



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 25, 2023 – 06:31 AM EDT

PDB ID : 5VVL  
Title : Cas1-Cas2 bound to full-site mimic with Ni  
Authors : Wright, A.V.; Knott, G.J.; Doxzen, K.D.; Doudna, J.A.  
Deposited on : 2017-05-19  
Resolution : 3.31 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35.1  
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) : 2.35.1

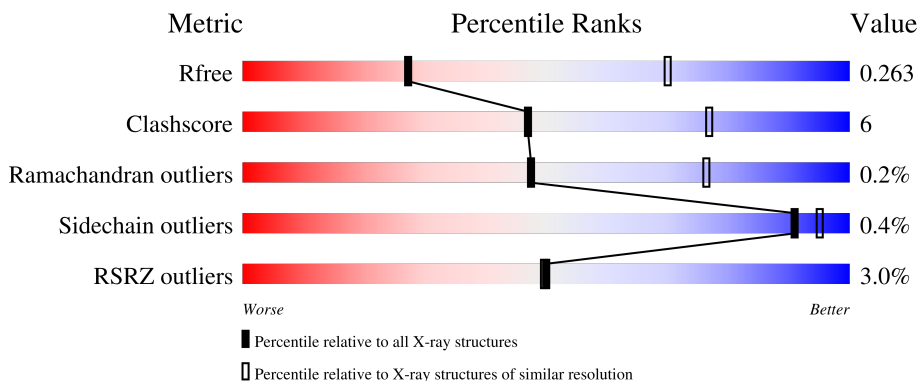
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.31 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.






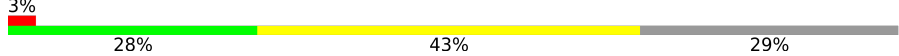
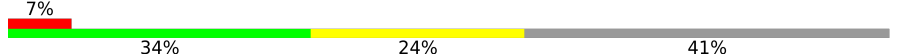
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1089 (3.36-3.28)
Clashscore	141614	1137 (3.36-3.28)
Ramachandran outliers	138981	1115 (3.36-3.28)
Sidechain outliers	138945	1114 (3.36-3.28)
RSRZ outliers	127900	1059 (3.36-3.28)

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  $\geq 3$ , 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	308	<div style="display: flex; align-items: center;"> <div style="width: 4%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 76%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 10%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 14%; height: 10px; background-color: grey;"></div> </div>
1	B	308	<div style="display: flex; align-items: center;"> <div style="width: 2%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 77%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 11%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 11%; height: 10px; background-color: grey;"></div> </div>
1	C	308	<div style="display: flex; align-items: center;"> <div style="width: 0%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 74%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 12%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 13%; height: 10px; background-color: grey;"></div> </div>
1	D	308	<div style="display: flex; align-items: center;"> <div style="width: 2%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 76%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 12%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 12%; height: 10px; background-color: grey;"></div> </div>
2	E	103	<div style="display: flex; align-items: center;"> <div style="width: 0%; height: 10px; background-color: red; margin-right: 5px;"></div> <div style="width: 76%; height: 10px; background-color: green; margin-right: 5px;"></div> <div style="width: 15%; height: 10px; background-color: yellow; margin-right: 5px;"></div> <div style="width: 9%; height: 10px; background-color: grey;"></div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	103	 <p>78% 15% 8%</p>
3	G	11	 <p>36% 55% 45%</p>
4	H	11	 <p>27% 36% 64%</p>
5	J	58	 <p>3% 28% 43% 29%</p>
6	K	58	 <p>7% 34% 24% 41%</p>

## 2 Entry composition

There are 7 unique types of molecules in this entry. The entry contains 11762 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 CRISPR-associated endonuclease Cas1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	264	Total 2027	C 1296	N 359	O 365	S 7	0	0	0
1	B	273	Total 2105	C 1346	N 374	O 378	S 7	0	0	0
1	C	267	Total 2052	C 1311	N 364	O 370	S 7	0	0	0
1	D	271	Total 2091	C 1338	N 370	O 376	S 7	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	SER	-	expression tag	UNP Q46896
A	-1	PHE	-	expression tag	UNP Q46896
A	0	THR	-	expression tag	UNP Q46896
B	-2	SER	-	expression tag	UNP Q46896
B	-1	PHE	-	expression tag	UNP Q46896
B	0	THR	-	expression tag	UNP Q46896
C	-2	SER	-	expression tag	UNP Q46896
C	-1	PHE	-	expression tag	UNP Q46896
C	0	THR	-	expression tag	UNP Q46896
D	-2	SER	-	expression tag	UNP Q46896
D	-1	PHE	-	expression tag	UNP Q46896
D	0	THR	-	expression tag	UNP Q46896

- Molecule 2 is a protein called CRISPR-associated endoribonuclease Cas2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	E	94	Total 739	C 475	N 128	O 132	S 4	0	0	0
2	F	95	Total 743	C 477	N 129	O 133	S 4	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	95	GLY	-	expression tag	UNP P45956
E	96	SER	-	expression tag	UNP P45956
E	97	SER	-	expression tag	UNP P45956
E	98	GLU	-	expression tag	UNP P45956
E	99	ASN	-	expression tag	UNP P45956
E	100	LEU	-	expression tag	UNP P45956
E	101	TYR	-	expression tag	UNP P45956
E	102	PHE	-	expression tag	UNP P45956
E	103	GLN	-	expression tag	UNP P45956
F	95	GLY	-	expression tag	UNP P45956
F	96	SER	-	expression tag	UNP P45956
F	97	SER	-	expression tag	UNP P45956
F	98	GLU	-	expression tag	UNP P45956
F	99	ASN	-	expression tag	UNP P45956
F	100	LEU	-	expression tag	UNP P45956
F	101	TYR	-	expression tag	UNP P45956
F	102	PHE	-	expression tag	UNP P45956
F	103	GLN	-	expression tag	UNP P45956

- Molecule 3 is a DNA chain called DNA (11-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	G	11	220	105	42	63	10	0	0	0

- Molecule 4 is a DNA chain called DNA (11-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	H	11	222	106	44	62	10	0	0	0

- Molecule 5 is a DNA chain called DNA (58-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	J	41	841	401	157	243	40	0	0	0

- Molecule 6 is a DNA chain called DNA (58-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	K	34	700	331	131	205	33	0	0	0

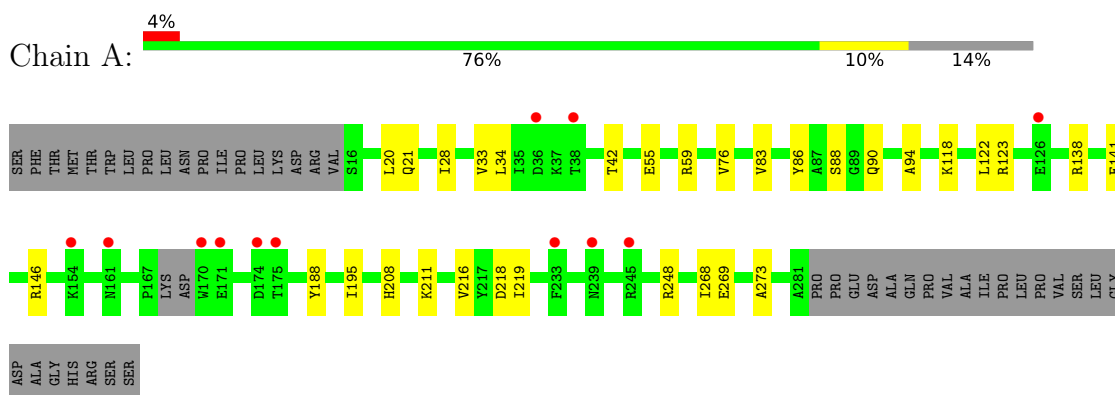
- Molecule 7 is NICKEL (II) ION (three-letter code: NI) (formula: Ni).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	A	2	Total 2	Ni 2	0	0
7	B	1	Total 1	Ni 1	0	0
7	C	2	Total 2	Ni 2	0	0
7	D	1	Total 1	Ni 1	0	0
7	E	1	Total 1	Ni 1	0	0
7	G	1	Total 1	Ni 1	0	0
7	J	7	Total 7	Ni 7	0	0
7	K	7	Total 7	Ni 7	0	0

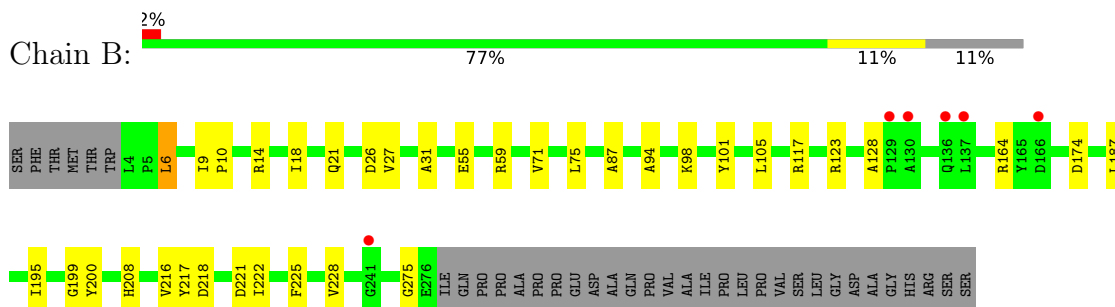
### 3 Residue-property plots [i](#)

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.

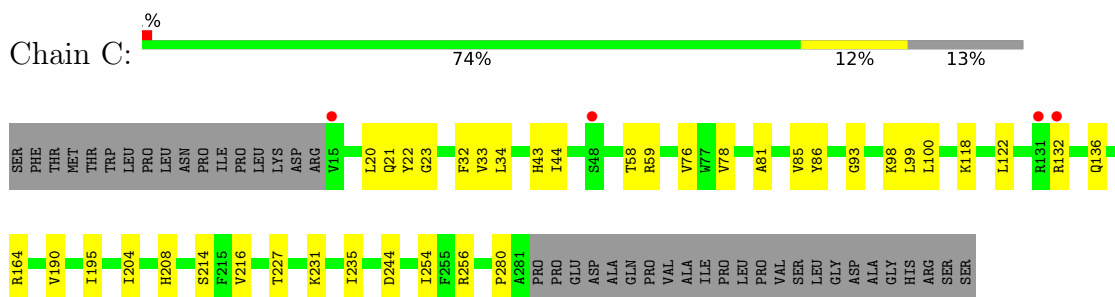
- Molecule 1: CRISPR-associated endonuclease Cas1



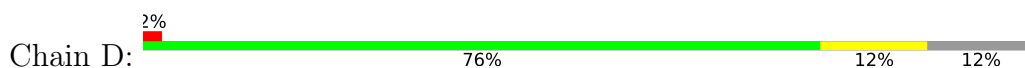
- Molecule 1: CRISPR-associated endonuclease Cas1

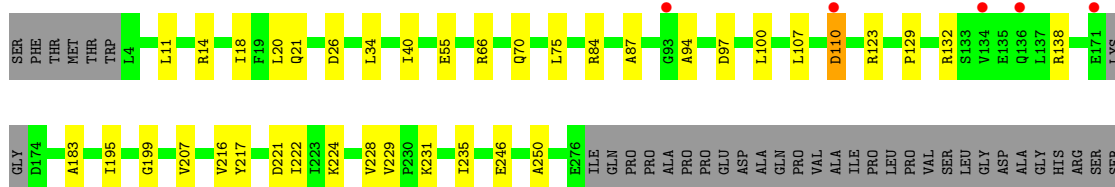


- Molecule 1: CRISPR-associated endonuclease Cas1

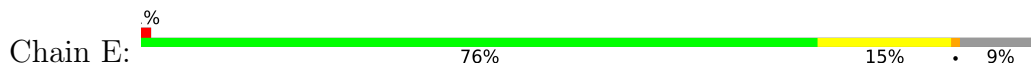


- Molecule 1: CRISPR-associated endonuclease Cas1





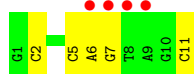
- Molecule 2: CRISPR-associated endoribonuclease Cas2



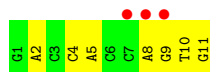
- Molecule 2: CRISPR-associated endoribonuclease Cas2



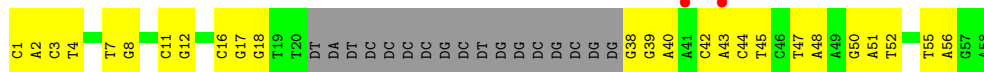
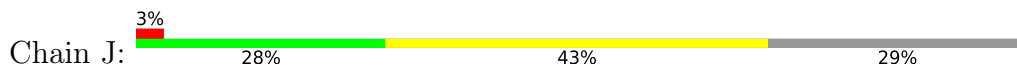
- Molecule 3: DNA (11-MER)



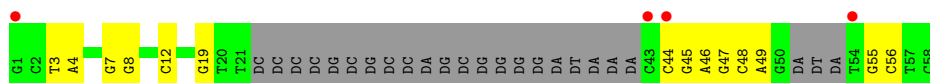
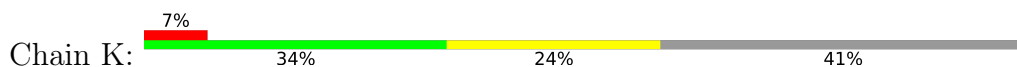
- Molecule 4: DNA (11-MER)



- Molecule 5: DNA (58-MER)



- Molecule 6: DNA (58-MER)





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	74.64Å 197.73Å 88.79Å 90.00° 111.35° 90.00°	Depositor
Resolution (Å)	39.50 – 3.31 39.50 – 2.96	Depositor EDS
% Data completeness (in resolution range)	98.8 (39.50-3.31) 96.3 (39.50-2.96)	Depositor EDS
$R_{merge}$	0.20	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.97 (at 2.95Å)	Xtrriage
Refinement program	PHENIX 1.11.1_2575	Depositor
R, $R_{free}$	0.224 , 0.263 0.224 , 0.263	Depositor DCC
$R_{free}$ test set	2502 reflections (5.24%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	87.8	Xtrriage
Anisotropy	0.530	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 70.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11762	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	106.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.27% 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

### 5.1 Standard geometry

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

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	0.25	0/2066	0.39	0/2800
1	B	0.24	0/2146	0.39	0/2910
1	C	0.25	0/2092	0.39	0/2837
1	D	0.24	0/2131	0.39	0/2889
2	E	0.25	0/753	0.45	0/1024
2	F	0.25	0/757	0.44	0/1029
3	G	0.53	0/246	0.81	0/377
4	H	0.53	0/249	0.80	0/382
5	J	0.55	0/943	0.92	0/1451
6	K	0.56	0/783	0.94	0/1204
All	All	0.32	0/12166	0.54	0/16903

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	2027	0	2077	26	0
1	B	2105	0	2169	21	0
1	C	2052	0	2106	25	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	2091	0	2150	26	0
2	E	739	0	756	9	0
2	F	743	0	759	10	0
3	G	220	0	124	5	0
4	H	222	0	124	5	0
5	J	841	0	463	23	0
6	K	700	0	385	16	0
7	A	2	0	0	0	0
7	B	1	0	0	0	0
7	C	2	0	0	0	0
7	D	1	0	0	0	0
7	E	1	0	0	0	0
7	G	1	0	0	0	0
7	J	7	0	0	0	0
7	K	7	0	0	0	0
All	All	11762	0	11113	140	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 6.

The worst 5 of 140 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:146:ARG:HH12	6:K:55:DG:H21	1.22	0.85
5:J:44:DC:H2'	5:J:45:DT:H71	1.69	0.75
5:J:47:DT:H2''	5:J:48:DA:C8	2.26	0.70
1:A:59:ARG:NH1	4:H:2:DA:OP1	2.25	0.69
1:C:100:LEU:HB3	1:D:107:LEU:HD21	1.75	0.68

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	260/308 (84%)	255 (98%)	5 (2%)	0	100	100
1	B	271/308 (88%)	262 (97%)	8 (3%)	1 (0%)	34	66
1	C	265/308 (86%)	259 (98%)	6 (2%)	0	100	100
1	D	267/308 (87%)	255 (96%)	12 (4%)	0	100	100
2	E	92/103 (89%)	90 (98%)	1 (1%)	1 (1%)	14	45
2	F	93/103 (90%)	90 (97%)	3 (3%)	0	100	100
All	All	1248/1438 (87%)	1211 (97%)	35 (3%)	2 (0%)	47	76

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	E	28	ALA
1	B	6	LEU

### 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	209/248 (84%)	209 (100%)	0	100	100
1	B	219/248 (88%)	218 (100%)	1 (0%)	88	93
1	C	212/248 (86%)	211 (100%)	1 (0%)	88	93
1	D	218/248 (88%)	217 (100%)	1 (0%)	88	93
2	E	79/87 (91%)	78 (99%)	1 (1%)	69	83
2	F	79/87 (91%)	79 (100%)	0	100	100
All	All	1016/1166 (87%)	1012 (100%)	4 (0%)	91	95

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	6	LEU
1	C	190	VAL
1	D	110	ASP

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
2	E	7	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
2	E	71	GLN
2	F	71	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 monosaccharides in this entry.

### 5.6 Ligand geometry [i](#)

Of 22 ligands modelled in this entry, 22 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 [i](#)

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	264/308 (85%)	0.30	12 (4%) 33 33	66, 106, 149, 161	0
1	B	273/308 (88%)	0.15	6 (2%) 62 60	64, 109, 157, 176	0
1	C	267/308 (86%)	0.01	4 (1%) 73 72	68, 90, 121, 158	0
1	D	271/308 (87%)	0.08	5 (1%) 68 67	62, 92, 136, 152	0
2	E	94/103 (91%)	-0.10	1 (1%) 80 81	63, 78, 104, 124	0
2	F	95/103 (92%)	-0.32	0 100 100	62, 77, 108, 116	0
3	G	11/11 (100%)	1.39	4 (36%) 0 0	73, 91, 189, 190	0
4	H	11/11 (100%)	1.35	3 (27%) 0 0	79, 90, 183, 186	0
5	J	41/58 (70%)	0.58	2 (4%) 29 29	93, 132, 191, 245	0
6	K	34/58 (58%)	0.71	4 (11%) 4 3	74, 135, 195, 229	0
All	All	1361/1576 (86%)	0.14	41 (3%) 50 49	62, 96, 151, 245	0

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	H	8	DA	4.6
3	G	9	DA	4.5
3	G	7	DG	4.5
3	G	8	DT	4.1
1	A	36	ASP	4.1

### 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 monosaccharides 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	NI	J	105	1/1	0.39	0.13	153,153,153,153	0
7	NI	J	101	1/1	0.62	0.12	148,148,148,148	0
7	NI	K	102	1/1	0.75	0.18	126,126,126,126	0
7	NI	K	107	1/1	0.77	0.26	117,117,117,117	1
7	NI	C	402	1/1	0.79	0.13	108,108,108,108	1
7	NI	K	106	1/1	0.79	0.08	148,148,148,148	0
7	NI	J	107	1/1	0.79	0.10	161,161,161,161	0
7	NI	K	101	1/1	0.81	0.21	110,110,110,110	0
7	NI	K	104	1/1	0.81	0.06	134,134,134,134	0
7	NI	E	201	1/1	0.82	0.59	153,153,153,153	0
7	NI	C	401	1/1	0.83	0.05	134,134,134,134	0
7	NI	J	102	1/1	0.84	0.06	125,125,125,125	0
7	NI	D	401	1/1	0.84	0.28	109,109,109,109	0
7	NI	J	106	1/1	0.84	0.09	129,129,129,129	0
7	NI	A	401	1/1	0.86	0.23	160,160,160,160	1
7	NI	K	105	1/1	0.89	0.09	119,119,119,119	0
7	NI	K	103	1/1	0.89	0.10	111,111,111,111	0
7	NI	B	401	1/1	0.89	0.33	123,123,123,123	0
7	NI	G	101	1/1	0.90	0.13	125,125,125,125	0
7	NI	J	104	1/1	0.90	0.04	129,129,129,129	0
7	NI	J	103	1/1	0.97	0.09	127,127,127,127	0
7	NI	A	402	1/1	0.98	0.04	115,115,115,115	1

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

There are no such residues in this entry.