

AlphaFold2 の「不老」 Type II でのインストールと試行的利用

Installation and trial use of AlphaFold2 on Flow Type II supercomputer

立教大理¹, 東大生研², 名大情基セ³, 東大農⁴

◦望月祐志^{1,2}, 大島聡史³, 森脇由隆⁴, 奥脇弘次¹, 秋澤和輝¹, 北原駿¹, 片桐孝洋³

Fac. Sci. Rikkyo Univ.¹, IIS Univ. Tokyo², ITC Nagoya Univ.³, Fac. Agri. Univ. Tokyo⁴

◦Yuji Mochizuki^{1,2}, Satoshi Ohshima³, Yoshitaka Moriwaki⁴, Koji Okuwaki¹,

Kazuki Akisawa¹, Shun Kitahara¹, Takahiro Katagiri³

E-mail: fullmoon@rikkyo.ac.jp

【序】 2021 年の 7 月 16 日、英 DeepMind 社が公開したタンパク質の構造推定システム AlphaFold の Ver. 2 (以下 AF2) [1]は、構造生物学や生物物理学の境界で大きな話題となり、国内でも多くの研究者が導入と利用を試み (例えば Qiita 記事[2])、その有効性と可能性が高く評価されました。AF2 の入力データは、一文字表記のアミノ酸残基の配列情報のみで、類縁構造のデータベース検索、深層学習と構造緩和の一連の処理が実行され、候補構造が PDB フォーマットで出力されます。今回の発表では、名古屋大学情報基盤センターのスパコン「不老」[3]の Type II サブシステムに AF2 をインストールし、試行的な利用を行った事例をご紹介します。

【インストール】 AF2 のデータベースの量は 2TB を超えるため、高速検索には SSD の利用が望ましいとされます[2]。また、深層学習の処理には GPU による加速が有効です。「不老」Type II は、SSD と GPU (NVIDIA Tesla V100) を共に備える特徴的なスパコンで、プラットフォームとして好適です。AF2 の実際のインストールは Qiita 記事[4]に詳細がありますが、作業用 SSD 領域の永続化と共有化などの設定を行いました。Q9KVZ7 例題では、データ検索のステップが SSD を使わないと 11 時間なのに対し、SSD 使用では 10 分程に短縮されることを確認しました。AF2 は、アカデミック向けに既に公開され、一般アカウントでもリーズナブルな課金で利用可能です[3]。

【試行的利用】 幾つかの例題を試みた後、感染症対策関連の FMO 計算プロジェクト (hp210026 課題) で利活用を模索中です。ここでは、以前解析したインフルエンザウイルスのヘマグルチニン (HA) と Fab 抗体の複合体 (PDB-ID: 1EO8) [5]を扱った例を記します。HA、Fab は各々 2 本の鎖から成りますが、X 線構造に比して HA、Fab 各々の構造はリーズナブルに再現されました。しかし、HA-Fab の接合面に関しては大きくずれた予測となってしまう、異種タンパク質の多量体の扱いについては注意が要ることも分かりました (結果の詳細は講演当日にお示しします)。

【今後】 AF2 のデータベースの補助データの拡充 (例えば[6]) もなされていますので、適用範囲の拡大が期待され、特に抗原-抗体系の予測精度の向上が望まれます。こうした改良も踏まえ、今後、AF2 を含めたタンパク質の計算化学のワークフローが常用されるようになると思われます。

【謝辞】 「不老」Type II は、JSPS 科研費 JP19H05662 のアカウントで利用しました。

【参考文献】 [1] J. Jumper et al., *Nature* **596** (2021) 583. [2] <<https://bit.ly/3qbbx1r>>. [3] <<https://icts.nagoya-u.ac.jp/ja/sc/>>. [4] <<https://bit.ly/33xpbV2>>. [5] K. Takematsu et al., *J. Phys. Chem. B* **113** (2009) 4991. [6] M. L. Hekkelman, et al., *bioRxiv* <<https://bit.ly/3mdQbPN>>.