



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 10, 2020 – 04:32 AM BST

PDB ID : 5VVV
Title : Structural Investigations of the Substrate Specificity of Human O-GlcNAcase
Authors : Li, B.; Jiang, J.; Li, H.; Hu, C.-W.
Deposited on : 2017-05-20
Resolution : 2.80 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 ⓘ 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 : 2.13.1
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) : 2.13.1

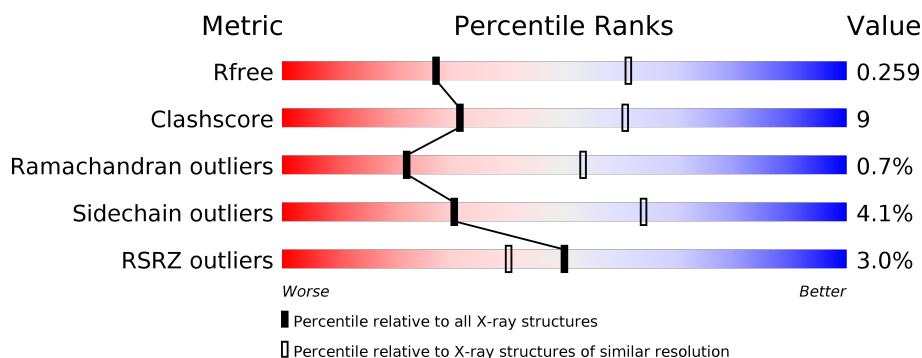
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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 ≥ 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	504	<div> <div>3%</div> <div> <div></div> <div>67%</div> <div>15%</div> <div>•</div> <div>17%</div> </div> </div>
1	C	504	<div> <div>2%</div> <div> <div></div> <div>65%</div> <div>18%</div> <div>•</div> <div>16%</div> </div> </div>
2	B	13	<div> <div>15%</div> <div>85%</div> </div>
2	D	13	<div> <div>8%</div> <div>31%</div> <div>62%</div> </div>

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7015 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	416	Total	C	N	O	S	0	0	0
			3428	2227	563	616	22			
1	C	423	Total	C	N	O	S	0	0	0
			3477	2253	574	627	23			

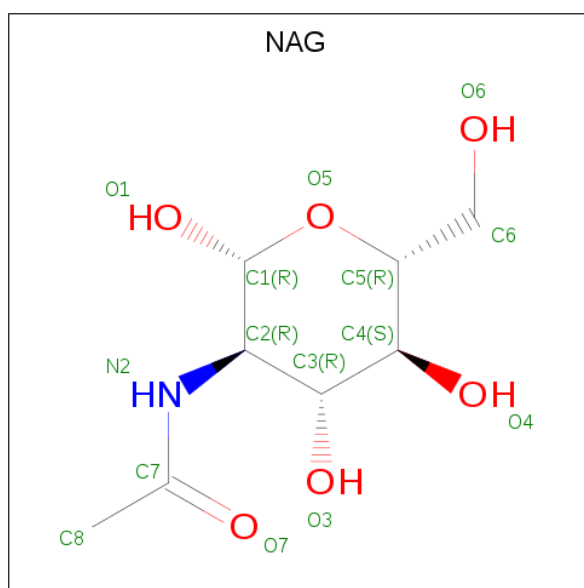
There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	59	HIS	-	expression tag	UNP O60502
A	175	ASN	ASP	engineered mutation	UNP O60502
A	543	GLY	-	linker	UNP O60502
A	544	GLY	-	linker	UNP O60502
A	545	GLY	-	linker	UNP O60502
A	546	GLY	-	linker	UNP O60502
A	547	SER	-	linker	UNP O60502
A	548	GLY	-	linker	UNP O60502
A	549	GLY	-	linker	UNP O60502
A	550	GLY	-	linker	UNP O60502
A	551	GLY	-	linker	UNP O60502
A	552	SER	-	linker	UNP O60502
C	59	HIS	-	expression tag	UNP O60502
C	175	ASN	ASP	engineered mutation	UNP O60502
C	543	GLY	-	linker	UNP O60502
C	544	GLY	-	linker	UNP O60502
C	545	GLY	-	linker	UNP O60502
C	546	GLY	-	linker	UNP O60502
C	547	SER	-	linker	UNP O60502
C	548	GLY	-	linker	UNP O60502
C	549	GLY	-	linker	UNP O60502
C	550	GLY	-	linker	UNP O60502
C	551	GLY	-	linker	UNP O60502
C	552	SER	-	linker	UNP O60502

- Molecule 2 is a protein called α -crystallin B.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
2	B	2	Total	C	N	O	0	0	0
			13	7	2	4			
2	D	5	Total	C	N	O	0	0	0
			34	20	5	9			

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	B	1	Total	C	N	O	0	0
			14	8	1	5		
3	D	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	21	Total	O	0	0
			21	21		
4	C	14	Total	O	0	0
			14	14		

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:

67% 15% 3% 17%

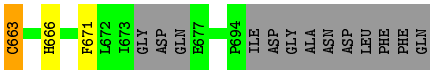
Legend: Green (Favorable), Yellow (Marginal), Red (Unfavorable), Grey (Not Defined)

Residue Number	Amino Acid	State
1	LEU	Grey
2	PHE	Grey
3	PHE	Grey
4	GLN	Grey
5	VAL	Grey
6	ASN	Grey
7	CYS	Grey
8	LYS	Grey
9	GLY	Grey
10	LYS	Grey
11	ASP	Grey
12	GLU	Grey
13	GLU	Grey
14	LYS	Grey
15	LEU	Grey
16	GLY	Grey
17	ASP	Grey
18	GLU	Grey
19	GLY	Grey
20	ASP	Grey
21	GLY	Grey
22	ASP	Grey
23	GLY	Grey
24	ASP	Grey
25	GLY	Grey
26	ASP	Grey
27	GLY	Grey
28	ASP	Grey
29	GLY	Grey
30	ASP	Grey
31	GLY	Grey
32	ASP	Grey
33	GLY	Grey
34	ASP	Grey
35	GLY	Grey
36	ASP	Grey
37	GLY	Grey
38	ASP	Grey
39	GLY	Grey
40	ASP	Grey
41	GLY	Grey
42	ASP	Grey
43	GLY	Grey
44	ASP	Grey
45	GLY	Grey
46	ASP	Grey
47	GLY	Grey
48	ASP	Grey
49	GLY	Grey
50	ASP	Grey
51	GLY	Grey
52	ASP	Grey
53	GLY	Grey
54	ASP	Grey
55	GLY	Grey
56	ASP	Grey
57	GLY	Grey
58	ASP	Grey
59	GLY	Grey
60	ASP	Grey
61	GLY	Grey
62	ASP	Grey
63	GLY	Grey
64	ASP	Grey
65	GLY	Grey
66	ASP	Grey
67	GLY	Grey
68	ASP	Grey
69	GLY	Grey
70	ASP	Grey
71	GLY	Grey
72	ASP	Grey
73	GLY	Grey
74	ASP	Grey
75	GLY	Grey
76	ASP	Grey
77	GLY	Grey
78	ASP	Grey
79	GLY	Grey
80	ASP	Grey
81	GLY	Grey
82	ASP	Grey
83	GLY	Grey
84	ASP	Grey
85	GLY	Grey
86	ASP	Grey
87	GLY	Grey
88	ASP	Grey
89	GLY	Grey
90	ASP	Grey
91	GLY	Grey
92	ASP	Grey
93	GLY	Grey
94	ASP	Grey
95	GLY	Grey
96	ASP	Grey
97	GLY	Grey
98	ASP	Grey
99	GLY	Grey
100	ASP	Grey
101	GLY	Grey
102	ASP	Grey
103	GLY	Grey
104	ASP	Grey
105	GLY	Grey
106	ASP	Grey
107	GLY	Grey
108	ASP	Grey
109	GLY	Grey
110	ASP	Grey
111	GLY	Grey
112	ASP	Grey
113	GLY	Grey
114	ASP	Grey
115	GLY	Grey
116	ASP	Grey
117	GLY	Grey
118	ASP	Grey
119	GLY	Grey
120	ASP	Grey
121	GLY	Grey
122	ASP	Grey
123	GLY	Grey
124	ASP	Grey
125	GLY	Grey
126	ASP	Grey
127	GLY	Grey
128	ASP	Grey
129	GLY	Grey
130	ASP	Grey
131	GLY	Grey
132	ASP	Grey
133	GLY	Grey
134	ASP	Grey
135	GLY	Grey
136	ASP	Grey
137	GLY	Grey
138	ASP	Grey
139	GLY	Grey
140	ASP	Grey
141	GLY	Grey
142	ASP	Grey
143	GLY	Grey
144	ASP	Grey
145	GLY	Grey
146	ASP	Grey
147	GLY	Grey
148	ASP	Grey
149	GLY	Grey
150	ASP	Grey
151	GLY	Grey
152	ASP	Grey
153	GLY	Grey
154	ASP	Grey
155	GLY	Grey
156	ASP	Grey
157	GLY	Grey
158	ASP	Grey
159	GLY	Grey
160	ASP	Grey
161	GLY	Grey
162	ASP	Grey
163	GLY	Grey

Chain C:

29% 65% 18% 16%

Amino Acid	Percentage
GLY	29%
VAL	65%
ARG	18%
LYS	16%



● Molecule 2: a-crystallin B



● Molecule 2: a-crystallin B



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	83.08Å 96.26Å 89.81Å 90.00° 114.29° 90.00°	Depositor
Resolution (Å)	50.00 – 2.80 46.81 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.9 (50.00-2.80) 99.9 (46.81-2.80)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.22 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.8.0131	Depositor
R, R_{free}	0.196 , 0.257 0.196 , 0.259	Depositor DCC
R_{free} test set	1576 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å ²)	68.8	Xtriage
Anisotropy	0.051	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 66.6	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	0.94	EDS
Total number of atoms	7015	wwPDB-VP
Average B, all atoms (Å ²)	74.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.82% 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 [i](#)

5.1 Standard geometry [i](#)

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

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	0.70	0/3522	0.83	2/4769 (0.0%)
1	C	0.67	0/3571	0.81	1/4832 (0.0%)
2	B	0.76	0/12	0.89	0/15
2	D	0.70	0/33	0.99	0/44
All	All	0.69	0/7138	0.82	3/9660 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	104	ARG	NE-CZ-NH1	-5.53	117.53	120.30
1	C	287	ASP	CB-CG-OD1	5.37	123.14	118.30
1	A	628	LEU	CB-CG-CD2	-5.12	102.29	111.00

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	3428	0	3357	48	0
1	C	3477	0	3403	70	0
2	B	13	0	10	1	0
2	D	34	0	33	10	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	14	0	13	0	0
3	D	14	0	13	0	0
4	A	21	0	0	1	0
4	C	14	0	0	1	0
All	All	7015	0	6829	124	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 9.

The worst 5 of 124 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:220:CYS:HB3	1:C:250:THR:HG22	1.38	1.01
1:C:266:GLU:O	1:C:269:LYS:HG2	1.65	0.97
1:C:266:GLU:O	1:C:269:LYS:CG	2.14	0.95
2:D:41:SER:O	2:D:42:THR:OG1	1.86	0.92
1:C:264:ILE:HD13	1:C:307:LEU:HD21	1.57	0.85

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	406/504 (81%)	371 (91%)	33 (8%)	2 (0%)	29	61
1	C	413/504 (82%)	389 (94%)	22 (5%)	2 (0%)	29	61
2	D	3/13 (23%)	1 (33%)	0	2 (67%)	0	0
All	All	822/1021 (80%)	761 (93%)	55 (7%)	6 (1%)	22	53

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	43	SER
1	A	140	GLY
2	D	42	THR
1	A	314	PRO
1	C	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	370/442 (84%)	359 (97%)	11 (3%)	41	75
1	C	376/442 (85%)	358 (95%)	18 (5%)	25	58
2	B	2/13 (15%)	2 (100%)	0	100	100
2	D	5/13 (38%)	3 (60%)	2 (40%)	0	0
All	All	753/910 (83%)	722 (96%)	31 (4%)	30	64

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	171	LEU
1	C	242	LEU
1	C	663	CYS
1	C	199	THR
1	C	250	THR

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

Mol	Chain	Res	Type
1	C	288	GLN

5.3.3 RNA ⓘ

There are no RNA molecules 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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

2 ligands are modelled in this entry.

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	NAG	D	101	2	14,14,15	2.88	5 (35%)	17,19,21	2.95	8 (47%)
3	NAG	B	101	2	14,14,15	2.66	3 (21%)	17,19,21	3.03	8 (47%)

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	NAG	D	101	2	-	1/6/23/26	0/1/1/1
3	NAG	B	101	2	-	2/6/23/26	0/1/1/1

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	101	NAG	C1-C2	7.55	1.63	1.52
3	B	101	NAG	C1-C2	7.45	1.63	1.52
3	D	101	NAG	C3-C2	4.89	1.62	1.52
3	B	101	NAG	C8-C7	4.24	1.59	1.50
3	B	101	NAG	C3-C2	3.53	1.60	1.52

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	101	NAG	C1-O5-C5	5.96	120.26	112.19
3	B	101	NAG	C3-C4-C5	-5.66	100.15	110.24
3	D	101	NAG	C3-C4-C5	-5.63	100.20	110.24
3	D	101	NAG	C1-O5-C5	5.51	119.66	112.19
3	D	101	NAG	O5-C1-C2	4.44	118.29	111.29

There are no chirality outliers.

All (3) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	B	101	NAG	C1-C2-N2-C7
3	D	101	NAG	C3-C2-N2-C7
3	B	101	NAG	C3-C2-N2-C7

There are no ring outliers.

No monomer is involved in short contacts.

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	416/504 (82%)	0.03	17 (4%) 37 27	34, 64, 107, 144	0
1	C	423/504 (83%)	0.00	8 (1%) 66 59	43, 78, 111, 146	0
2	B	2/13 (15%)	0.27	0 100 100	69, 69, 69, 106	0
2	D	5/13 (38%)	0.67	0 100 100	90, 100, 118, 132	0
All	All	846/1034 (81%)	0.02	25 (2%) 50 40	34, 73, 110, 146	0

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	187	VAL	5.6
1	A	145	PHE	4.6
1	A	591	VAL	4.2
1	A	225	TYR	4.0
1	A	607	TRP	3.7

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](#)

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(\AA^2)	Q<0.9
3	NAG	B	101	14/15	0.97	0.17	43,50,58,59	0
3	NAG	D	101	14/15	0.98	0.19	41,52,70,70	0

6.5 Other polymers [i](#)

There are no such residues in this entry.