SATYA MCL1 Open Results Collection

Overview of the collection's records, built 2026-10-06 from the collection tree.

Invariant Research, Jason Leonard Volk, founder and chief scientist. jason@invariant.pro, https://invariant.pro

Zenodo records

Pre-registration, version 1 (published 2026-10-01): https://zenodo.org/records/23076000 (DOI 10.5281/zenodo.23076000)

Version 2: https://zenodo.org/records/23084017 (DOI 10.5281/zenodo.23084017)

All versions (concept DOI): 10.5281/zenodo.23075999

Credits

OpenFold3 and the official MCL1 example outputs: the OpenFold consortium, Hugging Face OpenFold/OpenFold3, Apache-2.0.

OpenFold3 MCL1 ensemble (20 predictions): Zonca, Andrea, San Diego Supercomputer Center, DOI 10.5281/zenodo.21926059, CC-BY-4.0.

Boltz-2 (MIT) generated the Tier 1 models. Experimental structures: the Protein Data Bank (RCSB). UniProt Q07820 MCL1_HUMAN.

This document is a reading copy of the collection's records. The provenance table lists every source it reads, with its SHA-256 and whether that exact file is deposited on Zenodo. Deposited files govern; files marked NOT YET DEPOSITED will be in the next version.

SATYA MCL1 Open Results Collection: overview

What this collection is

SATYA is an independent checker for predicted protein-ligand structures. It does not generate structures and does not change the file it is given. It reads the coordinates of a structure, checks them against a structural contract without needing an experimental reference, gives a whole-structure verdict, points to the exact residues and ligand atoms behind each finding with the measured distances, and packages the result with SHA-256 digests so that it can be checked later.

This collection runs SATYA, through the same accepted route a customer would use, on structures of MCL1, a protein that cancer drugs target. Its rules (the protocol below) and its pinned sources were deposited on Zenodo on 2026-10-01, before any Boltz-2 model was generated. Refusals and machinery failures are kept as results. Nothing is excluded after the fact.

What it contains

The experimental control set: every MCL1 entry in the Protein Data Bank that has a small-molecule component, 125 entries, each run unchanged. Lab-determined structures are the control group: they show what SATYA's reference-free findings look like on structures that were not generated by a model.

Two published sets of OpenFold3 predictions, credited to their authors: the OpenFold consortium's example outputs for the official MCL1 example (Apache-2.0) and Andrea Zonca's ensemble of the same example (San Diego Supercomputer Center, DOI 10.5281/zenodo.21926059, CC-BY-4.0). The Zonca files are PDB format; the collection converts them to mmCIF under a pinned rule and measured, with a control experiment, what that conversion changes.

Boltz-2 models generated by Invariant Research under the frozen protocol: each ligand that appears in two or more crystal structures of MCL1, paired with its best-resolution crystal structure, three seeds, five samples per seed.

What it does not claim

No statement in this collection concerns the correctness of any structure, generator or verifier. No frequency, rate, cross-set or cross-generator statement is made. A SATYA finding is a violation of a structural contract on the coordinates as given; it is never called an error, and SATYA's own verdict is stated with every result. Per-model numbers are in the collection index and are reported there as data, not as characterizations of any generator or of the PDB.

Provenance of these pages

Each source this document reads, measured when it was built, compared with the files of every public Zenodo record of the collection.

Source	SHA-256 (first 16)	Status
Collection index (MCL1_COLLECTION_INDEX.json)	4923a6b61c6d3ab2	NOT YET DEPOSITED
Protocol MCL1_COLLECTION_PROTOCOL_V1.md	76fa4eab99f0049c	DEPOSITED in Zenodo record 23076000, 23084017
Protocol MCL1_COLLECTION_PROTOCOL_V1_1.md	527782c60f05692f	DEPOSITED in Zenodo record 23076000, 23084017
Protocol MCL1_COLLECTION_PROTOCOL_V1_2.md	7137f926465e827e	DEPOSITED in Zenodo record 23076000, 23084017

Source	SHA-256 (first 16)	Status
Protocol MCL1_COLLECTION_PROTOCOL_V1_3.md	21190ee4d00e1ce0	DEPOSITED in Zenodo record 23076000, 23084017
Protocol MCL1_COLLECTION_PROTOCOL_V1_4.md	f2c98026aa96003d	DEPOSITED in Zenodo record 23084017
Protocol MCL1_COLLECTION_PROTOCOL_V1_5.md	85bc2b00e6e25481	DEPOSITED in Zenodo record 23084017
Protocol MCL1_COLLECTION_PROTOCOL_V1_6.md	6f768302b3b2e264	NOT YET DEPOSITED
Conversion control (CONVERSION_CONTROL_V2_COMPARE.json)	9d848472f054e722	DEPOSITED in Zenodo record 23084017
Latest deposit README_DEPOSIT.md (ZENODO_DEPOSIT_5)	10c520f31c5fe931	DEPOSITED in Zenodo record 23084017
Latest deposit DEPOSIT_MANIFEST.json (ZENODO_DEPOSIT_5)	966b6f5c3569fdc2	DEPOSITED in Zenodo record 23084017

State of the collection

Computed from MCL1_COLLECTION_INDEX.json (sha256 4923a6b61c6d3ab2). Counts of models and of their terminal states; per-model verifier numbers are in the index.

Set	Models	Terminal states
Boltz-2 Tier 1: 5 ligand pairs, 3 seeds, 5 samples each, generated under Protocol V1.6 set name: boltz2_tier1	90	COMPLETE 75, UNAVAILABLE 15
Experimental control set: every MCL1 entry in the PDB with a small-molecule component, run unchanged set name: experimental	125	COMPLETE 124, MACHINERY_FAILURE 1
T1: the OpenFold consortium's published OpenFold3 example outputs (Apache-2.0), 5 models set name: openfold3_hf	5	COMPLETE 5
Conversion control, first attempt (Protocol V1.4): 5 round trips of T1 set name: openfold3_hf_roundtrip	5	INPUT_REFUSED 5
Conversion control, corrected (Protocol V1.5): 5 round trips of T1 set name: openfold3_hf_roundtrip_v2	5	COMPLETE 5
T2, first conversion form: Zonca 2026 ensemble (CC-BY-4.0), 20 models, refused by the parser as converted set name: openfold3_zonca	20	INPUT_REFUSED 20
T2, second conversion form (Protocol V1.3): Zonca 2026 ensemble, 20 models set name: openfold3_zonca_conv2	20	COMPLETE 20

Terminal state meanings: COMPLETE, the route delivered an accepted evidence package; INPUT_REFUSED, the route's parser refused the input by policy and the refusal is the result; MACHINERY_FAILURE, a browser-side delivery probe failed after the science completed; UNAVAILABLE, the model was never produced (recorded, never generated).

Conversion neutrality control (Protocols V1.4 and V1.5)

MCL1 CONVERSION CONTROL V2: 0 identical, 5 different (0 of them names only), 0 not comparable

Model	Outcome	First difference
of3hf_rt2_s42_sample1	DIFFERENT	$.verifier.num_violated: 99 vs 100
of3hf_rt2_s42_sample2	DIFFERENT	$.verifier.findings: 19 vs 20
of3hf_rt2_s42_sample3	DIFFERENT	$.verifier.findings: 20 vs 21
of3hf_rt2_s42_sample4	DIFFERENT	$.verifier.num_violated: 100 vs 99
of3hf_rt2_s42_sample5	DIFFERENT	$.verifier.findings: 42 vs 43

Reading, as frozen in Protocol V1.4 section 22: every result on the converted T2 files carries the measured difference as a caveat. The PDB format stores coordinates to three decimals; OpenFold3's mmCIF carries up to eight significant digits.

The protocol, every version

MCL1_COLLECTION_PROTOCOL_V1.md (sha256 76fa4eab99f0049c)

MCL1 OPEN RESULTS COLLECTION: PROTOCOL V1

Version: V1, frozen 2026-09-30 before the first run that counts. Governs: roadmap v3.0.0 Phase M, steps M1 and the experimental control set. The generation grid (generators, versions, ligand panel selection for generation, seeds, samples) is not in V1; it is added as V1.1 before any generated model runs. Amendment rule: this protocol changes only by a new version with a changelog line, written before the runs it affects. Runs already made under an earlier version stand as made.

1. Purpose

A public, inspectable record of what SATYA finds on every MCL1 protein-ligand model that can be generated or obtained, with SATYA's own verdict per model, and with experimental MCL1 structures run through the same route as controls. Every model that enters the collection is published with its outcome. There is no post-hoc exclusion: a refusal is a result with its code, a machinery failure is a result with its class, a clean model is a result.

2. Target

MCL1_HUMAN, UniProt accession Q07820. The canonical sequence is fetched by satya_mcl1_prep_v1.py from UniProt and pinned by SHA-256; the run refuses if the entry name is not MCL1_HUMAN. Residue numbering in every published result is the numbering of the file that was run (PDB author numbering for experimental entries, model-local numbering for generated models); no renumbering is applied.

3. The experimental control set

Selection rule, executed by satya_mcl1_prep_v1.py against the RCSB PDB search API and recorded verbatim in MCL1_SOURCES.json: every PDB entry with a polymer entity whose reference sequence identifier is UniProt Q07820 and whose non-polymer entity count is at least one, of experimental content type, released at or before the day the prep script runs. Entries are ordered by identifier. Nothing is added by hand and nothing is removed by hand.

Files: the mmCIF of each entry as served by files.rcsb.org, pinned by SHA-256 and size, stored unchanged under sources\pdb\. No cleaning: waters, alternate conformations, ions and every non-polymer component stay as delivered. The route's own reader decides what it accepts; an INPUT_REFUSED outcome (for example UNSUPPORTED_INSERTION_CODES) is a recorded result for that entry, not a reason to edit the file.

Metadata recorded per entry from the RCSB data API: experimental method, resolution, initial release date, title, and every non-polymer component (component id, name, heavy-atom count, SMILES).

Route: satya_pub1_bundle_export_v3.py, invoked exactly as the front door service invokes it (--root, --input, --model-source, --progress-file), with model source "PDB experimental". One delivery run per entry under runs\PUB1_DELIVERY_<m>\ in the public workbench package; the batch driver records exit code, terminal state, code and detail, the delivery outcome and the verifier's numbers into MCL1_COLLECTION_INDEX.json after every model.

Pilot: the first run is limited to a fixed number of entries to measure the route's time per model; the pilot's results count and are kept. The full set follows.

4. Ligand panel derivation (for V1.1)

Derived from the experimental set, not chosen by hand. From the distinct non-polymer components across all entries, remove those on the exclusion list below (solvents, buffers, cryoprotectants, ions), then keep components with at least 15 heavy atoms. The remaining components, with their SMILES as recorded by the

RCSB chemical component dictionary and the entries they occur in, form the candidate panel. V1.1 will state which panel members are generated, with the rule that selects them.

Exclusion list (component ids): HOH DOD SO4 PO4 GOL EDO PEG PG4 PGE 1PE P6G PE4 PE8 DMS ACT ACY FMT TRS MES EPE MPD BME DTT NA K CL BR IOD MG CA ZN MN NI CO CU FE CD NO3 NH4 AZI CIT TLA MLI EOH IPA BU3 PGO PDO FLC BCT SCN CAC IMD NHE HEZ UNX UNL

5. Reporting rules

Every count names its surface (entries, models, list entries, findings). SATYA's verdict is stated whenever findings are. No model, generator or experimental structure is called wrong, in error or incorrect on reference-free evidence (roadmap 2.2). No statement that a generator missed anything. No frequency, rate or cross-generator statement before roadmap step M6. The omega "cis" label question (roadmap M2) is resolved, by a verifier successor or a visible reading note, before any document from this collection is published.

6. Evidence

```
sigma\protein\mcl1_collection\sources\MCL1_SOURCES.json     the pinned sources and the derived panel
sigma\protein\mcl1_collection\sources\pdb\<ID>.cif          experimental files as served, unchanged
sigma\protein\mcl1_collection\sources\uniprot\Q07820.fasta  the pinned sequence
sigma\protein\mcl1_collection\MCL1_COLLECTION_INDEX.json    every model, its source, digest, outcome, delivery
sigma\protein\mcl1_collection\batch\<set>\<id>\             per-model progress and console from the route
sigma\protein\viewer_b7\public_workbench_v1\runs\PUB1_DELIVERY_<m>\   the route's own evidence per model
```

Changelog

```
V1  2026-09-30  frozen: target, experimental control set rule, files unchanged, route, pilot, panel derivation
                rule, reporting rules. Generation grid deferred to V1.1.
```

MCL1_COLLECTION_PROTOCOL_V1_1.md (sha256 527782c60f05692f)

MCL1 OPEN RESULTS COLLECTION: PROTOCOL V1.1 (the generation grid)

Version: V1.1, frozen 2026-10-01, before any generated model runs. Amends V1 by adding sections 7 to 12. V1 (MCL1_COLLECTION_PROTOCOL_V1.md) is unchanged and stays the governing text for the experimental control set; the batch driver binds V1's digest in every batch record, and the generation preparation script binds V1 and V1.1 together. Amendment rule as in V1: a later version is written before the runs it affects; runs already made stand.

7. Generators and environments

Every generator version is resolved inside the interpreter that runs it, through importlib.metadata or the generator's own version API, by the generation preparation script, and written to sources\MCL1_GENERATION_SOURCES.json before any generation command is issued. A version typed by hand is not a pin.

```
Boltz-2      interpreter C:\Dev\kha\.venv-boltz\Scripts\python.exe, the environment that produced the KRAS
             production candidates; expected version 2.2.1 and checkpoint hashes as recorded in the KRAS Tier 1
             record, re-resolved at pin time and required to agree. MSAs per construct from the same ColabFold
             endpoint and tooling the KRAS production used, pinned by digest, no --use_msa_server and no
             --use_potentials in production commands. 5 diffusion samples per invocation.
OpenFold3    the environment that produced mcl1_seed_42_sample_1_model.cif (the published case). Its location is
             not recorded in the tree (demo_data\hf_openfold3 is empty); Jason names it, and the preparation script
             pins its version and the input format it consumed. 5 samples per invocation as in the published run.
Chai-1, Protenix   added by amendment V1.2 once installed in their own environments and version-pinned; the grid
             below applies to them unchanged. Not part of V1.1 runs.
AlphaFold3   excluded: server outputs carry non-commercial terms (roadmap C2).
```

8. Inputs: one pair per experimental entry and panel ligand

The pairing rule makes every generated model comparable to one experimental entry.

Protein	the sequence of the polymer entity that references UniProt Q07820 in that entry, as given by the entry's own entity_poly canonical one-letter sequence (pinned from the RCSB data API, never typed); one copy; partner polymers (BH3 peptides, antibodies, fusion partners) excluded.
Ligand	the panel component of that entry, by its CCD identifier and the SMILES pinned in MCL1_SOURCES.json; ions, solvents, buffers and cryoprotectants excluded by the V1 exclusion list.
Pair id	<entry>_<component>, e.g. 6OQB_<comp>. An entry with two panel components yields two pairs.

Tiers, so the grid can start before it is finished:

Tier 1	pairs whose component occurs in at least two entries; for each such component, the entry with the best (lowest) resolution, ties by earliest identifier. One pair per component.
Tier 2	every remaining (entry, component) pair, in identifier order.

9. Sampling

seeds	1, 2, 3 for every generator
samples	5 per seed per generator (Boltz-2 diffusion samples; OpenFold3 samples)
models	per pair per generator: 15; per pair with two generators: 30

10. Companion files

Each generator's own summary files stay beside the model file under the names the generator wrote (Boltz-2 per-sample confidence JSON; OpenFold3 aggregated confidences), so the accepted route records them as companion files present and the index carries the generator-reported values next to SATYA's findings for every model. No value is transcribed by hand.

11. Route, sets and recording

As V1 section 3: the accepted route, satya_pub1_bundle_export_v3.py, exactly as the front door service invokes it; set names boltz2 and openfold3 (chai1, protenix later); the model source string is the generator name and pinned version; one delivery per model; every outcome recorded; no exclusion after the fact. The generation preparation script writes the model manifest the batch driver consumes, with each model's pair id, generator, seed, sample, digest and companion files, so every generated model is paired in the index with its experimental entry.

12. Reporting

As V1 section 5. In addition: no statement comparing generators, pairs or tiers is made before roadmap step M6, and any such statement after M6 names the exact subset it was measured on.

Changelog

V1.1	2026-10-01	frozen: generators and pinning rule, pairing rule and tiers, seeds and samples, companion files, sets and recording. OpenFold3's environment to be named by Jason and pinned by the preparation script before its first run. Chai-1 and Protenix deferred to V1.2.

MCL1_COLLECTION_PROTOCOL_V1_2.md (sha256 7137f926465e827e)

MCL1 OPEN RESULTS COLLECTION: PROTOCOL V1.2 (OpenFold3 provenance corrected; published OpenFold3 sets added)

Version: V1.2, frozen 2026-10-01, before any generated or third-party model runs. Amends V1.1. V1 and V1.1 stay unchanged; where V1.2 and V1.1 differ, V1.2 governs.

13. Correction to V1.1 section 7 (OpenFold3)

FACT. V1.1 section 7 assumed an OpenFold3 environment on Jason's machine had produced mcl1_seed_42_sample_1_model.cif. It did not. The seed 42 MCL1 files in sigma\protein\cofold\demo_data (samples 1 to 5, models and aggregated confidences) are the OpenFold consortium's own published example outputs, downloaded in May 2026 from the Hugging Face repository OpenFold/OpenFold3, path examples/common_examples/mcl1/seed_42 (record: conversation of 2026-05-28). The repository is licensed Apache-2.0. No OpenFold3 installation exists in the tree.

CORRECTION. OpenFold3 enters the collection in two ways: the published OpenFold3 sets in section 14, now; and a locally generated OpenFold3 set under the V1.1 grid, only after an installation exists (section 15).

COMPOSITION. The official MCL1 example is four MCL1 chains (A to D) with ATP on chains F, G and H and a small-molecule ligand on chain Z, in the context of PDB 5FDR. Sources: our 2026-05-28 analysis of sample 1 (ATP on F, G, H; ligand LIG0 on Z) and the published description of the same example in the Zonca record. Any statement that calls the published case "MCL1 with ATP" alone is incomplete and is corrected wherever it appears.

14. Published OpenFold3 sets (third-party, run as delivered, credited)

```
T1  OpenFold consortium example outputs. Source: Hugging Face OpenFold/OpenFold3,
    examples/common_examples/mcl1/seed_42; 5 models (mmCIF) with aggregated confidences; license Apache-2.0.
    Files: the local copies in demo_data, pinned by SHA-256 by the third-party preparation script, with the source
    path and the date obtained recorded as declared provenance; re-verified against the repository when an access
    token for the gated repository is available, otherwise reported as not re-verified. Set name openfold3_hf.
T2  Zonca, Andrea (San Diego Supercomputer Center), "OpenFold3 MCL1 protein-ligand ensemble predictions (20
    structures) on SDSC Expanse V100", Zenodo, DOI 10.5281/zenodo.21926059, published 2026-08-14, license
    CC-BY-4.0. OpenFold3 0.4.5; seeds 42, 1337, 2024, 2026; 5 diffusion samples each; MSAs from the ColabFold
    server; 20 models in PDB format with confidence JSONs. Files downloaded from the Zenodo record, each archive's
    MD5 checked against the record and pinned by SHA-256, every extracted file pinned. Set name openfold3_zonca.
```

Format rule for T2. The accepted route reads .cif and .mmcif only and refuses other formats by design (front door acceptance G00 and U03). T2 models are therefore converted from PDB to mmCIF by a pinned converter (gemmi, version resolved at preparation time), atom records unchanged and no other edit; each converted file records its source PDB's digest, the converter and version, and its own digest. Results on T2 are reported as results on the mmCIF conversion of the published PDB files. A file the converter cannot convert is recorded as such and not run.

Experimental pairing for T1 and T2: PDB 5FDR, already in the experimental control set.

Credit: every published result on T1 and T2 names its source and license as above; the collection's deposit carries the Apache-2.0 notice for T1 and the CC-BY-4.0 attribution for T2.

15. Locally generated OpenFold3 (deferred)

When an OpenFold3 installation exists (pip package openfold3 with its model parameters, which are gated behind the repository's terms form, on a CUDA GPU; the documented platform is Linux), its version is resolved in its own interpreter and pinned as V1.1 section 7 requires, and it runs the V1.1 grid. Until then, the V1.1 grid runs with Boltz-2 only.

16. Run order under V1.2

```
1  experimental control set (V1), complete except one entry whose delivery probes did not pass
2  T1, then T2 (third-party OpenFold3 sets)
3  Boltz-2 grid, Tier 1 then Tier 2 (V1.1)
4  OpenFold3 local grid, when section 15 is met
5  Chai-1 and Protenix, by a later amendment
```

Changelog

```
V1.2  2026-10-01  frozen: OpenFold3 provenance corrected (the seed 42 files are the OpenFold consortium's published
                  example outputs, Apache-2.0; no local installation exists); composition of the official example
                  stated (four MCL1 chains, ATP on F, G, H, a small-molecule ligand on Z); published OpenFold3 sets
                  T1 and T2 added with credit and a pinned PDB-to-mmCIF conversion rule for T2; local OpenFold3
                  generation deferred; run order. Chai-1 and Protenix move to a later amendment.
```

MCL1_COLLECTION_PROTOCOL_V1_3.md (sha256 21190ee4d00e1ce0)

MCL1 OPEN RESULTS COLLECTION: PROTOCOL V1.3 (T2 conversion, second form)

Version: V1.3, frozen 2026-10-01, before the T2 rerun. Amends V1.2 section 14's format rule. V1, V1.1 and V1.2 stay unchanged; where V1.3 and V1.2 differ, V1.3 governs.

17. What happened under V1.2

The V1.2 conversion (gemmi 0.6.5, satya_mcl1_pdb_to_mmcif_converter_v1: read, set up entities, write) produced mmCIF files whose polymer atoms carry no label_seq_id, because the published PDB files carry no SEQRES records. The accepted route's parser refuses a polymer atom without label_seq_id by design (sigma.protein.cofold.cofold_ingest, Amendment 4 A18.7: peptide topology is read only from label_seq_id). All 20 converted T2 models were refused, INPUT_REFUSED_BY_PARSER_POLICY, on 2026-10-01. Those 20 refusals stand as recorded results of the V1.2 conversion, with their files and digests; nothing is deleted.

18. The second conversion form

PDB exposes one namespace for chain and residue number; mmCIF separates the author namespace from the label (topology) namespace. SATYA's own PDB reader (cofold_ingest.parse_cofold_pdb) states the rule for this case: the PDB chain fills both chain roles and the PDB residue number fills both number roles, because the format does not distinguish them, and that is not fabrication. The second conversion form carries exactly that rule into the mmCIF file:

```
auth_asym_id and label_asym_id   the PDB chain identifier
auth_seq_id                      the PDB residue number
label_seq_id                     the PDB residue number, for polymer residues only
everything else                  as written by gemmi from the PDB atom records, unchanged
```

Converter: satya_mcl1_pdb_to_mmcif_converter_v2 (embedded in satya_mcl1_thirdparty_prep_v2.py, written beside its outputs and pinned by digest), gemmi resolved in the Boltz environment's interpreter. It refuses, rather than inventing identifiers for, any chain that mixes polymer and non-polymer residues or holds more than one non-polymer residue; a refused file is listed as unavailable with its reason and not run. Each converted file records its source PDB's digest (verified equal to the digest recorded at download), the converter, gemmi's version, and its own digest.

Set name openfold3_zonca_conv2. Results are reported as results on the second-form mmCIF conversion of the published PDB files, with the V1.2 refusals reported beside them.

Changelog

```
V1.3  2026-10-01  frozen: the V1.2 conversion's 20 parser-policy refusals recorded and kept; second conversion form
                  defined from SATYA's own one-namespace PDB rule (chain into both chain fields, residue number into
                  both number fields for polymer residues); converter v2 with its refusal cases; set
                  openfold3_zonca_conv2.
```

MCL1_COLLECTION_PROTOCOL_V1_4.md (sha256 f2c98026aa96003d)

MCL1 OPEN RESULTS COLLECTION: PROTOCOL V1.4 (conversion neutrality control)

Version: V1.4, frozen 2026-10-01, before the control runs. Adds sections 19 to 22. V1 to V1.3 unchanged.

19. Why

T2 results (set openfold3_zonca_conv2) come from mmCIF files converted from the published PDB files. On 2026-10-01 the 18 completed T2 models carried 0 to 5 findings each, against 14 to 42 for the 5 T1 models (native mmCIF). The difference has several possible sources: the OpenFold3 version, the MSAs, the hardware, the seeds, and the conversion path. Only the conversion path is ours, and this control measures it.

20. Procedure

For each T1 model (5 native mmCIF files, already run through the route): gemmi (Boltz environment interpreter) writes it as PDB; converter v2, byte-identical to the T2 converter (digest checked against the T2 manifest), converts that PDB back to mmCIF; the companion confidences file is placed beside it as for T2; the result runs through the accepted route as set openfold3_hf_roundtrip. Every step's digest is recorded.

21. Comparison

Per model, against its native run: the verifier's numbers (verdict, constraints, violated, critical, findings) and the delivered payload's viewer entries, with atom serial numbers, digests and paths removed. Outcome per model IDENTICAL or DIFFERENT, with the first difference; a model whose run did not complete is NOT_COMPARABLE.

22. Reading

All five IDENTICAL: the conversion path is recorded as neutral on these files, and T2 results are read as results on the published predictions, with the remaining sources (version, MSAs, hardware, seeds) still confounded and not separable by this collection. Any DIFFERENT: every T2 result carries the measured difference as a stated caveat. In both cases no cross-set, cross-version or cross-generator statement is made before roadmap step M6.

Changelog

```
V1.4  2026-10-01  frozen: conversion neutrality control on the five T1 models (PDB round trip through converter v2),
                  comparison rule and reading rule.
```

MCL1_COLLECTION_PROTOCOL_V1_5.md (sha256 85bc2b00e6e25481)

MCL1 OPEN RESULTS COLLECTION: PROTOCOL V1.5 (conversion control, first step corrected)

Version: V1.5, frozen 2026-10-01, before the corrected control runs. Amends V1.4 section 20. V1 to V1.4 unchanged.

23. What happened under V1.4

The control's first step (native mmCIF to PDB) read zero atoms from all five T1 files without raising an error. The five PDB files it wrote hold only a CRYST1 line and END and share one digest, c8ba6687139a; converter v2 converted those empty files without refusing them, and the route refused the empty results ("No _atom_site loop found"). The control was NOT_COMPARABLE for all five. Cause, read from the files: OpenFold3's mmCIF output has no _atom_site.occupancy column; gemmi's mmCIF reader requires that column and, when it is absent, returns an atom table with no rows instead of failing. SATYA's own mmCIF parser does not read occupancy (it sets 1.0 for every mmCIF atom), which is why the native files ran. T2 is not affected: its gemmi step read the published PDB files, which carry occupancy, and its 20 conversions each recorded all eight chains.

24. Corrected first step

Before gemmi builds the structure, a file whose _atom_site table has no occupancy column gets one, with the value 1.0 for every atom, the same value SATYA's parser assumes for mmCIF atoms; whether the column was

supplied is recorded per file. Then gemmi writes the PDB.

Both steps now refuse empty or partial output: the structure gemmi builds must hold exactly as many atoms as the native file's _atom_site rows, the PDB must hold that many ATOM and HETATM records, and the converted mmCIF must hold that many _atom_site rows; a mismatch makes the model NOT_COMPARABLE with the counts recorded, and it is not run.

Converter v2 is the object under test and stays byte-identical to the T2 converter. Its own lack of an empty-input refusal is recorded here as a known gap; it did not affect T2, whose inputs all carried atoms.

Set name openfold3_hf_roundtrip_v2. The V1.4 run's refusals and NOT_COMPARABLE records stay as made.

Changelog

```
V1.5  2026-10-01  frozen: the V1.4 control failure and its cause (no occupancy column in OpenFold3's mmCIF, which
                  gemmi requires and silently treats as zero atoms); occupancy 1.0 supplied before gemmi reads,
                  recorded per file; atom counts verified at every step; converter v2 unchanged; set
                  openfold3_hf_roundtrip_v2.
```

MCL1_COLLECTION_PROTOCOL_V1_6.md (sha256 6f768302b3b2e264)

MCL1 OPEN RESULTS COLLECTION: PROTOCOL V1.6 (Boltz-2 inputs, made exact)

Version: V1.6, frozen 2026-10-01, before any Boltz-2 input is built. Makes V1.1 sections 7 to 11 exact where they left a choice open. V1 to V1.5 unchanged. The Boltz-2 grid follows the KRAS production path, by importing its functions, not by re-implementing them.

25. Tier 1 selection

Panel components (MCL1_SOURCES.json, ligand_panel) present in two or more entries. Per component, the entry with the lowest resolution (the smallest value in resolution_combined); an entry without a resolution ranks after every entry with one; ties by earliest identifier. Pair id <entry>_<component>.

26. Protein sequence

From the RCSB data API: the entry's polymer entities whose reference sequence identifiers include UniProt Q07820. Exactly one distinct canonical one-letter sequence (entity_poly.pdbx_seq_one_letter_code_can, whitespace removed) is required, made of the twenty standard residues only; otherwise the pair is REFUSED with its reason and is not replaced by another entry. One copy, chain A. Expression tags stay as deposited, because the construct is what the entry contains.

27. Ligand declaration

Chain B, one copy. By the component's CCD code when Boltz-2's own component library holds it (cache\mols\<code>.pkl, the library that resolved the KRAS ligand 6IC); otherwise by the pinned stereo SMILES; otherwise by the pinned SMILES; a pair with none of these is REFUSED. The declaration used is recorded per pair.

28. MSAs

One per distinct sequence, by the KRAS procedure: make_kras_msas.build_argv on an MSA build job of the protein and its pair's ligand with no msa key (--use_msa_server, one diffusion sample, plumbing seed 1000, a build tree outside the predictions tree marked NOT_A_CANDIDATE); the generated CSV is copied into the inputs, re-hashed, and accepted only when msa_min.assert_query binds its query row exactly to the sequence. As KRAS records: the remote MSA service and its databases are not frozen; the pinned file is what makes generation reproducible, and production generation reads it from disk with no network request. An already pinned MSA whose binding passes is reused, never rebuilt.

29. Generator and commands

Interpreter, executable, cache, accelerator, kernel policy and output format are inherited from the frozen Mpro generation freeze through make_kras_msas.inherit_generator_environment; the Boltz version is resolved through importlib.metadata in that interpreter and must equal the version the freeze records. Production commands: 5 diffusion samples, seeds 1, 2 and 3, one output directory per pair and seed under generation\boltz2\predictions, no --use_msa_server, no --use_potentials, and nothing the inherited freeze forbids.

30. Freeze before generation

Every input (selection, sequences, ligand declarations, job files, pinned MSAs, production commands) is written with its digest to GENERATION_INPUTS_BOLTZ2_TIER1.json before any production command runs. A frozen record is never rewritten; the generation runner verifies every digest before it starts and runs only what the record lists.

Changelog

V1.6	2026-10-01	frozen: Tier 1 selection with resolution ranking; sequence source and refusal cases; ligand declaration rule (CCD from Boltz's own library, else stereo SMILES, else SMILES); MSAs by the KRAS procedure with exact query-row binding; generator identity inherited and version-checked; productio flags; freeze before generation.

Latest deposit on the tree: ZENODO_DEPOSIT_5

SATYA MCL1 Open Results Collection: version 2

Deposited 2026-10-01 by Invariant Research (Jason Leonard Volk). Version 1, published 2026-10-01 as DOI 10.5281/zenodo.23076000, pre-registered Protocols V1 to V1.3 before any Boltz-2 model was generated. This version adds Protocols V1.4 and V1.5, the published OpenFold3 sets run through the accepted SATYA route, and the conversion neutrality control.

State at deposit, terminal states by set:

```
experimental                COMPLETE 124, MACHINERY_FAILURE 1
openfold3_hf                COMPLETE 5
openfold3_hf_roundtrip      INPUT_REFUSED 5
openfold3_hf_roundtrip_v2   COMPLETE 5
openfold3_zonca             INPUT_REFUSED 20
openfold3_zonca_conv2       COMPLETE 20
```

Conversion neutrality control (Protocols V1.4 and V1.5): MCL1 CONVERSION CONTROL V2: 0 identical, 5 different (0 of them names only), 0 not comparable.

```
of3hf_rt2_s42_sample1    DIFFERENT  $.verifier.num_violated: 99 vs 100
of3hf_rt2_s42_sample2    DIFFERENT  $.verifier.findings: 19 vs 20
of3hf_rt2_s42_sample3    DIFFERENT  $.verifier.findings: 20 vs 21
of3hf_rt2_s42_sample4    DIFFERENT  $.verifier.num_violated: 100 vs 99
of3hf_rt2_s42_sample5    DIFFERENT  $.verifier.findings: 42 vs 43
```

Reading, as frozen in V1.4 section 22: every result on the converted T2 files carries the measured difference as a caveat. On the five T1 files the PDB round trip left every constraint total unchanged and moved the violated count by at most two and the finding count by at most one per model; the PDB format stores coordinates to three decimals, against up to eight significant digits in OpenFold3's mmCIF.

Credits: T1 models are the OpenFold consortium's published OpenFold3 example outputs (Hugging Face OpenFold/OpenFold3, Apache-2.0); T2 models are Zonca, Andrea (San Diego Supercomputer Center), DOI 10.5281/zenodo.21926059 (CC-BY-4.0). This record carries manifests, digests and SATYA's numbers for those models, not the model files themselves.

Not yet deposited: per-model SATYA evidence documents, which wait for roadmap step M2 (the omega label). No statement in this record concerns the correctness of any structure, generator or verifier, and no frequency, rate, cross-set or cross-generator statement is made. SATYA's findings are reference-free, with SATYA's own verdict stated per model.

Files:

```
MCL1_COLLECTION_PROTOCOL_V1.md                    5280 bytes  sha256 76fa4eab99f0049c
MCL1_COLLECTION_PROTOCOL_V1_1.md                  4927 bytes  sha256 527782c60f05692f
MCL1_COLLECTION_PROTOCOL_V1_2.md                  4869 bytes  sha256 7137f926465e827e
MCL1_COLLECTION_PROTOCOL_V1_3.md                  2847 bytes  sha256 21190ee4d00e1ce0
MCL1_COLLECTION_PROTOCOL_V1_4.md                  2060 bytes  sha256 f2c98026aa96003d
MCL1_COLLECTION_PROTOCOL_V1_5.md                  2461 bytes  sha256 85bc2b00e6e25481
satya_mcl1_batch_route_v1.py                     15052 bytes  sha256 28b910de37f793a4
satya_mcl1_batch_route_v2.py                     16243 bytes  sha256 b668d4d9fc790e1f
satya_mcl1_controls_audit_v1.py                  17974 bytes  sha256 9c6000aa156faf7e
satya_mcl1_conversion_control_v1.py              13710 bytes  sha256 99a66daadb05da57
satya_mcl1_conversion_control_v2.py              16908 bytes  sha256 8bd9fa5968e394d6
satya_mcl1_prep_v1.py                            16068 bytes  sha256 81ffffbe999768fe
satya_mcl1_thirdparty_prep_v1.py                 24121 bytes  sha256 6de78193b3e952b4
satya_mcl1_thirdparty_prep_v2.py                 11329 bytes  sha256 f99f9d195ad0caca
satya_mcl1_zenodo_deposit_v1.py                  16616 bytes  sha256 a353c58aed4ae46a
satya_mcl1_zenodo_deposit_v2.py                  16085 bytes  sha256 61323fb62e3d6bb4
MCL1_SOURCES.json                               179875 bytes  sha256 6b2cfa8de219a3fc
PREP_CONSOLE.txt                                  8556 bytes  sha256 8089b67faab2caee
MCL1_COLLECTION_INDEX.json                     1040281 bytes  sha256 5d2e50e3fb54b38e
MODELS_OPENFOLD3_HF.json                         20009 bytes  sha256 81add450d5a19b0a
MODELS_OPENFOLD3_HF_ROUNDTRIP.json                9880 bytes  sha256 4eabbdb61663cdbc
MODELS_OPENFOLD3_HF_ROUNDTRIP_V2.json            16392 bytes  sha256 78588e6313a677d8
MODELS_OPENFOLD3_ZONCA.json                      88406 bytes  sha256 3d4092593a8925b1
MODELS_OPENFOLD3_ZONCA_CONV2.json               102344 bytes  sha256 2a0c915d4c9f8cf4
```

```
THIRDPARTY_SOURCES.json                              2055 bytes  sha256 ba1c4d65caede3b7
THIRDPARTY_CONSOLE.txt                               4476 bytes  sha256 03dc9371e5ee8750
THIRDPARTY_CONV2_CONSOLE.txt                         2415 bytes  sha256 572baf9d20664c7c
CONVERSION_CONTROL_COMPARE.json                      1288 bytes  sha256 3c327c51da323aef
CONVERSION_CONTROL_PREPARE.json                      9304 bytes  sha256 1873efad31a3ad5d
CONVERSION_CONTROL_COMPARE_CONSOLE.txt                715 bytes  sha256 453f86018fa86930
CONVERSION_CONTROL_PREPARE_CONSOLE.txt                738 bytes  sha256 c7f4d6f69a59ec6c
CONVERSION_CONTROL_V2_COMPARE.json                   3436 bytes  sha256 9d848472f054e722
CONVERSION_CONTROL_V2_PREPARE.json                  15636 bytes  sha256 d59ab60ef984ffe6
CONVERSION_CONTROL_V2_COMPARE_CONSOLE.txt             887 bytes  sha256 cb432bf294a1c92b
CONVERSION_CONTROL_V2_PREPARE_CONSOLE.txt            1012 bytes  sha256 46ebd81f2029f486
CONTROLS_AUDIT.json                               1661434 bytes  sha256 4e4dadea9df9bdac
CONTROLS_AUDIT_CONSOLE.txt                          24871 bytes  sha256 2ed03d442d47aea1
BATCH_CONSOLE_20260930T212454Z.txt                   1836 bytes  sha256 33795950fd369c2f
BATCH_CONSOLE_20261001T003704Z.txt                  16262 bytes  sha256 0307e4382fd52dda
BATCH_CONSOLE_20261001T013358Z.txt                   9297 bytes  sha256 b578c1183d6ddb84
BATCH_CONSOLE_20261001T060459Z.txt                   7969 bytes  sha256 57c6c08ea44bbc71
BATCH_CONSOLE_20261001T071942Z.txt                   1402 bytes  sha256 128979a6e03255f9
BATCH_CONSOLE_20261001T075910Z.txt                   1719 bytes  sha256 3f57a8cdbabf295e
BATCH_CONSOLE_20261001T132034Z.txt                   1496 bytes  sha256 c304237a6d7e41cd
BATCH_CONSOLE_20261001T072000Z.txt                   4568 bytes  sha256 da9d0d3c2c07de64
BATCH_CONSOLE_20261001T074808Z.txt                   3574 bytes  sha256 bd6e6964c80c1787
BATCH_CONSOLE_20261001T080111Z.txt                   2324 bytes  sha256 df1bd865e6cb982d
BATCH_CONSOLE_20261001T132157Z.txt                   2196 bytes  sha256 50eb06c56d4c1a19
```

Files in the deposit

File	Bytes	SHA-256 (first 24)
MCL1_COLLECTION_PROTOCOL_V1.md	5280	76fa4eab99f0049c4bda45a5
MCL1_COLLECTION_PROTOCOL_V1_1.md	4927	527782c60f05692f9544b87e
MCL1_COLLECTION_PROTOCOL_V1_2.md	4869	7137f926465e827e984ad494
MCL1_COLLECTION_PROTOCOL_V1_3.md	2847	21190ee4d00e1ce036bebb40
MCL1_COLLECTION_PROTOCOL_V1_4.md	2060	f2c98026aa96003d82b9634c
MCL1_COLLECTION_PROTOCOL_V1_5.md	2461	85bc2b00e6e25481d3d73001
satya_mcl1_batch_route_v1.py	15052	28b910de37f793a481a1bbf4
satya_mcl1_batch_route_v2.py	16243	b668d4d9fc790e1fe3dd80ab
satya_mcl1_controls_audit_v1.py	17974	9c6000aa156faf7e6cafd133
satya_mcl1_conversion_control_v1.py	13710	99a66daadb05da5728fefebf
satya_mcl1_conversion_control_v2.py	16908	8bd9fa5968e394d661d1ba6d
satya_mcl1_prep_v1.py	16068	81ffffbe999768fe3a712059
satya_mcl1_thirdparty_prep_v1.py	24121	6de78193b3e952b40f4b3bba
satya_mcl1_thirdparty_prep_v2.py	11329	f99f9d195ad0caca4e7dc3ad
satya_mcl1_zenodo_deposit_v1.py	16616	a353c58aed4ae46ac32e8a25
satya_mcl1_zenodo_deposit_v2.py	16085	61323fb62e3d6bb4df8fcd72
MCL1_SOURCES.json	179875	6b2cfa8de219a3fcbc59052e
PREP_CONSOLE.txt	8556	8089b67faab2caee6717d3ec
MCL1_COLLECTION_INDEX.json	1040281	5d2e50e3fb54b38ea7409bac
MODELS_OPENFOLD3_HF.json	20009	81add450d5a19b0a87212d5f
MODELS_OPENFOLD3_HF_ROUNDTRIP.json	9880	4eabbdb61663cdbce8ce8b9a
MODELS_OPENFOLD3_HF_ROUNDTRIP_V2.json	16392	78588e6313a677d8ad5cdafe

File	Bytes	SHA-256 (first 24)
MODELS_OPENFOLD3_ZONCA.json	88406	3d4092593a8925b1962676bb
MODELS_OPENFOLD3_ZONCA_CONV2.json	102344	2a0c915d4c9f8cf405e7ee42
THIRDPARTY_SOURCES.json	2055	ba1c4d65caede3b77faeaa53
THIRDPARTY_CONSOLE.txt	4476	03dc9371e5ee8750fea593a3
THIRDPARTY_CONV2_CONSOLE.txt	2415	572baf9d20664c7c05a01195
CONVERSION_CONTROL_COMPARE.json	1288	3c327c51da323aefd7d433f9
CONVERSION_CONTROL_PREPARE.json	9304	1873efad31a3ad5daa337616
CONVERSION_CONTROL_COMPARE_CONSOLE.txt	715	453f86018fa86930d9d2d7c4
CONVERSION_CONTROL_PREPARE_CONSOLE.txt	738	c7f4d6f69a59ec6c9ec986cf
CONVERSION_CONTROL_V2_COMPARE.json	3436	9d848472f054e7229c396ef5
CONVERSION_CONTROL_V2_PREPARE.json	15636	d59ab60ef984ffe6a49e5f6f
CONVERSION_CONTROL_V2_COMPARE_CONSOLE.txt	887	cb432bf294a1c92b4f10b12f
CONVERSION_CONTROL_V2_PREPARE_CONSOLE.txt	1012	46ebd81f2029f486ed60d5f6
CONTROLS_AUDIT.json	1661434	4e4dadea9df9bdacdd7468ef
CONTROLS_AUDIT_CONSOLE.txt	24871	2ed03d442d47aea17d833237
BATCH_CONSOLE_20260930T212454Z.txt	1836	33795950fd369c2f5faad39c
BATCH_CONSOLE_20261001T003704Z.txt	16262	0307e4382fd52dda5ceeae28
BATCH_CONSOLE_20261001T013358Z.txt	9297	b578c1183d6ddb84e9c18f86
BATCH_CONSOLE_20261001T060459Z.txt	7969	57c6c08ea44bbc71ab50726d
BATCH_CONSOLE_20261001T071942Z.txt	1402	128979a6e03255f99298d47d
BATCH_CONSOLE_20261001T075910Z.txt	1719	3f57a8cdbabf295e9a93987d
BATCH_CONSOLE_20261001T132034Z.txt	1496	c304237a6d7e41cd46d8cdac
BATCH_CONSOLE_20261001T072000Z.txt	4568	da9d0d3c2c07de64a4bf0bd9
BATCH_CONSOLE_20261001T074808Z.txt	3574	bd6e6964c80c178780906223
BATCH_CONSOLE_20261001T080111Z.txt	2324	df1bd865e6cb982db9446022
BATCH_CONSOLE_20261001T132157Z.txt	2196	50eb06c56d4c1a1944818455