wwPDB X-ray Structure Validation Summary Report (i)

Aug 18, 2026 – 12:32 pm BST

PDB ID	:	5NQD / pdb_00005nqd
Title	:	Arsenite oxidase AioAB from Rhizobium sp. str. NT-26 mutant AioBF108A
Authors	:	Santos-Silva, T.; Romao, M.; Vieira, M.; Marques, A.T.
Deposited on	:	2017-04-20
Resolution	:	2.20 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 (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:

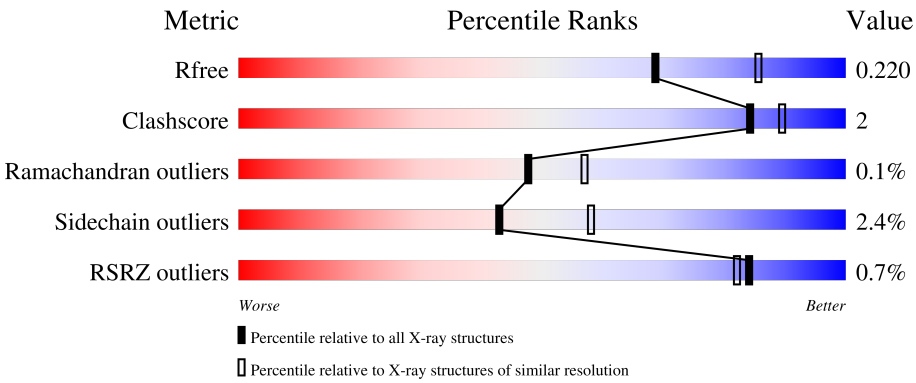
MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Buster-report	:	wwPDB partial adaption of 1.1.7 (2018)
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.015 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.50

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 2.20 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}	180053	6164 (2.20-2.20)
Clashscore	190562	6851 (2.20-2.20)
Ramachandran outliers	187476	6768 (2.20-2.20)
Sidechain outliers	187428	6769 (2.20-2.20)
RSRZ outliers	180081	6166 (2.20-2.20)

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	843	.% 92% 7% •
1	C	843	93% 6%
1	E	843	.% 93% 6% •
1	G	843	.% 92% 7% •
2	B	132	97% •

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	132	2% 97%
2	F	132	2% 92% 8%
2	H	132	% 94% 6%

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
11	EDO	G	2012	-	-	-	X
9	GOL	G	2010	-	X	-	-

2 Entry composition (i)

There are 14 unique types of molecules in this entry. The entry contains 32195 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 AroA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	843	6545	4088	1167	1253	37	0	0	0
1	C	843	6545	4088	1167	1253	37	0	0	0
1	E	843	6545	4088	1167	1253	37	0	0	0
1	G	843	6545	4088	1167	1253	37	0	0	0

- Molecule 2 is a protein called Arsenite oxidase small subunit AioB Rieske [2Fe-2S] cluster.

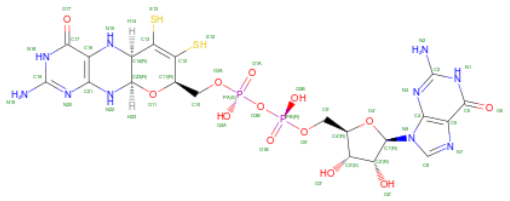
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	132	989	623	166	198	2	0	0	0
2	D	132	989	623	166	198	2	0	0	0
2	F	132	989	623	166	198	2	0	0	0
2	H	132	989	623	166	198	2	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	108	ALA	PHE	conflict	UNP L0NMC5
D	108	ALA	PHE	conflict	UNP L0NMC5
F	108	ALA	PHE	conflict	UNP L0NMC5
H	108	ALA	PHE	conflict	UNP L0NMC5

- Molecule 3 is 2-AMINO-5,6-DIMERCAPTO-7-METHYL-3,7,8A,9-TETRAHYDRO-8-OXA-1,3,9,10-TETRAAZA-ANTHRACEN-4-ONE GUANOSINE DINUCLEOTIDE (CCD ID: MGD) (formula: $C_{20}H_{26}N_{10}O_{13}P_2S_2$).

MGD



Mol	Chain	Residues	Atoms						ZeroOcc	AltConf
			Total	C	N	O	P	S		
3	A	1	47	20	10	13	2	2	0	0
3	A	1	47	20	10	13	2	2	0	0
3	C	1	47	20	10	13	2	2	0	0
3	C	1	47	20	10	13	2	2	0	0
3	E	1	47	20	10	13	2	2	0	0
3	E	1	47	20	10	13	2	2	0	0
3	G	1	47	20	10	13	2	2	0	0
3	G	1	47	20	10	13	2	2	0	0

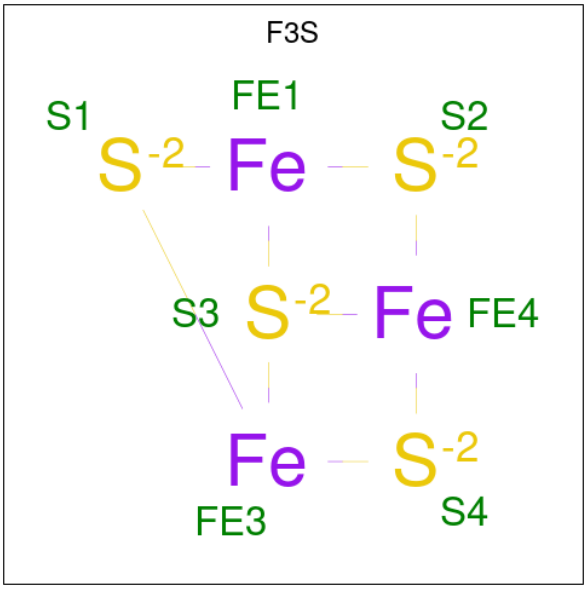
- Molecule 4 is OXYGEN ATOM (CCD ID: O) (formula: O).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	1	1	1	0	0
4	C	1	1	1	0	0
4	E	1	1	1	0	0
4	G	1	1	1	0	0

- Molecule 5 is MOLYBDENUM(IV) ION (CCD ID: 4MO) (formula: Mo).

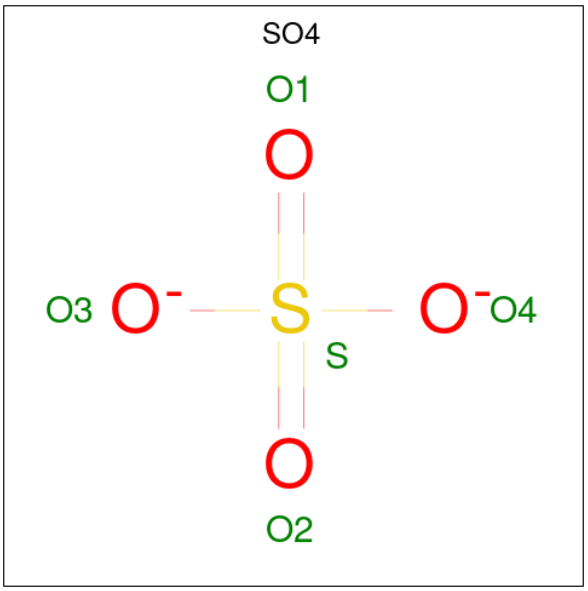
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	1	Total Mo 1 1	0	0
5	C	1	Total Mo 1 1	0	0
5	E	1	Total Mo 1 1	0	0
5	G	1	Total Mo 1 1	0	0

- Molecule 6 is FE3-S4 CLUSTER (CCD ID: F3S) (formula: Fe_3S_4).



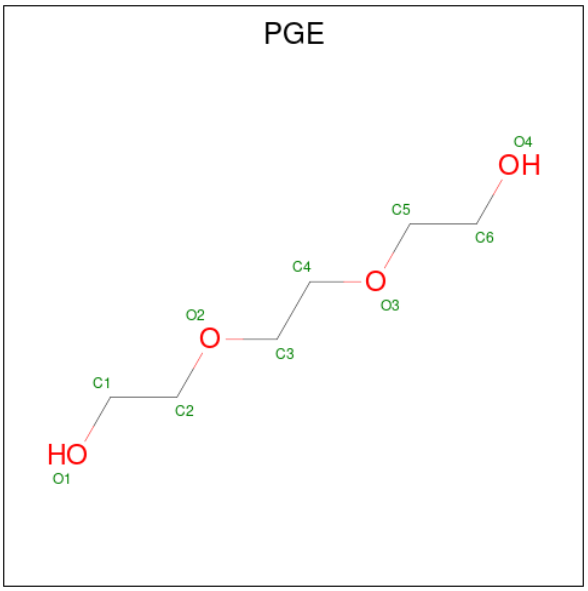
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total Fe S 7 3 4	0	0
6	C	1	Total Fe S 7 3 4	0	0
6	E	1	Total Fe S 7 3 4	0	0
6	G	1	Total Fe S 7 3 4	0	0

- Molecule 7 is SULFATE ION (CCD ID: SO4) (formula: O_4S).



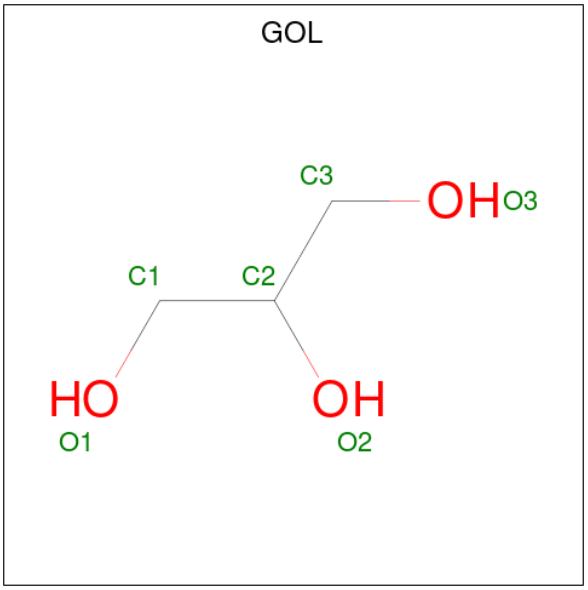
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	O	S			
7	A	1	5	4	1		0	0
7	A	1	5	4	1		0	0
7	A	1	5	4	1		0	0
7	A	1	5	4	1		0	0
7	C	1	5	4	1		0	0
7	C	1	5	4	1		0	0
7	C	1	5	4	1		0	0
7	C	1	5	4	1		0	0
7	E	1	5	4	1		0	0
7	E	1	5	4	1		0	0
7	G	1	5	4	1		0	0
7	G	1	5	4	1		0	0

- Molecule 8 is TRIETHYLENE GLYCOL (CCD ID: PGE) (formula: $C_6H_{14}O_4$).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
8	A	1	10	6	4	0	0

- Molecule 9 is GLYCEROL (CCD ID: GOL) (formula: $C_3H_8O_3$).



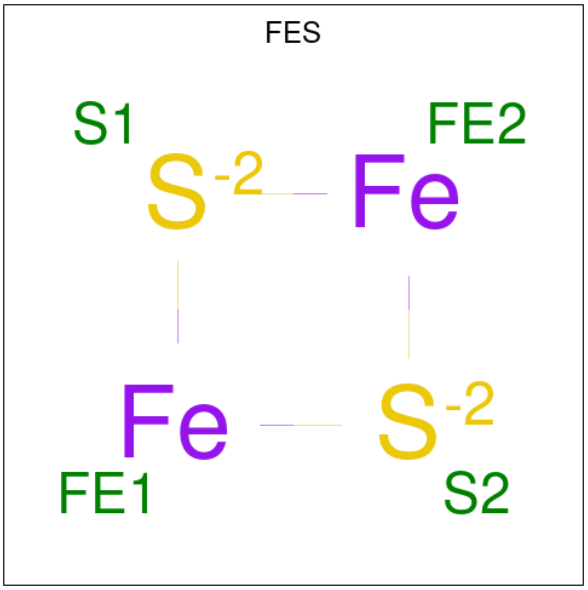
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
9	A	1	6	3	3	0	0
9	C	1	6	3	3	0	0
9	E	1	6	3	3	0	0

Continued on next page...

Continued from previous page...

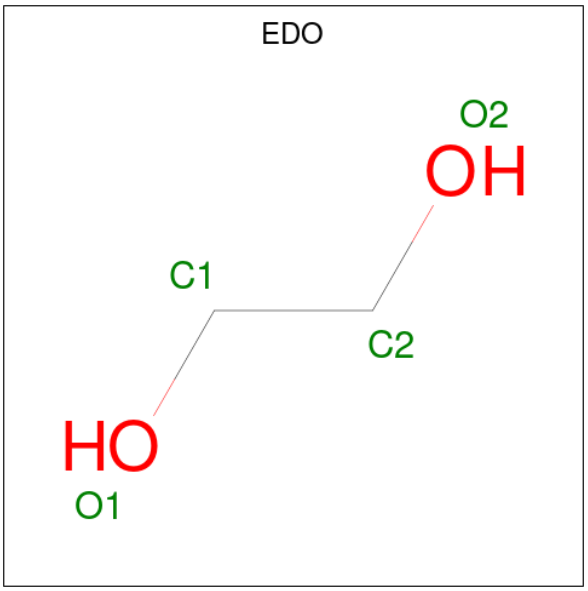
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
9	G	1	6	3	3	0	0
			Total	C	O		
9	G	1	6	3	3	0	0

- Molecule 10 is FE2/S2 (INORGANIC) CLUSTER (CCD ID: FES) (formula: Fe_2S_2).



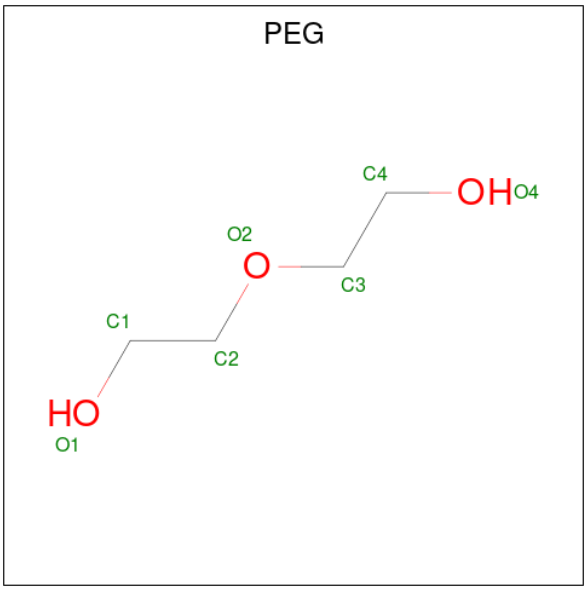
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	Fe	S		
10	B	1	4	2	2	0	0
			Total	Fe	S		
10	D	1	4	2	2	0	0
			Total	Fe	S		
10	F	1	4	2	2	0	0
			Total	Fe	S		
10	H	1	4	2	2	0	0

- Molecule 11 is 1,2-ETHANEDIOL (CCD ID: EDO) (formula: $C_2H_6O_2$).



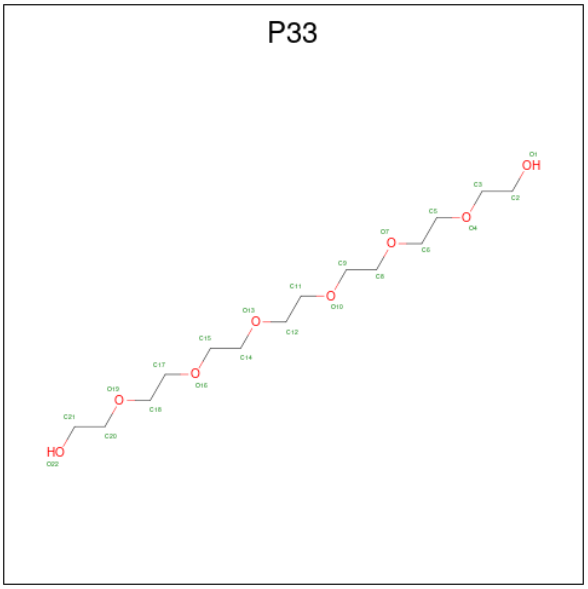
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
11	B	1	4	2	2	0	0
11	C	1	4	2	2	0	0
11	G	1	4	2	2	0	0

- Molecule 12 is DI(HYDROXYETHYL)ETHER (CCD ID: PEG) (formula: $C_4H_{10}O_3$).



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

- Molecule 13 is 3,6,9,12,15,18-HEXAOXAICOSANE-1,20-DIOL (CCD ID: P33) (formula: $C_{14}H_{30}O_8$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
13	G	1	Total C O 22 14 8	0	0
13	G	1	Total C O 22 14 8	0	0

- Molecule 14 is water.

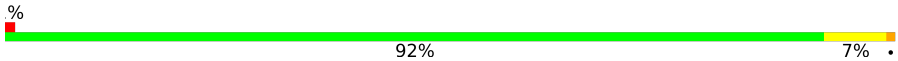
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
14	A	383	Total O 383 383	0	0
14	B	25	Total O 25 25	0	0
14	C	308	Total O 308 308	0	0
14	D	33	Total O 33 33	0	0
14	E	284	Total O 284 284	0	0
14	F	25	Total O 25 25	0	0
14	G	374	Total O 374 374	0	0
14	H	36	Total O 36 36	0	0

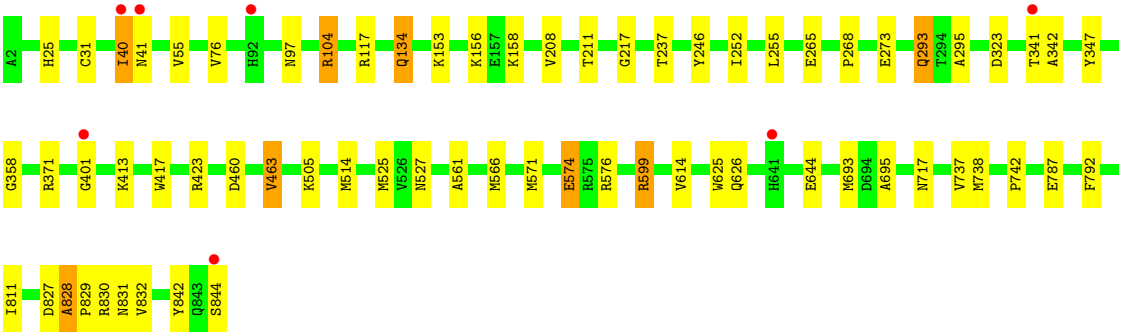
3 Residue-property plots (i)

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: AroA

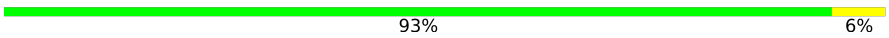
Chain A:

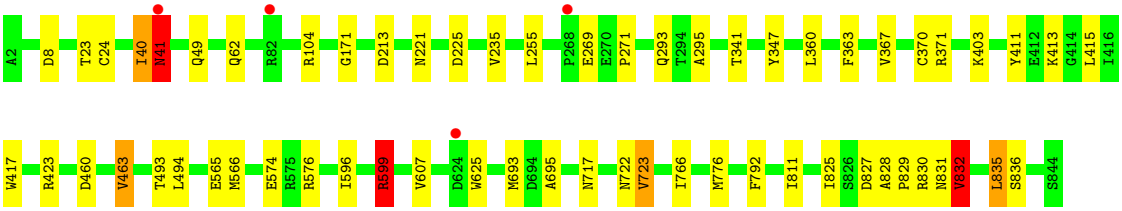




- Molecule 1: AroA

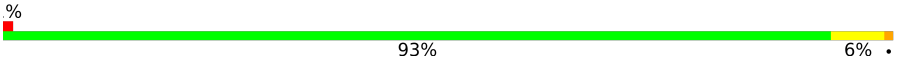
Chain C:

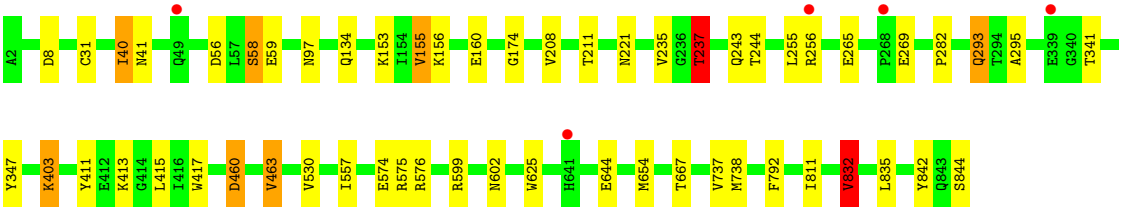




- Molecule 1: AroA

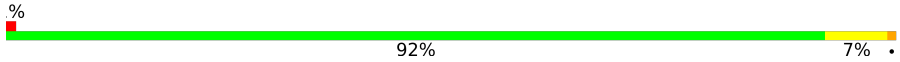
Chain E:

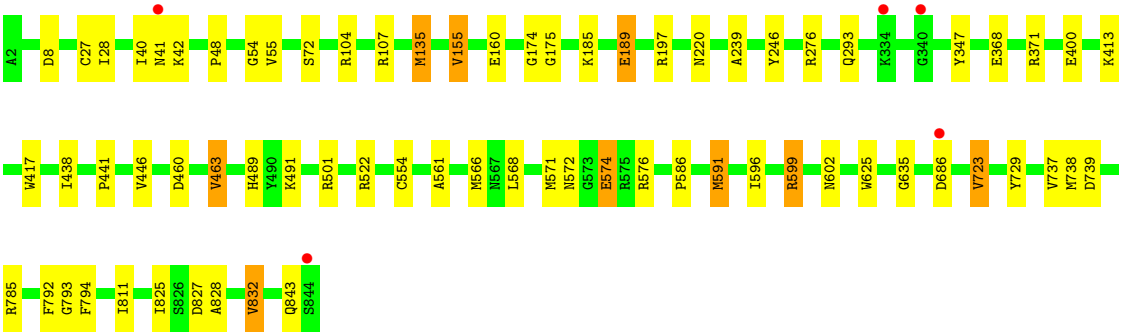




- Molecule 1: AroA

Chain G:



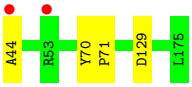


- Molecule 2: Arsenite oxidase small subunit AioB Rieske [2Fe-2S] cluster

Chain B: 97%

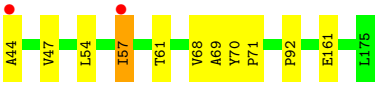
- Molecule 2: Arsenite oxidase small subunit AioB Rieske [2Fe-2S] cluster

Chain D: 2% 97%



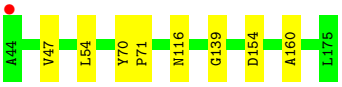
- Molecule 2: Arsenite oxidase small subunit AioB Rieske [2Fe-2S] cluster

Chain F: 2% 92% 8%



- Molecule 2: Arsenite oxidase small subunit AioB Rieske [2Fe-2S] cluster

Chain H: % 94% 6%



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α, β, γ	141.41Å 148.66Å 232.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.55 – 2.20 49.55 – 2.20	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.55-2.20) 100.0 (49.55-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.80 (at 2.20Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.172 , 0.217 0.179 , 0.220	Depositor DCC
R_{free} test set	12218 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	22.6	Xtriage
Anisotropy	0.040	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 27.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.018 for k,h,-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	32195	wwPDB-VP
Average B, all atoms (Å^2)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.32% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: FES, GOL, F3S, PEG, SO4, O, 4MO, P33, EDO, MGD, PGE

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	1.07	1/6693 (0.0%)	1.05	13/9068 (0.1%)
1	C	1.08	3/6693 (0.0%)	1.09	13/9068 (0.1%)
1	E	1.06	2/6693 (0.0%)	1.06	6/9068 (0.1%)
1	G	1.05	3/6693 (0.0%)	1.06	12/9068 (0.1%)
2	B	0.94	0/1011	0.93	0/1378
2	D	0.97	0/1011	0.95	0/1378
2	F	0.97	0/1011	0.99	0/1378
2	H	0.99	0/1011	0.97	1/1378 (0.1%)
All	All	1.06	9/30816 (0.0%)	1.05	45/41784 (0.1%)

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

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	41	ASN	N-CA	8.05	1.55	1.46
1	G	175	GLY	C-O	6.92	1.28	1.23
1	A	252	ILE	CA-C	6.30	1.59	1.52
1	C	171	GLY	N-CA	6.15	1.50	1.44
1	G	591	MET	N-CA	-5.83	1.41	1.45

The worst 5 of 45 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	723	VAL	CB-CA-C	-7.89	99.47	112.26
1	E	237	THR	CB-CA-C	-7.59	94.33	109.35
1	G	723	VAL	CB-CA-C	-7.49	100.13	112.26
1	C	40	ILE	CA-C-N	-7.47	113.24	126.45
1	C	40	ILE	C-N-CA	-7.47	113.24	126.45

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	401	GLY	Peptide

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	6545	0	6296	35	0
1	C	6545	0	6296	30	0
1	E	6545	0	6296	24	0
1	G	6545	0	6296	32	1
2	B	989	0	943	1	0
2	D	989	0	943	3	0
2	F	989	0	943	6	0
2	H	989	0	943	4	0
3	A	94	0	44	1	0
3	C	94	0	44	1	0
3	E	94	0	44	1	0
3	G	94	0	44	2	0
4	A	1	0	0	0	0
4	C	1	0	0	0	0
4	E	1	0	0	0	0
4	G	1	0	0	1	0
5	A	1	0	0	0	0
5	C	1	0	0	0	0
5	E	1	0	0	0	0
5	G	1	0	0	1	0
6	A	7	0	0	0	0
6	C	7	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	E	7	0	0	0	0
6	G	7	0	0	0	0
7	A	20	0	0	1	0
7	C	20	0	0	0	0
7	E	10	0	0	0	0
7	G	10	0	0	0	0
8	A	10	0	14	0	0
9	A	6	0	8	0	0
9	C	6	0	8	0	0
9	E	6	0	8	0	0
9	G	12	0	16	0	0
10	B	4	0	0	0	0
10	D	4	0	0	0	0
10	F	4	0	0	0	0
10	H	4	0	0	0	0
11	B	4	0	6	0	0
11	C	4	0	6	0	0
11	G	4	0	6	0	0
12	C	7	0	10	0	0
13	G	44	0	60	3	0
14	A	383	0	0	6	0
14	B	25	0	0	0	0
14	C	308	0	0	3	0
14	D	33	0	0	1	0
14	E	284	0	0	0	0
14	F	25	0	0	1	0
14	G	374	0	0	6	1
14	H	36	0	0	0	0
All	All	32195	0	29274	134	2

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 2.

The worst 5 of 134 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:E:574:GLU:OE1	1:E:576:ARG:NH1	1.95	1.00
4:G:2003:O:O	5:G:2004:4MO:MO	1.35	0.97
1:A:566:MET:HE1	14:A:2104:HOH:O	1.75	0.87
1:G:574:GLU:OE1	1:G:576:ARG:NH1	2.08	0.86
1:G:599:ARG:NH1	1:G:602:ASN:OD1	2.10	0.85

All (2) 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	Interatomic distance (Å)	Clash overlap (Å)
14:G:2463:HOH:O	14:G:2463:HOH:O[2_545]	1.27	0.93
1:G:686:ASP:O	1:G:686:ASP:O[2_545]	1.98	0.22

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	841/843 (100%)	808 (96%)	32 (4%)	1 (0%)	48	57
1	C	841/843 (100%)	805 (96%)	35 (4%)	1 (0%)	48	57
1	E	841/843 (100%)	807 (96%)	33 (4%)	1 (0%)	48	57
1	G	841/843 (100%)	804 (96%)	36 (4%)	1 (0%)	48	57
2	B	130/132 (98%)	124 (95%)	6 (5%)	0	100	100
2	D	130/132 (98%)	123 (95%)	7 (5%)	0	100	100
2	F	130/132 (98%)	122 (94%)	8 (6%)	0	100	100
2	H	130/132 (98%)	122 (94%)	8 (6%)	0	100	100
All	All	3884/3900 (100%)	3715 (96%)	165 (4%)	4 (0%)	48	57

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	G	811	ILE
1	C	811	ILE
1	E	811	ILE
1	A	811	ILE

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	677/677 (100%)	662 (98%)	15 (2%)	45	61
1	C	677/677 (100%)	661 (98%)	16 (2%)	43	58
1	E	677/677 (100%)	657 (97%)	20 (3%)	36	49
1	G	677/677 (100%)	659 (97%)	18 (3%)	39	53
2	B	104/104 (100%)	102 (98%)	2 (2%)	50	66
2	D	104/104 (100%)	104 (100%)	0	100	100
2	F	104/104 (100%)	101 (97%)	3 (3%)	37	51
2	H	104/104 (100%)	103 (99%)	1 (1%)	68	81
All	All	3124/3124 (100%)	3049 (98%)	75 (2%)	43	58

5 of 75 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	G	42	LYS
1	G	785	ARG
1	G	155	VAL
1	G	438	ILE
1	C	403	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	G	293	GLN
1	G	478	GLN
2	H	143	GLN
1	C	472	GLN
1	C	293	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 48 ligands modelled in this entry, 8 are monoatomic - leaving 40 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	$\#\|Z\| > 2$
8	PGE	A	2010	-	9,9,9	1.14	0	8,8,8	1.30	0
7	SO4	C	2006	-	4,4,4	0.60	0	6,6,6	0.71	0
13	P33	G	2008	-	21,21,21	1.04	0	20,20,20	1.54	4 (20%)
3	MGD	A	2001	5	45,52,52	1.50	7 (15%)	54,81,81	1.48	9 (16%)
7	SO4	E	2007	-	4,4,4	0.48	0	6,6,6	0.53	0
7	SO4	G	2007	-	4,4,4	0.55	0	6,6,6	0.54	0
3	MGD	E	2002	5	45,52,52	1.31	7 (15%)	54,81,81	1.85	14 (25%)
9	GOL	E	2008	-	5,5,5	0.50	0	5,5,5	0.37	0
10	FES	D	201	2	0,4,4	-	-	-		
3	MGD	C	2001	5	45,52,52	1.49	8 (17%)	54,81,81	1.96	17 (31%)
6	F3S	E	2005	1	0,9,9	-	-	-		
9	GOL	C	2011	-	5,5,5	0.71	0	5,5,5	1.01	0
7	SO4	A	2006	-	4,4,4	0.46	0	6,6,6	0.88	0
12	PEG	C	2010	-	6,6,6	0.65	0	5,5,5	0.63	0
7	SO4	C	2007	-	4,4,4	0.57	0	6,6,6	0.83	0
3	MGD	E	2001	5	45,52,52	1.61	7 (15%)	54,81,81	1.81	11 (20%)
10	FES	F	201	2	0,4,4	-	-	-		
7	SO4	A	2008	-	4,4,4	0.79	0	6,6,6	1.06	0
7	SO4	C	2009	-	4,4,4	0.67	0	6,6,6	0.57	0
7	SO4	A	2009	-	4,4,4	0.46	0	6,6,6	1.00	0
10	FES	H	201	2	0,4,4	-	-	-		

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
9	GOL	G	2010	-	5,5,5	1.31	1 (20%)	5,5,5	1.76	2 (40%)
6	F3S	C	2005	1	0,9,9	-	-	-		
11	EDO	B	202	-	3,3,3	0.39	0	2,2,2	0.83	0
6	F3S	A	2005	1	0,9,9	-	-	-		
3	MGD	G	2002	5	45,52,52	1.67	9 (20%)	54,81,81	1.67	10 (18%)
9	GOL	G	2011	-	5,5,5	1.11	0	5,5,5	1.85	1 (20%)
3	MGD	G	2001	5	45,52,52	1.37	5 (11%)	54,81,81	1.69	12 (22%)
11	EDO	G	2012	-	3,3,3	0.64	0	2,2,2	0.45	0
7	SO4	G	2006	-	4,4,4	0.42	0	6,6,6	0.75	0
7	SO4	A	2007	-	4,4,4	0.41	0	6,6,6	0.89	0
7	SO4	E	2006	-	4,4,4	0.32	0	6,6,6	0.46	0
11	EDO	C	2012	-	3,3,3	0.48	0	2,2,2	0.45	0
13	P33	G	2009	-	21,21,21	0.69	0	20,20,20	0.82	1 (5%)
3	MGD	A	2002	5	45,52,52	1.48	6 (13%)	54,81,81	1.75	11 (20%)
9	GOL	A	2011	-	5,5,5	1.07	0	5,5,5	1.45	1 (20%)
3	MGD	C	2002	5	45,52,52	1.75	10 (22%)	54,81,81	1.75	12 (22%)
6	F3S	G	2005	1	0,9,9	-	-	-		
7	SO4	C	2008	-	4,4,4	0.45	0	6,6,6	0.23	0
10	FES	B	201	2	0,4,4	-	-	-		

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
8	PGE	A	2010	-	-	4/7/7/7	-
13	P33	G	2008	-	-	10/19/19/19	-
3	MGD	A	2001	5	-	4/22/66/66	0/6/6/6
3	MGD	E	2002	5	-	6/22/66/66	0/6/6/6
9	GOL	E	2008	-	-	4/4/4/4	-
10	FES	D	201	2	-	-	0/1/1/1
3	MGD	C	2001	5	-	4/22/66/66	0/6/6/6
9	GOL	C	2011	-	-	3/4/4/4	-
6	F3S	E	2005	1	-	-	0/3/3/3
12	PEG	C	2010	-	-	3/4/4/4	-
10	FES	F	201	2	-	-	0/1/1/1
10	FES	H	201	2	-	-	0/1/1/1
9	GOL	G	2010	-	-	3/4/4/4	-

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
11	EDO	B	202	-	-	1/1/1/1	-
6	F3S	C	2005	1	-	-	0/3/3/3
6	F3S	A	2005	1	-	-	0/3/3/3
3	MGD	G	2002	5	-	3/22/66/66	0/6/6/6
9	GOL	G	2011	-	-	2/4/4/4	-
3	MGD	G	2001	5	-	5/22/66/66	0/6/6/6
11	EDO	G	2012	-	-	1/1/1/1	-
11	EDO	C	2012	-	-	1/1/1/1	-
13	P33	G	2009	-	-	11/19/19/19	-
6	F3S	G	2005	1	-	-	0/3/3/3
3	MGD	A	2002	5	-	3/22/66/66	0/6/6/6
9	GOL	A	2011	-	-	4/4/4/4	-
3	MGD	C	2002	5	-	3/22/66/66	0/6/6/6
3	MGD	E	2001	5	-	5/22/66/66	0/6/6/6
10	FES	B	201	2	-	-	0/1/1/1

The worst 5 of 60 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	2001	MGD	C16-C21	6.55	1.49	1.38
3	C	2002	MGD	C16-C21	5.58	1.48	1.38
3	C	2001	MGD	C16-C21	5.56	1.48	1.38
3	G	2002	MGD	C16-C21	5.46	1.48	1.38
3	A	2002	MGD	C16-C21	5.09	1.47	1.38

The worst 5 of 105 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	E	2001	MGD	C19-N20-C21	4.80	122.10	113.43
3	G	2001	MGD	C19-N20-C21	4.77	122.04	113.43
3	C	2001	MGD	C19-N20-C21	4.67	121.86	113.43
3	C	2002	MGD	C5-C4-N3	-4.55	121.09	128.46
3	C	2001	MGD	C5-C4-N3	-4.41	121.31	128.46

There are no chirality outliers.

5 of 80 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	2001	MGD	PA-O3B-PB-O5'
3	A	2001	MGD	C5'-O5'-PB-O3B

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms
3	A	2002	MGD	C5'-O5'-PB-O1B
3	C	2001	MGD	C5'-O5'-PB-O3B
3	E	2001	MGD	PA-O3B-PB-O5'

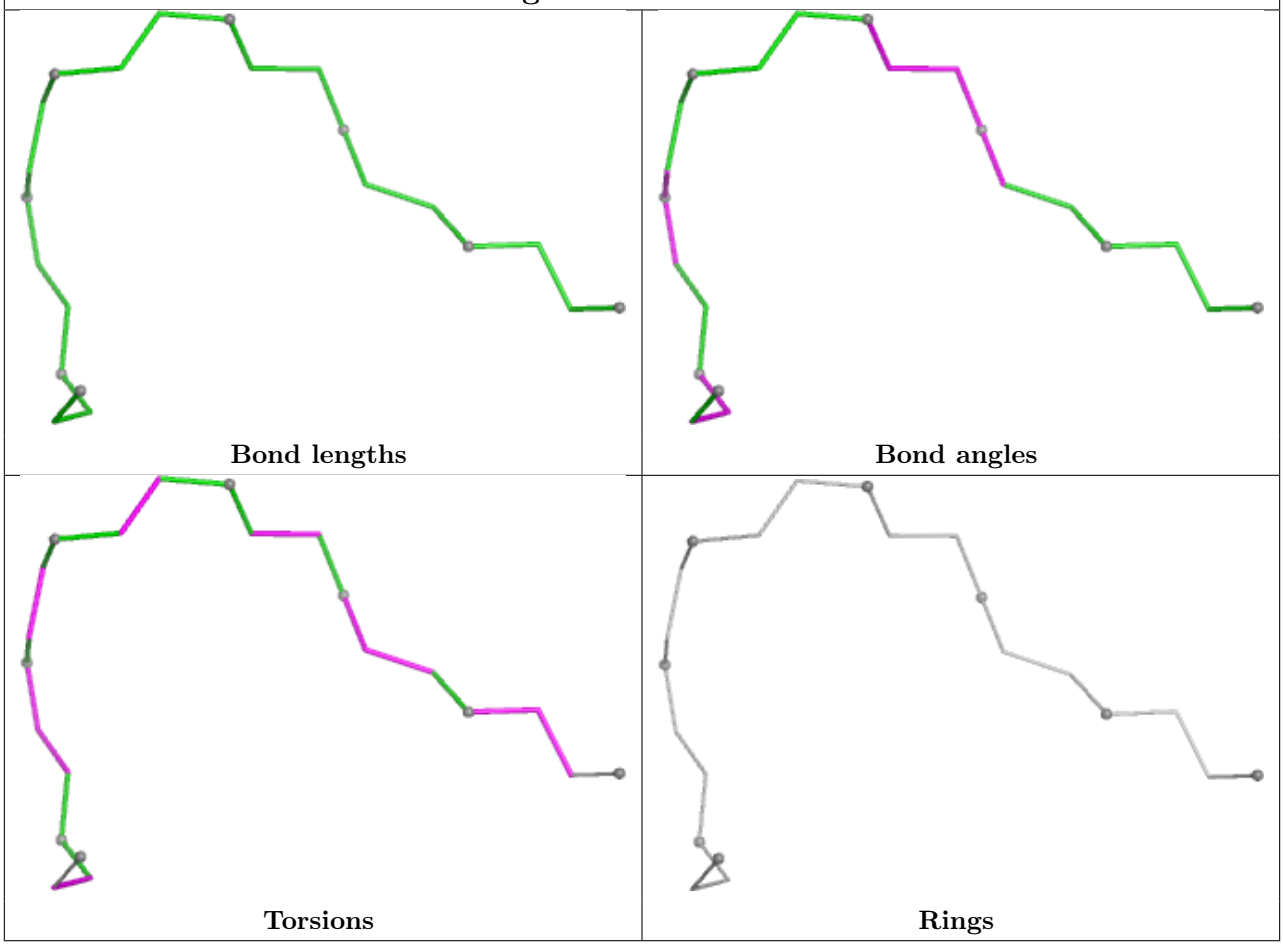
There are no ring outliers.

6 monomers are involved in 9 short contacts:

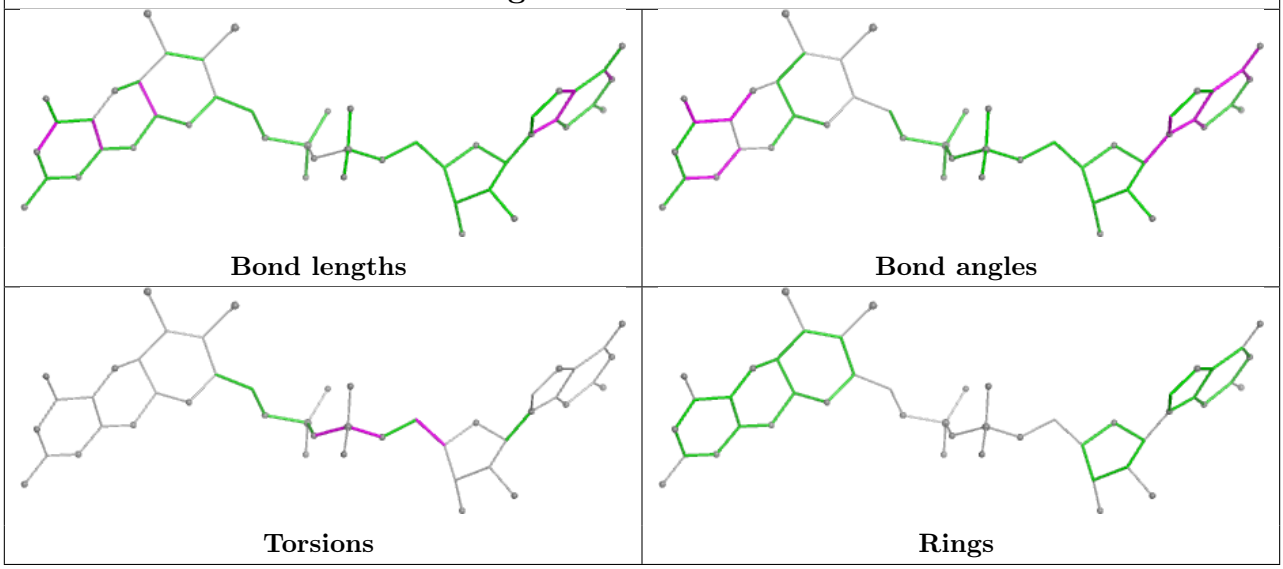
Mol	Chain	Res	Type	Clashes	Symm-Clashes
13	G	2008	P33	3	0
3	E	2001	MGD	1	0
7	A	2008	SO4	1	0
3	G	2001	MGD	2	0
3	A	2002	MGD	1	0
3	C	2002	MGD	1	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

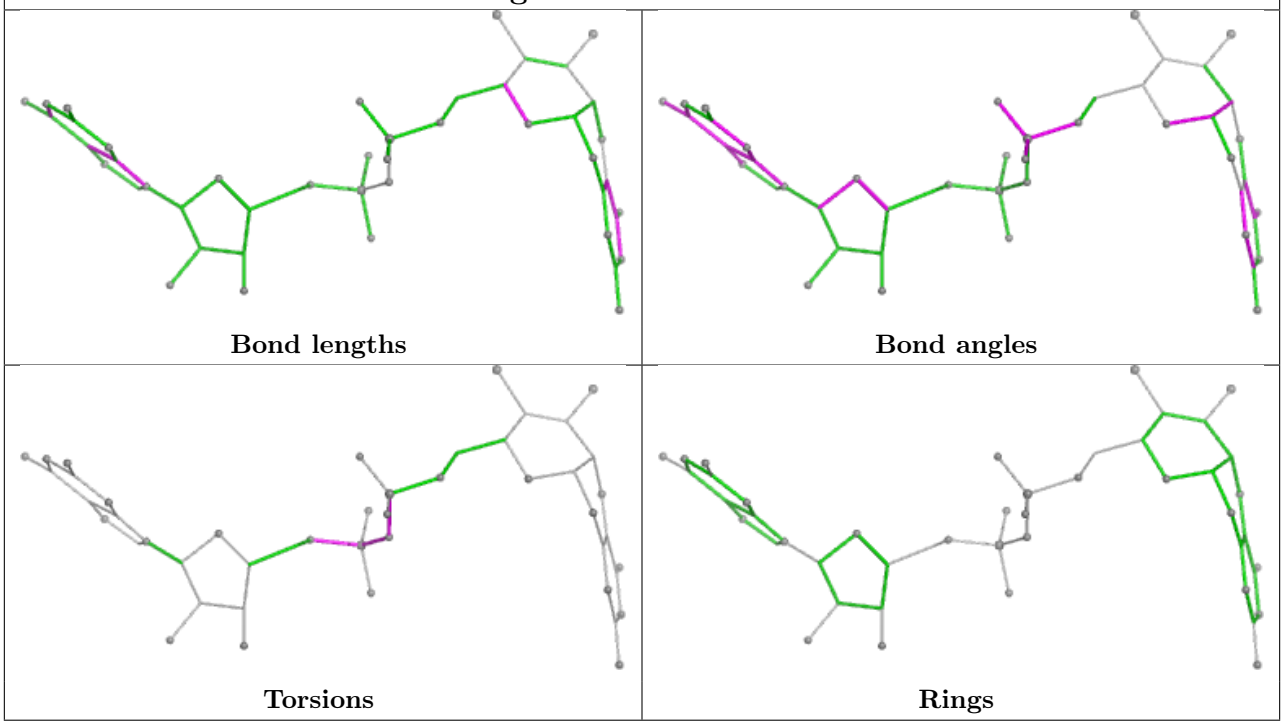
Ligand P33 G 2008



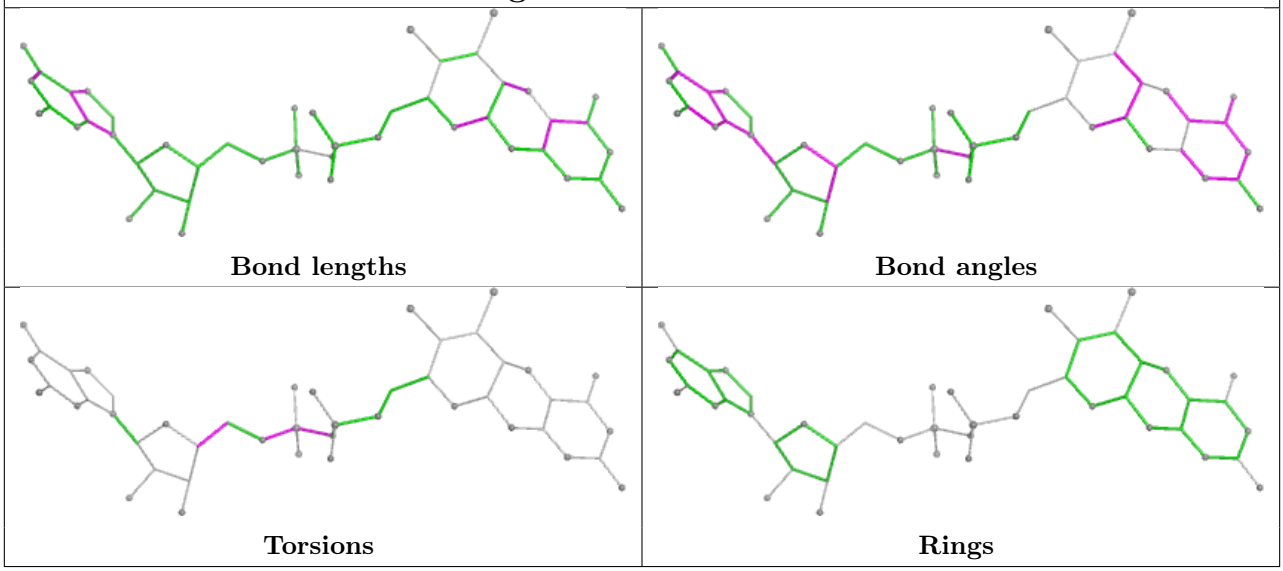
Ligand MGD A 2001



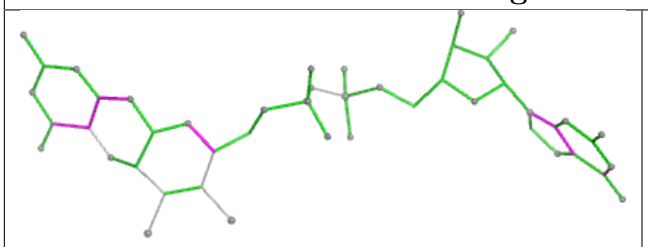
Ligand MGD E 2002



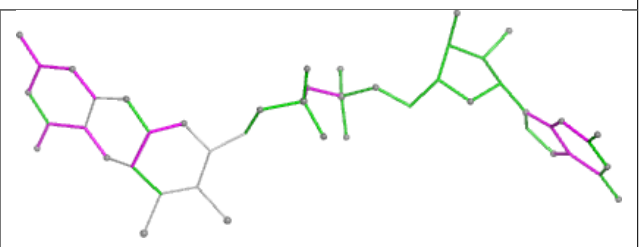
Ligand MGD C 2001



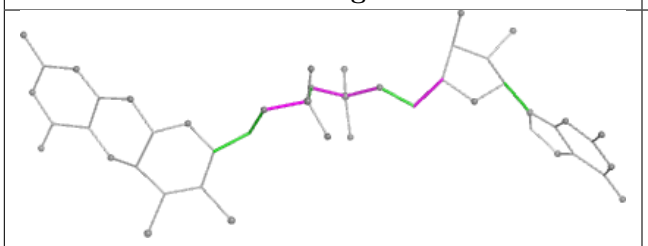
Ligand MGD E 2001



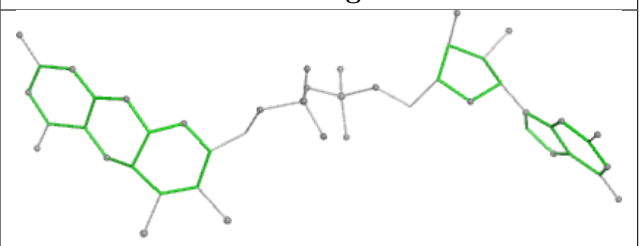
Bond lengths



Bond angles

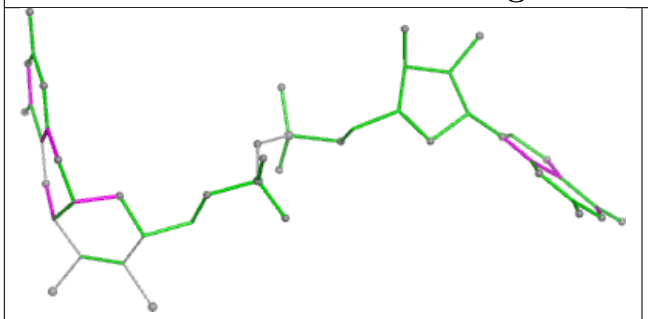


Torsions

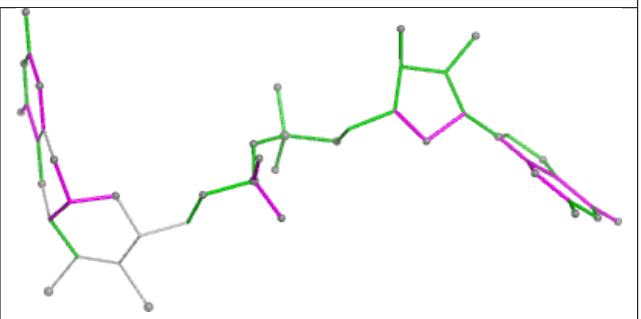


Rings

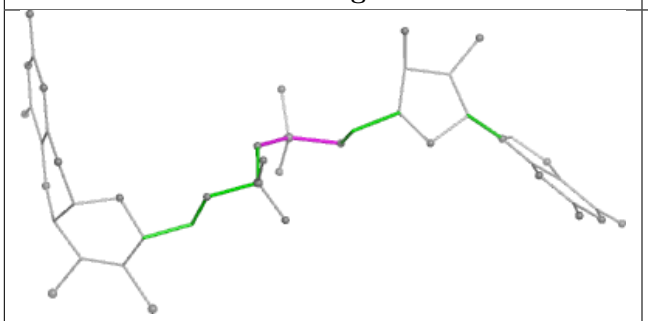
Ligand MGD G 2002



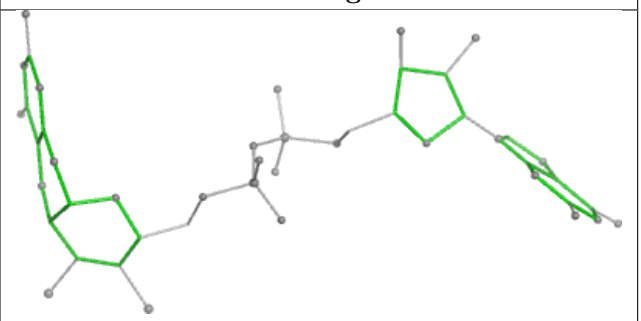
Bond lengths



Bond angles

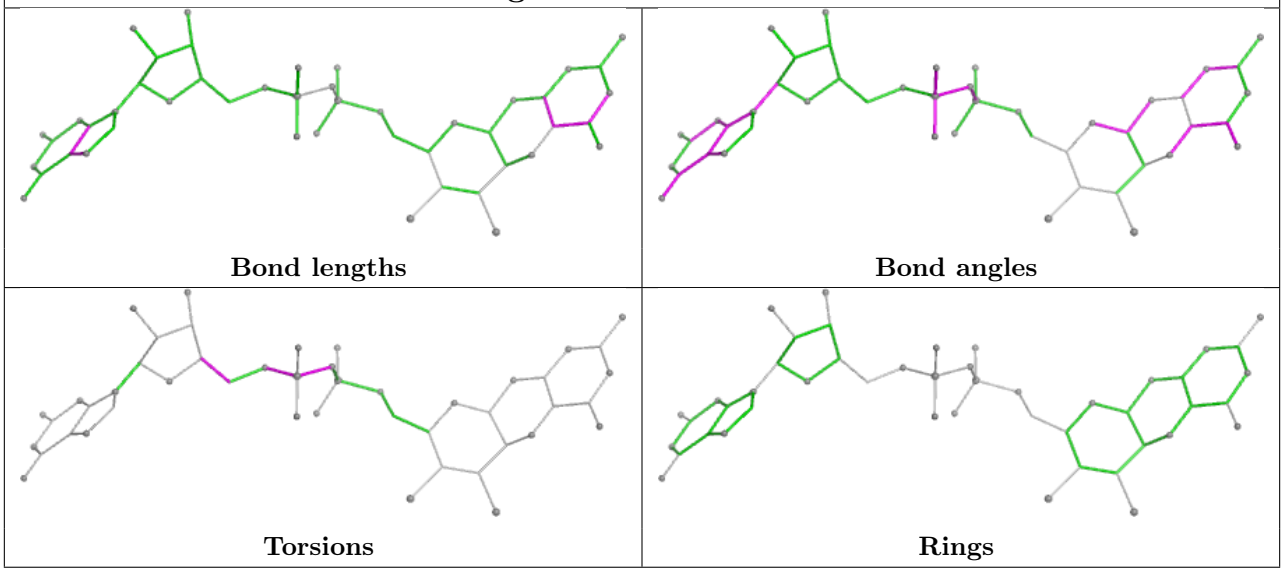


Torsions

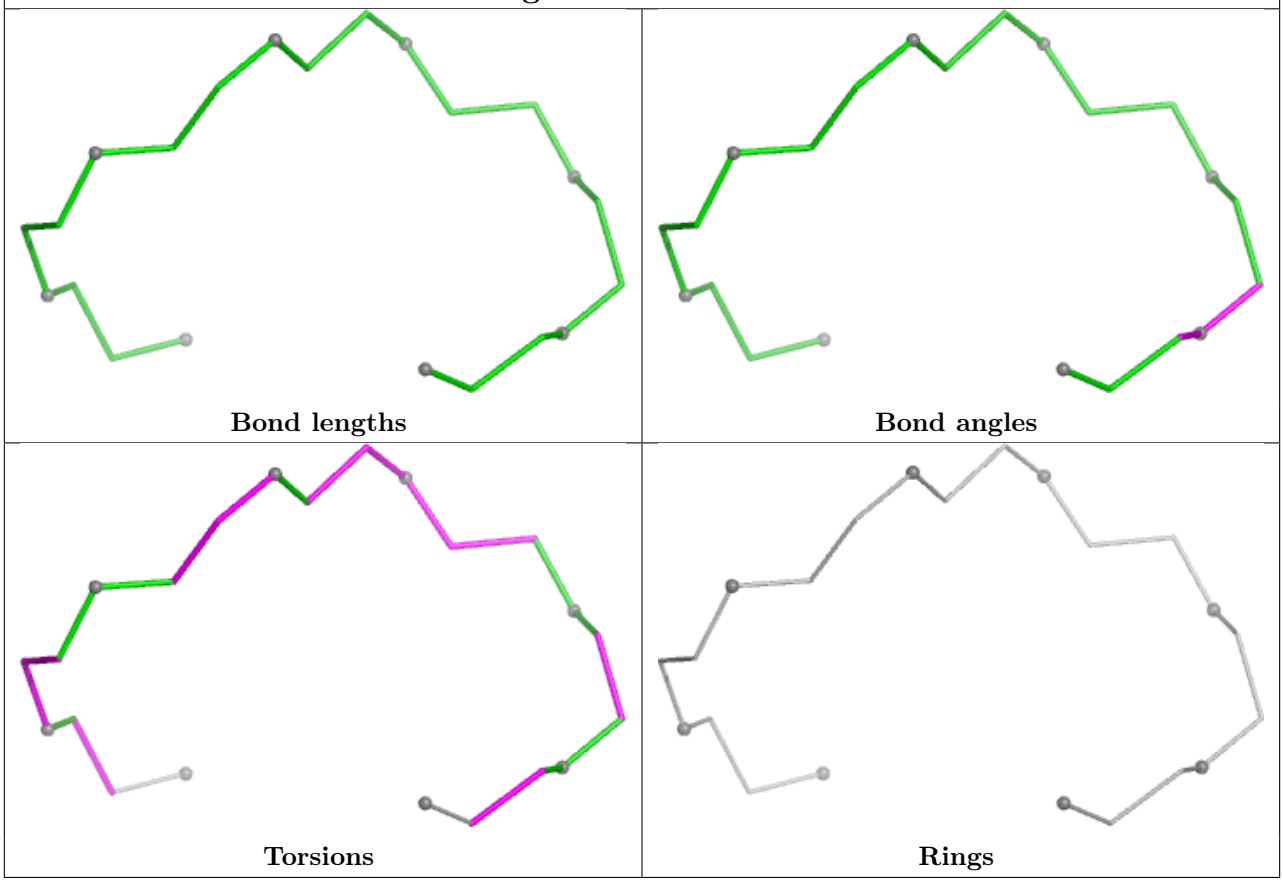


Rings

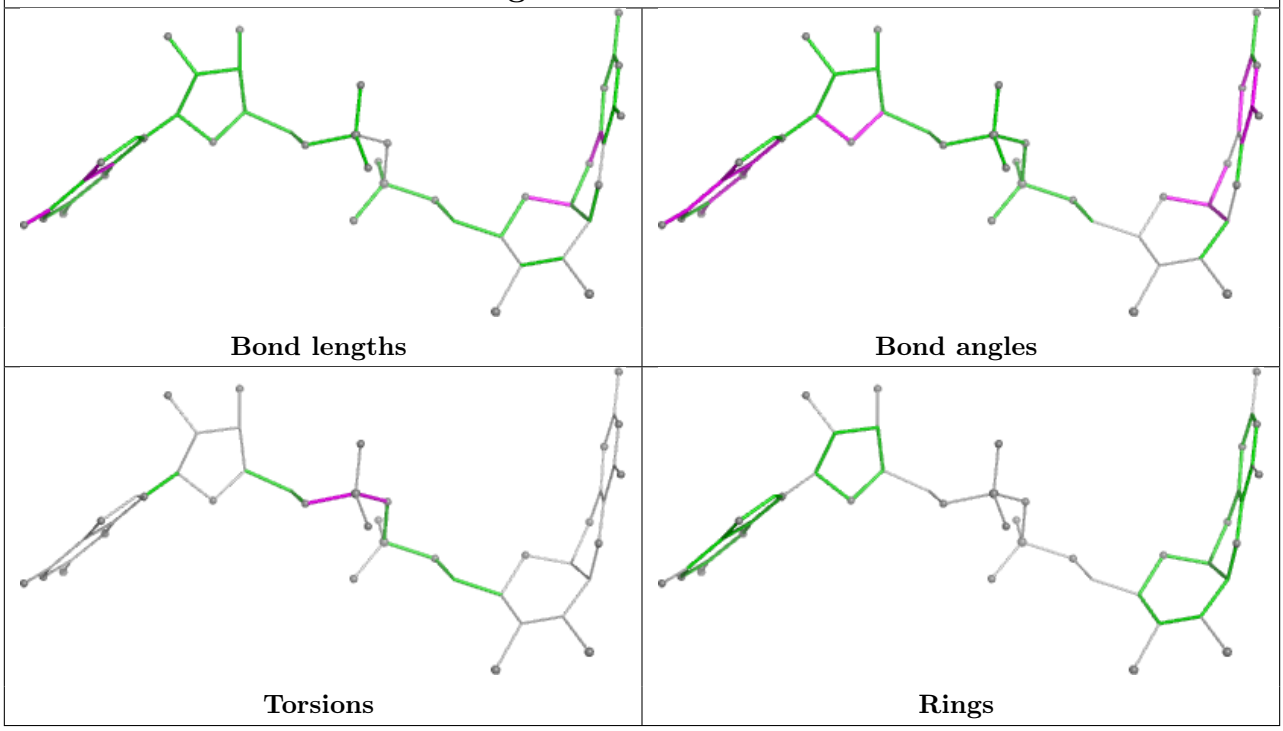
Ligand MGD G 2001



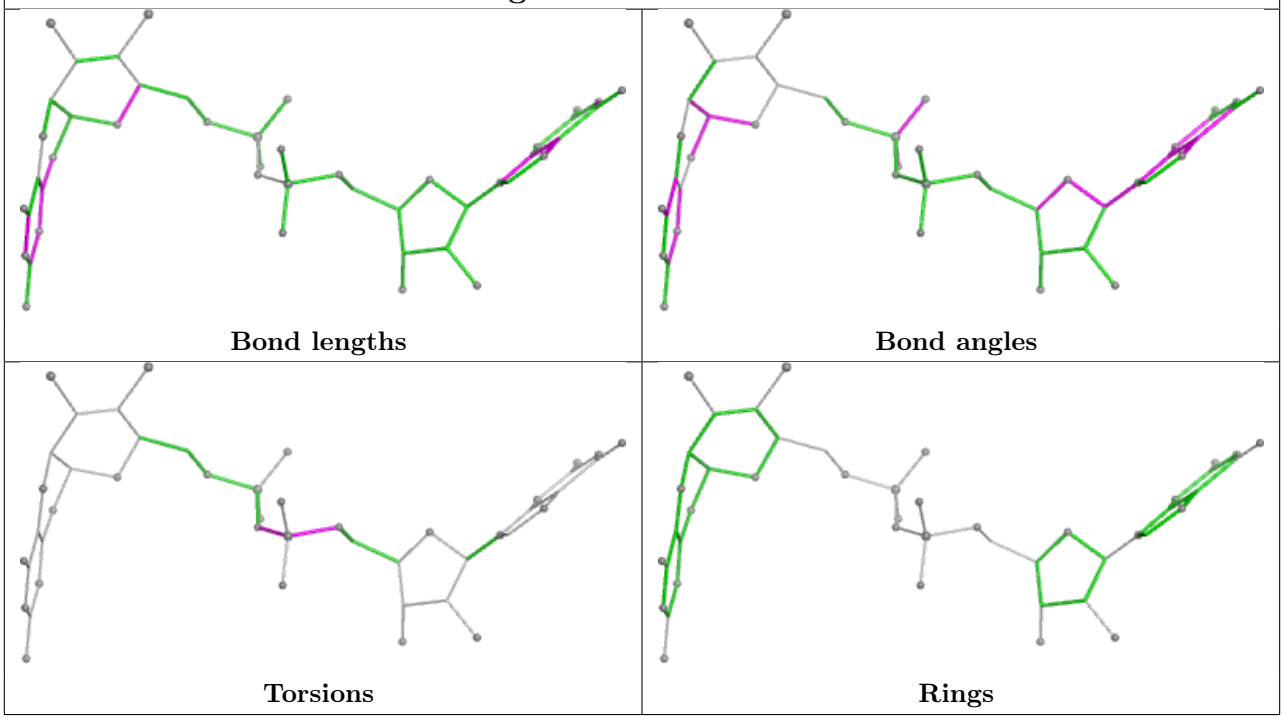
Ligand P33 G 2009



Ligand MGD A 2002



Ligand MGD C 2002



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	843/843 (100%)	-0.55	7 (0%)	82	80	11, 21, 39, 65	0
1	C	843/843 (100%)	-0.40	4 (0%)	87	85	14, 24, 43, 73	0
1	E	843/843 (100%)	-0.38	5 (0%)	85	83	13, 24, 44, 72	0
1	G	843/843 (100%)	-0.52	5 (0%)	85	83	12, 22, 40, 72	0
2	B	132/132 (100%)	-0.11	0	100	100	19, 32, 51, 67	0
2	D	132/132 (100%)	-0.13	2 (1%)	72	69	20, 31, 50, 68	0
2	F	132/132 (100%)	0.29	2 (1%)	72	69	22, 36, 57, 65	0
2	H	132/132 (100%)	-0.23	1 (0%)	82	80	19, 29, 46, 60	0
All	All	3900/3900 (100%)	-0.40	26 (0%)	84	82	11, 24, 45, 73	0

The worst 5 of 26 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	G	340	GLY	3.9
1	E	49	GLN	3.1
1	G	41	ASN	3.0
1	C	41	ASN	2.9
2	F	44	ALA	2.9

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 oligosaccharides 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
11	EDO	G	2012	4/4	0.44	0.42	99,99,100,100	0
7	SO4	C	2009	5/5	0.64	0.18	64,72,76,87	0
9	GOL	A	2011	6/6	0.74	0.24	49,58,60,63	0
12	PEG	C	2010	7/7	0.77	0.24	49,53,60,65	0
9	GOL	C	2011	6/6	0.79	0.20	49,55,57,60	0
9	GOL	E	2008	6/6	0.80	0.20	47,55,61,62	0
9	GOL	G	2010	6/6	0.81	0.16	29,39,40,43	0
9	GOL	G	2011	6/6	0.81	0.16	35,43,47,48	0
7	SO4	E	2007	5/5	0.82	0.13	66,69,74,81	0
7	SO4	G	2007	5/5	0.85	0.11	54,64,68,72	0
7	SO4	C	2008	5/5	0.86	0.11	66,72,76,76	0
11	EDO	B	202	4/4	0.86	0.18	45,47,49,49	0
8	PGE	A	2010	10/10	0.87	0.14	29,35,36,38	0
13	P33	G	2009	22/22	0.88	0.15	37,47,51,54	0
13	P33	G	2008	22/22	0.90	0.13	26,37,43,44	0
7	SO4	C	2007	5/5	0.90	0.13	57,57,64,67	0
7	SO4	A	2007	5/5	0.91	0.14	56,57,61,69	0
7	SO4	A	2006	5/5	0.92	0.14	50,57,66,66	0
11	EDO	C	2012	4/4	0.92	0.12	31,34,36,36	0
7	SO4	E	2006	5/5	0.93	0.08	61,65,68,73	0
7	SO4	A	2008	5/5	0.97	0.08	26,32,36,36	0
7	SO4	A	2009	5/5	0.98	0.06	29,32,34,35	0
7	SO4	C	2006	5/5	0.98	0.07	26,28,33,34	0
3	MGD	E	2001	47/47	0.98	0.04	14,17,20,22	0
3	MGD	E	2002	47/47	0.98	0.04	18,20,22,24	0
4	O	E	2003	1/1	0.98	0.08	32,32,32,32	0
7	SO4	G	2006	5/5	0.99	0.05	29,30,31,33	0
6	F3S	C	2005	7/7	0.99	0.03	21,22,24,24	0
6	F3S	E	2005	7/7	0.99	0.03	23,24,26,31	0
6	F3S	G	2005	7/7	0.99	0.03	20,22,25,25	0
3	MGD	C	2002	47/47	0.99	0.04	18,20,22,24	0
3	MGD	A	2001	47/47	0.99	0.04	13,16,17,18	0
3	MGD	A	2002	47/47	0.99	0.04	11,14,17,19	0
3	MGD	G	2001	47/47	0.99	0.04	14,18,19,21	0
10	FES	B	201	4/4	0.99	0.02	22,24,25,27	0
10	FES	D	201	4/4	0.99	0.03	27,28,29,31	0
10	FES	H	201	4/4	0.99	0.03	25,25,27,28	0

Continued on next page...

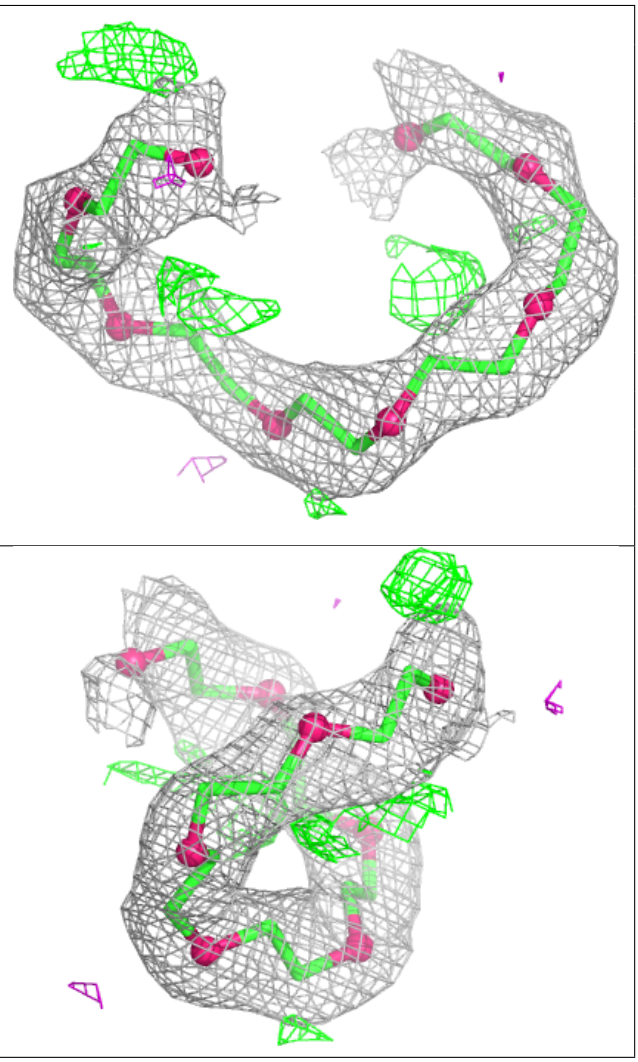
Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	MGD	G	2002	47/47	0.99	0.04	14,17,19,20	0
4	O	A	2003	1/1	0.99	0.04	17,17,17,17	0
4	O	C	2003	1/1	0.99	0.07	22,22,22,22	0
3	MGD	C	2001	47/47	0.99	0.04	15,18,20,21	0
4	O	G	2003	1/1	0.99	0.05	23,23,23,23	0
6	F3S	A	2005	7/7	0.99	0.03	19,20,22,24	0
5	4MO	E	2004	1/1	1.00	0.01	23,23,23,23	0
5	4MO	G	2004	1/1	1.00	0.01	22,22,22,22	0
10	FES	F	201	4/4	1.00	0.02	29,30,31,31	0
5	4MO	A	2004	1/1	1.00	0.01	18,18,18,18	0
5	4MO	C	2004	1/1	1.00	0.01	23,23,23,23	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.

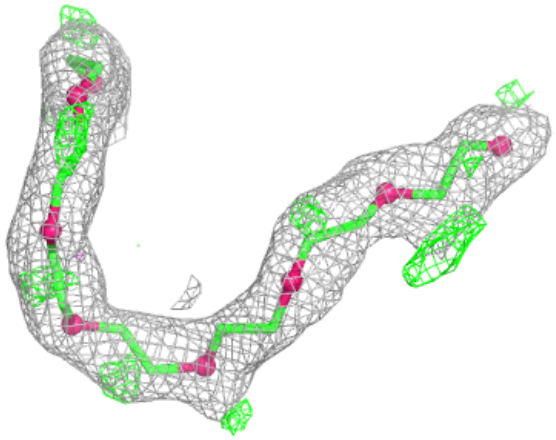
Electron density around P33 G 2009:

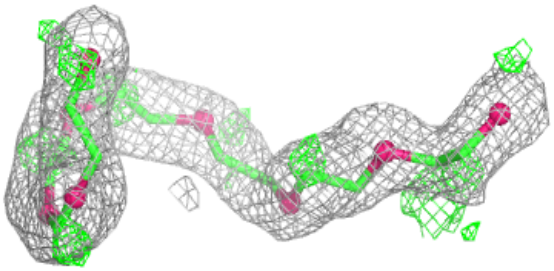
2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)

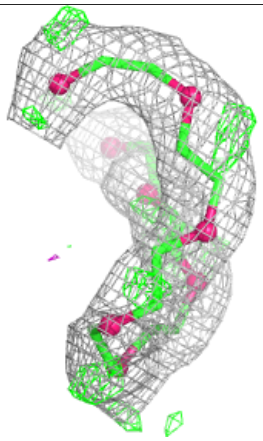


Electron density around P33 G 2008:

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

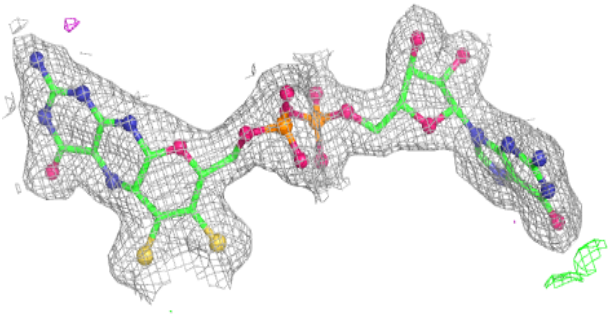


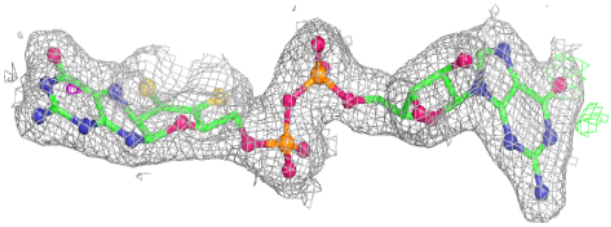


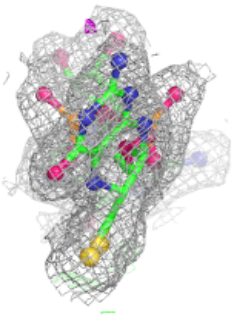


Electron density around MGD E 2001:

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

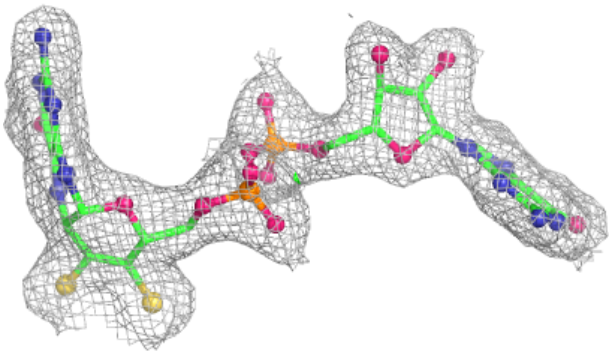


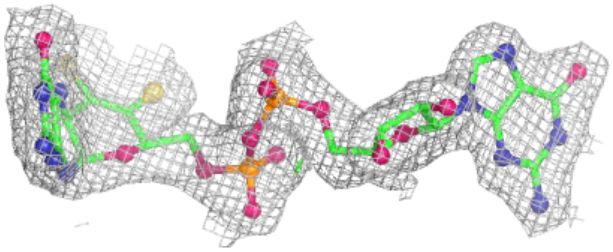


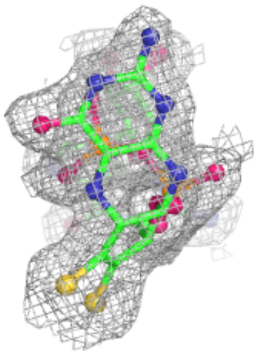


Electron density around MGD E 2002:

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

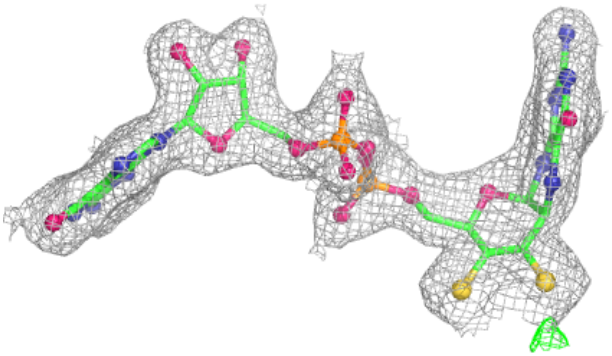


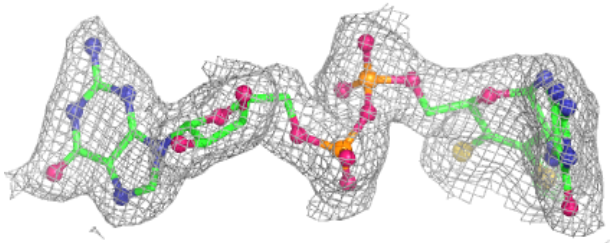


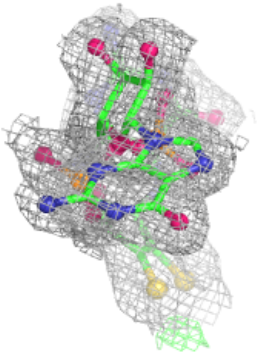


Electron density around MGD C 2002:

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

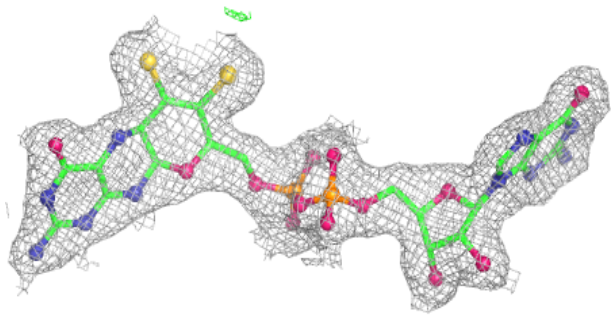


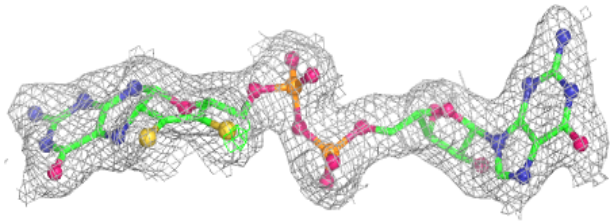


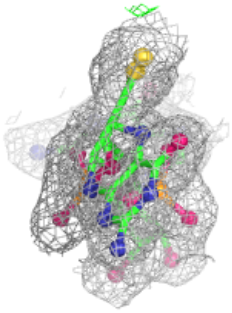


Electron density around MGD A 2001:

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

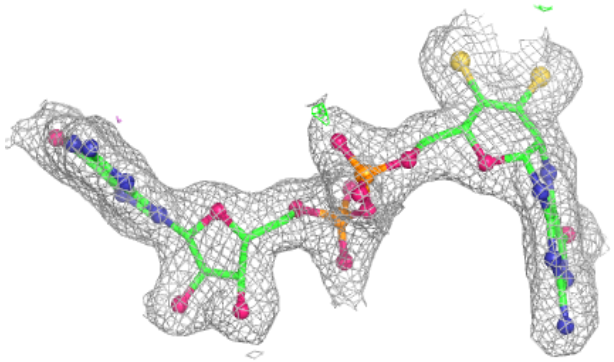


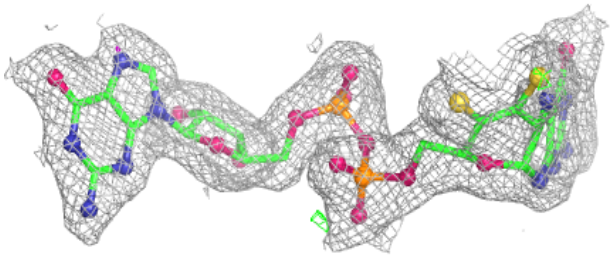


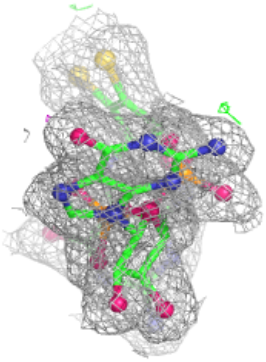


Electron density around MGD A 2002:

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

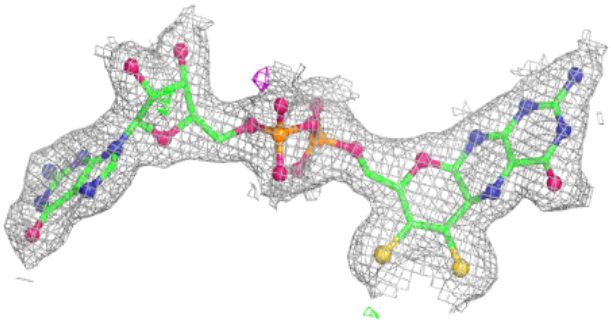


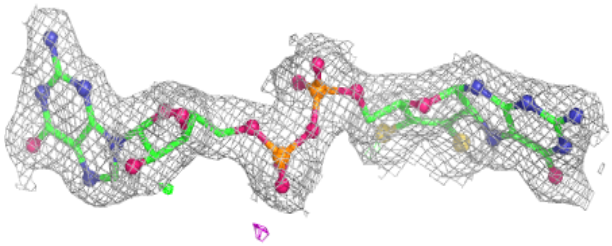


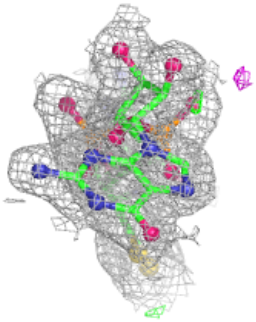


Electron density around MGD G 2001:

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

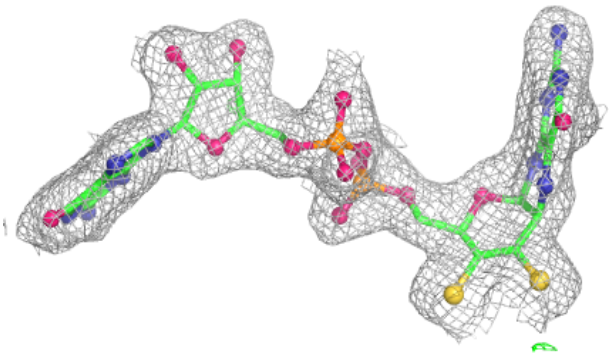


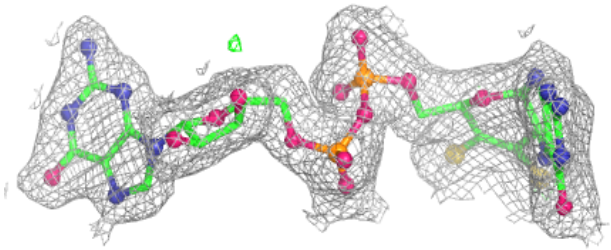


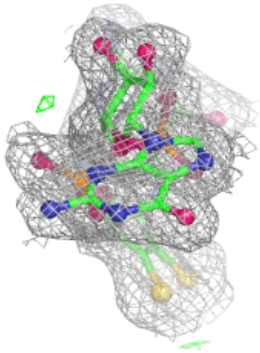


Electron density around MGD G 2002:

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

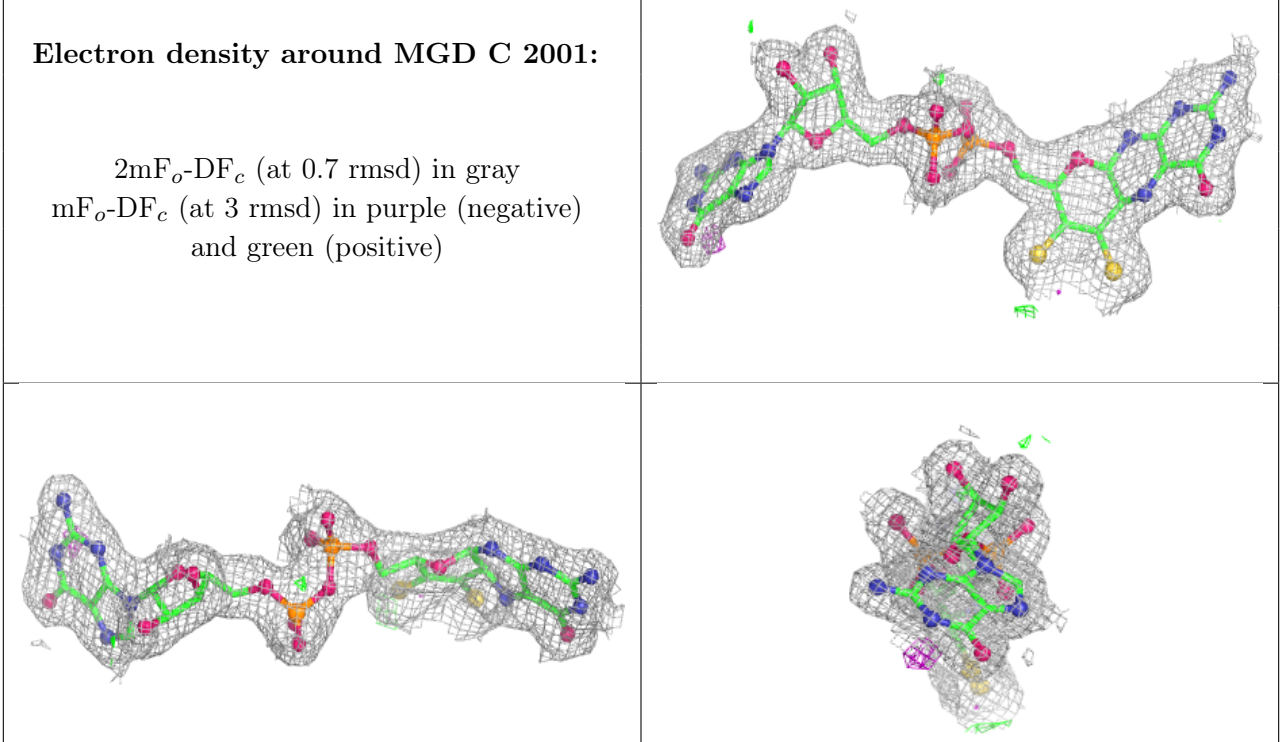






Electron density around MGD C 2001:

$2mF_o$-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.