Full wwPDB NMR Structure Validation Report (i)

May 28, 2020 – 10:55 pm BST

PDB ID	:	2L31
Title	:	Human PARP-1 zinc finger 2
Authors	:	Neuhaus, D.; Eustermann, S.; Yang, J.; Videler, H.
Deposited on	:	2010-08-30

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 (1)) were used in the production of this report:

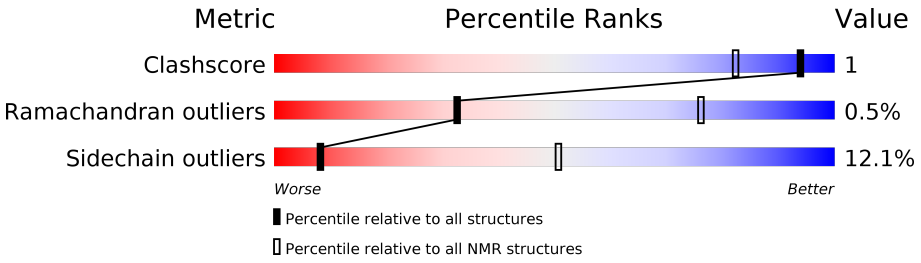
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance

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

The overall completeness of chemical shifts assignment is 80%.

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	112	58% 13% 29%

2 Ensemble composition and analysis

This entry contains 30 models. Model 5 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:109-A:142, A:155-A:200 (80)	0.21	5

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 6 clusters and 9 single-model clusters were found.

Cluster number	Models
1	2, 5, 8, 13, 14
2	1, 4, 15, 17
3	6, 12, 23, 25
4	10, 19, 24
5	20, 22, 28
6	3, 11
Single-model clusters	7; 9; 16; 18; 21; 26; 27; 29; 30

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1771 atoms, of which 890 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Poly [ADP-ribose] polymerase 1.

Mol	Chain	Residues	Atoms						Trace
1	A	112	Total	C	H	N	O	S	0
			1770	549	890	156	169	6	

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

Mol	Chain	Residues	Atoms	
2	A	1	Total	Zn
			1	1

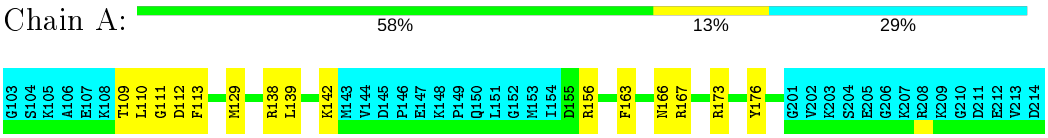
4 Residue-property plots

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: Poly [ADP-ribose] polymerase 1

Chain A:



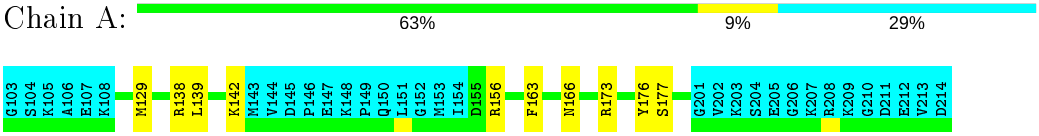
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Poly [ADP-ribose] polymerase 1

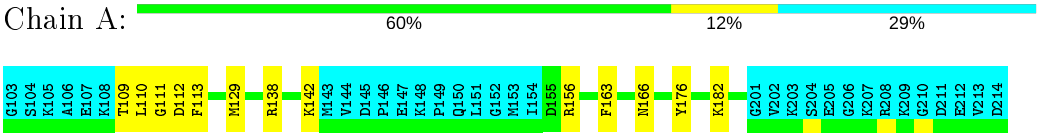
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Poly [ADP-ribose] polymerase 1

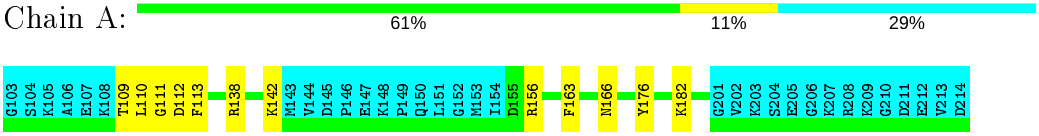
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Poly [ADP-ribose] polymerase 1

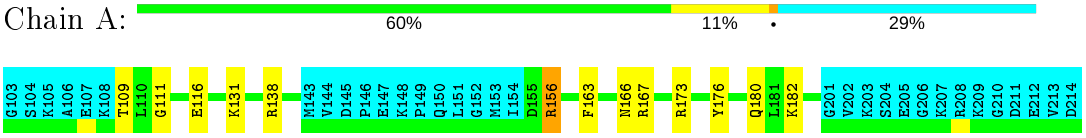
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Poly [ADP-ribose] polymerase 1

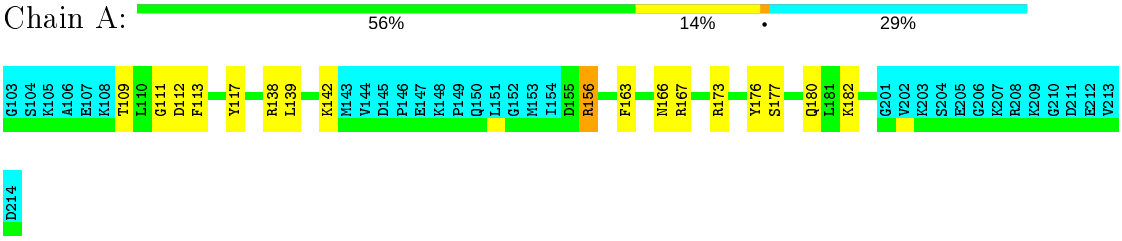
Chain A:



4.2.5 Score per residue for model 5 (medoid)

• Molecule 1: Poly [ADP-ribose] polymerase 1

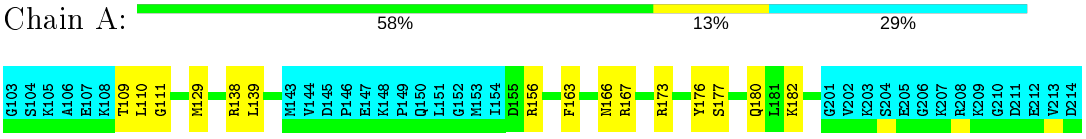
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Poly [ADP-ribose] polymerase 1

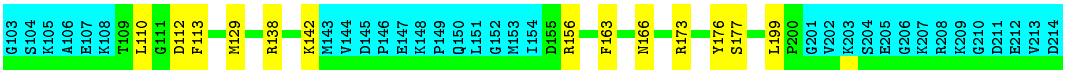
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Poly [ADP-ribose] polymerase 1

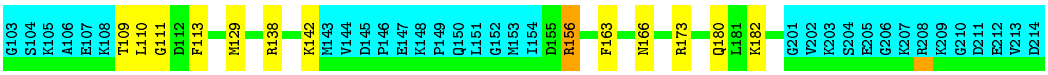
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Poly [ADP-ribose] polymerase 1

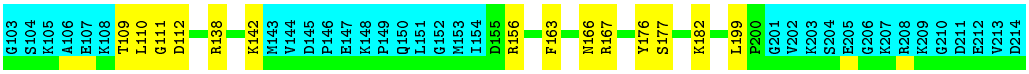
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Poly [ADP-ribose] polymerase 1

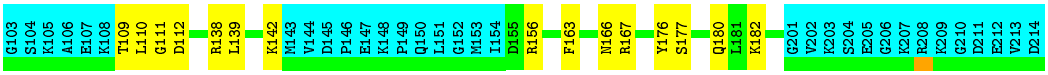
Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: Poly [ADP-ribose] polymerase 1

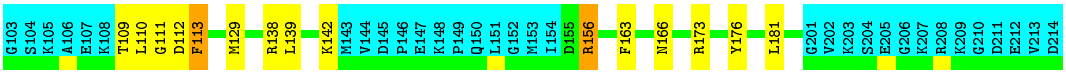
Chain A:



4.2.11 Score per residue for model 11

- Molecule 1: Poly [ADP-ribose] polymerase 1

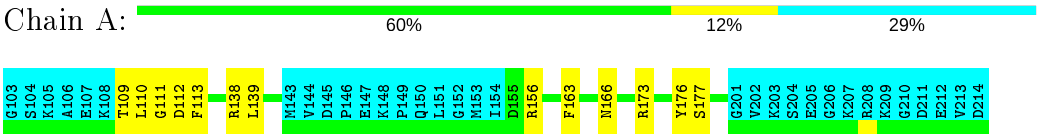
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Poly [ADP-ribose] polymerase 1

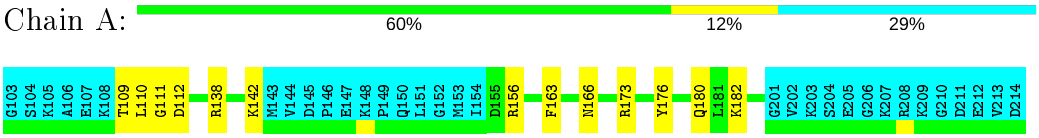
Chain A:



4.2.13 Score per residue for model 13

- Molecule 1: Poly [ADP-ribose] polymerase 1

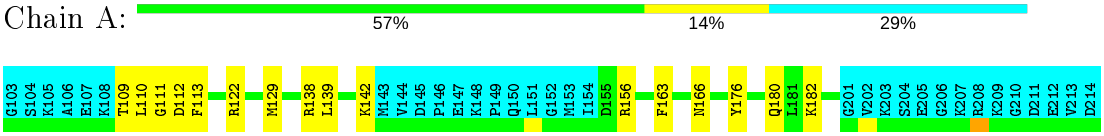
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Poly [ADP-ribose] polymerase 1

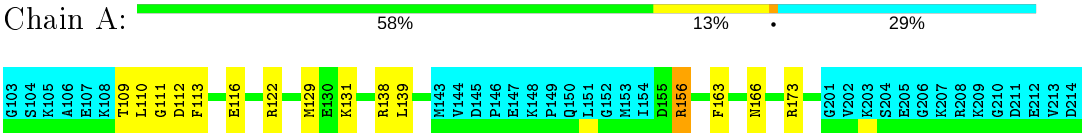
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Poly [ADP-ribose] polymerase 1

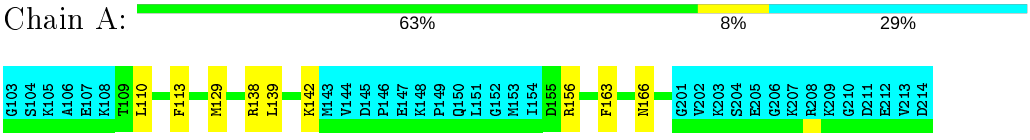
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Poly [ADP-ribose] polymerase 1

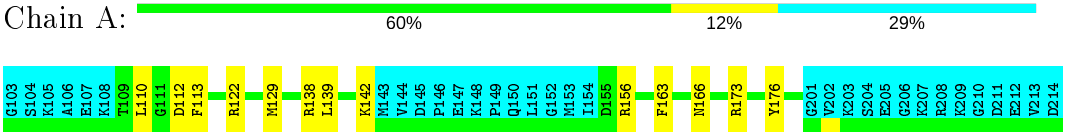
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Poly [ADP-ribose] polymerase 1

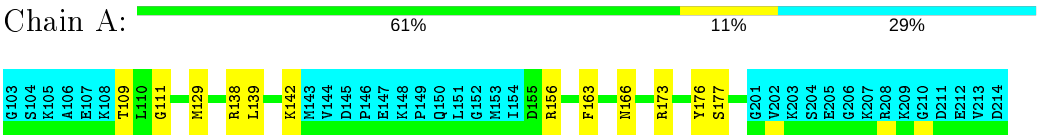
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Poly [ADP-ribose] polymerase 1

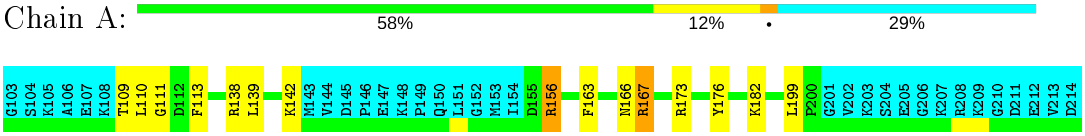
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Poly [ADP-ribose] polymerase 1

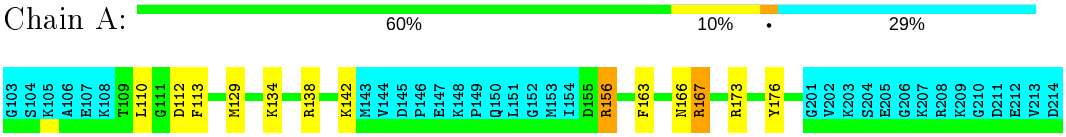
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Poly [ADP-ribose] polymerase 1

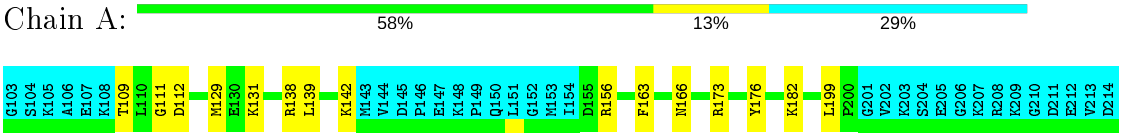
Chain A:



4.2.21 Score per residue for model 21

• Molecule 1: Poly [ADP-ribose] polymerase 1

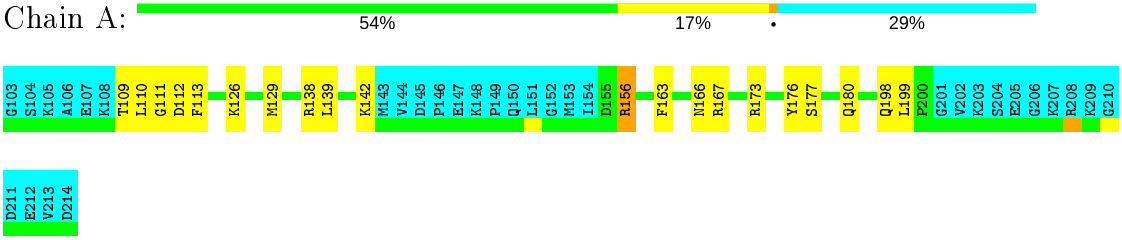
Chain A:



4.2.22 Score per residue for model 22

• Molecule 1: Poly [ADP-ribose] polymerase 1

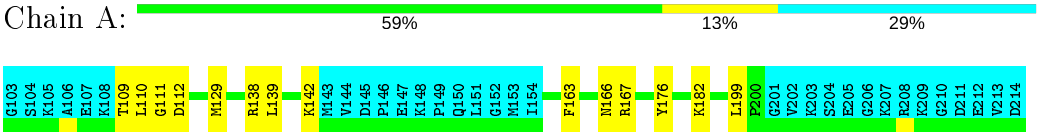
Chain A:



4.2.23 Score per residue for model 23

• Molecule 1: Poly [ADP-ribose] polymerase 1

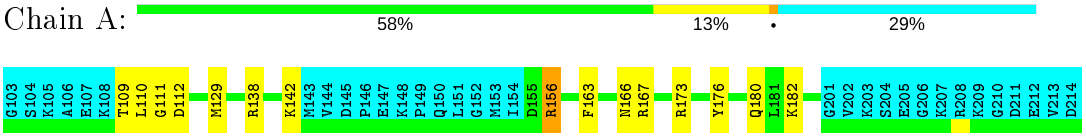
Chain A:



4.2.24 Score per residue for model 24

• Molecule 1: Poly [ADP-ribose] polymerase 1

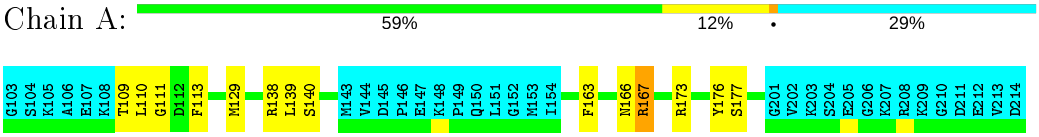
Chain A:



4.2.25 Score per residue for model 25

• Molecule 1: Poly [ADP-ribose] polymerase 1

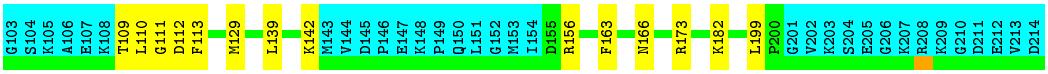
Chain A:



4.2.26 Score per residue for model 26

• Molecule 1: Poly [ADP-ribose] polymerase 1

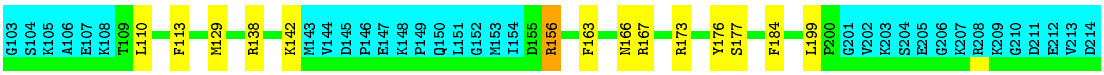
Chain A:



4.2.27 Score per residue for model 27

- Molecule 1: Poly [ADP-ribose] polymerase 1

Chain A:



4.2.28 Score per residue for model 28

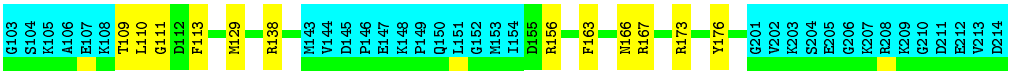
- Molecule 1: Poly [ADP-ribose] polymerase 1

Chain A:

4.2.29 Score per residue for model 29

- Molecule 1: Poly [ADP-ribose] polymerase 1

Chain A:



4.2.30 Score per residue for model 30

- Molecule 1: Poly [ADP-ribose] polymerase 1

Chain A:

5 Refinement protocol and experimental data overview

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

Of the 50 calculated structures, 30 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
ATNOSCANDID	structure solution	
X-PLOR NIH	structure solution	
Amber	refinement	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1169
Number of shifts mapped to atoms	1169
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	80%

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

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	641	642	642	2±1
All	All	19260	19260	19260	48

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:109:THR:HG22	1:A:111:GLY:H	0.51	1.65	8	23
1:A:109:THR:HG23	1:A:180:GLN:O	0.48	2.08	8	9
1:A:109:THR:HG21	1:A:182:LYS:CG	0.46	2.40	8	15
1:A:113:PHE:CD1	1:A:181:LEU:HD23	0.45	2.46	11	1

5.2 Torsion angles (i)

5.2.1 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	80/112 (71%)	75±1 (94±2%)	5±1 (6±2%)	0±0 (0±1%)	32	76
All	All	2400/3360 (71%)	2250 (94%)	139 (6%)	11 (0%)	32	76

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	177	SER	11

5.2.2 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	68/94 (72%)	60±2 (88±3%)	8±2 (12±3%)	8	51
All	All	2040/2820 (72%)	1794 (88%)	246 (12%)	8	51

All 17 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	163	PHE	30
1	A	166	ASN	30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	176	TYR	26
1	A	110	LEU	24
1	A	142	LYS	22
1	A	129	MET	21
1	A	113	PHE	19
1	A	112	ASP	18
1	A	139	LEU	18
1	A	156	ARG	17
1	A	199	LEU	9
1	A	167	ARG	5
1	A	131	LYS	3
1	A	184	PHE	1
1	A	134	LYS	1
1	A	126	LYS	1
1	A	198	GLN	1

5.2.3 RNA

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

5.4 Carbohydrates

There are no carbohydrates in this entry.

5.5 Ligand geometry

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

5.6 Other polymers

There are no such molecules in this entry.

5.7 Polymer linkage issues

There are no chain breaks in this entry.

6 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 80% for the well-defined parts and 78% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chem_shift_list_1*

6.1.1 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	1169
Number of shifts mapped to atoms	1169
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	8

6.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	111	-0.41 ± 0.08	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	101	-0.12 ± 0.11	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	102	0.43 ± 0.29	None needed (< 0.5 ppm)

6.1.3 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 80%, i.e. 822 atoms were assigned a chemical shift out of a possible 1024. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	309/394 (78%)	154/157 (98%)	80/160 (50%)	75/77 (97%)
Sidechain	440/551 (80%)	276/329 (84%)	158/191 (83%)	6/31 (19%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	73/79 (92%)	37/42 (88%)	35/35 (100%)	1/2 (50%)
Overall	822/1024 (80%)	467/528 (88%)	273/386 (71%)	82/110 (75%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 78%, i.e. 1091 atoms were assigned a chemical shift out of a possible 1404. 0 out of 14 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	425/550 (77%)	212/219 (97%)	111/224 (50%)	102/107 (95%)
Sidechain	593/775 (77%)	370/462 (80%)	216/272 (79%)	7/41 (17%)
Aromatic	73/79 (92%)	37/42 (88%)	35/35 (100%)	1/2 (50%)
Overall	1091/1404 (78%)	619/723 (86%)	362/531 (68%)	110/150 (73%)

6.1.4 Statistically unusual chemical shifts (i)

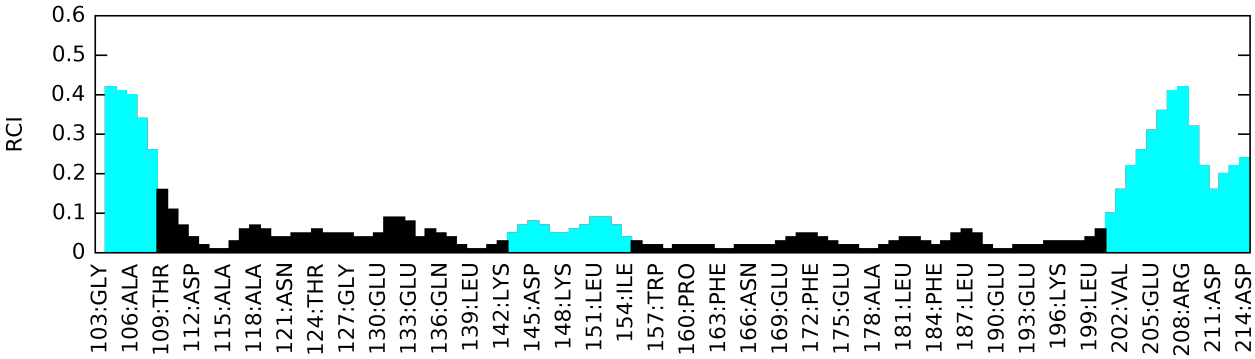
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	167	ARG	NE	119.85	92.63 – 76.73	22.1
1	A	173	ARG	HB3	-0.15	3.17 – 0.37	-6.9
1	A	155	ASP	HB3	0.94	4.07 – 1.27	-6.2
1	A	140	SER	HB3	2.22	5.25 – 2.45	-5.8
1	A	138	ARG	HD3	1.79	4.36 – 1.86	-5.3
1	A	192	LYS	HG3	-0.10	2.76 – -0.04	-5.2
1	A	140	SER	HB2	2.52	5.18 – 2.58	-5.2
1	A	179	SER	HA	2.34	6.53 – 2.43	-5.2

6.1.5 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:


RCI
0.6
0.5
0.4
0.3
0.2
0.1
0
103:GLY
106:ALA
109:THR
112:ASP
115:ALA
118:ALA
121:ASN
124:THR
127:GLY
130:GLU
133:GLU
136:GLN
139:LEU
142:LYS
145:ASP
148:LYS
151:LEU
154:ILE
157:TRP
160:PRO
163:PHE
166:ASN
169:GLU
172:PHE
175:GLU
178:ALA
181:LEU
184:PHE
187:LEU
190:GLU
193:GLU
196:LYS
199:LEU
202:VAL
205:GLU
208:ARG
211:ASP
214:ASP