Full wwPDB NMR Structure Validation Report (i)

Mar 6, 2022 – 11:09 PM EST

PDB ID	:	2KP8
Title	:	Ligand bound to a model peptide that mimics the open fusogenic form
Authors	:	Olejniczak, E.T.
Deposited on	:	2009-10-09

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:

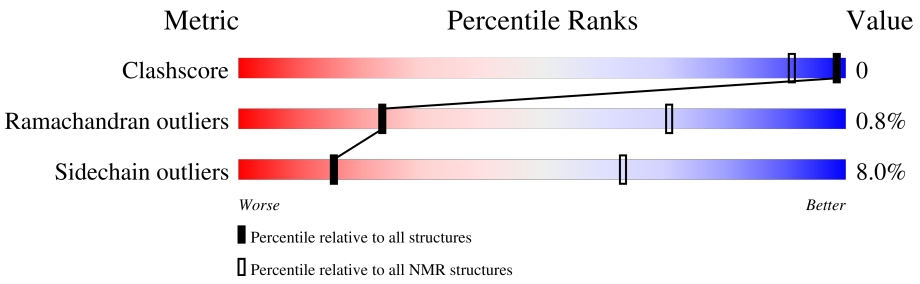
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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 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	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	72	92%
1	B	72	96%
1	C	72	96%

2 Ensemble composition and analysis (i)

This entry contains 5 models. Model 2 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:72, B:101-B:172, C:201-C:272 (213)	0.60	2

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 and 1 single-model cluster was found.

Cluster number	Models
1	2, 3
2	1, 4
Single-model clusters	5

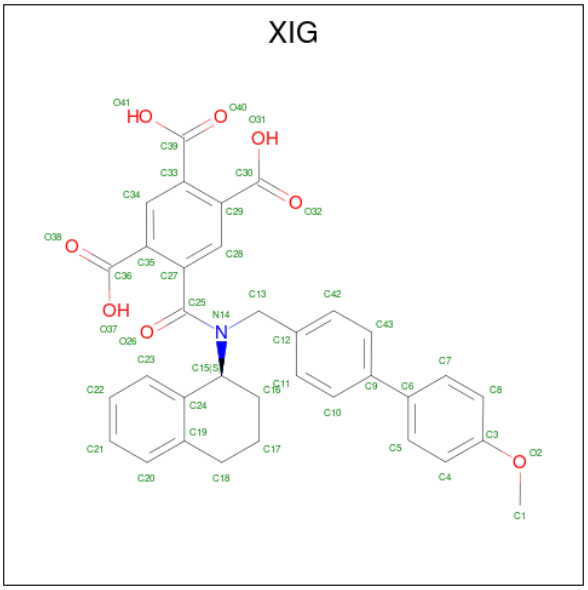
3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3525 atoms, of which 1763 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Model peptide.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	72	1152	356	579	102	115	0
1	B	72	1152	356	579	102	115	0
1	C	72	1152	356	579	102	115	0

- Molecule 2 is 5-{[(4'-methoxybiphenyl-4-yl)methyl][(1S)-1,2,3,4-tetrahydronaphthalen-1-yl] carbamoyl}benzene-1,2,4-tricarboxylic acid (three-letter code: XIG) (formula: $C_{34}H_{29}NO_8$).



Mol	Chain	Residues	Atoms				
			Total	C	H	N	O
2	B	1	69	34	26	1	8

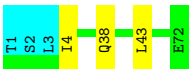
4 Residue-property plots (i)

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.

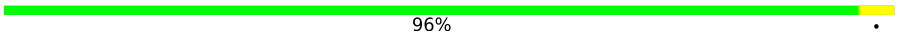
• Molecule 1: Model peptide

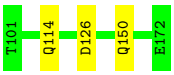
Chain A:



• Molecule 1: Model peptide

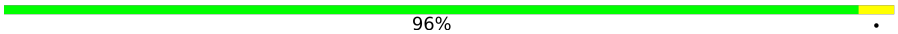
Chain B:





• Molecule 1: Model peptide

Chain C:



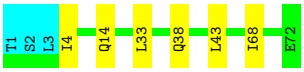
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

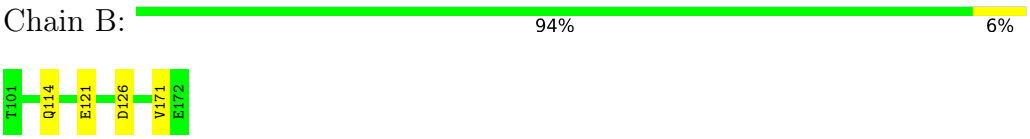
• Molecule 1: Model peptide

Chain A:



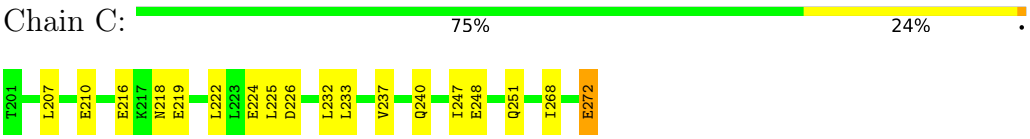
• Molecule 1: Model peptide

Chain B:



• Molecule 1: Model peptide

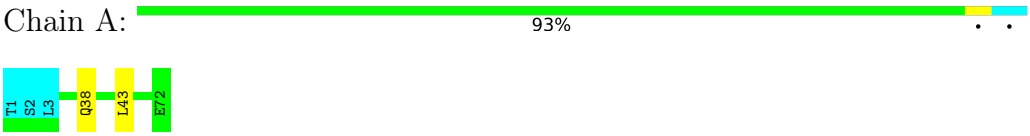
Chain C:



4.2.2 Score per residue for model 2 (medoid)

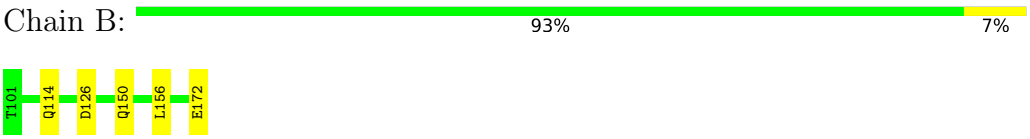
• Molecule 1: Model peptide

Chain A:



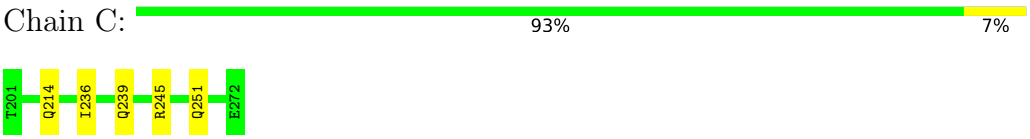
• Molecule 1: Model peptide

Chain B:



• Molecule 1: Model peptide

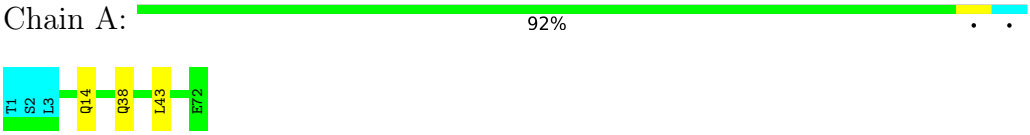
Chain C:



4.2.3 Score per residue for model 3

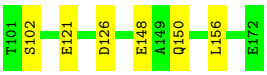
• Molecule 1: Model peptide

Chain A:



• Molecule 1: Model peptide

Chain B: 92% 8%



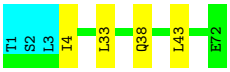
• Molecule 1: Model peptide

Chain C: 92% 8%

4.2.4 Score per residue for model 4

• Molecule 1: Model peptide

Chain A: 90% 6%



• Molecule 1: Model peptide

Chain B: 96%

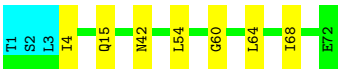
• Molecule 1: Model peptide

Chain C: 90% 8%

4.2.5 Score per residue for model 5

• Molecule 1: Model peptide

Chain A: 86% 10%



• Molecule 1: Model peptide

Chain B: 97%

• Molecule 1: Model peptide

Chain C:

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing, Final minimization step.*

Of the 50 calculated structures, 5 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
X-PLOR	geometry optimization	
X-PLOR	refinement	
Schrodinger	geometry optimization	

No chemical shift data was provided.

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: XIG

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	0.62±0.01	0±0/556 (0.0± 0.0%)	0.89±0.01	0±0/751 (0.0± 0.0%)
1	B	0.62±0.01	0±0/577 (0.0± 0.0%)	0.93±0.02	0±0/780 (0.0± 0.0%)
1	C	0.67±0.07	0±0/577 (0.1± 0.1%)	0.95±0.04	0±0/780 (0.0± 0.1%)
All	All	0.64	2/8550 (0.0%)	0.92	1/11555 (0.0%)

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	C	272	GLU	CD-OE1	9.71	1.36	1.25	4	2

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	C	245	ARG	NE-CZ-NH2	-7.36	116.62	120.30	2	1

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	C	573	579	579	0±0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes
All	All	8705	8705	8705	1

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:C:222:LEU:HD22	1:C:237:VAL:HG11	0.48	1.84	1	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	68/72 (94%)	65±1 (95±1%)	3±1 (4±2%)	0±0 (0±1%)	44	80
1	B	70/72 (97%)	66±1 (95±2%)	3±1 (4±2%)	1±1 (1±1%)	15	61
1	C	70/72 (97%)	66±2 (95±2%)	3±2 (5±2%)	0±0 (1±1%)	29	74
All	All	1040/1080 (96%)	987 (95%)	45 (4%)	8 (1%)	24	71

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	B	126	ASP	4
1	C	226	ASP	1
1	B	102	SER	1
1	A	60	GLY	1
1	C	260	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	61/64 (95%)	57±2 (93±3%)	4±2 (7±3%)	19 68
1	B	64/64 (100%)	61±1 (95±1%)	3±1 (5±1%)	30 79
1	C	64/64 (100%)	56±4 (88±6%)	8±4 (12±6%)	8 50
All	All	945/960 (98%)	869 (92%)	76 (8%)	16 63

All 40 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	38	GLN	4
1	A	43	LEU	4
1	C	232	LEU	4
1	A	4	ILE	3
1	B	114	GLN	3
1	C	251	GLN	3
1	C	272	GLU	3
1	B	150	GLN	3
1	A	14	GLN	2
1	A	33	LEU	2
1	A	68	ILE	2
1	B	121	GLU	2
1	C	207	LEU	2
1	C	210	GLU	2
1	C	218	ASN	2
1	C	219	GLU	2
1	C	233	LEU	2
1	C	240	GLN	2
1	C	247	ILE	2
1	C	268	ILE	2
1	B	156	LEU	2
1	B	172	GLU	2
1	C	214	GLN	2
1	C	236	ILE	2
1	C	239	GLN	2
1	B	171	VAL	1
1	C	216	GLU	1
1	C	224	GLU	1
1	C	225	LEU	1
1	C	248	GLU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	B	148	GLU	1
1	A	15	GLN	1
1	A	42	ASN	1
1	A	54	LEU	1
1	A	64	LEU	1
1	B	117	LYS	1
1	C	220	GLN	1
1	C	222	LEU	1
1	C	250	GLN	1
1	C	253	LEU	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)

1 ligand 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	XIG	B	999	-	41,47,47	1.43±0.08	3±0 (7±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	XIG	B	999	-	50,67,67	0.72±0.15	1±1 (1±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	XIG	B	999	-	-	0±0,22,44,44	0±0,5,5,5

All 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	999	XIG	C35-C36	4.33	1.51	1.47	3	5
2	B	999	XIG	C33-C39	4.10	1.51	1.47	3	5
2	B	999	XIG	C29-C30	3.95	1.51	1.47	5	5

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	999	XIG	C12-C13-N14	3.54	120.02	113.54	5	2
2	B	999	XIG	C13-N14-C25	2.96	125.19	116.97	3	1

There are no chirality outliers.

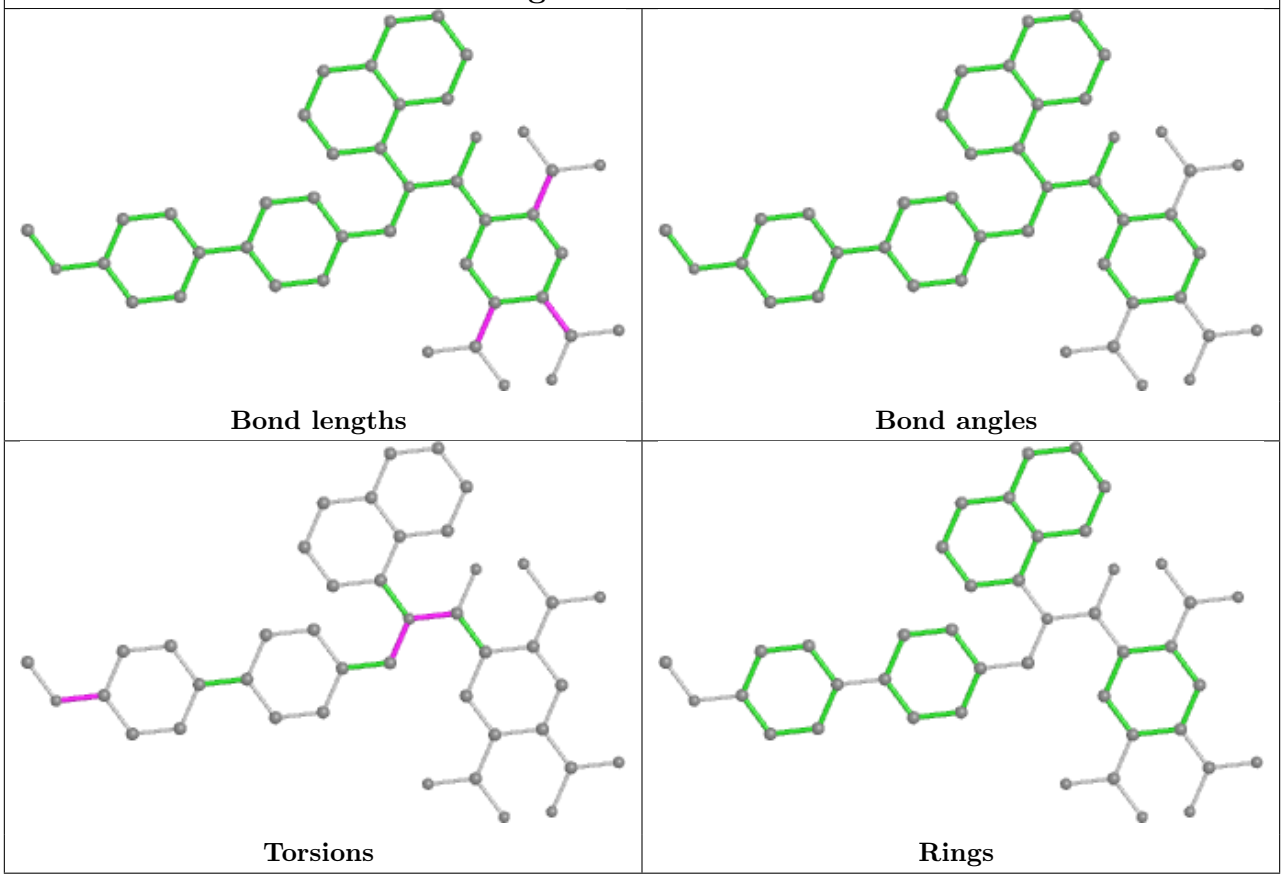
There are no torsion outliers.

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 all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is

within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand XIG B 999



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)

No chemical shift data were provided