



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 06:12 PM UTC

PDB ID : 5WZR / pdb_00005wzr
Title : Alpha-N-acetylgalactosaminidase NagBb from Bifidobacterium bifidum - Gal-NHAc-DNJ complex
Authors : Sato, M.; Arakawa, T.; Ashida, H.; Fushinobu, S.
Deposited on : 2017-01-18
Resolution : 2.79 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 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
Mogul	:	2022.3.0, CSD as543be (2022)
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

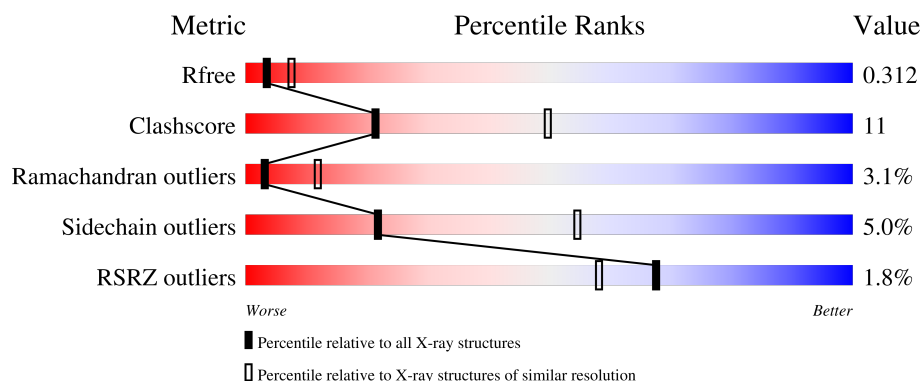
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.79 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	3866 (2.80-2.80)
Clashscore	190562	4276 (2.80-2.80)
Ramachandran outliers	187476	4196 (2.80-2.80)
Sidechain outliers	187428	4198 (2.80-2.80)
RSRZ outliers	180081	3869 (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 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	640	% 72% 23% ..
1	B	640	3% 71% 25% ..

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10008 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 Alpha-N-acetylgalactosaminidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	630	Total	C	N	O	S	0	0	0
			4984	3126	891	929	38			
1	B	630	Total	C	N	O	S	0	0	0
			4988	3129	893	928	38			

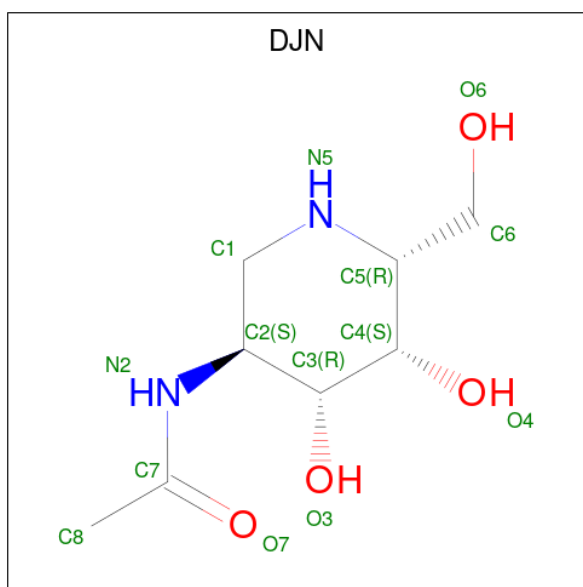
There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	635	HIS	-	expression tag	UNP G5ELM1
A	636	HIS	-	expression tag	UNP G5ELM1
A	637	HIS	-	expression tag	UNP G5ELM1
A	638	HIS	-	expression tag	UNP G5ELM1
A	639	HIS	-	expression tag	UNP G5ELM1
A	640	HIS	-	expression tag	UNP G5ELM1
B	635	HIS	-	expression tag	UNP G5ELM1
B	636	HIS	-	expression tag	UNP G5ELM1
B	637	HIS	-	expression tag	UNP G5ELM1
B	638	HIS	-	expression tag	UNP G5ELM1
B	639	HIS	-	expression tag	UNP G5ELM1
B	640	HIS	-	expression tag	UNP G5ELM1

- Molecule 2 is ZINC ION (CCD ID: ZN) (formula: Zn).

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

- Molecule 3 is N-[(3S,4R,5S,6R)-4,5-dihydroxy-6-(hydroxymethyl)piperidin-3-yl]acetamide (CCD ID: DJN) (formula: C₈H₁₆N₂O₄).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	2	4		
3	B	1	Total	C	N	O	0	0
			14	8	2	4		

- Molecule 4 is CALCIUM ION (CCD ID: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	1	Total	Ca	0	0
			1	1		
4	B	1	Total	Ca	0	0
			1	1		

- Molecule 5 is water.

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

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:

Segment ID	Percentage	Color
M1	1%	Green
M2	1%	Green
M3	1%	Green
M4	1%	Green
M5	1%	Green
M6	1%	Green
M7	1%	Green
M8	1%	Green
M9	1%	Green
M10	1%	Green
M11	1%	Green
M12	1%	Green
M13	1%	Green
M14	1%	Green
M15	1%	Green
M16	1%	Green
M17	1%	Green
M18	1%	Green
M19	1%	Green
M20	1%	Green
M21	1%	Green
M22	1%	Green
M23	1%	Green
M24	1%	Green
M25	1%	Green
M26	1%	Green
M27	1%	Green
M28	1%	Green
M29	1%	Green
M30	1%	Green
M31	1%	Green
M32	1%	Green
M33	1%	Green
M34	1%	Green
M35	1%	Green
M36	1%	Green
M37	1%	Green
M38	1%	Green
M39	1%	Green
M40	1%	Green
M41	1%	Green
M42	1%	Green
M43	1%	Green
M44	1%	Green
M45	1%	Green
M46	1%	Green
M47	1%	Green
M48	1%	Green
M49	1%	Green
M50	1%	Green
M51	1%	Green
M52	1%	Green
M53	1%	Green
M54	1%	Green
M55	1%	Green
M56	1%	Green
M57	1%	Green
M58	1%	Green
M59	1%	Green
M60	1%	Green
M61	1%	Green
M62	1%	Green
M63	1%	Green
M64	1%	Green
M65	1%	Green
M66	1%	Green
M67	1%	Green
M68	1%	Green
M69	1%	Green
M70	1%	Green
M71	1%	Green
M72	1%	Green
M73	1%	Green
M74	1%	Green
M75	1%	Green
M76	1%	Green
M77	1%	Green
M78	1%	Green
M79	1%	Green
M80	1%	Green
M81	1%	Green
M82	1%	Green
M83	1%	Green
M84	1%	Green
M85	1%	Green
M86	1%	Green
M87	1%	Green
M88	1%	Green
M89	1%	Green
M90	1%	Green
M91	1%	Green
M92	1%	Green
M93	1%	Green
M94	1%	Green
M95	1%	Green
M96	1%	Green
M97	1%	Green
M98	1%	Green
M99	1%	Green
M100	1%	Green

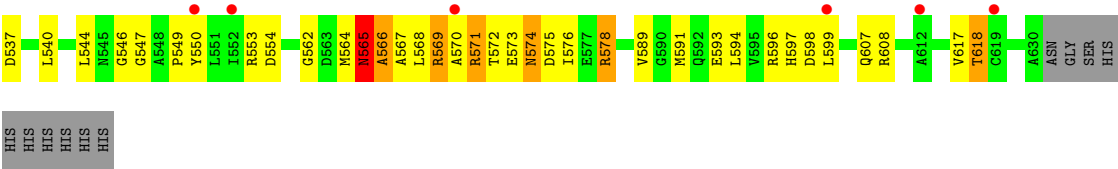
Chain B:

3% 71% 25%

T378
F379
L387
P393
T404
R416
R417
S420
I427
Y433
L434
D435
V436
R454
Y467
H471
G472
S476
S477
E478
E479
V480
V485
F490
Y493
R507
H508
G509
I510
P511
V512
L517
V518
Y519
H520
D521
G523
V523
P526
W527
R531

P210
R214
L225
L226
D230
H231
T232
A233
L234
G235
T236
T237
Y238
R239
L251
A255
N258
P259
R262
L265
G266
R267
S268
W269
V270
H271
V272
G273
L274
K275
D294
V297
Q305
S306
R307
T308
L309
A314
H320
W324
T353
H366
P377

P118
F121
D122
R123
A124
D125
H127
T133
H134
E135
M139
N142
S143
W144
P145
T146
E147
V148
G149
T150
G157
R158
F159
Y165
M166
P167
Q171
S174
D175
G176
H177
A178
Y179
A189
G190
Y191
D192
L193
D194
H195
P196
A197
G198
V204
G205
H206
V207
F208



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	64.63Å 128.48Å 176.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	103.95 – 2.79 103.95 – 2.79	Depositor EDS
% Data completeness (in resolution range)	98.7 (103.95-2.79) 98.7 (103.95-2.79)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.08 (at 2.81Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.229 , 0.313 0.232 , 0.312	Depositor DCC
R_{free} test set	1852 reflections (4.93%)	wwPDB-VP
Wilson B-factor (Å ²)	72.1	Xtriage
Anisotropy	0.045	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 24.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	10008	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 3.49% 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: CA, ZN, DJN

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	0.69	0/5124	0.95	7/6963 (0.1%)
1	B	0.66	0/5129	0.97	6/6971 (0.1%)
All	All	0.67	0/10253	0.96	13/13934 (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	B	0	1

There are no bond length outliers.

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	99	VAL	N-CA-C	7.93	116.09	107.60
1	B	99	VAL	CB-CA-C	-6.05	104.22	111.18
1	B	485	VAL	CA-C-N	-6.04	113.40	119.56
1	B	485	VAL	C-N-CA	-6.04	113.40	119.56
1	A	117	ALA	CA-C-N	5.88	125.89	119.90

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	174	SER	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	4984	0	4675	105	0
1	B	4988	0	4685	108	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
3	A	14	0	16	1	0
3	B	14	0	16	1	0
4	A	1	0	0	0	0
4	B	1	0	0	0	0
5	A	2	0	0	0	0
5	B	2	0	0	0	0
All	All	10008	0	9392	212	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 11.

The worst 5 of 212 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:B:124:ALA:O	1:B:126:ALA:N	1.98	0.94
1:A:74:PHE:O	1:A:76:GLY:N	2.05	0.89
1:B:573:GLU:O	1:B:574:ASN:OD1	1.91	0.86
1:B:107:GLU:HB3	1:B:108:PRO:HA	1.60	0.83
1:A:533:ALA:N	1:A:568:LEU:O	2.19	0.76

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	626/640 (98%)	542 (87%)	61 (10%)	23 (4%)	2	9
1	B	628/640 (98%)	537 (86%)	75 (12%)	16 (2%)	4	16
All	All	1254/1280 (98%)	1079 (86%)	136 (11%)	39 (3%)	3	12

5 of 39 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	73	GLY
1	A	75	ALA
1	A	78	ASP
1	A	106	ALA
1	A	125	ASP

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	511/520 (98%)	490 (96%)	21 (4%)	27	62
1	B	511/520 (98%)	481 (94%)	30 (6%)	18	48
All	All	1022/1040 (98%)	971 (95%)	51 (5%)	22	54

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	109	ARG
1	B	297	VAL
1	B	599	LEU
1	B	125	ASP
1	B	214	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	171	GLN
1	B	574	ASN
1	B	258	ASN
1	B	597	HIS
1	B	368	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](#)

Of 6 ligands modelled in this entry, 4 are monoatomic - leaving 2 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	DJN	A	702	-	14,14,14	1.56	2 (14%)	15,19,19	1.59	5 (33%)
3	DJN	B	702	-	14,14,14	1.62	1 (7%)	15,19,19	1.87	3 (20%)

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	DJN	A	702	-	-	0/6/23/23	0/1/1/1
3	DJN	B	702	-	-	0/6/23/23	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	702	DJN	C1-C2	-5.16	1.47	1.52
3	A	702	DJN	C1-C2	-4.64	1.47	1.52
3	A	702	DJN	O4-C4	2.00	1.47	1.43

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	702	DJN	C2-C1-N5	-4.36	103.34	112.47
3	B	702	DJN	C3-C4-C5	-3.71	105.58	111.02
3	B	702	DJN	O6-C6-C5	-3.38	102.90	111.10
3	A	702	DJN	C3-C4-C5	-2.95	106.70	111.02
3	A	702	DJN	O7-C7-C8	-2.71	117.23	122.05

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	702	DJN	1	0
3	B	702	DJN	1	0

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	630/640 (98%)	0.05	6 (0%) 79 72	40, 73, 113, 164	0
1	B	630/640 (98%)	0.34	17 (2%) 56 46	47, 84, 122, 175	0
All	All	1260/1280 (98%)	0.19	23 (1%) 67 58	40, 79, 118, 175	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	269	TRP	4.2
1	A	567	ALA	3.1
1	B	552	ILE	3.0
1	B	1	MET	2.8
1	B	114	LEU	2.8

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 oligosaccharides 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($\text{\AA}^2$)	Q<0.9
3	DJN	B	702	14/14	0.91	0.08	56,65,68,71	0
3	DJN	A	702	14/14	0.95	0.10	43,47,58,59	0
4	CA	B	703	1/1	0.98	0.04	49,49,49,49	0
4	CA	A	703	1/1	0.99	0.05	37,37,37,37	0
2	ZN	A	701	1/1	1.00	0.04	50,50,50,50	0
2	ZN	B	701	1/1	1.00	0.01	56,56,56,56	0

6.5 Other polymers [i](#)

There are no such residues in this entry.