WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 23, 2024 – 09:13 AM EDT

PDB ID	:	1YGZ
Title	:	Crystal Structure of Inorganic Pyrophosphatase from Helicobacter pylori
Authors	:	Wu, C.A.; Lokanath, N.K.; Kim, D.Y.; Park, H.J.; Hwang, H.Y.; Kim, S.T.; Suh, S.W.; Kim, K.K.
Deposited on	:	2005-01-06
Resolution	:	2.60 Å(reported)

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

|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

1   Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.60 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.
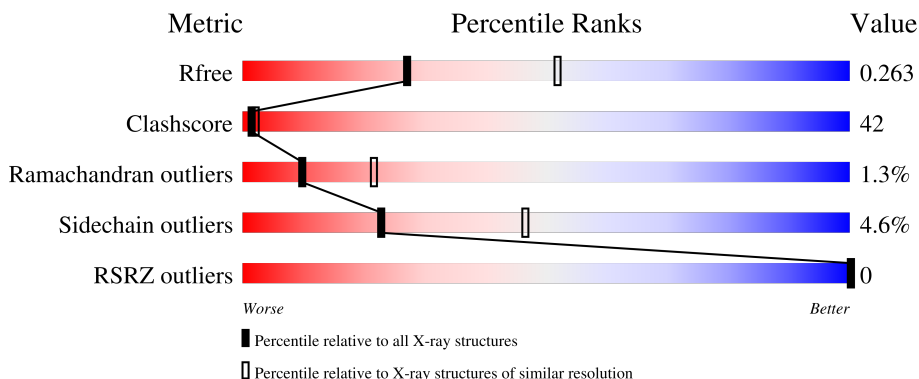
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	164625	3775 (2.60-2.60)
Clashscore	180529	4181 (2.60-2.60)
Ramachandran outliers	177936	4129 (2.60-2.60)
Sidechain outliers	177891	4129 (2.60-2.60)
RSRZ outliers	164620	3775 (2.60-2.60)

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	173	46% 49% 5% ·
1	B	173	40% 57% · · ·
1	C	173	43% 53% · ·
1	D	173	46% 50% · ·
1	E	173	43% 53% · ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	173	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 8552 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 Inorganic pyrophosphatase.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	172	Total	C	N	O	S	Se	0	0	0
			1351	861	219	266	1	4			
1	B	172	Total	C	N	O	S	Se	0	0	0
			1351	861	219	266	1	4			
1	C	172	Total	C	N	O	S	Se	0	0	0
			1351	861	219	266	1	4			
1	D	172	Total	C	N	O	S	Se	0	0	0
			1351	861	219	266	1	4			
1	E	172	Total	C	N	O	S	Se	0	0	0
			1351	861	219	266	1	4			
1	F	172	Total	C	N	O	S	Se	0	0	0
			1351	861	219	266	1	4			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MSE	MET	modified residue	UNP P56153
A	39	MSE	MET	modified residue	UNP P56153
A	94	MSE	MET	modified residue	UNP P56153
A	100	MSE	MET	modified residue	UNP P56153
B	1	MSE	MET	modified residue	UNP P56153
B	39	MSE	MET	modified residue	UNP P56153
B	94	MSE	MET	modified residue	UNP P56153
B	100	MSE	MET	modified residue	UNP P56153
C	1	MSE	MET	modified residue	UNP P56153
C	39	MSE	MET	modified residue	UNP P56153
C	94	MSE	MET	modified residue	UNP P56153
C	100	MSE	MET	modified residue	UNP P56153
D	1	MSE	MET	modified residue	UNP P56153
D	39	MSE	MET	modified residue	UNP P56153
D	94	MSE	MET	modified residue	UNP P56153
D	100	MSE	MET	modified residue	UNP P56153
E	1	MSE	MET	modified residue	UNP P56153

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	39	MSE	MET	modified residue	UNP P56153
E	94	MSE	MET	modified residue	UNP P56153
E	100	MSE	MET	modified residue	UNP P56153
F	1	MSE	MET	modified residue	UNP P56153
F	39	MSE	MET	modified residue	UNP P56153
F	94	MSE	MET	modified residue	UNP P56153
F	100	MSE	MET	modified residue	UNP P56153

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	67	Total 67	O 67	0	0
2	B	64	Total 64	O 64	0	0
2	C	74	Total 74	O 74	0	0
2	D	83	Total 83	O 83	0	0
2	E	72	Total 72	O 72	0	0
2	F	86	Total 86	O 86	0	0

3   Residue-property plots (i)

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.
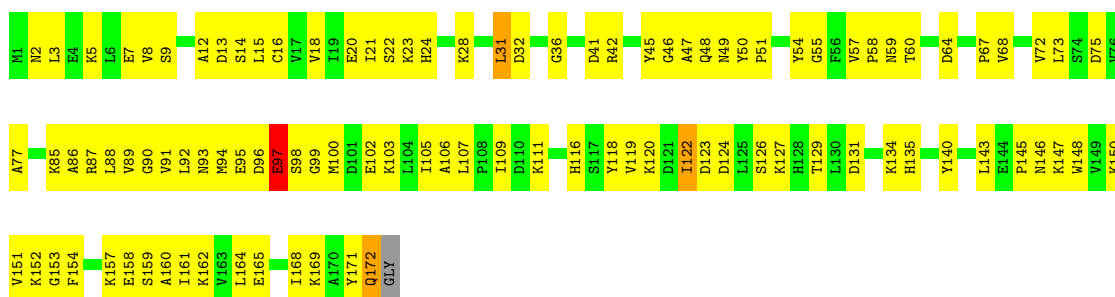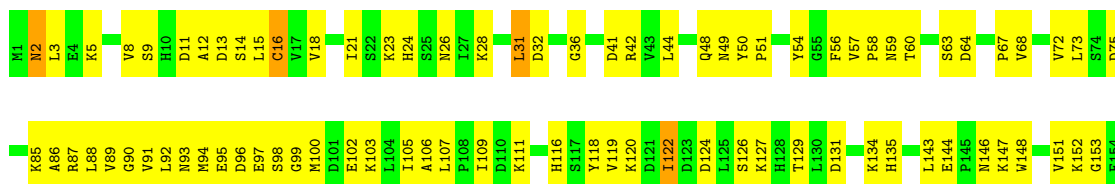
• Molecule 1: Inorganic pyrophosphatase

Chain A:
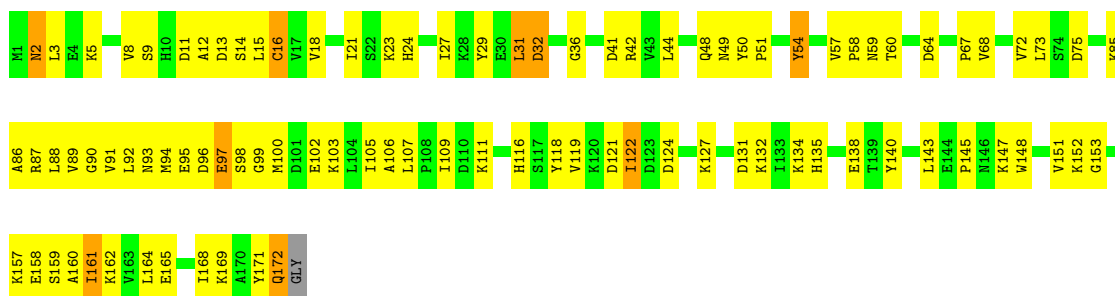
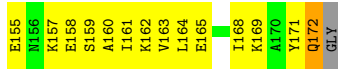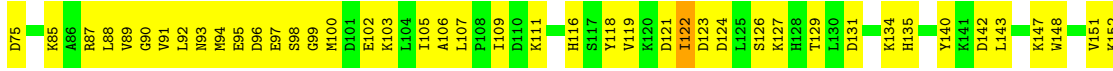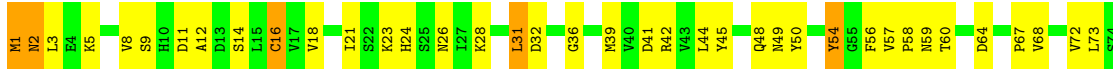• Molecule 1: Inorganic pyrophosphatase
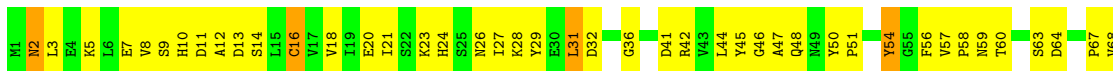
Chain B:

• Molecule 1: Inorganic pyrophosphatase

Chain C:

E155 N156 K157 E158 S159 A160 I161 K162 V163 L164 E165 I168 K169 A170 Y171 Q172 GLY

• Molecule 1: Inorganic pyrophosphatase

Chain D:

46%   50%

M1 N2 L3 E4 K5 V8 S9 H10 D11 A12 D13 S14 L15 C16 V17 V18 I21 S22 K23 H24 S25 N26 I27 K28 L31 D32 G36 M39 V40 D41 R42 V43 L44 Y45 Q48 N49 Y50 Y54 G55 F56 V57 P58 N59 T60 D64 P67 V68 V72 L73 S74

D75 K85 A86 R87 L88 V89 G90 V91 L92 N93 M94 E95 D96 E97 S98 G99 M100 D101 E102 K103 L104 I105 A106 L107 P108 I109 D110 K111 H116 S117 Y118 V119 K120 D121 I122 D123 D124 L125 S126 K127 H128 T129 L130 D131 K134 H135 Y140 K141 D142 L143 K147 W148 V151 K152

G153 K157 E158 S159 A160 I161 K162 V163 L164 E165 I168 K169 A170 Y171 Q172 GLY

• Molecule 1: Inorganic pyrophosphatase

Chain E:

43%   53%

M1 N2 L3 E4 K5 V8 S9 H10 D11 A12 D13 S14 L15 C16 V17 V18 I21 S22 K23 H24 K28 L31 D32 S35 G36 A37 D41 R42 V43 L44 Y45 G46 A47 Q48 N49 Y50 P51 Y54 G55 F56 V57 P58 N59 T60 D64 P67 V68 V72 L73 S74

D75 V76 A77 F78 Q79 A80 K85 A86 R87 L88 V89 G90 V91 L92 N93 M94 E95 D96 E97 S98 G99 M100 D101 E102 K103 L104 I105 A106 L107 P108 I109 D110 K111 H116 S117 Y118 V119 I122 D123 D124 K127 H128 D131 K132 I133 K134 H135 Y140 E144 K147 W148

V151 K152 G153 K157 E158 S159 A160 I161 K162 V163 L164 E165 I168 K169 A170 Y171 Q172 GLY

• Molecule 1: Inorganic pyrophosphatase

Chain F:

37%   58%

M1 N2 L3 E4 K5 L6 E7 V8 S9 H10 D11 A12 D13 S14 L15 C16 V17 V18 I19 E20 I21 S22 K23 H24 S25 N26 I27 K28 Y29 E30 L31 D32 G36 D41 R42 V43 L44 Y45 G46 A47 Q48 N49 Y50 P51 Y54 G55 F56 V57 P58 N59 T60 S63 D64 P67 V68

V72 L73 S74 D75 Q79 V84 K85 A86 R87 L88 V89 G90 V91 L92 N93 M94 E95 D96 E97 S98 G99 M100 D101 E102 I105 A106 L107 P108 I109 D110 K111 H116 S117 Y118 V119 K120 D121 I122 D123 D124 L125 S126 K127 H128 T129 L130 D131 K132 I133 K134 H135 F136 F137 E138 T139

Y140 K141 D142 L143 E144 K147 W148 V151 K152 G153 K157 E158 S159 A160 I161 K162 V163 L164 E165 I168 K169 A170 Y171 Q172 GLY

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 3	Depositor
Cell constants a, b, c, α, β, γ	115.06Å 115.06Å 73.78Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	20.00 – 2.60 20.00 – 2.60	Depositor EDS
% Data completeness (in resolution range)	98.5 (20.00-2.60) 98.5 (20.00-2.60)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	9.47 (at 2.61Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.206 , 0.265 0.204 , 0.263	Depositor DCC
R_{free} test set	3117 reflections (9.44%)	wwPDB-VP
Wilson B-factor (Å^2)	14.7	Xtriage
Anisotropy	0.025	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 67.1	EDS
L-test for twinning[2]	$< |L| > = 0.34, < L^2 > = 0.17$	Xtriage
Estimated twinning fraction	0.100 for -h,-k,l 0.499 for h,-h-k,-l 0.100 for -k,-h,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	8552	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.93% 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 (i)

5.1   Standard geometry (i)

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	0.44	0/1372	0.63	0/1848
1	B	0.48	0/1372	0.64	1/1848 (0.1%)
1	C	0.45	0/1372	0.61	0/1848
1	D	0.49	0/1372	0.62	1/1848 (0.1%)
1	E	0.46	0/1372	0.62	0/1848
1	F	0.53	2/1372 (0.1%)	0.64	1/1848 (0.1%)
All	All	0.47	2/8232 (0.0%)	0.63	3/11088 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	97	GLU	CD-OE1	-8.62	1.16	1.25
1	F	97	GLU	CD-OE2	-6.29	1.18	1.25

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	F	97	GLU	OE1-CD-OE2	-8.22	113.44	123.30
1	D	1	MSE	CG-SE-CE	5.96	112.00	98.90
1	B	97	GLU	OE1-CD-OE2	-5.02	117.27	123.30

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1351	0	1354	115	0
1	B	1351	0	1354	119	0
1	C	1351	0	1354	139	0
1	D	1351	0	1354	120	0
1	E	1351	0	1354	113	1
1	F	1351	0	1354	129	0
2	A	67	0	0	5	0
2	B	64	0	0	4	0
2	C	74	0	0	9	0
2	D	83	0	0	10	0
2	E	72	0	0	5	0
2	F	86	0	0	10	0
All	All	8552	0	8124	686	1

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 42.

The worst 5 of 686 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:97:GLU:OE2	1:D:147:LYS:HG3	1.54	1.05
1:E:95:GLU:HG2	1:E:152:LYS:HE2	1.42	1.01
1:A:152:LYS:HG3	1:C:127:LYS:HE3	1.39	1.01
1:D:95:GLU:HG2	1:D:152:LYS:HE2	1.45	0.99
1:F:95:GLU:HG2	1:F:152:LYS:HE2	1.45	0.98

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:45:TYR:OH	1:E:111:LYS:O[3_675]	2.19	0.01

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	170/173 (98%)	142 (84%)	25 (15%)	3 (2%)	7	14
1	B	170/173 (98%)	141 (83%)	27 (16%)	2 (1%)	11	24
1	C	170/173 (98%)	142 (84%)	26 (15%)	2 (1%)	11	24
1	D	170/173 (98%)	144 (85%)	24 (14%)	2 (1%)	11	24
1	E	170/173 (98%)	141 (83%)	27 (16%)	2 (1%)	11	24
1	F	170/173 (98%)	144 (85%)	24 (14%)	2 (1%)	11	24
All	All	1020/1038 (98%)	854 (84%)	153 (15%)	13 (1%)	10	21

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	64	ASP
1	B	64	ASP
1	D	64	ASP
1	E	64	ASP
1	F	64	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	151/147 (103%)	141 (93%)	10 (7%)	14	30
1	B	151/147 (103%)	145 (96%)	6 (4%)	27	52
1	C	151/147 (103%)	146 (97%)	5 (3%)	33	59
1	D	151/147 (103%)	143 (95%)	8 (5%)	19	40
1	E	151/147 (103%)	145 (96%)	6 (4%)	27	52
1	F	151/147 (103%)	144 (95%)	7 (5%)	23	46
All	All	906/882 (103%)	864 (95%)	42 (5%)	23	46

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	172	GLN
1	F	2	ASN
1	E	2	ASN
1	E	75	ASP
1	F	31	LEU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 22 such sidechains are listed below:

Mol	Chain	Res	Type
1	E	10	HIS
1	E	172	GLN
1	E	128	HIS
1	F	10	HIS
1	B	172	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 oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

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(Å²)	Q<0.9
1	A	168/173 (97%)	-1.77	0	100	100	1, 16, 38, 48	0
1	B	168/173 (97%)	-1.85	0	100	100	0, 15, 35, 53	0
1	C	168/173 (97%)	-1.73	0	100	100	1, 25, 40, 52	0
1	D	168/173 (97%)	-1.76	0	100	100	1, 20, 38, 46	0
1	E	168/173 (97%)	-1.78	0	100	100	1, 18, 37, 57	0
1	F	168/173 (97%)	-1.80	0	100	100	0, 16, 34, 52	0
All	All	1008/1038 (97%)	-1.78	0	100	100	0, 18, 38, 57	0

There are no RSRZ outliers to report.

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 ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.