

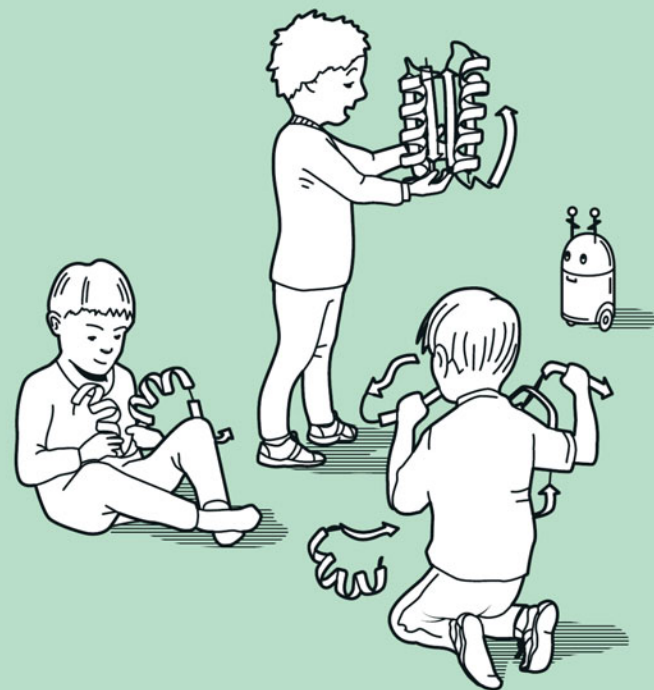
第49回 けいはんな「エジソンの会」  
公益財団法人 国際高等研究所  
2025年4月25日（土）15:10～16:10

# AlphaFold

タンパク質の立体構造予測がもたらす  
未来の創薬・医療

東京科学大学 情報理工学院 准教授

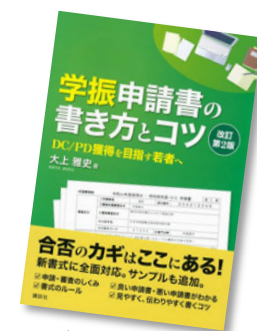
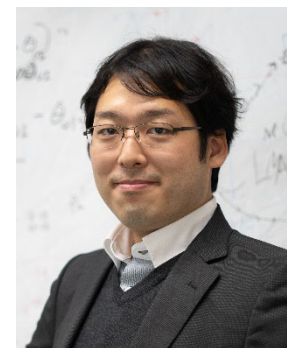
おおうえ まさひと  
大上 雅史



## 東京科学大学 情報理工学院 情報工学系 准教授

略歴（1987年3月生まれ・石川県出身）

- 2007年 石川工業高等専門学校 電子情報工学科 卒業
- 2007年～2009年 東工大 工学部 情報工学科（3年次編入学）
- 2009年～2011年 東工大 大学院情報理工学研究科 計算工学専攻 修士課程
- 2011年～2014年 東工大 大学院情報理工学研究科 計算工学専攻 博士課程、博士（工学）  
（指導教員：秋山 泰 教授）
- 日本学術振興会 特別研究員-DC1
- 2014年～2015年 日本学術振興会 特別研究員-PD
- 2015年～2016年 東工大 大学院情報理工学研究科 助教
- 2016年～2020年 東工大 情報理工学院 助教
- 2020年～2023年 東工大 情報理工学院 助教（テニユアトラック適用、PI化）
- 研究室開設（@すすかけ台キャンパス）  <https://www.li.c.titech.ac.jp>
- 2024年～現在 東工大 情報理工学院 准教授
- （2024年10月～ **東京科学大学** 情報理工学院 准教授）



代表著作

- 2015年～2021年 一般社団法人 日本生物物理学会 理事
- 2018年～現在 アヘッド・バイオコンピューティング株式会社 取締役
- 2019年～現在 特定非営利活動法人 並列生物情報処理イニシアティブ(IPAB) 理事
- 2020年～現在 特定非営利活動法人 日本バイオインフォマティクス学会 理事
- 学術変革領域研究(A)「潜在空間分子設計」計画班代表者&総括班分担者

# 計算でタンパク質の構造を予測し、新しいタンパク質を作る

タンパク質の理解と創造の革新を導いた研究者たち —— 2024年ノーベル化学賞



NOBELPRISET I KEMI 2024  
THE NOBEL PRIZE IN CHEMISTRY 2024



**David Baker**  
University of Washington  
USA

*"för datorbaserad proteindesign"*

*"for computational protein design"*



**Demis Hassabis**  
Google DeepMind  
United Kingdom

*"för proteinstrukturprediktion"*

*"for protein structure prediction"*



**John M. Jumper**  
Google DeepMind  
United Kingdom

## AlphaFold2/3

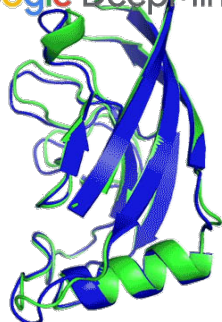


T1037 / 6vr4  
90.7 GDT  
(RNA polymerase domain)



Demis Hassabis, III, Niklas Elmehed © Nobel Prize Outreach

## Google DeepMind

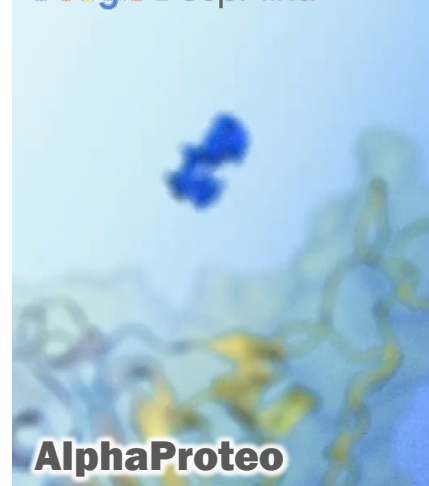


T1049 / 6y4f  
93.3 GDT  
(adhesin tip)



John Jumper, III, Niklas Elmehed © Nobel Prize Outreach

## Google DeepMind



**AlphaProteo**



Demis Hassabis, III, Niklas Elmehed © Nobel Prize Outreach



**RFdiffusion  
ProteinMPNN**



David Baker, III, Niklas Elmehed © Nobel Prize Outreach



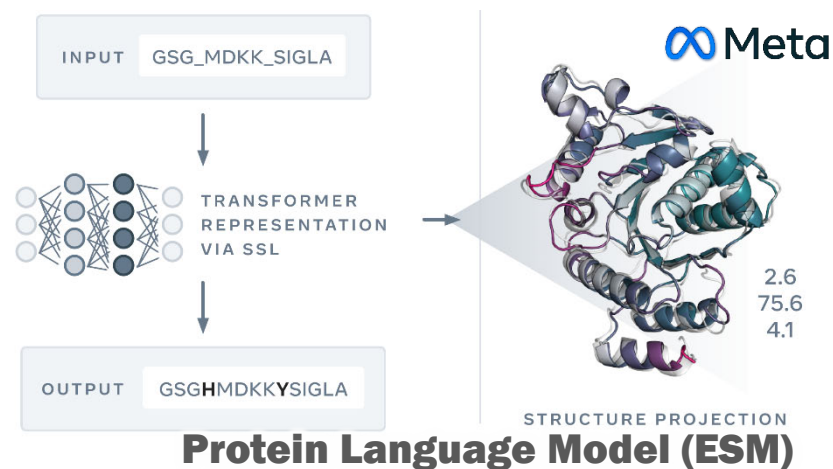
**AlphaMissense**

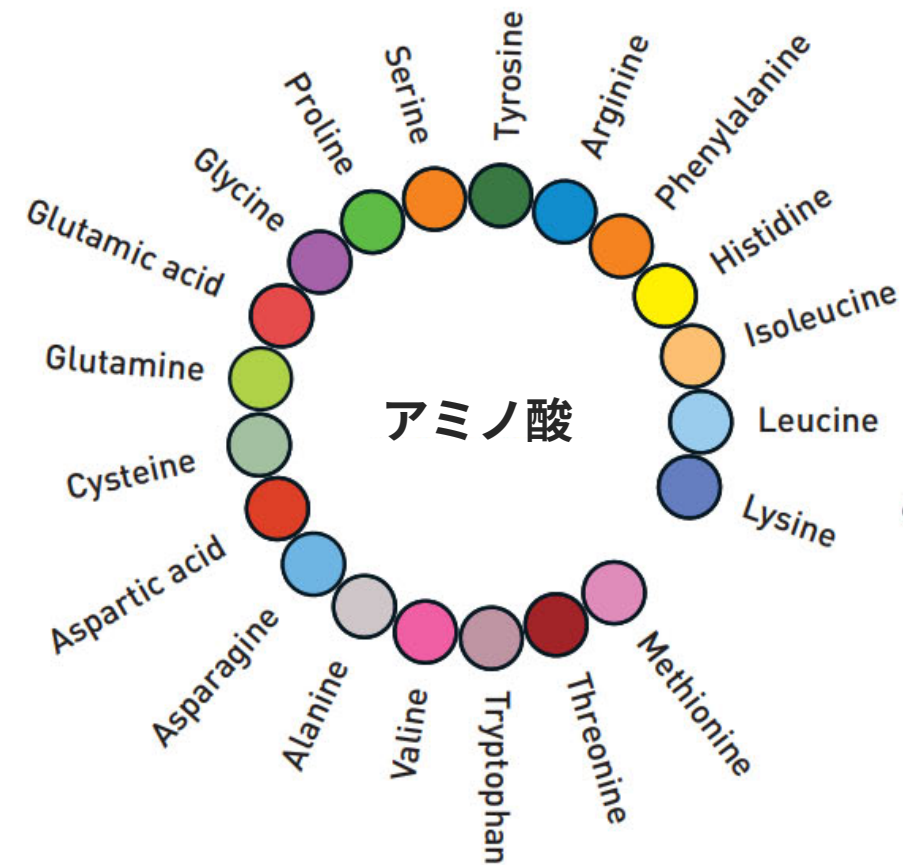


Demis Hassabis, III, Niklas Elmehed © Nobel Prize Outreach



John Jumper, III, Niklas Elmehed © Nobel Prize Outreach





## タンパク質

立体構造

$\alpha$ ヘリックス

$\beta$ シート

— アミノ酸配列

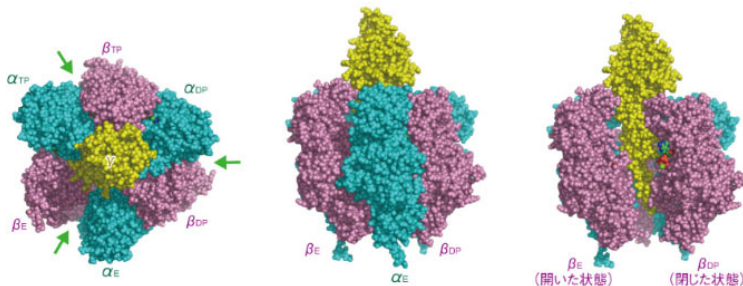


## • 実際に見る = 「構造生物学」

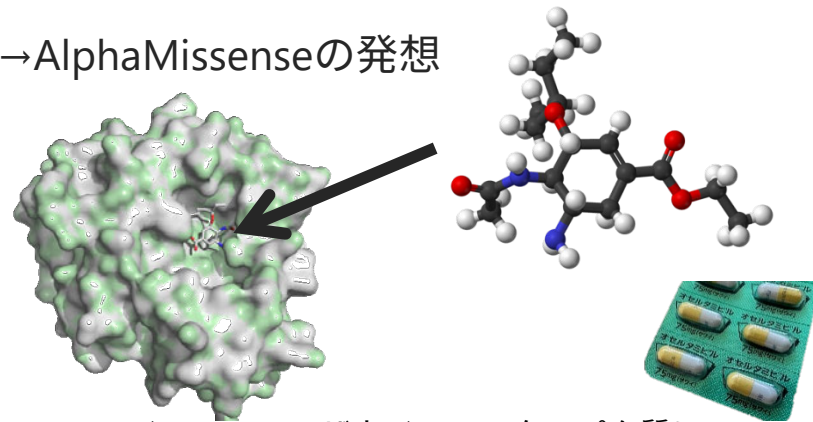
- X線結晶構造解析、NMR、クライオ電子顕微鏡などの実験技術
- 多くの構造生物学者がタンパク質の構造解明に骨身を削る
  - 日本もタンパク3000PJやターゲットタンパクPJなど国プロで貢献
  - 大型放射光施設や、構造データ管理 (PDBj/阪大) などの国際的貢献も
- 論文を発表する際には「必ず」構造データ(PDB)の公開を義務付け
  - のちにAIが活躍する布石に

## • 構造がわかると何が嬉しい？

- 姿かたちがわかる、機能が推定できる（似たもの同士の比較）
- 疾病理解 & 薬剤設計のヒントに
  - 変異があったらヤバそうな場所もわかる→AlphaMissenseの発想



分子モーターの構造  
[Iino et al., 2011]



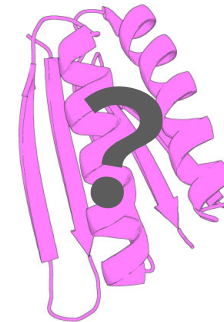
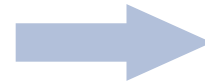
インフルエンザウイルスのタンパク質に  
結合する治療薬タミフル [Kim et al., 1997]

## • 実験でしかわからないの？

– Yes、だが推測ならできる＝「タンパク質立体構造予測」

DIQVQVNIDDNGKNFDYTYTVTTESELQKVLNEL  
MDYIKKQGAKRVRISITARTKKEAEKFAAILIKV  
FAELGYNDINVTFDGDVTVEGQLEGGSL

タンパク質のアミノ酸配列



タンパク質の立体構造

## • タンパク質立体構造に関する経験則

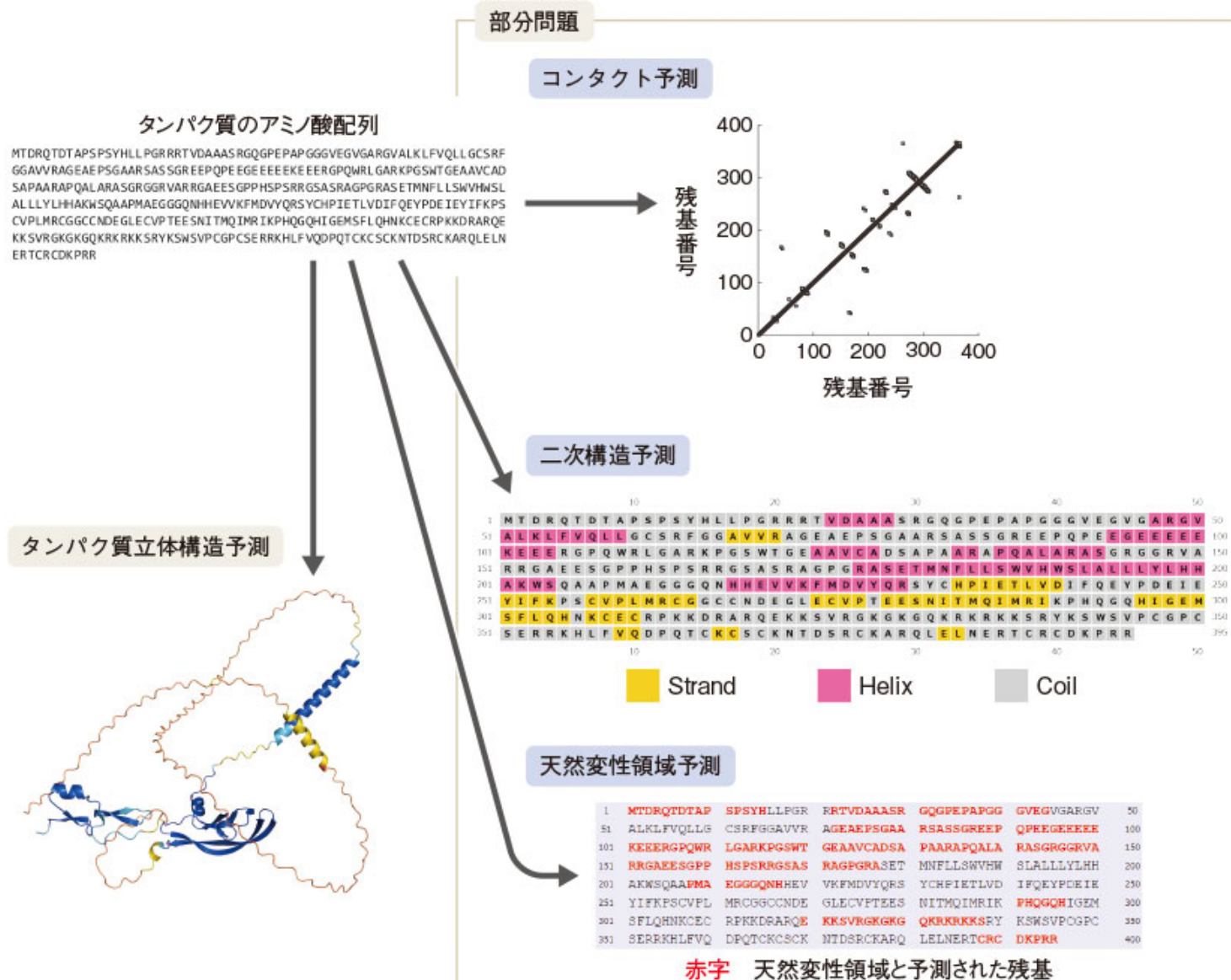
– 配列が似ていると、構造も似ている（ことが多い）

- だいたい40%くらいの配列の一致があると同一構造
- ゲノムや分子の進化と関係

＝類縁配列（親戚）がいると構造がわかる（＝ホモロジーモデリング法）

– 親戚がいないとよくわからない

- 部分的な構造（ $\alpha$ ヘリックスや $\beta$ シート、2次構造）なら当たる



## • CASP

- タンパク質構造予測の国際コンペ
- 配列を出題して、その構造を当てる  
(主催側は、裏で構造解析を行って、答えとなる構造を得ておく)
- 2年に1回開催、2024年にCASPの16回目 (12月頭に結果発表)
- 様々な問題カテゴリが存在
  - 立体構造予測が難しかったときにあったカテゴリ
    - コンタクトマップ予測
    - 2次構造予測
    - 天然変性領域予測
  - より難易度の高いカテゴリ
    - 多量体構造予測 (CAPRIコンペから合流)
    - RNA立体構造予測
    - 他の分子 (RNA, 低分子, 金属イオン等) との複合体構造予測



The image is a screenshot of a news article from the journal Nature. The top of the page features the 'nature' logo on the left and a link to 'View all Nature Research' on the right. Below the logo is a navigation bar with three items: 'Explore our content' with a dropdown arrow, 'Journal information' with a dropdown arrow, and 'Subscribe'. The main content area has a breadcrumb trail 'nature > news > article'. The article is dated 'NEWS • 30 NOVEMBER 2020'. The headline is in large, bold, black font: '‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures'. Below the headline is a sub-headline in a smaller, regular black font: 'Google’s deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.'

nature

View all Nature Research

Explore our content ▾ Journal information ▾ Subscribe

nature > news > article

NEWS • 30 NOVEMBER 2020

## ‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures

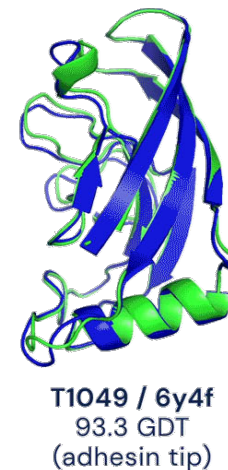
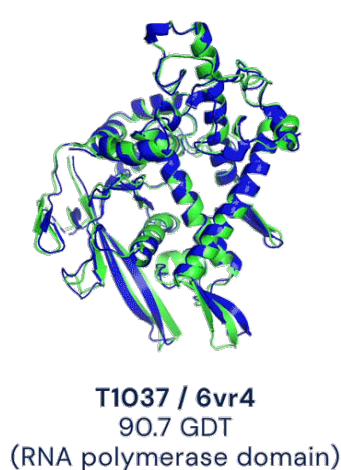
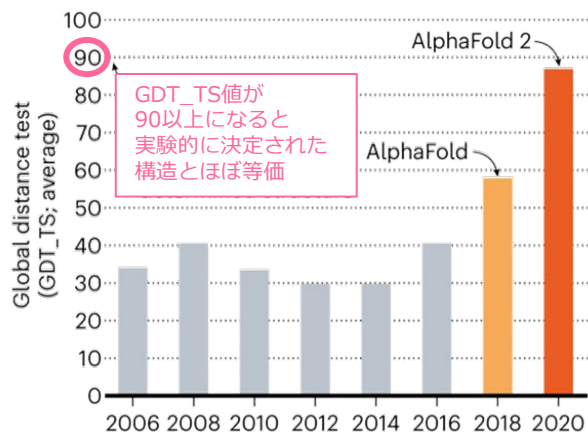
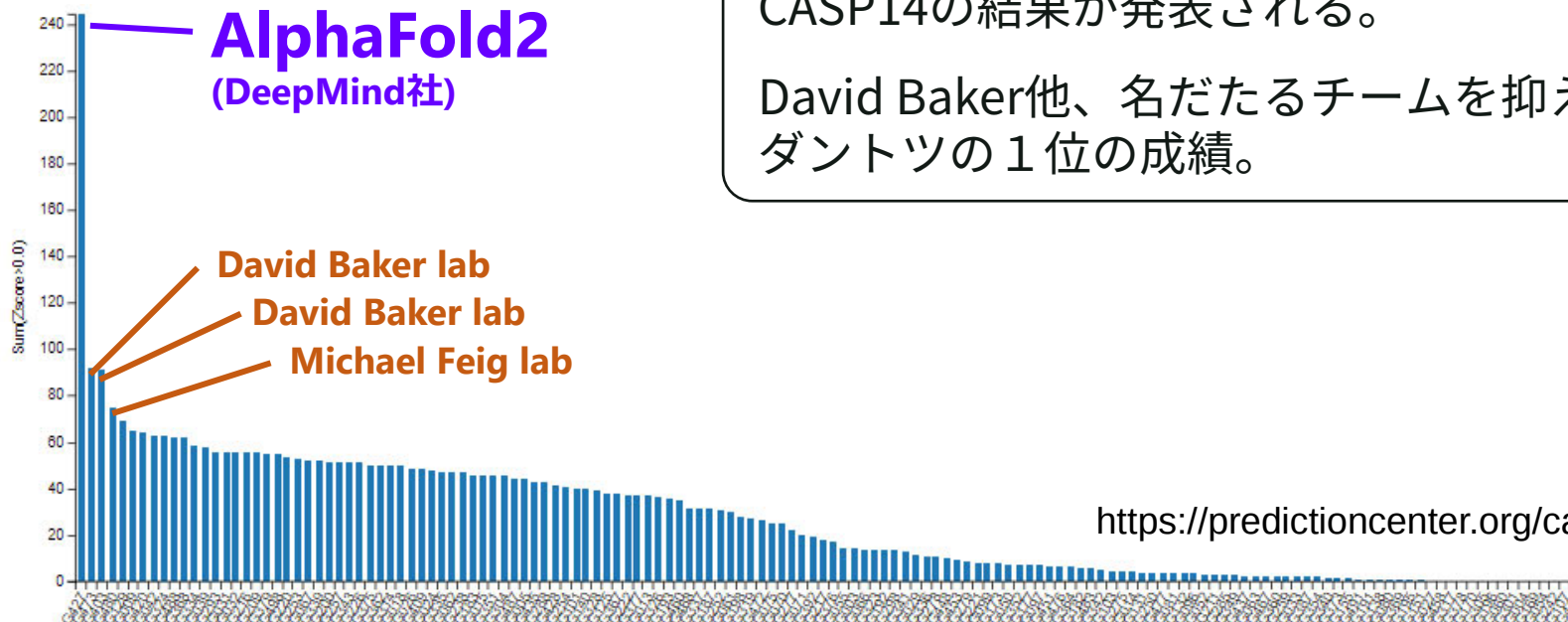
Google’s deep-learning program for determining the 3D shapes of proteins stands to transform biology, say scientists.

‘It will change everything’: DeepMind’s AI makes gigantic leap in solving protein structures  
<https://www.nature.com/articles/d41586-020-03348-4>

2020年11月30日

タンパク質立体構造予測コンペティション  
CASP14の結果が発表される。

David Baker他、名だたるチームを抑えて  
ダントツの1位の成績。



2021年7月16日

AlphaFold2のNature論文公開

Jumper J, et al. *Nature*, 2021.

## Article

# Highly accurate protein structure prediction with AlphaFold

<https://doi.org/10.1038/s41586-021-03819-2>

Received: 11 May 2021

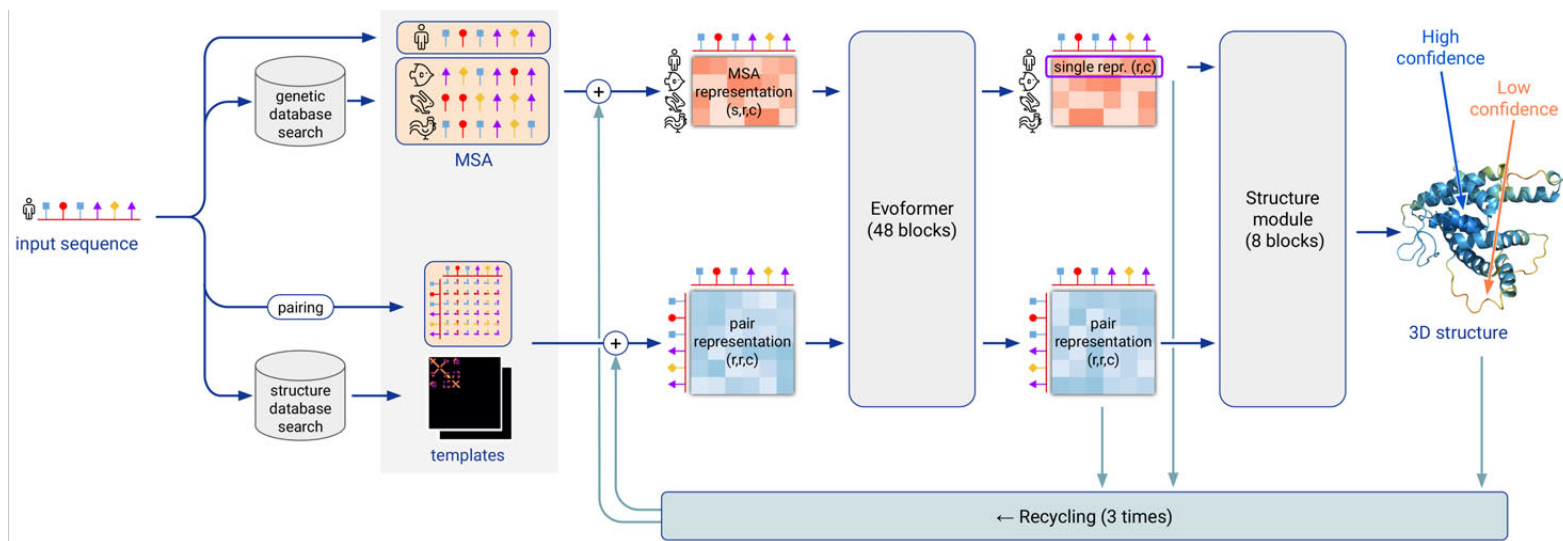
Accepted: 12 July 2021

Published online: 15 July 2021

Open access

 Check for updates

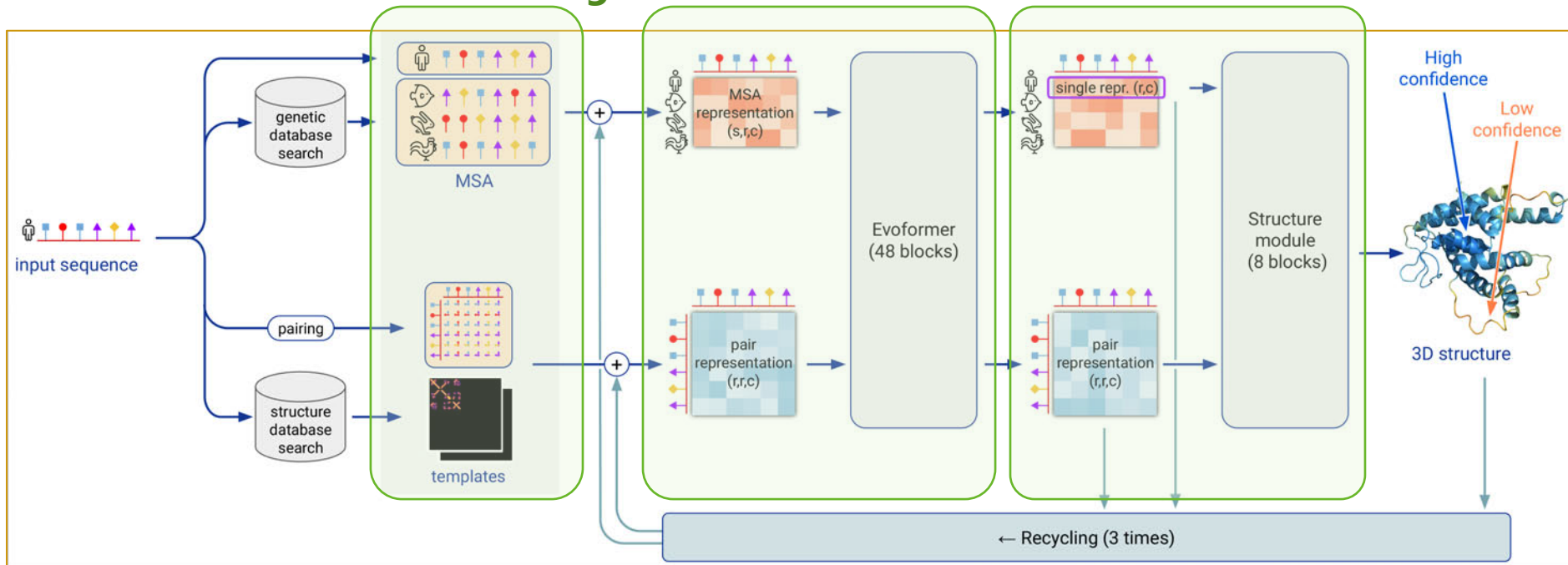
**John Jumper**<sup>1,4</sup>✉, Richard Evans<sup>1,4</sup>, Alexander Pritzel<sup>1,4</sup>, Tim Green<sup>1,4</sup>, Michael Figurnov<sup>1,4</sup>, Olaf Ronneberger<sup>1,4</sup>, Kathryn Tunyasuvunakool<sup>1,4</sup>, Russ Bates<sup>1,4</sup>, Augustin Žídek<sup>1,4</sup>, Anna Potapenko<sup>1,4</sup>, Alex Bridgland<sup>1,4</sup>, Clemens Meyer<sup>1,4</sup>, Simon A. A. Kohl<sup>1,4</sup>, Andrew J. Ballard<sup>1,4</sup>, Andrew Cowie<sup>1,4</sup>, Bernardino Romera-Paredes<sup>1,4</sup>, Stanislav Nikolov<sup>1,4</sup>, Rishub Jain<sup>1,4</sup>, Jonas Adler<sup>1</sup>, Trevor Back<sup>1</sup>, Stig Petersen<sup>1</sup>, David Reiman<sup>1</sup>, Ellen Clancy<sup>1</sup>, Michal Zielinski<sup>1</sup>, Martin Steinegger<sup>2,3</sup>, Michalina Pacholska<sup>1</sup>, Tamas Berghammer<sup>1</sup>, Sebastian Bodenstein<sup>1</sup>, David Silver<sup>1</sup>, Oriol Vinyals<sup>1</sup>, Andrew W. Senior<sup>1</sup>, Koray Kavukcuoglu<sup>1</sup>, Pushmeet Kohli<sup>1</sup> & **Demis Hassabis**<sup>1,4</sup>✉



## 1. Embedding

## 2. Evoformer

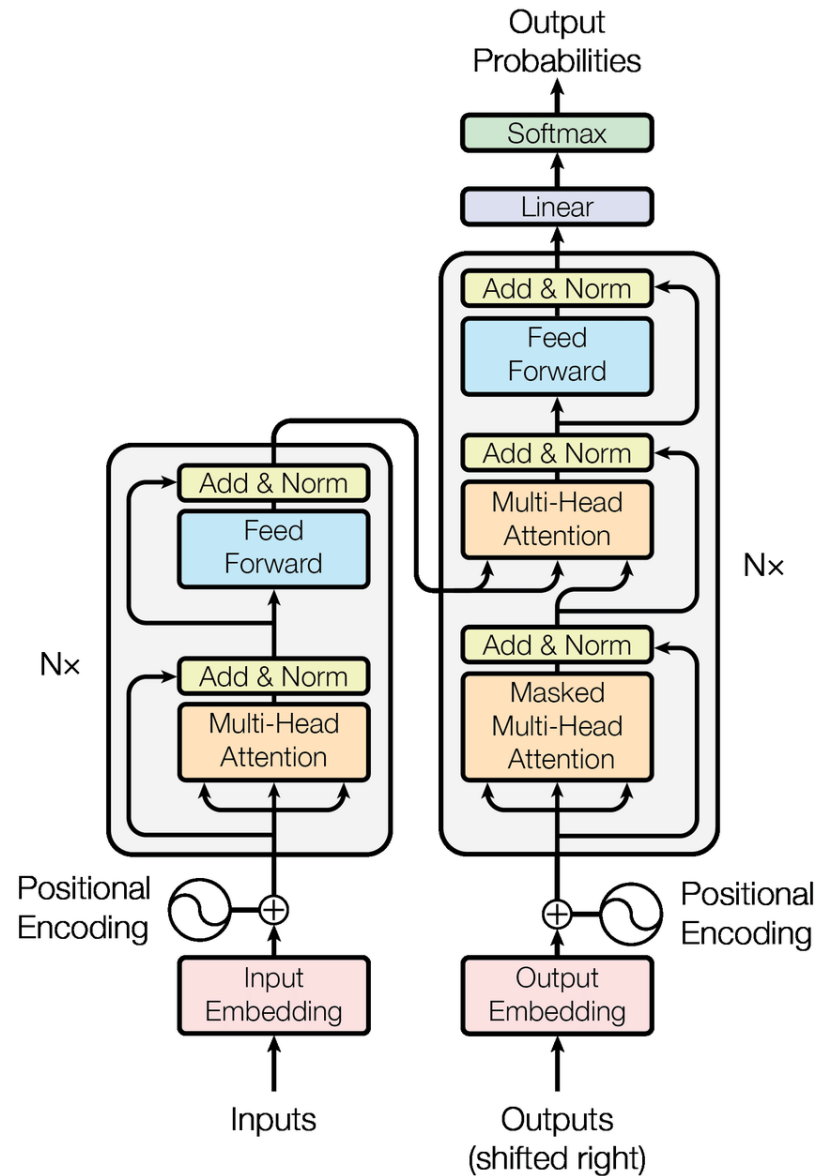
## 3. Structure



## 要はTransformerベースの深層ニューラルネットワークモデル

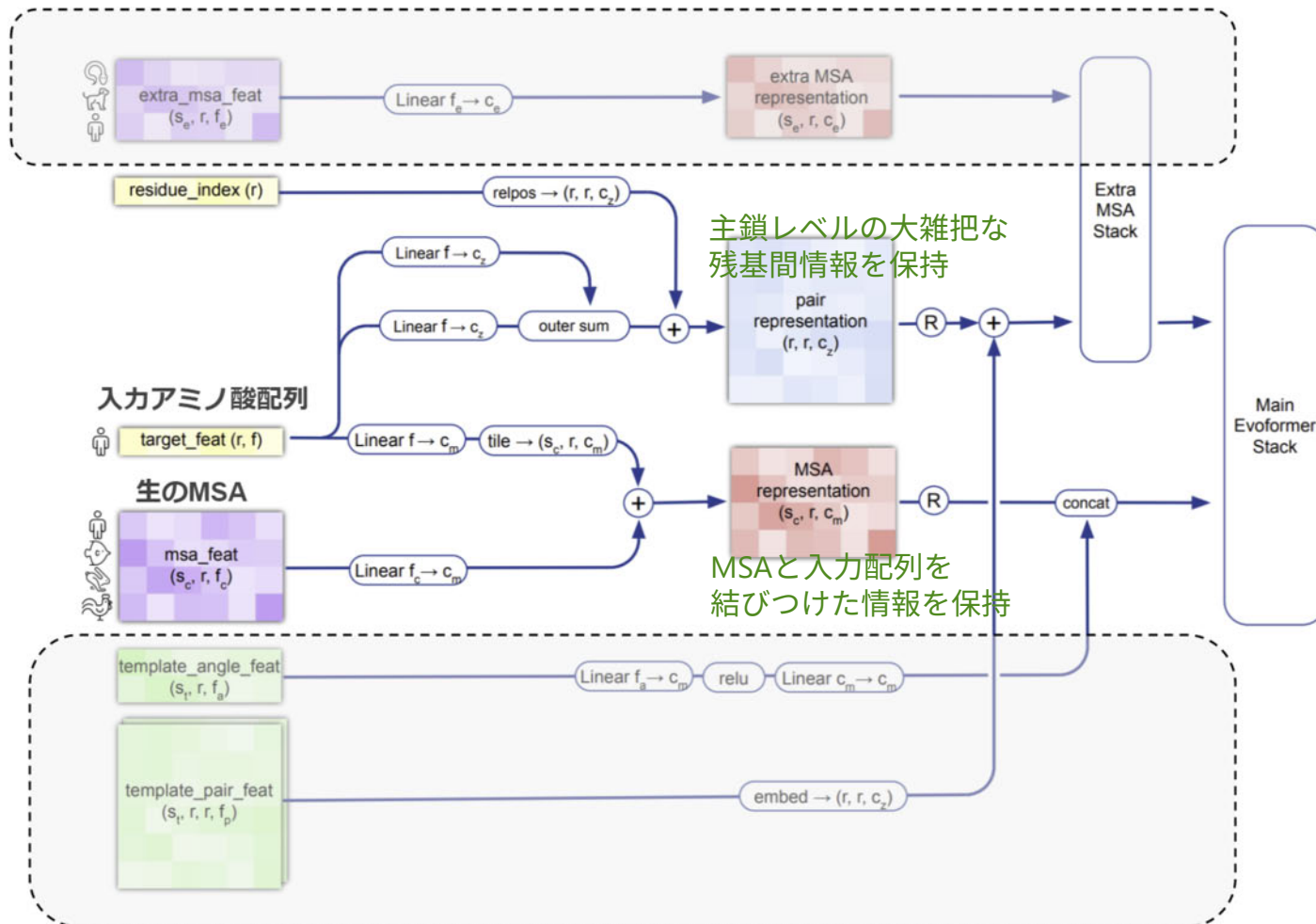
- 入力配列から出力構造までEnd-to-Endで学習
  - AlphaFold1ではコンタクトマップと二面角の予測をNNで行い、Rosettaで最適化
- 最初に多重配列アラインメント (MSA) を構築
  - メタゲノムを含む大規模な配列DBに対して相同性検索
  - Evoformer blockでMSAの改善
- 3つの主要なモジュール：Embedding, Evoformer, Structure
- Recyclingプロセスにより予測構造を最適化

(Vaswani+, 2017)



# AlphaFold2 最初のブロック

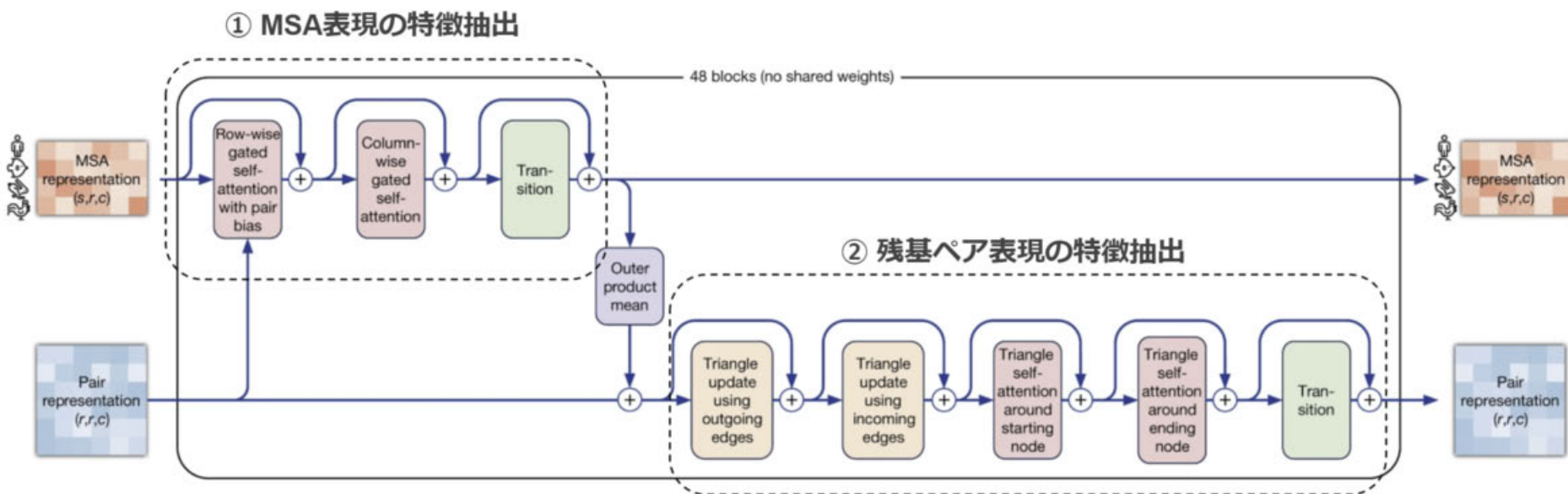
やっていること=入力配列のembedding (denseベクトル化)



(眺めて雰囲気を感じるだけでOKです)

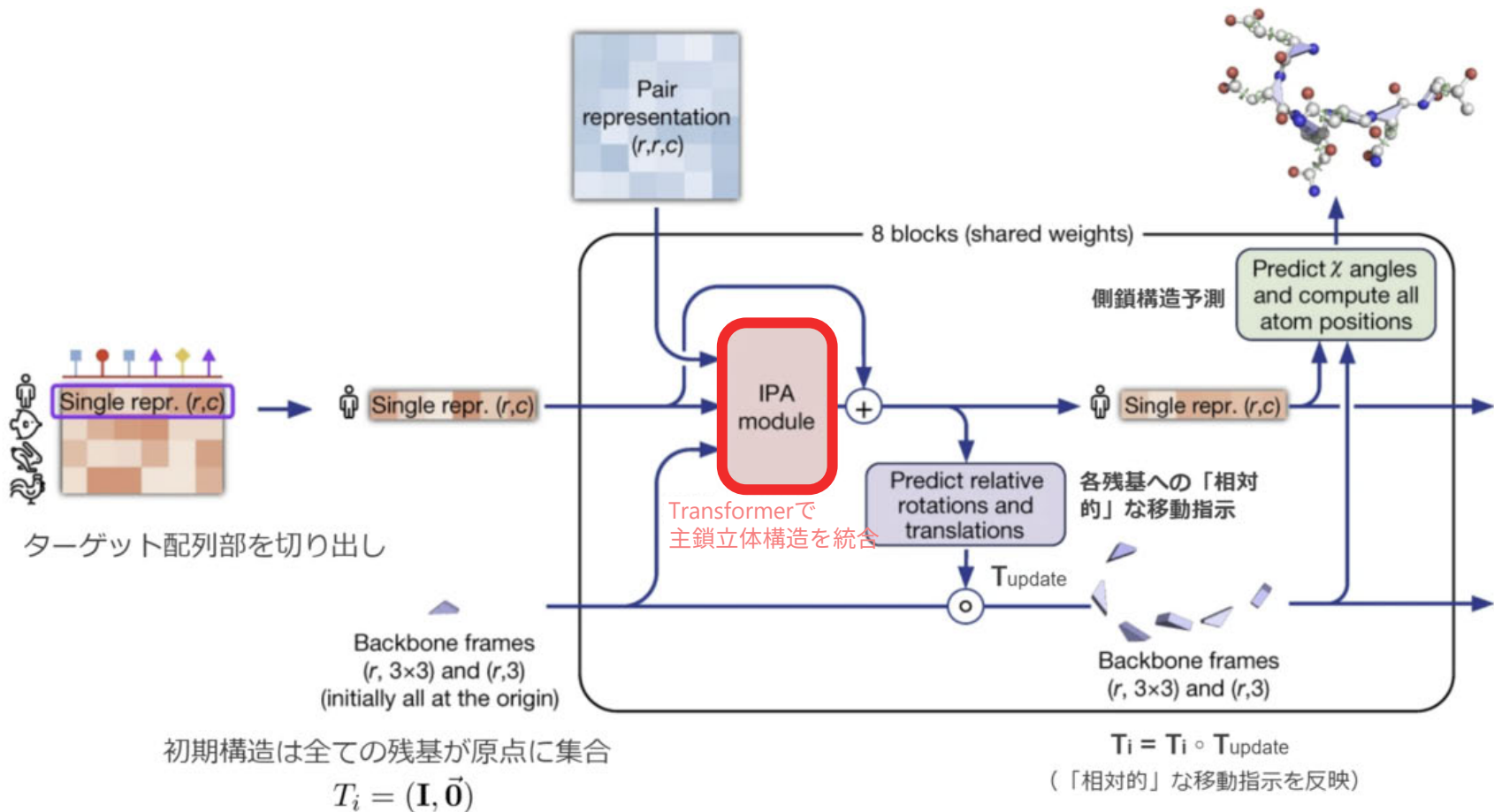
Evoformer = Evolution（分子進化）のためのTransformer

- MSAからの特徴抽出
- Pair representationからの特徴抽出
- MSAとPair representationでの情報交換



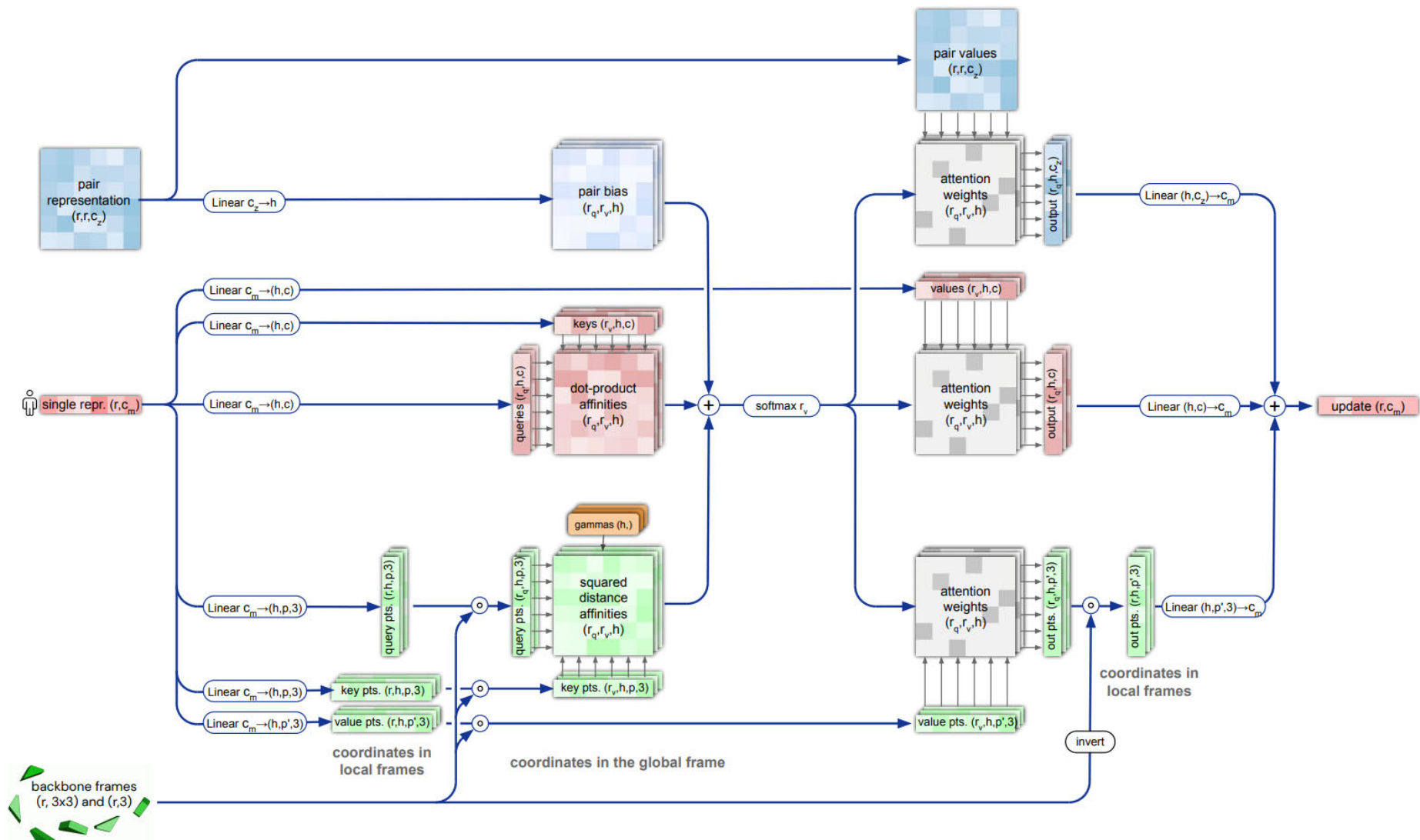
(眺めて雰囲気を感じるだけでOKです)

Evoformerで得た情報をもとに、残基位置を更新



(眺めて雰囲気を感じるだけでOKです)

## Invariant Point Attention (IPA) moduleの構造



(眺めて雰囲気を感じるだけでOKです)

(MIT Sergey Ovchinnikov氏のスライドより一部改変)

<https://t.co/wjnw3uhMcV>

## 共進化残基ペア

予測したいタンパク質のアミノ酸配列

配列同源性検索

BFD, MGnify, UniRefなど  
(メタゲノム配列含む)

配列DB

多重配列アラインメント  
(MSA)

共進化残基ペアは、立体構造上で近くにいるはず  
(Direct Coupling Analysis, DCA)

立体構造

PDB

## AlphaFold2の予測精度を支えるエッセンス

- 共進化残基ペアが立体構造上で近くに位置しやすい (DCA) (2009年)
- MSA構築の際、メタゲノム配列からアセンブルされた身元不明の配列も使うとよい (2017年)
- Transformerをベースとした深層学習により、DCAの考え方を予測モデルとして集成 (2020年)

2021年7月23日

AlphaFold Databaseの公開

Tunyasuvunakool K, *et al. Nature*, 2021.

## Article

# Highly accurate protein structure prediction for the human proteome

<https://doi.org/10.1038/s41586-021-03828-1>

Received: 11 May 2021

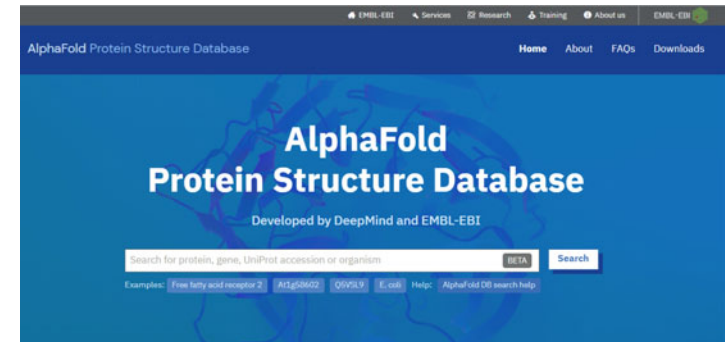
Accepted: 16 July 2021

Published online: 22 July 2021

Open access

Check for updates

Kathryn Tunyasuvunakool<sup>1,2,3</sup>, Jonas Adler<sup>1</sup>, Zachary Wu<sup>1</sup>, Tim Green<sup>1</sup>, Michal Zielinski<sup>1</sup>, Augustin Židek<sup>1</sup>, Alex Bridgland<sup>1</sup>, Andrew Cowie<sup>1</sup>, Clemens Meyer<sup>1</sup>, Agata Laydon<sup>1</sup>, Sameer Velankar<sup>1</sup>, Gerard J. Kleywegt<sup>2</sup>, Alex Bateman<sup>2</sup>, Richard Evans<sup>1</sup>, Alexander Pritzel<sup>1</sup>, Michael Figurnov<sup>1</sup>, Olaf Ronneberger<sup>1</sup>, Russ Bates<sup>1</sup>, Simon A. A. Kohl<sup>1</sup>, Anna Potapenko<sup>1</sup>, Andrew J. Ballard<sup>1</sup>, Bernardino Romera-Paredes<sup>1</sup>, Stanislav Nikolov<sup>1</sup>, Rishub Jain<sup>1</sup>, Ellen Clancy<sup>1</sup>, David Reiman<sup>1</sup>, Stig Petersen<sup>1</sup>, Andrew W. Senior<sup>1</sup>, Koray Kavukcuoglu<sup>1</sup>, Ewan Birney<sup>2</sup>, Pushmeet Kohli<sup>1</sup>, John Jumper<sup>1,3,2,3</sup> & Demis Hassabis<sup>1,3,2,3</sup>

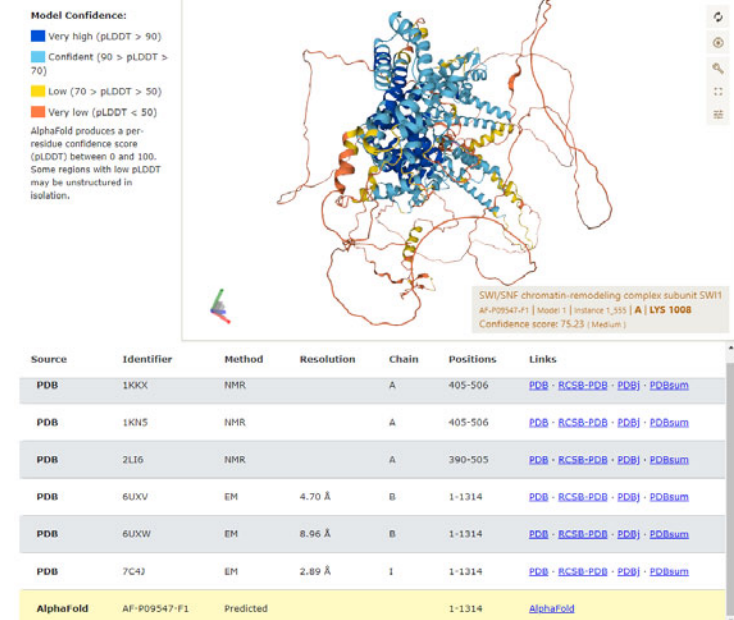


<https://alphafold.ebi.ac.uk/>

| Species                              | Common Name     | Reference Proteome | Predicted Structures | Download           |
|--------------------------------------|-----------------|--------------------|----------------------|--------------------|
| <i>Arabidopsis thaliana</i>          | Arabidopsis     | UP000006548        | 27,434               | Download (3642 MB) |
| <i>Caenorhabditis elegans</i>        | Nematode worm   | UP000001940        | 19,694               | Download (2601 MB) |
| <i>Candida albicans</i>              | C. albicans     | UP000000559        | 5,974                | Download (965 MB)  |
| <i>Danio rerio</i>                   | Zebrafish       | UP000000437        | 24,664               | Download (4141 MB) |
| <i>Dictyostelium discoideum</i>      | Dictyostelium   | UP000002195        | 12,622               | Download (2150 MB) |
| <i>Drosophila melanogaster</i>       | Fruit fly       | UP000000803        | 13,458               | Download (2174 MB) |
| <i>Escherichia coli</i>              | E. coli         | UP000000625        | 4,363                | Download (448 MB)  |
| <i>Glycine max</i>                   | Soybean         | UP000008827        | 55,799               | Download (7142 MB) |
| <i>Homo sapiens</i>                  | Human           | UP000005640        | 23,391               | Download (4784 MB) |
| <i>Leishmania infantum</i>           | L. infantum     | UP000008153        | 7,924                | Download (1481 MB) |
| <i>Methanocaldococcus jannaschii</i> | M. jannaschii   | UP000000805        | 1,773                | Download (171 MB)  |
| <i>Mus musculus</i>                  | Mouse           | UP000000589        | 21,615               | Download (3547 MB) |
| <i>Mycobacterium tuberculosis</i>    | M. tuberculosis | UP000001584        | 3,988                | Download (421 MB)  |
| <i>Oryza sativa</i>                  | Asian rice      | UP000059680        | 43,649               | Download (4416 MB) |
| <i>Plasmodium falciparum</i>         | P. falciparum   | UP000001450        | 5,187                | Download (1132 MB) |
| <i>Rattus norvegicus</i>             | Rat             | UP000002494        | 21,272               | Download (3404 MB) |
| <i>Saccharomyces cerevisiae</i>      | Budding yeast   | UP000002311        | 6,040                | Download (960 MB)  |
| <i>Schizosaccharomyces pombe</i>     | Fission yeast   | UP000002485        | 5,128                | Download (776 MB)  |
| <i>Staphylococcus aureus</i>         | S. aureus       | UP000008816        | 2,888                | Download (268 MB)  |
| <i>Trypanosoma cruzi</i>             | T. cruzi        | UP000002296        | 19,036               | Download (2905 MB) |
| <i>Zea mays</i>                      | Maize           | UP000007305        | 39,299               | Download (5014 MB) |

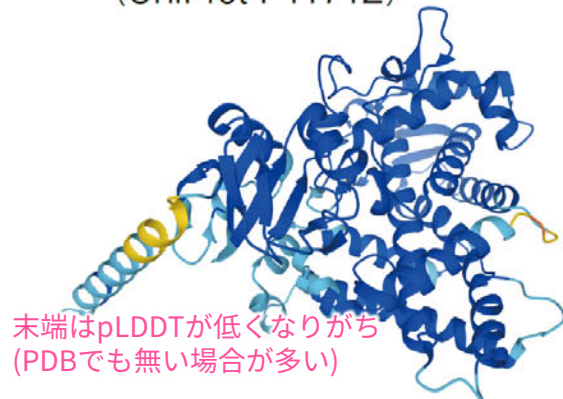
ほぼすべてのタンパク質構造予測結果が収録

## Structure

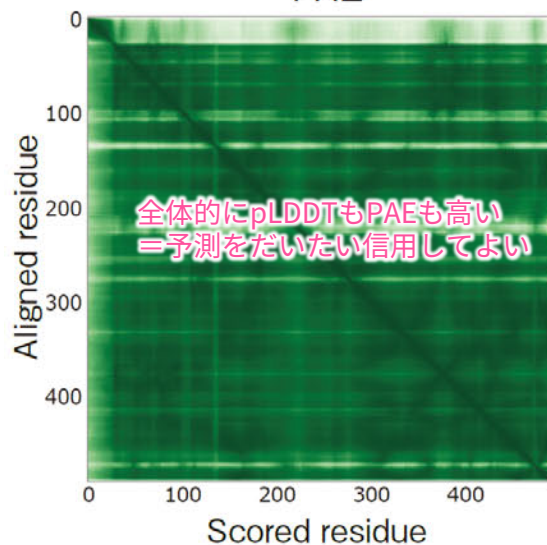


# AlphaFold2はどんな予測を行うのか？

**A** Cytochrome P450 2C9  
(UniProt P11712)



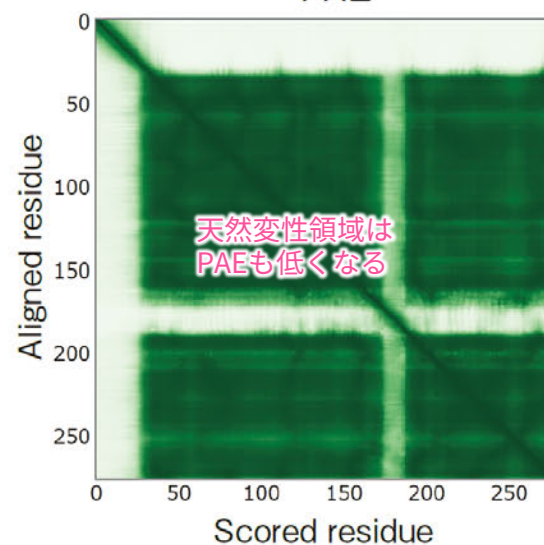
PAE



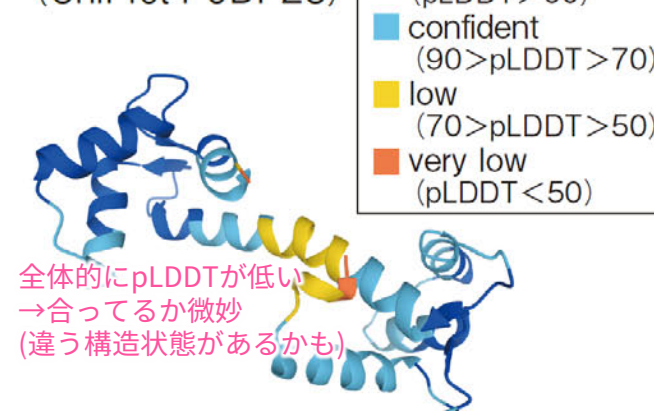
**B** Caspase-3  
(UniProt P42574)



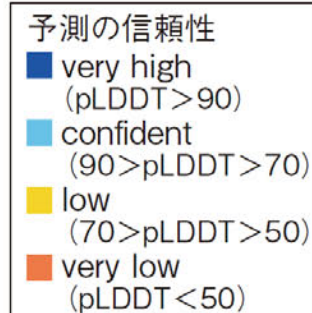
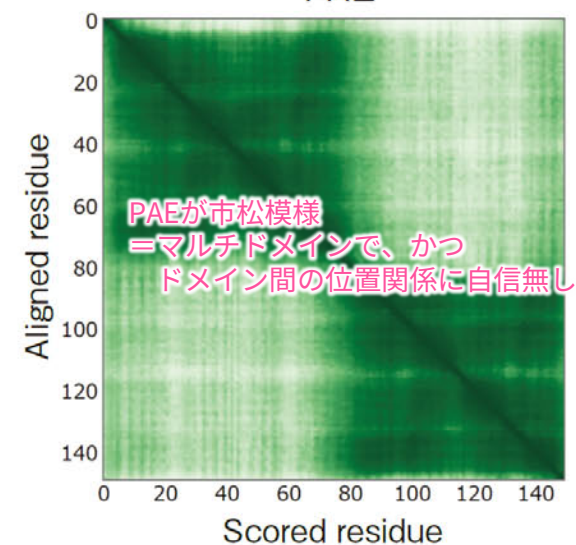
PAE



**C** Calmodulin-1  
(UniProt P0DP23)



PAE



## AlphaFold2関連プレプリント (2021年7月~9月)

New Results **「AF2予測構造をTransformerに突っ込んでDNA結合部位予測」** [Follow this preprint](#)

**AlphaFold-aware prediction of protein-DNA binding sites using graph transformer**

Qianmu Yuan, Sheng Chen, Jiahua Rao, Shuangjia Zheng, Huiying Zhao, Yuedong Yang

doi: <https://doi.org/10.1101/2021.08.25.457661> **August 26, 2021.**

New Results **「AF2複合体構造予測を改良」** [Follow this preprint](#)

**Improved Docking of Protein Models by a Combination of Alphafold2 and ClusPro**

Usman Ghani, Israel Desta, Akhil Jindal, Omeir Khan, George Jones, Sergey Kotelnikov, Dzimtry Padhorny, Sandor Vajda, Dima Kozakov

doi: <https://doi.org/10.1101/2021.09.07.459290> **September 7, 2021.**

New Results **「AF2複合体構造予測を改良」** [Follow this preprint](#)

**Improved prediction of protein-protein interactions using AlphaFold2 and extended multiple-sequence alignments**

P. Bryant, G. Pozzati, A. Elofsson

doi: <https://doi.org/10.1101/2021.09.15.460468> **September 15, 2021.**

New Results **「AF2タンパク質-ペプチド複合体構造予測のベンチマーキング」** [Follow this preprint](#)

**Can AlphaFold2 predict protein-peptide complex structures accurately?**

Junsu Ko, Juyong Lee

doi: <https://doi.org/10.1101/2021.07.27.453972> **July 28, 2021.**

New Results **「AF2をいろんなタスクに使ったら良い感じだった」** [Follow this preprint](#)

**A structural biology community assessment of AlphaFold 2 applications**

Mehmet Akdel, Douglas EV Pires, Eduard Porta-Pardo, Jurgen Janes, Arthur O Zalevsky, Balint Meszaros, Patrick Bryant, Lydia L Good, Roman A Laskowski, Gabriele Pozzati, Aditi Shenoy, Wensi Zhu, Petras Kundrotas, Victoria Ruiz-Serra, Carlos HM Rodrigues, Alistair S Dunham, David Burke, Neera Borkakoti, Sameer Velankar, Adam Frost, Kresten Lindorff-Larsen, Alfonso Valencia, Sergey Ovchinnikov, Janani Durairaj, David B Ascher, Janet M Thornton, Norman E Davey, Amelie Stein, Arne Elofsson, Tristan I Croll, Pedro Beltrao

doi: <https://doi.org/10.1101/2021.09.26.461876> **September 26, 2021.**

New Results **「AF2で主鎖構造のde novoデザイン」** [Follow this preprint](#)

**Using AlphaFold for Rapid and Accurate Fixed Backbone Protein Design**

Lewis Moffat, Joe G. Greener, David T. Jones

doi: <https://doi.org/10.1101/2021.08.24.457549> **August 26, 2021.**

New Results **「AF2は点変異の安定性( $\Delta\Delta G$ )予測は無理っぽい」** [Follow this preprint](#)

**Using AlphaFold to predict the impact of single mutations on protein stability and function**

Marina A. Pak, Karina A. Markhieva, Mariia S. Novikova, Dmitry S. Petrov, Ilya S. Vorobyev, Ekaterina S. Maksimova, Fyodor A. Kondrashov, Dmitry N. Ivankov

doi: <https://doi.org/10.1101/2021.09.19.460937> **September 20, 2021.**

New Results **「予測構造はフォールディングパスウェイをあんまり埋められてないよ」** [Follow this preprint](#)

**Current protein structure predictors do not produce meaningful folding pathways**

Carlos Outeiral, Daniel A. Nissley, Charlotte M. Deane

doi: <https://doi.org/10.1101/2021.09.20.461137> **September 20, 2021.**

## AlphaFold2によるタンパク質複合体構造予測



@inferist

てか複合体作ると分かってるタンパク質ペアを適当にリンカー配列で繋げてあるはほうど式に放り込んだら複合体の構造が出てくるのだろうか

(預言者) 2020-12-01 2:07:10



**Yoshitaka Moriwaki** @Ag\_smith · Jul 19

AlphaFold2 can also predict heterocomplexes. All you have to do is input the two sequences you want to predict and connect them with a long linker.

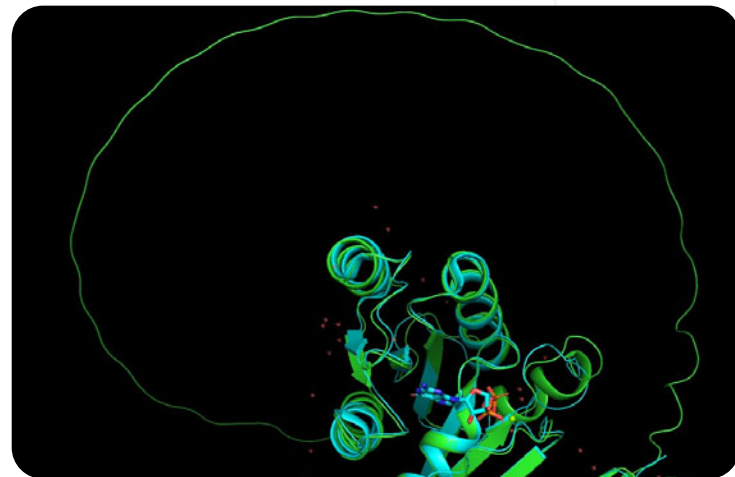
(その後AlphaFold-Multimerとして  
本家に取り込まれ、AlphaFold3論文でpost引用)



Many single protein chains can be accurately predicted using AlphaFold, but multi-chain protein complexes remain challenging. Introducing AlphaFold-Multimer, a model that can predict the structure of multi-chain protein complexes with increased accuracy [dpmd.ai/alphafold\\_multimer/](https://dpmd.ai/alphafold_multimer/)



午後7:31 · 2021年10月5日



- Moriwaki, Y. Post on X. X [https://x.com/Ag\\_smith/status/1417063635000598528?lang=en-GB](https://x.com/Ag_smith/status/1417063635000598528?lang=en-GB) (2021).
- Baek, M. Post on X. X <https://x.com/minkbaek/status/1417538291709071362?lang=en> (2021).

Twitter (現X) 等のSNSの活用、オープンサイエンスの流れが、以前よりかなり加速したように思われる。

## • ColabFold

(Mirdita, *et al.* Nat Methods 2023)

AlphaFold2をGoogle Colabで  
気軽に利用できる  
(3クリックで構造予測！)



OPEN

## ColabFold: making protein folding accessible to all

Milot Mirdita<sup>1,10</sup>, Konstantin Schütze<sup>2</sup>, Yoshitaka Moriwaki<sup>3,4</sup>, Lim Heo<sup>5</sup>,  
Sergey Ovchinnikov<sup>6,7,10</sup> and Martin Steinegger<sup>2,8,9,10</sup>

東京科学大学 総合研究院 難治疾患研究所 森脇由隆先生も貢献！

AlphaFold2.ipynb

ファイル 編集 表示 挿入 ランタイム ツール ヘルプ

+ コード + テキスト ドライブにコピー

接続 T4 ハイメモリ + Gemini

ColabFold v1.5.5: AlphaFold2 using MMseqs2

Easy to use protein structure and complex prediction using [AlphaFold2](#) and [Alphafold2-multimer](#). Sequence alignments/templates are generated through [MMseqs2](#) and [HHsearch](#). For more details, see [bottom](#) of the notebook, checkout the [ColabFold GitHub](#) and read our manuscript. Old versions: [v1.4](#), [v1.5.1](#), [v1.5.2](#), [v1.5.3-patch](#)

[Mirdita M, Schütze K, Moriwaki Y, Heo L, Ovchinnikov S, Steinegger M. ColabFold: Making protein folding accessible to all. Nature Methods. 2022](#)

Input protein sequence(s), then hit Runtime -> Run all

query\_sequence: PIAQIHILEGRSDEQKETLIREVSEAIRSLDAPLTSVRVIITEMAKGHFGIGGELASK

Use : to specify inter-protein chainbreaks for **modeling complexes** (supports homo- and hetro-oligomers). For example **PI...SK:PI...SK** for a homodimer

jobname: test

num\_relax: 0

specify how many of the top ranked structures to relax using amber

template\_mode: none

none = no template information is used. pdb100 = detect templates in pdb100 (see [notes](#)). custom - upload and search own templates (PDB or mmCIF format, see [notes](#))

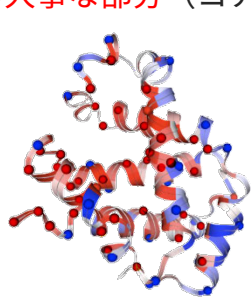
コードの表示

**MACHINE LEARNING** (Cheng, *et al. Science* 2023)

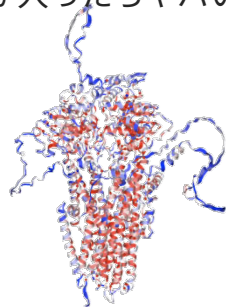
## Accurate proteome-wide missense variant effect prediction with AlphaMissense

Jun Cheng\*, Guido Novati, Joshua Pan†, Clare Bycroft†, Akvilė Žemgulytė†, Taylor Applebaum†, Alexander Pritzel, Lai Hong Wong, Michal Zielinski, Tobias Sargeant, Rosalia G. Schneider, Andrew W. Senior, John Jumper, Demis Hassabis, Pushmeet Kohli\*, Žiga Avsec\*

「構造の**大事な部分**（コア）に変異が入ったらヤバイ」→AlphaFold2っぽく予測



Hemoglobin subunit beta (HBB)

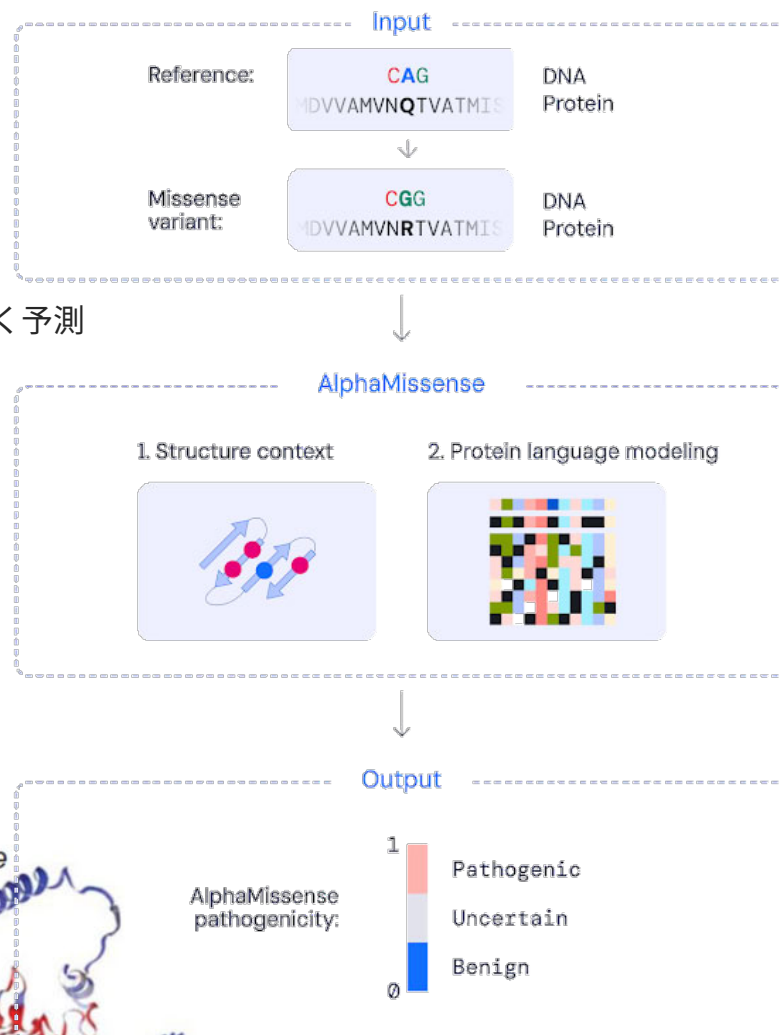
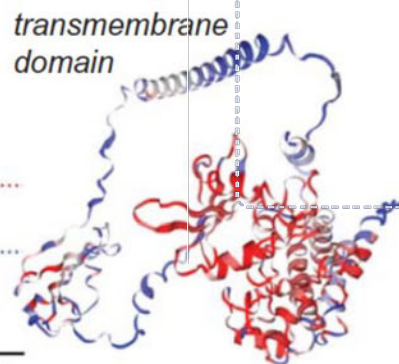
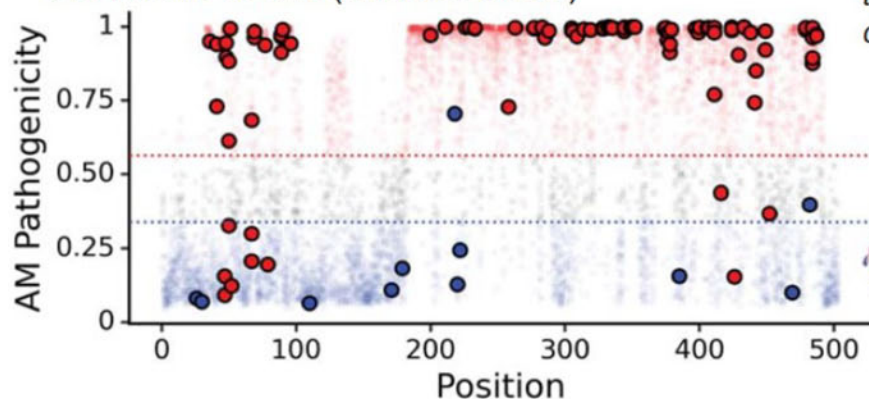


Cystic fibrosis transmembrane conductance regulator (CFTR)

## ACVRL1: Activin A Receptor Like Type 1

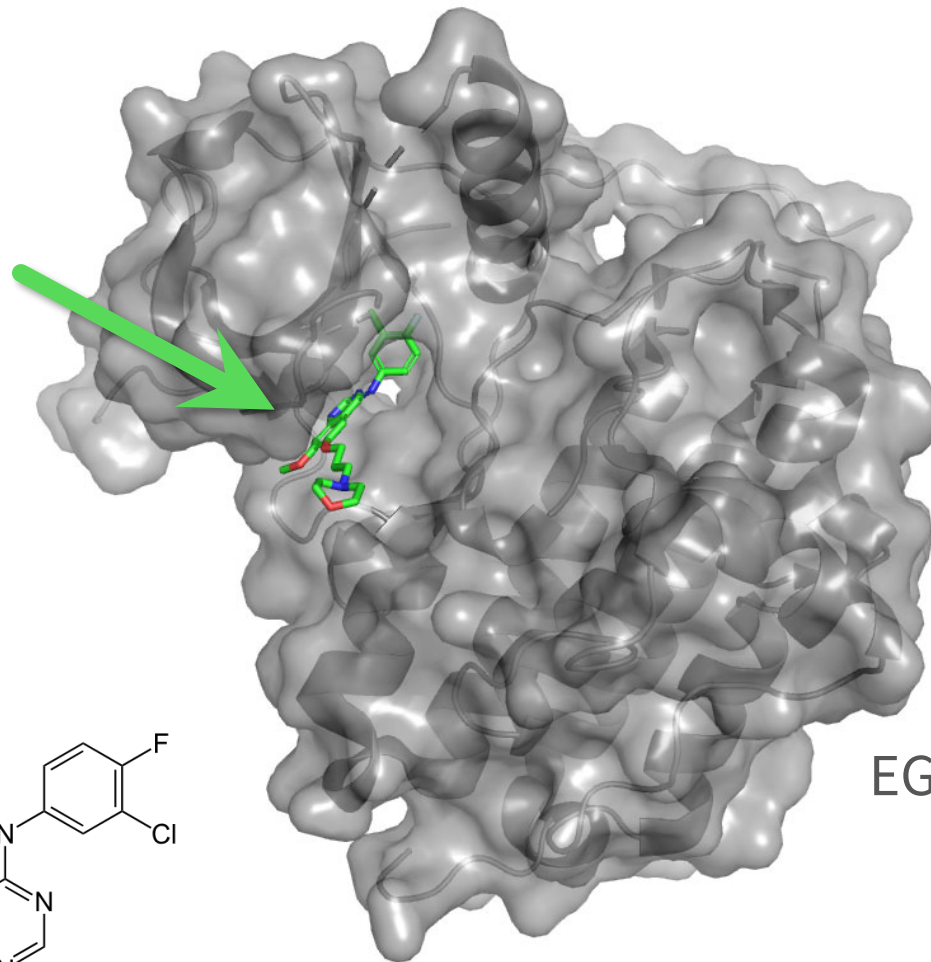
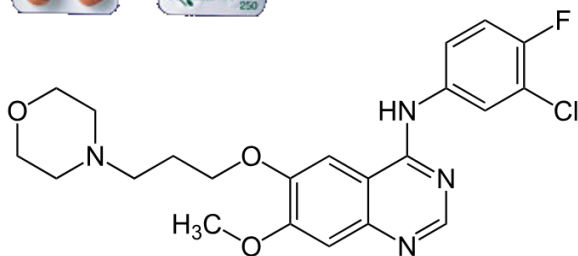
Hereditary hemorrhagic telangiectasia (HHT, 遺伝性出血性末梢血管拡張症, オスラー病)

ACVRL1/P37023 (auROC=0.965)

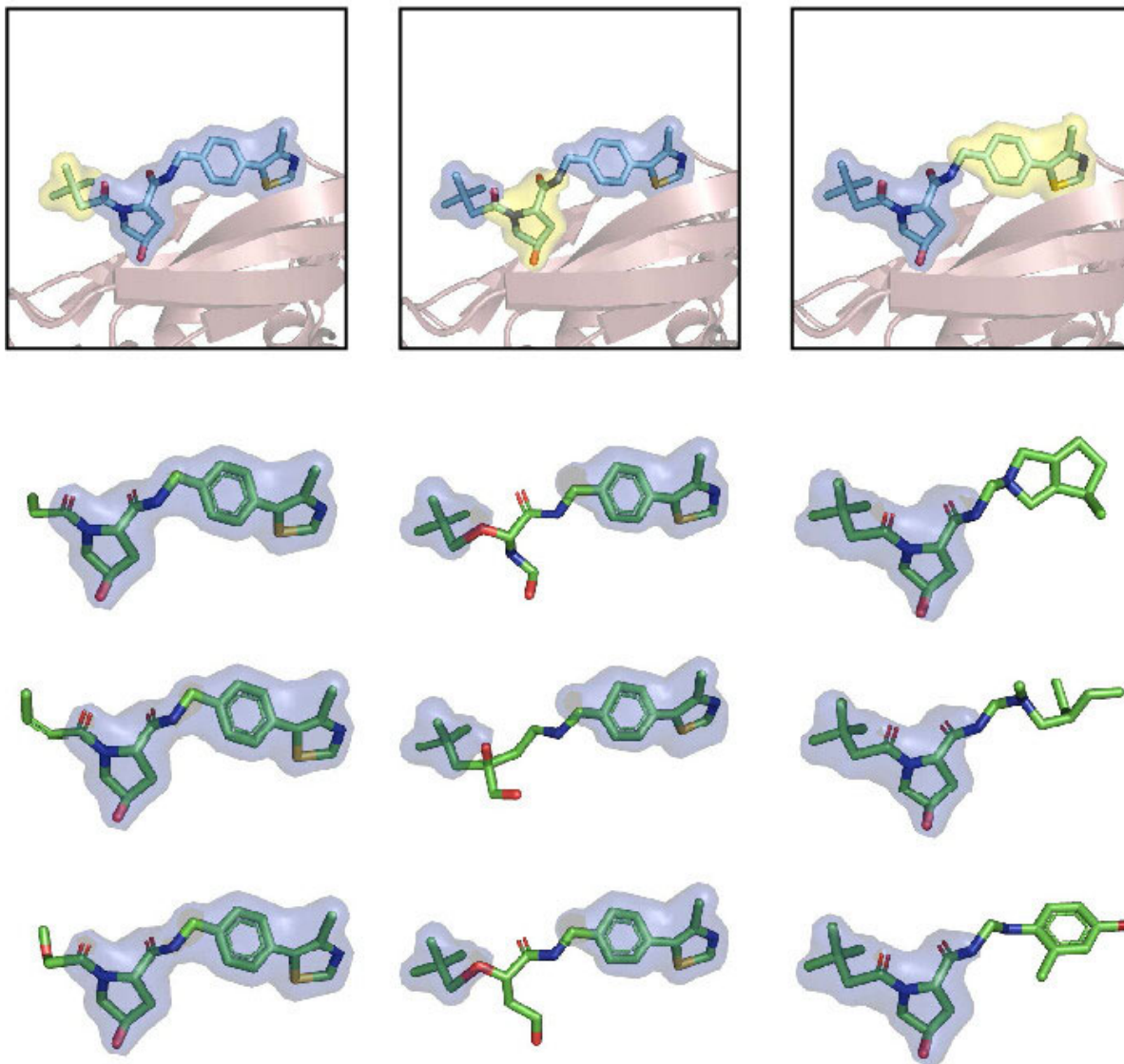


(例) 非小細胞肺がんの治療薬イレッサ® (Gefitinib)  
がん細胞の上皮成長因子受容体(EGFR)タンパク質が伝達しているシグナルを遮断することで、  
腫瘍の増殖を止め、アポトーシス(細胞死)を誘導する。

Gefitinib  
(イレッサ®)



EGFRタンパク質





eLife, 11:e75751, 2022.

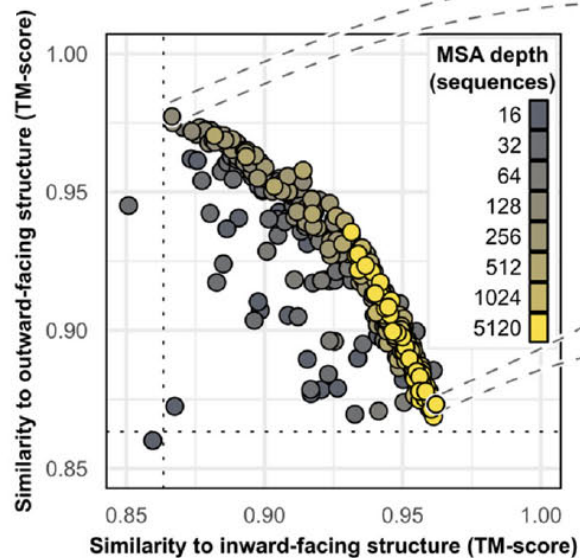
SHORT REPORT



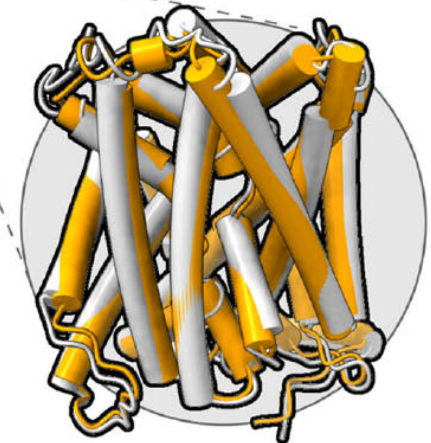
## Sampling alternative conformational states of transporters and receptors with AlphaFold2

Diego del Alamo<sup>1,2†</sup>, Davide Sala<sup>3†</sup>, Hassane S Mchaourab<sup>1\*</sup>, Jens Meiler<sup>2,3\*</sup>

「MSAの配列数を変える（減らす）ことで、構造を無理やり変える」



LAT1, inward-facing (PDB: 6IRSb)



LAT1, outward-facing (PDB: 7DSQb)

# AlphaFold2の予測構造をいろいろ生成する

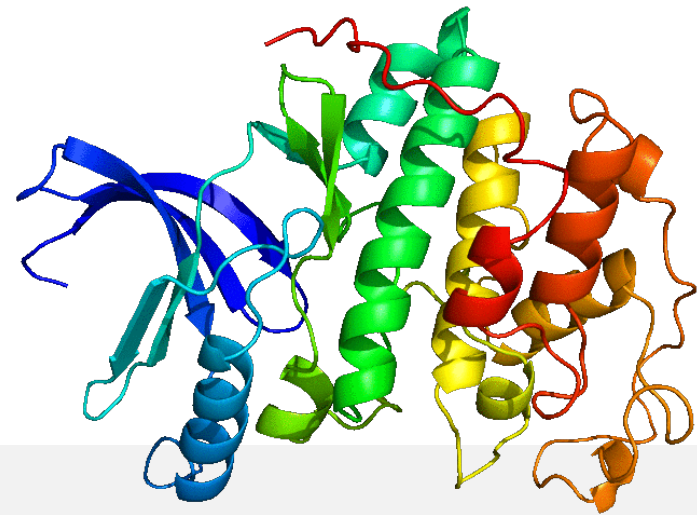
## 構造ベーススクリーニング技術の高度化

AlphaFold2に、構造ベーススクリーニングに向いている構造を出させる

アミノ酸配列

...ACDEFGHI...

AlphaFold2



変えられるパラメータ

- 乱数シード
- 多重配列アラインメントの本数
- 多重配列アラインメントの改変 (Ala置換)
- recyclingの回数

全部やるのは大変  
→ **目標**に向かって推移させる  
(遺伝的アルゴリズム等で)

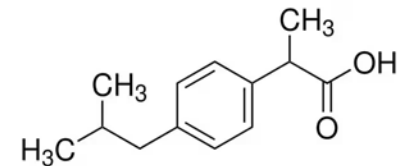
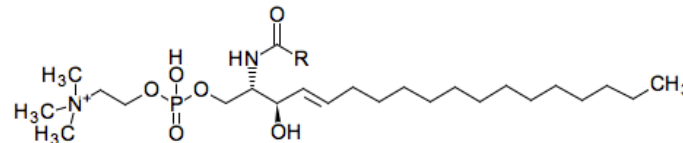
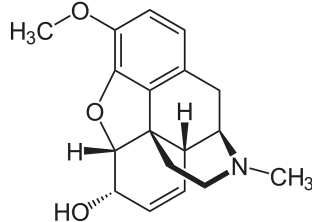
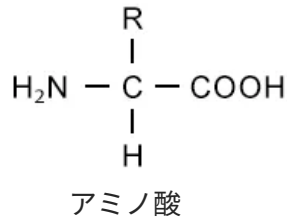
### 「目標」の例

ドッキングの成績が良くなるようにする  
特定の構造に近くなる (RMSDが小さくなる) ようにする  
未知の構造になる (RMSDが大きくなる) ようにする、etc.

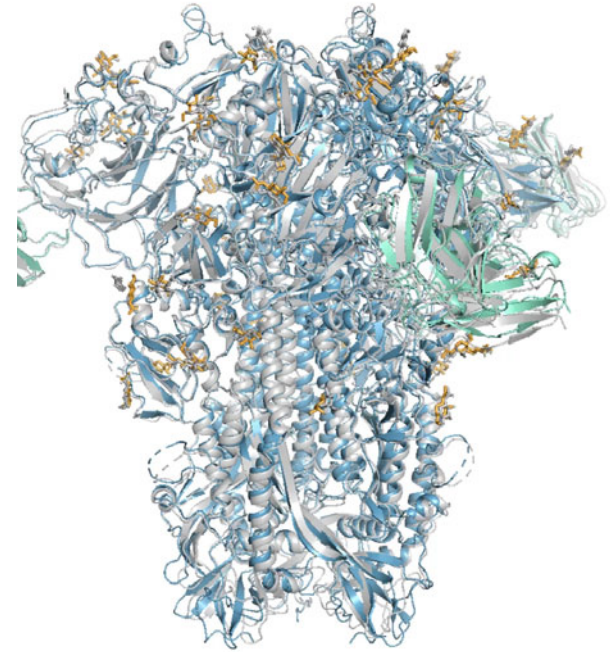
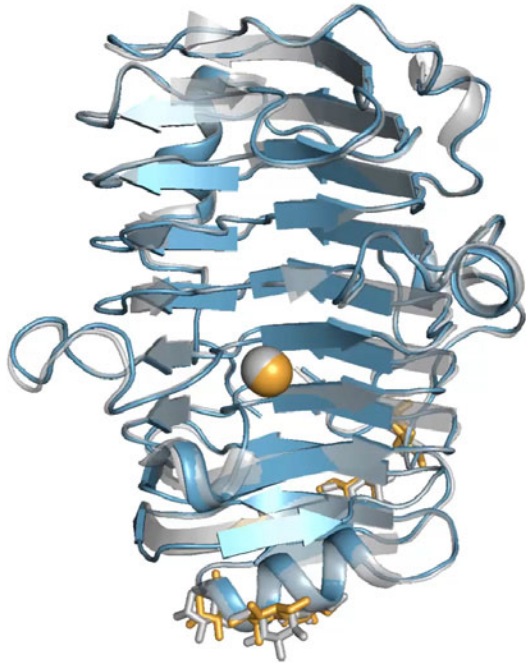
## • AlphaFold2は多くの研究で日常的に利用

- RoseTTAFold (David Baker) など、類似技術も多く発表される。
- ESMFold等のタンパク質言語モデルに基づく構造予測も。  
(だが、計算は速いが精度はAlphaFold2に勝てず)
- なんで「2」？ → 「1」があった (Nature 2019)
  - あまり流行らなかったが、インパクトはそれなりに大きかった
  - Jumper, HassabisはAlphaFold1からコアメンバー

## • 標準アミノ酸(20種類)以外の分子は入力/予測できない



→ AlphaFold3で解決へ



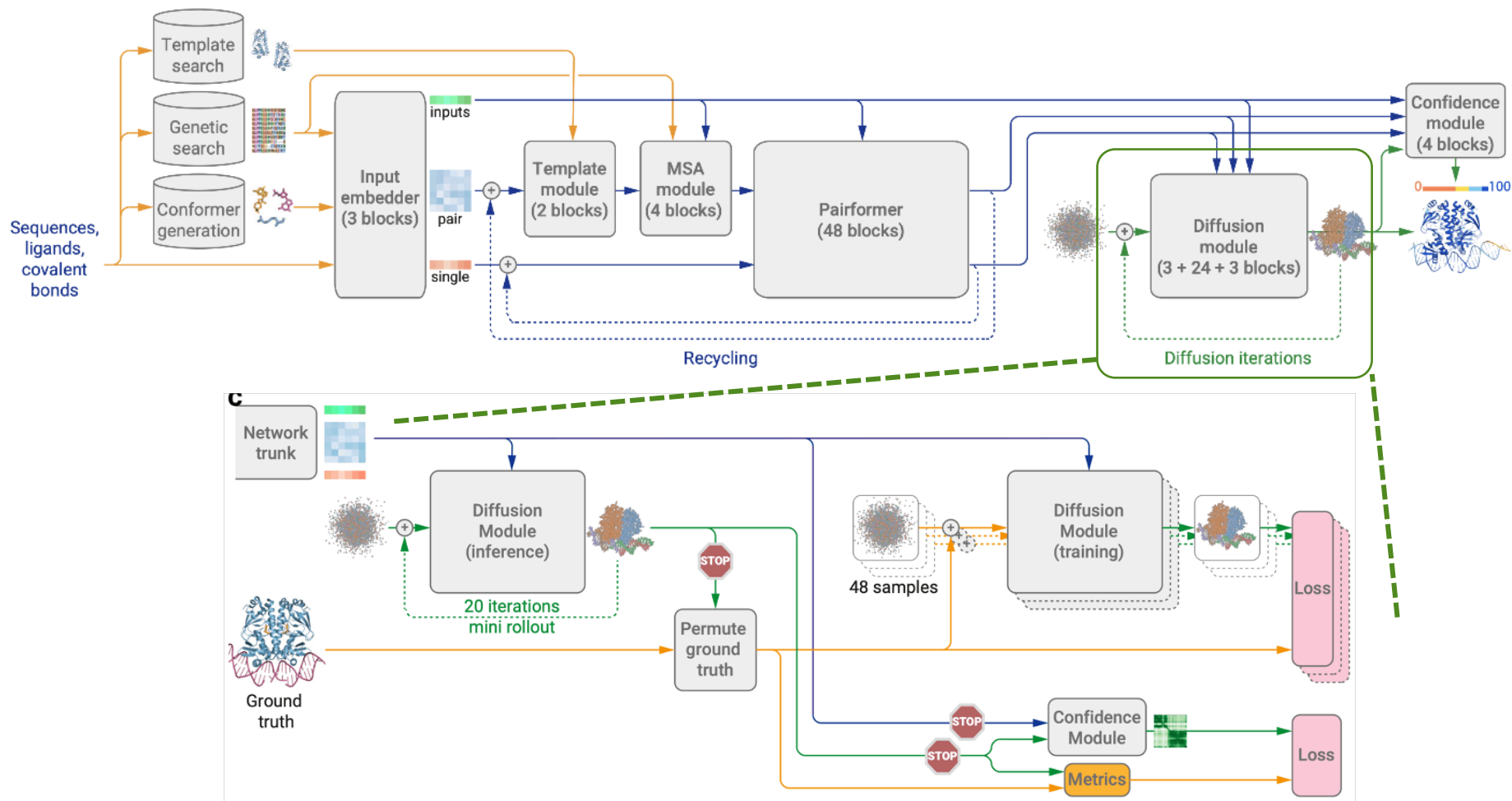
Abramson J, et al. (DeepMind, Isomorphic Labs, Stanford Univ, Princeton Univ)  
**Accurate structure prediction of biomolecular interactions with AlphaFold 3**, *Nature* 630, 493-500, 2024.

## 「AlphaFold流」のヘテロ複合体構造予測AIの歴史

- 核酸** • (2022.9.10) David Bakerら、RoseTTAFold NAをbioRxivに発表
- なんでも** • (2023.10.1) David Bakerら、RoseTTAFold All-AtomをbioRxivに発表
- なんでも** • (2023.10.30) DeepMind社、“AlphaFold-latest” white paperを公開
- 低分子** • (2023.11.5) Patrick Bryantら (ベルリン自由大学)、UmolをbioRxivに発表
- 核酸** • (2023.11.23) RoseTTAFold NA、Nature Methods掲載
- なんでも** • (2024.5.7) RoseTTAFold All-Atom、Science掲載
- なんでも** • (2024.5.8) AlphaFold3、Nature掲載 商用不可・サーバー版のみ
- 低分子** • (2024.5.28) Umol、Nature Communications掲載
- なんでも** • (2024.8.30) Baidu社、HelixFold3をarXivに発表 商用不可
- なんでも** • (2024.9.5) Ligo Biosciences社、AlphaFold3 cloneを公開(不完全) 商用不可
- なんでも** • (2024.9.16) Chai Discovery社、Chai-1をGitHubに公開 商用可
- なんでも** • (2024.11.9) ByteDance社、ProtenixをGitHubに公開 商用可
- なんでも** • (2024.11.11) AlphaFold3、コード公開 商用不可
- なんでも** • (2024.11.18) Jeremy Wohlwendら (MIT)、Boltz-1をGitHubに公開 商用可
- OpenFoldコンソーシアムが完全オープンソース・商用可のOpenFold3を作るはず  
<https://openfold.io/>

DeepMind/Isomorphic LabsによるAlphaFold新バージョン  
タンパク質と低分子/核酸/金属原子等との複合体予測に対応。

Abramson J, et al. *Nature*, 2024.



基本的なアイデアは変わっていない。構造生成に**拡散モデル**を利用。

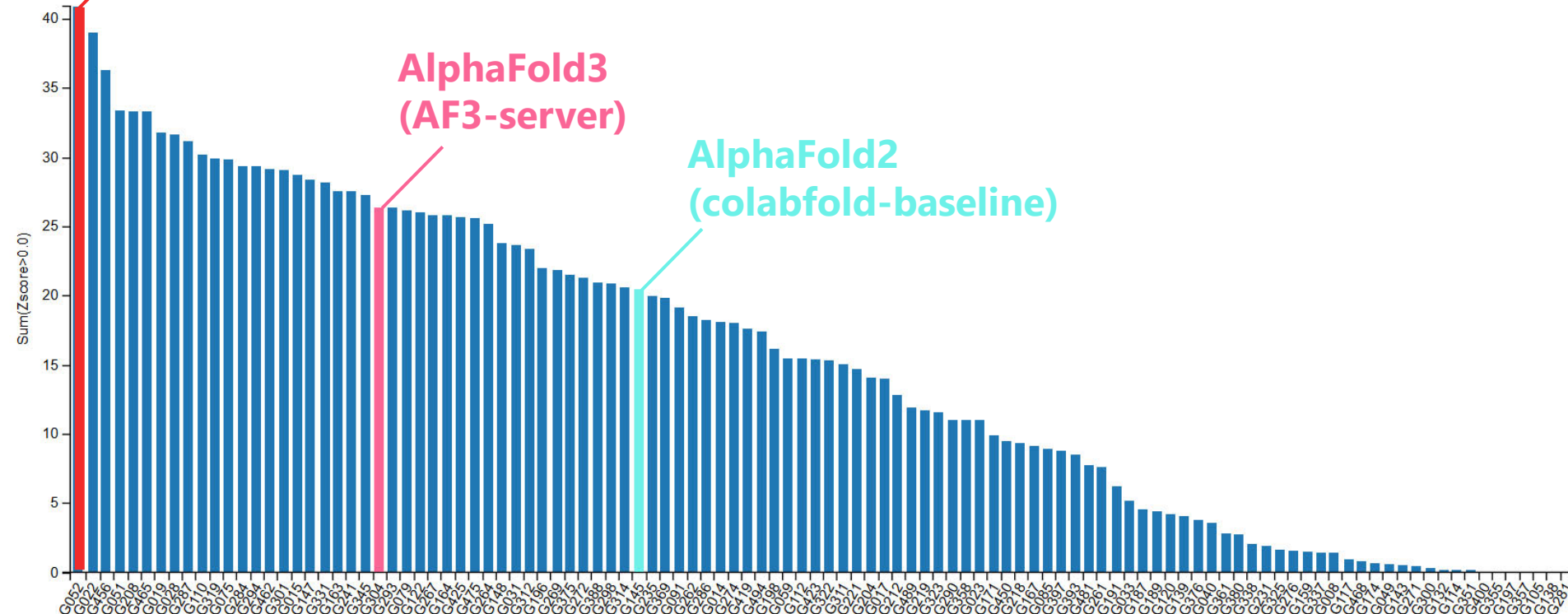
## CASP16 (2024年) の結果 – Protein Domainsカテゴリ

<https://predictioncenter.org/casp16/index.cgi>

**Jianyi Yang**ら (regular targetでCASP15, 16 連続1位)  
(AlphaFold2/3利用+色々工夫)

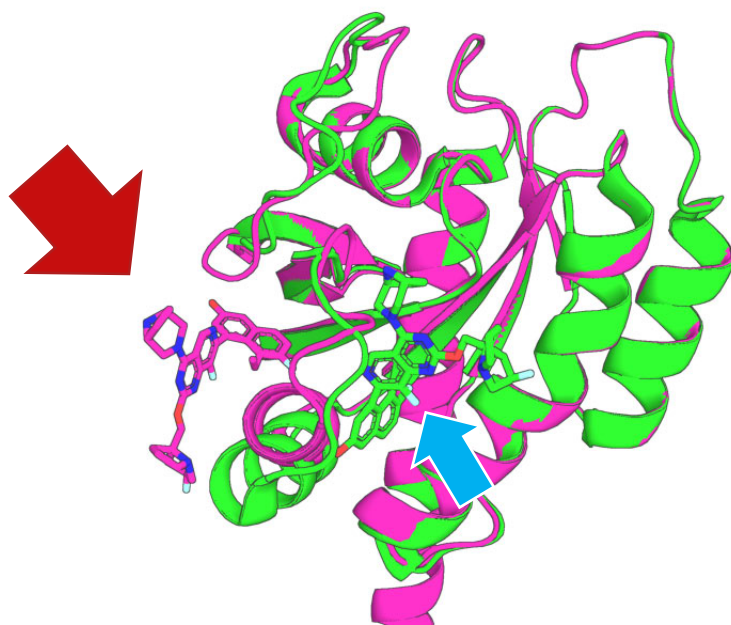
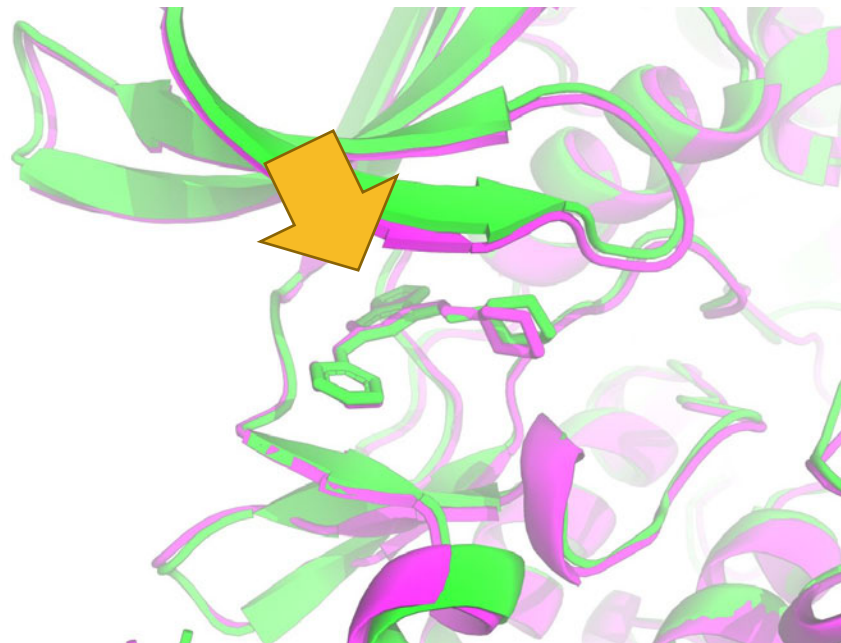
AlphaFold3  
(AF3-server)

AlphaFold2  
(colabfold-baseline)



- 正解構造(PDB)
- AlphaFold3予測構造

バッチリ当たることもある



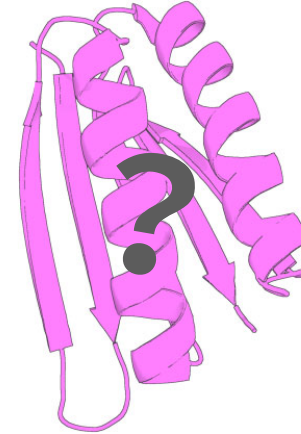
外れることもある.....

**使い物になるか、評価はこれから**

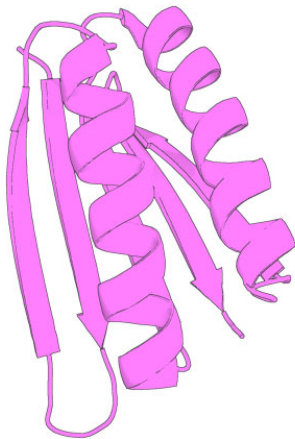
(AlphaFold3は昨年11月11日ようやく学術利用で自由に使えるようになった)

## タンパク質立体構造予測

DIQVQVNIDDNGKNFDYTYT VTTESEL  
QKVLNELMDYIKKQGAKRVRISITART  
KKEAEKFAAILIKVFAELGYNDINVT  
FDGDTVTVEGQLEGGSL



## タンパク質デザイン



DIQVQVNIDDNGKNFDYTYT VTTESEL  
QKVLNELMDYIKKQGAKRVRISITART  
KKEAEKFAAILIKVFAELGYNDINVT  
FDGDTVTVEGQLEGGSL

- アミノ酸は20種類
- タンパク質は数百個のアミノ酸(40~2000超)

アミノ酸の選び方 (200アミノ酸として)

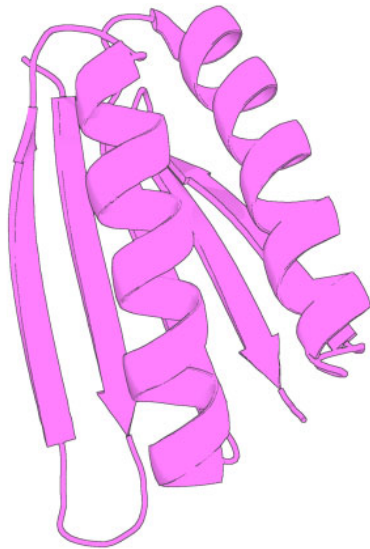
$$20^{200} \doteq 1.6 \times 10^{260}$$

DIQVQVNIDDNGKNFDYTYTVTTESEL  
QKVLNELMDYIKKQGAKRVRISITART  
KKEAEKFAAILIKVFAELGYNDINVT  
DGDTVTVEGQLEGGSL

宇宙の原子の数 →  $10^{80}$  個  
将棋の手の数 →  $10^{220}$  通り

## 人工タンパク質 Top7 のデザイン

(Bakerら, *Science* 2003)



### Design of a Novel Globular Protein Fold with Atomic-Level Accuracy

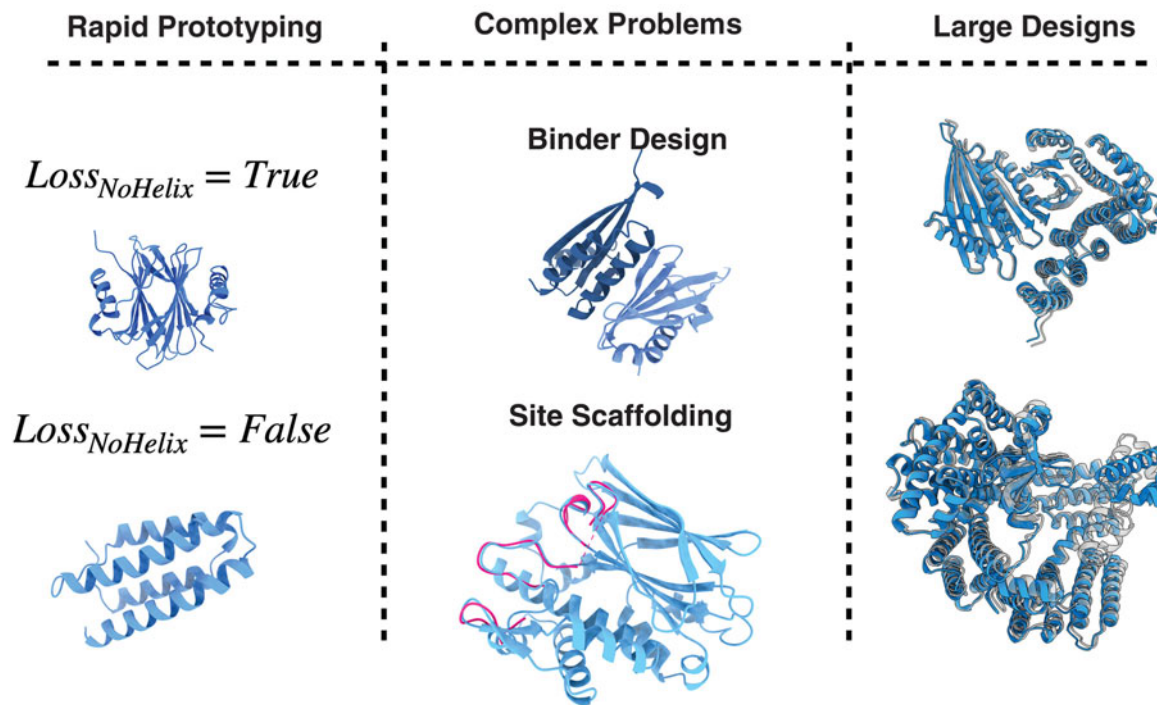
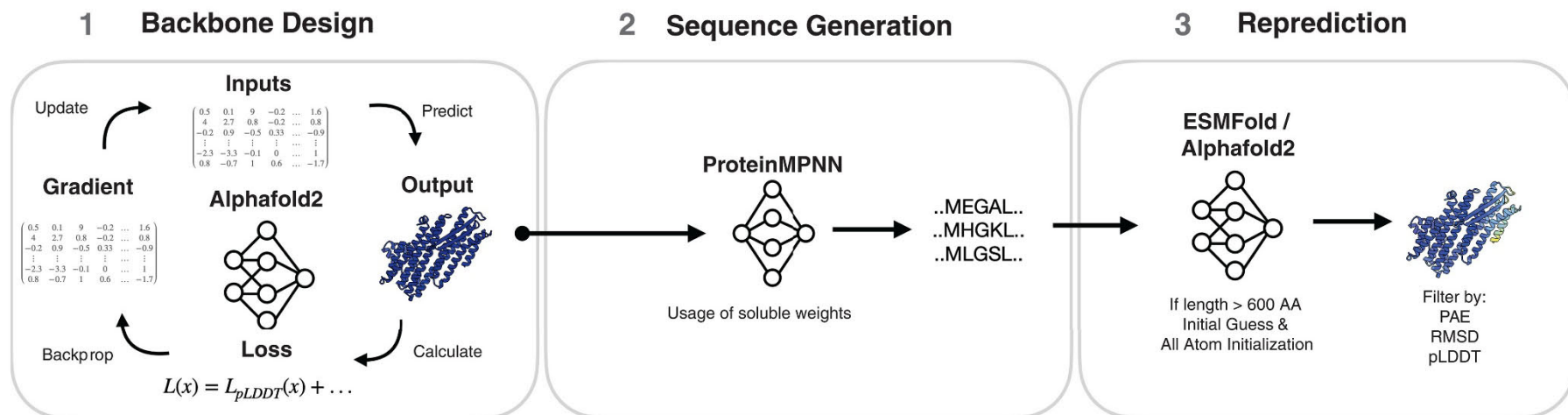
Brian Kuhlman,<sup>1\*</sup> Gautam Dantas,<sup>1\*</sup> Gregory C. Ireton,<sup>4</sup>  
Gabriele Varani,<sup>1,2</sup> Barry L. Stoddard,<sup>4</sup> David Baker<sup>1,3†</sup>

DIQVQVNIDDNGKNFDYTYTVTTESEL  
QKVLNELMDYIKKQGAKRVRISITART  
KKEAEKFAAILIKVFAELGYNDINVT  
DGDTVTVEGQLEGGSL

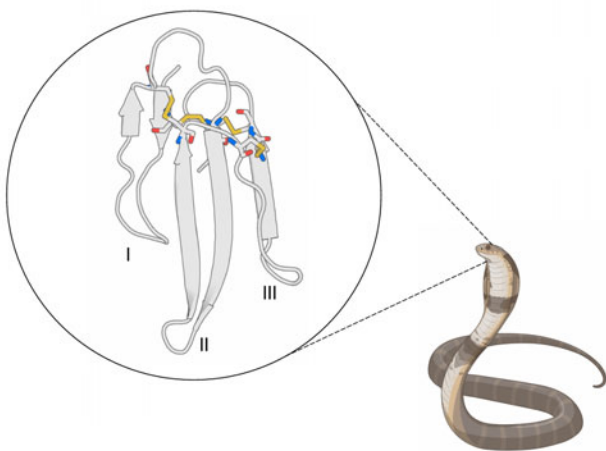
### 何がすごい？

- 知ってる配列を少し変えたとかではなく、ゼロから作った
- 親戚がない（既知のアミノ酸配列との相同性がない）
- それまでにみたことのない構造パターン（フォールド）

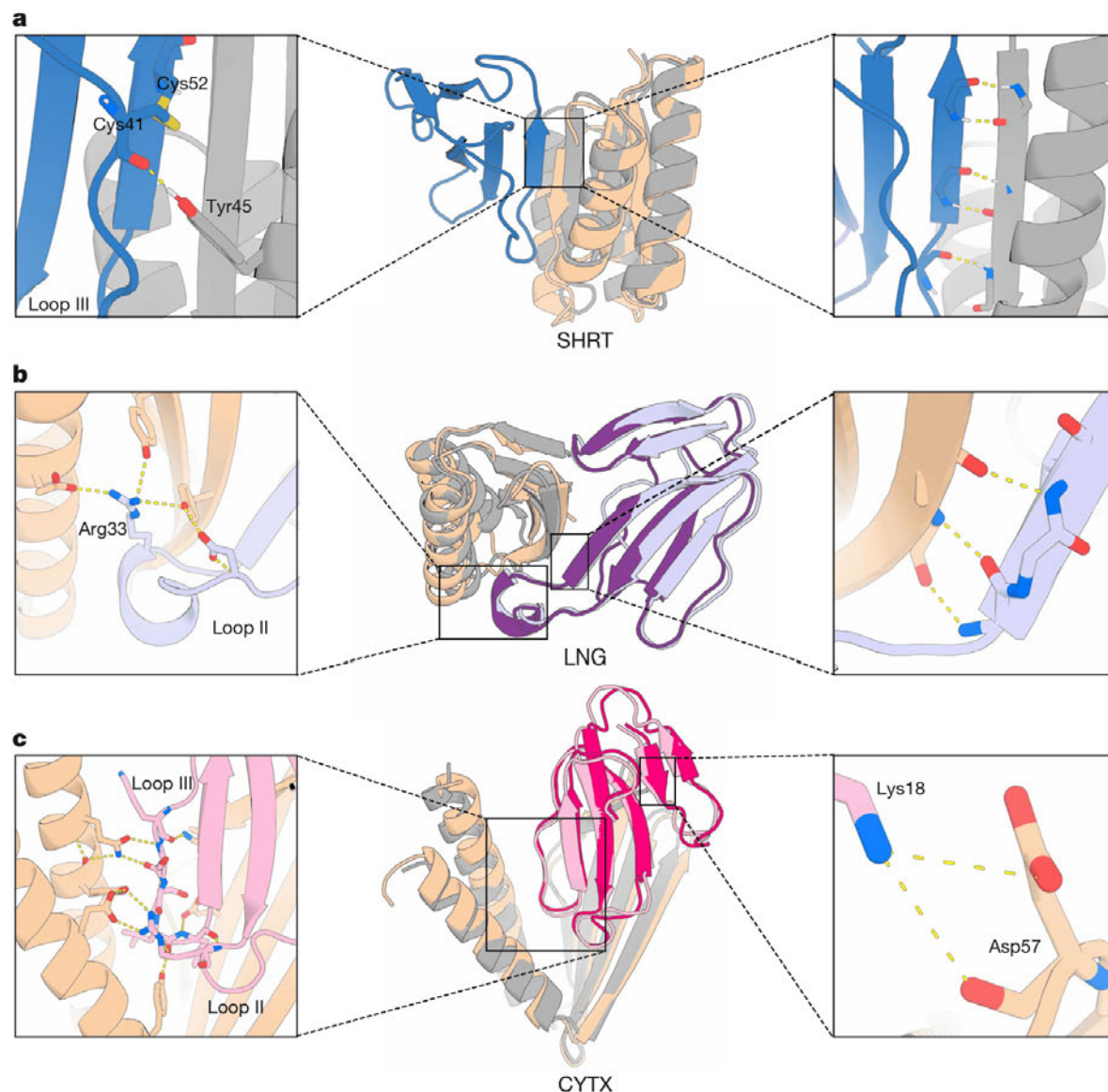
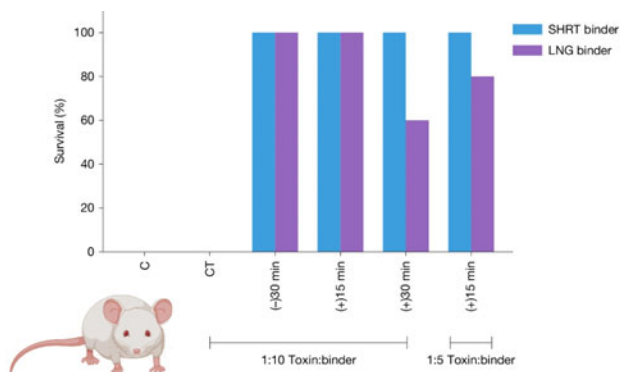
なんで7？ → 6までも作ってた。が、あんまりいい感じにならなかったとの談



ヘビ毒のタンパク質群 (3FTx) に結合する人工タンパク質を設計 (Bakerら, 2025)

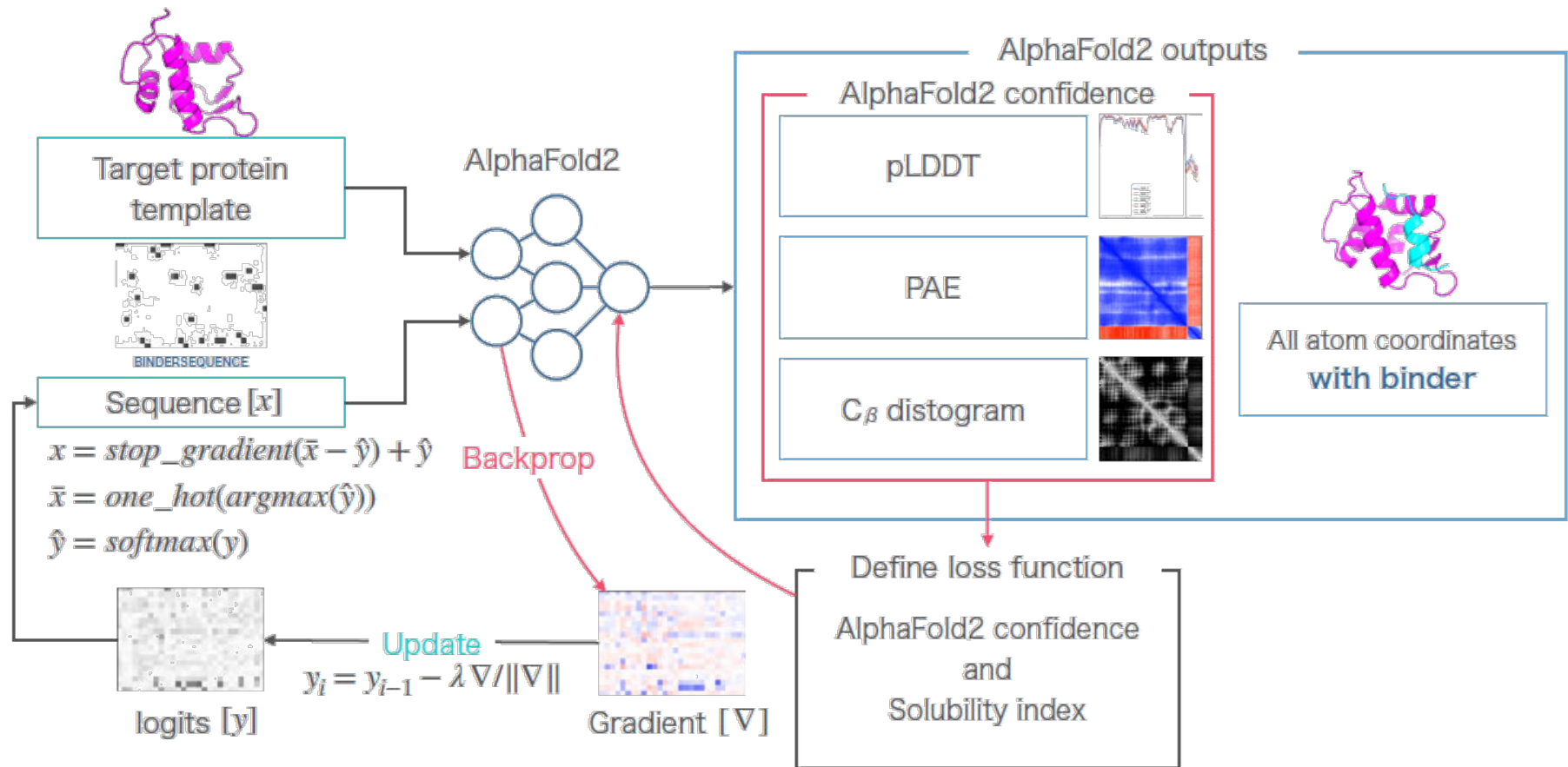


SHRT, LNGは  
致死量毒素投与後の生存を確認  
(CYTXは有意な効果が得られず)



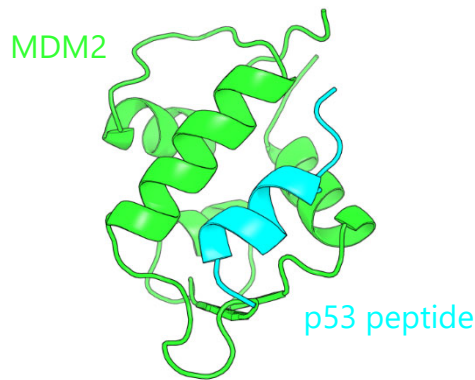
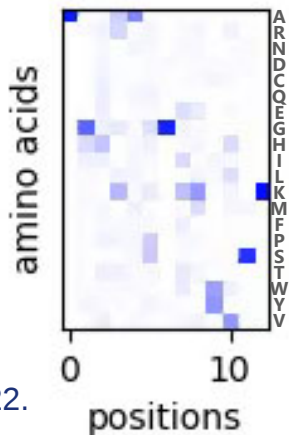
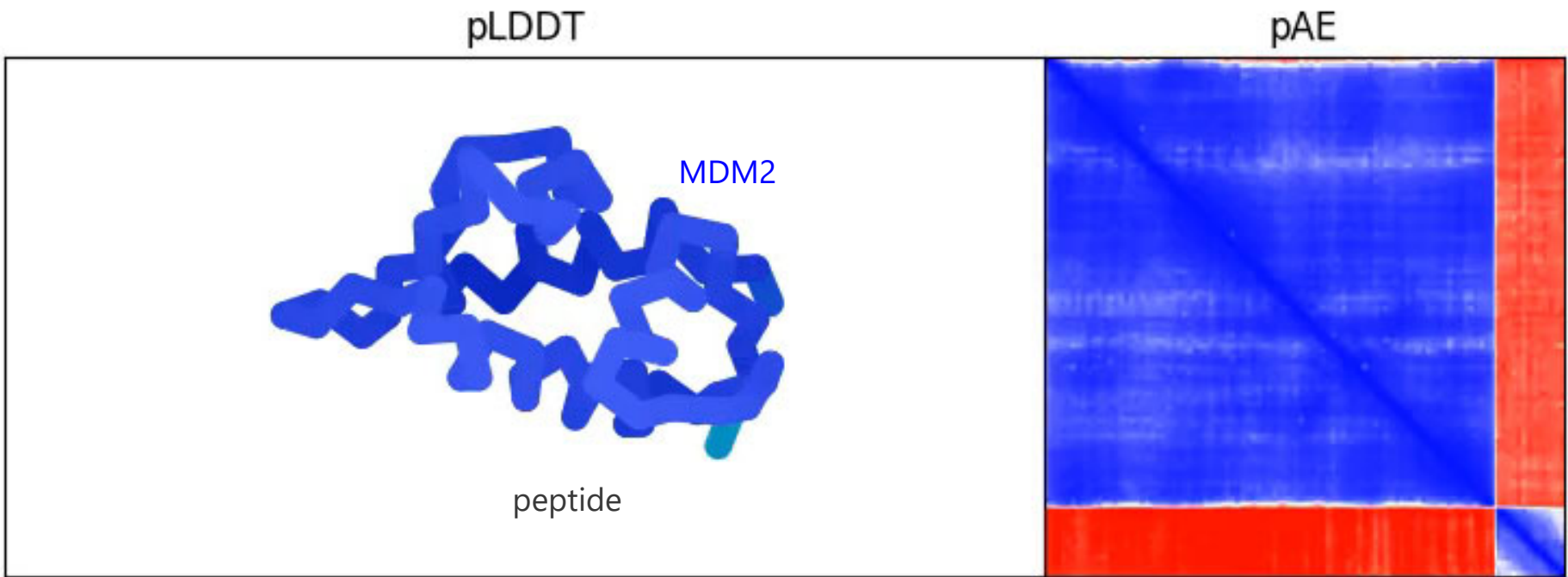
# AlphaFold2によるペプチド設計

AlphaFold2の繰り返し予測で、結合しそうなペプチド配列を設計

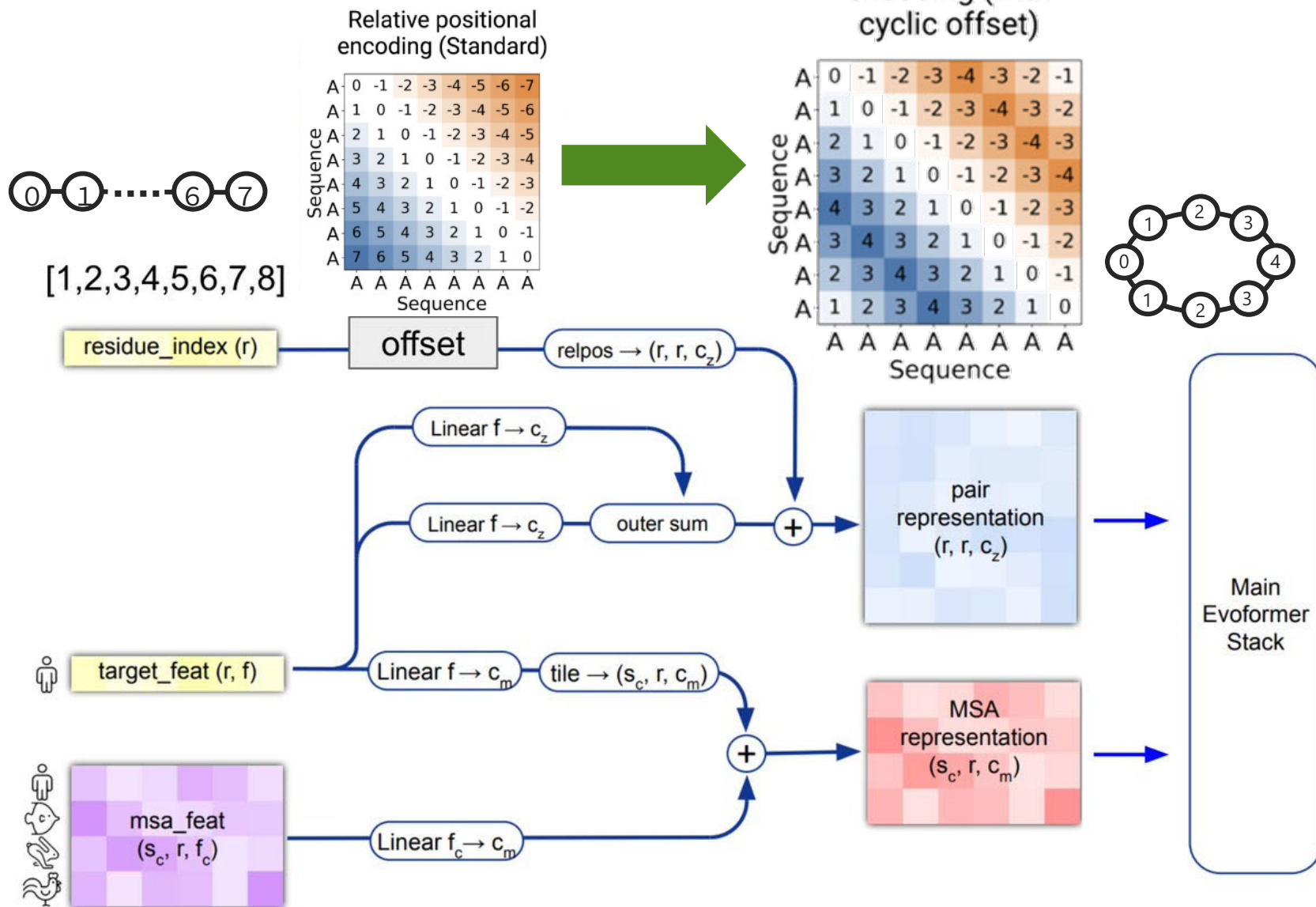


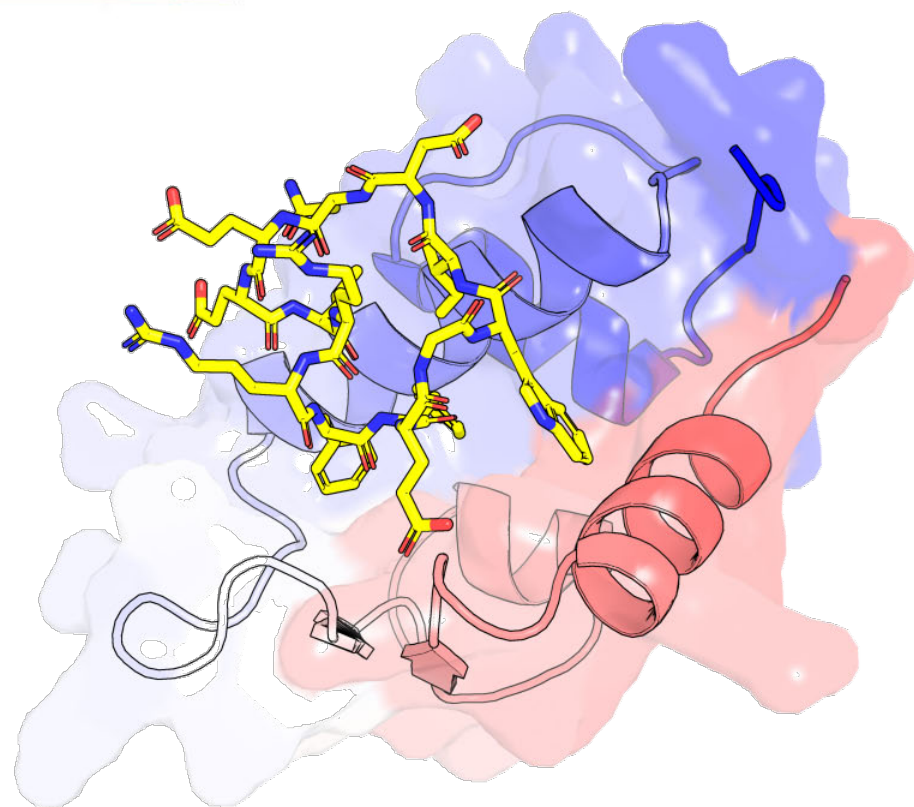
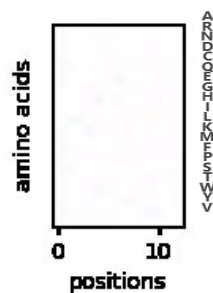
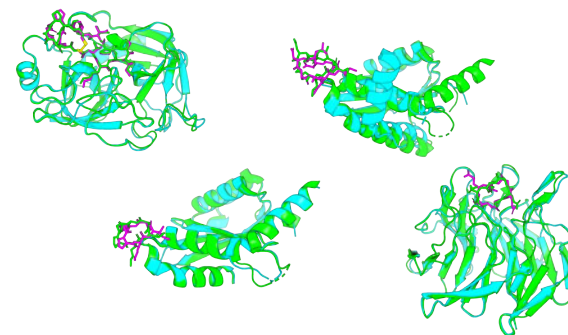
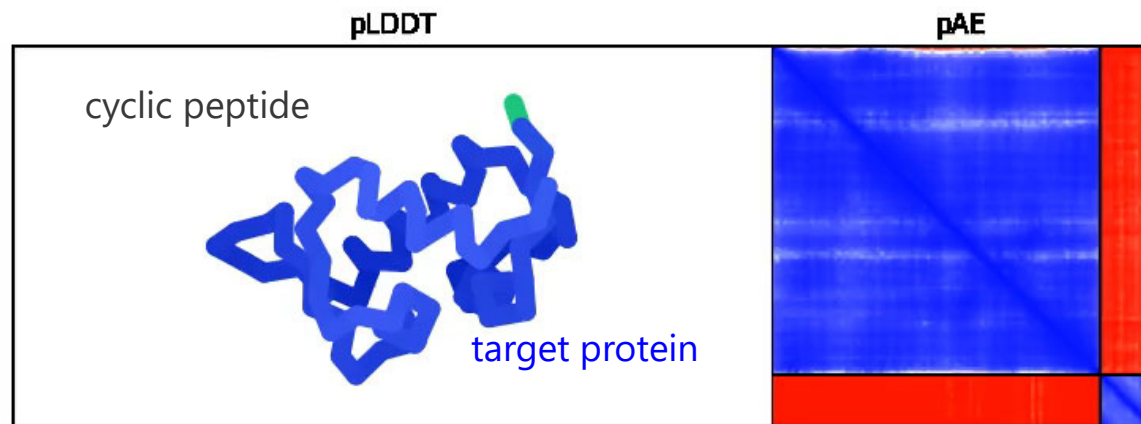
ColabDesignを改変して実装

<https://github.com/sokrypton/ColabDesign>



## Evoformerの手前のfeature embedding block

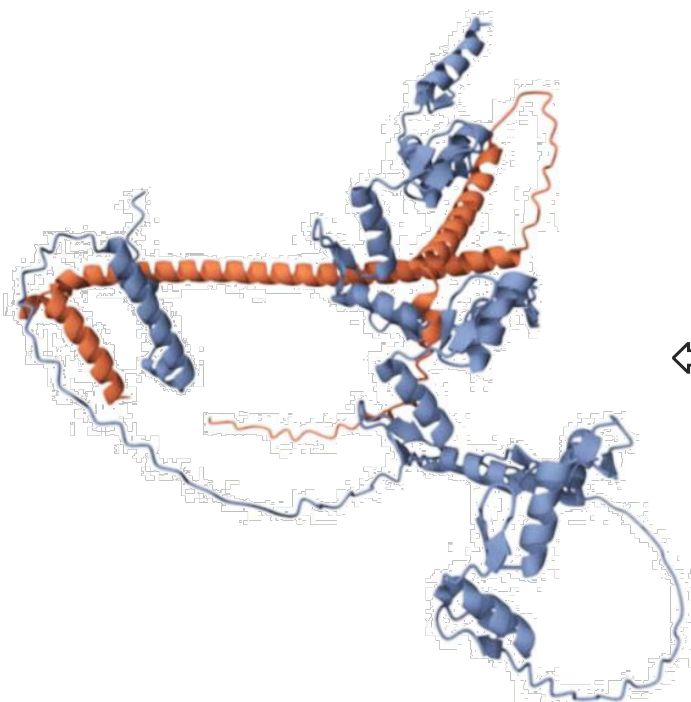




## AlphaFoldの予測構造はホンモノ？

AlphaFoldに2つのタンパク質を入力したらそれっぽい形が得られた。

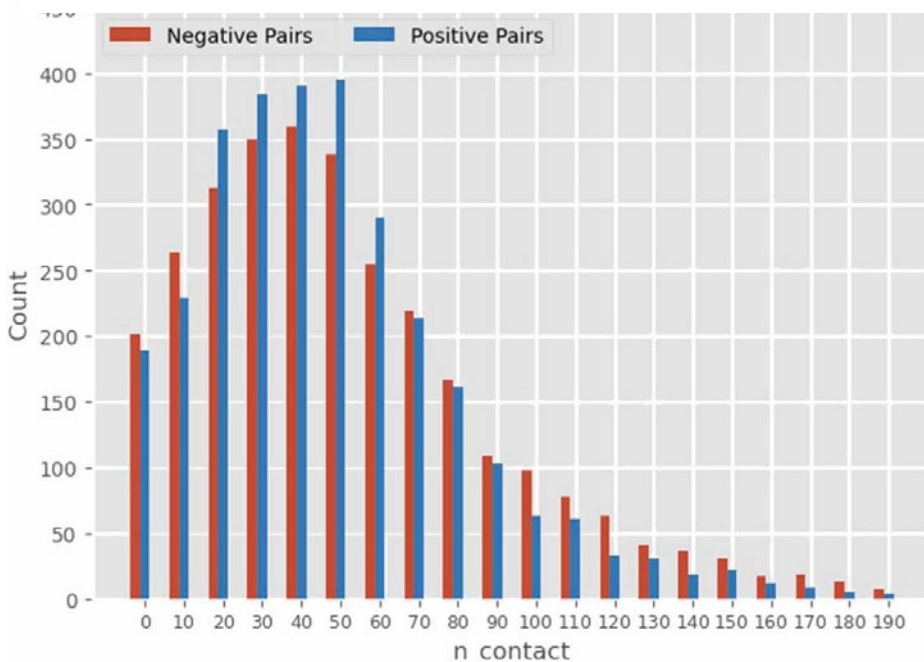
でもその2つのタンパク質は本当に相互作用するの？



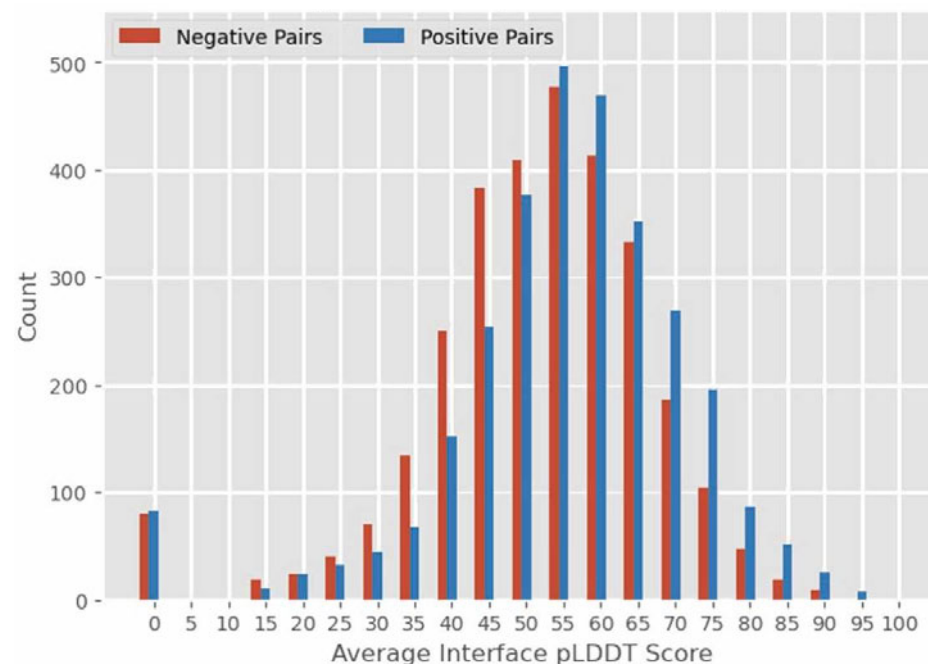
⇐ AlphaFold2が予測したタンパク質AとBの相互作用。  
**だが実際はAとBは相互作用しない。**  
(wet実験で分かっている事実)

## 偽物ペアと本物ペアの統計比較 → イマイチ違いがわからない

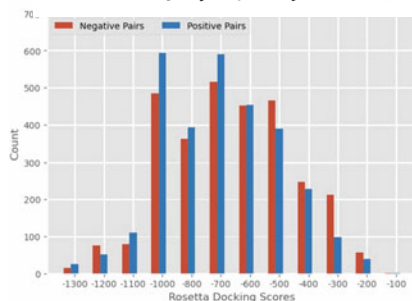
予測複合体の相互作用面のコンタクト残基の数



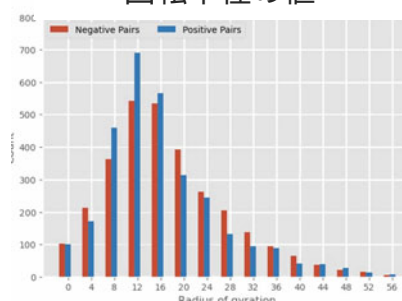
予測複合体の相互作用面の平均pLDDT



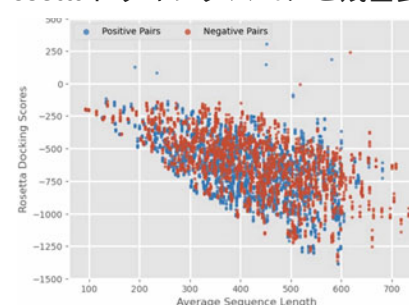
Rosettaドッキングスコア



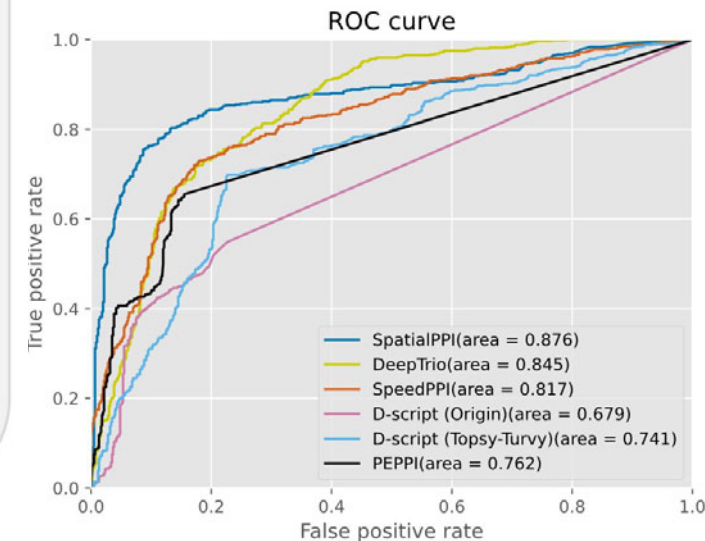
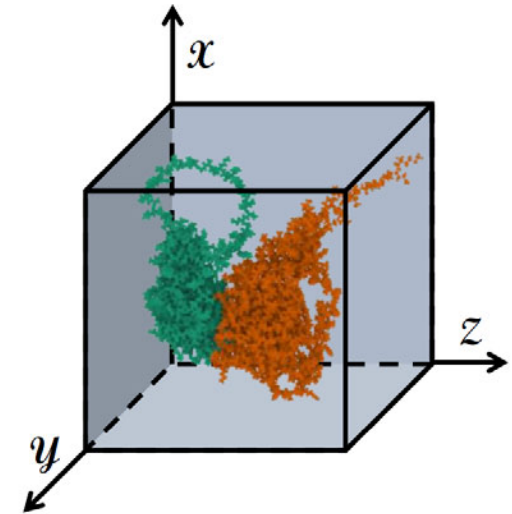
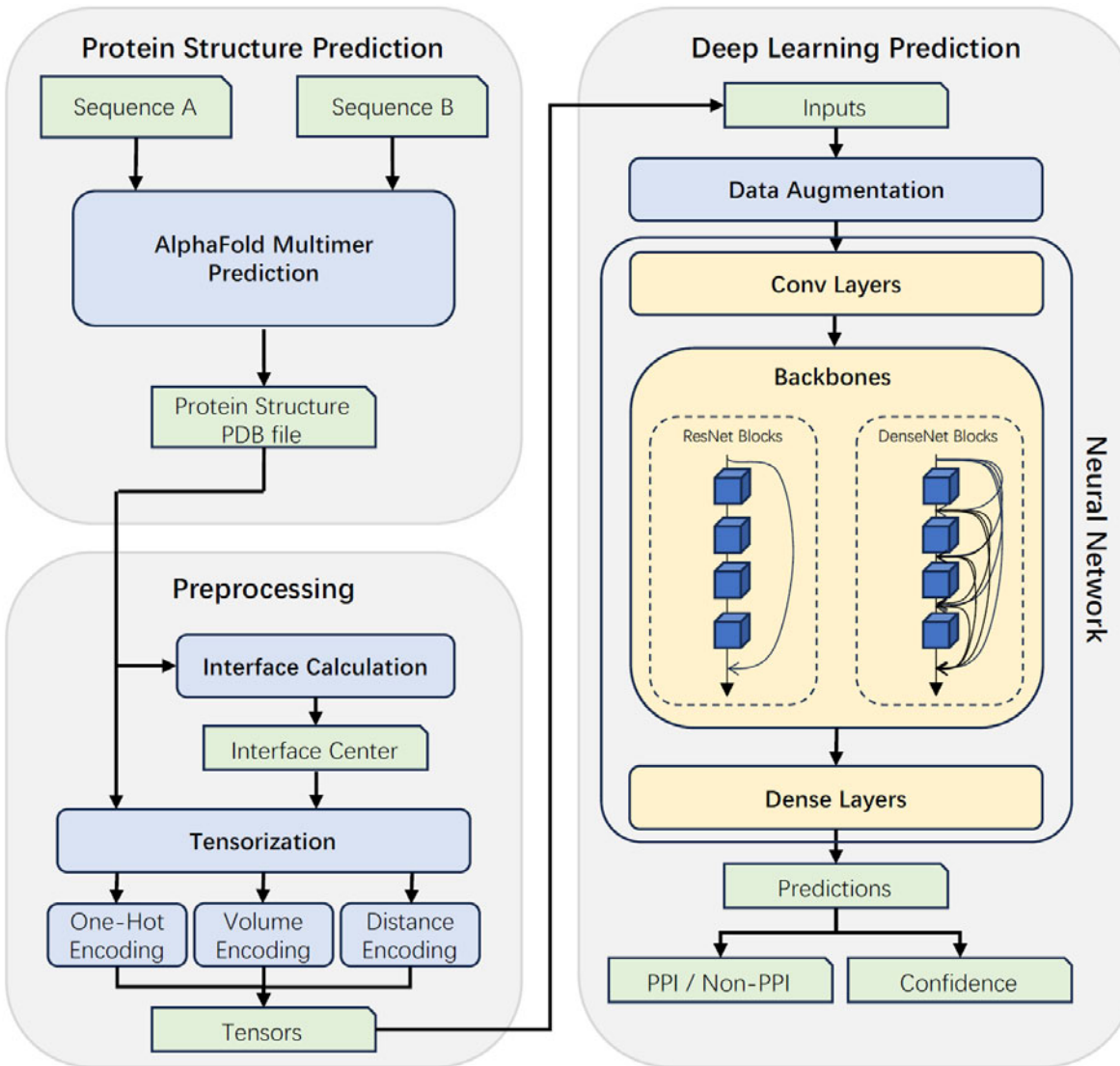
回転半径の値

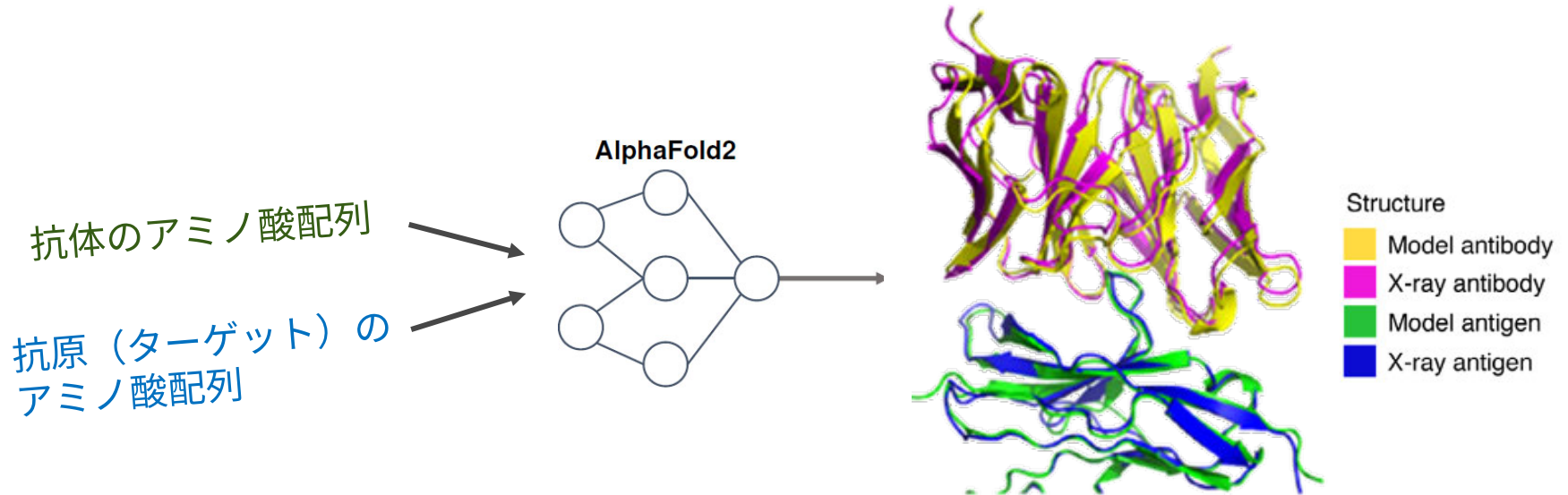


Rosettaドッキングスコアと残基長の分布

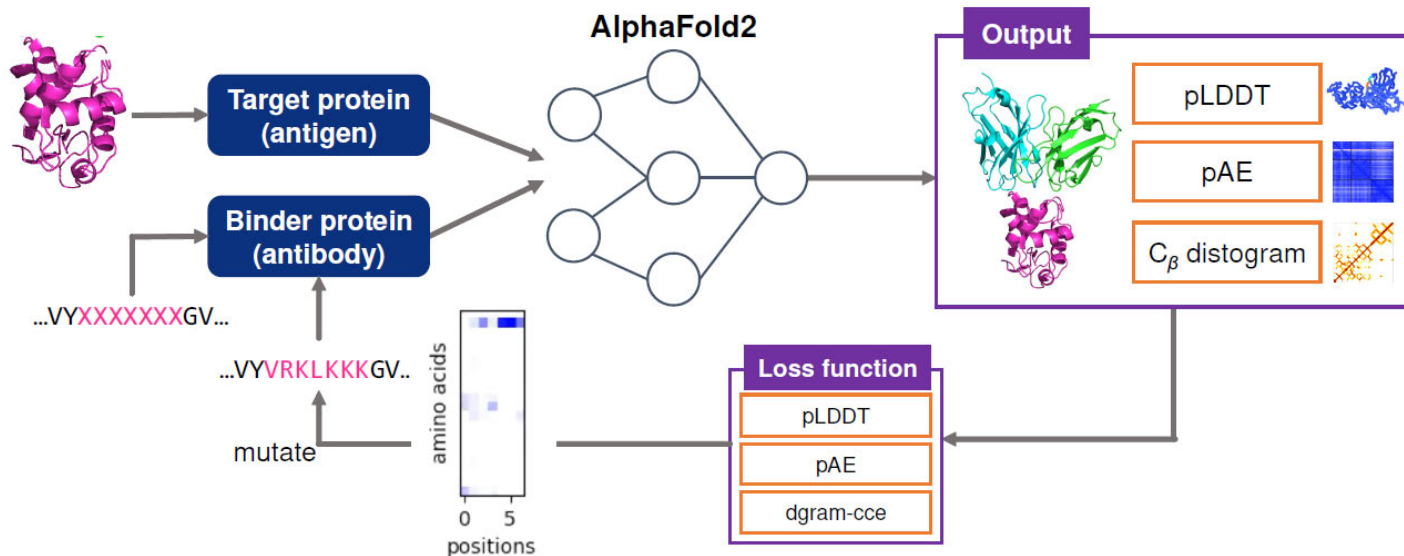


## SpatialPPI AlphaFold2が出力したPPIが「ホンモノ」か「ウソ」かを見分ける深層学習

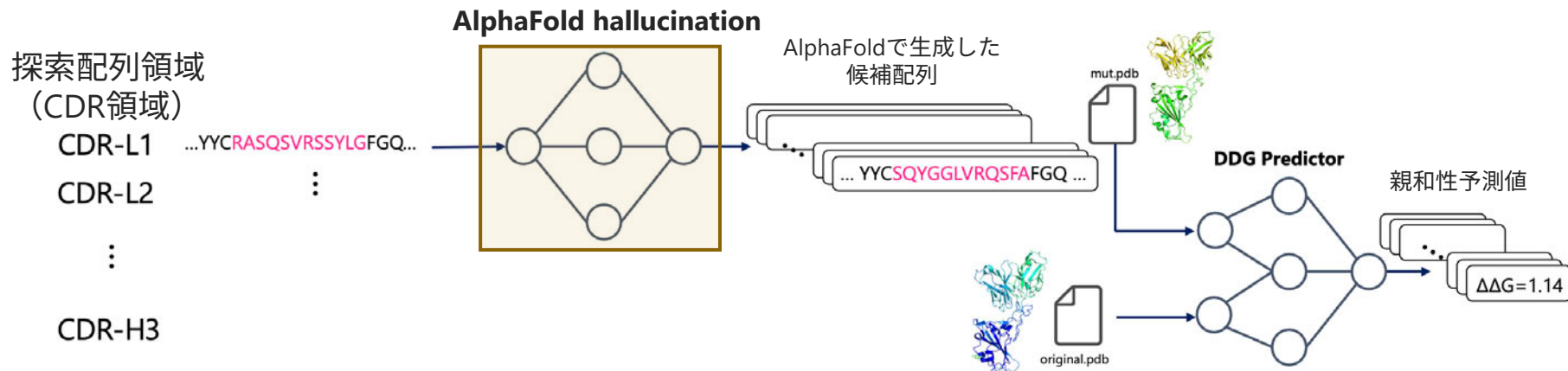




AlphaFold2は抗体の立体構造も精度良く予測できる。  
→配列を色々入れ替えて予測すればより良い抗体を探すことができるのでは？



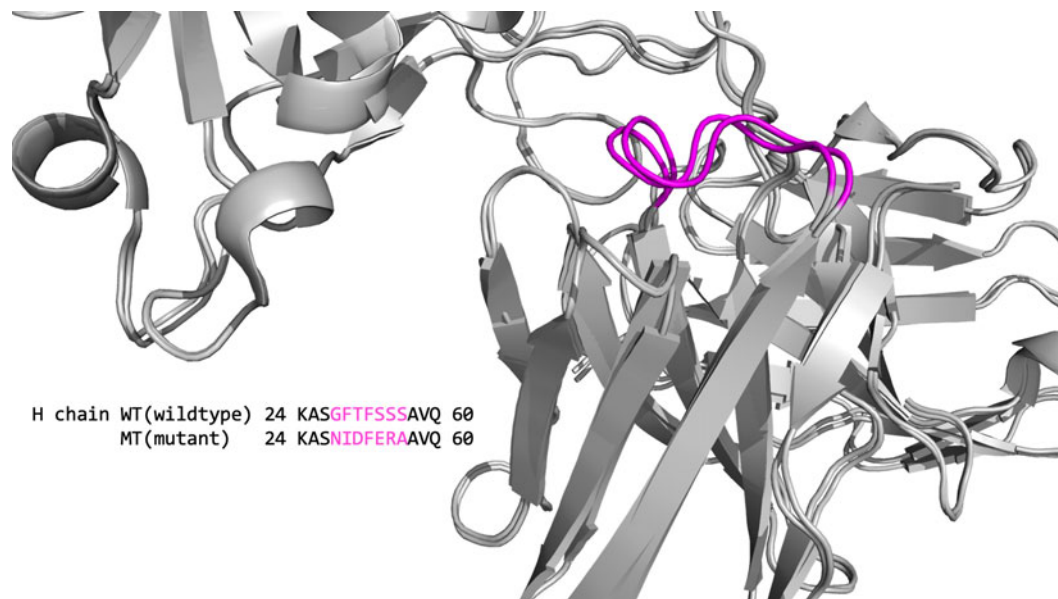
## より良い抗体をデザインする



## 予測結果の例

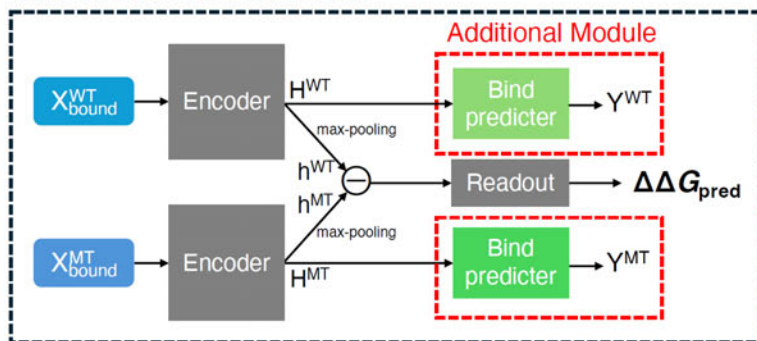
結合能予測値

$$\Delta\Delta G_{\text{pred}} = -0.49 \text{ kcal/mol}$$



## 深層強化学習とRosettaによる計算を併用した抗体デザイン

### Modified RDE-Network



### 1. Model Training

### Training Dataset

$\Delta\Delta G_{\text{FlexddG}}$

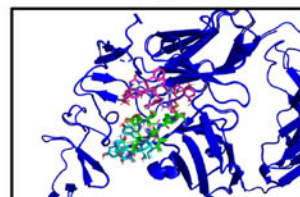
|      |
|------|
| -1.1 |
| 0.3  |
| 1.2  |
| -3.4 |
| 8.0  |
| 2.2  |

### 4. Data Augmentation

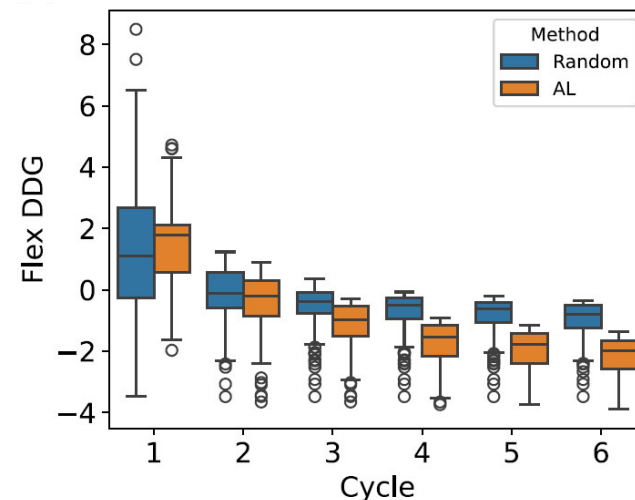
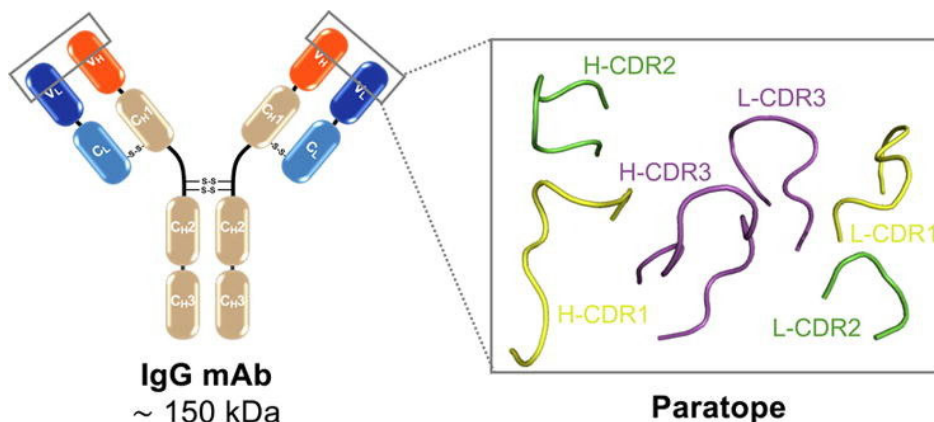
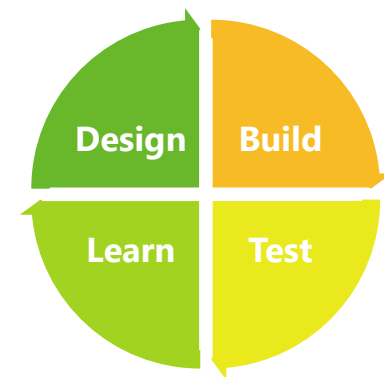
### Generated Mutants

### 2. Mutant Selection

YRWGGDGFYANDY



### 3. Flex ddG Calculation



## • AlphaFoldとはなんだったのか？ → 新しい生命科学を拓いた道具

- 気軽に構造を予測して機能を考える時代に（一般攻撃魔法）
- 構造予測は9割方当たる（凄い） ↔ 1割は外す
  - 似ている配列は似た構造であると覚えている → 1アミノ酸違いに弱い
  - 知ってる構造に引っ張られる → ほんとうに未知のものの予測はできてるか？
- タンパク質以外の予測もAlphaFold3でできるようになってきた
  - 予測が当たってるかどうかの検証はこれから
  - 知らない結合様式の予測ができるか？ たぶんできない  
（アロステリックサイト、未知エピトープ）
- タンパク質にもそれぞれの「事情」があるが無視している
  - 存在している環境はどのような状況か？
  - 翻訳後修飾、リガンドや糖鎖など、さまざまな分子の影響は？
  - 実験的に解かれた構造も「あり得る1つのスナップショット」のはず
- 「予測結果は正しいか」を見極める能力が人間に必要
- AIで予測できるから実験は要らない、ではない。（さらに重要に）
- タンパク質の構造を予測できる → タンパク質のデザインができるように

## ・タンパク質デザインと構造予測は表裏一体

- 構造予測精度が上がるとデザインもできる
- 構造は設計できるようになってきた。機能は？
  - 知っている機能の構造を作れば、似た機能になる
  - 新しい機能をデザインするには？
  - 実用上は一部だけアミノ酸を変えるだけでも良い、が、予測できるか？
- 機能は他者との関わりで決まる＝複合体のデザインへ
  - 新しい生化学反応を起こす人工タンパク質
  - 薬になる人工タンパク質
  - 毒になる人工タンパク質
- 作ったタンパク質が実在するかは実験で作って確かめる
  - 100%作れるかはわからない（1個できたらすごい＝成功率が表現しにくい）
  - 作りやすいタンパク質かはわからない
  - 作ってみないと、思った通りの機能を持っているかわからない
- タンパク質は生命システムの部品 ＝ 新しい生命のデザインへ

- AIによって創薬・生命科学でできることが増えてきた
  - 「無理なものは無理」を見極める能力が必要
  - 「予測結果は正しいか」を見極める能力も必要
  - Wet実験の重要性がより高まる＝DBTL, DMTAサイクル
- 次世代の創薬を開拓するために
  - 言語モデル技術は創薬においても注目
    - LLM、タンパク質言語モデル、化合物言語モデル、ゲノム言語モデル
  - どのような実験データがあるか、どのような生物評価系が使えるか
  - データサイエンス、シミュレーションを駆使する総合格闘技へ



## Funding



文部科学省 科学研究費助成事業 学術変革領域研究 (A)  
天然物が織り成す化合物潜在空間が拓く  
生物活性分子デザイン



第49回 けいはんな「エジソンの会」  
公益財団法人 国際高等研究所  
2025年4月25日（土）15:10～16:10

# AlphaFold

タンパク質の立体構造予測がもたらす  
未来の創薬・医療

東京科学大学 情報理工学院 准教授

おおうえ まさひと  
大上 雅史

