



# wwPDB X-ray Structure Validation Summary Report

Jun 10, 2014 – 12:06 PM EDT

PDB ID : 4PYS  
Title : The crystal structure of beta-N-acetylhexosaminidase from *Bacteroides fragilis* NCTC 9343  
Authors : Tan, K.; Hatzos-Skintges, C.; Clancy, S.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)  
Deposited on : 2014-03-27  
Resolution : 1.82 Å (reported)

This is a wwPDB 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 <http://wwpdb.org/ValidationPDFNotes.html>

---

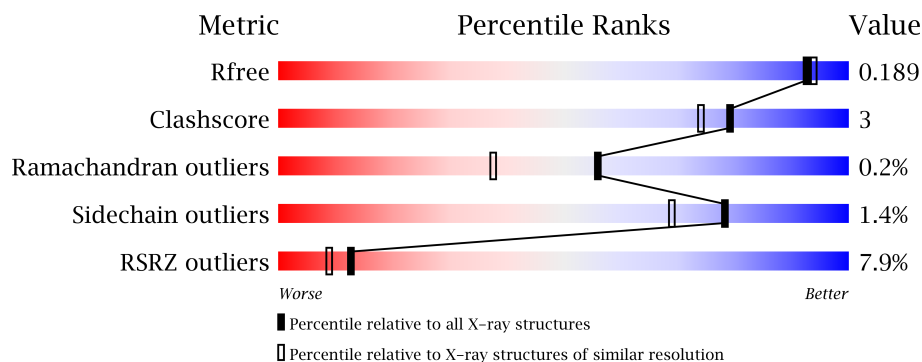
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.16 November 2013  
Xtriage (Phenix) : dev-1439  
EDS : stable23161  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable23161

# 1 Overall quality at a glance

The reported resolution of this entry is 1.82 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))
$R_{free}$	66092	4101 (1.84-1.80)
Clashscore	79885	5140 (1.84-1.80)
Ramachandran outliers	78287	5077 (1.84-1.80)
Sidechain outliers	78261	5077 (1.84-1.80)
RSRZ outliers	66119	4103 (1.84-1.80)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	506	
1	B	506	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	GOL	A	605	-	X
4	FMT	A	608	-	X
4	FMT	A	609	-	X
4	FMT	A	610	-	X
4	FMT	B	606	-	X
4	FMT	B	607	-	X

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Geometry	Electron density
4	FMT	B	608	-	X

## 2 Entry composition i

There are 5 unique types of molecules in this entry. The entry contains 8553 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

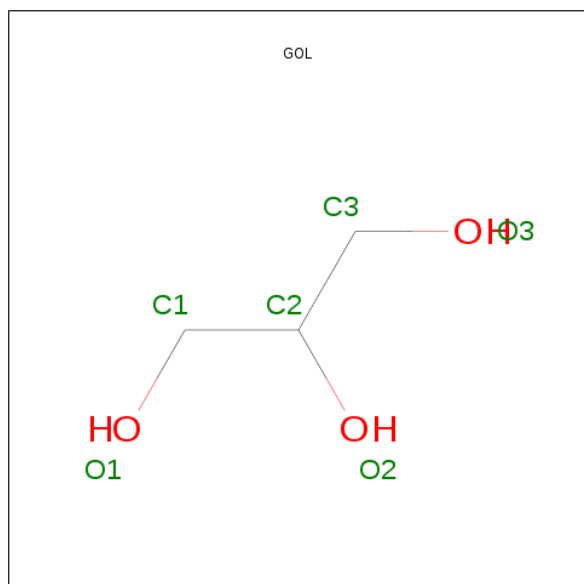
- Molecule 1 is a protein called beta-N-acetylhexosaminidase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	492	Total	C	N	O	S	Se	0	1	0
			3972	2564	675	713	8	12			
1	B	483	Total	C	N	O	S	Se	0	4	0
			3909	2517	674	698	8	12			

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

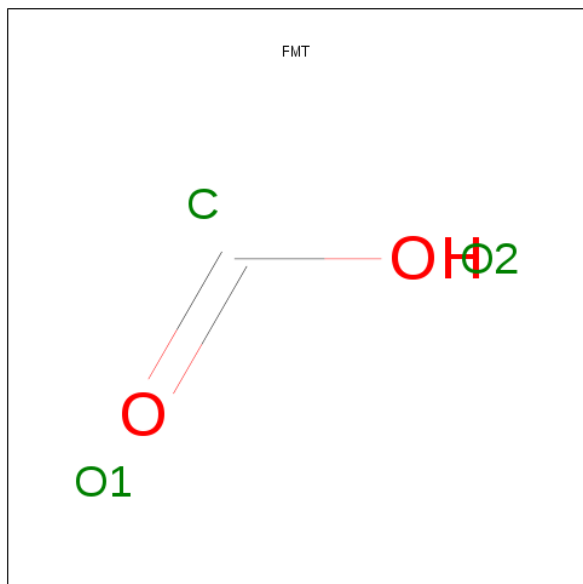
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Zn	0	0
			1	1		
2	A	1	Total	Zn	0	0
			1	1		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	A	1	Total	C	O	0	0
			6	3	3		
3	B	1	Total	C	O	0	0
			6	3	3		
3	B	1	Total	C	O	0	0
			6	3	3		

- Molecule 4 is FORMIC ACID (three-letter code: FMT) (formula:  $\text{CH}_2\text{O}_2$ ).



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

*Continued on next page...*

*Continued from previous page...*

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

- Molecule 5 is water.

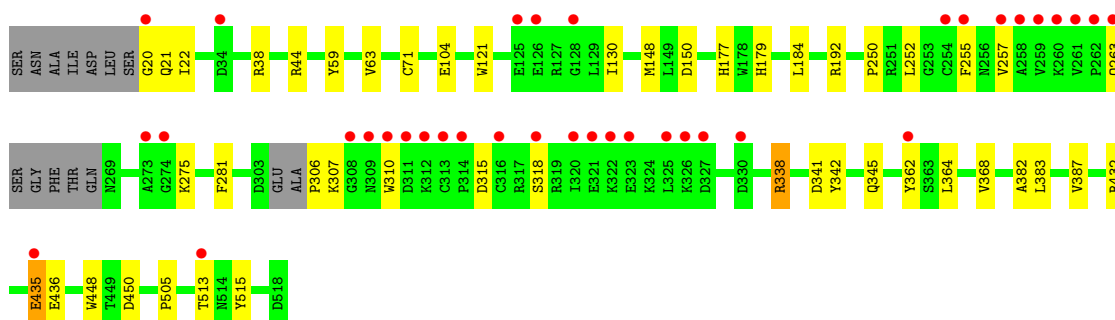
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	280	Total	O	0	0
			280	280		
5	B	324	Total	O	0	0
			324	324		

### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

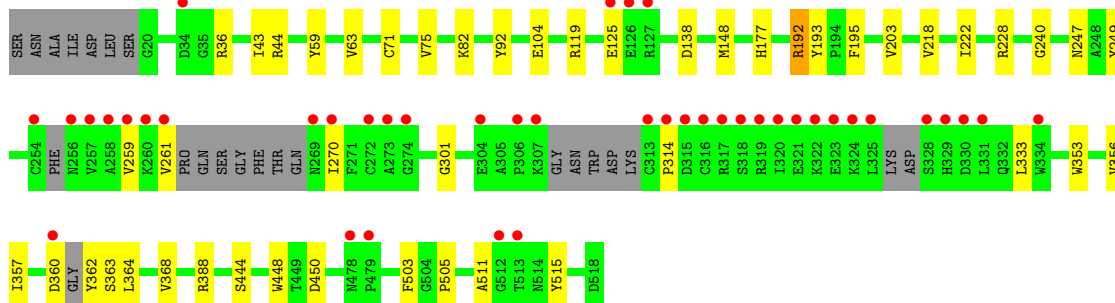
- Molecule 1: beta-N-acetylhexosaminidase

Chain A: 



- Molecule 1: beta-N-acetylhexosaminidase

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	96.82Å 96.82Å 315.97Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	34.03 – 1.82 37.97 – 1.82	Depositor EDS
% Data completeness (in resolution range)	99.6 (34.03-1.82) 99.6 (37.97-1.82)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.31 (at 1.82Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, $R_{free}$	0.165 , 0.190 0.164 , 0.189	Depositor DCC
$R_{free}$ test set	6739 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	26.8	Xtriage
Anisotropy	0.015	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.37 , 48.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 134269 reflections (0.001%)	Xtriage
$F_o, F_c$ correlation	0.97	EDS
Total number of atoms	8553	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	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 40.01 % 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 2.9169e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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

## 5 Model quality

### 5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, ZN, FMT

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	0/4078	0.53	0/5512
1	B	0.35	0/4016	0.54	2/5423 (0.0%)
All	All	0.35	0/8094	0.54	2/10935 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	119	ARG	NE-CZ-NH1	-6.35	117.12	120.30
1	B	119	ARG	NE-CZ-NH2	5.75	123.18	120.30

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	3972	0	3798	29	0
1	B	3909	0	3766	26	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	24	0	32	1	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	12	0	16	0	0
4	A	15	0	5	2	0
4	B	15	0	5	2	0
5	A	280	0	0	2	0
5	B	324	0	0	2	0
All	All	8553	0	7622	54	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 3.

The worst 5 of 54 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:184:LEU:HD21	1:A:263:GLN:HA	1.69	0.74
1:B:388[A]:ARG:NH2	5:B:1022:HOH:O	2.23	0.71
1:A:21:GLN:HG3	1:A:22:ILE:H	1.61	0.64
1:B:203:VAL:HB	1:B:261:VAL:HG21	1.82	0.60
1:A:275:LYS:NZ	1:A:315:ASP:OD2	2.37	0.57

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	487/506 (96%)	471 (97%)	15 (3%)	1 (0%)	56	38
1	B	475/506 (94%)	460 (97%)	14 (3%)	1 (0%)	56	38
All	All	962/1012 (95%)	931 (97%)	29 (3%)	2 (0%)	56	38

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	435	GLU
1	B	314	PRO

### 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	405/425 (95%)	400 (99%)	5 (1%)	82	74
1	B	403/425 (95%)	396 (98%)	7 (2%)	73	61
All	All	808/850 (95%)	796 (98%)	12 (2%)	78	66

5 of 12 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	75	VAL
1	B	192[A]	ARG
1	B	333	LEU
1	A	513	THR
1	B	228	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	478	ASN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 ⓘ

Of 18 ligands modelled in this entry, 2 are monoatomic - leaving 16 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	$\# Z  > 2$
3	GOL	A	602	-	5,5,5	0.31	0	5,5,5	0.31	0
3	GOL	A	603	-	5,5,5	0.37	0	5,5,5	0.64	0
3	GOL	A	604	-	5,5,5	0.26	0	5,5,5	0.44	0
3	GOL	A	605	-	5,5,5	0.30	0	5,5,5	0.33	0
4	FMT	A	606	-	2,2,2	0.58	0	1,1,1	0.28	0
4	FMT	A	607	-	2,2,2	0.70	0	1,1,1	0.27	0
4	FMT	A	608	-	2,2,2	0.61	0	1,1,1	0.26	0
4	FMT	A	609	-	2,2,2	0.63	0	1,1,1	0.16	0
4	FMT	A	610	-	2,2,2	0.67	0	1,1,1	0.15	0
3	GOL	B	602	-	5,5,5	0.37	0	5,5,5	0.68	0
3	GOL	B	603	-	5,5,5	0.34	0	5,5,5	0.28	0
4	FMT	B	604	-	2,2,2	0.72	0	1,1,1	0.27	0
4	FMT	B	605	-	2,2,2	0.75	0	1,1,1	0.21	0
4	FMT	B	606	-	2,2,2	0.61	0	1,1,1	0.15	0
4	FMT	B	607	-	2,2,2	0.62	0	1,1,1	0.23	0
4	FMT	B	608	-	2,2,2	0.58	0	1,1,1	0.14	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
3	GOL	A	602	-	-	0/4/4/4	0/0/0/0
3	GOL	A	603	-	-	0/4/4/4	0/0/0/0
3	GOL	A	604	-	-	0/4/4/4	0/0/0/0
3	GOL	A	605	-	-	0/4/4/4	0/0/0/0
4	FMT	A	606	-	-	0/0/0/0	0/0/0/0
4	FMT	A	607	-	-	0/0/0/0	0/0/0/0
4	FMT	A	608	-	-	0/0/0/0	0/0/0/0
4	FMT	A	609	-	-	0/0/0/0	0/0/0/0
4	FMT	A	610	-	-	0/0/0/0	0/0/0/0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	GOL	B	602	-	-	0/4/4/4	0/0/0/0
3	GOL	B	603	-	-	0/4/4/4	0/0/0/0
4	FMT	B	604	-	-	0/0/0/0	0/0/0/0
4	FMT	B	605	-	-	0/0/0/0	0/0/0/0
4	FMT	B	606	-	-	0/0/0/0	0/0/0/0
4	FMT	B	607	-	-	0/0/0/0	0/0/0/0
4	FMT	B	608	-	-	0/0/0/0	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 5.7 Other polymers ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues

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<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	492/506 (97%)	0.13	36 (7%)	15 11	16, 34, 88, 118	0
1	B	483/506 (95%)	0.08	42 (8%)	10 7	16, 32, 81, 104	0
All	All	975/1012 (96%)	0.11	78 (8%)	13 9	16, 33, 83, 118	0

The worst 5 of 78 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	316	CYS	8.0
1	B	320	ILE	8.0
1	B	313	CYS	7.9
1	A	262	PRO	7.4
1	A	310	TRP	7.2

### 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 ⓘ

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
3	GOL	A	605	6/6	0.32	34.04	44,51,62,64	0
4	FMT	B	607	3/3	0.21	12.86	38,38,44,52	0
4	FMT	B	606	3/3	0.20	7.37	43,43,44,52	0
4	FMT	B	608	3/3	0.22	5.42	40,40,44,45	0
4	FMT	A	610	3/3	0.16	4.48	32,32,48,50	0
4	FMT	A	608	3/3	0.16	2.68	47,47,53,66	0
4	FMT	A	609	3/3	0.19	2.50	39,39,41,46	0
3	GOL	A	602	6/6	0.09	1.39	28,29,33,35	0
3	GOL	B	602	6/6	0.15	1.29	29,46,52,52	0
3	GOL	A	603	6/6	0.14	0.96	32,42,46,49	0
4	FMT	A	607	3/3	0.11	0.94	21,21,35,36	0
4	FMT	B	605	3/3	0.09	-0.00	22,22,30,35	0
4	FMT	A	606	3/3	0.07	-0.05	22,22,33,34	0
4	FMT	B	604	3/3	0.08	-0.36	24,24,31,33	0
3	GOL	B	603	6/6	0.07	-0.64	29,33,36,36	0
3	GOL	A	604	6/6	0.06	-0.69	26,31,33,34	0
2	ZN	A	601	1/1	0.09	-1.76	66,66,66,66	1
2	ZN	B	601	1/1	0.13	-2.07	91,91,91,91	1

## 6.5 Other polymers ⓘ

There are no such residues in this entry.