Full wwPDB X-ray Structure Validation Report (i)

Sep 20, 2023 – 09:33 PM EDT

PDB ID	:	5JD8
Title	:	Crystal structure of the serine endoprotease from Yersinia pestis
Authors	:	Filippova, E.V.; Wawrzsak, Z.; Sandoval, J.; Skarina, T.; Grimshaw, S.; Savchenko, A.; Anderson, W.F.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on	:	2016-04-15
Resolution	:	1.85 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 (i) 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 (i)) were used in the production of this report:

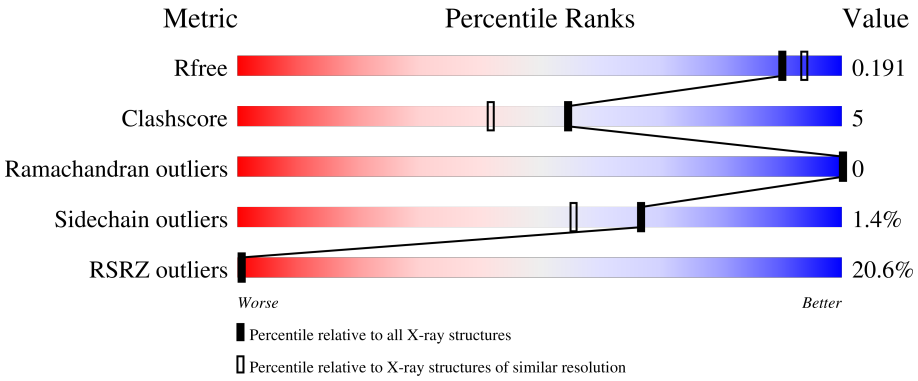
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 1.85 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	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	334	17% 73% 8% · 19%

2 Entry composition

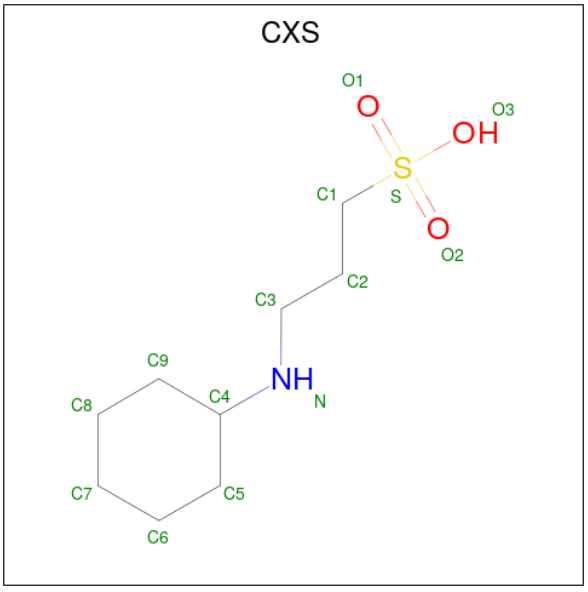
There are 5 unique types of molecules in this entry. The entry contains 2218 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Periplasmic serine peptidase DegS.

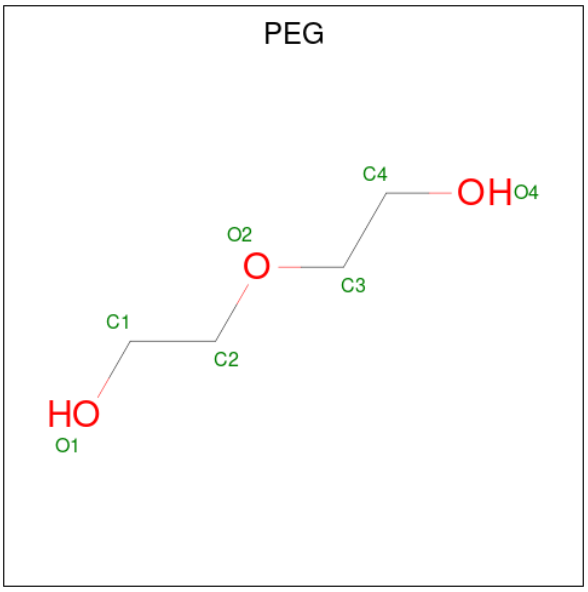
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	272	2036	1278	361	392	5	0	2	0

- Molecule 2 is 3-CYCLOHEXYL-1-PROPYLSULFONIC ACID (three-letter code: CXS) (formula: $C_9H_{19}NO_3S$).



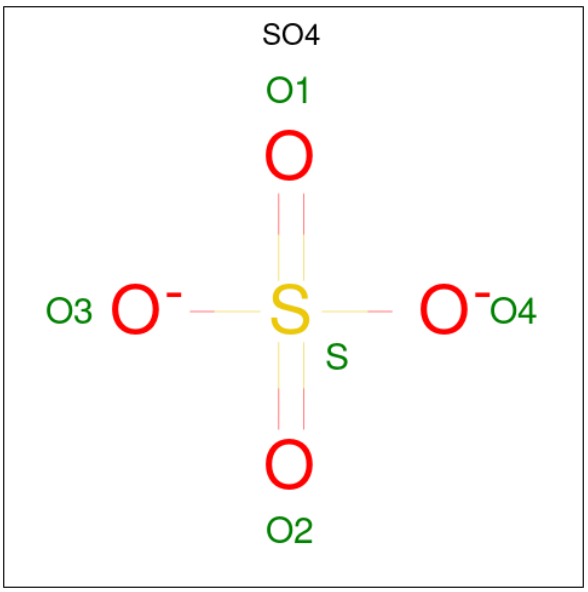
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	S		
2	A	1	14	9	1	3	1	0	0

- Molecule 3 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



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

- Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



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

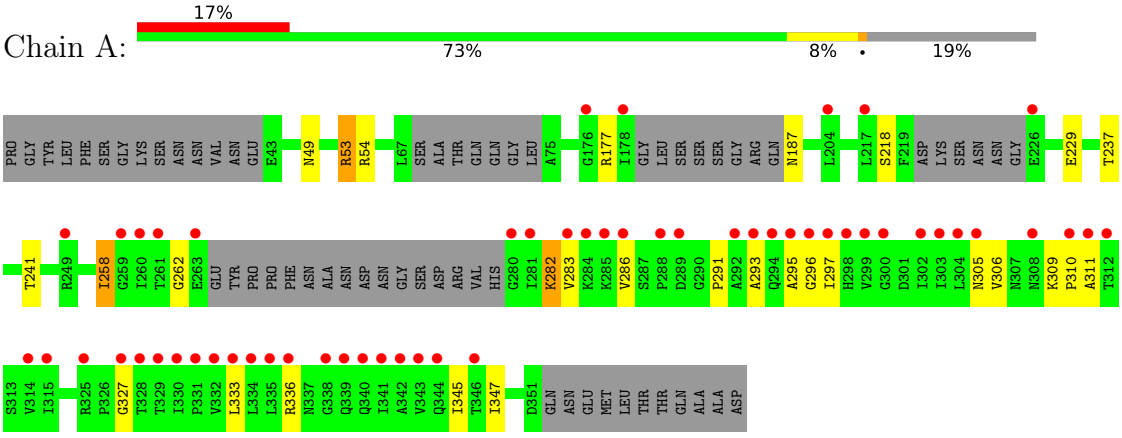
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	154	Total	O	0	2
			156	156		

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: Periplasmic serine peptidase DegS



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	104.42Å 104.42Å 76.02Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 1.85 29.09 – 1.85	Depositor EDS
% Data completeness (in resolution range)	99.9 (30.00-1.85) 99.9 (29.09-1.85)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.29 (at 1.85Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.148 , 0.183 0.161 , 0.191	Depositor DCC
R_{free} test set	1292 reflections (4.88%)	wwPDB-VP
Wilson B-factor (Å^2)	20.8	Xtriage
Anisotropy	0.121	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 80.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.032 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	2218	wwPDB-VP
Average B, all atoms (Å^2)	46.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.62% 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: CXS, SO4, PEG

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	1.01	2/2055 (0.1%)	1.00	5/2793 (0.2%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	310	PRO	N-CD	5.27	1.55	1.47
1	A	229	GLU	CD-OE2	-5.07	1.20	1.25

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	53[A]	ARG	NE-CZ-NH1	6.93	123.77	120.30
1	A	53[B]	ARG	NE-CZ-NH1	6.93	123.77	120.30
1	A	177	ARG	NE-CZ-NH1	-6.79	116.91	120.30
1	A	309	LYS	C-N-CD	5.59	140.15	128.40
1	A	54	ARG	NE-CZ-NH1	5.08	122.84	120.30

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	2036	0	2115	22	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	A	14	0	18	0	0
3	A	7	0	10	0	0
4	A	5	0	0	0	0
5	A	156	0	0	1	0
All	All	2218	0	2143	22	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.

All (22) 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:258:ILE:CD1	1:A:347:ILE:CD1	2.63	0.77
1:A:258:ILE:HD13	1:A:347:ILE:CD1	2.21	0.71
1:A:291:PRO:O	1:A:295:ALA:HB2	1.91	0.70
1:A:286:VAL:HG11	1:A:293:ALA:HB2	1.80	0.64
1:A:297:ILE:CD1	1:A:345:ILE:HD11	2.33	0.58
1:A:286:VAL:CG1	1:A:293:ALA:HB2	2.35	0.56
1:A:49:ASN:OD1	1:A:53[A]:ARG:HD3	2.05	0.56
1:A:258:ILE:HD13	1:A:347:ILE:HD12	1.88	0.55
1:A:258:ILE:CD1	1:A:347:ILE:HD13	2.38	0.53
1:A:283:VAL:O	1:A:283:VAL:HG12	2.09	0.52
1:A:296:GLY:O	1:A:336:ARG:NH2	2.42	0.51
1:A:297:ILE:HD11	1:A:345:ILE:HD11	1.95	0.48
1:A:187:ASN:ND2	5:A:504:HOH:O	2.47	0.46
1:A:297:ILE:HD12	1:A:345:ILE:HD11	1.96	0.45
1:A:291:PRO:O	1:A:295:ALA:CB	2.63	0.45
1:A:305:ASN:HB2	1:A:333:LEU:HB2	1.99	0.45
1:A:237:THR:O	1:A:241:THR:HG23	2.16	0.45
1:A:283:VAL:HG11	1:A:286:VAL:HG22	1.99	0.44
1:A:306:VAL:HG23	1:A:311:ALA:HB2	2.01	0.43
1:A:262:GLY:HA2	1:A:282:LYS:O	2.21	0.40
1:A:291:PRO:O	1:A:345:ILE:HG23	2.21	0.40
1:A:327:GLY:N	1:A:347:ILE:O	2.52	0.40

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	264/334 (79%)	255 (97%)	9 (3%)	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	224/273 (82%)	221 (99%)	3 (1%)	69 58

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

Mol	Chain	Res	Type
1	A	218	SER
1	A	258	ILE
1	A	282	LYS

Sometimes 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 monosaccharides in this entry.

5.6 Ligand geometry (i)

3 ligands are 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	$\#\|Z\| > 2$
3	PEG	A	402	-	6,6,6	0.66	0	5,5,5	0.59	0
2	CXS	A	401	-	14,14,14	2.01	2 (14%)	18,18,18	2.50	6 (33%)
4	SO4	A	403	-	4,4,4	0.34	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
3	PEG	A	402	-	-	2/4/4/4	-
2	CXS	A	401	-	-	2/8/16/16	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	401	CXS	C1-S	-5.53	1.69	1.77
2	A	401	CXS	O1-S	4.74	1.59	1.45

All (6) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	401	CXS	O1-S-C1	6.24	114.43	106.92
2	A	401	CXS	C3-N-C4	5.41	124.75	114.14
2	A	401	CXS	O2-S-C1	-3.10	103.19	106.92
2	A	401	CXS	C7-C6-C5	2.96	117.46	111.42
2	A	401	CXS	C8-C9-C4	2.83	116.43	111.11
2	A	401	CXS	C6-C5-C4	2.47	115.76	111.11

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	401	CXS	C1-C2-C3-N
2	A	401	CXS	C5-C4-N-C3
3	A	402	PEG	O2-C3-C4-O4
3	A	402	PEG	C4-C3-O2-C2

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

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, 95^{th} 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	272/334 (81%)	0.76	56 (20%) 1 1	10, 28, 112, 133	0

All (56) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	329	THR	7.1
1	A	288	PRO	6.8
1	A	263	GLU	6.0
1	A	312	THR	6.0
1	A	298	HIS	5.5
1	A	333	LEU	5.3
1	A	285	LYS	5.1
1	A	284	LYS	5.0
1	A	310	PRO	4.9
1	A	335	LEU	4.8
1	A	261	THR	4.2
1	A	340	GLN	4.1
1	A	325	ARG	4.0
1	A	342	ALA	3.9
1	A	327	GLY	3.9
1	A	286	VAL	3.8
1	A	331	PRO	3.8
1	A	300	GLY	3.7
1	A	311	ALA	3.7
1	A	330	ILE	3.7
1	A	299	VAL	3.5
1	A	339	GLN	3.5
1	A	304	LEU	3.5
1	A	283	VAL	3.5
1	A	295	ALA	3.4
1	A	315	ILE	3.4
1	A	314	VAL	3.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	302	ILE	3.1
1	A	305	ASN	3.0
1	A	289	ASP	2.9
1	A	346	THR	2.9
1	A	344	GLN	2.8
1	A	294	GLN	2.8
1	A	280	GLY	2.7
1	A	259	GLY	2.7
1	A	334	LEU	2.6
1	A	343	VAL	2.6
1	A	281	ILE	2.6
1	A	293	ALA	2.6
1	A	332	VAL	2.6
1	A	204	LEU	2.5
1	A	292	ALA	2.5
1	A	176	GLY	2.5
1	A	260	ILE	2.5
1	A	338	GLY	2.5
1	A	297	ILE	2.4
1	A	296	GLY	2.3
1	A	341	ILE	2.2
1	A	308	ASN	2.2
1	A	226	GLU	2.2
1	A	328	THR	2.2
1	A	178	ILE	2.1
1	A	303	ILE	2.1
1	A	336	ARG	2.1
1	A	217	LEU	2.1
1	A	249	ARG	2.0

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 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, 95^{th} 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
2	CXS	A	401	14/14	0.87	0.16	29,38,47,48	0
3	PEG	A	402	7/7	0.92	0.12	31,43,53,57	0
4	SO4	A	403	5/5	0.97	0.17	73,73,75,79	5

6.5 Other polymers

There are no such residues in this entry.