



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 18, 2017 – 05:34 PM EDT

PDB ID : 3E3J
Title : Crystal Structure of an Intermediate Complex of T7 RNAP and 8nt of RNA
Authors : Durniak, K.J.; Bailey, S.; Steitz, T.A.
Deposited on : unknown
Resolution : 6.70 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

<http://wwpdb.org/validation/2016/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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

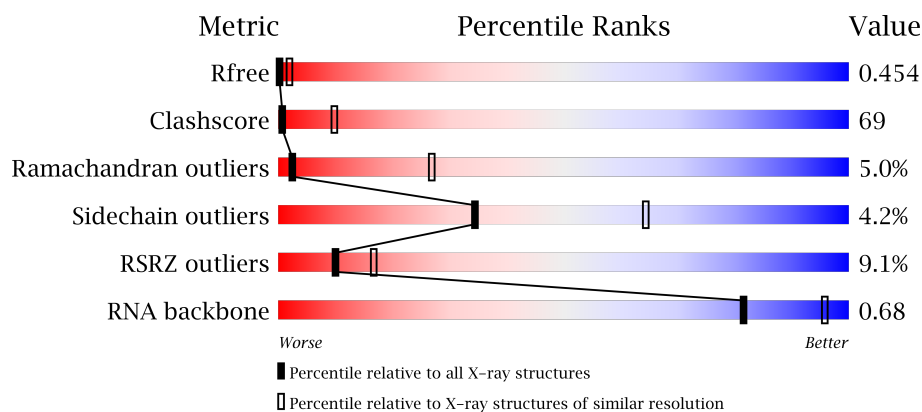
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 6.70 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}	100719	1099 (9.50-3.70)
Clashscore	112137	1033 (9.50-3.80)
Ramachandran outliers	110173	1002 (9.50-3.76)
Sidechain outliers	110143	1097 (9.50-3.70)
RSRZ outliers	101464	1002 (9.50-3.72)
RNA backbone	2435	1053 (10.00-2.90)

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. 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 $\leq 5\%$. 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	B	889	<div> <div>9%</div> <div>52% 35% 6% • 6%</div> </div>
1	C	889	<div> <div>7%</div> <div>57% 32% 5% 6%</div> </div>
2	T	32	<div> <div>9%</div> <div>19% 47% 22% 13%</div> </div>
2	X	32	<div> <div>16%</div> <div>6% 81% 13%</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	N	32	<div><div></div><div></div><div></div><div></div></div> <div>25%28%9%38%</div>
3	Y	32	<div><div></div><div></div><div></div><div></div></div> <div>13%38%13%38%</div>
4	R	8	<div><div></div><div></div><div></div><div></div></div> <div>13%75%88%</div>
4	Z	8	<div><div></div><div></div><div></div><div></div></div> <div>13%63%88%</div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 15406 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-directed RNA polymerase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	C	833	Total	C	N	O	S	0	0	0
			6567	4191	1133	1207	36			
1	B	833	Total	C	N	O	S	0	0	0
			6545	4176	1126	1206	37			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-5	HIS	-	EXPRESSION TAG	UNP P00573
C	-4	HIS	-	EXPRESSION TAG	UNP P00573
C	-3	HIS	-	EXPRESSION TAG	UNP P00573
C	-2	HIS	-	EXPRESSION TAG	UNP P00573
C	-1	HIS	-	EXPRESSION TAG	UNP P00573
C	0	HIS	-	EXPRESSION TAG	UNP P00573
C	266	LEU	PRO	ENGINEERED	UNP P00573
B	-5	HIS	-	EXPRESSION TAG	UNP P00573
B	-4	HIS	-	EXPRESSION TAG	UNP P00573
B	-3	HIS	-	EXPRESSION TAG	UNP P00573
B	-2	HIS	-	EXPRESSION TAG	UNP P00573
B	-1	HIS	-	EXPRESSION TAG	UNP P00573
B	0	HIS	-	EXPRESSION TAG	UNP P00573
B	266	LEU	PRO	ENGINEERED	UNP P00573

- Molecule 2 is a DNA chain called DNA (28-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	T	28	Total	C	N	O	P	0	0	0
			569	273	93	175	28			
2	X	28	Total	C	N	O	P	0	0	0
			569	273	93	175	28			

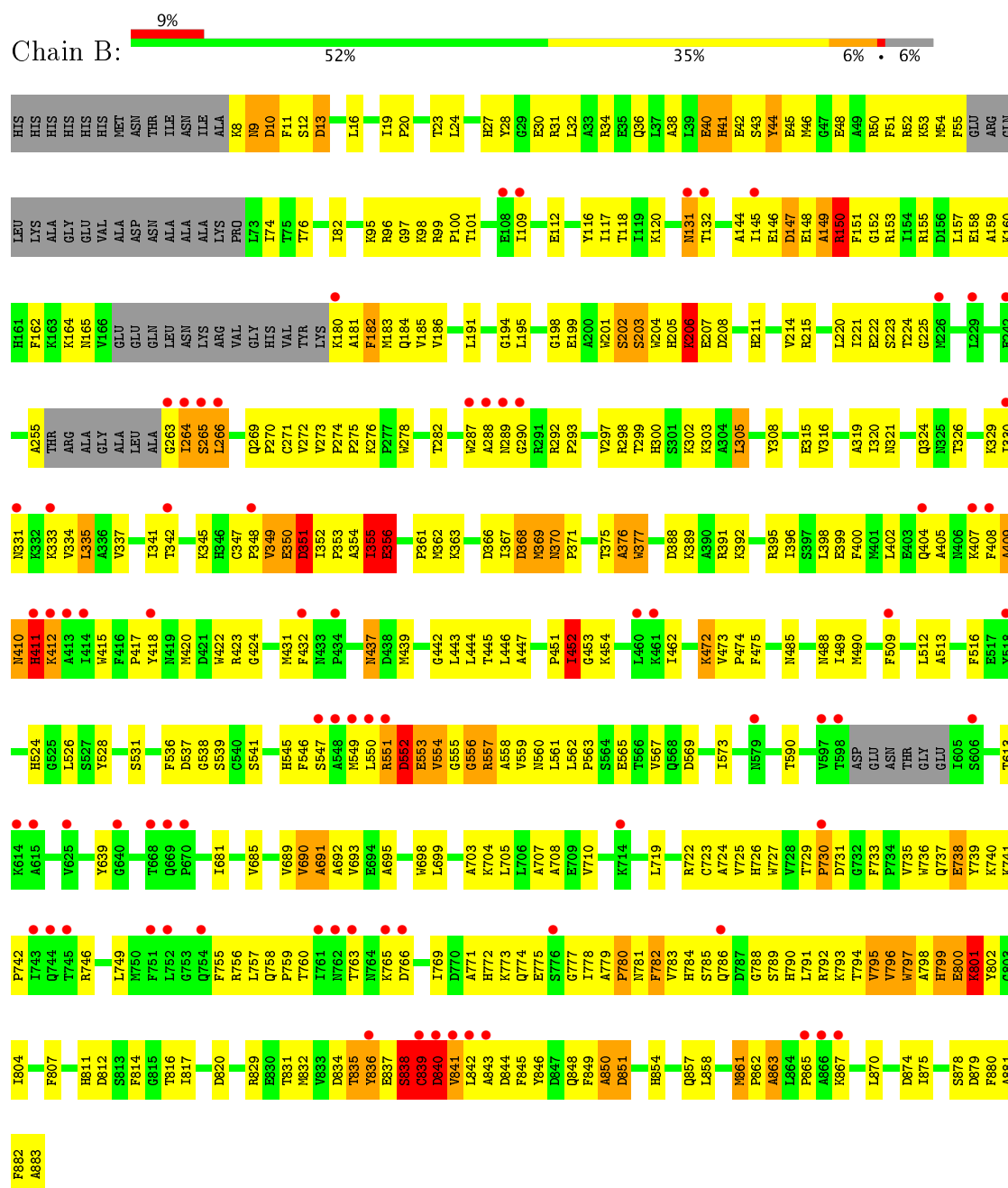
- Molecule 3 is a DNA chain called DNA (5'-D(*DTP*DAP*DAP*DTP*DAP*DCP*DGP*DAP*DCP*DTP*DCP*DAP*DCP*DTP*DAP*DTP*DAP*DTP*DTP*DCP*DTP*

DGP*DCP*DCP*DAP*DAP*DAP*DCP*DGP*DGP*DC)-3').

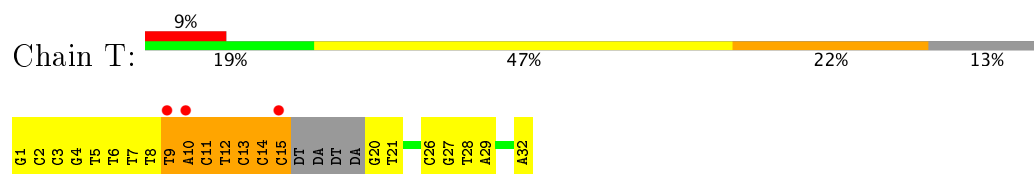
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	N	20	Total 404	C 194	N 79	O 112	P 19	0	0	0
3	Y	20	Total 404	C 194	N 79	O 112	P 19	0	0	0

- Molecule 4 is a RNA chain called RNA (5'-R(*GP*GP*GP*AP*GP*UP*AP*A)-3').

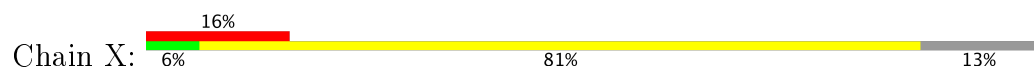
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	R	8	Total 174	C 79	N 37	O 51	P 7	0	0	0
4	Z	8	Total 174	C 79	N 37	O 51	P 7	0	0	0

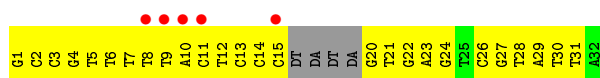


• Molecule 2: DNA (28-MER)

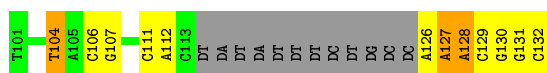
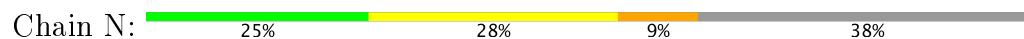


• Molecule 2: DNA (28-MER)

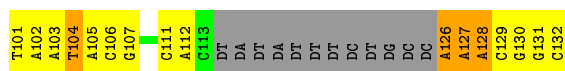




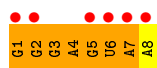
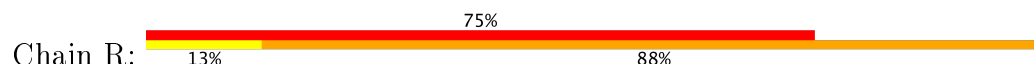
- Molecule 3: DNA (5'-D(*DTP*DAP*DAP*DTP*DAP*DCP*DGP*DAP*DCP*DTP*DCP*DAP*DCP*DTP*DAP*DTP*DAP*DTP*DTP*DTP*DCP*DTP*DGP*DCP*DCP*DAP*DAP*DAP*DCP*DGP*DGP*DC)-3')



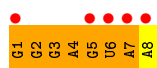
- Molecule 3: DNA (5'-D(*DTP*DAP*DAP*DTP*DAP*DCP*DGP*DAP*DCP*DTP*DCP*DAP*DCP*DTP*DAP*DTP*DAP*DTP*DTP*DTP*DCP*DTP*DGP*DCP*DCP*DAP*DAP*DAP*DCP*DGP*DGP*DC)-3')



- Molecule 4: RNA (5'-R(*GP*GP*GP*AP*GP*UP*AP*A)-3')



- Molecule 4: RNA (5'-R(*GP*GP*GP*AP*GP*UP*AP*A)-3')



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	99.22Å 180.60Å 198.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.81 – 6.70 49.82 – 6.70	Depositor EDS
% Data completeness (in resolution range)	87.2 (49.81-6.70) 88.3 (49.82-6.70)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.50 (at 6.68Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.392 , 0.454 0.390 , 0.454	Depositor DCC
R_{free} test set	273 reflections (4.78%)	DCC
Wilson B-factor (Å ²)	305.5	Xtriage
Anisotropy	0.170	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.25 , 44.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.42$, $\langle L^2 \rangle = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.76	EDS
Total number of atoms	15406	wwPDB-VP
Average B, all atoms (Å ²)	94.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.00% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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 ⓘ

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	B	0.34	0/6694	0.54	0/9055
1	C	0.34	0/6716	0.53	0/9082
2	T	1.10	6/633 (0.9%)	1.46	15/972 (1.5%)
2	X	0.74	0/633	1.10	0/972
3	N	0.88	2/453 (0.4%)	1.10	1/694 (0.1%)
3	Y	0.88	3/453 (0.7%)	1.11	1/694 (0.1%)
4	R	1.03	1/196 (0.5%)	2.15	14/306 (4.6%)
4	Z	1.02	1/196 (0.5%)	2.15	14/306 (4.6%)
All	All	0.49	13/15974 (0.1%)	0.76	45/22081 (0.2%)

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
1	C	0	2
All	All	0	3

The worst 5 of 13 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	R	6	U	C1'-N1	6.94	1.59	1.48
2	T	15	DC	C1'-N1	6.93	1.58	1.49
2	T	11	DC	C1'-N1	6.91	1.58	1.49
2	T	13	DC	C1'-N1	6.91	1.58	1.49
2	T	14	DC	C1'-N1	6.90	1.58	1.49

The worst 5 of 45 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	Z	1	G	OP2-P-O3'	7.38	121.44	105.20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	T	13	DC	OP2-P-O3'	7.25	121.16	105.20
4	Z	2	G	OP2-P-O3'	7.23	121.12	105.20
2	T	10	DA	OP2-P-O3'	7.23	121.11	105.20
4	R	6	U	OP2-P-O3'	7.22	121.09	105.20

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	40	GLU	Mainchain
1	C	149	ALA	Peptide
1	C	150	ARG	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	B	6545	0	6473	1142	1
1	C	6567	0	6519	776	2
2	T	569	0	321	126	6
2	X	569	0	321	96	7
3	N	404	0	223	28	4
3	Y	404	0	223	106	4
4	R	174	0	85	42	0
4	Z	174	0	86	21	0
All	All	15406	0	14251	2057	12

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 69.

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:28:TYR:CE1	1:B:182:PHE:HA	1.31	1.66
1:B:28:TYR:CD1	1:B:182:PHE:CB	1.77	1.64

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:724:ALA:CB	1:C:738:GLU:HB3	1.17	1.64
1:C:54:MET:CA	1:C:255:ALA:HB2	1.21	1.62
1:C:553:GLU:HB3	1:C:554:VAL:CG1	1.14	1.62

The worst 5 of 12 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)
3:N:132:DC:N4	2:X:1:DG:N3[2_454]	1.48	0.72
1:C:873:ARG:NH2	1:B:801:LYS:CE[3_444]	1.68	0.52
1:C:711:LYS:NZ	2:X:2:DC:C2'[2_454]	1.69	0.51
3:N:132:DC:N4	2:X:1:DG:C4[2_454]	1.71	0.49
2:T:1:DG:O6	2:X:1:DG:N1[2_454]	1.79	0.41

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	B	823/889 (93%)	684 (83%)	91 (11%)	48 (6%)	2	24
1	C	823/889 (93%)	704 (86%)	84 (10%)	35 (4%)	3	29
All	All	1646/1778 (93%)	1388 (84%)	175 (11%)	83 (5%)	2	27

5 of 83 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	150	ARG
1	C	153	ARG
1	C	265	SER
1	C	353	PRO
1	C	412	LYS

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	B	684/735 (93%)	650 (95%)	34 (5%)	28	60
1	C	687/735 (94%)	663 (96%)	24 (4%)	41	69
All	All	1371/1470 (93%)	1313 (96%)	58 (4%)	34	64

5 of 58 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	147	ASP
1	B	264	ILE
1	B	838	SER
1	B	150	ARG
1	B	182	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 54 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	786	GLN
1	B	86	ASN
1	B	764	ASN
1	C	823	ASN
1	B	22	ASN

5.3.3 RNA ⓘ

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
4	R	7/8 (87%)	0	0
4	Z	7/8 (87%)	0	0
All	All	14/16 (87%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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 carbohydrates 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 ⓘ

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(Å ²)	Q<0.9
1	B	833/889 (93%)	0.50	76 (9%) 10 17	20, 100, 100, 101	0
1	C	833/889 (93%)	0.47	66 (7%) 13 19	20, 100, 100, 101	0
2	T	28/32 (87%)	0.47	3 (10%) 7 14	20, 54, 100, 100	0
2	X	28/32 (87%)	0.57	5 (17%) 2 9	20, 20, 100, 100	0
3	N	20/32 (62%)	-0.29	0 100 100	20, 100, 100, 100	0
3	Y	20/32 (62%)	-0.29	0 100 100	20, 100, 100, 100	0
4	R	8/8 (100%)	2.56	6 (75%) 0 3	45, 50, 52, 55	0
4	Z	8/8 (100%)	2.69	5 (62%) 0 3	45, 50, 52, 55	0
All	All	1778/1922 (92%)	0.49	161 (9%) 10 17	20, 100, 100, 101	0

The worst 5 of 161 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	263	GLY	7.4
1	C	851	ASP	5.6
1	C	763	THR	5.1
1	B	413	ALA	5.1
1	B	288	ALA	4.9

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 ⓘ

There are no carbohydrates 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.