Full wwPDB NMR Structure Validation Report (i)

Jun 11, 2024 – 09:03 PM EDT

PDB ID	:	2K2J
BMRB ID	:	15707
Title	:	NMR solution structure of the split PH domain from Phospholipase C gamma 2
Authors	:	Harris, R.; Bunney, T.D.; Katan, M.; Driscoll, P.C.
Deposited on	:	2008-04-02

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:

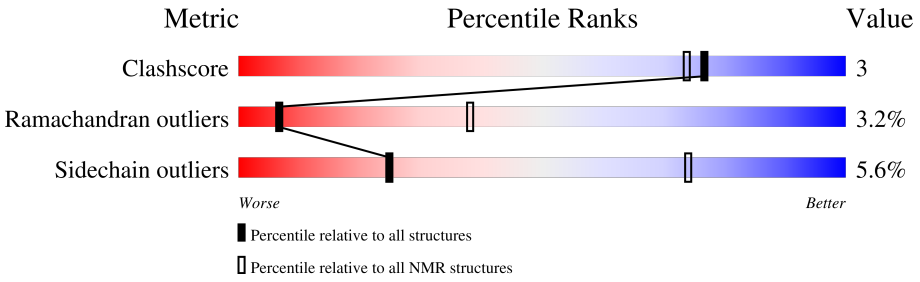
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson 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 92%.

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	124	

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 16 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:14-A:47, A:61-A:119 (93)	0.53	16

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 4 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1905 atoms, of which 934 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	117	1905	616	934	161	190	4	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP P16885
A	2	GLY	-	expression tag	UNP P16885
A	3	GLY	-	expression tag	UNP P16885
A	4	SER	-	expression tag	UNP P16885
A	5	GLY	-	expression tag	UNP P16885
A	6	GLY	-	expression tag	UNP P16885
A	7	SER	-	expression tag	UNP P16885
A	94	ASP	TYR	variant	UNP P16885

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: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A: 70% • • 19% 6%

GLY GLY GLY SER GLY GLY SER K8 K9 D10 E11 H12 K13 D37 Q48 T49 M50 E51 E52 D53 N54 P55 L56 G57 S58 L59 C60 V72 Q92 G93 D94 F108 Q112 K120 I121 D122 T123 K124

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: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A: 67% 6% • 19% 6%

GLY GLY GLY SER GLY GLY SER K8 K9 D10 E11 H12 K13 I24 D37 Q48 T49 M50 E51 E52 D53 N54 P55 L56 G57 S58 L59 C60 N67 V72 K73 A74 V84 Q92 G93 D94 F108 Q112 K120 I121 D122 T123 K124

4.2.2 Score per residue for model 2

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A: 67% 7% • 19% 6%

GLY GLY GLY SER GLY GLY SER K8 K9 D10 E11 H12 K13 Q14 I24 I35 A36 D37 Q48 T49 M50 E51 E52 D53 N54 P55 L56 G57 S58 L59 C60 V72 P75 Q92 G93 D94 F108 Q112 K120 I121 D122 T123 K124

4.2.3 Score per residue for model 3

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A: 69% 5% • 19% 6%

GLY GLY GLY SER GLY GLY SER K8 K9 D10 E11 H12 K13 Q48 T49 M50 E51 E52 D53 N54 P55 L56 G57 S58 L59 C60 N67 V72 P75 K78 Q92 G93 D94 F108 Q112 K120 I121 D122 T123 K124

4.2.4 Score per residue for model 4

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A: 65% 7% • 19% 6%

GLY GLY GLY SER GLY GLY SER K8 K9 D10 E11 H12 K13 Q14 Q26 K27 D37 Q48 T49 M50 E51 E52 D53 N54 P55 L56 G57 S58 L59 C60 V72 P75 N79 Q92 G93 D94 P95 F108 Q112 K120 I121 D122 T123 K124

4.2.5 Score per residue for model 5

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A: 62% 11% • 19% 6%

GLY GLY GLY SER GLY GLY SER K8 K9 D10 E11 H12 K13 L18 D37 A38 K39 E47 Q48 T49 M50 E51 E52 D53 N54 P55 L56 G57 S58 L59 C60 D65 V71 V72 F85 Q92 G93 D94 F108 F111 Q112 S113 E116 I117 K120 I121 D122 T123 K124

4.2.6 Score per residue for model 6

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A: 65% 7% • 19% 6%

GLY GLY GLY SER GLY GLY SER K8 K9 D10 E11 H12 K13 Q26 D37 E47 Q48 T49 M50 E51 E52 D53 N54 P55 L56 G57 S58 L59 C60 V71 V72 F85 Q92 G93 D94 E98 F108 F111 Q112 K120 I121 D122 T123 K124

4.2.7 Score per residue for model 7

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

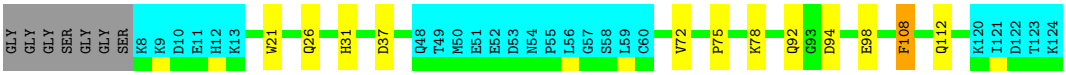
Chain A: 66% 8% • 19% 6%

GLY GLY GLY SER GLY GLY SER K8 K9 D10 E11 H12 K13 L18 Q26 D37 Q48 T49 M50 E51 E52 D53 N54 P55 L56 G57 S58 L59 C60 D65 V72 F85 Q92 G93 D94 F108 F111 Q112 K120 I121 D122 T123 K124

4.2.8 Score per residue for model 8

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

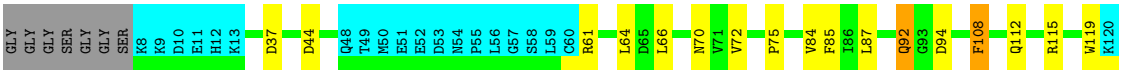
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

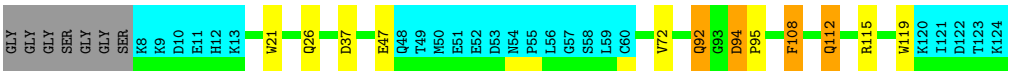
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

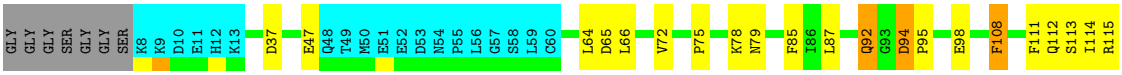
Chain A:

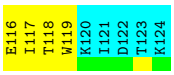


4.2.11 Score per residue for model 11

• Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A:

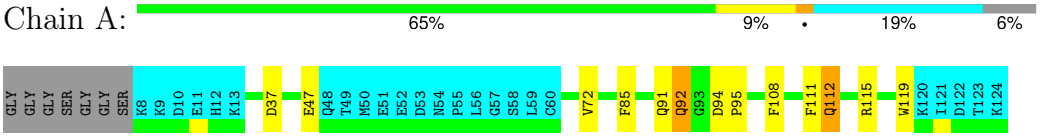




4.2.12 Score per residue for model 12

- Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

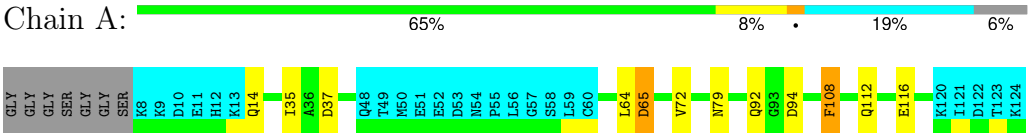
Chain A:



4.2.13 Score per residue for model 13

- Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

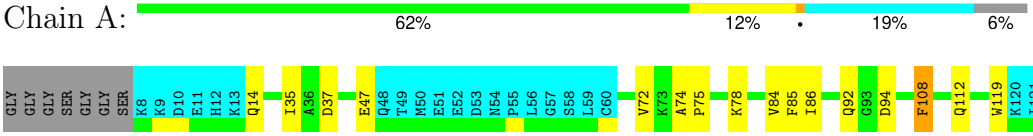
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A:



4.2.15 Score per residue for model 15

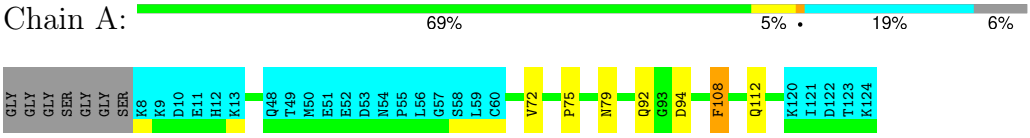
- Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A:

4.2.16 Score per residue for model 16 (medoid)

- Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

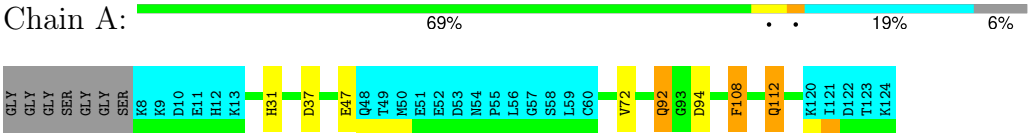
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

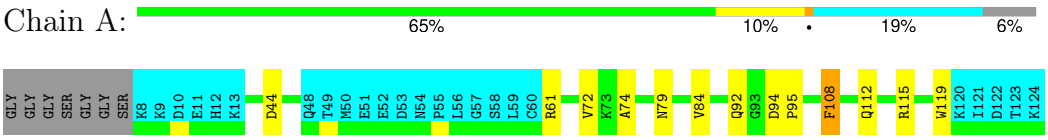
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

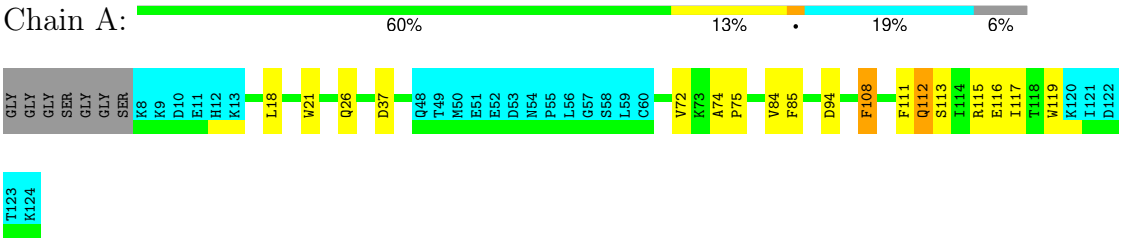
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

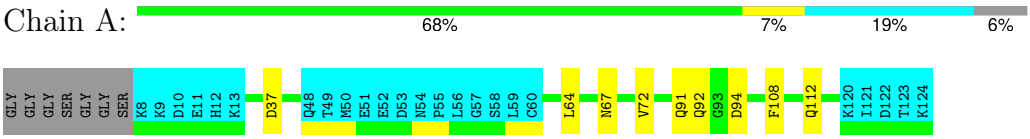
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: 1-phosphatidylinositol-4,5-bisphosphate phosphodiesterase gamma-2

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing.*

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *20 structures for lowest energy.*

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

Software name	Classification	Version
CNS	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	1474
Number of shifts mapped to atoms	1474
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	92%

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	778	741	740	5±2
All	All	15560	14820	14800	91

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 3.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:108:PHE:O	1:A:112:GLN:HB2	0.59	1.98	7	17
1:A:21:TRP:CZ2	1:A:26:GLN:HG3	0.51	2.41	19	4
1:A:64:LEU:HD21	1:A:87:LEU:HG	0.49	1.83	11	1
1:A:115:ARG:O	1:A:119:TRP:HB2	0.47	2.09	12	2
1:A:108:PHE:O	1:A:112:GLN:CB	0.47	2.62	13	15
1:A:108:PHE:O	1:A:112:GLN:HB3	0.45	2.10	13	1
1:A:94:ASP:HB3	1:A:95:PRO:HD2	0.45	1.88	11	5
1:A:78:LYS:HG2	1:A:98:GLU:HG2	0.45	1.88	11	2
1:A:66:LEU:HB2	1:A:117:ILE:HG21	0.45	1.89	11	1
1:A:14:GLN:HB2	1:A:35:ILE:HD12	0.44	1.88	14	3
1:A:115:ARG:O	1:A:119:TRP:HB3	0.44	2.12	9	2
1:A:24:ILE:HD12	1:A:24:ILE:N	0.44	2.28	2	2
1:A:117:ILE:HD12	1:A:117:ILE:N	0.44	2.27	19	1
1:A:66:LEU:HD22	1:A:114:ILE:HG23	0.43	1.89	11	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:26:GLN:OE1	1:A:26:GLN:N	0.43	2.51	7	3
1:A:85:PHE:HB2	1:A:111:PHE:CE1	0.43	2.49	5	6
1:A:44:ASP:HB3	1:A:61:ARG:HE	0.43	1.73	18	1
1:A:39:LYS:HG2	1:A:65:ASP:HA	0.43	1.89	5	1
1:A:115:ARG:HB3	1:A:119:TRP:HB3	0.43	1.90	11	2
1:A:78:LYS:HD2	1:A:86:ILE:HD11	0.43	1.90	14	1
1:A:78:LYS:HG3	1:A:79:ASN:ND2	0.42	2.29	11	1
1:A:64:LEU:O	1:A:66:LEU:HD22	0.42	2.15	9	1
1:A:113:SER:O	1:A:116:GLU:HB2	0.41	2.14	5	3
1:A:44:ASP:HB3	1:A:61:ARG:NE	0.41	2.30	18	2
1:A:74:ALA:N	1:A:84:VAL:O	0.41	2.53	19	4
1:A:64:LEU:HD22	1:A:64:LEU:N	0.41	2.30	20	1
1:A:24:ILE:HD12	1:A:24:ILE:H	0.41	1.74	1	1
1:A:112:GLN:O	1:A:116:GLU:HG3	0.41	2.15	13	1
1:A:115:ARG:O	1:A:119:TRP:N	0.41	2.54	10	1
1:A:117:ILE:N	1:A:117:ILE:HD12	0.40	2.31	5	1
1:A:84:VAL:CG1	1:A:85:PHE:N	0.40	2.85	9	2
1:A:70:ASN:O	1:A:87:LEU:HA	0.40	2.17	9	1
1:A:91:GLN:HG2	1:A:92:GLN:HG2	0.40	1.94	12	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	93/124 (75%)	80±2 (86±2%)	10±2 (11±2%)	3±1 (3±1%)	7	38
All	All	1860/2480 (75%)	1593 (86%)	207 (11%)	60 (3%)	7	38

All 8 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	92	GLN	17
1	A	37	ASP	16
1	A	75	PRO	10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	47	GLU	7
1	A	79	ASN	5
1	A	65	ASP	3
1	A	119	TRP	1
1	A	91	GLN	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	84/109 (77%)	79±2 (94±2%)	5±2 (6±2%)	25 74
All	All	1680/2180 (77%)	1586 (94%)	94 (6%)	25 74

All 17 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	72	VAL	20
1	A	94	ASP	18
1	A	108	PHE	16
1	A	92	GLN	11
1	A	112	GLN	10
1	A	67	ASN	4
1	A	18	LEU	3
1	A	71	VAL	2
1	A	31	HIS	2
1	A	78	LYS	1
1	A	14	GLN	1
1	A	27	LYS	1
1	A	98	GLU	1
1	A	118	THR	1
1	A	64	LEU	1
1	A	65	ASP	1
1	A	80	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)

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 92% for the well-defined parts and 90% 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 (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	1474
Number of shifts mapped to atoms	1474
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	12

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	114	-0.06 ± 0.12	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	109	0.01 ± 0.23	None needed (< 0.5 ppm)
$^{13}C'$	107	-0.01 ± 0.10	None needed (< 0.5 ppm)
^{15}N	111	0.09 ± 0.28	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 92%, i.e. 1209 atoms were assigned a chemical shift out of a possible 1312. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	455/461 (99%)	186/186 (100%)	180/186 (97%)	89/89 (100%)
Sidechain	628/708 (89%)	427/454 (94%)	194/225 (86%)	7/29 (24%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	126/143 (88%)	63/70 (90%)	59/67 (88%)	4/6 (67%)
Overall	1209/1312 (92%)	676/710 (95%)	433/478 (91%)	100/124 (81%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 90%, i.e. 1474 atoms were assigned a chemical shift out of a possible 1633. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

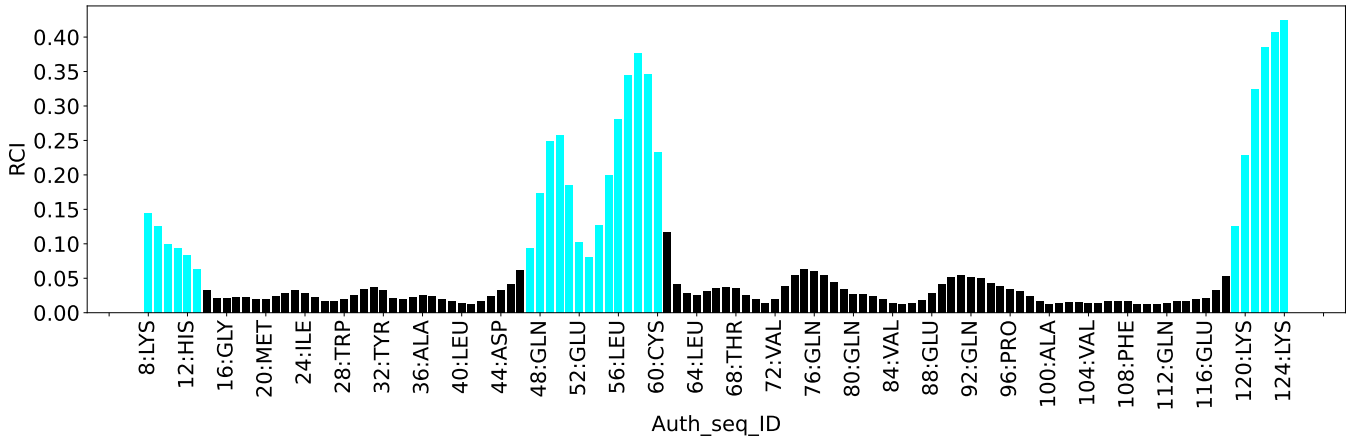
	Total	^{1}H	^{13}C	^{15}N
Backbone	563/580 (97%)	231/234 (99%)	221/234 (94%)	111/112 (99%)
Sidechain	783/902 (87%)	533/576 (93%)	241/290 (83%)	9/36 (25%)
Aromatic	128/151 (85%)	64/74 (86%)	60/69 (87%)	4/8 (50%)
Overall	1474/1633 (90%)	828/884 (94%)	522/593 (88%)	124/156 (79%)

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	81	LYS	HE2	0.92	1.95 – 3.88	-10.3
1	A	81	LYS	HG2	-0.52	0.13 – 2.61	-7.6
1	A	14	GLN	HG2	0.35	1.01 – 3.62	-7.5
1	A	81	LYS	HD2	0.19	0.58 – 2.64	-6.9
1	A	81	LYS	HE3	1.61	1.92 – 3.89	-6.6
1	A	14	GLN	HE21	4.32	5.02 – 9.43	-6.6
1	A	99	PHE	HB2	0.85	1.20 – 4.80	-5.9
1	A	82	SER	HB2	2.39	2.61 – 5.13	-5.9
1	A	21	TRP	HB2	1.40	1.51 – 4.87	-5.3
1	A	18	LEU	HD11	-0.66	-0.61 – 2.12	-5.2
1	A	18	LEU	HD12	-0.66	-0.61 – 2.12	-5.2
1	A	18	LEU	HD13	-0.66	-0.61 – 2.12	-5.2

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: