wwPDB X-ray Structure Validation Summary Report (i)

Aug 30, 2023 – 08:10 AM EDT

PDB ID	:	3M5R
Title	:	Crystal Structure of Swine Flu Virus NS1 Effector Domain from H1N1 Influenza A/California/07/2009
Authors	:	Fremont, D.H.; Yu, Y.Y.L.; Nelson, C.A.; Center for Structural Genomics of Infectious Diseases (CSGID)
Deposited on	:	2010-03-13
Resolution	:	2.00 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
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 (1)) were used in the production of this report:

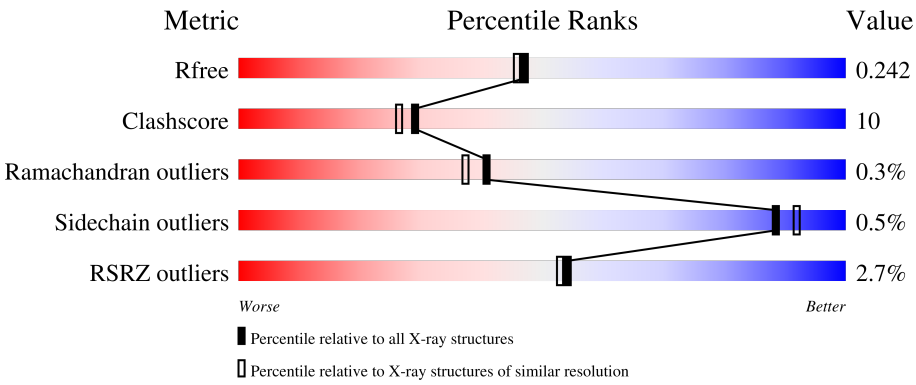
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.00 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	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (2.00-2.00)

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	133	% 73% 22% 5%
1	B	133	% 80% 15% 5%
1	D	133	73% 18% 9%
1	E	133	5% 74% 17% 9%
1	F	133	8% 73% 17% • 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	G	133	% 69% 21% • 9%

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 6136 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 Nonstructural protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	126	1000	641	171	181	7	0	0	0
			Total	C	N	O	S			
1	B	126	1000	641	171	181	7	0	0	0
			Total	C	N	O	S			
1	D	121	962	617	164	175	6	0	0	0
			Total	C	N	O	S			
1	E	121	962	617	164	175	6	0	0	0
			Total	C	N	O	S			
1	F	121	962	617	164	175	6	0	0	0
			Total	C	N	O	S			
1	G	121	962	617	164	175	6	0	0	0

There are 18 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	76	GLY	-	expression tag	UNP D2Y6Z6
A	77	SER	-	expression tag	UNP D2Y6Z6
A	78	HIS	-	expression tag	UNP D2Y6Z6
B	76	GLY	-	expression tag	UNP D2Y6Z6
B	77	SER	-	expression tag	UNP D2Y6Z6
B	78	HIS	-	expression tag	UNP D2Y6Z6
D	76	GLY	-	expression tag	UNP D2Y6Z6
D	77	SER	-	expression tag	UNP D2Y6Z6
D	78	HIS	-	expression tag	UNP D2Y6Z6
E	76	GLY	-	expression tag	UNP D2Y6Z6
E	77	SER	-	expression tag	UNP D2Y6Z6
E	78	HIS	-	expression tag	UNP D2Y6Z6
F	76	GLY	-	expression tag	UNP D2Y6Z6
F	77	SER	-	expression tag	UNP D2Y6Z6
F	78	HIS	-	expression tag	UNP D2Y6Z6
G	76	GLY	-	expression tag	UNP D2Y6Z6
G	77	SER	-	expression tag	UNP D2Y6Z6

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	78	HIS	-	expression tag	UNP D2Y6Z6

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
2	A	49	49	49	0	0
			Total	O		
2	B	49	49	49	0	0
			Total	O		
2	D	62	62	62	0	0
			Total	O		
2	E	54	54	54	0	0
			Total	O		
2	F	27	27	27	0	0
			Total	O		
2	G	47	47	47	0	0

3 Residue-property plots (i)

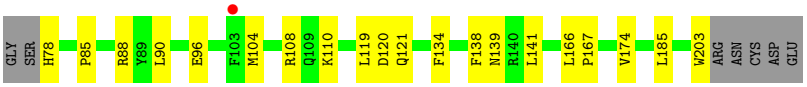
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Nonstructural protein 1

Chain A:

• Molecule 1: Nonstructural protein 1

Chain B:



• Molecule 1: Nonstructural protein 1

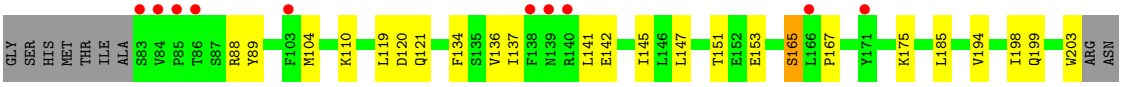
Chain D:

• Molecule 1: Nonstructural protein 1

Chain E:

• Molecule 1: Nonstructural protein 1

Chain F:



CYS
ASP
GLU

• Molecule 1: Nonstructural protein 1

Chain G:

GLY
SER
HIS
MET
THR
ILE
ALA

S83 V84 P85 T86 S87 R88 M98 Q109 K110 I111 V117 I123 K131 A132 N133 F134 I137 E142 L147 V157 L163 P164 S165 L166 P167 G168 Y171 K175 W187 V194 S195 E196 N197 I198 Q199 R200 F201 A202 W203

ARG
ASN
CYS
ASP
GLU

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	73.53Å 70.80Å 96.69Å 90.00° 110.30° 90.00°	Depositor
Resolution (Å)	50.00 – 2.00 48.77 – 2.01	Depositor EDS
% Data completeness (in resolution range)	91.8 (50.00-2.00) 91.8 (48.77-2.01)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.82 (at 2.01Å)	Xtriage
Refinement program	PHENIX 1.6_289	Depositor
R, R_{free}	0.207 , 0.244 0.204 , 0.242	Depositor DCC
R_{free} test set	1694 reflections (2.73%)	wwPDB-VP
Wilson B-factor (Å^2)	33.5	Xtriage
Anisotropy	0.341	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 52.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.020 for h,-k,-h-l	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6136	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 15.31% 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)

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	0/1020	0.55	0/1383
1	B	0.39	0/1020	0.56	0/1383
1	D	0.45	0/981	0.58	0/1330
1	E	0.41	0/981	0.55	0/1330
1	F	0.40	0/981	0.56	0/1330
1	G	0.42	0/981	0.59	0/1330
All	All	0.41	0/5964	0.56	0/8086

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	1000	0	1018	27	0
1	B	1000	0	1018	13	0
1	D	962	0	979	19	0
1	E	962	0	979	20	0
1	F	962	0	979	18	0
1	G	962	0	979	26	0
2	A	49	0	0	3	0
2	B	49	0	0	1	0
2	D	62	0	0	1	0
2	E	54	0	0	2	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	F	27	0	0	0	0
2	G	47	0	0	2	0
All	All	6136	0	5952	113	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 10.

The worst 5 of 113 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:139:ASN:HA	1:E:201:PHE:HE1	1.31	0.96
1:E:139:ASN:HA	1:E:201:PHE:CE1	2.20	0.75
1:D:89:TYR:HE1	1:D:145:ILE:HD12	1.56	0.70
1:B:85:PRO:HB3	1:B:138:PHE:CE1	2.27	0.70
1:A:139:ASN:HD21	1:A:200:ARG:NH2	1.88	0.69

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	124/133 (93%)	122 (98%)	2 (2%)	0	100	100
1	B	124/133 (93%)	117 (94%)	7 (6%)	0	100	100
1	D	119/133 (90%)	115 (97%)	4 (3%)	0	100	100
1	E	119/133 (90%)	114 (96%)	4 (3%)	1 (1%)	19	13
1	F	119/133 (90%)	116 (98%)	3 (2%)	0	100	100
1	G	119/133 (90%)	114 (96%)	4 (3%)	1 (1%)	19	13
All	All	724/798 (91%)	698 (96%)	24 (3%)	2 (0%)	41	37

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	165	SER
1	G	168	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	111/117 (95%)	111 (100%)	0	100	100
1	B	111/117 (95%)	110 (99%)	1 (1%)	78	83
1	D	107/117 (92%)	107 (100%)	0	100	100
1	E	107/117 (92%)	107 (100%)	0	100	100
1	F	107/117 (92%)	106 (99%)	1 (1%)	78	83
1	G	107/117 (92%)	106 (99%)	1 (1%)	78	83
All	All	650/702 (93%)	647 (100%)	3 (0%)	88	92

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	139	ASN
1	F	165	SER
1	G	137	ILE

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	D	139	ASN
1	G	133	ASN

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)

There are no ligands in this entry.

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	126/133 (94%)	-0.16	1 (0%)	86	85	25, 38, 55, 64	0
1	B	126/133 (94%)	-0.08	1 (0%)	86	85	27, 39, 54, 64	0
1	D	121/133 (90%)	-0.05	0	100	100	21, 33, 49, 65	0
1	E	121/133 (90%)	0.02	7 (5%)	23	22	23, 37, 63, 87	0
1	F	121/133 (90%)	0.23	10 (8%)	11	10	26, 41, 70, 89	0
1	G	121/133 (90%)	0.08	1 (0%)	86	85	25, 39, 58, 67	0
All	All	736/798 (92%)	0.01	20 (2%)	54	53	21, 38, 58, 89	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	84	VAL	7.1
1	F	84	VAL	6.6
1	F	139	ASN	5.1
1	F	138	PHE	4.8
1	E	166	LEU	4.5

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)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.