wwPDB NMR Structure Validation Summary Report

Jun 14, 2020 – 01:22 am BST

PDB ID	:	1E8E
Title	:	Solution Structure of Methylophilus methylotrophus Cytochrome c". Insights into the Structural Basis of Haem-Ligand Detachment
Authors	:	Brennan, L.; Turner, D.L.; Fareleira, P.; Santos, H.
Deposited on	:	2000-09-20

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 following versions of software and data (see references (1)) were used in the production of this report:

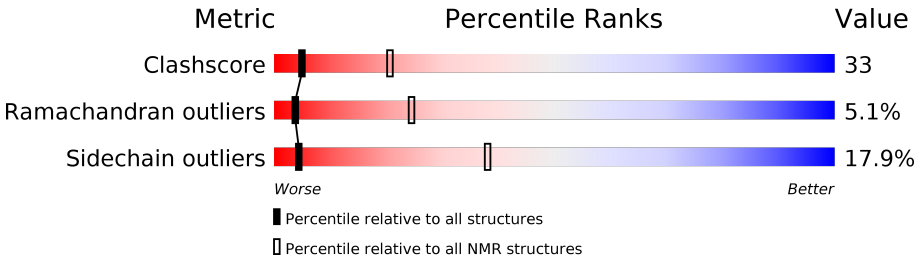
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)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

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

The overall completeness of chemical shifts assignment is 48%.

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	124	33% 56% 5% 6%

2 Ensemble composition and analysis

This entry contains 20 models. Model 10 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:3-A:118 (116)	0.29	10

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 3 clusters and 3 single-model clusters were found.

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

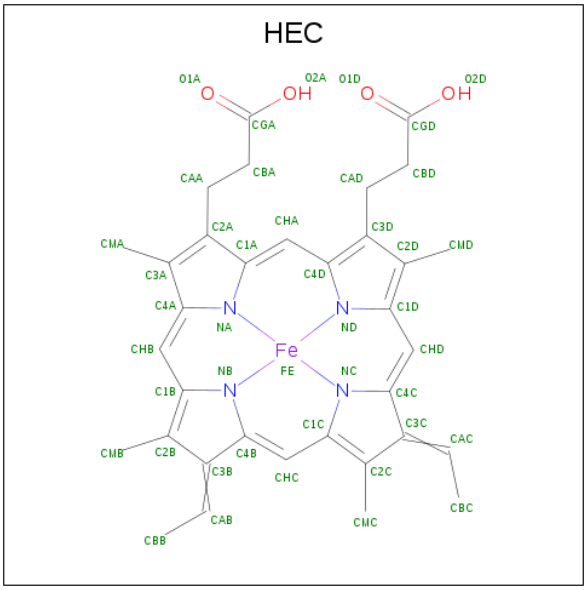
3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1986 atoms, of which 982 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called CYTOCHROME C".

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	124	1911	604	950	164	188	5	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	H	N	O
2	A	1	75	34	1	32	4	4

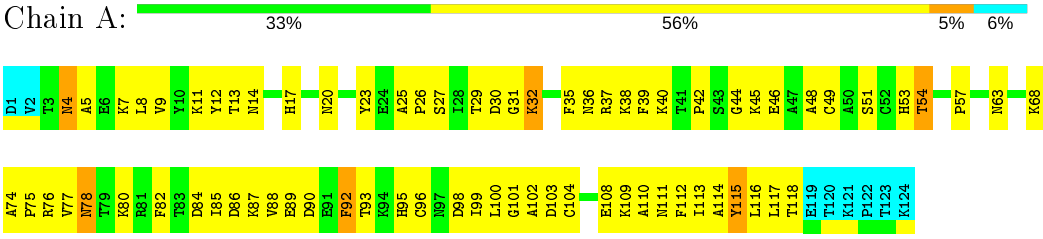
4 Residue-property plots (i)

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:

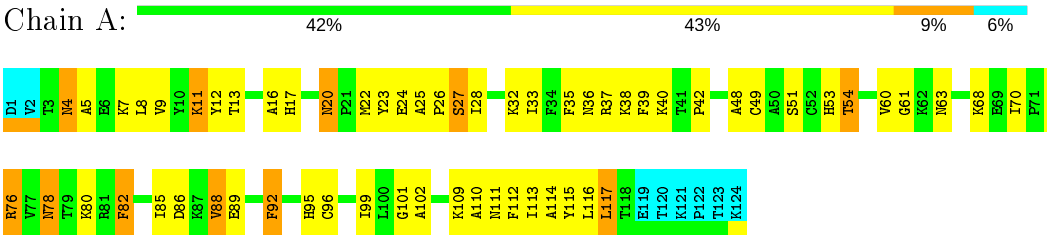


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 10. Colouring as in section 4.1 above.

- Molecule 1: CYTOCHROME C"

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *TORSION ANGLE DYNAMICS WITH SIMULATED ANNEALING*.

Of the 400 calculated structures, 20 were deposited, based on the following criterion: *LEAST RESTRAINT VIOLATION*.

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

Software name	Classification	Version
PARADYANA	refinement	
PARADYANA	structure solution	

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)	input_cs.cif
Number of chemical shift lists	2
Total number of shifts	885
Number of shifts mapped to atoms	857
Number of unparsed shifts	0
Number of shifts with mapping errors	28
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	48%

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

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	897	882	882	59±6
2	A	43	32	30	8±2
All	All	18800	18280	18240	1217

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 33.

5 of 329 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:73:LEU:HD12	1:A:73:LEU:O	0.92	1.65	20	2
1:A:5:ALA:HB1	1:A:117:LEU:HD21	0.86	1.48	11	4
1:A:5:ALA:CB	1:A:117:LEU:HD21	0.86	2.00	6	3
1:A:63:ASN:ND2	1:A:70:ILE:HD11	0.86	1.85	8	18
1:A:53:HIS:CE1	1:A:73:LEU:HD22	0.84	2.07	9	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	116/124 (94%)	86±3 (74±3%)	24±3 (21±2%)	6±1 (5±1%)	4 24
All	All	2320/2480 (94%)	1714 (74%)	487 (21%)	119 (5%)	4 24

5 of 23 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	54	THR	20
1	A	38	LYS	18
1	A	44	GLY	13
1	A	45	LYS	9
1	A	88	VAL	9

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	99/107 (93%)	81±2 (82±2%)	18±2 (18±2%)	4 38
All	All	1980/2140 (93%)	1626 (82%)	354 (18%)	4 38

5 of 48 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	92	PHE	20
1	A	78	ASN	19
1	A	4	ASN	19
1	A	32	LYS	18
1	A	7	LYS	15

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 carbohydrates 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	HEC	A	125	1	26,50,50	1.69±0.01	2±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	HEC	A	125	1	18,82,82	1.35±0.01	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	HEC	A	125	1	-	0±0,6,54,54	-

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	125	HEC	C3B-C2B	5.21	1.35	1.40	7	20
2	A	125	HEC	C3C-C2C	5.16	1.35	1.40	4	20

There are no bond-angle outliers.

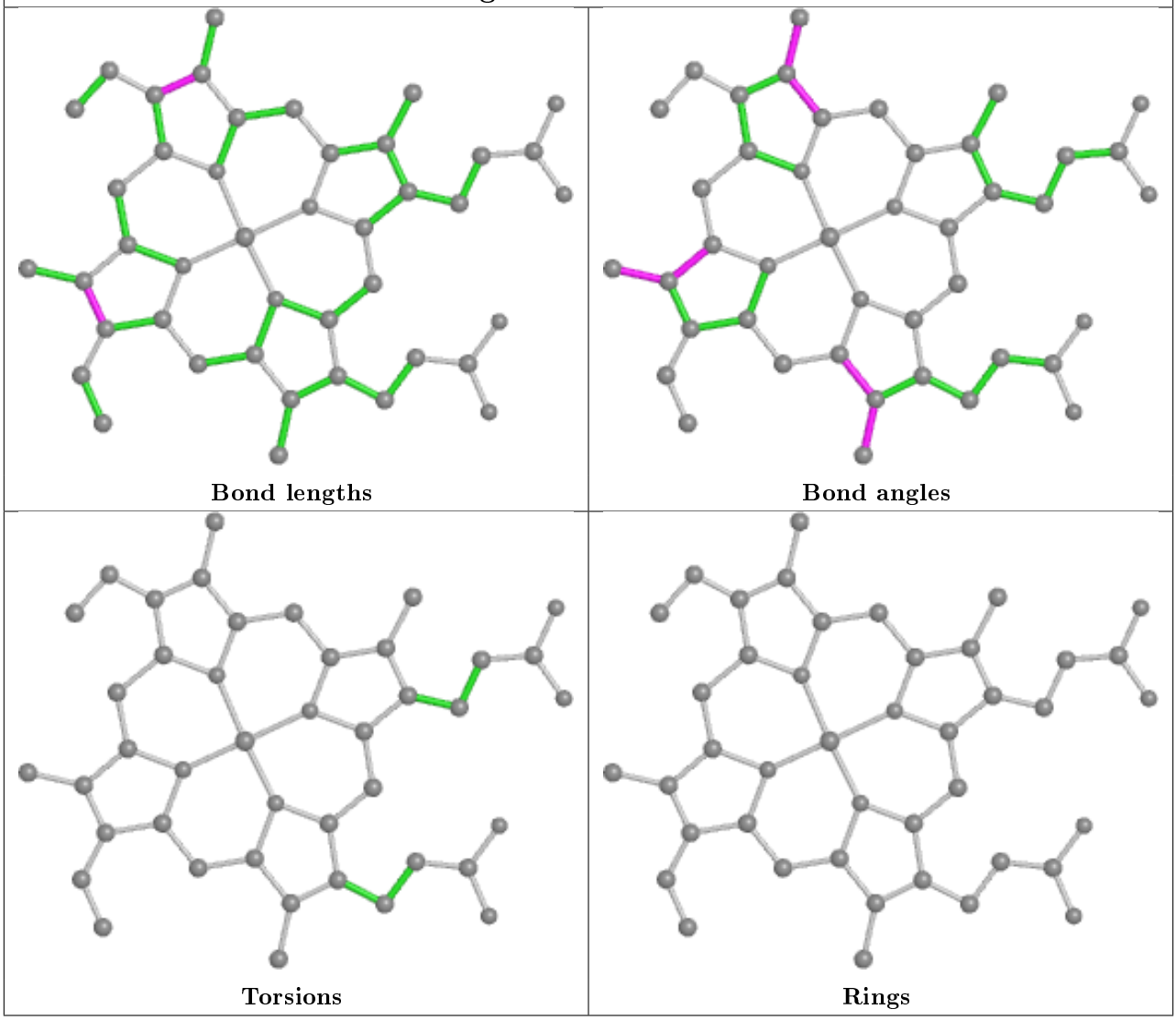
There are no chirality outliers.

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 HEC A 125



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 48% for the well-defined parts and 48% for the entire structure.

7.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_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	857
Number of shifts mapped to atoms	857
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	41

7.1.2 Chemical shift referencing

No chemical shift referencing corrections were calculated (not enough data).

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 48%, i.e. 674 atoms were assigned a chemical shift out of a possible 1391. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	223/564 (40%)	223/224 (100%)	0/232 (0%)	0/108 (0%)
Sidechain	402/720 (56%)	402/424 (95%)	0/264 (0%)	0/32 (0%)
Aromatic	49/107 (46%)	49/58 (84%)	0/46 (0%)	0/3 (0%)
Overall	674/1391 (48%)	674/706 (95%)	0/542 (0%)	0/143 (0%)

7.1.4 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	95	HIS	HB2	20.64	4.91 – 1.31	48.7
1	A	53	HIS	HB2	15.43	4.91 – 1.31	34.2
1	A	95	HIS	HB3	11.89	5.00 – 1.10	22.7
1	A	53	HIS	HB3	10.83	5.00 – 1.10	19.9
1	A	53	HIS	HA	11.38	6.81 – 2.41	15.4
1	A	95	HIS	HA	11.24	6.81 – 2.41	15.1
1	A	73	LEU	HB2	5.90	3.32 – -0.08	12.6
1	A	99	ILE	HG12	5.56	3.27 – -0.73	10.7
1	A	98	ASP	HB2	5.56	4.07 – 1.37	10.5
1	A	108	GLU	HG3	0.07	3.31 – 1.21	-10.4
1	A	73	LEU	HD23	3.41	2.14 – -0.66	9.5
1	A	73	LEU	HD21	3.41	2.14 – -0.66	9.5
1	A	73	LEU	HD22	3.41	2.14 – -0.66	9.5
1	A	99	ILE	HG23	-1.49	2.13 – -0.57	-8.4
1	A	99	ILE	HG22	-1.49	2.13 – -0.57	-8.4
1	A	99	ILE	HG21	-1.49	2.13 – -0.57	-8.4
1	A	109	LYS	HD2	-0.30	2.76 – 0.46	-8.3
1	A	109	LYS	HD3	-0.30	2.75 – 0.45	-8.3
1	A	98	ASP	HB3	4.92	4.07 – 1.27	8.0
1	A	57	PRO	HA	6.90	6.05 – 2.75	7.6
1	A	94	LYS	HG2	3.26	2.67 – 0.07	7.3
1	A	94	LYS	HB2	3.58	3.03 – 0.53	7.2
1	A	35	PHE	HZ	9.98	9.11 – 4.91	7.1
1	A	53	HIS	H	12.81	11.68 – 4.78	6.6
1	A	61	GLY	HA2	6.48	5.87 – 2.07	6.6
1	A	54	THR	HG23	2.62	2.29 – -0.01	6.4
1	A	54	THR	HG21	2.62	2.29 – -0.01	6.4
1	A	54	THR	HG22	2.62	2.29 – -0.01	6.4
1	A	49	CYS	HB3	-0.03	5.25 – 0.55	-6.2
1	A	81	ARG	HD3	1.57	4.36 – 1.86	-6.2
1	A	49	CYS	HB2	0.22	5.20 – 0.70	-6.1
1	A	95	HIS	H	12.26	11.68 – 4.78	5.8
1	A	61	GLY	HA3	6.10	5.80 – 2.00	5.8
1	A	94	LYS	HB3	3.27	3.10 – 0.40	5.6
1	A	35	PHE	HE1	8.87	8.69 – 5.49	5.6
1	A	35	PHE	HE2	8.87	8.69 – 5.49	5.6
1	A	75	PRO	HG2	0.24	3.48 – 0.38	-5.5
1	A	75	PRO	HB3	0.09	3.81 – 0.21	-5.3

Continued on next page...

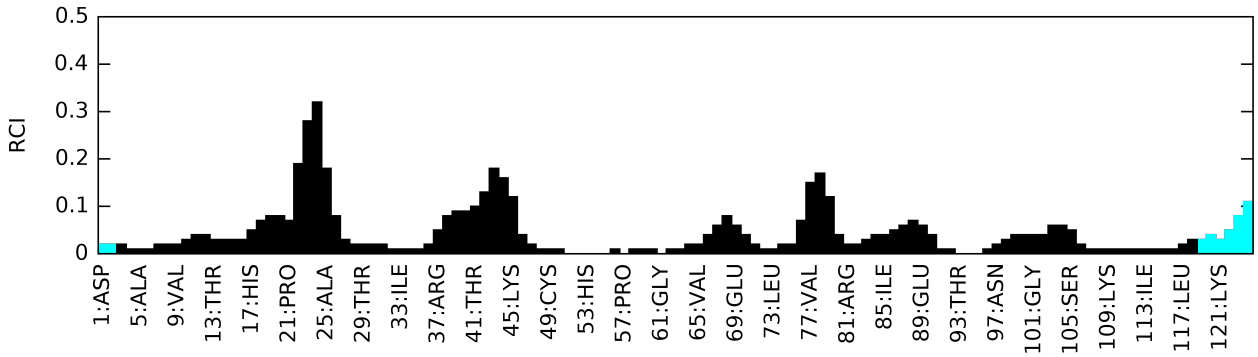
Continued from previous page...

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	100	LEU	HD23	-0.70	2.14 – -0.66	-5.1
1	A	100	LEU	HD21	-0.70	2.14 – -0.66	-5.1
1	A	100	LEU	HD22	-0.70	2.14 – -0.66	-5.1

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.

Random coil index (RCI) for chain A:



7.2 Chemical shift list 2

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_2*

7.2.1 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	28
Number of shifts mapped to atoms	0
Number of unparsed shifts	0
Number of shifts with mapping errors	28
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

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

- Chain not found in structure. First 5 (of 28) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
UNMAPPED	1	HEC_ox	3HMC	17.57	-1.0	1
UNMAPPED	1	HEC_ox	3HBC	0.21	-1.0	1
UNMAPPED	1	HEC_ox	2HAA	4.63	-1.0	2
UNMAPPED	1	HEC_ox	HHB	-2.38	-1.0	1
UNMAPPED	1	HEC_ox	2HMC	17.57	-1.0	1

7.2.2 Chemical shift referencing

No chemical shift referencing corrections were calculated (not enough data).

7.2.3 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 0%, i.e. 0 atoms were assigned a chemical shift out of a possible 1391. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	0/564 (0%)	0/224 (0%)	0/232 (0%)	0/108 (0%)
Sidechain	0/720 (0%)	0/424 (0%)	0/264 (0%)	0/32 (0%)
Aromatic	0/107 (0%)	0/58 (0%)	0/46 (0%)	0/3 (0%)
Overall	0/1391 (0%)	0/706 (0%)	0/542 (0%)	0/143 (0%)

7.2.4 Statistically unusual chemical shifts

There are no statistically unusual chemical shifts.

7.2.5 Random Coil Index (RCI) plots

No *random coil index* (RCI) plot could be generated from the current chemical shift list (assigned_chem_shift_list_2). RCI is only applicable to proteins.