

# タンパク質の構造の理解に向けて

京都産業大学 吉田賢右

宇宙の生命は・・・つまり地球の生命および地球外生命すべては・・・炭素と水の独特の化学的な性質に依存して存在しています。炭素は混成軌道により、シグマ（一重）結合、2重結合、3重結合、共鳴結合、環形成、などさまざまな化学結合が可能で、膨大な種類の化合物を作ることができます。液体の水は、遷移的な分子間水素結合による独特な物性（高い沸点など）および水素イオンの平衡的な解離によって、生命の分子にかけがえのない環境を提供します。炭素の代わりにケイ素を使い、水の代わりに液体アンモニアや液体硫化水素の中で生きる生命は存在するのでしょうか。現代の化学はそのわずかな可能性も示すことができません。

どんな宇宙生命でも、個体が活動し、子供を作る（遺伝）ことが必要でしょう。地球の生命は、遺伝にはリン酸ジエステル結合でつながれた核酸ポリマーである DNA（まれに RNA）を採用し、個体の活動にはアミノ酸がペプチド結合でつながれたタンパク質を採用しています。遺伝子の機能としては必要なのはコピー可能な1次元の文字列であり、それは DNA に限らずいろいろな高分子が採用可能でしょう。しかし、活動分子としてのタンパク質の機能は3次元的な立体構造に依存しており、そうすると、どんな宇宙生命もタンパク質を採用せざるを得ないと思われます（もちろん、20種のL型アミノ酸を使用しているのは地球生命に独特のものでしょうか）。それは、ペプチド結合の半2重結合性、 $-CO:HN-$ の主鎖間の水素結合、そして側鎖の多様性による広大な物性と多様な立体構造の可能性、などがタンパク質以外的高分子では不可能のように思えるからです。核酸でもきちんとした立体構造を持つものが知られています。しかし、その構造の多様性はごく限定されています。タンパク質は宇宙において生命を可能とした奇跡の化合物である、と言ってもいいでしょう。ですから、タンパク質の理解、とりわけタンパク質にとって命とも言える立体構造の研究は、予言的に言えば、宇宙生命的な重要性がある、と私は思っています。この分野において最近、革新的な開拓が進んでいるので、簡単に紹介します。

タンパク質は20種のアミノ酸がいろいろな順番で重合したヒモ状の高分子で、ヒモは折りたたまれて（folding）さまざまな立体構造を形成します。

「天然のタンパク質の立体構造はそのアミノ酸配列を持つタンパク質が水溶液中で取りうる構造の中で熱力学的に最も安定な構造である」というのがタンパク質の基本原理です。それなら、立体構造を計算できそうなものですが、ダメです。少なくともごく最近まではダメでした。

アミノ酸配列は、立体構造の暗号である Anfinsen の原理

天然に存在するタンパク質の立体構造形成 (folding) には、外から情報やエネルギーを与える必要はない。アミノ酸配列だけで立体構造は決まる

## 天然タンパク質の基本方程式

$$y = f(x)$$

y : 立体構造  
F : folding 関数  
x : アミノ酸配列

問題 (x) と答 (y) は12万セット知られているが、関数 (f) は未知である

この暗号解読は、難攻不落 (だった)

それなら、立体構造は実験で決めるしかありません。そのため、この 60 年間、X 線結晶解析が主要な方法として使われてきました。この方法では、結晶を作ることが今も昔も最大の難関ですが、解析手段・方法は長足の進歩をしています。例えば、時間を追って構造変化を見ることが出来ます。呼吸鎖酵素では酸素分子が銅に立ち寄ってからヘムに結合する過程がわかりました。また、膜脂質の原子の位置までわかる革新的な方法が開発され、膜タンパク質とリン脂質の相互作用が見えるようになりました。

ごく最近、低温電子顕微鏡 + 1 電子検出像 で多数の 1 分子 (単粒子) 画像解析を行う強力な方法が登場しました。結晶解析とは逆に、大きな分子を得意とし、結晶化は不要です。時間分解、複数の構造異性の解析も可能です。これにより、今までどうしても結晶化ができなかったタンパク質 (およびタンパク質複合体) の構造が次々と解明されています。

タンパク質の立体構造を計算によって予測しよう、という試みも何十年と続けられてきました。既知の立体構造を持つタンパク質に似たアミノ酸配列を持つタンパク質はやっぱり似た立体構造を持つだろう、ということで、この作業を計算機にやらせる、というのが今まで、一番当たりそうな方法でした。計算だけに頼るアプローチは 2 つあります。1 分子の構造変化の時間追跡 (MD, molecular dynamics)、多分子の統計力学的解析です。

MD はニュートン力学の計算です。1 つ 1 つの原子に化学結合の拘束と原子間力と電荷を与えて全原子の運動をフェムト秒刻みで追跡するのです。多くの星がお互いに重力を及ぼしながら運動する天文学の多体問題と似ています。そして、計算の出発点となる初期構造が必要です。

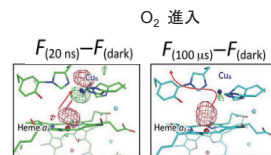
## タンパク質の立体構造は、1 つ 1 つ実験で決める

### X 線による結晶解析

#### X 線自由電子レーザー で動画

シトクロームオキシダーゼに CO (O<sub>2</sub> analog) を結合させておいて、閃光で CO を追い出す (O<sub>2</sub> 結合の逆)。nano sec の間に起きる構造変化と O<sub>2</sub> の進入経路を解明

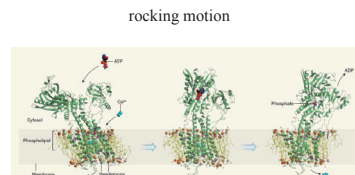
Shimada, ..., Yoshikawa, Tsukihara  
Sci Adv. 2017



#### 結晶解析で膜も見える

Ca pump のほぼ全中間体の構造を、周囲の膜 (~50 分子のリン脂質) とともに解明

Norimatsu, ..., Toyoshima  
Nature 2017

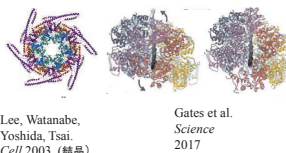


### 低温電子顕微鏡による単粒子解析

結晶困難なタンパク質の構造決定。  
大きい分子ほど有利。早い。

複数の構造の混合物でも  
分類できれば、構造がわかる

Hsp104; プリオン凝集体やアミロイドを溶かす



Lee, Watanabe,  
Yoshida, Tsai,  
Cell 2003 (結晶)

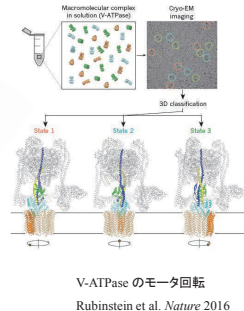
Gates et al.  
Science  
2017

さらに

時間分解の構造 (~msec)

連続的な構造変化

ATP synthase  
Meier et al. Mol  
Cell 2016



V-ATPase のモーター回転  
Rubinstein et al. Nature 2016

## 立体構造を計算で予測する

### 分子動力学 MD : 時間

運動力学:

1 分子の構造の動きの時間変化を追ってゆく

### 自由エネルギー計算 : 空間

統計力学:

ある一瞬の多数の分子の構造の分布

## タンパク質の分子動力学 (MD) シミュレーション

### ■ 力場の方程式 (古典力学)

$F =$  化学結合力 + ファンデルワール力 + 静電的力

$$F_i = m_i \frac{d^2 x_i}{dt^2} \quad i: \text{each atom}$$

### ■ 初期構造が必要

### ■ 計算: タンパク質全原子 (1 万 - 1000 万) と水原子

1 femto (10<sup>-15</sup>) sec の時間刻み

MDにはいくつか問題点があります。まず、計算の基礎となる力場の方程式ですが、量子論は含まれず、あくまで近似です。タンパク質の構造と言ったら誰でも知っている水素結合も疎水相互作用も含まれていません。中性のはずの炭素原子にもマイナス電荷を割り振ったりしています。小さい分子の計算結果が実際の分子の構造に合うように数値を割り振っているのです。はたしてこれが大きなタンパク質分子に有効なのか、予断できません。

初期構造も問題です。ふつう結晶解析などで決まった構造を初期構造として採用します。構造未知のタンパク質には使えません。水素の位置などはわからないのがふつうなので、推定です。また、実際のタンパク質の構造は揺らいでいます。知りたい構造が大きく揺らいだときの構造である場合、計算してもそこにたどりつけない可能性が大きい。多数の異なった初期構造から計算を繰り返すためには、多数の既知の（あるいは仮定の）構造が必要ですし、また計算量が膨大になります。

一番の問題は、計算量です。10 フェムト秒以上の時間刻みだとうまくいきません。1 フェムト秒後の（最低でも）数万の原子の位置を数値的に計算し、その計算結果の構造を元にまた 1 フェムト秒後の原子の位置を計算する、という繰り返しです。タンパク質の folding や生物的に意味のある構造変化は大抵の場合、ミリ秒以上のオーダーですから、繰り返しは膨大になります。世界最速の計算機でもマイクロ秒の計算がせいぜいです。

しかたがないので、この分野の研究者は「粗視化」をやります。大胆な単純化によって計算量を減らし、ミリ秒くらいの長時間の構造変化を見ようというのです。しかし、都合のいい粗視化の方法は未だ見つからず、計算結果はあまり信頼性がありません。薬の候補となる低分子とタンパク質の結合のスクリーニング計算でも当たるのは数パーセント以下だそうです。さらに問題なのは、これでは力場の方程式が有効なのかどうか、それもわからないことです。

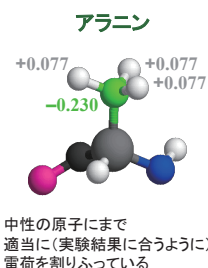
## 「力場の方程式」の問題

### 正しいのか？

この方程式は、  
水素結合も疎水結合も含まない！

物理学的な根拠が不明確

小さい分子にはOK、  
しかし、タンパク質に使えるのか？



## 「初期構造」の問題

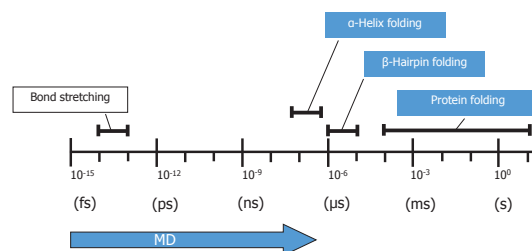
結晶構造、単粒子低温電顕、NMR

### ■ 初期構造を設定

実際には揺らぎがある。  
多くの初期構造で計算  
する必要

## 「計算」の問題

タンパク質の機能的な構造変化は 1 msec ( $10^{-3}$ ) 以上  
1兆回の計算



## 「計算」の問題

### 今まで

計算力不足で、1  $\mu$ s の計算もたいへん

さまざまな「粗視化」(乱暴な単純化)で  
計算量を 1000分の1 以下に

1. 計算結果が頼りにならない
2. 力場方程式が正しいのか、不明

ところが最近大きな進歩がありました。米国の計算科学者 David E. Shaw は、計算技術を使って株式で大富豪になり、その資金で私立研究所を設立し MD 専用計算機 (Anton と称す) を作り上げたのです。今までの巨大な汎用計算機は膨大な数の計算コアを並列させたものですが、コア間の通信にますます時間がかかるようになっていきます。そこで Shaw たちは CPU チップから専用のものを開発し無駄な計算をはぶき、今までよりも数桁速い計算速度を実現したのです。

## MD専用計算機の開発

David E. Shaw

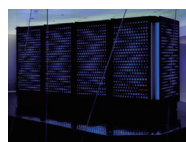
元コロンビア大学・計算機科学教授。  
モルガンスタンレー入社後、個人の投資会社を設立。資産2000億円以上。



富士通 90nm テクノロジー

普通のCPUだと無駄な計算が多いので、専用CPUにすれば速い。

並列計算機(「京」60万のユニット)は、並列が多くなるとユニット間の通信が速度制限



MD専用計算機 Anton

2010年 個人の研究所  
D.E. Shaw Research (DESRES) を設立  
MD 専用計算機 ANTON-1、-2 を開発

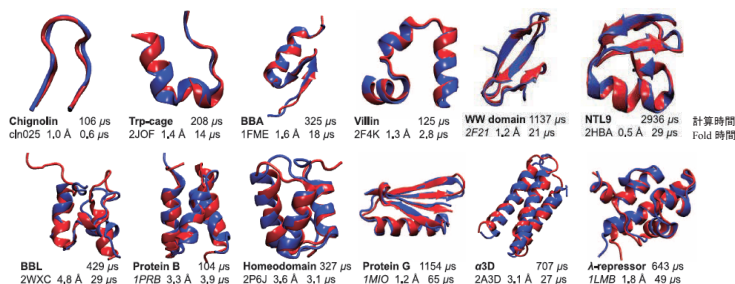
■ 2万原子の系で1日に10 $\mu$ sを計算。  
日本のスパコン「京」の100倍以上

開発は大きな賭けでした。というのは、力場の方程式が大きなタンパク質にも有効かどうか誰も知らなかったわけで、もし無効なら何の役にも立たない計算機を何百億円もかけて作ったことになりまますから。しかし、まず、ごく小さなタンパク質の構造を、アミノ酸配列だけの情報を与えて計算したところ、百発百中でかつてない正確さで予測できました。

## 小さなタンパク質の fast folding 計算の成功

赤(実験)、青(計算)

Science 2011, 334, 517

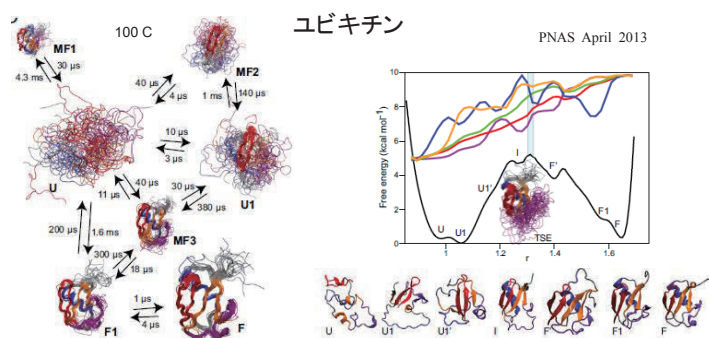


1. 計算した構造は、実際の構造とよく一致している。
2. Folding の経路は、1つ、あるいは2つ( $\beta$ -sheet のできる順番が違う)
3. Unfold でも、native-like の部分構造( $\alpha$ -helix が多い)が出現消滅している。  
native-like topology で folding が始まる。sequential stabilization 起きている
4. 律速の過程はない(大きなenergy barrier はない)

次に Shaw たちは本格的なタンパク質としてユビキチン (アミノ酸 76 個) の folding を計算しました。すると、自由に動くヒモ状の unfold の状態から天然の(native)立体構造まですべての過程が計算で出現しました。変性温度 (100 度) で計算すると、ユビキチンは unfold と native の間を行き来します。また、これで力場の方程式がタンパク質にも適用可能であることが証明されました。ついにアミノ酸配列から立体構造を計算できる時代が来たのです。

興味のある方のために以下の4枚の図で Anton の計算の威力をいくつか示します。

## 普通のタンパク質の slow folding 計算の成功



- native-like の部分構造が早くから出現
- Energy barrier (~5 kcal/mol)がある (= folding 中間体がある)
- **力場の仮定が 正当化された**
- folding は過去の履歴に依らない

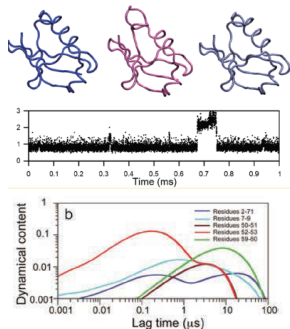


## 構造の揺らぎ(minor conformational state)の計算

"Picosecond to Millisecond Structural Dynamics in Human Ubiquitin" *J. Phys. Chem.* 2016

at 27 °C

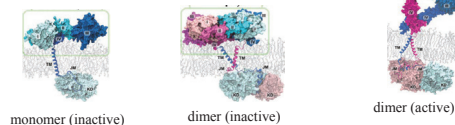
ユビキチン



各々の構造揺らぎは、いろいろな時間スケールで起きている

これらの揺らぎ構造は、他のタンパク質と複合体を形成した時のユビキチンの構造と一致。また、NMR構造とも合う。

"Architecture and Membrane Interactions of the EGF Receptor," *Cell*, 2013



"EGFR oligomerization organizes kinase-active dimers into competent signalling platforms" *Nature Communication*, 2016

膜の中の上皮細胞増殖因子EGFの受容体(EGFR, 995 アミノ酸)の、モノマー、ダイマー、テトラマー 形成のMDシミュレーションをおこなった。膜脂質も含む。

115万原子、40 μsec のsimulation

テトラマーになって初めて効率的な自己リン酸化が起こる。

ダイマーからテトラマーになるには、空のEGF結合表面を使う。EGF結合ダイマーとEGFなしダイマーが、テトラマーを作る。

EGFが増えれば、テトラマーは出来にくくなる。

EGFR (= HER family) の複雑なリガンド濃度依存性が、説明できた。

Tetramer-EGFRは、膜を凹ませる。膜厚も薄くなる

"Structural Basis for Modulation of a G-Protein-Coupled Receptor by Allosteric Drugs," *Nature*, 2013

M2 ムスカリン性アセチルコリン受容体(M2 受容体)のいくつかのアロステリック調節薬について結合部位、結合時の構造、薬剤-受容体相互作用を解明

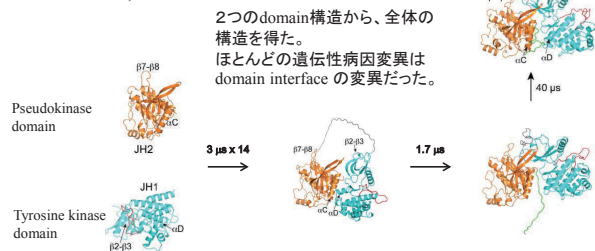
薬の構造がかなり多様であるにもかかわらず、全ての薬が、古典的で「オルソステリックな」リガンド結合部位から約15 Å 離れた、受容体の細胞外部分で、芳香族残基のクラスターとの間で陽イオン-π 相互作用を形成していた

このシミュレーションを基に、薬の親和性を増加、または減少させるように変異受容体を設計し、これらの変異体への放射性リガンド結合実験により、これを実証

シミュレーションにより、古典的なリガンド結合の正と負のアロステリック調節に寄与する機構が明らかになった。この機構には、2つの結合部位の共役したコンホメーション変化や、これらの部位におけるリガンド間の静電的相互作用が含まれる。

これらの観察から、調節薬のアロステリック効果を本質的に変える化学修飾の設計が可能となる。

"Molecular Basis for Pseudokinase-Dependent Autoinhibition of JAK2 Tyrosine Kinase," *Nature Struct. Mol. Biol.*, 2014

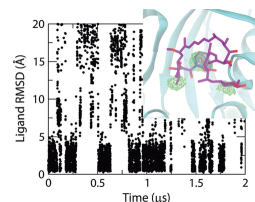


2つのdomain構造から、全体の構造を得た。ほとんどの遺伝性病変異はdomain interfaceの変異だった。

"Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations," *J Chem Theory Comput.* 2017

FKBPに対する低分子断片の結合を長時間MD。結合と解離を繰り返す、結合力と結合場所がわかった。

自由エネルギーの計算から求める結合力と一致した。その計算法の改良にも使える。



MD専用機で切り開かれたタンパク質の計算化学は今後どのような発展をするだろうか。まず、力場の方程式の改良が格段に進むでしょう。計算結果の良し悪しの評価が正確に出来れば、計算方法の改良は計算機自体にやらせることができます。例えば最近の計算将棋の進歩は、棋面の評価の改良によるところが大きいのです。さらなる方法の改良と計算機の開発により、タンパク質すべてのうち 30%くらいは計算で立体構造がわかるようになる、と Shaw は述べています。

## 分子動力学(MD)によるタンパク質研究の今後

### ANTON で力場の方程式の改良

例：計算ではfolding 最初期のポリペプチド鎖の凝集が過度に進む

$$F = \text{化学結合力} + \text{ファンデルワール力} + \text{静電的力}$$

引力ばかりで斥力がないから？(吉田)

評価関数があれば、改良は限界まで進む

### 計算力

Shaw said; 今でもPDB(protein data bank)の10%は、計算可能。10年以内にはこれは30%になるだろう。

1000万原子の長時間(msec)シミュレーションをめざして最高の研究者を集めている

Anton-1 → Anton-2 (x2 速度) → Anton-3 (x4 速度)

しかし、大きな問題があります。Anton の開発はすべて Shaw の私費によって行われました。得られた情報や計算機の使用は、外部に対して閉ざされています。科学の公開性はここにはありません。この分野は大きな実用的な需要があります。独占的な新薬の開発です。外部のアクセスを遮断する理由があるのです。科学界が今まで経験しなかった資本主義的な科学知識の独占です。私は、日本でも Anton 級の MD 専用機の開発が急務だと思います。

最後に、MD とまったく異なる「統計力学」的なアプローチについて紹介します。ここからは、元分子研で現豊田理化学研究所の平田文男さんの理論の紹介です。私も議論に加わって最近、一つの定式に表現されました。タンパク質の構造は水溶液中で水分子と相互作用しながらその時の平衡構造の付近でゆらいでいます。天然構造が非常にまれに極端にゆらいで完全変性してもまたもとの天然構造にもどることができます。すなわち、タンパク質の構造変化は弾性変形です。

弾性変形ならば線形代数が使える、そして平衡構造からのずれの大きさに比例した復元力が働く、その結果、タンパク質の構造ゆらぎはガウス分布を含む関数で表現できます。その関数に大きな摂動を与えれば大きなゆらぎの構造を知ることができます。さらに、溶液の条件が大きく変るときにはタンパク質の構造も大きく変わりますが、解析接続という方法で途中の構造を知ることができます。原理的には、アミノ酸配列から立体構造が計算できるといいます。しかも少ない計算量で。まだ開発中ですが、期待しています。

## 問題

ANTON 24台。外部に使わせない。売らない。

ビル・ゲイツ, ジェフ・ベゾス (Amazon) などの協力

米国の薬会社と 1.2 billion dollar (1000億円超) の契約

日本 理研で Shaw の 100 分の 1 (?) の予算で開発中

2014年に完成予定だったが……

新たに組織と予算を集中して日本でも ANTON 級の MD 専用機を

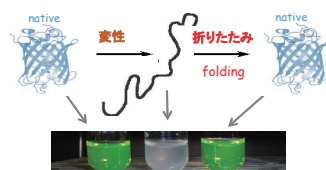
## タンパク質の自由エネルギー計算

### 変形 (ものの形が変わる)

塑性変形: 非可逆的 (破壊現象), 非線形, 地震

弾性変形: 可逆的 (もとに戻る), 線形の数学的記述, 力学的エネルギー 弾性; バネ, ボール  
エントロピー 弾性; ゴム

### タンパク質の構造は? 自由エネルギー 弾性変形



クラゲの蛍光タンパク質(GFP)

$y = f(x)$  は解明できる可能性

## 水中のタンパク質の構造ゆらぎは弾性変形であり、線形応答理論が使える

F. Hirata, M. Sugita, M. Yoshida, K. Akasaka.  
*J. Chem. Phys.* **145**, 234106 (2016)

タンパク質の構造のゆらぎ (変位  $\propto$  力、ばねのフックの法則と同じ)

溶媒の水分子も考慮したタンパク質の自由エネルギー曲面上での基準振動解析

$$\mathbf{R}_\alpha(t) = \mathbf{R}_\alpha(t = \infty) - \frac{1}{k_B T} \sum_\gamma \left\langle \Delta \mathbf{R}_\alpha(t) \Delta \mathbf{R}_\gamma \right\rangle_{eq}^{(0)} \cdot \mathbf{f}_\gamma(0)$$

平衡構造に摂動を与えてゆらぎ構造を計算

大きなゆらぎの構造を取り出すためには、大きな摂動をかける

数秒に 1 回出現する構造も、短い計算時間でわかる