

若手研究者インターナショナル・トレーニング・プログラム(ITP)

バイオインフォマティクスとシステムズバイオロジーの国際連携教育研究プログラム 応募書類

Name: 鎌田 真由美
Title: タンパク質の立体構造揺らぎの時系列解析と機械学習による化合物分類に関する研究
Institute: 京都大学化学研究所バイオインフォマティクスセンター
Partner institute of your choice : Macromolecular Modeling Group, Free University of Berlin
Duration of your choice: 平成 22 年 9 月 27 日～平成 22 年 12 月 20 日
Plan : <u>滞在目的</u> 今回私が滞在を希望する Free University of Berlin の Knapp 研究室では、現在私が所属している阿久津研究室と同様、計算機を用いた生体高分子に対する解析・予測の計算手法の開発を行っている。阿久津研究室では情報科学の基礎理論（アルゴリズム論、学習理論）および数理科学の観点に基づいて手法の開発を行っているのに対し、Knapp 研究室では、高分子の化学物理学的な特性をベースとしている。生体高分子の大規模な系に対する解析手法において、精度や計算資源の点から、自然科学と情報科学双方に基づいた手法の開発は必須である。ゆえに、自然科学をベースに手法の開発を行う Knapp 研究室への滞在は、上記のような手法開発に求められる視点や知識を養うに最適であると考えられる。 また Knapp 研究室では、分子動力学に基づくシミュレーション・解析用パッケージである CHARMM を用いた研究を、その開発者と共同で行っている。この分子動力学 (MD) シミュレーションは、これまでの私の研究テーマ（下記参照）であるタンパク質の動的挙動に大いに関連している。更に、Knapp 研では創薬デザインの為の機械学習を用いた化合物分類なども行っている。阿久津研でも機械学習を用いて様々な研究を行っており、カーネル関数の設計なども行っている。 現在多くの創薬コンセプトにおいてタンパク質の構造情報が活用されており、構造情報を直接創薬に応用する Structure Based Drug Design などもある。これにも当然、構造の動的性質の考慮が必要になると考えられる。故にこれまで私が行ってきたタンパク質の動的挙動に関する研究と現在の所属している阿久津研での機械学習に関する研究、そして Knapp 研での物理化学的な生物情報学に関する研究すべてを包括し考慮した研究発展は、創薬研究においてとても有用であると考えられる。以上のことから今回の滞在によってこれまでの研究の発展だけでなく今後の新たな研究課題への取り組みにおいても有益な議論・研究が行えると考え、Knapp 研への研究滞在を強く希望する。
<u>研究計画</u> 私はこれまで経時的なタンパク質の動的挙動と機能との関係性を主なテーマとし、時系列解析などを行ってきた。近年の研究により、生体内におけるタンパク質のダイナミクスが様々な機能発現において重要な役割を果たしていることが明らかにされており、タンパク質のダイナミクスについて詳細な解明が求められている。タンパク質のダイナミクスには、フェムト秒～ピコ秒で起こるような個々の原子の振動から、マイクロ秒～秒で起こるフォールディングなど様々なタイムスケールでの動きが含まれている。タンパク質の機能発現や複合体形成に関連するような骨格構造や二次構造での変化は、ナノ秒～マイクロ秒のタイムスケールで起こるとされている。しかし、このようなタイムスケールで

の動きは、基準振動解析で表現できるような振幅一定の動きだけではなく、複数の振幅が入り混じった非調和的な動きをしている。これまで私はこのような非調和性を含んだタンパク質の動きをターゲットとし、MD シミュレーションによる時系列データから特徴を抽出して時系列解析を行ってきた。しかし、上記で述べたような複合体形成などに関するナノ秒～マイクロ秒での時系列データには適用しておらず課題として残っていた。そこで実際に MD シミュレーションなどを行っている Knapp 研で長時間の MD シミュレーションを行い、これまでの解析手法への適用・検証を行いたい。

また Knapp 研ではドッキング予測などに用いる化合物分類の研究も行っている。分類には機械学習を用いており、新たに与えられた対象がどのクラスに属するのか(分類されるのか)を予測する。学習データや予測データをうまく表現しようとするとその特徴ベクトルは膨大な数となる。しかし実際にはそのすべてが重要な特徴というわけではなく、更に、データセットよりも大きな特徴セットは過学習と呼ばれる問題を引き起こす要因ともなる。この過学習を回避するための手法に正則化というものがあり、正則化による予測精度の向上や不要な特徴の制約も期待される。タンパク質のダイナミクスを考慮した解析や予測に膨大な数の特徴が必要となることは明らかであり、今後の研究発展に機械学習を用いたいと考えていることから、学習における正則化について学び議論を行いたい。そして新たな分類手法の提案を行っていきたいと考える。

以上のことから、MD シミュレーション（これまでの研究課題の発展）と化合物分類に関する研究（新たな研究課題へ取り組み）の2つを軸に、各々以下にあげる点を目標に研究を行っていきたい。

I) 長時間分子動力学シミュレーションによるタンパク質構造変化の時系列解析

- ・ 10ns 以上の MD シミュレーションの実行
- ・ シミュレーション結果をこれまで行ってきた時系列解析への適用し、今後の発展に向けた考察

II) 創薬デザインに向けた化合物分類手法の開発

- ・ 機械学習における正則化に関する議論と新たな手法の提案
- ・ 提案手法の実装と検証

また、Knapp 研究室での定期セミナーなどにも参加し、様々なバックグラウンド・国籍を持つメンバーとの交流を通して、研究において積極的な議論を可能にするコミュニケーション能力の取得も目標とする。そして、今後の研究発展に繋がるよう様々なことに自発的に挑戦し、3ヶ月間有意義に研究を行っていきたい。