Full wwPDB X-ray Structure Validation Report ⓘ

May 27, 2020 – 03:33 am BST

PDB ID	:	3GV2
Title	:	X-ray Structure of Hexameric HIV-1 CA
Authors	:	Kelly, B.N.
Deposited on	:	2009-03-30
Resolution	:	7.00 Å(reported)

The following versions of software and data (see references ⓘ) were used in the production of this report:

|		
MolProbity	:	4.02b-467
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 7.00 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.
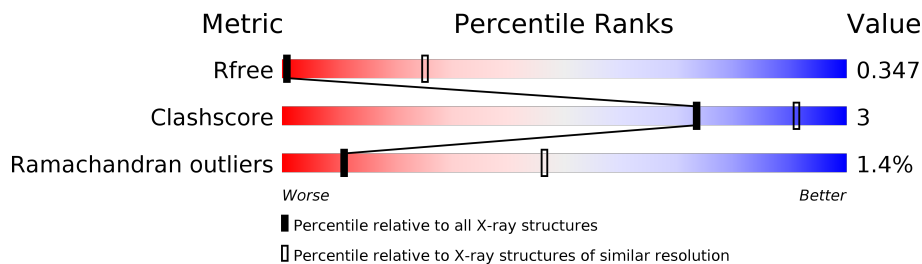
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1004 (10.00-3.90)
Clashscore	141614	1069 (10.00-3.90)
Ramachandran outliers	138981	1002 (10.00-3.90)

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%

Mol	Chain	Length	Quality of chain
1	A	342	61% · · 36%
1	B	342	63% · 36%
1	C	342	62% · · 36%
1	D	342	61% · 36%
1	E	342	63% · 36%
1	F	342	62% · · 36%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 5256 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 protein p24,Carbon dioxide-concentrating mechanism protein CcmK homolog 4.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	219	Total 876	C 438	N 219	O 219	0	0	0
1	B	219	Total 876	C 438	N 219	O 219	0	0	0
1	C	219	Total 876	C 438	N 219	O 219	0	0	0
1	D	219	Total 876	C 438	N 219	O 219	0	0	0
1	E	219	Total 876	C 438	N 219	O 219	0	0	0
1	F	219	Total 876	C 438	N 219	O 219	0	0	0

There are 102 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	184	ALA	TRP	engineered mutation	UNP P12493
A	185	ALA	MET	engineered mutation	UNP P12493
A	220	GLY	-	linker	UNP P12493
A	221	VAL	-	linker	UNP P12493
A	222	GLY	-	linker	UNP P12493
A	223	GLY	-	linker	UNP P12493
A	224	THR	-	linker	UNP P12493
A	225	ARG	-	linker	UNP P12493
A	226	PRO	-	linker	UNP P12493
A	227	GLU	-	linker	UNP P12493
A	228	LEU	-	linker	UNP P12493
A	332	TYR	GLU	conflict	UNP P73407
A	338	GLU	ASN	conflict	UNP P73407
A	339	VAL	-	expression tag	UNP P73407
A	340	LEU	-	expression tag	UNP P73407
A	341	PHE	-	expression tag	UNP P73407

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	342	GLN	-	expression tag	UNP P73407
B	184	ALA	TRP	engineered mutation	UNP P12493
B	185	ALA	MET	engineered mutation	UNP P12493
B	220	GLY	-	linker	UNP P12493
B	221	VAL	-	linker	UNP P12493
B	222	GLY	-	linker	UNP P12493
B	223	GLY	-	linker	UNP P12493
B	224	THR	-	linker	UNP P12493
B	225	ARG	-	linker	UNP P12493
B	226	PRO	-	linker	UNP P12493
B	227	GLU	-	linker	UNP P12493
B	228	LEU	-	linker	UNP P12493
B	332	TYR	GLU	conflict	UNP P73407
B	338	GLU	ASN	conflict	UNP P73407
B	339	VAL	-	expression tag	UNP P73407
B	340	LEU	-	expression tag	UNP P73407
B	341	PHE	-	expression tag	UNP P73407
B	342	GLN	-	expression tag	UNP P73407
C	184	ALA	TRP	engineered mutation	UNP P12493
C	185	ALA	MET	engineered mutation	UNP P12493
C	220	GLY	-	linker	UNP P12493
C	221	VAL	-	linker	UNP P12493
C	222	GLY	-	linker	UNP P12493
C	223	GLY	-	linker	UNP P12493
C	224	THR	-	linker	UNP P12493
C	225	ARG	-	linker	UNP P12493
C	226	PRO	-	linker	UNP P12493
C	227	GLU	-	linker	UNP P12493
C	228	LEU	-	linker	UNP P12493
C	332	TYR	GLU	conflict	UNP P73407
C	338	GLU	ASN	conflict	UNP P73407
C	339	VAL	-	expression tag	UNP P73407
C	340	LEU	-	expression tag	UNP P73407
C	341	PHE	-	expression tag	UNP P73407
C	342	GLN	-	expression tag	UNP P73407
D	184	ALA	TRP	engineered mutation	UNP P12493
D	185	ALA	MET	engineered mutation	UNP P12493
D	220	GLY	-	linker	UNP P12493
D	221	VAL	-	linker	UNP P12493
D	222	GLY	-	linker	UNP P12493
D	223	GLY	-	linker	UNP P12493
D	224	THR	-	linker	UNP P12493

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	225	ARG	-	linker	UNP P12493
D	226	PRO	-	linker	UNP P12493
D	227	GLU	-	linker	UNP P12493
D	228	LEU	-	linker	UNP P12493
D	332	TYR	GLU	conflict	UNP P73407
D	338	GLU	ASN	conflict	UNP P73407
D	339	VAL	-	expression tag	UNP P73407
D	340	LEU	-	expression tag	UNP P73407
D	341	PHE	-	expression tag	UNP P73407
D	342	GLN	-	expression tag	UNP P73407
E	184	ALA	TRP	engineered mutation	UNP P12493
E	185	ALA	MET	engineered mutation	UNP P12493
E	220	GLY	-	linker	UNP P12493
E	221	VAL	-	linker	UNP P12493
E	222	GLY	-	linker	UNP P12493
E	223	GLY	-	linker	UNP P12493
E	224	THR	-	linker	UNP P12493
E	225	ARG	-	linker	UNP P12493
E	226	PRO	-	linker	UNP P12493
E	227	GLU	-	linker	UNP P12493
E	228	LEU	-	linker	UNP P12493
E	332	TYR	GLU	conflict	UNP P73407
E	338	GLU	ASN	conflict	UNP P73407
E	339	VAL	-	expression tag	UNP P73407
E	340	LEU	-	expression tag	UNP P73407
E	341	PHE	-	expression tag	UNP P73407
E	342	GLN	-	expression tag	UNP P73407
F	184	ALA	TRP	engineered mutation	UNP P12493
F	185	ALA	MET	engineered mutation	UNP P12493
F	220	GLY	-	linker	UNP P12493
F	221	VAL	-	linker	UNP P12493
F	222	GLY	-	linker	UNP P12493
F	223	GLY	-	linker	UNP P12493
F	224	THR	-	linker	UNP P12493
F	225	ARG	-	linker	UNP P12493
F	226	PRO	-	linker	UNP P12493
F	227	GLU	-	linker	UNP P12493
F	228	LEU	-	linker	UNP P12493
F	332	TYR	GLU	conflict	UNP P73407
F	338	GLU	ASN	conflict	UNP P73407
F	339	VAL	-	expression tag	UNP P73407
F	340	LEU	-	expression tag	UNP P73407

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	341	PHE	-	expression tag	UNP P73407
F	342	GLN	-	expression tag	UNP P73407

3 Residue-property plots ⓘ

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. 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.
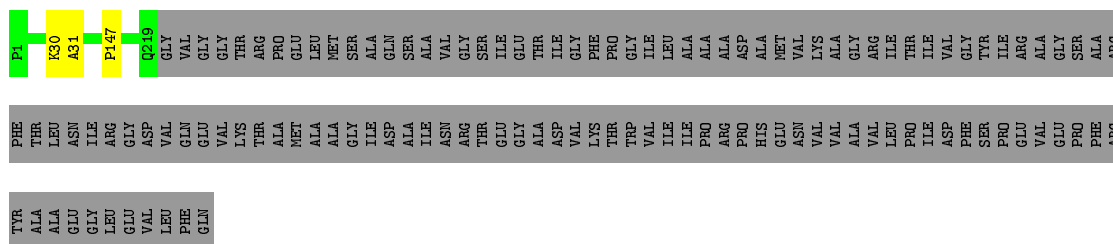
• Molecule 1: Capsid protein p24,Carbon dioxide-concentrating mechanism protein CcmK homolog 4

• Molecule 1: Capsid protein p24,Carbon dioxide-concentrating mechanism protein CcmK homolog 4

• Molecule 1: Capsid protein p24,Carbon dioxide-concentrating mechanism protein CcmK homolog 4

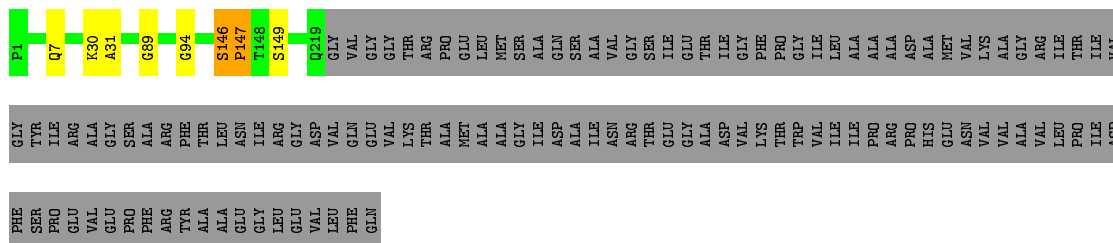• Molecule 1: Capsid protein p24,Carbon dioxide-concentrating mechanism protein CcmK homolog 4
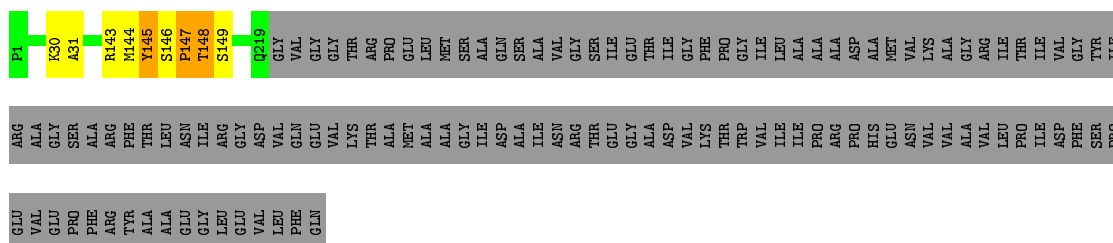
Chain D:

• Molecule 1: Capsid protein p24,Carbon dioxide-concentrating mechanism protein CcmK homolog 4

Chain E:

• Molecule 1: Capsid protein p24,Carbon dioxide-concentrating mechanism protein CcmK homolog 4

Chain F:

4   Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	90.62Å 156.44Å 196.64Å 90.00° 100.35° 90.00°	Depositor
Resolution (Å)	49.75 – 7.00 49.75 – 6.95	Depositor EDS
% Data completeness (in resolution range)	92.7 (49.75-7.00) 81.2 (49.75-6.95)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	6.10 (at 6.68Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.282 , 0.323 0.337 , 0.347	Depositor DCC
R_{free} test set	435 reflections (10.06%)	wwPDB-VP
Wilson B-factor (Å^2)	161.2	Xtriage
Anisotropy	1.361	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 341.8	EDS
L-test for twinning[2]	$<|L|> = 0.47, <L^2> = 0.30$	Xtriage
Estimated twinning fraction	0.116 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.71	EDS
Total number of atoms	5256	wwPDB-VP
Average B, all atoms (Å^2)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.99% 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	0.57	0/875	1.02	0/1092
1	B	0.56	0/875	1.02	0/1092
1	C	0.57	0/875	1.02	0/1092
1	D	0.57	0/875	1.03	0/1092
1	E	0.57	0/875	1.03	0/1092
1	F	0.57	0/875	1.03	0/1092
All	All	0.57	0/5250	1.02	0/6552

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	876	0	230	7	0
1	B	876	0	230	1	0
1	C	876	0	230	3	2
1	D	876	0	230	5	0
1	E	876	0	230	2	0
1	F	876	0	230	4	0
All	All	5256	0	1380	22	2

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 3.

All (22) 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:147:PRO:CA	1:A:149:SER:N	2.40	0.84
1:F:147:PRO:O	1:F:148:THR:O	2.02	0.78
1:D:149:SER:O	1:D:151:LEU:N	2.24	0.69
1:A:143:ARG:C	1:A:145:TYR:H	1.96	0.68
1:A:147:PRO:CA	1:A:149:SER:H	2.13	0.62
1:C:146:SER:O	1:C:147:PRO:C	2.39	0.62
1:A:143:ARG:O	1:A:145:TYR:N	2.35	0.59
1:F:146:SER:O	1:F:147:PRO:C	2.41	0.58
1:A:147:PRO:CA	1:A:148:THR:C	2.75	0.53
1:D:149:SER:O	1:D:150:ILE:C	2.46	0.52
1:D:149:SER:C	1:D:151:LEU:N	2.63	0.51
1:E:149:SER:O	1:E:152:ASP:N	2.37	0.51
1:F:146:SER:O	1:F:147:PRO:O	2.30	0.49
1:C:146:SER:O	1:C:147:PRO:O	2.30	0.49
1:A:143:ARG:C	1:A:145:TYR:N	2.63	0.46
1:D:149:SER:O	1:D:152:ASP:N	2.32	0.45
1:B:30:LYS:O	1:B:31:ALA:C	2.57	0.42
1:A:30:LYS:O	1:A:31:ALA:C	2.57	0.42
1:C:30:LYS:O	1:C:31:ALA:C	2.57	0.42
1:D:30:LYS:O	1:D:31:ALA:C	2.57	0.41
1:F:30:LYS:O	1:F:31:ALA:C	2.57	0.40
1:E:30:LYS:O	1:E:31:ALA:C	2.57	0.40

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:94:GLY:O	1:C:94:GLY:O[2_756]	1.21	0.99
1:C:7:GLN:O	1:C:89:GLY:CA[2_756]	1.82	0.38

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	217/342 (64%)	204 (94%)	8 (4%)	5 (2%)	6	34
1	B	217/342 (64%)	207 (95%)	9 (4%)	1 (0%)	29	69
1	C	217/342 (64%)	206 (95%)	8 (4%)	3 (1%)	11	46
1	D	217/342 (64%)	205 (94%)	8 (4%)	4 (2%)	8	40
1	E	217/342 (64%)	206 (95%)	10 (5%)	1 (0%)	29	69
1	F	217/342 (64%)	206 (95%)	7 (3%)	4 (2%)	8	40
All	All	1302/2052 (64%)	1234 (95%)	50 (4%)	18 (1%)	11	46

All (18) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	144	MET
1	A	147	PRO
1	C	146	SER
1	C	147	PRO
1	C	149	SER
1	D	146	SER
1	D	147	PRO
1	D	148	THR
1	F	147	PRO
1	F	148	THR
1	A	145	TYR
1	A	148	THR
1	D	150	ILE
1	B	147	PRO
1	F	149	SER
1	A	146	SER
1	F	146	SER
1	E	147	PRO

5.3.2   Protein sidechains (i)

There are no protein residues with a non-rotameric sidechain to report in this entry.

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)

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.