wwPDB NMR Structure Validation Summary Report

Jun 6, 2023 – 02:21 pm BST

PDB ID	:	5MCS
BMRB ID	:	34064
Title	:	Solution structure and dynamics of the outer membrane cytochrome OmcF from Geobacter sulfurreducens
Authors	:	Dantas, J.M.; Silva, M.A.; Morgado, L.; Pantoja-Uceda, D.; Turner, D.L.; Bruix, M.; Salgueiro, C.A.
Deposited on	:	2016-11-10

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:

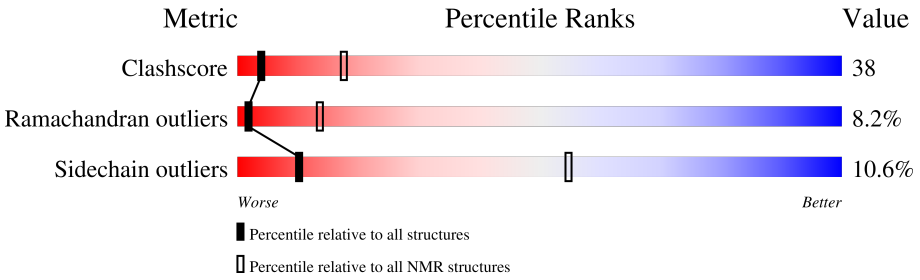
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
BMRB Restraints Analysis	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.33

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 56%.

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	80	55% 36% 6%

2 Ensemble composition and analysis

This entry contains 20 models. Model 18 is the overall representative, medoid model (most similar to other models). The authors have identified model 20 as representative, based on the following criterion: *target function*.

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:26-A:104 (79)	0.28	18

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 2 single-model clusters were found.

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

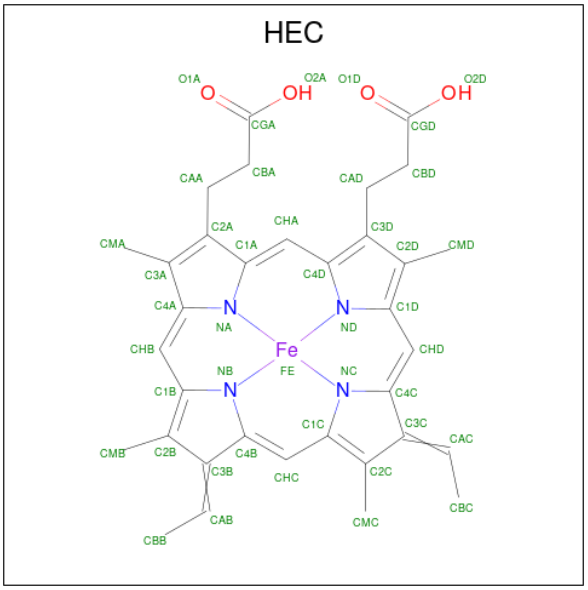
3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1176 atoms, of which 550 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Lipoprotein cytochrome c, 1 heme-binding site.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	80	1133	364	550	110	105	4	0

- Molecule 2 is HEME C (three-letter code: HEC) (formula: $C_{34}H_{34}FeN_4O_4$).



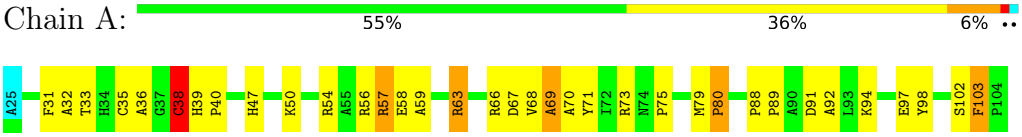
Mol	Chain	Residues	Atoms				
			Total	C	Fe	N	O
2	A	1	43	34	1	4	4

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, 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.

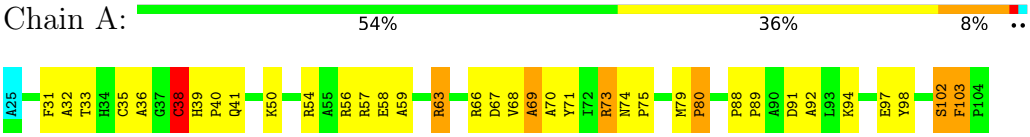
• Molecule 1: Lipoprotein cytochrome c, 1 heme-binding site



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

The representative model is number 18. Colouring as in section 4.1 above.

• Molecule 1: Lipoprotein cytochrome c, 1 heme-binding site



5 Refinement protocol and experimental data overview

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

Of the 100 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
DYANA	refinement	
DYANA	structure calculation	

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	604
Number of shifts mapped to atoms	512
Number of unparsed shifts	0
Number of shifts with mapping errors	92
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	56%

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: HEC

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	578	545	388	35±4
2	A	43	0	30	6±2
All	All	12420	10900	8360	780

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 38.

5 of 94 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:32:ALA:HA	1:A:36:ALA:HB2	0.78	1.55	3	20
1:A:88:PRO:HG2	1:A:91:ASP:CG	0.74	2.03	9	20
1:A:31:PHE:CG	1:A:98:TYR:CE1	0.70	2.79	19	4
1:A:88:PRO:HG2	1:A:91:ASP:OD1	0.67	1.89	5	8
1:A:61:GLY:C	1:A:63:ARG:HH11	0.66	1.93	17	4

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	114/80 (142%)	79±2 (69±1%)	26±2 (23±2%)	9±1 (8±1%)	2	13
All	All	2280/1600 (142%)	1579 (69%)	514 (23%)	187 (8%)	2	13

5 of 15 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	33[1]	THR	20
1	A	33[2]	THR	20
1	A	33[3]	THR	20
1	A	38	CYS	20
1	A	69	ALA	20

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	92/56 (164%)	82±2 (89±2%)	10±2 (11±2%)	10	55
All	All	1840/1120 (164%)	1645 (89%)	195 (11%)	10	55

5 of 27 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	CYS	20
1	A	94	LYS	20
1	A	103	PHE	20
1	A	63	ARG	18
1	A	56	ARG	14

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	HEC	A	201	1	32,50,50	1.59±0.02	2±0 (7±1%)

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	HEC	A	201	1	24,82,82	1.36±0.00	3±0 (12±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	HEC	A	201	1	-	0±0,10,54,54	-

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
2	A	201	HEC	C3C-C2C	5.22	1.35	1.40	3	20
2	A	201	HEC	C2B-C3B	5.16	1.35	1.40	16	20
2	A	201	HEC	C1C-CHC	2.57	1.48	1.41	17	6

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	201	HEC	CMC-C2C-C1C	2.36	124.84	128.46	13	20
2	A	201	HEC	CMB-C2B-C1B	2.35	124.85	128.46	7	20
2	A	201	HEC	CMD-C2D-C1D	2.35	124.86	128.46	12	20

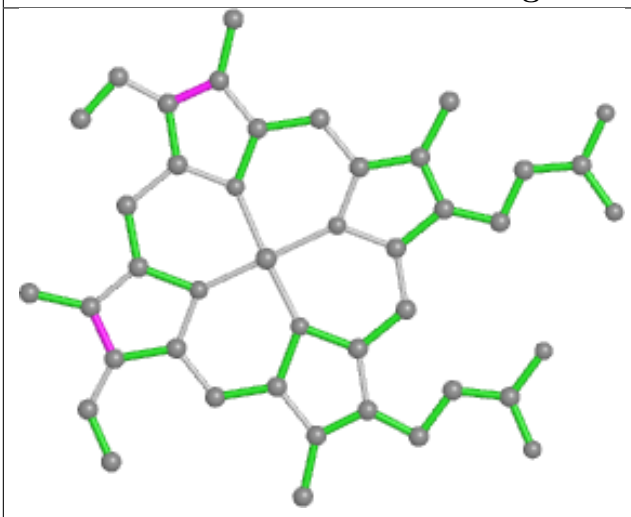
There are no chirality outliers.

There are no torsion outliers.

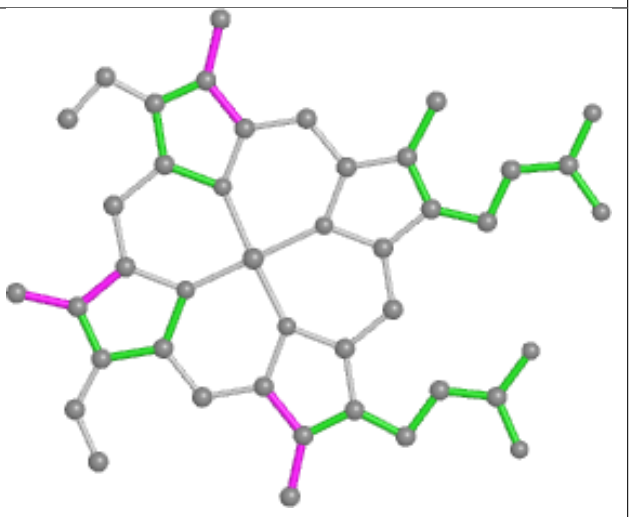
There are no ring outliers.

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.

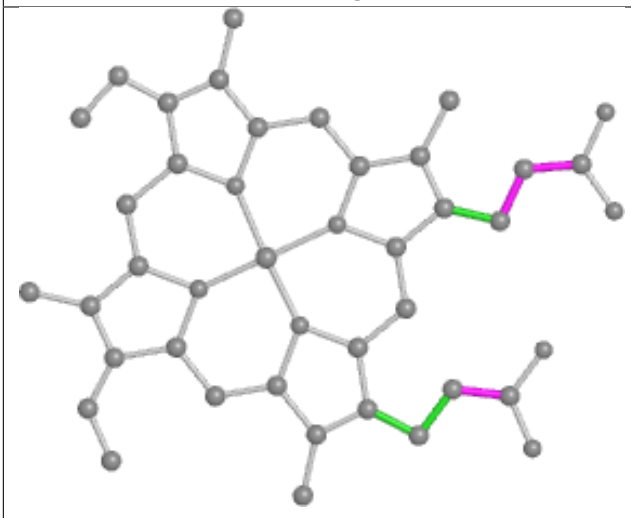
Ligand HEC A 201



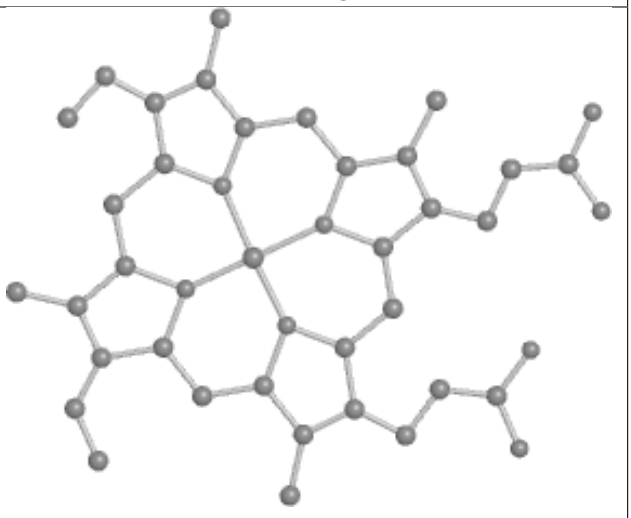
Bond lengths



Bond angles



Torsions



Rings

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

The completeness of assignment taking into account all chemical shift lists is 56% for the well-defined parts and 56% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chemical_shifts_1*

7.1.1 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	604
Number of shifts mapped to atoms	512
Number of unparsed shifts	0
Number of shifts with mapping errors	92
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	20

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 92) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	30	LEU	HD12	0.985	0.003	.
1	A	30	LEU	HD13	0.985	0.003	.
1	A	30	LEU	HD21	1.135	0.004	.
1	A	30	LEU	HD23	1.135	0.004	.
1	A	33	THR	HG21	0.662	0.003	.
1	A	33	THR	HG23	0.662	0.003	.
1	A	41	GLN	HE22	6.839	0.004	.
1	A	44	ASN	HD22	7.268	0.004	.
1	A	45	THR	HG21	0.904	0.003	.
1	A	45	THR	HG23	0.904	0.003	.
1	A	46	VAL	HG12	1.214	0.002	.
1	A	46	VAL	HG13	1.214	0.002	.
1	A	46	VAL	HG21	1.41	0.002	.
1	A	46	VAL	HG23	1.41	0.002	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	51	THR	HG21	1.055	0.004	.
1	A	51	THR	HG23	1.055	0.004	.
1	A	52	LEU	HD12	-0.434	0.003	.
1	A	52	LEU	HD13	-0.434	0.003	.
1	A	52	LEU	HD21	-1.072	0.003	.
1	A	52	LEU	HD23	-1.072	0.003	.
1	A	60	ASN	HD22	7.58	0.007	.
1	A	62	ILE	HD12	0.786	0.003	.
1	A	62	ILE	HD13	0.786	0.003	.
1	A	62	ILE	HG21	0.909	0.002	.
1	A	62	ILE	HG23	0.909	0.002	.
1	A	64	THR	HG21	1.317	0.003	.
1	A	64	THR	HG23	1.317	0.003	.
1	A	65	VAL	HG12	1.145	0.003	.
1	A	65	VAL	HG13	1.145	0.003	.
1	A	65	VAL	HG21	1.155	0.002	.
1	A	65	VAL	HG23	1.155	0.002	.
1	A	68	VAL	HG12	1.535	0.003	.
1	A	68	VAL	HG13	1.535	0.003	.
1	A	68	VAL	HG21	1.207	0.004	.
1	A	68	VAL	HG23	1.207	0.004	.
1	A	72	ILE	HD12	0.696	0.006	.
1	A	72	ILE	HD13	0.696	0.006	.
1	A	72	ILE	HG21	0.827	0.018	.
1	A	72	ILE	HG23	0.827	0.018	.
1	A	74	ASN	HD22	6.615	0.002	.
1	A	87	ILE	HD12	0.135	0.002	.
1	A	87	ILE	HD13	0.135	0.002	.
1	A	87	ILE	HG21	0.72	0.002	.
1	A	87	ILE	HG23	0.72	0.002	.
1	A	93	LEU	HD12	0.955	0.005	.
1	A	93	LEU	HD13	0.955	0.005	.
1	A	93	LEU	HD21	0.887	0.004	.
1	A	93	LEU	HD23	0.887	0.004	.
1	A	95	ILE	HD12	1.104	0.007	.
1	A	95	ILE	HD13	1.104	0.007	.
1	A	95	ILE	HG21	1.591	0.004	.
1	A	95	ILE	HG23	1.591	0.004	.
1	A	99	VAL	HG12	1.278	0.004	.
1	A	99	VAL	HG13	1.278	0.004	.
1	A	99	VAL	HG21	1.677	0.003	.

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	99	VAL	HG23	1.677	0.003	.
1	A	100	VAL	HG12	0.962	0.002	.
1	A	100	VAL	HG13	0.962	0.002	.
1	A	100	VAL	HG21	1.015	0.003	.
1	A	100	VAL	HG23	1.015	0.003	.
1	A	201	HEC	HAD1	4.152	0.004	.
1	A	201	HEC	HAD2	4.446	0.002	.
1	A	201	HEC	HAA1	2.693	0.002	.
1	A	201	HEC	HAA2	4.114	0.003	.
1	A	201	HEC	HHA	10.097	0.003	.
1	A	201	HEC	HBD1	2.887	0.010	.
1	A	201	HEC	HBD2	3.241	0.004	.
1	A	201	HEC	HBA1	2.014	0.004	.
1	A	201	HEC	HBA2	3.119	0.002	.
1	A	201	HEC	HHB	9.417	0.002	.
1	A	201	HEC	HHD	9.334	0.003	.
1	A	201	HEC	HHC	9.588	0.002	.
1	A	201	HEC	HAB	5.479	0.003	.
1	A	201	HEC	HAC	6.195	0.003	.
1	A	201	HEC	HMB1	3.638	0.002	.
1	A	201	HEC	HMB2	3.638	0.002	.
1	A	201	HEC	HMB3	3.638	0.002	.
1	A	201	HEC	HMC1	4.074	0.003	.
1	A	201	HEC	HMC2	4.074	0.003	.
1	A	201	HEC	HMC3	4.074	0.003	.
1	A	201	HEC	HMD1	3.197	0.004	.
1	A	201	HEC	HMD2	3.197	0.004	.
1	A	201	HEC	HMD3	3.197	0.004	.
1	A	201	HEC	HMA1	3.024	0.004	.
1	A	201	HEC	HMA2	3.024	0.004	.
1	A	201	HEC	HMA3	3.024	0.004	.
1	A	201	HEC	HBB1	1.141	0.003	.
1	A	201	HEC	HBB2	1.141	0.003	.
1	A	201	HEC	HBB3	1.141	0.003	.
1	A	201	HEC	HBC1	2.5	0.002	.
1	A	201	HEC	HBC2	2.5	0.002	.
1	A	201	HEC	HBC3	2.5	0.002	.

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	69	1.29 ± 0.45	Should be applied

7.1.3 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 56%, i.e. 566 atoms were assigned a chemical shift out of a possible 1005. 0 out of 8 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	228/390 (58%)	160/161 (99%)	0/158 (0%)	68/71 (96%)
Sidechain	318/546 (58%)	314/358 (88%)	0/164 (0%)	4/24 (17%)
Aromatic	20/69 (29%)	20/35 (57%)	0/31 (0%)	0/3 (0%)
Overall	566/1005 (56%)	494/554 (89%)	0/353 (0%)	72/98 (73%)

7.1.4 Statistically unusual chemical shifts

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.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	79	MET	HB2	-2.99	0.42 – 3.63	-15.7
1	A	79	MET	HG3	-3.29	0.54 – 4.26	-15.3
1	A	39	HIS	HD2	0.52	4.65 – 9.35	-13.8
1	A	79	MET	HE1	-2.85	-0.03 – 3.80	-12.4
1	A	79	MET	HE2	-2.85	-0.03 – 3.80	-12.4
1	A	79	MET	HE3	-2.85	-0.03 – 3.80	-12.4
1	A	39	HIS	HE1	1.15	5.13 – 10.76	-12.1
1	A	79	MET	HG2	-1.25	0.65 – 4.19	-10.4
1	A	102	SER	HB3	1.59	2.49 – 5.20	-8.3
1	A	79	MET	HB3	-0.52	0.33 – 3.66	-7.5
1	A	75	PRO	HB3	-0.28	0.25 – 3.76	-6.5
1	A	52	LEU	HD21	-1.07	-0.65 – 2.13	-6.5
1	A	52	LEU	HD22	-1.07	-0.65 – 2.13	-6.5
1	A	52	LEU	HD23	-1.07	-0.65 – 2.13	-6.5
1	A	102	SER	HB2	2.27	2.61 – 5.13	-6.4
1	A	39	HIS	HB2	1.00	1.36 – 4.85	-6.0
1	A	75	PRO	HB2	0.04	0.37 – 3.78	-6.0

Continued on next page...

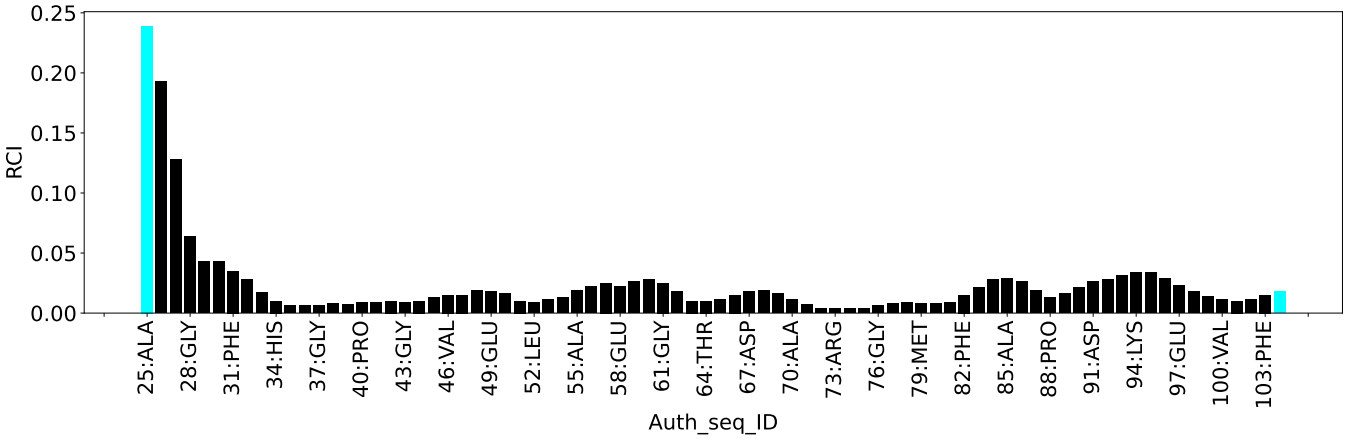
Continued from previous page...

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	42	GLY	HA3	1.76	2.08 – 5.71	-5.9
1	A	38	CYS	HB2	0.63	0.81 – 5.11	-5.4
1	A	75	PRO	HA	2.77	2.78 – 6.00	-5.0

7.1.5 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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:



8 NMR restraints analysis (i)

No restraints data found

9 Distance violation analysis (i)

No distance restraints data found

10 Dihedral-angle violation analysis (i)

No dihedral-angle restraints found