



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 22, 2020 – 03:21 am BST

PDB ID : 2II8  
Title : Anabaena sensory rhodopsin transducer  
Authors : Vogeley, L.; Sineshchekov, O.A.; Trivedi, V.D.; Spudich, E.N.; Spudich, J.L.; Luecke, H.  
Deposited on : 2006-09-27  
Resolution : 2.10 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

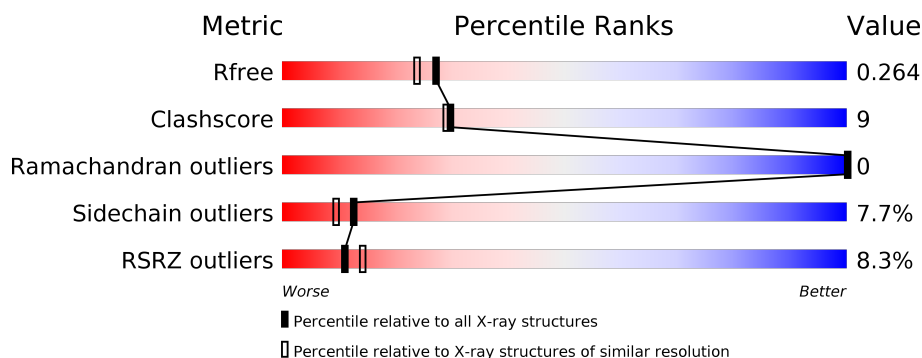
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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>6%</div> <div> <div></div> <div>60%</div> <div>13%</div> <div>•</div> <div>26%</div> </div> </div>
1	B	131	<div> <div>6%</div> <div> <div></div> <div>63%</div> <div>14%</div> <div>•</div> <div>21%</div> </div> </div>
1	C	131	<div> <div>6%</div> <div> <div></div> <div>57%</div> <div>10%</div> <div>•</div> <div>30%</div> </div> </div>
1	D	131	<div> <div>2%</div> <div> <div></div> <div>60%</div> <div>15%</div> <div>•</div> <div>25%</div> </div> </div>
1	E	131	<div> <div>3%</div> <div> <div></div> <div>55%</div> <div>12%</div> <div>••</div> <div>30%</div> </div> </div>
1	F	131	<div> <div>7%</div> <div> <div></div> <div>56%</div> <div>11%</div> <div>•</div> <div>31%</div> </div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	131	<div> <div>6%</div> <div> <div></div> <div>56%</div> <div>10%</div> <div>••</div> <div>30%</div> </div> </div>
1	H	131	<div> <div>11%</div> <div> <div></div> <div>59%</div> <div>11%</div> <div>•</div> <div>30%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 6525 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	97	Total	C	N	O	S	0	0	0
			764	485	132	145	2			
1	B	104	Total	C	N	O	S	0	0	0
			815	517	140	156	2			
1	C	92	Total	C	N	O	S	0	0	0
			723	457	126	138	2			
1	D	98	Total	C	N	O	S	0	0	0
			768	486	133	147	2			
1	E	92	Total	C	N	O	S	0	0	0
			720	456	124	138	2			
1	F	91	Total	C	N	O	S	0	0	0
			713	452	123	136	2			
1	G	92	Total	C	N	O	S	0	0	0
			720	456	124	138	2			
1	H	92	Total	C	N	O	S	0	0	0
			720	456	124	138	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	71	Total O 71 71	0	0
2	B	81	Total O 81 81	0	0

*Continued on next page...*

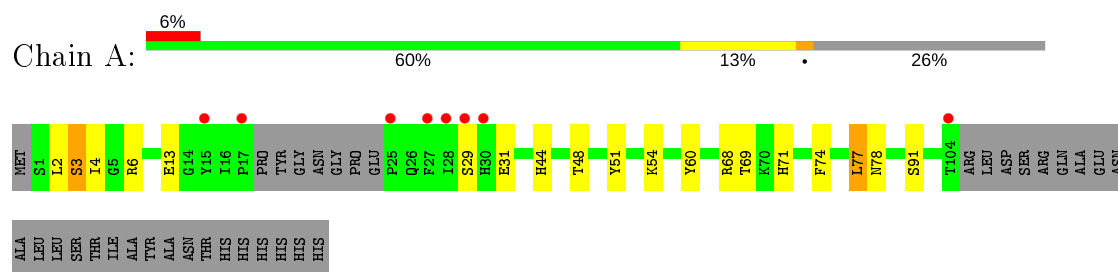
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	C	86	Total 86	O 86	0	0
2	D	73	Total 73	O 73	0	0
2	E	70	Total 70	O 70	0	0
2	F	73	Total 73	O 73	0	0
2	G	80	Total 80	O 80	0	0
2	H	48	Total 48	O 48	0	0

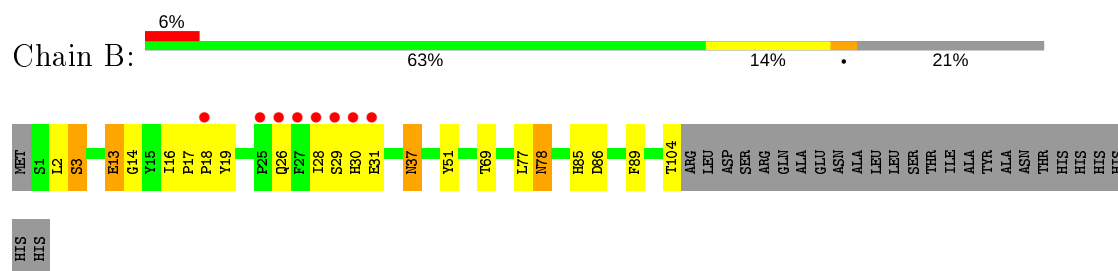
### 3 Residue-property plots [i](#)

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.

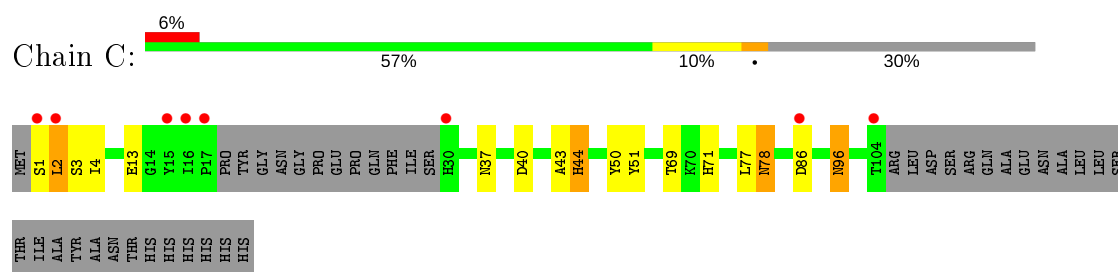
- Molecule 1: Anabaena sensory rhodopsin transducer protein



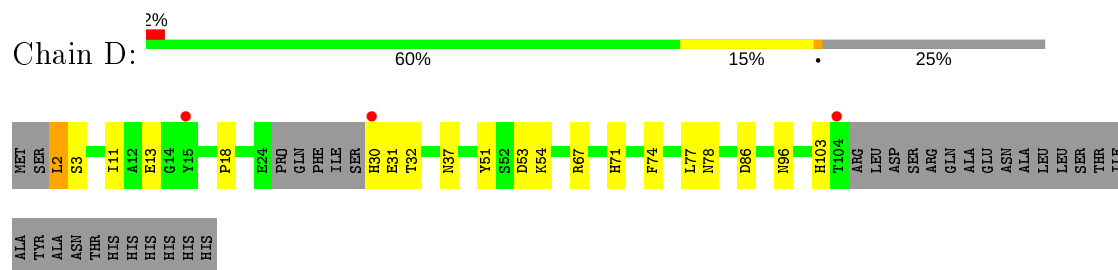
- Molecule 1: Anabaena sensory rhodopsin transducer protein



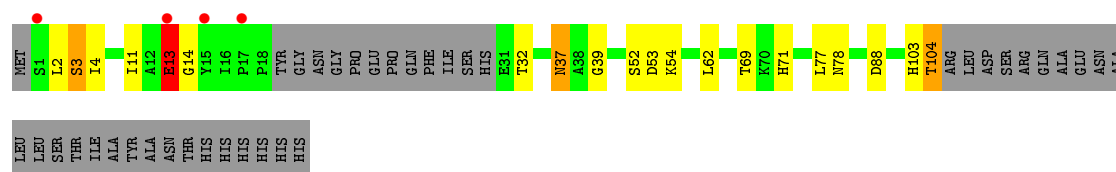
- Molecule 1: Anabaena sensory rhodopsin transducer protein



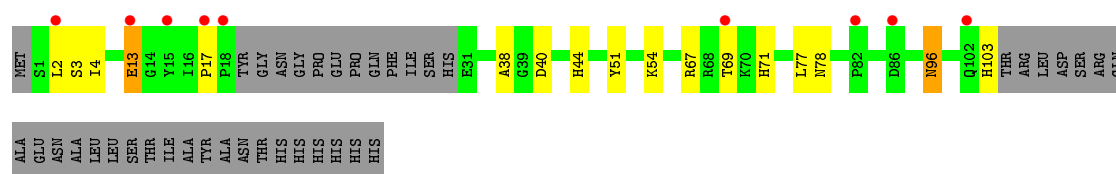
- Molecule 1: Anabaena sensory rhodopsin transducer protein



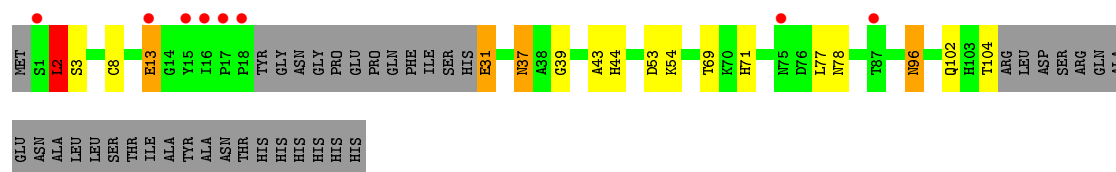
- Molecule 1: Anabaena sensory rhodopsin transducer protein



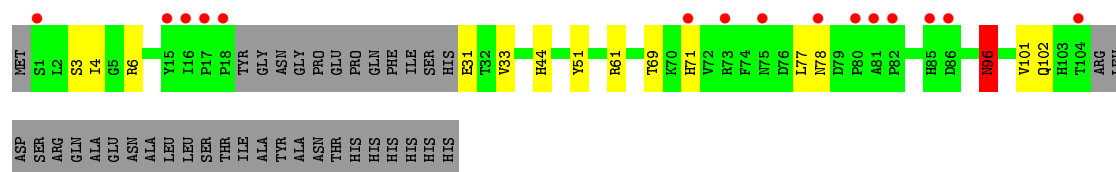
- Molecule 1: Anabaena sensory rhodopsin transducer protein



- Molecule 1: Anabaena sensory rhodopsin transducer protein



- Molecule 1: Anabaena sensory rhodopsin transducer protein





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.80 Å 119.24 Å 120.79 Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	35.22 – 2.10 35.23 – 2.10	Depositor EDS
% Data completeness (in resolution range)	97.5 (35.22-2.10) 97.4 (35.23-2.10)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.43 (at 2.10 Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.228 , 0.263 0.228 , 0.264	Depositor DCC
$R_{free}$ test set	4287 reflections (7.09%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	34.5	Xtriage
Anisotropy	0.016	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 53.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.013 for -h,l,k	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	6525	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.78% 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	1.20	1/785 (0.1%)	0.97	2/1075 (0.2%)
1	B	1.08	1/840 (0.1%)	0.91	0/1154
1	C	1.06	1/742 (0.1%)	0.89	0/1017
1	D	1.04	1/790 (0.1%)	0.98	3/1084 (0.3%)
1	E	1.04	1/739 (0.1%)	0.93	2/1014 (0.2%)
1	F	1.08	1/732 (0.1%)	0.92	1/1004 (0.1%)
1	G	1.13	2/739 (0.3%)	1.00	2/1014 (0.2%)
1	H	1.05	1/739 (0.1%)	0.88	0/1014
All	All	1.09	9/6106 (0.1%)	0.94	10/8376 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	F	0	1

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	G	13	GLU	CB-CG	8.05	1.67	1.52
1	E	13	GLU	CB-CG	6.90	1.65	1.52
1	H	96	ASN	CG-ND2	-6.24	1.17	1.32
1	F	13	GLU	CB-CG	6.01	1.63	1.52
1	G	8	CYS	CB-SG	5.83	1.92	1.82

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	68	ARG	NE-CZ-NH1	8.16	124.38	120.30
1	D	2	LEU	CA-CB-CG	7.32	132.15	115.30
1	E	88	ASP	CB-CG-OD2	-6.37	112.57	118.30

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	68	ARG	NE-CZ-NH2	-6.08	117.26	120.30
1	G	2	LEU	CA-CB-CG	6.03	129.18	115.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	F	17	PRO	Peptide

## 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	764	0	741	15	0
1	B	815	0	782	19	0
1	C	723	0	700	16	1
1	D	768	0	733	11	0
1	E	720	0	700	20	0
1	F	713	0	693	18	0
1	G	720	0	700	20	0
1	H	720	0	700	15	1
2	A	71	0	0	7	0
2	B	81	0	0	3	0
2	C	86	0	0	4	0
2	D	73	0	0	0	0
2	E	70	0	0	2	0
2	F	73	0	0	3	0
2	G	80	0	0	6	0
2	H	48	0	0	0	0
All	All	6525	0	5749	110	1

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 9.

The worst 5 of 110 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:G:104:THR:HG21	2:G:143:HOH:O	1.37	1.20
1:D:71:HIS:CE1	1:G:13:GLU:OE1	1.97	1.18
1:C:71:HIS:CE1	1:E:13:GLU:OE1	2.04	1.11
1:G:31:GLU:HB3	2:G:160:HOH:O	1.66	0.95
1:H:96:ASN:H	1:H:96:ASN:HD22	1.17	0.90

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:86:ASP:OD2	1:H:61:ARG:NH2[2_545]	2.15	0.05

## 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	93/131 (71%)	89 (96%)	4 (4%)	0	100	100
1	B	102/131 (78%)	99 (97%)	3 (3%)	0	100	100
1	C	88/131 (67%)	84 (96%)	4 (4%)	0	100	100
1	D	94/131 (72%)	91 (97%)	3 (3%)	0	100	100
1	E	88/131 (67%)	84 (96%)	4 (4%)	0	100	100
1	F	87/131 (66%)	84 (97%)	3 (3%)	0	100	100
1	G	88/131 (67%)	85 (97%)	3 (3%)	0	100	100
1	H	88/131 (67%)	85 (97%)	3 (3%)	0	100	100
All	All	728/1048 (70%)	701 (96%)	27 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	86/114 (75%)	81 (94%)	5 (6%)	20	17
1	B	91/114 (80%)	83 (91%)	8 (9%)	10	6
1	C	81/114 (71%)	74 (91%)	7 (9%)	10	7
1	D	85/114 (75%)	79 (93%)	6 (7%)	14	11
1	E	81/114 (71%)	74 (91%)	7 (9%)	10	7
1	F	80/114 (70%)	75 (94%)	5 (6%)	18	15
1	G	81/114 (71%)	73 (90%)	8 (10%)	8	5
1	H	81/114 (71%)	76 (94%)	5 (6%)	18	15
All	All	666/912 (73%)	615 (92%)	51 (8%)	13	9

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	54	LYS
1	E	37	ASN
1	H	31	GLU
1	D	78	ASN
1	E	54	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	71	HIS
1	E	78	ASN
1	H	78	ASN
1	D	78	ASN
1	B	75	ASN

### 5.3.3 RNA ⓘ

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 [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	97/131 (74%)	0.85	8 (8%) 11 15	21, 33, 78, 93	0
1	B	104/131 (79%)	0.57	8 (7%) 13 17	22, 35, 69, 83	0
1	C	92/131 (70%)	0.29	8 (8%) 10 13	21, 33, 56, 66	0
1	D	98/131 (74%)	0.36	3 (3%) 49 55	22, 34, 57, 70	0
1	E	92/131 (70%)	0.28	4 (4%) 35 41	22, 33, 55, 62	0
1	F	91/131 (69%)	0.64	9 (9%) 7 9	21, 33, 56, 62	0
1	G	92/131 (70%)	0.32	8 (8%) 10 13	21, 33, 55, 61	0
1	H	92/131 (70%)	0.73	15 (16%) 1 2	22, 33, 58, 62	0
All	All	758/1048 (72%)	0.51	63 (8%) 11 14	21, 34, 58, 93	0

The worst 5 of 63 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	27	PHE	14.1
1	F	17	PRO	7.1
1	A	28	ILE	7.1
1	F	18	PRO	6.4
1	B	30	HIS	5.9

### 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 [i](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.