Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

GluePlex: harnessing a co-folding model to predict molecular-glueinduced CRBN-neosubstrate ternary complex structures

Jin Wang

wang.j@bcm.edu

Baylor College of Medicine https://orcid.org/0000-0003-3625-7919

Hanfeng Lin

Baylor College of Medicine

Haiyang Zheng

Baylor College of Medicine

Article

Keywords:

Posted Date: November 11th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11580478/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: Yes there is potential Competing Interest. H.L. and J.W. conceptualized the study. H.L. and H.Z. conducted experiments. All authors contributed to the development or design of methodology. H.L. and H.Z. developed the code, supporting algorithms and curated the data and repositories. H.L. and H.Z. analyzed the data. H.L. drafted the manuscript. J.W. guided and reviewed the manuscript. All authors read and approved of the final manuscript.

GluePlex: harnessing a co-folding model to predict molecular-glue-induced CRBN-neosubstrate ternary complex structures

Hanfeng Lin[1,2], Haiyang Zheng[1,2], Jin Wang[1,2,3]

1. The Verna and Marrs McLean Department of Biochemistry and Molecular Pharmacology, Baylor College of Medicine, Houston, TX, USA
2. Center for NextGen Therapeutics and Department of Pharmacology and Chemical Biology, Baylor College of Medicine, Houston, TX, USA.
3. Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, TX, USA

Correspondence: hanfeng.lin@bcm.edu, wangj@bcm.edu

Abstract

Molecular glue degraders act through a single induced geometry, and predicting that geometry in silico is the step that would allow degrader design to become rational rather than empirical. However, current co-folding performance on structurally novel neosubstrates remains uncertain. Here, we assembled 52 cereblon (CRBN) target–compound pairs with published ternary structures and stratified them by training-set exposure. On the 13 novel pairs, plain Boltz-2 co-folding produced an acceptable pose in 12 of 39 independent runs, with a median DockQ (protein-interface accuracy score) of 0.01. GluePlex combines ligand-free, restrained protein–protein docking with two passes of template-guided Boltz-2 co-folding. It increased success to 23 of 39 runs, with a median DockQ of 0.43 and 19 rank-1 successes among them. Ablation analysis, which tests the contribution of pipeline components by removing or replacing them, attributed most of this gain to docking and template transfer. Despite this improvement, ligand accuracy remained limited: the CRBN-binding warhead had a median root-mean-square deviation (RMSD) of 0.7 Å, compared with 2.8 Å for the distal moiety, whose placement remained inaccurate even when the protein interface was correct. Exploratory ligand truncations rescued some complexes but worsened others. The parent model's target–ligand confidence helped identify cases likely to benefit, although these comparisons assume that truncation preserves the parent's ternary geometry. Changing the docked conformer affected individual compounds without improving overall success, supporting the use of multiple conformers. Across configurations and truncations, an acceptable pose was recovered for 12 of 13 novel families while the remaining 1 can be recovered by plain co-folding, but no single configuration solved them all. In general, GluePlex improves prediction of novel CRBN ternary complexes, while model ranking and distal ligand placement remain key limitations.

Introduction

Molecular glue degraders are a class of targeted protein degradation agents [1,2]. Unlike bivalent PROTACs, they recruit neosubstrates without a second target-binding pharmacophore, instead modifying an E3 ligase surface to favor a new protein interaction [3]. Cereblon (CRBN) is the best-characterized and most actively pursued ligase for molecular glue discovery. Thalidomide and its clinically used analogues demonstrate that low-molecular-weight compounds can induce therapeutically useful protein degradation. These compounds bind the CRBN CULT domain [4] and recruit neosubstrates like IKZF1 and IKZF3 [5,6], casein kinase 1α [7], GSPT1 [8] and other C2H2 zinc-finger proteins [9–11]. These shallow interfaces are formed largely by the two proteins but the presence of the glue is sufficient and necessary for stable ternary complex formation, making recruitment sensitive to ligand chemistry and cooperative interactions [12–14].

Molecular glue discovery remains largely empirical. Approaches include deconvolution of phenotypic hits [15,16], chemical profiling in ligase-deficient cell lines [17,18] and high-throughput proteomics [19–21]. In each case, the ternary complex structure is usually determined after a potent compound has been identified through medicinal chemistry campaign. Predicting this ternary complex structure with weak hits before the medicinal chemistry campaign could significantly facilitate rational design of molecular glues [22].

Co-folding offers a direct approach to ternary complex structure prediction. Building on protein structure prediction methods [23], AlphaFold 3 [24] and other AF3-like co-folding models (Boltz-2 [25,26], Chai-1 [27], and Protenix [28]) jointly model proteins and ligands from sequences and a SMILES string encoding the ligand's chemical structure. MGBench reports 50.6% success on the protein–protein interface and 32.9% recovery of the glue's own contacts for the best method across 88 complexes, but on the degrader subset those rates fall to roughly 20% and 15%, respectively [29]. It concludes that current co-folding methods do not meet the precision required for degrader complexes [29]. Dedicated glue models have reported higher rates [30]. However, training-set overlap complicates these estimates, since many CRBN ternary structures containing C2H2 zinc-finger degrons predate current model training cutoffs. Date and sequence filters also miss structurally conserved interfaces: WIZ belongs to MGBench's low-homology split but retains the C2H2 G-loop motif represented in several training structures and is modeled well across methods [17,29]. Performance on such targets may therefore overestimate accuracy for neosubstrates without a known ternary homologue. Similar limitations have been reported for ligand pose prediction on novel sequences [31]. Recently, a number of non-G-loop CRBN neosubstrates were discovered, greatly expanding the scope of targets accessible to CRBN molecular glues [20,21,32–35]. However, due to the lack of co-evolution signal between CRBN and chemical-induced neosubstrates, the co-folding performance on these pairs is expected to be challenging, but so far has not been systematically studied.

To address this gap, we assembled a benchmark from the extensive collection of experimentally determined CRBN ternary structures, selecting one representative per target–compound pair and excluding PROTACs to obtain 52 pairs. We classified 28 pairs as SEEN (direct), 11 as SEEN (homologue) with ≥40% sequence identity over ≥50% coverage, and 13 as novel. Both SEEN strata consist entirely of C2H2 G-loop degrons and show near-saturated performance. On the novel pairs, plain Boltz-2 co-folding (providing only sequences, SMILES, and a multiple sequence alignment (MSA)) produced an acceptable pose in only 12 of 39 independent runs, with a median DockQ of 0.01, and all successes involved G-loop degrons [20,36,37]. GluePlex is a workflow harnessing the same Boltz-2 co-folding model with structural templates generated

by PeSTo and HADDOCK3 supplied as structural guidance for co-folding. PeSTo [38], a geometric deep learning model that scores protein-protein interaction (PPI) hotspots from structures alone, provides prior information on the target surface. HADDOCK3 [39,40] docks CRBN against the target under the PeSTo prior with no ligand present. Every docked cluster representative is then supplied to Boltz-2 as a structural template, and the first-pass structures that score highest on chain-pair ipTM (a model-derived confidence score for the relative placement of two chains) serve as templates for a second, refining pass. On the same novel pairs this reaches CAPRI-acceptable quality in 23 of 39 runs (19 of them rank-1 successes), with a median DockQ of 0.43.

GluePlex was developed during a prospective VAV1 degrader campaign, when no experimental CRBN–VAV1 ternary structure was publicly available. Plain Boltz-2 co-folding placed VAV1 away from the CRBN recruitment surface. Adding interface restraints and docking templates produced models that supported the discovery of NGT-201-12, which was validated experimentally [21]. The NGT-201-12-induced ternary structure has not been determined, so that compound is excluded from this benchmark. Here, we evaluate how GluePlex workflow generalizes for CRBN-based glues beyond the VAV1 campaign.

We observed that the small, three-ring NGT-201-12 was modeled more robustly than the larger VAV1 ligand that was subsequently deposited in PDB, suggesting that ligand structure influences prediction accuracy. To test this, we generated 29 in-silico truncations of benchmark glues and scored them against their parent structures. These smaller compounds also approximate the chemical complexity of early screening hits. Because no experimental structures are available for the truncated analogues, the analysis assumes that truncation preserves the parent's ternary complex geometry. We therefore report these results separately from the main benchmark.

We also used ablation experiments to assess the contributions of interface restraints and docking templates, compared target conformers (alternative structures of the same target) under a fixed interface prior, and evaluated ligand accuracy separately for the CRBN-binding warhead and the distal moiety. These analyses identify which pipeline choices improve prediction and where errors remain.

Results

GluePlex pipeline uses physics-based prior information as co-folding templates

The default GluePlex workflow (**Fig. 1**) takes a target sequence and a glue SMILES string. PeSTo scores AlphaFold models of the target domain and CRBN, and solvent-accessible residues above PeSTo interface propensity of 0.3 define HADDOCK3 restraints. On CRBN, excluding the constitutive DDB1-binding interface leaves three PeSTo-filtered residues on the glue-binding face (F150, E377 and W386) used for every target. Using these restraints, HADDOCK3 docks the two proteins without the ligand. Each cluster representative then serves as a Boltz-2 template, together with the protein sequences and glue SMILES. The ten first-pass structures with the highest CRBN–target chain-pair ipTM become templates for a second, refining co-folding stage (pass 2).

We chose Boltz-2 because its template interface allows the relative geometry of a docked protein pair to inform co-folding. Supplying CRBN and the target in a single mmCIF coordinate file preserves inter-chain template information, whereas separate single-chain templates provide no information about their relative orientation. By comparison, the AlphaFold 3 template interface only accepts templates for individual chains and cannot transfer a docked interface through this route. Boltz-2 therefore provides a direct connection between protein–protein docking and ternary-complex prediction. This conditioning also allows the model to improve upon the supplied geometry after the ligand is introduced, rather than requiring the ligand to fit a fixed docking pose.

For evaluation, models from both passes are re-clustered by RMSD. Each cluster is represented and ranked by its highest chain-pair ipTM member. We report the highest-confidence cluster (rank 1), the most accurate of the ten highest-ranked clusters (best-of-top-10), and the model with the lowest neosubstrate RMSD across all generated models (whole-ensemble oracle). These comparisons help distinguish failure to generate a suitable pose (sampling) from failure to rank it highly (ranking). Models are evaluated against experimental ternary complex structures in PDB using DockQ (a 0–1 interface-accuracy score, with higher values indicating better agreement) and CAPRI docking-quality thresholds: incorrect (<0.23), acceptable (≥ 0.23), medium (≥ 0.49) and high (≥ 0.80) [41,42]. Neosubstrate chain and ligand RMSDs are measured after superposition on the CRBN CULT backbone.

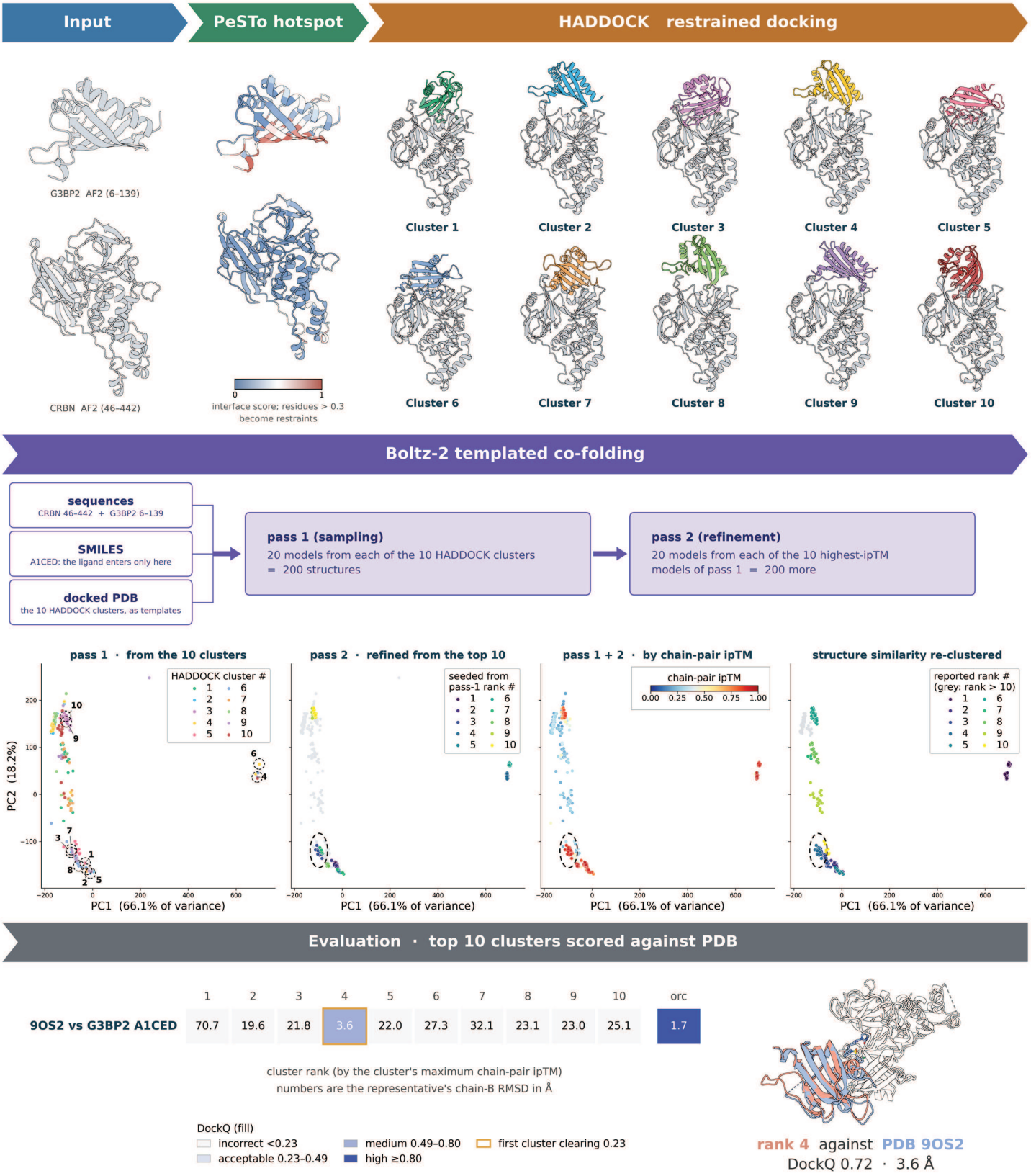


Figure 1. The complete GluePlex workflow illustrated using G3BP2 and compound A1CED as an example. AlphaFold models of G3BP2 (residues 6–139) and CRBN (residues 46–442) are scored by PeSTo. Surface-exposed residues with interface propensity above 0.3 define docking restraints. HADDOCK cluster representatives serve as Boltz-2 templates together with the glue SMILES. Pass 1 generates 20 models per template; pass 2 generates 20

models from each of the ten first-pass structures with the highest chain-pair ipTM. The four scatter plots share a principal component analysis (PCA) projection, which summarizes structural variation in two dimensions, and are colored by HADDOCK cluster, pass-1 ipTM rank, pass-2 ipTM and final cluster rank. Dashed outlines enclose models with DockQ ≥ 0.49 in this case. The evaluation row shows neosubstrate chain RMSD against PDB 9OS2 for the top ten clusters, colored by CAPRI tier. The reference-like pose ranks fourth (ipTM 0.914, versus 0.938 at rank 1), with DockQ 0.72 and RMSD 3.6 Å; the lowest RMSD in the ensemble (orc, oracle) is 1.7 Å. In the overlay, blue denotes the reference, red the prediction and white CRBN.

GluePlex performance is near-saturated on established G-loop degrons

Most known CRBN neosubstrates contain a G-loop degron, first resolved in casein kinase 1α (CSNK1A) [7] and GSPT1 [8]. The same motif supports recruitment of IKZF1, IKZF3 and other C2H2 zinc-finger proteins [9], and occurs more broadly across the proteome [18,20]. In this benchmark, 14 of 19 targets contain a G-loop and account for 47 of 52 target–compound pairs: six C2H2 zinc fingers, seven other β-hairpin loops and one helical variant.

GluePlex produced an acceptable or above rank-1 pose in 130 of 141 G-loop runs (47 pairs, three runs each; **Fig. 2A**). Considering the top ten clusters added two successes, and the whole-ensemble oracle added none. Thus, ranking contributed little to the remaining failures in this group. Performance was near saturation in both SEEN strata. Rank-1 poses were acceptable in 83 of 84 runs for SEEN (direct) and all 33 runs for SEEN (homologue), and the ensemble oracle added no successes beyond the top ten. By contrast, the eight novel G-loop pairs were less accurate, with 14 of 24 rank-1 successes and a median best DockQ of 0.45, compared with 0.92 and 0.88 for the two SEEN strata.

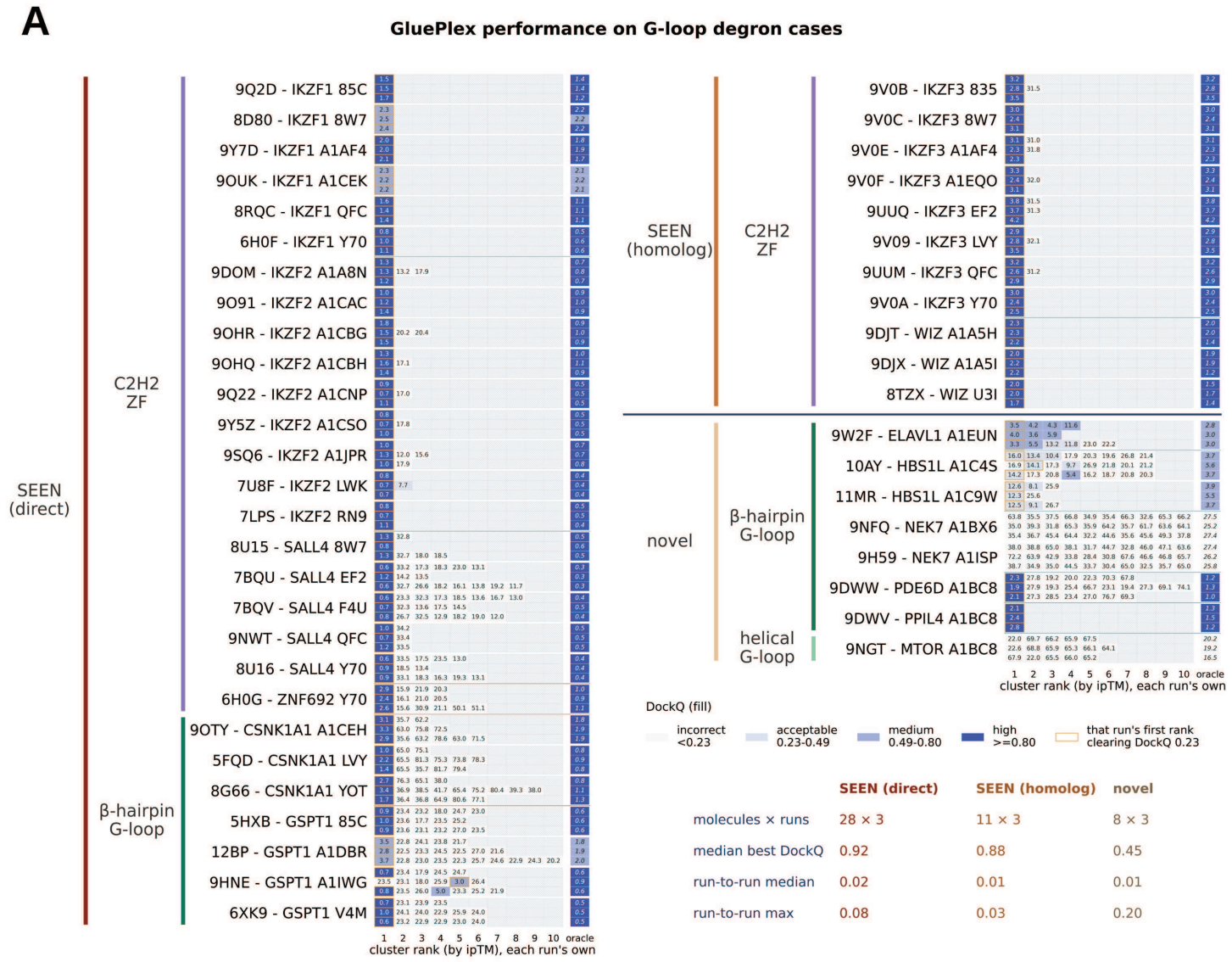
Protein and ligand errors differed among novel G-loop pairs (**Fig. 2B**). For MTOR–A1BC8, the target was misplaced (neosubstrate RMSD 22.0 Å; DockQ 0.05), while the ligand remained close to its reference pose in CRBN (RMSD 1.62 Å). For the two NEK7 compounds, both components were inaccurate: neosubstrate RMSDs were 28–32 Å and ligand RMSDs were 4.7 and 5.7 Å.

ELAVL1–A1EUN prompted closer examination of the reference ligand pose (**Fig. 2B**). Although GluePlex reproduced the protein interface (DockQ 0.751), the ligand differed from 9W2F by 3.29 Å RMSD, with a warhead centroid displacement of 2.1 Å. Inspection of 9W2F revealed a glutarimide pose distinct from those in the other CRBN structures surveyed. This pose would severely clash with Trp380 in its usual conformation within the tri-tryptophan pocket (**Fig. S1A**); the deposited model accommodates it through an unusual Trp380 rotamer which was never seen in 106 other CRBN chains with glues in PDB. By contrast, GluePlex picked the canonical Trp380 rotamer and glutarimide pose (**Fig. S1A**). We therefore compared the two ligand–Trp380 arrangements against the 3.4 Å cryo-EM map, EMD-65569, and its half maps to assess whether the predicted arrangement better explained the density.

The ligand and Trp380 gave mixed results when evaluated separately. For the ligand alone, map–model correlation favored the deposited pose in the full map over the GluePlex prediction in both half maps (**Fig. S1B**). For the Trp380 side chain alone, the full map favored the GluePlex prediction by 0.070, whereas the half maps differed by only 0.002 in opposite directions. However, evaluating the ligand and the entire Trp380 residue together favored GluePlex in all three

maps (**Fig. S1B**). Of 600 GluePlex models, 597 exceeded the deposited correlation in all three maps for this selection. These observations suggest an alternative model of ligand–Trp380 position from the deposited model. ELAVL1's ligand RMSD should therefore be interpreted cautiously, as it may reflect uncertainty in the reference pose as well as prediction error.

Accurate degron placement did not, however, extend to domains flanking the degron (**Fig. S2**). Across seven multi-zinc-finger structures of IKZF2 and SALL4, rank-1 DockQ was 0.85–0.96, but only the CRBN-contacting finger was positioned accurately (RMSD 0.6–2.0 Å). Flanking fingers had RMSDs of 6.8–28.4 Å; for IKZF2–RN9, the three fingers had RMSDs of 28.4, 0.8 and 6.8 Å. These poorly positioned flanking fingers nevertheless contribute to the CRBN interface: five of eight make 11–20 heavy-atom contacts with CRBN in the reference structures. Such contacts are sparsely represented in the training structures. In 8DEY, the only multi-finger CRBN ternary released before the Boltz-2 training cutoff, the flanking finger makes just one contact with CRBN. This limited representation may contribute to the poor placement of flanking domains. We use the degron-containing domain for subsequent analyses; accurate placement of a full-length or multi-domain construct should not be assumed.



	SEEN (direct)	SEEN (homolog)	novel
molecules × runs	28 × 3	11 × 3	8 × 3
median best DockQ	0.92	0.88	0.45
run-to-run median	0.02	0.01	0.01
run-to-run max	0.08	0.03	0.20

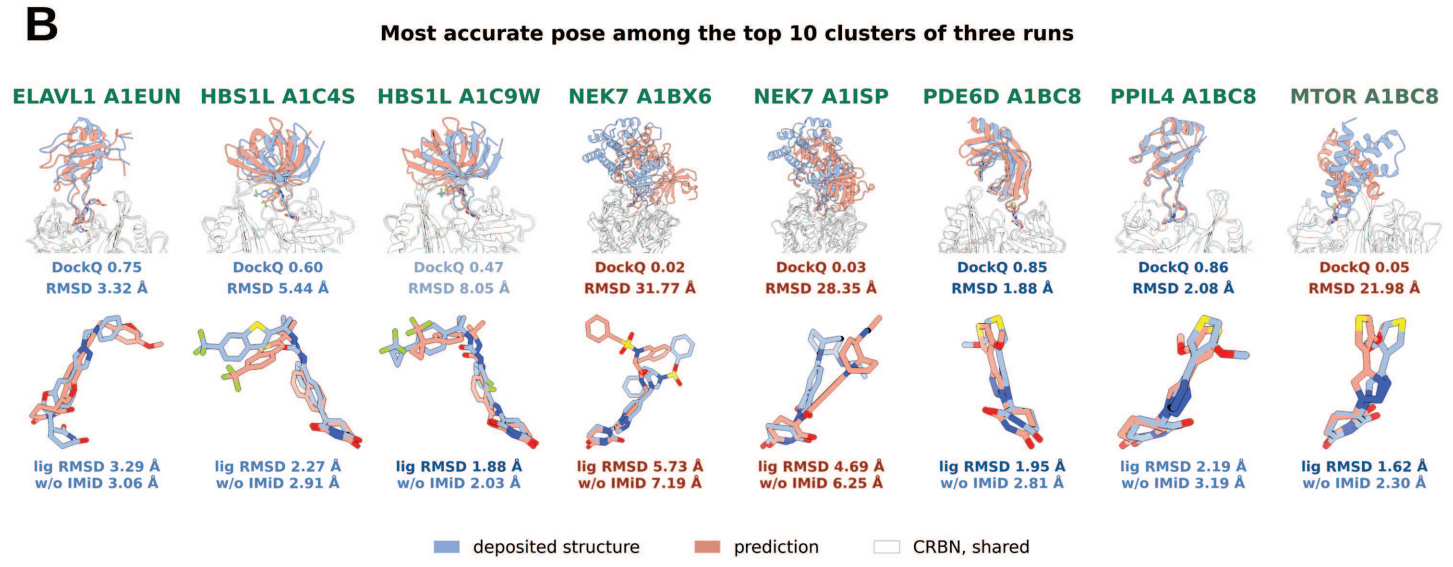


Figure 2. G-loop degron targets, stratified by training-set exposure, across three independent GluePlex runs. **(A)** GluePlex predictions for G-loop-containing target–compound pairs, grouped by training-set exposure and degron

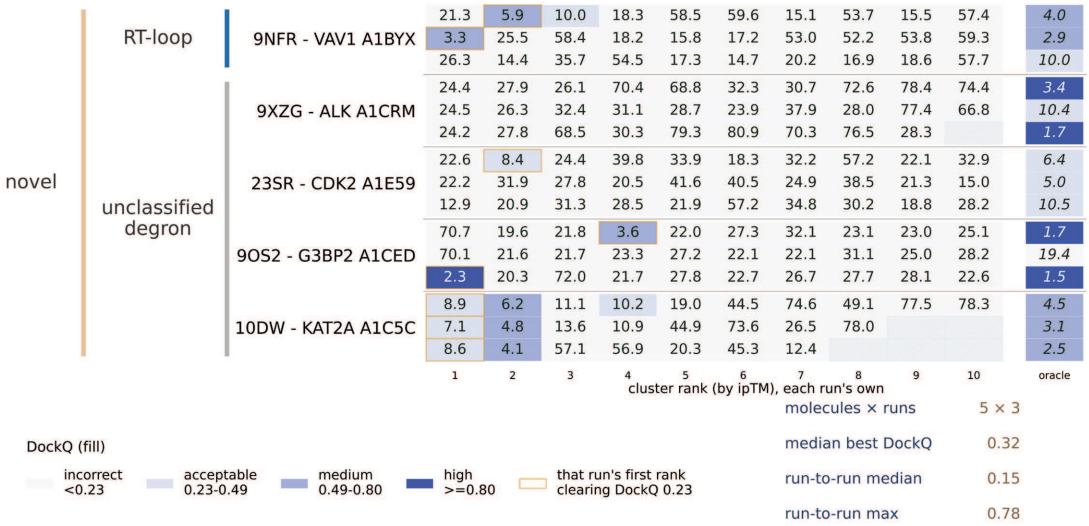
class. Each pair is labeled by its reference PDB entry and ligand identifier and shown as three stacked rows, one per independent run. Columns 1–10 show clusters in descending order of their representatives' CRBN–target chain-pair ipTM; fewer columns are populated when a run returns fewer than ten clusters. Each representative is the highest-ipTM model within its cluster. The final column shows the whole-ensemble oracle, defined as the model with the lowest neosubstrate chain RMSD among all structures generated in that run, irrespective of confidence rank. Cell colors indicate DockQ-based CAPRI quality tiers: incorrect (<0.23), acceptable (0.23 ≤ DockQ < 0.49), medium (0.49 ≤ DockQ < 0.80) and high (≥0.80). Numbers report neosubstrate chain RMSD (Å) after superposition on the CRBN CULT backbone. Amber outlines mark the highest-ranked cluster reaching DockQ ≥ 0.23 in each run. Comparing the ranked clusters with the oracle shows whether a reference-like pose was sampled and how highly it was ranked. **(B)** Best models for the eight novel G-loop parent compounds, selected by lowest neosubstrate chain RMSD among the top ten clusters of all three runs. Complexes are superposed on the CRBN CULT domain (white), with reference neosubstrates in blue and predictions in red. Ligands are shown separately in the same frame; carbons match the corresponding chain color and heteroatoms are colored by element. DockQ and neosubstrate RMSD are colored by CAPRI tier. Whole-ligand and warhead-excluded RMSDs are blue below 2 Å, pale blue from 2 to 4 Å and red above 4 Å.

GluePlex improves prediction for targets without a known degron motif

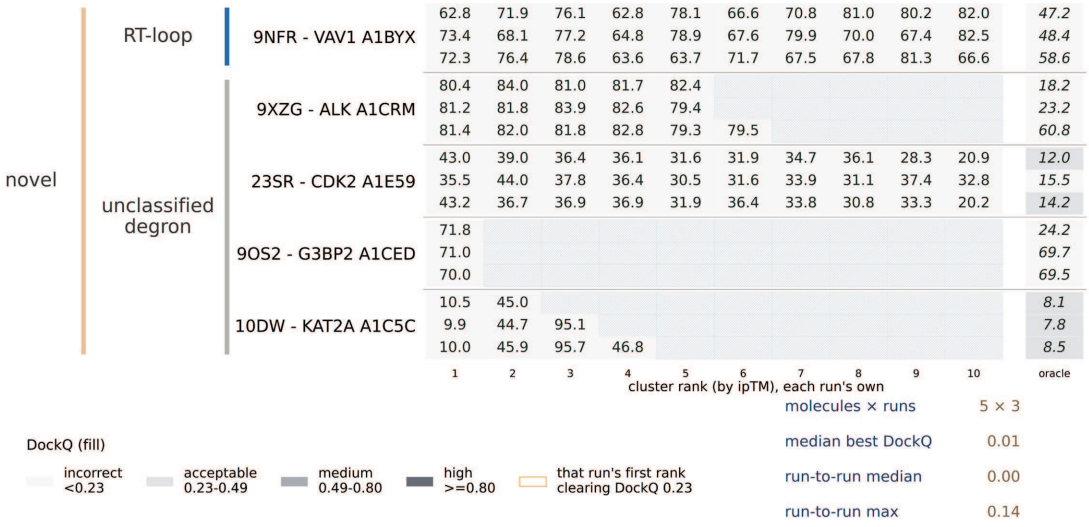
For the five targets without a known G-loop degron, GluePlex recovered acceptable poses (**Fig. 3A, C**) that baseline (plain Boltz-2 co-folding) missed (**Fig. 3B**). VAV1 reached best DockQ values of 0.55, 0.72 and 0.10 across three runs; the baseline returned 0.01 in each run, with an oracle no higher than 0.013. Across all 15 runs, GluePlex generated an acceptable pose in 14 ensembles, but retained one in the top ten in only eight runs and at rank 1 in five. Ranking was therefore the main limitation. For ALK–A1CRM, oracle DockQ values were 0.82, 0.39 and 0.91, whereas rank-1 DockQ never exceeded 0.04. The baseline produced no acceptable top 10 pose in these 15 runs, with a median best DockQ of 0.01 versus 0.32 for GluePlex.

As with the novel G-loop pairs, protein-interface accuracy did not imply ligand accuracy (**Fig. 3C**). For ALK–A1CRM, the target was misplaced (RMSD 23.9 Å; DockQ 0.04), while the ligand remained near its CRBN-bound pose (RMSD 2.23 Å). Conversely, G3BP2–A1CED had an accurate interface (DockQ 0.83; neosubstrate RMSD 2.34 Å) but a poorly placed ligand (RMSD 7.44 Å overall and 9.20 Å excluding the warhead).

A GluePlex performance on novel degron cases



B Boltz-2 MSA baseline performance on novel degron cases



C Most accurate pose among the top 10 clusters of three runs

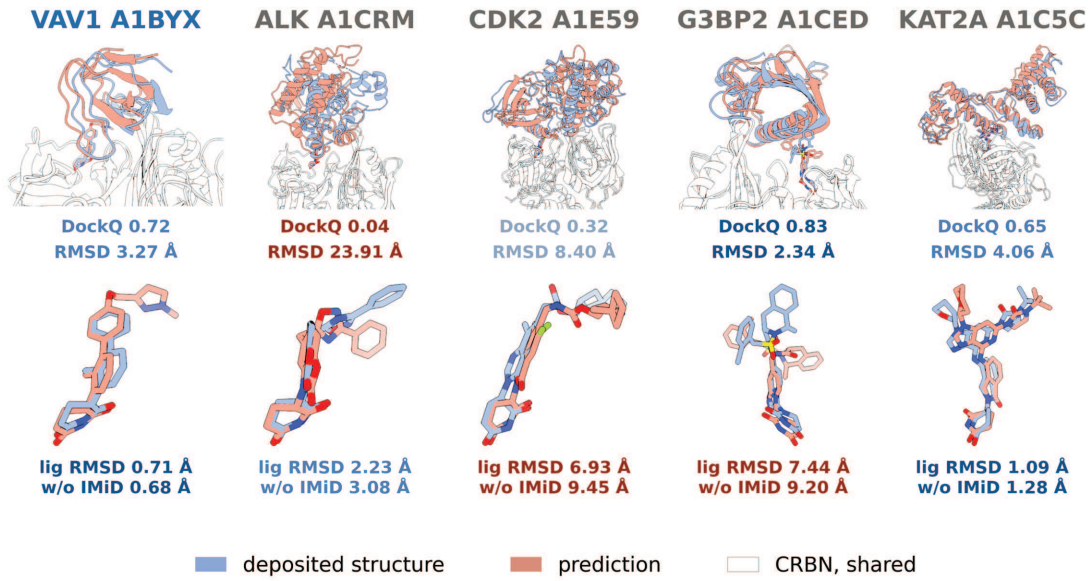


Figure 3. Targets without a known degron motif: GluePlex against the plain co-folding baseline, across three independent runs. (A) GluePlex predictions for five target–compound pairs without a known G-loop degron. Each pair is labeled by its reference PDB entry and ligand identifier and shown as three stacked rows, one per independent run. Columns 1–10 show clusters in descending order of their representatives' CRBN–target chain-pair ipTM; fewer columns are populated when a run returns fewer than ten clusters. Each representative is the highest-ipTM model within its cluster. The final column shows the whole-ensemble oracle, defined as the model with the lowest neosubstrate chain RMSD among all structures generated in that run, irrespective of confidence rank. Cell colors indicate DockQ-based CAPRI quality tiers: incorrect (<0.23), acceptable (0.23 ≤ DockQ < 0.49), medium (0.49 ≤ DockQ < 0.80) and high (≥0.80). Numbers report neosubstrate chain RMSD (Å) after superposition on the CRBN CULT backbone. Amber outlines mark the highest-ranked cluster reaching DockQ ≥ 0.23 in each run. Comparing the ranked clusters with the oracle shows whether a reference-like pose was sampled and how highly it was ranked. **(B)** Plain Boltz-2 co-folding with a server-generated multiple sequence alignment (MSA), no templates and no restraints, using identical scoring and a sampling budget (number of generated models) at or above that of the pipeline. **(C)** Best models selected by lowest neosubstrate chain RMSD among the top ten clusters of all three GluePlex runs. Complexes are superposed on CRBN CULT (white); reference neosubstrates are blue and predictions red. Ligands are shown separately in the same frame, with matching carbon colors and heteroatoms colored by element. DockQ and neosubstrate RMSD are colored by CAPRI tier. Whole-ligand and warhead-excluded RMSDs are blue below 2 Å, pale blue from 2 to 4 Å and red above 4 Å.

Truncating a glue rescues some failures and creates others

VAV1 predictions differed between the present benchmark and the earlier prospective campaign [21]. The benchmark compound reached an acceptable pose in two of three runs, whereas NGT-201-12 was placed at rank 1 in every attempt in the earlier campaign [21]. NGT-201-12 is a three-ring phenyl-glutarimide with 23 heavy atoms and no deposited ternary structure. The benchmark ligand A1BYX from PDB 9NFR has four rings and 29 heavy atoms. This difference suggested that ligand chemistry contributes to prediction reliability.

More broadly, the distal ligand moiety was a common source of error. In 9NFR, only 21 of 29 ligand heavy atoms are resolved; the N-methylpyrazole and OCH2 linker lack density. Across 87 models, splitting ligands into the CRBN-binding warhead and distal moiety (a median of 15 of roughly 30 heavy atoms per part) showed larger distal errors in 77 cases, with a paired median difference of 1.34 Å. Even at DockQ ≥ 0.80, median RMSDs were 0.61 Å for the warhead and 2.82 Å for the remainder (**Table 1**). A contact-based split (atoms within 4.5 Å of CRBN versus those within 4.5 Å of the target) gave the same pattern across 90 models (1.02 versus 2.67 Å). These results motivated testing whether removal of distal groups could improve prediction.

To isolate the ligand contribution, each truncation shared its parent's HADDOCK ensemble within a run, holding the docking input fixed while changing the ligand. Smaller analogues also resemble early screening hits more closely than optimized degraders. However, scoring a truncation against its parent's reference assumes that the two induce the same

ternary geometry. No experimental structures of the truncated analogues are available to test this assumption, so this analysis is exploratory and separate from the benchmark.

Ligand truncation improved prediction accuracy or consistency for some compounds but reduced accuracy for others (**Fig. 4**). The co-folding results of ALK, VAV1 and KAT2A glue truncations illustrate these three outcomes, respectively. Across 29 truncations, 15 changed DockQ by less than 0.05, whereas the six largest changes ranged from 0.21 to 0.88 in magnitude. Averaging across three runs, the paired ΔDockQ had a median of +0.010 and a mean of +0.033, with no significant overall shift at either the truncation level (Wilcoxon signed-rank test, n = 29, p = 0.16) or family level (n = 17, p = 0.17). These heterogeneous responses motivated identifying which compounds are likely to benefit from truncation.

Table 1. Ligand accuracy for novel target–compound pairs, split into the CRBN-binding warhead (the imide ring and its attached ring system) and distal moiety. Both RMSDs are measured after superposition on CRBN CULT domain.

	n	Warhead RMSD	Distal-moiety RMSD	n with distal-moiety RMSD < 2 Å
All	87	0.66 Å	2.82 Å	35 (40%)
Parent glues	36	0.69 Å	3.16 Å	7 (19%)
Truncations	51	0.55 Å	1.46 Å	28 (55%)
DockQ ≥ 0.80	11	0.61 Å	2.82 Å	4 (36%)

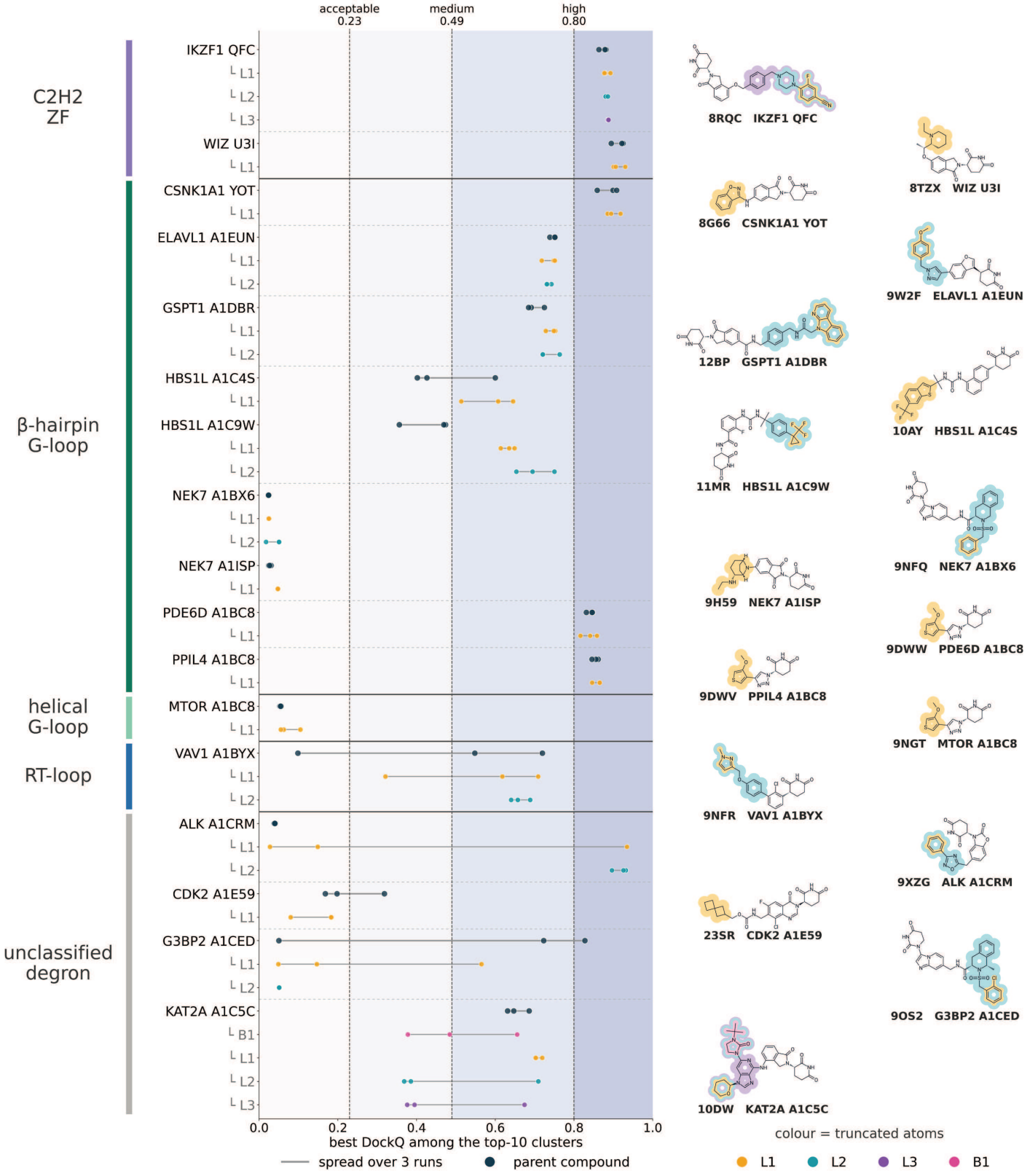


Figure 4. Effect of glue truncation on complex accuracy, across three independent runs. Forty-six compounds (17 parents and 29 truncations), each modeled in three independent runs. Points show individual runs and bars span their range. Each truncation shares its parent's HADDOCK ensemble within a run, allowing paired comparison of ligand changes. The right column shows truncations on the parent's two-dimensional structure; ring colors and

highlight radii identify the removed atoms. For KAT2A A1C5C truncations, L1 and B1 each removes one ring branch, L2 removes both branches, while L3 removes the whole distal arm.

For ALK, the depth of truncation mattered. The parent A1CRM returned DockQ 0.04 in every run. Removing only the terminal phenyl group (L1), while retaining the oxadiazole, gave DockQ values of 0.15, 0.03 and 0.93. Removing the entire arm with a remaining methyl group (L2) gave 0.90–0.93 in all three runs. Thus, the deeper cut improved both accuracy and consistency in this case. For VAV1, truncation mainly improved consistency. Parent A1BYX DockQ values of 0.55, 0.72 and 0.10 spanned medium-quality and incorrect poses, whereas A1BYX-L2 returned 0.69, 0.66 and 0.64. A best-of-three comparison would miss this benefit. The result is consistent with the more reproducible modeling of the smaller NGT-201-12 in the earlier campaign, although it does not establish the cause of that difference. In contrast, truncation in KAT2A often reduced accuracy. The parent consistently reached DockQ 0.63–0.69. Three of four truncations reduced DockQ to 0.37–0.48 in two of three runs, while the fourth had little effect. Truncation therefore requires a criterion for identifying likely benefits and losses.

The best models across three runs illustrate the range of outcomes for all 35 novel-stratum compounds (**Fig. 5**). ALK–A1CRM improved from DockQ 0.04 to 0.93 for both truncations, with ligand RMSD falling from 2.23 Å to 0.56 and 0.55 Å. VAV1–A1BYX changed little (0.72 to 0.71 and 0.69); MTOR–A1BC8 and NEK7 remained inaccurate (0.05 to 0.10 and 0.02–0.03, respectively). G3BP2–A1CED worsened from 0.83 to 0.57 and 0.05, and CDK2 from 0.32 to 0.18. Truncation to the bare warhead also had target-dependent outcomes: DockQ was 0.87 versus 0.86 for the PPIL4 parent, 0.82 versus 0.85 for PDE6D, 0.74 versus 0.75 for ELAVL1, and 0.10 versus 0.05 for MTOR. Three of these use the same 13-atom fragment, yet span DockQ 0.10–0.87. The effect therefore depends on the protein complex itself as well as the removed chemistry, consistent with molecular glues strengthening pre-existing weak protein interactions [43].

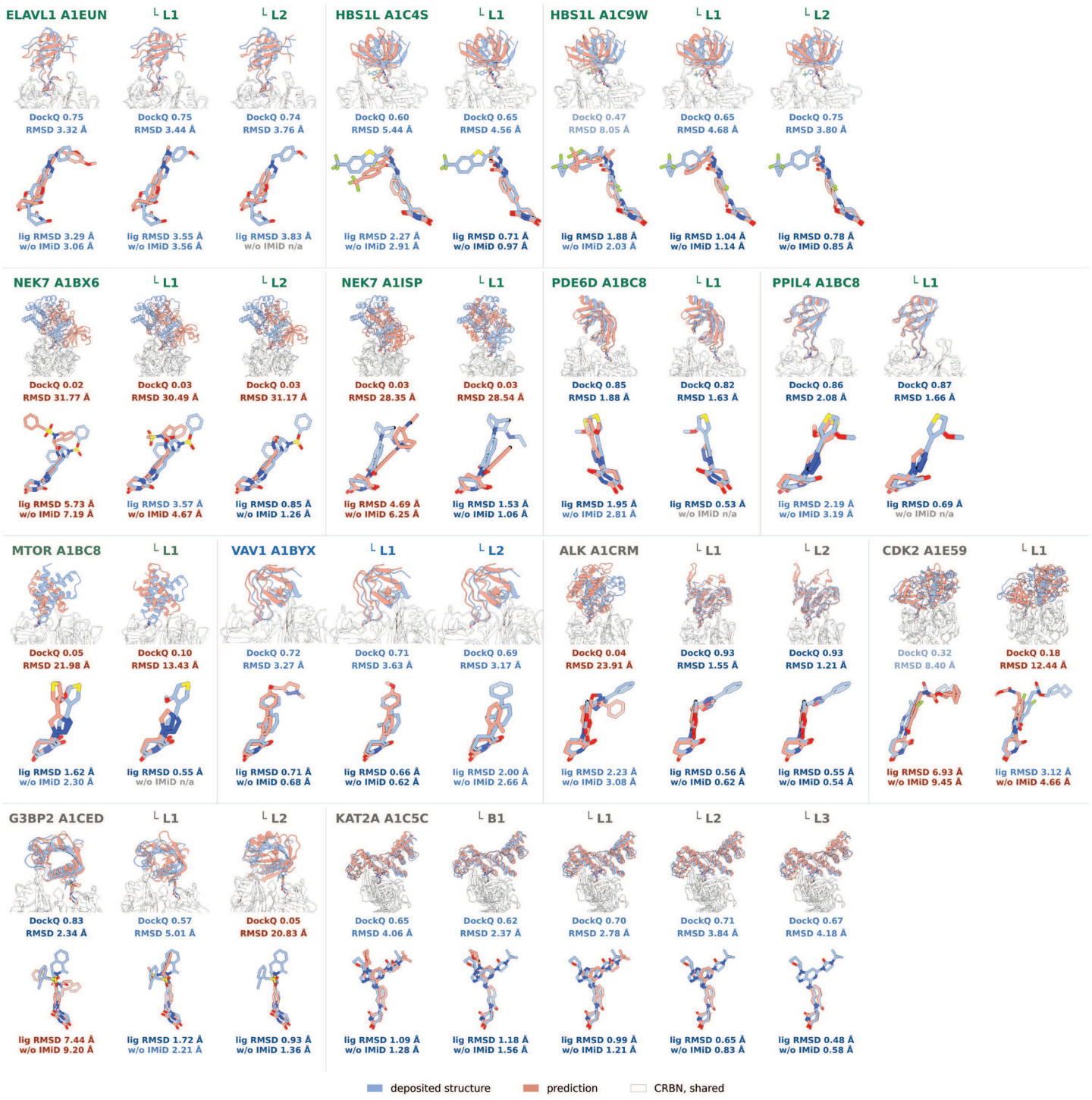


Figure 5. The most accurate complex reached for every compound of the novel stratum, with each parent glue shown beside the truncations derived from it. All 35 compounds of the novel stratum, arranged so that each parent is adjacent to its own truncations, with families separated by rules. Parents are named in full and truncations by the cut applied. Each panel shows the cluster representative with the smallest neosubstrate chain RMSD among the top ten of all three runs. Four truncations retain nothing beyond the imide ring and the ring system bonded to it; for these the warhead-excluded ligand RMSD is undefined and is reported as n/a.

We next tested whether these compound-specific effects persisted across docking inputs and whether confidence scores from the parent model could identify truncations likely to help.

Docking and template transfer account for most of the improvement

We compared five docking configurations that varied the source of the target interface prior, the target structure supplied to HADDOCK, or both. PeSTo priors came from an AlphaFold model, a union of AlphaFold and experimental structures, a selected experimental chain, or in the "nopesto" control, the solvent-accessible surface alone. Docked structures were the AlphaFold model, the experimental medoid (the chain with the lowest average structural deviation from the other experimental chains), or the chain with the strongest PeSTo signal. A sixth configuration used plain Boltz-2 co-folding without docking or templates as baseline (**Fig. 6**; Methods).

Each configuration was run three times on the 13 novel benchmark pairs and their 22 truncations (35 molecules; **Fig. 6**). Plain Boltz-2 co-folding ("msa": server MSA, no templates or restraints) used a sampling budget at or above that of the pipeline and identical scoring. On the benchmark pairs, it reached 12 of 39 acceptable poses, with median DockQ 0.01, versus 23 of 39 and median DockQ 0.43 for default GluePlex. Across all 35 molecules, the corresponding counts were 30 of 105 versus 67 of 105, including 30 versus 59 rank-1 successes. All the successful cases in plain co-folding still happen in targets with a G-loop degron.

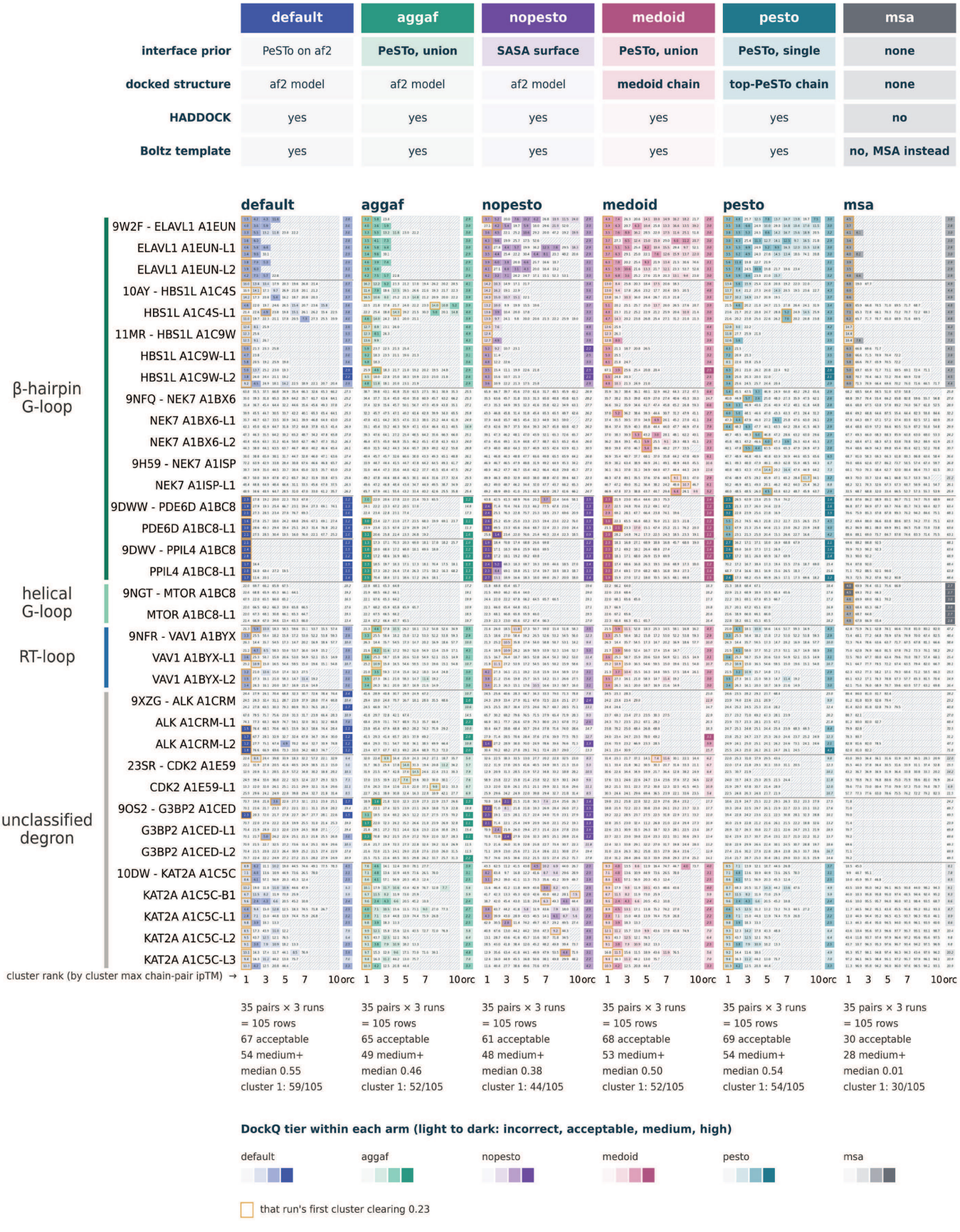


Figure 6. Pipeline ablation across 13 novel benchmark pairs and 22 truncations, with three independent runs per configuration. The table defines six pipeline configurations. Each target–compound pair has three stacked bands for independent runs. Columns show up to ten clusters ranked by representative chain-pair ipTM, followed by

the whole-ensemble oracle. Cell color indicates DockQ CAPRI tier within each configuration's color scale; values give neosubstrate chain RMSD (Å). Amber outlines mark the first cluster with DockQ ≥ 0.23. Summaries below each panel report acceptable and medium-or-better counts, median DockQ and rank-1 successes across 105 runs. The "msa" baseline uses a plain Boltz-2 co-folding with server MSA enabled, no templates or restraints, a sampling budget at or above that of the pipeline, and identical scoring.

Replacing PeSTo with solvent-accessibility restraints retained 23 of 39 benchmark successes but reduced rank-1 successes from 19 to 17. Across all 35 molecules, success fell from 67 to 61 and rank-1 success from 59 to 44; the larger difference arose from truncations. Thus, docking and template transfer accounted for most of the gain over plain co-folding, while PeSTo mainly improved ranking.

Across configurations and truncations, all 13 novel families had an acceptable pose among the top ten clusters. GluePlex configurations recovered 12 families, while MTOR–A1BC8 was recovered only by plain co-folding (DockQ 0.79–0.82 in all three runs). Conversely, the two NEK7 compounds required docking an experimental chain. Nevertheless, no single configuration recovered all 13 families, and the default recovered ten. Configuration selection and ranking therefore limit how much of the available accuracy is realized in any one run.

Parent-model confidence helps identify beneficial truncations

To identify when truncation might help, we evaluated 33 truncation pairs, each comparing a compound with a more truncated analogue of the same target (e.g. parent→L1 or L1→L2). Here, parent refers to the less truncated member of each comparison. We evaluated these pairs across five docking configurations and three runs, yielding 495 paired observations (**Fig. S3A, B**). This comparison tested whether truncation effects persisted when docking inputs or prior changed. Plain co-folding was excluded because it has no docking stage. Because the 15 observations (5 configurations × 3 runs) from each truncation pair are not independent, we estimated confidence intervals by resampling whole truncation pairs rather than individual observations (cluster bootstrap).

Truncation outcomes depended strongly on parent accuracy (**Fig. S4A**). For parents with medium-or-better DockQ, truncation improved 3 of 257 observations and worsened 71. Below medium quality, it improved 66 of 238 and worsened 2, a 53-percentage-point difference in the net improvement rate. The largest benefit occurred for acceptable-but-not-medium parents (37 of 81 improved, two worsened). Because DockQ requires a reference structure, we next examined confidence scores available from the parent model alone.

The parent's target–ligand chain-pair ipTM correlated with DockQ (Spearman $\rho = 0.60$), suggesting that it could help identify beneficial truncations without an experimental reference. We used the median ipTM across three runs to classify each parent–configuration combination because individual runs crossed the 0.80 threshold in 71 of 165 combinations. Below a median ipTM of 0.80, truncation improved 24% of 177 observations and worsened 3%. At or above 0.80, only 8% of 318 observations improved, while 21% worsened. Unconditional truncation improved 14% and worsened 15% (**Fig. S4B**). Thus, low parent ipTM identified a subset with a more favorable balance of improvements and losses.

The difference in net improvement rate between the two ipTM groups was 32.9 percentage points (95% confidence interval, 16.4–50.9, estimated by resampling whole truncation pairs; no bootstrap resamples below zero). This separation persisted when each target was omitted in turn (largest $p = 1.1 \times 10^{-5}$). Only in a single compound family under the "nopesto" configuration, 5 of the 6 losses among observations selected by the criterion occurred. Across the other four configurations, the criterion selected 138 observations, with 31 improvements and one loss.

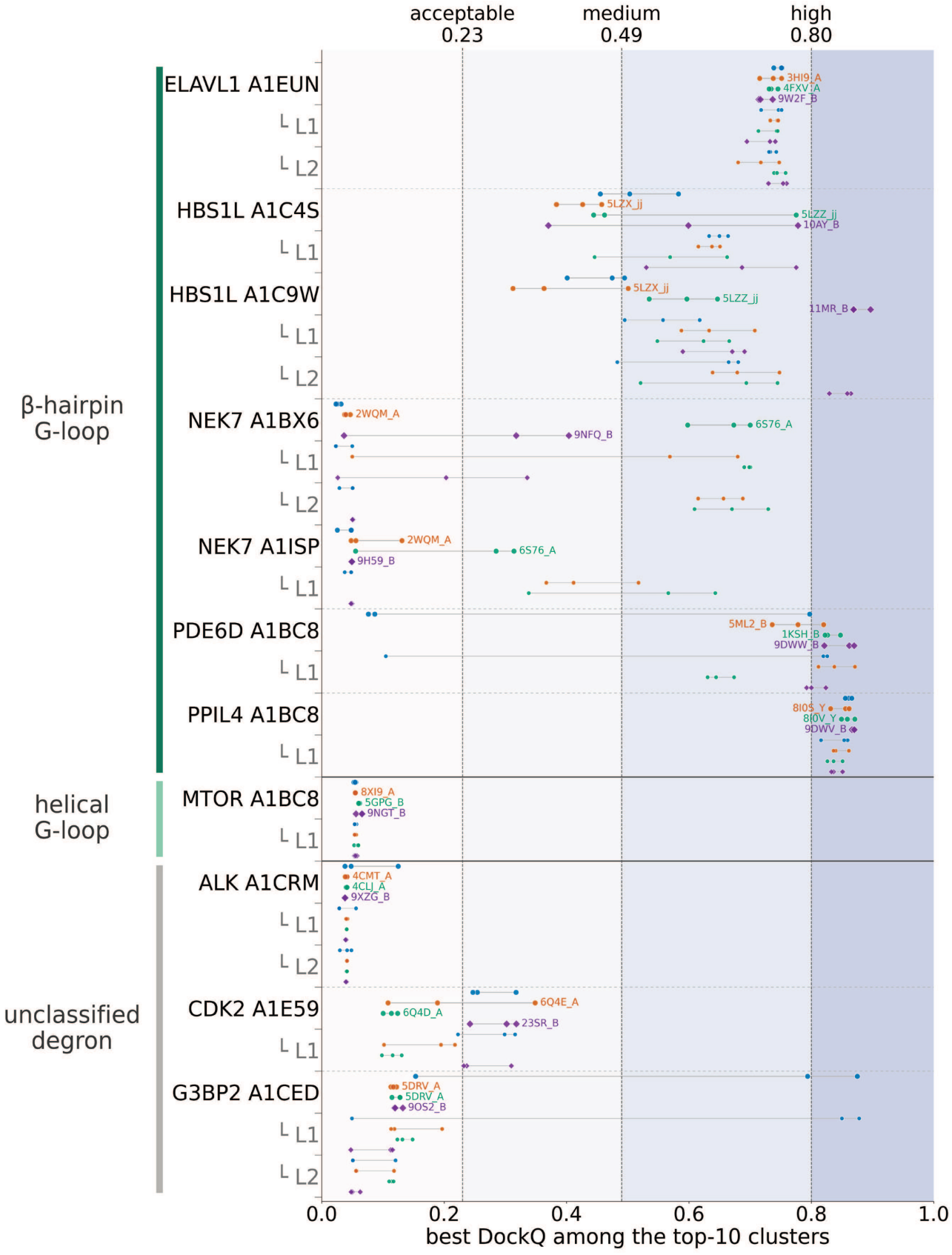
Because the 0.80 threshold was selected from these data, we also evaluated threshold selection by leave-one-gene-out validation. In each split, we selected the threshold using ten targets and evaluated it on the held-out eleventh. The resulting difference in net improvement rate was 33.2 percentage points, close to the in-sample value of 32.9. This supports the criterion's robustness within this dataset, although prospective validation is still needed.

Confidence in the moiety to be removed provided a second indicator (**Fig. S3, Fig. S4B**), showcased in **Fig. S5**. We measured this as the mean predicted local distance difference test (pLDDT) score, a measure of local structural confidence, across those atoms in the parent model. A threshold near 83 was selected in nine of eleven leave-one-gene-out validation splits. Removed-moiety pLDDT added little discrimination among observations already selected by ipTM < 0.80 (**Fig. S4B**). For parents with ipTM between 0.80 and 0.90, however, removing confidently modeled moieties was substantially more likely to reduce accuracy (**Fig. S4B**). We therefore retained parent ipTM as the primary criterion because it selected a smaller subset with a higher improvement rate and a lower loss rate, while removed-moiety pLDDT helped flag potentially detrimental cuts.

Conformer choice affects individual compounds more than overall success rate

Beyond ligand chemistry, the target conformation supplied to docking may also affect prediction accuracy. However, the **Fig. 6** ablation experiment varied both the docked conformer and the interface prior, while unseeded HADDOCK runs introduced additional sampling variability. To separate these effects, we compared target conformers under a fixed interface prior across repeated runs.

We docked four target conformers under the same PeSTo union prior (interface restraints pooled across target structures), holding the compound, reference and scoring fixed (**Fig. 7; Fig. S6**): the AlphaFold model, experimental medoid, experimental chain with the strongest PeSTo signal, and the target chain of the deposited ternary reference. The last conformer was a bound-conformer control, supplying the target conformation but no ternary interface. The comparison covered 9 targets, 11 parent glues and 16 truncations, with three runs per conformer input.



chain selection (compounds ranked first / ranked first by more than 0.15)

af2 model (6 / 2) · medoid chain (1 / 0) · top-PeSTo chain (5 / 3) · ternary PDB chain (9 / 2)

spread over 3 runs

Figure 7. Effect of the docked conformer under a single interface prior. Each row shows one compound, with four conformer inputs at fixed positions. Points represent independent runs, and bars span the three runs for each input. Experimental chains are labeled by identifier. The four inputs are the AlphaFold model, experimental medoid, chain with the strongest PeSTo signal, and reference-bound target chain. All use the same PeSTo union prior, compound, reference and scoring. Reference-bound target chain is a control rather than a blind prediction input. The comparison includes 27 compounds across 9 targets, with three runs per input.

Conformer effects were small for most compounds. Each pairwise comparison left 17–22 of 27 molecules within $|\Delta\text{DockQ}| \leq 0.15$, and the median paired difference was below 0.01. The conformer yielding the highest DockQ was consistent across all three runs for only eight molecules; it changed between runs for the other 19. **Table 2** therefore compares median DockQ across three runs for the 21 molecules with an acceptable pose under at least one input.

Table 2. Conformer comparison based on median DockQ across three runs for 21 compounds with DockQ ≥ 0.23 under at least one input. Columns count compounds with the highest median, a lead greater than 0.15, medium-or-better quality and high quality. The ternary PDB chain is the reference-bound control.

Chain used	highest median DockQ	median DockQ lead > 0.15	median DockQ	DockQ ≥ 0.49	DockQ ≥ 0.80
AlphaFold model	6	2	0.299	12	4
medoid chain	1	0	0.411	12	3
top-PeSTo chain	5	3	0.569	15	3
reference-bound target chain	9	2	0.302	12	5

Despite similar aggregate performance, 10 of 27 compounds crossed the acceptable threshold when the conformer changed. Both NEK7 parents and their three truncations failed with AlphaFold inputs (DockQ 0.02–0.05) but reached 0.28–0.70 with experimental chains. PDE6D–A1BC8 similarly improved from 0.09 to 0.78–0.86. The reverse occurred for G3BP2–A1CED and its first truncation: DockQ fell from 0.79 and 0.85 with AlphaFold inputs to 0.11–0.13 with experimental chains. CDK2 showed the same direction of change. Relative to AlphaFold, the best blind experimental input converted six failures to successes and four successes to failures.

Among comparisons exceeding $|\Delta\text{DockQ}| = 0.15$, the top-PeSTo chain outperformed the medoid for 4 molecules versus 1 across all 27, and 3 versus 0 among the 11 parents. It also had the highest median DockQ and most medium-or-better results. However, the reference-bound control did not consistently outperform this input: it won 4 of 9 comparisons overall and 2 of 4 among parents. Its oracle DockQ was more than 0.15 lower for 8 molecules and higher for 1, and it rescued none of the six molecules that failed with every blind input. Nevertheless, it did produce more high-quality poses: 8 oracle results versus 3-6 for blind inputs, and 5 top-10 results versus 3-4. Thus, the bound conformer improved high-quality recovery in some cases without increasing the range of compounds solved.

These results support sampling several target conformers when resources permit. No single input was consistently best, and conformers recovered different subsets of compounds. If only one input can be used, the top-PeSTo experimental chain performed best in this dataset; an AlphaFold model remains the option when no experimental structure is available. Experimental medoid and top-PeSTo inputs exceeded AlphaFold by $|\Delta\text{DockQ}| \geq 0.15$ in 4 versus 2 and 6 versus 4 comparisons, respectively.

Discussion

GluePlex improved prediction of ternary complex structures for novel CRBN neosubstrates, increasing acceptable runs from 12 to 23 of 39 and median DockQ from 0.01 to 0.43. Performance on SEEN G-loop pairs was near saturation. For previously uncharacterized G-loop targets, GluePlex may help identify the degron, but aligning candidate G-loop motifs to known CRBN ternary structures provides a more reliable starting point when these motifs can be recognized and localized. This approach specifies the candidate recruitment interface without relying on co-folding to select it. For example, in our earlier VAV1 study, we modeled LIMD1 recruitment by aligning its ZBD3 G-loop to that of CK1α, with subsequent mutagenesis supporting the predicted interface [21]. GluePlex is therefore most useful when no suitable degron template can be assigned, as was the case for the non-canonical VAV1 degron.

The ablation experiment in the GluePlex pipeline identifies restrained docking and template transfer as the main sources of improvement. By comparison, PeSTo provided a smaller benefit, most apparent at rank 1. Implementing docking and template transfer should therefore take priority over tuning the interface prior.

From a medicinal chemistry perspective, this retrospective benchmark extends the earlier prospective VAV1 campaign [21] and suggests that GluePlex may be particularly useful for structural exploration based on weak hits in early medicinal chemistry. At this stage, a model that identifies a plausible recruitment interface can help prioritize substitutions and sites for compound growth before an experimental ternary complex structure can be determined, which usually demands molecular glues with high affinities. The truncation results support this use: simplifying the ligand rescued some failed predictions and improved consistency across runs in others, indicating that small fragment-like ligands can provide useful starting points for modeling even when larger analogues are difficult to place. This is consistent with our earlier VAV1 campaign, in which GluePlex supported rationalizing the small, three-ring degrader hit NGT-201-12. However, truncation also reduced accuracy for complexes already with high DockQ or high confidence, so ligand size alone does not determine whether modeling will succeed. These predictions should guide testable chemical and mutational hypotheses; their practical value requires experimental evaluation, particularly because the truncation analysis assumes that each analogue retains its parent's ternary complex geometry, a hypothesis that needs to be tested experimentally.

A complementary route to interface restraints is to measure them experimentally, for example by hydrogen–deuterium exchange or cross-linking mass spectrometry, and to supply the resulting contacts to a co-folding model [44]. This is attractive where it is feasible, but it carries a practical circularity: obtaining interpretable Hydrogen-Deuterium exchange or cross-linking signal for a glue-induced ternary complex generally requires a compound potent enough to populate the ternary complex at the concentrations the experiment demands, and a compound of that quality is often one for which a cryo-EM or crystallographic structure is already attainable. The structural guidance therefore arrives at the stage of a campaign where it is least needed. For prospective medicinal chemistry the requirement is the reverse: the useful question

is what to synthesize next from a weak, fragment-like hit identified in a biochemical ternary complex assay. Such hits give a clear positive signal for complex formation, and may not yet show degradation in cells, but they bind far too weakly to support cryo-EM, crystallography or hydrogen–deuterium exchange. GluePlex is positioned at that stage by design, because its restraints are derived from computed interface propensities and restrained docking rather than from measurements on the complex itself, so it requires only the identity of the target and the compound rather than a structurally characterized complex.

Despite these gains, ranking is a major remaining limitation. For targets without a known degron motif, acceptable poses occurred in 14 of 15 ensembles but appeared in the top ten in only eight runs and at rank 1 in five. Pooling three runs before re-clustering recovered additional acceptable poses but did not improve rank-1 success. These results suggest that better pose selection could be more useful than additional sampling alone. One direction is to complement model confidence with physics-based or learned scores that assess the plausibility of both protein–protein and protein–ligand interactions. Such scores should be tested for their ability to recover accurate poses already present in the ensemble. Distal ligand placement also requires explicit evaluation, because an accurate protein interface does not guarantee an accurate ligand pose.

Conformational coverage remains another limitation, both between domains and within individual domains. In multi-zinc-finger constructs, GluePlex accurately placed the degron-containing finger but failed to reproduce the relative arrangement of adjacent fingers (**Fig. S2**), supporting the use of isolated degron domains. At the local level, CDK2 illustrates a failure to recover the bound activation-loop conformation. In the deposited ternary complex (PDB 23SR), activation-loop residues 145–172 differ from all 481 other CDK2 chains in PDB that resolve the full loop: the closest loop RMSD is 4.86 Å and the median is 8.96 Å. Among 1380 GluePlex models, 1363 were closest to inactive CDK2 which is the most abundant conformation in PDB, and none placed the loop within 2 Å of the ternary reference. Even the 17 models reaching DockQ 0.30 had loops close to the inactive structure and far from the ternary reference. DockQ correlated negatively with distance to the inactive structure (Spearman $\rho = -0.37$). This case shows that a plausible interface score can coexist with failure to recover the ligand-associated target conformation.

Overall, GluePlex shows how a docking-guided workflow can harness an existing co-folding foundation model to predict CRBN ternary complexes that plain MSA-based co-folding often misses, without retraining the model. Its current limits are model ranking, target conformational coverage and distal ligand placement. Multiple conformers and confidence-guided truncations can recover additional cases, although neither guarantees an accurate pose. Prospective studies should test whether these strategies improve compound design and whether truncated ligands retain their parent's ternary geometry. More broadly, harnessing foundation models for molecular discovery requires connecting their predictive capabilities with structural sampling, physical information and task-specific evaluation. GluePlex provides one such framework, whose components can evolve alongside advances in docking, scoring and co-folding. Although the present implementation uses CRBN-specific docking restraints, the strategy of transferring candidate protein–protein interfaces into co-folding is not inherently restricted to CRBN and could be adapted to other ligases and to non-degrading molecular glues.

Methods

Benchmark construction

We combined RCSB entries containing a polymer mapped to human CRBN (UniProt Q96SW2) with a full-text search for "cereblon". Entries were retained if they contained CRBN, a neosubstrate of at least 20 residues, and a non-polymer entity above 200 Da contacting both proteins within 5 Å. Crystallization additives, PROTACs, ligase machinery (DDB1, DDA1, CUL1/4A/4B, RBX1, NEDD8, ubiquitin and SKP1/2) and common fusion tags were excluded.

Reference files contain exactly three chains: CRBN (the copy that contacts the glue), the neosubstrate (the copy contacted by that same glue molecule) and the glue as a single HETATM residue.

IKZF3 was excluded from the multi-zinc-finger analysis because all eight deposited structures (9UUM, 9UUQ, 9V09, 9V0A, 9V0B, 9V0C, 9V0E and 9V0F) are ubiquitylation assemblies containing UbcH5a. Its second finger contacts a component absent from the predictions, confounding the displacement comparison. The only two-finger CRBN ternary released before the Boltz-2 training cutoff was 8DEY.

Strata, modality and exclusions

We used the Boltz-2 structural training cutoff of 2023-06-01. Each neosubstrate was classified by the release date of its earliest CRBN ternary structure. Targets without a pre-cutoff ternary were screened for homologues among exposed targets using MMseqs2, requiring ≥ 40% identity over ≥ 50% of the shorter domain [45]. The resulting strata were SEEN (direct), SEEN (homolog) and novel.

Docking inputs

Observed neosubstrate residues were mapped to the canonical UniProt sequence by local alignment, and the contiguous span defined the modeled domain. Default docking inputs were AlphaFold Protein Structure Database models cropped to this span. Cropped sequences matched the UniProt subsequence for every target.

The GluePlex pipeline

Inputs. Each run used the target UniProt sequence and residue range, the glue SMILES string, and a reference structure used only for scoring. Target models were cropped as described above. The CRBN input was the AlphaFold model of human cereblon (UniProt Q96SW2), residues 46–442; this construct also defined the restraint numbering and superposition frame.

Step 1, interface prior. PeSTo was run locally, with per-residue propensities stored in the B-factor column. Active target residues required Cα propensity ≥0.3 and solvent accessibility by NACCESS [46]. If none qualified, the solvent-accessible surface was used. This fallback applied to eight targets, including six C2H2 zinc-finger degrons of 26–55 residues whose maximum PeSTo scores were 0.01–0.19. Passive residues were generated with haddock3-restraints passive_from_active, and active/passive sets from both proteins were combined into one ambiguous interaction restraint table.

CRBN PeSTo scores were dominated by the DDB1-binding interface. Excluding this interface left three above-threshold residues on the glue-binding face: F150, E377 and W386. These restraints were fixed across targets.

Step 2, restrained docking. HADDOCK3 docked CRBN and the target without the ligand. The protocol used topology generation with autohis off, rigid-body sampling of 1000 orientations at restraint tolerance 5, selection of the best 200, semi-flexible refinement, short restrained molecular-dynamics refinement at the same tolerance, and clustering by fraction of common contacts (FCC), which groups models with similar inter-chain contact patterns. The four best models per cluster were retained. The benchmark produced 6–19 clusters per target; the first model of each cluster was exported as an mmCIF co-folding template. HADDOCK3 ran locally, with CNS jobs distributed by multiprocessing on one node.

Steps 3 and 4, templated co-folding. Each HADDOCK cluster representative was supplied to a separate Boltz-2 job with both protein sequences and the glue SMILES. Both chains used "msa: empty", meaning that no multiple sequence alignment was supplied. Pass 1 generated 20 diffusion structures per template. The ten structures with the highest CRBN–target chain-pair ipTM became templates for pass 2, which generated 20 structures per template. This yielded approximately 20n + 200 structures per compound, where n is the HADDOCK cluster count. Seeded repeats used a fixed random-number seed to control stochastic sampling.

Ablation arms

Five ablation configurations shared the docking and co-folding workflow but varied the target structure used to define the prior, the structure docked by HADDOCK, or both. All used the same three CRBN restraint residues. The sixth configuration, "msa", used plain Boltz-2 co-folding.

The "aggaf" and "medoid" configurations used a union prior to compensate for weak signals from individual structures. For example, the NEK7 AlphaFold model had only one residue above PeSTo 0.3, whereas experimental structures of the domain had stronger signals. The union applied the same threshold to the AlphaFold model and available PDB structures of the gene, then merged qualifying residues into one prior.

The "medoid" configuration docked the experimental chain with the lowest mean Cα RMSD to the other experimental chains and used the union prior. By contrast, the "pesto" configuration docked the chain with the strongest PeSTo signal and used that chain's own prior. Since both coordinates and prior differed, this comparison did not isolate conformer choice.

The "nopesto" configuration replaced PeSTo restraints with the solvent-accessible surface. It matched the default prior for targets where no residue reached 0.3 and the default already used this fallback. Eight targets met this condition; only CDK2 belonged to the novel stratum used for ablation.

The "msa" configuration used Boltz-2 with a server MSA, greedy pairing and no template. It was the only configuration without "msa: empty". It sampled 440 diffusion structures, matching the median count of a full two-pass run, and used the same scoring, clustering and ranking.

Scoring, ranking and the superposition anchor

DockQ and the CAPRI tiers are computed as defined in refs [41] and [42]. Models are clustered by average linkage on chain-B Cα at 5.0 Å with a minimum size of 5, scoped to the degron for multi-domain constructs, and clusters are ordered by the maximum CRBN–target chain-pair ipTM among their members.

All reference-based RMSDs, including ligand RMSD, were measured after superposing CRBN CULT backbone residues 319–425. Whole-CRBN alignment can be affected by motion between the Lon and CULT domains, whose relative position depends on the glue [47]. Similarly, alignment to ligand-lining residues from both proteins, as used by MGBench, can shift the ligand frame when the neosubstrate is misplaced. CULT provides a common local frame for the binding pocket; restricting the fit to nine pocket residues was considered under-determined. The ligand-lining convention was not retained in the released code, so no numerical comparison of alignment methods is reported.

Sampling recall measures whether an acceptable pose occurs anywhere in the ensemble (i.e. whether the oracle is at or above acceptable). Ranking recall measures whether ipTM ranking retains one at rank 1 or within the top 10.

The fixed-prior conformer comparison

Four conformers were docked under a fixed PeSTo union prior: the AlphaFold model, experimental medoid, chain with the strongest PeSTo signal, and target protein chain of the reference structure. The last supplied the bound target conformation but no ternary interface and served as a bound-conformer control. The comparison covered 9 targets, 11 parent glues and 16 truncations, with three runs per input. Run labels were not shared across inputs, so runs were paired by position. This pairing was arbitrary rather than a matched-seed design.

Ligand accuracy of the warhead and distal moiety

Ligand RMSD was evaluated separately for the CRBN-binding warhead and distal moiety (all remaining atoms). The warhead comprised the six-membered imide ring (glutarimide or dihydrouracil) plus one fused or bonded ring system. Chemical features were assigned from each run's SMILES and mapped to the reference by bond-generic substructure matching, because reference molecules were inferred from coordinates without bond orders. When the reference resolved fewer atoms, matching was reversed and analysis was restricted to resolved atoms. A complementary contact-based split compared atoms within 4.5 Å of CRBN with those within 4.5 Å of the target.

Overlay panels

Each panel shows the cluster representative with the smallest neosubstrate chain RMSD among the top ten of all three runs. Structures are superposed on CRBN's CULT domain.

Computational cost

Wall-clock cost per stage is given in **Table 3**, measured over the 142 compound-runs of the fixed-prior conformer comparison, on NVIDIA H100 NVL GPUs and a 192-thread host.

Table 3. Wall-clock cost per stage, measured over the 142 compound-runs of the fixed-prior conformer comparison, on NVIDIA H100 NVL GPUs and a 192-thread host.

stage	resource	wall time per job

PeSTo prior and HADDOCK (1000 rigid-body models, 60 CNS workers)	60 CPU threads	34 min median, IQR 32 to 46, max 69, per target
Boltz-2 passes 1 and 2 (20 samples from each of 10 templates, twice: 400 structures)	1 GPU	65 min median, IQR 49 to 72, max 99, per compound

Truncations reused their parent's PeSTo and HADDOCK outputs through symbolic links. A parent with k truncations therefore required one docking job and k + 1 GPU jobs. A full target–compound run cost approximately 1.1 GPU-hours and 0.6 CPU-hours using 60 threads. Storage is about 0.5 GB per compound-run once Boltz-2 per-model predicted aligned error (PAE) and predicted distance error (PDE) arrays are deleted, which the pipeline does after scoring.

Statistics

ΔDockQ was calculated as truncation minus parent DockQ within the same run and docking configuration, sharing the sampling seed and docked input. The decision-rule analysis included 495 observations from 33 truncation pairs across five configurations and three runs. Runs were analyzed individually. Outcomes were classified as improved (ΔDockQ > 0.15), worsened (ΔDockQ < −0.15), or unchanged. Net improvement rate was the fraction improved minus the fraction worsened, with unchanged observations retained in the denominator.

Rule performance was measured as the difference in net improvement rate between observations meeting the parent target–ligand ipTM criterion (<0.80) and the remainder. Confidence intervals were estimated from 4000 cluster-bootstrap replicates, resampling the 33 truncation pairs with replacement and retaining all 15 observations per selected pair. Replicates with fewer than five observations in either group were excluded; 95% confidence intervals used the 2.5th and 97.5th percentiles.

Threshold selection was evaluated by leave-one-gene-out validation, selecting the threshold on ten targets and evaluating it on the held-out eleventh. Associations between rule selection and outcome were assessed using two-sided Fisher's exact tests restricted to improved and worsened observations. Separately, overall truncation effects were assessed using two-sided Wilcoxon signed-rank tests on run-averaged ΔDockQ at the truncation (n = 29) and compound-family (n = 17) levels.

Data availability

Deposited structures used as references are available from the Protein Data Bank under the accession codes listed in the supplementary methods. AlphaFold models of the targets are available from the AlphaFold Protein Structure Database. Derived score tables for every run reported here are provided as supplementary data.

Code availability

Structure prediction used Boltz-2 v2.2.1 with Python 3.12.12, PyTorch 2.9.1 and CUDA 12.8 (NVIDIA driver 565.57.01). Protein–protein docking used HADDOCK3 v2025.11.0, and interface prediction used PeSTo commit ba651aa with checkpoint i_v4_1_2021-09-07_11-21, in a Python 3.9.25 environment with PyTorch 2.8.0 and NumPy 2.0.2. Target

structures were AlphaFold DB monomer v2.0 models retrieved on 1 August 2025. Scoring and analysis used Python 3.12.4, NumPy 1.26.4, SciPy 1.16.3, pandas 2.3.3, Biopython 1.86 and RDKit 2024.09.6; ligand processing within the prediction environment used RDKit 2025.9.3. Plots and molecular renderings were generated with Matplotlib 3.10.8 and UCSF ChimeraX 1.10.1, respectively. The GluePlex pipeline and the benchmark scripts are available at https://github.com/Hanfeng-Lin/GluePlex

Acknowledgements

HADDOCK3 was carried out using the SBGrid software environment, which provides curated, version-controlled access to structural biology software and manages dependencies to support reproducible research. Software execution was supported through the SBGrid Capsules framework, which standardizes application environments and workflows across platforms, as described in [48].

Funding Statement

The research was supported in part by NIH R01-AI198974 (to J.W.), the Michael E. DeBakey, M.D., Professorship in Pharmacology and the seed funding for Center for NextGen Therapeutics. Haiyang Zheng was supported by a CPRIT fellowship (RP210027).

Author Contributions Statement

H.L. and J.W. conceptualized the study. H.L. and H.Z. conducted experiments. All authors contributed to the development or design of methodology. H.L. and H.Z. developed the code, supporting algorithms and curated the data and repositories. H.L. and H.Z. analyzed the data. H.L. drafted the manuscript. J.W. guided and reviewed the manuscript. All authors read and approved of the final manuscript.

Competing Interest Statement

J.W. is the co-founder of Chemical Biology Probes LLC. J.W. has stock ownership in CoRegen Inc and serves as a consultant for this company. J.W. is a co-founder of Fortitude Biomedicines, Inc. and holds equity interest in this company. The remaining authors declare no competing interests.

References

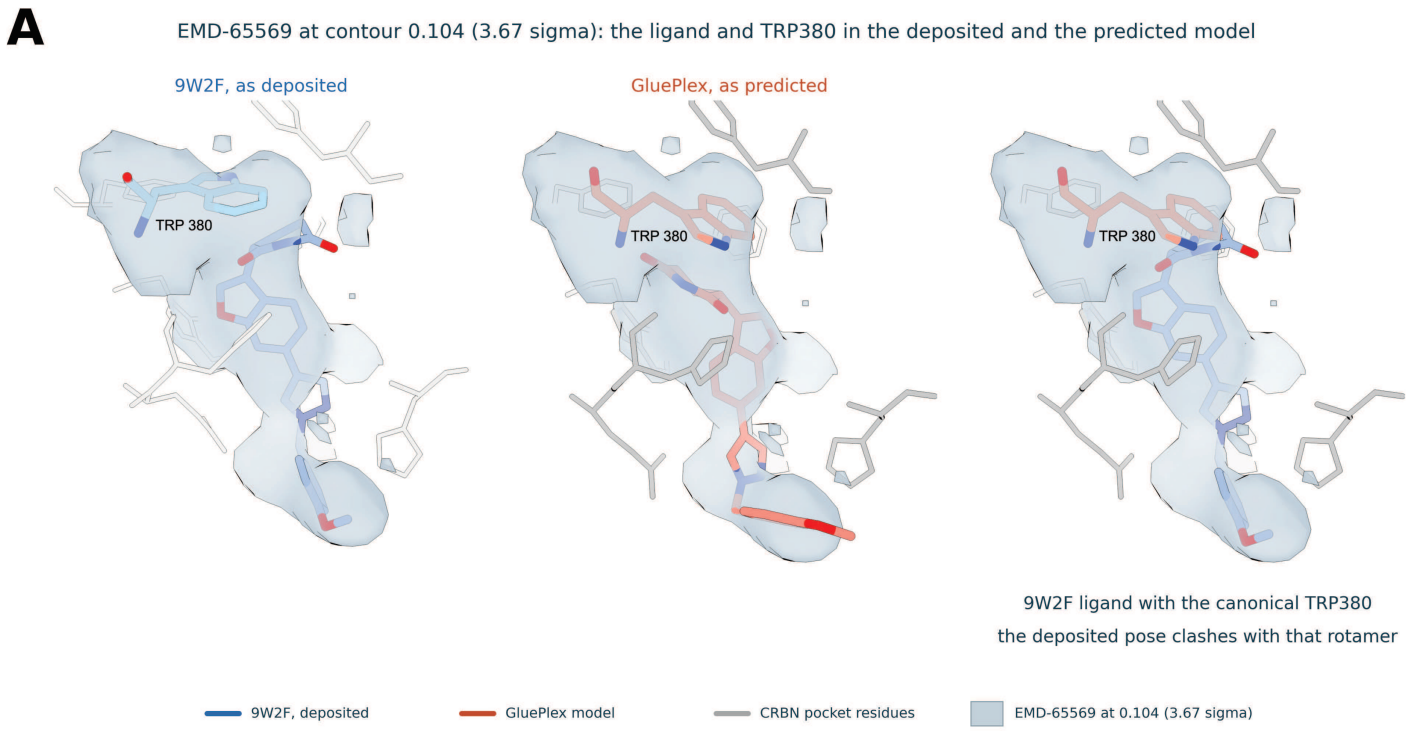
1. Békés, M., Langley, D. R. & Crews, C. M. PROTAC targeted protein degraders: the past is prologue. *Nat. Rev. Drug Discov.* **21**, 181–200 (2022).
2. Tsai, J. M., Nowak, R. P., Ebert, B. L. & Fischer, E. S. Targeted protein degradation: from mechanisms to clinic. *Nat. Rev. Mol. Cell Biol.* **25**, 740–757 (2024).
3. Oleinikovas, V., Gainza, P., Ryckmans, T., Fasching, B. & Thomä, N. H. From Thalidomide to Rational Molecular Glue Design for Targeted Protein Degradation. *Annu. Rev. Pharmacol. Toxicol.* **64**, 291–312 (2024).
4. Ito, T. *et al.* Identification of a primary target of thalidomide teratogenicity. *Science* **327**, 1345–1350 (2010).
5. Krönke, J. *et al.* Lenalidomide causes selective degradation of IKZF1 and IKZF3 in multiple myeloma cells. *Science* **343**, 301–305 (2014).
6. Lu, G. *et al.* The myeloma drug lenalidomide promotes the cereblon-dependent destruction of Ikaros proteins. *Science* **343**, 305–309 (2014).
7. Petzold, G., Fischer, E. S. & Thomä, N. H. Structural basis of lenalidomide-induced CK1α degradation by the CRL4(CRBN) ubiquitin ligase. *Nature* **532**, 127–130 (2016).
8. Matyskiela, M. E. *et al.* A novel cereblon modulator recruits GSPT1 to the CRL4CRBN ubiquitin ligase. *Nature* **535**, 252–257 (2016).
9. Sievers, Q. L. *et al.* Defining the human C2H2 zinc finger degrome targeted by thalidomide analogs through CRBN. *Science* **362**, eaat0572 (2018).
10. Słabicki, M. *et al.* Expanding the druggable zinc-finger proteome defines properties of drug-induced degradation. *Mol. Cell* **85**, 3184-3201.e14 (2025).
11. Galli, P. *et al.* Proteome-wide identification of the druggable CRBN interactome. *Nat. Biotechnol.* 1–10 (2026) doi:10.1038/s41587-026-03237-7.
12. Kozicka, Z. & Thomä, N. H. Haven't got a glue: Protein surface variation for the design of molecular glue degraders. *Cell Chem. Biol.* **28**, 1032–1047 (2021).
13. Hanzl, A., Inghelram, C., Schmitt, S. & Thomä, N. H. Primed for degradation: How weak protein interactions enable molecular glue degraders. *Curr. Opin. Struct. Biol.* **92**, 103052 (2025).

14. Dudas, B., Athanasiou, C., Mobarec, J. C. & Rosta, E. Quantifying Cooperativity through Binding Free Energies in Molecular Glue Degraders. *J. Chem. Theory Comput.* https://doi.org/10.1021/acs.jctc.5c00064 (2025) doi:10.1021/acs.jctc.5c00064.
15. Słabicki, M. *et al.* The CDK inhibitor CR8 acts as a molecular glue degrader that depletes cyclin K. *Nature* **585**, 293–297 (2020).
16. Han, T. *et al.* Anticancer sulfonamides target splicing by inducing RBM39 degradation via recruitment to DCAF15. *Science* **356**, eaal3755 (2017).
17. Ting, P. Y. *et al.* A molecular glue degrader of the WIZ transcription factor for fetal hemoglobin induction. *Science* **385**, 91–99 (2024).
18. Mayor-Ruiz, C. *et al.* Rational discovery of molecular glue degraders via scalable chemical profiling. *Nat. Chem. Biol.* **16**, 1199–1207 (2020).
19. Steger, M. *et al.* Unbiased mapping of cereblon neosubstrate landscape by high-throughput proteomics. *Nat. Commun.* **16**, 7773 (2025).
20. Petzold, G. *et al.* Mining the CRBN target space redefines rules for molecular glue–induced neosubstrate recognition. *Science* **389**, eadt6736 (2025).
21. Lin, H. *et al.* Leveraging high-throughput proteomics and AI-based protein folding to accelerate VAV1 molecular glue discovery. *Nat. Commun.* https://doi.org/10.1038/s41467-026-77657-z (2026) doi:10.1038/s41467-026-77657-z.
22. Wang, B., Cao, S. & Zheng, N. Emerging strategies for prospective discovery of molecular glue degraders. *Curr. Opin. Struct. Biol.* **86**, 102811 (2024).
23. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
24. Abramson, J. *et al.* Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature* **630**, 493–500 (2024).
25. Wohlwend, J. *et al.* Boltz-1 Democratizing Biomolecular Interaction Modeling. 2024.11.19.624167 Preprint at https://doi.org/10.1101/2024.11.19.624167 (2025).
26. Passaro, S. *et al.* Boltz-2: Towards Accurate and Efficient Binding Affinity Prediction. 2025.06.14.659707 Preprint at https://doi.org/10.1101/2025.06.14.659707 (2025).
27. Discovery, C. *et al.* Chai-1: Decoding the molecular interactions of life. 2024.10.10.615955 Preprint at https://doi.org/10.1101/2024.10.10.615955 (2024).

28. Team, P. *et al.* Protenix-v1: Toward High-Accuracy Open-Source Biomolecular Structure Prediction. 2026.02.05.703733 Preprint at https://doi.org/10.64898/2026.02.05.703733 (2026).
29. Liao, Y., Zhu, J., Xie, J., Lai, L. & Pei, J. Benchmarking Cofolding Methods for Molecular Glue Ternary Structure Prediction. *J. Chem. Inf. Model.* **65**, 11136–11148 (2025).
30. Che, X. & Team, T. S. M. I. YDS-GlueFold: Surpassing AlphaFold 3-Type Models for Molecular Glue-Induced Ternary Complex Prediction. 2024.12.23.630090 Preprint at https://doi.org/10.1101/2024.12.23.630090 (2025).
31. Buttenschoen, M., Morris, G. M. & Deane, C. M. PoseBusters: AI-based docking methods fail to generate physically valid poses or generalise to novel sequences. *Chem. Sci.* **15**, 3130–3139 (2024).
32. Pei, Y. *et al.* Selective CDK2 Degradation via Noncanonical Recruitment. *J. Med. Chem.* **69**, 15996–16024 (2026).
33. Conery, A. R. *et al.* TRI-611, a selective, brain-penetrant molecular glue degrader of ALK. *Nature* 1–9 (2026) doi:10.1038/s41586-026-10998-3.
34. Annunziato, S. *et al.* Cereblon induces G3BP2 neosubstrate degradation using molecular surface mimicry. *Nat. Struct. Mol. Biol.* **33**, 479–487 (2026).
35. Ojeda, S. *et al.* Degron-independent recruitment of KAT2A expands the target space of CRBN molecular glues. *Science* **393**, 188–194 (2026).
36. Lu, X. *et al.* Molecular glue degraders of HuR suppress BRAF-mutant colorectal cancer. *Nature* **655**, 780–789 (2026).
37. Nicholson, H. E. *et al.* TNG961 Is a Selective Oral HBS1L Molecular Glue Degrader for the Treatment of FOCAD-Deleted Cancers. *Cancer Discov.* **16**, 1884–1902 (2026).
38. Krapp, L. F., Abriata, L. A., Cortés Rodriguez, F. & Dal Peraro, M. PeSTo: parameter-free geometric deep learning for accurate prediction of protein binding interfaces. *Nat. Commun.* **14**, 2175 (2023).
39. Honorato, R. V. *et al.* The HADDOCK2.4 web server for integrative modeling of biomolecular complexes. *Nat. Protoc.* **19**, 3219–3241 (2024).
40. Giulini, M. *et al.* HADDOCK3: A Modular and Versatile Platform for Integrative Modeling of Biomolecular Complexes. *J. Chem. Inf. Model.* **65**, 7315–7324 (2025).
41. Basu, S. & Wallner, B. DockQ: A Quality Measure for Protein-Protein Docking Models. *PLOS ONE* **11**, e0161879 (2016).

42. Méndez, R., Leplae, R., De Maria, L. & Wodak, S. J. Assessment of blind predictions of protein-protein interactions: current status of docking methods. *Proteins* **52**, 51–67 (2003).
43. Cao, S. *et al.* Defining molecular glues with a dual-nanobody cannabidiol sensor. *Nat. Commun.* **13**, 815 (2022).
44. Shtrikman, A. *et al.* Co-folding model guided by structural proteomics. Preprint at https://doi.org/10.48550/arXiv.2605.26192 (2026).
45. Steinegger, M. & Söding, J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. *Nat. Biotechnol.* **35**, 1026–1028 (2017).
46. Hubbard, S. J. & Thornton, J. M. NACCESS, Computer Program. Department of Biochemistry and Molecular Biology, University College London (1993).
47. Watson, E. R. *et al.* Molecular glue CELMoD compounds are regulators of cereblon conformation. *Science* **378**, 549–553 (2022).
48. Herre, C. *et al.* Introduction of the Capsules environment to support further growth of the SBGrid structural biology software collection. *Acta Crystallogr. Sect. Struct. Biol.* **80**, 439–450 (2024).

Supplementary Figures



B

map-model correlation

each cell is 9W2F / GluePlex, the higher value in bold

measured on	full map	half map 1	half map 2
ligand alone	**0.551** / 0.516	0.703 / **0.721**	0.719 / **0.725**
TRP380 side chain alone	0.452 / **0.522**	0.546 / **0.548**	**0.569** / 0.567
ligand + TRP380 side chain	0.556 / **0.562**	0.686 / **0.704**	0.703 / **0.710**
ligand + TRP380 whole residue	0.563 / **0.585**	0.690 / **0.711**	0.713 / **0.725**
ligand + tri-Trp cage	0.576 / **0.580**	0.698 / **0.709**	0.699 / **0.704**

Figure S1. Alternative ligand–Trp380 arrangements in the ELAVL1–CRBN complex.

(A) The deposited 9W2F model (blue), the GluePlex prediction (red), and a hybrid overlay combining the deposited ligand with predicted Trp380. The same EMD-65569 density surface is shown in all three panels at contour 0.104 (3.67 σ). Other CRBN pocket residues are shown as thin neutral-colored sticks. The hybrid overlay illustrates the incompatibility of the deposited ligand with the predicted Trp380 canonical rotamer. The GluePlex model is the ELAVL1 representative shown in **Fig. 2B**, selected by the lowest neosubstrate chain RMSD among the top ten clusters across three runs.

(B) Map–model correlations for the deposited and predicted structures against the full map and two half maps. Density was simulated at 3.4 Å, and correlations were calculated over voxels occupied by the simulated density, independently of the display contour in panel A. Each cell lists deposited/predicted values, with the higher value in bold. Rows evaluate the ligand alone, Trp380 side chain alone, ligand plus Trp380 side chain, ligand plus the entire Trp380 residue, and ligand plus the tri-tryptophan pocket. Separate ligand and side-chain comparisons show mixed preferences across maps; all three combined selections favor the prediction in all three maps.

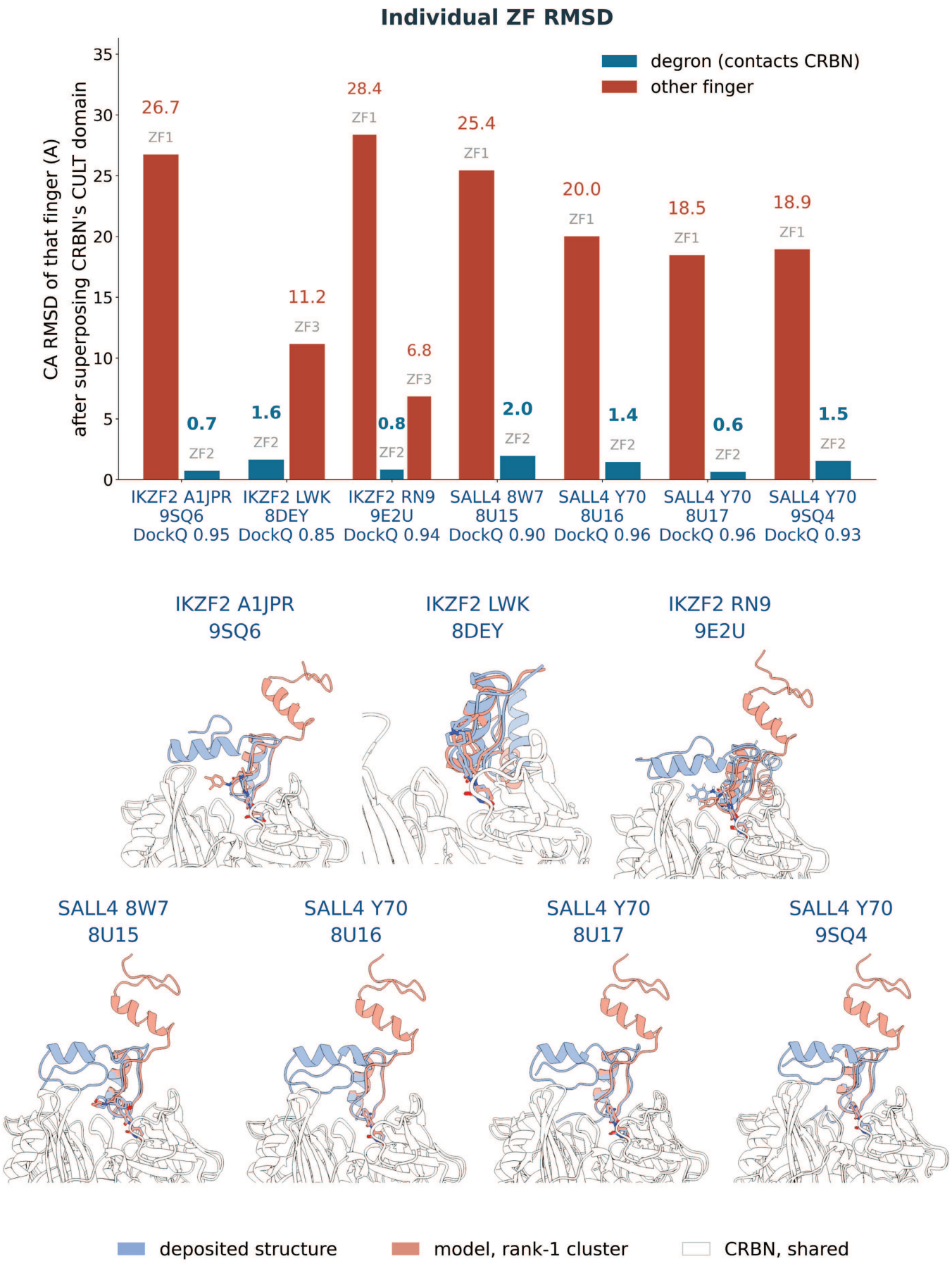


Figure S2. Accurate degron placement does not extend to flanking zinc fingers.

Per-finger displacement in seven multi-zinc-finger structures from two targets, measured as the median over the three highest-ranked clusters of the default configuration. Bars show individual fingers, colored by CRBN contact in the reference and labeled with displacement. Rank-1 DockQ is shown below each pair. IKZF2–RN9 (9E2U) contains three fingers. Overlays align each prediction to its reference through CRBN CULT and place all seven complexes in a common CRBN frame. N = 1 for this experiment.

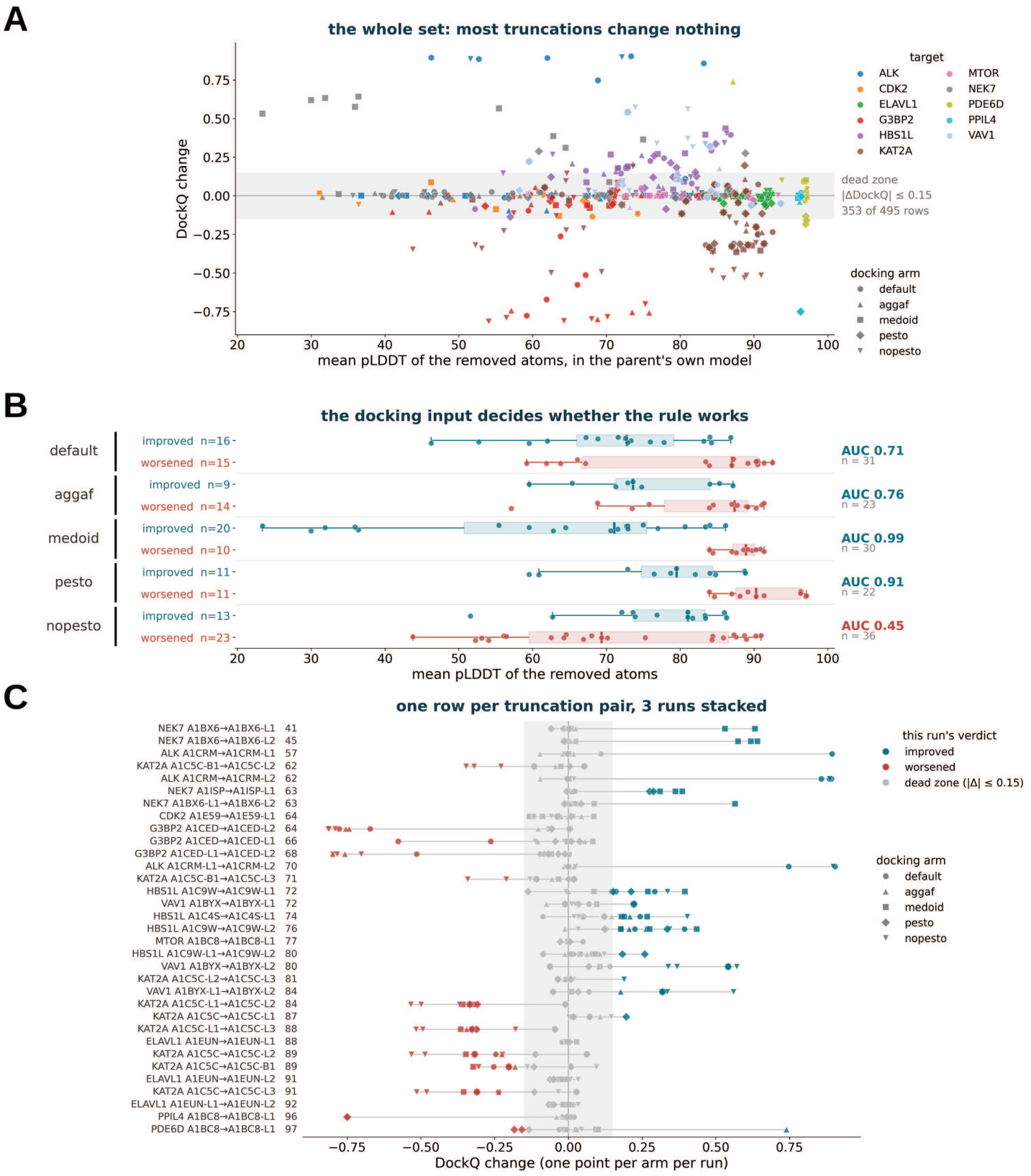


Figure S3. Truncation outcome against the confidence of the moiety removed, over three independent runs of five docking configurations.

Each point represents one of 33 truncation pairs in one of five docking configurations and three independent runs. **(A)** ΔDockQ versus mean pLDDT of the removed atoms in the parent model, with the dead zone shaded. Marker shape

indicates configuration and color indicates target. **(B)** Removed-moiety pLDDT for improved and worsened observations, separated by configuration, with corresponding ROC AUC values. **(C)** Three stacked runs per truncation pair, ordered by removed-moiety pLDDT (shown after the name).

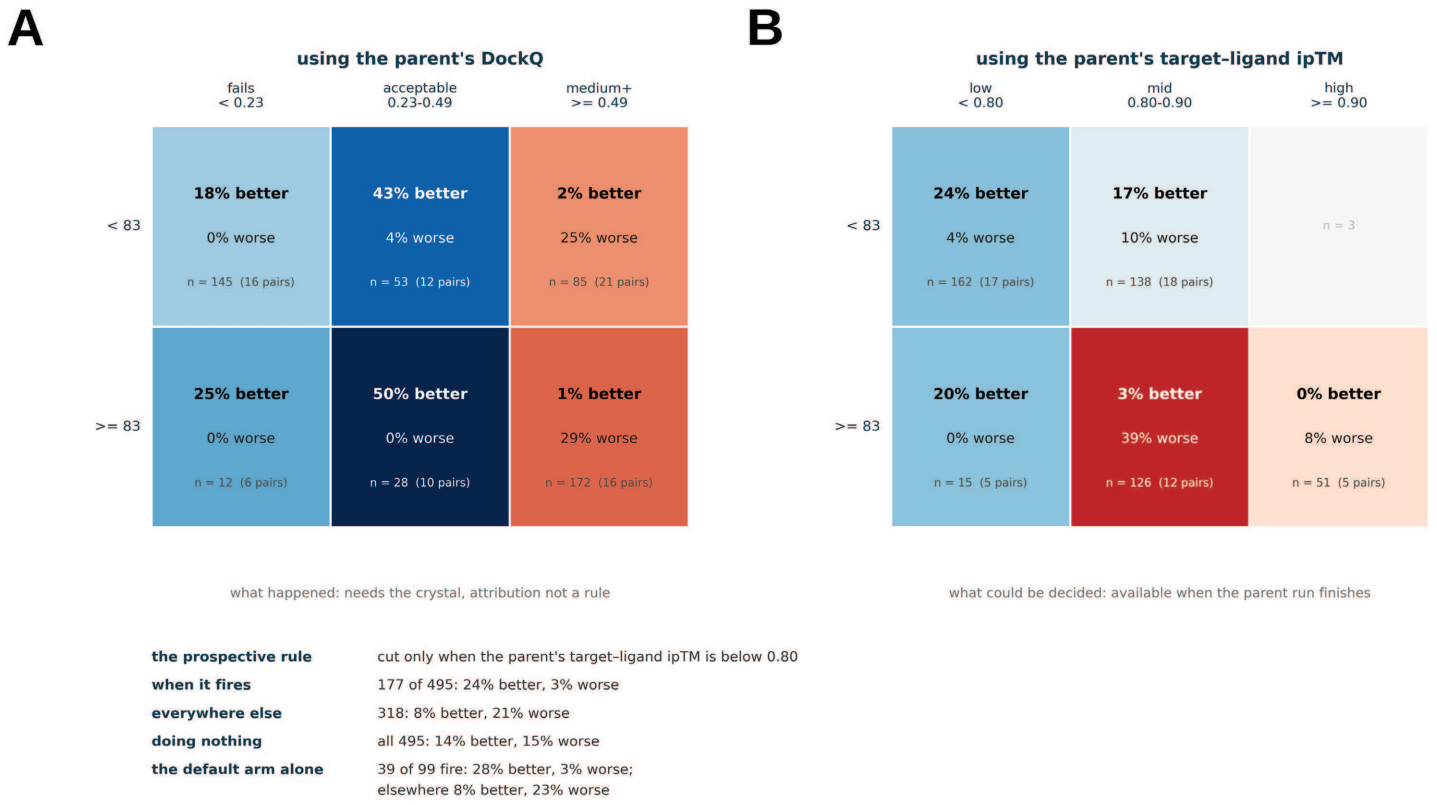


Figure S4. Truncation outcome against the parent model's confidence, over three independent runs of five docking configurations.

Both grids show 495 observations: 33 truncation pairs × five docking configurations × three runs. ΔDockQ is the truncation score minus its parent's score in the same run; improvement and loss require changes beyond ±0.15. Cell color shows the percentage improved minus the percentage worsened. Labels report both percentages, observation counts and distinct pair counts. Both grids split removed-moiety pLDDT at 83. **(A)** Outcomes grouped by run-specific parent DockQ and removed-moiety pLDDT. DockQ requires the reference structure. **(B)** Outcomes grouped by median parent target–ligand chain-pair ipTM and removed-moiety pLDDT across three runs per configuration. The table evaluates truncation when parent ipTM < 0.80, compared with unselected observations, unconditional truncation and the default configuration alone.

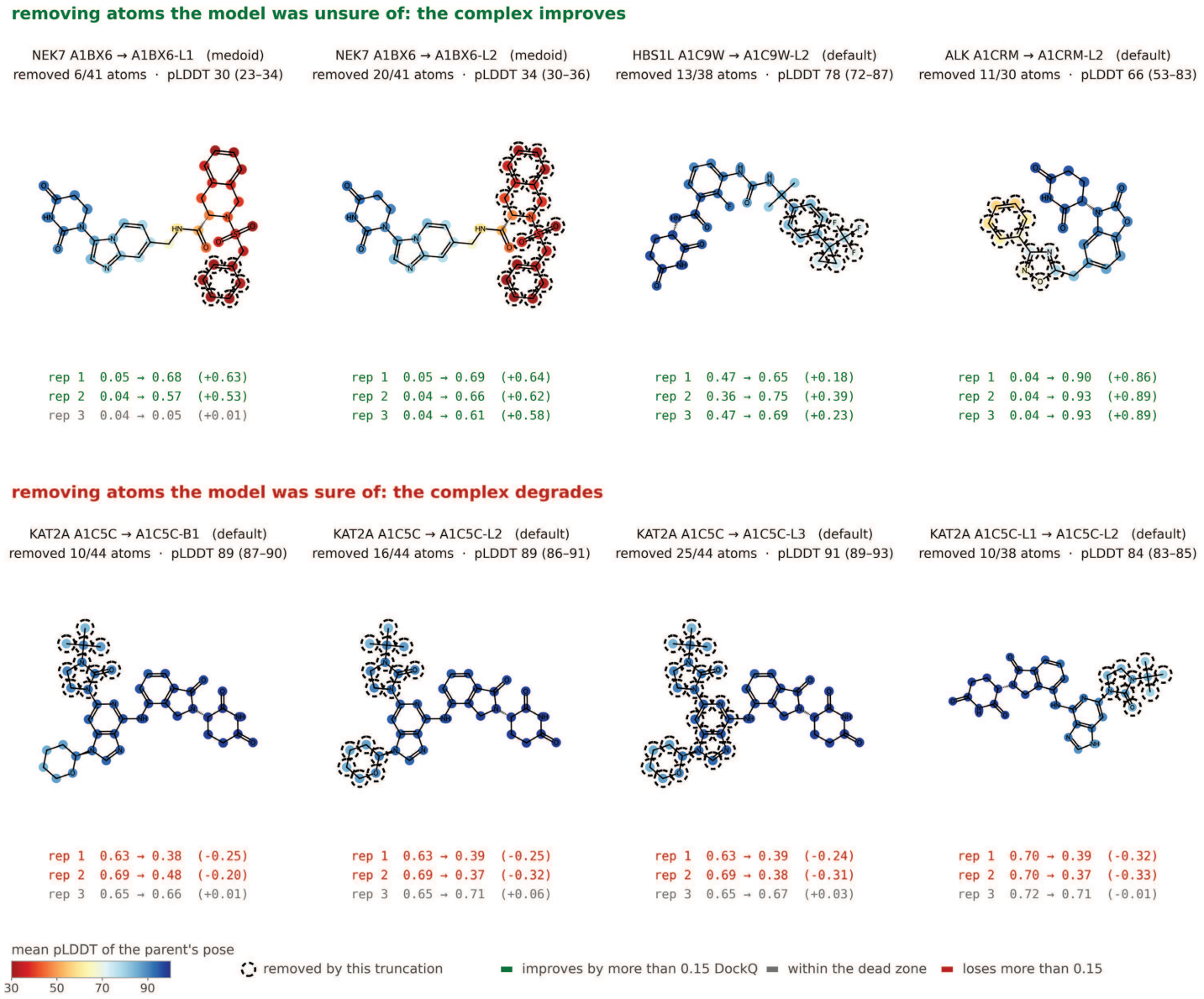


Figure S5. Atoms removed by each truncation, and reproducibility of the outcome across independent runs.

Eight truncations are shown with removed atoms colored by parent-model per-atom pLDDT averaged across runs; brackets give the range. DockQ values for all three runs appear beneath each panel to show variation in outcome. Runs 1–3 differ by seed.



Figure S6. Four docked conformers under a single interface prior, across three independent runs.

The four conformer inputs from **Fig. 7**. Only the docked coordinates vary; all inputs use the PeSTo union prior and identical compound, reference and scoring. Each compound has three stacked bands for independent runs. Columns show up to ten clusters ranked by representative chain-pair ipTM, followed by the whole-ensemble oracle. Cell color indicates DockQ CAPRI tier within each input's color scale; values give neosubstrate chain RMSD (Å). Amber outlines mark the first cluster with DockQ ≥ 0.23. Summaries below each panel report acceptable and medium-or-better counts, median DockQ and rank-1 successes across 81 runs. The final input is the reference-bound target conformer.