



wwPDB X-ray Structure Validation Summary Report

Nov 6, 2014 – 02:35 PM EST

PDB ID : 4CD9
Title : Structure of SasG A-domain (residues 163-419) from Staphylococcus aureus
Authors : Atkin, K.E.; Brentnall, A.S.; Dodson, E.J.; Turkenburg, J.P.; Potts, J.R.
Deposited on : 2013-10-30
Resolution : 1.65 Å(reported)

This is a wwPDB validation summary report for a publicly released PDB entry.

We welcome your comments at validation@mail.wwpdb.org

A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

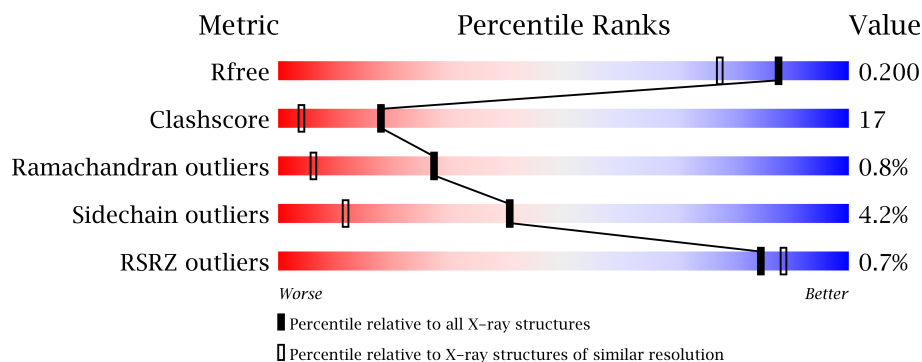
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.16 November 2013
Xtriage (Phenix)	:	dev-1439
EDS	:	stable24103
Percentile statistics	:	21963
Refmac	:	5.8.0049
CCP4	:	6.1.3
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et. al. (1996)
Validation Pipeline (wwPDB-VP)	:	stable24103

1 Overall quality at a glance

The reported resolution of this entry is 1.65 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}	66092	1404 (1.68-1.64)
Clashscore	79885	1001 (1.66-1.66)
Ramachandran outliers	78287	1581 (1.68-1.64)
Sidechain outliers	78261	1580 (1.68-1.64)
RSRZ outliers	66119	1404 (1.68-1.64)

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. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	259	
1	B	259	
1	C	259	
1	D	259	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
3	EDO	A	1420	-	X

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 8263 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

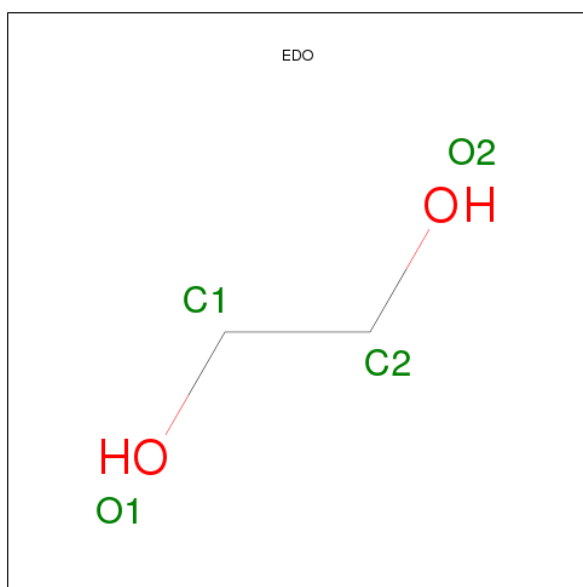
- Molecule 1 is a protein called SURFACE PROTEIN G.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	257	Total	C	N	O	S	0	0	0
			1981	1234	343	400	4			
1	B	255	Total	C	N	O	S	0	0	0
			1966	1225	340	397	4			
1	C	256	Total	C	N	O	S	0	0	0
			1975	1231	342	398	4			
1	D	254	Total	C	N	O	S	0	0	0
			1960	1222	339	395	4			

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	Ca	0	0
			1	1		
2	A	1	Total	Ca	0	0
			1	1		
2	D	1	Total	Ca	0	0
			1	1		
2	C	1	Total	Ca	0	0
			1	1		

- Molecule 3 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: C₂H₆O₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			4	2	2		
3	C	1	Total	C	O	0	0
			4	2	2		
3	C	1	Total	C	O	0	0
			4	2	2		

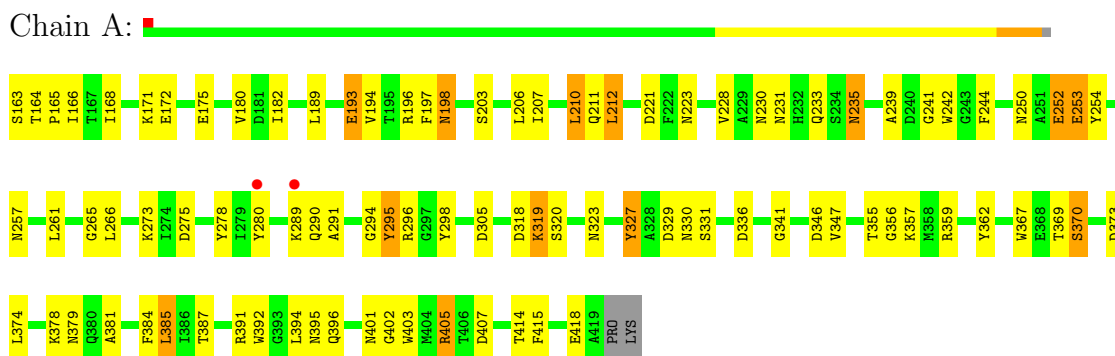
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	95	Total	O	0	3
			98	98		
4	B	80	Total	O	0	1
			81	81		
4	C	101	Total	O	0	1
			102	102		
4	D	84	Total	O	0	0
			84	84		

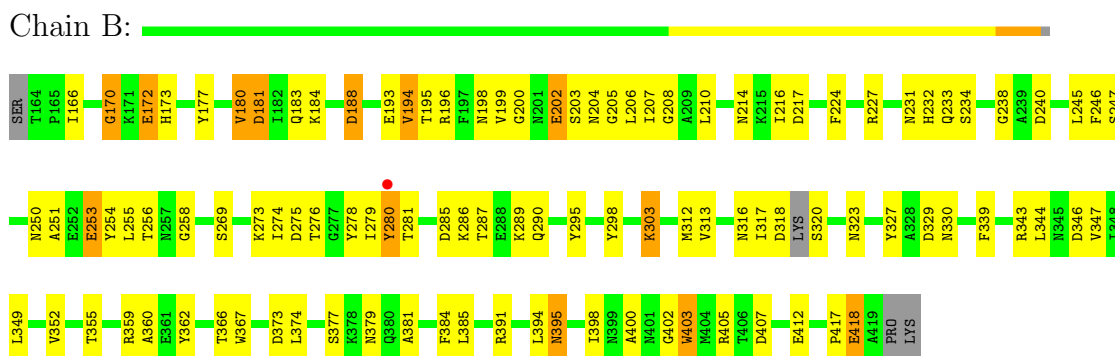
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

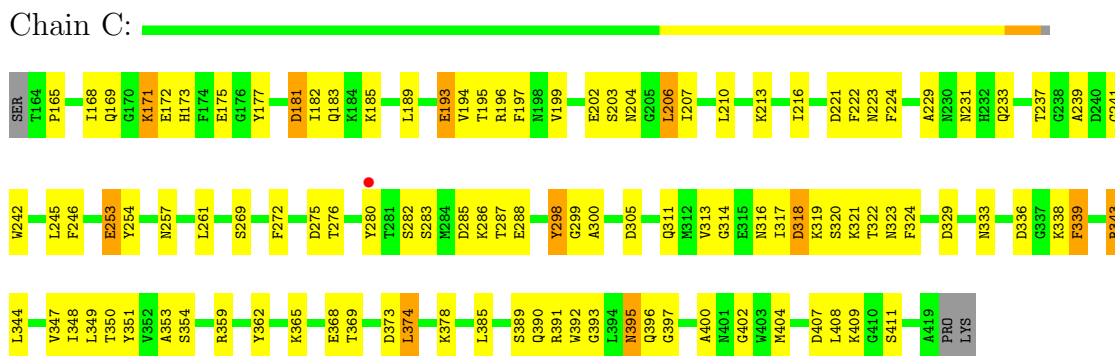
• Molecule 1: SURFACE PROTEIN G



• Molecule 1: SURFACE PROTEIN G



• Molecule 1: SURFACE PROTEIN G



• Molecule 1: SURFACE PROTEIN G

L385	L261	SER
R391	D275	T164
L394	Y278	T167
N395	T281	T168
Q396	D285	Q169
Q397	K286	E172
L398	T287	H173
N399	E288	F174
A400	K289	E175
N401	Q290	S179
G402	A291	V180
W403	G292	D181
W404	Q293	I182
R405	Q299	P186
T406	A300	T187
D407	D305	D188
S411	Q311	L189
T414	N316	E193
E418	I317	V194
A419	D318	T195
PRO	K319	R196
LYS	SER	E202
	LYS	K203
	T322	N204
	N323	G205
	F324	L206
	S335	I207
	F339	K213
	R343	N214
	L344	K215
	T348	I216
	L349	D217
	T350	F218
	Y351	S219
	M358	K220
	R359	K225
	A360	N231
	E361	N235
	Y362	T236
	A363	T237
	G364	G238
	K365	V242
	A381	L245
	Y382	E253
	N383	Y254
	F384	L255

4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α , β , γ	63.21Å 63.21Å 273.36Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	57.37 – 1.65 57.37 – 1.65	Depositor EDS
% Data completeness (in resolution range)	99.1 (57.37-1.65) 99.1 (57.37-1.65)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.50 (at 1.65Å)	Xtriage
Refinement program	REFMAC 5.8.0049	Depositor
R, R_{free}	0.185 , 0.243 0.206 , 0.200	Depositor DCC
R_{free} test set	6365 reflections (5.29%)	DCC
Wilson B-factor (Å ²)	15.0	Xtriage
Anisotropy	0.049	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 15.7	EDS
Estimated twinning fraction	0.506 for H, K, L 0.494 for K, H, -L 0.478 for h,-k,-l	Xtriage
Reported twinning fraction	0.506 for H, K, L 0.494 for K, H, -L	Depositor
L-test for twinning	$\langle L \rangle = 0.40$, $\langle L^2 \rangle = 0.22$	Xtriage
Outliers	0 of 126731 reflections	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	8263	wwPDB-VP
Average B, all atoms (Å ²)	16.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 67.22 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 5.3298e-06.*

¹Intensities estimated from amplitudes.

5 Model quality

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CA, EDO

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	1.40	7/2018 (0.3%)	1.40	14/2718 (0.5%)
1	B	1.34	5/2002 (0.2%)	1.43	19/2696 (0.7%)
1	C	1.44	6/2012 (0.3%)	1.49	23/2710 (0.8%)
1	D	1.36	7/1996 (0.4%)	1.41	17/2688 (0.6%)
All	All	1.38	25/8028 (0.3%)	1.43	73/10812 (0.7%)

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	1
1	B	0	4
1	D	0	1
All	All	0	6

The worst 5 of 25 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	253	GLU	CD-OE1	7.94	1.34	1.25
1	D	242	TRP	CE3-CZ3	-7.50	1.25	1.38
1	B	303	LYS	C-O	-7.21	1.09	1.23
1	B	269	SER	CB-OG	7.17	1.51	1.42
1	D	362	TYR	CB-CG	-6.76	1.41	1.51

The worst 5 of 73 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	343	ARG	NE-CZ-NH2	-12.10	114.25	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	343	ARG	NE-CZ-NH1	9.59	125.09	120.30
1	D	285	ASP	CB-CG-OD2	-9.42	109.82	118.30
1	D	181	ASP	CB-CG-OD1	8.65	126.08	118.30
1	C	221	ASP	CB-CG-OD1	8.34	125.81	118.30

There are no chirality outliers.

5 of 6 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	405	ARG	Mainchain
1	B	194	VAL	Peptide
1	B	238	GLY	Peptide
1	B	247	SER	Peptide
1	B	320	SER	Peptide

5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1981	0	1894	67	0
1	B	1966	0	1875	76	0
1	C	1975	0	1889	73	0
1	D	1960	0	1870	47	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	4	0	6	2	0
3	C	8	0	12	7	0
4	A	98	0	0	5	0
4	B	81	0	0	1	0
4	C	102	0	0	6	1
4	D	84	0	0	1	1
All	All	8263	0	7546	264	1

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 17.

The worst 5 of 264 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:C:231:ASN:HD21	1:C:239:ALA:H	1.07	0.96
1:B:233:GLN:HG3	1:B:234:SER:H	1.27	0.95
1:C:175:GLU:HG2	1:C:213:LYS:HE3	1.48	0.93
1:B:233:GLN:HG3	1:B:234:SER:N	1.84	0.92
1:D:203:SER:O	1:D:395:ASN:C	2.11	0.89

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	Distance(Å)	Clash(Å)
4:C:2047[B]:HOH:O	4:D:2034:HOH:O[4_554]	2.12	0.08

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	255/259 (98%)	228 (89%)	22 (9%)	5 (2%)	11	1
1	B	251/259 (97%)	233 (93%)	17 (7%)	1 (0%)	43	19
1	C	254/259 (98%)	231 (91%)	21 (8%)	2 (1%)	27	6
1	D	250/259 (96%)	226 (90%)	24 (10%)	0	100	100
All	All	1010/1036 (98%)	918 (91%)	84 (8%)	8 (1%)	27	6

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	295	TYR
1	A	198	ASN
1	A	319	LYS
1	B	198	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	290	GLN

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	211/213 (99%)	207 (98%)	4 (2%)	69	43
1	B	209/213 (98%)	200 (96%)	9 (4%)	40	11
1	C	210/213 (99%)	200 (95%)	10 (5%)	35	9
1	D	208/213 (98%)	196 (94%)	12 (6%)	28	5
All	All	838/852 (98%)	803 (96%)	35 (4%)	40	12

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	203	SER
1	C	298	TYR
1	D	311	GLN
1	C	222	PHE
1	C	254	TYR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 38 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	323	ASN
1	C	231	ASN
1	D	333	ASN
1	B	383	ASN
1	C	311	GLN

5.3.3 RNA ⓘ

There are no RNA chains 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 7 ligands modelled in this entry, 4 are monoatomic - leaving 3 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	$\# Z > 2$
3	EDO	A	1420	-	3,3,3	0.76	0	2,2,2	0.50	0
3	EDO	C	1420	-	3,3,3	0.44	0	2,2,2	0.78	0
3	EDO	C	1421	-	3,3,3	0.50	0	2,2,2	0.40	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	EDO	A	1420	-	-	0/1/1/1	0/0/0/0
3	EDO	C	1420	-	-	0/1/1/1	0/0/0/0
3	EDO	C	1421	-	-	0/1/1/1	0/0/0/0

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.

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	257/259 (99%)	-0.04	2 (0%) 83 87	6, 15, 28, 40	0
1	B	255/259 (98%)	0.03	1 (0%) 90 94	7, 16, 29, 34	0
1	C	256/259 (98%)	-0.11	1 (0%) 90 94	7, 15, 28, 37	0
1	D	254/259 (98%)	0.03	3 (1%) 75 79	6, 17, 29, 40	0
All	All	1022/1036 (98%)	-0.02	7 (0%) 84 88	6, 16, 29, 40	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	293	GLN	3.9
1	A	280	TYR	3.5
1	D	238	GLY	2.9
1	C	280	TYR	2.9
1	A	289	LYS	2.7

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	RSR	LLDF	B-factors(\AA^2)	Q<0.9
3	EDO	A	1420	4/4	0.16	2.04	20,24,24,27	0
3	EDO	C	1421	4/4	0.08	-0.85	16,16,17,18	0
2	CA	B	500	1/1	0.07	-1.36	18,18,18,18	0
2	CA	C	500	1/1	0.06	-2.03	16,16,16,16	0
3	EDO	C	1420	4/4	0.06	-2.09	10,13,13,14	0
2	CA	D	500	1/1	0.06	-2.64	19,19,19,19	0
2	CA	A	500	1/1	0.04	-2.88	16,16,16,16	0

6.5 Other polymers ⓘ

There are no such residues in this entry.