



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 10:42 pm BST

PDB ID : 5HLG  
Title : Structure of reduced AbfR bound to DNA  
Authors : Liu, G.; Liu, X.; Gan, J.; Yang, C.G.  
Deposited on : 2016-01-15  
Resolution : 3.00 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

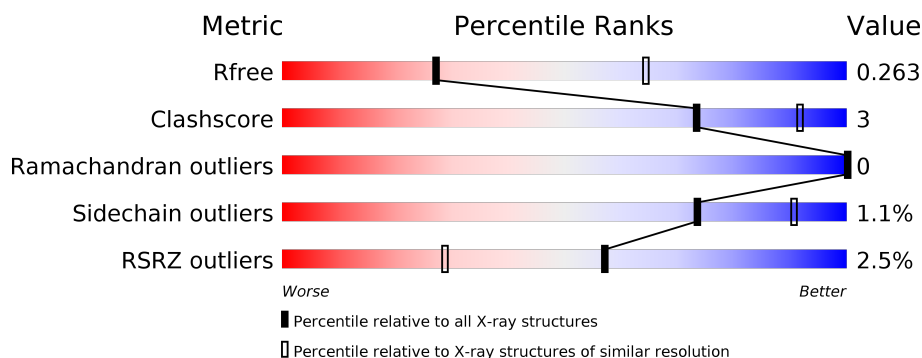
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.00 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	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (3.00-3.00)

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	I	24	<div> <div>4%</div> <div> <div></div> <div>92%</div> <div>8%</div> </div> </div>
1	J	24	<div> <div>4%</div> <div> <div></div> <div>92%</div> <div>8%</div> </div> </div>
1	K	24	<div> <div></div> <div> <div></div> <div>88%</div> <div>13%</div> </div> </div>
1	L	24	<div> <div></div> <div> <div></div> <div>83%</div> <div>17%</div> </div> </div>
1	M	24	<div> <div></div> <div> <div></div> <div>92%</div> <div>8%</div> </div> </div>
1	N	24	<div> <div></div> <div> <div></div> <div>88%</div> <div>13%</div> </div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	O	24	<div><div></div><div>79%21%</div></div>
1	P	24	<div><div>4%</div><div></div><div>88%13%</div></div>
2	A	148	<div><div>5%</div><div></div><div>82%8%9%</div></div>
2	B	148	<div><div>%</div><div></div><div>89%8%</div></div>
2	C	148	<div><div>%</div><div></div><div>82%9%9%</div></div>
2	D	148	<div><div>2%</div><div></div><div>88%6%6%</div></div>
2	E	148	<div><div>2%</div><div></div><div>85%7%7%</div></div>
2	F	148	<div><div>%</div><div></div><div>86%7%7%</div></div>
2	G	148	<div><div>6%</div><div></div><div>83%9%8%</div></div>
2	H	148	<div><div>%</div><div></div><div>84%9%7%</div></div>

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 13024 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 DNA chain called DNA (5'-D(\*TP\*AP\*AP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*GP\*CP\*GP\*CP\*GP\*CP\*GP\*AP\*TP\*TP\*GP\*AP\*GP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	I	24	Total	C	N	O	P	0	0	0
			489	234	90	142	23			
1	J	24	Total	C	N	O	P	0	0	0
			489	234	90	142	23			
1	K	24	Total	C	N	O	P	0	0	0
			489	234	90	142	23			
1	L	24	Total	C	N	O	P	0	0	0
			489	234	90	142	23			
1	M	24	Total	C	N	O	P	0	0	0
			489	234	90	142	23			
1	N	24	Total	C	N	O	P	0	0	0
			489	234	90	142	23			
1	O	24	Total	C	N	O	P	0	0	0
			489	234	90	142	23			
1	P	24	Total	C	N	O	P	0	0	0
			489	234	90	142	23			

- Molecule 2 is a protein called MarR family transcriptional regulator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	134	Total	C	N	O	S	0	0	0
			1114	706	192	212	4			
2	B	143	Total	C	N	O	S	0	0	0
			1180	750	201	224	5			
2	C	134	Total	C	N	O	S	0	0	0
			1116	708	190	214	4			
2	D	139	Total	C	N	O	S	0	0	0
			1137	720	196	217	4			
2	E	137	Total	C	N	O	S	0	0	0
			1135	718	195	217	5			
2	F	138	Total	C	N	O	S	0	0	0
			1147	726	197	220	4			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	G	136	Total	C	N	O	S	0	0	0
			1134	718	193	219	4			
2	H	138	Total	C	N	O	S	0	0	0
			1149	729	196	219	5			

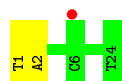
There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	CYS	-	expression tag	UNP A0A0N1EJ89
A	0	ASN	-	expression tag	UNP A0A0N1EJ89
A	72	MET	LEU	engineered mutation	UNP A0A0N1EJ89
B	-1	CYS	-	expression tag	UNP A0A0N1EJ89
B	0	ASN	-	expression tag	UNP A0A0N1EJ89
B	72	MET	LEU	engineered mutation	UNP A0A0N1EJ89
C	-1	CYS	-	expression tag	UNP A0A0N1EJ89
C	0	ASN	-	expression tag	UNP A0A0N1EJ89
C	72	MET	LEU	engineered mutation	UNP A0A0N1EJ89
D	-1	CYS	-	expression tag	UNP A0A0N1EJ89
D	0	ASN	-	expression tag	UNP A0A0N1EJ89
D	72	MET	LEU	engineered mutation	UNP A0A0N1EJ89
E	-1	CYS	-	expression tag	UNP A0A0N1EJ89
E	0	ASN	-	expression tag	UNP A0A0N1EJ89
E	72	MET	LEU	engineered mutation	UNP A0A0N1EJ89
F	-1	CYS	-	expression tag	UNP A0A0N1EJ89
F	0	ASN	-	expression tag	UNP A0A0N1EJ89
F	72	MET	LEU	engineered mutation	UNP A0A0N1EJ89
G	-1	CYS	-	expression tag	UNP A0A0N1EJ89
G	0	ASN	-	expression tag	UNP A0A0N1EJ89
G	72	MET	LEU	engineered mutation	UNP A0A0N1EJ89
H	-1	CYS	-	expression tag	UNP A0A0N1EJ89
H	0	ASN	-	expression tag	UNP A0A0N1EJ89
H	72	MET	LEU	engineered mutation	UNP A0A0N1EJ89

### 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.

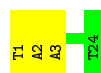
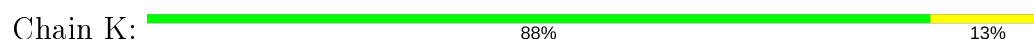
- Molecule 1: DNA (5'-D(\*TP\*AP\*AP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*GP\*CP\*GP\*CP\*GP\*CP\*GP\*AP\*TP\*TP\*GP\*AP\*GP\*T)-3')



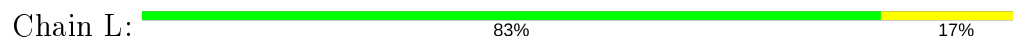
- Molecule 1: DNA (5'-D(\*TP\*AP\*AP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*GP\*CP\*GP\*CP\*GP\*CP\*GP\*AP\*TP\*TP\*GP\*AP\*GP\*T)-3')



- Molecule 1: DNA (5'-D(\*TP\*AP\*AP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*GP\*CP\*GP\*CP\*GP\*CP\*GP\*AP\*TP\*TP\*GP\*AP\*GP\*T)-3')



- Molecule 1: DNA (5'-D(\*TP\*AP\*AP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*GP\*CP\*GP\*CP\*GP\*CP\*GP\*AP\*TP\*TP\*GP\*AP\*GP\*T)-3')



- Molecule 1: DNA (5'-D(\*TP\*AP\*AP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*GP\*CP\*GP\*CP\*GP\*CP\*GP\*AP\*TP\*TP\*GP\*AP\*GP\*T)-3')





- Molecule 1: DNA (5'-D(\*TP\*AP\*AP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*GP\*CP\*GP\*CP\*GP\*CP\*GP\*AP\*TP\*TP\*GP\*AP\*GP\*T)-3')

Chain N: 88% 13%



- Molecule 1: DNA (5'-D(\*TP\*AP\*AP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*GP\*CP\*GP\*CP\*GP\*CP\*GP\*AP\*TP\*TP\*GP\*AP\*GP\*T)-3')

Chain O: 79% 21%



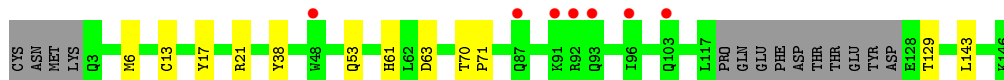
- Molecule 1: DNA (5'-D(\*TP\*AP\*AP\*CP\*TP\*CP\*AP\*AP\*TP\*CP\*GP\*CP\*GP\*CP\*GP\*CP\*GP\*AP\*TP\*TP\*GP\*AP\*GP\*T)-3')

Chain P: 4% 88% 13%



- Molecule 2: MarR family transcriptional regulator

Chain A: 5% 82% 8% 9%



- Molecule 2: MarR family transcriptional regulator

Chain B: % 89% 8% .

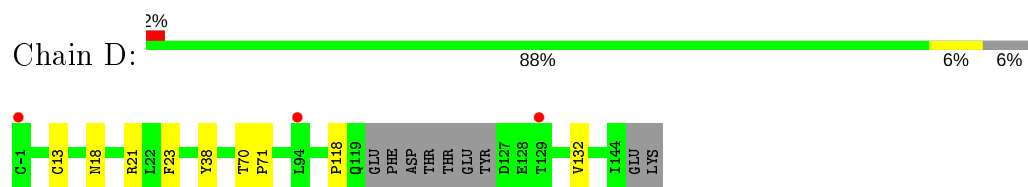


- Molecule 2: MarR family transcriptional regulator

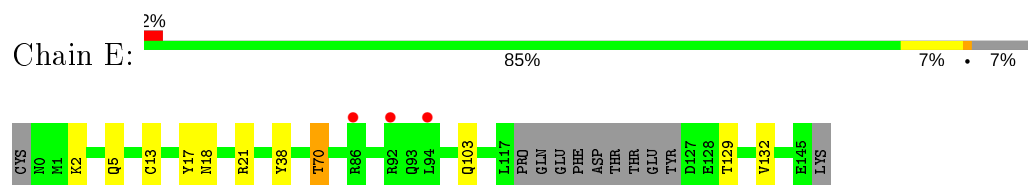
Chain C: % 82% 9% 9%



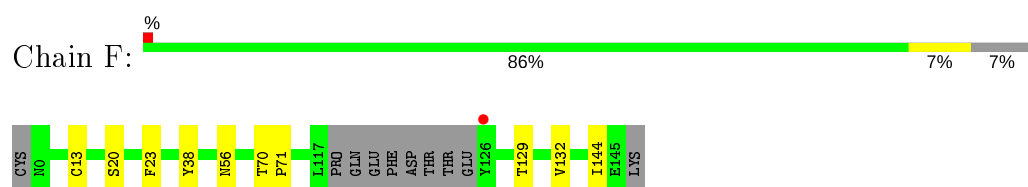
- Molecule 2: MarR family transcriptional regulator



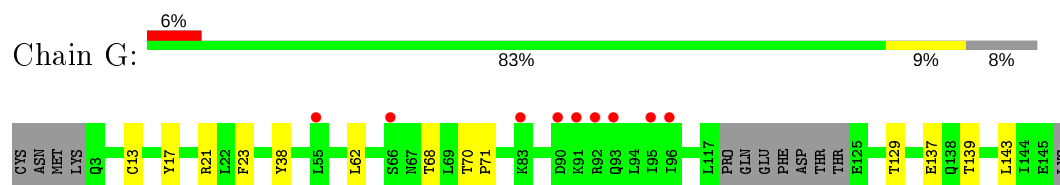
- Molecule 2: MarR family transcriptional regulator



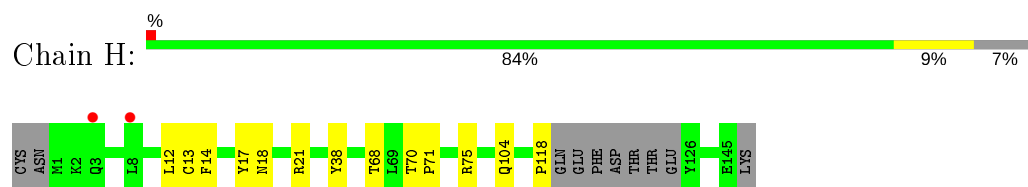
- Molecule 2: MarR family transcriptional regulator



- Molecule 2: MarR family transcriptional regulator



- Molecule 2: MarR family transcriptional regulator





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	150.32Å 290.50Å 52.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 3.00 49.63 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.0 (30.00-3.00) 98.8 (49.63-3.00)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.01 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.226 , 0.257 0.232 , 0.263	Depositor DCC
$R_{free}$ test set	2381 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	66.5	Xtriage
Anisotropy	0.683	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.31 , 53.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	13024	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	92.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.49% 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	I	0.41	0/548	0.79	0/844
1	J	0.42	0/548	0.78	0/844
1	K	0.52	1/548 (0.2%)	0.85	1/844 (0.1%)
1	L	0.56	0/548	0.84	0/844
1	M	0.44	0/548	0.77	0/844
1	N	0.47	0/548	0.77	0/844
1	O	0.39	0/548	0.77	0/844
1	P	0.50	0/548	0.79	0/844
2	A	0.45	0/1131	0.64	0/1521
2	B	0.45	0/1200	0.66	0/1616
2	C	0.46	0/1134	0.66	0/1527
2	D	0.48	0/1155	0.66	0/1557
2	E	0.46	0/1152	0.67	0/1549
2	F	0.47	0/1165	0.63	0/1568
2	G	0.41	0/1152	0.66	0/1551
2	H	0.52	0/1168	0.69	0/1572
All	All	0.46	1/13641 (0.0%)	0.71	1/19213 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	K	1	DT	C5'-C4'	-7.04	1.43	1.51

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	K	1	DT	C5'-C4'-C3'	5.82	124.58	114.10

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	I	489	0	272	3	0
1	J	489	0	272	1	0
1	K	489	0	272	2	0
1	L	489	0	272	5	0
1	M	489	0	272	2	0
1	N	489	0	272	3	0
1	O	489	0	272	5	0
1	P	489	0	272	3	0
2	A	1114	0	1116	11	0
2	B	1180	0	1165	13	0
2	C	1116	0	1108	13	0
2	D	1137	0	1123	10	0
2	E	1135	0	1134	11	0
2	F	1147	0	1140	10	0
2	G	1134	0	1122	16	0
2	H	1149	0	1148	19	0
All	All	13024	0	11232	81	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 3.

The worst 5 of 81 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)
2:E:2:LYS:HA	2:E:5:GLN:HG2	1.49	0.94
2:G:143:LEU:HD21	2:H:118:PRO:HD3	1.49	0.93
2:E:38:TYR:HH	2:F:13:CYS:HG	1.24	0.84
2:H:18:ASN:HA	2:H:21:ARG:HH21	1.44	0.82
2:A:13:CYS:SG	2:B:38:TYR:OH	2.30	0.78

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	
2	A	130/148 (88%)	127 (98%)	3 (2%)	0	100	100
2	B	139/148 (94%)	137 (99%)	2 (1%)	0	100	100
2	C	130/148 (88%)	128 (98%)	2 (2%)	0	100	100
2	D	135/148 (91%)	130 (96%)	5 (4%)	0	100	100
2	E	133/148 (90%)	131 (98%)	2 (2%)	0	100	100
2	F	134/148 (90%)	132 (98%)	2 (2%)	0	100	100
2	G	132/148 (89%)	131 (99%)	1 (1%)	0	100	100
2	H	134/148 (90%)	133 (99%)	1 (1%)	0	100	100
All	All	1067/1184 (90%)	1049 (98%)	18 (2%)	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	
2	A	125/139 (90%)	124 (99%)	1 (1%)	81	93
2	B	130/139 (94%)	129 (99%)	1 (1%)	81	93
2	C	125/139 (90%)	124 (99%)	1 (1%)	81	93
2	D	125/139 (90%)	124 (99%)	1 (1%)	81	93
2	E	127/139 (91%)	125 (98%)	2 (2%)	62	86
2	F	128/139 (92%)	125 (98%)	3 (2%)	50	80

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	G	127/139 (91%)	126 (99%)	1 (1%)	81	93
2	H	129/139 (93%)	128 (99%)	1 (1%)	81	93
All	All	1016/1112 (91%)	1005 (99%)	11 (1%)	73	90

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	70	THR
2	E	103	GLN
2	F	56	ASN
2	D	23	PHE
2	F	23	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	101	ASN
2	G	105	GLN
2	E	93	GLN
2	C	5	GLN
2	D	105	GLN

### 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	I	24/24 (100%)	0.38	1 (4%)	36	14	62, 83, 107, 121	0
1	J	24/24 (100%)	0.46	1 (4%)	36	14	65, 80, 111, 125	0
1	K	24/24 (100%)	0.08	0	100	100	54, 76, 103, 108	0
1	L	24/24 (100%)	0.34	0	100	100	58, 75, 82, 102	0
1	M	24/24 (100%)	0.07	0	100	100	51, 75, 96, 101	0
1	N	24/24 (100%)	0.16	0	100	100	56, 75, 84, 87	0
1	O	24/24 (100%)	0.33	0	100	100	58, 87, 96, 104	0
1	P	24/24 (100%)	0.57	1 (4%)	36	14	67, 79, 100, 103	0
2	A	134/148 (90%)	0.43	7 (5%)	27	10	66, 114, 152, 177	0
2	B	143/148 (96%)	0.21	2 (1%)	75	49	54, 89, 154, 199	0
2	C	134/148 (90%)	0.14	2 (1%)	73	46	61, 99, 141, 156	0
2	D	139/148 (93%)	0.11	3 (2%)	62	33	55, 82, 136, 158	0
2	E	137/148 (92%)	0.14	3 (2%)	62	33	55, 88, 133, 155	0
2	F	138/148 (93%)	0.09	1 (0%)	87	69	53, 80, 131, 150	0
2	G	136/148 (91%)	0.43	9 (6%)	18	5	65, 117, 145, 165	0
2	H	138/148 (93%)	0.10	2 (1%)	75	49	57, 86, 157, 180	0
All	All	1291/1376 (93%)	0.22	32 (2%)	57	29	51, 90, 143, 199	0

The worst 5 of 32 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	5	GLN	6.4
2	A	93	GLN	4.6
2	G	95	ILE	4.0
2	A	92	ARG	3.7
2	D	-1	CYS	3.7

## 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.