Full wwPDB X-ray Structure Validation Report (i)

Jun 17, 2024 – 07:16 AM EDT

PDB ID	:	5IF2
Title	:	Crystal structure of polymerase acid protein (PA) from Influenza A virus, WILSON-SMITH/1933 (H1N1) bound to fragment hit EBSI-576 (5,6-dichloro-1H-1,3-benzodiazol-2-yl)methanol
Authors	:	Seattle Structural Genomics Center for Infectious Disease (SSGCID)
Deposited on	:	2016-02-25
Resolution	:	2.35 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

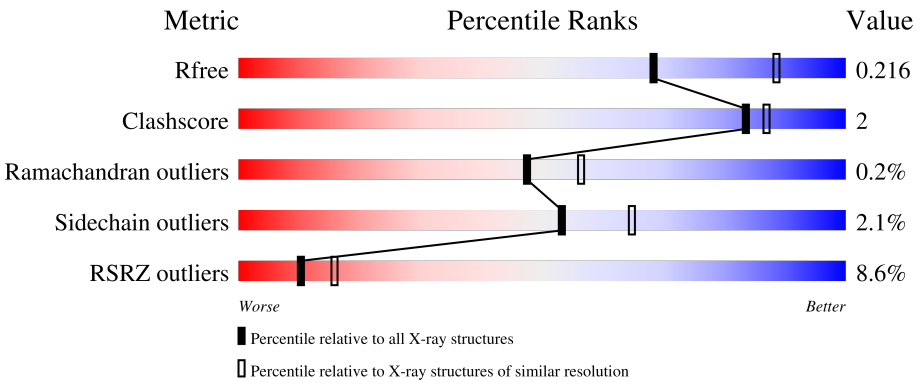
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 2.35 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	1164 (2.36-2.36)
Clashscore	141614	1232 (2.36-2.36)
Ramachandran outliers	138981	1211 (2.36-2.36)
Sidechain outliers	138945	1212 (2.36-2.36)
RSRZ outliers	127900	1150 (2.36-2.36)

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	464	8% 82% 5% 12%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3339 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.

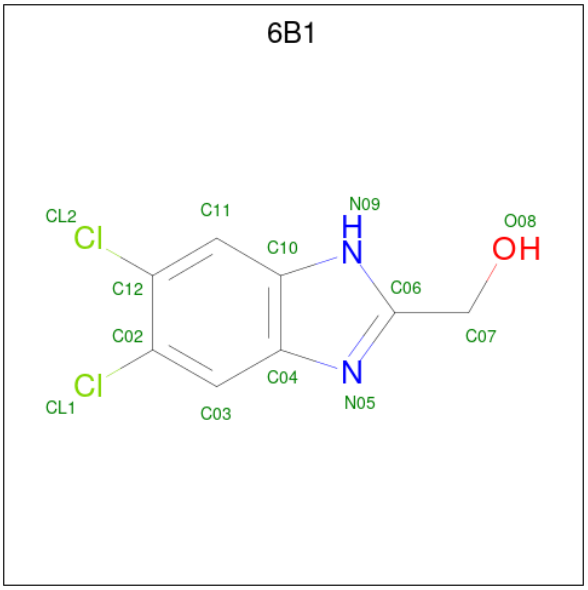
- Molecule 1 is a protein called Polymerase acidic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	409	3192	2030	543	594	25	0	4	0

There is a discrepancy between the modelled and reference sequences:

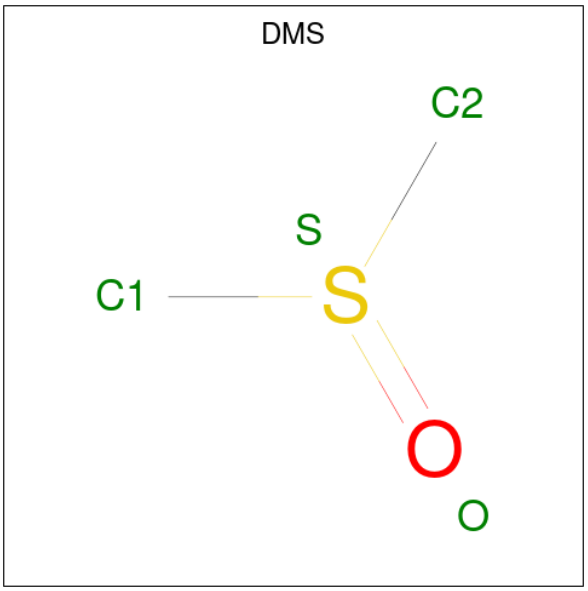
Chain	Residue	Modelled	Actual	Comment	Reference
A	253	GLY	-	expression tag	UNP P15659

- Molecule 2 is (5,6-dichloro-1H-benzimidazol-2-yl)methanol (three-letter code: 6B1) (formula: $C_8H_6Cl_2N_2O$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	Cl	N	O		
2	A	1	13	8	2	2	1	0	0

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



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

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
4	A	1	1	1	0	0

- Molecule 5 is water.

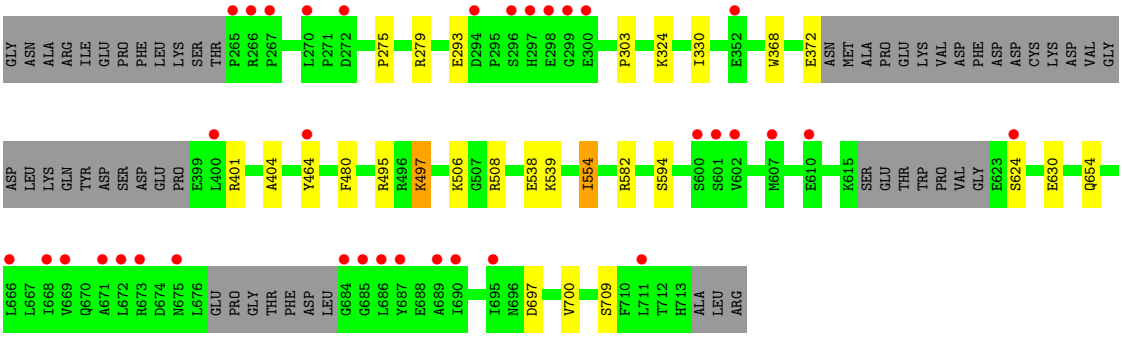
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	129	129	129	0	0

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: Polymerase acidic protein

Chain A:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	68.89Å 68.89Å 396.92Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	44.30 – 2.35 47.69 – 2.35	Depositor EDS
% Data completeness (in resolution range)	99.9 (44.30-2.35) 100.0 (47.69-2.35)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.59 (at 2.34Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.176 , 0.216 0.175 , 0.216	Depositor DCC
R_{free} test set	1949 reflections (7.91%)	wwPDB-VP
Wilson B-factor (Å^2)	42.2	Xtriage
Anisotropy	0.204	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 49.4	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	0.95	EDS
Total number of atoms	3339	wwPDB-VP
Average B, all atoms (Å^2)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.17% 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)

Bond lengths and bond angles in the following residue types are not validated in this section: DMS, 6B1, CL

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	0.39	0/3269	0.52	0/4424

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 (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	3192	0	3115	15	0
2	A	13	0	0	0	0
3	A	4	0	6	0	0
4	A	1	0	0	0	0
5	A	129	0	0	3	0
All	All	3339	0	3121	15	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 2.

All (15) 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:293:GLU:HG3	1:A:497:LYS:HD2	1.85	0.59
1:A:330:ILE:HD12	1:A:368:TRP:NE1	2.27	0.50
1:A:279:ARG:NH1	1:A:697:ASP:OD2	2.44	0.49
1:A:293:GLU:OE2	1:A:495:ARG:HD3	2.13	0.48
1:A:303:PRO:HB3	1:A:464[A]:TYR:OH	2.15	0.47
1:A:582:ARG:HB2	5:A:938:HOH:O	2.15	0.46
1:A:630:GLU:OE1	1:A:630:GLU:N	2.49	0.46
1:A:495:ARG:HD2	5:A:902:HOH:O	2.15	0.45
1:A:275:PRO:HB3	1:A:404:ALA:HB2	2.00	0.44
1:A:480:PHE:HA	1:A:506:LYS:O	2.18	0.44
1:A:654:GLN:HB3	1:A:700:VAL:HG21	1.99	0.44
1:A:464[A]:TYR:OH	5:A:901:HOH:O	2.16	0.43
1:A:401[B]:ARG:HH12	1:A:538:GLU:CD	2.23	0.42
1:A:372:GLU:OE2	1:A:508:ARG:HD2	2.21	0.40
1:A:401[B]:ARG:NH1	1:A:538:GLU:OE2	2.43	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	405/464 (87%)	394 (97%)	10 (2%)	1 (0%)	47 56

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	554	ILE

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	342/413 (83%)	335 (98%)	7 (2%)	55 66

All (7) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	324	LYS
1	A	497	LYS
1	A	539	LYS
1	A	554	ILE
1	A	594	SER
1	A	624	SER
1	A	709	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

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)

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	DMS	A	802	-	3,3,3	0.62	0	3,3,3	0.83	0
2	6B1	A	801	-	12,14,14	1.59	2 (16%)	11,20,20	0.88	1 (9%)

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	6B1	A	801	-	-	0/0/2/2	0/2/2/2

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	801	6B1	C12-CL2	2.90	1.80	1.73
2	A	801	6B1	C02-CL1	2.53	1.79	1.73

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	801	6B1	N09-C06-N05	-2.38	108.61	115.89

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 (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	409/464 (88%)	0.18	35 (8%) 10 16	22, 44, 90, 113	0

All (35) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	299	GLY	9.1
1	A	297	HIS	6.6
1	A	296	SER	5.3
1	A	686	LEU	4.6
1	A	672	LEU	4.4
1	A	298	GLU	4.4
1	A	602	VAL	4.3
1	A	624	SER	3.3
1	A	610	GLU	3.3
1	A	685	GLY	3.1
1	A	267	PRO	2.9
1	A	690	ILE	2.9
1	A	669	VAL	2.9
1	A	711	LEU	2.9
1	A	601	SER	2.7
1	A	294	ASP	2.7
1	A	675	ASN	2.6
1	A	272	ASP	2.6
1	A	687	TYR	2.5
1	A	265	PRO	2.5
1	A	300	GLU	2.5
1	A	666	LEU	2.5
1	A	689	ALA	2.4
1	A	352	GLU	2.3
1	A	464[A]	TYR	2.3
1	A	671	ALA	2.3
1	A	400	LEU	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	266	ARG	2.2
1	A	695	ILE	2.2
1	A	270	LEU	2.2
1	A	607	MET	2.1
1	A	668	ILE	2.0
1	A	673	ARG	2.0
1	A	684	GLY	2.0
1	A	600	SER	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 monosaccharides 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, 95th 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
2	6B1	A	801	13/13	0.62	0.35	59,76,84,89	13
4	CL	A	803	1/1	0.88	0.36	85,85,85,85	0
3	DMS	A	802	4/4	0.92	0.32	68,69,75,75	0

6.5 Other polymers (i)

There are no such residues in this entry.