



Full wwPDB X-ray Structure Validation Report ⓘ

May 26, 2020 – 07:13 am BST

PDB ID : 6E1I
Title : HLA-A*0201 single chain trimer with murine H2K alpha 3 domain and HPV.16
E7 peptide YMLDLQPET
Authors : Finton, K.A.K.; Rusnac, D.V.
Deposited on : 2018-07-09
Resolution : 1.99 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

<https://www.wwpdb.org/validation/2017/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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

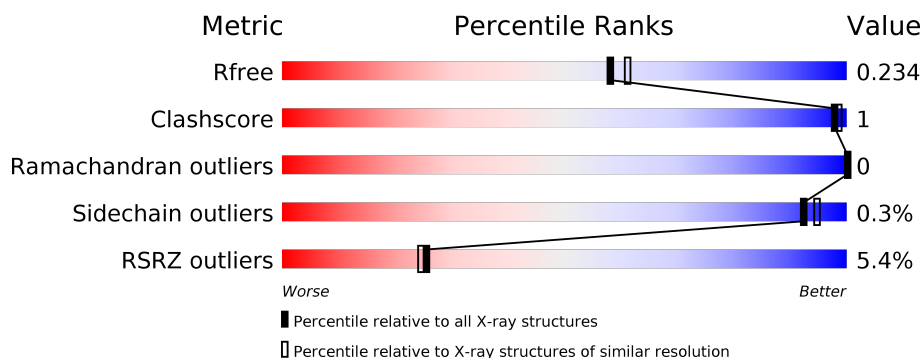
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.99 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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 $\leq 5\%$. 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	442	<div> <div>5%</div> <div> <div></div> <div>86%</div> <div>•</div> <div>12%</div> </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 3240 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 HLA-A*0201 single chain trimer with murine H2K alpha 3 domain and HPV.16 E7 peptide YMLDLQPET.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	389	Total	C	N	O	S	0	2	0
			3089	1956	548	570	15			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	10	GLY	-	linker	PDB ?
A	11	GLY	-	linker	PDB ?
A	12	GLY	-	linker	PDB ?
A	13	GLY	-	linker	PDB ?
A	14	SER	-	linker	PDB ?
A	15	GLY	-	linker	PDB ?
A	16	GLY	-	linker	PDB ?
A	17	GLY	-	linker	PDB ?
A	18	GLY	-	linker	PDB ?
A	19	SER	-	linker	PDB ?
A	20	GLY	-	linker	PDB ?
A	21	GLY	-	linker	PDB ?
A	22	GLY	-	linker	PDB ?
A	23	GLY	-	linker	PDB ?
A	24	SER	-	linker	PDB ?
A	25	ILE	-	linker	PDB ?
A	26	GLN	-	linker	PDB ?
A	27	ARG	-	linker	PDB ?
A	57	SER	PRO	conflict	UNP P01887
A	58	ASP	HIS	conflict	UNP P01887
A	78	LEU	MET	conflict	UNP P01887
A	124	GLY	-	linker	UNP P01887
A	125	GLY	-	linker	UNP P01887
A	126	GLY	-	linker	UNP P01887
A	127	GLY	-	linker	UNP P01887
A	128	SER	-	linker	UNP P01887

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	129	GLY	-	linker	UNP P01887
A	130	GLY	-	linker	UNP P01887
A	131	GLY	-	linker	UNP P01887
A	132	GLY	-	linker	UNP P01887
A	133	SER	-	linker	UNP P01887
A	134	GLY	-	linker	UNP P01887
A	135	GLY	-	linker	UNP P01887
A	136	GLY	-	linker	UNP P01887
A	137	GLY	-	linker	UNP P01887
A	138	SER	-	linker	UNP P01887
A	139	GLY	-	linker	UNP P01887
A	140	GLY	-	linker	UNP P01887
A	141	GLY	-	linker	UNP P01887
A	142	GLY	-	linker	UNP P01887
A	143	SER	-	linker	UNP P01887
A	227	ALA	TYR	conflict	UNP P01892
A	407	GLU	LYS	conflict	UNP P01902
A	418	GLU	-	expression tag	UNP P01902
A	419	ASN	-	expression tag	UNP P01902
A	420	LEU	-	expression tag	UNP P01902
A	421	TYR	-	expression tag	UNP P01902
A	422	PHE	-	expression tag	UNP P01902

- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	151	Total O 151 151	0	0

4 Data and refinement statistics

Property	Value	Source
Space group	P 2 21 21	Depositor
Cell constants a, b, c, α , β , γ	52.47Å 82.19Å 107.17Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 1.99 47.12 – 1.99	Depositor EDS
% Data completeness (in resolution range)	99.8 (50.00-1.99) 99.8 (47.12-1.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	4.48 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, R_{free}	0.204 , 0.233 0.211 , 0.234	Depositor DCC
R_{free} test set	1664 reflections (5.16%)	wwPDB-VP
Wilson B-factor (Å ²)	27.2	Xtriage
Anisotropy	0.050	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.36 , 41.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	3240	wwPDB-VP
Average B, all atoms (Å ²)	32.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.93% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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	0.37	0/3187	0.62	1/4341 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	178	ARG	NE-CZ-NH1	5.67	123.14	120.30

There are no chirality outliers.

There are no planarity outliers.

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	3089	0	2850	4	0
2	A	151	0	0	0	0
All	All	3240	0	2850	4	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 1.

All (4) 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:334:TYR:CE1	1:A:342:VAL:HG11	2.44	0.52

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:84:TRP:CE2	1:A:260:ALA:HB2	2.50	0.47
1:A:191[A]:ARG:HA	1:A:191[A]:ARG:HD3	1.82	0.45
1:A:403:HIS:ND1	1:A:414:THR:HG22	2.36	0.41

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	385/442 (87%)	375 (97%)	10 (3%)	0	100	100

There are no Ramachandran outliers to report.

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	311/362 (86%)	310 (100%)	1 (0%)	92	95

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	94	PHE

Some 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 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](#)

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, 95th 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	389/442 (88%)	0.50	21 (5%) 25 24	17, 30, 58, 89	0

All (21) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	11	GLY	6.2
1	A	364	GLY	5.6
1	A	334	TYR	4.8
1	A	368	THR	4.0
1	A	10	GLY	3.6
1	A	400	TYR	3.4
1	A	366	ASP	3.3
1	A	340	VAL	3.1
1	A	12	GLY	3.0
1	A	422	PHE	2.9
1	A	281	MET	2.7
1	A	279	ALA	2.6
1	A	363	ASN	2.6
1	A	417	TRP	2.5
1	A	288	HIS	2.5
1	A	361	GLN	2.4
1	A	335	HIS	2.3
1	A	229	ASN	2.3
1	A	367	LEU	2.2
1	A	99	THR	2.1
1	A	394	LEU	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.