

日本医療研究開発機構 創薬等ライフサイエンス研究支援基盤事業 事後評価報告書



I 基本情報

補助事業課題名：（日本語）創薬等ライフサイエンス研究支援基盤事業
（プログラム名）（英語）Platform Project for Supporting Drug Discovery and Life Science Research

実施期間：平成29年4月1日～令和4年3月31日

補助事業担当者 氏名：（日本語）富井 健太郎
（英語）Kentaro TOMII

補助事業担当者 所属機関・部署・役職：
（日本語）産業技術総合研究所・人工知能研究センター・研究チーム長
（英語）Team Leader・Artificial Intelligence Research Center (AIRC)・National Institute of Advanced Industrial Science and Technology (AIST)

II 補助事業の概要

創薬、医療をはじめライフサイエンス分野の優れた研究成果の実用化へ向け、タンパク質の立体構造及び複合体構造予測、タンパク質と化合物等との相互作用予測等われわれのグループが開発した、あるいは既存のバイオインフォマティクス技術を用いて、外部及び事業内部研究者に対する支援を実施した。期間内に30件以上の支援を実施し、研究支援に関する論文14報が刊行された。事業の一環として、クライオ電子顕微鏡の観測データを利用した全原子モデル構築に関する技術開発と、それを活用した研究支援に注力しており、支援の成果の一つが国際誌に刊行された。さらに、これに関連する追加支援に関する研究成果がまとめられた原稿も国際誌へ投稿されており、クライオ電子顕微鏡による観測データの利用に関する経験、知見と成果を積み重ねつつある。以下に、これまでの主な支援内容とそれらにより得られた成果を記す。

ミトコンドリアではたらくタンパク質の多くはサイトゾルで合成されてから、ミトコンドリアに輸送される。ミトコンドリアの正常な機能発現には、こうしたタンパク質がミトコンドリア内の適切な位置に局在する必要がある。そこで重要な役割を果たすのが、トランスロケーターである。ミトコンドリア外膜に存在するトランスロケーターの一つであるSAM(sorting and assembly machinery)複合体は、ミトコンドリアβ-barrel外膜タンパク質を折りたたんで外膜に挿入するSam50とSam35及びSam37から成るタンパク質複合体である。この詳細なメカニズム解明に向け、遠藤先生(京都産業大学)のグループは、クライオ電子顕微鏡を用いてSAM複合体の立体構造を明らかにすること成功した(東京大学吉川研究室との共同研究)。しかし一部にモデリングを進めるのが困難な低解像度領域が存在していた。そのため、われわれの開発した構造

モデル構築とその検証に関する技術を用いて、クライオ電子顕微鏡の観測データに基づくモデリングの支援を行った。その結果、モデルの健全性及びデータへのあてはまりの両面から妥当と考えられる構造モデルを提供することができた。この結果を含む成果は、国際誌に発表された(Takeda et al., *Nature*, 2021)。また、基質タンパク質を含む SAM 複合体の立体構造解析に関する追加支援を行い、該当する部位の推定が困難であった低解像度領域に対する構造モデルを提供し、その結果の妥当性が検証実験により示されている。こうした結果を含む研究成果についての論文原稿が国際誌へ投稿され、査読中である。また村田先生(生理学研究所)への支援では、同先生のグループにより得られたクライオ電顕マップに基づき、イオン輸送性 V-ATPase 複合体の a-subunit 膜貫通部分をはじめ、モデリングを進めるのが困難な領域が存在していた複数のサブユニットの全原子モデルの構築を行った。

岡澤先生(東京医科歯科大)への支援では、異常リン酸化を受けたタンパク質 Ku70 と DNA の相互作用について、MD シミュレーションを利用した解析の支援を実施し、原子レベルでの相互作用変化に関する情報を得ることができた。こうした成果を含む論文が、国際誌に発表された(Tanaka et al., *Communications Biology*, 2021)。田村先生(山形大学)への支援では、原核生物型 CDP-ジアシルグリセロール合成酵素 Tam41 の系統解析を実施するとともに、AlphaFold の予測結果を利用し、立体構造決定が困難な領域の配置推定を行なった。推定結果は検証実験により確認され、こうした成果を含む論文が発表された(Kimura et al., *Journal of Biochemistry*, 2022)。池中先生(大阪大学)への支援では、 α -シヌクレインモノマータンパク質の構造変化について、同先生のグループによる NMR の測定結果をもとに、より詳細なコンフォメーションを明らかにするためのバイオインフォマティクス技術による支援を実施した。こうした成果を含む論文のプレプリントが刊行された(Aguirre et al., bioRxiv, 2022)。五嶋先生(横浜市立大学)への支援では、対象受容体の立体構造予測と基質との相互作用予測を行った。対象タンパク質に対する基質は一部を除き不明であり、アッセイ系も限られているため、スクリーニング法の開発が必要とされている。このため、対象受容体の立体構造及び基質との結合様式を予測し、その妥当性を確認した。澤崎先生(愛媛大学)への支援では、既知情報を利用した基質-タンパク質の複合体構造モデルのサンプルとその解析を通して、対象タンパク質に対するゲニステインの有力な結合モードを明らかにした。ゲニステイン及び類似化合物であるダイゼインと TAGK2(及び TAGK3)間の結合能の差を明らかにするために、当該化合物の基本骨格構造の一つであるベンゾピロン構造を有する阻害剤の既知結合構造を基に、実験で得られた結合能の差を説明可能な有力な結合モード候補を明らかにした(Nemoto et al., *Nature Communications*, 2017)。また、同先生のグループによる、アブジジン酸(ABA)受容体の新規アゴニストの同定に関する研究では、既知情報を利用した基質-タンパク質の複合体構造モデルの推定とその解析を通して、標的毎の基質結合能に差が生じる機構を明らかにした。対象分子と類似骨格構造を有する ABA 受容体アゴニストの既知結合構造を利用して、基質-タンパク質の複合体構造モデルを構築するとともに、結合の鍵となる残基を同定した(Nemoto et al., *Scientific Reports*, 2018)。野崎先生(東京大学)への支援では、赤痢アメーバ感染後の症状と深い関係があることが示された機能未知遺伝子がコードするタンパク質の立体構造予測や配列情報解析を通して、関連する他の 2 つの遺伝子がコードする同じファミリータンパク質と比較した場合の独自の安定化機構や機能に重要と考えられる部位を明らかにした(Nakada-Tsukui et al., *PLoS Pathogens*, 2018)。また、バイオインフォマティクス技術による赤痢アメーバゲノムの解析により、赤痢アメーバ特異的な新規 β 型膜タンパク質候補二つを同定した。両タンパク質ともに N-末端シグナルペプチド配列と周期的に疎水性残基を含む複数の反復配列を有しており、膜貫通型 β -ストランドペアからなる β -ヘアピンモジュールの存在を示唆するものであった。こうした解析結果に基づき行われた検証実験の結果、 β -シートに富む構造を示すスペクトルが観測された。また各々のタンパク質の小胞体タンパク質との部分的共局在化、及び、オルガネラ膜への組み込みが示唆された(Santos et al., *FEBS Journal*, 2019)。疾病のモデルマウスの病態機序解明を目指した研究では、CRIM1 タンパク質の pCys140Ser ミスセンス変異の及ぼす影響を、バイオインフォマティクス技術の適用により明らかにした。

CRIM1 は 1 回膜貫通型タンパク質であり、細胞外領域に Insulin-like growth factor binding protein (IGFBP)様ドメインと von Willebrand factor C (vWC)ドメインを持つことが知られていた。脊椎動物間で高度に保存されている 140 番目の Cys 残基はこれら両ドメイン間に位置しており、この領域のドメイン構成は不明であった。われわれの解析によってこの領域における新たな複数のドメインの存在を明らかになり、当該変異は、われわれが同定した比較的小さなドメインの内部に位置し、ドメイン内のジスルフィド架橋形成に関わる可能性が示された。モデルマウスの表現型は Crim1 欠損マウスと類似しているが Crim1 欠損マウスより軽度であり、当該変異は CRIM1 の機能を部分的に低下させると考えられており、今回の解析結果と対応するものであった。こうした成果をまとめた論文が発表された(Furuichi et al., *Mammalian Genome*, 2019)。新たな糖尿病早期診断マーカー開発に向けた研究では、核内受容体に対する脂質酸化生成物 (HODEs) の結合予測を行った。HODE には 6 種類の幾何異性体が存在し、各々生体内で異なる活性種によって産生される。そこで、糖尿病治療薬の標的でもある核内受容体 PPAR γ に着目し、HODE 異性体との結合可能性を探ることになった。しかし一部の HODEs を除き PPAR γ に対する結合能は未知であったため、既知結合様式の解析とそれに基づくドッキングシミュレーションを行い、対象の HODE 異性体と PPAR γ との結合を予測した。その後 HODE 異性体の結合は、共同研究者の実験により裏付けられた。こうした結果を含む成果が発表された(Umeno et al., *Bioscience Reports*, 2020)。医薬品化を目指し開発が進められているヒトラクトフェリン(hLF)-Fc 融合タンパク質に関する研究では、複合体構造モデルの構築とその解析を通して、ヒンジ領域欠失 Fc 融合タンパク質に対する免疫細胞活性化低減の機構の一端を明らかにした。本研究では、われわれが構築した複合体構造モデルにより、hLF-CH2-CH3 中の hLF が、ヒンジ領域の欠損のため、hLF-CH2-CH3 と Fc γ RIIIa の相互作用の障害に寄与し得ることが示された(Shiga et al., *Molecular Pharmaceutics*, 2017)。

高度化では、創薬・医療等の分野における課題解決や研究加速に向け、配列情報解析、複合体構造予測、相互作用予測等に関するバイオインフォマティクス手法の開発ならびに改良を実施した。これまでに開発した高感度アミノ酸スコア行列 MIQS や、PSI-BLAST の改良コード PSI-BLASTexB、プロファイル比較に基づく高感度タンパク質類似性検索法 FORTE 等を統合した構造予測パイプラインを構築している。このパイプラインと MD シミュレーションを利用したフィッティングを用いて、クライオ電子顕微鏡の観測データに基づく、主に中～低解像度領域に対する全原子モデルを構築するための技術開発と、それを活用した研究支援を進めており、支援の成果が国際誌に発表される等、クライオ電子顕微鏡の観測データの利用に関する知見と成果を蓄積した。クライオ電子顕微鏡の観測データを利用したわれわれの全原子モデル構築技術に関する PoC を含む成果は、国際誌に発表された(Yamamori et al., *International Journal of Molecular Sciences*, 2022)。またこのパイプラインを用いて、タンパク質立体構造予測研究コミュニティにより開催されている CASP (Critical Assessment of Techniques for Protein Structure Prediction)等に参加し、手法の実証と応用を実施している。これまでに CASP12 のアセンブリ部門で最高位の成績をおさめ(Nakamura et al., *Proteins*, 2018)るとともに、CASP13, CASP14 でも会議での登壇講演の機会を得ており、国際的な評価を獲得している。2020 年には、CASP_Commons ROUND 2 - 2020 (SARS-2-CoV targets)へ参加し、新型コロナウイルスの構造未知の 10 タンパク質について、予測構造モデルを投稿し、それらの評価、解析等を通して、研究コミュニティによる構造情報の提供に寄与した(Kryshtafovych et al., *Proteins*, 2021)。また、タンパク質に関する大規模データのインシリコスクリーニングへの応用に向け、これまでに開発した基質結合(候補)部位の類似性を列挙したデータベース PoSSuM(<https://possum.cbrc.jp/PoSSuM/>)の拡充を進めるとともに、深層学習を利用した新たなタンパク質と化合物の相互作用予測法提案(Tsubaki et al., *Bioinformatics*, 2019)と、研究対象に適した深層学習モデル構築のためのデータセットの抽出に関する研究開発を行った(Ikeda et al., *Frontiers in Molecular Biosciences*, 2021)。

上記に加え、ユニット内のグループに開発技術の提供を行うとともに、AMED-BINDS セミナーや BINDS

ユニット連携講習会等で発表を行う等事業内連携にもつとめた。

To support drug discovery and life science research, we have developed computational methods and databases for protein structure prediction, multimer structure prediction, and interaction prediction for proteins and compounds. We have continued supporting diverse studies conducted in the field of life science research by external and internal researchers using existing and newly developed tools. In addition, we have provided over 30 cases of support, with 14 published papers on research support during that term. Moreover, based on observations from cryo-electron microscopy (cryo-EM) and using our modeling method with profile-profile comparison, we have developed methods and protocols for building an all-atom model. Using those methods, we have demonstrated the possibility of constructing an appropriate structural model for the medium-resolution to low-resolution region.

A representative example of our support using those methods is structural analysis of the mitochondrial sorting and assembly machinery (SAM). The SAM complex, comprising Sam35, Sam37, and Sam50, plays an important role in the localization of β -barrel proteins in the outer membranes of mitochondria. To elucidate a mechanism for the sorting and membrane integration of β -barrel proteins, Prof. Endo's group, collaborating with Prof. Kikkawa's group, used cryo-EM to ascertain the three-dimensional structure of the SAM complex. However, progress with modeling was difficult for some low-resolution regions. Therefore, we supported modeling, based on the density map obtained using cryo-EM, with our methods. We were able to provide appropriate structural models in terms of fitting to the cryo-EM density map and quality as a protein segment, for "missing" regions of Sam37 and Sam50. Additionally, we were able to provide a structural model for an "unknown" region for which it was difficult to infer the relevant site. Our model was validated for this case by experiments conducted by Prof. Endo's group. In this way, we have accumulated experience, knowledge, and results related to the use of cryo-EM data. In addition to this example, our support extends to widely various issues associated with sequence analysis, structure prediction, and interaction prediction supporting life science research.

Regarding "sophistication", we have developed and improved bioinformatics methods to accelerate life science research. Using the protein structure prediction pipeline we have constructed, which integrates various methods developed to date, we can construct all-atomic models of a target protein. Also, based on cryo-EM data, we have developed methods for obtaining appropriate models, as described above. Using this pipeline, we participated in past Critical Assessments of Techniques for Protein Structure Prediction (CASP) and achieved the highest grade in the assembly category of CASP12 (Nakamura et al., *Proteins*, 2018). Moreover, we obtained opportunities to present our approaches at CASP13 and CASP14. In 2020, we participated in CASP_Commons ROUND 2 (SARS-CoV-2 targets) and contributed to structural models for 10 proteins of SARS-CoV-2 having unknown structure. For sequence analysis, we improved and developed methods including the acceleration of MAFFT, complex structure prediction, interaction prediction between proteins and compounds, and protein crystal improvement. Additionally, we tried to expand our database PoSSuM to cover protein-protein interaction and AlphaFold models.