Full wwPDB X-ray Structure Validation Report (i)

Oct 31, 2023 – 07:37 AM EDT

PDB ID	:	3MRI
Title	:	Crystal Structure of MHC class I HLA-A2 molecule complexed with HCV NS3-1073-1081 nonapeptide G4M-V5W variant
Authors	:	Reiser, J.-B.; Le Gorrec, M.; Chouquet, A.; Debeaupuis, E.; Echasserieau, K.; Saulquin, X.; Bonneville, M.; Housset, D.
Deposited on	:	2010-04-29
Resolution	:	2.10 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

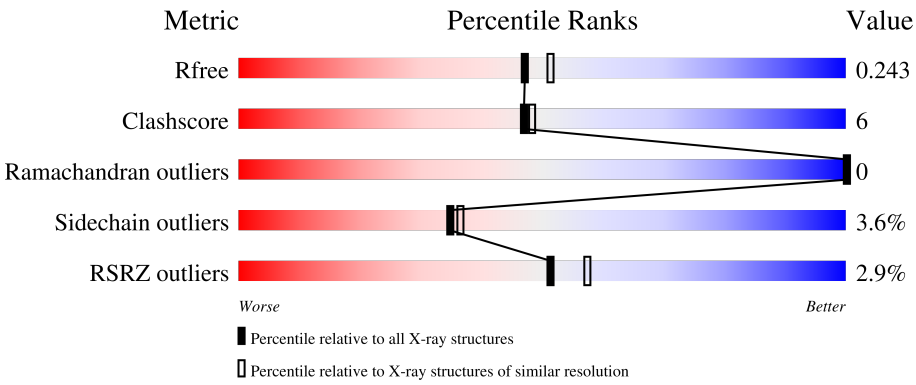
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.10 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	5197 (2.10-2.10)
Clashscore	141614	5710 (2.10-2.10)
Ramachandran outliers	138981	5647 (2.10-2.10)
Sidechain outliers	138945	5648 (2.10-2.10)
RSRZ outliers	127900	5083 (2.10-2.10)

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 >=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 <=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	293	% 80% 12% • 6%
2	B	100	5% 84% 15% •
3	P	9	22% 56% 33% 11%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 3197 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
			Total	C	N	O	S			
1	A	274	2248	1406	408	425	9	0	2	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	245	VAL	ALA	ENGINEERED MUTATION	UNP P01892
A	277	GLY	-	EXPRESSION TAG	UNP P01892
A	278	SER	-	EXPRESSION TAG	UNP P01892
A	279	LEU	-	EXPRESSION TAG	UNP P01892
A	280	HIS	-	EXPRESSION TAG	UNP P01892
A	281	HIS	-	EXPRESSION TAG	UNP P01892
A	282	ILE	-	EXPRESSION TAG	UNP P01892
A	283	LEU	-	EXPRESSION TAG	UNP P01892
A	284	ASP	-	EXPRESSION TAG	UNP P01892
A	285	ALA	-	EXPRESSION TAG	UNP P01892
A	286	GLN	-	EXPRESSION TAG	UNP P01892
A	287	LYS	-	EXPRESSION TAG	UNP P01892
A	288	MET	-	EXPRESSION TAG	UNP P01892
A	289	VAL	-	EXPRESSION TAG	UNP P01892
A	290	TRP	-	EXPRESSION TAG	UNP P01892
A	291	ASN	-	EXPRESSION TAG	UNP P01892
A	292	HIS	-	EXPRESSION TAG	UNP P01892
A	293	ARG	-	EXPRESSION TAG	UNP P01892

- 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	-	EXPRESSION TAG	UNP P61769

- Molecule 3 is a protein called 9-meric peptide from Serine protease/NTPase/helicase NS3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	P	9	84	56	12	12	4	0	1	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	4	MET	GLY	ENGINEERED MUTATION	UNP Q03463
P	5	TRP	VAL	ENGINEERED MUTATION	UNP Q03463

- Molecule 4 is water.

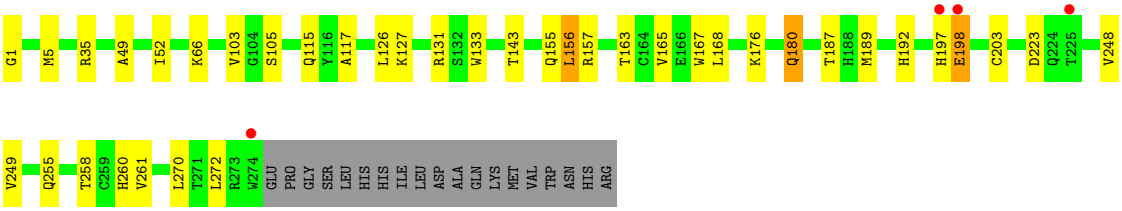
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	22	22	22	0	0
4	B	5	5	5	0	0
4	P	1	1	1	0	0

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.

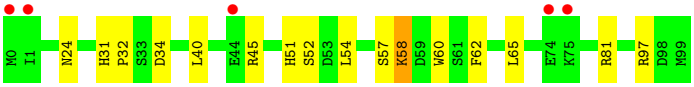
• Molecule 1: HLA class I histocompatibility antigen, A-2 alpha chain

Chain A:



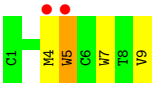
• Molecule 2: Beta-2-microglobulin

Chain B:



• Molecule 3: 9-meric peptide from Serine protease/NTPase/helicase NS3

Chain P:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	52.41Å 80.51Å 55.90Å 90.00° 112.44° 90.00°	Depositor
Resolution (Å)	15.00 – 2.10 41.51 – 2.10	Depositor EDS
% Data completeness (in resolution range)	100.0 (15.00-2.10) 99.7 (41.51-2.10)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.56 (at 2.10Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.239 , 0.296 0.241 , 0.243	Depositor DCC
R_{free} test set	2506 reflections (10.00%)	wwPDB-VP
Wilson B-factor (Å^2)	24.8	Xtriage
Anisotropy	0.773	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 41.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	3197	wwPDB-VP
Average B, all atoms (Å^2)	39.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.13% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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	0.61	1/2319 (0.0%)	0.68	0/3148
2	B	0.56	0/860	0.64	0/1162
3	P	0.51	0/90	0.79	0/121
All	All	0.60	1/3269 (0.0%)	0.67	0/4431

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	203	CYS	CB-SG	-5.01	1.73	1.81

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	A	2248	0	2105	30	0
2	B	837	0	803	9	0
3	P	84	0	83	5	0
4	A	22	0	0	2	0
4	B	5	0	0	0	0
4	P	1	0	0	0	0
All	All	3197	0	2991	38	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 6.

All (38) 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:198:GLU:OE1	1:A:248:VAL:HG13	1.28	1.28
1:A:198:GLU:OE1	1:A:248:VAL:CG1	2.14	0.95
1:A:197:HIS:CE1	1:A:198:GLU:HB2	2.08	0.88
1:A:258:THR:OG1	1:A:260:HIS:HE1	1.65	0.80
1:A:133:TRP:HA	4:A:314:HOH:O	1.83	0.78
1:A:197:HIS:ND1	1:A:198:GLU:HB2	2.01	0.75
3:P:5:TRP:HZ3	3:P:7:TRP:O	1.72	0.72
1:A:198:GLU:CD	1:A:248:VAL:HG13	2.11	0.69
1:A:258:THR:OG1	1:A:260:HIS:CE1	2.47	0.68
1:A:115:GLN:HG3	2:B:60:TRP:CH2	2.32	0.65
1:A:127:LYS:N	4:A:314:HOH:O	2.31	0.64
2:B:51:HIS:HD2	2:B:52:SER:O	1.81	0.62
1:A:1:GLY:N	1:A:105:SER:OG	2.33	0.61
1:A:155:GLN:NE2	3:P:7:TRP:HZ2	2.00	0.59
1:A:117:ALA:HB2	2:B:60:TRP:CE2	2.41	0.56
1:A:66:LYS:HG2	3:P:4[B]:MET:HG2	1.88	0.56
1:A:126:LEU:HD22	1:A:156:LEU:HD13	1.89	0.54
3:P:5:TRP:CZ3	3:P:7:TRP:O	2.58	0.54
2:B:24:ASN:HB3	2:B:65:LEU:HD11	1.93	0.51
1:A:103:VAL:CG1	1:A:165:VAL:HG13	2.41	0.50
2:B:31:HIS:ND1	2:B:32:PRO:HA	2.27	0.49
1:A:189:MET:HE3	1:A:272:LEU:HB3	1.94	0.48
1:A:143:THR:HG23	3:P:9:VAL:HA	1.95	0.47
1:A:189:MET:HE3	1:A:272:LEU:CB	2.45	0.46
1:A:198:GLU:HG2	1:A:249:VAL:O	2.16	0.45
1:A:189:MET:CE	1:A:272:LEU:HB2	2.47	0.45
2:B:54:LEU:HD11	2:B:62:PHE:CD1	2.53	0.44
1:A:131:ARG:NH2	1:A:157:ARG:HH12	2.15	0.43
2:B:40:LEU:HD11	2:B:81:ARG:HB2	2.01	0.42
1:A:49:ALA:O	1:A:52:ILE:HG22	2.20	0.41
2:B:40:LEU:HD23	2:B:45:ARG:HA	2.02	0.41
1:A:255:GLN:CD	1:A:255:GLN:H	2.23	0.41
1:A:187:THR:HG21	1:A:261:VAL:HG21	2.02	0.41
1:A:189:MET:CE	1:A:272:LEU:CB	2.98	0.41
1:A:5:MET:HB2	1:A:168:LEU:HD13	2.03	0.41
1:A:115:GLN:HE22	2:B:58:LYS:HD2	1.86	0.40
1:A:176:LYS:HA	1:A:180:GLN:HG3	2.04	0.40
1:A:163:THR:O	1:A:167:TRP:HD1	2.05	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	274/293 (94%)	262 (96%)	12 (4%)	0	100	100
2	B	98/100 (98%)	96 (98%)	2 (2%)	0	100	100
3	P	8/9 (89%)	8 (100%)	0	0	100	100
All	All	380/402 (94%)	366 (96%)	14 (4%)	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	233/248 (94%)	226 (97%)	7 (3%)	41	44
2	B	95/95 (100%)	91 (96%)	4 (4%)	30	30
3	P	10/9 (111%)	9 (90%)	1 (10%)	7	5
All	All	338/352 (96%)	326 (96%)	12 (4%)	35	36

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	35	ARG
1	A	156	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	180	GLN
1	A	192	HIS
1	A	198	GLU
1	A	223	ASP
1	A	270	LEU
2	B	34	ASP
2	B	57	SER
2	B	58	LYS
2	B	97	ARG
3	P	5	TRP

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	32	GLN
1	A	115	GLN
1	A	155	GLN
1	A	260	HIS
2	B	51	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 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 (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, 95th 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^2)	Q<0.9
1	A	274/293 (93%)	0.25	4 (1%)	73	77	20, 37, 50, 56	0
2	B	100/100 (100%)	0.39	5 (5%)	28	34	26, 43, 56, 61	0
3	P	9/9 (100%)	1.02	2 (22%)	0	0	39, 44, 46, 48	0
All	All	383/402 (95%)	0.30	11 (2%)	51	57	20, 38, 53, 61	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	1	ILE	4.0
2	B	0	MET	3.3
3	P	4[A]	MET	2.6
1	A	198	GLU	2.6
1	A	225	THR	2.3
2	B	75	LYS	2.2
3	P	5	TRP	2.2
1	A	197	HIS	2.2
2	B	74	GLU	2.1
2	B	44	GLU	2.1
1	A	274	TRP	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 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.