wwPDB NMR Structure Validation Summary Report

Apr 20, 2024 – 08:41 AM EDT

PDB ID	:	5UE5
Title	:	proMMP-7 with heparin octasaccharide bound to the catalytic domain
Authors	:	Fulcher, Y.G.; Prior, S.H.; Linhardt, R.J.; Van Doren, S.R.
Deposited on	:	2016-12-29

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 (1)) were used in the production of this report:

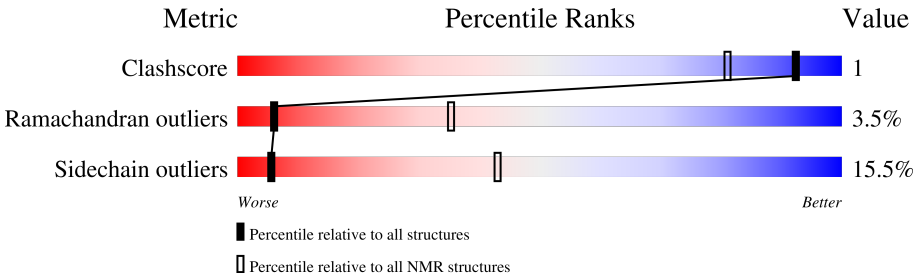
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 86%.

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	247	65% 13% • 20%
2	B	8	100%

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	B	IDS	4	1	-
2	B	SGN	7	1	-

2 Ensemble composition and analysis

This entry contains 16 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:15-A:23, A:35-A:71, A:81-A:216, A:224-A:239 (198)	0.65	1

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 5 single-model clusters were found.

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

3 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 4032 atoms, of which 1945 are hydrogens and 0 are deuteriums.

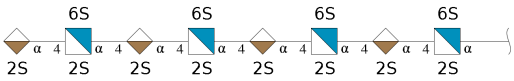
- Molecule 1 is a protein called Matrilysin.

Mol	Chain	Residues	Atoms						Trace
1	A	247	Total	C	H	N	O	S	0
			3828	1234	1885	338	362	9	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	195	ALA	GLU	conflict	UNP P09237

- Molecule 2 is an oligosaccharide called 2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose.



Mol	Chain	Residues	Atoms						Trace
2	B	8	Total	C	H	N	O	S	0
			200	48	60	4	76	12	

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	
3	A	2	Total	Ca
			2	2

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
4	A	2	Total	Zn
			2	2

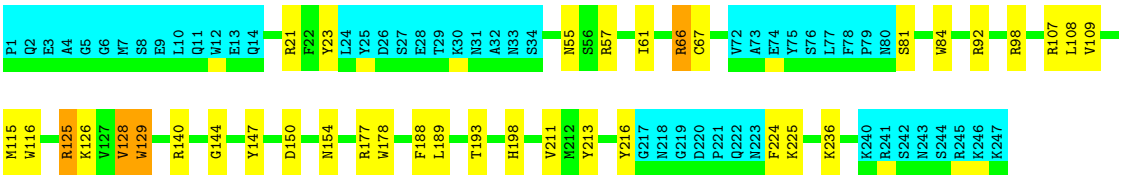
4 Residue-property plots

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.

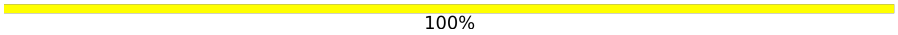
• Molecule 1: Matrilysin

Chain A:



• Molecule 2: 2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose

Chain B:

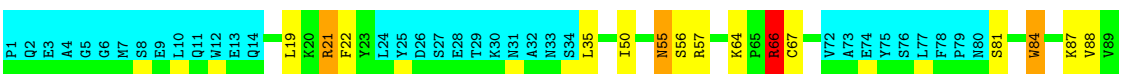


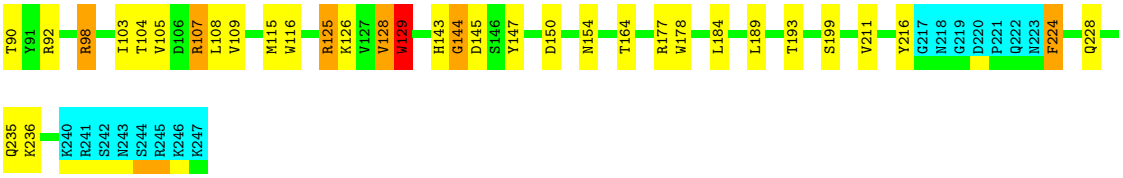
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

• Molecule 1: Matrilysin

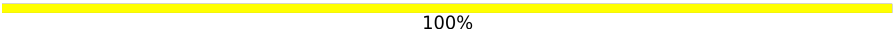
Chain A:





• Molecule 2: 2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose-(1-4)-2-O-sulfo-alpha-L-idopyranuronic acid-(1-4)-2-deoxy-6-O-sulfo-2-(sulfoamino)-alpha-D-glucopyranose

Chain B:



5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 16 were deposited, based on the following criterion: *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 calculation	2.1
SYBYL	refinement	X 2.1.1

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	2746
Number of shifts mapped to atoms	2746
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	86%

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: SGN, ZN, IDS, CA

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	0.99±0.01	0±0/1603 (0.0± 0.0%)	1.19±0.02	13±1/2171 (0.6± 0.0%)
All	All	0.99	1/25648 (0.0%)	1.19	210/34736 (0.6%)

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	1.2±1.1
All	All	0	19

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	84	TRP	CG-CD2	5.06	1.52	1.43	5	1

5 of 20 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	125	ARG	NE-CZ-NH1	8.65	124.63	120.30	10	16
1	A	107	ARG	NE-CZ-NH1	8.21	124.41	120.30	15	12
1	A	21	ARG	NE-CZ-NH1	7.62	124.11	120.30	13	15
1	A	116	TRP	CE2-CD2-CG	-7.49	101.31	107.30	7	16
1	A	57	ARG	NE-CZ-NH1	7.40	124.00	120.30	6	16

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	A	216	TYR	Sidechain	7
1	A	213	TYR	Sidechain	5
1	A	23	TYR	Sidechain	3
1	A	147	TYR	Sidechain	1
1	A	160	PHE	Sidechain	1

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	1557	1528	1527	4±2
2	B	140	60	45	0±0
All	All	27216	25408	25152	74

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 1.

5 of 49 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:46:PHE:CD2	1:A:61:ILE:HD12	0.60	2.32	3	2
1:A:23:TYR:CG	1:A:23:TYR:O	0.57	2.58	2	2
1:A:194:HIS:CD2	1:A:198:HIS:NE2	0.54	2.75	5	3
1:A:172:PHE:CD2	1:A:188:PHE:CE2	0.53	2.96	5	1
1:A:172:PHE:CE2	1:A:188:PHE:CZ	0.52	2.98	11	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	198/247 (80%)	169±3 (85±1%)	22±3 (11±2%)	7±2 (3±1%)	6 35
All	All	3168/3952 (80%)	2705 (85%)	353 (11%)	110 (3%)	6 35

5 of 20 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	55	ASN	16
1	A	147	TYR	16
1	A	154	ASN	16
1	A	144	GLY	10
1	A	66	ARG	8

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	163/205 (80%)	138±3 (85±2%)	25±3 (15±2%)	5 43
All	All	2608/3280 (80%)	2205 (85%)	403 (15%)	5 43

5 of 81 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	128	VAL	16
1	A	129	TRP	16
1	A	211	VAL	16
1	A	236	LYS	16
1	A	108	LEU	15

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)

8 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	SGN	B	1	2	18,19,20	3.77±0.05	4±1 (19±3%)
2	IDS	B	2	2	16,16,17	1.29±0.03	1±0 (6±0%)
2	SGN	B	3	2	18,19,20	3.72±0.12	4±1 (22±3%)
2	IDS	B	4	2	16,16,17	1.30±0.04	1±0 (6±0%)
2	SGN	B	5	2	18,19,20	3.78±0.05	5±0 (26±2%)
2	IDS	B	6	2	16,16,17	1.28±0.03	1±0 (6±1%)
2	SGN	B	7	2	18,19,20	3.73±0.05	4±1 (23±3%)
2	IDS	B	8	2	16,16,17	1.19±0.02	1±0 (7±2%)

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	SGN	B	1	2	22,29,31	1.42±0.16	3±1 (15±5%)
2	IDS	B	2	2	17,24,26	1.22±0.12	1±1 (6±4%)
2	SGN	B	3	2	22,29,31	1.90±0.21	7±1 (31±6%)
2	IDS	B	4	2	17,24,26	1.09±0.12	1±1 (5±5%)
2	SGN	B	5	2	22,29,31	1.57±0.17	4±1 (19±5%)

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	IDS	B	6	2	17,24,26	1.00±0.14	0±1 (2±5%)
2	SGN	B	7	2	22,29,31	1.55±0.16	4±1 (17±6%)
2	IDS	B	8	2	17,24,26	1.12±0.13	1±0 (6±1%)

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	SGN	B	1	2	-	0±0,11,28,31	0±0,1,1,1
2	IDS	B	2	2	-	0±0,9,26,29	0±0,1,1,1
2	SGN	B	3	2	-	0±0,11,28,31	0±0,1,1,1
2	IDS	B	4	2	-	0±0,9,26,29	0±0,1,1,1
2	SGN	B	5	2	-	0±0,11,28,31	0±0,1,1,1
2	IDS	B	6	2	-	0±0,9,26,29	0±0,1,1,1
2	SGN	B	7	2	-	0±0,11,28,31	0±0,1,1,1
2	IDS	B	8	2	-	0±0,9,26,29	0±0,1,1,1

5 of 26 unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
2	B	3	SGN	S1-N2	15.70	1.80	1.59	1	16
2	B	7	SGN	S1-N2	15.35	1.80	1.59	16	16
2	B	5	SGN	S1-N2	15.23	1.80	1.59	4	16
2	B	1	SGN	S1-N2	15.14	1.80	1.59	9	16
2	B	5	SGN	O1S-S1	3.46	1.46	1.42	5	16

5 of 57 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	3	SGN	O6-C6-C5	5.62	118.11	107.62	11	13
2	B	3	SGN	C1-O5-C5	4.93	118.87	112.19	3	13
2	B	8	IDS	O2-C2-C3	4.65	113.45	106.95	7	12
2	B	3	SGN	O1S-S1-O2S	4.43	109.69	120.16	4	8
2	B	3	SGN	O5-C5-C6	4.30	117.13	107.61	11	1

All unique chiral outliers are listed below. They are sorted by the frequency of occurrence in the

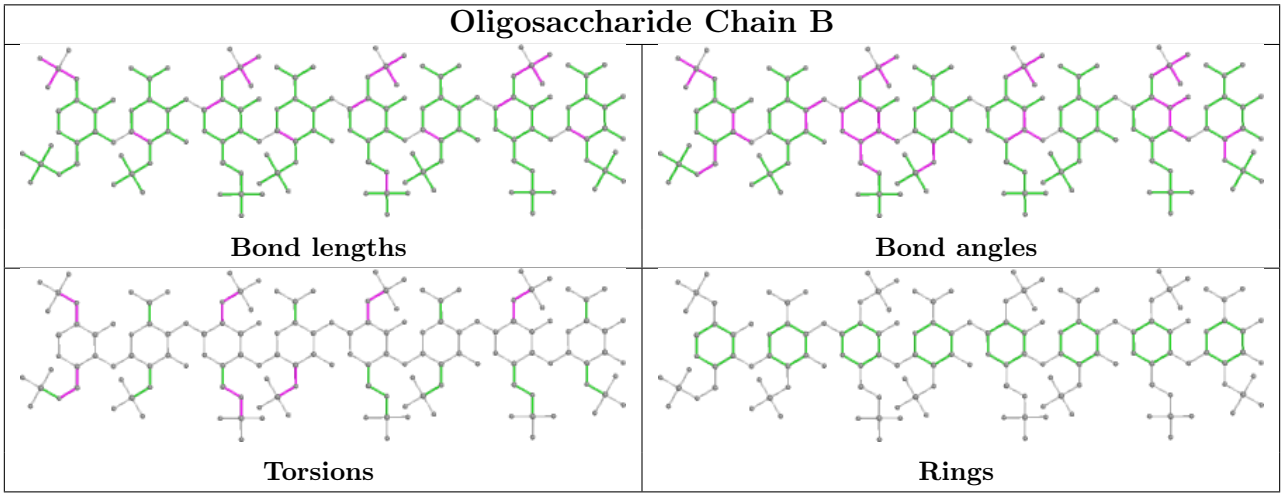
ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	B	4	IDS	C1	1
2	B	7	SGN	C1	1

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.



6.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

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

The completeness of assignment taking into account all chemical shift lists is 86% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

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	2746
Number of shifts mapped to atoms	2746
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	15

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	236	-0.28 ± 0.17	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	212	0.28 ± 0.13	None needed (< 0.5 ppm)
$^{13}C'$	232	0.03 ± 0.11	None needed (< 0.5 ppm)
^{15}N	220	-0.11 ± 0.24	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 86%, i.e. 2302 atoms were assigned a chemical shift out of a possible 2675. 0 out of 27 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	956/989 (97%)	389/406 (96%)	386/396 (97%)	181/187 (97%)
Sidechain	1178/1401 (84%)	797/916 (87%)	371/435 (85%)	10/50 (20%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	168/285 (59%)	87/142 (61%)	77/131 (59%)	4/12 (33%)
Overall	2302/2675 (86%)	1273/1464 (87%)	834/962 (87%)	195/249 (78%)

7.1.4 Statistically unusual chemical shifts (i)

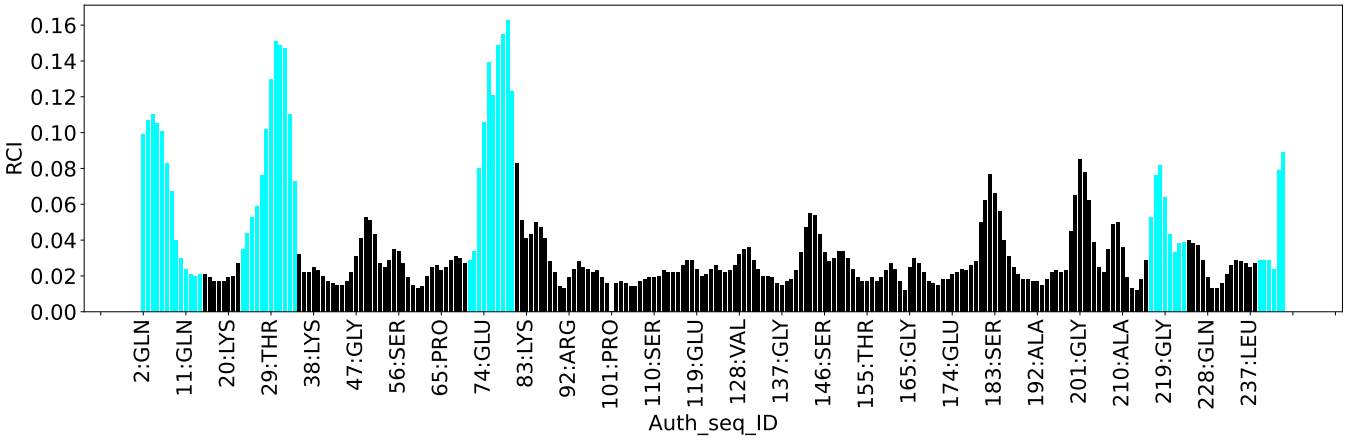
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	107	ARG	NE	116.85	76.53 – 92.65	20.0
1	A	57	ARG	NE	115.24	76.53 – 92.65	19.0
1	A	92	ARG	NE	114.04	76.53 – 92.65	18.3
1	A	66	ARG	NE	112.31	76.53 – 92.65	17.2
1	A	162	PRO	HB2	-0.32	0.37 – 3.78	-7.0
1	A	60	GLU	HB2	0.60	1.00 – 3.05	-6.9
1	A	247	LYS	CE	36.10	37.57 – 46.21	-6.7
1	A	60	GLU	HB3	0.64	0.95 – 3.05	-6.5
1	A	60	GLU	CG	43.68	30.20 – 42.01	6.4
1	A	65	PRO	HA	2.44	2.78 – 6.00	-6.1
1	A	3	GLU	CG	43.11	30.20 – 42.01	5.9
1	A	62	MET	HG2	0.38	0.65 – 4.19	-5.8
1	A	242	SER	N	135.65	99.14 – 133.45	5.6
1	A	211	VAL	H	11.59	4.98 – 11.56	5.0
1	A	153	GLY	H	11.43	5.23 – 11.42	5.0

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:


RCI
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0.00
2:GLN
11:GLN
20:LYS
29:THR
38:LYS
47:GLY
56:SER
65:PRO
74:GLU
83:LYS
92:ARG
101:PRO
110:SER
119:GLU
128:VAL
137:GLY
146:SER
155:THR
165:GLY
174:GLU
183:SER
192:ALA
201:GLY
210:ALA
219:GLY
228:GLN
237:LEU
Auth_seq_ID