



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 14, 2017 – 09:51 pm GMT

PDB ID : 4HDG
Title : Crystal Structure of viral RdRp in complex with GTP
Authors : Surana, P.; Nair, D.T.
Deposited on : 2012-10-02
Resolution : 2.38 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
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Xtriage (Phenix) : 1.9-1692
EDS : trunk28620
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) : recalc28949

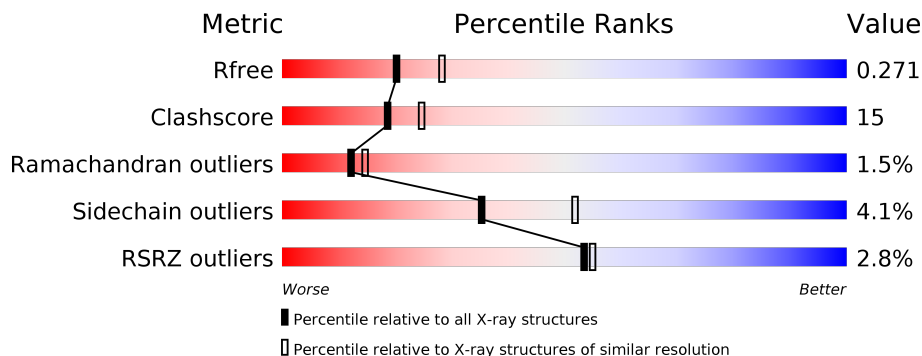
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.38 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	4388 (2.40-2.36)
Clashscore	112137	4984 (2.40-2.36)
Ramachandran outliers	110173	4907 (2.40-2.36)
Sidechain outliers	110143	4909 (2.40-2.36)
RSRZ outliers	101464	4423 (2.40-2.36)

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	A	639	<div> <div>3%</div> <div> <div></div> <div>72%</div> <div>21%</div> <div>..</div> </div> </div>
1	B	639	<div> <div>3%</div> <div> <div></div> <div>75%</div> <div>17%</div> <div>...</div> </div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
2	GTP	A	1001	-	-	X	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 10361 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.

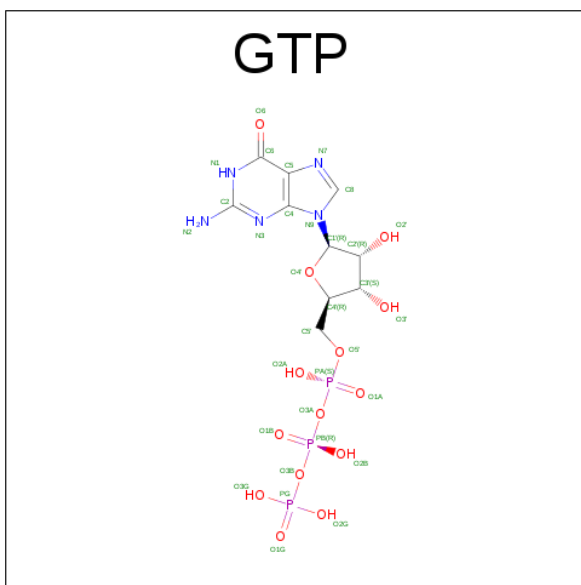
- Molecule 1 is a protein called Polypeptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	612	Total	C	N	O	S	0	0	0
			4924	3105	889	900	30			
1	B	612	Total	C	N	O	S	0	0	0
			4914	3100	884	899	31			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	267	GLY	-	EXPRESSION TAG	UNP G3LHD9
A	268	PRO	-	EXPRESSION TAG	UNP G3LHD9
A	269	LEU	-	EXPRESSION TAG	UNP G3LHD9
A	270	GLY	-	EXPRESSION TAG	UNP G3LHD9
A	271	SER	-	EXPRESSION TAG	UNP G3LHD9
A	373	ARG	LYS	CONFLICT	UNP G3LHD9
A	429	ASN	ASP	CONFLICT	UNP G3LHD9
A	836	ALA	THR	CONFLICT	UNP G3LHD9
B	267	GLY	-	EXPRESSION TAG	UNP G3LHD9
B	268	PRO	-	EXPRESSION TAG	UNP G3LHD9
B	269	LEU	-	EXPRESSION TAG	UNP G3LHD9
B	270	GLY	-	EXPRESSION TAG	UNP G3LHD9
B	271	SER	-	EXPRESSION TAG	UNP G3LHD9
B	373	ARG	LYS	CONFLICT	UNP G3LHD9
B	429	ASN	ASP	CONFLICT	UNP G3LHD9
B	836	ALA	THR	CONFLICT	UNP G3LHD9

- Molecule 2 is GUANOSINE-5'-TRIPHOSPHATE (three-letter code: GTP) (formula: $C_{10}H_{16}N_5O_{14}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			32	10	5	14	3		
2	B	1	Total	C	N	O	P	0	0
			32	10	5	14	3		

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	2	Total	Zn	0	0
			2	2		
3	A	2	Total	Zn	0	0
			2	2		

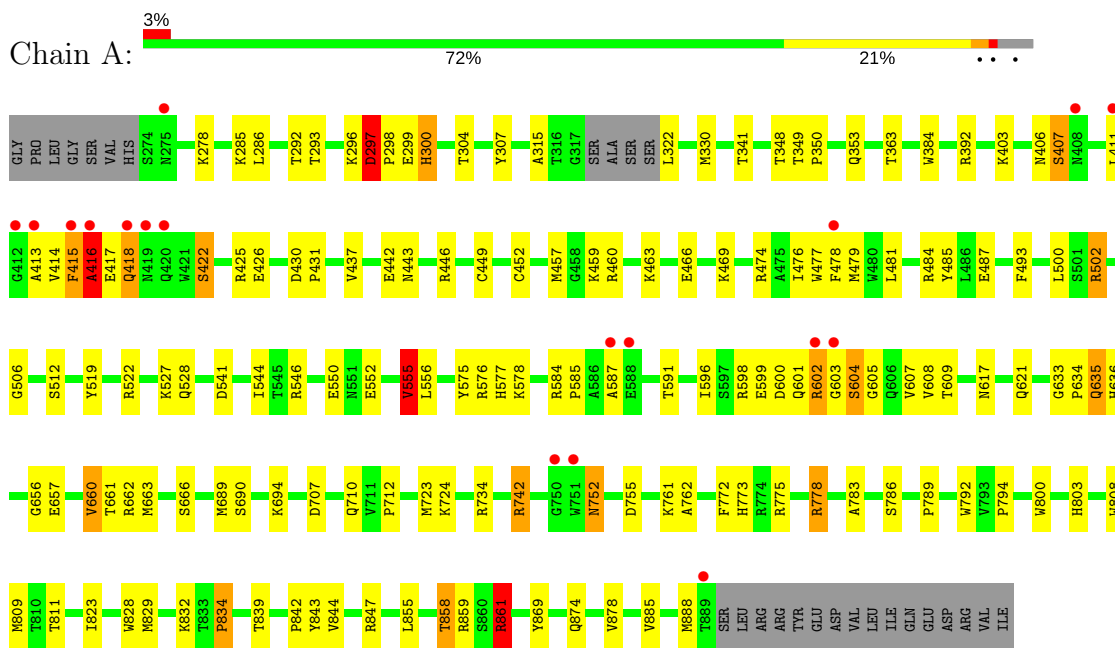
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	211	Total	O	0	0
			211	211		
4	B	244	Total	O	0	0
			244	244		

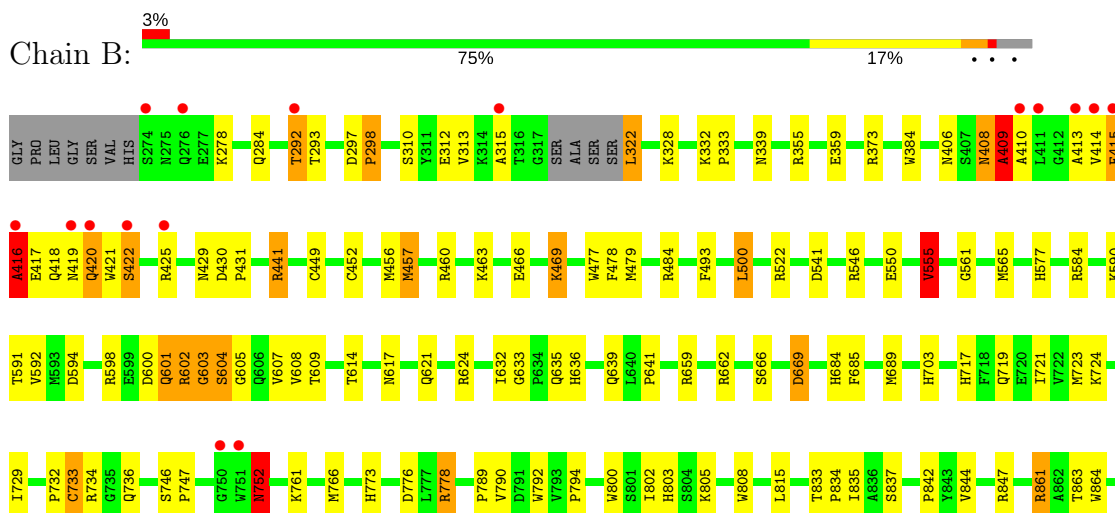
3 Residue-property plots

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: Polyprotein



• Molecule 1: Polyprotein



1867	1868	1872	1885	1889	SER	LEU	ARG	ARG	TYR	GLU	ASP	VAL	LEU	ILE	GLN	GLU	ASP	ARG	VAL	ILE
------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

4 Data and refinement statistics

Property	Value	Source
Space group	P 2 ₁ 2 ₁ 2	Depositor
Cell constants a, b, c, α , β , γ	144.57Å 86.55Å 112.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.38 49.77 – 2.38	Depositor EDS
% Data completeness (in resolution range)	99.7 (50.00-2.38) 99.8 (49.77-2.38)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.42 (at 2.37Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.220 , 0.276 0.216 , 0.271	Depositor DCC
R_{free} test set	4034 reflections (7.58%)	DCC
Wilson B-factor (Å ²)	32.0	Xtriage
Anisotropy	0.780	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 41.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	10361	wwPDB-VP
Average B, all atoms (Å ²)	35.0	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 50.41 % 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 6.4982e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GTP, ZN

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	0.85	3/5041 (0.1%)	0.82	6/6823 (0.1%)
1	B	0.83	2/5031 (0.0%)	0.81	6/6811 (0.1%)
All	All	0.84	5/10072 (0.0%)	0.82	12/13634 (0.1%)

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	2
1	B	0	2
All	All	0	4

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	422	SER	CB-OG	10.26	1.55	1.42
1	B	733	CYS	CB-SG	-8.15	1.68	1.82
1	B	298	PRO	N-CD	7.44	1.58	1.47
1	A	416	ALA	CA-CB	7.16	1.67	1.52
1	A	834	PRO	N-CD	7.13	1.57	1.47

The worst 5 of 12 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	392	ARG	NE-CZ-NH2	-9.02	115.79	120.30
1	A	500	LEU	CA-CB-CG	-7.24	98.65	115.30
1	A	502	ARG	NE-CZ-NH2	-7.19	116.70	120.30
1	A	861	ARG	NE-CZ-NH1	6.25	123.42	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	409	ALA	N-CA-C	6.25	127.86	111.00

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	300	HIS	Mainchain
1	A	416	ALA	Peptide
1	B	408	ASN	Peptide
1	B	416	ALA	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	4924	0	4826	156	0
1	B	4914	0	4808	131	0
2	A	32	0	12	9	0
2	B	32	0	12	2	0
3	A	2	0	0	0	0
3	B	2	0	0	0	0
4	A	211	0	0	7	0
4	B	244	0	0	15	0
All	All	10361	0	9658	289	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 15.

The worst 5 of 289 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:527:LYS:NZ	1:A:661:THR:OG1	1.79	1.13
1:A:742:ARG:HH11	1:A:742:ARG:HG2	0.96	1.12
1:A:487:GLU:CD	1:A:607:VAL:HG12	1.70	1.11
1:A:416:ALA:HB3	1:A:478:PHE:CE2	1.86	1.10
1:A:414:VAL:HG11	1:A:478:PHE:CE1	1.86	1.09

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	608/639 (95%)	572 (94%)	28 (5%)	8 (1%)	14	18
1	B	608/639 (95%)	570 (94%)	28 (5%)	10 (2%)	11	13
All	All	1216/1278 (95%)	1142 (94%)	56 (5%)	18 (2%)	12	15

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	297	ASP
1	A	416	ALA
1	B	409	ALA
1	B	416	ALA
1	B	420	GLN

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	516/544 (95%)	494 (96%)	22 (4%)	33	49
1	B	514/544 (94%)	494 (96%)	20 (4%)	37	54
All	All	1030/1088 (95%)	988 (96%)	42 (4%)	35	52

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	847	ARG
1	B	322	LEU
1	B	778	ARG
1	A	858	THR
1	A	861	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	710	GLN
1	B	339	ASN
1	B	765	GLN
1	A	717	HIS
1	A	803	HIS

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 for Mogul analysis.

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	GTP	A	1001	-	27,34,34	2.02	5 (18%)	27,54,54	2.20	7 (25%)
2	GTP	B	1001	-	27,34,34	1.87	8 (29%)	27,54,54	1.98	7 (25%)

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
2	GTP	A	1001	-	-	0/18/38/38	0/3/3/3
2	GTP	B	1001	-	-	0/18/38/38	0/3/3/3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	1001	GTP	PG-O3B	-2.74	1.55	1.60
2	B	1001	GTP	O2'-C2'	2.03	1.47	1.43
2	B	1001	GTP	PA-O5'	2.13	1.68	1.59
2	B	1001	GTP	C5'-C4'	2.13	1.58	1.51
2	A	1001	GTP	O2'-C2'	2.17	1.47	1.43

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1001	GTP	N3-C2-N1	-6.32	118.24	127.46
2	B	1001	GTP	N3-C2-N1	-5.49	119.44	127.46
2	A	1001	GTP	C4'-O4'-C1'	-3.22	106.35	109.77
2	A	1001	GTP	C5-C6-N1	-3.04	119.16	123.48
2	B	1001	GTP	C4'-O4'-C1'	-2.96	106.62	109.77

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1001	GTP	9	0
2	B	1001	GTP	2	0

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	612/639 (95%)	0.05	18 (2%) 52 53	15, 33, 60, 81	0
1	B	612/639 (95%)	0.02	16 (2%) 56 58	16, 33, 57, 76	0
All	All	1224/1278 (95%)	0.04	34 (2%) 53 55	15, 33, 58, 81	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	416	ALA	9.0
1	A	415	PHE	6.1
1	B	416	ALA	5.2
1	A	412	GLY	4.6
1	A	413	ALA	4.4

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 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(\AA^2)	Q<0.9
2	GTP	B	1001	32/32	0.90	0.20	1.70	52,71,78,78	0
2	GTP	A	1001	32/32	0.90	0.19	0.77	44,72,78,80	0
3	ZN	B	1002	1/1	0.99	0.07	-1.96	31,31,31,31	0
3	ZN	A	1003	1/1	0.98	0.09	-1.98	36,36,36,36	1
3	ZN	A	1002	1/1	0.99	0.08	-2.02	33,33,33,33	0
3	ZN	B	1003	1/1	0.99	0.07	-4.26	37,37,37,37	1

6.5 Other polymers [i](#)

There are no such residues in this entry.