



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 6, 2020 – 10:04 PM BST

PDB ID : 6LYV  
Title : The crystal structure of SAUGI/KSHVUDG complex  
Authors : Liao, Y.T.; Ko, T.P.; Wang, H.C.  
Deposited on : 2020-02-16  
Resolution : 2.70 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 : 2.13.1  
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) : 2.13.1

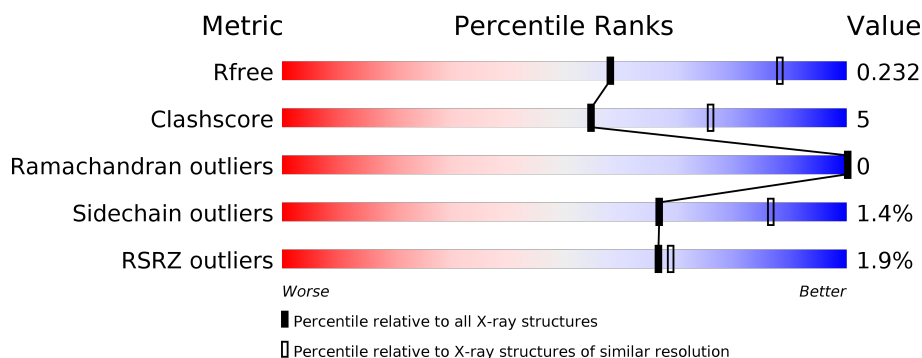
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	255	<div> <div>80%</div> <div>11%</div> <div>8%</div> </div>
1	C	255	<div> <div>79%</div> <div>11%</div> <div>8%</div> </div>
1	E	255	<div> <div>77%</div> <div>14%</div> <div>8%</div> </div>
1	G	255	<div> <div>80%</div> <div>10%</div> <div>9%</div> </div>
1	I	255	<div> <div>78%</div> <div>13%</div> <div>9%</div> </div>
1	K	255	<div> <div>78%</div> <div>13%</div> <div>8%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	M	255	 2% 82% 9% 9%
1	O	255	 6% 76% 15% 9%
2	B	112	 2% 77% 13% 11%
2	D	112	 % 70% 19% 11%
2	F	112	 % 77% 9% 13%
2	H	112	 3% 75% 13% 12%
2	J	112	 3% 78% 12% 11%
2	L	112	 3% 69% 19% 12%
2	N	112	 71% 16% 13%
2	P	112	 83% 5% 11%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 22519 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 Uracil-DNA glycosylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	234	Total	C	N	O	S	0	0	0
			1883	1206	338	332	7			
1	C	234	Total	C	N	O	S	0	0	0
			1883	1206	338	332	7			
1	E	234	Total	C	N	O	S	0	0	0
			1881	1206	338	330	7			
1	G	233	Total	C	N	O	S	0	0	0
			1875	1202	337	329	7			
1	I	233	Total	C	N	O	S	0	0	0
			1874	1201	337	329	7			
1	K	234	Total	C	N	O	S	0	0	0
			1885	1207	339	332	7			
1	M	233	Total	C	N	O	S	0	0	0
			1874	1201	337	329	7			
1	O	233	Total	C	N	O	S	0	0	0
			1874	1201	337	329	7			

- Molecule 2 is a protein called SAUGI.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	100	Total	C	N	O	S	0	0	0
			846	551	128	163	4			
2	D	100	Total	C	N	O	S	0	0	0
			846	551	128	163	4			
2	F	97	Total	C	N	O	S	0	0	0
			823	535	125	160	3			
2	H	99	Total	C	N	O	S	0	0	0
			839	546	127	162	4			
2	J	100	Total	C	N	O	S	0	0	0
			846	551	128	163	4			
2	L	99	Total	C	N	O	S	0	0	0
			839	546	127	162	4			

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	N	97	Total	C	N	O	S	0	0	0
			823	535	125	160	3			
2	P	100	Total	C	N	O	S	0	0	0
			846	551	128	163	4			

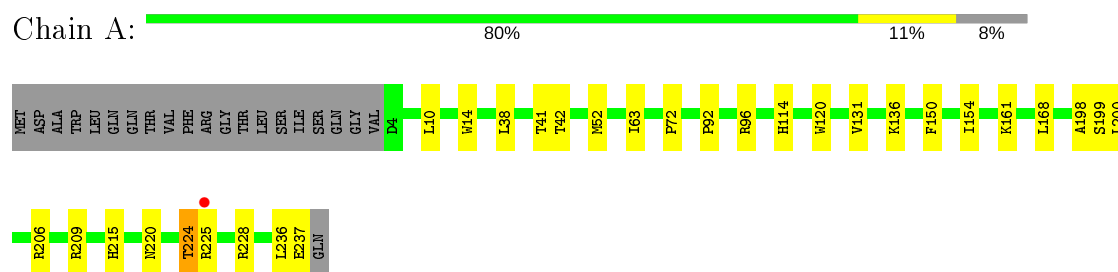
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	104	Total	O	0	0
			104	104		
3	B	22	Total	O	0	0
			22	22		
3	C	96	Total	O	0	0
			96	96		
3	D	34	Total	O	0	0
			34	34		
3	E	73	Total	O	0	0
			73	73		
3	F	29	Total	O	0	0
			29	29		
3	G	90	Total	O	0	0
			90	90		
3	H	32	Total	O	0	0
			32	32		
3	I	86	Total	O	0	0
			86	86		
3	J	10	Total	O	0	0
			10	10		
3	K	68	Total	O	0	0
			68	68		
3	L	28	Total	O	0	0
			28	28		
3	M	49	Total	O	0	0
			49	49		
3	N	21	Total	O	0	0
			21	21		
3	O	31	Total	O	0	0
			31	31		
3	P	9	Total	O	0	0
			9	9		

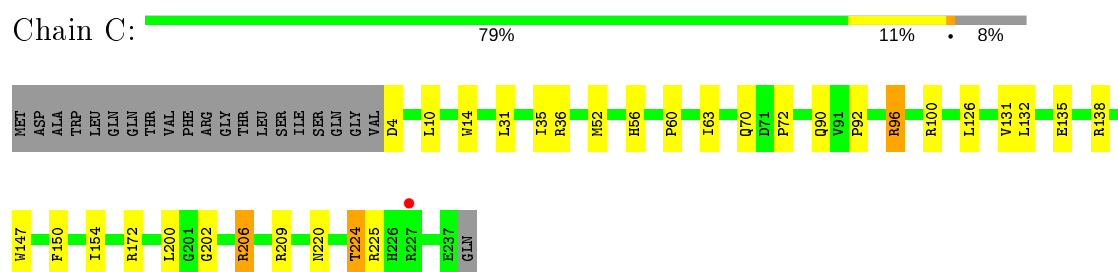
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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.

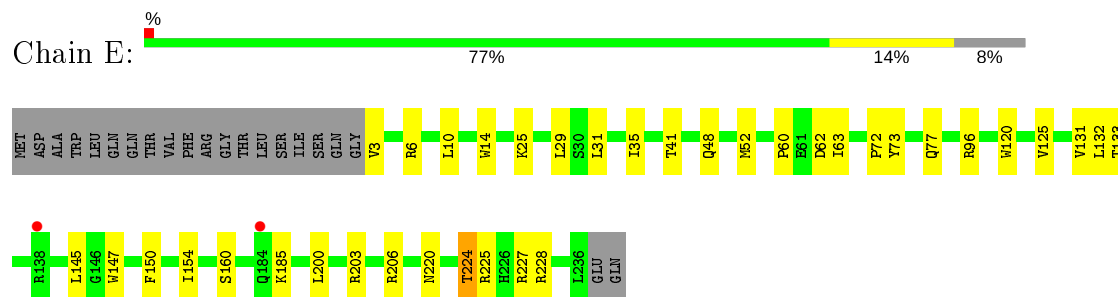
- Molecule 1: Uracil-DNA glycosylase



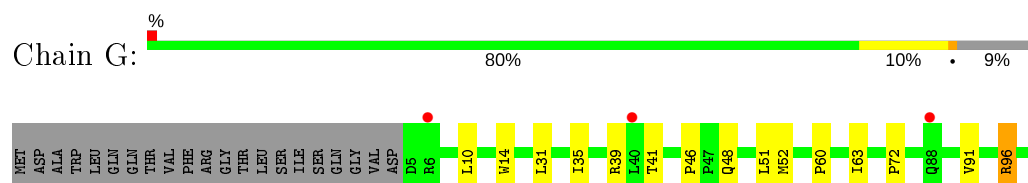
- Molecule 1: Uracil-DNA glycosylase



- Molecule 1: Uracil-DNA glycosylase

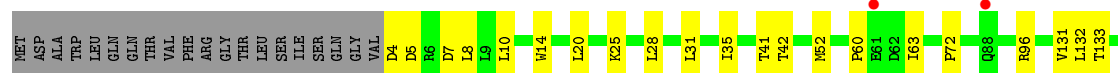
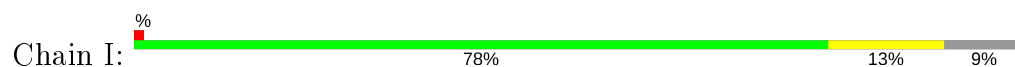


- Molecule 1: Uracil-DNA glycosylase

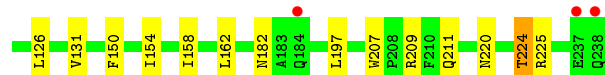
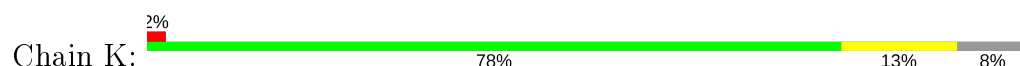




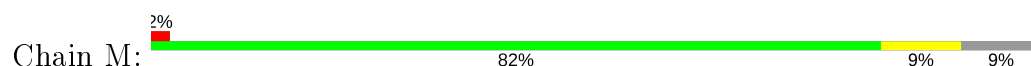
- Molecule 1: Uracil-DNA glycosylase



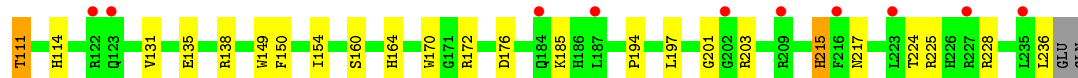
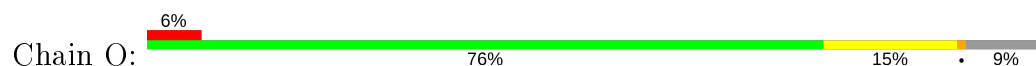
- Molecule 1: Uracil-DNA glycosylase



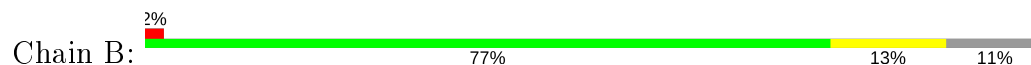
- Molecule 1: Uracil-DNA glycosylase



- Molecule 1: Uracil-DNA glycosylase



- Molecule 2: SAUGI

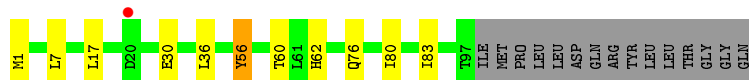




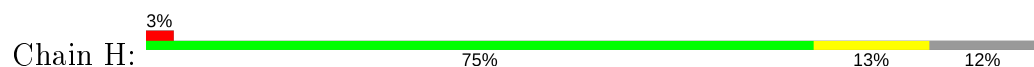
• Molecule 2: SAUGI



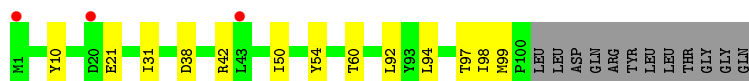
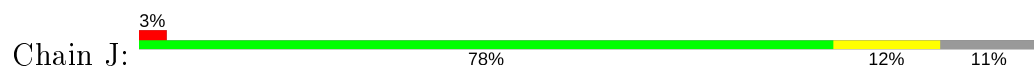
• Molecule 2: SAUGI



• Molecule 2: SAUGI



• Molecule 2: SAUGI



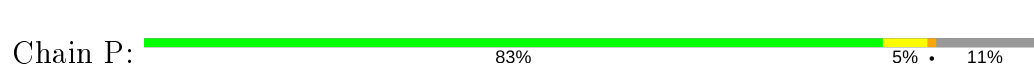
• Molecule 2: SAUGI



• Molecule 2: SAUGI



• Molecule 2: SAUGI



W1	T12	L17	E21	I35	Y56	Y57	T60	P100	LEU	LEU	ASP	GLN	ARG	TYR	LEU	LEU	THR	GLY	GLN
----	-----	-----	-----	-----	-----	-----	-----	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	85.89Å 92.20Å 98.11Å 108.14° 90.61° 92.23°	Depositor
Resolution (Å)	19.96 – 2.70 19.96 – 2.70	Depositor EDS
% Data completeness (in resolution range)	93.8 (19.96-2.70) 93.9 (19.96-2.70)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.17 (at 2.71Å)	Xtriage
Refinement program	PHENIX 1.17.1_3660	Depositor
R, $R_{free}$	0.192 , 0.232 0.192 , 0.232	Depositor DCC
$R_{free}$ test set	3675 reflections (4.99%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	41.6	Xtriage
Anisotropy	0.041	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 30.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.046 for h,-k,-l	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	22519	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.18% 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	0.29	0/1937	0.45	0/2635
1	C	0.27	0/1937	0.45	0/2635
1	E	0.26	0/1935	0.43	0/2633
1	G	0.26	0/1929	0.43	0/2624
1	I	0.29	0/1928	0.45	0/2623
1	K	0.26	0/1939	0.43	0/2636
1	M	0.28	0/1928	0.45	0/2623
1	O	0.26	0/1928	0.43	0/2623
2	B	0.25	0/868	0.43	0/1181
2	D	0.26	0/868	0.46	0/1181
2	F	0.25	0/844	0.43	0/1148
2	H	0.28	0/860	0.44	0/1169
2	J	0.25	0/868	0.46	0/1181
2	L	0.29	0/860	0.47	0/1169
2	N	0.28	0/844	0.46	0/1148
2	P	0.25	0/868	0.41	0/1181
All	All	0.27	0/22341	0.44	0/30390

There are no bond length outliers.

There are no bond angle outliers.

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	1883	0	1875	20	0
1	C	1883	0	1875	24	0
1	E	1881	0	1878	26	0
1	G	1875	0	1871	21	0
1	I	1874	0	1869	23	0
1	K	1885	0	1879	21	0
1	M	1874	0	1869	16	0
1	O	1874	0	1869	23	0
2	B	846	0	818	8	1
2	D	846	0	818	15	0
2	F	823	0	791	9	0
2	H	839	0	811	12	0
2	J	846	0	818	8	0
2	L	839	0	811	13	1
2	N	823	0	791	9	0
2	P	846	0	818	5	0
3	A	104	0	0	3	0
3	B	22	0	0	1	0
3	C	96	0	0	3	0
3	D	34	0	0	2	0
3	E	73	0	0	2	0
3	F	29	0	0	4	0
3	G	90	0	0	2	0
3	H	32	0	0	2	0
3	I	86	0	0	2	0
3	J	10	0	0	0	0
3	K	68	0	0	5	0
3	L	28	0	0	3	0
3	M	49	0	0	1	0
3	N	21	0	0	1	0
3	O	31	0	0	1	0
3	P	9	0	0	0	0
All	All	22519	0	21461	223	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 5.

The worst 5 of 223 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:D:30:GLU:HB3	2:D:35:ILE:HD11	1.60	0.81
1:C:225:ARG:HG2	1:M:105:SER:HA	1.63	0.78

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:M:99:PHE:HB3	1:M:111:THR:HG23	1.66	0.76
1:I:198:ALA:O	1:I:209:ARG:NH2	2.19	0.75
1:G:209:ARG:NH1	3:G:302:HOH:O	2.20	0.73

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)
2:B:10:TYR:OH	2:L:6:GLN:OE1[1_655]	2.18	0.02

## 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	232/255 (91%)	226 (97%)	6 (3%)	0	100	100
1	C	232/255 (91%)	226 (97%)	6 (3%)	0	100	100
1	E	232/255 (91%)	226 (97%)	6 (3%)	0	100	100
1	G	231/255 (91%)	225 (97%)	6 (3%)	0	100	100
1	I	231/255 (91%)	222 (96%)	9 (4%)	0	100	100
1	K	232/255 (91%)	226 (97%)	6 (3%)	0	100	100
1	M	231/255 (91%)	226 (98%)	5 (2%)	0	100	100
1	O	231/255 (91%)	224 (97%)	7 (3%)	0	100	100
2	B	98/112 (88%)	97 (99%)	1 (1%)	0	100	100
2	D	98/112 (88%)	97 (99%)	1 (1%)	0	100	100
2	F	95/112 (85%)	94 (99%)	1 (1%)	0	100	100
2	H	97/112 (87%)	95 (98%)	2 (2%)	0	100	100
2	J	98/112 (88%)	96 (98%)	2 (2%)	0	100	100

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	L	97/112 (87%)	96 (99%)	1 (1%)	0	100	100
2	N	95/112 (85%)	94 (99%)	1 (1%)	0	100	100
2	P	98/112 (88%)	96 (98%)	2 (2%)	0	100	100
All	All	2628/2936 (90%)	2566 (98%)	62 (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	
1	A	209/227 (92%)	205 (98%)	4 (2%)	57	82
1	C	209/227 (92%)	205 (98%)	4 (2%)	57	82
1	E	209/227 (92%)	207 (99%)	2 (1%)	76	91
1	G	208/227 (92%)	204 (98%)	4 (2%)	57	82
1	I	208/227 (92%)	205 (99%)	3 (1%)	67	86
1	K	209/227 (92%)	206 (99%)	3 (1%)	67	86
1	M	208/227 (92%)	206 (99%)	2 (1%)	76	91
1	O	208/227 (92%)	204 (98%)	4 (2%)	57	82
2	B	95/105 (90%)	94 (99%)	1 (1%)	73	90
2	D	95/105 (90%)	94 (99%)	1 (1%)	73	90
2	F	92/105 (88%)	91 (99%)	1 (1%)	73	90
2	H	94/105 (90%)	94 (100%)	0	100	100
2	J	95/105 (90%)	94 (99%)	1 (1%)	73	90
2	L	94/105 (90%)	92 (98%)	2 (2%)	53	80
2	N	92/105 (88%)	90 (98%)	2 (2%)	52	79
2	P	95/105 (90%)	94 (99%)	1 (1%)	73	90
All	All	2420/2656 (91%)	2385 (99%)	35 (1%)	67	86

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	111	THR
1	I	96	ARG
1	O	215	HIS
1	G	224	THR
1	I	41	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	211	GLN
1	G	226	HIS
2	L	62	HIS
2	F	95	HIS
1	K	108	ASN

### 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 monosaccharides 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 ⓘ

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	234/255 (91%)	-0.43	1 (0%) 92 93	24, 35, 56, 76	0
1	C	234/255 (91%)	-0.44	1 (0%) 92 93	23, 36, 56, 98	0
1	E	234/255 (91%)	-0.30	2 (0%) 84 85	27, 39, 60, 69	0
1	G	233/255 (91%)	-0.33	3 (1%) 77 78	28, 39, 61, 108	0
1	I	233/255 (91%)	-0.28	3 (1%) 77 78	28, 40, 64, 81	0
1	K	234/255 (91%)	-0.23	6 (2%) 56 57	28, 44, 68, 110	0
1	M	233/255 (91%)	-0.03	6 (2%) 56 57	32, 52, 76, 94	0
1	O	233/255 (91%)	0.31	15 (6%) 19 18	36, 63, 99, 130	0
2	B	100/112 (89%)	-0.27	2 (2%) 65 67	29, 41, 61, 85	0
2	D	100/112 (89%)	-0.29	1 (1%) 82 83	29, 42, 62, 88	0
2	F	97/112 (86%)	-0.27	1 (1%) 82 83	30, 44, 60, 77	0
2	H	99/112 (88%)	-0.26	3 (3%) 50 51	29, 39, 58, 84	0
2	J	100/112 (89%)	-0.09	3 (3%) 50 51	32, 44, 70, 84	0
2	L	99/112 (88%)	-0.30	3 (3%) 50 51	30, 40, 63, 86	0
2	N	97/112 (86%)	-0.14	0 100 100	31, 43, 67, 75	0
2	P	100/112 (89%)	0.11	0 100 100	38, 57, 80, 85	0
All	All	2660/2936 (90%)	-0.21	50 (1%) 66 69	23, 42, 74, 130	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	O	227	ARG	4.3
1	G	6	ARG	4.2
2	L	99	MET	3.9
1	M	235	LEU	3.7
2	J	20	ASP	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 monosaccharides 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.