old_pdb

Nathan Baker

Feb 27, 2021

CONTENTS

This code reads and writes the old PDB format.

Development of this code was greatly helped by the Chimera documentation.

CHAPTER

ONE

ORIGIN OF THIS SOFTWARE

This software is derived, in part, from the PDB2PQR code which is supported by NIH grant GM069702.

CHAPTER

TWO

INSTALLATION

This python package can be installed via setuptools, `pip install .`

CHAPTER

THREE

TESTING

The software can be tested with pytest by running:

```
python -m pytest
```

from the top-level directory.

CHAPTER

FOUR

INDICES AND TABLES

- genindex
- modindex

CHAPTER

FIVE

CONTENTS

5.1 API reference

Note: The API is still changing. We use semantic versioning and our *Change log* to document changes between versions.

5.1.1 PDB data structures and I/O

PDB entry top-level class

`pdb_entry` module

Top-level module for PDB structure entries.

The specifications used in this class are derived from the Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description, Version 3.3.

class old_pdb.pdb_entry.**Entry**

Top-level class for PDB structure entry.

annotate_link(*record*) → *old_pdb.secondary.Link*

Annotate LINK to indicate whether the named atoms are elements.

Creates two new Boolean attributes in record: is_element1 and is_element2.

Parameters **record** (secondary.Link) – record to annotate

Returns annotated record

check_master()

Check the contents against internal bookkeeping records.

Raises **AssertionError** – if checks fail

find_atom_by_name(*chain_id*, *residue_id*, *atom_name*, *model_num=1*) → *old_pdb.coordinates.Atom*

Find a specific atom by name.

Parameters

- **chain_id** (*str*) – chain ID to find
- **residue_id** (*int*) – residue ID to find

- **atom_name** (*str*) – name of atom to find
- **model_num** (*int*) – model number to use

Returns ATOM or HETATM object

find_residue(*chain_id*, *residue_id*, *model_num=1*) → list
Find a specific residue.

Parameters

- **chain_id** (*str*) – chain ID to find
- **residue_id** (*int*) – residue ID to find
- **model_num** (*int*) – model number to use

Returns list of `coordinates.Atom`-like objects

num_atoms(*heavy_only=True*) → int
Number of ATOM and HETATM entries in all chains in entry.

Parameters **heavy_only** (*bool*) – exclude hydrogen atoms from count

num_chains() → int
Number of chains in entry.

num_residues(*count_hetatm=False*) → int
Number of residues in entry.

Parameters **count_hetam** (*bool*) – include heterogen residues in count

num_ter() → int
Number of TER records in entry.

num_transforms() → int
Return the number of optional transform records in entry.

Returns number of ORGIXn + SCALEn + MTRIXn

parse_file(*file_*)
Parse a PDB file.

Parameters **file** (*file*) – file open for reading.

parse_line(*line*)
Parse a line of a PDB file.

Parameters **line** (*str*) – line of PDB file

PDB records

annotation module

Classes for PDB records that provide annotation information.

class old_pdb.annotation.**Author**
AUTHOR field

The AUTHOR record contains the names of the people responsible for the contents of the entry.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"AUTHOR"	
9-10	Continuation	continuation	Allows concatenation of multiple records.
11-79	List	authorList	List of the author names, separated by commas.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Caveat**
CAVEAT field

CAVEAT warns of severe errors in an entry. Use caution when using an entry containing this record.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"CAVEAT"	
9-10	Continuation	continuation	Allows concatenation of multiple records.
12-15	IDcode	idCode	PDB ID code of this entry.
20-79	String	comment	Free text giving the reason for the CAVEAT.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Compound**
COMPND field

The COMPND record describes the macromolecular contents of an entry. Each macromolecule found in the entry is described by a set of token: value pairs, and is referred to as a COMPND record component. Since the concept of a molecule is difficult to specify exactly, PDB staff may exercise editorial judgment in consultation with depositors in assigning these names.

For each macromolecular component, the molecule name, synonyms, number assigned by the Enzyme Commission (EC), and other relevant details are specified.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"COMPND"	
8-10	Continuation	continuation	Allows concatenation of multiple records.
11-80	Specification list	compound	Description of the molecular components.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**ExperimentalData**
EXPDTA field

The EXPDTA record identifies the experimental technique used. This may refer to the type of radiation and sample, or include the spectroscopic or modeling technique. Permitted values include:

- ELECTRON DIFFRACTION
- FIBER DIFFRACTION
- FLUORESCENCE TRANSFER

- NEUTRON DIFFRACTION
- NMR
- THEORETICAL MODEL
- X-RAY DIFFRACTION

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"EXPDTA"	
9-10	Continuation	continuation	Allows concatenation of multiple records.
11-79	SList	technique	The experimental technique(s) with optional comment describing the sample or experiment.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Header**
HEADER field

The HEADER record uniquely identifies a PDB entry through the id_code field. This record also provides a classification for the entry. Finally, it contains the date the coordinates were deposited at the PDB.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"HEADER"	
11-50	String(40)	classification	Classifies the molecule(s).
51-59	Date	depDate	Deposition date. This is the date the coordinates were received at the PDB.
63-66	IDcode	idCode	This identifier is unique within the PDB.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Journal**
JRNL field

The JRNL record contains the primary literature citation that describes the experiment which resulted in the deposited coordinate set. There is at most one JRNL reference per entry. If there is no primary reference, then there is no JRNL reference. Other references are given in REMARK 1.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"JRNL"	
13-79	LString	text	See details in PDB specification.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Keywords**
KEYWDS field

The KEYWDS record contains a set of terms relevant to the entry. Terms in the KEYWDS record provide a simple means of categorizing entries and may be used to generate index files. This record addresses some of the limitations found in the classification field of the HEADER record. It provides the opportunity to add further annotation to the entry in a concise and computer-searchable fashion.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"KEYWDS"	
9-10	Continuation	continuation	Allows concatenation of records if necessary.
11-79	List	keywds	Comma-separated list of keywords relevant to the entry.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**ModelType**
MDLTYP field.

The MDLTYP record contains additional annotation pertinent to the coordinates presented in the entry.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"MDLTYP"	
9-10	Continuation	continuation	Allows concatenation of multiple records.
11-80	SList	comment	Free Text providing additional structural annotation.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**NumModels**
NUMMDL field

The NUMMDL record indicates total number of models in a PDB entry.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"NUMMDL"	
11-14	Integer	modelNumber	Number of models.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Obsolete**
OBSLTE field

This record acts as a flag in an entry which has been withdrawn from the PDB's full release. It indicates which, if any, new entries have replaced the withdrawn entry.

The format allows for the case of multiple new entries replacing one existing entry.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"OBSLTE"	
9-10	Continuation	continuation	Allows concatenation of multiple records
12-20	Date	repDate	Date that this entry was replaced.
22-25	IDcode	idCode	ID code of this entry.
32-35	IDcode	rIdCode	ID code of entry that replaced this one.
37-40	IDcode	rIdCode	ID code of entry that replaced this one.
42-45	IDcode	rIdCode	ID code of entry that replaced this one.
47-50	IDcode	rIdCode	ID code of entry that replaced this one.
52-55	IDcode	rIdCode	ID code of entry that replaced this one.
57-60	IDcode	rIdCode	ID code of entry that replaced this one.
62-65	IDcode	rIdCode	ID code of entry that replaced this one.
67-70	IDcode	rIdCode	ID code of entry that replaced this one.
72-75	IDcode	rIdCode	ID code of entry that replaced this one.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Remark**
REMARK field

REMARK records present experimental details, annotations, comments, and information not included in other records. In a number of cases, REMARKs are used to expand the contents of other record types. A new level of structure is being used for some REMARK records. This is expected to facilitate searching and will assist in the conversion to a relational database.

parse_line(*line*)
Initialize by parsing line.

COLUMNS	TYPE	FIELD	DEFINITION
8-10	int	remark_num	Remark number. It is not an error for remark n to exist in an entry when remark n-1 does not.
12-79	str	remark_text	Left as white space in first line of each new remark.

Parameters **line** (*str*) – line with PDB class

class old_pdb.annotation.**Revision**
Class to store contents of a single REVDAT modification.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"REVDAT"	
8-10	Integer	modNum	Modification number.
11-12	Continuation	continuation	Allows concatenation of multiple records.
14-22	Date	modDate	Date of modification (or for new entries) in DD-MMM-YY format. This is not repeated on continued lines.
24-27	IDCode	modId	ID code of this entry. This is not repeated on continuation lines.
32	Integer	modType	An integer identifying the type of modification. For all revisions, the modification type is listed as 1
40-45	LString(6)	record	Modification detail.
47-52	LString(6)	record	Modification detail.
54-59	LString(6)	record	Modification detail.
61-66	LString(6)	record	Modification detail.

parse_line(*line*)
Parse PDB-format line for specific revision.

Parameters **line** (*str*) – line to parse.

class old_pdb.annotation.**RevisionData**
REVDAT field

REVDAT records contain a history of the modifications made to an entry since its release.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"REVDAT"	
8-10	Integer	modNum	Modification number.
11-12	Continuation	continuation	Allows concatenation of multiple records.
14-22	Date	modDate	Date of modification (or for new entries) in DD-MMM-YY format. This is not repeated on continued lines.
24-27	IDCode	modId	ID code of this entry. This is not repeated on continuation lines.
32	Integer	modType	An integer identifying the type of modification. For all revisions, the modification type is listed as 1
40-45	LString(6)	record	Modification detail.
47-52	LString(6)	record	Modification detail.
54-59	LString(6)	record	Modification detail.
61-66	LString(6)	record	Modification detail.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Site**
SITE class

The SITE records supply the identification of groups comprising important sites in the macromolecule.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	“SITE “	
8-10	Integer	seqNum	Sequence number.
12-14	LString(3)	siteID	Site name.
16-17	Integer	numRes	Number of residues that compose the
			site.
19-21	Residue name	resName1	Residue name for first residue that creates the site.
23	Character	chainID1	Chain identifier for first residue of site.
24-27	Integer	seq1	Residue sequence number for first residue of the site.
28	AChar	iCode1	Insertion code for first residue of the site.
30-32	Residue name	resName2	Residue name for second residue that creates the site.
34	Character	chainID2	Chain identifier for second residue of the site.
35-38	Integer	seq2	Residue sequence number for second residue of the site.
39	AChar	iCode2	Insertion code for second residue of the site.
41-43	Residue name	resName3	Residue name for third residue that creates the site.
45	Character	chainID3	Chain identifier for third residue of the site.
46-49	Integer	seq3	Residue sequence number for third residue of the site.
50	AChar	iCode3	Insertion code for third residue of the site.
52-54	Residue name	resName4	Residue name for fourth residue that creates the site.
56	Character	chainID4	Chain identifier for fourth residue of the site.
57-60	Integer	seq4	Residue sequence number for fourth residue of the site.
61	AChar	iCode4	Insertion code for fourth residue of the site.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Source**
SOURCE field

The SOURCE record specifies the biological and/or chemical source of each biological molecule in the entry. Sources are described by both the common name and the scientific name, e.g., genus and species. Strain and/or cell-line for immortalized cells are given when they help to uniquely identify the biological entity studied.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	“SOURCE”	
8-10	Continuation	continua-tion	Allows concatenation of multiple records.
11-79	Specification List	srcName	Identifies the source of the macromolecule in a token: value format.

parse_line(*line*)
Parse a PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Split**
SPLIT field

The SPLIT record is used in instances where a specific entry composes part of a large macromolecular complex. It will identify the PDB entries that are required to reconstitute a complete complex.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"SPLIT "	
9-10	Continuation	continuation	Allows concatenation of multiple records.
12-15	IDcode	idCode	ID code of related entry.
17-20	IDcode	idCode	ID code of related entry.
22-25	IDcode	idCode	ID code of related entry.
27-30	IDcode	idCode	ID code of related entry.
32-35	IDcode	idCode	ID code of related entry.
37-40	IDcode	idCode	ID code of related entry.
42-45	IDcode	idCode	ID code of related entry.
47-50	IDcode	idCode	ID code of related entry.
52-55	IDcode	idCode	ID code of related entry.
57-60	IDcode	idCode	ID code of related entry.
62-65	IDcode	idCode	ID code of related entry.
67-70	IDcode	idCode	ID code of related entry.
72-75	IDcode	idCode	ID code of related entry.
77-80	IDcode	idCode	ID code of related entry.

parse_line(*line*)
Parse input line.

Parameters **line** (*str*) – PDB-format line to parse

class old_pdb.annotation.**Supersedes**
SPRSDE field

The SPRSDE records contain a list of the ID codes of entries that were made obsolete by the given coordinate entry and withdrawn from the PDB release set. One entry may replace many. It is PDB policy that only the principal investigator of a structure has the authority to withdraw it.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"SPRSDE"	
9-10	Continuation	continuation	Allows for multiple ID codes.
12-20	Date	sprsdeDate	Date this entry superseded the listed entries. This field is not copied on continuations.
22-25	IDcode	idCode	ID code of this entry. This field is not copied on continuations.
32-35	IDcode	sIdCode	ID code of a superseded entry.
37-40	IDcode	sIdCode	ID code of a superseded entry.
42-45	IDcode	sIdCode	ID code of a superseded entry.
47-50	IDcode	sIdCode	ID code of a superseded entry.
52-55	IDcode	sIdCode	ID code of a superseded entry.
57-60	IDcode	sIdCode	ID code of a superseded entry.
62-65	IDcode	sIdCode	ID code of a superseded entry.
67-70	IDcode	sIdCode	ID code of a superseded entry.
72-75	IDcode	sIdCode	ID code of a superseded entry.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.annotation.**Title**
TITLE field

The TITLE record contains a title for the experiment or analysis that is represented in the entry. It should identify an entry in the PDB in the same way that a title identifies a paper.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"TITLE "	
9-10	Continuation	continuation	Allows concatenation of multiple records.
11-80	String	title	Title of the experiment.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

primary module

Classes for PDB records that provide primary structure information.

class old_pdb.primary.**DatabaseReference**
DBREF record.

The DBREF record provides cross-reference links between PDB sequences (what appears in SEQRES record) and a corresponding database sequence.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"DBREF "	
8-11	IDcode	idCode	ID code of this entry.
13	Character	chainID	Chain identifier.
15-18	Integer	seqBegin	Initial sequence number of the PDB sequence segment.
19	AChar	insertBegin	Initial insertion code of the PDB sequence segment.
21-24	Integer	seqEnd	Ending sequence number of the PDB sequence segment.
25	AChar	insertEnd	Ending insertion code of the PDB sequence segment.
27-32	LString	database	Sequence database name.
34-41	LString	dbAccession	Sequence database accession code.
43-54	LString	dbIdCode	Sequence database identification code.
56-60	Integer	dbseqBegin	Initial sequence number of the database seqment.
61	AChar	idbnsBeg	Insertion code of initial residue the segment, if PDB is the reference.
63-67	Integer	dbseqEnd	Ending sequence number of the segment.
68	AChar	dbinsEnd	Insertion code of the ending of the segment, if PDB is the reference.

parse_line(*line*)
Parse DBREF line.

Parameters **line** (*str*) – line to parse

class old_pdb.primary.**DatabaseReference1**

Provides cross-reference links between PDB sequences (what appears in SEQRES record) and a corresponding database sequence.

This updated two-line format is used when the accession code or sequence numbering does not fit the space allotted in the standard DBREF format. This includes some GenBank sequence numbering (greater than 5 characters) and UNIMES accession numbers (greater than 12 characters).

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"DBREF1"	
8-11	IDcode	idCode	ID code of this entry.
13	Character	chainID	Chain identifier.
15-18	Integer	seqBegin	Initial sequence number of the PDB sequence segment, right justified.
19	AChar	insertBegin	Initial insertion code of the PDB sequence segment.
21-24	Integer	seqEnd	Ending sequence number of the PDB sequence segment, right justified.
25	AChar	insertEnd	Ending insertion code of the PDB sequence segment.
27-32	LString	database	Sequence database name.
48-67	LString	dbIdCode	Sequence database identification code, left justified.

parse_line(*line*)

Parse PDB-format line.

Parameters **line** (*str*) – line with PDB class

class old_pdb.primary.**DatabaseReference2**

Provides cross-reference links between PDB sequences (what appears in SEQRES record) and a corresponding database sequence.

This updated two-line format is used when the accession code or sequence numbering does not fit the space allotted in the standard DBREF format. This includes some GenBank sequence numbering (greater than 5 characters) and UNIMES accession numbers (greater than 12 characters).

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"DBREF2"	
8-11	IDcode	idCode	ID code of this entry.
13	Character	chainID	Chain identifier.
19-40	LString	dbAccession	Sequence database accession code, left justified.
46-55	Integer	seqBegin	Initial sequence number of the Database segment, right justified.
58-67	Integer	seqEnd	Ending sequence number of the Database segment, right justified.

parse_line(*line*)

Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.primary.**ModifiedResidue**
MODRES field

The MODRES record provides descriptions of modifications (e.g., chemical or post-translational) to protein and nucleic acid residues. Included are a mapping between residue names given in a PDB entry and standard residues.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"MODRES"	
8-11	IDcode	idCode	ID code of this entry.
13-15	Residue name	resName	Residue name used in this entry.
17	Character	chainID	Chain identifier.
19-22	Integer	seqNum	Sequence number.
23	AChar	iCode	Insertion code.
25-27	Residue name	stdRes	Standard residue name.
30-70	String	comment	Description of the residue modification.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.primary.**SequenceDifferences**
SEQADV field

The SEQADV record identifies conflicts between sequence information in the ATOM records of the PDB entry and the sequence database entry given on DBREF. Please note that these records were designed to identify differences and not errors. No assumption is made as to which database contains the correct data. PDB may include REMARK records in the entry that reflect the depositor's view of which database has the correct sequence.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"SEQADV"	
8-11	IDcode	idCode	ID code of this entry.
13-15	Residue name	resName	Name of the PDB residue in conflict.
17	Character	chainID	PDB chain identifier.
19-22	Integer	seqNum	PDB sequence number.
23	AChar	iCode	PDB insertion code.
25-28	LString	database	
30-38	LString	dbAccession	Sequence database accession number.
40-42	Residue name	dbRes	Sequence database residue name.
44-48	Integer	dbSeq	Sequence database sequence number.
50-70	LString	conflict	Conflict comment.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.primary.**SequenceResidues**
SEQRES field

SEQRES records contain the amino acid or nucleic acid sequence of residues in each chain of the macromolecule that was studied.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"SEQRES"	
8-10	Integer	serNum	Serial number of the SEQRES record for the current chain. Starts at 1 and increments by one each line. Reset to 1 for each chain.
12	Character	chainID	Chain identifier. This may be any single legal character, including a blank which is is used if there is only one chain.
14-17	Integer	numRes	Number of residues in the chain. This value is repeated on every record.
20-22	Residue name	resName	Residue name.
24-26	Residue name	resName	Residue name.
28-30	Residue name	resName	Residue name.
32-34	Residue name	resName	Residue name.
36-38	Residue name	resName	Residue name.
40-42	Residue name	resName	Residue name.
44-46	Residue name	resName	Residue name.
48-50	Residue name	resName	Residue name.
52-54	Residue name	resName	Residue name.
56-58	Residue name	resName	Residue name.
60-62	Residue name	resName	Residue name.
64-66	Residue name	resName	Residue name.
68-70	Residue name	resName	Residue name.

num_chains() → int
Number of chains in sequence.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

heterogen module

Classes for PDB records that provide heterogen information.

class old_pdb.heterogen.**Formula**
FORMUL field

The FORMUL record presents the chemical formula and charge of a non-standard group.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"FORMUL"	
9-10	Integer	compNum	Component number.
13-15	LString(3)	hetID	Het identifier.
17-18	Integer	continuation	Continuation number.
19	Character	asterisk	"*" for water.
20-70	String	text	Chemical formula.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.heterogen.**Heterogen**
HET field

HET records are used to describe non-standard residues, such as prosthetic groups, inhibitors, solvent molecules, and ions for which coordinates are supplied. Groups are considered HET if they are:

- not one of the standard amino acids, and
- not one of the nucleic acids (C, G, A, T, U, and I), and
- not one of the modified versions of nucleic acids (+C, +G, +A, +T, +U, and +I), and
- not an unknown amino acid or nucleic acid where UNK is used to indicate the unknown residue name.

Het records also describe heterogens for which the chemical identity is unknown, in which case the group is assigned the hetatm_id UNK.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"HET "	
8-10	LString(3)	hetID	Het identifier, right-justified.
13	Character	ChainID	Chain identifier.
14-17	Integer	seqNum	Sequence number.
18	AChar	iCode	Insertion code.
21-25	Integer	numHetAtoms	Number of HETATM records for the group present in the entry.
31-70	String	text	Text describing Het group.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.heterogen.**HeterogenName**
HETNAM field

This record gives the chemical name of the compound with the given hetatm_id.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"HETNAM"	
9-10	Continuation	continuation	Allows concatenation of multiple records.
12-14	LString(3)	hetID	Het identifier, right- justified.
16-70	String	text	Chemical name.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.heterogen.**HeterogenSynonym**
HETSYN field

This record provides synonyms, if any, for the compound in the corresponding (i.e., same hetatm_id) HETNAM record. This is to allow greater flexibility in searching for HET groups.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"HETSYN"	
9-10	Continuation	continuation	Allows concatenation of multiple records.
12-14	LString(3)	hetID	Het identifier, right- justified.
16-70	SList	hetSynonyms	List of synonyms.

parse_line(*line*)
Parse line of PDB file.

Parameters **line** (*str*) – PDB file line to parse

secondary module

Classes for records with secondary structure and connectivity information.

class old_pdb.secondary.**CisPeptide**
CISPEP field

CISPEP records specify the prolines and other peptides found to be in the cis conformation. This record replaces the use of footnote records to list cis peptides.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"CISPEP"	
8-10	Integer	serNum	Record serial number.
12-14	LString(3)	pep1	Residue name.
16	Character	chainID1	Chain identifier.
18-21	Integer	seqNum1	Residue sequence number.
22	AChar	icode1	Insertion code.
26-28	LString(3)	pep2	Residue name.
30	Character	chainID2	Chain identifier.
32-35	Integer	seqNum2	Residue sequence number.
36	AChar	icode2	Insertion code.
44-46	Integer	modNum	Identifies the specific model.
54-59	Real(6.2)	measure	Angle measurement in degrees.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.secondary.**DisulfideBond**
SSBOND field

The SSBOND record identifies each disulfide bond in protein and polypeptide structures by identifying the two residues involved in the bond.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"SSBOND"	
8-10	Integer	serNum	Serial number.
12-14	LString(3)	"CYS"	Residue name.
16	Character	chainID1	Chain identifier.
18-21	Integer	seqNum1	Residue sequence number.
22	AChar	icode1	Insertion code.
26-28	LString(3)	"CYS"	Residue name.
30	Character	chainID2	Chain identifier.
32-35	Integer	seqNum2	Residue sequence number.
36	AChar	icode2	Insertion code.
60-65	SymOP	sym1	Symmetry operator for residue 1.
67-72	SymOP	sym2	Symmetry operator for residue 2.
74-78	Real(5.2)	Length	Disulfide bond distance

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.secondary.**Helix**
HELIX field

HELIX records are used to identify the position of helices in the molecule. Helices are both named and numbered. The residues where the helix begins and ends are noted, as well as the total length.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"HELIX "	
8-10	Integer	serNum	Serial number of the helix. This starts at 1 and increases incrementally.
12-14	LString(3)	helixID	Helix identifier. In addition to a serial number, each helix is given an alphanumeric character helix identifier.
16-18	Residue name	initResName	Name of the initial residue.
20	Character	initChainID	Chain identifier for the chain containing this helix.
22-25	Integer	initSeqNum	Sequence number of the initial residue.
26	AChar	initICode	Insertion code of the initial residue.
28-30	Residue name	endResName	Name of the terminal residue of the helix.
32	Character	endChainID	Chain identifier for the chain containing this helix.
34-37	Integer	endSeqNum	Sequence number of the terminal residue.
38	AChar	endICode	Insertion code of the terminal residue.
39-40	Integer	helixClass	Helix class (see below).
41-70	String	comment	Comment about this helix.
72-76	Integer	length	Length of this helix.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.secondary.**Link**
LINK field

The LINK records specify connectivity between residues that is not implied by the primary structure. Connectivity is expressed in terms of the atom names. This record supplements information given in CONECT records and is provided here for convenience in searching.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	“LINK “	
13-16	Atom	name1	Atom name.
17	Character	altLoc1	Alternate location indicator.
18-20	Residue name	resName1	Residue name.
22	Character	chainID1	Chain identifier.
23-26	Integer	resSeq1	Residue sequence number.
27	AChar	iCode1	Insertion code.
43-46	Atom	name2	Atom name.
47	Character	altLoc2	Alternate location indicator.
48-50	Residue name	resName2	Residue name.
52	Character	chainID2	Chain identifier.
53-56	Integer	resSeq2	Residue sequence number.
57	AChar	iCode2	Insertion code.
60-65	SymOP	sym1	Symmetry operator atom 1.
67-72	SymOP	sym2	Symmetry operator atom 2.
74-78	Real(5.2)	Length	Link distance

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.secondary.**Sheet**
SHEET field

SHEET records are used to identify the position of sheets in the molecule. Sheets are both named and numbered. The residues where the sheet begins and ends are noted.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	“SHEET “	
8-10	Integer	strand	Strand number which starts at 1 for each strand within a sheet and increases by one.
12-14	LString(3)	sheetID	Sheet identifier.
15-16	Integer	numStrands	Number of strands in sheet.
18-20	Residue name	initResName	Residue name of initial residue.
22	Character	initChainID	Chain identifier of initial residue in strand.
23-26	Integer	initSeqNum	Sequence number of initial residue in strand.
27	AChar	initICode	Insertion code of initial residue in strand.
29-31	Residue name	endResName	Residue name of terminal residue
33	Character	endChainID	Chain identifier of terminal residue
34-37	Integer	endSeqNum	Sequence number of terminal residue.
38	AChar	endICode	Insertion code of terminal residue.
39-40	Integer	sense	Sense of strand with respect to previous strand in the sheet. 0 if first strand, 1 if parallel, and -1 if anti-parallel.
42-45	Atom	curAtom	Registration. Atom name in curent strand.
46-48	Residue name	curResName	Registration. Residue name in current strand
50	Character	curChainId	Registration. Chain identifier in current strand.
51-54	Integer	curResSeq	Registration. Residue sequence number in current strand.
55	AChar	curICode	Registration. Insertion code in current strand.
57-60	Atom	prevAtom	Registration. Atom name in previous strand.
61-63	Residue name	prevResName	Registration. Residue name in previous strand.
65	Character	prevChainId	Registration. Chain identifier in previous strand.
66-69	Integer	prevResSeq	Registration. Residue sequence number in previous strand.
70	AChar	prevICode	Registration. Insertion code in previous strand.

parse_line(*line*)

Parse PDB-format line.

Parameters **line** (*str*) – line to parse

crystallography module

Classes for records with crystallographic information.

class old_pdb.crystallography.**FractionalTransform**(*n*)
SCALEn baseclass

The SCALEn (n = 1, 2, or 3) records present the transformation from the orthogonal coordinates as contained in the entry to fractional crystallographic coordinates. Non-standard coordinate systems should be explained in the remarks.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"SCALEn" n=1, 2, or 3	
11 - 20	Real(10.6)	s[n][1]	Sn1
21 - 30	Real(10.6)	s[n][2]	Sn2
31 - 40	Real(10.6)	s[n][3]	Sn3
46 - 55	Real(10.5)	u[n]	Un

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.crystallography.**NoncrystalTransform**(*n*)
MTRIXn baseclass

The MTRIXn (n = 1, 2, or 3) records present transformations expressing non-crystallographic symmetry.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"MTRIXn"	n=1, 2, or 3
8-10	Integer	serial	Serial number.
11-20	Real(10.6)	m[n][1]	Mn1
21-30	Real(10.6)	m[n][2]	Mn2
31-40	Real(10.6)	m[n][3]	Mn3
46-55	Real(10.5)	v[n]	Vn
60	Integer	iGiven	1 if coordinates for the representations which are approximately related by the transformations of the molecule are contained in the entry. Otherwise, blank.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.crystallography.**OriginalTransform**(*n*)
ORIGXn class

The ORIGXn (n = 1, 2, or 3) records present the transformation from the orthogonal coordinates contained in the entry to the submitted coordinates.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"ORIGXn"	n=1, 2, or 3
11-20	Real(10.6)	o[n][1]	On1
21-30	Real(10.6)	o[n][2]	On2
31-40	Real(10.6)	o[n][3]	On3
46-55	Real(10.5)	t[n]	Tn

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.crystallography.**UnitCell**
CRYST1 class

The CRYST1 record presents the unit cell parameters, space group, and Z value. If the structure was not determined by crystallographic means, CRYST1 simply defines a unit cube.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"CRYST1"	
7-15	Real(9.3)	a	a (Angstroms).
16-24	Real(9.3)	b	b (Angstroms).
25-33	Real(9.3)	c	c (Angstroms).
34-40	Real(7.2)	alpha	alpha (degrees).
41-47	Real(7.2)	beta	beta (degrees).
48-54	Real(7.2)	gamma	gamma (degrees).
56-66	LString	sGroup	Space group.
67-70	Integer	z	Z value.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

coordinates module

Classes for records with coordinate information.

class old_pdb.coordinates.**Atom**
ATOM class

The ATOM records present the atomic coordinates for standard residues. They also present the occupancy and temperature factor for each atom. Heterogen coordinates use the HETATM record type. The element symbol is always present on each ATOM record; segment identifier and charge are optional.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	“ATOM “	
7-11	Integer	serial	Atom serial number.
13-16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18-20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23-26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31-38	Real(8.3)	x	Orthogonal coordinates for X in Angstroms.
39-46	Real(8.3)	y	Orthogonal coordinates for Y in Angstroms.
47-54	Real(8.3)	z	Orthogonal coordinates for Z in Angstroms.
55-60	Real(6.2)	occupancy	Occupancy.
61-66	Real(6.2)	tempFactor	Temperature factor.
77-78	LString(2)	element	Element symbol, right-justified.
79-80	LString(2)	charge	Charge on the atom.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.coordinates.**ChainTerminus**
TER class

The TER record indicates the end of a list of ATOM/HETATM records for a chain.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	“TER “	
7-11	Integer	serial	Serial number.
18-20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23-26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Insertion code.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.coordinates.**HeterogenAtom**
HETATM class

The HETATM records present the atomic coordinate records for atoms within “non-standard” groups. These records are used for water molecules and atoms presented in HET groups.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"HETATM"	
7-11	Integer	serial	Atom serial number.
13-16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18-20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23-26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31-38	Real(8.3)	x	Orthogonal coordinates for X.
39-46	Real(8.3)	y	Orthogonal coordinates for Y.
47-54	Real(8.3)	z	Orthogonal coordinates for Z.
55-60	Real(6.2)	occupancy	Occupancy.
61-66	Real(6.2)	tempFactor	Temperature factor.
77-78	LString(2)	element	Element symbol; right-justified.
79-80	LString(2)	charge	Charge on the atom.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.coordinates.**Model**
MODEL class.

The MODEL record specifies the model serial number when multiple structures are presented in a single coordinate entry, as is often the case with structures determined by NMR.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"MODEL "	
11-14	Integer	serial	Model serial number.

property all_atoms
Get all atoms in model.

Returns list of *Atom*-like objects

property atoms
Get ATOM atoms in model.

Returns list of *Atom*-like objects

property het_atoms
Get HETATM atoms in model.

Returns list of *Atom*-like objects

num_atoms(*heavy_only*) → int
Number of ATOM and HETATM entries in all chains in model.

Parameters **heavy_only** (*bool*) – exclude hydrogen atoms from count

num_chains() → int
Count number of chains in model.

num_residues(*count_hetatm*) → int
Number of residues in entry.

Parameters **count_hetatm** (*bool*) – include heterogen residues in count

num_ter() → int
Count number of termini in entry.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.coordinates.**TemperatureFactor**
ANISOU class

The ANISOU records present the anisotropic temperature factors.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"ANISOU"	
7-11	Integer	serial	Atom serial number.
13-16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator
18-20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23-26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Insertion code.
29-35	Integer	u[0][0]	U(1,1)
36-42	Integer	u[1][1]	U(2,2)
43-49	Integer	u[2][2]	U(3,3)
50-56	Integer	u[0][1]	U(1,2)
57-63	Integer	u[0][2]	U(1,3)
64-70	Integer	u[1][2]	U(2,3)
77-78	LString(2)	element	Element symbol, right-justified.
79-80	LString(2)	charge	Charge on the atom.

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

bookkeeping module

Classes for records with connectivity and bookkeeping information.

class old_pdb.bookkeeping.**Connection**
CONECT class

The CONECT records specify connectivity between atoms for which coordinates are supplied. The connectivity is described using the atom serial number as found in the entry. CONECT records are mandatory for HET groups (excluding water) and for other bonds not specified in the standard residue connectivity table which involve atoms in standard residues (see Appendix 4 for the list of standard residues). These records are generated by the PDB.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"CONECT"	
7-11	Integer	serial	Atom serial number
12-16	Integer	serial	Serial number of bonded atom
17-21	Integer	serial	Serial number of bonded atom
22-26	Integer	serial	Serial number of bonded atom
27-31	Integer	serial	Serial number of bonded atom

parse_line(*line*)
Parse PDB-format line.

Parameters **line** (*str*) – line to parse

class old_pdb.bookkeeping.**Master**
MASTER class

The MASTER record is a control record for bookkeeping. It lists the number of lines in the coordinate entry or file for selected record types.

COLUMNS	DATA TYPE	FIELD	DEFINITION
1-6	Record name	"MASTER"	
11-15	Integer	numRemark	Number of REMARK records
16-20	Integer	"0"	
21-25	Integer	numHet	Number of HET records
26-30	Integer	numHelix	Number of HELIX records
31-35	Integer	numSheet	Number of SHEET records
36-40	Integer	numTurn	deprecated
41-45	Integer	numSite	Number of SITE records
46-50	Integer	numXform	Number of coordinate transformation records (ORIGX+SCALE+MTRIX)
51-55	Integer	numCoord	Number of atomic coordinate records records (ATOM+HETATM)
56-60	Integer	numTer	Number of TER records
61-65	Integer	numConect	Number of CONECT records
66-70	Integer	numSeq	Number of SEQRES records

parse_line(*line*)
Parse a PDB-format line

Parameters **line** (*str*) – line to parse

5.2 Change log

5.2.1 Current version

- Fixed readthedocs.io rendering of documentation (#3)

5.2.2 0.0.1

Initial release.

PYTHON MODULE INDEX

O

INDEX

O

P

R

S

T

U