SEMAGLUTIDE — N-TERMINAL CONJUGATION OF A SHORT DISCRETE PEG2 (8-AMINO-3,6-DIOXAOCTANOIC ACID, AEEA) SPACER TO THE A-AMINE OF HIS-1 VIA A STABLE AMIDE BOND, YIELDING AEEA-HAEGTFTSDVSSYLEGQAAK(ΓGLU-ΓGLU-C18DIACID)EFIAWLVRGRG. NATIVE AIB-2 AND LYS-20 LIPIDATION ARE PRESERVED.

generated 2026-05-05T04:34:49.933371+00:00

DISCARDED | METABOLIC

N-TERMINAL CONJUGATION OF A SHORT DISCRETE PEG2 (8-AMINO-3,6-DIOXAOCTANOIC ACID, AEEA) SPACER TO THE A-AMINE OF HIS-1 VIA A STABLE AMIDE BOND, YIELDING AEEA-HAEGTFTSDVSSYLEGQAAK(ΓGLU-ΓGLU-C18DIACID)EFIAWLVRGRG. NATIVE AIB-2 AND LYS-20 LIPIDATION ARE PRESERVED.

GLUCAGON-LIKE PEPTIDE 1 RECEPTOR

AVERAGE CONFIDENCE	PTM / IPTM	VERDICT
68.4%	0.657 / 0.463	DISCARDED
TARGET	**UNIPROT**	**BINDING PROBABILITY**
Glucagon-like peptide 1 receptor	P43220	—

TLDR

Fold №81 tested whether appending a discrete PEG2 (AEEA) spacer to the α-amine of His-1 in semaglutide could improve oral/intranasal bioavailability by shielding the

N-terminus from luminal aminopeptidases, while preserving GLP-1R engagement via the intact His-1 imidazole. The structural prediction returned an ipTM of 0.46 — below the threshold required for confident complex modeling — meaning the receptor-bound pose of the N-terminally extended peptide could not be reliably adjudicated by current in silico tools. This is a tool-limit discard, not a biological invalidation: the hypothesis is mechanistically grounded and chemically novel, but the alpha-amine modification of His-1 represents a pharmacophoric perturbation that AlphaFold-family models are not well-equipped to evaluate for Class B GPCR binding. The delivery rationale remains scientifically compelling and warrants wet-lab investigation.

EXECUTIVE SUMMARY

Fold №81 (DISCARDED): AEEA-PEG2 N-terminal conjugation of semaglutide could not be structurally adjudicated — ipTM 0.46, tool-limit failure on non-canonical linker chemistry. The delivery rationale is sound; wet-lab GLP-1R functional assay needed.

DETAILED ANALYSIS

Semaglutide is a benchmark GLP-1 receptor agonist engineered for once-weekly subcutaneous dosing through two structural innovations: an Aib substitution at position 2 (blocking DPP-IV cleavage) and a γGlu-γGlu-C18 fatty diacid chain at Lys-26 (conferring high albumin affinity and a ~46-hour half-life). Despite these modifications, oral semaglutide (Rybelsus) achieves only ~1% bioavailability even in the presence of the absorption enhancer SNAC, a limitation driven in part by luminal aminopeptidase degradation of the exposed His-1 N-terminus. Fold №81 addresses this bottleneck directly by proposing the conjugation of a discrete PEG2 (AEEA, 8-amino-3,6-dioxaoctanoic acid) spacer to the α-amine of His-1 via a stable amide bond — a delivery-focused N-terminal modification not previously explored in this lab's semaglutide series.

The modification rationale is specific and layered. AEEA is a short (~10 Å), rigid-flexible discrete PEG unit that is substantially less immunogenic and sterically disruptive than polymeric PEG chains. By masking the α-amine of His-1 with an amide bond to AEEA, the hypothesis predicts that the principal aminopeptidase recognition site is occluded, reducing the rate of N-terminal proteolytic cleavage in gastric and mucosal environments. The AEEA unit is short enough to project away from the orthosteric binding pocket into bulk solvent, leaving the His-1 imidazole side chain — the critical pharmacophore for TM6 engagement in GLP-1R — geometrically accessible. The Aib-2 and Lys-26 lipidation elements are preserved unchanged, meaning albumin binding and DPP-IV resistance should be unaffected by the N-terminal addition.

However, there is a fundamental chemical tension at the core of this hypothesis. Cryo-EM structures of GLP-1 bound to GLP-1R consistently show that the α-amine of His-1 forms a direct hydrogen bond to Glu387 in TM6. The AEEA amide conjugation replaces the primary α-amine (pKa ~8, positively charged at physiological pH, strong H-bond donor) with a secondary amide nitrogen (non-protonatable, weakened H-bond donor capacity). This is not a trivial perturbation — it removes a charged pharmacophoric element that makes a specific polar contact in the transmembrane bundle. The literature supports the view that the mid-helical and C-terminal regions of GLP-1 engage the ECD, while the N-terminal residues 1-6 are indispensable for receptor activation: this has been called the 'two-domain' model of GLP-1R activation. Modifications distal to His-1 (as in the Lys-26 lipidation) are tolerated; modifications at His-1 itself are high-risk for potency.

The structural prediction from Boltz-2 returned a pLDDT of 0.68, a pTM of 0.66, and critically an ipTM of 0.46. The ipTM score is the primary metric for complex confidence in AlphaFold-family models, and values below ~0.50 are generally considered unreliable for structural interpretation. No Chai-1 ensemble agreement was obtained, and the Boltz-2 affinity module did not return binding change values. These metrics together constitute a tool-limit failure rather than a biological verdict: the model could not confidently place the AEEA-extended N-terminus relative to the GLP-1R binding interface, likely because (a) the AEEA linker introduces non-canonical chemistry that is poorly represented in training data, and (b) the N-terminal binding geometry in Class B GPCRs involves a narrow transmembrane cavity that is difficult to model with high confidence even for native peptides. The heuristic sequence-based profile (aggregation propensity 0.19, stability score 0.20, BBB penetration 0.11) is consistent with a hydrophilic, albumin-bound peptide with low CNS penetration — unremarkable for this class.

This fold connects meaningfully to prior semaglutide work in the lab. Fold №52 tested α-methyl-L-histidine at position 1 — a Cα-methylation that rigidifies the N-cap conformation while preserving the imidazole — and returned a PROMISING verdict (pLDDT 0.72, ipTM above threshold). The contrast is instructive: Cα-methylation at His-1 modifies the backbone geometry without altering the α-amine charge state, whereas AEEA conjugation converts the α-amine into an amide, a chemically more disruptive change. Fold №75 tested C-terminal truncation (PROMISING, pLDDT 0.78), and Fold №15 tested a mid-helix homoglutamate substitution (PROMISING, pLDDT 0.71) — both modifications distal to the receptor-critical N-terminus and both evaluable by current tools. Fold №36 was discarded (pLDDT 0.70, poor ipTM) for the β-Ala linker substitution at the Lys-26 lipidation site, another modification involving non-canonical linker chemistry that challenged structural prediction tools.

The clinical agent context is important: semaglutide is already approved for both subcutaneous and oral routes, and the PIONEER trials have demonstrated meaningful glycemic and weight outcomes with oral dosing despite the low bioavailability. The incremental benefit of N-terminal AEEA modification must therefore be weighed against the real risk of potency reduction at GLP-1R — a

concern amplified by the fact that oral semaglutide is dosed at 14 mg (vs. 0.5-2 mg for subcutaneous) precisely because bioavailability is low. If AEEA conjugation reduces receptor affinity by even 5-10 fold (plausible given α-amine conversion to amide), the formulation advantage could be offset.

This fold is discarded as a tool-limit result. The hypothesis remains scientifically novel — no published study has assessed N-terminal AEEA conjugation on semaglutide or any close GLP-1 analogue — and the delivery rationale is well-founded. What is needed is not better computational prediction but wet-lab functional assay: a cAMP accumulation assay or GLP-1R binding competition assay on the AEEA-His-1 conjugate would definitively answer whether the α-amine-to-amide conversion ablates, reduces, or preserves receptor activation. The AEEA peptide can be synthesized by standard Fmoc SPPS with an AEEA building block, making this experimentally accessible.

RESEARCH BRIEF

TLDR

Fold №81 was **DISCARDED** due to insufficient complex-modeling confidence (ipTM 0.46), a tool-limit failure driven by the non-canonical AEEA linker chemistry and the difficulty of modeling N-terminal Class B GPCR contacts in silico. This is **not** a biological invalidation of the hypothesis.

WHAT WE TRIED

This fold asked whether appending a discrete PEG2 spacer — specifically AEEA (8-amino-3,6-dioxaoctanoic acid) — to the α-amine of His-1 in semaglutide could improve oral and intranasal bioavailability by shielding the N-cap from luminal aminopeptidases, while preserving GLP-1R engagement through the intact His-1 imidazole side chain. The modified sequence is **AEEA-H-Aib-EGTFTSDVSSYLEGQAAK(γGlu-γGlu-C18-diacid)EFIAWLVRGRG**, retaining the native Aib-2 DPP-IV block and the Lys-26 C18 fatty diacid albumin anchor entirely unchanged.

The mechanistic rationale is that AEEA (~10 Å, flexible-rigid) is short enough to project into bulk solvent from the His-1 α-amine without steric intrusion into the GLP-1R transmembrane binding pocket, while the amide conjugation to the α-amine blocks the primary aminopeptidase recognition site responsible for the low (~1%) oral bioavailability of Rybelsus. This represents a delivery-focused N-terminal modification — a combination of focus and modification category not previously explored in the lab's semaglutide series (Folds #15, #36, #52, #75 tested mid-

helix, lipid linker, N-cap rigidification, and C-terminal truncation strategies respectively).

WHY IT WAS DISCARDED

The Boltz-2 structural prediction returned **ipTM 0.46** — below the ~0.50 threshold required for reliable complex-mode interpretation in AlphaFold-family models. No Chai-1 ensemble was obtained for independent agreement, and the Boltz-2 affinity module did not produce a binding change value. The pLDDT of 0.68 for the peptide chain itself is borderline acceptable, but without a confident interface score the receptor-bound pose of the AEEA-extended N-terminus cannot be interpreted structurally.

The likely drivers of this tool-limit failure are: (1) **Non-canonical AEEA chemistry** — the dioxaoctanoic acid unit is poorly represented in protein structure databases used to train AlphaFold-family models, degrading confidence wherever the linker contacts the receptor ECD; and (2) **Class B GPCR N-terminal binding geometry** — the insertion of GLP-1 residues 1-6 into a narrow transmembrane cavity (TM6 contacts) is a notoriously difficult modeling problem even for native sequences, with ipTM scores frequently marginal even in published GLP-1R structure predictions. Together, these factors place the fold outside reliable in silico resolution.

There is also a substantive chemical concern that the literature raises but the structural tools cannot adjudicate: the α-amine of His-1 in GLP-1R-bound GLP-1 makes a direct hydrogen bond to **Glu387 in TM6** (cryo-EM evidence). AEEA conjugation converts the primary α-amine (charged at pH 7.4, strong H-bond donor) to a secondary amide (non-protonatable, weaker H-bond donor). This pharmacophoric change is distinct from the Cα-methylation tested in Fold №52 (αMe-His, PROMISING, pLDDT 0.72), which preserved the α-amine charge while rigidifying the backbone. Whether this chemical perturbation ablates, reduces, or is tolerated by GLP-1R is genuinely unknown — no published study has tested it.

WHAT THIS DOESN'T MEAN

DISCARDED does not mean disproved. The structural tools could not confidently model the AEEA-His-1/GLP-1R complex — this is a resolution failure of the prediction pipeline, not evidence that the modified peptide is inactive or structurally incoherent. The delivery hypothesis (aminopeptidase shielding at the N-cap) is well-founded: the PK literature confirms that oral semaglutide bioavailability is low and condition-sensitive (PMID 38952487), and AEEA is chemically accessible via standard Fmoc SPPS. The α-amine-to-amide chemical concern is real and important, but it is a hypothesis to be tested — not a proven disqualifier. The Lys-26 fatty diacid modification in native semaglutide (also a significant structural addition)

reduced GLP-1R affinity only ~3-fold relative to liraglutide (PMID 26308095), suggesting the peptide scaffold tolerates conjugation chemistry even at pharmacophore-adjacent positions. The N-terminal case is higher-risk, but not foreclosed.

WHAT WOULD ANSWER THE QUESTION

- **cAMP accumulation assay (GLP-1R-transfected HEK293T cells):** Directly measures receptor activation potency (EC_{50}) of the AEEA-His-1 conjugate vs. native semaglutide. This is the most decisive single experiment — if EC_{50} is preserved within 10-fold, the modification is worth pursuing for delivery optimization.
- **Competitive radioligand or TR-FRET binding assay:** Quantifies GLP-1R binding affinity (Ki) of the AEEA conjugate, separating binding from activation and identifying whether α-amine conversion affects ECD engagement or TM6 insertion specifically.
- **In vitro proteolytic stability assay (rat or human intestinal homogenate, or purified aminopeptidase N):** Tests the core delivery hypothesis — whether AEEA shielding measurably extends N-terminal half-life under luminal conditions relevant to oral/intranasal dosing.
- **Molecular dynamics (MD) or FEP simulation with explicit AEEA parameterization:** Standard CHARMM36 or GAFF2 force-field parameterization of the AEEA unit would enable multi-microsecond MD of the AEEA-His-1/GLP-1R complex in a lipid bilayer environment, providing a physics-based (not ML-based) assessment of whether the imidazole side chain can still reach Glu387 in TM6 with the α-amine blocked.

RAW METRICS

Metric	Value
pLDDT	0.684
pTM	0.657
ipTM	0.463
Chai-1 agreement	None obtained
Boltz-2 affinity module	No values returned
Predicted binding change	None
Aggregation propensity (heuristic)	0.194

Metric	Value
Stability score (heuristic)	0.203
BBB penetration (heuristic)	0.111
Half-life estimate (heuristic)	Long (>6 h, modification-dependent)

SEQUENCES

NATIVE

```
HAEGTFTSDVSSYLEGQAAKEFIAWLVRGRG
```

MODIFIED

```
AEEA-H-Aib-EGTFTSDVSSYLEGQAAK(γGlu-γGlu-C18-diacid)EFIAWLVRGRG
```

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
- DISCARDED verdict reflects tool-limit failure (ipTM 0.46, non-canonical AEEA chemistry, Class B GPCR TM binding resolution limit) — not biological invalidation of the hypothesis
- AEEA (dioxaoctanoic acid) linker is poorly represented in AlphaFold training data; structural predictions involving this unit carry elevated uncertainty
- α-amine-to-amide conversion at His-1 removes a charged pharmacophoric element (Glu387/TM6 H-bond donor); potency impact is chemically plausible but unquantified — heuristic property estimates do not capture this effect
- heuristic aggregation, stability, and BBB scores are sequence-based estimates only and do not account for the AEEA conjugate or the C18 fatty diacid lipidation chemistry
- no Chai-1 ensemble or Boltz-2 affinity module output was obtained; structural confidence cannot be independently verified

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 —** (2022) — — Wegovy (semaglutide): a new weight loss drug for chronic weight management
6. **PMID —** (2023) — — Semaglutide for the treatment of obesity
7. **PMID —** (2025) — — Efficacy and Safety of Glucagon-Like Peptide-1 Receptor Agonists for Weight Loss Among Adults Without Diabetes: A Systematic Review of Randomized Controlled Trials
8. **PMID —** (2026) — — A Retrospective Study Evaluating the Efficacy and Safety of Oral Semaglutide Compared to Injectables in Controlling Diabetes and Weight Reduction

SOLANA SIGNATURE 3b7ribWeMGM4j6Grrga9mi7F3tk41fCLhy58XriYU3pKEvhyePEuzJyPSN5uARi5an6DPB25tdhNsudPu2G8RN6V

DATA SHA-256 3cc1761e179d29fc09428561baa92d0bd1c2697970360b1ab040a1881b67f1f0

VERIFY https://solscan.io/tx/3b7ribWeMGM4j6Grrga9mi7F3tk41fCLhy58XriYU3pKEvhyePEuzJyPSN5uARi5an6DPB25tdhNsudPu2G8RN6V