



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 14, 2018 – 06:38 am GMT

PDB ID : 4YF9  
Title : Structure of N-acylhomoserine lactone acylase MacQ  
Authors : Yasutake, Y.; Kusada, H.; Kimura, N.  
Deposited on : 2015-02-25  
Resolution : 2.60 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	:	trunk31020
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk31020

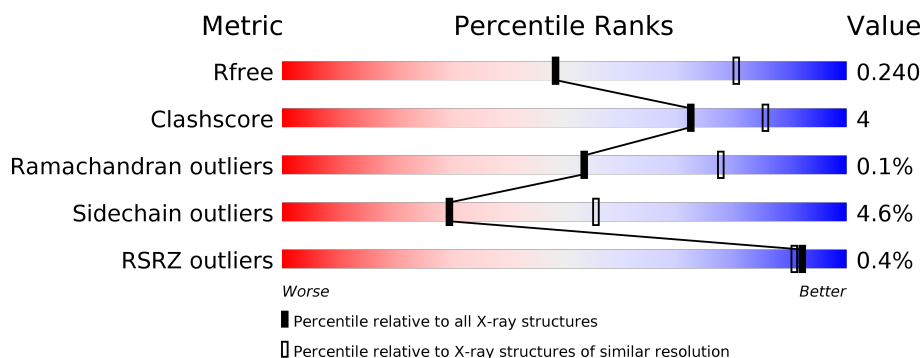
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.60 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))
$R_{free}$	111664	2767 (2.60-2.60)
Clashscore	122126	3110 (2.60-2.60)
Ramachandran outliers	120053	3062 (2.60-2.60)
Sidechain outliers	120020	3062 (2.60-2.60)
RSRZ outliers	108989	2706 (2.60-2.60)

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	178	<div> <div style="width: 88%;"></div> <div style="width: 6%;"></div> <div style="width: 6%;"></div> </div> <div>88% 6% • 6%</div>
1	D	178	<div> <div style="width: 86%;"></div> <div style="width: 8%;"></div> <div style="width: 6%;"></div> </div> <div>86% 8% • 6%</div>
1	G	178	<div> <div style="width: 85%;"></div> <div style="width: 8%;"></div> <div style="width: 6%;"></div> </div> <div>85% 8% • 6%</div>
1	J	178	<div> <div style="width: 80%;"></div> <div style="width: 12%;"></div> <div style="width: 6%;"></div> </div> <div>80% 12% • 6%</div>
2	B	27	<div> <div style="width: 52%;"></div> <div style="width: 44%;"></div> </div> <div>52% 44%</div>
2	E	27	<div> <div style="width: 7%;"></div> <div style="width: 52%;"></div> <div style="width: 48%;"></div> </div> <div>7% 52% 48%</div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	27	<div><div><div></div><div></div><div></div></div><div>4%44%11%44%</div></div>
2	K	27	<div><div><div></div><div></div><div></div></div><div>48%11%41%</div></div>
3	C	581	<div><div><div></div><div></div><div></div></div><div>%85%12%..</div></div>
3	F	581	<div><div><div></div><div></div><div></div></div><div>87%10%..</div></div>
3	I	581	<div><div><div></div><div></div><div></div></div><div>86%11%..</div></div>
3	L	581	<div><div><div></div><div></div><div></div></div><div>%85%13%..</div></div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 23165 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 related to penicillin acylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	167	Total	C	N	O	S	0	0	0
			1289	802	243	238	6			
1	D	168	Total	C	N	O	S	0	0	0
			1295	805	244	240	6			
1	G	167	Total	C	N	O	S	0	0	0
			1289	802	243	238	6			
1	J	167	Total	C	N	O	S	0	0	0
			1289	802	243	238	6			

- Molecule 2 is a protein called Protein related to penicillin acylase.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	15	Total	C	N	O	0	0	0
			110	68	22	20			
2	E	14	Total	C	N	O	0	0	0
			106	66	21	19			
2	H	15	Total	C	N	O	0	0	0
			110	68	22	20			
2	K	16	Total	C	N	O	0	0	0
			119	73	23	23			

- Molecule 3 is a protein called Protein related to penicillin acylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	575	Total	C	N	O	S	0	0	0
			4353	2739	763	832	19			
3	F	575	Total	C	N	O	S	0	0	0
			4353	2739	763	832	19			
3	I	574	Total	C	N	O	S	0	0	0
			4344	2734	762	829	19			
3	L	575	Total	C	N	O	S	0	0	0
			4353	2739	763	832	19			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	574	LEU	-	expression tag	UNP A0A0A1VBK6
C	575	GLU	-	expression tag	UNP A0A0A1VBK6
C	576	HIS	-	expression tag	UNP A0A0A1VBK6
C	577	HIS	-	expression tag	UNP A0A0A1VBK6
C	578	HIS	-	expression tag	UNP A0A0A1VBK6
C	579	HIS	-	expression tag	UNP A0A0A1VBK6
C	580	HIS	-	expression tag	UNP A0A0A1VBK6
C	581	HIS	-	expression tag	UNP A0A0A1VBK6
F	574	LEU	-	expression tag	UNP A0A0A1VBK6
F	575	GLU	-	expression tag	UNP A0A0A1VBK6
F	576	HIS	-	expression tag	UNP A0A0A1VBK6
F	577	HIS	-	expression tag	UNP A0A0A1VBK6
F	578	HIS	-	expression tag	UNP A0A0A1VBK6
F	579	HIS	-	expression tag	UNP A0A0A1VBK6
F	580	HIS	-	expression tag	UNP A0A0A1VBK6
F	581	HIS	-	expression tag	UNP A0A0A1VBK6
I	574	LEU	-	expression tag	UNP A0A0A1VBK6
I	575	GLU	-	expression tag	UNP A0A0A1VBK6
I	576	HIS	-	expression tag	UNP A0A0A1VBK6
I	577	HIS	-	expression tag	UNP A0A0A1VBK6
I	578	HIS	-	expression tag	UNP A0A0A1VBK6
I	579	HIS	-	expression tag	UNP A0A0A1VBK6
I	580	HIS	-	expression tag	UNP A0A0A1VBK6
I	581	HIS	-	expression tag	UNP A0A0A1VBK6
L	574	LEU	-	expression tag	UNP A0A0A1VBK6
L	575	GLU	-	expression tag	UNP A0A0A1VBK6
L	576	HIS	-	expression tag	UNP A0A0A1VBK6
L	577	HIS	-	expression tag	UNP A0A0A1VBK6
L	578	HIS	-	expression tag	UNP A0A0A1VBK6
L	579	HIS	-	expression tag	UNP A0A0A1VBK6
L	580	HIS	-	expression tag	UNP A0A0A1VBK6
L	581	HIS	-	expression tag	UNP A0A0A1VBK6

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	4	Total O 4 4	0	0
4	C	31	Total O 31 31	0	0
4	D	15	Total O 15 15	0	0

*Continued on next page...*

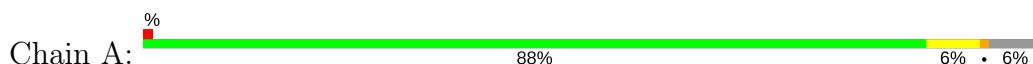
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	F	29	Total 29	O 29	0	0
4	G	18	Total 18	O 18	0	0
4	H	1	Total 1	O 1	0	0
4	I	33	Total 33	O 33	0	0
4	J	8	Total 8	O 8	0	0
4	L	16	Total 16	O 16	0	0

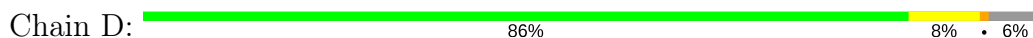
### 3 Residue-property plots [i](#)

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.

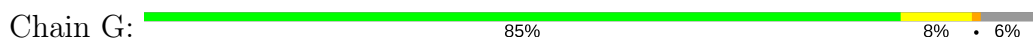
- Molecule 1: Protein related to penicillin acylase



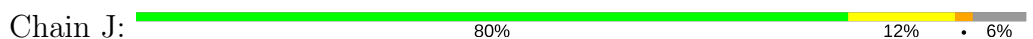
- Molecule 1: Protein related to penicillin acylase



- Molecule 1: Protein related to penicillin acylase



- Molecule 1: Protein related to penicillin acylase

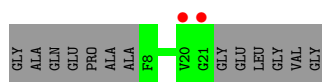


- Molecule 2: Protein related to penicillin acylase

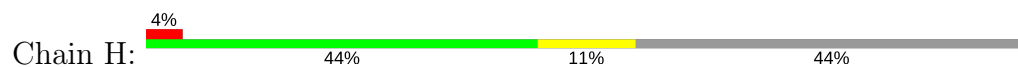


- Molecule 2: Protein related to penicillin acylase

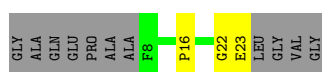




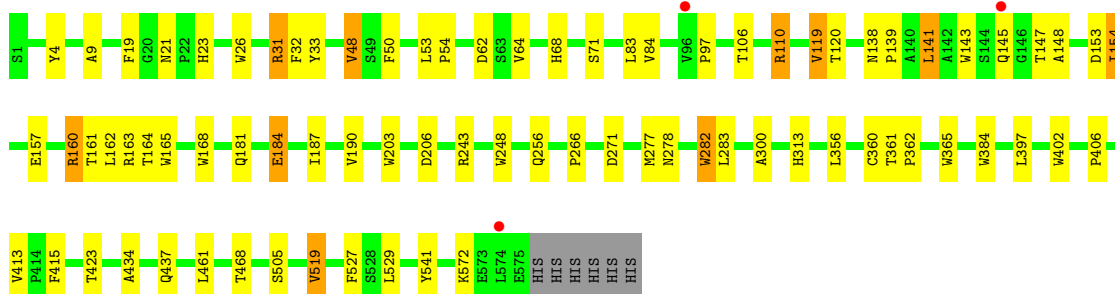
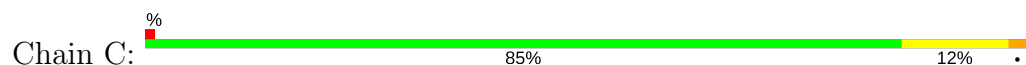
- Molecule 2: Protein related to penicillin acylase



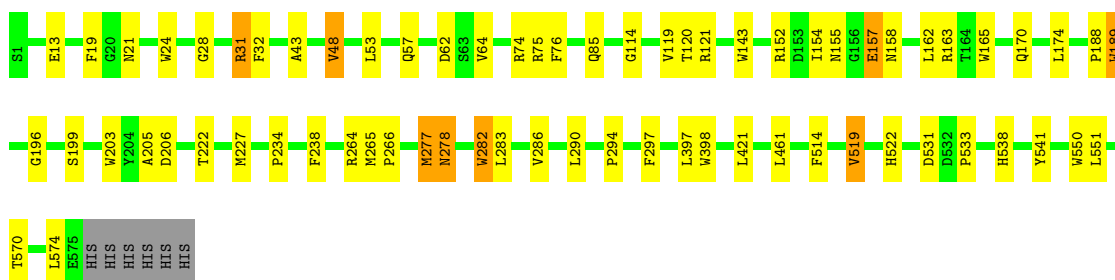
- Molecule 2: Protein related to penicillin acylase



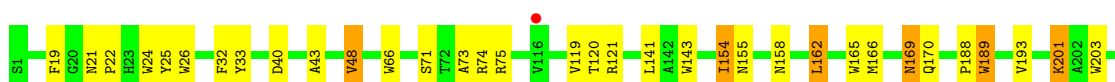
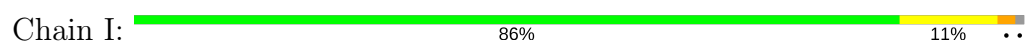
- Molecule 3: Protein related to penicillin acylase



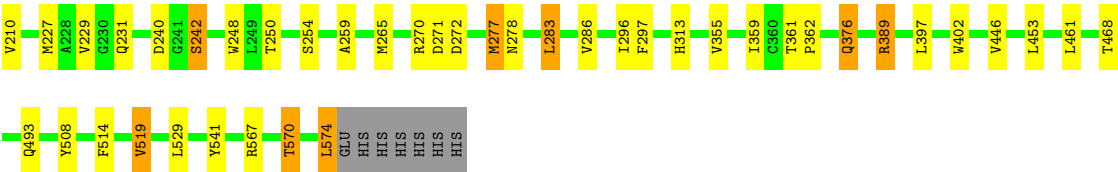
- Molecule 3: Protein related to penicillin acylase



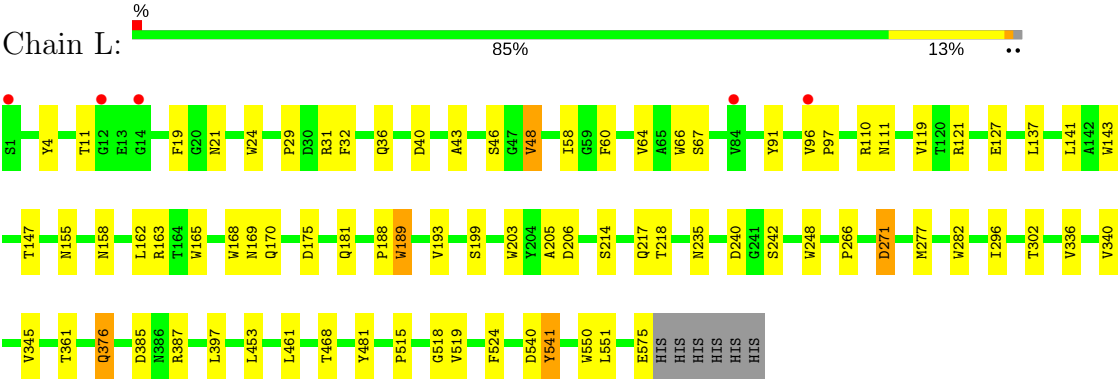
- Molecule 3: Protein related to penicillin acylase







● Molecule 3: Protein related to penicillin acylase



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	85.62Å 90.11Å 123.17Å 103.49° 104.96° 105.95°	Depositor
Resolution (Å)	38.32 – 2.60 38.32 – 2.60	Depositor EDS
% Data completeness (in resolution range)	98.7 (38.32-2.60) 98.8 (38.32-2.60)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.94 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.196 , 0.241 0.196 , 0.240	Depositor DCC
$R_{free}$ test set	4943 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	40.3	Xtriage
Anisotropy	0.026	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 31.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.004 for -h,-k,h+k+l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	23165	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.24% 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	0.50	0/1320	0.55	0/1788
1	D	0.50	1/1326 (0.1%)	0.54	0/1796
1	G	0.51	1/1320 (0.1%)	0.55	0/1788
1	J	0.52	1/1320 (0.1%)	0.54	0/1788
2	B	0.38	0/112	0.47	0/150
2	E	0.38	0/108	0.48	0/145
2	H	0.38	0/112	0.49	0/150
2	K	0.37	0/121	0.48	0/162
3	C	0.55	8/4467 (0.2%)	0.58	0/6094
3	F	0.55	7/4467 (0.2%)	0.59	0/6094
3	I	0.55	8/4458 (0.2%)	0.60	0/6082
3	L	0.54	8/4467 (0.2%)	0.59	0/6094
All	All	0.54	34/23598 (0.1%)	0.58	0/32131

The worst 5 of 34 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	L	282	TRP	CD2-CE2	5.63	1.48	1.41
3	I	165	TRP	CD2-CE2	5.63	1.48	1.41
3	L	24	TRP	CD2-CE2	5.46	1.47	1.41
3	I	24	TRP	CD2-CE2	5.43	1.47	1.41
3	C	402	TRP	CD2-CE2	5.42	1.47	1.41

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	1289	0	1228	10	0
1	D	1295	0	1233	7	0
1	G	1289	0	1228	7	0
1	J	1289	0	1228	14	0
2	B	110	0	108	1	0
2	E	106	0	105	0	0
2	H	110	0	108	3	0
2	K	119	0	114	2	0
3	C	4353	0	4161	46	0
3	F	4353	0	4161	31	0
3	I	4344	0	4155	40	0
3	L	4353	0	4161	35	0
4	A	4	0	0	0	0
4	C	31	0	0	0	0
4	D	15	0	0	0	0
4	F	29	0	0	0	0
4	G	18	0	0	0	0
4	H	1	0	0	0	0
4	I	33	0	0	0	0
4	J	8	0	0	0	0
4	L	16	0	0	0	0
All	All	23165	0	21990	165	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 4.

The worst 5 of 165 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)
3:I:376:GLN:HG2	3:I:453:LEU:HD12	1.60	0.83
3:C:32:PHE:HB3	3:C:48:VAL:CG2	2.21	0.70
1:G:106:GLN:HG3	1:G:153:TRP:CH2	2.27	0.69
3:F:74:ARG:HH21	3:F:234:PRO:HG2	1.57	0.69
3:C:32:PHE:HB3	3:C:48:VAL:HG21	1.76	0.67

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	165/178 (93%)	158 (96%)	7 (4%)	0	100	100
1	D	166/178 (93%)	159 (96%)	7 (4%)	0	100	100
1	G	165/178 (93%)	160 (97%)	5 (3%)	0	100	100
1	J	165/178 (93%)	158 (96%)	6 (4%)	1 (1%)	27	51
2	B	13/27 (48%)	12 (92%)	1 (8%)	0	100	100
2	E	12/27 (44%)	11 (92%)	1 (8%)	0	100	100
2	H	13/27 (48%)	11 (85%)	2 (15%)	0	100	100
2	K	14/27 (52%)	14 (100%)	0	0	100	100
3	C	573/581 (99%)	547 (96%)	26 (4%)	0	100	100
3	F	573/581 (99%)	545 (95%)	27 (5%)	1 (0%)	49	74
3	I	572/581 (98%)	545 (95%)	26 (4%)	1 (0%)	49	74
3	L	573/581 (99%)	549 (96%)	23 (4%)	1 (0%)	49	74
All	All	3004/3144 (96%)	2869 (96%)	131 (4%)	4 (0%)	53	78

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	I	40	ASP
3	L	40	ASP
1	J	139	ARG
3	F	278	ASN

### 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	125/130 (96%)	123 (98%)	2 (2%)	65	85
1	D	126/130 (97%)	121 (96%)	5 (4%)	34	61
1	G	125/130 (96%)	118 (94%)	7 (6%)	23	45
1	J	125/130 (96%)	119 (95%)	6 (5%)	28	53
2	B	11/17 (65%)	11 (100%)	0	100	100
2	E	11/17 (65%)	11 (100%)	0	100	100
2	H	11/17 (65%)	11 (100%)	0	100	100
2	K	12/17 (71%)	12 (100%)	0	100	100
3	C	446/452 (99%)	427 (96%)	19 (4%)	32	58
3	F	446/452 (99%)	427 (96%)	19 (4%)	32	58
3	I	445/452 (98%)	420 (94%)	25 (6%)	23	45
3	L	446/452 (99%)	421 (94%)	25 (6%)	23	45
All	All	2329/2396 (97%)	2221 (95%)	108 (5%)	29	55

5 of 108 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	72	LEU
3	I	201	LYS
3	L	277	MET
1	G	155	ARG
3	I	119	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	176	GLN
3	L	111	ASN
3	I	21	ASN
3	C	21	ASN
3	I	231	GLN

### 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	167/178 (93%)	-0.50	1 (0%) 89 88	31, 46, 62, 82	0
1	D	168/178 (94%)	-0.70	0 100 100	26, 37, 52, 70	0
1	G	167/178 (93%)	-0.65	0 100 100	27, 36, 49, 57	0
1	J	167/178 (93%)	-0.61	0 100 100	29, 38, 57, 77	0
2	B	15/27 (55%)	0.20	0 100 100	58, 65, 77, 94	0
2	E	14/27 (51%)	0.62	2 (14%) 2 1	65, 79, 83, 85	0
2	H	15/27 (55%)	0.16	1 (6%) 18 13	50, 68, 80, 81	0
2	K	16/27 (59%)	-0.03	0 100 100	59, 68, 74, 85	0
3	C	575/581 (98%)	-0.56	3 (0%) 90 89	24, 39, 69, 88	0
3	F	575/581 (98%)	-0.57	0 100 100	26, 40, 64, 80	0
3	I	574/581 (98%)	-0.65	1 (0%) 94 95	21, 36, 56, 75	0
3	L	575/581 (98%)	-0.37	5 (0%) 84 81	28, 48, 71, 96	0
All	All	3028/3144 (96%)	-0.54	13 (0%) 92 91	21, 41, 67, 96	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	574	LEU	3.5
2	E	21	GLY	2.5
3	L	12	GLY	2.5
3	L	14	GLY	2.4
3	L	1	SER	2.4

### 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.