



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 29, 2020 – 06:44 am BST

PDB ID : 5CWS  
Title : Crystal structure of the intact *Chaetomium thermophilum* Nsp1-Nup49-Nup57 channel nucleoporin heterotrimer bound to its Nic96 nuclear pore complex attachment site  
Authors : Bley, C.J.; Petrovic, S.; Paduch, M.; Lu, V.; Kossiakoff, A.A.; Hoelz, A.  
Deposited on : 2015-07-28  
Resolution : 3.77 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](mailto: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 ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

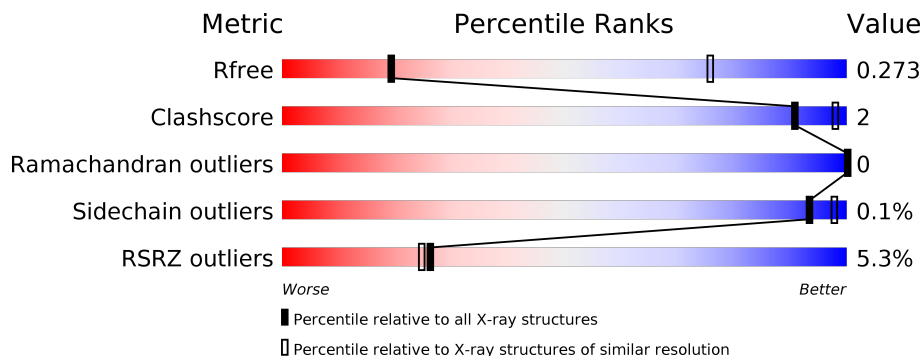
MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)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

# 1 Overall quality at a glance

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

The reported resolution of this entry is 3.77 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	1038 (3.96-3.60)
Clashscore	141614	1100 (3.96-3.60)
Ramachandran outliers	138981	1062 (3.96-3.60)
Sidechain outliers	138945	1058 (3.96-3.60)
RSRZ outliers	127900	1009 (3.98-3.58)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	237	
1	G	237	
2	B	266	
2	H	266	
3	C	208	
3	I	208	

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
4	D	227	<p>2% 76% 21%</p>
4	J	227	<p>75% 21%</p>
5	E	247	<p>2% 89% 9%</p>
5	K	247	<p>2% 90% 8%</p>
6	F	74	<p>50% 46%</p>
6	L	74	<p>51% 46%</p>

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 32898 atoms, of which 16334 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 sAB-158 Fab Light Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
1	A	208	3077	990	1503	260	319	5	0	4	0
1	G	208	3077	990	1503	260	319	5	0	4	0

- Molecule 2 is a protein called sAB-158 Fab Heavy Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
2	B	223	3240	1043	1588	278	326	5	0	0	0
2	H	223	3240	1043	1588	278	326	5	0	0	0

- Molecule 3 is a protein called Nucleoporin NSP1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
3	C	169	2736	855	1364	235	276	6	0	0	0
3	I	169	2736	855	1364	235	276	6	0	0	0

- Molecule 4 is a protein called Nucleoporin NUP49.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	H	N	O				S
4	D	180	2803	876	1402	236	282	7	0	0	0
4	J	180	2803	876	1402	236	282	7	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	244	MET	-	initiating methionine	UNP G0S4X2
D	245	SER	-	expression tag	UNP G0S4X2
J	244	MET	-	initiating methionine	UNP G0S4X2
J	245	SER	-	expression tag	UNP G0S4X2

- Molecule 5 is a protein called Nucleoporin NUP57.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
5	E	241	Total	C	H	N	O	S	0	0	0
			3947	1239	1976	360	368	4			
5	K	241	Total	C	H	N	O	S	0	0	0
			3947	1239	1976	360	368	4			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	70	MET	-	initiating methionine	UNP G0S0R2
K	70	MET	-	initiating methionine	UNP G0S0R2

- Molecule 6 is a protein called Nucleoporin NIC96.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	F	40	Total	C	H	N	O	0	0	0
			645	195	334	59	57			
6	L	40	Total	C	H	N	O	0	0	0
			645	195	334	59	57			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
F	140	SER	-	expression tag	UNP G0S024
L	140	SER	-	expression tag	UNP G0S024

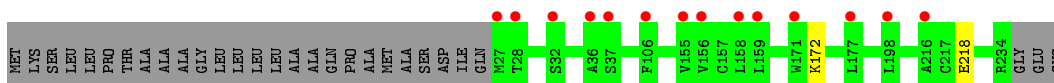
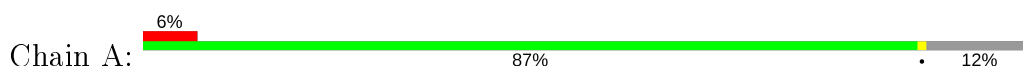
- Molecule 7 is OSMIUM ION (three-letter code: OS) (formula: Os).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	I	1	Total	Os	0	0
			1	1		
7	C	1	Total	Os	0	0
			1	1		

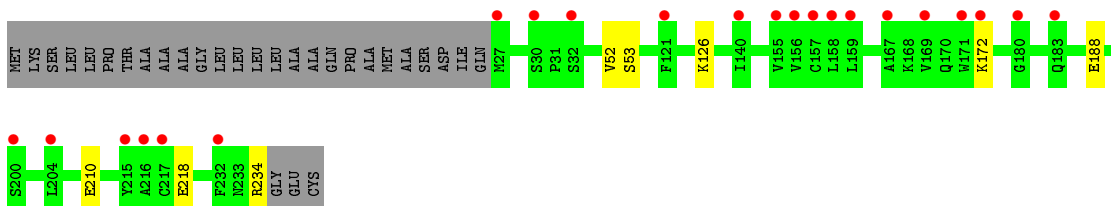
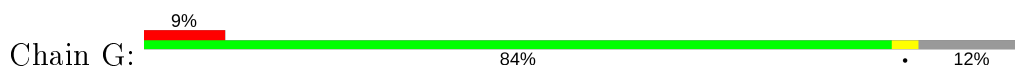
### 3 Residue-property plots [i](#)

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.

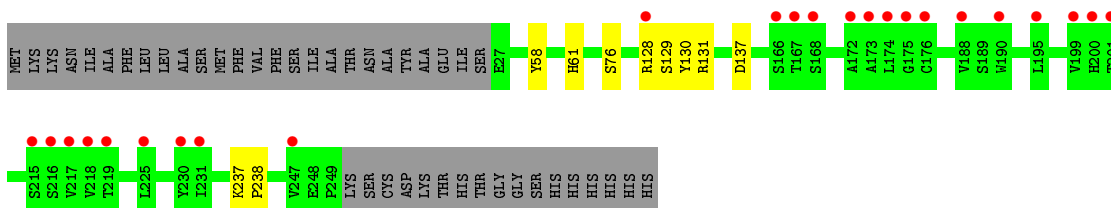
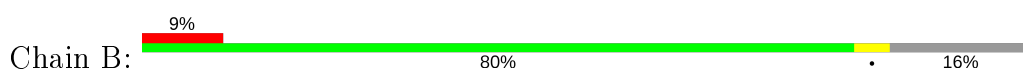
- Molecule 1: sAB-158 Fab Light Chain



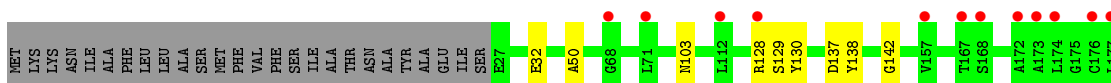
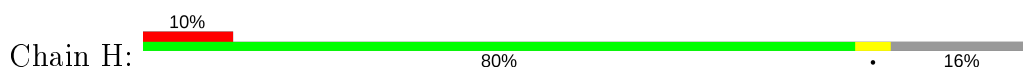
- Molecule 1: sAB-158 Fab Light Chain

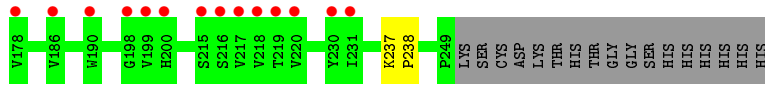


- Molecule 2: sAB-158 Fab Heavy Chain

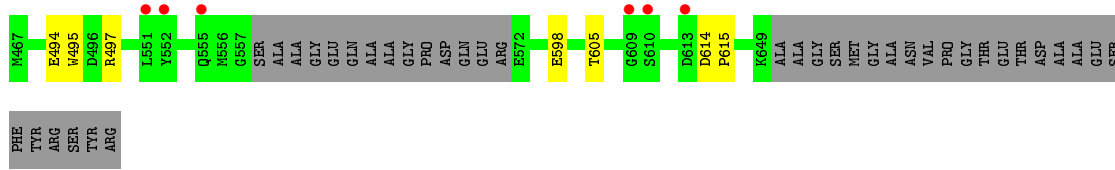
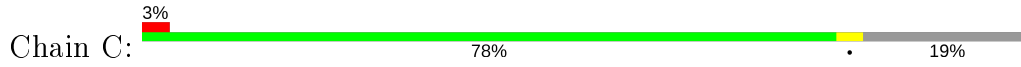


- Molecule 2: sAB-158 Fab Heavy Chain

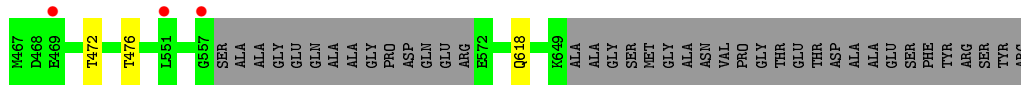
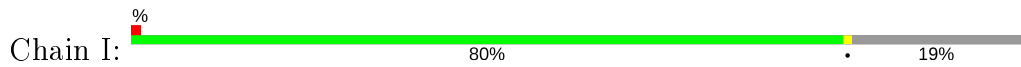




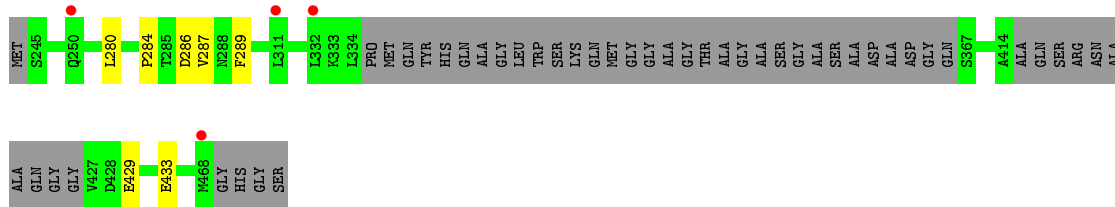
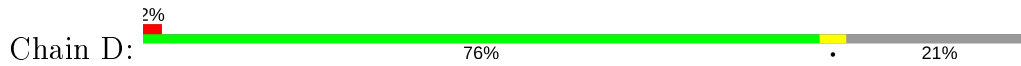
• Molecule 3: Nucleoporin NSP1



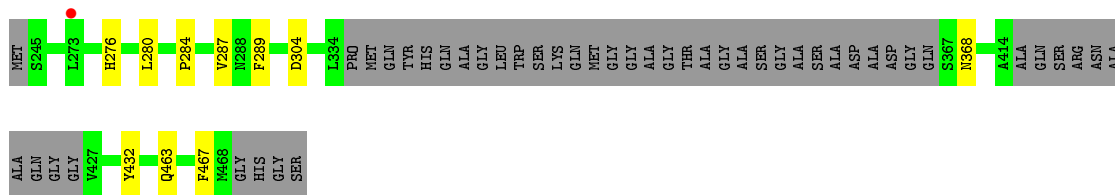
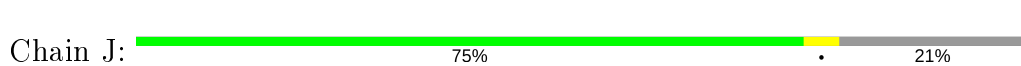
• Molecule 3: Nucleoporin NSP1



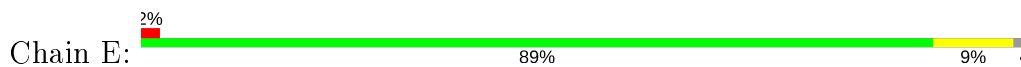
• Molecule 4: Nucleoporin NUP49



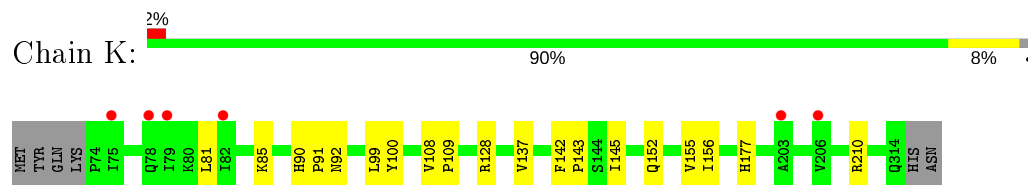
• Molecule 4: Nucleoporin NUP49



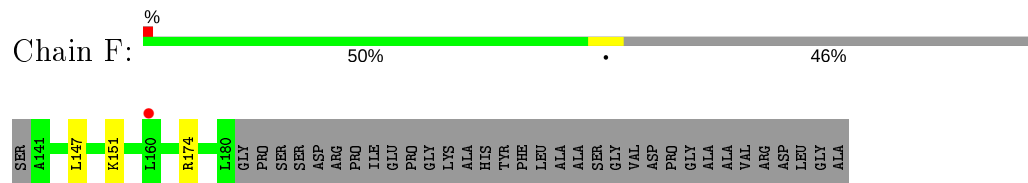
• Molecule 5: Nucleoporin NUP57



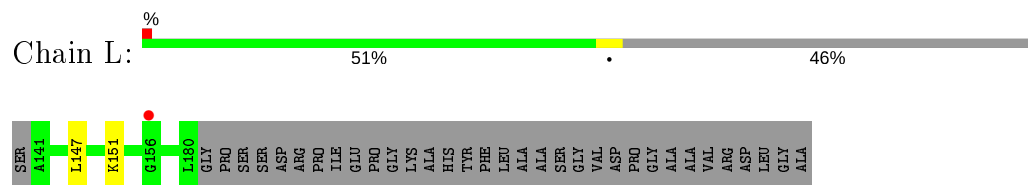
- Molecule 5: Nucleoporin NUP57



- Molecule 6: Nucleoporin NIC96



- Molecule 6: Nucleoporin NIC96





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	123.18Å 162.82Å 212.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.67 – 3.77 49.67 – 3.77	Depositor EDS
% Data completeness (in resolution range)	99.9 (49.67-3.77) 100.0 (49.67-3.77)	Depositor EDS
$R_{merge}$	0.20	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.11 (at 3.77Å)	Xtrriage
Refinement program	PHENIX 1.8.2_1309	Depositor
R, $R_{free}$	0.229 , 0.265 0.235 , 0.273	Depositor DCC
$R_{free}$ test set	1998 reflections (4.55%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	179.4	Xtrriage
Anisotropy	0.344	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 193.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.43$ , $\langle L^2 \rangle = 0.26$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	32898	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	264.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.42% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 0.2 for perfectly twinned datasets.

## 5 Model quality [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: OS

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	0.23	0/1621	0.39	0/2207
1	G	0.22	0/1621	0.39	0/2207
2	B	0.22	0/1694	0.40	0/2317
2	H	0.23	0/1694	0.39	0/2317
3	C	0.26	0/1388	0.39	0/1866
3	I	0.24	0/1388	0.38	0/1866
4	D	0.24	0/1413	0.38	0/1898
4	J	0.26	0/1413	0.40	0/1898
5	E	0.26	0/2011	0.41	0/2715
5	K	0.25	0/2011	0.40	0/2715
6	F	0.23	0/312	0.38	0/416
6	L	0.23	0/312	0.39	0/416
All	All	0.24	0/16878	0.40	0/22838

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	1574	1503	1488	1	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	G	1574	1503	1488	4	0
2	B	1652	1588	1584	7	0
2	H	1652	1588	1584	7	0
3	C	1372	1364	1363	5	0
3	I	1372	1364	1363	1	3
4	D	1401	1402	1398	6	0
4	J	1401	1402	1398	8	3
5	E	1971	1976	1968	14	0
5	K	1971	1976	1968	13	0
6	F	311	334	334	2	0
6	L	311	334	334	2	0
7	C	1	0	0	0	0
7	I	1	0	0	0	0
All	All	16564	16334	16270	57	3

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 2.

The worst 5 of 57 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:B:58:TYR:OH	5:E:171:ASP:OD2	2.10	0.58
1:G:172:LYS:NZ	1:G:218:GLU:OE1	2.33	0.58
3:C:598:GLU:OE1	5:E:254:ARG:NH1	2.40	0.55
4:J:368:ASN:OD1	5:K:210:ARG:NH2	2.40	0.54
1:G:210:GLU:O	1:G:234:ARG:NH1	2.41	0.54

All (3) 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)
3:I:472:THR:O	4:J:432:TYR:OH[4_455]	2.07	0.13
3:I:476:THR:HG1	4:J:432:TYR:OH[4_455]	1.54	0.06
3:I:472:THR:O	4:J:432:TYR:HH[4_455]	1.59	0.01

## 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	210/237 (89%)	197 (94%)	13 (6%)	0	100	100
1	G	210/237 (89%)	197 (94%)	13 (6%)	0	100	100
2	B	221/266 (83%)	204 (92%)	17 (8%)	0	100	100
2	H	221/266 (83%)	206 (93%)	15 (7%)	0	100	100
3	C	165/208 (79%)	158 (96%)	7 (4%)	0	100	100
3	I	165/208 (79%)	158 (96%)	7 (4%)	0	100	100
4	D	174/227 (77%)	172 (99%)	2 (1%)	0	100	100
4	J	174/227 (77%)	171 (98%)	3 (2%)	0	100	100
5	E	239/247 (97%)	212 (89%)	27 (11%)	0	100	100
5	K	239/247 (97%)	215 (90%)	24 (10%)	0	100	100
6	F	38/74 (51%)	31 (82%)	7 (18%)	0	100	100
6	L	38/74 (51%)	30 (79%)	8 (21%)	0	100	100
All	All	2094/2518 (83%)	1951 (93%)	143 (7%)	0	100	100

There are no Ramachandran outliers to report.

### 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	174/204 (85%)	174 (100%)	0	100	100
1	G	174/204 (85%)	174 (100%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	180/221 (81%)	179 (99%)	1 (1%)	86	92
2	H	180/221 (81%)	180 (100%)	0	100	100
3	C	146/171 (85%)	146 (100%)	0	100	100
3	I	146/171 (85%)	146 (100%)	0	100	100
4	D	150/176 (85%)	150 (100%)	0	100	100
4	J	150/176 (85%)	150 (100%)	0	100	100
5	E	213/219 (97%)	213 (100%)	0	100	100
5	K	213/219 (97%)	213 (100%)	0	100	100
6	F	32/55 (58%)	32 (100%)	0	100	100
6	L	32/55 (58%)	32 (100%)	0	100	100
All	All	1790/2092 (86%)	1789 (100%)	1 (0%)	93	98

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	137	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	200	HIS
5	K	164	HIS
5	K	296	HIS
5	K	300	GLN

### 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

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

## 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 [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	208/237 (87%)	0.30	14 (6%) 17 15	147, 258, 302, 346	0
1	G	208/237 (87%)	0.37	22 (10%) 6 5	157, 241, 289, 312	0
2	B	223/266 (83%)	0.40	24 (10%) 5 5	144, 226, 349, 417	0
2	H	223/266 (83%)	0.32	26 (11%) 4 5	152, 233, 333, 402	0
3	C	169/208 (81%)	-0.14	6 (3%) 42 38	134, 220, 318, 365	0
3	I	169/208 (81%)	-0.05	3 (1%) 68 65	146, 239, 350, 397	0
4	D	180/227 (79%)	-0.16	4 (2%) 62 57	131, 222, 313, 333	0
4	J	180/227 (79%)	-0.15	1 (0%) 89 88	140, 254, 332, 369	0
5	E	241/247 (97%)	-0.21	4 (1%) 70 66	121, 234, 344, 393	0
5	K	241/247 (97%)	-0.17	6 (2%) 57 52	121, 255, 341, 386	0
6	F	40/74 (54%)	0.12	1 (2%) 57 52	187, 231, 323, 332	0
6	L	40/74 (54%)	0.44	1 (2%) 57 52	183, 250, 342, 361	0
All	All	2122/2518 (84%)	0.07	112 (5%) 26 25	121, 239, 333, 417	0

The worst 5 of 112 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	216	SER	11.0
2	H	216	SER	9.5
2	B	217	VAL	7.7
1	G	27	MET	7.3
1	G	216	ALA	7.2

### 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 carbohydrates in this entry.

### 6.4 Ligands [i](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<sup>th</sup> percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
7	OS	C	701	1/1	0.84	0.28	296,296,296,296	0
7	OS	I	701	1/1	0.96	0.29	301,301,301,301	0

### 6.5 Other polymers [i](#)

There are no such residues in this entry.