wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 12:07 pm BST

PDB ID	:	5XRC
Title	:	A Trimodular GH5_4 Subfamily Endoglucanase Structure with Large Unit Cell
Authors	:	Zhang, H.D.
Deposited on	:	2017-06-08
Resolution	:	2.90 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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

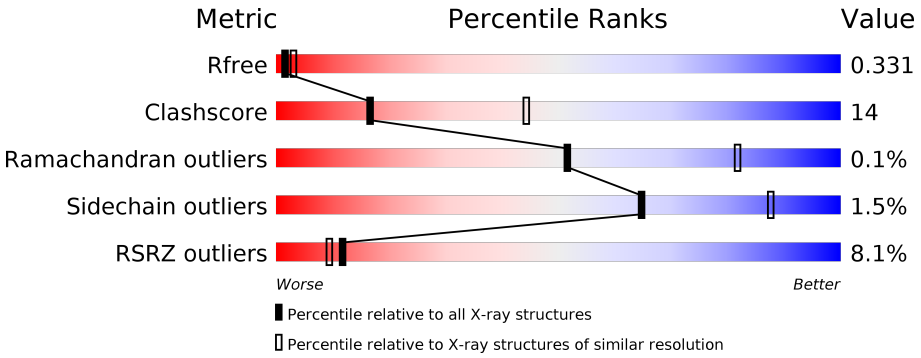
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.90 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	1957 (2.90-2.90)
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)
RSRZ outliers	127900	1906 (2.90-2.90)

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 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 <=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	585	% 71% 19% 9%
1	B	585	20% 55% 34% • 9%
1	C	585	% 69% 21% • 9%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 12970 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 Cellulase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	531	Total	C	N	O	S	0	0	0
			4312	2724	727	848	13			
1	B	530	Total	C	N	O	S	0	0	0
			4303	2719	725	846	13			
1	C	532	Total	C	N	O	S	0	0	0
			4321	2730	729	849	13			

There are 147 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-33	MET	-	expression tag	UNP D4P8C6
A	-32	GLY	-	expression tag	UNP D4P8C6
A	-31	SER	-	expression tag	UNP D4P8C6
A	-30	SER	-	expression tag	UNP D4P8C6
A	-29	HIS	-	expression tag	UNP D4P8C6
A	-28	HIS	-	expression tag	UNP D4P8C6
A	-27	HIS	-	expression tag	UNP D4P8C6
A	-26	HIS	-	expression tag	UNP D4P8C6
A	-25	HIS	-	expression tag	UNP D4P8C6
A	-24	HIS	-	expression tag	UNP D4P8C6
A	-23	SER	-	expression tag	UNP D4P8C6
A	-22	SER	-	expression tag	UNP D4P8C6
A	-21	GLY	-	expression tag	UNP D4P8C6
A	-20	LEU	-	expression tag	UNP D4P8C6
A	-19	VAL	-	expression tag	UNP D4P8C6
A	-18	PRO	-	expression tag	UNP D4P8C6
A	-17	ARG	-	expression tag	UNP D4P8C6
A	-16	GLY	-	expression tag	UNP D4P8C6
A	-15	SER	-	expression tag	UNP D4P8C6
A	-14	HIS	-	expression tag	UNP D4P8C6
A	-13	MET	-	expression tag	UNP D4P8C6
A	-12	ALA	-	expression tag	UNP D4P8C6
A	-11	SER	-	expression tag	UNP D4P8C6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-10	MET	-	expression tag	UNP D4P8C6
A	-9	THR	-	expression tag	UNP D4P8C6
A	-8	GLY	-	expression tag	UNP D4P8C6
A	-7	GLY	-	expression tag	UNP D4P8C6
A	-6	GLN	-	expression tag	UNP D4P8C6
A	-5	GLN	-	expression tag	UNP D4P8C6
A	-4	MET	-	expression tag	UNP D4P8C6
A	-3	GLY	-	expression tag	UNP D4P8C6
A	-2	ARG	-	expression tag	UNP D4P8C6
A	-1	GLY	-	expression tag	UNP D4P8C6
A	0	SER	-	expression tag	UNP D4P8C6
A	537	VAL	-	expression tag	UNP D4P8C6
A	538	ASP	-	expression tag	UNP D4P8C6
A	539	LYS	-	expression tag	UNP D4P8C6
A	540	LEU	-	expression tag	UNP D4P8C6
A	541	ALA	-	expression tag	UNP D4P8C6
A	542	ALA	-	expression tag	UNP D4P8C6
A	543	ALA	-	expression tag	UNP D4P8C6
A	544	LEU	-	expression tag	UNP D4P8C6
A	545	GLU	-	expression tag	UNP D4P8C6
A	546	HIS	-	expression tag	UNP D4P8C6
A	547	HIS	-	expression tag	UNP D4P8C6
A	548	HIS	-	expression tag	UNP D4P8C6
A	549	HIS	-	expression tag	UNP D4P8C6
A	550	HIS	-	expression tag	UNP D4P8C6
A	551	HIS	-	expression tag	UNP D4P8C6
B	-33	MET	-	expression tag	UNP D4P8C6
B	-32	GLY	-	expression tag	UNP D4P8C6
B	-31	SER	-	expression tag	UNP D4P8C6
B	-30	SER	-	expression tag	UNP D4P8C6
B	-29	HIS	-	expression tag	UNP D4P8C6
B	-28	HIS	-	expression tag	UNP D4P8C6
B	-27	HIS	-	expression tag	UNP D4P8C6
B	-26	HIS	-	expression tag	UNP D4P8C6
B	-25	HIS	-	expression tag	UNP D4P8C6
B	-24	HIS	-	expression tag	UNP D4P8C6
B	-23	SER	-	expression tag	UNP D4P8C6
B	-22	SER	-	expression tag	UNP D4P8C6
B	-21	GLY	-	expression tag	UNP D4P8C6
B	-20	LEU	-	expression tag	UNP D4P8C6
B	-19	VAL	-	expression tag	UNP D4P8C6
B	-18	PRO	-	expression tag	UNP D4P8C6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	-17	ARG	-	expression tag	UNP D4P8C6
B	-16	GLY	-	expression tag	UNP D4P8C6
B	-15	SER	-	expression tag	UNP D4P8C6
B	-14	HIS	-	expression tag	UNP D4P8C6
B	-13	MET	-	expression tag	UNP D4P8C6
B	-12	ALA	-	expression tag	UNP D4P8C6
B	-11	SER	-	expression tag	UNP D4P8C6
B	-10	MET	-	expression tag	UNP D4P8C6
B	-9	THR	-	expression tag	UNP D4P8C6
B	-8	GLY	-	expression tag	UNP D4P8C6
B	-7	GLY	-	expression tag	UNP D4P8C6
B	-6	GLN	-	expression tag	UNP D4P8C6
B	-5	GLN	-	expression tag	UNP D4P8C6
B	-4	MET	-	expression tag	UNP D4P8C6
B	-3	GLY	-	expression tag	UNP D4P8C6
B	-2	ARG	-	expression tag	UNP D4P8C6
B	-1	GLY	-	expression tag	UNP D4P8C6
B	0	SER	-	expression tag	UNP D4P8C6
B	537	VAL	-	expression tag	UNP D4P8C6
B	538	ASP	-	expression tag	UNP D4P8C6
B	539	LYS	-	expression tag	UNP D4P8C6
B	540	LEU	-	expression tag	UNP D4P8C6
B	541	ALA	-	expression tag	UNP D4P8C6
B	542	ALA	-	expression tag	UNP D4P8C6
B	543	ALA	-	expression tag	UNP D4P8C6
B	544	LEU	-	expression tag	UNP D4P8C6
B	545	GLU	-	expression tag	UNP D4P8C6
B	546	HIS	-	expression tag	UNP D4P8C6
B	547	HIS	-	expression tag	UNP D4P8C6
B	548	HIS	-	expression tag	UNP D4P8C6
B	549	HIS	-	expression tag	UNP D4P8C6
B	550	HIS	-	expression tag	UNP D4P8C6
B	551	HIS	-	expression tag	UNP D4P8C6
C	-33	MET	-	expression tag	UNP D4P8C6
C	-32	GLY	-	expression tag	UNP D4P8C6
C	-31	SER	-	expression tag	UNP D4P8C6
C	-30	SER	-	expression tag	UNP D4P8C6
C	-29	HIS	-	expression tag	UNP D4P8C6
C	-28	HIS	-	expression tag	UNP D4P8C6
C	-27	HIS	-	expression tag	UNP D4P8C6
C	-26	HIS	-	expression tag	UNP D4P8C6
C	-25	HIS	-	expression tag	UNP D4P8C6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-24	HIS	-	expression tag	UNP D4P8C6
C	-23	SER	-	expression tag	UNP D4P8C6
C	-22	SER	-	expression tag	UNP D4P8C6
C	-21	GLY	-	expression tag	UNP D4P8C6
C	-20	LEU	-	expression tag	UNP D4P8C6
C	-19	VAL	-	expression tag	UNP D4P8C6
C	-18	PRO	-	expression tag	UNP D4P8C6
C	-17	ARG	-	expression tag	UNP D4P8C6
C	-16	GLY	-	expression tag	UNP D4P8C6
C	-15	SER	-	expression tag	UNP D4P8C6
C	-14	HIS	-	expression tag	UNP D4P8C6
C	-13	MET	-	expression tag	UNP D4P8C6
C	-12	ALA	-	expression tag	UNP D4P8C6
C	-11	SER	-	expression tag	UNP D4P8C6
C	-10	MET	-	expression tag	UNP D4P8C6
C	-9	THR	-	expression tag	UNP D4P8C6
C	-8	GLY	-	expression tag	UNP D4P8C6
C	-7	GLY	-	expression tag	UNP D4P8C6
C	-6	GLN	-	expression tag	UNP D4P8C6
C	-5	GLN	-	expression tag	UNP D4P8C6
C	-4	MET	-	expression tag	UNP D4P8C6
C	-3	GLY	-	expression tag	UNP D4P8C6
C	-2	ARG	-	expression tag	UNP D4P8C6
C	-1	GLY	-	expression tag	UNP D4P8C6
C	0	SER	-	expression tag	UNP D4P8C6
C	537	VAL	-	expression tag	UNP D4P8C6
C	538	ASP	-	expression tag	UNP D4P8C6
C	539	LYS	-	expression tag	UNP D4P8C6
C	540	LEU	-	expression tag	UNP D4P8C6
C	541	ALA	-	expression tag	UNP D4P8C6
C	542	ALA	-	expression tag	UNP D4P8C6
C	543	ALA	-	expression tag	UNP D4P8C6
C	544	LEU	-	expression tag	UNP D4P8C6
C	545	GLU	-	expression tag	UNP D4P8C6
C	546	HIS	-	expression tag	UNP D4P8C6
C	547	HIS	-	expression tag	UNP D4P8C6
C	548	HIS	-	expression tag	UNP D4P8C6
C	549	HIS	-	expression tag	UNP D4P8C6
C	550	HIS	-	expression tag	UNP D4P8C6
C	551	HIS	-	expression tag	UNP D4P8C6

- Molecule 2 is water.

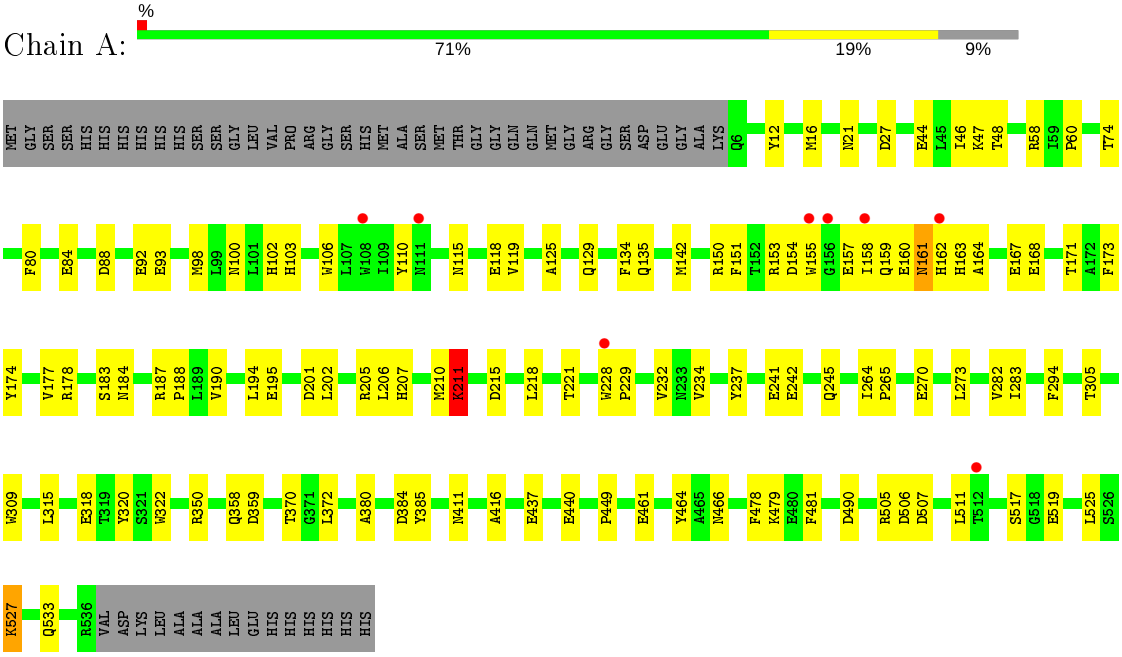
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	8	Total O 8 8	0	0
2	B	1	Total O 1 1	0	0
2	C	25	Total O 25 25	0	0

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.

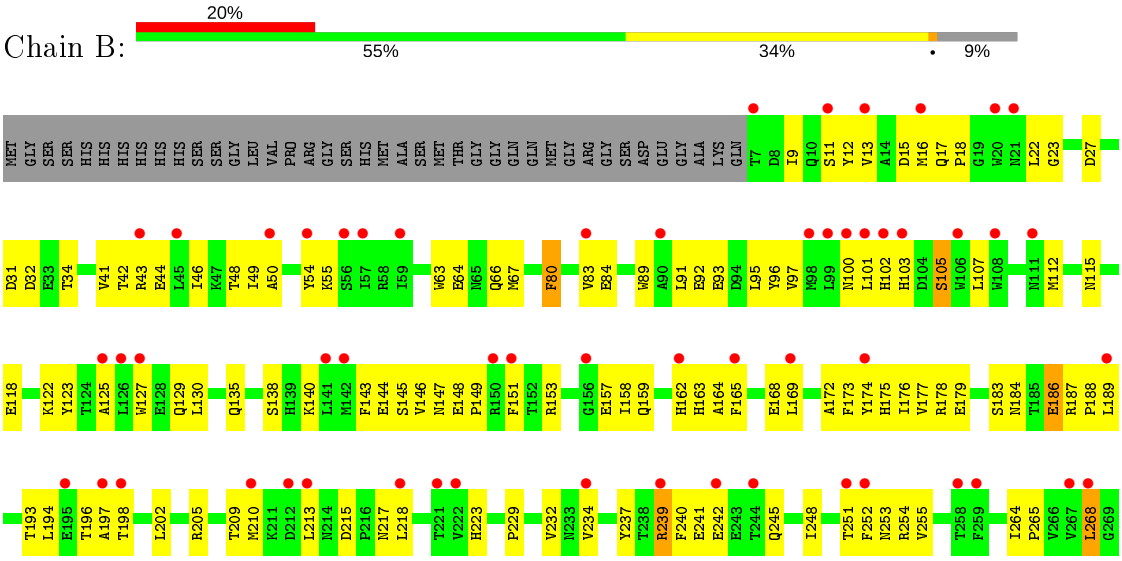
- Molecule 1: Cellulase

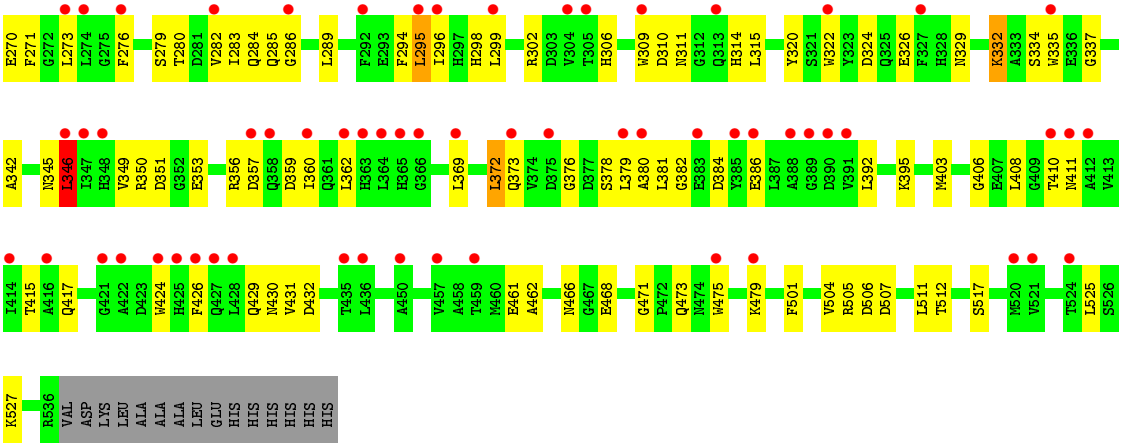
Chain A:



- Molecule 1: Cellulase

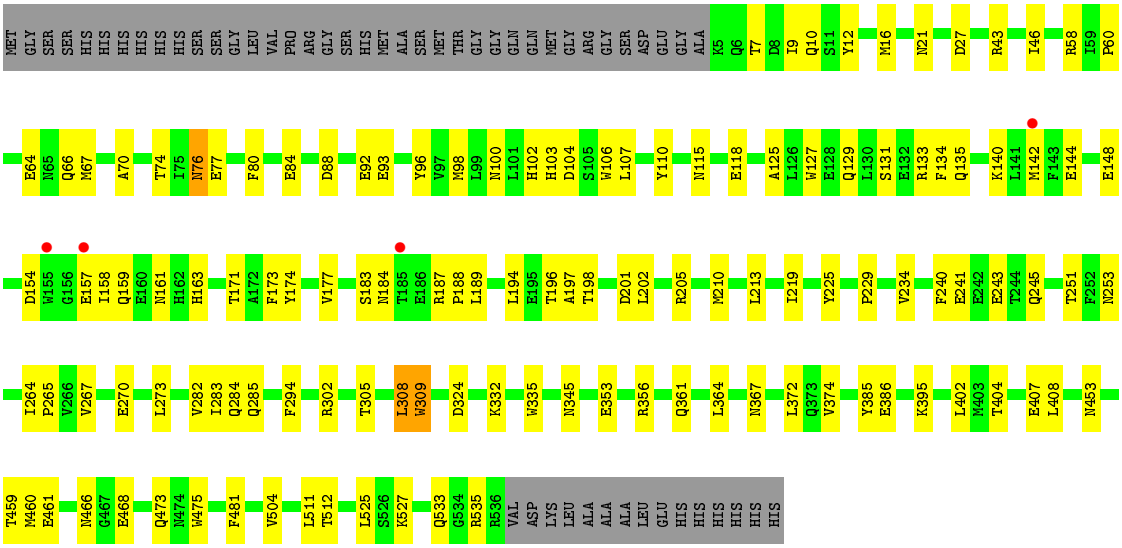
Chain B:





- Molecule 1: Cellulase

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	I 4 2 2	Depositor
Cell constants a, b, c, α, β, γ	120.14Å 120.14Å 615.34Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	85.97 – 2.90 85.97 – 2.90	Depositor EDS
% Data completeness (in resolution range)	98.7 (85.97-2.90) 99.0 (85.97-2.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.48 (at 2.91Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575)	Depositor
R, R_{free}	0.276 , 0.332 0.277 , 0.331	Depositor DCC
R_{free} test set	2408 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å^2)	33.3	Xtriage
Anisotropy	0.030	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 58.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	12970	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 79.99 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 4.9383e-07. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[1] Intensities estimated from amplitudes.

[2] Theoretical values of $< |L| >$, $< L^2 >$ for acentric reflections are 0.5, 0.333 respectively for untwinned datasets, and 0.375, 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	0.30	0/4430	0.53	2/6032 (0.0%)
1	B	0.37	1/4421 (0.0%)	0.64	6/6020 (0.1%)
1	C	0.30	1/4439 (0.0%)	0.53	2/6043 (0.0%)
All	All	0.33	2/13290 (0.0%)	0.57	10/18095 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	308	LEU	C-N	-9.01	1.13	1.34
1	B	186	GLU	CB-CG	6.95	1.65	1.52

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	309	TRP	O-C-N	10.00	138.70	122.70
1	B	239	ARG	NE-CZ-NH2	-9.15	115.72	120.30
1	C	309	TRP	CA-C-N	-7.82	99.99	117.20
1	A	211	LYS	CA-CB-CG	7.51	129.93	113.40
1	B	239	ARG	CG-CD-NE	-7.41	96.25	111.80

There are no chirality outliers.

There are no planarity outliers.

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	4312	0	3976	83	0
1	B	4303	0	3968	189	0
1	C	4321	0	3988	83	0
2	A	8	0	0	1	0
2	B	1	0	0	0	0
2	C	25	0	0	0	0
All	All	12970	0	11932	353	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 14.

The worst 5 of 353 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:B:379:LEU:HD21	1:B:384:ASP:CG	1.51	1.30
1:B:360:ILE:HD11	1:B:426:PHE:CE2	1.69	1.28
1:B:424:TRP:CD1	1:B:426:PHE:CE1	2.20	1.27
1:B:379:LEU:HD21	1:B:384:ASP:OD2	1.09	1.19
1:B:381:LEU:HD23	1:B:382:GLY:N	1.57	1.17

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	529/585 (90%)	507 (96%)	21 (4%)	1 (0%)	47 78
1	B	528/585 (90%)	500 (95%)	28 (5%)	0	100 100
1	C	530/585 (91%)	505 (95%)	24 (4%)	1 (0%)	47 78
All	All	1587/1755 (90%)	1512 (95%)	73 (5%)	2 (0%)	51 82

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	70	ALA
1	A	160	GLU

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	462/503 (92%)	453 (98%)	9 (2%)	57 84
1	B	461/503 (92%)	452 (98%)	9 (2%)	55 82
1	C	463/503 (92%)	460 (99%)	3 (1%)	86 96
All	All	1386/1509 (92%)	1365 (98%)	21 (2%)	65 87

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	43	ARG
1	B	84	GLU
1	B	517	SER
1	A	527	LYS
1	C	76	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 16 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	345	ASN
1	B	365	HIS
1	C	253	ASN
1	B	217	ASN
1	C	411	ASN

5.3.3 RNA

There are no RNA molecules 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

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	C	1

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	C	308:LEU	C	309:TRP	N	1.13

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^{th} 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	531/585 (90%)	0.44	8 (1%) 73 73	26, 55, 82, 119	0
1	B	530/585 (90%)	1.34	117 (22%) 0 0	30, 90, 108, 126	0
1	C	532/585 (90%)	0.27	4 (0%) 86 86	14, 29, 49, 83	0
All	All	1593/1755 (90%)	0.68	129 (8%) 12 9	14, 56, 101, 126	0

The worst 5 of 129 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	414	ILE	4.4
1	B	426	PHE	3.8
1	B	99	LEU	3.7
1	B	57	ILE	3.6
1	B	90	ALA	3.6

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.