

タンパク質構造と進化と 情報幾何(生物)

小山洋平

理研CDBシステムバイオロジー研究チーム
東京大学総合文化研究科友田研究室

離散力学系の分子細胞生物学への応用数理
2009年1月8日

泰地

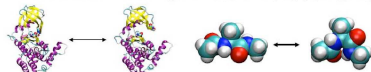
「最初の発表は小山さんで、生物サイドからのお話となります。1 日目、2 日目にも生物サイドから発表はあったのですが、かなり物理に近い内容でした。本セッションでは、どちらかというと数理寄りという構成にしたいと考えています。その辺は、生物の方でもわかるようにお願い致します。では小山さん、発表をお願いします。」

小山

「ありがとうございます。それでは発表を始めさせていただきます。」

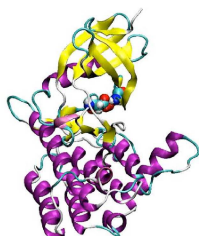
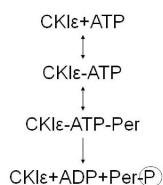
自己紹介&Outline

- 四年前に博士課程として理研CDBシステムバイオロジー研究チームに所属
- 今年度学位取得見込み
- 当初の研究目的
 - 細胞の分化のような状態変化を理解したい
 - ~D1, 7月頃止まる
 - MDシミュレーションによる転写因子のデザイン
 - ~D2 難しいことが分かった。
 - 後半で難しかった点について話します。
- 現在の研究内容
 - MDシミュレーションを用いて分子の状態変化を理解すること



まず自己紹介と概要についてです。私は 4 年前に博士課程の学生として理研 CDB のシステムバイオロジー研究チームに所属しました。今年度に学位取得見込みとなっています。4 年前に入ってきた時の研究目的は二点あって、一点目には、細胞分化のような状態変化を理解したいということ、これは D1 のときに止まってしまいました。二点目には、MD シミュレーションを用いて転写因子をデザインし、複雑な転写因子のネットワークを作るということで、これは D2 のときまで続けたんですが、難しいということがわかりました。何が難しかったかを後半で紹介したいと思います。現在は何をやっているかという、この二つの中間のような形で、MD シミュレーションを用いて分子の状態を理解しようとしています。本来は左下図のようなタンパク質機能の解析をお話出来ればよかったのですが、今は方法を開発している段階で、右下図のような小さい分子しか試せていないので拍子抜けかもしれませんが、タンパク質立体構造と統計学の解析手法がどういう風に絡み合っていくかということが伝わればと考えています。

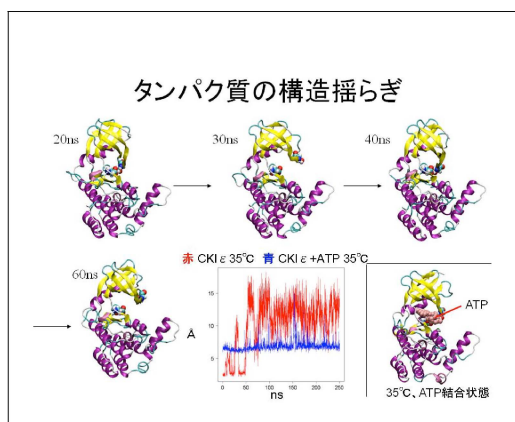
タンパク質の構造揺らぎ



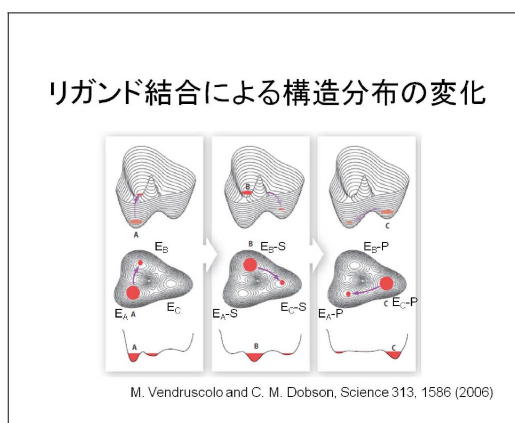
このタンパク質は哺乳類の概日時計の周期決定などに重要な機能を果たすタンパク質として知られている「CKI・」というリン酸化する酵素で、左図のように CKI・ が ATP に結合した状態になって、基質であるペリオドタンパク質と結合するようになり、最後にペリオドタンパク質にリン酸基が転移し、ペリオドタンパク質がリン酸化されるという反応を起こします。私はこの CKI・ の MD シミュレーションを行っています。

(動画デモ)

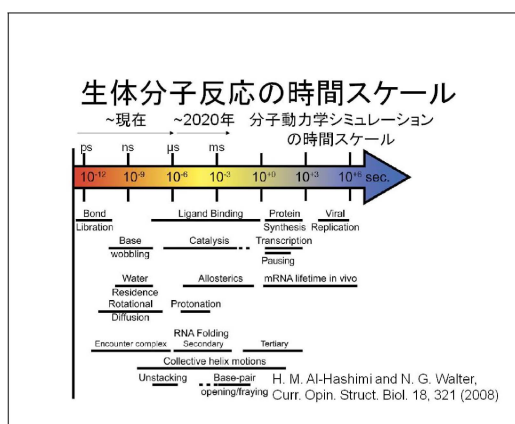
ATP バインディングサイトに ATP が結合していない状態で MD シミュレーションを行うと、ATP バインディングサイトがたまに開いたり閉じたりするのですが、最後は開き放しになります。今のアニメーションは 100 ナノ秒ぐらいの MD シミュレーションの結果です。



スナップショットで書くと、ATP のない状態でこのように ATP バインディングサイトの開閉運動が起きています。右下図は ATP が結合した状態ですが、ちょうど ATP が入り込むような開いた構造が自発的に、ATP がいない状態でも形成されていることが示されています。これが構造揺らぎというものです。中央下のグラフは、ATP バインディングサイトの（ファンデルワールスモデルで示した）原子間の距離を縦軸に、横軸を時間としたものですが、赤線は ATP が結合していないシミュレーション、開いたり閉じたりということが続き、後はずっと開いています。青線は ATP が結合してシミュレーションで、たまに開く場合がある、という風になっています。平衡状態で揺らいでいて、バインディングサイトは開いたり閉じたりしていますが、ATP が開いた状態になったときに結合すると、開いた状態に落ち着くという風になっています。



図は酵素基質反応の Energy Landscape の模式図です。最初の基質が結合していない状態 A では左図のように状態 A と基質との親和性の高い状態 B とを行き来していますが、基質が結合すると Energy Landscape が変化し、中央図のように基質が結合した状態 B が安定になります。次に、状態 B と基質を生成物にすることができる活性化状態 C との間で構造揺らぎが生じ（今回はこれは示せなかったが）、状態 C のときに触媒反応が起こり、生成物 P が生じると今度は右図のように状態 C が安定化されます。最後に、状態 C から反応物との親和性が低い状態 A との間で構造揺らぎが生じ、状態 A のときに反応物 P が放出されると、左図のように状態 A が安定になり、最初の状態に戻る、というわけです。リガンドなどが特定の分子の状態を選択し、構造揺らぎを制御することで分子の機能が実現されていると考えられています（Population Shift モデル）。



小山

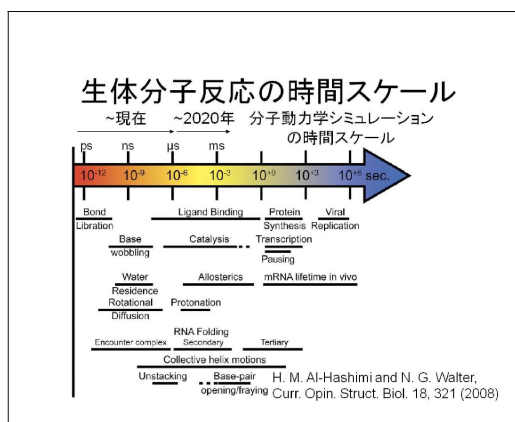
「先ほどのアニメーションで示しましたのは、100 ナノ秒とか 250 ナノ秒とかの MD シミュレーションだったんですが、いま MD シミュレーションで取り扱える時間スケールは、最先端のもので 1 マイクロ秒ほどは可能です。ちょうど大きな構造変化ぐらいは見えるようになってきた、という状況です。」

質問者

「すいません、MD シミュレーションというのは、何ですか？」

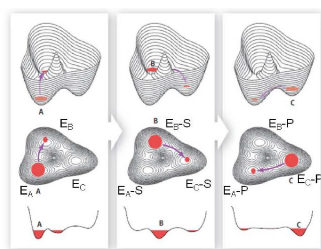
小山

「MD シミュレーションは分子動力学シミュレーションのことで、分子の結晶構造などを初期構造として、ニュートン方程式などを解いて分子のダイナミクスをコンピュータの中でシミュレーションするというものです。先ほどお見せした動画が、MD シミュレーションです。」



これから 10 年以内でマイクロ秒からミリ秒の MD シミュレーションができるようになると考えられています。速いタンパク質フォールディングはマイクロ秒のスケールで起こりますし、ミリ秒であれば普通のフォールディング速度なので、今後 10 年ぐらいでフォールディングやタンパク質の機能に関わるような大きな構造変化のシミュレーションが可能になるのではないかと考えられています。

リガンド結合による構造分布の変化

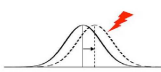


M. Vendruscolo and C. M. Dobson, Science 313, 1586 (2006)

先程は状態変化を Energy Landscape の模式図で示したのですが、マイクロからミリ秒の MD シミュレーションができるようになってくると分子の安定状態間の遷移であるとかリガンドの選択的な結合で状態が変化するというのを、模式図ではなく原子レベルで詳細に観察できるだろうと考えられます。その場合、分子の安定状態やその状態変化というのを原子レベルから理解するために、MD シミュレーションを行ったあとの解析手法の開発というのが、今後益々重要になっていくだろうと考えられます。それを具体的にどうやるかということなのですが、私は原子間相互作用への摂動によって構造揺らぎがどう変化するかを考えることで分子状態の同定や制御メカニズムが理解できるのではないかと考えています。

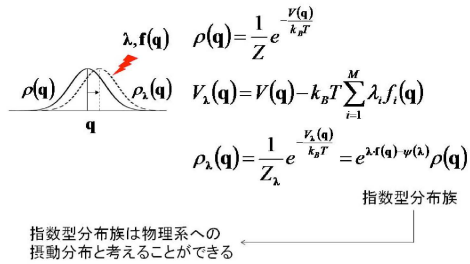
摂動による分子状態の同定と制御

- 分子状態の同定
 - 分子内相互作用の摂動
 - 構造を大きく変化させる分子内相互作用の同定
 - 対応する状態の同定
- 分子状態の制御
 - 分子間相互作用の摂動
 - 構造を大きく変化させる分子間相互作用の同定
 - 対応する状態の比率変化予測



分子状態の同定や制御メカニズムを理解するためにはどうしたらいいかということですが、ここでは、分子内の原子間相互作用の強さを変えたときの構造の分布の変化を考えます (図)。例えば、分子中のある原子間相互作用の強さを変えて分布があまり変化しなかったら、その相互作用の詳細は元の構造をとる上で重要でなかったということがわかりますし、逆にその強さを変えて分布が大きく変化したら、その相互作用は重要であったということがわかります。そこで、構造の分布が大きく変化するような原子間相互作用の組み合わせを探すことで、重要な原子間相互作用の組み合わせとその相互作用が規定している分子状態を同定でき、これらの原子間相互作用の変化がその分子状態の制御に関わっていると考えます。

分子への摂動



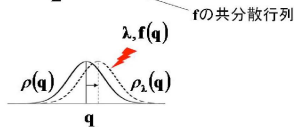
まず分子への摂動とは何かということですが、 q というのが分子の構造だと
して、 $\bullet(q)$ というのが構造の分布、 $f(q)$ というのが構造に対する摂動で、
 $\bullet$ が摂動の大きさを表すパラメータです。 $\bullet$ の摂動を掛けたときの構造揺ら
ぎの分布が点線で示したもので、 $\bullet_\lambda(q)$ と変化するものと仮定します。まず
この q の分布がどうなるかということですが、これは統計力学のカノニカ
ル分布として知られているもので、タンパク質のポテンシャルエネルギーが
 $V(q)$ だったときに、右上式のような分布となるということがわかります。
次に摂動をどう表すかということですが、元のポテンシャルエネルギーに右
中央式の様に色々な項を足していき、これを摂動として考えます。このとき
の分布は先程と同じように V_λ のカノニカル分布を変形していくと最終的に
右下のような式となります。この分布の形は統計学で「指数型分布族」
として知られています。つまり指数型分布族というのは物理系への摂動
を掛けたときの分布と一般的に考えることができます。

摂動による構造分布変化の定量化

Kullback-Leibler divergence

$$D(\rho_\lambda \parallel \rho) = \int \rho_\lambda(q) \log \frac{\rho_\lambda(q)}{\rho(q)} dq \geq 0$$

$$D(\rho_\lambda \parallel \rho) = \frac{1}{2} \lambda^T \text{cov}(f) \lambda + \dots$$



次に、摂動を掛けたときに構造分布の変化が大きい摂動を探していきたいの
で、それを定量化するというのをします。用いるのは Kullback-Leibler
divergence、または相対エントロピーあるいは情報利得という言い方をさ
れる量で、定義は上式のようなものです。二つの分布がどれくらい離れたも
のであるかというものを表す量で、両分布が同じ場合にだけ 0 となり、違う
分布だと 0 より大きくなります。この KL divergence の値が大きいほど分
布の変化は大きいという風に用います。 $\bullet$ に関して展開すると、二次の項に
は共分散行列と呼ばれるものが出てきます。

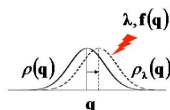
大きな構造分布変化を引き起こす
摂動の組み合わせ

$$\max_{\|\lambda\|=\delta} D(\rho_\lambda \parallel \rho) \approx \max_{\|\lambda\|=\delta} \frac{1}{2} \lambda^T \text{cov}(f) \lambda$$

$$= \frac{1}{2} (\delta u_1)^T \text{cov}(f) (\delta u_1)$$

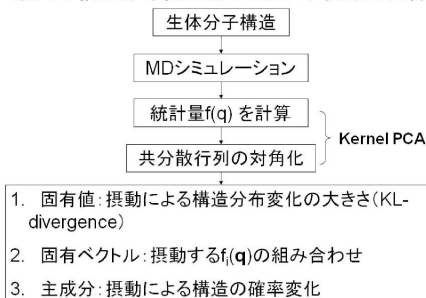
$$= \frac{1}{2} \delta^2 \sigma_1^2$$

$u_1; \text{cov}(f)$ の第一固有ベクトル
 $\sigma_1^2; \text{cov}(f)$ の第一固有値



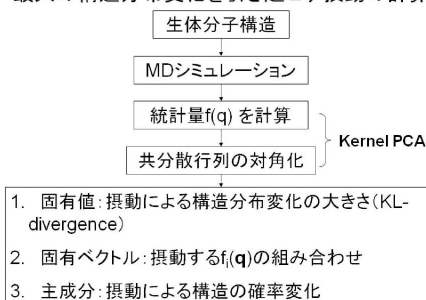
次に、大きな構造変化を起こす摂動の組み合わせを探します。 $\bullet$ がバラバラ
に変えられるといくらでも組み合わせが出てきてしまいますので、 $\bullet$ がある
一定の長さの摂動を考え、先程の摂動前の分布から摂動後の分布が一番大き
くなるような $\bullet$ を探してくるということをします。摂動の大きさが小さけ
れば先程の二次の近似が成り立ちますので、上式のように書けます。上式右
辺の最大化は、共分散行列の固有値問題として捉えることができ、共分散
行列の第 1 固有ベクトルの方向に $\bullet$ という摂動を掛けたときに最大になる、
ということが一般的に示せます。このときの分布の変化の大きさがどれぐ
らいになるかというのが、共分散行列の固有値で表すことができます。
ここで、最大の構造分布変化を起こす摂動をどのように探すかという一連の
流れをまとめます。まず、X 線結晶解析などで得られた生体分子構造を初
期構造として MD シミュレーションを行い、先程のある統計量 $f(q)$ を構
造データから計算します。

最大の構造分布変化を引き起こす摂動の計算



その統計量から共分散行列を計算した後対角化し、第1固有ベクトルをとってくればそれが一番大きい摂動になっているということがわかります。これは q という空間から f という空間に写像して、そこでの主成分分析(PCA: Principal Component Analysis)をやっているということに対応しています。このような写像した空間でのPCAはKernel PCAとして知られており、今のように統計力学的に考えるとKernel PCAというのは物理的な系への摂動を考えたときに、分布を最も大きく変化させる摂動を探す手法と考えることができます。このとき固有値は摂動を与えたときの構造分布変化の大きさ(KL divergence)に対応しており、固有値が大きいほど分布は大きく変化しているということに対応します。また、固有ベクトルは摂動する組み合わせに対応していることがわかります。今回は説明いたしませんでしたが、摂動によって構造の比率がどのように変化するかというのが主成分といわれる値で決まるということもわかります。

最大の構造分布変化を引き起こす摂動の計算



質問者 A

「MD シミュレーションは、分子内の相互作用というのをどの程度網羅しているんですか？ほぼ全てを網羅しているのですか？」

小山

「そうです。全部の原子間の相互作用に関して、これは古典力学ですけれども、計算しています」

質問者 B

「水分子は？」

小山

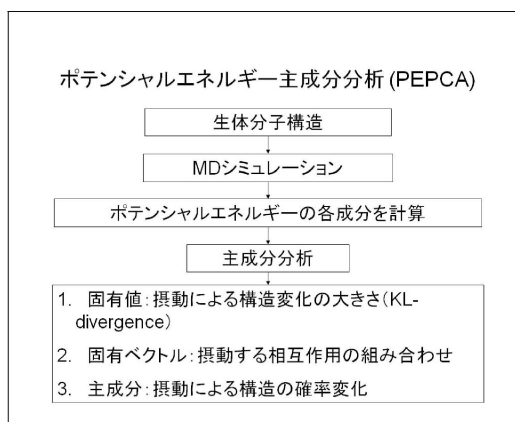
「最初の例では水は入れていませんでしたが、二つ目の例では水を入れていきます。水の効果についても後でお話し致します」

生体分子ポテンシャルエネルギー
(相互作用)

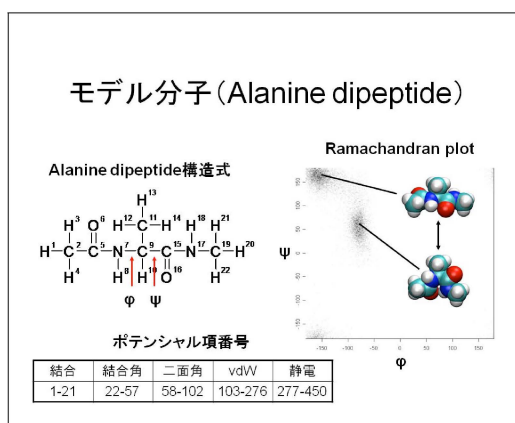
$$U = \sum_{\text{All Bonds}} \frac{1}{2} K_b (b - b_0)^2 + \sum_{\text{All Angles}} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{\text{All Torsion Angles}} K_\phi [1 - \cos(n\phi + \delta)] + \sum_{\text{All nonbonded pairs}} \epsilon \left[\left(\frac{r_0}{r} \right)^{12} - 2 \left(\frac{r_0}{r} \right)^6 \right] + \sum_{\text{All partial charges}} \frac{332 q_i q_j}{r}$$

Michael Levitt, Nat. Struct. Biol. 8, 392 (2001)

左に書きましたのは先程のポテンシャルエネルギーで、実際はこのように個別の相互作用を仮定して、シミュレーションを行います。第1項は結合の強さ、フック力のようなもので、第2項は角度によるポテンシャル、第3項は捻れたりするときの角度によるポテンシャル、第4項はファン・デル・ワールス力と呼ばれる、距離が近いと反発し、距離が遠いと引力が働くというものです。最後の項は静電相互作用です。今まで摂動を掛けるというしてきましたけれども、分子の状態がどういう風にできているかというのを調べるためにはまず分子がこの相互作用で成り立っているということを念頭に置くと、この相互作用の強さを少し変えてみたりして構造の分布がどのように変わるかということを見てやれば、どの相互作用が構造に一番重要であるかがわかるのではないかという風に考えました。



そこで、統計量としてポテンシャルエネルギー項を用いたポテンシャルエネルギー主成分分析 (PEPCA) というのを考えました。図は一般の統計量をポテンシャルエネルギーに書き直しただけですけれども、まず初期構造を用意し、それについて MD シミュレーションを行って、ポテンシャルエネルギーの各成分を計算し、その後に主成分分析を行います。そうすると、第一固有ベクトルが構造の分布を最も大きく変化される原子間相互作用の組み合わせに対応し、固有値はそのときの分布の変化を表します。また、主成分は摂動によって構造の比率がどのように変化するかということを表します。



第一主成分 (青: 一、赤: 十)

PCA plot showing two clusters of points, A (blue) and B (red), in a 2D space. The x-axis ranges from -100 to 100, and the y-axis ranges from -100 to 100. Cluster A is centered around (-50, 50) and Cluster B is centered around (50, -50).

第一固有ベクトル

NMR spectrum of the first principal component. The x-axis is labeled '化学作用番号' (Chemical shift number) and ranges from 0 to 600. The y-axis is labeled '強度' (Intensity) and ranges from -0.4 to 0.4. Two peaks are labeled: '4.6-18' and '4.6-16'.

**正の相互作用を強める
負の相互作用を弱める**

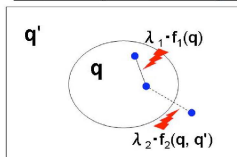
NMR spectrum and molecular model for the first component. The spectrum shows peaks at 4.6-18 and 4.6-16 ppm. The molecular model shows the interaction between the two clusters, A and B, with positive interactions (red arrows) and negative interactions (blue arrows).

**正の相互作用を強める
負の相互作用を弱める**

NMR spectrum and molecular model for the second component. The spectrum shows peaks at 4.6-18 and 4.6-16 ppm. The molecular model shows the interaction between the two clusters, A and B, with positive interactions (red arrows) and negative interactions (blue arrows).

Figure 1 consists of four panels illustrating the interaction between the first main component (Q1) and the first fixed vector (V1).
 - Top-left: A 2D scatter plot showing the correlation between Q1 (x-axis, ranging from -100 to 100) and V1 (y-axis, ranging from -0.2 to 0.2). The data points show a positive correlation.
 - Top-right: A 3D molecular model showing the interaction between the first main component (Q1) and the first fixed vector (V1). The model is labeled with 'Q1' and 'V1' and shows a positive correlation.
 - Bottom-left: A 3D molecular model showing the interaction between the first main component (Q1) and the first fixed vector (V1). The model is labeled with 'Q1' and 'V1' and shows a positive correlation.
 - Bottom-right: A 3D molecular model showing the interaction between the first main component (Q1) and the first fixed vector (V1). The model is labeled with 'Q1' and 'V1' and shows a positive correlation.

	$D(\rho_{\mathbf{z}}(\mathbf{q}, \mathbf{q}') \ \rho(\mathbf{q}, \mathbf{q}')) = D(\rho_{\mathbf{z}}(\mathbf{q}) \ \rho(\mathbf{q})) + \langle D(\rho_{\mathbf{z}}(\mathbf{q}' \mathbf{q}) \ \rho(\mathbf{q}' \mathbf{q})) \rangle_{\mathbf{q}}$
$\lambda_2 = 0$	$\frac{D(\rho_{\mathbf{z}}(\mathbf{q}, \mathbf{q}') \ \rho(\mathbf{q}, \mathbf{q}'))}{\frac{1}{2} \lambda_1^T \text{cov}(\mathbf{f}_1) \lambda_1} \quad \frac{D(\rho_{\mathbf{z}}(\mathbf{q}) \ \rho(\mathbf{q}))}{\frac{1}{2} \lambda_1^T \text{cov}(\mathbf{f}_1) \lambda_1} \quad \frac{\langle D(\rho_{\mathbf{z}}(\mathbf{q}' \mathbf{q}) \ \rho(\mathbf{q}' \mathbf{q})) \rangle_{\mathbf{q}}}{0}$
$\lambda_2 \neq 0$	$\frac{D(\rho_{\mathbf{z}}(\mathbf{q}, \mathbf{q}') \ \rho(\mathbf{q}, \mathbf{q}'))}{\frac{1}{2} \lambda^T \text{cov}(\mathbf{f}) \lambda} \quad \frac{D(\rho_{\mathbf{z}}(\mathbf{q}) \ \rho(\mathbf{q}))}{\frac{1}{2} \lambda^T \text{cov}(\mathbf{f}) \lambda} \quad \frac{\langle D(\rho_{\mathbf{z}}(\mathbf{q}' \mathbf{q}) \ \rho(\mathbf{q}' \mathbf{q})) \rangle_{\mathbf{q}}}{\frac{1}{2} \lambda_2^T \text{cov}(\mathbf{f}_2 \mathbf{q}) \lambda_2}$

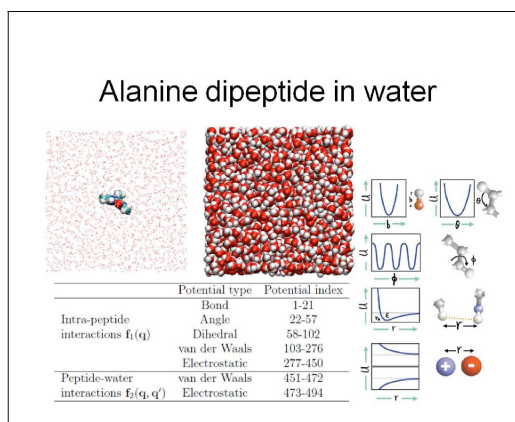


分子間相互作用による
分子状態の制御

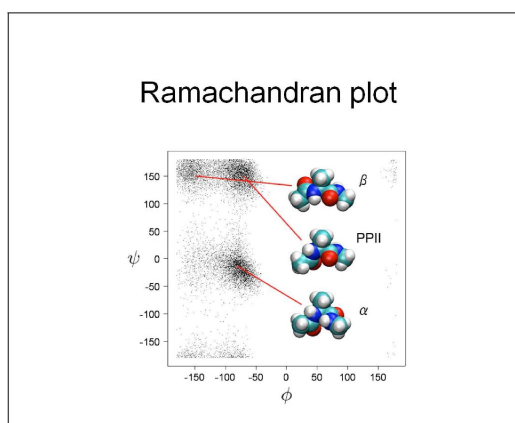
PCA with $\text{cov}(\langle \mathbf{V}_i | \mathbf{q} \rangle / k_B T)$

左上図のラマチャンドランプロットで赤点が主成分が正の構造、青点が主成分が負の構造を示しています。赤と青の点がそれぞれ二つの安定状態 A と B に対応していることが分かります。このことから第一固有ベクトルの相互作用の組み合わせは分子の二つの安定状態およびその比率決定に関わっていることが分かります。実際に、固有ベクトルの成分が正である e-6-18 や負である el-8-16 は引力的な静電相互作用ですが、それぞれ主成分が負である状態 B や正である状態 A の安定化に重要な相互作用であることが右上図の分子構造からわかります。また、その他の固有ベクトルの成分が負である el-5-18 や el-6-17 が斥力的な静電相互作用ですがこれらを強めると主成分が負である状態 B が不安定化することが分かります。このように、主成分の符号により固有ベクトルの原子間相互作用の組み合わせがどのような分子状態を安定化あるいは不安定化しているかを同定することができます。

今までは分子内への摂動を考え、分子の状態や状態変化を理解することが目的でした。次に、外部の分子による分子状態の制御メカニズムを理解するために、分子間への摂動を考えます。左下図では、タンパク質の座標を q として、それ以外の水などの座標を q^* と表しています。分子内への摂動パラメーターを $\bullet_1$ 、統計量を f_1 と書いて、分子間への摂動を $\bullet_2$ 、 f_2 と書きます。それらをまとめて、 $\bullet$ が全系への摂動と書きます。全系の分布が摂動によりどれくらい変化するかは上の式の左辺で書けます。次にこのタンパク質だけの変化を表すというのが第1項で、知りたいのはこの第1項です。しかし chain rule というのをを使うと、第2項のような余分な項が出てきます。これらの二次近似をまとめたものが中央の表です。分子内摂動による分子の構造分布変化を考えたいときには統計量の共分散行列を用い、分子間摂動では統計量の条件付期待値の共分散行列を用いればよいことが分かります。



以上の考え方を試すために、先ほどの Alanine dipeptide が水の中に入った系を考えます（上段図）。分子内のポテンシャルに加えてペプチドと水の間の分子間相互作用の項が新しく出てきます（下段表）。



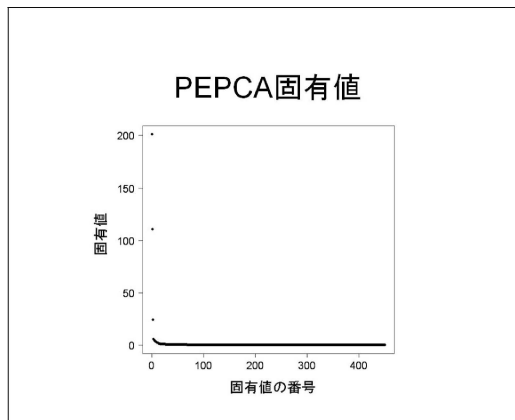
図は水中での Alanine dipeptide の Ramachandran plot です。真空中では安定状態が二つでしたが、水中では三つあることが分かります。

分子内相互作用による 分子状態の同定

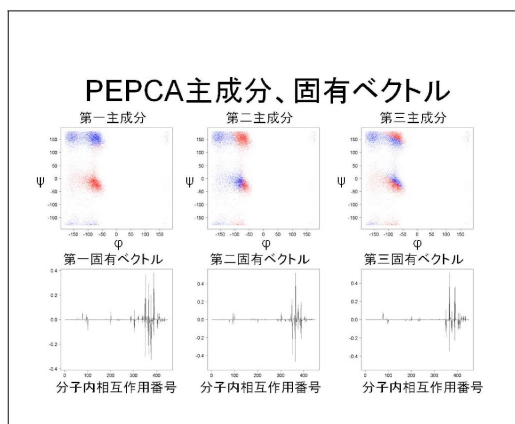
$$\text{PCA with cov}\left(\frac{\mathbf{V}_1(\mathbf{q})}{k_B T}\right)$$

ポテンシャルエネルギー主成分分析
PEPCA: Potential energy principal
component analysis

まず、分子内での相互作用による状態の同定を考えます。そのためには先の分子内相互作用の共分散を考えればよいので、その主成分分析を行います (PEPCA)。



そうすると、摂動を掛けたときの分布の変化が特に大きい固有値が3つあるということがわかります。



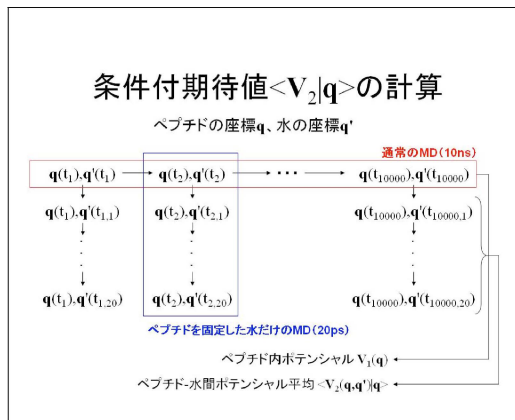
対応する主成分がどのような状態を規定しているかという、第一主成分は左上図のように上下の状態の安定化に関与しており、第二主成分は左右2つの状態の安定化に関与しています（下の安定状態は無視した場合）。つまり第一と第二主成分で3つの安定状態を規定していることがわかります。第三主成分では、左上の安定状態を無視すると右上と右下の安定状態の中間状態を識別しているので、この状態間の遷移状態を決定していることが分かります。また、下図の固有ベクトルにより具体的にどの原子間相互作用がどの状態を安定化・不安定化させているのかを詳細に解析することができます。

分子間相互作用による 分子状態の制御

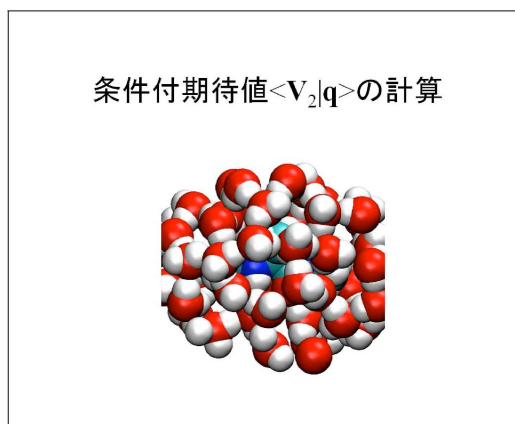
$$\text{PCA with cov}\left(\frac{\langle V_2(\mathbf{q}, \mathbf{q}') | \mathbf{q} \rangle}{k_B T}\right)$$

条件付期待値ポテンシャルエネルギー主成分分析
cPEPCA: Conditional expectation potential
energy principal component analysis

次に、どのような分子間相互作用の組み合わせが分子の構造変化に重要であるかについて考えます。これは分子間相互作用のポテンシャルエネルギーの条件付期待値に対する主成分分析 (cPEPCA) で解析することができます。



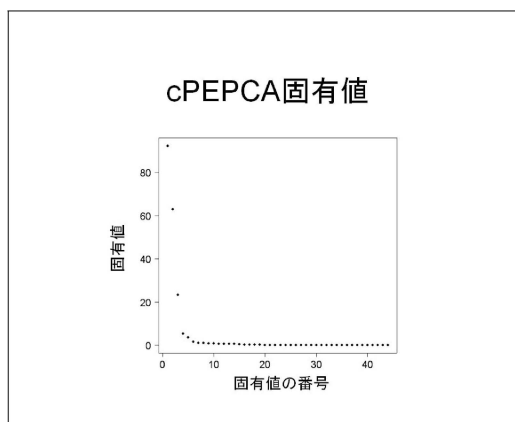
分子間相互作用のポテンシャルエネルギーの条件付期待値は以下のようにして求めることができます。まず通常の MD シミュレーションを行います (図の赤枠)。次に、得られた複数の構造を初期構造として水などの環境分子だけを動かす複数の独立な MD シミュレーションを行います (図の青枠)。この後、分子間相互作用の平均値をそれぞれの MD シミュレーションごとに計算することで条件付期待値を計算することができます。



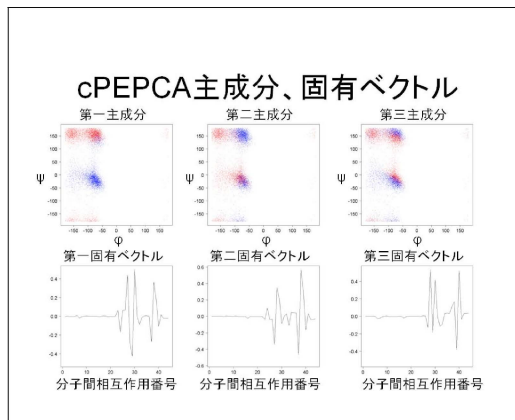
条件付き期待値を求めるための MD がこちらです。

(動画デモ)

ペプチドは完全に止まった状態ですが、水分子はペプチドと相互作用しています。



この条件付き期待値に対して主成分分析を行いました (cPEPCA)。ここでは水は一まとまりとして考えており、ファンデルワールスと静電相互作用をあわせて水との相互作用は 44 個しかないんですけども、先ほどと同じように 3 つの大きな固有値が出てきました。



対応する主成分の符号は先ほどとほとんど同じパターンとなり、第一主成分が上下の状態、第二主成分が左右の状態、第三主成分が右側の状態間の遷移状態となりました。下段は固有ベクトルで、真空中のときと同様に、正の分子間相互作用を強め、負の分子間相互作用を弱めると、主成分が負の構造の比率が増える、と推測することができます。第一から第三固有ベクトルで分子間相互作用のパターンが異なっていますが、異なる分子間相互作用のパターンにより異なる状態の比率が決定されているということがわかります。

まとめ: 分子状態の同定と制御

- 分子内相互作用による分子状態の同定
 - ポテンシャルエネルギー主成分分析 (PEPCA)
- 分子間相互作用による分子状態の制御
 - 条件付期待値ポテンシャルエネルギー主成分分析 (cPEPCA)

- 固有値: 構造の分布の変化の大きさ
- 固有ベクトル: 原子間相互作用の振動の組み合わせ
- 主成分: 振動による比率の変化

**構造揺らぎにより特異的な刺激に対する
選択的な状態変化が可能**

これまでのまとめです。分子の異なる状態が異なる分子の機能を実現しており、分子状態はその原子間相互作用で決定されることから、この分子状態を分子内相互作用から理解するための手法としてポテンシャルエネルギー主成分分析 (PEPCA) を紹介しました。次に、この分子状態を制御している分子間相互作用を明らかにするために条件付期待値ポテンシャルエネルギー主成分分析 (cPEPCA) を紹介しました。また、これらの解析手法から得られる固有値は構造の分布変化の大きさ、固有ベクトルは原子間相互作用の組み合わせ、主成分は振動による構造の比率変化、という解釈ができます。また、異なる原子間相互作用の組み合わせが異なる分子状態の比率を決定しており、このような性質は分子が細胞中のような様々な分子が存在する中で、特異的な反応を行うことができることに関係しているかもしれません。

分子デザイン

- アミノ酸や塩基の置換も振動と考えることができる
- 大きな振動なので二次近似では不十分であることが多い

次に、分子デザインが難しい点と思った点についてお話いたします。アミノ酸や塩基の変異も振動と考えることができますが、先ほどの分子の構造揺らぎ解析と異なり、変異は一般には大きい振動となるために、二次近似では不十分となる点が多いです。

変異による分布の変化

$$\frac{\rho_B(\mathbf{q})}{\rho_A(\mathbf{q})} = e^{\frac{\Delta V(\mathbf{q}) - \Delta F}{k_B T}} = \frac{\rho_B(\Delta V)}{\rho_A(\Delta V)}$$

$$\Delta V(\mathbf{q}) = V_B(\mathbf{q}) - V_A(\mathbf{q})$$

$$\Delta F = F_B - F_A$$

$$D(\rho_B(\mathbf{q}) \| \rho_A(\mathbf{q})) = D(\rho_B(\Delta V) \| \rho_A(\Delta V))$$

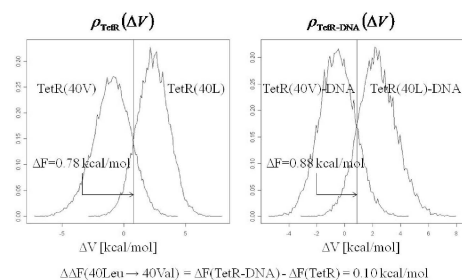
$\mathbf{q}$: 3N次元

ΔV : 1次元

自由エネルギー変化 ΔF の推定誤差は摂動前後の分布の重なり具合に関係
C. H. Bennett, J. Comp. Phys. 22, 245 (1976)

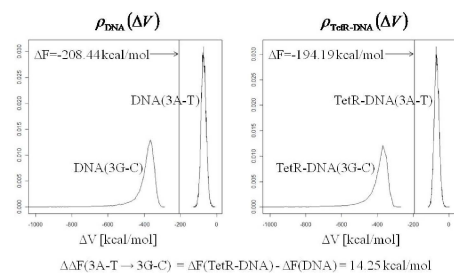
分子デザインをする際にデザインの指標となる分子間の親和性（結合定数）を求めるために、変異による自由エネルギー変化の推定を行う必要があります。A という分子から B という分子への変異を考えます。両者での構造の比率は上式左辺のように書けますが、分子 A での分子 B と分子 A との間のポテンシャルエネルギーの差の分布と、B での差の分布の比率に等しい（右辺）ということがわかります。この性質を用いると、分子 A から分子 B に変異させたときの 3N 次元の構造分布の変化を一次元のポテンシャルエネルギーの差の分布で実際に計算することができます。3N 次元となる $\mathbf{q}$ の積分から、一次元の V の積分へと圧縮できるという点が不思議なところですが、これは統計の「十分統計量」というものの性質に関係してくるものと考えられます。また、 $\bullet V = \bullet F$ のときに二つの分布の値が一致し、二つの分布の重なり具合が大きいほど自由エネルギー変化が精度よく推定できることがわかります。

アミノ酸置換 (40Leu→40Val)

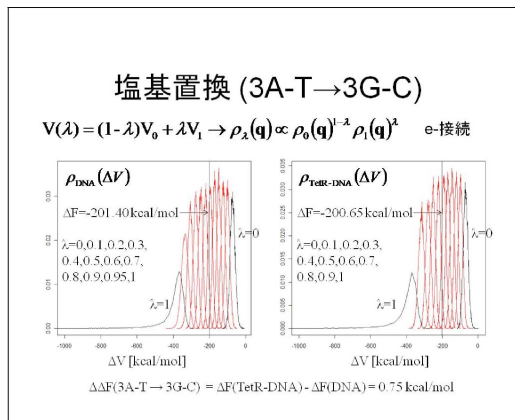


図は以前に行った TetR という転写因子と DNA の間の結合定数をデザインしようとした時の自由エネルギー計算です。TetR の 40 番目のロイシンをバリンに変異させています。変異前後の MD シミュレーションを行った後、 $\bullet V$ の分布が図のようになります。分布が交差している部分が自由エネルギー変化に対応します。分布の重なりが十分大きいので自由エネルギー変化が精度よく推定できることがわかります。

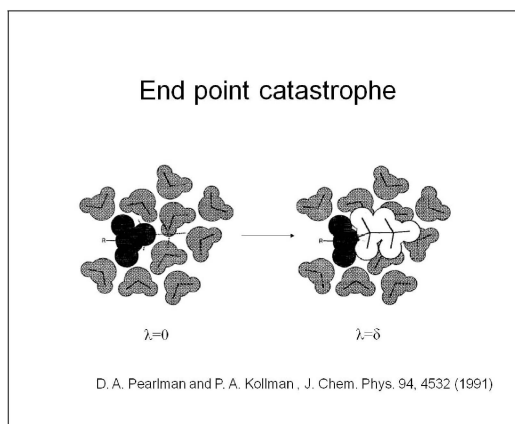
塩基置換 (3A-T→3G-C)



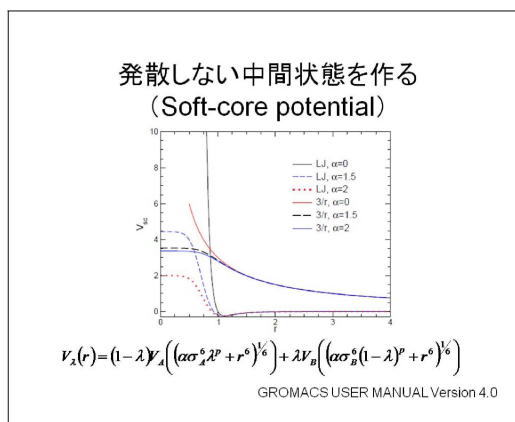
これは、DNA の 3 番目の A-T を G-C に置換した時の $\bullet V$ の分布です。変異前後で分布が全然重なり合わないため、自由エネルギー変化が精度良く計算できないということがわかります。



このような場合には、変異前後のポテンシャルエネルギーを上式左辺のように混合した中間的な状態を作り、分布の重なりができるようにする、ということがよくやられています。分布の形で書くと上式右辺のようになりますが、これは情報幾何でe-接続と言われているもので、中間の状態を分布の掛け算で作るというものです。ただ、分布の重なりを十分にするためには多くの中間状態が必要となり、多くの計算量が必要となります。また、 $\lambda = 1$ のときにロングテイルになっていますが、自由エネルギー変化が収束するためには長時間の計算時間が必要となります。



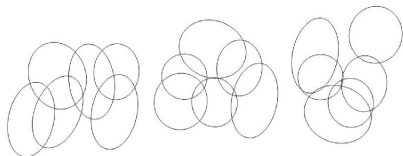
このようなロングテイルが生じる原因は end point catastrophe というもので、変異を入れたときに新たな原子が突然増えると、元にあった原子との重なりができてしまうので (右図)、そこで反発が起きエネルギーが発散してしまうというものです。



この問題点を回避するために、原子に重なりがあっても発散しない仮想的なポテンシャルを作る、ということがされています。

分子デザイン

N箇所のアミノ酸の網羅的な置換をすると 20^N の構造が存在する



- ・自由度が変わる分布を統一的に扱うやり方は？
- ・最適な中間状態の作り方、サンプリングの仕方は？
- ・最適な計算をしたときの計算のオーダーは？
- ・実際の進化ダイナミクスと構造揺らぎの関係は？

分子デザインを行うために、N 残基のアミノ酸を網羅的に置換しようとする、 20^N の MD シミュレーションを行わなければいけません。しかし、図のように構造の分布が似ている変異に関しては、まとめて一度に計算することができれば計算量が指数関数的に多くなるということにはならなさそうな気がします。このときに、昨日に時田先生のお話でも出ていたと思いますが、変異のように自由度が変わるダイナミクスをどうやって扱ったらいいか、という点も難しいです。また、計算手法以外にも、実際の進化（変異）のダイナミクスと構造揺らぎに関係があるかどうか、という点も気になるところです。

構造揺らぎと Kullback-Leibler divergence

分子の構造揺らぎ

Kullback-Leibler divergence

分子の非平衡統計力学
分子の情報理論
分子の計算理論
分子の統計学（推定、検定）

最後ですが、Kullback-Leibler divergence は情報理論、統計学などの様々な分野で重要な役割を果たすことが知られており、また、非平衡統計力学や計算の熱力学においても使われています。今回、構造揺らぎ解析でも Kullback-Leibler divergence が役立つことをお話しましたが、分子の非平衡統計力学、情報理論、計算理論、統計学などが分子構造揺らぎの原子間相互作用のレベルから理解することができれば面白いと考えています。

Acknowledgement

- Hiroki R. Ueda @ RIKEN CDB
- Tetsuya J. Kobayashi @ Tokyo Univ.
- Shuji Tomoda @ Tokyo Univ.

最後に謝辞ですけれども、上田さん、小林さん、友田先生に感謝いたします。以上です、ありがとうございました。

質疑応答

● 質問者 A

水分子はどれぐらい入れているんですか？

● 小山

十分、入れています。Alanine dipeptide は 2000 個ほどで、CKI ϵ は 9000 個ほどです。

● 泰地

タンパクの原子数の 10 倍ぐらい入れるのが普通ではないでしょうかね。

● 小山

そうですね、3 層から 5 層ぐらい水の層ができるぐらい入れるのが、今は普通ですね。

● 質問者 B

ある程度小さい分子であればシミュレーション可能だと思うんですけども、例えば普通のタンパク質で、小さいドメインであっても、数百アミノ酸ぐらいにはなりますよね。現状だと一次配列から一意に立体構造が決まるわけではないという中で、こういった摂動の中で相互作用を探するという試みはどの程度可能なのでしょうか。

● 小山

フォールディングは X 線結晶構造解析で決めて、その中で安定状態があるわけです。最初に図示しましたように、キナーゼの開状態と閉状態があるのですが、その比率をどう変えたらいいのかということをシミュレーションしますが、フォールディングまではやろうとは考えていないです。

● 質問者 B

それは NMR を使ったときの実測の揺らぎと一致するものなのですか？

● 小山

まだそこまで詳細には対応していないのですが、最近少しずつやられていますけれども、どれぐらい対応するかという事は私もよくわかりません。そういう反応速度が構造揺らぎに対応しているということは NMR で見えてきています。

● 泰地

X 線結晶構造解析だと B ファクターと言って、原子の揺らぎの大きさが見えるんですけども、この B ファクターの値と、MD で出した原子の揺らぎの大きさというのはかなり一致します。今は、判別をしようという段階だと思うんですが、こういうものと熱力学量との関係は見えてきているのでしょうか？

● 小山

そこは興味のあるところで、先ほど一分子の熱力学で Kullback-Leibler divergence が関係していると言いましたが、どういう風に関係しているかというと、それが自由エネルギーよりも余分にしなければいけない仕事量（散逸仕事）に関係していて、どれぐらい熱を捨てることによって状態変化をしなければならないかということにもしかすると関係するかもしれないと思っています。

● 質問者 C

最初の CKI というタンパク質は、体内時計に関わるタンパク質なんですよ？あの ATP binding site が開閉するというのが、体内時計の周期に対応するのですか？

● 小山

いえ、あの過程は重要かどうかはわかりません。周期決定に本当に重要なのは、おそらくは ATP が結合した後のピリオドタンパク質のリン酸化の過程で、そのシミュレーションはまだやっていません。

● 質問者 C

ATP って電荷を持っていますよね？電荷を持っている ATP がくつついても、その binding site は段々揺らがなくなるものなんですか？タンパク質の binding site はプラスの電荷を持っているものだと思うんですけど、マイナスとプラスの電荷の相互作用はタンパク質の構造が揺らいでしまうほどに弱いものなのですか？

● 小山

いえ、一旦 binding すると揺らぎは小さくなります。ATP が結合する前は揺らぐんですが、ATP が結合するとほとんど揺らがなくなります（2 ページ 1 番目スライド）。

● 泰地

さっきの Kullback-Leibler に戻ると、ああいうのが重要になってきたのは、MD でも昔は 1 個やっておしまい、みたいだったのが、かなり長くやれるようになって、分布が見れるようになった。1 分子のものでも分布が見れるようになったんで、Kullback-Leibler divergence のようなものを使って議論ができるようになった、ということなんだと思うんですが、他にこういった分布が見れるようになったという例というのはないんでしょうか？このあと江口先生の話でも測定値の分布という話は色々出てくると思うのですが、測定値がたくさんとれるというものから分布が取れるという例は他にないんですかね？

● 質問者 D

細胞というのはできますよね？細胞レベルでの発現状態というのと集団レベルでの発現状態の分布であるとか。

● 質問者 E

逆に、分布がわかることでどういったことが数学的に示せるんですか？分布は取ろうと思えばどのようなものでも取れるんですか？

● 小山

その分布から、どういう方向に変化が起こりやすいか、であるとか、ということはわかります。どういう風に刺激を与えるとどういう遷移が起こりやすいかということは、揺らぎの情報があると推定できます。

● 泰地

他にございますでしょうか？他になければそろそろ時間ですので、小山さん、ありがとうございました。