

SEMAGLUTIDE — HIS-1 → A-METHYL-L-HISTIDINE (AME-HIS); CA-METHYLATION OF THE N-TERMINAL HISTIDINE TO RIGIDIFY THE BIOACTIVE N-CAP CONFORMATION WHILE PRESERVING THE IMIDAZOLE PHARMACOPHORE

generated 2026-05-04T05:29:41.384297+00:00

PROMISING

METABOLIC

HIS-1 → A-METHYL-L-HISTIDINE (AME-HIS); CA-METHYLATION OF THE N-TERMINAL HISTIDINE TO RIGIDIFY THE BIOACTIVE N-CAP CONFORMATION WHILE PRESERVING THE IMIDAZOLE PHARMACOPHORE

GLUCAGON-LIKE PEPTIDE 1 RECEPTOR

| AVERAGE CONFIDENCE | PTM / IPTM | VERDICT |
|----------------------------------|---------------|---------------------|
| 72.1% | 0.822 / 0.843 | PROMISING |
| TARGET | UNIPROT | BINDING PROBABILITY |
| Glucagon-like peptide 1 receptor | P43220 | — |

TLDR

Fold №52 introduces α -methyl-L-histidine at position 1 of semaglutide, aiming to extend the existing Aib-2 helix-nucleating element into a consecutive $C\alpha, \alpha$ -disubstituted N-terminal dipeptide. The structural prediction returned a moderate pLDDT of 0.72 with strong pTM/ipTM scores, but the absence of Chai-1 agreement data and Boltz-2 affinity values left no binding signal to evaluate, warranting a DISCARDED verdict. The heuristic peptide profile shows low aggregation propensity and moderate stability but cannot substitute for direct binding evidence. The fold is biologically uninformative for this modification at this position, though the negative result meaningfully narrows the experimental space for GLP-1R N-cap engineering.

EXECUTIVE SUMMARY

Fold №52 tests α Me-His at the semaglutide N-cap: pLDDT 0.72, strong pTM/ipTM, but no binding signal from any affinity module. A third consecutive semaglutide DISCARDED — the pipeline lacks coverage for non-canonical residues on this scaffold; the conformational hypothesis remains biologically open.

DETAILED ANALYSIS

Semaglutide is a second-generation GLP-1 receptor agonist whose molecular architecture was deliberately engineered for receptor potency and metabolic stability. The foundational design element most relevant here is Aib-2 (α -aminoisobutyric acid at position 2), a $C\alpha,\alpha$ -disubstituted residue introduced in the Lau et al. 2015 scaffold to nucleate helical structure at the N-terminus and block DPP-IV cleavage at the His1-Xaa2 bond. Semaglutide achieves sub-nanomolar GLP-1R affinity (IC₅₀ ~0.38 nM) and a clinically validated once-weekly dosing profile, establishing a high baseline against which any modification must be measured.

The modification hypothesis in Fold №52 is conceptually elegant: if Aib-2 contributes N-terminal helical pre-organization, then extending the $C\alpha,\alpha$ -disubstituted segment one residue further — by replacing His-1 with α -methyl-L-histidine (α Me-His) — should cooperatively reinforce the bioactive N-cap conformation while preserving the imidazole side chain essential for TMD engagement. The rationale draws on well-established $C\alpha$ -methylation chemistry, which restricts backbone ϕ/ψ angles toward helical space, and on literature precedent in other receptor-targeted peptides (GHRH, CGRP analogs). Crucially, the imidazole pharmacophore is retained, distinguishing this from substitutions that ablate the His-1 contact entirely.

The structural prediction produced a pLDDT of 0.72 — sitting at the lower boundary of the 'moderate confidence' band — alongside pTM of 0.82 and ipTM of 0.84, which are ostensibly encouraging interface scores. However, these figures exist without the cross-model validation that Chai-1 agreement would provide, and the Boltz-2 affinity module returned no values. In practice this means the prediction engine produced a plausible fold but no actionable binding signal. For a modification whose entire rationale rests on improved TMD engagement, the absence of any quantified predicted binding change renders the fold informationally silent on the key question.

The heuristic peptide profile — aggregation propensity 0.129 (low, favorable), stability score 0.439 (moderate), BBB penetration 0.086 (very low, expected for a 33-mer with lipidation), half-life moderate-to-long — describes a peptide that is not obviously degraded or aggregation-prone, but these are sequence-derived estimates rather than structural observations. They cannot compensate for the missing binding data. The profile is consistent with semaglutide's known

physicochemical class but adds no discriminating information about whether α Me-His-1 helps or hurts.

Placing this fold within the lab's running narrative: Fold №52 is the third semaglutide distillation, following Fold №15 (Glu-16 \rightarrow homoglutamate, DISCARDED, pLDDT 0.71) and Fold №36 (γ Glu- γ Glu \rightarrow β -Ala- β -Ala spacer at Lys-20, DISCARDED, pLDDT 0.70). All three have landed in the DISCARDED bin at similar confidence levels — a pattern that may reflect an inherent limitation of single-run AlphaFold-class prediction for heavily modified 33-mer peptides with non-canonical residues, rather than a universal verdict against semaglutide modifications. The structural predictor handles canonical sequences well but non-canonical backbone substitutions (Aib, α Me-His) are typically represented as their nearest canonical surrogate, likely degrading position-1 confidence specifically. The broader METABOLIC class context (Folds №23, №31 on tirzepatide; Fold №34, №45 on retatrutide) shows that Aib substitutions and C α -methylation elsewhere in GLP family peptides have yielded PROMISING verdicts — this reinforces the interpretation that the DISCARDED outcome here is a tool-coverage issue as much as a biological signal.

The sterics question flagged in the literature review — whether the GLP-1R transmembrane pocket can accommodate a C α , α -disubstituted imidazole at position 1 — remains genuinely open. Cryo-EM structures (e.g., PDB 7KI0) show relatively tight packing of His-1 in the TMD bundle, and the added methyl group at the α -carbon does increase effective steric volume. This is a legitimate biological concern that in silico tools at the current state-of-the-art cannot resolve for non-canonical residues. The fold's uninformative outcome should therefore be read as 'this question requires better tools or wet lab,' not as 'this modification is inactive.'

In summary, Fold №52 is discarded on technical grounds: the predictor produced a moderate-confidence structure but yielded no binding metrics, making it impossible to evaluate the core hypothesis. The modification rationale remains scientifically coherent and the conformational chemistry is sound. The lab-wide pattern of semaglutide DISCARDS across three structurally distinct hypotheses suggests that the most productive next step is either ensemble prediction with explicit non-canonical residue parameterization, molecular dynamics simulation against the GLP-1R TMD, or direct synthesis of α Me-His-1 semaglutide for cAMP/ β -arrestin functional assays.

RESEARCH BRIEF

FOLD №52 — SEMAGLUTIDE AME-HIS-1 N-CAP

Verdict: DISCARDED | pLDDT 0.72 | pTM 0.82 | ipTM 0.84 | No binding signal

MECHANISM OF ACTION (BACKGROUND)

Semaglutide is a GLP-1 receptor agonist that activates GLP-1R on pancreatic β -cells, hypothalamic nuclei, and cardiovascular tissue to suppress appetite, reduce fasting glucose, and improve cardiometabolic outcomes. Receptor activation proceeds through a two-domain mechanism: the peptide C-terminus docks to the extracellular domain (ECD) of GLP-1R, anchoring the peptide, while the N-terminal segment — most critically His-1 — inserts into the transmembrane (7TM) bundle and triggers $G\alpha_s$ coupling and cAMP elevation. The imidazole side chain of His-1 makes direct contacts within the TMD that are essential for agonist efficacy; truncation or ablation of His-1 converts full agonists to partial agonists or antagonists.

Semaglutide's medicinal chemistry diverges from native GLP-1(7-36) at three points: Aib at position 2 (helix nucleation + DPP-IV blockade), Arg at position 34 (reduced renal clearance), and a C18 fatty diacid via γ Glu- γ Glu spacer at Lys-20/26 (albumin binding, \sim 1-week half-life in humans). The Aib-2 substitution is directly relevant here: it establishes that a $C\alpha,\alpha$ -disubstituted residue at the N-terminal dipeptide is not only tolerated but deliberately engineered, providing the conceptual foundation for the Fold N \circ 52 hypothesis.

MODIFICATION HYPOTHESIS (WHAT WE TESTED)

Fold N \circ 52 replaces His-1 with **α -methyl-L-histidine (α Me-His)** — a $C\alpha,\alpha$ -disubstituted histidine bearing a methyl group at the α -carbon. The hypothesis: combining α Me-His-1 with the existing Aib-2 creates a **consecutive $C\alpha,\alpha$ -disubstituted dipeptide** that cooperatively pre-organizes the N-terminal cap into the helical conformation observed in the GLP-1R-bound state, reducing the entropic cost of binding and potentially improving intrinsic receptor activation efficiency. The imidazole side chain is chemically intact, so all direct His-1-TMD contacts are nominally preserved.

This modification targets **conformation rather than stability** — a deliberate departure from Folds N \circ 15 and N \circ 36, which tested central-helix salt-bridge geometry (Glu-16 \rightarrow homoglutamate) and lipidation spacer chemistry (γ Glu- γ Glu \rightarrow β -Ala- β -Ala), respectively. All three semaglutide distillations have now returned DISCARDED verdicts, but for distinct mechanistic reasons.

WHY THE PREDICTION WAS UNINFORMATIVE (TECHNICAL ANALYSIS OF THE METRICS)

pLDDT = 0.72 places the predicted structure at the lower edge of moderate confidence. For a 33-residue peptide, this is not catastrophic, but it is noticeably below the pLDDT > 0.75 threshold the Researcher identified as the target for a

confident verdict. The moderate score almost certainly reflects the predictor's handling of non-canonical residues: AlphaFold-class models are trained predominantly on canonical amino acids, and α Me-His is typically represented as its nearest canonical surrogate (histidine) with no information about the α -methyl constraint. The structural geometry of the N-terminal cap is therefore likely mis-modeled at precisely the residue of interest.

pTM = 0.82 / ipTM = 0.84 are superficially encouraging interface scores. However, in the absence of Chai-1 agreement data — which would cross-validate the predicted binding mode against an independent model — these values cannot be interpreted as evidence of productive receptor engagement. High ipTM in a single-model, single-run prediction without ensemble agreement is not a reliable binding signal.

Boltz-2 affinity module: no values. This is the critical gap. The entire biological hypothesis rests on whether α Me-His-1 improves receptor binding affinity. Without a predicted binding change ($\Delta\Delta G$ or binding probability), the fold cannot speak to its own core question. The two prior semaglutide folds faced the same outcome:

| Fold | Modification | pLDDT | Binding signal |
|------------------------|--|-------------|----------------|
| N _e 15 | Glu-16 → homoglutamate | 0.71 | None |
| N _e 36 | γ Glu- γ Glu → β -Ala- β -Ala spacer | 0.70 | None |
| N_e52 | His-1 → αMe-His | 0.72 | None |

The pattern across all three semaglutide distillations is consistent: moderate pLDDT, no binding metrics, DISCARDED. This strongly suggests a **systematic tool-coverage limitation** for heavily modified, lipidated 33-mer peptides with non-canonical residues — not a biological verdict against the modification itself.

The heuristic peptide profile (aggregation propensity 0.129, stability 0.439, half-life moderate-to-long) is sequence-derived and not structurally informative. It confirms semaglutide-class physicochemistry but cannot substitute for binding data.

WHAT THIS TELLS US (NEGATIVE RESULTS ARE DATA — WHAT DOES IT RULE OUT?)

The DISCARDED verdict **does not rule out** that α Me-His-1 improves GLP-1R engagement. It rules out that current single-run in silico predictors can answer this question for this modification on this scaffold. That is a meaningful result for lab-wide methodology.

The biological uncertainty it surfaces is real: the GLP-1R TMD binding pocket is sterically constrained around His-1, and the added α -methyl group increases the

effective steric volume at the α -carbon. Two outcomes are plausible and cannot be distinguished computationally without better tools:

1. **Favorable:** Conformational pre-organization dominates; the entropic gain from locking the N-cap outweighs any small steric penalty, net improvement in K_d and/or EC_{50} .
2. **Unfavorable:** Steric clash in the TMD pocket reduces affinity despite favorable conformational bias; over-rigidification of the N-terminus disrupts the dynamic insertion mechanism.

Cryo-EM structures of GLP-1R in complex with peptide agonists (e.g., PDB 7KI0) show relatively tight packing around His-1 — the steric concern is legitimate and should be a primary consideration in any follow-up experimental design. This fold has not resolved the question; it has confirmed that resolving it requires either MD simulation with explicit non-canonical residue parameterization or direct synthesis.

For the broader semaglutide modification program: three DISCARDED folds across lipidation chemistry (N₃₆), central-helix geometry (N₁₅), and N-cap conformation (N₅₂) collectively suggest that **the AlphaFold-class pipeline is not the right primary tool for semaglutide SAR** given its non-canonical residue content. This contrasts with tirzepatide Folds N₂₃ and N₃₁, where C α -methylation and Aib substitutions yielded PROMISING verdicts — likely because tirzepatide's shorter, more canonical scaffold is better handled by the predictor.

ALTERNATIVE HYPOTHESES TO TEST (AVOID THE FAILURE MODE)

Better computational approaches for this specific question: - **Molecular dynamics simulation** (GROMACS/AMBER with GAFF or CHARMM-CGenFF parameters for α Me-His) against the GLP-1R TMD (starting from PDB 7KI0) to directly measure N-terminal conformational dynamics, binding free energy (FEP or MM-GBSA), and steric clash potential. - **Ensemble prediction** (≥ 5 AlphaFold runs with varied random seeds) to assess conformational variance at the N-terminus and distinguish genuine structural flexibility from prediction noise. - **Rosetta FlexPepDock** or similar peptide-receptor docking with explicit non-canonical residue parameterization for α Me-His, which handles backbone constraints more rigorously than standard AlphaFold input.

Alternative modification strategies to test the N-cap hypothesis: - **His-1 \rightarrow D-His:** epimerization is a simpler, better-parameterized modification for AlphaFold surrogates, and D-His has precedent in GLP-1 analogs as a DPP-IV-resistant, partial-agonist probe that could inform TMD tolerance at position 1. - **Desamino-His-1:** removal of the α -amino group eliminates DPP-IV cleavage while keeping backbone flexibility — the result would isolate the pharmacophore contribution of His-1 without conformational constraint. - **N τ -methyl-His-1:** methylation of the imidazole

nitrogen rather than the α -carbon probes a different aspect of the His-1-TMD interaction (H-bond donor capacity) without introducing steric bulk at the backbone.

- **Single-point cAMP assay on synthetic α Me-His-1 semaglutide**: given that this fold cannot be resolved computationally, direct synthesis and functional testing against HEK293-GLP1R is the definitive experiment. The modification is chemically feasible via solid-phase peptide synthesis with Fmoc- α Me-His(Trt)-OH building blocks.

For the broader semaglutide program, the lab should consider whether future distillations on this scaffold are better served by switching to the MD/FEP computational track rather than the AlphaFold-binding-predictor pipeline, given the consistent tool-coverage limitation observed across Folds №15, №36, and №52.

SEQUENCES

NATIVE

```
HAEGTFTSDVSSYLEGQAAKEFIAWLVRGRG
```

MODIFIED

```
( $\alpha$ Me-H)A(Aib)EGTFTSDVSSYLEGQAAKEFIAWLVRGRG
```

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
- α -methyl-histidine is not natively parameterized in AlphaFold-class models; the predictor likely treated it as canonical histidine, degrading structural accuracy at the residue of interest
- Chai-1 agreement data unavailable — pTM/ipTM scores cannot be cross-validated and should not be interpreted as reliable binding signals
- Boltz-2 affinity module returned no values; the core binding hypothesis (improved TMD engagement) cannot be evaluated from this prediction
- heuristic peptide profile (aggregation, stability, half-life) is sequence-derived, not structurally derived, and does not account for the lipidation chain or non-canonical residue content
- steric tolerance of the GLP-1R TMD pocket for α Me-His at position 1 is a genuine unresolved biological concern not addressed by this fold

- 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** — (2023) — — Semaglutide for the treatment of obesity
3. **PMID** — (2020) — — Semaglutide lowers body weight in rodents via distributed neural pathways
4. **PMID** — (2024) — — Clinical Pharmacokinetics of Semaglutide: A Systematic Review
5. **PMID** — (2021) — — Safety of Semaglutide
6. **PMID** — (2024) — — Glucagon-like peptide-1 receptor agonist semaglutide reduces atrial fibrillation incidence: A systematic review and meta-analysis
7. **PMID** — (2022) — — Wegovy (semaglutide): a new weight loss drug for chronic weight management
8. **PMID** — (2021) — — Semaglutide 2·4 mg once a week in adults with overweight or obesity, and type 2 diabetes (STEP 2)

SOLANA SIGNATURE QG18KkCC8hfkEhyPYpfiYtt7YHvhDdWvjU2yBbnN6czv9EzKnDdtErUh3a
QUeMim36gcLzJawWxwDBFWCEv2Wmc
DATA SHA-256 18c7081ac90d4e4cd619730a1dd01e9dade0a3edb02aec4e5a8a44aba19b75
63
VERIFY [https://solscan.io/tx/
QG18KkCC8hfkEhyPYpfiYtt7YHvhDdWvjU2yBbnN6czv9EzKnDdtErUh3aQUeMim36gcLzJawW
xwDBFWCEv2Wmc](https://solscan.io/tx/QG18KkCC8hfkEhyPYpfiYtt7YHvhDdWvjU2yBbnN6czv9EzKnDdtErUh3aQUeMim36gcLzJawWxwDBFWCEv2Wmc)