wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 04:15 AM BST

PDB ID	:	2MY1
Title	:	Solution structure of Bud31p
Authors	:	van Roon, A.M.; Yang, J.; Mathieu, D.; Bermel, W.; Nagai, K.; Neuhaus, D.
Deposited on	:	2015-01-19

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:

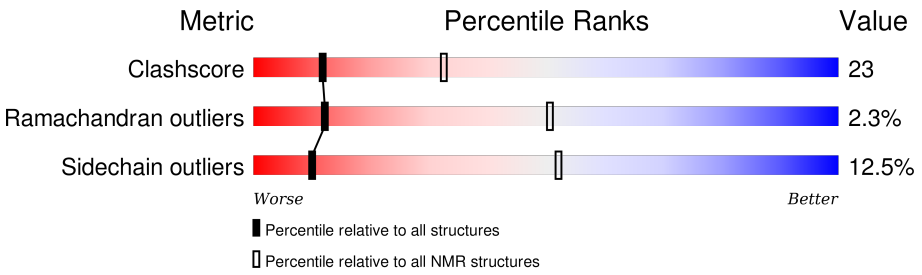
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 76%.

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	159	

2 Ensemble composition and analysis (i)

This entry contains 35 models. Model 2 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:12-A:36, A:45-A:110, A:119-A:155 (128)	0.35	2

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 6 clusters and 5 single-model clusters were found.

Cluster number	Models
1	2, 4, 6, 7, 9, 10, 11, 14, 15, 16, 20, 22, 26, 27, 30, 35
2	1, 8, 17, 29
3	12, 18, 28, 34
4	5, 23
5	31, 32
6	3, 24
Single-model clusters	13; 19; 21; 25; 33

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2614 atoms, of which 1314 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Pre-mRNA-splicing factor BUD31.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	159	2611	810	1314	242	235	10	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	GLY	-	EXPRESSION TAG	UNP P25337
A	0	GLY	-	EXPRESSION TAG	UNP P25337
A	1	SER	-	EXPRESSION TAG	UNP P25337

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	3	3	3

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: Pre-mRNA-splicing factor BUD31

Chain A: 52% 25% • 19%

G-1 G0 S1 P2 R3 I4 K5 T6 R7 R8 S9 K10 P11 A12 I19 K20 P21 T22 L23 F26 L30 A33 K37 S38 S39 K40 L41 A42 A43 K44 W50 E51 I52 M53 Q54 L55 H56 H57 Q58 R59 Y62 I63 L66 Y67 I73 L77 W80 L81

K85 Y86 A87 L91 R96 K97 K102 L103 C104 C105 L106 R107 C108 I109 Q110 K111 N112 E113 T114 N115 N116 G117 S118 T119 C120 I121 C122 R123 L129 F142 C148 G152 S155 T156 D157

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

The representative model is number 2. Colouring as in section 4.1 above.

- Molecule 1: Pre-mRNA-splicing factor BUD31

Chain A: 47% 30% • 19%

G-1 G0 S1 P2 R3 I4 K5 T6 R7 R8 S9 K10 P11 A12 I19 K20 P21 T22 L23 T24 D25 F26 L30 R31 D32 A33 K37 S38 S39 K40 L41 A42 A43 K44 W50 E51 I52 L55 H56 H57 Q58 R59 Y62 I63 L66 Y67 I73 L77 W80 L81

I82 K83 E84 K85 Y86 A87 D88 K89 L90 R96 K97 Y100 E101 K102 L103 C104 C105 L106 R107 C108 I109 Q110 K111 N112 E113 T114 N115 N116 G117 S118 T119 C120 I121 C122 R123 L129 E132 A133 T138 F142 H143 Q144 C148 R151 T156 D157

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 50 calculated structures, 35 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
X-PLOR NIH	structure solution	2.28
X-PLOR NIH	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)	2my1_cs.str
Number of chemical shift lists	1
Total number of shifts	1698
Number of shifts mapped to atoms	1698
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	76%

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: ZN

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 (average) 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	1.09±0.00	0±0/1088 (0.0±0.0%)	1.19±0.00	1±0/1461 (0.1±0.0%)
All	All	1.09	0/38080 (0.0%)	1.19	30/51135 (0.1%)

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
1	A	108	CYS	N-CA-CB	-5.58	100.55	110.60	31	30

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	1065	1066	1066	49±4
All	All	37380	37310	37310	1704

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 23.

5 of 178 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:30:LEU:HD11	1:A:52:ILE:HG23	1.05	1.26	13	15
1:A:22:THR:HG21	1:A:62:TYR:CE2	0.86	2.05	32	34
1:A:121:ILE:HD12	1:A:142:PHE:CD1	0.86	2.06	15	17
1:A:63:ILE:HD12	1:A:87:ALA:HB2	0.84	1.47	26	29
1:A:30:LEU:CD1	1:A:52:ILE:HG23	0.83	2.03	29	15

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	128/159 (81%)	111±2 (87±2%)	14±2 (11±2%)	3±1 (2±1%)	12	51
All	All	4480/5565 (81%)	3894 (87%)	483 (11%)	103 (2%)	12	51

5 of 8 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	97	LYS	32
1	A	109	ILE	30
1	A	152	GLY	27
1	A	12	ALA	7
1	A	14	ASP	3

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	115/141 (82%)	101±3 (87±3%)	14±3 (13±3%)	10	52
All	All	4025/4935 (82%)	3520 (87%)	505 (13%)	10	52

5 of 59 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	121	ILE	31
1	A	102	LYS	31
1	A	155	SER	29
1	A	96	ARG	28
1	A	30	LEU	23

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, 3 are monoatomic - leaving 0 for Mogul analysis.

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 76% for the well-defined parts and 73% for the entire structure.

7.1 Chemical shift list 1

File name: 2my1_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	1698
Number of shifts mapped to atoms	1698
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	158	-0.47 ± 0.14	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	149	0.17 ± 0.07	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	151	-0.31 ± 0.38	None needed (< 0.5 ppm)

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 76%, i.e. 1291 atoms were assigned a chemical shift out of a possible 1706. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	502/634 (79%)	251/253 (99%)	127/256 (50%)	124/125 (99%)
Sidechain	654/925 (71%)	408/549 (74%)	236/322 (73%)	10/54 (19%)

Continued on next page...

Continued from previous page...

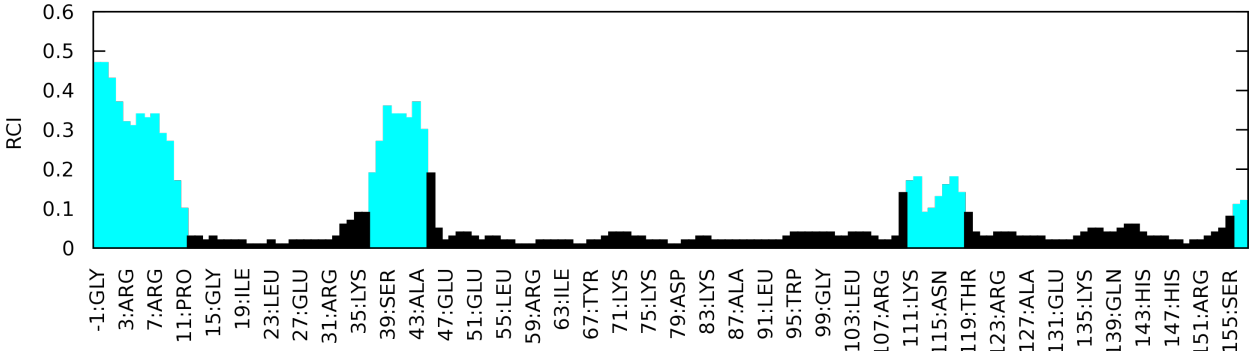
	Total	^{1}H	^{13}C	^{15}N
Aromatic	135/147 (92%)	69/77 (90%)	63/63 (100%)	3/7 (43%)
Overall	1291/1706 (76%)	728/879 (83%)	426/641 (66%)	137/186 (74%)

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	84	GLU	HG2	0.50	3.33 – 1.23	-8.5
1	A	84	GLU	HG3	1.07	3.31 – 1.21	-5.7
1	A	12	ALA	HB3	0.01	2.61 – 0.11	-5.4
1	A	12	ALA	HB1	0.01	2.61 – 0.11	-5.4
1	A	12	ALA	HB2	0.01	2.61 – 0.11	-5.4
1	A	91	LEU	HG	-0.20	3.16 – -0.14	-5.2

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: