

## Journal Club (2023 年 12 月 18 日) まとめ

担当: 清水 秀幸

### 発表論文:

Watson JL, ..., David Baker.

De novo design of protein structure and function with RFdiffusion.

Nature. 2023 Aug;620(7976):1089-1100

doi: 10.1038/s41586-023-06415-8

### 研究目的および概要:

タンパクデザインを in silico で精密に行うことは、ライフサイエンスの謎を紐解くという点はもちろん、将来的な治療薬創出にもつながりうる。近年 Stable diffusion などの生成 AI が台頭しており、拡散モデルを用いてタンパク質のデザインに挑戦した。

筆者らのグループがこれまでに開発していた RoseTTaFold (RF) と拡散モデルを組み合わせた RFdiffusion の開発に成功し、タンパク質のバックボーンを生成できるようになった。PDB に登録されている小さめのタンパク質を題材に訓練をしており、さらに例えば標的タンパク質に対する binder の設計などは RFdiffusion の微調整を行うことでこの課題に取り組んだ (図 1)。

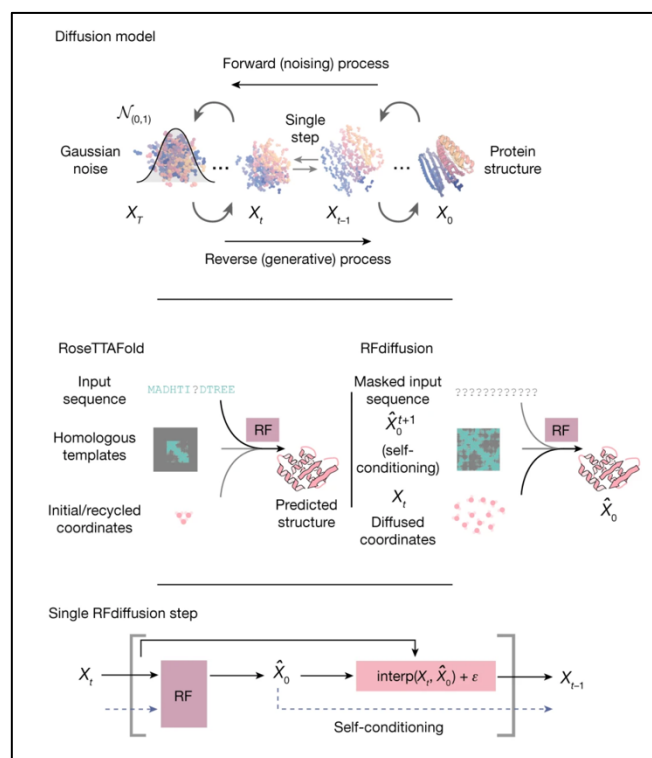


図 1

RFdiffusion が行うのはタンパク質のバックボーンを生成するところのみであり、その主鎖構造になりそうなアミノ酸配列の生成についてはすでに開発済みの ProteinMPNN を用いて行う。そして生成されたアミノ酸配列について、AlphaFold2 や RoseTTaFold などによって目的の構造をとるかをフィルタリングした後に、そのタンパク質を発現させ実験的検証を行っている。筆者らは多様な構造を持つタンパク質の生成に成功しており、さらに驚くべきことに任意の標的タンパク質に結合するタンパク質もおよそ 10 から 20%ほどの成功率でデザインできるようにな

った。これはこれまでの成功率を大きく凌駕するパフォーマンスである (図 2)。

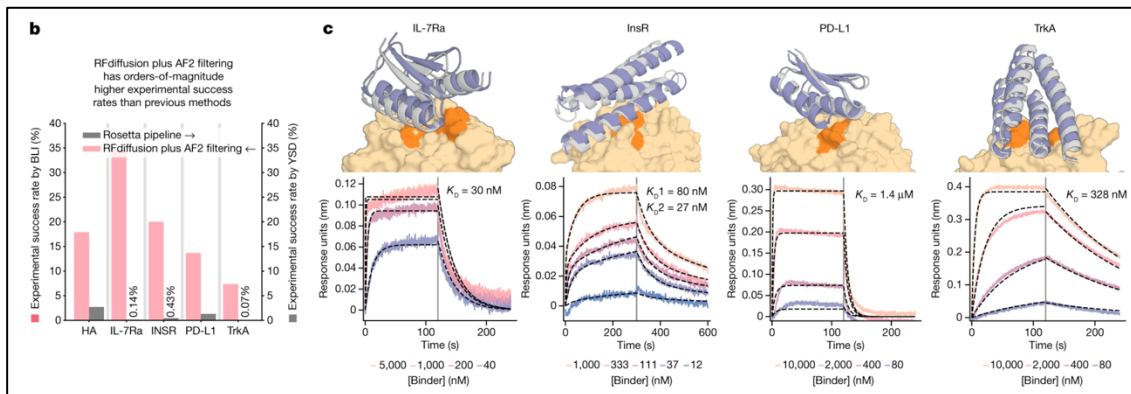


図 2

### 先行研究と比べて何がすごい？ 技術やアプローチのキモはどこ？：

- 業界で標準的な DDPM と呼ばれる拡散モデルを使っているのは同じだが、DDPM に入力させるデータと筆者らのタンパク構造予測アルゴリズムである RoseTTaFold との類似性を見つけそこを掘り下げている。逆拡散過程におけるネットワーク構造を SE(3)-transformer を活用するなどタンパク質の回転や並進性にも対応したものを使っている点は筆者らのタンパク質科学に関するドメイン知識が使われている。

### どうやってこの手法/仮説の有効性を検証したのか：

- RFdiffusion+ProteinMPNN により提案されたタンパク質を発見させ、CD 法により 2 次構造に大きな違いがないとか、BLI を用いた種々のハイスループット結合アッセイにより目的のタンパク質に結合することを通して多様な構造を意図したとおりに作成できることを示した。

### その他、議論した内容 (ネガティブコメントや limitation もあれば)：

- 熱安定性を示しているのはごく一部の例だけであるからもう少し見たい
- Binder を生成するところにおいて、アフィニティーは抗体等に比べてまだ大きく劣る。多様なタンパクを作るという方向性ではなく、指定した分子に対しより高いアフィニティーで結合できる分子生成ができないと医療応用としては弱い
- 成功率はよくて 2 割程度であり、まだ extensive な実験検証が不可欠な分野
- 拡散モデルに限らず、新しいイノベーションは分野外からやってくる。アンテナを張るのが大事

### この研究をさらに発展させるとしたら：

- 抗体もしくはアプタマーにおけるデータを取り込んだモデルを作製する