Full wwPDB X-ray Structure Validation Report

May 25, 2020 – 01:54 am BST

PDB ID	:	3CH3
Title	:	Crystal Structure Analysis of SERA5E from plasmodium falciparum
Authors	:	Smith, B.J.; Malby, R.L.; Colman, P.M.; Clarke, O.B.
Deposited on	:	2008-03-07
Resolution	:	1.79 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 (i) symbol.

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

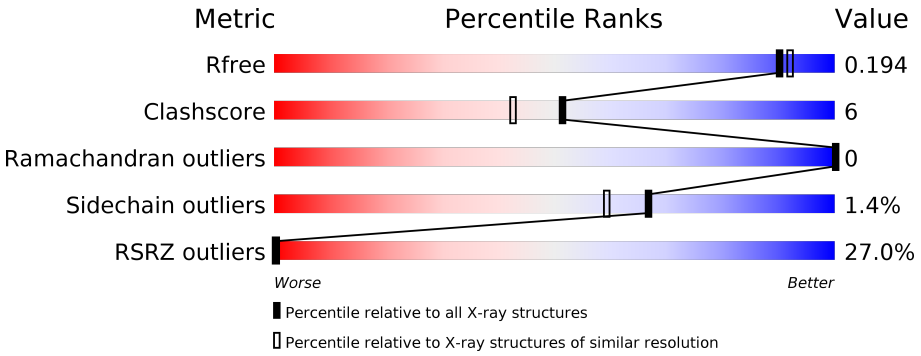
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 1.79 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)
RSRZ outliers	127900	5850 (1.80-1.80)

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 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	X	265	26% 87% 9%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 2265 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.

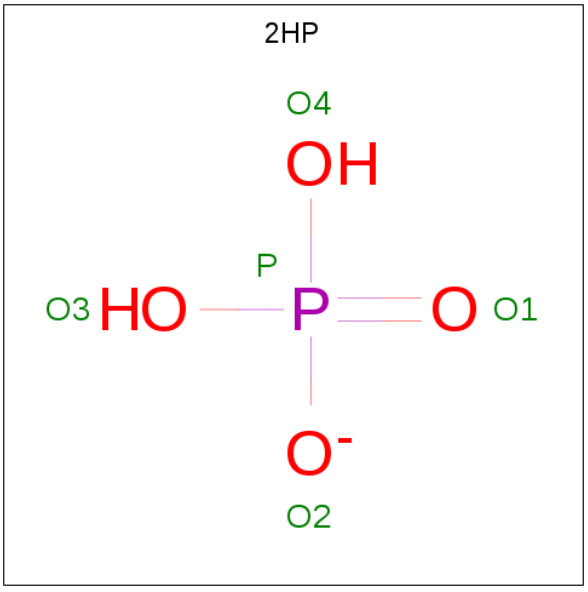
- Molecule 1 is a protein called Serine-repeat antigen protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	X	256	2065	1314	333	394	24	0	10	0

- Molecule 2 is POTASSIUM ION (three-letter code: K) (formula: K).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	K		
2	X	1	1	1	0	0

- Molecule 3 is DIHYDROGENPHOSPHATE ION (three-letter code: 2HP) (formula: H_2O_4P).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	O	P		
3	X	1	5	4	1	0	0
			Total	O	P		
3	X	1	5	4	1	0	0

- Molecule 4 is water.

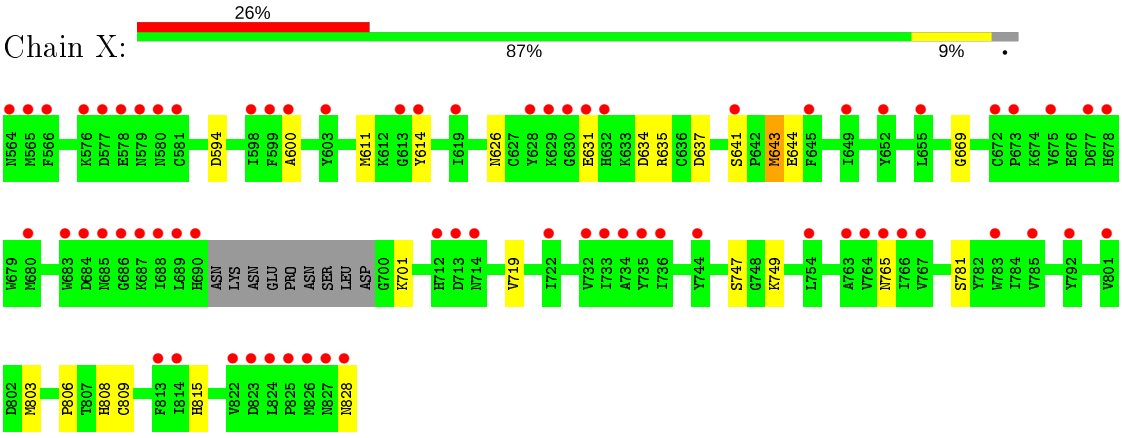
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	X	188	Total O 189 189	0	3

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: Serine-repeat antigen protein

Chain X:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	H 3	Depositor
Cell constants a, b, c, α, β, γ	103.37Å 103.37Å 72.51Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	24.74 – 1.79 24.74 – 1.79	Depositor EDS
% Data completeness (in resolution range)	98.5 (24.74-1.79) 98.5 (24.74-1.79)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.12 (at 1.79Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.168 , 0.198 0.165 , 0.194	Depositor DCC
R_{free} test set	1344 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	32.6	Xtriage
Anisotropy	0.046	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 54.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.024 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	2265	wwPDB-VP
Average B, all atoms (Å^2)	37.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.52% 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

Bond lengths and bond angles in the following residue types are not validated in this section: 2HP, K

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	X	0.72	0/2167	0.68	0/2932

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	X	2065	0	1867	22	0
2	X	1	0	0	0	0
3	X	10	0	0	1	0
4	X	189	0	0	8	1
All	All	2265	0	1867	22	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 6.

All (22) 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:X:635:ARG:NH2	1:X:644:GLU:OE2	2.13	0.82
1:X:643:MET:HE2	1:X:643:MET:H	1.50	0.76
1:X:701:LYS:NZ	4:X:927:HOH:O	2.18	0.76
1:X:626:ASN:HD21	1:X:669:GLY:H	1.42	0.65
1:X:626:ASN:ND2	1:X:669:GLY:H	1.94	0.65
1:X:600:ALA:HA	1:X:765:ASN:HD21	1.66	0.60
1:X:611[B]:MET:HE3	4:X:1015:HOH:O	2.02	0.59
1:X:781[B]:SER:OG	1:X:803[B]:MET:HE2	2.05	0.56
1:X:594:ASP:HB3	3:X:829:2HP:O4	2.06	0.56
1:X:781[B]:SER:OG	1:X:803[B]:MET:CE	2.55	0.55
1:X:643:MET:H	1:X:643:MET:CE	2.20	0.53
1:X:749:LYS:C	4:X:851:HOH:O	2.49	0.50
1:X:611[A]:MET:HE3	4:X:923:HOH:O	2.13	0.48
1:X:701:LYS:NZ	4:X:849:HOH:O	2.47	0.47
1:X:614:TYR:CD2	1:X:828:ASN:HB3	2.51	0.46
1:X:611[A]:MET:CE	4:X:923:HOH:O	2.67	0.43
1:X:719:VAL:HG11	1:X:803[B]:MET:HE3	2.00	0.42
1:X:815:HIS:HE1	4:X:941:HOH:O	2.03	0.42
1:X:808:HIS:HD2	4:X:921:HOH:O	2.03	0.41
1:X:634:ASP:CG	1:X:637:ASP:HB2	2.41	0.41
1:X:806:PRO:HD2	1:X:809:CYS:SG	2.61	0.40
1:X:641:SER:OG	1:X:643:MET:HG2	2.22	0.40

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)
4:X:848:HOH:O	4:X:943:HOH:O[2_555]	2.09	0.11

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	X	262/265 (99%)	254 (97%)	8 (3%)	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	X	223/235 (95%)	219 (98%)	4 (2%)	59 48

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

Mol	Chain	Res	Type
1	X	631	GLU
1	X	643	MET
1	X	747[A]	SER
1	X	747[B]	SER

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

Mol	Chain	Res	Type
1	X	626	ASN
1	X	660	ASN
1	X	752	GLN
1	X	765	ASN
1	X	808	HIS

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 (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 1 is monoatomic - leaving 2 for Mogul analysis.

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	$\#\|Z\| > 2$
3	2HP	X	830	2	4,4,4	5.06	2 (50%)	6,6,6	0.95	0
3	2HP	X	829	-	4,4,4	5.56	2 (50%)	6,6,6	1.41	0

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	X	829	2HP	P-O4	8.08	1.78	1.54
3	X	829	2HP	P-O3	7.31	1.76	1.54
3	X	830	2HP	P-O3	7.20	1.76	1.54
3	X	830	2HP	P-O4	7.03	1.75	1.54

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	X	829	2HP	1	0

5.7 Other polymers (i)

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	X	256/265 (96%)	1.35	69 (26%) 0 0	24, 34, 50, 57	0

All (69) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	X	828	ASN	10.3
1	X	564	ASN	8.3
1	X	825	PRO	6.5
1	X	685	ASN	6.2
1	X	827	ASN	5.7
1	X	826	MET	5.6
1	X	565	MET	5.4
1	X	678	HIS	5.3
1	X	566	PHE	5.3
1	X	631	GLU	5.0
1	X	652	TYR	5.0
1	X	677	ASP	4.9
1	X	733	ILE	4.9
1	X	689	LEU	4.8
1	X	824	LEU	4.8
1	X	613	GLY	4.8
1	X	619	ILE	4.7
1	X	785	VAL	4.6
1	X	577	ASP	4.4
1	X	579	ASN	4.3
1	X	766	ILE	4.3
1	X	714	ASN	4.3
1	X	683	TRP	4.2
1	X	632	HIS	4.2
1	X	814	ILE	4.1
1	X	813	PHE	4.1
1	X	686	GLY	4.0

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	X	614	TYR	3.9
1	X	764	VAL	3.9
1	X	823	ASP	3.9
1	X	576	LYS	3.5
1	X	630	GLY	3.4
1	X	801	VAL	3.4
1	X	734	ALA	3.3
1	X	712	HIS	3.3
1	X	792	TYR	3.2
1	X	687	LYS	3.1
1	X	581	CYS	3.0
1	X	645	PHE	3.0
1	X	822	VAL	3.0
1	X	736	ILE	3.0
1	X	744	TYR	3.0
1	X	578	GLU	3.0
1	X	688	ILE	2.9
1	X	684	ASP	2.9
1	X	628	TYR	2.9
1	X	580	ASN	2.8
1	X	783	TRP	2.7
1	X	754	LEU	2.6
1	X	732	VAL	2.6
1	X	767	VAL	2.5
1	X	655	LEU	2.4
1	X	763	ALA	2.4
1	X	713	ASP	2.4
1	X	690	HIS	2.4
1	X	629	LYS	2.3
1	X	765	ASN	2.3
1	X	675	VAL	2.3
1	X	649	ILE	2.3
1	X	600	ALA	2.2
1	X	673	PRO	2.2
1	X	722	ILE	2.2
1	X	599	PHE	2.1
1	X	672	CYS	2.1
1	X	598	ILE	2.1
1	X	735	TYR	2.1
1	X	603	TYR	2.1
1	X	641	SER	2.1
1	X	680	MET	2.0

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, 95^{th} 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(Å^2)	Q<0.9
3	2HP	X	829	5/5	0.96	0.19	35,41,43,46	0
3	2HP	X	830	5/5	0.97	0.18	31,35,39,39	0
2	K	X	1	1/1	0.99	0.16	34,34,34,34	0

6.5 Other polymers (i)

There are no such residues in this entry.