Full wwPDB X-ray Structure Validation Report ⓘ

May 17, 2020 – 07:34 am BST

PDB ID	:	6G3J
Title	:	MHC A02 Allele presenting MTSAIGILVP
Authors	:	Rizkallah, P.J.; Sewell, A.K.
Deposited on	:	2018-03-26
Resolution	:	2.45 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 ⓘ symbol.

The following versions of software and data (see references ①) were used in the production of this report:

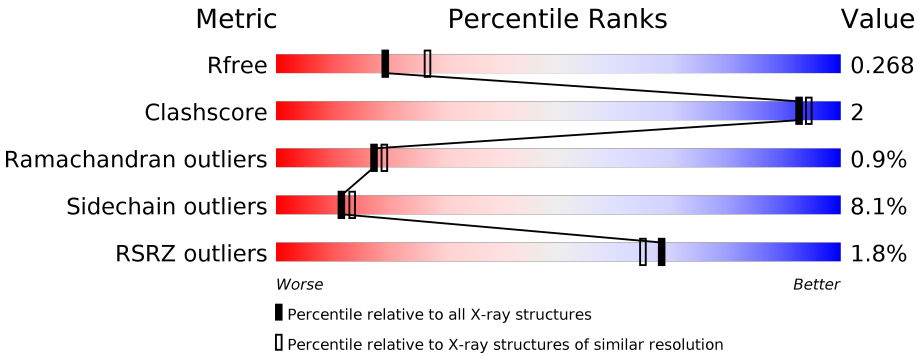
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.45 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	1544 (2.48-2.44)
Clashscore	141614	1613 (2.48-2.44)
Ramachandran outliers	138981	1598 (2.48-2.44)
Sidechain outliers	138945	1598 (2.48-2.44)
RSRZ outliers	127900	1523 (2.48-2.44)

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 >=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	276	% 87% 12%
1	D	276	4% 88% 11%
2	B	100	% 87% 12%
2	E	100	% 88% 11%
3	C	10	80% 20%
3	F	10	60% 30% 10%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6354 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
1	A	276	Total	C	N	O	S	0	1	0
			2264	1414	413	428	9			
1	D	276	Total	C	N	O	S	0	0	0
			2254	1408	410	427	9			

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

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

There are 2 discrepancies between the modelled and reference sequences:

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

- Molecule 3 is a protein called MET-THR-SER-ALA-ILE-GLY-ILE-LEU-PRO-VAL.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	10	Total	C	N	O	S	0	0	0
			69	45	10	13	1			
3	F	10	Total	C	N	O	S	0	0	0
			69	45	10	13	1			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	6	Total	O	0	0
			6	6		

Continued on next page...

Continued from previous page...

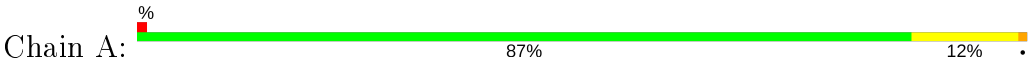
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	4	Total O 4 4	0	0
4	C	1	Total O 1 1	0	0
4	D	6	Total O 6 6	0	0
4	E	7	Total O 7 7	0	0

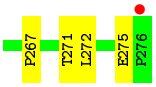
3 Residue-property plots (i)

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

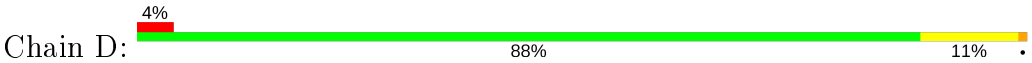
Chain A:

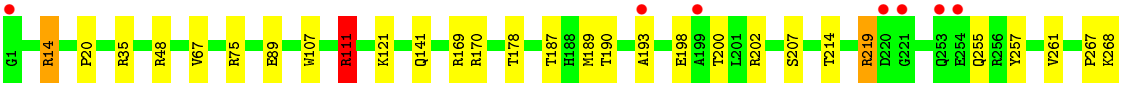


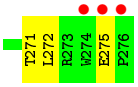


- Molecule 1: HLA class I histocompatibility antigen, A-2 alpha chain

Chain D:

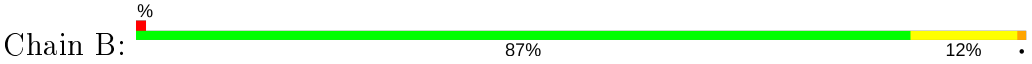


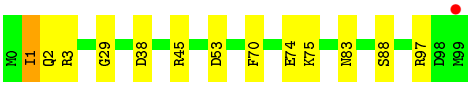




- Molecule 2: Beta-2-microglobulin

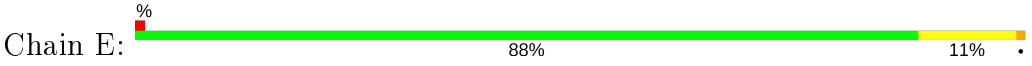
Chain B:

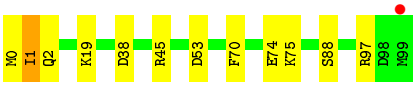




- Molecule 2: Beta-2-microglobulin

Chain E:





- Molecule 3: MET-THR-SER-ALA-ILE-GLY-ILE-LEU-PRO-VAL

Chain C:

- Molecule 3: MET-THR-SER-ALA-ILE-GLY-ILE-LEU-PRO-VAL

Chain F:

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	202.94Å 50.28Å 119.07Å 90.00° 122.86° 90.00°	Depositor
Resolution (Å)	41.05 – 2.45 41.05 – 2.45	Depositor EDS
% Data completeness (in resolution range)	98.3 (41.05-2.45) 98.3 (41.05-2.45)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.48 (at 2.45Å)	Xtriage
Refinement program	REFMAC 5.8.0189	Depositor
R, R_{free}	0.223 , 0.270 0.227 , 0.268	Depositor DCC
R_{free} test set	1818 reflections (4.91%)	wwPDB-VP
Wilson B-factor (Å^2)	62.8	Xtriage
Anisotropy	0.125	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 40.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6354	wwPDB-VP
Average B, all atoms (Å^2)	86.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.50% 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

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	0.88	0/2331	1.01	7/3164 (0.2%)
1	D	0.87	0/2320	0.98	4/3149 (0.1%)
2	B	1.00	1/860 (0.1%)	1.00	1/1162 (0.1%)
2	E	0.98	0/860	1.01	3/1162 (0.3%)
3	C	0.83	0/69	1.20	1/92 (1.1%)
3	F	1.00	0/69	1.20	1/92 (1.1%)
All	All	0.91	1/6509 (0.0%)	1.00	17/8821 (0.2%)

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	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	53	ASP	CB-CG	6.30	1.65	1.51

All (17) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	14	ARG	NE-CZ-NH1	-7.08	116.76	120.30
3	C	1	MET	CG-SD-CE	6.84	111.14	100.20
1	D	14	ARG	NE-CZ-NH2	6.79	123.70	120.30
1	D	111	ARG	NE-CZ-NH2	6.32	123.46	120.30
1	A	122	ASP	CB-CG-OD2	-6.31	112.62	118.30
2	E	45	ARG	NE-CZ-NH1	6.27	123.44	120.30
1	D	14	ARG	NE-CZ-NH1	-6.14	117.23	120.30
1	A	14	ARG	NE-CZ-NH2	6.00	123.30	120.30
3	F	6	GLY	N-CA-C	-5.89	98.39	113.10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	202	ARG	NE-CZ-NH2	-5.60	117.50	120.30
1	A	122	ASP	CB-CG-OD1	5.34	123.11	118.30
1	D	48	ARG	NE-CZ-NH2	-5.28	117.66	120.30
1	A	48	ARG	NE-CZ-NH2	-5.21	117.70	120.30
1	A	21	ARG	NE-CZ-NH1	5.20	122.90	120.30
2	E	53	ASP	CB-CG-OD1	5.11	122.90	118.30
2	B	45	ARG	CB-CA-C	-5.06	100.28	110.40
2	E	45	ARG	CB-CA-C	-5.01	100.38	110.40

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	14	ARG	Peptide
1	D	193	ALA	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	2264	0	2109	7	0
1	D	2254	0	2103	7	0
2	B	837	0	803	2	0
2	E	837	0	803	2	0
3	C	69	0	80	0	0
3	F	69	0	80	1	0
4	A	6	0	0	0	0
4	B	4	0	0	0	0
4	C	1	0	0	0	0
4	D	6	0	0	1	0
4	E	7	0	0	0	0
All	All	6354	0	5978	19	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 (19) 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:190:THR:OG1	1:A:202:ARG:HB3	2.09	0.53
1:D:190:THR:OG1	1:D:202:ARG:HB3	2.09	0.52
3:F:7:ILE:HG22	3:F:7:ILE:O	2.12	0.49
1:D:111:ARG:HA	1:D:111:ARG:CZ	2.43	0.48
1:A:74:HIS:HE1	1:A:97:ARG:HE	1.62	0.48
1:A:74:HIS:CE1	1:A:97:ARG:HE	2.32	0.47
1:A:187:THR:HG21	1:A:261:VAL:HG21	1.95	0.47
1:D:187:THR:HG21	1:D:261:VAL:HG21	1.98	0.46
1:A:219:ARG:HG2	1:A:257:TYR:CE2	2.51	0.45
1:D:107:TRP:HB3	1:D:169:ARG:HD3	1.99	0.45
1:A:107:TRP:HB3	1:A:169:ARG:HD3	2.00	0.43
2:B:1:ILE:HD13	2:B:1:ILE:O	2.19	0.42
2:E:1:ILE:HD13	2:E:1:ILE:O	2.20	0.42
4:D:305:HOH:O	2:E:0:MET:HG2	2.19	0.42
2:B:3:ARG:HG2	2:B:29:GLY:O	2.20	0.41
1:D:20:PRO:HG2	1:D:75:ARG:HG3	2.03	0.41
1:A:259:CYS:HB3	1:A:272:LEU:HB2	2.03	0.41
1:D:187:THR:HB	1:D:272:LEU:HD21	2.03	0.40
1:D:219:ARG:HG2	1:D:257:TYR:CE2	2.56	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	275/276 (100%)	259 (94%)	14 (5%)	2 (1%)	22 25
1	D	274/276 (99%)	257 (94%)	15 (6%)	2 (1%)	22 25
2	B	98/100 (98%)	95 (97%)	2 (2%)	1 (1%)	15 16
2	E	98/100 (98%)	95 (97%)	2 (2%)	1 (1%)	15 16
3	C	8/10 (80%)	7 (88%)	1 (12%)	0	100 100
3	F	8/10 (80%)	7 (88%)	0	1 (12%)	0 0

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	761/772 (99%)	720 (95%)	34 (4%)	7 (1%)	17 19

All (7) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	97	ARG
2	E	97	ARG
1	A	275	GLU
1	D	275	GLU
1	A	267	PRO
1	D	267	PRO
3	F	7	ILE

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	233/232 (100%)	215 (92%)	18 (8%)	13 15
1	D	232/232 (100%)	215 (93%)	17 (7%)	14 16
2	B	95/95 (100%)	87 (92%)	8 (8%)	11 12
2	E	95/95 (100%)	87 (92%)	8 (8%)	11 12
3	C	8/8 (100%)	7 (88%)	1 (12%)	4 3
3	F	8/8 (100%)	6 (75%)	2 (25%)	0 0
All	All	671/670 (100%)	617 (92%)	54 (8%)	11 14

All (54) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	34	VAL
1	A	35	ARG
1	A	67	VAL
1	A	86	ASN
1	A	111	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	121	LYS
1	A	141	GLN
1	A	176	LYS
1	A	178	THR
1	A	189	MET
1	A	200	THR
1	A	207	SER
1	A	214	THR
1	A	219	ARG
1	A	223	ASP
1	A	224	GLN
1	A	255	GLN
1	A	271	THR
2	B	1	ILE
2	B	2	GLN
2	B	38	ASP
2	B	70	PHE
2	B	74	GLU
2	B	75	LYS
2	B	83	ASN
2	B	88	SER
3	C	5	ILE
1	D	35	ARG
1	D	67	VAL
1	D	89	GLU
1	D	111	ARG
1	D	121	LYS
1	D	141	GLN
1	D	170	ARG
1	D	178	THR
1	D	189	MET
1	D	198	GLU
1	D	200	THR
1	D	207	SER
1	D	214	THR
1	D	219	ARG
1	D	255	GLN
1	D	268	LYS
1	D	271	THR
2	E	1	ILE
2	E	2	GLN
2	E	19	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	E	38	ASP
2	E	70	PHE
2	E	74	GLU
2	E	75	LYS
2	E	88	SER
3	F	1	MET
3	F	10	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (15) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	54	GLN
1	A	74	HIS
1	A	86	ASN
1	A	155	GLN
1	A	253	GLN
1	A	262	GLN
2	B	31	HIS
2	B	83	ASN
1	D	54	GLN
1	D	74	HIS
1	D	86	ASN
1	D	155	GLN
1	D	253	GLN
2	E	31	HIS
2	E	83	ASN

5.3.3 RNA

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.5 Carbohydrates

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

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^{th} 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	276/276 (100%)	0.01	2 (0%) 87 88	50, 86, 128, 149	0
1	D	276/276 (100%)	0.16	10 (3%) 42 39	51, 92, 137, 157	0
2	B	100/100 (100%)	-0.26	1 (1%) 82 83	51, 72, 108, 122	0
2	E	100/100 (100%)	-0.22	1 (1%) 82 83	48, 72, 107, 128	0
3	C	10/10 (100%)	0.39	0 100 100	74, 77, 90, 91	0
3	F	10/10 (100%)	0.34	0 100 100	73, 79, 88, 98	0
All	All	772/772 (100%)	0.01	14 (1%) 68 65	48, 84, 130, 157	0

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	276	PRO	5.2
2	B	99	MET	5.1
2	E	99	MET	3.4
1	D	199	ALA	3.1
1	D	1	GLY	3.0
1	D	221	GLY	2.9
1	D	276	PRO	2.8
1	D	275	GLU	2.8
1	D	254	GLU	2.6
1	D	193	ALA	2.5
1	D	220	ASP	2.5
1	D	274	TRP	2.3
1	A	220	ASP	2.1
1	D	253	GLN	2.0

6.2 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates

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.