

令和 6 年 5 月 20 日現在

機関番号：11301

研究種目：基盤研究(C)（一般）

研究期間：2021～2023

課題番号：21K06107

研究課題名（和文）「水和」創薬理論の創生

研究課題名（英文）A Construction of Drug Design Theory based on the Hydration of Proteins

研究代表者

吉留 崇 (Yoshidome, Takashi)

東北大学・工学研究科・准教授

研究者番号：90456830

交付決定額（研究期間全体）：（直接経費） 3,200,000円

研究成果の概要（和文）：まず、タンパク質水和分布を高速かつ正確に予測する深層学習モデル「gr Predictor」を完成させ、論文を出版するとともに、GitHubで公開した。次に、gr Predictorを発展させ、グリッド不均一溶液理論(Grid-Inhomogeneous Solvation Theory, GIST)で得られるタンパク質周りの水和自由エネルギー分布を高速に予測する深層学習モデル「Deep GIST」を開発した。さらに、gr PredictorとDeep GISTを用いて、リガンド結合部位の水分子の自由エネルギー解析を行い、リガンド結合に伴い移動する水分子の特徴の一部を明らかにした。

研究成果の学術的意義や社会的意義

開発したgr PredictorやDeep GISTの創薬への応用が期待できる。既存のドッキングソフトは高速にリガンド結合部位を予測できる一方、予測精度が高くないことが問題である。その原因の1つとして水を顧みず考慮していないことが挙げられる。gr PredictorやDeep GISTを導入することで、計算速度はそのまま予測精度が向上することが期待される。

研究成果の概要（英文）：First, we completed the implementation of a deep-learning model that fast and accurate prediction of the hydration structures around proteins is possible. Our deep-learning model, referred to as the “gr Predictor”, was shared with the world. Then, by developing gr Predictor, we also implemented a deep-learning model, referred to as the “Deep GIST”, enabling us to predict the distributions of hydration-free energy around proteins. Furthermore, free-energy analysis of water molecules at the ligand-binding pocket was performed using gr Predictor and Deep GIST, and a feature of water molecules replaced with ligand upon ligand binding was elucidated.

研究分野：生物物理学

キーワード：水和 深層学習 積分方程式理論

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

これまでに様々なリガンド結合予測ソフトウェアが提案されてきたが、予測精度が高くないことが問題となっている。このため、現在は、例えばリガンド結合予測ソフトウェアで得られた複数のリガンド結合構造の中から目視でリガンド結合構造を選んでいる。予測精度が高くない原因の一つとして、既存のリガンド結合予測ソフトウェアでは水を顕に考慮していないことが挙げられる。

リガンド結合予測に水を顕に考慮することを阻むのは、既存のタンパク質周りの水の分布(水和分布)を計算する手法(分子動力学(MD)シミュレーションや溶液理論(3D-RISM理論))では、タンパク質1構造あたり数時間程度の計算が必要であることである。リガンド結合予測では膨大な量のタンパク質-リガンド複合体を扱う必要があるため、既存の水和分布を計算する手法はリガンド結合予測に向かない。

一方で、申請者(吉留)らは、タンパク質水和分布を高速に予測する深層学習モデル「gr Predictor」を開発した。gr Predictorは、水の酸素原子の分布を1分程度で予測することができる。また、3D-RISM理論で得られた水の酸素原子の分布を定量的に再現できる。よってgr Predictorは高速かつ高精度に水の酸素原子の分布を予測することが出来る。このため、gr Predictorを用いれば水を顕に考慮したリガンド結合予測が可能になるのではないかと考えた。

### 2. 研究の目的

水を顕に考慮したリガンド結合予測法を開発することが研究の目的である。

### 3. 研究の方法

#### (1) gr Predictor

gr Predictorの公開に向け、ハイパーパラメータサーチ、水分子の位置の予測精度の議論、水の水素原子の分布を予測するモデルの開発、を行った。また、予測結果を詳細に解析し、予測精度が悪い箇所とその原因を明らかにするとともに、その箇所の訓練データを増やすことによって精度向上するかどうかを調べた。さらに、これまではタンパク質単量体でのみ予測精度を議論していたが、複合体でも予測できるかどうかを検討した。

#### (2) 水和熱力学量の計算法の開発

リガンド結合予測を行うためには、リガンド結合に伴う自由エネルギー変化である「結合自由エネルギー」を計算する必要がある。これを高速に計算することを実現するために、gr Predictorを改良することによって、水和熱力学量(水和自由エネルギー、水和エネルギー、水和エントロピー)を高速に計算する深層学習モデルの開発を行った。MDシミュレーションを用いて水和熱力学量を計算する手法の一つであるグリッド不均一溶液理論(Grid-Inhomogeneous Solvation Theory, GIST)[1]に着目し、gr Predictorを改良することにより、水和熱力学量を高速で計算する深層学習モデルの研究に取り組んだ。まず、78個のタンパク質の水和熱力学量の分布を水和熱力学量の分布をMDとGISTを用いて求めた。次に、68個のタンパク質の水和熱力学量分布を訓練データとして用いて、深層学習モデルを構築した。モデル構築の際、ハイパーパラメータサーチを行った。最後に残り10個のタンパク質を用いて予測を行い、正解のデータと比較した。比較は、水和自由エネルギー分布を用いて行った。開発した深層学習モデルには「Deep GIST」と名付けた。

また、水和エントロピーを定量的に計算する手法の開発を行った。水和エントロピーを正確に計算できる理論は分子性流体用積分方程式理論[2]であるが、その理論の数学的複雑さのためにタンパク質のような複雑な形状の溶質には適用できない。そこで、複雑な形状の溶質の溶媒和エントロピーの計算が可能である剛体球+レナード-ジョーンズ・ポテンシャルの6乗引力のポテンシャルを有する単純流体モデルを用いて、分子性流体用積分方程式理論の結果を定量的に再現する手法の開発を行った。

#### (3) リガンド結合予測に向けた取り組み

水を顕に考慮したリガンド結合予測に向け、gr Predictorと(2)で開発した水和熱力学量計算法を用いて、リガンド結合部位に存在する水分子の自由エネルギー解析を行った。創薬への応用に向け、リガンド結合部位に存在する水分子をタンパク質表面からバルク領域に移動するのに必要な自由エネルギー変化を計算した。具体的には、まず2718個程度のタンパク質の水和分布をgr Predictorで求めた。次に、得られた水和分布とプログラム Placevent[3]を用いて水分子をタンパク質表面に置いた。最後に、Deep GISTを用いて、リガンド結合部位の水分子に対し、結合部位から移動するのに必要な自由エネルギー変化を計算した。

また、2つのドッキングソフトを組み合わせることでリガンド結合予測精度が向上するかにつ

いても研究した。まず、2つのドッキングソフトを816個のタンパク質に対して適用し、リガンド結合ポーズを正しく予測できるタンパク質の種類が異なるかどうかを調べた。もし異なるのであれば、2つのドッキングソフトを組み合わせることでリガンド結合予測を正しく行うことができるタンパク質の数が増えるので、組み合わせる方法について研究した。

最後に、タンパク質立体構造予測を高速に行うAlphaFold2 [4]を用いたリガンド結合予測に向け、予測構造の側鎖の向きについて検討した。具体的には、低温電子顕微鏡実験で構造が既知のタンパク質を用意し、その立体構造をAlphaFold2で予測した。その後、実験構造と側鎖の向きについて議論した。

#### 4. 研究成果

##### (1) gr Predictor の公開と展開

gr Predictor のハイパーパラメータを行い、最適なハイパーパラメータを決定した。その後、2696個のタンパク質の水和分布を予測したところ、決定係数は0.98以上であることがわかった。すなわち、水和分布を高精度に予測できることがわかった。次に、プログラムPlaceventと予測結果を用いて水分子をタンパク質表面に置き、結晶水の位置と比較することにより、水分子の位置をどの程度正確に予測できるかを調べた。その結果、結晶水の位置からのずれは1-2Åであり、この程度の精度で水分子の位置を予測できることがわかった。最後に、水の水素原子の分布を予測する深層学習モデルを構築した。予測精度は水の酸素原子の分布の予測精度と同程度だった。以上の結果を論文としてまとめ、論文の投稿と同時にgr PredictorのプログラムをGitHub [5]で公開した。論文は2022年に出版した。

その後、gr Predictorの予測結果を解析した。その結果、システイン残基付近の水和の精度が悪いことがわかった。そこでシステイン残基付近の水和の訓練データを増やしたところ、gr Predictorの予測精度が向上した。

さらに、2量体界面の水和分布を「gr Predictor」で予測できるかを調べた。gr Predictorの訓練にはタンパク質の単量体のみを用いていたが、2量体界面の水和分布を高精度で予測できることが明らかとなった。これは2量体界面の構造が単量体の表面と類似していることを意味している。

##### (2) 水和自由エネルギー分布を高速に予測する深層学習モデル「Deep GIST」の開発

Deep GISTの水和自由エネルギー分布の予測精度を決定係数で評価したところ、決定係数の値は平均0.75だった。このため、高精度で予測できる箇所とそうでない箇所が見られた。次に、リガンドが結合する際にタンパク質から移動する水分子の自由エネルギーを計算した。Deep GISTとGISTの両方で行い、結果を比較した。その結果、Deep GISTで得た水分子の自由エネルギーのGIST計算結果からの誤差率は、約4%であった。この結果はDeep GISTが高精度でGISTの計算結果を再現できることを意味する。

Deep GISTは、水和自由エネルギー分布を約1分程度で計算することが出来、この計算時間はGIST計算(2日程度)の約1/1000であった。よってDeep GISTは、高速に水和自由エネルギー分布を計算でき、リガンドが結合する際にタンパク質から移動する水分子の自由エネルギーを高精度に計算できることが分かった。

##### (3) 単純流体モデルを用いた水和エントロピーの定量的計算法の提案

分子性流体用積分方程式理論の結果との誤差が最小となるように単純流体モデルのパラメータを決定した。分子性流体用積分方程式理論と形態計測学アプローチで得られた水和エントロピーの値と、単純流体モデルで計算した溶媒和エントロピーの値を比較したところ、誤差は数%であることがわかった。

##### (4) リガンド結合部位に存在する水分子の自由エネルギー解析

創薬への応用に向け、リガンド結合部位に存在する水分子をタンパク質表面からバルク領域に移動するのに必要な自由エネルギー変化を計算した。計算結果を解析するために、リガンド結合部位の水分子を「リガンドが結合する際に移動する水分子」とそうでない水分子に分類したところ、前者の水分子は自由エネルギー変化が小さいものが多いことが分かった。これは、リガンド結合の際に比較的移動しやすいことを意味しており、移動しやすい水分子がリガンド結合の際にリガンド原子に置き換わることが分かった。

##### (5) ドッキングソフトの精度向上方法の提案

ドッキングソフトの精度向上を目指し、2つのドッキングソフトを組み合わせる手法について検討した。ドッキングソフトとして、smiina [6]と畳み込みニューラルネットワークに基づく手法 [7]を用いた。まず、結合ポーズが既知のタンパク質-リガンド複合体に2つのドッキングソフトを適用したところ、2つのドッキングソフトが予測できるタンパク質-リガンド複合体は異なることが分かった。そこで、2つの手法を組み合わせる新たな手法の開発を行った。2つの手法を1つのタンパク質-リガンド複合体に適用すると結合ポーズが2種類得られるので、どちらを選ぶかを定める指標を開発した。その指標を用いて結合ポーズ予測を行ったところ、1つのド

ッキングソフトを用いる場合よりも精度が向上することが分かった。

(6) AlphaFold2 における側鎖の予測精度の議論  
約 500 残基の立体構造が既知のタンパク質に対し、AlphaFold2 で立体構造を予測し、正解の立体構造と比較したところ、約 20% の側鎖の 2 面角が大きくずれていることが分かった (右図)。また、そのずれが 100 ns の MD シミュレーションで補正できるかを調べたところ、一部の側鎖の向きは補正できないことがわかった。したがって、拡張アンサンブル法で修正する必要があると結論づけた。

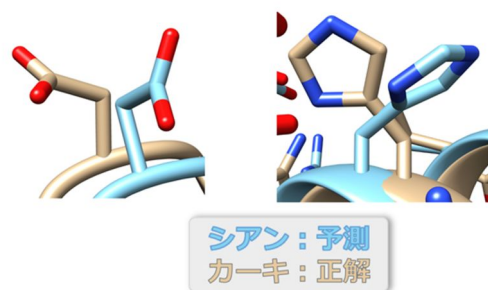


図 1 F<sub>1</sub>-ATPase のβサブユニットの正解と予測。

#### <引用文献>

- [1] C.N.Nguyen, T.Kurtzman Young, and M.K. Gilson, J. Chem. Phys. **137**, 044101 (2012).
- [2] M. Kinoshita, J. Chem. Phys. **128**, 024507 (2008).
- [3] D.J. Sindhikara, N. Yoshida, and F. Hirata, J. Comput. Chem. **33**, 1536 (2012).
- [4] J. Jumper, *et al.*, Nature **596**, 583 (2021).
- [5] <https://github.com/YoshidomeGroup-Hydration/gr-predictor>
- [6] D.R. Koes, M.P. Baumgartner and C.J. Camacho, J. Chem. Inf. Model. **53**, 1893 (2013).
- [7] A.H. Mahmoud, M.R. Masters, Y. Yang and M.A. Lill, Commun. Chem. **3**, 19 (2020).

## 5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 4件/うち国際共著 0件/うちオープンアクセス 0件）

1. 著者名 Yoshidome Takashi	4. 巻 391
2. 論文標題 Mechanism of a high free-energy transduction efficiency of Bacillus PS3 F1-ATPase from the perspective of solvent entropy	5. 発行年 2023年
3. 雑誌名 Journal of Molecular Liquids	6. 最初と最後の頁 123346 ~ 123346
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.molliq.2023.123346	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Shiono Daiki, Yoshidome Takashi	4. 巻 91
2. 論文標題 AlphaFold-predicted Protein Structure vs Experimentally Obtained Protein Structure: An Emphasis on the Side Chains	5. 発行年 2022年
3. 雑誌名 Journal of the Physical Society of Japan	6. 最初と最後の頁 064804-1-4
掲載論文のDOI (デジタルオブジェクト識別子) 10.7566/JPSJ.91.064804	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Yoshidome Takashi	4. 巻 91
2. 論文標題 Simple-Fluid Model for Accurate Reproduction of Hydration Entropy	5. 発行年 2022年
3. 雑誌名 Journal of the Physical Society of Japan	6. 最初と最後の頁 094802-1-6
掲載論文のDOI (デジタルオブジェクト識別子) 10.7566/JPSJ.91.094802	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Kawama Kosuke, Fukushima Yusaku, Ikeguchi Mitsunori, Ohta Masateru, Yoshidome Takashi	4. 巻 62
2. 論文標題 gr Predictor: A Deep Learning Model for Predicting the Hydration Structures around Proteins	5. 発行年 2022年
3. 雑誌名 Journal of Chemical Information and Modeling	6. 最初と最後の頁 4460 ~ 4473
掲載論文のDOI (デジタルオブジェクト識別子) 10.1021/acs.jcim.2c00987	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計24件（うち招待講演 3件 / うち国際学会 2件）

1. 発表者名 吉留 崇
2. 発表標題 タンパク質水和の理論
3. 学会等名 第12回分子モーター討論会（招待講演）
4. 発表年 2023年

1. 発表者名 吉留 崇
2. 発表標題 タンパク質水和自由エネルギー分布を予測する深層学習モデルの開発
3. 学会等名 「ソフトマテリアル理論研究の最前線」研究会
4. 発表年 2023年

1. 発表者名 木村 啓太、吉留 崇
2. 発表標題 タンパク質-リガンド結合様式予測の精度向上法の提案：複数の結合予測法のハイブリッドアプローチ
3. 学会等名 第37回分子シミュレーション討論会
4. 発表年 2023年

1. 発表者名 福島 悠朔、伊藤 祐希、吉留 崇
2. 発表標題 深層学習を用いたGISTマップの高速計算：リガンド結合に伴う水の自由エネルギー変化の計算への応用
3. 学会等名 第61回日本生物物理学会年会
4. 発表年 2023年

1. 発表者名 福島 悠朔、吉留 崇
2. 発表標題 深層学習を用いたグリッドベースの水和熱力学量計算の高速化
3. 学会等名 第23回日本蛋白質科学会年会
4. 発表年 2023年

1. 発表者名 吉留崇
2. 発表標題 Physical Chemistry of Protein Hydration
3. 学会等名 2022年度日本化学会化学系学協会東北大会（招待講演）
4. 発表年 2022年

1. 発表者名 河間 光祐、福島 悠朔、池口 満徳、大田 雅照、吉留 崇
2. 発表標題 gr Predictor : 深層学習を活用したタンパク質水和分布の高速計算法
3. 学会等名 第60回日本生物物理学会年会
4. 発表年 2022年

1. 発表者名 福島 悠朔、吉留 崇
2. 発表標題 深層学習によるグリッドベースの水和自由エネルギー計算
3. 学会等名 第60回日本生物物理学会年会
4. 発表年 2022年

1. 発表者名 福島 悠朔、吉留 崇
2. 発表標題 水和自由エネルギーを高速に計算する深層学習モデルの開発
3. 学会等名 第36回分子シミュレーション討論会
4. 発表年 2022年

1. 発表者名 Yusaku Fukushima and Takashi Yoshidome
2. 発表標題 A Deep-Learning Model for the Computation of Grid-Based Solvation Free Energy
3. 学会等名 The 67th Biophysical Society Annual Meeting (国際学会)
4. 発表年 2022年

1. 発表者名 伊藤 祐希、吉留 崇
2. 発表標題 タンパク質-タンパク質界面における水和分布の深層学習を用いた高速予測
3. 学会等名 2022年度 生物物理学会 北海道支部 - 東北支部合同例会
4. 発表年 2022年

1. 発表者名 木村 啓太、吉留 崇
2. 発表標題 水を顕に考慮したタンパク質-リガンド結合様式の高速予測の研究：深層学習によるアプローチ
3. 学会等名 2022年度 生物物理学会 北海道支部 - 東北支部合同例会
4. 発表年 2022年



1. 発表者名 河間 光祐、福島 悠朔、吉留 崇、池口 満徳、大田 雅照
2. 発表標題 深層学習と溶液理論のハイブリッドアプローチによるタンパク質水和分布予測
3. 学会等名 第21回 日本蛋白質科学会年会
4. 発表年 2021年

1. 発表者名 福島 悠朔、河間 光祐、吉留 崇
2. 発表標題 深層学習の視点に基づいたタンパク質水和分布法の高度化
3. 学会等名 14th Mini-Symposium on Liquids
4. 発表年 2021年

1. 発表者名 河間 光祐、福島 悠朔、吉留 崇、池口 満徳、大田 雅照
2. 発表標題 深層学習を活用したタンパク質水和分布計算法の開発
3. 学会等名 14th Mini-Symposium on Liquids
4. 発表年 2021年

1. 発表者名 吉留 崇
2. 発表標題 溶液理論で得たタンパク質水和の包括的解析
3. 学会等名 構造活性フォーラム2021 (招待講演)
4. 発表年 2021年

1. 発表者名 河間 光祐、福島 悠朔、吉留 崇、池口 満徳、大田 雅照
2. 発表標題 深層学習と溶液理論のハイブリッドアプローチによるタンパク質水和分布予測
3. 学会等名 日本物理学会2021年秋季大会
4. 発表年 2021年

1. 発表者名 河間 光祐、福島 悠朔、吉留 崇、池口 満徳、大田 雅照
2. 発表標題 A hybrid approach of deep learning and solvation theory for predicting the hydration structures around proteins
3. 学会等名 第59回日本生物物理学会年会
4. 発表年 2021年

1. 発表者名 福島悠朔、吉留崇
2. 発表標題 A Fast Calculation Method for the Grid Inhomogeneous Solvation Theory via Deep Learning
3. 学会等名 第59回日本生物物理学会年会
4. 発表年 2021年

1. 発表者名 .Kosuke Kawama, Yusaku Fukushima, Takashi Yoshidome, Mitsunori Ikeguchi, and Masateru Ohta
2. 発表標題 Development of a Deep-Learning Model for Predicting the Hydration Structures around Proteins
3. 学会等名 65th Biophysical Society Annual Meeting (国際学会)
4. 発表年 2022年

1. 発表者名 塩野大輝、吉留 崇
2. 発表標題 AlphaFold2で得られたタンパク質立体構造の熱揺らぎ解析
3. 学会等名 2021年度 生物物理学会 北海道支部 - 東北支部合同例会
4. 発表年 2022年

1. 発表者名 大橋 暖、吉留 崇
2. 発表標題 タンパク質水和分布を予測する深層学習モデルの精度向上
3. 学会等名 2021年度 生物物理学会 北海道支部 - 東北支部合同例会
4. 発表年 2022年

1. 発表者名 福島 悠朔、吉留 崇
2. 発表標題 深層学習によるGISTの高速生成法の研究
3. 学会等名 日本物理学会 第77回年次大会
4. 発表年 2022年

1. 発表者名 吉留 崇
2. 発表標題 タンパク質水和の理論研究
3. 学会等名 京都工芸繊維大学 松ヶ崎サイエンスフォーラム
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------