wwPDB X-ray Structure Validation Summary Report (i)

May 14, 2020 – 04:29 am BST

PDB ID	:	4AM6
Title	:	C-TERMINAL DOMAIN OF ACTIN-RELATED PROTEIN ARP8 FROM S. CEREVISIAE
Authors	:	Wuerges, J.; Saravanan, M.; Bose, D.; Cook, N.J.; Zhang, X.; Wigley, D.B.
Deposited on	:	2012-03-07
Resolution	:	2.70 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:

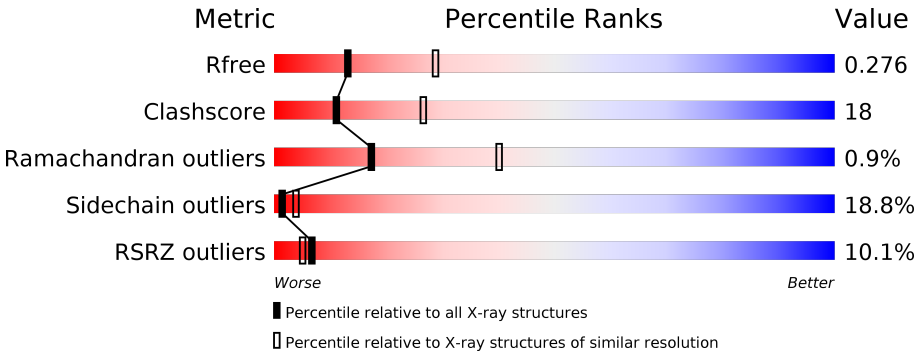
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.70 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	2808 (2.70-2.70)
Clashscore	141614	3122 (2.70-2.70)
Ramachandran outliers	138981	3069 (2.70-2.70)
Sidechain outliers	138945	3069 (2.70-2.70)
RSRZ outliers	127900	2737 (2.70-2.70)

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	655	9% 55% 30% 9% • 5%
1	B	655	11% 56% 30% 9% 5%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 10114 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 ACTIN-LIKE PROTEIN ARP8.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	623	Total	C	N	O	S	0	0	0
			5036	3219	840	963	14			
1	B	623	Total	C	N	O	S	0	0	0
			5036	3219	840	963	14			

There are 42 discrepancies between the modelled and reference sequences:

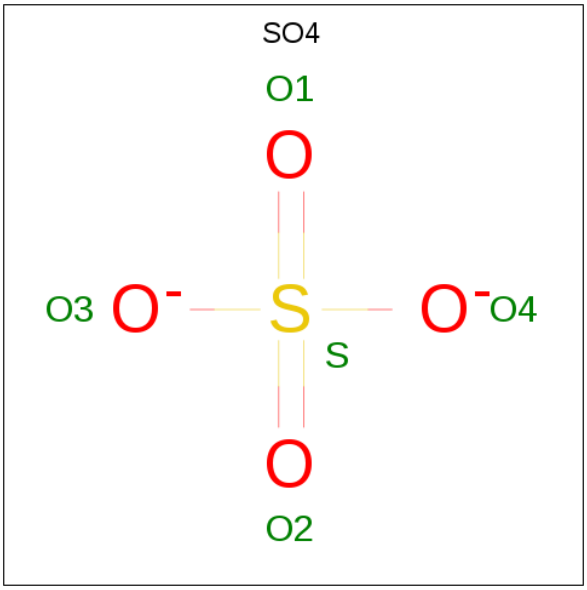
Chain	Residue	Modelled	Actual	Comment	Reference
A	227	MET	-	expression tag	UNP Q12386
A	228	GLY	-	expression tag	UNP Q12386
A	229	SER	-	expression tag	UNP Q12386
A	230	SER	-	expression tag	UNP Q12386
A	231	HIS	-	expression tag	UNP Q12386
A	232	HIS	-	expression tag	UNP Q12386
A	233	HIS	-	expression tag	UNP Q12386
A	234	HIS	-	expression tag	UNP Q12386
A	235	HIS	-	expression tag	UNP Q12386
A	236	HIS	-	expression tag	UNP Q12386
A	237	SER	-	expression tag	UNP Q12386
A	238	SER	-	expression tag	UNP Q12386
A	239	GLY	-	expression tag	UNP Q12386
A	240	LEU	-	expression tag	UNP Q12386
A	241	VAL	-	expression tag	UNP Q12386
A	242	PRO	-	expression tag	UNP Q12386
A	243	ARG	-	expression tag	UNP Q12386
A	244	GLY	-	expression tag	UNP Q12386
A	245	SER	-	expression tag	UNP Q12386
A	246	HIS	-	expression tag	UNP Q12386
A	247	MET	-	expression tag	UNP Q12386
B	227	MET	-	expression tag	UNP Q12386
B	228	GLY	-	expression tag	UNP Q12386
B	229	SER	-	expression tag	UNP Q12386
B	230	SER	-	expression tag	UNP Q12386

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	231	HIS	-	expression tag	UNP Q12386
B	232	HIS	-	expression tag	UNP Q12386
B	233	HIS	-	expression tag	UNP Q12386
B	234	HIS	-	expression tag	UNP Q12386
B	235	HIS	-	expression tag	UNP Q12386
B	236	HIS	-	expression tag	UNP Q12386
B	237	SER	-	expression tag	UNP Q12386
B	238	SER	-	expression tag	UNP Q12386
B	239	GLY	-	expression tag	UNP Q12386
B	240	LEU	-	expression tag	UNP Q12386
B	241	VAL	-	expression tag	UNP Q12386
B	242	PRO	-	expression tag	UNP Q12386
B	243	ARG	-	expression tag	UNP Q12386
B	244	GLY	-	expression tag	UNP Q12386
B	245	SER	-	expression tag	UNP Q12386
B	246	HIS	-	expression tag	UNP Q12386
B	247	MET	-	expression tag	UNP Q12386

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



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

- Molecule 3 is water.

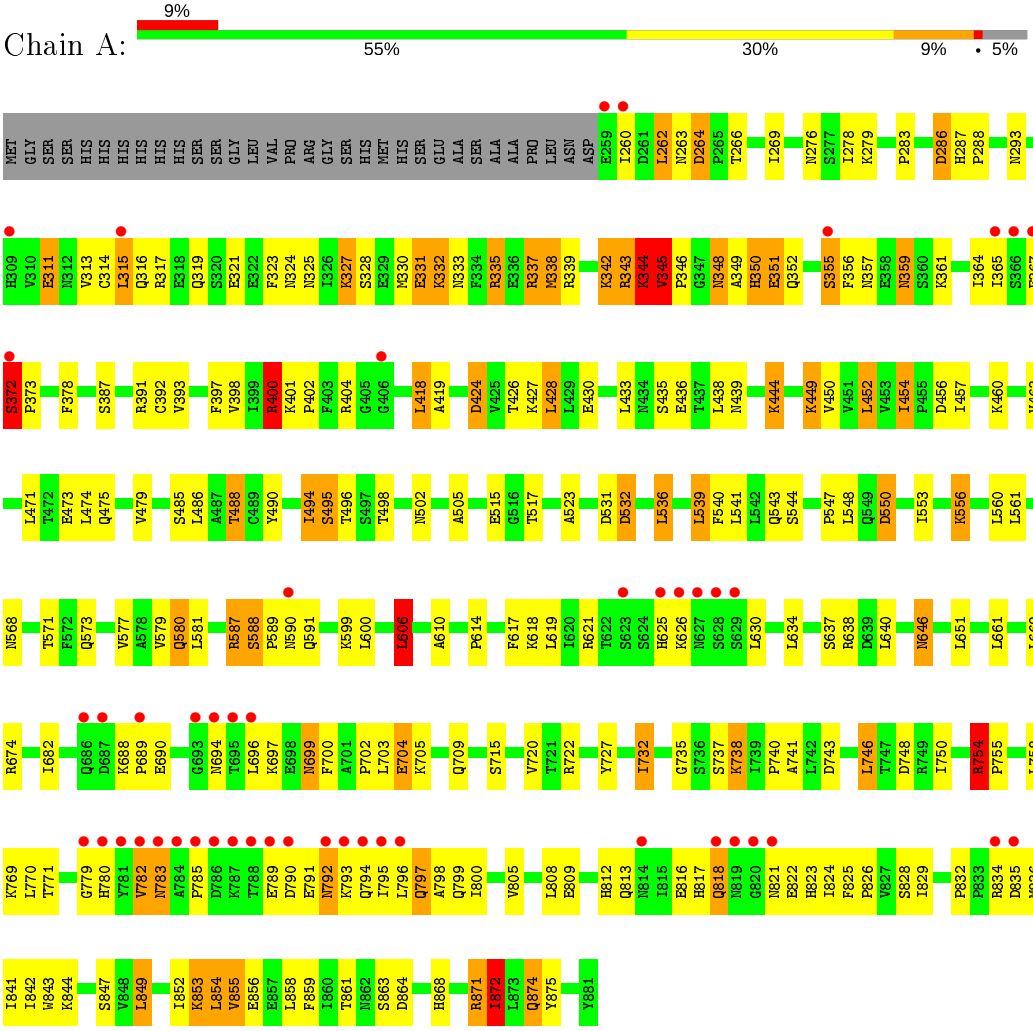
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	18	Total O 18 18	0	0
3	B	14	Total O 14 14	0	0

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.

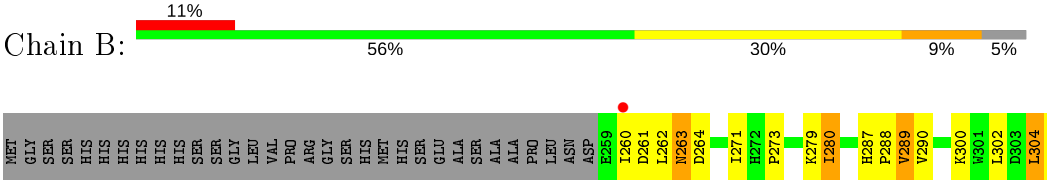
- Molecule 1: ACTIN-LIKE PROTEIN ARP8

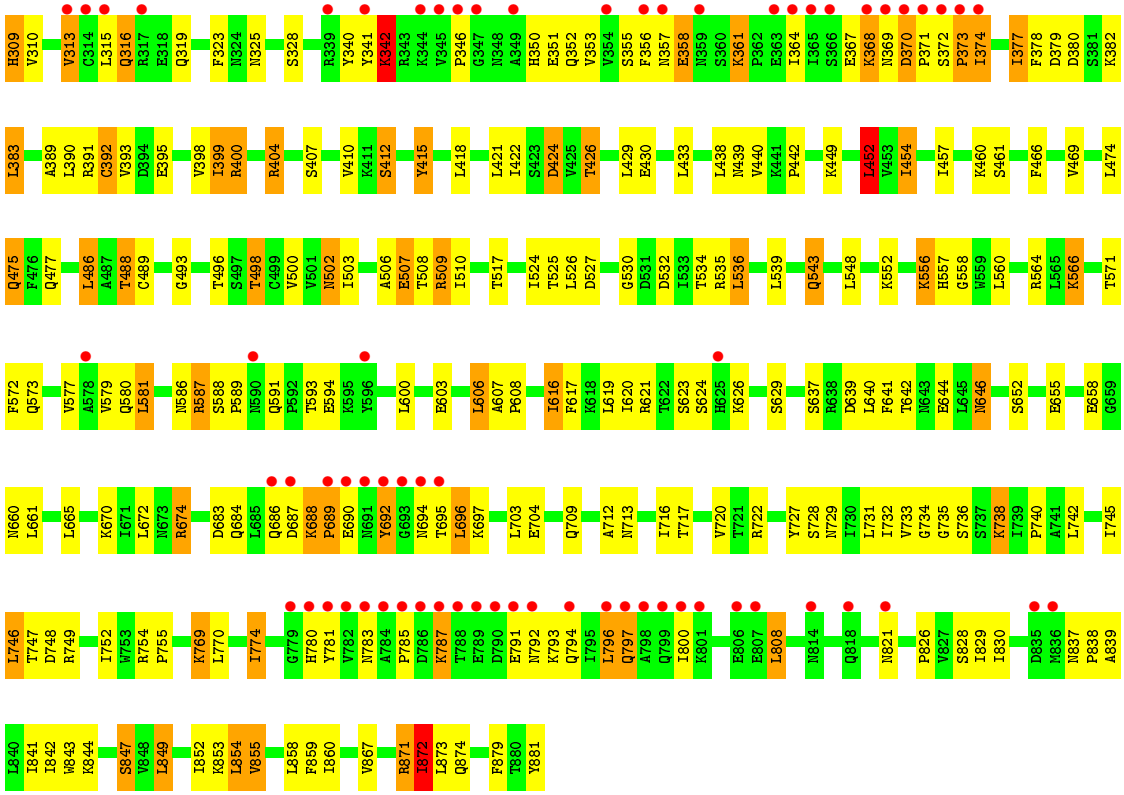
Chain A:



- Molecule 1: ACTIN-LIKE PROTEIN ARP8

Chain B:




H309 V310 V313 C314 L315 Q316 R317 E318 Q319 F323 N324 N325 S328 R339 Y340 Y341 K342 R343 K344 V345 P346 G347 N348 A349 H350 E351 Q352 V353 V354 S355 F356 N357 E358 N359 S360 K361 P362 E363 I364 I365 S366 E367 K368 N369 D370 P371 S372 P373 I374 I377 F378 D379 D380 S381 K382
L383 A389 L390 R391 C392 V393 D394 E395 V398 I399 R400 R404 S407 V410 K411 S412 V415 L418 L421 I422 S423 D424 V425 T426 L429 E430 L433 L438 N439 V440 K441 P442 K449 L452 V453 I454 I457 K460 S461 F466 V469 L474
Q475 F476 Q477 L486 A487 T488 C489 G493 T496 S497 T498 C499 V500 V501 N502 I503 A506 E507 T508 R509 I510 T517 I524 T525 L526 D527 G530 D531 D532 I533 T534 R535 L536 L539 Q543 L548 K552 K556 H557 G558 W559 L560 R564 L565 K566 T571
F572 Q573 V577 A578 V579 Q580 L581 N586 R587 S588 P589 N590 Q591 P592 T593 E594 K595 Y596 L600 E603 L606 A607 P608 I616 F617 K618 L619 I620 R621 T622 S623 S624 H625 K626 S629 S637 E638 D639 L640 F641 T642 N643 E644 L645 N646 S652 E655 E658 G659
N660 L661 L665 K670 I671 L672 N673 R674 D683 Q684 L685 Q686 D687 K688 P689 E690 N691 Y692 G693 N694 T695 L696 K697 L703 E704 Q709 A712 N713 I716 T717 V720 T721 R722 Y727 S728 N729 I730 L731 I732 V733 G734 G735 S736 S737 K738 I739 P740 A741 L742 I745
L746 T747 D748 R749 I752 W753 R754 P755 K769 L770 I774 G779 H780 Y781 V782 N783 A784 P785 D786 K787 T788 E789 D790 E791 N792 K793 Q794 I795 L796 Q797 A798 Q799 I800 K801 E806 E807 L808 N814 Q818 N821 P826 V827 S828 I829 I830 D835 M836 N837 P838 A839
L840 I841 I842 W843 K844 S847 V848 L849 I852 K853 L854 V855 L858 F859 I860 V867 R871 I872 L873 Q874 F879 T880 Y881

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	138.20Å 87.86Å 149.37Å 90.00° 115.40° 90.00°	Depositor
Resolution (Å)	28.89 – 2.70 28.89 – 2.70	Depositor EDS
% Data completeness (in resolution range)	98.9 (28.89-2.70) 98.9 (28.89-2.70)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.83 (at 2.72Å)	Xtriage
Refinement program	REFMAC 5.5.0088	Depositor
R, R_{free}	0.223 , 0.286 0.221 , 0.276	Depositor DCC
R_{free} test set	2190 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	70.3	Xtriage
Anisotropy	0.025	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 71.4	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.015 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	10114	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.02% 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	0.90	4/5149 (0.1%)	0.94	14/6982 (0.2%)
1	B	0.75	5/5149 (0.1%)	0.86	11/6982 (0.2%)
All	All	0.83	9/10298 (0.1%)	0.90	25/13964 (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	2
1	B	0	1
All	All	0	3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	331	GLU	CD-OE1	15.34	1.42	1.25
1	A	331	GLU	CD-OE2	13.05	1.40	1.25
1	A	344	LYS	CE-NZ	10.53	1.75	1.49
1	A	327	LYS	CD-CE	6.42	1.67	1.51
1	B	424	ASP	CB-CG	6.02	1.64	1.51

The worst 5 of 25 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	424	ASP	CB-CG-OD1	10.67	127.91	118.30
1	A	424	ASP	CB-CG-OD1	-7.80	111.28	118.30
1	A	754	ARG	NE-CZ-NH1	7.54	124.07	120.30
1	A	342	LYS	N-CA-C	6.94	129.74	111.00

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	344	LYS	CD-CE-NZ	6.92	127.63	111.70

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	349	ALA	Peptide
1	A	587	ARG	Peptide
1	B	342	LYS	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	5036	0	4985	206	0
1	B	5036	0	4985	159	0
2	A	10	0	0	1	0
3	A	18	0	0	2	0
3	B	14	0	0	0	0
All	All	10114	0	9970	361	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 18.

The worst 5 of 361 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:344:LYS:CE	1:A:344:LYS:NZ	1.75	1.46
1:A:785:PRO:HB2	1:A:793:LYS:HD2	1.21	1.19
1:A:488:THR:HG22	1:A:842:ILE:CD1	1.79	1.11
1:A:315:LEU:HD23	1:A:316:GLN:N	1.66	1.09
1:A:315:LEU:CG	1:A:316:GLN:H	1.71	1.03

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	621/655 (95%)	554 (89%)	64 (10%)	3 (0%)	29 54
1	B	621/655 (95%)	544 (88%)	69 (11%)	8 (1%)	12 30
All	All	1242/1310 (95%)	1098 (88%)	133 (11%)	11 (1%)	17 40

5 of 11 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	346	PRO
1	A	372	SER
1	B	373	PRO
1	B	688	LYS
1	B	371	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	569/595 (96%)	464 (82%)	105 (18%)	1 4
1	B	569/595 (96%)	460 (81%)	109 (19%)	1 4
All	All	1138/1190 (96%)	924 (81%)	214 (19%)	1 4

5 of 214 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	828	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	358	GLU
1	B	780	HIS
1	A	853	LYS
1	B	287	HIS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	794	GLN
1	B	316	GLN
1	B	818	GLN
1	A	818	GLN
1	B	369	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)

2 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	$\#\|Z\| > 2$
2	SO4	A	1882	-	4,4,4	0.32	0	6,6,6	0.52	0
2	SO4	A	1883	-	4,4,4	0.35	0	6,6,6	0.39	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.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1883	SO4	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^{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	623/655 (95%)	0.39	56 (8%) 9 7	32, 64, 146, 240	0
1	B	623/655 (95%)	0.60	70 (11%) 5 4	41, 85, 174, 263	0
All	All	1246/1310 (95%)	0.49	126 (10%) 7 5	32, 74, 162, 263	0

The worst 5 of 126 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	786	ASP	14.9
1	B	373	PRO	13.9
1	B	689	PRO	12.3
1	B	372	SER	11.6
1	A	782	VAL	11.5

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
2	SO4	A	1883	5/5	0.95	0.10	70,73,73,73	0
2	SO4	A	1882	5/5	0.96	0.22	62,63,64,67	0

6.5 Other polymers (i)

There are no such residues in this entry.