Has AlphaFold 3 Solved the Protein Folding Problem for D-Peptides?

Henry Childs[1, 2], Pei Zhou[3], and Bruce R. Donald[1, 2, 3, 4, *]

[1]Department of Chemistry, Duke University, 124 Science Drive, Durham, NC 27708, United States
[2]Department of Computer Science, Duke University, 308 Research Drive, Durham, NC 27708, United States
[3]Department of Biochemistry, Duke University School of Medicine, 307 Research Drive, Durham, NC 22710, United States
[4]Department of Mathematics, Duke University, 120 Science Drive, Durham, NC 27708, United States
[*]Correspondence: brd+bioRxiv25@cs.duke.edu

Abstract

Due to the favorable chemical properties of mirrored chiral centers (such as improved stability, bioavailability, and membrane permeability) the computational design of D-peptides targeting biological L-proteins is a valuable area of research. To design these structures *in silico*, a computational workflow should correctly dock and fold a peptide while maintaining chiral centers. The latest AlphaFold 3 (AF3) from Abramson et al. (2024) enforces a strict chiral violation penalty to maintain chiral centers from model inputs and is reported to have a low chiral violation rate of only 4.4% on a PoseBusters benchmark containing diverse chiral molecules. Herein, we report the results of 3,255 experiments with AF3 to evaluate its ability to predict the fold, chirality, and binding pose of D-peptides in heterochiral complexes. Despite our inputs specifying explicit D-stereocenters, we report that the AF3 chiral violation for D-peptide binders is much higher at 51% across all evaluated predictions; on average the model is as accurate as chance (random chirality choice, L or D, for each peptide residue). Increasing the number of seeds failed to improve this violation rate. The AF3 predictions exhibit incorrect folds and binding poses, with D-peptides commonly oriented incorrectly in the L-protein binding pocket. Confidence metrics returned by AF3 also fail to distinguish predictions with low chirality violation and correct docking vs. predictions with high chirality violation and incorrect docking. We conclude that AF3 is a poor predictor of D-peptide chirality, fold, and binding pose.

Summary

A crucial task in computational protein design is predicting fold, as this property determines the structure and function of a protein. Abramson et al.[1] published in *Nature* on AlphaFold 3 (AF3), a powerful deep learning framework for predicting chemical structures in both bound and unbound states. This architecture is tuned to respect chiral centers, which are atoms (in proteins, backbone α-carbons) covalently bound to four different chemical species[2]. These centers adopt two non-superposable forms, often called "handedness," termed L (all biological proteins adopt this form) and D (the mirror image of L). L and D chiral centers exert significant influence on chemical function; changing the chirality of even a single residue can dramatically alter chemical properties

such as enantioselective binding (e.g., antifolate resistance[3]) and stability[4]. Additionally, D-peptides (small proteins containing exclusively D chiral centers) exhibit many advantages compared to their L-peptide counterparts, such as protease evasion[5], and are therefore therapeutically relevant modalities. Due to vastly differing chemical properties, an algorithm should respect chiral center inputs and exhibit an error rate of 0%. Although Abramson et al.[1] reports a low 4.4% chirality violation across diverse chiral centers, we have found that the chiral violation rate for D-peptides with D chiral center inputs explicitly specified is much higher at 51%. Increasing the number of seeds fails to improve this rate. Our data highlights a crucial structural prediction error in AF3 and demonstrates the model is as accurate on average as chance (random chirality choice, L or D, for each peptide residue). Compared to empirical structures, AF3 is also highly inaccurate when folding and docking D-peptide:L-protein complexes. The failure of AF3 to accurately predict these chemical interactions indicates more work is need for high-quality prediction of D-peptides.

1 Introduction

The design of high-affinity peptides with the mirror handedness of canonical L-peptides is an ongoing, high-value area of research[6–9]. D-peptides, comprised of D-amino acids that are the mirror images of L-amino acids, offer unique therapeutic advantages. In particular, D-peptides improve peptide stability by evading degradation by proteolytic enzymes[5] and have been shown to have improved bioavailability and membrane permeability compared to their L counterparts[4,10]. To efficiently design D-peptide binders, a software workflow must accurately respect chirality (stereocenter configuration), determine the fold (ligand and target conformation), and predict the binding pose (ligand docking and orientation in the binding pocket). Herein, we report that AlphaFold 3 (AF3) fails to accurately predict each of these chemical features of D-peptides in complex with L-proteins.

Due to the importance of correct stereochemistry, many machine learning architectures implement a strict *chirality violation* penalty: model ranking scores are penalized for predictions with incorrect chirality. A model that contains incorrect stereocenters is unable to be reliably integrated into drug discovery pipelines, particularly for D-peptide binders. Due to the importance of an *in silico* framework to correctly predict stereocenters, the latest AF3 model includes this penalty for chirality violation; in a PoseBusters benchmark, AF3 produces[1] a chirality violation rate of 4.4%. However, we report that AF3's chirality violation specifically for D-peptide chiral centers is significantly higher at an average of 51% across all evaluated predictions of D-peptide:L-protein complexes. This is a shockingly high number, especially as model inputs contained explicit D-chirality specification for each stereocenter. This rate, on par with a random coin flip, is unacceptable for a drug discovery pipeline and indicates that structural prediction of D-peptides is an unresolved research problem. Due to widespread use of this model, the precise shortcomings of AF3 on D-peptides warrant a thorough investigation to diagnose the capability of this framework on D-peptides.

Similar to prior publications reviewing AlphaFold accuracy on specific biochemical systems[11–13], we demonstrate this chirality violation with three crystal structures of D-peptide:L-target, one apo D-protein, and nine synthetic D-peptide:L-protein complexes. Although many Protein Data Bank[14] (PDB) entries contain D-amino acids, we discovered only ten deposited crystal structures of D-peptides in complex with an L-protein. Of these exceptionally few D-peptide:L-protein crystal structures in the PDB, we selected three systems based on criteria of high resolution (1.61 Å to 2.20 Å) and chemical diversity (fold, binding, and peptide length). Tested systems were DP19:L-19437 (PDB ID 7YH8), DP9:Streptavidin (PDB ID 5N8T), and DP12:MDM2 (PDB ID 3IWY) where DPX refers to a D-

peptide of length X. See SI Table 2 for full sequences. All D-peptides were input to AlphaFold 3 with the stereocenters explicitly specified: in the ideal case, a prediction algorithm should never change even a single residue to the opposite chirality. In our experiments, however, we observed high chirality violation for all evaluated complexes, which are reported in Table 1. Top-ranked predictions generated using higher seed volume yielded no improvement in chirality violation rate.

By exploiting properties of mirror symmetry[15], we also queried AF3 to predict the fold of an apo D-space SH3 domain (PDB ID 1SHG, reflected). The fold predicted by AF3 was not accurate and had a chirality violation rate of 44%. We additionally tested seven synthetic systems, where D-peptides were either (1) placed in complex with a decoy target to investigate chirality violation patterns, or (2) fluorinated, where each α-hydrogen in the ligand was replaced with fluorine to evaluate stereocenters with non-hydrogen atoms. For computational experiments (1) and (2), AF3 still produced peptides with high chirality violation (average of 44% for (1) and 33% for (2), see SI Section 4.2). In total, we queried AF3 with 13 systems to produce $3,255$ predictions.

2 Results

2.1 Chirality Violation for Retrospective Prediction of Crystal Structures

D-space chiral centers were explicitly specified in the input to AF3; these stereocenters should never have been different in the prediction outputs. Despite this proscriptive input, AF3 generates ligands with high chirality violation (average 51%, see Table 1) across all diffusion samples ("sample" refers to a structure prediction for the corresponding seed. Each seed produces five samples.) for DP19, DP9, and DP12 (see Figure 1 for a visualization of this chirality violation). The most accurate samples had a chirality violation rate of 37%, 33%, and 17% for DP19, DP9, and DP12, respectively. The least accurate samples had a chirality violation rate of 63%, 67%, and 83% for DP19, DP9, and DP12, respectively. The average chirality violation was 52% for DP19, 51% for DP9, and 50% for DP12. No single residue was predicted in the correct D chirality for all samples (0-4) in a given system. While some residue violation rates were lower in select systems (e.g., DP9 residue 1 is only predicted in the incorrect L chirality in the top-ranked sample 4; see Table 1), each system displays inconsistent patterns of chirality violation across the five samples for each residue. Additionally, no samples in any system contained a chirality error-free ligand, meaning the per-peptide chirality violation across these 15 samples is 100%. We also reviewed AF3's poor performance on D-peptide fold and binding pose (visualized in Figures 2 - 4). Although AF3 accurately predicts the L-target, it largely fails to correctly predict the fold and binding pose for DP19, DP9, and DP12. To assess if chiral violation was due to the potential inclusion of a hydrogen stripping module in AF3 input processing, we additionally specified inputs with another small atom, fluorine, substituted for each α-hydrogen. The average chirality violation rate was 33% across all systems (see SI Section 4.2).

System	Res 1	Res 2	Res 3	Res 4	Res 5	Res 6	Res 7	Res 8	Res 9	Res 10	Res 11	Res 12	Res 13	Res 14	Res 15	Res 16	Res 17	Res 18	Res 19	Chirality Violation Rate
DP19 Sample 0	L	D	D	D	D	D	D	D	D	Planar	L	L	L	L	L	L	L	L	L	58%
DP19 Sample 1	D	D	D	L	D	D	L	L	L	L	L	D	D	D	D	D	D	D	L	37%
DP19 Sample 2	L	D	L	D	D	D	D	D	D	L	L	L	L	L	L	L	L	L	L	63%
DP19 Sample 3*	D	L	Planar	L	L	L	L	L	L	D	D	D	D	D	D	D	D	D	L	47%
DP19 Sample 4	D	D	D	D	D	D	D	D	L	L	L	L	Planar	L	L	L	L	L	D	53%
DP9 Sample 0	D	L	D	Planar	D	D	L	D	D											33%
DP9 Sample 1	D	L	D	D	D	D	D	L	L											33%
DP9 Sample 2	D	Planar	L	L	L	L	D	L	D											67%
DP9 Sample 3	D	D	L	L	L	L	L	D	L											67%
DP9 Sample 4*	L	D	L	L	D	L	L	D	D											56%
DP12 Sample 0	D	L	L	D	D	L	L	D	D	D	D	L								42%
DP12 Sample 1	L	L	L	D	L	D	L	L	L	L	L	L								83%
DP12 Sample 2	D	L	L	D	D	D	D	D	D	D	D	D								17%
DP12 Sample 3*	D	D	D	L	L	L	L	D	D	D	D	D								33%
DP12 Sample 4	L	L	L	D	D	D	L	L	L	L	L	L								75%

Table 1: Per-residue chirality violation for AlphaFold 3 predictions of D-peptides DP19, DP9, and DP12 in complex with L-target. Chirality for the L-targets (L-19437, Streptavidin, and MDM2, respectively) are not shown in this table. The System column includes the D-peptide name and sample number ("sample" is an AF3 term that labels a prediction for a specific seed). Each seed produces five samples. Each prediction was run with one seed, so five samples are shown. Top-ranked predictions are starred (*). Entries under the Res # columns indicate the chirality of that residue index (D (correct) or L (incorrect)). Entries of "Planar" indicate that AlphaFold 3 incorrectly predicted a flattened protein backbone that could not be assigned to either chirality (see Figure 1 for a visualization of chiral centers). Correct D stereocenters are in green while incorrect L and Planar stereocenters are in red. Chirality violation rate is calculated by dividing the number of residues with incorrect chirality (L or Planar) by the length of the peptide. DP19 exhibits the highest average chirality at 52% across all five seeds; DP9 has an average rate of 51% while DP12 has an average rate of 50%. The average chirality violation rate across all three systems is 51%. No AF3 prediction had a chirality violation rate even approaching 4.4%, the reported AF3 violation rate on the PoseBusters benchmark for chiral centers[1]. Further, no AF3 samples produced models with zero chiral errors, so the per-peptide chirality violation for these systems is 100%.

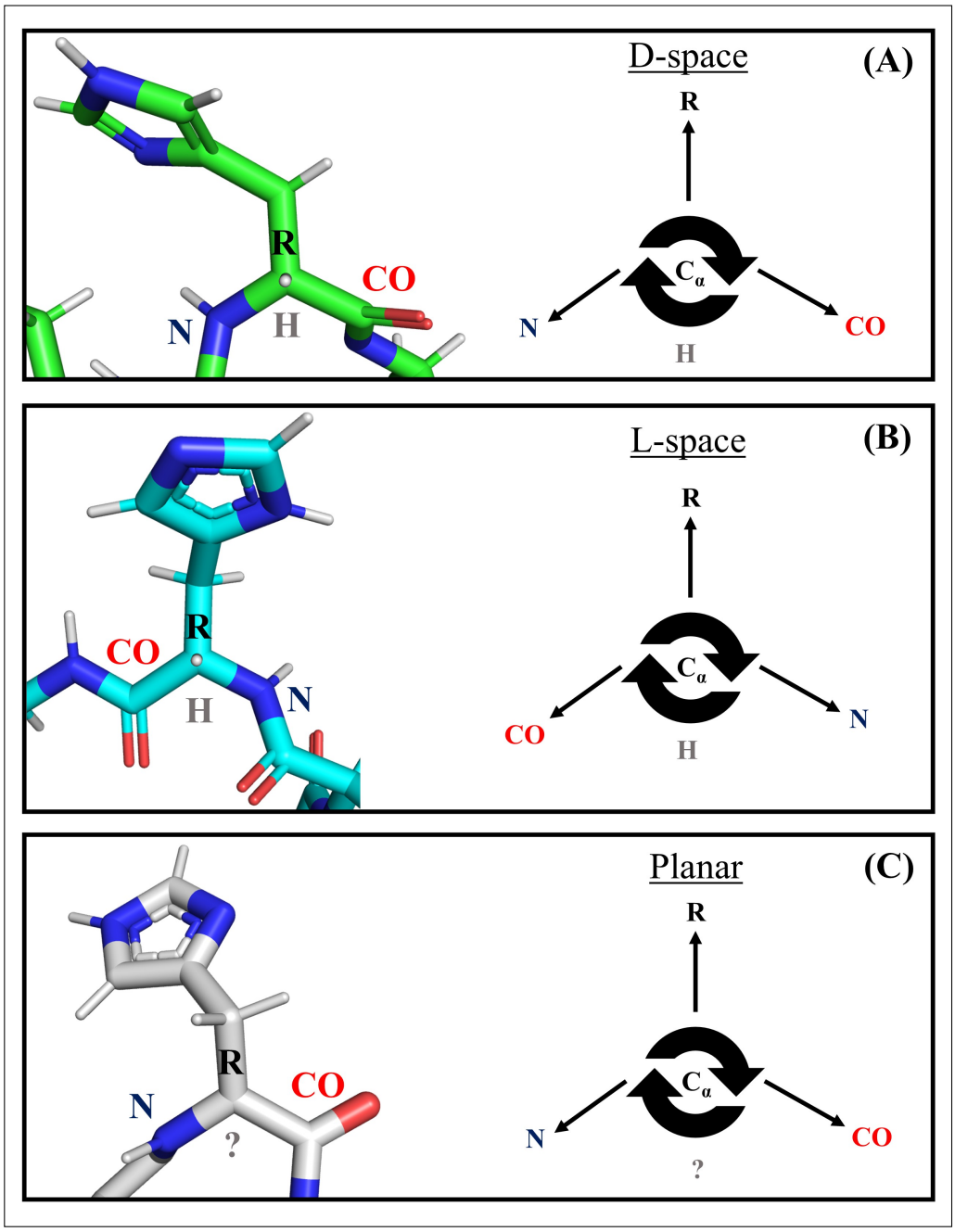

Figure 1: Comparison of DP19 residue His3 chirality between crystal structure (PDB ID 7YH8), AF3 sample 2, and AF3 sample 3. The description herein for L and D-space amino acids adheres to standard chemical definition[2]. **(A)**: the correct D-space residue from the crystal structure (green). **(B)**: the incorrect L-space stereocenter from AF3 sample 2 (cyan). **(C)**: the incorrect planar stereocenter from AF3 sample 3 (grey). In a correct D-space configuration (A), the backbone nitrogen, sidechain group (R), and backbone carboxyl are arranged clockwise relative to the α-hydrogen - α-carbon vector. Conversely, in the incorrect L-space configuration (B), the backbone carboxyl, sidechain, and backbone nitrogen are arranged in a clockwise direction relative to the α-hydrogen - α-carbon vector (the mirror image of D-space). In a planar stereocenter (C), the incorrect flattening of the protein backbone prevents accurate hydrogen placement: chirality can not be assigned by protonation software[16,17] or visual inspection. These planar stereocenters are not an accurate structural prediction.

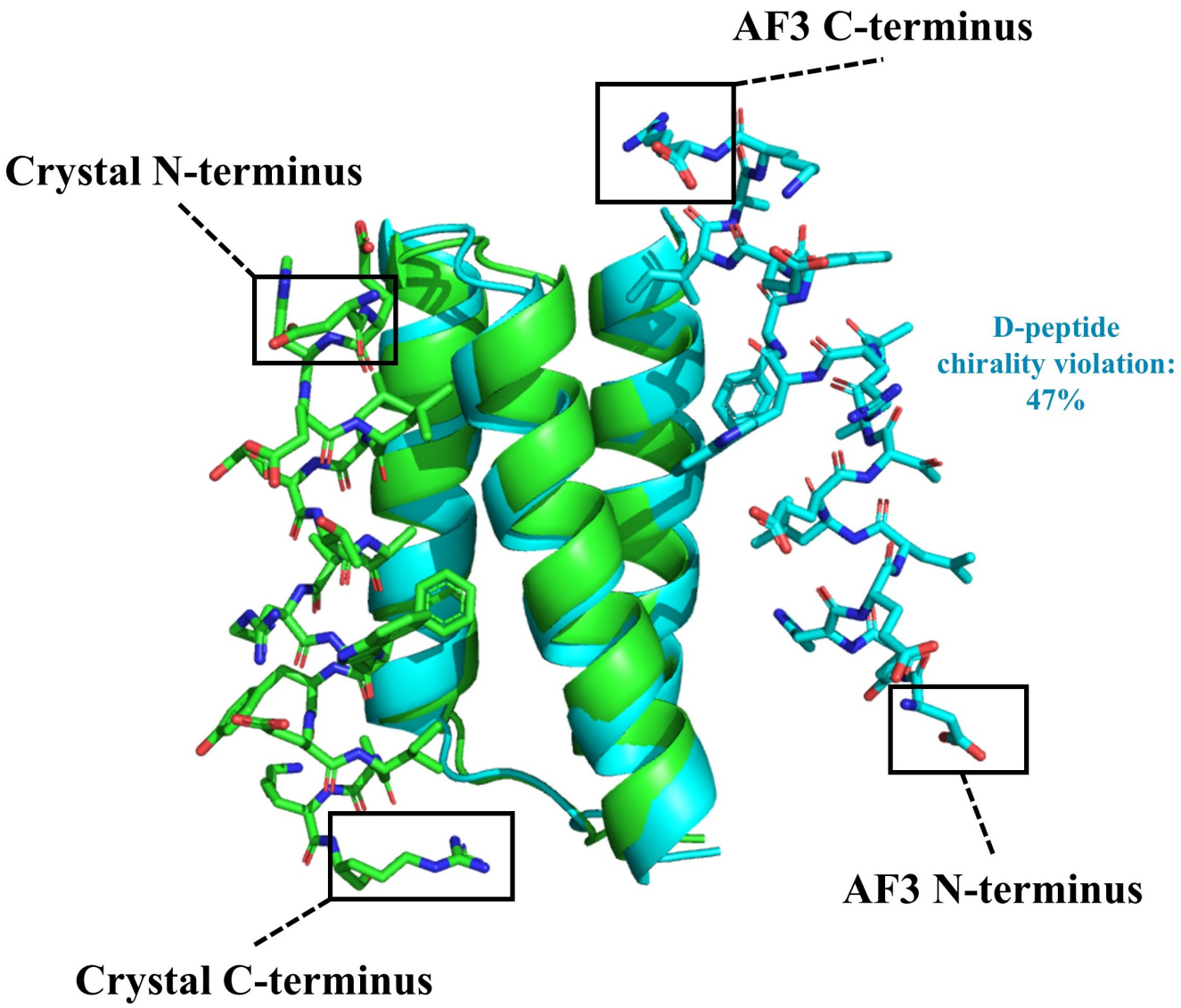

Figure 2: AF3 top-ranked DP19:L-19437 prediction (sample 3) incorrectly docks and orients the ligand compared to crystal structure (PDB ID 7YH8). The top-ranked sample is a representative prediction that demonstrates the poor performance of AF3 on DP19; for a visualization of all samples, refer to SI Section 4.4. The crystal structure is shown in green; the AF3 prediction is shown in cyan. Hydrogens are omitted for clarity. Although AF3 is generally successful predicting the fold of the L-target (cartoon fold), it fails to correctly fold and dock the D-peptide binder (sticks). AF3 both incorrectly places the D-ligand on the opposite side of the L-target and flips the ligand upside down relative to the target. The chirality violation rate for the AF3 top-ranked (sample 3) prediction of DP19 is 47%.

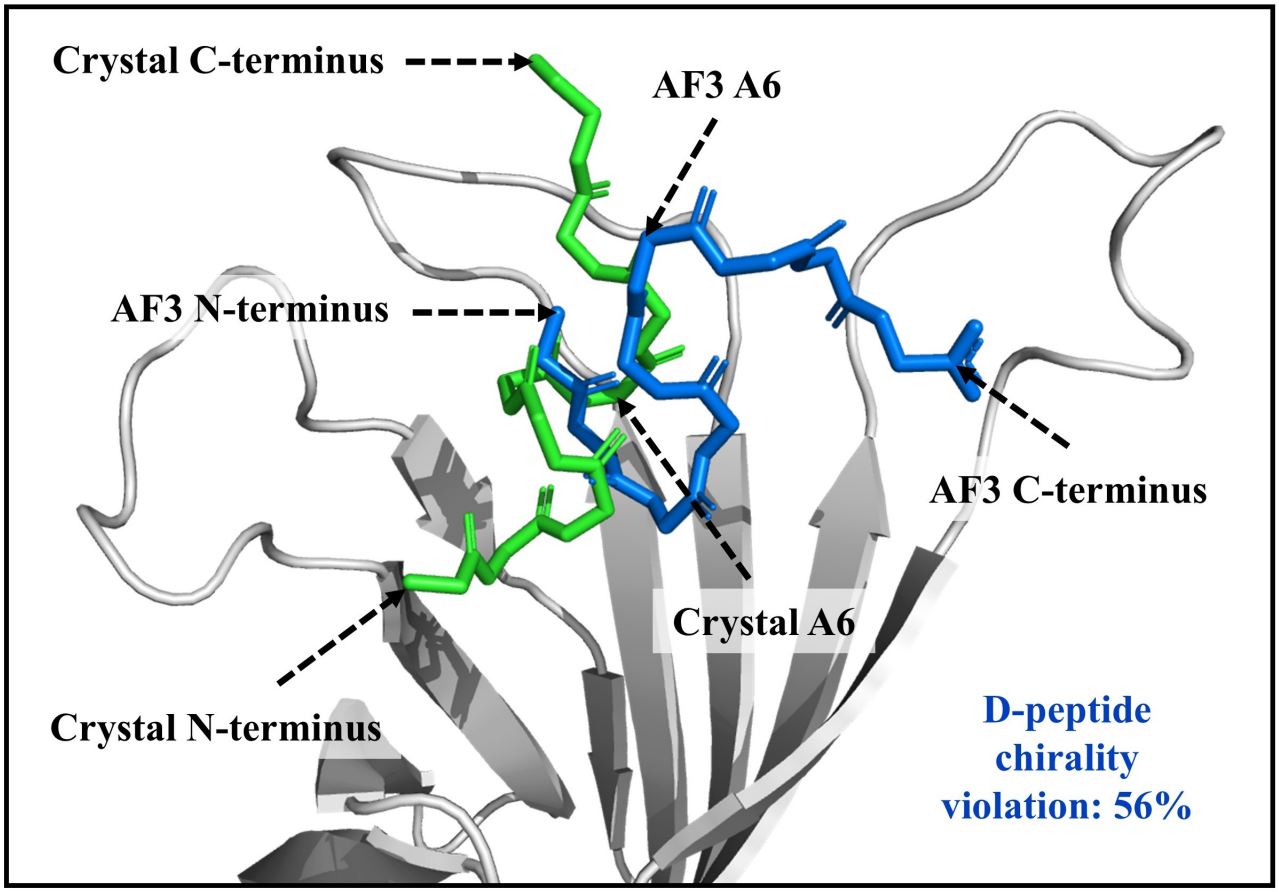

Figure 3: Comparison of DP9:Streptavidin crystal structure (PDB ID 5N8T) to AlphaFold 3 top-ranked (sample 4) prediction. The top-ranked sample is a representative prediction that demonstrates the poor performance of AF3 on DP9; for a visualization of all samples, refer to SI Section 4.4. This sample has a chirality violation rate of 56%. The crystal structure DP9 is in green sticks and the AF3 DP9 prediction is in blue sticks; sidechains are hidden for clarity. Streptavidin is shown in gray cartoon. Due to close agreement of the Streptavidin crystal structure and prediction (see SI Section 4.4), only the crystal structure of the L-target is shown. Dashed arrows indicate the N-term, C-term, and alanine at position 6 (A6) for both crystal structure and AF3 prediction. Due to the significant conformational deviations between DP9 crystal structure and prediction, AF3 predicts many unfavorable residue interactions that conflict with known tendencies of proteins. Two notable deviations occur at A6 and K9 (the C-term lysine). In the crystal structure, the charged C-term lysine is favorably oriented away from the hydrophobic β-barrel binding pocket and towards charged solvent. K9 likely does not directly interact with any Streptavidin residues (no L-target residues are within 4.0 Å of K9). However, AF3 orients K9 to the opposite side of the binding pocket, limiting favorable solvent exposure and placing the residue 3.6 Å from a nonpolar leucine at position 110 on Streptavidin (not shown). Similarly, AF3 incorrectly orients nonpolar A6 away from the hydrophobic pocket and towards charged solvent (no Streptavidin residues are $\leq$ 4.0 Å from A6). In the crystal structure, A6 is favorably buried 8.7 Å deeper into the pocket than the AF3 prediction A6, establishing inferred chemical contacts (by proximity, $\leq$ 3.5 Å) to W65, A72, and S74 on Streptavidin (not shown). The tendency of chemical residues to avoid or favor charged solvent is a key driver of protein behavior (including fold)[18,19]; therefore, the tendency of AF3 to predict unfavorable (and hence, unlikely) hydrophobic and hydrophilic interactions indicates a failure of the model to accurately predict D-peptides.

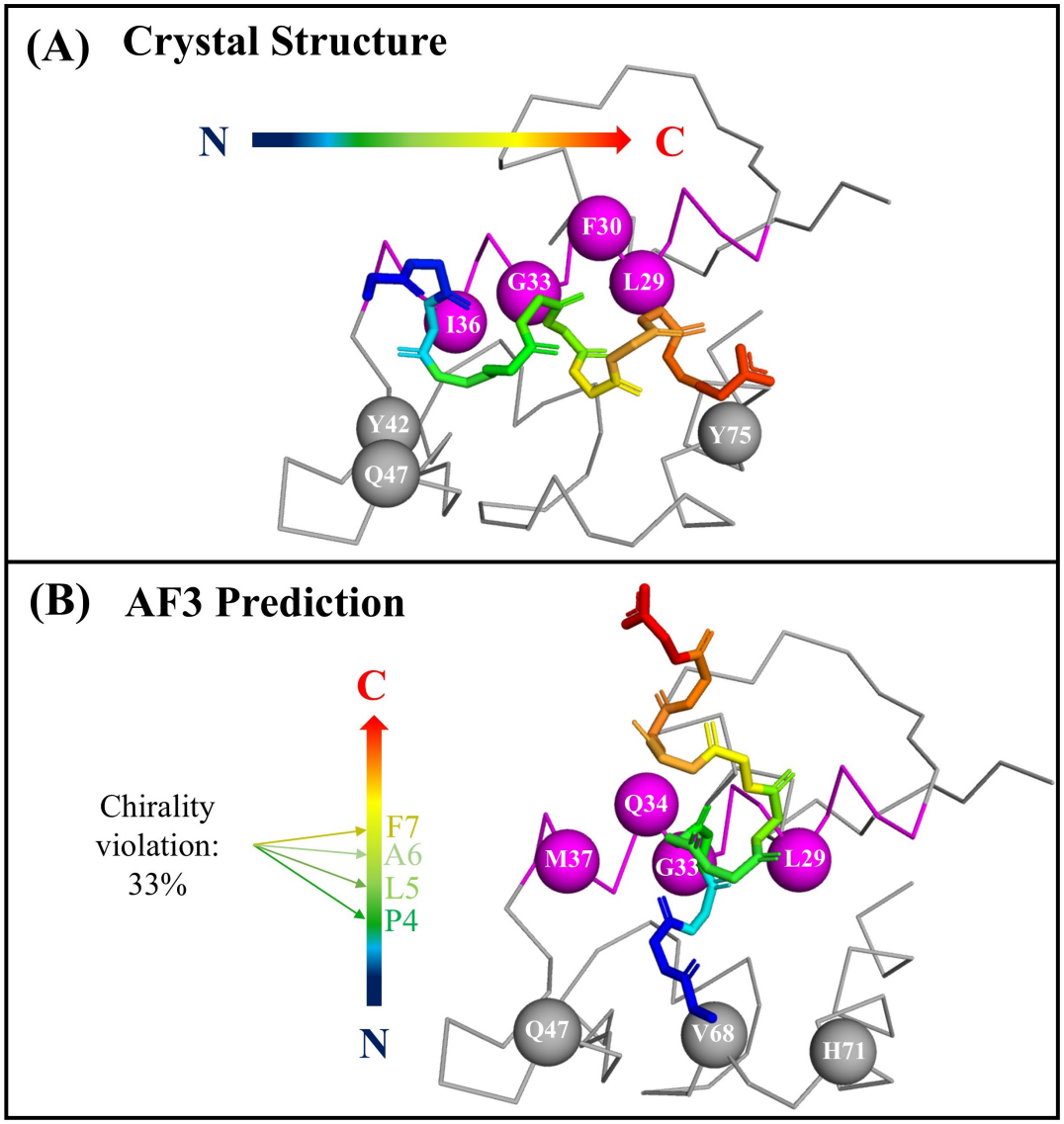

Figure 4: Comparison of DP12:MDM2 crystal structure (PDB ID 3IWY) to AlphaFold 3 top-ranked (sample 3) prediction. The top-ranked sample is a representative prediction that demonstrates the poor performance of AF3 on DP12; for a visualization of all samples, refer to SI Section 4.4. This sample has a chirality violation rate of 33%. In each frame, the D-peptide is shown in rainbow sticks; the N-terminus of the D-peptide is blue and the C-terminus is red. D-peptide sidechains are omitted for clarity. The L-target is represented as gray ribbon. MDM2 residues $\leq$ 3.5 Å from DP12 residues are noted with spheres. For reference, the second α-helix from the N-terminus of MDM2 (residues M25 to K39) is magenta. **(A)**: the crystal structure of DP12 in complex with MDM2. The D-peptide adopts a helix shape oriented antiparallel to the 2nd α-helix in MDM2. DP12 makes inferred chemical contacts (by proximity, $\leq$ 3.5 Å) to MDM2 residues L29, F30, G33, I36, Y42, Q47, and Y75. **(B)**: the AF3 top-ranked (sample 3) prediction of DP12 in complex with MDM2. Arrows indicate D-peptide prediction residues P4 to F7 where incorrect L-chirality occurs (see Table 1). The D-peptide prediction is incorrectly oriented perpendicular to the 2nd α-helix in MDM2. The DP12 prediction makes inferred chemical contacts (by proximity, $\leq$ 3.5 Å) with MDM2 residues L29, G33, Q34, M37, Q47, V68, and H71. The sets of inferred contacts only intersect on three residues: L29, G33, Q47. AF3 therefore only recapitulates 43% of target residue contacts and, due to inaccurate fold and chirality of the peptide, commonly establishes these chemical contacts with different residues relative to crystal structure.

2.2 Increased Number of Predictions for DP19

Although we use one seed above, Abramson et al.[1] report that for some challenging systems it may be necessary to generate a large number of predictions at the cost of additional computational time. We therefore selected the system with the highest average chirality violation (DP19, 52%) and ran predictions three times with total seed numbers of 1 (5 samples), 4 (20 samples), 16 (80 samples), 64 (320 samples) and 128 (640 samples) for a total of 3,195 samples. Each run contained non-overlapping seed intervals. As shown in Figure 5, increasing the number of seeds does not improve the chirality violation rate; the average chirality violation rate across all top-ranked samples is 63.5%. The average chirality violation rate across the entire population (3,195 samples) is $50.3 \pm 26.7\%$. These results show that inaccurate AF3 stereochemistry predictions are unable to be improved with additional computational power.

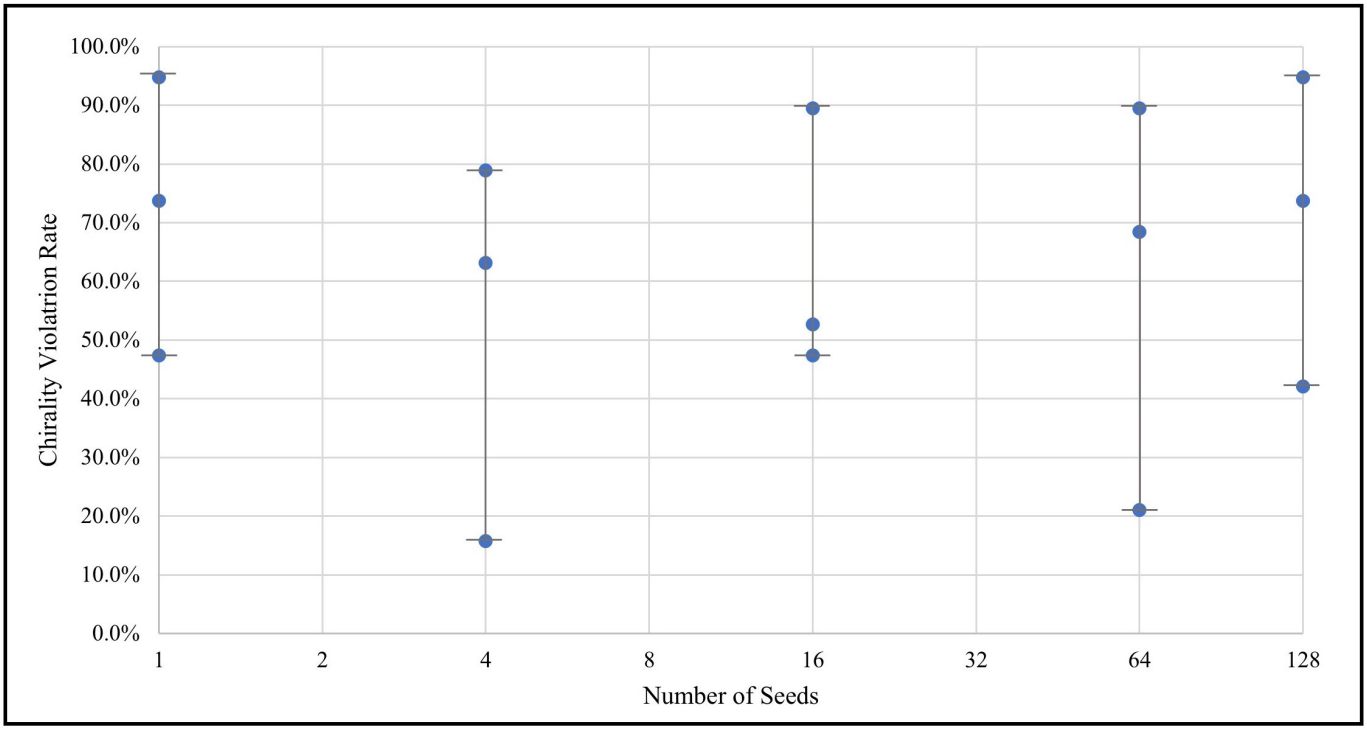

Figure 5: Increasing the number of AF3 seeds for known D-peptide binder DP19 (PDB ID 7YH8) fails to improve the chirality violation rate for top-ranked predictions. Three trials were run with total seed numbers of 1 (5 samples), 4 (20 samples), 16 (80 samples), 64 (320 samples) and 128 (640 samples). Seeds were run with non-overlapping intervals; no trial for a given D-peptide shares a seed number with another trial. A total of 3,195 samples were obtained. The x-axis is the number of seeds; each seed produces five samples (structural predictions). The y-axis is the chirality violation rate of the top-ranked sample, which is calculated by dividing the number of incorrect chiral centers (L or planar) by the total number of residues on the D-peptide. Although obtaining more samples has been considered a viable strategy for improving structural predictions on challenging structures[1], D-peptide top-ranked structural predictions do not improve with more seeds. The average chirality violation rate across all top-ranked samples is 63.5%. The average chirality violation rate across the entire population (3,195 samples) is $50.3 \pm 26.7\%$.

2.3 Apo D-Protein Using Mirror Symmetry

Of interest is to investigate if AF3 chirality violation is due to the presence of an L-target. While we would like to predict an apo (unbound) D-peptide, there are no known crystal structures of this type in the PDB due to the peptide's small size. However, as demonstrated in mirror image phage display[15], reflection is an energy-equivariant geometric operation that corresponds to symmetry in the energy field[20]. That is, an apo D-protein adopts the exact mirror fold of an apo L-protein, and vice versa[21]. Therefore, to accurately recapitulate known physics, AF3 should predict apo D-center inputs in the mirror conformation of the known apo L-counterpart, thereby maintaining correct stereocenter conformations.

To evaluate this capability, we obtained the crystal structure of the L-space Src-homology 3 (SH3) domain (PDB ID 1SHG) and input the mirror image of this structure (D-SH3) into AF3. If AF3 architecture correctly predicts D-peptide fold, it should predict D-SH3 in the mirror image of known L-SH3. However, the AF3 D-peptide prediction reported a markedly different fold (visualized in SI Section 4.3) and a chiral violation rate of 44%. This suggests that high chiral violation rate is not due to the presence of an L-target. We additionally investigated DP19, DP9, and DP12 in complex with decoy L-targets (Ubiquitin or GB1) and discovered the average chiral violation rate was 44% across all systems (see SI Section 4.2 for the full dataset).

2.4 Confidence Metrics

Abramson et al.[1] reported that AF3 confidence metrics are well calibrated with accuracy. Therefore, we reviewed the confidence scores for each prediction of Section 2.1 (see SI Section 4.5 for the full dataset) to investigate if any confidence metric returned by AF3 was predictive of structural accuracy for D-peptides. While the per-chain confidence scores were accurately low for D-peptide predictions, AF3 interface predicted template modeling (ipTM) score and ranking formula score fail to distinguish predictions with high chirality violation and correct docking vs. predictions with low chirality violation and incorrect docking.

The predicted template modeling (pTM) score indicates AF3 confidence in the entire structure. The mean D-peptide chain pTM for DP19, DP9, and DP12 are 0, 0.42, and 0, respectively. The mean L-protein chain pTM for L-19437, Streptavidin, and MDM2 are 0.80, 0.87, and 0.82, respectively. While all 12 L-protein chains reported high-confidence pTM scores (> 0.5), all 12 D-peptide chain predictions reported low-confidence pTM scores (< 0.5). Further, the per-chain ipTM score provides additional information about AF3 confidence in the D-peptide:L-protein interface. Each D-peptide:L-protein complex contains only one chain:chain interaction, so the ipTM score for the D-peptide, L-protein, and complex for each sample are identical. The average ipTM score for DP19:L-19437, DP9:Streptavidin, and DP12:MDM2 are 0.65, 0.57, and 0.60, respectively. These scores for each sample border the range between poor prediction (≤ 0.6) and "grey zone" ($0.6 - 0.8$), where prediction accuracy is unknown[22]. These results indicate that the ranges of AF3 confidence scores correctly predict that the overall structure of each L-protein is likely correct, and each D-peptide prediction is likely incorrect. However, despite these correct ranges, AF3 incorrectly ranks each seed and individual D-peptide samples.

Despite per-chain confidence scores predictive of structural accuracy, AF3 does not properly rank summary confidences (all samples for a given seed). Commonly, top-ranked predictions returned by AF3 exhibit the worst chirality violation and are predicted in the incorrect fold and binding pose. For example, DP19:L-19437 contains the best ranking score (0.77, 7YH8 Sample 3 in SI Section 4.5) across all systems, but this system has the highest average chirality violation rate (52%, see Table 1), and was the only system with a ligand predicted in the wrong binding site (see Figure 2). The Spearman

correlation coefficient (ρ) between ranking score and chirality violation rate across all systems is -0.30, and the Pearson correlation coefficient (R) is -0.31. Therefore, the correlation between ranking score and chirality violation is not statistically significant (see Figure 6 (A) for a visualization).

In addition to exhibiting poor seed ranking performance, the per-sample confidence value ranks exhibit no correlation to chiral violation rate. Neither D-peptide pTM nor D-peptide ipTM exhibit a significant statistical correlation to chirality violation rate for D-peptide samples. The pTM score and chirality violation were not correlated, where $\rho = -0.05$ and $R = -0.02$. Further, the ipTM score and chirality violation rate were also not correlated, with $\rho = -0.28$ and $R = -0.30$. None of these correlation coefficients are considered statistically significant; therefore, AF3 confidence metrics fail to distinguish predictions with low and high chirality violations. See Figure 6 (B) for a plot of the correlation between D-peptide ipTM and chirality violation. Additionally, the D-peptide ipTM and pTM scores converge on low confidence for each system (visualized in Figure 6 (C)).

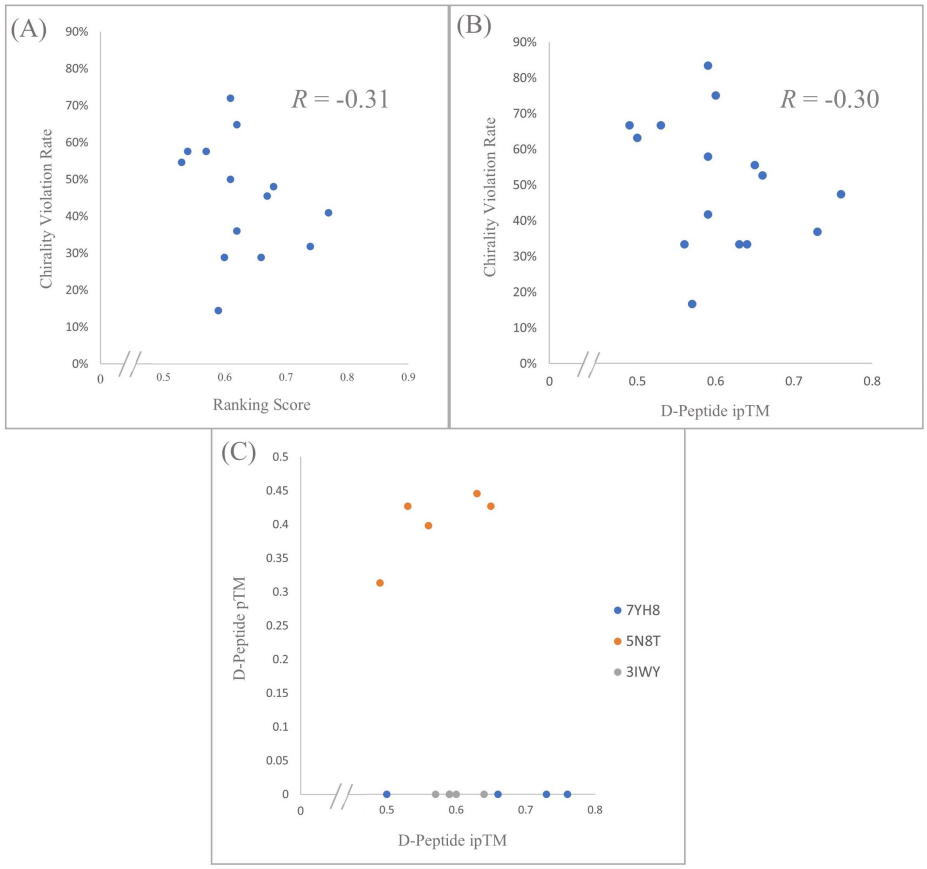

Figure 6: Plots of AF3 confidence metrics for D-peptide predictions demonstrate low correlation between correct chirality and confidence scores. **(A)**: plot of ranking score (all samples) versus chirality violation rate (%). The Spearman correlation coefficient (ρ) for (A) is -0.30 and the Pearson correlation coefficient (R) is -0.31; neither correlation is statistically significant. **(B)**: plot of D-peptide interface predicted template modeling (ipTM) score versus chirality violation rate (%). For (B), $\rho = -0.28$ and $R = -0.30$. This is not a statistically significant correlation. Plots (A) and (B) visualize how AF3 fails to establish statistical significance between D-peptide confidence metrics and correct chirality. **(C)**: plot of D-peptide ipTM score versus D-peptide predicted template modeling (pTM) visually confirms convergence of confidence metrics for each D-peptide sample at low confidence. Each system's D-peptide overall structure confidence (pTM) and ipTM sample scores group distinctly. Values are reported in Table 1 and SI Section 4.5.

3 Discussion

While the reported[1] AF3 chirality violation rate was low (4.4%) across all chiral centers, we have discovered that the model fails to maintain chiral centers demanded on input for D-peptides. The average chirality violation rate across all samples is 51%, so the model is as accurate on average as chance (random chirality choice, L or D, for each peptide residue). Additionally, AF3 inaccurately folds and docks D-peptides in complex with L-targets. As shown in Figures 2 - 4, AF3 typically either places the D-peptide in the incorrect binding position or improperly orients the D-peptide in the correct binding pocket. For the three D-peptides evaluated, each was folded incorrectly and was not predicted in the correct binding pose in any of the 15 samples (see SI Section 4.4 for a visualization of all samples). The top-ranked DP19 sample was oriented upside down and on the opposite side of L-19437 (Figure 2). The top-ranked DP9 sample was rotated in the binding pocket such that hydrophobic residues were unfavorably solvent exposed (Figure 3). Further, DP12 sample 3's backbone was rotated such that only three out of seven chemical contacts with MDM2 were recapitulated (Figure 4).

Interestingly, removing (apo) or changing the D-peptide's target (GB1 or Ubiquitin, see SI Table 3) fails to reduce average chirality violation to anything approaching the 4.4% originally reported[1]. It is possible that AF3's L-residue database statistics result in stereocenter bias, improperly producing samples with L-ligand residues. Additionally, describing D-stereocenters with fluorines in place of α-hydrogens did not eliminate chirality violation in any tested system (SI Table 3). Thus, AF3 architecture may have difficulty maintaining stereocenters for organic molecules resembling D-peptides. These results indicate AF3 struggles with D-peptide chirality regardless of target, even when hydrogens are replaced with other small atoms.

Analysis of confidence metrics revealed no statistical correlation between chirality violation rate and confidence scoring (Figure 6). Further, although the fold and binding pose of each D-peptide sample fails to converge in a structural ensemble (refer to SI Section 4.4 for a visualization), each D-peptide seed converges on low-confidence scoring (Figure 6 (C)). Because of the lack of correlation between D-peptide sample confidence rank and chirality violation, AF3 is unsuitable for D-peptides because its predictions are not only incorrect in chirality, fold, and pose, but also converge to low-confidence scoring. Unfortunately, these errors do not resolve with more seeds; as presented in Figure 5, the top-ranked and average chirality violation rate do not improve with additional predictions. This indicates the computational failures for this class of molecule cannot be be resolved with larger prediction volumes.

While AF3 has enabled a wide range of research for varying design objectives[23–25], the model is currently unable to reliably predict D-peptides. These shortcomings may be attributed to D-peptide crystal structure scarcity or inherent model flaws. AF3 correctly predicts low confidence ranges for incorrect D-peptide structures, but it nevertheless fails to properly rank and predict the correct chirality, fold, and binding pose of D-peptide predictions, regardless of number of predictions generated. This inability to produce high-quality D-peptide predictions indicates that further work is needed to effectively predict D-peptide:L-protein complexes. It is possible that performance will improve with comprehensive chirality violation parameters and more empirical D-peptide:L-protein structural data to be used in future model training.

Method Details

Evaluated D-peptide:L-protein systems were obtained from the Protein Data Bank[14]. These included DP19:L-19437 (PDB ID 7YH8), DP9:Streptavidin (PDB ID 5N8T), and DP12:MDM2 (PDB ID 3IWY). Water molecules and duplicate chains were removed. The D-peptide and L-target were split and saved into separate PDB files. Open Babel[26] was used to obtain the sequence of the L-protein

and D-peptide. The sequence for SH3 and decoy L-targets GB1 and Ubiquitin were obtained with the same methodology. The D-peptide amino acid sequence was converted to simplified molecular input line entry system (SMILES) data format (as used for AlphaFold 3 baseline models[1]) using NovoPro PepSMI[27]. To ensure D-peptide SMILES data was correctly converted, chiral centers were confirmed by two methods: (1) confirmed in 2D using RCSB PDB Chemical Sketch Tool[28] and (2) confirmed in 3D by converting SMILES to PDB using NovoPro smiles2pdb[29] and visualizing in PyMOL[30]. For D-peptides with fluorines replacing α-hydrogens, SMILES string were manually revised to represent this substitution. These data inputs were similarly evaluated in 2D and 3D. The L-target sequence and D-peptide SMILES string were input to AF3 in the required JSON format[31]. Model parameters were obtained from Google. Unless otherwise specified, each job was run with one seed. Outputs were protonated with Protoss[16,17] and inspected for L, D, and planar chiral centers.

Author Contributions

HC performed calculations, analyzed results, and wrote the manuscript with input from all authors. PZ and BRD designed experiments and provided research oversight.

Funding Information

We received funding from the NIH (grant R35 GM-144042 to BRD; grants R01 GM-115355 and R01 AI-094475 to PZ).

Competing Interests

BRD is a founder of Ten63 Therapeutics, Inc. HC and PZ have no competing interests to declare.

Materials & Correspondence

Requests for materials and correspondence should be directed to the corresponding author BRD (brd+bioRxiv25@cs.duke.edu).

References

[1] Josh Abramson et al. "Accurate structure prediction of biomolecular interactions with AlphaFold 3". In: *Nature* (2024). DOI: 10.1038/s41586-024-07487-w.

[2] Carl Branden and John Tooze. *Introduction to Protein Structure*. Garland Publishing, 1999.

[3] Siyu Wang et al. "Chiral evasion and stereospecific antifolate resistance in Staphylococcus aureus". In: *PLOS Computational Biology* (2022). DOI: https://doi.org/10.1371/journal.pcbi.1009855.

[4] S Chen et al. "Improving Binding Affinity and Stability of Peptide Ligands by Substituting Glycines with D-Amino Acids". In: *ChemBioChem* (2013). DOI: https://doi.org/10.1002/cbic.201300228.

[5] Li Di. "Strategic Approaches to Optimizing Peptide ADME Properties". In: *The AAPS Journal* (2015). DOI: 10.1208/s12248-014-9687-3.

[6] Nathan Guerin et al. "DexDesign: an OSPREY-based algorithm for designing de novo D-peptide inhibitors". In: *Protein Engineering, Design and Selection* (2024). DOI: 10.1093/protein/gzae007.

[7] PD Renfrew et al. "Incorporation of noncanonical amino acids into Rosetta and use in computational protein-peptide interface design". In: *PLoS ONE* (2012). DOI: https://doi.org/10.1371/journal.pone.0032637.

[8] G Bhardwaj, VK Mulligan, and CD Bahl. "Accurate de novo design of hyperstable constrained peptides". In: *Nature* (2016). DOI: 10.1038/nature19791.

[9] Pedro A. Valiente et al. "Computational Design of Potent D-Peptide Inhibitors of SARS-CoV-2". In: *Journal of Medicinal Chemistry* 64.20 (2021), pp. 14955–14967. DOI: 10.1021/acs.jmedchem.1c00655.

[10] Alessandro Angelini et al. "Bicyclic Peptide Inhibitor Reveals Large Contact Interface with a Protease Target". In: *ACS Chemical Biology* (2012). DOI: 10.1021/cb200478t.

[11] Gwen R. Buel and Kylie J. Walters. "Can AlphaFold2 predict the impact of missense mutations on structure?" In: *Nature Structural & Molecular Biology* (2022). DOI: https://doi.org/10.1038/s41594-021-00714-2.

[12] D Chakravarty et al. "AlphaFold predictions of fold-switched conformations are driven by structure memorization". In: *Nature Communications* (2024). DOI: https://doi.org/10.1038/s41467-024-51801-z.

[13] Robert T. McDonnell, Aaron N. Henderson, and Adrian H. Elcock. "Structure Prediction of Large RNAs with AlphaFold3 Highlights its Capabilities and Limitations". In: *Journal of Molecular Biology* (2024). DOI: https://doi.org/10.1016/j.jmb.2024.168816.

[14] H.M. Berman et al. "The Protein Data Bank". In: *Nucleic Acids Research* (2000). DOI: https://doi.org/10.1093/nar/28.1.235.

[15] T. Schumacher et al. "Identification of D-peptide ligands through mirror-image phage display". In: *Science* (1996). DOI: 10.1126/science.271.5257.1854.

[16] Katrin Schöning-Stierand et al. "ProteinsPlus: a comprehensive collection of web-based molecular modeling tools". In: *Nucleic Acid Research* (2022). DOI: 10.1093/nar/gkac305.

[17] Stefan Bietz et al. "Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes". In: *Journal of Cheminformatics* (2014). DOI: https://doi.org/10.1186/1758-2946-6-12.

[18] Anupam Nath Jha, Saraswathi Vishveshwara, and Jayanth R. Banavar. "Amino acid interaction preferences in proteins". In: *Protein Science* (2010). DOI: https://doi.org/10.1002/pro.339.

[19] Ken Dill. "Dominant Forces in Protein Folding". In: *Biochemistry* (1990). DOI: 10.1021/bi00483a001.

[20] E. Noether. "Gesammelte Abhandlungen-Collected Papers, Springer Collected Works in Mathematics". In: *Springer* (1983). DOI: https://doi.org/10.1007/978-3-642-39990-9.

[21] Alexander J. Lander, Yi Jin, and Louis Y. P. Luk. "D-Peptide and D-Protein Technology: Recent Advances, Challenges, and Opportunities". In: *ChemBioChem* (2022). DOI: https://doi.org/10.1002/cbic.202200537.

[22] EMBL-EBI. *Confidence scores in Alphafold-Multimer*. URL: https://www.ebi.ac.uk/training/online/courses/alphafold/inputs-and-outputs/evaluating-alphafolds-predicted-structures-using-confidence-scores/confidence-scores-in-alphafold-multimer/.

[23] Joseph Kreitz et al. "Programmable protein delivery with a bacterial contractile injection system". In: *Nature* (2023). DOI: https://doi.org/10.1038/s41586-023-05870-7.

[24] Yang Lim et al. "In silico protein interaction screening uncovers DONSON's role in replication initiation". In: *Science* (2023). DOI: 10.1126/science.adi3448.

[25] Shyamal Mosalaganti et al. "AI-based structure prediction empowers integrative structural analysis of human nuclear pores". In: *Science* (2022). DOI: https://doi.org/10.1126/science.abm9506.

[26] Noel M O'Boyle et al. "Open Babel: An open chemical toolbox". In: *Journal of Cheminformatics* (2011). DOI: 10.1186/1758-2946-3-33.

[27] NovoPro Bioscience Inc. *PepSMI: Convert Peptide to SMILES string*. URL: https://www.novoprolabs.com/tools/convert-peptide-to-smiles-string.

[28] RCSB. *Chemical Sketch Tool*. URL: https://www.rcsb.org/chemical-sketch.

[29] NovoPro Bioscience Inc. *Convert SMILES to 3D Structure*. URL: https://www.novoprolabs.com/tools/smiles2pdb.

[30] Schrödinger, LLC. "The PyMOL Molecular Graphics System, Version 2.5".

[31] Google DeepMind. *Documentation*. URL: https://github.com/google-deepmind.

[32] Yang Zhang and Jeffrey Skolnick. "Scoring function for automated assessment of protein structure template quality". In: *Proteins* (2004). DOI: 10.1002/prot.20264.

4 Supplementary Information

4.1 D-Peptide and L-Protein Sequences

Name	Sequence
DP19	DEHELLETAARWFYEIAKR
L-19437	LPVEKIIREAKKILDELLKRGLIDPELARIAREVLERARKLGNEEAARFVLELIERLRRELS
DP9	LWQHEATWK
Streptavidin	AGITGTWYNQLGSTFIVTAGADGALTGTYESAVGNAESRYVLTGRYDSAPATDGSGTALGWTVAWKNNYRNAHSATTWSG QYVGGAEARINTQWLLTSGTTEANAWKSTLVGHDTFTKVK
DP12	DWWPLAFEALLR
MDM2	TLVRPKPLLLKLLKSVGAQKDTYTMKEVLFYLGQYIMTKRLYDEKQQHIVYCSNDLLGDLFGVPSFSVKEHRKIYTMIYRNL
SH3	KELVLALYDYQEKSPREVTMKKGDILTLLNSTNKDWWKVEVNDRQGFVPAAYVKKLD
Ubiquitin	GMQIFVDTVQWKTITLEVEPSDTIENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESALILLLTLR
GB1	MTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDAATKTFTVTE

Table 2: Full sequences for DP19:L-19437 (PDB ID 7YH8), DP9:Streptavidin (PDB ID 5N8T), DP12:MDM2 (PDB ID 3IWY), SH3 (PDB ID 1SHG), and decoy targets Ubiquitin (PDB ID 6NXL) and GB1 (PDB ID 2J52). The Name column indicates protein label and the Sequence column contains the sequence in single-letter amino acid notation.

4.2 Decoy Targets and Fluorinated Peptides

In addition to testing each D-peptide in complex with the endogenous target (Section 2.1) and apo (Section 2.3), we placed each D-peptide in complex with decoy target Ubiquitin (PDB ID 6NXL, chain A) or GB1 (PDB ID 2J52, chain A) to further investigate if target affects ligand stereochemistry in AF3. Although predicting D-peptides with endogenous target and apo failed to improve chirality violation rate to anything approaching 4.4%, we evaluated if non-endogenous L-targets would improve predictions. While incorrect ligand chirality for an AF3 D-peptide:L-endogenous target may indicate a faulty chirality interpretation module, an incorrect D-peptide:L-decoy provides further evidence that AF3 has a fundamental flaw with D-peptides. We report that AF3 predicts models with high chirality violation (average 44% across all decoy systems) and again displays inconsistent patterns of D/L residue chirality for a given D-peptide. Additionally, to test whether AF3 chirality violation was due to hydrogen stripping in AF3's input processing, we replaced each D-peptide α-hydrogen with a different small atom, fluorine. When placed in complex with the respective endogenous ligand, AF3 still produces models with a high chirality violation (average 33% across all three systems) despite non-hydrogen stereocenters. These results are displayed in Table 3.

System	Res 1	Res 2	Res 3	Res 4	Res 5	Res 6	Res 7	Res 8	Res 9	Res 10	Res 11	Res 12	Res 13	Res 14	Res 15	Res 16	Res 17	Res 18	Res 19	Chirality Violation Rate
DP19 Sample 3	D	L	Planar	L	L	L	L	L	L	D	D	D	D	D	D	D	D	D	L	47%
DP19:Ubiquitin	Planar	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	D	5%
DP19:GB1	D	D	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	L	89%
DP19(Fluorinated):L-19437	L	L	L	D	L	L	L	L	L	D	D	D	D	D	D	L	D	D	L	53%
DP9 Sample 4	L	D	L	L	D	L	L	D	D											56%
DP9:Ubiquitin	Planar	D	L	L	L	D	L	L	Planar											78%
DP9:GB1	D	L	D	D	D	L	D	L	L											44%
DP9(Fluorinated):Streptavidin	D	D	D	D	D	L	L	D	D											22%
DP12 Sample 3	D	D	D	L	L	L	L	D	D	D	D	D								33%
DP12:Ubiquitin	Planar	D	L	D	D	D	D	D	D	D	D	L								25%
DP12:GB1	D	L	L	D	D	D	D	L	D	D	D	D								25%
DP12(Fluorinated):MDM2	L	L	L	D	D	D	D	D	D	D	D	D								25%

Table 3: Per-residue chirality violations for AlphaFold 3 predictions of D-peptides with synthetic L-targets or fluorine. Chirality for the L-targets (L-19437, Streptavidin, MDM2) are not shown in this table. The System column details the evaluated complex. DPX:Ubiquitin is the D-peptide DPX (DP19, DP9, or DP12) in complex with Ubiquitin. DPX:GB1 is the D-peptide DPX (DP19, DP9, or DP12) in complex with GB1. Systems with "(Fluorinated)" indicate the D-peptide's α-hydrogens were replaced with flourine. Top-ranked samples from Table 1 are repeated here for reference. For brevity, only the top-ranked sample for each experiment is displayed. The Res # columns indicate the chirality of the residue index (D (correct) or L (incorrect)). Entries of "Planar" indicate that AlphaFold 3 incorrectly predicted a flattened protein backbone that could not be assigned to either chirality (see Figure 1 for a visualization of chiral centers). Correct D stereocenters are in green while incorrect L and Planar stereocenters are in red. Chirality violation is calculated by dividing the number of peptide residues with incorrect chirality (L or Planar) by the length of the peptide. Decoy systems (peptides targeting GB1 or Ubiquitin) display inconsistent patterns of D/L chirality violation, per residue, on the top-ranked model. Describing chiral centers with fluorines in place of α-hydrogens failed to eliminate chirality violation in any system.

4.3 Apo D-protein Prediction Using SH3

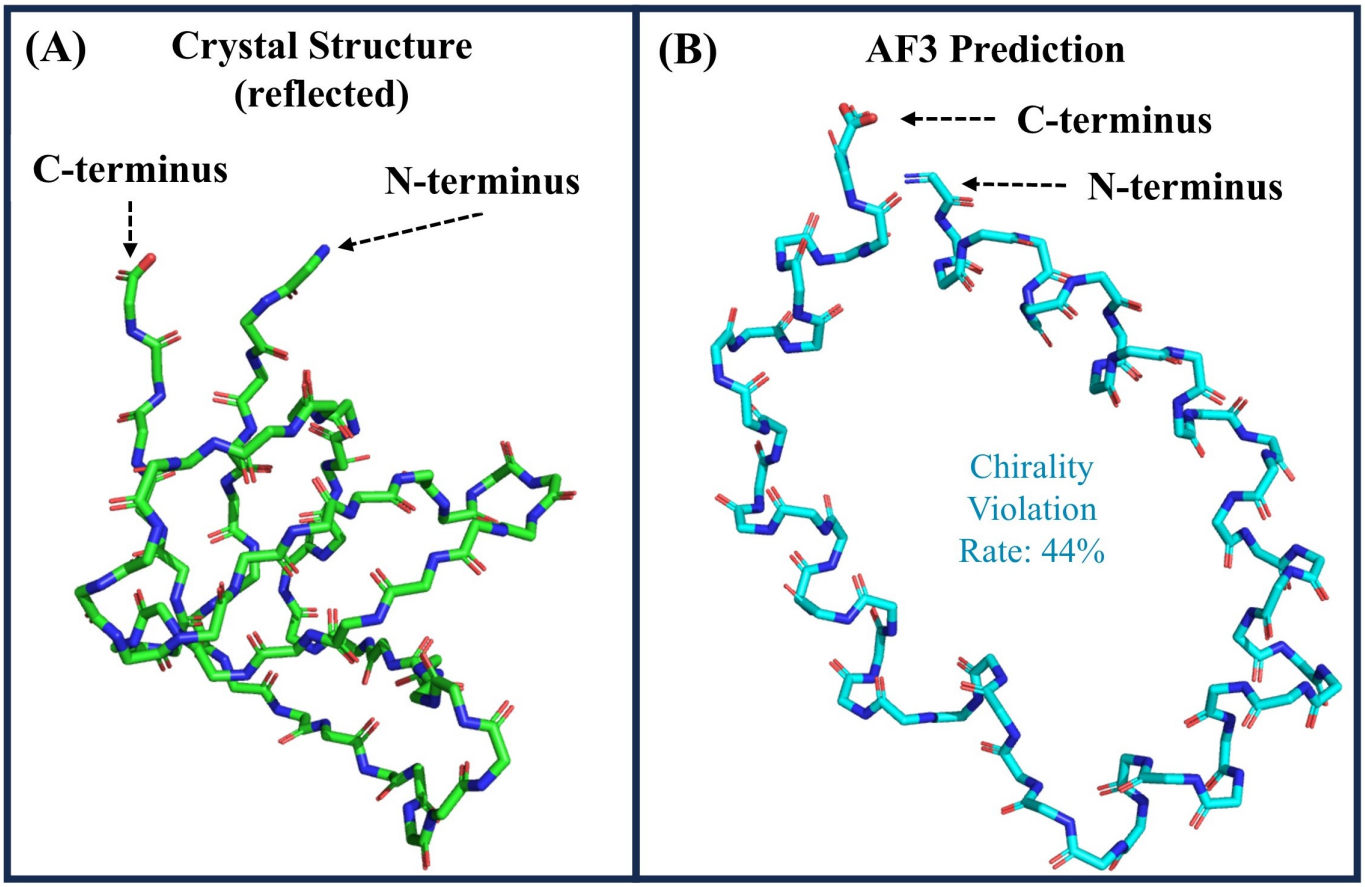

Figure 7: Comparison of D-SH3 crystal structure (PDB ID 1SHG, reflected across the z-axis) to AF3 top-ranked (sample 0) prediction of D-SH3. Sidechains are hidden for clarity. **(A)**: The crystal structure of Src-homology 3 (SH3) reflected into D-space (D-SH3, green sticks). L-SH3 and D-SH3 should adopt the same fold, but with opposite handedness[20,21]. **(B)**: AF3 prediction of D-SH3 (cyan sticks). This structure differs significantly from known crystal structure, and has a chiral violation rate of 44%.

4.4 AF3 D-peptide Predictions Fail to Converge to a Consensus Structure

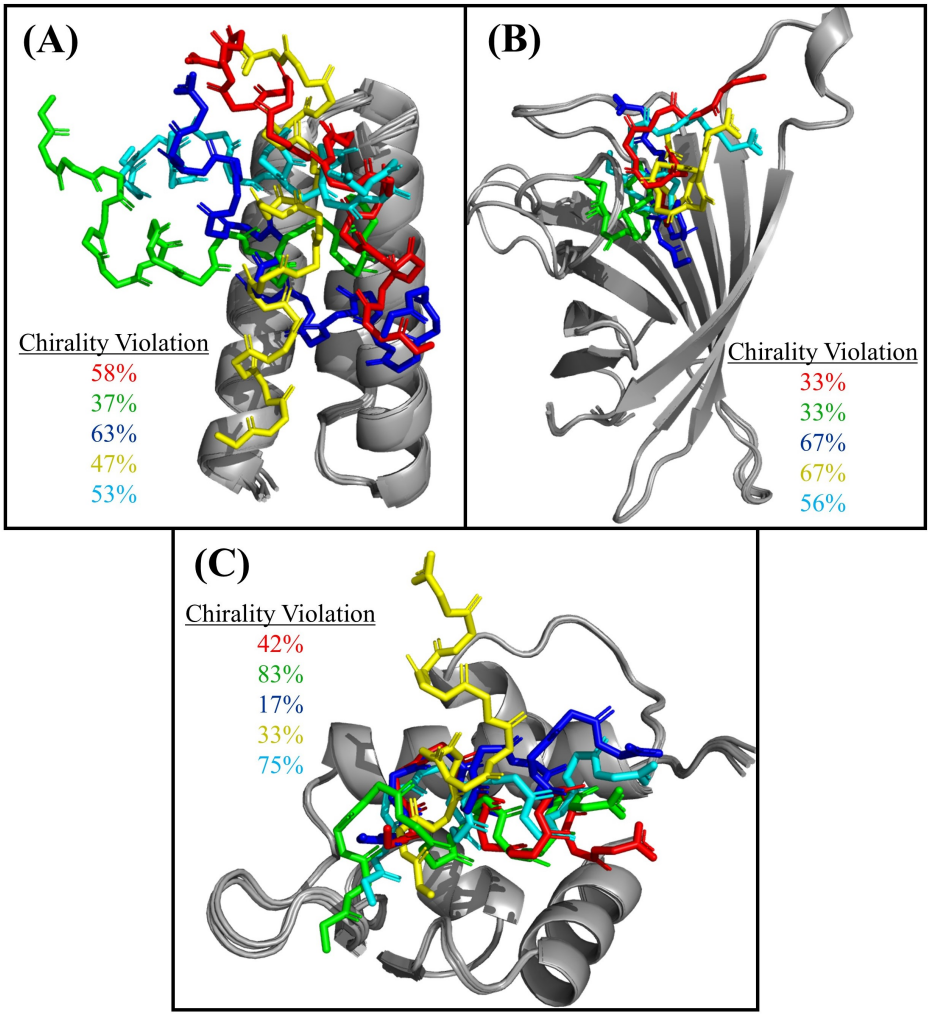

Figure 8: Samples from each AlphaFold 3 D-peptide prediction (Section 2.1) fail to converge. **(A)**: AF3 predictions for DP19:L-19437 (PDB ID 7YH8). **(B)**: AF3 predictions for DP9:Streptavidin (PDB ID 5N8T). **(C)**: AF3 predictions for DP12:MDM2 (PDB ID 3IWY). Each frame includes the five samples produced by AF3 (see Table 1 for chirality violation rates). All L-targets are shown as gray cartoons, while D-peptide samples are shown as sticks. Sidechains are hidden for clarity. Sample 0 is red, sample 1 is green, sample 2 is blue, sample 3 is yellow, and sample 4 is cyan. Chirality violation rates for the D-peptides are reported in each frame as the corresponding sample color. The L-target predictions largely converge; the set of samples for each system demonstrate few structural deviations. While AF3 docks each ligand prediction sample in the same general binding site, the D-peptide samples in each system fail to achieve a consensus fold or binding pose. (A): predictions for DP19 deviate in structure significantly between each sample. In addition to docking the predictions in the incorrect binding location (see Figure 2 for more details), each sample is predicted in a different orientation in the binding pocket, resulting in a C-term distance between sample 0 and sample 1 of 32.7 Å. (B): similar to DP19, predictions for DP9 fail to reach a fold and pose consensus (see Figure 3 for more details). The C-term distance between sample 0 and sample 1 is 14.5 Å. (C): DP12 fails to converge in fold or binding pose (see Figure 4 for more details). The C-term distance between sample 0 and sample 3 is 18.4 Å.

4.5 AF3 Confidence Metrics

System	Ranking Score	D-Peptide Chain pTM	L-Protein Chain pTM	Complex pTM	D-peptide Chain ipTM	L-Protein Chain ipTM	Complex ipTM	fraction_disordered	has_clash
7YH8 Sample 0	0.61	0	0.80	0.68	0.59	0.59	0.59	0	0
7YH8 Sample 1	0.74	0	0.83	0.79	0.73	0.73	0.73	0	0
7YH8 Sample 2	0.53	0	0.77	0.62	0.50	0.50	0.50	0	0
7YH8 Sample 3	0.77	0	0.79	0.81	0.76	0.76	0.76	0	0
7YH8 Sample 4	0.67	0	0.79	0.73	0.66	0.66	0.66	0	0
5N8T Sample 0	0.60	0.42	0.88	0.75	0.56	0.56	0.56	0	0
5N8T Sample 1	0.66	0.47	0.87	0.78	0.63	0.63	0.63	0	0
5N8T Sample 2	0.54	0.33	0.87	0.72	0.49	0.49	0.49	0	0
5N8T Sample 3	0.57	0.45	0.87	0.74	0.53	0.53	0.53	0	0
5N8T Sample 4	0.68	0.45	0.87	0.79	0.65	0.65	0.65	0	0
3IWY Sample 0	0.62	0	0.82	0.72	0.59	0.59	0.59	0	0
3IWY Sample 1	0.61	0	0.81	0.72	0.59	0.59	0.59	0	0
3IWY Sample 2	0.59	0	0.81	0.70	0.57	0.57	0.57	0	0
3IWY Sample 3	0.66	0	0.83	0.75	0.64	0.64	0.64	0	0
3IWY Sample 4	0.62	0	0.82	0.73	0.60	0.60	0.60	0	0

Table 4: Confidence metrics returned by AlphaFold 3 for predictions of DP19:L-19437 (PDB ID 7YH8), DP9:Streptavidin (PDB ID 5N8T), and DP12:MDM2 (PDB ID 3IWY). The System column indicates the evaluated complex and sample number. Ranking score is calculated by AF3 using the formula $0.8(\text{ipTM}) + 0.2(\text{pTM}) + 0.5(\text{fraction_disordered}) - 100(\text{has_clash})$. The D-peptide pTM, L-protein pTM, and Complex pTM columns indicate the predicted template modeling (pTM) score for the peptide, protein, and complex, respectively. Derived from template modeling[32] (TM) score, the pTM score reports how confident AF3 is in the accuracy of the entire structure. A pTM score ≥ 0.5 means the overall structure is likely similar to the crystal structure[22]. The D-peptide ipTM, L-protein ipTM, and Complex ipTM report the interface predicted template modeling (ipTM) score for the peptide, protein, and complex, respectively. The ipTM score reports the confidence in predicted interfaces. Because AF3 is only evaluating a single D-chain:L-chain interface, these values will be identical for a given sample. ipTM scores > 0.8 are considered high-quality predictions, values from $0.6 - 0.8$ are a "grey zone" where accuracy is uncertain, and values ≤ 0.6 likely indicate a poor prediction[22]. The fraction_disordered column reports a scalar in the range $0 - 1$ that indicates the fraction of prediction structure that is disordered. The has_clash column is a boolean indicating if the predicted sample contains either a chain with 100 clashing atoms or a chain where $> 50\%$ of the chain is clashing. AF3 correctly predicts high confidence for the L-target samples. AF3 correctly reports low confidence ranges for each D-peptide sample, but incorrectly ranks individual samples according to chirality violation rate.