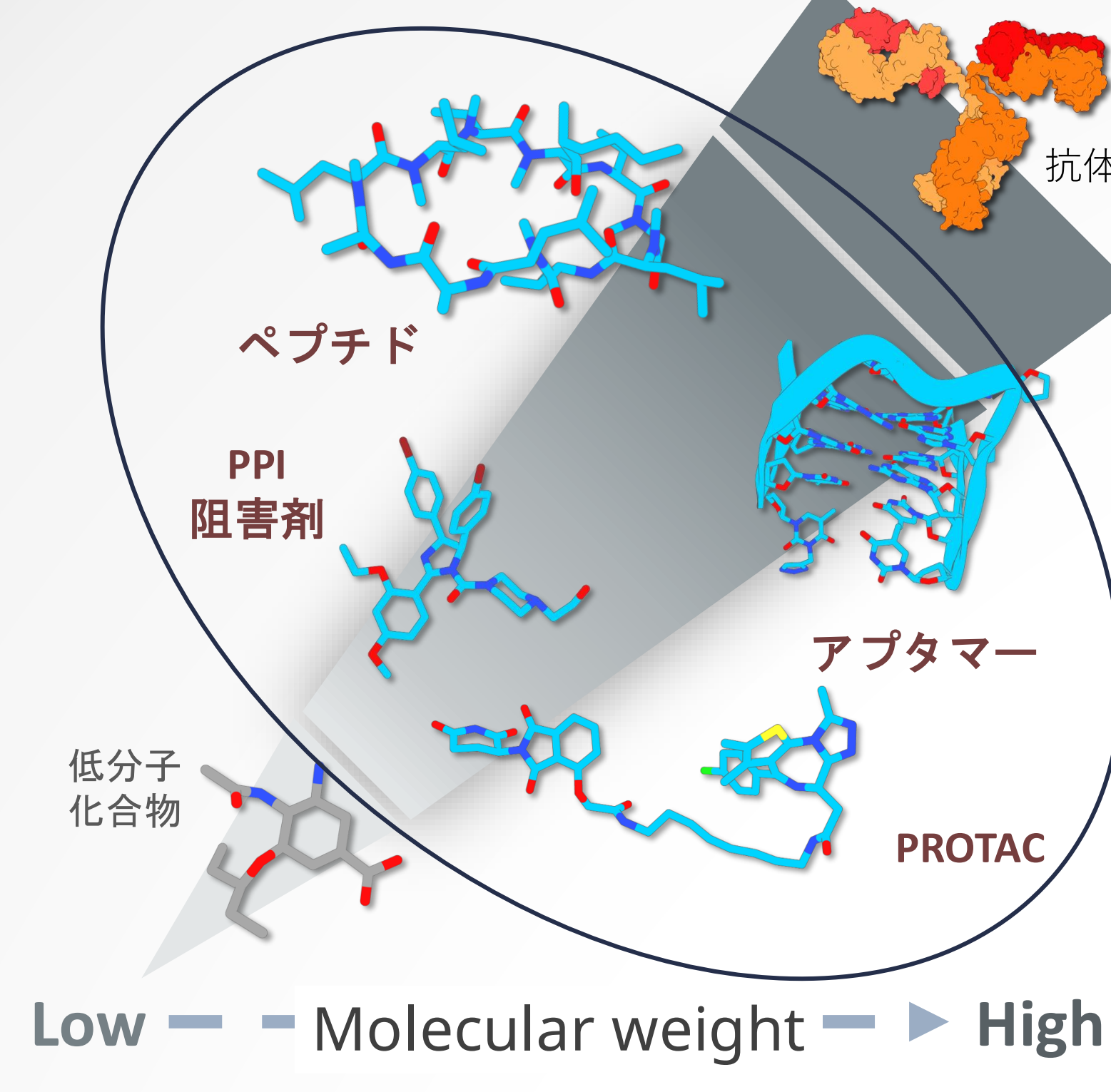


中分子創薬を支援する相互作用データベースMIIDB-AIの構築：立体構造注釈とAI予測の統合

池田和由（理研）、永江翼（産総研）、清水祐吾（理研）、富井健太郎（産総研）

Background



中分子（ペプチド、合成化合物・大環状化合物、核酸など）は、低分子医薬とバイオ医薬の中間を埋める重要な創薬モダリティとして近年注目されている。

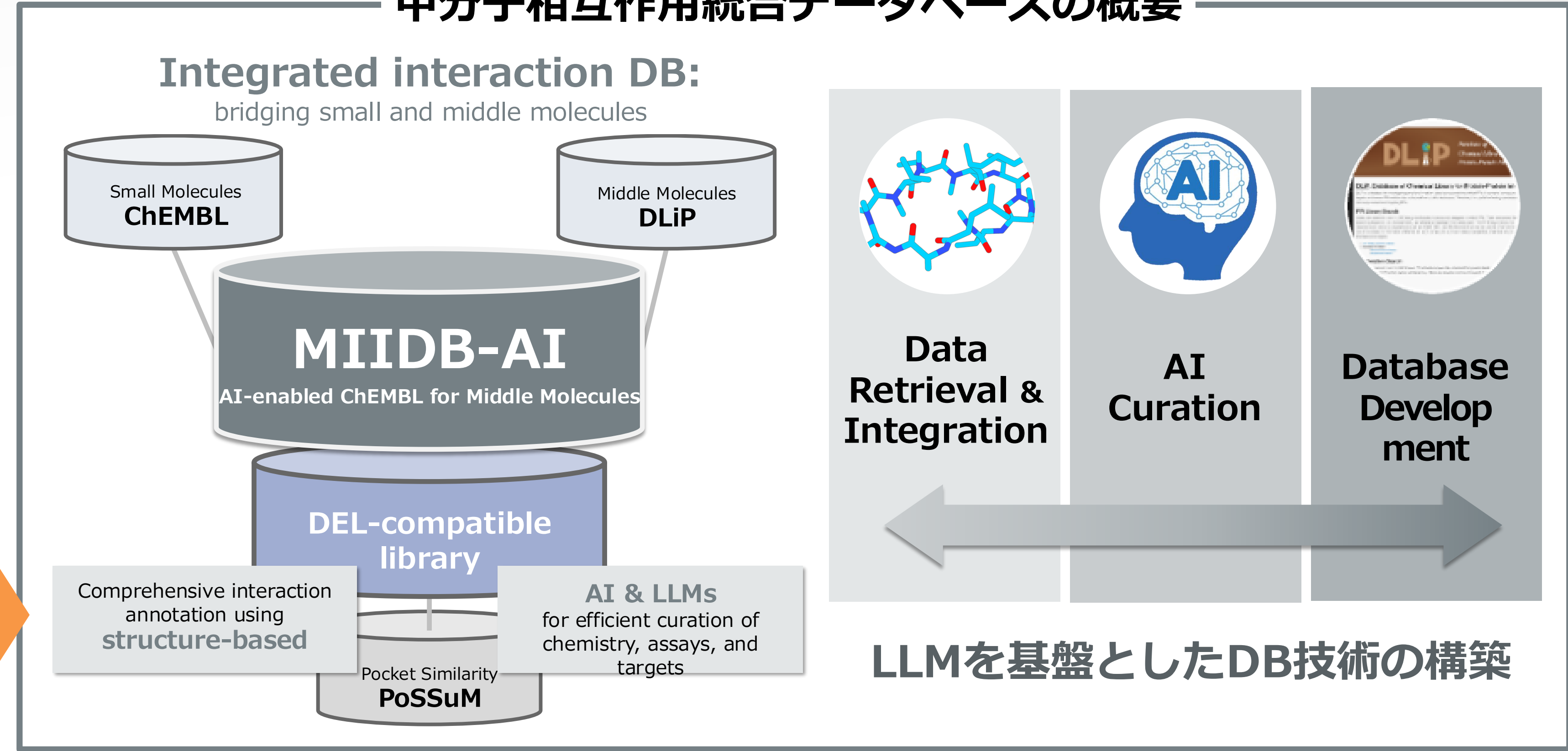
中分子は両者の利点を併せ持ち、比較的良好な物性と合成容易性を保ちながらタンパク質間相互作用（PPI）を標的にできることから、創薬における重要性が急速に高まっている。

一方、既存のデータベースは化学・生物・立体構造の各領域に分散しており、活性・構造・標的情報を統合的に解析することが困難である。

本研究では、中分子-標的相互作用に関する活性・立体構造・配列・標的情報を統合し検索可能としたプラットフォーム MIIDB-AI を構築した。

AIによる立体構造および親和性のアノテーションを組み込むことで、実験的に構造決定された複合体の範囲を超えて網羅性を拡張し、データ駆動型創薬を支援する。

中分子相互作用統合データベースの概要



Results & Developments

PDBからの中分子相互作用の抽出¹

Extract targets: middle molecule (asymmetric unit)* and its neighboring pocket residues**

*: Chain (_struct_asym) where 650 ≤ _entity.formula_weight ≤ 5000
**: Protein residues within 5.0 Å of the target molecule

non-polymer

- ex 1R1V: PPI inhibitors (Nutlin-2 – MDM2)
- 6B0Y: PROTACS (dBET6 PROTAC – CRBN – BRD4)

branched

- ex 2JCP: Oligosaccharides (Hyaluronan – CD44)
- 1E00: Oligosaccharides (Heparin – FGFR2 – FGF1)

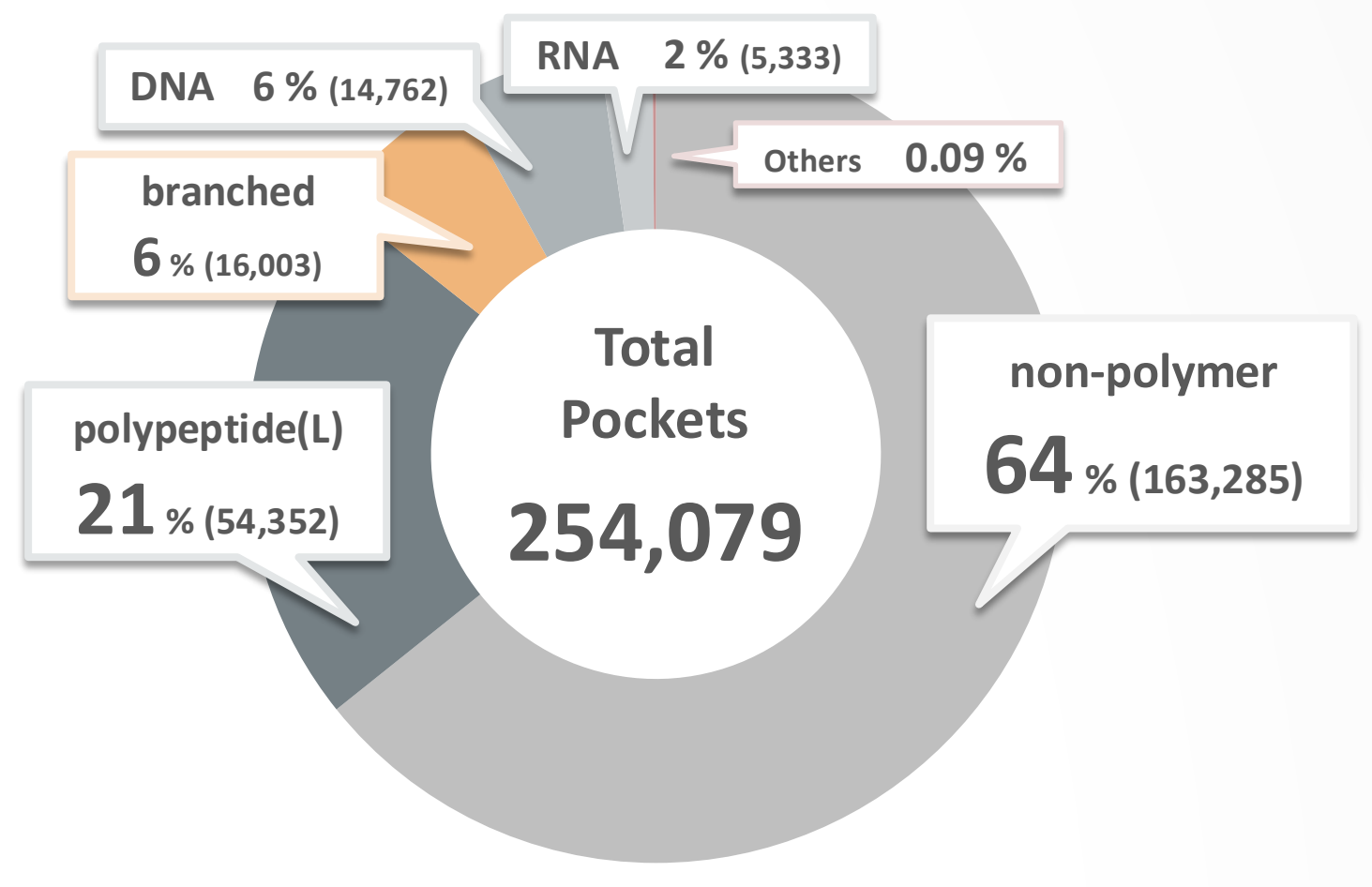
polymer

- ex 6Z8V: Aptamers (TBA-3L – Thrombin)
- 2Z6W: Peptides (Cyclosporin A – Cyclophilin D)

New Arrivals

Statistics

	Pocket	Unique PDB
non-polymer	163,285	19,616
branched	16,003	5,303
polymer	74,791	21,682
Total	254,079	46,420



Utilizing PDBx/mmCIF Format & PDB NextGen Archive²

Reconstructed & generate

What does “reconstructed” mean?

In pocket coordinates

Entity id	molecule
1	Cyclophilin D
2	Cyclosporin A

e.g.)_entity category

Category	Count
1 polymer 'PEPTIDYL-PROLYL CIS-TRANS ISOMERASE' 17783.322	1
2 polymer 'CYCLOSPORIN A' 1220.625	1
3 non-polymer 'CITRIC ACID' 192.124	1
4 water water 18.015	1

Derived categories provide information about the pocket

_ligand_info	_pocket_asym
asym_id: C	asym_id: A
entity_id: 1	auth_asym_id: A
entity_kind: polypeptide(L)	entity_id: 1
n_ideal_atoms: 85	entity_kind: polypeptide(L)
n_obs_atoms: 85	sifts_db_name: UNP
n_ideal_residues: 0	sifts_db_acc: P30405
n_obs_residues: 11	n_residues: 16
n_unobs_residues: 11	
has_covalent_to_pocket: N	
has_covalent_to_self: Y	
prd_id: PRD_000142	

Derived categories provide at-a-glance **summary and detailed** information for both pockets and ligands.

Leveraging PDB NextGen Archive

PDB NextGen Archive provides enriched annotations — including UniProt, SCOP2, and Pfam mappings — directly within the mmCIF.

Reconstructed from NextGen

- _pdbx_sifts_unp_segments
- _pdbx_sifts_xref_db
- _atom_site
- _pdbx_sifts_xref_db_name
- _pdbx_sifts_xref_db_acc
- _pdbx_sifts_xref_db_num
- _pdbx_sifts_xref_db_res

External DB Resources

- UniProt
- Pfam
- SCOP2
- CATH

This work leverages these annotations to incorporate **SIFTS cross-references** (e.g., UniProt accessions) into extracted pocket data without additional lookup.

MIIDB-AI 公開版とWebインターフェースの機能

MIIDB-AI β版から公開版へ

— 検索・詳細表示・3D可視化を備えたデータベース

URL: <https://miidb-ai.skb-insilico.com/>

トップページ：検索の入口

一覧表示：検索可能な中分子エントリ

詳細表示：活性データと立体構造相互作用

LLM

本データベースはLLMを用いた対話的アクセスに対応する予定である。

参考文献

1. wwPDB consortium. Nucleic Acids Res. 2019, 47, D520–D528. doi:10.1093/nar/gky949
2. Choudhary, P. et al. bioRxiv, 2023, 2023.10.24.563739. doi:10.1101/2023.10.24.563739
3. Passaro, S. et al. bioRxiv, 2025, 2025.06.14.659707. doi:10.1101/2025.06.14.659707
4. Shimizu, Y. et al. CBI Journal, 2026, 26, 11–27. doi:10.1273/cbij.26.11

Boltz-2³（AlphaFold3系）による立体構造・親和性予測結果

実験的に構造決定された複合体が存在しない中分子については、タンパク質-中分子複合体の立体構造モデルを生成し、Boltz-2による親和性スコアを立体構造に基づくアノテーションとして算出。妥当性を、公開データとの比較により標的横断的に評価した。

未学習の低・中分子と標的のペア → Boltz-2による複合体構造モデリング → 親和性スコアの算出 → 標的横断的な検証

独立ベンチマーク

- 356標的
- 10,933化合物
- 15,465化合物-標的のペア

予測性能

- MAE = 0.9
- 標的の21.2% : Pearson r > 0.6
- 標的の68.7% : MAE < 1.0

スケラビリティ

- NIMIにより実行時間を約60–90%短縮（1.9–7.6倍高速、NIMI-lightでは2.4–14.1倍）
- Shimizu, Y. et al. 2026⁴

汎化性

- 誤差と標的配列同一性・化合物の新規性との間に明確な相関は認められない

9CHX vs FK5 (magenta: X-ray, cyan: boltz2)

With potential option

MIW 804

pChEMBL value 9.7

Boltz-2 p(C50 (pred.)) 8.3

これらの結果は、既存の立体構造ベース予測手法と同等以上の性能を示した。

Conclusions

MIIDB-AIでは、中分子を対象として、立体構造に基づくアノテーションとAIによる構造・親和性予測情報を統合し、検索・詳細表示・3D可視化機能を備えたWeb DBを公開した。今後は、収載データを順次更新し、データベースのさらなる拡充を進める。

謝辞

本研究は、科学技術振興機構（JST）パイオサイエンスデータベースセンター（NBDC）（JPMJND2401）及び日本医療研究開発機構（AMED）「創薬等ライフサイエンス研究支援基盤事業」（JP26ama121028）の支援を受けて実施した。

