様 式 C－19、F－19－1、Z－19 （共通）

科学研究費助成事業　　研究成果報告書

令和 5 年 6 月 13 日現在

機関番号：８２４０１
研究種目：基盤研究(C)（一般）
研究期間：2019～2022
課題番号：１９Ｋ０６５３５

研究課題名（和文）全原子モデルの併用と機械学習手法の適用による粗視化分子モデルの高精度化

研究課題名（英文）Improving Coarse-Grained Molecular Models through the Combination of All-Atom Models and Machine Learning Methods

研究代表者
金田　亮（Kanada, Ryo）

国立研究開発法人理化学研究所・計算科学研究センター・上級研究員

研究者番号：４０４２３１３１

交付決定額（研究期間全体）：（直接経費）　2,100,000円

研究成果の概要（和文）：構造ベースの粗視化MDは全原子MDに比して計算コストが軽く生体分子の長時間動態の調査に適している。しかし全原子MDに比べて力場精度が低く、相互作用に関連するモデルパラメータに曖昧さが残る問題が存在する。また、力場構築の際に活用する参照構造周辺にサンプリング構造がトラップされる傾向がある。
これらの問題を解消する為、短時間の全原子MDで得られた動的揺らぎの情報と機械学習手法を組み合わせた新しい粗視化分子シミュレーション技法を構築した。この手法を適用する事で、大きな蛋白系の多様な構造アンサンブルを従来の全原子MDや粗視化モデルに比して効率的かつ探索的にサンプリングする事に成功した。

研究成果の学術的意義や社会的意義
本研究で開発した新しい粗視化分子シミュレーション技法を適用する事で、実験で観測される限られた構造を起点にして、幅広い構造アンサンブルを効率的に探索・サンプリングする事が出来る。この事は、標的蛋白質の長時間動態や多様な順安定状態の解明、相互作用機構の理解等が、必要不可欠な実践的創薬の分野において意義がある。特に標的蛋白質のサイズが大きい場合は、全原子MDシミュレーションよりも計算コストが低減できる為、効率的な創薬開発に寄与する可能性が期待される。

研究成果の概要（英文）：Structure based coarse-grained (CG) molecular dynamics (MD) offers a computationally efficient approach for investigating the long-timescale dynamics of biomolecules compared to all-atom molecular dynamics (AAMD). However, conventional CGMD suffers from lower force field accuracy and ambiguity in model parameters related to interactions. Additionally, there is a tendency for sampling structures to be trapped near the reference structure.
To address these issues, we developed a new CG-model by combining short-timescale AAMD simulations, capturing fluctuation correlations, and machine learning techniques to enable efficient model-parameter exploration. By applying this new model, we successfully achieved efficient and exploratory sampling of diverse structural ensembles in large protein systems, surpassing the sampling performance via conventional AA-MD and CG-MD.

研究分野：生物物理

キーワード：粗視化分子モデル　全原子モデル　機械学習　ベイズ最適化　能動学習

※科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属します。

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

近年の実験技術の発展により生命活動を担う様々な生体高分子の構造が解かれつつある。しかし機能発現に関わる全ての準安定構造を実験のみで精度良く取得する事は困難である。生体分子の時間発展を調査するアプローチとして全原子 MD シミュレーションが存在するが、現在の全原子 MD シミュレーション技術では ANTON 等を用いても機能発現に至るまでのミリ秒～数秒のオーダーの長時間動態を調査・再現する事は計算コストの観点から難しい。その為、蛋白質を構成する各アミノ酸に対して 1 個の自由度のみを付与する様な粗視化分子シミュレーションによるアプローチが必要となる。粗視化分子動力学シミュレーションは全原子分子動力学シミュレーションに比して計算コストを大幅に抑える事が出来る為、将来的にはより実践的な創薬応用に向けて有用な計算手法となる事が期待される。しかしながら、現状の構造ベースの粗視化分子シミュレーションの場合、その精度は全原子力場に比べて低い。特に分子間相互作用に関連するパラメータに曖昧さが残る。また、粗視化力場構築の際の参照となる結晶構造周辺にサンプリング構造がトラップされる傾向を持ち、サンプリング効率があまり高くないという問題も存在する。その為、これらの問題を解消する為に全原子モデルと粗視化モデルを組み合わせた新しいシミュレーション技法の構築が望まれる。

２．研究の目的

現状の構造ベースの粗視化分子モデルにおける力場の精度の低さ、相互作用パラメータの曖昧さ、サンプリング性能に関する問題等を解決・改善する為に、全原子モデル及び機械学習手法を組み合わせた新しい粗視化分子シミュレーション技法を確立する事。

３．研究の方法

(1) 粗視化分子モデルにおける相互作用パラメータの曖昧さはパラメータ空間が広大である事に起因する。本研究では、広範なパラメータ空間から実験で観測されている生体分子の動態を定性的に再現する事が出来る領域(successful パラメータ領域)を見出す為に、二つの機械学習：ベイズ最適化(BO: Bayesian Optimization)と能動学習手法(US: Uncertainty Sampling)を組み合わせた探索手法(BOUS)を適用し、効率的な探索を試みた。

(2) 構造ベースの粗視化分子モデルにおける、サンプリング構造が参照構造周辺にトラップされる問題を解決する為に、短い時間スケール(ns オーダー)の全原子 MD シミュレーションのトラジェクトリから評価される全残基ペア間の動的揺らぎの相関(DCCM: Dynamic correlation coefficient map)情報を活用して新しい適応型の粗視化弾性バネモデル(adaptive Coarse-Grained Elastic Network Model: adaptive CG-ENM)を構築した。この適用型の粗視化弾性モデルを中規模系(200 残基程度)や大規模系(1000 残基程度)に適用し、構造サンプリングの性能を従来型の全原子モデルや粗視化モデルと比較・調査した。(adaptive CG-ENM における残基ペア間のバネ定数の強度は機械学習手法(ベイズ最適化)により効率的に探索・チューニングをした。)

４．研究成果

(1) 機械学習手法によるパラメータ探索のターゲットとして、生体分子モーターF1-atp ase の ATP 加水分解反応(ms スケール)に伴う中央の回転子(γ-サブユニット)の回転運動に焦点を当てた。F1 分子モーターは、ATP 加水分解に伴い、回転子をほぼ確実に約 120 度反時計回りに回転させる。この F1 モーター系に対して構造ベースの粗視化分子モデル(Off-lattice Go モデル)を構築し、力場ポテンシャルを化学反応の進行に伴ってスイッチさせる事で回転運動の定性的な再現を目指す。特に中央の回転子(γ-サブユニット)と周りのαβサブユニット(固定子)の間の斥力相互作用の強度：EVI、溶媒の温度：T 等から成るパラメータ空間の中で、粗視化分子モデルの回転運動が定性的に成功する領域(successful 領域)を探索した。

その結果、二つの機械学習手法を組み合わせた BOUS 法は、網羅的探索やランダムサンプリング法(RS 法)のみならず、ベイズ最適化単独(BO 法)や能動学習手法単独(US 法)と比して有意に少ない探索回数で成功領域を探索出来る事が示された。図 1 は典型的な探索の例である。この図において、他の RS 法や BO 法では 80 回の探索をしても成功領域を全て見出す事が出来ていないのに対して、BOUS 法は 41 回で全ての成功領域を探索出来ている。更に、定量的な性能比較の為に、初期の選択パラメータセット（10 個）を 100 回ランダムに変えて探索を試行し、その 100 回の試行による平均値を評価した。図 2 は検知された成功領域(successful パラメータ)のパラメータ探索回数依存性を表す。この図から BOUS 法は US 単独法、BO 単独法や RS 法に比して有意に探索効率が高い事が見て取れる。(BOUS の探索コストは RS や網羅的グリッドサーチのコストの約 12%程度に抑えられた。)

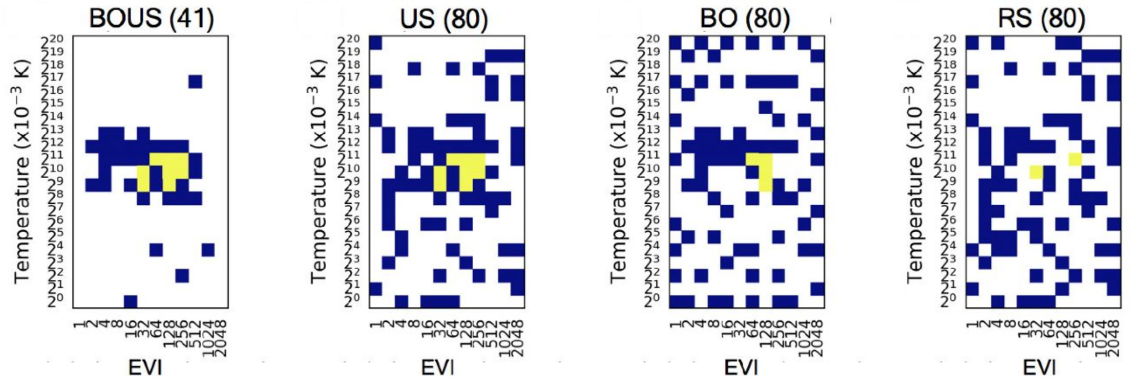

図１．BOUS 法、US 単独法、BO 単独法、RS(ランダムサーチ)法を用いた成功領域の典型的な探索の例。カッコの中の数字は各々の手法でパラメータサーチされた回数を示す。サーチされたパラメーターの内、青色の領域は回転が失敗する領域、黄色の領域は F1 モーターの回転が成功する領域を表す。(本図は発表論文より抜粋したもの)

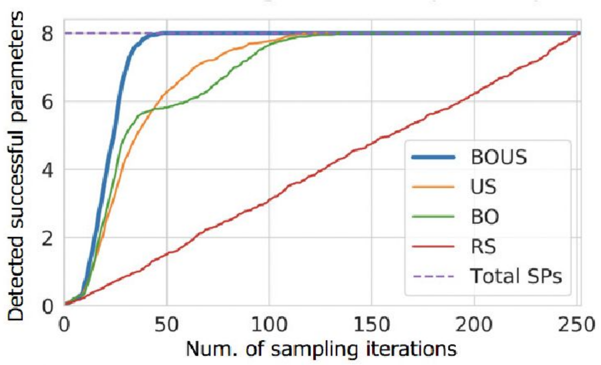

図 2 BOUS、US、BO 法と RS 法を用いた粗視化分子シミュレーションの成功領域(successful パラメータ)探索の性能。横軸がパラメータ探索回数、縦軸が検知された成功パラメータ数の平均。(本図は発表論文より抜粋したもの)

二つの機械学習手法を組み合わせる事で実現された効率的なパラメータ探索：BOUS は F1-atpase 系に限らず他の粗視化分子モデルに対しても幅広く一般的に適用可能である。広範なパラメータ探索に基づきモデルパラメータの曖昧さを排する事で、今後、粗視化分子シミュレーションにより新しい分子設計・デザインの指針等を与たり、機能発現や構造変化ダイナミクスの環境依存性調査をする事が可能になる。

(2)新しい粗視化分子モデル：adaptive　CG-ENM の適用対象としてまず、中規模分子系の adenylate kinase（ADK）及び glutamine binding protein（GBP）を選択した。高々ns 程度の短時間の全原子 MD シミュレーションのトラジェクトリに基づき DCCM を評価した結果が図３である。図から短い全原子 MD により対象系蛋白質の硬いドメインと柔らかい領域を十分に特定・判別できる事が示唆される。

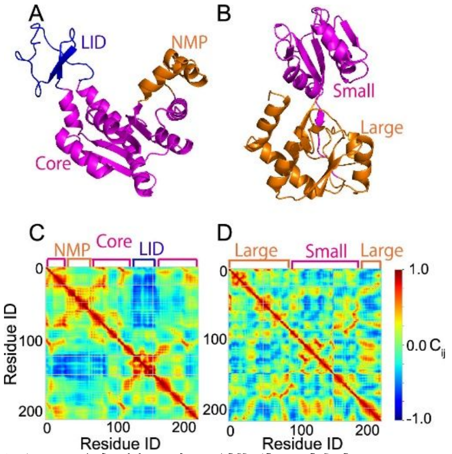

図 3 蛋白質の参照構造と対応する DCCM。(A、B) ADK と GBP の初期構造。実験的に示唆されるドメイン毎に色を分けている。(C、D)対応する ADK と GBP の DCCM。(本図は発表論文より抜粋したもの)

DCCM に基づき adaptive　CG-ENM の最適パラメータセットを探索・特定する際は、ベイズ最適

化を適用した。蛋白のドメイン内の構造安定性をある程度保持しつつ幅広い構造アンサンブルを実現する為に、パラメータ探索の目的関数として蛋白の Q-score の時間平均値と RMSD の分散の積を採用した。結果、パラメータ探索コストを RS や網羅的探索法に比べて大幅に抑える事が出来た。

パラメータ最適化された adaptive CG-ENM の構造サンプリング性能を調査する為､従来型の AA-MD や構造ベースの粗視化 MD で得られた構造アンサンブルとの比較をした。図 4A から adaptive CG-ENM による構造アンサンブルは TREMD による従来型の粗視化モデル(ENM)や AA-MD による 1μs の構造アンサンブル群に比して十分に幅広い事が見て取れる。更に図 8B に示す様に adaptive CG-ENM は基準となる apo 状態から遠く離れた holo 状態に近い構造（ターゲット構造）もサンプリング出来ている事が見て取れる。サンプリングされた構造の内、最も holo（ターゲット）に近い構造は図 4C の S1 構造であるが，この構造 S1 の holo 構造に対する RMSD 値は僅か 3.8Åとなっており有意に小さい。また，図 4C の S1-S5 の様に holo 構造や apo 構造に似た構造だけでなく、伸長した構造等、多様性のある構造群が adaptive CG-ENM によりサンプリング出来ている事が確認された。

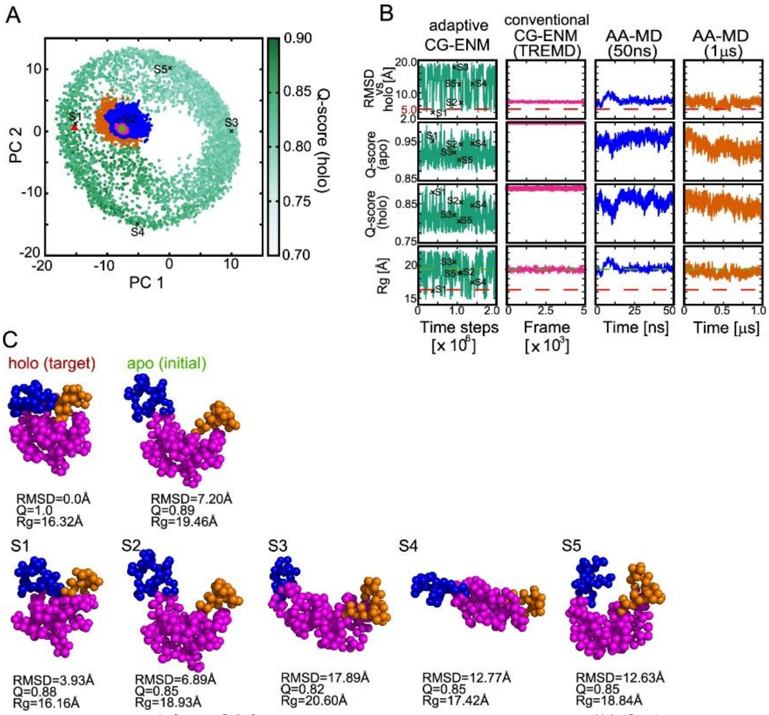

図 4 ADK 蛋白を対象とした adaptive CG-ENM と従来型の CG-ENM 及び全原子シミュレーションの比較。(A) 主成分解析の PC12 平面における構造アンサンブルの比較。Adaptive CG-ENM、従来型の CG-ENM、AA-MD（50ns と 1μs）のサンプリングポイントが各々、緑、マゼンタ、青、オレンジで色付けされている。(B) ターゲット構造（holo）に対する RMSD, Q-score（apo, holo）と慣性半径（Rg）の時系列。(C) 初期構造（apo）とターゲット構造（holo）と adaptive CG-ENM でサンプリングされた代表的構造群（S1-S5)。(本図は発表論文より抜粋したもの)

Adaptive CG-ENM による多様性のある構造サンプリング、及び従来型の CG-ENM や AA-MD に対するサンプリングの優位性は ADK 蛋白だけでなく、GBP 蛋白やより大きな 1000 残基程度の蛋白系(Integrin 等)においても同様に確認された。この事は新しい粗視化シミュレーション技法(adaptive-CGENM)が実験的に未知の構造を探索的に調査する際に広く適用可能である事を示唆する。今後の展開として、この新しい粗視化力場の精度を AA-MD の構造アンサンブルと対応する(全原子力場の)エネルギーを教師データとして学習する事で向上させ、実践的な創薬応用への適用を目指す。

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　2件）

１．著者名 Kanada Ryo、Terayama Kei、Tokuhisa Atsushi、Matsumoto Shigeyuki、Okuno Yasushi	４．巻 18
２．論文標題 Enhanced Conformational Sampling with an Adaptive Coarse-Grained Elastic Network Model Using Short-Time All-Atom Molecular Dynamics	５．発行年 2022年
３．雑誌名 Journal of Chemical Theory and Computation	６．最初と最後の頁 2062～2074
掲載論文のDOI（デジタルオブジェクト識別子） 10.1021/acs.jctc.1c01074	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

１．著者名 Kanada Ryo、Tokuhisa Atsushi、Tsuda Koji、Okuno Yasushi、Terayama Kei	４．巻 10
２．論文標題 Exploring Successful Parameter Region for Coarse-Grained Simulation of Biomolecules by Bayesian Optimization and Active Learning	５．発行年 2020年
３．雑誌名 Biomolecules	６．最初と最後の頁 482～482
掲載論文のDOI（デジタルオブジェクト識別子） 10.3390/biom10030482	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 －

〔学会発表〕　計5件（うち招待講演　0件／うち国際学会　0件）

１．発表者名 Seiryo Sugiura, Ryo Kanada, Takumi Washio, Xiaoke Cui, Jun-ichi Okada, Yasushi Okuno, Toshiaki Hisada
２．発表標題 A Heart simulator coupling the molecular dynamics with the finite element model: Cross-scale integration of our knowledge on heart
３．学会等名 The 50th NIPS International Symposium ‘MIRACLES’ in Cardiovascular Physiology -Metabolism, Interactions, Regulation, Application, Chemical Biology, Longevity, Exercise and Signaling-
４．発表年 2019年

１．発表者名 Ryo Kanada, Takumi Washio,Seiryo Sugiura,Jun-ichi Okada,Shoji Takada,Yasushi Okuno,and Toshiaki Hisada
２．発表標題 Cross-linking of UT-Heart simulator and coarse-grained molecular simulation
３．学会等名 LSACJ2019
４．発表年 2019年

１．発表者名 Ryo Kanada, Kei Terayama, Atsushi Tokuhisa, Shigeyuki Matsumoto, and Yasushi Okuno
２．発表標題 Efficient Conformational Sampling with an Adaptive Coarse-Grained Elastic Network Model using Dynamic Cross-Correlation Coefficient
３．学会等名 The 60th Annual Meeting of the Biophysical Society of Japan
４．発表年 2022年

１．発表者名 Ryo Kanada, Kei Terayama, Atsushi Tokuhisa and Yasushi Okuno
２．発表標題 Efficient Conformational Sampling with an Adaptive Coarse-Grained Elastic Network Model using Bayesian Optimization
３．学会等名 The 5th R-CCS International Symposium
４．発表年 2023年

１．発表者名 Ryo Kanada, Kei Terayama, and Atsushi Tokuhisa
２．発表標題 Efficient Parameter Search for Coarse Grained Simulation by Machine Learning Methods
３．学会等名 The 4th R-CCS International Symposium
４．発表年 2022年

〔図書〕　計0件

〔産業財産権〕

〔その他〕

－

６．研究組織

	氏名 （ローマ字氏名） （研究者番号）	所属研究機関・部局・職 （機関番号）	備考

７．科研費を使用して開催した国際研究集会

〔国際研究集会〕　計0件

８．本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関