Full wwPDB NMR Structure Validation Report (i)

Feb 12, 2017 – 11:23 pm GMT

PDB ID	:	2KVH
Title	:	Structure of the three-Cys2His2 domain of mouse testis zinc finger protein
Authors	:	Chou, C.-C.; Lou, Y.-C.; Chen, C.
Deposited on	:	2010-03-15

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

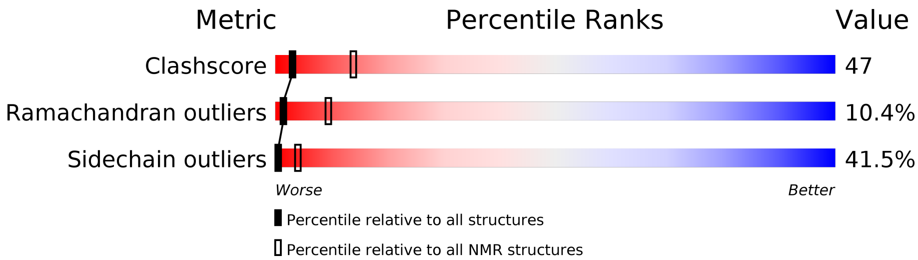
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (proteins)	:	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 83%.

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	27	

2 Ensemble composition and analysis (i)

This entry contains 15 models. Model 11 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:32-A:54 (23)	0.22	11

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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	2, 3, 4, 5, 6, 10, 13, 14
2	1, 7, 8, 9, 11, 12, 15

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 430 atoms, of which 212 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Zinc finger and BTB domain-containing protein 32.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	27	429	132	212	43	39	3	0

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

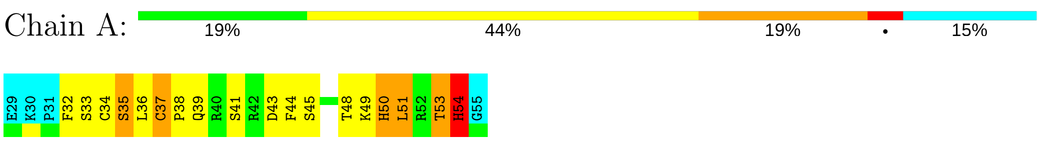
Mol	Chain	Residues	Atoms	
			Total	Zn
2	A	1	1	1

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: Zinc finger and BTB domain-containing protein 32

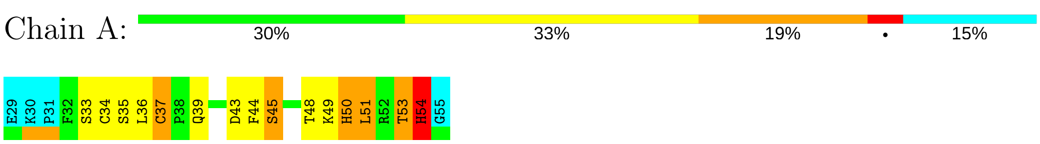


4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

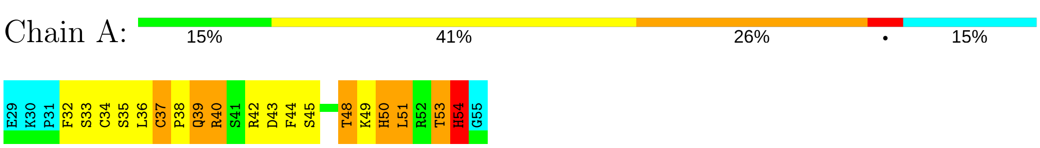
4.2.1 Score per residue for model 1

- Molecule 1: Zinc finger and BTB domain-containing protein 32



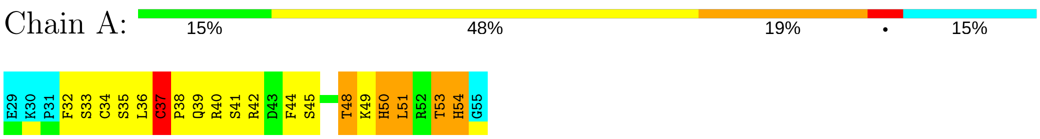
4.2.2 Score per residue for model 2

- Molecule 1: Zinc finger and BTB domain-containing protein 32



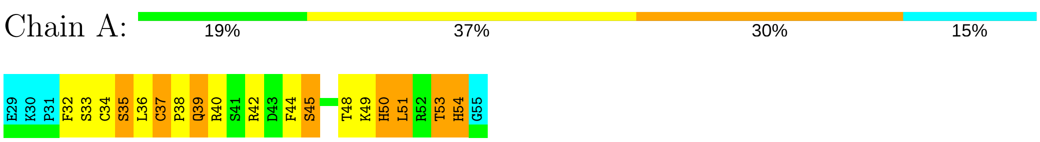
4.2.3 Score per residue for model 3

- Molecule 1: Zinc finger and BTB domain-containing protein 32



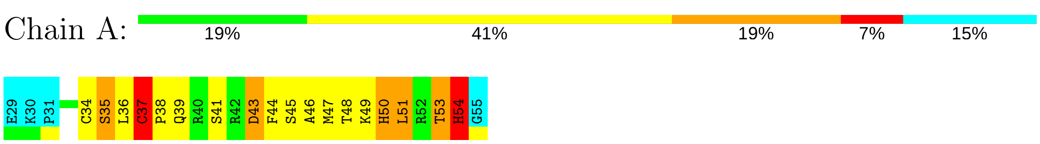
4.2.4 Score per residue for model 4

- Molecule 1: Zinc finger and BTB domain-containing protein 32



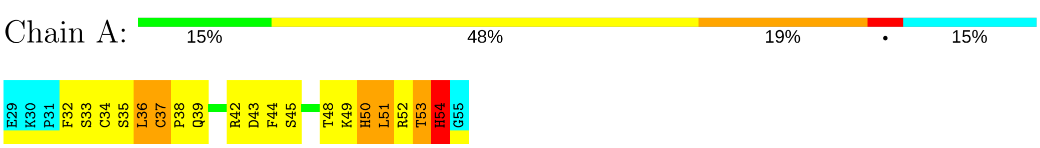
4.2.5 Score per residue for model 5

- Molecule 1: Zinc finger and BTB domain-containing protein 32



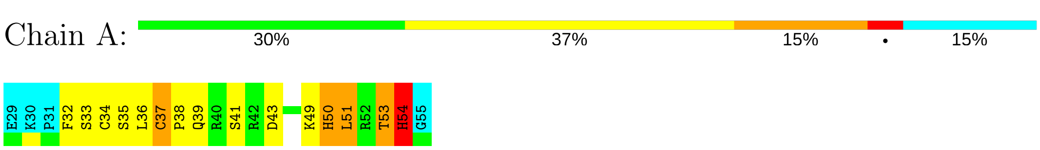
4.2.6 Score per residue for model 6

- Molecule 1: Zinc finger and BTB domain-containing protein 32



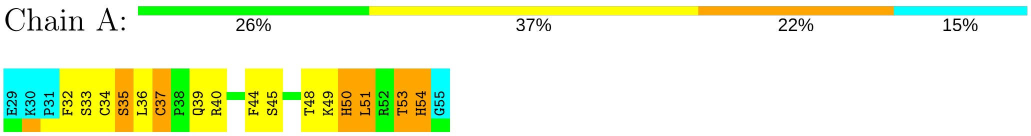
4.2.7 Score per residue for model 7

- Molecule 1: Zinc finger and BTB domain-containing protein 32



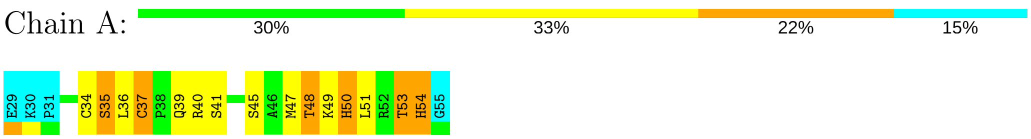
4.2.8 Score per residue for model 8

• Molecule 1: Zinc finger and BTB domain-containing protein 32



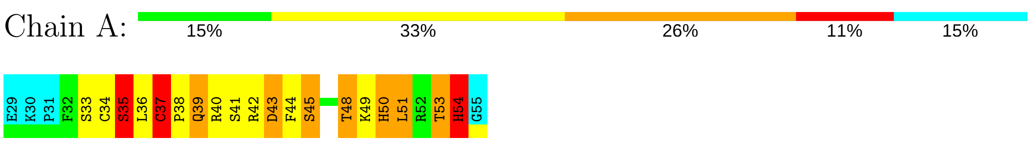
4.2.9 Score per residue for model 9

• Molecule 1: Zinc finger and BTB domain-containing protein 32



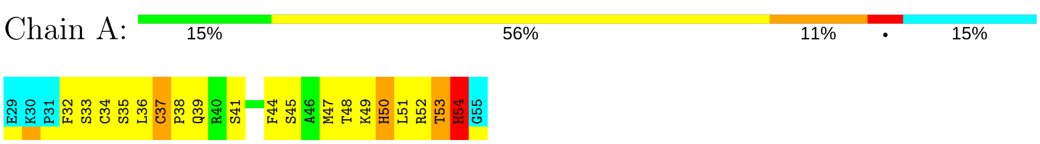
4.2.10 Score per residue for model 10

• Molecule 1: Zinc finger and BTB domain-containing protein 32



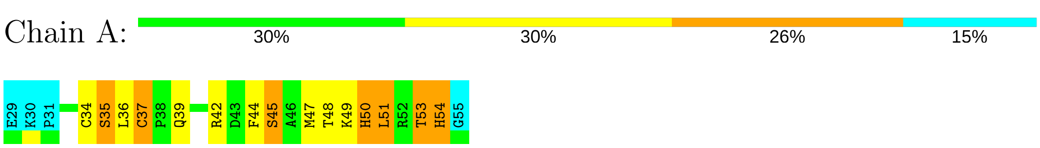
4.2.11 Score per residue for model 11 (medoid)

• Molecule 1: Zinc finger and BTB domain-containing protein 32



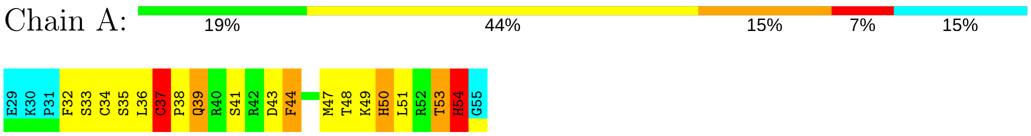
4.2.12 Score per residue for model 12

• Molecule 1: Zinc finger and BTB domain-containing protein 32



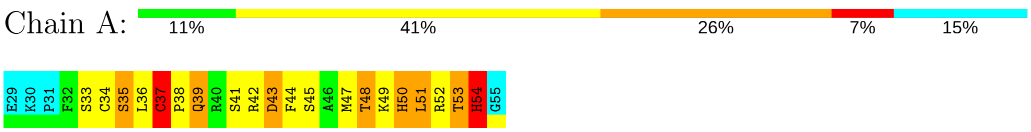
4.2.13 Score per residue for model 13

- Molecule 1: Zinc finger and BTB domain-containing protein 32



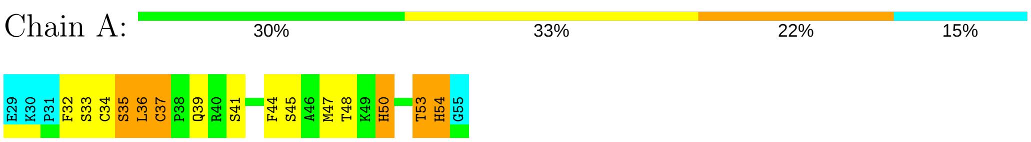
4.2.14 Score per residue for model 14

- Molecule 1: Zinc finger and BTB domain-containing protein 32



4.2.15 Score per residue for model 15

- Molecule 1: Zinc finger and BTB domain-containing protein 32



5 Refinement protocol and experimental data overview (i)

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

Of the 50 calculated structures, 15 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
X-PLOR NIH	structure solution	2.18
X-PLOR NIH	refinement	2.18

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 16797
Number of chemical shift lists	1
Total number of shifts	307
Number of shifts mapped to atoms	307
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	83%

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

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	187	181	181	17±2
All	All	2820	2715	2715	262

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 47.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:39:GLN:NE2	1:A:50:HIS:CE1	0.79	2.51	5	4
1:A:39:GLN:HE21	1:A:50:HIS:CD2	0.78	1.97	15	1
1:A:34:CYS:SG	1:A:35:SER:N	0.76	2.58	12	15
1:A:50:HIS:C	1:A:50:HIS:ND1	0.76	2.39	13	8
1:A:50:HIS:ND1	1:A:50:HIS:C	0.74	2.41	1	7
1:A:39:GLN:NE2	1:A:50:HIS:NE2	0.74	2.36	15	3
1:A:44:PHE:CE2	1:A:48:THR:CG2	0.73	2.71	12	5
1:A:36:LEU:O	1:A:37:CYS:SG	0.71	2.49	12	15
1:A:39:GLN:OE1	1:A:50:HIS:CG	0.70	2.44	11	5
1:A:50:HIS:ND1	1:A:50:HIS:O	0.70	2.25	10	9
1:A:39:GLN:OE1	1:A:50:HIS:CE1	0.69	2.45	9	3
1:A:39:GLN:HE22	1:A:50:HIS:CE1	0.69	2.05	15	4

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:50:HIS:O	1:A:50:HIS:ND1	0.68	2.26	5	6
1:A:44:PHE:CD2	1:A:48:THR:CG2	0.68	2.76	6	4
1:A:40:ARG:N	1:A:40:ARG:CD	0.68	2.57	2	1
1:A:39:GLN:OE1	1:A:50:HIS:CD2	0.67	2.48	3	3
1:A:39:GLN:NE2	1:A:50:HIS:CD2	0.67	2.61	15	1
1:A:36:LEU:C	1:A:37:CYS:SG	0.66	2.73	14	15
1:A:36:LEU:O	1:A:37:CYS:CB	0.66	2.44	1	15
1:A:39:GLN:NE2	1:A:50:HIS:CG	0.65	2.64	9	2
1:A:37:CYS:O	1:A:39:GLN:N	0.64	2.31	14	10
1:A:51:LEU:HD12	1:A:51:LEU:N	0.61	2.10	13	1
1:A:36:LEU:C	1:A:36:LEU:HD23	0.60	2.17	13	3
1:A:40:ARG:CD	1:A:40:ARG:H	0.59	2.10	2	1
1:A:51:LEU:N	1:A:51:LEU:HD12	0.59	2.12	9	2
1:A:39:GLN:C	1:A:39:GLN:HE21	0.59	2.01	2	1
1:A:53:THR:O	1:A:54:HIS:C	0.58	2.42	15	15
1:A:45:SER:O	1:A:48:THR:OG1	0.58	2.20	1	10
1:A:41:SER:OG	1:A:47:MET:SD	0.58	2.55	13	2
1:A:37:CYS:SG	1:A:39:GLN:OE1	0.57	2.62	9	1
1:A:36:LEU:O	1:A:36:LEU:HD23	0.57	1.99	7	2
1:A:39:GLN:O	1:A:39:GLN:OE1	0.57	2.22	10	1
1:A:49:LYS:O	1:A:53:THR:N	0.56	2.38	2	14
1:A:51:LEU:HD23	1:A:51:LEU:N	0.54	2.16	2	7
1:A:44:PHE:CD2	1:A:48:THR:HG23	0.54	2.37	15	3
1:A:50:HIS:O	1:A:54:HIS:N	0.53	2.42	15	3
1:A:44:PHE:O	1:A:48:THR:N	0.51	2.42	2	5
1:A:44:PHE:O	1:A:48:THR:HG23	0.51	2.06	11	5
1:A:43:ASP:O	1:A:44:PHE:C	0.51	2.48	5	1
1:A:51:LEU:N	1:A:51:LEU:HD23	0.50	2.20	10	4
1:A:42:ARG:O	1:A:43:ASP:OD1	0.50	2.29	10	1
1:A:36:LEU:HD23	1:A:36:LEU:C	0.50	2.27	2	2
1:A:39:GLN:HE21	1:A:40:ARG:N	0.50	2.03	2	1
1:A:42:ARG:CG	1:A:43:ASP:N	0.49	2.75	14	1
1:A:47:MET:O	1:A:51:LEU:HD21	0.49	2.07	14	1
1:A:39:GLN:C	1:A:39:GLN:OE1	0.47	2.52	10	1
1:A:51:LEU:HD12	1:A:51:LEU:H	0.47	1.70	9	2
1:A:47:MET:O	1:A:51:LEU:CD1	0.45	2.65	9	2
1:A:45:SER:OG	1:A:46:ALA:N	0.45	2.49	5	1
1:A:51:LEU:CD1	1:A:51:LEU:N	0.45	2.79	13	2
1:A:39:GLN:C	1:A:39:GLN:NE2	0.45	2.68	2	1
1:A:47:MET:O	1:A:51:LEU:CD2	0.45	2.65	14	1
1:A:34:CYS:CB	1:A:37:CYS:SG	0.45	3.05	15	4

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:44:PHE:CE2	1:A:48:THR:HG23	0.45	2.45	14	3
1:A:39:GLN:OE1	1:A:47:MET:CE	0.44	2.65	13	1
1:A:53:THR:O	1:A:54:HIS:O	0.44	2.35	7	1
1:A:34:CYS:HB3	1:A:37:CYS:SG	0.43	2.54	15	5
1:A:36:LEU:O	1:A:37:CYS:HB3	0.43	2.13	1	4
1:A:44:PHE:CE2	1:A:48:THR:HG22	0.42	2.50	6	1
1:A:51:LEU:N	1:A:51:LEU:CD2	0.42	2.80	8	1
1:A:36:LEU:O	1:A:36:LEU:CD2	0.42	2.66	7	2
1:A:43:ASP:O	1:A:45:SER:N	0.42	2.52	5	1
1:A:39:GLN:OE1	1:A:50:HIS:NE2	0.41	2.52	8	1
1:A:36:LEU:C	1:A:36:LEU:CD2	0.41	2.88	13	3
1:A:50:HIS:CD2	1:A:51:LEU:HD23	0.41	2.51	4	1
1:A:51:LEU:CD2	1:A:51:LEU:N	0.40	2.78	2	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	23/27 (85%)	19±1 (81±6%)	2±1 (8±4%)	2±1 (10±4%)	1 9
All	All	345/405 (85%)	281 (81%)	28 (8%)	36 (10%)	1 9

All 5 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	37	CYS	15
1	A	38	PRO	10
1	A	54	HIS	9
1	A	35	SER	1
1	A	32	PHE	1

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	22/25 (88%)	13±2 (58±10%)	9±2 (42±10%)	0 4
All	All	330/375 (88%)	193 (58%)	137 (42%)	0 4

All 19 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	54	HIS	15
1	A	53	THR	15
1	A	50	HIS	15
1	A	51	LEU	11
1	A	33	SER	11
1	A	43	ASP	8
1	A	35	SER	8
1	A	41	SER	7
1	A	48	THR	6
1	A	40	ARG	6
1	A	45	SER	6
1	A	32	PHE	5
1	A	37	CYS	5
1	A	42	ARG	5
1	A	39	GLN	5
1	A	52	ARG	3
1	A	47	MET	3
1	A	36	LEU	2
1	A	44	PHE	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 (i)

The completeness of assignment taking into account all chemical shift lists is 83% for the well-defined parts and 84% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 16797

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	307
Number of shifts mapped to atoms	307
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	27	-0.35 ± 0.29	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	26	0.51 ± 0.37	None needed (imprecise)
$^{13}C'$	24	—	None (insufficient data)
^{15}N	25	-0.60 ± 1.02	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 83%, i.e. 250 atoms were assigned a chemical shift out of a possible 301. 2 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	111/113 (98%)	45/45 (100%)	44/46 (96%)	22/22 (100%)
Sidechain	122/156 (78%)	78/95 (82%)	43/50 (86%)	1/11 (9%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	17/32 (53%)	16/18 (89%)	1/12 (8%)	0/2 (0%)
Overall	250/301 (83%)	139/158 (88%)	88/108 (81%)	23/35 (66%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 84%, i.e. 294 atoms were assigned a chemical shift out of a possible 348. 2 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	128/131 (98%)	52/52 (100%)	51/54 (94%)	25/25 (100%)
Sidechain	149/185 (81%)	96/113 (85%)	52/60 (87%)	1/12 (8%)
Aromatic	17/32 (53%)	16/18 (89%)	1/12 (8%)	0/2 (0%)
Overall	294/348 (84%)	164/183 (90%)	104/126 (83%)	26/39 (67%)

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	54	HIS	CD2	67.11	137.40 – 103.40	-15.7
1	A	47	MET	CG	43.06	38.33 – 25.73	8.8
1	A	47	MET	CE	27.29	26.97 – 7.37	5.2

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:


0.3
0.2
0.1
0
RCI