wwPDB X-ray Structure Validation Summary Report (i)

Jun 20, 2024 – 04:25 pm BST

PDB ID	:	8PEL
Title	:	Structure of C. thermophilum RNA exosome core
Authors	:	Lazzaretti, D.; Holdermann, I.; Sprangers, R.
Deposited on	:	2023-06-14
Resolution	:	3.81 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 (i)) were used in the production of this report:

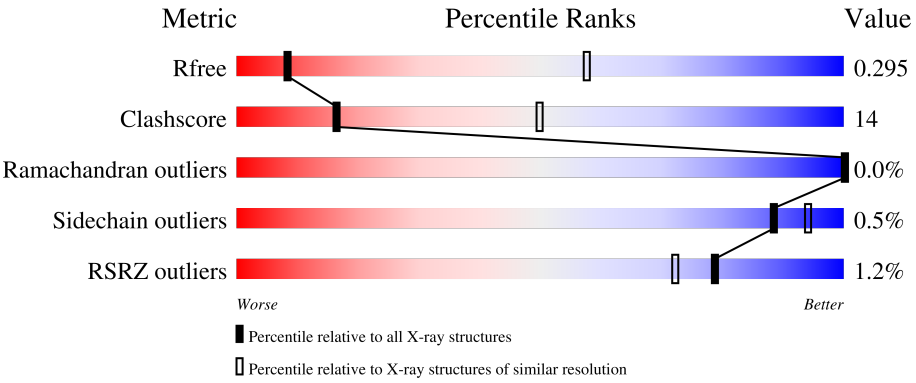
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.81 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	1231 (4.04-3.60)
Clashscore	141614	1031 (4.02-3.62)
Ramachandran outliers	138981	1261 (4.04-3.60)
Sidechain outliers	138945	1255 (4.04-3.60)
RSRZ outliers	127900	1139 (4.04-3.60)

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	293	2% 73% 23% .
2	B	284	57% 17% 26%
3	C	357	2% 62% 24% 13%
4	D	258	66% 24% 10%
5	E	413	47% 26% 26%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
6	F	284	2% 64% 31% 5%
7	G	256	% 64% 34%
8	H	358	% 54% 18% 27%
9	I	220	% 55% 25% 19%

2 Entry composition (i)

There are 9 unique types of molecules in this entry. The entry contains 35554 atoms, of which 17898 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 Rrp45.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	281	4376	1378	2195	375	417	11	4	0	0

- Molecule 2 is a protein called Exoribonuclease phosphorolytic domain-containing protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	211	3204	986	1610	290	307	11	4	0	0

- Molecule 3 is a protein called Exoribonuclease-like protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	C	311	4939	1560	2490	435	445	9	0	0	0

- Molecule 4 is a protein called Exoribonuclease phosphorolytic domain-containing protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
4	D	232	3540	1117	1765	310	342	6	0	0	0

- Molecule 5 is a protein called Exoribonuclease phosphorolytic domain-containing protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
5	E	304	4726	1499	2393	395	436	3	0	0	0

- Molecule 6 is a protein called Exosome complex component MTR3.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
6	F	269	4068	1264	2040	359	393	12	3	0	0

- Molecule 7 is a protein called Ribosomal RNA-processing protein 40.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
7	G	251	Total	C	H	N	O	S	0	0	0
			3897	1210	1971	353	355	8			

- Molecule 8 is a protein called Putative exosome complex protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
8	H	261	Total	C	H	N	O	S	0	0	0
			4013	1247	2020	362	376	8			

- Molecule 9 is a protein called Putative exosome 3'->5 protein.

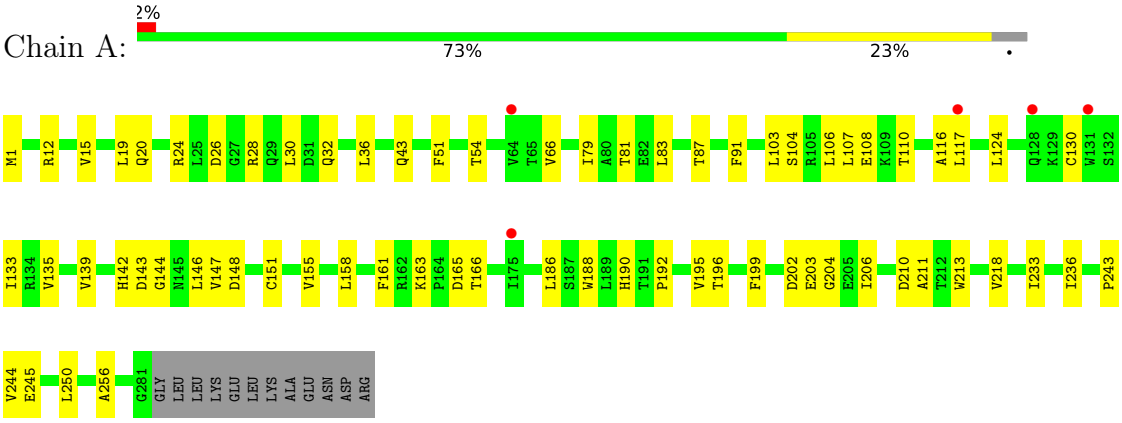
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
9	I	178	Total	C	H	N	O	S	0	0	0
			2791	872	1414	243	259	3			

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.

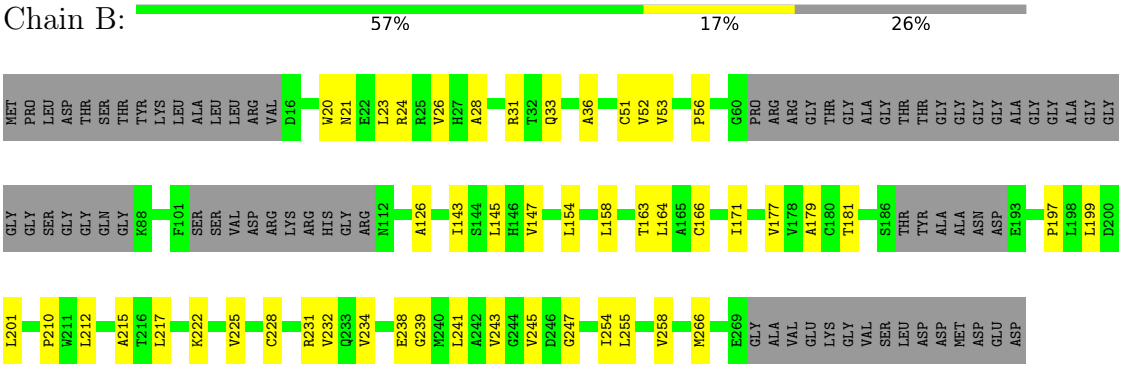
• Molecule 1: Rrp45

Chain A:



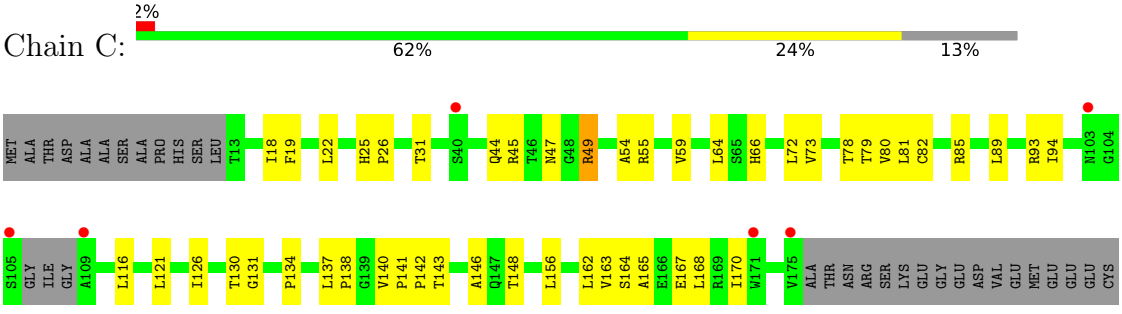
• Molecule 2: Exoribonuclease phosphorolytic domain-containing protein

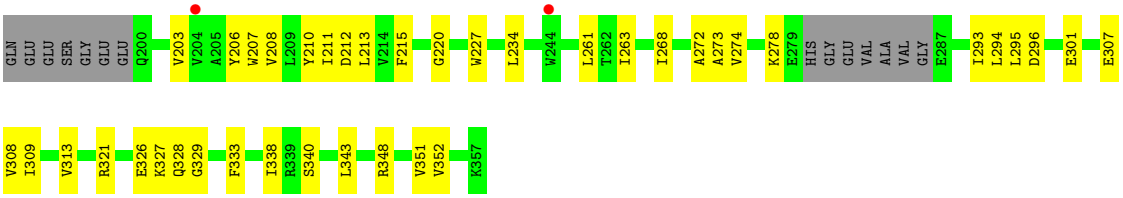
Chain B:

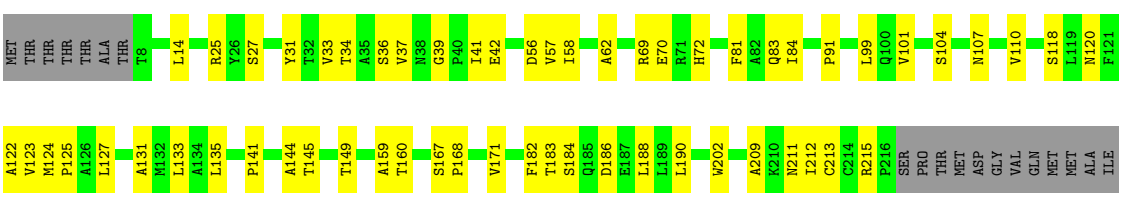


• Molecule 3: Exoribonuclease-like protein

Chain C:





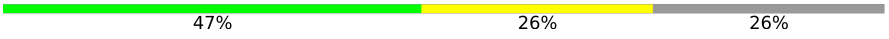


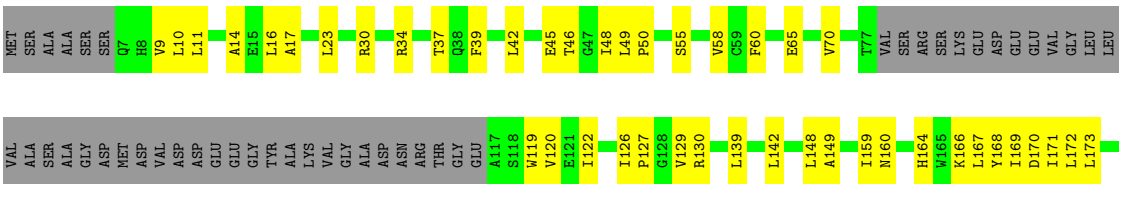
• Molecule 4: Exoribonuclease phosphorolytic domain-containing protein

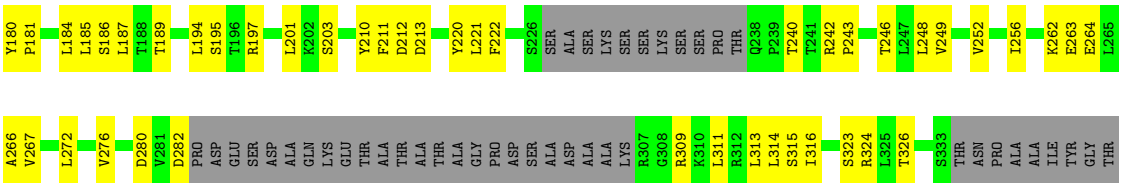
Chain D:

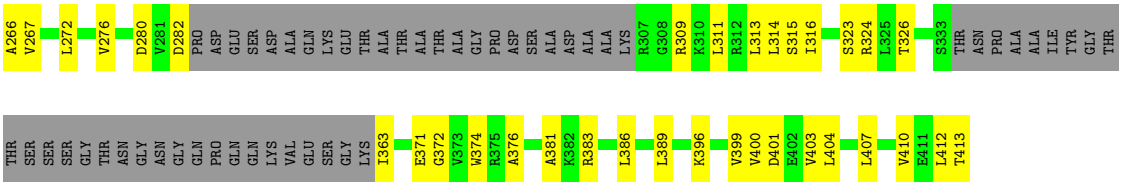
• Molecule 5: Exoribonuclease phosphorolytic domain-containing protein

Chain E:



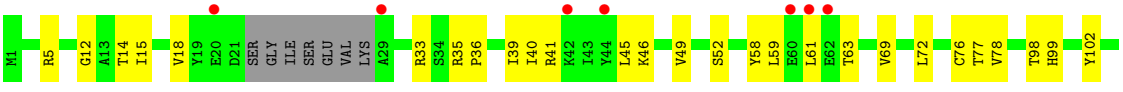


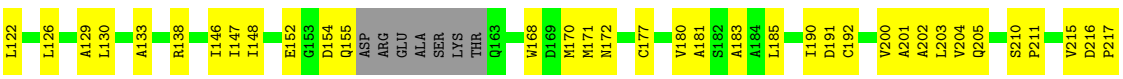




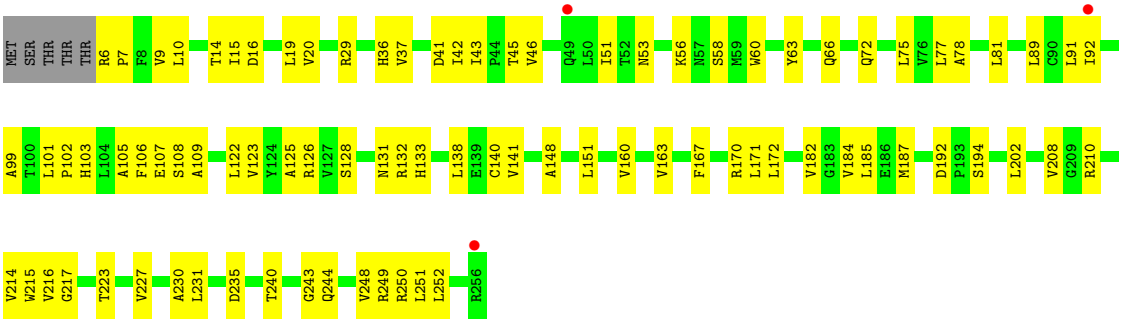
• Molecule 6: Exosome complex component MTR3

Chain F:

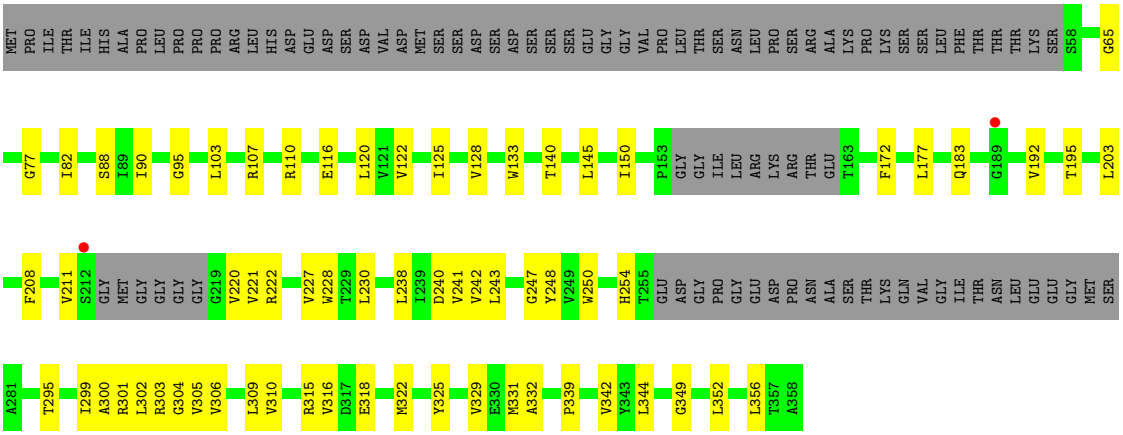




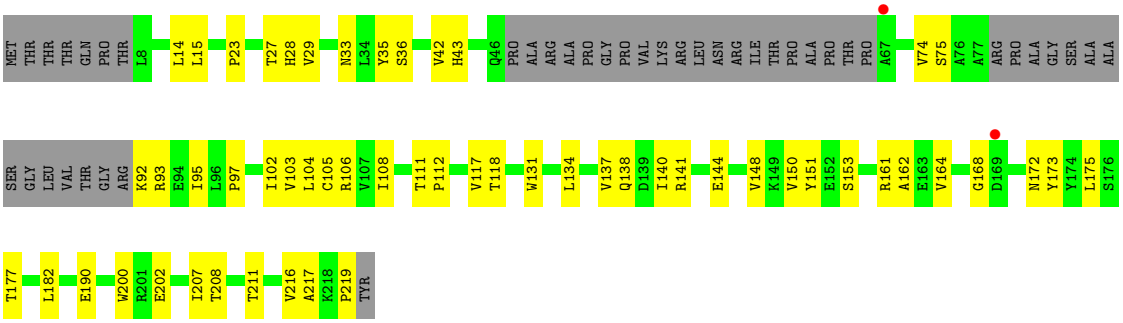
• Molecule 7: Ribosomal RNA-processing protein 40



• Molecule 8: Putative exosome complex protein



• Molecule 9: Putative exosome 3'->5 protein



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	100.53Å 148.38Å 195.06Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.90 – 3.81 48.90 – 3.81	Depositor EDS
% Data completeness (in resolution range)	99.6 (48.90-3.81) 99.5 (48.90-3.81)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.13 (at 3.77Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.243 , 0.292 0.247 , 0.295	Depositor DCC
R_{free} test set	1458 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	102.7	Xtriage
Anisotropy	0.610	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 60.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	35554	wwPDB-VP
Average B, all atoms (Å^2)	113.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.01% 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	0.28	0/2223	0.51	0/3021
2	B	0.27	0/1612	0.51	0/2182
3	C	0.29	0/2502	0.53	0/3399
4	D	0.27	0/1810	0.51	0/2462
5	E	0.31	0/2383	0.54	0/3253
6	F	0.28	0/2063	0.54	0/2804
7	G	0.30	0/1966	0.57	0/2673
8	H	0.31	0/2021	0.58	0/2734
9	I	0.29	0/1400	0.55	0/1906
All	All	0.29	0/17980	0.54	0/24434

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	2181	2195	2195	56	0
2	B	1594	1610	1606	42	0
3	C	2449	2490	2486	71	0
4	D	1775	1765	1764	43	0
5	E	2333	2393	2391	85	0
6	F	2028	2040	2038	83	0
7	G	1926	1971	1970	74	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
8	H	1993	2020	2016	44	0
9	I	1377	1414	1412	46	0
All	All	17656	17898	17878	506	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 506 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)
6:F:185:LEU:HD22	6:F:190:ILE:HG21	1.49	0.95
2:B:28:ALA:HB3	2:B:258:VAL:HG21	1.53	0.91
7:G:14:THR:HG22	7:G:43:ILE:HG12	1.59	0.84
9:I:14:LEU:HD11	9:I:33:ASN:HB3	1.60	0.84
3:C:274:VAL:HG12	3:C:293:ILE:HD13	1.57	0.84

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	279/293 (95%)	272 (98%)	7 (2%)	0	100	100
2	B	203/284 (72%)	200 (98%)	3 (2%)	0	100	100
3	C	303/357 (85%)	295 (97%)	8 (3%)	0	100	100
4	D	228/258 (88%)	224 (98%)	4 (2%)	0	100	100
5	E	294/413 (71%)	287 (98%)	6 (2%)	1 (0%)	41	74
6	F	263/284 (93%)	252 (96%)	11 (4%)	0	100	100
7	G	249/256 (97%)	242 (97%)	7 (3%)	0	100	100
8	H	253/358 (71%)	250 (99%)	3 (1%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
9	I	172/220 (78%)	169 (98%)	3 (2%)	0	100	100
All	All	2244/2723 (82%)	2191 (98%)	52 (2%)	1 (0%)	100	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
5	E	181	PRO

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	243/253 (96%)	243 (100%)	0	100	100
2	B	175/222 (79%)	175 (100%)	0	100	100
3	C	269/303 (89%)	268 (100%)	1 (0%)	91	95
4	D	189/211 (90%)	187 (99%)	2 (1%)	73	85
5	E	257/339 (76%)	255 (99%)	2 (1%)	81	89
6	F	223/235 (95%)	221 (99%)	2 (1%)	78	88
7	G	212/217 (98%)	210 (99%)	2 (1%)	78	88
8	H	212/293 (72%)	212 (100%)	0	100	100
9	I	152/184 (83%)	151 (99%)	1 (1%)	84	91
All	All	1932/2257 (86%)	1922 (100%)	10 (0%)	88	94

5 of 10 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
7	G	103	HIS
7	G	108	SER
9	I	75	SER
5	E	60	PHE
5	E	282	ASP

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

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

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	281/293 (95%)	0.01	5 (1%) 68 61	85, 104, 129, 160	4 (1%)
2	B	211/284 (74%)	0.06	0 100 100	77, 97, 125, 147	3 (1%)
3	C	311/357 (87%)	0.14	8 (2%) 56 46	74, 99, 132, 158	0
4	D	232/258 (89%)	-0.08	0 100 100	80, 97, 123, 138	0
5	E	304/413 (73%)	-0.01	0 100 100	72, 97, 120, 138	0
6	F	269/284 (94%)	0.22	7 (2%) 56 46	78, 98, 134, 146	1 (0%)
7	G	251/256 (98%)	0.19	3 (1%) 79 72	87, 107, 135, 152	0
8	H	261/358 (72%)	0.12	2 (0%) 86 81	74, 97, 119, 141	0
9	I	178/220 (80%)	0.32	2 (1%) 80 74	88, 115, 139, 146	0
All	All	2298/2723 (84%)	0.10	27 (1%) 79 72	72, 101, 130, 160	8 (0%)

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
8	H	212	SER	3.4
6	F	20	GLU	3.3
1	A	128	GLN	3.1
6	F	29	ALA	2.8
1	A	131	TRP	2.7

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

There are no monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers

There are no such residues in this entry.