



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:54 PM GMT

PDB ID : 4ATW  
Title : The crystal structure of Arabinofuranosidase  
Authors : Dumbrepatil, A.; Song, H.-N.; Jung, T.-Y.; Kim, T.-J.; Woo, E.-J.  
Deposited on : 2012-05-10  
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  
<http://wwpdb.org/validation/2016/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  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

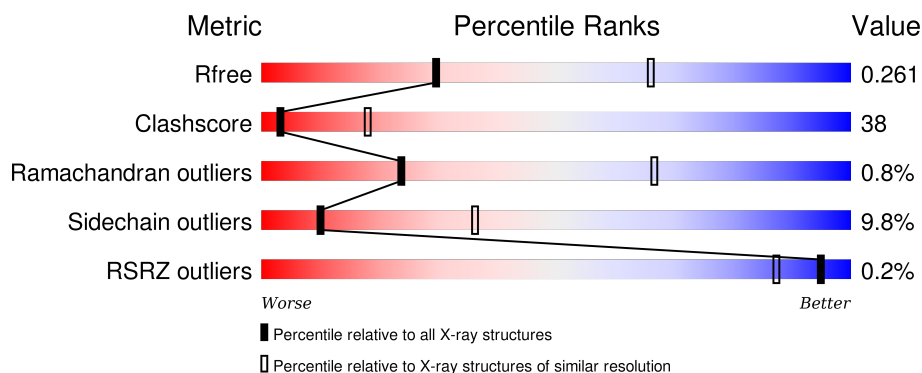
# 1 Overall quality at a glance

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))
$R_{free}$	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (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. 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	482	<div> <div>51%</div> <div>39%</div> <div>8%</div> <div>.</div> </div>
1	B	482	<div> <div>51%</div> <div>39%</div> <div>8%</div> <div>.</div> </div>
1	C	482	<div> <div>50%</div> <div>40%</div> <div>9%</div> <div>.</div> </div>
1	D	482	<div> <div>50%</div> <div>39%</div> <div>9%</div> <div>.</div> </div>
1	E	482	<div> <div>50%</div> <div>42%</div> <div>7%</div> <div>.</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	482	 A horizontal bar chart showing the quality of the chain. The bar is divided into three segments: green (53%), yellow (39%), and red (7%). The segments are labeled with their respective percentages: 53%, 39%, and 7%. A small black dot is visible at the end of the red segment.

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 23256 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 ALPHA-L-ARABINOFURANOSIDASE DOMAIN PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	481	Total	C	N	O	S	0	0	0
			3876	2486	646	729	15			
1	B	481	Total	C	N	O	S	0	0	0
			3876	2486	646	729	15			
1	C	481	Total	C	N	O	S	0	0	0
			3876	2486	646	729	15			
1	D	481	Total	C	N	O	S	0	0	0
			3876	2486	646	729	15			
1	E	481	Total	C	N	O	S	0	0	0
			3876	2486	646	729	15			
1	F	481	Total	C	N	O	S	0	0	0
			3876	2486	646	729	15			

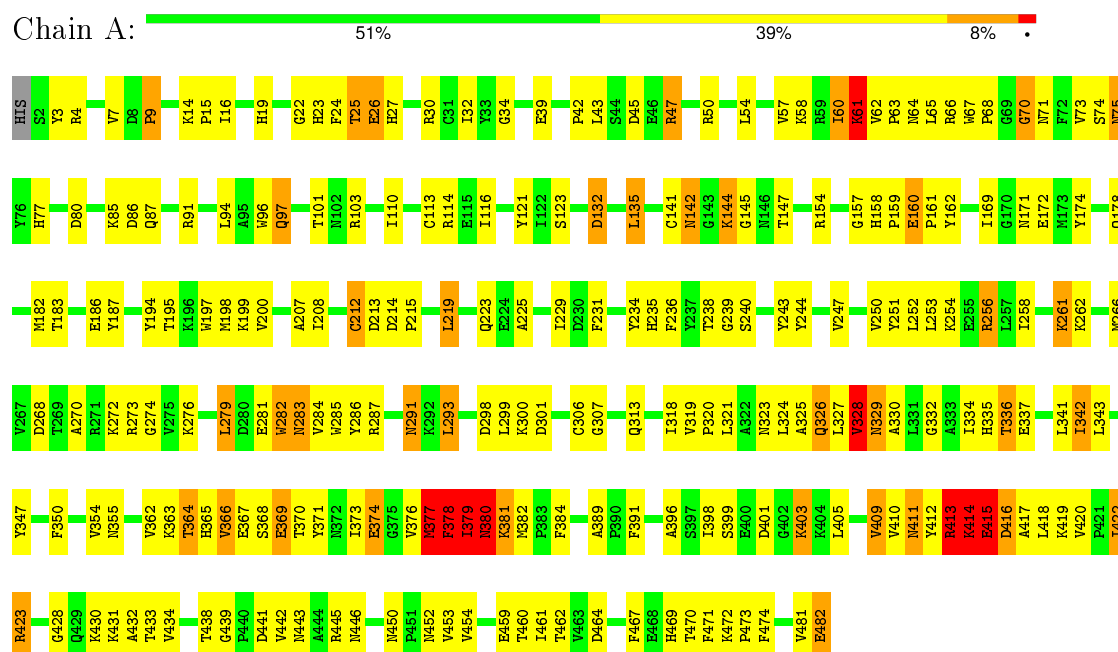
There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	HIS	-	EXPRESSION TAG	UNP G4FHJ5
B	1	HIS	-	EXPRESSION TAG	UNP G4FHJ5
C	1	HIS	-	EXPRESSION TAG	UNP G4FHJ5
D	1	HIS	-	EXPRESSION TAG	UNP G4FHJ5
E	1	HIS	-	EXPRESSION TAG	UNP G4FHJ5
F	1	HIS	-	EXPRESSION TAG	UNP G4FHJ5

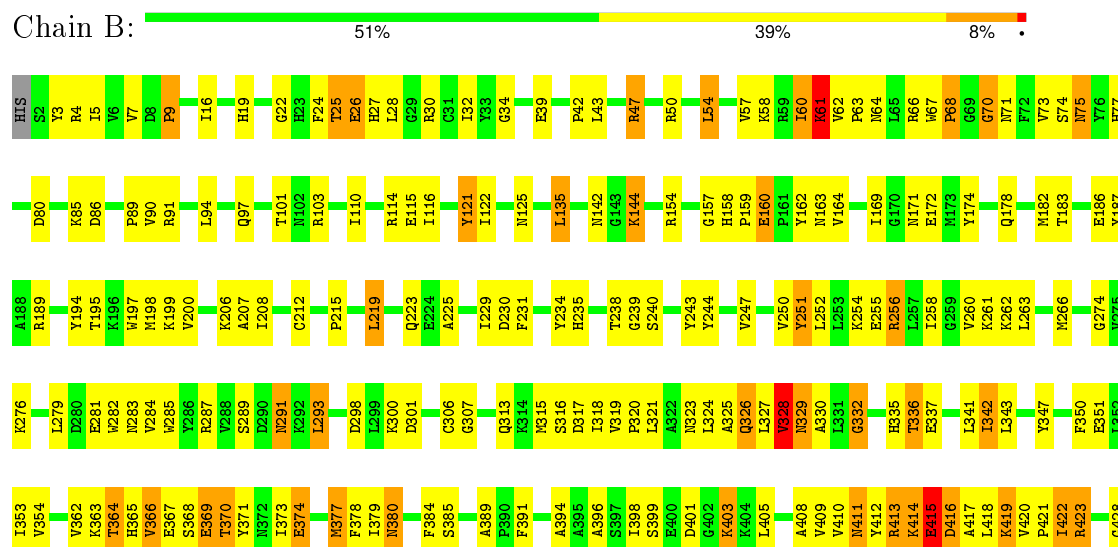
### 3 Residue-property plots

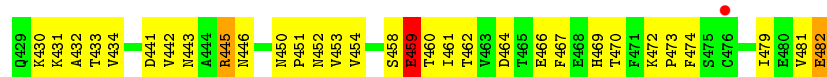
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according 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 ( $\text{RSRZ} > 2$ ). 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-L-ARABINOFURANOSIDASE DOMAIN PROTEIN



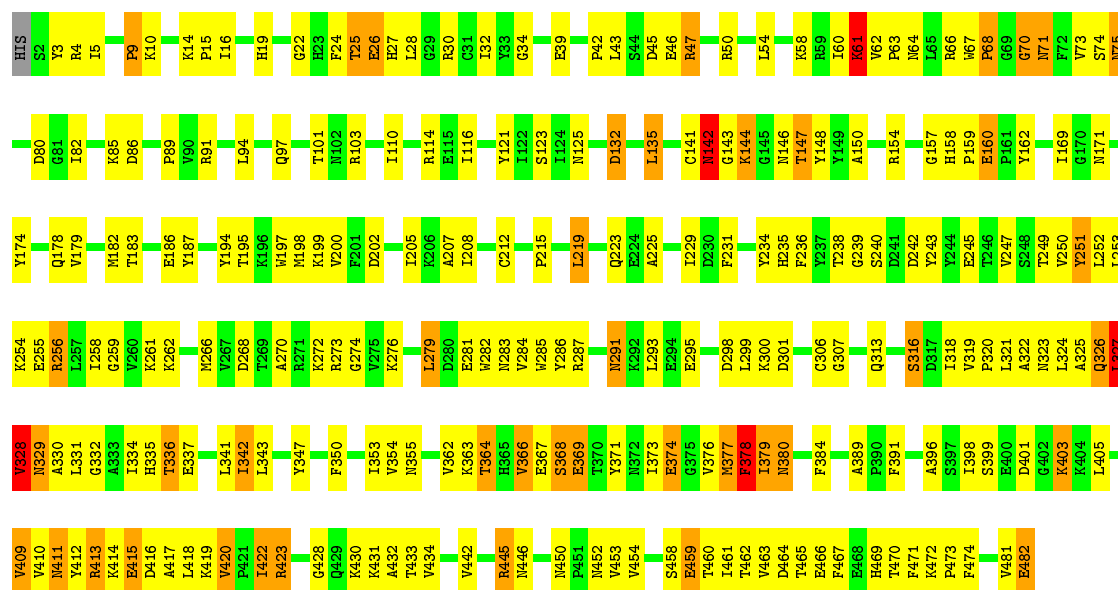
#### • Molecule 1: ALPHA-L-ARABINOFURANOSIDASE DOMAIN PROTEIN





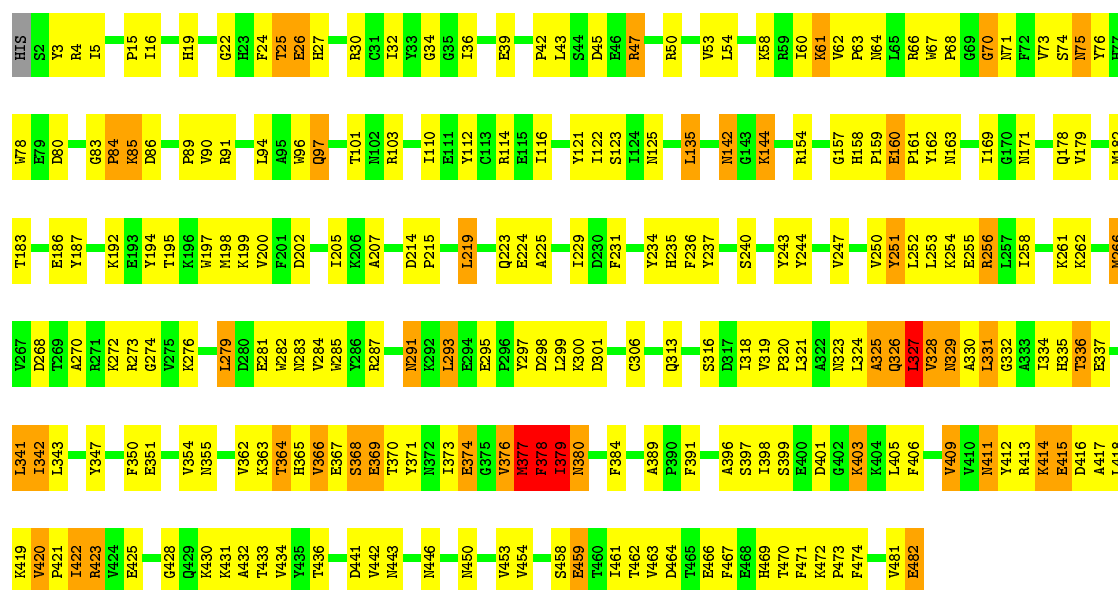
• Molecule 1: ALPHA-L-ARABINOFURANOSIDASE DOMAIN PROTEIN

Chain C: 50% 40% 9%



• Molecule 1: ALPHA-L-ARABINOFURANOSIDASE DOMAIN PROTEIN

Chain D: 50% 39% 9%



• Molecule 1: ALPHA-L-ARABINOFURANOSIDASE DOMAIN PROTEIN

Chain E: 50% 42% 7%



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	103.71Å 161.54Å 112.60Å 90.00° 106.30° 90.00°	Depositor
Resolution (Å)	29.94 – 3.00 29.94 – 2.99	Depositor EDS
% Data completeness (in resolution range)	84.5 (29.94-3.00) 84.2 (29.94-2.99)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.24 (at 3.00Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, $R_{free}$	0.227 , 0.262 0.227 , 0.261	Depositor DCC
$R_{free}$ test set	4429 reflections (7.37%)	DCC
Wilson B-factor (Å <sup>2</sup> )	45.1	Xtriage
Anisotropy	0.271	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.28 , 40.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 67140 reflections	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	23256	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.39% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	1.04	21/3971 (0.5%)	1.01	31/5383 (0.6%)
1	B	0.73	5/3971 (0.1%)	0.98	22/5383 (0.4%)
1	C	0.99	24/3971 (0.6%)	0.96	24/5383 (0.4%)
1	D	1.10	37/3971 (0.9%)	1.04	31/5383 (0.6%)
1	E	0.84	16/3971 (0.4%)	0.95	17/5383 (0.3%)
1	F	0.81	11/3971 (0.3%)	0.91	16/5383 (0.3%)
All	All	0.93	114/23826 (0.5%)	0.97	141/32298 (0.4%)

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	A	1	1
1	B	0	1
1	C	0	1
1	D	0	2
1	E	0	1
1	F	0	3
All	All	1	9

The worst 5 of 114 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	328	VAL	CB-CG1	-19.15	1.12	1.52
1	A	328	VAL	CA-CB	-18.55	1.15	1.54
1	D	26	GLU	C-N	-17.58	0.93	1.34
1	A	378	PHE	CE2-CZ	-16.47	1.06	1.37
1	D	328	VAL	CB-CG2	-16.30	1.18	1.52

The worst 5 of 141 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	329	ASN	N-CA-C	15.22	152.09	111.00
1	D	25	THR	O-C-N	-13.37	101.31	122.70
1	A	379	ILE	CB-CA-C	-12.96	85.69	111.60
1	D	325	ALA	CB-CA-C	12.94	129.50	110.10
1	E	376	VAL	CB-CA-C	-12.41	87.82	111.40

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	377	MET	CA

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	377	MET	Mainchain
1	B	251	TYR	Sidechain
1	C	251	TYR	Sidechain
1	D	251	TYR	Sidechain
1	D	327	LEU	Mainchain

## 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	3876	0	3788	292	0
1	B	3876	0	3788	290	0
1	C	3876	0	3788	276	2
1	D	3876	0	3787	308	0
1	E	3876	0	3789	300	2
1	F	3876	0	3788	302	0
All	All	23256	0	22728	1727	2

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 38.

The worst 5 of 1727 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:E:24:PHE:CD2	1:E:325:ALA:HB1	1.34	1.62
1:E:419:LYS:CG	1:E:470:THR:CG2	1.80	1.58
1:B:410:VAL:HG11	1:B:412:TYR:CZ	1.23	1.57
1:F:419:LYS:CG	1:F:470:THR:CG2	1.77	1.57
1:E:24:PHE:HD2	1:E:325:ALA:CB	1.17	1.57

All (2) 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:144:LYS:CE	1:E:431:LYS:CE[1_556]	1.60	0.60
1:C:144:LYS:CE	1:E:431:LYS:NZ[1_556]	1.90	0.30

## 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	476/482 (99%)	409 (86%)	62 (13%)	5 (1%)	17	58
1	B	476/482 (99%)	409 (86%)	61 (13%)	6 (1%)	15	53
1	C	476/482 (99%)	410 (86%)	61 (13%)	5 (1%)	17	58
1	D	476/482 (99%)	403 (85%)	70 (15%)	3 (1%)	30	72
1	E	476/482 (99%)	407 (86%)	66 (14%)	3 (1%)	30	72
1	F	476/482 (99%)	406 (85%)	68 (14%)	2 (0%)	39	80
All	All	2856/2892 (99%)	2444 (86%)	388 (14%)	24 (1%)	24	66

5 of 24 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	328	VAL
1	A	379	ILE
1	B	172	GLU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	C	465	THR
1	E	328	VAL

### 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	417/418 (100%)	376 (90%)	41 (10%)	10	36
1	B	417/418 (100%)	375 (90%)	42 (10%)	9	34
1	C	417/418 (100%)	374 (90%)	43 (10%)	9	33
1	D	417/418 (100%)	376 (90%)	41 (10%)	10	36
1	E	417/418 (100%)	378 (91%)	39 (9%)	11	39
1	F	417/418 (100%)	377 (90%)	40 (10%)	10	38
All	All	2502/2508 (100%)	2256 (90%)	246 (10%)	10	36

5 of 246 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	378	PHE
1	D	256	ARG
1	F	342	ILE
1	C	403	LYS
1	C	459	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 77 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	388	ASN
1	D	178	GLN
1	F	380	ASN
1	C	411	ASN
1	D	19	HIS

### 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	A	481/482 (99%)	-0.70	0 100 100	14, 42, 75, 115	0
1	B	481/482 (99%)	-0.54	1 (0%) 95 87	15, 47, 92, 117	0
1	C	481/482 (99%)	-0.69	0 100 100	17, 40, 66, 88	0
1	D	481/482 (99%)	-0.57	0 100 100	17, 54, 85, 110	0
1	E	481/482 (99%)	-0.31	3 (0%) 90 73	31, 67, 93, 115	0
1	F	481/482 (99%)	-0.31	2 (0%) 93 80	33, 64, 91, 129	0
All	All	2886/2892 (99%)	-0.52	6 (0%) 95 87	14, 52, 88, 129	0

The worst 5 of 6 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	146	ASN	3.3
1	F	402	GLY	3.2
1	E	325	ALA	2.4
1	F	156	TYR	2.3
1	B	476	CYS	2.2

### 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 [i](#)

There are no such residues in this entry.