



wwPDB X-ray Structure Validation Summary Report ⓘ

May 15, 2020 – 03:50 pm BST

PDB ID : 5XYR
Title : Crystal structure of a serine protease from Streptococcus species
Authors : Jobichen, C.; Sivaraman, J.
Deposited on : 2017-07-10
Resolution : 2.80 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

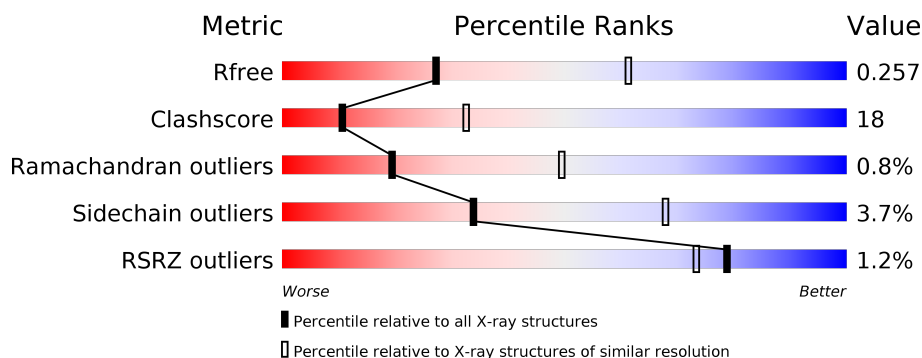
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (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 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	1647	<div> <div></div> <div>57%</div> <div>25%</div> <div>16%</div> </div>

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
2	CL	A	1701	-	-	X	-

2 Entry composition [i](#)

There are 6 unique types of molecules in this entry. The entry contains 10518 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 Chemokine protease C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1380	Total	C	N	O	Se	0	0	0
			10481	6584	1799	2074	24			

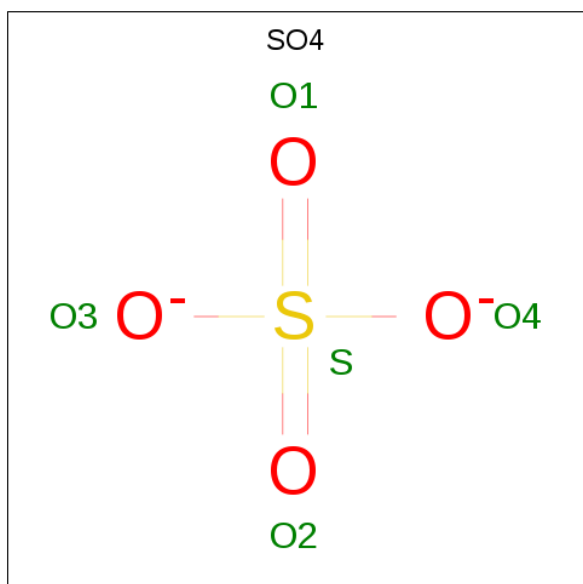
- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	1	Total	Cl	0	0
			1	1		

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

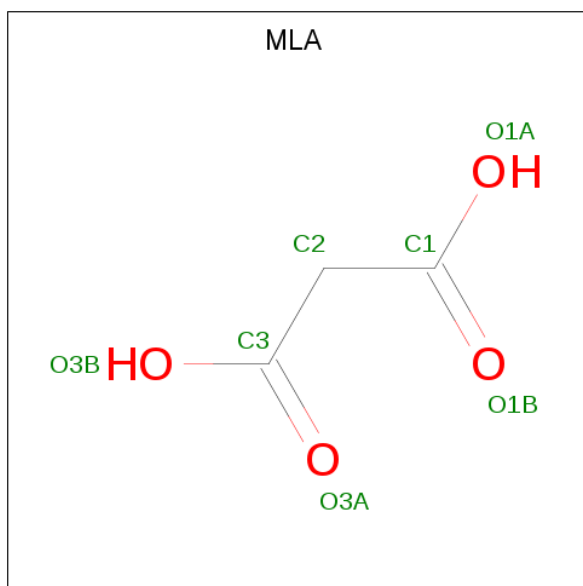
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	3	Total	Ca	0	0
			3	3		

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

- Molecule 5 is MALONIC ACID (three-letter code: MLA) (formula: $C_3H_4O_4$).



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

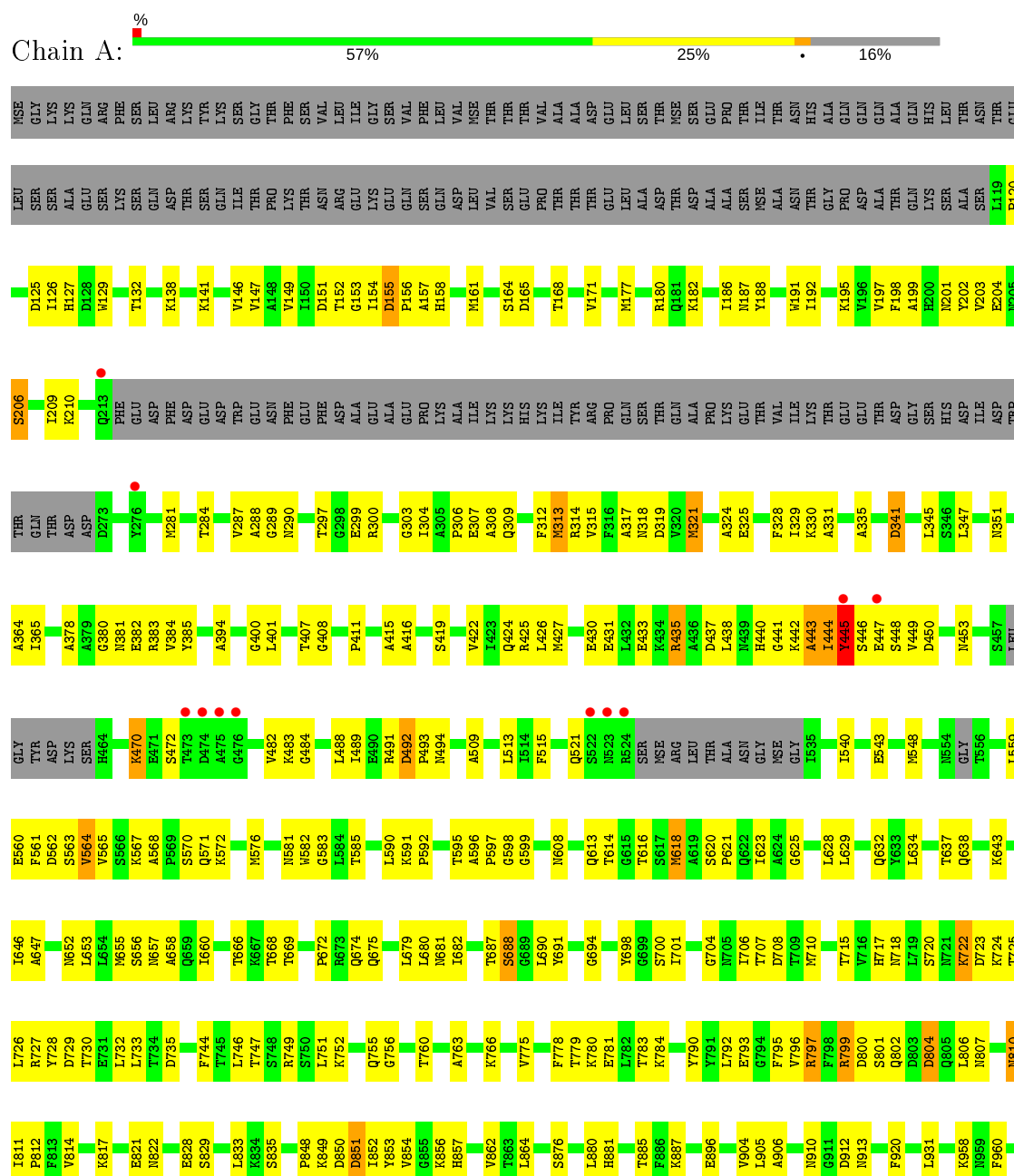
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	21	Total	O	0	0
			21	21		

3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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: Chemokine protease C



ARG	S969	L1094	M1257	K1392	T1560
ILE	Y994	F1095	M1268	A1393	L1561
VAL	Y995	L1096	R1269	V1394	P1562
GLY	Y996	M1107	D1273	L1407	L1570
LEU	Y999	V1108	E1274	R1408	R1571
VAL		E1109		V1409	L1572
LEU		D1110		P1410	V1573
LEU		F1111		E1411	K1574
GLY	G1004	A1112	I1281	D1412	VAL
LEU	R1007	V1115	D1285	I1415	THR
THR		A1116	K1286	E1453	CYS
CYS	M1010	I1117	P1288	L1454	ALA
VAL	T1011	A1118	M1289	T1456	VAL
PHE	F1012	L1124	I1291	Y1457	SER
SER	D1013	P1125	Q1292	A1461	ASP
ARG	M1014	G1129	R1293	A1462	HIS
LYS	D1017	I1133	G1294	K1463	LYS
SER	K1018	K1134	D1302	L1464	VAL
THR	Q1019	L1135	H1303	I1469	MSE
LYS	V1022	K1136	D1307	D1477	SER
ASP	L1023	L1152	K1310	M1487	ASN
LYS	S1024	Q1156	S1314	T1488	ASN
ASP	Q1025	T1159	I1317	M1489	GLN
THR	F1028	G1160	V1318	S1493	ALA
LEU	N1033	L1161	R1319	Q1494	LEU
THR	E1038	L1180	E1321	D1500	THR
ALA	K1041	M1183	Y1324	L1503	PRO
SER	D1042	D1186	K1339	A1514	THR
THR	R1043	F1187	K1353	Q1515	LYS
THR	A1046	F1188	Y1359	Q1516	ALA
LYS	R1049	K1197	T1360	Q1517	THR
THR	K1050	V1200	D1365	Y1526	ALA
THR	D1051	Y1210	D1370	V1527	ALA
ALA	F1054	D1212	S1370	K1533	LYS
THR	R1058	M1216	V1376	G1538	LEU
LEU	D1060	S1231	E1377	V1544	PRO
LEU	Y1074	Q1233	D1378	L1546	SER
THR	V1077	A1234	R1379	Y1550	THR
LYS	E1078	G1235	V1383	R1551	GLY
GLY	D1079	A1236	I1389	I1552	THR
LYS	N1080	G1237		E1553	GLU
LYS	K1081	A1238			LYS
THR	E1085				MSE
LEU	R1086				GLY
LEU					LEU
LEU					LYS
LEU					LEU

4 Data and refinement statistics

Property	Value	Source
Space group	P 62 2 2	Depositor
Cell constants a, b, c, α , β , γ	191.62Å 191.62Å 250.96Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.99 – 2.80 19.99 – 2.80	Depositor EDS
% Data completeness (in resolution range)	99.3 (19.99-2.80) 99.3 (19.99-2.80)	Depositor EDS
R_{merge}	0.21	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.77 (at 2.79Å)	Xtriage
Refinement program	PHENIX (dev_2733: ???)	Depositor
R, R_{free}	0.209 , 0.257 0.211 , 0.257	Depositor DCC
R_{free} test set	1956 reflections (2.94%)	wwPDB-VP
Wilson B-factor (Å ²)	65.5	Xtriage
Anisotropy	0.035	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.23 , 20.7	EDS
L-test for twinning ²	$\langle L \rangle = 0.45$, $\langle L^2 \rangle = 0.27$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	10518	wwPDB-VP
Average B, all atoms (Å ²)	68.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.59% 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, MLA, 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	0.57	0/10657	0.72	2/14404 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	445	TYR	CB-CA-C	-5.47	99.45	110.40
1	A	155	ASP	N-CA-C	-5.03	97.43	111.00

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.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	10481	0	10083	378	0
2	A	1	0	0	9	0
3	A	3	0	0	0	0
4	A	5	0	0	0	0
5	A	7	0	2	0	0
6	A	21	0	0	3	0
All	All	10518	0	10085	378	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 18.

The worst 5 of 378 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:383:ARG:NH2	1:A:746:LEU:CD1	1.77	1.45
1:A:383:ARG:NH2	1:A:746:LEU:HD12	1.06	1.36
1:A:447:GLU:OE2	1:A:453:ASN:CB	1.74	1.34
1:A:445:TYR:HE1	1:A:561:PHE:CE1	1.44	1.34
1:A:383:ARG:CZ	1:A:746:LEU:CD1	2.20	1.19

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	1370/1647 (83%)	1206 (88%)	153 (11%)	11 (1%)	19	49

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	482	VAL
1	A	540	ILE
1	A	564	VAL
1	A	1412	ASP
1	A	449	VAL

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	1104/1369 (81%)	1063 (96%)	41 (4%)	34 68

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	802	GLN
1	A	887	LYS
1	A	1477	ASP
1	A	810	ASN
1	A	835	SER

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

Mol	Chain	Res	Type
1	A	351	ASN
1	A	571	GLN
1	A	929	GLN
1	A	1494	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 carbohydrates in this entry.

5.6 Ligand geometry ⓘ

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	$\# Z > 2$
4	SO4	A	1705	-	4,4,4	0.47	0	6,6,6	0.94	0
5	MLA	A	1706	-	0,6,6	0.00	-	0,7,7	0.00	-

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
5	MLA	A	1706	-	-	0/0/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

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	1355/1647 (82%)	-0.36	16 (1%) 79 73	35, 65, 118, 148	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	1237	GLY	4.6
1	A	473	THR	4.6
1	A	447	GLU	4.0
1	A	522	SER	3.9
1	A	474	ASP	3.7

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(Å ²)	Q<0.9
3	CA	A	1703	1/1	0.79	0.24	91,91,91,91	0
5	MLA	A	1706	7/7	0.89	0.27	57,64,69,71	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
4	SO4	A	1705	5/5	0.90	0.25	66,71,88,105	0
3	CA	A	1704	1/1	0.94	0.11	51,51,51,51	0
3	CA	A	1702	1/1	0.99	0.07	44,44,44,44	0
2	CL	A	1701	1/1	1.00	0.20	45,45,45,45	0

6.5 Other polymers [i](#)

There are no such residues in this entry.