



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2018 – 10:44 AM EDT

PDB ID : 5NQK  
Title : human 199.16 TCR in complex with Melan-A/MART-1 (26-35) peptide and HLA-A2  
Authors : Exertier, C.; Reiser, J.-B.; Lantez, V.; Chouquet, A.; Bonneville, M.; Saulquin, X.; Housset, D.  
Deposited on : 2017-04-20  
Resolution : 3.25 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](mailto: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	:	rb-20031021
Percentile statistics	:	20171227.v01 (using entries in the PDB archive December 27th 2017)
Refmac	:	5.8.0158
CCP4	:	7.0 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20031021

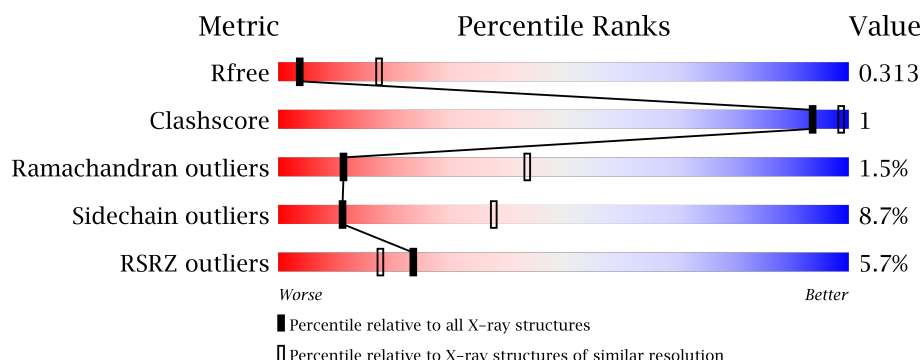
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.25 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))
$R_{free}$	111664	1008 (3.30-3.22)
Clashscore	122126	1066 (3.30-3.22)
Ramachandran outliers	120053	1046 (3.30-3.22)
Sidechain outliers	120020	1045 (3.30-3.22)
RSRZ outliers	108989	1993 (3.32-3.20)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. 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	H	276	<div> <div>7%</div> <div> <div></div> <div>87%</div> <div>12%</div> <div>.</div> </div> </div>
2	L	100	<div> <div>%</div> <div> <div></div> <div>86%</div> <div>13%</div> <div>.</div> </div> </div>
3	P	10	<div> <div></div> <div>100%</div> </div>
4	A	211	<div> <div>8%</div> <div> <div></div> <div>82%</div> <div>10%</div> <div>7%</div> </div> </div>
5	B	251	<div> <div>5%</div> <div> <div></div> <div>84%</div> <div>12%</div> <div>..</div> </div> </div>

## 2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 6604 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 class I histocompatibility antigen, A-2 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	H	275	Total	C	N	O	S	0	0	0
			2248	1405	409	425	9			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	245	VAL	ALA	conflict	UNP P01892

- Molecule 2 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	L	100	Total	C	N	O	S	0	0	0
			837	533	141	159	4			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	0	MET	-	initiating methionine	UNP P61769

- Molecule 3 is a protein called Melanoma antigen recognized by T-cells 1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	P	10	Total	C	N	O	0	0	0
			69	45	10	14			

- Molecule 4 is a protein called T-cell receptor alpha variable 12-2,T-cell receptor alpha joining 45,T-cell receptor alpha chain C region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	A	197	Total	C	N	O	S	0	0	0
			1518	945	251	314	8			

There are 11 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	SER	initiating methionine	UNP A0A075B6T6
A	49	SER	PHE	conflict	UNP A0A075B6T6
A	91	GLY	VAL	conflict	UNP A0A075B6T6
A	158	CYS	THR	engineered mutation	UNP A0A1B0GUM0
A	203	ASN	-	expression tag	UNP A0A1B0GUM0
A	204	ASP	-	expression tag	UNP A0A1B0GUM0
A	205	GLY	-	expression tag	UNP A0A1B0GUM0
A	206	GLY	-	expression tag	UNP A0A1B0GUM0
A	207	GLY	-	expression tag	UNP A0A1B0GUM0
A	208	CYS	-	expression tag	UNP A0A1B0GUM0
A	209	LYS	-	expression tag	UNP A0A1B0GUM0

- Molecule 5 is a protein called T-cell receptor beta variable 19,TRB protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
5	B	243	Total	C	N	O	S	0	0	1
			1931	1219	334	371	7			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	2	MET	-	initiating methionine	UNP A0A5B3
B	95	GLN	-	linker	UNP A0A5B3
B	96	GLY	-	linker	UNP A0A5B3
B	97	LEU	-	linker	UNP A0A5B3
B	98	ALA	-	linker	UNP A0A5B3
B	99	GLY	-	linker	UNP A0A5B3
B	100	ALA	-	linker	UNP A0A5B3
B	171	CYS	SER	conflict	UNP A0A0C4ZKA8
B	245	GLN	-	expression tag	UNP A0A0C4ZKA8
B	246	ASP	-	expression tag	UNP A0A0C4ZKA8
B	247	ARG	-	expression tag	UNP A0A0C4ZKA8
B	248	GLY	-	expression tag	UNP A0A0C4ZKA8
B	249	GLY	-	expression tag	UNP A0A0C4ZKA8
B	250	GLY	-	expression tag	UNP A0A0C4ZKA8
B	251	CYS	-	expression tag	UNP A0A0C4ZKA8
B	252	ASP	-	expression tag	UNP A0A0C4ZKA8

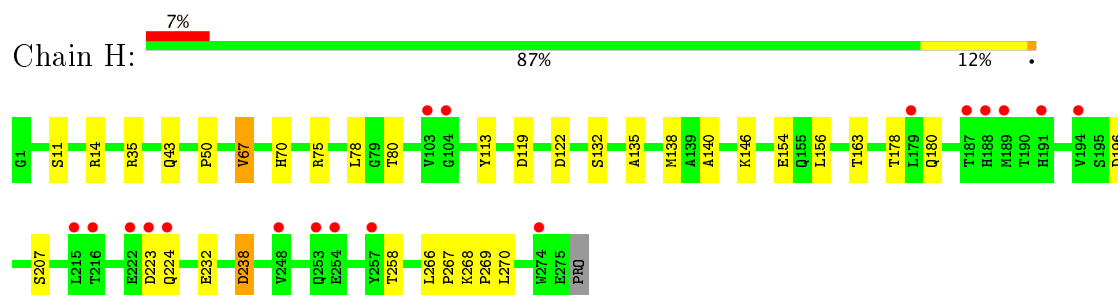
- Molecule 6 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	B	1	Total	Na	0	0
			1	1		

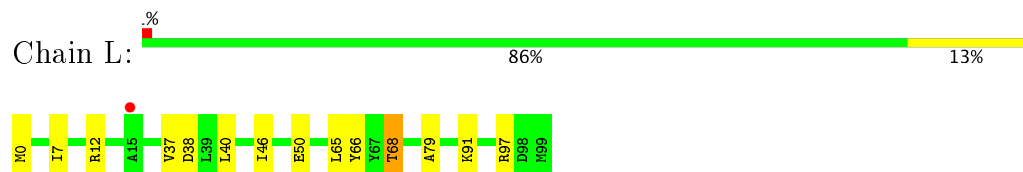
### 3 Residue-property plots [i](#)

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 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: HLA class I histocompatibility antigen, A-2 alpha chain



- Molecule 2: Beta-2-microglobulin

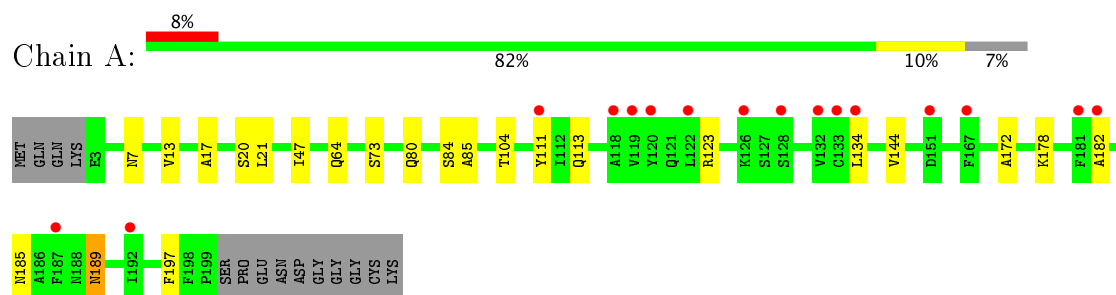


- Molecule 3: Melanoma antigen recognized by T-cells 1

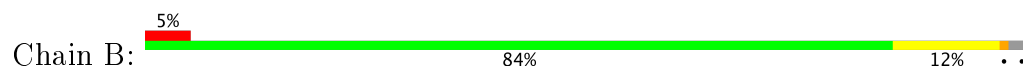


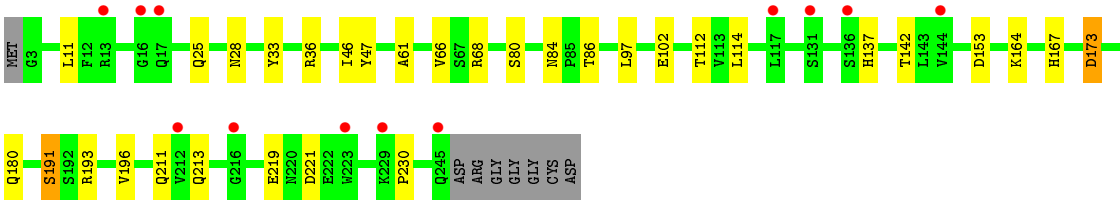
There are no outlier residues recorded for this chain.

- Molecule 4: T-cell receptor alpha variable 12-2, T-cell receptor alpha joining 45, T-cell receptor alpha chain C region



- Molecule 5: T-cell receptor beta variable 19, TRB protein





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	225.18 Å   49.72 Å   95.72 Å 90.00°   97.69°   90.00°	Depositor
Resolution (Å)	48.53 – 3.25 48.53 – 3.25	Depositor EDS
% Data completeness (in resolution range)	85.6 (48.53-3.25) 85.7 (48.53-3.25)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.17	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	0.89 (at 3.25 Å)	Xtriage
Refinement program	REFMAC 5.8.0158	Depositor
R, $R_{free}$	0.224   ,   0.316 0.227   ,   0.313	Depositor DCC
$R_{free}$ test set	832 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	77.0	Xtriage
Anisotropy	0.693	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 86.8	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.92	EDS
Total number of atoms	6604	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	112.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.59% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>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: NA

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	H	0.45	0/2313	0.68	0/3140
2	L	0.44	0/860	0.67	0/1162
3	P	0.42	0/68	0.71	0/90
4	A	0.46	0/1551	0.65	0/2101
5	B	0.46	0/1982	0.67	0/2694
All	All	0.45	0/6774	0.67	0/9187

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	H	2248	0	2100	3	0
2	L	837	0	803	5	0
3	P	69	0	79	0	0
4	A	1518	0	1421	7	0
5	B	1931	0	1833	6	0
6	B	1	0	0	0	0
All	All	6604	0	6236	19	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.

The worst 5 of 19 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)
4:A:13:VAL:HG13	4:A:17:ALA:HB3	1.78	0.64
2:L:46:ILE:HG21	2:L:68:THR:HG21	1.85	0.58
4:A:197:PHE:CB	5:B:137:HIS:HE2	2.20	0.54
4:A:144:VAL:HG21	4:A:172:ALA:CB	2.38	0.52
2:L:7:ILE:HD12	2:L:91:LYS:HD3	1.93	0.50

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	H	273/276 (99%)	238 (87%)	29 (11%)	6 (2%)	7	35
2	L	98/100 (98%)	91 (93%)	7 (7%)	0	100	100
3	P	8/10 (80%)	8 (100%)	0	0	100	100
4	A	195/211 (92%)	167 (86%)	27 (14%)	1 (0%)	31	66
5	B	241/251 (96%)	205 (85%)	31 (13%)	5 (2%)	8	36
All	All	815/848 (96%)	709 (87%)	94 (12%)	12 (2%)	11	43

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	H	119	ASP
1	H	163	THR
1	H	238	ASP
1	H	269	PRO
5	B	68	ARG

### 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	H	232/233 (100%)	207 (89%)	25 (11%)	7	28
2	L	95/95 (100%)	89 (94%)	6 (6%)	20	52
3	P	7/7 (100%)	7 (100%)	0	100	100
4	A	171/182 (94%)	160 (94%)	11 (6%)	19	51
5	B	209/215 (97%)	189 (90%)	20 (10%)	9	32
All	All	714/732 (98%)	652 (91%)	62 (9%)	11	38

5 of 62 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	L	50	GLU
4	A	64	GLN
5	B	196	VAL
4	A	7	ASN
4	A	80	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
4	A	114	ASN
4	A	143	ASN
5	B	84	ASN
4	A	113	GLN
5	B	56	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

## 5.5 Carbohydrates ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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 ⓘ

There are no such residues in this entry.

## 5.8 Polymer linkage issues ⓘ

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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	H	275/276 (99%)	0.41	18 (6%) 19 15	62, 105, 173, 213	0
2	L	100/100 (100%)	0.19	1 (1%) 82 77	62, 86, 135, 166	0
3	P	10/10 (100%)	0.15	0 100 100	70, 76, 90, 98	0
4	A	197/211 (93%)	0.52	16 (8%) 12 9	73, 110, 173, 197	0
5	B	243/251 (96%)	0.33	12 (4%) 29 23	66, 112, 162, 195	0
All	All	825/848 (97%)	0.38	47 (5%) 24 18	62, 107, 171, 213	0

The worst 5 of 47 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	A	181	PHE	9.7
1	H	254	GLU	7.1
1	H	253	GLN	6.0
1	H	191	HIS	4.6
4	A	182	ALA	4.3

### 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](#)

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, 95<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
6	NA	B	301	1/1	0.93	0.20	49,49,49,49	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.