Full wwPDB X-ray Structure Validation Report (i)

Oct 2, 2016 – 04:53 AM EDT

PDB ID	:	5IBV
Title	:	Crystal Structure of Human Astrovirus capsid protein
Authors	:	Tao, Y.J.; Toh, Y.S.
Deposited on	:	2016-02-22
Resolution	:	2.15 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*
A user guide is available at
http://wwpdb.org/validation/2016/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:

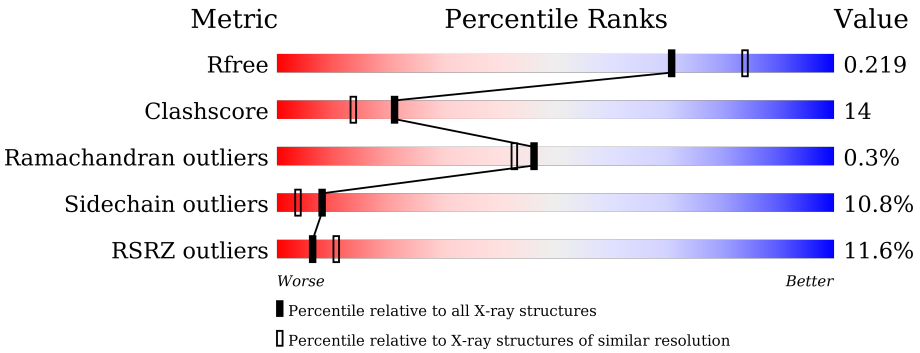
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027939

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 2.15 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}	91344	1045 (2.16-2.16)
Clashscore	102246	1152 (2.16-2.16)
Ramachandran outliers	100387	1131 (2.16-2.16)
Sidechain outliers	100360	1131 (2.16-2.16)
RSRZ outliers	91569	1050 (2.16-2.16)

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. 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	354	10%, 65%, 22%, ·, 8%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2525 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 Capsid polyprotein VP90.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	324	2452	1550	422	469	5	6	0	2	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	416	GLY	ALA	conflict	UNP Q9IFX1
A	417	GLY	-	expression tag	UNP Q9IFX1
A	418	GLY	-	expression tag	UNP Q9IFX1
A	419	HIS	-	expression tag	UNP Q9IFX1
A	420	HIS	-	expression tag	UNP Q9IFX1
A	421	HIS	-	expression tag	UNP Q9IFX1
A	422	HIS	-	expression tag	UNP Q9IFX1
A	423	HIS	-	expression tag	UNP Q9IFX1
A	424	HIS	-	expression tag	UNP Q9IFX1

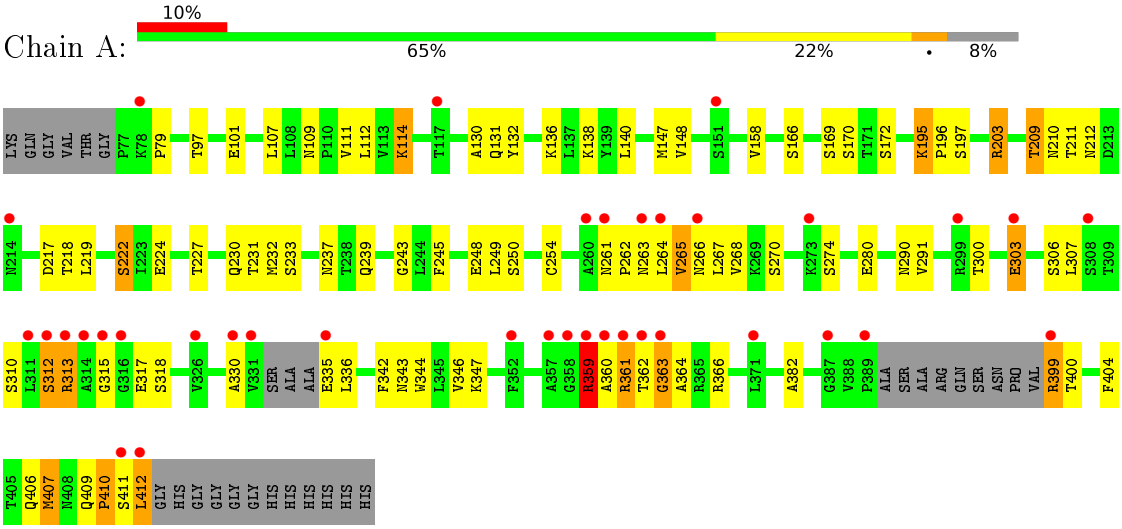
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	73	73	73	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 errors 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: Capsid polyprotein VP90



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	51.98Å 59.24Å 56.20Å 90.00° 91.45° 90.00°	Depositor
Resolution (Å)	30.00 – 2.15 29.62 – 2.16	Depositor EDS
% Data completeness (in resolution range)	97.5 (30.00-2.15) 98.4 (29.62-2.16)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.80 (at 2.16Å)	Xtriage
Refinement program	PHENIX (1.10_2155: ???)	Depositor
R, R_{free}	0.209 , 0.246 0.206 , 0.219	Depositor DCC
R_{free} test set	956 reflections (5.24%)	DCC
Wilson B-factor (Å^2)	28.9	Xtriage
Anisotropy	0.997	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 64.1	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.054 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2525	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.89% 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

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	0.56	0/2504	0.68	0/3410

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	9

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (9) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	264	LEU	Peptide
1	A	312	SER	Peptide
1	A	315	GLY	Peptide
1	A	359	ARG	Peptide
1	A	362	THR	Peptide
1	A	363	GLY	Peptide
1	A	407	MSE	Peptide
1	A	410	PRO	Peptide
1	A	411	SER	Peptide

5.2 Too-close contacts

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	2452	0	2431	69	0
2	A	73	0	0	7	0
All	All	2525	0	2431	69	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 14.

All (69) 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:412:LEU:HD21	2:A:507:HOH:O	1.44	1.12
1:A:412:LEU:CD2	2:A:507:HOH:O	2.00	0.99
1:A:291:VAL:O	1:A:364:ALA:HB1	1.77	0.84
1:A:360:ALA:HB1	1:A:361:ARG:O	1.76	0.84
1:A:360:ALA:HA	1:A:361:ARG:HB2	1.60	0.83
1:A:360:ALA:HA	1:A:361:ARG:CB	2.11	0.80
1:A:211:THR:HA	1:A:218:THR:HG21	1.65	0.76
1:A:195:LYS:HG3	1:A:197:SER:H	1.54	0.73
1:A:132:TYR:O	1:A:209:THR:HG21	1.87	0.73
1:A:262:PRO:O	1:A:263:ASN:HB2	1.91	0.70
1:A:361:ARG:N	1:A:363:GLY:O	2.25	0.68
1:A:412:LEU:C	1:A:412:LEU:HD13	2.15	0.67
1:A:290:ASN:HD22	1:A:366:ARG:HG2	1.59	0.66
1:A:203:ARG:NH1	1:A:217:ASP:OD2	2.26	0.66
1:A:310:SER:OG	2:A:502:HOH:O	2.14	0.66
1:A:79:PRO:O	2:A:501:HOH:O	2.13	0.65
1:A:158:VAL:HG22	1:A:227:THR:HG22	1.80	0.63
1:A:210:ASN:HD21	1:A:412:LEU:HD22	1.64	0.62
1:A:359:ARG:N	1:A:360:ALA:HB2	2.15	0.61
1:A:212:ASN:HB2	1:A:412:LEU:HD23	1.81	0.61
1:A:312:SER:OG	1:A:313:ARG:HB2	2.02	0.60
1:A:290:ASN:ND2	1:A:366:ARG:HG2	2.16	0.60
1:A:147[A]:MSE:HA	1:A:147[A]:MSE:HE2	1.85	0.58
1:A:336:LEU:HB2	1:A:346:VAL:HG21	1.86	0.58
1:A:330:ALA:HB3	1:A:400:THR:HA	1.86	0.57
1:A:224:GLU:HG2	2:A:566:HOH:O	2.06	0.55
1:A:306:SER:CB	1:A:410:PRO:HG2	2.36	0.55
1:A:232[B]:MSE:CE	1:A:239:GLN:HA	2.37	0.54
1:A:412:LEU:HD23	2:A:507:HOH:O	1.85	0.53
1:A:138:LYS:O	1:A:196:PRO:HG3	2.09	0.53

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:265:VAL:HG22	1:A:409:GLN:OE1	2.10	0.52
1:A:342:PHE:CZ	1:A:382:ALA:HB1	2.45	0.52
1:A:360:ALA:CA	1:A:361:ARG:CB	2.85	0.52
1:A:147[B]:MSE:HE2	1:A:245:PHE:N	2.23	0.52
1:A:231:THR:O	1:A:232[B]:MSE:HE2	2.10	0.51
1:A:268:VAL:HG11	1:A:300:THR:HG21	1.92	0.51
1:A:97:THR:HG23	1:A:101:GLU:OE2	2.10	0.50
1:A:114:LYS:O	1:A:114:LYS:HG2	2.10	0.50
1:A:195:LYS:HG3	1:A:197:SER:N	2.25	0.50
1:A:195:LYS:HG3	1:A:196:PRO:N	2.24	0.50
1:A:131:GLN:NE2	2:A:504:HOH:O	2.31	0.48
1:A:109:ASN:OD1	1:A:111:VAL:HG22	2.14	0.47
1:A:166:SER:HB2	1:A:219:LEU:HD23	1.96	0.47
1:A:330:ALA:CB	1:A:400:THR:HA	2.45	0.47
1:A:306:SER:HB2	1:A:410:PRO:HG2	1.97	0.47
1:A:230:GLN:CD	1:A:232[B]:MSE:HE3	2.35	0.47
1:A:107:LEU:HD23	1:A:222:SER:OG	2.16	0.46
1:A:265:VAL:CG2	1:A:409:GLN:OE1	2.64	0.46
1:A:300:THR:HA	1:A:303:GLU:HG3	1.99	0.45
1:A:360:ALA:HA	1:A:361:ARG:HB3	1.98	0.44
1:A:270:SER:HB3	1:A:404:PHE:CE1	2.52	0.44
1:A:140:LEU:HD11	1:A:249:LEU:HD11	1.99	0.44
1:A:147[B]:MSE:HE3	1:A:243:GLY:C	2.38	0.44
1:A:138:LYS:HG3	1:A:254:CYS:SG	2.59	0.43
1:A:210:ASN:HD21	1:A:412:LEU:CD2	2.31	0.43
1:A:232[B]:MSE:HE1	1:A:239:GLN:HA	2.00	0.43
1:A:312:SER:O	1:A:313:ARG:C	2.54	0.43
1:A:109:ASN:HB3	1:A:112:LEU:HD12	2.01	0.42
1:A:195:LYS:CG	1:A:197:SER:H	2.27	0.42
1:A:130:ALA:O	1:A:262:PRO:HA	2.20	0.42
1:A:219:LEU:HD11	1:A:407:MSE:CE	2.49	0.42
1:A:399:ARG:HG2	1:A:400:THR:HG23	2.02	0.41
1:A:270:SER:HB3	1:A:404:PHE:CZ	2.55	0.41
1:A:359:ARG:H	1:A:360:ALA:HB2	1.85	0.41
1:A:267:LEU:HD23	1:A:406:GLN:OE1	2.21	0.40
1:A:360:ALA:CA	1:A:361:ARG:HB2	2.35	0.40
1:A:232[B]:MSE:HE2	1:A:239:GLN:HA	2.02	0.40
1:A:336:LEU:HD23	1:A:399:ARG:HH11	1.86	0.40
1:A:170:SER:HB3	1:A:344:TRP:CE3	2.56	0.40

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	320/354 (90%)	309 (97%)	10 (3%)	1 (0%)	46 42

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	361	ARG

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	271/284 (95%)	242 (89%)	29 (11%)	8 4

All (29) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	114	LYS
1	A	136	LYS
1	A	148	VAL
1	A	169	SER
1	A	172	SER
1	A	195	LYS
1	A	203	ARG
1	A	209	THR
1	A	222	SER
1	A	233	SER

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	237	ASN
1	A	248	GLU
1	A	250	SER
1	A	261	ASN
1	A	265	VAL
1	A	266	ASN
1	A	274	SER
1	A	280	GLU
1	A	303	GLU
1	A	307	LEU
1	A	313	ARG
1	A	317	GLU
1	A	318	SER
1	A	335	GLU
1	A	343	ASN
1	A	347	LYS
1	A	359	ARG
1	A	399	ARG
1	A	412	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	210	ASN

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 carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

5.7 Other polymers

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	320/354 (90%)	0.56	37 (11%)	6	10	18, 35, 76, 90	0

All (37) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	314	ALA	9.7
1	A	313	ARG	7.0
1	A	316	GLY	6.6
1	A	359	ARG	6.5
1	A	361	ARG	6.2
1	A	312	SER	5.9
1	A	266	ASN	5.2
1	A	315	GLY	5.1
1	A	358	GLY	4.9
1	A	261	ASN	4.5
1	A	303	GLU	4.2
1	A	362	THR	4.1
1	A	263	ASN	3.9
1	A	264	LEU	3.6
1	A	117	THR	3.5
1	A	260	ALA	3.4
1	A	360	ALA	3.3
1	A	78	LYS	3.3
1	A	273	LYS	3.2
1	A	214	ASN	3.1
1	A	411	SER	3.0
1	A	363	GLY	2.9
1	A	389	PRO	2.9
1	A	330	ALA	2.7
1	A	299	ARG	2.6
1	A	335	GLU	2.6
1	A	326	VAL	2.5

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	412	LEU	2.3
1	A	331	VAL	2.3
1	A	151	SER	2.2
1	A	311	LEU	2.2
1	A	357	ALA	2.2
1	A	352	PHE	2.1
1	A	371	LEU	2.1
1	A	399	ARG	2.1
1	A	308	SER	2.1
1	A	387	GLY	2.1

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 (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.