Full wwPDB X-ray Structure Validation Report ⓘ

Jan 24, 2021 – 02:34 PM EST

PDB ID	:	3D3A
Title	:	Crystal structure of a beta-galactosidase from Bacteroides thetaiotaomicron
Authors	:	Palani, K.; Kumaran, D.; Burley, S.K.; Swaminathan, S.; New York SGX Research Center for Structural Genomics (NYSGXRC)
Deposited on	:	2008-05-09
Resolution	:	2.15 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ symbol.

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

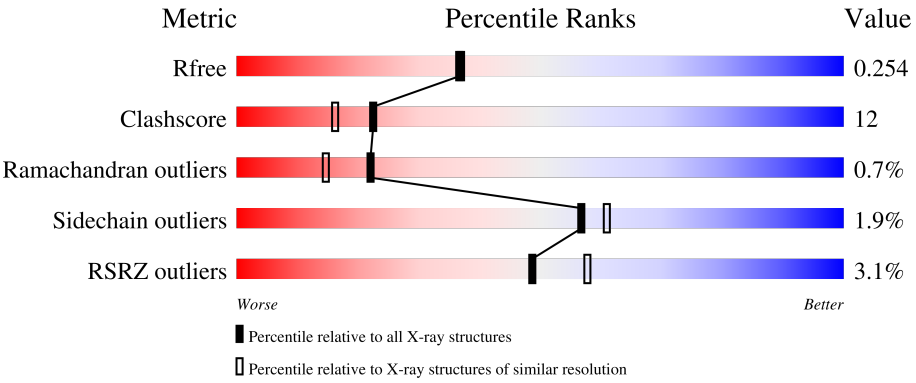
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.16
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.16

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 2.15 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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)
RSRZ outliers	127900	1456 (2.16-2.16)

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	612	3% 77% 20% ..

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 5257 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 Beta-galactosidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	604	4873	3126	819	899	8	21	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	21	MSE	-	expression tag	UNP Q8AB22
A	22	SER	-	expression tag	UNP Q8AB22
A	23	LEU	-	expression tag	UNP Q8AB22
A	626	GLY	-	expression tag	UNP Q8AB22
A	627	HIS	-	expression tag	UNP Q8AB22
A	628	HIS	-	expression tag	UNP Q8AB22
A	629	HIS	-	expression tag	UNP Q8AB22
A	630	HIS	-	expression tag	UNP Q8AB22
A	631	HIS	-	expression tag	UNP Q8AB22
A	632	HIS	-	expression tag	UNP Q8AB22

- Molecule 2 is water.

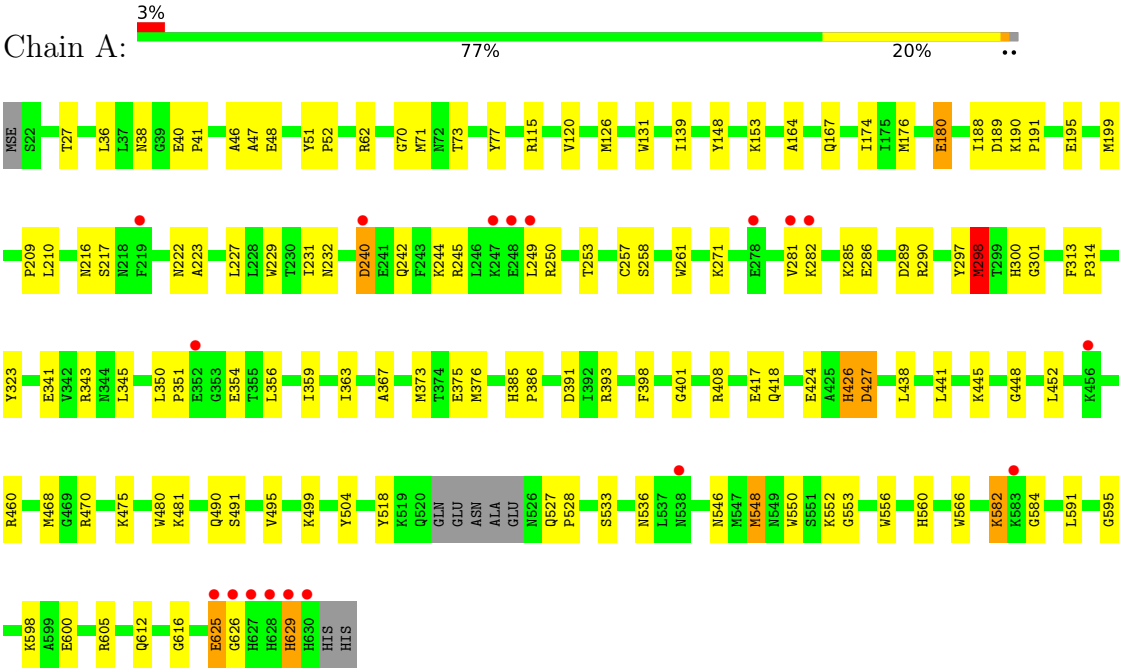
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	384	384	384	0	0

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: Beta-galactosidase

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	117.02Å 117.02Å 59.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.46 – 2.15 48.42 – 2.13	Depositor EDS
% Data completeness (in resolution range)	90.7 (32.46-2.15) 89.5 (48.42-2.13)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.54 (at 2.12Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.214 , 0.254 0.214 , 0.254	Depositor DCC
R_{free} test set	2005 reflections (4.74%)	wwPDB-VP
Wilson B-factor (Å^2)	18.7	Xtriage
Anisotropy	0.573	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 50.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.032 for h,-k,-l	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	5257	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.75% 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	0.35	0/4986	0.62	0/6714

There are no bond length outliers.

There are no bond angle outliers.

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	4873	0	4733	111	0
2	A	384	0	0	7	0
All	All	5257	0	4733	111	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 12.

All (111) 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:298:MSE:HE2	1:A:298:MSE:N	1.64	1.12
1:A:48:GLU:HG2	1:A:298:MSE:HE1	1.23	1.10
1:A:126:MSE:HE2	1:A:470:ARG:HD2	1.24	1.08
1:A:48:GLU:CG	1:A:298:MSE:HE1	2.01	0.89

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:126:MSE:HE1	1:A:131:TRP:HA	1.54	0.87
1:A:297:TYR:CG	1:A:298:MSE:HE3	2.11	0.85
1:A:126:MSE:HE1	1:A:131:TRP:CA	2.10	0.81
1:A:375:GLU:HB3	1:A:518:TYR:HB3	1.63	0.81
1:A:297:TYR:CD2	1:A:298:MSE:HE3	2.18	0.78
1:A:298:MSE:HE2	1:A:298:MSE:H	1.45	0.76
1:A:351:PRO:HB2	1:A:354:GLU:HG3	1.67	0.76
1:A:126:MSE:HE2	1:A:470:ARG:CD	2.10	0.75
1:A:245:ARG:NE	1:A:249:LEU:HD21	2.02	0.75
1:A:176:MSE:HG2	1:A:209:PRO:HG2	1.67	0.74
1:A:626:GLY:HA2	1:A:629:HIS:NE2	2.02	0.74
1:A:298:MSE:SE	2:A:990:HOH:O	2.56	0.72
1:A:418:GLN:HE21	1:A:491:SER:HB3	1.57	0.70
1:A:297:TYR:CD1	1:A:298:MSE:HE3	2.29	0.67
1:A:126:MSE:HE3	2:A:696:HOH:O	1.93	0.67
1:A:536:ASN:HA	1:A:584:GLY:O	1.96	0.66
1:A:261:TRP:HA	1:A:297:TYR:HB3	1.79	0.65
1:A:582:LYS:HG3	2:A:827:HOH:O	1.97	0.63
1:A:298:MSE:N	1:A:298:MSE:CE	2.53	0.63
1:A:285:LYS:HG2	1:A:289:ASP:OD2	1.99	0.62
1:A:470:ARG:HG2	1:A:481:LYS:HE2	1.82	0.62
1:A:297:TYR:O	1:A:298:MSE:HB2	1.98	0.62
1:A:475:LYS:NZ	2:A:983:HOH:O	2.31	0.62
1:A:27:THR:H	1:A:38:ASN:ND2	1.97	0.62
1:A:351:PRO:CB	1:A:354:GLU:HG3	2.31	0.61
1:A:216:ASN:HD22	1:A:242:GLN:CD	2.05	0.60
1:A:351:PRO:HB2	1:A:354:GLU:CG	2.32	0.59
1:A:598:LYS:HG2	1:A:600:GLU:HG2	1.84	0.59
1:A:351:PRO:HD2	1:A:354:GLU:HG3	1.86	0.58
1:A:350:LEU:HD11	1:A:356:LEU:HG	1.86	0.58
1:A:281:VAL:HG21	1:A:341:GLU:HG2	1.86	0.58
1:A:189:ASP:H	1:A:222:ASN:ND2	2.02	0.57
1:A:48:GLU:HB3	1:A:298:MSE:HE2	1.87	0.57
1:A:189:ASP:H	1:A:222:ASN:HD21	1.52	0.57
1:A:281:VAL:HG13	1:A:345:LEU:HD22	1.87	0.56
1:A:343:ARG:HH12	1:A:356:LEU:HB3	1.72	0.54
1:A:373:MSE:HA	1:A:533:SER:HB2	1.89	0.54
1:A:313:PHE:CD1	1:A:314:PRO:HA	2.43	0.54
1:A:441:LEU:N	1:A:441:LEU:HD22	2.23	0.54
1:A:223:ALA:HB1	1:A:229:TRP:CE2	2.43	0.53
1:A:297:TYR:CE2	1:A:298:MSE:HE3	2.43	0.53

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:536:ASN:ND2	1:A:584:GLY:HA2	2.23	0.52
1:A:553:GLY:HA3	1:A:591:LEU:O	2.09	0.52
1:A:351:PRO:CD	1:A:354:GLU:HG3	2.40	0.52
1:A:48:GLU:HB3	1:A:298:MSE:CE	2.40	0.52
1:A:391:ASP:OD1	1:A:499:LYS:HD3	2.09	0.52
1:A:297:TYR:CD1	1:A:298:MSE:CE	2.93	0.51
1:A:445:LYS:HE3	1:A:616:GLY:O	2.10	0.51
1:A:232:ASN:OD1	1:A:258:SER:HB3	2.11	0.51
1:A:426:HIS:O	1:A:427:ASP:HB2	2.11	0.50
1:A:552:LYS:HE2	1:A:566:TRP:CD2	2.47	0.50
1:A:195:GLU:O	1:A:199:MSE:HG3	2.12	0.50
1:A:282:LYS:O	1:A:286:GLU:HG3	2.12	0.50
1:A:51:TYR:CG	1:A:52:PRO:HD3	2.47	0.50
1:A:546:ASN:OD1	1:A:548:MSE:HB2	2.12	0.49
1:A:164:ALA:O	1:A:167:GLN:HG3	2.12	0.49
1:A:343:ARG:NE	1:A:359:ILE:HD11	2.28	0.49
1:A:426:HIS:HD2	1:A:480:TRP:O	1.95	0.49
1:A:350:LEU:HB3	1:A:354:GLU:HB2	1.94	0.49
1:A:386:PRO:HB3	1:A:504:TYR:CE1	2.48	0.48
1:A:298:MSE:HE2	1:A:298:MSE:CA	2.42	0.48
1:A:297:TYR:C	1:A:298:MSE:HE2	2.30	0.48
1:A:582:LYS:HB3	1:A:582:LYS:NZ	2.28	0.48
1:A:240:ASP:O	1:A:244:LYS:HB2	2.14	0.48
1:A:188:ILE:HA	1:A:222:ASN:HD21	1.80	0.47
1:A:70:GLY:HA2	1:A:356:LEU:HD21	1.97	0.47
1:A:216:ASN:ND2	1:A:242:GLN:NE2	2.62	0.47
1:A:271:LYS:HA	1:A:612:GLN:NE2	2.29	0.46
1:A:490:GLN:HB2	1:A:495:VAL:HG22	1.97	0.46
1:A:625:GLU:OE1	1:A:626:GLY:N	2.48	0.46
1:A:120:VAL:HG12	2:A:959:HOH:O	2.14	0.46
1:A:550:TRP:CD1	1:A:550:TRP:N	2.84	0.46
1:A:386:PRO:HB3	1:A:504:TYR:CZ	2.52	0.45
1:A:261:TRP:HD1	1:A:297:TYR:CG	2.34	0.45
1:A:595:GLY:HA2	2:A:716:HOH:O	2.15	0.45
1:A:46:ALA:HA	1:A:73:THR:O	2.15	0.45
1:A:286:GLU:O	1:A:290:ARG:HG3	2.16	0.45
1:A:527:GLN:HB2	1:A:528:PRO:CD	2.48	0.44
1:A:62:ARG:HE	1:A:300:HIS:CE1	2.35	0.44
1:A:139:ILE:HD11	1:A:148:TYR:HB2	1.99	0.44
1:A:297:TYR:CE1	1:A:298:MSE:HE3	2.53	0.44
1:A:408:ARG:NH2	1:A:460:ARG:HE	2.16	0.44

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:210:LEU:HB2	1:A:227:LEU:HD22	2.00	0.44
1:A:47:ALA:HB2	1:A:71:MSE:HG3	1.98	0.44
1:A:367:ALA:HB2	1:A:605:ARG:HG3	1.99	0.43
1:A:250:ARG:O	1:A:253:THR:HG22	2.18	0.43
1:A:363:ILE:N	1:A:363:ILE:HD12	2.33	0.43
1:A:301:GLY:HA3	1:A:323:TYR:O	2.19	0.43
1:A:36:LEU:HA	1:A:40:GLU:O	2.19	0.42
1:A:424:GLU:OE1	1:A:426:HIS:HE1	2.02	0.42
1:A:556:TRP:HA	1:A:560:HIS:O	2.19	0.42
1:A:216:ASN:OD1	1:A:217:SER:N	2.53	0.42
1:A:245:ARG:O	1:A:249:LEU:HD23	2.19	0.42
1:A:468:MSE:SE	1:A:552:LYS:HZ1	2.53	0.41
1:A:231:ILE:O	1:A:257:CYS:HA	2.19	0.41
1:A:393:ARG:HB2	1:A:398:PHE:CE2	2.54	0.41
1:A:180:GLU:HA	2:A:647:HOH:O	2.20	0.41
1:A:36:LEU:HD23	1:A:41:PRO:HA	2.02	0.41
1:A:131:TRP:CD2	1:A:401:GLY:HA3	2.55	0.41
1:A:167:GLN:HA	1:A:174:ILE:HB	2.01	0.41
1:A:216:ASN:ND2	1:A:242:GLN:CD	2.73	0.41
1:A:417:GLU:HG3	1:A:452:LEU:O	2.21	0.41
1:A:77:TYR:HA	1:A:115:ARG:HB2	2.02	0.41
1:A:51:TYR:N	1:A:52:PRO:CD	2.85	0.41
1:A:153:LYS:HB2	1:A:199:MSE:HE2	2.01	0.40
1:A:190:LYS:N	1:A:191:PRO:CD	2.84	0.40
1:A:48:GLU:CG	1:A:298:MSE:CE	2.88	0.40

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	600/612 (98%)	568 (95%)	28 (5%)	4 (1%)	22 15

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	427	ASP
1	A	629	HIS
1	A	298	MSE
1	A	448	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	518/503 (103%)	508 (98%)	10 (2%)	57 61

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	180	GLU
1	A	240	ASP
1	A	298	MSE
1	A	376	MSE
1	A	385	HIS
1	A	426	HIS
1	A	438	LEU
1	A	548	MSE
1	A	582	LYS
1	A	625	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (14) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	38	ASN
1	A	95	GLN
1	A	216	ASN
1	A	221	ASN
1	A	222	ASN
1	A	300	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	400	GLN
1	A	418	GLN
1	A	426	HIS
1	A	472	ASN
1	A	536	ASN
1	A	606	GLN
1	A	612	GLN
1	A	620	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)

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^{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	583/612 (95%)	0.22	18 (3%) 49 58	7, 21, 46, 102	0

All (18) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	630	HIS	14.5
1	A	628	HIS	12.9
1	A	626	GLY	11.1
1	A	627	HIS	10.3
1	A	629	HIS	9.2
1	A	583	LYS	3.1
1	A	625	GLU	2.8
1	A	538	ASN	2.7
1	A	248	GLU	2.3
1	A	352	GLU	2.3
1	A	249	LEU	2.2
1	A	219	PHE	2.2
1	A	281	VAL	2.2
1	A	247	LYS	2.1
1	A	456	LYS	2.1
1	A	240	ASP	2.1
1	A	282	LYS	2.1
1	A	278	GLU	2.0

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 (i)

There are no monosaccharides 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.