



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 12, 2026 – 03:36 PM UTC

PDB ID : 5A6J / pdb_00005a6j
Title : GH20C, Beta-hexosaminidase from Streptococcus pneumoniae
Authors : Cid, M.; Robb, C.S.; Higgins, M.A.; Boraston, A.B.
Deposited on : 2015-06-26
Resolution : 1.94 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

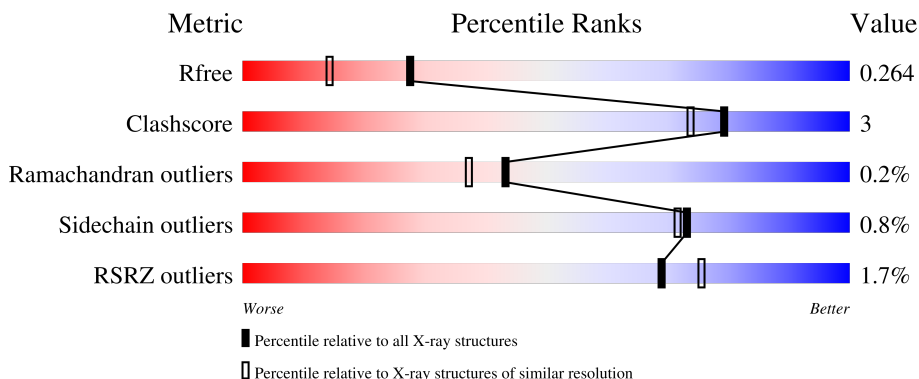
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.94 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	1452 (1.94-1.94)
Clashscore	190562	1494 (1.94-1.94)
Ramachandran outliers	187476	1479 (1.94-1.94)
Sidechain outliers	187428	1479 (1.94-1.94)
RSRZ outliers	180081	1453 (1.94-1.94)

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	649	90% 5% 5%
1	B	649	88% 6% 5%
1	C	649	3% 86% 9% 5%
1	D	649	2% 87% 6% 6%

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 21163 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 N-ACETYL-BETA-D-GLUCOSAMINIDASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	618	Total	C	N	O	S	0	3	0
			4927	3148	825	933	21			
1	B	615	Total	C	N	O	S	0	4	0
			4924	3140	826	938	20			
1	C	614	Total	C	N	O	S	0	1	0
			4858	3102	813	924	19			
1	D	609	Total	C	N	O	S	0	2	0
			4831	3088	810	914	19			

There are 100 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-22	MET	-	expression tag	UNP C1CH75
A	-21	GLY	-	expression tag	UNP C1CH75
A	-20	SER	-	expression tag	UNP C1CH75
A	-19	SER	-	expression tag	UNP C1CH75
A	-18	HIS	-	expression tag	UNP C1CH75
A	-17	HIS	-	expression tag	UNP C1CH75
A	-16	HIS	-	expression tag	UNP C1CH75
A	-15	HIS	-	expression tag	UNP C1CH75
A	-14	HIS	-	expression tag	UNP C1CH75
A	-13	HIS	-	expression tag	UNP C1CH75
A	-12	SER	-	expression tag	UNP C1CH75
A	-11	SER	-	expression tag	UNP C1CH75
A	-10	GLY	-	expression tag	UNP C1CH75
A	-9	LEU	-	expression tag	UNP C1CH75
A	-8	VAL	-	expression tag	UNP C1CH75
A	-7	PRO	-	expression tag	UNP C1CH75
A	-6	ARG	-	expression tag	UNP C1CH75
A	-5	GLY	-	expression tag	UNP C1CH75
A	-4	SER	-	expression tag	UNP C1CH75
A	-3	HIS	-	expression tag	UNP C1CH75
A	-2	MET	-	expression tag	UNP C1CH75

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	ALA	-	expression tag	UNP C1CH75
A	0	SER	-	expression tag	UNP C1CH75
A	9	LEU	PRO	conflict	UNP C1CH75
A	158	VAL	MET	conflict	UNP C1CH75
B	-22	MET	-	expression tag	UNP C1CH75
B	-21	GLY	-	expression tag	UNP C1CH75
B	-20	SER	-	expression tag	UNP C1CH75
B	-19	SER	-	expression tag	UNP C1CH75
B	-18	HIS	-	expression tag	UNP C1CH75
B	-17	HIS	-	expression tag	UNP C1CH75
B	-16	HIS	-	expression tag	UNP C1CH75
B	-15	HIS	-	expression tag	UNP C1CH75
B	-14	HIS	-	expression tag	UNP C1CH75
B	-13	HIS	-	expression tag	UNP C1CH75
B	-12	SER	-	expression tag	UNP C1CH75
B	-11	SER	-	expression tag	UNP C1CH75
B	-10	GLY	-	expression tag	UNP C1CH75
B	-9	LEU	-	expression tag	UNP C1CH75
B	-8	VAL	-	expression tag	UNP C1CH75
B	-7	PRO	-	expression tag	UNP C1CH75
B	-6	ARG	-	expression tag	UNP C1CH75
B	-5	GLY	-	expression tag	UNP C1CH75
B	-4	SER	-	expression tag	UNP C1CH75
B	-3	HIS	-	expression tag	UNP C1CH75
B	-2	MET	-	expression tag	UNP C1CH75
B	-1	ALA	-	expression tag	UNP C1CH75
B	0	SER	-	expression tag	UNP C1CH75
B	9	LEU	PRO	conflict	UNP C1CH75
B	158	VAL	MET	conflict	UNP C1CH75
C	-22	MET	-	expression tag	UNP C1CH75
C	-21	GLY	-	expression tag	UNP C1CH75
C	-20	SER	-	expression tag	UNP C1CH75
C	-19	SER	-	expression tag	UNP C1CH75
C	-18	HIS	-	expression tag	UNP C1CH75
C	-17	HIS	-	expression tag	UNP C1CH75
C	-16	HIS	-	expression tag	UNP C1CH75
C	-15	HIS	-	expression tag	UNP C1CH75
C	-14	HIS	-	expression tag	UNP C1CH75
C	-13	HIS	-	expression tag	UNP C1CH75
C	-12	SER	-	expression tag	UNP C1CH75
C	-11	SER	-	expression tag	UNP C1CH75
C	-10	GLY	-	expression tag	UNP C1CH75

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	-9	LEU	-	expression tag	UNP C1CH75
C	-8	VAL	-	expression tag	UNP C1CH75
C	-7	PRO	-	expression tag	UNP C1CH75
C	-6	ARG	-	expression tag	UNP C1CH75
C	-5	GLY	-	expression tag	UNP C1CH75
C	-4	SER	-	expression tag	UNP C1CH75
C	-3	HIS	-	expression tag	UNP C1CH75
C	-2	MET	-	expression tag	UNP C1CH75
C	-1	ALA	-	expression tag	UNP C1CH75
C	0	SER	-	expression tag	UNP C1CH75
C	9	LEU	PRO	conflict	UNP C1CH75
C	158	VAL	MET	conflict	UNP C1CH75
D	-22	MET	-	expression tag	UNP C1CH75
D	-21	GLY	-	expression tag	UNP C1CH75
D	-20	SER	-	expression tag	UNP C1CH75
D	-19	SER	-	expression tag	UNP C1CH75
D	-18	HIS	-	expression tag	UNP C1CH75
D	-17	HIS	-	expression tag	UNP C1CH75
D	-16	HIS	-	expression tag	UNP C1CH75
D	-15	HIS	-	expression tag	UNP C1CH75
D	-14	HIS	-	expression tag	UNP C1CH75
D	-13	HIS	-	expression tag	UNP C1CH75
D	-12	SER	-	expression tag	UNP C1CH75
D	-11	SER	-	expression tag	UNP C1CH75
D	-10	GLY	-	expression tag	UNP C1CH75
D	-9	LEU	-	expression tag	UNP C1CH75
D	-8	VAL	-	expression tag	UNP C1CH75
D	-7	PRO	-	expression tag	UNP C1CH75
D	-6	ARG	-	expression tag	UNP C1CH75
D	-5	GLY	-	expression tag	UNP C1CH75
D	-4	SER	-	expression tag	UNP C1CH75
D	-3	HIS	-	expression tag	UNP C1CH75
D	-2	MET	-	expression tag	UNP C1CH75
D	-1	ALA	-	expression tag	UNP C1CH75
D	0	SER	-	expression tag	UNP C1CH75
D	9	LEU	PRO	conflict	UNP C1CH75
D	158	VAL	MET	conflict	UNP C1CH75

- Molecule 2 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: C₂H₆O₂).



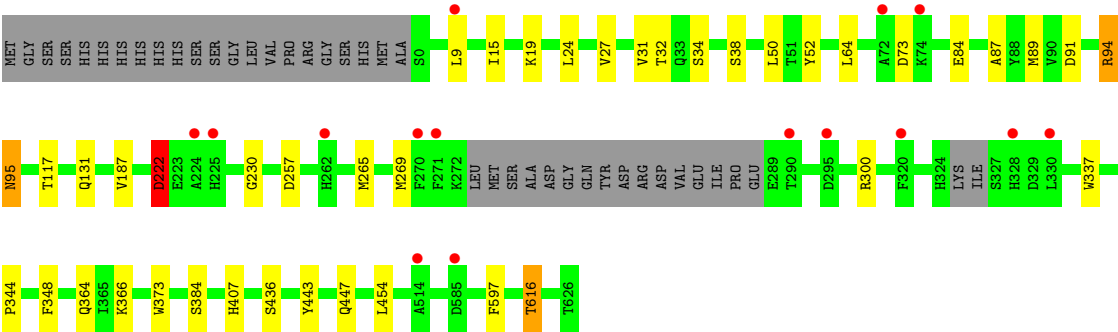
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	510	Total	O	0	0
			510	510		
3	B	427	Total	O	0	0
			427	427		
3	C	334	Total	O	0	0
			334	334		
3	D	348	Total	O	0	0
			348	348		

● Molecule 1: N-ACETYL-BETA-D-GLUCOSAMINIDASE





4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	92.84Å 136.94Å 202.35Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	113.41 – 1.94 113.41 – 1.94	Depositor EDS
% Data completeness (in resolution range)	97.2 (113.41-1.94) 97.2 (113.41-1.94)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.82 (at 1.94Å)	Xtriage
Refinement program	REFMAC 5.8.0071	Depositor
R, R_{free}	0.206 , 0.258 0.214 , 0.264	Depositor DCC
R_{free} test set	9269 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	14.4	Xtriage
Anisotropy	0.194	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 41.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.46$, $\langle L^2 \rangle = 0.28$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	21163	wwPDB-VP
Average B, all atoms (Å ²)	20.0	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 25.38 % 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 3.2085e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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

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

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	0.68	0/5043	0.84	3/6838 (0.0%)
1	B	0.67	0/5043	0.83	6/6837 (0.1%)
1	C	0.67	0/4965	0.83	3/6741 (0.0%)
1	D	0.64	0/4940	0.84	3/6703 (0.0%)
All	All	0.66	0/19991	0.84	15/27119 (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	D	0	1

There are no bond length outliers.

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	222	ASP	N-CA-C	-7.21	101.02	110.53
1	B	222	ASP	N-CA-C	-6.29	101.80	110.35
1	A	428	LEU	CA-C-N	5.77	126.17	119.47
1	A	428	LEU	C-N-CA	5.77	126.17	119.47
1	A	348	PHE	N-CA-C	-5.76	105.08	111.36

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	222	ASP	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	4927	0	4772	21	0
1	B	4924	0	4766	27	0
1	C	4858	0	4658	36	0
1	D	4831	0	4664	22	0
2	B	4	0	6	0	0
3	A	510	0	0	2	0
3	B	427	0	0	6	0
3	C	334	0	0	5	0
3	D	348	0	0	3	0
All	All	21163	0	18866	105	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.

The worst 5 of 105 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:D:265:MET:HE1	1:D:269:MET:HE3	1.70	0.73
1:D:616:THR:HG21	3:D:2324:HOH:O	1.88	0.72
1:B:222:ASP:HB2	1:B:223:GLU:CB	2.22	0.70
1:D:257:ASP:OD2	1:D:300:ARG:NH2	2.26	0.68
1:A:1:MET:HE3	1:A:26:ASP:HB2	1.79	0.64

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	617/649 (95%)	594 (96%)	22 (4%)	1 (0%)	43	37
1	B	615/649 (95%)	592 (96%)	21 (3%)	2 (0%)	36	30
1	C	611/649 (94%)	584 (96%)	25 (4%)	2 (0%)	36	30
1	D	605/649 (93%)	583 (96%)	21 (4%)	1 (0%)	43	37
All	All	2448/2596 (94%)	2353 (96%)	89 (4%)	6 (0%)	43	37

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	327	SER
1	A	95	ASN
1	B	95	ASN
1	B	223	GLU
1	C	95	ASN

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	517/557 (93%)	515 (100%)	2 (0%)	84	84
1	B	519/557 (93%)	516 (99%)	3 (1%)	78	78
1	C	502/557 (90%)	498 (99%)	4 (1%)	73	71
1	D	503/557 (90%)	496 (99%)	7 (1%)	59	52
All	All	2041/2228 (92%)	2025 (99%)	16 (1%)	73	71

5 of 16 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	222	ASP
1	D	94	ARG
1	C	417	GLU
1	D	34	SER
1	C	294	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	49	GLN
1	D	317	ASN
1	C	363	ASN
1	D	513	GLN
1	C	612	ASN

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](#)

1 ligand is 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	$\# Z > 2$
2	EDO	B	1627	-	3,3,3	0.43	0	2,2,2	0.70	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
2	EDO	B	1627	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	1627	EDO	O1-C1-C2-O2

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	618/649 (95%)	-0.18	4 (0%) 85 89	6, 16, 31, 45	6 (0%)
1	B	615/649 (94%)	-0.08	4 (0%) 84 88	8, 18, 32, 42	7 (1%)
1	C	614/649 (94%)	0.16	19 (3%) 51 57	7, 20, 41, 59	8 (1%)
1	D	609/649 (93%)	0.12	15 (2%) 58 66	8, 21, 41, 57	4 (0%)
All	All	2456/2596 (94%)	0.00	42 (1%) 69 75	6, 19, 37, 59	25 (1%)

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	328	HIS	5.5
1	C	326	ILE	5.1
1	D	290	THR	4.2
1	C	273	LEU	3.7
1	D	328	HIS	3.6

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
2	EDO	B	1627	4/4	0.81	0.14	23,24,24,27	0

6.5 Other polymers [i](#)

There are no such residues in this entry.