wwPDB X-ray Structure Validation Summary Report (i)

Jun 16, 2024 – 09:50 AM EDT

PDB ID	:	5DJW
Title	:	Crystal structure of Family 31 alpha-glucosidase (BT_3299) from Bacteroides thetaiotaomicron
Authors	:	Chaudet, M.M.; Rose, D.R.
Deposited on	:	2015-09-02
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
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

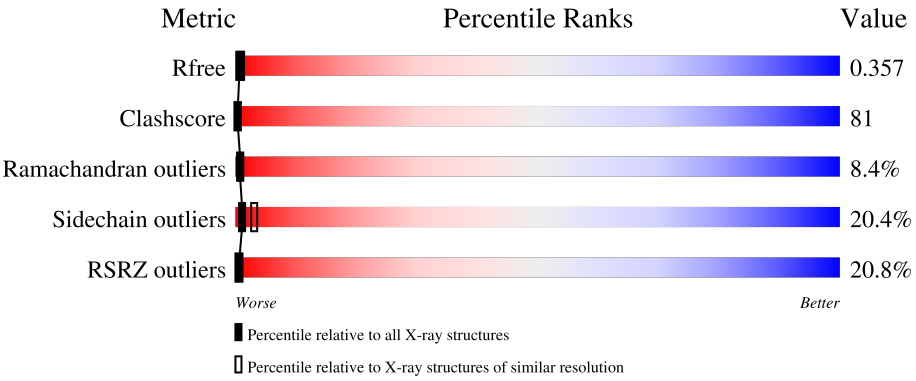
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.70 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}	130704	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (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 of 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	697	18% 22% 45% 15% · 16%
1	B	697	17% 22% 45% 14% · 16%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 9139 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-glucosidase II.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	588	Total	C	N	O	S	0	0	0
			4569	2924	782	843	20			
1	B	587	Total	C	N	O	S	0	0	0
			4547	2913	783	831	20			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP Q8A2K6
A	685	HIS	-	expression tag	UNP Q8A2K6
A	686	HIS	-	expression tag	UNP Q8A2K6
A	687	HIS	-	expression tag	UNP Q8A2K6
A	688	HIS	-	expression tag	UNP Q8A2K6
A	689	HIS	-	expression tag	UNP Q8A2K6
A	690	HIS	-	expression tag	UNP Q8A2K6
A	691	LEU	-	expression tag	UNP Q8A2K6
A	692	ARG	-	expression tag	UNP Q8A2K6
A	693	VAL	-	expression tag	UNP Q8A2K6
A	694	PRO	-	expression tag	UNP Q8A2K6
A	695	ARG	-	expression tag	UNP Q8A2K6
A	696	GLY	-	expression tag	UNP Q8A2K6
A	697	SER	-	expression tag	UNP Q8A2K6
B	1	MET	-	initiating methionine	UNP Q8A2K6
B	685	HIS	-	expression tag	UNP Q8A2K6
B	686	HIS	-	expression tag	UNP Q8A2K6
B	687	HIS	-	expression tag	UNP Q8A2K6
B	688	HIS	-	expression tag	UNP Q8A2K6
B	689	HIS	-	expression tag	UNP Q8A2K6
B	690	HIS	-	expression tag	UNP Q8A2K6
B	691	LEU	-	expression tag	UNP Q8A2K6
B	692	ARG	-	expression tag	UNP Q8A2K6
B	693	VAL	-	expression tag	UNP Q8A2K6
B	694	PRO	-	expression tag	UNP Q8A2K6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	695	ARG	-	expression tag	UNP Q8A2K6
B	696	GLY	-	expression tag	UNP Q8A2K6
B	697	SER	-	expression tag	UNP Q8A2K6

- Molecule 2 is water.

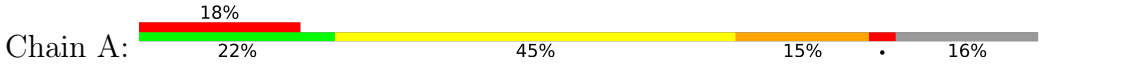
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	15	Total O 15 15	0	0
2	B	8	Total O 8 8	0	0

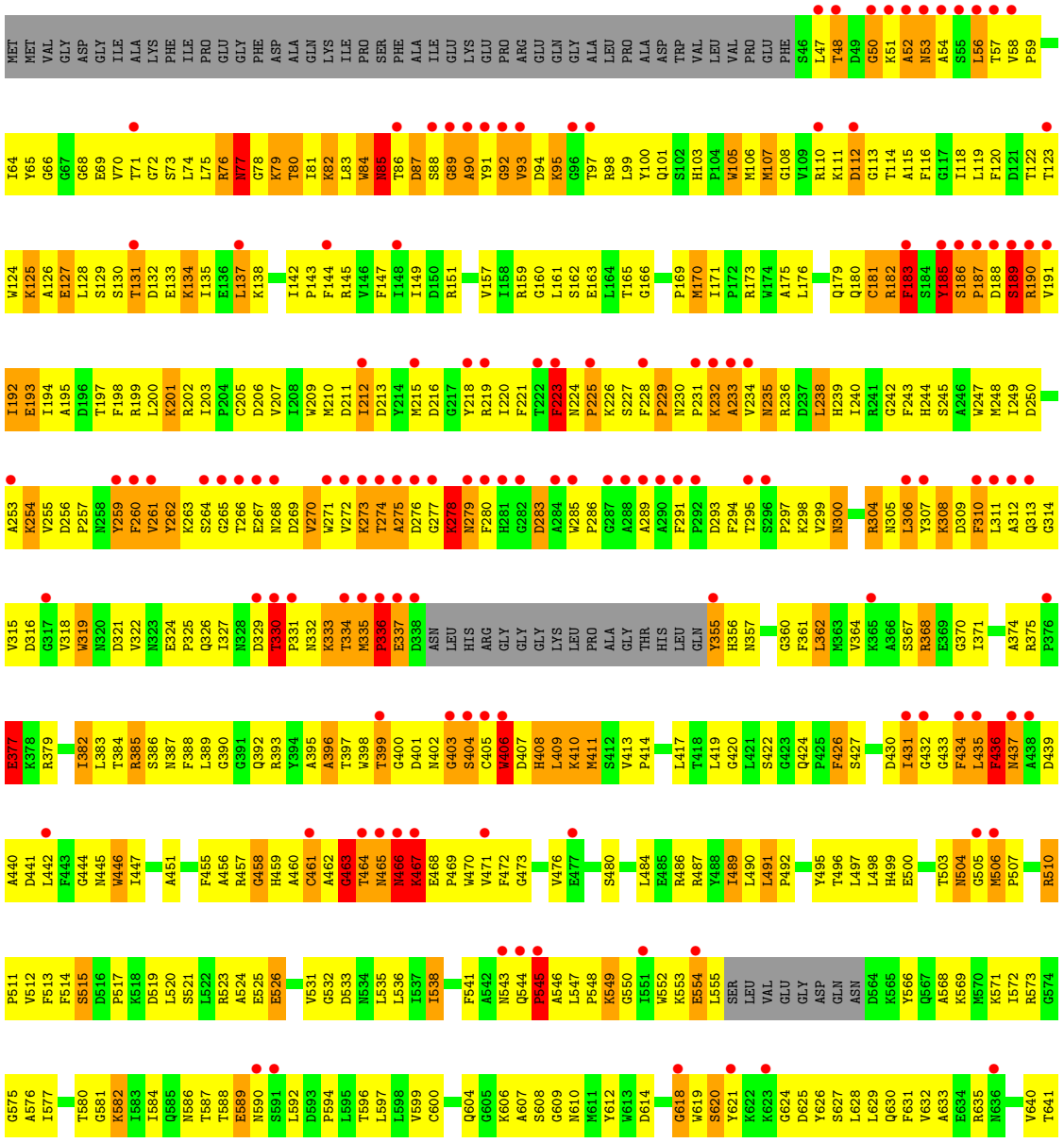
3 Residue-property plots (i)

These plots are drawn for all protein, RNA, DNA and oligosaccharide 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-glucosidase II

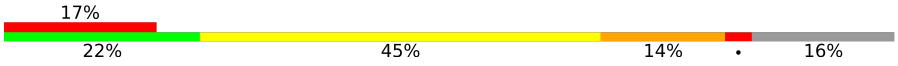
Chain A:

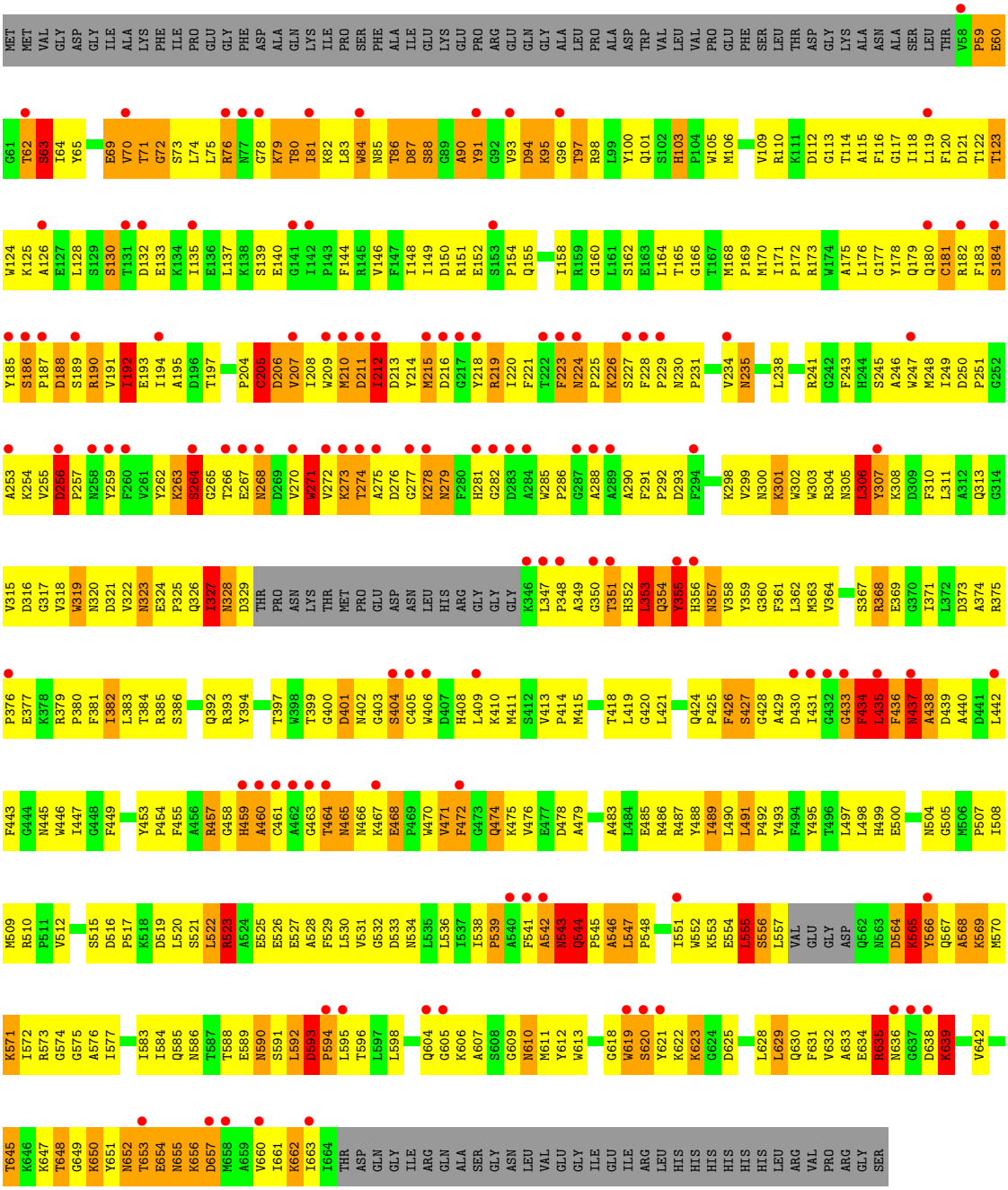




- Molecule 1: Alpha-glucosidase II

Chain B:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	89.26Å 74.86Å 94.43Å 90.00° 95.70° 90.00°	Depositor
Resolution (Å)	47.55 – 2.70 47.55 – 2.70	Depositor EDS
% Data completeness (in resolution range)	100.0 (47.55-2.70) 100.0 (47.55-2.70)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.66 (at 2.69Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.330 , 0.390 0.314 , 0.357	Depositor DCC
R_{free} test set	1716 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	44.0	Xtriage
Anisotropy	0.424	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 89.9	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.81	EDS
Total number of atoms	9139	wwPDB-VP
Average B, all atoms (Å^2)	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.64% of the height of the origin peak. No significant pseudotranslation is detected.*

[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	0.84	3/4694 (0.1%)	0.98	16/6370 (0.3%)
1	B	1.10	2/4672 (0.0%)	1.00	16/6342 (0.3%)
All	All	0.98	5/9366 (0.1%)	0.99	32/12712 (0.3%)

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	0	18
1	B	0	20
All	All	0	38

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	434	PHE	C-N	-49.64	0.19	1.34
1	B	594	PRO	C-N	13.36	1.64	1.34
1	A	50	GLY	C-N	9.53	1.55	1.34
1	A	143	PRO	N-CD	5.13	1.55	1.47
1	A	545	PRO	N-CD	5.13	1.55	1.47

The worst 5 of 32 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	50	GLY	O-C-N	-18.38	93.30	122.70
1	B	326	GLN	C-N-CA	14.11	156.97	121.70
1	A	309	ASP	CB-CA-C	-13.73	82.94	110.40
1	B	434	PHE	CA-C-N	-11.30	92.34	117.20
1	B	327	ILE	CA-C-N	-10.55	93.98	117.20

There are no chirality outliers.

5 of 38 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	185	TYR	Peptide
1	A	235	ASN	Peptide
1	A	259	TYR	Peptide
1	A	50	GLY	Mainchain
1	A	77	ASN	Peptide

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	4569	0	4348	730	5
1	B	4547	0	4343	709	1
2	A	15	0	0	8	1
2	B	8	0	0	3	0
All	All	9139	0	8691	1435	6

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 81.

The worst 5 of 1435 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:272:VAL:CG1	1:A:336:PRO:HG2	1.40	1.51
1:B:553:LYS:CB	1:B:572:ILE:HG22	1.36	1.50
1:B:256:ASP:O	1:B:259:TYR:CD2	1.73	1.40
1:B:552:TRP:CB	1:B:571:LYS:HE2	1.49	1.39
1:A:256:ASP:O	1:A:259:TYR:CD2	1.73	1.39

The worst 5 of 6 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:A:554:GLU:OE1	1:A:653:THR:OG1[2_557]	0.80	1.40
1:A:521:SER:OG	1:B:377:GLU:CB[2_557]	1.36	0.84

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:554:GLU:CD	1:A:653:THR:OG1[2_557]	1.72	0.48
1:A:554:GLU:OE2	1:A:652:ASN:O[2_557]	1.90	0.30
1:A:554:GLU:OE1	1:A:653:THR:CB[2_557]	1.98	0.22

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	582/697 (84%)	471 (81%)	66 (11%)	45 (8%)	1 1
1	B	581/697 (83%)	464 (80%)	64 (11%)	53 (9%)	1 0
All	All	1163/1394 (83%)	935 (80%)	130 (11%)	98 (8%)	1 1

5 of 98 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	84	TRP
1	A	87	ASP
1	A	90	ALA
1	A	93	VAL
1	A	112	ASP

5.3.2 Protein sidechains (i)

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	464/579 (80%)	369 (80%)	95 (20%)	1 3

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
1	B	462/579 (80%)	368 (80%)	94 (20%)	1 3
All	All	926/1158 (80%)	737 (80%)	189 (20%)	1 3

5 of 189 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	192	ILE
1	B	401	ASP
1	B	219	ARG
1	B	306	LEU
1	B	437	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 17 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	604	GLN
1	B	655	ASN
1	A	424	GLN
1	A	655	ASN
1	B	281	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 monosaccharides 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)

The following chains have linkage breaks:

Mol	Chain	Number of breaks
1	B	2

All chain breaks are listed below:

Model	Chain	Residue-1	Atom-1	Residue-2	Atom-2	Distance (Å)
1	B	594:PRO	C	595:LEU	N	1.64
1	B	434:PHE	C	435:LEU	N	0.19

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	588/697 (84%)	1.22	128 (21%)	0	0	10, 42, 120, 162	0
1	B	587/697 (84%)	1.18	116 (19%)	1	0	11, 43, 110, 166	0
All	All	1175/1394 (84%)	1.20	244 (20%)	1	0	10, 42, 117, 166	0

The worst 5 of 244 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	274	THR	11.1
1	A	260	PHE	8.7
1	B	637	GLY	8.3
1	B	58	VAL	8.0
1	B	211	ASP	7.1

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 monosaccharides 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.