



wwPDB X-ray Structure Validation Summary Report ⓘ

May 17, 2020 – 01:36 pm BST

PDB ID : 2X4S
Title : Crystal structure of MHC CClass I HLA-A2.1 bound to a peptide representing the epitope of the H5N1 (Avian Flu) Nucleoprotein
Authors : Celie, P.H.N.; Toebes, M.; Rodenko, B.; Ovaa, H.; Perrakis, A.; Schumacher, T.N.M.
Deposited on : 2010-02-02
Resolution : 2.55 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
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

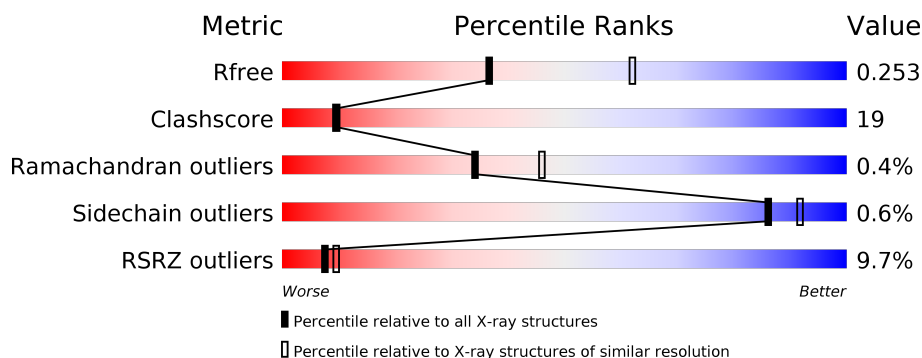
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.55 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	1284 (2.56-2.52)
Clashscore	141614	1332 (2.56-2.52)
Ramachandran outliers	138981	1315 (2.56-2.52)
Sidechain outliers	138945	1315 (2.56-2.52)
RSRZ outliers	127900	1272 (2.56-2.52)

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 $\leq 5\%$. 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	275	<div> <div>5%</div> <div> <div></div> <div>70%</div> <div>29%</div> </div> <div></div> </div>
1	D	275	<div> <div>13%</div> <div> <div></div> <div>68%</div> <div>32%</div> </div> <div></div> </div>
2	B	100	<div> <div>12%</div> <div> <div></div> <div>67%</div> <div>32%</div> </div> <div></div> </div>
2	E	100	<div> <div>9%</div> <div> <div></div> <div>73%</div> <div>27%</div> </div> <div></div> </div>
3	C	9	<div> <div>11%</div> <div> <div></div> <div>44%</div> <div>44%</div> <div>11%</div> </div> <div></div> </div>
3	F	9	<div> <div>33%</div> <div> <div></div> <div>33%</div> <div>67%</div> </div> <div></div> </div>

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 6474 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-2.1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	275	Total	C	N	O	S	0	0	0
			2247	1403	409	426	9			
1	D	275	Total	C	N	O	S	0	0	0
			2247	1403	409	426	9			

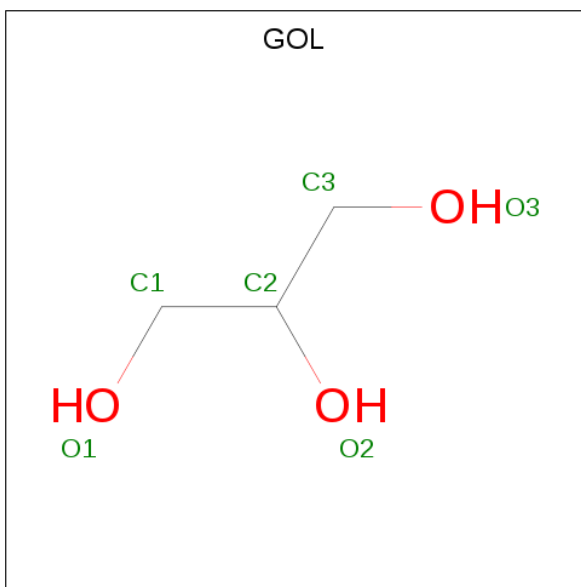
- Molecule 2 is a protein called BETA-2-MICROGLOBULIN.

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

- Molecule 3 is a protein called H5N1 INFLUENZA A NUCLEOPROTEIN.

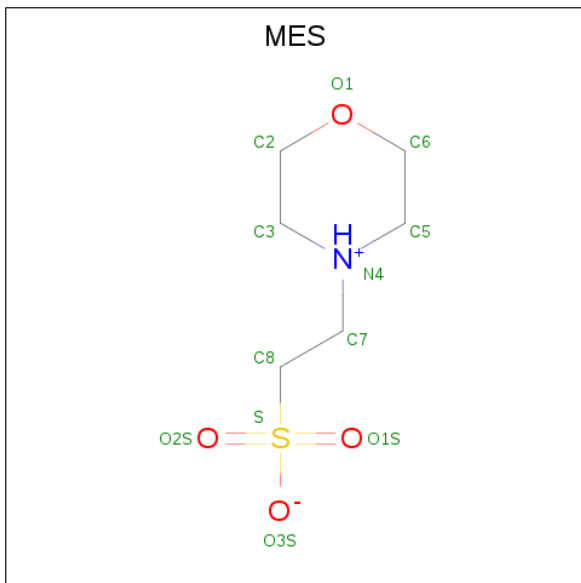
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	9	Total	C	N	O	S	0	0	0
			68	40	10	17	1			
3	F	9	Total	C	N	O	S	0	0	0
			68	40	10	17	1			

- Molecule 4 is GLYCEROL (three-letter code: GOL) (formula: C₃H₈O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	O	0	0
			6	3	3		

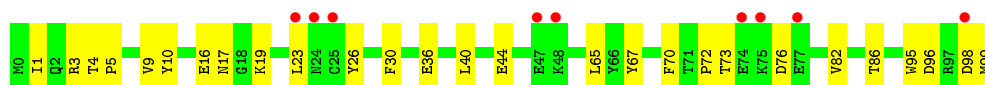
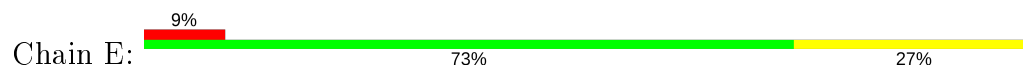
- Molecule 5 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).



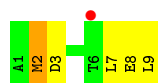
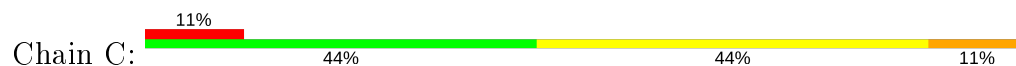
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
5	B	1	Total	C	N	O	S	0	0
			12	6	1	4	1		
5	E	1	Total	C	N	O	S	0	0
			12	6	1	4	1		

- Molecule 6 is water.

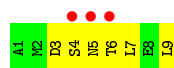
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	59	Total 59	O 59	0	0
6	B	20	Total 20	O 20	0	0
6	D	45	Total 45	O 45	0	0
6	E	16	Total 16	O 16	0	0



- Molecule 3: H5N1 INFLUENZA A NUCLEOPROTEIN



- Molecule 3: H5N1 INFLUENZA A NUCLEOPROTEIN



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	62.38Å 84.11Å 80.03Å 90.00° 90.97° 90.00°	Depositor
Resolution (Å)	19.93 – 2.55 19.93 – 2.55	Depositor EDS
% Data completeness (in resolution range)	99.6 (19.93-2.55) 99.6 (19.93-2.55)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.18 (at 2.56Å)	Xtriage
Refinement program	PHENIX (PHENIX.REFINE)	Depositor
R, R_{free}	0.178 , 0.247 0.195 , 0.253	Depositor DCC
R_{free} test set	1349 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	56.7	Xtriage
Anisotropy	0.031	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.37 , 62.5	EDS
L-test for twinning ²	$\langle L \rangle = 0.47$, $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.027 for -h,-l,-k 0.010 for -h,l,k 0.043 for h,-k,-l	Xtriage
Reported twinning fraction	0.048 for -h,-k,l	Depositor
Outliers	0 of 26956 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	6474	wwPDB-VP
Average B, all atoms (Å ²)	69.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.64% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, MES

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	0.34	0/2312	0.55	2/3137 (0.1%)
1	D	0.31	0/2312	0.47	0/3137
2	B	0.32	0/859	0.46	0/1159
2	E	0.30	0/859	0.47	0/1159
3	C	0.25	0/67	0.56	0/88
3	F	0.28	0/67	0.55	0/88
All	All	0.32	0/6476	0.50	2/8768 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
1	A	267	PRO	CA-N-CD	-7.95	100.37	111.50
1	A	269	PRO	CA-N-CD	-7.94	100.39	111.50

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	2247	0	2096	86	0
1	D	2247	0	2096	97	0
2	B	837	0	803	37	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	837	0	803	36	0
3	C	68	0	66	8	0
3	F	68	0	66	10	0
4	B	6	0	8	1	0
5	B	12	0	12	1	0
5	E	12	0	12	1	0
6	A	59	0	0	8	0
6	B	20	0	0	3	0
6	D	45	0	0	8	0
6	E	16	0	0	0	0
All	All	6474	0	5962	239	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 19.

The worst 5 of 239 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)
2:E:96:ASP:OD2	2:E:99:MSE:HG3	1.33	1.22
1:A:121:LYS:HB2	2:B:1:ILE:CD1	1.78	1.13
2:E:96:ASP:CG	2:E:99:MSE:HG3	1.80	1.00
1:A:121:LYS:HB2	2:B:1:ILE:HD11	1.43	0.99
1:D:51:TRP:CZ3	1:D:171:TYR:HB3	1.97	0.99

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	273/275 (99%)	252 (92%)	19 (7%)	2 (1%)	22	30
1	D	273/275 (99%)	252 (92%)	21 (8%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	B	98/100 (98%)	90 (92%)	8 (8%)	0	100	100
2	E	98/100 (98%)	93 (95%)	5 (5%)	0	100	100
3	C	7/9 (78%)	6 (86%)	0	1 (14%)	0	0
3	F	7/9 (78%)	3 (43%)	4 (57%)	0	100	100
All	All	756/768 (98%)	696 (92%)	57 (8%)	3 (0%)	34	46

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	C	2	MET
1	A	152	VAL
1	A	15	PRO

5.3.2 Protein sidechains ⓘ

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	231/231 (100%)	230 (100%)	1 (0%)	91	95
1	D	231/231 (100%)	230 (100%)	1 (0%)	91	95
2	B	95/93 (102%)	94 (99%)	1 (1%)	73	83
2	E	95/93 (102%)	94 (99%)	1 (1%)	73	83
3	C	8/8 (100%)	8 (100%)	0	100	100
3	F	8/8 (100%)	8 (100%)	0	100	100
All	All	668/664 (101%)	664 (99%)	4 (1%)	86	92

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	113	TYR
2	B	70	PHE
1	D	113	TYR
2	E	70	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 21 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	42	ASN
1	D	96	GLN
1	D	253	GLN
1	A	260	HIS
1	D	260	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 [i](#)

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\# Z > 2$	Counts	RMSZ	$\# Z > 2$
5	MES	E	1099	-	12,12,12	1.77	1 (8%)	14,16,16	7.73	10 (71%)
4	GOL	B	1099	-	5,5,5	0.42	0	5,5,5	0.32	0
5	MES	B	1100	-	12,12,12	2.10	1 (8%)	14,16,16	7.77	10 (71%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns.

'-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
5	MES	E	1099	-	-	1/6/14/14	0/1/1/1
4	GOL	B	1099	-	-	1/4/4/4	-
5	MES	B	1100	-	-	2/6/14/14	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
5	B	1100	MES	C8-S	-6.68	1.68	1.77
5	E	1099	MES	C8-S	-5.39	1.69	1.77

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
5	B	1100	MES	O1S-S-C8	-18.57	84.55	106.92
5	E	1099	MES	O2S-S-C8	-16.38	87.19	106.92
5	B	1100	MES	O2S-S-C8	-16.37	87.20	106.92
5	E	1099	MES	O1S-S-C8	-16.07	87.57	106.92
5	E	1099	MES	O3S-S-C8	-13.23	84.37	105.77

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	B	1100	MES	N4-C7-C8-S
5	E	1099	MES	C8-C7-N4-C3
5	B	1100	MES	C8-C7-N4-C3
4	B	1099	GOL	O1-C1-C2-O2

There are no ring outliers.

3 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	E	1099	MES	1	0
4	B	1099	GOL	1	0
5	B	1100	MES	1	0

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, 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	275/275 (100%)	0.40	14 (5%)	28 33	32, 62, 92, 120	0
1	D	275/275 (100%)	0.80	35 (12%)	3 4	36, 76, 108, 138	0
2	B	98/100 (98%)	0.59	12 (12%)	4 5	39, 66, 94, 113	0
2	E	98/100 (98%)	0.51	9 (9%)	9 10	37, 61, 91, 115	0
3	C	9/9 (100%)	0.47	1 (11%)	5 7	65, 84, 100, 100	0
3	F	9/9 (100%)	1.29	3 (33%)	0 0	66, 98, 125, 140	0
All	All	764/768 (99%)	0.59	74 (9%)	7 9	32, 68, 102, 140	0

The worst 5 of 74 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	91	GLY	7.4
1	D	225	THR	5.8
1	D	16	GLY	5.7
1	D	193	ALA	5.0
1	A	149	ALA	4.5

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](#)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(\AA^2)	Q<0.9
4	GOL	B	1099	6/6	0.78	0.28	56,63,69,72	0
5	MES	E	1099	12/12	0.91	0.21	66,79,97,98	0
5	MES	B	1100	12/12	0.93	0.17	53,64,82,89	0

6.5 Other polymers [i](#)

There are no such residues in this entry.