

# 分子スケール研究開発チーム成果報告

# 分子スケールチーム成果総括報告と Platypus MM/CG の開発

## 木寺 詔紀

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 チームリーダー



### 発表者紹介

- 1982年3月 京都大学工学研究科高分子化学専攻博士課程退学
- 1982年8月 Cornell 大学 化学科 博士研究員
- 1986年2月 京都工芸繊維大学 工芸学部 助手
- 1988年4月 蛋白工学研究所 主任研究員
- 1996年4月 京都大学大学院 理学研究科 助教授
- 2001年4月 横浜市立大学大学院 総合理学研究科 教授
- 2006年10月 理化学研究所次世代計算機科学研究開発プログラム チームリーダー（兼任）
- 2011年4月 理化学研究所 HPCI 計算生命科学推進プログラム 副プログラムディレクター（兼任）

### 研究分野

計算生物学、生物物理学

# 分子スケールチーム成果総括報告と Platypus MM/CG の開発

木寺詔紀

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 チームリーダー

## 1. 目的

生命現象の最下層にある生体分子の機能発現過程のシミュレーションは、それ自体として時間、空間の階層性を持っており、それぞれの階層に対応したシミュレーション方法を必要とする。短時間の比較的少数の原子が関わる化学反応を含む高精度の相互作用のレベルでは量子化学計算 (QM) を用いる。より長時間のタンパク質の全原子の動きのシミュレーションは、環境として存在する水や脂質分子をあらわに取り入れた系で、経験的な力場 (MM) に基づいた古典分子動力学計算を用いて行われる。生理学的なレベルに近く時間スケールでの現象をより巨大な系においてシミュレーションする場合は、MM をさらに粗視化したモデル (CG) に基づいて、環境を陰的に考慮するランジャンバン方程式などを用いた方法によって行う。分子スケールでは、生体分子に関わるあらゆる現象を対象としたシミュレーションを可能とするために QM-MM-CG という 3 階層にわたる方法論を開発する。さらに、それらの階層を接続し、それぞれの方法の限界を超えるマルチスケール法として、QM と MM とを連成した QM/MM 法、さらに MM と CG を練成した MM/CG 法を開発する。これらの方法論によるシミュレーションをスーパーコンピュータ京の性能を十分に生かして行うことを可能とするソフトウェアを開発することが分子スケールチームの目的である。さらに、実証研究の対象として、多剤排出トランスポーター、脂質代謝多酵素複合体を分子スケールの共通ターゲットとして選び、スーパーコンピュータ京上で多階層でのシミュレーションを行うことで、これらのソフトウェアの有用性を示す。

## 2. 現時点での成果

### 2.1 ProteinDF (QM)

巨大複雑系であるタンパク質の状態を最も精密なレベルでシミュレートする目的で、タンパク質全体を含んだ全電子カノニカル分子軌道計算を行う。計算法には基底関数にガウス型関数を使用

する密度汎関数法 (DFT) を採用した。実績のある汎関数 B3LYP を実装し、高度並列化を行っている。現時点で 8,000 並列を達成している。適用例として、インシュリン 6 量体、ミオグロビン、チトクローム c などのカノニカル分子軌道を求め、カノニカル分子軌道の重要性を示し、分子軌道と分子機能との関わりを明らかにした。

### 2.2 Platypus-QM/MM (QM/MM)

QM と MM を連成することで、QM/MM レベルの第一原理分子動力学計算を高精度で高速に実行することができる。現在、QM 部分 (Platypus-QM) の開発に注力している。基底状態を計算する HF 法、DFT 法ばかりでなく、励起状態を計算する方法を組み入れることで化学反応にも対応することができる。QM 部分は 8,000 並列超を達成している。また、MM からの寄与を QM へ取り込む相互作用計算において、新たな高精度の手法として Zero-Dipole Summation 法を開発した。適用例としてプロリン異性化酵素 Pin1 によるプロリン異性化反応の水中での QM/MD 計算を行い反応経路における自由エネルギーを計算することで、タンパク質原子の反応への寄与を明らかにした。

### 2.3 Platypus-QM/MM-FE (QM/MM)

生体分子機能に関わる化学反応をハイブリッド QM/MM 法を用いて解析し機能発現の分子機構の解明を可能にする方法論を開発した。QM/MM-FE 法では、MM 分子力場を用いた MD 計算による MM 領域の構造サンプリングを行い、その MM 分布によって定義される自由エネルギー曲面上で QM 領域の反応経路探索を行うことで、MM 領域のゆらぎの中で起こる酵素反応を定量的に評価することができる。本手法の特徴は、QM 計算と MM 計算を同時に実行する困難を避け、ふたつの計算を完全に分離することができる。α-アミラーゼの酵素反応機構に適用し、タンパク質の構造揺らぎの反応自由エネルギーへの寄与を明らかにした。

### 2.3 Marble (MM)

全原子分子動力学法のソフトウェアである。空間分割、負荷分散、PME法などの高効率、高精度のアルゴリズムを用いている。現在、京で8000並列超を実現している。応用としてMDとNMR、X線溶液散乱実験との連携技術の開発を行った。また、共通ターゲットである多剤排出トランスポーターの分子動力学シミュレーションを行い、プロトン移動が、膜貫通ヘリックス部位でのアミノ酸側鎖のプロトン化によって媒介され、プロトン化状態の変化によって電荷バランスが変化することで、構造変化をもたらしていることを明らかにした。

### 2.4 Platypus-REIN (MM/CG)

レプリカ交換分子動力学法を行うためのソフトウェアである。レプリカ交換分子動力学法は異なった環境変数で行われる分子動力学計算で駆動される多数のレプリカが確率的に他の環境変数に移行することによって、通常分子動力学計算では到達できない広範囲の構造アンサンブルを精密に求めることができる。ここではさらに、多次元の環境変数を選択することでさらなるサンプリング空間の拡張をめざす多次元レプリカ交換法を採用している。現在、京で3万並列超を実現している。適用例として、細胞膜中でのアミロイド前駆体のシミュレーションを行い、その会合過程を予測した。またモデルペプチド(AAQAA)<sub>3</sub>の水中でのヘリックス形成シミュレーションを行っている。

### 2.5 Platypus-MM/CG (MM/CG)

MMとCGの二つのモデルを共役項で制約させながらひとつのシミュレーションとして行う。それによって、MMの精度とCGの効率を両立させることを目指している。主要な方法論は、マルチスケール拡張サンプリング(MSES)法、有限温度ストリング法である。MSES法は、上記のシミュレーションをそのまま分子動力学計算として実施し、最終的にCGからの影響のないMMの構造アンサンブルを得るために、共役項の強度を変数とするHamiltonian交換法を用いている。有限温度ストリング法は、同様なシミュレーションを経路探索の問題として行う。CGのポテンシャルをflatにして、共役項の強度を大きくすることによってCG自由度の上でのMMポテンシャルの平均力場を求めることができ、最小自由エネルギー経路を得ることができる。MSES法は天然変性タンパク質Sortaseのdisorder-order転移に、有限温度ストリング法はAdenylate kinaseの構造変化に適用した。

### 2.5 Platypus-MM/CG (CG)

粗視化ポテンシャルを用いて、タンパク質、核酸、脂質を含んだ系の大規模、長時間の分子動力学シミュレーションを行う。これまでに、蛋白質、DNAおよび脂質それぞれの単体系およびDNA-蛋白質、脂質-蛋白質の複合体の粗視化モデルの開発および実装を完了した。レプリカ交換法を実装している。現在、8000並列と京への実装を達成している。これまでに、タンパク質のフォールディング、キネシンの微小管上の歩行、多剤排出トランスポーターの薬剤排出などのシミュレーションに適用し、実験または既存のシミュレーション手法では困難であった、それぞれの運動の機構を明らかにしてきた。特に、多剤排出トランスポーターでは、非対称性を生み出すアロステリック効果および機能的回転の経路についての知見を与えた。

## 2. プロジェクト終了時の達成目標

それぞれの方法論によるシミュレーションをスーパーコンピュータ京の性能を十分に生かして行うことを可能とするソフトウェアを開発する。実証研究の目標を以下にあげる。

#### ProteinDF

光合成反応中心、光合成アンテナ蛋白質のカノニカル全電子計算

#### Platypus-QM

数万電子系の基底状態/励起状態の電子構造の計算

#### Platypus-QM/MM

数千電子系とそれを取り巻く数10万古典原子による反応自由エネルギー地形の計算

#### Platypus-QM/MM-FE

多剤排出トランスポーターのプロトン移動の自由エネルギー計算

#### MARBLE

多剤排出トランスポーターの薬剤排出過程のシミュレーション

#### Platypus-REIN

タンパク質の構造変化、構造予測および、タンパク質間の結合自由エネルギープロファイル(数万~十万粒子系)



#### Platypus-MM/CG

多剤排出トランスポーターの構造変化経路計算(有限温度ストリング法)

脂質β酸化多酵素複合体の構造変化計算(MSES法)

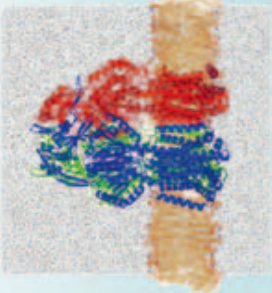
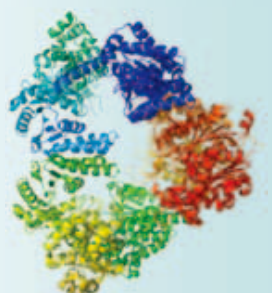
#### CafeMol


クロマチン(DNAヒストン複合体)のシミュレーション(1千万原子・ミリ秒の現象の再現)以上

# 分子スケールチーム成果総括報告と Platypus MM/CGの開発



理化学研究所  
次世代計算科学研究開発プログラム  
分子スケールチームリーダー  
木寺 詔紀




ISLiM成果報告会2011

1



## 背景・目的

- 背景
  - 生命現象の最下層にある生体分子の機能発現過程のシミュレーションは、それ自体として時間、空間の階層性を持っており、それぞれの階層に対応したシミュレーション方法が必要とする。短時間の比較的少数の原子が関わる化学反応を含む高精度の相互作用のレベルでは量子化学計算(QM)を用いる。より長時間のタンパク質の全原子の動きのシミュレーションは、環境として存在する水や脂質分子をあらわに取り入れた系で、経験的な力場(MM)に基づいた古典分子動力学計算を用いて行われる。生理学的なレベルに近づく時間スケールでの現象をより巨大な系においてシミュレーションする場合は、MMをさらに粗視化したモデル(CG)に基づいて、環境を陰的に考慮するランジャンバン方程式などを用いた方法によって行う。分子スケールでは、生体分子に関わるあらゆる現象を対象としたシミュレーションを可能とするためにQM/MM-CGという3階層にわたる方法論を開発する。さらに、それらの階層を接続し、それぞれの方法の限界を超えるマルチスケール法として、QMとMMとを連成したQM/MM法、さらにMMとCGを接続したMM/CG法を開発する。
- 目的
  - これらの方法論によるシミュレーションをスーパーコンピュータの性能を十分に生かして行うことを可能とするソフトウェアを開発する。

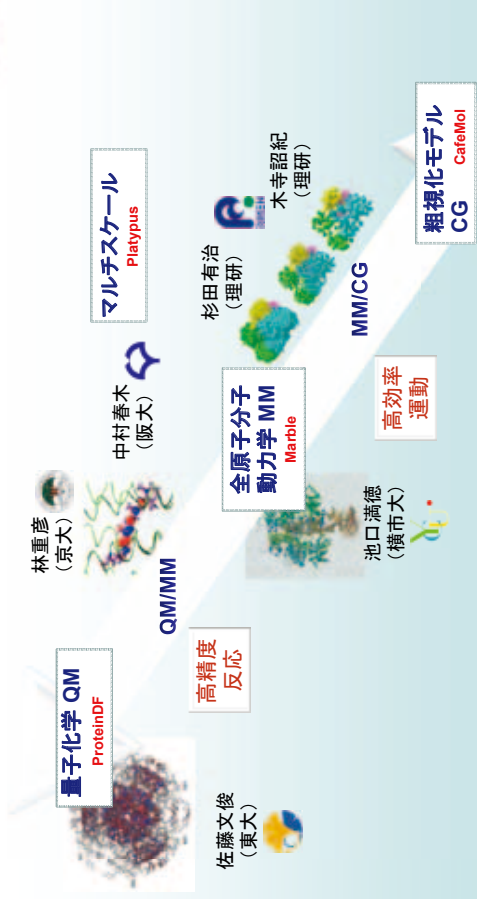


ISLiM成果報告会2011

2

## 概要・アプローチ



高効率運動

高精度反応

粗視化モデル CG CafeMol

高田彰二 (京大)

量子化学 QM  
ProteinDF

佐藤文俊 (東大)

全原子分子 MM  
Marble


池口満徳 (横浜市大)

マルチスケール  
Platypus

杉田有治 (理研)



MM/CG

木寺詔紀 (理研)



ISLiM成果報告会2011

3

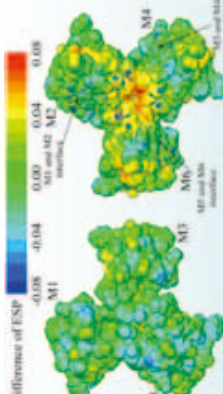



## QM

### ProteinDF

密度汎関数法に基づくタンパク質全電子波動関数計算


- 項番: M-5
- 開発責任者: 東京大学生産技術研究所教授 佐藤文俊
- 概要:
  - 複雑かつ大規模・高精度が要求されるタンパク質の全電子を密度汎関数法に基づきカニカル軌道で計算する。
  - 励起状態も含めた第一原理分子動力学計算が可能。
  - 離散化(計算モデル化)の方法
- 分子動力学
- 計算方法:
  - 直接法による密行列の対角化
  - 並列化の方法
  - 原子分割法、RT アルゴリズム、+n-シェルタイプ分類均等分割法。
  - 開発言語とライブラリ: C++, MPI, OpenMP, ScalAPACK
- コードの公開状況:
  - バイナリコードを無償で利用可能。
  - <http://www.ciss.is.u-tokyo.ac.jp/dl/>
- 現状の計算規模:
  - 306 残基、27,000 軌道、基底状態計算(世界最大)
  - Ahx3700 (64 CPU) で実行、300 GFlops
  - メモリ容量 256 GB、ディスク容量 1 TB
- 次世代機「京」での計算規模:
  - ほぼ全てのタンパク質に対応するため計算規模を3倍に
  - 基底状態だけでなく、励起状態のダイナミクスを解析
  - メモリ容量 2PB、ディスク容量 10PB (励起状態)



インスリン6量体静電ポテンシャルの古典計算との差

■ できること


- 創薬の信頼性の高い基礎研究のみならず、医薬品開発の高度化および高効率化、ならびに次世代の医薬品研究開発モデルの創成といった応用に貢献できる。
- 方法特許やビジネスモデル特許などを含め欧米に対する優位性を実現できる。
- 脂質、分子素子、環境物質などへの応用も大いに期待できる。



ISLiM成果報告会2011

4






## Platypus-QM

### 量子化学計算



- ・ 項番: M-7
- ・ 開発責任者: 大阪大学蛋白質研究所教授 中村春木
- ・ 概要:
  - 分子軌道法および密度汎関数理論 (DFT) を用いた蛋白質等分子系の全電子第一原理計算を行う。
  - 離散化 (計算モジュール) の方法
  - 分子軌道法、密度汎関数理論、Hartree-Fock との Hybrid DFT
- ・ 計算方法:
  - 多中心Euler-Maclaurin-Lebedev の求積分法、固有値問題等
- ・ 並列化の方法:
  - MPI 並列化 (ハイブリッド並列化→拡張中)
- ・ 開発言語とライブラリ: Fortran77, Fortran90, MPI, HDF, ScalLAPACK, LAPACK 等
- ・ コードの公開状況: 未公開
- ・ 現状での計算規模:
  - バクテリアオロロロロ等の色素
  - 数自~千電子系をPCクラス98192コアで解析
  - メモリ容量 1.2GB/コア、ディスク容量 100 MB
- ・ 次世代機「京」での計算規模:
  - 数100~数万電子系の基底状態/励起状態の電子構造を64万コアで解析
  - メモリ容量 1.6GB/コア、ディスク容量数GB



β-hemolysin subunit of Staphylococcus aureus の基底状態/励起状態の電子構造 (バクテリアオロロロロ) の色素

- ・ どのようなことができるか
  - 遷移金属を含む蛋白質系の基底状態および励起状態の電子構造の解明
  - 電子構造解析による反応機構の解明
  - 磁束的相互作用など、蛋白質構造が不明な場合のブロープ (EPR等) と付けられる、パラメータの計算

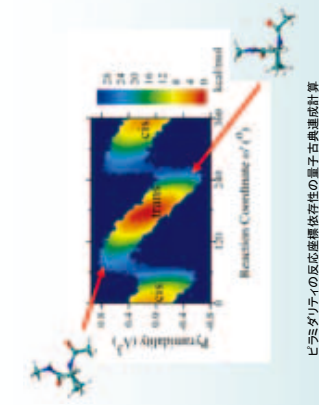
ISLiM 成果報告会 2011

## Platypus-QM/MM



### 量子化学計算/分子動力学連成計算

- ・ 項番: M-8
- ・ 開発責任者: 大阪大学蛋白質研究所教授 中村春木
- ・ 概要:
  - 電子状態の効果を分子動力学に取り込むことで酵素反応機構などの生体高分子の精密解析を行う。
  - 離散化 (計算モジュール) の方法
  - 分子軌道法、分子動力学法
- ・ 計算方法:
  - 直接法による密行列対角化、反復法による疎行列連立方程式求解、改良 Wozniak
- ・ 並列化の方法:
  - 領域分割 (分子軌道法)、原子分割 (MD)
- ・ 開発言語とライブラリ: Fortran77, Fortran90, C, MPI, OpenMP, ScalLAPACK, LAPACK
- ・ コードの公開状況: 未公開
- ・ 現状での計算規模:
  - プロトン及びUnimelicapsinids (毒素系) の水溶液中 (水分子は古典軌道) における異性化熱力学過程の解明
  - 20000原子を蛋白研PCクラス256コアで解析
  - メモリ容量 300MByte/コア、ディスク容量 3GB
- ・ 次世代機「京」での計算規模:
  - 数100~数万電子系と、それを取り巻く数10万古典軌道による反応自由エネルギー地形を64万コアで解析
  - 基底状態だけでなく、励起状態のダイナミクスを解析
  - メモリ容量 1.6GB/コア、ディスク容量数GB



ピラニダリチの反応座標依存性の量子古典連成計算

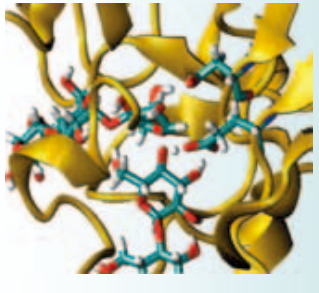
- ・ どのようなことができるか
  - 生命活動に重要な酵素反応機構や、その反応を制御する薬物などの開発に有用な知見を得ることができる。

## Platypus-QM/MM-FE

### ハイブリッド QM/MM 反応自由エネルギー計算

- ・ 項番: M-6
- ・ 開発責任者: 京都大学大学院理学研究科准教授 林重彦
- ・ 概要:
  - 量子化学的手法 (QM) と分子力場法 (MM) のハイブリッドである QM/MM 法により、生体分子内の反応基質分子の自由エネルギー最適構造の決定及び反応経路解析を行う。
  - 離散化 (計算モジュール) の方法
  - 分子軌道法、分子力場法
- ・ 計算方法:
  - Reweighting法 (自由エネルギー)、Ewald法 (クーロン)。
- ・ 並列化の方法:
  - MM 構造サンプリング
- ・ 開発言語とライブラリ: Fortran77, GAMESS の socks ライブラリ
- ・ コードの公開状況: ソースコードを無償で利用可能。要相談。
- ・ 現状での計算規模:
  - a-amyase 酵素における polysaccharide (二量体部分) の自由エネルギー最適構造。
  - 7万原子系、QM 原子数 69、基底関数 650、MM 構造サンプリング 140,000
- ・ 次世代機「京」での計算規模:
  - 多刺排出トランスポーター
  - 約50万原子、MM 構造サンプリング 100 万
  - メモリ容量 4 TB、ディスク容量 4 TB



a-amyase 酵素中のpolysaccharide 基質の加水分解反応遷移状態

- ・ どのようなことができるか
  - 酵素反応における基質の反応性解析が行える。
  - 生体エネルギー一変換で最も重要であるプロトン濃度の配を用いた分子動力学が明らかになる。




## MARBLE

### 全原子分子動力学計算

- ・ 項番: M-3
- ・ 開発責任者: 横浜市立大学大学院総合理学研究科准教授 池口尚徳
- ・ 概要:
  - タンパク質や核酸など生体分子の立体構造で、水素まで含んだ全原子モデルに対する分子動力学計算を行う。
  - 離散化 (計算モジュール) の方法
- ・ 計算方法:
  - EFTを用いたParticle Mesh Ewald法
  - シンプレクティック数値積分法
- ・ 並列化の方法:
  - 空間分割、スレッド並列はルーブ分割
- ・ 開発言語とライブラリ: C, MPI, OpenMP
- ・ コードの公開状況: 「京」において無償で利用可能
- ・ 現状での計算規模:
  - 10万~100万原子系
  - Cray XE6で3000コア並列、京で8000コア並列 (近距離力)
  - メモリ容量 4 GB/node、ディスク容量 1 TB
- ・ 次世代機「京」での計算規模:
  - 多刺排出トランスポーター (約50万原子)
  - 反復回数10'以上
  - メモリ容量 16 GB/node、ディスク容量 10 TB



多刺排出トランスポーター

- ・ どのようなことができるか
  - 膜タンパク質が大規模な構造変化を行う様子をシミュレートすることができる。
  - 薬剤耐性菌の原因因子である多刺排出トランスポーターのシミュレーションを行うことにより、薬が薬物を排出する際の分子機構を明らかにする。



## MM/CG

# Platypus-REIN

レプリカ交換分子動力学計算インターフェース

- ・ 項番: M-2
- ・ 開発責任者: 理化学研究所基礎研究主任 杉田有治
- ・ 概要: タンパク質のまわりの環境を含めたタンパク質の構造予測や、構造変化予測、自由エネルギー計算等を効率良く行うためのプログラム。多次元レプリカ交換分子動力学計算を行う。1原子に対応したMarcherを分子動力学計算に用いるが、その他のプログラムにも対応可。
- ・ 離散化(計算モデル化)の方法: 古典粒子(MD)
- ・ 計算方法: レプリカ交換法
- ・ 並列化の方法: 複数のレプリカの並列実行
- ・ 開発言語とライブラリ: Fortran90, C, MPI
- ・ コードの公開状況: ハイブリッドを無償で利用可能
- ・ 現状での計算規模: (AAGAA)3タンパク質(1万原子系)をRCCの8102コアで解析
- ・ 次世代機「京」での計算規模: タンパク質の構造変化、構造予測および、タンパク質間の結合自由エネルギー計算
- ・ 1-100万原子系を64コアで解析


ISLIM成果報告会2011

9



レプリカ交換法によるタンパク質の分解

- できること
  - 膜タンパク質や大規模な系の構造変化予測や、これまで予測できなかった系の構造予測ができる。
  - 原子解像度でタンパク質の複合体形成機構の理解が深まる。
  - 例えば応用として、タンパク質と疾患との関係等の理解が深まり、将来の医学や創薬に役立つ事が期待できる。



## MM/CG

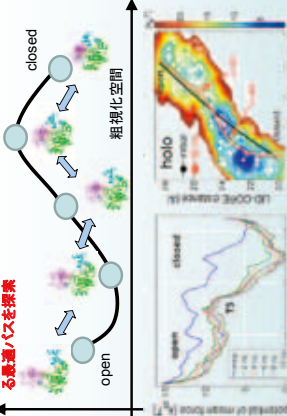
# Platypus-MM/CG

マルチコピー・マルチスケール分子シミュレーション法開発の基盤となるクラスライブラリ

- ・ 項番: M-1
- ・ 開発責任者: 理化学研究所分子スケール研究開発チーム
- ・ チームリーダー: 木寺昭紀
- ・ 概要: タンパク質や核酸、薬剤など生体分子を計算対象とする。反応の時間スケールが短い(ミリ秒~秒)ために、通常のMDでは追跡が困難な、生体分子同士の相互作用やこれに伴う立体構造変化過程の再現を目的とする。
- ・ 離散化(計算モデル化)の方法: マルチコピー・マルチスケール連成分子動力学法
- ・ 計算方法: レプリカ交換法, MSES法, ストリング法, 結合フィルタ法
- ・ 並列化の方法: コピー間並列(MPI並列), ループ分割(MPI・スレッド並列)
- ・ 開発言語とライブラリ: C++, MPI, OpenMP, LAPACK, FFTW, NetCDF
- ・ コードの公開状況: ハイブリッドを無償で利用可能
- ・ 現状での計算規模:
  - アニマルタンパク質系(6万原子系 × 256コピー)
  - 京 30720コア, メモリ容量 52 GB, ディスク容量 10 TB
  - 次世代機「京」での計算規模:
    - 多刺排出トランスポーター(AcB)
    - 1機2,800万原子系(50万原子系 × 256コピー)
    - 65,536コア(256コア77コピー)使用
    - メモリ容量 416 GB, ディスク容量 1 PB

ISLIM成果報告会2011

12



全原子モデル計算を用いて組織化空間における最適化を探索

closed

組織化空間

open

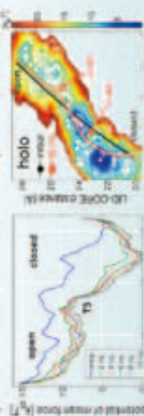



図. ストリング法はの概念図(上)とタンパク質への応用例(下)

- できること
  - 薬剤などの生理活性物質がタンパク質など生体高分子に結合する過程を再現できるようになる
  - 薬剤の結合やタンパク質同士の結合に伴う立体構造変化を予測できるようになる
  - 結合過程や立体構造変化過程の解析から、活性の高い物質の創出やタンパク質の機能改善につながる手がかりが得られる。



## CG

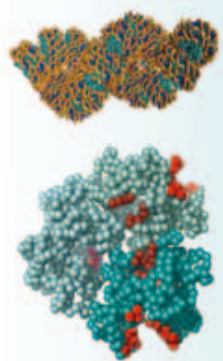
# CafeMol

組視化分子モデル計算

- ・ 項番: M-4
- ・ 開発責任者: 京大大学院理学研究科准教授 高田彰二
- ・ 概要: 組視化分子モデル計算により大規模生体分子の長時間シミュレーションを行う。
- ・ 離散化(計算モデル化)の方法: 粗視化された分子モデルによる古典分子動力学法
- ・ 計算方法: Langevin方程式の時間発展を数値的に積分する。
- ・ 並列化の方法: Neighbor list方式, レプリカ交換法
- ・ 開発言語とライブラリ: Fortran90, MPI, OpenMP
- ・ コードの公開状況: <https://www.cafemol.org> からダウンロード可。ただし生体膜、核酸については要調整。
- ・ 現状での計算規模:
  - 1万残基タンパク質のミリ秒相当のシミュレーションが可能
  - 京で8000コア並列
  - メモリ容量 2 GB/コア, ディスク容量 1 TB
- ・ 次世代機「京」での計算規模:
  - 10万粒子(100万原子)系の秒相当のシミュレーション
  - 反復回数10<sup>10</sup>回以上
  - メモリ容量 2 GB/コア, ディスク容量 1 PB


ISLIM成果報告会2011

12



トランスポーター(左)とDNAにストン複合体(右)

- できること
  - X線回折やNMRによる構造情報が存在するタンパク質、核酸などの動態の分子動力学シミュレーション。とくに構造変化に伴うタンパク質ドッキング、分子モーターやトランスポーターの大規模構造変化などをシミュレーションできる。
  - 例として、キネシンの動態、多刺排出トランスポーター、DNAにストン複合体(ヌクレオソーム)など。



## プロジェクト終了時の達成目標

それぞれの方法論によるシミュレーションをスーパーコンピュータの性能を十分に生かして行うことを可能とするソフトウェアを開発する。実証研究の目標を以下にあげる。

- ・ ProteinDF 光合成反応中心、光合成アンテナ蛋白質のカノニカル全電子計算
- ・ Platypus-QM 数万電子系の基底状態/励起状態の電子構造の計算
- ・ Platypus-QM/MM 数千電子系とそれを取り巻く数10万古典原子による反応自由エネルギー地形の計算
- ・ Platypus-QM/MM/FE 多刺排出トランスポーターのプロトン移動の自由エネルギー計算
- ・ MARBLE 多刺排出トランスポーターの薬剤排出過程のシミュレーション
- ・ Platypus-REIN タンパク質の構造変化、構造予測および、タンパク質間の結合自由エネルギープロファイル(数万~十萬粒子系)
- ・ Platypus-MM/CG 多刺排出トランスポーターの構造変化経路計算(有限温度ストリング法) 脂質β酸化多酵素複合体の構造変化計算(MSES法)
- ・ CafeMol クロマチン(DNAヒストン複合体)のシミュレーション(1千万原子・ミリ秒の現象の再現)

ISLIM成果報告会2011

12

# タンパク質カノニカル分子軌道法プログラム ProteinDF の京コンピュータへの実装

佐藤文俊

東京大学生産技術研究所  
革新的シミュレーション研究センター 教授




## 発表者紹介



1990年 3月 慶應義塾大学大学院理工学研究科物理学専攻博士課程修了  
1990年 4月 科学技術庁特別研究員  
1991年 3月 九州工業大学情報工学部助手  
2002年 11月 東京大学生産技術研究所客員助教授  
2004年 12月 東京大学生産技術研究所助教授  
2008年 4月 東京大学生産技術研究所教授

## 研究分野

計算生体分子科学、量子分子生物学

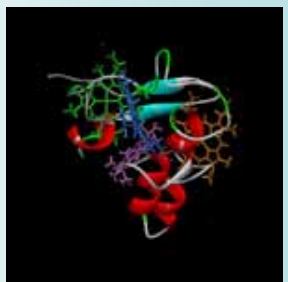






# タンパク質カノニカル分子軌道法プログラム ProteinDFの京コンピュータへの実装


東京大学生産技術研究所  
革新的シミュレーション研究センター  
教授 佐藤文俊








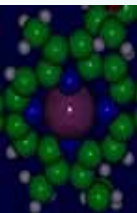
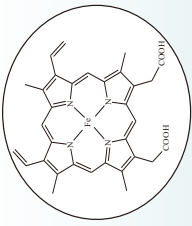
ISLIM成果報告会2011


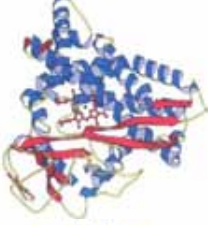
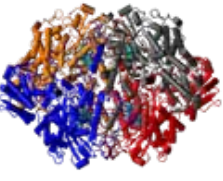

1



## ヘムタンパク質







**Hemoglobin**  
Oxygen molecule  
transportation

**P450, CYP**  
Oxidation/epoxidation

**Catalase**  
Hydrogen peroxide  
detoxification


**Cytochrome**  
Electron transfer





平成23年12月4日

ISLIM成果報告会2011

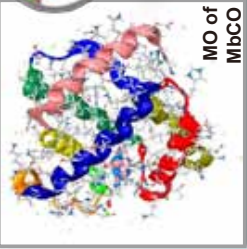
2



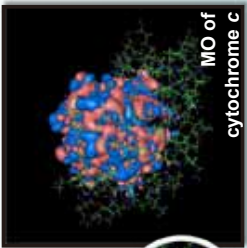



## 背景・目的

- 背景  
ProteinDFによるタンパク質の分子軌道計算




**MO of MbCO**  
94.7%: Feのd軌道  
99.9%: Heme+近位His  
Fe(III)で失活




**MO of cytochrome c**  
85.2%: Feのd軌道  
Fe(II)⇌Fe(III)で電子移動



- 目的  
ProteinDFを京コンピュータに実装・汎用的な巨大カノニカル分子軌道計算ライブラリへ



ISLIM成果報告会2011

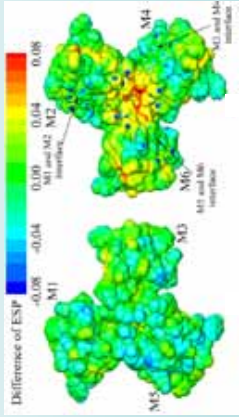
3




## 概要・アプローチ

- ProteinDFとは
  - タンパク質のカノニカル分子軌道を計算する超大規模量子化学計算プログラム
  - C++でコーディング
- 主な機能と特徴
  - HF, DFT(含ハイブリッド汎関数)
  - ガウス型基底関数
  - 分子種別にRI法やRT法アルゴリズムなどが選択可能
  - 大規模SCF計算を成功させるための高精度初期値作成法
  - エネルギー勾配計算





Difference of ESP  
-0.08 0.04 0.08  
M1 M2 M3 M4  
M1 and M2 interface M3 and M4 interface

- インスリン6量体の静電ポテンシャル古典法との比較




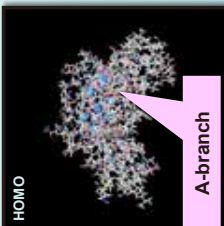
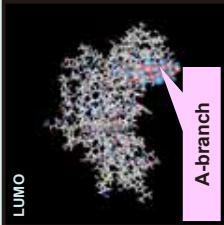
ISLIM成果報告会2011

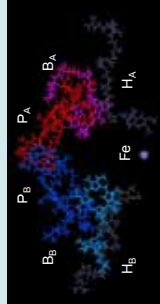
4






## 現在までの研究開発成果



- 計算事例
  - シトクロム<sub>c</sub><sub>3</sub>
  - 光合成反応中心(部分系)
 








ISLiM 成果報告会 2011 5






## 機能・性能がどこまで達成しているか

- 機能
  - 基本的な計算機能は「イノベーション基盤シミュレーションソフトウェアの研究開発」プロジェクトで開発(済み)
  - バージョン管理
    - ブランチを作らずに、様々なハードウェアに対応
    - 本プロジェクトでは、「京」対応のために新しい動作形式を研究開発
- 性能
  - 分散並列計算機における数千ハイブリッド並列計算
  - 「京」での正常動作
    - MPI / OpenMPでのハイブリッド並列で動作可能
    - 「京」での演算性能向上・チューニング作業中
    - 分子積分・数値積分の単体性能の向上






ISLiM 成果報告会 2011 6






## その際に克服すべき問題は何だったか

- 「京」で正常に動作させること
  - コンパイラの仕様に沿ったC++コードの記述を行う
  - 分散ローカルファイルシステムの対応
- 効率の良い大規模並列分子軌道計算アルゴリズムの開発
  - 理論上  $O(N^4)$  の分子積分アルゴリズムの計算量の削減
  - ロードバランスの均等化
    - 1つ1つの分子積分の演算量が、基底関数の組み合わせにより異なる
    - 総演算量を前もって見積もることが困難
  - 通信コストの削減
    - 分子軌道計算において、全通信は避けられない
- 巨大行列の取り扱い
  - 1,000残基規模タンパク質の軌道数  $\approx 100,000$
  - 行列1つに必要なメモリ量  $\approx 10^5 \times 10^5 \times 8 \text{ byte} = 80 \text{ GB}$
  - 行列演算に最低限必要なメモリ量  $\approx 80 \text{ GB} \times 3 = 240 \text{ GB}$
  - 最低限必要なメモリ量が1ノードに収まりきらない

ISLiM 成果報告会 2011 7



## どのように解決したか(最適化など)

- 「京」で正常に動作させる対策
  - OpenMPコードの書き換え
  - 新しいコンパイラでOpenMP3.0仕様に対応
  - 分散ローカルファイルシステム用I/Oルーチンの構築
- 大規模分子・大規模並列用分子積分アルゴリズムの開発
  - 高精度・高効率カットオフ法採用による計算コストの低減
  - セルの組み合わせの並び替えによるタスク均等化
  - 動的ロードバランスの採用
- 巨大行列の取り扱い
  - 大域分散行列の採用
  - ScaLAPACKによる高速密行列演算
  - 大域分散行列用分子積分・数値積分アルゴリズムの実装




7x8 行列を2x2小行列に分割

R <sub>0,0</sub>	R <sub>0,1</sub>	R <sub>0,2</sub>	R <sub>0,3</sub>	R <sub>0,4</sub>	R <sub>0,5</sub>	R <sub>0,6</sub>	R <sub>0,7</sub>	R <sub>0,8</sub>
R <sub>1,0</sub>	R <sub>1,1</sub>	R <sub>1,2</sub>	R <sub>1,3</sub>	R <sub>1,4</sub>	R <sub>1,5</sub>	R <sub>1,6</sub>	R <sub>1,7</sub>	R <sub>1,8</sub>
R <sub>2,0</sub>	R <sub>2,1</sub>	R <sub>2,2</sub>	R <sub>2,3</sub>	R <sub>2,4</sub>	R <sub>2,5</sub>	R <sub>2,6</sub>	R <sub>2,7</sub>	R <sub>2,8</sub>
R <sub>3,0</sub>	R <sub>3,1</sub>	R <sub>3,2</sub>	R <sub>3,3</sub>	R <sub>3,4</sub>	R <sub>3,5</sub>	R <sub>3,6</sub>	R <sub>3,7</sub>	R <sub>3,8</sub>
R <sub>4,0</sub>	R <sub>4,1</sub>	R <sub>4,2</sub>	R <sub>4,3</sub>	R <sub>4,4</sub>	R <sub>4,5</sub>	R <sub>4,6</sub>	R <sub>4,7</sub>	R <sub>4,8</sub>
R <sub>5,0</sub>	R <sub>5,1</sub>	R <sub>5,2</sub>	R <sub>5,3</sub>	R <sub>5,4</sub>	R <sub>5,5</sub>	R <sub>5,6</sub>	R <sub>5,7</sub>	R <sub>5,8</sub>
R <sub>6,0</sub>	R <sub>6,1</sub>	R <sub>6,2</sub>	R <sub>6,3</sub>	R <sub>6,4</sub>	R <sub>6,5</sub>	R <sub>6,6</sub>	R <sub>6,7</sub>	R <sub>6,8</sub>
R <sub>7,0</sub>	R <sub>7,1</sub>	R <sub>7,2</sub>	R <sub>7,3</sub>	R <sub>7,4</sub>	R <sub>7,5</sub>	R <sub>7,6</sub>	R <sub>7,7</sub>	R <sub>7,8</sub>

node 1      node 2      node 3      node 4

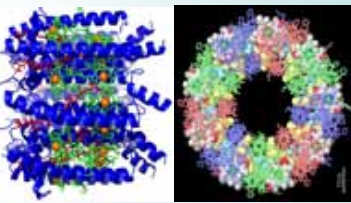



ISLiM 成果報告会 2011 8






## プロジェクト終了時の達成目標

- アンテナタンパク質(集光性タンパク質複合体)の全電子計算
  - A, B銀18量体構造の20000原子巨大タンパク質
  - 長寿命量子コヒーレンスの正体に迫る
  - (色素増感)太陽電池の効率向上設計などへの波及効果も期待
- 必要とする計算機資源
  - 京の1/10の資源で1ヶ月程度



Light-harvesting protein



Gregory A. Engel<sup>1,2</sup>, Thomas B. Cahoon<sup>1,2</sup>, Elizabeth L. Read<sup>1,2</sup>, Tao-Kyo Ahn<sup>1,2</sup>, Tomoko Maehata<sup>1,4</sup>, Yuan-Chung Cheng<sup>1,2</sup>, Robert E. Blankenship<sup>1,2</sup>, Graham R. Fleming<sup>1,2</sup>

**Evidence for wavelike energy transfer through quantum coherence in photosynthetic systems**

*Nature*, 446, 782-786 (2007).

G. Engel, T. Cahoun, E. Read, et al, *Nature*, 446, 782-786 (2007).

ISLiM成果報告会2011





## 謝辞・参考文献

- ProteinDF開発メンバー(敬称略)
 

次世代生命体統合シミュレーションソフトウェアの研究開発	超並列
平野 敏行	東大生研
上村 典子	東大生研
恒川 直樹	東大生研
松田 潤一	東大生研
- イノベーシヨシ基盤シミュレーションソフトウェアの研究開発
 

平野 敏行	東大生研	サブリーダー, エンジン等
阿部 敏彦	東大生研	NMR等
下向 智美	東大生研	GUI, モデリング等
谷村 景貴	東大生研	IRC等




**研究協力**

千葉 貢治	九大院シス生	ヘムタンパク質研究
田村 隆	岡山大院自然科	実験酵素研究

- 京速コンピュータ東京の試験利用を利用いたしました。

ISLiM成果報告会2011

# Platypus-QM/MM の開発と並列性能

中村 春木

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





## 発表者紹介

1980年 3月 東京大学大学院理学系研究科物理学専攻博士課程修了 理学博士  
1980年 4月 東京大学工学部物理工学科助手  
1987年 8月 蛋白工学研究所・第二研究部  
1995年 10月 生物分子工学研究所・情報解析研究部門  
1999年 4月 大阪大学蛋白質研究所 教授

## 研究分野

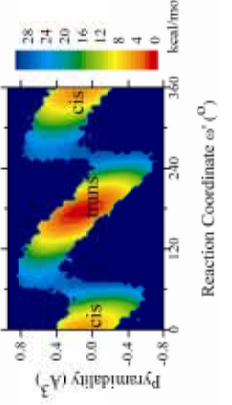
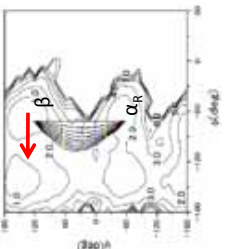
生物物理学、蛋白質科学、蛋白質計算科学、構造バイオインフォマティクス








## Platyplus-QM/MMの開発と並列性能

大阪大学蛋白質研究所  
附属プロテオミクス総合研究センター  
センター長  
教授 中村 春木

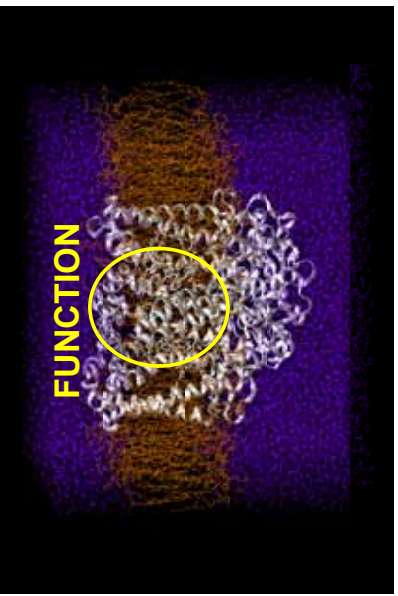
ISLiM成果報告会2011



## 背景・目的

- 1) 生体高分子は巨大で、ヘテロで、ダイナミックな系
- 2) 生体高分子の機能の発揮は局所部位の電子状態に依存する

上記2つの特徴を持つ生体高分子をリアルなモデルによって扱い、高速に計算できる「新規QM/MM 連成計算手法」を開発して機能を解明し、分子設計へ



2

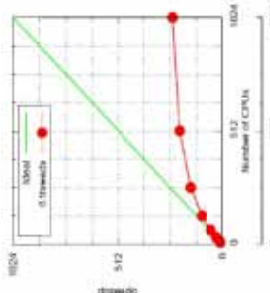



## 生体高分子を対象とした 超並列 *ab initio* QM/MM連成計算 プログラム (*platyplus*) の開発

PLATform for dYnamic Protein Unified Simulation



Parallelized DFT computation by *Platyplus-QM* for a system with cc-pVDZ (2728 AO) (整備中の「京」システムによる暫定的な値)

- 1) オリジナルなコードによって超並列を実現する *Ab initio* QM (量子化学: RHF, UHF, R-DFT, U-DFT, CASSCF, CIS, CIS(D), CIS-DFT, MP2) と、古典力学による MD (分子動力学) との連成計算
- 2) 周期境界系/非周期境界系に対して、PME法に替わる新たな長距離相互作用計算手法 (Zero-Multipole Summation Method) を導入し、*ab initio* QM/MM連成計算を高精度で高速に実行
- 3) Multicanonical MDやUmbrella Sampling法等の手法による *ab initio* QM/MM の自由エネルギー計算が可能
- 4) 4000コア以上までスケールアップにチューニングされた並列コード  
(PCクラスターでは並列化率 99.82 %  
→ 京では99.97 %を達成: 整備中の「京」システムによる暫定的な値)

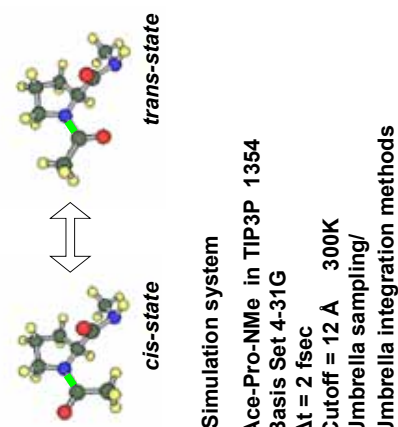


Bacterial chlorophyll dimer of Photosynthetic Reaction Center with two His side-chains.

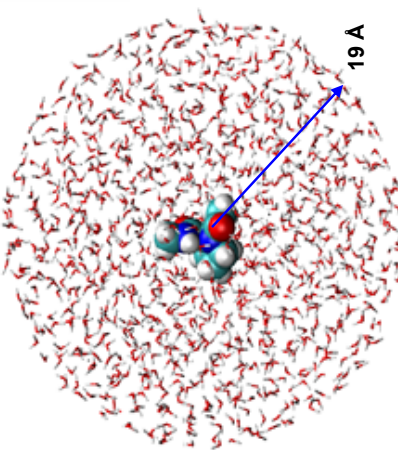
3

## Cis-trans isomerization process of a solvated proline dipeptide revealed by hybrid-*ab initio* QM/MM

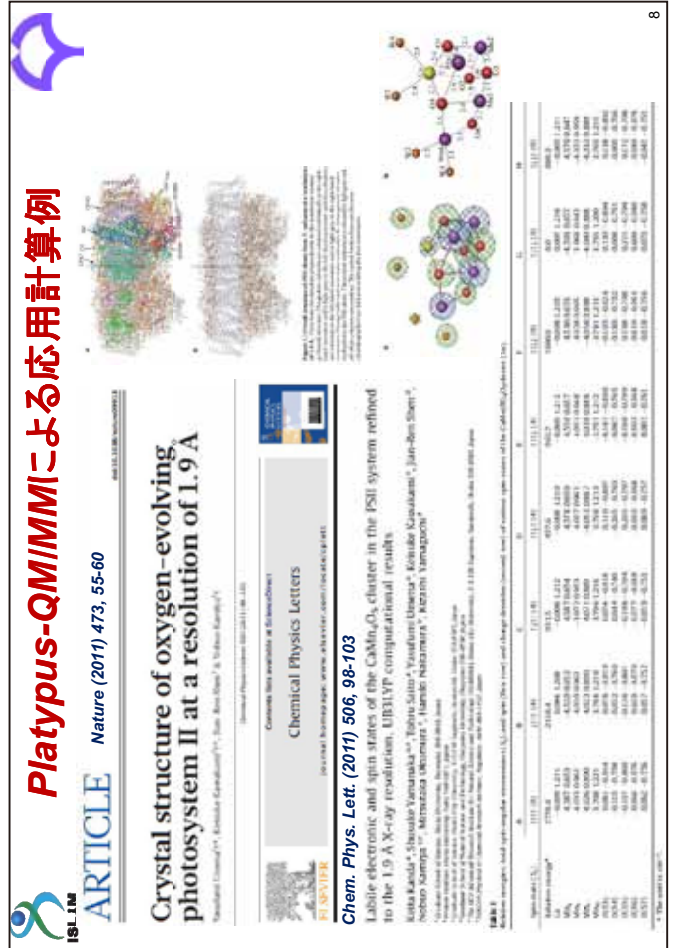
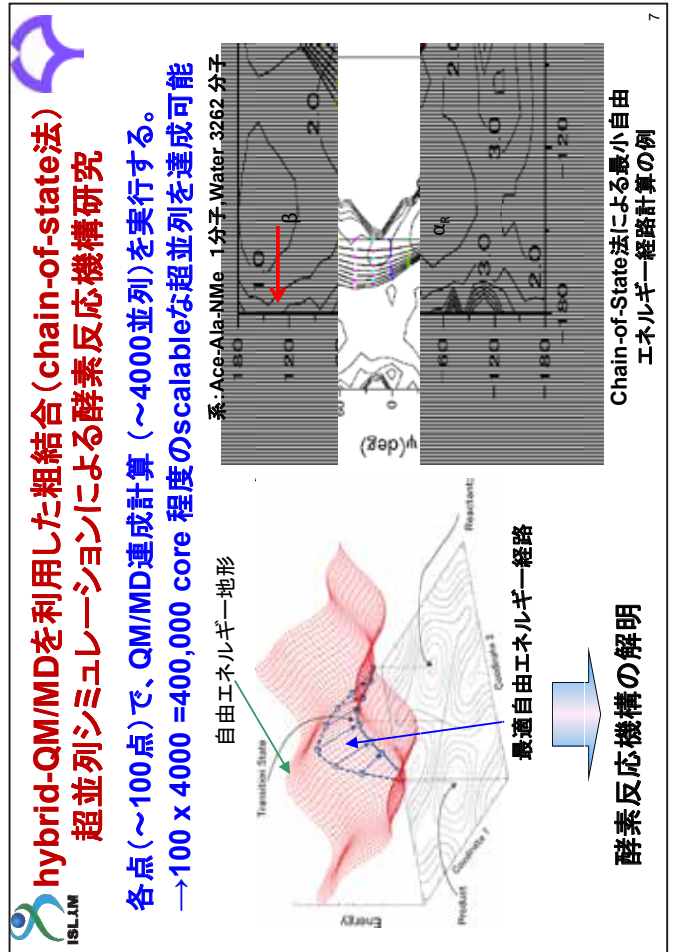
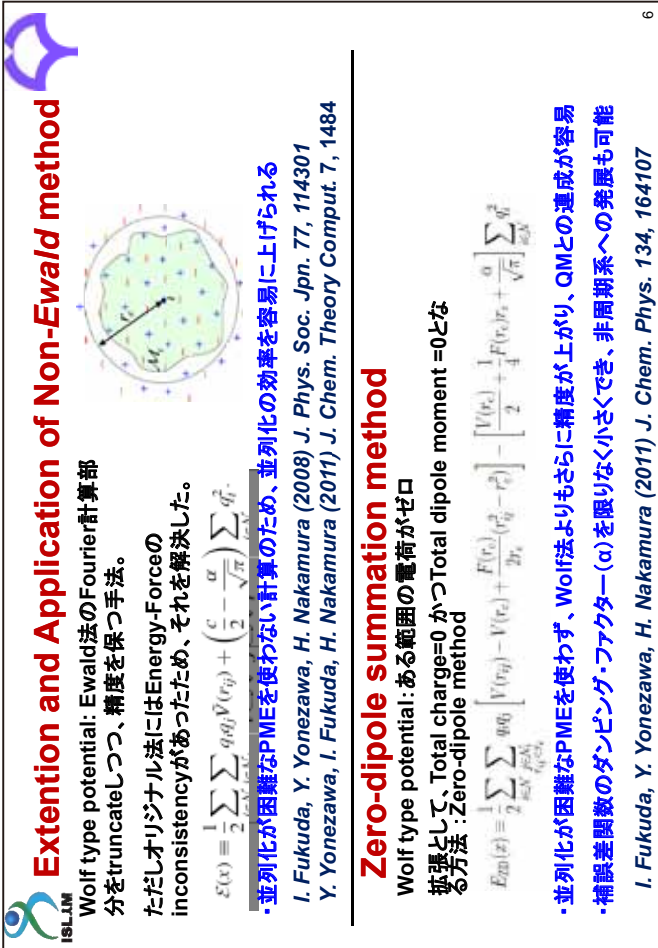
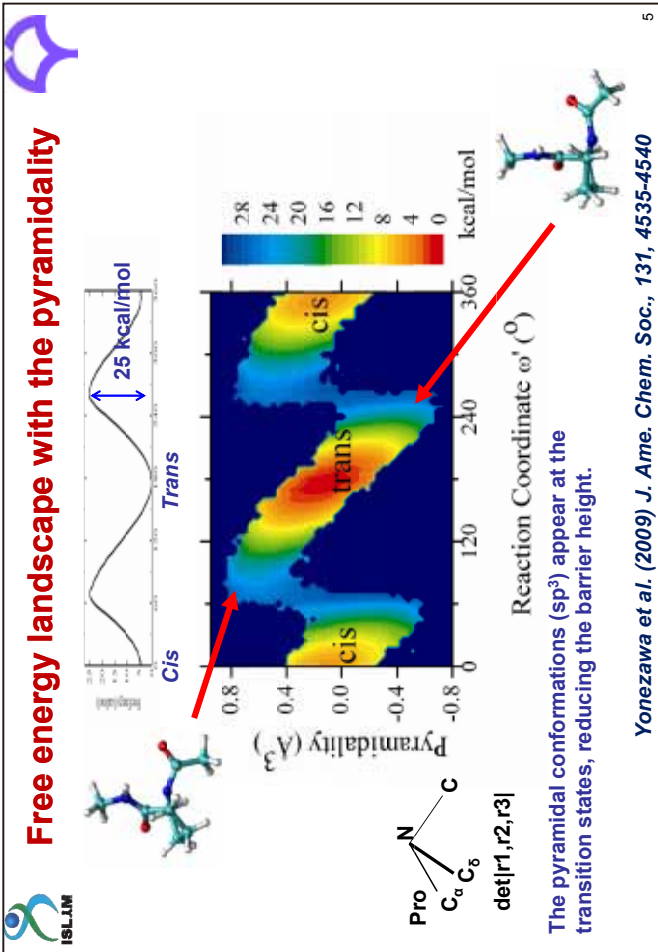


Simulation system  
Ace-Pro-NME in TIP3P 1354  
Basis Set 4-31G  
 $\Delta t = 2$  fsec  
Cutoff = 12 Å 300K  
Umbrella sampling/  
Umbrella integration methods  
Computation with *Platyplus* program



Yonezawa et al. (2009) *J. Am. Chem. Soc.*, 131, 4535-4540

4





## プロジェクト終了時の達成目標

- 1) オリジナルなプログラム (Platypus-QM/MD) によって、種々の *Ab initio* QM (量子化学計算: RHF, UHF, R-DFT, U-DFT, CASSCF, CIS, CIS(D), CIS-DFT, MP2) と、古典力学による MD (分子動力学) との連成計算を、「京」の上で高速に実施できる仕組みを確立する
- 2) 周期境界系および非周期境界系に対して、PME法に替わる新たな長距離相互作用計算手法 (Zero-Multipole Summation Method) を導入し、上記 *ab initio* QM/MD 連成計算を高精度で高速に実行し、反応の自由エネルギー地形を描く仕組みを構築する
- 3) 粗結合 (chain-of-state法) の超並列シミュレーションにより、種々の酵素反応機構を、電子状態と蛋白質・溶媒のダイナミクスを含めて、京の上で計算できる仕組みを構築する
- 4) 上記の仕組みを用いた実証研究を行う

9



## 共同研究者

米澤 康滋 (阪大蛋白研) → 現・近畿大)  
 山中 秀介 (阪大蛋白研) → 現・阪大理)  
 下山 紘充 (阪大蛋白研) → 現・北里大薬)  
 鷹野 優 (阪大蛋白研)  
 山崎 秀樹 (阪大蛋白研)  
 大島 勘二 (阪大蛋白研)  
 福田 育夫 (理研)  
 中田 一人 (NEC、阪大蛋白研)

なお、本発表の一部は、理化学研究所が実施している京速コンピュータ「京」の試験利用によるものです。

ISLiM成果報告会2011

10

# 全原子分子動力学計算ソフトウェア MARBLE の開発と 多剤排出トランスポーターAcrB への応用

池口 満徳

横浜市立大学 大学院生命ナノシステム研究科  
分子スケール研究開発チーム





## 発表者紹介

- 1994 年 3 月 東京大学大学院農学系研究科応用生命工学専攻博士課程修了
- 1994 年 4 月 東京大学農学部助手
- 2001 年 4 月 横浜市立大学大学院総合理学研究科助教授
- 2009 年 4 月 横浜市立大学大学院生命ナノシステム科学研究科准教授

## 研究分野


生物物理学, 分子シミュレーション





# 全原子分子動力学計算ソフトウェア MARBLEの開発と 多剤排出トランスポーターAcrBへの応用

横浜市立大学 大学院生命ナノシステム研究科  
分子スケール研究開発チーム  
池口 満徳, 山根 努



ISLiM成果報告会2011

1






## 背景・目的

- 背景**
  - タンパク質やDNAなどの生体高分子の立体構造を原子レベルで明らかにする  
構造生物学の進展
  - 実験的に決定される立体構造は、静的な構造であることがほとんど
  - しかし、生体分子が機能を果たすときには、動的であることが重要
  - 例：分子モーター(ATP等のエネルギーを使って構造変化することで力を出す)
  - 計算機中で、物理学法則に基づき、生体分子の動的構造をシミュレーションする  
方法が有効
  - 特に、超並列化時代に突入して、計算対象が、方法開発に重きを置いたモデル  
タンパク質から、生物学的に注目されるタンパク質に移行したことが重要な点
- 目的**
  - 「京」で効率的に動作する全原子分子動力学(MD)ソフトウェアMARBLEを作製
  - 実用性検証のため、多剤排出トランスポーターAcrBの分子シミュレーションを行い、  
その機能メカニズムを解明

ISLiM成果報告会2011

2






## 概要・アプローチ

- 対象モデル:**
  - 水素まで含んだ全原子モデルを対象とする。
  - 周囲の水分子、イオンなどの溶媒は露わに扱い、誘電率一定などの近似は  
用いない。
  - 膜タンパク質の計算時には、膜分子も原子レベルで扱う。
- シンプレクティック部分剛体時間積分法**
  - 精度の高い全エネルギー保存
- 並列化アプローチ**
  - 空間分割を用いて、通信量の削減を図る
- 相互作用計算**
  - 近距離力: 空間分割により、近接ノード通信を用いる
  - 遠距離力: Particle Mesh Ewald法によりFFTを用いて計算
- 実験データによる検証**
  - NMR, SAXS


ISLiM成果報告会2011

3

## 研究・開発のステップチャート

出力：座標の履歴



**生体超分子用全原子  
シミュレーション技術の開発**

- ✓ 超並列化ソフトウェア  
MARBLE
- ✓ 世界標準アルゴリズムを装  
備 (PMIE等)
- ✓ 先端アルゴリズムも装備  
(シンプレクティック剛体法  
等)


**シミュレーション結果の解析  
技術の開発**

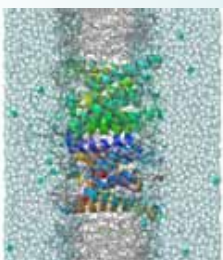
- ✓ 立体構造変化
- ✓ 立体構造ゆらぎ
- ✓ 自由エネルギー
- ✓ 実験との対応  
NMR, SAXSなど

**多剤排出トランスポーター  
AcrBの全原子分子動力学シ  
ミュレーション**

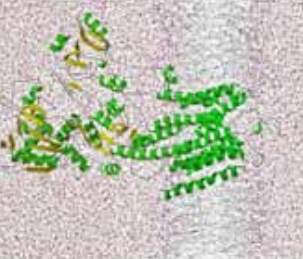
- ✓ 分子スケールチームで共  
通のターゲット
- ✓ 膜タンパク質
- ✓ 構造変化が機能に重要
- ✓ ゆらぎと構造変化

### MARBLEを用いてMDを行った例







Aqueporin  
(FEBS Lett 2005; Biophys J 2007)



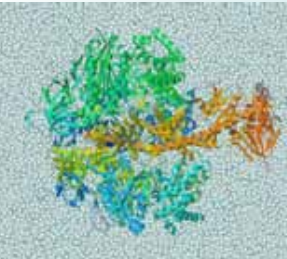
Ca<sup>2+</sup> ATPase  
(JACS 2006, PNAS 2010)




PhoB  
(Proteins 2008; JACS 2010)



EcoO109I  
(Biophys J 2009)




F<sub>1</sub>-ATPase (GPL 2010; JCC 2010; JACS 2011)




$\alpha$  Lactalbumin  
(JMB 2008; 2007)

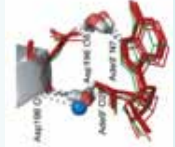
### 実験データによるMARBLEの検証例



#### NMRでの検証

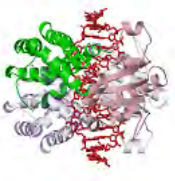


DNA結合タンパク質  
PhoB

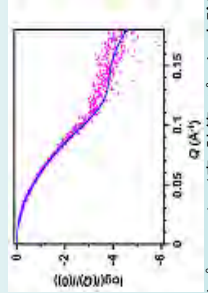


水を介した  
分子認識

#### X線澄液散乱実験での検証



制限酵素EcoO109I




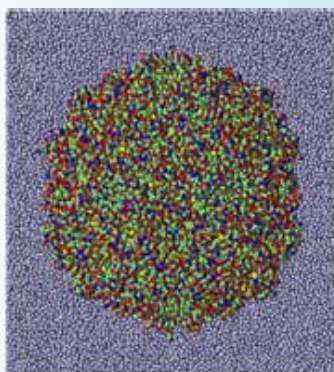
構造ゆらぎ方向(赤)と  
構造変化方向(青)の一致

散乱プロファイル(青:計算、ピンク:実験)

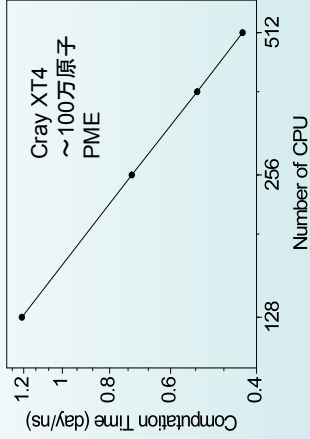
(Biophys J, 2009)

### 前バージョンでの並列化効率





水中のウイルス




前バージョン

Cray XT4  
~100万原子  
PME

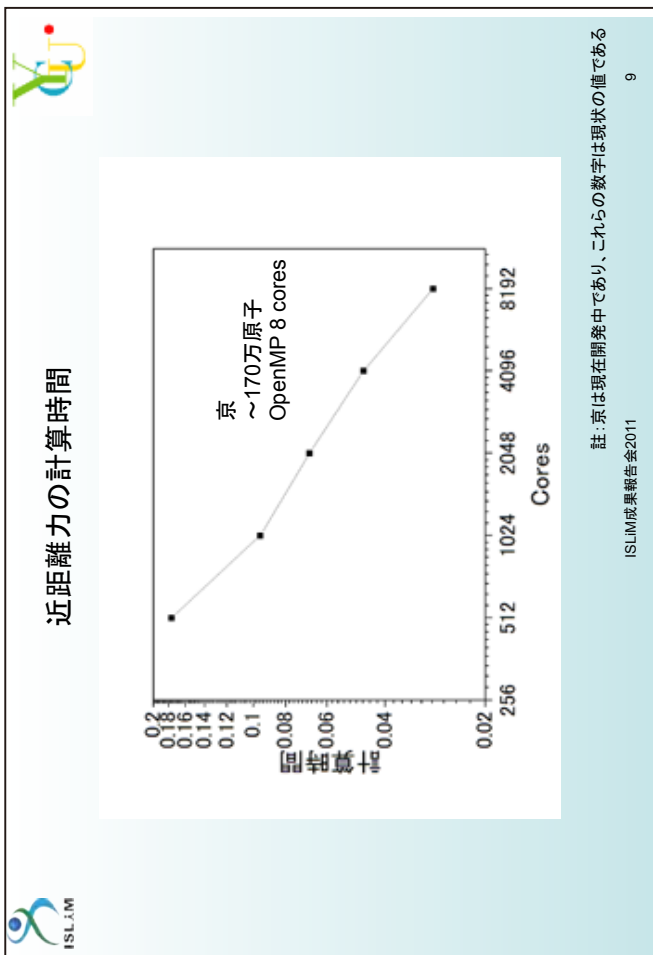
Number of CPU	Computation Time (day/ns)
128	~1.1
256	~0.7
512	~0.45

1000並列くらいで頭打ち (理研青山さん)  
Flat MPIで書かれていた

### 現在までの研究開発成果



- 旧バージョン
  - Flat MPIで書かれており、並列化効率は、1000コア程度まで。
- 通信の最適化
  - 3Dトランスネットワークに合わせて、分割法や通信法を抜本的に見直した
- OpenMP対応
  - ノード内は共有メモリ利用、ノード間はMPI利用することで、MPIプロセスの削減、通信の効率化を目指す
- MDの計算は、短距離力と長距離力の計算からなる。短距離力に適用
  - 短距離力の京でのコンパイル、テスト実行
  - 遠距離部分の高速化中
    - 長距離力には3DFFTを利用。3DFFTの2次元分割化
- 実験データによる精度検証
  - NMR, SAXS
- 多剤排出トランスポーターAcrBでの検証



### AcrBの全原子分子動力学シミュレーション

ISLIM

Program : MARBLE  
Force field: CHARMM27  
Electrostatic interaction: Non cut-off (Particle-mesh Ewald)

Ensemble: NPT  
Temperature: 300K,  
Time step: 2fs

System  
Lipid: POPE 809 molecules (101,125 atoms)  
water: TIP3P model 108,385 molecules (325,155 atoms)  
protein: (PDB ID: 2DHH) (No drug containing structure)  
Missing part (499-512) was added.  
(47,724 atoms)  
Ion : Na<sup>+</sup> (36 atoms)  
**Total: 474,040 atoms**  
Box size: 160Å X 160Å X 180Å

### 膜貫通部位のプロトン化状態の決定

ISLIM

プロトン濃度勾配のエネルギー利用には、  
 ✓ 水分子がそれぞれ側からアクセス可能  
 ✓ 同時には両方からアクセスは不可  
 ✓ プロトン着脱可能な基 (カルボキシル基等) を経由して、プロトンは透過

膜貫通部位で、プロトン着脱可能な基は、  
**Asp407, Asp408** のカルボキシル基  
 それぞれの構造でのプロトン化状態を決定

Binding: Asp407, Asp408, Lys940, Thr978  
 Extrusion: Asp407, Asp408, Lys940, Thr978  
 Access: Asp407, Asp408, Lys940, Thr978

Extrusion構造のLys940のコンフォメーションが異なる

### プロジェクト終了時の達成目標

ISLIM

- 「京」で効率的に動作する全原子分子動力学(MD)ソフトウェアMARBLEを完成させる
- 特に、遠距離力まで含んだ形で、OpenMP化して、公開する
- 公開に向けたドキュメント作成を行う
- 多剤排出トランスポーターAcrBでは、薬剤を入れたシミュレーションを行い、プロトン濃度勾配のエネルギーをどのように利用して、薬剤排出を行っているのか、そのメカニズムの解明を目指す
- 全原子モデルは、疎視化モデルと電子モデルの中間にあるので、双方の側との連携を目指す。
- 具体的には、2nd Runnerとして、Platyplus-MM/CGと連携して、薬剤排出過程の全原子構造変化MDを行う。また、Platyplus-QM/MM-MEと連携して、プロトン濃度勾配利用の際のプロトン脱着の量子過程の計算を行う。

謝辞:京での計算に関しては、京速コンピュータ東京の試験利用での結果です。

ISLIM成果報告会2011

12



# 粗視化分子ソフトウェア CafeMol による 多剤排出トランスポーターと遺伝子動態の シミュレーション研究

高田彰二

京都大学大学院  
理学研究科生物科学専攻





## 発表者紹介

- 1990年3月 京都大学理学研究科化学専攻修士課程修了
- 1991年3月 岡崎国立共同研究機構分子科学研究所技官
- 1998年4月 神戸大学理学部化学科講師・助教授
- 2007年3月 京都大学理学研究科生物学専攻准教授（現在に至る）

## 研究分野

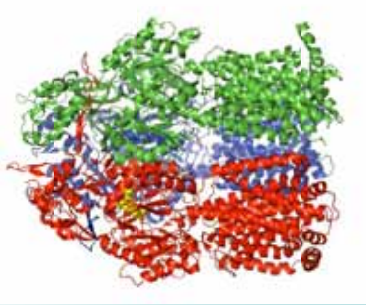
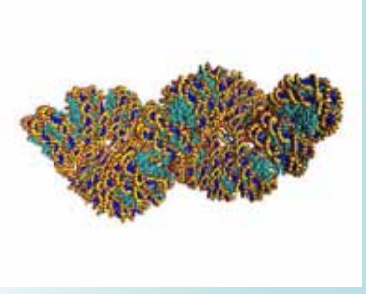
理論生物物理学





## 粗視化分子ソフトウェアCafeMolによる 多剤排出トランスポーターと遺伝子動態の シミュレーション研究

京都大学理学研究科、准教授  
高田彰二

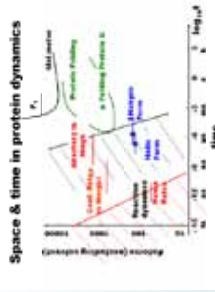
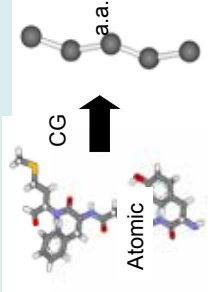
ISLIM成果報告会2011

1



## 背景・目的

- 背景
  - 生命体は階層的：分子スケールでは、原子レベルの構造的知見が分野を革新してきた。しかし、原子レベルから細胞レベルまでにはギャップがある。また、ゆらぎが大きいために構造解析が困難な場合が多い。
  - 大規模長時間分子シミュレーション実現のために、**粗視化モデル**が有用
- 目的
  - 原子レベルのシミュレーション法と構造生物学的実験データをもとにして、
  - 1)粗視化分子モデルの構築、
  - 2)それを実装したシミュレーションソフトウェア**CafeMol**の開発、および、
  - 3)細胞生物学の基礎的生命現象の分子スケールでの研究を目的とする。

ISLIM成果報告会2011

2






## CafeMol:概要・アプローチ

- 研究開発コードCafeMolの概要
  - 粗視化分子モデル計算により大規模生体分子の長時間シミュレーションを行う。
- アプローチ
  - 離散化(計算モデル化)の方法
    - 粗視化された分子モデルによる古典分子動力学法
  - 計算方法
    - Languevin方程式の時間発展を数値的に積分する。
  - 並列化の方法
    - Neighbor list方式、レプリカ交換法
  - 開発言語とライブラリ
    - Fortran90, MPI, OpenMP
  - コードの公開状況
    - <http://www.cafemol.org> からダウンロード可。ただし生体膜、核酸については要相談。
- 現状での計算規模
  - 1万残基タンパク質のミリ秒相当のシミュレーションが可能
  - PCクラスターで8192コア並列
  - メモリ容量 2 GB/コア、ディスク容量 1 TB

ISLIM成果報告会2011

3

## CafeMol: 現在までの研究開発成果

- 現時点の開発ソフトウェアCafeMolについて
  - 達成済み、
    - 蛋白質、核酸、膜の粗視化モデルの実装
    - PCクラスターでの8000並列(RICC)
    - 「京」への実装、単体性能32%(2コア利用)
      - 京は現在開発中であり、これらの数字は現状の値である
  - 克服すべき問題と解決策
    - 粗視化モデルの合理的構築、精度向上 → 原子レベルの分子力場からのマルチスケール技法によるパラメータ導出
    - 並列化 → スレッド並列、レプリカ交換法
  - 計算事例
    - 多剤排出トランスポーター、スクレオソーム、キネシン(分子モーター)他多数
  - 実験検証
    - 多剤排出トランスポーター: 村上聡氏(東工大)と共同研究中
    - キネシン: 富重雄氏(東大)らと共同研究中

ISLIM成果報告会2011

4

## 原子モデルに基づく粗視化モデルの構築

Atomic interaction-based CG model,  
Li, Wolyynes, Takada PNAS 2011

Atomic → CG

a.a.

$$V = \sum_i k_{ij}(r_i - r_j)^2 + \sum_i k_{\theta}(\theta_i - \theta_0)^2 + \sum_i \{ \epsilon_{\phi_i}^1 [1 - \cos(\phi_i - \phi_0)] + \epsilon_{\phi_i}^3 [1 - \cos 3(\phi_i - \phi_0)] \} + \sum_{i>j>3} \epsilon^{ij} [5(\sigma_{ij}^{ij}/r^{ij})^{12} - 6(\sigma_{ij}^{ij}/r^{ij})^{10}] + \sum_{i>j>3} \delta(C_i/r^{\delta_i})^{12}$$

1) 全原子モデル(AA)の対エネルギーから、アミノ酸対エネルギー  $\epsilon_{ij}$  を導く  
 $E^U(R_{ij}) = \sum_{i \in A, j \in B} u_{AA}(r_{ij}) \quad u_{AA}(r) = V(r) + \Delta G_{pol}^{GB}(r) + \Delta G^{SA}(r)$

2) 全原子モデルのゆらぎから粗視化モデルのパラメータを導く (40 蛋白質)

param	$K_{\phi}$	$K_{\theta}$	$K_{ij}^U$	$K_{ij}^E$	$K_{ij}^F$	$\sigma_{ij}^{\phi}$	$\sigma_{ij}^{\theta}$	$\sigma_{ij}^E$	$\sigma_{ij}^F$	$\sigma_{ij}^T$	$\sigma_{ij}^C$	$\epsilon_{ij}^{GB}$	$\epsilon_{ij}^{SA}$
Av.	109.94	13.40	40.03	17.32	19.35	11.70	0.28	1.76	1.32	0.82	0.81	0.81	0.37

Protein O

Protein L

Probability vs C-C

0.694  
0.758

別のタンパク質のゆらぎ、構造変化、フォールディング経路予測で実証

## CafeMol: 実装とHPC

さらに、核酸、膜、複合体のシミュレーション

Fortran 90, MPI

Towards 10<sup>6</sup> atoms & ms-time events on K-computer

Version 1.0 (only for proteins), available at [www.cafemol.org](http://www.cafemol.org)

Protein: Structure-based, switching

DNA: Structure-based+ P-Chem

Membrane: P-Chem

単体性能のチューニング

MPIによる並列と OpenMPIによるスレッド並列

RICC、ヌクレオソーム20で並列化率99.995%

多次元レプリカ交換法、インバランスの解消

## 実証研究1) 多剤排出トランスポーター

ToIC docking domain

Porter domain

Trans-membrane domain

ligand

Binding

Access

Extrusion

Yao

Murakami 2006

Yao et al., Nature Comm 2010

## 薬剤の侵入経路

Tunnel 1 膜界面からの長い経路  
Tunnel 2 ベリプラズムからの短い経路

Tunnel 3

Tunnel 2

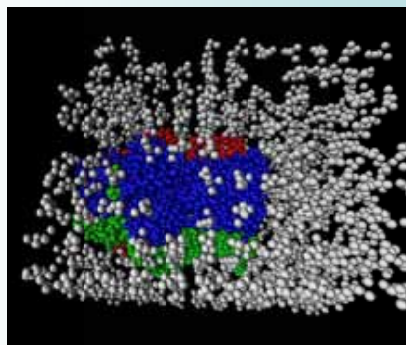
Tunnel 1

Inner membrane

A B E

### 薬剤動態シミュレーション

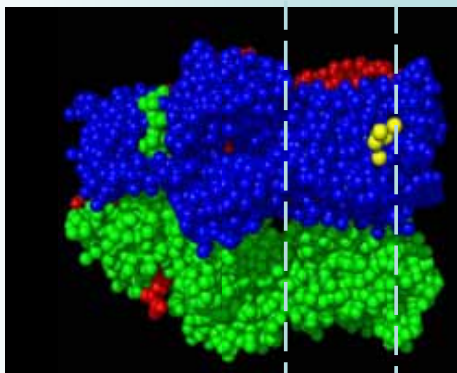
- 273 ligand molecules in a 160x160x300 (Å) box
- 300K Langevin
- 100 trajectories
- Switching  
ABE → BEA → EAB



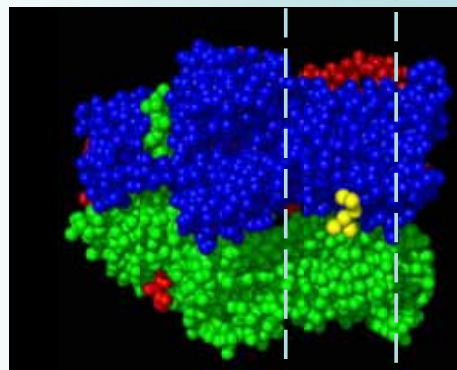
### 薬剤取り込みと排出

$C_{HP}=0.7, C_{MEM}=0.2$  (intermediate hydrophobicity & lipophilicity)

Tunnel 2

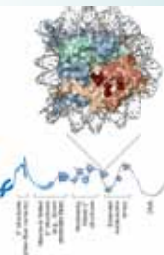


Tunnel 1

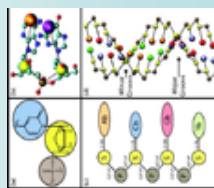


### 実証研究2) 遺伝子動態:ヌクレオソーム

クロマチン高次構造  
 高度圧縮と塩基配列読み取り



熱的および力学的ヌクレオソーム分解と再構成の解析



DNA  
 粗視化モデル



### 遺伝子動態「京」による計算課題

ヌクレオソームの可逆的巻き戻り



転写に必須: RNA-Pのアクセス  
 秒オーダーの現象



ヒストンシヤペロのアクセス



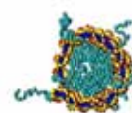
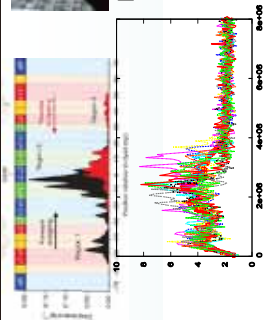
ヌクレオソームからクロマチンへ





転写因子などの塩基配列探索過程



Kenzaki





## プロジェクト終了時の達成目標

- CafeMolの「京」での計算規模
  - 10万粒子(100万原子)系の秒相当のシミュレーション
  - 反復回数 $10^{10}$ 回以上
  - メモリ容量 2 GB/コア、ディスク容量 1 PB
- クロマチン構造モデリング、遺伝子動態の解析

### 謝辞

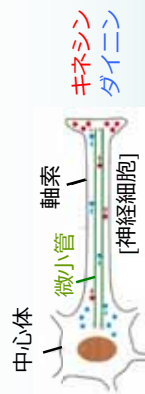
京での計算に関しては京速コンピュータ東京の試験利用の結果です。また、PCクラスタでの性能計測に関しては理化学研究所情報基盤センターのRICCを使用しています。


検崎博生、金田亮、Xin-Qiu Yao  
Wenfei Li、寺川剛

ISLiM成果報告会2011


14

## 実証研究3)キネシン(分子モーター)





**複合体の高精度モデリング  
実験家(富重氏、吉川氏)と共同研究**



Kanada



(謝辞)

本資料集に記載されている「京」での計算は、2011年3月の「京」の特別運用およびその後の試験利用によって行われたものです。

また、本資料集に記載されている「京」を使用した測定値は、開発整備中の「京」による、測定時点での数値です。