Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2022 – 03:18 PM EDT

PDB ID	:	7MLA
Title	:	Solution NMR structure of HDMX in complex with Zn and MCo-52-2
Authors	:	Ramirez, L.S.; Theophall, G.; Chaudhuri, D.; Camarero, J.C.; Shekhtman, A.
Deposited on	:	2021-04-28

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

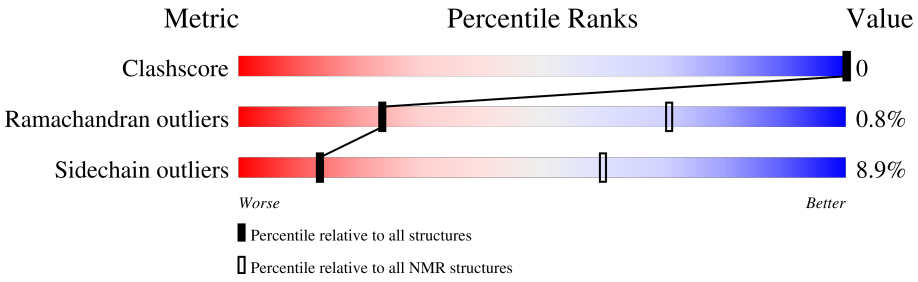
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.28.1
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.28.1

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 81%.

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	70	71% 7% 17% .
2	B	34	65% 35%

2 Ensemble composition and analysis

This entry contains 20 models. Model 19 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 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:434-A:488, B:8-B:29 (77)	0.42	19

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 4 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 1524 atoms, of which 757 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protein Mdm4.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	67	1060	325	540	99	88	8	0

There are 7 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	421	GLY	-	expression tag	UNP O15151
A	422	SER	-	expression tag	UNP O15151
A	423	HIS	-	expression tag	UNP O15151
A	424	MET	-	expression tag	UNP O15151
A	425	TYR	-	expression tag	UNP O15151
A	426	SER	-	expression tag	UNP O15151
A	427	GLY	-	expression tag	UNP O15151

- Molecule 2 is a protein called MCo-52-2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	34	462	144	217	46	49	6	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
3	A	2	2	2

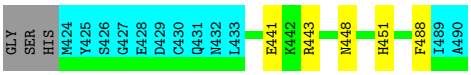
4 Residue-property plots

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.

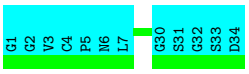
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

Chain B:



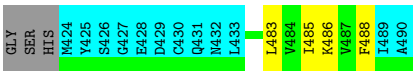
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

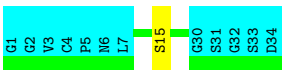
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

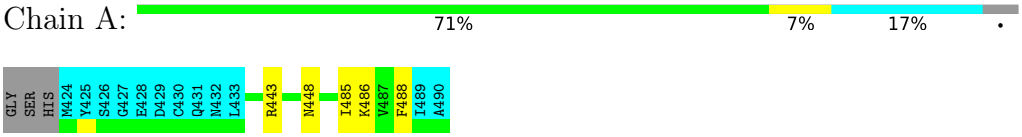
Chain B:



4.2.2 Score per residue for model 2

• Molecule 1: Protein Mdm4

Chain A:



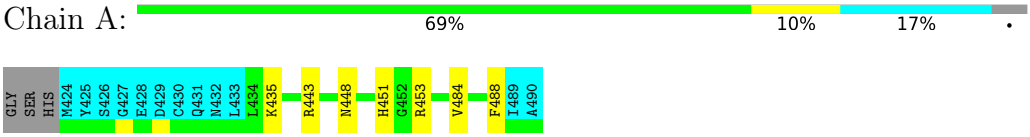
• Molecule 2: MCo-52-2

Chain B:

4.2.3 Score per residue for model 3

• Molecule 1: Protein Mdm4

Chain A:



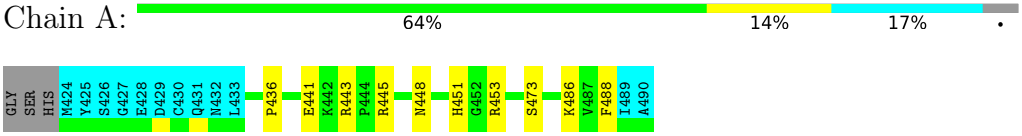
• Molecule 2: MCo-52-2

Chain B:

4.2.4 Score per residue for model 4

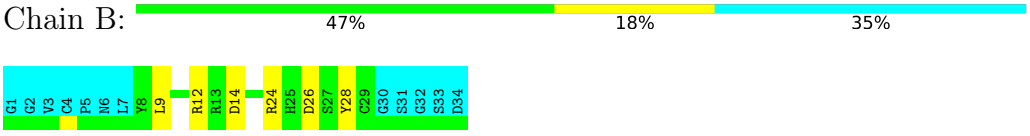
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

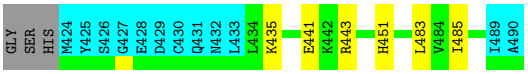
Chain B:



4.2.5 Score per residue for model 5

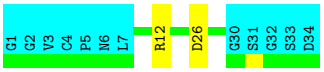
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

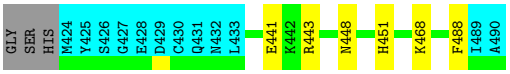
Chain B:



4.2.6 Score per residue for model 6

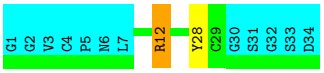
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

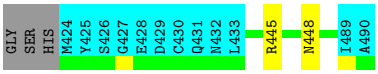
Chain B:



4.2.7 Score per residue for model 7

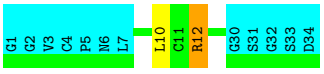
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

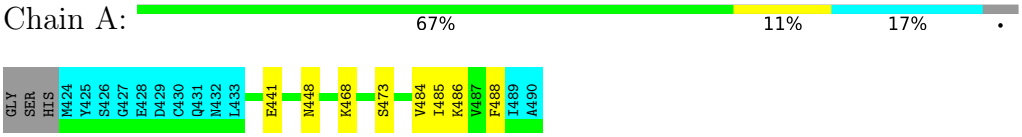
Chain B:



4.2.8 Score per residue for model 8

• Molecule 1: Protein Mdm4

Chain A:



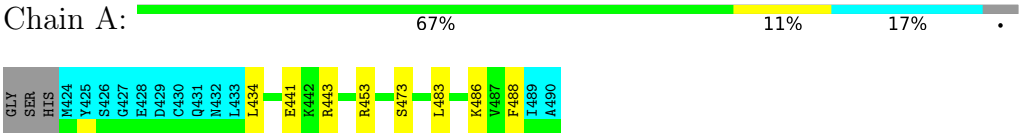
• Molecule 2: MCo-52-2

Chain B:

4.2.9 Score per residue for model 9

• Molecule 1: Protein Mdm4

Chain A:



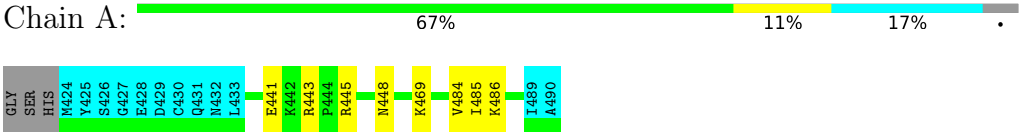
• Molecule 2: MCo-52-2

Chain B:

4.2.10 Score per residue for model 10

• Molecule 1: Protein Mdm4

Chain A:



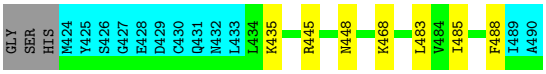
• Molecule 2: MCo-52-2

Chain B:

4.2.11 Score per residue for model 11

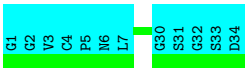
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

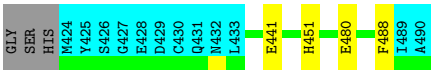
Chain B:



4.2.12 Score per residue for model 12

• Molecule 1: Protein Mdm4

Chain A:



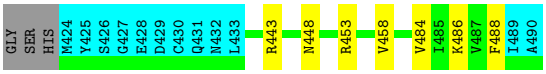
• Molecule 2: MCo-52-2

Chain B:

4.2.13 Score per residue for model 13

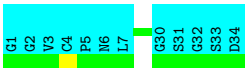
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

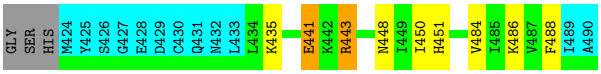
Chain B:



4.2.14 Score per residue for model 14

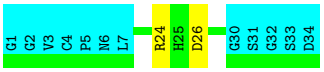
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

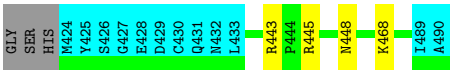
Chain B:



4.2.15 Score per residue for model 15

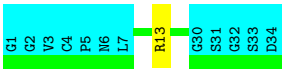
• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

Chain B:



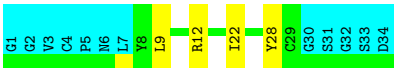
4.2.16 Score per residue for model 16

• Molecule 1: Protein Mdm4

Chain A:

• Molecule 2: MCo-52-2

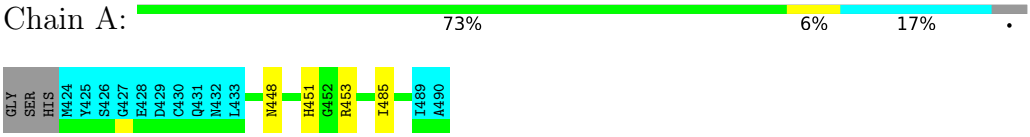
Chain B:



4.2.17 Score per residue for model 17

• Molecule 1: Protein Mdm4

Chain A:



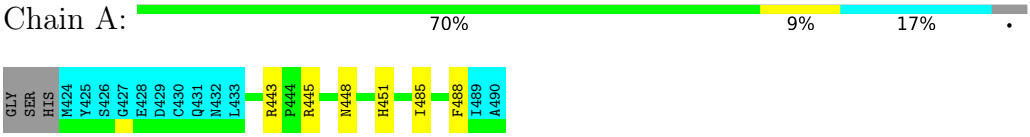
• Molecule 2: MCo-52-2

Chain B:

4.2.18 Score per residue for model 18

• Molecule 1: Protein Mdm4

Chain A:



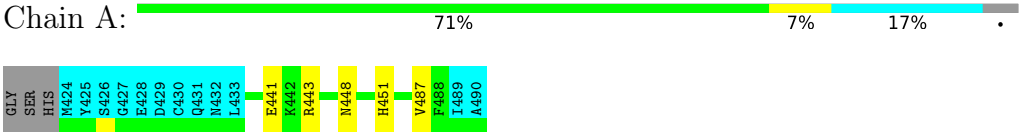
• Molecule 2: MCo-52-2

Chain B:

4.2.19 Score per residue for model 19 (medoid)

• Molecule 1: Protein Mdm4

Chain A:



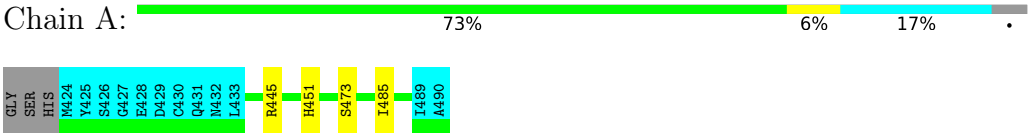
• Molecule 2: MCo-52-2

Chain B:

4.2.20 Score per residue for model 20

• Molecule 1: Protein Mdm4

Chain A:



• Molecule 2: MCo-52-2

Chain B:

5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 20 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
CYANA	structure calculation	3.98.13
YASARA	refinement	20.4.24

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	2
Total number of shifts	1100
Number of shifts mapped to atoms	1100
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	81%

6 Model quality

6.1 Standard geometry

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

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	0.64±0.01	0±0/436 (0.0± 0.0%)	0.89±0.02	1±1/583 (0.1± 0.1%)
2	B	0.66±0.02	0±0/177 (0.0± 0.0%)	0.92±0.05	0±0/239 (0.1± 0.2%)
All	All	0.65	0/12260 (0.0%)	0.90	22/16440 (0.1%)

There are no bond-length outliers.

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
1	A	443	ARG	NE-CZ-NH1	5.77	123.18	120.30	14	3
2	B	12	ARG	NE-CZ-NH1	5.75	123.17	120.30	6	4
1	A	453	ARG	NE-CZ-NH1	5.70	123.15	120.30	9	5
1	A	445	ARG	NE-CZ-NH1	5.68	123.14	120.30	18	7
2	B	24	ARG	NE-CZ-NH1	5.57	123.08	120.30	12	2
2	B	13	ARG	NE-CZ-NH1	5.04	122.82	120.30	15	1

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	428	458	458	0±0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes
All	All	12060	12240	12240	1

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:448:ASN:HD21	1:A:450:ILE:HD11	0.43	1.74	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	55/70 (79%)	50±1 (91±1%)	4±1 (8±2%)	1±1 (1±1%)	18 66
2	B	22/34 (65%)	21±1 (97±3%)	1±1 (3±3%)	0±0 (0±0%)	100 100
All	All	1540/2080 (74%)	1431 (93%)	97 (6%)	12 (1%)	24 71

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)
1	A	441	GLU	10
1	A	436	PRO	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	48/60 (80%)	43±1 (90±3%)	5±1 (10±3%)	12	57
2	B	20/28 (71%)	19±2 (93±8%)	1±2 (7±8%)	19	67
All	All	1360/1760 (77%)	1239 (91%)	121 (9%)	13	60

All 26 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	448	ASN	14
1	A	488	PHE	12
1	A	443	ARG	11
1	A	451	HIS	11
1	A	485	ILE	10
1	A	486	LYS	8
2	B	12	ARG	6
1	A	484	VAL	5
2	B	9	LEU	5
2	B	26	ASP	5
1	A	468	LYS	5
1	A	483	LEU	4
1	A	435	LYS	4
1	A	473	SER	4
2	B	28	TYR	3
2	B	15	SER	2
2	B	14	ASP	2
2	B	24	ARG	2
2	B	10	LEU	1
2	B	23	CYS	1
1	A	469	LYS	1
1	A	480	GLU	1
1	A	458	VAL	1
1	A	441	GLU	1
2	B	22	ILE	1
1	A	487	VAL	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 monosaccharides in this entry.

6.6 Ligand geometry

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

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 81% for the well-defined parts and 82% 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

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	769
Number of shifts mapped to atoms	769
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	67	-0.28 ± 0.27	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	62	-0.05 ± 0.23	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	64	0.10 ± 0.43	None needed (< 0.5 ppm)

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 59%, i.e. 576 atoms were assigned a chemical shift out of a possible 977. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	214/377 (57%)	107/150 (71%)	55/154 (36%)	52/73 (71%)
Sidechain	335/538 (62%)	208/322 (65%)	127/183 (69%)	0/33 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	27/62 (44%)	14/34 (41%)	13/24 (54%)	0/4 (0%)
Overall	576/977 (59%)	329/506 (65%)	195/361 (54%)	52/110 (47%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 57%, i.e. 685 atoms were assigned a chemical shift out of a possible 1212. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	262/495 (53%)	131/197 (66%)	67/202 (33%)	64/96 (67%)
Sidechain	388/647 (60%)	241/386 (62%)	147/225 (65%)	0/36 (0%)
Aromatic	35/70 (50%)	18/38 (47%)	17/28 (61%)	0/4 (0%)
Overall	685/1212 (57%)	390/621 (63%)	231/455 (51%)	64/136 (47%)

7.1.4 Statistically unusual chemical shifts (i)

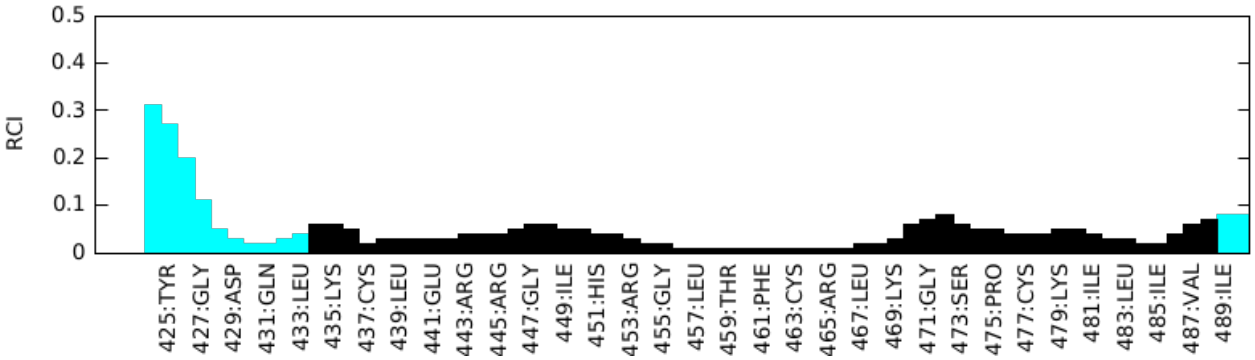
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	486	LYS	HB3	0.00	3.10 – 0.40	-6.5
1	A	476	ILE	CG2	8.78	24.63 – 10.43	-6.2

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 A:



7.2 Chemical shift list 2

File name: working_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	331
Number of shifts mapped to atoms	331
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.2.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	34	-0.05 ± 0.50	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	29	0.03 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	32	1.19 ± 1.03	None needed (imprecise)

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 22%, i.e. 216 atoms were assigned a chemical shift out of a possible 977. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned

stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	86/377 (23%)	43/150 (29%)	22/154 (14%)	21/73 (29%)
Sidechain	110/538 (20%)	69/322 (21%)	38/183 (21%)	3/33 (9%)
Aromatic	20/62 (32%)	10/34 (29%)	10/24 (42%)	0/4 (0%)
Overall	216/977 (22%)	122/506 (24%)	70/361 (19%)	24/110 (22%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 25%, i.e. 303 atoms were assigned a chemical shift out of a possible 1212. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	131/495 (26%)	65/197 (33%)	34/202 (17%)	32/96 (33%)
Sidechain	152/647 (23%)	95/386 (25%)	53/225 (24%)	4/36 (11%)
Aromatic	20/70 (29%)	10/38 (26%)	10/28 (36%)	0/4 (0%)
Overall	303/1212 (25%)	170/621 (27%)	97/455 (21%)	36/136 (26%)

7.2.4 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	24	ARG	NE	72.16	92.63 – 76.73	-7.9

7.2.5 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:


0.5
0.4
0.3
0.2
0.1
0
RCI