WORLDWIDE wwPDB PROTEIN DATA BANK

Full wwPDB NMR Structure Validation Report (i)

Feb 10, 2022 – 08:58 AM EST

PDB ID	:	1EIO
Title	:	ILEAL LIPID BINDING PROTEIN IN COMPLEX WITH GLYCO-CHOLATE
Authors	:	Luecke, C.; Zhang, F.; Hamilton, J.A.; Sacchettini, J.C.; Rueterjans, H.
Deposited on	:	2000-02-27

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)
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	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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.
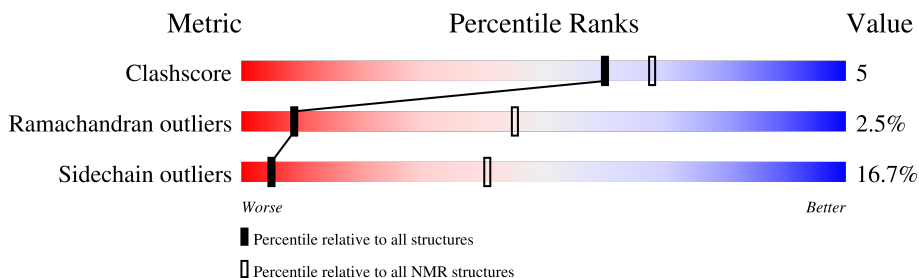
Metric	Whole archive (#Entries)	NMR archive (#Entries)
Clashscore	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428

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	127	65% 29% • 5%

2 Ensemble composition and analysis ⓘ

This entry contains 5 models. Model 5 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:2-A:22, A:28-A:127 (121)	0.67	5

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	1, 2
2	4, 5
Single-model clusters	3

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

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 2056 atoms, of which 1032 are hydrogens and 0 are deuteriums.
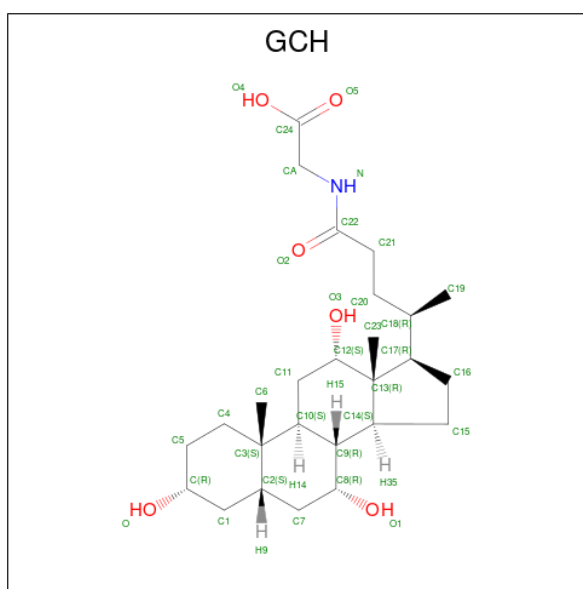
- Molecule 1 is a protein called ILEAL LIPID BINDING PROTEIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	127	1980	624	989	168	196	3	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	118	SER	THR	conflict	UNP P10289

- Molecule 2 is GLYCOCHOLIC ACID (three-letter code: GCH) (formula: $C_{26}H_{43}NO_6$).

| Mol	Chain	Residues	Atoms				
			Total	C	H	N	O
2	A	1	76	26	43	1	6

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.

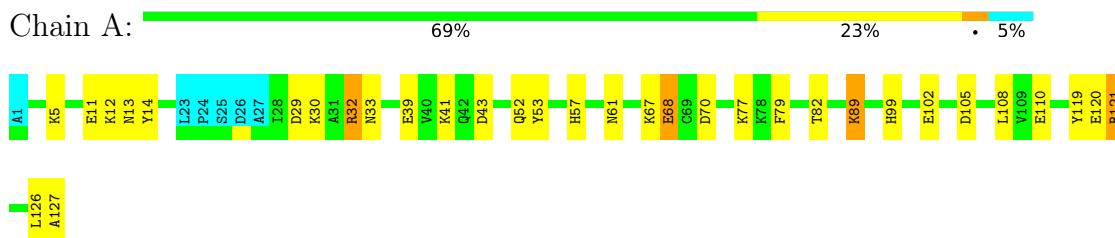• Molecule 1: ILEAL LIPID BINDING PROTEIN
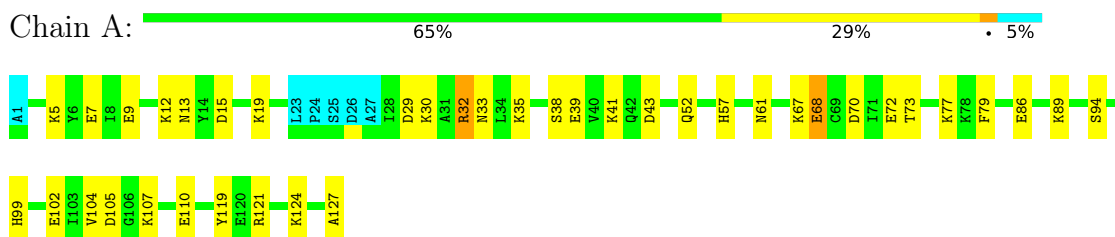
4.2 Scores per residue for each member of the ensemble

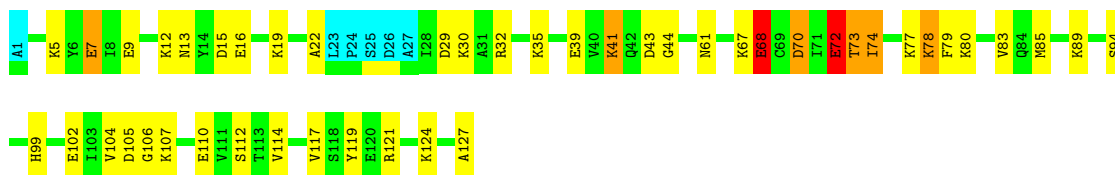Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: ILEAL LIPID BINDING PROTEIN

4.2.2 Score per residue for model 2

• Molecule 1: ILEAL LIPID BINDING PROTEIN

A1 K5 Y6 E7 I8 E9 K12 N13 Y14 D15 E16 K19 A22 L23 P24 S25 D26 A27 I28 D29 K30 A31 R32 K35 E39 V40 K41 Q42 D43 G44 N61 K67 E68 C69 D70 I71 E72 T73 I74 K77 K78 F79 K80 V83 Q84 M85 K89 S94

H99 E102 I103 V104 D105 G106 K107 E110 V111 S112 T113 V114 V117 S118 Y119 E120 R121 K124 A127

4.2.3 Score per residue for model 3

● Molecule 1: ILEAL LIPID BINDING PROTEIN

Chain A:

A1 K5 Y6 E7 I8 E9 D15 K19 L23 P24 S25 D26 A27 K30 A31 R32 N33 L34 K35 S38 E39 V40 K41 Q42 D43 Q51 Q52 Y53 P54 G55 G56 H57 N61 K67 E68 C69 D70 T73 K77 K78 F79 K80 E86 S94 P95 N96

H99 T100 V104 D105 G106 K107 E110 R121 V122 S123 K124 A127
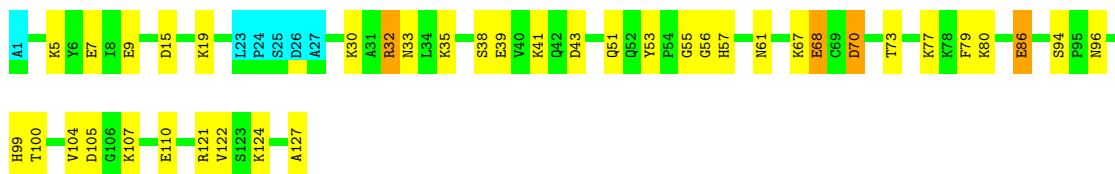
4.2.4 Score per residue for model 4

● Molecule 1: ILEAL LIPID BINDING PROTEIN

Chain A:

A1 K5 Y6 E7 I8 E9 N13 R20 L23 P24 S25 D26 A27 I28 D29 K30 N33 L34 K35 S38 E39 M46 Q51 Q52 H57 N61 K67 E68 I71 E72 T73 K77 K78 F79 M85 E86 G87 G88 K89 S94 H98 E102 I103

V104 D105 G106 Y119 E120 R121 V122 S123 K124 A127
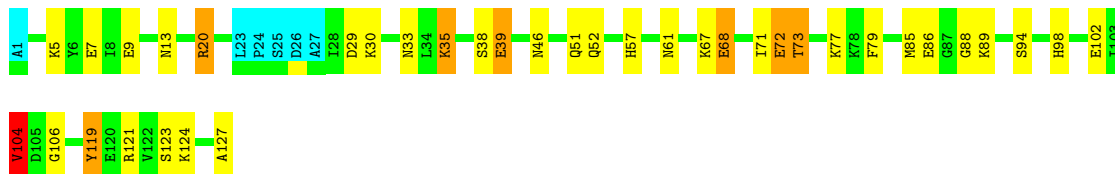
4.2.5 Score per residue for model 5 (medoid)

● Molecule 1: ILEAL LIPID BINDING PROTEIN

Chain A:

A1 K5 Y6 E7 K12 D15 K19 L23 P24 S25 D26 A27 I28 D29 K30 A31 R32 N33 L34 K35 S38 E39 V40 K41 Q42 D43 Q52 G55 G56 H57 N61 K67 E68 E72 F79 T82 E86 K89 N96 T100 A101 E102

D105 G106 K107 S112 T113 Y119 E120 R121 V122 S123 K124 A127
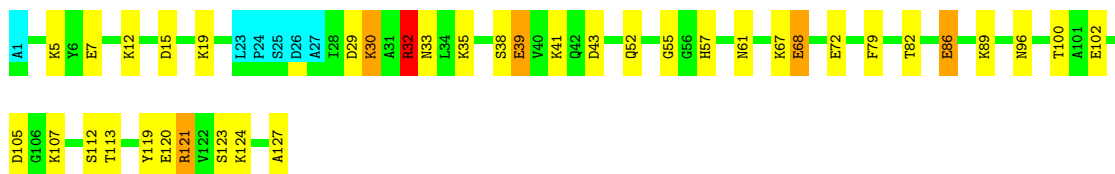
5 Refinement protocol and experimental data overview ⓘ

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

Of the 8 calculated structures, 5 were deposited, based on the following criterion: *structures with the lowest energies after docking of ligand.*

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

Software name	Classification	Version
DIANA	structure solution	2.8
SYBYL	refinement	6.4

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: GCH

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	0.96±0.01	1±0/968 (0.1± 0.0%)	1.18±0.04	4±2/1298 (0.3± 0.1%)
All	All	0.96	5/4840 (0.1%)	1.18	22/6490 (0.3%)

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	2.4±2.1
All	All	0	12

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	127	ALA	C-OXT	6.61	1.35	1.23	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
1	A	121	ARG	NE-CZ-NH1	11.07	125.83	120.30	5	5
1	A	121	ARG	NE-CZ-NH2	-8.39	116.11	120.30	5	3
1	A	32	ARG	NE-CZ-NH1	8.01	124.30	120.30	5	2
1	A	119	TYR	CB-CG-CD2	-7.35	116.59	121.00	4	4
1	A	119	TYR	CB-CG-CD1	7.14	125.29	121.00	4	3
1	A	121	ARG	CD-NE-CZ	6.58	132.81	123.60	5	1
1	A	20	ARG	NE-CZ-NH1	5.81	123.20	120.30	4	1
1	A	86	GLU	CB-CA-C	5.41	121.23	110.40	5	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models Worst	Total
1	A	102	GLU	OE1-CD-OE2	5.03	129.34	123.30	5	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	79	PHE	Peptide	2
1	A	52	GLN	Peptide	1
1	A	53	TYR	Sidechain	1
1	A	68	GLU	Peptide	1
1	A	72	GLU	Peptide	1
1	A	74	ILE	Peptide	1
1	A	83	VAL	Peptide	1
1	A	106	GLY	Peptide	1
1	A	71	ILE	Peptide	1
1	A	88	GLY	Peptide	1
1	A	32	ARG	Peptide	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	952	951	947	9±3
All	All	4925	4970	4945	46

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 5.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
1:A:41:LYS:NZ	1:A:43:ASP:OD1	0.57	2.35	5	4
1:A:86:GLU:OE2	1:A:98:HIS:NE2	0.57	2.35	4	1
1:A:67:LYS:NZ	1:A:68:GLU:OE2	0.56	2.38	3	5
1:A:105:ASP:OD1	1:A:124:LYS:NZ	0.55	2.40	5	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:99:HIS:NE2	1:A:110:GLU:OE2	0.54	2.37	3	3
1:A:5:LYS:NZ	1:A:39:GLU:OE2	0.52	2.40	3	4
1:A:7:GLU:OE2	1:A:35:LYS:NZ	0.51	2.38	3	3
1:A:89:LYS:NZ	1:A:102:GLU:OE1	0.51	2.43	2	1
1:A:89:LYS:NZ	1:A:102:GLU:OE2	0.50	2.41	1	2
1:A:29:ASP:OD1	1:A:32:ARG:NH1	0.50	2.44	5	1
1:A:105:ASP:OD2	1:A:124:LYS:NZ	0.48	2.41	2	2
1:A:29:ASP:OD2	1:A:32:ARG:NH1	0.47	2.48	2	1
1:A:7:GLU:OE1	1:A:35:LYS:NZ	0.46	2.49	2	1
1:A:15:ASP:OD1	1:A:19:LYS:NZ	0.46	2.44	2	2
1:A:9:GLU:OE1	1:A:107:LYS:NZ	0.46	2.46	2	2
1:A:68:GLU:OE1	1:A:80:LYS:NZ	0.45	2.45	2	1
1:A:104:VAL:O	1:A:106:GLY:N	0.45	2.50	4	1
1:A:74:ILE:O	1:A:74:ILE:HG22	0.43	2.14	2	1
1:A:72:GLU:OE1	1:A:73:THR:N	0.42	2.53	2	1
1:A:29:ASP:OD1	1:A:30:LYS:N	0.41	2.53	1	1
1:A:70:ASP:OD2	1:A:78:LYS:NZ	0.41	2.41	2	1
1:A:15:ASP:OD2	1:A:32:ARG:NE	0.41	2.50	3	2
1:A:70:ASP:OD1	1:A:80:LYS:NZ	0.41	2.43	3	1
1:A:11:GLU:OE2	1:A:121:ARG:NH1	0.40	2.49	1	1
1:A:72:GLU:OE2	1:A:77:LYS:NZ	0.40	2.49	4	1
1:A:5:LYS:NZ	1:A:39:GLU:OE1	0.40	2.53	2	1
1:A:12:LYS:NZ	1:A:120:GLU:OE2	0.40	2.41	5	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	120/127 (94%)	102±3 (85±2%)	15±4 (13±3%)	3±1 (2±1%)	9	45
All	All	600/635 (94%)	509 (85%)	76 (13%)	15 (2%)	9	45

All 9 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	73	THR	3
1	A	104	VAL	3
1	A	33	ASN	2
1	A	55	GLY	2
1	A	22	ALA	1
1	A	44	GLY	1
1	A	56	GLY	1
1	A	13	ASN	1
1	A	96	ASN	1

6.3.2   Protein sidechains ⓘ

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	104/108 (96%)	87±3 (83±3%)	17±3 (17±3%)	5	41
All	All	520/540 (96%)	433 (83%)	87 (17%)	5	41

All 51 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	61	ASN	5
1	A	68	GLU	5
1	A	70	ASP	3
1	A	77	LYS	3
1	A	72	GLU	3
1	A	94	SER	3
1	A	38	SER	3
1	A	79	PHE	3
1	A	12	LYS	2
1	A	13	ASN	2
1	A	32	ARG	2
1	A	82	THR	2
1	A	89	LYS	2
1	A	85	MET	2
1	A	112	SER	2
1	A	30	LYS	2
1	A	51	GLN	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	86	GLU	2
1	A	100	THR	2
1	A	33	ASN	2
1	A	39	GLU	2
1	A	52	GLN	2
1	A	57	HIS	2
1	A	123	SER	2
1	A	14	TYR	1
1	A	105	ASP	1
1	A	108	LEU	1
1	A	120	GLU	1
1	A	126	LEU	1
1	A	7	GLU	1
1	A	16	GLU	1
1	A	41	LYS	1
1	A	78	LYS	1
1	A	114	VAL	1
1	A	117	VAL	1
1	A	19	LYS	1
1	A	53	TYR	1
1	A	96	ASN	1
1	A	122	VAL	1
1	A	9	GLU	1
1	A	20	ARG	1
1	A	29	ASP	1
1	A	35	LYS	1
1	A	46	ASN	1
1	A	73	THR	1
1	A	104	VAL	1
1	A	119	TYR	1
1	A	124	LYS	1
1	A	107	LYS	1
1	A	113	THR	1
1	A	121	ARG	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 ⓘ

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	GCH	A	128	-	33,36,36	0.74±0.04	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	GCH	A	128	-	53,56,56	0.78±0.09	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	GCH	A	128	-	-	0±0,12,79,79	0±0,4,4,4

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
2	A	128	GCH	C20-C18-C17	2.23	114.89	110.28	5	1
2	A	128	GCH	CA-N-C22	2.10	119.06	122.34	4	1

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 ⓘ

No chemical shift data were provided