Full wwPDB NMR Structure Validation Report (i)

Feb 12, 2017 – 11:38 pm GMT

PDB ID	:	2L30
Title	:	Human PARP-1 zinc finger 1
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
http://wwpdb.org/validation/2016/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:

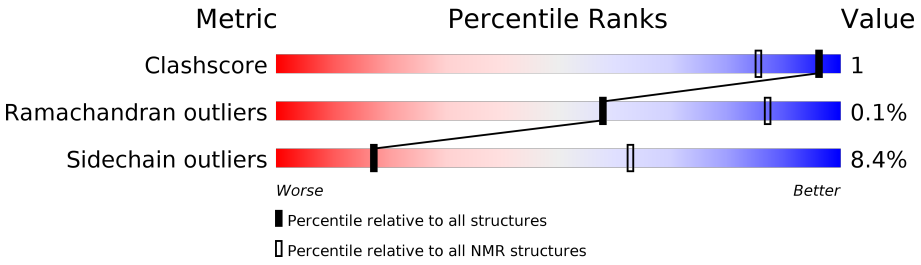
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (proteins)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

1 Overall quality at a glance (i)

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

The overall completeness of chemical shifts assignment is 84%.

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 >=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	108	71% 7% • 19%

2 Ensemble composition and analysis

This entry contains 30 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: *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:7-A:93 (87)	0.31	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 and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1677 atoms, of which 827 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	108	1676	531	827	151	160	7	0

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

Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	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

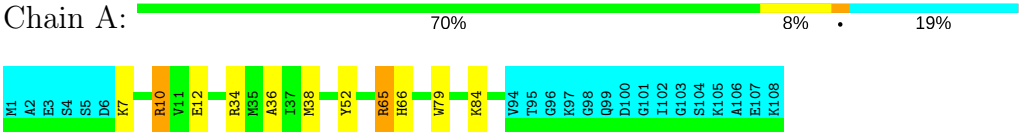
• Molecule 1: Poly [ADP-ribose] polymerase 1

Chain A:

4.2.3 Score per residue for model 3

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

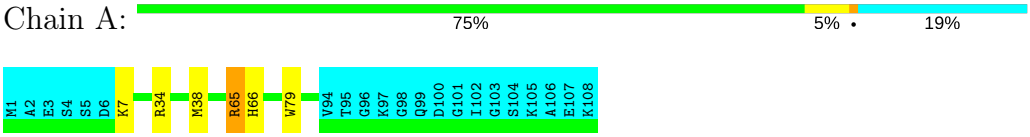
Chain A:



4.2.4 Score per residue for model 4

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

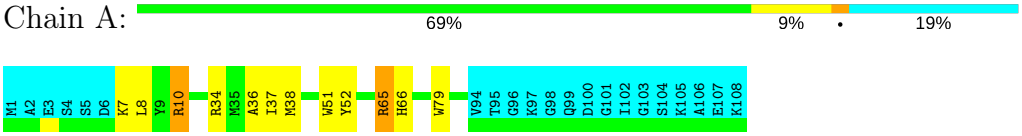
Chain A:



4.2.5 Score per residue for model 5

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

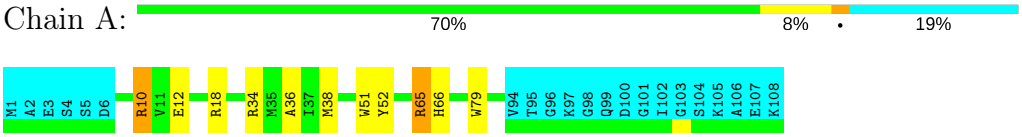
Chain A:



4.2.6 Score per residue for model 6

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

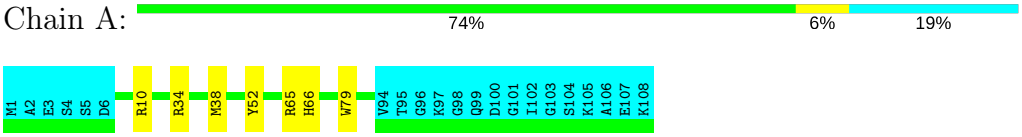
Chain A:



4.2.7 Score per residue for model 7

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

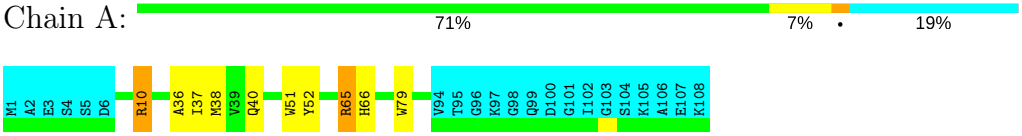
Chain A:



4.2.8 Score per residue for model 8

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

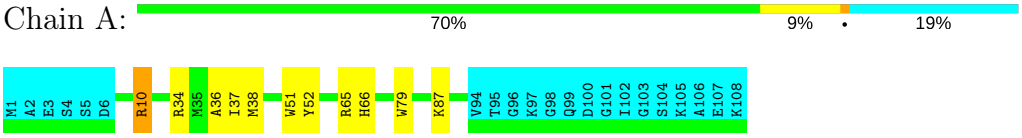
Chain A:



4.2.9 Score per residue for model 9

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

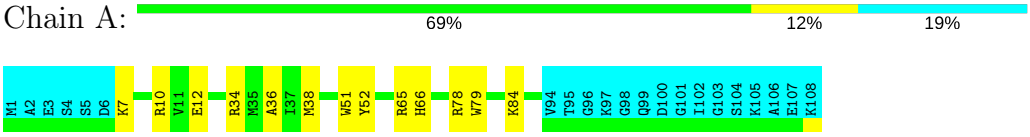
Chain A:



4.2.10 Score per residue for model 10

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

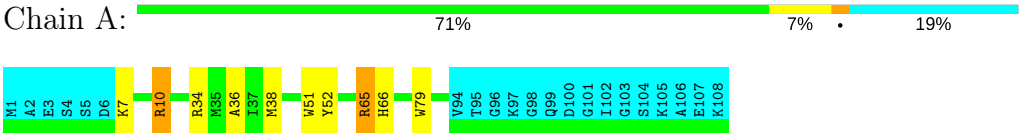
Chain A:



4.2.11 Score per residue for model 11

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

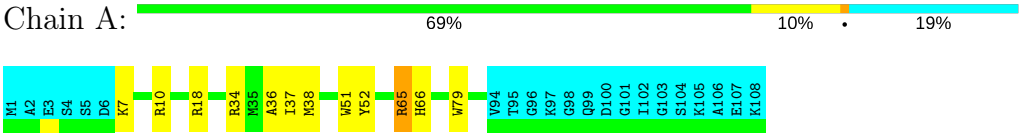
Chain A:



4.2.12 Score per residue for model 12

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

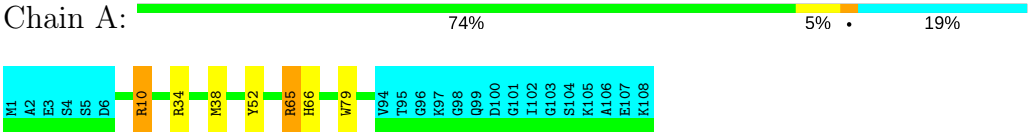
Chain A:



4.2.13 Score per residue for model 13

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

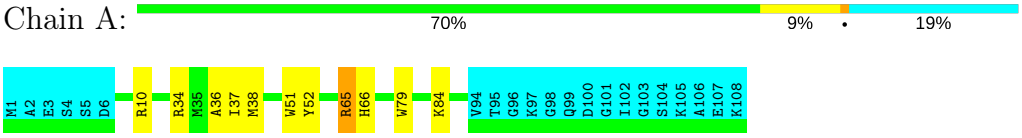
Chain A:



4.2.14 Score per residue for model 14

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

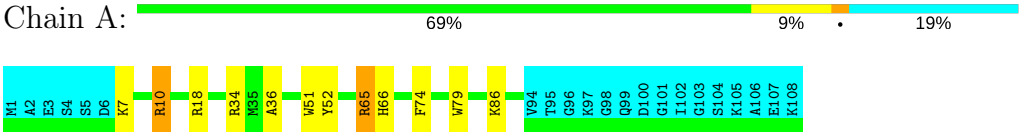
Chain A:



4.2.15 Score per residue for model 15

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

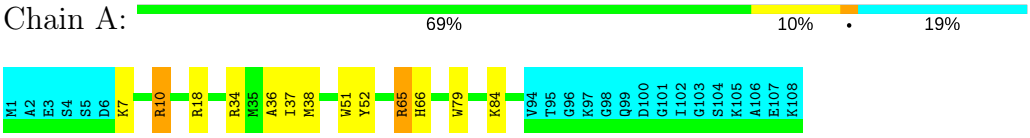
Chain A:



4.2.16 Score per residue for model 16 (medoid)

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

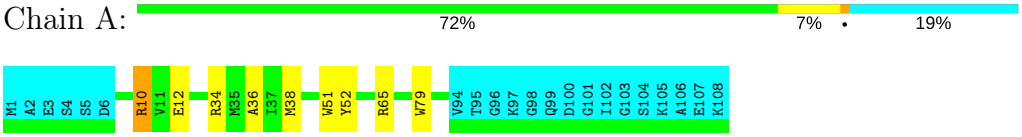
Chain A:



4.2.17 Score per residue for model 17

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

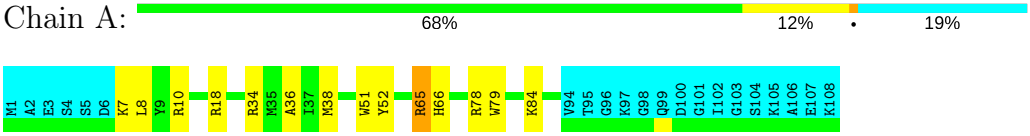
Chain A:



4.2.18 Score per residue for model 18

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

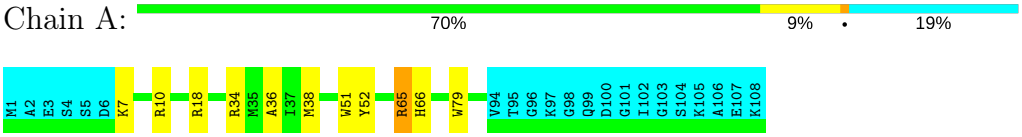
Chain A:



4.2.19 Score per residue for model 19

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

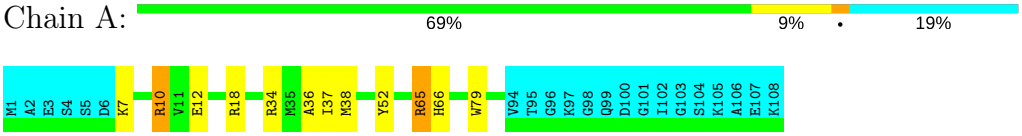
Chain A:



4.2.20 Score per residue for model 20

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

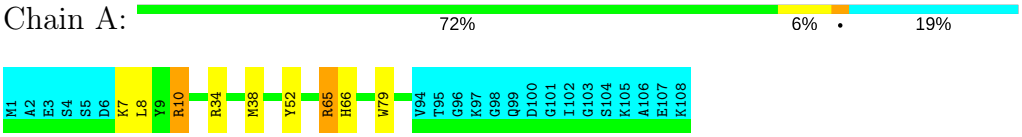
Chain A:



4.2.21 Score per residue for model 21

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

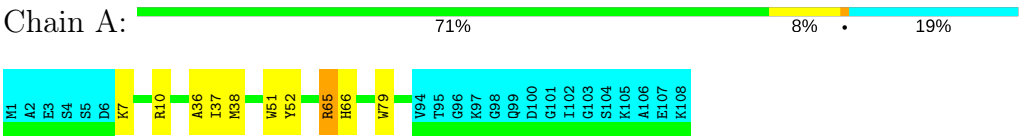
Chain A:



4.2.22 Score per residue for model 22

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

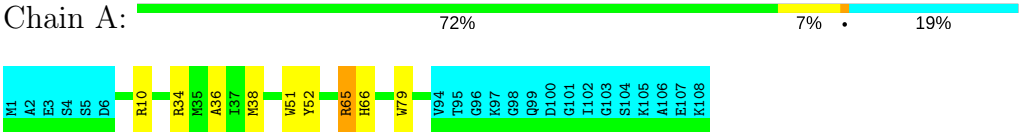
Chain A:



4.2.23 Score per residue for model 23

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

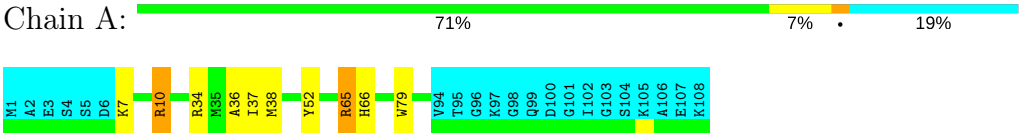
Chain A:



4.2.24 Score per residue for model 24

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

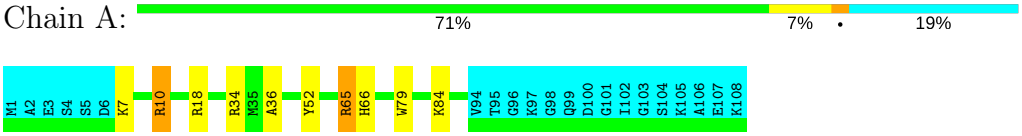
Chain A:



4.2.25 Score per residue for model 25

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

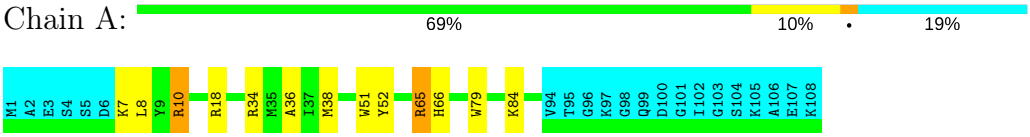
Chain A:



4.2.26 Score per residue for model 26

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

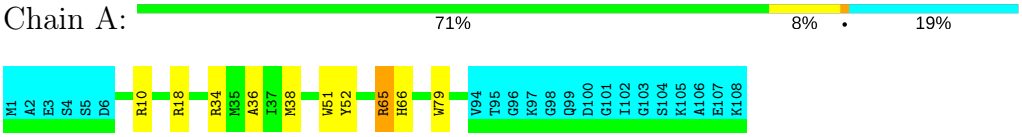
Chain A:



4.2.27 Score per residue for model 27

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

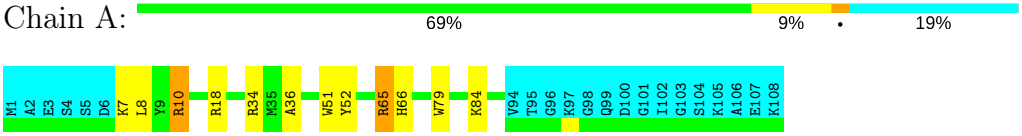
Chain A:



4.2.28 Score per residue for model 28

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

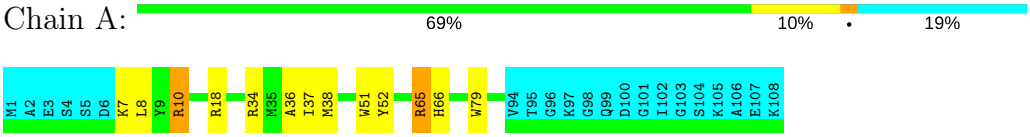
Chain A:



4.2.29 Score per residue for model 29

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

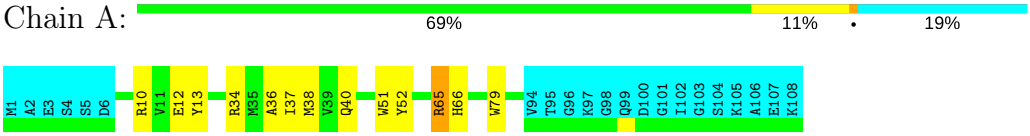
Chain A:



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 7 of this report.

Chemical shift file(s)	BMRB entry 17157
Number of chemical shift lists	1
Total number of shifts	1168
Number of shifts mapped to atoms	1168
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	84%

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: 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	0.73±0.00	0±0/725 (0.0±0.0%)	1.08±0.02	3±1/974 (0.4±0.1%)
All	All	0.73	0/21750 (0.0%)	1.08	103/29220 (0.4%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.3±0.5
All	All	0	9

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	10	ARG	NE-CZ-NH1	9.66	125.13	120.30	8	28
1	A	65	ARG	NE-CZ-NH1	8.24	124.42	120.30	1	26
1	A	18	ARG	NE-CZ-NH1	7.87	124.23	120.30	12	12
1	A	34	ARG	NE-CZ-NH1	7.75	124.18	120.30	19	27
1	A	78	ARG	NE-CZ-NH1	6.21	123.41	120.30	10	3
1	A	13	TYR	CB-CG-CD2	-5.60	117.64	121.00	30	1
1	A	10	ARG	CD-NE-CZ	5.41	131.17	123.60	30	1
1	A	65	ARG	NE-CZ-NH2	-5.40	117.60	120.30	14	2
1	A	10	ARG	NH1-CZ-NH2	-5.37	113.49	119.40	3	2
1	A	78	ARG	NE-CZ-NH2	-5.35	117.63	120.30	2	1

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	A	12	GLU	Peptide	6
1	A	10	ARG	Sidechain	3

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	704	685	685	1±1
All	All	21150	20550	20550	40

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:36:ALA:HB2	1:A:51:TRP:CE3	0.55	2.36	28	21
1:A:8:LEU:H	1:A:8:LEU:HD23	0.52	1.64	5	3
1:A:10:ARG:HB3	1:A:36:ALA:HB3	0.48	1.86	9	15
1:A:74:PHE:CZ	1:A:86:LYS:HE3	0.41	2.50	15	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	87/108 (81%)	85±1 (98±1%)	2±1 (2±1%)	0±0 (0±0%)	58	87
All	All	2610/3240 (81%)	2563 (98%)	45 (2%)	2 (0%)	58	87

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	7	LYS	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	76/91 (84%)	70±1 (92±1%)	6±1 (8±1%)	17	63
All	All	2280/2730 (84%)	2089 (92%)	191 (8%)	17	63

All 13 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	79	TRP	30
1	A	65	ARG	30
1	A	66	HIS	29
1	A	52	TYR	29
1	A	38	MET	27
1	A	7	LYS	18
1	A	37	ILE	12
1	A	84	LYS	8
1	A	8	LEU	3
1	A	40	GLN	2
1	A	10	ARG	1
1	A	18	ARG	1
1	A	87	LYS	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 carbohydrates in this entry.

6.6 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is 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

The completeness of assignment taking into account all chemical shift lists is 84% for the well-defined parts and 81% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 17157

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	1168
Number of shifts mapped to atoms	1168
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	9

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	105	0.21 ± 0.17	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	95	0.24 ± 0.10	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	99	0.23 ± 0.32	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 84%, i.e. 927 atoms were assigned a chemical shift out of a possible 1108. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	338/427 (79%)	169/170 (99%)	87/174 (50%)	82/83 (99%)
Sidechain	477/557 (86%)	302/334 (90%)	175/195 (90%)	0/28 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	112/124 (90%)	58/66 (88%)	51/51 (100%)	3/7 (43%)
Overall	927/1108 (84%)	529/570 (93%)	313/420 (75%)	85/118 (72%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 81%, i.e. 1069 atoms were assigned a chemical shift out of a possible 1324. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	408/532 (77%)	204/212 (96%)	105/216 (49%)	99/104 (95%)
Sidechain	549/668 (82%)	348/399 (87%)	201/237 (85%)	0/32 (0%)
Aromatic	112/124 (90%)	58/66 (88%)	51/51 (100%)	3/7 (43%)
Overall	1069/1324 (81%)	610/677 (90%)	357/504 (71%)	102/143 (71%)

7.1.4 Statistically unusual chemical shifts (i)

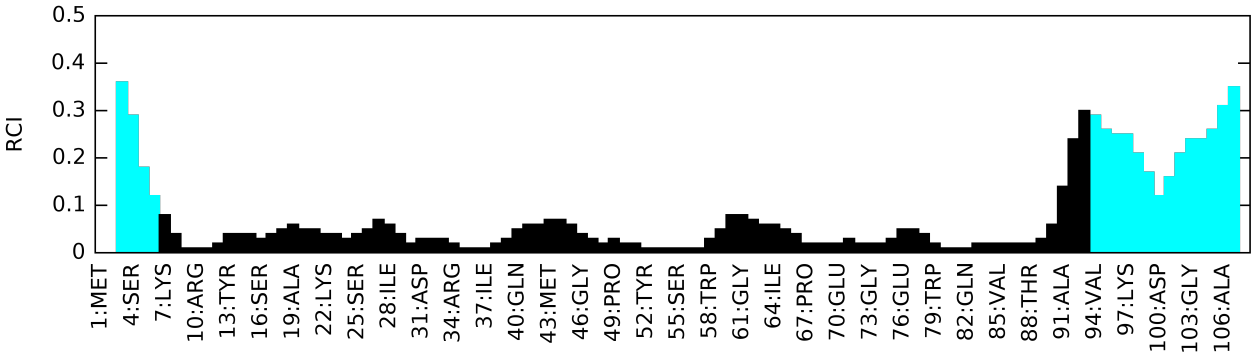
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	58	TRP	HE1	6.88	12.85 – 7.35	-5.8
1	A	57	PHE	HD1	5.41	8.56 – 5.56	-5.5
1	A	57	PHE	HD2	5.41	8.56 – 5.56	-5.5
1	A	36	ALA	HB1	-0.01	2.61 – 0.11	-5.5
1	A	36	ALA	HB2	-0.01	2.61 – 0.11	-5.5
1	A	36	ALA	HB3	-0.01	2.61 – 0.11	-5.5
1	A	88	THR	HG21	-0.04	2.29 – -0.01	-5.1
1	A	88	THR	HG23	-0.04	2.29 – -0.01	-5.1
1	A	88	THR	HG22	-0.04	2.29 – -0.01	-5.1

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:


RCI
0.5
0.4
0.3
0.2
0.1
0
1:MET
4:SER
7:LYS
10:ARG
13:TYR
16:SER
19:ALA
22:LYS
25:SER
28:ILE
31:ASP
34:ARG
37:ILE
40:GLN
43:MET
46:GLY
49:PRO
52:TYR
55:SER
58:TRP
61:GLY
64:ILE
67:PRO
70:GLU
73:GLY
76:GLU
79:TRP
82:GLN
85:VAL
88:THR
91:ALA
94:VAL
97:LYS
100:ASP
103:GLY
106:ALA