

#### マルチコピーシミュレーションによる タンパク質構造サンプリング

〇松永康佑、寺田透、森次圭、古田忠臣、木寺 詔紀

#### 理化学研究所 次世代計算科学研究開発プログラム 分子スケール研究開発チーム

2011年2月16日 2010年度HPCに関する理研シンポジウム

### Platypus-MM/CGの設計思想

ユーザが自由にオリジナルのシミュレータを構築

MM/CGシミュレーションを構成する、データ構造・相互作用・アルゴリズムをライブラリ化

マルチコピーシミュレーションを容易に実現

シミュレーションシステムをオブジェクト化

## オブジェクト指向(1)

- プログラムを相互に「メッセージ」を送りあう
   「オブジェクト」の集まりとみなす
- 「オブジェクト」はそれに関連するデータと手 続き(メソッド)の集まり



## オブジェクト指向(2)

オブジェクト指向は、
 コードの再利用が容易

 →オリジナルのシミュレータを構築する上で有利
 複数のインスタンスを生成できる
 →マルチコピーシミュレーションを容易に実現
 コードの修正の影響が及ぶ範囲が限定される
 →共同作業に適している

次世代スーパーコンピュータに標準搭載されるオブジェクト指向言語はC++
 →C++を使ってコード

作成済みのクラス

Dynamics





ベンチマーク on RICC

#### ストリング法のベンチマーク



系: Adenylate kinase + 溶媒, 62,475原子 strong scaling: 256レプリカ固定

## 成果1: Multi-Scale Essential Sampling 法

Brute-force MDの代わりに、CGモデルとカップリングした多数 のレプリカMD計算を並列に動かすことでサンプリングする

カップリングの強さは異なる



### Hamiltonian Exchange

 $H(\mathbf{x}, \mathbf{z}) = K(\mathbf{p}_{\mathbf{x}}) + K(\mathbf{p}_{\mathbf{z}}) + V(\mathbf{x}) + U(\mathbf{z}) + kW(\mathbf{z}, \theta[\mathbf{x}])$ 

Time *t*:



keeping detailed balance condition for canonical distribution

transition probability between n and  $m = \min(1, \exp(\Delta_{mn}))$ 9

$$\Delta_{mn} = \beta \left( k_m - k_n \right) \left\{ W(\theta[\mathbf{x}_m], \mathbf{z}_m) - W(\theta[\mathbf{x}_n], \mathbf{z}_n) \right\}$$

## MSES: Summary

 $H(\mathbf{x}, \mathbf{z}) = K(\mathbf{p}_{\mathbf{x}}) + K(\mathbf{p}_{\mathbf{z}}) + V(\mathbf{x}) + U(\mathbf{z}) + kW(\mathbf{z}, \theta[\mathbf{x}])$ 

$$W(\mathbf{z}, \boldsymbol{\theta}[\mathbf{x}]) \equiv \frac{1}{2} \left[ \mathbf{f} \left( \mathbf{z} \right) - \mathbf{f} \left( \boldsymbol{\theta}[\mathbf{x}] \right) \right]^2$$

in analogy with Maragliano and Vanden-Eijnden, CPL 446, 182 (2007)

 $W(\mathbf{z}, \boldsymbol{\theta}[\mathbf{x}])$ 

- Multiscale simulation: sampling of **x** is accelerated by **z**
- Prior knowledge: U(z) will determine the essential subspace in x of specific interest
- Scalable FES calculation :
  - $k \rightarrow 0$  via Hamiltonian exchange leads to unbiased conformations of **x**
  - exchange probability:  $min(1, exp(\Delta_{mn}))$  is determined simply by

 $\Delta_{mn} = \beta (k_m - k_n) \{ W(\theta[\mathbf{x}_m], \mathbf{z}_m) - W(\theta[\mathbf{x}_n], \mathbf{z}_n) \}$ 

# Example: CG Pulling MM in Solution



## Chignolin in Implicit Solvent



- six replica for  $P(V_{MMCG})$  overlap
- sampled from native to extended struct
- MSES FES in good agreement with multicanonical MD

Terada et al, FEBS 580, 3422, (2006)



#### Sortase A

- a family of Gram-positive transpeptidases
  - functions for cell wall anchored protein (CWAP)
  - recognizes the sorting signal with a conserved LPXTG motif in the CWAP
  - LPXT / G is cleaved and amide-linked to peptideglycan
- intrinsically disordered protein (IDP)
  - database of protein disorder, DisProt (DP00208)

Dunker et al., Nucleic Acids Res. 2007 (Database issue):D786-93 (2006)

- residue 162 to174 (length is 13 in total 226)



Nagamine et al., Chem. Comun., 1022 (2009)



#### Structure of Sortase A



Suree at al., J. Biol. Chem. 284, 24465 (2009)

- NMR structures: sortase A from *S. aureus* 
  - ligand-unbound: 1IJA, 25 models
  - ligand-bound: 2KID, 20 models
     Ca<sup>2+</sup> and peptide analog (LPAT)
- Eight-stranded  $\beta$  barrel fold with two short helices
- N-terminal 58 residues are lacking
  - membrane anchor
  - shown to be unrelated to enzyme activity
- Two loops undergo large structural changes
  - disordered loop (101-119): from disorder to order
  - dynamic loop (124-140): keeps order form



RMSD between the averaged bound and freestructures

#### Comparative Study: Calcium Ion Function



- Bound peptide and Ca<sup>2+</sup> restrict the disordered loop flexibility.
- Bound Ca<sup>2+</sup> decreases entropy change upon peptide binding, favoring binding affinity.

#### **Barster Structures**



bs rmsdMM < 4 Å (gray structure as an example)

成果2: ストリング法

多数のレプリカMD計算を並列に動かし、自由エネルギーの低い 構造変化経路を効率よくサンプリングする

ストリング法によるパスサンプリング



全原子モデル情報をもとにある粗視化自由度で自由エネ ルギー的に有利な構造変化パスを探索する方法であり、タ ンパク質のような大自由度系にも応用可能である





#### Alanine-dipeptideでのテスト



#### Alanine-dipeptideでのテスト



#### Adenylate kinaseへの応用

#### Adenylate kinase (214残基、3343原子)

ATP と AMP から 2 つの ADP を合成する反応を触媒するリン酸基転移酵素 2ADP ⇔ ATP + AMP リガンド結合に伴い大規模な構造変化を生じることが結晶構造解析などによ り示されている



エネルギーの低い反応経路の探索結果



#### Atomistic view of Ligand-AMPbd interaction



成果3: 逐次モンテカルロ法

実験データがあったときに、それを再現するような構造アンサンブルの時間変化(パス)を求める方法



逐次モンテカルロ法

目的

1.1次元の1分子計測実験データから、多数のレプリカのシミュレーションによって 高次元の情報(タンパク質構造情報)を推定し構造変化過程の詳細を得る

2.1分子計測データからより現実に近いシミュレーションモデルのパラメータを推定 する



逐次モンテカルロ法



#### 粗視化モデルでのテスト

逐次モンテカルロ法 によるフィルタリング

タンパク質粗視化モデル



粗視化モデルのシミュレーション により人工的に作成されたFRET時 系列 躍  $y_{1:t} = \{y_1, \dots, y_t\}$ 問  $y_{1:t} = \{y_1, \dots, y_t\}$ 

レプリカの数=1600 赤:元の実験データ 黒:逐次モンテカルロ法で再現したもの





 $x_{1:t} = \{x_1, \dots, x_t\}$ 



#### Adenylate kinaseへの応用

Prof. Haw Yang (Princeton Univ.)との共同研究



distance (A)

J-W. Chu, H. Yang, *Proc. Natl. Acad. Sci. USA* **104**, 18055 (2007)

まとめ

- 近年の超並列計算機を有効に利用するために、Brute-forceシミュレーションの代わりに、統計力学理論に基づいて多数のコピーによりタンパク質構造空間をサンプリングするプログラムを開発した。
- 実際、RICC上(8,000コア超)で動かし、実 証研究を行った。今後K-computer(64万コ ア)上の実証に向けて開発を進めていく。