

パネルディスカッション 「タンパク質構造と進化と情報幾何」

泰地真弘人 ノート 二階堂 愛

1 復習と議論

1.1 タンパク質構造解析概論

物理出身で、計算機作製を行ってきましたが、息抜きに A-Life、物理乱数の研究もやってきました。

タンパク質の立体構造を古典力学で考えると、 $6N$ 自由度のハミルトン力学と考えることができます。実際は溶媒があるのでタンパクだけだと散逸があります。では古典力学となにが違うのでしょうか。タンパク質は非均一な系なので気体・液体で使われる近似問題が使えません。よってシミュレーションが必要になります。しかし、タンパク質の立体構造研究は、動物園、博物学的で理論研究らしくないと感じます。またタンパク質はガラスっぽいと思います。というもののエネルギー最小の状態で安定なわけではないからです。

1.2 タンパク質解析の数理解析

今日はタンパク質の構造解析で使われる数理解析の俯瞰したいと思います。この分野では様々な数理が使われます。例えば、力学系、Integrators、データ解析 (PCA, Clustering)、サンプリング法 (Extended Ensembl etc...), 構造 bioinformatics/chemoinformatics (= 機械学習) などです。シミュレーションの分野では、連続できるシミュレーション時間が短い (ms 以下) ことが問題です。これは次世代スーパーコンピュータでもっと改善されるはずですが。

構造解析の課題を述べたいと思います。これは次の4つに分類できます。1. タンパク質の熱統計力学 (分布)、2. 安定構造 (基底状態/最適化, アフィーゼンのドグマ)、3. タンパク質のダイナミクス (時系列)、4. タンパク質の構造決定 (最適化)。この4つが立体構造解析の世界で課題とされている数理的な問題です。

1.3 統計物理とダイナミクスの関係

さて、統計物理とダイナミクスの関係について考えたいと思います。統計力学は分布を扱う学問です。

$$p(\{x_i\}) = \frac{1}{Z} \exp\left\{-\frac{U(x_i)}{k_b T}\right\}$$

またダイナミクスは時系列を扱うわけです。

$$\{x_1(t), p_1(t), x_2(t), p_2(t), \dots\}$$

この2つを関係付けるには、Landscape を最適化するような視点、例えるならばGoogle maps と Google street view (Daynamics) のような関係と捉えるのが良いでしょう。つまりDaynamics を Landscape 的にみるわけです。関連する数理としては、経路積分, Path Ensemble, transition path sampling (Braids (C. Moore)), 最適化を Dynamical に行うなどかキーワードになると思います。伊庭「無時間の思想」が参考になります。

1.4 本質的なダイナミクスについて

次に、Essential Dynamicsがあるのかを考えたいと思います。タンパク質の構造解析を、大自由度力学系のまま扱う必要があるか、それとも「Dynamicsの本質」を抽出するのか、という問題があります。仮に、時間スケールに開きがあれば、断熱近似できるかもしれませんが、そもそも Dynamics の本質というものが抽出できるのでしょうか？

例えば、タンパク質のダイナミクスの主成分の分布を見てみますと、第10主成分以下はほとんどガウシアンになってしまいます。このようにタンパク質の構造ダイナミクスを低次元化するモデルとしては、PCA, GMM, Markov Chain, e-Machine, Coarse-Grain Models, Go model (高田さん@京大) などがあります。

さて、そもそも、Dynamicsはどこまで必要なのでしょうか。タンパク質は、熱ゆらぎがありますが運動が拡散的です。そもそも生物がこのようなゆらぎを使っていないと考えて、人間が勝手に、くりこみをしているようなものですが、このような仮定が正しいのか、わかりません。

1.5 測定分野と数理的研究

次に、タンパク質の構造決定の分野とそこに関わる数理について紹介します。測定データである電子密度をフーリエ変換して強度スペクトル $|S(\vec{k})|^2$ を解くという問題は、いわゆる逆問題です。X線自由電子レーザーにより1分子の回折像が得られるようになり、これは非常に革命的 (原研 郷先生) です。

1.6 理論系と実験系

最後に、生物における理論系の悩みについてまとめます。これは「モデルに十分自信を持てるか」、「理論だけで説得できるか」、「理論に普遍性をもたせられるか」という3つの悩みに大別できます。これはすべて達成できていないと思います。理論には一定の予測能力はあり実験で見えないものも見えます。理論の研究者としては、実験の後追いにならないような仕事したいと思います。例としては、タンパク質の設計問題などはそのひとつだと考えています。

1.7 質疑

質問1: 細かいダイナミクスをみる必要が本当はないのでしょうか回答1: 遠くの軌道に移る問題が難しいです。これはピコ秒で起きる現象だが、とんでいくのはミリ秒オーダーです。これを扱うのが難しい。

質問2: 時計・砂時計的ゆらぎと構造変化の関係は調べられていますか回答2: それはいわゆる高次元力学系でやられています。ダイナミクスと kernel PCA の統合や3自由度の力学で扱えると思います。

質問3: 計算機の能力向上があっても、数学にできて計算機にできないことがありますか回答3: 生物が有限なので計算機向上すればできようになると思います。一方、進化などは数学の問題として残ると考えています。

質問4: 不変量をつくるとかは数学の仕事ではありませんか回答4: 不変量は好きではないです。うまい座標系を取るもののほうが重要だと考えています。計算機から出てくるデータは時系列なので、データ解析のような手法は重要だと思います。

質問5: 情報を捨てる、というトピックスがあるがそれをコンピュータができるのでしょうか回答5: それに関連した問題として、シミュレーションのまわりに面白い数理があります。日本でももっと研究すべきですが、あまりされていないようです。

質問6: まれなダイナミクスの遷移の特徴はありますか? 回答6: まず計算してみる必要があります。しかし計算量がとてもかかります。誰かやりませんか?

質問7: 近藤: わかるとは? 回答7: 非線形問題のわかりかたには理論枠組みがあります。ちよつとくりこみで、ちよつと厳密解。シミュレーションの予測性で評価という考えかたもあります。

コメント: 杉浦さんの解析の授業に出たときに、彼が、代数はわかるが解析はわかった気になりませんね(笑) と言っていたが、いまは気持ちがわかります。人間にわかるのは線形代数ぐらいなのではないでしょうか?

質問8: X線自由レーザーについて実現性はありますか? 回答:8 理論的方法があるので実現性があると思います。