W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 21, 2024 – 11:23 PM EST

PDB ID	:	4RI9
Title	:	FAN1 Nuclease bound to 5' phosphorylated p(dT)/3'(dT-dT-dT-dT-dT-dT-dT-dT) double flap DNA
Authors	:	Pavletich, N.P.; Wang, R.
Deposited on	:	2014-10-05
Resolution	:	2.90 Å(reported)

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

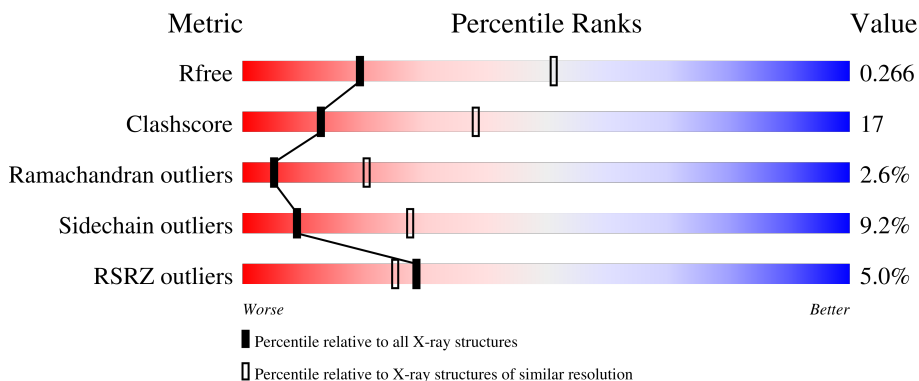|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.90 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.
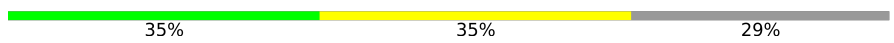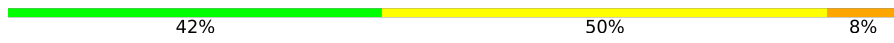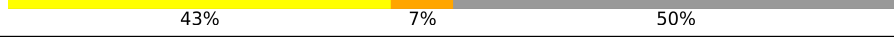
Metric	Percentile Ranks	Value
Rfree		0.266
Clashscore		17
Ramachandran outliers		2.6%
Sidechain outliers		9.2%
RSRZ outliers		5.0%

Worse   *Better*

▌ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 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	652	4% / 58% / 31% / 5% / 6%
1	B	652	6% / 60% / 30% / · / 6%
2	S	11	9% / 91%
2	W	11	18% / 82%
3	U	17	53% / 18% / 29%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	Z	17	35% 35% 29%
4	V	12	42% 50% 8%
4	X	12	25% 67% 8%
5	T	14	43% 7% 50%
5	Y	14	21% 29% 50%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 11606 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 Fanconi-associated nuclease 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	615	Total 4942	C 3145	N 880	O 890	S 27	0	0	0
1	B	615	Total 4942	C 3145	N 880	O 890	S 27	0	0	0

There are 46 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	357	SER	-	expression tag	UNP Q9Y2M0
A	358	GLY	-	expression tag	UNP Q9Y2M0
A	359	ALA	-	expression tag	UNP Q9Y2M0
A	360	HIS	-	expression tag	UNP Q9Y2M0
A	361	MET	-	expression tag	UNP Q9Y2M0
A	362	THR	-	expression tag	UNP Q9Y2M0
A	363	ARG	-	expression tag	UNP Q9Y2M0
A	364	ASN	-	expression tag	UNP Q9Y2M0
A	365	GLY	-	expression tag	UNP Q9Y2M0
A	366	PRO	-	expression tag	UNP Q9Y2M0
A	367	GLY	-	expression tag	UNP Q9Y2M0
A	368	GLN	-	expression tag	UNP Q9Y2M0
A	369	THR	-	expression tag	UNP Q9Y2M0
A	487	ALA	VAL	engineered mutation	UNP Q9Y2M0
A	?	-	CYS	deletion	UNP Q9Y2M0
A	?	-	THR	deletion	UNP Q9Y2M0
A	?	-	TRP	deletion	UNP Q9Y2M0
A	?	-	GLY	deletion	UNP Q9Y2M0
A	?	-	LYS	deletion	UNP Q9Y2M0
A	?	-	ASN	deletion	UNP Q9Y2M0
A	?	-	LYS	deletion	UNP Q9Y2M0
A	?	-	PRO	deletion	UNP Q9Y2M0
A	?	-	GLY	deletion	UNP Q9Y2M0
B	357	SER	-	expression tag	UNP Q9Y2M0
B	358	GLY	-	expression tag	UNP Q9Y2M0

WORLDWIDE PDB
PROTEIN DATA BANK

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	359	ALA	-	expression tag	UNP Q9Y2M0
B	360	HIS	-	expression tag	UNP Q9Y2M0
B	361	MET	-	expression tag	UNP Q9Y2M0
B	362	THR	-	expression tag	UNP Q9Y2M0
B	363	ARG	-	expression tag	UNP Q9Y2M0
B	364	ASN	-	expression tag	UNP Q9Y2M0
B	365	GLY	-	expression tag	UNP Q9Y2M0
B	366	PRO	-	expression tag	UNP Q9Y2M0
B	367	GLY	-	expression tag	UNP Q9Y2M0
B	368	GLN	-	expression tag	UNP Q9Y2M0
B	369	THR	-	expression tag	UNP Q9Y2M0
B	487	ALA	VAL	engineered mutation	UNP Q9Y2M0
B	?	-	CYS	deletion	UNP Q9Y2M0
B	?	-	THR	deletion	UNP Q9Y2M0
B	?	-	TRP	deletion	UNP Q9Y2M0
B	?	-	GLY	deletion	UNP Q9Y2M0
B	?	-	LYS	deletion	UNP Q9Y2M0
B	?	-	ASN	deletion	UNP Q9Y2M0
B	?	-	LYS	deletion	UNP Q9Y2M0
B	?	-	PRO	deletion	UNP Q9Y2M0
B	?	-	GLY	deletion	UNP Q9Y2M0

- Molecule 2 is a DNA chain called DNA (5'-D(P*TP*AP*GP*CP*CP*AP*CP*GP*CP*CP *T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	S	11	Total	C	N	O	P	0	0	0
			222	105	39	67	11			
2	W	11	Total	C	N	O	P	0	0	0
			222	105	39	67	11			

- Molecule 3 is a DNA chain called DNA (5'-D(P*AP*GP*AP*CP*TP*CP*CP*TP*CP*TP *TP*TP*TP*TP*TP*TP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	U	12	Total	C	N	O	P	0	0	0
			240	116	37	75	12			
3	Z	12	Total	C	N	O	P	0	0	0
			240	116	37	75	12			

- Molecule 4 is a DNA chain called DNA (5'-D(P*GP*CP*TP*GP*AP*GP*GP*AP*GP*T P*CP*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	V	12	Total	C	N	O	P	3	0	0
			250	118	47	73	12			
4	X	12	Total	C	N	O	P	3	0	0
			250	118	47	73	12			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	Y	7	Total	C	N	O	P	0	0	0
			148	69	30	42	7			
5	T	7	Total	C	N	O	P	0	0	0
			148	69	30	42	7			

- Molecule 6 is BARIUM ION (three-letter code: BA) (formula: Ba).

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	1	Total	Ba	0	0
			1	1		
6	B	1	Total	Ba	0	0
			1	1		

3 Residue-property plots ⓘ

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: Fanconi-associated nuclease 1
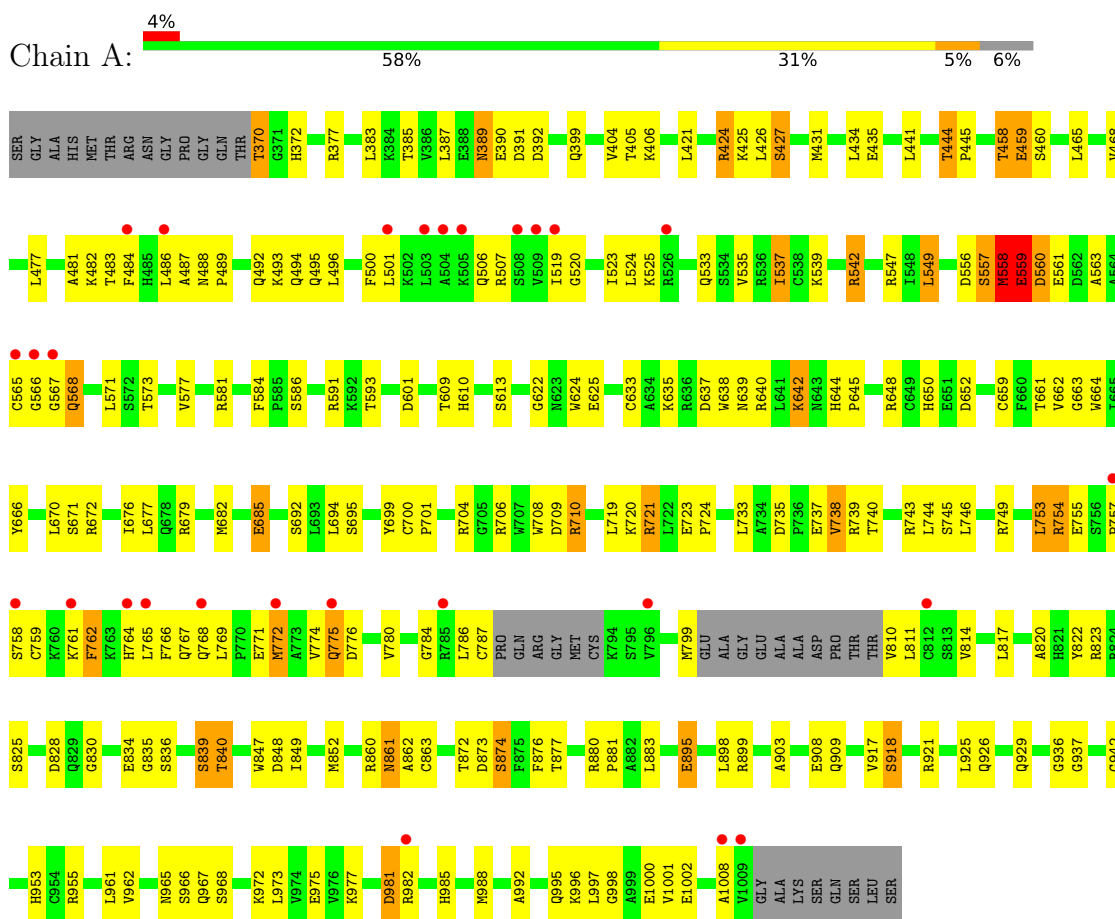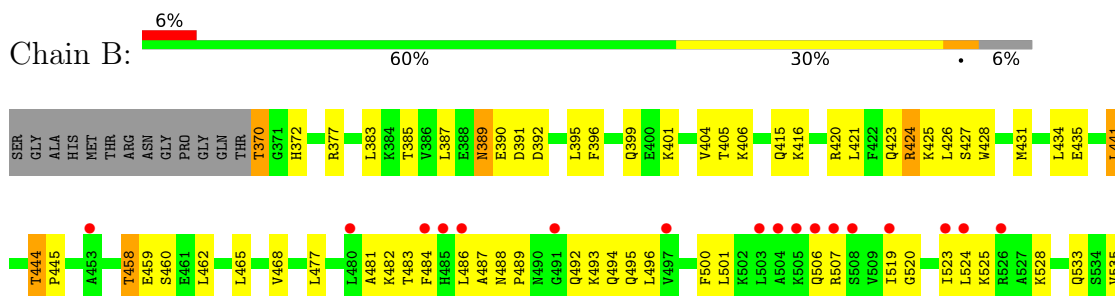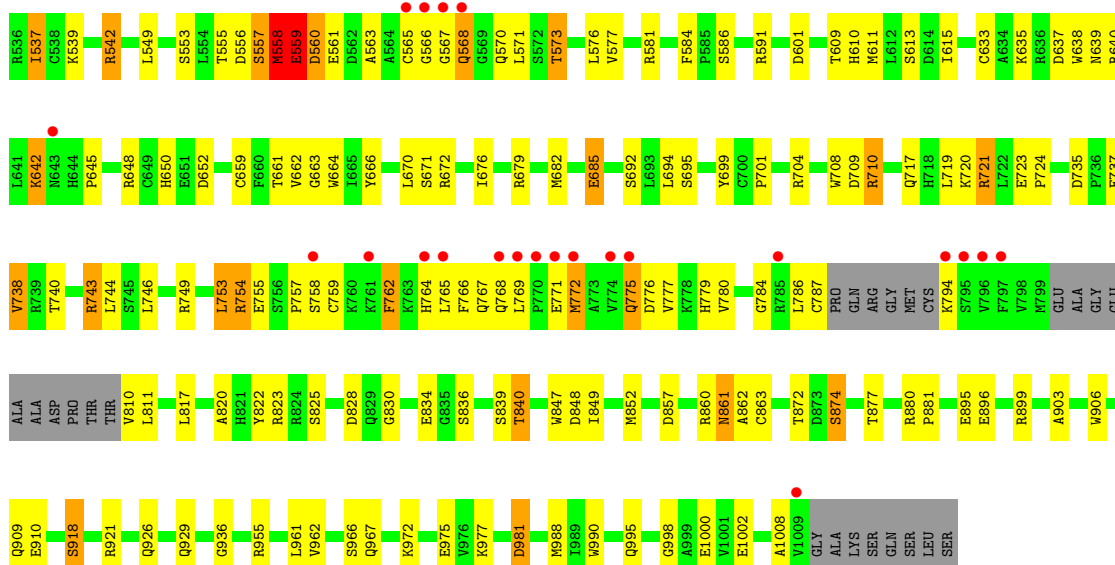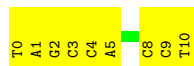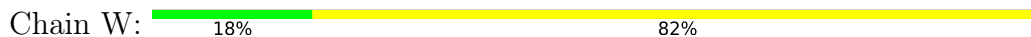
Chain A:

• Molecule 1: Fanconi-associated nuclease 1

Chain B:

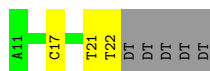• Molecule 2: DNA (5'-D(P*TP*AP*GP*CP*CP*AP*CP*GP*CP*CP*T)-3')

Chain S:
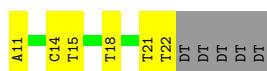

• Molecule 2: DNA (5'-D(P*TP*AP*GP*CP*CP*AP*CP*GP*CP*CP*T)-3')

Chain W:

• Molecule 3: DNA (5'-D(P*AP*GP*AP*CP*TP*CP*CP*TP*CP*TP*TP*TP*TP*TP*TP*T P*T)-3')

Chain U:

• Molecule 3: DNA (5'-D(P*AP*GP*AP*CP*TP*CP*CP*TP*CP*TP*TP*TP*TP*TP*TP*T P*T)-3')

Chain Z:

• Molecule 4: DNA (5'-D(P*GP*CP*TP*GP*AP*GP*GP*AP*GP*TP*CP*T)-3')

Chain V:

42% 50% 8%

G1 C2 T3 G4 A5 G6 G7 A8 T12

● Molecule 4: DNA (5'-D(P*GP*CP*TP*GP*AP*GP*GP*AP*GP*TP*CP*T)-3')

Chain X:

25% 67% 8%

G1 C2 T3 G4 A5 G6 G7 A8 G9 T12
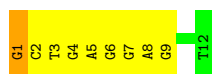
● Molecule 5: DNA (5'-D(*TP*TP*TP*TP*TP*TP*GP*AP*GP*GP*CP*GP*TP*G)-3')

Chain Y:

21% 29% 50%

DT DT DT DT DT DG A8 G9 G10 C11 G14
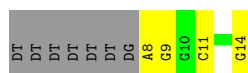
● Molecule 5: DNA (5'-D(*TP*TP*TP*TP*TP*TP*GP*AP*GP*GP*CP*GP*TP*G)-3')

Chain T:

43% 7% 50%

DT DT DT DT DT DG A8 G9 G10 C11 G12 T13 G14
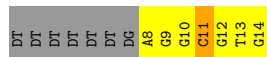
4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	66.86Å 211.57Å 69.27Å 90.00° 99.15° 90.00°	Depositor
Resolution (Å)	50.00 – 2.90 68.39 – 2.90	Depositor EDS
% Data completeness (in resolution range)	91.6 (50.00-2.90) 91.6 (68.39-2.90)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.69 (at 2.91Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.227 , 0.266 0.227 , 0.266	Depositor DCC
R_{free} test set	1627 reflections (4.06%)	wwPDB-VP
Wilson B-factor (Å^2)	62.0	Xtriage
Anisotropy	0.387	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 26.8	EDS
L-test for twinning[2]	$< |L| > = 0.48, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.078 for l,-k,h	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	11606	wwPDB-VP
Average B, all atoms (Å^2)	79.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.88% 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)

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

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	0.51	0/5044	0.71	1/6818 (0.0%)
1	B	0.49	0/5044	0.71	1/6818 (0.0%)
2	S	0.71	1/247 (0.4%)	0.87	1/376 (0.3%)
2	W	0.75	1/247 (0.4%)	0.87	0/376
3	U	0.50	0/266	0.82	1/407 (0.2%)
3	Z	0.41	0/266	0.82	0/407
4	V	0.83	1/280 (0.4%)	0.97	2/431 (0.5%)
4	X	2.18	1/280 (0.4%)	1.39	4/431 (0.9%)
5	T	0.58	0/166	0.93	2/255 (0.8%)
5	Y	0.56	0/166	0.83	1/255 (0.4%)
All	All	0.61	4/12006 (0.0%)	0.76	13/16574 (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 (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	X	1	DG	P-O5'	35.95	1.95	1.59
4	V	1	DG	P-O5'	12.60	1.72	1.59
2	W	0	DT	OP3-P	-10.29	1.48	1.61
2	S	0	DT	OP3-P	-9.46	1.49	1.61

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	X	1	DG	P-O5'-C5'	-20.55	88.01	120.90
4	X	1	DG	O5'-P-OP1	10.82	123.68	110.70
4	V	1	DG	P-O5'-C5'	-10.05	104.81	120.90
4	X	1	DG	O5'-P-OP2	-6.59	99.77	105.70
5	T	11	DC	C1'-O4'-C4'	-6.30	103.80	110.10

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	567	GLY	Peptide
1	A	861	ASN	Peptide
1	B	567	GLY	Peptide
1	B	861	ASN	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	4942	0	4978	157	0
1	B	4942	0	4978	156	0
2	S	222	0	124	6	0
2	W	222	0	124	11	1
3	U	240	0	138	4	0
3	Z	240	0	138	11	1
4	V	250	0	136	9	0
4	X	250	0	136	12	0
5	T	148	0	79	8	0
5	Y	148	0	79	5	0
6	A	1	0	0	0	0
6	B	1	0	0	0	0
All	All	11606	0	10910	367	1

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 17.

The worst 5 of 367 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:679:ARG:NH1	5:Y:14:DG:OP2	1.87	1.05
1:B:836:SER:O	1:B:840:THR:HG23	1.60	1.01
1:B:565:CYS:HB3	1:B:609:THR:HG23	1.49	0.94
1:A:836:SER:O	1:A:840:THR:HG23	1.68	0.94
2:W:1:DA:H2"	2:W:2:DG:OP2	1.69	0.93

All (1) 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)
2:W:10:DT:O3'	3:Z:11:DA:P[2_645]	1.91	0.29

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	609/652 (93%)	530 (87%)	60 (10%)	19 (3%)	4	16
1	B	609/652 (93%)	525 (86%)	71 (12%)	13 (2%)	7	26
All	All	1218/1304 (93%)	1055 (87%)	131 (11%)	32 (3%)	5	20

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	427	SER
1	A	558	MET
1	A	560	ASP
1	A	720	LYS
1	A	767	GLN

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	538/564 (95%)	489 (91%)	49 (9%)	9	28
1	B	538/564 (95%)	488 (91%)	50 (9%)	9	27
All	All	1076/1128 (95%)	977 (91%)	99 (9%)	9	27

5 of 99 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	435	GLU
1	B	573	THR
1	B	444	THR
1	B	537	ILE
1	B	662	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	832	HIS
1	B	953	HIS
1	A	926	GLN
1	A	953	HIS
1	B	488	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 (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	615/652 (94%)	0.06	27 (4%)	34	30	39, 66, 143, 171	0
1	B	615/652 (94%)	0.18	39 (6%)	20	16	39, 71, 150, 171	0
2	S	11/11 (100%)	-0.84	0	100	100	81, 87, 94, 98	0
2	W	11/11 (100%)	-0.81	0	100	100	89, 108, 126, 129	0
3	U	12/17 (70%)	-0.74	0	100	100	68, 88, 130, 148	0
3	Z	12/17 (70%)	-0.62	0	100	100	61, 85, 128, 139	0
4	V	12/12 (100%)	-0.98	0	100	100	74, 89, 114, 136	1 (8%)
4	X	12/12 (100%)	-0.94	0	100	100	80, 90, 105, 112	1 (8%)
5	T	7/14 (50%)	-0.83	0	100	100	63, 74, 93, 101	0
5	Y	7/14 (50%)	-0.75	0	100	100	55, 66, 81, 94	0
All	All	1314/1412 (93%)	0.06	66 (5%)	28	25	39, 70, 147, 171	2 (0%)

The worst 5 of 66 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	506	GLN	7.9
1	B	770	PRO	7.2
1	B	519	ILE	6.9
1	B	567	GLY	6.6
1	B	774	VAL	6.3

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 ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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(Å^2)	Q<0.9
6	BA	A	1101	1/1	1.00	0.21	71,71,71,71	0
6	BA	B	1101	1/1	1.00	0.20	73,73,73,73	0

6.5 Other polymers ⓘ

There are no such residues in this entry.