



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 28, 2016 – 05:54 PM EST

PDB ID : 5LSJ  
Title : CRYSTAL STRUCTURE OF THE HUMAN KINETOCHORE MIS12-CENP-C delta-HEAD2 COMPLEX  
Authors : Vetter, I.R.; Petrovic, A.; Keller, J.; Liu, Y.  
Deposited on : 2016-09-02  
Resolution : 3.25 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

<http://wwpdb.org/validation/2016/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
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028320
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20028320

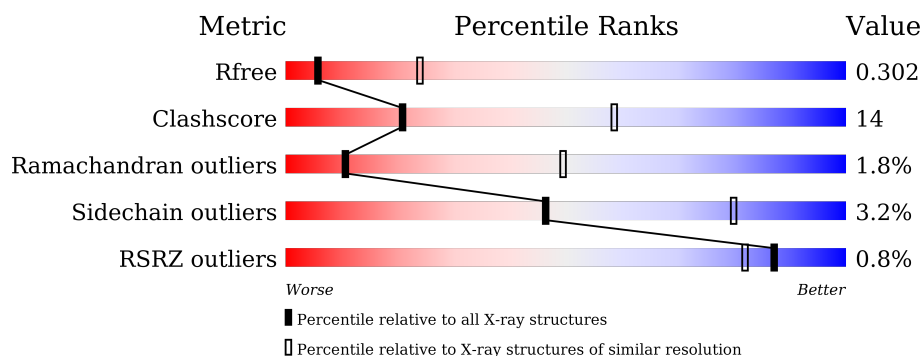
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.25 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))
$R_{free}$	91344	1624 (3.32-3.20)
Clashscore	102246	1806 (3.32-3.20)
Ramachandran outliers	100387	1773 (3.32-3.20)
Sidechain outliers	100360	1771 (3.32-3.20)
RSRZ outliers	91569	1632 (3.32-3.20)

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. 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	205	<div> <div>2%</div> <div>66%</div> <div>29%</div> <div>• •</div> </div>
1	C	205	<div> <div>60%</div> <div>34%</div> <div>• •</div> </div>
2	B	176	<div> <div>61%</div> <div>35%</div> <div>• •</div> </div>
2	E	176	<div> <div>66%</div> <div>30%</div> <div>• • •</div> </div>
3	D	178	<div> <div>39%</div> <div>17%</div> <div>43%</div> <div>•</div> </div>
3	F	178	<div> <div>2%</div> <div>42%</div> <div>15%</div> <div>43%</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	G	116	<div><div><div></div><div></div><div></div><div></div></div><div>3%54%33%•12%</div></div>
4	N	116	<div><div><div></div><div></div><div></div><div></div></div><div>%51%33%••13%</div></div>
5	P	76	<div><div><div></div><div></div><div></div><div></div></div><div>11%7%83%</div></div>
5	Q	76	<div><div><div></div><div></div><div></div><div></div></div><div>8%11%•80%</div></div>

## 2 Entry composition [i](#)

There are 5 unique types of molecules in this entry. The entry contains 9621 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 Protein MIS12 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	199	Total	C	N	O	S	0	0	0
			1642	1049	270	312	11			
1	C	199	Total	C	N	O	S	0	0	0
			1642	1049	270	312	11			

- Molecule 2 is a protein called Polyamine-modulated factor 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	173	Total	C	N	O	S	0	0	0
			1412	886	256	265	5			
2	E	173	Total	C	N	O	S	0	0	0
			1412	886	256	265	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	30	MET	-	initiating methionine	UNP Q6P1K2
E	30	MET	-	initiating methionine	UNP Q6P1K2

- Molecule 3 is a protein called Kinetochores-associated protein DSN1 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	101	Total	C	N	O	S	0	0	0
			815	509	134	166	6			
3	F	101	Total	C	N	O	S	0	0	0
			815	509	134	166	6			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	185	MET	-	initiating methionine	UNP Q9H410
D	357	HIS	-	expression tag	UNP Q9H410

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	358	HIS	-	expression tag	UNP Q9H410
D	359	HIS	-	expression tag	UNP Q9H410
D	360	HIS	-	expression tag	UNP Q9H410
D	361	HIS	-	expression tag	UNP Q9H410
D	362	HIS	-	expression tag	UNP Q9H410
F	185	MET	-	initiating methionine	UNP Q9H410
F	357	HIS	-	expression tag	UNP Q9H410
F	358	HIS	-	expression tag	UNP Q9H410
F	359	HIS	-	expression tag	UNP Q9H410
F	360	HIS	-	expression tag	UNP Q9H410
F	361	HIS	-	expression tag	UNP Q9H410
F	362	HIS	-	expression tag	UNP Q9H410

- Molecule 4 is a protein called Kinetochore-associated protein NSL1 homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	N	101	Total	C	N	O	S	0	0	0
			810	516	135	154	5			
4	G	102	Total	C	N	O	S	0	0	0
			821	525	136	155	5			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
N	91	MET	-	initiating methionine	UNP Q96IY1
G	91	MET	-	initiating methionine	UNP Q96IY1

- Molecule 5 is a protein called Centromere protein C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	P	13	Total	C	N	O	S	0	0	0
			117	73	26	17	1			
5	Q	15	Total	C	N	O	S	0	0	0
			135	84	31	19	1			

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	-4	GLY	-	expression tag	UNP Q03188
P	-3	PRO	-	expression tag	UNP Q03188
P	-2	LEU	-	expression tag	UNP Q03188
P	-1	GLY	-	expression tag	UNP Q03188

*Continued on next page...*

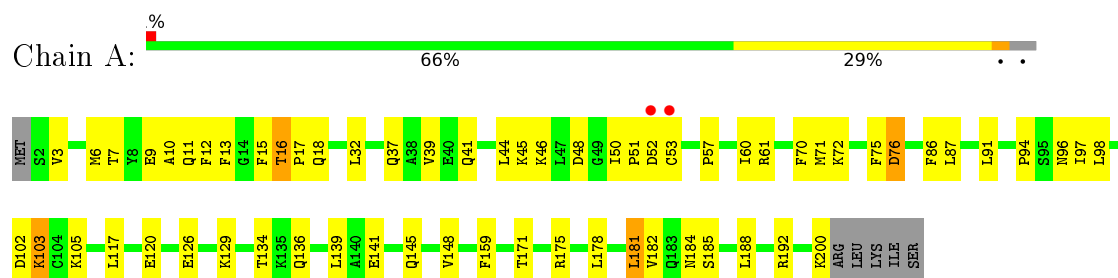
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
P	0	SER	-	expression tag	UNP Q03188
Q	-4	GLY	-	expression tag	UNP Q03188
Q	-3	PRO	-	expression tag	UNP Q03188
Q	-2	LEU	-	expression tag	UNP Q03188
Q	-1	GLY	-	expression tag	UNP Q03188
Q	0	SER	-	expression tag	UNP Q03188

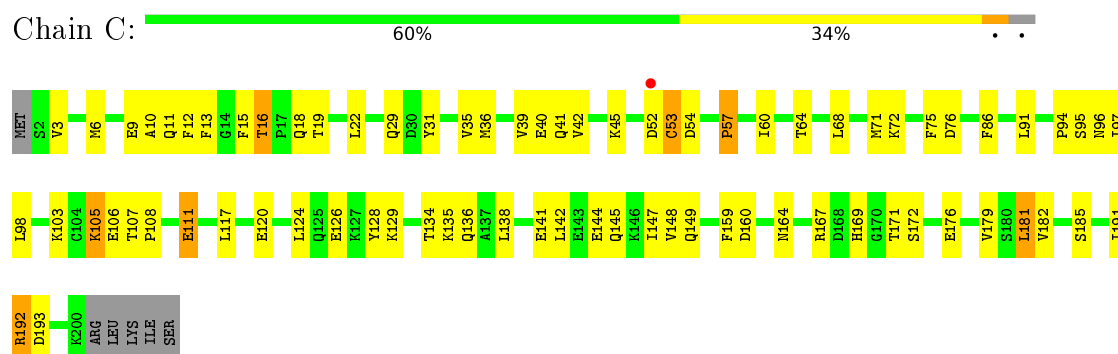
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

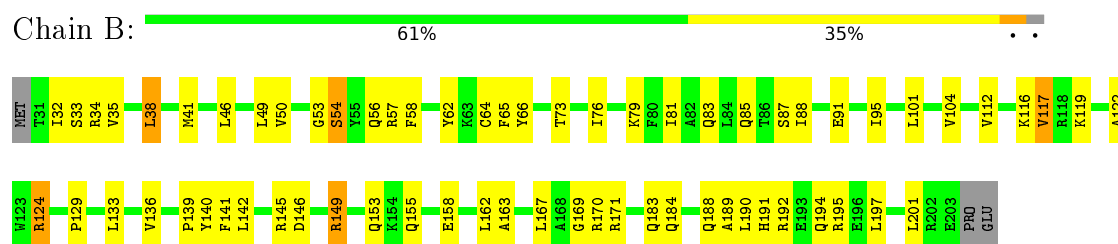
- Molecule 1: Protein MIS12 homolog



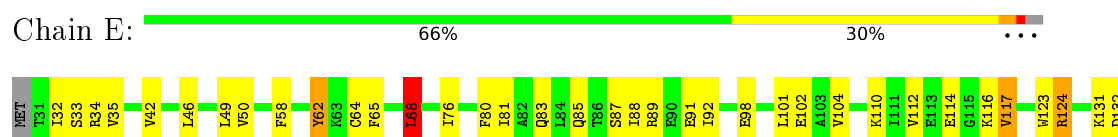
- Molecule 1: Protein MIS12 homolog

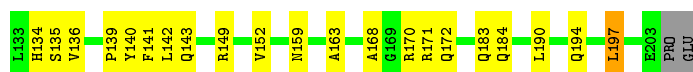


- Molecule 2: Polyamine-modulated factor 1



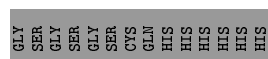
- Molecule 2: Polyamine-modulated factor 1





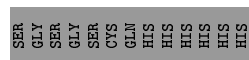
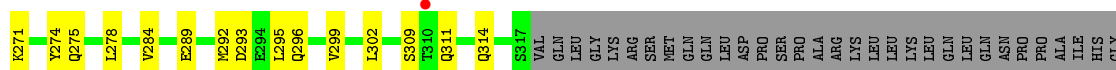
- Molecule 3: Kinetochore-associated protein DSN1 homolog

Chain D: 39% 17% 43%



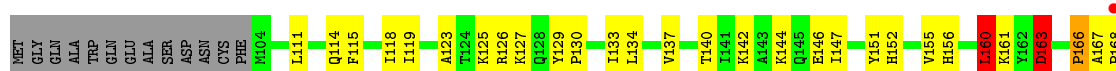
- Molecule 3: Kinetochore-associated protein DSN1 homolog

Chain F: 2% 42% 15% 43%



- Molecule 4: Kinetochore-associated protein NSL1 homolog

Chain N: 51% 33% 13%



- Molecule 4: Kinetochore-associated protein NSL1 homolog

Chain G: 3% 54% 33% 12%



- Molecule 5: Centromere protein C



SER  
THR  
ARG  
LYS  
ILE  
LYS  
ASP  
THR  
CYS  
ILE

LYS  
SER  
VAL  
PRO  
ASN  
SER  
THR  
ARG  
LYS  
ILE  
LYS  
ASP  
THR  
CYS  
ILE

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	68.16 Å   156.40 Å   76.00 Å 90.00°   102.82°   90.00°	Depositor
Resolution (Å)	19.99 – 3.25 45.55 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.7 (19.99-3.25) 99.8 (45.55-2.80)	Depositor EDS
$R_{merge}$	0.35	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.83 (at 2.81 Å)	Xtriage
Refinement program	PHENIX (1.10 _2155: ???)	Depositor
R, $R_{free}$	0.246   ,   0.297 0.249   ,   0.302	Depositor DCC
$R_{free}$ test set	1219 reflections (5.00%)	DCC
Wilson B-factor (Å <sup>2</sup> )	37.7	Xtriage
Anisotropy	0.653	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 28.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.38$ , $\langle L^2 \rangle = 0.21$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	9621	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.08% 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

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	0.29	0/1671	0.48	0/2246
1	C	0.35	2/1671 (0.1%)	0.49	0/2246
2	B	0.26	0/1433	0.47	1/1933 (0.1%)
2	E	0.27	0/1433	0.49	1/1933 (0.1%)
3	D	0.26	0/824	0.49	0/1104
3	F	0.26	0/824	0.48	0/1104
4	G	0.27	0/836	0.46	0/1126
4	N	0.40	1/824 (0.1%)	0.58	2/1110 (0.2%)
5	P	0.24	0/119	0.59	0/156
5	Q	0.24	0/138	0.55	0/182
All	All	0.30	3/9773 (0.0%)	0.49	4/13140 (0.0%)

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	N	152	HIS	C-N	8.33	1.50	1.34
1	C	111	GLU	CD-OE1	-5.93	1.19	1.25
1	C	111	GLU	CD-OE2	-5.37	1.19	1.25

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	N	173	LEU	CA-CB-CG	6.95	131.28	115.30
4	N	160	LEU	CA-CB-CG	6.46	130.16	115.30
2	E	68	LEU	CA-CB-CG	5.67	128.35	115.30
2	B	38	LEU	CA-CB-CG	5.21	127.29	115.30

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	1642	0	1646	65	0
1	C	1642	0	1646	77	0
2	B	1412	0	1419	54	0
2	E	1412	0	1419	48	0
3	D	815	0	801	29	0
3	F	815	0	801	24	0
4	G	821	0	842	33	0
4	N	810	0	833	29	0
5	P	117	0	116	6	0
5	Q	135	0	136	7	0
All	All	9621	0	9659	277	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 14.

The worst 5 of 277 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:181:LEU:HD21	3:D:302:LEU:HD22	1.63	0.79
1:A:6:MET:HG3	1:A:18:GLN:HE22	1.47	0.79
2:B:46:LEU:HD21	2:B:88:ILE:HG21	1.68	0.74
2:E:46:LEU:HD21	2:E:88:ILE:HG21	1.71	0.72
1:A:7:THR:OG1	4:N:126:ARG:NH1	2.23	0.72

There are no symmetry-related clashes.

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	197/205 (96%)	187 (95%)	8 (4%)	2 (1%)	19	62
1	C	197/205 (96%)	181 (92%)	14 (7%)	2 (1%)	19	62
2	B	171/176 (97%)	161 (94%)	6 (4%)	4 (2%)	8	42
2	E	171/176 (97%)	164 (96%)	4 (2%)	3 (2%)	11	49
3	D	97/178 (54%)	96 (99%)	1 (1%)	0	100	100
3	F	97/178 (54%)	95 (98%)	2 (2%)	0	100	100
4	G	100/116 (86%)	90 (90%)	7 (7%)	3 (3%)	5	35
4	N	99/116 (85%)	86 (87%)	9 (9%)	4 (4%)	4	26
5	P	11/76 (14%)	8 (73%)	3 (27%)	0	100	100
5	Q	13/76 (17%)	8 (62%)	2 (15%)	3 (23%)	0	0
All	All	1153/1502 (77%)	1076 (93%)	56 (5%)	21 (2%)	11	49

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	117	VAL
2	E	117	VAL
5	Q	7	ASP
4	N	166	PRO
4	G	155	VAL

### 5.3.2 Protein sidechains ⓘ

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	188/194 (97%)	182 (97%)	6 (3%)	46	80
1	C	188/194 (97%)	183 (97%)	5 (3%)	52	83
2	B	151/155 (97%)	146 (97%)	5 (3%)	45	79
2	E	151/155 (97%)	144 (95%)	7 (5%)	33	72
3	D	94/161 (58%)	92 (98%)	2 (2%)	61	86

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	F	94/161 (58%)	93 (99%)	1 (1%)	80	92
4	G	93/104 (89%)	91 (98%)	2 (2%)	60	86
4	N	92/104 (88%)	87 (95%)	5 (5%)	27	67
5	P	12/67 (18%)	12 (100%)	0	100	100
5	Q	14/67 (21%)	13 (93%)	1 (7%)	18	56
All	All	1077/1362 (79%)	1043 (97%)	34 (3%)	46	80

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	N	163	ASP
1	C	86	PHE
4	G	163	ASP
4	N	174	LYS
2	B	79	LYS

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

Mol	Chain	Res	Type
1	A	37	GLN
1	C	29	GLN

### 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

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	199/205 (97%)	-0.22	2 (1%) 84 78	29, 48, 74, 143	0
1	C	199/205 (97%)	-0.21	1 (0%) 91 88	30, 51, 81, 122	0
2	B	173/176 (98%)	-0.29	0 100 100	30, 45, 65, 79	0
2	E	173/176 (98%)	-0.24	0 100 100	27, 47, 65, 73	0
3	D	101/178 (56%)	-0.19	0 100 100	33, 55, 77, 111	0
3	F	101/178 (56%)	0.00	3 (2%) 54 44	34, 60, 96, 129	0
4	G	102/116 (87%)	0.05	3 (2%) 55 46	30, 68, 104, 124	0
4	N	101/116 (87%)	-0.11	1 (0%) 84 78	29, 61, 93, 105	0
5	P	13/76 (17%)	-0.34	0 100 100	52, 64, 74, 79	0
5	Q	15/76 (19%)	-0.18	0 100 100	56, 69, 96, 101	0
All	All	1177/1502 (78%)	-0.18	10 (0%) 87 82	27, 52, 89, 143	0

The worst 5 of 10 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	52	ASP	9.2
3	F	245	THR	4.4
1	C	52	ASP	3.5
4	G	168	PRO	3.1
3	F	310	THR	2.9

### 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.