nature portfolio

Peer Review File

Cyclic peptide structure prediction and design using AlphaFold2

Corresponding Author: Professor Gaurav Bhardwaj

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript by Rettie et al. focuses on the development of a new method based on AlphaFold for the design of cyclic peptides. The authors enhance the existing methodology by introducing relative positional encoding to allow sampling and design of structures with cyclic backbones. The approach presented is innovative and significant for the broad scientific community that would eventually merit its publication in this journal. However, there are several issues regarding the quantification of the accuracy of this new method that need to be addressed before the manuscript can be recommended for publication. These points mostly revolve around the use of the metrics selected to assess the accuracy of AfCycDesign and its advantage over its alternatives. There are a few other points that need to be mentioned in the manuscript. First, the authors need to comment on the advantages of their method beyond its accuracy, such as a measure of its speed, ease-of-use, etc. since the existing methods can also do a decent job in modeling macrocycles in many cases due to the limited conformational space of these peptides. Second, the authors over-present their successes and have no clear discussion on the failures of their method (which can amount to over half of their designs or predictions in some cases) and how these can be amended. Third, later sections of the manuscript lack detail and need to be expanded to deliver their full effect. Below are my point-by-point comments on what I consider to be the most important shortcomings of the manuscript.
"…however, AfCycDesign had lower confidence in those predictions (pLDDT < 0.85) (Figure 1B). The lower confidence in these predictions stemmed from the loop regions that are also flexible in the NMR ensembles"

> The regional or overall RMSD variation in NMR ensembles needs to be compared with the pLDDT predictions to confirm that the low pLDDT values are indeed caused by macrocycle flexibility. In other words, a plot that shows pLDDT vs NMR RMSD should be added as a supplementary figure to corroborate the authors' arguments.

"For each of the cyclic peptides from the PDB, the highest confidence model was used based on pLDDT, and the reported RMSD is lowest of the backbone heavy atom RMSD calculated using the RMSDMetric Simple Metric from the Rosetta software suite against all members of the deposited NMR ensemble."
> pLDDT is not a metric of absolute structural accuracy; it is a metric of how "consistent" a structure is expected to be with respect to a PDB structure. Since the PDB structures AF were trained on are typically globular proteins, there is no a priori reason to assume that it will be valid for peptides. Further, McDonald et al. (2023) have shown that AF-predicted pLDDT rankings do not correlate with prediction accuracy for peptides. Therefore, the comparisons between macrocycles and the NMR structures must be all vs. all (i.e., all five AF models vs. all NMR structures) in order to reflect the real accuracy of the method.

"Out of the 80 test cases predicted by AfCycDesign with single sequence and single recycle settings, 52 and 49 peptides were predicted close to the native structure (RMSD < 1.5 Å) and with high confidence (pLDDT > 0.7) by offset Types 1 and 2, respectively (Supplementary Figure S3C and Supplementary Table S2). These data suggested that positional encodings from residues far away in the sequence space do not significantly affect the structure prediction accuracy. To test this further, we explored the offset Type 3, where relative positional encodings were provided only for the two adjacent residues, and other relative positional encodings in the offset matrix were set to the maximum distance of 32 (Supplementary Figure S3A right panel). We observed that 48 out of the same 80 test cases were now predicted close to the native structure with high confidence, suggesting that applying a cyclic offset with relative positional encodings of adjacent residues is sufficient to predict the structure of cyclic peptides correctly."
> The underlined interpretation is not necessarily true since this depends on the size and the fold of a macrocycle. There are

at least two exceptions to the proposed conclusion. The first is that smaller macrocycles are likelier to be affected by long-range interactions more than larger macrocycles since the cyclic nature of these peptides inevitably results in a “tighter” arrangement of the residues within the scaffold. For example, in a seven-aa-long linear peptide, the residues 3 and 7 might be unrelated, but in a cyclic structure, these two residues can be within the interaction range. The second exception is stapled peptides. The dataset includes many macrocycles with disulfide bonds, and the cysteine residues, however far they may be in terms of the primary sequence, can form a disulfide bond in the tertiary structure of the macrocycle. This also applies to non-stapled structures with tightly packed tertiary folds. In order to deconvolute these factors, the authors need to provide three additional lines of data: the effect of the alternative cyclic offsets in relation to 1) the size of the tested macrocycles, 2) disulfide stapled vs. non-stapled structures, and 3) how tightly packed a structure is (perhaps SASA or a similar metric to measure).

“The sequence designed by AfCycDesign differed significantly from the Rosetta-designed sequence, with 12 mutations in the sequence and only a singular alanine in the core of the peptide being conserved.”
> This is expected since the Rosetta sequence design method is not a very accurate way of designing sequences. Why did the authors opt for Rosetta instead of a more accurate method like ProteinMPNN for their comparisons? This is later done for the hallucinated peptides, so I see no reason why the previous comparisons are confined to Rosetta.

“Many of these hallucinated sequences demonstrated promising folding propensity in these calculations, with 114 7-mers, 186 8-mers, 139 9-mers, and 76 10-mer sequences showing Rosetta Pnear values greater than 0.6. We selected one hallucinated design model per size between 7–10 amino acids with AlphaFold pLDDT > 0.9 and Rosetta Pnear > 0.9 for experimental validation and structural characterization.”
> Since the authors are using Rosetta as their gold standard, what are the Rosetta structures predicted for the tested sequences? Figure 3 provides only the predictions made by their method, which does not address questions regarding whether Rosetta is worse at folding or designing these macrocyclic structures compared to their method.

“For 11-mer, 12-mer, and 13-mer macrocycles, we identified 28457, 27715, and 27056 unique structural clusters, respectively (Figure 4, first column) from large-scale design calculations (see Methods).”
> I am surprised that the number of clusters is roughly the same although the increasing size should result in a larger conformational space, thus more scaffold diversity. Do the authors have an explanation for this?

“Next, we explored if our set of hallucinated peptides could be used as starting scaffolds for designing cyclic peptide binders to protein targets of interest. As a proof of concept, we chose to design binders against MDM2 given that its natural binding interface to p53 is a single helix, and our scaffold set contains multiple members with small helical motifs. We also expanded our scaffold set to sequence length up to 16 residues. We grafted a 5-residue motif from p53 (PDB: 4HFZ) on our set of hallucinated scaffolds using the MotifGraft mover from the Rosetta Software Suite 32 . Post-grafting, we designed the sequence of cyclic peptides using ProteinMPNN, followed by the energy minimization using the Rosetta energy function. For in silico filtering of design candidates, the peptide-target complex structure was predicted using AfCycDesign and down-selected based on the PAE of the interface residues (iPAE), Rosetta-based binding energy, aggregation propensity score, and contact molecular surface area 33 . We synthesized 14 designs and screened them first at a single concentration (50 μM) in an MDM2 AlphaLISA assay that measures the disruption of the interaction between MDM2 and a previously described ligand. (Figure 5). Of the 14 tested, 5 designs showed greater than 50% reduction in signal. The original 5-residue motif grafted onto our scaffold set showed no activity up to 50 μM in the same assay (Supplementary Figure S6). Next, we calculated the IC50 of the three best designs (designs D14, D15, D16) from the screening assay, and the best binder, design D14, showed an IC50 of 338.4 nM in these assays, highlighting the ability of our pipeline to identify binders even with low-throughput testing of a small number of design candidates”
> This is the most critical section of the paper since it shows the actual utility of the authors’ method, but the text is uninformative and is written in a sloppy manner contrary to the rest of the manuscript. The authors should either remove this section or elaborate. For the latter option, this section has to include a more thorough discussion of the specific interactions (and a supporting figure showing them) and how and why these designs are effective. There should also be a few sentences on existing inhibitors and how they compare with the hallucinated peptides designed by the authors in terms of IC50 or other metrics.

Minor
> Figure 1A needs to be redone to include the full x-axis legend.
> “IC50” should be changed into IC50 (50 subscript) in the last paragraphs of the results.

(Remarks on code availability)
I tested the Colab version to run the code, which seems to generate reasonable backbones.

Reviewer #2

(Remarks to the Author)
The authors have developed an improved approach for the design of small constrained peptides and determined how design predictions are influenced by sequence. The work is well presented but fails to fully convince on the broad applicability of the approach.

To back up the author claims that the new approach is faster and less computationally expensive than earlier approaches, the authors need to benchmark the approach against others and additionally consider how the different approaches

influence the number of molecules that need to be synthesised to generate early leads like the example shown.

Comparison of NMR structures not covered in AlphaFold are by their nature often restraint-underdefined but may better approximate solution structures. RMSD comparisons to the lowest energy structure to the best fitted structure need to be shown in a 2-way plot to gain deeper insight into the prediction strength of the approach. The authors need to discuss how the approach takes into account the solution structure of macrocells and ideally show how similar the determined X-ray structures of the macrocycles are theiir respective NMR structures.

The data in Fig. 5 is promising but the authors need to show that these modestly-potent early leads can be further optimised to prove the broad value of the approach.

The authors need to show the approach can be easily adapted to develop leads at related targets and conversely can be used to develop selectivity away from related targets to avoid unwanted interactions.

To validate the broad impact of this approach data in Fig. 5 should be extended to an unrelated target (eg a GPCR or ion channel) and be shown to be able to include D amino acids in the design approach.

To ensure the approach can be broadly implanted, the supporting software, databases and actual scripts used in the present study need to be made broadly available at time of publication.

Minor points:
I would prefer another term tohallucinated which suggests they were designed from a biologically altered state.

(Remarks on code availability)
Unfortunately, I am not in a position to assess the code.

Reviewer #3

(Remarks to the Author)
In the manuscript by Rettie et al. the authors describe a modification of the AlphaFold deep learning protein structure prediction model to better its performance on the task of predicting the structure of protegenic cyclic peptides. The authors modify the positional encoding portion of the sequence embedding portion of the AlphaFold architecture to encode the relative positions of a cyclic sequence. The pre-print of this draft manuscript and accompanying Colab notebooks represent the first published demonstration and characterization of this approach.

It should be noted that the publication of much of the code both as available github repositories as well as live and functional Colab Notebooks islaudable and democratizes many elements of this approach to practitioners in a variety of adjacent fields without the need to be able to set up a DL inference environment, something which can be challenging for computational and non-computational scientists alike.

The authors go on to use this modified model, dubbed AfCycDesign, as base model for a variety of structure prediction and design tasks. These tasks include validating the prediction on test of set of NMR structures in the PDB (excluded from the initial AlphaFold model training), sequence redesign where a new sequence is proposed for a known structure which engenders a lower calculated energy for the same conformation (ie a higher degree of preference for the target conformation than the parent sequence), “hallucination” of diverse geometry and rigid cyclic peptides of varying lengths, and finally, grafting known binding motifs of MDM2 into one of the hallucinated scaffolds, using a variety of previously described DL and non-DL tools to design novel binders and the resulting validation of binding via competitive biochemical assay. These results and approaches should be of interest to the field and provide a novel sampling and design strategy for proteinogenic cyclic peptides.

While the structure prediction and hallucination design portion of the manuscript is sound and well supported by the experimental data, the MDM2 binding design portion of the manuscript has some issues that should be addressed before publication:

1 - The established AlphaLISA is a reasonable way to initially screen and quantify the disruption of the interaction, a secondary biochemical assay technique is required to validate the results. Preferably a label free technique capable of establishing the Kd of the ligands to the target such as ITC or SPR, both of which are readily available to academic laboratories in most institutions and the reported IC50s should
allow for either technique to be employed.

2 - The authors provide crystallographic evidence of designed cyclic peptide structures for the hallucinated peptides but not their resulting grafted, designed binders, this would strengthen the results substantially.

3 - Along these lines, MDM2 readily crystallizes and a number of academic and commercial labs are capable of co-crystallization of these hits with the target, while more challenging than apo structures, this would provide much more compelling evidence of a successful designed binder

4 - A relevant negative control such as the enantiomeric form of one of the best binders should be included.

5 - AMG-232's binding isotherm should be included in the IC50 calculations

6 - A number of designs show no appreciable binding even at high concentrations, a rationale for the failure of these designs should be given, ideally supported by structural information

7 - While the cyclic peptide scaffold is cited as being advantageous from a stability and biophysical property standpoint, neither are assessed on the resulting scaffolds which are very unlikely to have any cell permeability. This should either be assayed and included or clarified in the text.

8 - The linear p53 motif that is used as a comparator for the cyclic scaffolds is much smaller than the cyclic peptides, using an equivalent length of residues to the cyclic peptide would be a more fair comparison and would likely show similar binding affinity to the best designed compounds based on previously affinities for reported p53/MDM2 FP assays.

9 - While this approach should be of value to the field, it is reliant on a known binding motif, compatible scaffold, grafting to said scaffold, and resulting reoptimization. Why was a hallucinated binder approach not tried as well or was it tried but failed due to insufficient sampling?

(Remarks on code availability)
As mentioned in the remarks to author, the inclusion of working Colab notebooks is a great addition to the code availability.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The changes made by the authors addressed all my concerns. I think the manuscript can be published as-is.

(Remarks on code availability)
See the first round reviews.

Reviewer #2

(Remarks to the Author)
I am happy that the authors have addressed my concerns adequately.

(Remarks on code availability)

Reviewer #1 (Remarks to the Author):

This manuscript by Rettie et al. focuses on the development of a new method based on AlphaFold for the design of cyclic peptides. The authors enhance the existing methodology by introducing relative positional encoding to allow sampling and design of structures with cyclic backbones. The approach presented is innovative and significant for the broad scientific community that would eventually merit its publication in this journal. However, there are several issues regarding the quantification of the accuracy of this new method that need to be addressed before the manuscript can be recommended for publication. These points mostly revolve around the use of the metrics selected to assess the accuracy of AfCycDesign and its advantage over its alternatives. There are a few other points that need to be mentioned in the manuscript. First, the authors need to comment on the advantages of their method beyond its accuracy, such as a measure of its speed, ease-of-use, etc. since the existing methods can also do a decent job in modeling macrocycles in many cases due to the limited conformational space of these peptides. Second, the authors over-present their successes and have no clear discussion on the failures of their method (which can amount to over half of their designs or predictions in some cases) and how these can be amended. Third, later sections of the manuscript lack detail and need to be expanded to deliver their full effect. Below are my point-by-point comments on what I consider to be the most important shortcomings of the manuscript.

We thank Reviewer 1 for their feedback and recommendation for publication after addressing their listed concerns. We have made several changes to the original manuscript based on the suggestions by Reviewer 1.

- We discuss the additional advantages of the AfCycDesign approach, including the improved speed, ease of use, lack of requirements around pre-specifying disulfide connectivities for stapled peptides, and improved ability to design larger cyclic peptides without additional crosslinks.
- We discuss the limitations and failures more in the manuscript now. While the success rates in hallucinating monomeric cyclic peptides are excellent, with the X-ray structures for 8/8 hallucinated designs closely matching their design models, only 5 out of 15 Mdm2 designs show binding. We now discuss the similarities between the successful and unsuccessful designs in the revised manuscript. These success rates are comparable to those observed in the design of larger-sized protein binders[1].
- We have rewritten and considerably expanded the binder design section with additional biochemical and structural validation. We also include design and validation against a second target, Keap1, in the revised manuscript. Three out of three designs tested against Keap1 show binding, with IC_{50}s ranging between 80–200 nM.

"…however, AfCycDesign had lower confidence in those predictions (pLDDT < 0.85) (Figure 1B). The lower confidence in these predictions stemmed from the loop regions that are also flexible in the NMR ensembles"

> The regional or overall RMSD variation in NMR ensembles needs to be compared with the pLDDT predictions to confirm that the low pLDDT values are indeed caused by macrocycle flexibility. In other words, a plot that shows pLDDT vs NMR RMSD should be added as a supplementary figure to corroborate the authors' arguments.

We agree with reviewer 1 on this point. While the data presented originally for a couple of test cases seems to show a correlation between low pLDDT and flexible regions in NMR structures; it cannot be said for sure that the lower confidence in those predictions is due to ("stems from") the flexibility given the limited number of such cases and overall size of the data analyzed in the benchmark set. Additional factors can affect the model confidence in those regions. Therefore, we removed this section and the related figure.

"For each of the cyclic peptides from the PDB, the highest confidence model was used based on pLDDT, and the reported RMSD is lowest of the backbone heavy atom RMSD calculated using the RMSDMetric Simple Metric from the Rosetta software suite against all members of the deposited NMR ensemble."
> pLDDT is not a metric of absolute structural accuracy; it is a metric of how "consistent" a structure is expected to be with respect to a PDB structure. Since the PDB structures AF were trained on are typically globular proteins, there is no a priori reason to assume that it will be valid for peptides. Further, McDonald et al. (2023) have shown that AF-predicted pLDDT rankings do not correlate with prediction accuracy for peptides. Therefore, the comparisons between macrocycles and the NMR structures must be all vs. all (i.e., all five AF models vs. all NMR structures) in order to reflect the real accuracy of the method.

We have added supplemental figures S2 and S3 showing the "all five models vs. all NMR structures" comparison for every example in the benchmark set (S2). We have also included a more detailed view of one outlier case observed during this analysis (S3). To address the concern that pLDDT may not be the best metric to choose a model to use for comparisons, we analyzed the high-confidence predictions (pLDDT >= 0.7) to identify cases where the highest pLDDT model is not within 1.5 Å of the experimental structure, but one of the other four models *is* within 1.5 Å to the same experimental structure. We found only seven such cases where choosing a model that was not the highest pLDDT would have given us a prediction that passed the 1.5 Å cutoff used in the manuscript. Therefore, considering all five models for comparisons would have improved our success in structure prediction from 50/80 cases to 57/80 cases for single sequence predictions. However, we note that in 6 out of the 7 such cases, the highest pLDDT models do not differ significantly from the best RMSD model as all of them are within 1.0–2 Å of the experimental structure (Figure S3). There is only one clear case with a substantial difference in prediction, where the best pLDDT model has an RMSD of 3.2 Å to the

experimental structure while another model correctly captures the experimental structure with an RMSD of 0.98 Å. We have added this analysis to the revised manuscript (line 155) and added a note recommending to consider all models for downstream work.

“Out of the 80 test cases predicted by AfCycDesign with single sequence and single recycle settings, 52 and 49 peptides were predicted close to the native structure (RMSD < 1.5 Å) and with high confidence (pLDDT > 0.7) by offset Types 1 and 2, respectively (Supplementary Figure S3C and Supplementary Table S2). These data suggested that positional encodings from residues far away in the sequence space do not significantly affect the structure prediction accuracy. To test this further, we explored the offset Type 3, where relative positional encodings were provided only for the two adjacent residues, and other relative positional encodings in the offset matrix were set to the maximum distance of 32 (Supplementary Figure S3A right panel). We observed that 48 out of the same 80 test cases were now predicted close to the native structure with high confidence, suggesting that applying a cyclic offset with relative positional encodings of adjacent residues is sufficient to predict the structure of cyclic peptides correctly.”
> The underlined interpretation is not necessarily true since this depends on the size and the fold of a macrocycle. There are at least two exceptions to the proposed conclusion. The first is that smaller macrocycles are likelier to be affected by long-range interactions more than larger macrocycles since the cyclic nature of these peptides inevitably results in a “tighter” arrangement of the residues within the scaffold. For example, in a seven-aa-long linear peptide, the residues 3 and 7 might be unrelated, but in a cyclic structure, these two residues can be within the interaction range. The second exception is stapled peptides. The dataset includes many macrocycles with disulfide bonds, and the cysteine residues, however far they may be in terms of the primary sequence, can form a disulfide bond in the tertiary structure of the macrocycle. This also applies to non-stapled structures with tightly packed tertiary folds. In order to deconvolute these factors, the authors need to provide three additional lines of data: the effect of the alternative cyclic offsets in relation to 1) the size of the tested macrocycles, 2) disulfide stapled vs. non-stapled structures, and 3) how tightly packed a structure is (perhaps SASA or a similar metric to measure).

We have now added all three additional comparisons suggested by reviewer 1 to Figure S4 in panels E-G. However, we do not observe any significant difference between Type 2 and Type 3 based on the size of the macrocycles (Figure S4E), the presence of staples in the macrocycle (Figure S4F), or the compactness of the macrocycles (Figure S4G). Notably, there is only one case that is substantially affected by the change in offset (Figure S4 legend). We discuss these data and observations in the main text (starts line 170). However, given the scenarios outlined by Reviewer 1, we use Type 2 offsets as default in AfCycDesign. We also note in the manuscript that Type 2 offset should be preferred and used with AfCycDesign.

“The sequence designed by AfCycDesign differed significantly from the Rosetta-designed sequence, with 12 mutations in the sequence and only a singular alanine in the core of the peptide being conserved.”

> This is expected since the Rosetta sequence design method is not a very accurate way of designing sequences. Why did the authors opt for Rosetta instead of a more accurate method like ProteinMPNN for their comparisons? This is later done for the hallucinated peptides, so I see no reason why the previous comparisons are confined to Rosetta.

We initially opted to compare against the Rosetta sequence design methods because ProteinMPNN had not been benchmarked for monomeric cyclic peptides at the time of the original work. In contrast, we and others had shown considerable success in designing accurate cyclic peptides with Rosetta in the past[2–7]. However, we agree with the reviewers' suggestion and have now added Figure S9 and additional analysis comparing hallucination to ProteinMPNN (line 389). We compared the *in silico* metrics for sequences designed by hallucination or ProteinMPNN for the same set of peptide backbones. While ProteinMPNN shifted the overall distribution of pLDDT higher, both methods (hallucination and ProteinMPNN) were comparable at generating high-confidence designs with pLDDT > 0.9. Interestingly, we found that different backbones benefited from different design methods. When we evaluated the backbones that resulted in high-confidence sequences for each method, only ~20% of backbones ended up with high-confidence sequences by both methods (Figure S9), suggesting ProteinMPNN and AfCycDesign may be better at designing sequences for different types of backbones. We note this in the manuscript text now and suggest designing with both methods to get more high-confidence designs.

“Many of these hallucinated sequences demonstrated promising folding propensity in these calculations, with 114 7-mers, 186 8-mers, 139 9-mers, and 76 10-mer sequences showing Rosetta Pnear values greater than 0.6. We selected one hallucinated design model per size between 7–10 amino acids with AlphaFold pLDDT > 0.9 and Rosetta Pnear > 0.9 for experimental validation and structural characterization.”
> Since the authors are using Rosetta as their gold standard, what are the Rosetta structures predicted for the tested sequences? Figure 3 provides only the predictions made by their method, which does not address questions regarding whether Rosetta is worse at folding or designing these macrocyclic structures compared to their method.

Since the designs for experimental validation were selected to have AlphaFold pLDDT > 0.9 and Rosetta Pnear > 0.9, the Rosetta structures also closely match the design models. This is shown in the structure-energy landscapes from Rosetta in Figures 3 and 4, the lowest energy structures from Rosetta are within 1 angstrom of the hallucinated models.

Notably, the hallucination is much faster and efficient at generating well-folded structures than Rosetta. We now also show the comparison of compute requirements to predict a structure using Rosetta vs. AfCycDesign in Figure S5 and discuss this comparison starting line 195. Ease of use, open source, and less compute requirements are all significant advantages of AfCycDesign over Rosetta for cyclic peptide structure prediction and design.

“For 11-mer, 12-mer, and 13-mer macrocycles, we identified 28457, 27715, and 27056 unique structural clusters, respectively (Figure 4, first column) from large-scale design calculations (see Methods).”
> I am surprised that the number of clusters is roughly the same although the increasing size should result in a larger conformational space, thus more scaffold diversity. Do the authors have an explanation for this?

We think this is due to the limited size of generated structures (N=48,000) for each size. For larger sizes, 48,000 structures is likely not enough to generate all possible structural clusters for that size comprehensively. If we were to continue to generate more hallucinated structures at those size ranges the pattern would be consistent with larger conformational space for larger sizes. We evaluated the sets of 48,000 hallucinated structures and have now included a new Figure S7, which shows that for those bigger (> 10 amino acids) macrocycles, the unique clusters had not plateaued within the 48,000 generated structures, and we are sampling many new and unique structural clusters as the sampling progressed.

“Next, we explored if our set of hallucinated peptides could be used as starting scaffolds for designing cyclic peptide binders to protein targets of interest. As a proof of concept, we chose to design binders against MDM2 given that its natural binding interface to p53 is a single helix, and our scaffold set contains multiple members with small helical motifs. We also expanded our scaffold set to sequence length up to 16 residues. We grafted a 5-residue motif from p53 (PDB: 4HFZ) on our set of hallucinated scaffolds using the MotifGraft mover from the Rosetta Software Suite 32 . Post-grafting, we designed the sequence of cyclic peptides using ProteinMPNN, followed by the energy minimization using the Rosetta energy function. For in silico filtering of design candidates, the peptide-target complex structure was predicted using AfCycDesign and down-selected based on the PAE of the interface residues (iPAE), Rosetta-based binding energy, aggregation propensity score, and contact molecular surface area 33 . We synthesized 14 designs and screened them first at a single concentration (50 μM) in an MDM2 AlphaLISA assay that measures the disruption of the interaction between MDM2 and a previously described ligand. (Figure 5). Of the 14 tested, 5 designs showed greater than 50% reduction in signal. The original 5-residue motif grafted onto our scaffold set showed no activity up to 50 μM in the same assay (Supplementary Figure S6). Next, we calculated the IC50 of the three best designs (designs D14, D15, D16) from the screening assay, and the best binder, design D14, showed an IC50 of 338.4 nM in these assays, highlighting the ability of our pipeline to identify binders even with low-throughput testing of a small number of design candidates”
> This is the most critical section of the paper since it shows the actual utility of the authors’ method, but the text is uninformative and is written in a sloppy manner contrary to the rest of the manuscript. The authors should either remove this section or elaborate. For the latter option, this section has to include a more thorough discussion of the specific interactions (and a supporting figure showing them) and how and why these designs are effective. There should also be a few sentences on existing inhibitors and

how they compare with the hallucinated peptides designed by the authors in terms of IC50 or other metrics.

We regret the lack of clarity and details in this section of the original manuscript. We have rewritten and significantly expanded this section in the revised manuscript now. Following is the list of key changes/additions to this section in the updated manuscript.

- We have now added details about the MDM2 binder and its interaction with the target. We have added text comparing the designed MDM2 binder sequence to previous binders for this target.
- We determined a high-resolution X-ray crystal structure of our best binder bound to MDM2. The experimentally determined structure closely matches the design model with an RMSD of 1.0 Å, confirming the accuracy of the design model.
- We optimized our best binder further by incorporating non-canonical amino acids, leading to a 10–fold improvement in binding affinity measured using SPR.
- Going beyond Mdm2 as a target, we demonstrate that the set of 24,104 hallucinated scaffolds can be used to scaffold 798 out of the 1,014 previously identified loop-mediated PPIs [8] to design cyclic peptide binders against hundreds of these targets.
- We have experimentally validated three cyclic peptide binders to Keap1, with IC_{50}s between 80-200 nM.

Minor
> Figure 1A needs to be redone to include the full x-axis legend.

We have redone the figure as requested.

> "IC50" should be changed into IC50 (50 subscript) in the last paragraphs of the results.

Thanks for noting it. We have now corrected it throughout the manuscript text and figures.

Reviewer #1 (Remarks on code availability):

I tested the Colab version to run the code, which seems to generate reasonable backbones.

Thank you for confirming the code availability and usability.

Reviewer #2 (Remarks to the Author):

The authors have developed an improved approach for the design of small constrained peptides and determined how design predictions are influenced by sequence. The work is well presented but fails to fully convince on the broad applicability of the approach.

We have made several additions to the revised manuscript to highlight the broad applicability of the described approach. First, we designed and validated high-affinity binders for an additional target, demonstrating the application of this approach to multiple targets now. Second, we have included data showing how the hallucinated approach can accommodate 798 loop-mediated protein-protein interfaces[8]. Third, we included X-ray structure validation and optimization of MDM2 hits with non-canonical amino acids. Further, since our original submission, the pre-print version of this manuscript has already been used and cited by many research groups. The development and experimental validation of the cyclic offset enabled the first-in-class use of deep learning for generating macrocyclic peptides in this manuscript. The conceptual advances presented here with cyclic offsets and hallucination have since inspired (and appropriately cited by) multiple other tools for peptide design by others and others. We believe the tools, scaffolds, and conceptual advances presented here will be broadly applicable for designing binders for multiple therapeutic targets and for non-therapeutic applications as well.

To back up the author claims that the new approach is faster and less computationally expensive than earlier approaches, the authors need to benchmark the approach against others and additionally consider how the different approaches influence the number of molecules that need to be synthesised to generate early leads like the example shown.

We have added Figure S5 to show the comparison of AfCycDesign vs. Rosetta cyclic peptide structure prediction simple_cycpep_predict method. This shows the time difference of 120 compute hours for Rosetta vs. 2 minutes for AfCycDesign to generate and validate a structure. Further, AfCycDesign is open source, hosted online on Google Colabs, and only requires sequence as an input compared to the significantly more complex submission using simple_cycpep_predict in Rosetta.
In the earlier approaches using Rosetta, 7.5 million structures were sampled (using the same method that took 120 hours each) to obtain ~1200 good designs of various sizes that were predicted to fold into a stable conformation[2]. While in this work, in just 48,000 sampled 10-mers (2 minutes each) we obtain ~1200 good designs based on our metrics. For binder design, the success rates we have observed are comparable to state of the art protein design[1]. The numbers tested here are substantially lower than library based methods that frequently test trillions of peptides per target. Overall, benchmarking different binder design methods is challenging given that the published data is from very different targets, and some tools may excel versus different targets.

Comparison of NMR structures not covered in AlphaFold are by their nature often restraint-underdefined but may better approximate solution structures. RMSD comparisons to the lowest energy structure to the best fitted structure need to be shown in a 2-way plot to gain deeper insight into the prediction strength of the approach. The authors need to discuss how the approach takes into account the solution structure of macrocells and ideally show how similar the determined X-ray structures of the macrocycles are theiir respective NMR structures.

The 80 peptides used for the structural prediction benchmark in the first part of the manuscript were all NMR solution structures that were not part of AlphaFold training. So, we are comparing the structure predicted by the AfCycDesign with the solution structure of those peptides. We have now also added Figure S2, which shows the comparison of every structure in the NMR ensemble to all five predicted models by NMR (also discussed starting at line 155). These, all against all comparisons, show that models predicted with high confidence capture the solution structure very closely. Since we used high-resolution crystal structures (Resolutions below 1.5 Å for each structure) to validate our hallucinated peptides, we have added a note in the discussion about possible low-abundance alternate states that may also exist and do not appear in these crystal structures.

The data in Fig. 5 is promising but the authors need to show that these modestly-potent early leads can be further optimised to prove the broad value of the approach.

We have optimized our initial hit by incorporating non-canonical and D-amino acids, improving the binding affinity by ~10-fold. We have also solved the X-ray crystal structure of the macrocycle-bound complex for the original binder, which matches very closely with the design model, confirming the accuracy of the design models and their applicability for structure-based optimization efforts.

The authors need to show the approach can be easily adapted to develop leads at related targets and conversely can be used to develop selectivity away from related targets to avoid unwanted interactions.

To validate the broad impact of this approach data in Fig. 5 should be extended to an unrelated target (eg a GPCR or ion channel) and be shown to be able to include D amino acids in the design approach.

We have added a second validated example in the revised manuscript by designing against the unrelated target Keap1, where we successfully designed an 87.7 nM inhibitor of the Keap1–Nrf2 interaction with only three designs experimentally tested. Additionally, we demonstrate that our hallucinated scaffolds can be used to mimic 798 motifs from loop-mediated protein-protein interfaces[8]. Collectively, these data show that the design methods and scaffolds generated in this work can be applied against a broad range of targets.

To ensure the approach can be broadly implanted, the supporting software, databases and actual scripts used in the present study need to be made broadly available at time of publication.

We agree with the reviewer on open and broad access. The tools and scripts used to generate the designs, including AfCycDesign and ProteinMPNN, were openly and publicly made available at the time of original submission. The tools have already been used and cited broadly by other research groups, and the conceptual advances presented here have already inspired the development of other tools for peptide structure prediction and design in the community.

Minor points:
I would prefer another term tohallucinated which suggests they were designed from a biologically altered state.

We agree with the reviewer. Unfortunately, and to our disappointment too, “hallucination” has become a commonly used term in deep learning and protein structure prediction/design field, that appropriately applies to the method implemented here. Using that term here places this work in the correct context of different DL approaches being developed in our labs and others.

Reviewer #2 (Remarks on code availability):

Unfortunately, I am not in a position to assess the code.

Working GitHub and Colab Notebooks are included and have been successfully used by other reviewers and many other researchers since the original submission. We are happy to provide any additional information.

Reviewer #3 (Remarks to the Author):

In the manuscript by Rettie et al. the authors describe a modification of the AlphaFold deep learning protein structure prediction model to better its performance on the task of predicting the structure of protegenic cyclic peptides. The authors modify the positional encoding portion of the sequence embedding portion of the AlphaFold architecture to encode the relative positions of a cyclic sequence. The pre-print of this draft manuscript and accompanying Colab notebooks represent the first published demonstration and characterization of this approach.

It should be noted that the publication of much of the code both as available github repositories as well as live and functional Colab Notebooks islaudable and democratizes many elements of this approach to practitioners in a variety of adjacent fields without the need to be able to set up a DL inference environment, something which can be challenging for computational and non-computational scientists alike.

We appreciate Reviewer 3’s remarks and supportive comments on the code availability. The open access and broad usage of these tools are very important to us, too. We are glad to see widespread usage of the publicly available AfCycDesign versions since the original submission.

The authors go on to use this modified model, dubbed AfCycDesign, as base model for a variety of structure prediction and design tasks. These tasks include validating the prediction on test of set of NMR structures in the PDB (excluded from the initial AlphaFold model training), sequence redesign where a new sequence is proposed for a known structure which engenders a lower calculated energy for the same conformation (ie a higher degree of preference for the target conformation than the parent sequence),

“hallucination” of diverse geometry and rigid cyclic peptides of varying lengths, and finally, grafting known binding motifs of MDM2 into one of the hallucinated scaffolds, using a variety of previously described DL and non-DL tools to design novel binders and the resulting validation of binding via competitive biochemical assay. These results and approaches should be of interest to the field and provide a novel sampling and design strategy for proteinogenic cyclic peptides.

While the structure prediction and hallucination design portion of the manuscript is sound and well supported by the experimental data, the MDM2 binding design portion of the manuscript has some issues that should be addressed before publication:

1 - The established AlphaLISA is a reasonable way to initially screen and quantify the disruption of the interaction, a secondary biochemical assay technique is required to validate the results. Preferably a label free technique capable of establishing the Kd of the ligands to the target such as ITC or SPR, both of which are readily available to academic laboratories in most institutions and the reported IC50s should allow for either technique to be employed.

We now include SPR data for label-free binding assays in the revised manuscript.

2 - The authors provide crystallographic evidence of designed cyclic peptide structures for the hallucinated peptides but not their resulting grafted, designed binders, this would strengthen the results substantially.

3 - Along these lines, MDM2 readily crystallizes and a number of academic and commercial labs are capable of co-crystallization of these hits with the target, while more challenging than apo structures, this would provide much more compelling evidence of a successful designed binder

We determined a high-resolution crystal structure of Mdm2-bound cyclic peptide and have added all that crystallographic data in the revised manuscript. The X-ray structure matches closely with the design model (RMSD 1.0 Å), confirming the designed macrocycle structure and binding mode.

4 - A relevant negative control such as the enantiomeric form of one of the best binders should be included.

We hope that the addition of a secondary binding assay (SPR) and X-ray crystal structure confirms the accuracy of the designed cyclic peptide binder structure and binding mode. However, if reviewer 3 thinks additional confirmation with an enantiomeric form is still required, we can add those assays to the manuscript.

5 - AMG-232’s binding isotherm should be included in the IC50 calculations

We have added the dose-response of AMG-232 to Figure S10 panel A. The calculated IC_{50} of 3.9 nM for AMG-232 in our assays matches the reported IC_{50} of 3.3 nM in the kit.

6 - A number of designs show no appreciable binding even at high concentrations, a rationale for the failure of these designs should be given, ideally supported by structural information

We now include an analysis of the working and failed designs in the revised manuscript (starting at line 470). While the success rates (5/15 for MDM2 and 3/3 for Keap1) are comparable to the design tools for larger protein binders[1], the key issue in our failures is the sequence design step. The best three MDM2 binding designs capture the established motif FXXXWXX{L/M}. Designs that showed no or low inhibition in the screen (Figure S10) also include similar structural motifs but have a Phe or Ile at the Leu/Met position. Additional factors like solubility may also be contributing to failures. However, we believe that improvements to sequence design methods are required to achieve the truly single-shot design of peptide binders.

7 - While the cyclic peptide scaffold is cited as being advantageous from a stability and biophysical property standpoint, neither are assessed on the resulting scaffolds which are very unlikely to have any cell permeability. This should either be assayed and included or clarified in the text.

We now clarify this in the text (line 589).

8 - The linear p53 motif that is used as a comparator for the cyclic scaffolds is much smaller than the cyclic peptides, using an equivalent length of residues to the cyclic peptide would be a more fair comparison and would likely show similar binding affinity to the best designed compounds based on previously affinities for reported p53/MDM2 FP assays.

We apologize for the lack of clarity on this point in the original submission. Our rationale for including the five residue motifs as a comparison/control was to evaluate if the binding observed from our designs was coming solely from the motif used for grafting and if the design steps were actually contributing to achieving improved target binding. The lack of binding from the motif alone vs. the dose-dependent binding displayed by the designed binders supports the benefit of computational design in this case.

9 - While this approach should be of value to the field, it is reliant on a known binding motif, compatible scaffold, grafting to said scaffold, and resulting reoptimization. Why was a hallucinated binder approach not tried as well or was it tried but failed due to insufficient sampling?

While the publicly available AfCycDesign allows for hallucinating a binder, we found that *de novo* hallucination of binders currently is computationally expensive and requires extensive

sampling. We believe the development of an optimized, efficient, and extensively validated *de novo* binder tool in AFCycDesign is outside of the scope of this work and will be a focus of future work.

Reviewer #3 (Remarks on code availability):

As mentioned in the remarks to author, the inclusion of working Colab notebooks is a great addition to the code availability.

Thank you for the supportive comments on code sharing and availability.

1. Watson, J. L. *et al.* De novo design of protein structure and function with RFdiffusion. *Nature* **620**, 1089–1100 (2023).
2. Hosseinzadeh, P. *et al.* Comprehensive computational design of ordered peptide macrocycles. *Science* **358**, 1461–1466 (2017).
3. Bhardwaj, G. *et al.* Accurate de novo design of membrane-traversing macrocycles. *Cell* **185**, 3520–3532.e26 (2022).
4. Hosseinzadeh, P. *et al.* Anchor extension: a structure-guided approach to design cyclic peptides targeting enzyme active sites. *Nat. Commun.* **12**, 3384 (2021).
5. Mulligan, V. K. *et al.* Computationally designed peptide macrocycle inhibitors of New Delhi metallo-β-lactamase 1. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
6. Mulligan, V. K. *et al.* Computational design of mixed chirality peptide macrocycles with internal symmetry. *Protein Sci.* **29**, 2433–2445 (2020).
7. Said, M. Y. *et al.* Exploration of structured symmetric cyclic peptides as ligands for metal-organic frameworks. *Chem. Mater.* **34**, 9736–9744 (2022).
8. Gavenonis, J., Sheneman, B. A., Siegert, T. R., Eshelman, M. R. & Kritzer, J. A. Comprehensive analysis of loops at protein-protein interfaces for macrocycle design. *Nat. Chem. Biol.* **10**, 716–722 (2014).