (2017)
- 41) 今井哲哉. ホエータンパク質の健康機能と利用. *ミルクサイエンス*. **55(4)**, 227–235. (2007)
- 42) 安枝武彦ら. ミルク MFG-E8/lactadherin の乳脂肪品質管理における役割. *ミルクサイエンス*. **62(3)**, 125–129. (2013)
- 43) Ogg S. et al. Expression of butyrophilin (Btn1a1) in lactating mammary gland is essential for the regulated secretion of milk-lipid droplets. *Proc. Natl Acad. Sci. U. S. A.* **101(27)**, 10084–10089. (2004)
- 44) Monks J. et al. Xanthine oxidoreductase mediates membrane docking of milk-fat droplets but is not essential for apocrine lipid secretion. *J. Physiol.* **594(20)**, 5899–5921. (2016)
- 45) Yasueda T. et al. A protective effect of milk fat globule EGF factor VIII (MFG-E8) on the spontaneous fusion of milk fat globules in breast milk. *J. Biochem.* **158(1)**, 25–35. (2015).
- 46) Norris G. E. et al. Molecular replacement solution of the structure of apolactoferrin, a protein displaying large-scale conformational change. *Acta Crystallogr. B.* **47(Pt 6)**, 998–1004. (1991)
- 47) Haridas M. et al. Structure of human diferric lactoferrin refined at 2.2 Å resolution. *Acta Crystallogr. D Biol. Crystallogr.* **51(Pt 5)**, 629–46. (1995)
- 48) Smith C. A. et al. Metal substitution in transferrins: the crystal structure of human copper-lactoferrin at 2.1-Å resolution. *Biochemistry*. **31(18)**, 4527–4533. (1992)
- 49) Mir R. et al. Specific interactions of C-terminal half (C-lobe) of lactoferrin protein with edible sugars: binding and structural studies with implications on diabetes. *Int. J. Biol. Macromol.* **47(1)**, 50–59. (2010)
- 50) Hunter H. N. et al. Human lactoferricin is partially folded in aqueous solution and is better stabilized in a membrane mimetic solvent. *Antimicrob. Agents Chemother.* **49(8)**, 3387–3395. (2005)
- 51) Mercurio F. A. et al. The antimicrobial peptides



- casocidins I and II: Solution structural studies in water and different membrane-mimetic environments. *Peptides*. **114**, 50–58. (2019)
- 52) 瀬戸泰幸. 牛乳のタンパク質. オレオサイエンス. **23(8)**, 415–421. (2023)
- 53) Andersen M. H. et al. Functional analyses of two cellular binding domains of bovine lactadherin. *Biochemistry*. **39(20)**, 6200–6206. (2000)
- 54) Hanayama R. et al. Identification of a factor that links apoptotic cells to phagocytes. *Nature*. **417(6885)**, 182–187. (2002)
- 55) Hanayama R. & Nagata S. Impaired involution of mammary glands in the absence of milk fat globule EGF factor 8. *Proc. Natl. Acad. Sci. U. S. A.* **102(46)**, 16886–16891. (2005)
- 56) Shao C. et al. Crystal structure of lactadherin C2 domain at 1.7Å resolution with mutational and computational analyses of its membrane-binding motif. *J. Biol. Chem.* **283(11)**, 7230–7241. (2008)
- 57) Zhao L. et al. TDP-43 facilitates milk lipid secretion by post-transcriptional regulation of Btn1a1 and Xdh. *Nat. Commun.* **11(1)**, 341. (2020)
- 58) Ishii T. et al. Carboxy-terminal cytoplasmic domain of mouse butyrophilin specifically associates with a 150-kDa protein of mammary epithelial cells and milk fat globule membrane. *Biochim. Biophys. Acta*. **1245(3)**, 285–292. (1995)
- 59) Jeong J. et al. The PRY/SPRY/B30.2 domain of butyrophilin 1A1 (BTN1A1) binds to xanthine oxidoreductase: implications for the function of BTN1A1 in the mammary gland and other tissues. *J. Biol. Chem.* **284(33)**, 22444–22456. (2009)
- 60) Zhu W. et al. Evaluation of AlphaFold-Multimer prediction on multi-chain protein complexes. *Bioinformatics*. **39(7)**, btad424. (2023)
- 61) Monks J. et al. Organellar Contacts of Milk Lipid Droplets. *Contact Thousand Oaks*). <https://doi.org/10.1177/2515256419897226> (2020)
- 62) Holt C. A quantitative calcium phosphate nanocluster model of the casein micelle: the average size, size distribution and surface properties. *Eur. Biophys. J.* **50(6)**, 847–866. (2021)
- 63) Takagi H. et al. Temperature dependence of the casein micelle structure in the range of 10–40 °C: An in-situ SAXS study. *Food Chem.* **393**, 133389. (2022)
- 64) Holt C. et al. Invited review: Caseins and the casein micelle: their biological functions, structures, and behavior in foods. *J. Dairy Sci.* **96(10)**, 6127–6146. (2013)
- 65) Silva S. V. & Malcata F. X. Caseins as source of bioactive peptides. *International Dairy Journal*. **15(1)**, 1–15. (2005)
- 66) Pisano A. et al. Characterization of O-linked glycosylation motifs in the glycopeptide domain of bovine kappa-casein. *Glycobiology*. **4(6)**, 837–844. (1994)
- 67) 堂迫俊一. 牛乳中の機能性タンパク質. ミルクサイエンス. **59(3)**, 283–294. (2010)
- 68) Córdova-Dávalos L. E. et al. Glycomacropeptide Bioactivity and Health: A Review Highlighting Action Mechanisms and Signaling Pathways. *Nutrients*. **11(3)**, 598. (2019)
- 69) 小野伴忠. カルシウム摂取における乳, カゼインミセルの役割とカゼインホスホペプチドについて. *酪農科学・食品の研究*. **43(4)**, A-65-A-71. (1994)
- 70) De Kruif C. & Holt C. Casein Micelle Structure, Functions and Interactions. *In: Advanced Dairy Chemistry—1 Proteins*, Springer, 233–276. (2003)