

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

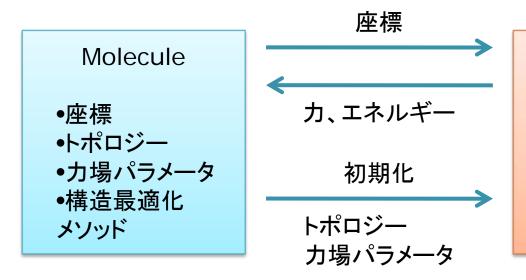
マルチコピーシミュレーションを容易に実現



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



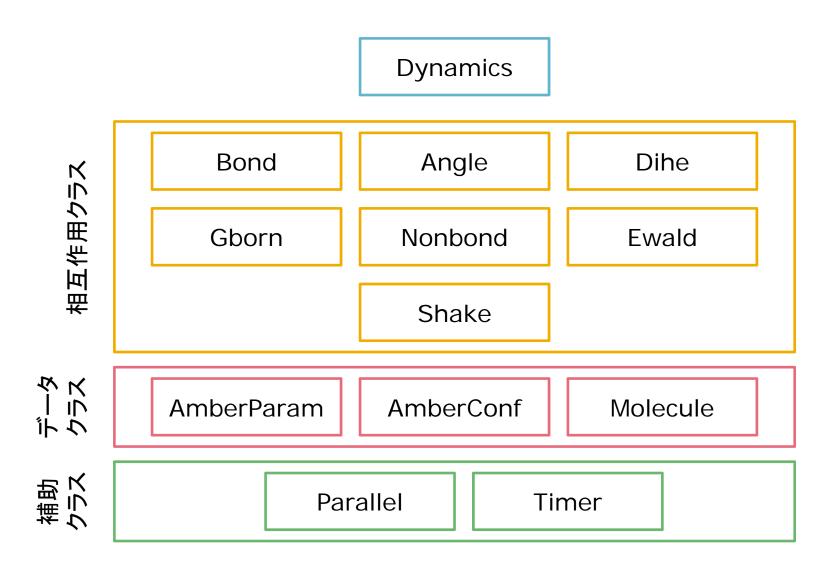
Bond

- •ペアリスト
- •力場パラメータ
- •初期化メソッド
- •力の計算メソッド

オブジェクト指向(2)

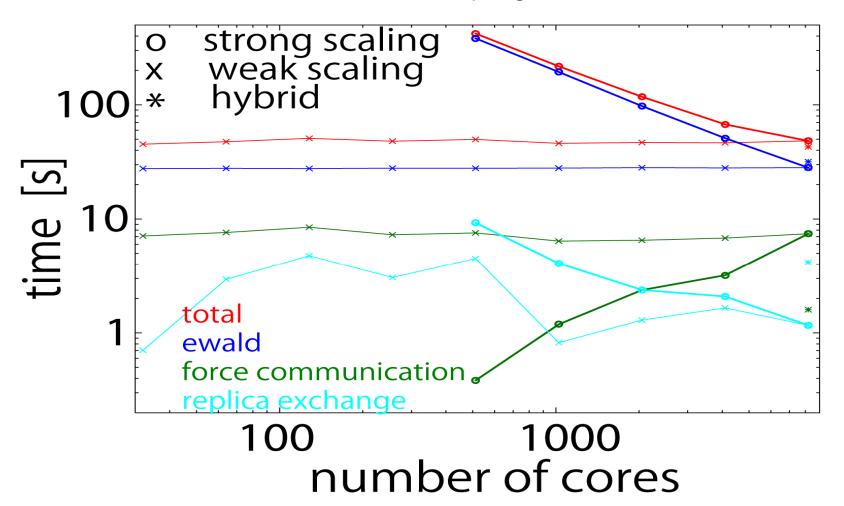
- オブジェクト指向は、
 - コードの再利用が容易→オリジナルのシミュレータを構築する上で有利
 - 複数のインスタンスを生成できる→マルチコピーシミュレーションを容易に実現
 - コードの修正の影響が及ぶ範囲が限定される →共同作業に適している
- 次世代スーパーコンピュータに標準搭載されるオブジェクト指向言語はC++
 - →C++を使ってコード

作成済みのクラス



ベンチマーク on RICC

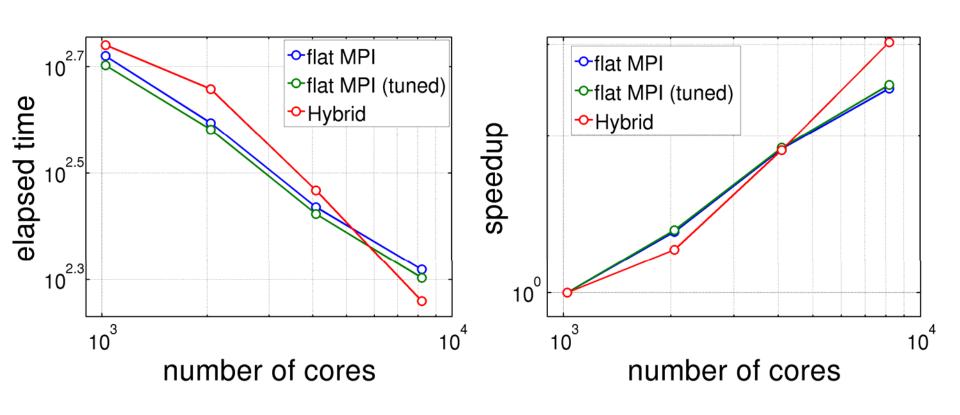
Multi-Scale Essential Sampling法のベンチマーク



系: sortase A + 溶媒, 27,810原子 weak scaling: 16コア/1レプリカ固定 strong scaling: 512レプリカ固定

ベンチマーク on RICC

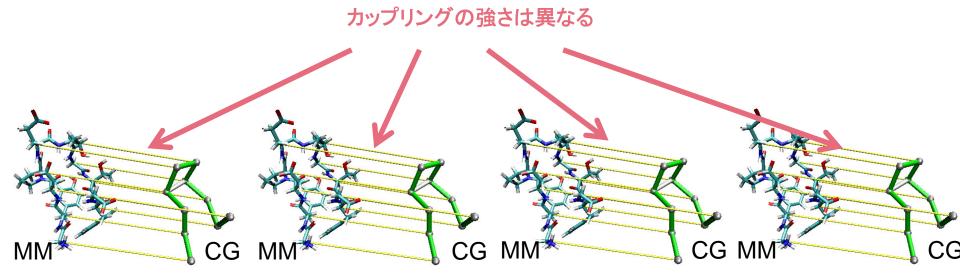
ストリング法のベンチマーク



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

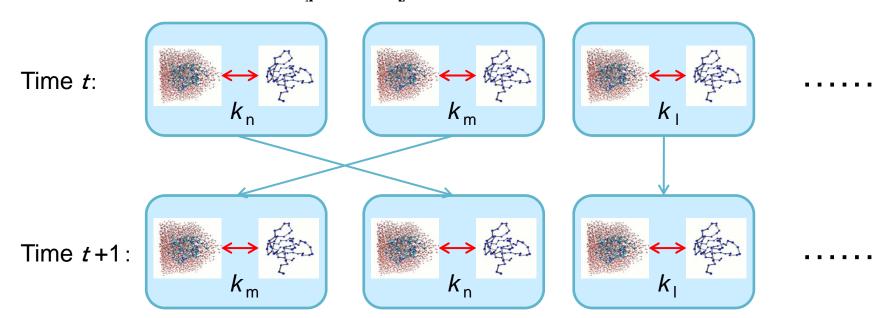
成果1: Multi-Scale Essential Sampling 法

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



Hamiltonian Exchange

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



keeping detailed balance condition for canonical distribution



transition probability between n and $m = \min(1, \exp(\Delta_{mn}))$ 9 $\Delta_{mn} = \beta (k_m - k_n) \{ W(\theta[\mathbf{x}_m], \mathbf{z}_m) - W(\theta[\mathbf{x}_n], \mathbf{z}_n) \}$

MSES: Summary

$$H(\mathbf{x}, \mathbf{z}) = K(\mathbf{p}_{\mathbf{x}}) + K(\mathbf{p}_{\mathbf{z}}) + V(\mathbf{x}) + U(\mathbf{z}) + kW(\mathbf{z}, \boldsymbol{\theta}[\mathbf{x}])$$
$$W(\mathbf{z}, \boldsymbol{\theta}[\mathbf{x}]) = \frac{1}{2} \left[\mathbf{f}(\mathbf{z}) - \mathbf{f}(\boldsymbol{\theta}[\mathbf{x}]) \right]^{2}$$

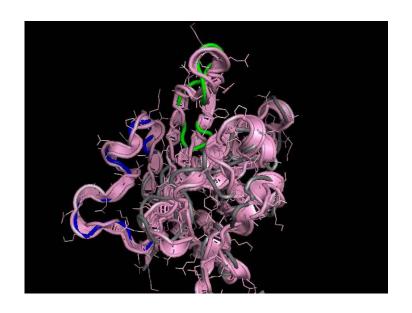
in analogy with Maragliano and Vanden-Eijnden, CPL 446, 182 (2007)

- 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

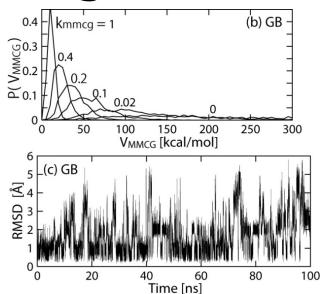
$$\Delta_{mn} = \beta (k_m - k_n) \{ W(\theta[\mathbf{x}_m], \mathbf{z}_m) - W(\theta[\mathbf{x}_n], \mathbf{z}_n) \}$$

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

Example: CG Pulling MM in Solution

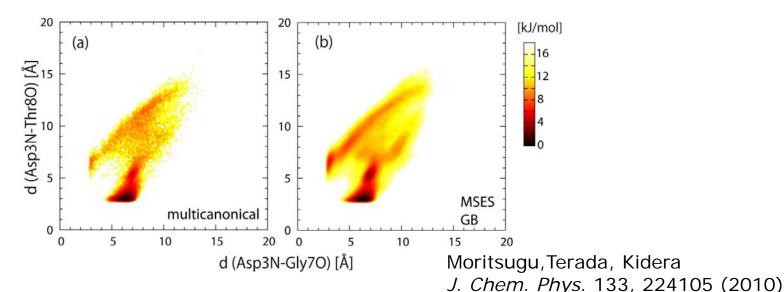


Chignolin in Implicit Solvent



- 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

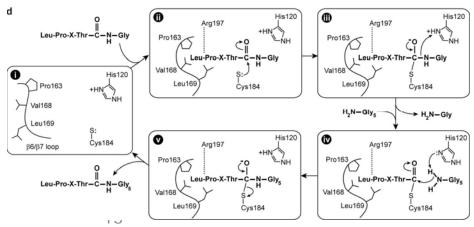
- a family of Gram-positive transpeptidases
 - functions for cell wall anchored protein (CWAP) functions
 - recognizes the sorting signal with a conserved LPXTG motif in the CWAP
 - LPXT / G is cleaved and amide-linked to peptideglycan
- intrinsically disordered protein (IDP)

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

Membrane

- database of protein disorder, DisProt (DP00208)

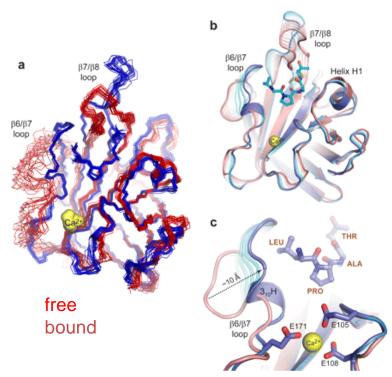
 Dunker et al., Nucleic Acids Res. 2007 (Database issue):D786-93 (2006)
- residue 162 to 174 (length is 13 in total 226)



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

Sortase

Structure of Sortase A

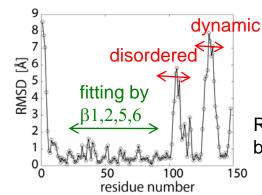


Suree at al., J. Biol. Chem. 284, 24465 (2009)

- NMR structures: sortase A from S. aureus
 - ligand-unbound: 1IJA, 25 models
 - ligand-bound: 2KID, 20 models

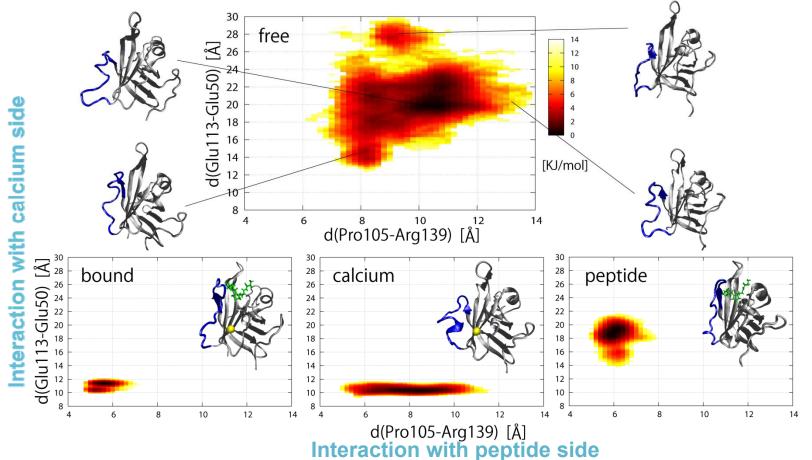
Ca²⁺ and peptide analog (LPAT)

- Eight-stranded β 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



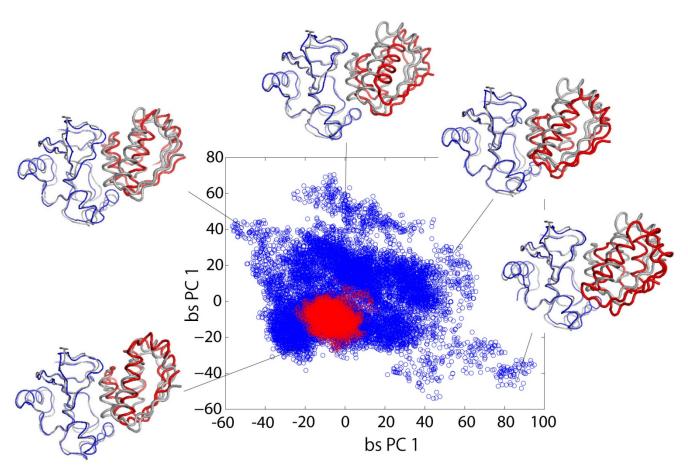
RMSD between the averaged bound and freestructures

Comparative Study: Calcium Ion Function



- Bound peptide and Ca²⁺ restrict the disordered loop flexibility.
- Bound Ca²⁺ decreases entropy change upon peptide binding, favoring binding affinity.

Barster Structures

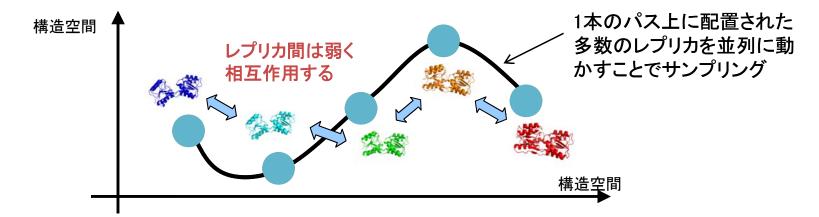


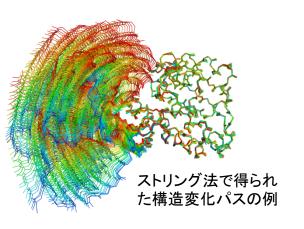
bs rmsdMM < 4 Å (gray structure as an example)

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

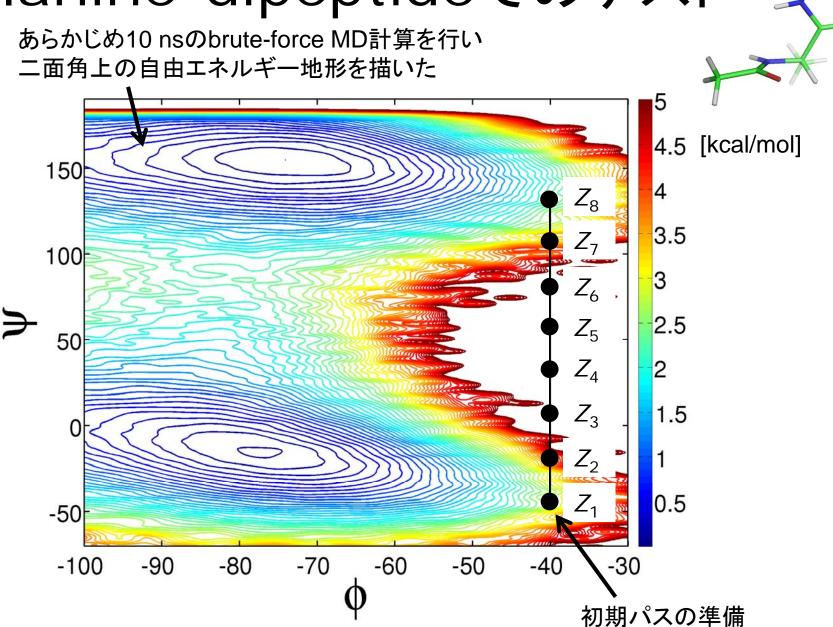
ストリング法によるパスサンプリング





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

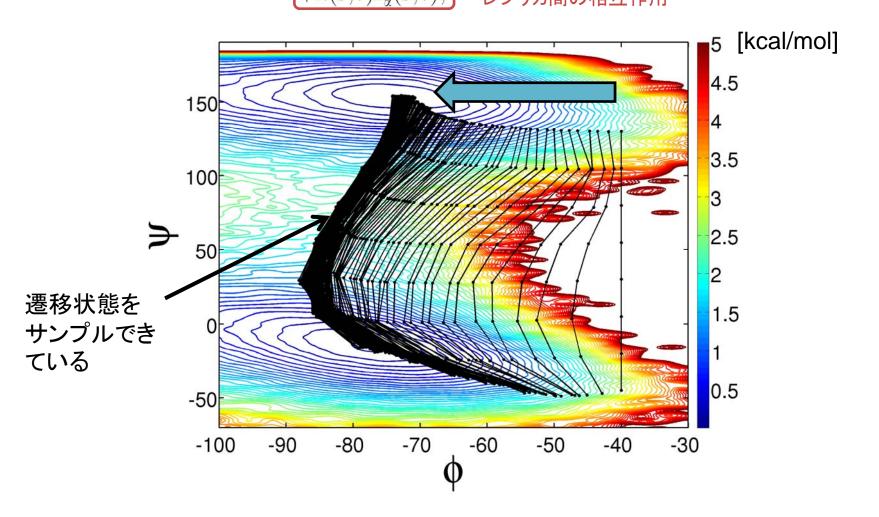
Alanine-dipeptideでのテスト



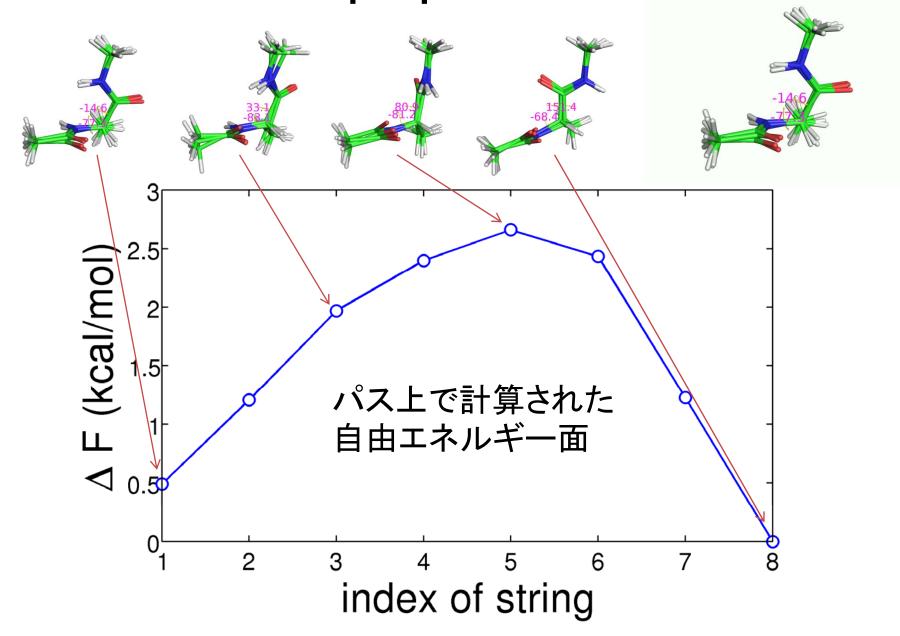
Alanine-dipeptideでのテスト

$$\gamma \dot{z}_{\alpha}(s,t) = \sum_{\beta=1}^{m} \widetilde{M}_{\alpha\beta}(x(s,t))\kappa(\theta_{\beta}(y(s,t)) - z_{\beta}(s,t))$$
 + $\lambda(s,t)z'_{\alpha}(s,t)$, レプリカ間の相互作用

自由エネルギー勾配に従っ て拘束長を動かす力



Alanine-dipeptideでのテスト

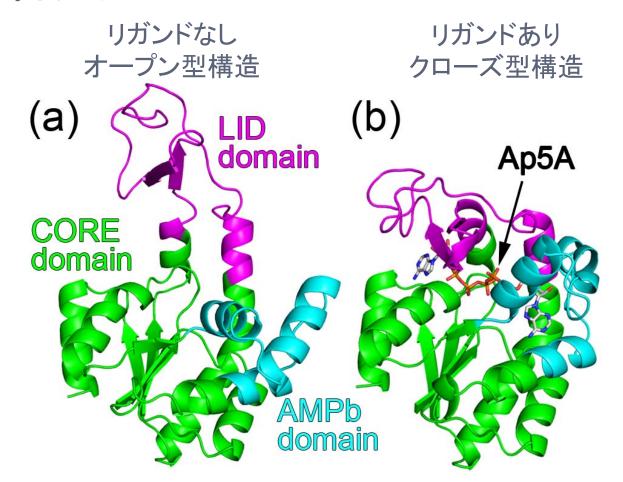


Adenylate kinaseへの応用

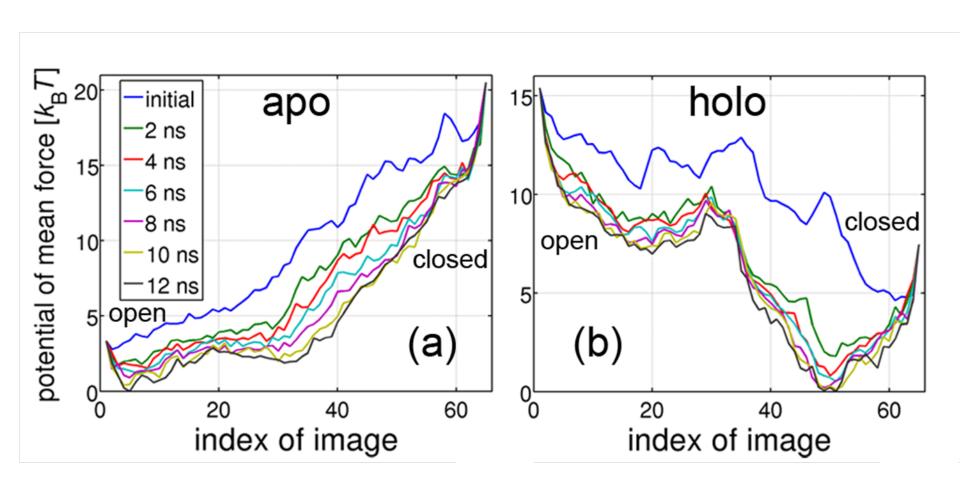
Adenylate kinase (214残基、3343原子)

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

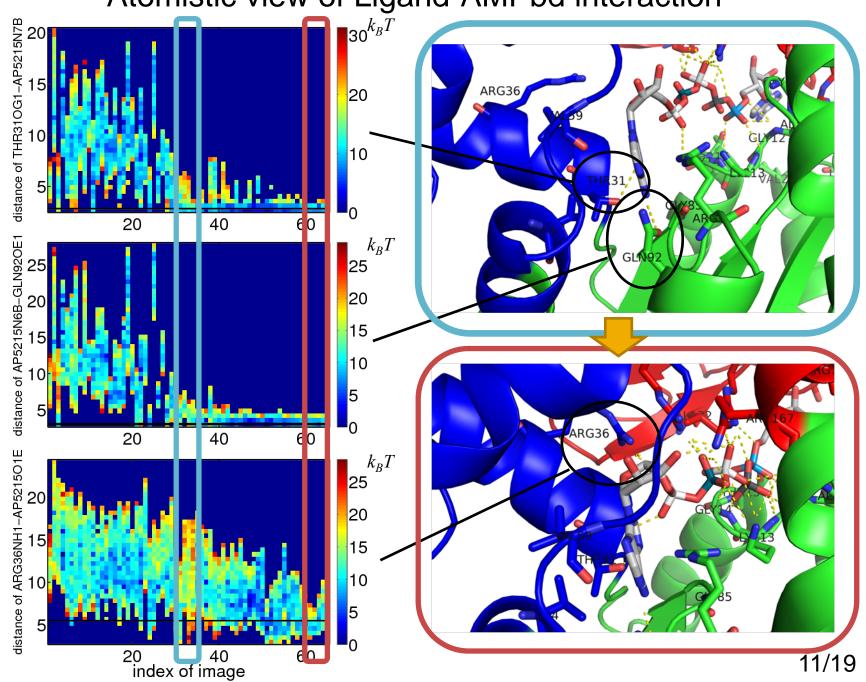
リガンド結合に伴い大規模な構造変化を生じることが結晶構造解析などにより示されている



エネルギーの低い反応経路の探索結果



Atomistic view of Ligand-AMPbd interaction

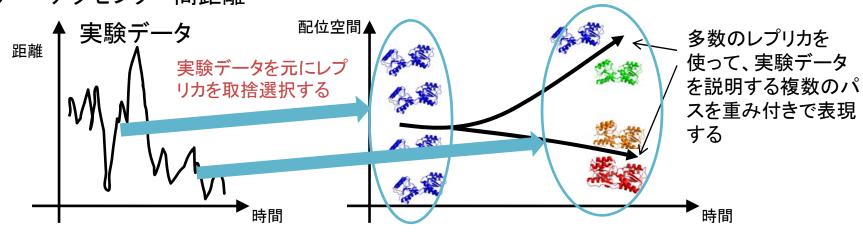


成果3:逐次モンテカルロ法

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

1分子FRET実験で観測される

ドナー・アクセプター間距離

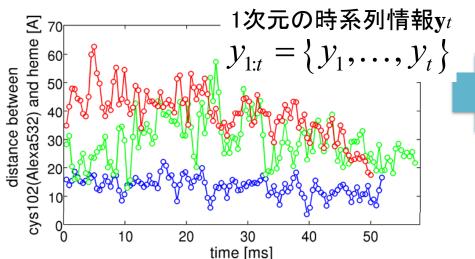


逐次モンテカルロ法

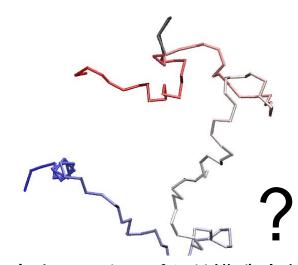
目的

- 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

1分子FRET時系列



M. Kinoshita, K. Kamagata, A. Maeda, Y. Goto, T. Komatsuzaki, and S. Takahashi, *Proc. Natl. Acad. Sci. USA* **104**, 10453 (2007)



高次元のタンパク質構造変数 x_t $x_{1:t} = \{x_1, \dots, x_t\}$

逐次モンテカルロ法

ベイズの定理

xについてマルコフ性を仮定することで逐次的な式に変形

$$p(x_{1:t} | y_{1:t}) = \frac{p(y_{1:t} | x_{1:t}) p(x_{1:t})}{p(y_{1:t})}.$$



$$p(x_{1:t} | y_{1:t}) = p(x_{1:t-1} | y_{1:t-1}) \frac{f(x_t | x_{t-1})g(y_t | x_t)}{p(y_t | y_{1:t-1})}.$$

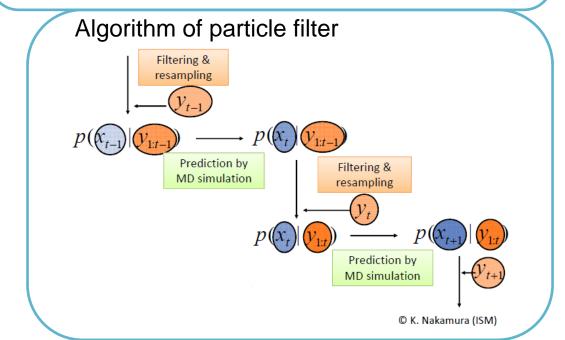
Simulation and observation model

$$x_{t} = f(x_{t-1}, \varepsilon) + \alpha_{t},$$

$$y_{t} = h(x_{t}) + w_{t}.$$

Langevin dynamics of coarse grained model observation model

 α_t, w_t : noise terms



 $f(\bullet | x_t)$: transition probability

 $g(\bullet | x_t)$: observation model

G. Kitagawa, *Journal of Computational* and *Graphical Statistics* **5**, 1 (1996).

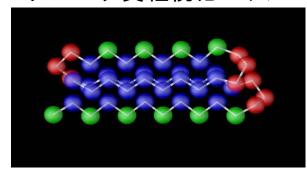
N.J. Gordon, D.J. Salmond, A.F.M.

Smith, *IEEE Proceedings F*, **140**, 107 (1993)

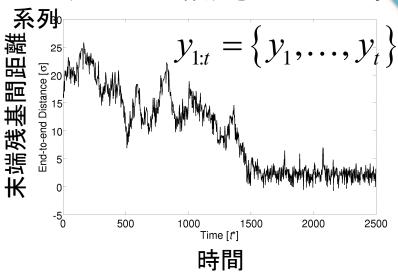
A. Doucet, and A. M. Johansen, "A Tutorial on Particle Filtering and Smoothing: Fifteen years Later", http://www.cs.ubc.ca/%7Earnaud/doucet_johansen_tutorialPF.pdf

粗視化モデルでのテスト

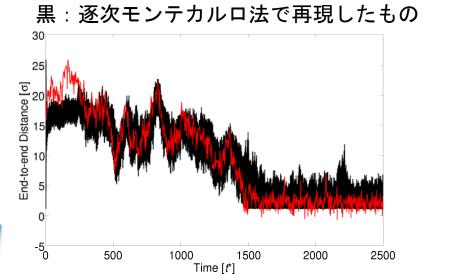
タンパク質粗視化モデル



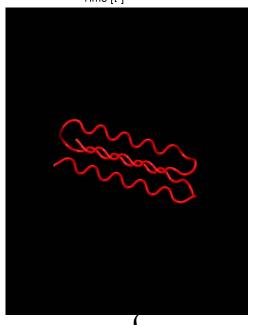
粗視化モデルのシミュレーション により人工的に作成されたFRET時



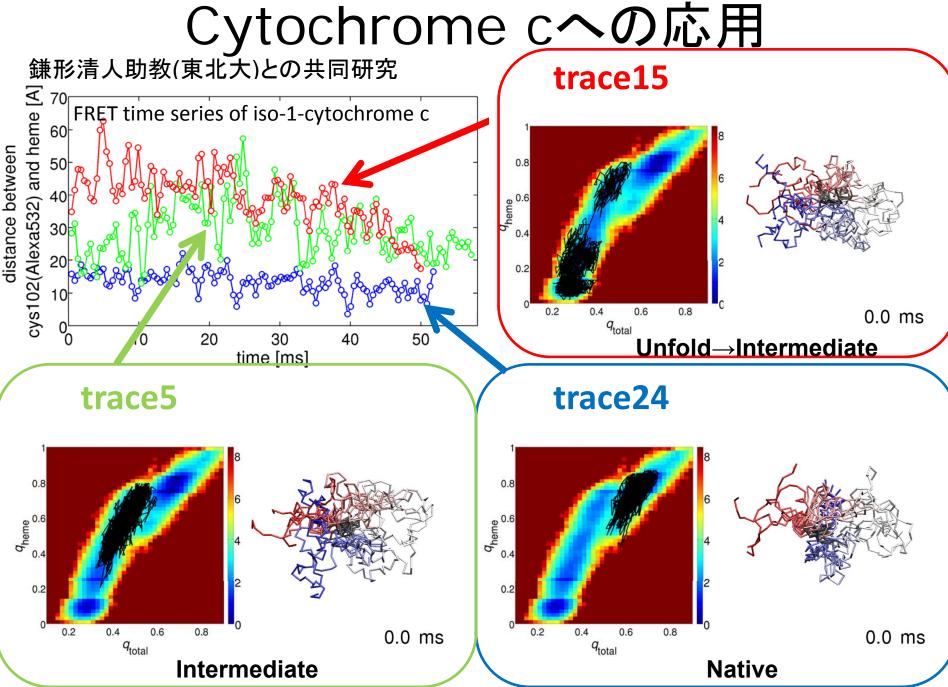
逐次モンテカルロ法 によるフィルタリング



レプリカの数=1600 赤:元の実験データ

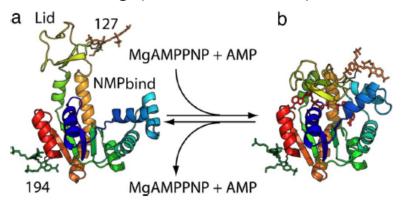


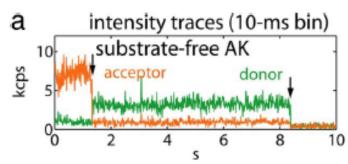
$$\overline{x_{1:t}} = \{x_1, \dots, x_t\}$$

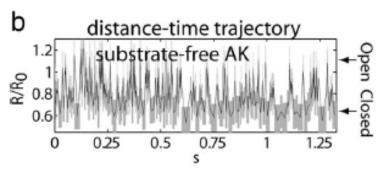


Adenylate kinaseへの応用

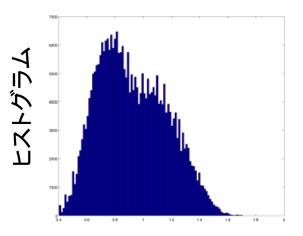
Prof. Haw Yang (Princeton Univ.)との共同研究







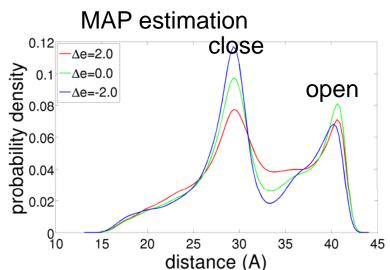
J. A. Hanson, K. Duderstadt, L. P. Watkins, S. Bhattacharyya, J. Brokaw, J-W. Chu, H. Yang, *Proc. Natl. Acad. Sci. USA* **104**, 18055 (2007)



プローブ間距離



逐次モンテカルロ法 によるフィルタリング



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。