



トーゴの日シンポジウム2024
2024年10月5日（土）品川ザ・グランドホール

生命科学の未来を予想するーデータベースはもう要らなくなる…ってコト?!ー

AIでなんでも解決できるのか？

東京科学大学 情報理工学院 准教授

おおうえ まさひと

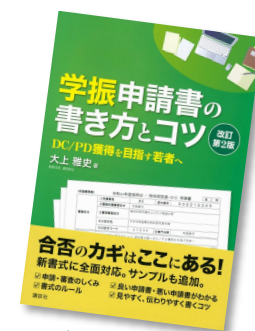
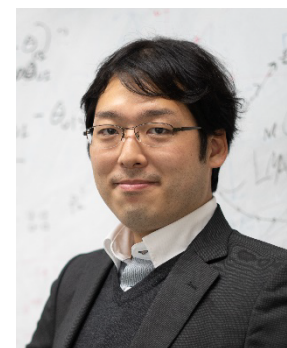
大上 雅史



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 (@すずかけ台キャンパス)  <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
- 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 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



医療用医薬品が発売されるまでには...

9～17年

新薬が発売されるまでには
およそ10年の歳月がかかる

数百億円

新薬を1個開発するには数
百億円～、中には1千億円
以上の研究開発費が必要

0.003%

化合物が新薬として発売さ
れる可能性は、およそ3万
分の1 (0.003%)

出典：日本製薬工業協会調べ（2005年～2009年）

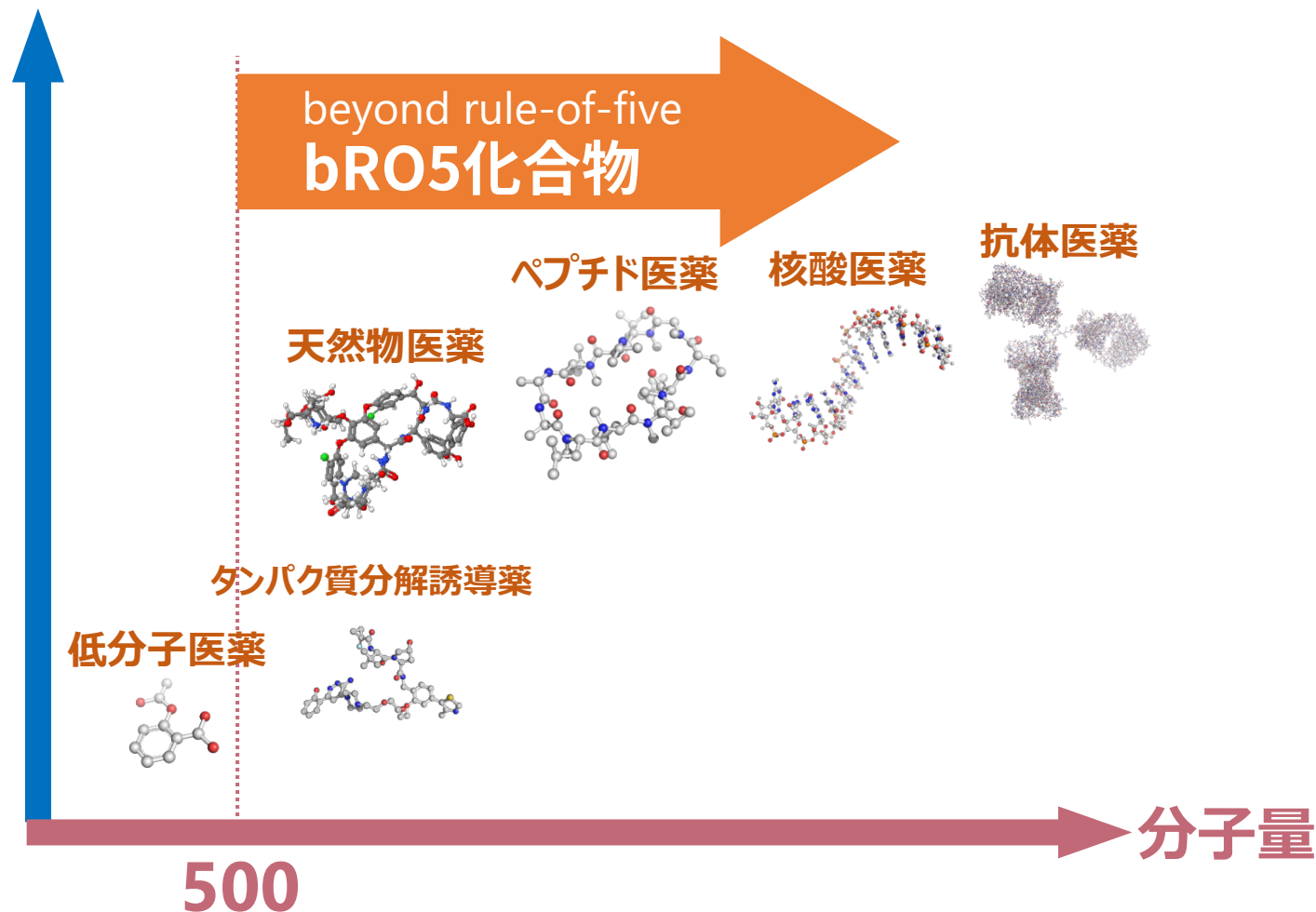


「有効」で「安全」で「安定」な「新しいモノ」が求められる

多様化する創薬モダリティを扱う情報科学技術の開発

※本講演の「モダリティ」は、分子の種類とか治療手段の種類という意味合いです

標的の多様性



(ニューモダリティ)

細胞療法



マイクロバイオー姆医療



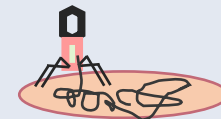
遺伝子治療



デジタル医療



ファージ治療



- なんでもできる？

- なんでもはできない、できることだけ

- AIで、できそうなことと、できなさそうなことの
目利き（経験）は重要

- データがあればできそう ←どんなデータが必要？

- データがあってもできなさそう

- できない理由は？

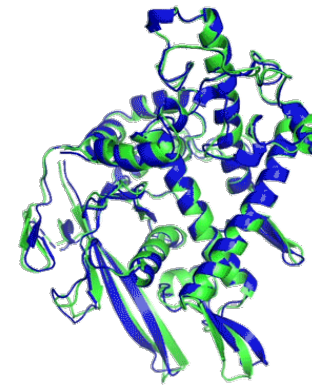
- よくある例 サンプル空間が広すぎる＝データが足りてない

- ＝知らない惑星の生物の写真を見て、名前を当てるのは無理

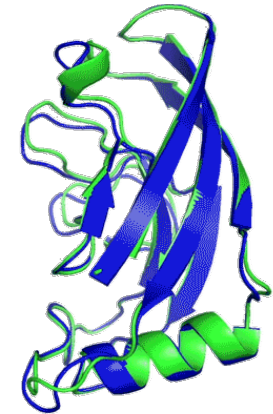
- データの量、多様性、適用範囲をきちんと考える必要がある

- できそうなこと
→ データがたくさんある
= 予測できそう！

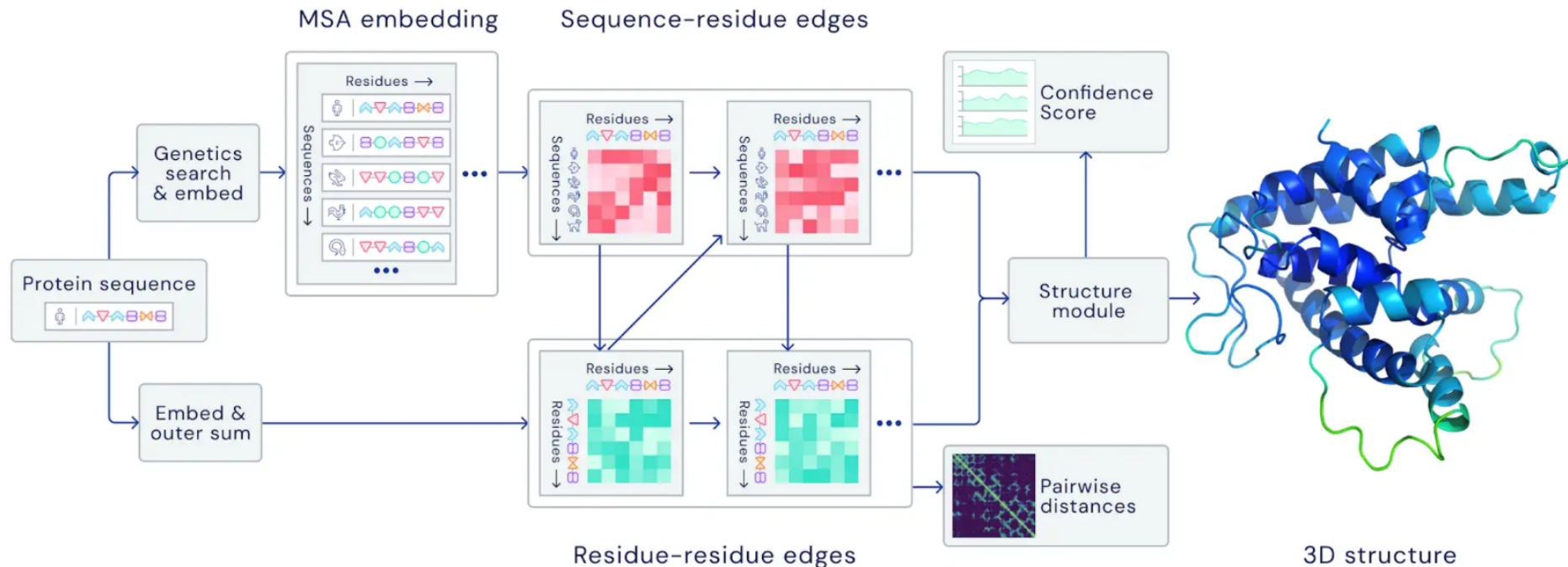
AlphaFold2 (2021)



T1037 / 6vr4
90.7 GDT
(RNA polymerase domain)



T1049 / 6y4f
93.3 GDT
(adhesin tip)



(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

配列相同性検索

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

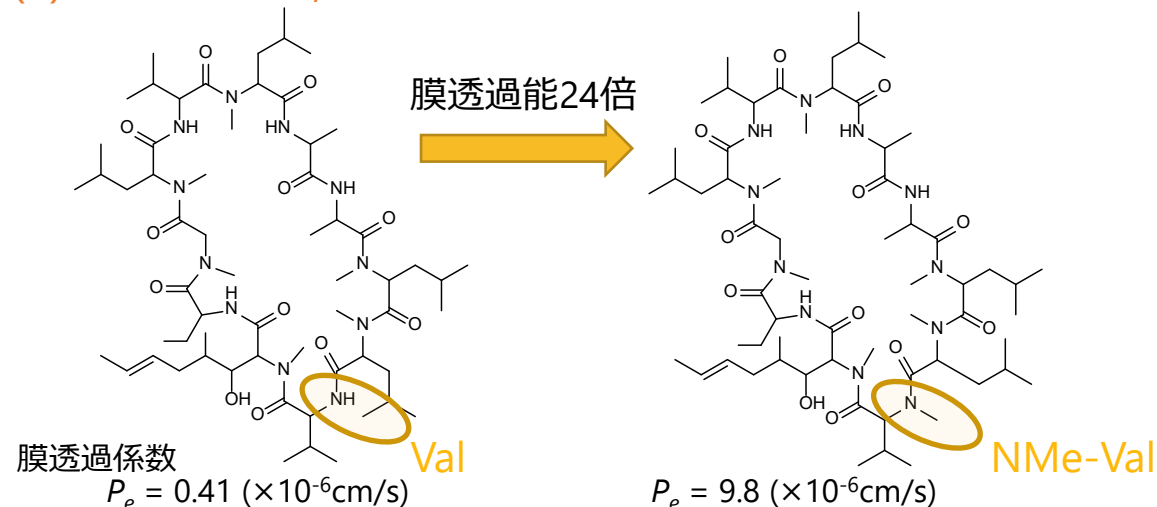
AlphaFold2の予測精度を支えるエッセンス

- 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)

- できなさそうなこと→データがあんまりない (あるけど不十分)
(細かい条件振ってない)
(違う惑星の生物が出てくる)

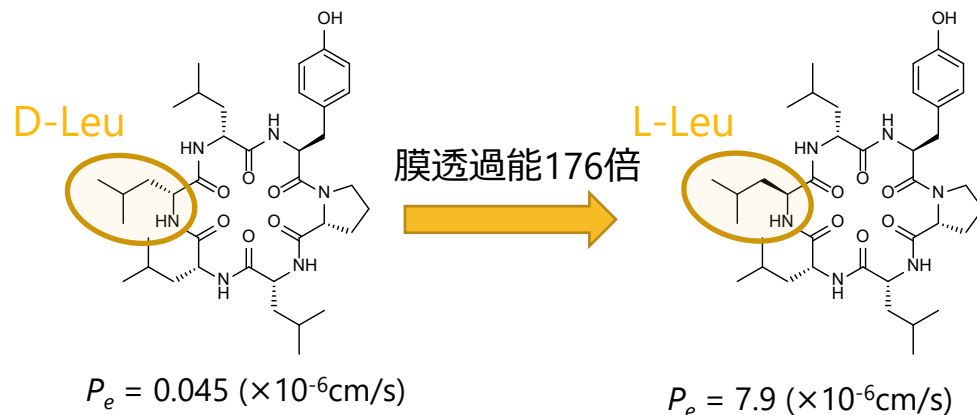
例 環状ペプチドの細胞膜透過能のアッセイ結果をしてみる

(a) Ahlback CL, *et al. Future Med. Chem.* 2015



Nメチル化のみの違いで
膜透過能が**24倍**変化

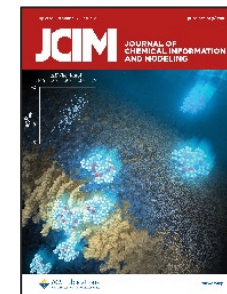
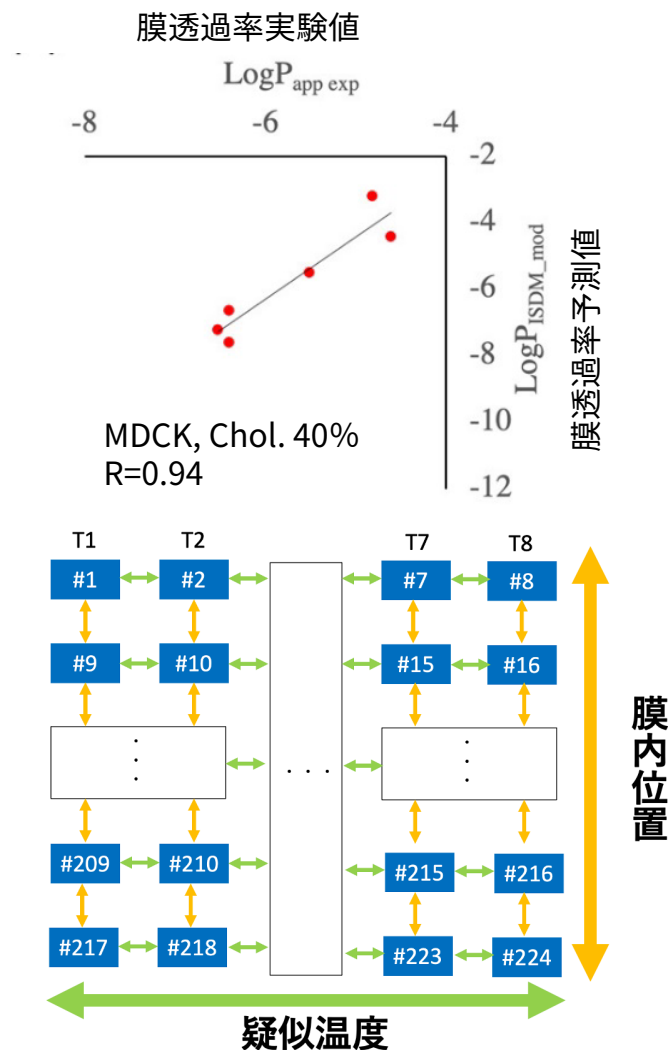
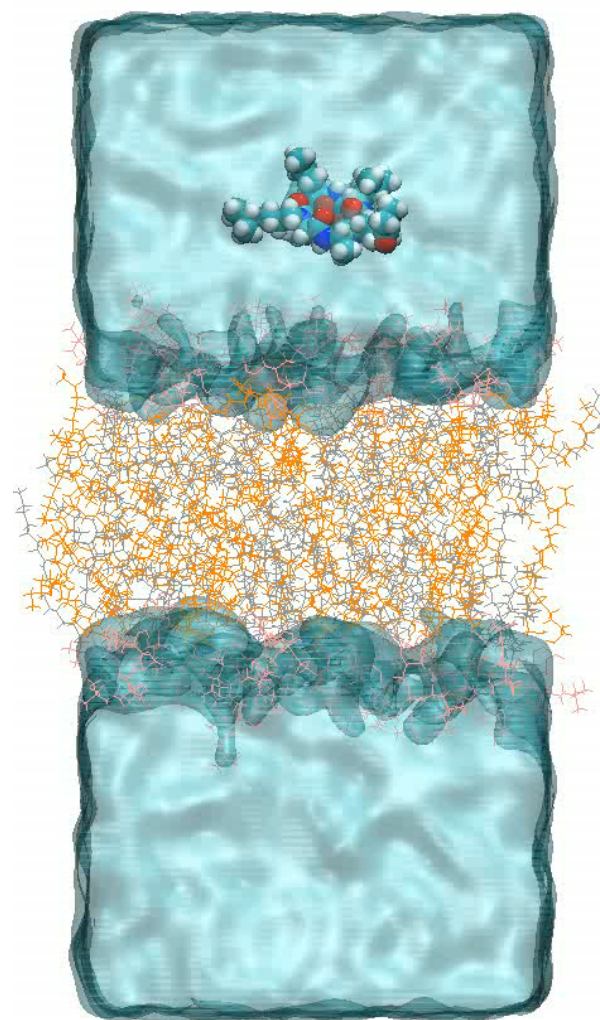
(b) Hewitt WM, *et al. J. Am. Chem. Soc.* 2015



L体/D体のみの違いで
膜透過能が**176倍**変化

ベクトルで表現すると多分至近距離。
どうやって区別する？

分子動力学シミュレーションでペプチドの膜透過過程を詳細に再現
TSUBAMEスパコン上で224枚のGPUが同時計算（※ただし1ペプチド10万円）



(Sugita+ *JCIM* 2021)



(Sugita+ *JCIM* 2022)



SARS-CoV-2のRdRPを阻害する経口投与可能な薬剤の候補となる化合物構造を生成してください。



SARS-CoV-2のRdRPを阻害する経口投与可能な薬剤の候補となる化合物構造を生成します。

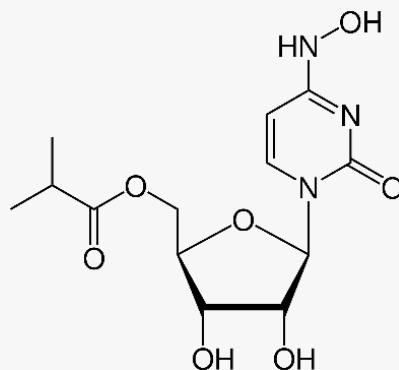


SMILES

Copy code

```
CC(C)C(=O)OC[C@H]1O[C@@H](n2cc/c(=N#O)[nH]c2=O)[C@H](O)[C@@H]1O=C
```

この化合物の構造式を示します。



※捏造ChatGPTです

AIでなんでも解決できるのか？

・ 情報科学・AIでできることとできないことがある

- データがあればできる／なくてもできることを探す／無理なものは無理
- 占いよりマシな予測（意外と難しい）、無茶振りに応える努力
- AI予測結果の解釈性、DBTLサイクルの拡充（Dry-Wet協働）、シミュレーション併用

・ バイオインフォマティクスの進展

- NLPやCV分野からの技術輸入で様々なタスクが解けた、ように見える、が...？
- 目的変数が大きく変わる微小変化の検出が重要（adversarial attackの逆）
- AIは役に立つようになった
 - ←以前からだいたいのことはSVMでできていた。何が変わった？
- やっぱりウェブインターフェースがないと浸透しない

・ AI技術の未来

- かゆいところに手が届く、きめ細やかなAIが求められる？
- 知りたいのは点変異の構造変化、ダイナミクス、メカニズム（予測根拠）
- どんなデータが、どれだけ必要か？（あんまり誰もよくわかってない）
- もっと大量の計算機資源