Full wwPDB NMR Structure Validation Report (i)

Apr 21, 2024 – 06:58 AM EDT

PDB ID	:	2LZE
BMRB ID	:	18749
Title	:	Ligase 10C
Authors	:	Chao, F.; Morelli, A.; Haugner, J.; Churchfield, L.; Hagmann, L.; Shi, L.; Masterson, L.; Sarangi, R.; Veglia, G.; Seelig, B.
Deposited on	:	2012-10-01

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 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:

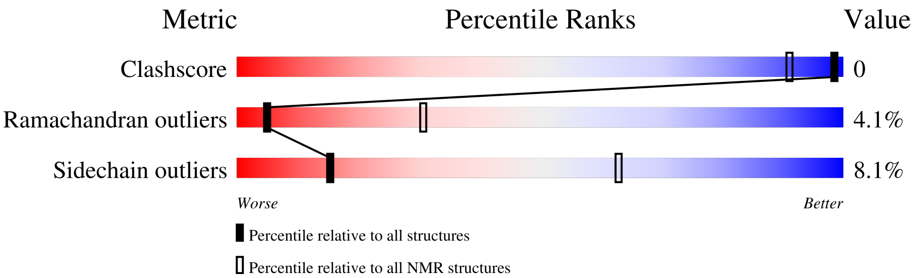
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 63%.

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	87	57% 6% 37%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 16 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:16-A:70 (55)	2.70	16

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. No single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1294 atoms, of which 611 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called a primordial catalytic fold generated by in vitro evolution.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	87	1292	415	611	123	133	10	0

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	2	2	2

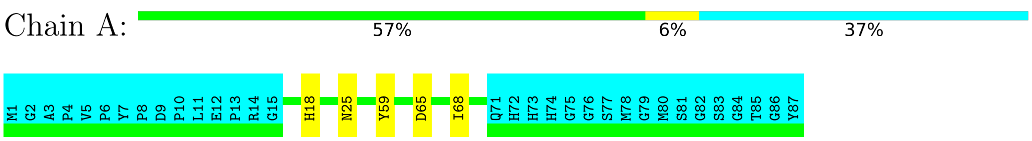
4 Residue-property plots (i)

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: a primordial catalytic fold generated by in vitro evolution

Chain A:



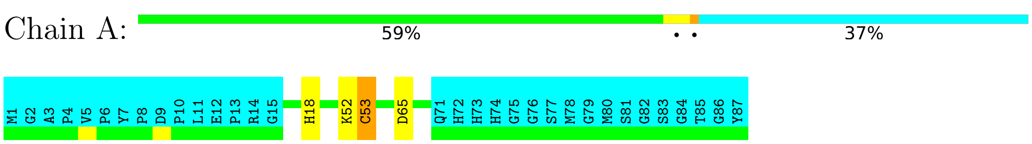
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: a primordial catalytic fold generated by in vitro evolution

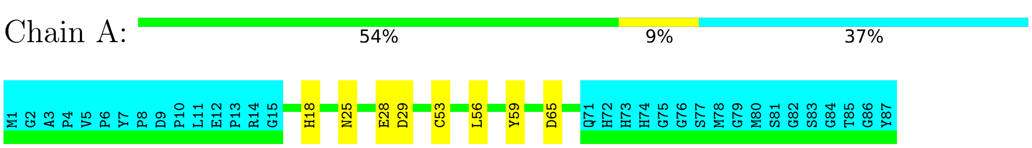
Chain A:



4.2.2 Score per residue for model 2

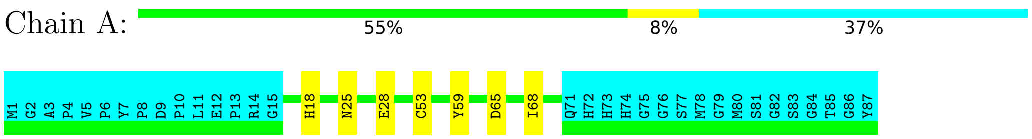
- Molecule 1: a primordial catalytic fold generated by in vitro evolution

Chain A:



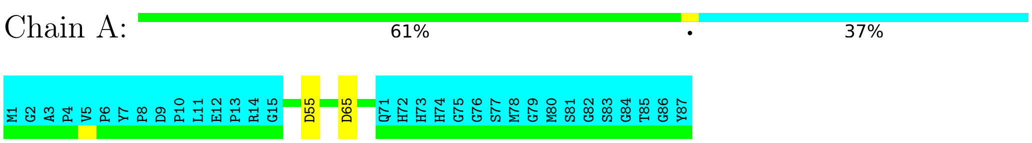
4.2.3 Score per residue for model 3

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



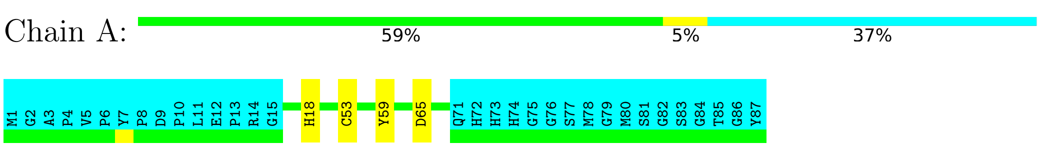
4.2.4 Score per residue for model 4

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



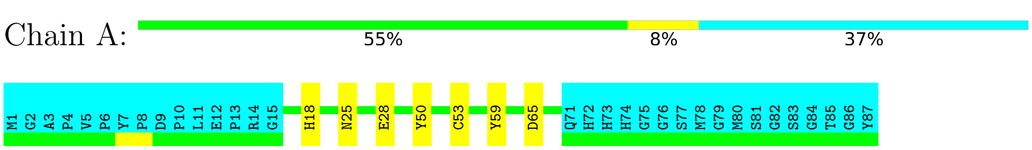
4.2.5 Score per residue for model 5

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



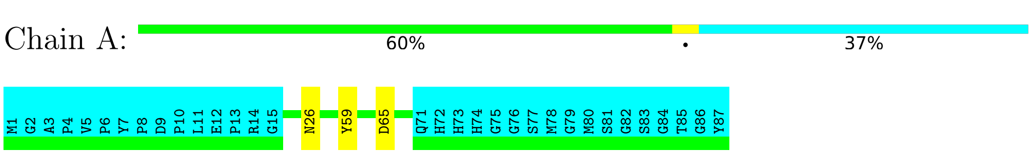
4.2.6 Score per residue for model 6

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



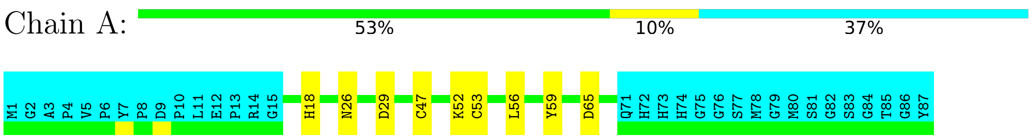
4.2.7 Score per residue for model 7

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



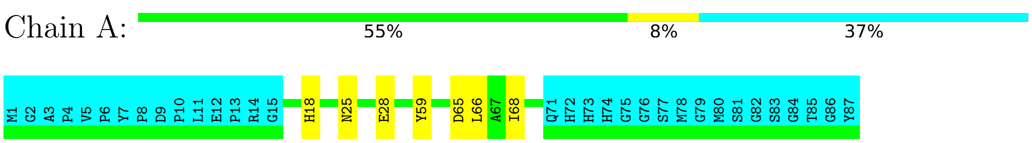
4.2.8 Score per residue for model 8

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



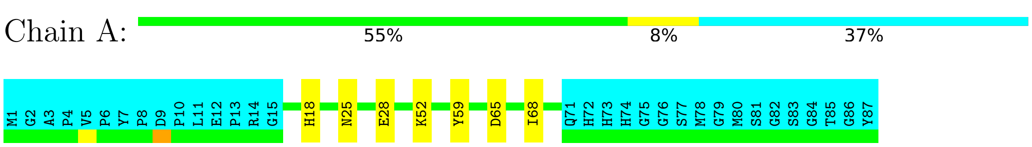
4.2.9 Score per residue for model 9

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



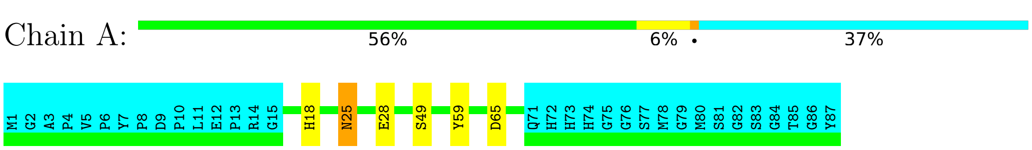
4.2.10 Score per residue for model 10

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



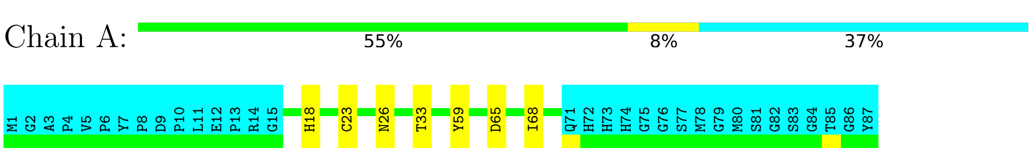
4.2.11 Score per residue for model 11

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



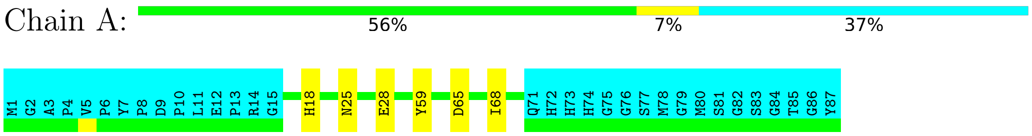
4.2.12 Score per residue for model 12

- Molecule 1: a primordial catalytic fold generated by in vitro evolution



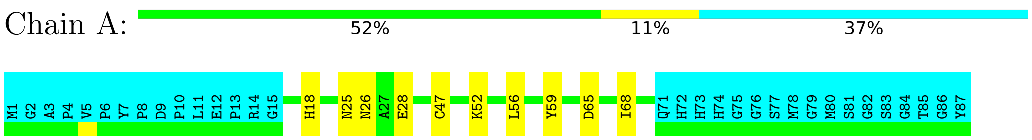
4.2.13 Score per residue for model 13

• Molecule 1: a primordial catalytic fold generated by in vitro evolution



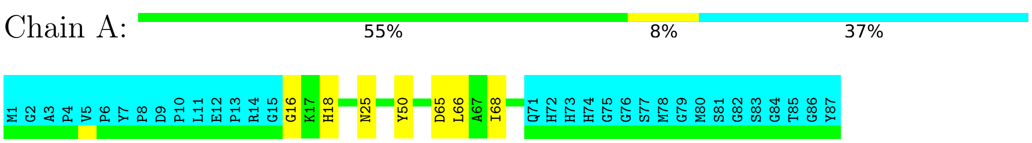
4.2.14 Score per residue for model 14

• Molecule 1: a primordial catalytic fold generated by in vitro evolution



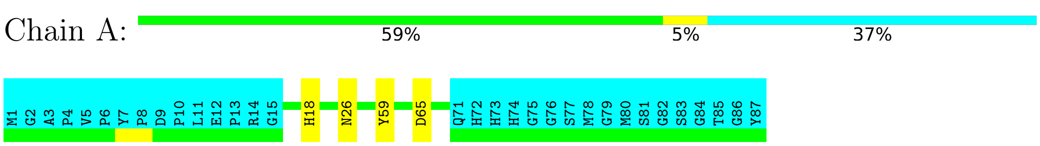
4.2.15 Score per residue for model 15

• Molecule 1: a primordial catalytic fold generated by in vitro evolution



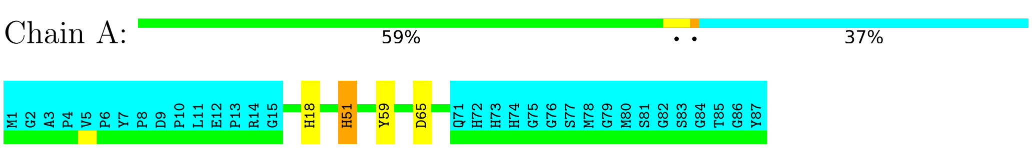
4.2.16 Score per residue for model 16 (medoid)

• Molecule 1: a primordial catalytic fold generated by in vitro evolution



4.2.17 Score per residue for model 17

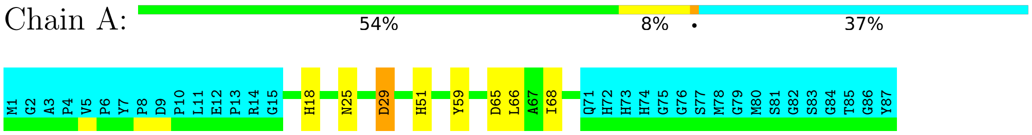
• Molecule 1: a primordial catalytic fold generated by in vitro evolution



4.2.18 Score per residue for model 18

- Molecule 1: a primordial catalytic fold generated by in vitro evolution

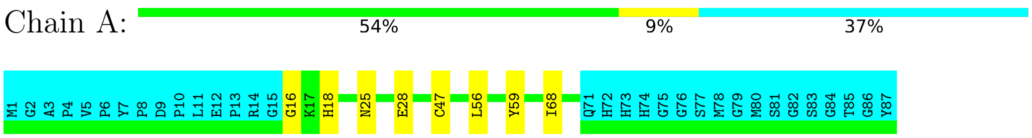
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: a primordial catalytic fold generated by in vitro evolution

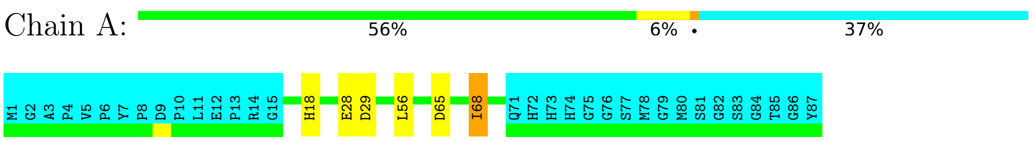
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: a primordial catalytic fold generated by in vitro evolution

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *simulated annealing*.

Of the 200 calculated structures, 20 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	608
Number of shifts mapped to atoms	608
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	63%

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: ZN

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 (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	453	406	407	0±0
All	All	9100	8120	8140	2

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 0.

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:ASN:HD21	1:A:33:THR:H	0.48	1.50	12	1
1:A:26:ASN:HD22	1:A:26:ASN:N	0.40	2.14	14	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	55/87 (63%)	45±2 (82±3%)	8±2 (14±3%)	2±2 (4±3%)	5	31
All	All	1100/1740 (63%)	905 (82%)	150 (14%)	45 (4%)	5	31

All 10 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	25	ASN	11
1	A	28	GLU	9
1	A	68	ILE	5
1	A	52	LYS	4
1	A	53	CYS	3
1	A	29	ASP	3
1	A	66	LEU	3
1	A	56	LEU	3
1	A	16	GLY	2
1	A	51	HIS	2

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	50/73 (68%)	46±1 (92±2%)	4±1 (8±2%)	15	63
All	All	1000/1460 (68%)	919 (92%)	81 (8%)	15	63

All 16 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	65	ASP	19
1	A	18	HIS	18
1	A	59	TYR	16
1	A	68	ILE	6
1	A	53	CYS	4
1	A	26	ASN	3
1	A	47	CYS	3
1	A	29	ASP	2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	56	LEU	2
1	A	50	TYR	2
1	A	55	ASP	1
1	A	25	ASN	1
1	A	28	GLU	1
1	A	49	SER	1
1	A	23	CYS	1
1	A	51	HIS	1

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)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 63% 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_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	608
Number of shifts mapped to atoms	608
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
$^{13}C_{\alpha}$	66	-0.18 ± 0.25	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	61	0.15 ± 0.20	None needed (< 0.5 ppm)
$^{13}C'$	59	0.03 ± 0.14	None needed (< 0.5 ppm)
^{15}N	61	-0.61 ± 0.48	None needed (imprecise)

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 63%, i.e. 453 atoms were assigned a chemical shift out of a possible 724. 0 out of 3 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	240/277 (87%)	90/112 (80%)	99/110 (90%)	51/55 (93%)
Sidechain	213/359 (59%)	139/228 (61%)	74/115 (64%)	0/16 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/88 (0%)	0/44 (0%)	0/40 (0%)	0/4 (0%)
Overall	453/724 (63%)	229/384 (60%)	173/265 (65%)	51/75 (68%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 56%, i.e. 608 atoms were assigned a chemical shift out of a possible 1092. 0 out of 5 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	304/435 (70%)	118/179 (66%)	125/174 (72%)	61/82 (74%)
Sidechain	304/530 (57%)	202/341 (59%)	102/169 (60%)	0/20 (0%)
Aromatic	0/127 (0%)	0/64 (0%)	0/56 (0%)	0/7 (0%)
Overall	608/1092 (56%)	320/584 (55%)	227/399 (57%)	61/109 (56%)

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 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: