



# Full wwPDB X-ray Structure Validation Report ⓘ

Feb 13, 2017 – 04:26 am GMT

PDB ID : 2XFX  
Title : cattle MHC class I N01301 presenting an 11mer from Theileria parva  
Authors : Macdonald, I.K.; Harkiolaki, M.; Hunt, L.; Morrison, W.I.; Connelley, T.;  
Graham, S.P.; Jones, E.Y.; Flower, D.R.; Ellis, S.A.  
Deposited on : 2010-05-28  
Resolution : 1.90 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

<http://wwpdb.org/validation/2016/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)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

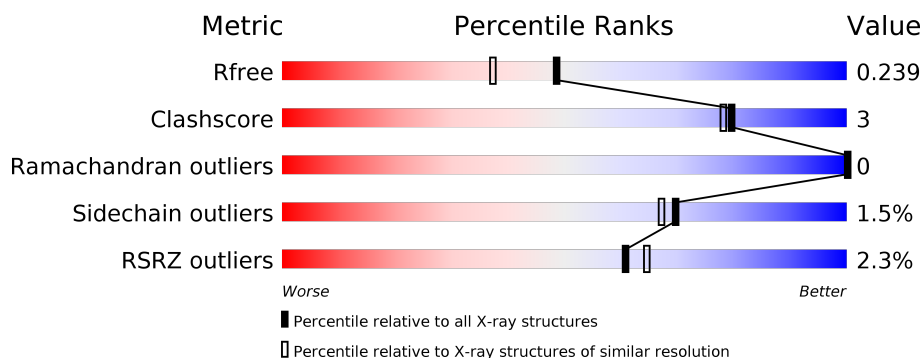
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.90 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))
$R_{free}$	100719	5047 (1.90-1.90)
Clashscore	112137	5731 (1.90-1.90)
Ramachandran outliers	110173	5669 (1.90-1.90)
Sidechain outliers	110143	5670 (1.90-1.90)
RSRZ outliers	101464	5100 (1.90-1.90)

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	A	277	<div> <div>2%</div> <div> <div></div> <div>87%</div> <div>13%</div> </div> </div>
2	B	99	<div> <div>4%</div> <div> <div></div> <div>92%</div> <div>8%</div> </div> </div>
3	C	11	<div> <div></div> <div> <div></div> <div>82%</div> <div>18%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3341 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 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	277	Total	C	N	O	S	0	0	0
			2273	1412	414	439	8			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	277	GLY	-	EXPRESSION TAG	UNP Q30291

- Molecule 2 is a protein called BETA-2-MICROGLOBULIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	99	Total	C	N	O	S	0	0	0
			829	531	143	153	2			

- Molecule 3 is a protein called UNCHARACTERIZED PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	11	Total	C	N	O	S	0	0	0
			90	59	13	17	1			

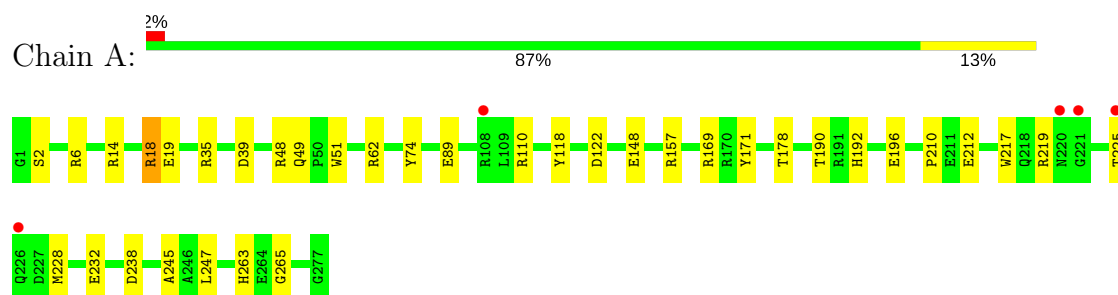
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	100	Total	O	0	0
			100	100		
4	B	39	Total	O	0	0
			39	39		
4	C	10	Total	O	0	0
			10	10		

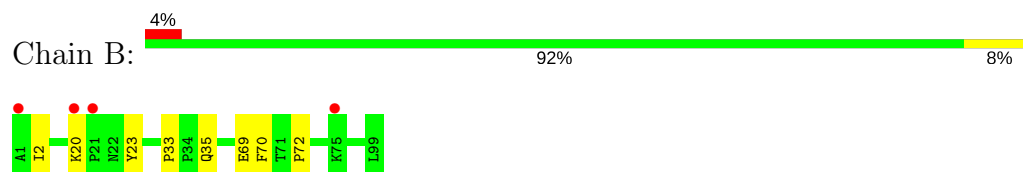
### 3 Residue-property plots

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.

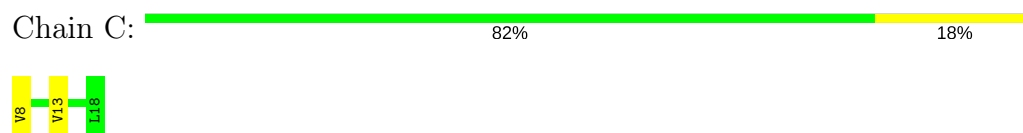
#### • Molecule 1: MHC CLASS 1



#### • Molecule 2: BETA-2-MICROGLOBULIN



#### • Molecule 3: UNCHARACTERIZED PROTEIN



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	44.61Å 97.89Å 123.72Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	45.51 – 1.90 45.51 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.9 (45.51-1.90) 99.4 (45.51-1.75)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.01 (at 1.75Å)	Xtriage
Refinement program	REFMAC 5.6.0070	Depositor
R, $R_{free}$	0.198 , 0.238 0.204 , 0.239	Depositor DCC
$R_{free}$ test set	2208 reflections (5.34%)	DCC
Wilson B-factor (Å <sup>2</sup> )	25.7	Xtriage
Anisotropy	0.247	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 40.6	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	3341	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	33.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.21% 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

### 5.1 Standard geometry

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	1.18	7/2333 (0.3%)	1.04	12/3163 (0.4%)
2	B	1.04	0/855	0.88	0/1158
3	C	1.47	1/91 (1.1%)	0.99	0/119
All	All	1.15	8/3279 (0.2%)	1.00	12/4440 (0.3%)

All (8) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	232	GLU	CD-OE1	7.66	1.34	1.25
1	A	74	TYR	CE1-CZ	-7.17	1.29	1.38
1	A	232	GLU	CG-CD	6.72	1.62	1.51
1	A	89	GLU	CG-CD	6.64	1.61	1.51
1	A	118	TYR	CD1-CE1	6.13	1.48	1.39
3	C	13	VAL	CB-CG1	5.45	1.64	1.52
1	A	118	TYR	CG-CD1	-5.39	1.32	1.39
1	A	148	GLU	CD-OE2	-5.24	1.19	1.25

All (12) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	157	ARG	NE-CZ-NH2	-9.11	115.75	120.30
1	A	157	ARG	NE-CZ-NH1	7.91	124.25	120.30
1	A	219	ARG	NE-CZ-NH1	7.34	123.97	120.30
1	A	62	ARG	NE-CZ-NH2	-6.78	116.91	120.30
1	A	122	ASP	CB-CG-OD1	6.66	124.29	118.30
1	A	62	ARG	NE-CZ-NH1	5.66	123.13	120.30
1	A	18	ARG	NE-CZ-NH1	5.64	123.12	120.30
1	A	238	ASP	CB-CG-OD1	5.36	123.13	118.30
1	A	169	ARG	NE-CZ-NH1	5.10	122.85	120.30
1	A	6	ARG	NE-CZ-NH2	-5.04	117.78	120.30
1	A	74	TYR	CB-CG-CD1	-5.03	117.98	121.00
1	A	14	ARG	NE-CZ-NH1	-5.01	117.79	120.30

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	2273	0	2123	17	0
2	B	829	0	806	3	0
3	C	90	0	94	1	0
4	A	100	0	0	0	0
4	B	39	0	0	0	0
4	C	10	0	0	0	0
All	All	3341	0	3023	20	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 (20) 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:228:MET:HE3	1:A:245:ALA:HB1	1.61	0.83
1:A:228:MET:CE	1:A:245:ALA:HB1	2.11	0.80
1:A:171:TYR:HH	3:C:8:VAL:N	1.82	0.77
1:A:263:HIS:CD2	1:A:265:GLY:H	2.10	0.69
1:A:225:THR:HA	1:A:228:MET:HG2	1.75	0.68
2:B:23:TYR:CE1	2:B:69:GLU:HG2	2.32	0.65
2:B:20:LYS:O	2:B:72:PRO:HD2	2.00	0.61
1:A:18:ARG:HG2	1:A:19:GLU:HG2	1.83	0.61
1:A:263:HIS:HD2	1:A:265:GLY:H	1.54	0.53
1:A:49:GLN:HE21	1:A:51:TRP:HE1	1.56	0.52
1:A:35:ARG:HE	1:A:48:ARG:NH1	2.09	0.51
1:A:49:GLN:NE2	1:A:178:THR:HG21	2.25	0.50
1:A:228:MET:HE1	1:A:245:ALA:HB1	1.90	0.48
1:A:210:PRO:O	1:A:263:HIS:HE1	1.96	0.48
1:A:217:TRP:CG	1:A:247:LEU:HD22	2.52	0.44
1:A:190:THR:OG1	1:A:192:HIS:HE1	2.01	0.44
1:A:2:SER:OG	1:A:110:ARG:NH2	2.52	0.43

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:228:MET:CE	1:A:245:ALA:CB	2.92	0.42
2:B:33:PRO:HB2	2:B:35:GLN:OE1	2.21	0.41
1:A:217:TRP:CD2	1:A:247:LEU:HD22	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	275/277 (99%)	268 (98%)	7 (2%)	0	100	100
2	B	97/99 (98%)	96 (99%)	1 (1%)	0	100	100
3	C	9/11 (82%)	9 (100%)	0	0	100	100
All	All	381/387 (98%)	373 (98%)	8 (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	236/236 (100%)	233 (99%)	3 (1%)	73	72
2	B	94/94 (100%)	92 (98%)	2 (2%)	59	53
3	C	10/10 (100%)	10 (100%)	0	100	100

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
All	All	340/340 (100%)	335 (98%)	5 (2%)	70	67

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

Mol	Chain	Res	Type
1	A	39	ASP
1	A	196	GLU
1	A	212	GLU
2	B	2	ILE
2	B	70	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	9	HIS
1	A	32	GLN
1	A	49	GLN
1	A	96	GLN
1	A	174	ASN
1	A	192	HIS
1	A	218	GLN
1	A	262	GLN
1	A	263	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 carbohydrates in this entry.

## 5.6 Ligand geometry

There are no ligands in this entry.

## 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	A	277/277 (100%)	-0.23	5 (1%) 69 72	14, 30, 59, 90	0
2	B	99/99 (100%)	-0.24	4 (4%) 39 43	17, 33, 57, 66	0
3	C	11/11 (100%)	-0.47	0 100 100	21, 23, 28, 36	0
All	All	387/387 (100%)	-0.24	9 (2%) 61 64	14, 31, 58, 90	0

All (9) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	225	THR	5.4
1	A	221	GLY	3.5
2	B	1	ALA	3.2
1	A	108	ARG	2.6
1	A	220	ASN	2.6
1	A	226	GLN	2.6
2	B	75	LYS	2.2
2	B	21	PRO	2.1
2	B	20	LYS	2.1

### 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.