wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:14 AM GMT

PDB ID	:	2Q06
Title	:	Crystal structure of Influenza A Virus H5N1 Nucleoprotein
Authors	:	Ng, A.K.L.; Zhang, H.; Tan, K.; Wang, J.; Shaw, P.C.
Deposited on	:	2007-05-21
Resolution	:	3.30 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
A user guide is available at
http://wwpdb.org/validation/2016/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:

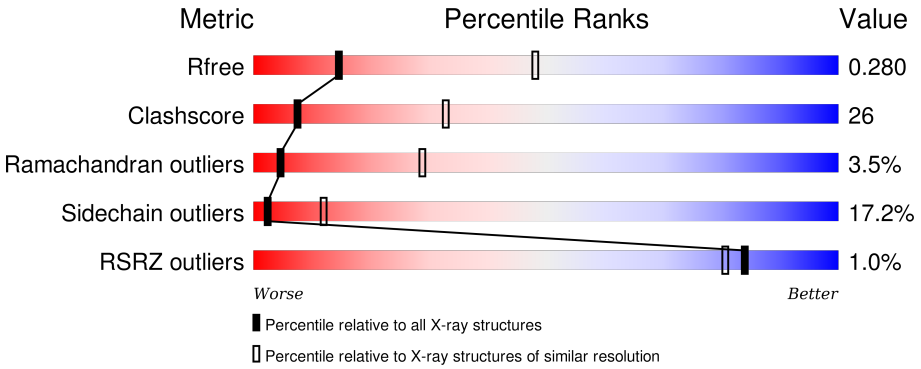
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.30 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	2060 (3.40-3.20)
Clashscore	102246	1058 (3.38-3.22)
Ramachandran outliers	100387	1038 (3.38-3.22)
Sidechain outliers	100360	1037 (3.38-3.22)
RSRZ outliers	91569	2070 (3.40-3.20)

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. 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	504	% 48% 35% 9% • 7%
1	B	504	% 47% 36% 9% • 8%

2 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 7343 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 Nucleoprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	467	Total	C	N	O	S	0	0	0
			3678	2269	696	686	27			
1	B	465	Total	C	N	O	S	0	0	0
			3665	2260	695	683	27			

There are 12 discrepancies between the modelled and reference sequences:

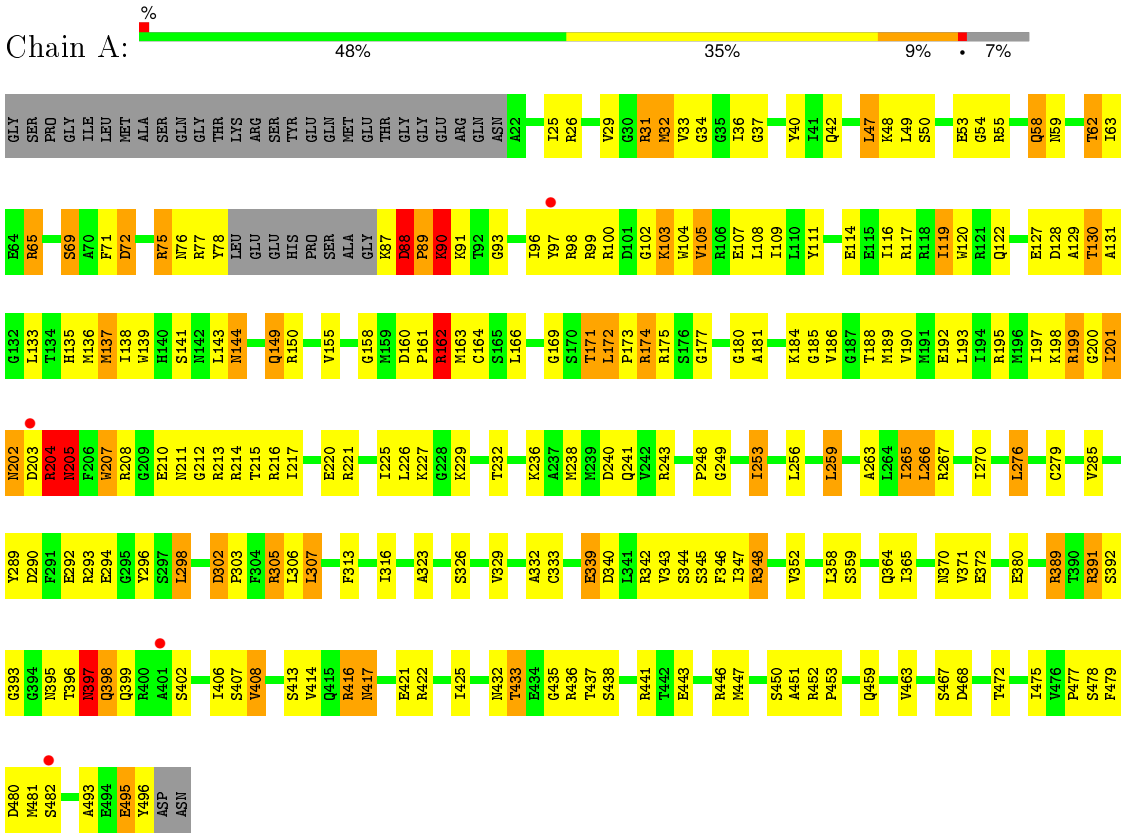
Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	GLY	-	EXPRESSION TAG	UNP Q9PX50
A	-4	SER	-	EXPRESSION TAG	UNP Q9PX50
A	-3	PRO	-	EXPRESSION TAG	UNP Q9PX50
A	-2	GLY	-	EXPRESSION TAG	UNP Q9PX50
A	-1	ILE	-	EXPRESSION TAG	UNP Q9PX50
A	0	LEU	-	EXPRESSION TAG	UNP Q9PX50
B	-5	GLY	-	EXPRESSION TAG	UNP Q9PX50
B	-4	SER	-	EXPRESSION TAG	UNP Q9PX50
B	-3	PRO	-	EXPRESSION TAG	UNP Q9PX50
B	-2	GLY	-	EXPRESSION TAG	UNP Q9PX50
B	-1	ILE	-	EXPRESSION TAG	UNP Q9PX50
B	0	LEU	-	EXPRESSION TAG	UNP Q9PX50

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 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.

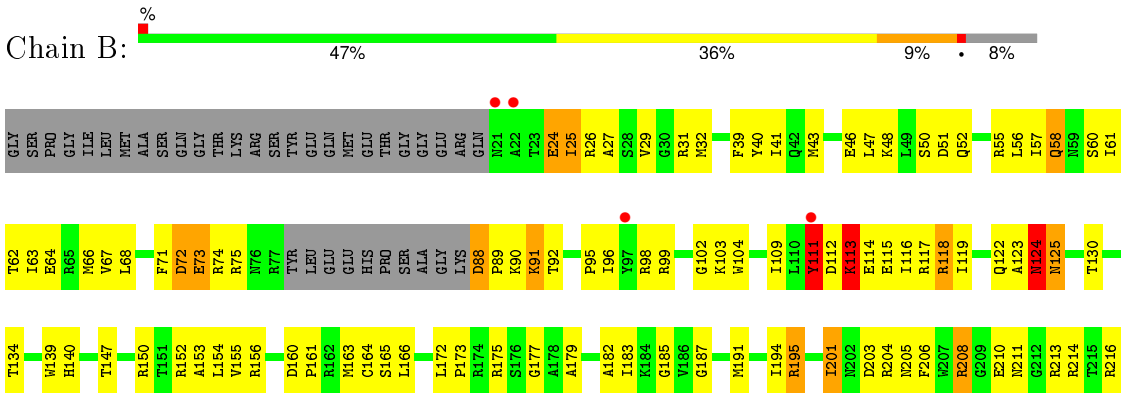
- Molecule 1: Nucleoprotein

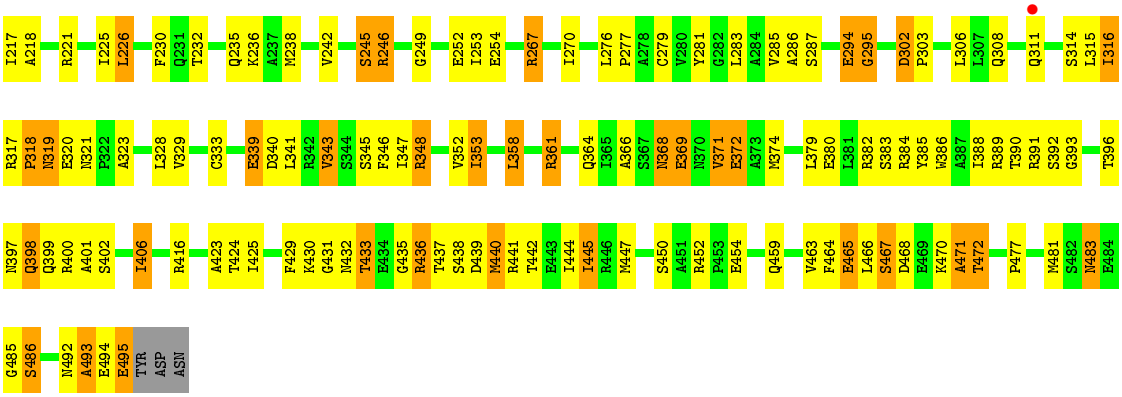
Chain A:



- Molecule 1: Nucleoprotein

Chain B:




I217 A218 R221 I225 L226 F230 Q231 T232 Q235 K236 A237 M238 V242 S245 R246 G249 E252 I253 E254 R267 I270 L276 P277 A278 C279 V280 Y281 C282 L283 A284 V285 A286 S287 E294 G295 D302 P303 L306 L307 Q308 Q311 S314 L315 I316
R317 P318 N319 E320 N321 F322 A323 L328 V329 C333 E339 D340 L341 R342 V343 S344 S345 F346 I347 R348 V352 I353 L358 R361 Q364 I365 A366 S367 N368 E369 N370 V371 E372 A373 M374 L379 E380 L381 R382 S383 R384 Y385 W386 A387 I388 R389 T390 R391 S392 G393 T396
N397 Q398 Q399 R400 A401 S402 I406 R416 A423 T424 I425 F429 K430 G431 N432 T433 E434 G435 R436 T437 S438 D439 M440 R441 T442 E443 I444 I445 R446 M447 S450 A451 R452 P453 E454 Q459 V463 F464 E465 L466 S467 D468 E469 K470 A471 T472 P477 M481 S482 N483 E484
G485 S486 N492 A493 E494 E495 TYR ASP ASN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, α, β, γ	153.58Å 153.58Å 153.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.32 – 3.30 46.31 – 3.30	Depositor EDS
% Data completeness (in resolution range)	99.7 (46.32-3.30) 99.7 (46.31-3.30)	Depositor EDS
R_{merge}	0.14	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.43 (at 3.32Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.202 , 0.279 0.203 , 0.280	Depositor DCC
R_{free} test set	945 reflections (5.42%)	DCC
Wilson B-factor (Å^2)	83.9	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 76.0	EDS
Estimated twinning fraction	0.041 for l,-k,h	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	1 of 18379 reflections (0.005%)	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	7343	wwPDB-VP
Average B, all atoms (Å^2)	93.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.32% 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, 0.375 respectively for untwinned datasets, and 0.333, 0.2 for perfectly twinned datasets.

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	0.45	0/3737	0.64	0/5024
1	B	0.46	0/3723	0.62	0/5003
All	All	0.45	0/7460	0.63	0/10027

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	A	0	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	31	ARG	Sidechain
1	A	479	PHE	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	3678	0	3639	180	0
1	B	3665	0	3639	197	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	7343	0	7278	373	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 26.

The worst 5 of 373 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:204:ARG:HA	1:A:204:ARG:HE	1.20	1.06
1:A:90:LYS:HD3	1:A:90:LYS:H	1.18	1.02
1:A:103:LYS:HD2	1:A:103:LYS:H	1.24	0.99
1:B:152:ARG:O	1:B:156:ARG:HG2	1.69	0.92
1:B:62:THR:HG22	1:B:66:MET:HE3	1.51	0.91

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	463/504 (92%)	389 (84%)	58 (12%)	16 (4%)	4	29
1	B	461/504 (92%)	389 (84%)	56 (12%)	16 (4%)	4	29
All	All	924/1008 (92%)	778 (84%)	114 (12%)	32 (4%)	4	29

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	398	GLN
1	A	493	ALA
1	B	372	GLU
1	A	88	ASP

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	89	PRO

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	387/420 (92%)	321 (83%)	66 (17%)	2 12
1	B	387/420 (92%)	320 (83%)	67 (17%)	2 11
All	All	774/840 (92%)	641 (83%)	133 (17%)	2 12

5 of 133 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	472	THR
1	B	72	ASP
1	B	439	ASP
1	A	478	SER
1	B	25	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	124	ASN
1	B	149	GLN
1	B	399	GLN
1	B	140	HIS
1	B	144	ASN

5.3.3 RNA (i)

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

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^{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	467/504 (92%)	-0.10	4 (0%)	85	82	31, 93, 109, 119	0
1	B	465/504 (92%)	-0.07	5 (1%)	82	78	73, 92, 106, 116	0
All	All	932/1008 (92%)	-0.09	9 (0%)	84	80	31, 93, 109, 119	0

The worst 5 of 9 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	21	ASN	10.5
1	B	97	TYR	2.5
1	A	401	ALA	2.4
1	B	22	ALA	2.3
1	A	482	SER	2.3

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.