Full wwPDB NMR Structure Validation Report (i)

Jun 3, 2020 – 06:57 am BST

PDB ID	:	2N3Y
Title	:	NMR structure of the Y48pCMF variant of human cytochrome c in its reduced state
Authors	:	Moreno-Beltran, B.; Del Conte, R.; Diaz-Quintana, A.; De la Rosa, M.A.; Turano, P.; Diaz-Moreno, I.
Deposited on	:	2015-06-15

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 (1)) were used in the production of this report:

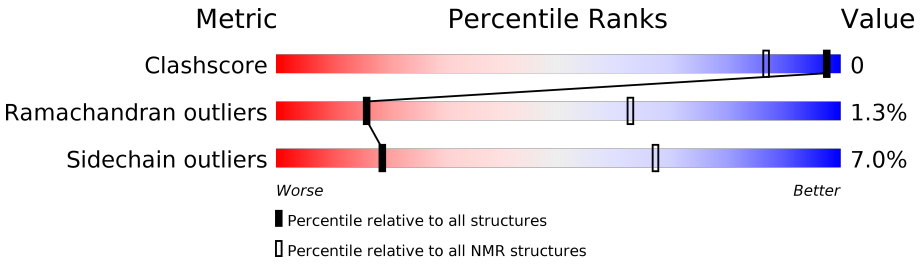
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
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

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

The overall completeness of chemical shifts assignment is 88%.

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	104	78% 6% 16%

2 Ensemble composition and analysis

This entry contains 20 models. Model 20 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:1-A:40, A:58-A:104 (87)	0.46	20

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 5 clusters and 5 single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1738 atoms, of which 876 are hydrogens and 0 are deuteriums.

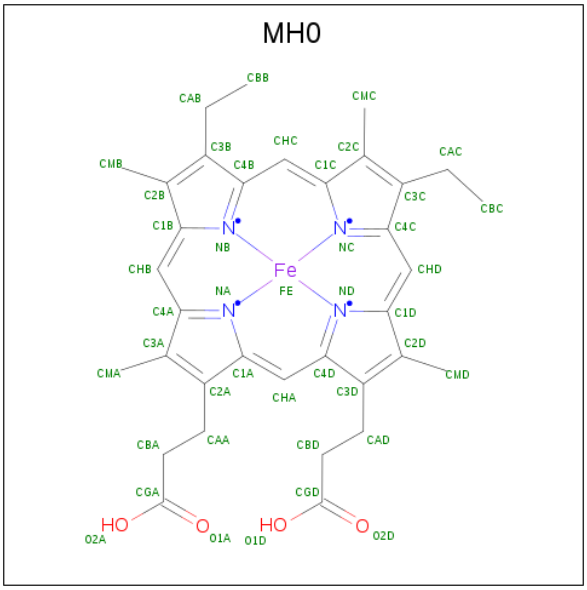
- Molecule 1 is a protein called Cytochrome c.

Mol	Chain	Residues	Atoms						Trace
1	A	104	Total	C	H	N	O	S	0
			1663	523	844	142	149	5	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	48	1PA	TYR	ENGINEERED MUTATION	UNP P99999

- Molecule 2 is Mesoheme (three-letter code: MH0) (formula: $C_{34}H_{36}FeN_4O_4$).



Mol	Chain	Residues	Atoms					
2	A	1	Total	C	Fe	H	N	O
			75	34	1	32	4	4

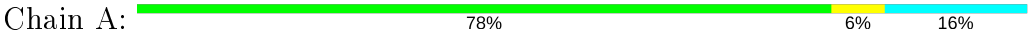
4 Residue-property plots (i)

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: Cytochrome c

Chain A:



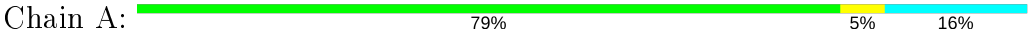
4.2 Scores per residue for each member of the ensemble

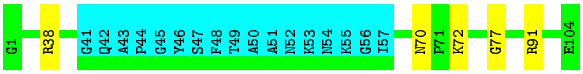
Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Cytochrome c

Chain A:

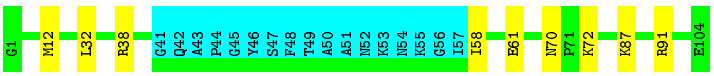




4.2.2 Score per residue for model 2

- Molecule 1: Cytochrome c

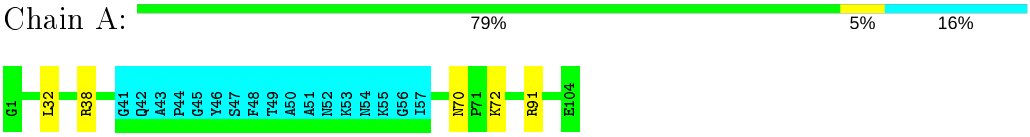
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Cytochrome c

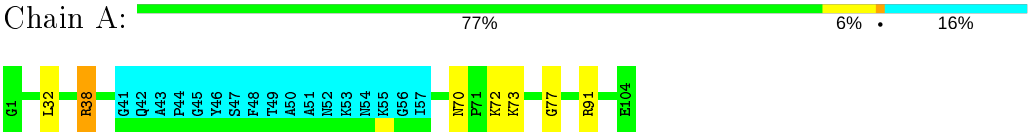
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Cytochrome c

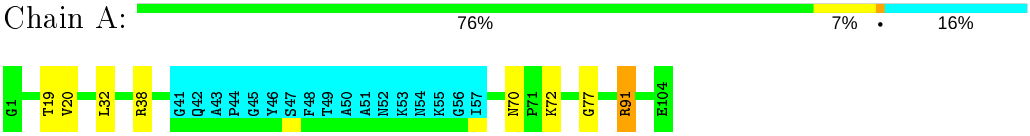
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Cytochrome c

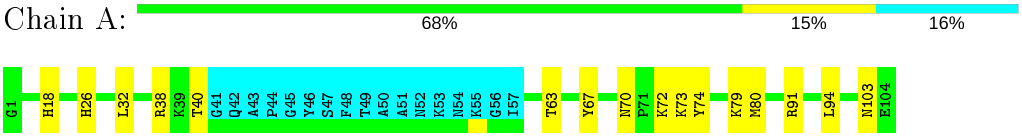
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Cytochrome c

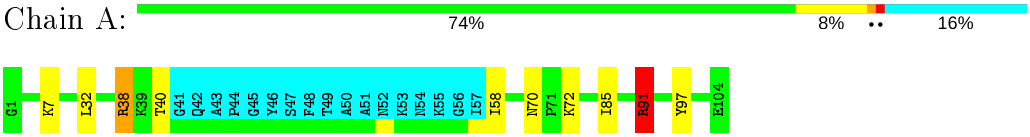
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Cytochrome c

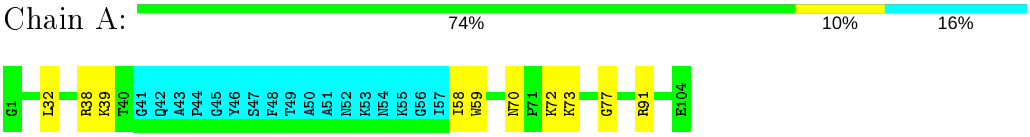
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Cytochrome c

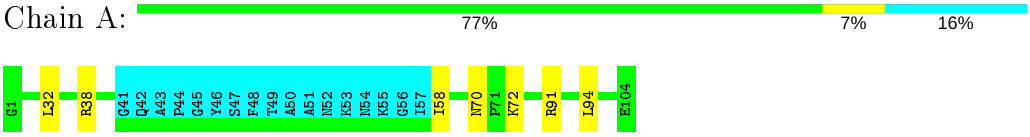
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Cytochrome c

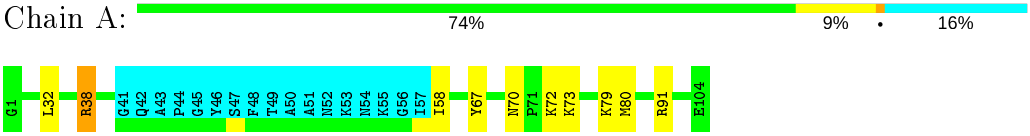
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Cytochrome c

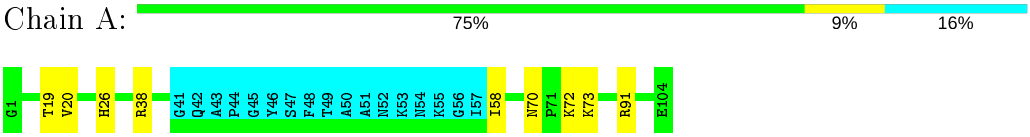
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Cytochrome c

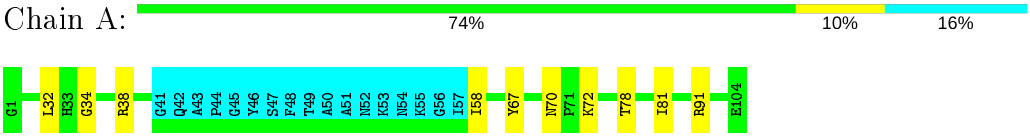
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Cytochrome c

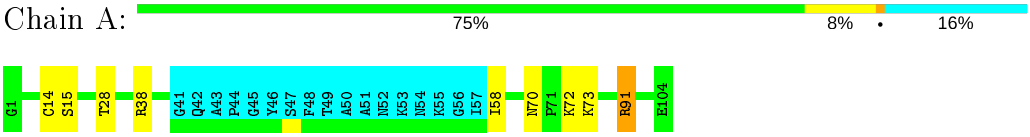
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Cytochrome c

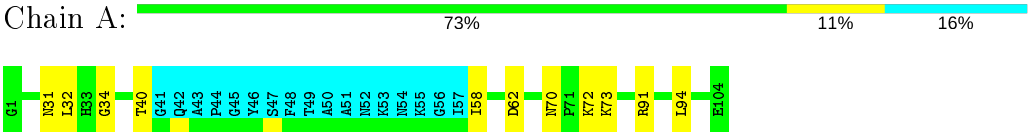
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Cytochrome c

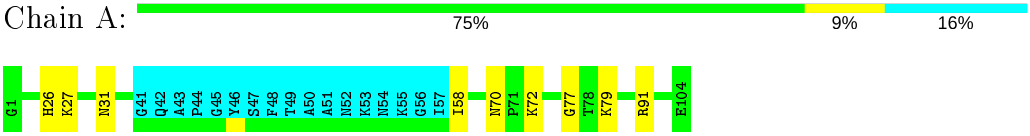
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Cytochrome c

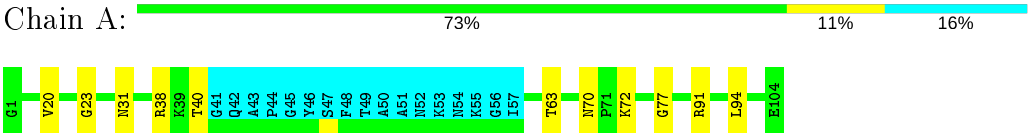
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Cytochrome c

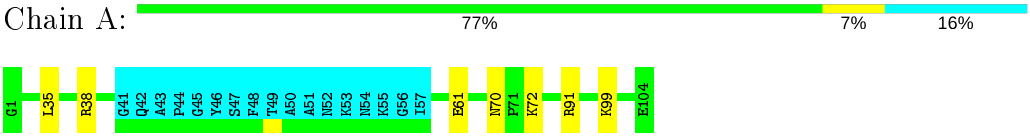
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Cytochrome c

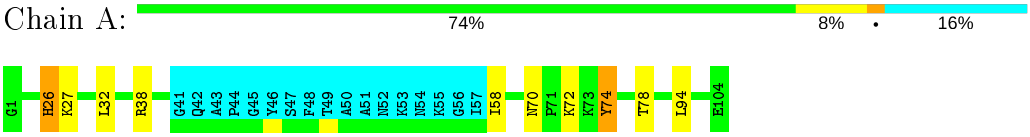
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Cytochrome c

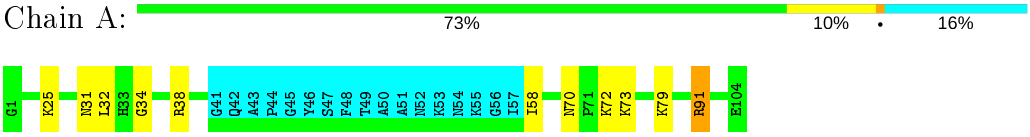
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: Cytochrome c

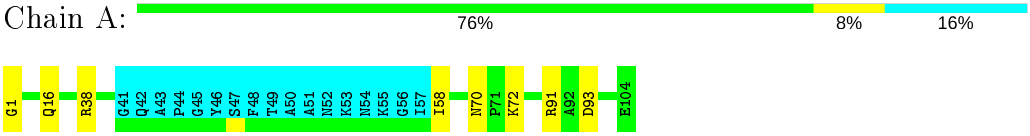
Chain A:



4.2.20 Score per residue for model 20 (medoid)

- Molecule 1: Cytochrome c

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 20 were deposited, based on the following criterion: *target function.*

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

Software name	Classification	Version
CYANA	structure solution	2.1
AMBER	refinement	
TALOS	geometry optimization	
CYANA	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1263
Number of shifts mapped to atoms	1263
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	88%

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	694	725	725	1±1
2	A	43	32	32	0±1
All	All	14740	15140	15140	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:26:HIS:CD2	1:A:27:LYS:H	0.68	2.06	18	1
1:A:7:LYS:HE3	1:A:97:TYR:CZ	0.52	2.40	7	1
2:A:201:MH0:H29	2:A:201:MH0:CGD	0.50	2.36	4	1
1:A:35:LEU:CD1	2:A:201:MH0:H7	0.47	2.39	17	1
1:A:94:LEU:HD11	2:A:201:MH0:H13	0.46	1.88	6	4
1:A:26:HIS:CG	1:A:27:LYS:H	0.44	2.28	18	1
1:A:61:GLU:CG	1:A:99:LYS:HE3	0.43	2.44	17	1
1:A:35:LEU:HD11	2:A:201:MH0:H7	0.43	1.90	17	1
1:A:1:GLY:HA2	1:A:93:ASP:OD1	0.41	2.15	20	1
1:A:85:ILE:HG21	1:A:91:ARG:NH1	0.41	2.31	7	1
1:A:14:CYS:SG	2:A:201:MH0:H13	0.40	2.56	13	1
1:A:7:LYS:HE3	1:A:97:TYR:CE1	0.40	2.52	7	1

5.2 Torsion angles (i)

5.2.1 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	85/104 (82%)	76±2 (89±2%)	8±2 (10±2%)	1±1 (1±1%)	16	63
All	All	1700/2080 (82%)	1510 (89%)	168 (10%)	22 (1%)	16	63

All 11 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	77	GLY	6
1	A	34	GLY	3
1	A	20	VAL	3
1	A	80	MET	2
1	A	78	THR	2
1	A	81	ILE	1
1	A	25	LYS	1
1	A	27	LYS	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	23	GLY	1
1	A	59	TRP	1
1	A	58	ILE	1

5.2.2 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	74/84 (88%)	69±2 (93±2%)	5±2 (7±2%)	19	67
All	All	1480/1680 (88%)	1376 (93%)	104 (7%)	19	67

All 23 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	LYS	20
1	A	70	ASN	20
1	A	32	LEU	13
1	A	58	ILE	12
1	A	73	LYS	8
1	A	91	ARG	4
1	A	31	ASN	4
1	A	26	HIS	3
1	A	40	THR	3
1	A	63	THR	2
1	A	79	LYS	2
1	A	19	THR	2
1	A	15	SER	1
1	A	61	GLU	1
1	A	12	MET	1
1	A	62	ASP	1
1	A	94	LEU	1
1	A	28	THR	1
1	A	74	TYR	1
1	A	16	GLN	1
1	A	87	LYS	1
1	A	39	LYS	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	38	ARG	1

5.2.3 RNA

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains

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
1	1PA	A	48	1	11,15,16	0.92±0.06	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
1	1PA	A	48	1	12,19,21	0.73±0.16	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
1	1PA	A	48	1	-	0±0,7,10,12	0±0,1,1,1

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.

5.4 Carbohydrates (i)

There are no carbohydrates in this entry.

5.5 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	MH0	A	201	1	28,50,50	0.49±0.02	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	MH0	A	201	1	14,82,82	0.99±0.14	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	MH0	A	201	1	-	0±0,10,54,54	-

There are no bond-length outliers.

There are no bond-angle outliers.

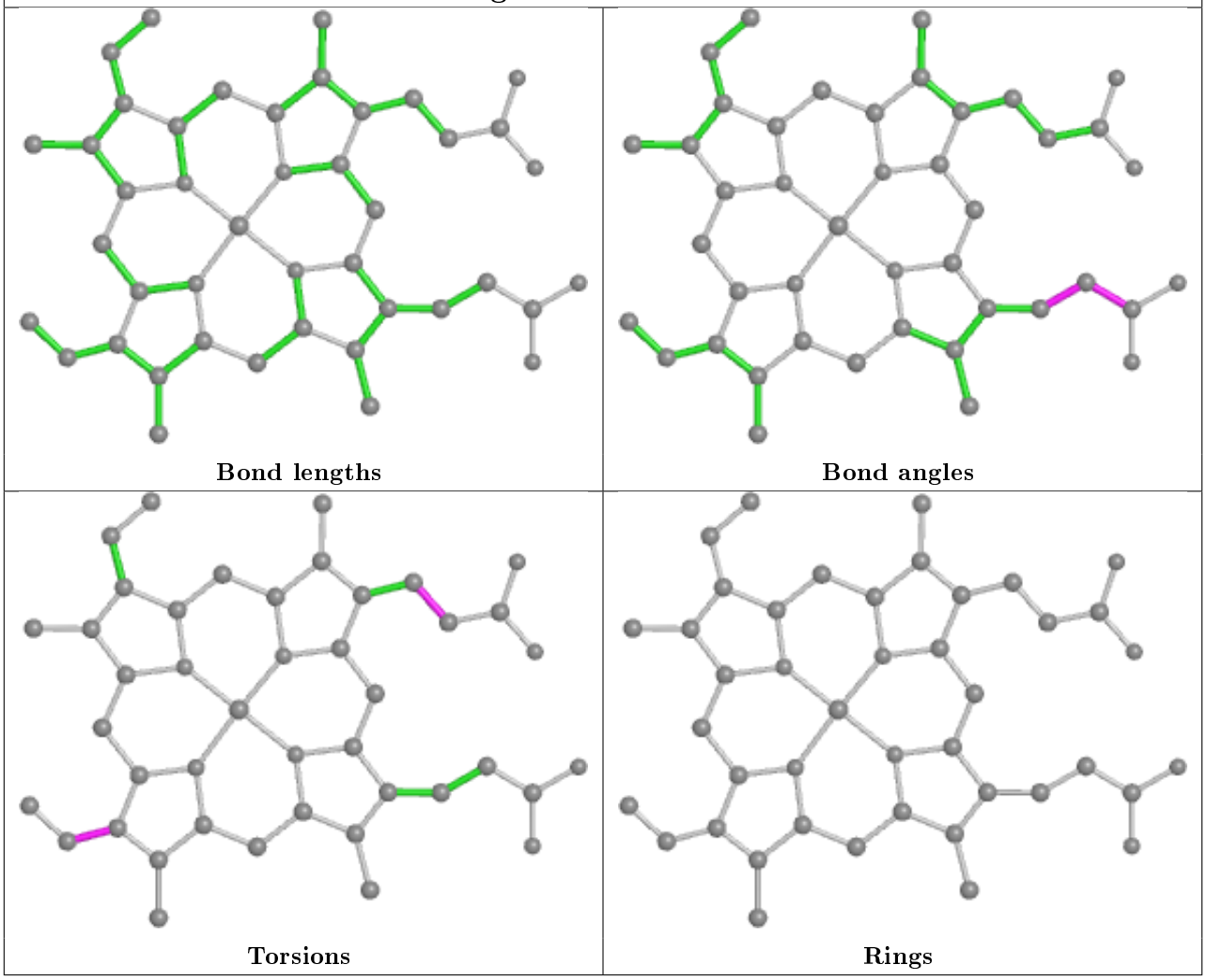
There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 MH0 A 201



5.6 Other polymers (i)

There are no such molecules in this entry.

5.7 Polymer linkage issues (i)

There are no chain breaks in this entry.

6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 88% for the well-defined parts and 87% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 Bookkeeping

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	1263
Number of shifts mapped to atoms	1263
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	23

6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	102	0.20 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	89	0.34 ± 0.13	None needed (< 0.5 ppm)
$^{13}C'$	100	0.26 ± 0.15	None needed (< 0.5 ppm)
^{15}N	88	-0.01 ± 0.28	None needed (< 0.5 ppm)

6.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 88%, i.e. 981 atoms were assigned a chemical shift out of a possible 1111. 9 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	417/429 (97%)	166/171 (97%)	173/174 (99%)	78/84 (93%)
Sidechain	525/598 (88%)	330/354 (93%)	191/218 (88%)	4/26 (15%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	39/84 (46%)	38/45 (84%)	0/35 (0%)	1/4 (25%)
Overall	981/1111 (88%)	534/570 (94%)	364/427 (85%)	83/114 (73%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 87%, i.e. 1119 atoms were assigned a chemical shift out of a possible 1281. 9 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	483/507 (95%)	193/202 (96%)	202/206 (98%)	88/99 (89%)
Sidechain	593/682 (87%)	373/404 (92%)	214/247 (87%)	6/31 (19%)
Aromatic	43/92 (47%)	42/49 (86%)	0/39 (0%)	1/4 (25%)
Overall	1119/1281 (87%)	608/655 (93%)	416/492 (85%)	95/134 (71%)

6.1.4 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	80	MET	HG3	-3.75	4.30 – 0.50	-16.2
1	A	18	HIS	HD2	0.01	9.28 – 4.78	-15.6
1	A	18	HIS	HE1	0.38	10.53 – 5.43	-14.9
1	A	80	MET	HB3	-2.81	3.70 – 0.30	-14.2
1	A	80	MET	HG2	-2.05	4.23 – 0.63	-12.4
1	A	80	MET	HE2	-3.43	4.28 – -0.52	-11.1
1	A	80	MET	HE3	-3.43	4.28 – -0.52	-11.1
1	A	80	MET	HE1	-3.43	4.28 – -0.52	-11.1
1	A	29	GLY	HA2	-0.07	5.87 – 2.07	-10.6
1	A	80	MET	HB2	-0.26	3.73 – 0.33	-6.7
1	A	82	PHE	HB3	0.42	4.85 – 1.05	-6.6
1	A	18	HIS	HB3	0.66	5.00 – 1.10	-6.1
1	A	18	HIS	HB2	1.00	4.91 – 1.31	-5.9
1	A	32	LEU	HD11	-0.87	2.16 – -0.64	-5.8
1	A	32	LEU	HD12	-0.87	2.16 – -0.64	-5.8
1	A	32	LEU	HD13	-0.87	2.16 – -0.64	-5.8
1	A	71	PRO	HG3	-0.01	3.56 – 0.26	-5.8
1	A	30	PRO	HD3	1.52	5.52 – 1.72	-5.5
1	A	97	TYR	HE2	5.46	7.86 – 5.56	-5.4
1	A	32	LEU	HD23	-0.76	2.14 – -0.66	-5.4
1	A	32	LEU	HD22	-0.76	2.14 – -0.66	-5.4
1	A	32	LEU	HD21	-0.76	2.14 – -0.66	-5.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	17	CYS	HB3	0.43	5.25 – 0.55	-5.3

6.1.5 Random Coil Index (RCI) plots

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: