



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 25, 2023 – 11:04 AM EDT

PDB ID : 5VVK  
Title : Cas1-Cas2 bound to full-site mimic  
Authors : Wright, A.V.; Knott, G.J.; Doxzen, K.D.; Doudna, J.A.  
Deposited on : 2017-05-19  
Resolution : 2.90 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

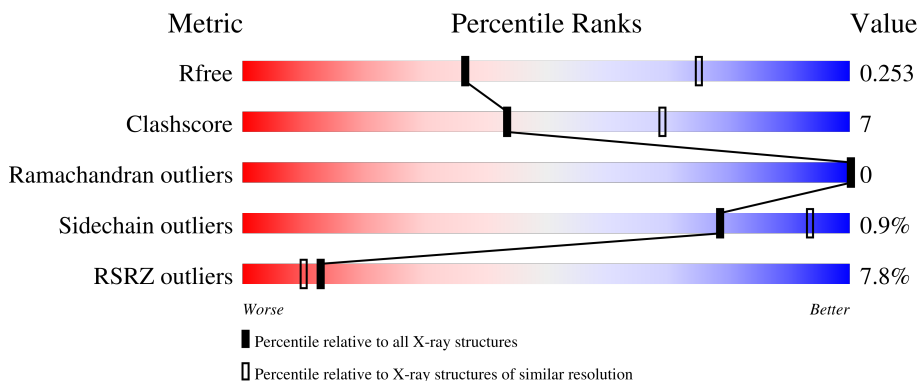
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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.



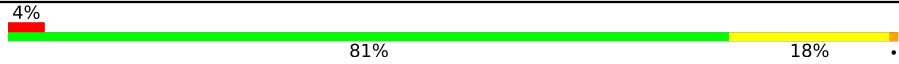

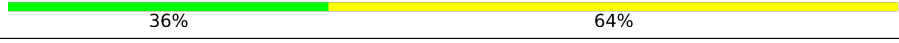
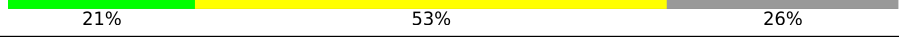
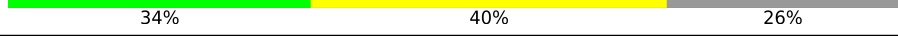
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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	 11% 76% 10% 14%
1	B	308	 11% 72% 13% 15%
1	C	308	 5% 72% 14% 13%
1	D	308	 5% 76% 13% 12%
2	E	94	 6% 83% 17%

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	F	94	 4% 81% 18%
3	G	11	 64% 36%
4	H	11	 36% 64%
5	J	58	 21% 53% 26%
6	K	58	 34% 40% 26%

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 11896 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	266	Total	C	N	O	S	0	0	0
			2045	1306	363	369	7			
1	B	263	Total	C	N	O	S	0	0	0
			2017	1290	361	359	7			
1	C	267	Total	C	N	O	S	0	0	0
			2052	1311	364	370	7			
1	D	272	Total	C	N	O	S	0	0	0
			2096	1340	372	377	7			

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	C	N	O	S	0	0	0
			739	475	128	132	4			
2	F	94	Total	C	N	O	S	0	0	0
			739	475	128	132	4			

- Molecule 3 is a DNA chain called DNA (5'-D(\*GP\*CP\*CP\*CP\*CP\*AP\*GP\*TP\*AP\*GP\*C)-3').

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 (5'-D(\*GP\*AP\*CP\*CP\*AP\*CP\*CP\*AP\*GP\*TP\*G)-3').

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	43	883	421	164	256	42	0	0	0

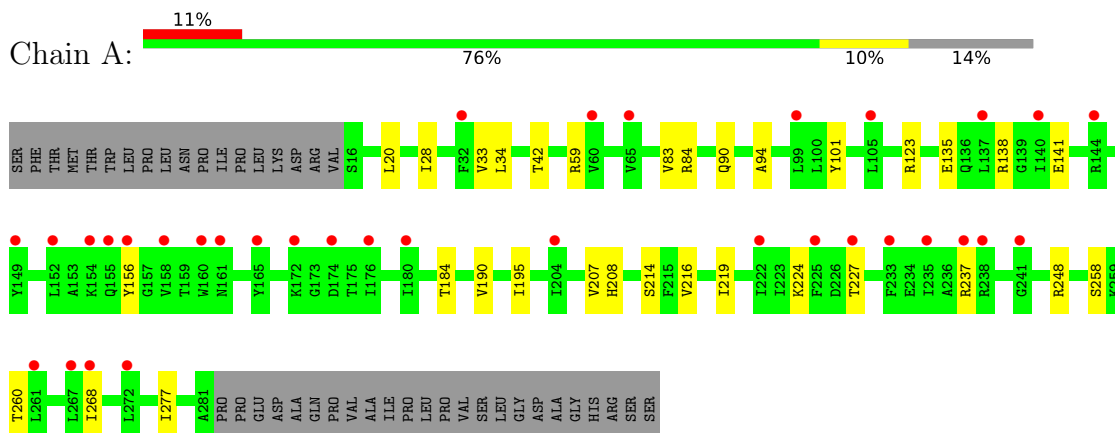
- Molecule 6 is a DNA chain called DNA (58-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
6	K	43	883	419	166	256	42	0	0	0

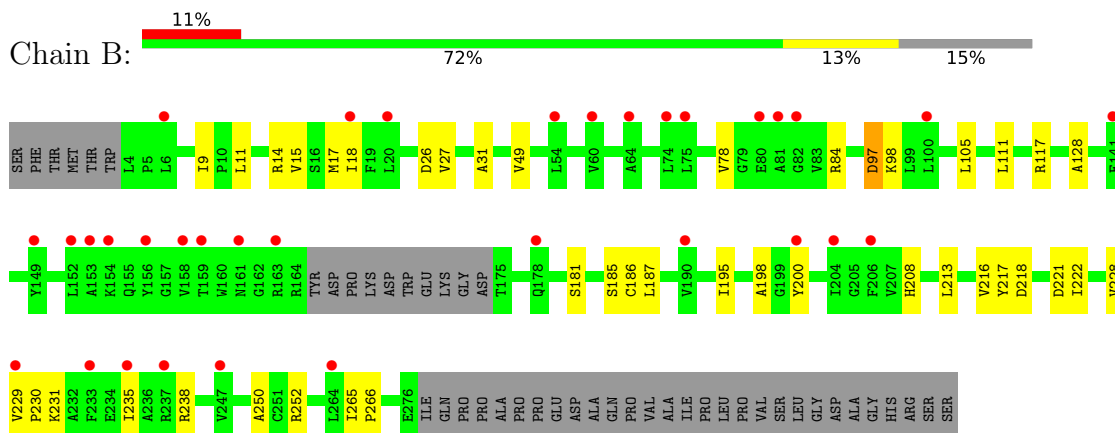
### 3 Residue-property plots [i](#)

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.

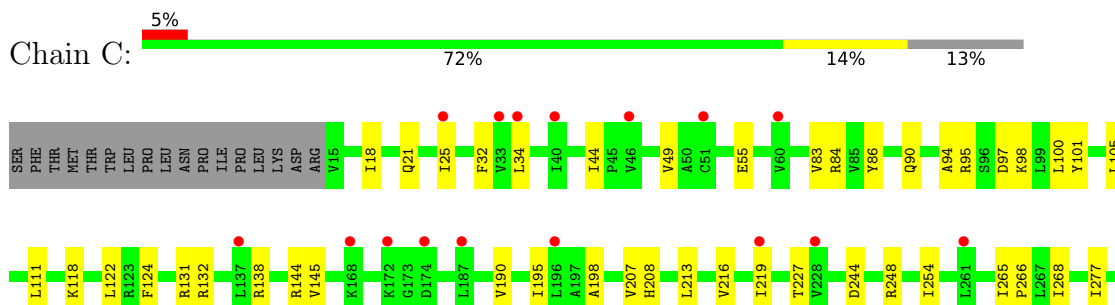
#### ● Molecule 1: CRISPR-associated endonuclease Cas1



#### ● Molecule 1: CRISPR-associated endonuclease Cas1



#### ● Molecule 1: CRISPR-associated endonuclease Cas1



A281  
 PRO  
 PRO  
 GLU  
 MET  
 THR  
 ASP  
 THR  
 ALA  
 GLN  
 PRO  
 VAL  
 VAL  
 ILE  
 ILE  
 PRO  
 LEU  
 LEU  
 PRO  
 VAL  
 SER  
 LEU  
 LEU  
 GLY  
 ASP  
 ALA  
 GLY  
 GLY  
 HIS  
 HIS  
 ARG  
 SER  
 SER

- Molecule 1: CRISPR-associated endonuclease Cas1

Chain D: 5% 76% 13% 12%

SER PHE THR MET THR ASP THR ALA GLN PRO VAL VAL ILE ILE PRO LEU LEU PRO VAL SER LEU LEU GLY ASP ALA GLY GLY HIS HIS ARG SER SER

E141 K154 V158 T159 W160 R163 E171 LYS G173 I176 C186 V190 I195 G199 Y200 V207 H208 L213 D218 I222 F233 R238 R245 R252 T260 A274 G275 E276 ILE GLN PRO PRO PRO PRO PRO PRO G11 ASP ALA GLN PRO

VAL ALA ILE PRO LEU VAL SER LEU ASP ALA GLY GLY HIS ARG SER SER

- Molecule 2: CRISPR-associated endonuclease Cas2

Chain E: 6% 83% 17%

M1 S2 M3 L4 V7 R16 G17 R18 E25 V26 V30 V32 I39 Q46 V57 M58 A59 M60 E65 F70 D84 G85 L86 R87 L88 V94

- Molecule 2: CRISPR-associated endonuclease Cas2

Chain F: 4% 81% 18%

M1 V7 M10 L15 R16 G17 R18 I21 E25 V26 V30 V32 M42 Q46 L50 V57 M58 F70 R78 L86 R87 L88 V94

- Molecule 3: DNA (5'-D(\*GP\*CP\*CP\*CP\*AP\*GP\*TP\*AP\*GP\*C)-3')

Chain G: 64% 36%

G1 C5 A6 G7 T8 A9 G10 C11

- Molecule 4: DNA (5'-D(\*GP\*AP\*CP\*CP\*AP\*CP\*CP\*AP\*GP\*TP\*G)-3')

Chain H: 36% 64%

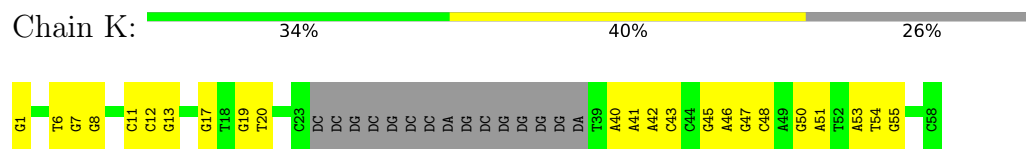
G1 A2 C3 T4 A5 C6 C7 A8 G9 T10 G11

- Molecule 5: DNA (58-MER)

Chain J: 21% 53% 26%

C1 A2 C3 T4 G5 G6 T7 G8 G9 T10 T11 C12 C13 C14 G17 G18 T21 A22 D1 D2 D3 D4 D5 D6 D7 D8 D9 D10 D11 D12 D13 D14 D15 D16 D17 D18 D19 D20 D21 D22 G38 G39 A40 A41 A42 C42 A43 C44 T45 C46 T47 A48 A49 G50 G51 A51 T52 T55 A56 A57 A58

● Molecule 6: DNA (58-MER)





## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	74.90Å 197.61Å 95.34Å 90.00° 112.70° 90.00°	Depositor
Resolution (Å)	98.81 – 2.90 98.81 – 2.76	Depositor EDS
% Data completeness (in resolution range)	99.8 (98.81-2.90) 98.1 (98.81-2.76)	Depositor EDS
$R_{merge}$	0.15	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.14 (at 2.77Å)	Xtrriage
Refinement program	PHENIX 1.11.1	Depositor
R, $R_{free}$	0.217 , 0.254 0.216 , 0.253	Depositor DCC
$R_{free}$ test set	3349 reflections (5.19%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	72.3	Xtrriage
Anisotropy	0.708	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 67.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.023 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	11896	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	101.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.98% 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](#)

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	0.24	0/2085	0.39	0/2827
1	B	0.24	0/2053	0.39	0/2782
1	C	0.25	0/2092	0.41	0/2837
1	D	0.25	0/2136	0.41	0/2896
2	E	0.26	0/753	0.48	0/1024
2	F	0.25	0/753	0.47	0/1024
3	G	0.55	0/246	0.81	0/377
4	H	0.55	0/249	0.79	0/382
5	J	0.53	0/990	0.91	0/1525
6	K	0.57	0/990	0.90	0/1525
All	All	0.33	0/12347	0.55	0/17199

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	2045	0	2097	17	0
1	B	2017	0	2095	27	0
1	C	2052	0	2106	26	0
1	D	2096	0	2155	25	0
2	E	739	0	756	15	0
2	F	739	0	756	16	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	G	220	0	124	4	0
4	H	222	0	124	4	0
5	J	883	0	487	25	0
6	K	883	0	485	18	0
All	All	11896	0	11185	155	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 7.

The worst 5 of 155 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:D:18:ILE:HD11	2:E:86:LEU:HD23	1.64	0.80
5:J:39:DG:H2''	5:J:40:DA:C8	2.17	0.79
5:J:47:DT:H2''	5:J:48:DA:C8	2.23	0.74
1:C:145:VAL:HG21	5:J:58:DA:H5'	1.71	0.72
1:A:59:ARG:NH1	4:H:2:DA:OP1	2.23	0.72

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	264/308 (86%)	262 (99%)	2 (1%)	0	100	100
1	B	259/308 (84%)	253 (98%)	6 (2%)	0	100	100
1	C	265/308 (86%)	260 (98%)	5 (2%)	0	100	100
1	D	268/308 (87%)	262 (98%)	6 (2%)	0	100	100
2	E	92/94 (98%)	89 (97%)	3 (3%)	0	100	100
2	F	92/94 (98%)	88 (96%)	4 (4%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
All	All	1240/1420 (87%)	1214 (98%)	26 (2%)	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	211/248 (85%)	208 (99%)	3 (1%)	67	89
1	B	210/248 (85%)	209 (100%)	1 (0%)	88	96
1	C	212/248 (86%)	210 (99%)	2 (1%)	78	93
1	D	218/248 (88%)	216 (99%)	2 (1%)	78	93
2	E	79/79 (100%)	79 (100%)	0	100	100
2	F	79/79 (100%)	78 (99%)	1 (1%)	69	90
All	All	1009/1150 (88%)	1000 (99%)	9 (1%)	78	93

5 of 9 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	260	THR
2	F	86	LEU
1	B	97	ASP
1	C	97	ASP
1	C	190	VAL

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

### 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](#)

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 [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	266/308 (86%)	0.81	34 (12%) 3 2	59, 99, 139, 158	0
1	B	263/308 (85%)	0.92	33 (12%) 3 3	55, 105, 164, 171	0
1	C	267/308 (86%)	0.69	16 (5%) 21 18	57, 85, 116, 157	0
1	D	272/308 (88%)	0.67	14 (5%) 28 24	47, 85, 125, 149	0
2	E	94/94 (100%)	0.72	6 (6%) 19 15	48, 70, 96, 108	0
2	F	94/94 (100%)	0.53	4 (4%) 35 31	52, 68, 101, 116	0
3	G	11/11 (100%)	-0.03	0 100 100	63, 88, 177, 189	0
4	H	11/11 (100%)	0.06	0 100 100	68, 93, 199, 202	0
5	J	43/58 (74%)	-0.37	0 100 100	74, 129, 191, 208	0
6	K	43/58 (74%)	-0.25	0 100 100	67, 145, 198, 210	0
All	All	1364/1558 (87%)	0.67	107 (7%) 13 10	47, 90, 158, 210	0

The worst 5 of 107 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	152	LEU	10.3
1	A	158	VAL	5.0
1	A	272	LEU	4.9
1	A	156	TYR	4.7
1	A	155	GLN	4.6

### 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](#)

There are no ligands in this entry.

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

There are no such residues in this entry.