wwPDB X-ray Structure Validation Summary Report (i)

May 26, 2020 – 06:19 am BST

PDB ID	:	5OSH
Title	:	Structure of retromer VPS29-VPS35C subunits complexed with RidL N-terminal domain (1-236)
Authors	:	Romano-Moreno, M.; Rojas, A.L.; Lucas, M.; Isupov, M.N.; Hierro, A.
Deposited on	:	2017-08-17
Resolution	:	4.30 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 following versions of software and data (see references (1)) were used in the production of this report:

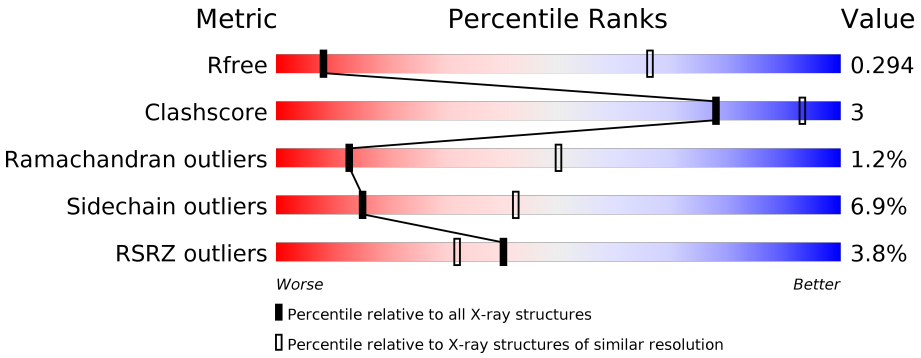
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 4.30 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.



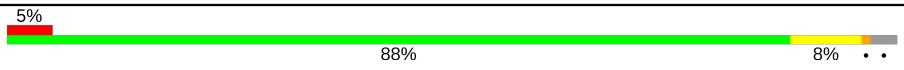
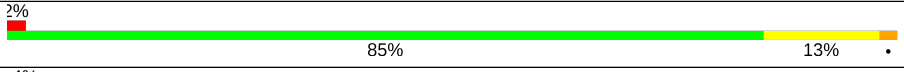
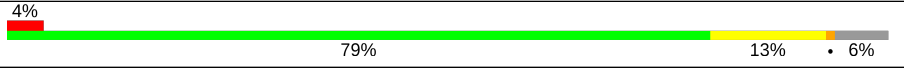
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1014 (4.80-3.80)
Clashscore	141614	1077 (4.80-3.80)
Ramachandran outliers	138981	1029 (4.80-3.80)
Sidechain outliers	138945	1012 (4.80-3.80)
RSRZ outliers	127900	1075 (4.90-3.70)

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 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	182	83% 16%
1	D	182	81% 19%
1	G	182	5% 85% 15%
1	J	182	10% 92% 8%
2	B	299	2% 85% 13%
2	E	299	86% 11%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	299	
2	K	299	
3	C	223	
3	F	223	
3	I	223	
3	L	223	

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 22276 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 Vacuolar protein sorting-associated protein 29.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	182	Total	C	N	O	S	0	0	0
			1447	936	242	263	6			
1	D	182	Total	C	N	O	S	0	0	0
			1447	936	242	263	6			
1	G	182	Total	C	N	O	S	0	0	0
			1447	936	242	263	6			
1	J	182	Total	C	N	O	S	0	0	0
			1447	936	242	263	6			

- Molecule 2 is a protein called Vacuolar protein sorting-associated protein 35.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	299	Total	C	N	O	S	0	0	0
			2424	1539	423	452	10			
2	E	294	Total	C	N	O	S	0	0	0
			2383	1514	418	441	10			
2	H	290	Total	C	N	O	S	0	0	0
			2344	1491	408	435	10			
2	K	293	Total	C	N	O	S	0	0	0
			2374	1509	417	438	10			

- Molecule 3 is a protein called Interaptin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	217	Total	C	N	O	S	0	0	0
			1750	1103	299	344	4			
3	F	214	Total	C	N	O	S	0	0	0
			1723	1085	294	340	4			
3	I	223	Total	C	N	O	S	0	0	0
			1802	1135	306	356	5			
3	L	209	Total	C	N	O	S	0	0	0
			1688	1062	288	334	4			

There are 8 discrepancies between the modelled and reference sequences:

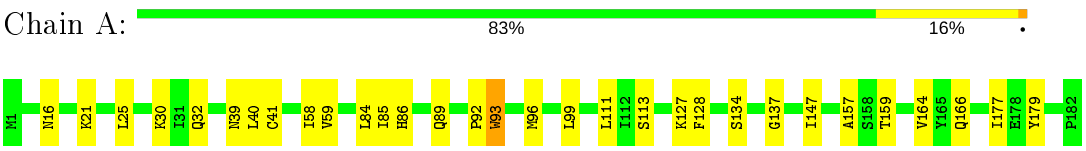
Chain	Residue	Modelled	Actual	Comment	Reference
C	1	MET	-	initiating methionine	UNP G8UZ99
C	2	ALA	-	expression tag	UNP G8UZ99
F	1	MET	-	initiating methionine	UNP G8UZ99
F	2	ALA	-	expression tag	UNP G8UZ99
I	1	MET	-	initiating methionine	UNP G8UZ99
I	2	ALA	-	expression tag	UNP G8UZ99
L	1	MET	-	initiating methionine	UNP G8UZ99
L	2	ALA	-	expression tag	UNP G8UZ99

3 Residue-property plots (i)

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.

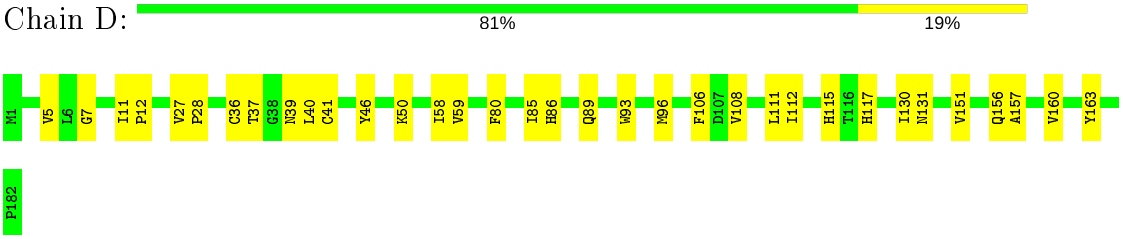
- Molecule 1: Vacuolar protein sorting-associated protein 29

Chain A:



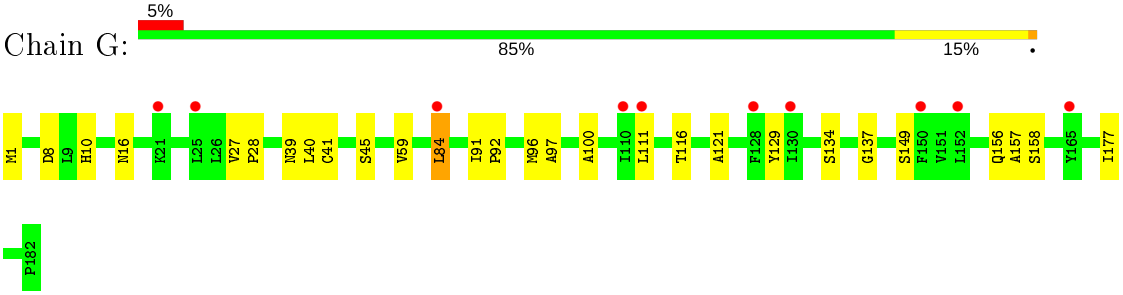
- Molecule 1: Vacuolar protein sorting-associated protein 29

Chain D:



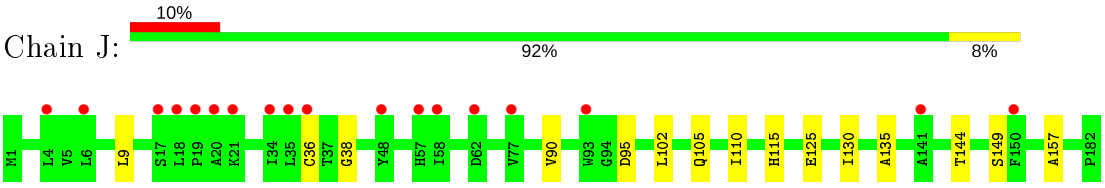
- Molecule 1: Vacuolar protein sorting-associated protein 29

Chain G:



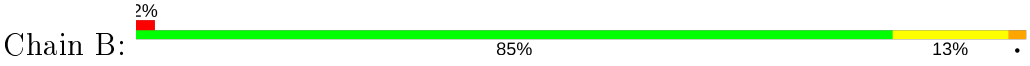
- Molecule 1: Vacuolar protein sorting-associated protein 29

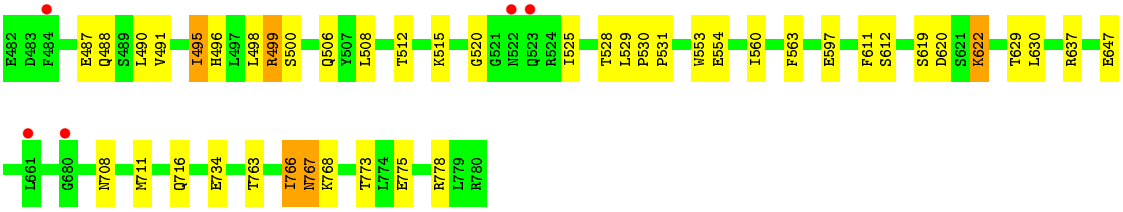
Chain J:



- Molecule 2: Vacuolar protein sorting-associated protein 35

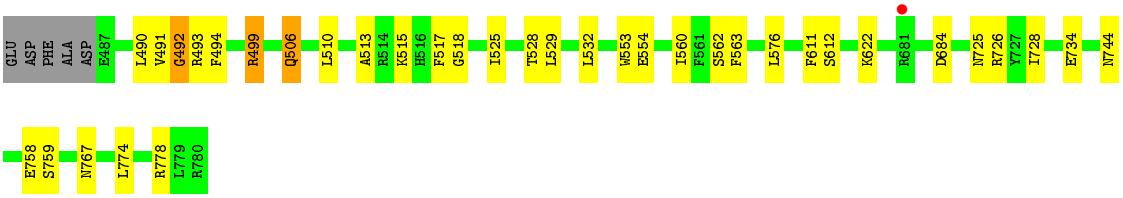
Chain B:





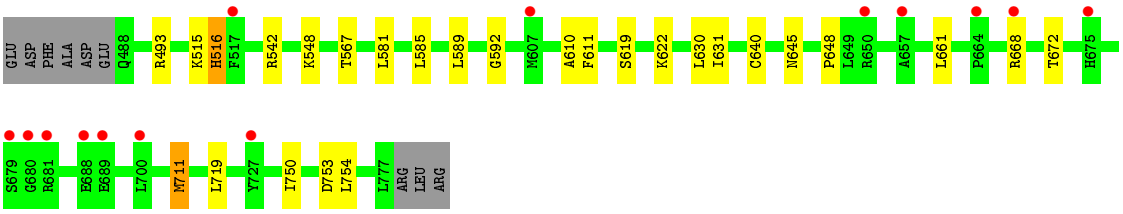
• Molecule 2: Vacuolar protein sorting-associated protein 35

Chain E:



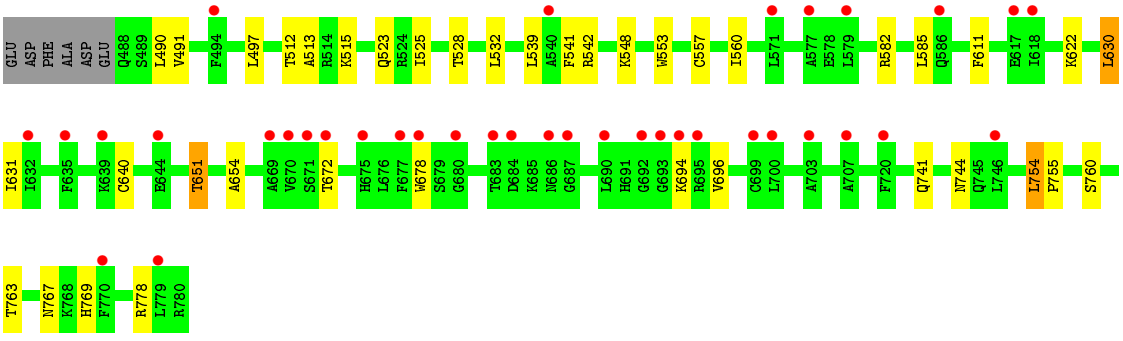
• Molecule 2: Vacuolar protein sorting-associated protein 35

Chain H:



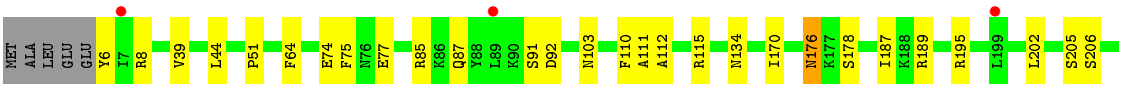
• Molecule 2: Vacuolar protein sorting-associated protein 35

Chain K:



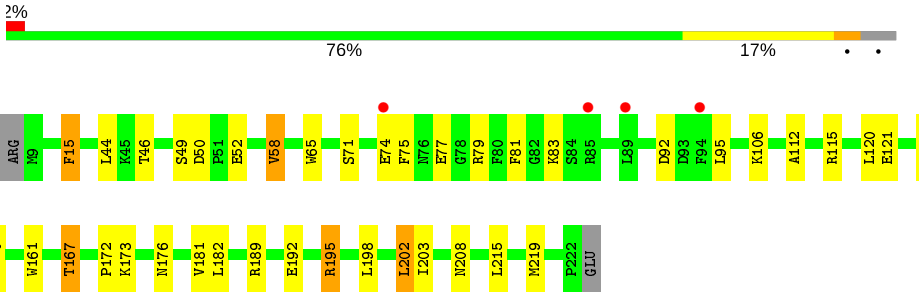
• Molecule 3: Interaptin

Chain C:



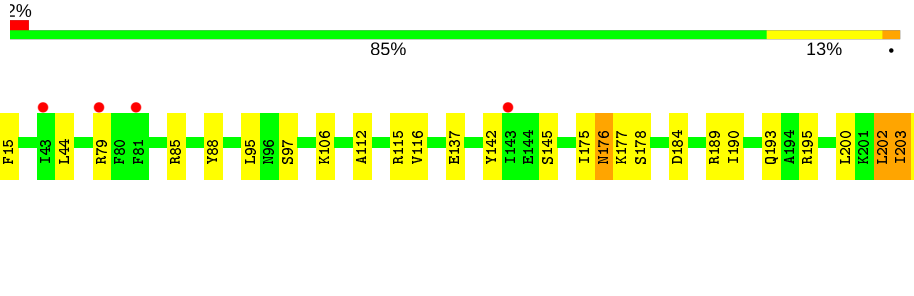
- Molecule 3: Interaptin

Chain F:



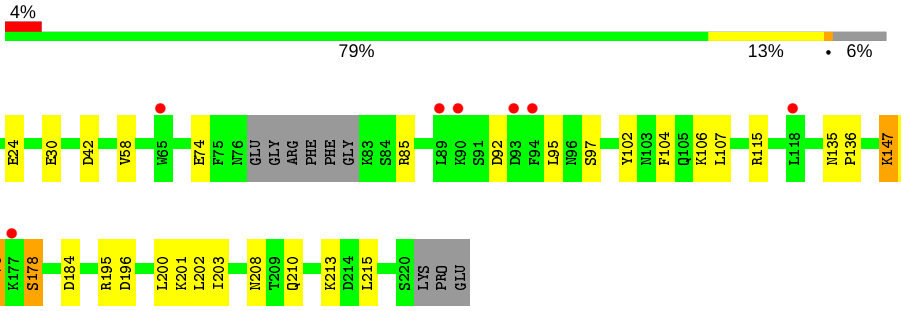
- Molecule 3: Interaptin

Chain I:



- Molecule 3: Interaptin

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	90.20Å 173.24Å 445.77Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	222.89 – 4.30 112.79 – 4.30	Depositor EDS
% Data completeness (in resolution range)	100.0 (222.89-4.30) 100.0 (112.79-4.30)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.24 (at 4.30Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.254 , 0.311 0.249 , 0.294	Depositor DCC
R_{free} test set	2460 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	216.4	Xtriage
Anisotropy	0.285	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 238.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.88	EDS
Total number of atoms	22276	wwPDB-VP
Average B, all atoms (Å^2)	266.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.58% 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	0.54	0/1481	0.78	0/2008
1	D	0.56	0/1481	0.77	0/2008
1	G	0.53	0/1481	0.71	0/2008
1	J	0.45	0/1481	0.63	0/2008
2	B	0.56	0/2471	0.69	0/3327
2	E	0.54	0/2429	0.69	0/3270
2	H	0.50	0/2390	0.64	0/3219
2	K	0.54	0/2420	0.64	0/3258
3	C	0.55	0/1783	0.71	0/2401
3	F	0.56	0/1755	0.71	0/2362
3	I	0.57	0/1835	0.70	0/2469
3	L	0.51	0/1717	0.65	0/2311
All	All	0.54	0/22724	0.69	0/30649

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

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	1447	0	1459	13	0
1	D	1447	0	1459	15	0
1	G	1447	0	1459	10	0
1	J	1447	0	1459	7	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	2424	0	2407	16	0
2	E	2383	0	2379	10	0
2	H	2344	0	2336	9	0
2	K	2374	0	2373	18	0
3	C	1750	0	1726	8	0
3	F	1723	0	1704	16	0
3	I	1802	0	1783	14	0
3	L	1688	0	1673	7	0
All	All	22276	0	22217	131	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 131 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:I:11:LYS:NZ	3:I:88:TYR:OH	2.19	0.75
3:L:176:ASN:ND2	3:L:178:SER:OG	2.26	0.69
1:A:134:SER:OG	1:A:137:GLY:N	2.30	0.64
1:G:27:VAL:HG22	1:G:28:PRO:HD2	1.84	0.60
3:L:58:VAL:HG21	3:L:95:LEU:HD11	1.83	0.59

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	180/182 (99%)	160 (89%)	17 (9%)	3 (2%)	9 43
1	D	180/182 (99%)	157 (87%)	21 (12%)	2 (1%)	14 52
1	G	180/182 (99%)	160 (89%)	18 (10%)	2 (1%)	14 52

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	J	180/182 (99%)	157 (87%)	20 (11%)	3 (2%)	9 43
2	B	297/299 (99%)	259 (87%)	34 (11%)	4 (1%)	12 48
2	E	292/299 (98%)	265 (91%)	25 (9%)	2 (1%)	22 62
2	H	288/299 (96%)	265 (92%)	20 (7%)	3 (1%)	15 54
2	K	291/299 (97%)	264 (91%)	25 (9%)	2 (1%)	22 62
3	C	215/223 (96%)	190 (88%)	24 (11%)	1 (0%)	29 68
3	F	212/223 (95%)	182 (86%)	26 (12%)	4 (2%)	8 41
3	I	221/223 (99%)	198 (90%)	19 (9%)	4 (2%)	8 42
3	L	205/223 (92%)	184 (90%)	19 (9%)	2 (1%)	15 54
All	All	2741/2816 (97%)	2441 (89%)	268 (10%)	32 (1%)	13 50

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	157	ALA
2	B	496	HIS
1	D	157	ALA
3	F	50	ASP
3	F	71	SER

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	160/160 (100%)	151 (94%)	9 (6%)	21 48
1	D	160/160 (100%)	157 (98%)	3 (2%)	57 75
1	G	160/160 (100%)	151 (94%)	9 (6%)	21 48
1	J	160/160 (100%)	158 (99%)	2 (1%)	69 82
2	B	260/260 (100%)	240 (92%)	20 (8%)	13 39
2	E	256/260 (98%)	237 (93%)	19 (7%)	13 40
2	H	252/260 (97%)	241 (96%)	11 (4%)	28 54

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	K	255/260 (98%)	237 (93%)	18 (7%)	14 41
3	C	192/198 (97%)	172 (90%)	20 (10%)	7 27
3	F	190/198 (96%)	169 (89%)	21 (11%)	6 25
3	I	198/198 (100%)	184 (93%)	14 (7%)	14 41
3	L	187/198 (94%)	166 (89%)	21 (11%)	6 25
All	All	2430/2472 (98%)	2263 (93%)	167 (7%)	15 42

5 of 167 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	F	58	VAL
1	G	16	ASN
3	L	155	GLU
3	F	75	PHE
3	F	126	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	598	ASN
3	F	176	ASN
3	L	134	ASN
3	F	68	HIS
3	F	208	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

There are no ligands in this entry.

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

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	182/182 (100%)	0.21	0 100 100	124, 191, 258, 292	0
1	D	182/182 (100%)	0.07	0 100 100	126, 182, 255, 298	0
1	G	182/182 (100%)	0.29	10 (5%) 25 22	185, 263, 320, 362	0
1	J	182/182 (100%)	0.33	18 (9%) 7 7	219, 314, 372, 462	0
2	B	299/299 (100%)	0.21	5 (1%) 70 61	140, 230, 338, 449	0
2	E	294/299 (98%)	0.13	1 (0%) 94 90	121, 229, 313, 387	0
2	H	290/299 (96%)	0.29	14 (4%) 30 26	165, 275, 377, 435	0
2	K	293/299 (97%)	0.62	37 (12%) 3 5	235, 343, 441, 539	0
3	C	217/223 (97%)	0.04	3 (1%) 75 66	132, 216, 314, 371	0
3	F	214/223 (95%)	0.17	4 (1%) 66 58	148, 226, 334, 438	0
3	I	223/223 (100%)	-0.00	4 (1%) 68 60	210, 307, 384, 453	0
3	L	209/223 (93%)	0.20	10 (4%) 30 26	209, 315, 401, 451	0
All	All	2767/2816 (98%)	0.22	106 (3%) 40 32	121, 261, 385, 539	0

The worst 5 of 106 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	679	SER	9.8
2	H	681	ARG	7.1
2	K	687	GLY	6.1
2	K	694	LYS	5.9
2	K	686	ASN	5.2

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 (i)

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.