

CCDCから新たなワークフローの提案

～PROTAC Conformer Generator～

2026年3月11日

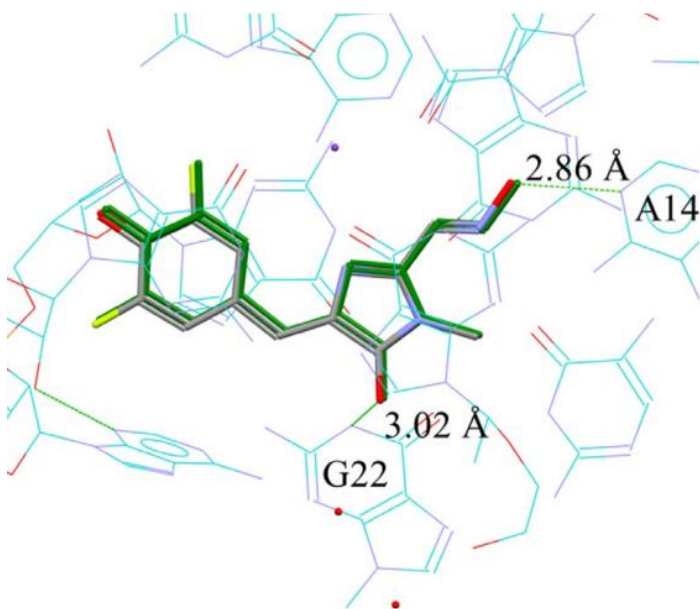
化学情報協会 坂本敏昭

低分子創薬の多様化

結晶構造情報を“次世代型”低分子創薬に活かす



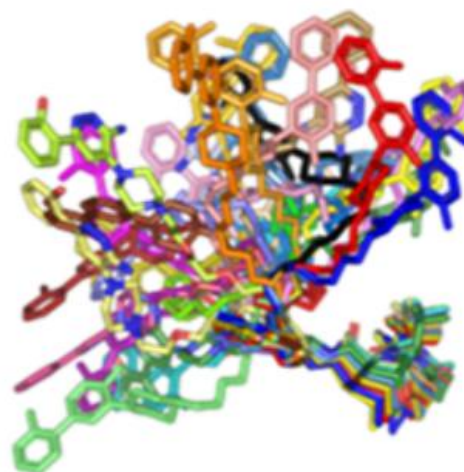
RNA – ligand docking



<https://www.ccdc.cam.ac.uk/nucleic-acids-gold/>



PROTAC Conformer Generator (PCG)



Key word

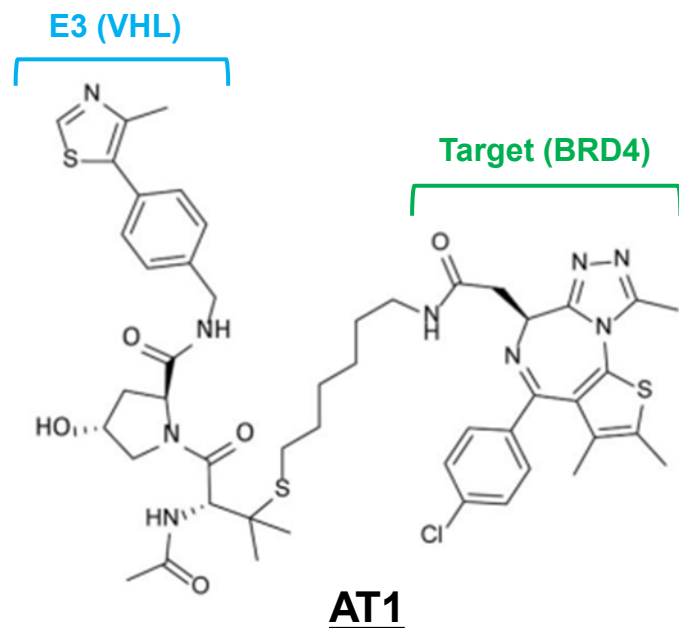
- ✓ Knowledge-based
- ✓ Conformer Ensemble
- ✓ Ternary Complex

J. Chem. Inf. Model. 2025, 65, 18, 9541–9556

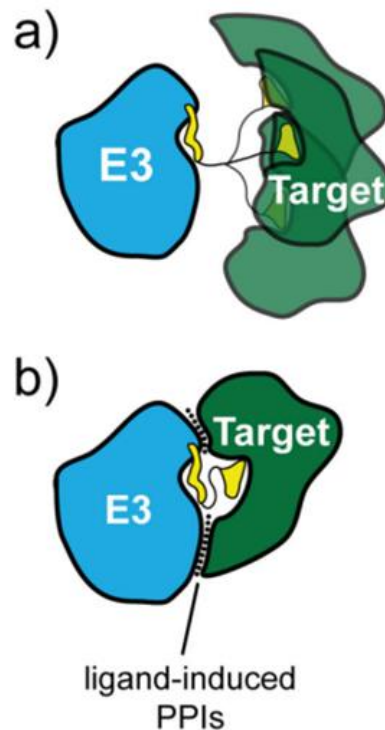
PROTAC (Proteolysis Targeting Chimera)

分解活性には、安定なTernary Complex (TC) の形成が鍵

➤ 標的タンパク質とE3を適切な位置に配置するリンカー設計が鍵を握る



Nat Chem Biol. 2017, 13, 514–521



配座予測, TCモデル取得

難易度：高

✓ Rotatable Bondが多い

✓ 2つのタンパク質が関与する

- 実験的手法 (X-ray, cryo-EM)
- 計算科学的手法
- 構造予測

CSD Conformer Generator

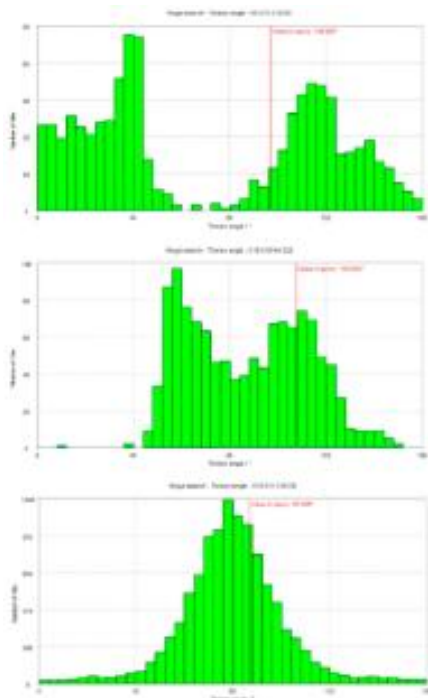
Mogulに蓄積されたデータに基づき、分子配座のアンサンブルを高速発生

➤ 計算化学では得られない“多様な”配座群を“高速で”生成する *Knowledge-based* プログラム

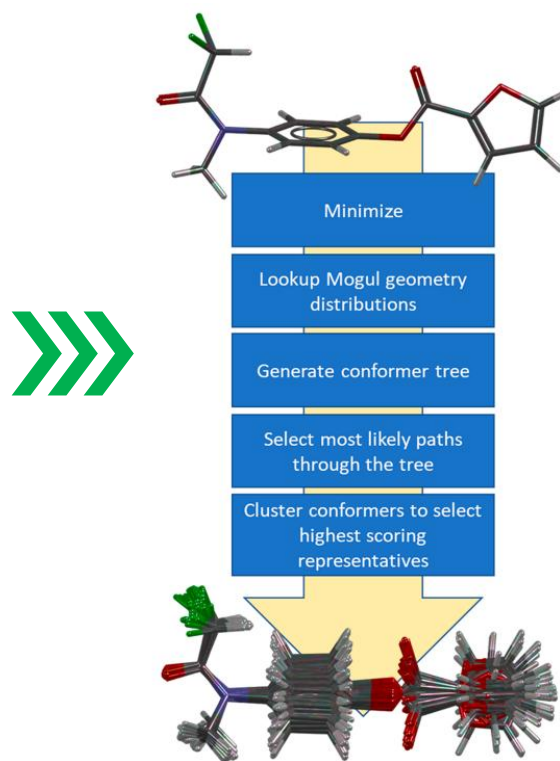


Mogul

分子ジオメトリライブラリ

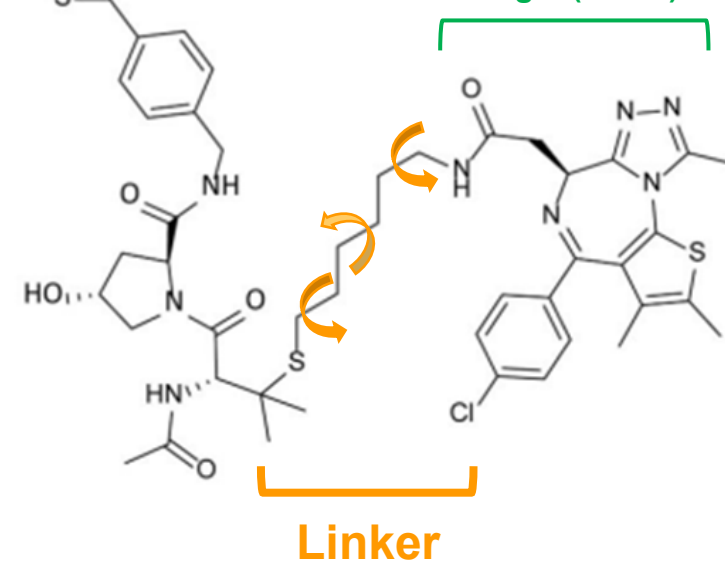


Conformer Generator



E3 (VHL)

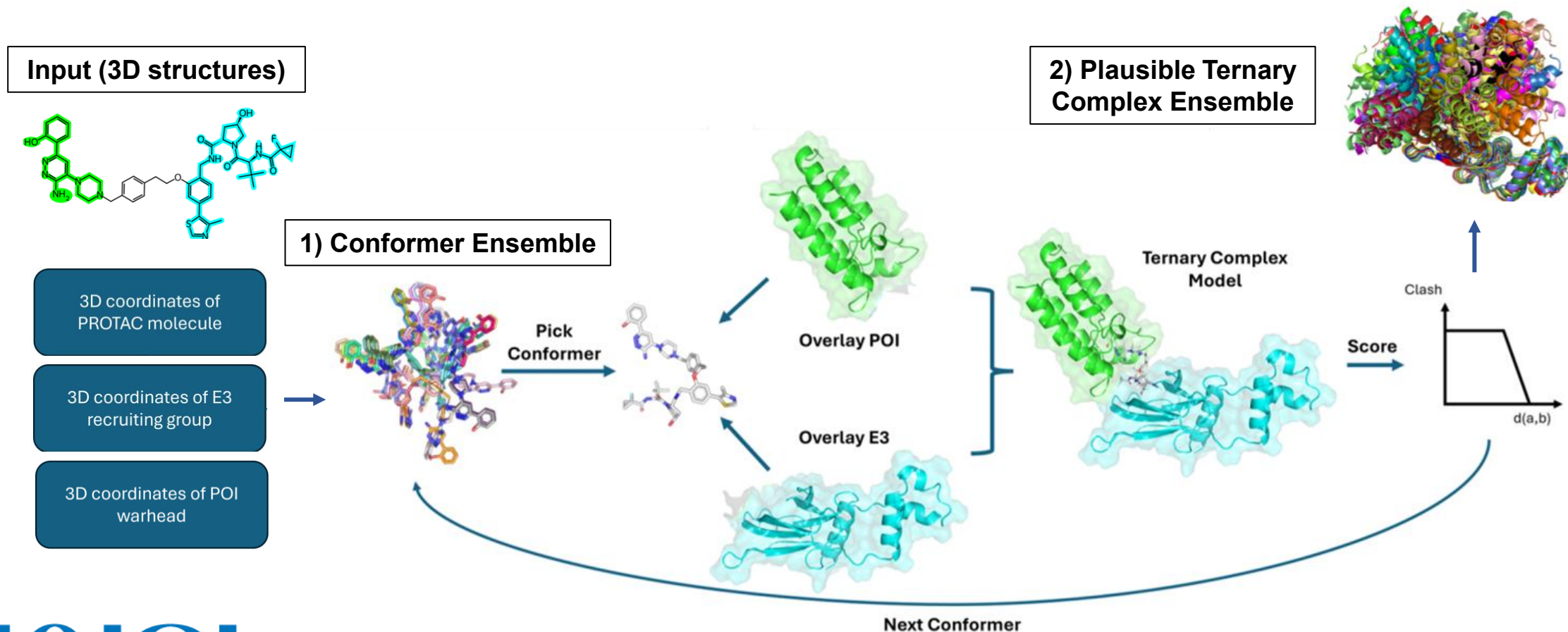
Target (BRD4)



PROTAC Conformer Generator (PCG) の概要

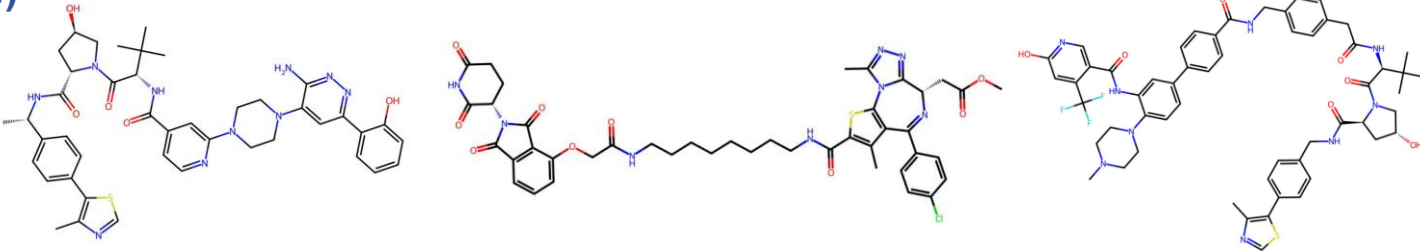
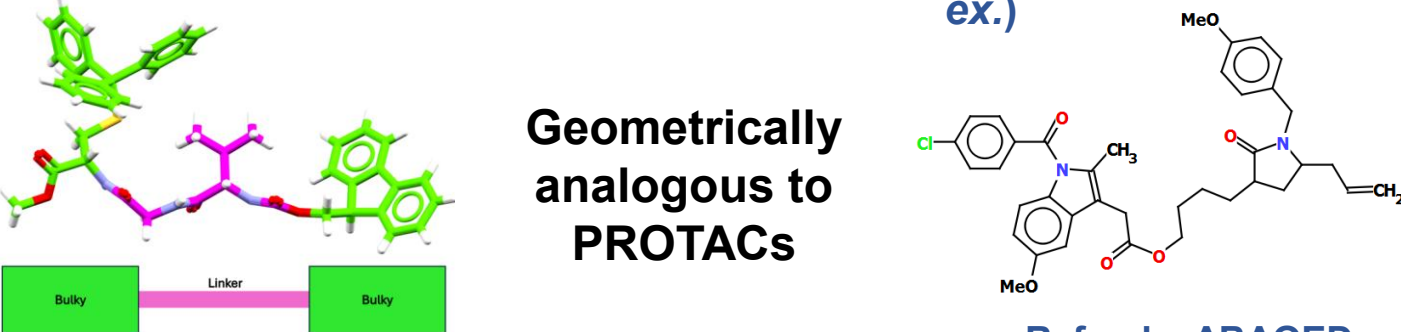
Python API scriptベースの自動化ワークフロー

➤ CCDCのGitHub repositoryからscriptを入手可能



Validationに用いたデータセット

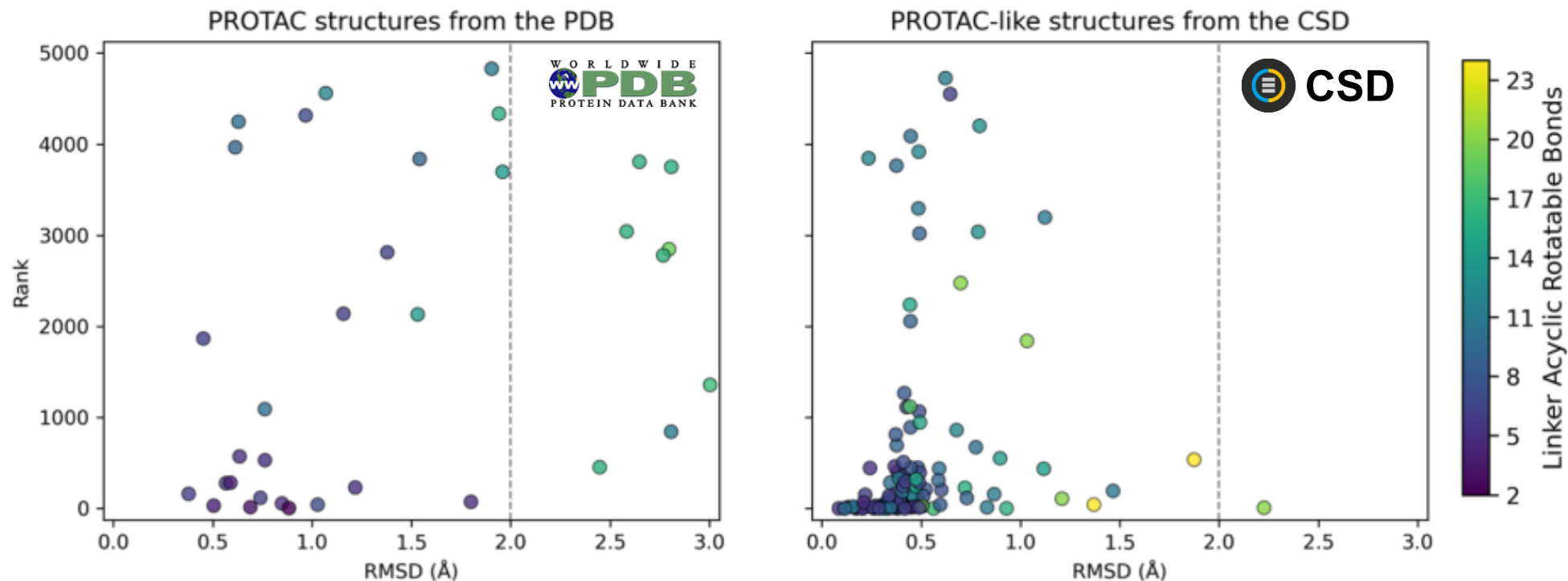
PDBおよびCSDからPROTAC (-like) 150分子を抽出

Set-1	PROTACs 34 cpds WORLDWIDE PDB PROTEIN DATA BANK	ex.)  <u>PDB code: 8G1Q</u> <u>PDB code: 6BN7</u> <u>PDB code: 8BB5</u>
Set-2	“PROTAC-like” molecules 116 cpds  CSD	ex.)  Geometrically analogous to PROTACs <u>Refcode: ABAQED</u>

Conformer生成の検証結果

実験構造を近似する能力あり / Set-2についてはRank付けも満足できる結果

➤ 5000個の立体配座を約6分で高速生成 [24 GB RAM and 16 Intel(R) Xeon(R) Gold 5218R physical cores]

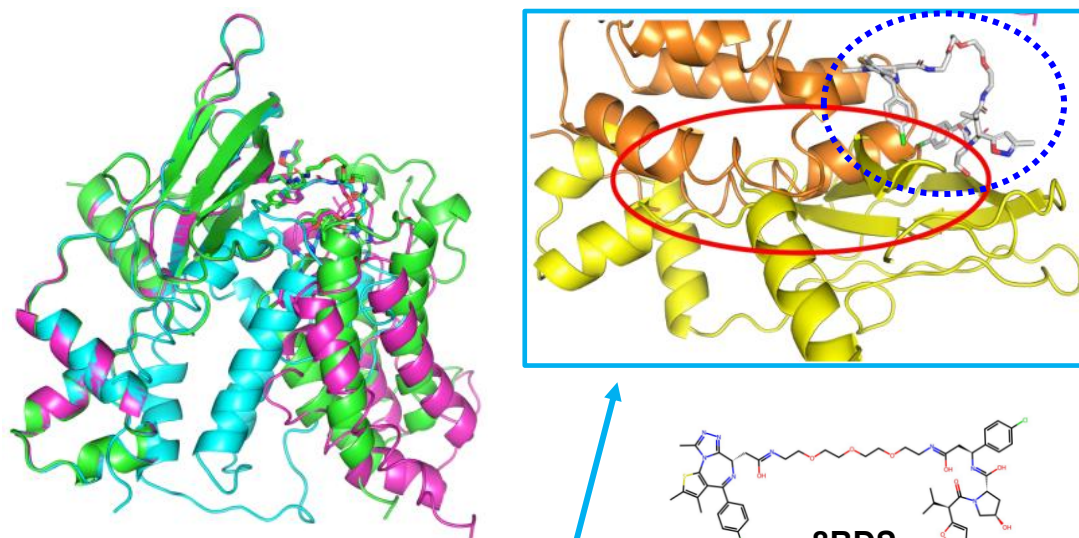


	RMSD $\leq 0.5\text{\AA}$	RMSD $\leq 1\text{\AA}$	RMSD $\leq 2\text{\AA}$	Top 10 Rank	Top 100 Rank	Top 1000 Rank
Set-1	6%	46%	76%	3%	18%	44%
Set-2	75%	93%	99%	17%	47%	84%

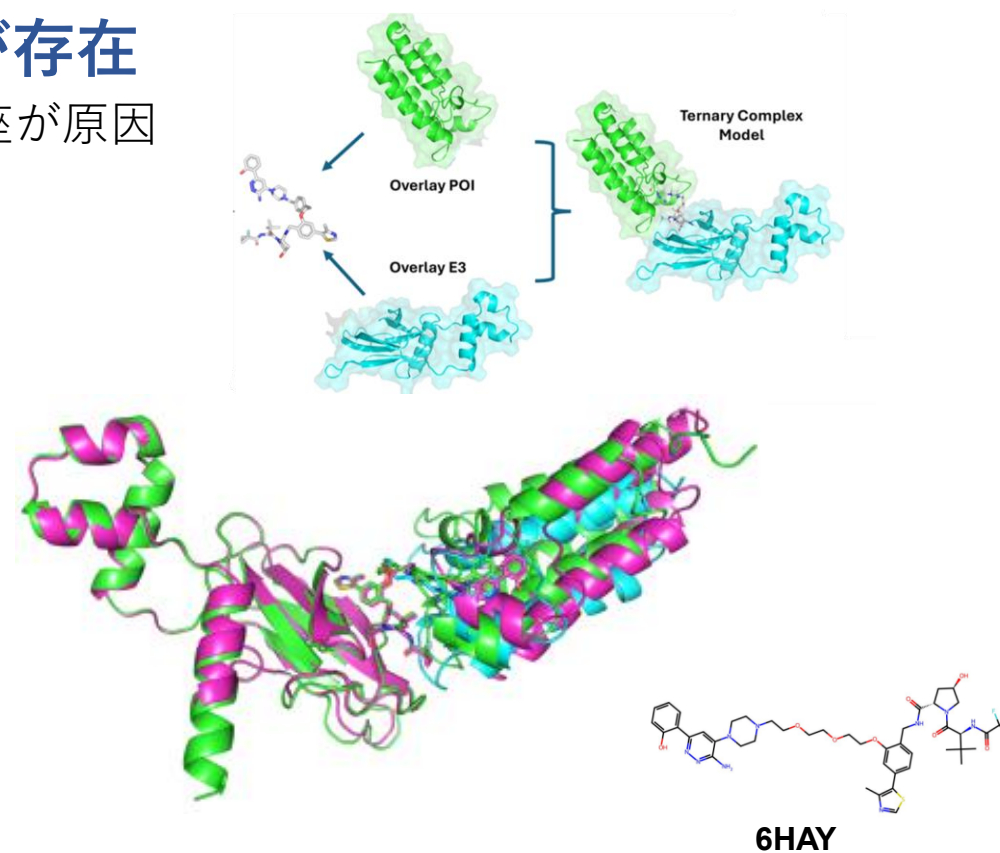
Ternary Complexモデル構築によるフィルタリング

立体障害が生じたり、PPIを欠くモデルが存在

- 過度にコンパクトな配座や過度に拡張された配座が原因
- 実際の三者複合体に酷似したモデルも存在



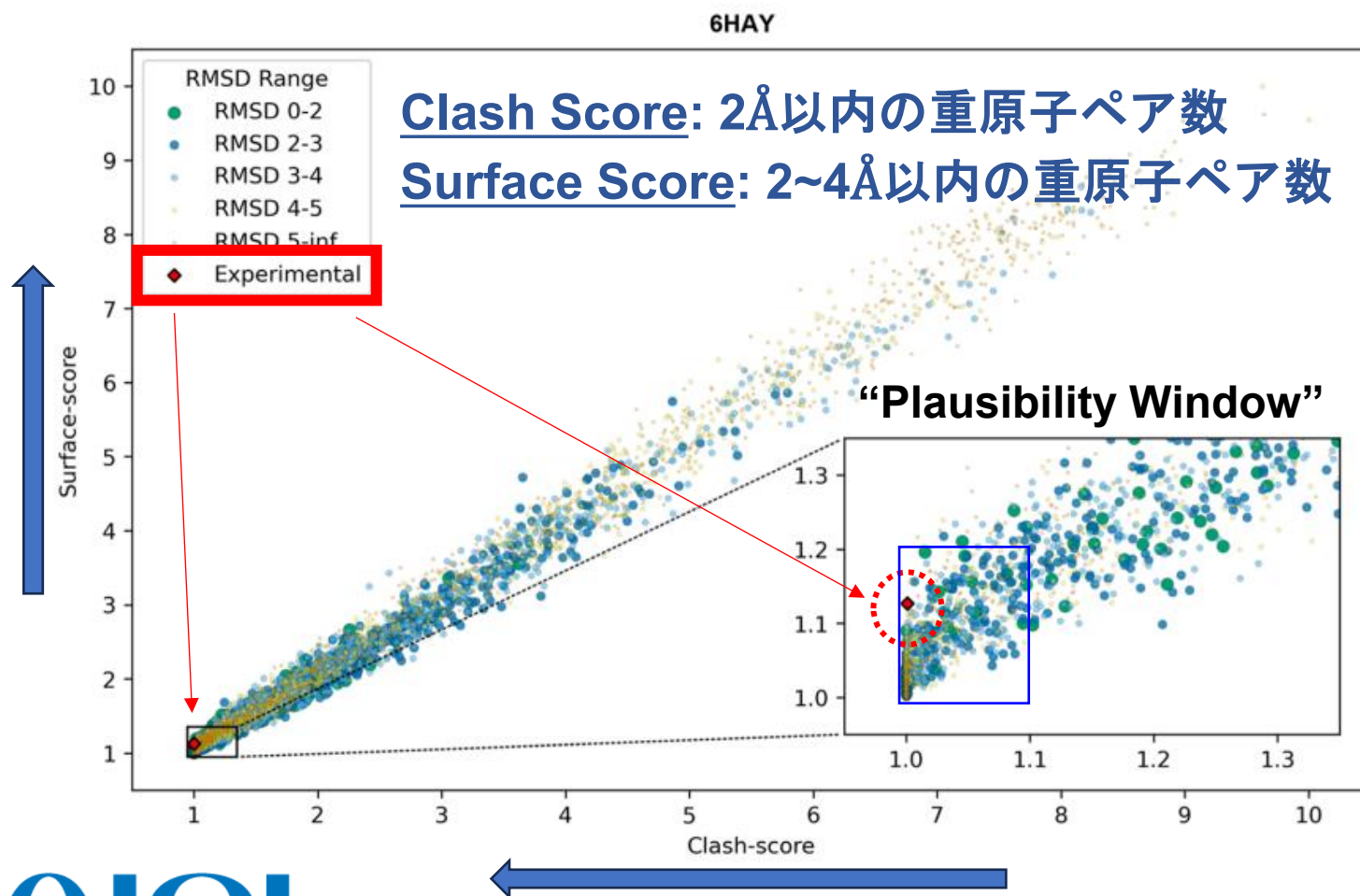
TC crystal structure (green)
Closest conformer (cyan, lowest lig_RMSD)
Best POI-E3 orientation (magenta, lowest pp_RMSD)



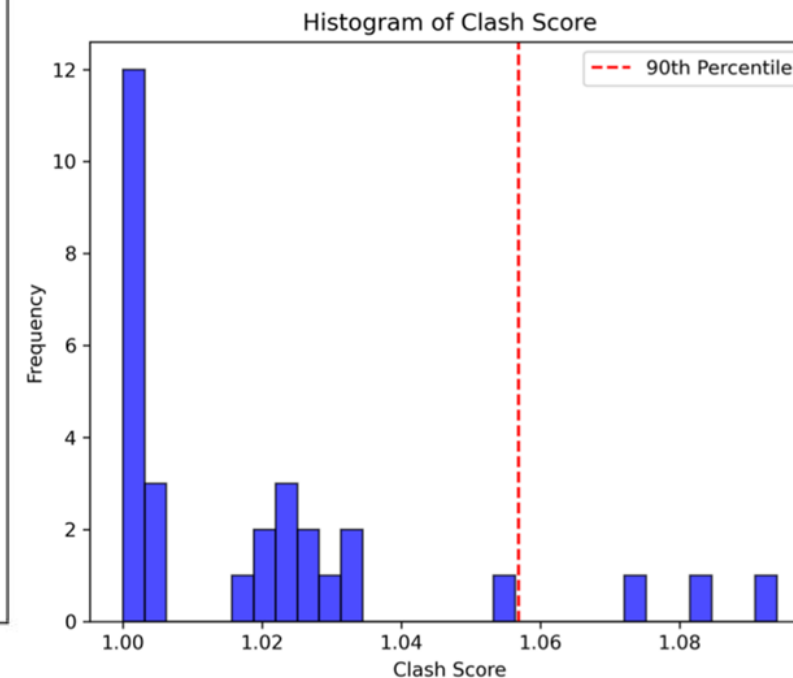
TC crystal structure (green)
Closest conformer (cyan, lowest lig_RMSD)
Best POI-E3 orientation (magenta, lowest pp_RMSD)

Clash Score / Surface Score

低Clash Score領域で、高Surface Scoreを示すものが理想的



Set-1の90%が Clash Score < 1.06



Ternary Complexモデルの評価

90%のケースで実験構造に近いTernary Complexを1つ以上生成

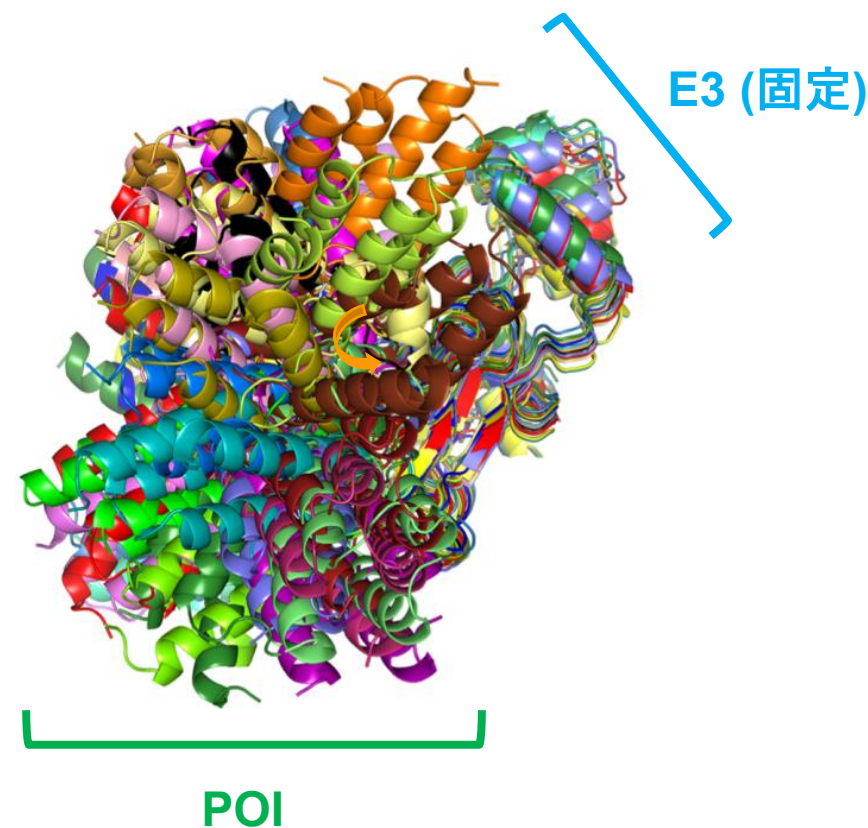
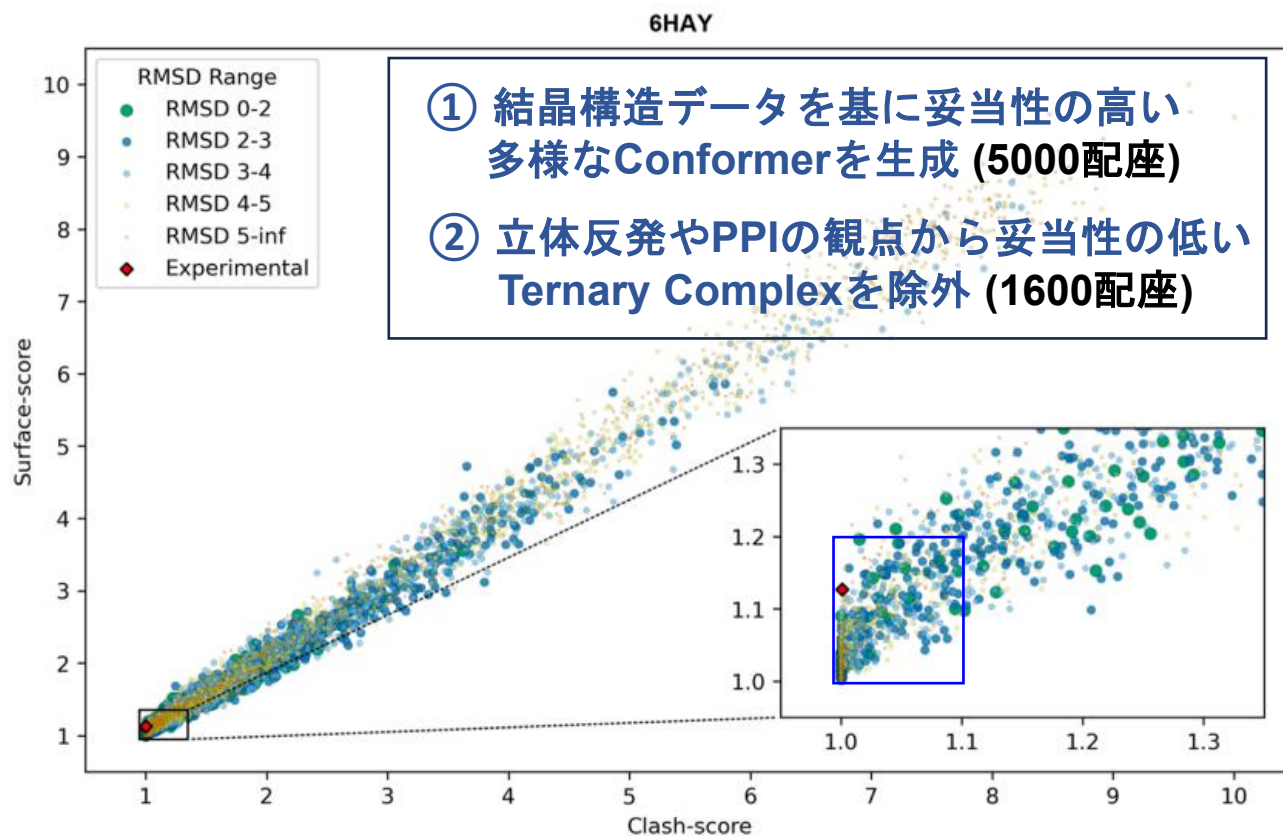
- ▶ サンプルング*した場合には、多くのケースで3割以上のモデルがpp_RMSD < 10Åを達成
- ✓ 高品質のモデルを生成する能力を持っている一方で、それらを特定することが依然として課題

PROTAC	pp_RMSD (Å)	rank	$n < 10 \text{ Å}$	n_{sampled}	PROTAC	pp_RMSD (Å)	rank	$n < 10 \text{ Å}$	n_{sampled}
5T35	3.36	2836	52	100	8BB4	9.57	313	2	100
6BN7	6.71	4528	6	21	8BB5	3.22	191	32	100
6BOY	10.97	2131	0	100	7KHH	9.41	1083	2	100
6HAX	2.57	279	75	100	7S4E	3.20	4267	35	100
6HAY	3.60	624	53	100	7Z6L	2.46	579	51	100
6HR2	3.04	120	51	100	7Z76	1.41	1092	72	100
6W7O	4.26	3021	17	100	7Z77	4.26	34	11	98
6W8I	13.96	151	0	100	7ZNT	2.75	2582	67	100
7JTO	7.72	4566	2	100	8BDS	5.07	4899	37	100
7JTP	5.16	68	3	17	8BDT	3.85	4483	77	100
7PI4	8.84	46	1	8	8BDX	4.25	3164	44	100
7Q2J	2.22	1415	39	100	8BEB	6.89	2113	5	100
8BB2	7.03	4422	1	100	8G1P	2.15	292	62	100
8BB3	9.81	2513	2	100	8G1Q	4.21	35	3	6
6ZHC	6.10	2827	2	100					

アンサンブルの整理

標的タンパク質とE3の相対配置で系統を分類 (リンカー形状の違いは考慮しない)

- 構造的多様性と妥当性を保持しながら、系統を分類する工程をフローに追加

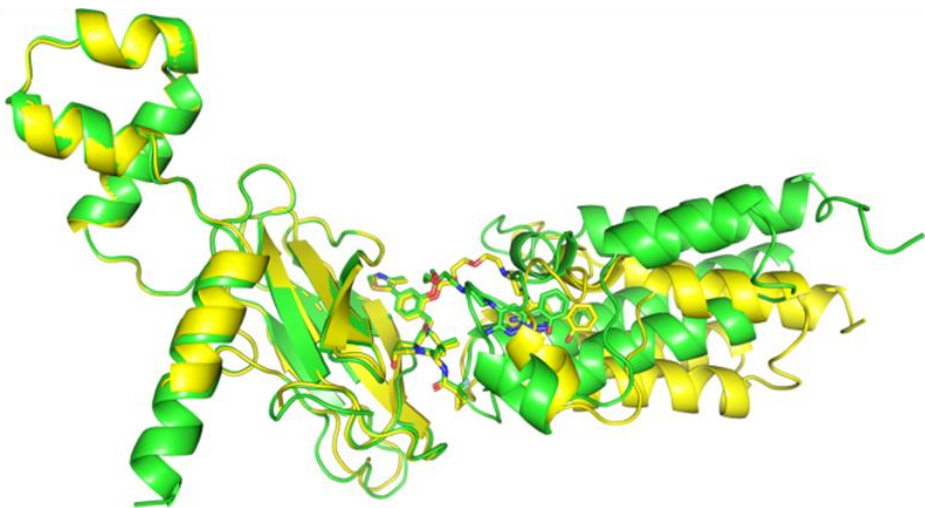


クラスタリングによるアンサンブル数の削減

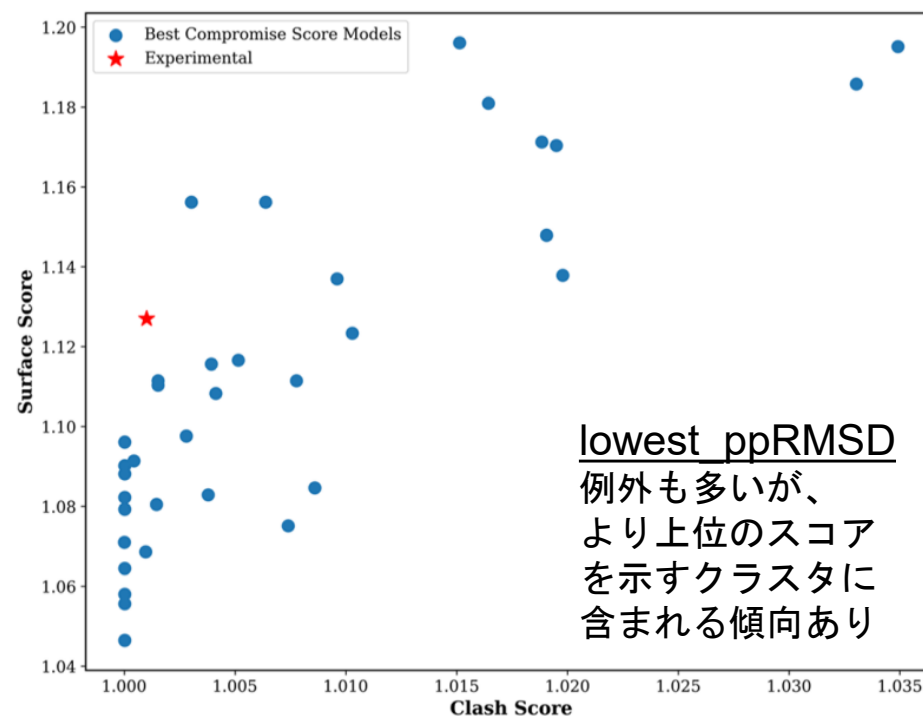
標的タンパク質とE3の相対配置を基に、35クラスターに分類

- 各クラスターの中で最もCrash / Surface Scoreが優れるモデルを代表モデルに選定

モデルの一つは、実際のTernary Complexと高い構造類似性を示す (pp_RMSD = 5.12Å)



各クラスターの代表モデルの散布図



PROTAC Conformer Generatorの位置付け

より精密なモデルを構築するための多様な初期モデルを迅速に提供するツール

➤ MDシミュレーションやタンパク質-タンパク質ドッキングワークフローを補完するツール

◆他のTernary Complexモデリング手法

✓タンパク質-タンパク質のドッキング、配座のサンプリング、スコアリングをさまざまな方法で統合
✓構造の最適化やエネルギーベースの評価に重点が置かれ高精度 / 計算コストが高い

- **PRosettaC**; *J. Chem. Inf. Model.* 2020, 60 (10), 4894–4903.
- **PROTAC-Model**; *J. Med. Chem.* 2021, 64 (21), 16271–16281.
- **PRODE**; *ACS Omega* 2024, 12611–12621.



補完的な役割

◆PROTAC Conformer Generator

✓CSDのKnowledge (結晶構造のねじれ角分布) の活用と剛体ドッキングにより、迅速にモデルを構築
✓リガンド側からのアプローチ

まとめ

■ Knowledge-based

- CSDに蓄積されたKnowledgeを基に、PROTACのConformer Ensembleと三者複合体モデルのEnsembleを高速で生成するツール
- Python API ベースの自動化ワークフローで、CCDCのGitHub repositoryからscriptを入手可能

■ Conformer Ensemble

- 活性コンフォメーションの予測には課題が残るが、PROTAC分子単独の配座予測には優れている
- 新規デザイン化合物の三者複合体形成能の予測やSARの解釈に用いることができる

■ Ternary Complex

- Clash ScoreとSurface Scoreを用いて、妥当性の低い構造をフィルタリング
- 多くのケースで実験構造に近いTernary Complex を生成する
- 最適なTernary Complexの特定には課題が残るものの、クラスタリングによりアンサンブルの削減が可能
- より精密なモデルを構築するための多様な初期モデルを迅速に提供するツールとして利用可能