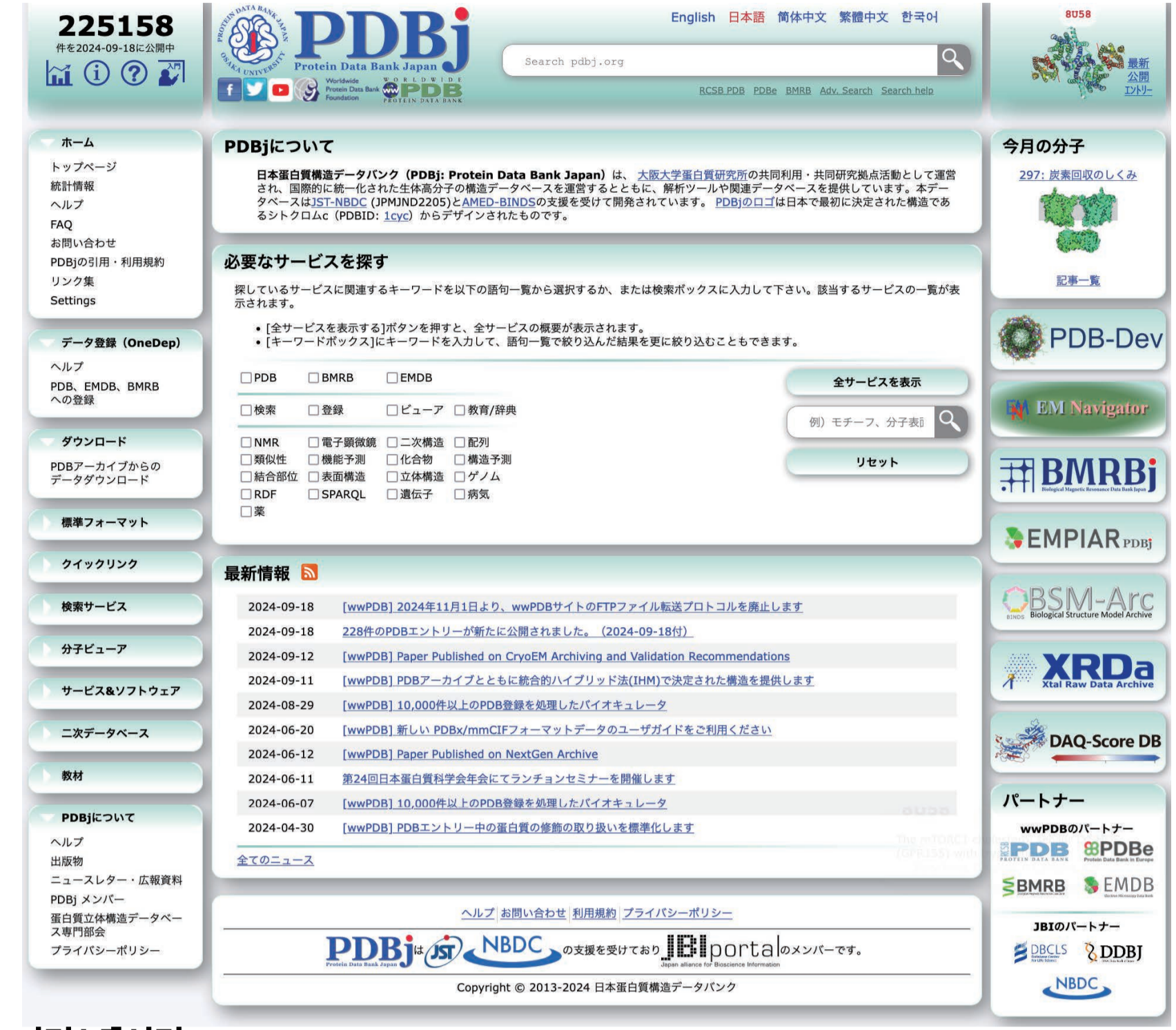


Poster  
 10

# 蛋白質構造データベースのデータ駆動型研究基盤への拡張

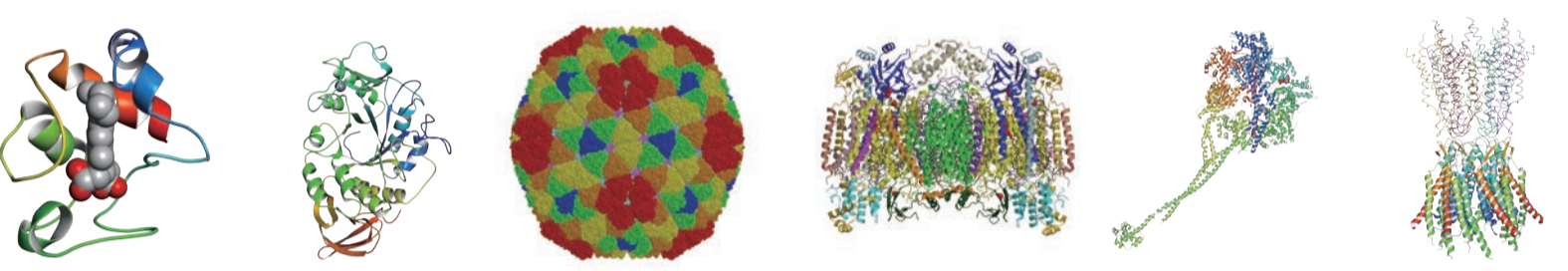
栗栖源嗣<sup>(1)</sup>、工藤高裕<sup>(2)</sup>、Gert-Jan Bekker<sup>(1)</sup>、山下鈴子<sup>(1)</sup>、横地政志<sup>(1)</sup>

(1) 大阪大学蛋白質研究所 (2) 一般財団法人 蛋白質研究奨励会

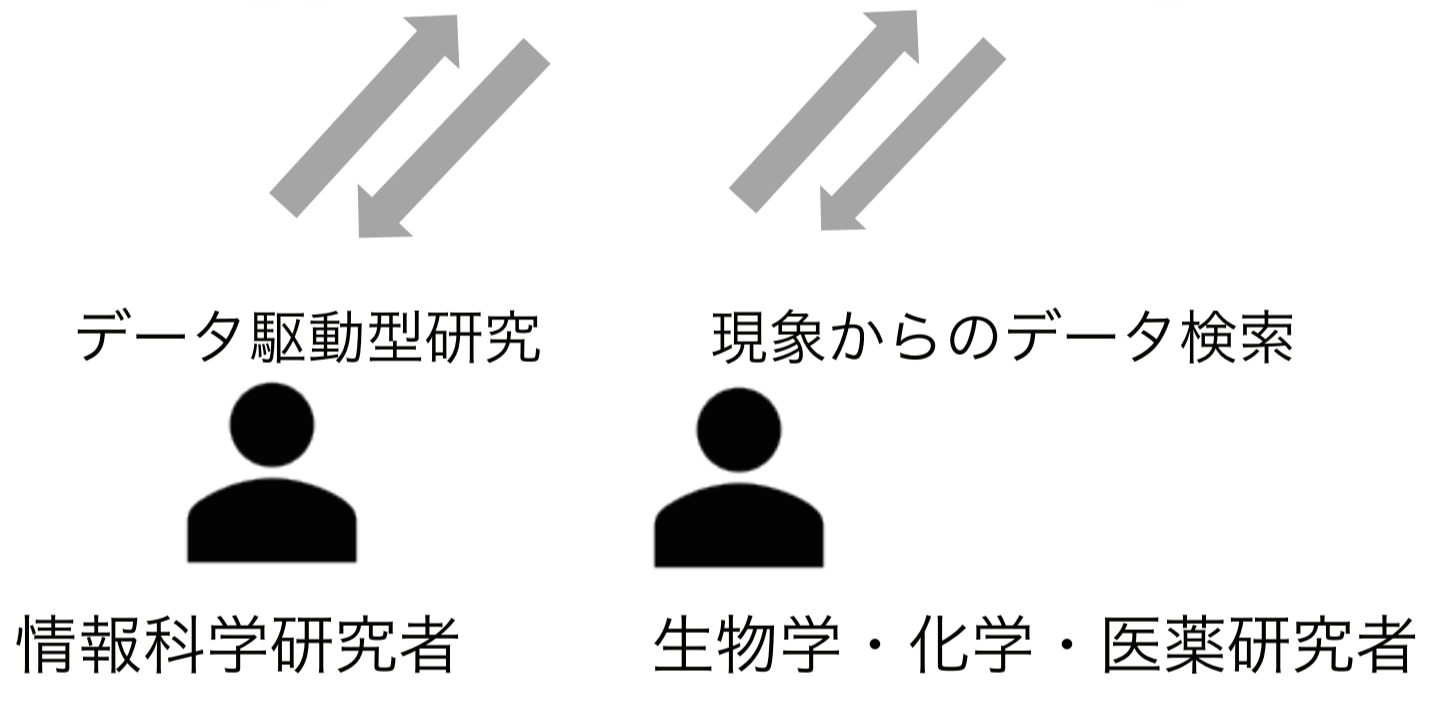


<https://pdbj.org/>

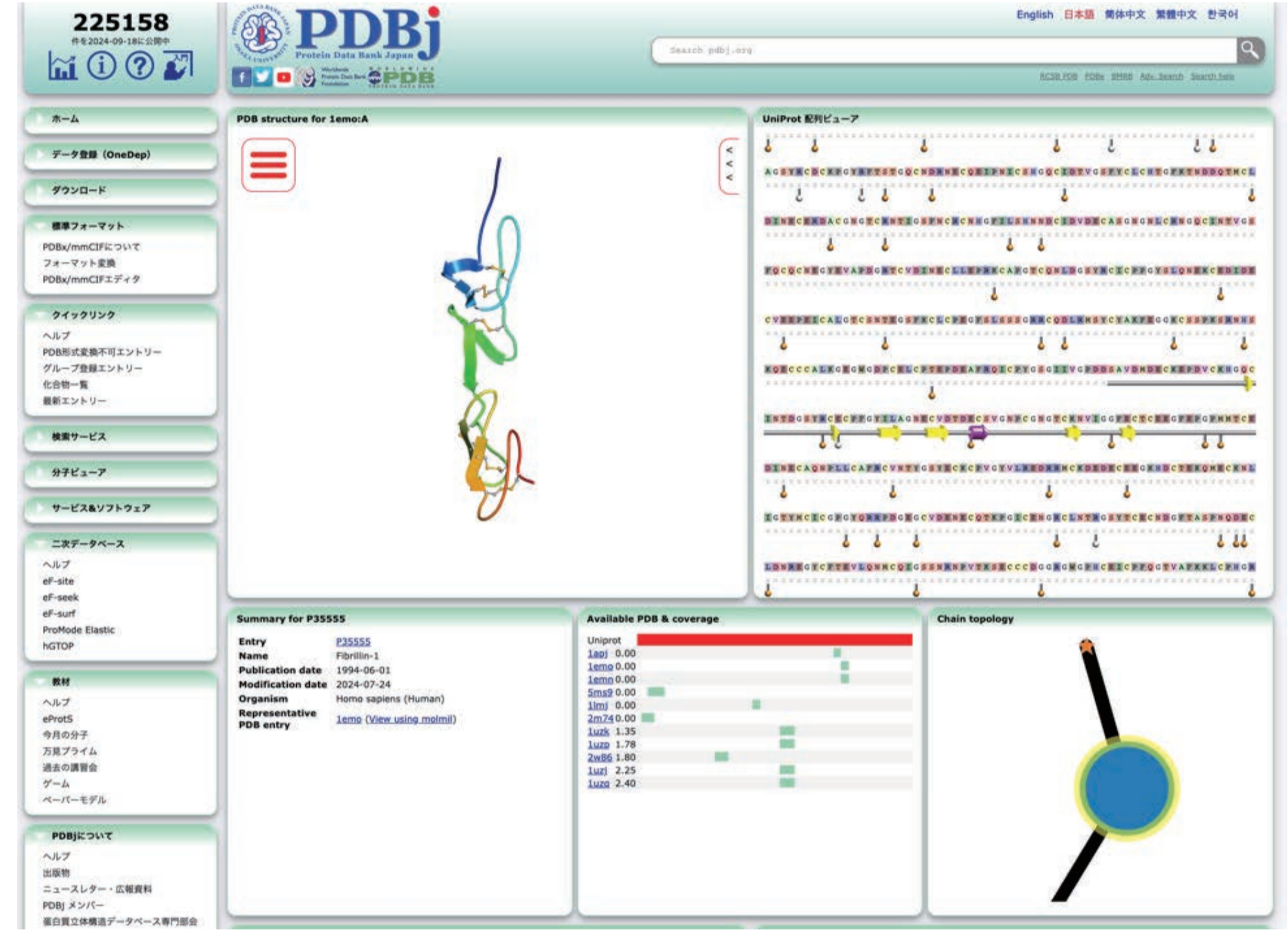
蛋白質構造データベース(Protein Data Bank:PDB)は蛋白質や核酸、糖鎖等の生体高分子の原子分解能の構造情報を集めた世界で唯一のデータアーカイブです。PDBの構造情報に基づいて、高次の生命機序までを合理的に説明しようとする構造生物学が、医学・薬学、農学等と強く結びつき、創薬等バイオ産業への応用によりインベーションに大きく貢献しています。本発表では、構造生物学の基盤データベースであるPDBとBMRB(NMR実験情報データベース)を、日米欧の国際協力により、安定的に構築・公開する活動の最近の話題を紹介するとともに、データ品質管理のための検証レポートをデータ利用にも拡張する計画について紹介します。また、PDBjにさまざまな統合利用ポータルを構築して機械学習用選抜データセットを公開する計画や、他のデータベースと統合的に利用できるデータベース利用サービスの開発状況についても紹介します。さらに、データベースへの登録と利用を促進するために継続的にPDBjで開発してきたサービスやツール類についても詳しくご紹介いたします。



## 利用者の動向調査に基づいた統合利用ポータルの制作



2022年度に日本蛋白質科学会、生命医薬情報学連合大会、日本生物物理学会、CBI学会、日本結晶学会、ライフインテリジェンスコンソーシアムなどで、利用者がこれまでどのようなデータベースと関連付けてPDBを利用してきたか、また今後したいと考えているかについて詳細な調査を実施し意見を収集した。この意見をもとに、ゲノム情報との統合利用の需要が大きいと判断し、まずは日本発のコホート調査のデータを基に作成されたjMorp(日本人多層オミクス参照パネル)との連携を進めた。東北大学のjMorpグループにDBCLSの片山招聘教授にも加わって頂いて、毎月定期的にオンライン会議を実施してjMorpのデータをPDBデータと統合利用するためには、まずUniprotベースのサマリページを用意し、どのPDBエントリを選ぶのが最適であるかサジェストする必要があると判断した。エントリ毎のRegion explorerやChain Topologyの表示などの新機能を開発し、PDBjへの実装準備を進めた。



<https://pdbjlc1.pdbj.org/next/uniprot/P35555/5ms9>

創薬候補化合物のターゲット蛋白質へのドッキングシミュレーションの精度を高めたり、補欠分子属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制度は目的によって異なってくる。化合物に特化した機械学習用のデータセットを選択する上で、目的別に実験データとの整合性をどの程度、どういった視点で検証し、フィルタリングするのが最適であるのか、利用者の声を聞くために、2022年度に日本蛋白質科学会、生命医薬情報学連合大会、日本生物物理学会、CBI学会、日本結晶学会で詳細な利用動向調査を実施した。目的ごとの機械学習用データセット選定のための基準作りを進めており、検証方法の確立しているX線結晶解析については概ね方法を固定した。

## NMR距離制限情報

PDBアーカイブの検証レポート(wwPDB validation reports)は NMRデータファイルは、多くの異なるソフトウェア固有のフォーマットでNMRデータを持つPDBエントリに対して、NEFおよび NMR-STARフォーマット [Ulrich, 2019]で標準化されています。この改善プロジェクトでは、PDBエントリのNMRデータを1つのNMR-STAR/NEFファイルに統合して提供します。wwPDBでは、NMRデータ(距離制限、化学シフト、場合によってはピークリスト)のNMR-STAR/NEF形式での単一ファイルのアップロードを進めており、将来的にはOneDepでのソフトウェア固有の形式でのアップロードを段階的に廃止する予定です。移行期間中、OneDepは特定の距離制限形式を登録時にNMR-STAR/NEFに変換し、化学シフトと合理的な制限を割り当てたNMRエントリのwwPDB検証レポートにおいて、以下のサポート形式でNMR距離制限の検証をユーザーに提供します:

AMBER, BIOSYM, CHARMM, CNS, CYANA, DY-NAMO/TALOS/PALES, GROMACS, ISD, ROSETTA, SYBYL, XPLOR-NIH

距離制限の複雑さとスペクトルピークリストの登録の有無により、初回の公開では、サポートされているフォーマットで単純な距離制限を持つNMRエントリだけが含まれ、スペクトルピークリストは含まれません。次の公開では、スペクトルピークリストを持つエントリが含まれる予定です。右の条件を満たしたエントリは、改善対象となります:

- ・割り当てられた化学シフトと距離制限の両方がPDBに登録されていること
- ・特定のフォーマット(AMBER, CHARMM, GROMACSなど)の距離制限ファイルを解釈するために、有効なトポロジーファイルまたは特定のコメントが存在する必要があります。
- ・NMRデータに記述されたすべての原子(割り当てられた化学シフトと制限)がモデルの原子と一致すること
- ・モデルと距離制限の間の配列アライメントが一致し、末端配列の拡張が可能であること

既存のNMRエントリのNEFおよびNMR-STAR形式の新しい統一NMRデータファイルは、単一のNMRデータファイルを伴うエントリと同様に、PDBアーカイブの「nmr\_data」ディレクトリ([https://files.wwpdb.org/pub/pdb/data/structures/divided/nmr\\_data/](https://files.wwpdb.org/pub/pdb/data/structures/divided/nmr_data/)) 及びBMRBアーカイブ([https://bmr.io/ftp/pub/bmr/nmr\\_pdb\\_integrated\\_data/coordinates\\_restraints\\_chemshifts/remediated\\_restraints/](https://bmr.io/ftp/pub/bmr/nmr_pdb_integrated_data/coordinates_restraints_chemshifts/remediated_restraints/)) で配布されます。

wwPDB検証レポートは、対象エントリのNMR距離制限の検証を行うために再計算されます。NMR距離制限は、構造計算の際に様々なソフトウェアによって異なる解釈や重み付けをされることがあります。NMR距離制限の検証ソフトは、NMR-VTFが推奨するより一般的なアプローチを採用し、すべてのモデルで各距離制限の満足度を独立して検証しています。このため、実際の構造決定ソフトウェアが異なるアプローチを採用したり、構造計算時の重み付けを少なくしたりすると、多数の違反が発生することがあります。

