

28 MDDDB-AI Hub: MDデータの共有と再利用を支えるデータ・モデル基盤の構築

○松永康佑^{1,2}、小林千草²、Diego Ugarte La Torre^{2,3}、坪内秀人³、川合 毅¹、杉田有治^{2,3}

¹ 埼玉大学 ² 理化学研究所 ³ 東京大学

背景と目的

分子動力学（MD）シミュレーションは、生体分子の「動き」を原子レベルで捉え、静的構造だけでは見えない機能の理解や創薬標的の探索に役立つ。一方、大規模なMDデータは研究室ごとに個別管理され、登録・共有・再利用の基盤が十分に整っていない。さらに、MDを高速に代替するAIサロゲートモデルの開発には、機械学習に利用しやすい訓練データと、性能を公平に比較できる評価基盤が求められる。

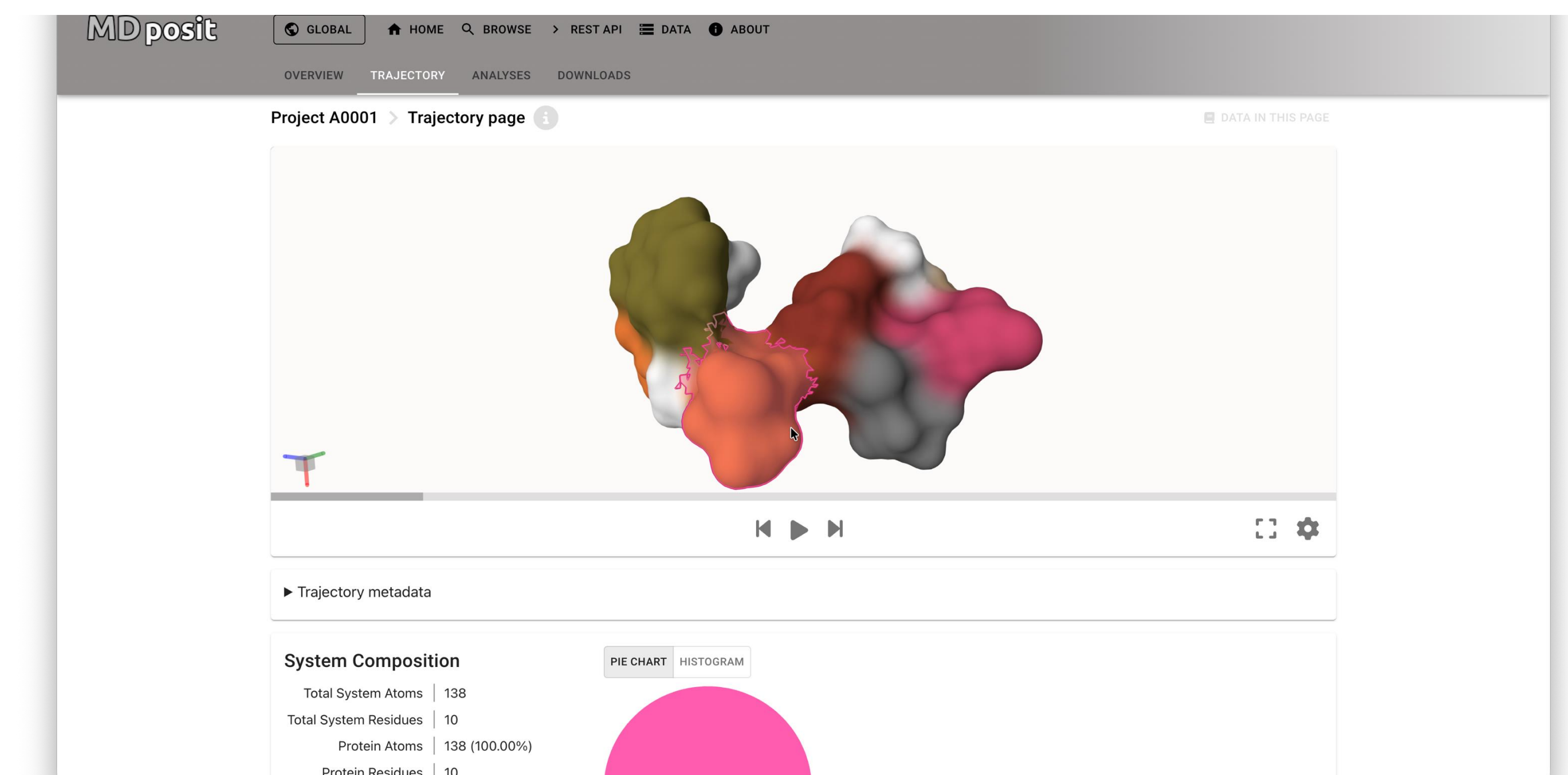
そこで本課題では、欧州のMolecular Dynamics Data Bank（MDDDB）と連携し、日本・アジア太平洋地域の拠点となるMDDDB-AI Hubの構築を進めている。実験研究者にはWeb上での動態閲覧・解析を、計算科学者にはMDデータの登録・公開を、AI研究者には訓練データ・学習済みモデル・ベンチマークを提供することを目指す。

MDDDB-AI Hubの全体構想



データ登録・可視化の進捗

MDDDBのローカルテストベッドを構築し、Docker環境でデータの登録から閲覧までの一連の動作を検証した。さらに、Mol*を用いたトラジェクトリ可視化機能を実装し、Web上で動きを閲覧できるようにした。



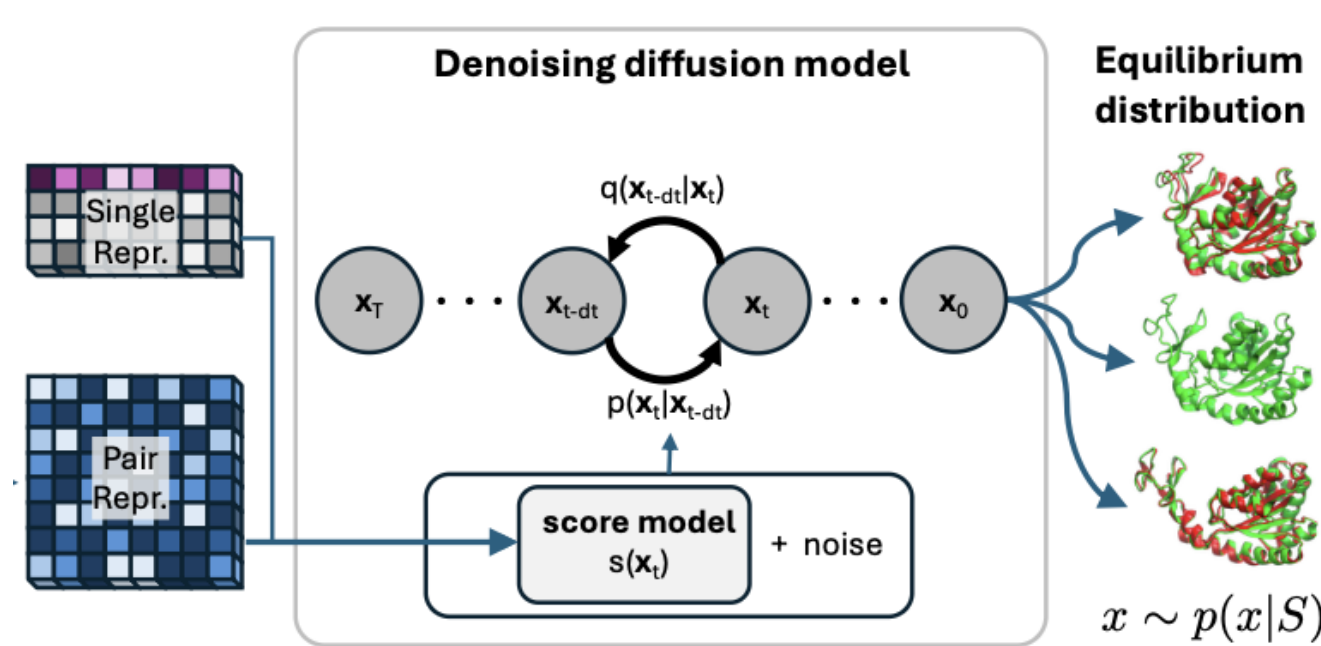
登録したトラジェクトリの表示例

HPCI共用ストレージへの試験展開を進めている。欧州MDDDBとの接続方式や管理・権限分担を検討し、外部公開に向けた運用体制を整える。

サロゲートモデルの評価基盤

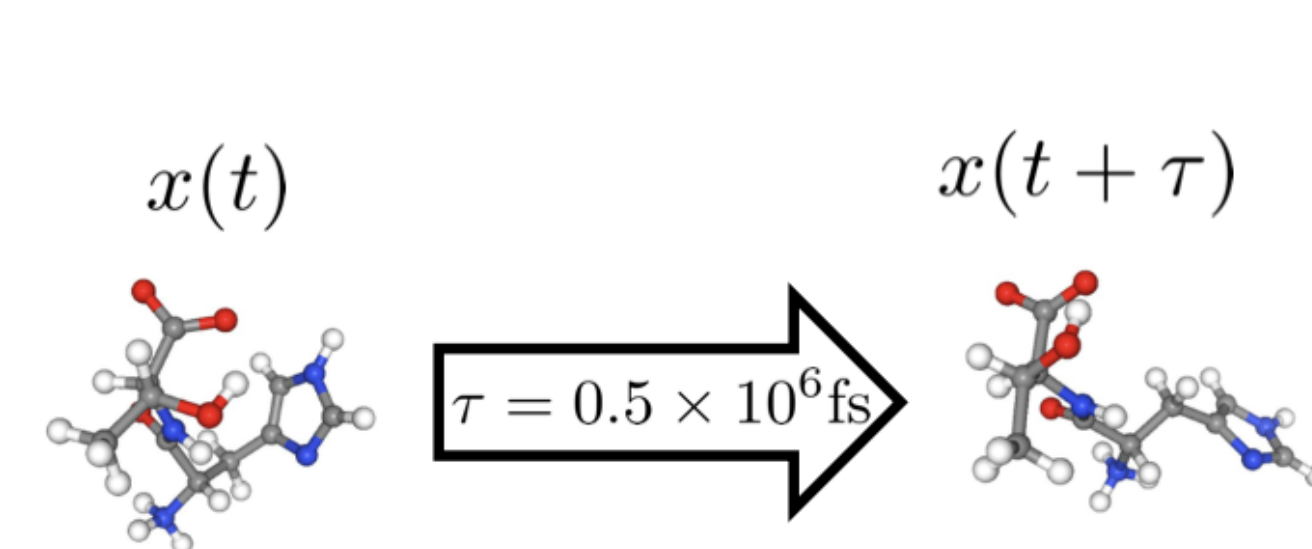
MDサロゲートモデルは、MDデータを学習し、構造の分布や時間を隔てた構造を高速に生成するAIモデルである。以下のように出力の目的が異なる。

平衡分布から構造を生成



例：BioEmu（平衡構造群）
Lewis et al., Science (2025)

時間を隔てた構造を提案



例：Timewarp（遷移の提案）
Klein et al., NeurIPS (2023)

目的に応じた共通の評価軸

構造妥当性	結合長・結合角・立体衝突
計算効率	構造生成に要する計算時間
平衡分布	構造の出現確率の再現性
動力学	状態間の遷移の速さ
構造被覆	多様な構造をどこまで捉えるか

モデルの目的に応じて評価項目を選び、公平に比較。共通ベンチマークと暫定リーダーボードを設計中。

まとめと今後の展望

MDデータの登録・動態閲覧を基盤に、ML-Readyな訓練データ、学習済みモデル、評価基盤を整備する。今後は、GENESISで取得した拡張サンプリングやマルチスケールのデータも活用し、共有・再利用を広げる。欧州MDDDBとの国際連携とHPCI環境への展開を進め、実験・計算・AI研究者をつなぐ基盤を目指す。

謝辞

本研究は、JST NBDC 統合化推進プログラム（ Grant 番号：JPMJND2603 ）の支援を受けたものである。

