



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 20, 2023 – 05:19 PM EST

PDB ID : 6P2S  
Title : Structure of a nested set of N-terminally extended MHC I-peptides provide novel insights into antigen processing and presentation  
Authors : Li, L.; Batliwala, M.; Bouvier, M.  
Deposited on : 2019-05-22  
Resolution : 1.65 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 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 : 2.32.1  
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) : 2.32.1

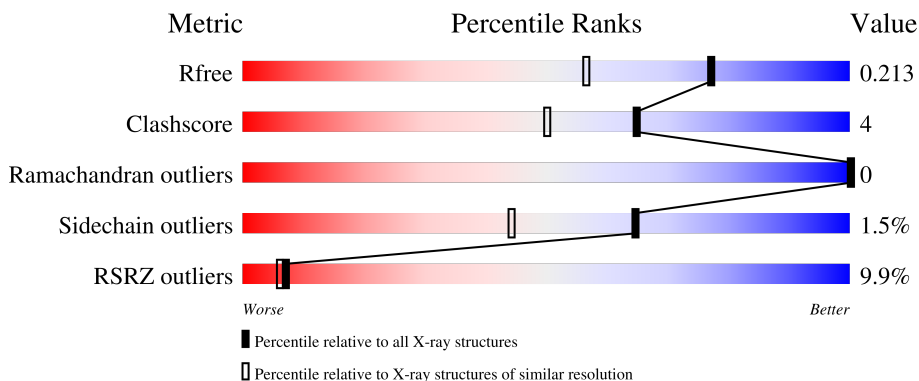
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*


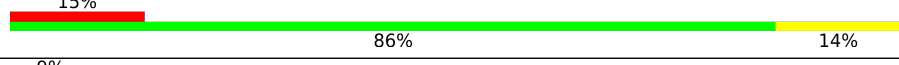

The reported resolution of this entry is 1.65 Å.

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	1827 (1.66-1.66)
Clashscore	141614	1931 (1.66-1.66)
Ramachandran outliers	138981	1891 (1.66-1.66)
Sidechain outliers	138945	1891 (1.66-1.66)
RSRZ outliers	127900	1791 (1.66-1.66)

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  $\geq 3$ , 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	A	276	
2	B	100	
3	C	11	

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 3390 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 MHC class I antigen.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	275	2241	1388	410	435	8	0	0	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	76	CYS	GLU	engineered mutation	UNP R4ZGR5

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

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

There is a discrepancy between the modelled and reference sequences:

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

- Molecule 3 is a protein called MHC I-peptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	9	69	45	12	11	1	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	197	Total	O	0	0
			197	197		
4	B	39	Total	O	0	0
			39	39		

*Continued on next page...*

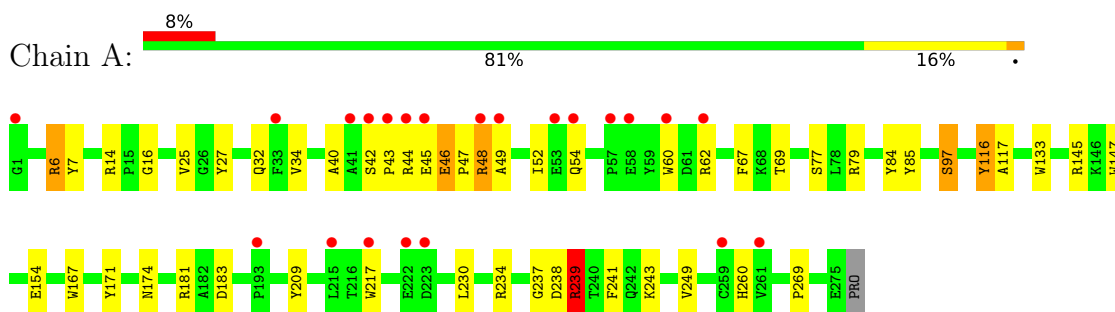
*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
4	C	7	Total	O	0	0
			7	7		

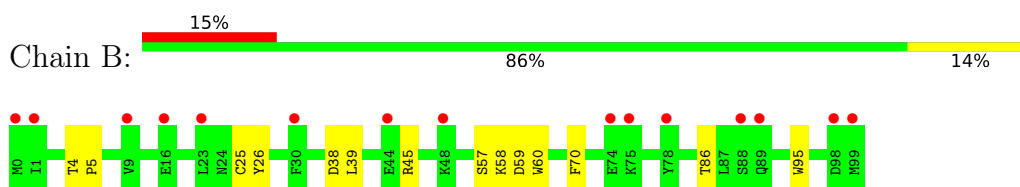
### 3 Residue-property plots [i](#)

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: MHC class I antigen



- Molecule 2: Beta-2-microglobulin



- Molecule 3: MHC I-peptide



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	50.55Å 80.91Å 108.71Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	64.91 – 1.65 19.39 – 1.65	Depositor EDS
% Data completeness (in resolution range)	98.9 (64.91-1.65) 99.0 (19.39-1.65)	Depositor EDS
$R_{merge}$	0.07	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.04 (at 1.65Å)	Xtrriage
Refinement program	REFMAC 5.6.0117	Depositor
R, $R_{free}$	0.204 , 0.214 0.204 , 0.213	Depositor DCC
$R_{free}$ test set	2694 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	26.0	Xtrriage
Anisotropy	0.255	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.39 , 48.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtrriage
Estimated twinning fraction	No twinning to report.	Xtrriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3390	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	34.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.33% 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	A	1.68	25/2302 (1.1%)	1.28	12/3130 (0.4%)
2	B	1.53	2/860 (0.2%)	1.20	2/1162 (0.2%)
3	C	1.35	0/69	1.47	1/88 (1.1%)
All	All	1.64	27/3231 (0.8%)	1.26	15/4380 (0.3%)

All (27) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	97	SER	CB-OG	-9.79	1.29	1.42
1	A	239	ARG	CD-NE	-9.45	1.30	1.46
1	A	77	SER	CB-OG	8.10	1.52	1.42
1	A	167	TRP	CD2-CE2	7.52	1.50	1.41
2	B	57	SER	CA-CB	7.46	1.64	1.52
1	A	154	GLU	CD-OE1	7.24	1.33	1.25
1	A	154	GLU	CD-OE2	6.47	1.32	1.25
1	A	133	TRP	CD2-CE2	6.44	1.49	1.41
1	A	25	VAL	CB-CG2	-6.39	1.39	1.52
1	A	237	GLY	N-CA	6.32	1.55	1.46
1	A	27	TYR	CG-CD1	6.29	1.47	1.39
1	A	116	TYR	CE2-CZ	-6.01	1.30	1.38
1	A	147	TRP	CD2-CE2	5.83	1.48	1.41
2	B	95	TRP	CD2-CE2	5.69	1.48	1.41
1	A	85	TYR	CG-CD1	5.65	1.46	1.39
1	A	260	HIS	CG-CD2	5.62	1.45	1.35
1	A	16	GLY	N-CA	5.59	1.54	1.46
1	A	217	TRP	CD2-CE2	5.52	1.48	1.41
1	A	133	TRP	CG-CD1	5.42	1.44	1.36
1	A	241	PHE	CE2-CZ	5.39	1.47	1.37
1	A	27	TYR	CZ-OH	5.30	1.46	1.37
1	A	69	THR	CB-OG1	5.29	1.53	1.43
1	A	209	TYR	CZ-OH	5.26	1.46	1.37
1	A	249	VAL	CB-CG2	5.16	1.63	1.52
1	A	7	TYR	CZ-OH	5.10	1.46	1.37

*Continued on next page...*

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	269	PRO	N-CA	5.05	1.55	1.47
1	A	209	TYR	CD1-CE1	5.03	1.46	1.39

All (15) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	239	ARG	NE-CZ-NH2	-16.39	112.10	120.30
2	B	59	ASP	CB-CG-OD1	9.03	126.42	118.30
1	A	239	ARG	NE-CZ-NH1	8.04	124.32	120.30
1	A	14	ARG	NE-CZ-NH2	-7.18	116.71	120.30
1	A	27	TYR	CG-CD1-CE1	-6.99	115.71	121.30
1	A	145	ARG	NE-CZ-NH2	-6.38	117.11	120.30
1	A	79	ARG	NE-CZ-NH1	6.19	123.39	120.30
1	A	181	ARG	NE-CZ-NH1	6.01	123.30	120.30
1	A	48	ARG	NE-CZ-NH2	-5.91	117.34	120.30
1	A	116	TYR	CD1-CE1-CZ	-5.72	114.65	119.80
1	A	234	ARG	NE-CZ-NH2	-5.58	117.51	120.30
2	B	26	TYR	CG-CD1-CE1	-5.48	116.92	121.30
3	C	9	LEU	CB-CG-CD2	-5.41	101.81	111.00
1	A	84	TYR	CB-CG-CD2	-5.23	117.86	121.00
1	A	84	TYR	CD1-CE1-CZ	-5.08	115.23	119.80

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	2241	0	2084	19	0
2	B	837	0	803	7	0
3	C	69	0	80	2	0
4	A	197	0	0	0	0
4	B	39	0	0	0	0
4	C	7	0	0	1	0
All	All	3390	0	2967	25	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 4.

All (25) 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:45:GLU:HG3	1:A:67:PHE:HE2	1.34	0.92
1:A:44:ARG:NH2	1:A:46:GLU:OE2	2.21	0.73
1:A:49:ALA:O	1:A:52:ILE:HG22	1.89	0.73
1:A:45:GLU:OE1	1:A:60:TRP:CE3	2.43	0.70
2:B:4:THR:HG22	2:B:5:PRO:HD2	1.80	0.61
2:B:4:THR:HG23	2:B:86:THR:OG1	2.00	0.61
1:A:45:GLU:HG3	1:A:67:PHE:CE2	2.26	0.60
1:A:42:SER:O	1:A:44:ARG:HG3	2.01	0.60
1:A:47:PRO:HB3	1:A:60:TRP:CH2	2.38	0.59
1:A:230:LEU:HD22	1:A:243:LYS:HE3	1.89	0.54
1:A:40:ALA:O	1:A:43:PRO:HD3	2.08	0.52
1:A:171:TYR:OH	3:C:2:ALA:HB2	2.10	0.51
1:A:34:VAL:CG2	1:A:45:GLU:HG2	2.40	0.51
1:A:54:GLN:HE22	1:A:174:ASN:HB3	1.77	0.49
3:C:5:LYS:NZ	4:C:101:HOH:O	2.37	0.48
2:B:4:THR:CG2	2:B:5:PRO:HD2	2.45	0.47
1:A:47:PRO:HB3	1:A:60:TRP:HH2	1.79	0.45
1:A:6:ARG:HE	2:B:58:LYS:NZ	2.15	0.44
1:A:238:ASP:OD1	1:A:238:ASP:C	2.56	0.44
2:B:25:CYS:HB2	2:B:39:LEU:HD21	1.99	0.44
1:A:97:SER:HB3	1:A:116:TYR:CD2	2.54	0.42
1:A:117:ALA:HB2	2:B:60:TRP:CE2	2.55	0.42
1:A:183:ASP:OD2	1:A:239:ARG:NH1	2.53	0.41
2:B:38:ASP:OD1	2:B:45:ARG:HD2	2.20	0.41
1:A:32:GLN:OE1	1:A:48:ARG:HG3	2.21	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	273/276 (99%)	267 (98%)	6 (2%)	0	100	100
2	B	98/100 (98%)	97 (99%)	1 (1%)	0	100	100
3	C	7/11 (64%)	7 (100%)	0	0	100	100
All	All	378/387 (98%)	371 (98%)	7 (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	A	234/235 (100%)	230 (98%)	4 (2%)	60	39
2	B	95/95 (100%)	94 (99%)	1 (1%)	73	57
3	C	6/7 (86%)	6 (100%)	0	100	100
All	All	335/337 (99%)	330 (98%)	5 (2%)	65	44

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	6	ARG
1	A	46	GLU
1	A	62	ARG
1	A	239	ARG
2	B	70	PHE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	54	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](#)

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

### 6.1 Protein, DNA and RNA chains

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	A	275/276 (99%)	0.41	22 (8%) <b>12</b>   <b>12</b>	19, 29, 54, 67	0
2	B	100/100 (100%)	0.83	15 (15%) <b>2</b>   <b>1</b>	21, 39, 58, 71	0
3	C	9/11 (81%)	0.72	1 (11%) <b>5</b>   <b>4</b>	26, 30, 42, 60	0
All	All	384/387 (99%)	0.53	38 (9%) <b>7</b>   <b>6</b>	19, 31, 56, 71	0

All (38) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	1	ILE	6.5
2	B	0	MET	6.4
1	A	43	PRO	6.2
3	C	1	ALA	5.2
1	A	41	ALA	4.7
1	A	45	GLU	4.1
2	B	48	LYS	4.1
1	A	60	TRP	3.9
1	A	53	GLU	3.9
1	A	44	ARG	3.5
1	A	42	SER	3.3
2	B	44	GLU	3.1
2	B	74	GLU	3.0
1	A	57	PRO	3.0
1	A	48	ARG	3.0
2	B	88	SER	2.9
1	A	217	TRP	2.9
1	A	33	PHE	2.8
1	A	223	ASP	2.8
1	A	54	GLN	2.8
2	B	16	GLU	2.7
1	A	49	ALA	2.7
1	A	261	VAL	2.5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
2	B	75	LYS	2.5
1	A	259	CYS	2.4
2	B	78	TYR	2.4
2	B	23	LEU	2.3
1	A	58	GLU	2.3
1	A	1	GLY	2.3
2	B	89	GLN	2.2
2	B	9	VAL	2.1
2	B	99	MET	2.1
2	B	98	ASP	2.1
1	A	62	ARG	2.0
1	A	222	GLU	2.0
1	A	215	LEU	2.0
2	B	30	PHE	2.0
1	A	193	PRO	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 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.