W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Feb 20, 2022 – 12:39 AM EST

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 following versions of software and data (see references (i)) were used in the production of this report:
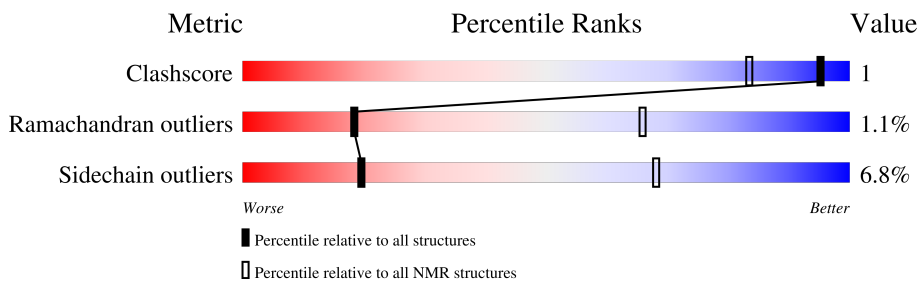
1 Overall quality at a glance ⓘ

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

The overall completeness of chemical shifts assignment was not calculated.

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		1
Ramachandran outliers		1.1%
Sidechain outliers		6.8%

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	61	82% • 16%

2   Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 14 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:441-A:491 (51)	0.30	14

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 883 atoms, of which 422 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called F-spondin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	61	883	277	422	80	95	9	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	439	GLY	-	cloning artifact	UNP P35446
A	440	SER	-	cloning artifact	UNP P35446

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.
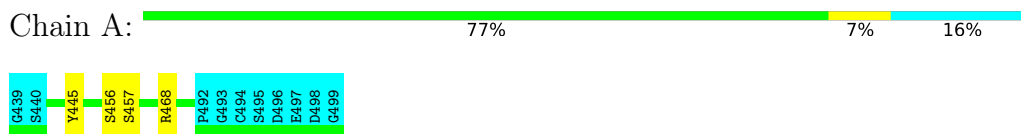
- Molecule 1: F-spondin

Chain A:

82% · 16%

G439 S440 Y445 P492 G493 C494 S495 D496 E497 D498 G499
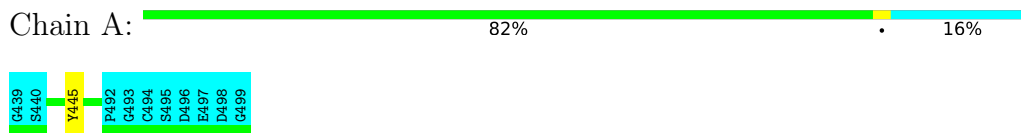
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: F-spondin

Chain A:

77% 7% 16%

G439 S440 Y445 S456 S457 R468 P492 G493 C494 S495 D496 E497 D498 G499

4.2.2 Score per residue for model 2

- Molecule 1: F-spondin

Chain A:

72% 11% 16%

G439 S440 Y445 S446 N447 W448 S452 C459 R468 D485 P492 G493 C494 S495 D496 E497 D498 G499
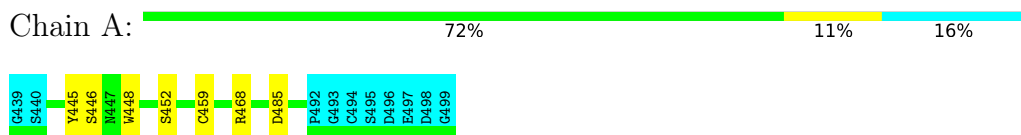
4.2.3 Score per residue for model 3

• Molecule 1: F-spondin

Chain A:

4.2.4 Score per residue for model 4

• Molecule 1: F-spondin

Chain A:

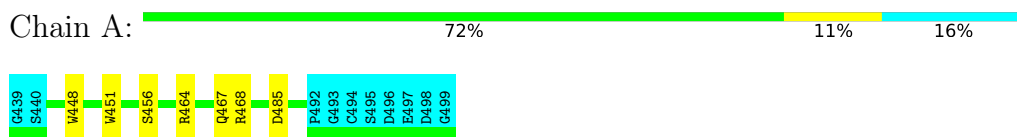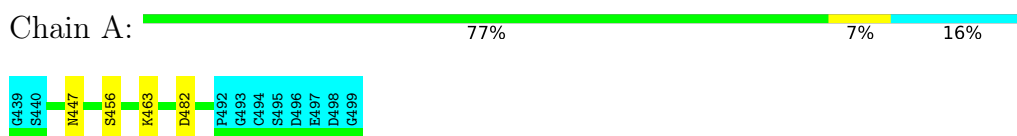### 4.2.5 Score per residue for model 5

• Molecule 1: F-spondin

Chain A:
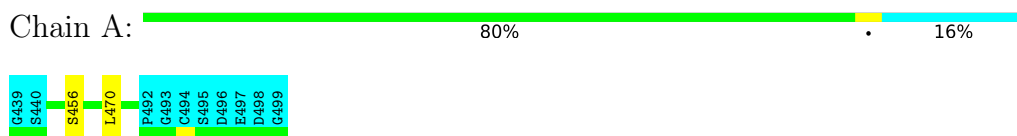
4.2.6 Score per residue for model 6

• Molecule 1: F-spondin

Chain A:
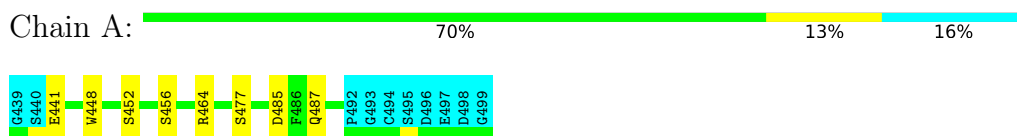
4.2.7 Score per residue for model 7

• Molecule 1: F-spondin

Chain A:
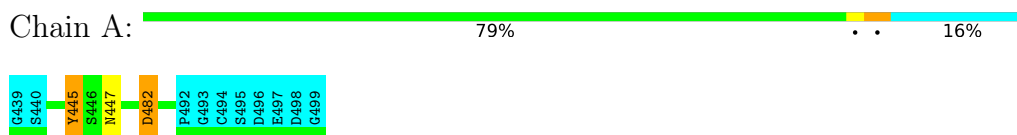
4.2.8 Score per residue for model 8

• Molecule 1: F-spondin

Chain A: 79%   5%   16%

G439 S440 S455 R464 D482 P492 G493 C494 S495 D496 E497 D498 G499
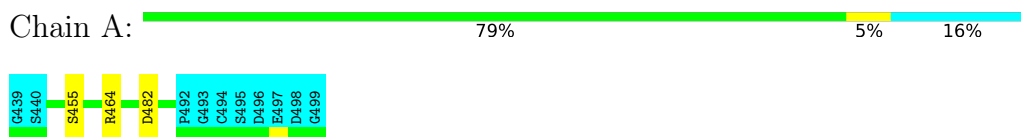
4.2.9 Score per residue for model 9

• Molecule 1: F-spondin

Chain A: 82%   •   16%

G439 S440 Q487 P492 G493 C494 S495 D496 E497 D498 G499
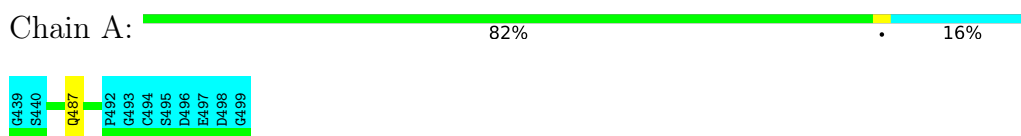
4.2.10 Score per residue for model 10

• Molecule 1: F-spondin

Chain A: 74%   10%   16%

G439 S440 Y445 W448 S455 Q467 D485 F486 Q487 P492 G493 C494 S495 D496 E497 D498 G499
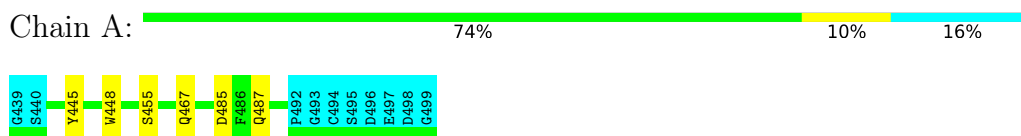
4.2.11 Score per residue for model 11

• Molecule 1: F-spondin

Chain A: 72%   10%   •   16%

G439 S440 C443 I444 Y445 S446 N447 E460 S477 Q487 G491 P492 G493 C494 S495 D496 E497 D498 G499
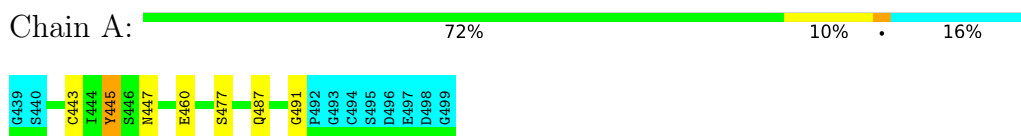
4.2.12 Score per residue for model 12

• Molecule 1: F-spondin

Chain A: 75%   8%   16%

G439 S440 E441 T442 N447 S457 R464 D482 P492 G493 C494 S495 D496 E497 D498 G499
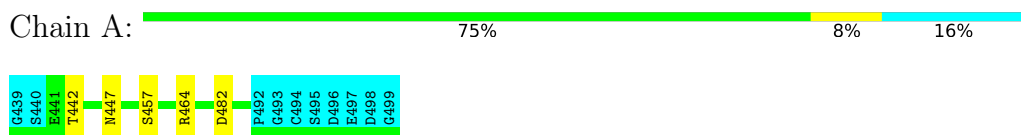
4.2.13 Score per residue for model 13

• Molecule 1: F-spondin

Chain A:

74%   10%   16%

G439 S440 E441 T442 W448 Q467 R468 S477 D485 P492 G493 C494 S495 D496 E497 D498 G499
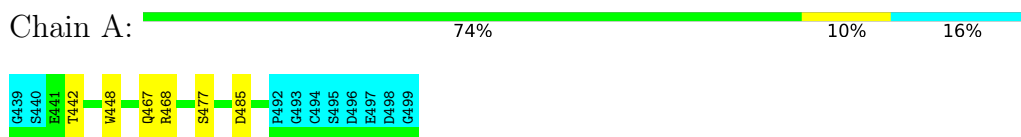
4.2.14 Score per residue for model 14 (medoid)

• Molecule 1: F-spondin

Chain A:

70%   10%   •   16%

G439 S440 E441 Y445 W448 S457 R468 S477 T483 Q484 D485 P492 G493 C494 S495 D496 E497 D498 G499
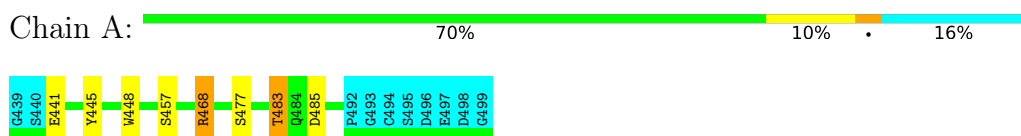
4.2.15 Score per residue for model 15

• Molecule 1: F-spondin

Chain A:

72%   7%   5%   16%

G439 S440 Y445 S446 N447 S456 Q467 R468 D482 T483 P492 G493 C494 S495 D496 E497 D498 G499
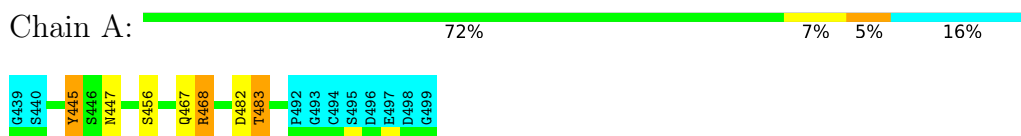
4.2.16 Score per residue for model 16

• Molecule 1: F-spondin

Chain A:

75%   8%   16%

G439 S440 Y445 C459 E460 K461 R468 M490 G491 P492 G493 C494 S495 D496 E497 D498 G499
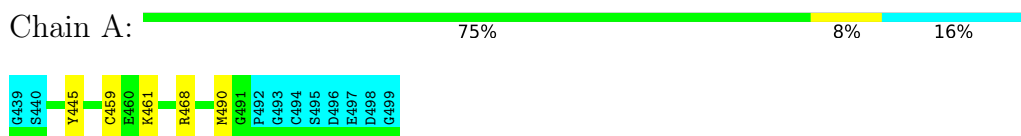
4.2.17 Score per residue for model 17

• Molecule 1: F-spondin

Chain A:

77%   7%   16%

G439 S440 S456 R468 L474 Q487 P492 G493 C494 S495 D496 E497 D498 G499
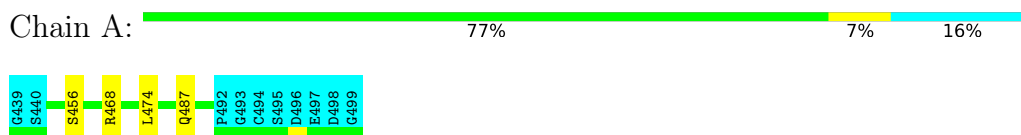
4.2.18 Score per residue for model 18

• Molecule 1: F-spondin

Chain A:

72%   11%   16%

G439 S440 Y445 W451 S452 K463 R464 D482 Q487 P492 G493 C494 S495 D496 E497 D498 G499
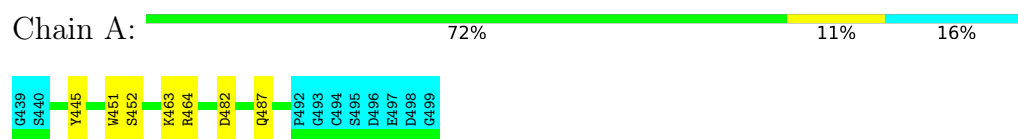
4.2.19 Score per residue for model 19

• Molecule 1: F-spondin

Chain A:

80%   •   16%

G439 S440 S456 D475 P492 G493 C494 S495 D496 E497 D498 G499
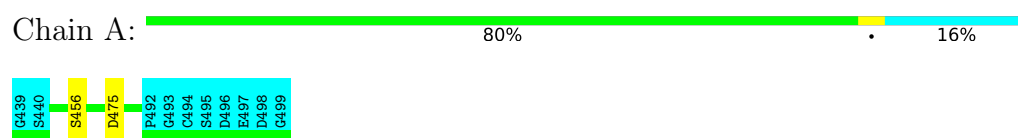
4.2.20 Score per residue for model 20

• Molecule 1: F-spondin

Chain A:

72%   11%   16%

G439 S440 N447 W451 S455 S456 S457 R466 K471 Q487 P492 G493 C494 S495 D496 E497 D498 G499
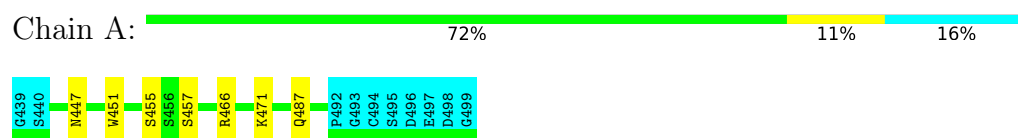
5 Refinement protocol and experimental data overview ⓘ

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

Of the 200 calculated structures, 20 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 solution	2.0.30
OPALp	refinement	1.3

No chemical shift data was provided.

6 Model quality ⓘ

6.1 Standard geometry ⓘ

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	0.57±0.01	0±0/409 (0.0± 0.0%)	1.09±0.03	0±0/554 (0.1± 0.1%)
All	All	0.57	0/8180 (0.0%)	1.09	7/11080 (0.1%)

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	0.1±0.4
All	All	0	3

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	468	ARG	CD-NE-CZ	5.90	131.86	123.60	13	4
1	A	466	ARG	NE-CZ-NH2	-5.43	117.58	120.30	20	1
1	A	468	ARG	NE-CZ-NH2	-5.24	117.68	120.30	1	1
1	A	464	ARG	NE-CZ-NH2	-5.23	117.68	120.30	8	1

There are no chirality outliers.

All 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	445	TYR	Sidechain	2
1	A	468	ARG	Sidechain	1

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	399	378	378	1±1
All	All	7980	7560	7560	15

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.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:448:TRP:CH2	1:A:468:ARG:NH1	0.60	2.70	14	1
1:A:451:TRP:CE3	1:A:464:ARG:HB2	0.49	2.42	18	2
1:A:448:TRP:CH2	1:A:485:ASP:HB2	0.46	2.46	13	6
1:A:461:LYS:HE3	1:A:490:MET:SD	0.43	2.53	16	1
1:A:451:TRP:CH2	1:A:487:GLN:HB2	0.42	2.49	20	1
1:A:468:ARG:HH11	1:A:468:ARG:HG2	0.42	1.73	14	1
1:A:443:CYS:HB2	1:A:445:TYR:CZ	0.42	2.49	11	1
1:A:448:TRP:CZ2	1:A:468:ARG:CZ	0.41	3.04	3	1
1:A:445:TYR:CD2	1:A:468:ARG:NH1	0.41	2.88	15	1

6.3   Torsion angles ⓘ

6.3.1   Protein backbone ⓘ

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	51/61 (84%)	48±1 (94±3%)	3±1 (5±3%)	1±1 (1±1%)	18	66
All	All	1020/1220 (84%)	954 (94%)	55 (5%)	11 (1%)	18	66

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	482	ASP	4
1	A	452	SER	2
1	A	459	CYS	2
1	A	483	THR	2
1	A	491	GLY	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	47/54 (87%)	44±1 (93±3%)	3±1 (7±3%)	19	68
All	All	940/1080 (87%)	876 (93%)	64 (7%)	19	68

All 22 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	456	SER	8
1	A	445	TYR	8
1	A	447	ASN	6
1	A	487	GLN	6
1	A	457	SER	4
1	A	467	GLN	4
1	A	477	SER	4
1	A	482	ASP	3
1	A	455	SER	3
1	A	463	LYS	2
1	A	441	GLU	2
1	A	464	ARG	2
1	A	442	THR	2
1	A	483	THR	2
1	A	446	SER	1
1	A	470	LEU	1
1	A	452	SER	1
1	A	460	GLU	1
1	A	468	ARG	1
1	A	474	LEU	1
1	A	475	ASP	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	471	LYS	1

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 ⓘ

There are no monosaccharides in this entry.

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 (i)

No chemical shift data were provided