



# Full wwPDB X-ray Structure Validation Report ⓘ

May 25, 2020 – 03:30 pm BST

PDB ID : 5GSR  
Title : Mouse MHC class I H-2Kd with a MERS-CoV-derived peptide I5A  
Authors : Liu, K.; Chai, Y.; Qi, J.; Tan, W.; Liu, W.J.; Gao, G.F.  
Deposited on : 2016-08-17  
Resolution : 2.20 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](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 : 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

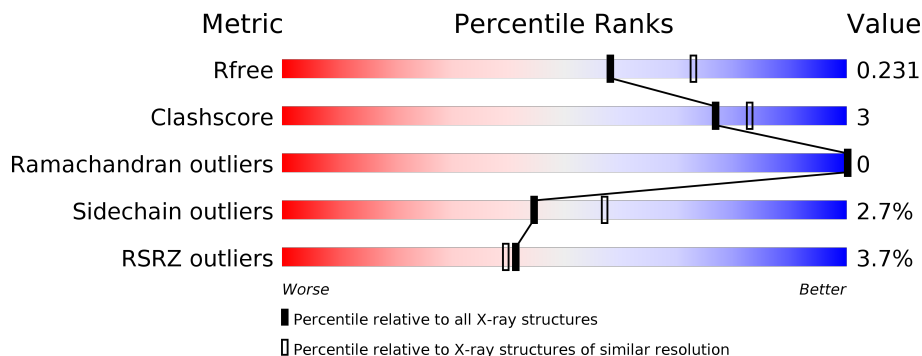
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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 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	B	100	5% (Poor fit), 89% (0-1 outliers), 11% (2-3+ outliers)
1	D	100	2% (Poor fit), 96% (0-1 outliers), 2% (2-3+ outliers)
2	A	274	4% (Poor fit), 89% (0-1 outliers), 10% (2-3+ outliers)
2	C	274	4% (Poor fit), 92% (0-1 outliers), 7% (2-3+ outliers)
3	P	9	78% (0-1 outliers), 22% (2-3+ outliers)
3	Q	9	78% (0-1 outliers), 11% (2 outliers), 11% (3+ outliers)

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 6690 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 Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	D	100	833	531	141	158	3	0	0	0
1	B	100	833	531	141	158	3	0	0	0

- Molecule 2 is a protein called H-2 class I histocompatibility antigen, K-D alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	274	2261	1432	401	421	7	0	0	0
2	C	274	2261	1432	401	421	7	0	0	0

- Molecule 3 is a protein called 9-mer peptide from Spike protein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	P	9	74	50	11	13	0	0	0
3	Q	9	74	50	11	13	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	5	ALA	ILE	engineered mutation	UNP A0A0U2W1D8
Q	5	ALA	ILE	engineered mutation	UNP A0A0U2W1D8

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	D	60	Total	O	0	0
			60	60		

*Continued on next page...*

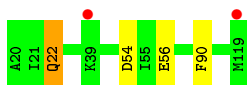
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>	<b>ZeroOcc</b>	<b>AltConf</b>
4	B	61	Total O 61 61	0	0
4	A	127	Total O 127 127	0	0
4	C	101	Total O 101 101	0	0
4	P	4	Total O 4 4	0	0
4	Q	1	Total O 1 1	0	0

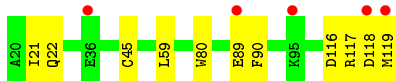
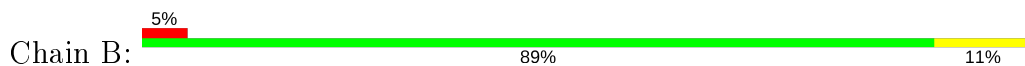
### 3 Residue-property plots [i](#)

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.

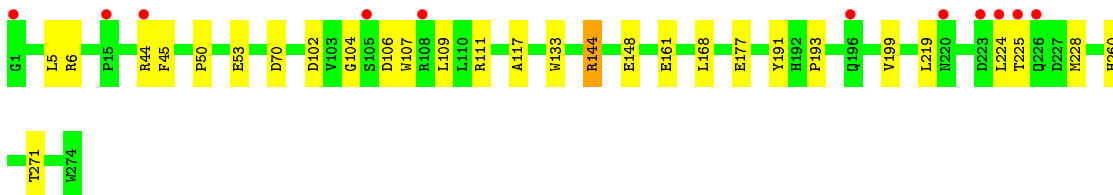
- Molecule 1: Beta-2-microglobulin



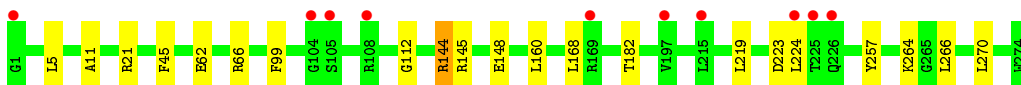
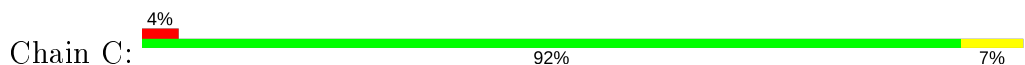
- Molecule 1: Beta-2-microglobulin



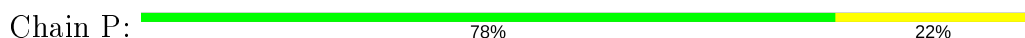
- Molecule 2: H-2 class I histocompatibility antigen, K-D alpha chain



- Molecule 2: H-2 class I histocompatibility antigen, K-D alpha chain

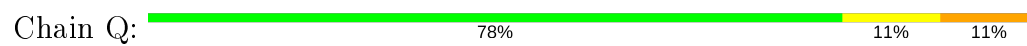


- Molecule 3: 9-mer peptide from Spike protein





- Molecule 3: 9-mer peptide from Spike protein



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	50.35Å 79.66Å 122.71Å 90.00° 90.30° 90.00°	Depositor
Resolution (Å)	31.83 – 2.20 48.61 – 2.20	Depositor EDS
% Data completeness (in resolution range)	97.8 (31.83-2.20) 96.9 (48.61-2.20)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.42 (at 2.20Å)	Xtrriage
Refinement program	PHENIX 1.10.1_2155	Depositor
R, $R_{free}$	0.201 , 0.231 0.201 , 0.231	Depositor DCC
$R_{free}$ test set	2391 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	35.8	Xtrriage
Anisotropy	0.537	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 22.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.447 for h,-k,-l	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6690	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.77% 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](#)

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	B	0.49	0/856	0.52	0/1159
1	D	0.37	0/856	0.48	0/1159
2	A	0.41	0/2329	0.52	0/3169
2	C	0.38	0/2329	0.50	0/3169
3	P	0.37	0/77	0.64	0/105
3	Q	0.24	0/77	0.36	0/105
All	All	0.41	0/6524	0.50	0/8866

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	B	833	0	796	5	0
1	D	833	0	796	3	0
2	A	2261	0	2125	18	0
2	C	2261	0	2125	11	0
3	P	74	0	71	1	0
3	Q	74	0	71	1	0
4	A	127	0	0	4	0
4	B	61	0	0	1	0
4	C	101	0	0	0	0
4	D	60	0	0	3	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	P	4	0	0	1	0
4	Q	1	0	0	0	0
All	All	6690	0	5984	37	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 3.

All (37) 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)
2:A:148:GLU:OE1	4:A:301:HOH:O	2.01	0.77
1:B:89:GLU:O	4:B:201:HOH:O	2.17	0.59
2:C:144:ARG:HD2	2:C:148:GLU:OE2	2.06	0.56
2:A:102:ASP:OD2	4:A:302:HOH:O	2.17	0.56
2:A:191:TYR:CZ	2:A:199:VAL:HG21	2.42	0.55
2:A:177:GLU:O	4:A:303:HOH:O	2.18	0.55
2:A:104:GLY:O	2:A:107:TRP:N	2.37	0.53
2:A:5:LEU:HB2	2:A:168:LEU:HD13	1.91	0.53
1:D:56:GLU:OE2	4:D:201:HOH:O	2.18	0.52
2:C:62:GLU:OE2	2:C:66:ARG:NE	2.37	0.50
1:B:116:ASP:C	1:B:118:ASP:H	2.16	0.49
2:C:112:GLY:HA3	2:C:160:LEU:HD13	1.96	0.48
1:D:54:ASP:OD1	4:D:202:HOH:O	2.20	0.48
2:A:191:TYR:CE2	2:A:199:VAL:HG21	2.48	0.48
2:C:219:LEU:HD13	2:C:257:TYR:CZ	2.48	0.48
2:A:104:GLY:O	2:A:107:TRP:HD1	1.96	0.48
2:A:50:PRO:O	2:A:53:GLU:HG2	2.12	0.48
2:A:177:GLU:C	4:A:303:HOH:O	2.52	0.47
1:B:45:CYS:HB2	1:B:59:LEU:HD21	1.96	0.47
2:A:109:LEU:HD22	2:A:161:GLU:HG2	1.96	0.47
2:C:99:PHE:CE1	3:Q:3:SER:HB3	2.51	0.46
2:C:266:LEU:HD13	2:C:270:LEU:HG	1.98	0.46
2:A:193:PRO:HA	2:A:199:VAL:HG23	1.98	0.45
2:C:5:LEU:HB2	2:C:168:LEU:HD13	1.99	0.45
1:D:22:GLN:NE2	4:D:203:HOH:O	2.29	0.44
2:C:11:ALA:HA	2:C:21:ARG:O	2.19	0.43
2:C:182:THR:OG1	2:C:264:LYS:NZ	2.51	0.43
2:A:225:THR:HG23	2:A:228:MET:CE	2.48	0.43
1:B:80:TRP:CE2	2:A:117:ALA:HB2	2.53	0.43
2:A:260:HIS:HD1	2:A:271:THR:HG22	1.85	0.42
3:P:3:SER:HB3	4:P:102:HOH:O	2.20	0.42

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:219:LEU:HB3	2:A:224:LEU:HD21	2.02	0.42
1:B:21:ILE:HG13	1:B:21:ILE:H	1.69	0.41
2:A:106:ASP:O	2:A:107:TRP:HB2	2.21	0.41
2:C:223:ASP:OD1	2:C:224:LEU:N	2.54	0.41
2:C:266:LEU:HA	2:C:266:LEU:HD23	1.84	0.41
2:A:133:TRP:HB2	2:A:144:ARG:HD3	2.03	0.40

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	B	98/100 (98%)	97 (99%)	1 (1%)	0	100	100
1	D	98/100 (98%)	97 (99%)	1 (1%)	0	100	100
2	A	272/274 (99%)	264 (97%)	8 (3%)	0	100	100
2	C	272/274 (99%)	265 (97%)	7 (3%)	0	100	100
3	P	7/9 (78%)	7 (100%)	0	0	100	100
3	Q	7/9 (78%)	7 (100%)	0	0	100	100
All	All	754/766 (98%)	737 (98%)	17 (2%)	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	B	94/94 (100%)	90 (96%)	4 (4%)	29	36
1	D	94/94 (100%)	92 (98%)	2 (2%)	53	67
2	A	233/233 (100%)	227 (97%)	6 (3%)	46	58
2	C	233/233 (100%)	230 (99%)	3 (1%)	69	81
3	P	8/8 (100%)	7 (88%)	1 (12%)	4	4
3	Q	8/8 (100%)	6 (75%)	2 (25%)	0	0
All	All	670/670 (100%)	652 (97%)	18 (3%)	44	57

All (18) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	22	GLN
1	D	90	PHE
1	B	22	GLN
1	B	90	PHE
1	B	117	ARG
1	B	119	MET
2	A	6	ARG
2	A	44	ARG
2	A	45	PHE
2	A	70	ASP
2	A	111	ARG
2	A	144	ARG
2	C	45	PHE
2	C	144	ARG
2	C	145	ARG
3	P	1	TYR
3	Q	1	TYR
3	Q	3	SER

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, 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	B	100/100 (100%)	0.34	5 (5%) 28 27	24, 37, 67, 85	0
1	D	100/100 (100%)	0.29	2 (2%) 65 63	26, 36, 65, 77	0
2	A	274/274 (100%)	0.31	11 (4%) 38 36	24, 41, 74, 100	0
2	C	274/274 (100%)	0.38	10 (3%) 42 41	22, 41, 77, 96	0
3	P	9/9 (100%)	0.29	0 100 100	32, 36, 45, 46	0
3	Q	9/9 (100%)	0.39	0 100 100	31, 35, 43, 44	0
All	All	766/766 (100%)	0.34	28 (3%) 41 39	22, 39, 72, 100	0

All (28) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	C	225	THR	6.7
2	C	108	ARG	4.4
2	C	226	GLN	3.9
2	A	225	THR	3.9
1	B	119	MET	3.6
2	A	226	GLN	3.5
2	A	224	LEU	3.5
2	C	105	SER	3.1
1	B	36	GLU	3.1
2	C	197	VAL	3.0
2	A	1	GLY	2.9
1	D	39	LYS	2.7
1	D	119	MET	2.6
2	A	105	SER	2.5
2	C	224	LEU	2.4
1	B	95	LYS	2.4
2	A	15	PRO	2.4
1	B	89	GLU	2.4
2	A	108	ARG	2.4

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	A	220	ASN	2.3
2	C	1	GLY	2.3
2	C	215	LEU	2.3
2	A	196	GLN	2.2
2	C	104	GLY	2.1
2	A	44	ARG	2.1
2	C	169	ARG	2.1
1	B	118	ASP	2.0
2	A	223	ASP	2.0

## 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.