

拡張アンサンブル法による生体分子系の 計算機シミュレーション

岡本 祐幸 物質理学専攻教授

仮想空間での実験とは

実験は実際の生体分子を直接観測するが、計算機シミュレーションは計算機上の仮想空間で生体分子を観測する。実験は複雑な系を扱うことができるが、計算機が扱える生体系には計算機の能力に限度があるので、扱える系は比較的小さなものになってしまう。たとえば、普通の実験では、 10^{-3} モルとか 10^{-6} モルとか(すなわち、 10^{21} 個とか 10^{18} 個とかの分子数)のサンプルを扱うが、計算機上では、世界最大のスーパーコンピュータをもってしても 10^8 個とか 10^9 個の原子数しか扱えない。また、実験では、

よく数時間とか数日間の現象を観測するが、計算機上では、ニュートンの運動方程式を差分方程式として解き、時間ステップを1フェムト秒(10^{-15} 秒)として、時間発展を追っていくので、膨大なステップ数の計算が必要になり、現在のスーパーコンピュータの計算能力ではマイクロ秒(10^{-6} 秒)とか最長でもミリ秒(10^{-3} 秒)ぐらいの間に起こる現象しか観測できない。

それでは、計算機シミュレーションは生体分子系のような多自由度複雑系には全く歯が立たないのかといえば、そうではなく、現在では、実験屋と計算屋の

共同研究が普通に行われるようになっているのである。計算機シミュレーションの強みは、各原子の位置と速度をフェムト秒の時間ステップで時々刻々追えることであり、仮想空間上ではあるが、時間および空間の超高解像度をもつ顕微鏡で観測した情報を提供することができることである。筆者らは、素粒子物理学や物性物理学の分野で開発された拡張アンサンブル法と総称される手法を生体分子系に導入するとともに、生体分子系に適した新手法を開発して、さまざまな問題に挑戦してきた。

従来の計算機シミュレーションは、実



Yuko Okamoto

1956年三重県尾鷲市須賀利町生まれ。1975年静岡県立浜松北高校卒業。1979年ブラウン大学卒業。1984年コーネル大学大学院修了、Ph.D.の学位を取得。1984-1986年バージニア工科大学博士研究員。1986年奈良女子大学理学部助手(1993年、助教授)。1995年分子科学研究所助教授(総合研究大学院大学助教授併任)を経て、2005年より現職。専門は生物物理、計算科学。

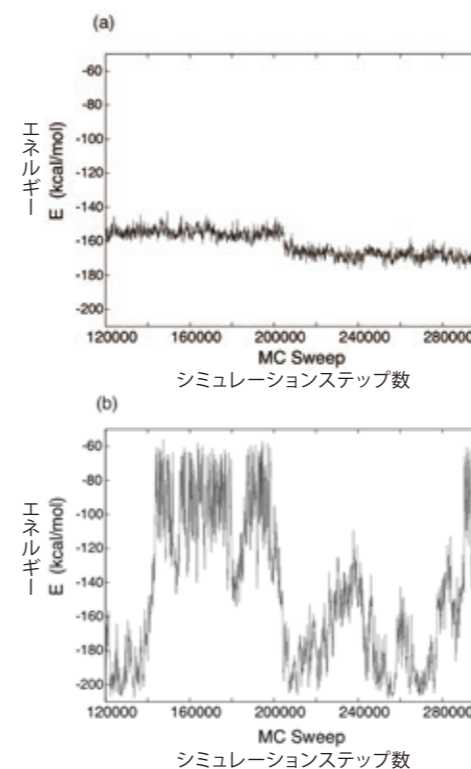


図1 計算機シミュレーション中のポテンシャルエネルギーの「時間発展」
リボヌクレアーゼT1というタンパク質の一部である17アミノ酸残基のペプチドフラグメントの結果。
(a)は温度 $T = 200\text{K}$ における従来のシミュレーション。
(b)は拡張アンサンブル法によるシミュレーション。従来のシミュレーションでは、エネルギーが約20の幅をゆらいでいるだけなのに対し、拡張アンサンブルシミュレーションでは、約140もの広い幅のエネルギーの酔歩が実現している。

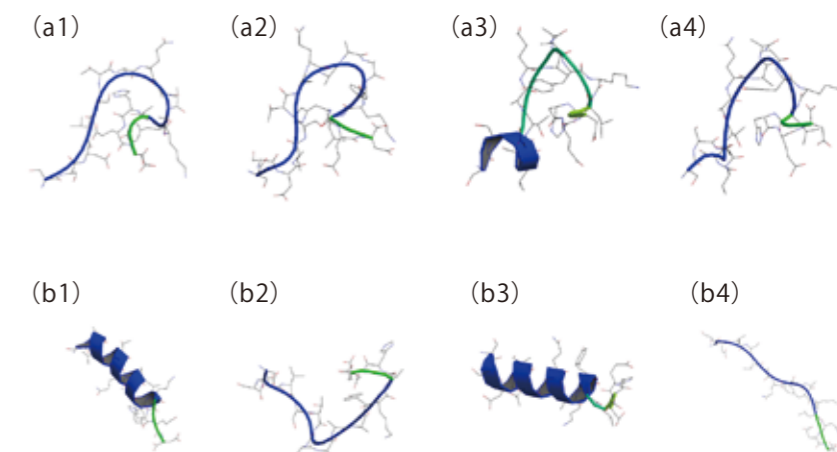


図2 計算機シミュレーション中の立体構造のスナップショット
リボヌクレアーゼT1というタンパク質の一部である17アミノ酸残基のペプチドフラグメントの結果。
(a1)-(a4)は図1(a)の従来のシミュレーションからのスナップショット。
(b1)-(b4)は図1(b)の拡張アンサンブルシミュレーションからのスナップショット。
対応するシミュレーションステップ数は(a1)と(b1)は13万8000、(a2)と(b2)は19万、(a3)と(b3)は24万3000、(a4)と(b4)は29万5000。

際の実験条件に合わせて、温度や圧力などの環境パラメータの値を固定して実行される。しかし、多自由度複雑系では、シミュレーション中に、無数に存在するエネルギー極小状態に系がとどまってしまいうために、誤った結果を出してしまう困難がある。言い換えれば、計算機で到達できるマイクロ秒とかミリ秒とかの計算時間は短すぎて、系が十分に熱平衡に到達できないということである。拡張アンサンブル法は温度や圧力などのランダムウォーク(酔歩)を実現することによって、系の熱平衡への到達を大幅に早める。ここで酔歩とは、ある変数の値がシミュレーション中に上がったたり下がったりを繰り返すことをいう。すなわち、シミュレーション中に温度を上げたり下げたりして、室温のシミュレーションでは一つの状態にとどまってしまうのに対し、温度が上がった時に、熱振動が増し、そこから脱出し、幅広い立体構造空間を効率良く探索するのを可能にするのである。拡張アンサンブル法の長所は一回のシミュレーション結果から、任意の温度や圧力などにおける熱力学量を精度良く計算できることである。

拡張アンサンブル法の原理

まず拡張アンサンブル法の原理を小ペプチドの例で説明する。リボヌクレアーゼT1というタンパク質の一部である17個のアミノ酸残基からなるペプチドフラグメントは、 α ヘリックスとよばれる螺旋構造をもっていることが実験によって知られている。しかし、従来の温度一定のシミュレーションを実行しても、 α ヘリックス構造はできない。それに対し、拡張アンサンブルシミュレーションを実行すれば、エネルギーが高い時にはコイル状態であるが、エネルギーが下がってくると α ヘリックス構造が形成される。図1にこれら2種類のシミュレーション中のエネルギーの時間発展を示す。従来のシミュレーションでは、エネルギーが約20の幅をゆらいでいるだけなのに対し、拡張アンサンブルシミュレーションでは、約140もの広い幅の酔歩が実現していることがわかる。そして、この2シミュレーションのシミュレーションステップ数が12万から30万に至るまでの間でそれぞれ四つの構造を抽出して図2に示した。

図2(a1)-(a4)の従来のシミュレーションでは、図2(a3)でペプチドの左端

が α ヘリックス構造をつくりかけているが、エネルギー極小状態にとどまってしまって、長いヘリックス構造ができないのに対し、拡張アンサンブルシミュレーションでは、図2(b1)、(b2)、(b3)、(b4)と時間発展するにつれて、 α ヘリックス構造とコイル状態の間を行き来して、幅広い立体構造空間をサンプルしていることがわかる。すなわち、従来のシミュレーションでは一つの状態にとどまってしまうのに対し、拡張アンサンブルシミュレーションでは、エネルギー(温度)が上がった時に熱振動が増すため、エネルギー極小状態から脱出することができ、幅広い立体構造空間の効率良い探索が可能になるのである。

レプリカ交換法の適用例

次に拡張アンサンブル法の一つであるレプリカ交換法の適用例を紹介する。レプリカ交換法では、系のコピー(レプリカ)と温度の値をそれぞれM個ずつ用意して、それぞれのレプリカに一つの温度値を対応させる。平行して独立な従来の温度一定シミュレーションを実行しながら、一定時間間隔ごとに隣り合う温度

対を交換していく。これによってすべてのレプリカにおいて温度空間の酔歩が実現されて、シミュレーションがエネルギー極小状態にとどまるのを避けることができる。

一つ目の例は、膜タンパク質の立体構造予測である。膜タンパク質は細胞膜などの脂質2重膜を貫通しているタンパク質であり、膜貫通部分は多くの場合、 α ヘリックス構造をとっている。タンパク質の動きの観測には前稿の高速原子間力顕微鏡が有力であるが、自然の立体構造はX線回折、核磁気共鳴(NMR)、電子顕微鏡などの実験で決定され、プロテインデータバンク(PDB)に登録されるが、現在、16万個を超えるデータが蓄積され、無料で公開され、他の研究者の研究に利用されている。ゲノム解析によると各生命体が合成するタンパク質の約4分1は膜タンパク質であるにも関わらず、膜タンパク質では膜も一緒に結晶化するのが難しいので、X線回折実験等で構造決定するのが困難で、PDBのデータはほとんどが水溶性タンパク質のものであり、膜タンパク質のものは全体の数パーセントにすぎない。

実験で構造決定が難しいなら、計算機シミュレーションでやろうというのが筆者らの考えであった。この問題に我々は大胆な近似を導入した。すなわち、膜の外側の水分子も膜を構成する脂質分子も省略し、膜貫通ヘリックスだけを考慮した。シミュレーションの操作としては、ヘリックスの平行移動と回転、側鎖の構造変化、そしてヘリックスのたわみの4種類である。図3にバクテリオロドプシンという7本のヘリックスからなる系のレプリカ交換シミュレーションの時間発展を示す。温度が高い時は空間的にも離れるとともにヘリックスが大きく歪むが、低温ではヘリックス同士が会合していることがわかる。図4にPDBに登録されて

いるX線回折実験で決定された構造と我々の予測結果を比較した。実は、7本のヘリックスの並び順ばかりでなく、各ヘリックスの理想形からの曲がり具合も良い一致が得られた。

二つ目の例は、抗体(免疫グロブリン、IgG)の糖鎖の構造ダイナミクスに関する、我々が独自に開発した、レプリカ交換分子動力学(REMD)法の適用例である。抗体タンパク質は、Fc γ 受容体(Fc γ IIIa)とよばれるタンパク質に結合することで機能を発揮する。IgGとFc γ IIIaの相互作用には糖鎖が関わっており、抗体の糖鎖からフコース残基を取り除くことによってガン細胞を殺す活性が大きくなる。我々は糖鎖を含むIgGとFc γ IIIaの複合体のうち、フコース残基を含む系と取り除いた系の二つの場合において、温度の酔歩を実現するREMDシミュレーションを実行した。ここでは、主に糖鎖の構造ゆらぎを調べた。図5に計算の初期構造を示す。その結果、X線回折実験の結果と同じように、フコース残基が付いた系ではFc γ

IIIaの糖鎖の構造が大きくゆらいでおり、IgGとFc γ IIIaの糖鎖同士の接触が失われていることが示された(参照動画)。この計算結果により、抗体の糖鎖からフコース残基を取り除くことにより、抗体の制ガン作用が促進される仕組みの理解を深められた。また、本計算の副産物として、構造ゆらぎが大きく、X線回折実験では決定できなかった糖鎖の一部の立体構造を決めることもできた。

期待される創薬設計への利用

本稿では、筆者らの生体分子系の拡張アンサンブルシミュレーションの結果を紹介した。はじめに述べたように、これらは仮想空間上のものであり、実験結果を再現できない場合も多い。すなわち、計算機上で扱う系のエネルギー関数が正確でなければ、計算結果は信用できないということである。現在の計算は古典力学に基づいているが、エネルギー関数の精度を上げるためには、より精度が高い量子力学に基づく計算を実行する必要がある。現在のスーパーコン

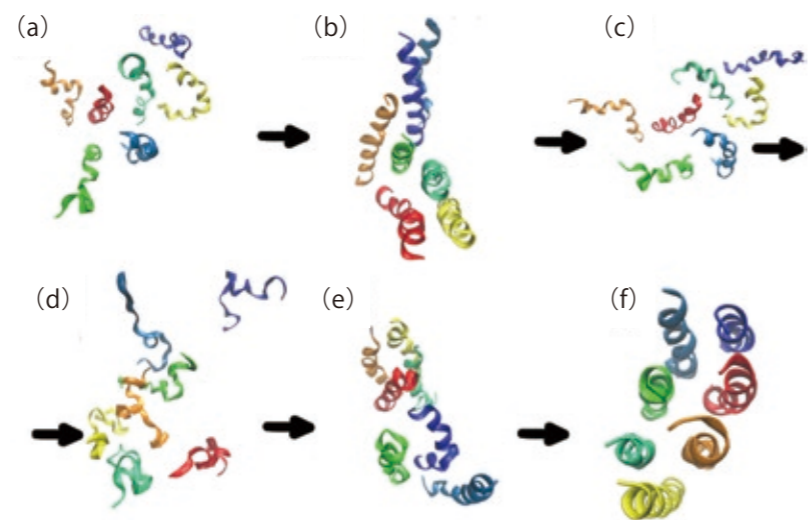


図3 レプリカ交換モンテカルロシミュレーション中の立体構造のスナップショット
バクテリオロドプシンの7本の膜貫通ヘリックスのレプリカ交換モンテカルロシミュレーションからの立体構造のスナップショットの時間発展。温度が高い時(a)、(c)、(d)は空間的にも離れるとともにヘリックスが大きく歪むが、低温(b)、(e)、(f)ではヘリックス同士が会合している。

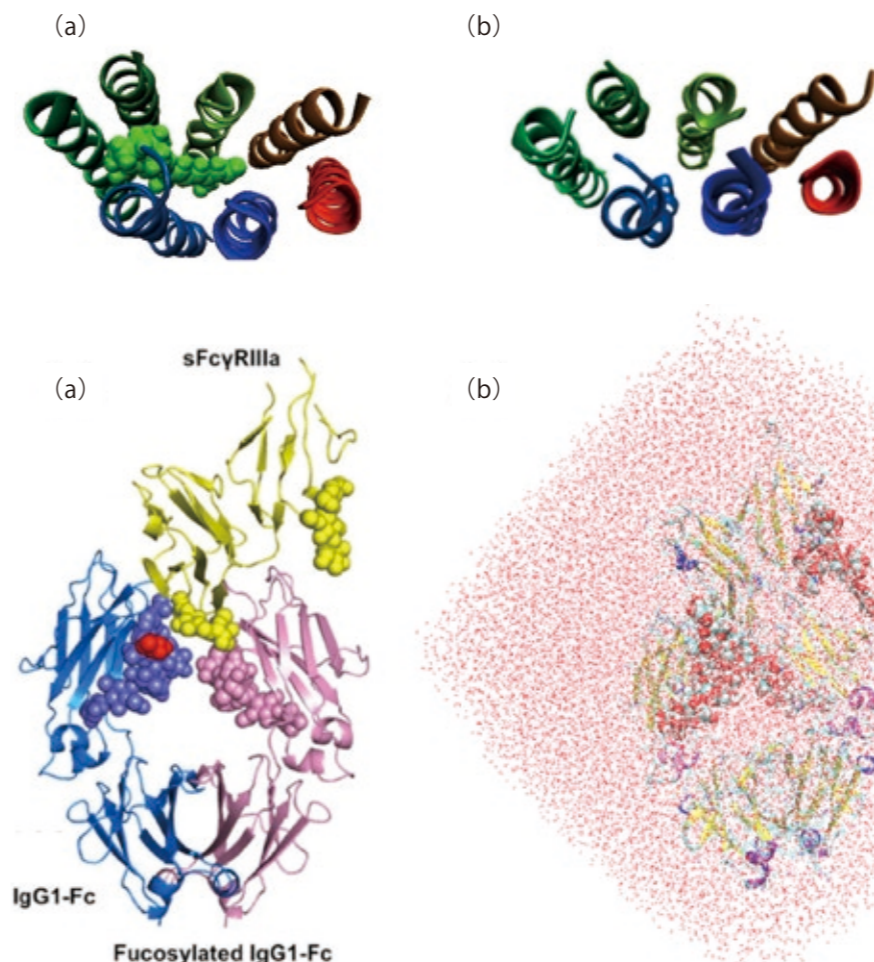


図4 実験結果と計算結果の比較
(a)はX線回折実験で決定された、バクテリオロドプシンの7本の膜貫通ヘリックスの自然の構造(PDB code: 1PY6)。レチナル分子も空間充填図で示した。
(b)はレプリカ交換モンテカルロシミュレーションで予測された構造。両者において、7本のヘリックスの並び順ばかりでなく、各ヘリックスの理想形からの曲がり具合も良い一致が得られた。

図5 レプリカ交換分子動力学シミュレーションの初期状態
二つのタンパク質IgGとFc γ IIIaとそれらの糖鎖のレプリカ交換分子動力学シミュレーションの初期状態。
(a)はPDBデータから構築したフコース残基(赤色)を含む系。タンパク質の部分(リボン図で青色と紫色はIgG、黄色はFc γ IIIa)は583個のアミノ酸からなる。糖鎖は球の空間充填図で表した。水分子は省略した。
(b)は水分子も表示した図。水分子の数は約9000個であり、系の全原子数は、フコースを含む系と含まない系で、それぞれ、9万6323個と9万6797個であった。

ピュータの能力がまだ低すぎるのである。ただし、計算機技術の進歩は絶え間なく、これまで、大体、30年間で約100万倍の計算能力の向上がずっと得られてきた。よって、正確なエネルギー関数の計算はいずれ可能になるであろう。精度の高いエネルギー関数と幅広い立体構造空間を探索できる拡張アンサンブル法を合わせることによって、計算機上で百発百中の実験結果の再現が可能となると期待される。

本稿では、拡張アンサンブル法の最も基本的な例として、温度空間の酔歩を引き起こす手法である、レプリカ交換法による結果を紹介した。我々は拡張アンサンブル法を他の変数空間の酔歩を実現する手法へと一般化してきた。たとえ

ば、圧力(と体積)空間の酔歩を実現する拡張アンサンブル法の開発によって、タンパク質の高圧変性のメカニズムを解明することに成功した。また、薬剤候補分子とタンパク質の距離の酔歩を実現する拡張アンサンブル法を開発することによって、どの薬剤候補分子が標的とするタンパク質に結合しやすいかを予測できるとともに、その結合自由エネルギーを高い精度で計算できるようになった。これは、製薬会社の創薬設計に大きな革命を引き起こすと期待している。すなわち、創薬においては多くの薬剤候補分子が標的とするタンパク質に結合するかどうかを実験で決定する必要がある(スクリーニングとよばれる)。これには長い時間と膨大な費用がかかる。これ

を計算機シミュレーションによってやっつけてしまえば、創薬設計にかかる時間と費用を大幅に削減できる訳である。

実験屋と計算屋が緊密な共同研究を繰り返すことによって、生物物理学の分野の将来の発展はめざましいものになるであろう。

参照動画
フコース残基を含まない場合と含む場合の糖鎖、IgG、Fc γ IIIaの複合体のREMDシミュレーション
<https://www.nature.com/articles/s41598-017-13845-8>
中のElectronic supplementary materialの二つの動画を参照。
フコース残基を含まない糖鎖、IgG、Fc γ IIIaの複合体の場合 (Video S1)
フコース残基を含む糖鎖、IgG、Fc γ IIIaの複合体の場合 (Video S2)