wwPDB X-ray Structure Validation Summary Report (i)

Mar 7, 2022 – 10:26 AM EST

PDB ID	:	7S0X
Title	:	Fab352 in complex with the C-terminal alphaTSR domain of P. falciparum
Authors	:	Pholcharee, T.; Oyen, D.; Wilson, I.A.
Deposited on	:	2021-08-31
Resolution	:	2.80 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 following versions of software and data (see references (1)) were used in the production of this report:

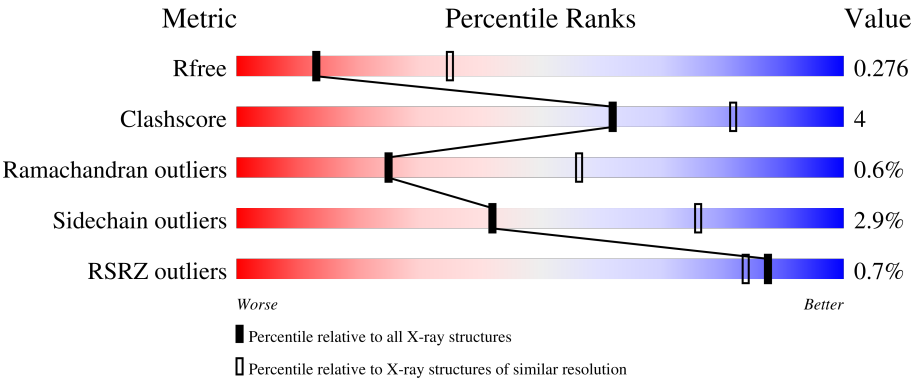
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.27
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 2.80 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.



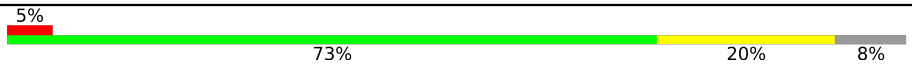
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	3140 (2.80-2.80)
Clashscore	141614	3569 (2.80-2.80)
Ramachandran outliers	138981	3498 (2.80-2.80)
Sidechain outliers	138945	3500 (2.80-2.80)
RSRZ outliers	127900	3078 (2.80-2.80)

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	D	213	85% 12% ·
1	L	213	84% 14% ··
2	C	223	87% 12% ·
2	H	223	91% 8% ·
3	E	66	2% 79% 12% · 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	66	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 14345 atoms, of which 7023 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 Fab352 light chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	L	209	Total	C	H	N	O	S	0	0	0
			3005	964	1470	257	310	4			
1	D	208	Total	C	H	N	O	S	0	0	0
			3037	968	1495	263	307	4			

- Molecule 2 is a protein called Fab352 heavy chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	H	221	Total	C	H	N	O	S	0	0	0
			3221	1030	1586	280	318	7			
2	C	221	Total	C	H	N	O	S	0	0	0
			3190	1027	1563	278	315	7			

- Molecule 3 is a protein called Circumsporozoite protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	E	61	Total	C	H	N	O	S	0	0	0
			921	289	455	81	92	4			
3	F	61	Total	C	H	N	O	S	0	0	0
			922	290	454	80	94	4			

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	L	11	Total	O	0	0
			11	11		
4	H	9	Total	O	0	0
			9	9		
4	D	16	Total	O	0	0
			16	16		
4	C	8	Total	O	0	0
			8	8		

Continued on next page...

Continued from previous page...

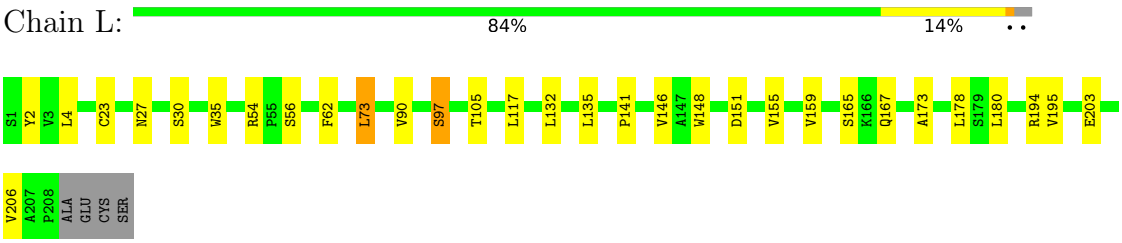
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	F	5	Total O 5 5	0	0

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.

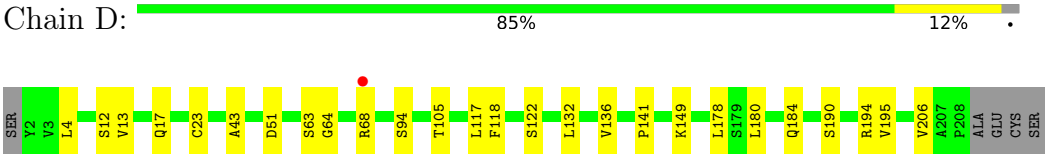
• Molecule 1: Fab352 light chain

Chain L:



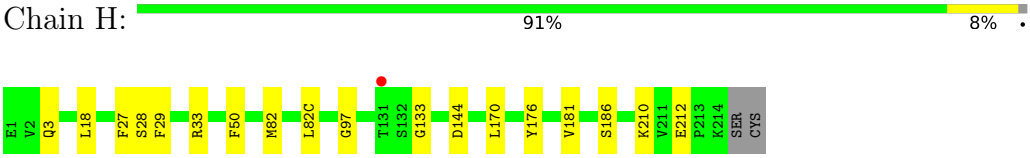
• Molecule 1: Fab352 light chain

Chain D:



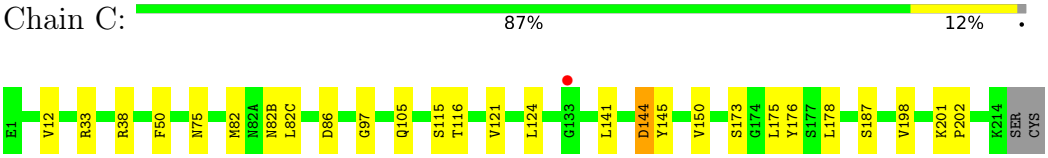
• Molecule 2: Fab352 heavy chain

Chain H:



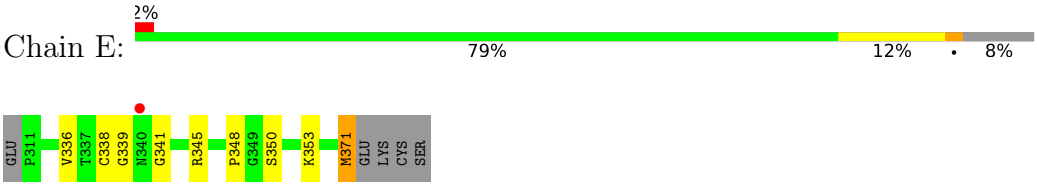
• Molecule 2: Fab352 heavy chain

Chain C:



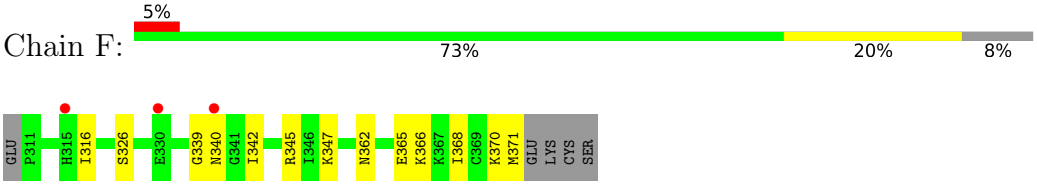
• Molecule 3: Circumsporozoite protein

Chain E:



• Molecule 3: Circumsporozoite protein

Chain F:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	84.06Å 70.50Å 105.96Å 90.00° 100.23° 90.00°	Depositor
Resolution (Å)	41.36 – 2.80 41.92 – 2.77	Depositor EDS
% Data completeness (in resolution range)	96.1 (41.36-2.80) 96.4 (41.92-2.77)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.21	Depositor
$< I/\sigma(I) >$ [1]	1.49 (at 2.77Å)	Xtriage
Refinement program	PHENIX (1.19_4092: ???)	Depositor
R, R_{free}	0.225 , 0.276 0.227 , 0.276	Depositor DCC
R_{free} test set	1503 reflections (4.95%)	wwPDB-VP
Wilson B-factor (Å^2)	38.6	Xtriage
Anisotropy	0.265	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 29.8	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	0.92	EDS
Total number of atoms	14345	wwPDB-VP
Average B, all atoms (Å^2)	45.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 72.54 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 2.1632e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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	D	0.28	0/1581	0.51	0/2165
1	L	0.28	0/1574	0.48	0/2159
2	C	0.36	0/1665	0.57	0/2268
2	H	0.28	0/1673	0.53	0/2278
3	E	0.26	0/474	0.46	0/639
3	F	0.26	0/476	0.45	0/642
All	All	0.30	0/7443	0.52	0/10151

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

5.2 Too-close contacts

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	D	1542	1495	1495	14	0
1	L	1535	1470	1471	16	0
2	C	1627	1563	1561	18	0
2	H	1635	1586	1586	8	0
3	E	466	455	455	3	0
3	F	468	454	452	8	0
4	C	8	0	0	0	0
4	D	16	0	0	0	0
4	F	5	0	0	0	0
4	H	9	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	L	11	0	0	0	0
All	All	7322	7023	7020	63	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 4.

The worst 5 of 63 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:C:121:VAL:HG11	2:C:198:VAL:HG11	1.72	0.72
2:C:12:VAL:HG11	2:C:82(C):LEU:HD12	1.79	0.65
1:D:132:LEU:HD12	1:D:178:LEU:HD23	1.78	0.64
1:D:149:LYS:HE2	1:D:194:ARG:HD2	1.80	0.64
2:C:38:ARG:NH1	2:C:86:ASP:OD1	2.31	0.61

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	D	206/213 (97%)	196 (95%)	9 (4%)	1 (0%)	29 61
1	L	207/213 (97%)	189 (91%)	17 (8%)	1 (0%)	29 61
2	C	219/223 (98%)	204 (93%)	14 (6%)	1 (0%)	29 61
2	H	219/223 (98%)	213 (97%)	4 (2%)	2 (1%)	17 46
3	E	59/66 (89%)	56 (95%)	2 (3%)	1 (2%)	9 29
3	F	59/66 (89%)	56 (95%)	3 (5%)	0	100 100
All	All	969/1004 (96%)	914 (94%)	49 (5%)	6 (1%)	25 56

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	H	28	SER
2	C	144	ASP
1	L	151	ASP
1	D	68	ARG
3	E	348	PRO

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	D	170/177 (96%)	166 (98%)	4 (2%)	49	81
1	L	168/177 (95%)	161 (96%)	7 (4%)	30	63
2	C	174/186 (94%)	170 (98%)	4 (2%)	50	82
2	H	179/186 (96%)	176 (98%)	3 (2%)	60	87
3	E	53/61 (87%)	50 (94%)	3 (6%)	20	50
3	F	53/61 (87%)	51 (96%)	2 (4%)	33	67
All	All	797/848 (94%)	774 (97%)	23 (3%)	42	76

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	50	PHE
2	C	187	SER
2	C	173	SER
3	E	338	CYS
1	L	165	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	L	17	GLN
1	L	27	ASN
1	L	167	GLN
2	C	82(B)	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 monosaccharides 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	D	208/213 (97%)	-0.17	1 (0%) 91 88	23, 36, 55, 77	0
1	L	209/213 (98%)	-0.24	0 100 100	25, 37, 53, 67	0
2	C	221/223 (99%)	-0.19	1 (0%) 91 88	26, 38, 55, 75	0
2	H	221/223 (99%)	-0.16	1 (0%) 91 88	22, 38, 56, 72	0
3	E	61/66 (92%)	0.25	1 (1%) 72 66	40, 52, 66, 76	0
3	F	61/66 (92%)	0.50	3 (4%) 29 20	43, 56, 70, 82	0
All	All	981/1004 (97%)	-0.12	7 (0%) 87 84	22, 39, 61, 82	0

The worst 5 of 7 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	F	340	ASN	3.5
3	F	315	HIS	3.4
1	D	68	ARG	3.1
2	H	131	THR	2.6
3	E	340	ASN	2.4

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 monosaccharides in this entry.

6.4 Ligands

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.