W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Jun 26, 2025 – 10:05 AM EDT

PDB ID	:	5J2Q / pdb_00005j2q
Title	:	HIV-1 reverse transcriptase in complex with DNA that has incorporated a mismatched EFdA-MP at the N-(pre-translocation) site
Authors	:	Salie, Z.L.; Kirby, K.A.; Sarafianos, S.G.
Deposited on	:	2016-03-29
Resolution	:	2.79 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the ⓘ symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references ⓘ) were used in the production of this report:

|		
MolProbity	:	4-5-2 with Phenix2.0rc1
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0rc1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.006 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.44

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.79 Å.

Percentile scores (ranging between 0-100) for global validation metrics of the entry are shown in the following graphic. The table shows the number of entries on which the scores are based.

Metric	Percentile Ranks	Value
Rfree		0.253
Clashscore		3
Ramachandran outliers		0
Sidechain outliers		0.4%
RSRZ outliers		5.0%

Worse   *Better*

▮ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution
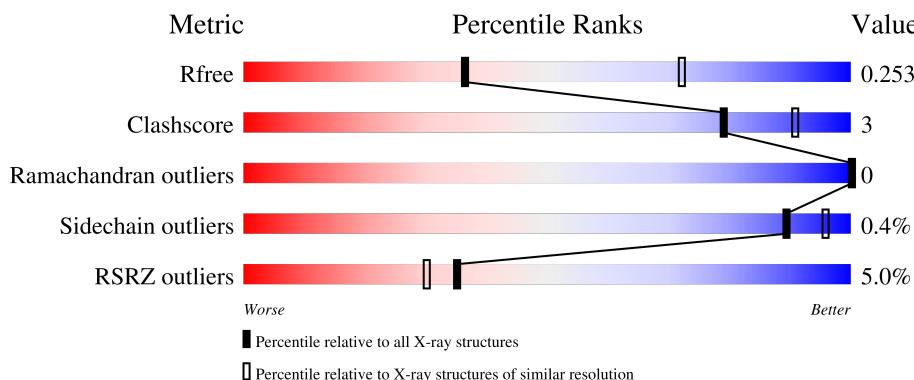
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	4924 (2.80-2.76)
Clashscore	180529	5458 (2.80-2.76)
Ramachandran outliers	177936	5386 (2.80-2.76)
Sidechain outliers	177891	5388 (2.80-2.76)
RSRZ outliers	164620	4926 (2.80-2.76)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments of the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria respectively. A grey segment represents the fraction of residues that are not modelled. The numeric value for each fraction is indicated below the corresponding segment, with a dot representing fractions <=5% The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density. The numeric value is given above the bar.

Mol	Chain	Length	Quality of chain
1	A	560	6% / 90% / 8% / ·
2	B	440	3% / 84% / 7% / 9%
3	P	22	64% / 23% / 14%
4	T	27	4% / 59% / 19% / 22%
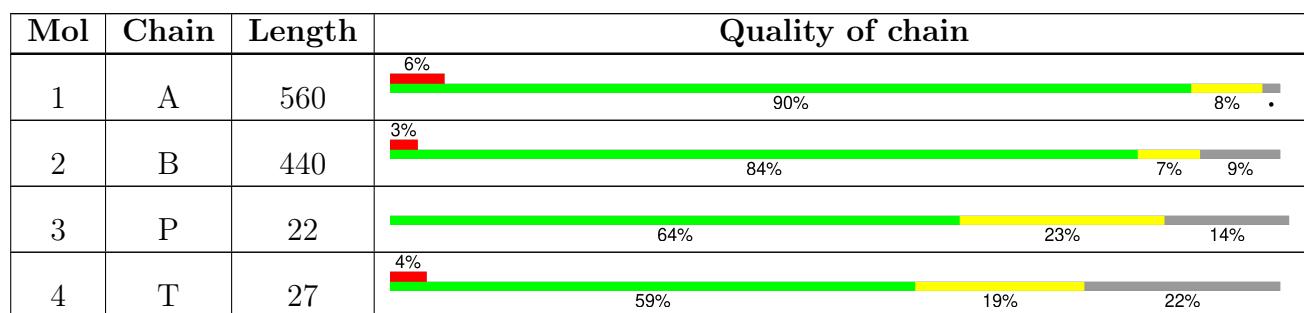
2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 8595 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called HIV-1 reverse transcriptase p66 subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	547	Total 4462	C 2885	N 747	O 822	S 8	48	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	258	CYS	GLN	engineered mutation	UNP P04585
A	280	SER	CYS	engineered mutation	UNP P04585

- Molecule 2 is a protein called HIV-1 reverse transcriptase p51 domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	400	Total 3297	C 2140	N 551	O 600	S 6	7	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	280	SER	CYS	engineered mutation	UNP P04585

- Molecule 3 is a DNA chain called DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP* TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(6FM)P*(6FM))-3').

Mol	Chain	Residues	Atoms							ZeroOcc	AltConf	Trace
3	P	19	Total 397	C 190	F 2	N 69	O 116	P 19	S 1	0	0	0

- Molecule 4 is a DNA chain called DNA (27-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	T	21	Total 435	C 204	N 87	O 123	P 21	0	0	0

- Molecule 5 is MAGNESIUM ION (CCD ID: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Mg 1 1	0	0

- Molecule 6 is water.

| Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	2	Total O 2 2	0	0
6	B	1	Total O 1 1	0	0

3 Residue-property plots ⓘ

These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density (RSRZ > 2). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

- Molecule 1: HIV-1 reverse transcriptase p66 subunit
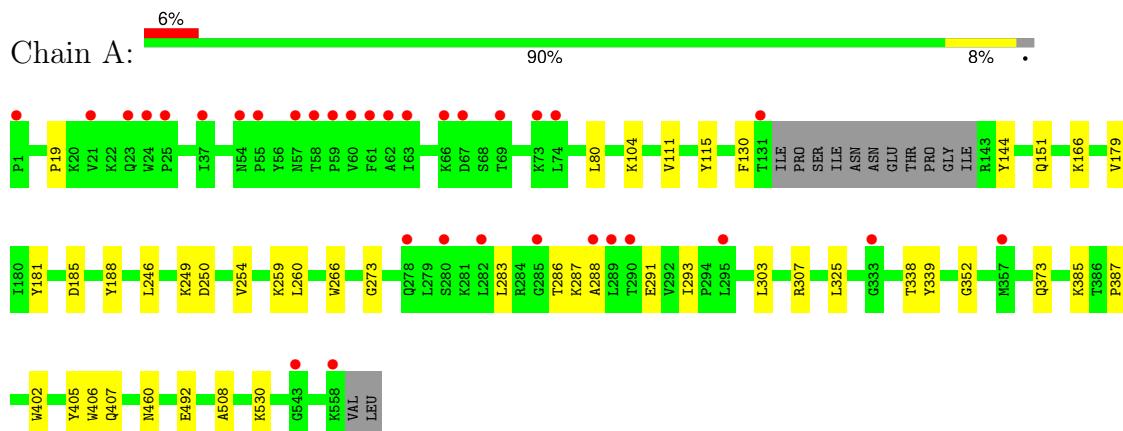
- Molecule 2: HIV-1 reverse transcriptase p51 domain
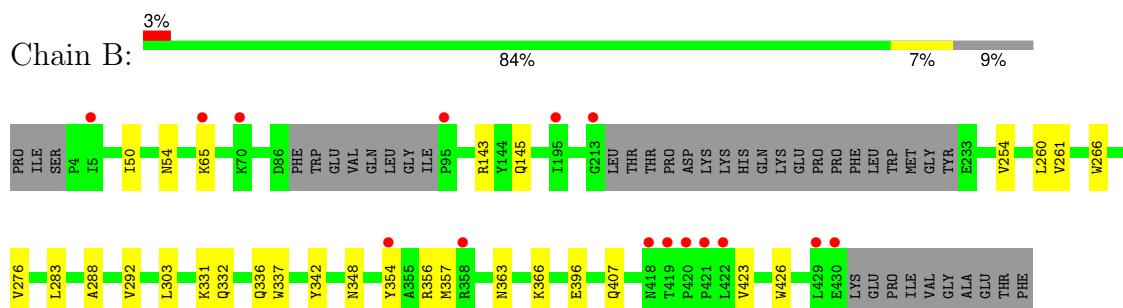
- Molecule 3: DNA (5'-D(*AP*CP*AP*GP*TP*CP*CP*CP*TP*GP*TP*TP*CP*GP*GP*(MRG)P*CP*GP*CP*CP*(6FM)P*(6FM))-3')
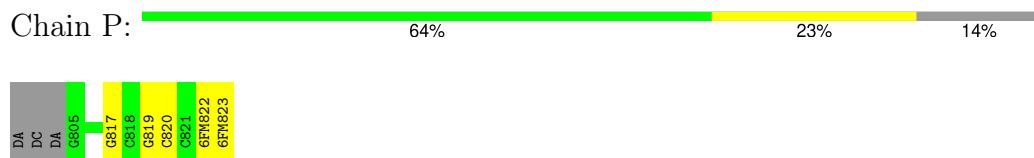
- Molecule 4: DNA (27-MER)

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, α, β, γ	165.56Å 170.51Å 102.23Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	64.33 – 2.79 64.33 – 2.79	Depositor EDS
% Data completeness (in resolution range)	99.7 (64.33-2.79) 99.7 (64.33-2.79)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.12 (at 2.77Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.212 , 0.252 0.214 , 0.253	Depositor DCC
R_{free} test set	1818 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	62.2	Xtriage
Anisotropy	0.599	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 32.9	EDS
L-test for twinning[2]	$< |L| > = 0.50, < L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.016 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	8595	wwPDB-VP
Average B, all atoms (Å^2)	71.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.93% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

5 Model quality ⓘ

5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: MRG, MG, 6FM

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$|Z|$>5	RMSZ	#$|Z|$>5
1	A	0.10	0/4576	0.27	0/6211
2	B	0.10	0/3387	0.25	0/4598
3	P	0.16	0/358	0.34	0/547
4	T	0.15	0/489	0.30	0/753
All	All	0.11	0/8810	0.27	0/12109

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts ⓘ

In the following table, the Non-H and H(model) columns list the number of non-hydrogen atoms and hydrogen atoms in the chain respectively. The H(added) column lists the number of hydrogen atoms added and optimized by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, whereas Symm-Clashes lists symmetry-related clashes.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4462	0	4518	25	0
2	B	3297	0	3329	17	0
3	P	397	0	196	2	0
4	T	435	0	234	4	0
5	A	1	0	0	0	0
6	A	2	0	0	0	0
6	B	1	0	0	0	0
All	All	8595	0	8277	43	0

The all-atom clashscore is defined as the number of clashes found per 1000 atoms (including

hydrogen atoms). The all-atom clashscore for this structure is 3.

The worst 5 of 43 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:492:GLU:HG2	1:A:530:LYS:HB2	1.78	0.65
1:A:115:TYR:HD2	1:A:151:GLN:HG2	1.64	0.61
1:A:288:ALA:HB3	1:A:291:GLU:HB2	1.81	0.61
2:B:363:ASN:HB3	2:B:366:LYS:HB3	1.87	0.56
1:A:286:THR:HG21	1:A:293:ILE:HD11	1.90	0.54

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	543/560 (97%)	523 (96%)	20 (4%)	0	100	100
2	B	394/440 (90%)	377 (96%)	17 (4%)	0	100	100
All	All	937/1000 (94%)	900 (96%)	37 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains ⓘ

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	488/500 (98%)	486 (100%)	2 (0%)	89	96
2	B	362/400 (90%)	361 (100%)	1 (0%)	91	97
All	All	850/900 (94%)	847 (100%)	3 (0%)	89	96

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	179	VAL
1	A	287	LYS
2	B	292	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	418	ASN
2	B	278	GLN
2	B	137	ASN
1	A	539	HIS
2	B	174	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

3 non-standard protein/DNA/RNA residues are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
3	6FM	P	822	3	18,26,27	2.34	8 (44%)	12,39,42	3.43	7 (58%)
3	6FM	P	823	3	18,26,27	2.32	8 (44%)	12,39,42	3.38	6 (50%)

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	MRG	P	817	1,3,4	21,28,29	1.84	4 (19%)	19,39,42	1.72	4 (21%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	6FM	P	822	3	-	3/4/37/38	0/3/3/3
3	6FM	P	823	3	-	0/4/37/38	0/3/3/3
3	MRG	P	817	1,3,4	-	3/8/26/27	0/3/3/3

The worst 5 of 20 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	P	817	MRG	O6-C6	5.28	1.35	1.23
3	P	823	6FM	O4'-C1'	5.04	1.50	1.42
3	P	822	6FM	O4'-C1'	4.96	1.50	1.42
3	P	822	6FM	C2'-C3'	-3.46	1.45	1.53
3	P	823	6FM	C2'-C3'	-3.37	1.45	1.53

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	P	822	6FM	N1-C2-N3	-6.70	120.92	130.60
3	P	823	6FM	N1-C2-N3	-6.60	121.05	130.60
3	P	822	6FM	F32-C2-N3	5.09	119.58	114.69
3	P	823	6FM	F32-C2-N3	5.07	119.56	114.69
3	P	822	6FM	F32-C2-N1	5.02	119.51	114.69

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	P	817	MRG	N3-C2-N2-C21
3	P	817	MRG	N1-C2-N2-C21
3	P	822	6FM	C3'-C4'-C5'-O5'
3	P	822	6FM	C33-C4'-C5'-O5'
3	P	822	6FM	O4'-C4'-C5'-O5'

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

There are no chain breaks in this entry.

6 Fit of model and data ⓘ

6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95th percentile and maximum values of the occupancy-weighted average B-factor per residue. The column labelled 'Q< 0.9' lists the number of (and percentage) of residues with an average occupancy less than 0.9.

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2			OWAB(Å²)	Q<0.9
1	A	547/560 (97%)	0.60	33 (6%)	29	24	2, 58, 122, 161	40 (7%)
2	B	400/440 (90%)	0.16	15 (3%)	44	39	37, 63, 124, 154	3 (0%)
3	P	16/22 (72%)	-0.28	0	100	100	62, 76, 97, 102	0
4	T	21/27 (77%)	-0.14	1 (4%)	36	31	53, 77, 96, 123	0
All	All	984/1049 (93%)	0.39	49 (4%)	35	30	2, 61, 123, 161	43 (4%)

The worst 5 of 49 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	61	PHE	47.1
1	A	60	VAL	43.4
1	A	59	PRO	34.6
1	A	62	ALA	31.0
1	A	24	TRP	18.8

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å²)	Q<0.9
3	MRG	P	817	26/27	0.86	0.15	81,90,108,114	0
3	6FM	P	823	24/25	0.89	0.13	75,85,99,104	0
3	6FM	P	822	24/25	0.94	0.09	64,75,81,83	0

6.3 Carbohydrates ⓘ

There are no oligosaccharides in this entry.

6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
5	MG	A	601	1/1	0.89	0.07	111,111,111,111	0

6.5 Other polymers ⓘ

There are no such residues in this entry.