Reactome LNP Agent v3 — Dual-Model Pipeline
Claude Haiku 3.5 (fast) + Claude Sonnet 4.5 (reasoning) | LangGraph + FAISS + AWS Bedrock

User Input

START

Model Assignment

Fast nodes: Claude 3.5 Haiku
Experts: Claude Sonnet 4.5
Lead Agent: Claude Sonnet 4.5
Embeddings: Titan Embed v2
Vector DB: FAISS (local)
Framework: LangGraph
Cloud: AWS Bedrock (us-west-2)

User Query
"Design a 3-tail ionizable lipid with amine heads..."

input

Infrastructure Nodes (Claude 3.5 Haiku — fast, low cost)

Query Rewrite
(self-contained rewrite using chat history)

rewritten query

Router
(synthesis / lookup / general)

query type

RAG Data Sources

Research Papers
(33 PDFs)

LNP Design Rules
(MCTS, constraints)

Reaction Templates
(13 SMARTS)

Compound Data
(217K blocks, 293 scores)

FAISS Retrieval + LLM Rerank
(454 vectors, top-8 -> rerank -> top-4)

synthesis queries

all queries

Parallel Expert Workers (Claude Sonnet 4.5 — strong reasoning)

Reaction Expert

* SMARTS template matching
* Functional group analysis
* Feasibility assessment

Lipid Design Expert

* Retrosynthesis routes
* SAR analysis
* Design rule compliance

Generative AI Expert

* De novo generation
* RL optimization
* Reward functions

Property Prediction Expert

* ML model recommendations
* Uncertainty quantification

Literature Scout

* PubMed search
* PubChem lookup
* Web search

Supervisor (Claude Sonnet 4.5)

Lead Reasoning Agent

* Critical evaluation of all experts
* Conflict resolution
* Confidence levels (HIGH/MED/LOW)
* Final synthesized answer

final answer

Output

Final Answer
(SSE streamed to client)

END