wwPDB X-ray Structure Validation Summary Report ⓘ

May 21, 2020 – 10:44 am BST

PDB ID	:	4O2F
Title	:	A peptide complexed with HLA-B*3901
Authors	:	Sun, M.; Liu, J.; Qi, J.; Tefsen, B.; Shi, Y.; Yan, J.; Gao, G.F.
Deposited on	:	2013-12-17
Resolution	:	1.90 Å(reported)

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

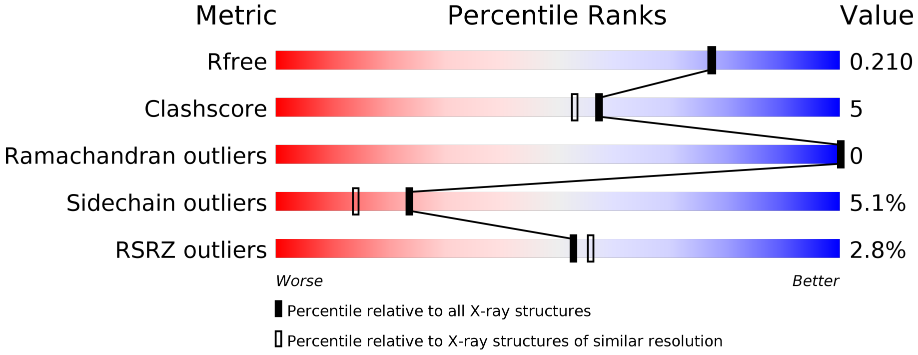
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation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

1 Overall quality at a glance

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

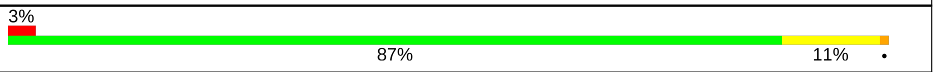
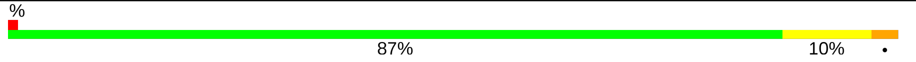
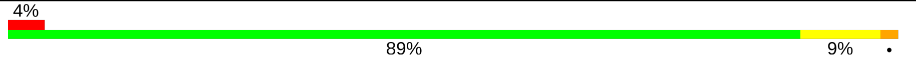
The reported resolution of this entry is 1.90 Å.

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}	130704	6207 (1.90-1.90)
Clashscore	141614	6847 (1.90-1.90)
Ramachandran outliers	138981	6760 (1.90-1.90)
Sidechain outliers	138945	6760 (1.90-1.90)
RSRZ outliers	127900	6082 (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 >=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	274	
1	D	274	
2	B	99	
2	E	99	
3	C	8	
3	F	8	

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 7133 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-39 alpha chain.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
1	A	274	Total C N O S 2236 1388 410 431 7	0	0	0
1	D	274	Total C N O S 2237 1388 410 432 7	0	0	0

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

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

- Molecule 3 is a protein called Peptide from ATP-dependent RNA helicase DDX3X.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf	Trace
3	C	8	Total C N O 60 37 11 12	0	0	0
3	F	8	Total C N O 60 37 11 12	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	311	Total O 311 311	0	0
4	B	132	Total O 132 132	0	0
4	C	9	Total O 9 9	0	0
4	D	293	Total O 293 293	0	0

Continued on next page...

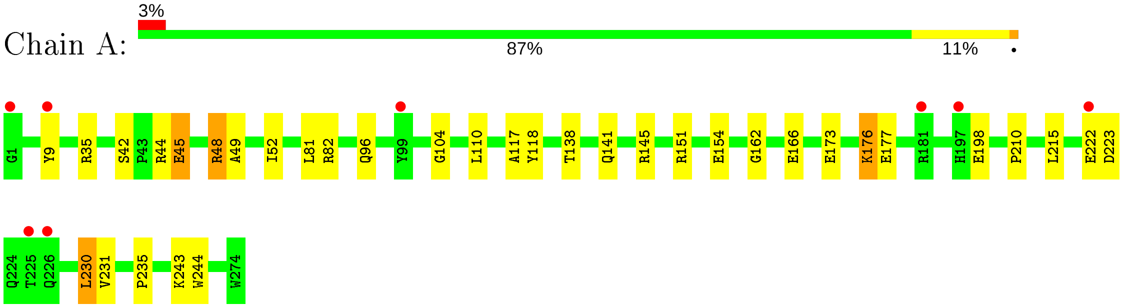
Continued from previous page...

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	E	128	Total O 128 128	0	0
4	F	9	Total O 9 9	0	0

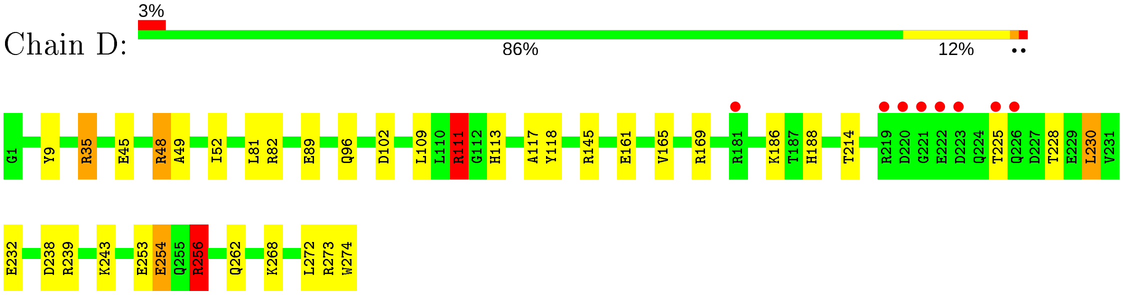
3 Residue-property plots

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.

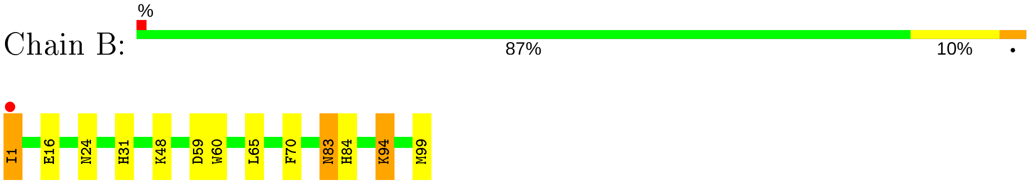
- Molecule 1: HLA class I histocompatibility antigen, B-39 alpha chain



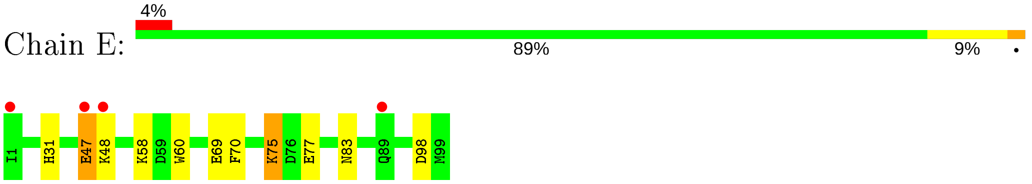
- Molecule 1: HLA class I histocompatibility antigen, B-39 alpha chain



- Molecule 2: Beta-2-microglobulin

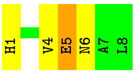


- Molecule 2: Beta-2-microglobulin



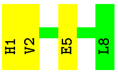
- Molecule 3: Peptide from ATP-dependent RNA helicase DDX3X

Chain C:



- Molecule 3: Peptide from ATP-dependent RNA helicase DDX3X

Chain F:



4 Data and refinement statistics ⓘ

Property	**Value**	**Source**
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	78.37Å 95.92Å 122.93Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.60 – 1.90 31.24 – 1.90	Depositor EDS
% Data completeness (in resolution range)	92.6 (29.60-1.90) 96.6 (31.24-1.90)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.89 (at 1.91Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.7_650)	Depositor
R, R_{free}	0.185 , 0.213 0.184 , 0.210	Depositor DCC
R_{free} test set	3587 reflections (5.05%)	wwPDB-VP
Wilson B-factor (Å^2)	19.5	Xtriage
Anisotropy	0.145	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 52.6	EDS
L-test for twinning[2]	$< \|L\| >$ = 0.49, $< L^2 >$ = 0.33	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	7133	wwPDB-VP
Average B, all atoms (Å^2)	24.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.92% 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	0.53	2/2297 (0.1%)	0.52	3/3124 (0.1%)
1	D	0.49	1/2298 (0.0%)	0.58	5/3124 (0.2%)
2	B	0.54	1/852 (0.1%)	0.47	0/1152
2	E	0.25	0/852	0.42	0/1152
3	C	0.53	0/60	0.44	0/80
3	F	1.27	1/60 (1.7%)	0.69	0/80
All	All	0.50	5/6419 (0.1%)	0.53	8/8712 (0.1%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	99	MET	C-OXT	-8.90	1.06	1.23
1	A	48	ARG	CB-CG	-6.79	1.34	1.52
1	D	254	GLU	CD-OE2	-5.70	1.19	1.25
1	A	82	ARG	CB-CG	-5.69	1.37	1.52
3	F	2	VAL	CB-CG1	-5.37	1.41	1.52

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	145	ARG	NE-CZ-NH2	-9.15	115.73	120.30
1	D	48	ARG	NE-CZ-NH2	-8.48	116.06	120.30
1	D	48	ARG	NE-CZ-NH1	8.29	124.45	120.30
1	A	48	ARG	NE-CZ-NH1	7.83	124.21	120.30
1	A	48	ARG	NE-CZ-NH2	-6.54	117.03	120.30

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	2236	0	2094	17	0
1	D	2237	0	2094	23	0
2	B	829	0	794	11	0
2	E	829	0	794	8	0
3	C	60	0	60	5	0
3	F	60	0	60	2	0
4	A	311	0	0	2	0
4	B	132	0	0	2	0
4	C	9	0	0	0	0
4	D	293	0	0	2	0
4	E	128	0	0	3	0
4	F	9	0	0	0	0
All	All	7133	0	5896	58	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 5.

The worst 5 of 58 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:B:1:ILE:N	2:B:1:ILE:HD13	1.70	1.06
2:B:1:ILE:H3	2:B:1:ILE:HD13	1.35	0.92
3:C:5:GLU:O	3:C:5:GLU:HG3	1.68	0.92
2:B:1:ILE:CD1	2:B:1:ILE:N	2.39	0.84
1:D:35:ARG:HG2	1:D:48:ARG:HD2	1.61	0.82

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	272/274 (99%)	264 (97%)	8 (3%)	0	100	100
1	D	272/274 (99%)	266 (98%)	6 (2%)	0	100	100
2	B	97/99 (98%)	95 (98%)	2 (2%)	0	100	100
2	E	97/99 (98%)	96 (99%)	1 (1%)	0	100	100
3	C	6/8 (75%)	5 (83%)	1 (17%)	0	100	100
3	F	6/8 (75%)	5 (83%)	1 (17%)	0	100	100
All	All	750/762 (98%)	731 (98%)	19 (2%)	0	100	100

There are no Ramachandran outliers to report.

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	235/235 (100%)	225 (96%)	10 (4%)	29	19
1	D	235/235 (100%)	225 (96%)	10 (4%)	29	19
2	B	94/94 (100%)	88 (94%)	6 (6%)	17	8
2	E	94/94 (100%)	88 (94%)	6 (6%)	17	8
3	C	6/6 (100%)	5 (83%)	1 (17%)	2	0
3	F	6/6 (100%)	5 (83%)	1 (17%)	2	0
All	All	670/670 (100%)	636 (95%)	34 (5%)	24	14

5 of 34 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	94	LYS
1	D	186	LYS
2	E	83	ASN
1	D	35	ARG
1	A	198	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 23 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	32	GLN
1	D	113	HIS
3	F	1	HIS
1	D	65	GLN
1	D	115	GLN

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, 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	274/274 (100%)	0.08	8 (2%)	51	54	10, 19, 40, 77	0
1	D	274/274 (100%)	0.03	8 (2%)	51	54	9, 21, 46, 71	0
2	B	99/99 (100%)	-0.11	1 (1%)	82	84	10, 20, 38, 49	0
2	E	99/99 (100%)	-0.02	4 (4%)	38	41	11, 21, 42, 59	0
3	C	8/8 (100%)	0.56	0	100	100	16, 21, 34, 56	0
3	F	8/8 (100%)	0.10	0	100	100	14, 18, 29, 41	0
All	All	762/762 (100%)	0.03	21 (2%)	53	56	9, 21, 42, 77	0

The worst 5 of 21 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	197	HIS	3.6
1	D	221	GLY	3.5
1	D	222	GLU	3.5
2	B	1	ILE	3.4
1	A	1	GLY	3.4

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.