



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 14, 2022 – 02:38 PM EST

PDB ID : 7SQP
Title : Single chain trimer HLA-A*02:01 (Y108C) with HPV.16 E7 peptide
YMLDLQPETTDL
Authors : Finton, K.A.K.; Rupert, P.B.
Deposited on : 2021-11-05
Resolution : 2.53 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
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

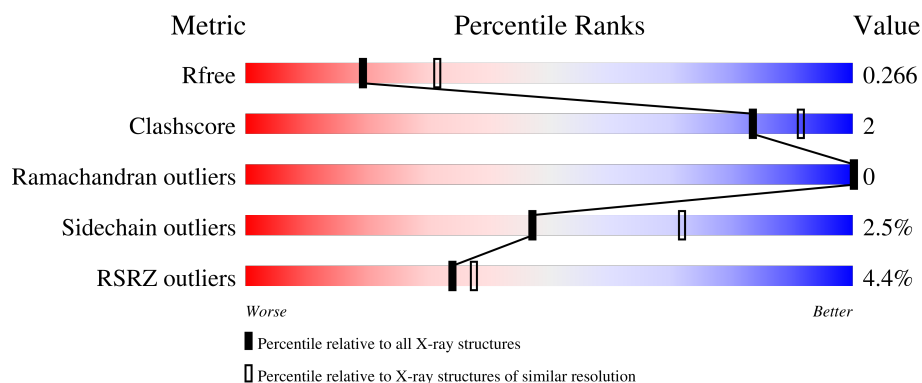
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION


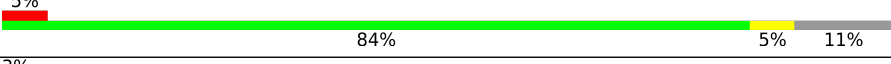
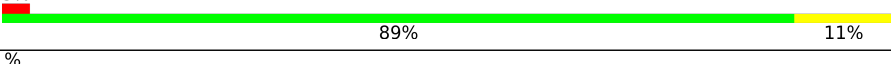
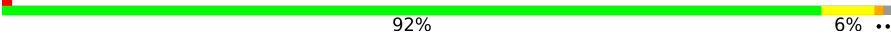
The reported resolution of this entry is 2.53 Å.

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	5743 (2.54-2.50)
Clashscore	141614	6463 (2.54-2.50)
Ramachandran outliers	138981	6335 (2.54-2.50)
Sidechain outliers	138945	6337 (2.54-2.50)
RSRZ outliers	127900	5630 (2.54-2.50)

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	427	 4% 83% 5% • 11%
1	C	427	 5% 84% 5% 11%
2	B	116	 3% 89% 11%
2	D	116	 % 92% 6% ••

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 7611 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	382	Total	C	N	O	S	0	0	0
			2910	1842	514	542	12			
1	C	380	Total	C	N	O	S	0	0	0
			2880	1827	508	534	11			

There are 84 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	13	GLY	-	linker	UNP P03129
A	13A	GLY	-	linker	UNP P03129
A	13B	GLY	-	linker	UNP P03129
A	13C	GLY	-	linker	UNP P03129
A	13D	SER	-	linker	UNP P03129
A	13E	GLY	-	linker	UNP P03129
A	13F	GLY	-	linker	UNP P03129
A	13G	GLY	-	linker	UNP P03129
A	13H	GLY	-	linker	UNP P03129
A	13I	SER	-	linker	UNP P03129
A	13J	GLY	-	linker	UNP P03129
A	13K	GLY	-	linker	UNP P03129
A	13L	GLY	-	linker	UNP P03129
A	13M	GLY	-	linker	UNP P03129
A	13N	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	227	ALA	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	12A	GLY	-	linker	UNP P03129
C	12B	GLY	-	linker	UNP P03129
C	12C	GLY	-	linker	UNP P03129
C	12D	GLY	-	linker	UNP P03129
C	12E	SER	-	linker	UNP P03129
C	12F	GLY	-	linker	UNP P03129
C	12G	GLY	-	linker	UNP P03129
C	12H	GLY	-	linker	UNP P03129
C	12I	GLY	-	linker	UNP P03129
C	12J	SER	-	linker	UNP P03129
C	12K	GLY	-	linker	UNP P03129
C	12L	GLY	-	linker	UNP P03129
C	12M	GLY	-	linker	UNP P03129
C	12N	GLY	-	linker	UNP P03129
C	12O	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
C	132	GLY	-	linker	UNP P16213

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
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	227	ALA	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	116	Total	C	N	O	S	0	0	0
			865	537	148	176	4			
2	D	115	Total	C	N	O	S	0	0	0
			848	523	147	174	4			

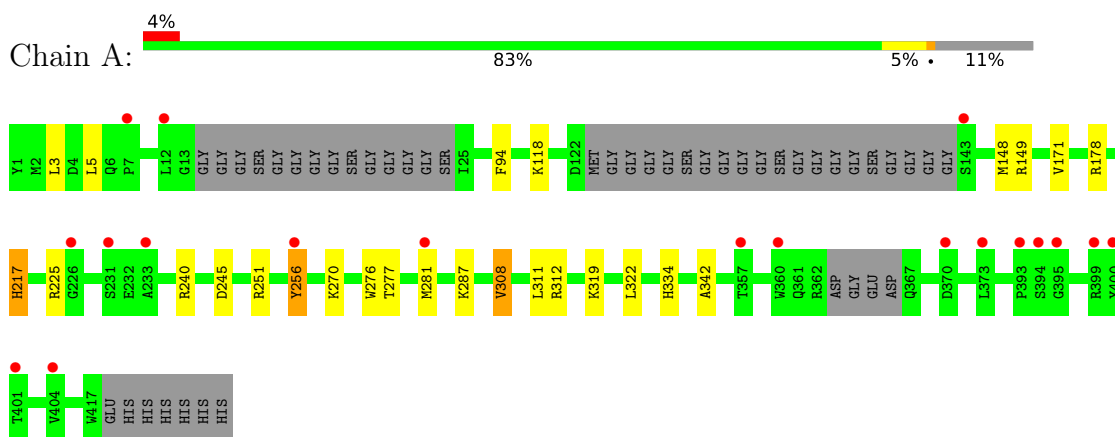
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	44	Total	O	0	0
			44	44		
3	B	25	Total	O	0	0
			25	25		
3	C	22	Total	O	0	0
			22	22		
3	D	17	Total	O	0	0
			17	17		

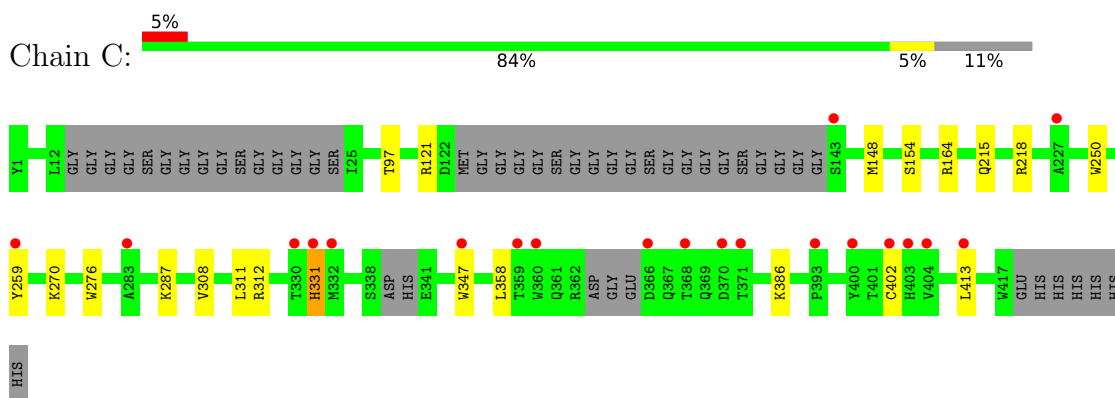
3 Residue-property plots [i](#)

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 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.

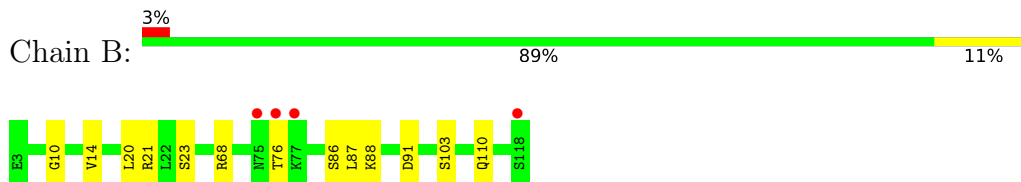
- Molecule 1: Protein E7 peptide,Beta-2-microglobulin,MHC class I antigen chimera



- Molecule 1: Protein E7 peptide,Beta-2-microglobulin,MHC class I antigen chimera

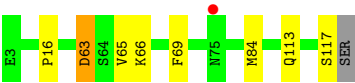


- Molecule 2: VHH



- Molecule 2: VHH





4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	117.98Å 117.98Å 262.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.96 – 2.53 48.96 – 2.53	Depositor EDS
% Data completeness (in resolution range)	98.0 (48.96-2.53) 97.9 (48.96-2.53)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	3.02 (at 2.54Å)	Xtriage
Refinement program		Depositor
R, R_{free}	0.239 , 0.260 0.245 , 0.266	Depositor DCC
R_{free} test set	3089 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å ²)	56.2	Xtriage
Anisotropy	0.222	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.29 , 38.9	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	7611	wwPDB-VP
Average B, all atoms (Å ²)	67.0	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 38.65 % 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 3.5716e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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 [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	0.23	0/2993	0.49	0/4087
1	C	0.23	0/2961	0.49	0/4045
2	B	0.25	0/881	0.49	0/1195
2	D	0.24	0/863	0.51	0/1170
All	All	0.24	0/7698	0.49	0/10497

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	2910	0	2566	15	0
1	C	2880	0	2544	10	0
2	B	865	0	812	5	0
2	D	848	0	787	5	0
3	A	44	0	0	0	0
3	B	25	0	0	1	0
3	C	22	0	0	0	0
3	D	17	0	0	1	0
All	All	7611	0	6709	34	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 34 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:118:LYS:NZ	2:B:103:SER:O	2.12	0.82
1:A:149:ARG:NH2	1:A:245:ASP:OD2	2.18	0.75
2:B:68:ARG:NH2	2:B:91:ASP:OD2	2.28	0.63
1:C:276:TRP:HB2	1:C:287:LYS:HD2	1.82	0.61
2:D:16:PRO:HD3	2:D:117:SER:HB2	1.83	0.60

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	374/427 (88%)	360 (96%)	14 (4%)	0	100	100
1	C	370/427 (87%)	356 (96%)	14 (4%)	0	100	100
2	B	114/116 (98%)	111 (97%)	3 (3%)	0	100	100
2	D	113/116 (97%)	109 (96%)	4 (4%)	0	100	100
All	All	971/1086 (89%)	936 (96%)	35 (4%)	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	273/349 (78%)	266 (97%)	7 (3%)	46	70
1	C	269/349 (77%)	264 (98%)	5 (2%)	57	79
2	B	92/97 (95%)	87 (95%)	5 (5%)	22	40
2	D	89/97 (92%)	88 (99%)	1 (1%)	73	88
All	All	723/892 (81%)	705 (98%)	18 (2%)	47	72

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	331	HIS
2	D	63	ASP
1	C	413	LEU
2	B	23	SER
1	C	270	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	197	GLN
1	A	217	HIS
1	C	215	GLN
1	C	331	HIS

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](#)

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	382/427 (89%)	0.17	19 (4%) 28 31	43, 65, 118, 147	0
1	C	380/427 (88%)	0.23	20 (5%) 26 28	43, 64, 130, 178	0
2	B	116/116 (100%)	0.03	4 (3%) 45 49	41, 53, 96, 121	0
2	D	115/116 (99%)	0.04	1 (0%) 84 86	41, 51, 98, 125	0
All	All	993/1086 (91%)	0.16	44 (4%) 34 37	41, 60, 121, 178	0

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	347	TRP	5.7
1	C	400	TYR	4.7
1	C	370	ASP	3.9
1	C	143	SER	3.8
1	A	370	ASP	3.8

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.