Full wwPDB NMR Structure Validation Report (i)

Nov 9, 2024 – 02:05 pm GMT

PDB ID	:	2YHH
BMRB ID	:	17625
Title	:	Microvirin:mannobiose complex
Authors	:	Hussan, S.; Bewley, C.A.; Clore, G.M.; Gustchina, E.; Ghirlando, R.
Deposited on	:	2011-05-02

This is a Full wwPDB NMR Structure Validation 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:

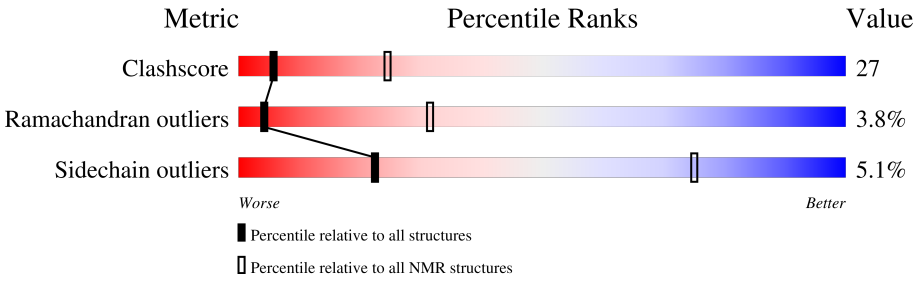
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

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 61%.

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	210492	14027
Ramachandran outliers	207382	12486
Sidechain outliers	206894	12463

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	108	63% 30% 6% ·
2	B	2	100%

2 Ensemble composition and analysis (i)

This entry contains 2 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1657 atoms, of which 784 are hydrogens and 0 are deuteriums.

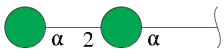
- Molecule 1 is a protein called MANNAN-BINDING LECTIN.

Mol	Chain	Residues	Atoms						Trace
1	A	108	Total	C	H	N	O	S	0
			1612	516	762	141	185	8	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	64	ARG	HIS	conflict	UNP Q2MDE2

- Molecule 2 is an oligosaccharide called alpha-D-mannopyranose-(1-2)-alpha-D-mannopyranose.



Mol	Chain	Residues	Atoms				Trace
2	B	2	Total	C	H	O	0
			45	12	22	11	

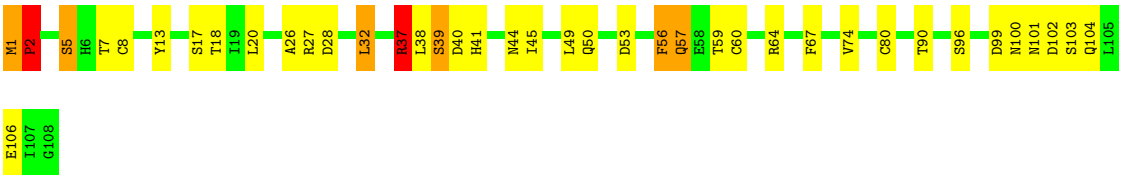
4 Residue-property plots

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.

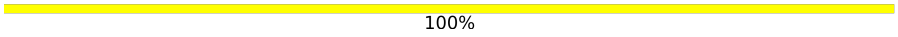
• Molecule 1: MANNAN-BINDING LECTIN

Chain A:



• Molecule 2: alpha-D-mannopyranose-(1-2)-alpha-D-mannopyranose

Chain B:



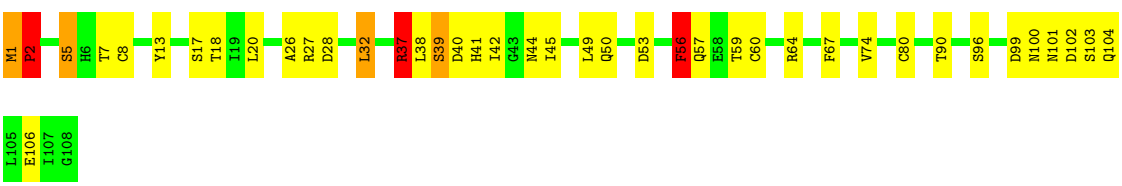
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

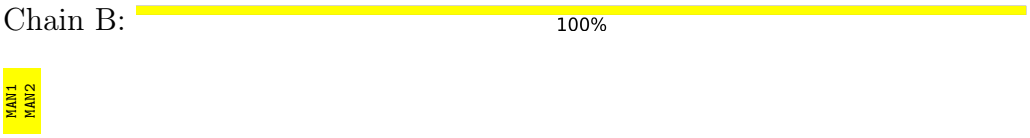
• Molecule 1: MANNAN-BINDING LECTIN

Chain A:



• Molecule 2: alpha-D-mannopyranose-(1-2)-alpha-D-mannopyranose

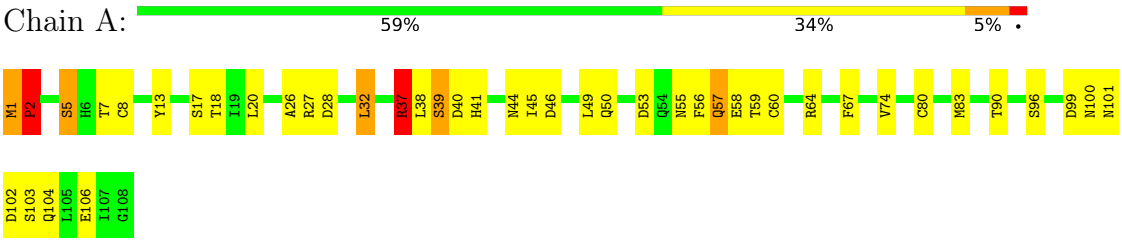
Chain B:



4.2.2 Score per residue for model 2

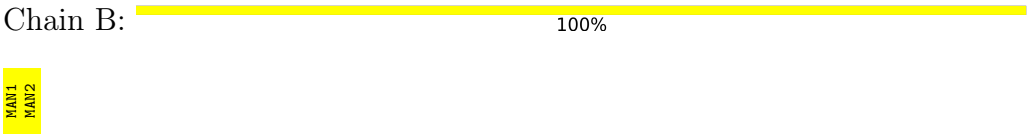
- Molecule 1: MANNAN-BINDING LECTIN

Chain A:



- Molecule 2: alpha-D-mannopyranose-(1-2)-alpha-D-mannopyranose

Chain B:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *TORSION ANGLE, SIMULATED ANNEALING*.

Of the 40 calculated structures, 2 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

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

Software name	Classification	Version
Xplor-NIH	refinement	
X-PLOR	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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	840
Number of shifts mapped to atoms	839
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	61%

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: MAN

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	1.19±0.00	0±0/865 (0.0± 0.0%)	0.99±0.05	2±2/1172 (0.2± 0.1%)
All	All	1.19	0/1730 (0.0%)	0.99	5/2344 (0.2%)

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
1	A	0.0±0.0	1.0±0.0
All	All	0	2

There are no bond-length outliers.

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
1	A	56	PHE	CB-CG-CD2	-10.13	113.71	120.80	1	1
1	A	56	PHE	CB-CG-CD1	9.32	127.33	120.80	1	1
1	A	56	PHE	CB-CA-C	-6.63	97.14	110.40	1	1
1	A	64	ARG	NE-CZ-NH2	-5.34	117.63	120.30	1	2

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	37	ARG	Sidechain	2

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	850	762	762	44±1
All	All	1746	1568	1566	88

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 27.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:37:ARG:CG	1:A:37:ARG:HH11	0.73	1.96	1	2
1:A:45:ILE:HG23	1:A:45:ILE:O	0.72	1.84	2	2
1:A:57:GLN:N	1:A:57:GLN:OE1	0.63	2.31	2	1
1:A:49:LEU:HD21	1:A:80:CYS:SG	0.63	2.33	1	2
1:A:37:ARG:CG	1:A:37:ARG:NH1	0.62	2.61	1	2
1:A:37:ARG:NH1	1:A:39:SER:OG	0.62	2.32	1	2
1:A:17:SER:O	1:A:37:ARG:NE	0.61	2.33	1	2
1:A:42:ILE:CG2	1:A:56:PHE:CD2	0.59	2.85	1	1
1:A:50:GLN:NE2	1:A:53:ASP:OD2	0.59	2.35	2	2
1:A:1:MET:O	1:A:2:PRO:O	0.57	2.23	1	2
1:A:56:PHE:CD1	1:A:57:GLN:N	0.57	2.73	1	1
1:A:44:ASN:ND2	1:A:59:THR:OG1	0.56	2.39	1	2
1:A:53:ASP:OD1	1:A:53:ASP:N	0.55	2.35	2	2
1:A:5:SER:O	1:A:8:CYS:O	0.55	2.24	1	2
1:A:103:SER:O	1:A:104:GLN:NE2	0.54	2.40	1	2
1:A:49:LEU:HD12	1:A:90:THR:OG1	0.54	2.03	1	2
1:A:46:ASP:OD1	1:A:83:MET:CE	0.53	2.57	2	1
1:A:45:ILE:O	1:A:45:ILE:CG2	0.53	2.54	1	2
1:A:38:LEU:O	1:A:40:ASP:N	0.53	2.42	1	2
1:A:37:ARG:NH1	1:A:37:ARG:HG2	0.52	2.20	1	2
1:A:37:ARG:NH1	1:A:39:SER:CB	0.49	2.75	1	2
1:A:42:ILE:HG22	1:A:56:PHE:CD2	0.48	2.44	1	1
1:A:8:CYS:SG	1:A:100:ASN:ND2	0.48	2.86	1	2
1:A:49:LEU:CD1	1:A:90:THR:OG1	0.48	2.61	1	2
1:A:57:GLN:N	1:A:57:GLN:CD	0.48	2.67	2	1
1:A:38:LEU:O	1:A:39:SER:C	0.47	2.52	1	2
1:A:56:PHE:HD1	1:A:56:PHE:H	0.47	1.53	1	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:13:TYR:CD2	1:A:13:TYR:O	0.47	2.68	1	2
1:A:50:GLN:CG	1:A:53:ASP:OD2	0.46	2.63	2	2
1:A:50:GLN:CD	1:A:53:ASP:OD2	0.46	2.55	2	2
1:A:7:THR:OG1	1:A:100:ASN:ND2	0.45	2.48	1	2
1:A:99:ASP:OD1	1:A:99:ASP:C	0.45	2.55	1	2
1:A:57:GLN:OE1	1:A:58:GLU:N	0.44	2.48	2	1
1:A:67:PHE:CE2	1:A:74:VAL:HG12	0.44	2.47	1	2
1:A:101:ASN:ND2	1:A:106:GLU:OE2	0.43	2.47	1	2
1:A:38:LEU:C	1:A:40:ASP:N	0.43	2.71	1	2
1:A:67:PHE:CD2	1:A:74:VAL:HG12	0.42	2.49	1	2
1:A:55:ASN:N	1:A:55:ASN:OD1	0.42	2.52	2	1
1:A:59:THR:C	1:A:60:CYS:SG	0.42	2.98	1	2
1:A:32:LEU:HD12	1:A:32:LEU:N	0.42	2.30	1	2
1:A:44:ASN:CG	1:A:59:THR:OG1	0.42	2.58	2	2
1:A:27:ARG:NH1	1:A:102:ASP:OD2	0.42	2.52	1	2
1:A:38:LEU:O	1:A:41:HIS:N	0.41	2.51	1	2
1:A:20:LEU:HD11	1:A:38:LEU:HD11	0.41	1.91	1	2
1:A:26:ALA:HB3	1:A:28:ASP:OD1	0.41	2.14	1	2
1:A:44:ASN:ND2	1:A:60:CYS:SG	0.41	2.93	2	1
1:A:42:ILE:HG21	1:A:56:PHE:CD2	0.41	2.50	1	1
1:A:57:GLN:CD	1:A:57:GLN:H	0.41	2.18	2	1
1:A:18:THR:O	1:A:18:THR:OG1	0.40	2.34	1	2
1:A:56:PHE:CD1	1:A:56:PHE:N	0.40	2.89	1	1
1:A:56:PHE:CG	1:A:57:GLN:N	0.40	2.90	2	1

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	106/108 (98%)	92±0 (87±0%)	10±0 (9±0%)	4±0 (4±0%)	4	31
All	All	212/216 (98%)	184 (87%)	20 (9%)	8 (4%)	4	31

All 4 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	2	PRO	2
1	A	5	SER	2
1	A	39	SER	2
1	A	96	SER	2

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	98/98 (100%)	93±0 (95±0%)	5±0 (5±0%)	22 75
All	All	196/196 (100%)	186 (95%)	10 (5%)	22 75

All 6 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	1	MET	2
1	A	2	PRO	2
1	A	32	LEU	2
1	A	37	ARG	2
1	A	56	PHE	1
1	A	57	GLN	1

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)

2 monosaccharides 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	MAN	B	1	2	12,12,12	0.71±0.28	0±0 (0±0%)
2	MAN	B	2	2	11,11,12	0.91±0.24	0±0 (4±4%)

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	MAN	B	1	2	17,17,17	1.37±0.13	2±0 (8±2%)
2	MAN	B	2	2	15,15,17	2.58±0.66	4±2 (23±16%)

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	MAN	B	1	2	-	0±0,2,22,22	0±0,1,1,1
2	MAN	B	2	2	-	0±0,2,19,22	0±0,1,1,1

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	B	2	MAN	C2-C3	2.22	1.55	1.52	1	1

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	2	MAN	O5-C1-C2	9.49	96.12	110.77	2	1
2	B	2	MAN	C1-O5-C5	7.06	121.75	112.19	1	2
2	B	1	MAN	C1-O5-C5	4.69	122.51	113.66	1	2
2	B	1	MAN	C4-C3-C2	3.89	104.04	110.82	2	1
2	B	2	MAN	C2-C3-C4	3.65	104.58	110.89	2	1
2	B	2	MAN	C6-C5-C4	2.57	106.99	113.00	2	1
2	B	2	MAN	C1-C2-C3	2.44	106.67	109.67	2	1
2	B	2	MAN	O5-C5-C6	2.19	110.63	107.20	2	1

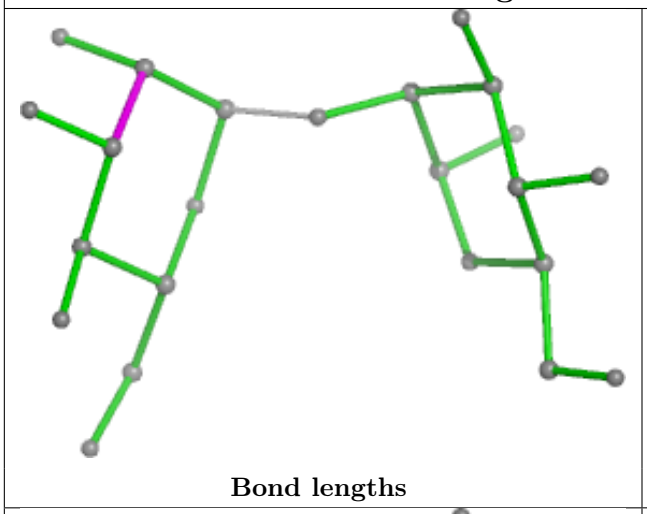
There are no chirality outliers.

There are no torsion outliers.

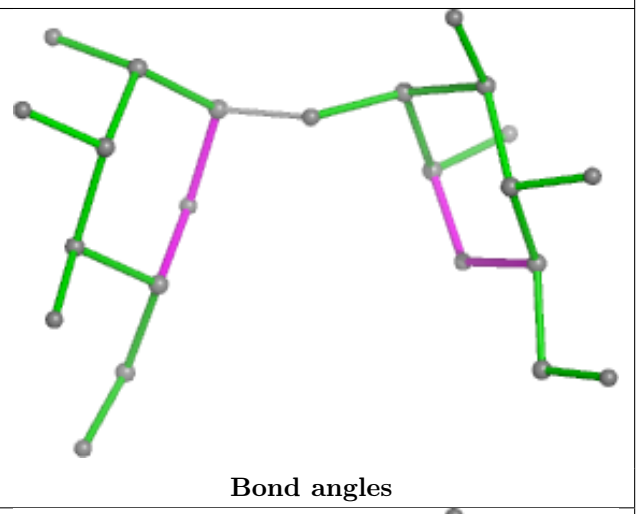
There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for oligosaccharide.

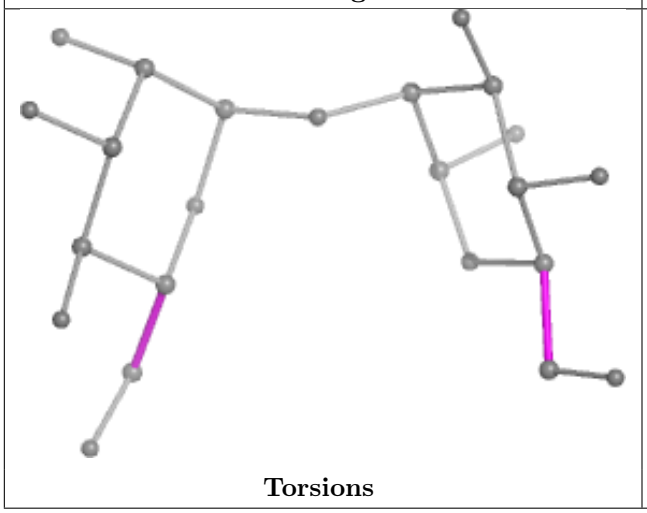
Oligosaccharide Chain B



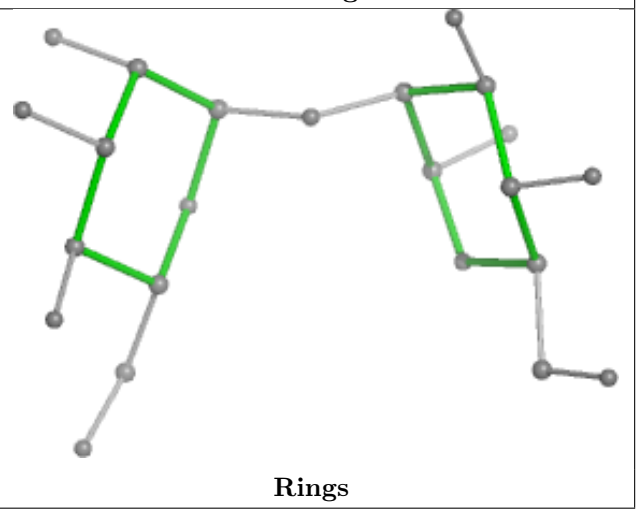
Bond lengths



Bond angles



Torsions



Rings

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 61% for the well-defined parts and 61% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chemical_shifts_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	840
Number of shifts mapped to atoms	839
Number of unparsed shifts	0
Number of shifts with mapping errors	1
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.

- No matching atom found in the structure. All 1 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	41	HIS	HE2	7.054	.	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	108	-0.21 ± 0.14	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	100	-0.03 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	106	0.33 ± 0.11	None needed (< 0.5 ppm)
^{15}N	102	-0.02 ± 0.43	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 61%, i.e. 840 atoms were assigned a chemical shift out of a possible 1377. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	517/542 (95%)	201/221 (91%)	214/216 (99%)	102/105 (97%)
Sidechain	290/736 (39%)	99/470 (21%)	175/240 (73%)	16/26 (62%)
Aromatic	33/99 (33%)	22/50 (44%)	8/44 (18%)	3/5 (60%)
Overall	840/1377 (61%)	322/741 (43%)	397/500 (79%)	121/136 (89%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 61%, i.e. 840 atoms were assigned a chemical shift out of a possible 1377. 0 out of 12 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	517/542 (95%)	201/221 (91%)	214/216 (99%)	102/105 (97%)
Sidechain	290/736 (39%)	99/470 (21%)	175/240 (73%)	16/26 (62%)
Aromatic	33/99 (33%)	22/50 (44%)	8/44 (18%)	3/5 (60%)
Overall	840/1377 (61%)	322/741 (43%)	397/500 (79%)	121/136 (89%)

7.1.4 Statistically unusual chemical shifts

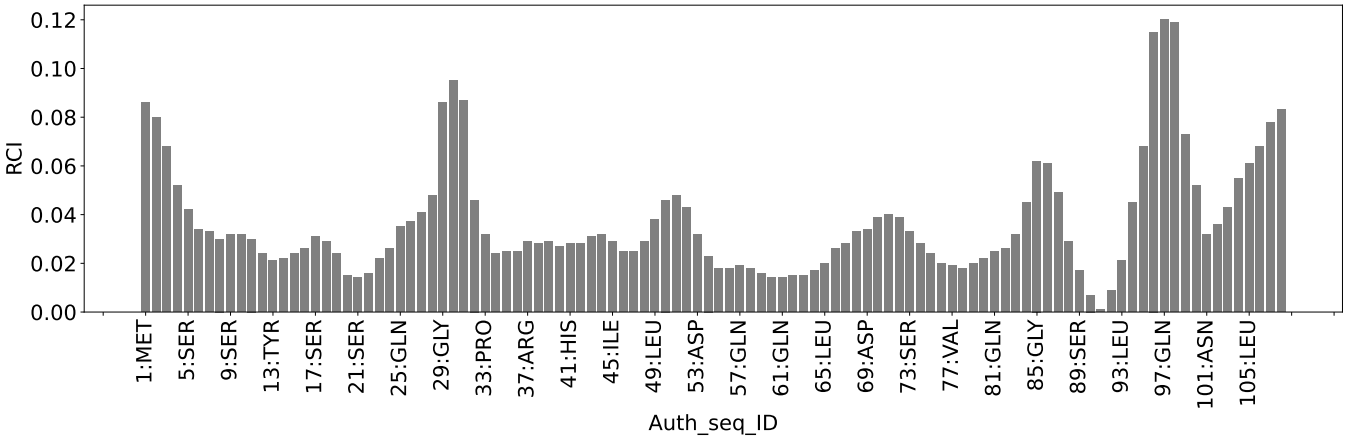
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	91	GLN	C	77.39	166.94 – 185.80	-52.5

7.1.5 Random Coil Index (RCI) plots

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:


RCI
0.12
0.10
0.08
0.06
0.04
0.02
0.00
1:MET
5:SER
9:SER
13:TYR
17:SER
21:SER
25:GLN
29:GLY
33:PRO
37:ARG
41:HIS
45:ILE
49:LEU
53:ASP
57:GLN
61:GLN
65:LEU
69:ASP
73:SER
77:VAL
81:GLN
85:GLY
89:SER
93:LEU
97:GLN
101:ASN
105:LEU
Auth_seq_ID