Full wwPDB NMR Structure Validation Report (i)

Apr 20, 2024 – 11:36 AM EDT

PDB ID	:	1S1N
Title	:	SH3 domain of human nephrocystin
Authors	:	Le Maire, A.; Weber, T.; Saunier, S.; Antignac, C.; Ducruix, A.; Dardel, F.
Deposited on	:	2004-01-07

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

The following versions of software and data (see references (i)) were used in the production of this report:

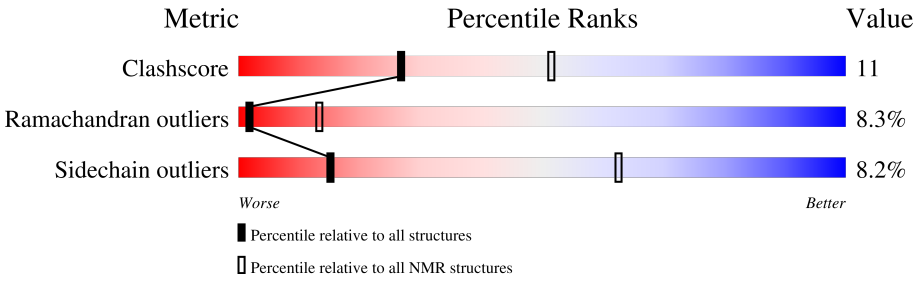
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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	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	68	59% 19% 6% • 12%

2 Ensemble composition and analysis (i)

This entry contains 17 models. Model 1 is the overall representative, medoid model (most similar to other models).

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:154-A:210 (57)	0.62	1

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 and 11 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 942 atoms, of which 466 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Nephrocystin 1.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	60	942	308	466	74	94	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	145	GLY	-	cloning artifact	UNP O15259
A	146	SER	-	cloning artifact	UNP O15259

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Nephrocystin 1

Chain A: 59% 19% 6% • 12%

GLY SER GLU SER HIS LYS TRP SER T153 Y157 I158 F173 K174 K175 I178 V181 I182 E183 K184 K185 W189 W190 I191 E199 R204 T205 Y206 L207 Y210 S211 E212

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 (medoid)

• Molecule 1: Nephrocystin 1

Chain A: 60% 19% • • 12%

GLY SER GLU SER HIS LYS TRP SER T153 Y157 F163 F173 K174 K175 V181 I182 E183 K184 K185 W189 W190 I191 E199 R204 T205 Y206 L207 Y210 S211 E212

4.2.2 Score per residue for model 2

• Molecule 1: Nephrocystin 1

Chain A: 59% 21% • • 12%

GLY SER GLU SER HIS LYS TRP SER T153 Y157 I158 V168 F173 K174 K175 V181 I182 E183 K184 K185 P186 W189 W190 I191 E199 R204 T205 Y210 S211 E212

4.2.3 Score per residue for model 3

- Molecule 1: Nephrocystin 1

Chain A: 57% 21% 6% • 12%

GLY SER GLU SER HIS LYS TRP SER T153 Y157 I158 V168 K175 I178 V181 I182 E183 K184 K185 P186 W189 W190 E199 P203 R204 T205 Y206 L207 Y210 S211 E212

4.2.4 Score per residue for model 4

- Molecule 1: Nephrocystin 1

Chain A: 53% 24% 6% • • 12%

GLY SER GLU SER HIS LYS TRP SER T153 Y157 I158 A159 F173 K174 K175 G176 E177 I178 V181 I182 E183 K184 K185 W189 W190 I191 D194 N198 E199 R204 L207 Y210 S211 E212

4.2.5 Score per residue for model 5

- Molecule 1: Nephrocystin 1

Chain A: 54% 22% 6% • • 12%

GLY SER GLU SER HIS LYS TRP SER T153 Y157 I158 G169 K175 G176 E177 L180 V181 I182 E183 K184 K185 G188 W189 W190 I191 A192 K193 E199 P203 R204 T205 Y206 L207 S211 E212

4.2.6 Score per residue for model 6

- Molecule 1: Nephrocystin 1

Chain A: 57% 18% 9% • 12%

GLY SER GLU SER HIS LYS TRP SER T153 E156 Y157 E177 I178 L179 L180 V181 I182 E183 K184 K185 P186 W189 W190 I191 D194 R204 T205 Y206 L207 S211 E212

4.2.7 Score per residue for model 7

- Molecule 1: Nephrocystin 1

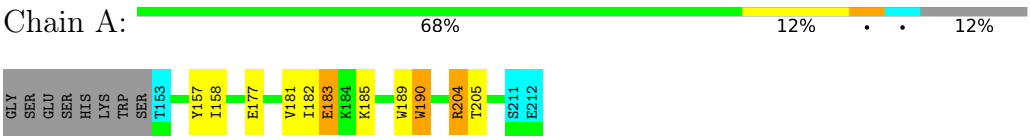
Chain A: 56% 24% • • • 12%

GLY SER GLU SER HIS LYS TRP SER T153 G154 E155 V168 F173 K174 K175 I178 L179 L180 V181 I182 E183 K184 K185 P186 D187 G188 W189 W190 E199 R204 E208 P209 Y210 S211 E212

4.2.8 Score per residue for model 8

- Molecule 1: Nephrocystin 1

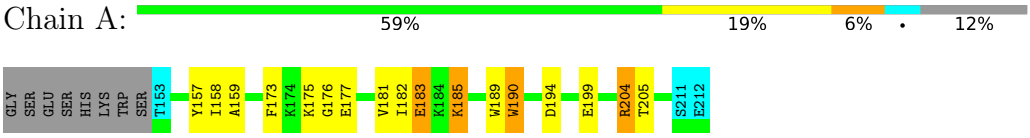
Chain A:



4.2.9 Score per residue for model 9

- Molecule 1: Nephrocystin 1

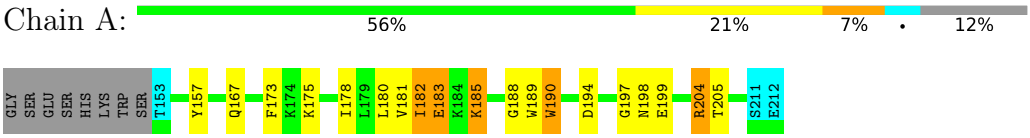
Chain A:



4.2.10 Score per residue for model 10

- Molecule 1: Nephrocystin 1

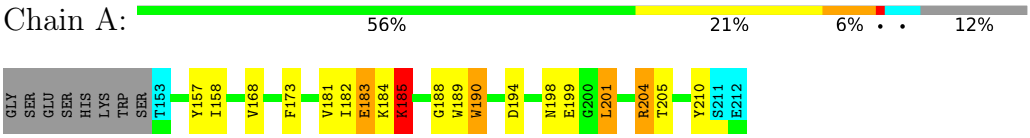
Chain A:



4.2.11 Score per residue for model 11

- Molecule 1: Nephrocystin 1

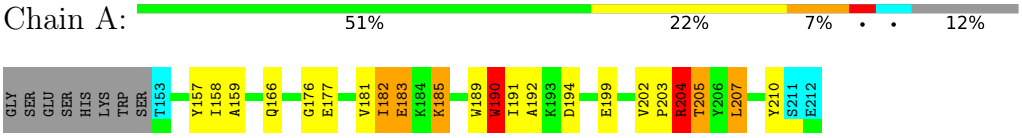
Chain A:



4.2.12 Score per residue for model 12

- Molecule 1: Nephrocystin 1

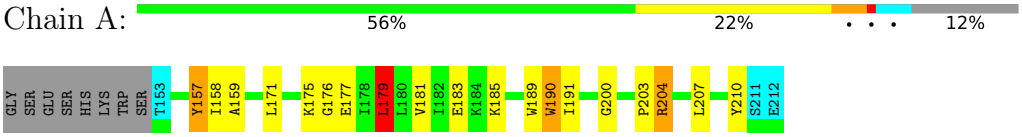
Chain A:



4.2.13 Score per residue for model 13

- Molecule 1: Nephrocystin 1

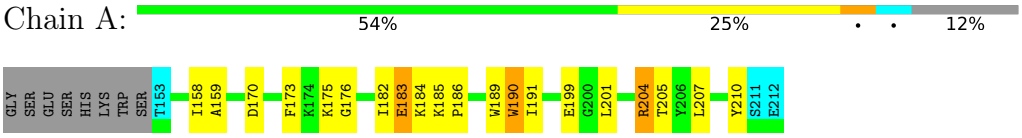
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Nephrocystin 1

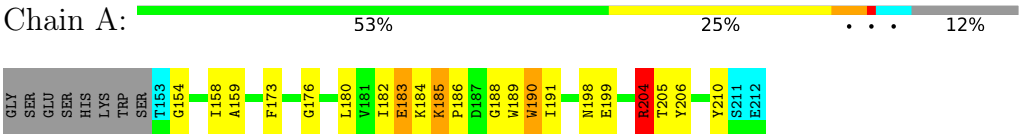
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Nephrocystin 1

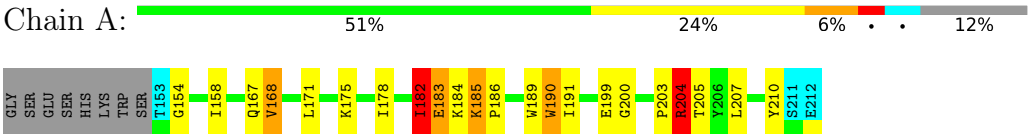
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Nephrocystin 1

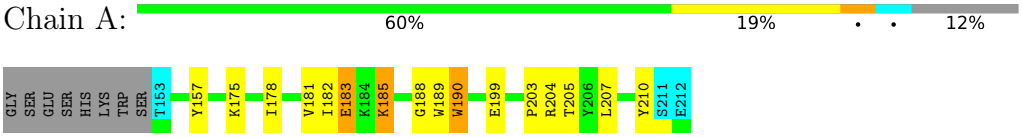
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Nephrocystin 1

Chain A:



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *distance geometry follwed by restrained simulated annealing.*

Of the 20 calculated structures, 17 were deposited, based on the following criterion: *structures with the least restraint violations,structures with the lowest energy.*

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

Software name	Classification	Version
DIANA	structure solution	2.8
X-PLOR	refinement	3.89

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	1.22±0.02	1±1/465 (0.2± 0.1%)	1.63±0.05	8±2/631 (1.3± 0.3%)
All	All	1.22	16/7905 (0.2%)	1.63	142/10727 (1.3%)

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	1.0±0.0
All	All	0	17

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	190	TRP	CG-CD2	-5.95	1.33	1.43	12	3
1	A	189	TRP	CG-CD2	-5.38	1.34	1.43	2	13

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	190	TRP	CD1-NE1-CE2	9.75	117.78	109.00	14	17
1	A	190	TRP	NE1-CE2-CZ2	9.52	140.87	130.40	13	12
1	A	189	TRP	CD1-NE1-CE2	8.94	117.05	109.00	14	17
1	A	157	TYR	CB-CG-CD2	-8.57	115.86	121.00	4	2
1	A	189	TRP	NE1-CE2-CZ2	8.12	139.34	130.40	15	17
1	A	179	LEU	CB-CG-CD1	-7.66	97.97	111.00	13	1
1	A	189	TRP	CG-CD1-NE1	-7.42	102.68	110.10	14	17
1	A	190	TRP	CG-CD1-NE1	-7.39	102.71	110.10	9	17

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	178	ILE	CA-CB-CG1	-7.13	97.45	111.00	4	1
1	A	157	TYR	CB-CG-CD1	6.55	124.93	121.00	4	2
1	A	190	TRP	NE1-CE2-CD2	-6.38	100.92	107.30	13	12
1	A	189	TRP	NE1-CE2-CD2	-6.12	101.19	107.30	15	17
1	A	190	TRP	CG-CD2-CE3	-6.03	128.47	133.90	13	1
1	A	194	ASP	CB-CG-OD2	-5.96	112.94	118.30	12	6
1	A	182	ILE	CA-CB-CG2	-5.25	100.41	110.90	16	1
1	A	189	TRP	CA-CB-CG	-5.22	103.78	113.70	8	2

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	204	ARG	Sidechain	17

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	453	448	448	10±3
All	All	7701	7616	7616	169

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 11.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:158:ILE:HG22	1:A:178:ILE:HG12	0.86	1.46	4	1
1:A:190:TRP:CD1	1:A:190:TRP:N	0.69	2.61	12	7
1:A:185:LYS:HB2	1:A:190:TRP:NE1	0.67	2.03	3	2
1:A:158:ILE:HG22	1:A:178:ILE:CG1	0.65	2.20	4	1
1:A:185:LYS:HB3	1:A:190:TRP:CD1	0.64	2.28	15	4
1:A:185:LYS:N	1:A:190:TRP:CD1	0.62	2.68	16	2
1:A:158:ILE:CG2	1:A:178:ILE:HG12	0.60	2.25	4	1
1:A:190:TRP:CE2	1:A:204:ARG:HB2	0.58	2.34	10	5

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:183:GLU:O	1:A:190:TRP:CE3	0.55	2.59	9	11
1:A:157:TYR:CD1	1:A:181:VAL:HG21	0.55	2.36	9	12
1:A:182:ILE:O	1:A:191:ILE:HB	0.52	2.04	4	9
1:A:185:LYS:HB2	1:A:188:GLY:CA	0.52	2.34	5	4
1:A:182:ILE:O	1:A:183:GLU:HB2	0.51	2.06	12	3
1:A:178:ILE:CD1	1:A:178:ILE:N	0.50	2.75	3	6
1:A:181:VAL:HG22	1:A:190:TRP:CZ3	0.49	2.42	12	1
1:A:185:LYS:O	1:A:185:LYS:CG	0.49	2.61	12	1
1:A:182:ILE:O	1:A:183:GLU:CB	0.49	2.60	2	15
1:A:190:TRP:CZ3	1:A:192:ALA:HB2	0.48	2.43	12	1
1:A:159:ALA:O	1:A:176:GLY:N	0.48	2.47	15	6
1:A:183:GLU:HG2	1:A:184:LYS:N	0.48	2.24	16	5
1:A:185:LYS:HD2	1:A:204:ARG:HB3	0.48	1.85	16	1
1:A:158:ILE:HG21	1:A:210:TYR:CD1	0.48	2.44	13	5
1:A:177:GLU:N	1:A:177:GLU:CD	0.47	2.67	6	2
1:A:184:LYS:O	1:A:186:PRO:HD2	0.46	2.10	16	1
1:A:158:ILE:CG2	1:A:210:TYR:CD1	0.46	2.98	13	5
1:A:185:LYS:CB	1:A:190:TRP:NE1	0.46	2.75	3	1
1:A:190:TRP:CE2	1:A:204:ARG:CB	0.46	3.00	16	1
1:A:190:TRP:CE3	1:A:202:VAL:HG23	0.45	2.47	12	1
1:A:179:LEU:N	1:A:179:LEU:HD23	0.45	2.27	13	1
1:A:173:PHE:N	1:A:173:PHE:CD1	0.44	2.85	2	2
1:A:185:LYS:CD	1:A:204:ARG:HB2	0.44	2.43	5	1
1:A:180:LEU:HD13	1:A:181:VAL:N	0.44	2.28	7	1
1:A:185:LYS:CG	1:A:190:TRP:HE1	0.44	2.26	10	6
1:A:204:ARG:O	1:A:205:THR:C	0.44	2.56	12	1
1:A:185:LYS:CB	1:A:190:TRP:CD1	0.43	3.01	12	1
1:A:171:LEU:HG	1:A:200:GLY:CA	0.43	2.42	13	2
1:A:156:GLU:H	1:A:156:GLU:CD	0.43	2.16	6	1
1:A:158:ILE:HA	1:A:177:GLU:O	0.43	2.13	5	4
1:A:190:TRP:CD2	1:A:202:VAL:HG23	0.43	2.49	12	1
1:A:178:ILE:N	1:A:178:ILE:HD12	0.42	2.28	16	3
1:A:155:GLU:N	1:A:155:GLU:CD	0.42	2.73	7	1
1:A:173:PHE:CD1	1:A:173:PHE:N	0.42	2.87	11	2
1:A:185:LYS:HB3	1:A:190:TRP:NE1	0.42	2.29	5	1
1:A:181:VAL:CG1	1:A:190:TRP:HB3	0.42	2.44	8	1
1:A:190:TRP:CZ2	1:A:204:ARG:HA	0.42	2.49	13	1
1:A:190:TRP:CZ3	1:A:202:VAL:CG2	0.42	3.03	12	1
1:A:190:TRP:NE1	1:A:204:ARG:HB2	0.42	2.30	14	1
1:A:182:ILE:N	1:A:182:ILE:HD13	0.42	2.28	3	1
1:A:201:LEU:CD2	1:A:201:LEU:N	0.42	2.83	11	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:184:LYS:O	1:A:186:PRO:CD	0.42	2.68	16	1
1:A:190:TRP:CE2	1:A:204:ARG:HA	0.41	2.50	5	1
1:A:175:LYS:CG	1:A:176:GLY:N	0.41	2.83	14	1
1:A:180:LEU:HB2	1:A:193:LYS:HB2	0.41	1.90	5	1
1:A:181:VAL:CG1	1:A:190:TRP:CE3	0.41	3.03	9	1
1:A:208:GLU:CD	1:A:208:GLU:N	0.41	2.73	7	1
1:A:204:ARG:O	1:A:204:ARG:HD3	0.41	2.15	12	1
1:A:185:LYS:HB2	1:A:190:TRP:HE1	0.41	1.73	3	1
1:A:185:LYS:HG2	1:A:188:GLY:CA	0.41	2.46	11	2
1:A:191:ILE:HA	1:A:200:GLY:O	0.41	2.15	13	1
1:A:157:TYR:CE1	1:A:181:VAL:HG21	0.41	2.51	13	1
1:A:163:PHE:HB3	1:A:173:PHE:CZ	0.41	2.51	1	1
1:A:185:LYS:HB2	1:A:190:TRP:CD1	0.41	2.51	3	1
1:A:185:LYS:HD3	1:A:190:TRP:NE1	0.41	2.30	5	1
1:A:183:GLU:O	1:A:190:TRP:CD2	0.41	2.74	17	1
1:A:157:TYR:CG	1:A:207:LEU:HD23	0.40	2.51	12	1
1:A:204:ARG:O	1:A:206:TYR:N	0.40	2.55	15	1
1:A:185:LYS:HG3	1:A:190:TRP:HE1	0.40	1.76	11	1
1:A:167:GLN:O	1:A:168:VAL:C	0.40	2.60	16	1
1:A:190:TRP:CE3	1:A:192:ALA:HB2	0.40	2.51	12	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	57/68 (84%)	42±2 (74±3%)	10±2 (17±3%)	5±1 (8±2%)	2	13
All	All	969/1156 (84%)	720 (74%)	169 (17%)	80 (8%)	2	13

All 13 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	183	GLU	17
1	A	185	LYS	17

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	205	THR	14
1	A	175	LYS	11
1	A	186	PRO	5
1	A	168	VAL	4
1	A	173	PHE	4
1	A	154	GLY	3
1	A	169	GLY	1
1	A	197	GLY	1
1	A	198	ASN	1
1	A	182	ILE	1
1	A	201	LEU	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	46/56 (82%)	42±2 (92±4%)	4±2 (8±4%)	15 62
All	All	782/952 (82%)	718 (92%)	64 (8%)	15 62

All 22 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	199	GLU	14
1	A	207	LEU	10
1	A	204	ARG	6
1	A	203	PRO	6
1	A	182	ILE	3
1	A	198	ASN	3
1	A	185	LYS	3
1	A	180	LEU	3
1	A	158	ILE	2
1	A	177	GLU	2
1	A	157	TYR	1
1	A	156	GLU	1
1	A	178	ILE	1
1	A	183	GLU	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	167	GLN	1
1	A	168	VAL	1
1	A	201	LEU	1
1	A	166	GLN	1
1	A	190	TRP	1
1	A	179	LEU	1
1	A	170	ASP	1
1	A	186	PRO	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 monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

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)

No chemical shift data were provided