



# Full wwPDB X-ray Structure Validation Report ⓘ

May 23, 2020 – 01:08 pm BST

PDB ID : 5Z08  
Title : The crystal structure of kinetochore subunits Cenp-H/I/K triple complex  
Authors : Tian, W.; Hu, L.Q.; He, X.  
Deposited on : 2017-12-18  
Resolution : 2.20 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

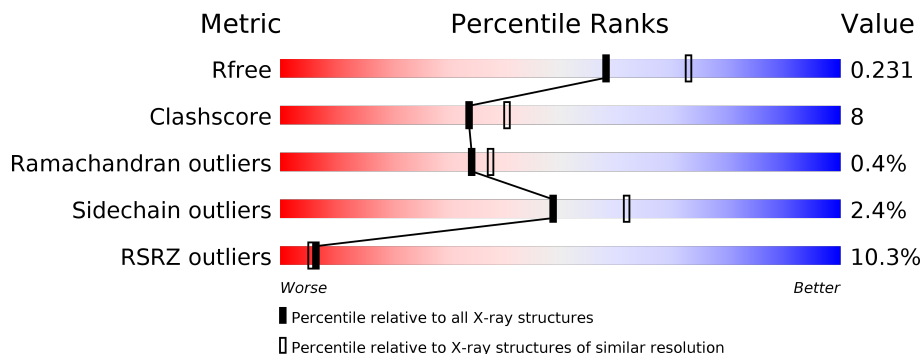
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	229	 4% 79% 14% • 7%
1	B	229	 3% 80% 13% • 7%
2	C	168	 21% 37% 24% 39%
3	D	45	 16% 73% 24% •

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4699 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 Cenp-I.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	214	Total	C	N	O	S	0	0	0
			1709	1097	313	297	2			
1	B	214	Total	C	N	O	S	0	0	0
			1709	1097	313	297	2			

- Molecule 2 is a protein called Cenp-K.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	C	102	Total	C	N	O	S	0	0	0
			806	503	147	151	5			

- Molecule 3 is a protein called Cenp-H.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	D	44	Total	C	N	O	S	0	0	0
			355	219	72	62	2			

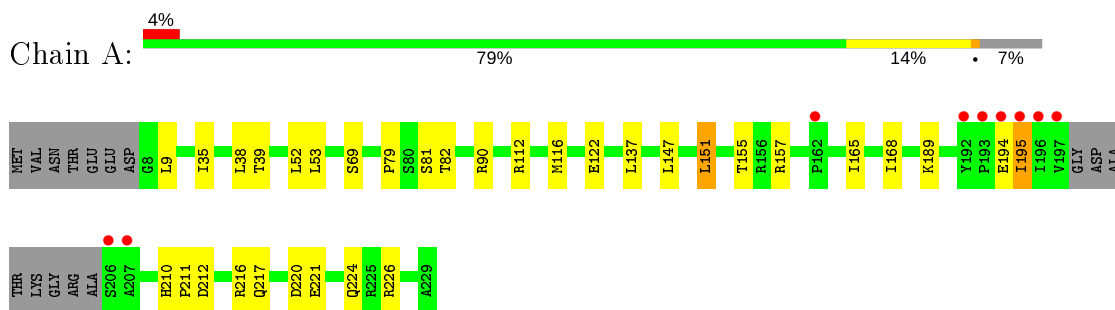
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	67	Total	0	0
			67		
4	B	43	Total	0	0
			43		
4	C	5	Total	0	0
			5		
4	D	5	Total	0	0
			5		

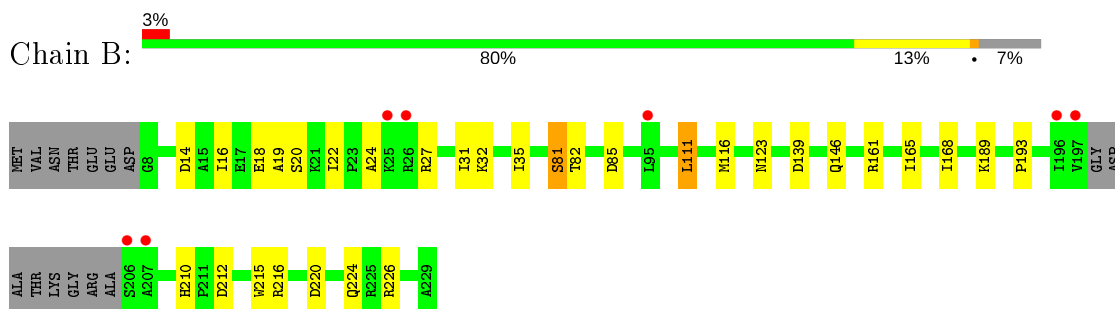
### 3 Residue-property plots [i](#)

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.

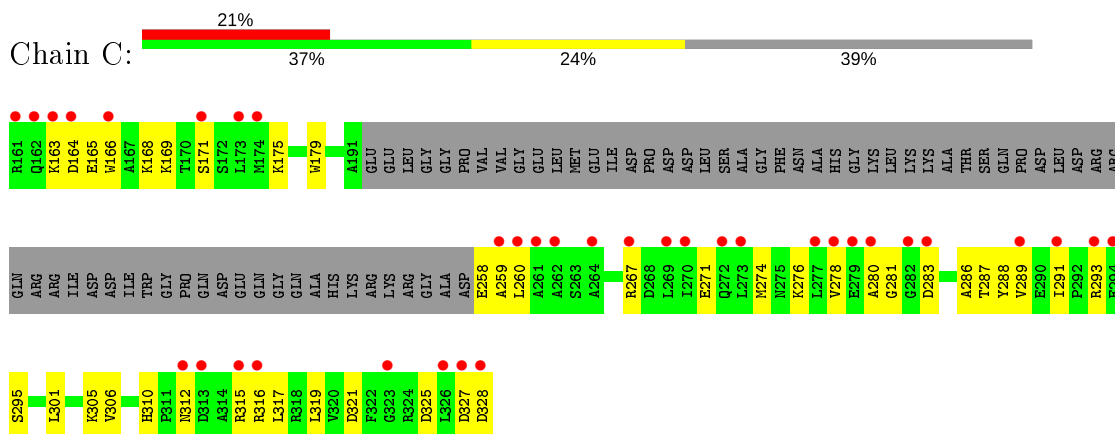
- Molecule 1: Cnp-I



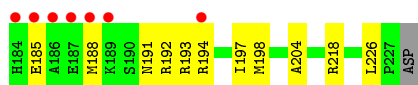
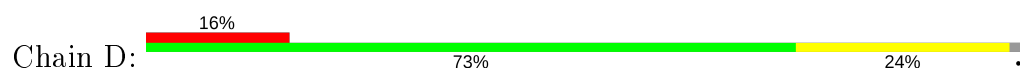
- Molecule 1: Cnp-I



- Molecule 2: Cnp-K



- Molecule 3: Cnp-H



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	67.85Å 67.85Å 186.27Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.98 – 2.20 47.98 – 2.20	Depositor EDS
% Data completeness (in resolution range)	99.7 (47.98-2.20) 99.7 (47.98-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.56 (at 2.20Å)	Xtrriage
Refinement program	PHENIX (dev_2463)	Depositor
R, $R_{free}$	0.188 , 0.230 0.188 , 0.231	Depositor DCC
$R_{free}$ test set	2143 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	46.4	Xtrriage
Anisotropy	0.463	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 52.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.057 for h,-k,-l	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4699	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	69.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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	0.40	0/1746	0.56	0/2374
1	B	0.40	0/1746	0.55	0/2374
2	C	0.34	0/817	0.52	1/1095 (0.1%)
3	D	0.32	0/360	0.56	0/482
All	All	0.39	0/4669	0.55	1/6325 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	317	LEU	CA-CB-CG	5.21	127.28	115.30

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	1709	0	1777	21	0
1	B	1709	0	1777	25	0
2	C	806	0	801	32	0
3	D	355	0	361	11	0
4	A	67	0	0	0	0
4	B	43	0	0	1	0
4	C	5	0	0	0	0
4	D	5	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	4699	0	4716	79	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 8.

All (79) 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:B:14:ASP:O	1:B:18:GLU:OE1	1.86	0.93
1:A:116:MET:HE1	1:A:212:ASP:HB3	1.52	0.91
2:C:288:TYR:HB3	2:C:316:ARG:HB3	1.62	0.80
2:C:274:MET:HE1	3:D:198:MET:HG2	1.68	0.76
1:B:27:ARG:O	4:B:701:HOH:O	2.04	0.75
2:C:319:LEU:HD23	2:C:321:ASP:H	1.50	0.75
2:C:289:VAL:HG23	2:C:291:ILE:HD11	1.74	0.70
2:C:258:GLU:HG2	2:C:259:ALA:H	1.56	0.70
2:C:258:GLU:HG2	2:C:259:ALA:N	2.06	0.69
1:A:116:MET:CE	1:A:212:ASP:HB3	2.22	0.69
1:A:81:SER:HA	1:A:226:ARG:HD2	1.76	0.67
2:C:276:LYS:HD2	2:C:287:THR:HA	1.76	0.67
2:C:166:TRP:CE2	3:D:192:ARG:HG3	2.33	0.64
1:B:111:LEU:HB3	1:B:146:GLN:HB3	1.80	0.63
2:C:165:GLU:N	2:C:165:GLU:OE1	2.32	0.63
2:C:306:VAL:HG12	3:D:204:ALA:HB3	1.81	0.62
2:C:163:LYS:HA	3:D:188:MET:HE2	1.82	0.62
1:A:165:ILE:O	1:A:168:ILE:HG22	2.01	0.60
1:B:31:ILE:HG12	1:B:35:ILE:HG12	1.84	0.60
1:B:116:MET:HE1	1:B:212:ASP:HB3	1.84	0.60
1:A:189:LYS:HE3	1:A:195:ILE:HG22	1.85	0.59
2:C:301:LEU:HD22	2:C:306:VAL:HG21	1.84	0.58
1:B:22:ILE:HD11	1:B:27:ARG:HG2	1.84	0.58
1:B:116:MET:CE	1:B:210:HIS:HE1	2.16	0.58
1:B:116:MET:HE2	1:B:210:HIS:HE1	1.68	0.58
1:B:116:MET:CE	1:B:212:ASP:HB3	2.34	0.57
2:C:325:ASP:HB3	2:C:328:ASP:HB3	1.85	0.57
1:A:137:LEU:HB2	1:A:147:LEU:HD11	1.87	0.56
1:B:81:SER:OG	1:B:82:THR:N	2.32	0.56
2:C:289:VAL:O	2:C:316:ARG:HA	2.07	0.55
2:C:169:LYS:HE2	3:D:226:LEU:HD21	1.89	0.54
2:C:171:SER:O	2:C:175:LYS:HG3	2.08	0.53
1:B:161:ARG:NH1	2:C:325:ASP:OD2	2.41	0.53

*Continued on next page...*



*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:D:194:ARG:HA	3:D:197:ILE:HD12	1.91	0.53
1:A:217:GLN:O	1:A:221:GLU:HG3	2.10	0.52
1:B:81:SER:HA	1:B:226:ARG:HD2	1.92	0.52
1:A:220:ASP:O	1:A:224:GLN:HG3	2.10	0.52
2:C:276:LYS:HZ2	2:C:287:THR:C	2.12	0.52
1:A:216:ARG:HD2	1:B:224:GLN:HG2	1.91	0.51
1:B:220:ASP:O	1:B:224:GLN:HG3	2.11	0.51
2:C:276:LYS:NZ	2:C:289:VAL:HG13	2.27	0.50
1:A:52:LEU:HD11	1:A:79:PRO:HB3	1.94	0.50
2:C:258:GLU:OE2	2:C:260:LEU:HD23	2.12	0.50
1:B:189:LYS:HG2	1:B:193:PRO:HA	1.93	0.50
1:A:35:ILE:HD13	1:A:69:SER:HB3	1.95	0.49
1:A:151:LEU:O	1:A:155:THR:HG22	2.11	0.49
2:C:267:ARG:HH12	2:C:271:GLU:CD	2.16	0.49
1:B:116:MET:HE1	1:B:215:TRP:HB3	1.93	0.49
1:B:116:MET:HG2	1:B:210:HIS:CE1	2.48	0.48
1:A:210:HIS:CE1	1:A:216:ARG:HB2	2.47	0.48
1:B:22:ILE:CD1	1:B:27:ARG:HG2	2.43	0.47
2:C:328:ASP:C	3:D:192:ARG:HH22	2.18	0.47
1:A:194:GLU:O	1:A:195:ILE:HB	2.15	0.47
1:B:210:HIS:CD2	1:B:216:ARG:HD3	2.50	0.46
1:B:16:ILE:HD12	1:B:31:ILE:HD11	1.96	0.46
3:D:193:ARG:O	3:D:197:ILE:HD12	2.16	0.45
2:C:293:ARG:HD3	2:C:315:ARG:HD2	1.99	0.45
2:C:310:HIS:CD2	2:C:312:ASN:HB2	2.51	0.45
1:A:82:THR:HB	1:A:122:GLU:HG3	2.00	0.44
2:C:293:ARG:CD	2:C:315:ARG:HD2	2.48	0.44
1:A:116:MET:HE2	1:A:210:HIS:HE1	1.83	0.44
2:C:179:TRP:CH2	3:D:218:ARG:HD3	2.53	0.44
2:C:278:VAL:HG11	3:D:194:ARG:NH1	2.33	0.43
1:A:116:MET:HE3	1:A:211:PRO:HG2	1.99	0.43
1:B:116:MET:CE	1:B:215:TRP:HB3	2.48	0.43
2:C:327:ASP:O	2:C:327:ASP:CG	2.57	0.42
1:A:53:LEU:HD11	1:A:90:ARG:CZ	2.50	0.42
1:B:85:ASP:OD1	1:B:123:ASN:HB3	2.20	0.42
2:C:164:ASP:O	2:C:168:LYS:N	2.42	0.42
3:D:185:GLU:O	3:D:185:GLU:HG2	2.19	0.42
1:B:165:ILE:O	1:B:168:ILE:HG22	2.20	0.41
1:A:9:LEU:HD11	1:A:38:LEU:HD12	2.02	0.41
2:C:289:VAL:CG2	2:C:291:ILE:HD11	2.48	0.41
1:B:19:ALA:CB	1:B:31:ILE:HD12	2.51	0.41

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:189:LYS:HE2	1:A:189:LYS:HA	2.02	0.41
2:C:276:LYS:HD2	2:C:286:ALA:O	2.21	0.41
1:A:116:MET:CE	1:A:210:HIS:HE1	2.35	0.40
1:B:24:ALA:HA	1:B:27:ARG:HG3	2.04	0.40
2:C:280:ALA:HB1	2:C:283:ASP:O	2.22	0.40

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	210/229 (92%)	205 (98%)	4 (2%)	1 (0%)	29	31
1	B	210/229 (92%)	206 (98%)	4 (2%)	0	100	100
2	C	98/168 (58%)	93 (95%)	4 (4%)	1 (1%)	15	14
3	D	42/45 (93%)	40 (95%)	2 (5%)	0	100	100
All	All	560/671 (84%)	544 (97%)	14 (2%)	2 (0%)	34	37

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	C	281	GLY
1	A	195	ILE

### 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	186/197 (94%)	182 (98%)	4 (2%)	52	65
1	B	186/197 (94%)	181 (97%)	5 (3%)	44	57
2	C	82/135 (61%)	80 (98%)	2 (2%)	49	62
3	D	36/37 (97%)	35 (97%)	1 (3%)	43	56
All	All	490/566 (87%)	478 (98%)	12 (2%)	49	62

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	39	THR
1	A	112	ARG
1	A	151	LEU
1	A	157	ARG
1	B	20	SER
1	B	32	LYS
1	B	81	SER
1	B	111	LEU
1	B	139	ASP
2	C	295	SER
2	C	305	LYS
3	D	191	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	190	ASN
1	A	210	HIS
1	B	210	HIS

### 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 [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<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	214/229 (93%)	0.15	9 (4%) 36 34	40, 53, 89, 162	0
1	B	214/229 (93%)	0.38	7 (3%) 46 44	42, 55, 98, 137	0
2	C	102/168 (60%)	1.56	36 (35%) 0 0	64, 100, 137, 159	0
3	D	44/45 (97%)	0.76	7 (15%) 1 1	54, 67, 130, 141	0
All	All	574/671 (85%)	0.53	59 (10%) 6 5	40, 60, 125, 162	0

All (59) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	206	SER	5.7
1	A	193	PRO	5.7
1	A	197	VAL	5.4
3	D	186	ALA	5.2
2	C	162	GLN	5.2
1	A	192	TYR	4.9
1	B	206	SER	4.8
2	C	328	ASP	4.6
2	C	260	LEU	4.5
2	C	161	ARG	4.2
3	D	185	GLU	4.2
2	C	293	ARG	4.1
2	C	166	TRP	3.9
1	A	196	ILE	3.8
3	D	184	HIS	3.7
2	C	278	VAL	3.7
2	C	326	LEU	3.5
2	C	279	GLU	3.4
2	C	277	LEU	3.3
2	C	163	LYS	3.3
3	D	188	MET	3.2

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	C	283	ASP	3.2
2	C	282	GLY	3.2
1	B	25	LYS	3.2
2	C	264	ALA	3.1
1	A	194	GLU	3.1
2	C	164	ASP	3.1
2	C	327	ASP	3.0
1	B	26	ARG	2.8
2	C	289	VAL	2.8
1	A	207	ALA	2.8
2	C	262	ALA	2.8
2	C	291	ILE	2.6
2	C	267	ARG	2.6
2	C	171	SER	2.5
2	C	270	ILE	2.5
1	A	195	ILE	2.5
2	C	259	ALA	2.5
1	B	197	VAL	2.5
1	B	207	ALA	2.4
2	C	174	MET	2.4
3	D	189	LYS	2.4
2	C	323	GLY	2.4
2	C	269	LEU	2.3
2	C	316	ARG	2.3
3	D	194	ARG	2.3
2	C	280	ALA	2.2
2	C	261	ALA	2.2
1	A	162	PRO	2.2
1	B	196	ILE	2.2
3	D	187	GLU	2.1
2	C	272	GLN	2.1
1	B	95	LEU	2.1
2	C	173	LEU	2.1
2	C	313	ASP	2.1
2	C	315	ARG	2.1
2	C	294	GLU	2.0
2	C	312	ASN	2.0
2	C	273	LEU	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 [i](#)

There are no carbohydrates 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.