wwPDB X-ray Structure Validation Summary Report (i)

Oct 8, 2023 – 10:35 AM EDT

PDB ID	:	6ECO
Title	:	Hexamer-2-Foldon HIV-1 capsid platform
Authors	:	Summers, B.J.; Xiong, Y.
Deposited on	:	2018-08-08
Resolution	:	4.20 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 (1)) were used in the production of this report:

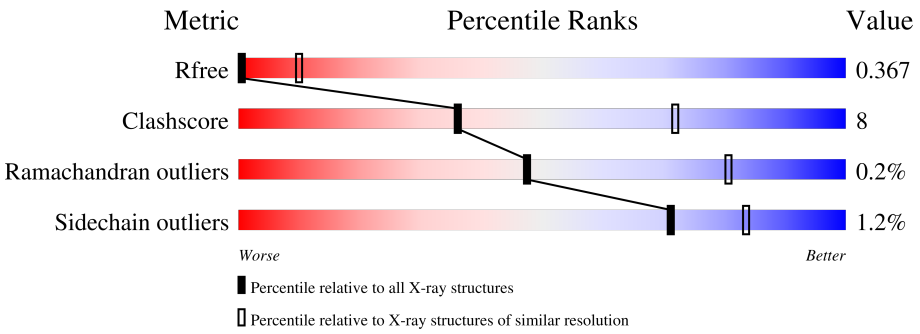
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 4.20 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	1005 (4.62-3.78)
Clashscore	141614	1044 (4.60-3.80)
Ramachandran outliers	138981	1000 (4.60-3.80)
Sidechain outliers	138945	1007 (4.62-3.78)

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%

Mol	Chain	Length	Quality of chain
1	A	253	77% 6% 17%
2	D	221	90% 6% • •

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3252 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 HIV-1 capsid platform protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	210	1598	1004	279	302	13	0	0	0

There are 33 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	6	LEU	ILE	conflict	UNP P04591
A	45	CYS	GLU	engineered mutation	UNP P04591
A	54	GLU	THR	engineered mutation	UNP P04591
A	83	LEU	VAL	conflict	UNP P04591
A	120	HIS	ASN	conflict	UNP P04591
A	208	GLY	ALA	conflict	UNP P04591
A	227	GLY	-	expression tag	UNP P04591
A	228	TYR	-	expression tag	UNP P04591
A	229	ILE	-	expression tag	UNP P04591
A	230	PRO	-	expression tag	UNP P04591
A	231	GLU	-	expression tag	UNP P04591
A	232	ALA	-	expression tag	UNP P04591
A	233	PRO	-	expression tag	UNP P04591
A	234	ARG	-	expression tag	UNP P04591
A	235	ASP	-	expression tag	UNP P04591
A	236	GLY	-	expression tag	UNP P04591
A	237	GLN	-	expression tag	UNP P04591
A	238	ALA	-	expression tag	UNP P04591
A	239	TYR	-	expression tag	UNP P04591
A	240	VAL	-	expression tag	UNP P04591
A	241	ARG	-	expression tag	UNP P04591
A	242	LYS	-	expression tag	UNP P04591
A	243	ASP	-	expression tag	UNP P04591
A	244	GLY	-	expression tag	UNP P04591
A	245	GLU	-	expression tag	UNP P04591
A	246	TRP	-	expression tag	UNP P04591
A	247	VAL	-	expression tag	UNP P04591

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	248	LEU	-	expression tag	UNP P04591
A	249	LEU	-	expression tag	UNP P04591
A	250	SER	-	expression tag	UNP P04591
A	251	THR	-	expression tag	UNP P04591
A	252	PHE	-	expression tag	UNP P04591
A	253	LEU	-	expression tag	UNP P04591

- Molecule 2 is a protein called HIV-1 capsid platform protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	217	1654	1045	286	311	12	0	0	0

There are 7 discrepancies between the modelled and reference sequences:

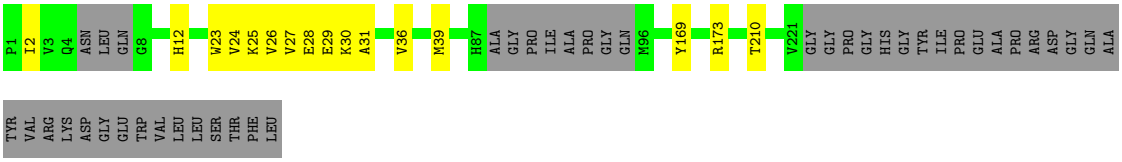
Chain	Residue	Modelled	Actual	Comment	Reference
D	6	LEU	ILE	conflict	UNP P04591
D	14	CYS	ALA	engineered mutation	UNP P04591
D	42	GLU	ALA	engineered mutation	UNP P04591
D	83	LEU	VAL	conflict	UNP P04591
D	120	HIS	ASN	conflict	UNP P04591
D	204	ASP	ALA	engineered mutation	UNP P04591
D	208	GLY	ALA	conflict	UNP P04591

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. 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. 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: HIV-1 capsid platform protein

Chain A:



- Molecule 2: HIV-1 capsid platform protein

Chain D:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 3 2 1	Depositor
Cell constants a, b, c, α, β, γ	95.25Å 95.25Å 122.57Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	30.00 – 4.20 29.03 – 4.20	Depositor EDS
% Data completeness (in resolution range)	98.1 (30.00-4.20) 98.6 (29.03-4.20)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.46 (at 4.26Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.286 , 0.316 0.321 , 0.367	Depositor DCC
R_{free} test set	214 reflections (4.36%)	wwPDB-VP
Wilson B-factor (Å^2)	148.6	Xtriage
Anisotropy	0.594	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 204.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.055 for -h,-k,l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	3252	wwPDB-VP
Average B, all atoms (Å^2)	264.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.39% 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 (i)

5.1 Standard geometry (i)

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	0.60	0/1631	0.64	0/2215
2	D	0.59	0/1691	0.64	0/2302
All	All	0.60	0/3322	0.64	0/4517

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	1598	0	1545	40	0
2	D	1654	0	1607	15	0
All	All	3252	0	3152	51	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 8.

The worst 5 of 51 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:30:LYS:CD	1:A:36:VAL:HG22	1.60	1.31
1:A:26:VAL:HG13	1:A:30:LYS:CE	1.61	1.29

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:26:VAL:HG13	1:A:30:LYS:CD	1.61	1.29
1:A:26:VAL:CG1	1:A:30:LYS:HD2	1.68	1.23
1:A:30:LYS:NZ	1:A:39:MET:HG2	1.51	1.22

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	204/253 (81%)	200 (98%)	4 (2%)	0	100	100
2	D	213/221 (96%)	205 (96%)	7 (3%)	1 (0%)	29	68
All	All	417/474 (88%)	405 (97%)	11 (3%)	1 (0%)	47	80

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	147	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	166/212 (78%)	166 (100%)	0	100	100
2	D	174/191 (91%)	170 (98%)	4 (2%)	50	70

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	340/403 (84%)	336 (99%)	4 (1%)	71 83

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	141	ILE
2	D	145	TYR
2	D	150	ILE
2	D	175	GLU

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

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 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 (i)

6.1 Protein, DNA and RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

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

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.