

プロジェクト名(タイトル): 深層学習などによる NMR データおよび生体高分子の解析

利用者氏名: 小林直宏

理研における所属研究室名: 生命機能科学研究センター先端 NMR 開発・応用研究チーム

### 1. 本プロジェクトの研究の背景、目的、関係するプロジェクトとの関係

HOKUSAI のような大規模 PC クラスタの利用は、NMR データ解析の自動化および高度化、深層学習ツール開発への応用に加え、MD 計算など複合的な活用を目的としている。NMR 解析法は現状として他の構造解析法である X 線結晶解析法、Cryo 電顕法に比して自動化・効率化について大きく遅れをとっている。代表者による研究プロジェクトは HOKUSAI に搭載された 1 ノード Intel 40 コアを有効活用することで NMR 信号帰属の自動化や構造モデル計算の高度化を実行し、NMR 解析自動化の効率化を目的とした。

### 2. 具体的な利用内容、計算方法

代表者は NMR 化学シフトを既知構造座標から推定する深層学習器の開発に成功している。HOKUSAI では学習済みモデルを利用して多数のベンチマーク計算を行う。深層学習ツールとしてはマイクロソフトが無償公開している CNTK を用いる。また、タンパク質の動的状態を NMR 実験データとの整合性を得るために、Amber20 による計算を行う。具体的には一般化ボルン(GB)によるアニーリング法による構造最適化、Sander.MPI あるいは PMEMD.MPI による explicit water での等温等圧(NPT)計算を行って、構造変化の統計的解析を行う。更に、レプリカ交換法を用いた MD 計算(REMD)により、NMR 実験データと動的状態の議論も行う。

### 3. 結果

化学シフトを深層学習によって構造座標から推定させるシステムは既に完成しているが、その推定精度を評価するベンチマークテストには HOKUSAI のような大規模計算機なくして短期間で完了することは無かったと言える。昨年 6 月より開始した計算結果をまとめ、現在はこれらの解析結果と共に国際誌へ投稿すべく鋭意執筆中である。

また、昨年末より開始したレプリカ交換法による MD 計算はラボスケールではとても実現不可能なスケールでの計算を行う事が出来ている。

### 4. まとめ

代表者により開発された構造座標から化学シフトを高精度

に予測する深層学習器は完成し、その性能を評価するためのベンチマーク計算は完了できた。また、新たなステップとして REMD 計算を組み入れた動的状態の解析へコマを進める事が出来た意義は大変大きい。HOKUSAI における Amber18, Amber20 の計算機環境(ライブラリ等)の適切な整備と大規模計算資源を提供いただいた事に感謝したい。

### 5. 今後の計画・展望

現在、200~400 コアを用いた小規模の REMD 計算を行っているが、これらの有効性を確認すれば他の大型計算機(富岳など)へ移行して更に大規模な REMD 計算が可能であると期待している。実際に、数 10nsec (40k atoms) の規模の REMD では NMR 実験で得られるミリ秒スケールの現象説明には不十分と言える。理研計算機施設である富岳、RAIDEN, HOKUSAI を上手く活用しつつ、これまでに実現困難であったより大規模な生体高分子系をミリ秒スケールで議論できれば極めてインパクトの高い研究成果につながるだろう。

### 6. 利用がなかった場合の理由