LIFE SCIENCES DEEP AGENTS

From Fuzzy Questions to Validated Facts

Multi-agent biomedical research grounded in canonical identifiers, powered by MCP

ORCHESTRATION

lifesciences-deepagents

LangGraph supervisor + 7 specialist subagents

DATA LAYER

lifesciences-research

12 MCP servers, 34+ tools, 691 tests

Fuzzy-to-Fact Protocol | ACVR1/FOP Drug Repurposing Example

The Problem: LLMs Hallucinate in Biomedicine

- LLMs invent trial IDs, fabricate drug mechanisms, conflate gene symbols
- In biomedicine, wrong answers have real consequences
- Core issue: no grounding in canonical identifiers (CURIEs)

WHAT THE LLM SAYS

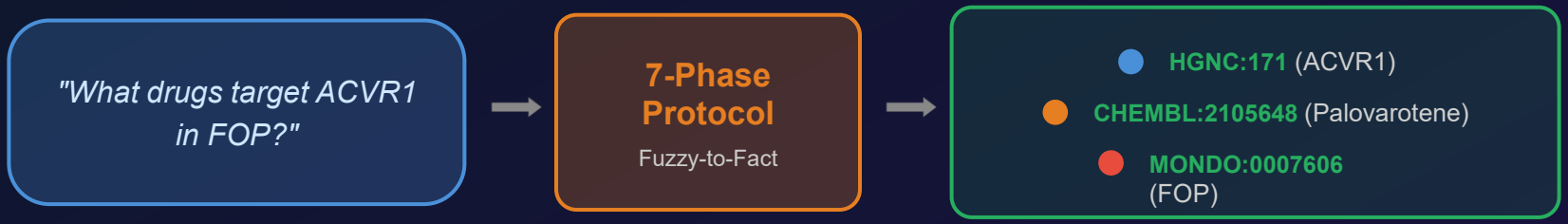
ACVR1 inhibitor Saracatinib (Phase 3)

WHAT THE DATABASE SAYS

Saracatinib: Src/ABL inhibitor, FOP Phase 2 (NCT04307953)

Every unchecked claim is a potential patient safety risk

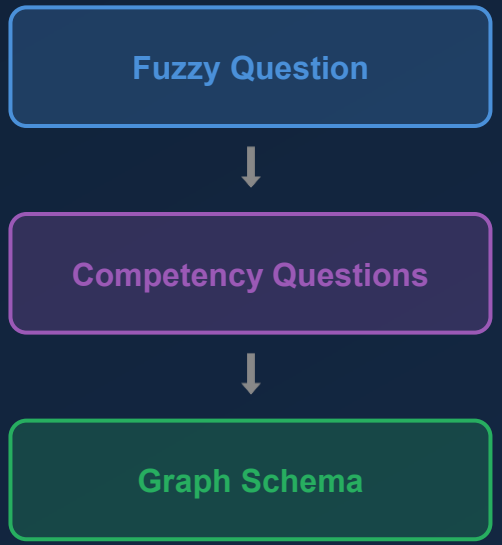
The Vision: Every Claim Traceable



Every node has a CURIE. Every edge has a source. Every claim is verifiable.

What Are Competency Questions?

- A CQ is a natural-language question that defines what the system MUST be able to answer
- From ontology engineering — like unit tests for a knowledge graph
- Analogy: A restaurant DB needs CQs like "Which restaurants serve Italian within 2 miles?" to know what nodes/edges to build
- CQs bridge fuzzy user intent and graph schema



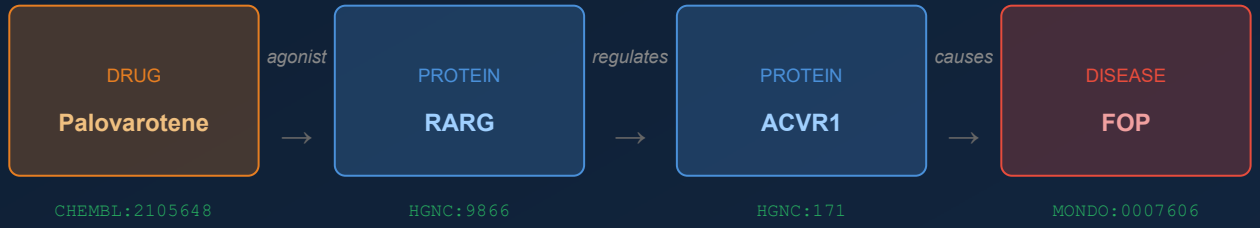
CQs Applied: ACVR1 / FOP

"What drugs target the ACVR1 pathway in FOP?"

ANCHOR	What is the canonical ID for ACVR1?
ENRICH	What protein does it encode?
EXPAND	What proteins interact with it?
TRAVERSE_DRUGS	What drugs target ACVR1 or its pathway?
TRAVERSE_TRIALS	Are there active clinical trials?
VALIDATE	Can we verify each fact?

15 CQs in catalog — each maps to one or more protocol phases

CQs as Acceptance Criteria



PASS — Generated graph matches gold standard

How It Works

Each competency question defines a gold standard path — the expected chain of entities and relationships with canonical CURIEs.

After the system runs, compare the generated graph against the gold standard. This transforms subjective evaluation into measurable pass/fail tests.

15 CQs in our catalog cover drug mechanisms, gene networks, clinical trials, and safety profiling.

Source: reference/competency_questions/competency-questions-catalog.md

Two Repositories, One Protocol

lifesciences-deepagents

Brain: Orchestration + Reasoning

- LangGraph supervisor
- 7 specialist subagents
- React UI
- think_tool reflection

↔

MCP

(JSON-RPC)

lifesciences-research

Backbone: 12 MCP Servers

- 34+ tools
- 691 tests
- FastMCP Cloud gateway
- Fuzzy search → CURIE resolution

Separation of concerns: API volatility isolated at the MCP boundary

The MCP Data Layer: 12 Databases

Each server: fuzzy search → candidate ranking → strict CURIE lookup

- **HGNC** — Fast, reliable
- **UniProt** — Protein data
- **STRING** — 1 req/s limit
- **BioGRID** — 0.5s delay
- **ChEMBL** — Unreliable (500s)
- **PubChem** — Chemical data
- **IUPHAR** — Pharmacology
- **Open Targets** — ChEMBL fallback
- **ClinicalTrials.gov** — v2 API, stable
- **WikiPathways** — Pathway data
- **Ensembl** — Genomic data
- **Entrez / PubMed** — Literature

Gateway composes all 12 into single FastMCP Cloud endpoint

Supervisor + Specialist Pattern

SUPERVISOR
No tools — routes only via task() calls

1	2	3	4a	4b	5	6
ANCHOR	**ENRICH**	**EXPAND**	**TRAVERSE DRUGS**	**TRAVERSE TRIALS**	**VALIDATE**	**PERSIST**
HGNC, ChEMBL search	UniProt, Ensembl	STRING, BioGRID, WikiPathways	ChEMBL, Open Targets	ClinicalTrials.gov	Cross-check all sources	Graphiti / Neo4j

think_tool forces reflection after each API call (Think-Act-Observe)

PHASE 1

ANCHOR — What are we talking about?

- Resolve fuzzy terms to canonical identifiers
- HGNC is always the first call (fastest, most reliable)
- Max 3 search attempts per entity
- If ChEMBL fails → fall back to Open Targets

RESOLUTION EXAMPLE

```
"ACVR1"        → HGNC:171
"Saracatinib"  → CHEMBL:217092
"FOP"          → MONDO:0007606
```

Every downstream phase depends on correct anchoring

PHASE 2

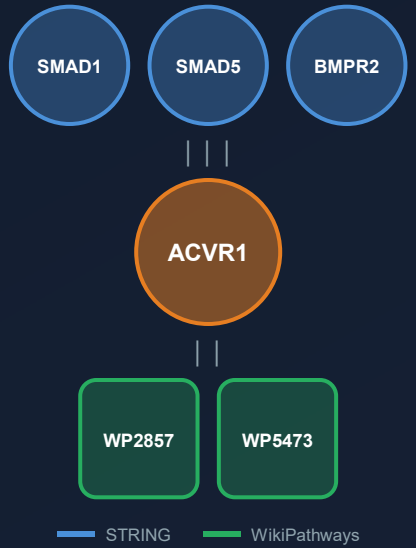
ENRICH — What do we know about these?

- HGNC:171 → UniProt Q04771, Ensembl ENSG00000115170
- Function: BMP type-1 receptor serine/threonine kinase
- FOP: ultra-rare, gain-of-function ACVR1 mutations
- Cross-database ID mapping builds the entity profile

PHASE 3

EXPAND — What connects to ACVR1?

- **STRING:** ACVR1 phosphorylates SMAD regulators
- **WikiPathways:** Mesodermal commitment (WP2857), Cytokine receptor interaction (WP5473)
- **BioGRID:** Physical + genetic interactions
- Network grows from single node to full neighborhood



PHASE 4A

TRAVERSE DRUGS — What drugs target this network?

ChEMBL API call

→ **500 Error**

Open Targets GraphQL

Think-Act-Observe loop evaluates sufficiency

DRUG CANDIDATES

Palovarotene
CHEMBL:2105648
RAR-gamma agonist — **Approved**

Eptotermin Alfa
ACVR1 agonist — **Phase 2**

Dibotermin Alfa
ACVR1 agonist — **Phase 2**

ChEMBL unreliable → Open Targets is the automatic fallback

PHASE 4B

TRAVERSE TRIALS — Are these in clinical trials?

- Drug names + disease context: "Palovarotene AND FOP"
- ClinicalTrials.gov v2 API: stable, structured JSON
- No FOP trials for Eptotermin / Dibotermin
- *Absence of evidence identifies repurposing gaps*

SEARCH RESULTS

Palovarotene + FOP ✓
NCT04307953
(Phase 2)

Palovarotene + FOP ✓
NCT03312634
(Phase 3)

Eptotermin Alfa + FOP ✗
No results

Dibotermin Alfa + FOP ✗
No results

PHASE 5

VALIDATE — Can we verify every claim?

- NCT ID verification against ClinicalTrials.gov
- Mechanism cross-checks across ChEMBL + Open Targets
- PubMed literature confirmation
- **Caught 2 hallucinated identifiers in actual run**
- Each fact: VALIDATED or INVALID with reason

VALIDATION CHECKLIST

HGNC:171 = ACVR1 ✓ **VALIDATED**

NCT04307953 exists ✓ **VALIDATED**

NCT03312634 exists ✓ **VALIDATED**

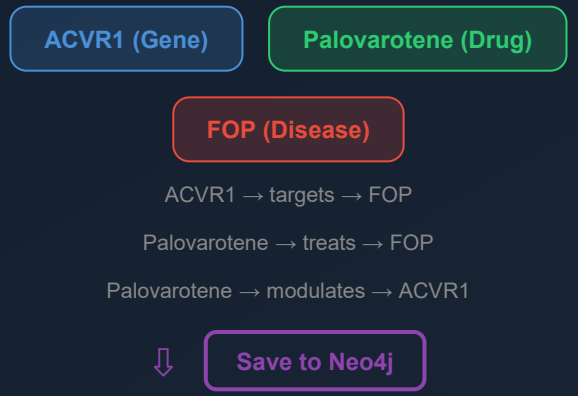
CHEMBL:999999 ✗ **INVALID — not found**

Mechanism match ✓ **VALIDATED**

PHASE 7 (OPTIONAL)

PERSIST — Save to Knowledge Graph

- Structure findings as typed nodes and edges
- Save to Graphiti/Neo4j with full provenance
- System fails soft if Graphiti unavailable
- Validated answer complete after Phase 5
- Enables cross-query knowledge accumulation



OPTIONAL — The validated answer is complete without persistence

The React UI: Transparency at Every Step

Real-Time Status

SubAgentIndicator shows which specialist is active

Phase progress visible throughout run

Tool Transparency

Expandable tool call args and results

See exactly what APIs were called

Human-in-the-Loop

Approve, reject, or edit tool arguments

Interrupt handling via `resumeInterrupt()`

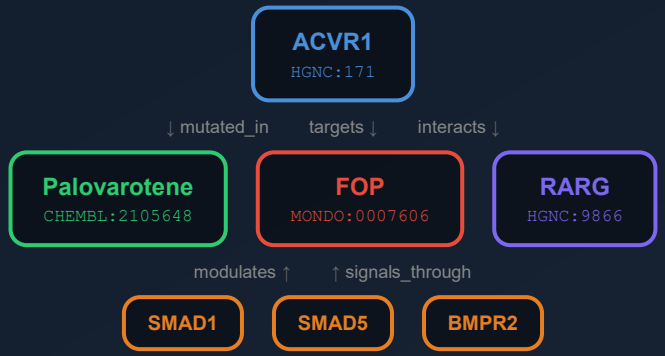
Thread Persistence

Thread ID in URL via nuqs

Resume conversations across sessions

Built with Next.js + @langchain/langgraph-sdk

Final Output: The Validated Knowledge Graph



6 entities resolved **3** drugs found **4** trials checked **5** facts validated

Run time: **~3.5 min** Cost: **~$0.49**

Every claim traces to HGNC, Open Targets, WikiPathways, or PubMed

Ecosystem: Three Implementations

Claude Code Skills

ORCHESTRATION
Skill scripts

DURABILITY
Session-only

UI
Terminal

BEST FOR
Prototyping

Deep Agents
CURRENT

ORCHESTRATION
LangGraph

DURABILITY
Thread persistence

UI
React

BEST FOR
Interactive research

Temporal Workflows

ORCHESTRATION
Temporal

DURABILITY
Built-in replay

UI
Dashboard

BEST FOR
Production pipelines

All share the same **lifesciences-research MCP** data layer

Key Takeaways

1. Grounding in CURIEs is non-negotiable in biomedicine
2. Competency questions = measurable acceptance criteria for knowledge graphs
3. Supervisor + specialist pattern with tool isolation prevents cross-contamination
4. MCP decouples agent logic from API volatility