

wwPDB データ不整合報告: _struct_asym エントリの欠損

報告日: 2026-01-18

アーカイブスキャン日: 2026-01-17

リポジトリ: <https://github.com/N283T/fix-struct-asym>

概要

PDBアーカイブの体系的なスキャンにより、_atom_site.label_asym_id の値に対応する _struct_asym カテゴリのエントリが欠損している **3件** のエントリを特定しました（調査対象: 243,083件）。該当する全てのケースは water の asym_id に関するものです。

該当エントリ: 1TS6, 2G10, 2K9Y 影響率: 調査対象の0.001%

用語について: 本報告における “asym_id” は _atom_site.label_asym_id および _struct_asym.id の値を指します（auth_asym_id ではありません）。

問題の説明

概要

mmCIF辞書によれば、_struct_asym カテゴリは非対称単位の内容を定義し、_atom_site で参照される全ての label_asym_id に対応するエントリを含む必要があります。該当エントリには、対応する _struct_asym.id 定義のない label_asym_id 値を持つ原子レコードが含まれています。

技術的詳細

項目	説明
期待される動作	全ての一意な _atom_site.label_asym_id 値に対応する _struct_asym.id エントリが存在する

項目	説明
観測された動作	特定の水の asym_id が _atom_site で参照されているが _struct_asym に存在しない
影響カテゴリ	_struct_asym
パターン	欠損エントリは全て water の asym_id に対応

該当エントリ

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

詳細

1TS6 asym_id C には水分子が _atom_site で参照されているが、対応する _struct_asym エントリが欠損している。

2G10 asym_id F には水分子が _atom_site で参照されているが、対応する _struct_asym エントリが欠損している。

2K9Y asym_id C および D には水分子が _atom_site で参照されているが、対応する _struct_asym エントリが欠損している。

例: 2G10

asym_id F を参照する _atom_site レコード（147件中最初の5件）：

```
_atom_site.group_PDB
_atom_site.id
_atom_site.label_atom_id
_atom_site.label_comp_id
_atom_site.label_asym_id      # <-- "F" が _atom_site に存在
_atom_site.label_entity_id    # <-- entity 5 (water) を参照
_atom_site.label_seq_id
...
HETATM 2706  O HOH  F  5 .  ...
HETATM 2707  O HOH  F  5 .  ...
HETATM 2708  O HOH  F  5 .  ...
HETATM 2709  O HOH  F  5 .  ...
HETATM 2710  O HOH  F  5 .  ...
```

_struct_asym カテゴリ (全体) :

```
_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” は _atom_site で参照されているが、対応する _struct_asym エントリが存在しない。

検出方法

検出プロセス

1. 各エントリから全ての一意な _atom_site.label_asym_id 値を抽出
2. 各エントリから全ての _struct_asym.id 値を抽出
3. atom_site集合にあってstruct_asym集合にない値を持つエントリを特定

スキャンパラメータ

パラメータ	値
スキャン総数	243,083
ミラー同期日	2025-10-10
スキャン日	2026-01-17
検証日	2026-01-18
アーカイブソース	PDB mmCIFアーカイブ（ミラー） / RCSB（検証）

注: 初回スキャンは約3ヶ月前に同期されたローカルミラーを使用しました。検出結果の有効性を確認するため、2026-01-18にRCSBから該当エントリを直接再ダウンロードし、不整合が現在も継続していることを確認しました。

検証

1. **初回検出:** ローカルPDBミラーの体系的スキャン（2026-01-17）
2. **確認:** RCSBから該当エントリを再ダウンロードし、不整合が現在のアーカイブでも継続していることを確認（2026-01-18）
3. **修正検証:** 修正ファイルを生成し、再スキャンにより不整合がゼロであることを確認

提言

1. **是正について:** 該当エントリに欠損している `_struct_asym` 定義を追加していただけますと幸いです
2. **予防について:** デポジション/処理パイプラインにおいて、全ての `_atom_site.label_asym_id` 値に対応する `_struct_asym.id` エントリが存在することをリリース前に検証するバリデーションステップの追加をご検討いただければ幸いです

補足資料

修正済みmmCIFファイルおよびスキャン結果はGitHubリポジトリで公開しています：

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

リポジトリ: <https://github.com/N283T/fix-struct-async>

再現性

本報告は自動検出ツールを使用して生成されました。検出方法、スキャン結果、および修正ファイルは上記リポジトリで公開されており、検証可能です。