



Full wwPDB X-ray Structure Validation Report ⓘ

Mar 1, 2026 – 12:58 PM UTC

PDB ID : 5M7R / pdb_00005m7r
Title : Structure of human O-GlcNAc hydrolase
Authors : Roth, C.; Chan, S.; Offen, W.A.; Hemsworth, G.R.; Willems, L.I.; King, D.;
Varghese, V.; Britton, R.; Vocadlo, D.J.; Davies, G.J.
Deposited on : 2016-10-28
Resolution : 2.35 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4-5-2 with Phenix2.0
Xtriage (Phenix) : 2.0
EDS : 3.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4 : 9.0.010 (Gargrove)
Density-Fitness : 1.0.12
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.49

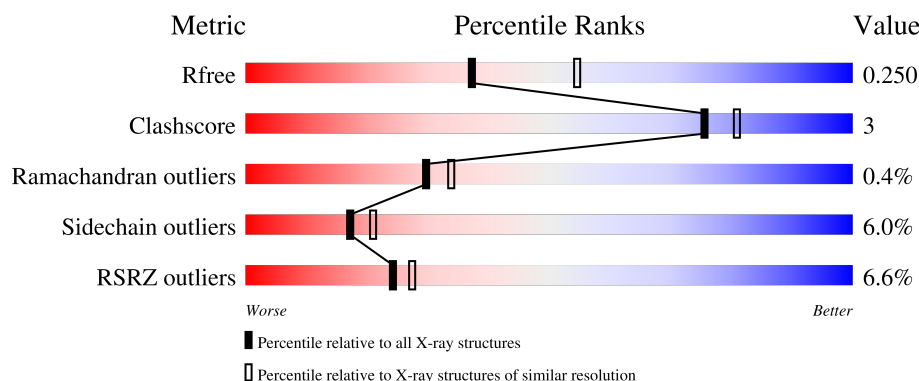
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.35 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}	180053	1596 (2.36-2.36)
Clashscore	190562	1663 (2.36-2.36)
Ramachandran outliers	187476	1646 (2.36-2.36)
Sidechain outliers	187428	1646 (2.36-2.36)
RSRZ outliers	180081	1598 (2.36-2.36)

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 $\leq 5\%$. 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	916	
1	B	916	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 7704 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 Protein O-GlcNAcase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	466	Total	C	N	O	S	0	0	0
			3813	2468	628	693	24			
1	B	470	Total	C	N	O	S	0	0	0
			3839	2481	634	699	25			

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	21	Total	O	0	0
			21	21		
2	B	31	Total	O	0	0
			31	31		

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.

Chain A:

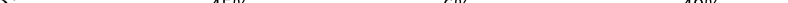
4%

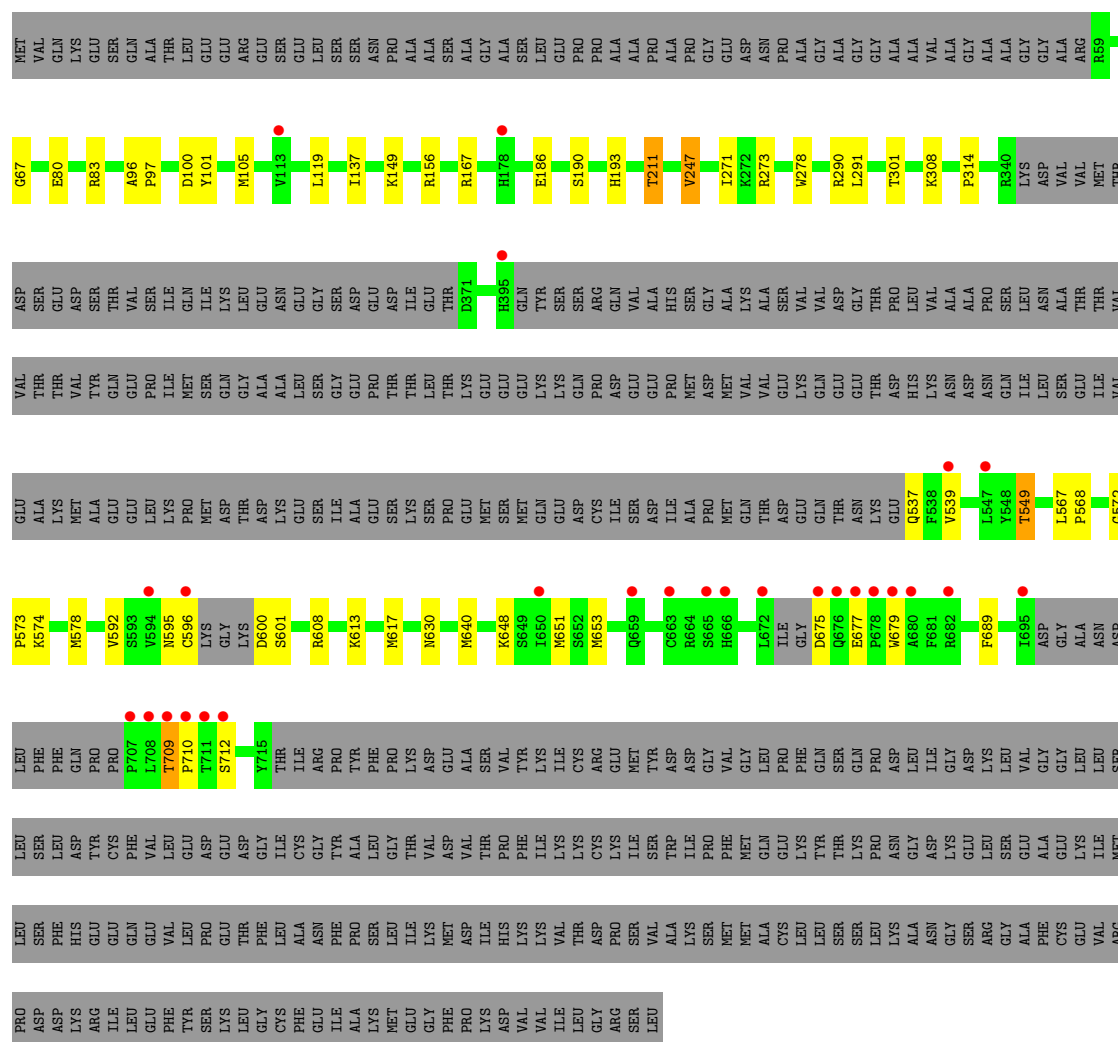
43%

6%

49%

- Molecule 1: Protein O-GlcNAcase

Chain B: 



4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α , β , γ	102.23Å 102.23Å 285.49Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	96.25 – 2.35 96.25 – 2.35	Depositor EDS
% Data completeness (in resolution range)	100.0 (96.25-2.35) 100.0 (96.25-2.35)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.51 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0155, REFMAC 5.8.0155	Depositor
R, R_{free}	0.216 , 0.249 0.219 , 0.250	Depositor DCC
R_{free} test set	1909 reflections (2.98%)	wwPDB-VP
Wilson B-factor (Å ²)	59.3	Xtriage
Anisotropy	0.630	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.30 , 41.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	7704	wwPDB-VP
Average B, all atoms (Å ²)	80.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.08% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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	1.06	1/3915 (0.0%)	1.14	14/5302 (0.3%)
1	B	0.89	0/3942	1.02	5/5339 (0.1%)
All	All	0.98	1/7857 (0.0%)	1.08	19/10641 (0.2%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	142	ASP	CA-C	5.61	1.59	1.52

All (19) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	167	ARG	CB-CG-CD	8.30	130.39	111.30
1	B	601	SER	N-CA-C	7.58	120.21	111.11
1	B	595	ASN	N-CA-C	7.34	122.24	107.41
1	A	167	ARG	CG-CD-NE	7.19	127.82	112.00
1	A	601	SER	N-CA-C	7.18	120.02	111.33
1	A	679	TRP	N-CA-C	7.16	120.91	111.75
1	A	626	THR	N-CA-CB	6.96	120.47	110.16
1	A	159	ASP	CB-CA-C	6.37	123.38	110.38
1	A	142	ASP	N-CA-CB	-5.92	101.83	111.18
1	A	671	PHE	N-CA-C	5.77	117.44	111.03
1	B	679	TRP	N-CA-C	5.66	118.22	111.71
1	B	101	TYR	N-CA-C	5.64	118.97	111.75
1	A	677	GLU	CB-CG-CD	5.62	122.16	112.60
1	A	101	TYR	N-CA-C	5.52	118.82	111.75
1	A	114	GLU	CB-CA-C	-5.24	102.62	110.90
1	A	159	ASP	N-CA-CB	5.20	118.31	110.30
1	A	574	LYS	N-CA-C	5.12	116.87	111.28
1	B	574	LYS	N-CA-C	5.06	116.80	111.28
1	A	543	ASN	N-CA-C	5.02	118.56	112.23

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	3813	0	3741	29	0
1	B	3839	0	3759	22	0
2	A	21	0	0	0	0
2	B	31	0	0	1	0
All	All	7704	0	7500	45	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 3.

All (45) 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:75:MET:HE1	1:A:122:LEU:HD22	1.25	1.09
1:B:596:CYS:HG	1:B:600:ASP:N	1.61	0.98
1:A:75:MET:CE	1:A:122:LEU:HD22	2.16	0.68
1:A:689:PHE:CE2	1:B:651:MET:HE1	2.29	0.67
1:A:105:MET:HE3	1:B:630:ASN:OD1	1.98	0.63
1:A:595:ASN:O	1:A:599:LYS:N	2.32	0.62
1:A:709:THR:HG22	1:A:710:PRO:HD2	1.85	0.57
1:B:578:MET:HE2	1:B:617:MET:HB3	1.88	0.56
1:A:75:MET:HE3	1:A:75:MET:HA	1.89	0.55
1:A:709:THR:HG22	1:A:710:PRO:CD	2.37	0.55
1:A:143:ILE:HG13	1:A:150:GLU:HG2	1.88	0.54
1:B:596:CYS:SG	1:B:600:ASP:N	2.74	0.54
1:B:592:VAL:O	1:B:592:VAL:HG12	2.08	0.54
1:A:676:GLN:HE21	1:A:676:GLN:HA	1.71	0.53
1:A:190:SER:OG	1:A:193:HIS:HD2	1.92	0.52
1:A:578:MET:HE2	1:A:617:MET:HB3	1.90	0.52
1:A:106:PHE:CE1	1:B:549:THR:HG21	2.45	0.51
1:A:592:VAL:HG12	1:A:592:VAL:O	2.10	0.51
1:A:651:MET:HE1	1:B:689:PHE:CE2	2.45	0.51
1:A:75:MET:HE3	1:A:78:ARG:HD2	1.94	0.50

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:622:MET:N	1:A:622:MET:HE3	2.26	0.50
1:B:190:SER:OG	1:B:193:HIS:HD2	1.95	0.50
1:B:709:THR:HG22	1:B:710:PRO:CD	2.42	0.50
1:B:80:GLU:OE2	1:B:83:ARG:NH1	2.45	0.49
1:A:143:ILE:HG13	1:A:150:GLU:CG	2.43	0.49
1:A:67:GLY:HA2	1:A:96:ALA:O	2.13	0.48
1:A:684:GLY:CA	1:B:290:ARG:HD3	2.44	0.48
1:A:385:THR:HG22	1:A:555:LEU:HD22	1.95	0.47
1:B:709:THR:HG22	1:B:710:PRO:HD2	1.97	0.47
1:B:67:GLY:HA2	1:B:96:ALA:O	2.16	0.46
1:A:97:PRO:HG2	1:A:100:ASP:HB2	1.97	0.46
1:A:630:ASN:ND2	1:B:105:MET:HE3	2.30	0.46
1:A:291:LEU:HD21	1:A:640:MET:HE3	1.98	0.45
1:A:572:GLY:O	1:A:573:PRO:C	2.59	0.45
1:A:567:LEU:HB3	1:A:568:PRO:HD2	1.99	0.45
1:B:567:LEU:HB3	1:B:568:PRO:HD2	1.99	0.44
1:B:572:GLY:O	1:B:573:PRO:C	2.60	0.44
1:A:296:TYR:CE2	1:A:299:ARG:HG2	2.52	0.44
1:B:97:PRO:HG2	1:B:100:ASP:HB2	1.99	0.44
1:B:211:THR:HG22	2:B:1014:HOH:O	2.19	0.42
1:B:247:VAL:CG2	1:B:271:ILE:HD13	2.49	0.42
1:B:96:ALA:N	1:B:97:PRO:CD	2.83	0.41
1:A:326:HIS:CD2	1:A:383:ALA:HB2	2.56	0.41
1:B:291:LEU:HD21	1:B:640:MET:HE3	2.03	0.41
1:A:96:ALA:N	1:A:97:PRO:CD	2.84	0.41

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	454/916 (50%)	437 (96%)	15 (3%)	2 (0%)	30	34

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	B	458/916 (50%)	440 (96%)	16 (4%)	2 (0%)	30	34
All	All	912/1832 (50%)	877 (96%)	31 (3%)	4 (0%)	30	34

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	314	PRO
1	B	314	PRO
1	B	712	SER
1	A	710	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	416/797 (52%)	388 (93%)	28 (7%)	15	17
1	B	419/797 (53%)	397 (95%)	22 (5%)	20	25
All	All	835/1594 (52%)	785 (94%)	50 (6%)	17	21

All (50) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	75	MET
1	A	137	ILE
1	A	142	ASP
1	A	149	LYS
1	A	167	ARG
1	A	186	GLU
1	A	262	GLU
1	A	273	ARG
1	A	278	TRP
1	A	289	LYS
1	A	299	ARG
1	A	301	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	303	LEU
1	A	308	LYS
1	A	381	LYS
1	A	539	VAL
1	A	599	LYS
1	A	603	LYS
1	A	622	MET
1	A	626	THR
1	A	648	LYS
1	A	656	SER
1	A	673	ILE
1	A	676	GLN
1	A	677	GLU
1	A	678	PRO
1	A	682	ARG
1	A	709	THR
1	B	119	LEU
1	B	137	ILE
1	B	149	LYS
1	B	156	ARG
1	B	167	ARG
1	B	186	GLU
1	B	211	THR
1	B	247	VAL
1	B	273	ARG
1	B	278	TRP
1	B	301	THR
1	B	308	LYS
1	B	537	GLN
1	B	539	VAL
1	B	549	THR
1	B	608	ARG
1	B	613	LYS
1	B	648	LYS
1	B	653	MET
1	B	675	ASP
1	B	677	GLU
1	B	709	THR

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	103	HIS
1	A	118	GLN
1	A	193	HIS
1	A	288	GLN
1	A	326	HIS
1	A	335	ASN
1	A	537	GLN
1	A	630	ASN
1	A	676	GLN
1	B	103	HIS
1	B	118	GLN
1	B	178	HIS
1	B	193	HIS
1	B	288	GLN

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 oligosaccharides 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, 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	466/916 (50%)	0.47	35 (7%) 20 23	57, 83, 124, 164	0
1	B	470/916 (51%)	0.38	27 (5%) 29 34	45, 68, 127, 160	0
All	All	936/1832 (51%)	0.43	62 (6%) 24 27	45, 78, 126, 164	0

All (62) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	695	ILE	8.6
1	B	672	LEU	7.0
1	A	695	ILE	6.4
1	B	666	HIS	6.3
1	B	707	PRO	6.2
1	A	711	THR	6.0
1	A	673	ILE	5.8
1	A	679	TRP	5.1
1	B	665	SER	4.8
1	B	711	THR	4.8
1	A	708	LEU	4.7
1	B	678	PRO	4.7
1	B	710	PRO	4.3
1	A	672	LEU	4.3
1	B	663	CYS	4.2
1	A	660	TRP	4.2
1	B	679	TRP	4.0
1	A	663	CYS	3.9
1	A	676	GLN	3.6
1	A	666	HIS	3.4
1	A	395	HIS	3.3
1	B	113	VAL	3.3
1	A	715	TYR	3.2
1	B	594	VAL	3.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	709	THR	3.0
1	A	678	PRO	3.0
1	B	659	GLN	2.8
1	B	677	GLU	2.8
1	A	59	ARG	2.8
1	B	708	LEU	2.8
1	B	395	HIS	2.7
1	A	710	PRO	2.7
1	A	607	TRP	2.7
1	A	372	VAL	2.6
1	A	592	VAL	2.6
1	A	665	SER	2.6
1	A	373	LEU	2.6
1	A	677	GLU	2.6
1	A	594	VAL	2.6
1	A	662	GLY	2.5
1	B	712	SER	2.5
1	A	664	ARG	2.4
1	B	547	LEU	2.4
1	A	709	THR	2.4
1	B	539	VAL	2.4
1	B	650	ILE	2.4
1	A	378	MET	2.4
1	A	177	ASP	2.4
1	A	671	PHE	2.3
1	B	596	CYS	2.3
1	A	179	ASN	2.3
1	A	713	LYS	2.2
1	A	712	SER	2.2
1	A	657	PHE	2.2
1	B	680	ALA	2.1
1	B	682	ARG	2.1
1	B	178	HIS	2.1
1	A	668	SER	2.1
1	A	178	HIS	2.1
1	A	661	LEU	2.1
1	B	676	GLN	2.0
1	B	675	ASP	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 oligosaccharides 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.