wwPDB X-ray Structure Validation Summary Report (i)

Sep 20, 2023 – 01:06 AM EDT

PDB ID	:	5INC
Title	:	Crystal structure of HLA-B5801, a protective HLA allele for HIV-1 infection
Authors	:	Li, X.; Wang, J.-H.
Deposited on	:	2016-03-07
Resolution	:	2.88 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 (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:

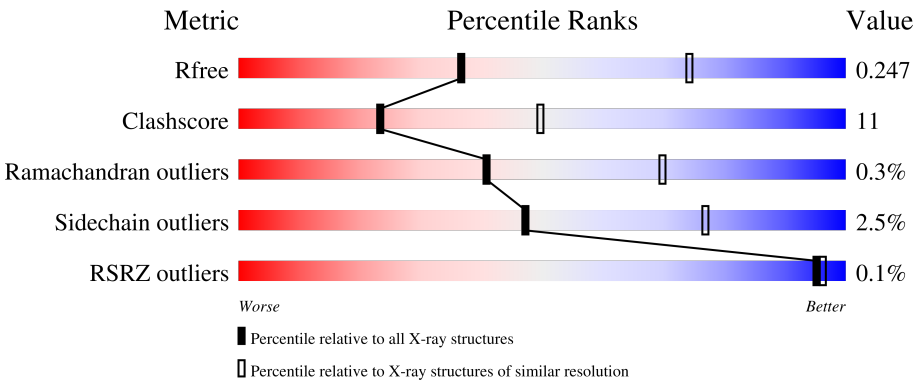
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 2.88 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	2691 (2.90-2.86)
Clashscore	141614	2947 (2.90-2.86)
Ramachandran outliers	138981	2868 (2.90-2.86)
Sidechain outliers	138945	2871 (2.90-2.86)
RSRZ outliers	127900	2629 (2.90-2.86)

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	277	77% 22%
1	C	277	77% 23%
2	B	99	70% 30%
2	D	99	70% 29% •
3	E	9	67% 11% 22%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	9	11% 33% 56% 11%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6334 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, B-58 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	277	2254	1405	412	429	8	0	0	0
1	C	277	2254	1405	412	429	8	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	194	ILE	VAL	variant	UNP P10319
A	277	HIS	-	expression tag	UNP P10319
C	194	ILE	VAL	variant	UNP P10319
C	277	HIS	-	expression tag	UNP P10319

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	99	829	528	140	158	3	0	0	0
2	D	99	829	528	140	158	3	0	0	0

- Molecule 3 is a protein called GLN-ALA-THR-GLN-GLU-VAL-LYS-ASN-TRP.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
3	E	9	74	45	13	16	0	0	0
3	F	9	78	48	14	16	0	0	0

- Molecule 4 is water.

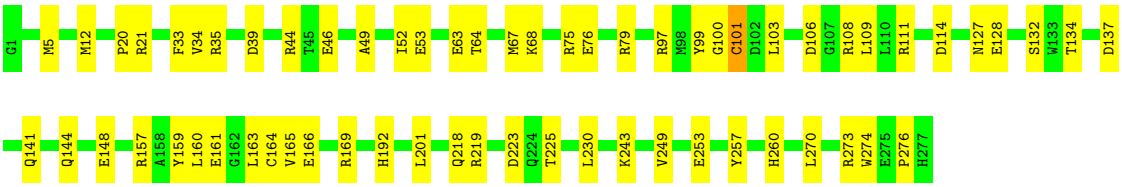
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
4	A	2	2	2	0	0
			Total	O		
4	C	5	5	5	0	0
			Total	O		
4	E	6	6	6	0	0
			Total	O		
4	F	3	3	3	0	0

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.

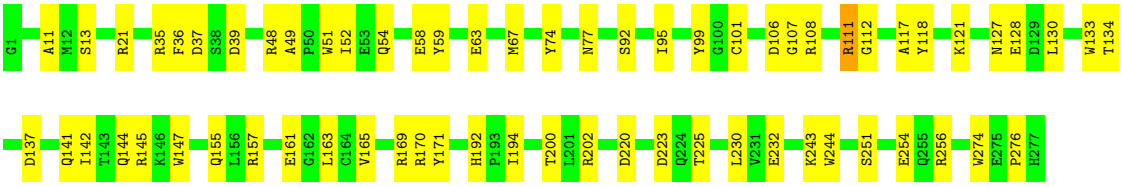
• Molecule 1: HLA class I histocompatibility antigen, B-58 alpha chain

Chain A:



• Molecule 1: HLA class I histocompatibility antigen, B-58 alpha chain

Chain C:

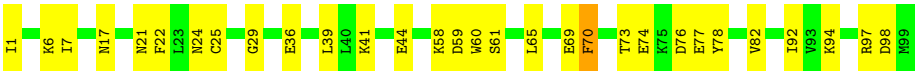


• Molecule 2: Beta-2-microglobulin

Chain B:

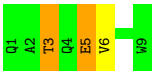
• Molecule 2: Beta-2-microglobulin

Chain D:



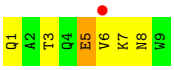
• Molecule 3: GLN-ALA-THR-GLN-GLU-VAL-LYS-ASN-TRP

Chain E:



- Molecule 3: GLN-ALA-THR-GLN-GLU-VAL-LYS-ASN-TRP

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	68.40Å 82.41Å 154.62Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.70 – 2.88 43.70 – 2.88	Depositor EDS
% Data completeness (in resolution range)	91.8 (43.70-2.88) 86.5 (43.70-2.88)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.81 (at 2.90Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.195 , 0.248 0.195 , 0.247	Depositor DCC
R_{free} test set	947 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	49.9	Xtriage
Anisotropy	0.136	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 22.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	6334	wwPDB-VP
Average B, all atoms (Å^2)	36.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.08% 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	0.58	2/2317 (0.1%)	0.71	2/3150 (0.1%)
1	C	0.54	0/2317	0.71	1/3150 (0.0%)
2	B	0.57	0/852	0.69	0/1152
2	D	0.76	4/852 (0.5%)	0.66	0/1152
3	E	0.74	0/75	0.84	0/101
3	F	0.68	0/79	0.48	0/105
All	All	0.60	6/6492 (0.1%)	0.70	3/8810 (0.0%)

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	69	GLU	CD-OE1	-7.36	1.17	1.25
1	A	164	CYS	CB-SG	-6.77	1.70	1.82
2	D	74	GLU	CD-OE1	-6.56	1.18	1.25
2	D	69	GLU	CD-OE2	-5.66	1.19	1.25
2	D	74	GLU	CD-OE2	-5.65	1.19	1.25

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	101	CYS	CA-CB-SG	-8.52	98.67	114.00
1	C	101	CYS	CA-CB-SG	-7.92	99.75	114.00
1	A	164	CYS	CA-CB-SG	-6.18	102.88	114.00

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	2254	0	2119	43	1
1	C	2254	0	2119	56	1
2	B	829	0	794	22	0
2	D	829	0	794	26	0
3	E	74	0	63	11	0
3	F	78	0	74	9	0
4	A	2	0	0	1	0
4	C	5	0	0	2	0
4	E	6	0	0	1	0
4	F	3	0	0	0	0
All	All	6334	0	5963	139	1

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 11.

The worst 5 of 139 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)
3:F:5:GLU:OE2	3:F:7:LYS:NZ	1.88	1.06
2:B:58:LYS:HD2	2:B:58:LYS:H	1.23	1.04
1:C:111:ARG:NH1	1:C:128:GLU:O	1.93	1.01
1:C:111:ARG:HA	1:C:111:ARG:HE	1.30	0.96
2:B:42:ASN:ND2	2:B:76:ASP:OD1	2.03	0.91

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:53:GLU:OE2	1:C:157:ARG:NH2[3_554]	2.19	0.01

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%)	271 (98%)	4 (2%)	0	100	100
1	C	275/277 (99%)	269 (98%)	5 (2%)	1 (0%)	34	64
2	B	97/99 (98%)	97 (100%)	0	0	100	100
2	D	97/99 (98%)	97 (100%)	0	0	100	100
3	E	7/9 (78%)	6 (86%)	0	1 (14%)	0	0
3	F	7/9 (78%)	7 (100%)	0	0	100	100
All	All	758/770 (98%)	747 (98%)	9 (1%)	2 (0%)	41	70

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	106	ASP
3	E	5	GLU

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	232/232 (100%)	228 (98%)	4 (2%)	60	84
1	C	232/232 (100%)	228 (98%)	4 (2%)	60	84
2	B	94/94 (100%)	91 (97%)	3 (3%)	39	71
2	D	94/94 (100%)	92 (98%)	2 (2%)	53	80
3	E	7/8 (88%)	6 (86%)	1 (14%)	3	9
3	F	8/8 (100%)	5 (62%)	3 (38%)	0	0
All	All	667/668 (100%)	650 (98%)	17 (2%)	47	77

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	F	1	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
3	F	6	VAL
1	C	111	ARG
1	C	170	ARG
1	C	251	SER

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

Mol	Chain	Res	Type
1	C	54	GLN
1	C	192	HIS
1	C	260	HIS
3	F	1	GLN
3	F	8	ASN

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	277/277 (100%)	-0.34	0	100	100	21, 34, 55, 85	0
1	C	277/277 (100%)	-0.32	0	100	100	23, 34, 54, 83	0
2	B	99/99 (100%)	-0.25	0	100	100	23, 36, 62, 76	0
2	D	99/99 (100%)	-0.38	0	100	100	22, 34, 53, 66	0
3	E	9/9 (100%)	0.04	0	100	100	26, 37, 61, 67	0
3	F	9/9 (100%)	0.28	1 (11%)	5	4	29, 36, 61, 62	0
All	All	770/770 (100%)	-0.31	1 (0%)	95	96	21, 34, 57, 85	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	6	VAL	2.3

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.