W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report (i)

Mar 18, 2024 – 10:45 AM JST

PDB ID	:	5XGX
Title	:	Crystal structure of colwellia psychrerythraea strain 34H isoaspartyl dipeptidase E80Q mutant complexed with beta-isoaspartyl lysine
Authors	:	Lee, J.H.; Lee, C.W.; Park, S.H.
Deposited on	:	2017-04-18
Resolution	:	2.33 Å(reported)

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

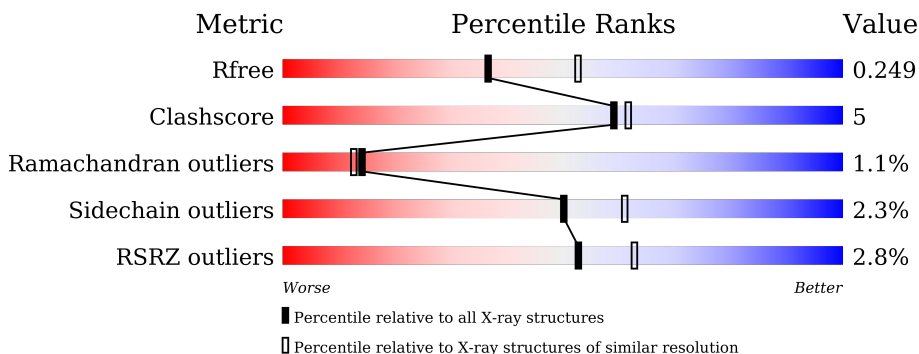|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 2.33 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	Percentile Ranks	Value
Rfree		0.249
Clashscore		5
Ramachandran outliers		1.1%
Sidechain outliers		2.3%
RSRZ outliers		2.8%

Worse ← → Better

▌ Percentile relative to all X-ray structures

❙ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2096 (2.36-2.32)
Clashscore	141614	2193 (2.36-2.32)
Ramachandran outliers	138981	2159 (2.36-2.32)
Sidechain outliers	138945	2160 (2.36-2.32)
RSRZ outliers	127900	2067 (2.36-2.32)

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 <=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	395	3% / 84% / 10% • • •
1	B	395	3% / 88% / 8% • • •

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
4	DLY	B	404	-	-	-	X

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 6154 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 Isoaspartyl dipeptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	381	Total	C	N	O	S	0	0	0
			2890	1832	494	550	14			
1	B	381	Total	C	N	O	S	0	0	0
			2890	1832	494	550	14			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	80	GLN	GLU	engineered mutation	UNP Q484B6
B	80	GLN	GLU	engineered mutation	UNP Q484B6

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

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

- Molecule 3 is D-ASPARTIC ACID (three-letter code: DAS) (formula: $C_4H_7NO_4$).

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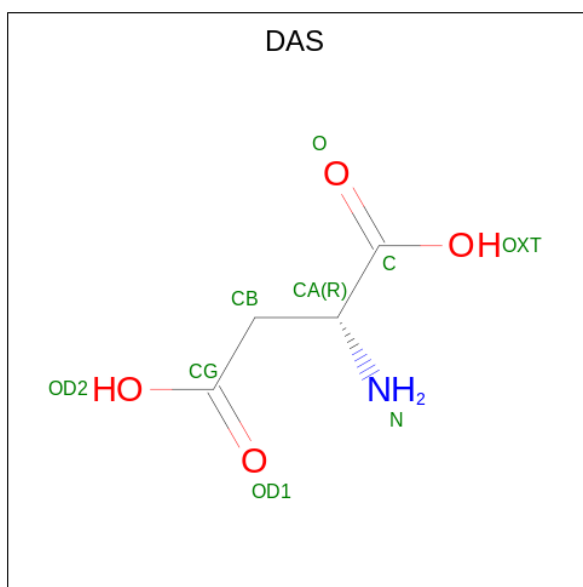
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			8	4	1	3		
3	B	1	Total	C	N	O	0	0
			8	4	1	3		

• Molecule 4 is D-LYSINE (three-letter code: DLY) (formula: $C_6H_{14}N_2O_2$).

DLY
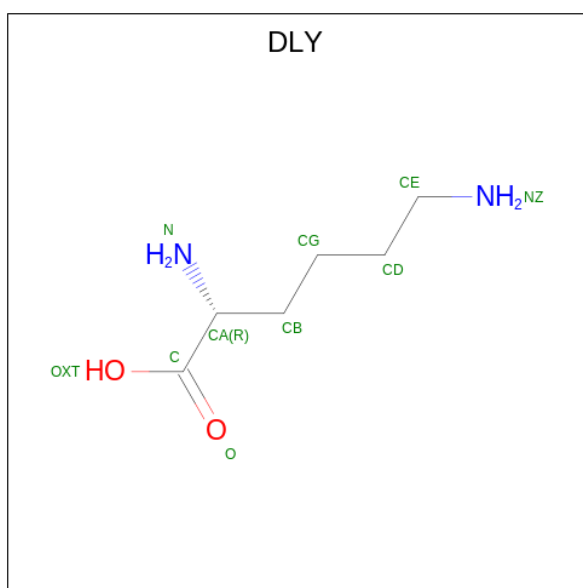
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			10	6	2	2		
4	B	1	Total	C	N	O	0	0
			10	6	2	2		

• Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	165	Total O 165 165		0	0
5	B	169	Total O 169 169		0	0

3 Residue-property plots ⓘ

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.

- Molecule 1: Isoaspartyl dipeptidase
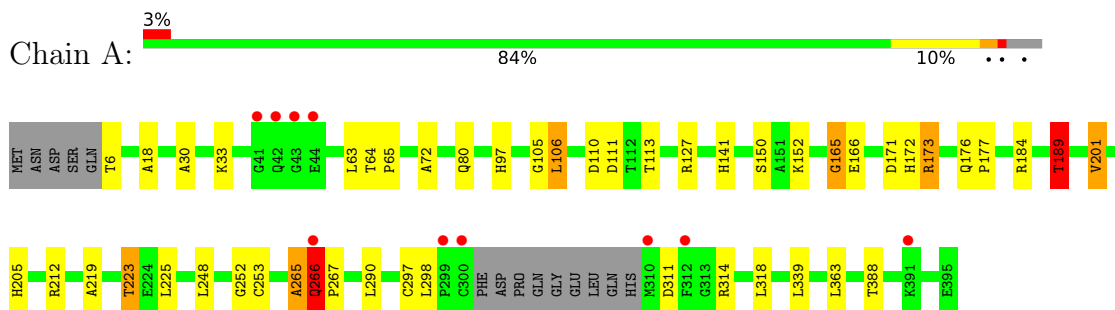
- Molecule 1: Isoaspartyl dipeptidase
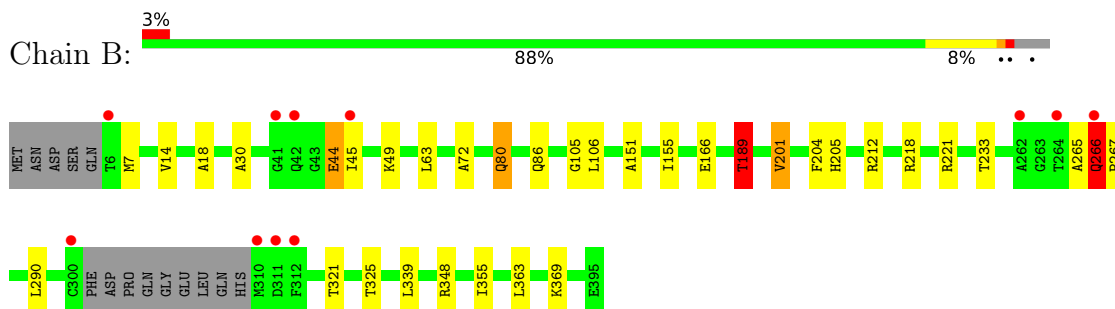
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 4 21 2	Depositor
Cell constants a, b, c, α, β, γ	107.88Å 107.88Å 155.88Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.01 – 2.33 35.35 – 2.33	Depositor EDS
% Data completeness (in resolution range)	95.9 (50.01-2.33) 95.9 (35.35-2.33)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	6.72 (at 2.34Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.189 , 0.246 0.196 , 0.249	Depositor DCC
R_{free} test set	1910 reflections (4.98%)	wwPDB-VP
Wilson B-factor (Å^2)	33.2	Xtriage
Anisotropy	0.001	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 36.5	EDS
L-test for twinning[2]	$<|L|>= 0.51, <L^2>= 0.35$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6154	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.19% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $<|L|>$, $<L^2>$ 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: ZN, DLY, DAS

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	0.81	1/2948 (0.0%)	1.01	12/4004 (0.3%)
1	B	0.82	1/2948 (0.0%)	0.94	7/4004 (0.2%)
All	All	0.81	2/5896 (0.0%)	0.97	19/8008 (0.2%)

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	A	0	3
1	B	0	1
All	All	0	4

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	189	THR	CB-CG2	-5.62	1.33	1.52
1	A	189	THR	CB-CG2	-5.15	1.35	1.52

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	127	ARG	NE-CZ-NH2	13.71	127.16	120.30
1	A	127	ARG	NE-CZ-NH1	-12.75	113.92	120.30
1	A	184	ARG	NE-CZ-NH1	-12.65	113.97	120.30
1	B	212	ARG	NE-CZ-NH1	-12.59	114.00	120.30
1	A	184	ARG	NE-CZ-NH2	11.25	125.92	120.30

There are no chirality outliers.

All (4) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	165	GLY	Mainchain,Peptide
1	A	265	ALA	Peptide
1	B	265	ALA	Peptide

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	2890	0	2881	36	0
1	B	2890	0	2881	18	0
2	A	2	0	0	0	0
2	B	2	0	0	0	0
3	A	8	0	5	1	0
3	B	8	0	5	2	0
4	A	10	0	11	2	0
4	B	10	0	11	1	0
5	A	165	0	0	5	0
5	B	169	0	0	3	0
All	All	6154	0	5794	54	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 5.

The worst 5 of 54 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:290:LEU:HD23	1:B:339:LEU:HD21	1.48	0.93
1:B:266:GLN:HB3	1:B:267:PRO:HD3	1.72	0.72
1:A:248:LEU:O	5:A:501:HOH:O	2.08	0.72
1:A:219:ALA:O	1:A:223:THR:HB	1.90	0.72
1:B:80:GLN:HE22	3:B:403:DAS:HB3	1.55	0.71

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	377/395 (95%)	363 (96%)	10 (3%)	4 (1%)	14	13
1	B	377/395 (95%)	362 (96%)	11 (3%)	4 (1%)	14	13
All	All	754/790 (95%)	725 (96%)	21 (3%)	8 (1%)	14	13

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	266	GLN
1	B	266	GLN
1	A	311	ASP
1	A	265	ALA
1	B	44	GLU

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	311/324 (96%)	305 (98%)	6 (2%)	57	68
1	B	311/324 (96%)	303 (97%)	8 (3%)	46	56
All	All	622/648 (96%)	608 (98%)	14 (2%)	50	61

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	49	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	80	GLN
1	B	369	LYS
1	B	201	VAL
1	B	266	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	80	GLN
1	A	97	HIS
1	A	254	HIS
1	A	351	ASN

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 ⓘ

Of 8 ligands modelled in this entry, 4 are monoatomic - leaving 4 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	$\#|Z| > 2$
4	DLY	A	404	3	8,9,9	1.73	1 (12%)	9,10,10	0.67	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#|Z| > 2$	Counts	RMSZ	$\#|Z| > 2$
4	DLY	B	404	3	8,9,9	1.71	1 (12%)	9,10,10	0.61	0
3	DAS	B	403	2,4	6,7,8	1.22	0	6,8,10	1.51	1 (16%)
3	DAS	A	403	2,4	6,7,8	1.27	1 (16%)	6,8,10	0.50	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
4	DLY	A	404	3	-	3/9/9/9	-
4	DLY	B	404	3	-	4/9/9/9	-
3	DAS	B	403	2,4	-	5/7/7/8	-
3	DAS	A	403	2,4	-	6/7/7/8	-

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	404	DLY	CB-CA	3.23	1.60	1.53
4	A	404	DLY	CB-CA	3.12	1.59	1.53
3	A	403	DAS	OXT-C	-2.34	1.22	1.30

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	403	DAS	O-C-CA	-2.19	114.41	122.14

There are no chirality outliers.

5 of 18 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	403	DAS	O-C-CA-N
3	A	403	DAS	N-CA-CB-CG
3	B	403	DAS	O-C-CA-N
3	B	403	DAS	C-CA-CB-CG
4	A	404	DLY	N-CA-CB-CG

There are no ring outliers.

4 monomers are involved in 6 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	404	DLY	2	0
4	B	404	DLY	1	0
3	B	403	DAS	2	0
3	A	403	DAS	1	0

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, 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	381/395 (96%)	-0.28	10 (2%)	56	64	27, 36, 63, 127	0
1	B	381/395 (96%)	-0.41	11 (2%)	51	61	26, 36, 66, 110	0
All	All	762/790 (96%)	-0.34	21 (2%)	53	63	26, 36, 66, 127	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	310	MET	7.0
1	A	300	CYS	6.5
1	A	42	GLN	5.4
1	B	310	MET	4.9
1	B	300	CYS	3.9

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 monosaccharides 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. 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(Å²)	Q<0.9
2	ZN	B	401	1/1	0.72	0.10	102,102,102,102	0
4	DLY	B	404	10/10	0.74	0.47	87,103,112,126	0
3	DAS	B	403	8/9	0.76	0.19	40,61,79,88	0
4	DLY	A	404	10/10	0.82	0.35	78,99,107,107	0
3	DAS	A	403	8/9	0.88	0.19	37,60,67,68	0
2	ZN	A	401	1/1	0.91	0.11	94,94,94,94	0
2	ZN	A	402	1/1	1.00	0.07	38,38,38,38	0
2	ZN	B	402	1/1	1.00	0.05	39,39,39,39	0

6.5 Other polymers ⓘ

There are no such residues in this entry.