wwPDB X-ray Structure Validation Summary Report (i)

Jun 24, 2024 – 02:43 PM EDT

PDB ID	:	6QH7
Title	:	AP2 clathrin adaptor mu2T156-phosphorylated core with two cargo peptides in open+ conformation
Authors	:	Wrobel, A.G.; Owen, D.J.; McCoy, A.J.; Evans, P.R.
Deposited on	:	2019-01-15
Resolution	:	3.40 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:

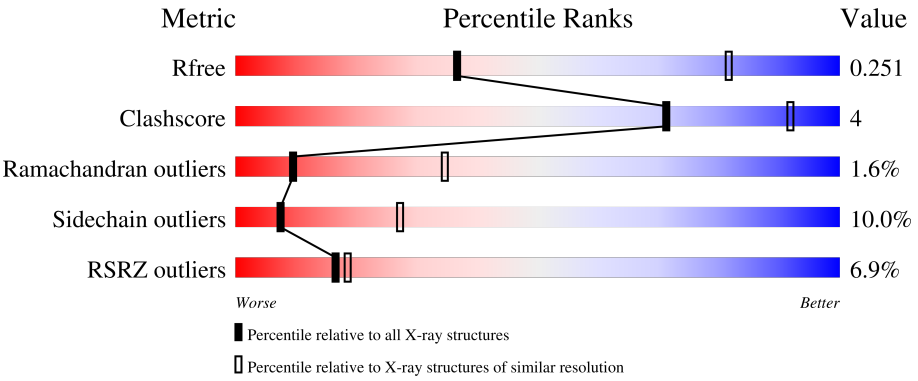
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.40 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.



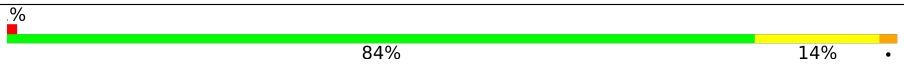
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1026 (3.48-3.32)
Clashscore	141614	1055 (3.48-3.32)
Ramachandran outliers	138981	1038 (3.48-3.32)
Sidechain outliers	138945	1038 (3.48-3.32)
RSRZ outliers	127900	2173 (3.50-3.30)

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	621	2% 81% 15%
2	B	592	7% 68% 15% 15%
3	M	446	% 47% 9% 43%
3	N	446	13% 27% 5% 68%
4	P	6	100%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	Q	10	
6	S	142	

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 13342 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 AP-2 complex subunit alpha.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	614	Total	C	N	O	S	0	0	0
			4844	3085	834	904	21			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	8	ASP	GLU	conflict	UNP Q66HM2

- Molecule 2 is a protein called AP-2 complex subunit beta.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	502	Total	C	N	O	S	0	0	0
			3974	2529	661	763	21			

- Molecule 3 is a protein called ADAPTOR-RELATED PROTEIN COMPLEX 2, MU 2 SUBUNIT, C-TERMINAL DOMAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	M	255	Total	C	N	O	S	0	0	0
			2033	1310	350	359	14			
3	N	143	Total	C	N	O	S	0	0	0
			1157	742	203	207	5			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	237	MET	-	insertion	UNP Q3ZC13
M	238	GLU	-	insertion	UNP Q3ZC13
M	239	GLN	-	insertion	UNP Q3ZC13
M	240	LYS	-	insertion	UNP Q3ZC13
M	241	LEU	-	insertion	UNP Q3ZC13
M	242	ILE	-	insertion	UNP Q3ZC13

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
M	243	SER	-	insertion	UNP Q3ZC13
M	244	GLU	-	insertion	UNP Q3ZC13
M	245	GLU	-	insertion	UNP Q3ZC13
M	246	ASP	-	insertion	UNP Q3ZC13
M	247	LEU	-	insertion	UNP Q3ZC13
N	237	MET	-	insertion	UNP Q3ZC13
N	238	GLU	-	insertion	UNP Q3ZC13
N	239	GLN	-	insertion	UNP Q3ZC13
N	240	LYS	-	insertion	UNP Q3ZC13
N	241	LEU	-	insertion	UNP Q3ZC13
N	242	ILE	-	insertion	UNP Q3ZC13
N	243	SER	-	insertion	UNP Q3ZC13
N	244	GLU	-	insertion	UNP Q3ZC13
N	245	GLU	-	insertion	UNP Q3ZC13
N	246	ASP	-	insertion	UNP Q3ZC13
N	247	LEU	-	insertion	UNP Q3ZC13

- 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 CD4 CARGO PEPTIDE.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
5	Q	9	Total	C	N	O	P	S	0	0	0
			77	46	14	15	1	1			

- Molecule 6 is a protein called AP-2 complex subunit sigma.

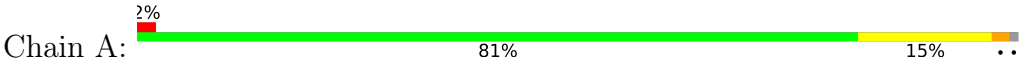
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
6	S	142	Total	C	N	O	S	0	0	0
			1200	778	200	215	7			

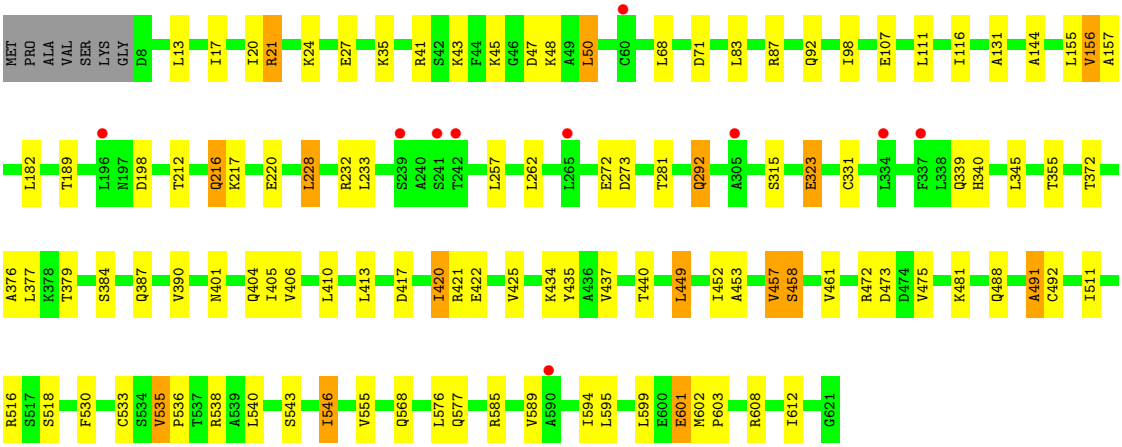
3 Residue-property plots (i)

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: AP-2 complex subunit alpha

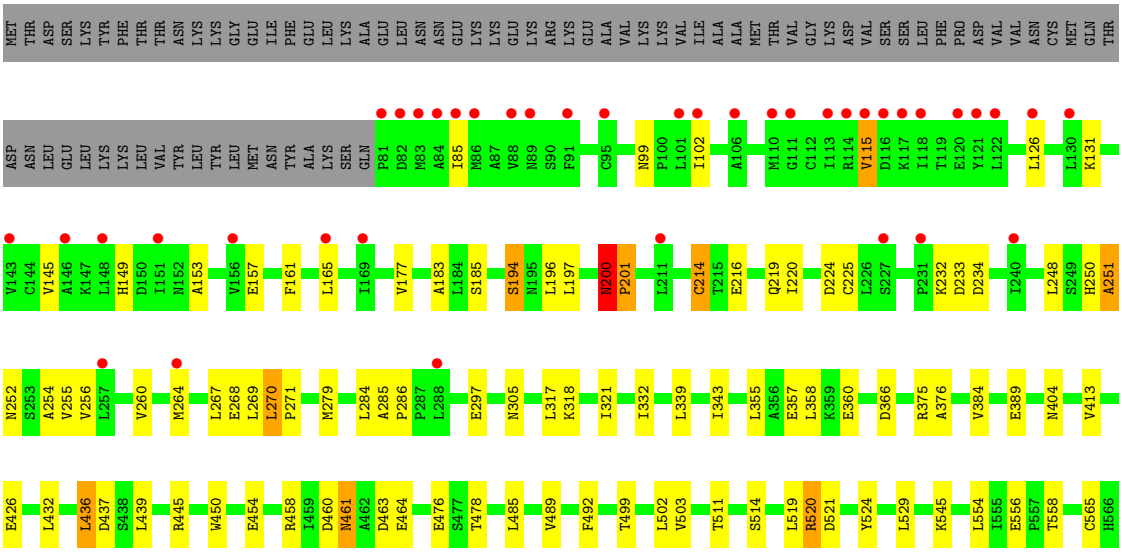
Chain A:





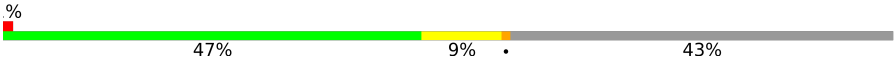
• Molecule 2: AP-2 complex subunit beta

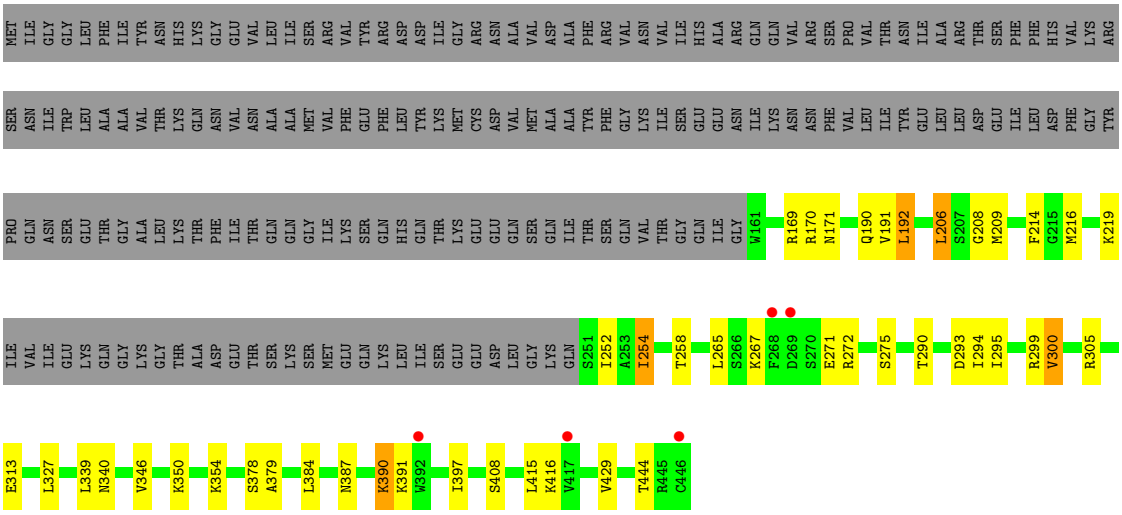
Chain B:



• Molecule 3: ADAPTOR-RELATED PROTEIN COMPLEX 2, MU 2 SUBUNIT, C-TERMINAL DOMAIN

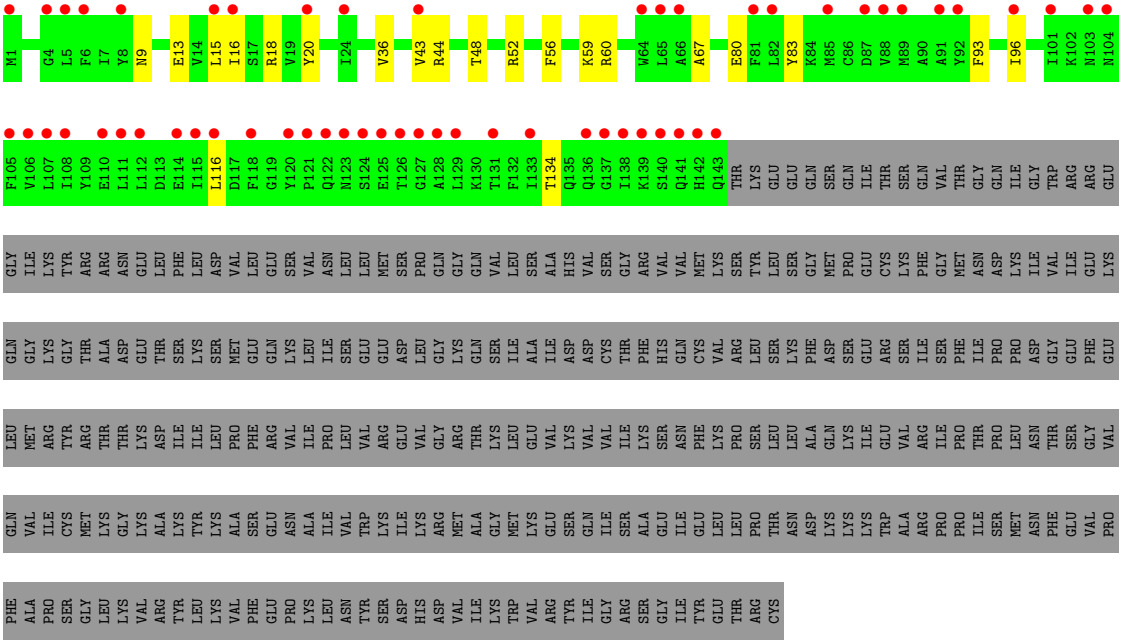
Chain M:





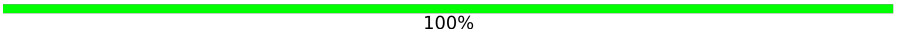
• Molecule 3: ADAPTOR-RELATED PROTEIN COMPLEX 2, MU 2 SUBUNIT, C-TERMINAL DOMAIN

Chain N:



• Molecule 4: TGN38 CARGO PEPTIDE

Chain P:



There are no outlier residues recorded for this chain.

• Molecule 5: CD4 CARGO PEPTIDE

Chain Q:

• Molecule 6: AP-2 complex subunit sigma

Chain S:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α, β, γ	182.18Å 182.18Å 211.60Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	110.00 – 3.40 110.03 – 3.40	Depositor EDS
% Data completeness (in resolution range)	99.8 (110.00-3.40) 99.7 (110.03-3.40)	Depositor EDS
R_{merge}	0.28	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.25 (at 3.41Å)	Xtriage
Refinement program	REFMAC 5.8.0189	Depositor
R, R_{free}	0.208 , 0.248 0.211 , 0.251	Depositor DCC
R_{free} test set	2526 reflections (5.11%)	wwPDB-VP
Wilson B-factor (Å^2)	119.9	Xtriage
Anisotropy	0.061	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 103.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	13342	wwPDB-VP
Average B, all atoms (Å^2)	147.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.59% 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: SEP

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	0.54	0/4929	0.77	4/6680 (0.1%)
2	B	0.55	0/4038	0.74	1/5488 (0.0%)
3	M	0.55	0/2076	0.77	2/2797 (0.1%)
3	N	0.57	0/1180	0.70	1/1592 (0.1%)
4	P	0.51	0/57	0.63	0/74
5	Q	0.67	0/65	0.99	0/82
6	S	0.56	0/1224	0.77	0/1650
All	All	0.55	0/13569	0.76	8/18363 (0.0%)

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
2	B	0	2
3	M	0	1
6	S	0	1
All	All	0	5

There are no bond length outliers.

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	585	ARG	NE-CZ-NH1	6.12	123.36	120.30
1	A	21	ARG	NE-CZ-NH1	5.96	123.28	120.30
2	B	520	ARG	NE-CZ-NH2	-5.55	117.52	120.30
3	M	192	LEU	CA-CB-CG	5.25	127.37	115.30
1	A	601	GLU	N-CA-C	5.24	125.15	111.00

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	491	ALA	Peptide
2	B	251	ALA	Peptide
2	B	271	PRO	Peptide
3	M	293	ASP	Peptide
6	S	138	LEU	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	4844	0	4956	36	0
2	B	3974	0	4061	39	0
3	M	2033	0	2097	11	0
3	N	1157	0	1158	8	0
4	P	57	0	53	0	0
5	Q	77	0	83	0	0
6	S	1200	0	1195	7	0
All	All	13342	0	13603	95	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 4.

The worst 5 of 95 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)
2:B:251:ALA:HB3	2:B:255:VAL:HG23	1.62	0.80
1:A:198:ASP:O	1:A:232:ARG:NH2	2.27	0.68
1:A:387:GLN:HE22	1:A:420:ILE:HG21	1.62	0.64
2:B:200:ASN:N	2:B:200:ASN:HD22	1.97	0.61
2:B:332:ILE:HG21	3:N:43:VAL:HG13	1.84	0.60

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	612/621 (99%)	563 (92%)	35 (6%)	14 (2%)	6 28
2	B	500/592 (84%)	449 (90%)	44 (9%)	7 (1%)	11 37
3	M	251/446 (56%)	211 (84%)	35 (14%)	5 (2%)	7 30
3	N	141/446 (32%)	123 (87%)	18 (13%)	0	100 100
4	P	4/6 (67%)	4 (100%)	0	0	100 100
5	Q	6/10 (60%)	4 (67%)	2 (33%)	0	100 100
6	S	140/142 (99%)	128 (91%)	11 (8%)	1 (1%)	22 55
All	All	1654/2263 (73%)	1482 (90%)	145 (9%)	27 (2%)	9 34

5 of 27 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	458	SER
1	A	601	GLU
2	B	200	ASN
3	M	252	ILE
3	M	254	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	538/543 (99%)	490 (91%)	48 (9%)	9 33
2	B	452/533 (85%)	404 (89%)	48 (11%)	6 24

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
3	M	227/398 (57%)	200 (88%)	27 (12%)	5 19
3	N	124/398 (31%)	114 (92%)	10 (8%)	11 38
4	P	6/6 (100%)	6 (100%)	0	100 100
5	Q	8/9 (89%)	5 (62%)	3 (38%)	0 0
6	S	131/131 (100%)	118 (90%)	13 (10%)	8 27
All	All	1486/2018 (74%)	1337 (90%)	149 (10%)	7 27

5 of 149 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	M	350	LYS
6	S	109	LYS
3	M	390	LYS
3	N	93	PHE
2	B	115	VAL

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 12 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	305	ASN
2	B	319	GLN
3	N	123	ASN
2	B	461	ASN
1	A	401	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)

1 non-standard protein/DNA/RNA residue is 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	$\#\|Z\| > 2$
5	SEP	Q	3	5	8,9,10	0.71	0	8,12,14	1.20	1 (12%)

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
5	SEP	Q	3	5	-	2/5/8/10	-

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
5	Q	3	SEP	P-OG-CB	2.00	123.81	118.30

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	Q	3	SEP	CA-CB-OG-P
5	Q	3	SEP	CB-OG-P-O3P

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates (i)

There are no monosaccharides in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	614/621 (98%)	0.32	10 (1%) 72 70	79, 121, 176, 229	0
2	B	502/592 (84%)	0.52	40 (7%) 12 13	81, 137, 260, 311	0
3	M	255/446 (57%)	0.04	5 (1%) 65 64	104, 162, 224, 270	0
3	N	143/446 (32%)	1.72	56 (39%) 0 0	121, 190, 237, 280	0
4	P	6/6 (100%)	0.10	0 100 100	158, 182, 208, 215	0
5	Q	8/10 (80%)	1.25	2 (25%) 0 0	142, 169, 190, 193	0
6	S	142/142 (100%)	0.39	2 (1%) 75 74	84, 117, 160, 185	0
All	All	1670/2263 (73%)	0.47	115 (6%) 16 18	79, 136, 236, 311	0

The worst 5 of 115 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	N	143	GLN	9.4
3	N	124	SER	7.2
2	B	85	ILE	6.8
3	N	65	LEU	6.4
3	N	138	ILE	6.0

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

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
5	SEP	Q	3	10/11	0.90	0.14	149,224,292,333	0

6.3 Carbohydrates (i)

There are no monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.