

# 1 分子時系列情報から解読される生体分子系の状態空間構造

神戸大学理学部地球惑星科学科(非線形科学分野)

小松崎民樹

tamiki@kobe-u.ac.jp

現在、報告されているほとんどの生体分子系のキネティクスは、多様な時間・空間スケールに渡り複雑に絡み合った個々の分子ダイナミクスの集団平均として観測されている。近年、1 分子計測技術等の飛躍的な進展により、「観測」の在り方が大きな変貌を遂げようとしている。すなわち、サブミリ秒程度の時間分解能で、1 分子レベルの大規模構造変形や細胞の分化の計時変化を直接観測することが可能になってきた。たとえば、Steven Chuらは他のリボザイムの特定部位を切断する機能を保有するヘアピンリボザイムが 4 種のDocked状態と 1 種のUndocked状態のあいだを構造転移するプロセスにおいて、確率的には等しい確率で 4 通りの経路がありえる状況下、リボザイムはひとつの経路を数時間に渡り遍歴するといった“分子記憶”を保有していることを初めて明らかにしている[1]。このようなメソスケールで生起する“動的”現象に対してどのようなアプローチがあり得るのであろうか？ 計算機のなかでこれらの現象を再現しようとするのであれば、大胆な近似を導入しない限り数時間に及ぶ動的な振る舞いを追うのは現実問題不可能であるし、仮になんらかの粗視化されたシミュレーションが実現できたとしても、1 分子計測実験と対比して実験結果では観測に掛からない自然現象をどれくらい正しく”模写”できたか判断しえるのであろうか？ このような現象の解明には従来の理論・手法とは異なる新しい研究手法が必要となる。

我々はなんらかの粗視化・近似をアプリオリに仮定した動力学シミュレーションを開発する代わりに「系を構成するハミルトニアンかつ準安定状態の数および準安定状態間の計量関係が未知の状況下、唯一、1 分子レベルの時系列情報があるとき、如何にして状態を規定し背後に存在するエネルギー地形・状態空間を推定できるであろうか？」を問う立場をとる。仮にマイクロ秒以下の高時間分解能で 1 分子を追跡することが可能になるとすると、どのような新現象が発見され、更にはそれに伴ってどのような指導原理が新たに創出されるであろうか？ 不連続な“階段”としてのみ観測されてきた過渡的な遷移過程の直接観測が可能になるであろう。その場合には、遷移の非統計性 [2]が実験的に論じることが可能になるかもしれない。また、階層的なエネルギー地形 [3]の存在、階層間の非隷属性、更には、「自由エネルギー地形」概念の妥当性の再考、など興味深い研究対象が現れることが期待される。また、観測する時間分解能をマイクロ秒以下に上げることは、従来の時間分解能に比べて熱揺らぎの時間スケールに近づく。マイクロ秒の時間領域で起こる蛋白質ダイナミクスと熱揺らぎの拮抗関係、時間・空間スケールの異なる階層間の情報伝達・相関性など様々な魅力的なテーマが待ち構えている。1 分子時系列から構成される記憶地形(Memory Landscape)[4]などの時系列解析手法が提案されているが、自由エネルギー地形ならびに状態遷移ネットワーク構造との関係が不明瞭であり、上記の展望に応えるだけの解析理論は現時点では未開拓状態といえる。

本分科会では、現在、我々が開発している局所平衡仮定に依拠しない計算力学(Computational Mechanics)[5]に基づく時系列解析理論を紹介し、時系列情報を通して観測されるダイナミクス描像を概観し、本解析理論のもつ将来性を展望する。通常、観測量(例えば、FRET 効率)から評価したい物理量(例えば、色素分子間距離)に変換する必要がある(例えば、フィッシャー情報基準を用いた文献[6]などを参照していただきたい)が、ここでは、変換された後の 1 分子時系列情報から、背後に潜む生体分子系の動態構造を抽出するための方法論に焦点を絞って解説する。また、1 分子観察(および、計算機実験)時系列データへ実際の適用した例は講義のなかで紹介する予定である。

## 目次

1. 情報理論入門 (分科会では説明しない)
2. 1 分子時系列情報から定義する Causal State(因果状態)
3. 状態間遷移ネットワーク( $\epsilon$ マシンの)の構成とネットワークの複雑さ解析
4. Leu-Enkephalin のターン形成ダイナミクスにおける時系列情報から構成される状態空間構造
5. 展望

# 1. 情報理論入門

紙面の都合上、情報理論に関しては必要最低限の概念を紹介するに留める(幾つかの良い参考書[7,8,9]が出版されているので、参照されたい)。

## 1.1 Shannon エントロピー

$n$ 個の実現可能な事象が確率 $p_1, p_2, \dots, p_n$ で生起するとする。どの事象が生起するか否かに関して知りえる全情報は確率 $\{p_i\}$ のみとする。このとき、確率事象 $\{p_i\}$ に応じて、出力に対する“不確実さ”に差異が存在しえる。例えば、すべての $p_i$ が $1/n$ である事象と $p_1 \approx 1, p_{i \neq 1} \approx 0$ である事象を比べた場合、仮にある出力を得られたとしても、前者はすべての事象が等確率で統計的に生起するので不確実さがより大きいであろう。さて、与えられた確率事象 $\{p_i\}$ に対する出力に関してどれくらい平均として“不確実であるか”を評価・定量する指標はないであろうか？ この問いに対しShannonは3つの公理(Khinchin公理と呼ばれる)を設定し、その3つの公理を同時に満たす物理量 $H$ が一義的に決定されることを示した[8]。

①  $H$ は $p_1, p_2, \dots, p_n$ に関する連続関数である

② すべての $p_i$ が $1/n$ である場合には、 $H$ は $n$ に関して単調増加関数である

$$\textcircled{3} H(p_1, p_2, \dots, p_n) = H(\omega_1, \omega_2, \dots, \omega_m) + \sum_{j=1}^m \omega_j H(\{p_i / \omega_j\}_j), \quad \omega_j = \sum_{i=n_{j-1}+1}^{n_j} p_i \quad (1)$$

ここで、 $\{p_i / \omega_j\}_j$ は $\omega_j$ を構成する $p_i$ を意味する。分かりづらい表記であるが、以下の例を考えればよい。図1(左)は確率 $p_1=1/2, p_2=1/3, p_3=1/6$ で生起する3つの事象を表している。この確率事象に対して $H(1/2, 1/3, 1/6)$ が付与されることになる。図1(右)は、同じ確率事象を対象にするが、選択手順を変えている。すなわち、まず確率 $\omega_1=p_1=1/2, \omega_2=p_2+p_3=1/2$ をもつ2つの選択可能な“ルート”を考える。もし、 $\omega_2$ の“ルート”を選択した場合には、更に確率 $p'_2=2/3, p'_3=1/3$ をもつ2つの選択を行う。この場合、3つの異なる事象が最終的に生起する確率は図1(左)と等しい。Shannonは、物理量 $H$ は確率分布そのものの関数であるべきで、確率事象の選択手続きに依るべきではないと考えて、図1(左)と図1(右)のあいだに

$$H\left(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}\right) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{2}{3}, \frac{1}{3}\right) + \frac{1}{2} H(1) = H\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{2} H\left(\frac{2}{3}, \frac{1}{3}\right)$$

の関係があるという条件を加えた。ここで、不確実性が存在しない事象1つの場合の $H(1)$ は0となる。

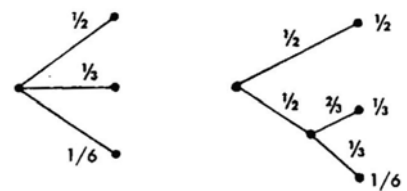


図 1

Shannon は Khinchin 公理を満足する  $H$  関数は一義的に決定され、

$$H(\mathbf{p}) = -K \sum_{i=1}^N p_i \log p_i \quad \sum_{i=1}^N p_i = 1 \quad (2)$$

で表されることを示した( $K$ は任意の正の定数)[8]。通常、 $K=1$ で $\log$ の底は2をとり、そのときの単位をbits(ビット)と表す。物理量 $H$ は一般にShannonエントロピーと呼ばれている。本稿ではShannonのオリジナルの導出方法[8]を紹介する。

まず、 $A(n) \equiv H(1/n, 1/n, \dots, 1/n)$ とおき、 $A(n)$ が $n$ についてどういう関数になるかを考える。 $A(n)$ は $H$ 関数の等確率バージョンであり、条件③を満足することが前提となる。いま、 $s^m$ 個の等確率で生起する事象を $s$ 個の等確率で生起する $m$ 事象に条件③を用いて分割する( $s, m$ :正の整数)。

$$A(s^m) = H(1/s^m, 1/s^m, \dots, 1/s^m) = mH(1/s, 1/s, \dots, 1/s) = mA(s) \quad (3)$$

となる( $s=2, m=3$ などの場合で条件③を展開してみれば理解しやすい)。同様にある正の整数 $t, n$ に対しても、

$$A(t^n) = H(1/t^n, 1/t^n, \dots, 1/t^n) = nH(1/t, 1/t, \dots, 1/t) = nA(t) \quad (4)$$

我々は $n$ を任意に大きくとることができ、かつ、 $s^m \leq t^n < s^{m+1}$ を満たす $m$ を見つけることができる。両辺の $\log$ をとり、

$n \log s$ で割ると、

$$\frac{m}{n} \leq \frac{\log t}{\log s} < \frac{m}{n} + \frac{1}{n} \quad \text{もしくは} \quad 0 \leq \frac{\log t}{\log s} - \frac{m}{n} < \varepsilon \quad (\varepsilon \text{は任意に小さい数}) \quad (5)$$

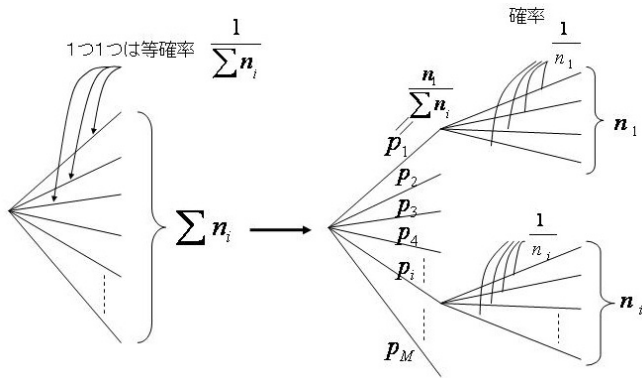
一方、 $A(n)$ の $n$ に対する単調増加性(条件②)より、 $A(s^m) \leq A(t) < A(s^{m+1})$ 、すなわち(3)(4)式を用いて  $mA(s) \leq nA(t) < (m+1)A(s)$ が満足される。両辺を $nA(s)$ で割ると、

$$\frac{m}{n} \leq \frac{A(t)}{A(s)} < \frac{m}{n} + \frac{1}{n} \quad \text{もしくは} \quad 0 \leq \frac{A(t)}{A(s)} - \frac{m}{n} < \varepsilon \quad (\varepsilon \text{は任意に小さい数}) \quad (6)$$

が成り立つ。(5)(6)式を合わせて、

$$0 \leq \frac{A(t)}{A(s)} - \frac{\log t}{\log s} < \varepsilon \quad (7)$$

ここで、 $\varepsilon$ は任意に小さい値を取れるので、 $A(s) = K \log s$  ( $K > 0$ )が成立する。ここで、単調増加性(条件②)を保証するため、 $K$ は正の定数となる。



さて、等確率 $1/N$ で生起する $N$ 事象から成る確率分布を考える。ここで、 $N = \sum_i^M n_i$ とする。 $N$ を $M$ 個に分割し、最初に $M$ 個に分割した各集合を確率  $p_j = n_j / \sum_i^M n_i$  で選択し、その後、確率 $p_j$ で選択された $j$ 番目の集合に対して、 $1/n_j$ の等確率で個々の事象を選択することを考える(左図参照)。

このとき、条件③より、

$$K \log \sum_i n_i = H(p_1, p_2, \dots, p_n) + K \sum_i p_i \log n_i \quad (8)$$

すなわち、

$$H(p_1, p_2, \dots, p_n) = K \log \sum_i n_i - K \sum_i p_i \log n_i = -(-K \sum_i p_i \log \sum_i n_i + K \sum_i p_i \log n_i) = -K \sum_i p_i \log p_i \quad (9)$$

となり、(2)式が導出される。以上が Shannon のオリジナルの導出方法である。

Shannonエントロピー $H(\mathbf{p})$ は $-\log p_i$ の期待値であることに注意されたい。 $-\log p_i$ はShannon情報量 $I(p_i)$ と呼ばれ、

- ① 意外な情報(確率の低い情報)ほど情報量が多い( $p$ に対して $I(p)$ は単調減少関数)
- ② 情報を一度に受けても、分割して受けても、最終的に得られる情報量は同じ(すなわち、 $I(pq) = I(p) + I(q)$ )
- ③ 情報量は非負である(すなわち、 $I(p) \geq 0$ )

という3つの公理から一義的に導かれる。Shannon情報量は端的に言うと、確率 $p_i$ の事象がもつ不確かさ(換言すると、情報の「量」)を評価するもので、Shannonエントロピーは平均の不確かさ(情報量)と解釈することができる。

すべての事象が等確率で生起する場合のShannonエントロピー $H(\mathbf{p})$ はBoltzmannの熱力学的エントロピー $S(E)$ と形式的に等価である。 $S(E)$ は

$$S(E) = \log N(E) \quad (10)$$

( $N(E)$ :エネルギー $E$ における到達可能な微視的状態数)で表される。等しいエネルギー $E$ における微視的な状態はすべて等重率に存在する場合、微視的状態の出現確率 $p_i$ は、みな $p_i = 1/N(E)$ となる。 $p_i = 1/N(E)$ を(9)式に代入することによって、(10)式を得ることが分かる。

**問題1** 情報量  $I(p)$ が上記の3つの条件から導出されることを示しなさい。

**問題2** Shannonエントロピー  $H[p]$ はすべての  $p_i$ が  $1/n$ である場合、 $(p$ に関して)最大値を取ること、 $H[p]$ は非負の値をとる( $H[p] \geq 0$ )ことを示しなさい。

### 1.2 結合エントロピーと条件付きエントロピー

いま、2つの確率変数  $X, Y$  を考える。 $x$  および  $y$  は  $X$  および  $Y$  の取りうる任意の値を表すものとする。このとき、 $X=x$  および  $Y=y$  を同時に満足する結合確率を  $p(x, y)$  とすると、 $X, Y$  の同時分布に対する Shannon エントロピーは

$$H(X, Y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x, y)$$

と表される。 $H(X, Y)$  を結合エントロピー (Joint entropy) と呼び、 $X, Y$  の結合確率分布のもつ平均の不確かさに対応する。また、 $Y=y$  のときの  $X=x$  についての条件つき確率を  $p(x|y)$  とするとき、 $Y=y$  のときの  $X$  の条件つき分布のエントロピー  $H(X|y)$  は

$$H(X|y) = - \sum_{x \in X} p(x|y) \log_2 p(x|y) \quad p(x, y) = p(x|y)p(y) = p(y|x)p(x)$$

と書ける。 $H(X|y)$  を  $y$  に関して平均をとったものは、

$$H(X|Y) = \sum_{y \in Y} p(y) H(X|y) = - \sum_{x \in X} \sum_{y \in Y} p(x, y) \log_2 p(x|y)$$

と表される。 $H(X|Y)$  を  $Y$  が知らされたときの  $X$  の条件つきエントロピー (Conditional Entropy) とよび、 $Y$  を知った後に (平均的に) 残る  $X$  の不確かさに対応する。

**問題3** 結合エントロピーと条件つきエントロピーのあいだに

$$H(X, Y) = H(X) + H(X|Y) = H(Y) + H(Y|X)$$

$$H(X_1, X_2, \dots, X_n) = H(X_1) + H(X_2|X_1) + H(X_3|X_1, X_2) + \dots + H(X_n|X_1, X_2, \dots, X_{n-1})$$

が成り立つことを示しなさい。ただし、 $H(X) = - \sum_{x \in X} p(x) \log_2 p(x)$  である。

### 1.3 相互情報量

$X$  が何かという不確かさは  $H(X)$  で与えられる。 $Y$  を知るによって、 $H(X|Y)$  まで不確かさは減少する。換言すると、その差

$$I(X, Y) = H(X) - H(X|Y)$$

は  $Y$  を知るによってもたらされた  $X$  に関する平均の情報量ということが出来る。また、

$$I(Y, X) = H(Y) - H(Y|X) = I(X, Y)$$

も成り立ち、 $Y$  を知るによってもたらされた  $X$  に関する平均の情報量は  $X$  を知ることによってもたらされた  $Y$  に関する平均の情報量と等しい。 $I(X, Y)$  を相互情報量 (Mutual Information) と呼ぶ。

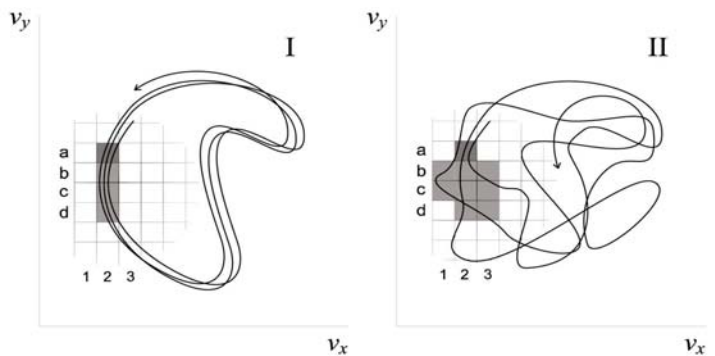


図2 2次元時系列データ  $(v_x(t), v_y(t))$  のシンボル化

## 2. 1 分子時系列情報から定義する Causal State (因果状態)

まず、1分子時系列からシンボル化された時系列  $X(i)$  を作成する。図2に2次元時系列の場合のシンボル化を例示している。

Iでは  $\{a2, b2, c2, d2\}$ 、IIでは  $\{a2, b1, c1, d2\}$ 、 $\{a2, b2, c2, d2\}$  および  $\{a2, b2, b3, c3\}$  というシンボル列を生成する。さて、シンボル列に変換された時系列情報からどのように、状態を規定したらよいであろうか？





## 統計的複雑さ

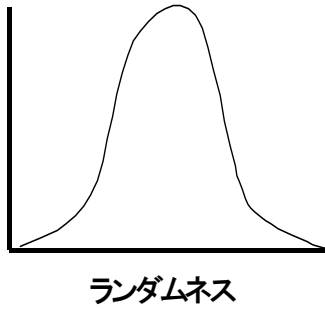


図7 統計的複雑さとランダムネスの定性的な関係

時系列情報の背後に存在する生体分子の動態構造を正しく理解するためには、シンボル時系列のランダムネスに相当する指標(例えば、自己相関関数、リヤプノフ指数解析、平均相互情報量、測度エントロピー(metric entropy)[5])を複雑さ解析と相補に評価することが必要不可欠となる。

数学的に、 $\epsilon$ マシンは定常なシンボル時系列を統計的に再現・予測する最適かつ唯一・最小モデルであることが証明されている[5]。紙面の関係上、原著に証明は委ねるが、

$$\textcircled{1} H[\text{Future} | \text{any partition of history}] \geq H[\text{Future} | \text{Casual States}]$$

Causal State を知ったうえで、なお残る未来時系列に対する条件つきエントロピー(不確かさ)は、過去のシンボル時系列を任意に分割して得られる他の情報を知ったうえで、なお残る未来時系列に対するあいまいさよりも

小さい。→最適な予測モデル

$$\textcircled{2} \text{統計的複雑さ(other machines)} \geq \text{統計的複雑さ}(\epsilon\text{-machine})$$

$\epsilon$ マシンの統計的複雑さは、他のマシンのそれよりも小さい。→唯一・最小モデル

**問題4** これらの3つの例において、シンボル列の長さを2, 3個にしたときの $\epsilon$ マシンを論じ、すべての例において Topological Complexity および Statistical Complexity が不変であることを確認しなさい。

**問題5** 変性状態と天然状態のあいだを2状態転移する時系列を想定し、自己相関関数から評価される特徴的な(“記憶”)時間スケールが $\tau$ であるものとする。時間きざみ  $t > \tau$ で、記録された時系列が作り出す $\epsilon$ マシンがどのようなネットワークになるかを論じなさい。また、時間きざみ  $t \approx \tau$ では、 $\epsilon$ マシンがどう変わるかを予想しなさい。

**問題6** 実際の1分子時系列データは有限の長さを持っている。例1のコイン投げを例にして、 $\epsilon$ マシンにおける時系列長の有限性を論じなさい。

## 4. Leu-Enkephalin のターン形成ダイナミクスにおける時系列情報から構成される状態空間構造

1分子時系列情報から(状態を規定する時間長の関数として) $\epsilon$ マシンを局所的に構成し、 $\epsilon$ マシンの状態遷移ネットワークの時間変化を追跡することを通して、生体分子ダイナミクスの動態構造を解析することができる。ここでは、 $\epsilon$ マシンを生体分子系ダイナミクスに最初に応用した英国・Cambridge 大のグループによる Leu-Enkephalin の全原子等温 MD シミュレーションに関する原子速度時系列解析の結果[10]を紹介する。

彼らは等温 MD 計算を温度 300K、水分子 5826 個、周期境界条件の下、3ns 実行し、図8に示すような $\beta$ ターン形成過程における状態遷移ネットワークの計時変化を解析した。1分子計測実験データを解析する立場では、得られた1分子時系列データそのものと対峙し、“その”時系列情報から如何に背後に存在する動態構造を解読できるかを問う。しかしながら、本来、系のどういった時系列情報に着眼するかによって、実際に見えてくる状態空間構造は異なる。また、一般にシンボル化が細かくなるほど、着眼するダイナミクスの時間スケールは小さくなるが、細かすぎると、時系列の長さが有限であることから、シンボル列の統計が十分取れなくなるという問題も発生する。

彼らは原子の速度ベクトルが作る多次元時系列情報から $\beta$ ターン形成における状態遷移ネットワークの計時変化を論じ、原子 $k$ の極座標( $r_k, \phi_k, \theta_k$ )において、 $r_k, \phi_k$ および $\theta_k$ を $m, n$ および $n/2$ 個の異なる領域(sector)に分割し、各sectorを番号づけし、どのsectorに属するかに応じて、そのときの速度の局所的な期待値から(高次元の)シンボル化を行った。彼らは $n$ として8から16を、 $m$ として1から2を用い、 $m, n$ に応じて、 $\beta$ ターン形成に沿ってどのように局所的な状態遷移ネットワークが変化するかを調べた。その結果、特徴的な時間スケールが相対的に小さい水分子の原子速度時系列情報に関しては、 $\beta$ ターン形成に沿って有意な変化を観測するためには、より細かいシンボル化が必要であると帰結した。Leu-Enkephalinのターン形成時、前後において得られた代表的な $\epsilon$ マシンの計時変化を図9に示す。図より明らかなように、ターン形成時近傍における $\epsilon$ マシンが相対的に“単純化”していることが分かる。

図10に Leu-Enkephalin および Leu-Enkephalin 近傍の水分子群の(平均化された)Topological

Complexity の計時変化を示す。ターン形成時近傍で Leu-Enkephalin および Leu-Enkephalin 近傍の水分子群とも、状態遷移ネットワークの複雑さが減少していることがわかる。水中における Zwitterion の構造転移反応においても、彼らは「遷移過程における複雑さの減少」を観測している。

エネルギー地形などが詳細に調べられている 46 ビーズたんぱく質モデルの折れ畳み温度におけるポテンシャルエネルギー揺らぎ等の時系列データに対しても、折れ畳みの前後に比べて、折れ畳み時にはεマシンの統計的複雑さが顕著に減少することが示唆されている [12]。

これらの「遷移過程における複雑さの減少」の発見は、気相[2]および凝縮相[13]化学反応において、遷移状態を越えてゆく運動が遷移状態の前後に比べて、ダイナミクスの記憶を持った規則的過程でありえることを原理的に証明した近年の我々の基礎研究[2]を具現化した一例として捉えることができる。

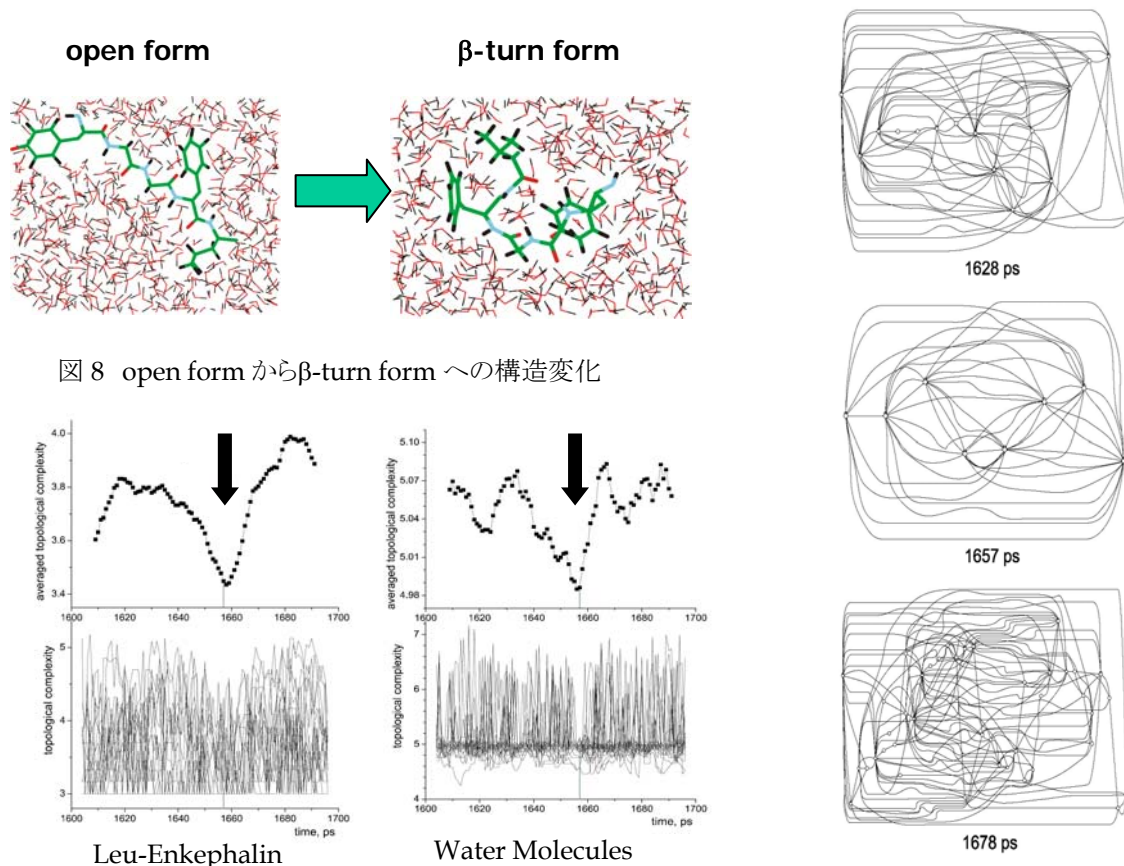


図 8 open form からβ-turn form への構造変化

図 10 Topological Complexity の計時変化(图中、矢印はβターン形成時を指し示している)

上:平均化された Topological Complexity

下:Topological Complexity

左:Leu-Enkephalin (総シンボル数 8、シンボル長 3ps)

右:Leu-Enkephalin 近傍の水分子群 (総シンボル数 32、シンボル長 4ps)

図 9 βターン形成時(1657ps)とその前後における原子速度時系列から構成された代表的なεマシンの各ノードは Causal State を、弧は Causal State 間遷移を意味する

## 5. 展望

自然科学研究において革命的な発展をもたらすのは、多くの場合、新しい実験技術とその新しい実験事実に基づいた理論・概念の転回である。近年、1 分子計測技術等の飛躍的な進展により、「観測」の在り方が大きな変貌を遂げ、1分子レベルでの生体分子の大規模構造転移、細胞の分化などが直接観測されるようになった。しかしながら“時系列情報から如何にしてその背後に潜むメカニズムに関する知見を抽出しえるのか”は未踏の未解決問題である。本稿で紹介した計算力学 (Computational Mechanics) は埋め込み理論[14]の適用



が困難な短い 1 分子時系列情報に対しても、その集合を用いることで、局所平衡を予め仮定しないで、状態および状態間遷移を規定することができる“第一原理的”時系列解析理論である。このとき、分子記憶の原理は

- ① Causal State の内部構造: マルコフ過程とみなしえるためにはどれくらい長いシンボル列(どれくらい長い記憶を保有しえるかに相当する)を必要とするか等
- ②  $\epsilon$ マシンにおける Causal State 間の“つながり”の様相: つながりに規則的なパターンが存在するかどうか等

から探求することができる。例えば、運動の記憶を持たない時間スケールでは、一般的な自由エネルギー地形から類推される状態遷移ネットワーク構造が得られる。しかしながら、記憶を持つ場合には、ダイナミクスの情報を捉えた状態遷移ネットワーク構造に取ってかわることになる。また、1 分子時系列をある長さで断片化することで時間依存の $\epsilon$ マシンを構成することができる: 例えば、天然状態、変性状態、その間の遷移過程における局所的な動態構造とその変化過程を抽出することができる[10]。更には、(今回、紙面の都合上、割愛したが)1 分子時系列データを Wavelet 多重分割で異なる時間スケールのダイナミクスに分解することができる。そこでは、階層毎の $\epsilon$ マシンを構成し、異なる階層における異なる状態遷移ネットワーク構造ならびに階層間の相互情報を調べることができる。また、統計的複雑さ(Statistical complexity)と力学系理論における $(\epsilon, \tau)$ エントロピーのあいだに類似性があることが分かっており[12]、情報理論に立脚して開発された計算力学の手法を力学系理論の概念で解釈できる可能性がある。

我々は、最近、状態間遷移のサンプルが数個しか取れない“短時間の”1 分子時系列の集団から背後に存在する多次元自由エネルギー地形を抽出する解析手法も開発した[15]。この方法では、第ゼロ近似としてさまざまなエネルギー障壁の高低に依らずにすべて平衡統計に従うものと仮定し、観測される物理量に対する各準安定状態の分布関数(一般に非正規分布)を自己無撞着に推定する。各準安定状態を時系列情報から順に抽出した後、各々の滞在確率から対応する準安定状態の自由エネルギーを、また、状態間の遷移確率から対応する“活性化”自由エネルギーを算定する。分布間距離から状態間距離を新たに定義することで、計量多次元尺度法によりその距離関係を最大限保存するように多次元自由エネルギー地形を構成することが可能になる。局所平衡を仮定する“自由エネルギー地形”上のマルコフ的遷移時系列と実際の遷移時系列との対応関係を調べることを通して、分子記憶の所在(どこからどこへの遷移の際に存在するかなど)を評価できることが期待される。

本稿で解説した(局所平衡仮定を予め必要としない)計算力学に立脚する解析理論は、(必要に応じて)上記の多次元自由エネルギー地形と比較することを通して、どのような時間スケールで自由エネルギー地形の概念が成り立つか、壊れるか、更には、分子記憶がなぜ保持しえるかなどを解明する上で大きな手掛かりを与えるものと期待される。

## 謝辞:

共同研究者の神戸大学理学部 Chun Bui Li 氏、馬場昭典氏、ケンブリッジ大 Unilever Centre for Molecular Informatics Dmitry Nerukh 氏、カリフォルニア大バークレー校 Haw Yang 氏らに感謝する。特に、Li 氏は $\epsilon$ マシンの電子移動反応を用いた 1 分子時系列データ[3]などへの具体的応用ならびに情報理論に依拠する $\epsilon$ マシンと $(\epsilon, \tau)$ エントロピーなどの力学系理論に依拠する概念との相補性に関する共同研究を実施している。また、いつも有益な議論をしてくれる神戸大学理学部郡司幸夫氏、奈良女子大学理学部戸田幹人氏らにもこの場をお借りして感謝する。本稿に関係する当該研究分野における研究活動は JST/CREST、JSPS、特定領域「強光子場科学」、神戸大学理学部地球惑星科学科 COE、分子科学研究所等から財政的支援を受けており、併せてここに感謝する。

## 参考文献

- [1] X. Zhuang, H. Kim, M. J.B. Pereira, H. P. Babcock, N. G. Walter, and S. Chu *Science*, **296** 1473(2002)
- [2a] T. Komatsuzaki and R. S. Berry, *Adv. Chem. Phys.* **123**(2002) 79
- [2b] T. Komatsuzaki and R. S. Berry, *Proc. Natl. Acad. Sci. USA* **98**,7666 (2001)[2c]小松崎民樹、松永康佑 *生物物理* **42**,285 (2002)
- [3] H. Yang, G. Luo, P. Karnchanaphanurach, T.-M. Louie, I. Rech, S. Cova, L. Xun, and X. S. Xie, *Science*, **302**, 262 (2003)
- [4] L. Edman and R. Rigler, *Proc. Natl. Acad. Sci. USA*, **97**, 8266(2000)[5] C.R. Shalizi and J.P. Crutchfield,

*Journal of Statistical Physics*, **104**, 817 (2001)

[6] L. P. Watkins, H. Yang, *Biophys. J.* **86**, 4015 (2004)[7] T.M. Cover and J.A. Thomas, *Elements of Information Theory* (Wiley Series in Telecommunications and Signal Processing) 1991

[8] C.E. Shannon and W. Weaver, *The mathematical theory of communication*, Univ. Illinois Press 1963

[9] 堀部安一、情報エントロピー論 森北出版(1989)

[10a] D. Nerukh, G. Karvounis and R.C. Glen, *J. Chem. Phys.***117**, 9611(2002); **117**, 9618(2002)

[10b] D. Nerukh, G. Karvounis and R.C. Glen *Complexity* **10**, 40 (2004)

[10c] G. Karvounis, D. Nerukh and R.C. Glen, *J. Chem. Phys.*, **121**, 4925 (2004)

[11] G. J. Rylance, R. L. Johnston, Y. Matsunaga, C. B. Li, A. Baba, T. Komatsuzaki 投稿準備中

[12]C. B. Li, A. Baba, and T. Komatsuzaki 投稿準備中

[13a] T. Bartsch, R. Hernandez, and T. Uzer, *J. Chem. Phys.* **123**, 204102 (2005)

[13b] T. Bartsch, R. Hernandez, and T. Uzer, *Phys. Rev. Lett.* **95**, 058301 (2005)

[14a] T. Komatsuzaki, K. Hoshino and Y. Matsunaga, *Adv. Chem. Phys.* **130B**, 257-313 (2005)

[14b]松永康佑 物性研究(修論特集号)**81** 571 (2004)

[14c] Y. Matsunaga, K. S. Kostov and T. Komatsuzaki, *J. Phys. Chem. A* **106**, 10898 (2002)

[15]A. Baba, C. B. Li and T. Komatsuzaki 投稿準備中