



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 2, 2023 – 06:10 pm GMT

PDB ID : 8OPR
Title : Structure of the EA1 surface layer of Bacillus anthracis
Authors : Sogues, A.; Remaut, H.
Deposited on : 2023-04-07
Resolution : 1.81 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

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

MolProbity : 4.02b-467
Mogul : 1.8.4, CSD as541be (2020)
Xtriage (Phenix) : 1.13
EDS : 2.36
buster-report : 1.1.7 (2018)
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) : 2.36

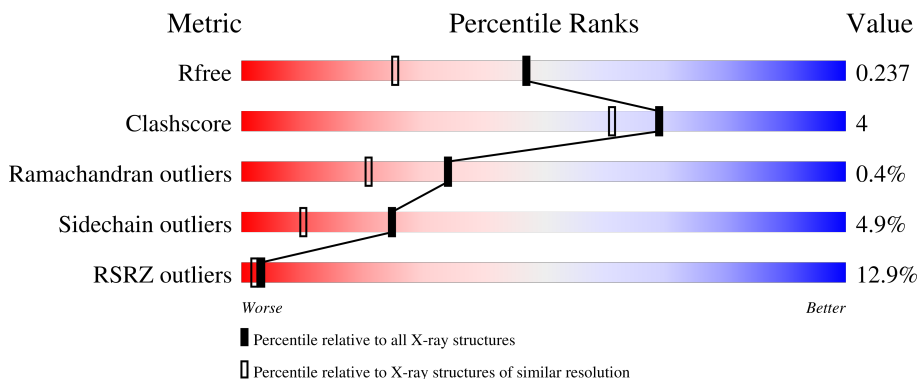
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.81 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	7484 (1.84-1.80)
Clashscore	141614	8401 (1.84-1.80)
Ramachandran outliers	138981	8290 (1.84-1.80)
Sidechain outliers	138945	8290 (1.84-1.80)
RSRZ outliers	127900	7371 (1.84-1.80)

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 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 $\leq 5\%$. 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	665	 88% 8% ••
2	B	127	 79% 13% • 6%
3	C	133	 77% 14% • 6%

2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 7588 atoms, of which 15 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 S-layer protein EA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	651	4793	3009	800	983	1	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	198	MET	-	initiating methionine	UNP P94217
A	199	HIS	-	expression tag	UNP P94217
A	200	HIS	-	expression tag	UNP P94217
A	201	HIS	-	expression tag	UNP P94217
A	202	HIS	-	expression tag	UNP P94217
A	203	HIS	-	expression tag	UNP P94217
A	204	HIS	-	expression tag	UNP P94217
A	205	ILE	-	expression tag	UNP P94217
A	206	THR	-	expression tag	UNP P94217
A	207	SER	-	expression tag	UNP P94217
A	208	LEU	-	expression tag	UNP P94217
A	209	TYR	-	expression tag	UNP P94217
A	210	LYS	-	expression tag	UNP P94217
A	211	LYS	-	expression tag	UNP P94217
A	212	GLY	-	expression tag	UNP P94217
A	213	GLN	-	expression tag	UNP P94217
A	214	MET	-	expression tag	UNP P94217
A	827	LEU	THR	conflict	UNP P94217

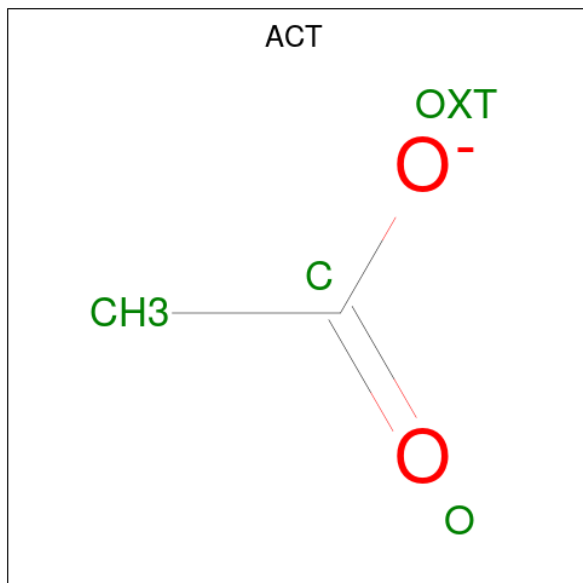
- Molecule 2 is a protein called Nanobody 643.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	120	897	555	159	180	3	0	0	0

- Molecule 3 is a protein called Nanobody 632.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	125	980	615	175	186	4	0	0	0

- Molecule 4 is ACETATE ION (three-letter code: ACT) (formula: C₂H₃O₂).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
4	A	1	7	2	3	2	3	0
4	A	1	7	2	3	2	3	0
4	A	1	7	2	3	2	3	0
4	A	1	7	2	3	2	3	0
4	A	1	7	2	3	2	3	0

- Molecule 5 is CALCIUM ION (three-letter code: CA) (formula: Ca) (labeled as "Ligand of Interest" by depositor).

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

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula: O₄S).



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

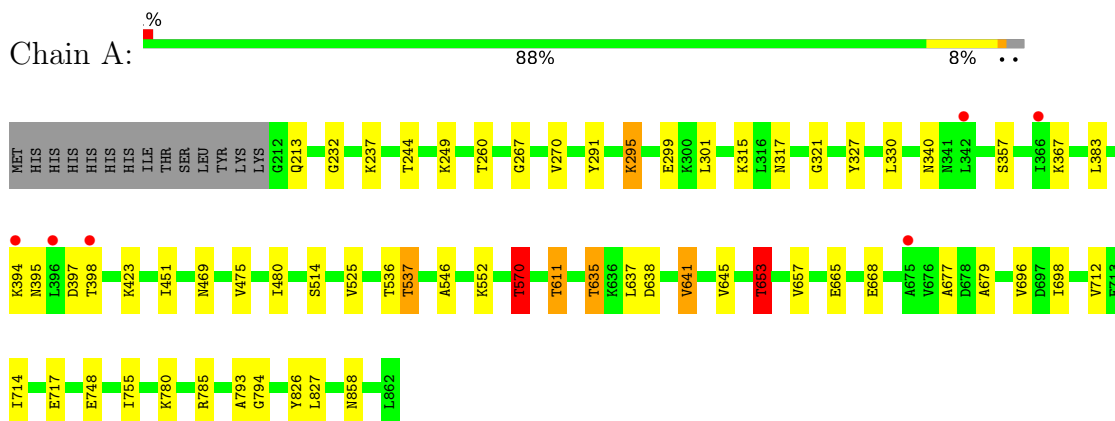
- Molecule 7 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
7	A	795	Total O 795 795	0	0
7	B	46	Total O 46 46	0	0
7	C	19	Total O 19 19	0	0

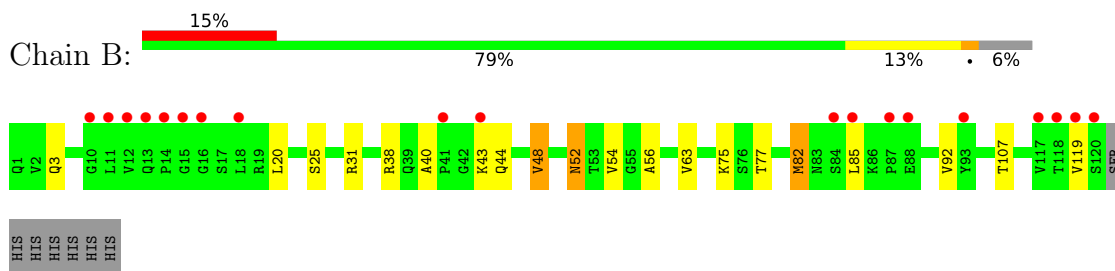
3 Residue-property plots [i](#)

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.

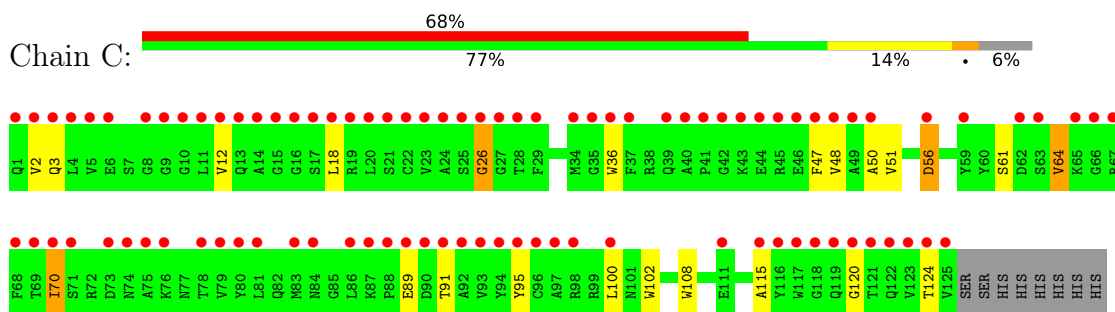
- Molecule 1: S-layer protein EA1



- Molecule 2: Nanobody 643



- Molecule 3: Nanobody 632



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	72.93Å 74.30Å 87.65Å 107.85° 101.14° 112.42°	Depositor
Resolution (Å)	63.37 – 1.81 63.37 – 1.81	Depositor EDS
% Data completeness (in resolution range)	53.0 (63.37-1.81) 53.1 (63.37-1.81)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.46 (at 1.81Å)	Xtrriage
Refinement program	BUSTER 2.10.4 (3-FEB-2022)	Depositor
R, R_{free}	0.204 , 0.243 0.199 , 0.237	Depositor DCC
R_{free} test set	3807 reflections (5.20%)	wwPDB-VP
Wilson B-factor (Å ²)	29.0	Xtrriage
Anisotropy	0.035	Xtrriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 56.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.51$, $\langle L^2 \rangle = 0.34$	Xtrriage
Estimated twinning fraction	0.002 for k,h,-h-k-l	Xtrriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	7588	wwPDB-VP
Average B, all atoms (Å ²)	55.0	wwPDB-VP

Xtrriage'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.*

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, ACT, SO4

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	0.46	0/4848	0.64	3/6589 (0.0%)
2	B	0.33	0/912	0.53	0/1238
3	C	0.24	0/1003	0.44	0/1356
All	All	0.41	0/6763	0.60	3/9183 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	826	TYR	C-N-CA	7.67	140.87	121.70
1	A	570	THR	N-CA-CB	-6.36	98.22	110.30
1	A	653	THR	N-CA-CB	-5.65	99.56	110.30

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.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	4793	0	4817	31	0
2	B	897	0	865	9	0
3	C	980	0	931	12	0
4	A	20	15	15	2	0
5	A	3	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	10	0	0	0	0
6	B	10	0	0	0	0
7	A	795	0	0	2	0
7	B	46	0	0	0	0
7	C	19	0	0	0	0
All	All	7573	15	6628	49	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 4.

The worst 5 of 49 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:546:ALA:H	1:A:570:THR:HG22	1.35	0.91
1:A:653:THR:HG21	7:A:1077:HOH:O	1.81	0.80
1:A:536:THR:OG1	1:A:570:THR:HG21	1.82	0.80
1:A:514:SER:HB3	1:A:537:THR:HG22	1.69	0.74
1:A:611:THR:HG23	7:A:1077:HOH:O	1.92	0.69

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	649/665 (98%)	630 (97%)	17 (3%)	2 (0%)	41	27
2	B	118/127 (93%)	112 (95%)	6 (5%)	0	100	100
3	C	123/133 (92%)	116 (94%)	5 (4%)	2 (2%)	9	2
All	All	890/925 (96%)	858 (96%)	28 (3%)	4 (0%)	34	21

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	397	ASP
1	A	793	ALA
3	C	56	ASP
3	C	26	GLY

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	526/557 (94%)	508 (97%)	18 (3%)	37	22
2	B	95/104 (91%)	85 (90%)	10 (10%)	7	1
3	C	100/109 (92%)	93 (93%)	7 (7%)	15	4
All	All	721/770 (94%)	686 (95%)	35 (5%)	25	10

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

Mol	Chain	Res	Type
3	C	3	GLN
3	C	48	VAL
3	C	70	ILE
1	A	653	THR
1	A	641	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	52	ASN
3	C	101	ASN
1	A	435	ASN
1	A	850	GLN
1	A	858	ASN

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](#)

Of 12 ligands modelled in this entry, 3 are monoatomic - leaving 9 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	ACT	A	905	-	3,3,3	0.74	0	3,3,3	1.68	1 (33%)
6	SO4	A	910	-	4,4,4	0.17	0	6,6,6	0.44	0
6	SO4	B	202	-	4,4,4	0.17	0	6,6,6	0.14	0
4	ACT	A	902	-	3,3,3	0.51	0	3,3,3	1.48	0
4	ACT	A	901	-	3,3,3	1.17	0	3,3,3	1.29	0
6	SO4	A	909	-	4,4,4	0.21	0	6,6,6	0.44	0
4	ACT	A	904	-	3,3,3	0.97	0	3,3,3	1.63	1 (33%)
6	SO4	B	201	-	4,4,4	0.26	0	6,6,6	0.20	0
4	ACT	A	903	-	3,3,3	1.01	0	3,3,3	1.38	0

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	A	905	ACT	OXT-C-O	2.35	130.73	122.05
4	A	904	ACT	OXT-C-O	2.12	129.85	122.05

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	902	ACT	1	0
4	A	901	ACT	1	0

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	651/665 (97%)	0.13	6 (0%) 84 82	16, 36, 67, 85	0
2	B	120/127 (94%)	0.95	19 (15%) 2 1	40, 71, 109, 117	0
3	C	125/133 (93%)	3.92	91 (72%) 0 0	74, 132, 160, 168	0
All	All	896/925 (96%)	0.77	116 (12%) 3 2	16, 44, 139, 168	0

The worst 5 of 116 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	11	LEU	14.0
3	C	124	THR	10.8
2	B	11	LEU	10.4
3	C	86	LEU	10.4
3	C	18	LEU	10.1

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 [i](#)

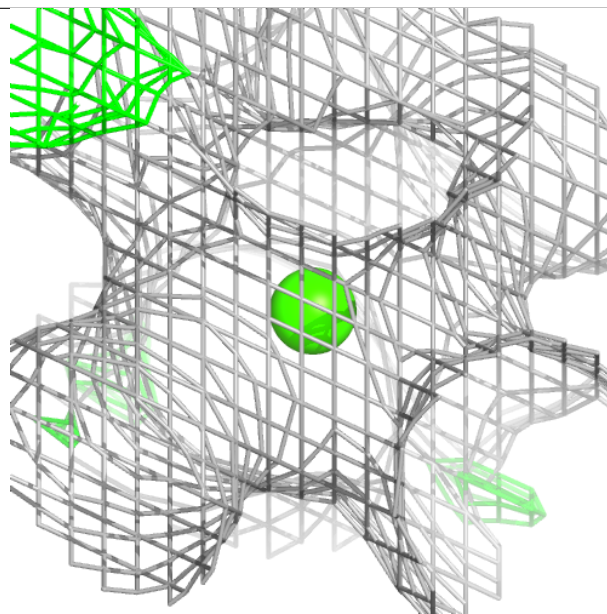
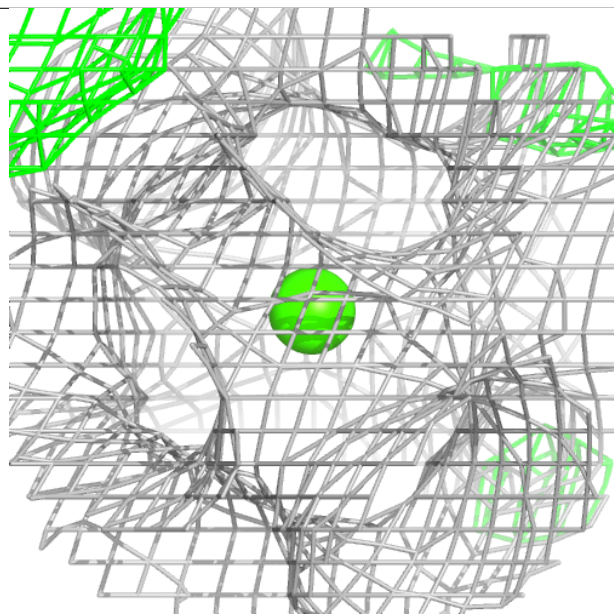
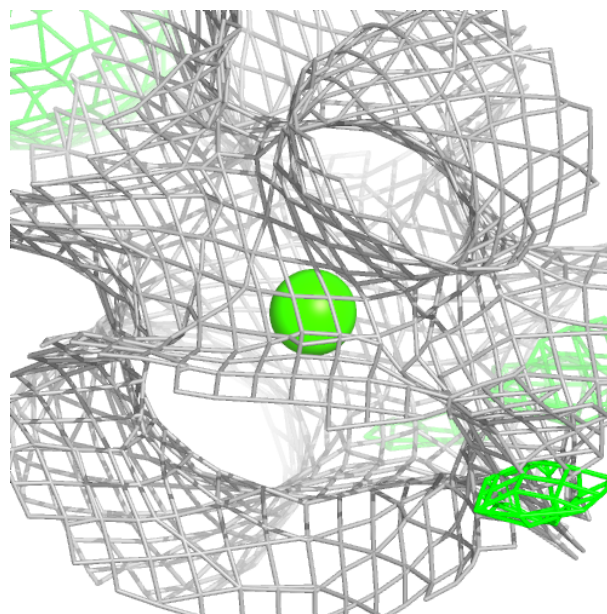
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q < 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
4	ACT	A	904	4/4	0.50	0.19	66,66,70,70	3
6	SO4	B	202	5/5	0.77	0.21	146,146,146,146	0
4	ACT	A	903	4/4	0.80	0.20	50,50,54,55	3
4	ACT	A	901	4/4	0.82	0.21	33,33,39,40	3
6	SO4	A	910	5/5	0.89	0.14	102,102,102,102	0
4	ACT	A	902	4/4	0.94	0.18	37,37,40,42	3
6	SO4	A	909	5/5	0.97	0.11	63,63,63,63	0
6	SO4	B	201	5/5	0.98	0.09	64,65,65,65	0
4	ACT	A	905	4/4	0.98	0.12	42,42,43,44	3
5	CA	A	907	1/1	1.00	0.13	38,38,38,38	0
5	CA	A	908	1/1	1.00	0.12	29,29,29,29	0
5	CA	A	906	1/1	1.00	0.13	25,25,25,25	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

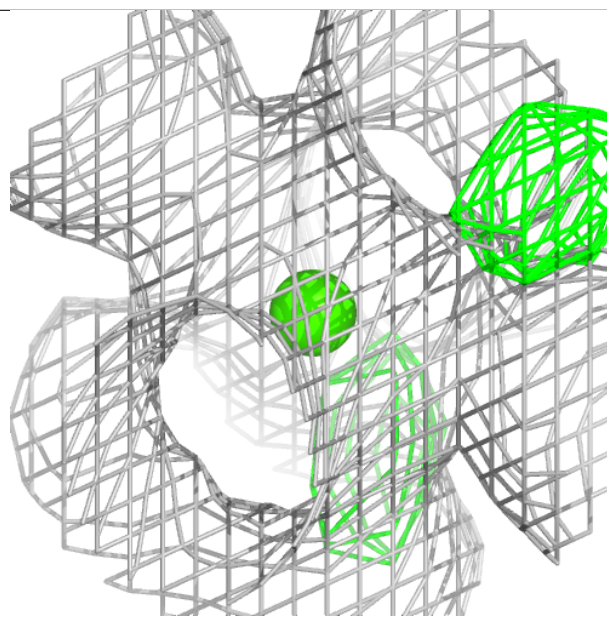
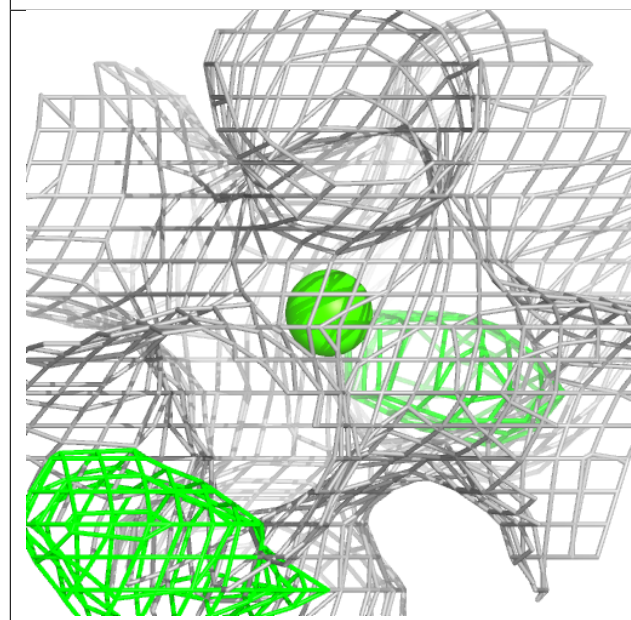
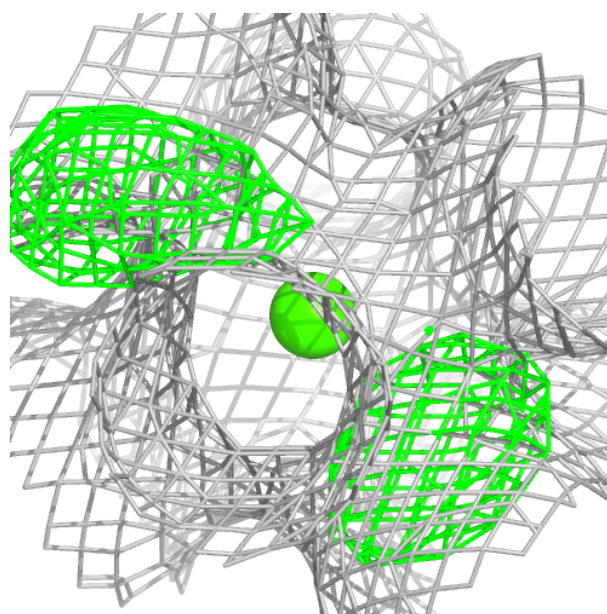
Electron density around CA A 907:

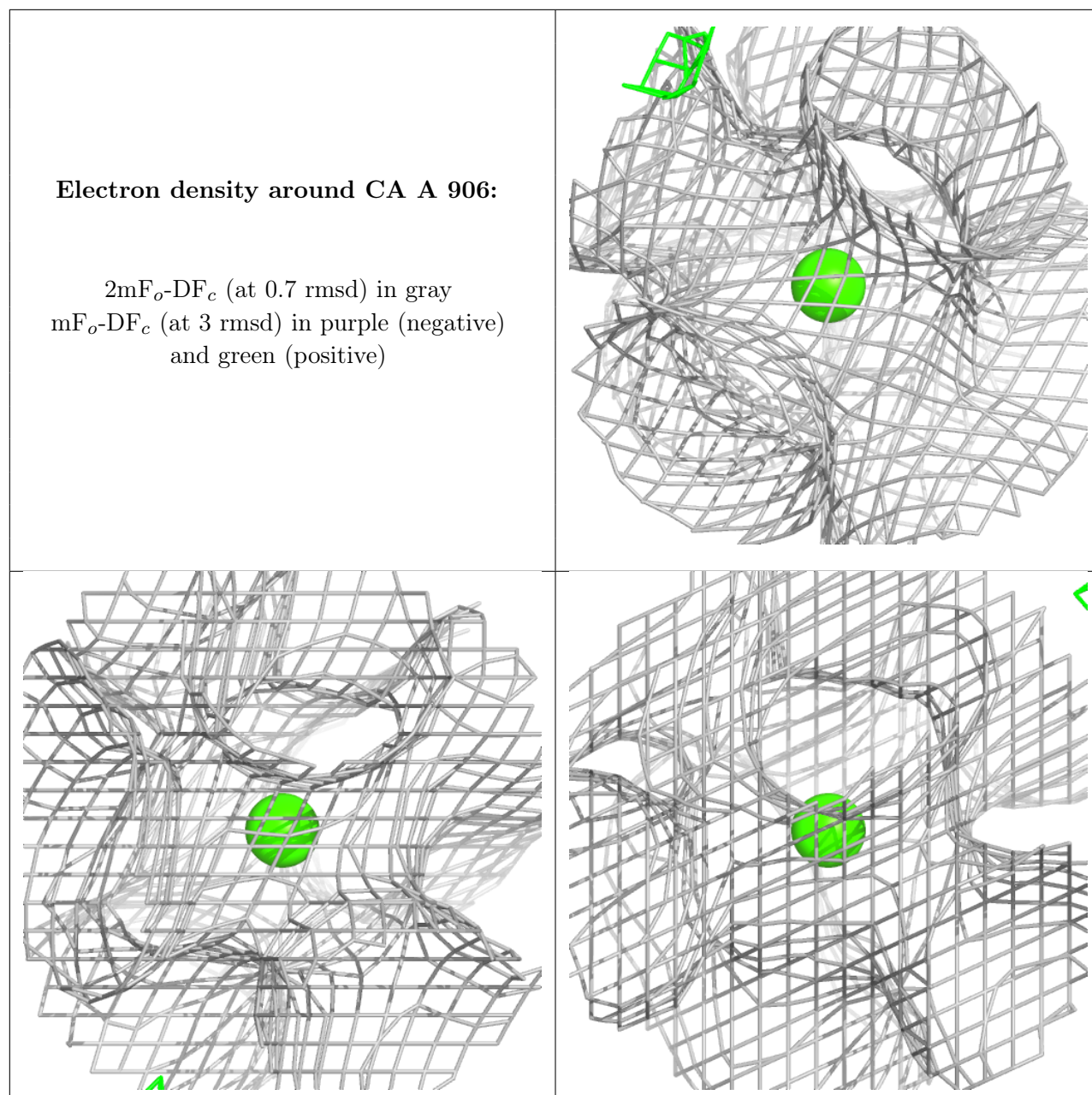
$2mF_o-DF_c$ (at 0.7 rmsd) in gray
 mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



Electron density around CA A 908:

$2mF_o-DF_c$ (at 0.7 rmsd) in gray
 mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)





6.5 Other polymers [i](#)

There are no such residues in this entry.