



wwPDB X-ray Structure Validation Summary Report

May 23, 2020 – 07:50 am BST

PDB ID : 5U2G
Title : 2.6 Angstrom Resolution Crystal Structure of Penicillin-Binding Protein 1A from Haemophilus influenzae
Authors : Minasov, G.; Wawrzak, Z.; Shuvalova, L.; Kiryukhina, O.; Dubrovskaya, I.; Grimshaw, S.; Kwon, K.; Anderson, W.F.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on : 2016-11-30
Resolution : 2.61 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 : 2.11
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) : 2.11

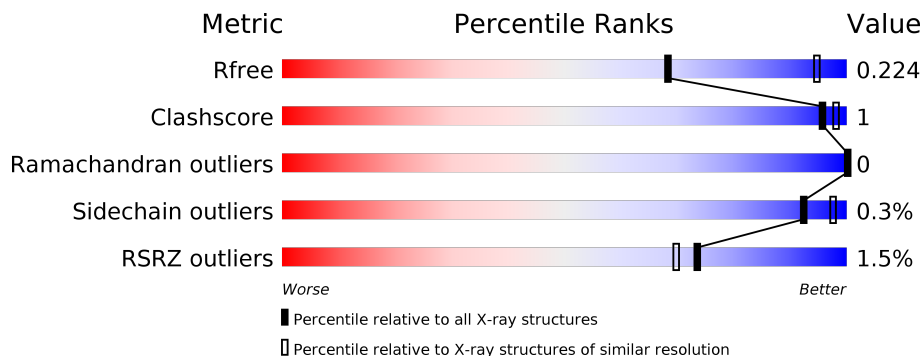
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.61 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	3797 (2.64-2.60)
Clashscore	141614	4168 (2.64-2.60)
Ramachandran outliers	138981	4093 (2.64-2.60)
Sidechain outliers	138945	4093 (2.64-2.60)
RSRZ outliers	127900	3731 (2.64-2.60)

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 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	838	 % 77% • 20%
1	B	838	 % 78% • 20%

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
5	PEG	A	914	-	-	-	X

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 10912 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 Penicillin-binding protein 1A.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	668	5242	3331	907	988	2	14	0	3	0
1	B	670	5233	3329	903	985	2	14	0	0	0

There are 30 discrepancies between the modelled and reference sequences:

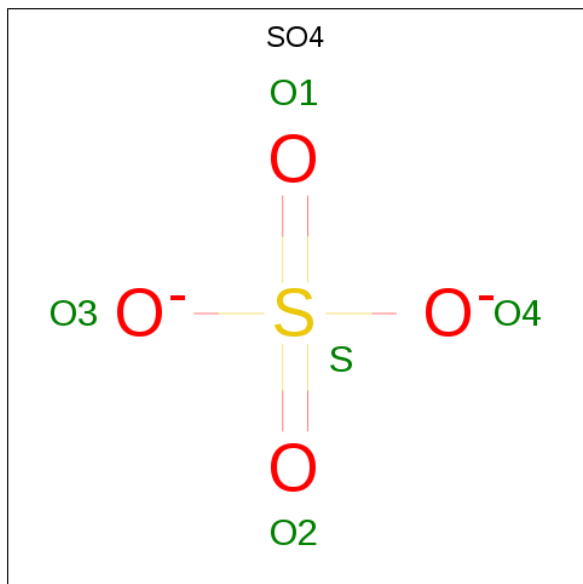
Chain	Residue	Modelled	Actual	Comment	Reference
A	27	SER	-	expression tag	UNP P31776
A	28	ASN	-	expression tag	UNP P31776
A	29	ALA	-	expression tag	UNP P31776
A	34	PRO	-	VARIANT	UNP P31776
A	35	SER	-	VARIANT	UNP P31776
A	36	VAL	-	VARIANT	UNP P31776
A	37	GLU	-	VARIANT	UNP P31776
A	38	THR	-	VARIANT	UNP P31776
A	39	LEU	-	VARIANT	UNP P31776
A	40	LYS	-	VARIANT	UNP P31776
A	41	THR	-	VARIANT	UNP P31776
A	42	VAL	-	VARIANT	UNP P31776
A	43	GLU	-	VARIANT	UNP P31776
A	44	LEU	-	VARIANT	UNP P31776
B	27	SER	-	expression tag	UNP P31776
B	28	ASN	-	expression tag	UNP P31776
B	29	ALA	-	expression tag	UNP P31776
B	32	GLU	-	VARIANT	UNP P31776
B	33	LEU	-	VARIANT	UNP P31776
B	34	PRO	-	VARIANT	UNP P31776
B	35	SER	-	VARIANT	UNP P31776
B	36	VAL	-	VARIANT	UNP P31776
B	37	GLU	-	VARIANT	UNP P31776
B	38	THR	-	VARIANT	UNP P31776
B	39	LEU	-	VARIANT	UNP P31776

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	40	LYS	-	VARIANT	UNP P31776
B	41	THR	-	VARIANT	UNP P31776
B	42	VAL	-	VARIANT	UNP P31776
B	43	GLU	-	VARIANT	UNP P31776
B	44	LEU	-	VARIANT	UNP P31776

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	S		
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	A	1	5	4	1	0	0
2	B	1	5	4	1	0	0

Continued on next page...

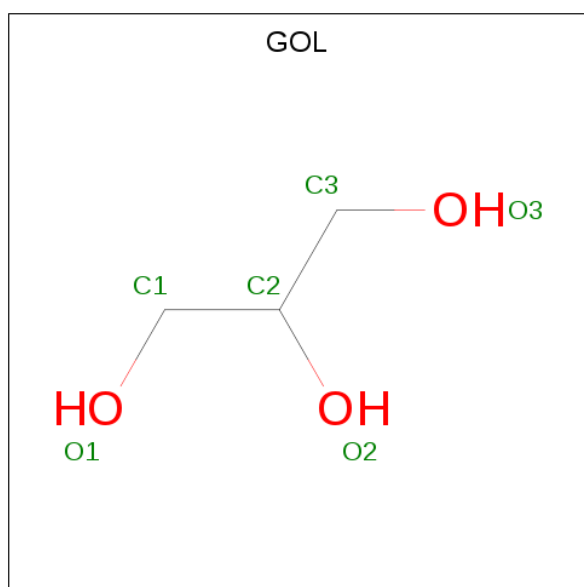
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0
2	B	1	Total O S 5 4 1	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

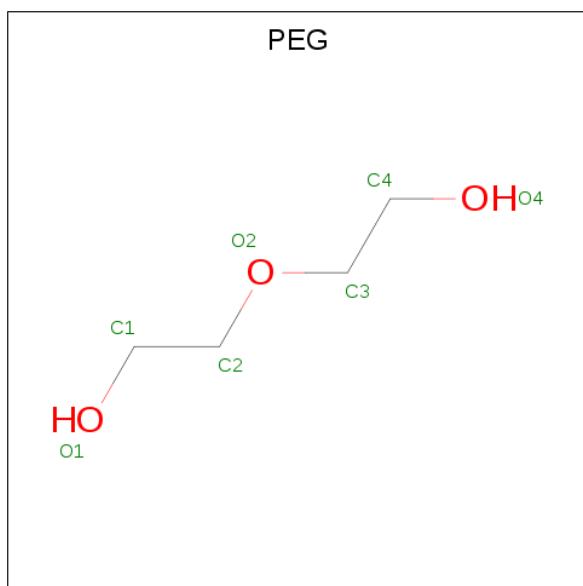
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	B	5	Total Cl 5 5	0	0
3	A	4	Total Cl 4 4	0	0

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	A	1	Total	C	O	0	0
			6	3	3		

- Molecule 5 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: C₄H₁₀O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total	C	O	0	0
			7	4	3		

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	159	Total	O	0	4
			163	163		
6	B	170	Total	O	0	2
			172	172		

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	87.45Å 111.93Å 107.69Å 90.00° 90.40° 90.00°	Depositor
Resolution (Å)	29.15 – 2.61 29.15 – 2.61	Depositor EDS
% Data completeness (in resolution range)	98.9 (29.15-2.61) 97.2 (29.15-2.61)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.08	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.61 (at 2.61Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.179 , 0.223 0.186 , 0.224	Depositor DCC
R_{free} test set	3003 reflections (4.78%)	wwPDB-VP
Wilson B-factor (Å ²)	63.0	Xtriage
Anisotropy	0.180	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 29.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.014 for -h,-l,-k 0.000 for -h,l,k 0.397 for h,-k,-l	Xtriage
F_o, F_c correlation	0.96	EDS
Total number of atoms	10912	wwPDB-VP
Average B, all atoms (Å ²)	77.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.29% of the height of the origin peak. No significant pseudotranslation is detected.*

¹ Intensities estimated from amplitudes.

² Theoretical 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: GOL, PEG, SO4, CL

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	0.45	0/5329	0.68	0/7176
1	B	0.44	0/5321	0.68	0/7166
All	All	0.45	0/10650	0.68	0/14342

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	5242	0	5264	13	0
1	B	5233	0	5259	10	0
2	A	40	0	0	0	0
2	B	40	0	0	0	0
3	A	4	0	0	0	0
3	B	5	0	0	0	0
4	A	6	0	8	0	0
5	A	7	0	10	0	0
6	A	163	0	0	0	0
6	B	172	0	0	0	0
All	All	10912	0	10541	23	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 1.

The worst 5 of 23 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:A:358:VAL:HG12	1:A:359:LYS:N	2.24	0.52
1:B:421:SER:HB3	1:B:568:ASN:HB3	1.91	0.52
1:A:421:SER:HB3	1:A:568:ASN:HB3	1.93	0.50
1:A:70:LYS:HA	1:A:70:LYS:HE2	1.95	0.49
1:B:442:ARG:HA	1:B:446:SER:HB2	1.95	0.48

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	663/838 (79%)	642 (97%)	21 (3%)	0	100	100
1	B	662/838 (79%)	636 (96%)	26 (4%)	0	100	100
All	All	1325/1676 (79%)	1278 (96%)	47 (4%)	0	100	100

There are no Ramachandran outliers to report.

5.3.2 Protein sidechains [i](#)

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	550/671 (82%)	548 (100%)	2 (0%)	91	97
1	B	548/671 (82%)	547 (100%)	1 (0%)	93	98
All	All	1098/1342 (82%)	1095 (100%)	3 (0%)	92	97

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	770	LYS
1	A	791	LYS
1	B	770	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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 carbohydrates in this entry.

5.6 Ligand geometry [i](#)

Of 27 ligands modelled in this entry, 9 are monoatomic - leaving 18 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
2	SO4	A	906	-	4,4,4	0.35	0	6,6,6	0.11	0
2	SO4	B	906	-	4,4,4	0.34	0	6,6,6	0.06	0
2	SO4	A	902	-	4,4,4	0.35	0	6,6,6	0.10	0
2	SO4	B	902	-	4,4,4	0.32	0	6,6,6	0.12	0
2	SO4	B	908	-	4,4,4	0.35	0	6,6,6	0.16	0
2	SO4	B	905	-	4,4,4	0.35	0	6,6,6	0.10	0
2	SO4	A	903	-	4,4,4	0.34	0	6,6,6	0.10	0
2	SO4	B	901	-	4,4,4	0.34	0	6,6,6	0.08	0
5	PEG	A	914	-	6,6,6	0.50	0	5,5,5	0.23	0
2	SO4	A	904	-	4,4,4	0.37	0	6,6,6	0.06	0
2	SO4	A	907	-	4,4,4	0.35	0	6,6,6	0.06	0
2	SO4	B	907	-	4,4,4	0.35	0	6,6,6	0.07	0
4	GOL	A	913	-	5,5,5	0.34	0	5,5,5	0.20	0
2	SO4	A	905	-	4,4,4	0.39	0	6,6,6	0.13	0
2	SO4	B	903	-	4,4,4	0.34	0	6,6,6	0.14	0
2	SO4	B	904	-	4,4,4	0.36	0	6,6,6	0.12	0
2	SO4	A	901	-	4,4,4	0.32	0	6,6,6	0.10	0
2	SO4	A	908	-	4,4,4	0.33	0	6,6,6	0.05	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	GOL	A	913	-	-	0/4/4/4	-
5	PEG	A	914	-	-	1/4/4/4	-

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
5	A	914	PEG	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	654/838 (78%)	-0.30	9 (1%) 75 71	46, 69, 133, 164	0
1	B	656/838 (78%)	-0.30	11 (1%) 70 66	49, 69, 135, 168	0
All	All	1310/1676 (78%)	-0.30	20 (1%) 73 70	46, 69, 133, 168	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	675	SER	7.6
1	B	252	TYR	4.8
1	A	245	LYS	4.5
1	B	247	PRO	3.8
1	B	248	ILE	3.3

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 carbohydrates 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
5	PEG	A	914	7/7	0.77	0.57	110,114,117,120	0
4	GOL	A	913	6/6	0.79	0.14	88,101,102,102	0
2	SO4	B	906	5/5	0.79	0.28	77,83,89,90	5
2	SO4	B	903	5/5	0.83	0.18	120,127,130,138	0
3	CL	B	909	1/1	0.84	0.19	69,69,69,69	1
2	SO4	B	907	5/5	0.85	0.16	81,81,83,89	5
2	SO4	A	907	5/5	0.87	0.19	80,81,86,88	5
3	CL	A	909	1/1	0.88	0.13	96,96,96,96	0
2	SO4	A	905	5/5	0.89	0.11	109,111,116,130	0
3	CL	B	913	1/1	0.89	0.20	95,95,95,95	0
2	SO4	B	908	5/5	0.90	0.19	54,57,58,62	5
2	SO4	A	902	5/5	0.91	0.07	103,111,122,125	0
2	SO4	B	901	5/5	0.91	0.14	116,120,126,130	0
2	SO4	B	904	5/5	0.91	0.18	101,105,115,126	0
2	SO4	A	908	5/5	0.92	0.14	75,76,79,80	5
2	SO4	A	904	5/5	0.93	0.09	96,112,122,123	0
2	SO4	A	903	5/5	0.94	0.15	105,107,117,119	0
2	SO4	B	902	5/5	0.94	0.07	103,109,112,113	0
2	SO4	B	905	5/5	0.94	0.09	100,111,118,119	0
2	SO4	A	906	5/5	0.95	0.11	99,102,109,110	0
2	SO4	A	901	5/5	0.95	0.10	115,118,120,121	0
3	CL	B	910	1/1	0.96	0.12	75,75,75,75	0
3	CL	A	911	1/1	0.97	0.08	75,75,75,75	0
3	CL	A	910	1/1	0.97	0.15	69,69,69,69	0
3	CL	B	912	1/1	0.97	0.11	62,62,62,62	0
3	CL	B	911	1/1	0.98	0.11	70,70,70,70	0
3	CL	A	912	1/1	0.99	0.16	72,72,72,72	0

6.5 Other polymers [i](#)

There are no such residues in this entry.