



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 19, 2017 – 12:36 PM EDT

PDB ID : 3MI6
Title : Crystal structure of the alpha-galactosidase from Lactobacillus brevis, Northeast Structural Genomics Consortium Target LbR11.
Authors : Vorobiev, S.; Chen, Y.; Seetharaman, J.; Belote, R.; Sahdev, S.; Xiao, R.; Acton, T.B.; Everett, J.K.; Montelione, G.T.; Hunt, J.F.; Tong, L.; Northeast Structural Genomics Consortium (NESG)
Deposited on : unknown
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

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
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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)	:	rb-20030345

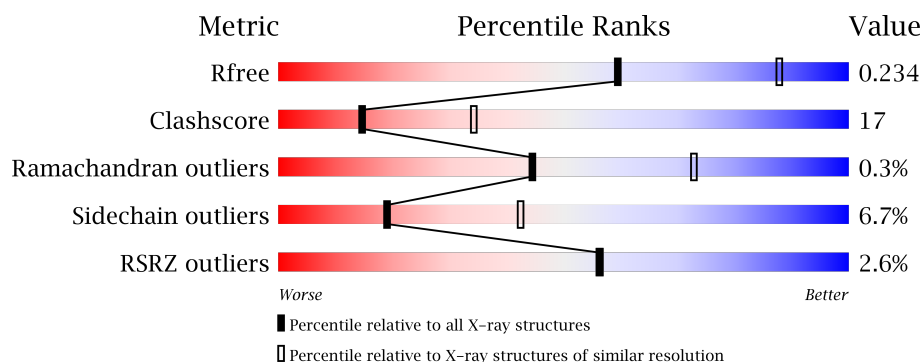
1 Overall quality at a glance

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}	100719	2259 (2.70-2.70)
Clashscore	112137	2590 (2.70-2.70)
Ramachandran outliers	110173	2550 (2.70-2.70)
Sidechain outliers	110143	2550 (2.70-2.70)
RSRZ outliers	101464	2275 (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 ≥ 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	745	<div> <div>2%</div> <div>65%</div> <div>32%</div> <div>.</div> </div>
1	B	745	<div> <div>71%</div> <div>25%</div> <div>..</div> </div>
1	C	745	<div> <div>5%</div> <div>64%</div> <div>31%</div> <div>..</div> </div>
1	D	745	<div> <div>2%</div> <div>71%</div> <div>25%</div> <div>.</div> </div>

2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 24417 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-galactosidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	742	Total	C	N	O	S	Se	0	0	0
			6007	3822	1022	1137	2	24			
1	B	736	Total	C	N	O	S	Se	0	0	0
			5947	3786	1008	1127	2	24			
1	C	737	Total	C	N	O	S	Se	0	0	0
			5946	3782	1008	1130	2	24			
1	D	743	Total	C	N	O	S	Se	0	0	0
			6018	3829	1027	1136	2	24			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	738	LEU	-	expression tag	UNP Q03PP7
A	739	GLU	-	expression tag	UNP Q03PP7
A	740	HIS	-	expression tag	UNP Q03PP7
A	741	HIS	-	expression tag	UNP Q03PP7
A	742	HIS	-	expression tag	UNP Q03PP7
A	743	HIS	-	expression tag	UNP Q03PP7
A	744	HIS	-	expression tag	UNP Q03PP7
A	745	HIS	-	expression tag	UNP Q03PP7
B	738	LEU	-	expression tag	UNP Q03PP7
B	739	GLU	-	expression tag	UNP Q03PP7
B	740	HIS	-	expression tag	UNP Q03PP7
B	741	HIS	-	expression tag	UNP Q03PP7
B	742	HIS	-	expression tag	UNP Q03PP7
B	743	HIS	-	expression tag	UNP Q03PP7
B	744	HIS	-	expression tag	UNP Q03PP7
B	745	HIS	-	expression tag	UNP Q03PP7
C	738	LEU	-	expression tag	UNP Q03PP7
C	739	GLU	-	expression tag	UNP Q03PP7
C	740	HIS	-	expression tag	UNP Q03PP7
C	741	HIS	-	expression tag	UNP Q03PP7
C	742	HIS	-	expression tag	UNP Q03PP7

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	743	HIS	-	expression tag	UNP Q03PP7
C	744	HIS	-	expression tag	UNP Q03PP7
C	745	HIS	-	expression tag	UNP Q03PP7
D	738	LEU	-	expression tag	UNP Q03PP7
D	739	GLU	-	expression tag	UNP Q03PP7
D	740	HIS	-	expression tag	UNP Q03PP7
D	741	HIS	-	expression tag	UNP Q03PP7
D	742	HIS	-	expression tag	UNP Q03PP7
D	743	HIS	-	expression tag	UNP Q03PP7
D	744	HIS	-	expression tag	UNP Q03PP7
D	745	HIS	-	expression tag	UNP Q03PP7

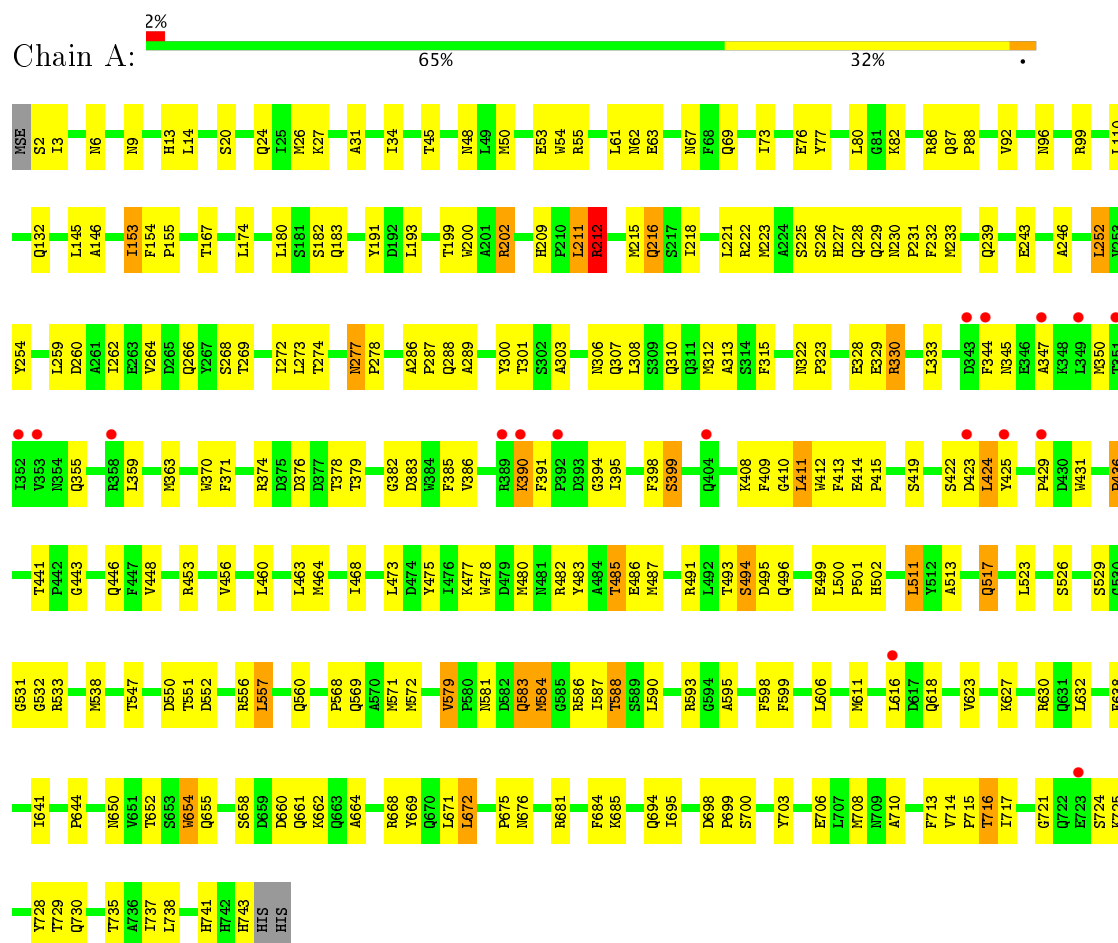
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	117	Total	O	0	0
			117	117		
2	B	179	Total	O	0	0
			179	179		
2	C	84	Total	O	0	0
			84	84		
2	D	119	Total	O	0	0
			119	119		

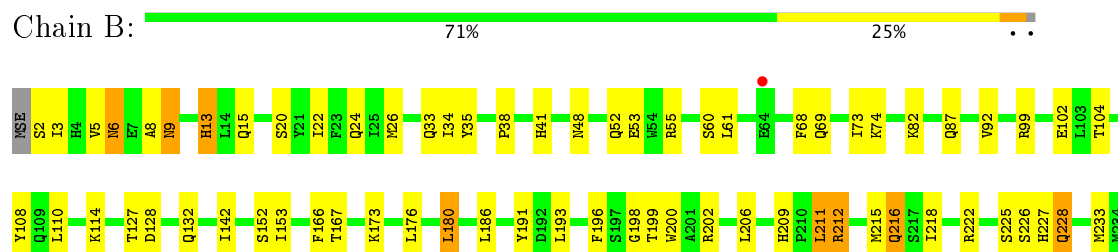
3 Residue-property plots

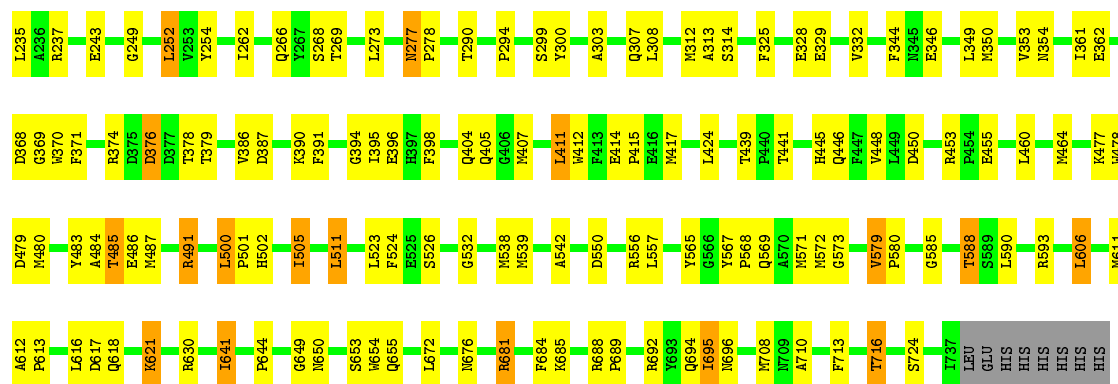
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: Alpha-galactosidase

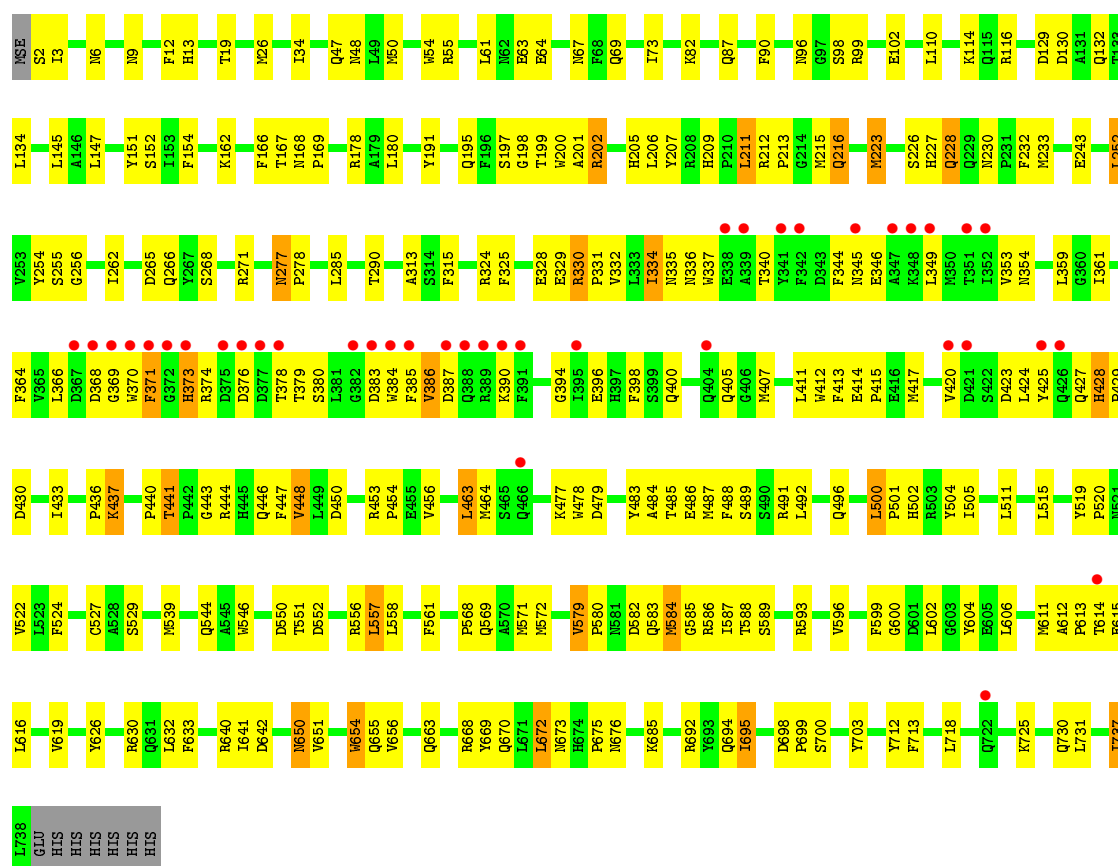


• Molecule 1: Alpha-galactosidase

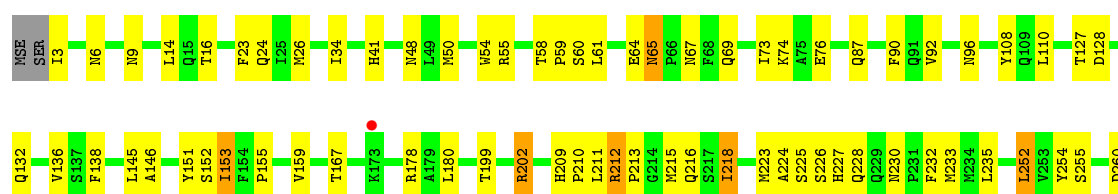


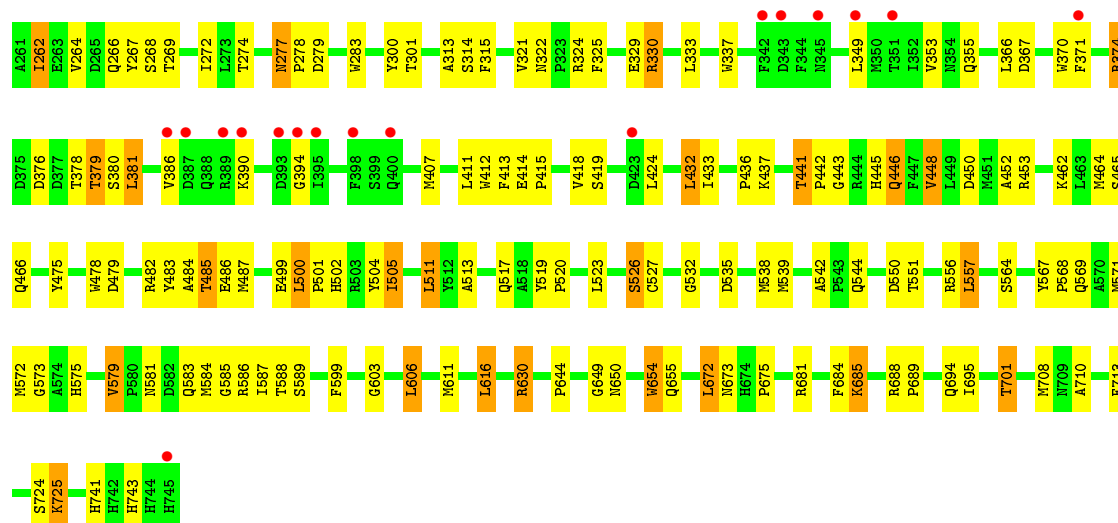


• Molecule 1: Alpha-galactosidase



• Molecule 1: Alpha-galactosidase





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	115.94Å 158.82Å 166.02Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.03 – 2.70 45.53 – 2.70	Depositor EDS
% Data completeness (in resolution range)	99.7 (40.03-2.70) 99.7 (45.53-2.70)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	6.11 (at 2.69Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.5_2), CNS	Depositor
R, R_{free}	0.173 , 0.241 0.164 , 0.234	Depositor DCC
R_{free} test set	4193 reflections (4.99%)	DCC
Wilson B-factor (Å ²)	41.8	Xtriage
Anisotropy	0.370	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 31.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.011 for -h,l,k	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	24417	wwPDB-VP
Average B, all atoms (Å ²)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.85% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	0.46	0/6150	0.62	2/8319 (0.0%)
1	B	0.49	0/6086	0.65	2/8232 (0.0%)
1	C	0.44	0/6083	0.60	2/8227 (0.0%)
1	D	0.47	0/6163	0.63	2/8336 (0.0%)
All	All	0.46	0/24482	0.62	8/33114 (0.0%)

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	695	ILE	CB-CA-C	-5.93	99.75	111.60
1	C	371	PHE	CB-CA-C	-5.66	99.09	110.40
1	A	695	ILE	CB-CA-C	-5.53	100.54	111.60
1	B	237	ARG	NE-CZ-NH1	-5.43	117.58	120.30
1	B	695	ILE	CB-CA-C	-5.42	100.75	111.60

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	6007	0	5701	230	0
1	B	5947	0	5654	192	0
1	C	5946	0	5650	237	0
1	D	6018	0	5706	187	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	117	0	0	9	0
2	B	179	0	0	13	0
2	C	84	0	0	4	0
2	D	119	0	0	10	0
All	All	24417	0	22711	799	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 17.

The worst 5 of 799 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:A:233:MSE:HE1	1:A:262:ILE:HD11	1.27	1.11
1:B:299:SER:HB3	1:B:312:MSE:HE2	1.10	1.05
1:C:584:MSE:HE1	1:C:586:ARG:HB2	1.41	1.00
1:D:226:SER:HB2	1:D:228:GLN:NE2	1.80	0.97
1:C:167:THR:HG23	1:C:290:THR:HG22	1.47	0.96

There are no symmetry-related clashes.

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	740/745 (99%)	698 (94%)	38 (5%)	4 (0%)	32	60
1	B	734/745 (98%)	697 (95%)	36 (5%)	1 (0%)	55	82
1	C	735/745 (99%)	677 (92%)	55 (8%)	3 (0%)	38	66
1	D	741/745 (100%)	691 (93%)	48 (6%)	2 (0%)	44	73
All	All	2950/2980 (99%)	2763 (94%)	177 (6%)	10 (0%)	44	73

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	386	VAL
1	D	374	ARG
1	A	77	TYR
1	A	480	MSE
1	C	337	TRP

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	645/624 (103%)	605 (94%)	40 (6%)	21	46
1	B	638/624 (102%)	597 (94%)	41 (6%)	20	45
1	C	638/624 (102%)	592 (93%)	46 (7%)	17	39
1	D	645/624 (103%)	600 (93%)	45 (7%)	18	40
All	All	2566/2496 (103%)	2394 (93%)	172 (7%)	19	42

5 of 172 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	654	TRP
1	C	268	SER
1	D	500	LEU
1	B	716	THR
1	C	180	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 117 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	618	GLN
1	C	216	GLN
1	D	445	HIS
1	B	655	GLN
1	C	48	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 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, 95th 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 ²)	Q<0.9
1	A	718/745 (96%)	-0.31	17 (2%) 59 60	24, 38, 84, 103	0
1	B	712/745 (95%)	-0.47	1 (0%) 95 96	23, 35, 49, 80	0
1	C	713/745 (95%)	-0.03	39 (5%) 26 24	25, 44, 97, 104	0
1	D	719/745 (96%)	-0.28	18 (2%) 58 58	25, 39, 67, 94	0
All	All	2862/2980 (96%)	-0.27	75 (2%) 56 56	23, 38, 81, 104	0

The worst 5 of 75 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	389	ARG	7.4
1	C	349	LEU	5.6
1	C	391	PHE	5.4
1	C	342	PHE	5.3
1	C	376	ASP	5.1

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.