



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 13, 2020 – 12:20 am BST

PDB ID : 2II7  
Title : Anabaena sensory rhodopsin transducer  
Authors : Vogeley, L.  
Deposited on : 2006-09-27  
Resolution : 2.80 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

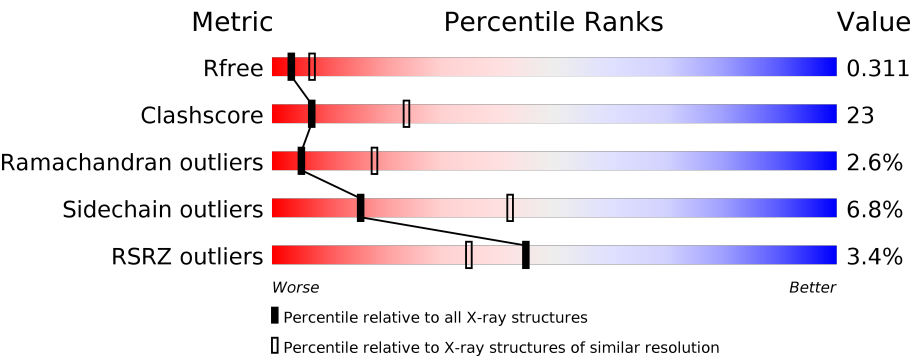
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	131	<div><div>8%</div><div><div></div><div></div><div></div><div></div><div></div></div><div>54%22%8% • 15%</div></div>
1	B	131	<div><div></div><div><div></div><div></div><div></div><div></div><div></div></div><div>50%31%5% • 12%</div></div>
1	C	131	<div><div>7%</div><div><div></div><div></div><div></div><div></div><div></div></div><div>50%26%8%16%</div></div>
1	D	131	<div><div>%</div><div><div></div><div></div><div></div><div></div><div></div></div><div>55%20% • • 21%</div></div>
1	E	131	<div><div>%</div><div><div></div><div></div><div></div><div></div><div></div></div><div>53%23%5% • 19%</div></div>
1	F	131	<div><div></div><div><div></div><div></div><div></div><div></div><div></div></div><div>50%15% • • 31%</div></div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	131	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>5%47%20%••31%</div></div>
1	H	131	<div><div><div></div><div></div><div></div><div></div><div></div></div><div>%55%20%5%•19%</div></div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 6555 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 Anabaena sensory rhodopsin transducer protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	112	Total	C	N	O	S	0	0	0
			881	555	155	169	2			
1	B	115	Total	C	N	O	S	0	0	0
			905	571	159	173	2			
1	C	110	Total	C	N	O	S	0	0	0
			858	538	152	166	2			
1	D	103	Total	C	N	O	S	0	0	0
			809	511	142	154	2			
1	E	106	Total	C	N	O	S	0	0	0
			830	520	147	161	2			
1	F	91	Total	C	N	O	S	0	0	0
			714	453	123	136	2			
1	G	91	Total	C	N	O	S	0	0	0
			714	453	123	136	2			
1	H	106	Total	C	N	O	S	0	0	0
			829	520	146	161	2			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	125	HIS	-	EXPRESSION TAG	UNP Q8YSC3
A	126	HIS	-	EXPRESSION TAG	UNP Q8YSC3
A	127	HIS	-	EXPRESSION TAG	UNP Q8YSC3
A	128	HIS	-	EXPRESSION TAG	UNP Q8YSC3
A	129	HIS	-	EXPRESSION TAG	UNP Q8YSC3
A	130	HIS	-	EXPRESSION TAG	UNP Q8YSC3
B	125	HIS	-	EXPRESSION TAG	UNP Q8YSC3
B	126	HIS	-	EXPRESSION TAG	UNP Q8YSC3
B	127	HIS	-	EXPRESSION TAG	UNP Q8YSC3
B	128	HIS	-	EXPRESSION TAG	UNP Q8YSC3
B	129	HIS	-	EXPRESSION TAG	UNP Q8YSC3
B	130	HIS	-	EXPRESSION TAG	UNP Q8YSC3
C	125	HIS	-	EXPRESSION TAG	UNP Q8YSC3

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	126	HIS	-	EXPRESSION TAG	UNP Q8YSC3
C	127	HIS	-	EXPRESSION TAG	UNP Q8YSC3
C	128	HIS	-	EXPRESSION TAG	UNP Q8YSC3
C	129	HIS	-	EXPRESSION TAG	UNP Q8YSC3
C	130	HIS	-	EXPRESSION TAG	UNP Q8YSC3
D	125	HIS	-	EXPRESSION TAG	UNP Q8YSC3
D	126	HIS	-	EXPRESSION TAG	UNP Q8YSC3
D	127	HIS	-	EXPRESSION TAG	UNP Q8YSC3
D	128	HIS	-	EXPRESSION TAG	UNP Q8YSC3
D	129	HIS	-	EXPRESSION TAG	UNP Q8YSC3
D	130	HIS	-	EXPRESSION TAG	UNP Q8YSC3
E	125	HIS	-	EXPRESSION TAG	UNP Q8YSC3
E	126	HIS	-	EXPRESSION TAG	UNP Q8YSC3
E	127	HIS	-	EXPRESSION TAG	UNP Q8YSC3
E	128	HIS	-	EXPRESSION TAG	UNP Q8YSC3
E	129	HIS	-	EXPRESSION TAG	UNP Q8YSC3
E	130	HIS	-	EXPRESSION TAG	UNP Q8YSC3
F	125	HIS	-	EXPRESSION TAG	UNP Q8YSC3
F	126	HIS	-	EXPRESSION TAG	UNP Q8YSC3
F	127	HIS	-	EXPRESSION TAG	UNP Q8YSC3
F	128	HIS	-	EXPRESSION TAG	UNP Q8YSC3
F	129	HIS	-	EXPRESSION TAG	UNP Q8YSC3
F	130	HIS	-	EXPRESSION TAG	UNP Q8YSC3
G	125	HIS	-	EXPRESSION TAG	UNP Q8YSC3
G	126	HIS	-	EXPRESSION TAG	UNP Q8YSC3
G	127	HIS	-	EXPRESSION TAG	UNP Q8YSC3
G	128	HIS	-	EXPRESSION TAG	UNP Q8YSC3
G	129	HIS	-	EXPRESSION TAG	UNP Q8YSC3
G	130	HIS	-	EXPRESSION TAG	UNP Q8YSC3
H	125	HIS	-	EXPRESSION TAG	UNP Q8YSC3
H	126	HIS	-	EXPRESSION TAG	UNP Q8YSC3
H	127	HIS	-	EXPRESSION TAG	UNP Q8YSC3
H	128	HIS	-	EXPRESSION TAG	UNP Q8YSC3
H	129	HIS	-	EXPRESSION TAG	UNP Q8YSC3
H	130	HIS	-	EXPRESSION TAG	UNP Q8YSC3

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total O 2 2	0	0
2	B	2	Total O 2 2	0	0

*Continued on next page...*

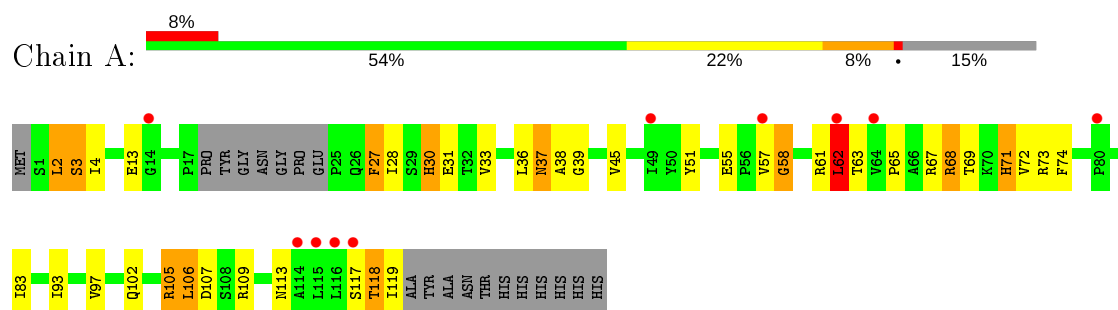
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	3	Total 3	O 3	0	0
2	E	5	Total 5	O 5	0	0
2	F	2	Total 2	O 2	0	0
2	H	1	Total 1	O 1	0	0

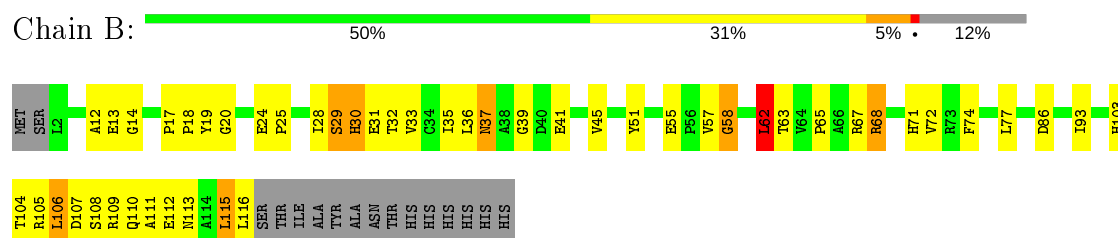
### 3 Residue-property plots

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

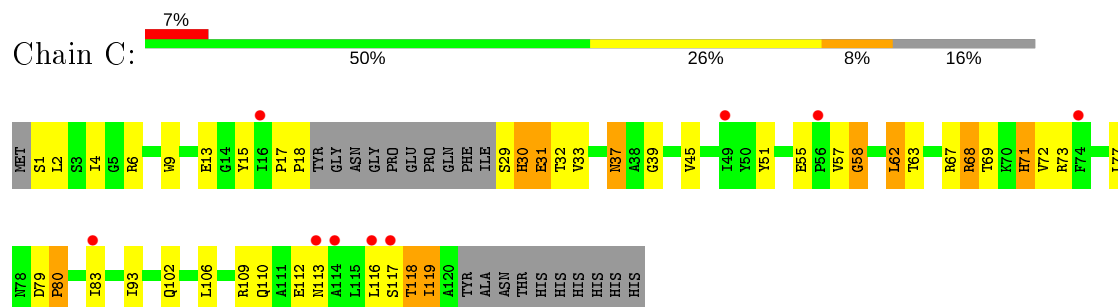
- Molecule 1: Anabaena sensory rhodopsin transducer protein



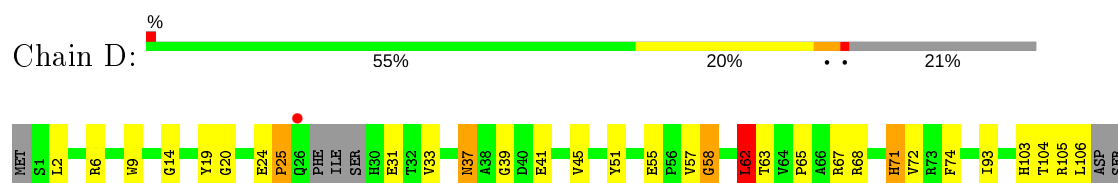
- Molecule 1: Anabaena sensory rhodopsin transducer protein



- Molecule 1: Anabaena sensory rhodopsin transducer protein



- Molecule 1: Anabaena sensory rhodopsin transducer protein



ARG  
GLN  
ALA  
GLU  
ASN  
ALA  
LEU  
LEU  
LEU  
SER  
THR  
ILE  
ALA  
TYR  
ALA  
ALA  
ASN  
THR  
THR  
HIS  
HIS  
HIS  
HIS  
HIS  
HIS

• Molecule 1: Anabaena sensory rhodopsin transducer protein



MET S1 S3 I4 I11 A12 E13 P18 Y19 G22 P23 GLU PRO GLN PHE ILE SER HIS E31 T32 V33 N37 A38 G39 V45 Y51 E55 P56 V57 G58 L62 P65 A66 R67 R68 T69 K70 H71 V72 R73 F74 N75 D76 L77 N78 A81 D86

I93 H103 T104 R106 L107 S108 Q109 A110 Q111 H112 H113 ALA LEU LEU SER THR ILE ALA TYR ASN THR HIS HIS HIS HIS HIS

• Molecule 1: Anabaena sensory rhodopsin transducer protein



MET SER L2 A12 E13 G14 Y15 I16 P17 P18 TYR GLY ASN GLY PRO GLU GLN PHE ILE SER HIS E31 N37 A38 G39 V45 Y51 E55 P56 V57 G58 L62 T63 V64 A66 R67 R68 T69 K70 H71 V72 L77 I93 T104 ARG LEU ASP SER ARG

GLN ALA GLU ALA LEU LEU LEU SER THR ILE TYR ALA ASN THR THR HIS HIS HIS HIS HIS HIS

• Molecule 1: Anabaena sensory rhodopsin transducer protein



MET SER L2 I7 I11 A12 E13 I16 P17 P18 TTR GLY ASN GLY PRO GLN PHE ILE SER HIS E31 T32 V33 N37 A38 G39 V45 Y51 K54 E55 P56 V57 G58 R61 L62 T63 V64 A66 R67 R68 T69 K70 H71 V72 R73 F74 L77

P80 T87 I93 N96 H103 T104 ARG LEU ASP SER ARG GLN ALA ALA ASN ALA LEU LEU SER THR ILE ALA TYR TTR ALA ALA ASN THR HIS HIS HIS HIS HIS HIS

• Molecule 1: Anabaena sensory rhodopsin transducer protein



MET S1 L2 I4 G14 Y15 I16 P17 TYR ASN GLY PRO GLU GLN PHE ILE SER HIS E31 N37 A38 G39 V45 Y51 E55 P56 V57 G58 L62 T63 R67 R68 T69 K70 H71 V72 I83 I93 Q102 H103 T104 R105 L106 R109

E112 H113 A114 I115 S117 T118 ILE TYR ASN THR HIS HIS HIS HIS HIS HIS HIS



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	73.83Å 122.33Å 130.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.80 40.88 – 2.60	Depositor EDS
% Data completeness (in resolution range)	99.8 (30.00-2.80) 99.8 (40.88-2.60)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.52 (at 2.61Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.259 , 0.315 0.257 , 0.311	Depositor DCC
$R_{free}$ test set	2602 reflections (7.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	55.7	Xtriage
Anisotropy	0.137	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 62.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.51$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	6555	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.44% 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	0.37	0/902	0.65	1/1233 (0.1%)
1	B	0.40	0/930	0.69	1/1275 (0.1%)
1	C	0.41	0/878	0.67	0/1202
1	D	0.42	0/832	0.68	1/1141 (0.1%)
1	E	0.48	0/851	0.72	1/1165 (0.1%)
1	F	0.40	0/733	0.66	1/1006 (0.1%)
1	G	0.38	0/733	0.64	1/1006 (0.1%)
1	H	0.39	0/848	0.68	1/1161 (0.1%)
All	All	0.41	0/6707	0.68	7/9189 (0.1%)

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	H	62	LEU	CA-CB-CG	6.14	129.42	115.30
1	B	62	LEU	CA-CB-CG	5.99	129.09	115.30
1	F	62	LEU	CA-CB-CG	5.94	128.96	115.30
1	A	62	LEU	CA-CB-CG	5.64	128.28	115.30
1	G	62	LEU	CA-CB-CG	5.59	128.16	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	881	0	862	56	0
1	B	905	0	872	55	0
1	C	858	0	838	52	0
1	D	809	0	780	35	0
1	E	830	0	799	37	0
1	F	714	0	692	31	0
1	G	714	0	692	30	0
1	H	829	0	810	41	0
2	A	2	0	0	1	0
2	B	2	0	0	1	0
2	D	3	0	0	1	0
2	E	5	0	0	0	0
2	F	2	0	0	1	0
2	H	1	0	0	0	0
All	All	6555	0	6345	295	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 23.

The worst 5 of 295 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:2:LEU:H	1:A:2:LEU:HD12	1.33	0.92
1:H:2:LEU:HD23	1:H:4:ILE:HD11	1.54	0.90
1:A:30:HIS:CD2	1:A:73:ARG:HH12	1.91	0.87
1:C:31:GLU:HG2	1:C:109:ARG:NH2	1.90	0.86
1:C:31:GLU:HG2	1:C:109:ARG:HH21	1.41	0.86

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	108/131 (82%)	94 (87%)	10 (9%)	4 (4%)	3	11
1	B	113/131 (86%)	99 (88%)	12 (11%)	2 (2%)	8	28
1	C	106/131 (81%)	97 (92%)	3 (3%)	6 (6%)	1	5
1	D	99/131 (76%)	89 (90%)	7 (7%)	3 (3%)	4	15
1	E	102/131 (78%)	91 (89%)	9 (9%)	2 (2%)	7	24
1	F	87/131 (66%)	80 (92%)	6 (7%)	1 (1%)	14	41
1	G	87/131 (66%)	81 (93%)	4 (5%)	2 (2%)	6	21
1	H	102/131 (78%)	93 (91%)	8 (8%)	1 (1%)	15	44
All	All	804/1048 (77%)	724 (90%)	59 (7%)	21 (3%)	5	18

5 of 21 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	31	GLU
1	A	58	GLY
1	A	118	THR
1	B	58	GLY
1	C	58	GLY

### 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	99/114 (87%)	90 (91%)	9 (9%)	9	27
1	B	100/114 (88%)	92 (92%)	8 (8%)	12	34
1	C	96/114 (84%)	91 (95%)	5 (5%)	23	55
1	D	90/114 (79%)	87 (97%)	3 (3%)	38	72
1	E	92/114 (81%)	85 (92%)	7 (8%)	13	36
1	F	80/114 (70%)	74 (92%)	6 (8%)	13	37
1	G	80/114 (70%)	75 (94%)	5 (6%)	18	46
1	H	93/114 (82%)	86 (92%)	7 (8%)	13	37

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	730/912 (80%)	680 (93%)	50 (7%)	16	42

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	37	ASN
1	E	68	ARG
1	H	71	HIS
1	D	62	LEU
1	E	37	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	85	HIS
1	E	75	ASN
1	H	37	ASN
1	D	102	GLN
1	E	85	HIS

### 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 [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	112/131 (85%)	0.32	10 (8%) <span>9</span> <span>5</span>	40, 69, 91, 97	0
1	B	115/131 (87%)	-0.05	0 <span>100</span> <span>100</span>	32, 56, 95, 113	0
1	C	110/131 (83%)	0.29	9 (8%) <span>11</span> <span>6</span>	20, 62, 97, 113	0
1	D	103/131 (78%)	-0.16	1 (0%) <span>82</span> <span>77</span>	24, 50, 96, 120	0
1	E	106/131 (80%)	-0.20	1 (0%) <span>84</span> <span>80</span>	17, 39, 101, 121	0
1	F	91/131 (69%)	-0.10	0 <span>100</span> <span>100</span>	28, 53, 87, 108	0
1	G	91/131 (69%)	0.28	6 (6%) <span>18</span> <span>11</span>	43, 69, 102, 113	0
1	H	106/131 (80%)	-0.16	1 (0%) <span>84</span> <span>80</span>	22, 56, 100, 113	0
All	All	834/1048 (79%)	0.03	28 (3%) <span>45</span> <span>35</span>	17, 59, 98, 121	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	117	SER	5.3
1	C	113	ASN	5.2
1	C	116	LEU	4.6
1	C	16	ILE	3.9
1	G	18	PRO	3.9

### 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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.