

大規模シミュレーションとAIで拓く ペプチド創薬への道



アヘッド・バイオコンピューティング株式会社
取締役 CTO 秋山 泰

学際ハブ拠点 「生命科学」チームの一員として

2

AI 時代における計算科学の社会実装を実現する学際ハブ拠点形成

材料・**生命科学**・気象などの分野を中心に、
企業・国研・大学間のスーパーコンピュータを用いた連携研究を促進し、
我が国におけるシミュレーション科学・工学の実用範囲を広げ、特に
企業における計算科学DXを大きく推進する、計算科学の社会実装を目指す。

・筑波大学 計算科学研究センター

HPC分野 朴泰祐教授、小林諒平助教ら
計算生命科学 重田育照教授、堀優太助教ら



・アヘッド・バイオコンピューティング株式会社

秋山泰、大上雅史、杉田昌岳、藤江拓哉ら



・エヌビディア合同会社 エンタープライズ事業本部

古家真之介氏ら



アヘッド・バイオコンピューティング株式会社

3

バイオインフォマティクス、機械学習、高性能計算等の技術を統合して、創薬や大規模バイオデータ処理で、皆様のお手伝いをする企業です。
先端技術の融合で迅速なソリューションを与えることを得意としています。

AI システム開発 納入例

- 創薬ターゲットに特化したAlphaFold2拡張
- 細胞培養を最適化する培地予測AI
- 化合物活性予測・誘導体推薦AI
- ΔG計算のための化合物生成モデル

IT創薬 納入例

- MEGADOCKによるタンパク質間相互作用予測
- ΔG計算による化合物スクリーニングの計画最適化
- 標的結合ペプチド配列予測
- 環状ペプチドの細胞膜透過MDシミュレーション計算

量子 コンピューティング 納入例

- 低分子SBDD向け量子アルゴリズム
- ペプチド配列最適化
- アロステリック制御予測
- 金融ポートフォリオ最適化

**受託研究開発、大量データ解析、
各種プログラム開発、調査など、承ります。**



アヘッド・バイオコンピューティング株式会社
<https://ahead-biocomputing.co.jp>

当社は2018年に設立された、東京科学大学認定ベンチャー(第T092号)です。

〒210-0007
神奈川県川崎市川崎区駅前本町11-2 川崎フロンティアビル4階

☎ 044-440-3122

✉ contact@ahead-biocomputing.co.jp



本日の内容

4

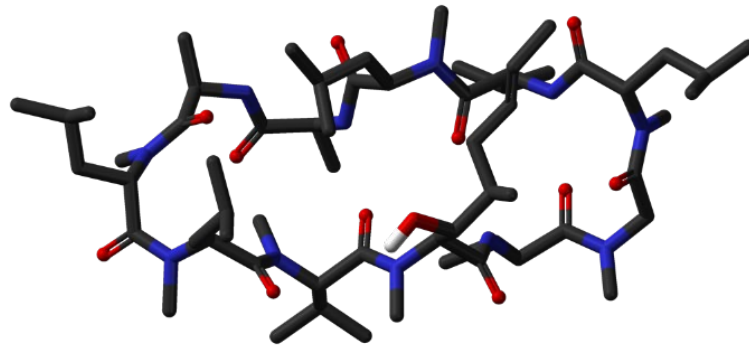
1. ペプチド創薬への期待と課題
2. 深層学習による細胞膜透過性予測
3. 大規模シミュレーションによる細胞膜透過性予測
4. MPS (Multi-Process Service)による
大規模シミュレーションの効率化



※当事業成果 ICPP2024 国際会議 (2024.8)にて発表

このうち1.~3.の内容は、東京科学大学で実施された学術研究の既発表成果をご紹介します。
アヘッドバイオ社は同大学成果を有償で技術移転して頂き、その社会実装を目指して活動しています。

1. ペプチド創薬への期待と課題

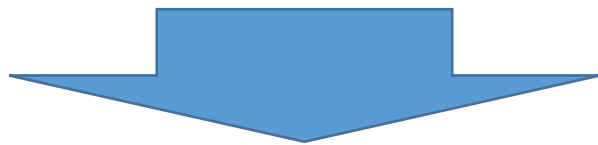


ペプチド創薬

タンパク質 = アミノ酸が約50～約3万個 連なった分子

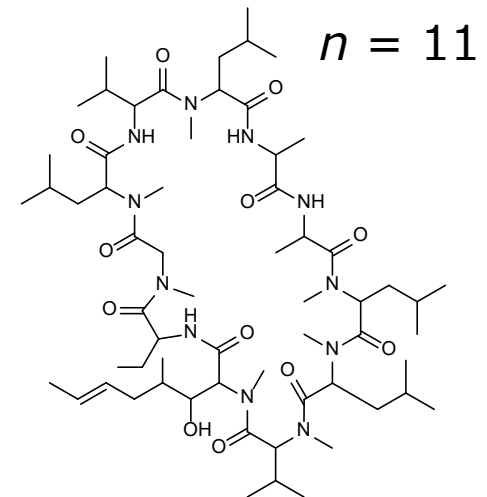
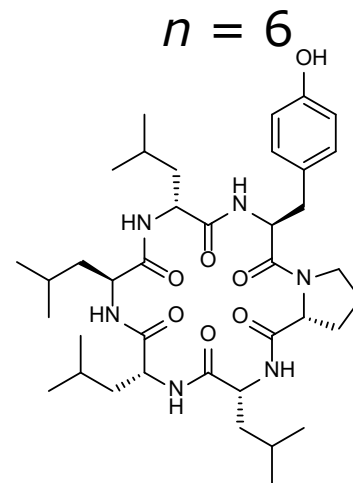
ペプチド = アミノ酸が2個～約50個 連なった分子

- ・栄養としてのペプチド
- ・機能性ペプチド



・ペプチド創薬

「薬剤設計の標準的骨組み」として
(特殊環状) ペプチド分子を活用



既存の創薬に
比べて様々な利点

低分子と抗体の利点を併せ持つ ペプチド医薬

7

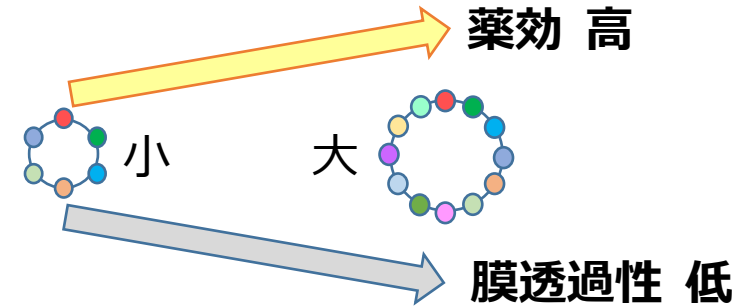
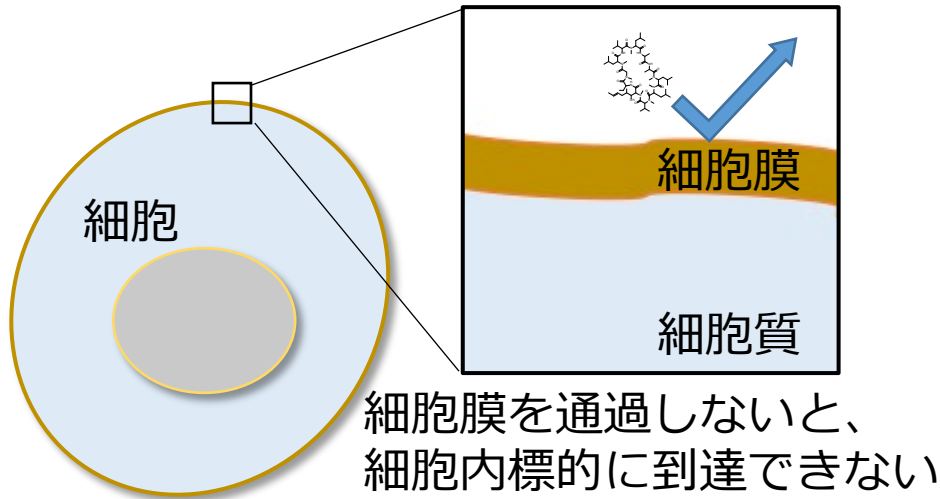
	低分子医薬品	中分子医薬品		高分子医薬品
	有機化合物 ～1,000	ペプチド医薬 600～2,500	核酸医薬 1,000～30,000	抗体医薬 150,000～
分子量	～1,000	600～2,500	1,000～30,000	150,000～
医薬の対象となる標的の数	◎ 多い（ブロックバスター）	○ 多い（アンメットニーズ）	○ 多い（アンメットニーズ）	△ 少ない
標的特異性	△	○	◎	◎
PPI阻害能	△	○	○	○
体内持続性 （体内安定性）	○	△	△	◎
細胞膜透過性	○	△	△	×
標的探索	○	△	△	◎
安全性・毒性	△	△	△	○
製造コスト	◎ 安価	○	○	× 高価

（ご参考）皮膚がん・肺がん等向け抗体医薬のオプジーボは、2018年4月の国内薬価引下げ後も患者一人当たり年間1000万円超の費用となる。製造には動物細胞を利用。

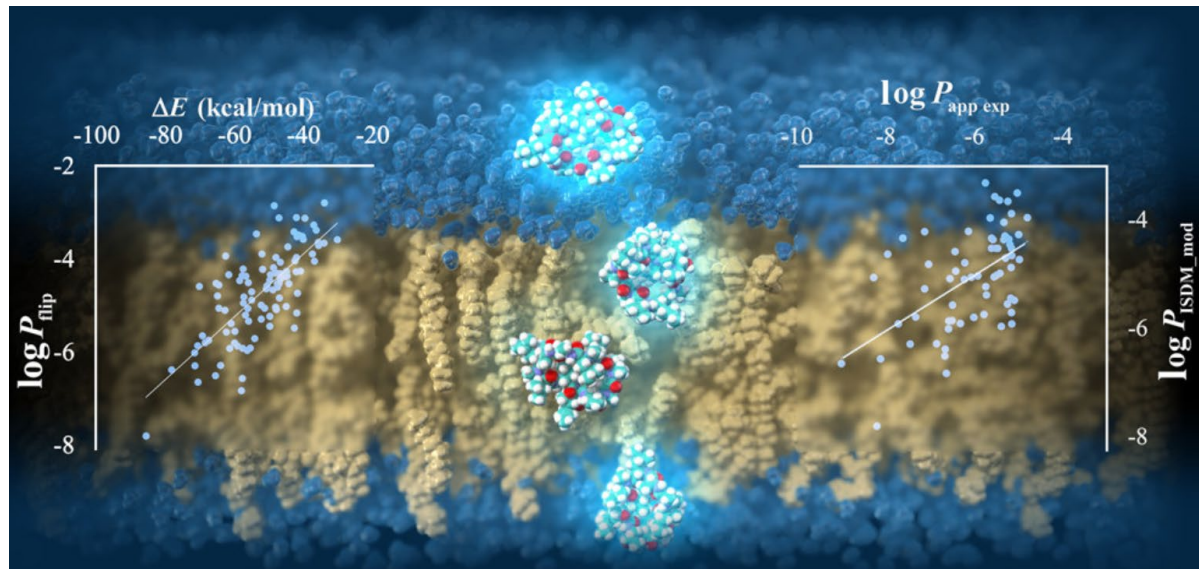
膜透過性はペプチド創薬の重大な未解決課題

8

問題点 サイズが大きいため、細胞膜を通過しにくい



しかも、薬効の高い大型ペプチドほど膜透過性の良いものは見つけにくい





2. 深層学習による 細胞膜透過性予測

深層学習による体内持続性予測 (2022) 10

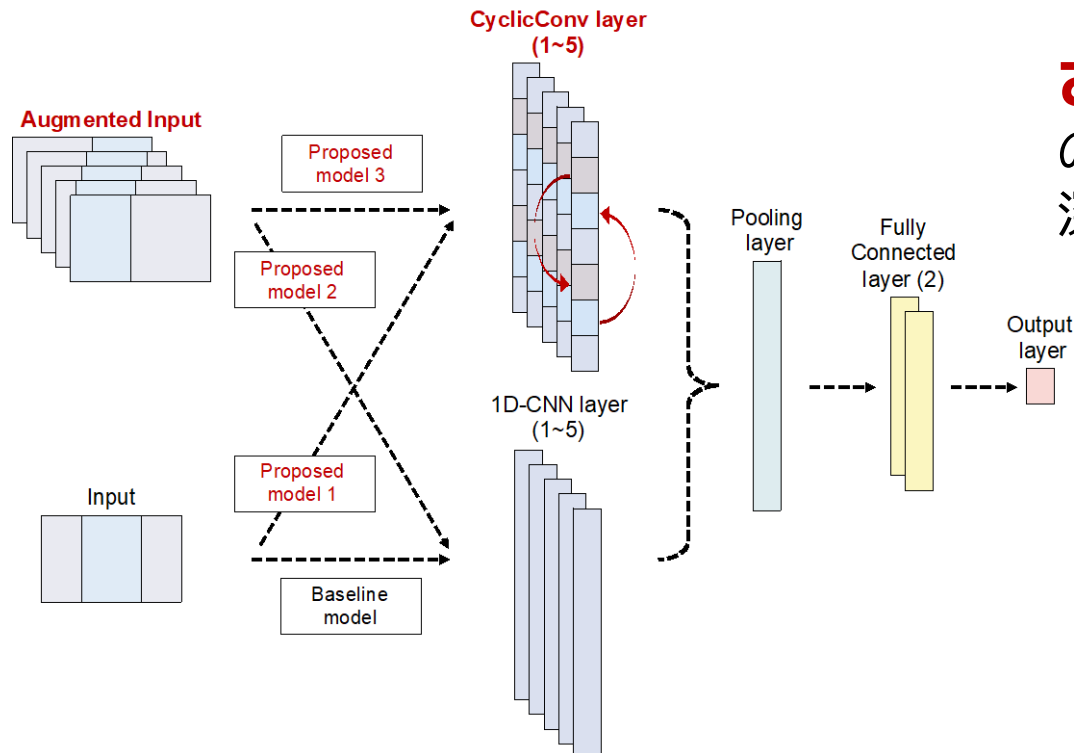
Bioinformatics, 2021, 1–8
<https://doi.org/10.1093/bioinformatics/btab726>
Original Paper

OXFORD

Data and text mining

Plasma protein binding prediction focusing on residue-level features and circularity of cyclic peptides by deep learning

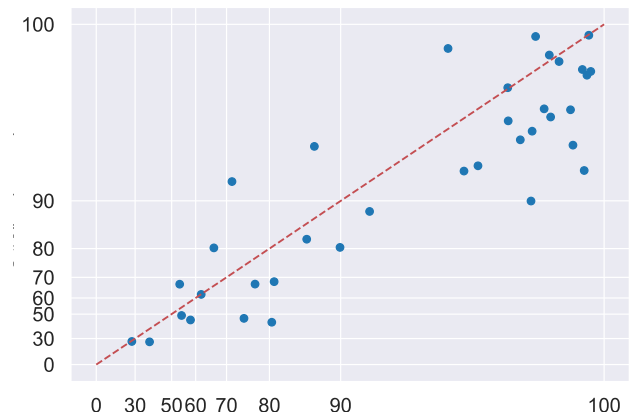
Li J., et al. *Bioinformatics* 38(4) 2022.
doi: 10.1093/bioinformatics/btab726



産業界との数年間の共同研究で、
深層学習用のデータ量を確保

**環状ペプチドのPPBデータを
約500件独自収集した**

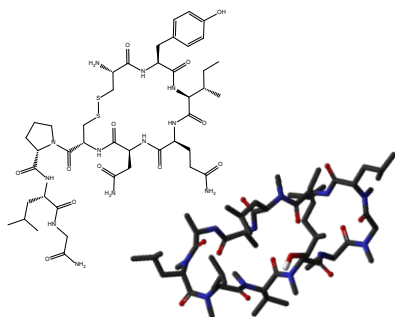
さらにデータ拡張（水増し）
の手法を独自開発したことにより
深層学習の適用がようやく可能に



機械学習による予測モデル構築の流れ

11

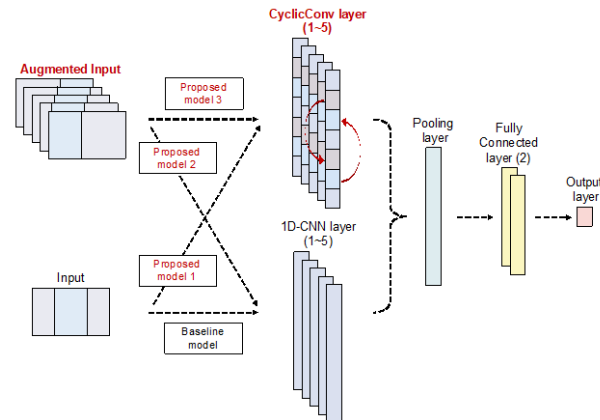
ペプチド構造情報
(2次元, 3次元, 他)



高次元特徴ベクトル表現,
化合物グラフ表現

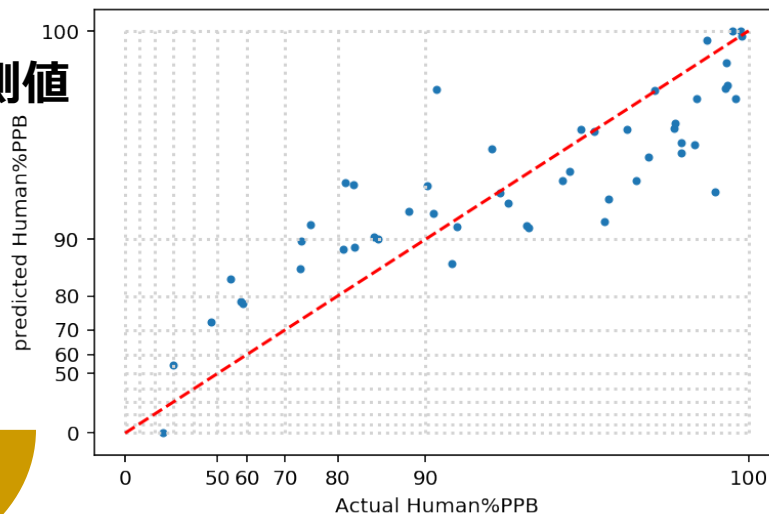
MW
LogS
LogP
#Ring
TPSA
:

機械学習アルゴリズム選択,
アーキテクチャ設計



予測モデル

予測値

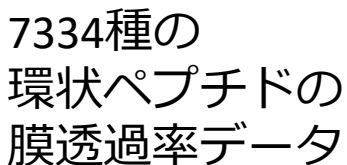


実測値

訓練データ
テストデータ
網羅的収集

評価・改良
繰り返し

実験データが
より多く集まれば
さらに精度が向上



- ①HELM記法の採用による正確な構造記述
- ②可能な三次元配座を計算済（クロロホルム中、水中、真空中）

CycPeptMPDB データベースとして国際公開

13

CycPeptMPDB: A Comprehensive Database of Membrane Permeability of Cyclic Peptides

Jianan Li, Keisuke Yanagisawa, Masatake Sugita, Takuya Fujie, Masahito Ohue, and Yutaka Akiyama*

Li J., et al. *J. Chem. Inf. Model.* 63(7), 2240–2250, 2023.
doi: 10.1021/acs.jcim.2c01573

CycPeptMPDB: Webから利用できる世界初の環状ペプチド膜透過性データベース

7,334種の環状ペプチドのデータを集積

立体配座



精密な
配座表現

PEPT1([dL]...)
\$PEPT1,PEPT1,
1:R1-6:R2\$\$\$



各種実験で計測された
膜透過性データ



物性値インデックス
(脂溶性など)



環状ペプチドを
構成する部分構造
(モノマー)の同定



閲覧・検索
および可視化
の強力なツール



無償で利用や
ダウンロード
が可能

CycPeptMPDBは、創薬研究者らに利用されるだけでなく、
膜透過性予測のための機械学習モデルの訓練データとして活用できます

深層学習による細胞膜透過性予測 (2024) 14

CycPeptMP: enhancing membrane permeability prediction of cyclic peptides with multi-level molecular features and data augmentation

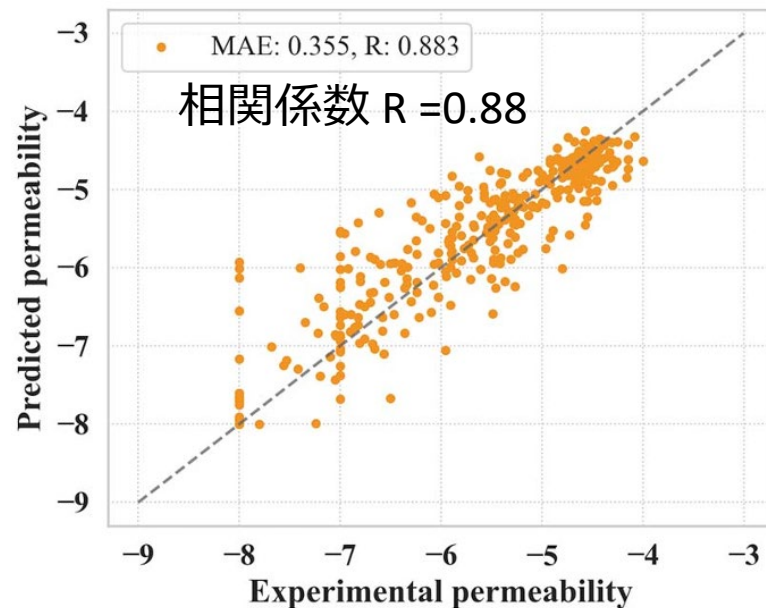
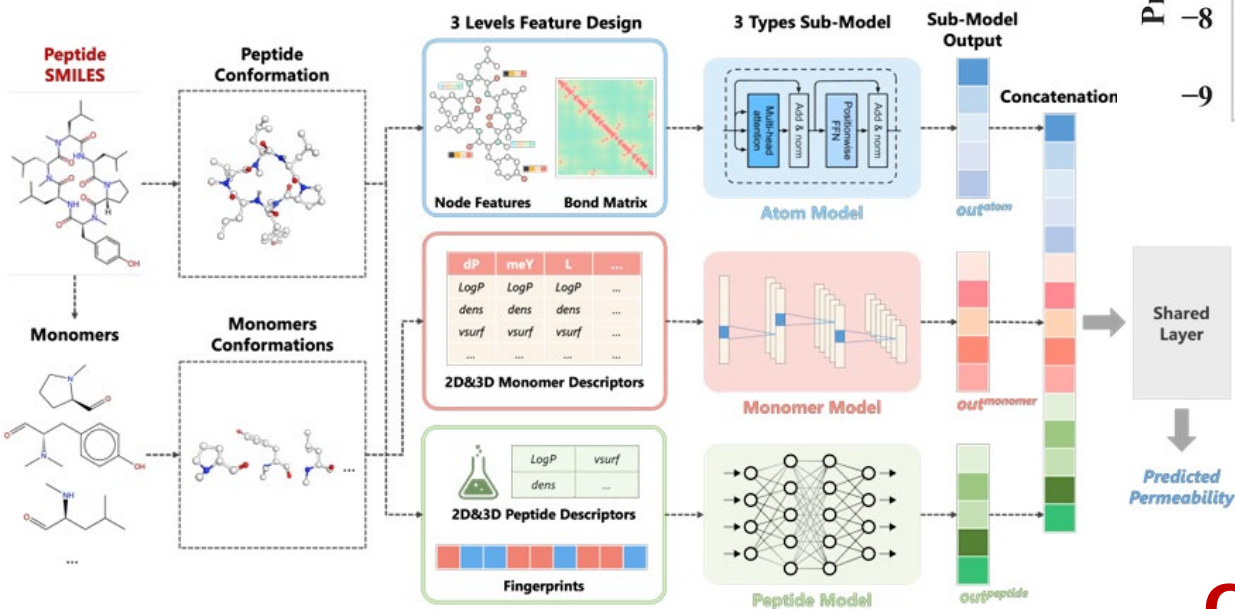
Jianan Li¹, Keisuke Yanagisawa^{1,2}, Yutaka Akiyama^{1,2,*}

2024年8月電子公開

Li J., et al.

Briefings in Bioinformatics, 25(5), 2024

doi: 10.1093/bib/bbae417



Brief. Bioinform, 2024.
doi;

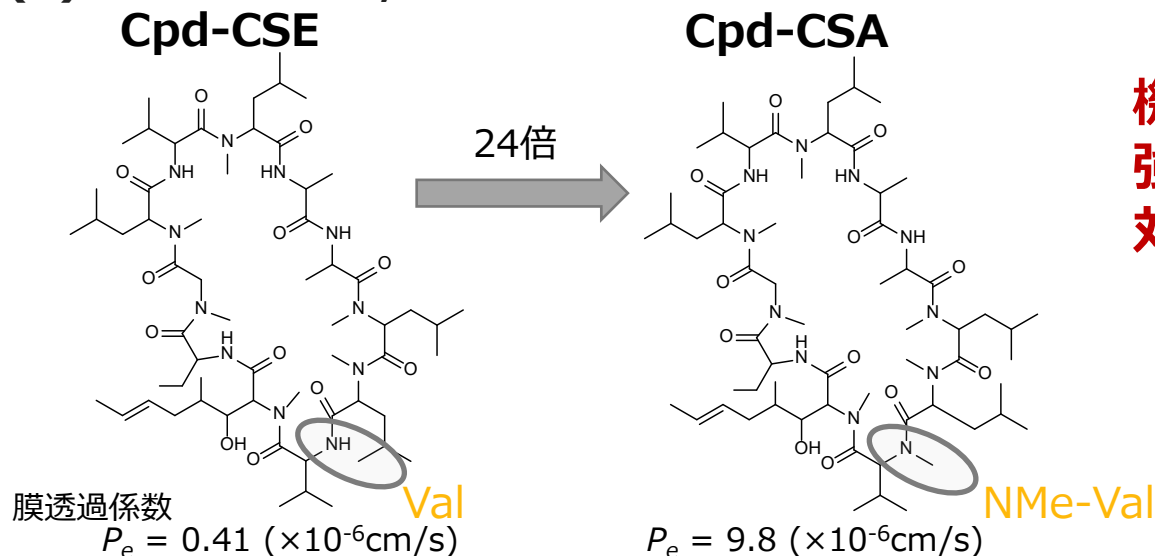
原子レベル、モノマー（残基）レベル、
ペプチド全体レベルの多層構造

**CycPeptMPDBを活用
さらに独自のデータ拡張**

しかし、機械学習を阻む化学修飾の多様性

15

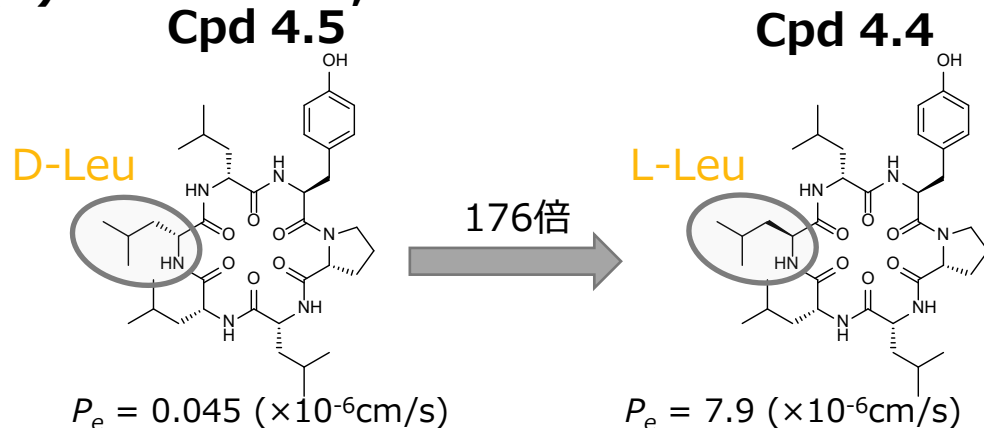
(a) Ahlbach CL, et al. *Future Med. Chem.* 2015



機械学習は設計シリーズ内では強力だが、新規パターンにまで対応できるとは限らない・・・

N-メチル化だけで
24倍 の膜透過率

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



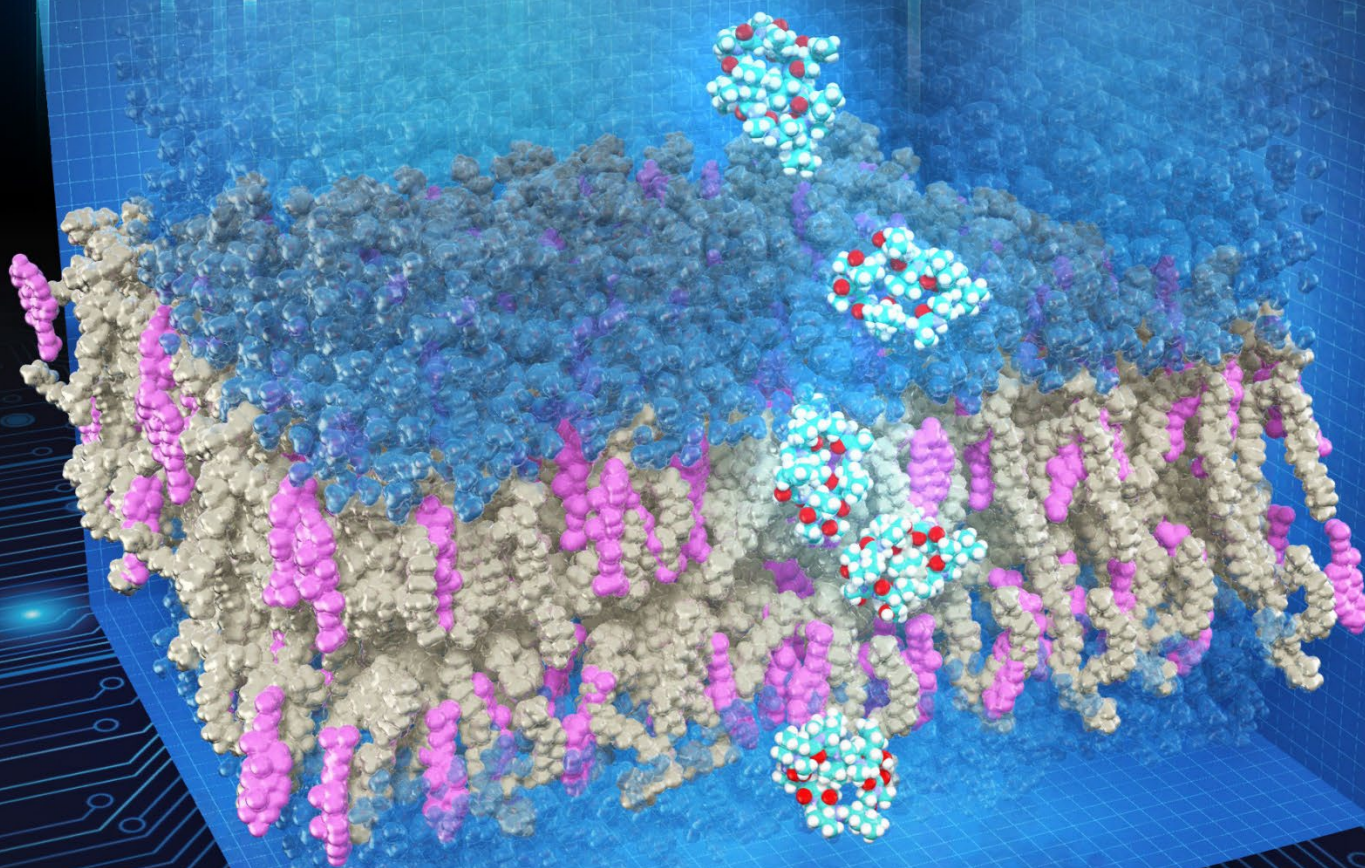
L/D 光学異性体で
176倍 の膜透過率



3. 大規模シミュレーション による細胞膜透過性予測

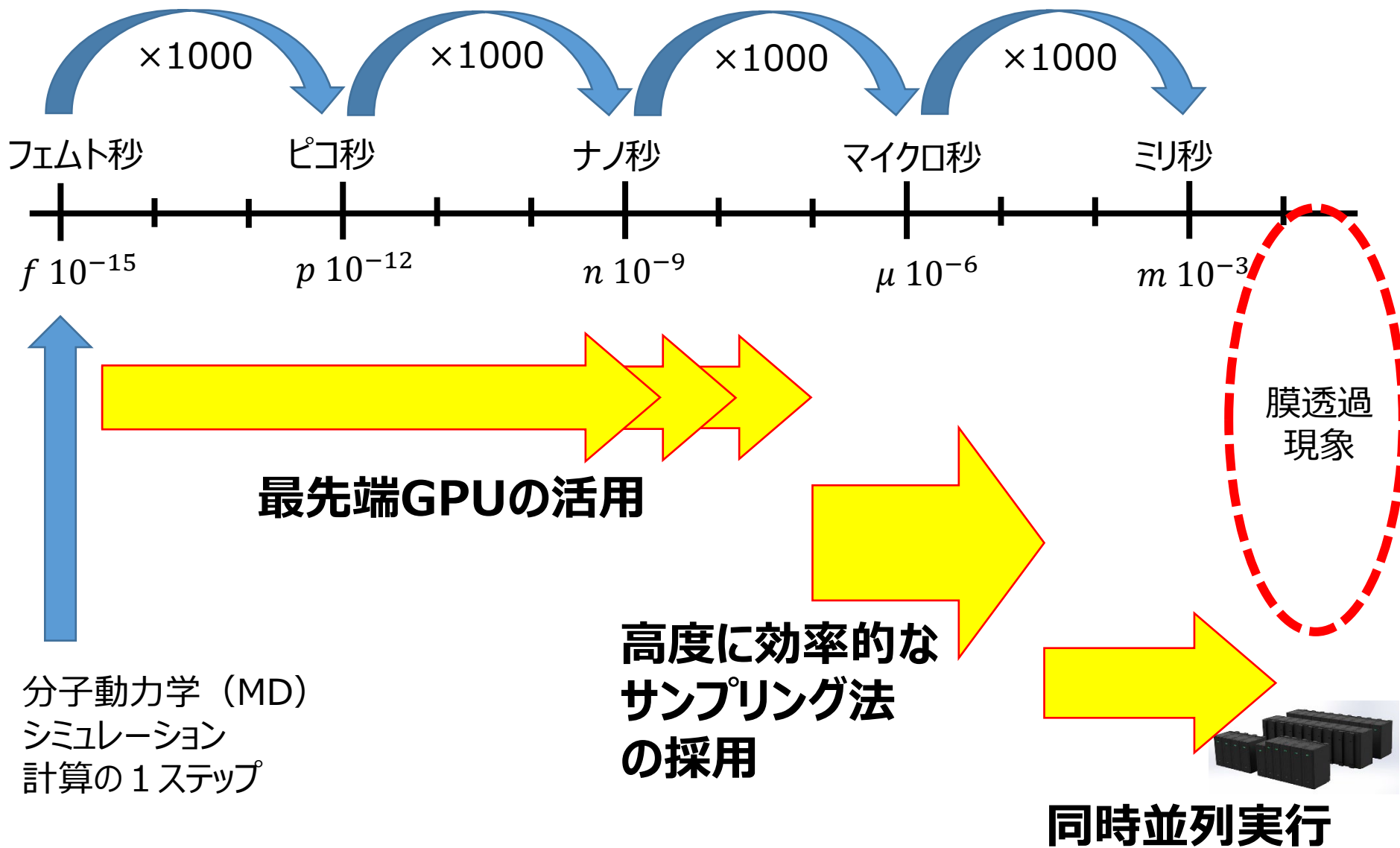
Sugita M. et al.
J. Chem. Inf. Model. 61(7), 2021.
doi: 10.1021/acs.jcim.1c00380

Sugita M. et al.
J. Chem. Inf. Model. 62(18), 2022.
doi: 10.1021/acs.jcim.2c00931



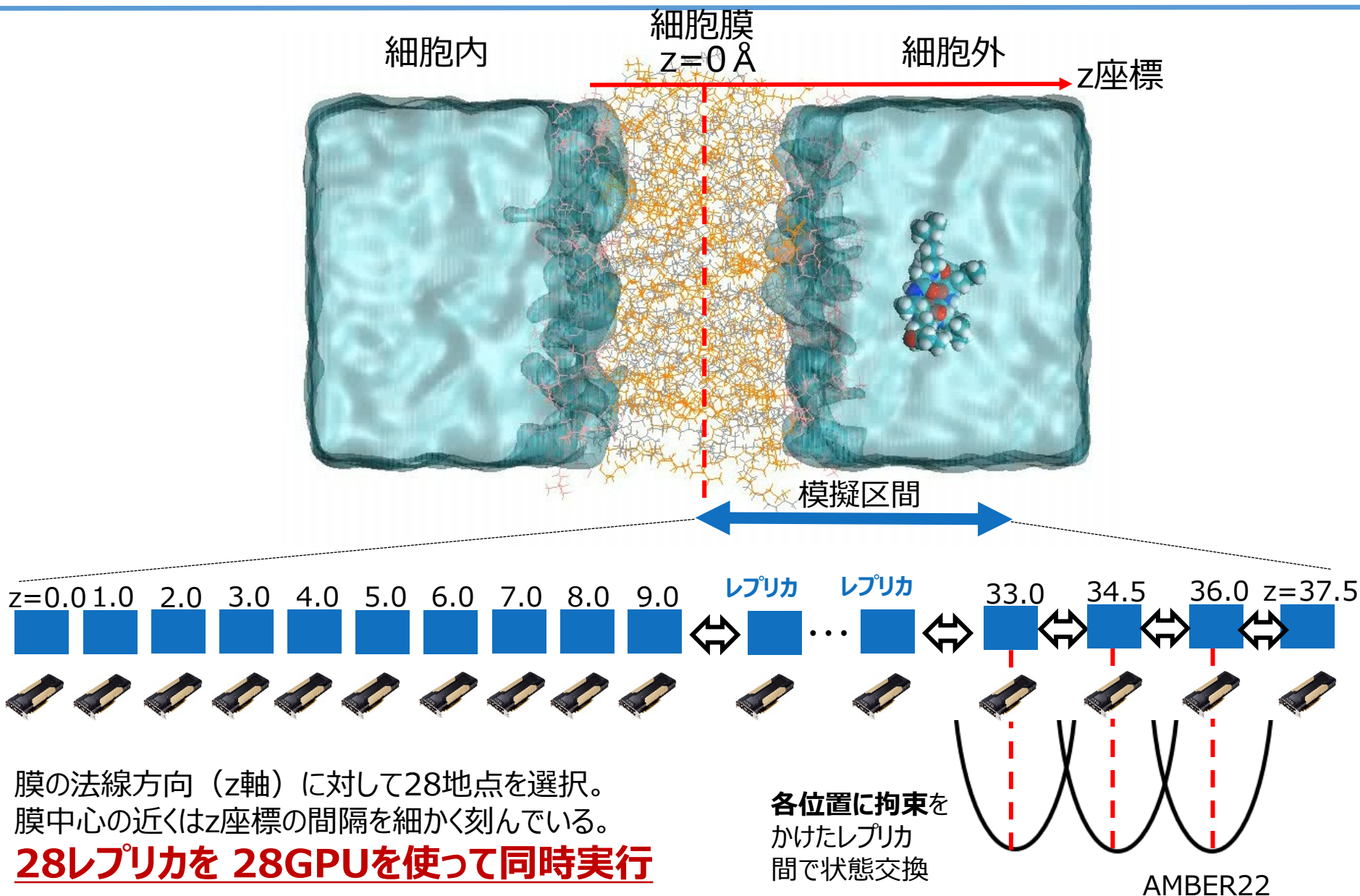
「膜透過」の時間スケールにどう到達するか？

18



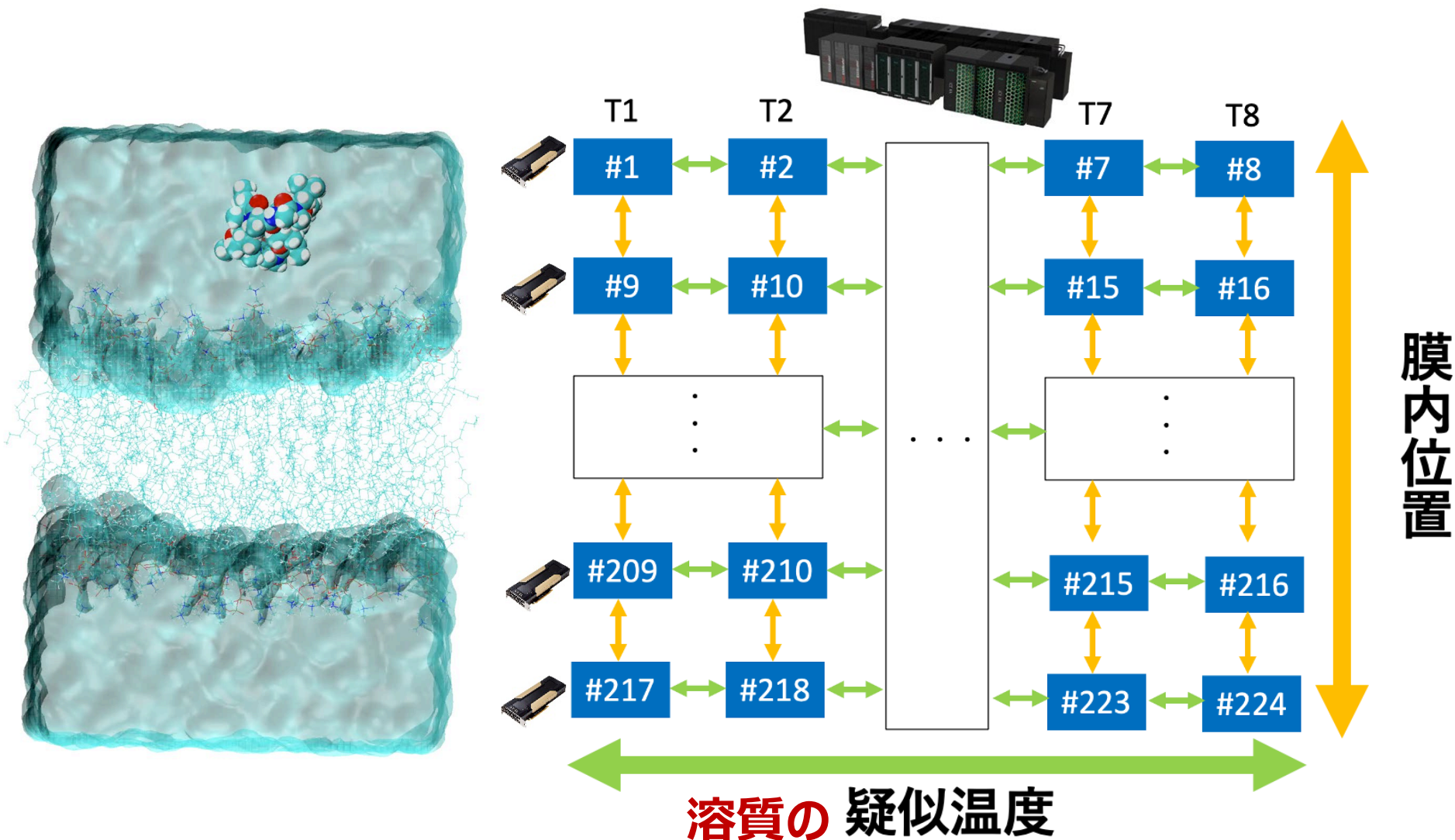
様々な位置での計算を同時並列実行するサンプリング法 (REUS)

19



より高度なサンプリング法への拡張 (REST/REUS)

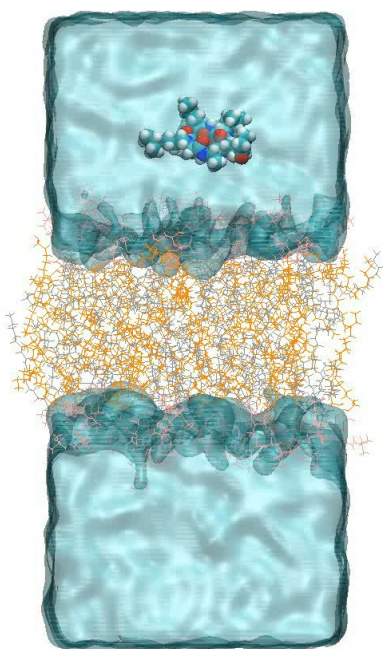
20



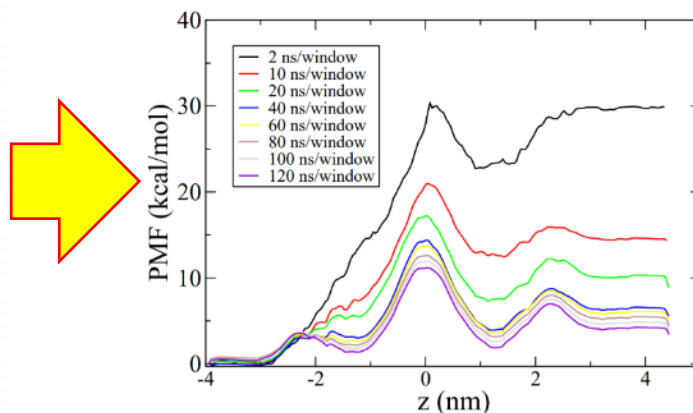
各ペプチド毎に、
224GPU × 約90時間 (P100) で 約500ns のシミュレーションを実施

シミュレーションで得たエネルギー地形から膜透過率を算出

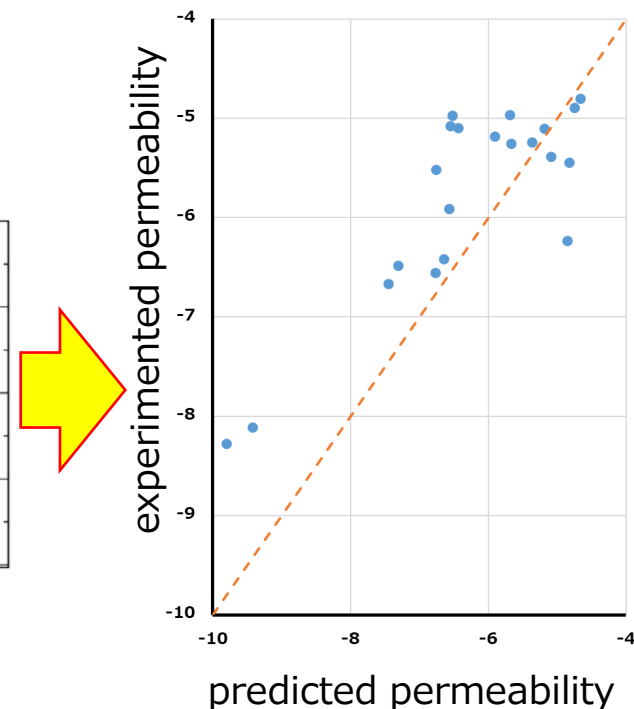
21



細胞膜透過を再現した
分子シミュレーションを
スパコン上で実施する



稀にしか到達しない状態をつなげる
傘サンプリング法とWHAMにより
平均力ポテンシャルPMF曲線を推定



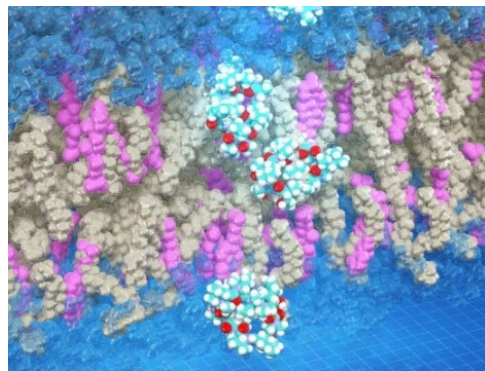
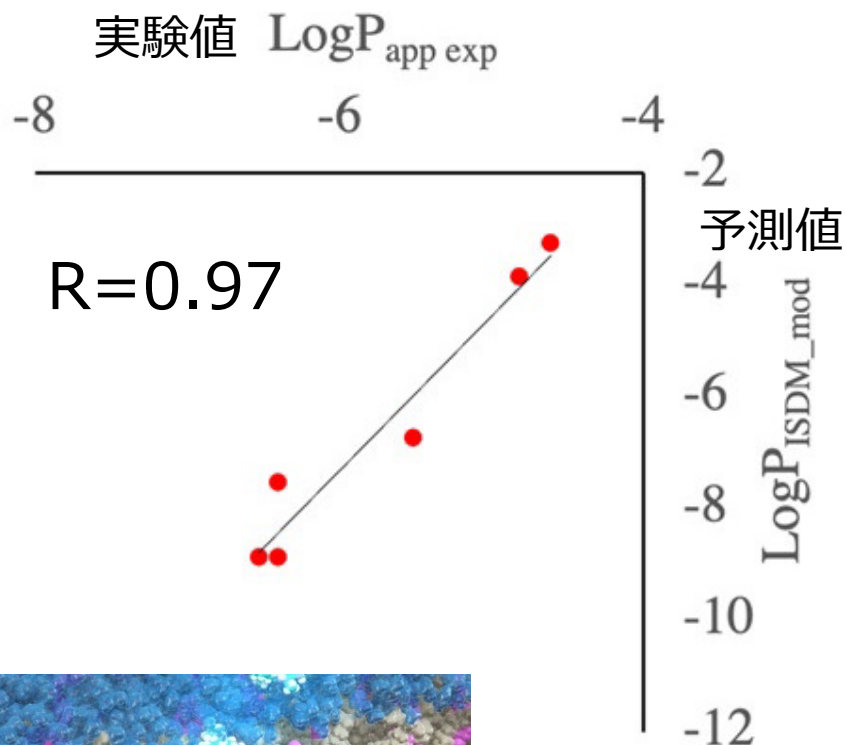
$$\frac{1}{P} = \int_{z_1}^{z_2} \frac{\exp[\beta w(z)]}{D(z)} dz$$

PMF推定値を利用しながら
上式右辺の積分を計算すると
膜透過率 (P) が求められる

予測値と実験値の相関係数 0.8~0.9程度 (2022)

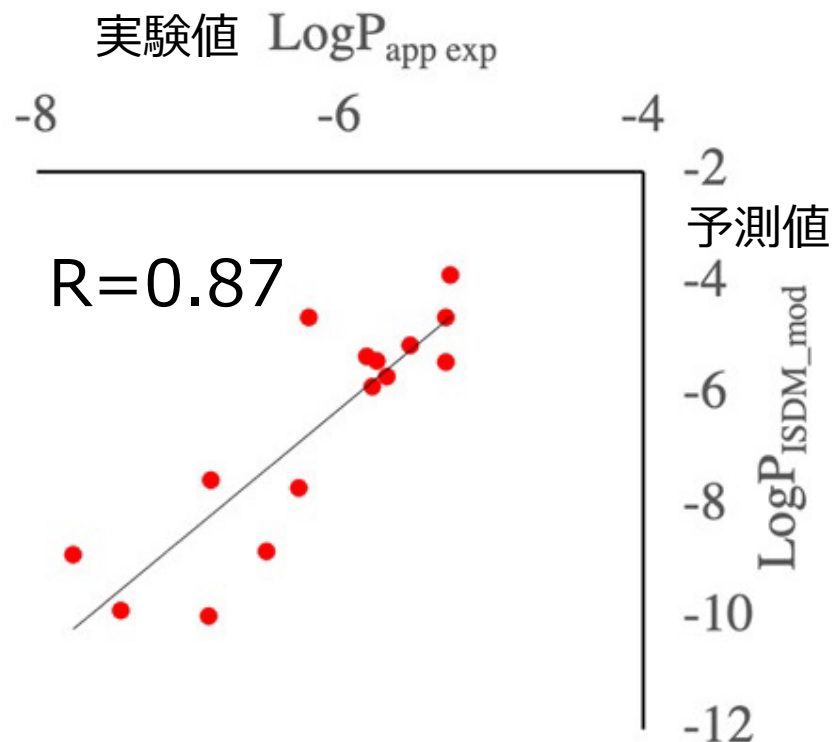
22

MDCK (イヌ腎臓由来)



膜モデルの精緻化
コレステロール
50 mol %
を添加した

PAMPA (人工膜)



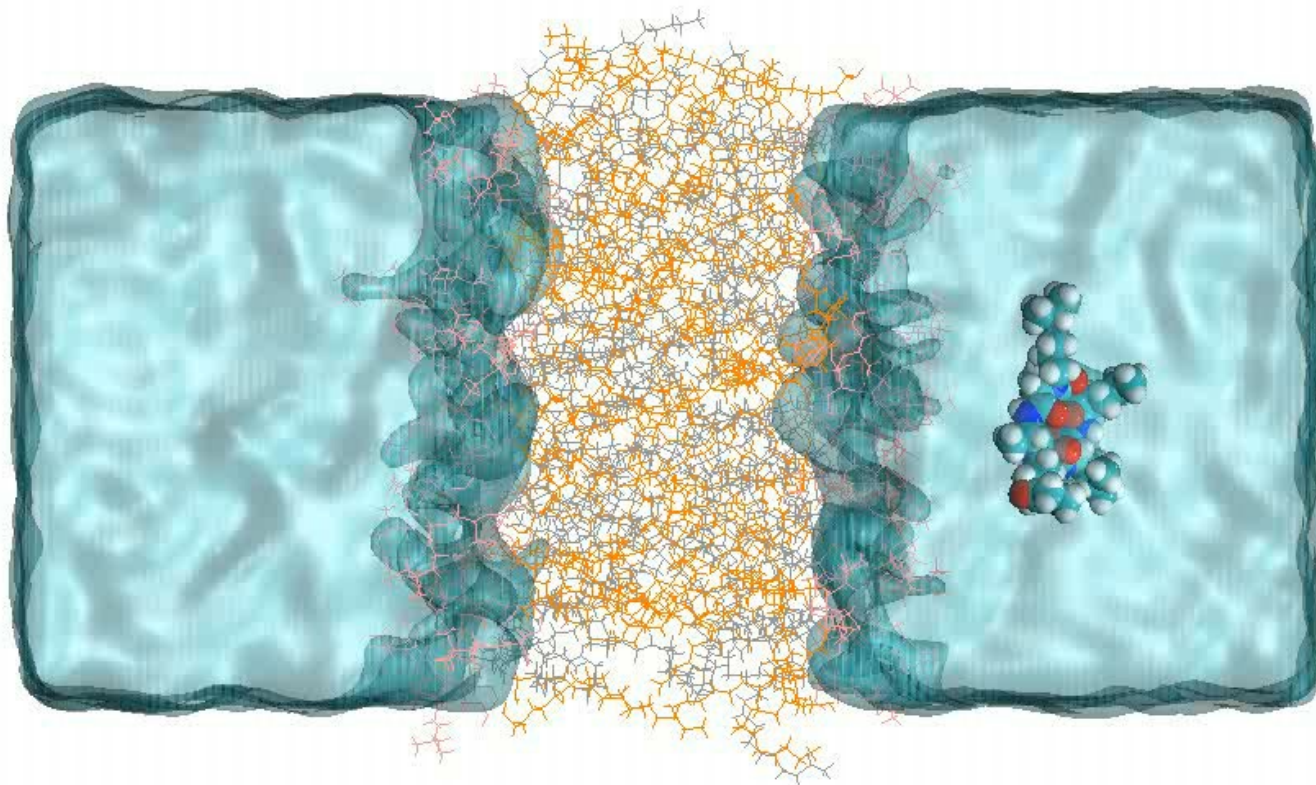
10残基ペプチド 18種類について
膜透過率実験値との相関を示した
(Sugita et al. 2022)
この論文だけで70万GPU時間



シミュレーションを通じて「見える」もの

23

VideoMach unregistered



Amide-to-ester substitution as a stable alternative to *N*-methylation for increasing membrane permeability in cyclic peptides

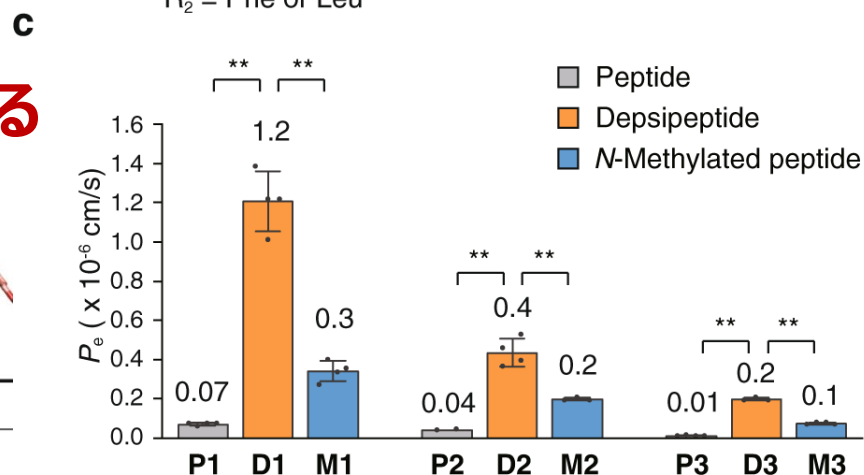
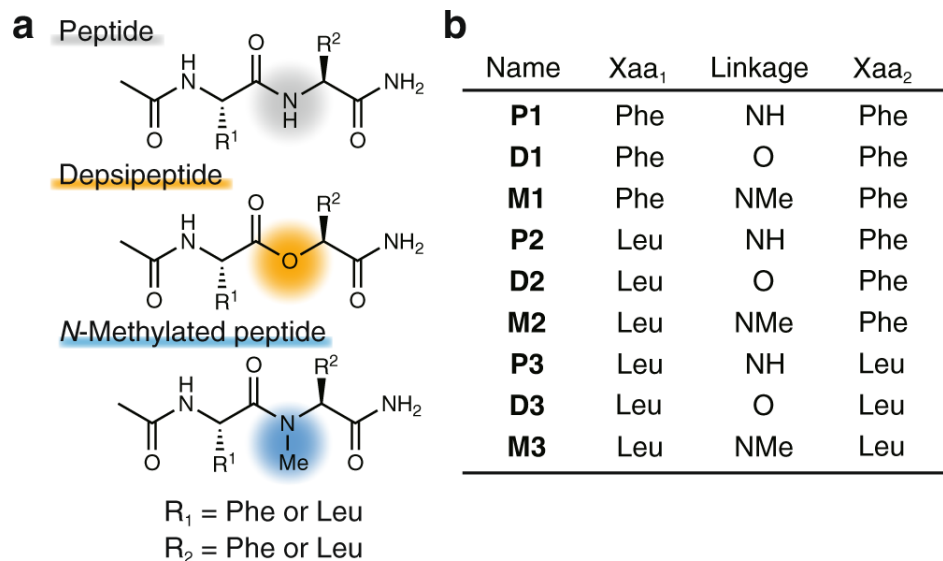
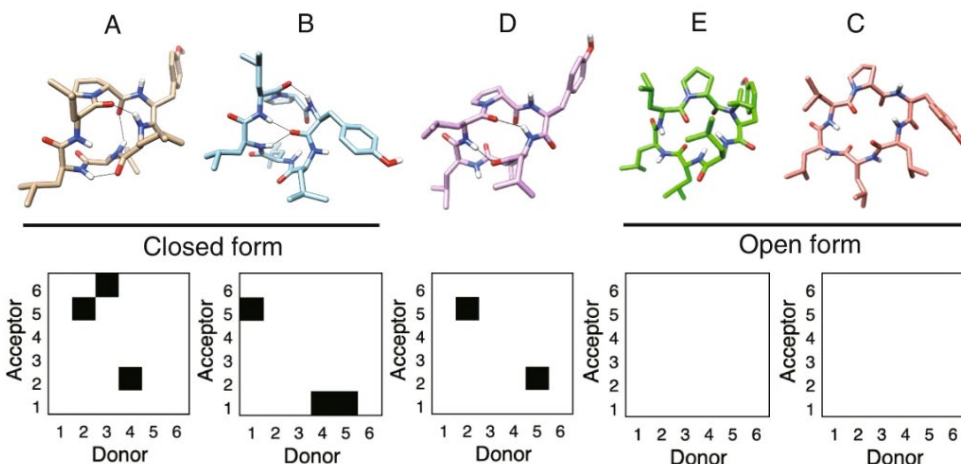
Hosono Y. *et al.*

Nat Commun 14(1416), 2023.

doi: 10.1038/s41467-023-36978-z

東大 山東信介教授、UCSC Prof. Scot Lokeyら

環状ペプチドの主鎖の一部をエステル化すると膜透過性が向上する



7~9残基 環状ペプチド

- 従来研究は「水中のふるまい」「油性環境中のふるまい」を単独に調査。
我々は**脂質二重膜と水の層を詳細にモデル化して全原子シミュレーション**
- 最新の計算プロトコルでは、細胞膜と鉛直方向（z軸）に**28地点**を取り、各点に**8種類の溶質温度**を設けた**REUS/REST法による二次元レプリカ交換**分子動力学シミュレーションを実施している。**28×8=224レプリカ必要**
- TSUBAME3.0では、**224GPUを用いて並列計算**させていた。



P100 GPU

TSUBAME3.0（東京工業大学、当時） 2024年3月末でリプレイスされた



4. MPS (Multi-Process Service)による 大規模シミュレーションの効率化



当事業の最新成果 ICPP'24にて発表

27

- ICPP'24国際会議に論文採択、2024年8月 Swedenにて発表
- 筑波大学計算科学研究センター、アヘッド・バイオコンピューティング株式会社、
東京工業大学、NVIDIAの共同研究成果
(当時)



Improving Performance on Replica-Exchange Molecular Dynamics Simulations by Optimizing GPU Core Utilization

Authors: [Taisuke Boku](#), [Masatake Sugita](#), [Ryohei Kobayashi](#), [Shinnosuke Furuya](#), [Takuya Fujie](#), [Masahito Ohue](#), [Yutaka Akiyama](#) | [Authors Info & Claims](#)

ICPP '24: Proceedings of the 53rd International Conference on Parallel Processing • Pages 1082 - 1091

<https://doi.org/10.1145/3673038.3673097>

Published: 12 August 2024 [Publication History](#)

**MPSにおける GPUコア利用率を適切に調整して
224並列MDシミュレーションを約2倍高速に実行**



H100搭載機の先駆け 筑波大学 Pegasus

28



アヘッド・バイオコンピューティング
藤江、杉田、大上、秋山



筑波大学CCS
朴泰祐先生、堀優太先生



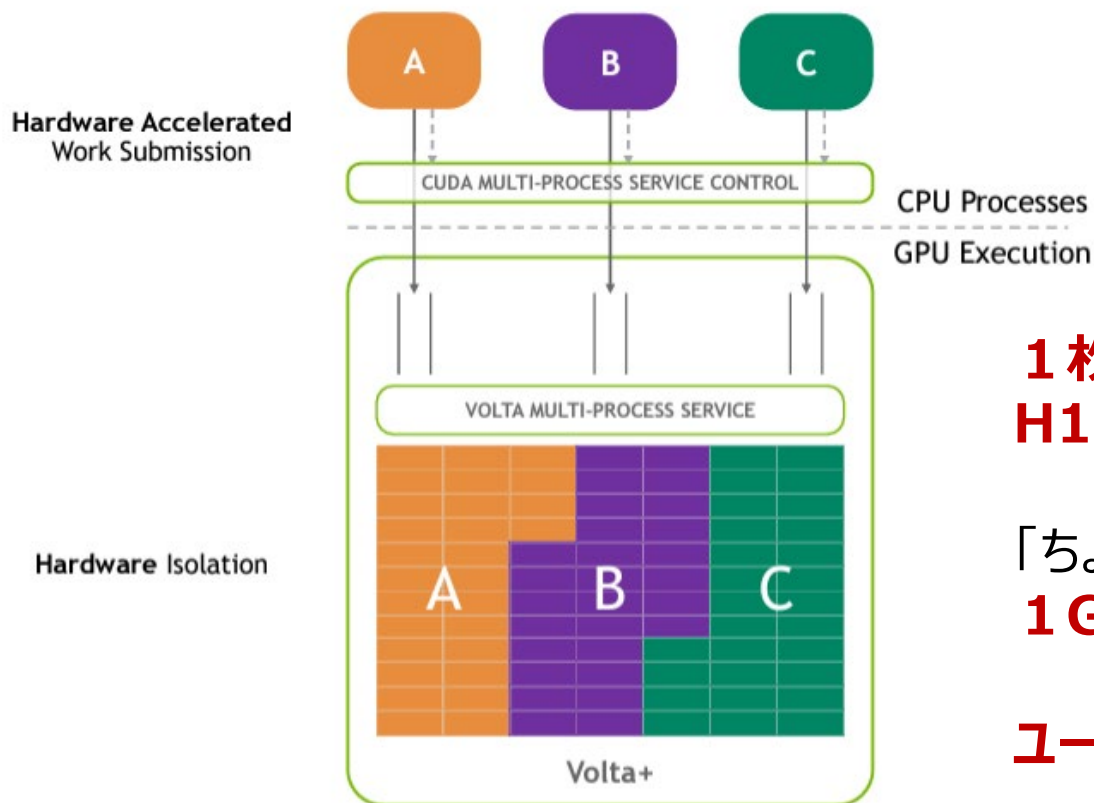
エヌビディア
古家真之介氏

2023年1月稼働 NVIDIA H100 GPU 120枚搭載

(撮影：2023.12.6 筑波大学)

MULTI-PROCESS SERVICE (MPS)

29



MPSは V100時代から存在したが、益々重要になっている

**1 枚のGPUに 1 つの計算では
H100 GPUの能力が大幅に余る**

「ちょうどよく、忙しくなるくらいに」
1 GPUに複数プロセスを割り当てる

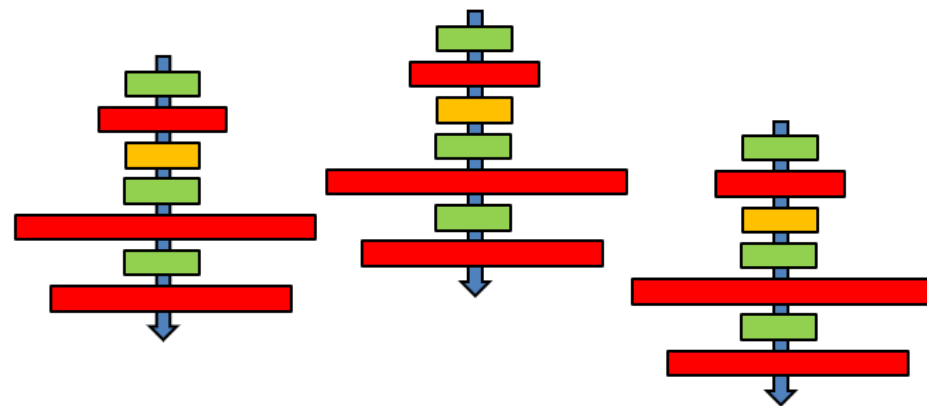
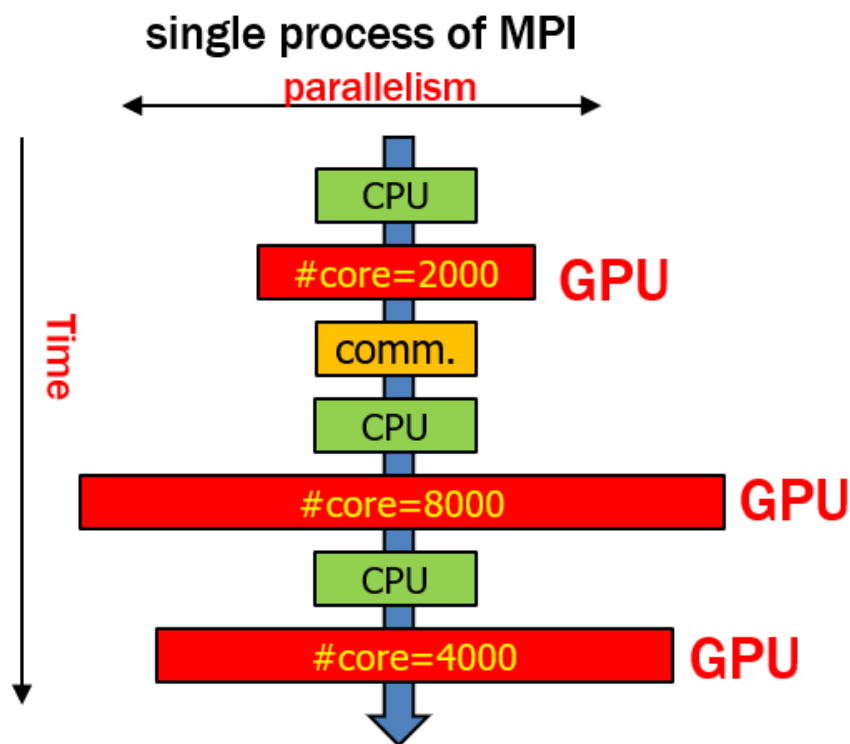
ユーザ自身が投入をコントロールする

CUDA MPS ACTIVE THREAD PERCENTAGE

パラメータの値を各プロセスごとにどう設定するかが鍵

便宜上、CMATPと略記します

- MPS内の各プロセスが、コアの何%までを使うことを許すかを定めるパラメータ
- **default = 100%** （あわよくば、コアの全部を使おうとしてしまう貪欲設定）
- このような貪欲な設定は、相当に良くない結果をもたらすのではないか？
しかし、 N プロセスだから $1/N$ ずつに設定するというのでは消極的すぎる

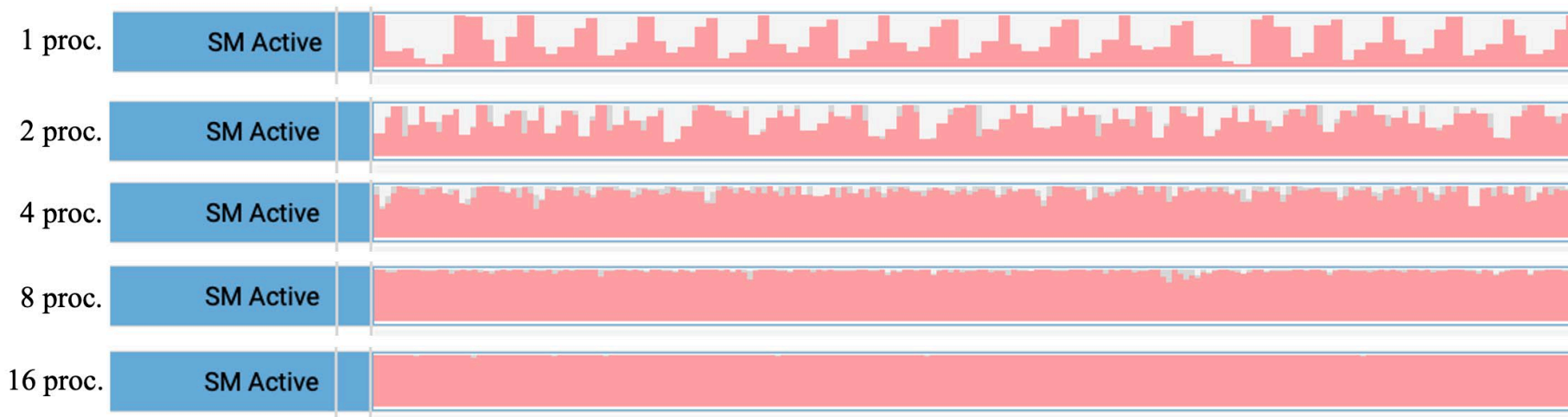


各プロセスがどの時点で
多くのGPUコアを使いたいのか
事前に正確に予測することは困難

Nsight systemsでコア利用を観測できる

31

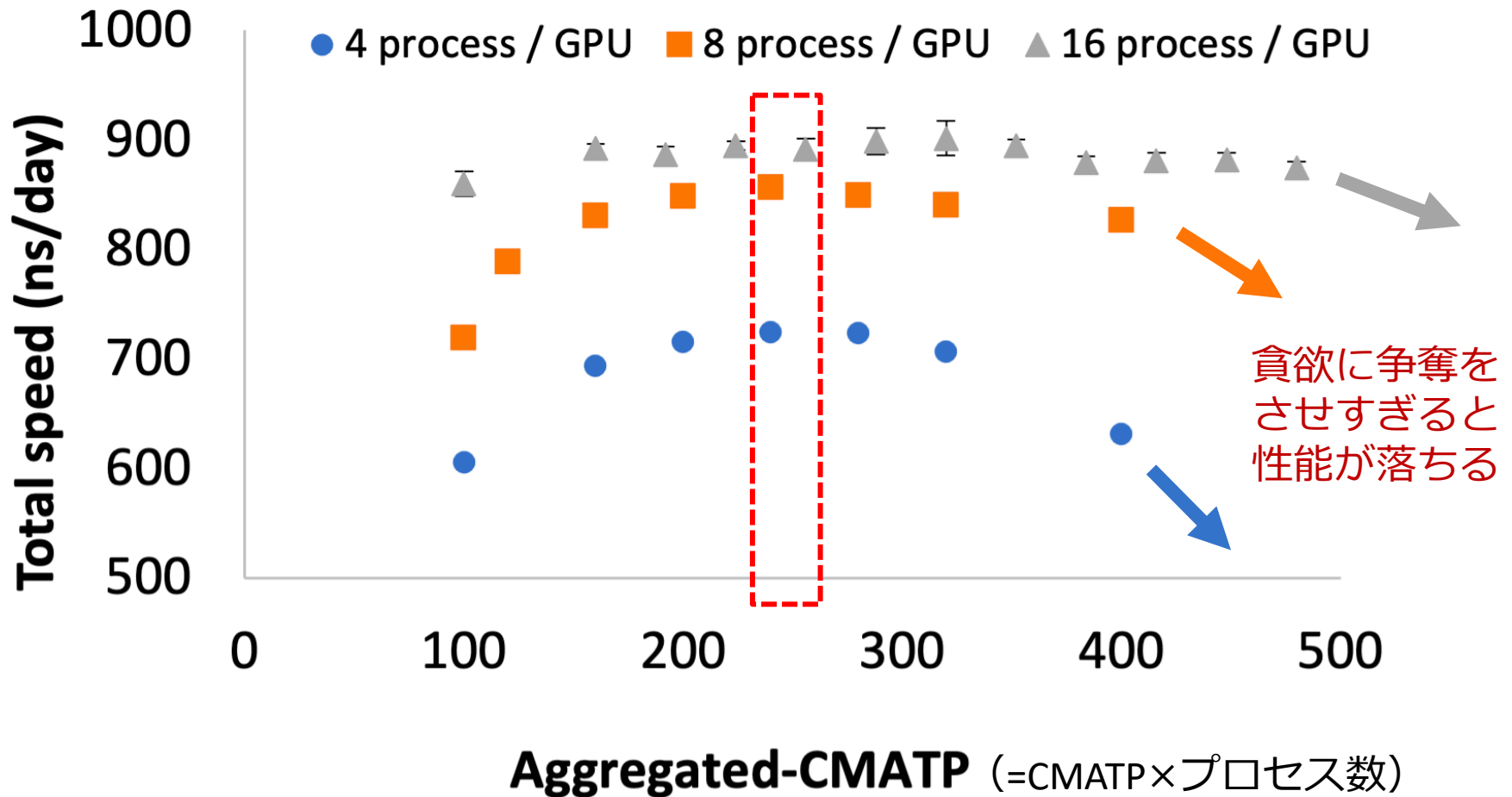
- **NVIDIAが提供するNsight systemsで「SM Active」（コア利用率）を観測**
- MPSに投入するプロセス数を増やすと、コア利用率が上昇する様子が確認できる
- 貪欲な組み合わせ（CMATP=100%）とすると、たしかにコアは埋まるが、処理が全体として効率的に進んでいるのか否かはこれを見ただけでは判らない。



AMBER22, この実験では、各CMATP = 100%（最も貪欲な設定）としている

累計CMATPを240%前後に設定する！

32



- **CMATP60% × 4プロセス、30% × 8プロセス、15% × 16プロセス**等に設定
- 応用により最適値は異なるはずだが、3～5万原子級のMDでは同傾向なのは

適切な設定により、効率は約2倍となった

33

224レプリカの大規模MD計算を実施

# GPU	14	28	56	112
# Proc/GPU	16	8	4	2
CMATP	15	30	60	100
Aggregated-CMATP	240	240	240	200
Speed/GPU (ns/day)	889.76	830.48	722.92	465.70
Efficiency (%)	100	93.3	81.2	53.6

ここが
重要→

効率 約2倍

- 16プロセス/GPU × 14GPU同時利用は
2プロセス/GPU × 112GPU同時利用よりも効率は約2倍
- ただし、多数のGPUを同時利用するほうが当然ながら結果を得るまでの時間は早い
したがって、8プロセス/GPU×28GPU同時利用（効率93%）は検討に値する。

- ペプチド創薬は、次世代創薬モダリティとして大きく期待される。
ただし、細胞膜透過性を与えることが大きな技術的課題となっている。
- AI（深層学習）のアプローチも続けているが、高速なGPUスパコンを活用した大規模分子動力学（MD）シミュレーションの手法を我々は開発した。膜透過性の**実験値との相関係数が0.8**を超えている。
224プロセスが並列計算する。
- 新世代のH100 GPUは工夫して使う必要がある。MPSにおける累計CMATP値を 240% 前後に設定すると**実行効率が約2倍向上**した。

東京科学大学の研究に関する謝辞（項目1～3）

35

- 文部科学省 イノベーションシステム整備事業 地域イノベーション・エコシステム形成プログラム「**IT創薬技術と化学合成技術の融合による革新的な中分子創薬フローの事業化**」（国立大学法人東京工業大学、川崎市、平成29年度採択地域）
- 科研費 基盤研究（B）「**中分子創薬に適した特性を有する環状ペプチド分子設計手法の開発**」17H01814（研究代表者 秋山泰）
- 科研費 基盤研究（B）「**部分構造の重複を利用した大規模化合物データベース向けバーチャルスクリーニング手法**」22H03684（研究代表者 秋山泰）
- 科研費 基盤研究（B）「**タンパク質間相互作用を標的とする薬剤分子設計技術の開発**」20H04280（研究代表者 大上雅史）
- 日本医療研究開発機構(AMED)「**スーパーコンピュータ資源及び大規模シミュレーションとAIに基づく創薬・生命科学の支援**」JP22ama121026
- 科学技術振興機構（JST）次世代研究者挑戦的研究プログラム「**殻を破るぞ！越境型理工系博士人材育成—総合知の創造と社会変革に貢献する**」JPMJSP2106
- 産業技術総合研究所（ABCIグランドチャレンジ採択）
- 東京工業大学（TSUBAMEグランドチャレンジ採択）
- さくらインターネット株式会社（無償共同研究）
- ペプチドリーム株式会社（無償共同研究）
- 塩野義株式会社（無償共同研究）
- 富士通株式会社（有償共同研究）



文部科学省



Regional
Innovation Ecosystems



AMED



アヘッド社の研究に関する謝辞（項目4）

36

- MPSの最適な利用条件の研究は、ICPP'24国際会議論文として2024年8月に発表済で、筑波大学計算科学研究センター、アヘッド・バイオコンピューティング株式会社、東京工業大学、NVIDIAの共同研究成果です。

Improving Performance on Replica-Exchange Molecular Dynamics Simulations by Optimizing GPU Core Utilization

Authors: [Taisuke Boku](#), [Masatake Sugita](#), [Ryohei Kobayashi](#), [Shinnosuke Furuya](#), [Takuya Fujie](#), [Masahito Ohue](#), [Yutaka Akiyama](#) | [Authors Info & Claims](#)

ICPP '24: Proceedings of the 53rd International Conference on Parallel Processing • Pages 1082 - 1091
<https://doi.org/10.1145/3673038.3673097>

Published: 12 August 2024 [Publication History](#)



- 2023年度 文部科学省 共同利用・共同研究システム形成事業「学際領域展開ハブ形成プログラム」（採択テーマ名：AI時代における計算科学の社会実装を実現する学際ハブ拠点形成）の支援を受けています。