



wwPDB NMR Structure Validation Summary Report ⓘ

Feb 12, 2017 – 10:41 pm GMT

PDB ID : 2K33
Title : Solution structure of an N-glycosylated protein using in vitro glycosylation
Authors : Slynko, V.; Schubert, M.; Numao, S.; Kowarik, M.; Aebe, M.; Allain, F.H.-T.
Deposited on : 2008-04-18

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 ⓘ symbol.

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

Cyrange : Kirchner and Güntert (2011)
NmrClust : Kelley et al. (1996)
MolProbity : 4.02b-467
Mogul : 1.7.2 (RC1), CSD as538be (2017)
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : trunk28760
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : recalc28949

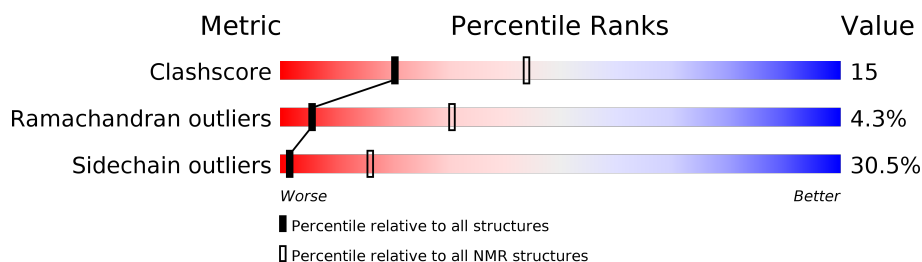
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 82%.

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	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	116	<div> <div style="width: 35%; background-color: green;"></div> <div style="width: 41%; background-color: yellow;"></div> <div style="width: 23%; background-color: cyan;"></div> </div>

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA and RNA chains that are outliers for geometric criteria:

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
2	A	A2G	118	5	-
2	A	A2G	119	9	-
2	A	A2G	120	5	-
2	A	A2G	121	9	-
2	A	A2G	122	10	-
2	A	B6D	117	7	-
2	A	BGC	123	14	-

2 Ensemble composition and analysis

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: *closest to the average*.

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:2-A:34, A:67-A:101 (68)	0.34	3
2	A:44-A:64 (21)	1.30	20

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 1 single-model cluster was found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1975 atoms, of which 979 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called AcrA.

Mol	Chain	Residues	Atoms					Trace
1	A	116	Total	C	H	N	O	0
			1787	566	888	157	176	

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	36	GLN	LYS	ENGINEERED	UNP Q8RTE5
A	50	GLN	LYS	ENGINEERED	UNP Q8RTE5
A	109	GLY	-	EXPRESSION TAG	UNP Q8RTE5
A	110	SER	-	EXPRESSION TAG	UNP Q8RTE5
A	111	HIS	-	EXPRESSION TAG	UNP Q8RTE5
A	112	HIS	-	EXPRESSION TAG	UNP Q8RTE5
A	113	HIS	-	EXPRESSION TAG	UNP Q8RTE5
A	114	HIS	-	EXPRESSION TAG	UNP Q8RTE5
A	115	HIS	-	EXPRESSION TAG	UNP Q8RTE5
A	116	HIS	-	EXPRESSION TAG	UNP Q8RTE5

- Molecule 2 is a polymer of unknown type called SUGAR (7-MER).

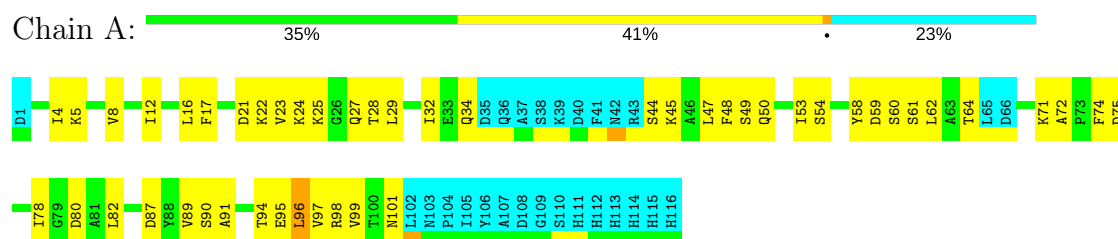
Mol	Chain	Residues	Atoms				
2	A	7	Total	C	H	N	O
			188	56	91	7	34

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.

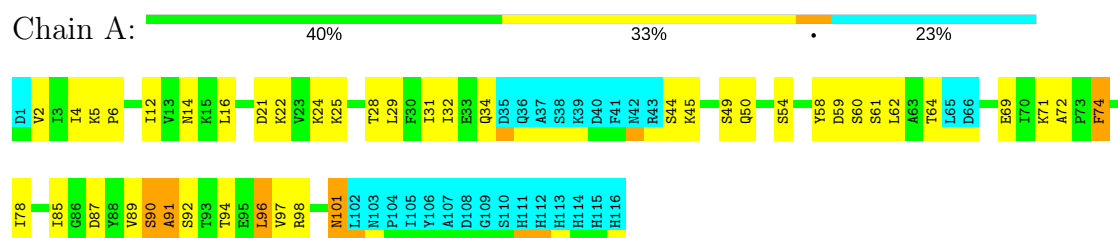
- Molecule 1: AcrA



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: AcrA



5 Refinement protocol and experimental data overview

The models were refined using the following method: *distance geometry*.

Of the 1000 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	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 15737
Number of chemical shift lists	1
Total number of shifts	1332
Number of shifts mapped to atoms	1221
Number of unparsed shifts	0
Number of shifts with mapping errors	111
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	82%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: B6D, BGC, A2G

There are no covalent bond-length or bond-angle outliers.

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
2	A	3.0±1.1	0.0±0.0
All	All	59	0

There are no bond-length outliers.

There are no bond-angle outliers.

5 of 7 unique chiral outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	123	BGC	C1	14
2	A	122	A2G	C1	10
2	A	119	A2G	C1	9
2	A	121	A2G	C1	9
2	A	117	B6D	C1	7

There are no planarity outliers.

6.2 Too-close contacts

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	674	700	700	23±6
2	A	97	91	85	2±1
All	All	15420	15820	15700	476

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 15.

5 of 244 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:29:LEU:HD11	1:A:72:ALA:HB2	0.98	1.35	16	14
1:A:89:VAL:HG12	1:A:94:THR:HG21	0.98	1.28	17	7
1:A:32:ILE:HD13	1:A:89:VAL:HG21	0.93	1.37	5	4
1:A:32:ILE:HD12	1:A:70:ILE:HD11	0.92	1.42	18	9
1:A:78:ILE:HD11	1:A:97:VAL:HG11	0.89	1.41	6	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	89/116 (77%)	70±3 (79±3%)	15±3 (17±3%)	4±1 (4±2%)	6	30
All	All	1780/2320 (77%)	1399 (79%)	304 (17%)	77 (4%)	6	30

5 of 18 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	91	ALA	15
1	A	25	LYS	10
1	A	101	ASN	8
1	A	44	SER	5
1	A	51	SER	4

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	76/100 (76%)	53±3 (69±5%)	23±3 (31±5%)	2	16
All	All	1520/2000 (76%)	1056 (69%)	464 (31%)	2	16

5 of 60 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	87	ASP	20
1	A	96	LEU	20
1	A	60	SER	19
1	A	34	GLN	17
1	A	27	GLN	17

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates ⓘ

7 carbohydrates are modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	B6D	A	117	1,2	14,16,17	1.42±0.01	0±0 (0±0%)
2	A2G	A	118	2	14,14,15	1.63±0.00	0±0 (0±0%)
2	A2G	A	119	2	14,14,15	1.64±0.00	0±0 (0±0%)
2	A2G	A	120	2	14,14,15	1.64±0.01	0±0 (0±0%)
2	A2G	A	121	2	14,14,15	1.64±0.01	0±0 (0±0%)
2	A2G	A	122	2	14,14,15	1.63±0.01	0±0 (0±0%)
2	BGC	A	123	2	11,11,12	1.30±0.01	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	B6D	A	117	1,2	16,22,24	0.84±0.00	0±0 (0±0%)
2	A2G	A	118	2	15,19,21	1.16±0.01	0±0 (0±0%)
2	A2G	A	119	2	15,19,21	1.16±0.01	0±0 (0±0%)
2	A2G	A	120	2	15,19,21	1.16±0.01	0±0 (0±0%)
2	A2G	A	121	2	15,19,21	1.16±0.01	0±0 (0±0%)
2	A2G	A	122	2	15,19,21	1.16±0.01	0±0 (0±0%)
2	BGC	A	123	2	13,15,17	1.30±0.01	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	B6D	A	117	1,2	-	0±0,8,25,28	0±0,1,1,1
2	A2G	A	118	2	-	0±0,6,23,26	0±0,1,1,1
2	A2G	A	119	2	-	0±0,6,23,26	0±0,1,1,1
2	A2G	A	120	2	-	0±0,6,23,26	0±0,1,1,1
2	A2G	A	121	2	-	0±0,6,23,26	0±0,1,1,1
2	A2G	A	122	2	-	0±0,6,23,26	0±0,1,1,1
2	BGC	A	123	2	-	0±0,2,19,22	0±0,1,1,1

There are no bond-length outliers.

There are no bond-angle outliers.

5 of 7 unique chiral outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	123	BGC	C1	14
2	A	122	A2G	C1	10
2	A	121	A2G	C1	9
2	A	119	A2G	C1	9
2	A	117	B6D	C1	7

There are no torsion outliers.

There are no ring outliers.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 82% for the well-defined parts and 77% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 15737

Chemical shift list name: *assigned_chem_shift_list_1*

7.1.1 Bookkeeping

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	1332
Number of shifts mapped to atoms	1221
Number of unparsed shifts	0
Number of shifts with mapping errors	111
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Chain not found in structure. First 5 (of 111) occurrences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
UNMAPPED	4	A2G	C2	52.196	0.2	1
UNMAPPED	1	X	C5	75.838	0.2	1
UNMAPPED	4	A2G	C8	24.962	0.2	1
UNMAPPED	6	BGC	H5	3.351	0.003	1
UNMAPPED	7	A2G	H8	2.003	0.002	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	112	-0.06 ± 0.27	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	104	0.02 ± 0.12	None needed (< 0.5 ppm)
$^{13}\text{C}'$	0	—	None (insufficient data)

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction \pm precision, ppm	Suggested action
^{15}N	106	-0.27 \pm 0.33	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 82%, i.e. 862 atoms were assigned a chemical shift out of a possible 1051. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	352/441 (80%)	176/176 (100%)	89/178 (50%)	87/87 (100%)
Sidechain	486/551 (88%)	296/318 (93%)	182/213 (85%)	8/20 (40%)
Aromatic	24/59 (41%)	24/32 (75%)	0/26 (0%)	0/1 (0%)
Overall	862/1051 (82%)	496/526 (94%)	271/417 (65%)	95/108 (88%)

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	42	ASN	ND2	130.75	124.24 – 101.34	7.8

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:

