Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 05:46 PM BST

PDB ID	:	1WNJ
Title	:	NMR structure of human coactosin-like protein
Authors	:	Liepinsh, E.; Rakonjac, M.; Boissonneault, V.; Provost, P.; Samuelsson, B.; Radmark, O.; Otting, G.
Deposited on	:	2004-08-05

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:

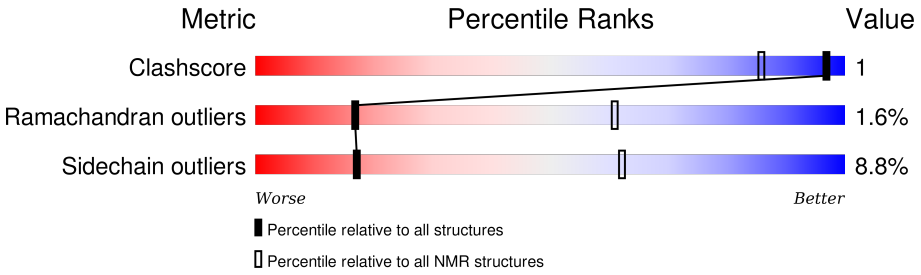
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 (i)

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

The overall completeness of chemical shifts assignment was not calculated.

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	145	79% 8% 13%

2 Ensemble composition and analysis

This entry contains 20 models. Model 13 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:5-A:68, A:74-A:135 (126)	0.39	13

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 3 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2277 atoms, of which 1131 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Coactosin-like protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	145	2277	722	1131	197	223	4	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	CLONING ARTIFACT	UNP Q14019
A	2	ILE	-	CLONING ARTIFACT	UNP Q14019
A	3	ARG	-	CLONING ARTIFACT	UNP Q14019

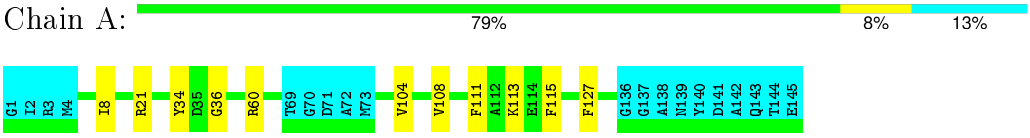
4 Residue-property plots

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: Coactosin-like protein

Chain A:



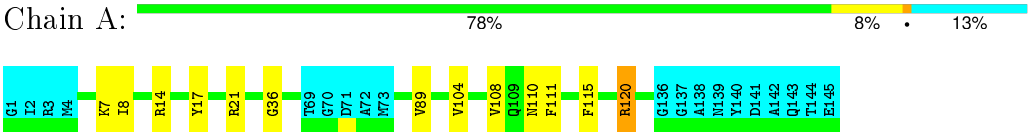
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

- Molecule 1: Coactosin-like protein

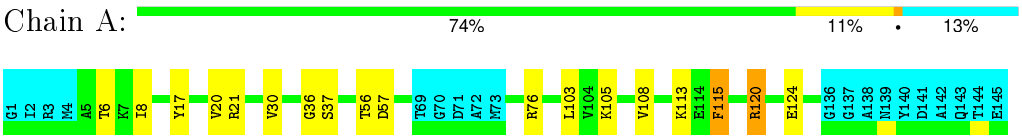
Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: Coactosin-like protein

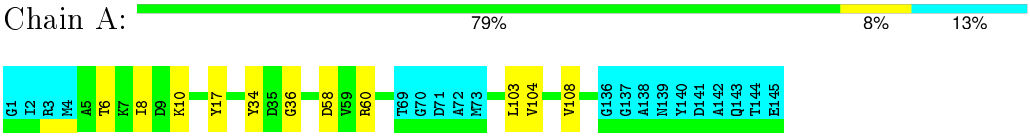
Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Coactosin-like protein

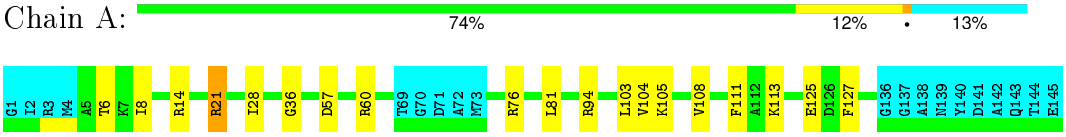
Chain A:



4.2.4 Score per residue for model 4

- Molecule 1: Coactosin-like protein

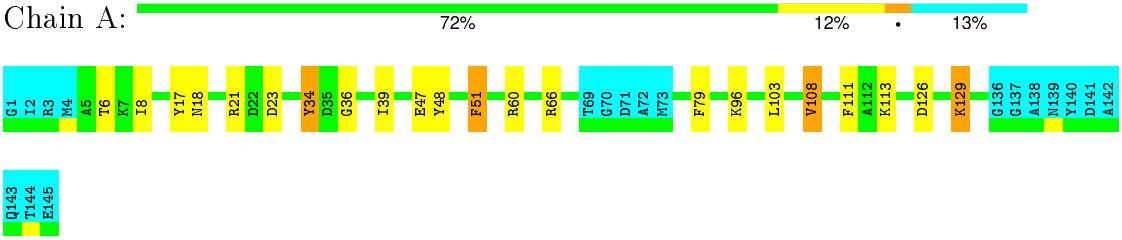
Chain A:



4.2.5 Score per residue for model 5

- Molecule 1: Coactosin-like protein

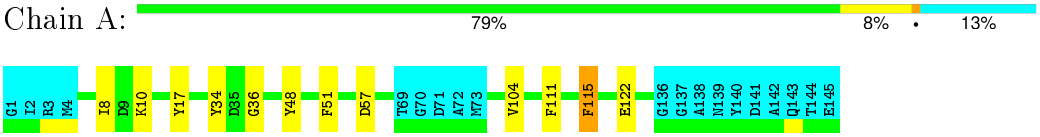
Chain A:



4.2.6 Score per residue for model 6

- Molecule 1: Coactosin-like protein

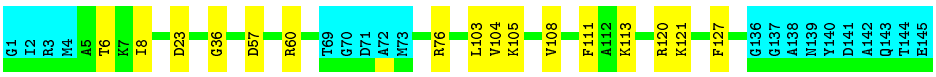
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Coactosin-like protein

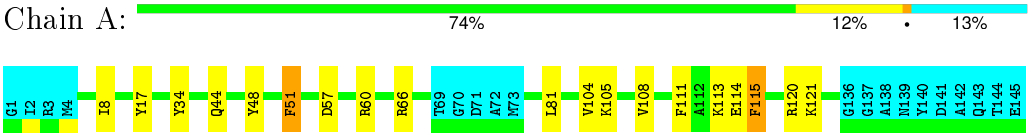
Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Coactosin-like protein

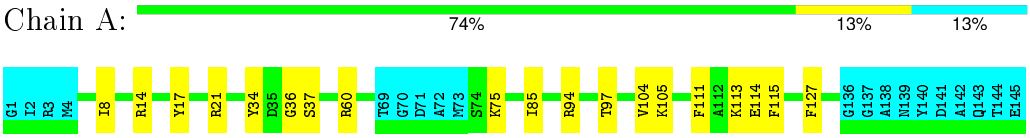
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Coactosin-like protein

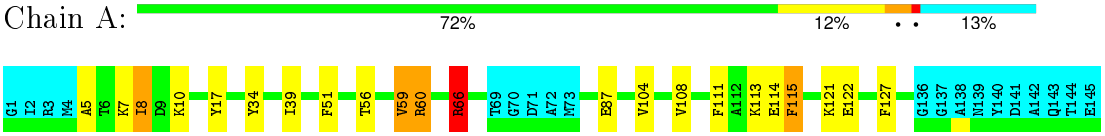
Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: Coactosin-like protein

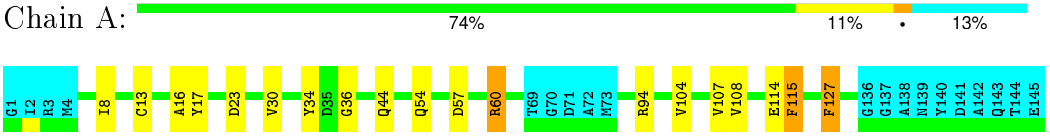
Chain A:



4.2.11 Score per residue for model 11

- Molecule 1: Coactosin-like protein

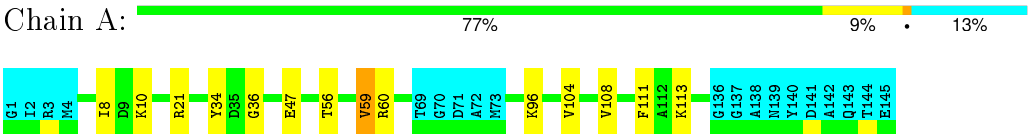
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Coactosin-like protein

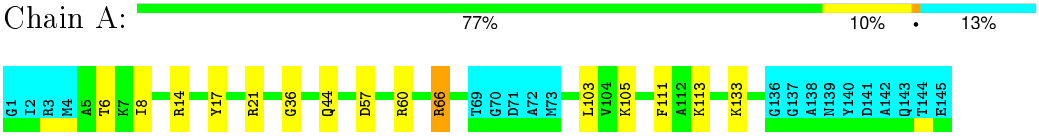
Chain A:



4.2.13 Score per residue for model 13 (medoid)

- Molecule 1: Coactosin-like protein

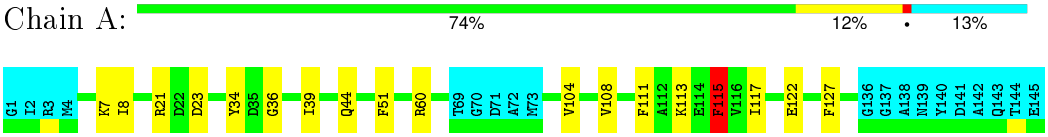
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Coactosin-like protein

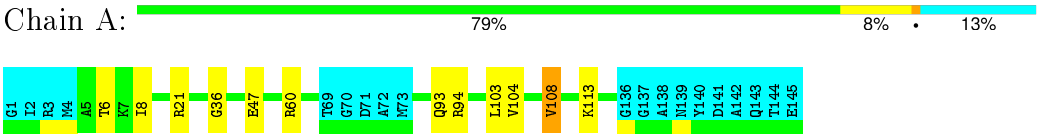
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Coactosin-like protein

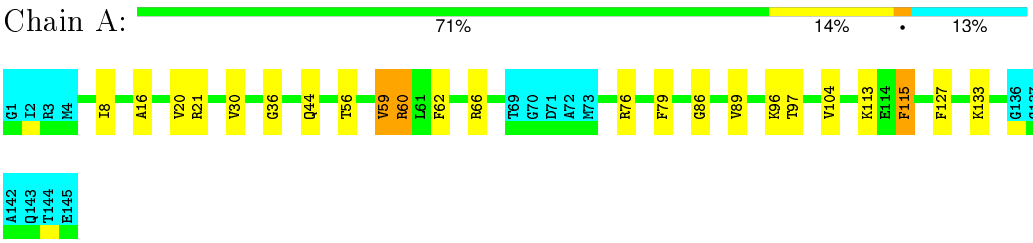
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Coactosin-like protein

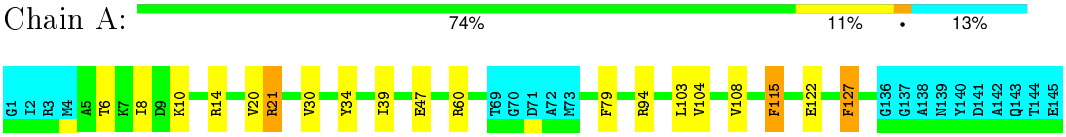
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Coactosin-like protein

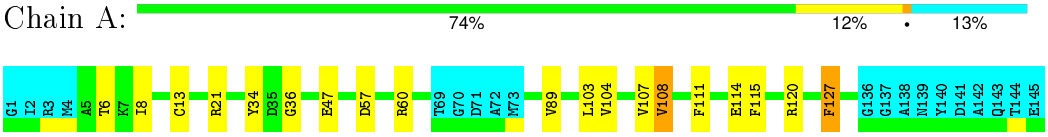
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Coactosin-like protein

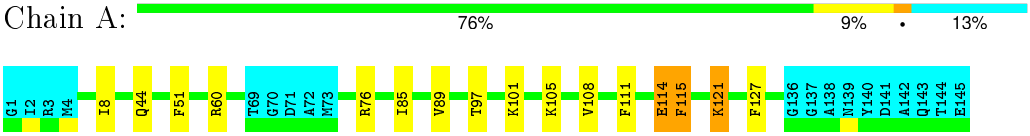
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Coactosin-like protein

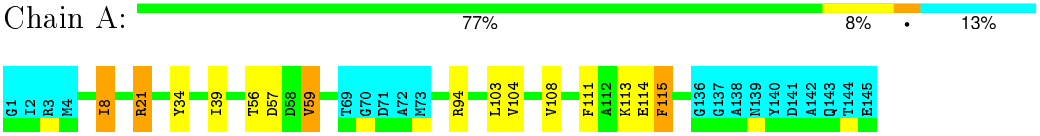
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Coactosin-like protein

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *all calculated structures submitted*.

The following table shows the software used for structure solution, optimisation and refinement.

Software name	Classification	Version
DYANA	structure solution	1.5
Opal	refinement	2.6

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

6 Model quality

6.1 Standard geometry

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	0.61±0.01	0±0/1029 (0.0±0.0%)	1.14±0.03	2±1/1385 (0.2±0.1%)
All	All	0.61	0/20580 (0.0%)	1.14	46/27700 (0.2%)

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	2.0±1.1
All	All	0	40

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	115	PHE	CB-CG-CD2	-9.22	114.34	120.80	16	12
1	A	115	PHE	CB-CG-CD1	7.64	126.15	120.80	16	8
1	A	66	ARG	NE-CZ-NH2	-6.74	116.93	120.30	5	4
1	A	59	VAL	CA-CB-CG2	6.67	120.90	110.90	10	4
1	A	60	ARG	NE-CZ-NH2	-6.19	117.20	120.30	11	1
1	A	17	TYR	CB-CG-CD1	-6.09	117.34	121.00	13	1
1	A	21	ARG	NE-CZ-NH1	6.02	123.31	120.30	2	1
1	A	127	PHE	CB-CG-CD2	-5.96	116.63	120.80	14	4
1	A	14	ARG	NE-CZ-NH2	-5.71	117.44	120.30	9	2
1	A	21	ARG	CD-NE-CZ	5.67	131.54	123.60	2	1
1	A	21	ARG	NE-CZ-NH2	-5.66	117.47	120.30	12	3
1	A	108	VAL	CA-CB-CG1	5.42	119.03	110.90	5	2
1	A	56	THR	C-N-CA	5.19	134.67	121.70	2	1
1	A	108	VAL	CG1-CB-CG2	-5.19	102.60	110.90	15	1
1	A	60	ARG	CB-CA-C	5.10	120.60	110.40	11	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	34	TYR	Sidechain	10
1	A	17	TYR	Sidechain	9
1	A	60	ARG	Sidechain	4
1	A	127	PHE	Sidechain	4
1	A	79	PHE	Sidechain	3
1	A	76	ARG	Sidechain	3
1	A	120	ARG	Sidechain	3
1	A	21	ARG	Sidechain	3
1	A	66	ARG	Sidechain	1

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	1011	1012	1012	2±1
All	All	20220	20240	20240	42

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:6:THR:HB	1:A:103:LEU:HD11	0.64	1.69	2	9
1:A:62:PHE:CZ	1:A:97:THR:HG23	0.54	2.37	16	1
1:A:86:GLY:O	1:A:89:VAL:HG12	0.54	2.02	16	1
1:A:85:ILE:HD11	1:A:97:THR:HG21	0.53	1.78	9	2
1:A:13:CYS:HB3	1:A:107:VAL:HG11	0.53	1.80	18	2
1:A:56:THR:OG1	1:A:59:VAL:HG22	0.53	2.04	10	4
1:A:16:ALA:HB1	1:A:30:VAL:HG11	0.50	1.82	16	2
1:A:121:LYS:H	1:A:121:LYS:HE3	0.48	1.68	19	1
1:A:6:THR:CB	1:A:103:LEU:HD11	0.45	2.40	2	2
1:A:8:ILE:HG22	1:A:39:ILE:HG12	0.45	1.89	10	2
1:A:34:TYR:CE2	1:A:39:ILE:CD1	0.45	2.99	5	3

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:120:ARG:HH12	1:A:124:GLU:CD	0.43	2.15	2	1
1:A:20:VAL:HG21	1:A:30:VAL:HG23	0.42	1.92	2	1
1:A:8:ILE:HG21	1:A:103:LEU:O	0.42	2.13	20	1
1:A:16:ALA:CB	1:A:30:VAL:HG11	0.42	2.44	11	1
1:A:13:CYS:CB	1:A:107:VAL:HG11	0.42	2.44	18	1
1:A:48:TYR:CE2	1:A:51:PHE:CD2	0.42	3.08	8	1
1:A:48:TYR:CE1	1:A:51:PHE:CD2	0.42	3.07	6	2
1:A:126:ASP:HA	1:A:129:LYS:CE	0.42	2.44	5	1
1:A:20:VAL:CG2	1:A:30:VAL:HG23	0.41	2.46	16	2
1:A:101:LYS:HE2	1:A:114:GLU:HG3	0.40	1.94	19	1
1:A:115:PHE:CD1	1:A:117:ILE:HD11	0.40	2.52	14	1

6.3 Torsion angles (i)

6.3.1 Protein backbone (i)

In the following table, the Percentiles column shows the percent Ramachandran outliers of the chain as a percentile score with respect to all PDB entries followed by that with respect to all NMR entries. The Analysed column shows the number of residues for which the backbone conformation was analysed and the total number of residues.

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	A	126/145 (87%)	117±1 (93±1%)	7±2 (6±1%)	2±1 (2±1%)	17	61
All	All	2520/2900 (87%)	2338 (93%)	141 (6%)	41 (2%)	17	61

All 4 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	108	VAL	16
1	A	36	GLY	15
1	A	57	ASP	9
1	A	5	ALA	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	108/120 (90%)	99±2 (91±2%)	10±2 (9±2%)	17 62
All	All	2160/2400 (90%)	1970 (91%)	190 (9%)	17 62

All 38 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	8	ILE	20
1	A	104	VAL	16
1	A	111	PHE	14
1	A	60	ARG	14
1	A	113	LYS	13
1	A	115	PHE	11
1	A	21	ARG	9
1	A	114	GLU	7
1	A	105	LYS	7
1	A	94	ARG	6
1	A	44	GLN	6
1	A	127	PHE	5
1	A	47	GLU	5
1	A	10	LYS	5
1	A	51	PHE	5
1	A	122	GLU	4
1	A	23	ASP	4
1	A	121	LYS	4
1	A	7	LYS	3
1	A	96	LYS	3
1	A	120	ARG	3
1	A	89	VAL	3
1	A	66	ARG	3
1	A	14	ARG	3
1	A	133	LYS	2
1	A	81	LEU	2
1	A	37	SER	2
1	A	93	GLN	1
1	A	75	LYS	1
1	A	18	ASN	1
1	A	58	ASP	1
1	A	87	GLU	1
1	A	110	ASN	1
1	A	129	LYS	1
1	A	76	ARG	1
1	A	54	GLN	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	125	GLU	1
1	A	28	ILE	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

No chemical shift data were provided