



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:24 PM GMT

PDB ID : 3ZDP  
Title : R416A Monomeric nucleoprotein of influenza A virus  
Authors : Chenavas, S.; Ruigrok, R.W.H.; Crepin, T.  
Deposited on : 2012-11-29  
Resolution : 2.69 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Mogul : 1.7 (RC4), CSD as536be (2015)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20026688  
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)  
Refmac : 5.8.0135  
CCP4 : 6.5.0  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : trunk26865

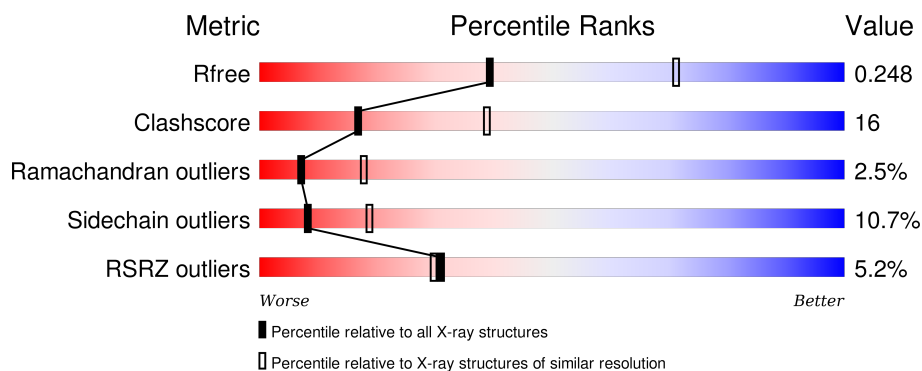
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.69 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))
$R_{free}$	91344	2103 (2.70-2.70)
Clashscore	102246	2422 (2.70-2.70)
Ramachandran outliers	100387	2382 (2.70-2.70)
Sidechain outliers	100360	2382 (2.70-2.70)
RSRZ outliers	91569	2107 (2.70-2.70)

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. 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  $\leq 5\%$ . 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	498	 6% 58% 28% 7% 7%
1	B	498	 5% 65% 24% • 6%
1	C	498	 4% 63% 25% • • 7%

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 11113 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 NUCLEOCAPSID PROTEIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	462	Total	C	N	O	S	0	0	0
			3670	2275	681	688	26			
1	B	467	Total	C	N	O	S	0	0	0
			3703	2294	688	695	26			
1	C	462	Total	C	N	O	S	0	0	0
			3677	2279	683	689	26			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	416	ALA	ARG	ENGINEERED MUTATION	UNP Q1K9H2
B	416	ALA	ARG	ENGINEERED MUTATION	UNP Q1K9H2
C	416	ALA	ARG	ENGINEERED MUTATION	UNP Q1K9H2

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

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	B	1	Total	K	0	0
			1	1		
2	A	1	Total	K	0	0
			1	1		
2	C	1	Total	K	0	0
			1	1		

- Molecule 3 is water.

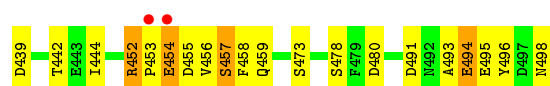
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	22	Total	O	0	0
			22	22		
3	B	21	Total	O	0	0
			21	21		

*Continued on next page...*

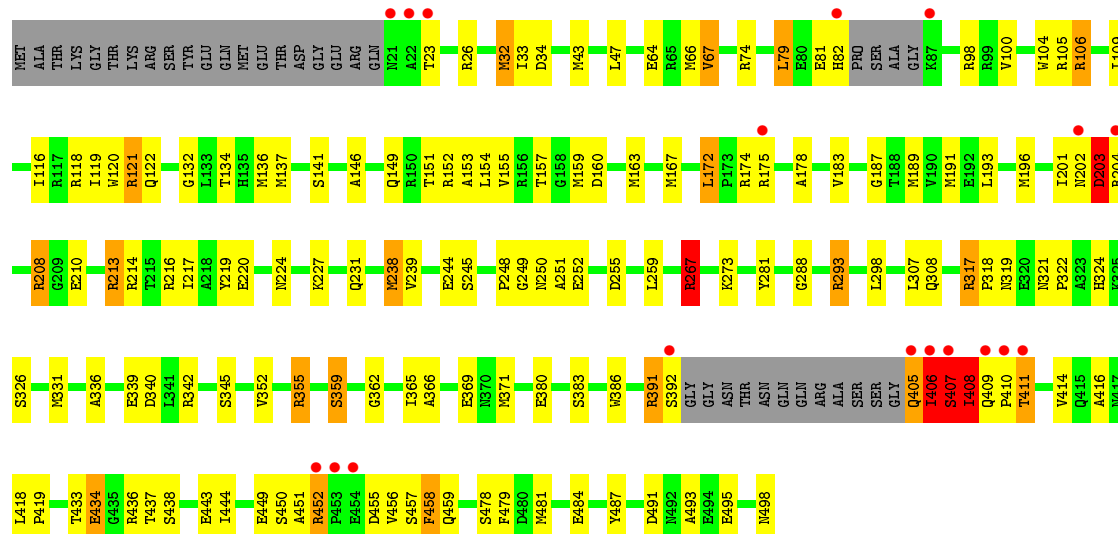
*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	C	17	Total	O	0	0
			17	17		





● Molecule 1: NUCLEOCAPSID PROTEIN



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 2 2 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	165.42Å 285.44Å 118.31Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	143.12 – 2.69 39.89 – 2.69	Depositor EDS
% Data completeness (in resolution range)	99.4 (143.12-2.69) 99.4 (39.89-2.69)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.98 (at 2.69Å)	Xtriage
Refinement program	REFMAC 5.6.0116	Depositor
R, $R_{free}$	0.198 , 0.250 0.197 , 0.248	Depositor DCC
$R_{free}$ test set	3914 reflections (5.32%)	DCC
Wilson B-factor (Å <sup>2</sup> )	46.8	Xtriage
Anisotropy	0.048	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 37.8	EDS
Estimated twinning fraction	0.015 for 1/2*h-1/2*k,-3/2*h-1/2*k,-l 0.018 for 1/2*h+1/2*k,3/2*h-1/2*k,-l	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Outliers	1 of 77393 reflections (0.001%)	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	11113	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.62% of the height of the origin peak. No significant pseudotranslation is detected.*

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

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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:  
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	A	0.76	0/3734	0.84	4/5022 (0.1%)
1	B	0.84	3/3768 (0.1%)	0.88	3/5070 (0.1%)
1	C	0.83	3/3740 (0.1%)	0.90	8/5030 (0.2%)
All	All	0.81	6/11242 (0.1%)	0.87	15/15122 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	B	0	2
1	C	0	3
All	All	0	5

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	120	TRP	CD2-CE2	6.80	1.49	1.41
1	B	104	TRP	CD2-CE2	6.22	1.48	1.41
1	C	386	TRP	CD2-CE2	5.68	1.48	1.41
1	B	120	TRP	CD2-CE2	5.63	1.48	1.41
1	B	393	GLY	N-CA	5.10	1.53	1.46

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	355	ARG	NE-CZ-NH2	-8.90	115.85	120.30
1	B	167	MET	CG-SD-CE	-7.41	88.34	100.20
1	C	267	ARG	NE-CZ-NH1	7.27	123.94	120.30

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	355	ARG	NE-CZ-NH1	6.82	123.71	120.30
1	C	317	ARG	NE-CZ-NH1	6.22	123.41	120.30

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	210	GLU	Peptide
1	B	411	THR	Peptide
1	C	288	GLY	Peptide
1	C	411	THR	Peptide
1	C	450	SER	Peptide

## 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	3670	0	3630	132	0
1	B	3703	0	3661	95	0
1	C	3677	0	3637	137	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
3	A	22	0	0	1	0
3	B	21	0	0	1	0
3	C	17	0	0	0	0
All	All	11113	0	10928	361	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 16.

The worst 5 of 361 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:C:355:ARG:HD3	1:C:493:ALA:HB2	1.26	1.13

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:208:ARG:HH11	1:C:208:ARG:HG2	0.98	1.13
1:C:324:HIS:CD2	1:C:359:SER:H	1.67	1.12
1:B:78:TYR:HA	1:B:137:MET:CE	1.80	1.10
1:A:369:GLU:C	1:A:370:ASN:CA	2.19	1.08

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	456/498 (92%)	398 (87%)	40 (9%)	18 (4%)	4	8
1	B	463/498 (93%)	423 (91%)	32 (7%)	8 (2%)	11	29
1	C	456/498 (92%)	419 (92%)	28 (6%)	9 (2%)	9	24
All	All	1375/1494 (92%)	1240 (90%)	100 (7%)	35 (2%)	7	18

5 of 35 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	410	PRO
1	A	455	ASP
1	B	87	LYS
1	B	211	ASN
1	B	408	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	391/421 (93%)	345 (88%)	46 (12%)	6	15
1	B	395/421 (94%)	352 (89%)	43 (11%)	8	18
1	C	393/421 (93%)	356 (91%)	37 (9%)	11	25
All	All	1179/1263 (93%)	1053 (89%)	126 (11%)	8	19

5 of 126 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	174	ARG
1	B	321	ASN
1	C	407	SER
1	B	199	ARG
1	B	217	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 29 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	272	HIS
1	B	324	HIS
1	C	324	HIS
1	B	308	GLN
1	B	409	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 ⓘ

There are no carbohydrates in this entry.

## 5.6 Ligand geometry

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

There are no bond length outliers.

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

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 [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, 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	462/498 (92%)	-0.17	29 (6%)	23 22	24, 45, 116, 153	0
1	B	467/498 (93%)	-0.24	26 (5%)	28 26	23, 40, 99, 156	0
1	C	462/498 (92%)	-0.25	18 (3%)	43 43	24, 41, 98, 161	0
All	All	1391/1494 (93%)	-0.22	73 (5%)	31 30	23, 42, 104, 161	0

The worst 5 of 73 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	453	PRO	7.7
1	A	453	PRO	7.2
1	C	22	ALA	6.0
1	A	86	GLY	5.9
1	C	407	SER	5.6

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	K	A	1499	1/1	0.95	0.08	-1.80	53,53,53,53	0
2	K	B	1499	1/1	0.99	0.08	-2.45	48,48,48,48	0
2	K	C	1499	1/1	0.97	0.06	-3.26	49,49,49,49	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.