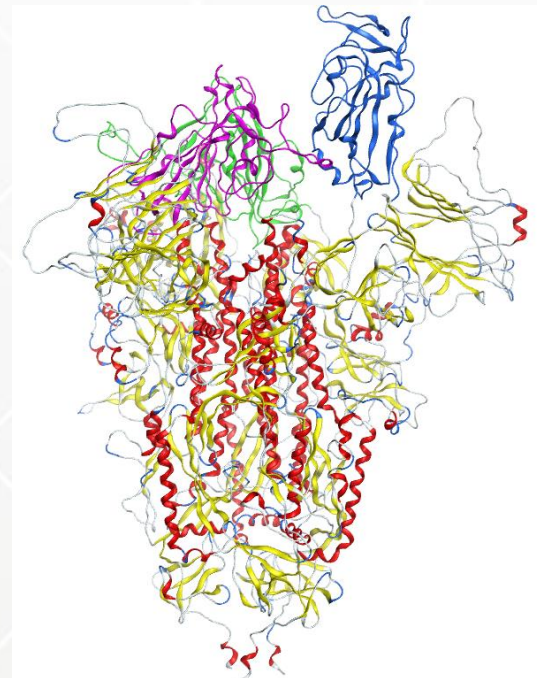
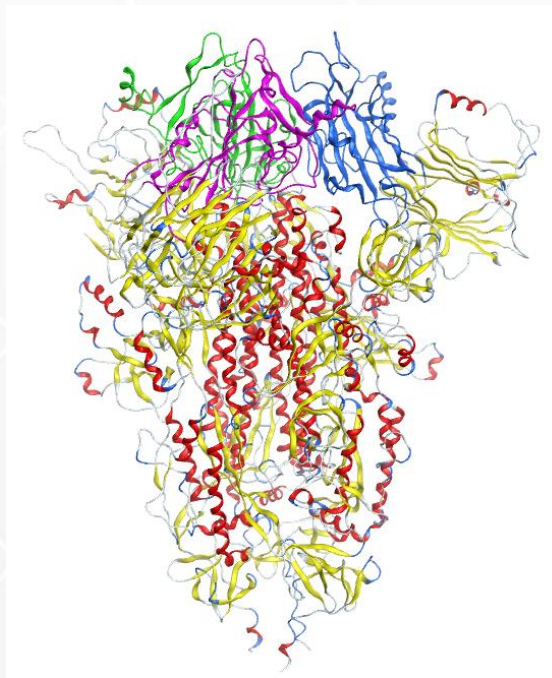
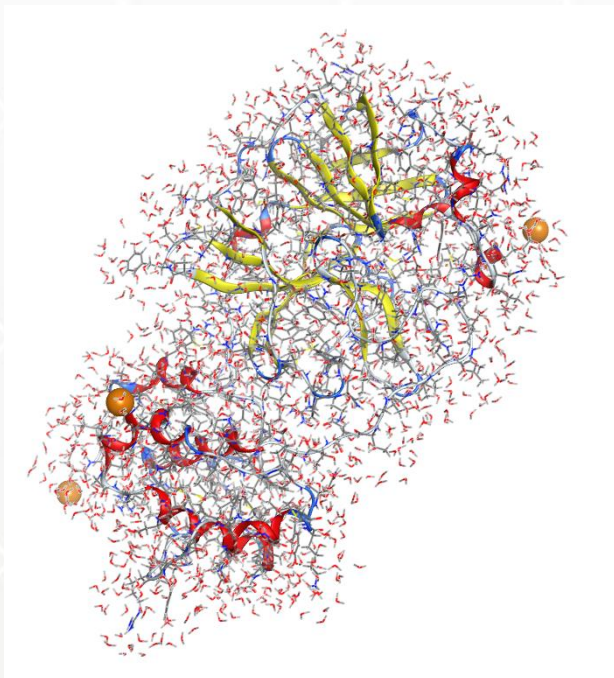


新型コロナウイルス関連タンパク質に対するフラグメント分子軌道計算

望月祐志* (立教大学), 田中成典 (神戸大学), 福澤薫 (星薬科大学)

1. 望月：FMO計算の概要、ABINIT-MPプログラム（富岳の計算力）
2. 福澤：メインプロテアーゼ／RNAポリメラーゼとリガンド（構造揺らぎの重要性）
3. 田中：スパイクタンパク質（ユニット間相互作用の差異）
4. 望月：まとめと今後
5. 補助資料

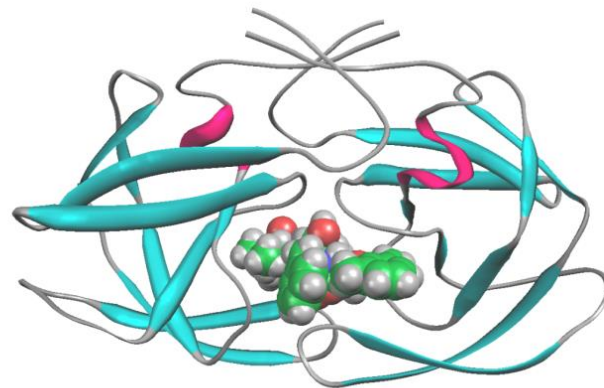


◇巨大分子系

生体高分子や凝集系では一般的

⇒ タンパク質、DNA（水和状態）

数千～数万原子、数千～数十万軌道



【HIVプロテアーゼとロピナビル】

◇分割&統合系の量子論的手法の一つ

北浦らが20年程前に2体展開で提案

⇒ フラグメントとその対で系のエネルギーを評価（FMO2）

⇒ 環境静電ポテンシャル（ESP）、直接結合切断（BDA）

⇒ 階層的な並列処理（フラグメントリスト&内部処理）

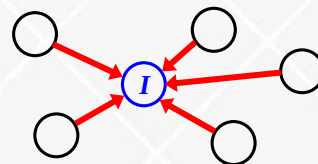
⇒ 2次摂動(MP2)などの相関補正も直截（分散力の記述に必須）

フラグメント間の相互作用エネルギー（IFIE）

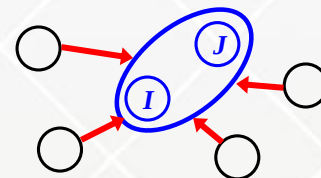
⇒ 生体分子の解析ツール

⇒ 特に創薬系には有用

⇒ 成分分析も可能（PIEDA）



モノマー（アミノ酸単位など）



ダイマー

- FMO-HF法の全エネルギー E の計算式

$$E = \sum_{I>J} E_{IJ} - (N_f - 2) \sum_I E_I$$

N_f : フラグメント(モノマー)の個数

E_I : モノマーのエネルギー, E_{IJ} : ダイマーのエネルギー

- FMO-HF法の全エネルギー E を書き換えると、全エネルギーを**フラグメント間相互作用エネルギー $\Delta\tilde{E}_{IJ}$** 、モノマーのエネルギー E_I から周囲のフラグメントとの静電相互作用エネルギーを除いたエネルギー E'_I の和で表せる。
- MP2などの**相関エネルギー補正**は加成的に行えばよい。

$$E = \sum_{I>J} \Delta\tilde{E}_{IJ} + \sum_I E'_I$$

IFIE解析 ⇒ 対象系の内部状態の理解
(Inter-Fragment Interaction Energy)

- さらに成分に分けることもよく行われる(**PIEDA**)。これにより水素結合による安定化、分散力による安定化が識別出来る。

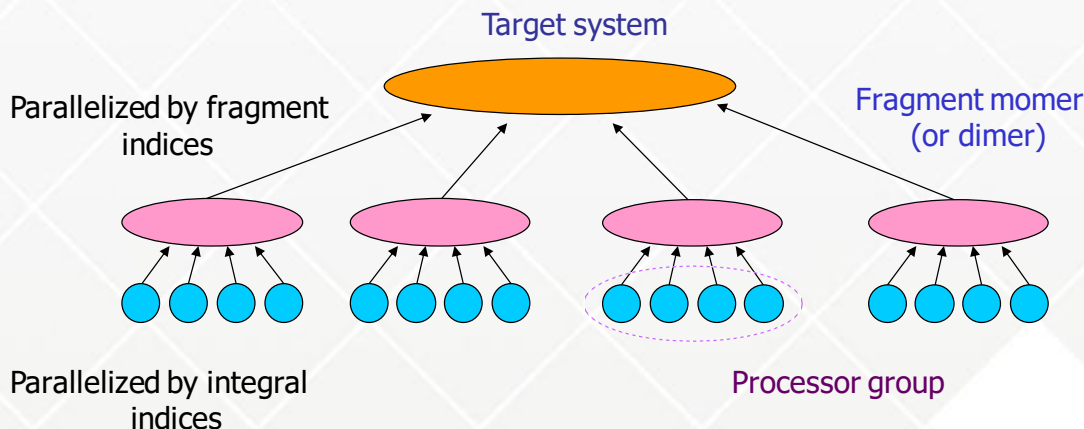
◇GAMESS-US [米国Gordonグループ]; Fedorov、北浦、中田ら
 GAUSSIANに抗し得る有力なフリーソフト、世界規模 (Fortran)
 ⇒ 様々な計算機能をFMO化、GDDI並列

◇ABINIT-MP; 望月、中野、福澤、田中ら
 実用機能は十分、東大系PJ・CREST-PJなどで開発 (Fortran)
 ⇒ MPI/OpenMP混成並列、4体展開FMO、高次電子相関、**高速**
 ⇒ 専用GUI、自動フラグメント分割で**簡便**

◇PAICS; 石川
 FMO-MP2(RI)に特化 (C)
 ⇒ MPI並列

◇OpenFMO; 稲富、鬼頭ら
 FMO-HFのみ (C)
 ⇒ 超並列指向

2階層の並列処理で計算資源を有効利用



(混成並列では数万コアも利用可能)

・エネルギー

- FMO4: HF, MP2
- FMO2: HF~CCSD(T), LRD
- Dimer-ESの連続多重極近似

・エネルギー微分

- FMO4: HF, MP2
- FMO2: MP2構造最適化, MD

・その他機能

- SCIFIE, PB水和, sp²-BDA切断
- CAFI, FILM

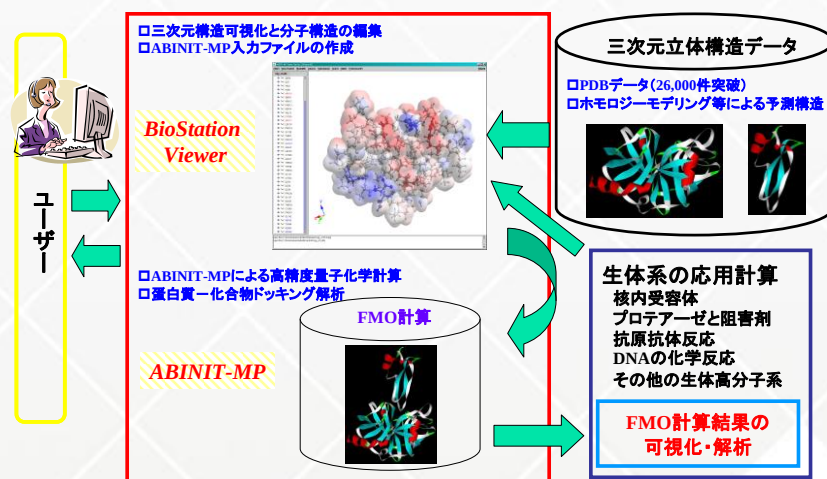
・並列化環境

- MPI, OpenMP/MPI混成
- 最深部はBLAS処理
- Xeon系小型サーバ~「富岳」まで

$$\begin{aligned}
 E^{\text{FMO4}} &= \sum_{I>J>K>L} E_{IJKL} - (N-4) \sum_{I>J>K} E_{IJK} + \frac{(N-3)(N-4)}{2} \sum_{I>J} E_{IJ} \\
 &\quad - \frac{(N-2)(N-3)(N-4)}{6} \sum_I E_I \\
 &= \sum_{I>J>K>L} \{ \Delta E_{IJKL} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{IL} - \Delta E_{JK} - \Delta E_{JL} - \Delta E_{KL} \\
 &\quad - [\Delta E_{IJK} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK}] - [\Delta E_{IJL} - \Delta E_{IJ} - \Delta E_{IL} - \Delta E_{JL}] \\
 &\quad - [\Delta E_{IKL} - \Delta E_{IK} - \Delta E_{IL} - \Delta E_{KL}] - [\Delta E_{JKL} - \Delta E_{JK} - \Delta E_{JL} - \Delta E_{KL}] \} \\
 &\quad + \sum_{I>J>K} [\Delta E_{IJK} - \Delta E_{IJ} - \Delta E_{IK} - \Delta E_{JK}] \\
 &\quad + \sum_{I>J} \Delta E_{IJ} + \sum_I E_I
 \end{aligned}$$

【4体のフラグメント展開 (FMO4)】

Ref.; Nakano et al., *Chem. Phys. Lett.* 523 (2012) 128.



【BioStationViewer / GUI】

■ 文部科学省 - 東京大学生産技術研究所拠点 / プロジェクトリーダー 加藤千幸

- ・「戦略的基盤ソフトウェアの開発」 2002年5月～2006年3月
- ・「革新的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」 2005年6月～2008年3月
- ・「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」 2008年10月～2013年3月
- ・「HPCI戦略プログラム - 分野4 次世代ものづくり」 2011年4月～2015年3月

■ 文部科学省 - 東京大学工学部拠点 / プロジェクトリーダー 吉村忍

- ・「ポスト「京」で重点的に取り組むべき社会的・科学的課題に関するアプリケーション開発・研究開発 - 重点課題6」 2015年2月～2020年3月

■ JST - CREST / プロジェクトリーダー 田中成典

- ・「フラグメント分子軌道法による生体分子計算システムの開発」 2004年9月～2010年3月

■ JSPS - 科学研究費助成事業 / 対象者 望月祐志

- ・「特定領域（代表 榊茂好） - 実在系の分子理論（公募班）」 2008年4月～2010年3月
- ・「アクチニド分子種と生体分子の相互作用に関する計算化学的研究」 2016年4月～2019年3月

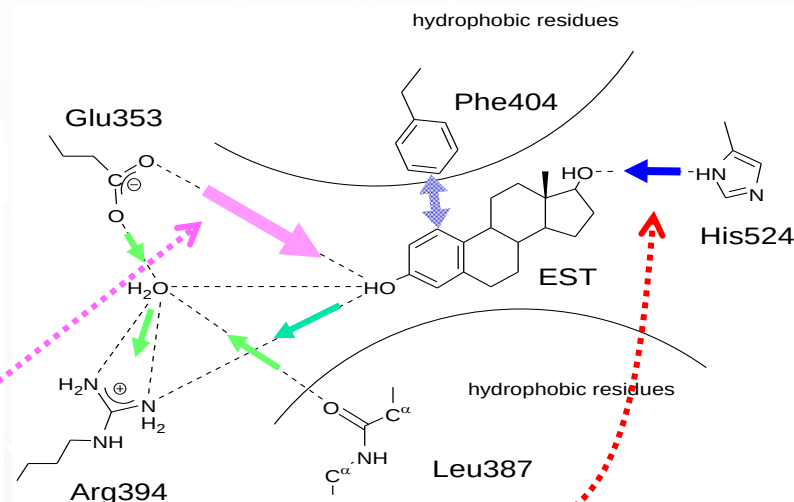
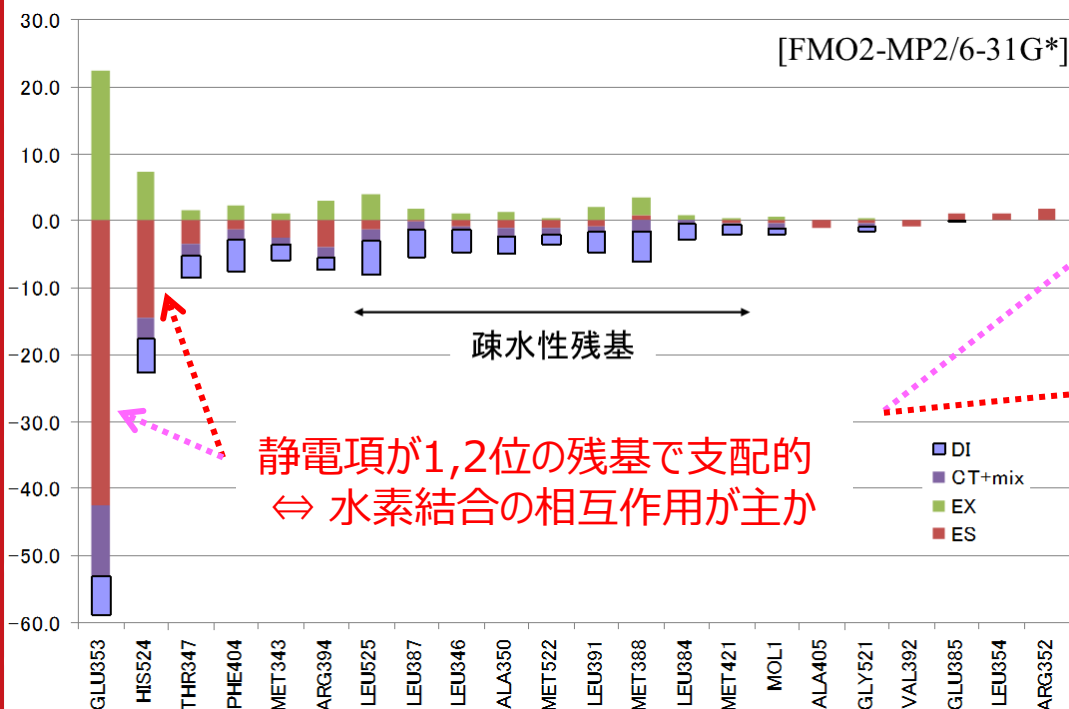
■ AMED - BINDS / 対象者 本間光貴 & 福澤薫

- ・「構造インフォマティクスとFMO計算を融合したインシリコスクリーニング支援研究」 2018年4月～2022年3月

■ 立教大学 - SFR / 対象者 望月祐志

- ・「複雑系の反応・物性の化学」 2010年4月～2013年3月
- ・「脂質およびタンパク質に関するマルチスケールシミュレーション手法の開発整備と応用展開」 2019年4月～2021年3月

女性ホルモン受容体のファーマコフォアの解析



IFIE	Lig. Chrg.
-107.74	-0.13

ES	EX	CT+mix	DI
-73.47	55.49	-29.61	-60.16

(ES: 静電的相互作用 / EX: パウリ交換反発 / CT+mix: 電荷移動など / DI: 分散力等の相関補正)

- ・ 静電相互作用と分散相互作用がオーバーオール安定化に寄与
- ・ 電荷移動も水素結合の安定化に寄与 (Glu553など)

■ 富岳の特徴

- ・ ノードあたりで48コア、利用者プログラムで20GB超のメモリ空間にアクセス出来る
- ・ 圧倒的なノード数、ラック単位(384ノード)で10程度を要求するジョブも容易に可
- ・ 端的に言えば「京」の100倍の計算力

■ 多数の分子構造による統計的な評価

- ・ capacity computing的な使い方
- ・ 構造の揺らぎを古典動力学(MD)シミュレーションで取込み、それらにFMO計算を一括実施
- ・ 静的な構造一つでの相互作用エネルギー評価と異なる動的な描像を取得可能
- ・ 本プロジェクトではメインプロテアーゼ(Mpro)の解析を対象に
- ・ FMO-MP2/6-31G*レベル、1.7千フラグメント（水含む）、1構造で0.6時間（半ラック）

■ 高コストな3次摂動計算の実行

- ・ capability computing的な使い方
- ・ MP2よりも高コストのMP3を実用的に実行、Dimer-ES近似を連続多重極展開で高速処理
- ・ スパイクタンパク質（3.3千フラグメント）を解析
- ・ FMO-MP3/cc-pVDZレベル、クローズ型(6VXX)は3.4時間で完了（8ラック）

■ 「京」から「富岳」で（今後）実現すること

- ・ 多数サンプル構造の計算が容易に可能 → 創薬検討の上でのリアリティの向上
- ・ 機械学習やデータ科学との親和性 → ビッグデータの処理、物理化学的描像の自動演繹
- ・ FMO-MP3計算の実用利用 → 定量性の向上
- ・ FMOによる重要領域の部分構造最適化 → 強化の予定（開発項目の中で高優先度）
- ・ 4体展開の高解像度FMO4計算の併用 → 薬物分子の局所相互作用の解析に利

FMO-MP2/cc-pVDZ ジョブ

FMO-MP3/cc-pVDZ ジョブ

TIME PROFILE

```
=====
Elapsed time: Monomer SCF      =          3892.7 seconds
Elapsed time: Monomer MP2      =           27.7 seconds
Elapsed time: Monomer (Total)  =          3965.3 seconds
Elapsed time: Dimer ES         =           282.4 seconds
Elapsed time: Dimer SCF        =           838.0 seconds
Elapsed time: Dimer MP2        =           473.1 seconds
Elapsed time: Dimer (Total)    =          2001.4 seconds
Elapsed time: FMO (Total)      =          5966.6 seconds
```

*** Write check point file (CPF) open1.0 rev10) ***

WriteGeom = 6vxx_nonag-min0430_50k-mp2-ccpvdz.cpf
Write coordinate data: done

Time profile

```
Number of cores (total)    =    3072
Number of cores (fragment) =        1

THREADS (FRAGMENT)        =        48
```

Total time = **6135.0 seconds** **1.7時間**

TIME PROFILE

```
=====
Elapsed time: Monomer SCF      =          3885.7 seconds
Elapsed time: Monomer MP3      =           280.6 seconds
Elapsed time: Monomer (Total)  =          4211.3 seconds
Elapsed time: Dimer ES         =           296.1 seconds
Elapsed time: Dimer SCF        =           763.6 seconds
Elapsed time: Dimer MP3        =          5701.6 seconds
Elapsed time: Dimer (Total)    =          7907.9 seconds
Elapsed time: FMO (Total)      =         12119.1 seconds
```

*** Write check point file (CPF) open1.0 rev10) ***

WriteGeom = 6vxx_nonag-min0430_50k-mp3-ccpvdz.cpf
Write coordinate data: done

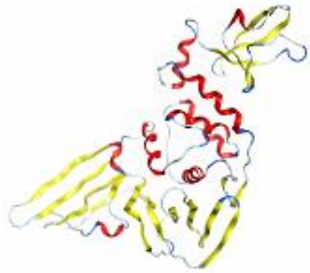
Time profile

```
Number of cores (total)    =    3072
Number of cores (fragment) =        1

THREADS (FRAGMENT)        =        48
```

Total time = **12291.0 seconds** **3.4時間**

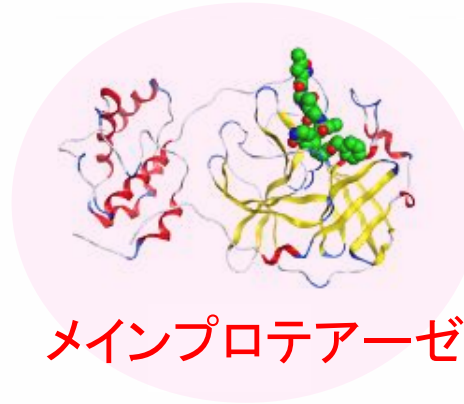
- ・ 総計では8ラックで3072プロセス×48スレッド、147,456コアを使用した超並列計算
- ・ 1ノード/フラグメント条件で実行、エネルギー成分分析(PIEDA)はオン
- ・ MP2ジョブではモノマー-SCF(HF)段階がジョブ時間の半分以上を占める
- ・ FMO-MP2に比してFMO-MP3の相対コストは2倍で済む → 「富岳」の計算力の証左
- ・ Dimer-ESは連続多重極展開で近似 → 未使用では計算は不可



Papain like protease
(1 entries)

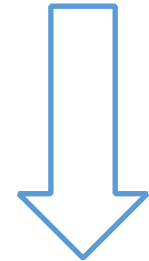


ADP ribose phosphatase
(6 entries)

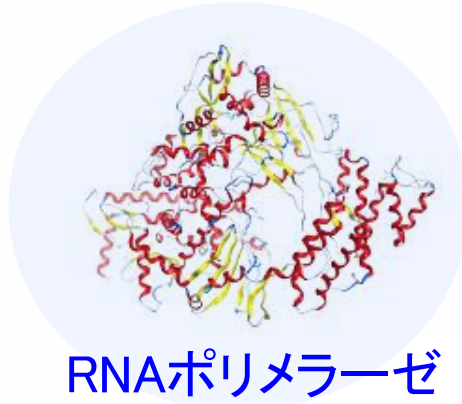


メインプロテアーゼ

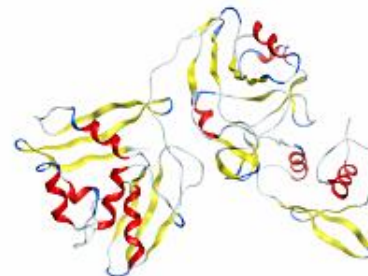
9種類のタンパク質の構造がProtein Data Bankで公開されている。



Nsp9 RNA binding protein
(4 entries)

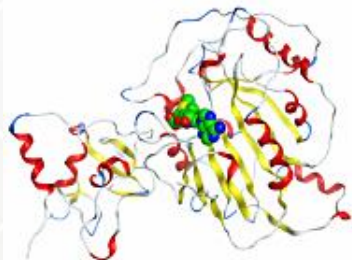


RNAポリメラーゼ

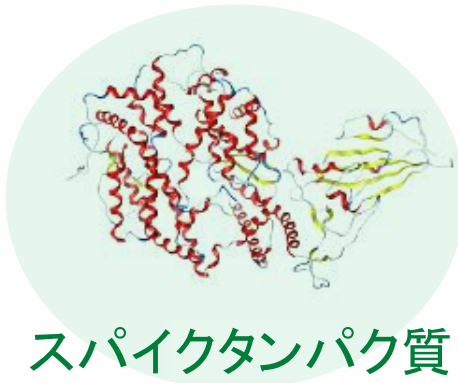


Endoribonuclease
(2 entries)

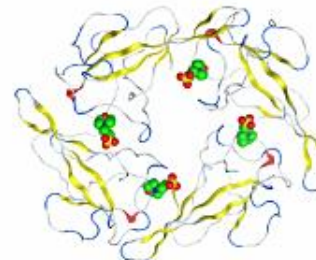
AMED-BINDSプロジェクト（本間、福澤ら）での計算結果をFMOデータベースで公開している。



2'-O-ribose methyltransferase
(2 entries)



スパイクタンパク質

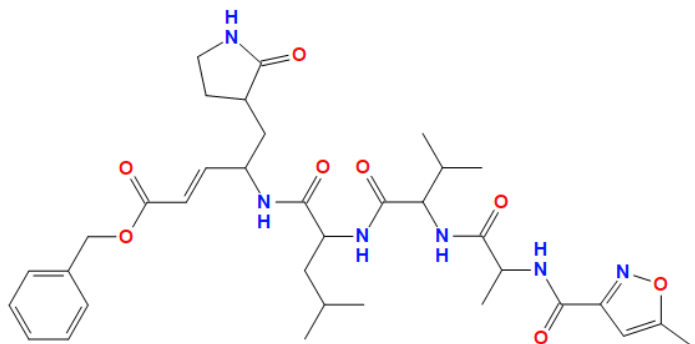


Nucleocapsid protein
(5 entries)

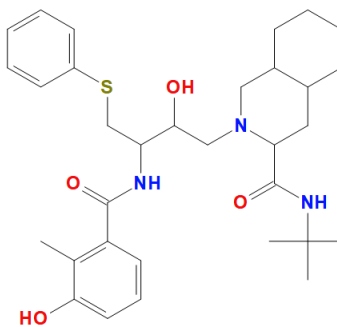


出典：FMO database

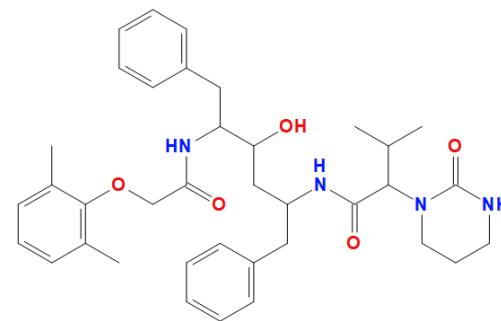
<https://drugdesign.riken.jp/FMODB/>



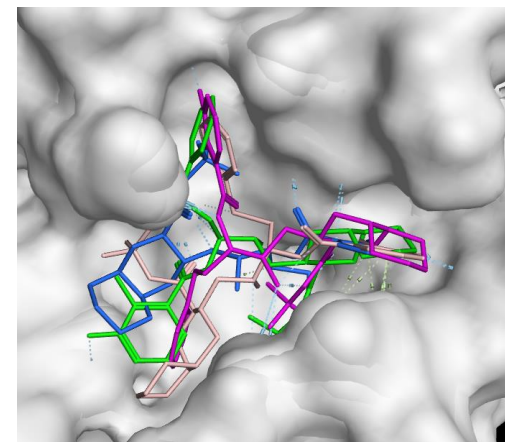
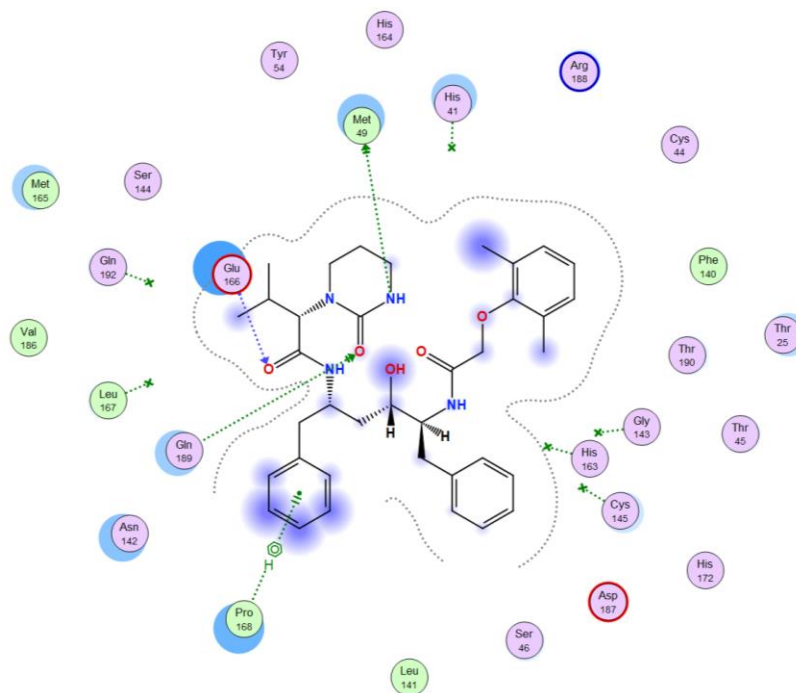
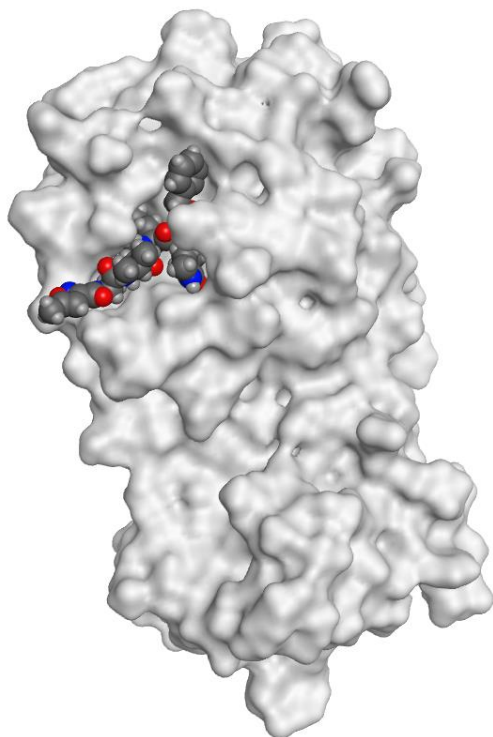
N3阻害剤(結晶構造あり)



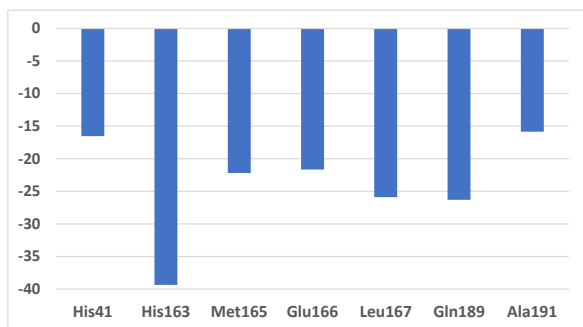
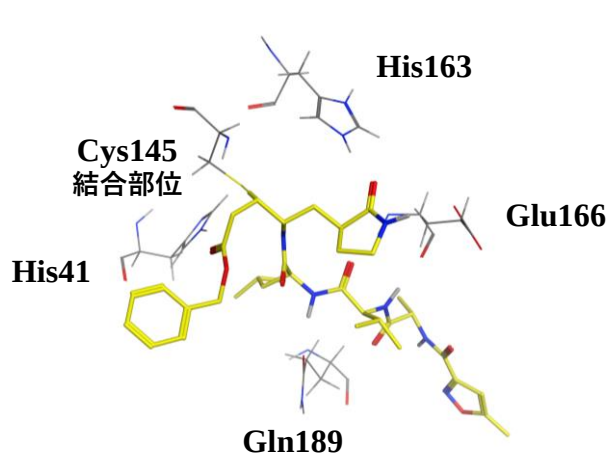
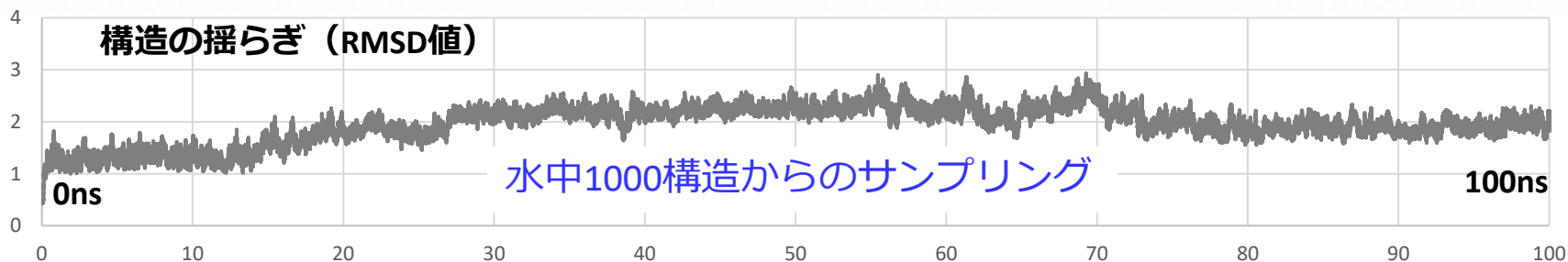
ネルフィナビル



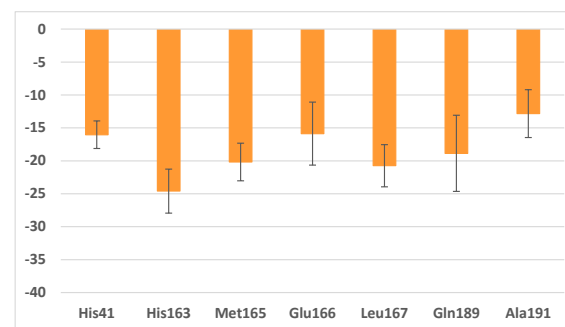
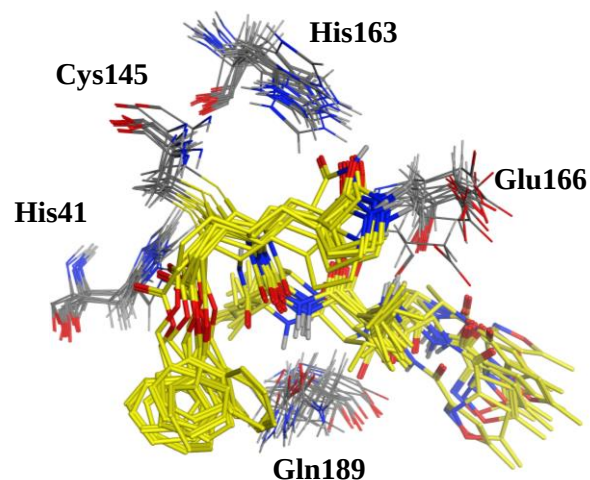
ロピナビル



既存薬との複合体構造
(ドッキング予測)

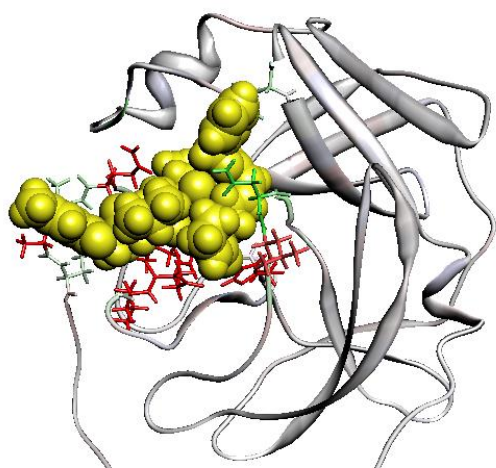


揺らぎ無しIFIE（結晶構造）

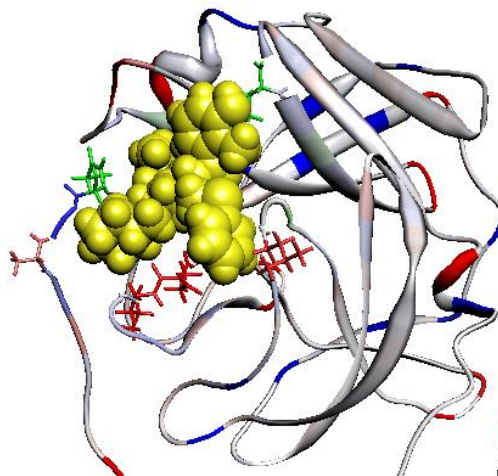


揺らぎありIFIE（生体環境）

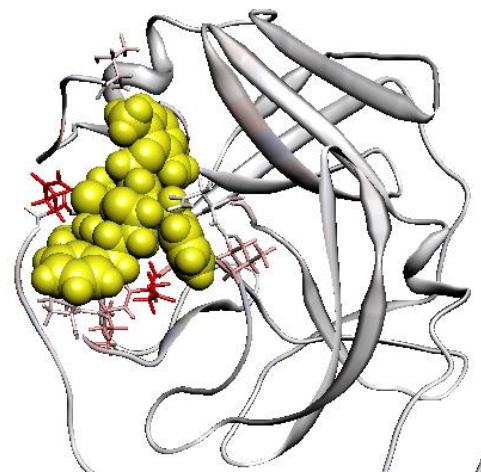
「富岳」の威力：0.6時間/1構造、100並列で1000構造が総計5時間（京の100倍超の性能）



N3阻害剤(結晶構造あり)

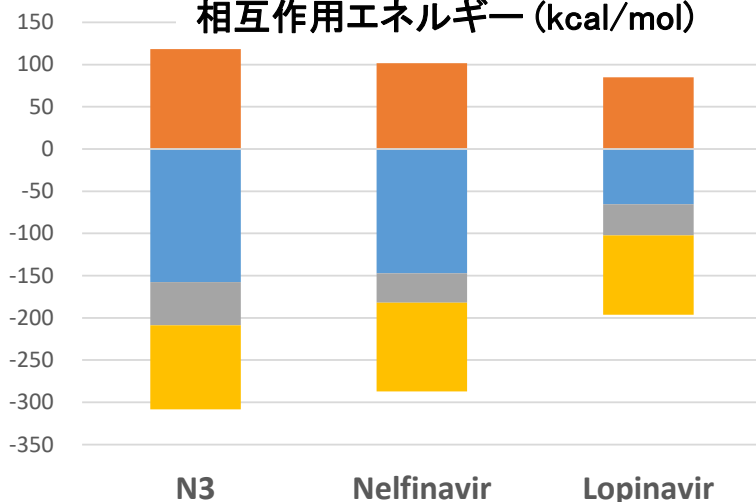


ネルフィナビル



ロピナビル

相互作用エネルギー (kcal/mol)



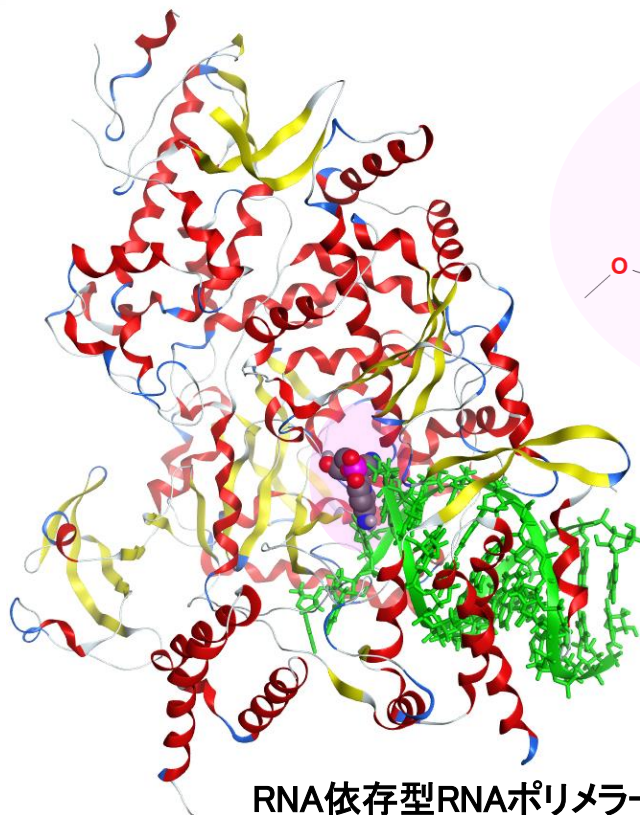
total -189.7 -185.6 -111.4

- ✓ ネルフィナビルはN3阻害剤と同程度の結合性
- ✓ ロピナビル(カレトラ)の結合性は6割程度

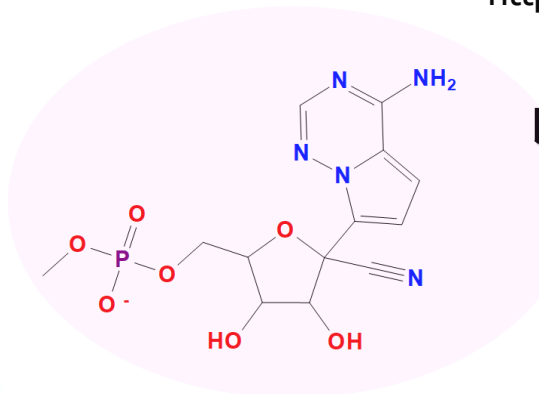


既存薬の結合性評価、新規化合物の相互作用ベースの設計に活かせる

「富岳」のパワーにより、静止画ではなく動画を用いた分子設計へと飛躍

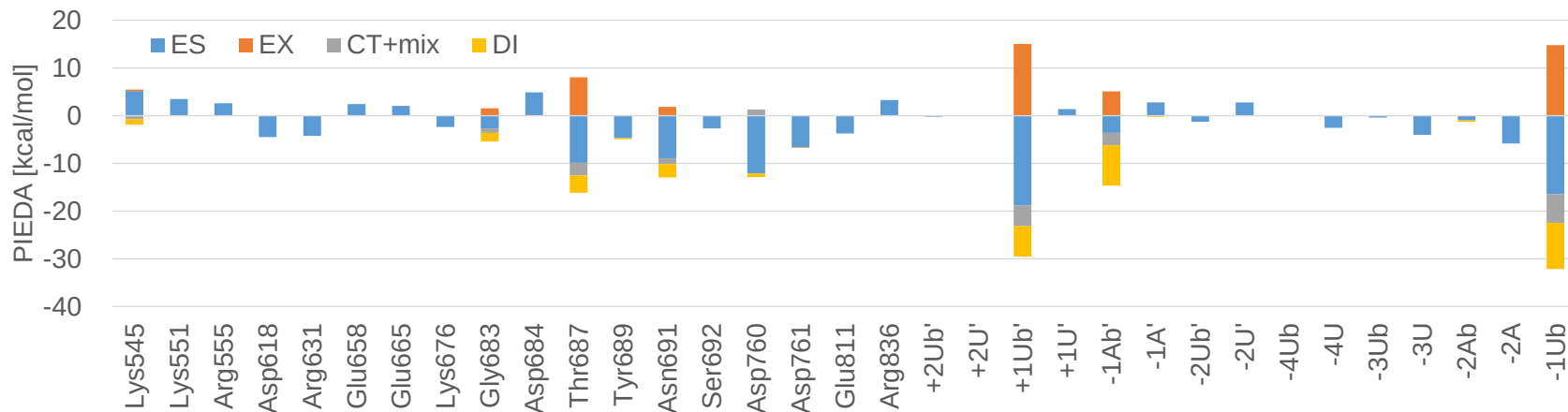
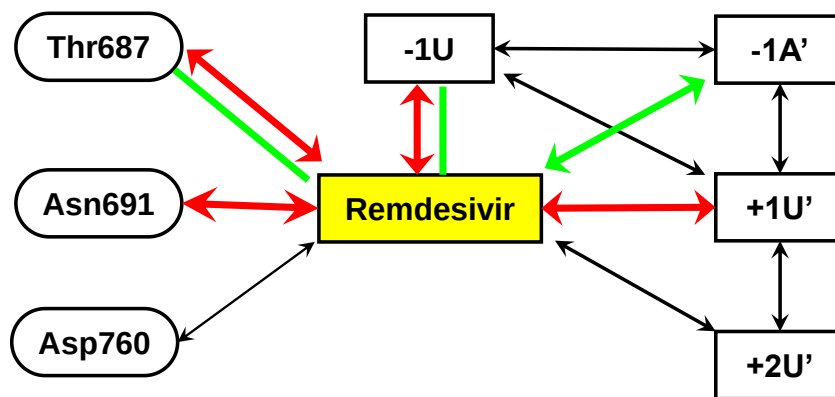


RNA依存型RNAポリメラーゼ



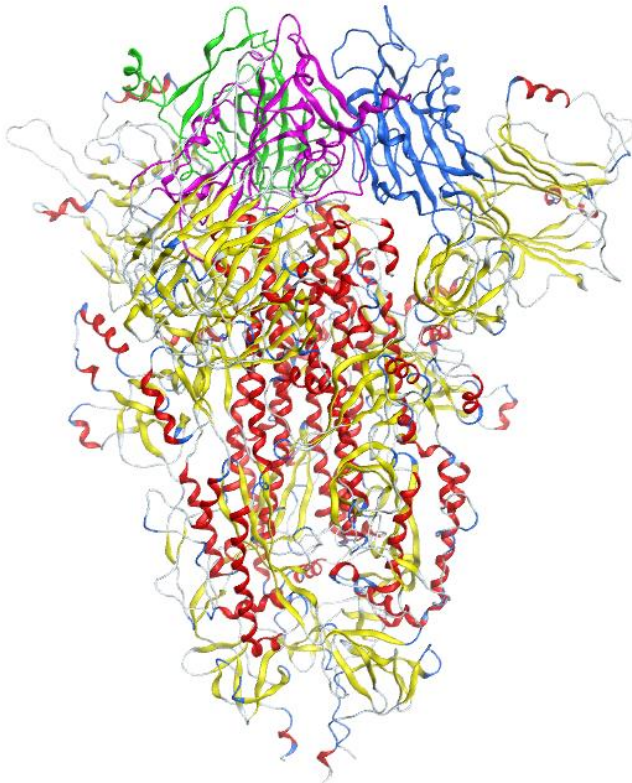
レムデシビル
(Active form)

通常のRNAとは異なる相互作用
によってRNAの伸長が止まる

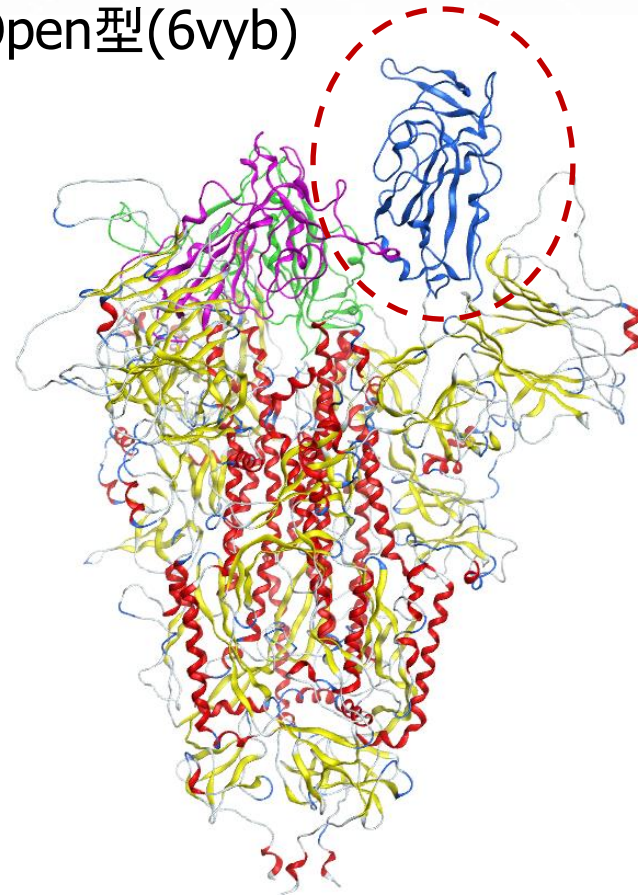


- 1100残基以上のタンパク質鎖(chain)が3ユニット絡み合って構成

Closed型(6vxx)



Open型(6vyb)



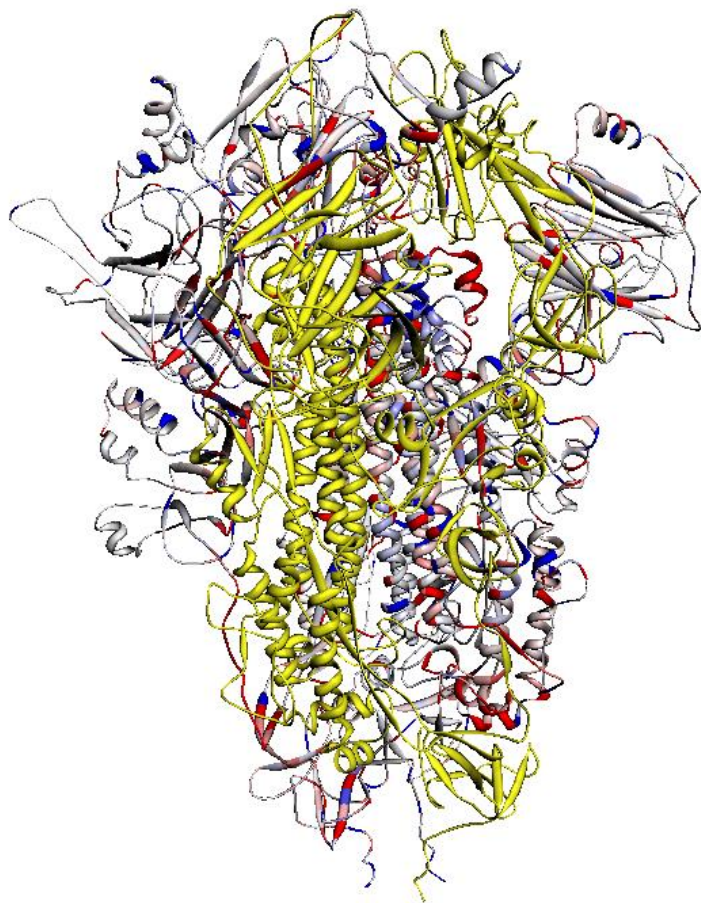
【RBD (Receptor Binding Domain): Thr333-Pro527】 A:紫、B:青、C:緑

Closed型 ⇒ Open型: Chain B のRBD領域がオープン

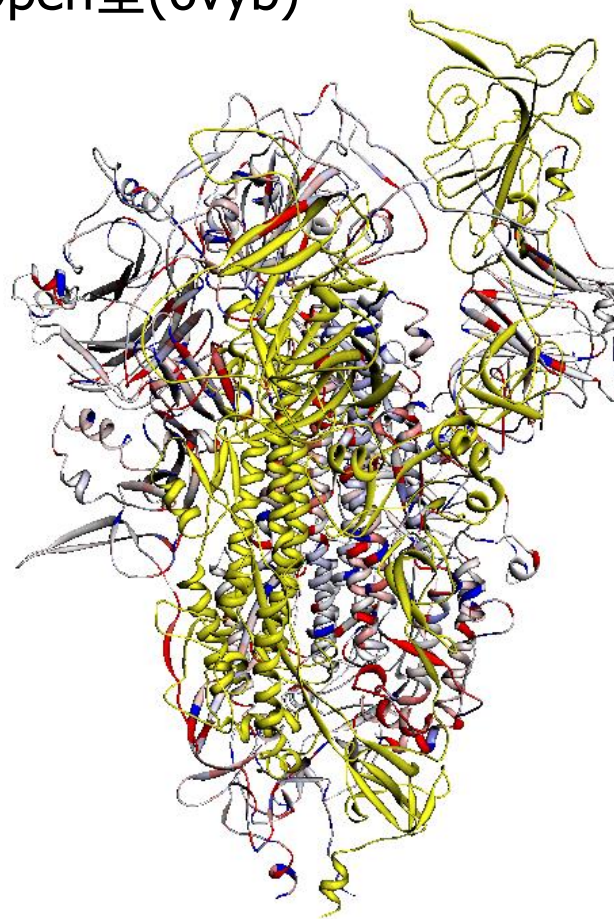
両者の相互作用様式の差異をFMO計算+データ科学的手法で検討

■ 黄色(選択領域)に対し安定化残基: 赤、不安定化部位: 青

Closed型(6vxx)

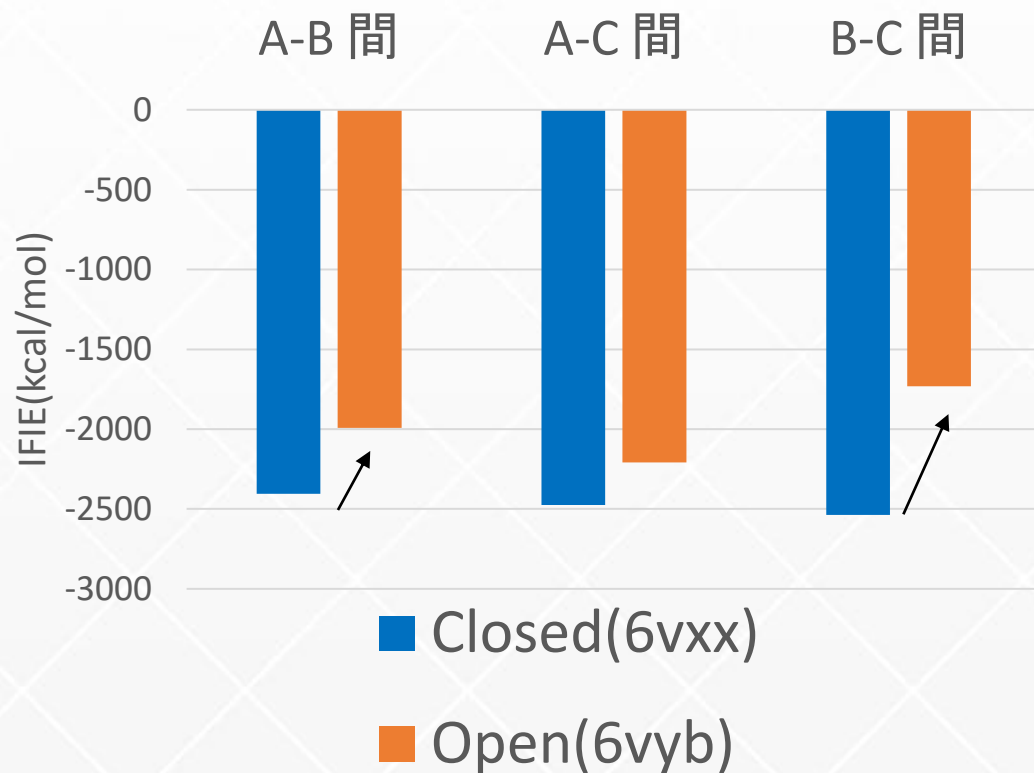


Open型(6vyb)

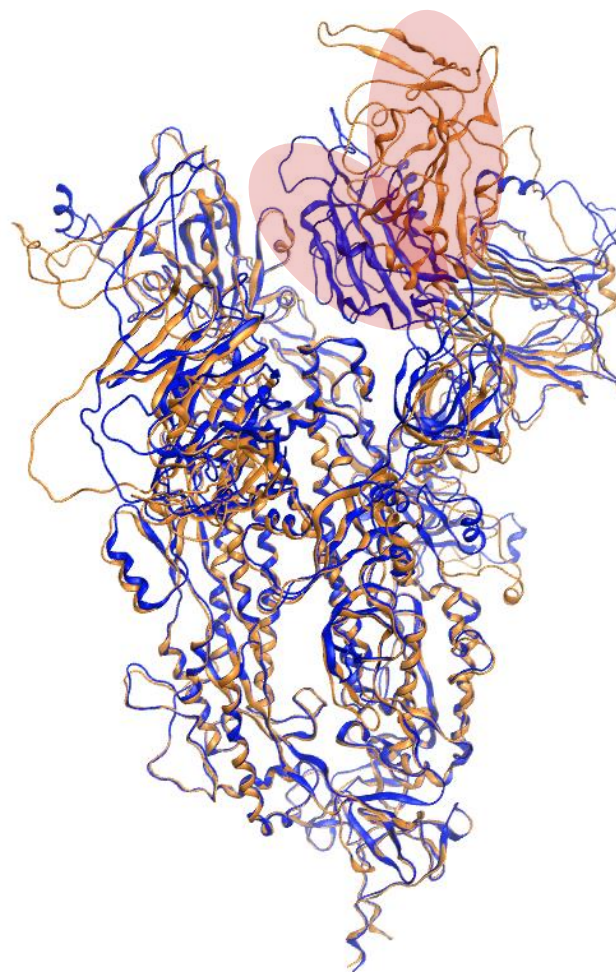


多くの残基間に強い相互作用を確認

■ 各Chain 間のIFIE(MP3/cc-pVDZ)

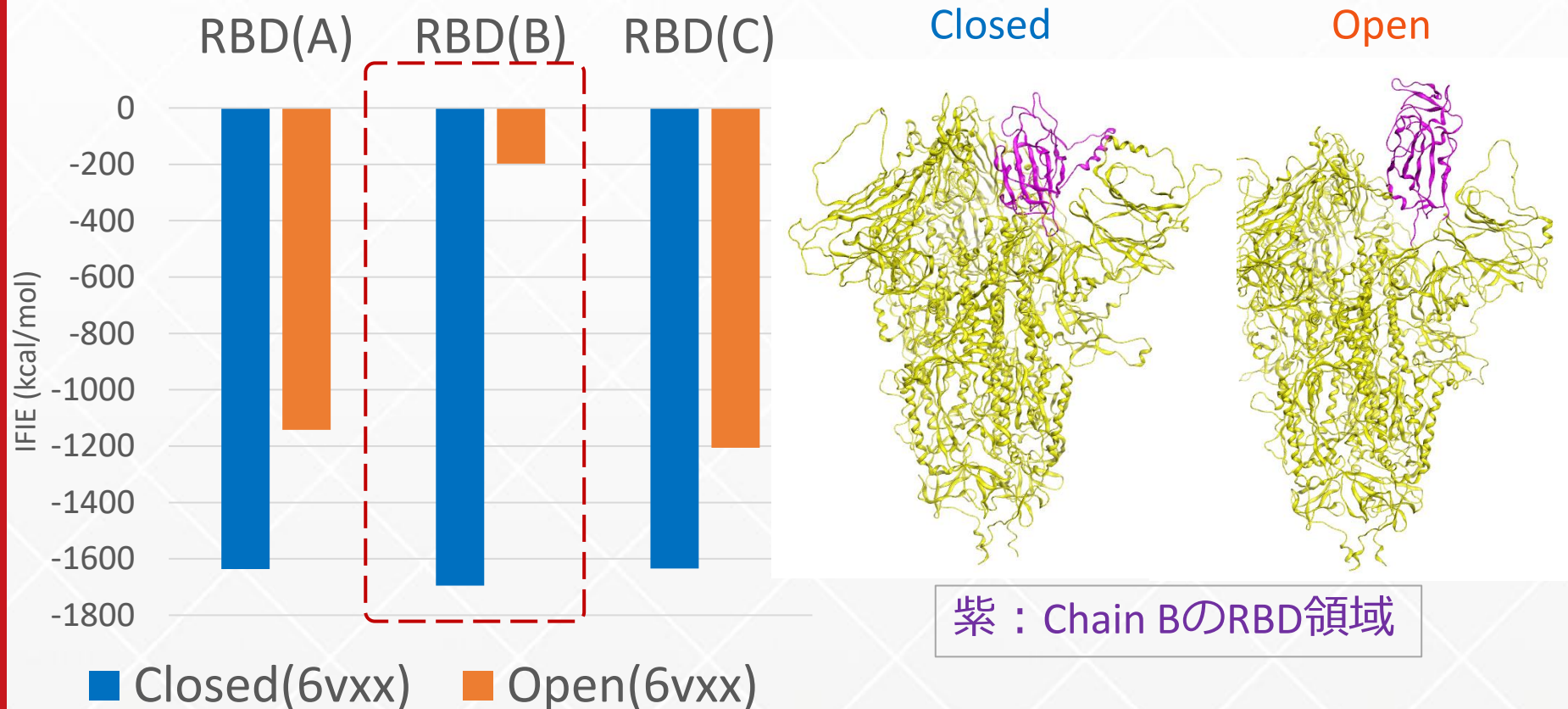


B-C間: Closed(青), Open(橙)
重ね合わせ



OpenしたChain Bとの相互作用が大きく減少

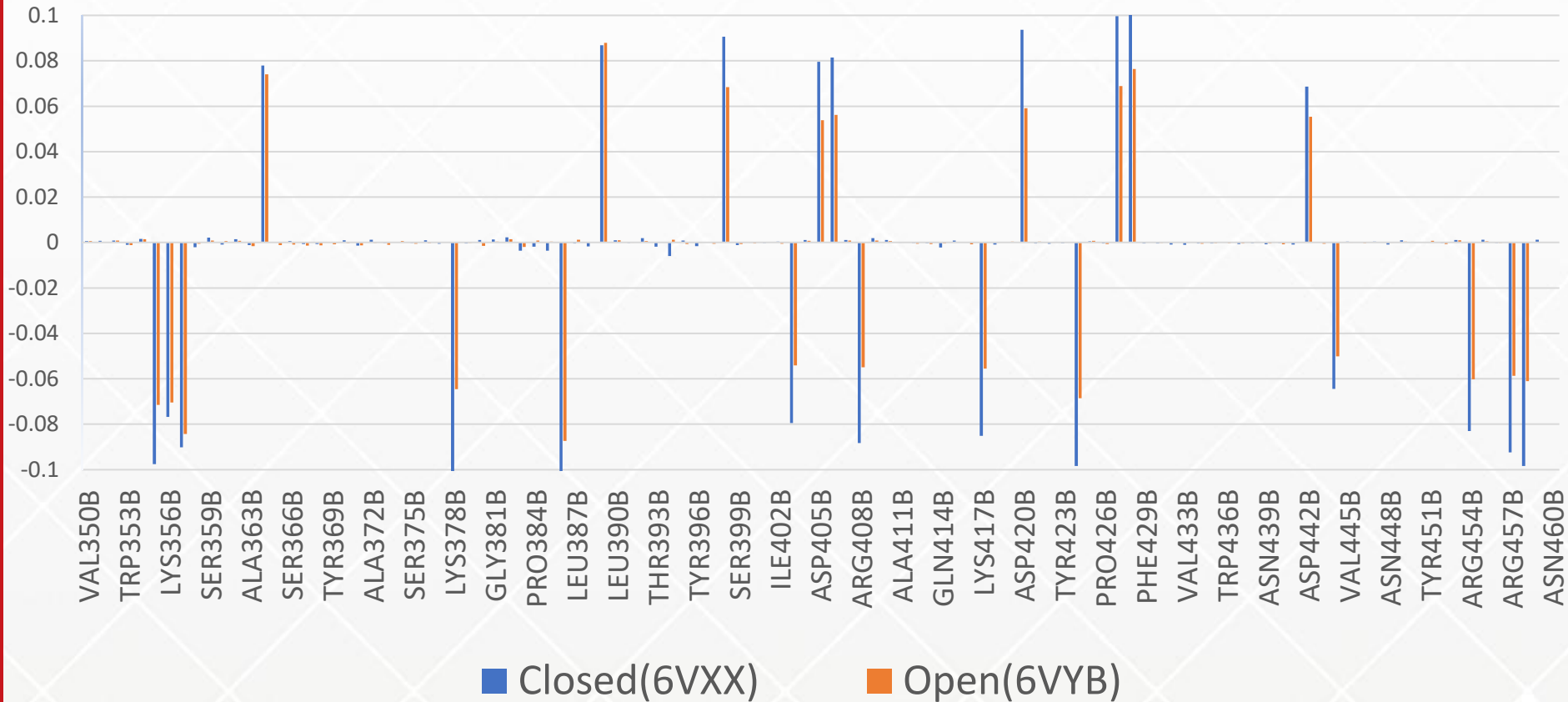
■ 各鎖RBD領域と全体のIFIE Open型-Closed型での比較 (MP3/cc-pVDZ)



OpenしたChain BのRBD領域: 周囲との相互作用が劇的に減少する
⇒ ヒトのACE2(アンジオテンシン変換酵素2)や抗体に結合可能となる

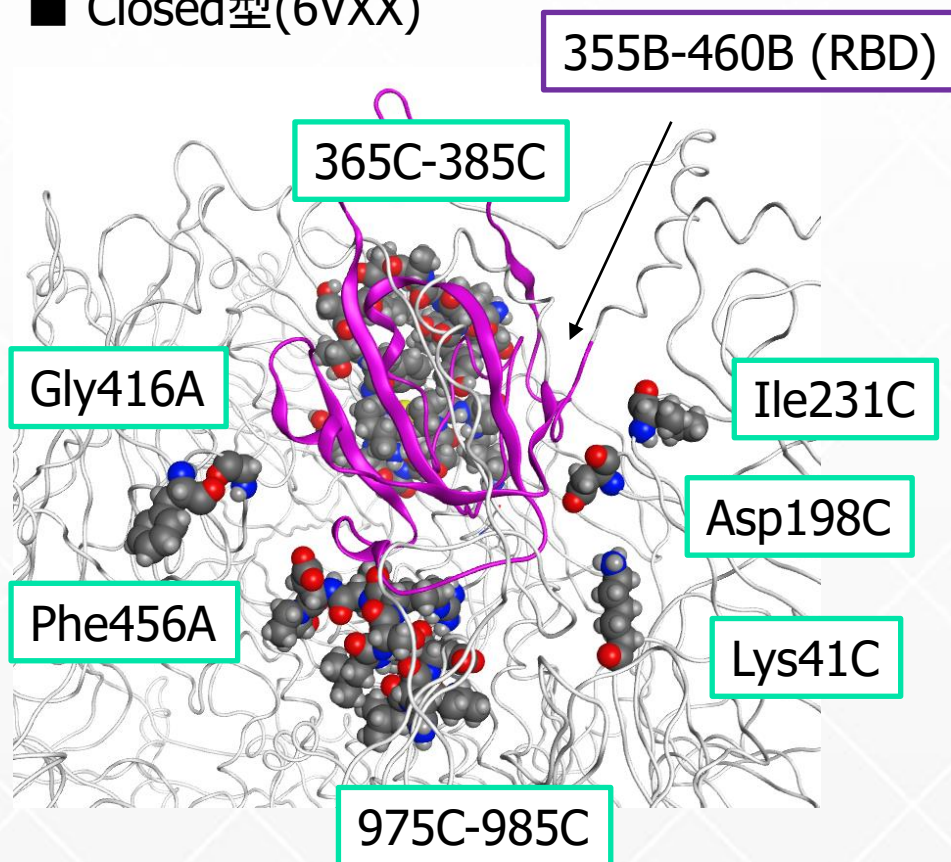
■ データ科学的手法で特徴残基を抽出

Chain B-C間 Closed, Open型での特徴量の比較



- ・RBD領域が特徴残基として抽出
- ・Closed, Openで特徴量の差が大きい

■ Closed型(6VXX)

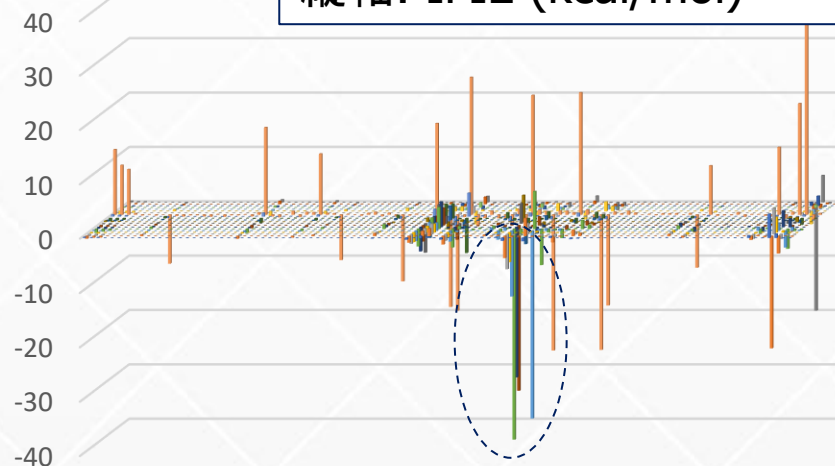


Open型になることで
特定部位の相互作用が消失

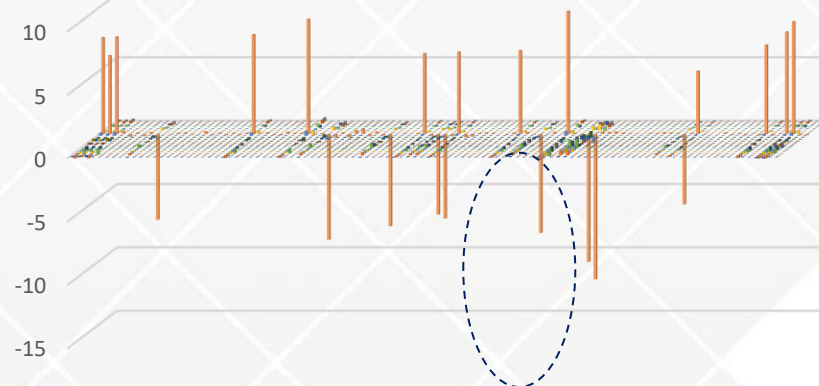
■ 特徴領域のIFIE

Closed型

横軸: Chain Bの特徴残基
奥行き: 接触残基(Chain C)
縦軸: IFIE (kcal/mol)



Open型



■ ABINIT-MPの状況

- ・ FMO-MP2ジョブは多構造サンプルの大量実行が可能 ⇔ capacity computing
- ・ FMO-MP3ジョブも実用化 ⇔ capability computing

■ メインプロテアーゼと阻害剤／既存薬

- ・ 「富岳」では「京」の100倍超の性能を用いて、静的相互作用解析から動的相互作用解析に発展させることが出来る。「結晶状態」⇒「生体環境」に
- ・ 既存薬の活性の定量的評価や新薬の設計へと繋がる

■ RNA依存型RNAポリメラーゼとレムデシビル

- ・ レムデシビルがRNA伸長に及ぼす影響を相互作用レベルで解明。「富岳」では高次電子相関による定量化や、金属原子を含むよりリアルなシミュレーションが可能に

■ スパイクタンパク質

- ・ 3量体においてOpen/Closed型の相互作用の差を残基レベルで定量化
- ・ 特にReceptor Binding Domain (RBD)に関して顕著な差 ⇒ ACE2

■ 今後

- ・ 論文の公開（プレプリントサーバ含む）
- ・ 計算結果のデータベース(FMODB)での公開 ⇔ オープンサイエンスの促進
- ・ ABINIT-MPの更なる高速化と機能強化 ⇔ ライブラリ化を意識
- ・ 主要プロテアーゼと他種リガンドとの組み合わせ ⇔ FMO4による詳細解析も
- ・ スパイクタンパク質に対する糖の影響評価。ACE2との結合

本プロジェクトの構成メンバー

■立教大学 理学部

望月祐志（教授）、奥脇弘次（助教）、畑田峻（博士前期課程2年）、秋澤和輝（学部4年）

■星薬科大学 薬学部

福澤薫（准教授）、川嶋裕介（特任助手）、半田佑磨（博士課程1年）

■神戸大学大学院 システム情報学科

田中成典（教授）

■産業技術総合研究所 健康医工学研究部門

古明地勇人（主任研究員）

■計算科学振興財団

坂倉耕太（担当課長）

■HPCシステムズ

渡邊啓正（マネージャー）

■役割分担

計算実行と解析： 奥脇、畑田、秋澤、川嶋、半田、古明地、田中、福澤、望月

ABINIT-MP実行環境整備、改良、ベンチマークテスト： 望月、坂倉、渡邊

全体統括： 望月、田中、福澤