



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 14, 2022 – 02:44 PM EST

PDB ID : 7SR0
Title : Single chain trimer HLA-A*02:01 (H98L, Y108C) with HPV.16 E7 peptide YMLDLQPET
Authors : Finton, K.A.K.; Rupert, P.B.
Deposited on : 2021-11-07
Resolution : 2.54 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 ⓘ 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 ⓘ](#)) 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	:	2.31.2
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)	:	2.31.2

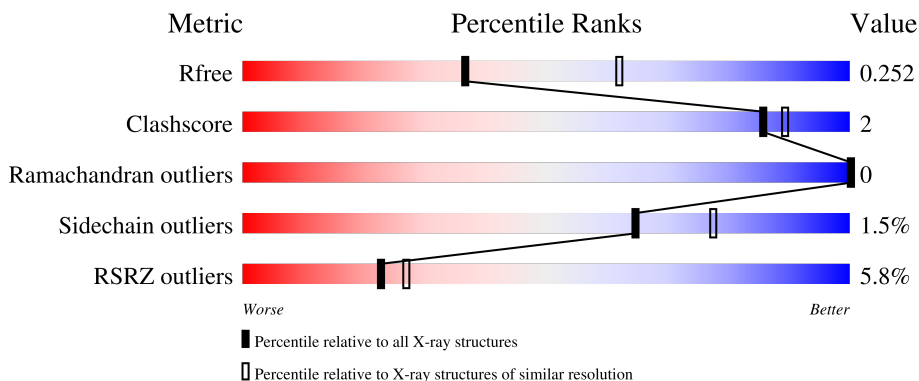
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION


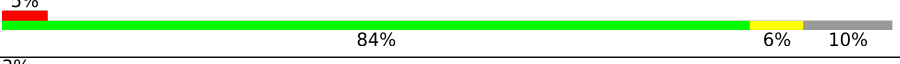
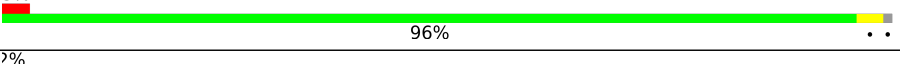
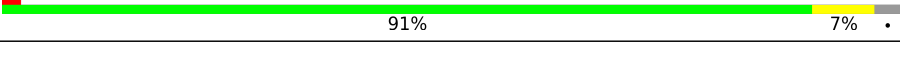
The reported resolution of this entry is 2.54 Å.

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	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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 $\leq 5\%$. 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	424	
1	C	424	
2	B	116	
2	D	116	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7626 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 Protein E7 peptide,Beta-2-microglobulin,MHC class I antigen chimera.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	370	Total	C	N	O	S	0	0	0
			2841	1794	501	533	13			
1	C	383	Total	C	N	O	S	0	0	0
			2927	1845	518	551	13			

There are 86 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	10	GLY	-	linker	UNP P03129
A	11	CYS	-	linker	UNP P03129
A	12	GLY	-	linker	UNP P03129
A	13	GLY	-	linker	UNP P03129
A	14	SER	-	linker	UNP P03129
A	15	GLY	-	linker	UNP P03129
A	16	GLY	-	linker	UNP P03129
A	17	GLY	-	linker	UNP P03129
A	18	GLY	-	linker	UNP P03129
A	19	SER	-	linker	UNP P03129
A	20	GLY	-	linker	UNP P03129
A	21	GLY	-	linker	UNP P03129
A	22	GLY	-	linker	UNP P03129
A	23	GLY	-	linker	UNP P03129
A	24	SER	-	linker	UNP P03129
A	124	GLY	-	linker	UNP P16213
A	125	GLY	-	linker	UNP P16213
A	126	GLY	-	linker	UNP P16213
A	127	GLY	-	linker	UNP P16213
A	128	SER	-	linker	UNP P16213
A	129	GLY	-	linker	UNP P16213
A	130	GLY	-	linker	UNP P16213
A	131	GLY	-	linker	UNP P16213
A	132	GLY	-	linker	UNP P16213

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	133	SER	-	linker	UNP P16213
A	134	GLY	-	linker	UNP P16213
A	135	GLY	-	linker	UNP P16213
A	136	GLY	-	linker	UNP P16213
A	137	GLY	-	linker	UNP P16213
A	138	SER	-	linker	UNP P16213
A	139	GLY	-	linker	UNP P16213
A	140	GLY	-	linker	UNP P16213
A	141	GLY	-	linker	UNP P16213
A	142	GLY	-	linker	UNP P16213
A	143	SER	-	linker	UNP P16213
A	217	LEU	HIS	engineered mutation	UNP A0A678ZGP6
A	227	CYS	TYR	engineered mutation	UNP A0A678ZGP6
A	419	HIS	-	expression tag	UNP A0A678ZGP6
A	420	HIS	-	expression tag	UNP A0A678ZGP6
A	421	HIS	-	expression tag	UNP A0A678ZGP6
A	422	HIS	-	expression tag	UNP A0A678ZGP6
A	423	HIS	-	expression tag	UNP A0A678ZGP6
A	424	HIS	-	expression tag	UNP A0A678ZGP6
C	10	GLY	-	linker	UNP P03129
C	11	CYS	-	linker	UNP P03129
C	12	GLY	-	linker	UNP P03129
C	13	GLY	-	linker	UNP P03129
C	14	SER	-	linker	UNP P03129
C	15	GLY	-	linker	UNP P03129
C	16	GLY	-	linker	UNP P03129
C	17	GLY	-	linker	UNP P03129
C	18	GLY	-	linker	UNP P03129
C	19	SER	-	linker	UNP P03129
C	20	GLY	-	linker	UNP P03129
C	21	GLY	-	linker	UNP P03129
C	22	GLY	-	linker	UNP P03129
C	23	GLY	-	linker	UNP P03129
C	24	SER	-	linker	UNP P03129
C	124	GLY	-	linker	UNP P16213
C	125	GLY	-	linker	UNP P16213
C	126	GLY	-	linker	UNP P16213
C	127	GLY	-	linker	UNP P16213
C	128	SER	-	linker	UNP P16213
C	129	GLY	-	linker	UNP P16213
C	130	GLY	-	linker	UNP P16213
C	131	GLY	-	linker	UNP P16213

Continued on next page...

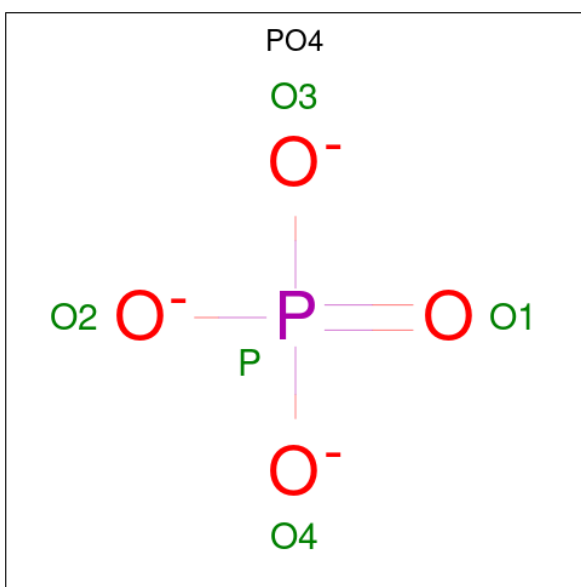
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	132	GLY	-	linker	UNP P16213
C	133	SER	-	linker	UNP P16213
C	134	GLY	-	linker	UNP P16213
C	135	GLY	-	linker	UNP P16213
C	136	GLY	-	linker	UNP P16213
C	137	GLY	-	linker	UNP P16213
C	138	SER	-	linker	UNP P16213
C	139	GLY	-	linker	UNP P16213
C	140	GLY	-	linker	UNP P16213
C	141	GLY	-	linker	UNP P16213
C	142	GLY	-	linker	UNP P16213
C	143	SER	-	linker	UNP P16213
C	217	LEU	HIS	engineered mutation	UNP A0A678ZGP6
C	227	CYS	TYR	engineered mutation	UNP A0A678ZGP6
C	419	HIS	-	expression tag	UNP A0A678ZGP6
C	420	HIS	-	expression tag	UNP A0A678ZGP6
C	421	HIS	-	expression tag	UNP A0A678ZGP6
C	422	HIS	-	expression tag	UNP A0A678ZGP6
C	423	HIS	-	expression tag	UNP A0A678ZGP6
C	424	HIS	-	expression tag	UNP A0A678ZGP6

- Molecule 2 is a protein called VHH.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	115	Total 846	C 522	N 146	O 174	S 4	0	0	0
2	D	113	Total 843	C 522	N 145	O 172	S 4	0	0	0

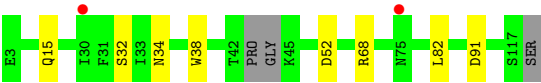
- Molecule 3 is PHOSPHATE ION (three-letter code: PO4) (formula: O₄P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	O	P	0	0
			5	4	1		
3	C	1	Total	O	P	0	0
			5	4	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	45	Total	O	0	0
			45	45		
4	B	29	Total	O	0	0
			29	29		
4	C	54	Total	O	0	0
			54	54		
4	D	31	Total	O	0	0
			31	31		



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	117.71Å 117.71Å 261.84Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.84 – 2.54 48.84 – 2.54	Depositor EDS
% Data completeness (in resolution range)	97.2 (48.84-2.54) 97.2 (48.84-2.54)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.32 (at 2.54Å)	Xtriage
Refinement program	REFMAC 1.19.1_4122, PHENIX 1.19.1_4122	Depositor
R, R_{free}	0.217 , 0.242 0.228 , 0.252	Depositor DCC
R_{free} test set	2956 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å ²)	47.5	Xtriage
Anisotropy	0.153	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 33.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	7626	wwPDB-VP
Average B, all atoms (Å ²)	56.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 24.38 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 3.8764e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹ Intensities estimated from amplitudes.

² Theoretical 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: PO4

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	0.24	0/2918	0.50	0/3977
1	C	0.24	0/3009	0.50	0/4110
2	B	0.24	0/861	0.51	0/1168
2	D	0.24	0/856	0.53	0/1157
All	All	0.24	0/7644	0.50	0/10412

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	2841	0	2532	13	0
1	C	2927	0	2582	15	0
2	B	846	0	786	2	0
2	D	843	0	797	4	0
3	A	5	0	0	0	0
3	C	5	0	0	0	0
4	A	45	0	0	0	0
4	B	29	0	0	0	0
4	C	54	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	D	31	0	0	0	0
All	All	7626	0	6697	31	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 2.

The worst 5 of 31 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:C:149:ARG:NH2	1:C:245:ASP:OD1	2.11	0.84
1:A:178:ARG:HG2	1:A:191:ARG:HD3	1.73	0.70
2:D:68:ARG:NH2	2:D:91:ASP:OD2	2.25	0.67
1:A:331:HIS:HD2	1:A:347:TRP:HB2	1.62	0.64
2:B:68:ARG:NH2	2:B:91:ASP:OD2	2.31	0.63

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	360/424 (85%)	352 (98%)	8 (2%)	0	100	100
1	C	377/424 (89%)	366 (97%)	11 (3%)	0	100	100
2	B	113/116 (97%)	109 (96%)	4 (4%)	0	100	100
2	D	109/116 (94%)	107 (98%)	2 (2%)	0	100	100
All	All	959/1080 (89%)	934 (97%)	25 (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	273/348 (78%)	269 (98%)	4 (2%)	65	77
1	C	278/348 (80%)	274 (99%)	4 (1%)	67	79
2	B	89/97 (92%)	88 (99%)	1 (1%)	73	83
2	D	90/97 (93%)	88 (98%)	2 (2%)	52	66
All	All	730/890 (82%)	719 (98%)	11 (2%)	65	77

5 of 11 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	178	ARG
1	C	375	GLU
2	D	32	SER
2	D	15	GLN
2	B	21	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 10 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	215	GLN
1	C	284	GLN
1	C	298	GLN
1	A	331	HIS
1	C	175	GLN

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](#)

2 ligands are 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	$\# Z > 2$
3	PO4	C	501	-	4,4,4	0.91	0	6,6,6	0.43	0
3	PO4	A	501	-	4,4,4	0.91	0	6,6,6	0.42	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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	370/424 (87%)	0.34	29 (7%) 13 16	36, 55, 112, 154	0
1	C	383/424 (90%)	0.24	23 (6%) 21 25	36, 53, 105, 139	0
2	B	115/116 (99%)	0.03	3 (2%) 56 62	34, 44, 92, 112	0
2	D	113/116 (97%)	0.06	2 (1%) 68 74	35, 43, 92, 108	0
All	All	981/1080 (90%)	0.23	57 (5%) 23 27	34, 51, 106, 154	0

The worst 5 of 57 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	370	ASP	5.9
1	A	360	TRP	5.1
1	A	359	THR	4.8
1	C	342	ALA	4.3
1	A	369	GLN	4.1

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 monosaccharides in this entry.

6.4 Ligands [i](#)

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, 95th 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(\AA^2)	Q<0.9
3	PO4	A	501	5/5	0.93	0.13	73,74,82,84	0
3	PO4	C	501	5/5	0.95	0.11	66,70,73,73	0

6.5 Other polymers [i](#)

There are no such residues in this entry.