

## **Reviewer 2**

### **Comment 1**

***Original comment:** However, coverage of the most consequential 2022–2025 methodological advances is incomplete for a 2026 survey of this scope: RFdiffusion, RFpeptides, ProteinMPNN, ProtGPT2, ESM-2/ESM-3, AlphaFold3, ProGen/ProGen2, PepFlow and PepGLAD are absent from the abstract, Table 4 contents, and section summaries, while the representative-method list reads as a convenience sample rather than a systematic selection. With only 86 references across 8 classes, 5 representation families, workflows, datasets and a 16-domain checklist, the evidentiary base is too thin for the claimed breadth.*

**Response:** We thank the reviewer and agree that the original version did not make either the methodological coverage or the basis for selecting representative methods sufficiently explicit. RFpeptides, RFdiffusion, ProteinMPNN, and AlphaFold 3 were already discussed in the main text, and the corresponding primary papers for ESM-2 and ESM3 were already cited as References 63 and 64, but their coverage was not sufficiently visible in the Abstract, comparative table, or section summaries. We have therefore revised the Abstract and Sections 4.1, 7, and 12 to integrate ProtGPT2, ESM-2/ESM3, ProGen/ProGen2, ProteinMPNN, RFdiffusion, AlphaFold 3, RFpeptides, PepFlow, and PepGLAD according to their actual methodological roles. The submitted Table 4, now Table 5 following the addition of the new Table 2 requested by Reviewer 1, has also been expanded so that all of these methods are represented. Importantly, the revised text distinguishes general protein foundation, inverse-folding, and structure-prediction models from methods developed specifically for peptides; protein-level validation is not presented as evidence of performance on short, flexible, or chemically modified therapeutic peptides. We have also added an explicit scope paragraph stating that this is a critical narrative rather than a systematic review and defining the criteria used to select representative studies. The revised manuscript now contains 105 references rather than the 86 in the submitted version, while the selection logic and the evidentiary limits of each method family are now explicit.

Location in the revised manuscript: Abstract; Section 1; Sections 4.1, 7, and 12; Table 5; References

67–69 and 86–87.

## Comment 2

**Original comment:** *The “Peptide-AI Evidence Grid” is a useful reframing but largely a repackaging of DOME/TRIPOD+AI/TRL-style maturity concepts, and its novelty should be calibrated accordingly.*

**Response:** We agree and thank the reviewer for this important clarification. We have revised the manuscript to avoid presenting the Peptide-AI Evidence Grid as a wholly new general framework for reporting or maturity assessment. The revised text now states explicitly that the Grid adapts established reporting and validation principles and the stage-based logic of Machine Learning Technology Readiness Levels (MLTRL). We also define its narrower contribution: (i) peptide-specific operationalisation; (ii) orthogonal assessment of computational rigour (M0-M3) and experimental/translational depth (T0-T3); and (iii) claim-level assignment, so that strength on one axis cannot compensate for weakness on the other. MLTRL follows the maturity of an ML system from research through productisation and deployment, whereas the T axis of the Grid grades evidence accrued for the peptide molecule. We have therefore recalibrated the wording in the Abstract, Introduction, Section 9, Section 10.1, and Conclusions and now describe the Grid as an operational peptide-specific synthesis rather than as a new universal maturity framework. The peer-reviewed MLTRL framework paper has been added as Reference 100. We also refined the wording of T0 and T2 so that the translational axis applies consistently to predictive and generative studies, therapeutic candidates, and class-specific material-design evidence.

**Location in the revised manuscript:** Abstract; Section 1; Section 9; Section 10.1; Section 12; References 59, 99, and 100.

## Comment 3

**Original comment:** *The core framework is sound and salvageable; the literature integration and novelty claims require more revision.*

**Response:** We thank the reviewer for this balanced assessment. We have retained the core framework and revised precisely the two areas identified. As detailed in our responses to Comments 1 and 2 above, we strengthened literature integration by expanding coverage of major 2022–2025 protein- and peptide-design methods, defining explicit selection criteria, expanding Table 5, and increasing the reference base from 86 to 105. We also recalibrated the novelty claim by positioning the Peptide-AI Evidence Grid as a peptide-specific operational synthesis of established reporting and maturity principles, rather than as a wholly new universal framework.

**Location in the revised manuscript:** Abstract; Section 1; Sections 4.1, 7, 9, 10.1, and 12; Table 5;

References 59, 67–69, 86–87, and 99–100.

#### Comment 4

**Original**

**comment:**

(2)Questions

:

(i) What inclusion criteria governed the selection of methods in Tables 3–4 (HELM-GPT/CycloPepper included vs RFdiffusion/ProteinMPNN excluded)?

Response: The submitted Tables 3–4 were intended to be representative rather than exhaustive, but the criteria were not stated clearly enough. The revised manuscript now defines them explicitly in Section 1: peer-reviewed studies available at the September 2026 cut-off were selected to span therapeutic classes, representation and model families, workflow types, and evidence levels, while prioritising peptide-specific methods, direct enabling technologies, sufficient reporting for M/T assessment, and prospective validation where available. HELM-GPT and CycloPepper were retained because they address peptide-specific macrocycle representation and generation and experimentally measured cyclization, respectively. RFdiffusion and ProteinMPNN are general protein-design methods, which explains why they were not originally treated as peptide-specific exemplars; however, we agree that their enabling role should be visible. Revised Table 5 (submitted Table 4; renumbered after addition of the new Table 2) now includes them in a separate contextual row, without a peptide-specific M/T score (n.a./n.a.), and Section 7 identifies RFpeptides as an adaptation of the RFdiffusion–ProteinMPNN lineage.

Location in the revised manuscript: Section 1; Section 7; revised Table 5 and its note.

#### Comment 5

**Original comment:** (ii) How do you position the grid against maturity/validation frameworks beyond DOME and TRIPOD+AI ?

**Response:** We thank the reviewer for this question. Beyond DOME and TRIPOD+AI, the revised manuscript now positions the Peptide-AI Evidence Grid explicitly in relation to Machine Learning Technology Readiness Levels (MLTRL) [100], reproducibility standards for machine learning in the life sciences [60], and the FAIR principles [105]. These frameworks address different but complementary dimensions. Reproducibility standards define the artefacts needed to make an ML result repeatable, while FAIR concerns the stewardship and reuse of data and models. MLTRL is the closest maturity comparator, but it follows the ML system from research through productisation and deployment. By contrast, the Grid separates methodological validation of the computational model (M axis) from the experimental and translational evidence accumulated for the peptide molecule (T axis). Its contribution is therefore not a replacement maturity framework, but a

peptide-specific, claim-level operationalisation that prevents software maturity, computational performance, and molecule-level evidence from being conflated.

**Location in the revised manuscript:** Sections 9 and 10.1; References 60, 100, and 105.

**Comment 6**

**Original comment:** (iii). *Can you substantiate “the M3 column is almost empty” with an explicit tally?*

**Response:** We agree that the original wording was qualitative. We now provide an explicit tally derived from the discrete evidence assignments in Tables 4 and 5. Among the 25 peptide-specific method-evidence entries assigned an M/T pair, none reached M3 (0/25): 0/11 in Table 4 and 0/14 in Table 5. We replaced the qualitative statement in Section 9.1 with this scoped count, revised the corresponding Table 7 cell, and report the same in the Conclusions. Because this is intended as a critical narrative rather than a systematic review, the revised wording refers specifically to the representative entries evaluated here and does not imply a field-wide prevalence estimate.

**Location in the revised manuscript:** Section 9.1 and Table 7; Section 12; Tables 4–5.