

2015年度バイオスーパーコンピューティング・ソフト講習会シリーズ

量子化学計算/分子動力学連成計算ソフト 「Platypus-QM/MM」 講習会

中田 一人

2016年3月28日

Orchestrating a brighter world

未来に向かい、人が生きる、豊かに生きるために欠かせないもの。
それは「安全」「安心」「効率」「公平」という価値が実現された社会です。

NECは、ネットワーク技術とコンピューティング技術をあわせ持つ
類のないインテグレーターとしてリーダーシップを発揮し、
卓越した技術とさまざまな知見やアイデアを融合することで、
世界の国々や地域の人々と協奏しながら、
明るく希望に満ちた暮らしと社会を実現し、未来につなげていきます。

本日の内容

1. Platypus-QM/MM プログラム概要
2. Platypus-QM プログラム概要

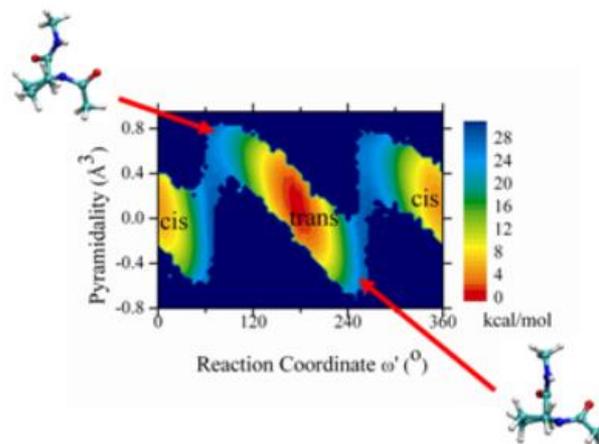
Platypus-QM/MM プログラム概要

どんなことができるのか

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（計算モデル化）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Wolf等
- 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, ScaLAPACK, LAPACK
- コードの公開状況
未公開
- 現状での計算規模
 - プロリン及びN-methylacetamide（量子系）の水溶液中（水分子は古典粒子）における異性化熱力学過程の解明
 - 20000原子を蛋白研PCクラスタ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1000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GPB



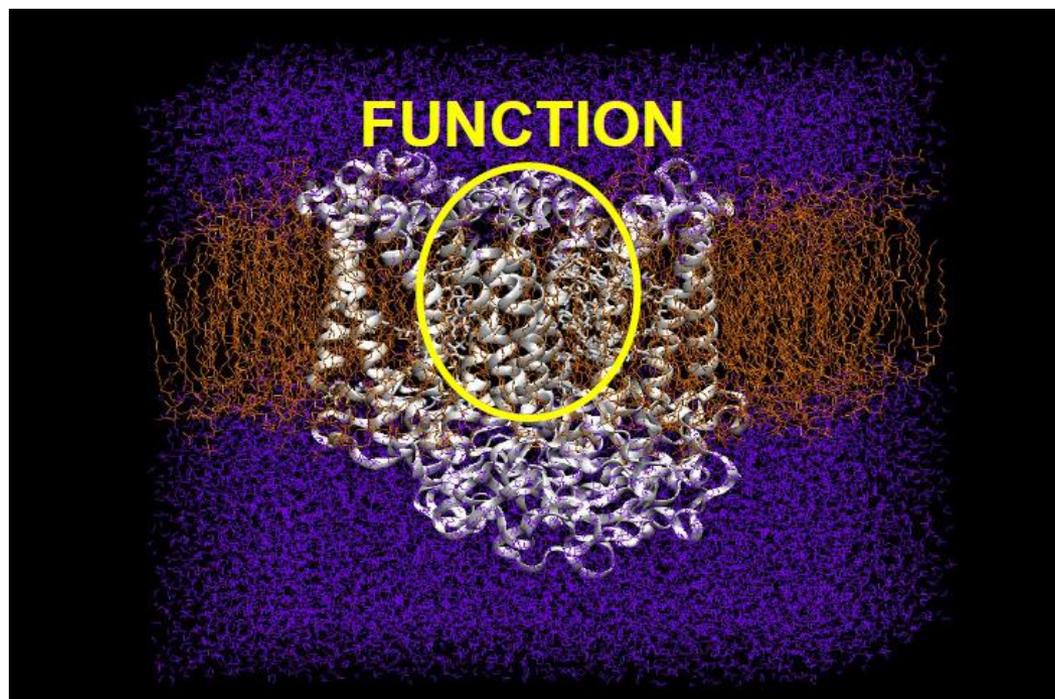
ピラミダリティの反応座標依存性の量子古典連成計算

- どんなことができるか
 - 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。

[このページのTOP▲](#)

次のような特徴をもつ生体高分子を、リアルなモデルで扱って、機能解明し、分子設計へ

- 巨大で、ヘテロで、ダイナミックな系
- 機能の発揮は局所部位の電子状態に依存



Platypus-QM/MM (PLATform for dYnamic Protein Unified Simulation)

古典論 : cosgene

大阪大学 中村春木 先生

近畿大学 米澤康滋 先生

広島市立大学 鷹野 優 先生

量子論 : Platypus-QM

大阪大学 山中秀介 先生 (DFT)

NEC 中田一人 (MO)

+ 数値計算ライブラリ (ご支援いただいております)

筑波大学 櫻井鉄也 先生 連立一次方程式 (Shifted CG法)

理研 今村俊幸 先生 固有値問題 (EigenExa, LOBPCG)
<http://www.aics.riken.jp/labs/lpnctr/EigenExa.html>

Platypus-QM/MM

$$E_{tot} = E_{QM} + E_{MM} + E_{QM/MM}$$

古典論：cosgene

String法

分子動力学

+

量子論：Platypus-QM

CASSCF ⇒

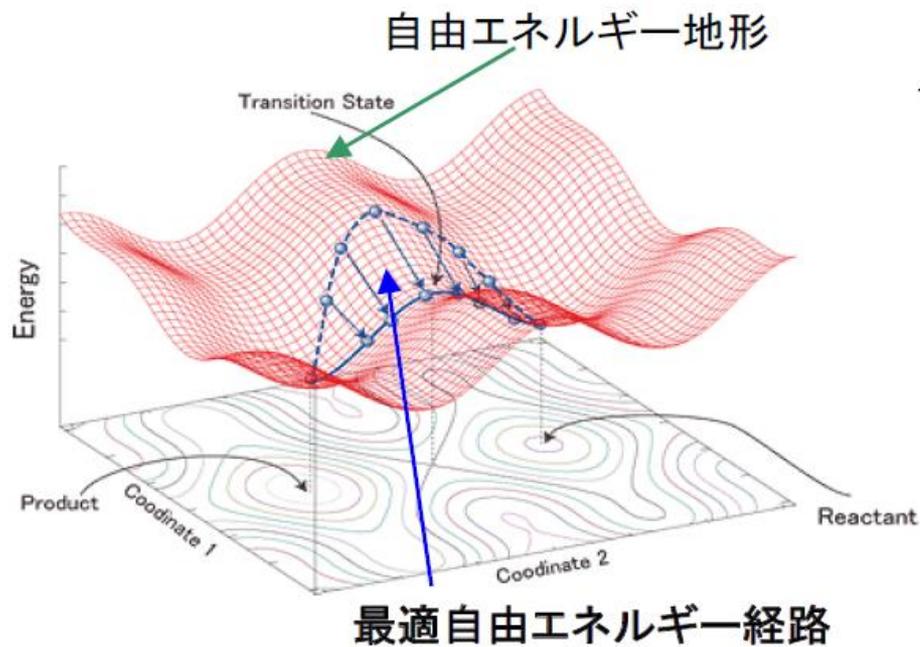
化学反応

励起状態

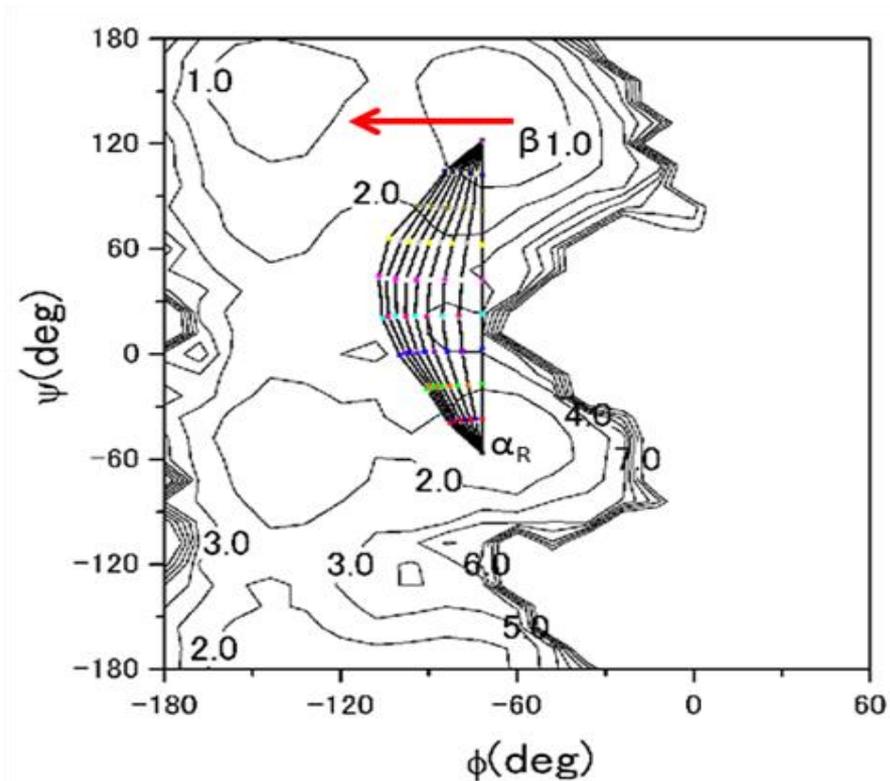
電子配置の重ね合わせ
が必要な状態

(Complete Active Space SCF, 完全活性空間SCF)

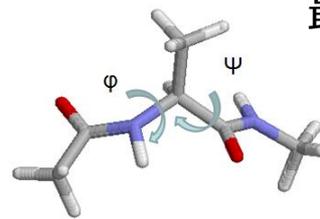
(MR ; Multireference, 多配置)



酵素反応機構の解明



String法による
最小自由エネルギー経路



整備でき次第、オープンソース公開へ …

- ① ソースコードの整備
- ② 使い勝手の向上
- ③ マニュアルの整備

プログラム詳細は論文をご参照ください：

“Development of massive multilevel molecular dynamics simulation program, platypus (PLATform for dYnamic protein unified simulation), for the elucidation of protein functions”,

Journal of Computational Chemistry に採択され掲載決定。

① Intel CPU のクラスタ

② 京コンピュータ

予定：Xeon Phi への対応

MPI + OpenMP のハイブリッド並列

DGEMM(行列-行列積) 利用によるSIMD対応

なぜ？ HFではないのか

シュレーディンガー方程式 $\hat{H}\Psi = E\Psi$

	CI 係数	MO 係数
	$\Psi = \sum_I^{N_{\text{CSF}}} C_I \Phi_I$	$\phi_i = \sum_{\mu}^{N_{\text{AO}}} C_{\mu i} \chi_{\mu}$
HF	不要	最適化
CI	最適化	固定
CASSCF	最適化	最適化

- Ψ : 波動関数
- Φ_I : 電子配置 (CSF; Configuration State Function)
- ϕ_i : 分子軌道 (MO; Molecular Orbital)
- χ_{μ} : 原子軌道 (AO; Atomic Orbital)

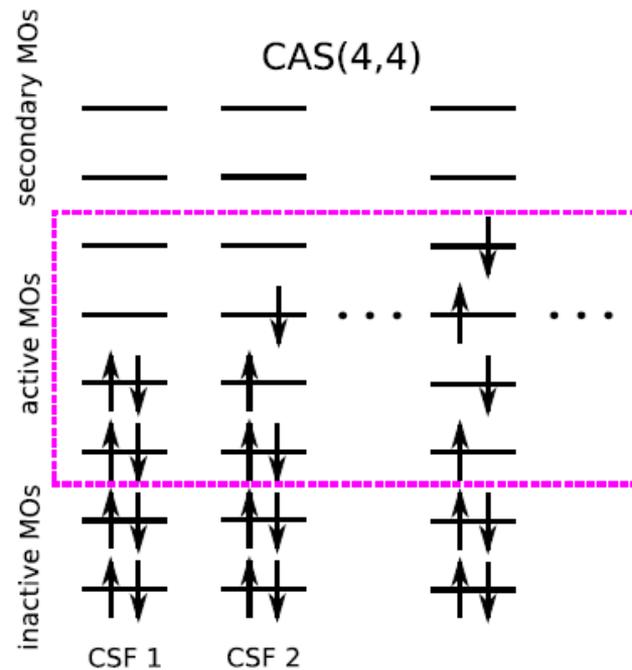
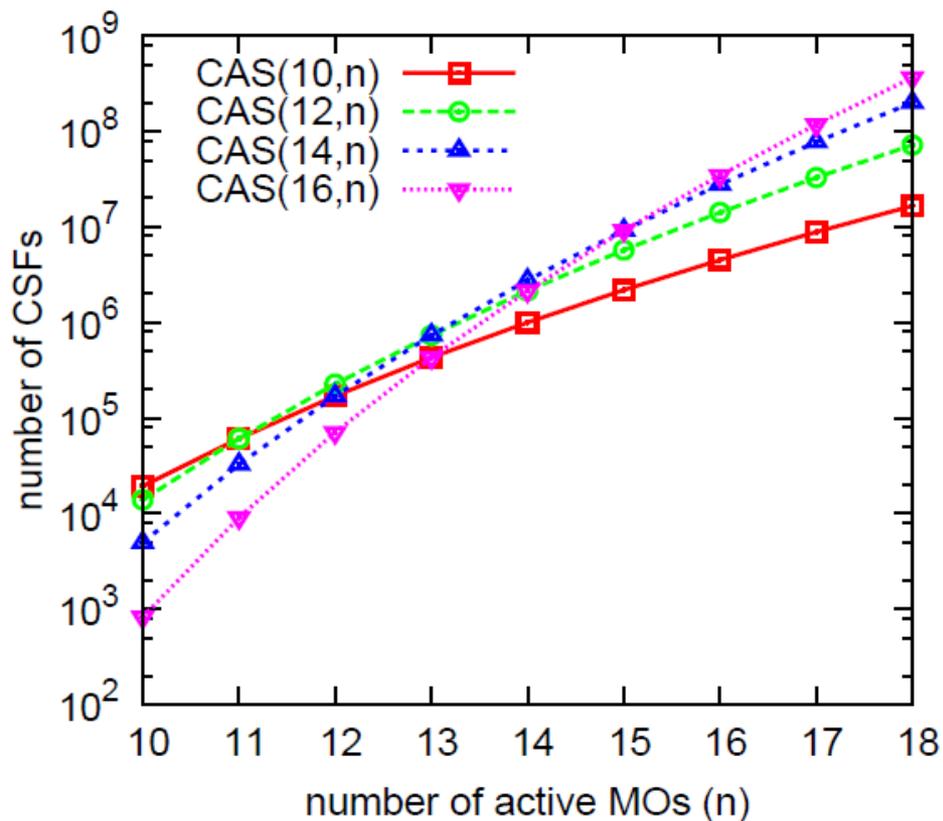
HF (Hartree-Fock) : 単スレーター行列式。基底状態を計算。
CI (Configuration Interaction) : 複数の電子配置。基底状態や励起状態を計算。
CASSCF (Complete Active Space Self-Consistent Field) : 同上。

両方とも最適化

CI 係数 : $\Psi = \sum_I C_I \Phi_I$ 線形方程式

MO 係数 : $\phi_i = \sum_{\mu} C_{\mu i} \chi_{\mu}$ 非線形方程式

- 収束まで反復計算
- CSF(電子配置関数)の空間を CAS (Complete Active Space) にすると、一般的な MCSCF よりも比較的収束しやすい



- 問題点：active空間を広げると、CSFの個数が急増。
- 表記法：CAS(active電子数, active MO数)

なぜ？ DFTの電子相関ではないのか

Kohn-Sham DFT (KS-DFT) は、よく使われている手法の一つ

波動関数理論 (WFT) における
Hartree-Fock (HF) と同等の計算コストで
動的電子相関を含んだ計算結果が得られる

⇒ 大きな分子の計算が可能

動的電子相関 とは、電子の短距離的な相互作用
電子密度 $\rho(r)$ の局所的な情報から見積もる

しかし、KS-DFT と HF は、単一スレーター行列式である

	単一スレーター 行列式	多参照(Multireference, MR) 多配置(Multiconfiguration, MC)
WFT (波動関数理論)	HF	多参照配置間相互作用 (MRCI)法 多配置自己無撞着場法(MCSCF)法など 定番手法 ⇒ <u>CASSCF</u>
DFT (密度汎関数理論)	KS-DFT	まだ定番手法がない ⇒ CASDFT

単一スレーター行列式

制限HF (Restricted Hartree-Fock, RHF)

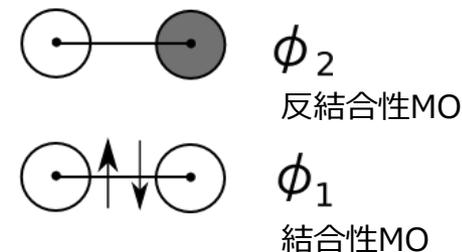
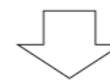
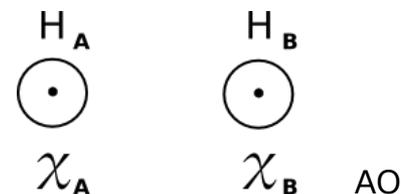
$$\Psi^{\text{RHF}} = |\varphi_1(1)\alpha(1)\varphi_1(2)\beta(2)\rangle$$

非制限HF (Unrestricted Hartree-Fock, UHF)

$$\Psi^{\text{UHF}} = |\varphi^+(1)\alpha(1)\varphi^-(2)\beta(2)\rangle$$

$$\varphi^+ = \cos\theta\varphi_1 + \sin\theta\varphi_2$$

$$\varphi^- = \cos\theta\varphi_1 - \sin\theta\varphi_2$$



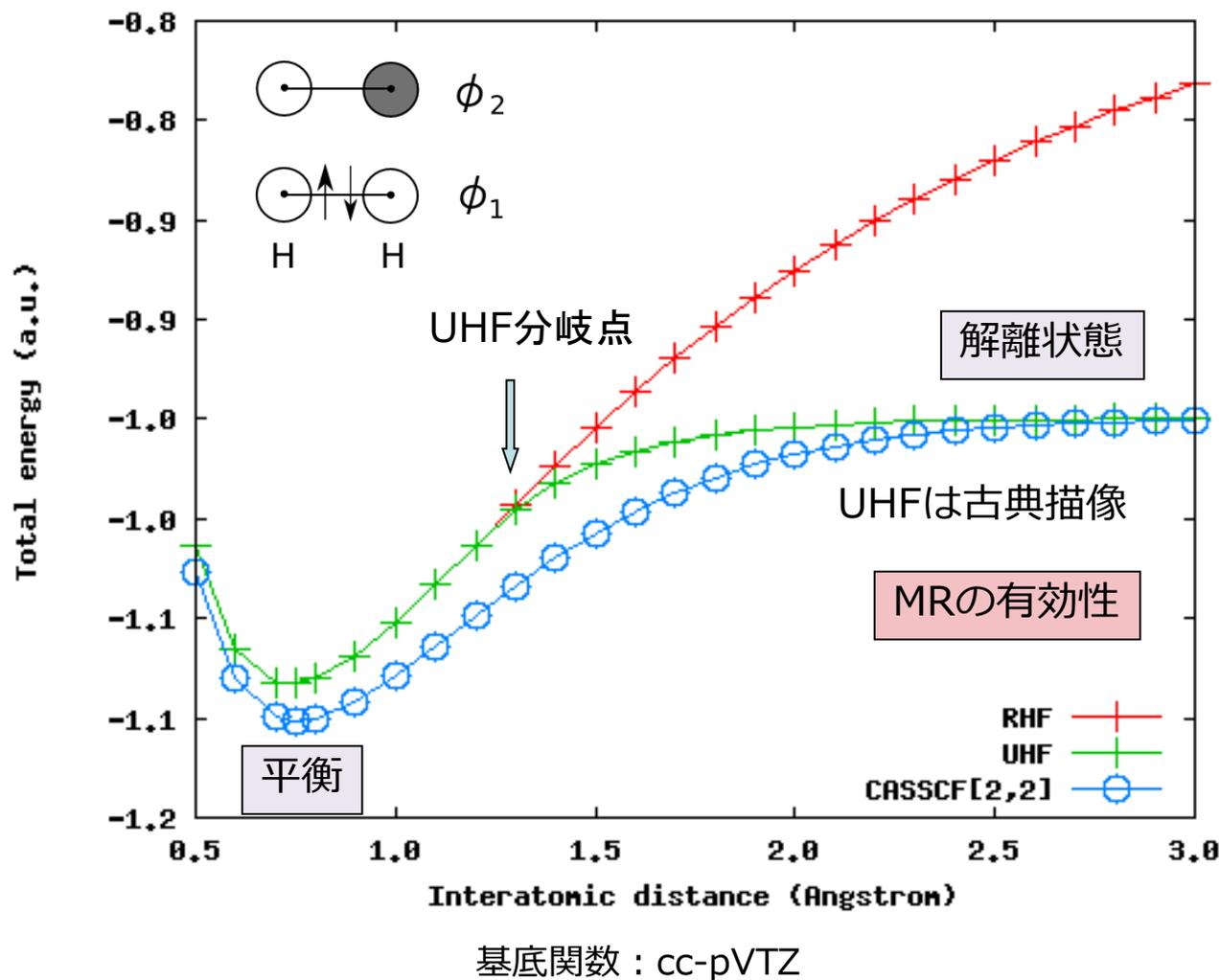
MR

CASSCF (Complete Active Space Self-consistent Field)

$$\Psi^{\text{CAS}} = C_1 |\varphi_1\bar{\varphi}_1\rangle + C_2 |\varphi_2\bar{\varphi}_2\rangle + \frac{C_3}{\sqrt{2}} (|\varphi_1\bar{\varphi}_2\rangle - |\bar{\varphi}_1\varphi_2\rangle)$$

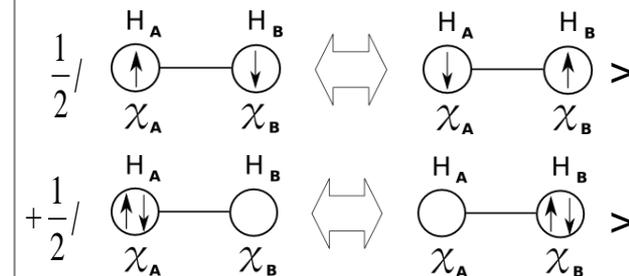


HF と CASSCF



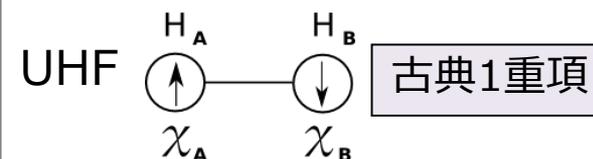
解離状態

RHFでは表現できない



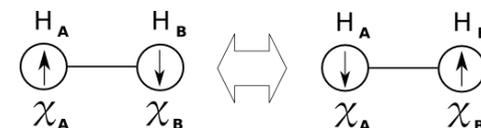
量子1重項

+ イオン項



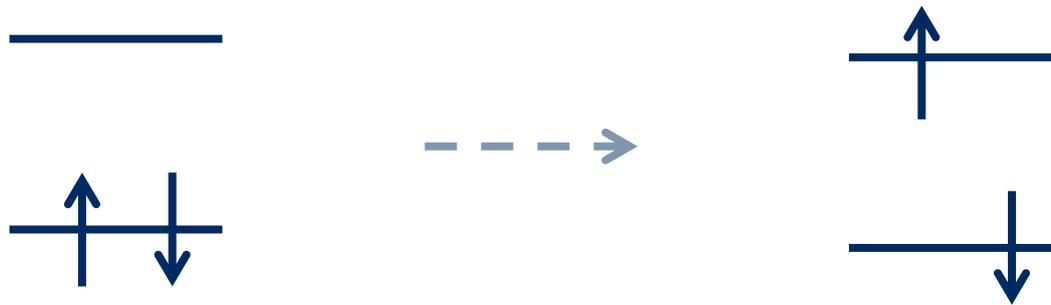
CASSCF

MRの有効性



量子1重項

なぜ？ CISやTDDFTではないのか



HF や KS-DFT
の基底状態

から見た

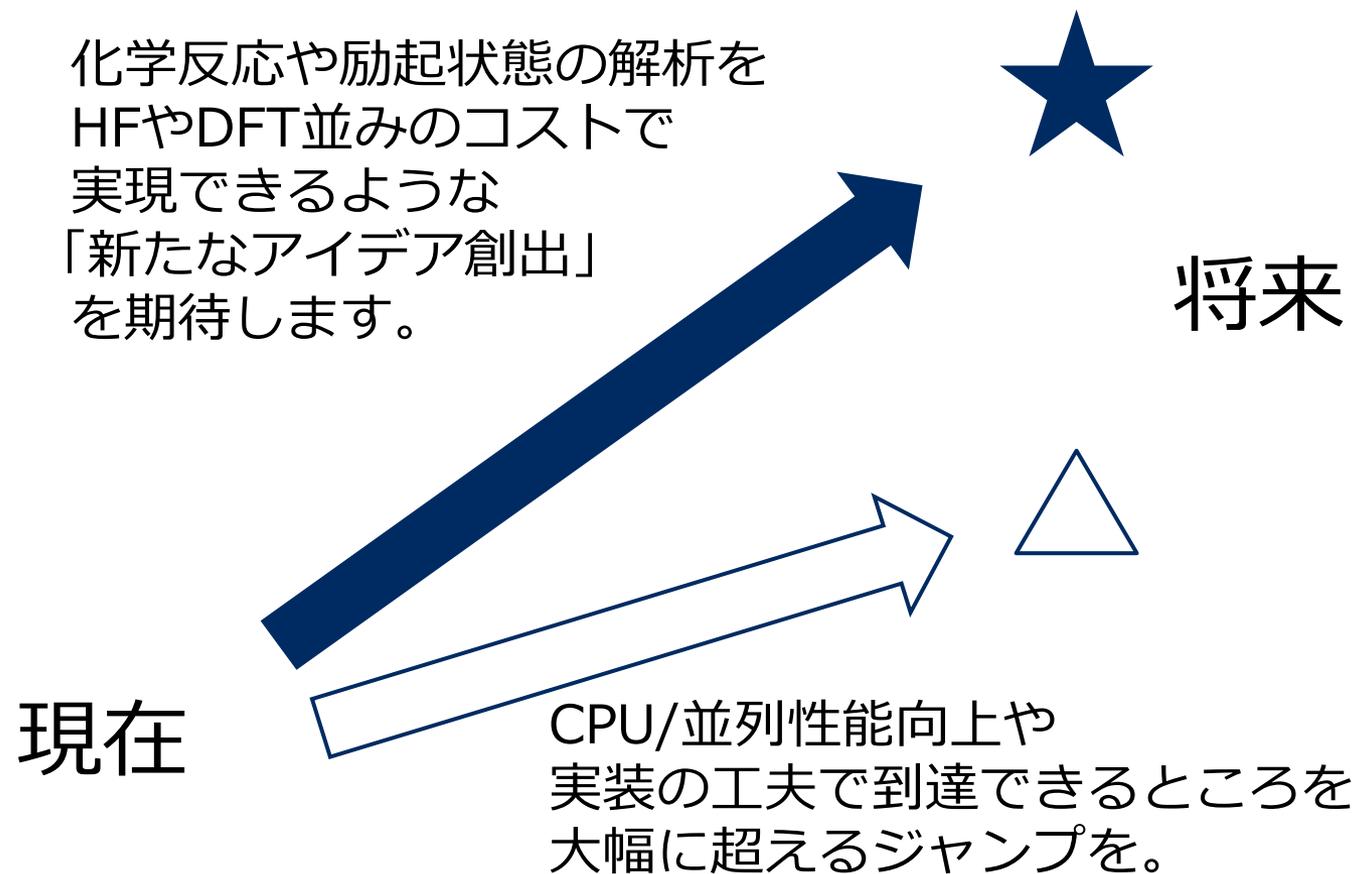
励起エネルギー

MO係数最適化での収束安定性 ⇒ 反復回数が増大

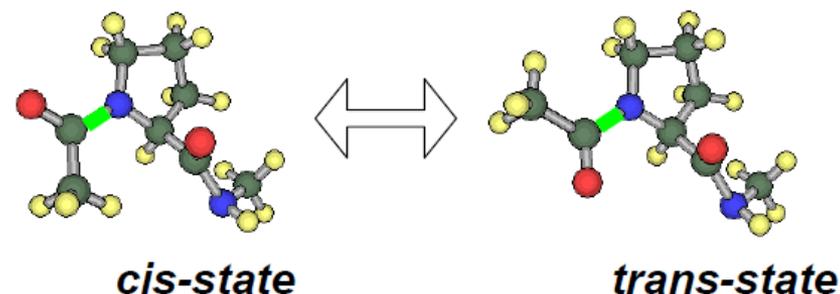
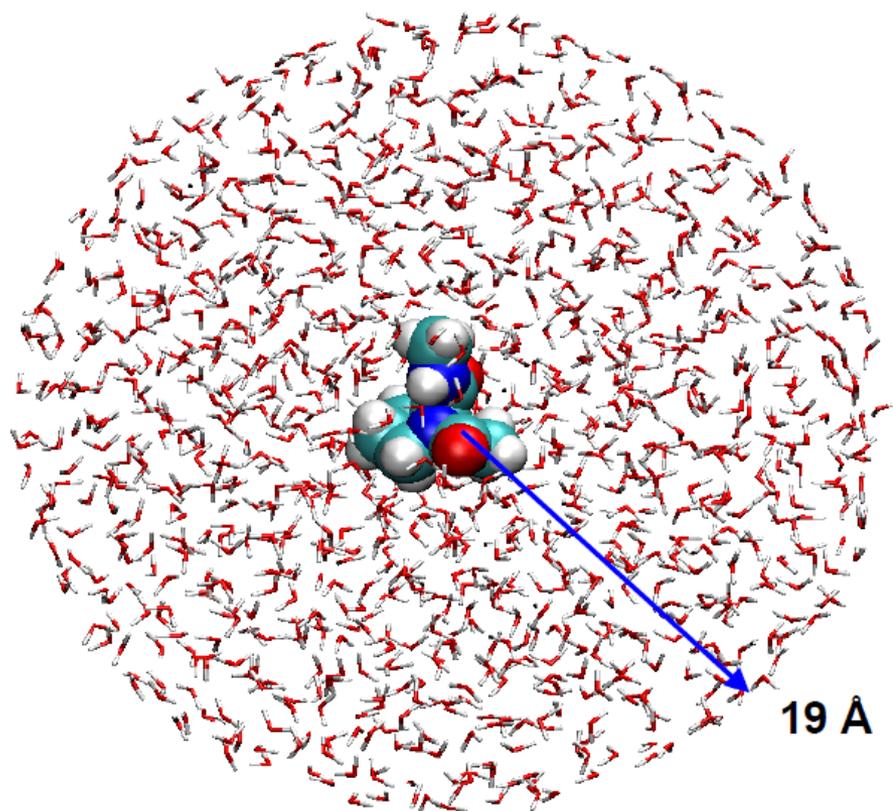
active空間の拡大 ⇒ 計算時間が増大



QM/MM分子動力学 : 計算時間の制約により、
計算できるものが限られる



研究事例の紹介① シス・トランス異性化



Simulation system

Ace-Pro-NMe in TIP3P 1354

Basis Set 4-31G

$\Delta t = 2$ fsec

Cutoff = 12 Å 300K

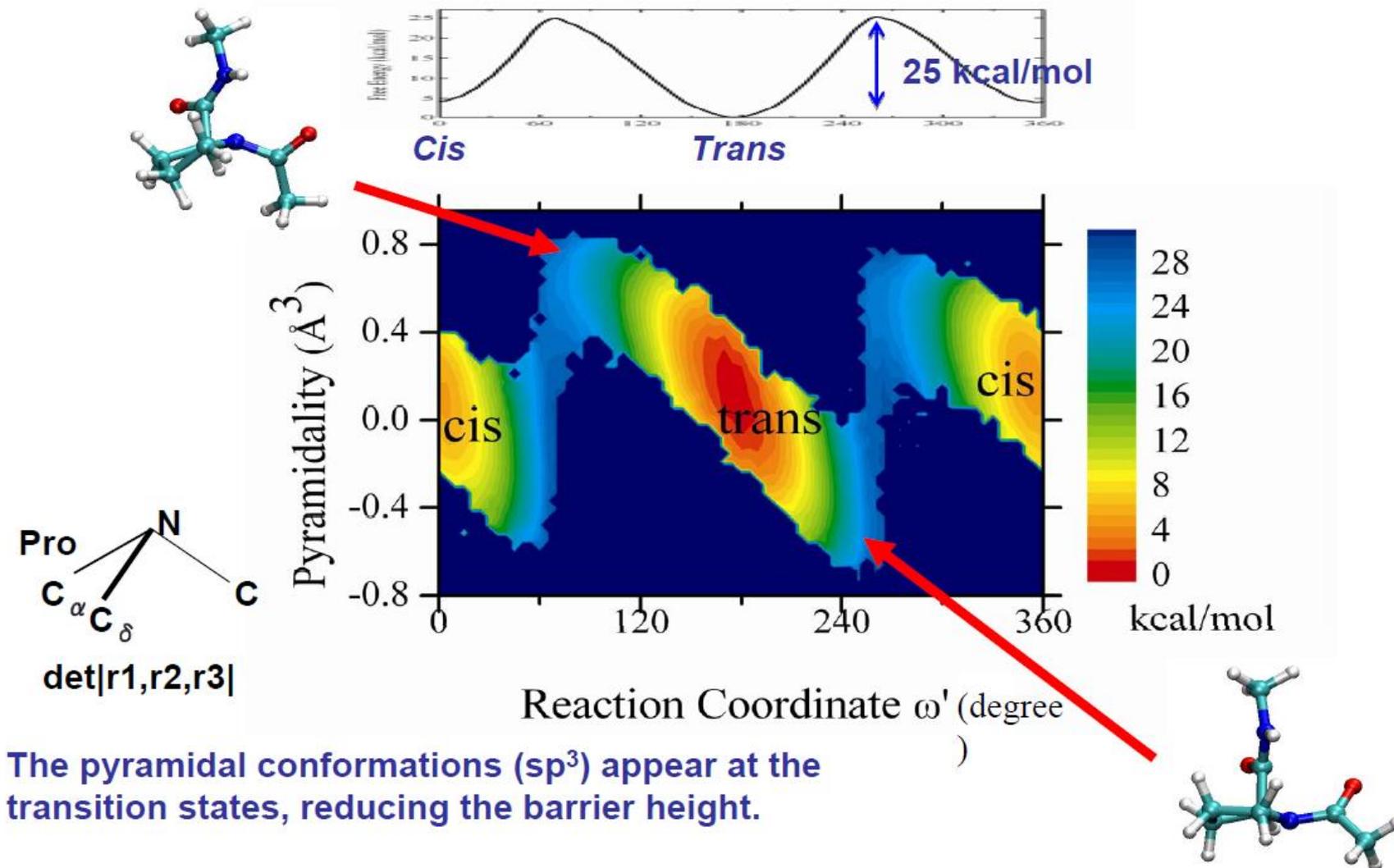
Umbrella sampling/

Umbrella integration methods

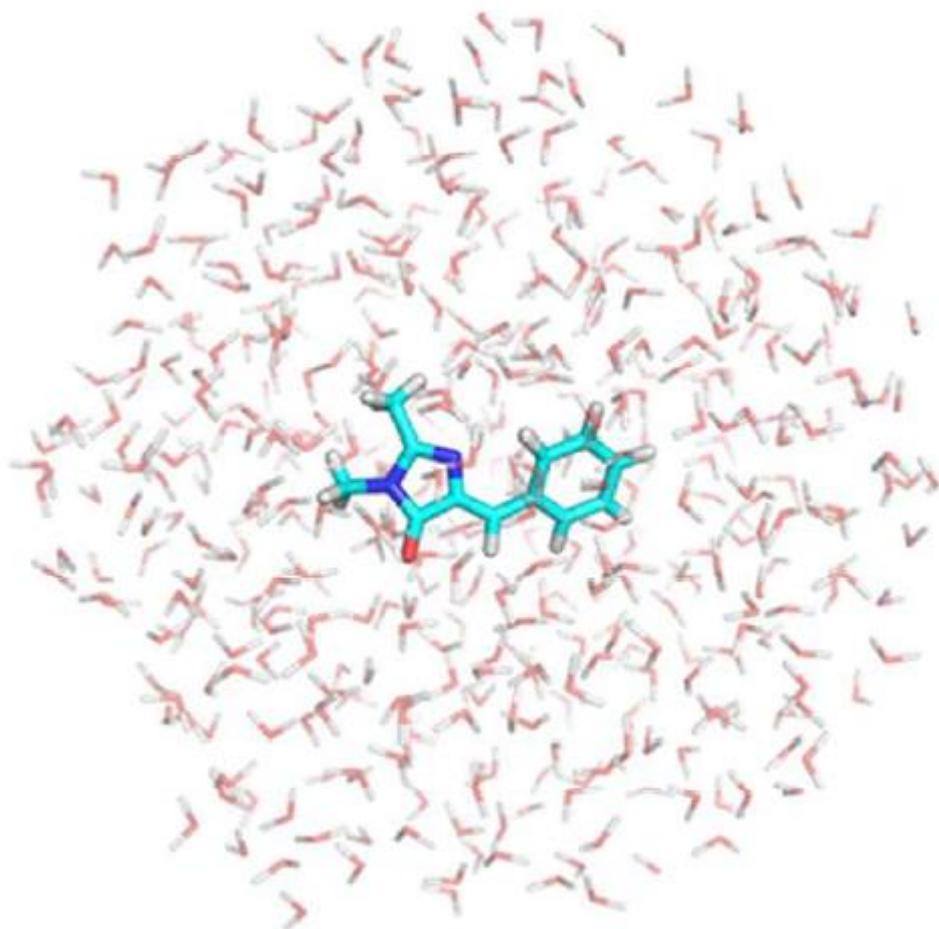
Computation with *Platypus* program

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

研究事例の紹介① シス・トランス異性化



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



蛍光タンパク質Sirius

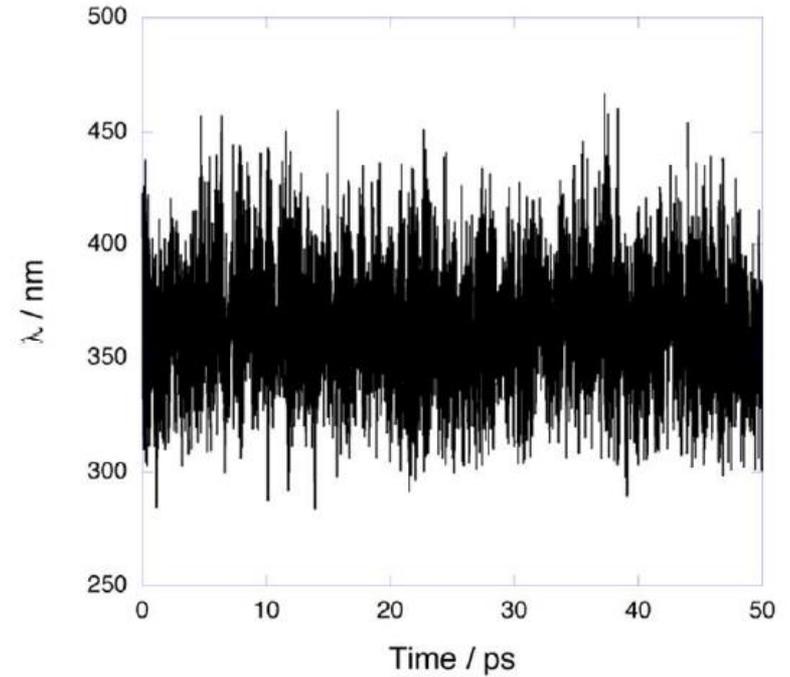
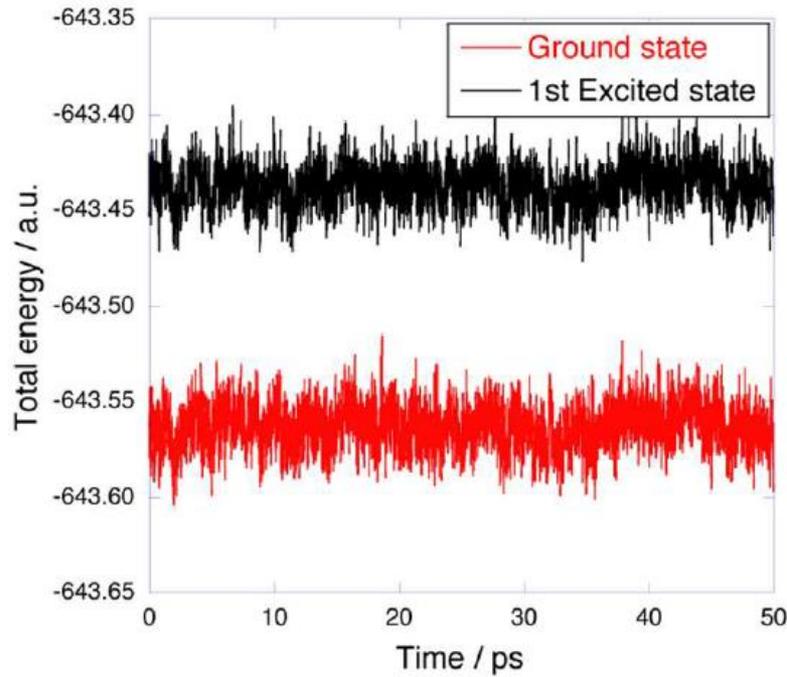
Matsuda et al.

Nature Methods **5**, 339-345 (2008).

QM : 27 原子

MM : 449 TIP3P

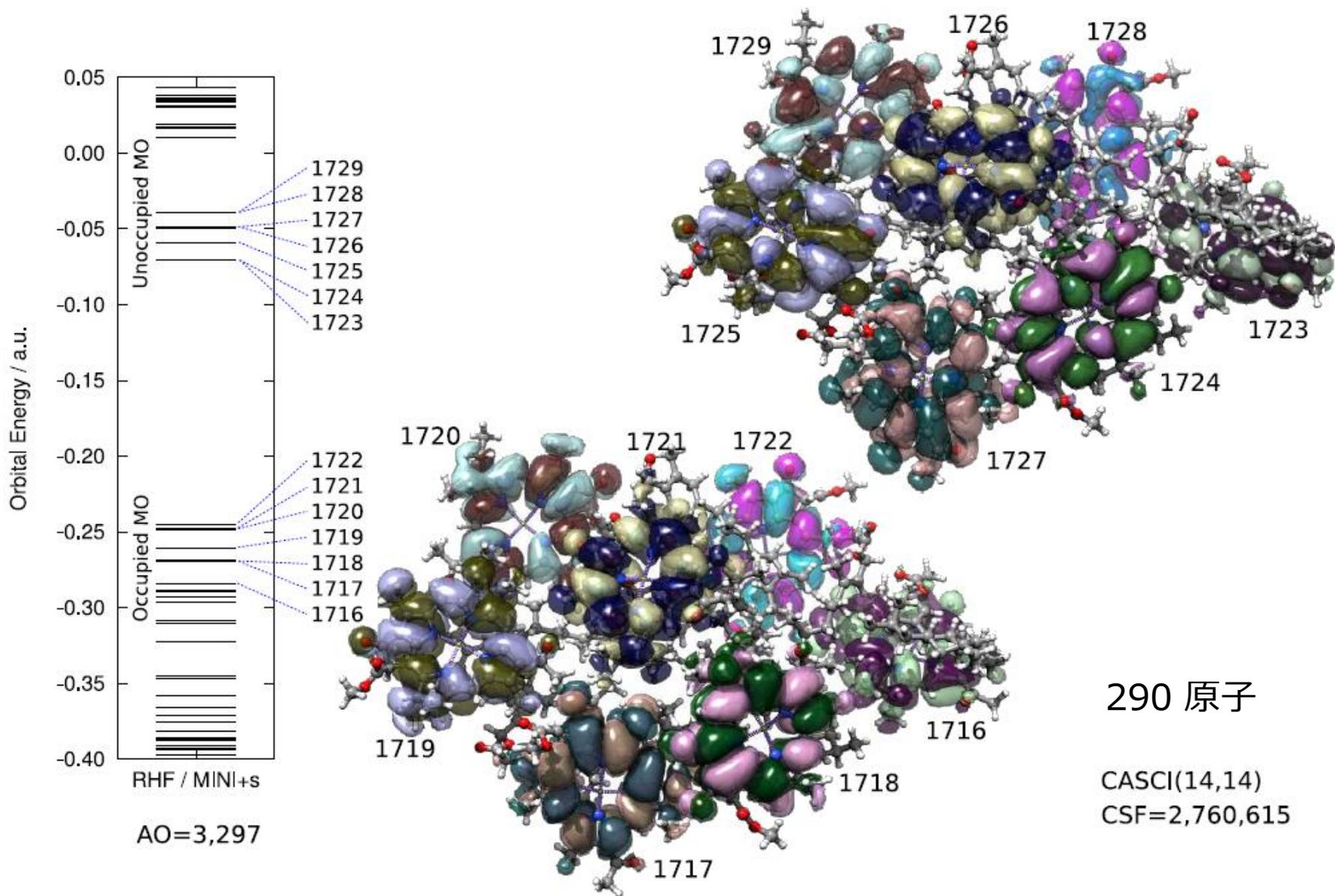
研究事例の紹介③ 励起状態のQM/MM-MD



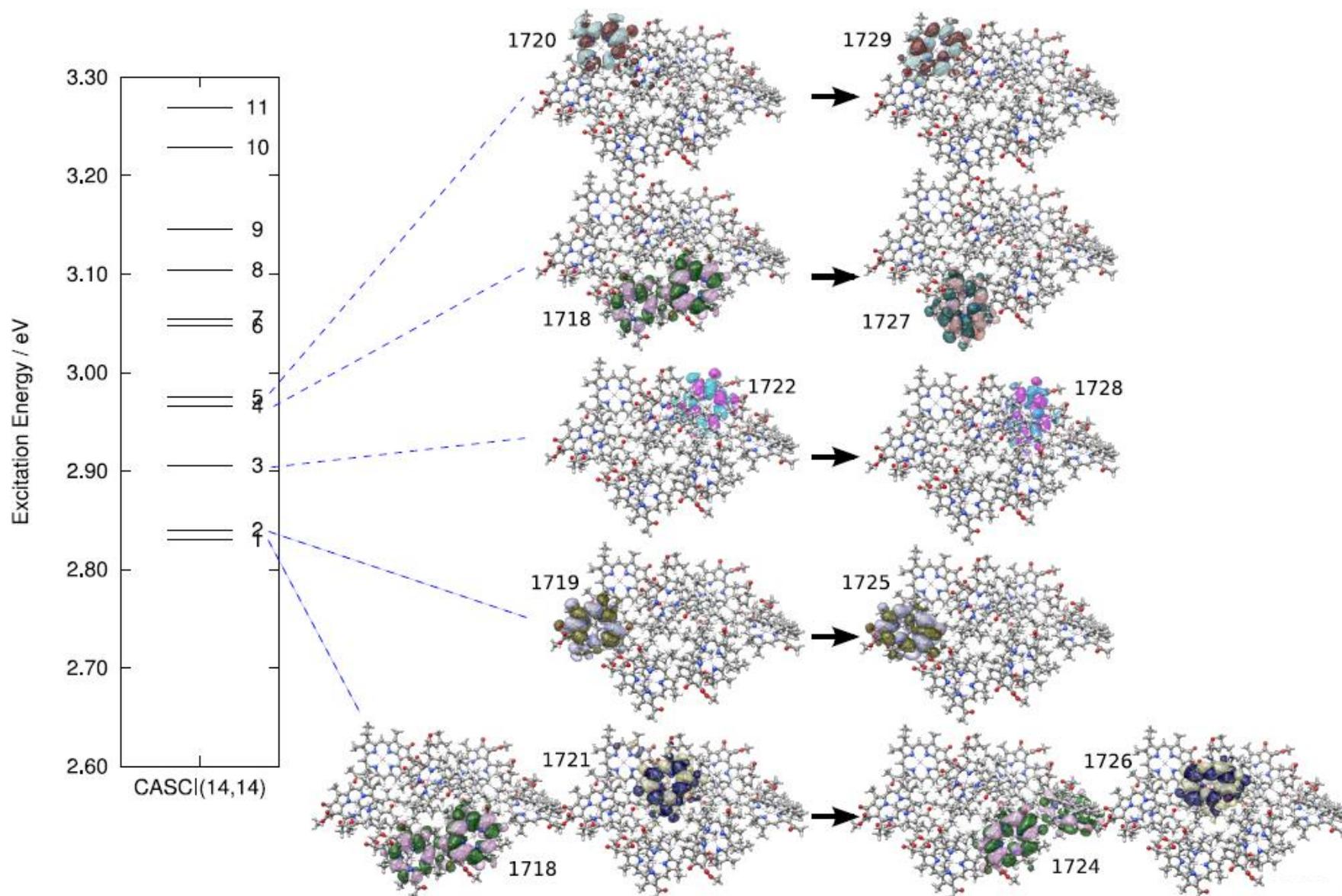
Emission energies

SA-CASSCF(8,8)/MINI-4

研究事例の紹介③ Platypus-QM 緑色硫黄細菌蛋白質



研究事例の紹介③ Platypus-QM 緑色硫黄細菌蛋白質

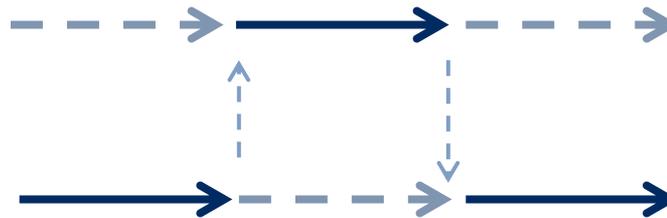


化学反応

遷移状態

励起状態

基底状態



Platypus-QMプログラム概要

工夫している点

SGA (Symmetric Group Approach, 対称群)を採用

疎行列の固有値問題 理研 今村俊幸 先生 (LOBPCG)

direct CI 法 :

$$\sigma_I = \sum_J^{N_{CSF}} H_{IJ} C'_J \quad \leftarrow \text{毎回 } H_{IJ} \text{ を再計算}$$

京コンピュータのメモリ容量の制限 (2GB/コア)で
CAS(16,16) の既約表現行列をメモリ上に格納すると
ギリギリであった。

Newton-Raphson (連立一次方程式)

$$\mathbf{H}\mathbf{x} = -\mathbf{g}$$

\mathbf{H} : Hessian 行列 (実、対称、密)

\mathbf{g} : gradient ベクトル

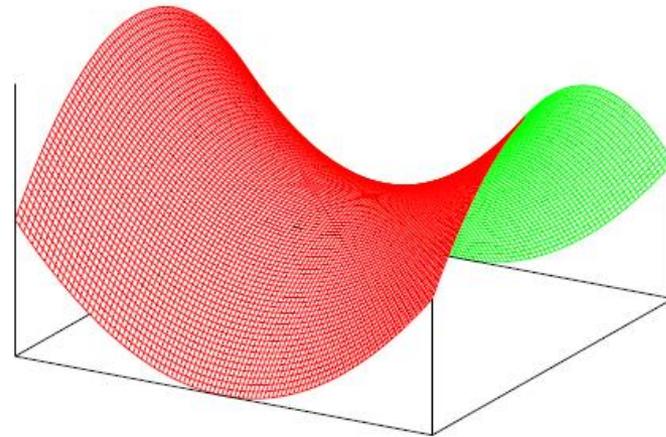
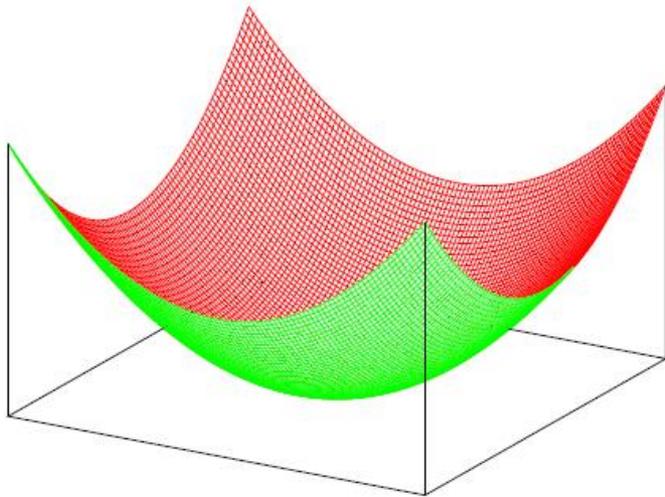
\mathbf{x} : MO 係数 $C_{\mu i}$ の変化量 (変換行列 $\mathbf{U} = e^{\mathbf{x}}$, $\mathbf{C}^{new} = \mathbf{C}\mathbf{U}$)

$$g_{pq} = \left. \frac{\partial E}{\partial x_{pq}} \right|_{x=0} \quad H_{pq,rs} = \left. \frac{\partial^2 E}{\partial x_{pq} \partial x_{rs}} \right|_{x=0}$$

- 2 次的な収束が期待できるが...
- Hessian 行列が正定値でないと...

Platypus-QM : CASSCFの収束安定性

- Hessian 行列が**正定値でない**
正定値であっても**固有値が非常に小さい**
↓
間違った方向、長すぎる、発散する...



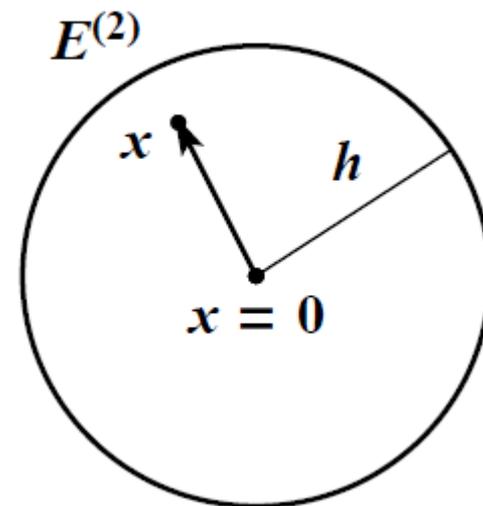
筑波大学 櫻井鉄也 先生 大野浩史さん (Shifted CG法)

■ 制約付き最小化問題

信頼半径 h : 2次までで十分近似できている範囲

$$\text{minimize } E^{(2)}(x) \quad \text{subject to } \|x\| \leq h$$

$$(H + \lambda I)x = -g$$



Platypus-QM : AO基底からMO基底への変換

MO基底

AO基底

$$(uv|xy) = \sum_{\mu} \sum_{\nu} \sum_{\rho} \sum_{\sigma} C_{\mu u} C_{\nu v} C_{\rho x} C_{\sigma y} (\mu\nu|\rho\sigma)$$

逐次処理版

並列処理版

step 1 : $(\mu\nu|\rho y) = \sum_{\sigma} C_{\sigma y} (\mu\nu|\rho\sigma)$

$(\tilde{\mu}\tilde{\nu}|\rho y) = \sum_{\sigma} C_{\sigma y} (\tilde{\mu}\tilde{\nu}|\rho\sigma)$

step 2 : $(\mu\nu|xy) = \sum_{\rho} C_{\rho x} (\mu\nu|\rho y)$

$(\tilde{\mu}\tilde{\nu}|xy) = \sum_{\rho} C_{\rho x} (\tilde{\mu}\tilde{\nu}|\rho y)$

step 3 : $(\mu v|xy) = \sum_{\nu} C_{\nu v} (\mu\nu|xy)$

$(\tilde{\mu}\tilde{v}|xy) = \sum_{\nu} C_{\nu v} (\tilde{\mu}\tilde{\nu}|xy)$

step 4 : $(uv|xy) = \sum_{\mu} C_{\mu u} (\mu v|xy)$

$(\tilde{u}\tilde{v}|xy) = \sum_{\mu} C_{\mu u} (\tilde{\mu}\tilde{v}|xy)$

step 5 : —————

$(uv|xy) = \sum_{n_{proc}} (\tilde{u}\tilde{v}|xy)$

- 各プロセッサは指定された $(\tilde{\mu}\tilde{\nu})$ の組 $(\tilde{\mu}\tilde{\nu}|\rho\sigma)$ を計算する。
- データ転送は step 5 のみ。

積分は漸化式で表現されている

⇒ 工夫 : RubyスクリプトでFortranコードを生成



- いくつか検証すれば、全てが(ほぼ)正しい
- コード書き直しは、スクリプト変更すれば容易
- いろいろな積分が、漸化式をスクリプトにできれば可能

Platypus-QM : 積分コードでの工夫

```
1  !$omp parallel ...
2  !$omp single
3    do i = i1, i2
4      ...
5    end do
6  !$omp end single nowait
7  ...
8  !$omp do schedule(dynamic)
9    do i = i1, i2
10     ...
11    end do
12  !$omp end do nowait
13  ...
14  !$omp end parallel
```

Platypus-QM : RHFエネルギー計算

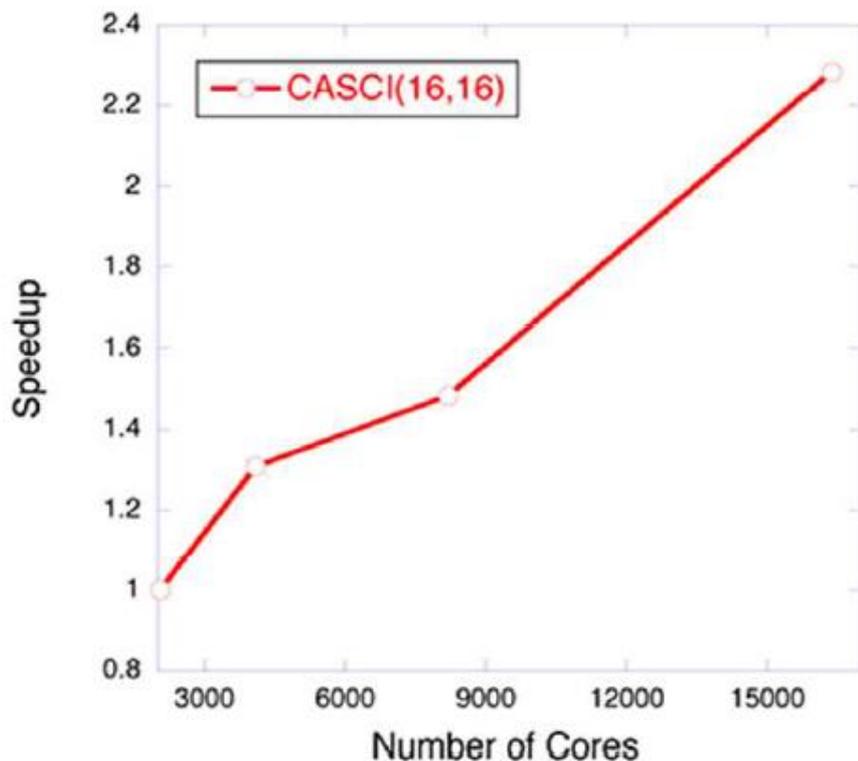
CPU数	スレッド数	コア数	経過時間 [sec]	Speedup
64	1	64	373.72683	1.0000
64	8	512	53.44365	6.9929

Platypus-QM : RHF力の計算

CPU数	スレッド数	コア数	経過時間 [sec]	Speedup
64	1	64	948.64316	1.0000
64	8	512	119.17338	7.9602

single でも nowait
dynamic で nowait

Platypus-QM : CAS(16,16)の実現



direct CI計算



クロロフィルダイマー
280原子、30904点電荷
CASI(16,16)/6-31G**

 **Orchestrating** a brighter world

NEC