



# Full wwPDB X-ray Structure Validation Report

Feb 28, 2014 – 10:34 PM GMT

PDB ID : 4J2R  
Title : Middle domain of influenza A virus RNA-dependent polymerase PB2  
Authors : Qiu, H.; Tsurumura, T.; Tsumori, Y.; Hatakeyama, D.; Kuzuhara, T.; Tsuge, H.  
Deposited on : 2013-02-05  
Resolution : 2.42 Å(reported)

This is a full wwPDB validation report for a publicly released PDB entry.  
We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)  
A user guide is available at <http://wwpdb.org/ValidationPDFNotes.html>

---

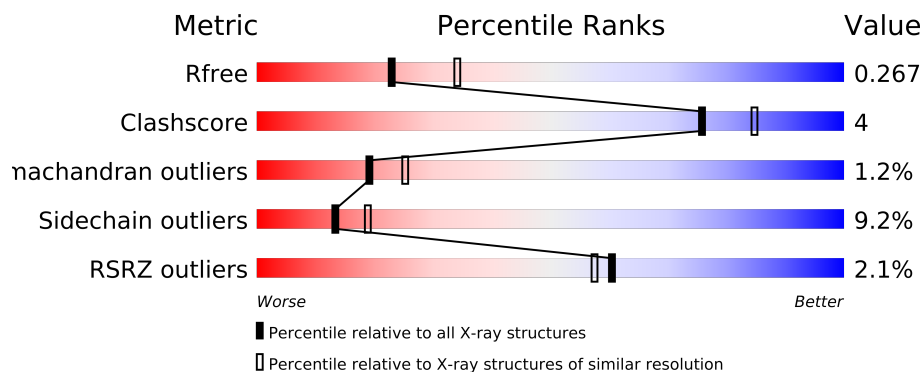
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

The reported resolution of this entry is 2.42 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))
$R_{free}$	66092	2496 (2.44-2.40)
Clashscore	79885	3124 (2.44-2.40)
Ramachandran outliers	78287	3067 (2.44-2.40)
Sidechain outliers	78261	3068 (2.44-2.40)
RSRZ outliers	66119	2499 (2.44-2.40)

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  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	174	
1	B	174	

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2653 atoms, of which 0 are hydrogen and 0 are deuterium.

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 basic protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	163	Total	C	N	O	S	0	0	0
			1291	809	238	234	10			
1	B	163	Total	C	N	O	S	0	0	0
			1291	809	238	234	10			

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	314	GLY	-	EXPRESSION TAG	UNP P03428
A	315	SER	-	EXPRESSION TAG	UNP P03428
A	316	HIS	-	EXPRESSION TAG	UNP P03428
A	317	MET	-	EXPRESSION TAG	UNP P03428
A	485	SER	-	EXPRESSION TAG	UNP P03428
A	486	GLY	-	EXPRESSION TAG	UNP P03428
A	487	CYS	-	EXPRESSION TAG	UNP P03428
B	314	GLY	-	EXPRESSION TAG	UNP P03428
B	315	SER	-	EXPRESSION TAG	UNP P03428
B	316	HIS	-	EXPRESSION TAG	UNP P03428
B	317	MET	-	EXPRESSION TAG	UNP P03428
B	485	SER	-	EXPRESSION TAG	UNP P03428
B	486	GLY	-	EXPRESSION TAG	UNP P03428
B	487	CYS	-	EXPRESSION TAG	UNP P03428

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	33	Total	O	0	0
			33	33		
2	B	38	Total	O	0	0
			38	38		

### 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 errors 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 basic protein 2

Chain A: 



- Molecule 1: Polymerase basic protein 2

Chain B: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	I 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	75.97Å 103.51Å 112.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	25.88 – 2.42 25.88 – 2.42	Depositor EDS
% Data completeness (in resolution range)	92.5 (25.88-2.42) 92.6 (25.88-2.42)	Depositor EDS
$R_{merge}$	0.04	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	5.22 (at 2.41Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, $R_{free}$	0.215 , 0.271 0.217 , 0.267	Depositor DCC
$R_{free}$ test set	803 reflections (5.27%)	DCC
Wilson B-factor (Å <sup>2</sup> )	51.1	Xtriage
Anisotropy	0.346	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 37.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Outliers	0 of 16044 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	2653	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.00% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

## 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	0.78	0/1309	0.92	0/1754
1	B	0.72	0/1309	0.85	0/1754
All	All	0.75	0/2618	0.88	0/3508

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	1291	0	0	7	0
1	B	1291	0	0	4	0
2	A	33	0	0	5	0
2	B	38	0	0	2	0
All	All	2653	0	0	10	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 4.

All (10) close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:351:THR:CG2	2:A:533:HOH:O	2.51	0.59

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:332:ARG:NH2	2:A:528:HOH:O	2.37	0.57
1:A:427:ARG:CG	2:A:522:HOH:O	2.56	0.54
1:A:439:GLN:NE2	2:A:503:HOH:O	2.43	0.51
1:A:379:ARG:N	2:A:514:HOH:O	2.46	0.48
1:B:334:SER:OG	2:B:538:HOH:O	2.20	0.47
1:A:456:ASN:ND2	1:B:459:GLY:O	2.50	0.45
1:A:333:THR:N	1:A:362:GLU:O	2.50	0.44
1:B:421:VAL:N	2:B:534:HOH:O	2.51	0.44
1:B:382:ILE:CD1	1:B:382:ILE:C	2.86	0.44

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone ⓘ

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	161/174 (92%)	152 (94%)	8 (5%)	1 (1%)	33	46
1	B	161/174 (92%)	153 (95%)	5 (3%)	3 (2%)	12	14
All	All	322/348 (92%)	305 (95%)	13 (4%)	4 (1%)	19	25

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	322	SER
1	B	378	THR
1	A	334	SER
1	B	455	ASP

### 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	142/150 (95%)	131 (92%)	11 (8%)	18	28
1	B	142/150 (95%)	127 (89%)	15 (11%)	10	13
All	All	284/300 (95%)	258 (91%)	26 (9%)	13	19

All (26) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	336	SER
1	A	338	VAL
1	A	345	LEU
1	A	386	VAL
1	A	387	SER
1	A	392	GLN
1	A	408	ASP
1	A	422	ASN
1	A	443	LYS
1	A	473	MET
1	A	482	LYS
1	B	320	SER
1	B	321	SER
1	B	341	GLU
1	B	353	LYS
1	B	354	ILE
1	B	382	ILE
1	B	387	SER
1	B	389	ARG
1	B	390	ASP
1	B	421	VAL
1	B	422	ASN
1	B	440	LYS
1	B	443	LYS
1	B	455	ASP
1	B	461	ILE

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

### 5.3.3 RNA ⓘ

There are no RNA chains 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	163/174 (93%)	-0.09	3 (1%) 65 62	40, 54, 92, 110	1 (0%)
1	B	163/174 (93%)	-0.16	4 (2%) 54 52	34, 52, 84, 102	1 (0%)
All	All	326/348 (93%)	-0.13	7 (2%) 60 57	34, 53, 90, 110	2 (0%)

All (7) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	320	SER	6.6
1	B	320	SER	5.4
1	A	321	SER	4.5
1	A	457	VAL	3.5
1	B	455	ASP	2.9
1	B	321	SER	2.6
1	B	423	ARG	2.5

### 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.