wwPDB X-ray Structure Validation Summary Report (i)

Oct 22, 2024 – 12:17 PM JST

PDB ID	:	8XIE
Title	:	Archaeal exosome complex (Rrp4-Rrp41-Rrp42)
Authors	:	Kim, H.S.; Park, S.H.; Kim, S.H.; Hwang, K.Y.
Deposited on	:	2023-12-19
Resolution	:	3.50 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:

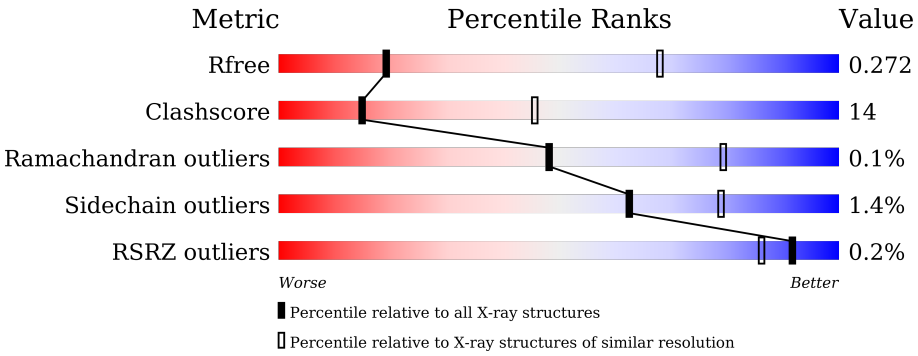
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 3.50 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.



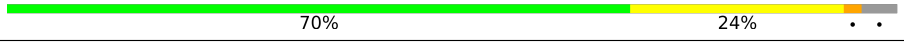
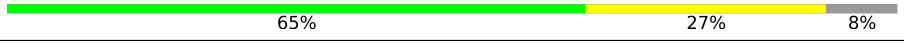
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	1094 (3.56-3.44)
Clashscore	180529	1045 (3.54-3.46)
Ramachandran outliers	177936	1032 (3.54-3.46)
Sidechain outliers	177891	1033 (3.54-3.46)
RSRZ outliers	164620	1093 (3.56-3.44)

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	248	67% 26% • •
1	C	248	62% 30% • 7%
1	D	248	69% 23% • 7%
2	B	260	66% 25% • 8%
2	E	260	% 59% 32% • 8%
2	F	260	63% 26% • 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	237	
3	H	237	
3	I	237	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 16183 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 Exosome complex component Rrp41.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	237	Total	C	N	O	S	0	0	0
			1834	1145	317	359	13			
1	C	231	Total	C	N	O	S	0	0	0
			1788	1117	310	348	13			
1	D	230	Total	C	N	O	S	0	0	0
			1779	1111	308	347	13			

- Molecule 2 is a protein called Exosome complex component Rrp42.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	238	Total	C	N	O	S	0	0	0
			1811	1146	308	351	6			
2	E	239	Total	C	N	O	S	0	0	0
			1820	1152	310	352	6			
2	F	239	Total	C	N	O	S	0	0	0
			1820	1152	310	352	6			

- Molecule 3 is a protein called Exosome complex component Rrp4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	G	227	Total	C	N	O	S	0	0	0
			1755	1138	285	321	11			
3	I	227	Total	C	N	O	S	0	0	0
			1752	1137	285	319	11			
3	H	219	Total	C	N	O	S	0	0	0
			1690	1096	275	309	10			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
G	0	ALA	-	expression tag	UNP Q9HIP3
I	0	ALA	-	expression tag	UNP Q9HIP3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
H	0	ALA	-	expression tag	UNP Q9HIP3

- Molecule 4 is water.

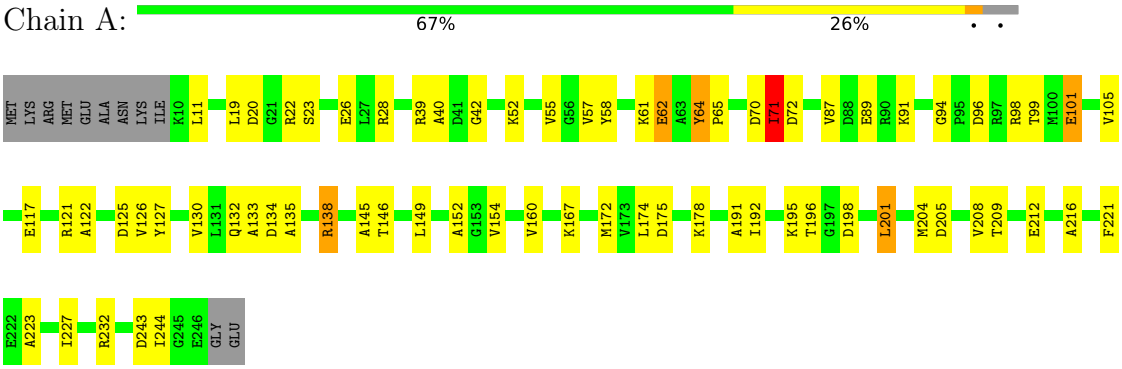
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	21	Total O 21 21	0	0
4	B	23	Total O 23 23	0	0
4	C	20	Total O 20 20	0	0
4	D	7	Total O 7 7	0	0
4	E	15	Total O 15 15	0	0
4	F	21	Total O 21 21	0	0
4	G	12	Total O 12 12	0	0
4	I	11	Total O 11 11	0	0
4	H	4	Total O 4 4	0	0

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.

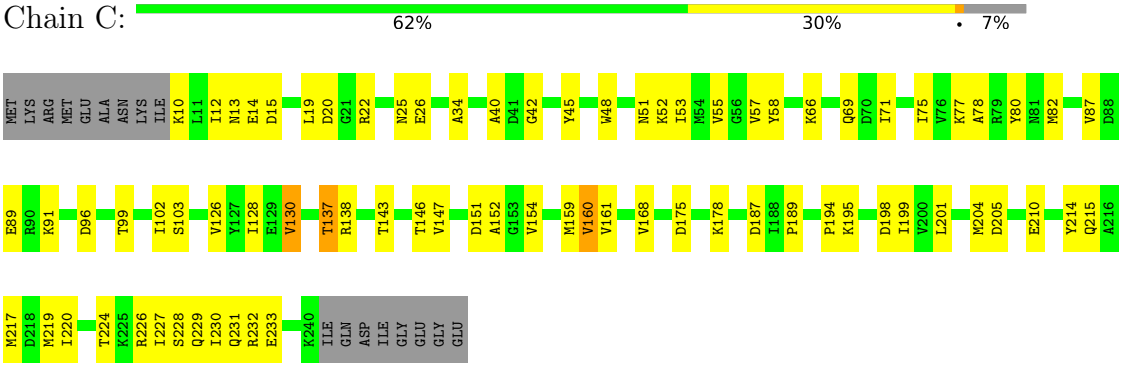
• Molecule 1: Exosome complex component Rrp41

Chain A:



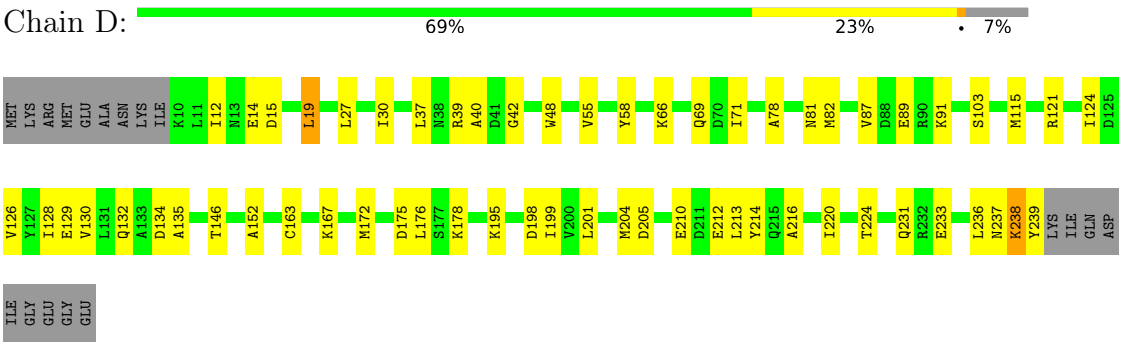
• Molecule 1: Exosome complex component Rrp41

Chain C:



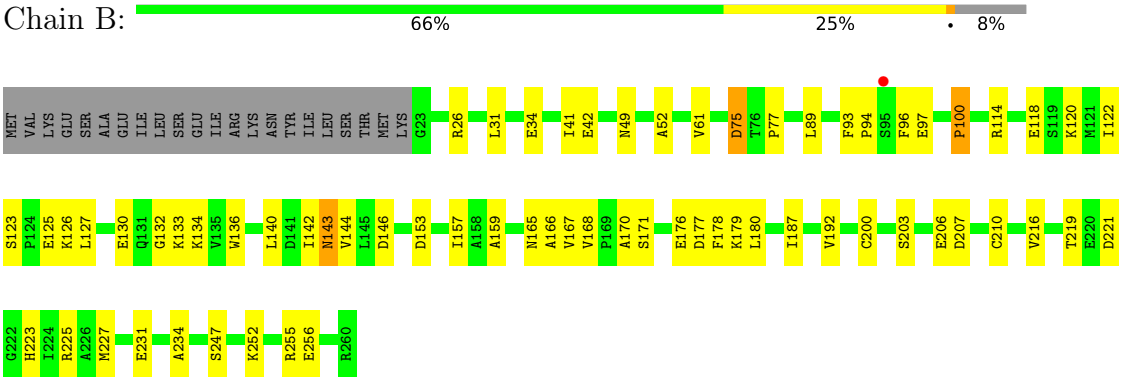
• Molecule 1: Exosome complex component Rrp41

Chain D:



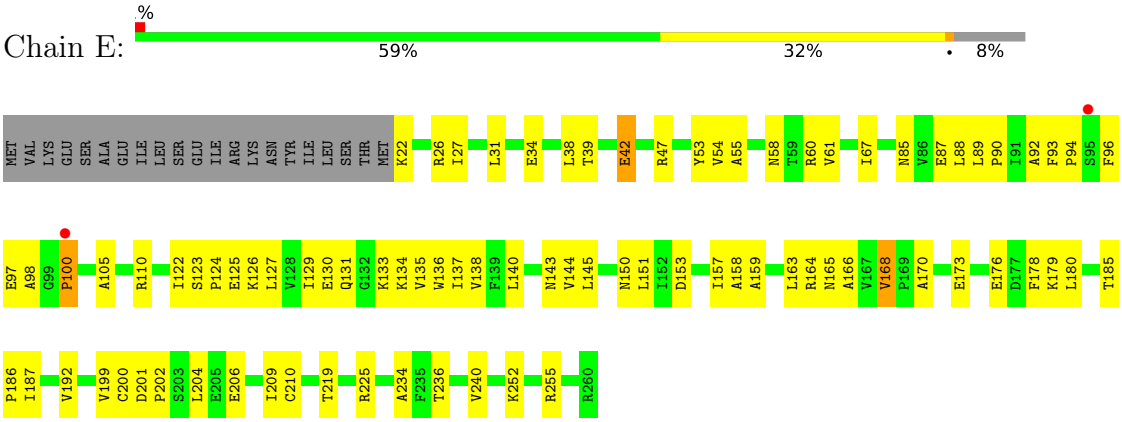
• Molecule 2: Exosome complex component Rrp42

Chain B:



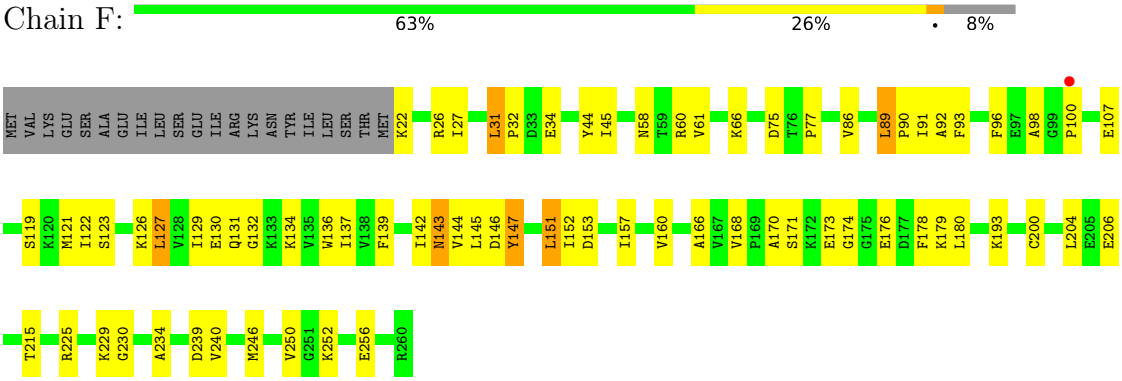
• Molecule 2: Exosome complex component Rrp42

Chain E:



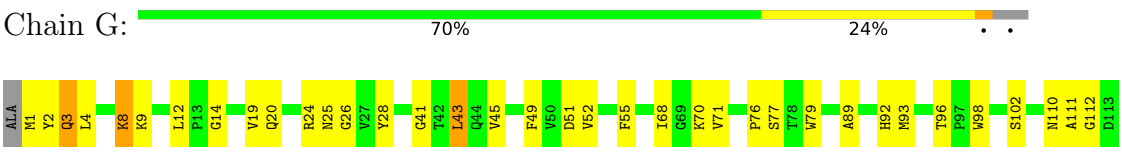
• Molecule 2: Exosome complex component Rrp42

Chain F:



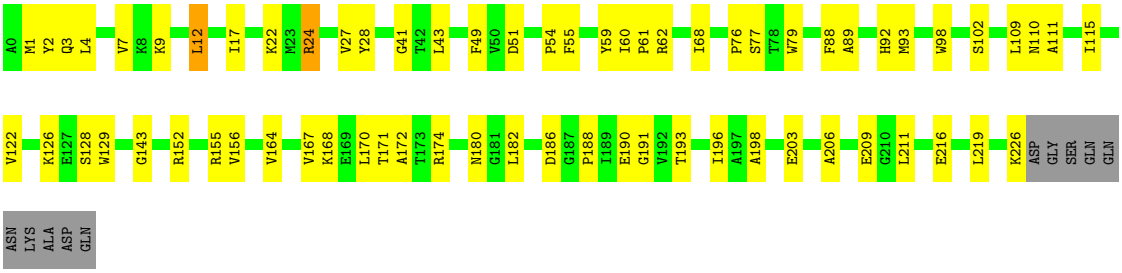
• Molecule 3: Exosome complex component Rrp4

Chain G:



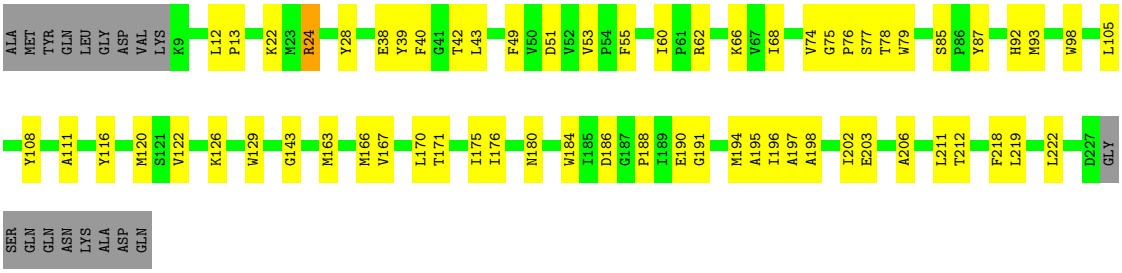
• Molecule 3: Exosome complex component Rrp4

Chain I:



• Molecule 3: Exosome complex component Rrp4

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	240.82Å 240.82Å 216.83Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	49.43 – 3.50 49.43 – 3.50	Depositor EDS
% Data completeness (in resolution range)	94.4 (49.43-3.50) 94.5 (49.43-3.50)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.96 (at 3.48Å)	Xtriage
Refinement program	PHENIX (1.20_4459: ???)	Depositor
R, R_{free}	0.229 , 0.273 0.231 , 0.272	Depositor DCC
R_{free} test set	84215 reflections (2.30%)	wwPDB-VP
Wilson B-factor (Å^2)	63.2	Xtriage
Anisotropy	0.462	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 106.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.31$, $< L^2 > = 0.14$	Xtriage
Estimated twinning fraction	0.178 for -h,-k,l	Xtriage
F_o,F_c correlation	0.87	EDS
Total number of atoms	16183	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.96% 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	0.88	2/1859 (0.1%)	1.12	6/2506 (0.2%)
1	C	0.87	1/1813 (0.1%)	1.01	5/2444 (0.2%)
1	D	0.78	0/1804	0.99	4/2433 (0.2%)
2	B	0.85	2/1839 (0.1%)	1.00	0/2491
2	E	0.74	2/1848 (0.1%)	0.94	3/2502 (0.1%)
2	F	0.82	2/1848 (0.1%)	0.95	5/2502 (0.2%)
3	G	0.68	0/1788	0.91	2/2403 (0.1%)
3	H	0.68	2/1722 (0.1%)	0.84	1/2315 (0.0%)
3	I	0.70	0/1785	0.91	2/2399 (0.1%)
All	All	0.78	11/16306 (0.1%)	0.97	28/21995 (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	C	0	2
2	B	0	1
2	E	0	2
2	F	0	2
3	H	0	1
All	All	0	10

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	101	GLU	CG-CD	7.13	1.62	1.51
2	F	107	GLU	CG-CD	6.57	1.61	1.51
2	E	168	VAL	CB-CG1	-6.42	1.39	1.52
2	E	42	GLU	CB-CG	5.85	1.63	1.52
2	F	147	TYR	CD2-CE2	5.71	1.48	1.39

The worst 5 of 28 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	71	ILE	CA-CB-CG1	-10.23	91.56	111.00
3	G	43	LEU	CB-CG-CD2	-8.78	96.08	111.00
1	D	201	LEU	CB-CG-CD1	-7.19	98.78	111.00
1	A	201	LEU	CB-CG-CD1	-7.10	98.94	111.00
2	F	151	LEU	CB-CG-CD2	-6.89	99.29	111.00

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	133	ALA	Peptide
1	A	71	ILE	Peptide
2	B	100	PRO	Peptide
1	C	160	VAL	Mainchain
1	C	226	ARG	Sidechain

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	1834	0	1837	60	0
1	C	1788	0	1794	53	1
1	D	1779	0	1781	45	0
2	B	1811	0	1852	45	0
2	E	1820	0	1865	60	0
2	F	1820	0	1865	56	0
3	G	1755	0	1756	53	0
3	H	1690	0	1687	47	0
3	I	1752	0	1757	64	1
4	A	21	0	0	5	0
4	B	23	0	0	1	0
4	C	20	0	0	7	0
4	D	7	0	0	0	0
4	E	15	0	0	1	0
4	F	21	0	0	3	0
4	G	12	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	H	4	0	0	0	0
4	I	11	0	0	1	0
All	All	16183	0	16194	437	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 14.

The worst 5 of 437 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)
2:E:166:ALA:HB1	2:E:180:LEU:HB3	1.35	1.06
1:A:98:ARG:HH12	1:A:138:ARG:HH12	1.04	0.98
2:B:166:ALA:HB1	2:B:180:LEU:HB3	1.49	0.95
2:F:166:ALA:HB1	2:F:180:LEU:HB3	1.51	0.92
3:H:198:ALA:HB2	3:H:219:LEU:HD21	1.54	0.90

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)
1:C:52:LYS:NZ	3:I:2:TYR:OH[3_655]	2.08	0.12

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	235/248 (95%)	228 (97%)	6 (3%)	1 (0%)	30 64
1	C	229/248 (92%)	227 (99%)	2 (1%)	0	100 100
1	D	228/248 (92%)	228 (100%)	0	0	100 100
2	B	236/260 (91%)	234 (99%)	2 (1%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	E	237/260 (91%)	234 (99%)	3 (1%)	0	100	100
2	F	237/260 (91%)	235 (99%)	2 (1%)	0	100	100
3	G	225/237 (95%)	222 (99%)	2 (1%)	1 (0%)	30	64
3	H	217/237 (92%)	216 (100%)	1 (0%)	0	100	100
3	I	225/237 (95%)	219 (97%)	6 (3%)	0	100	100
All	All	2069/2235 (93%)	2043 (99%)	24 (1%)	2 (0%)	48	79

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	64	TYR
3	G	3	GLN

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	195/204 (96%)	192 (98%)	3 (2%)	60	77
1	C	190/204 (93%)	186 (98%)	4 (2%)	48	71
1	D	189/204 (93%)	184 (97%)	5 (3%)	41	66
2	B	199/220 (90%)	194 (98%)	5 (2%)	42	67
2	E	200/220 (91%)	199 (100%)	1 (0%)	86	93
2	F	200/220 (91%)	197 (98%)	3 (2%)	60	77
3	G	192/199 (96%)	191 (100%)	1 (0%)	86	93
3	H	185/199 (93%)	184 (100%)	1 (0%)	86	93
3	I	191/199 (96%)	189 (99%)	2 (1%)	73	84
All	All	1741/1869 (93%)	1716 (99%)	25 (1%)	62	79

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	204	MET
2	E	130	GLU
3	H	24	ARG
1	D	238	LYS
2	F	130	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
3	I	94	ASN
3	G	3	GLN
2	E	131	GLN
1	D	231	GLN
2	E	143	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 oligosaccharides 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	237/248 (95%)	-1.51	0 100 100	19, 53, 130, 197	0
1	C	231/248 (93%)	-1.63	0 100 100	26, 55, 108, 199	0
1	D	230/248 (92%)	-1.59	0 100 100	27, 64, 123, 206	0
2	B	238/260 (91%)	-1.58	1 (0%) 89 79	23, 61, 124, 201	0
2	E	239/260 (91%)	-1.56	2 (0%) 82 67	31, 68, 136, 229	0
2	F	239/260 (91%)	-1.57	1 (0%) 89 79	25, 66, 133, 221	0
3	G	227/237 (95%)	-1.40	0 100 100	50, 105, 165, 216	0
3	H	219/237 (92%)	-0.90	0 100 100	71, 155, 223, 260	0
3	I	227/237 (95%)	-1.45	0 100 100	46, 99, 147, 182	0
All	All	2087/2235 (93%)	-1.47	4 (0%) 92 86	19, 75, 178, 260	0

All (4) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	100	PRO	3.6
2	E	95	SER	2.7
2	E	100	PRO	2.1
2	B	95	SER	2.0

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.