wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 06:23 PM EDT

PDB ID	:	2L7Y
BMRB ID	:	17381
Title	:	Solution structure of a putative surface protein
Authors	:	Wang, T.; Yuan, G.; Zhang, J.; Tu, X.
Deposited on	:	2010-12-27

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

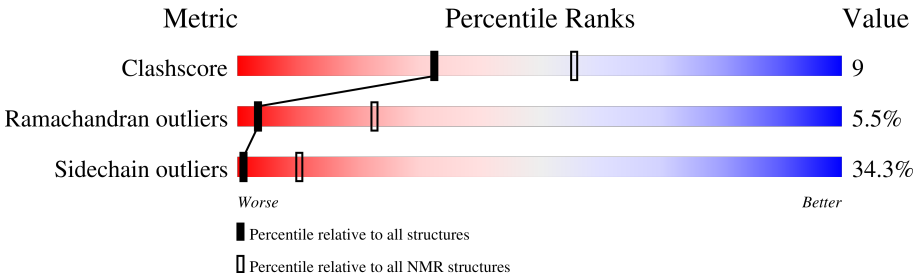
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 69%.

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	98	46% 32% • 10% 9%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 3 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:4-A:82 (79)	0.81	3

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 4 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1342 atoms, of which 652 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Putative endo-beta-N-acetylglucosaminidase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	89	1342	429	652	112	148	1	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	expression tag	UNP Q97S90
A	90	LEU	-	expression tag	UNP Q97S90
A	91	GLU	-	expression tag	UNP Q97S90
A	92	HIS	-	expression tag	UNP Q97S90
A	93	HIS	-	expression tag	UNP Q97S90
A	94	HIS	-	expression tag	UNP Q97S90
A	95	HIS	-	expression tag	UNP Q97S90
A	96	HIS	-	expression tag	UNP Q97S90
A	97	HIS	-	expression tag	UNP Q97S90

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: Putative endo-beta-N-acetylglucosaminidase

Chain A: 46% 32% • 10% 9%

MET S1 P2 K3 T4 I5 L6 Q12 K15 K16 D17 Y18 L19 S23 L24 D25 L26 S27 V33 T39 H43 S44 F45 T46 V50 E51 I52 S53 G54 Y55 Q58 K59 T60 G61 R62 L65 T66 L67 H68 Y69 Q70 V78 L79 V80 S81 P82 K83 A84 A85 L86

N87 D88 E89 LEU GLU HIS HIS HIS HIS HIS HIS

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

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: Putative endo-beta-N-acetylglucosaminidase

Chain A: 45% 32% • 10% 9%

MET S1 P2 K3 T4 I5 L6 G7 I8 E9 V10 S11 Q12 K15 K16 D17 Y18 L19 S23 L24 D25 L26 S27 E28 G29 T39 E42 H43 S44 F45 T46 D47 V50 S53 Q58 T66 L67 H68 Y69 V74 S75 L79 V80 S81 P82 K83 A84 A85 L86 N87

D88 E89 LEU GLU HIS HIS HIS HIS HIS HIS

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DGSA-distance geometry simulated annealing.*

Of the 250 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	
CYANA	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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	783
Number of shifts mapped to atoms	783
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	69%

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	616	577	577	11±3
All	All	12320	11540	11540	223

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 9.

5 of 115 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:52:ILE:HG21	1:A:65:LEU:HD23	0.89	1.43	18	11
1:A:24:LEU:HD21	1:A:78:VAL:HG21	0.85	1.47	19	1
1:A:5:ILE:HG21	1:A:33:VAL:HG22	0.80	1.54	2	2
1:A:65:LEU:HD11	1:A:78:VAL:HG12	0.65	1.66	16	1
1:A:52:ILE:CG2	1:A:65:LEU:HD23	0.62	2.25	9	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	79/98 (81%)	66±2 (84±3%)	8±2 (11±3%)	4±2 (6±2%)	3 23
All	All	1580/1960 (81%)	1324 (84%)	169 (11%)	87 (6%)	3 23

5 of 14 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	15	LYS	20
1	A	70	GLN	13
1	A	46	THR	12
1	A	25	ASP	10
1	A	24	LEU	7

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	69/86 (80%)	45±3 (66±5%)	24±3 (34±5%)	1 10
All	All	1380/1720 (80%)	906 (66%)	474 (34%)	1 10

5 of 56 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	53	SER	20
1	A	66	THR	20
1	A	39	THR	19
1	A	81	SER	18
1	A	4	THR	17

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)

The completeness of assignment taking into account all chemical shift lists is 69% for the well-defined parts and 68% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *17381_hy_zamtnutt.str*

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	783
Number of shifts mapped to atoms	783
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	85	0.25 ± 0.14	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	80	-0.00 ± 0.24	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	84	-0.53 ± 0.41	None needed (imprecise)

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 69%, i.e. 707 atoms were assigned a chemical shift out of a possible 1021. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	315/398 (79%)	161/163 (99%)	77/158 (49%)	77/77 (100%)
Sidechain	392/536 (73%)	265/347 (76%)	127/175 (73%)	0/14 (0%)

Continued on next page...

Continued from previous page...

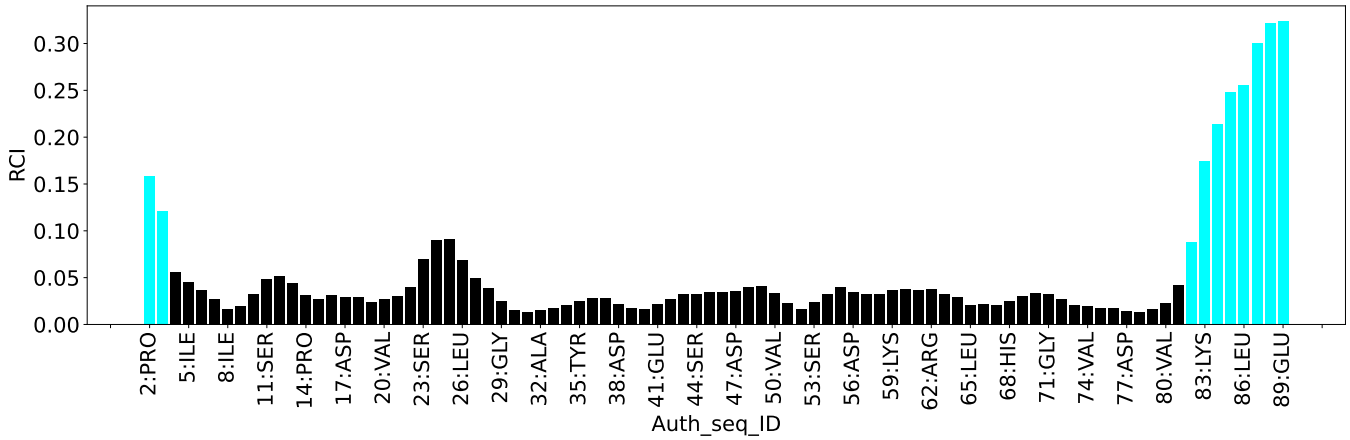
	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/87 (0%)	0/43 (0%)	0/41 (0%)	0/3 (0%)
Overall	707/1021 (69%)	426/553 (77%)	204/374 (55%)	77/94 (82%)

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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	43	HIS	HB3	0.66	1.18 – 4.91	-6.4
1	A	29	GLY	HA3	1.74	2.08 – 5.71	-5.9

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: