

課題名(タイトル):

## ALEX 計測による1分子 FRET データの最大エントロピー・クラスタリング解析

利用者氏名:

○岡本 憲二、佐甲 靖志

理研における所属研究室名:

開拓研究本部 佐甲細胞情報研究室

### 1. 本課題の研究の背景、目的、関係するプロジェクトとの関係

蛋白質分子の構造は機能と深く関わっており、その理解は極めて重要で、長年にわたって研究がおこなわれている。しかし、そのほとんどは精製分子を用いた *in vitro* 実験であり、実際の生きた細胞の中での構造変化の様子は明らかになっていない場合が多い。

われわれは、主に溶液中実験で用いられてきた ALEX (alternating-laser excitation) 法を生細胞中の蛋白質分子に応用することに成功した。ALEX 法は、溶液中を自由拡散する分子を対象とした 1 分子計測法であり、2色の色素で蛍光ラベルした蛋白質などの標的分子を蛍光バースト信号として検出する。バーストからは1分子毎に、FRET (Förster resonance energy transfer) 効率(色素間距離に対応するため分子構造の手がかりが分かる)、ストイキオメトリ(2色の蛍光ラベルのラベル比)、蛍光バースト強度、等の情報が得られる。生細胞中の CRAF 蛋白質を計測した例を図 1 に示す。ストイキオメトリから、図 1(a)の上下の分布は蛍光色素の一方(FRET ドナーまたはアクセプタ)が褪色した分子と考えられるため無視でき、中央破線部分の両方の色素が活性を保った分子だけに注目することができる(図 2)。

図 2(a)のバースト分布には、さまざまな状態の分子が含まれていると考えられる。たとえば、左右に大きく分かれた 2 つの分布はそれぞれ分子の開閉構造に対応する。それ以外にも、細かい構造の違いや多量体形成など、多様な分子状態が考えられる。しかし、これまでは ALEX データを詳細に分析する方法は存在しなかった。われわれは CRAF の細胞中でのダイマー形成に注目しており、蛍光バースト強度からモノマーとダイマーの混在状況を判別できると考え、最大エントロピー法とクラスタリングを組み合わせたデータ解析法を新たに開発した。

本課題は、開発した新たな解析法を、実際の ALEX 実験データに適用し、細胞中 CRAF の分子状態を明らかにする事を目的とする。

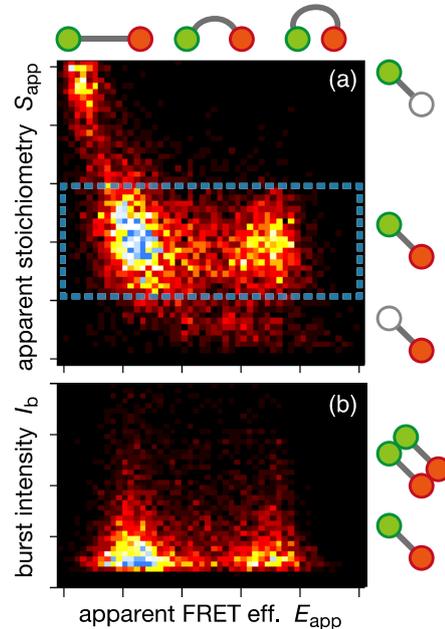


図 1: ALEX 計測データの例。バースト分布の 2 次元ヒストグラム。(a)ストイキオメトリ  $S_{app}$  vs FRET 効率  $E_{app}$ 。(b)蛍光バースト強度  $I_b$  vs FRET 効率  $E_{app}$ 。破線領域は 2 色の色素(FRET ドナーとアクセプタ)でラベルされた分子。

### 2. 具体的な利用内容、計算方法

図 2 のような FRET 効率  $E$  とストイキオメトリ  $S$  の 2 次元分布を考える時、各分子状態は固有の  $E, S$  値を持つと考えられる。しかし、実際の実験では、装置や不確定性のため分布が広がり、見かけ (apparent) の  $E_{app}-S_{app}$  分布が得られる(図 2(a), (b))。この  $E_{app}-S_{app}$  分布から、隠された (latent) 真の  $E_{lat}-S_{lat}$  分布(図 2(c))を求めたい。この時、 $E, S$  方向の広がり(ベータ分布になる(計算上は  $S$  に関してはガウス分布で近似)ことを考慮した上で、 $E_{lat}-S_{lat}$  分布のエントロピーが最大化されることを条件とする最大エントロピー法を用いた。従来の最大エントロピー法では図 1(a)のような分布(ヒストグラム)から図 2(c)のような分布(ヒストグラム)を推定する問題が取り扱われる。今回われわれは、図 2(a)のような分布から、確率分布を推定し図 2(c)を推定する解析法を開発した。これにより、比較的データ点数が限られる 1 分子計測実験データでも精度の良い推定を実現できた。

また今回われわれは、モノマー/ダイマーの分布に着目した。モノマーとダイマーの違いはバースト強度に表れる(ダイマーがモノマーの2倍)と考えられるが、図1(b)で見られるように、バースト強度分布は指数分布となり、モノマー/ダイマーの差は指数分布の減衰係数の差として表れる。そこで、モノマー/ダイマーそれぞれに  $E-S-I$  の3次元確率分布を考え、最大エントロピー法で最適化するクラスタリング解析法を考案した。

上記の解析法を実現するため、C++言語を用いたプログラムを自作した。基本的に、比較的簡単な数式を多重ループで繰り返す構造であるため、OpenMPを用いた並列化で高速化を図った。また特殊関数と逆行列計算(数千次元。実際の大きさはデータに依存)に Intel<sup>®</sup> Math Kernel Libraryを使用した。

### 3. 結果

実際に解析したデータと結果の例を図2に示す。

データ(図2(a))は、 $E, S, I$ 分布で、各点が1つのバースト(分子)に対応する。 $E_{\text{lat}}-S_{\text{lat}}$ 分布(図2(c))から実験による広がり考慮して計算される  $E_{\text{app}}-S_{\text{app}}$ 分布(図2(b))を確率分布と見なし、データに対する尤度を最大化するように、 $E_{\text{lat}}-S_{\text{lat}}$ 分布を最適化する。この時、クラスター化されたモノマーとダイマーそれぞれに  $E-S$ 分布を最適化している(図

2(d)-(g))。図2では、モノマー成分を青、ダイマー成分を赤として表示している。緑はマイナーなノイズ成分を表す。

### 4. まとめ

新たに開発したデータ解析法により、ALEX計測データから、モノマー/ダイマーを判別しつつ、真の  $E_{\text{lat}}-S_{\text{lat}}$  状態を推定できることを示した。

ただし、その後の対照実験で、検出されたバースト強度分布が、装置的な問題(顕微鏡の色収差)に依存したアーティファクトであった可能性が明らかになった。その問題の発見は解析結果の分析によるものであり、解析法自体はバースト強度を精度良く判別する性能を備えていた証拠であると考えている。しかし、今回検出されたクラスタリング結果が、実際のモノマー/ダイマー状態を反映しているとは言えなくなった。

### 5. 今後の計画・展望

解析法自体は十分な性能を備えていると考えられるが、今回は装置の問題により、本来の目的であった、分子情報の解明には至らなかった。今後は装置の問題を解決し、正しく計測し直したデータを用いて、新たな分子情報を解明する解析に改めてチャレンジしたい。

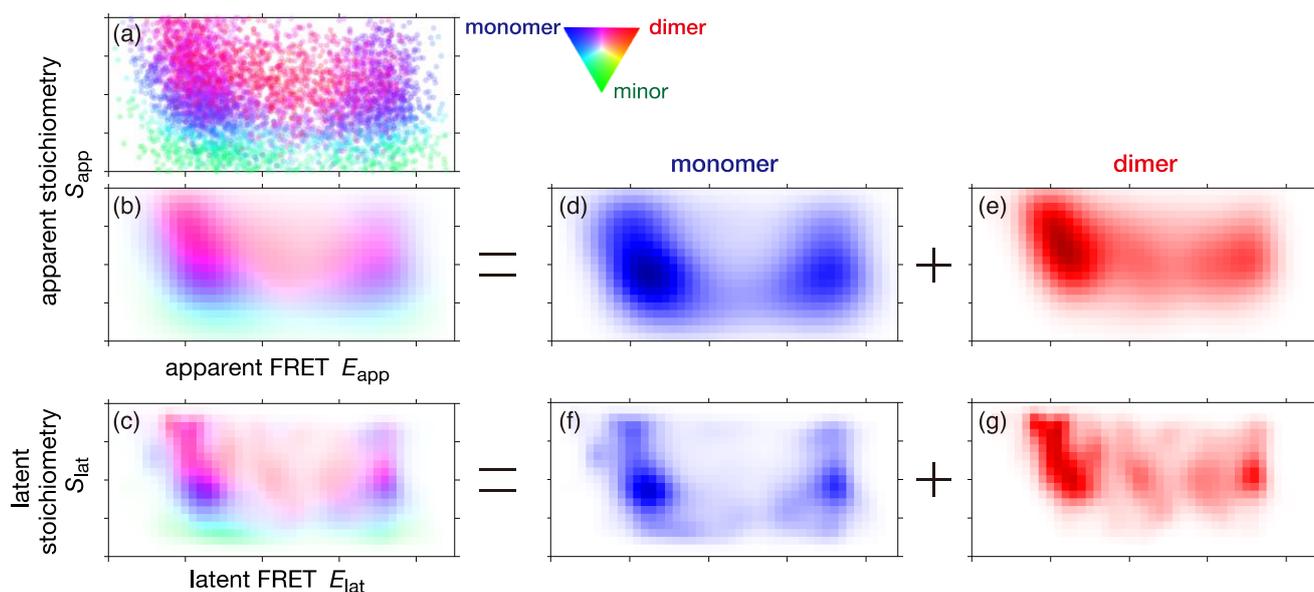


図2: 解析結果の例。(a)実験データのバースト分布。解析結果にもとづいて色付けしてある。(b),(c)解析によって復元した(b)  $S_{\text{app}}-E_{\text{app}}$ 分布と(c)  $S_{\text{lat}}-E_{\text{lat}}$ 分布。(b),(c)はそれぞれモノマー成分((d),(f))とダイマー成分((e),(g))で構成される。

## 2020年度 利用報告書

### 2020年度 利用研究成果リスト

#### 【ポスター発表】

岡本 憲二、佐甲 靖志:「最大エントロピー法と変分ベイズクラスタリングを用いた1分子 FRET データ解析による細胞質中 RAF のダイマー化状態の検出」日本生物物理学会第 58 回年会、20522S、2020 年 9 月 16-18 日(オンライン開催)。

#### 【その他(著書、プレスリリースなど)】

岡本 憲二:「1 分子計測データの隠れ状態を推定する統計的解析法—隠れマルコフモデルと最大エントロピークラスタリング」実験医学増刊 Vol. 38, No. 20 「機械学習を生命科学に使う! シークエンスや画像データをどう解析し、新たな生物学的発見につなげるか?」第 4 章「統計解析」。(スーパーコンピュータ利用の記載は失念)