wwPDB X-ray Structure Validation Summary Report (i)

Oct 3, 2021 – 03:33 PM EDT

PDB ID	:	3K5N
Title	:	Crystal structure of E.coli Pol II-abasic DNA binary complex
Authors	:	Yang, W.; Wang, F.
Deposited on	:	2009-10-07
Resolution	:	3.15 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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

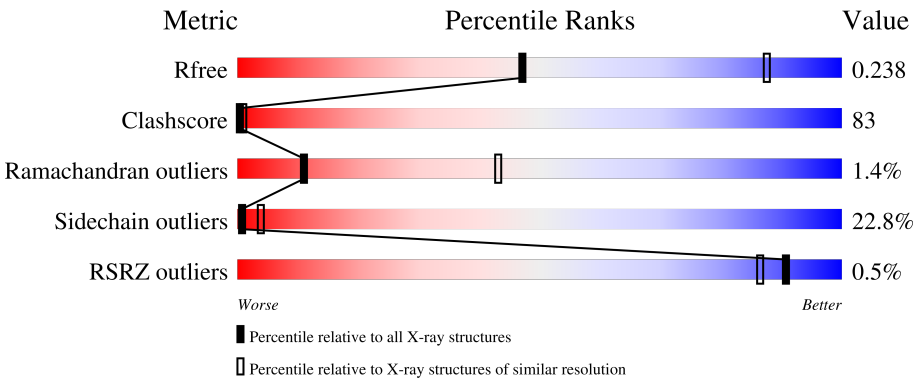
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.23.2
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 3.15 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	1665 (3.20-3.12)
Clashscore	141614	1804 (3.20-3.12)
Ramachandran outliers	138981	1770 (3.20-3.12)
Sidechain outliers	138945	1769 (3.20-3.12)
RSRZ outliers	127900	1616 (3.20-3.12)

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	786	.% 28% 50% 17% • •
1	B	786	27% 49% 17% • 6%
2	T	20	5% 30% 35% 35%
3	P	13	69% 31%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 12722 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 DNA polymerase II.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	759	6169	3934	1093	1117	25	0	0	0
1	B	738	6007	3832	1066	1086	23	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP P21189
A	-1	PRO	-	expression tag	UNP P21189
A	0	HIS	-	expression tag	UNP P21189
A	335	ASN	ASP	engineered mutation	UNP P21189
B	-2	GLY	-	expression tag	UNP P21189
B	-1	PRO	-	expression tag	UNP P21189
B	0	HIS	-	expression tag	UNP P21189
B	335	ASN	ASP	engineered mutation	UNP P21189

- Molecule 2 is a DNA chain called DNA (5'-D(*GP*TP*CP*CP*TP*GP*(3DR)*TP*AP*CP*GP*CP*TP*AP*GP*GP*CP*AP*CP*A)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	T	13	263	126	51	74	12	0	0	0

- Molecule 3 is a DNA chain called DNA (5'-D(*GP*TP*GP*CP*CP*TP*AP*GP*CP*GP*TP*AP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	P	13	255	122	45	76	12	0	0	0

- Molecule 4 is water.

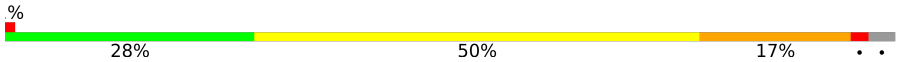
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	18	18	18	0	0
			Total	O		
4	B	10	10	10	0	0

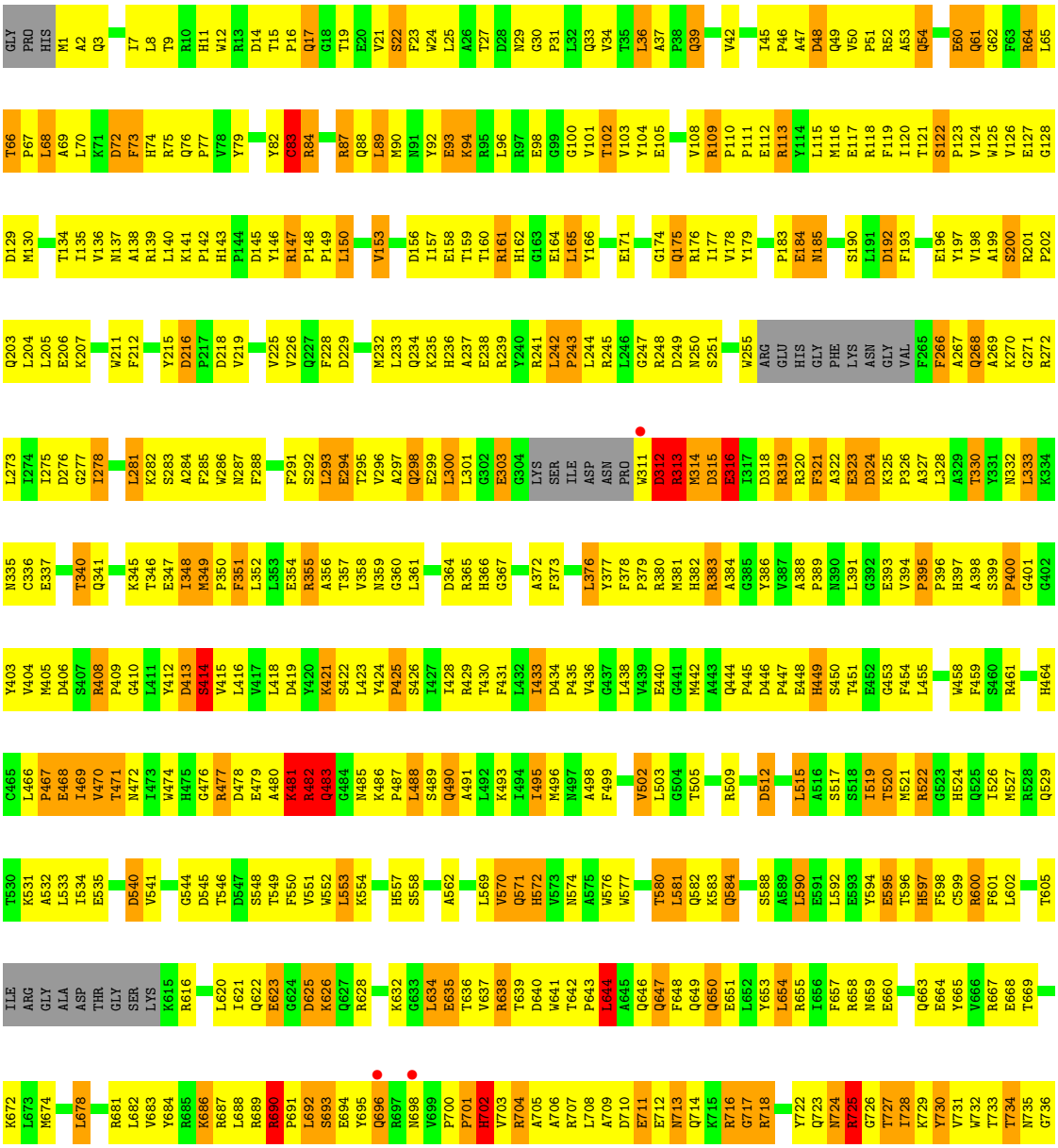
3 Residue-property plots (i)

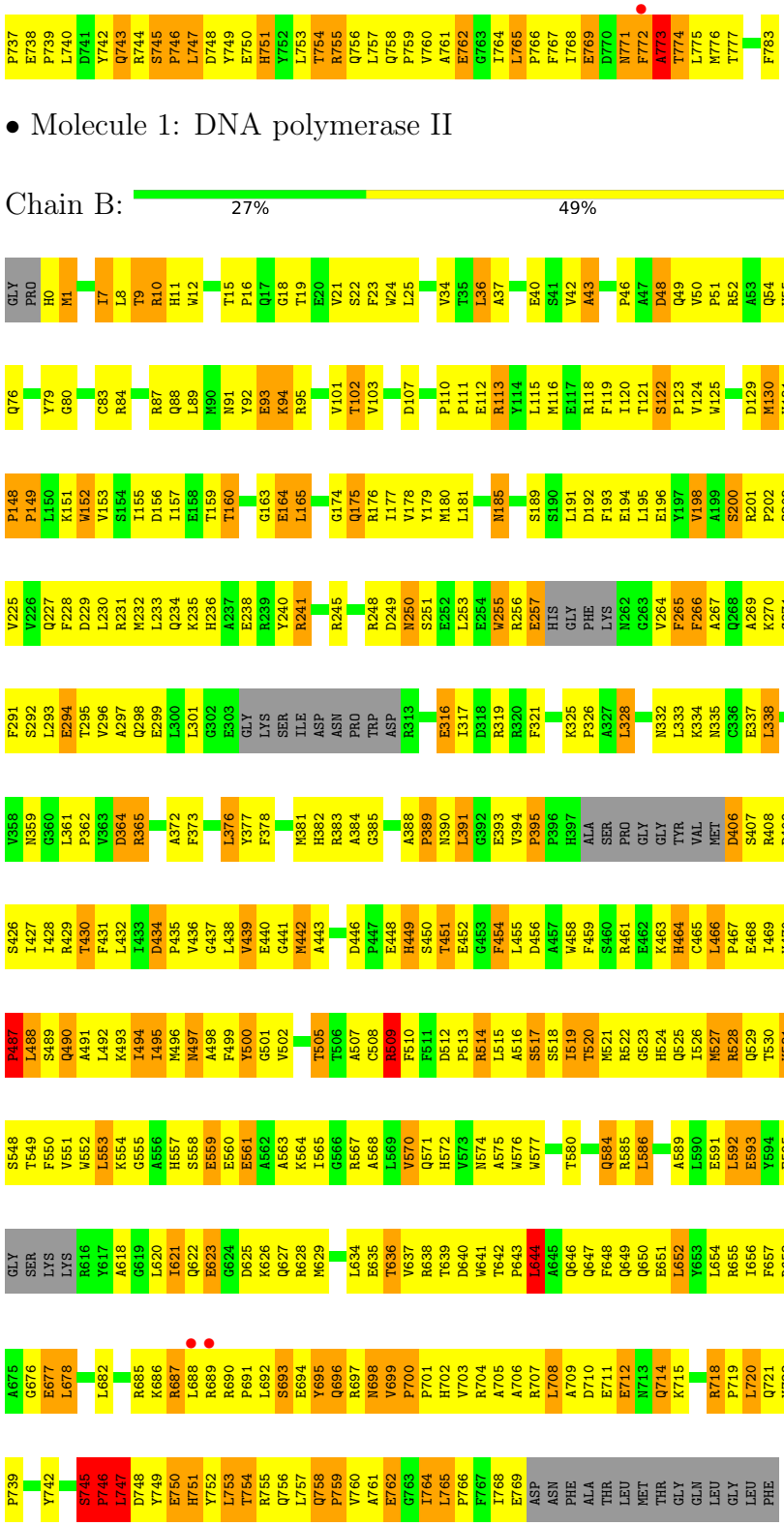
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: DNA polymerase II

Chain A:







• Molecule 1: DNA polymerase II

Chain B: 27% 49% 17% • 6%

• Molecule 2: DNA (5'-D(*GP*TP*CP*CP*TP*GP*(3DR)*TP*AP*CP*GP*CP*TP*AP*GP*GP*CP*AP*CP*A)-3')

Chain T: 5% 30% 35% 35%

• Molecule 3: DNA (5'-D(*GP*TP*GP*CP*CP*TP*AP*GP*CP*GP*TP*AP*G)-3')

Chain P:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	88.41Å 147.55Å 96.28Å 90.00° 99.98° 90.00°	Depositor
Resolution (Å)	31.31 – 3.15 46.18 – 3.15	Depositor EDS
% Data completeness (in resolution range)	95.2 (31.31-3.15) 95.2 (46.18-3.15)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.43 (at 3.12Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.232 , 0.236 0.232 , 0.238	Depositor DCC
R_{free} test set	1696 reflections (4.05%)	wwPDB-VP
Wilson B-factor (Å^2)	46.3	Xtriage
Anisotropy	0.117	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 48.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	12722	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.71% 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 (i)

5.1 Standard geometry (i)

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	0.98	3/6331 (0.0%)	1.13	32/8587 (0.4%)
1	B	0.94	2/6163 (0.0%)	1.05	18/8359 (0.2%)
2	T	1.27	2/295 (0.7%)	1.67	11/453 (2.4%)
3	P	1.50	4/285 (1.4%)	1.66	9/438 (2.1%)
All	All	0.98	11/13074 (0.1%)	1.13	70/17837 (0.4%)

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	B	0	1

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	725	ARG	CG-CD	-17.75	1.07	1.51
1	B	255	TRP	CB-CG	-8.48	1.34	1.50
3	P	902	DT	C5-C6	-6.43	1.29	1.34
1	B	152	TRP	CB-CG	-6.00	1.39	1.50
2	T	808	DT	C5-C6	-5.86	1.30	1.34

The worst 5 of 70 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	725	ARG	NE-CZ-NH1	-24.14	108.23	120.30
1	A	725	ARG	NE-CZ-NH2	15.19	127.90	120.30
1	A	725	ARG	CA-CB-CG	13.00	142.00	113.40
1	A	725	ARG	CB-CA-C	-9.47	91.46	110.40
1	B	746	PRO	CA-N-CD	-7.91	100.42	111.50

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	745	SER	Mainchain

5.2 Too-close contacts (i)

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	6169	0	6040	865	2
1	B	6007	0	5881	1075	4
2	T	263	0	147	67	0
3	P	255	0	143	59	0
4	A	18	0	0	1	0
4	B	10	0	0	0	0
All	All	12722	0	12211	2057	4

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 83.

The worst 5 of 2057 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:688:LEU:CD1	1:B:728:ILE:HD11	1.34	1.51
1:B:231:ARG:HD2	1:B:265:PHE:CE1	1.42	1.51
1:B:688:LEU:HD11	1:B:728:ILE:CD1	1.45	1.44
1:B:642:THR:CB	1:B:756:GLN:NE2	1.85	1.37
1:A:14:ASP:OD2	1:A:87:ARG:HG2	1.20	1.35

All (4) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:B:79:TYR:OH	1:B:671:ASP:OD1[2_554]	1.81	0.39
1:B:79:TYR:OH	1:B:667:ARG:NH1[2_554]	1.85	0.35
1:A:241:ARG:NH2	1:B:585:ARG:NH1[1_556]	2.14	0.06
1:A:201:ARG:NH1	1:B:714:GLN:O[2_555]	2.15	0.05

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	751/786 (96%)	698 (93%)	42 (6%)	11 (2%)	10 41
1	B	728/786 (93%)	678 (93%)	41 (6%)	9 (1%)	13 46
All	All	1479/1572 (94%)	1376 (93%)	83 (6%)	20 (1%)	11 43

5 of 20 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	324	ASP
1	A	483	GLN
1	A	690	ARG
1	A	769	GLU
1	B	1	MET

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	649/672 (97%)	496 (76%)	153 (24%)	1 3
1	B	632/672 (94%)	493 (78%)	139 (22%)	1 4
All	All	1281/1344 (95%)	989 (77%)	292 (23%)	1 4

5 of 292 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	497	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	751	HIS
1	B	520	THR
1	B	644	LEU
1	A	553	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 50 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	76	GLN
1	B	236	HIS
1	B	756	GLN
1	B	86	HIS
1	B	175	GLN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Fit of model and data (i)

6.1 Protein, DNA and RNA chains (i)

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(Å^2)	Q<0.9
1	A	759/786 (96%)	-0.44	4 (0%) 91 86	3, 29, 70, 98	0
1	B	738/786 (93%)	-0.33	2 (0%) 94 92	9, 46, 78, 93	0
2	T	13/20 (65%)	0.30	1 (7%) 13 6	85, 100, 131, 134	0
3	P	13/13 (100%)	0.45	0 100 100	60, 97, 131, 131	0
All	All	1523/1605 (94%)	-0.37	7 (0%) 91 86	3, 36, 78, 134	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	311	TRP	2.9
1	B	688	LEU	2.8
1	A	696	GLN	2.7
1	A	772	PHE	2.5
1	A	698	ASN	2.5

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

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

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.