Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 11:25 PM BST

PDB ID	:	2KKZ
Title	:	Solution NMR structure of the monomeric W187R mutant of A/Udorn NS1 effector domain. Northeast Structural Genomics target OR8C[W187R].
Authors	:	Aramini, J.M.; Ma, L.; Lee, H.; Zhao, L.; Cunningham, K.; Ciccosanti, C.; Janjua, H.; Fang, Y.; Xiao, R.; Krug, R.M.; Montelione, G.T.; Northeast Structural Genomics Consortium (NESG)
Deposited on	:	2009-06-29

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:

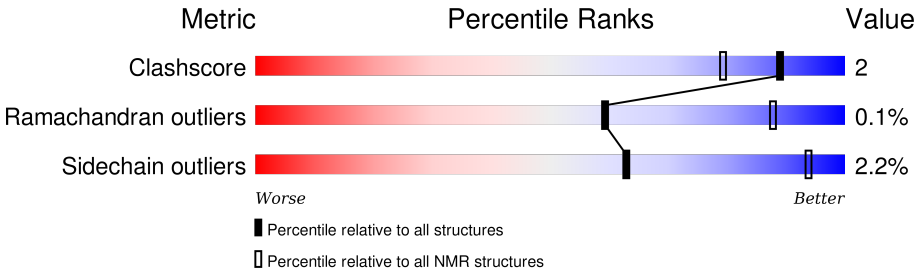
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 91%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	140	77% · 16% ·

2 Ensemble composition and analysis

This entry contains 20 models. Model 12 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:88-A:163, A:169-A:203 (111)	0.34	12

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 4 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2137 atoms, of which 1080 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Non-structural protein NS1.

Mol	Chain	Residues	Atoms						Trace
1	A	134	Total	C	H	N	O	S	0
			2137	672	1080	181	197	7	

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	84	MET	-	INITIATING METHIONINE	UNP Q6XSU2
A	187	ARG	TRP	ENGINEERED	UNP Q6XSU2
A	216	LEU	-	EXPRESSION TAG	UNP Q6XSU2
A	217	GLU	-	EXPRESSION TAG	UNP Q6XSU2
A	218	HIS	-	EXPRESSION TAG	UNP Q6XSU2
A	219	HIS	-	EXPRESSION TAG	UNP Q6XSU2
A	220	HIS	-	EXPRESSION TAG	UNP Q6XSU2
A	221	HIS	-	EXPRESSION TAG	UNP Q6XSU2
A	222	HIS	-	EXPRESSION TAG	UNP Q6XSU2
A	223	HIS	-	EXPRESSION TAG	UNP Q6XSU2

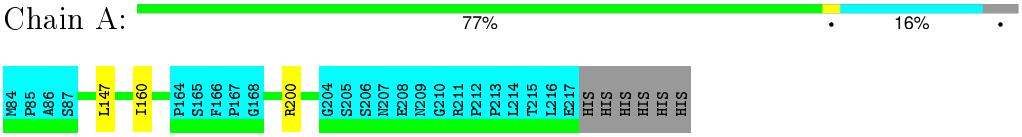
4 Residue-property plots (i)

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: Non-structural protein NS1

Chain A:



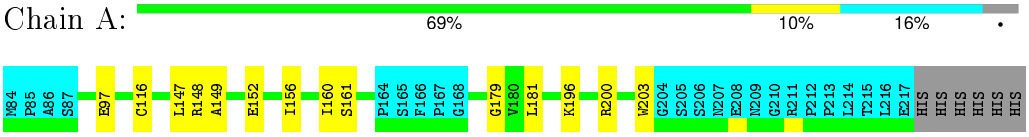
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Non-structural protein NS1

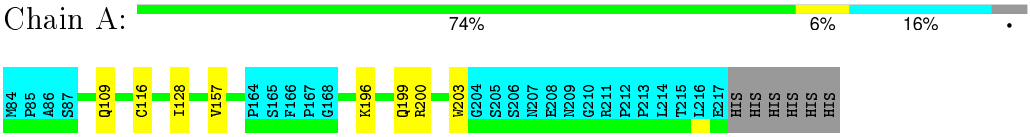
Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: Non-structural protein NS1

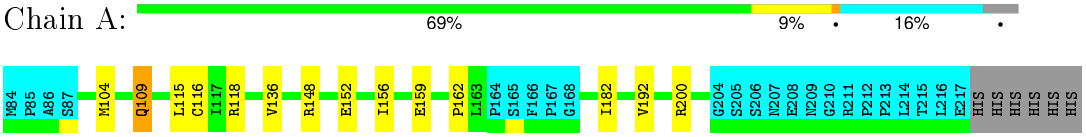
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Non-structural protein NS1

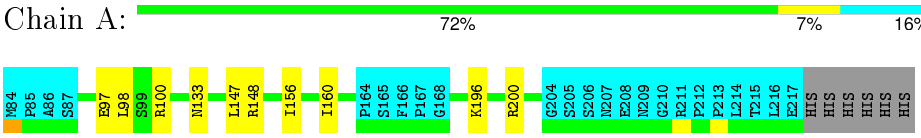
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Non-structural protein NS1

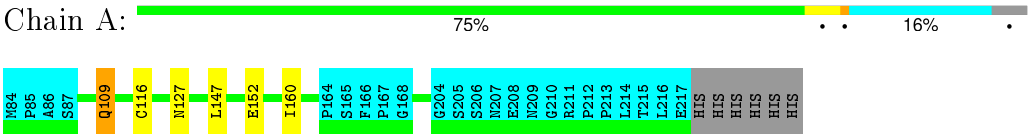
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Non-structural protein NS1

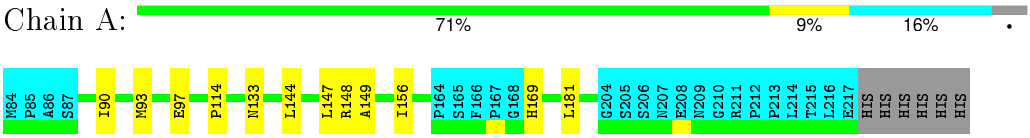
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Non-structural protein NS1

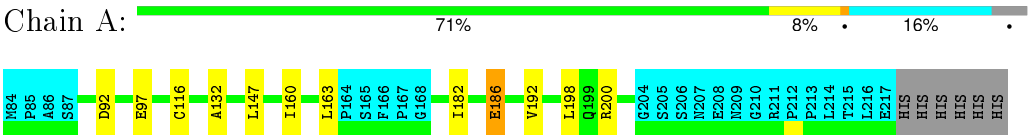
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Non-structural protein NS1

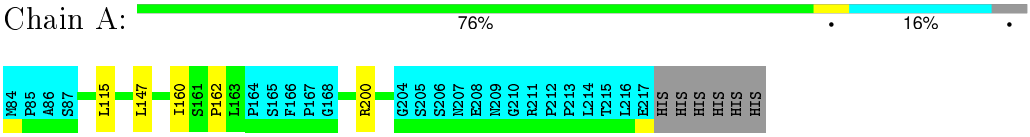
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Non-structural protein NS1

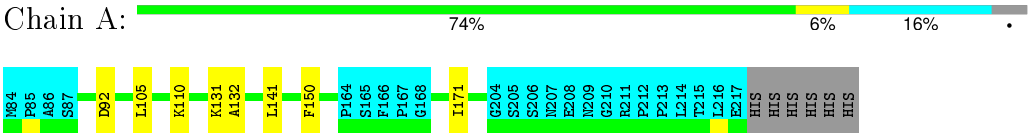
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Non-structural protein NS1

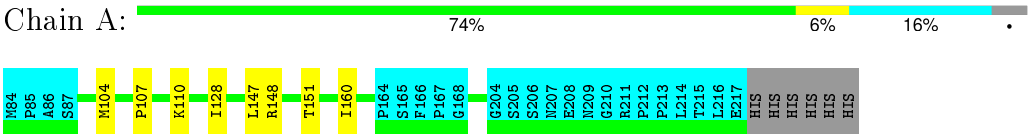
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Non-structural protein NS1

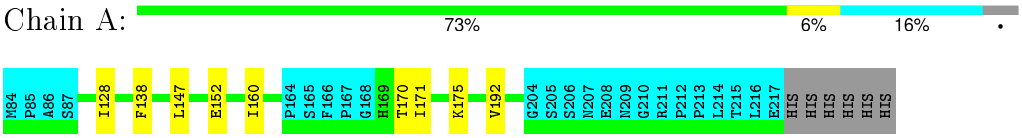
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Non-structural protein NS1

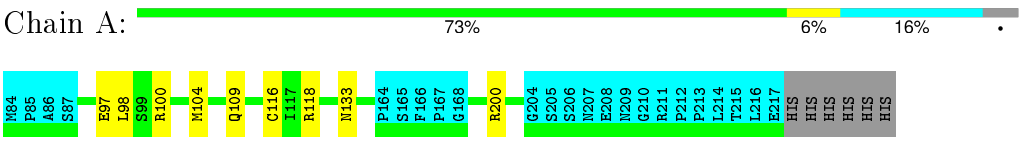
Chain A:



4.2.12 Score per residue for model 12 (medoid)

• Molecule 1: Non-structural protein NS1

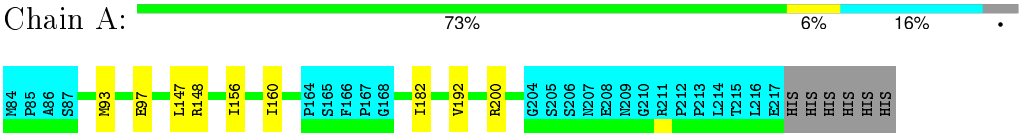
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Non-structural protein NS1

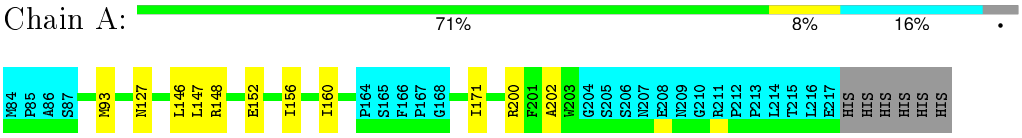
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Non-structural protein NS1

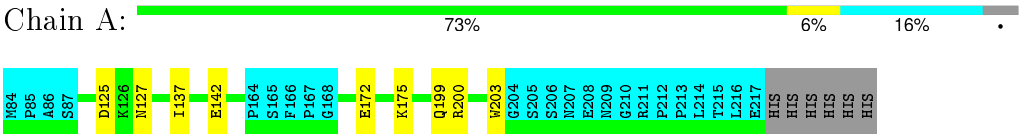
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Non-structural protein NS1

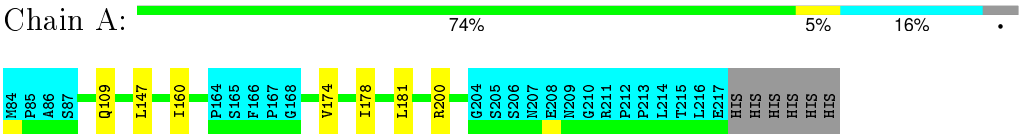
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Non-structural protein NS1

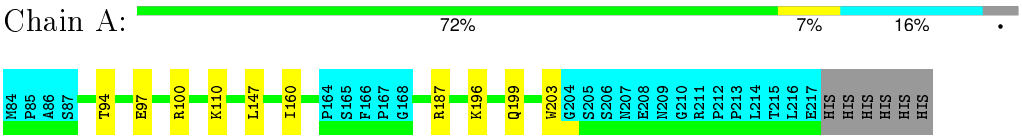
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Non-structural protein NS1

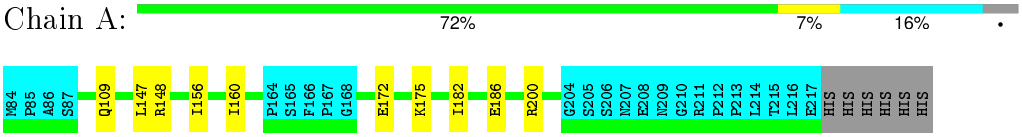
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Non-structural protein NS1

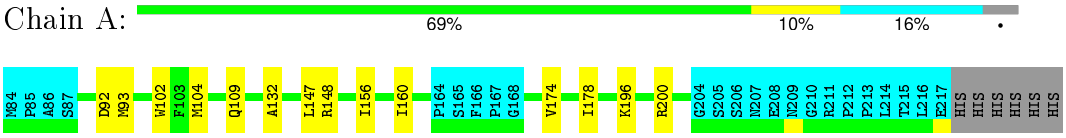
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Non-structural protein NS1

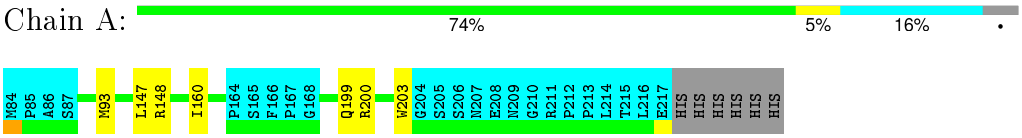
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Non-structural protein NS1

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 100 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 solution	3.0
CNS	refinement	1.1
AutoStructure	structure solution	2.2.1
TALOS	refinement	plus

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 16376
Number of chemical shift lists	1
Total number of shifts	1741
Number of shifts mapped to atoms	1741
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	91%

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

6 Model quality

6.1 Standard geometry

There are no covalent bond-length or bond-angle outliers.

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	0.1±0.2
All	All	0	1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	148	ARG	Sidechain	1

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	894	925	925	4±1
All	All	17880	18500	18500	83

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:148:ARG:HD3	1:A:156:ILE:HD13	0.58	1.73	14	2
1:A:182:ILE:HG23	1:A:192:VAL:HG21	0.57	1.74	3	3

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:93:MET:SD	1:A:148:ARG:HG2	0.57	2.40	20	1
1:A:93:MET:HG3	1:A:97:GLU:HG3	0.56	1.76	13	2
1:A:147:LEU:HB3	1:A:160:ILE:HB	0.56	1.77	7	13
1:A:148:ARG:HH11	1:A:156:ILE:HG21	0.54	1.63	13	5
1:A:116:CYS:SG	1:A:161:SER:HB2	0.52	2.43	1	1
1:A:98:LEU:HD11	1:A:133:ASN:ND2	0.52	2.19	4	1
1:A:102:TRP:CZ3	1:A:104:MET:SD	0.51	3.04	19	1
1:A:109:GLN:NE2	1:A:116:CYS:SG	0.50	2.85	5	3
1:A:97:GLU:HA	1:A:100:ARG:HB2	0.49	1.82	4	3
1:A:116:CYS:SG	1:A:163:LEU:HD21	0.49	2.48	7	1
1:A:116:CYS:SG	1:A:118:ARG:NH1	0.49	2.86	12	1
1:A:93:MET:SD	1:A:148:ARG:HB3	0.48	2.49	19	2
1:A:172:GLU:HA	1:A:175:LYS:NZ	0.48	2.23	15	1
1:A:131:LYS:HG3	1:A:150:PHE:CZ	0.48	2.44	9	1
1:A:141:LEU:O	1:A:171:ILE:HB	0.48	2.09	9	1
1:A:147:LEU:HD13	1:A:160:ILE:HD12	0.47	1.84	11	4
1:A:172:GLU:HA	1:A:175:LYS:HE2	0.47	1.85	18	1
1:A:132:ALA:HB3	1:A:198:LEU:HD21	0.47	1.85	7	1
1:A:149:ALA:HB2	1:A:181:LEU:HD21	0.46	1.86	6	2
1:A:93:MET:SD	1:A:148:ARG:NE	0.45	2.90	6	1
1:A:115:LEU:HA	1:A:162:PRO:HA	0.45	1.88	8	2
1:A:104:MET:SD	1:A:118:ARG:HD2	0.45	2.51	3	2
1:A:199:GLN:HA	1:A:203:TRP:HB3	0.45	1.89	2	4
1:A:104:MET:SD	1:A:107:PRO:HB3	0.44	2.53	10	1
1:A:132:ALA:HB2	1:A:147:LEU:HG	0.44	1.88	7	1
1:A:128:ILE:HB	1:A:192:VAL:HG12	0.44	1.90	11	1
1:A:118:ARG:NH2	1:A:156:ILE:HG22	0.43	2.29	3	1
1:A:92:ASP:HB3	1:A:132:ALA:O	0.43	2.13	9	3
1:A:146:LEU:HD21	1:A:148:ARG:NH2	0.43	2.29	14	1
1:A:144:LEU:HD11	1:A:147:LEU:HB2	0.43	1.91	6	1
1:A:160:ILE:HG13	1:A:181:LEU:HD22	0.43	1.90	1	2
1:A:128:ILE:HG12	1:A:151:THR:HG22	0.42	1.92	10	1
1:A:137:ILE:HG22	1:A:142:GLU:HG2	0.42	1.89	15	1
1:A:182:ILE:O	1:A:186:GLU:HG2	0.41	2.15	18	1
1:A:148:ARG:NH1	1:A:156:ILE:HD13	0.41	2.30	1	1
1:A:90:ILE:O	1:A:133:ASN:HA	0.41	2.16	6	1
1:A:174:VAL:O	1:A:178:ILE:HG13	0.41	2.15	19	2
1:A:171:ILE:HD11	1:A:202:ALA:HA	0.41	1.92	14	1
1:A:128:ILE:HD13	1:A:157:VAL:HG21	0.41	1.92	2	1
1:A:170:THR:HG22	1:A:171:ILE:HG22	0.40	1.92	11	1
1:A:98:LEU:HD12	1:A:133:ASN:ND2	0.40	2.31	12	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:118:ARG:HB3	1:A:159:GLU:HB3	0.40	1.94	3	1
1:A:114:PRO:HG2	1:A:169:HIS:HB3	0.40	1.92	6	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	111/140 (79%)	106±1 (96±1%)	4±1 (4±1%)	0±0 (0±0%)	59	88
All	All	2220/2800 (79%)	2129 (96%)	89 (4%)	2 (0%)	59	88

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	203	TRP	1
1	A	138	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	99/124 (80%)	97±1 (98±1%)	2±1 (2±1%)	63	94
All	All	1980/2480 (80%)	1936 (98%)	44 (2%)	63	94

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	200	ARG	14

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	109	GLN	6
1	A	152	GLU	5
1	A	196	LYS	5
1	A	127	ASN	3
1	A	110	LYS	3
1	A	97	GLU	2
1	A	187	ARG	1
1	A	94	THR	1
1	A	186	GLU	1
1	A	175	LYS	1
1	A	105	LEU	1
1	A	125	ASP	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 carbohydrates in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

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 91% for the well-defined parts and 91% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 16376

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	1741
Number of shifts mapped to atoms	1741
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	134	-0.29 ± 0.07	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	122	-0.03 ± 0.08	None needed (< 0.5 ppm)
$^{13}C'$	127	0.13 ± 0.17	None needed (< 0.5 ppm)
^{15}N	123	0.18 ± 0.18	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 91%, i.e. 1305 atoms were assigned a chemical shift out of a possible 1428. 19 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	544/549 (99%)	218/219 (100%)	219/222 (99%)	107/108 (99%)
Sidechain	679/794 (86%)	410/461 (89%)	259/295 (88%)	10/38 (26%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	82/85 (96%)	43/45 (96%)	35/36 (97%)	4/4 (100%)
Overall	1305/1428 (91%)	671/725 (93%)	513/553 (93%)	121/150 (81%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 91%, i.e. 1533 atoms were assigned a chemical shift out of a possible 1680. 21 out of 21 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	641/654 (98%)	257/260 (99%)	261/268 (97%)	123/126 (98%)
Sidechain	802/932 (86%)	489/546 (90%)	301/343 (88%)	12/43 (28%)
Aromatic	90/94 (96%)	47/50 (94%)	39/40 (98%)	4/4 (100%)
Overall	1533/1680 (91%)	793/856 (93%)	601/651 (92%)	139/173 (80%)

7.1.4 Statistically unusual chemical shifts (i)

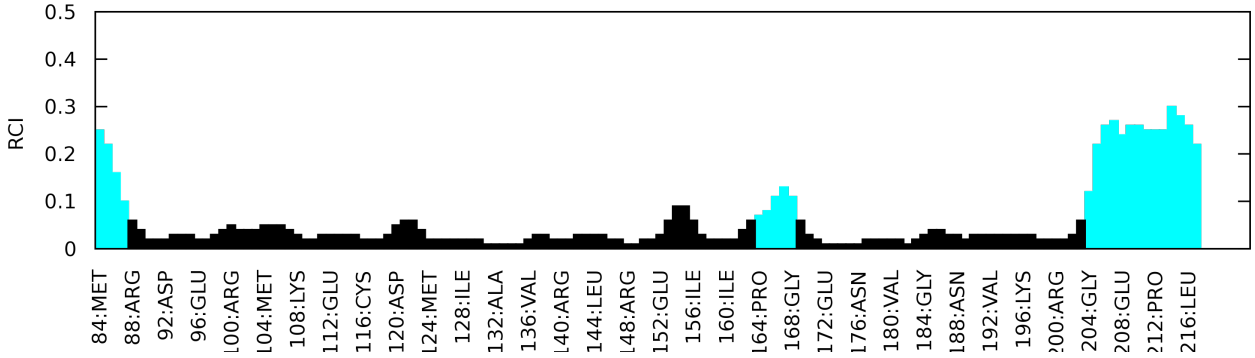
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	144	LEU	H	12.56	11.47 – 4.97	6.7
1	A	156	ILE	HG22	-0.72	2.13 – -0.57	-5.6
1	A	156	ILE	HG23	-0.72	2.13 – -0.57	-5.6
1	A	156	ILE	HG21	-0.72	2.13 – -0.57	-5.6
1	A	90	ILE	HB	0.30	3.24 – 0.34	-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
84:MET
88:ARG
92:ASP
96:GLU
100:ARG
104:MET
108:LYS
112:GLU
116:CYS
120:ASP
124:MET
128:ILE
132:ALA
136:VAL
140:ARG
144:LEU
148:ARG
152:GLU
156:ILE
160:ILE
164:PRO
168:GLY
172:GLU
176:ASN
180:VAL
184:GLY
188:ASN
192:VAL
196:LYS
200:ARG
204:GLY
208:GLU
212:PRO
216:LEU