Metastate Peptide-Lab

AI-assisted short-peptide discovery pipeline for GLP-1 secretion

Whitepaper | Date: February 7, 2026

Metastate Brain in action:

An end-to-end system that turns proteins into testable short-peptide candidates, using digestion simulation, explainable multi-factor scoring, 3D modeling, docking, and an experimental feedback loop.

Public summary - no proprietary sequences or confidential assay data included.

Executive Summary

Metastate Peptide-Lab is a modular, end-to-end platform for discovering short, food-like peptides that may trigger GLP-1 secretion from gut L-cells.
It combines digestion simulation, explainable multi-factor scoring (chemistry, safety, stability), peptide 3D modeling (AlphaFold/ColabFold), and peptide docking (Vina/ADCP) to prioritize candidates for wet-lab testing.

Key point: Computational scores are not proof of biological activity. Peptide-Lab is designed to accelerate validation by producing a small, high-quality shortlist and a clear rationale for each candidate.

What we built	How it works	Why it matters
A modular platform that turns proteins into ranked short-peptide candidates.	Digestion simulation -> scoring -> AlphaFold/ColabFold -> docking -> shortlist.	Fewer low-probability experiments; clearer rationale for validation.

Problem & Opportunity

Short peptide discovery is fragmented: digestion simulation, scoring, structure prediction, docking, and reporting often live in separate tools.
That fragmentation pushes too many low-probability candidates into the lab, creating slow and expensive experimental loops.

Solution: Peptide-Lab (Modular Platform)

Peptide-Lab unifies the workflow in one place: protein -> virtual digestion -> feature scoring -> 3D modeling -> docking -> shortlist -> assay feedback.
The result is a repeatable, auditable pipeline that reduces experimental load and improves hit-rate by filtering early.

Metastate Brain Capacity

"Metastate Brain" is how we describe the orchestration layer: a system that chains together mechanistic rules (protease cleavage), ML predictors (toxicity, bitterness, stability), and structure-based methods (AlphaFold + docking) into one explainable decision flow.

End-to-End Pipeline (MVP)

The MVP implements a repeatable, stage-gated workflow:

Stage 1 - Virtual Digestion

We simulate protease cleavage to generate plausible short peptide fragments from input proteins. We track enzymes, missed cleavages, and length filters. A practical target window is ~8-12 amino acids.

Digestion settings

- Missed cleavages: allow 'skipped' cuts to reflect real digestion and generate longer fragments.
- Length filters: keep peptides within a chosen window (e.g., 6-20 aa; practical sweet spot ~8-12).
- Combining enzymes: take the union of cut sites across selected proteases; merge unique sequences.
- Sequential digestion (pepsin -> pancreatic) is on the roadmap for improved realism.

Proteases simulated

Enzyme	Cuts After...	Exceptions / Notes	Why We Include
Trypsin	K or R	No cut if next residue is P	Enriches basic residues; improves electrostatic interactions
Chymotrypsin	F, W, or Y	Often blocked by next residue P	Boosts aromaticity for pocket interactions
Pepsin	Hydrophobic sites (F, L, W, Y)	Broad preference at low pH	Generates hydrophobic-leaning fragments
Elastase	A, G, S, or V	Prefers small neutrals	Diversifies candidates across hydro-phobicity
None	-	No cuts applied	For comparison / pre-digested inputs

Stage 2 - Feature Profiling & Scoring

Each peptide is profiled across chemistry, safety, stability, and bioactivity cues. Features are normalized and combined into an interpretable composite score. A Z-score compares candidates to shuffled baselines to catch 'too good to be true' artifacts.

Category	Examples	Why it matters
Basic chemistry	Aromaticity, basicity, charge/pI, hydrophobicity	Signals likely pocket interactions and solubility
Safety & stability	Toxicity, protease vulnerability, DPP-IV risk, half-life	Survival through digestion; avoid harmful motifs
Bioactivity cues	Bitterness, length & local motifs	Hints at receptor triggering and sensible size

Core feature set

Feature	Plain Definition	Why It Matters
Aromaticity	Share of F/Y/W	Hydrophobic/pi-pi pocket contacts
Basicity	Share of K/R/H	Electrostatic attraction to receptors
Charge/pI	Net charge, neutral point	Solubility & interaction balance
Hydrophobicity	Water-repelling tendency	Pocket fit vs aggregation
Length/Motifs	~8-12 aa; turns/helix	Present key residues for binding
Bitterness	Bitter-like pattern/ML	Potential gut taste activation
Stability	Resists degradation	Survives to site of action
Toxicity	Safety likelihood	Screen out risky motifs
Protease vuln.	Predicted cut rate	Digestive survivability
DPP-IV risk	N-terminal trimming	Avoid rapid inactivation

Explainability & safeguards: The composite score is a weighted fusion of normalized features. Missing predictors are skipped (not treated as zero). A Z-score compares each peptide to shuffled versions to sanity-check 'unusually good' sequences.

Benchmarking Against Known GLP-1–Relevant Peptides

To assess whether the GLiCo scoring framework captures biologically meaningful signal rather than random peptide properties, we performed an initial benchmarking analysis using a small reference set of peptides reported in the literature to induce GLP-1 secretion or related endocrine responses, compared against a randomized peptide background.

Dataset Construction

Two groups were evaluated:

- **Reference set**: 5 short peptides with reported biological relevance in GLP-1 secretion or closely related signaling contexts.
- **Randomized set**: 30 peptides generated without biological curation, matched loosely for length distribution but otherwise unconstrained.

This benchmark is intentionally conservative in scale and scope. The goal was not to demonstrate predictive power for clinical efficacy, but to test whether the scoring framework can distinguish known biologically active peptides from random sequences at all.

Scoring and Statistical Analysis

Each peptide was evaluated using the GLiCo composite score, which integrates physicochemical, sequence-derived, and stability-related features into a normalized score range.

A permutation-based significance test (3,000 permutations) was performed to evaluate whether the observed separation between the two groups could plausibly arise by chance.

Results:

- Mean score difference (Δ mean): **0.121**
- Effect size (Cohen's d): **1.20**
- Empirical p-value: **≈ 0.017**

The effect size indicates a large separation relative to within-group variance, despite the small reference set.

Group Statistics

Reference peptides

- Count: 5
- Mean score: **0.53**
- Median: 0.51
- Standard deviation: 0.12

Observed scores ranged from **0.37 to 0.68**, with most reference peptides clustering above 0.50.

Randomized peptides

- Count: 30
- Mean score: **0.41**
- Median: 0.39
- Standard deviation: 0.09

While individual random peptides occasionally scored highly, the overall distribution was shifted downward relative to the reference group.

Interpretation and Limitations

This benchmark demonstrates that the GLiCo scoring framework assigns systematically higher scores to literature-supported peptides than to a randomized background, with statistical significance despite limited sample size.

However, several limitations must be emphasized:

- The reference set is small and heterogeneous, reflecting the scarcity of well-characterized short peptides with validated GLP-1 secretion data.
- The analysis does not claim predictive validity for in vivo efficacy or clinical relevance.
- Docking, receptor proxy analysis, and digestion simulation were not used as independent validation signals in this benchmark and should not be interpreted as confirmatory evidence.

The result should therefore be interpreted as **evidence of non-random signal**, not proof of biological causality.

Based on this observed separation, subsequent structure modeling and docking steps are used strictly as secondary prioritization signals, not as independent validation.

Stages 3-5 - Structure Modeling & Docking

Stage 3 - Receptor selection: Today we use peptide-binding GPCRs as practical proxies while structures for specific L-cell targets mature. Current receptor set: MC1R, MC4R, B2R, AT1R, PAR1.

Stage 4 - Peptide 3D modeling: For each shortlisted sequence we generate a starting conformation with AlphaFold/ColabFold (fallback: a linearized conformation).

Stage 5 - Docking: We dock flexible peptides into predefined extracellular binding boxes using AutoDock Vina (and ADCP when needed). We record the best-scoring pose as a prioritization signal, not as proof of activation.

How to read docking scores: More negative docking energies generally imply stronger predicted binding, but binding is not the same as activation. Small score differences should not be over-interpreted.

Known Limitations

- Rigid receptor models: limited induced-fit and conformational changes.
- Surrogate receptors: not all are true L-cell targets yet.
- In-vivo journey is not modeled (mucus, transport, metabolism).

Planned Improvements

- Sequential digestion presets (e.g., pepsin -> pancreatic) and expanded protease sets.
- Data-driven weight tuning using wet-lab feedback; structure-aware features.
- Add L-cell targets (e.g., TAS2Rs, GPRC6A, CaSR) as structures mature.
- Induced-fit docking / MD for top hits; membrane-aware simulations.
- Simple peptide modifications (e.g., N-Ac, C-NH2) to improve stability.

Validation Plan & Partner Path

Peptide-Lab is validation-ready: every ranked candidate comes with an explainable score breakdown and clear next experiments.

In-vitro validation modules (planned / in progress)

- Peptide synthesis (SPPS) and QC (HPLC / LC-MS).
- Simulated digestion (SGF/SIF) and protease-panel stability.
- Serum stability and basic developability checks (solubility, logD).
- Permeability screening (PAMPA).
- Functional assays: GLP-1 ELISA, cAMP readouts, and taste-receptor (TAS2R) Ca2+ mobilization (via CRO).
- Safety screens: toxicity (MTT/LDH) and safety window estimation.
- Closed-loop learning: assay outcomes feed back into score weights, thresholds, and model explanations.

Platform engineering stack

- Backend API: FastAPI
- Frontend: Angular
- Cloud & orchestration: AWS + Kubernetes (K8S)
- 3D modeling: AlphaFold / ColabFold
- Docking: AutoDock Vina and ADCP
- Refinement: MM-GBSA and short MD (OpenMM) for top candidates
- Reporting: explainable scorecards + Z-score sanity checks

Target KPIs (project-level)

Success metric	Target
Score behavior matches known peptides (benchmark set)	>= 75% accuracy
Peptides passing in-silico pre-screen	>= 100,000
Peptides with detailed docking + MD-lite refinement	>= 1,000
Candidates tested in vitro	>= 20
Candidates showing GLP-1 increase (cell-based)	>= 3

Partnership ask: We are actively looking for validation partners who can run GLP-1 secretion assays (cells or organoids), receptor pharmacology readouts, and/or stability & developability panels. In return, partners get early access to prioritized candidates and the platform outputs that justify them.

Status, IP, and Roadmap

Project direction: build a TRL 3 -> TRL 7 prototype platform within a 16-month plan, with in-vitro feedback loops that calibrate scoring and improve hit-rate over time.

IP status

A US provisional patent application has been filed for 'Short Peptides for Enhancing GLP-1 Secretion' (Application No. 63/909,826; receipt date Nov 1, 2025).

Roadmap

- Lock default digestion presets and publish reference docs.
- Benchmark scoring on public peptide sets; release a demo dataset.
- Dock the top 100 candidates and shortlist ~10 for synthesis.
- Finalize a GLP-1 secretion assay plan (cells/organoids) with partners.

How to collaborate

We collaborate in 2 ways:

1. Validation partner: you run agreed assays (GLP-1 ELISA, cAMP, Ca2+ mobilization, organoids), we provide ranked candidates and full scorecards.
2. Platform partner: you bring targets, datasets, or screening constraints; we integrate them as a new module (receptors, proteases, scoring features, reports).

Contact

Oğuzcan (Can) Ünver
Founder, Metastate
can@metastate.bio

LinkedIn | *Email* | *Book a Call*