

# Two oral GLP-1 agonists, two evidence worlds

How Compound Insights assembles a drug-compound dossier so that every published number can be traced, checked — and disbelieved

Compound2 / Compound Insights · compoundinsights.dev · prepared 26 August 2026

## A routine question with a non-routine answer

Orforglipron and elecoglipron are both oral, non-peptide GLP-1 receptor agonists, dosed once daily as a tablet without food or fluid restriction. They are direct competitors from different sponsors, and within a few months of each other one was approved for chronic weight management while the other reported two positive phase 2b trials and moved into phase 3. Asking which is the better starting point — or simply what is known about each — is a question discovery and competitive-intelligence teams ask weekly.

The evidence available to answer it is not routine. The approved compound has a structured database identity and a substantial, fast-growing literature. The other has no structured database anchor at all and three records in PubMed. Both facts are true at once, and the gap between them reflects how far the public data infrastructure has caught up rather than any difference in scientific interest.

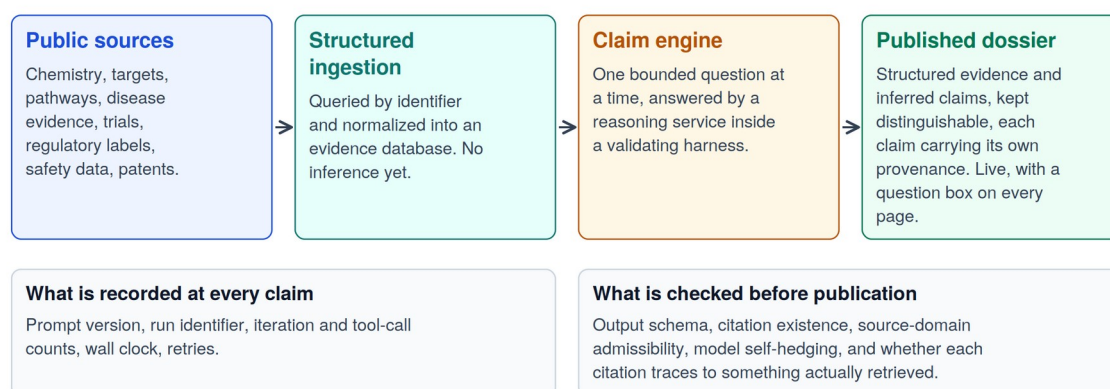
This is where conventional dossier assembly fails invisibly. It handles the first compound well and the second badly, and produces two documents that read with equal confidence. A fluent, uncited summary and a fluent, correctly cited summary are indistinguishable at a glance. Generative tools make the first kind faster to produce; they do not make it safer to act on. The distinguishing property is not prose quality — it is whether each statement can be traced back to the exact question that produced it and the exact retrieval that supports it.

## What Compound2 does

Compound2 is not a document generator. It is an assembly line whose atomic unit is a single claim answering one bounded question, and in which every stage is allowed to reject. The dossier is what remains after that process runs, rendered for a reader.

The governing principle is **attribute, never assert**. Every inferred value is published together with the prompt version, run identifier, tool-call count, per-citation validation state and grounding verdict that produced it. Nothing is endorsed beyond what its evidence supports — and where the evidence is thin, the dossier says so in the same voice it uses to report a well-supported result.

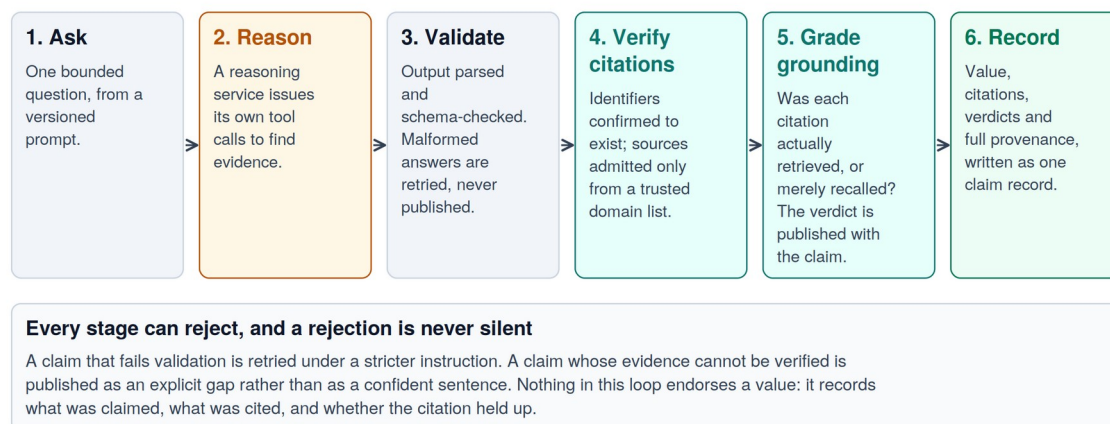
A third compound sits alongside the pair. **Imatinib is the pipe-cleaning compound**: evidence-saturated and twenty-five years old, it is where the machinery was built and validated, precisely because a failure against it is unambiguous. It is not the object of study.



**Figure 1.** From public sources to a published dossier. Structured evidence and inferred claims are produced by different machinery with different failure modes, and are kept distinguishable all the way to the page, so a reader always knows which they are looking at.

## How a single claim is made

Each claim begins as one bounded question drawn from a manifest that also fixes its permitted tools, the kind of citation expected, and the schema its answer must satisfy. A reasoning service answers it by issuing its own searches; the harness around that service is where the engineering lives, and it is what turns a plausible answer into a checkable one.



**Figure 2.** The life of one claim. Each stage can reject, and the rejection behavior at each stage is deliberate: a malformed answer is retried under a stricter instruction, while an unverifiable one is published as an explicit gap.

For claims where a single pass would conceal genuine disagreement, the same question is run twice from opposing analytical stances and a third run arbitrates between them. The whole interrogation is kept inside the claim record, so a reader can see not only the answer but the strongest case that was made against it. Where repeated runs disagree on a number and every distinct value carries its own support, the dossier publishes a provenanced range rather than a false point.

## Where the design earns its keep

| Mechanism                 | What it buys                                                                                                                      |
|---------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| One claim, one question   | A citation traces to the exact question that produced it — something a wholesale-generated document cannot offer                  |
| Schema validation         | Malformed output is rejected and retried rather than reaching a reader                                                            |
| Citation verification     | A cited identifier is confirmed to exist before it is persisted; sources are admitted from a trusted domain list                  |
| Grounding grade           | Records whether a citation traces to something actually retrieved, rather than recalled and stated fluently                       |
| Adversarial pass          | Exposes that two defensible readings of the same evidence exist, which a single pass conceals                                     |
| Reproducibility promotion | Publishes a provenanced range where repeated runs genuinely disagree and every value is supported, and refuses where they are not |
| Full provenance           | Run identifier, prompt version, iterations, tool calls, timing and retries, carried per claim                                     |

**Table 1.** Remove any one of these and the output starts to look like every other confident, unverifiable summary. The cost of the process is the price of the claim that distinguishes it.

## What the three dossiers show

Every claim record on disk was joined to its inference session record and graded offline, with the reasoning service switched off throughout, so the verdicts measure what was actually retrieved at the time each claim was made rather than what can be re-found today.

|                    | imatinib               | orforglipron          | elecoglipron          |
|--------------------|------------------------|-----------------------|-----------------------|
| Role               | pipe-cleaning compound | case-study compound A | case-study compound B |
| Structured anchor  | CHEMBL941              | CHEMBL4446782         | none, hand-seeded     |
| Literature corpus  | hundreds of records    | substantial           | 3 records             |
| Validated claims   | 459                    | 25                    | 18                    |
| Grounded share     | 41%                    | 44%                   | 22%                   |
| Invented citations | 0                      | 0                     | 0                     |

**Table 2.** 502 graded claim-steps in total, across all three compounds. No published citation was invented, and no claim lost its provenance.

The third column is the result that matters. A 22% grounded share is not a weakness in the output; it is a thin corpus reported as a measurement. Three records exist for that compound — a non-clinical and first-in-human characterization, and the two phase 2b trial reports — and those three are precisely the ones the engine cited. The remaining claims report gaps because the gaps are real. A dossier that says ‘three papers exist and here they are’ is worth more to a team making a decision than one that fills the same space with plausible sentences.

The two case-study compounds broke different assumptions, and both broke assumptions imatinib had quietly installed. One showed that evidence can exist while its provenance is still insufficient to promote a value — a judgment about what counts as evidence, which the dossier surfaces as unresolved rather than quietly publishing a number that would look identical to a well-supported one. The other showed what happens when there is no structured anchor at all. Thin-corpus and weak-provenance compounds are not edge cases to be handled once the main path works; they are the main path for anything not yet approved.

## What can be asked of a dossier today, and what comes next

Each published dossier carries a question box backed by a language model that reads the whole dossier rather than retrieved fragments — deliberately, because a fragment containing a number without its citation state and grounding verdict is exactly the artifact this system exists to avoid producing. Today that answers questions about one compound at a time.

The comparison across compounds is the next build. It needs a published schema over the claim record so that claims can be selected across dossiers before any model is invoked — given a protein, what is known about it everywhere; given several candidates, where their evidence actually differs and how strong each difference is. That is what turns a set of dossiers into a body of evidence that can be interrogated as one thing, with each borrowed claim still carrying its citations, its grounding verdict and its reproducibility decision.

## In one line

Compound Insights publishes what is known about a compound, what it rests on, and where the evidence runs out — in a form a scientist can check rather than take on trust.

---

Three dossiers are live at [compoundinsights.dev](https://compoundinsights.dev). Contact: [raminderpal@hitchhikersai.org](mailto:raminderpal@hitchhikersai.org)