



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 06:30 am GMT

PDB ID : 3HW3
Title : The crystal structure of avian influenza virus PA_N in complex with UMP
Authors : Zhao, C.; Lou, Z.; Guo, Y.; Ma, M.; Chen, Y.; Rao, Z.
Deposited on : 2009-06-17
Resolution : 1.90 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 following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467
Mogul : 1.7.3 (157068), CSD as539be (2018)
Xtriage (Phenix) : 1.13
EDS : trunk30967
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac : 5.8.0158
CCP4 : 7.0 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : trunk30967

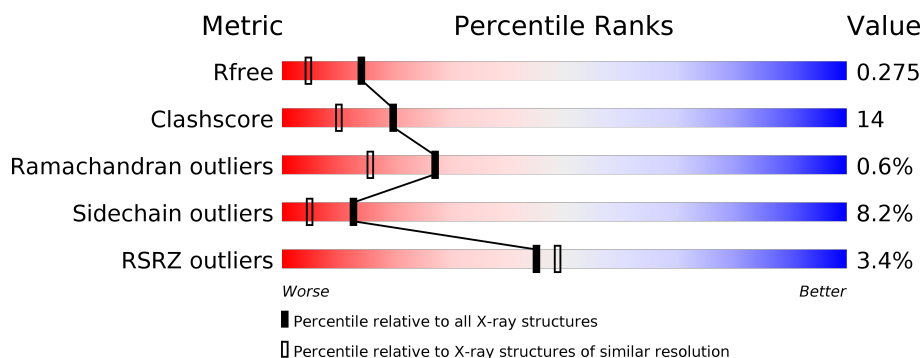
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.90 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}	111664	5502 (1.90-1.90)
Clashscore	122126	6115 (1.90-1.90)
Ramachandran outliers	120053	6048 (1.90-1.90)
Sidechain outliers	120020	6048 (1.90-1.90)
RSRZ outliers	108989	5379 (1.90-1.90)

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	A	261	<div> <div>48%</div> <div>16%</div> <div>• •</div> <div>32%</div> </div>
1	B	261	<div> <div>3%</div> <div>46%</div> <div>18%</div> <div>•</div> <div>31%</div> </div>
1	C	261	<div> <div>3%</div> <div>55%</div> <div>11%</div> <div>•</div> <div>32%</div> </div>
1	D	261	<div> <div>3%</div> <div>50%</div> <div>15%</div> <div>•</div> <div>33%</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 crite-

ria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	U5P	A	257	-	-	X	-

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6372 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	177	Total	C	N	O	S	0	0	0
			1462	924	251	276	11			
1	B	180	Total	C	N	O	S	0	0	0
			1496	948	258	279	11			
1	C	177	Total	C	N	O	S	0	0	0
			1463	925	253	274	11			
1	D	176	Total	C	N	O	S	0	0	0
			1453	919	250	273	11			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-4	GLY	-	EXPRESSION TAG	UNP Q9Q0U9
A	-3	PRO	-	EXPRESSION TAG	UNP Q9Q0U9
A	-2	LEU	-	EXPRESSION TAG	UNP Q9Q0U9
A	-1	GLY	-	EXPRESSION TAG	UNP Q9Q0U9
A	0	SER	-	EXPRESSION TAG	UNP Q9Q0U9
A	201	ILE	VAL	SEE REMARK 999	UNP Q9Q0U9
B	-4	GLY	-	EXPRESSION TAG	UNP Q9Q0U9
B	-3	PRO	-	EXPRESSION TAG	UNP Q9Q0U9
B	-2	LEU	-	EXPRESSION TAG	UNP Q9Q0U9
B	-1	GLY	-	EXPRESSION TAG	UNP Q9Q0U9
B	0	SER	-	EXPRESSION TAG	UNP Q9Q0U9
B	201	ILE	VAL	SEE REMARK 999	UNP Q9Q0U9
C	-4	GLY	-	EXPRESSION TAG	UNP Q9Q0U9
C	-3	PRO	-	EXPRESSION TAG	UNP Q9Q0U9
C	-2	LEU	-	EXPRESSION TAG	UNP Q9Q0U9
C	-1	GLY	-	EXPRESSION TAG	UNP Q9Q0U9
C	0	SER	-	EXPRESSION TAG	UNP Q9Q0U9
C	201	ILE	VAL	SEE REMARK 999	UNP Q9Q0U9
D	-4	GLY	-	EXPRESSION TAG	UNP Q9Q0U9
D	-3	PRO	-	EXPRESSION TAG	UNP Q9Q0U9
D	-2	LEU	-	EXPRESSION TAG	UNP Q9Q0U9

Continued on next page...

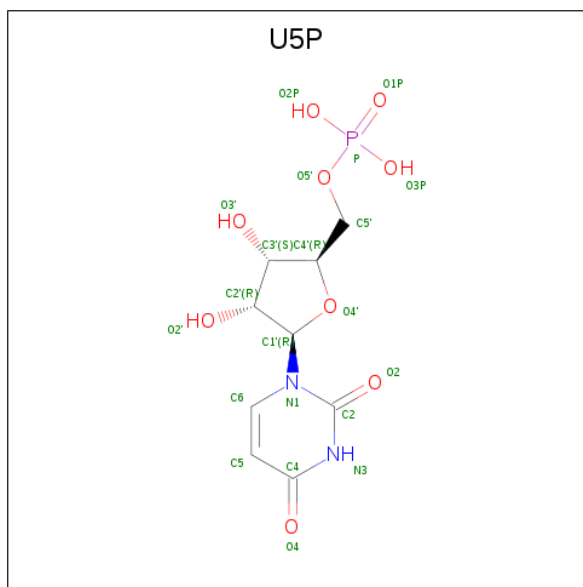
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	-1	GLY	-	EXPRESSION TAG	UNP Q9Q0U9
D	0	SER	-	EXPRESSION TAG	UNP Q9Q0U9
D	201	ILE	VAL	SEE REMARK 999	UNP Q9Q0U9

- Molecule 2 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Mg	0	0
			1	1		
2	A	1	Total	Mg	0	0
			1	1		
2	D	1	Total	Mg	0	0
			1	1		
2	C	1	Total	Mg	0	0
			1	1		

- Molecule 3 is URIDINE-5'-MONOPHOSPHATE (three-letter code: U5P) (formula: C₉H₁₃N₂O₉P).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
3	A	1	Total	C	N	O	P	0	0
			21	9	2	9	1		

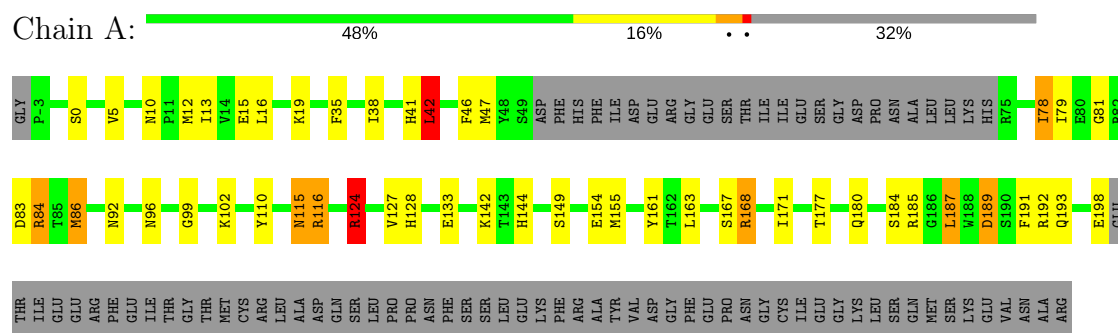
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	144	Total 144	O 144	0	0
4	B	112	Total 112	O 112	0	0
4	C	111	Total 111	O 111	0	0
4	D	106	Total 106	O 106	0	0

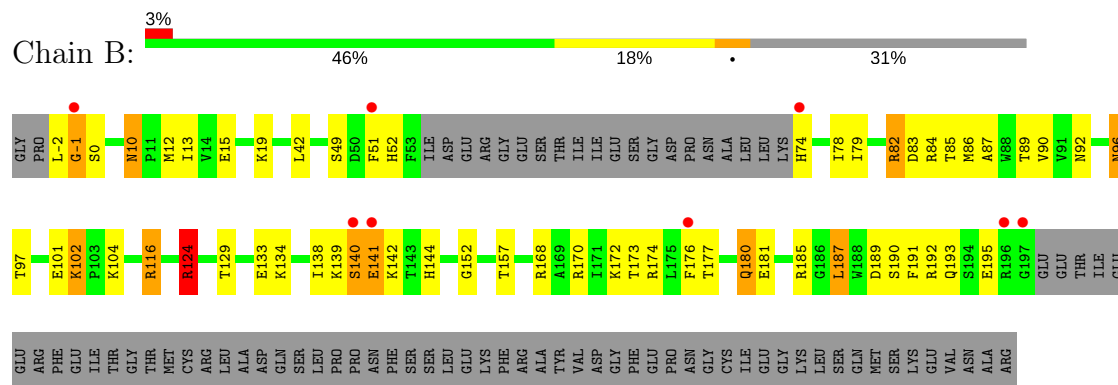
3 Residue-property plots

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.

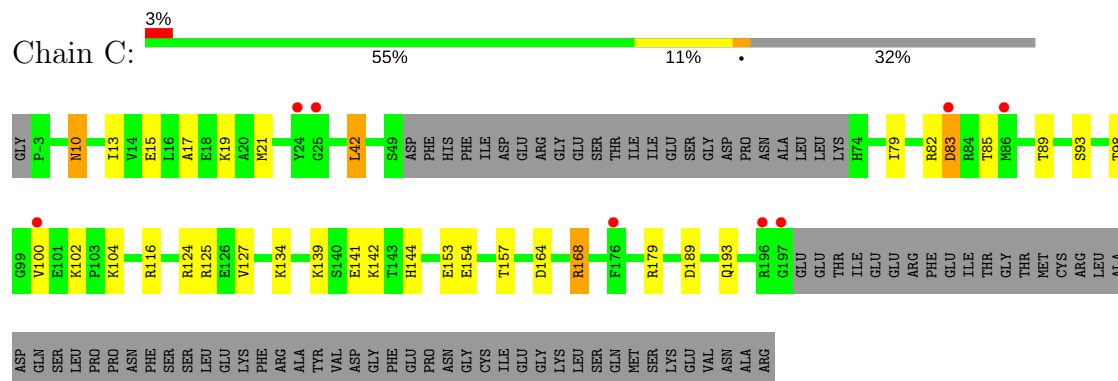
• Molecule 1: Polymerase acidic protein



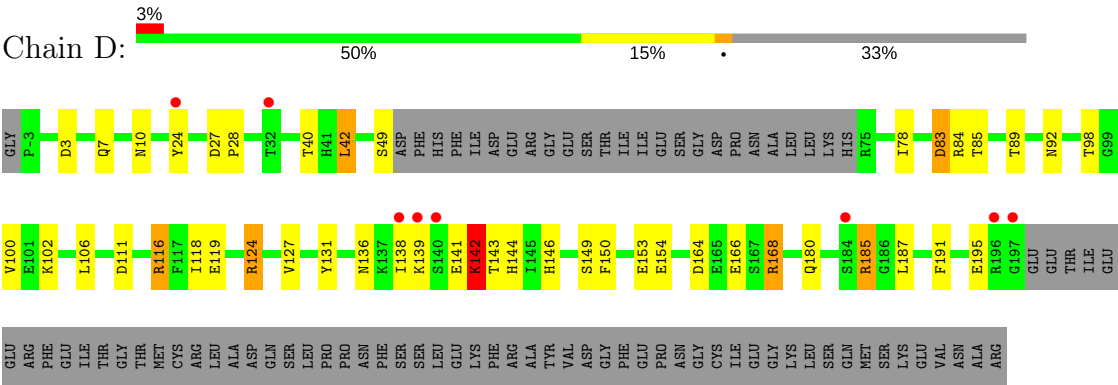
• Molecule 1: Polymerase acidic protein



• Molecule 1: Polymerase acidic protein



● Molecule 1: Polymerase acidic protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	51.13Å 59.70Å 66.85Å 96.28° 96.65° 109.44°	Depositor
Resolution (Å)	50.00 – 1.90 47.58 – 1.90	Depositor EDS
% Data completeness (in resolution range)	95.9 (50.00-1.90) 95.9 (47.58-1.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.42 (at 1.91Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.218 , 0.276 0.217 , 0.275	Depositor DCC
R_{free} test set	2806 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å ²)	26.7	Xtriage
Anisotropy	0.035	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.38 , 51.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	6372	wwPDB-VP
Average B, all atoms (Å ²)	28.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.63% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: MG, U5P

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	0.78	0/1491	0.89	7/2000 (0.3%)
1	B	0.77	0/1528	0.96	6/2050 (0.3%)
1	C	0.70	0/1493	0.86	6/2003 (0.3%)
1	D	0.74	0/1482	0.92	7/1988 (0.4%)
All	All	0.75	0/5994	0.91	26/8041 (0.3%)

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

There are no bond length outliers.

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	116	ARG	NE-CZ-NH1	10.59	125.59	120.30
1	B	168	ARG	NE-CZ-NH2	-10.21	115.19	120.30
1	C	168	ARG	NE-CZ-NH2	-10.01	115.29	120.30
1	D	116	ARG	NE-CZ-NH1	9.74	125.17	120.30
1	D	168	ARG	NE-CZ-NH2	-9.51	115.55	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	-1	GLY	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	1462	0	1435	49	0
1	B	1496	0	1458	52	0
1	C	1463	0	1436	21	0
1	D	1453	0	1429	39	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	21	0	11	10	0
4	A	144	0	0	3	0
4	B	112	0	0	13	0
4	C	111	0	0	4	0
4	D	106	0	0	9	0
All	All	6372	0	5769	160	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 14.

The worst 5 of 160 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)
3:A:257:U5P:H3'	3:A:257:U5P:C6	1.67	1.18
1:A:38:ILE:HG12	3:A:257:U5P:H2'	1.23	1.13
3:A:257:U5P:H6	3:A:257:U5P:C3'	1.73	1.13
1:B:170:ARG:O	1:B:173:THR:HG22	1.58	1.04
1:B:173:THR:HB	4:B:400:HOH:O	1.62	0.98

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	173/261 (66%)	168 (97%)	4 (2%)	1 (1%)	27	15
1	B	176/261 (67%)	170 (97%)	4 (2%)	2 (1%)	16	5
1	C	173/261 (66%)	168 (97%)	5 (3%)	0	100	100
1	D	172/261 (66%)	163 (95%)	8 (5%)	1 (1%)	27	15
All	All	694/1044 (66%)	669 (96%)	21 (3%)	4 (1%)	27	15

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	142	LYS
1	B	141	GLU
1	B	140	SER
1	A	99	GLY

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	159/232 (68%)	145 (91%)	14 (9%)	11	4
1	B	162/232 (70%)	145 (90%)	17 (10%)	7	3
1	C	159/232 (68%)	150 (94%)	9 (6%)	23	12
1	D	158/232 (68%)	146 (92%)	12 (8%)	14	6
All	All	638/928 (69%)	586 (92%)	52 (8%)	12	5

5 of 52 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	104	LYS
1	B	187	LEU
1	D	166	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	124	ARG
1	B	139	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	52	HIS
1	B	92	ASN
1	C	74	HIS
1	B	10	ASN
1	C	33	ASN

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 5 ligands modelled in this entry, 4 are monoatomic - leaving 1 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	$\# Z > 2$
3	U5P	A	257	2	17,22,22	1.69	5 (29%)	21,33,33	3.43	10 (47%)

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	U5P	A	257	2	-	0/6/26/26	0/2/2/2

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	257	U5P	C2'-N3	-3.01	1.32	1.38
3	A	257	U5P	O4'-C4'	-2.59	1.39	1.45
3	A	257	U5P	C3'-C4'	-2.24	1.47	1.53
3	A	257	U5P	O5'-C5'	-2.17	1.36	1.44
3	A	257	U5P	P-O3P	-2.01	1.46	1.54

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	A	257	U5P	C2'-C3'-C4'	-7.72	87.80	102.62
3	A	257	U5P	C4'-O4'-C1'	-6.47	103.08	109.83
3	A	257	U5P	O3P-P-O5'	-4.13	95.73	106.73
3	A	257	U5P	O4'-C4'-C3'	-3.26	98.70	105.15
3	A	257	U5P	O2P-P-O5'	-3.18	98.26	106.73

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 10 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	257	U5P	10	0

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 [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	177/261 (67%)	0.07	0 100 100	14, 23, 38, 42	0
1	B	180/261 (68%)	0.27	8 (4%) 34 37	14, 25, 42, 52	0
1	C	177/261 (67%)	0.20	8 (4%) 33 36	18, 29, 43, 55	0
1	D	176/261 (67%)	0.35	8 (4%) 33 36	16, 28, 49, 60	0
All	All	710/1044 (68%)	0.22	24 (3%) 45 49	14, 26, 43, 60	0

The worst 5 of 24 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	196	ARG	7.4
1	D	197	GLY	7.1
1	D	196	ARG	5.3
1	B	140	SER	4.4
1	B	176	PHE	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. 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(\AA^2)	Q<0.9
3	U5P	A	257	21/21	0.80	0.24	26,47,51,52	0
2	MG	B	999	1/1	0.93	0.05	23,23,23,23	0
2	MG	D	999	1/1	0.95	0.06	35,35,35,35	0
2	MG	C	999	1/1	0.97	0.04	23,23,23,23	0
2	MG	A	999	1/1	0.97	0.06	24,24,24,24	0

6.5 Other polymers [i](#)

There are no such residues in this entry.