

# SEMAGLUTIDE — C-TERMINAL TRUNCATION: DELETE GLY-31, YIELDING A 30-RESIDUE PEPTIDE ENDING IN ...VRGRG → ...VRGR. NATIVE AIB-2 AND LYS-20 (ΓGLU-ΓGLU-C18 DIACID) LIPIDATION ARE PRESERVED.

generated 2026-05-04T22:59:07.550163+00:00

PROMISING

METABOLIC

C-TERMINAL TRUNCATION: DELETE GLY-31, YIELDING A 30-RESIDUE PEPTIDE ENDING IN ...VRGRG → ...VRGR. NATIVE AIB-2 AND LYS-20 (ΓGLU-ΓGLU-C18 DIACID) LIPIDATION ARE PRESERVED.

GLUCAGON-LIKE PEPTIDE 1 RECEPTOR

AVERAGE CONFIDENCE	PTM / IPTM	VERDICT
<b>77.7%</b>	0.841 / 0.924	PROMISING
TARGET	UNIPROT	BINDING PROBABILITY
Glucagon-like peptide 1 receptor	P43220	—

## TLDR

Fold №75 tested whether removing the C-terminal Gly-31 from semaglutide would yield a 30-residue analog retaining full GLP-1R engagement, with the literature strongly supporting the dispensability of this residue given that endogenous GLP-1(7-36) amide lacks it and remains pharmacologically active. The structural prediction returned pLDDT 0.78 and strong interface confidence metrics (ipTM 0.92), which are not the problem — the fold was discarded because lipidated peptides of this class remain outside reliable AlphaFold-family resolution due to the non-canonical C18 diacid moiety. This marks the fourth semaglutide fold discarded in the lab (Folds #15, #36, #52, #75), consistently reflecting tool-limit failures rather than

biological invalidation of the underlying hypotheses. The biological rationale for this truncation remains plausible and merits wet-lab evaluation.

## EXECUTIVE SUMMARY

---

Fold №75 tests Gly-31 deletion from semaglutide — biologically plausible given GLP-1(7-36) amide precedent — but is discarded for the fourth time on this scaffold: C18 diacid lipidation remains outside Boltz-2's chemical resolution, making interface metrics uninterpretable. Hypothesis intact; tools insufficient.

## DETAILED ANALYSIS

---

Semaglutide is a 31-residue GLP-1 receptor agonist built on the GLP-1(7-37) backbone, distinguished from native GLP-1 by an Aib substitution at position 2 conferring DPP-4 resistance, an Arg-34 substitution preventing proteolysis at that locus, and a C18 fatty diacid appended via a  $\gamma$ Glu- $\gamma$ Glu-miniPEG linker at Lys-26 (Lys-20 by semaglutide internal numbering) to enable high-affinity albumin binding and extend plasma half-life to approximately 165 hours. Fold №75 asks a targeted pharmacological question: is the terminal Gly-31 — corresponding to Gly-37 of proglucagon — functionally dispensable, and can its removal yield a 30-mer with equivalent or improved properties?

The biological rationale is exceptionally well-grounded. The dominant circulating form of endogenous GLP-1 is GLP-1(7-36) amide, a 30-residue peptide that naturally lacks the Gly-37 extension and is fully active at the GLP-1 receptor. This means the Gly-31 in semaglutide recapitulates a residue that evolution has already shown to be dispensable for GLP-1R engagement — it is present in semaglutide only because the drug was developed from the GLP-1(7-37) isoform rather than the amide. Cryo-EM and crystallographic structural data on GLP-1R agonist complexes consistently show that the ECD contact surface terminates around position 24-26 of GLP-1, with C-terminal residues 35-37 projecting into solvent without making direct receptor contacts. The truncation hypothesis is therefore structurally motivated, not speculative.

A secondary hypothesis concerns carboxypeptidase susceptibility. Removing Gly-31 exposes a new C-terminal Arg-30, which is a potential substrate for plasma carboxypeptidases N and M. The literature review notes that native GLP-1(7-36) amide avoids this vulnerability through C-terminal amidation; the proposed analog presents a free C-terminal Arg, which introduces a genuine pharmacokinetic risk. However, the literature also makes clear that semaglutide's half-life is overwhelmingly determined by albumin binding via the lipid chain, which sterically shields the peptide from protease access. The carboxypeptidase risk exists but is likely modest in the context of the heavily lipidated, albumin-bound molecule.

The structural prediction produced pLDDT 0.78, pTM 0.84, and ipTM 0.93 — metrics that, considered in isolation, would suggest a confident complex model. However, the discard verdict reflects a known and persistent limitation: Boltz-2 and Chai-1 cannot model the C18 diacid lipidation chemistry with chemical fidelity. The fatty acid chain,  $\gamma$ Glu- $\gamma$ Glu-miniPEG linker, and albumin-bound conformation collectively define semaglutide's dominant solution-state geometry, and none of these are represented in the structure prediction. The 3D output models the peptide backbone in an idealized aqueous context that does not reflect the conformational landscape of the lipidated drug, making any binding interface assessment unreliable regardless of the confidence scores.

This result is the fourth consecutive semaglutide discard in the Alembic lab. Fold #52 ( $\alpha$ Me-His at position 1, pLDDT 0.72) and Fold #15 (homoglutamate at Glu-16, pLDDT 0.71) tested backbone and side-chain modifications; Fold #36 ( $\beta$ -Ala- $\beta$ -Ala spacer replacement, pLDDT 0.70) directly probed the linker chemistry. All four were blocked by the same fundamental limitation: semaglutide's pharmacological identity is inseparable from its lipidation, and current generative structure prediction tools cannot model this. The pattern is consistent and instructive — it points toward a class-level tool gap rather than any problem with individual modification hypotheses.

For this specific truncation, the most informative next step is a competitive radioligand binding assay comparing the 30-mer analog against full-length semaglutide at the GLP-1R, combined with a cAMP dose-response curve to confirm functional agonism. These experiments are straightforward, well-validated in the GLP-1 analog literature, and would definitively answer whether Gly-31 is dispensable for binding affinity and efficacy. A secondary carboxypeptidase stability assay in plasma would address the C-terminal Arg vulnerability question. Molecular dynamics simulation using a force-field parameterized for the full lipidated peptide (including albumin docking) would provide the computational structural insight that current fold predictors cannot.

The heuristic sequence-based profile (aggregation propensity 0.14, stability 0.53, half-life moderate-to-long) is noted for transparency but carries minimal interpretive weight for a lipidated peptide of this class — these estimates are derived from sequence alone and do not account for the albumin-binding pharmacokinetics that dominate semaglutide's in vivo behavior. The biological hypothesis underlying Fold №75 remains scientifically sound; the discard reflects only the limits of the prediction toolkit, not the limits of the idea.

## RESEARCH BRIEF

---

# FOLD №75 — SEMAGLUTIDE C-TERMINAL GLY-31 TRUNCATION

**Verdict: DISCARDED (tool-limit failure — lipidated peptide outside current predictor resolution)**

---

## TLDR

Fold №75 was **DISCARDED** because the dominant structural and pharmacokinetic determinant of semaglutide — its C18 diacid lipidation at Lys-20 — cannot be modelled with chemical fidelity by Boltz-2 or Chai-1, rendering the binding interface prediction unreliable regardless of backbone confidence scores. This is a **tool-limit failure**, not a biological invalidation. The underlying hypothesis — that Gly-31 is dispensable for GLP-1R engagement — is strongly supported by existing pharmacological literature (endogenous GLP-1[7-36] amide is the dominant active form and naturally lacks this residue) and remains testable by wet-lab methods.

---

## WHAT WE TRIED

Semaglutide is a 31-residue GLP-1 receptor agonist based on GLP-1(7-37), distinguished from the native peptide by Aib-2 (DPP-4 resistance), Arg-34 (proteolytic stability), and a C18 fatty diacid at Lys-26 via a  $\gamma$ Glu- $\gamma$ Glu-miniPEG linker conferring albumin binding and a ~165-hour plasma half-life. Fold №75 asked whether the terminal Gly-31 — corresponding to Gly-37 of proglucagon — could be deleted to yield a 30-residue analog (HAEGTFTSDVSSYLEGQAAKEFIAWLVRGR) while preserving full GLP-1R engagement.

The hypothesis was motivated by two independent lines of reasoning. First, structural biology of GLP-1R agonist complexes places the ECD contact surface at positions  $\leq 26$  of GLP-1, with the GRG tail extending into solvent without direct receptor contacts. Second, and more compellingly, the dominant circulating form of endogenous GLP-1 is GLP-1(7-36) amide — a 30-residue peptide that naturally terminates at the equivalent of semaglutide's Arg-30 and is fully pharmacologically active. The C-terminal Gly-31 in semaglutide is therefore an artefact of the GLP-1(7-37) template used in drug development, not a biological necessity for receptor engagement. A secondary hypothesis proposed that exposing C-terminal Arg-30 might modestly alter carboxypeptidase susceptibility and clearance kinetics, though the albumin-bound lipidated fraction was expected to remain largely shielded.

---

## WHY IT WAS DISCARDED

The structural prediction produced pLDDT 0.78, pTM 0.84, and ipTM 0.93 — numerically acceptable values that would ordinarily suggest a confident complex model. However, these metrics reflect only the backbone confidence of the unlipidated peptide sequence; they do not and cannot capture the conformational geometry of semaglutide as it actually exists in solution or in its receptor-bound state. The C18 diacid lipid chain,  $\gamma$ Glu- $\gamma$ Glu-miniPEG spacer, and albumin-engaged conformation collectively define the pharmacologically relevant structure of semaglutide, and none of these chemical moieties are represented in the Boltz-2/Chai-1 prediction. The output models an unlipidated peptide backbone docked to GLP-1R — a physically unrealistic scenario that cannot support conclusions about binding affinity, interface geometry, or conformational stability of the actual drug.

This is the same tool-limit failure that discarded Folds **#15** (Glu-16  $\rightarrow$  homoglutamate, pLDDT 0.71), **#36** ( $\beta$ -Ala- $\beta$ -Ala spacer replacement, pLDDT 0.70), and **#52** ( $\alpha$ Me-His at position 1, pLDDT 0.72). The pattern across four semaglutide folds is consistent: the backbone prediction scores are borderline-acceptable, but the lipidation chemistry that governs the molecule's conformational ensemble and pharmacological identity falls outside the chemical space these tools can resolve. Chai-1 agreement data was not returned for this fold, removing a secondary confidence check.

---

## WHAT THIS DOESN'T MEAN

**DISCARDED does not mean disproved.** This verdict reflects a tool-limit failure — the current prediction infrastructure cannot model lipidated peptides of this chemical complexity with sufficient fidelity to adjudicate binding hypotheses. The biological rationale for the Gly-31 truncation is independently supported by decades of GLP-1 pharmacology: GLP-1(7-36) amide, which endogenously lacks the terminal Gly, is the dominant active circulating form and engages GLP-1R with equivalent potency to GLP-1(7-37). No evidence in the literature contradicts the core hypothesis, and no structural data places Gly-31 within the receptor binding footprint. The carboxypeptidase concern (C-terminal Arg exposure) is a real but likely modest pharmacokinetic risk in the context of the albumin-bound lipidated molecule. The hypothesis is scientifically sound; it simply requires tools and assays better suited to the chemistry than current in silico predictors.

---

## WHAT WOULD ANSWER THE QUESTION

- **Competitive radioligand binding assay (GLP-1R):** Direct IC<sub>50</sub> comparison of the 30-mer analog vs. full-length semaglutide using [<sup>125</sup>I]-GLP-1 or fluorescent tracer displacement at HEK293 cells overexpressing GLP-1R — the gold-standard

assay for GLP-1 analog affinity, well-established in the discovery literature (Lau et al., 2015 framework).

- **cAMP dose-response (functional agonism):** EC<sub>50</sub> comparison via HTRF or BRET cAMP assay to confirm that Gly-31 deletion does not impair receptor activation, independent of binding affinity.
- **Plasma carboxypeptidase stability assay:** Incubation of the 30-mer analog in pooled human plasma at 37°C with LC-MS/MS monitoring for des-Arg-29 and des-Arg-28 cleavage products — directly tests the C-terminal Arg vulnerability hypothesis.
- **Full-atom MD simulation with explicit lipidation:** Force-field parameterisation of the complete lipidated 30-mer (AMBER or CHARMM with custom GAFF2 parameters for the C18 diacid and  $\gamma$ Glu- $\gamma$ Glu linker) docked to the cryo-EM GLP-1R structure, with explicit solvent and optional albumin inclusion — the appropriate computational tool for this chemistry class, as distinct from generative fold predictors.

---

## RAW METRICS

Metric	Value
pLDDT	0.777
pTM	0.841
ipTM	0.924
Chai-1 agreement	Not returned
Boltz-2 affinity	Not returned
Predicted binding change	Not determined
Aggregation propensity (heuristic)	0.142
Stability score (heuristic)	0.533
BBB penetration (heuristic)	0.05
Half-life estimate (heuristic)	Moderate-to-long (~1–6 h, sequence-only; not reflective of lipidated PK)

Heuristic properties are sequence-based estimates only and do not account for the C18 diacid lipidation that dominates semaglutide's actual pharmacokinetic profile.

---

**Lab context:** This is the fourth semaglutide fold discarded due to lipidation-related tool limits (see also Fold #15, Fold #36, Fold #52). The pattern strongly suggests that semaglutide modifications require either wet-lab validation or purpose-built MD

workflows with explicit lipid parameterisation before in silico verdicts can be trusted. The biological hypotheses across all four folds — including this truncation — remain independently plausible.

## SEQUENCES

---

### NATIVE

```
HAEGTFTSDVSSYLEGQAAKEFIAWLVRGRG
```

### MODIFIED

```
HAEGTFTSDVSSYLEGQAAKEFIAWLVRGR
```

## CAVEATS

---

- in silico prediction only — requires wet lab validation
- single-run prediction (not ensembled)
- predicted properties may not reflect real-world biological behavior
- this is research, not medical advice
- lipidated peptides (C18 diacid,  $\gamma$ Glu- $\gamma$ Glu-miniPEG linker) are outside reliable AlphaFold-family chemical resolution — backbone pLDDT/ipTM values do not reflect the conformational ensemble of the actual drug
- Chai-1 agreement data was not returned for this fold, removing secondary confidence validation
- heuristic half-life estimate (~1-6 h) reflects unlipidated sequence only and grossly underestimates semaglutide's albumin-mediated ~165 h half-life
- carboxypeptidase susceptibility of C-terminal Arg-30 in the lipidated, albumin-bound context is not experimentally characterised and cannot be assessed in silico with current tools
- DISCARDED verdict indicates tool-limit failure, not biological invalidation of the truncation hypothesis
- Verdict reclassified: DISCARDED → PROMISING. Raw metrics (pLDDT/pTM/ipTM) permit at least the higher tier; the original LLM discard reflected modification chemistry the predictor cannot represent (D-AA, lipid moiety, non-canonical residue). Per the metric-floor rule this is a caveat, not a verdict downgrade. Report text below pre-dates the rule and may still describe the fold as DISCARDED — the structural verdict shown is the authoritative one.

## CITATIONS

---

1. **PMID** — (2015) — — Discovery of the Once-Weekly Glucagon-Like Peptide-1 (GLP-1) Analogue Semaglutide
2. **PMID** — (2024) — — Clinical Pharmacokinetics of Semaglutide: A Systematic Review
3. **PMID** — (2020) — — Semaglutide lowers body weight in rodents via distributed neural pathways
4. **PMID** — (2021) — — Safety of Semaglutide
5. **PMID** — (2023) — — Semaglutide for the treatment of obesity
6. **PMID** — (2021) — — Semaglutide 2·4 mg once a week in adults with overweight or obesity, and type 2 diabetes (STEP 2)

SOLANA SIGNATURE 45y6fyoCMajcmia8cumpkL9ha5uQE9kFEsC1KfdgRZnCLU7ajQTC6VdgA  
UzpACwtykn8gAEbwnJQ6psBcZJGDHfu  
DATA SHA-256 9b5b68378543fe019041fad0adca23811459bd6963fea5e98fa8cf5736bbca61  
VERIFY [https://solscan.io/tx/  
45y6fyoCMajcmia8cumpkL9ha5uQE9kFEsC1KfdgRZnCLU7ajQTC6VdgAUzpACwtykn8gAEbwn  
JQ6psBcZJGDHfu](https://solscan.io/tx/45y6fyoCMajcmia8cumpkL9ha5uQE9kFEsC1KfdgRZnCLU7ajQTC6VdgAUzpACwtykn8gAEbwnJQ6psBcZJGDHfu)