

構造活性相関部会・ニュースレター <1 October, 2022>

SAR News No.43

「目次」

///// Perspective/Retrospective /////			
構造生物学と情報科学をつなぐ Protein Data Bank			
果栖	源嗣	•••	1
///// Cutting Edge /////			
計算化学による抗体の立体構造モデリング			
田黒	大祐	•••	9
CASP におけるモデル構造評価法にみるタンパク質立体構造予測法	の発展		
清田 泰臣、竹田一志鷹 勇	〔由子	•••	24
///// Activities /////			
<報告>			
構造活性フォーラム 2022			
植沢	芳広	•••	29
<会告>			
第 50 回構造活性相関シンポジウム	-t+ () [)		
小田	彰史	•••	31
編集後記			
			33
			-

///// Perspective/Retrospective /////

構造生物学と情報科学をつなぐ Protein Data Bank

大阪大学蛋白質研究所 栗栖源嗣

1. はじめに

蛋白質構造データバンク(Protein Data Bank: PDB)は、1971年にスタートした老舗のデータ ベースである[1]。世界中で唯一の生体高分子の構造データベースであり、毎日世界中で200万 件以上ダウンロードされ、基礎から創薬などの応用まで幅広く活用されている。日本では1979 年から磁気テープ、CD-ROM、インターネットと媒体を変えながら大阪大学蛋白質研究所で継続 的にデータ配布が行われてきた[2]。50年にわたるPDBの歴史のなかで、フォーマット変更や運 営形態の変更、収蔵するコンテンツの拡充や整理など、様々な高度化が実施されている[3-6]。関 連する UniProt やゲノムデータベースとの連携も進み、PDBを利用しやすい環境整備が日米欧で 継続的に行われてきた[7-9]。本稿では、構造生物学のデータベースとして PDB の品質を担保す る取り組みと、機械学習による大規模な構造予測時代への対応、そして生命科学の基幹データベ ースの一翼として、いかに統合的な利用を推進しようとしているかについて、現状と背景、そし て今後の見通しをご紹介したい。

2. 構造生物学からの Data-in と情報科学への Data-out

PDB のコアとなるデータは構造生物学者が実験的に構造決定した原子座標と実験情報、それ らに紐づくメタデータである。我々PDB 関係者は、国際的に決めた統一基準でデータを編集・ 登録・検証する作業を行なっており、この活動を Data-in と呼んでいる。Data-in の活動では、原 子座標とメタデータは全て PDB に収録され、結晶構造解析の場合は実験情報(結晶学的構造因 子)も座標と一緒に PDB に登録される。これに対して、NMR 実験で構造が決定された場合は、 実験情報は BMRB (Biological Magnetic Resonance Data Bank) へ登録され、クライオ電子顕微鏡 構造解析の場合には、密度マップが EMDB (Electron Microscopy Data Bank) に登録される(図 1)。



図 1. 構造生物学のコアアーカイブへのデータ割り振り。PDB, BMRB, EMDB の 3 つのア ーカイブにデータが蓄積される。^{#1}溶液構造決定に用いられる NOE などの制限データの一 部は PDB に蓄積されているものもあるが、今後 BMRB に統一される予定である。

アジア地区では大阪大学蛋白質研究所が設置した Protein Data Bank Japan (PDBj)が PDB, BMRB, EMDB の 3 つのアーカイブを維持・管理している。Data-in の活動の他、データベースを構築し て検索機能を付与したり、派生する二次データベースを構築したりするデータ公開に関する活動 も実施している。これを Data-out と呼び、Data-out に関しては日米欧の各地域 PDB 拠点で独自 に活動する裁量が認められている。すなわち、アーカイブの構築は国際連携で統一的に実施する ものの、オリジナルのWebサービスや二次データベースの提供には競争原理が導入されている のである。データ利用に関するサービスは大きく2つに分けて考えられる。一つ一つのエントリ ーについて深掘りする実験研究者によるWebベースの利用と、機械的にアーカイブ全体へアク セスしAIなどを活用して網羅的に分析・解析する情報科学的な利用である。これらの背景と現 状認識の下で、構造生物学からのData-inと情報科学へのData-outの取り組みがどのように実施 されているのか順に見ていくことにする。

2.1 世界でひとつの PDB を構築する Data-in の仕組み

世界中の研究者が構造解析した蛋白質などの構造情報を集めた PDB は、1971 年に米国でスタ ートした[1]。1998 年までは米国 Brookhaven 国立研究所が世界中のデータを収集し、米国の予算 で登録・編集が行なわれていた[2]。90 年代後半になると、放射光施設を利用した構造ゲノムプ ロジェクトが日米欧で立ち上がり、PDB への登録件数が飛躍的に増加した。そこで日本では大 阪大学に日本蛋白質構造データバンク(PDBj:代表、中村春木教授)が、欧州では European Bioinformatics Institute (EBI)に Protein Data Bank in Europe (PDBe:代表、Kim Henrick 教授)が設 置され、米国だけでなく日本や欧州でデータ登録と編集が開始された[6]。2002 年スイス・ジュ ネーブで開催された国際結晶学連合会議(IUCr)の際、1999 年に Brookhaven 国立研究所から PDB を引き継いだ米国 Research Collaboratory for Protein Informatics (RCSB) PDB の Helen Berman 教授 (Rutgers 大学)と中村教授、Henrick 教授とが集まって、RCSB PDB, PDBj の 3 拠点で構 成される国際蛋白質構造データバンク(worldwide Protein Data Bank: wwPDB)を設立し、国際連 携で PDB を運営することが合意された。2003 年に wwPDB の正式な協定書が締結され[10]、国 際的統一基準で PDB を維持・管理し、無料でデータを公開する枠組みが確立した。

しかし歴史的経緯により各拠点でバラバラの登録・編集システムを採用していたため、2003 年当時「国際的統一基準」を謳うことは"言うは易し行うは難し"の状況であった。そこで wwPDB の各メンバーが協力し、日米欧のどこでも使えて、結晶、NMR、電子顕微鏡のどの解析手法に 対しても共通で使える登録システムの開発が開始された。常時10人を超えるスタッフで数年に 及ぶ共同開発を行い OneDep という共通のデータ登録・編集システムを完成させ、2014 年に X 線結晶構造解析のエントリーに対して先行導入した[11]。現在は NMR、クライオ電子顕微鏡に も利用を拡張し、各拠点で同じシステムによる編集が可能となっている。OneDep の導入によっ て、登録する PI が所属する国・地域を基に登録セッションをリダイレクトする自動的な地域分 担割り当てが可能となった。現在 PDBj は地域のデータ登録数を基にアジア・中東地区を担当し ている(図 2)。



図 2. Data-in の地域分担と構築するアーカイブの種類を示した地図。 PDBj は 3 つのアーカイブ全て担当している。

2.2 データベースの利用許諾条件の成文化

1971年に7件のエントリーで開始したPDBエントリーは、利用許諾の考え方が浸透する以前 から存在していたため、これまでエントリーコンテンツの利用許諾について明確な規定が存在し ていなかった。そこで、2021年のwwPDB定款(Charter)更新の際に、データアクセスライセ ンスをクリエイティブ・コモンズの標準的なオープンソースライセンスである CC0 1.0全世界 (CC0 1.0)に設定し(<u>https://www.wwpdb.org/about/agreement</u>)、PDBエントリーのコンテンツは引 用元を記述する必要なく、営利・非営利を問わず自由に利用することが可能であると明文化され た。PDBエントリーのコンテンツの他に、PDBjホームページ中の2次データベースの情報、文 章、構造の画像なども、アカデミア、企業、個人を問わず、無料で自由に利用することが可能で ある。転載・引用の際には、引用元として「PDBj」を明示していただくことをお願いしている。 但し、Molecule of the Month(今月の分子)の図の著作権は米国の RCSB PDB に帰属している。 利用許諾条件の詳細は、PDBjのホームページに情報を掲載しているのでご参照いただきたい (<u>https://pdbj.org/info/terms-conditions</u>)。

2.3 PDB に登録・編集されるデータの種類と実験データに対する検証

今では考えられないことだが 90 年代前半までは、構造解析の論文が発表された後も PDB にデ ータが公開されない (登録されない) 例が多数あり、大きな問題となっていた[12]。IUCr の生体 高分子委員会から「生体高分子の構造を論文発表する際に PDB への座標登録を必須とする」こ とが勧告され[13]、構造生物学コミュニティの努力もあって、その後徐々に「PDB に登録しなけ れば論文投稿できない」というコンセンサスが浸透していった。PDB は現在、結晶構造解析 (X 線、中性子、電子線)、NMR 構造解析、クライオ電子顕微鏡構造解析 (螺旋平均化、単粒子平 均化)の何れかの手法による構造を収録していることが広く認知されている。しかし、2002 年 7 月 1 日までは、ホモロジーモデルや第一原理計算など純粋に計算のみで予測された構造も PDB アーカイブに含まれていた (正確には、今も PDB アーカイブから分離された状態で公開されて いる[/pub/pdb/data/structures/models/])。理論モデルを公開するデータベースが存在しない頃の名 残であるが、そもそも"理論モデル (Theoretical Model)"と実験で決定した構造の定義も曖昧で あった。そこで 2006 年 10 月 15 日以降は、上記 3 実験で構造決定された"検証可能な"エント リーのみを PDB で受け付けている[14]。「検証可能」とは、実験データを用いて第三者が構造エ ントリーの品質を評価できる検証レポート (Validation Report)を作成できるという意味で用い ている (図 3)。



図 3. 検証レポートの一例。PDB ID 7CKA の検証レポートの一部を示す。(A) 同レベルの 解像度の解析例に対する相対評価。(B) アミノ酸毎の実験データとの一致度。(C) リガン ドの座標と実験データ(この場合は電子密度マップ)の一致度を wwPDB の基準で計算し たもの。NMR や電子顕微鏡による構造についても、検証レポートが用意されている。

2008年以降、結晶構造解析の場合には結晶構造因子(Structure Factors)、NMRの場合にはNOE などの制限情報ファイル、そして電子顕微鏡構造解析の場合にはクーロンポテンシャル密度マッ プを座標データと一緒に登録することが必須となった。これら実験データを基に、実験データと の一致度や構造の品質を評価する検証レポートを全てのエントリーに対して計算し、wwPDBと してエントリーごとに提供している。現在では、主だった Journal において論文投稿の際、PDB ID の他に wwPDB が提供する検証レポートを一緒に提出することが求められている。これにより、 査読者が構造データを客観的に評価することが可能となっている。

2.4 フォーマット変更など PDB の方針決定の仕組み

検証レポートの全エントリーへの導入のような大きな変更は、wwPDBの運営諮問委員会 (wwPDB Advisory Committee)の下に設置するワーキンググループを通じて、コミュニティの声 を聞きながら実施される。例えば、上述の理論モデルを PDB アーカイブから切り離し、3 つの 検証可能な実験手法で決定された構造のみに限定するという基本方針の変更は、2005 年 11 月に 米国 Rutgers 大学で開催されたワーキンググループでの議論を経て、wwPDB ACの承認の下で決 定された[14]。その他にも、古い PDB フォーマットに代わる新フォーマットを何にするかとい う課題についても、2011 年に wwPDB AC の下にワーキンググループを設置し、構造解析ソフト や分子ビューア、解析ツールなど主だった構造生物学のソフトウェア開発者を招待して定期的に 会議を重ねた。その結果、STAR ベースの PDBx/mmCIF をマスターフォーマットとすることを 決定し、2013 年にアナウンスされている[15]。この紙面を借りてアナウンスさせていただくが、 古い PDB フォーマットは既に限界にきており、2023 年度中にも予想される化合物 ID (3 桁)の 枯渇によって、新しい PDB エントリーは全て PDBx/mmCIF で記述する必要が生じる(図 4)。 PDB フォーマットのみに対応したソフトをお使いの場合は、早急に PDBx/mmCIF 対応の新ソフ トへの移行をご検討いただきたい。構造解析系の Phenix, CCP4, CCPEM, Coot, 分子ビューア の PyMol や Chimera など主だったソフトウェアは既に対応済みである[16]。

# 100	~								
# 100j									
	_site.		wla a l						
_atom	_site.	type_sy							
_atom	_site.	label_a	tom_id						
_atom	_site.	label_a	lt_id						
_atom	_site.	label_o	omp_id						
_atom	_site.	label_a	sym_id						
_atom	_site.	label_e	ntity_id						
_atom	_site.	label_s	eq_id						
_atom	_site.	pdbx_PD	B_ins_code						
_atom	_site.	Cartn_x							
_atom	site.	Cartn_y							
atom	site.	Cartn z							
atom	site.	occupan	су						
atom	site.	B iso o	r equiv						
atom	site.	pdbx fo	rmal charge						
atom	site.	auth se	q id						
atom	_ site.	auth co	mo id						
atom	site.	auth as	vm id						
atom	site.	auth at	omid						
atom	site.	pdbx PD	B model num						
	site	calc fl	- · · · - ·						
_atom	 site	ndby t1	s aroun id						
	1	N N	GLY A 1 1	2 -24 035 68 606	85 927	1 000 70 366	0 -1	CLY AAA N	1 2 1
ATOM	2	C CA	CIVA11	2 -22 673 68 622	86 530	1 000 68 902	0 -1	CLY ANA CA	1 2 1
ATOM	3	CC	CIVA11	2 -21 603 68 394	85 484	1 000 67 058	0 -1	CLY AND C	1 2 1
ATOM	1	0.0	. ODIAII	2 _21 412 67 227	05.101	1 000 65 622	0 _1	CIVANO	1 0 1
ATOM	4		OPD 3 1 0	2 20 044 60 461	05.070	1 000 67 455	0 -1	CED ANA N	1 0 1
ATOM	5		. SERAIZ	: -20.944 09.461	00.034	1,000 66 560		SER AAA N	1 2 1
ATOM	b	CCA	. SERAIZ	: -19.898 69.39/	03.9/8	1.000 65.568	0 0	SER AAA CA	1 : 1
MO'L'A	1	CC	. SER A 1 2	? -18.676 68.618	84.472	1.000 65.234	υΟ	SER AAA C	1 ? 1

図 4. PDB フォーマットに変換できない PDBx/mmCIF フォーマットの例。管理者側で定義 する厳密な data には.*label* という item が付され、登録者が定義する曖昧な data には.*auth* という item が割り振られている。この例では、chain ID が3桁となっており PDB フォー マットには変換できない。PDBx/mmCIF フォーマットではコピー&ペーストはできず、専 用ソフトによる編集が必須である。

2.5 PDB の品質を向上させるバージョニング

wwPDB が進めてきた活動のうち新たな局面を迎えているとして期待されるのが、登録済みデ ータに遡って構造データの品質を向上させる活動が挙げられる。コロナ禍に直面した世界中の構 造生物学者は、創薬研究を加速することを目的に多くの新型コロナウイルスの蛋白質構造を報告 した。しかし、あまりにも急いで構造解析をおこなったため、中には不十分な精密化のものも多 く含まれていた。こうしたコロナ禍に起こった PDB に関連した大きな動きは、ボランティアベ ースで PDB の構造を再精密化し、データベースの品質を向上させようという試みを生んだ。実 験データを同時に寄託する wwPDB のデータ登録形式だからこそ可能な仕組みである。構造解析 の専門家が最良の構造に再精密化するという取り組みは、wwPDB で versioning と呼ぶデータ差 し替え機能を使えば、PDB ID を変更することなく、より高品質の構造データにエントリーを差 し替えることが可能である(古いデータも version を付して残る)[17]。

特に新型コロナウイルス蛋白質に特化した取り組みは「The Coronavirus Structural Task Force」 と呼ばれ、データ駆動型の新しい研究スタイルとして注目されている[18]。PDBj では wwPDB 必須の実験データである Data Reduction された結晶構造因子や EMDB に収録される処理済み密 度マップだけでなく、エントリーに紐づいた X 線の回折イメージや電子顕微鏡の生画像もデー タ寄託できるよう、X 線回折イメージのアーカイブ XRDa (<u>https://xrda.pdbj.org</u>) と電子顕微鏡イ メージのアーカイブ EMPIAR-PDBj (<u>https://empiar.pdbj.org</u>) を開設した。今後、PDB データベー スの品質向上も大きなインパクトを与えていくであろう。

2.6 PDB の統合的利用を加速すると期待される Data-out の活動

PDB では PDBx/mmCIF のフォーマット[15]、BMRB では NMR-STAR フォーマットを標準とし て採用している[19]。どちらも PDBj と PDBj の NMR 部門である BMRBj のイニシアチブによっ て XML と RDF 版が世界標準として用意され、週次更新に併せて公開される仕組みができてい る。特に RDF については DBCLS RDF 化ガイドラインに沿っており、既に JST NBDC の RDF portal から公開されている。収録されているメタデータは、実験の対象となった分子の由来する生物 種・器官・合成法・精製法・構造解析実験手法・解析手法・生物学的機能・複合体情報、他の関 連するデータベース(配列・taxonomy その他)など、多岐にわたっており、辞書(PDBx/mmCIF Dictionary Resources, http://mmcif.wwpdb.org/)にまとめられてオントロジーが整備されている(表 1)。

フォーマット	URL	オントロジーの説明
名称		
PDBx/mmCIF	https://mmcif.wwpdb.org	wwPDB 共通の定義辞書
NMR-STAR	https://bmrb.io	NMR データのリマスターフ
		オーマット
CCD	https://wwpdb.org/data/ccd	化合物辞書
BIRD	https://wwpdb.org/data/bird	生物学的に重要な分子に関
		する情報を集めた参照辞書
VRPT	https://wwpdb.org/validation/validation-reports	wwPDB の検証レポート
SIFTS	https://www.ebi.ac.uk/pdbe/docs/sifts	PDB に登録されたエンティ
		ティーの外部データベース
		の参照情報

表 1. PDB に収録されているメタデータのオントロジー一覧

2.7 AI などを用いた機械的・網羅的利用を想定した Data-out の展開

PDB には蛋白質ポリペプチド鎖のみが含まれている訳ではない。PDB データの実に 75%近く のエントリーが非蛋白質分子、すなわち「リガンド化合物」を含んでいる。創薬ターゲットの場 合には、このリガンド化合物の構造情報が極めて重要である。創薬候補化合物のターゲット蛋白

質へのドッキング予測の精度を高めたり、補欠分子族や基質分子の結合様式を考察する化学的知 見を高めたりする上で、利用者が必要とする構造精度は目的によって異なってくる。例えば、量 子化学計算を伴う QM/MM 法などでは、結合長や結合角まで構造解析の精度が大きく計算結果 に影響するであろうし、MD をベースとした結合シミュレーションであれば実験的な根拠の低い ノイズのような実験根拠を基に、いくら結合をシミュレーションしたとしても、結合途中の様子 を正確にトレースすることは困難であろう。そこで予測精度の向上に資する機械学習(AI)向 きの高精度構造データの提供が重要である。しかし、実際に PDB に含まれるリガンド化合物の 原子座標を実験的に得られた密度マップとの整合性で評価するのは、高度な専門性を必要とした。 具体的には結晶学や NMR 分光学、電子顕微鏡学を専門としない一般の生命科学研究者がデータ を利用する際に、「とりあえず2Åよりも高分解能」といった密度マップとの整合性を用いず、 分解能のみを用いた単一的な尺度で構造データをフィルタリングし、機械学習やデータ分析に用 いられている例が多い。しかし時には、2Åを超える高分解能でも、リガンド化合物と蛋白質と の相互作用が弱くリガンド化合物の占有率が低い場合などは、実験データと原子座標との整合性 は著しく低い(図 4)。



図5.2 Åを超える分解能でのリガンド化合物と実験で得られた密度マップデータの一致例。 (A) 1.8 Å 分解能の実験データとの一致度が良い例と、(B) 1.9 Å 分解能で一致度の悪い例。 結晶学では実験で得られた構造因子を F-observed: |Fo|,原子座標から計算される予測構造 因子を F-calculated: |Fc|と表記する。灰色のマップは、測定データ(Fo)に測定と計算の 2 つの構造因子の差([Fo|-|Fc])を上乗せし、測定誤差で重み付けしたマップ(2m|Fo|-D|Fc|map) を表す。緑色と紫色のマップは差マップ(m|Fo|-D|Fc|map)と呼ばれ、|Fo|と|Fc|の差分をプ ラスとマイナスで表している。マップが、ほぼ灰色で表記されていれば実験と構造モデル の一致度が高いことを示し、紫色のマイナスの差マップが表示されていれば、測定データ |Fo|の方に該当する情報が少なく、|Fc|の寄与でのみマップが計算されていることを示唆し、 実験とモデルの一致度が低いと考えられる。

現在では、2.3 で紹介した通りリガンド化合物の実験データとの一致度を残基単位で評価でき る指標として「実空間信頼度因子(Real Space Reliability Factor: RSR)」や「マップファイルとの 相関係数(CC)」、化学結合距離や結合角度の標準的な値からのずれを示す統計値などが検証レ ポートとして整備されている[20]。2022年4月にスタートしたJST-NBDCの統合化プロジェクト において、PDBに実装済みのRDF化した検証レポートを最大限に活用し、目的ごとに化合物に 特化した機械学習用データセット選抜基準を作成する。そして機械学習用データセットを予め選 抜して、ユーザーがダウンロードできるサービスを開始する。データ公開は、毎週のPDBのデ ータ更新に合わせて行い、事前にデータセットも自動更新して配布できるように準備中である。 需要が高いと思われるデータセット(例えば、「分解能2Å以上で、活性データが計測されてい る医薬品様化合物が蛋白質の単一の鎖に結合している全てのエントリー」や、「相互作用してい るドメインを配列類似度でクラスタリングしたときの代表構造の複合体」)などを準備して、標 準データセットとして配布する計画である。

3. AlphaFold2 のインパクトと AlphaFoldDB と PDB の関係

構造予測の分野では、DeepMind 社が開発した AlphaFold2 と呼ばれる機械学習システムによる 構造予測技術が注目を集めている。ゲノム情報とともに PDB の 19 万件をこえる膨大な数の構造 データから品質の高いエントリーを学習データとして取り込み、アミノ酸配列から機械学習によ り未知構造を高精度で予測することが可能となってきている[21]。予測精度を高めるための開発 競争は既にスタートしており、ワシントン大学の Baker 教授らによる RoseTTAFold にも注目が 集まっている [22]。予測構造の精度を高めるため、より高精度で広範な実験構造を蓄積してい くことも求められているが、PDB 事業者として構造予測システムを提供する複数の研究グルー プとどう連携していくのかも重要な課題である。既に wwPDB のパートナーである PDBe(PDB in Europe) が DeepMind 社との共同研究として、UniProt の配列から Google 社のクラウド計算機で 網羅的に予測した 3 次元構造を AlphaFoldDB として一般に公開し[23]、今現在も対象とする予測 構造の数は増え続けている。その際、wwPDB で標準フォーマットとして開発した PDBx/mmCIF 形式に準拠した ModelCIF が用いられ、PDBe が蓄積してきた蛋白質構造データベース構築技術 が援用されたのは、20 年以上に渡る国際協力で培ってきた国際的なノウハウが生かされたとも 言えるであろう。

AlphaFold2 による予測構造が特に注目をあびているが、対抗するワシントン大学の Baker 教授 らによる RosseTTAFold のエントリーは、RCSB PDB の Web サイトから AlphaFold2 のエントリ ーと並行して公開されている。今現在は、構造を予測している各当事者が自身の基準で信頼性を 評価しているが、より客観的かつ公平な基準で構造予測の精度を評価することが重要となるであ ろう。PDB の運用で培った構造検証のノウハウを活かし、日米欧の 3 拠点共同で客観的な予測 構造の検証基準をつくることを検討している。急速に重要性の高まった構造予測の分野でも、 PDB データベースの運用で培ったノウハウの価値がますます増しているといえるであろう。

謝辞

PDBj の活動は、国立研究開発法人・科学技術振興機構(JST) バイオサイエンスデータベー ス事業部(NBDC:グラント番号 JPMJND2205)と国立研究開発法人・日本医療研究開発機構 (AMED) 創薬等先端技術支援基盤プラットフォーム(BINDS:グラント番号 22ama121001) およ び、文部科学省から共同利用・共同研究拠点として大阪大学蛋白質研究所に措置された経費によ って支援されています。また、日頃一緒に活動している PDBj の現在のメンバーと、RCSB PDB, PDBe, BMRB, そして EMDB のメンバーの皆さんに感謝したい。最後に、wwPDB のメンバー として活動する PDBj の枠組みは、大阪大学名誉教授の中村春木先生により確立されました。先 生の永年にわたるご貢献に改めて感謝の意を表したいと思います。

参考文献

- [1] Protein Data Bank. Crystallography:Protein Data Bank. Nature New Biol., 233, 223 (1971). doi:10.1038/newbio233223b0
- [2] Sussman, J.L., Lin, D., Jiang, J., et al. The Protein Data Bank at Brookhaven, Int. Tables Cryst., F, 649-656 (2006). doi:10.1107/97809553602060000718
- [3] Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., et al. The Protein Data Bank: A Computer-based Archival File for Macromolecular Structures, J. Mol. Biol., 112, 535-542 (1977). doi: 10.1016/s0022-2836(77)80200-3
- [4] Berman, H.M., Westbrook, J., Feng, Z., et al. The Protein Data Bank. *Nucleic Acids Res.*, 28, 235-242 (2000). doi:<u>10.1093/nar/28.1.235</u>
- [5] wwPDB consortium. Protein Data Bank: the single global archive for macromolecular structure data, *Nucleic Acids Res.*, 47, D520-D528 (2019). doi:<u>10.1093/nar/gky949</u>
- [6] Velankar, S., Burley, S.K., Kurisu, G., et al. The Protein Data Bank Archive. *Methods Mol. Biol.*, 2305, 3-21 (2021). doi:10.1007/978-1-0716-1406-8_1
- [7] Burley, S.K., Berman, H.M., Bhikadiya, C. et al., RCSB protein data bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.*, 47, D464-D474 (2019). doi:10.1093/nar/gky1004

- [8] Armstrong, D.R., Berrisford, J.M., Conroy, M.J., et al., PDBe: improved findabiligy of macromolecular structure data in the PDB. *Nucleic Acids Res.*, 48, D335-D343 (2020). doi:10.1093/nar/gkz990
- [9] Bekker, G.J., Yokochi, M., Suzuki, H., et al., Protein Data Bank Japan: Celebrating our 20th anniversary during a global pandemic as the Asian hub of three-dimensional macromolecular structural data. *Protein Sci.*, **31**, 173-186 (2022). doi: <u>10.1002/pro.4211</u>
- [10] Berman, H., Henrick, K., Nakamura, H. Announcing the worldwide protein data bank. *Nature Struct. Biol.*, **10**, 980 (2003). doi: <u>10.1038/nsb1203-980</u>
- [11] Young, J.Y., Westbrook, J.D., Feng, Z., et al., OneDep: Unified wwPDB System for Deposition, Biocuration, and Validation of Macromolecular Structures in the PDB Archive. *Structure*, 25, 536-545 (2017). doi:10.1016/j.str.2017.01.004
- [12] Barinaga, M. The Missing Crystallography Data. Science, 245, 4923 (1989). doi:<u>10.1126/science.2781276</u>
- [13] Commission on Biological Macromolecules. Policy on Publication and the Deposition of Data from Crystallographic Studies of Biological Macromolecules., Act Cryst. A45, 658 (1989). doi:10.1107/S0108767389007695
- [14] Berman, H.M., Burley, S.K., Chiu, W. et al., Outcome of a workshop on archiving structural models of biological macromolecules. *Structure*, 14, 1211-1217 (2006). doi: <u>10.1016/j.str.2006.06.005</u>
- [15] https://www.wwpdb.org/news/news?year=2013#5764490799cccf749a90cdc5
- [16] Adams, P.D., Afonine, P.V., Baskaran, K., et al., Announcing mandatory submission of PDBx/mmCIF format files for crystallographic depositions to the Protein Data Bank (PDB). Acta Cryst. D75, 451-454 (2019). doi: 10.1107/S2059798319004522
- [17] <u>http://www.wwpdb.org/news/news?year=2017#5979ef4161fd3d50915a4b00</u>
- [18] Croll, T., Diederichs, K., Fischer, F., et al., Making the invisible enemy visible. *Nature Struct. Mol. Biol.*, 28, 404-408 (2021). doi: <u>10.1038/s41594-021-00593-7</u>
- [19] Ulrich, E.L., Baskaran, K., Dashti, H., et al., NMR-STAR: comprehensive ontology for representing, archiving and exchanging data from nuclear magnetic resonance spectroscopic experiments. J. *Biomol. NMR.*, **73**, 5-9 (2019). doi: 10.1007/s10858-018-0220-3
- [20] Read, R.J., Adams, P.D., Arendall III, W.B., et al., A New Generation of Crystallographic Validation Tools for the Protein Data Bank. *Structure*, **19**, 1395-1412 (2011). doi: <u>10.1016/j.str.2011.08.006</u>
- [21] Jumper, J., Evans, R., Pritzel, A., et al., Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583-589 (2021). doi: <u>10.1038/s41586-021-03819-2</u>
- [22] Baek, M., DiMaio, F., Anishchenko, I., et al., Accurate prediction of protein structures and interactions using a 3-track network. *Science*, **373**, 871-876 (2021). doi: <u>10.1126/science.abj8754/</u>
- [23] Varadi, M., Anyango, S., Deshpande, M., et al. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, 50, D439-D444 (2022). doi: 10.1093/nar/gkab1061

///// Cutting Edge /////

計算化学による抗体の立体構造モデリング

国立感染症研究所 黒田 大祐

1. はじめに

医薬品にはさまざまなモダリティが存在する。その中でも「抗体」は、標的抗原に対する高い 結合親和性と特異性から、副作用の少ない分子標的薬として、創薬研究が進められている。昨今 問題となっている、コロナウイルスに対しても、抗体医薬品の使用が既に始まっている[1]。

それでは、抗体創薬を進める上で、抗体の「立体構造」は重要だろうか?図1に、抗体創薬に おける計算化学の課題をまとめた。立体構造を起点とすることで、計算化学を用いた抗体の物性 や機能の予測が可能となり、より高品質な抗体医薬の開発につながる[2-4]。ワクチン設計の際 も、抗体の構造情報は有用である。抗体-抗原複合体の立体構造を起点として、ワクチン設計の際 デザインしようという試みが盛んに行われている[5,6]。ワクチンの目的は、標的抗原に対する B 細胞の免疫応答、つまり、抗体産生を効率的に誘導することにある。そこで、例えば、抗体-抗 原複合体の立体構造から、抗原上の抗体が認識する領域(エピトープと呼ばれる)のみを切り出 し、それを安定化するようにデザインすることで、より効率的なワクチンを開発できる可能性が ある。こうした過程では、デザイン対象となる分子の立体構造に基づく計算化学が欠かせない技 術となっている。

本稿では、計算化学を用いた最近の抗体研究を概観する。近年話題の AlphaFold2 と抗体研究 との関わりについても述べる。



図1. 抗体創薬における計算化学の課題。

2. 抗体の分子構造と機能

2.1 抗体の立体構造

抗体は重(Heavy; H)鎖と軽(Light; L)鎖から構成され、教科書ではY字型をした分子として描か れている。そうした抗体全長の分子構造もX線結晶構造解析により既に明らかにされている[7]。 図 2A に示すように、抗体は、H 鎖とL 鎖が二量体を形成したものが、さらに二量体を形成し、 全体としてY字型になっている。Y字の先端部分である可変領域(Variable domain)で抗原を認識 している。つまり、1つの IgG 抗体に対して、2 つの抗原認識部位が存在する。こうした特徴を 活かし、特異性の異なる 2 つの L/H 鎖ペアを、計算化学の技術で二量体化することで、Y 字の 先端でそれぞれ異なる抗原を認識できる、二重特性を持つ抗体の分子設計も行われている[8]。 また、抗体は複数のドメイン構造がヒンジで繋がれた構造をしており、当然ドメイン間の相対配 置は動きやすくなっている。こうしたドメイン間の相対配置の変化も、抗原認識に重要な役割を 果たしている。抗体は様々な分子フォーマットが提案されているが、医薬品として使われている 抗体は、図 2A のような IgG 型が多い。IgG 型抗体の分子量は 10 万を超える。こうした巨大分 子のダイナミクスを評価するには、計算化学の立場からは、全原子分子動力学(Molecular Dynamics; MD)計算を用いることができる。しかし、その計算コストを考えると、容易な仕事で はない。代替手法として、粗視化シミュレーション[9]や幾何学的シミュレーションなども使わ れている[10]。

一方で、IgG 型の抗体をドメイン単位で切り出しても、抗原への結合能を維持していることが 知られている。タンパク質立体構造データバンク(PDB)に登録されている抗体の多くは、可変領 域と定常領域がヒンジで繋がれた Fab 型である。これらのドメイン間の角度は Elbow angle と呼 ばれる。同一抗体でも、登録されている PDB ファイルによって、Elbow angle が異なることもあ る。また、抗体の L 鎖は Kappa 鎖と Lambda 鎖の 2 つのサブタイプが存在するが、こうしたサブ タイプの違いによって、Elbow angle が異なるという報告もある[11]。

近年、アルパカやラマ由来の抗体が注目を集めている。これらの動物は、L鎖H鎖を持つIgG型の抗体も産生するが、ヒトとは異なり、H鎖のみでY字を構成する抗体も産生できる。こうしたH鎖のみから構成される抗体の可変領域部分を切り出した V_HH (Variable domain of Heavy chain of Heavy chain antibody)抗体は、1つのドメインで抗原認識能を発揮し、分子量も小さく、溶媒中で安定に存在できる。そのため、センサーなどのバイオ素子としての応用が期待されている。

コンピュータを用いて抗体を研究する上で、しばしば問題となるのが、そのアミノ酸配列のナ ンバリングである[12]。後述するように、抗体間で、抗原に応じて、抗原認識部位のアミノ酸配 列の長さが異なる(図 2B)。この違いは、抗体が由来する gernline 遺伝子や、親和性成熟時の塩 基配列の挿入/欠失に由来する。こうした抗体間での部分配列長の違いに対応するために、挿入/ 欠失部位を数字+アルファベット(100A,100B など)を用いて表し、抗体間で保存されている部 位を数字のみで表すということが古くから行われてきた。このような挿入/欠失部位を同定する には、配列情報だけよりも、立体構造を考慮した方が精度は高くなる。そのため、主に挿入/欠 失の取り扱い方の違いから、複数の抗体ナンバリング方法が提案されている。



図 2. 抗体の立体構造と抗原認識部位の多様性。(A) 抗体の立体構造。緑:L 鎖、青:H
 鎖、黄: CDR-L1/L2/L3、シアン: CDR-H1/H2、マゼンタ: CDR-H3、黒: 糖鎖。
 (B) ヒト抗体における相補性決定領域(CDR)の長さの分布。

2.2 抗体の抗原認識部位の多様性と画一性

可変領域に位置する抗原認識部位は相補性決定領域(Complementarity determining regions; CDRs)と呼ばれ、6つのループ構造(L1-L3, H1-H3)から構成される(図 2A)。CDR は歴史的には 抗体の全体配列をアライメントした時に「多様性のある領域」として定義されている[13]。その 後、立体構造が明らかになるにつれ、構造情報に基づく CDR の定義づけが行われている[14]。

無数に存在する抗原を考えると、CDR の配列や構造もそれに合わせて多様でなければいけな いように思える。しかしながら、抗体の立体構造を観察すると、中央に位置する CDR-H3 以外の 5 つの CDR は、比較的その配列や立体構造が保存されている。こうした保存された構造は「カ ノニカル構造」と呼ばれている[15]。カノニカル構造を維持するために必要なアミノ酸も既に明 らかにされており、鋳型に基づくアミノ酸配列からの抗体モデリングに活用されてきた。また、 医薬品としての抗原性の問題を回避するために、しばしば抗体の「ヒト化」を行うが、その際に、 目的の機能を持つ CDR と同一のカノニカル構造を持つヒト抗体のフレームワーク領域を選ぶと いうことが行われてきた。図 2B に Chothia らの定義に基づく 6 つの CDR の長さの分布を示し た。H3 以外は、ある限られた長さの CDR しか存在しないことが分かる。

CDR-H3 は抗体 H 鎖をコードしている VDJ 遺伝子の接合部位から成る。そのため、その配列 と立体構造は非常に多様である(図 2A)。CDR-H3 には、いわゆるカノニカル構造は定義されて いない。しかし、その根本の部分構造は、Kink 型と Extend 型に分類でき、アミノ酸配列から高 精度な予測が可能となっている[16,17]。また、近年、アルパカ由来の単ドメイン抗体でも、その CDR-H3 の立体構造は、ある限られた構造クラスに分類できることも報告されている[18]。

一般的に、程度の差はあれど、タンパク質は分子認識に伴い構造変化を起こす。それでは、抗体の立体構造は、抗原認識に応じて、その立体構造を変化させるのだろうか?こうした問いに答えるために、抗原結合状態および抗体単独で結晶構造が取得されている49個の抗体の立体構造を比較した報告がある[19]。この研究では CDR の構造変化は比較的小さく、最も構造変化が起きていた CDR-H3 でも、Cα原子の RMSD が平均1.3Å程度であった。一方で、抗原認識部位から遠くに位置するH鎖定常領域のCH1ループに関して大きな構造変化(平均 Cα-RMSD=1.8Å)が観察されており、定常領域が可変領域に影響を与えるという、抗体の抗原認識におけるアロステリック制御の可能性を指摘している。図3に、著者が独自に収集した PDB 中の93 個の抗体の立体構造に基づき、抗原の有無での CDR の構造変化の度合いを主鎖の RMSD として定量化した結果を示す。L 鎖の CDR では、90%以上が、CDR-H1 と H2 でも 80%以上が 1Å以内の構造の差異に収まっていることが分かる。CDR-H3 では、データセット中の 60%の抗体で構造変化が 1Å以下となっている。しかし、このデータはあくまで、抗原結合状態と非結合状態の「結晶構造」間での比較であり、そもそも柔軟な CDR 構造を持つ抗体は結晶化が困難である可能性もあるため、解釈に注意が必要である。



図 3. 抗原認識に伴う CDR(L1-L3, H1-H3)の構造変化。フレームワーク領域を重ね合わせ た時の各 CDR の構造のずれを CDR 主鎖の RMSD として表している。

3. 抗体の立体構造モデリング

3.1 鋳型に基づく構造モデリング

抗体の立体構造モデリングとは、通常は抗体の可変領域の構造予測のことを指している。上述 したように、抗体は CDR-H3 以外は比較的よく保存されている。従って、抗体の構造モデリング では、CDR-H3 以外は立体構造が既知の抗体を鋳型として用いることができる。そのため、CDR-H3 以外の構造は、ほとんどの場合、それなりの精度(Cα-RMSD < 1.0Å)で予測することができ る。鋳型の選択は、CDR やフレームワーク領域といった部分配列ごとの BLAST 検索で十分であ る[20]。注意点としては、抗体配列全長ではなく、部分構造に分割して考えるという点にある。 また、BLAST を用いた単純なアミノ酸配列の相同性に基づく手法以外にも、機械学習(勾配ブ ースティング)による CDR の鋳型選択も提案されている[21]。

抗体の構造モデリングの際には、抗体はL鎖とH鎖の二量体ということも考慮する必要がある。つまり、可変領域を構成するL鎖とH鎖の相対配置も予測する必要がある。こうした問題に対し、計算化学を活用した取り組みも、複数報告されている[22–25]。RosettaAntibody[20]では、 CDRを除いたL鎖とH鎖を人工的につなげた1本のアミノ酸配列に対してBLAST検索を行い、 最も配列相同性の高いPDB中の抗体をL鎖/H鎖の相対配置の鋳型としている。古典分子力場を 用いた評価による最安定構造が、結晶構造中に見られるL鎖/H鎖の相対配置をよく再現してい たという報告もある[22]。L鎖/H鎖の相対配置と、アミノ酸配列や認識する抗原の種類との相関 も指摘されている[24]。

非常に多様性に富む CDR-H3 に対しても、鋳型探索に基づく構造予測が試みられている[26-29]。後述するように、現状では、どの技術を用いても CDR-H3 の立体構造予測は難しいが、鋳型に基づく手法の最大の利点は、その計算コストにある。例えば、抗体の「安定化デザイン」や「ヒト化」では、CDR-H3 の立体構造は重要ではないことも多い。こうした目的では、鋳型に基づく構造予測手法で十分である。また、近年、実験技術の進展も目覚ましく、PDB 中のタンパク質の構造データも爆発的に増加している。そのため、鋳型に基づく構造モデリングにおけるサンプリング量も時間と共に増加し、カバーできる構造空間が広がり、予測精度の向上も期待できる。分子認識における構造-機能相関を考える上でも、抗体以外のタンパク質に CDR-H3 との類似構造が存在しうるという事実は、大変興味深い。

3.2 CDR-H3 のデノボ構造モデリング

鋳型に頼らずに、シミュレーションに基づく手法で CDR-H3 の立体構造を予測しようという 試みも多い[30-33]。MD 計算以外にも、ラマチャンドランプロットに基づく二面角探索によるモ デル構築も行われている[34]。しかし、シミュレーションによるモデリングでは、ループが長く なると探索空間も膨大になり、現状のシミュレーション技術では限界がある。探索空間を制限す るために、シミュレーションの拘束条件として CDR-H3 の配列-構造相関ルールを利用している 例もあるが、他の 5 つの CDR で見られるような高い予測精度には至っていない[35,36]。

タンパク質のループ構造は本来柔軟なはずである。程度の差こそあれ、溶媒中では常に揺らい でいると考えられる。「単一構造」を予測しようとするのではなく、「アンサンブル予測」として CDR-H3の構造予測問題を捉えるのが正しい姿勢であろう。こうした目的では、網羅的に構造空 間を探索することが重要となる。従って、拡張アンサンブルなどのシミュレーション手法がより 重要になってくると思われる[33]。こうした網羅的なサンプリング技術を教師データにして学習 することで、より高精度な CDR-H3 の立体構造予測を実現できる可能性がある。

3.3 AlphaFold2/ディープラーニングによる抗体の構造モデリング

近年話題の AlphaFold2[37]も抗体の立体構造予測に適用できる。しかし、AlphaFold2 をもって しても、CDR-H3 の立体構造を予測することは、現状では難しい。AlphaFold2 以外にも、 DeepAb[38]や、IgFold[39]、ABlooper[40]など、抗体に特化したディープラーニングに基づく立体 構造予測手法も提案されているが、やはり CDR-H3 の予測精度が課題となっている。こうした 手法をベンチマークした論文によると、正解構造に対する RMSD の平均値が 2Å を超えている [39]。抗体のモデル構造がこれほど結晶構造からずれてしまうと、その後のドッキング計算や抗 原への結合親和性のデザイン計算も成功する可能性は極めて低い[41]。一方で、安定化のための デザイン計算やヒト化が目的なら、十分実用に耐えうる予測精度だと考えられる。

3.4 抗体-抗原相互作用のモデリング

一般的に、タンパク質-タンパク質ドッキング計算の精度はそれほど高くない。特に分子認識 に伴う構造変化をドッキング過程に取り組むことは、現状では容易ではない。通常、抗原は生体 「外」に存在する。従って、生体「内」に存在する抗体とは独立した進化過程を辿っている。 AlphaFold2 は、タンパク質同士の共進化に基づく構造予測手法であるため、抗体-抗原複合体の 構造予測に適用しても、その精度は極めて限定的となっている。

分子の構造変化に加え、ドッキング計算で問題となることの一つに、その探索空間の広さがあ る。しかし、抗体は CDR で抗原を認識している。そこで、こうした情報を利用し、抗体に特化 したドッキング計算手法も提案されているが[42-45]、抗原エピトープの情報が利用できない場 合には、予測精度は格段に落ちてしまう。従って、抗体-抗原複合体の構造情報を得たい場合に は、水素-重水素交換質量分析(HDX-MS)やアラニンスキャンなどの実験手法を用いてエピト ープを十分に絞り込み、ドッキング計算の探索空間を絞り込むことが重要となる。

計算化学を活用した抗原エピトープの絞り込みも、さまざまな手法が提案されている。この「B 細胞エピトープ予測問題」は、抗体科学の中でも、機械学習の応用が盛んな領域の一つとなって いる[46-54]。ここでの問題設定は「抗原配列/構造を入力として、抗体が一般的に認識しやすい 領域(アミノ酸配列)を予測する」となる。教師データは、PDBに登録されている抗体-抗原複 合体の立体構造や IEDB[55]に登録されているエピトープ配列である。しかしながら、言うまで もなく抗原は無数に存在するため、こうしたデータベースに登録されているエピトープ情報は 極めて限定的と考えられる。免疫系は、原則どんな抗原配列に対しても抗体を作製できるように 思える。つまり、任意の抗体の抗原エピトープ配列を実験的に同定することは比較的容易な一方 で、免疫系にとって、真に「エピトープにならない配列」を定義することは極めて困難であり、 エピトープ予測問題を難しくしている。そこで、抗原/抗体両方のアミノ酸配列の組成に基づき、 抗体ごとに個別の抗原エピトープを予測する手法も提案されている[56-58]。

相補性決定領域という名前が示すように、抗体の抗原認識には、「相補性」が重要となる。その相補性を決めているのは、当然、CDRのアミノ酸配列と立体構造である。PDB中の抗体-抗原 複合体に基づき、共通した配列/構造上の特徴を統計的に見出そうという研究も多くの報告がある。こうした特徴量は機械学習による抗体-抗原相互作用予測に利用されている[59]。

PDB 中のタンパク質-タンパク質相互作用の「かたちの相補性」を定量化するために、「Sc」という指標が広く使われている[60]。Sc が考案された当初は、抗原が生体内に存在する抗体とは独立して進化しているため、抗体-抗原複合体のかたちの相補性は、一般的なタンパク質間相互作用面よりも劣ると考えられていた。しかしその後、こうした低い Sc 値は、計算に用いた結晶構造の分解能の影響であることが示され、現在では、抗体-抗原複合体の Sc は、他のタンパク質複合体と同等とされている[61]。

H3 を除く5 つの CDR では、そのカノニカル構造の「組み合わせ」も限られている[62]。つま り、H3 以外の CDR が構成する抗原認識部位は、その全体の形状もある程度限られていることに なる。そこで、抗原認識部位の形状の違いと、認識する抗原の大まかな種類(タンパク質/ペプ チド/ハプテンなど)との相関関係が指摘されている[63-66]。特に CDR-L1 が長いと抗原認識部 位に凹凸が生まれ、ハプテンやペプチドなどの分子量が小さい抗原を認識しやすくなる傾向が ある。一方で、CDR-L1 が短くなると、抗原認識部位が平らになり、タンパク質抗原を認識しや すくなる。こうした情報は、例えば抗体の合成ライブラリの作製や、既存の抗体の異なる抗原へ のリポジショニングの際に有用である。

近年の次世代シークエンサー(NGS: Next Generation Sequencing)の発展により、B細胞が産生す る抗体の情報を丸ごと読めるような時代になっている。その結果得られるのは、大量の抗体遺伝 子情報である。これらは「B細胞レパトア」あるいは「抗体レパトア」と呼ばれている。そこで、 抗体配列の情報から、それらの抗体が認識する抗原を予測しようというモチベーションが生ま れる。しかしながら、抗体の遺伝子情報のみから、各抗体が認識する抗原をゼロベースで予測す ることは、現実的な問題設定ではない。ある程度抗原の候補が限られていれば、ドッキング計算 や機械学習に基づき、抗体が認識しうる抗原を予測することは可能かもしれない。例えば、ドッ キング計算は常にタンパク質同士を「結合」させてしまうが、評価関数により、その結合のしや すさの傾向を評価することができる。また、同じ抗原エピトープを認識する抗体は、相同性の高 い配列や構造上の特徴を持つと考えられる。そこで、抗原エピトープが既知の抗体を教師データ とすることで、抗原エピトープが未知の抗体の特異性を予測できる可能性がある。

4. 抗体の立体構造と「進化」

抗原が抗体とは独立した進化過程を辿っていることは既に述べた。一方で、抗体は抗原に適応して生体内で「進化」できる。この過程は抗体の「親和性成熟」と呼ばれる。その名前が示す通り、成熟後の抗体は、抗原に対する「結合親和性」が向上していると考えられている。この成熟の過程で、抗体がどのように変化していくのかを理解することは、抗体創薬においても、基礎免疫学においても、意義が大きい。多くの研究では、10個以下の少数の抗体に対して、成熟前後の立体構造を、実験あるいは計算化学の手法を用いて比較することで、「抗体は成熟するにつれ、その分子構造が硬くなる」と結論づけている[67-84]。これは、抗体があらかじめ剛直になることで、抗原認識に伴うエントロピーのロスが減少し、結合親和性を向上させるためであると説明されている。しかし、より最近の、数千規模の抗体のアミノ酸配列およびその立体構造を、剛性理論により比較解析した研究によると、必ずしも剛直になるわけではないとされている[85]。

上述したように、6 つの CDR やL 鎖および H 鎖のフレームワーク領域など、抗体の立体構造は少数の部分構造に分けて考えることができる。抗原認識をはじめとした抗体の機能発現においても、それぞれの部分構造が異なる役割を果たす。親和性成熟により CDR-H3 は硬くなる一方で、他の CDR の柔軟性が増すなど、抗原の物性・形状に応じて、抗体成熟の過程で各部分構造が異なる挙動を示すと考えるのが自然であろう。

計算化学により抗体を「進化」させようという研究も盛んに行われている。コンピュータを 使ったアミノ酸配列のデザイン計算の基本は、与えられたタンパク質の主鎖構造に対して、それ らを最も「安定化」させる側鎖構造の組み合わせを探索するという、組合せ最適化問題である。 つまり、アミノ酸配列からその立体構造を予測するという、立体構造予測の逆問題となっている (図1)。シミュレーションに基づく生体分子のデザイン計算に関する研究は 1990 年代から報告 があり、その適用範囲が徐々に拡張されている。立体構造に基づく、計算化学を用いた抗体の分 子設計も、2006 年ごろから登場している。それらの多くはシミュレーションに基づく分子設計 であったが[86-93]、より最近では、ディープラーニングや言語モデルを筆頭に機械学習に基づ く抗体アミノ酸配列の改変・生成に多くの研究者が挑んでいる[94-97]。ここでは、近年爆発的に 増えている B 細胞レパトア情報やファージディスプレイなどの進化工学の実験結果を教師デー タとし、新たなアミノ酸配列をコンピュータ上でサンプリングすることを目的としている。タン パク質はわずか 20 種類の限られた文字列で表現できることから、自然言語処理における言語モ デルとの相性が良い[98]。特に抗体は、限られた数の germline 遺伝子からスタートし、抗原に適 応して生体内での進化を経て、膨大な数の抗体遺伝子が創出される。このことから、免疫系、特 に B 細胞が産生する抗体レパトアは、言語モデルによる表現に適したよいモデル系であると言 える[99-101]。NGS により得られる膨大なレパトア情報に加え、臨床試験まで到達した 137 の抗 体医薬品候補の配列情報や実験による物性データの測定値なども公開されており[102]、動物の 免疫系ではなく、機械により創製された抗体医薬品が上市される日も近いのかもしれない。

5. おわりに

本稿では、医薬品のモダリティとして「抗体」にフォーカスし、計算化学を用いた抗体研究を 概観した。AlphaFold2は構造生物学の分野に大きなインパクトを与えているが、抗体の立体構造 モデリングは、まだまだ未解決の課題である。抗体に特化したディープラーニングによる構造予 測手法も提案されているが、計算化学だけで完結できるほどの予測精度はまだない。抗体創薬に おいても、実験による評価と計算化学を用いた予測のサイクルを回すことが重要となっている。 コンピュータの演算速度は年々格段に速くなっており、計算化学による予測や分子デザインは 取り組みやすい環境になっている。従って、それを評価する実験側のスループット性が重要な鍵 となる。

低分子創薬における「ルール・オブ・ファイブ」に相当する、抗体の「薬らしさの指標」や 「Developability」も、抗体の構造情報に基づき提案されており、抗体の立体構造は創薬を進める 上で欠かすことのできない情報となっている[103]。実験的に明らかにできる立体構造の数は限 られている。そうした実験のスループット性を補うのも、計算化学の重要な役割である。

抗体医薬品の欠点は、その大きさにある。細胞膜を透過することができないため、抗体医薬品 の標的は細胞外タンパク質に限定される。こうした問題点を克服するために、抗体を低分子量化 しようという試みもある。抗原認識への寄与の大きい一部の CDR のみを切り出すなど、抗体か らペプチド医薬品の開発につなげようという研究も複数報告されている[104,105]。抗原認識には、 抗原タンパク質の立体構造と相補的な CDR の立体構造が重要である。単に CDR を切り出すだ けでは、その立体構造を維持することが難しく、CDR ペプチドを環状化するなどの工夫が必要 となる。このように、医薬品のモダリティを横断した研究が今後も盛んに行われていくと思われ る。そうした中で、今後ますます計算化学が重要な役割を果たすのは間違いない。

謝辞

本稿を執筆する機会を与えてくださいました SAR News 編集委員長である幸瞳先生(理化学研究所)および高谷大輔先生(大阪大学)に感謝いたします。

参考文献

- P.C. Taylor, A.C. Adams, M.M. Hufford, et al. Neutralizing monoclonal antibodies for treatment of COVID-19, *Nat Rev Immunol.* 21 (2021) 382–393. https://doi.org/10.1038/s41577-021-00542-x
- D. Kuroda, H. Shirai, M.P. Jacobson, et al. Computer-aided antibody design, *Protein Engineering*, Design and Selection. 25 (2012) 507–521. https://doi.org/10.1093/protein/gzs024
- [3] D. Kuroda, K. Tsumoto, Engineering Stability, Viscosity, and Immunogenicity of Antibodies by Computational Design, J Pharm Sci. 109 (2020) 1631–1651. https://doi.org/10.1016/j.xphs.2020.01.011
- [4] A.M. Hummer, B. Abanades, C.M. Deane, Advances in computational structure-based antibody design, *Curr Opin Struct Biol.* 74 (2022) 102379. https://doi.org/10.1016/j.sbi.2022.102379
- [5] C. Havenar-Daughton, R.K. Abbott, W.R. Schief, et al. When designing vaccines, consider the starting material: the human B cell repertoire, *Curr Opin Immunol.* 53 (2018) 209–216. https://doi.org/10.1016/j.coi.2018.08.002
- [6] P.D. Kwong, B.J. DeKosky, J.B. Ulmer, Antibody-guided structure-based vaccines, *Semin Immunol.* 50 (2020) 101428. https://doi.org/10.1016/j.smim.2020.101428
- [7] L.J. Harris, S.B. Larson, K.W. Hasel, et al. Refined structure of an intact IgG2a monoclonal antibody, *Biochemistry*. 36 (1997) 1581–1597. https://doi.org/10.1021/bi962514+
- [8] S.M. Lewis, X. Wu, A. Pustilnik, et al. Generation of bispecific IgG antibodies by structure-based design of an orthogonal Fab interface, *Nat Biotechnol.* 32 (2014) 191–198. https://doi.org/10.1038/nbt.2797
- [9] A. Chaudhri, I.E. Zarraga, T.J. Kamerzell, et al. Coarse-Grained Modeling of the Self-Association of Therapeutic Monoclonal Antibodies, *J Phys Chem B.* 116 (2012) 8045–8057.

https://doi.org/10.1021/jp301140u

- [10] A. Tucs, K. Tsuda, A. Sljoka, Probing conformational dynamics of antibodies with geometric simulations, arXiv (2021). https://doi.org/10.48550/arXiv.2109.14724
- [11] R.L. Stanfield, A. Zemla, I.A. Wilson, et al. Antibody elbow angles are influenced by their light chain class, *J Mol Biol.* 357 (2006) 1566–1574. https://doi.org/10.1016/j.jmb.2006.01.023
- [12] M. Dondelinger, P. Filée, E. Sauvage, et al. Understanding the significance and implications of antibody numbering and antigen-binding surface/residue definition, *Front Immunol.* 9 (2018). https://doi.org/10.3389/fimmu.2018.02278
- [13] T. te Wu, E.A. Kabat, An analysis of the sequences of the variable regions of bence jones proteins and myeloma light chains and their implications for antibody complementarity, *Journal of Experimental Medicine*. **132** (1970) 211–250. https://doi.org/10.1084/jem.132.2.211
- [14] B. Al-Lazikani, A.M. Lesk, C. Chothia, Standard conformations for the canonical structures of immunoglobulins, *J Mol Biol.* 273 (1997) 927–948. https://doi.org/10.1006/jmbi.1997.1354
- C. Chothia, A.M. Lesk, Canonical structures for the hypervariable regions of immunoglobulins, J Mol Biol. 196 (1987) 901–917. https://doi.org/10.1016/0022-2836(87)90412-8
- H. Shirai, A. Kidera, H. Nakamura, H3-rules: identification of CDR-H3 structures in antibodies, FEBS Lett. 455 (1999) 188–197. https://doi.org/10.1016/S0014-5793(99)00821-2
- [17] D. Kuroda, H. Shirai, M. Kobori, et al. Structural classification of CDR-H3 revisited: A lesson in antibody modeling, *Proteins: Structure, Function and Bioinformatics*. **73** (2008) 608–620. https://doi.org/10.1002/prot.22087
- [18] D. Kuroda, K. Tsumoto, Structural Classification of CDR-H3 in Single-Domain VHH Antibodies, Methods in Molecular Biology, 2022: in press.
- [19] I. Sela-Culang, S. Alon, Y. Ofran, A Systematic Comparison of Free and Bound Antibodies Reveals Binding-Related Conformational Changes, *The Journal of Immunology*. **189** (2012) 4890–4899. https://doi.org/10.4049/jimmunol.1201493
- [20] B.D. Weitzner, J.R. Jeliazkov, S. Lyskov, et al. Modeling and docking of antibody structures with Rosetta, *Nat Protoc.* 12 (2017) 401–416. https://doi.org/10.1038/nprot.2016.180
- [21] X. Long, J.R. Jeliazkov, J.J. Gray, Non-H3 CDR template selection in antibody modeling through machine learning, *PeerJ.* 7 (2019) e6179. https://doi.org/10.7717/peerj.6179
- [22] A. Narayanan, B.D. Sellers, M.P. Jacobson, Energy-Based Analysis and Prediction of the Orientation between Light- and Heavy-Chain Antibody Variable Domains, *J Mol Biol.* 388 (2009) 941–953. https://doi.org/10.1016/j.jmb.2009.03.043
- [23] K.R. Abhinandan, A.C.R. Martin, Analysis and prediction of V_H/V_L packing in antibodies, *Protein Engineering, Design and Selection.* 23 (2010) 689–697. https://doi.org/10.1093/protein/gzq043
- [24] A. Chailyan, P. Marcatili, A. Tramontano, The association of heavy and light chain variable domains in antibodies: Implications for antigen specificity, *FEBS Journal.* 278 (2011) 2856–2866.

https://doi.org/10.1111/j.1742-4658.2011.08207.x

- [25] N.A. Marze, S. Lyskov, J.J. Gray, Improved prediction of antibody V_L-V_H orientation, *Protein Engineering, Design and Selection.* 29 (2016) 409–418. https://doi.org/10.1093/protein/gzw013
- [26] M. Lis, T. Kim, J.J. Sarmiento, et al. Bridging the gap between single-template and fragment based protein structure modeling using Spanner, *Immunome Res.* 7 (2011) 1–8. https://doi.org/10.4172/1745-7580.1000041
- [27] M.A. Messih, R. Lepore, P. Marcatili, et al. Improving the accuracy of the structure prediction of the third hypervariable loop of the heavy chains of antibodies, *Bioinformatics*. **30** (2014) 2733– 2740. https://doi.org/10.1093/bioinformatics/btu194
- [28] R. Lepore, P.P. Olimpieri, M.A. Messih, et al. PIGSPro: prediction of immunoGlobulin structures v2, *Nucleic Acids Res.* 45 (2017) W17–W23. https://doi.org/10.1093/nar/gkx334
- [29] D. Schritt, S. Li, J. Rozewicki, et al. Repertoire Builder: high-throughput structural modeling of B and T cell receptors, *Mol Syst Des Eng.* 4 (2019) 761–768. https://doi.org/10.1039/C9ME00020H
- [30] R.E. Bruccoleri, J. Novotny, Antibody modeling using the conformational search program CONGEN, *Immunomethods.* 1 (1992) 96–106. https://doi.org/10.1016/S1058-6687(05)80033-6
- [31] B.D. Weitzner, D. Kuroda, N. Marze, et al. Blind prediction performance of RosettaAntibody 3.0: Grafting, relaxation, kinematic loop modeling, and full CDR optimization, *Proteins: Structure, Function and Bioinformatics.* 82 (2014) 1611–1623. https://doi.org/10.1002/prot.24534
- [32] K. Zhu, T. Day, D. Warshaviak, et al. Antibody structure determination using a combination of homology modeling, energy-based refinement, and loop prediction, *Proteins: Structure, Function* and Bioinformatics. 82 (2014) 1646–1655. https://doi.org/10.1002/prot.24551
- [33] H. Nishigami, N. Kamiya, H. Nakamura, Revisiting antibody modeling assessment for CDR-H3 loop, *Protein Engineering, Design and Selection.* 29 (2016) 477–484. https://doi.org/10.1093/protein/gzw028
- [34] M.P. Jacobson, D.L. Pincus, C.S. Rapp, et al. A Hierarchical Approach to All-Atom Protein Loop Prediction, *Proteins: Structure, Function and Genetics*. 55 (2004) 351–367. https://doi.org/10.1002/prot.10613
- [35] J.A. Finn, J. Koehler Leman, J.R. Willis, et al. Improving Loop Modeling of the Antibody Complementarity-Determining Region 3 Using Knowledge-Based Restraints, *PLoS One.* 11 (2016) e0154811. https://doi.org/10.1371/journal.pone.0154811
- [36] B.D. Weitzner, J.J. Gray, Accurate Structure Prediction of CDR H3 Loops Enabled by a Novel Structure-Based C-Terminal Constraint, *The Journal of Immunology.* 198 (2017) 505–515. https://doi.org/10.4049/jimmunol.1601137
- [37] J. Jumper, R. Evans, A. Pritzel, et al. Highly accurate protein structure prediction with AlphaFold, *Nature*. 596 (2021) 583–589. https://doi.org/10.1038/s41586-021-03819-2
- [38] J.A. Ruffolo, J. Sulam, J.J. Gray, Antibody structure prediction using interpretable deep learning,

Patterns. 3 (2022) 100406. https://doi.org/10.1016/j.patter.2021.100406

- [39] J.A. Ruffolo, L.-S. Chu, S.P. Mahajan, et al. Fast, accurate antibody structure prediction from deep learning on massive set of natural antibodies, *bioRxiv.* (2022). https://doi.org/10.1101/2022.04.20.488972
- [40] B. Abanades, G. Georges, A. Bujotzek, et al. ABlooper: fast accurate antibody CDR loop structure prediction with accuracy estimation, *Bioinformatics*. 38 (2022) 1877–1880. https://doi.org/10.1093/bioinformatics/btac016
- [41] D. Kuroda, J.J. Gray, Pushing the Backbone in Protein-Protein Docking, *Structure*. 24 (2016) 1821–1829. https://doi.org/10.1016/j.str.2016.06.025
- [42] A. Sircar, J.J. Gray, SnugDock: Paratope Structural Optimization during Antibody-Antigen Docking Compensates for Errors in Antibody Homology Models, *PLoS Comput Biol.* 6 (2010) e1000644. https://doi.org/10.1371/journal.pcbi.1000644
- [43] R. Brenke, D.R. Hall, G.Y. Chuang, et al. Application of asymmetric statistical potentials to antibody-protein docking, *Bioinformatics*. 28 (2012) 2608–2614. https://doi.org/10.1093/bioinformatics/bts493
- [44] K. Krawczyk, T. Baker, J. Shi, et al. Antibody i-Patch prediction of the antibody binding site improves rigid local antibody-antigen docking, *Protein Engineering, Design and Selection.* 26 (2013) 621–629. https://doi.org/10.1093/protein/gzt043
- [45] A. Davila, Z. Xu, S. Li, et al. AbAdapt: an adaptive approach to predicting antibody-antigen complex structures from sequence, *Bioinformatics Advances.* 2 (2022) vbac015. https://doi.org/10.1093/bioadv/vbac015
- [46] S. Saha, G.P.S. Raghava, Prediction of continuous B-cell epitopes in an antigen using recurrent neural network, *Proteins: Structure, Function and Genetics.* 65 (2006) 40–48. https://doi.org/10.1002/prot.21078
- [47] M.J. Sweredoski, P. Baldi, COBEpro: A novel system for predicting continuous B-cell epitopes, *Protein Engineering, Design and Selection.* 22 (2009) 113–120. https://doi.org/10.1093/protein/gzn075
- [48] K. Krawczyk, X. Liu, T. Baker, et al. Improving B-cell epitope prediction and its application to global antibody-antigen docking, *Bioinformatics*. 30 (2014) 2288–2294. https://doi.org/10.1093/bioinformatics/btu190
- [49] M.C. Jespersen, B. Peters, M. Nielsen, et al. BepiPred-2.0: Improving sequence-based B-cell epitope prediction using conformational epitopes, *Nucleic Acids Res.* 45 (2017) W24–W29. https://doi.org/10.1093/nar/gkx346
- [50] T. Liu, K. Shi, W. Li, Deep learning methods improve linear B-cell epitope prediction, *BioData Min.* 13 (2020) 1. https://doi.org/10.1186/s13040-020-00211-0
- [51] M. Collatz, F. Mock, E. Barth, et al. Epidope: A deep neural network for linear b-cell epitope

prediction, Bioinformatics. 37 (2021) 448-455. https://doi.org/10.1093/bioinformatics/btaa773

- [52] B.M. da Silva, Y. Myung, D.B. Ascher, et al. Epitope3D: A machine learning method for conformational B-cell epitope prediction, *Brief Bioinform.* 23 (2022) bbab423. https://doi.org/10.1093/bib/bbab423
- [53] W. Alghamdi, M. Attique, E. Alzahrani, et al. LBCEPred: a machine learning model to predict linear
 B-cell epitopes, *Brief Bioinform.* 23 (2022) bbac035. https://doi.org/10.1093/bib/bbac035
- [54] S. Pittala, C. Bailey-Kellogg, Learning context-aware structural representations to predict antigen and antibody binding interfaces, *Bioinformatics*. 36 (2020) 3996–4003. https://doi.org/10.1093/bioinformatics/btaa263
- [55] R. Vita, S. Mahajan, J.A. Overton, et al. The Immune Epitope Database (IEDB): 2018 update, *Nucleic Acids Res.* 47 (2019) D339–D343. https://doi.org/10.1093/nar/gky1006
- [56] S. Soga, D. Kuroda, H. Shirai, et al. Use of amino acid composition to predict epitope residues of individual antibodies, *Protein Engineering, Design and Selection.* 23 (2010) 441–448. https://doi.org/10.1093/protein/gzq014
- [57] I. Sela-Culang, S. Ashkenazi, B. Peters, et al. PEASE: Predicting B-cell epitopes utilizing antibody sequence, *Bioinformatics*. **31** (2015) 1313–1315. https://doi.org/10.1093/bioinformatics/btu790
- [58] M.C. Jespersen, S. Mahajan, B. Peters, et al. Antibody specific B-cell epitope predictions: Leveraging information from antibody-antigen protein complexes, *Front Immunol.* 10 (2019). https://doi.org/10.3389/fimmu.2019.00298
- [59] R. Akbar, P.A. Robert, M. Pavlović, et al. A compact vocabulary of paratope-epitope interactions enables predictability of antibody-antigen binding, *Cell Rep.* 34 (2021) 108856. https://doi.org/10.1016/j.celrep.2021.108856
- [60] M.C. Lawrence, P.M. Colman, Shape complementarity at protein/protein interfaces., *J Mol Biol.* 234 (1993) 946–50. https://doi.org/10.1006/jmbi.1993.1648
- [61] D. Kuroda, J.J. Gray, Shape complementarity and hydrogen bond preferences in protein-protein interfaces: Implications for antibody modeling and protein-protein docking, *Bioinformatics*. 32 (2016) 2451–2456. https://doi.org/10.1093/bioinformatics/btw197
- [62] E. Vargas-Madrazo, F. Lara-Ochoa, J. Carlos Almagro, Canonical Structure Repertoire of the Antigen-binding Site of Immunoglobulins Suggests Strong Geometrical Restrictions Associated to the Mechanism of Immune Recognition, J Mol Biol. 254 (1995) 497–504. https://doi.org/10.1006/jmbi.1995.0633
- [63] M. Lee, P. Lloyd, X. Zhang, et al. Shapes of antibody binding sites: Qualitative and quantitative analyses based on a geomorphic classification scheme, *Journal of Organic Chemistry.* 71 (2006) 5082–5092. https://doi.org/10.1021/j0052659z
- [64] A.V.J. Collis, A.P. Brouwer, A.C.R. Martin, Analysis of the antigen combining site: Correlations between length and sequence composition of the hypervariable loops and the nature of the antigen,

J Mol Biol. 325 (2003) 337-354. https://doi.org/10.1016/S0022-2836(02)01222-6

- [65] R.M. MacCallum, A.C.R. Martin, J.M. Thornton, Antibody-antigen interactions: Contact analysis and binding site topography, J Mol Biol. 262 (1996) 732–735. https://doi.org/10.1006/jmbi.1996.0548
- [66] G. Raghunathan, J. Smart, J. Williams, et al. Antigen-binding site anatomy and somatic mutations in antibodies that recognize different types of antigens, *Journal of Molecular Recognition*. 25 (2012) 103–113. https://doi.org/10.1002/jmr.2158
- [67] G.J. Wedemayer, P.A. Patten, L.H. Wang, et al. Structural Insights into the Evolution of an Antibody Combining Site, *Science* 276 (1997) 1665–1669. https://doi.org/10.1126/science.276.5319.1665
- [68] L.T. Chong, Y. Duan, L. Wang, et al. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7, *Proc Natl Acad Sci U S A.* 96 (1999) 14330–14335. https://doi.org/10.1073/pnas.96.25.14330
- [69] E.C. Mundorff, M.A. Hanson, A. Varvak, et al. Conformational effects in biological catalysis: An antibody-catalyzed oxy-Cope rearrangement, *Biochemistry.* **39** (2000) 627–632. https://doi.org/10.1021/bi9924314
- [70] V. Manivel, N.C. Sahoo, D.M. Salunke, et al. Maturation of an antibody response is governed by modulations in flexibility of the antigen-combining site, *Immunity.* 13 (2000) 611–620. https://doi.org/10.1016/S1074-7613(00)00061-3
- [71] J. Yin, E.C. Mundorff, P.L. Yang, et al. A comparative analysis of the immunological evolution of antibody 28B4, *Biochemistry*. 40 (2001) 10764–10773. https://doi.org/10.1021/bi010536c
- [72] R. Jimenez, G. Salazart, K.K. Baldridge, et al. Flexibility and molecular recognition in the immune system, *Proc Natl Acad Sci U S A*. 100 (2003) 92–97. https://doi.org/10.1073/pnas.262411399
- [73] J. Yin, A.E. Beuscher, S.E. Andryski, et al. Structural plasticity and the evolution of antibody affinity and specificity., J Mol Biol. 330 (2003) 651–656. https://doi.org/10.1016/S0022-2836(03)00631-4
- [74] R. Jimenez, G. Salazar, J. Yin, et al. Protein dynamics and the immunological evolution of molecular recognition, *Proc Natl Acad Sci U S A.* 101 (2004) 3803–3808. https://doi.org/10.1073/pnas.0305745101
- [75] J. Zimmermann, E.L. Oakman, I.F. Thorpe, et al. Antibody evolution constrains conformational heterogeneity by tailoring protein dynamics, *Proceedings of the National Academy of Sciences*. 103 (2006) 13722–13727. https://doi.org/10.1073/pnas.0603282103
- [76] I.F. Thorpe, C.L. Brooks, Molecular evolution of affinity and flexibility in the immune system, *Proceedings of the National Academy of Sciences.* 104 (2007) 8821–8826. https://doi.org/10.1073/pnas.0610064104
- [77] M. Babor, T. Kortemme, Multi-constraint computational design suggests that native sequences of germline antibody H3 loops are nearly optimal for conformational flexibility, *Proteins: Structure*,

Function and Bioinformatics. 75 (2009) 846-858. https://doi.org/10.1002/prot.22293

- [78] J. Zimmermann, F.E. Romesberg, C.L. Brooks, et al. Molecular description of flexibility in an antibody combining site, *Journal of Physical Chemistry B.* 114 (2010) 7359–7370. https://doi.org/10.1021/jp906421v
- [79] S.E. Wong, B.D. Sellers, M.P. Jacobson, Effects of somatic mutations on CDR loop flexibility during affinity maturation, *Proteins: Structure, Function and Bioinformatics*. **79** (2011) 821–829. https://doi.org/10.1002/prot.22920
- [80] R. Adhikary, W. Yu, M. Oda, et al. Protein dynamics and the diversity of an antibody response, *Journal of Biological Chemistry.* 287 (2012) 27139–27147. https://doi.org/10.1074/jbc.M112.372698
- [81] A.G. Schmidt, H. Xu, A.R. Khan, et al. Preconfiguration of the antigen-binding site during affinity maturation of a broadly neutralizing influenza virus antibody, *Proceedings of the National Academy* of Sciences. 110 (2013) 264–269. https://doi.org/10.1073/pnas.1218256109
- [82] J.R. Willis, B.S. Briney, S.L. DeLuca, et al. Human germline antibody gene segments encode polyspecific antibodies., *PLoS Comput Biol.* 9 (2013) e1003045. https://doi.org/10.1371/journal.pcbi.1003045
- [83] T. Li, M.B. Tracka, S. Uddin, et al. Rigidity Emerges during Antibody Evolution in Three Distinct Antibody Systems: Evidence from QSFR Analysis of Fab Fragments, *PLoS Comput Biol.* 11 (2015) e1004327. https://doi.org/10.1371/journal.pcbi.1004327
- [84] M.C. Demirel, A.M. Lesk, Molecular forces in antibody maturation, *Phys Rev Lett.* 95 (2005) 208106. https://doi.org/10.1103/PhysRevLett.95.208106
- [85] J.R. Jeliazkov, A. Sljoka, D. Kuroda, et al. Repertoire Analysis of Antibody CDR-H3 Loops Suggests Affinity Maturation Does Not Typically Result in Rigidification, *Front Immunol.* 9 (2018). https://doi.org/10.3389/fimmu.2018.00413
- [86] L.A. Clark, P.A. Boriack-Sjodin, J. Eldredge, et al. Affinity enhancement of an in vivo matured therapeutic antibody using structure-based computational design, *Protein Science*. 15 (2006) 949– 960. https://doi.org/10.1110/ps.052030506
- [87] S.M. Lippow, K.D. Wittrup, B. Tidor, Computational design of antibody-affinity improvement beyond in vivo maturation, *Nat Biotechnol.* 25 (2007) 1171–1176. https://doi.org/10.1038/nbt1336
- [88] R. Barderas, J. Desmet, P. Timmerman, et al. Affinity maturation of antibodies assisted by in silico modeling., *Proc. Nat. Acad. Sci. USA.* 105 (2008) 9029–9034. https://doi.org/10.1073/pnas.0801221105
- [89] C.J. Farady, B.D. Sellers, M.P. Jacobson, et al. Improving the species cross-reactivity of an antibody using computational design, *Bioorg Med Chem Lett.* 19 (2009) 3744–3747. https://doi.org/10.1016/j.bmcl.2009.05.005
- [90] M. Kiyoshi, J.M.M. Caaveiro, E. Miura, et al. Affinity improvement of a therapeutic antibody by

structure-based computational design: Generation of electrostatic interactions in the transition state stabilizes the antibody-antigen complex, *PLoS One.* **9** (2014) e87099. https://doi.org/10.1371/journal.pone.0087099

- [91] J.R. Willis, G. Sapparapu, S. Murrell, et al. Redesigned HIV antibodies exhibit enhanced neutralizing potency and breadth, *Journal of Clinical Investigation*. **125** (2015) 2523–2531. https://doi.org/10.1172/JCI80693
- [92] D. Baran, M.G. Pszolla, G.D. Lapidoth, et al. Principles for computational design of binding antibodies, *Proceedings of the National Academy of Sciences*. **114** (2017) 10900–10905. https://doi.org/10.1073/pnas.1707171114
- [93] G. Nimrod, S. Fischman, M. Austin, et al. Computational Design of Epitope-Specific Functional Antibodies, *Cell Rep.* 25 (2018) 2121-2131.e5. https://doi.org/10.1016/j.celrep.2018.10.081
- [94] G. Liu, H. Zeng, J. Mueller, et al. Antibody complementarity determining region design using high-capacity machine learning, *Bioinformatics*. 36 (2020) 2126–2133. https://doi.org/10.1093/bioinformatics/btz895
- [95] K. Saka, T. Kakuzaki, S. Metsugi, et al. Antibody design using LSTM based deep generative model from phage display library for affinity maturation, *Sci Rep.* 11 (2021) 5852. https://doi.org/10.1038/s41598-021-85274-7
- [96] D.M. Mason, S. Friedensohn, C.R. Weber, et al. Optimization of therapeutic antibodies by predicting antigen specificity from antibody sequence via deep learning, *Nat Biomed Eng.* 5 (2021) 600–612. https://doi.org/10.1038/s41551-021-00699-9
- [97] S. Shan, S. Luo, Z. Yang, et al. Deep learning guided optimization of human antibody against SARS-CoV-2 variants with broad neutralization, *Proc Natl Acad Sci U S A*. **119** (2022) e2122954119. https://doi.org/10.1073/pnas.2122954119
- [98] T. Bepler, B. Berger, Learning the protein language: Evolution, structure, and function, *Cell Syst.* 12 (2021) 654-669.e3. https://doi.org/10.1016/j.cels.2021.05.017
- [99] R.W. Shuai, J.A. Ruffolo, J.J. Gray, Generative Language Modeling for Antibody Design, *bioRxiv*.
 (2021). https://doi.org/10.1101/2021.12.13.472419
- [100] T.H. Olsen, I.H. Moal, C.M. Deane, AbLang: an antibody language model for completing antibody sequences, *Bioinformatics Advances*. 2 (2022) vbac046. https://doi.org/10.1093/bioadv/vbac046
- [101] J. Leem, L.S. Mitchell, J.H.R. Farmery, et al. Deciphering the language of antibodies using selfsupervised learning, *Patterns.* 3 (2022) 100513. https://doi.org/10.1016/j.patter.2022.100513
- [102] T. Jain, T. Sun, S. Durand, et al. Biophysical properties of the clinical-stage antibody landscape, Proceedings of the National Academy of Sciences. 114 (2017) 944–949. https://doi.org/10.1073/pnas.1616408114
- [103] M.I.J. Raybould, C. Marks, K. Krawczyk, et al. Five computational developability guidelines for therapeutic antibody profiling, *Proceedings of the National Academy of Sciences*. **116** (2019) 4025–

4030. https://doi.org/10.1073/pnas.1810576116

- [104] R.U. Kadam, J. Juraszek, B. Brandenburg, et al. Potent peptidic fusion inhibitors of influenza virus., *Science.* 358 (2017) 496–502. https://doi.org/10.1126/science.aan0516
- [105] A.M. Sevy, I.M. Gilchuk, B.P. Brown, et al. Computationally Designed Cyclic Peptides Derived from an Antibody Loop Increase Breadth of Binding for Influenza Variants., *Structure*. 28 (2020) 1114-1123.e4. https://doi.org/10.1016/j.str.2020.04.005

///// Cutting Edge /////

CASP におけるモデル構造評価法にみる タンパク質立体構造予測法の発展

北里大学薬学部 清田泰臣、竹田-志鷹真由子

1. はじめに

創薬においてタンパク質立体構造情報は言うまでもなく重要であり、その予測技術の向上は 常に望まれていた。そして近年、人工知能を用いた AlphaFold2 [1]の登場により、タンパク質立 体構造予測技術にブレイクスルーが起きたことは記憶に新しい。これにより、予測されたタンパ ク質立体構造情報の重要性、利用価値はさらに高まると考えられる。本稿では、タンパク質立体 構造予測技術の発展を支え、AlphaFold2 の躍進の舞台ともなった Critical Assessment of Structure Prediction (CASP) について、その概要やモデル構造に対する予測精度の評価法、近年の傾向を 伝えたい。

2. Critical Assessment of Structure Prediction (CASP)について

CASP は、「アミノ酸配列情報からタンパク質立体構造を予測する手法の発展を図る共同実験 (community experiment)」[2]であると運営組織からは紹介されている。CASP は 1994 年より隔年 で開かれ、直近では 2022 年 5 月~8 月にかけて 15 回目の CASP である CASP15 [3]が開催され ている。CASP 開催期間中は、参加グループに対してモデリングターゲット(運営組織が集めた、 実験構造が解明されているがまだ公開されていないタンパク質群)に対応するアミノ酸配列が、 数日の間隔を空けて提供される。参加グループは、出題日からおおよそ 3 週間以内に、予測した モデル構造を提出する必要がある。自動予測サーバーとして参加する場合は、さらに短く 72 時 間以内にモデル構造を提出する必要がある。提出された構造は、カテゴリ毎に決められた評価者 によって細かく評価される。これらの評価を基に、グループとしての成績が決定される。このよ うに参加者にとっては、厳密なブラインドテストを実施できる場にもなっている。

CASP は、評価法によって複数のカテゴリに分けられている。2020 年に開催された CASP14[2] では、High Accuracy Modeling category(単体・ドメイン構造予測カテゴリ)、Topology category(異 なる評価法での単体・ドメイン構造予測カテゴリ)、CASP12 [4]から加わった Assembly category (複合体構造予測カテゴリ)、Contact and Distance Prediction category(残基間コンタクト予測カテ ゴリ)、Refinement category(リファインメントカテゴリ)、Accuracy Estimation category(モデル 精度推定カテゴリ)が実施された。本稿では特に、High Accuracy Modeling category および Assembly category について、その評価法と共に詳細を解説する。

2.1 High Accuracy Modeling category とその評価法

CASP14 における High Accuracy Modeling category は、68 個のモデリングターゲット(ドメインとしては96 個)で構成されていた[2]。AlphaFold2 はこのカテゴリにおいて、圧倒的な成績を収めたことで有名だが、実際にはどのような点が評価されたのか。まずは、このカテゴリにおける各参加グループの成績を決めるためのスコアを解説する。

参加グループは各モデリングターゲットについて、5 個まで予測構造を提出することができる。 その5 個を参加グループ自身で順位付けし、1 位として提出したモデル構造について、評価者は 以下の式(1)に基づいたスコアを与える。CASP13 までは、このようなスコアを全モデリング ターゲットで総和した値をグループの成績としていたが[5]、提出構造数にばらつきのあった CASP14 においてはその中央値を成績としていた[6]。

$$S_{\text{CASP14}} = \left(\frac{1}{16}(Z_{\text{IDDT}} + Z_{\text{CADaa}} + Z_{\text{SG}} + Z_{\text{sidechain}}) + \frac{1}{12}(Z_{\text{MolPrb}} + Z_{\text{backbone}} + Z_{\text{DipDiff}}) + \frac{1}{12}(Z_{\text{MolPrb}} + Z_{\text{backbone}} + Z_{\text{DipDiff}})\right)$$

$$\frac{1}{4}(Z_{\rm GDT-HA} + Z_{\rm ASE})\right) \tag{1}$$

各種 Z は、その評価項目における標準得点を意味するが、CASP においては調整がなされている[5]。まず、全参加グループのデータを用いた標準得点を算出し、外れ値に相当する Z < -2 の データを取り除く。その後、もう一度標準得点を計算し、Z < 0 となった参加者について、最初 に取り除いた参加者も含めて、Z=0 とする。つまり、予測精度の悪い側には下限が存在するが、 予測精度が良い側には上限がないようなスコアとなっている。

また、各種評価項目を大別してみると、①主鎖の予測精度に関わる評価項目: Z_{IDDT} (Local Distance Difference Test [7])、 $Z_{backbone}$ (主鎖ねじれ角の偏差スコア[5])、 Z_{GDT-HA} (Global Distance Test - High Accuracy [8])、 Z_{ASE} (Accuracy Self-Estimate score [9])。②側鎖の予測精度に関わる評価項目: Z_{CADaa} (residue-residue Contact Area Difference score, All Atoms [10])、 Z_{SG} (Sphere-Grinder [11])、 $Z_{sidechain}$ (側鎖ねじれ角の偏差スコア[5])。③原子間の衝突など物理的・幾何学的な評価項目: Z_{MolPrb} (MolProbity clash score [12])、 $Z_{DipDiff}$ (DipScore [13]の差分)。これらモデル構造を評価するための指標をバランス良く含んでおり、このスコア関数を用いた成績は図 1 のようになる[6]。AlphaFold2の好成績は、上記のほぼ全ての評価項目について平均以上でなければ実現できない。CASP13 (2018 年開催) における評価法も近いスコア関数であったが、CASP13 では、ここまで突出したグループは存在しなかった[5]。このように、評価法という観点からみても、AlphaFold2の革新性を垣間見ることができる。



図 1. CASP14 High Accuracy Modeling category の上位 50 グループの成績. https://github.com/JoanaMPereira/CASP14_high_accuracy/より公開されているデータおよびス クリプトを用いて、上位 50 グループの成績に基づいて作成した。評価者による公式なグ ラフは Ref.6 の Fig.4 (A)を参照されたい。

2.2 Assembly category とその評価法

CASP14 における Assembly category は、22 個のモデリングターゲット(巨大な複合体は分割 されて評価されるため、評価ユニットは29 個)で構成されていた[14]。内訳として、ヘテロター ゲットが 14 個、ホモターゲットが 8 個であり、また量体数で分類すると、2 量体ターゲットが 9 個、3 量体ターゲットが 4 個、それ以上が 9 個であった。最も大きなモデリングターゲットは、 27 量体、約 6500 残基にもなる。

(2)

Assembly category においても、各モデリングターゲットに対して、各参加グループが1位として提出したモデル構造を用いて、以下の式(2)に基づいたスコアが与えられる[14]。

$S_{\text{CASP14-assembly}} = Z_{\text{Jaccard}} + Z_{\text{F1}} + Z_{\text{IDDT}} + Z_{\text{TM}}$

各種 Z は High Accuracy Modeling category と同様の調整がなされるが、これらの評価法に関して知るためには、まず Assembly category における相互作用界面の定義について理解しなければならない。CASP においては、自身とは別のボリペプチド鎖から 5 Å 以内、これを相互作用界面に属する残基として Interface amino acids (I) で定義している。Jaccard スコアは IPS (Interface Patch Score) とも呼ばれ、この I を用いて式(3)のように定義される[15]。

$$IPS(M,T) = Jc(M,T) = \frac{|I_M \cap I_T|}{|I_M \cup I_T|}$$
(3)

M および T は予測構造 (Model) と実験構造 (Target) を表す。IPS は 2 つの相互作用界面に属 するアミノ酸残基を正確に予測できているほど大きな値となり、完全に一致していれば 1 とな る。F1 スコアは ICS (Interface Contact Score) とも呼ばれ、式(4) で定義される[15]。

$$ICS(M,T) = F1(P,R) = 2\frac{P(M,T) \times R(M,T)}{P(M,T) + R(M,T)}$$
(4)

PはPrecision(適合率)、RはRecall(再現率)であり、Pはモデル構造の持つ界面コンタクトのうち、正しく予測できた界面コンタクトの割合を示し、Rは実験構造の持つ界面コンタクトのうちどれだけ正しく予測できたかの割合を示すことになる。ICSに関しても、正確に予測しているほど値は大きくなり、完全に一致していれば1となる。一般的にIPS>ICSという傾向が見られ、相互作用界面は予測できていても、そのコンタクトまで予測できていないことが伺える[14]。また、Z_{IDDT}やZ_{TM}は複合体全体構造の主鎖に関しての予測精度の評価項目である。それらを総合した成績は図2の通りである。本研究室は、既知複合体構造(複合体テンプレート、複合体データベース)を利用した手法でこのカテゴリに参加していたが、3位の成績を収めることができた。



図 2. CASP14 Assembly category の成績.

https://predictioncenter.org/casp14/zscores_multimer.cgi より公開されているデータを基に作成 した。評価者による公式なグラフは Ref.14 の Fig.1 を参照されたい。

3. CAPRI とその評価法について

CAPRI (Critical Assessment of PRedicted Interactions) は 2001 年より始まった複合体立体構造を 予測する手法の発展を図る共同実験である。CASP14 の Assembly category は、CAPRI の Round 50 と共同で開催されていた。CASP の評価法はタンパク質複合体の全体構造の予測精度に着目 しているのに対し、CAPRI では相互作用界面をより重視している。CAPRI における評価法[16, 17]では、評価する相互作用界面で複合体を分離した際に、大きいユニットをレセプター、小さ いユニットをリガンドと定義し、相手側から 10 Å 以内にあるアミノ酸残基を界面残基としてい る。予測精度に関しては、3 つの指標 f(nat)、i_rms、L_rms を用いて4 段階に分類される。f(nat) は、実験構造の界面残基のコンタクトをどれだけ再現できているかを意味する。また、i_rms は 界面残基のみの RMSD を、L_rms はレセプターを重ね合わせた際のリガンドの RMSD を意味す る。これらの 3 つの指標は、1 つの相互作用界面に対して計算することができる指標のため、3 量体以上の複合体では、様々な相互作用界面に対して評価されることになる。4 段階の分類は、 モデルの予測精度が高い順に High、Medium、Acceptable、Incorrect とされる。複数の相互作用界 面を持つターゲットについては平均値によって判定し、Acceptable 以上のモデル構造を予測でき たターゲット数に応じて、式(5) のようにスコアが計算される[18]。

 $Score_G = \omega_1 N_{ACC} + \omega_2 N_{MED} + \omega_3 N_{HIGH}$ (5)

CASP14 と共催された CAPRI Round 50 においては、12 個のモデリングターゲットを CASP と 共有しており、評価された相互作用界面は 29 個であった。式(5)の重みは $\omega_1 = 1, \omega_2 = 2, \omega_3 = 3$ であり、成績としてまとめると、表1のようになる。CASP のように複合体を全体構造で捉えるにしても、CAPRI のように複合体を相互作用界面で捉えるにしても、CASP14の段階では、突出した構造予測手法がないことが分かる。

Rank	Group	相互作用界面毎の評価 (ACC/MED/HIGH の個数)	ターゲット毎の評価 (ACC/MED/HIGH の個数)	Score
1	Seok	9 / 4 / 0	5/4/0	13
	Venclovas, Baker	8/3/1	4/3/1	13
4	Zou, Chang	8 / 3 / 0	5/3/0	11
	MDOCKPP	7 / 2 / 1	4 / 2 / 1	11
7	Pierce, Kihara	7 / 3 / 0	4/3/0	10

表 1. CAPRI Round 50 の上位グループの成績.

4. まとめ

発展の目覚ましい分野ではあるが、一口にタンパク質立体構造予測手法と言っても、様々な切 り口、そして評価法が存在する。特に複合体立体構造予測は、上述したように考慮すべき点が多 く複雑である。この難解な問題に対し、CASP14 開催後であるが、AlphaFold2 は複合体立体構造 予測に対して対応可能になった[19]。AlphaFold2 の参入によって、今後どのように複合体構造予 測技術が発展していくかは注目に値するだろう。また、自身もこのような分野に貢献したいと考 えている。

参考文献

- [1] Jumper, J., Evans, R., Pritzel, A., et al. Highly accurate protein structure prediction with AlphaFold, *Nature*, **596**, 583-589 (2021).
- [2] Kryshtafovych, A., Schwede, T., Topf, M., et al. Critical assessment of methods of protein structure prediction (CASP)—Round XIV, *Proteins*, 89, 1607-1617 (2021).
- [3] Protein Structure Prediction Center, "CASP15 Experiment", https://predictioncenter.org/casp15/index.cgi (2022-08-21).
- [4] Moult, J., Fidelis, K., Kryshtafovych, A., et al. Critical assessment of methods of protein structure prediction (CASP)—Round XII, *Proteins*, **86**, 7-15 (2018).

- [5] Croll, T. I., Sammito, M. D., Kryshtafovych, A., et al. Evaluation of template-based modeling in CASP13. Proteins, 87, 1113-1127 (2019).
- [6] Pereira, J., Simpkin, A. J., Hartmann, M. D., et al. High-accuracy protein structure prediction in CASP14. *Proteins*, 89, 1687-1699 (2021).
- [7] Mariani, V., Biasini, M., Barbato, A., et al. IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests, *Bioinformatics*, **29**, 2722-2728 (2013).
- [8] Zemla, A., LGA: a method for finding 3D similarities in protein structures, *Nucleic Acids Res.*, **31**, 3370-3374 (2003).
- [9] Kryshtafovych, A., Monastyrskyy, B., Fidelis, K., CASP11 statistics and the prediction center evaluation system, *Proteins*, 84, 15-19 (2016).
- [10] Olechnovič, K., Kulberkytė, E., Venclovas, Č., CAD-score: A new contact area difference-based function for evaluation of protein structural models, *Proteins*, 81, 149-162 (2013).
- [11] Kryshtafovych, A., Monastyrskyy, B., Fidelis, K., CASP prediction center infrastructure and evaluation measures in CASP10 and CASP ROLL. *Proteins*, 82(82), 7-13 (2014).
- [12] Chen, V. B., Arendall, W. B. 3rd, Headd, J. J., et al. MolProbity: all-atom structure validation for macromolecular crystallography, *Acta Cryst.*, D66, 12-21 (2010).
- [13] Pereira J., Lamzin V. S., A distance geometry-based description and validation of protein main-chain conformation, *IUCrJ.*, 4, 657-670 (2017).
- [14] Ozden, B., Kryshtafovych, A., Karaca, E., Assessment of the CASP14 assembly predictions, *Proteins*, 89, 1787-1799 (2021).
- [15] Lafita, A., Bliven, S., Kryshtafovych, A., et al. Assessment of protein assembly prediction in CASP12. *Proteins*, 86, 247-256 (2018).
- [16] Lensink, M. F., Méndez, R., Wodak, S. J., Docking and scoring protein complexes: CAPRI 3rd Edition, *Proteins*, 69, 704-718 (2007).
- [17] Lensink, M. F., Wodak, S. J., Docking and scoring protein interactions: CAPRI 2009, Proteins, 78, 3073-3084 (2010).
- [18] Lensink, M. F., Brysbaert, G., Mauri, T., et al. Prediction of protein assemblies, the next frontier: The CASP14-CAPRI experiment, *Proteins*, 89, 1800-1823 (2021).
- [19] Evans, R., O'Neill, M., Pritzel, A., et al. Protein complex prediction with AlphaFold-Multimer, *bioRxiv*, doi: 2021.10.04.463034.

///// Activities /////

<開催報告>

構造活性フォーラム 2022

実行委員長 植沢 芳広 明治薬科大学医療分子解析学研究室

構造活性フォーラム 2022「In silico アプローチによる毒性予測研究および周辺分野の現状と 展望」を2022年6月3日(金)にオンライン開催いたしました。今回のフォーラムは、化学物 質の安全性を確保するために必須の課題である毒性の評価をテーマとしました。毒性物質を効 率的に識別する方法として、in silico 技術を用いた予測法の確立には大きな期待が寄せられてお ります。しかし、毒性は一般に複雑な発現メカニズムを伴うことから、予測が困難な対象として も知られています。一方で、予測モデルの説明性に対する要望は極めて強いこともこの分野の特 徴のひとつです。本フォーラムは、このような毒性関連の生理的なイベントを in silico 技術でひ もとき、解析し、予測する研究について意見交換できるような場をご提供できればと思い企画い たしました。

プログラムは以下のとおりです。

基調講演「毒性発現メカニズムに基づく一般化学品の毒性予測—AI-SHIPS プロジェクト—」 船津 公人(奈良先端科学技術大学院大学)

講演1 「Applicable Artificial Intelligence Method to Drug Metabolism and Pharmacokinetics -Comparison of Various Methods for Metabolic Active Sites-」 笹原 克則(Otsuka Pharmaceutical Development & Commercialization, Inc.)

講演2「大規模変異原性データを用いた第二回 Ames/QSAR 国際チャレンジプロジェクト」 古濱 彩子(国立医薬品食品衛生研究所)

講演3「AI創薬の基盤とデータ統合」 水口 賢司(大阪大学蛋白質研究所・医薬基盤健康栄養研究所)

講演4 「拡散方程式の ADMET 予測モデルへの適用」 日高中(武田薬品工業)

講演5「副作用研究における AI の可能性」 奥野 恭史 (京都大学大学院医学研究科)

はじめに、本フォーラムの基調講演をお引き受けくださいました船津先生から、経済産業省 AI-SHIPS (AI-based Substance Hazard Integrated Prediction System) プロジェクトについてご講演い ただきました。先生には説明性を重視した in silico 毒性予測戦略に基づく5年間にわたるプロジ ェクトの成果をご説明いただくとともに、最終的に構築された毒性予測システムについて詳細 に解説していただきました。

招待講演の第1席では、米国から参加していただいた笹原先生に、毒性発現において重要な代

謝物の in silico 予測に関してお話しいただきました。分子動力学法や量子化学計算を適用した高 精度 ADME 予測や活性代謝部位の可視化モデル等を用いることによる薬物代謝の多様な計算法 とその評価についてご紹介いただきました。

第2席では、古濱先生に国立医薬品食品衛生研究所主催の第二回 Ames/QSAR 国際チャレンジ プロジェクトについてご解説いただきました。当該プロジェクトの概要、結果、そして第一回 Ames/QSAR チャレンジとの詳細な比較検討を通して、Ames 試験の in silico 予測に関する現状と 今後の課題についてご説明いただきました。

第3席の水口先生からは、現在までに構築されてきた多様な薬物動態パラメーター予測モデ ルおよびデータ統合の話題とともに、データ駆動型と原理ベースのモデリングを組み合わせる 試み等についてご紹介いただきました。

第4席の日高先生からは、熱拡散方程式を用いた新規な QSAR 解析手法について解説していただきました。本法は生物活性・ADMET 特性の予測等の様々な場面で応用が期待されるとてもユニークなアプローチとなることを詳細にご説明いただきました。

第5席の奥野先生には、毒性研究の観点から創薬プロセスの各ステップにおける化合物デザインに対する取り組みや大規模な臨床データからの副作用要因の抽出等に関するAI技術の適用 例を解説していただきました。化学構造をグラフ構造として取り扱うGCNを用いたネットワーク構造解析等を用いた大変興味深い新規な知見の数々をご紹介いただきました。

以上のように、本フォーラムは毒性に関わる様々な課題の解決を目指した最新の in silico 研究 の集大成となった極めて高度な内容の講演会となりました。全ての演題に多くの活発なご質疑・ ご討論をいただき、盛会となりましたことをご報告させていただきます。毒性研究という広大な 領域の中で、分野の垣根を超えた多様な議論がなされたことは本フォーラムの大きな成果であ ったと感じております。

なお、本フォームは新型コロナウイルス感染症の蔓延を懸念して完全オンライン開催とさせ ていただきました。オンライン開催には参加が容易であるといった利点もありますが、対面での 会話のような情報交換の機会が制限されてしまいます。そこで今回は、フォーラム初の試みとし てオンライン上における情報交換会を開催いたしました。情報交換プラットフォームして oVice を導入し、実際の会話に近い環境の再現を試みました。多くの方々に参加していただき、こちら も盛況となりましたことをご報告いたします。

オンライン開催であるにもかかわらず、多くの方々のご協力によって大変活気のある充実し たフォーラムとなったことを心より感謝いたします。ご講演いただきました先生方、座長を務め ていただいた先生方、活発な議論を行っていただいた参加者の方々に厚く御礼申し上げます。フ ォーラムを成功に導いてくださった本フォーラム実行委員の粕谷敦先生(第一三共株式会社)、 頭金正博先生(名古屋市立大学)、吉成浩一先生(静岡県立大学)、古濱彩子先生(国立医薬品食 品衛生研究所)、浴本亨先生(横浜市立大学)、朝田瑞穂先生(明治薬科大学)のご助力、ご支援 に深く感謝いたします。また、お忙しい中ホームページを作成していただいた高木達也先生なら びに、ご支援をいただいた本間光貴先生、大田雅照先生、杉本学先生、幸瞳先生をはじめとする 日本薬学会構造活性相関部会常任世話人・常任幹事の先生方に感謝いたします。また、開催運営 のご援助をいただきました日本薬学会、協賛・後援いただきました日本化学会、CBI学会、日本 毒性学会に感謝いたします。

来年の構造活性フォーラム 2023 は、株式会社アグロデザイン・スタジオの西ヶ谷有輝先生が 実行委員長を担当されます。ぜひ多くの皆様が来年のフォーラムにご参加いただき、活発なご議 論の場となるよう、引き続きご支援のほど、よろしくお願い申し上げます。

///// Activities /////

<会告>

第50回構造活性相関シンポジウム

新型コロナウイルス感染症の蔓延が未だ収束しない中、研究者の皆様におかれましては、種々の制約のもとで研究活動に推進されていることと拝察いたします。

さて、1975年に初回が開催された構造活性相関シンポジウムは、今年度、第50回目の開催を 予定しております。昨年度と同様にオンライン開催ではありますが、特別講演、招待講演、一般 演題、ランチョンセミナーなど、充実した内容で開催予定です。また、節目の50回開催を記念 して部会長の本間光貴先生による記念講演も予定されています。

以下に概要をご案内いたします。現在、ホームページにて参加登録予約を受け付けております。 皆様のご参加を心よりお待ち申し上げます。

実行委員長 名城大学薬学部 小田 彰史

- 主催: 日本薬学会構造活性相関部会
- **会期:** 2022年11月10日(木)~11日(金)
- **会場:** オンライン開催
- **参加登録:** 受付中~ 2022 年 11 月 4 日(金)

人数上限(約250人)に達した場合、その時点で参加登録を終了します。

参加登録費: 日本薬学会会員、協賛学会会員、後援学会会員、学生の方(会員・非会員問わず) は無料。上記以外は 1,000 円。なお、協賛学会、後援学会の詳細につきましては、ホームページ (http://www.bio.info.hiroshima-cu.ac.jp/gsar2022/index.html)をご確認ください。

情報交換会参加費: 無料

第 50 回記念講演:

本間 光貴(日本薬学会構造活性相関部会長)(理化学研究所)

「シミュレーションと AI を組み合わせた医薬品設計定量予測の重要性と今後の展開」

特別講演:

立花 裕樹(塩野義製薬株式会社)

「新型コロナウイルス感染症治療薬 S-217622 の創製」

招待講演:

伊庭 斉志 (東京大学大学院情報理工学系研究科 教授)

「進化・創発と AI」

栗原 正明(湘南医療大学薬学部 教授)

「核内受容体に作用する薬物分子の創製」

高田 匠(京都大学複合原子力科学研究所 准教授)
「加齢に応じて進行するアミノ酸残基異性化と分離分析手法の現状」

その他、最新情報はホームページおよび公式 twitter にてご確認ください。

HP : <u>http://www.bio.info.hiroshima-cu.ac.jp/qsar2022/index.html</u>

Twitter : <u>@50th_SAR</u>

問い合わせ先:

〒468-8503 愛知県名古屋市天白区八事山 150
 第 50 回構造活性相関シンポジウム実行委員長
 名城大学薬学部 小田 彰史
 E-mail: SympSAR@ccml.meijo-u.ac.jp

部会役員人事			
2022 年度 常任世話人	2022/10/1 現在		
部会長	本間 光貴 (理化学研究所)		
副部会長	服部 一成(塩野義製薬(株))		
副部会長	竹田志鷹 真由子(北里大学 薬学部)		
会計幹事	川下 理日人(近畿大学 理工学部)		
庶務幹事	杉本 学(熊本大学大学院 先端科学研究部)		
広報幹事	加藤 博明(広島商船高等専門学校)		
SAR News 編集長	幸 瞳(理化学研究所)		
ホームページ委員長	高木 達也(大阪大学大学院 薬学研究科)		

構造活性相関部会の沿革と趣旨

1970年代の前半、医農薬を含む生理活性物質の活性発現の分子機構、立体構造・電子構造の計算や活性 データ処理に対するコンピュータの活用など、関連分野のめざましい発展にともなって、構造活性相関と 分子設計に対する新しい方法論が世界的に台頭してきた。このような情勢に呼応するとともに、研究者の 交流と情報交換、研究発表と方法論の普及の場を提供することを目的に設立されたのが本部会の前身の構 造活性相関懇話会である。1975年5月京都において第1回の「懇話会」(シンポジウム)が旗揚げされ、1980 年からは年1回の「構造活性相関シンポジウム」が関係諸学会の共催の下で定期的に開催されるようになっ た。

1993年より同シンポジウムは日本薬学会医薬化学部会の主催の下、関係学会の共催を得て行なわれることとなった。構造活性相関懇話会は1995年にその名称を同研究会に改め、シンポジウム開催の実務担当グループとしての役割を果すこととなった。2002年4月からは、日本薬学会の傘下組織の構造活性相関部会として再出発し、関連諸学会と密接な連携を保ちつつ、生理活性物質の構造活性相関に関する学術・研究の振興と推進に向けて活動している。現在それぞれ年1回のシンポジウムとフォーラムを開催するとともに、部会誌のSAR Newsを年2回発行し、関係領域の最新の情勢に関する啓蒙と広報活動を行っている。本部会の沿革と趣旨および最新の動向などの詳細に関してはホームページを参照頂きたい。(https://sar.pharm.or.jp/)

編集後記

日本薬学会構造活性相関部会誌 SAR News 第43 号をお届けいたします。今号では「タンパク質構造と計算化学」をテーマにしております。Perspective/Retrospective では大阪大学 蛋白質研究所の栗栖先生に PDB に蓄積されているデータや今後の展望についてご紹介いただきました。Cutting Edge では、国立感染症研究 所の黒田先生に抗体の立体構造モデリングの現状について、北里大学の清田先生、竹田-志鷹先生に CASP および CAPRI の評価方法について、それぞれわかりやすくご紹介いただきました。すべての先生が言及されていますとおり、AlphaFold2 の登場によってタンパク質のホモロジーモデルの利用がさらに身近になりました。一方、ホモロジーモデルだけではなく各種計算結果および PDB 登録構造について、付随するデータを理解しながら研究を進める重要性がますます増していると感じます。ご多忙の中、快くご執筆していただいた各先生に深く感謝申し上げます。6月に開催された構造活性フォーラムの報告および 11 月の構造活性相関シンポジウムの会告も掲載いたしましたので、お目通しいただければ幸いです。

(編集委員会)

SAR News No.43 2022 年 10 月 1 日

発行:日本薬学会 構造活性相関部会長 本間 光貴

SAR News 編集委員会

(委員長) 幸 瞳、河合 健太郎、清田 泰臣、合田 浩明、田上 宇乃、仲西 功

*本誌の全ての記事、図表等の無断複写・転載を禁じます。