wwPDB X-ray Structure Validation Summary Report (i)

Nov 21, 2023 – 11:29 PM JST

PDB ID	:	8HNI
Title	:	hnRNP A2/B1 RRMs in complex with telomeric DNA
Authors	:	Liu, Y.; Abula, A.; Xiao, H.; Guo, H.; Li, T.; Zheng, L.; Chen, B.; Nguyen, H.; Ji, X.
Deposited on	:	2022-12-07
Resolution	:	2.64 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

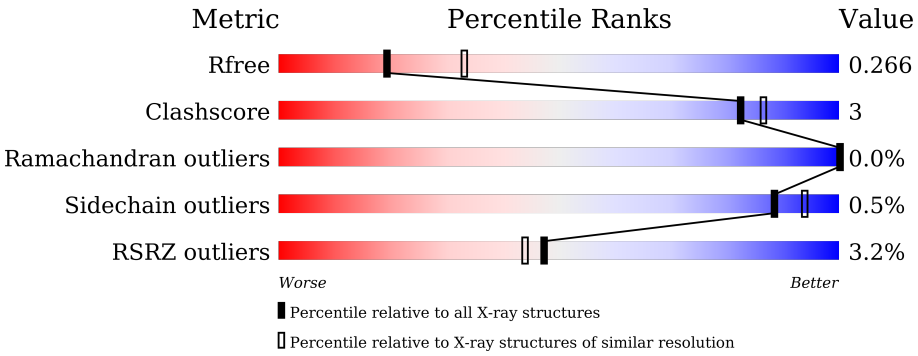
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.64 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.



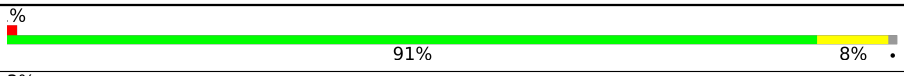
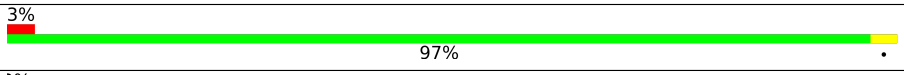
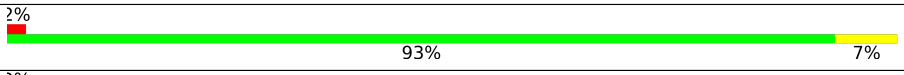
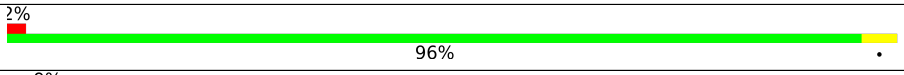
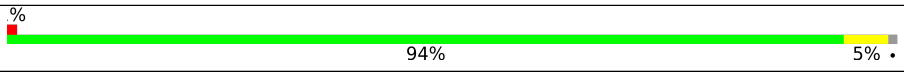
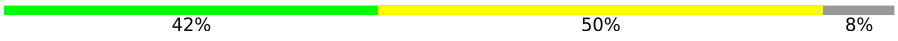
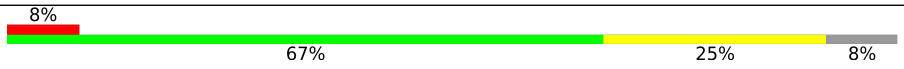
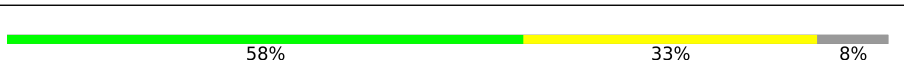
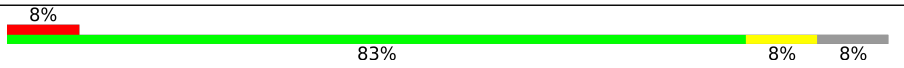
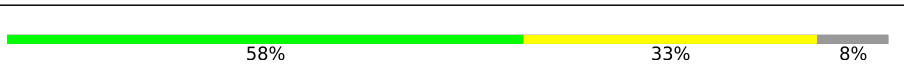
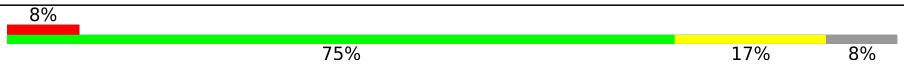
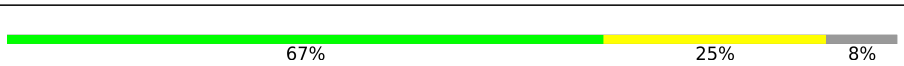
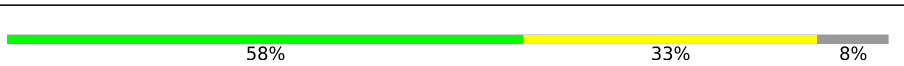
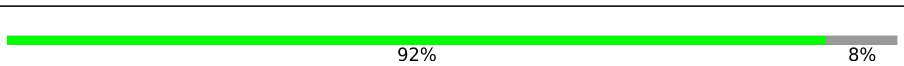
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1426 (2.66-2.62)
Clashscore	141614	1472 (2.66-2.62)
Ramachandran outliers	138981	1446 (2.66-2.62)
Sidechain outliers	138945	1446 (2.66-2.62)
RSRZ outliers	127900	1408 (2.66-2.62)

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	179	95% 5%
1	B	179	2% 97% .
1	C	179	3% 94% 6%
1	D	179	3% 96% .
1	E	179	% 94% 6%
1	F	179	11% 90% 5% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	179	
1	H	179	
1	I	179	
1	J	179	
1	K	179	
1	L	179	
2	M	12	
2	N	12	
2	O	12	
2	P	12	
2	Q	12	
2	R	12	
2	S	12	
2	T	12	
2	U	12	
2	V	12	
2	W	12	
2	X	12	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 35084 atoms, of which 16260 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 Heterogeneous nuclear ribonucleoproteins A2/B1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	179	Total	C	H	N	O	S	0	0	0
			2668	866	1284	252	261	5			
1	B	179	Total	C	H	N	O	S	0	0	0
			2750	885	1341	250	269	5			
1	C	179	Total	C	H	N	O	S	0	0	0
			2639	854	1269	250	261	5			
1	D	179	Total	C	H	N	O	S	0	0	0
			2555	840	1220	239	252	4			
1	E	179	Total	C	H	N	O	S	0	0	0
			2712	876	1319	249	263	5			
1	F	170	Total	C	H	N	O	S	0	0	0
			1981	692	877	198	211	3			
1	G	177	Total	C	H	N	O	S	0	0	0
			2676	865	1300	249	257	5			
1	H	179	Total	C	H	N	O	S	0	0	0
			2659	865	1287	247	256	4			
1	I	179	Total	C	H	N	O	S	0	0	0
			2711	872	1320	250	265	4			
1	J	179	Total	C	H	N	O	S	0	0	0
			2633	857	1269	250	252	5			
1	K	161	Total	C	H	N	O	S	0	0	0
			2111	717	976	202	211	5			
1	L	178	Total	C	H	N	O	S	0	0	0
			2690	872	1305	246	262	5			

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	M	12	Total	C	H	N	O	P	0	0	0
			389	120	138	48	72	11			
2	N	11	Total	C	H	N	O	P	0	0	0
			354	110	124	46	64	10			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	O	11	Total	C	H	N	O	P	0	0	0
			356	110	125	46	65	10			
2	P	11	Total	C	H	N	O	P	0	0	0
			357	110	124	46	66	11			
2	Q	11	Total	C	H	N	O	P	0	0	0
			358	110	124	46	67	11			
2	R	11	Total	C	H	N	O	P	0	0	0
			353	110	123	45	65	10			
2	S	11	Total	C	H	N	O	P	0	0	0
			354	110	120	46	67	11			
2	T	11	Total	C	H	N	O	P	0	0	0
			358	110	124	46	67	11			
2	U	11	Total	C	H	N	O	P	0	0	0
			355	110	121	46	67	11			
2	V	11	Total	C	H	N	O	P	0	0	0
			358	110	124	46	67	11			
2	W	11	Total	C	H	N	O	P	0	0	0
			358	110	124	46	67	11			
2	X	11	Total	C	H	N	O	P	0	0	0
			349	109	122	46	62	10			

3 Residue-property plots (i)

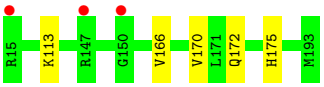
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Heterogeneous nuclear ribonucleoproteins A2/B1

Chain A:

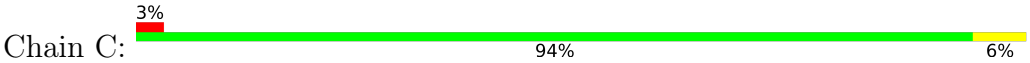
• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

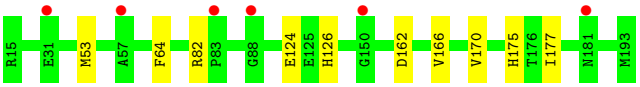
Chain B:



• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

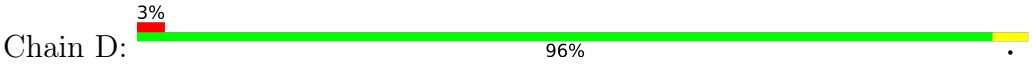
Chain C:

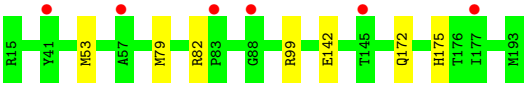




• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

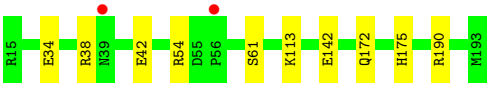
Chain D:





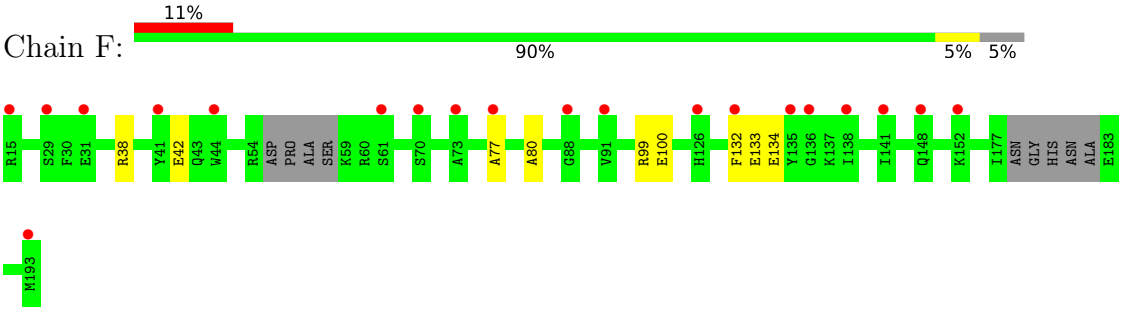
• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

Chain E:



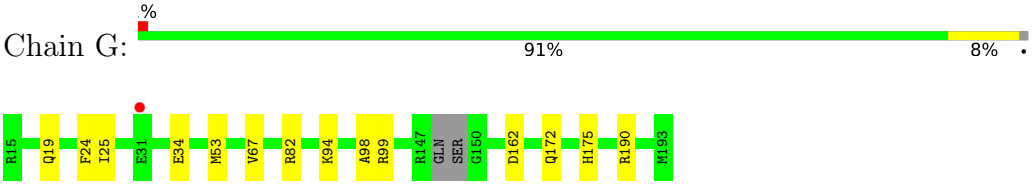
• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

Chain F:



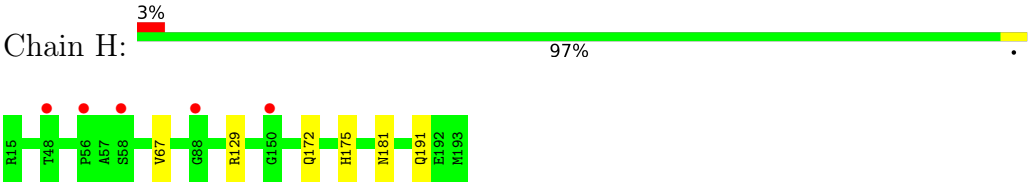
• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

Chain G:



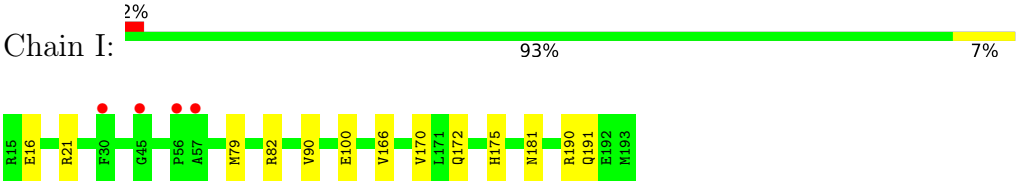
• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

Chain H:



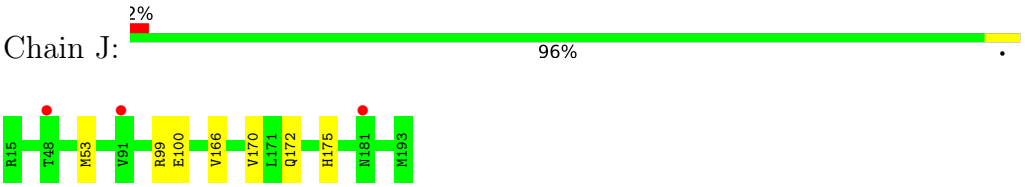
• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

Chain I:



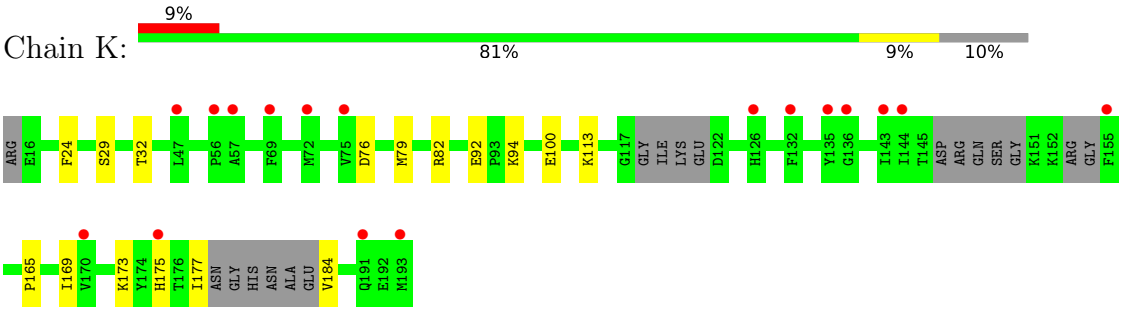
• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

Chain J:



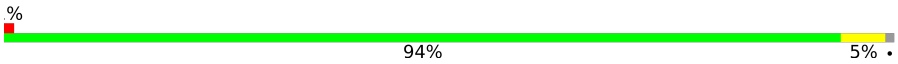
• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

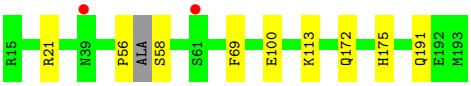
Chain K:



• Molecule 1: Heterogeneous nuclear ribonucleoproteins A2/B1

Chain L:



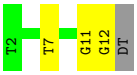


• Molecule 2: DNA (5'-D(P*TP*AP*GP*GP*GP*TP*TP*AP*GP*GP*GP*T)-3')

Chain M:

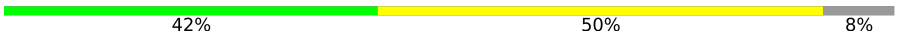
• Molecule 2: DNA (5'-D(P*TP*AP*GP*GP*GP*TP*TP*AP*GP*GP*GP*T)-3')

Chain N:



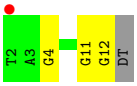
• Molecule 2: DNA (5'-D(P*TP*AP*GP*GP*GP*TP*TP*AP*GP*GP*GP*T)-3')

Chain O:



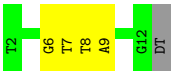
• Molecule 2: DNA (5'-D(P*TP*AP*GP*GP*GP*TP*TP*AP*GP*GP*GP*T)-3')

Chain P:



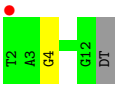
• Molecule 2: DNA (5'-D(P*TP*AP*GP*GP*GP*TP*TP*AP*GP*GP*GP*T)-3')

Chain Q:



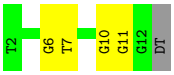
• Molecule 2: DNA (5'-D(P*TP*AP*GP*GP*GP*TP*TP*AP*GP*GP*GP*T)-3')

Chain R:



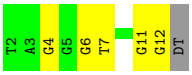
• Molecule 2: DNA (5'-D(P*TP*AP*GP*GP*GP*TP*TP*AP*GP*GP*GP*T)-3')

Chain S:



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

Chain T:

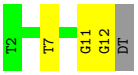


• Molecule 2: DNA (5'-D(P*TP*AP*GP*GP*GP*TP*TP*AP*GP*GP*GP*T)-3')

Chain U:

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

Chain V:



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

Chain W:

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

Chain X:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	225.25Å 98.31Å 145.93Å 90.00° 130.37° 90.00°	Depositor
Resolution (Å)	46.80 – 2.64 46.80 – 2.64	Depositor EDS
% Data completeness (in resolution range)	99.5 (46.80-2.64) 99.5 (46.80-2.64)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.50 (at 2.65Å)	Xtriage
Refinement program	PHENIX (1.10_2148: ???)	Depositor
R, R_{free}	0.237 , 0.262 0.241 , 0.266	Depositor DCC
R_{free} test set	1996 reflections (2.82%)	wwPDB-VP
Wilson B-factor (Å^2)	51.0	Xtriage
Anisotropy	0.139	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 36.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.024 for -h-2*l,-k,l	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	35084	wwPDB-VP
Average B, all atoms (Å^2)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.68% 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

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	A	0.26	0/1413	0.44	0/1902
1	B	0.26	0/1438	0.44	0/1931
1	C	0.26	0/1398	0.45	0/1882
1	D	0.25	0/1363	0.44	0/1842
1	E	0.28	0/1422	0.47	0/1913
1	F	0.33	0/1122	0.50	0/1527
1	G	0.25	0/1404	0.44	0/1888
1	H	0.28	0/1401	0.45	0/1887
1	I	0.28	0/1419	0.46	0/1909
1	J	0.26	0/1393	0.45	0/1877
1	K	0.29	0/1155	0.48	0/1561
1	L	0.24	0/1413	0.44	0/1901
2	M	0.47	0/282	1.00	0/436
2	N	0.43	0/259	0.89	0/401
2	O	0.43	0/260	0.89	0/402
2	P	0.46	0/262	0.90	0/404
2	Q	0.46	0/263	0.90	0/406
2	R	0.46	0/259	0.88	0/400
2	S	0.56	0/263	0.88	0/406
2	T	0.46	0/263	0.90	0/406
2	U	0.60	0/263	0.94	0/406
2	V	0.42	0/263	0.89	0/406
2	W	0.45	0/263	0.90	0/406
2	X	0.47	0/256	0.89	0/395
All	All	0.31	0/19497	0.56	0/26894

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	1384	1284	1282	6	0
1	B	1409	1341	1341	3	0
1	C	1370	1269	1269	5	0
1	D	1335	1220	1218	6	0
1	E	1393	1319	1319	5	0
1	F	1104	877	875	6	0
1	G	1376	1300	1300	9	0
1	H	1372	1287	1287	4	0
1	I	1391	1320	1320	8	0
1	J	1364	1269	1269	5	0
1	K	1135	976	969	15	0
1	L	1385	1305	1304	5	0
2	M	251	138	138	3	0
2	N	230	124	123	2	0
2	O	231	125	126	6	0
2	P	233	124	125	2	0
2	Q	234	124	125	3	0
2	R	230	123	124	1	0
2	S	234	120	125	4	0
2	T	234	124	125	4	0
2	U	234	121	125	3	0
2	V	234	124	125	2	0
2	W	234	124	125	3	0
2	X	227	122	122	0	0
All	All	18824	16260	16261	89	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 3.

The worst 5 of 89 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:L:21:ARG:NH1	1:L:69:PHE:O	2.20	0.73
1:F:38:ARG:O	1:F:42:GLU:CB	2.37	0.72
1:K:177:ILE:HD12	1:K:177:ILE:N	2.11	0.66

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:N:7:DT:O2	1:H:191:GLN:NE2	2.31	0.63
1:D:53:MET:HE3	2:S:11:DG:H5'	1.80	0.62

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	177/179 (99%)	176 (99%)	1 (1%)	0	100	100
1	B	177/179 (99%)	176 (99%)	1 (1%)	0	100	100
1	C	177/179 (99%)	176 (99%)	1 (1%)	0	100	100
1	D	177/179 (99%)	176 (99%)	1 (1%)	0	100	100
1	E	177/179 (99%)	175 (99%)	2 (1%)	0	100	100
1	F	164/179 (92%)	158 (96%)	5 (3%)	1 (1%)	25	37
1	G	173/179 (97%)	172 (99%)	1 (1%)	0	100	100
1	H	177/179 (99%)	175 (99%)	2 (1%)	0	100	100
1	I	177/179 (99%)	175 (99%)	2 (1%)	0	100	100
1	J	177/179 (99%)	173 (98%)	4 (2%)	0	100	100
1	K	151/179 (84%)	149 (99%)	2 (1%)	0	100	100
1	L	174/179 (97%)	173 (99%)	1 (1%)	0	100	100
All	All	2078/2148 (97%)	2054 (99%)	23 (1%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	F	133	GLU

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	135/153 (88%)	135 (100%)	0	100 100
1	B	144/153 (94%)	144 (100%)	0	100 100
1	C	134/153 (88%)	134 (100%)	0	100 100
1	D	126/153 (82%)	126 (100%)	0	100 100
1	E	140/153 (92%)	138 (99%)	2 (1%)	67 80
1	F	80/153 (52%)	80 (100%)	0	100 100
1	G	137/153 (90%)	135 (98%)	2 (2%)	65 79
1	H	134/153 (88%)	133 (99%)	1 (1%)	84 91
1	I	141/153 (92%)	140 (99%)	1 (1%)	84 91
1	J	131/153 (86%)	131 (100%)	0	100 100
1	K	95/153 (62%)	95 (100%)	0	100 100
1	L	139/153 (91%)	138 (99%)	1 (1%)	84 91
All	All	1536/1836 (84%)	1529 (100%)	7 (0%)	88 94

5 of 7 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	190	ARG
1	H	129	ARG
1	L	191	GLN
1	I	181	ASN
1	G	34	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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

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, 95^{th} 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	179/179 (100%)	0.16	0 100 100	31, 56, 79, 110	0
1	B	179/179 (100%)	0.32	3 (1%) 70 67	32, 54, 87, 113	0
1	C	179/179 (100%)	0.47	6 (3%) 45 41	35, 66, 102, 109	0
1	D	179/179 (100%)	0.43	6 (3%) 45 41	40, 69, 104, 125	0
1	E	179/179 (100%)	0.30	2 (1%) 80 78	33, 55, 104, 134	0
1	F	170/179 (94%)	0.72	20 (11%) 4 3	39, 90, 123, 135	0
1	G	177/179 (98%)	0.21	1 (0%) 89 88	31, 57, 86, 100	0
1	H	179/179 (100%)	0.26	5 (2%) 53 49	36, 55, 87, 118	0
1	I	179/179 (100%)	0.24	4 (2%) 62 58	35, 55, 89, 120	0
1	J	179/179 (100%)	0.35	3 (1%) 70 67	36, 65, 102, 121	0
1	K	161/179 (89%)	0.54	17 (10%) 6 4	40, 76, 104, 142	0
1	L	178/179 (99%)	0.27	2 (1%) 80 78	32, 51, 86, 110	0
2	M	12/12 (100%)	0.10	0 100 100	40, 54, 77, 96	0
2	N	11/12 (91%)	-0.02	0 100 100	34, 45, 55, 63	0
2	O	11/12 (91%)	-0.10	0 100 100	44, 49, 61, 64	0
2	P	11/12 (91%)	0.05	1 (9%) 9 7	42, 52, 74, 77	0
2	Q	11/12 (91%)	0.03	0 100 100	41, 50, 65, 77	0
2	R	11/12 (91%)	0.22	1 (9%) 9 7	54, 77, 87, 99	0
2	S	11/12 (91%)	0.17	0 100 100	45, 61, 73, 82	0
2	T	11/12 (91%)	0.33	0 100 100	42, 59, 81, 83	0
2	U	11/12 (91%)	0.12	1 (9%) 9 7	48, 64, 76, 90	0
2	V	11/12 (91%)	0.20	0 100 100	40, 57, 77, 78	0
2	W	11/12 (91%)	0.18	0 100 100	45, 60, 89, 95	0
2	X	11/12 (91%)	0.06	0 100 100	50, 64, 81, 99	0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	<RSRZ>	#RSRZ>2	OWAB(Å^2)	Q<0.9
All	All	2251/2292 (98%)	0.34	72 (3%) 47 44	31, 61, 104, 142	0

The worst 5 of 72 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	56	PRO	4.8
1	F	41	TYR	3.9
1	F	88	GLY	3.9
1	F	73	ALA	3.8
1	K	69	PHE	3.5

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.