

DiscoveryLab — Discovery Strategy

Peptiter · Technical specification of the eight-stage peptide discovery workflow, now extended with pathway-mechanism verification and Lean 4 audit receipts

DiscoveryLab is a condition-first peptide discovery application. This document summarises the workflow stages, the artefact each stage produces, the diagram elements shown in the product UI, and the underlying methods or data sources. The current implementation adds an explicit mechanism-verification layer: AI and graph methods can propose pathway hypotheses, Swift checks finite graph properties, Lean 4 artifacts provide an audit path, and perturbation or wet-lab evidence tests the assumptions.

NEW Pathway intelligence and formal mechanism checks

Candidate peptides can now carry typed pathway mechanisms: biological nodes, causal edges, intervention-blocked nodes, therapeutic reachability goals, adverse-pathway blockade claims, protected-pathway safety claims, conservation laws, and perturbation evidence anchors.

IMPLEMENTED ARTIFACTS

PathwayMechanismHypothesis	Structured candidate mechanism with proposal source, evidence anchors, causal graph, intervention context, reachability, blockade, safety, and conservation claims.
PathwayMechanismVerifier	Local graph-level verifier that checks reachability, blockade, protected-node safety, and reaction conservation before wet-lab handoff.
LeanVerificationArtifact	Generated Lean 4 module containing node types, baseline and intervention reachability, theorem names, and checksum-bound report identity.
peptiter-lean-verify	External CLI that writes Lean source, invokes Lean or dry-run mode, captures diagnostics, validates checksums, and emits JSON verification receipts.
PerturbationEvidenceRecord	Assay, omics, CRISPR, chemical perturbation, or partner wet-lab evidence attached to mechanism assumptions and scored for support, contradiction, or gaps.

Current assays: mechanismVerification · perturbationEvidence · receptor fit · stability · solubility · aggregation · synthesis · off-target · assay readiness

01 Select condition

Define therapeutic intent or phenotype as the entry point. Anchored to standard vocabularies so downstream queries are reproducible.

DIAGRAM ELEMENTS

Condition node (E66 / M06 / G56)	ICD-10 code with surrounding evidence ring; standard vocabulary anchor.
phenotype	Observable clinical traits (HPO terms) used to constrain the search.
intent	Agonism, antagonism, allosteric modulation, or biased signalling.
comorbid	Co-occurring conditions adjusting target prioritisation and selectivity requirements.
contra-ind.	Receptor or pathway interactions to avoid (off-target, safety liabilities).

References: ICD-10 · MeSH · MONDO · HPO

02 Map pathways & receptors

Build a directed graph from the condition through pathway layers to candidate receptors with ligand class metadata and candidate mechanism claims that can later be verified.

DIAGRAM ELEMENTS

Condition vertex	Entry point of the directed pathway graph.
Upstream / direct / downstream pathway	Signalling cascade reachable from the condition (e.g. incretin → GLP-1/cAMP → insulin secretion for obesity).
Receptor candidates (4 nodes)	Druggable receptors retrieved with ligand-class metadata, e.g. GLP1R, GIPR, GCGR, Y2R.
Convergence node	Receptors with overlapping endogenous ligands.

References: Reactome · KEGG · IUPHAR / GtoPdb · UniProt · Lean 4

03 BioScout source systems

Evidence-backed mimicry plans drawn from evolved peptide systems; converted into a curated motif library.

DIAGRAM ELEMENTS

Source organism rows	Program-specific lineages (e.g. Gila monster / amphibian / fish gut for obesity; cone snail / spider / scorpion for ion channels).
Motif library	Curated sequence motifs and pharmacophores indexed from APD3, DRAMP, ConoServer.

References: APD3 · DRAMP · ConoServer

04 Seeded evolution

Start from validated bioactive peptide seeds and evolve local analogs with full ancestry, parent IDs, and operator history.

DIAGRAM ELEMENTS

Seed node	Validated parent peptide with known bioactivity (e.g. exendin-4, magainin-2, ω -MVIIA).
Branch operator	Substitution / cyclisation / N-methylation under family constraints.
Analog leaves	Candidates with operator history and rationale for ranking.

References: Pfam / InterPro · Hopp & Woods 1981

05 Visualize peptide–receptor fit

Structure-aware 3D review of binding orientation; RealityKit-based visualisation on macOS / visionOS.

DIAGRAM ELEMENTS

Receptor scaffold	Schematic of the target (Class A/B GPCR 7TM bundle, cytokine receptor, ion channel).
Peptide backbone	C α trace of the candidate placed against the receptor pocket.
C α atoms	Per-residue selectable nodes for side-chain inspection in the 3D view.
Key contact	Predicted polar / hydrophobic interaction with a pocket residue (distance < 4 Å).

References: RCSB PDB · AlphaFold DB · PEP-FOLD

06 In-silico lab assessment + mechanism verification

Multi-criteria scoring with explicit rejection gates pre-wet-lab. Each gate is conservative until calibrated outcome data exists. Mechanism claims are also checked for pathway reachability, intervention blockade, protected-node safety, conservation, Lean auditability, and perturbation evidence support.

DIAGRAM ELEMENTS

fit	Composite score from docking pose quality and contact-residue agreement.
stab	

	Predicted resistance to proteolysis and conformational entropy penalty.
sol	CamSol / SolubiS-style intrinsic solubility proxy.
aggr	Zyggregator / Tango-style β -aggregation score.
synth	SPPS coupling-difficulty estimate plus length and modification penalties.
tox	ToxinPred-class classifier; failing candidates are gated out.
mech	Graph-level mechanism verification: desired endpoint reachable, adverse endpoint blocked, safety claim explicit.
Lean receipt	Checksum-bound verification artifact generated for CI or reviewer-facing audit.
perturb	Evidence coverage across wet-lab, omics, CRISPR, chemical perturbation, or partner assay records.

References: ATTRACT / AttractKit · Lean 4 · ToxinPred · CamSol · Tango / Zyggregator

07 Prepare wet-lab batch

Hand off through LabSpace to capability-matched partners using machine-readable batch manifests.

DIAGRAM ELEMENTS

Batch manifest	Sequences, modifications, and assay plan in a SiLA 2 / Allotrope ADF-compatible format.
Candidate IDs	Per-program identifier prefixes (e.g. DL-GLP-0421, DL-IL17-0438, DL-CAV-0455).
Vials	Synthesised quantity and assay readout — height encodes activity proxy (EC50 / IC50).

References: SiLA 2 · Allotrope ADF

08 Receive results & refine

Wet-lab and perturbation feedback update the surrogate model and the encoded mechanism assumptions; re-ranking selects the next batch or the next assay under an acquisition function.

DIAGRAM ELEMENTS

Closed loop	Bayesian / active-learning loop between wet-lab and re-rank nodes.
Wet-lab node	

Assay results (binding affinity, cytotoxicity, stability) returned through LabSpace.

Re-rank node

Surrogate model updated with new evidence; acquisition function selects next batch.

References: Shahriari et al. 2016 (Bayesian Opt.) · Settles 2009 (Active Learning)

Program variants

DiscoveryLab ships with three reference programs that swap condition codes, pathway labels, source organisms, seed peptides, gate thresholds, and candidate ID prefixes throughout the workflow.

Program	Condition	Primary target	Seed	Sources
Obesity	ICD-10 E66	GLP1R · Class B GPCR	exendin-4	Gila monster, amphibian, fish gut
Inflammation	ICD-10 M06	IL-17RA · cytokine receptor	magainin-2	amphibian, marine invert., human defensin
Neuropathic pain	ICD-10 G56	Cav2.2 · N-type Ca ²⁺ channel	ω -MVIIA	cone snail, spider venom, scorpion

Language and posture

The system uses constrained search, receptor-conditioned design, evidence-gated generation, candidate-family evolution, in-silico triage, pathway-mechanism verification, Lean audit receipts, perturbation evidence scoring, and a wet-lab feedback loop. It does not promise instant discovery, guaranteed binding, clinical efficacy, or fully automated drug discovery. Formal verification means the encoded claims follow from encoded assumptions; it does not prove that the biology is complete or clinically true. Scoring is intentionally conservative until calibrated outcome data exists, then specific gates are replaced by validated production packages and curated pathway importers.