

# ProteinDFの紹介

---

東京大学 生産技術研究所  
佐藤 文俊, 平野 敏行

2014/09/05  
ProteinDF利用講習会



# Agenda

- 背景
- 計算法
- タンパク質全電子計算達成への道筋





# 2013年 ノーベル化学賞



Photo © Harvard University

## Martin Karplus

**Martin Karplus**, U.S. and Austrian citizen. Born 1930 in Vienna, Austria. Ph.D. 1953 from California Institute of Technology, CA, USA. Professeur Conventionné, Université de Strasbourg, France and Theodore William Richards Professor of Chemistry, Emeritus, Harvard University, Cambridge, MA, USA.

▶ [More on Martin Karplus](#)



Photo: S. Fisch

## Michael Levitt

**Michael Levitt**, U.S., British and Israeli citizen. Born 1947 in Pretoria, South Africa. Ph.D. 1971 from University of Cambridge, UK. Robert W. and Vivian K. Cahill Professor in Cancer Research, Stanford University School of Medicine, Stanford, CA, USA.

▶ [Have a look at Michael Levitt's photo gallery](#)



Photo: Wikimedia Commons

## Arieh Warshel

**Arieh Warshel**, U.S. and Israeli citizen. Born 1940 in Kibbutz Sde-Nahum, Israel. Ph.D. 1969 from Weizmann Institute of Science, Rehovot, Israel. Distinguished Professor, University of Southern California, Los Angeles, CA, USA.

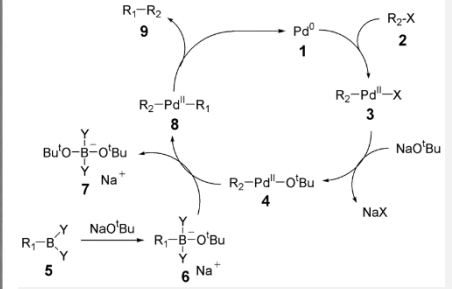
▶ [Interviews with Chemistry Laureate Arieh Warshel](#)



# ナノスケールの分子機械

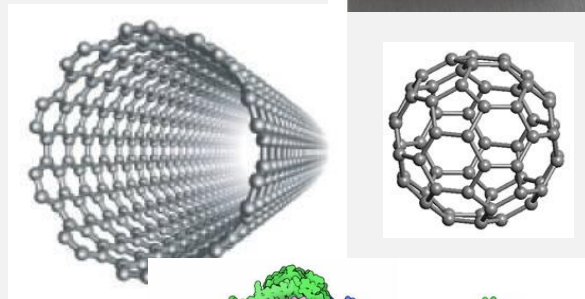
- 特異な反応を省エネルギーで
  - (密に詰まった)複雑かつユニークな電子状態を利用
- 大きく分けて3つのアプローチ
  - 重金属(レアメタル)

- **大量生産技術**
- **資源確保・汚染**



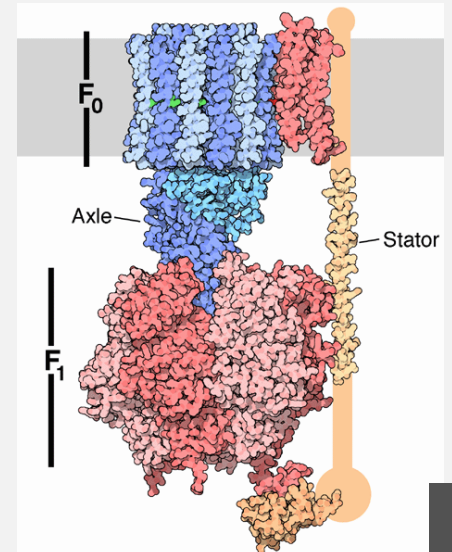
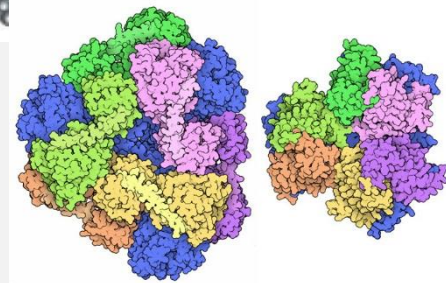
– ナノ構造

- **特徴ある構造**
- **生産・加工技術**



– タンパク質

- **タンパク質工学**
- **解析・設計技術**





# 定量性ある高精度計算の必要性

体内温度における標準自由エネルギー変化  
数~十数 kcal/mol

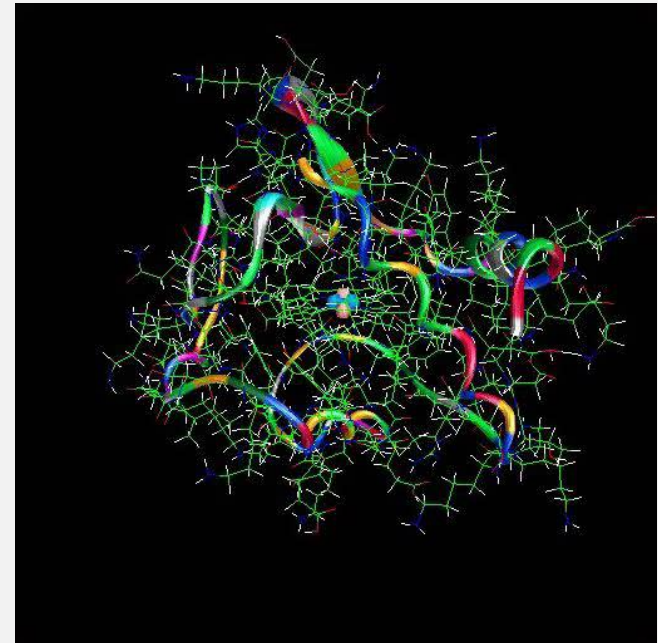
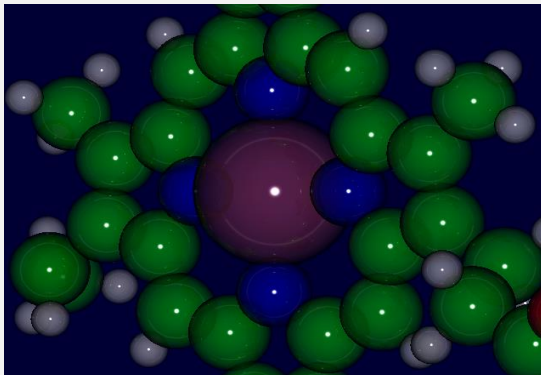
$$IP = E(N - 1) - E(N)$$

電子を抜くのに...

鉄イオン      ~1000 kcal/mol

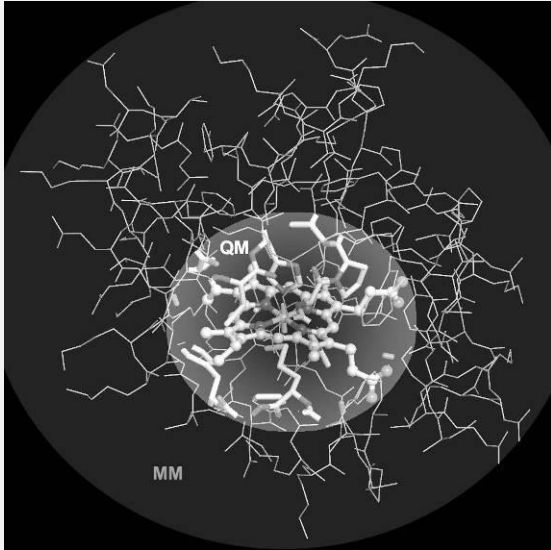
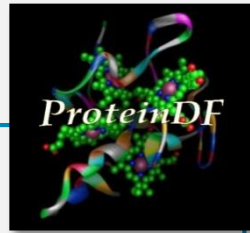
ヘム            ~100

ヘムタンパク質      <10



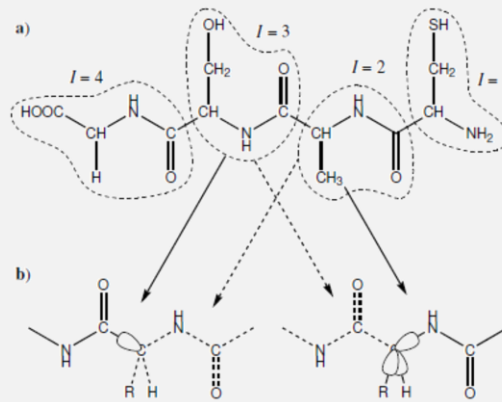


# 生体分子量子計算の挑戦



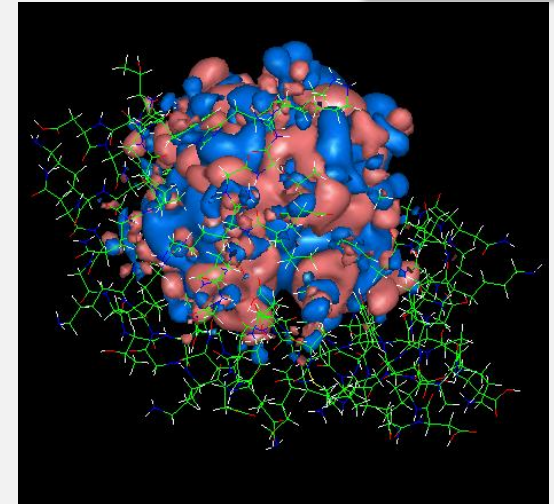
QM/MM, ONIOM:

- Gaussian



FMO:

- GAMESS
- ABINIT-MP



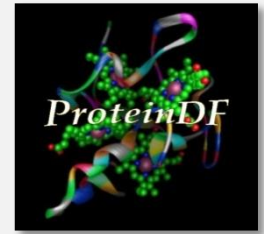
カノニカル分子軌道計算:

- ProteinDF

当方法



# ProteinDF の特徴



- 含金属タンパク質の全電子計算を対象
- 密度汎関数法に基づく  
カノニカル分子軌道計算プログラム
- **ガウス型基底関数**を使用
- **RI法、コレスキー分解法**により高速化
- 大規模SCF計算達成のための**QCLO法**
- オブジェクト指向言語**C++**で開発
- **MPI/OpenMP**ハイブリッド並列化対応





# ProteinDF の動作環境

- 必要なハード

- PCからスパコンまで幅広く使える

- PC & PCクラスタ
- 「京」

- 必要なソフト

- UNIX / Linux 環境

- C++コンパイラ

- GNU, intel, PGI, etc.

- 線形演算ライブラリ(LAPACK, ScaLAPACK)





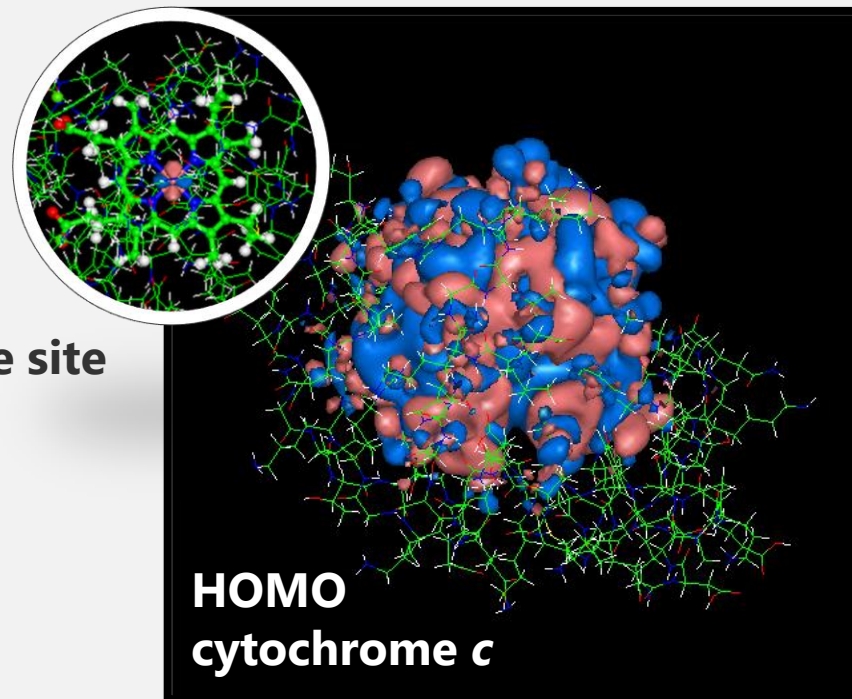
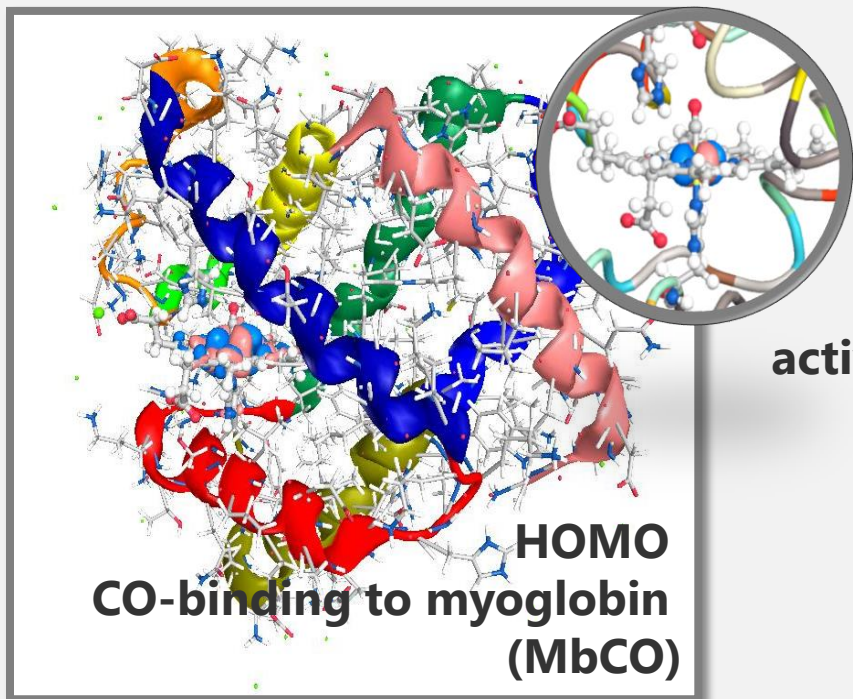
# ProteinDF 開発年表

- 1992 柏木研@九工大で開発はじまる
- 2000 金属タンパク質 Cytochrome c (9 600軌道)の全電子計算
- 2002 文科省「**戦略的基盤ソフトウェアの開発**」以降@東大生研
- 2006 Insulin 6量体 (26 790軌道)の全電子計算
- 2007 文科省「**革新的シミュレーションソフトウェアの研究開発**」
- 2010 文科省「**イノベーション基盤ソフトウェアの研究開発**」
- 2010 文科省「**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**」
- 2010 金属タンパク質 CO-Myoglobin の全電子計算
- 2010 金属タンパク質 Cytochrome  $c_3$  の全電子計算
- 2013 交換相関項の新グリッドフリー計算法の開発
- 2013 第3世代アルゴリズム法の開発
- 2014 GPL版リリース



# 全電子計算例

## - ヘムタンパク質の分子軌道 -



active site

Mb : Fe(III)で失活

HOMOの成分

94.7% : Feのd軌道

99.9% : Heme+近位His

Cyt.c : Fe(II) ⇌ Fe(III)で電子移動

HOMOの成分

85.2% : Feのd軌道

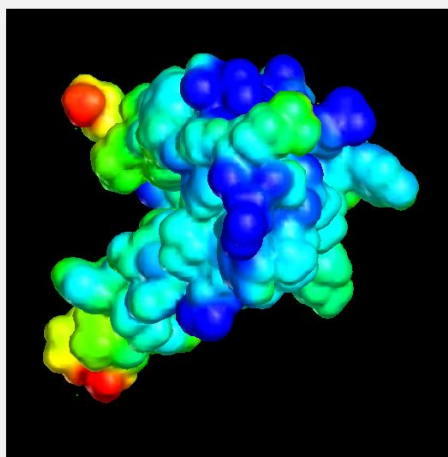
- 主成分の同じHOMOでも分子軌道の広がりには違いが見られる  
- 機能の違いと相関(?)



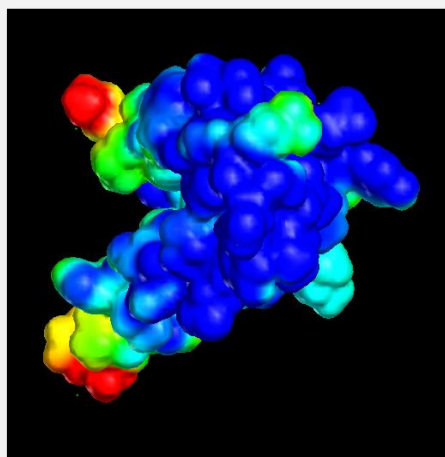
# 全電子計算例

## - Insulinの静電ポテンシャル -

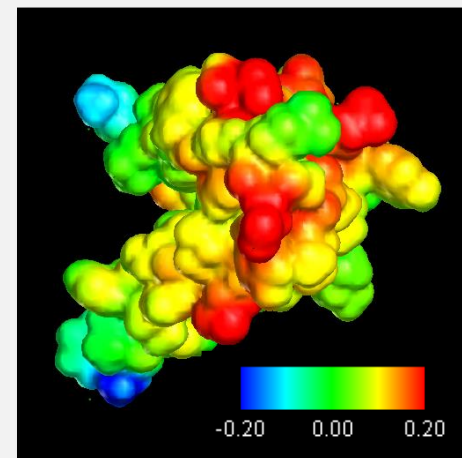
- タンパク質の相互作用に関わる重要な物理量
- 全電子計算から高精度な静電ポテンシャルが算出可能



ProteinDFによる  
**全電子計算結果**



**古典計算結果**  
(各原子の電荷は  
RESP電荷を使用)



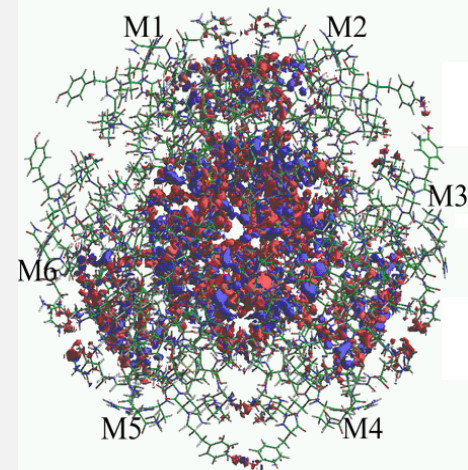
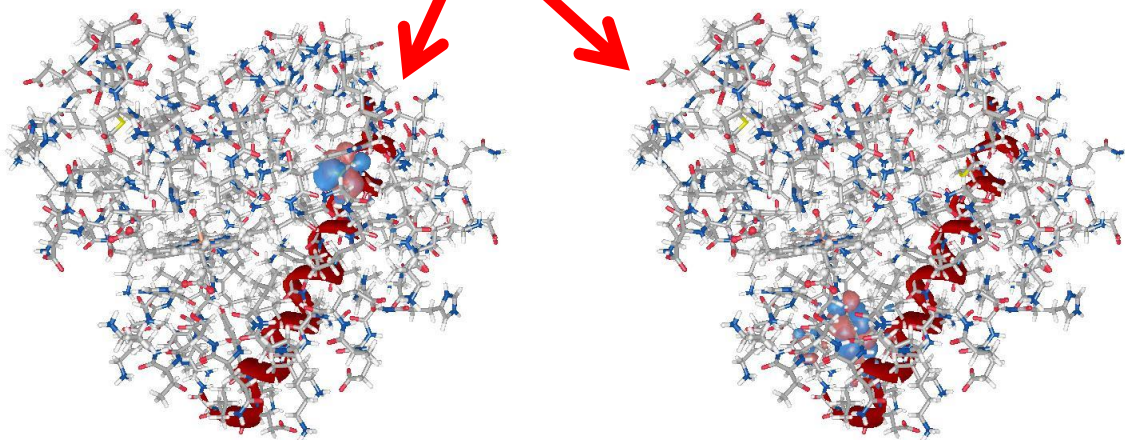
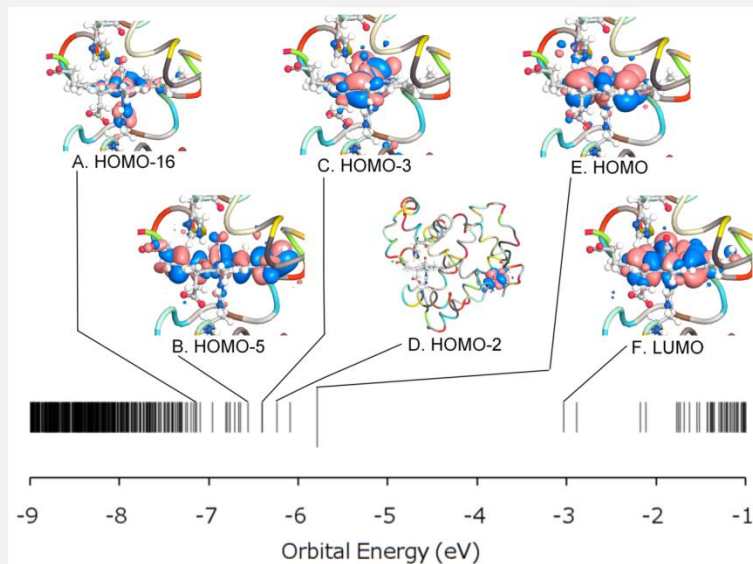
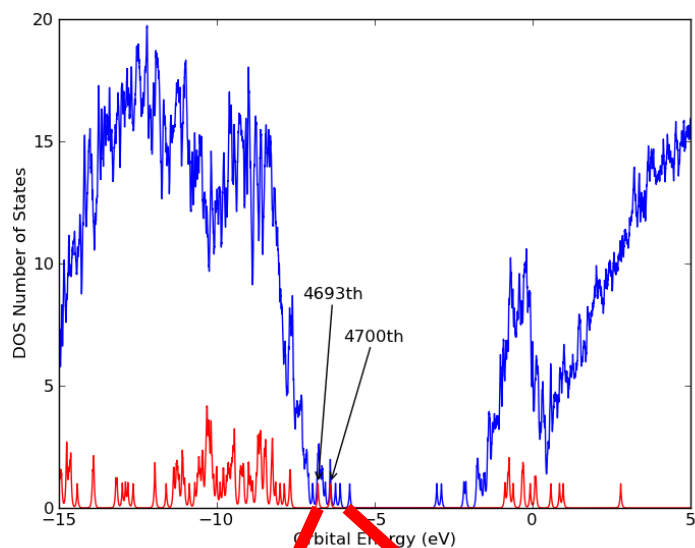
全電子計算と  
古典計算の**差**

- 主流の古典計算では、**misleading**を招く恐れがある



# 分子軌道、電子密度、状態密度解析

- タンパク質の階層構造と電子状態との間で、詳細かつ直接的な解析





# 計算法



# Kohn-Sham-Roothaan 方程式

$$FC = SC\epsilon$$

---

$$F_{pq} = h_{pq} + J_{pq} + \alpha K_{pq} + \beta \mu_{pq}$$

$$J_{pq} = \sum_{pq} P_{pq} \langle pq | rs \rangle$$

$$K_{pq} = \sum_{pq} P_{pq} \langle pr | qs \rangle$$

$$\mu_{pq} = \int g_p(\mathbf{r}) \mu_{XC}[\rho(\mathbf{r})] g_q(\mathbf{r}) d\mathbf{r}$$

$$\langle pq | rs \rangle = \int g_p(\mathbf{r}) g_q(\mathbf{r}) \frac{1}{|\mathbf{r} - \mathbf{r}'|} g_r(\mathbf{r}) g_s(\mathbf{r}) d\mathbf{r} d\mathbf{r}'$$



# タンパク質分子サイズと計算資源

- サイズに関するルール

残基数 : 原子数 : 電子数 : 軌道数(行列次元)  
≈ 100 : 2000 : 7000 : 10000

100残基タンパク質なら	1億要素	3GB/3行列
1000残基タンパク質なら	100億要素	300GB/3行列

- 律速計算

- 分子積分計算 : クーロン項、Fock交換項
- 数値積分計算 : 交換相関項
- 行列演算 : 行列積、対角化

- 並列計算

- 巨大行列を分散保持、律速計算毎にタスクを分割
- これまでは99.9x% → 数1000~10000並列規模





# 3G DFT 法の技術

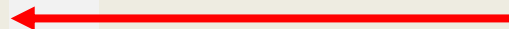
## コレスキー分解

- 4中心分子積分をコレスキーベクトルに分解
- CDAM法(Y. Okiyama, et. al, *CPL*, **490**, 84, 2010)でsuper-matrixを小さく
- pivot法でrank数を小さく
- コレスキーベクトルが求めれば、以降の計算は行列演算のみ

クーロン項、Fock交換項、**交換相関項**

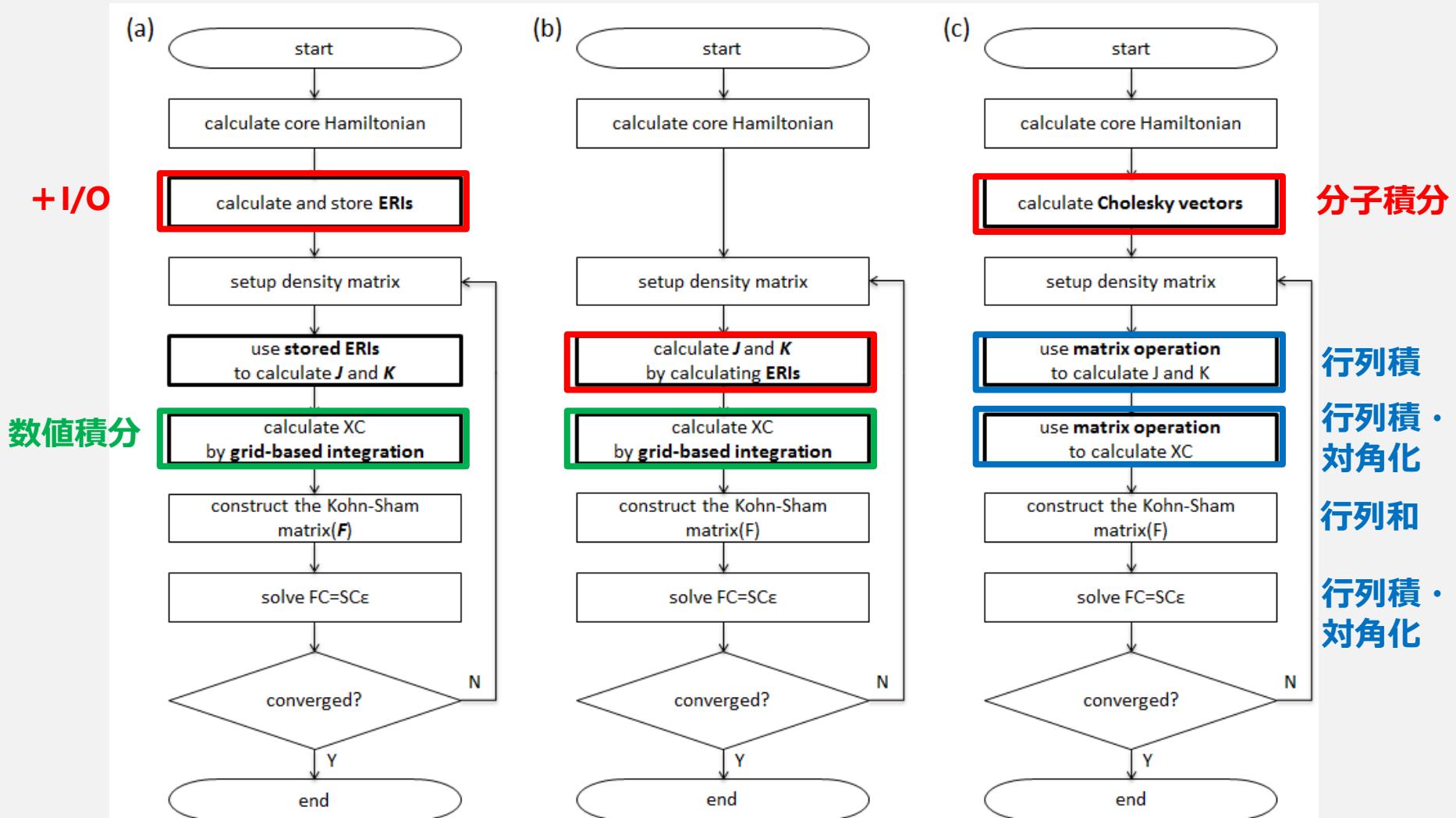
## dual-basis grid-free法

- 4中心重なり積分を用いて密度行列表現(MRD)から解析的に交換相関項を計算
- 超大規模・大量計算で懸案となるノイズ・数値不安定性を払拭、力も解析計算
- MRD専用の基底関数セットの利用でシミュレーション精度を向上





# 第3世代密度汎関数法アルゴリズム



ファイル法

ダイレクト法(現方法)

第3世代法



# 3G DFT 法の特徴

- 分子軌道計算特有の分子積分計算はSCF繰り返し計算前に1度だけ
- SCF繰り返し計算中は行列演算のみ
  - 計算機に最適化された線形演算ライブラリが利用可能
  - 超並列化(タスク最適均等化)が容易

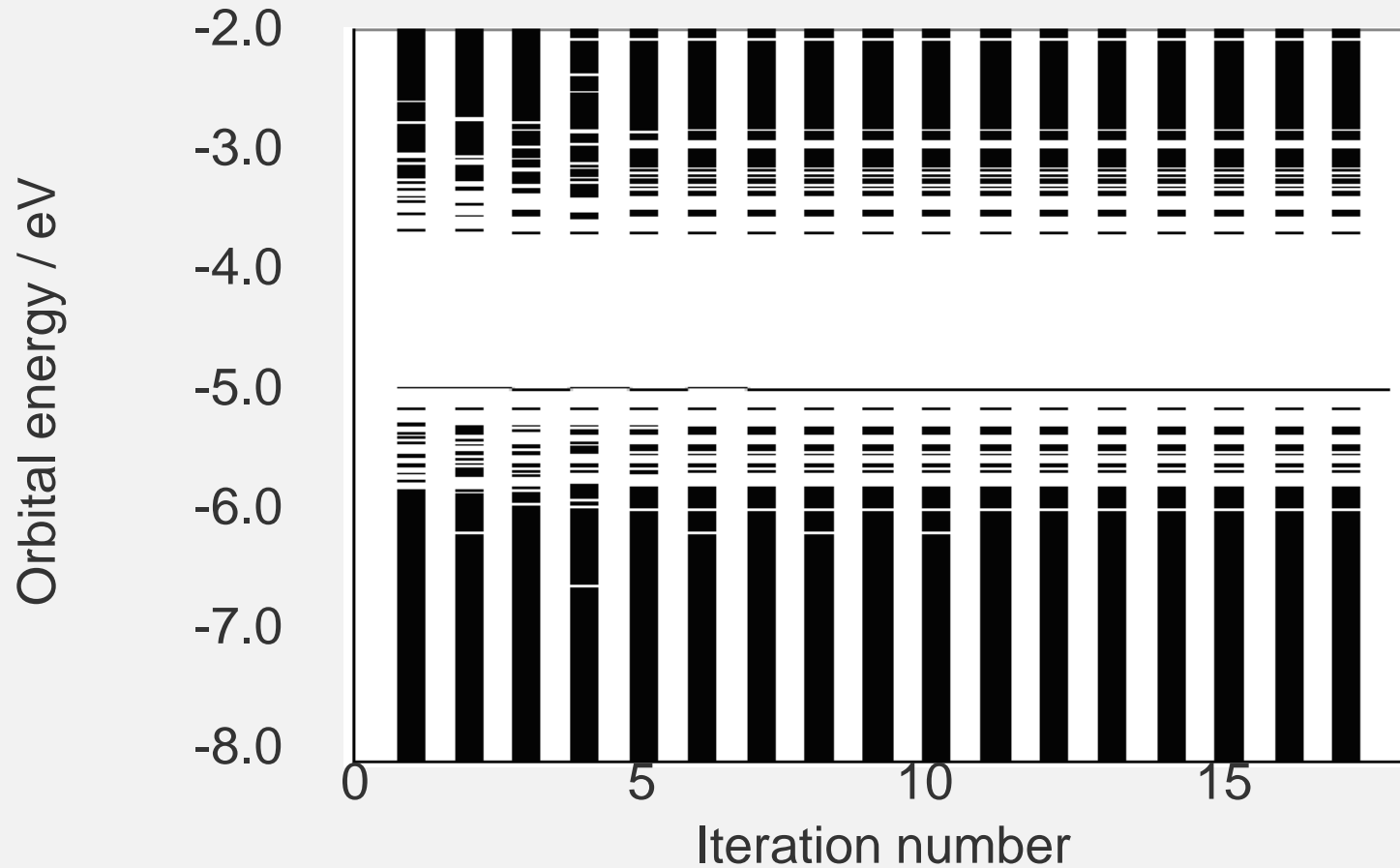


# タンパク質全電子計算達成への 道筋



# SCF の収束

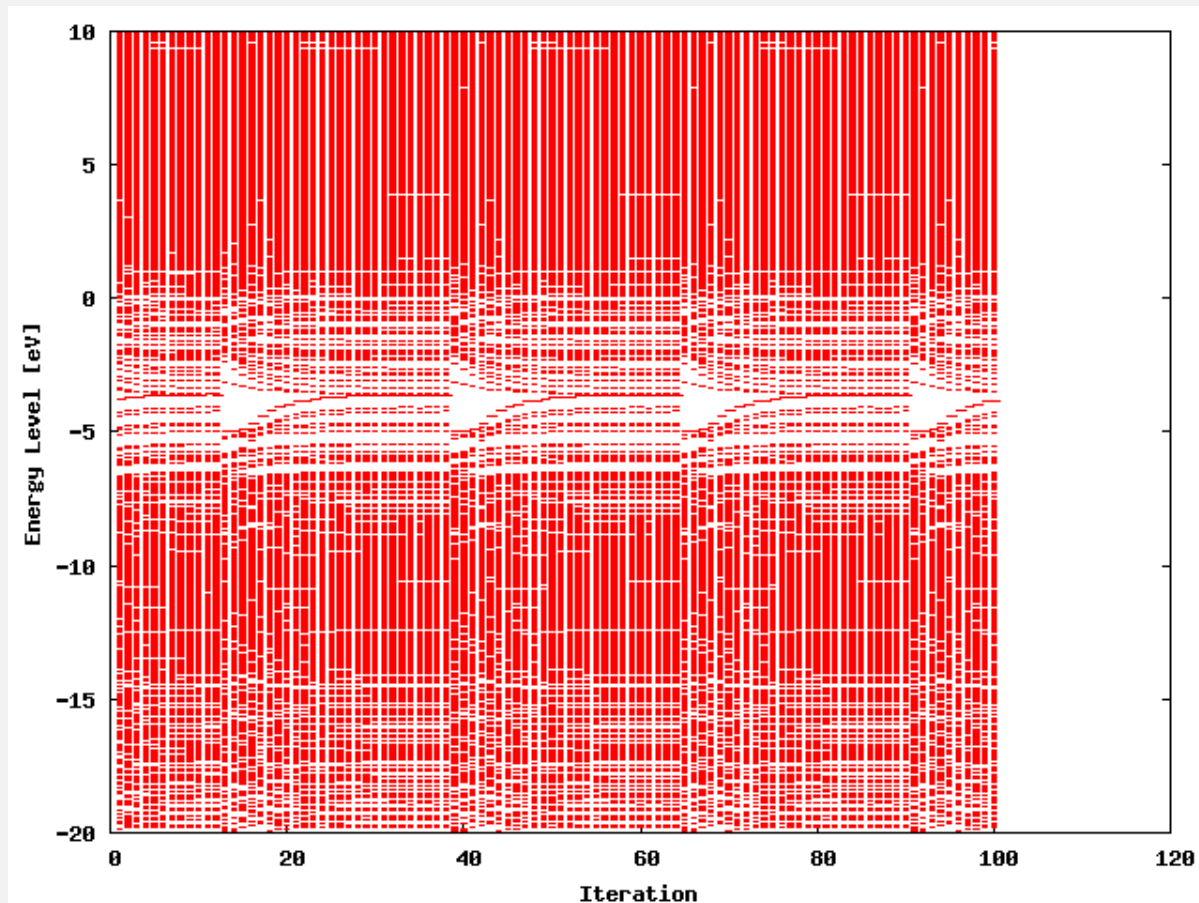
- 理想的な収束





# SCF 計算失敗例

- 被占軌道と空軌道の入れ替わり
- ギャップが空いていない





# 計算シナリオと収束法

- **タンパク質全電子計算達成のためのシナリオ**
  - タンパク質計算をより安全に達成する技術
  - 計算シナリオ、QCLO法
- **SCF計算収束法 (本日は省略)**
  - SCF計算を安全・安定に収束へと導く技術
    - 軌道の重なり, プロジェクション, レベルシフト, シンプルミキシング
    - 特にタンパク質のような巨大な系において重要
  - SCF計算収束を加速する技術
    - Anderson, DIIS
    - 本来、正しく収束するならば、SCF計算は何回転してもかまわない  
→ ...が、タンパク質の量子計算はコスト高
    - 収束加速はそのまま計算時間の短縮に貢献



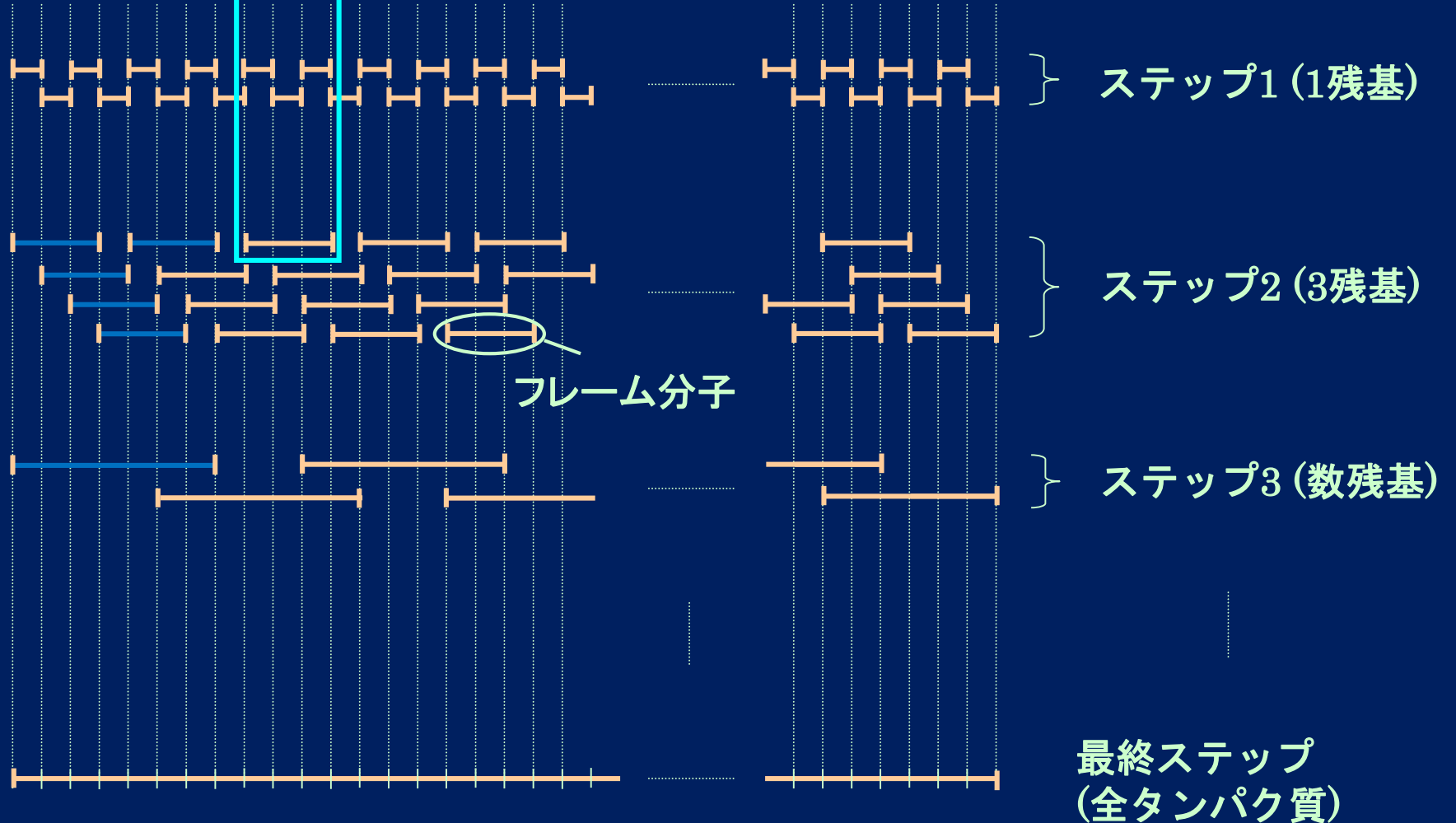
# タンパク質全電子計算のシナリオ

GRGQFPREHTVLAWQNTTY

...

FPRETQH

アミノ酸シーケンス

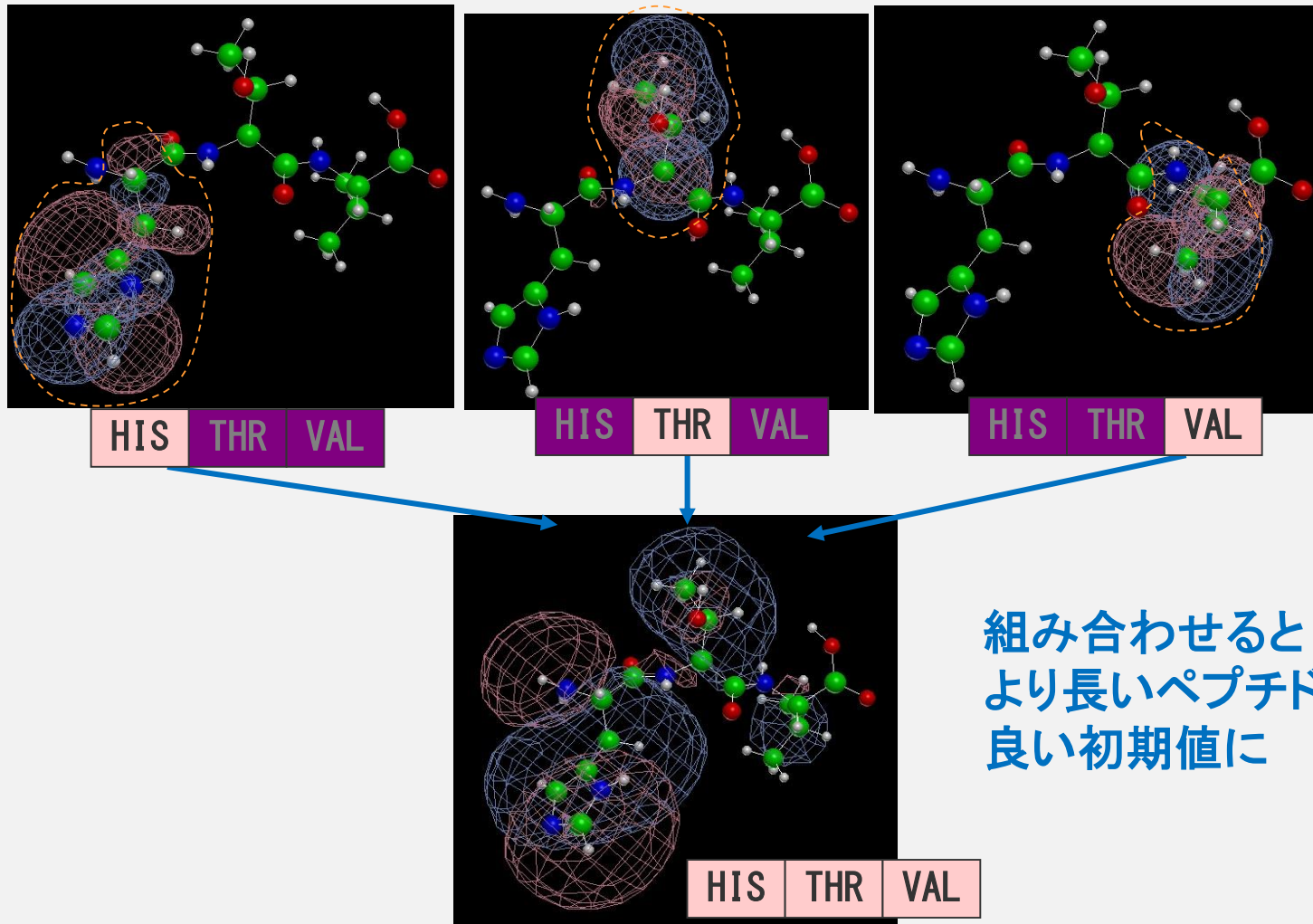






# QCLO (Quasi-Canonical Localized Orbital)

- 任意の空間で局在化し、その空間内ではカノニカル軌道





# 最後に

- 本日の演習
  - FOCUSスパコンを使用したProteinDFとQCLOの演習
- ご注意
  - 本日はあまり大きな計算はできません
  - QCLOはプリミティブ版で演習を行います
  - 今後、逐次機能をアップデートして行きます
    - スノーボールエフェクトも期待しております