wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 05:43 AM BST

PDB ID	:	2YZ0
Title	:	Solution Structure of RWD/GI domain of Saccharomyces cerevisiae GCN2
Authors	:	Ogura, K.; Torikai, S.; Kumeta, H.; Inagaki, F.
Deposited on	:	2007-05-02

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:

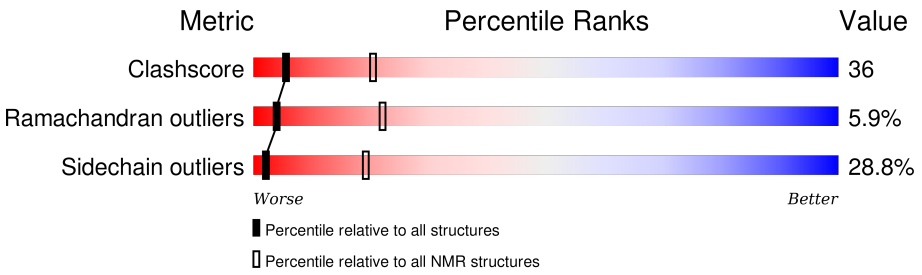
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 80%.

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	138	27% 42% 10% • 20%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 15 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:8-A:34, A:45-A:128 (111)	0.24	15

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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2261 atoms, of which 1118 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Serine/threonine-protein kinase GCN2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	138	2261	725	1118	183	228	7	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: Serine/threonine-protein kinase GCN2

Chain A: 27% 42% 10% • 20%

M1 S2 L3 S4 H5 L6 T7 L8 D9 Y12 N17 E18 I22 R23 S24 I25 D28 D29 F30 T31 D32 L33 T34 K35 R36 K37 S38 S39 W40 D41 K42 Q43 P44 Q45 I46 I47 F48 E49 L52 R53 S54 V55 D56 K57 E58 P59 V60 E61 S62 S63 I64 T65 L66 H67 F68

A69 M70 T71 P72 M73 Y74 P75 Y76 T77 A78 P79 E80 I81 E82 F83 K84 N85 V86 Q87 N88 V89 M90 D91 S92 Q93 L94 L97 K98 S99 E100 F101 K102 K103 I104 H105 N106 R109 I113 I114 F115 E116 I117 T118 S119 F120 T121 Q122 E123 K124 L125 D126 E127 F128 Q129 N130 V131 V132

N133 T134 Q135 S136 L137 E138

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

The representative model is number 15. Colouring as in section 4.1 above.

- Molecule 1: Serine/threonine-protein kinase GCN2

Chain A: 38% 33% 8% • 20%

M1 S2 L3 S4 H5 L6 T7 L8 N17 E18 I22 R23 S24 I25 L33 T34 K35 R36 K37 S38 S39 W40 D41 K42 Q43 P44 Q45 I46 I47 F48 E49 L52 R53 S54 E58 P59 V60 E61 S62 L66 H67 F68 A69 M70 T71 P72 M73 Y74 P75 Y76 T77 A78 P79

E82 F83 K84 N85 V86 Q87 D91 S92 Q93 L94 Q95 M96 L97 E100 F101 I104 H105 N106 R109 I113 I114 F115 E116 I117 T118 T121 K124 L125 D126 E127 F128 Q129 N130 V131 V132 N133 T134 Q135 S136 L137 E138

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
DYANA	structure solution	
DYANA	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)	BMRB entry 7395
Number of chemical shift lists	1
Total number of shifts	1473
Number of shifts mapped to atoms	1473
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	80%

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)

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	924	897	897	65±6
All	All	18480	17940	17940	1293

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 36.

5 of 394 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:64:ILE:HD11	1:A:89:VAL:HG21	1.06	1.26	8	8
1:A:66:LEU:HD11	1:A:121:THR:HG21	1.04	1.24	12	12
1:A:33:LEU:HD21	1:A:67:HIS:CE1	1.04	1.87	15	6
1:A:66:LEU:HD21	1:A:121:THR:HG21	1.02	1.27	3	5
1:A:97:LEU:HD22	1:A:121:THR:HG22	1.01	1.27	2	12

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	111/138 (80%)	92±2 (83±2%)	13±2 (11±2%)	7±1 (6±1%)	4 22
All	All	2220/2760 (80%)	1835 (83%)	253 (11%)	132 (6%)	4 22

5 of 13 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	46	ILE	20
1	A	62	SER	20
1	A	86	VAL	20
1	A	91	ASP	19
1	A	85	ASN	17

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	107/134 (80%)	76±4 (71±3%)	31±4 (29±3%)	2 19
All	All	2140/2680 (80%)	1523 (71%)	617 (29%)	2 19

5 of 73 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	47	ILE	20
1	A	74	TYR	20
1	A	68	PHE	20
1	A	87	GLN	20
1	A	71	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)

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 80% for the well-defined parts and 74% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 7395

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	1473
Number of shifts mapped to atoms	1473
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	126	-0.37 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	118	0.32 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	119	0.42 ± 0.19	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 80%, i.e. 1143 atoms were assigned a chemical shift out of a possible 1426. 12 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	435/547 (80%)	218/218 (100%)	111/222 (50%)	106/107 (99%)
Sidechain	631/753 (84%)	397/441 (90%)	226/285 (79%)	8/27 (30%)

Continued on next page...

Continued from previous page...

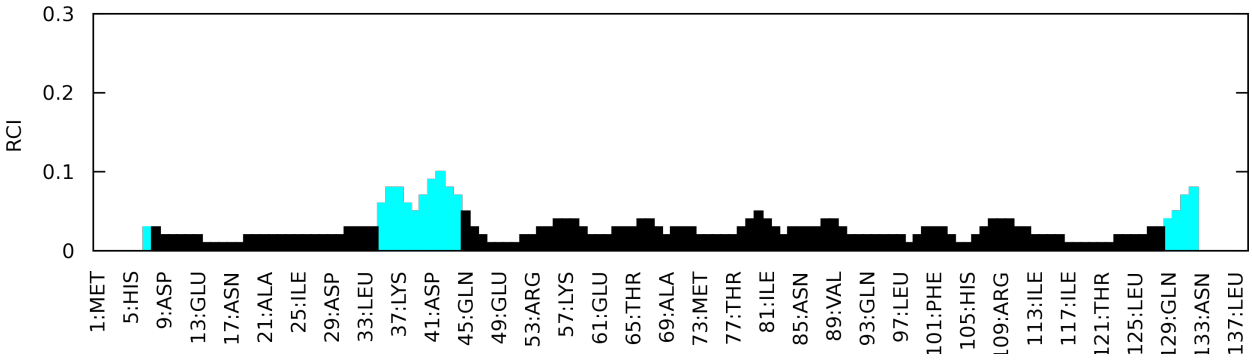
	Total	^{1}H	^{13}C	^{15}N
Aromatic	77/126 (61%)	53/68 (78%)	24/56 (43%)	0/2 (0%)
Overall	1143/1426 (80%)	668/727 (92%)	361/563 (64%)	114/136 (84%)

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	53	ARG	NE	111.05	92.63 – 76.73	16.6
1	A	43	GLN	HB2	0.73	3.30 – 0.80	-5.3
1	A	68	PHE	HB2	1.06	4.85 – 1.15	-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: