wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 04:43 AM BST

PDB ID	:	2M2U
Title	:	Binary complex of African Swine Fever Virus Pol X with MgdGTP
Authors	:	Wu, W.; Su, M.; Tsai, M.
Deposited on	:	2013-01-03

This is a wwPDB NMR 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/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

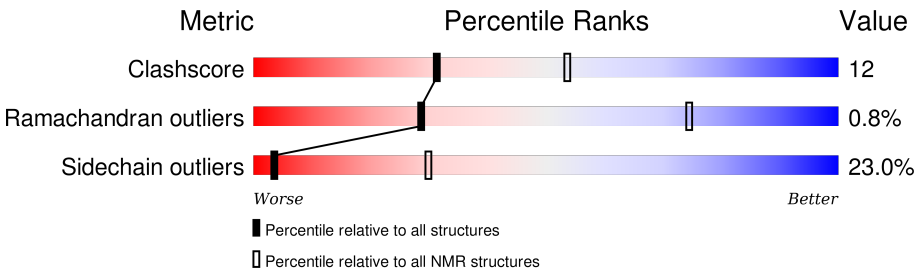
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 88%.

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)	NMR archive (#Entries)
Clashscore	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

The table below summarises the geometric issues observed across the polymeric chains and their fit to the experimental data. The red, orange, yellow and green segments indicate the fraction of residues that contain outliers for >=3, 2, 1 and 0 types of geometric quality criteria. A cyan segment indicates the fraction of residues that are not part of the well-defined cores, and 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%

Mol	Chain	Length	Quality of chain
1	A	174	59% 38% • •

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 6 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

The following residues are included in the computation of the global validation metrics.

Well-defined (core) protein residues			
Well-defined core	**Residue range (total)**	**Backbone RMSD (Å)**	**Medoid model**
1	A:2-A:174 (173)	0.28	6

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

The models can be grouped into 3 clusters and 6 single-model clusters were found.

Cluster number	Models
1	6, 10, 12, 13, 14, 17, 18
2	1, 5, 8, 11, 20
3	2, 16
Single-model clusters	3; 4; 7; 9; 15; 19

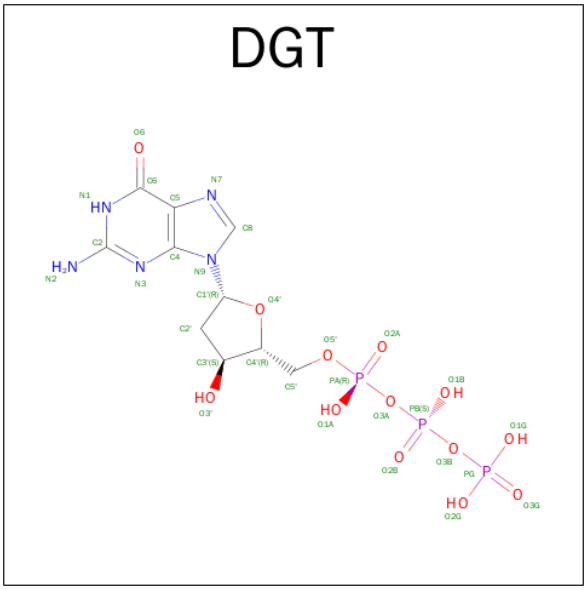
3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 3048 atoms, of which 1581 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Repair DNA polymerase X.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	174	3003	943	1569	248	239	4	0

- Molecule 2 is 2'-DEOXYGUANOSINE-5'-TRIPHOSPHATE (three-letter code: DGT) (formula: $C_{10}H_{16}N_5O_{13}P_3$).



Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	P
2	A	1	43	10	12	5	13	3

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	
			Total	Mg
3	A	2	2	2

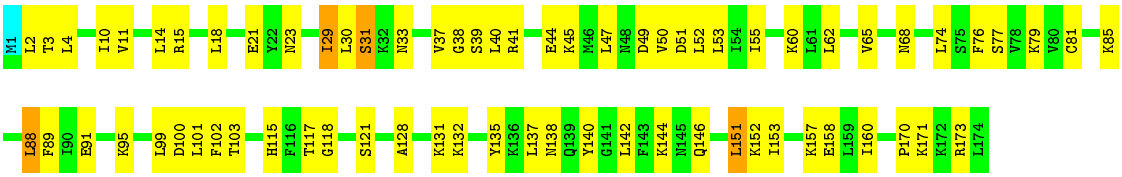
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-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. Stretches of 2 or more consecutive residues without any outliers are shown as green connectors. Residues which are classified as ill-defined in the NMR ensemble, are shown in cyan with an underline colour-coded according to the previous scheme. Residues which were present in the experimental sample, but not modelled in the final structure are shown in grey.

- Molecule 1: Repair DNA polymerase X

Chain A: 59% 38%

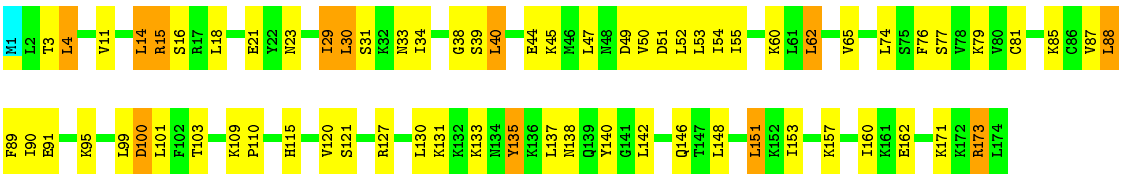


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 6. Colouring as in section 4.1 above.

- Molecule 1: Repair DNA polymerase X

Chain A: 60% 33% 7%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *distance geometry*.

Of the 2000 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
CYANA	structure solution	
Haddock	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	2m2u_cs.str
Number of chemical shift lists	1
Total number of shifts	2357
Number of shifts mapped to atoms	2357
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	88%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality (i)

6.1 Standard geometry (i)

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

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

6.2 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 each 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 averaged over the ensemble.

Mol	Chain	Non-H	H(model)	H(added)	Clashes
1	A	1426	1558	1558	32±4
2	A	31	12	12	9±2
3	A	2	0	0	1±0
All	All	29180	31400	31400	698

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 12.

5 of 223 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:A:201:DGT:O3G	3:A:203:MG:MG	0.78	1.27	13	17
1:A:91:GLU:HB2	1:A:96:THR:HG22	0.71	1.61	3	8
1:A:115:HIS:CD2	2:A:201:DGT:H8	0.69	2.22	3	12
1:A:10:ILE:HG22	1:A:92:TRP:HZ2	0.69	1.47	3	1
2:A:201:DGT:O3G	2:A:201:DGT:O2B	0.69	2.10	17	2

6.3 Torsion angles (i)

6.3.1 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 PDB entries followed by that with respect to all NMR entries. 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	172/174 (99%)	163±1 (95±1%)	8±1 (5±1%)	1±1 (1±0%)	29	74
All	All	3440/3480 (99%)	3251 (95%)	161 (5%)	28 (1%)	29	74

All 5 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	173	ARG	19
1	A	117	THR	4
1	A	145	ASN	2
1	A	119	PRO	2
1	A	146	GLN	1

6.3.2 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 PDB entries followed by that with respect to all NMR entries. 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	158/159 (99%)	122±3 (77±2%)	36±3 (23±2%)	3	30
All	All	3160/3180 (99%)	2433 (77%)	727 (23%)	3	30

5 of 97 unique residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	A	29	ILE	20
1	A	153	ILE	20
1	A	151	LEU	20
1	A	79	LYS	20
1	A	103	THR	20

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates (i)

There are no carbohydrates in this entry.

6.6 Ligand geometry (i)

Of 3 ligands modelled in this entry, 2 are monoatomic - leaving 1 for Mogul analysis.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds 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 is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	DGT	A	201	3	25,33,33	1.62±0.02	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of 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 angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	DGT	A	201	3	26,52,52	1.50±0.13	0±0 (0±1%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means

no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	DGT	A	201	3	-	0±0,18,34,34	0±0,3,3,3

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	A	201	DGT	C1'-N9-C4	5.90	118.45	127.07	7	3

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.7 Other polymers (i)

There are no such molecules in this entry.

6.8 Polymer linkage issues (i)

There are no chain breaks in this entry.

7 Chemical shift validation (i)

The completeness of assignment taking into account all chemical shift lists is 88% for the well-defined parts and 88% for the entire structure.

7.1 Chemical shift list 1

File name: 2m2u_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	2357
Number of shifts mapped to atoms	2357
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	19

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	174	2.46 ± 0.09	Should be applied
$^{13}C_{\beta}$	166	2.79 ± 0.13	Should be applied
$^{13}C'$	149	2.88 ± 0.11	Should be applied
^{15}N	167	0.69 ± 0.31	Should be applied

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 88%, i.e. 2071 atoms were assigned a chemical shift out of a possible 2350. 38 out of 38 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	829/853 (97%)	340/340 (100%)	322/346 (93%)	167/167 (100%)
Sidechain	1124/1336 (84%)	696/784 (89%)	416/485 (86%)	12/67 (18%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	118/161 (73%)	78/86 (91%)	39/71 (55%)	1/4 (25%)
Overall	2071/2350 (88%)	1114/1210 (92%)	777/902 (86%)	180/238 (76%)

7.1.4 Statistically unusual chemical shifts (i)

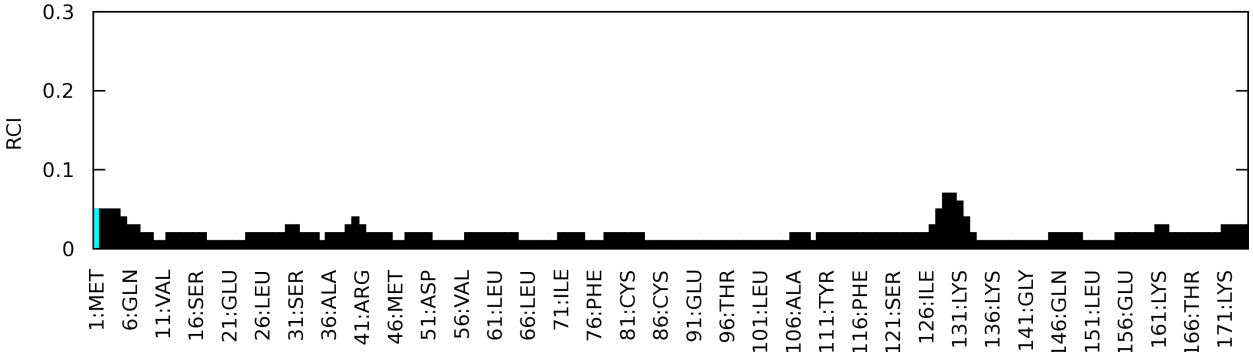
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	168	ARG	CD	33.07	47.57 – 38.77	-11.5
1	A	42	ARG	CD	34.13	47.57 – 38.77	-10.3
1	A	8	LYS	CD	40.58	34.86 – 23.06	9.9
1	A	170	PRO	CD	41.64	55.31 – 45.41	-8.8
1	A	8	LYS	HD3	3.14	2.75 – 0.45	6.7
1	A	8	LYS	HD2	3.14	2.76 – 0.46	6.7
1	A	141	GLY	HA3	1.50	5.80 – 2.00	-6.3
1	A	173	ARG	HD3	1.55	4.36 – 1.86	-6.2
1	A	168	ARG	HD2	1.72	4.27 – 1.97	-6.1
1	A	93	GLU	HG3	1.00	3.31 – 1.21	-6.0
1	A	70	ARG	HD2	1.77	4.27 – 1.97	-5.9
1	A	173	ARG	HD2	1.77	4.27 – 1.97	-5.8
1	A	127	ARG	HD2	1.78	4.27 – 1.97	-5.8
1	A	168	ARG	HD3	1.72	4.36 – 1.86	-5.6
1	A	42	ARG	CG	33.88	33.23 – 21.23	5.5
1	A	70	ARG	HD3	1.77	4.36 – 1.86	-5.4
1	A	127	ARG	HD3	1.78	4.36 – 1.86	-5.3
1	A	17	ARG	HD2	4.30	4.27 – 1.97	5.1
1	A	10	ILE	HG13	-0.86	3.26 – -0.84	-5.1

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain A:


RCI
0.3
0.2
0.1
0
1:MET
6:GLN
11:VAL
16:SER
21:GLU
26:LEU
31:SER
36:ALA
41:ARG
46:MET
51:ASP
56:VAL
61:LEU
66:LEU
71:ILE
76:PHE
81:CYS
86:CYS
91:GLU
96:THR
101:LEU
106:ALA
111:TYR
116:PHE
121:SER
126:ILE
131:LYS
136:LYS
141:GLY
146:GLN
151:LEU
156:GLU
161:LYS
166:THR
171:LYS