



# wwPDB Data Inconsistency Report: Missing \_struct\_asym Entries

**Report Date:** 2026-01-18

**Archive Scan Date:** 2026-01-17

**Repository:** <https://github.com/N283T/fix-struct-asym>

---

## Executive Summary

A systematic scan of the PDB archive identified **3 entries** (out of 243,083 examined) containing `_atom_site.label_asym_id` values that lack corresponding entries in the `_struct_asym` category. All affected cases involve water asym\_ids.

**Affected Entries:** 1TS6, 2G10, 2K9Y **Impact Rate:** 0.001% of entries examined

**Terminology note:** In this report, “asym\_id” refers to `_atom_site.label_asym_id` and `_struct_asym.id` values (not `auth_asym_id`).

## Problem Description

### Issue

According to the mmCIF dictionary, the `_struct_asym` category defines the asymmetric unit contents and should contain an entry for every `label_asym_id` referenced by `_atom_site`. The affected entries contain atom records with `label_asym_id` values that have no corresponding `_struct_asym.id` definition.

### Technical Details

Aspect	Description
Expected behavior	Every unique <code>_atom_site.label_asym_id</code> value has a matching <code>_struct_asym.id</code> entry

Aspect	Description
Observed behavior	Certain water asym_ids are referenced in _atom_site but absent from _struct_asym
Affected category	_struct_asym
Pattern	All missing entries correspond to water asym_ids

## Affected Entries

PDB ID	Missing _struct_asym.id	Entity ID	Entity Type
1TS6	C	3	water
2G10	F	5	water
2K9Y	C, D	2	water

## Detailed Findings

**1TS6** Asym\_id C contains water molecules referenced in \_atom\_site but lacks a corresponding \_struct\_asym entry.

**2G10** Asym\_id F contains water molecules referenced in \_atom\_site but lacks a corresponding \_struct\_asym entry.

**2K9Y** Asym\_ids C and D contain water molecules referenced in \_atom\_site but lack corresponding \_struct\_asym entries.

## Example: 2G10

\_atom\_site records referencing asym\_id F (first 5 of 147):

```
_atom_site.group_PDB
_atom_site.id
_atom_site.label_atom_id
_atom_site.label_comp_id
_atom_site.label_asym_id      # <-- "F" exists in _atom_site
_atom_site.label_entity_id    # <-- References entity 5 (water)
_atom_site.label_seq_id
...
HETATM 2706 0 HOH F 5 . ...
HETATM 2707 0 HOH F 5 . ...
HETATM 2708 0 HOH F 5 . ...
HETATM 2709 0 HOH F 5 . ...
HETATM 2710 0 HOH F 5 . ...
```

#### **\_struct\_asym category (complete):**

```
_struct_asym.id
_struct_asym.pdbx_blank_PDB_chainid_flag
_struct_asym.pdbx_modified
_struct_asym.entity_id
_struct_asym.details
A N N 1 ?
B N N 2 ?
C N N 3 ?
D N N 4 ?
E N N 5 ?
```

Asym\_id “F” is referenced in \_atom\_site but has no corresponding entry in \_struct\_asym.

## **Methodology**

### **Detection Process**

1. Extracted all unique \_atom\_site.label\_asym\_id values from each entry
2. Extracted all \_struct\_asym.id values from each entry
3. Identified entries where the atom site set contains values absent from the struct\_asym set

## Scan Parameters

Parameter	Value
Total entries scanned	243,083
Mirror sync date	2025-10-10
Scan date	2026-01-17
Verification date	2026-01-18
Archive source	PDB mmCIF archive (mirror) / RCSB (verification)

Note: The initial scan used a local mirror synced approximately 3 months prior. To ensure the findings remain valid, the affected entries were re-downloaded directly from RCSB on 2026-01-18 and the inconsistencies were confirmed to persist.

## Validation

1. **Initial detection:** Systematic scan of local PDB mirror (2026-01-17)
2. **Verification:** Re-downloaded affected entries from RCSB (2026-01-18) and confirmed the inconsistencies persist in the current archive
3. **Fix validation:** Generated corrected files and re-scanned to confirm zero remaining inconsistencies

## Recommendations

1. **Remediation:** We kindly suggest updating the affected entries to include the missing `_struct_asym` definitions
2. **Prevention:** It may be worth considering a validation step in the deposition/processing pipeline to verify that all `_atom_site.label_asym_id` values have corresponding `_struct_asym.id` entries before release

# Supplementary Materials

Corrected mmCIF files and scan results are available at the GitHub repository:

- fixed/1ts6.cif
- fixed/2g10.cif
- fixed/2k9y.cif
- reports/data/results.json

**Repository:** <https://github.com/N283T/fix-struct-asym>

# Reproducibility

This report was generated using automated detection tools. The methodology, scan results, and corrected files are publicly available for verification at the repository above.