



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 06:04 PM GMT

PDB ID : 4KDM  
Title : Crystal structure of the hemagglutinin of ferret-transmissible H5N1 virus  
Authors : Lu, X.; Shi, Y.; Zhang, W.; Zhang, Y.; Qi, J.; Gao, G.F.  
Deposited on : 2013-04-25  
Resolution : 2.50 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

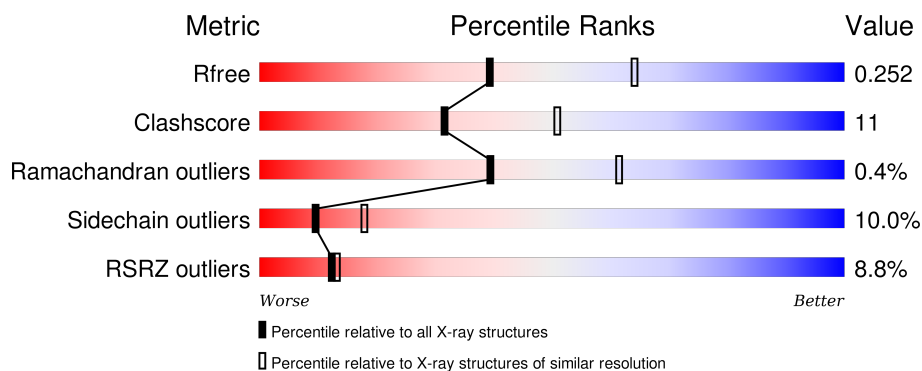
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.50 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	3553 (2.50-2.50)
Clashscore	102246	4242 (2.50-2.50)
Ramachandran outliers	100387	4156 (2.50-2.50)
Sidechain outliers	100360	4158 (2.50-2.50)
RSRZ outliers	91569	3562 (2.50-2.50)

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	322	<div> <div>72%</div> <div>23%</div> <div>5%</div> </div>
1	C	322	<div> <div>4%</div> <div>70%</div> <div>25%</div> <div>6%</div> </div>
1	E	322	<div> <div>5%</div> <div>66%</div> <div>29%</div> <div>5%</div> </div>
2	B	175	<div> <div>5%</div> <div>82%</div> <div>14%</div> <div>•</div> </div>
2	D	175	<div> <div>21%</div> <div>75%</div> <div>23%</div> <div>•</div> </div>

*Continued on next page...*

*Continued from previous page...*

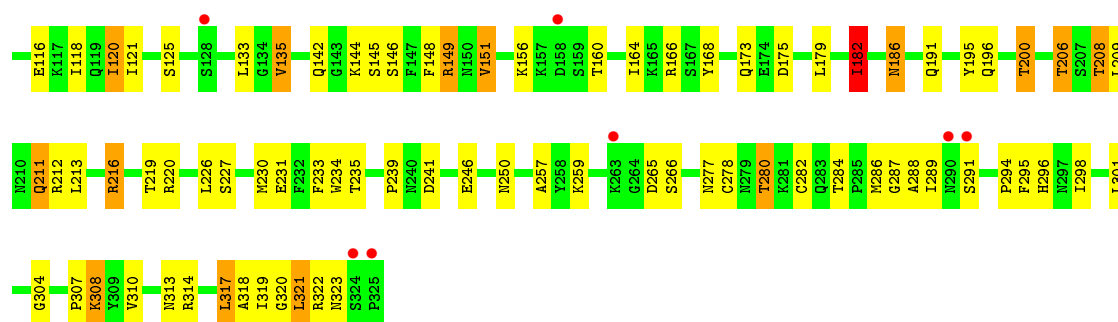
Mol	Chain	Length	Quality of chain
2	F	175	 A horizontal bar chart showing the quality of chain F. The bar is divided into four segments: red (30%), green (70%), yellow (26%), and orange (1%). The segments are labeled with their respective percentages: 30%, 70%, 26%, and 1%.



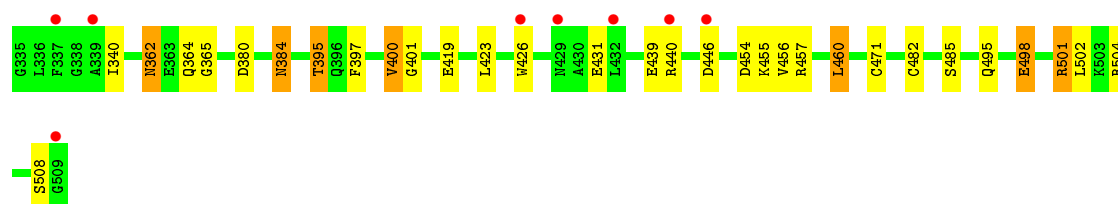
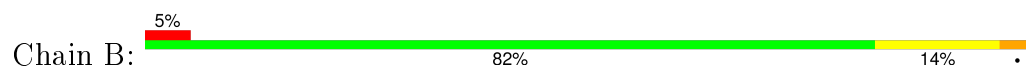


Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	53	Total 53	O 53	0	0
5	B	13	Total 13	O 13	0	0
5	C	42	Total 42	O 42	0	0
5	D	17	Total 17	O 17	0	0
5	E	33	Total 33	O 33	0	0
5	F	12	Total 12	O 12	0	0

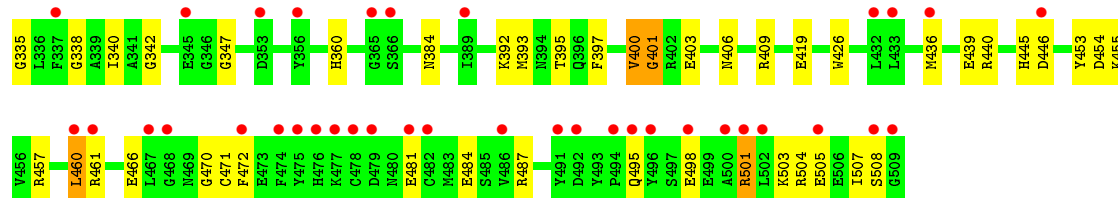
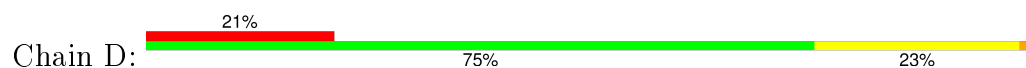




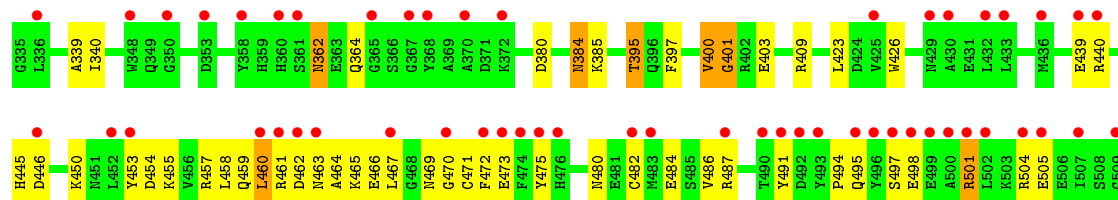
• Molecule 2: Hemagglutinin



• Molecule 2: Hemagglutinin



• Molecule 2: Hemagglutinin



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	66.76Å 243.85Å 71.61Å 90.00° 110.01° 90.00°	Depositor
Resolution (Å)	35.07 – 2.50 35.61 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.4 (35.07-2.50) 99.5 (35.61-2.50)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.60 (at 2.51Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8_1069)	Depositor
R, $R_{free}$	0.215 , 0.251 0.213 , 0.252	Depositor DCC
$R_{free}$ test set	3718 reflections (5.03%)	DCC
Wilson B-factor (Å <sup>2</sup> )	54.8	Xtriage
Anisotropy	0.293	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 43.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtriage
Outliers	0 of 73951 reflections	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	12221	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.41% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section:  
NAG

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	0.30	0/2621	0.55	0/3558
1	C	0.29	0/2621	0.53	0/3558
1	E	0.29	0/2621	0.55	1/3558 (0.0%)
2	B	0.37	0/1443	0.51	1/1939 (0.1%)
2	D	0.37	0/1443	0.48	1/1939 (0.1%)
2	F	0.36	0/1443	0.50	0/1939
All	All	0.32	0/12192	0.53	3/16491 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	460	LEU	CB-CA-C	7.13	123.75	110.20
1	E	182	ILE	CB-CA-C	-5.44	100.72	111.60
2	D	460	LEU	CB-CA-C	5.20	120.09	110.20

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	2559	0	2513	61	0
1	C	2559	0	2513	60	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	E	2559	0	2513	69	1
2	B	1416	0	1319	23	0
2	D	1416	0	1319	33	0
2	F	1416	0	1319	42	0
3	A	14	0	13	0	0
3	C	14	0	13	0	0
3	E	14	0	13	0	0
4	A	28	0	25	1	0
4	C	28	0	25	0	0
4	E	28	0	25	0	0
5	A	53	0	0	8	0
5	B	13	0	0	2	0
5	C	42	0	0	4	0
5	D	17	0	0	6	0
5	E	33	0	0	6	0
5	F	12	0	0	2	0
All	All	12221	0	11610	253	1

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 11.

The worst 5 of 253 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:E:206:THR:HG22	1:E:208:THR:H	1.37	0.89
4:A:602:NAG:O3	5:A:753:HOH:O	1.88	0.89
1:E:58:ASP:OD1	5:E:726:HOH:O	1.90	0.88
1:E:38:GLU:OE1	5:E:725:HOH:O	1.92	0.87
1:C:138:ALA:O	5:C:705:HOH:O	1.94	0.84

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:141:TYR:O	1:E:144:LYS:NZ[1_655]	2.14	0.06

## 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	320/322 (99%)	297 (93%)	22 (7%)	1 (0%)	46	68
1	C	320/322 (99%)	301 (94%)	18 (6%)	1 (0%)	46	68
1	E	320/322 (99%)	299 (93%)	20 (6%)	1 (0%)	46	68
2	B	173/175 (99%)	164 (95%)	8 (5%)	1 (1%)	30	50
2	D	173/175 (99%)	165 (95%)	7 (4%)	1 (1%)	30	50
2	F	173/175 (99%)	166 (96%)	6 (4%)	1 (1%)	30	50
All	All	1479/1491 (99%)	1392 (94%)	81 (6%)	6 (0%)	39	61

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	277	ASN
1	C	277	ASN
1	E	277	ASN
2	B	401	GLY
2	D	401	GLY

### 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	290/290 (100%)	258 (89%)	32 (11%)	8	14
1	C	290/290 (100%)	257 (89%)	33 (11%)	7	13
1	E	290/290 (100%)	256 (88%)	34 (12%)	7	12

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	B	149/149 (100%)	138 (93%)	11 (7%)	17	31
2	D	149/149 (100%)	140 (94%)	9 (6%)	24	43
2	F	149/149 (100%)	136 (91%)	13 (9%)	13	24
All	All	1317/1317 (100%)	1185 (90%)	132 (10%)	9	18

5 of 132 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	191	GLN
1	C	317	LEU
2	F	395	THR
1	C	200	THR
1	C	227	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
2	B	364	GLN
1	C	248	ASN
2	D	360	HIS
2	F	364	GLN
2	F	429	ASN

### 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 ⓘ

6 carbohydrates 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





## 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	322/322 (100%)	0.17	4 (1%) 81 83	20, 44, 72, 119	0
1	C	322/322 (100%)	0.27	14 (4%) 39 44	23, 51, 89, 121	0
1	E	322/322 (100%)	0.38	15 (4%) 35 40	30, 56, 97, 143	0
2	B	175/175 (100%)	0.54	8 (4%) 36 41	30, 57, 88, 140	0
2	D	175/175 (100%)	1.15	37 (21%) 1 1	25, 75, 123, 191	0
2	F	175/175 (100%)	1.38	53 (30%) 1 0	29, 84, 147, 196	0
All	All	1491/1491 (100%)	0.54	131 (8%) 12 13	20, 54, 115, 196	0

The worst 5 of 131 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	509	GLY	8.7
1	C	325	PRO	8.0
2	B	509	GLY	7.3
2	D	353	ASP	6.6
2	F	475	TYR	6.3

### 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](#)

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
4	NAG	E	602	14/15	0.93	0.26	1.88	46,62,71,78	0
4	NAG	C	602	14/15	0.91	0.15	-0.28	47,71,85,85	0
4	NAG	A	602	14/15	0.95	0.10	-	28,43,53,55	0
4	NAG	A	603	14/15	0.81	0.21	-	52,56,66,72	0
4	NAG	E	603	14/15	0.73	0.51	-	73,89,97,99	0
4	NAG	C	603	14/15	0.88	0.36	-	80,89,92,96	0

## 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
3	NAG	C	601	14/15	0.83	0.20	-	79,85,90,90	0
3	NAG	A	601	14/15	0.83	0.24	-	78,82,87,87	0
3	NAG	E	601	14/15	0.80	0.33	-	85,92,96,99	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.