



# Full wwPDB NMR Structure Validation Report ⓘ

Feb 12, 2017 – 09:56 pm GMT

PDB ID : 2FWL  
Title : The cytochrome c552/CuA complex from *Thermus thermophilus*  
Authors : Muresanu, L.; Pristovsek, P.; Loehr, F.; Maneg, O.; Mukrasch, M.D.; Rueterjans, H.; Ludwig, B.; Luecke, C.  
Deposited on : 2006-02-02

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/NMRValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

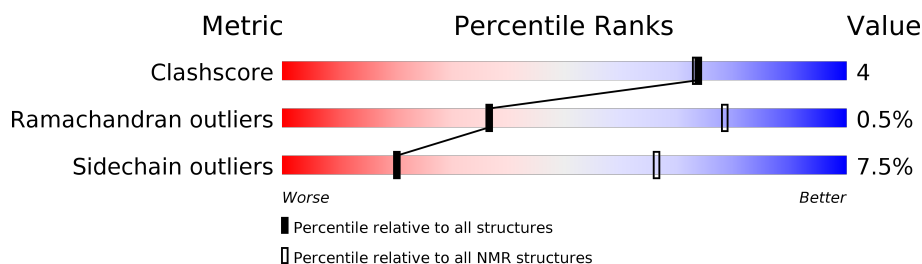
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 27%.

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	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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  $\geq 3$ , 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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	133	 79% 8% 10% •
2	B	136	 83% 15% ••

## 2 Ensemble composition and analysis

This entry contains 3 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *fewest violations*.

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:4-A:21, A:25-A:122 (116)	0.25	2
2	B:35-B:168 (134)	0.28	2

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 1 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3

### 3 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4150 atoms, of which 2079 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Cytochrome c-552.

Mol	Chain	Residues	Atoms						Trace
1	A	129	Total	C	H	N	O	S	0
			2012	632	1028	173	175	4	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-1	MET	-	CLONING ARTIFACT	UNP P04164
A	0	ALA	-	CLONING ARTIFACT	UNP P04164

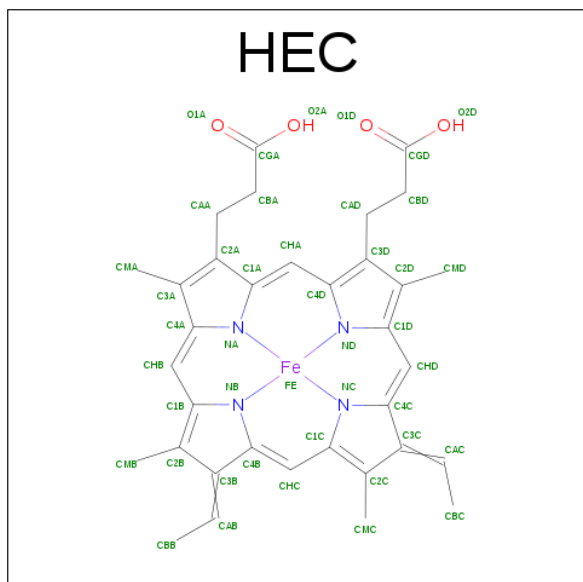
- Molecule 2 is a protein called Cytochrome c oxidase subunit II.

Mol	Chain	Residues	Atoms						Trace
2	B	134	Total	C	H	N	O	S	0
			2061	669	1019	176	194	3	

There is a discrepancy between the modelled and reference sequences:

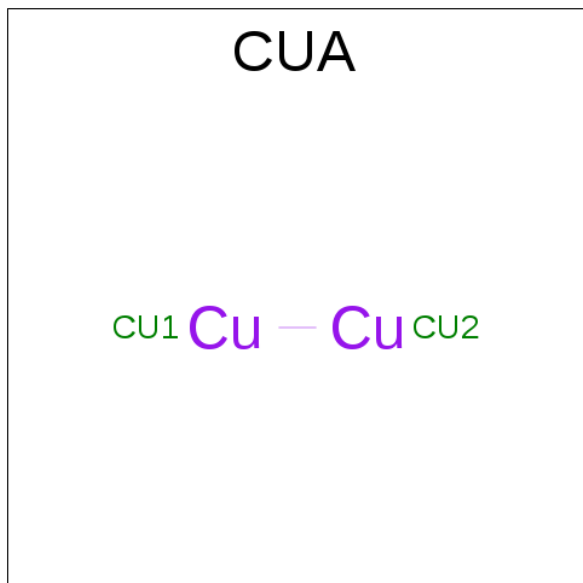
Chain	Residue	Modelled	Actual	Comment	Reference
B	33	MET	-	CLONING ARTIFACT	UNP P98052

- Molecule 3 is HEME C (three-letter code: HEC) (formula:  $C_{34}H_{34}FeN_4O_4$ ).



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

- Molecule 4 is DINUCLEAR COPPER ION (three-letter code: CUA) (formula: Cu<sub>2</sub>).



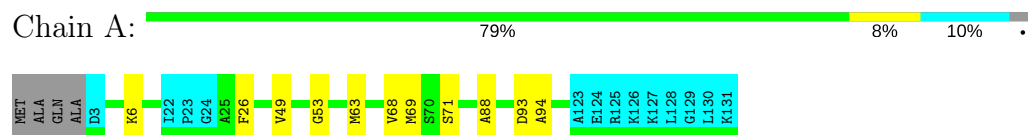
Mol	Chain	Residues	Atoms	
4	B	1	Total	Cu
			2	2

## 4 Residue-property plots

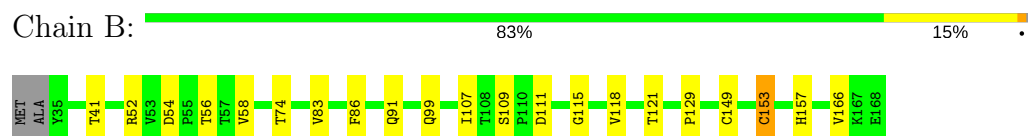
### 4.1 Average score per residue in the NMR ensemble

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-552



- Molecule 2: Cytochrome c oxidase subunit II

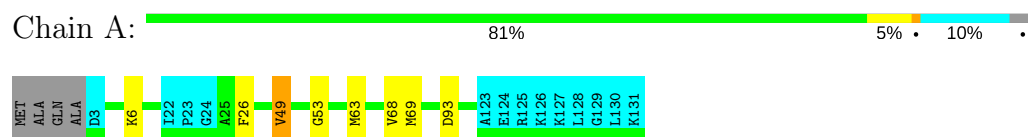


### 4.2 Scores per residue for each member of the ensemble

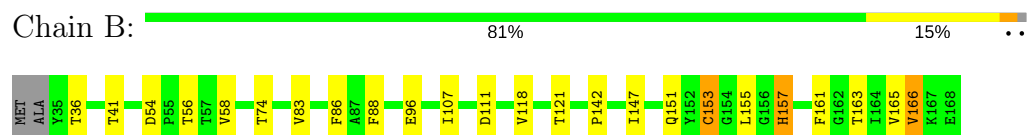
Colouring as in section 4.1 above.

#### 4.2.1 Score per residue for model 1

- Molecule 1: Cytochrome c-552

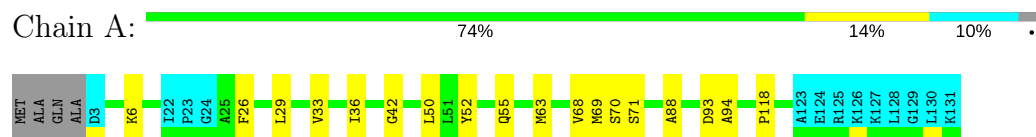


- Molecule 2: Cytochrome c oxidase subunit II

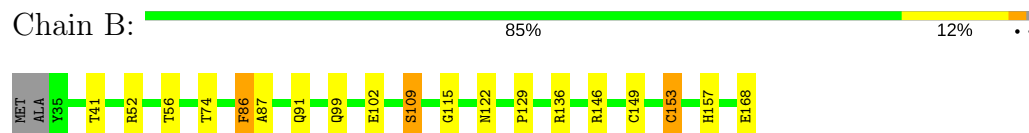


### 4.2.2 Score per residue for model 2 (medoid)

- Molecule 1: Cytochrome c-552

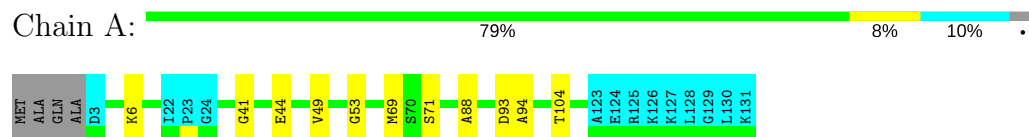


- Molecule 2: Cytochrome c oxidase subunit II

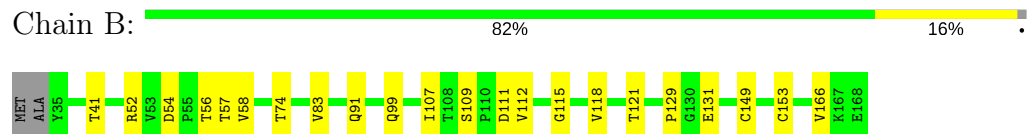


### 4.2.3 Score per residue for model 3

- Molecule 1: Cytochrome c-552



- Molecule 2: Cytochrome c oxidase subunit II



## 5 Refinement protocol and experimental data overview

The models were refined using the following method: *docking calculations*.

Of the 10 calculated structures, 3 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
HADDOCK	structure solution	1.3
HADDOCK	refinement	1.3

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)	BMRB entry 6965
Number of chemical shift lists	1
Total number of shifts	993
Number of shifts mapped to atoms	993
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	27%

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: CUA, 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	884	915	913	6±2
2	B	1042	1019	1016	7±1
3	A	43	32	30	5±2
All	All	5913	5898	5877	45

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 4.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
3:A:132:HEC:HBC3	3:A:132:HEC:HMC1	0.80	1.51	1	2
3:A:132:HEC:HBB3	3:A:132:HEC:HMB1	0.71	1.61	2	1
3:A:132:HEC:HMC1	3:A:132:HEC:HBC3	0.67	1.66	3	1
2:B:109:SER:O	2:B:129:PRO:HA	0.58	1.97	2	2
1:A:68:VAL:HB	2:B:88:PHE:CG	0.56	2.35	1	1
2:B:153:CYS:SG	2:B:157:HIS:HB2	0.55	2.41	1	2
1:A:26:PHE:CD1	3:A:132:HEC:HMD3	0.53	2.39	2	2
1:A:69:MET:HB2	3:A:132:HEC:C4D	0.51	2.36	3	2
3:A:132:HEC:HMB1	3:A:132:HEC:HBB3	0.50	1.83	3	2
2:B:96:GLU:HA	2:B:165:VAL:O	0.50	2.07	1	1

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:54:ASP:O	2:B:58:VAL:HB	0.49	2.08	3	2
2:B:142:PRO:HA	2:B:166:VAL:HG13	0.48	1.84	1	1
1:A:88:ALA:HB1	1:A:94:ALA:HB2	0.48	1.83	2	2
1:A:49:VAL:O	1:A:53:GLY:HA2	0.47	2.09	1	2
1:A:52:TYR:O	1:A:118:PRO:HG3	0.47	2.10	2	1
3:A:132:HEC:CMC	3:A:132:HEC:HBC3	0.46	2.35	1	2
2:B:118:VAL:HB	2:B:121:THR:OG1	0.45	2.11	1	2
1:A:41:GLY:HA2	1:A:44:GLU:OE1	0.45	2.12	3	1
2:B:111:ASP:CG	2:B:112:VAL:H	0.45	2.15	3	1
2:B:115:GLY:O	2:B:149:CYS:HA	0.44	2.12	3	2
1:A:68:VAL:N	3:A:132:HEC:HMD2	0.44	2.28	2	1
1:A:50:LEU:HD21	3:A:132:HEC:HMB2	0.44	1.88	2	1
2:B:102:GLU:OE2	2:B:136:ARG:HD3	0.44	2.12	2	1
1:A:29:LEU:O	1:A:33:VAL:HG23	0.43	2.12	2	1
1:A:26:PHE:HD1	3:A:132:HEC:HMD3	0.43	1.71	2	2
2:B:83:VAL:O	2:B:107:ILE:HA	0.43	2.14	1	2
2:B:147:ILE:O	2:B:161:PHE:HA	0.42	2.15	1	1
1:A:68:VAL:HB	2:B:88:PHE:CB	0.41	2.45	1	1
1:A:36:ILE:O	1:A:42:GLY:HA3	0.41	2.15	2	1
2:B:149:CYS:O	2:B:157:HIS:CE1	0.41	2.74	2	1
2:B:86:PHE:CZ	2:B:91:GLN:HB2	0.40	2.51	2	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	116/133 (87%)	109±2 (94±1%)	7±2 (6±1%)	0±0 (0±0%)	100	100
2	B	132/136 (97%)	119±1 (90±1%)	12±1 (9±1%)	1±0 (1±0%)	23	70
All	All	744/807 (92%)	683 (92%)	57 (8%)	4 (1%)	37	78

All 2 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	153	CYS	3
2	B	87	ALA	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	88/100 (88%)	83±1 (94±1%)	5±1 (6±1%)	28 74
2	B	112/113 (99%)	102±1 (91±1%)	10±1 (9±1%)	16 61
All	All	600/639 (94%)	555 (92%)	45 (8%)	20 66

All 29 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)
2	B	74	THR	3
1	A	93	ASP	3
2	B	41	THR	3
1	A	6	LYS	3
2	B	56	THR	3
1	A	71	SER	2
2	B	52	ARG	2
2	B	99	GLN	2
2	B	166	VAL	2
2	B	86	PHE	2
1	A	63	MET	2
2	B	163	THR	1
2	B	57	THR	1
2	B	111	ASP	1
1	A	49	VAL	1
2	B	151	GLN	1
1	A	70	SER	1
2	B	168	GLU	1
1	A	104	THR	1
2	B	109	SER	1
2	B	36	THR	1
2	B	122	ASN	1
1	A	69	MET	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
2	B	146	ARG	1
2	B	131	GLU	1
2	B	155	LEU	1
2	B	91	GLN	1
2	B	157	HIS	1
1	A	55	GLN	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 carbohydrates in this entry.

## 6.6 Ligand geometry ⓘ

2 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	HEC	A	132	1	28,50,50	1.56±0.01	0±0 (0±0%)
4	CUA	B	169	2	0,1,1	0.00±0.00	-

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	HEC	A	132	1	16,82,82	0.70±0.03	0±0 (0±0%)
4	CUA	B	169	2	0,0,0	0.00±0.00	-

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
3	HEC	A	132	1	-	0±0,6,54,54	0±0,0,8,8
4	CUA	B	169	2	-	0±0,0,0,0	0±0,0,0,0

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.

## 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 27% for the well-defined parts and 26% for the entire structure.

### 7.1 Chemical shift list 1

File name: BMRB entry 6965

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	993
Number of shifts mapped to atoms	993
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	13

#### 7.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	0	—	None (insufficient data)
$^{13}\text{C}_\beta$	0	—	None (insufficient data)
$^{13}\text{C}'$	0	—	None (insufficient data)
$^{15}\text{N}$	110	$-0.59 \pm 0.26$	Should be applied

#### 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 27%, i.e. 816 atoms were assigned a chemical shift out of a possible 3013. 0 out of 42 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	353/1220 (29%)	243/485 (50%)	0/500 (0%)	110/235 (47%)
Sidechain	401/1538 (26%)	390/899 (43%)	0/575 (0%)	11/64 (17%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	62/255 (24%)	61/134 (46%)	0/108 (0%)	1/13 (8%)
Overall	816/3013 (27%)	694/1518 (46%)	0/1183 (0%)	122/312 (39%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 26%, i.e. 816 atoms were assigned a chemical shift out of a possible 3182. 0 out of 44 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	353/1283 (28%)	243/510 (48%)	0/526 (0%)	110/247 (45%)
Sidechain	401/1644 (24%)	390/962 (41%)	0/612 (0%)	11/70 (16%)
Aromatic	62/255 (24%)	61/134 (46%)	0/108 (0%)	1/13 (8%)
Overall	816/3182 (26%)	694/1606 (43%)	0/1246 (0%)	122/330 (37%)

#### 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
2	B	114	HIS	HD2	31.99	9.28 – 4.78	55.5
2	B	156	GLY	HA3	0.91	5.80 – 2.00	-7.9
2	B	108	THR	HG22	-0.46	2.29 – -0.01	-7.0
2	B	108	THR	HG23	-0.46	2.29 – -0.01	-7.0
2	B	108	THR	HG21	-0.46	2.29 – -0.01	-7.0
2	B	63	PRO	HB3	-0.29	3.81 – 0.21	-6.4
2	B	155	LEU	HB2	3.59	3.32 – -0.08	5.8
2	B	155	LEU	HB3	3.59	3.34 – -0.26	5.7
2	B	114	HIS	HE2	24.47	23.17 – -3.53	5.5
2	B	112	VAL	H	11.81	11.69 – 4.89	5.2
2	B	147	ILE	HG21	-0.58	2.13 – -0.57	-5.0
2	B	147	ILE	HG23	-0.58	2.13 – -0.57	-5.0
2	B	147	ILE	HG22	-0.58	2.13 – -0.57	-5.0

#### 7.1.5 Random Coil Index (RCI) plots [i](#)

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 B:

