



# wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 11, 2021 – 02:17 PM JST

PDB ID : 6KHT  
Title : Chimeric beta-glucosidase Cel1b-H13  
Authors : Niu, K.L.  
Deposited on : 2019-07-16  
Resolution : 3.31 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](mailto: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
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

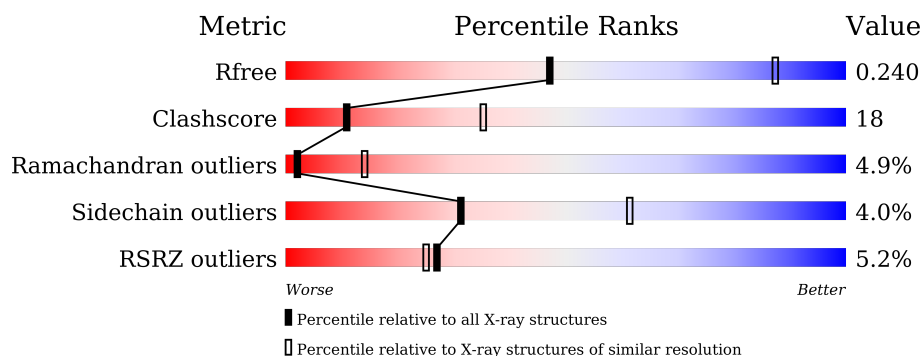
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.31 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	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	513	<div> <div>3%</div> <div> <div></div> <div>59%</div> <div>29%</div> <div>.. 7%</div> </div> </div>
1	B	513	<div> <div>4%</div> <div> <div></div> <div>62%</div> <div>27%</div> <div>.. 7%</div> </div> </div>
1	C	513	<div> <div>7%</div> <div> <div></div> <div>63%</div> <div>26%</div> <div>.. 7%</div> </div> </div>

## 2 Entry composition

There is only 1 type of molecule in this entry. The entry contains 11529 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 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	477	Total	C	N	O	S	0	0	0
			3843	2451	661	718	13			
1	B	477	Total	C	N	O	S	0	0	0
			3843	2451	661	718	13			
1	C	477	Total	C	N	O	S	0	0	0
			3843	2451	661	718	13			

There are 114 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-13	HIS	-	expression tag	UNP G0RIF5
A	-12	HIS	-	expression tag	UNP G0RIF5
A	-11	HIS	-	expression tag	UNP G0RIF5
A	-10	HIS	-	expression tag	UNP G0RIF5
A	-9	HIS	-	expression tag	UNP G0RIF5
A	-8	HIS	-	expression tag	UNP G0RIF5
A	-7	SER	-	expression tag	UNP G0RIF5
A	-6	SER	-	expression tag	UNP G0RIF5
A	-5	GLY	-	expression tag	UNP G0RIF5
A	-4	LEU	-	expression tag	UNP G0RIF5
A	-3	VAL	-	expression tag	UNP G0RIF5
A	-2	PRO	-	expression tag	UNP G0RIF5
A	-1	ARG	-	expression tag	UNP G0RIF5
A	0	GLY	-	expression tag	UNP G0RIF5
A	1	SER	-	expression tag	UNP G0RIF5
A	200	ASP	ASN	conflict	UNP G0RIF5
A	319	TYR	PHE	conflict	UNP G0RIF5
A	479	ILE	-	expression tag	UNP G0RIF5
A	480	PRO	-	expression tag	UNP G0RIF5
A	481	GLU	-	expression tag	UNP G0RIF5
A	482	GLU	-	expression tag	UNP G0RIF5
A	483	LEU	-	expression tag	UNP G0RIF5
A	484	ALA	-	expression tag	UNP G0RIF5

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	485	HIS	-	expression tag	UNP G0RIF5
A	486	LEU	-	expression tag	UNP G0RIF5
A	487	ALA	-	expression tag	UNP G0RIF5
A	488	ASP	-	expression tag	UNP G0RIF5
A	489	LEU	-	expression tag	UNP G0RIF5
A	490	LYS	-	expression tag	UNP G0RIF5
A	491	PHE	-	expression tag	UNP G0RIF5
A	492	LEU	-	expression tag	UNP G0RIF5
A	493	GLU	-	expression tag	UNP G0RIF5
A	494	HIS	-	expression tag	UNP G0RIF5
A	495	HIS	-	expression tag	UNP G0RIF5
A	496	HIS	-	expression tag	UNP G0RIF5
A	497	HIS	-	expression tag	UNP G0RIF5
A	498	HIS	-	expression tag	UNP G0RIF5
A	499	HIS	-	expression tag	UNP G0RIF5
B	-13	HIS	-	expression tag	UNP G0RIF5
B	-12	HIS	-	expression tag	UNP G0RIF5
B	-11	HIS	-	expression tag	UNP G0RIF5
B	-10	HIS	-	expression tag	UNP G0RIF5
B	-9	HIS	-	expression tag	UNP G0RIF5
B	-8	HIS	-	expression tag	UNP G0RIF5
B	-7	SER	-	expression tag	UNP G0RIF5
B	-6	SER	-	expression tag	UNP G0RIF5
B	-5	GLY	-	expression tag	UNP G0RIF5
B	-4	LEU	-	expression tag	UNP G0RIF5
B	-3	VAL	-	expression tag	UNP G0RIF5
B	-2	PRO	-	expression tag	UNP G0RIF5
B	-1	ARG	-	expression tag	UNP G0RIF5
B	0	GLY	-	expression tag	UNP G0RIF5
B	1	SER	-	expression tag	UNP G0RIF5
B	200	ASP	ASN	conflict	UNP G0RIF5
B	319	TYR	PHE	conflict	UNP G0RIF5
B	479	ILE	-	expression tag	UNP G0RIF5
B	480	PRO	-	expression tag	UNP G0RIF5
B	481	GLU	-	expression tag	UNP G0RIF5
B	482	GLU	-	expression tag	UNP G0RIF5
B	483	LEU	-	expression tag	UNP G0RIF5
B	484	ALA	-	expression tag	UNP G0RIF5
B	485	HIS	-	expression tag	UNP G0RIF5
B	486	LEU	-	expression tag	UNP G0RIF5
B	487	ALA	-	expression tag	UNP G0RIF5
B	488	ASP	-	expression tag	UNP G0RIF5

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	489	LEU	-	expression tag	UNP G0RIF5
B	490	LYS	-	expression tag	UNP G0RIF5
B	491	PHE	-	expression tag	UNP G0RIF5
B	492	LEU	-	expression tag	UNP G0RIF5
B	493	GLU	-	expression tag	UNP G0RIF5
B	494	HIS	-	expression tag	UNP G0RIF5
B	495	HIS	-	expression tag	UNP G0RIF5
B	496	HIS	-	expression tag	UNP G0RIF5
B	497	HIS	-	expression tag	UNP G0RIF5
B	498	HIS	-	expression tag	UNP G0RIF5
B	499	HIS	-	expression tag	UNP G0RIF5
C	-13	HIS	-	expression tag	UNP G0RIF5
C	-12	HIS	-	expression tag	UNP G0RIF5
C	-11	HIS	-	expression tag	UNP G0RIF5
C	-10	HIS	-	expression tag	UNP G0RIF5
C	-9	HIS	-	expression tag	UNP G0RIF5
C	-8	HIS	-	expression tag	UNP G0RIF5
C	-7	SER	-	expression tag	UNP G0RIF5
C	-6	SER	-	expression tag	UNP G0RIF5
C	-5	GLY	-	expression tag	UNP G0RIF5
C	-4	LEU	-	expression tag	UNP G0RIF5
C	-3	VAL	-	expression tag	UNP G0RIF5
C	-2	PRO	-	expression tag	UNP G0RIF5
C	-1	ARG	-	expression tag	UNP G0RIF5
C	0	GLY	-	expression tag	UNP G0RIF5
C	1	SER	-	expression tag	UNP G0RIF5
C	200	ASP	ASN	conflict	UNP G0RIF5
C	319	TYR	PHE	conflict	UNP G0RIF5
C	479	ILE	-	expression tag	UNP G0RIF5
C	480	PRO	-	expression tag	UNP G0RIF5
C	481	GLU	-	expression tag	UNP G0RIF5
C	482	GLU	-	expression tag	UNP G0RIF5
C	483	LEU	-	expression tag	UNP G0RIF5
C	484	ALA	-	expression tag	UNP G0RIF5
C	485	HIS	-	expression tag	UNP G0RIF5
C	486	LEU	-	expression tag	UNP G0RIF5
C	487	ALA	-	expression tag	UNP G0RIF5
C	488	ASP	-	expression tag	UNP G0RIF5
C	489	LEU	-	expression tag	UNP G0RIF5
C	490	LYS	-	expression tag	UNP G0RIF5
C	491	PHE	-	expression tag	UNP G0RIF5
C	492	LEU	-	expression tag	UNP G0RIF5

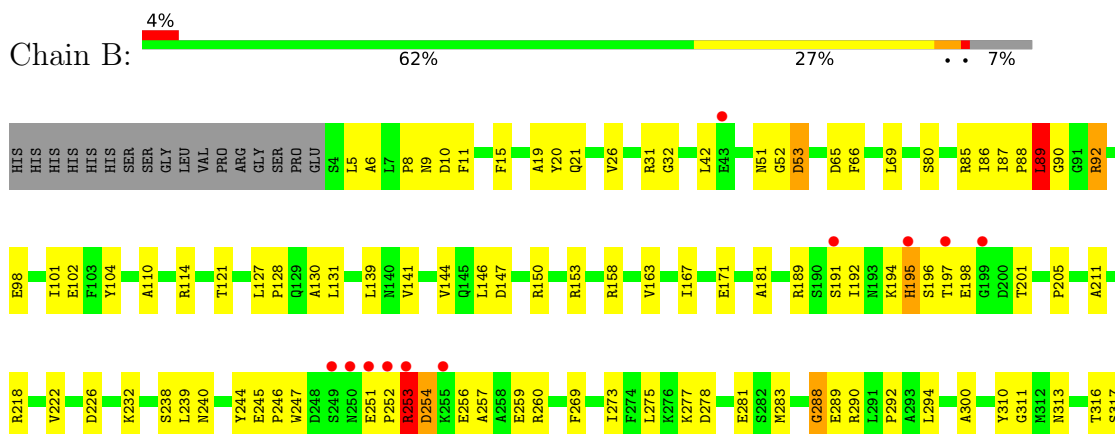
*Continued on next page...*

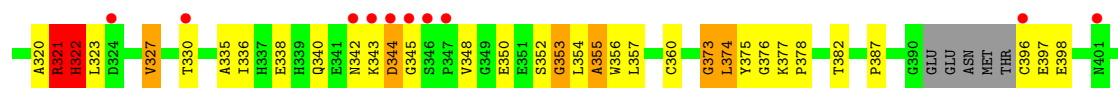
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	493	GLU	-	expression tag	UNP G0RIF5
C	494	HIS	-	expression tag	UNP G0RIF5
C	495	HIS	-	expression tag	UNP G0RIF5
C	496	HIS	-	expression tag	UNP G0RIF5
C	497	HIS	-	expression tag	UNP G0RIF5
C	498	HIS	-	expression tag	UNP G0RIF5
C	499	HIS	-	expression tag	UNP G0RIF5

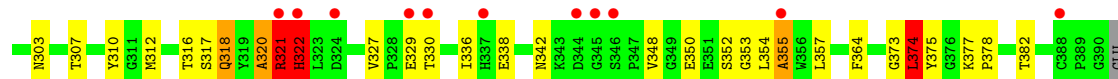
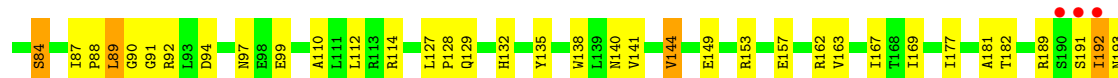
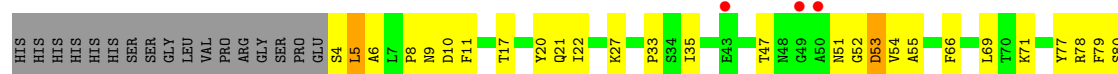


- Molecule 1: Glycoside hydrolase family 1





● Molecule 1: Glycoside hydrolase family 1





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	108.58Å 182.39Å 210.51Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	38.23 – 3.31 38.23 – 3.31	Depositor EDS
% Data completeness (in resolution range)	99.4 (38.23-3.31) 99.5 (38.23-3.31)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.17 (at 3.32Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496, PHENIX	Depositor
R, $R_{free}$	0.174 , 0.237 0.181 , 0.240	Depositor DCC
$R_{free}$ test set	1590 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	41.6	Xtriage
Anisotropy	0.024	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 43.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.012 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.022 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	11529	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.75% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

### 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	0.55	0/3960	0.75	5/5380 (0.1%)
1	B	0.53	0/3960	0.76	8/5380 (0.1%)
1	C	0.52	0/3960	0.75	4/5380 (0.1%)
All	All	0.53	0/11880	0.75	17/16140 (0.1%)

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	4
1	B	0	1
1	C	0	4
All	All	0	9

There are no bond length outliers.

The worst 5 of 17 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	89	LEU	CA-CB-CG	12.25	143.47	115.30
1	C	322	HIS	N-CA-C	7.91	132.37	111.00
1	B	485	HIS	N-CA-C	7.69	131.78	111.00
1	C	374	LEU	CA-CB-CG	7.67	132.95	115.30
1	B	484	ALA	CB-CA-C	7.56	121.44	110.10

There are no chirality outliers.

5 of 9 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	373	GLY	Peptide
1	A	374	LEU	Peptide

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	A	5	LEU	Peptide
1	A	88	PRO	Peptide
1	B	373	GLY	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	3843	0	3635	136	0
1	B	3843	0	3635	134	0
1	C	3843	0	3635	141	0
All	All	11529	0	10905	409	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 18.

The worst 5 of 409 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:479:ILE:CG2	1:C:480:PRO:HA	1.46	1.41
1:A:479:ILE:CG2	1:A:480:PRO:HA	1.57	1.33
1:C:478:GLN:O	1:C:479:ILE:HD13	1.17	1.30
1:C:479:ILE:CB	1:C:480:PRO:HA	1.57	1.30
1:C:478:GLN:O	1:C:479:ILE:CD1	1.80	1.30

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	473/513 (92%)	407 (86%)	40 (8%)	26 (6%)	2	11
1	B	473/513 (92%)	399 (84%)	53 (11%)	21 (4%)	2	16
1	C	473/513 (92%)	402 (85%)	49 (10%)	22 (5%)	2	14
All	All	1419/1539 (92%)	1208 (85%)	142 (10%)	69 (5%)	2	14

5 of 69 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	6	ALA
1	A	8	PRO
1	A	11	PHE
1	A	89	LEU
1	A	327	VAL

### 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	396/429 (92%)	385 (97%)	11 (3%)	43	70
1	B	396/429 (92%)	377 (95%)	19 (5%)	25	56
1	C	396/429 (92%)	378 (96%)	18 (4%)	27	58
All	All	1188/1287 (92%)	1140 (96%)	48 (4%)	31	61

5 of 48 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	344	ASP
1	B	478	GLN
1	C	459	LEU
1	B	360	CYS
1	B	458	THR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such

sidechains are listed below:

Mol	Chain	Res	Type
1	B	313	ASN
1	C	485	HIS
1	C	140	ASN
1	A	265	HIS
1	B	322	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 [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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	477/513 (92%)	-0.19	15 (3%) 49 48	11, 27, 68, 90	0
1	B	477/513 (92%)	-0.10	23 (4%) 30 28	11, 32, 75, 99	0
1	C	477/513 (92%)	0.02	36 (7%) 14 13	13, 32, 76, 97	0
All	All	1431/1539 (92%)	-0.09	74 (5%) 27 25	11, 30, 73, 99	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	197	THR	6.2
1	C	250	ASN	5.0
1	A	485	HIS	5.0
1	C	483	LEU	4.7
1	A	250	ASN	4.7

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.