wwPDB X-ray Structure Validation Summary Report (i)

May 15, 2020 – 07:52 pm BST

PDB ID	:	2XA7
Title	:	AP2 clathrin adaptor core in active complex with cargo peptides
Authors	:	Jackson, L.P.; Kelly, B.T.; McCoy, A.J.; Evans, P.R.; Owen, D.J.
Deposited on	:	2010-03-29
Resolution	:	3.10 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 following versions of software and data (see references (1)) were used in the production of this report:

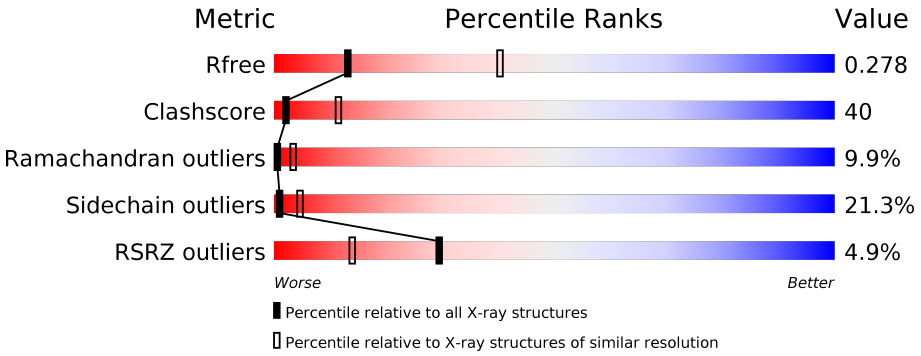
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.10 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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

The table below summarises the geometric issues observed across the polymeric chains and their fit to the electron density. The red, orange, yellow and green segments on the lower bar indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria 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	621	4% 37% 45% 16%
2	B	592	32% 44% 19%
3	M	446	13% 36% 40% 18%
4	P	6	17% 33% 67%
5	S	142	% 28% 53% 19%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 14137 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 ADAPTOR-RELATED PROTEIN COMPLEX 2, ALPHA 2 SUBUNIT.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	613	4836	3081	833	901	21	0	0	0

- Molecule 2 is a protein called AP-2 COMPLEX SUBUNIT BETA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	579	4596	2928	763	880	25	0	0	0

- Molecule 3 is a protein called AP-2 COMPLEX SUBUNIT MU,.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	M	428	3423	2192	591	621	19	0	0	0

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	237	MET	-	insertion	UNP P84092
M	238	GLU	-	insertion	UNP P84092
M	239	GLN	-	insertion	UNP P84092
M	240	LYS	-	insertion	UNP P84092
M	241	LEU	-	insertion	UNP P84092
M	242	ILE	-	insertion	UNP P84092
M	243	SER	-	insertion	UNP P84092
M	244	GLU	-	insertion	UNP P84092
M	245	GLU	-	insertion	UNP P84092
M	246	ASP	-	insertion	UNP P84092
M	247	LEU	-	insertion	UNP P84092

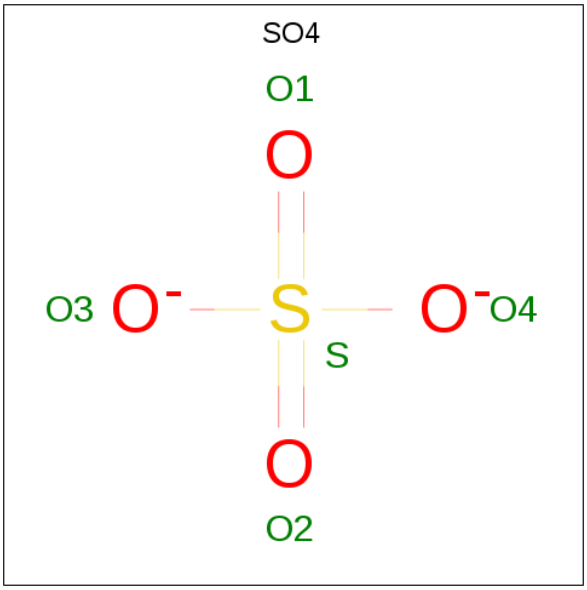
- Molecule 4 is a protein called TGN38 CARGO PEPTIDE.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	P	6	Total	C	N	O	0	0	0
			57	34	11	12			

- Molecule 5 is a protein called AP-2 COMPLEX SUBUNIT SIGMA.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	S	142	Total	C	N	O	S	0	0	0
			1200	778	200	215	7			

- Molecule 6 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



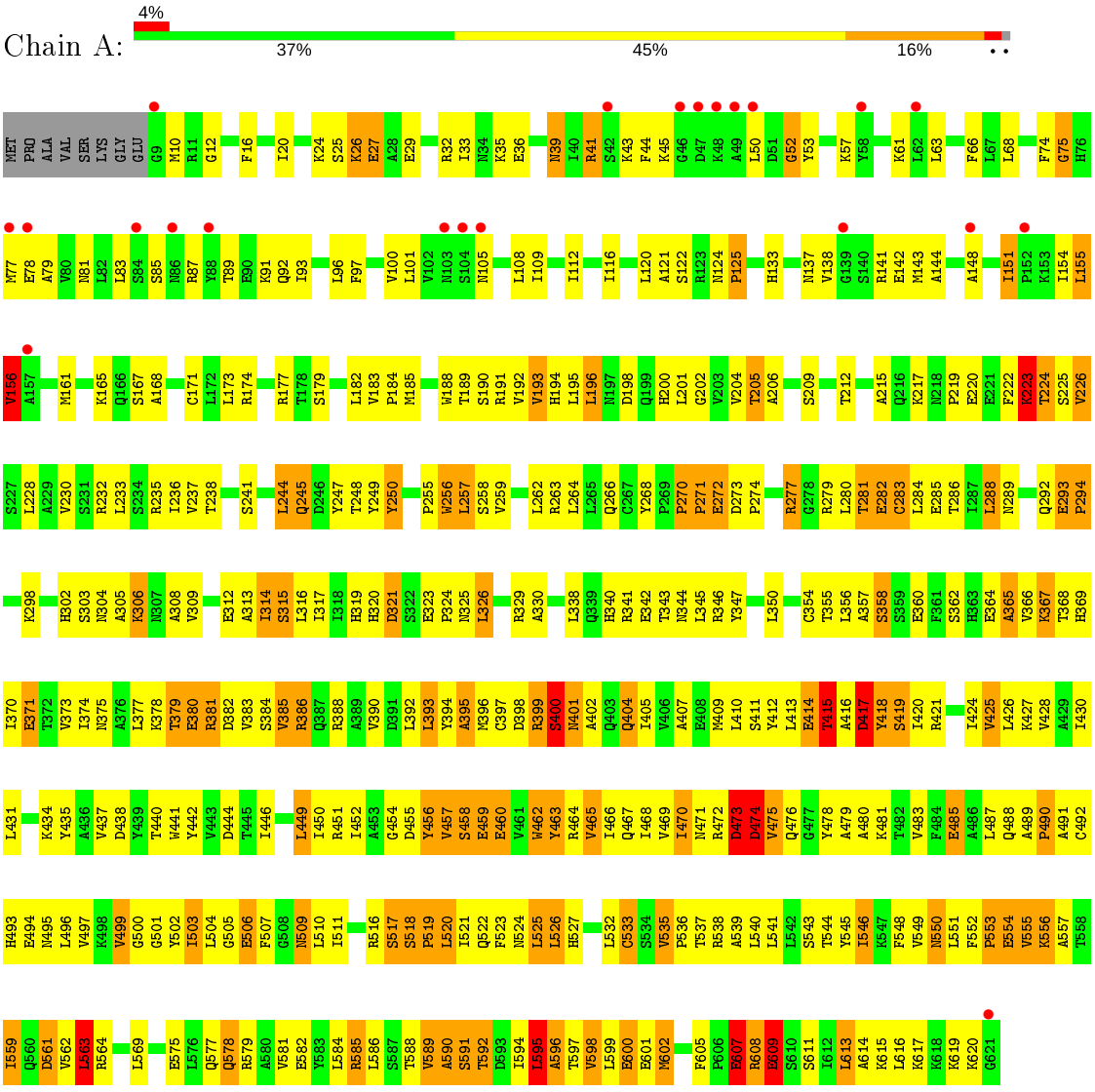
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	A	1	Total	O	S	0	0
			5	4	1		
6	A	1	Total	O	S	0	0
			5	4	1		
6	A	1	Total	O	S	0	0
			5	4	1		
6	S	1	Total	O	S	0	0
			5	4	1		
6	S	1	Total	O	S	0	0
			5	4	1		

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.

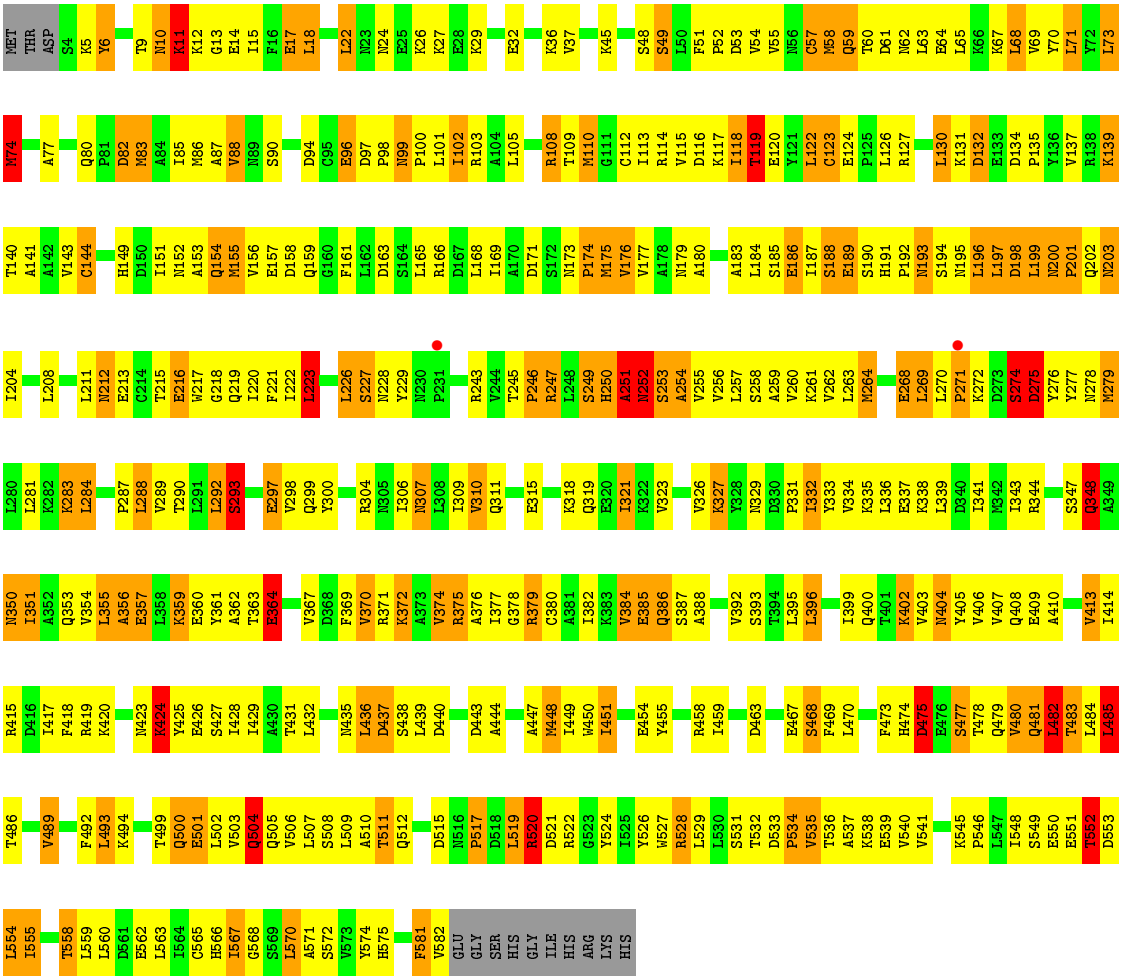
- Molecule 1: ADAPTOR-RELATED PROTEIN COMPLEX 2, ALPHA 2 SUBUNIT

Chain A:



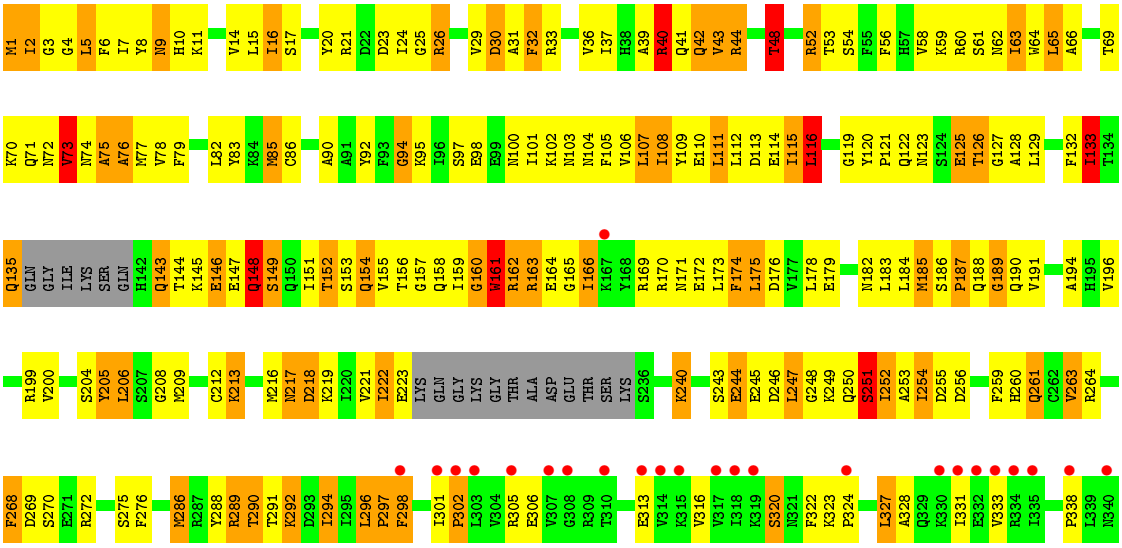
- Molecule 2: AP-2 COMPLEX SUBUNIT BETA

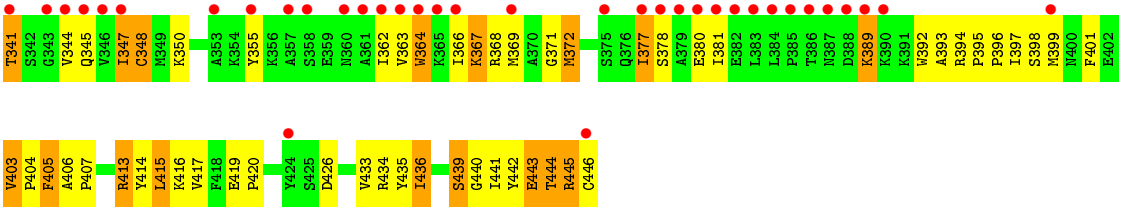
Chain B:



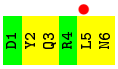
• Molecule 3: AP-2 COMPLEX SUBUNIT MU,

Chain M:

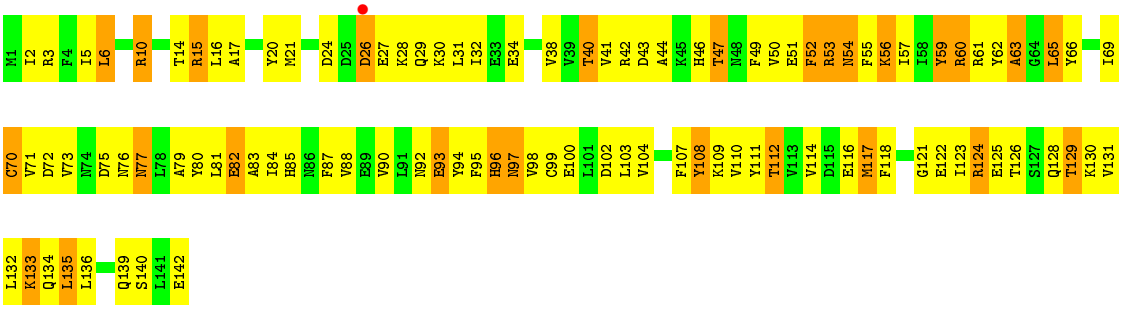




• Molecule 4: TGN38 CARGO PEPTIDE



• Molecule 5: AP-2 COMPLEX SUBUNIT SIGMA



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	255.32Å 255.32Å 156.75Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	127.88 – 3.10 42.63 – 3.10	Depositor EDS
% Data completeness (in resolution range)	96.2 (127.88-3.10) 96.2 (42.63-3.10)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.77 (at 3.12Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.236 , 0.285 0.231 , 0.278	Depositor DCC
R_{free} test set	3378 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	83.6	Xtriage
Anisotropy	0.500	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 82.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.44$, $< L^2 > = 0.26$	Xtriage
Estimated twinning fraction	0.027 for -1/3*h+1/3*k+4/3*l,-k,2/3*h+1/3*k+1/3*l 0.018 for -2/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+4/3*l,-1/3*h+1/3*k+1/3*l 0.028 for -h,1/3*h-1/3*k-4/3*l,-1/3*h-2/3*k+1/3*l 0.022 for -1/3*h-2/3*k+4/3*l,-2/3*h-1/3*k-4/3*l,1/3*h-1/3*k-1/3*l 0.024 for -h,2/3*h+1/3*k+4/3*l,1/3*h+2/3*k-1/3*l 0.026 for 1/3*h+2/3*k-4/3*l,-k,-2/3*h-1/3*k-1/3*l 0.043 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	14137	wwPDB-VP
Average B, all atoms (Å^2)	112.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.60% 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

5.1 Standard geometry

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

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#$\|Z\|$ >5	RMSZ	#$\|Z\|$ >5
1	A	1.01	12/4921 (0.2%)	1.05	14/6669 (0.2%)
2	B	0.83	2/4668 (0.0%)	0.97	10/6331 (0.2%)
3	M	0.81	0/3488	0.91	2/4701 (0.0%)
4	P	0.63	0/57	0.82	0/74
5	S	0.85	0/1224	0.89	0/1650
All	All	0.89	14/14358 (0.1%)	0.98	26/19425 (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	3
2	B	0	5
3	M	0	2
All	All	0	10

The worst 5 of 14 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	607	GLU	CB-CG	11.97	1.74	1.52
1	A	607	GLU	CG-CD	11.27	1.68	1.51
1	A	492	CYS	CB-SG	-8.39	1.68	1.82
1	A	85	SER	C-N	6.45	1.48	1.34
1	A	400	SER	C-O	6.22	1.35	1.23

The worst 5 of 26 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	595	LEU	CA-CB-CG	8.76	135.44	115.30
1	A	456	TYR	N-CA-C	-7.51	90.72	111.00
1	A	458	SER	N-CA-C	-7.25	91.43	111.00
2	B	196	LEU	CA-CB-CG	7.17	131.78	115.30
1	A	399	ARG	NE-CZ-NH1	7.06	123.83	120.30

There are no chirality outliers.

5 of 10 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	381	ARG	Peptide
1	A	401	ASN	Peptide
1	A	456	TYR	Peptide
2	B	249	SER	Peptide
2	B	251	ALA	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	4836	0	4952	403	0
2	B	4596	0	4715	435	0
3	M	3423	0	3485	260	0
4	P	57	0	53	4	0
5	S	1200	0	1195	120	0
6	A	15	0	0	0	0
6	S	10	0	0	2	0
All	All	14137	0	14400	1144	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 40.

The worst 5 of 1144 close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:607:GLU:CB	1:A:607:GLU:CG	1.74	1.56

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:155:MET:HB2	2:B:158:ASP:CB	1.67	1.24
1:A:457:VAL:HG13	1:A:458:SER:O	1.41	1.19
3:M:320:SER:HB2	3:M:322:PHE:HE1	1.03	1.17
5:S:57:ILE:HG22	5:S:70:CYS:HB2	1.29	1.15

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	611/621 (98%)	429 (70%)	120 (20%)	62 (10%)	0 3
2	B	577/592 (98%)	380 (66%)	139 (24%)	58 (10%)	0 3
3	M	422/446 (95%)	307 (73%)	72 (17%)	43 (10%)	0 3
4	P	4/6 (67%)	4 (100%)	0	0	100 100
5	S	140/142 (99%)	106 (76%)	24 (17%)	10 (7%)	1 6
All	All	1754/1807 (97%)	1226 (70%)	355 (20%)	173 (10%)	0 3

5 of 173 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	125	PRO
1	A	156	VAL
1	A	223	LYS
1	A	226	VAL
1	A	270	PRO

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	537/543 (99%)	445 (83%)	92 (17%)	2	9
2	B	522/533 (98%)	394 (76%)	128 (24%)	0	2
3	M	378/398 (95%)	291 (77%)	87 (23%)	1	3
4	P	6/6 (100%)	5 (83%)	1 (17%)	2	9
5	S	131/131 (100%)	103 (79%)	28 (21%)	1	4
All	All	1574/1611 (98%)	1238 (79%)	336 (21%)	1	4

5 of 336 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	292	LEU
2	B	443	ASP
5	S	10	ARG
2	B	319	GLN
2	B	372	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 48 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	99	ASN
2	B	205	ASN
5	S	48	ASN
2	B	149	HIS
2	B	173	ASN

5.3.3 RNA (i)

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

5 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	$\#\|Z\| > 2$
6	SO4	S	1143	-	4,4,4	0.32	0	6,6,6	0.50	0
6	SO4	S	1144	-	4,4,4	0.16	0	6,6,6	0.37	0
6	SO4	A	1622	-	4,4,4	0.17	0	6,6,6	0.30	0
6	SO4	A	1624	-	4,4,4	0.18	0	6,6,6	0.39	0
6	SO4	A	1623	-	4,4,4	0.16	0	6,6,6	0.37	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	S	1143	SO4	1	0
6	S	1144	SO4	1	0

5.7 Other polymers (i)

There are no such residues in this entry.

5.8 Polymer linkage issues

There are no chain breaks in this entry.

6 Fit of model and data

6.1 Protein, DNA and RNA chains

In the following table, the column labelled '#RSRZ> 2' contains the number (and percentage) of RSRZ outliers, followed by percent RSRZ outliers for the chain as percentile scores relative to all X-ray entries and entries of similar resolution. The OWAB column contains the minimum, median, 95^{th} 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	613/621 (98%)	-0.16	22 (3%) 42 22	40, 94, 189, 206	0
2	B	579/592 (97%)	-0.30	2 (0%) 94 88	57, 101, 139, 157	0
3	M	428/446 (95%)	0.49	60 (14%) 2 1	62, 118, 216, 252	0
4	P	6/6 (100%)	0.65	1 (16%) 1 1	121, 124, 129, 129	0
5	S	142/142 (100%)	-0.29	1 (0%) 87 75	70, 106, 149, 184	0
All	All	1768/1807 (97%)	-0.06	86 (4%) 29 14	40, 105, 188, 252	0

The worst 5 of 86 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	M	387	ASN	7.3
3	M	379	ALA	7.2
3	M	385	PRO	7.0
3	M	315	LYS	6.1
3	M	386	THR	5.7

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

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

6.3 Carbohydrates

There are no carbohydrates in this entry.

6.4 Ligands

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum,

median, 95^{th} 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
6	SO4	A	1622	5/5	0.88	0.13	162,162,163,163	0
6	SO4	A	1624	5/5	0.93	0.14	122,122,123,123	0
6	SO4	S	1144	5/5	0.95	0.13	104,105,107,108	0
6	SO4	S	1143	5/5	0.97	0.17	83,85,86,86	0
6	SO4	A	1623	5/5	0.98	0.16	77,78,79,79	0

6.5 Other polymers (i)

There are no such residues in this entry.