wwPDB NMR Structure Validation Summary Report (i)

May 19, 2016 – 10:48 PM EDT

PDB ID	:	2YH1
Title	:	MODEL OF HUMAN U2AF65 TANDEM RRM1 AND RRM2 DOMAINS WITH EIGHT-SITE URIDINE BINDING
Authors	:	Mackereth, C.D.; Madl, T.; Simon, B.; Zanier, K.; Gasch, A.; Sattler, M.
Deposited on	:	2011-04-26

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:

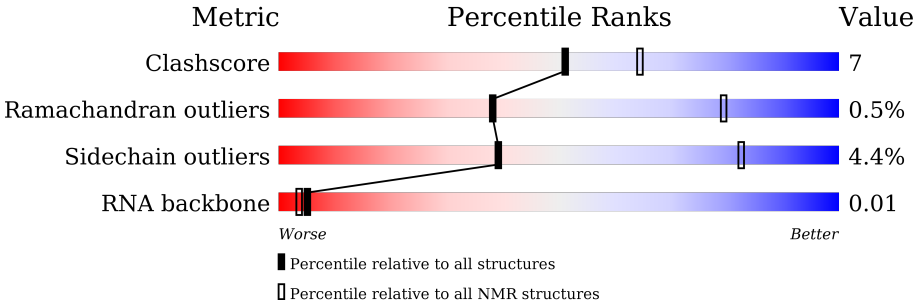
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 55%.

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
RNA backbone	3027	600

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	198	71% 10% • 17% •
2	B	9	67% 22% 11%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 5 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:149-A:230 (82)	0.05	5
2	A:259-A:337 (79)	0.03	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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	2, 3, 4, 5, 7, 10
2	1, 6, 8, 9

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3255 atoms, of which 1584 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called SPLICING FACTOR U2AF 65 KDA SUBUNIT.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	195	2986	952	1492	254	281	7	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	145	GLY	-	EXPRESSION TAG	UNP P26368
A	146	ALA	-	EXPRESSION TAG	UNP P26368
A	147	MET	-	EXPRESSION TAG	UNP P26368

- Molecule 2 is a RNA chain called 5'-R(*UP*UP*UP*UP*UP*UP*UP*UP*UP)-3'.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	9	269	81	92	18	70	8	0

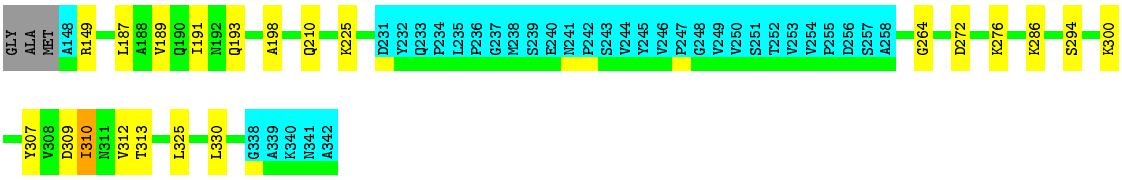
4 Residue-property plots (i)

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: SPLICING FACTOR U2AF 65 KDA SUBUNIT

Chain A: 71% 10% • 17% •



- Molecule 2: 5'-R(*UP*UP*UP*UP*UP*UP*UP*UP*UP)-3'

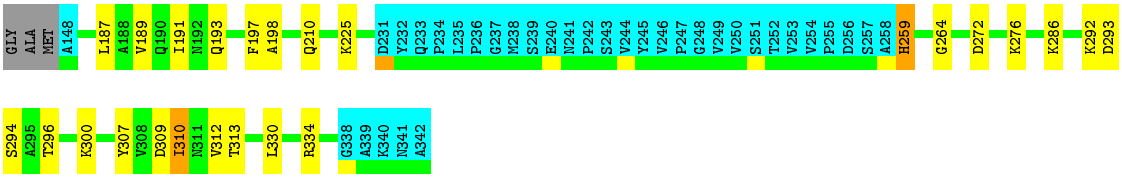
Chain B: 67% 22% 11%

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

The representative model is number 5. Colouring as in section 4.1 above.

- Molecule 1: SPLICING FACTOR U2AF 65 KDA SUBUNIT

Chain A: 69% 12% • 17% •



- Molecule 2: 5'-R(*UP*UP*UP*UP*UP*UP*UP*UP*UP)-3'

Chain B: 11% 44% 33% 11%

U601
U602
U603
U604
U605
U606
U607
U608
U609

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *MODIFIED ARIA*.

Of the 125 calculated structures, 10 were deposited, based on the following criterion: *LOWEST ENERGY*.

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

Software name	Classification	Version
CNS 1.2	refinement	
SPARKY	structure solution	

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)	2yh1_cs.cif
Number of chemical shift lists	1
Total number of shifts	1474
Number of shifts mapped to atoms	1473
Number of unparsed shifts	1
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	55%

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)

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	0.59±0.00	0±0/1279 (0.0±0.0%)	0.50±0.01	0±0/1726 (0.0±0.0%)
2	B	0.79±0.07	2±0/194 (0.9±0.2%)	0.88±0.04	0±0/298 (0.1±0.2%)
All	All	0.62	17/14730 (0.1%)	0.58	3/20240 (0.0%)

All unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	B	604	U	O3'-P	7.89	1.70	1.61	6	10
2	B	604	U	C3'-O3'	7.04	1.52	1.42	4	7

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	B	604	U	C3'-C2'-C1'	5.75	106.10	101.50	8	2
2	B	604	U	P-O3'-C3'	5.35	126.12	119.70	6	1

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	1255	1260	1258	18±5

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes
2	B	177	92	92	12±5
All	All	14320	13520	13500	198

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 7.

5 of 63 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:331:LEU:CD1	2:B:602:U:H3	1.34	1.34	4	3
1:A:331:LEU:HD13	2:B:602:U:N3	1.09	1.61	4	2
1:A:331:LEU:CD1	2:B:602:U:N3	1.08	2.16	4	3
1:A:331:LEU:HD13	2:B:602:U:H3	1.03	1.11	6	2
1:A:300:LYS:NZ	2:B:601:U:O2'	0.99	1.96	1	7

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	161/198 (81%)	150±1 (93±1%)	10±1 (6±0%)	1±1 (0±0%)	38	79
All	All	1610/1980 (81%)	1499 (93%)	103 (6%)	8 (0%)	38	79

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	337	VAL	2
1	A	188	ALA	2
1	A	281	SER	2
1	A	149	ARG	1
1	A	259	HIS	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	133/161 (83%)	127±1 (96±0%)	6±1 (4±0%)	39	82
All	All	1330/1610 (83%)	1271 (96%)	59 (4%)	39	82

5 of 10 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	210	GLN	10
1	A	310	ILE	10
1	A	189	VAL	10
1	A	187	LEU	10
1	A	193	GLN	10

6.3.3 RNA (i)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
2	B	8/9 (89%)	7±1 (85±15%)	2±1 (20±8%)	0.01±0.01
All	All	80/90 (89%)	68 (85%)	16 (20%)	0.01

The overall RNA backbone suiteness is 0.01.

5 of 8 unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	606	U	10
2	B	602	U	10
2	B	603	U	9
2	B	609	U	9
2	B	605	U	8

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
2	B	606	U	8
2	B	605	U	7

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
2	B	603	U	1

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)

There are no ligands in this entry.

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 55% for the well-defined parts and 51% for the entire structure.

7.1 Chemical shift list 1

File name: 2yh1_cs.cif

Chemical shift list name: *RRM12_bound_chemicalshifts_31.str.csh*

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	1474
Number of shifts mapped to atoms	1473
Number of unparsed shifts	1
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

The following errors were found when reading this chemical shift list.

- Entity instance (chain) must be specified. The only occurence is reported below.

Shift ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1474	?	345	VAL	N	121.445	?	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	129	-0.05 ± 0.16	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	118	0.02 ± 0.12	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	171	0.35 ± 0.30	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 55%, i.e. 1181 atoms were assigned a chemical shift out of a possible 2133. 0 out of 28 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

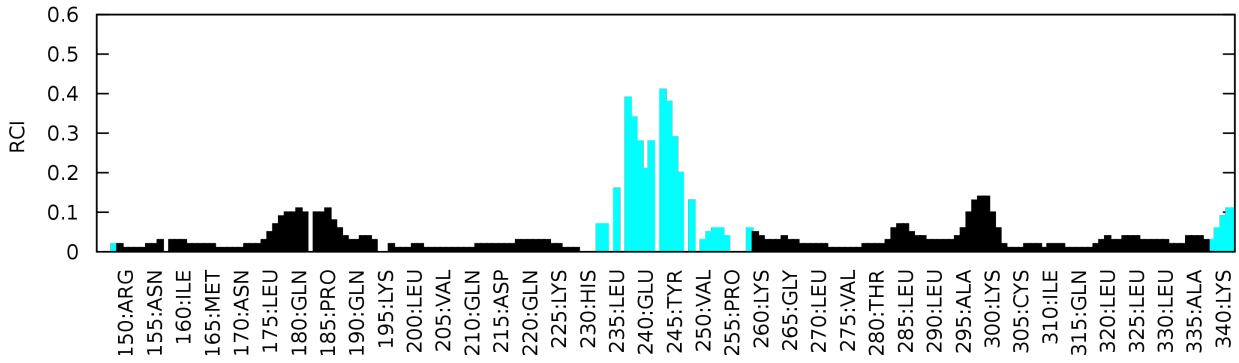
	Total	^{1}H	^{13}C	^{15}N
Backbone	535/793 (67%)	267/316 (84%)	121/322 (38%)	147/155 (95%)
Sidechain	607/1031 (59%)	366/602 (61%)	230/378 (61%)	11/51 (22%)
Aromatic	39/156 (25%)	39/84 (46%)	0/68 (0%)	0/4 (0%)
Overall	1181/2133 (55%)	672/1083 (62%)	351/831 (42%)	158/219 (72%)

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	276	LYS	HD3	0.15	2.75 – 0.45	-6.3
1	A	172	GLN	HG3	0.60	3.75 – 0.85	-5.9
1	A	208	THR	HA	1.95	6.86 – 2.06	-5.2
1	A	276	LYS	HD2	0.43	2.76 – 0.46	-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: