Full wwPDB NMR Structure Validation Report (i)

May 29, 2020 – 07:17 am BST

PDB ID : 5J7J

Title : NMR Derived Structure of Ca2+ Calmodulin bound to Phosphorylated PSD-95

Authors : Turner, M.L.; Ames, J.B.; Anderson, D.E.

Deposited on : 2016-04-06

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

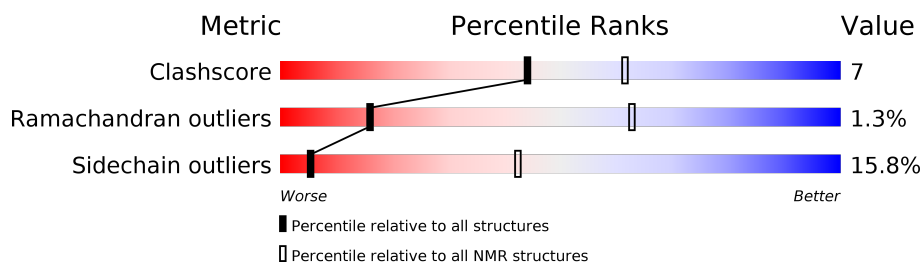|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 56%.

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	Percentile Ranks	Value
Clashscore		7
Ramachandran outliers		1.3%
Sidechain outliers		15.8%

Worse   *Better*

■ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	148	64% 32% • •
2	B	19	58% 37% 5%

2 Ensemble composition and analysis (i)

This entry contains 4 models. Model 4 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:147, B:1-B:18 (162)	0.36	4

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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	2, 3
2	1, 4

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2518 atoms, of which 1217 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Calmodulin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	144	2198	696	1064	182	247	9	0

- Molecule 2 is a protein called Disks large homolog 4.

Mol	Chain	Residues	Atoms							Trace
			Total	C	H	N	O	P	S	
2	B	19	316	98	153	25	36	1	3	0

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

Mol	Chain	Residues	Atoms	
			Total	Ca
3	A	4	4	4

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: Calmodulin

Chain A:

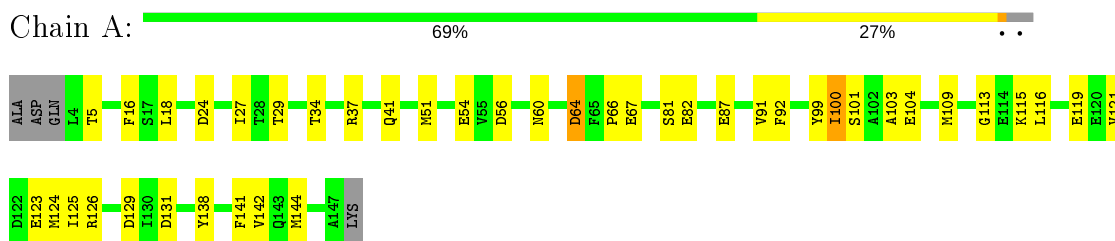- Molecule 2: Disks large homolog 4

Chain B:
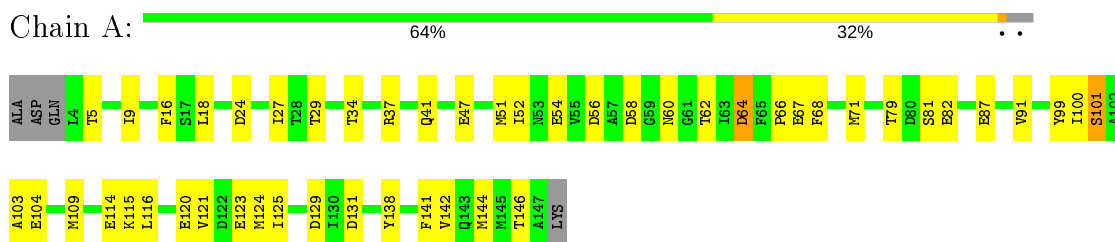
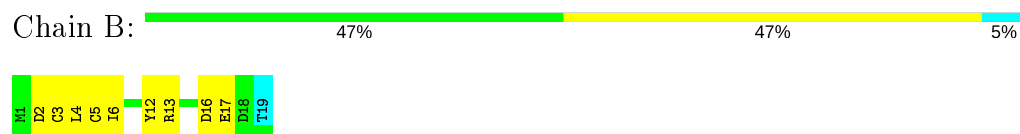## 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

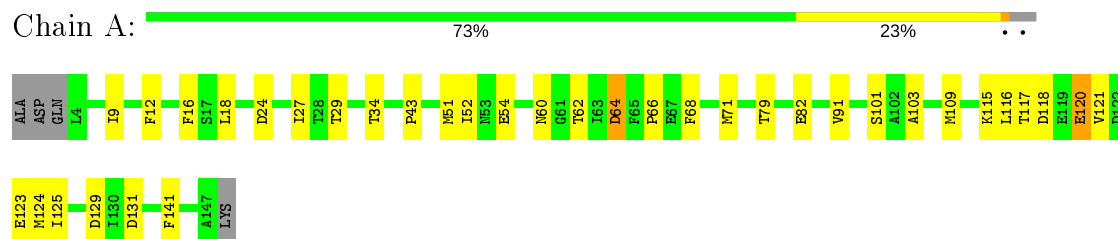### 4.2.1 Score per residue for model 1

- Molecule 1: Calmodulin

Chain A:

- Molecule 2: Disks large homolog 4

Chain B:　　　　　　　　　　47%　　　　　　　　　　47%　　　　　5%

M1 D2 C3 L4 C5 I6 Y12 R13 D16 E17 D18 T19

4.2.2　Score per residue for model 2

• Molecule 1: Calmodulin

Chain A:　　　　　　　　　　73%　　　　　　　　　　23%

ALA ASP GLN L4 I9 F12 F16 S17 L18 D24 I27 T28 T29 T34 P43 M51 I52 N53 E54 N60 G61 T62 I63 D64 F65 P66 E67 F68 M71 T79 E82 V91 S101 A102 A103 M109 K115 L116 T117 D118 E119 E120 V121 D122

E123 M124 I125 D129 I130 D131 F141 A147 LYS

• Molecule 2: Disks large homolog 4

Chain B:　　　　　　　　　　84%　　　　　　　　　　11%　　　5%

M1 D2 C3 Y12 T19
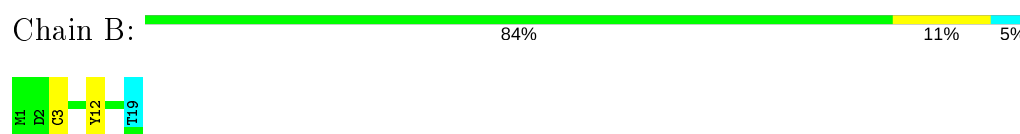
4.2.3　Score per residue for model 3

• Molecule 1: Calmodulin

Chain A:　　　　　　　　　　72%　　　　　　　　　　23%

ALA ASP GLN L4 E7 Q8 I9 F16 D24 T29 T44 E47 M51 I52 N53 E54 D58 G59 N60 D64 F65 P66 E67 F68 M71 T79 E83 E84 D93 I100 S101 A102 A103 E104 L105 V108 M109 E114 K115 L116 V121 D122 E123

M124 I125 D131 Y138 F141 V142 T146 A147 LYS

• Molecule 2: Disks large homolog 4

Chain B:　　　　　　　　　　74%　　　　　　　　16%　　5%　5%

M1 D2 C3 Y12 T19
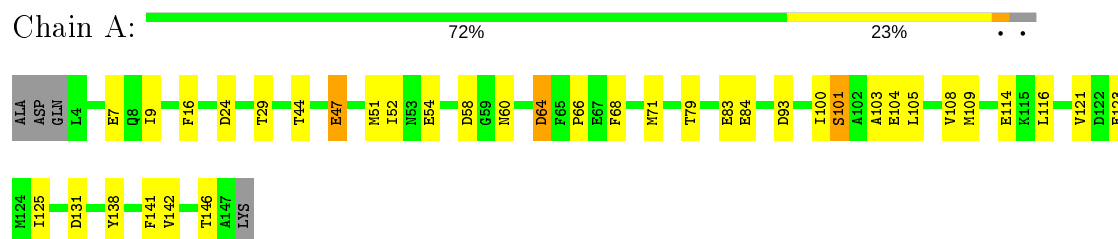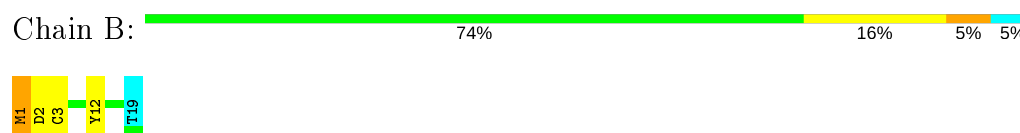
4.2.4　Score per residue for model 4 (medoid)

• Molecule 1: Calmodulin

Chain A:

Chain A sequence bar: 65% green, 31% yellow, · ·

Chain A residues: ALA ASP GLN L4 T5 E6 I9 F16 S17 L18 D24 I27 R37 Q41 M51 I52 N53 E54 V55 D56 A57 D58 G59 N60 G61 T62 I63 D64 E67 F68 L69 M72 K75 D78 T79 D80 S81 E82 E87 A88 V91 Y99 I100 S101 A102 A103

E104 M109 E114 K115 L116 E123 M124 I125 D129 I130 D131 G132 D133 F141 M144 M145 T146 A147 LYS

• Molecule 2: Disks large homolog 4

Chain B:

Chain B sequence bar: 47% green, 26% yellow, 21% orange, 5% cyan

Chain B residues: M1 D2 C3 L4 C5 I6 V7 T8 K11 Y12 R13 T19

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 4 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
HADDOCK	structure calculation	
HADDOCK	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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1164
Number of shifts mapped to atoms	1164
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	56%

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

6 Model quality (i)

6.1 Standard geometry (i)

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

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	1134	1064	1063	16±3
2	B	152	147	149	6±3
All	All	5160	4844	4848	69

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 7.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:116:LEU:HD23	1:A:120:GLU:HB3	0.70	1.61	2	1
1:A:116:LEU:HD11	2:B:13:ARG:HD3	0.65	1.69	1	2
1:A:109:MET:HG2	1:A:116:LEU:HD12	0.64	1.67	4	2
1:A:51:MET:O	1:A:54:GLU:HB3	0.63	1.92	4	4
1:A:141:PHE:HE2	2:B:12:TYR:CD2	0.59	2.15	4	3
1:A:141:PHE:CE2	2:B:12:TYR:CD2	0.58	2.92	3	4
1:A:75:LYS:HB3	2:B:7:VAL:HG11	0.57	1.76	4	1
1:A:37:ARG:HA	1:A:41:GLN:O	0.55	2.01	1	2
1:A:51:MET:SD	2:B:1:MET:HG2	0.55	2.42	3	1
1:A:12:PHE:HB3	1:A:68:PHE:CE2	0.52	2.39	2	1
1:A:124:MET:SD	2:B:12:TYR:CE1	0.52	3.03	2	2

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:A:87:GLU:O	1:A:91:VAL:HG23	0.52	2.05	4	2
1:A:64:ASP:O	1:A:68:PHE:HB2	0.51	2.05	2	1
1:A:64:ASP:OD1	1:A:66:PRO:HD2	0.51	2.05	1	3
1:A:101:SER:C	1:A:103:ALA:H	0.50	2.09	3	4
1:A:104:GLU:O	1:A:108:VAL:HG23	0.50	2.07	3	1
1:A:124:MET:SD	2:B:12:TYR:HE1	0.49	2.30	1	2
1:A:47:GLU:O	1:A:51:MET:HG2	0.48	2.08	3	1
2:B:2:ASP:HA	2:B:5:CYS:SG	0.48	2.48	1	2
1:A:121:VAL:O	1:A:125:ILE:HG12	0.48	2.09	3	3
1:A:100:ILE:HG22	1:A:104:GLU:OE1	0.47	2.08	4	1
1:A:145:MET:SD	2:B:8:THR:HG23	0.47	2.48	4	1
1:A:138:TYR:O	1:A:142:VAL:HG23	0.47	2.09	3	2
1:A:125:ILE:O	1:A:129:ASP:HB2	0.46	2.10	1	3
1:A:141:PHE:HE2	2:B:12:TYR:CE2	0.46	2.29	1	1
1:A:123:GLU:HA	1:A:126:ARG:NH2	0.45	2.26	1	1
2:B:4:LEU:HG	2:B:5:CYS:N	0.45	2.26	4	1
1:A:115:LYS:O	1:A:116:LEU:HG	0.44	2.12	2	1
1:A:44:THR:H	1:A:47:GLU:HB2	0.44	1.71	3	1
1:A:71:MET:SD	2:B:3:CYS:HB3	0.43	2.54	2	1
2:B:1:MET:N	2:B:4:LEU:HD23	0.43	2.28	4	1
1:A:92:PHE:O	1:A:100:ILE:HG12	0.43	2.14	1	1
1:A:69:LEU:O	1:A:72:MET:HB3	0.43	2.14	4	1
1:A:105:LEU:HD23	1:A:121:VAL:HG13	0.42	1.90	3	1
1:A:56:ASP:HA	1:A:67:GLU:CD	0.42	2.36	1	2
1:A:6:GLU:O	1:A:9:ILE:HB	0.41	2.16	4	1
1:A:88:ALA:CB	2:B:8:THR:HG21	0.41	2.45	4	1
1:A:145:MET:O	2:B:11:LYS:HE3	0.41	2.15	4	1
1:A:52:ILE:HG21	1:A:62:THR:HA	0.41	1.93	2	1
2:B:1:MET:SD	2:B:2:ASP:N	0.41	2.89	3	1
1:A:52:ILE:CG2	1:A:62:THR:HA	0.40	2.46	2	1
1:A:141:PHE:CE2	2:B:12:TYR:HD2	0.40	2.35	2	1
1:A:55:VAL:HG12	1:A:67:GLU:HB3	0.40	1.93	4	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	142/148 (96%)	122±2 (86±1%)	19±1 (13±1%)	2±1 (1±0%)	18	66
2	B	17/19 (89%)	14±1 (82±4%)	3±1 (15±3%)	1±1 (3±5%)	7	41
All	All	636/668 (95%)	543 (85%)	85 (13%)	8 (1%)	16	63

All 5 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	82	GLU	3
1	A	101	SER	2
2	B	17	GLU	1
1	A	113	GLY	1
2	B	16	ASP	1

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	123/126 (98%)	103±3 (84±2%)	20±3 (16±2%)	5	41
2	B	18/18 (100%)	16±1 (88±8%)	2±1 (12±8%)	8	50
All	All	564/576 (98%)	475 (84%)	89 (16%)	5	42

All 48 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	64	ASP	4
1	A	131	ASP	4
1	A	24	ASP	4
1	A	16	PHE	4
1	A	60	ASN	4
1	A	9	ILE	3
1	A	27	ILE	3
1	A	29	THR	3
1	A	79	THR	3
1	A	123	GLU	3
1	A	18	LEU	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	68	PHE	2
1	A	109	MET	2
1	A	5	THR	2
2	B	4	LEU	2
1	A	100	ILE	2
2	B	1	MET	2
1	A	144	MET	2
1	A	146	THR	2
1	A	81	SER	2
1	A	114	GLU	2
1	A	99	TYR	2
2	B	3	CYS	2
1	A	34	THR	2
1	A	58	ASP	2
1	A	78	ASP	1
1	A	120	GLU	1
1	A	83	GLU	1
1	A	71	MET	1
1	A	93	ASP	1
1	A	43	PRO	1
2	B	6	ILE	1
1	A	84	GLU	1
1	A	133	ASP	1
2	B	13	ARG	1
1	A	119	GLU	1
1	A	62	THR	1
1	A	7	GLU	1
1	A	52	ILE	1
2	B	5	CYS	1
1	A	115	LYS	1
1	A	104	GLU	1
1	A	116	LEU	1
1	A	91	VAL	1
1	A	53	ASN	1
1	A	117	THR	1
1	A	118	ASP	1
1	A	47	GLU	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)

1 non-standard protein/DNA/RNA residue is 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	TPO	B	19	2	8,10,11	1.01±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	TPO	B	19	2	10,14,16	1.22±0.02	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	TPO	B	19	2	-	0±0,9,11,13	-

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.5 Carbohydrates (i)

There are no carbohydrates in this entry.

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 (i)

The completeness of assignment taking into account all chemical shift lists is 56% for the well-defined parts and 56% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

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	1164
Number of shifts mapped to atoms	1164
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
$^{13}C_\alpha$	147	-0.26 ± 0.09	None needed ($<$ 0.5 ppm)
$^{13}C_\beta$	136	0.14 ± 0.06	None needed ($<$ 0.5 ppm)
$^{13}C'$	138	-0.47 ± 0.11	None needed ($<$ 0.5 ppm)
^{15}N	143	-0.33 ± 0.25	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 56%, i.e. 1080 atoms were assigned a chemical shift out of a possible 1942. 0 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	700/806 (87%)	279/322 (87%)	280/324 (86%)	141/160 (88%)
Sidechain	380/1025 (37%)	225/592 (38%)	155/391 (40%)	0/42 (0%)

Continued on next page...

Continued from previous page...

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Aromatic	0/111 (0%)	0/60 (0%)	0/50 (0%)	0/1 (0%)
Overall	1080/1942 (56%)	504/974 (52%)	435/765 (57%)	141/203 (69%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 56%, i.e. 1080 atoms were assigned a chemical shift out of a possible 1942. 0 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	700/806 (87%)	279/322 (87%)	280/324 (86%)	141/160 (88%)
Sidechain	380/1025 (37%)	225/592 (38%)	155/391 (40%)	0/42 (0%)
Aromatic	0/111 (0%)	0/60 (0%)	0/50 (0%)	0/1 (0%)
Overall	1080/1942 (56%)	504/974 (52%)	435/765 (57%)	141/203 (69%)

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	93	ASP	HB2	1.21	4.07 − 1.37	-5.6
1	A	20	ASP	HB2	1.35	4.07 − 1.37	-5.1

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: