



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 6, 2020 – 09:17 PM BST

PDB ID : 4YIZ  
Title : Crystal structure of engineered TgAMA1 lacking the DII loop in complex with an Eimeria tenella RON2D3 peptide  
Authors : Parker, M.L.; Boulanger, M.J.  
Deposited on : 2015-03-02  
Resolution : 2.20 Å(reported)

This is a Full wwPDB X-ray Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](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 : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.13.1  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
Refmac : 5.8.0158  
CCP4 : 7.0.044 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.13.1

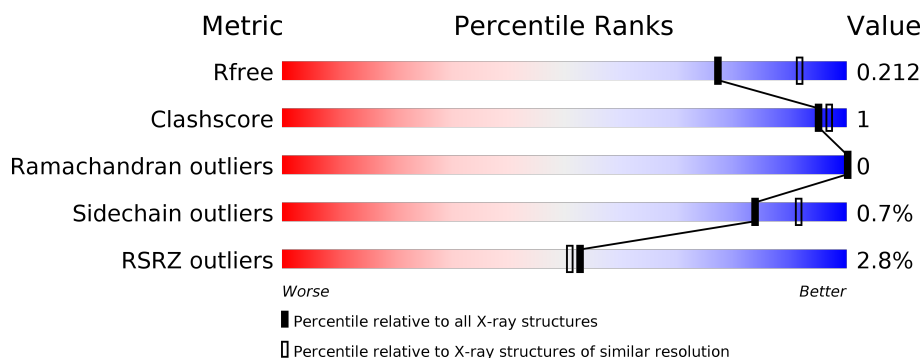
# 1 Overall quality at a glance

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))
$R_{free}$	130704	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (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 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	419	<div> <div>91%</div> <div>6%</div> </div>
1	C	419	<div>6%</div> <div>91%</div> <div>6%</div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 10604 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 Apical membrane antigen AMA1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	392	Total	C	N	O	S	0	3	0
			3091	1939	528	602	22			
1	C	392	Total	C	N	O	S	0	2	0
			3085	1935	528	600	22			
1	E	388	Total	C	N	O	S	0	1	0
			3047	1910	521	594	22			

There are 93 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	59	GLY	-	expression tag	UNP S8GKS3
A	60	SER	-	expression tag	UNP S8GKS3
A	61	ALA	-	expression tag	UNP S8GKS3
A	62	MET	-	expression tag	UNP S8GKS3
A	63	GLY	-	expression tag	UNP S8GKS3
A	?	-	HIS	deletion	UNP S8GKS3
A	?	-	THR	deletion	UNP S8GKS3
A	?	-	TYR	deletion	UNP S8GKS3
A	?	-	PRO	deletion	UNP S8GKS3
A	?	-	LEU	deletion	UNP S8GKS3
A	?	-	THR	deletion	UNP S8GKS3
A	?	-	SER	deletion	UNP S8GKS3
A	?	-	GLN	deletion	UNP S8GKS3
A	?	-	ALA	deletion	UNP S8GKS3
A	?	-	SER	deletion	UNP S8GKS3
A	?	-	TRP	deletion	UNP S8GKS3
A	?	-	ASN	deletion	UNP S8GKS3
A	?	-	ASP	deletion	UNP S8GKS3
A	353	GLY	TRP	linker	UNP S8GKS3
A	354	SER	TRP	linker	UNP S8GKS3
A	355	GLY	PRO	linker	UNP S8GKS3
A	356	SER	LEU	linker	UNP S8GKS3
A	357	GLY	HIS	linker	UNP S8GKS3

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	358	SER	GLN	linker	UNP S8GKS3
A	359	GLY	SER	linker	UNP S8GKS3
A	485	ALA	-	expression tag	UNP S8GKS3
A	486	ALA	-	expression tag	UNP S8GKS3
A	487	LEU	-	expression tag	UNP S8GKS3
A	488	VAL	-	expression tag	UNP S8GKS3
A	489	PRO	-	expression tag	UNP S8GKS3
A	490	ARG	-	expression tag	UNP S8GKS3
C	59	GLY	-	expression tag	UNP S8GKS3
C	60	SER	-	expression tag	UNP S8GKS3
C	61	ALA	-	expression tag	UNP S8GKS3
C	62	MET	-	expression tag	UNP S8GKS3
C	63	GLY	-	expression tag	UNP S8GKS3
C	?	-	HIS	deletion	UNP S8GKS3
C	?	-	THR	deletion	UNP S8GKS3
C	?	-	TYR	deletion	UNP S8GKS3
C	?	-	PRO	deletion	UNP S8GKS3
C	?	-	LEU	deletion	UNP S8GKS3
C	?	-	THR	deletion	UNP S8GKS3
C	?	-	SER	deletion	UNP S8GKS3
C	?	-	GLN	deletion	UNP S8GKS3
C	?	-	ALA	deletion	UNP S8GKS3
C	?	-	SER	deletion	UNP S8GKS3
C	?	-	TRP	deletion	UNP S8GKS3
C	?	-	ASN	deletion	UNP S8GKS3
C	?	-	ASP	deletion	UNP S8GKS3
C	353	GLY	TRP	linker	UNP S8GKS3
C	354	SER	TRP	linker	UNP S8GKS3
C	355	GLY	PRO	linker	UNP S8GKS3
C	356	SER	LEU	linker	UNP S8GKS3
C	357	GLY	HIS	linker	UNP S8GKS3
C	358	SER	GLN	linker	UNP S8GKS3
C	359	GLY	SER	linker	UNP S8GKS3
C	485	ALA	-	expression tag	UNP S8GKS3
C	486	ALA	-	expression tag	UNP S8GKS3
C	487	LEU	-	expression tag	UNP S8GKS3
C	488	VAL	-	expression tag	UNP S8GKS3
C	489	PRO	-	expression tag	UNP S8GKS3
C	490	ARG	-	expression tag	UNP S8GKS3
E	59	GLY	-	expression tag	UNP S8GKS3
E	60	SER	-	expression tag	UNP S8GKS3
E	61	ALA	-	expression tag	UNP S8GKS3

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	62	MET	-	expression tag	UNP S8GKS3
E	63	GLY	-	expression tag	UNP S8GKS3
E	?	-	HIS	deletion	UNP S8GKS3
E	?	-	THR	deletion	UNP S8GKS3
E	?	-	TYR	deletion	UNP S8GKS3
E	?	-	PRO	deletion	UNP S8GKS3
E	?	-	LEU	deletion	UNP S8GKS3
E	?	-	THR	deletion	UNP S8GKS3
E	?	-	SER	deletion	UNP S8GKS3
E	?	-	GLN	deletion	UNP S8GKS3
E	?	-	ALA	deletion	UNP S8GKS3
E	?	-	SER	deletion	UNP S8GKS3
E	?	-	TRP	deletion	UNP S8GKS3
E	?	-	ASN	deletion	UNP S8GKS3
E	?	-	ASP	deletion	UNP S8GKS3
E	353	GLY	TRP	linker	UNP S8GKS3
E	354	SER	TRP	linker	UNP S8GKS3
E	355	GLY	PRO	linker	UNP S8GKS3
E	356	SER	LEU	linker	UNP S8GKS3
E	357	GLY	HIS	linker	UNP S8GKS3
E	358	SER	GLN	linker	UNP S8GKS3
E	359	GLY	SER	linker	UNP S8GKS3
E	485	ALA	-	expression tag	UNP S8GKS3
E	486	ALA	-	expression tag	UNP S8GKS3
E	487	LEU	-	expression tag	UNP S8GKS3
E	488	VAL	-	expression tag	UNP S8GKS3
E	489	PRO	-	expression tag	UNP S8GKS3
E	490	ARG	-	expression tag	UNP S8GKS3

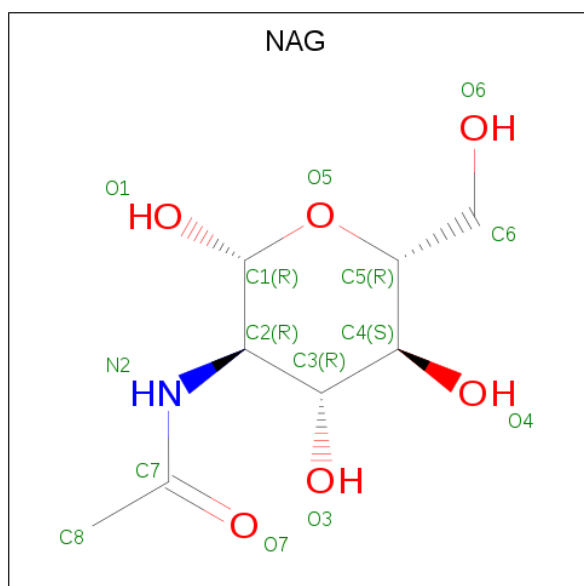
- Molecule 2 is a protein called Rhoptry neck protein 2, putative.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	39	Total	C	N	O	S	0	1	0
			271	168	45	55	3			
2	D	39	Total	C	N	O	S	0	1	0
			271	168	45	55	3			
2	F	39	Total	C	N	O	S	0	1	0
			271	168	45	55	3			

There are 9 discrepancies between the modelled and reference sequences:

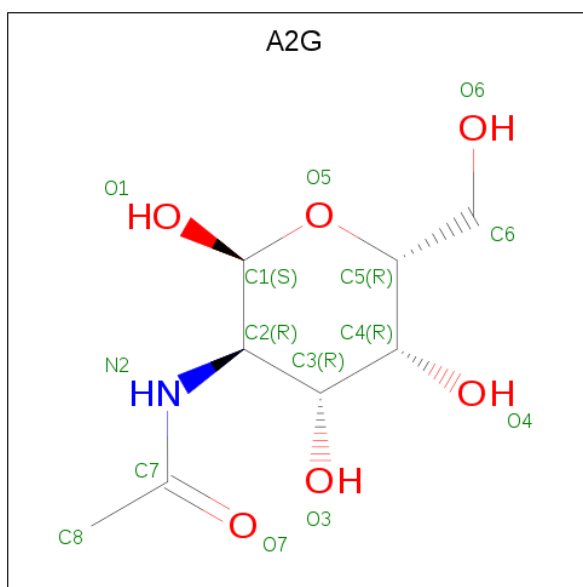
Chain	Residue	Modelled	Actual	Comment	Reference
B	1258	GLY	-	expression tag	UNP U6KQJ2
B	1259	SER	-	expression tag	UNP U6KQJ2
B	1260	ALA	-	expression tag	UNP U6KQJ2
D	1258	GLY	-	expression tag	UNP U6KQJ2
D	1259	SER	-	expression tag	UNP U6KQJ2
D	1260	ALA	-	expression tag	UNP U6KQJ2
F	1258	GLY	-	expression tag	UNP U6KQJ2
F	1259	SER	-	expression tag	UNP U6KQJ2
F	1260	ALA	-	expression tag	UNP U6KQJ2

- Molecule 3 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula:  $C_8H_{15}NO_6$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	0
			14	8	1	5		
3	C	1	Total	C	N	O	0	0
			14	8	1	5		
3	E	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 4 is 2-acetamido-2-deoxy-alpha-D-galactopyranose (three-letter code: A2G) (formula:  $C_8H_{15}NO_6$ ).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	A	1	Total	C	N	O	0	0
			14	8	1	5		

- Molecule 5 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



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

*Continued on next page...*

*Continued from previous page...*

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

- Molecule 6 is water.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	165	Total	O		0	0
			165	165			
6	C	133	Total	O		0	0
			133	133			
6	E	153	Total	O		0	0
			153	153			
6	B	11	Total	O		0	0
			11	11			
6	D	12	Total	O		0	0
			12	12			
6	F	13	Total	O		0	0
			13	13			



### 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.

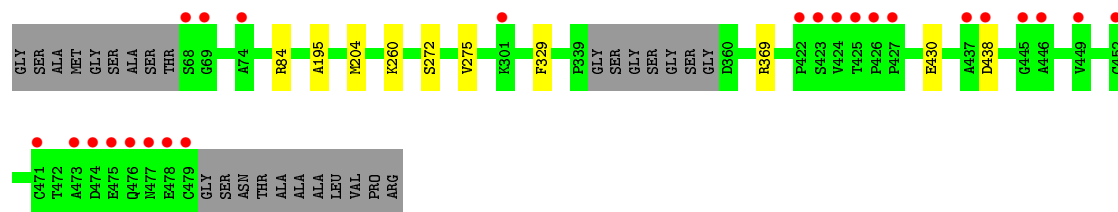
- Molecule 1: Apical membrane antigen AMA1

Chain A: 




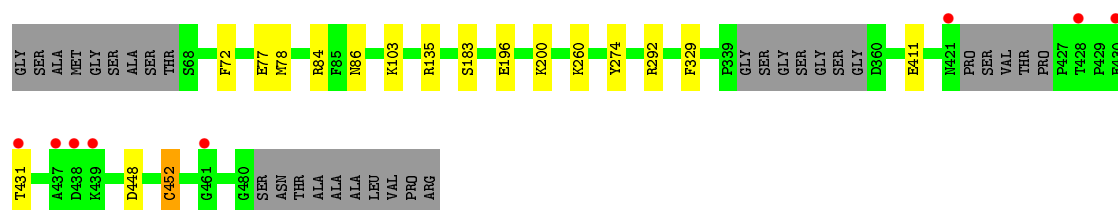
- Molecule 1: Apical membrane antigen AMA1

Chain C: 



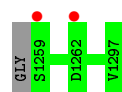
- Molecule 1: Apical membrane antigen AMA1

Chain E: 

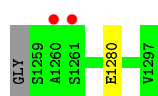
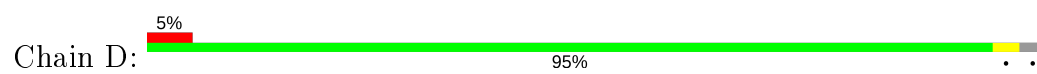


- Molecule 2: Rhoptry neck protein 2, putative

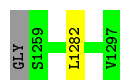
Chain B: 



- Molecule 2: Rhoptry neck protein 2, putative



- Molecule 2: Rhoptry neck protein 2, putative



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	265.93 Å   265.93 Å   94.16 Å 90.00°   90.00°   120.00°	Depositor
Resolution (Å)	59.50 – 2.20 66.48 – 2.20	Depositor EDS
% Data completeness (in resolution range)	93.5 (59.50-2.20) 91.7 (66.48-2.20)	Depositor EDS
$R_{merge}$	0.13	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	3.02 (at 2.20 Å)	Xtriage
Refinement program	PHENIX	Depositor
R, $R_{free}$	0.187 , 0.207 0.195 , 0.212	Depositor DCC
$R_{free}$ test set	9083 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	23.9	Xtriage
Anisotropy	0.359	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 41.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	0.032 for -h,-k,l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	10604	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.51% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4, NAG, A2G

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	0.27	0/3179	0.44	0/4310
1	C	0.28	0/3170	0.47	0/4298
1	E	0.26	0/3126	0.45	0/4233
2	B	0.25	0/275	0.41	0/380
2	D	0.24	0/275	0.44	0/380
2	F	0.24	0/275	0.40	0/380
All	All	0.27	0/10300	0.45	0/13981

There are no bond length outliers.

There are no bond angle outliers.

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	3091	0	2942	6	0
1	C	3085	0	2937	6	0
1	E	3047	0	2892	11	0
2	B	271	0	265	0	0
2	D	271	0	265	1	0
2	F	271	0	265	1	0
3	A	14	0	13	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	14	0	13	0	0
3	E	14	0	13	1	0
4	A	14	0	12	3	0
5	A	10	0	0	0	0
5	C	10	0	0	0	0
5	E	5	0	0	0	0
6	A	165	0	0	0	0
6	B	11	0	0	0	0
6	C	133	0	0	1	0
6	D	12	0	0	0	0
6	E	153	0	0	1	0
6	F	13	0	0	0	0
All	All	10604	0	9617	23	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 1.

All (23) 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:72:PHE:HA	1:E:78:MET:HG2	1.75	0.67
1:E:77:GLU:OE2	1:E:431:THR:HG22	1.92	0.67
1:E:135:ARG:NH1	6:E:708:HOH:O	2.37	0.58
1:C:195:ALA:HB1	1:C:275:VAL:HG22	1.91	0.53
1:E:292:ARG:HB2	1:E:411:GLU:HG2	1.91	0.52
1:E:196:GLU:OE1	1:E:200:LYS:NZ	2.45	0.48
1:C:195:ALA:CB	1:C:275:VAL:HG22	2.44	0.47
1:A:81:PHE:O	1:A:84:ARG:HG2	2.14	0.47
1:C:369[B]:ARG:NH1	6:C:731:HOH:O	2.48	0.47
1:E:448:ASP:O	1:E:452:CYS:N	2.45	0.47
1:A:425:THR:O	4:A:502:A2G:H8A	2.14	0.46
1:E:103:LYS:NZ	1:E:274:TYR:OH	2.50	0.44
1:A:292:ARG:HB2	1:A:411[A]:GLU:HG2	1.99	0.43
1:A:427:PRO:HD3	4:A:502:A2G:H8	2.01	0.43
1:C:272:SER:O	1:C:275:VAL:HG13	2.19	0.43
1:E:86:ASN:HD22	3:E:501:NAG:H83	1.84	0.43
1:E:77:GLU:OE2	1:E:431:THR:CG2	2.65	0.42
1:A:184:ASN:O	1:A:200:LYS:NZ	2.40	0.41
1:E:260:LYS:HE3	1:E:329:PHE:CG	2.55	0.41
1:E:200:LYS:HG2	2:F:1282:LEU:HD12	2.03	0.41
1:C:204:MET:SD	2:D:1280:GLU:HG2	2.60	0.41

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:260:LYS:HE3	1:C:329:PHE:CG	2.56	0.40
1:A:427:PRO:HA	4:A:502:A2G:O7	2.21	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles [i](#)

### 5.3.1 Protein backbone [i](#)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the backbone conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	391/419 (93%)	384 (98%)	7 (2%)	0	100	100
1	C	390/419 (93%)	380 (97%)	10 (3%)	0	100	100
1	E	383/419 (91%)	376 (98%)	7 (2%)	0	100	100
2	B	38/40 (95%)	38 (100%)	0	0	100	100
2	D	38/40 (95%)	38 (100%)	0	0	100	100
2	F	38/40 (95%)	38 (100%)	0	0	100	100
All	All	1278/1377 (93%)	1254 (98%)	24 (2%)	0	100	100

There are no Ramachandran outliers to report.

### 5.3.2 Protein sidechains [i](#)

In the following table, the Percentiles column shows the percent sidechain outliers of the chain as a percentile score with respect to all X-ray entries followed by that with respect to entries of similar resolution.

The Analysed column shows the number of residues for which the sidechain conformation was analysed, and the total number of residues.

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	A	344/356 (97%)	342 (99%)	2 (1%)	86	93

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	C	343/356 (96%)	340 (99%)	3 (1%)	78	88
1	E	337/356 (95%)	334 (99%)	3 (1%)	78	88
2	B	31/30 (103%)	31 (100%)	0	100	100
2	D	31/30 (103%)	31 (100%)	0	100	100
2	F	31/30 (103%)	31 (100%)	0	100	100
All	All	1117/1158 (96%)	1109 (99%)	8 (1%)	84	91

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	138	PHE
1	A	452	CYS
1	C	84	ARG
1	C	430	GLU
1	C	438	ASP
1	E	84	ARG
1	E	183	SER
1	E	452	CYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 5.6 Ligand geometry ⓘ

9 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z  > 2$	Counts	RMSZ	$\# Z  > 2$
4	A2G	A	502	1	14,14,15	0.28	0	17,19,21	0.60	0
3	NAG	E	501	1	14,14,15	0.32	0	17,19,21	0.58	0
5	SO4	C	503	-	4,4,4	0.14	0	6,6,6	0.05	0
3	NAG	C	501	1	14,14,15	0.26	0	17,19,21	0.58	0
5	SO4	A	504	-	4,4,4	0.15	0	6,6,6	0.05	0
3	NAG	A	501	1	14,14,15	0.32	0	17,19,21	0.60	0
5	SO4	C	502	-	4,4,4	0.14	0	6,6,6	0.05	0
5	SO4	E	502	-	4,4,4	0.14	0	6,6,6	0.05	0
5	SO4	A	503	-	4,4,4	0.14	0	6,6,6	0.06	0

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

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	NAG	E	501	1	-	3/6/23/26	0/1/1/1
3	NAG	C	501	1	-	2/6/23/26	0/1/1/1
4	A2G	A	502	1	-	2/6/23/26	0/1/1/1
3	NAG	A	501	1	-	2/6/23/26	0/1/1/1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (9) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	E	501	NAG	C8-C7-N2-C2
3	E	501	NAG	O7-C7-N2-C2
3	C	501	NAG	C8-C7-N2-C2
3	C	501	NAG	O7-C7-N2-C2
3	A	501	NAG	C8-C7-N2-C2

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms
3	A	501	NAG	O7-C7-N2-C2
4	A	502	A2G	C8-C7-N2-C2
4	A	502	A2G	O7-C7-N2-C2
3	E	501	NAG	C4-C5-C6-O6

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	A	502	A2G	3	0
3	E	501	NAG	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 ⓘ

### 6.1 Protein, DNA and RNA chains ⓘ

In the following table, the column labelled ‘#RSRZ> 2’ contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95<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	392/419 (93%)	-0.46	0 100 100	13, 26, 55, 87	1 (0%)
1	C	392/419 (93%)	-0.16	24 (6%) 21 20	15, 30, 88, 117	2 (0%)
1	E	388/419 (92%)	-0.29	8 (2%) 63 61	14, 30, 71, 120	2 (0%)
2	B	39/40 (97%)	-0.17	2 (5%) 28 26	25, 37, 87, 91	0
2	D	39/40 (97%)	-0.21	2 (5%) 28 26	23, 32, 78, 80	0
2	F	39/40 (97%)	-0.23	0 100 100	22, 34, 79, 87	0
All	All	1289/1377 (93%)	-0.29	36 (2%) 53 51	13, 29, 74, 120	5 (0%)

All (36) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	430	GLU	5.7
1	C	424	VAL	5.2
1	C	423	SER	4.4
1	E	438	ASP	4.4
1	C	479	CYS	4.4
1	C	473	ALA	4.1
1	C	477	ASN	4.0
1	C	422	PRO	3.7
1	C	438	ASP	3.5
1	E	437	ALA	3.4
1	C	452	CYS	3.4
1	C	68	SER	3.3
1	C	474	ASP	3.1
1	E	421	ASN	3.0
2	D	1261	SER	2.9
1	C	478	GLU	2.8
1	E	431	THR	2.8
2	B	1262	ASP	2.7
1	C	425	THR	2.7

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	E	461	GLY	2.7
1	C	74	ALA	2.6
1	C	301	LYS	2.5
1	E	439	LYS	2.5
1	C	476	GLN	2.5
1	C	449	VAL	2.3
2	B	1259	SER	2.2
1	C	446	ALA	2.2
1	C	475	GLU	2.2
1	C	437	ALA	2.1
1	C	426	PRO	2.1
1	C	445	GLY	2.1
1	C	427	PRO	2.1
1	C	471	CYS	2.1
1	E	428	THR	2.0
1	C	69	GLY	2.0
2	D	1260	ALA	2.0

## 6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

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

## 6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

## 6.4 Ligands ⓘ

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95<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
4	A2G	A	502	14/15	0.86	0.15	50,59,66,66	0
5	SO4	A	503	5/5	0.92	0.17	50,66,87,92	0
3	NAG	C	501	14/15	0.93	0.09	32,42,47,55	0
5	SO4	A	504	5/5	0.94	0.12	53,56,88,91	0
3	NAG	A	501	14/15	0.95	0.08	23,31,42,43	0
5	SO4	C	502	5/5	0.95	0.20	42,59,79,86	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
3	NAG	E	501	14/15	0.95	0.11	30,37,47,56	0
5	SO4	E	502	5/5	0.97	0.12	51,67,72,74	0
5	SO4	C	503	5/5	0.99	0.15	47,51,55,58	5

## 6.5 Other polymers [i](#)

There are no such residues in this entry.