Full wwPDB X-ray Structure Validation Report (i)

May 23, 2020 – 07:15 am BST

PDB ID	:	5IFC
Title	:	Crystal structure of polymerase acid protein (PA) from Influenza A virus, WILSON-SMITH/1933 (H1N1) bound to follow on fragment EBSI-4720 1-(4-bromophenyl)-1H-imidazole
Authors	:	Seattle Structural Genomics Center for Infectious Disease (SSGCID)
Deposited on	:	2016-02-25
Resolution	:	2.65 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:

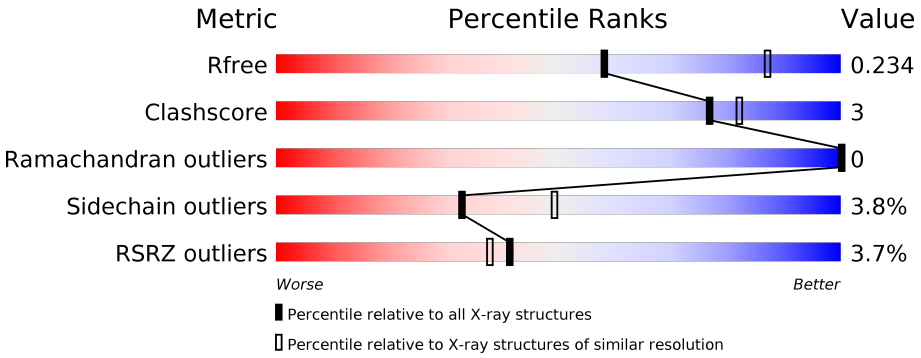
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, 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 2.65 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	1332 (2.68-2.64)
Clashscore	141614	1374 (2.68-2.64)
Ramachandran outliers	138981	1349 (2.68-2.64)
Sidechain outliers	138945	1349 (2.68-2.64)
RSRZ outliers	127900	1318 (2.68-2.64)

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	464	3% 79% 9% 12%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3296 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.

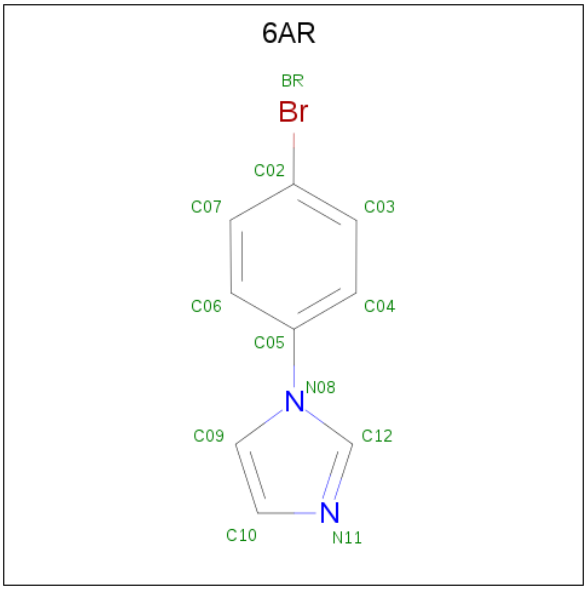
- Molecule 1 is a protein called Polymerase acidic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	410	Total	C	N	O	S	0	4	0
			3207	2045	549	588	25			

There is a discrepancy between the modelled and reference sequences:

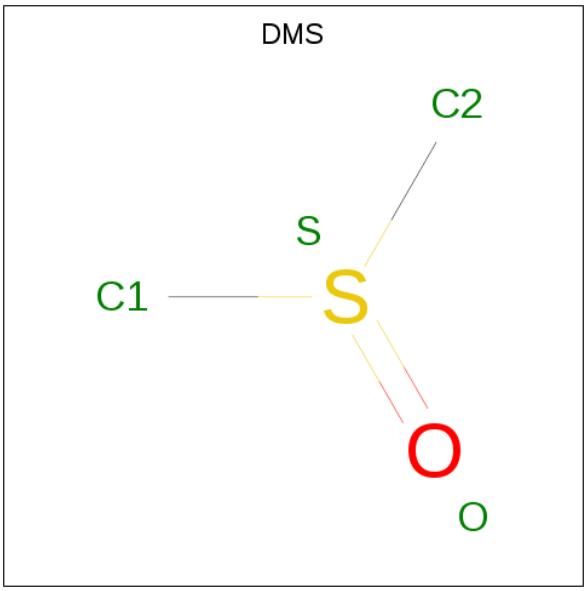
Chain	Residue	Modelled	Actual	Comment	Reference
A	253	GLY	-	expression tag	UNP P15659

- Molecule 2 is 1-(4-bromophenyl)-1H-imidazole (three-letter code: 6AR) (formula: $C_9H_7BrN_2$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	Br	C	N	0	0
			12	1	9	2		

- Molecule 3 is DIMETHYL SULFOXIDE (three-letter code: DMS) (formula: C_2H_6OS).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O S 4 2 1 1	0	0

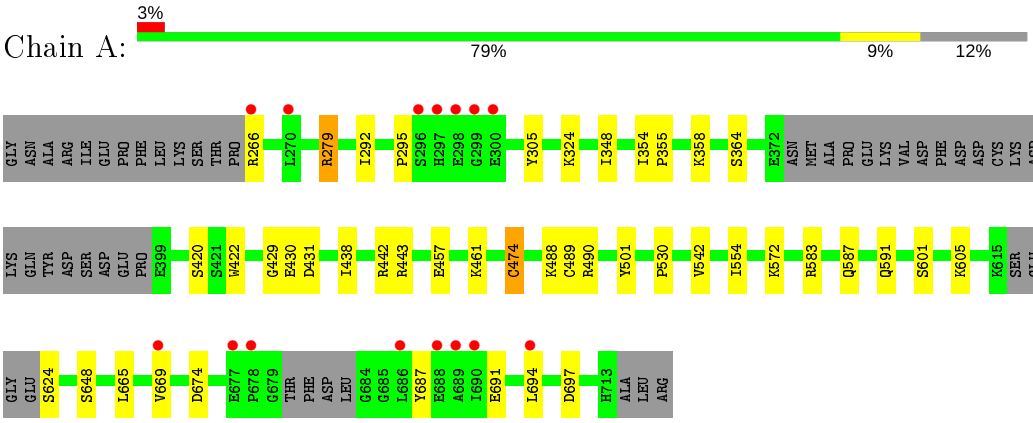
- Molecule 4 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	73	Total O 73 73	0	0

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: Polymerase acidic protein



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	69.08Å 69.08Å 397.97Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	41.21 – 2.65 41.21 – 2.65	Depositor EDS
% Data completeness (in resolution range)	99.8 (41.21-2.65) 99.8 (41.21-2.65)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.71 (at 2.65Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.178 , 0.234 0.178 , 0.234	Depositor DCC
R_{free} test set	865 reflections (4.94%)	wwPDB-VP
Wilson B-factor (Å^2)	42.3	Xtriage
Anisotropy	0.314	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 47.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	3296	wwPDB-VP
Average B, all atoms (Å^2)	48.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.55% 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: 6AR, DMS

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	0.41	1/3286 (0.0%)	0.55	0/4445

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	474	CYS	CB-SG	-5.68	1.72	1.81

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	A	3207	0	3155	19	0
2	A	12	0	0	0	0
3	A	4	0	6	0	0
4	A	73	0	0	0	0
All	All	3296	0	3161	19	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 3.

All (19) 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:A:305:TYR:OH	1:A:354:ILE:O	2.03	0.74
1:A:431:ASP:OD2	1:A:443[B]:ARG:NH1	2.26	0.69
1:A:279:ARG:NH1	1:A:697:ASP:OD2	2.35	0.59
1:A:530:PRO:HG3	1:A:542:VAL:HG11	1.90	0.54
1:A:348:ILE:HD11	1:A:358:LYS:HG2	1.92	0.52
1:A:292:ILE:HD12	1:A:295:PRO:HB3	1.93	0.51
1:A:430:GLU:OE2	1:A:583:ARG:NH2	2.43	0.51
1:A:422:TRP:CE2	1:A:457:GLU:HB2	2.46	0.50
1:A:438:ILE:O	1:A:442:ARG:HG3	2.15	0.47
1:A:587:GLN:O	1:A:591:GLN:HG3	2.15	0.47
1:A:665:LEU:O	1:A:669:VAL:HG23	2.16	0.45
1:A:691:GLU:HA	1:A:694:LEU:HG	1.98	0.45
1:A:461:LYS:HE3	1:A:501:TYR:CG	2.52	0.45
1:A:422:TRP:CE3	1:A:489:CYS:HB3	2.53	0.44
1:A:674:ASP:OD1	1:A:674:ASP:N	2.52	0.43
1:A:429:GLY:HA3	1:A:443[B]:ARG:HH11	1.85	0.42
1:A:266:ARG:HD3	1:A:687:TYR:CE2	2.55	0.42
1:A:348:ILE:HG23	1:A:355:PRO:HD2	2.03	0.41
1:A:422:TRP:CZ3	1:A:489:CYS:HB3	2.56	0.40

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	406/464 (88%)	397 (98%)	9 (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	344/413 (83%)	331 (96%)	13 (4%)	33 49

All (13) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	279	ARG
1	A	324	LYS
1	A	364	SER
1	A	420	SER
1	A	474	CYS
1	A	488	LYS
1	A	490	ARG
1	A	554	ILE
1	A	572	LYS
1	A	601	SER
1	A	605	LYS
1	A	624	SER
1	A	648	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	587	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 (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

2 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	$\#\|Z\| > 2$
3	DMS	A	802	-	3,3,3	0.64	0	3,3,3	0.59	0
2	6AR	A	801	-	13,13,13	3.08	4 (30%)	15,17,17	0.62	0

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
2	6AR	A	801	-	-	0/4/4/4	0/2/2/2

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	801	6AR	C05-N08	-8.45	1.31	1.45
2	A	801	6AR	C09-N08	-4.88	1.32	1.39
2	A	801	6AR	C12-N08	-4.34	1.31	1.36
2	A	801	6AR	BR-C02	2.64	1.95	1.90

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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

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	A	410/464 (88%)	-0.21	15 (3%) 41 38	25, 44, 90, 131	0

All (15) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	299	GLY	5.0
1	A	296	SER	4.2
1	A	686	LEU	3.3
1	A	689	ALA	3.3
1	A	690	ILE	3.2
1	A	297	HIS	2.8
1	A	300	GLU	2.6
1	A	270	LEU	2.5
1	A	688	GLU	2.5
1	A	678	PRO	2.4
1	A	298	GLU	2.4
1	A	677	GLU	2.4
1	A	266	ARG	2.3
1	A	669	VAL	2.1
1	A	694	LEU	2.0

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

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	DMS	A	802	4/4	0.92	0.40	99,103,105,107	0
2	6AR	A	801	12/12	0.97	0.17	43,49,60,88	0

6.5 Other polymers

There are no such residues in this entry.