



# wwPDB X-ray Structure Validation Summary Report ⓘ

Sep 25, 2018 – 04:24 PM EDT

PDB ID : 5NQD  
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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : rb-20031172  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20031172

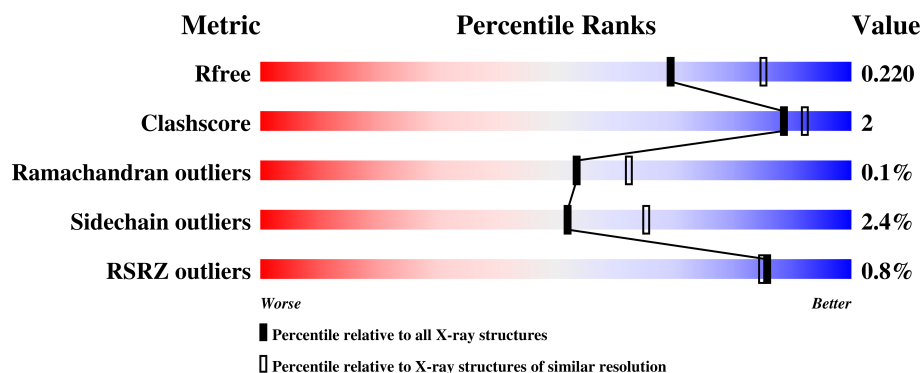
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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))
$R_{free}$	111664	4343 (2.20-2.20)
Clashscore	122126	5027 (2.20-2.20)
Ramachandran outliers	120053	4952 (2.20-2.20)
Sidechain outliers	120020	4953 (2.20-2.20)
RSRZ outliers	108989	4245 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	843	<div> <div>91%</div> <div>8%</div> </div>
1	C	843	<div> <div>2%</div> <div>93%</div> <div>6%</div> </div>
1	E	843	<div> <div>92%</div> <div>7%</div> </div>
1	G	843	<div> <div>93%</div> <div>6%</div> </div>
2	B	132	<div> <div>2%</div> <div>96%</div> </div>

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
2	D	132	<div> <div>5%</div> <div>97%</div> <div>•</div> </div>
2	F	132	<div> <div>10%</div> <div>91%</div> <div>8% •</div> </div>
2	H	132	<div> <div>%</div> <div>93%</div> <div>7%</div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
11	EDO	G	2012	-	-	-	X

## 2 Entry composition

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
1	A	843	Total	C	N	O	S	0	0	0
			6545	4088	1167	1253	37			
1	C	843	Total	C	N	O	S	0	0	0
			6545	4088	1167	1253	37			
1	E	843	Total	C	N	O	S	0	0	0
			6545	4088	1167	1253	37			
1	G	843	Total	C	N	O	S	0	0	0
			6545	4088	1167	1253	37			

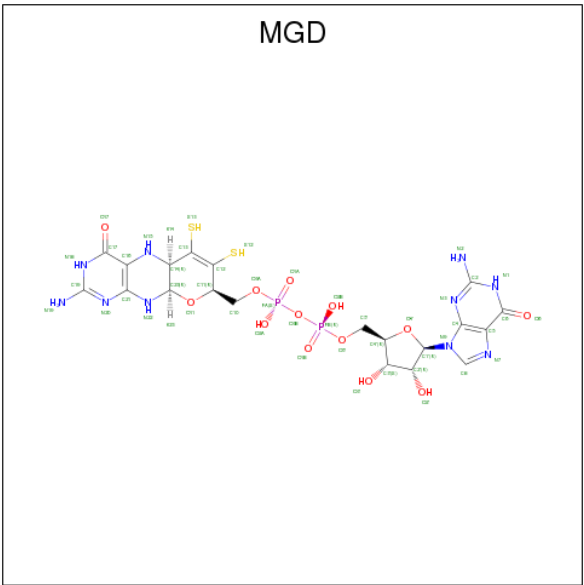
- Molecule 2 is a protein called Arsenite oxidase small subunit AioB Rieske [2Fe-2S] cluster.

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

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 (three-letter code: MGD) (formula: C<sub>20</sub>H<sub>26</sub>N<sub>10</sub>O<sub>13</sub>P<sub>2</sub>S<sub>2</sub>).



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

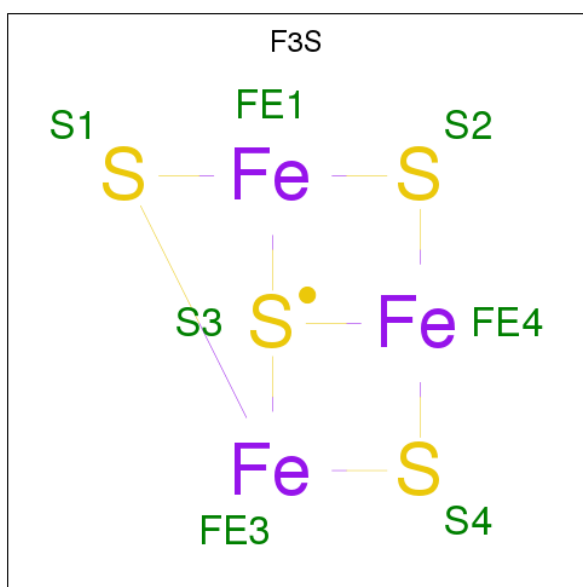
- Molecule 4 is OXYGEN ATOM (three-letter code: O) (formula: O).

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

- Molecule 5 is MOLYBDENUM(IV) ION (three-letter code: 4MO) (formula: Mo).

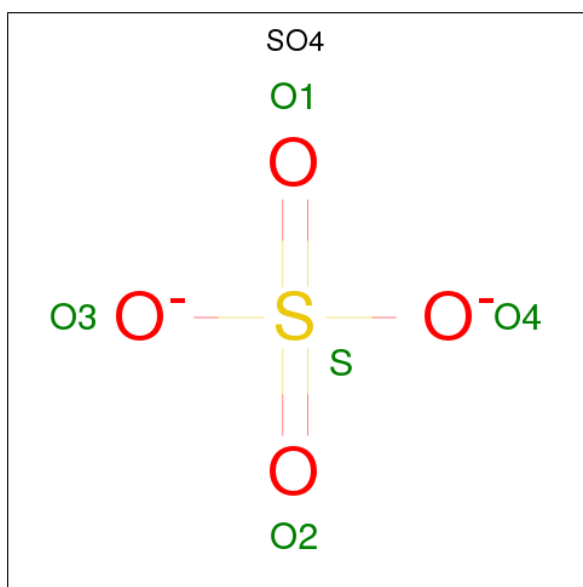
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	G	1	Total 1	Mo 1	0	0
5	A	1	Total 1	Mo 1	0	0
5	C	1	Total 1	Mo 1	0	0
5	E	1	Total 1	Mo 1	0	0

- Molecule 6 is FE3-S4 CLUSTER (three-letter code: F3S) (formula: Fe<sub>3</sub>S<sub>4</sub>).



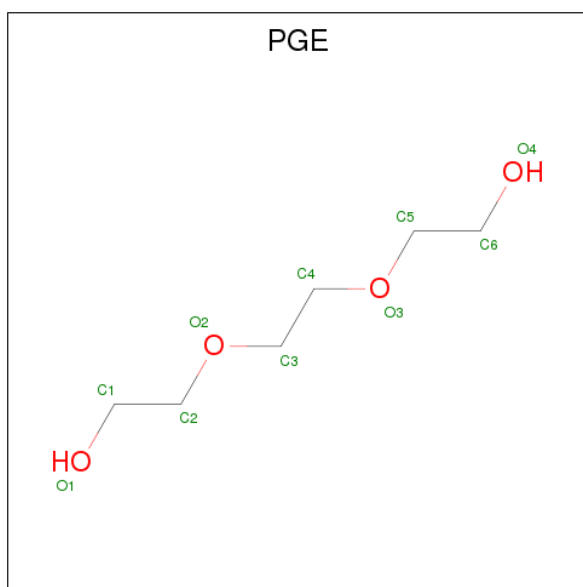
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total 7	Fe 3	S 4	0	0
6	C	1	Total 7	Fe 3	S 4	0	0
6	E	1	Total 7	Fe 3	S 4	0	0
6	G	1	Total 7	Fe 3	S 4	0	0

- Molecule 7 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



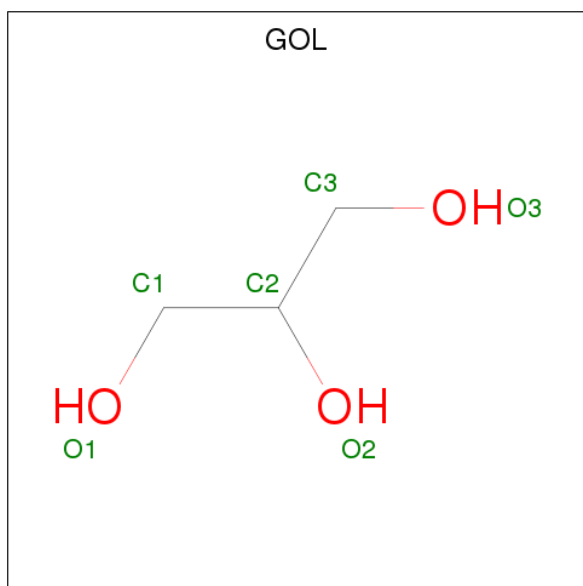
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total	O	S	0	0
			5	4	1		
7	A	1	Total	O	S	0	0
			5	4	1		
7	A	1	Total	O	S	0	0
			5	4	1		
7	A	1	Total	O	S	0	0
			5	4	1		
7	C	1	Total	O	S	0	0
			5	4	1		
7	C	1	Total	O	S	0	0
			5	4	1		
7	C	1	Total	O	S	0	0
			5	4	1		
7	E	1	Total	O	S	0	0
			5	4	1		
7	E	1	Total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

- Molecule 8 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: C<sub>6</sub>H<sub>14</sub>O<sub>4</sub>).



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

- Molecule 9 is GLYCEROL (three-letter code: GOL) (formula:  $C_3H_8O_3$ ).



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

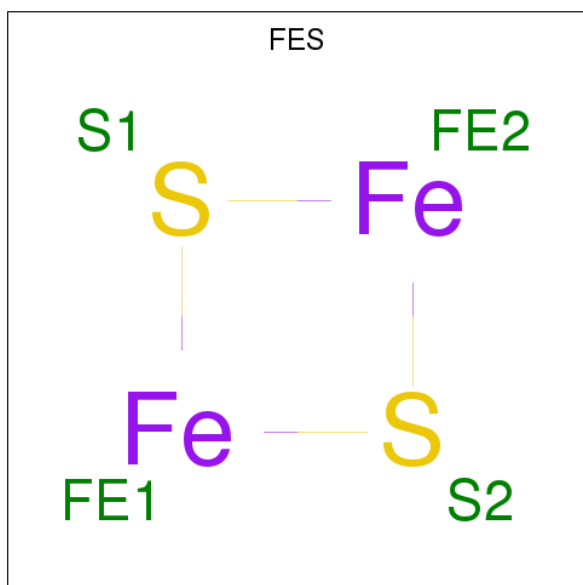
*Continued on next page...*



Continued from previous page...

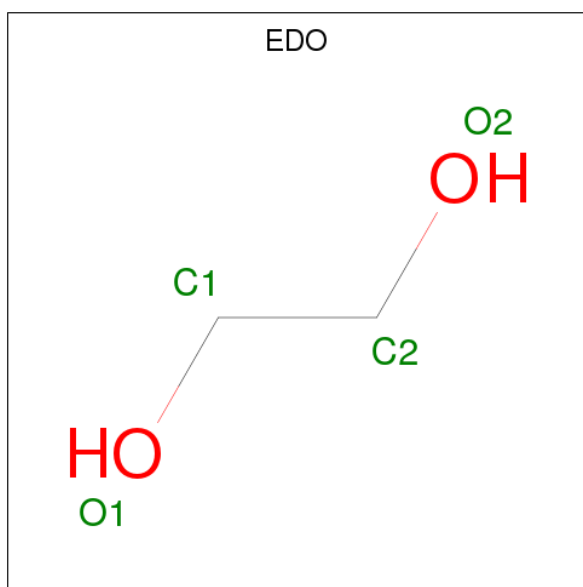
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
9	G	1	Total	C	O	0	0
			6	3	3		
9	G	1	Total	C	O	0	0
			6	3	3		

- Molecule 10 is FE2/S2 (INORGANIC) CLUSTER (three-letter code: FES) (formula:  $\text{Fe}_2\text{S}_2$ ).



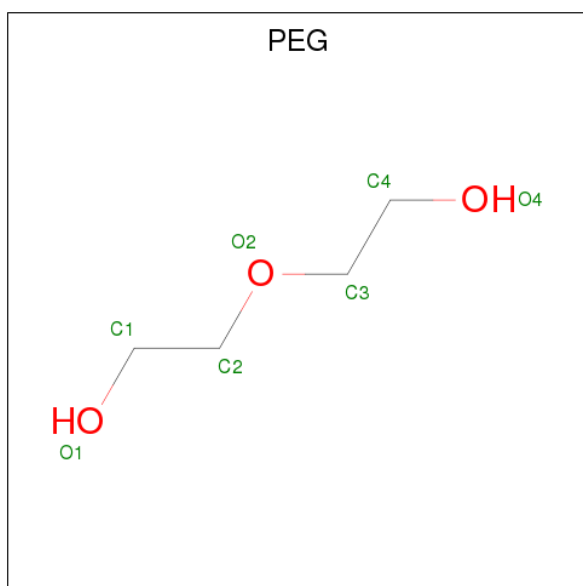
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
10	B	1	Total	Fe	S	0	0
			4	2	2		
10	D	1	Total	Fe	S	0	0
			4	2	2		
10	F	1	Total	Fe	S	0	0
			4	2	2		
10	H	1	Total	Fe	S	0	0
			4	2	2		

- Molecule 11 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula:  $\text{C}_2\text{H}_6\text{O}_2$ ).



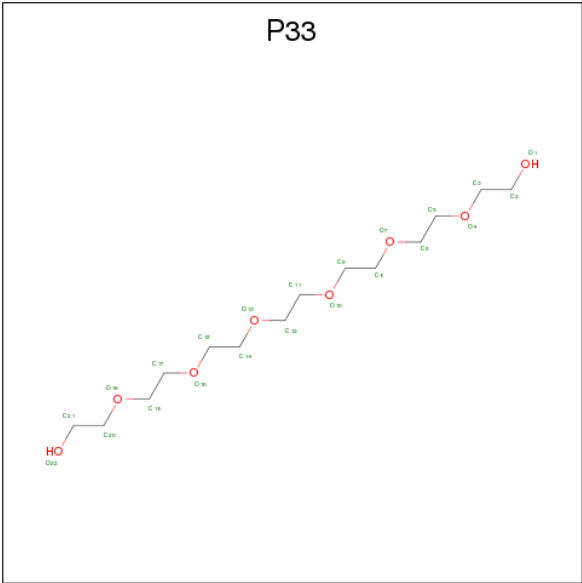
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
11	B	1	Total	C	O	0	0
			4	2	2		
11	C	1	Total	C	O	0	0
			4	2	2		
11	G	1	Total	C	O	0	0
			4	2	2		

- Molecule 12 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula:  $C_4H_{10}O_3$ ).



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

- Molecule 13 is 3,6,9,12,15,18-HEXAOSAICOSANE-1,20-DIOL (three-letter code: P33) (formula: C<sub>14</sub>H<sub>30</sub>O<sub>8</sub>).



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

- Molecule 14 is water.

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

Continued on next page...

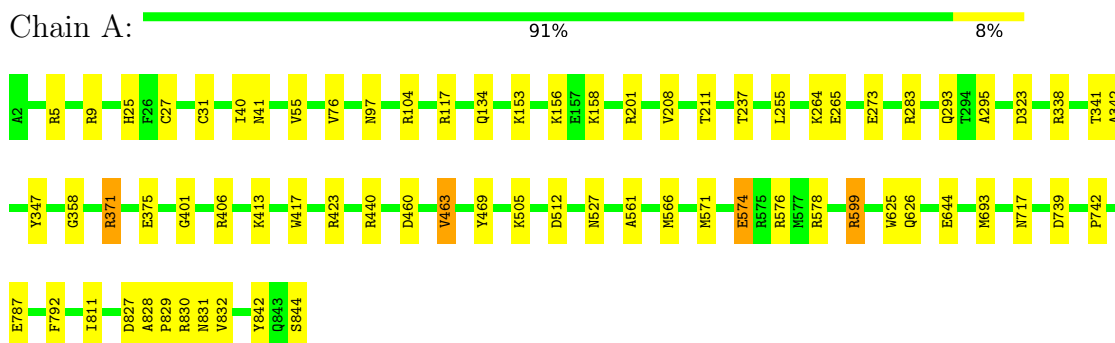
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
14	G	374	Total 374	O 374	0	0
14	H	36	Total 36	O 36	0	0

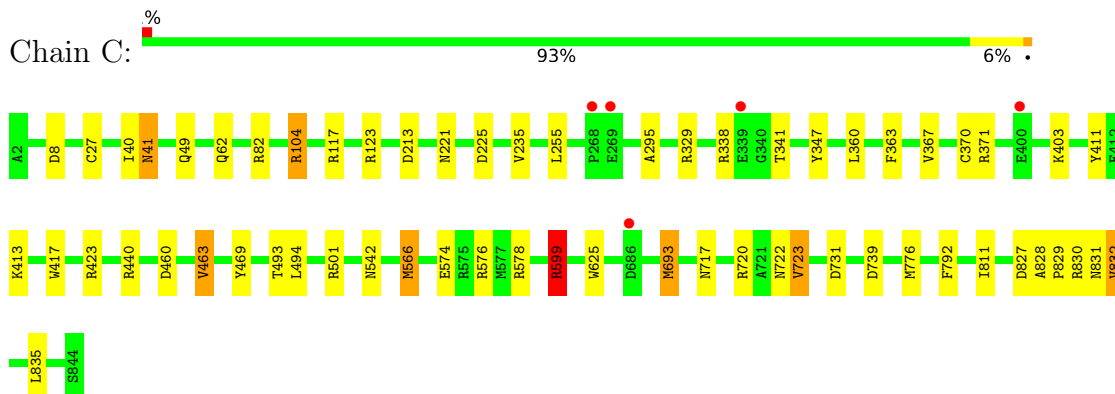
### 3 Residue-property plots [i](#)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are color-coded according to the number of geometric quality criteria for which they contain at least one outlier: green = 0, yellow = 1, orange = 2 and red = 3 or more. A red dot above a residue indicates a poor fit to the electron density ( $RSRZ > 2$ ). Stretches of 2 or more consecutive residues without any outlier are shown as a green connector. Residues present in the sample, but not in the model, are shown in grey.

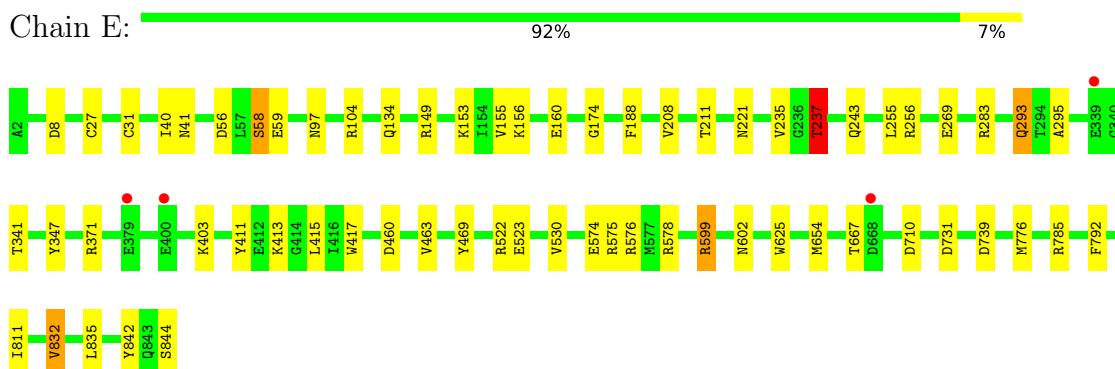
#### • Molecule 1: AroA



#### • Molecule 1: AroA



#### • Molecule 1: AroA



- 
- Sequence logo for the 10th position. The y-axis represents information content in bits, ranging from 0 to 1.2. The x-axis shows the amino acid sequence: A44, V47, L54, Y70, P71, R85, N116, R151, D154, A160, L175. The D154 position has a red dot above it, indicating a high information content.

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	141.41Å 148.66Å 232.01Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	49.55 – 2.20 49.24 – 2.20	Depositor EDS
% Data completeness (in resolution range)	100.0 (49.55-2.20) 100.0 (49.24-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	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 <sup>2</sup> )	22.6	Xtriage
Anisotropy	0.040	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 30.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	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 <sup>2</sup> )	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.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical 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

### 5.1 Standard geometry

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

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	0.83	1/6693 (0.0%)	0.92	18/9068 (0.2%)
1	C	0.81	0/6693	0.95	22/9068 (0.2%)
1	E	0.82	1/6693 (0.0%)	0.93	16/9068 (0.2%)
1	G	0.81	1/6693 (0.0%)	0.91	15/9068 (0.2%)
2	B	0.71	0/1011	0.82	1/1378 (0.1%)
2	D	0.74	0/1011	0.83	0/1378
2	F	0.73	0/1011	0.83	1/1378 (0.1%)
2	H	0.74	0/1011	0.85	1/1378 (0.1%)
All	All	0.80	3/30816 (0.0%)	0.92	74/41784 (0.2%)

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

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	375	GLU	CG-CD	5.37	1.60	1.51
1	E	523	GLU	CG-CD	5.34	1.59	1.51
1	G	787	GLU	CD-OE1	5.11	1.31	1.25

The worst 5 of 74 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	599	ARG	NE-CZ-NH2	-14.57	113.02	120.30
1	A	104	ARG	NE-CZ-NH2	-13.43	113.59	120.30

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	599	ARG	NE-CZ-NH1	13.06	126.83	120.30
1	G	104	ARG	NE-CZ-NH2	-12.18	114.21	120.30
1	A	599	ARG	NE-CZ-NH2	-10.54	115.03	120.30

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	34	0
1	C	6545	0	6296	27	0
1	E	6545	0	6296	23	0
1	G	6545	0	6296	29	1
2	B	989	0	943	1	0
2	D	989	0	943	2	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	1	0
4	C	1	0	0	1	0
4	E	1	0	0	0	0
4	G	1	0	0	1	0
5	A	1	0	0	1	0
5	C	1	0	0	1	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
6	E	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	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	2	0
14	A	383	0	0	7	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	7	1
14	H	36	0	0	0	0
All	All	32195	0	29274	128	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 128 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: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
1:C:574:GLU:OE1	1:C:576:ARG:NH1	2.09	0.84

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%)	53	62
1	C	841/843 (100%)	805 (96%)	35 (4%)	1 (0%)	53	62
1	E	841/843 (100%)	807 (96%)	33 (4%)	1 (0%)	53	62
1	G	841/843 (100%)	804 (96%)	36 (4%)	1 (0%)	53	62
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%)	53	62

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 ⓘ

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%)	660 (98%)	17 (2%)	50	63
1	C	677/677 (100%)	661 (98%)	16 (2%)	52	65
1	E	677/677 (100%)	658 (97%)	19 (3%)	47	59
1	G	677/677 (100%)	661 (98%)	16 (2%)	52	65
2	B	104/104 (100%)	102 (98%)	2 (2%)	60	74
2	D	104/104 (100%)	103 (99%)	1 (1%)	78	88
2	F	104/104 (100%)	101 (97%)	3 (3%)	45	58
2	H	104/104 (100%)	102 (98%)	2 (2%)	60	74
All	All	3124/3124 (100%)	3048 (98%)	76 (2%)	52	65

5 of 76 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	832	VAL
1	E	256	ARG
1	G	723	VAL
2	D	115	ASP
1	E	59	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	25	HIS
1	A	619	GLN
1	E	425	ASN
1	G	249	ASN
1	G	637	ASN

### 5.3.3 RNA ⓘ

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 carbohydrates 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	$\# Z  > 2$
3	MGD	A	2001	5	41,52,52	1.35	6 (14%)	39,81,81	2.17	9 (23%)
3	MGD	A	2002	5	41,52,52	1.38	6 (14%)	39,81,81	2.27	11 (28%)
6	F3S	A	2005	1	0,9,9	0.00	-	0,15,15	0.00	-
7	SO4	A	2006	-	4,4,4	0.54	0	6,6,6	0.85	0
7	SO4	A	2007	-	4,4,4	0.50	0	6,6,6	0.86	0
7	SO4	A	2008	-	4,4,4	0.87	0	6,6,6	1.06	0
7	SO4	A	2009	-	4,4,4	0.55	0	6,6,6	0.96	0
8	PGE	A	2010	-	9,9,9	1.12	0	8,8,8	1.27	0
9	GOL	A	2011	-	5,5,5	1.08	0	5,5,5	1.44	1 (20%)
10	FES	B	201	2	0,4,4	0.00	-	0,4,4	0.00	-
11	EDO	B	202	-	3,3,3	0.40	0	2,2,2	0.87	0
3	MGD	C	2001	5	41,52,52	1.57	6 (14%)	39,81,81	2.17	15 (38%)
3	MGD	C	2002	5	41,52,52	1.44	6 (14%)	39,81,81	2.38	12 (30%)
6	F3S	C	2005	1	0,9,9	0.00	-	0,15,15	0.00	-
7	SO4	C	2006	-	4,4,4	0.68	0	6,6,6	0.74	0
7	SO4	C	2007	-	4,4,4	0.66	0	6,6,6	0.80	0
7	SO4	C	2008	-	4,4,4	0.53	0	6,6,6	0.22	0
7	SO4	C	2009	-	4,4,4	0.76	0	6,6,6	0.55	0
12	PEG	C	2010	-	6,6,6	0.64	0	5,5,5	0.62	0
9	GOL	C	2011	-	5,5,5	0.70	0	5,5,5	0.99	0
11	EDO	C	2012	-	3,3,3	0.49	0	2,2,2	0.49	0

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
10	FES	D	201	2	0,4,4	0.00	-	0,4,4	0.00	-
3	MGD	E	2001	5	41,52,52	1.70	7 (17%)	39,81,81	2.45	15 (38%)
3	MGD	E	2002	5	41,52,52	1.48	5 (12%)	39,81,81	2.62	15 (38%)
6	F3S	E	2005	1	0,9,9	0.00	-	0,15,15	0.00	-
7	SO4	E	2006	-	4,4,4	0.41	0	6,6,6	0.45	0
7	SO4	E	2007	-	4,4,4	0.57	0	6,6,6	0.55	0
9	GOL	E	2008	-	5,5,5	0.48	0	5,5,5	0.38	0
10	FES	F	201	2	0,4,4	0.00	-	0,4,4	0.00	-
3	MGD	G	2001	5	41,52,52	1.57	5 (12%)	39,81,81	2.31	16 (41%)
3	MGD	G	2002	5	41,52,52	1.36	6 (14%)	39,81,81	2.22	14 (35%)
6	F3S	G	2005	1	0,9,9	0.00	-	0,15,15	0.00	-
7	SO4	G	2006	-	4,4,4	0.51	0	6,6,6	0.73	0
7	SO4	G	2007	-	4,4,4	0.64	0	6,6,6	0.53	0
13	P33	G	2008	-	21,21,21	1.01	0	20,20,20	1.51	4 (20%)
13	P33	G	2009	-	21,21,21	0.68	0	20,20,20	0.81	1 (5%)
9	GOL	G	2010	-	5,5,5	1.31	1 (20%)	5,5,5	1.73	1 (20%)
9	GOL	G	2011	-	5,5,5	1.11	0	5,5,5	1.83	1 (20%)
11	EDO	G	2012	-	3,3,3	0.65	0	2,2,2	0.43	0
10	FES	H	201	2	0,4,4	0.00	-	0,4,4	0.00	-

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	MGD	A	2001	5	-	0/18/66/66	0/6/6/6
3	MGD	A	2002	5	-	0/18/66/66	0/6/6/6
6	F3S	A	2005	1	-	0/0/24/24	0/3/3/3
7	SO4	A	2006	-	-	0/0/0/0	0/0/0/0
7	SO4	A	2007	-	-	0/0/0/0	0/0/0/0
7	SO4	A	2008	-	-	0/0/0/0	0/0/0/0
7	SO4	A	2009	-	-	0/0/0/0	0/0/0/0
8	PGE	A	2010	-	-	0/7/7/7	0/0/0/0
9	GOL	A	2011	-	-	0/4/4/4	0/0/0/0
10	FES	B	201	2	-	0/0/4/4	0/1/1/1
11	EDO	B	202	-	-	0/1/1/1	0/0/0/0
3	MGD	C	2001	5	-	0/18/66/66	0/6/6/6
3	MGD	C	2002	5	-	0/18/66/66	0/6/6/6
6	F3S	C	2005	1	-	0/0/24/24	0/3/3/3
7	SO4	C	2006	-	-	0/0/0/0	0/0/0/0

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
7	SO4	C	2007	-	-	0/0/0/0	0/0/0/0
7	SO4	C	2008	-	-	0/0/0/0	0/0/0/0
7	SO4	C	2009	-	-	0/0/0/0	0/0/0/0
12	PEG	C	2010	-	-	0/4/4/4	0/0/0/0
9	GOL	C	2011	-	-	0/4/4/4	0/0/0/0
11	EDO	C	2012	-	-	0/1/1/1	0/0/0/0
10	FES	D	201	2	-	0/0/4/4	0/1/1/1
3	MGD	E	2001	5	-	0/18/66/66	0/6/6/6
3	MGD	E	2002	5	-	0/18/66/66	0/6/6/6
6	F3S	E	2005	1	-	0/0/24/24	0/3/3/3
7	SO4	E	2006	-	-	0/0/0/0	0/0/0/0
7	SO4	E	2007	-	-	0/0/0/0	0/0/0/0
9	GOL	E	2008	-	-	0/4/4/4	0/0/0/0
10	FES	F	201	2	-	0/0/4/4	0/1/1/1
3	MGD	G	2001	5	-	0/18/66/66	0/6/6/6
3	MGD	G	2002	5	-	0/18/66/66	0/6/6/6
6	F3S	G	2005	1	-	0/0/24/24	0/3/3/3
7	SO4	G	2006	-	-	0/0/0/0	0/0/0/0
7	SO4	G	2007	-	-	0/0/0/0	0/0/0/0
13	P33	G	2008	-	-	0/19/19/19	0/0/0/0
13	P33	G	2009	-	-	0/19/19/19	0/0/0/0
9	GOL	G	2010	-	-	0/4/4/4	0/0/0/0
9	GOL	G	2011	-	-	0/4/4/4	0/0/0/0
11	EDO	G	2012	-	-	0/1/1/1	0/0/0/0
10	FES	H	201	2	-	0/0/4/4	0/1/1/1

The worst 5 of 48 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	C	2002	MGD	O11-C11	-2.95	1.39	1.43
3	G	2002	MGD	C14-N15	-2.70	1.41	1.45
3	G	2001	MGD	C16-N15	-2.56	1.32	1.38
3	C	2001	MGD	O11-C23	-2.54	1.40	1.43
3	C	2002	MGD	C8-N9	-2.47	1.33	1.36

The worst 5 of 115 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	E	2002	MGD	C4'-O4'-C1'	-9.22	100.21	109.83
3	G	2002	MGD	C4'-O4'-C1'	-6.72	102.82	109.83
3	C	2001	MGD	C6-C5-C4	-6.08	114.88	120.85
3	E	2001	MGD	C6-C5-C4	-6.07	114.89	120.85

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	C	2002	MGD	C4'-O4'-C1'	-6.01	103.57	109.83

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6 monomers are involved in 8 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	2002	MGD	1	0
7	A	2008	SO4	1	0
3	C	2002	MGD	1	0
3	E	2001	MGD	1	0
3	G	2001	MGD	2	0
13	G	2008	P33	2	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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	843/843 (100%)	-0.60	0 100 100	11, 21, 39, 65	0
1	C	843/843 (100%)	-0.52	5 (0%) 89 88	14, 24, 43, 73	0
1	E	843/843 (100%)	-0.59	4 (0%) 90 90	13, 24, 44, 72	0
1	G	843/843 (100%)	-0.60	1 (0%) 95 95	12, 22, 40, 72	0
2	B	132/132 (100%)	-0.18	3 (2%) 60 58	19, 32, 51, 67	0
2	D	132/132 (100%)	-0.08	6 (4%) 33 32	20, 31, 50, 68	0
2	F	132/132 (100%)	0.28	13 (9%) 7 6	22, 36, 57, 65	0
2	H	132/132 (100%)	-0.30	1 (0%) 86 85	19, 29, 46, 60	0
All	All	3900/3900 (100%)	-0.51	33 (0%) 86 85	11, 24, 45, 73	0

The worst 5 of 33 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	57	ILE	4.4
1	G	340	GLY	4.2
2	F	58	SER	4.0
2	D	57	ILE	3.2
2	F	92	PRO	3.2

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

There are no non-standard protein/DNA/RNA residues in this entry.

### 6.3 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 6.4 Ligands ⓘ

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
11	EDO	G	2012	4/4	0.24	0.98	99,99,100,100	0
9	GOL	A	2011	6/6	0.62	0.27	49,58,60,63	0
7	SO4	C	2009	5/5	0.75	0.35	64,72,76,87	0
9	GOL	G	2011	6/6	0.75	0.18	35,43,47,48	0
12	PEG	C	2010	7/7	0.76	0.22	49,53,60,65	0
9	GOL	C	2011	6/6	0.78	0.23	49,55,57,60	0
9	GOL	G	2010	6/6	0.83	0.30	29,39,40,43	0
9	GOL	E	2008	6/6	0.85	0.16	47,55,61,62	0
13	P33	G	2009	22/22	0.86	0.18	37,47,51,54	0
13	P33	G	2008	22/22	0.89	0.19	26,37,43,44	0
8	PGE	A	2010	10/10	0.89	0.18	29,35,36,38	0
7	SO4	E	2007	5/5	0.91	0.25	66,69,74,81	0
11	EDO	B	202	4/4	0.92	0.15	45,47,49,49	0
7	SO4	C	2008	5/5	0.93	0.25	66,72,76,76	0
7	SO4	G	2007	5/5	0.93	0.26	54,64,68,72	0
11	EDO	C	2012	4/4	0.94	0.13	31,34,36,36	0
7	SO4	C	2007	5/5	0.95	0.19	57,57,64,67	0
7	SO4	A	2006	5/5	0.96	0.17	50,57,66,66	0
7	SO4	A	2007	5/5	0.96	0.18	56,57,61,69	0
7	SO4	E	2006	5/5	0.97	0.18	61,65,68,73	0
3	MGD	E	2002	47/47	0.98	0.09	18,20,22,24	0
3	MGD	A	2002	47/47	0.98	0.11	11,14,17,19	0
7	SO4	C	2006	5/5	0.98	0.06	26,28,33,34	0
7	SO4	A	2008	5/5	0.98	0.08	26,32,36,36	0
3	MGD	E	2001	47/47	0.98	0.10	14,17,20,22	0
3	MGD	C	2002	47/47	0.98	0.12	18,20,22,24	0
3	MGD	G	2002	47/47	0.98	0.10	14,17,19,20	0
7	SO4	A	2009	5/5	0.98	0.08	29,32,34,35	0
3	MGD	C	2001	47/47	0.98	0.11	15,18,20,21	0
3	MGD	A	2001	47/47	0.99	0.12	13,16,17,18	0
6	F3S	C	2005	7/7	0.99	0.08	21,22,24,24	0
4	O	E	2003	1/1	0.99	0.10	32,32,32,32	0
6	F3S	G	2005	7/7	0.99	0.08	20,22,25,25	0
10	FES	F	201	4/4	0.99	0.04	29,30,31,31	0
7	SO4	G	2006	5/5	0.99	0.07	29,30,31,33	0
6	F3S	A	2005	7/7	0.99	0.07	19,20,22,24	0
10	FES	B	201	4/4	0.99	0.04	22,24,25,27	0

Continued on next page...

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
10	FES	D	201	4/4	0.99	0.05	27,28,29,31	0
3	MGD	G	2001	47/47	0.99	0.10	14,18,19,21	0
10	FES	H	201	4/4	0.99	0.06	25,25,27,28	0
6	F3S	E	2005	7/7	0.99	0.07	23,24,26,31	0
4	O	C	2003	1/1	1.00	0.07	22,22,22,22	0
5	4MO	C	2004	1/1	1.00	0.07	23,23,23,23	0
5	4MO	A	2004	1/1	1.00	0.07	18,18,18,18	0
4	O	G	2003	1/1	1.00	0.08	23,23,23,23	0
5	4MO	G	2004	1/1	1.00	0.07	22,22,22,22	0
5	4MO	E	2004	1/1	1.00	0.07	23,23,23,23	0
4	O	A	2003	1/1	1.00	0.09	17,17,17,17	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.