Full wwPDB X-ray Structure Validation Report ⓘ

May 21, 2020 – 09:30 pm BST

PDB ID	:	4AEX
Title	:	HCV-JFH1 NS5B POLYMERASE STRUCTURE AT 2.4 ANGSTROM in a primitive orthorhombic space group
Authors	:	Caillet-Saguy, C.; Bressanelli, S.
Deposited on	:	2012-01-12
Resolution	:	2.41 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 following versions of software and data (see references ⓘ) were used in the production of this report:

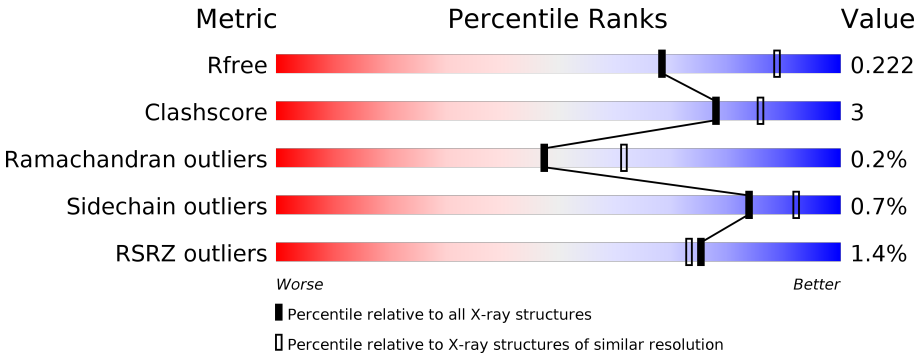
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

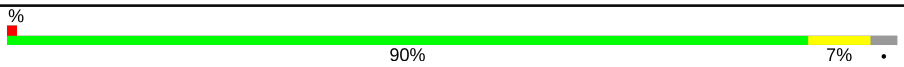
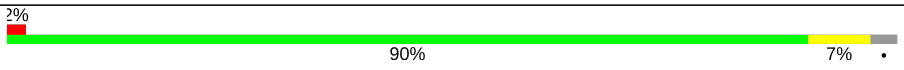
The reported resolution of this entry is 2.41 Å.

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	4647 (2.44-2.40)
Clashscore	141614	5161 (2.44-2.40)
Ramachandran outliers	138981	5073 (2.44-2.40)
Sidechain outliers	138945	5074 (2.44-2.40)
RSRZ outliers	127900	4543 (2.44-2.40)

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 on 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	579	
1	B	579	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9421 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.

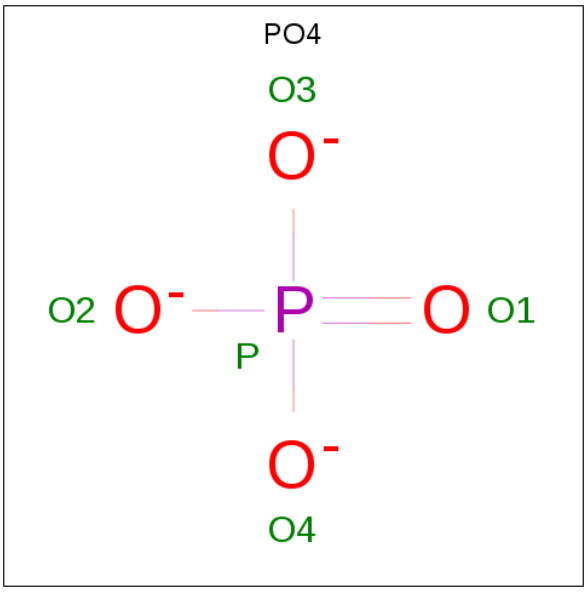
- Molecule 1 is a protein called RNA-DIRECTED RNA POLYMERASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	563	Total	C	N	O	S	0	14	0
			4489	2841	793	826	29			
1	B	564	Total	C	N	O	S	0	13	0
			4485	2839	791	826	29			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	expression tag	UNP Q99IB8
A	572	HIS	-	expression tag	UNP Q99IB8
A	573	HIS	-	expression tag	UNP Q99IB8
A	574	HIS	-	expression tag	UNP Q99IB8
A	575	HIS	-	expression tag	UNP Q99IB8
A	576	HIS	-	expression tag	UNP Q99IB8
A	577	HIS	-	expression tag	UNP Q99IB8
B	-1	MET	-	expression tag	UNP Q99IB8
B	572	HIS	-	expression tag	UNP Q99IB8
B	573	HIS	-	expression tag	UNP Q99IB8
B	574	HIS	-	expression tag	UNP Q99IB8
B	575	HIS	-	expression tag	UNP Q99IB8
B	576	HIS	-	expression tag	UNP Q99IB8
B	577	HIS	-	expression tag	UNP Q99IB8

- Molecule 2 is PHOSPHATE ION (three-letter code: PO4) (formula: O_4P).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total O P 5 4 1	0	0

- Molecule 3 is water.

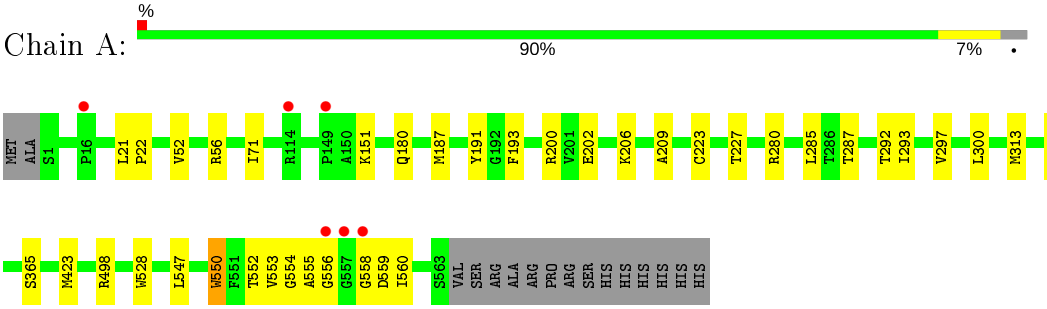
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	215	Total O 215 215	0	0
3	B	227	Total O 227 227	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA 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.

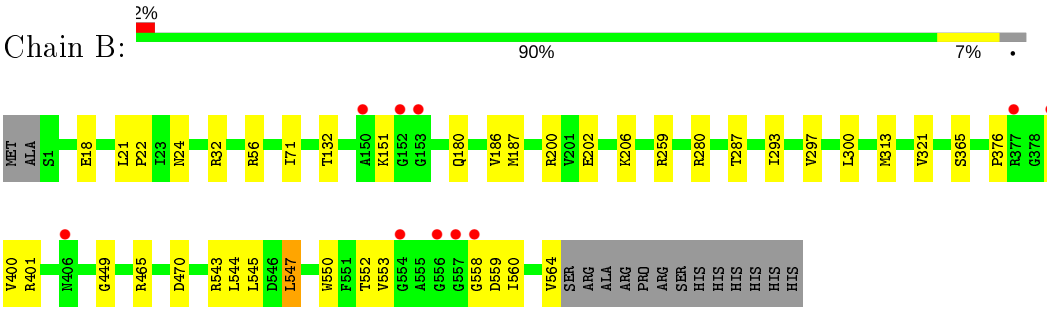
- Molecule 1: RNA-DIRECTED RNA POLYMERASE

Chain A:



- Molecule 1: RNA-DIRECTED RNA POLYMERASE

Chain B:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	98.10Å 114.70Å 115.80Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.75 – 2.41 42.02 – 2.29	Depositor EDS
% Data completeness (in resolution range)	99.9 (40.75-2.41) 99.6 (42.02-2.29)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.47 (at 2.29Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.187 , 0.230 0.183 , 0.222	Depositor DCC
R_{free} test set	1992 reflections (3.35%)	wwPDB-VP
Wilson B-factor (Å^2)	27.2	Xtriage
Anisotropy	0.518	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 42.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.000 for -h,l,k	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	9421	wwPDB-VP
Average B, all atoms (Å^2)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 22.80 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 5.2920e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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: PO4

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	0.24	0/4606	0.48	0/6252
1	B	0.24	0/4599	0.48	0/6244
All	All	0.24	0/9205	0.48	0/12496

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand.A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	1
1	B	0	1
All	All	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	151	LYS	Peptide
1	B	151	LYS	Peptide

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	4489	0	4516	25	0
1	B	4485	0	4511	30	0
2	A	5	0	0	0	0
3	A	215	0	0	1	0
3	B	227	0	0	2	0
All	All	9421	0	9027	53	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.

All (53) 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:B:200[B]:ARG:NH2	1:B:365:SER:OG	2.10	0.84
1:B:545:LEU:HG	1:B:547:LEU:HD11	1.61	0.82
1:A:200[B]:ARG:HH21	1:A:365:SER:HG	1.27	0.80
1:B:186:VAL:HG23	1:B:187:MET:HG3	1.67	0.76
1:B:381:ARG:NH1	1:B:470:ASP:OD2	2.20	0.74
1:B:465:ARG:HG2	1:B:547:LEU:HD13	1.72	0.72
1:A:558[B]:GLY:O	1:A:560:ILE:N	2.23	0.71
1:B:200[B]:ARG:HH21	1:B:365:SER:HG	1.37	0.69
1:A:200[B]:ARG:NH2	1:A:365:SER:OG	2.19	0.65
1:B:321:VAL:HG21	1:B:365:SER:HB2	1.80	0.64
1:B:558[B]:GLY:O	1:B:560:ILE:N	2.31	0.63
1:B:449:GLY:O	1:B:558[B]:GLY:N	2.31	0.63
1:A:321:VAL:HG21	1:A:365:SER:HB2	1.81	0.61
1:B:18:GLU:OE1	1:B:401:ARG:NH1	2.36	0.59
1:A:21:LEU:HD12	1:A:22:PRO:HD2	1.86	0.58
1:A:300:LEU:HB2	1:A:313:MET:HE1	1.88	0.56
1:A:187:MET:HE1	1:A:292:THR:HG22	1.90	0.54
1:A:280:ARG:HD2	1:A:287:THR:HA	1.89	0.54
1:B:545:LEU:HG	1:B:547:LEU:CD1	2.37	0.53
1:B:24:ASN:HB3	1:B:400:VAL:HG11	1.91	0.53
1:B:379:ARG:HD3	1:B:380:ARG:H	1.74	0.53
1:A:547:LEU:HD22	1:A:550[A]:TRP:CD2	2.44	0.53
1:B:300:LEU:HB2	1:B:313:MET:HE1	1.93	0.51
1:B:21:LEU:HD12	1:B:22:PRO:HD2	1.93	0.50
1:A:552[A]:THR:OG1	1:A:553[A]:VAL:N	2.45	0.50
1:B:293:ILE:O	1:B:297:VAL:HG23	2.13	0.49
1:B:379:ARG:HD3	1:B:380:ARG:N	2.27	0.49
1:A:313:MET:HG2	1:A:322:VAL:HG22	1.94	0.49
1:B:280:ARG:HD2	1:B:287:THR:HA	1.95	0.48

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:498:ARG:NE	3:A:2200:HOH:O	2.38	0.48
1:B:449:GLY:O	1:B:558[A]:GLY:HA2	2.15	0.47
1:A:293:ILE:O	1:A:297:VAL:HG23	2.15	0.46
1:A:202:GLU:HG2	1:A:206:LYS:HE3	1.98	0.45
1:B:132:THR:O	1:B:259:ARG:NE	2.36	0.45
1:B:32:ARG:NH1	3:B:2023:HOH:O	2.40	0.45
1:B:71:ILE:HD13	1:B:297:VAL:HG22	1.99	0.44
1:B:547:LEU:HD12	1:B:547:LEU:N	2.34	0.43
1:B:376:PRO:O	3:B:2172:HOH:O	2.21	0.43
1:B:202:GLU:HG2	1:B:206:LYS:HE3	2.00	0.42
1:A:209:ALA:O	1:B:543:ARG:NH1	2.52	0.42
1:A:180:GLN:OE1	1:A:553[B]:VAL:HA	2.20	0.42
1:B:200[B]:ARG:NH2	1:B:365:SER:HG	2.04	0.42
1:A:193:PHE:CG	1:A:556[B]:GLY:HA2	2.55	0.42
1:A:554[A]:GLY:HA2	1:A:560:ILE:HG13	2.02	0.42
1:B:180:GLN:OE1	1:B:553[B]:VAL:HA	2.19	0.41
1:A:191:TYR:OH	1:A:553[A]:VAL:HB	2.20	0.41
1:A:285:LEU:HD12	1:A:555[A]:ALA:HB2	2.03	0.41
1:A:206:LYS:HG2	1:B:544:LEU:HD13	2.02	0.41
1:B:71:ILE:CD1	1:B:297:VAL:HG22	2.50	0.41
1:A:71:ILE:HD13	1:A:297:VAL:HG22	2.03	0.41
1:A:227:THR:HB	1:A:347:SER:O	2.21	0.41
1:A:52:VAL:HG12	1:A:223[B]:CYS:SG	2.61	0.41
1:A:423:MET:HA	1:A:528:TRP:CZ2	2.57	0.40

There are no symmetry-related clashes.

5.3 Torsion angles (i)

5.3.1 Protein backbone (i)

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	575/579 (99%)	560 (97%)	14 (2%)	1 (0%)	47 61

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	575/579 (99%)	560 (97%)	14 (2%)	1 (0%)	47 61
All	All	1150/1158 (99%)	1120 (97%)	28 (2%)	2 (0%)	47 61

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	559	ASP
1	B	559	ASP

5.3.2 Protein sidechains (i)

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	488/493 (99%)	485 (99%)	3 (1%)	86 93
1	B	488/493 (99%)	481 (99%)	7 (1%)	67 81
All	All	976/986 (99%)	966 (99%)	10 (1%)	84 87

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	56	ARG
1	A	550[A]	TRP
1	A	550[B]	TRP
1	B	56	ARG
1	B	547	LEU
1	B	550[A]	TRP
1	B	550[B]	TRP
1	B	552[A]	THR
1	B	552[B]	THR
1	B	564	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	402	HIS

5.3.3 RNA

There are no RNA molecules in this entry.

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

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

There are no carbohydrates in this entry.

5.6 Ligand geometry

1 ligand is 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	$\#\|Z\| > 2$
2	PO4	A	1565	-	4,4,4	0.90	0	6,6,6	0.43	0

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, 95^{th} 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(Å^2)	Q<0.9
1	A	563/579 (97%)	-0.52	6 (1%)	80	78	19, 29, 53, 82	0
1	B	564/579 (97%)	-0.48	10 (1%)	68	65	18, 29, 53, 82	0
All	All	1127/1158 (97%)	-0.50	16 (1%)	75	73	18, 29, 53, 82	0

All (16) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	556[A]	GLY	5.8
1	A	556[A]	GLY	5.1
1	A	557[A]	GLY	5.0
1	B	152	GLY	4.1
1	B	558[A]	GLY	4.0
1	B	377	ARG	3.6
1	B	557[A]	GLY	3.6
1	A	558[A]	GLY	3.3
1	A	16	PRO	3.2
1	B	379	ARG	3.0
1	B	153	GLY	2.8
1	B	150	ALA	2.8
1	B	554[A]	GLY	2.5
1	A	114	ARG	2.4
1	B	406[A]	ASN	2.2
1	A	149	PRO	2.2

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

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, 95^{th} 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
2	PO4	A	1565	5/5	0.92	0.21	83,83,84,84	0

6.5 Other polymers (i)

There are no such residues in this entry.