wwPDB NMR Structure Validation Summary Report (i)

Aug 18, 2022 – 12:35 AM EDT

PDB ID	:	2M56
Title	:	The structure of the complex of cytochrome P450cam and its electron donor putidaredoxin determined by paramagnetic NMR spectroscopy
Authors	:	Hiruma, Y.; Hass, M.A.S.; Ubbink, M.
Deposited on	:	2013-02-14

This is a wwPDB NMR Structure Validation Summary 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:

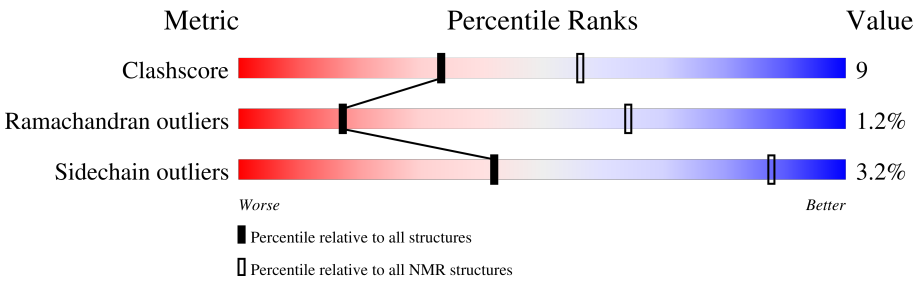
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.29
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.29

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 8%.

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	404	82% 15% •
2	B	106	78% 20% •

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 4 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:11-A:414 (404)	0.00	4
2	B:1-B:106 (106)	0.00	6

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. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 7, 8
2	9, 10

3 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 7977 atoms, of which 3925 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Camphor 5-monooxygenase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	404	6355	2029	3155	559	594	18	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	276	GLN	GLU	SEE REMARK 999	UNP P00183

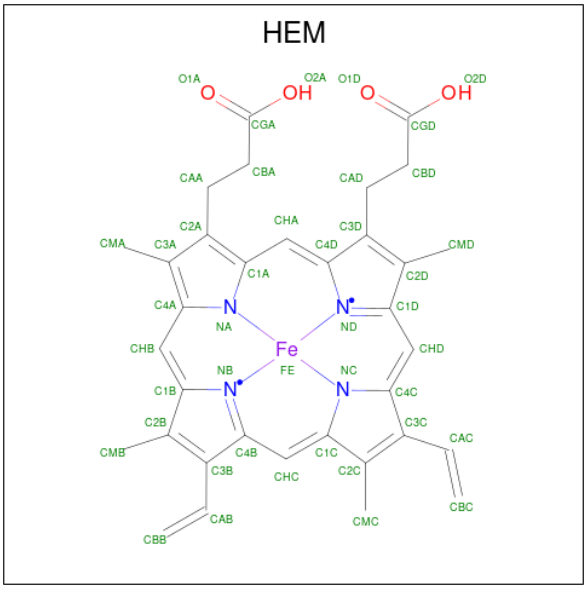
- Molecule 2 is a protein called Putidaredoxin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	106	1564	488	770	136	162	8	0

There is a discrepancy between the modelled and reference sequences:

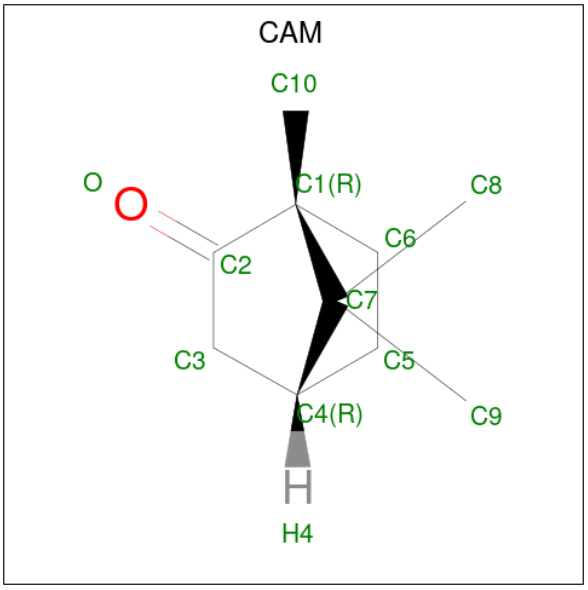
Chain	Residue	Modelled	Actual	Comment	Reference
B	73	SER	CYS	engineered mutation	UNP P00259

- Molecule 3 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula: $C_{34}H_{32}FeN_4O_4$).



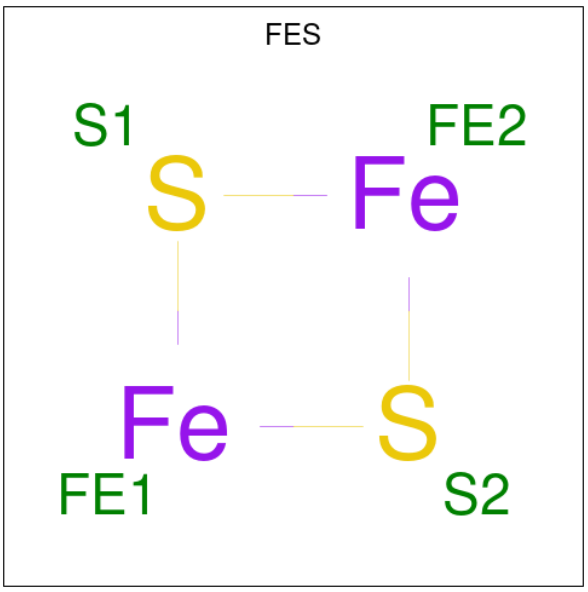
Mol	Chain	Residues	Atoms				
			Total	C	Fe	N	O
3	A	1	43	34	1	4	4

- Molecule 4 is CAMPHOR (three-letter code: CAM) (formula: $C_{10}H_{16}O$).



Mol	Chain	Residues	Atoms		
			Total	C	O
4	A	1	11	10	1

- Molecule 5 is FE2/S2 (INORGANIC) CLUSTER (three-letter code: FES) (formula: Fe_2S_2).



Mol	Chain	Residues	Atoms		
			Total	Fe	S
5	B	1	4	2	2

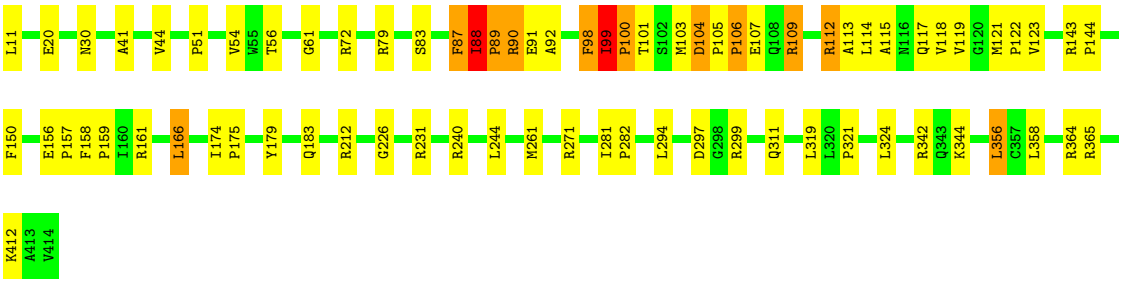
4 Residue-property plots (i)

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.

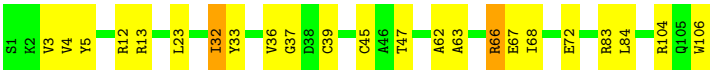
- Molecule 1: Camphor 5-monooxygenase

Chain A: 82% 15% •



- Molecule 2: Putidaredoxin

Chain B: 78% 20% •

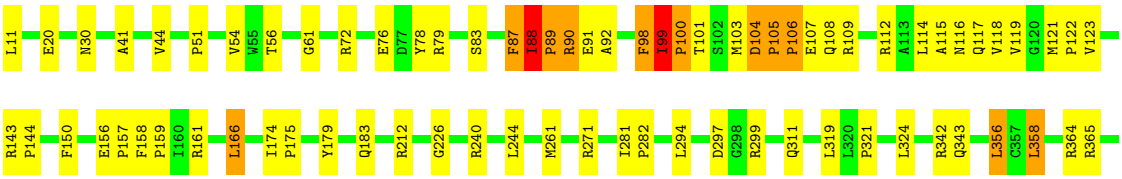


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

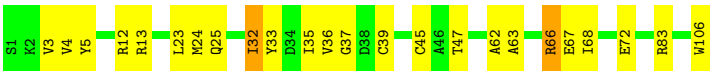
- Molecule 1: Camphor 5-monooxygenase

Chain A: 81% 16% •



• Molecule 2: Putidaredoxin

Chain B:



5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 10 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
X-PLOR NIH	structure solution	
X-PLOR NIH	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	494
Number of shifts mapped to atoms	494
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	8%

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: FES, CAM, HEM

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	1.10±0.00	1±0/3279 (0.0± 0.0%)	1.24±0.00	10±0/4454 (0.2± 0.0%)
2	B	1.08±0.00	0±0/805 (0.0± 0.0%)	1.21±0.00	0±0/1093 (0.0± 0.0%)
All	All	1.10	10/40840 (0.0%)	1.24	97/55470 (0.2%)

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	10.6±0.5
2	B	0.0±0.0	2.8±0.4
All	All	0	134

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	99	ILE	N-CA	5.80	1.57	1.46	1	10

5 of 10 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	107	GLU	N-CA-CB	-7.45	97.20	110.60	2	10
1	A	105	PRO	N-CA-C	7.21	130.85	112.10	1	10
1	A	88	ILE	CA-CB-CG1	-6.39	98.85	111.00	1	10
1	A	89	PRO	N-CA-C	-6.20	95.98	112.10	1	10
1	A	104	ASP	CB-CA-C	6.16	122.73	110.40	1	10

There are no chirality outliers.

5 of 14 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	72	ARG	Sidechain	10
1	A	79	ARG	Sidechain	10
1	A	109	ARG	Sidechain	10
1	A	112	ARG	Sidechain	10
1	A	212	ARG	Sidechain	10

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	3200	3155	3152	63±4
2	B	794	770	771	13±2
3	A	43	0	30	8±3
5	B	4	0	0	2±0
All	All	40520	39250	39687	731

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 9.

5 of 116 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:358:LEU:CD1	3:A:501:HEM:HBD2	1.21	1.65	8	3
1:A:358:LEU:HD11	3:A:501:HEM:HBD2	1.00	1.06	8	1
1:A:99:ILE:HD11	1:A:240:ARG:CZ	0.95	1.90	1	10
1:A:100:PRO:O	3:A:501:HEM:O2D	0.95	1.85	1	10
1:A:100:PRO:O	3:A:501:HEM:CGD	0.89	2.20	1	10

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	402/404 (100%)	381±0 (95±0%)	16±0 (4±0%)	5±0 (1±0%)	17	64
2	B	104/106 (98%)	99±0 (95±0%)	4±0 (4±0%)	1±0 (1±0%)	20	68
All	All	5060/5100 (99%)	4800 (95%)	200 (4%)	60 (1%)	17	64

5 of 6 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	88	ILE	10
1	A	90	ARG	10
1	A	99	ILE	10
1	A	100	PRO	10
1	A	106	PRO	10

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	349/349 (100%)	339±1 (97±0%)	10±1 (3±0%)	46	90
2	B	89/89 (100%)	84±1 (95±1%)	5±1 (5±1%)	27	76
All	All	4380/4380 (100%)	4238 (97%)	142 (3%)	42	88

5 of 30 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	20	GLU	10
1	A	30	ASN	10
1	A	166	LEU	10
1	A	311	GLN	10
2	B	13	ARG	10

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)

3 ligands are 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
3	HEM	A	501	1	41,50,50	1.44±0.00	4±0 (9±0%)
5	FES	B	201	2	0,4,4	0.00±0.00	-
4	CAM	A	502	-	12,12,12	0.53±0.00	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
3	HEM	A	501	1	45,82,82	1.37±0.00	7±0 (15±0%)
5	FES	B	201	2	-	-	-
4	CAM	A	502	-	20,21,21	2.72±0.00	7±0 (35±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
4	CAM	A	502	-	-	-	0±0,3,2,2
5	FES	B	201	2	-	-	0±0,1,1,1
3	HEM	A	501	1	-	0±0,12,54,54	-

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
3	A	501	HEM	C3C-C2C	4.37	1.34	1.40	1	10
3	A	501	HEM	CAA-C2A	3.35	1.57	1.52	1	10
3	A	501	HEM	C3C-CAC	3.12	1.54	1.47	1	10
3	A	501	HEM	CMD-C2D	2.22	1.55	1.50	1	10

5 of 14 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
4	A	502	CAM	C7-C1-C2	5.45	109.56	100.30	1	10
4	A	502	CAM	C3-C2-C1	5.44	96.54	107.34	1	10
4	A	502	CAM	C4-C3-C2	4.72	110.10	102.03	1	10
4	A	502	CAM	C5-C4-C3	4.60	93.73	106.40	1	10
4	A	502	CAM	O-C2-C3	4.19	135.18	126.47	1	10

There are no chirality outliers.

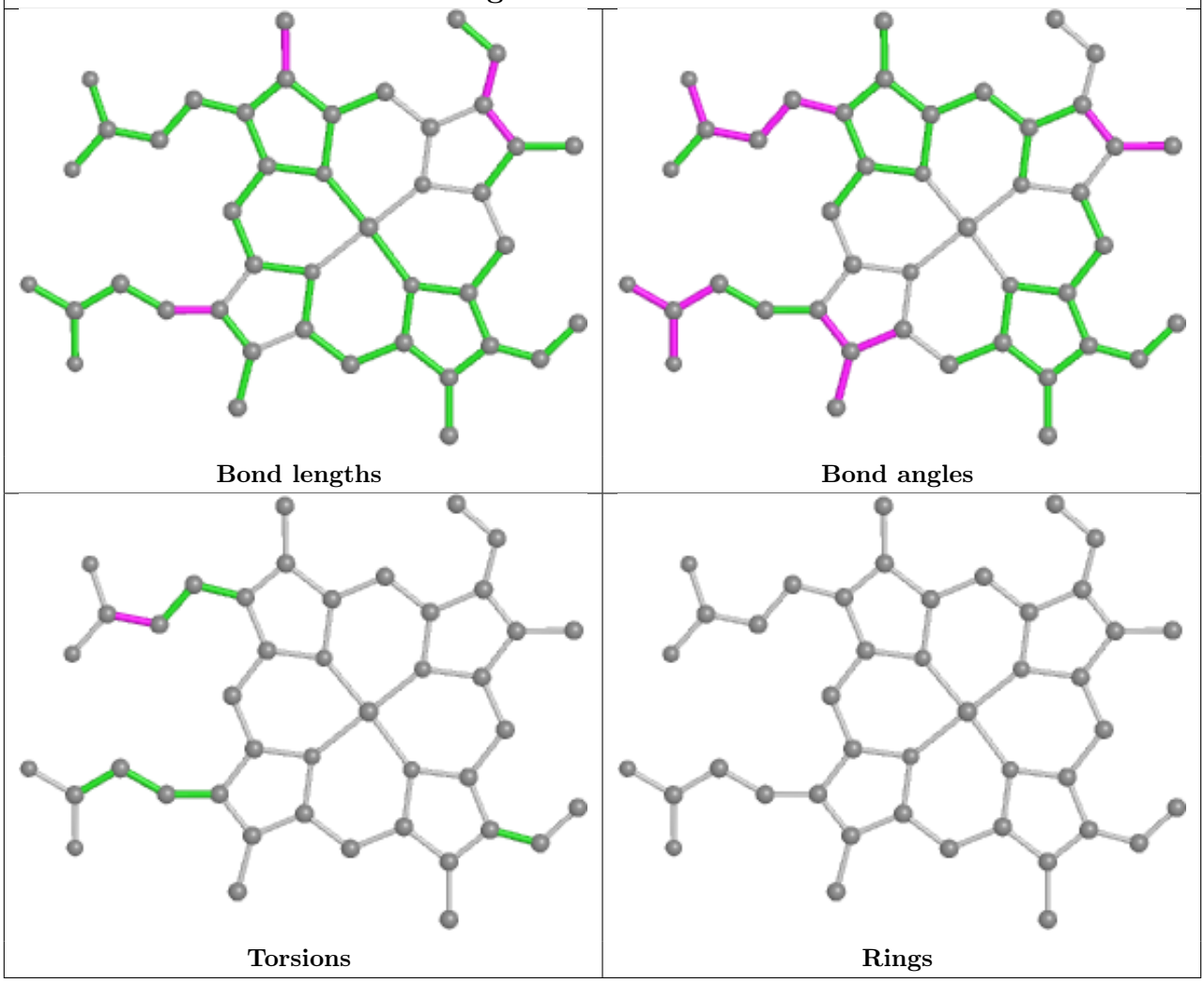
There are no torsion outliers.

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 HEM A 501



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 8% for the well-defined parts and 8% 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	494
Number of shifts mapped to atoms	494
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	None (insufficient data)
$^{13}C_{\beta}$	0	—	None (insufficient data)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	247	-0.01 ± 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 8%, i.e. 493 atoms were assigned a chemical shift out of a possible 6237. 0 out of 86 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	493/2482 (20%)	246/986 (25%)	0/1020 (0%)	247/476 (52%)
Sidechain	0/3311 (0%)	0/1944 (0%)	0/1218 (0%)	0/149 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/444 (0%)	0/239 (0%)	0/184 (0%)	0/21 (0%)
Overall	493/6237 (8%)	246/3169 (8%)	0/2422 (0%)	247/646 (38%)

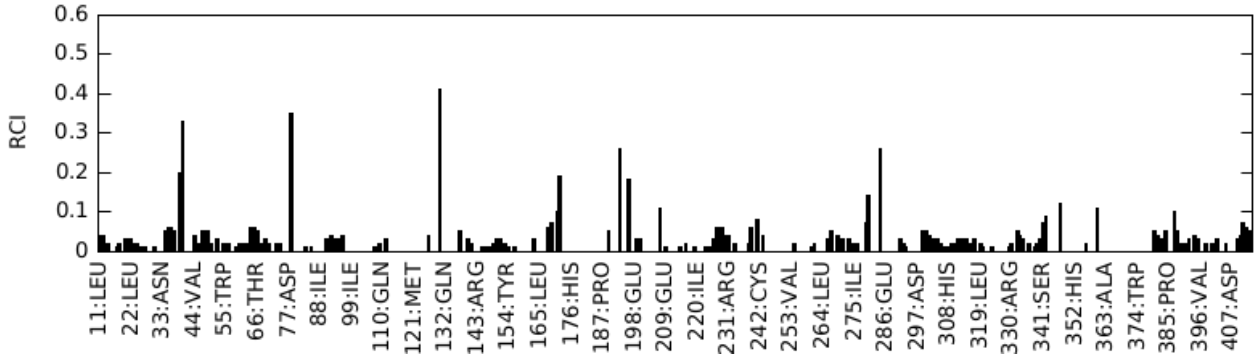
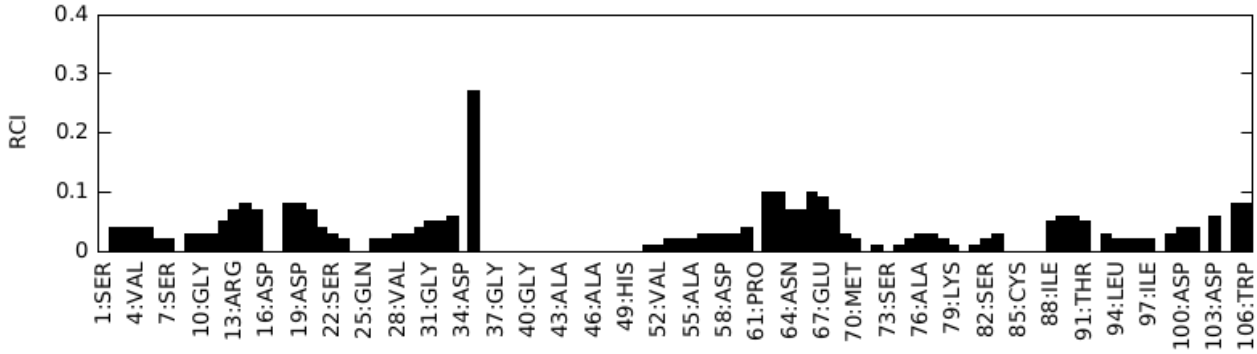
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The images below report *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:



Random coil index (RCI) for chain B: