wwPDB X-ray Structure Validation Summary Report (i)

Aug 20, 2023 – 09:44 AM EDT

PDB ID	:	2HZS
Title	:	Structure of the Mediator head submodule Med8C/18/20
Authors	:	Lariviere, L.; Geiger, S.; Hoeppner, S.; Rother, S.; Straesser, K.; Cramer, P.
Deposited on	:	2006-08-09
Resolution	:	2.70 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:

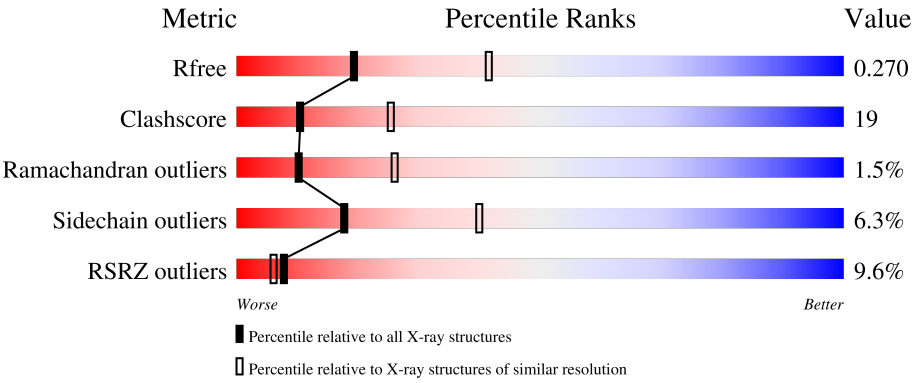
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.70 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.



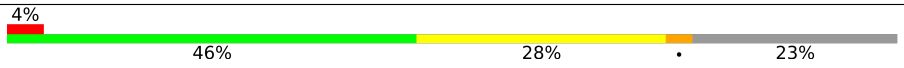
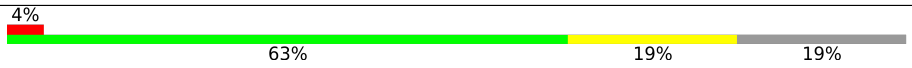
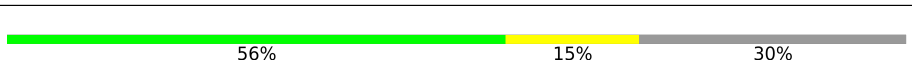
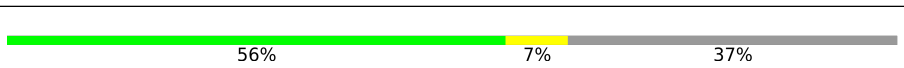
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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 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	209	16% 62% 30% 5%
1	C	209	13% 61% 32%
1	E	209	11% 62% 32%
1	G	209	19% 63% 30%
2	B	306	2% 50% 24% 23%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	306	
2	F	306	
2	H	306	
3	I	27	
3	J	27	
3	K	27	
3	L	27	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 14086 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 RNA polymerase II mediator complex subunit 20.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	203	Total	C	N	O	S	0	0	0
			1560	987	263	305	5			
1	C	202	Total	C	N	O	S	0	0	0
			1554	984	262	303	5			
1	E	203	Total	C	N	O	S	0	0	0
			1560	987	263	305	5			
1	G	203	Total	C	N	O	S	0	0	0
			1560	987	263	305	5			

- Molecule 2 is a protein called RNA polymerase II mediator complex subunit 18.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	235	Total	C	N	O	S	0	0	0
			1844	1175	302	358	9			
2	D	214	Total	C	N	O	S	0	0	0
			1671	1065	273	326	7			
2	F	235	Total	C	N	O	S	0	0	0
			1829	1156	298	367	8			
2	H	230	Total	C	N	O	S	0	0	0
			1803	1146	297	351	9			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	274	VAL	ALA	engineered mutation	UNP P32585
D	274	VAL	ALA	engineered mutation	UNP P32585
F	274	VAL	ALA	engineered mutation	UNP P32585
H	274	VAL	ALA	engineered mutation	UNP P32585

- Molecule 3 is a protein called RNA polymerase II mediator complex subunit 8.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	I	22	Total 192	C 123	N 37	O 32	0	0	0
3	J	19	Total 156	C 102	N 25	O 29	0	0	0
3	K	16	Total 131	C 87	N 19	O 25	0	0	0
3	L	17	Total 141	C 93	N 22	O 26	0	0	0

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
I	211	HIS	-	expression tag	UNP P38304
I	212	HIS	-	expression tag	UNP P38304
I	213	HIS	-	expression tag	UNP P38304
I	214	HIS	-	expression tag	UNP P38304
I	215	HIS	-	expression tag	UNP P38304
I	216	HIS	-	expression tag	UNP P38304
J	211	HIS	-	expression tag	UNP P38304
J	212	HIS	-	expression tag	UNP P38304
J	213	HIS	-	expression tag	UNP P38304
J	214	HIS	-	expression tag	UNP P38304
J	215	HIS	-	expression tag	UNP P38304
J	216	HIS	-	expression tag	UNP P38304
K	211	HIS	-	expression tag	UNP P38304
K	212	HIS	-	expression tag	UNP P38304
K	213	HIS	-	expression tag	UNP P38304
K	214	HIS	-	expression tag	UNP P38304
K	215	HIS	-	expression tag	UNP P38304
K	216	HIS	-	expression tag	UNP P38304
L	211	HIS	-	expression tag	UNP P38304
L	212	HIS	-	expression tag	UNP P38304
L	213	HIS	-	expression tag	UNP P38304
L	214	HIS	-	expression tag	UNP P38304
L	215	HIS	-	expression tag	UNP P38304
L	216	HIS	-	expression tag	UNP P38304

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	3	Total 3	O 3	0	0
4	B	34	Total 34	O 34	0	0

Continued on next page...

Continued from previous page...

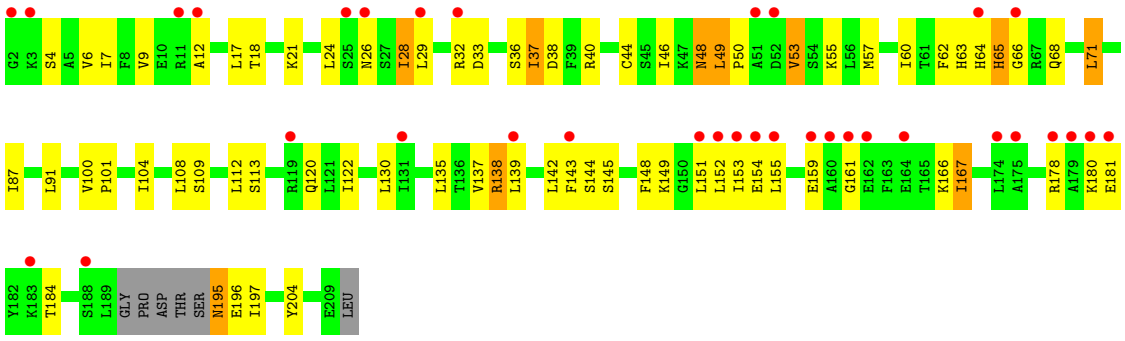
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	C	2	Total O 2 2	0	0
4	D	12	Total O 12 12	0	0
4	E	2	Total O 2 2	0	0
4	F	13	Total O 13 13	0	0
4	H	15	Total O 15 15	0	0
4	I	4	Total O 4 4	0	0

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.

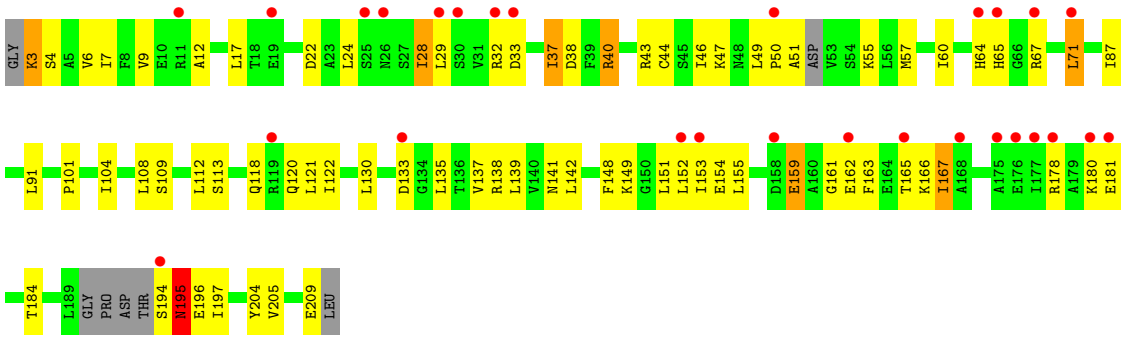
• Molecule 1: RNA polymerase II mediator complex subunit 20

Chain A:



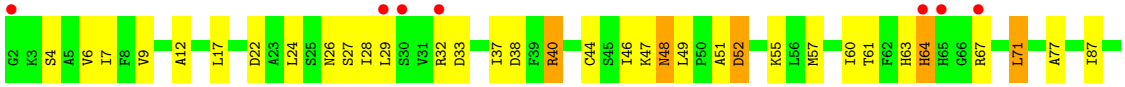
• Molecule 1: RNA polymerase II mediator complex subunit 20

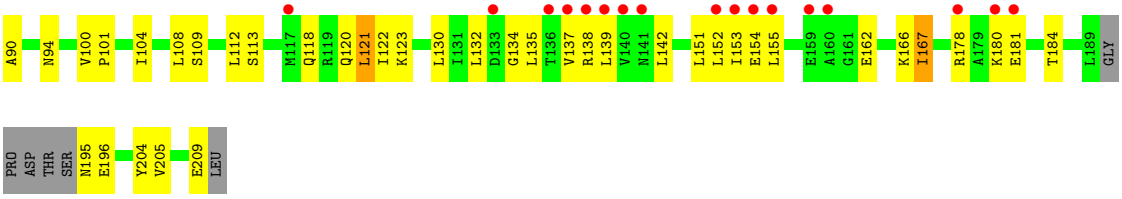
Chain C:



• Molecule 1: RNA polymerase II mediator complex subunit 20

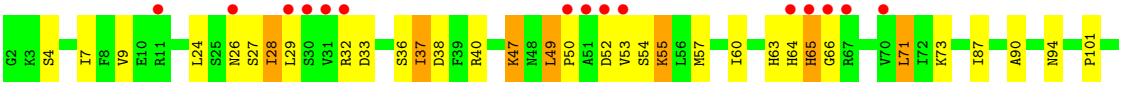
Chain E:

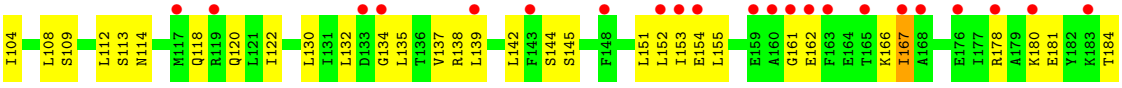


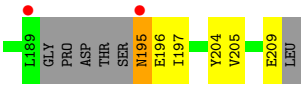


• Molecule 1: RNA polymerase II mediator complex subunit 20

Chain G:

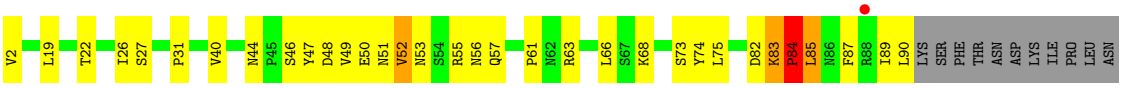


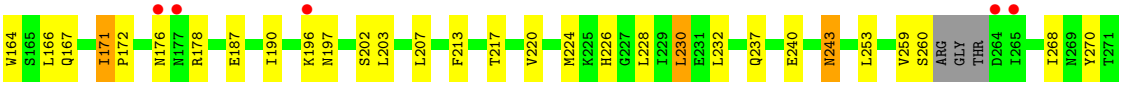


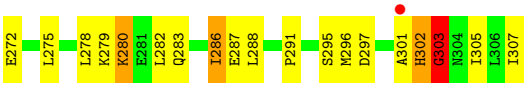


• Molecule 2: RNA polymerase II mediator complex subunit 18

Chain B:

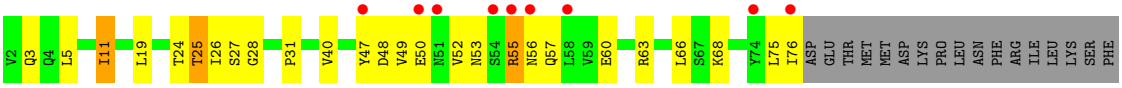


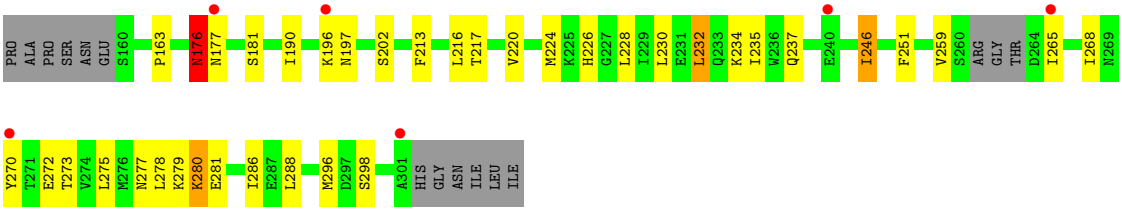




• Molecule 2: RNA polymerase II mediator complex subunit 18

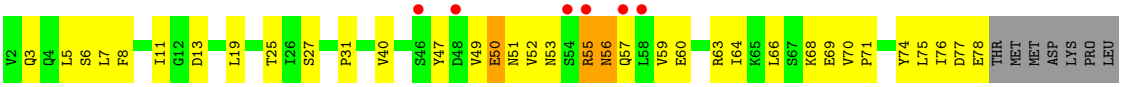
Chain D:

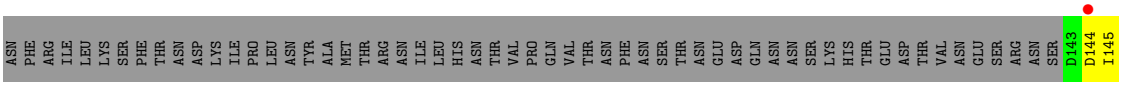


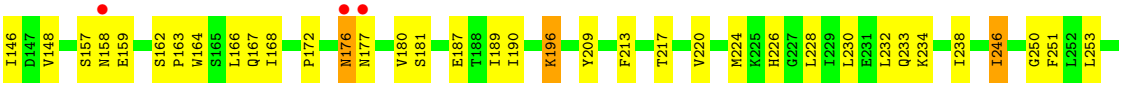


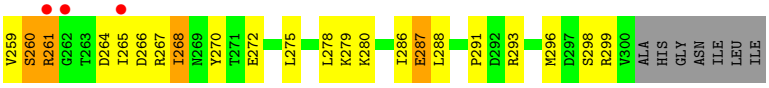
• Molecule 2: RNA polymerase II mediator complex subunit 18

Chain F:



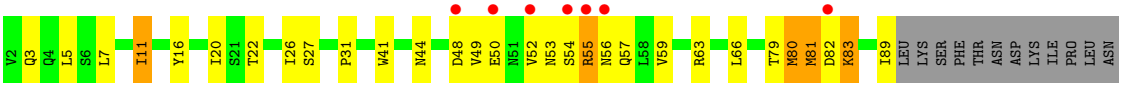


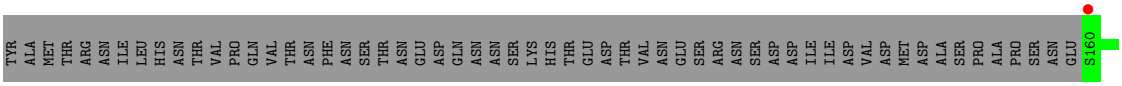


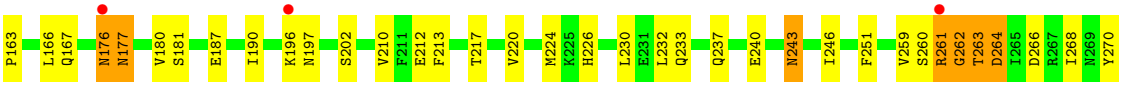


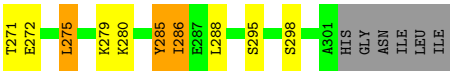
• Molecule 2: RNA polymerase II mediator complex subunit 18

Chain H:



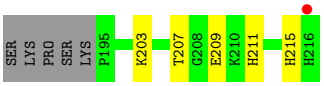






• Molecule 3: RNA polymerase II mediator complex subunit 8

Chain I:



• Molecule 3: RNA polymerase II mediator complex subunit 8

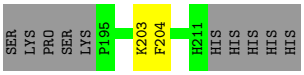
Chain J:

• Molecule 3: RNA polymerase II mediator complex subunit 8

Chain K:

• Molecule 3: RNA polymerase II mediator complex subunit 8

Chain L:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	75.36Å 115.76Å 129.16Å 90.00° 98.49° 90.00°	Depositor
Resolution (Å)	20.00 – 2.70 19.79 – 2.67	Depositor EDS
% Data completeness (in resolution range)	(Not available) (20.00-2.70) 99.1 (19.79-2.67)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.07	Depositor
$< I/\sigma(I) >$ [1]	2.26 (at 2.67Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.236 , 0.272 0.232 , 0.270	Depositor DCC
R_{free} test set	3117 reflections (5.07%)	wwPDB-VP
Wilson B-factor (Å^2)	57.6	Xtriage
Anisotropy	0.191	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 43.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	14086	wwPDB-VP
Average B, all atoms (Å^2)	64.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.72% 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	0.38	0/1584	0.87	8/2145 (0.4%)
1	C	0.38	0/1577	0.63	1/2134 (0.0%)
1	E	0.35	0/1584	0.64	1/2145 (0.0%)
1	G	0.36	0/1584	0.59	0/2145
2	B	0.46	0/1876	0.73	2/2541 (0.1%)
2	D	0.44	0/1700	0.68	1/2304 (0.0%)
2	F	0.45	0/1861	0.70	0/2526
2	H	0.45	0/1835	0.70	0/2486
3	I	0.45	0/201	0.60	0/269
3	J	0.49	0/160	0.59	0/214
3	K	0.55	0/134	0.65	0/179
3	L	0.52	0/145	0.59	0/194
All	All	0.42	0/14241	0.69	13/19282 (0.1%)

There are no bond length outliers.

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	138	ARG	NE-CZ-NH1	-14.18	113.21	120.30
1	A	40	ARG	NE-CZ-NH2	-13.12	113.74	120.30
1	A	138	ARG	NE-CZ-NH2	12.83	126.72	120.30
1	A	40	ARG	NE-CZ-NH1	9.33	124.97	120.30
2	B	303	GLY	N-CA-C	-7.56	94.19	113.10

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	1560	0	1576	70	0
1	C	1554	0	1573	66	0
1	E	1560	0	1576	61	0
1	G	1560	0	1576	52	0
2	B	1844	0	1851	79	0
2	D	1671	0	1676	53	1
2	F	1829	0	1811	87	1
2	H	1803	0	1809	67	0
3	I	192	0	168	5	0
3	J	156	0	150	5	0
3	K	131	0	126	2	0
3	L	141	0	133	2	0
4	A	3	0	0	0	0
4	B	34	0	0	2	0
4	C	2	0	0	0	0
4	D	12	0	0	0	0
4	E	2	0	0	0	0
4	F	13	0	0	1	0
4	H	15	0	0	0	0
4	I	4	0	0	0	0
All	All	14086	0	14025	522	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 19.

The worst 5 of 522 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:142:LEU:HD11	1:A:152:LEU:HD12	1.35	1.06
1:G:142:LEU:HD11	1:G:152:LEU:HD12	1.36	1.05
2:F:265:ILE:HD12	2:F:265:ILE:H	1.25	1.02
1:C:142:LEU:HD11	1:C:152:LEU:HD12	1.37	1.01
1:E:142:LEU:HD11	1:E:152:LEU:HD12	1.37	1.00

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:D:25:THR:OG1	2:F:25:THR:OG1[2_646]	2.06	0.14

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	199/209 (95%)	180 (90%)	17 (8%)	2 (1%)	15 37
1	C	196/209 (94%)	174 (89%)	18 (9%)	4 (2%)	7 19
1	E	199/209 (95%)	180 (90%)	18 (9%)	1 (0%)	29 54
1	G	199/209 (95%)	176 (88%)	21 (11%)	2 (1%)	15 37
2	B	229/306 (75%)	212 (93%)	12 (5%)	5 (2%)	6 17
2	D	208/306 (68%)	196 (94%)	10 (5%)	2 (1%)	15 37
2	F	231/306 (76%)	209 (90%)	19 (8%)	3 (1%)	12 30
2	H	226/306 (74%)	208 (92%)	10 (4%)	8 (4%)	3 8
3	I	20/27 (74%)	20 (100%)	0	0	100 100
3	J	17/27 (63%)	14 (82%)	3 (18%)	0	100 100
3	K	14/27 (52%)	13 (93%)	1 (7%)	0	100 100
3	L	15/27 (56%)	14 (93%)	1 (7%)	0	100 100
All	All	1753/2168 (81%)	1596 (91%)	130 (7%)	27 (2%)	10 26

5 of 27 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	83	LYS
2	B	84	PRO
2	F	56	ASN
2	F	261	ARG
2	H	56	ASN

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	172/177 (97%)	161 (94%)	11 (6%)	17	39
1	C	172/177 (97%)	163 (95%)	9 (5%)	23	49
1	E	172/177 (97%)	162 (94%)	10 (6%)	20	43
1	G	172/177 (97%)	160 (93%)	12 (7%)	15	35
2	B	213/280 (76%)	199 (93%)	14 (7%)	16	38
2	D	193/280 (69%)	179 (93%)	14 (7%)	14	33
2	F	212/280 (76%)	197 (93%)	15 (7%)	14	34
2	H	208/280 (74%)	193 (93%)	15 (7%)	14	34
3	I	21/26 (81%)	21 (100%)	0	100	100
3	J	18/26 (69%)	18 (100%)	0	100	100
3	K	15/26 (58%)	15 (100%)	0	100	100
3	L	16/26 (62%)	16 (100%)	0	100	100
All	All	1584/1932 (82%)	1484 (94%)	100 (6%)	18	40

5 of 100 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	196	GLU
2	F	280	LYS
2	H	286	ILE
2	F	55	ARG
2	F	217	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 51 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	173	HIS
2	F	226	HIS
2	H	269	ASN
1	E	195	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	F	176	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

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	203/209 (97%)	0.86	34 (16%) 1 1	35, 77, 120, 127	0
1	C	202/209 (96%)	0.79	28 (13%) 2 2	34, 81, 121, 127	0
1	E	203/209 (97%)	0.60	24 (11%) 4 3	34, 74, 117, 122	0
1	G	203/209 (97%)	0.81	39 (19%) 1 0	34, 79, 118, 122	0
2	B	235/306 (76%)	-0.11	7 (2%) 50 51	28, 45, 72, 111	0
2	D	214/306 (69%)	0.01	15 (7%) 16 14	29, 49, 100, 109	0
2	F	235/306 (76%)	-0.03	13 (5%) 25 24	29, 48, 91, 108	0
2	H	230/306 (75%)	-0.00	11 (4%) 30 28	30, 47, 92, 104	0
3	I	22/27 (81%)	0.10	1 (4%) 33 31	35, 44, 61, 82	0
3	J	19/27 (70%)	0.09	0 100 100	40, 50, 83, 93	0
3	K	16/27 (59%)	-0.30	0 100 100	37, 46, 64, 69	0
3	L	17/27 (62%)	-0.24	0 100 100	36, 46, 70, 85	0
All	All	1799/2168 (82%)	0.33	172 (9%) 8 6	28, 57, 115, 127	0

The worst 5 of 172 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	65	HIS	9.8
1	C	65	HIS	7.9
1	E	65	HIS	7.4
1	A	175	ALA	6.0
2	D	50	GLU	6.0

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 (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.