



# Full wwPDB X-ray Structure Validation Report ⓘ

May 14, 2020 – 04:00 pm BST

PDB ID : 4PGB  
Title : MHC Class I in complex with modified Sendai virus nucleoprotein peptide  
FAPGNWPAL  
Authors : Celie, P.H.N.; Joosten, R.P.; Perrakis, A.; Neefjes, J.  
Deposited on : 2014-05-01  
Resolution : 2.80 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

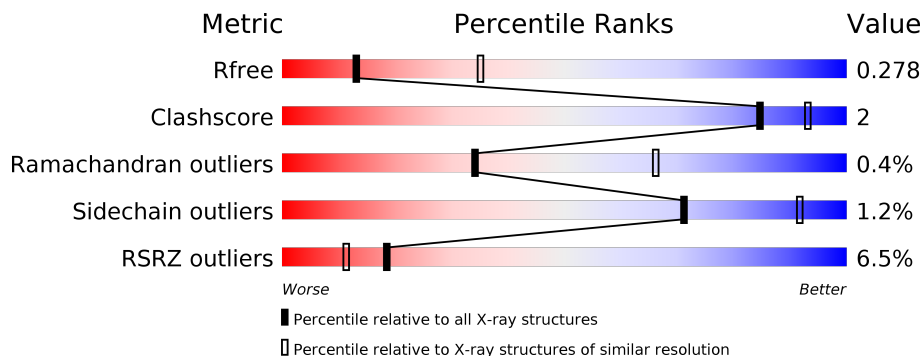
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.80 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	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	A	304	 6% 84% 8% 9%
1	D	304	 7% 84% 7% 9%
2	B	100	 2% 94% 5%
2	E	100	 10% 94% 5%
3	C	9	 11% 89% 11%
3	F	9	 78% 22%

## 2 Entry composition i

There are 3 unique types of molecules in this entry. The entry contains 6328 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 H-2 class I histocompatibility antigen, K-B alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	278	2262	1428	397	428	9	0	0	0
1	D	278	2262	1428	397	428	9	0	0	0

There are 52 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-25	MET	-	initiating methionine	UNP P01901
A	-24	GLY	-	expression tag	UNP P01901
A	-23	SER	-	expression tag	UNP P01901
A	-22	SER	-	expression tag	UNP P01901
A	-21	HIS	-	expression tag	UNP P01901
A	-20	HIS	-	expression tag	UNP P01901
A	-19	HIS	-	expression tag	UNP P01901
A	-18	HIS	-	expression tag	UNP P01901
A	-17	HIS	-	expression tag	UNP P01901
A	-16	HIS	-	expression tag	UNP P01901
A	-15	SER	-	expression tag	UNP P01901
A	-14	SER	-	expression tag	UNP P01901
A	-13	GLY	-	expression tag	UNP P01901
A	-12	LEU	-	expression tag	UNP P01901
A	-11	VAL	-	expression tag	UNP P01901
A	-10	PRO	-	expression tag	UNP P01901
A	-9	ARG	-	expression tag	UNP P01901
A	-8	GLY	-	expression tag	UNP P01901
A	-7	SER	-	expression tag	UNP P01901
A	-6	HIS	-	expression tag	UNP P01901
A	-5	MET	-	expression tag	UNP P01901
A	-4	LEU	-	expression tag	UNP P01901
A	-3	GLU	-	expression tag	UNP P01901
A	-2	ASP	-	expression tag	UNP P01901
A	-1	PRO	-	expression tag	UNP P01901

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	expression tag	UNP P01901
D	-25	MET	-	initiating methionine	UNP P01901
D	-24	GLY	-	expression tag	UNP P01901
D	-23	SER	-	expression tag	UNP P01901
D	-22	SER	-	expression tag	UNP P01901
D	-21	HIS	-	expression tag	UNP P01901
D	-20	HIS	-	expression tag	UNP P01901
D	-19	HIS	-	expression tag	UNP P01901
D	-18	HIS	-	expression tag	UNP P01901
D	-17	HIS	-	expression tag	UNP P01901
D	-16	HIS	-	expression tag	UNP P01901
D	-15	SER	-	expression tag	UNP P01901
D	-14	SER	-	expression tag	UNP P01901
D	-13	GLY	-	expression tag	UNP P01901
D	-12	LEU	-	expression tag	UNP P01901
D	-11	VAL	-	expression tag	UNP P01901
D	-10	PRO	-	expression tag	UNP P01901
D	-9	ARG	-	expression tag	UNP P01901
D	-8	GLY	-	expression tag	UNP P01901
D	-7	SER	-	expression tag	UNP P01901
D	-6	HIS	-	expression tag	UNP P01901
D	-5	MET	-	expression tag	UNP P01901
D	-4	LEU	-	expression tag	UNP P01901
D	-3	GLU	-	expression tag	UNP P01901
D	-2	ASP	-	expression tag	UNP P01901
D	-1	PRO	-	expression tag	UNP P01901
D	0	MET	-	expression tag	UNP P01901

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	99	821	524	138	152	7	0	0	0
2	E	99	821	524	138	152	7	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP P01887
B	85	ASP	ALA	conflict	UNP P01887
E	0	MET	-	initiating methionine	UNP P01887

*Continued on next page...*

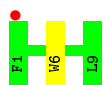
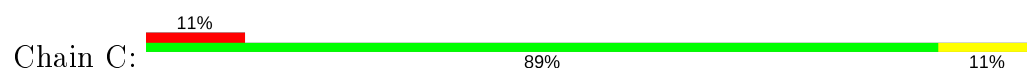
*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
E	85	ASP	ALA	conflict	UNP P01887

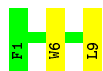
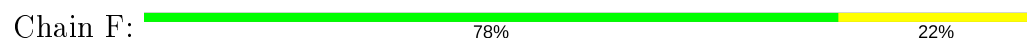
- Molecule 3 is a protein called Sendai virus nucleoprotein.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	9	Total	C	N	O	0	1	0
			81	58	12	11			
3	F	9	Total	C	N	O	0	1	0
			81	58	12	11			





- Molecule 3: Sendai virus nucleoprotein



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	65.55Å 89.26Å 88.46Å 90.00° 110.67° 90.00°	Depositor
Resolution (Å)	44.63 – 2.80 44.63 – 2.80	Depositor EDS
% Data completeness (in resolution range)	97.8 (44.63-2.80) 97.8 (44.63-2.80)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.07	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.48 (at 2.81Å)	Xtrriage
Refinement program	REFMAC 5.7.0032	Depositor
R, $R_{free}$	0.258 , 0.281 0.258 , 0.278	Depositor DCC
$R_{free}$ test set	1158 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	44.3	Xtrriage
Anisotropy	0.670	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.41 , 35.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtrriage
Estimated twinning fraction	0.000 for h,-k,-h-l	Xtrriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	6328	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 41.93 % 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 2.1982e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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	0.27	0/2326	0.48	0/3161
1	D	0.26	0/2326	0.46	0/3161
2	B	0.26	0/847	0.44	0/1148
2	E	0.26	0/847	0.44	0/1148
3	C	0.28	0/90	0.33	0/125
3	F	0.28	0/90	0.35	0/125
All	All	0.26	0/6526	0.46	0/8868

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	D	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	277	PRO	Peptide

### 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	2262	0	2150	14	0
1	D	2262	0	2150	12	0
2	B	821	0	796	3	0
2	E	821	0	796	3	0
3	C	81	0	75	0	0
3	F	81	0	75	2	0
All	All	6328	0	6042	28	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.

All (28) 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:181:ARG:O	1:A:182:THR:HG23	2.02	0.58
1:A:181:ARG:O	1:A:181:ARG:CG	2.54	0.55
1:A:28:VAL:HG23	1:A:33:PHE:CD1	2.43	0.54
1:D:28:VAL:HG23	1:D:33:PHE:CD1	2.43	0.53
1:A:181:ARG:O	1:A:181:ARG:HG3	2.09	0.53
1:D:147:TRP:CZ2	3:F:9:LEU:HD23	2.46	0.50
1:A:262:TYR:CD2	1:D:108:ARG:HD3	2.47	0.49
1:A:277:PRO:HB2	1:A:278:PRO:C	2.33	0.48
1:A:121:CYS:SG	2:B:1:ILE:HD12	2.54	0.47
1:A:81:LEU:HD12	1:A:95:ILE:HD11	1.97	0.47
1:D:121:CYS:SG	2:E:1:ILE:HD12	2.55	0.47
1:D:81:LEU:HD12	1:D:95:ILE:HD11	1.97	0.47
2:B:79:ALA:HB2	2:B:94:TYR:CD2	2.52	0.44
1:A:169:ARG:HD3	1:D:269:PRO:HD2	1.98	0.44
1:D:209:TYR:CD2	1:D:210:PRO:HA	2.53	0.43
2:E:79:ALA:HB2	2:E:94:TYR:CD2	2.52	0.43
1:A:19:GLU:OE1	1:A:75:ARG:NH2	2.48	0.43
1:D:194:ARG:O	1:D:196:GLU:N	2.52	0.43
1:A:224:LEU:O	1:A:228:MET:HB2	2.20	0.42
1:D:54:GLN:HA	1:D:54:GLN:HE21	1.85	0.42
1:D:224:LEU:O	1:D:228:MET:HB2	2.19	0.42
1:A:194:ARG:O	1:A:196:GLU:N	2.52	0.42
1:D:147:TRP:HZ2	3:F:9:LEU:HD23	1.85	0.41
2:B:7:ILE:HB	2:B:93:VAL:HG21	2.02	0.41
1:A:44:ARG:HA	1:A:64:THR:HG23	2.03	0.41
2:E:7:ILE:HB	2:E:93:VAL:HG21	2.02	0.41
1:D:44:ARG:HA	1:D:64:THR:HG23	2.03	0.41
1:A:209:TYR:CD1	1:A:210:PRO:HA	2.57	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	A	276/304 (91%)	266 (96%)	8 (3%)	2 (1%)	22	53
1	D	276/304 (91%)	267 (97%)	8 (3%)	1 (0%)	34	66
2	B	97/100 (97%)	94 (97%)	3 (3%)	0	100	100
2	E	97/100 (97%)	94 (97%)	3 (3%)	0	100	100
3	C	8/9 (89%)	7 (88%)	1 (12%)	0	100	100
3	F	8/9 (89%)	8 (100%)	0	0	100	100
All	All	762/826 (92%)	736 (97%)	23 (3%)	3 (0%)	34	66

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	181	ARG
1	A	195	PRO
1	D	195	PRO

### 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	236/259 (91%)	234 (99%)	2 (1%)	81	94
1	D	236/259 (91%)	232 (98%)	4 (2%)	60	87

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	94/95 (99%)	94 (100%)	0	100	100
2	E	94/95 (99%)	94 (100%)	0	100	100
3	C	7/6 (117%)	5 (71%)	2 (29%)	0	1
3	F	7/6 (117%)	5 (71%)	2 (29%)	0	1
All	All	674/720 (94%)	664 (98%)	10 (2%)	71	89

All (10) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	45	TYR
1	A	227	ASP
3	C	6[A]	TRP
3	C	6[B]	TRP
1	D	45	TYR
1	D	54	GLN
1	D	182	THR
1	D	227	ASP
3	F	6[A]	TRP
3	F	6[B]	TRP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	2	GLN
1	D	54	GLN
2	E	2	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 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

### 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	278/304 (91%)	0.61	17 (6%) 21 13	37, 48, 67, 74	0
1	D	278/304 (91%)	0.64	20 (7%) 15 8	40, 51, 61, 69	0
2	B	99/100 (99%)	0.36	2 (2%) 65 56	33, 46, 61, 73	0
2	E	99/100 (99%)	0.74	10 (10%) 7 4	40, 54, 70, 77	0
3	C	9/9 (100%)	1.01	1 (11%) 5 3	78, 90, 100, 123	0
3	F	9/9 (100%)	0.46	0 100 100	55, 62, 70, 89	0
All	All	772/826 (93%)	0.61	50 (6%) 18 11	33, 50, 67, 123	0

All (50) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	278	PRO	6.8
1	A	195	PRO	6.3
1	A	277	PRO	5.5
1	D	193	SER	4.9
1	A	1	GLY	3.9
1	D	240	THR	3.9
1	A	61	GLU	3.6
1	D	180	LEU	3.3
1	D	195	PRO	3.2
1	D	176	ASN	3.1
2	B	1	ILE	3.0
1	A	193	SER	2.8
1	A	180	LEU	2.8
1	D	224	LEU	2.8
1	D	54	GLN	2.7
1	D	226	GLN	2.7
1	A	2	PRO	2.6
2	E	69	GLU	2.6
1	D	2	PRO	2.6

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	RSRZ
1	D	211	ALA	2.6
1	A	128	GLU	2.5
1	D	209	TYR	2.5
1	D	225	ILE	2.5
1	A	196	GLU	2.4
2	E	1	ILE	2.4
2	E	75	THR	2.4
1	A	194	ARG	2.4
2	E	18	GLY	2.4
2	E	16	GLU	2.3
2	E	21	ASN	2.3
1	A	58	GLU	2.2
1	D	173	LYS	2.2
1	D	194	ARG	2.2
2	E	23	LEU	2.2
1	D	183	ASP	2.1
1	D	241	PHE	2.1
2	E	19	LYS	2.1
2	E	70	PHE	2.1
2	E	68	THR	2.1
1	A	211	ALA	2.1
1	A	176	ASN	2.1
1	D	128	GLU	2.1
1	A	200	THR	2.1
1	D	222	GLU	2.1
2	B	16	GLU	2.1
1	A	106	ASP	2.0
1	D	278	PRO	2.0
3	C	1	PHE	2.0
1	A	226	GLN	2.0
1	D	110	LEU	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

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.