wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2024 – 06:36 PM EST

PDB ID	:	4KC7
Title	:	Crystal Structure of Endo-1,5-alpha-L-arabinanase from Thermotoga petrophila RKU-1
Authors	:	Nascimento, A.F.Z.; Polo, C.C.; Santos, C.R.; Costa, M.C.M.F.; Mesa, A.N.; Prade, R.A.; Ruller, R.; Squina, F.M.; Murakami, M.T.
Deposited on	:	2013-04-24
Resolution	:	1.75 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 (i)) were used in the production of this report:

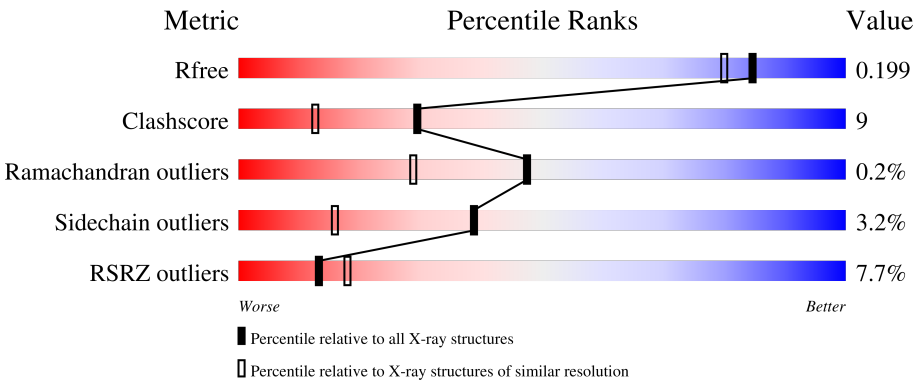
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 1.75 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	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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	474	84% 10% 5%
1	B	474	% 81% 12% • 5%
1	C	474	20% 68% 20% • 9%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 11493 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 Glycoside hydrolase, family 43.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	449	3623	2337	602	674	10	0	6	0
1	B	449	3632	2345	605	672	10	0	7	0
1	C	433	3476	2244	582	640	10	0	3	0

There are 69 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	MET	-	initiating methionine	UNP A5IKD4
A	-1	GLY	-	expression tag	UNP A5IKD4
A	0	SER	-	expression tag	UNP A5IKD4
A	1	SER	-	expression tag	UNP A5IKD4
A	2	HIS	-	expression tag	UNP A5IKD4
A	3	HIS	-	expression tag	UNP A5IKD4
A	4	HIS	-	expression tag	UNP A5IKD4
A	5	HIS	-	expression tag	UNP A5IKD4
A	6	HIS	-	expression tag	UNP A5IKD4
A	7	HIS	-	expression tag	UNP A5IKD4
A	8	SER	-	expression tag	UNP A5IKD4
A	9	SER	-	expression tag	UNP A5IKD4
A	10	GLY	-	expression tag	UNP A5IKD4
A	11	LEU	-	expression tag	UNP A5IKD4
A	12	VAL	-	expression tag	UNP A5IKD4
A	13	PRO	-	expression tag	UNP A5IKD4
A	14	ARG	-	expression tag	UNP A5IKD4
A	15	GLY	-	expression tag	UNP A5IKD4
A	16	SER	-	expression tag	UNP A5IKD4
A	17	HIS	-	expression tag	UNP A5IKD4
A	18	MET	-	expression tag	UNP A5IKD4
A	19	ALA	-	expression tag	UNP A5IKD4
A	20	SER	-	expression tag	UNP A5IKD4

Continued on next page...

Continued from previous page...

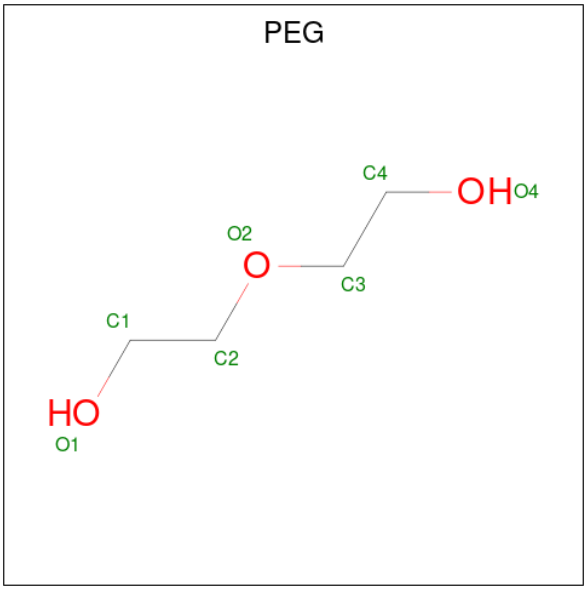
Chain	Residue	Modelled	Actual	Comment	Reference
B	-2	MET	-	initiating methionine	UNP A5IKD4
B	-1	GLY	-	expression tag	UNP A5IKD4
B	0	SER	-	expression tag	UNP A5IKD4
B	1	SER	-	expression tag	UNP A5IKD4
B	2	HIS	-	expression tag	UNP A5IKD4
B	3	HIS	-	expression tag	UNP A5IKD4
B	4	HIS	-	expression tag	UNP A5IKD4
B	5	HIS	-	expression tag	UNP A5IKD4
B	6	HIS	-	expression tag	UNP A5IKD4
B	7	HIS	-	expression tag	UNP A5IKD4
B	8	SER	-	expression tag	UNP A5IKD4
B	9	SER	-	expression tag	UNP A5IKD4
B	10	GLY	-	expression tag	UNP A5IKD4
B	11	LEU	-	expression tag	UNP A5IKD4
B	12	VAL	-	expression tag	UNP A5IKD4
B	13	PRO	-	expression tag	UNP A5IKD4
B	14	ARG	-	expression tag	UNP A5IKD4
B	15	GLY	-	expression tag	UNP A5IKD4
B	16	SER	-	expression tag	UNP A5IKD4
B	17	HIS	-	expression tag	UNP A5IKD4
B	18	MET	-	expression tag	UNP A5IKD4
B	19	ALA	-	expression tag	UNP A5IKD4
B	20	SER	-	expression tag	UNP A5IKD4
C	-2	MET	-	initiating methionine	UNP A5IKD4
C	-1	GLY	-	expression tag	UNP A5IKD4
C	0	SER	-	expression tag	UNP A5IKD4
C	1	SER	-	expression tag	UNP A5IKD4
C	2	HIS	-	expression tag	UNP A5IKD4
C	3	HIS	-	expression tag	UNP A5IKD4
C	4	HIS	-	expression tag	UNP A5IKD4
C	5	HIS	-	expression tag	UNP A5IKD4
C	6	HIS	-	expression tag	UNP A5IKD4
C	7	HIS	-	expression tag	UNP A5IKD4
C	8	SER	-	expression tag	UNP A5IKD4
C	9	SER	-	expression tag	UNP A5IKD4
C	10	GLY	-	expression tag	UNP A5IKD4
C	11	LEU	-	expression tag	UNP A5IKD4
C	12	VAL	-	expression tag	UNP A5IKD4
C	13	PRO	-	expression tag	UNP A5IKD4
C	14	ARG	-	expression tag	UNP A5IKD4
C	15	GLY	-	expression tag	UNP A5IKD4
C	16	SER	-	expression tag	UNP A5IKD4

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	17	HIS	-	expression tag	UNP A5IKD4
C	18	MET	-	expression tag	UNP A5IKD4
C	19	ALA	-	expression tag	UNP A5IKD4
C	20	SER	-	expression tag	UNP A5IKD4

- Molecule 2 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	O		0	0
			7	4	3			
2	B	1	Total	C	O		0	0
			7	4	3			
2	B	1	Total	C	O		0	0
			7	4	3			
2	C	1	Total	C	O		0	0
			7	4	3			

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Ca	0	0
			1	1		
3	B	1	Total	Ca	0	0
			1	1		
3	C	1	Total	Ca	0	0
			1	1		

- Molecule 4 is water.

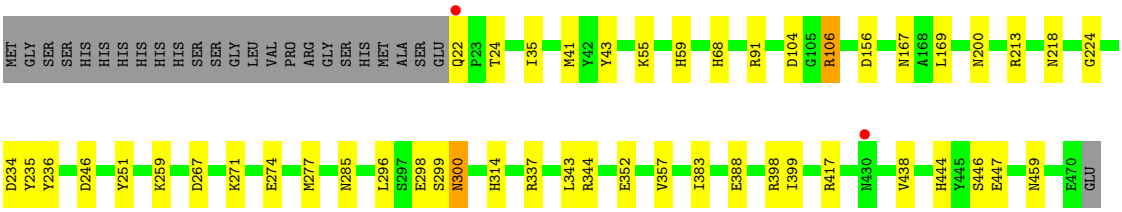
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	355	355	355	0	0
			Total	O		
4	B	176	176	176	0	0
			Total	O		
4	C	200	200	200	0	0

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.

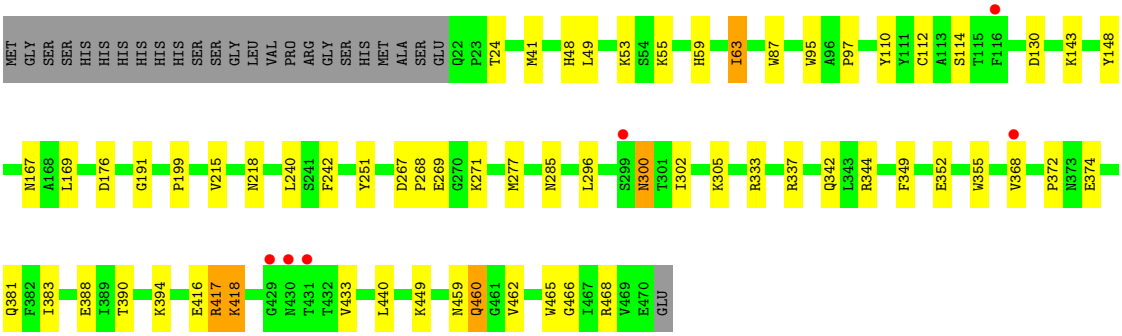
• Molecule 1: Glycoside hydrolase, family 43

Chain A:



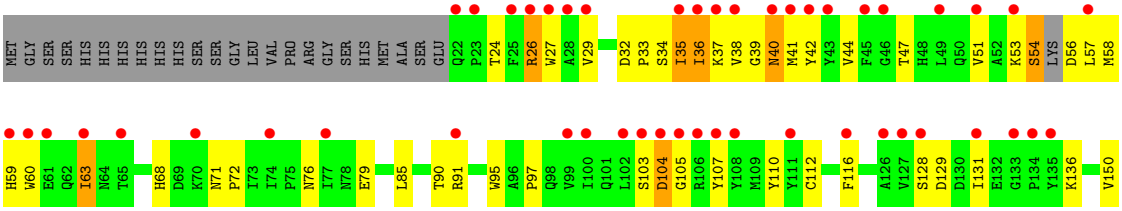
• Molecule 1: Glycoside hydrolase, family 43

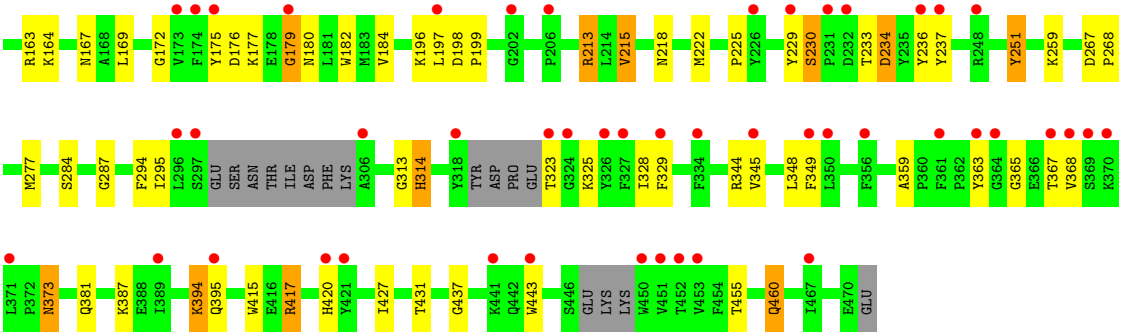
Chain B:



• Molecule 1: Glycoside hydrolase, family 43

Chain C:




R163 K164 N167 A168 L169 G172 V173 F174 Y175 D176 K177 E178 G179 N180 L181 W182 M183 V184 K196 L197 D198 P199 G202 P206 R213 L214 V215 N218 M222 P225 Y226 Y229 S230 P231 D232 T233 D234 Y235 Y236 Y237 R248 Y251 K259 D267 P268
M277 S284 G287 F294 I295 L296 S297 GLU SER ASN THR ILE ASP PHE LYS A306 G313 H314 Y318 TYR ASP PRO GLU T323 G324 K325 Y326 F327 I328 F329 F334 R344 V345 L348 F349 L350 F356 A359 P360 F361 P362 Y363 G364 G365 E366 T367 V368 S369 K370
L371 P372 N373 Q381 K387 E388 I389 K394 Q395 W415 E416 R417 H420 Y421 I427 T431 G437 K441 Q442 W443 S446 GLU LYS LYS W450 V451 T452 V453 F454 T455 Q460 I467 E470 GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	41.67Å 86.91Å 194.85Å 90.00° 90.15° 90.00°	Depositor
Resolution (Å)	40.77 – 1.75 40.77 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.1 (40.77-1.75) 99.1 (40.77-1.75)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.49 (at 1.75Å)	Xtriage
Refinement program	REFMAC 5.5.0102	Depositor
R, R_{free}	0.160 , 0.197 0.168 , 0.199	Depositor DCC
R_{free} test set	7010 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	20.1	Xtriage
Anisotropy	0.534	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 33.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.45$, $< L^2 > = 0.27$	Xtriage
Estimated twinning fraction	0.398 for h,-k,-l	Xtriage
Reported twinning fraction	0.612 for H, K, L 0.388 for -h,-k,l	Depositor
Outliers	0 of 139453 reflections	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	11493	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.27% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: PEG, CA

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	1.13	3/3766 (0.1%)	1.14	11/5117 (0.2%)
1	B	0.94	1/3780 (0.0%)	1.03	5/5134 (0.1%)
1	C	0.86	1/3598 (0.0%)	0.98	8/4885 (0.2%)
All	All	0.99	5/11144 (0.0%)	1.05	24/15136 (0.2%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	182	TRP	CB-CG	-7.43	1.36	1.50
1	A	344	ARG	CZ-NH2	5.92	1.40	1.33
1	A	224	GLY	N-CA	-5.74	1.37	1.46
1	B	465	TRP	CG-CD1	-5.37	1.29	1.36
1	A	344	ARG	NE-CZ	-5.15	1.26	1.33

The worst 5 of 24 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	213	ARG	NE-CZ-NH1	10.61	125.61	120.30
1	A	106	ARG	NE-CZ-NH1	9.51	125.06	120.30
1	A	106	ARG	NE-CZ-NH2	-9.16	115.72	120.30
1	C	213	ARG	NE-CZ-NH2	-8.22	116.19	120.30
1	C	35	ILE	CB-CA-C	-7.19	97.22	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	3623	0	3479	30	0
1	B	3632	0	3501	46	0
1	C	3476	0	3335	123	2
2	A	7	0	10	0	0
2	B	14	0	20	3	0
2	C	7	0	10	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
4	A	355	0	0	12	2
4	B	176	0	0	10	0
4	C	200	0	0	46	0
All	All	11493	0	10355	200	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 9.

The worst 5 of 200 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:C:58:MET:HB3	4:C:779:HOH:O	1.29	1.25
1:B:459:ASN:HB3	4:B:707:HOH:O	1.39	1.23
1:C:136:LYS:HB3	4:C:621:HOH:O	1.51	1.08
1:C:38:VAL:HG21	1:C:107:TYR:OH	1.55	1.05
1:C:175:TYR:CD1	4:C:789:HOH:O	2.12	1.02

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:213:ARG:NH2	4:A:725:HOH:O[2_555]	1.23	0.97
1:C:213:ARG:CZ	4:A:725:HOH:O[2_555]	1.92	0.28

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	453/474 (96%)	436 (96%)	17 (4%)	0	100	100
1	B	454/474 (96%)	441 (97%)	13 (3%)	0	100	100
1	C	426/474 (90%)	407 (96%)	16 (4%)	3 (1%)	22	8
All	All	1333/1422 (94%)	1284 (96%)	46 (4%)	3 (0%)	47	29

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	40	ASN
1	C	314	HIS
1	C	179	GLY

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	391/406 (96%)	382 (98%)	9 (2%)	50	28
1	B	392/406 (97%)	381 (97%)	11 (3%)	43	20
1	C	372/406 (92%)	355 (95%)	17 (5%)	27	8
All	All	1155/1218 (95%)	1118 (97%)	37 (3%)	39	16

5 of 37 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	164	LYS
1	C	431	THR
1	C	167	ASN
1	C	251	TYR
1	B	251	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	167	ASN
1	C	347	GLN
1	B	314	HIS
1	C	430	ASN
1	C	167	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 7 ligands modelled in this entry, 3 are monoatomic - leaving 4 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	PEG	B	502	-	6,6,6	0.61	0	5,5,5	0.73	0
2	PEG	C	501	-	6,6,6	0.47	0	5,5,5	1.02	0
2	PEG	B	501	-	6,6,6	0.43	0	5,5,5	1.47	1 (20%)
2	PEG	A	501	-	6,6,6	0.57	0	5,5,5	0.82	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	PEG	B	502	-	-	3/4/4/4	-
2	PEG	C	501	-	-	2/4/4/4	-
2	PEG	B	501	-	-	1/4/4/4	-
2	PEG	A	501	-	-	0/4/4/4	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	501	PEG	O2-C2-C1	2.42	120.71	110.07

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	502	PEG	O1-C1-C2-O2
2	C	501	PEG	C1-C2-O2-C3
2	B	502	PEG	C1-C2-O2-C3
2	B	501	PEG	C1-C2-O2-C3
2	C	501	PEG	C4-C3-O2-C2

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	502	PEG	2	0
2	B	501	PEG	1	0

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	449/474 (94%)	-0.02	2 (0%)	92	94	10, 16, 29, 54	2 (0%)
1	B	449/474 (94%)	0.11	6 (1%)	77	83	17, 25, 38, 54	1 (0%)
1	C	433/474 (91%)	1.20	95 (21%)	0	0	22, 34, 51, 63	5 (1%)
All	All	1331/1422 (93%)	0.42	103 (7%)	13	18	10, 25, 46, 63	8 (0%)

The worst 5 of 103 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	133	GLY	7.1
1	C	443	TRP	6.2
1	C	175	TYR	5.9
1	C	179	GLY	5.6
1	C	127	VAL	5.6

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 monosaccharides in this entry.

6.4 Ligands (i)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
2	PEG	C	501	7/7	0.84	0.11	40,41,42,42	0
2	PEG	B	501	7/7	0.85	0.14	33,35,43,48	0
2	PEG	B	502	7/7	0.87	0.25	31,36,40,46	0
2	PEG	A	501	7/7	0.93	0.13	28,29,34,39	0
3	CA	B	503	1/1	0.94	0.21	39,39,39,39	0
3	CA	A	502	1/1	0.95	0.13	39,39,39,39	0
3	CA	C	502	1/1	0.96	0.07	41,41,41,41	0

6.5 Other polymers (i)

There are no such residues in this entry.