wwPDB NMR Structure Validation Summary Report (i)

Feb 23, 2022 – 11:24 AM EST

PDB ID	:	1X50
Title	:	Solution structure of the C-terminal gal-bind lectin domain from human galectin-4
Authors	:	Tomizawa, T.; Kigawa, T.; Saito, K.; Koshiba, S.; Inoue, M.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2005-05-15

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
https://www.wwpdb.org/validation/2017/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:

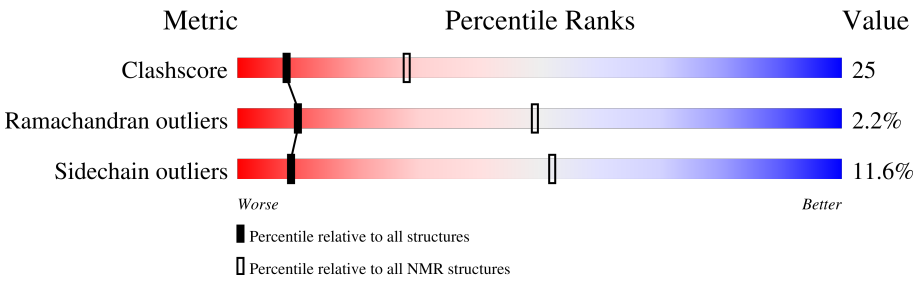
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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 was not calculated.

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	164	50% 27% • 18%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 18 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:22-A:76, A:80-A:158 (134)	0.14	18

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2481 atoms, of which 1231 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Galectin-4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	164	2481	791	1231	225	231	3	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	cloning artifact	UNP P56470
A	2	SER	-	cloning artifact	UNP P56470
A	3	SER	-	cloning artifact	UNP P56470
A	4	GLY	-	cloning artifact	UNP P56470
A	5	SER	-	cloning artifact	UNP P56470
A	6	SER	-	cloning artifact	UNP P56470
A	7	GLY	-	cloning artifact	UNP P56470
A	159	SER	-	cloning artifact	UNP P56470
A	160	GLY	-	cloning artifact	UNP P56470
A	161	PRO	-	cloning artifact	UNP P56470
A	162	SER	-	cloning artifact	UNP P56470
A	163	SER	-	cloning artifact	UNP P56470
A	164	GLY	-	cloning artifact	UNP P56470

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Galectin-4

Chain A: 50% 27% • 18%

G1 S2 S3 G4 S5 S6 G7 H8 Q9 Q10 L11 N12 S13 L14 P15 T16 M17 E18 G19 P20 P21 T22 F23 N24 P25 P26 V27 P28 R32 G36 L37 T38 R41 Y48 F56 A57 F60 K61 V62 D67 I68 A69 L70 H71 I72 N73 P74 R75 M76 G77 N78 G79

V82 R83 N84 S85 L86 L87 N88 G89 S90 W91 K96 K97 I98 T99 H100 F103 D110 L111 S112 C115 D118 R119 F120 K121 V122 F130 D131 F132 A133 F139 Q140 R141 V142 D143 T144 L145 Q148 S159 G160 P161 S162 S163 G164

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

The representative model is number 18. Colouring as in section 4.1 above.

- Molecule 1: Galectin-4

Chain A: 49% 26% 6% 18%

G1 S2 S3 G4 S5 S6 G7 H8 Q9 Q10 L11 N12 S13 L14 P15 T16 M17 E18 G19 P20 P21 N24 P25 P26 V27 P28 G36 L37 T38 A39 R40 R41 Y48 V49 F56 A57 F60 K61 V62 I68 A69 L70 H71 I72 N73 P74 R75 M76 G77 N78 G79 V82 R83

N84 S85 L86 L87 N88 G89 S90 W91 K96 K97 I98 T99 H100 N101 P102 F103 F108 F109 D110 L111 S112 C115 F120 K121 V122 F130 D131 F132 A133 A138 F139 Q140 R141 V142 L145 Q148 I158 S159 G160 P161 S162 S163 G164

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function, structures with the lowest energy, structures with the least restraint violations.*

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

Software name	Classification	Version
CYANA	structure solution	2.0.17
CYANA	refinement	2.0.17

No chemical shift data was provided.

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	1058	1059	1055	52±5
All	All	21160	21180	21100	1042

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 25.

5 of 212 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:57:ALA:HB2	1:A:73:ASN:OD1	0.87	1.68	8	16
1:A:36:GLY:O	1:A:38:THR:HG23	0.85	1.71	6	14
1:A:86:LEU:HD12	1:A:90:SER:O	0.84	1.72	8	20
1:A:98:ILE:HD11	1:A:100:HIS:O	0.83	1.74	6	20
1:A:86:LEU:HD13	1:A:91:TRP:NE1	0.81	1.90	20	20

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	134/164 (82%)	114±2 (85±1%)	18±2 (13±1%)	3±1 (2±1%)	10 49
All	All	2680/3280 (82%)	2272 (85%)	350 (13%)	58 (2%)	10 49

5 of 9 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	88	ASN	20
1	A	142	VAL	18
1	A	103	PHE	8
1	A	128	HIS	4
1	A	37	LEU	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	114/136 (84%)	101±2 (88±2%)	13±2 (12±2%)	9 52
All	All	2280/2720 (84%)	2016 (88%)	264 (12%)	9 52

5 of 27 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	76	MET	20
1	A	90	SER	20
1	A	132	PHE	20
1	A	73	ASN	19
1	A	110	ASP	19

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 monosaccharides 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)

No chemical shift data were provided