



# wwPDB X-ray Structure Validation Summary Report i

Feb 27, 2014 – 07:59 PM GMT

PDB ID : 2G3M  
Title : Crystal structure of the Sulfolobus solfataricus alpha-glucosidase MalA  
Authors : Ernst, H.A.; Lo Leggio, L.; Willemoes, M.; Leonard, G.; Blum, P.; Larsen, S.  
Deposited on : 2006-02-20  
Resolution : 2.55 Å(reported)

This is a wwPDB 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

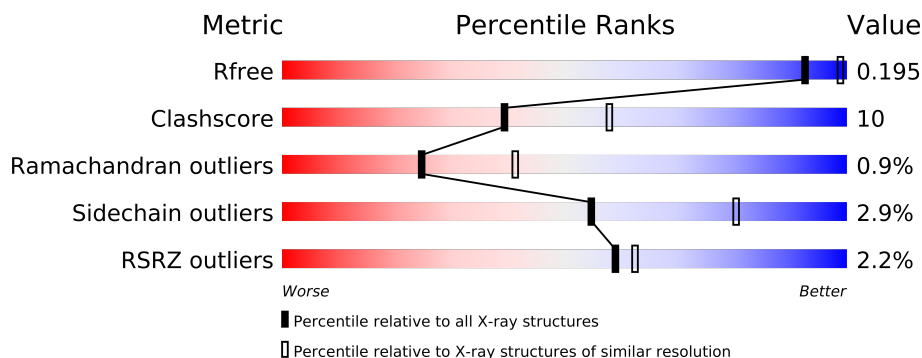
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.15 2013
Xtriage (Phenix)	:	dev-1323
EDS	:	stable22639
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.3.0 (Settle)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.55 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))
$R_{free}$	66092	3413 (2.58-2.50)
Clashscore	79885	4284 (2.58-2.50)
Ramachandran outliers	78287	4193 (2.58-2.50)
Sidechain outliers	78261	4195 (2.58-2.50)
RSRZ outliers	66119	3414 (2.58-2.50)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	693	
1	B	693	
1	C	693	
1	D	693	
1	E	693	
1	F	693	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 35378 atoms, of which 0 are hydrogen and 0 are deuterium.

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 Alpha-glucosidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	691	Total	C	N	O	S	0	1	0
			5684	3703	919	1047	15			
1	B	691	Total	C	N	O	S	0	1	0
			5684	3703	919	1047	15			
1	C	691	Total	C	N	O	S	0	1	0
			5684	3703	919	1047	15			
1	D	691	Total	C	N	O	S	0	1	0
			5684	3703	919	1047	15			
1	E	691	Total	C	N	O	S	0	1	0
			5684	3703	919	1047	15			
1	F	692	Total	C	N	O	S	0	1	0
			5695	3709	923	1048	15			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	CLONING ARTIFACT	UNP O59645
A	2	ARG	-	CLONING ARTIFACT	UNP O59645
A	3	ILE	-	CLONING ARTIFACT	UNP O59645
A	4	LEU	-	CLONING ARTIFACT	UNP O59645
B	1	MET	-	CLONING ARTIFACT	UNP O59645
B	2	ARG	-	CLONING ARTIFACT	UNP O59645
B	3	ILE	-	CLONING ARTIFACT	UNP O59645
B	4	LEU	-	CLONING ARTIFACT	UNP O59645
C	1	MET	-	CLONING ARTIFACT	UNP O59645
C	2	ARG	-	CLONING ARTIFACT	UNP O59645
C	3	ILE	-	CLONING ARTIFACT	UNP O59645
C	4	LEU	-	CLONING ARTIFACT	UNP O59645
D	1	MET	-	CLONING ARTIFACT	UNP O59645
D	2	ARG	-	CLONING ARTIFACT	UNP O59645
D	3	ILE	-	CLONING ARTIFACT	UNP O59645
D	4	LEU	-	CLONING ARTIFACT	UNP O59645
E	1	MET	-	CLONING ARTIFACT	UNP O59645

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	2	ARG	-	CLONING ARTIFACT	UNP O59645
E	3	ILE	-	CLONING ARTIFACT	UNP O59645
E	4	LEU	-	CLONING ARTIFACT	UNP O59645
F	1	MET	-	CLONING ARTIFACT	UNP O59645
F	2	ARG	-	CLONING ARTIFACT	UNP O59645
F	3	ILE	-	CLONING ARTIFACT	UNP O59645
F	4	LEU	-	CLONING ARTIFACT	UNP O59645

- Molecule 2 is water.

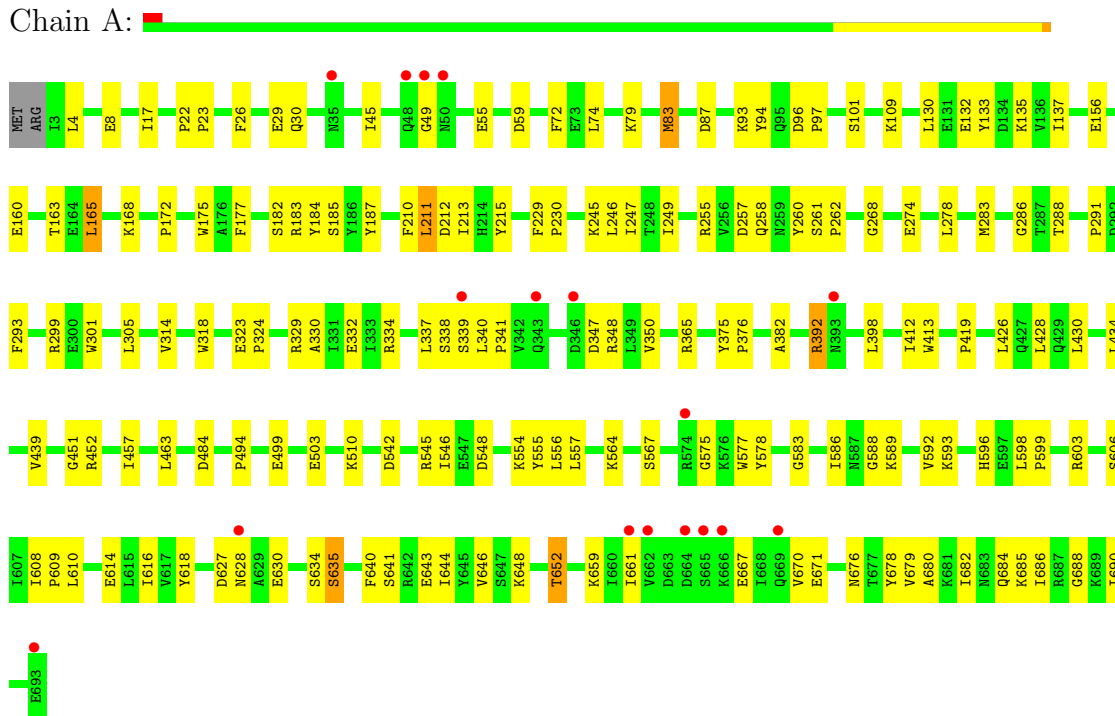
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	213	Total 213	O 213	0	0
2	B	210	Total 210	O 210	0	0
2	C	178	Total 178	O 178	0	0
2	D	221	Total 221	O 221	0	0
2	E	249	Total 249	O 249	0	0
2	F	192	Total 192	O 192	0	0

### 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 errors 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.

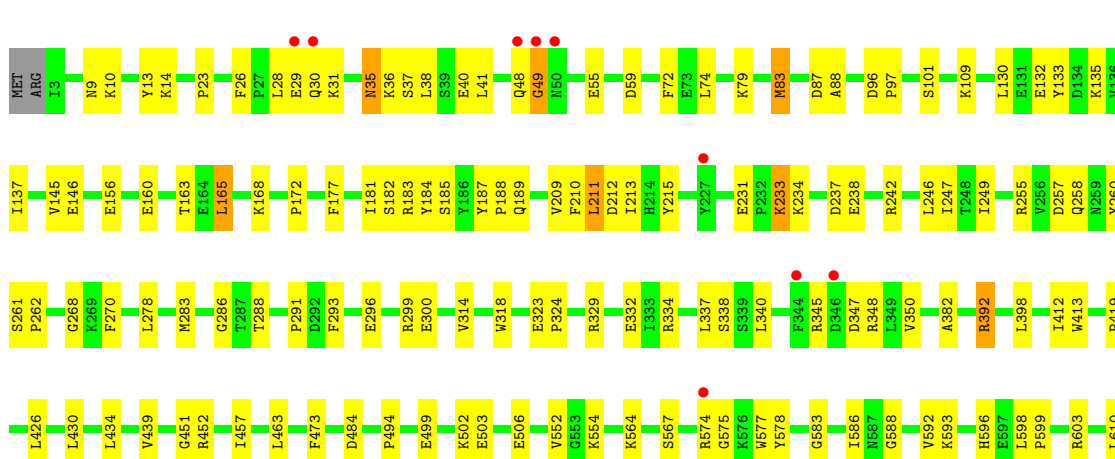
#### • Molecule 1: Alpha-glucosidase

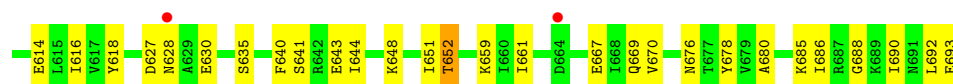
Chain A:



#### • Molecule 1: Alpha-glucosidase

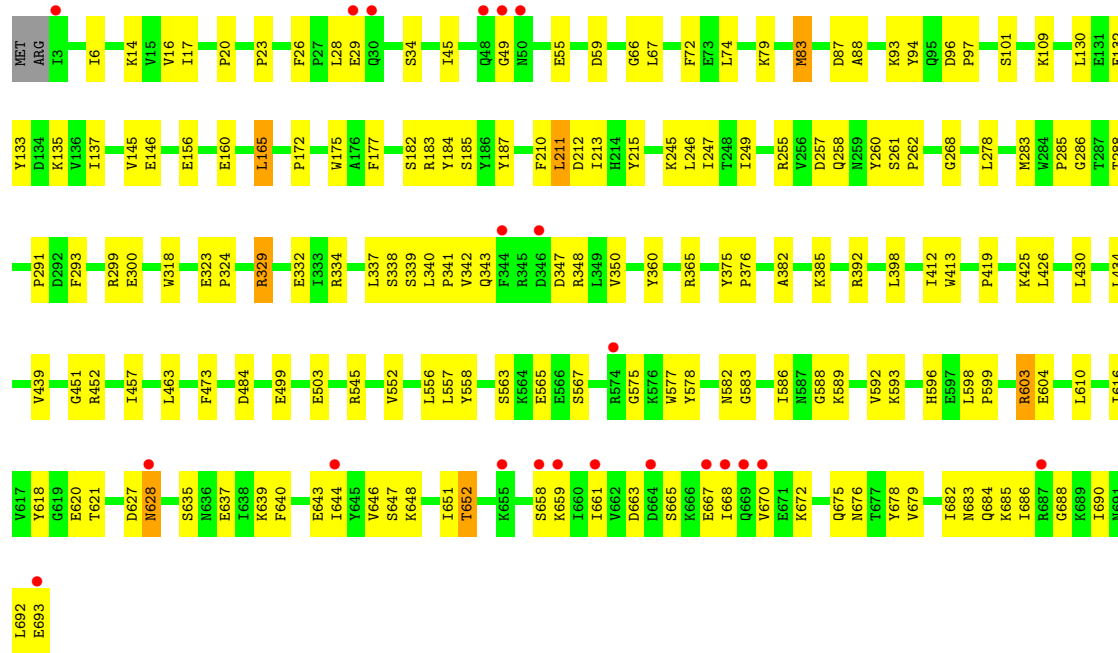
Chain B:





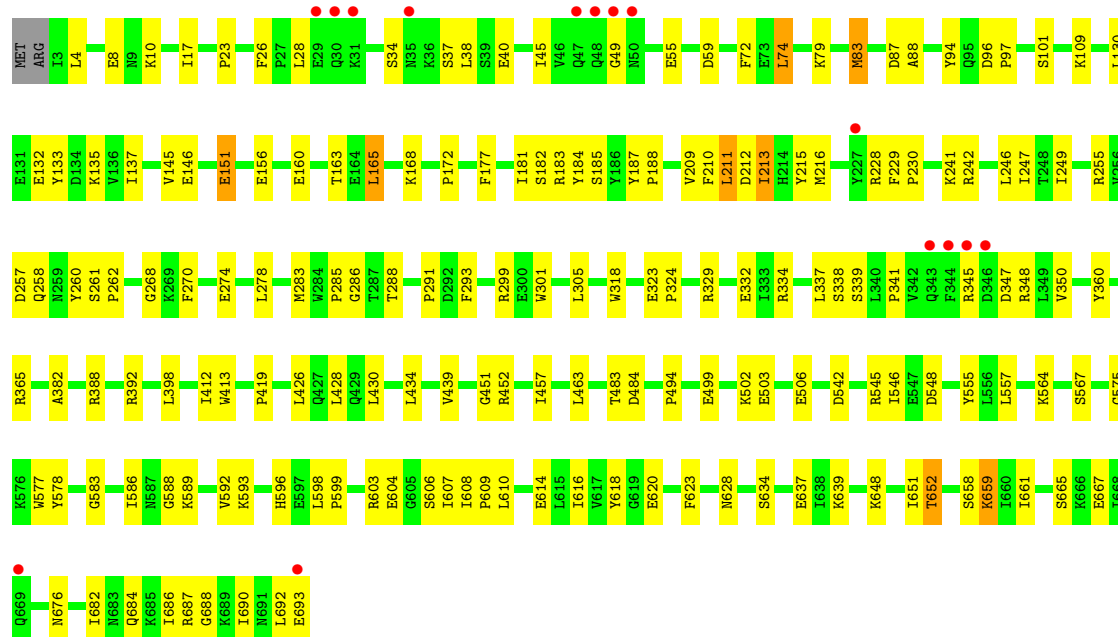
• Molecule 1: Alpha-glucosidase

Chain C:



• Molecule 1: Alpha-glucosidase

Chain D:

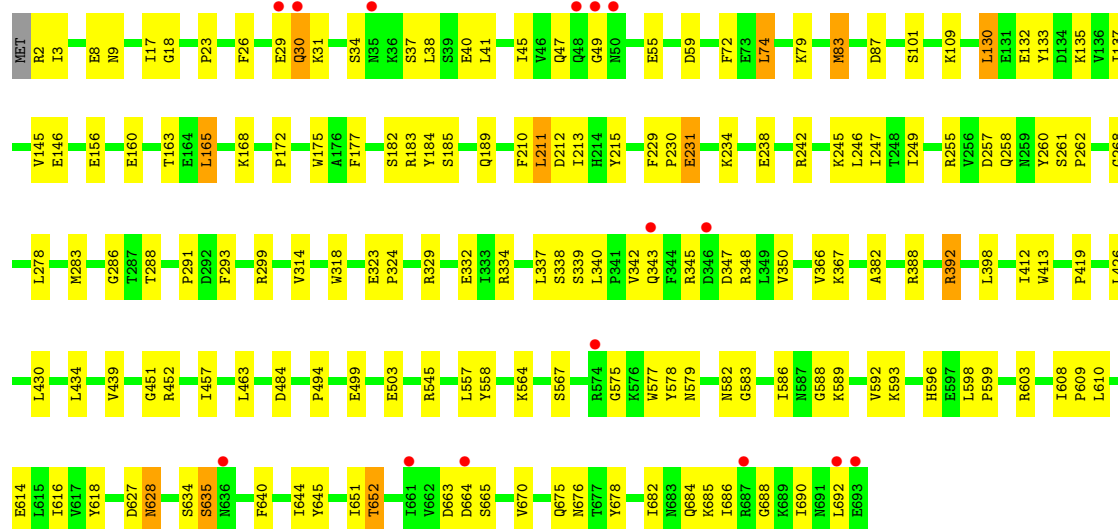


• Molecule 1: Alpha-glucosidase

Age Group	Percentage
18-24	~2%
25-34	~88%
35-44	~92%
45-54	~95%
55-64	~98%
65-74	~99%
75-84	~99%
85+	~99%



Age Group	Percentage
18-24	2%
25-34	78%
35-44	12%
45-54	5%
55-64	2%
65-74	1%
75-84	1%
85+	1%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	103.17Å 173.56Å 154.08Å 90.00° 108.00° 90.00°	Depositor
Resolution (Å)	34.25 – 2.55 34.25 – 2.55	Depositor EDS
% Data completeness (in resolution range)	99.9 (34.25-2.55) 100.0 (34.25-2.55)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.95 (at 2.54Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.170 , 0.195 0.171 , 0.195	Depositor DCC
$R_{free}$ test set	8380 reflections (5.01%)	DCC
Wilson B-factor (Å <sup>2</sup> )	33.7	Xtriage
Anisotropy	0.344	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 27.2	EDS
Estimated twinning fraction	0.018 for h,-k,-h-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	0 of 167412 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	35378	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.64% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.



## 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	0.36	0/5838	0.61	0/7902
1	B	0.36	0/5838	0.61	0/7902
1	C	0.36	0/5838	0.60	0/7902
1	D	0.37	0/5838	0.61	0/7902
1	E	0.37	0/5838	0.62	0/7902
1	F	0.36	0/5849	0.61	0/7916
All	All	0.36	0/35039	0.61	0/47426

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	5684	0	5651	109	0
1	B	5684	0	5651	115	0
1	C	5684	0	5651	125	0
1	D	5684	0	5651	127	0
1	E	5684	0	5651	111	0
1	F	5695	0	5664	123	0
2	A	213	0	0	2	0
2	B	210	0	0	3	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	178	0	0	2	0
2	D	221	0	0	3	0
2	E	249	0	0	4	0
2	F	192	0	0	4	0
All	All	35378	0	33919	651	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 10.

The worst 5 of 651 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:D:661:ILE:HG12	1:D:667:GLU:HG2	1.41	1.02
1:A:258:GLN:HG3	1:E:278:LEU:HD21	1.51	0.93
1:D:564:LYS:H	1:D:564:LYS:HD2	1.31	0.91
1:B:564:LYS:H	1:B:564:LYS:HD2	1.37	0.88
1:B:35:ASN:H	1:B:35:ASN:HD22	1.23	0.85

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	690/693 (100%)	651 (94%)	33 (5%)	6 (1%)	25	41
1	B	690/693 (100%)	653 (95%)	31 (4%)	6 (1%)	25	41
1	C	690/693 (100%)	648 (94%)	35 (5%)	7 (1%)	22	37
1	D	690/693 (100%)	652 (94%)	34 (5%)	4 (1%)	33	54
1	E	690/693 (100%)	649 (94%)	35 (5%)	6 (1%)	25	41
1	F	691/693 (100%)	649 (94%)	34 (5%)	8 (1%)	19	31
All	All	4141/4158 (100%)	3902 (94%)	202 (5%)	37 (1%)	25	41

5 of 37 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	29	GLU
1	C	29	GLU
1	E	29	GLU
1	F	635	SER
1	C	635	SER

### 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	623/624 (100%)	607 (97%)	16 (3%)	59	84
1	B	623/624 (100%)	604 (97%)	19 (3%)	53	79
1	C	623/624 (100%)	606 (97%)	17 (3%)	57	82
1	D	623/624 (100%)	604 (97%)	19 (3%)	53	79
1	E	623/624 (100%)	602 (97%)	21 (3%)	49	75
1	F	624/624 (100%)	606 (97%)	18 (3%)	55	80
All	All	3739/3744 (100%)	3629 (97%)	110 (3%)	55	80

5 of 110 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	603	ARG
1	D	398	LEU
1	F	392	ARG
1	C	652	THR
1	D	151	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	676	ASN
1	D	48	GLN
1	F	596	HIS
1	C	683	ASN
1	D	243	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

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	691/693 (99%)	-0.15	17 (2%) 54 58	21, 31, 60, 74	0
1	B	691/693 (99%)	-0.15	11 (1%) 68 72	21, 32, 59, 78	0
1	C	691/693 (99%)	-0.07	22 (3%) 45 48	22, 32, 66, 84	0
1	D	691/693 (99%)	-0.17	15 (2%) 59 62	20, 30, 54, 78	0
1	E	691/693 (99%)	-0.21	12 (1%) 67 70	18, 30, 50, 80	0
1	F	692/693 (99%)	-0.13	15 (2%) 59 62	19, 31, 63, 83	0
All	All	4147/4158 (99%)	-0.15	92 (2%) 59 62	18, 31, 60, 84	0

The worst 5 of 92 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	F	49	GLY	5.7
1	B	30	GLN	4.9
1	A	693	GLU	4.2
1	E	30	GLN	4.2
1	D	49	GLY	4.1

### 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates ⓘ

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.