



# Full wwPDB NMR Structure Validation Report ⓘ

May 29, 2020 – 08:37 am BST

PDB ID : 5VKV  
Title : Solution NMR structure of the membrane electron transporter CcdA  
Authors : Zhou, Y.; Bushweller, J.H.  
Deposited on : 2017-04-24

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto: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 ⓘ symbol.

---

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

Cyrange : Kirchner and Güntert (2011)  
NmrClust : Kelley et al. (1996)  
MolProbity : 4.02b-467  
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)  
RCI : v\_1n\_11\_5\_13\_A (Berjanski et al., 2005)  
PANAV : Wang et al. (2010)  
ShiftChecker : 2.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

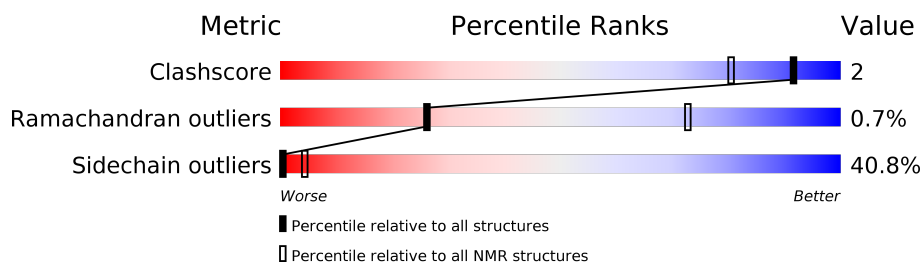
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 33%.

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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	232	

## 2 Ensemble composition and analysis

This entry contains 10 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:2-A:94, A:107-A:220 (207)	1.14	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 2 single-model clusters were found.

Cluster number	Models
1	1, 2, 4, 6, 7, 9
2	3, 5
Single-model clusters	8; 10

### 3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 3531 atoms, of which 1814 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Cytochrome c-type biogenesis protein CcdA.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	224	3531	1165	1814	279	269	4	0

There are 11 discrepancies between the modelled and reference sequences:

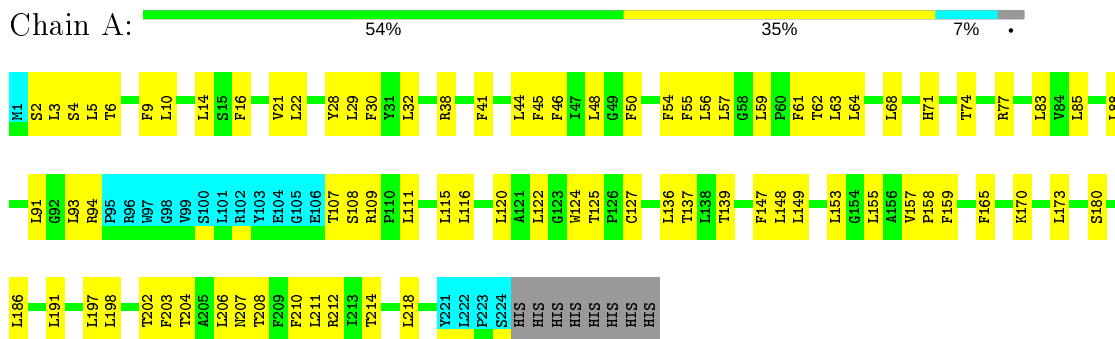
Chain	Residue	Modelled	Actual	Comment	Reference
A	35	ALA	GLU	engineered mutation	UNP Q5SIG0
A	223	PRO	-	expression tag	UNP Q5SIG0
A	224	SER	-	expression tag	UNP Q5SIG0
A	225	HIS	-	expression tag	UNP Q5SIG0
A	226	HIS	-	expression tag	UNP Q5SIG0
A	227	HIS	-	expression tag	UNP Q5SIG0
A	228	HIS	-	expression tag	UNP Q5SIG0
A	229	HIS	-	expression tag	UNP Q5SIG0
A	230	HIS	-	expression tag	UNP Q5SIG0
A	231	HIS	-	expression tag	UNP Q5SIG0
A	232	HIS	-	expression tag	UNP Q5SIG0

## 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: Cytochrome c-type biogenesis protein CcdA

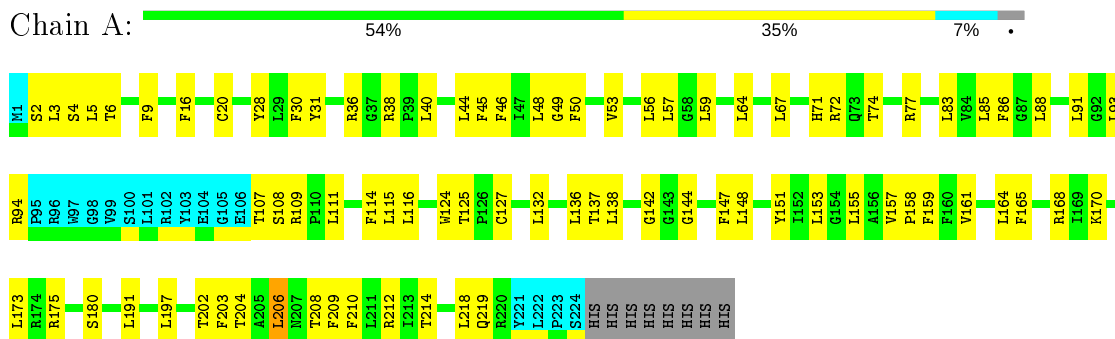


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

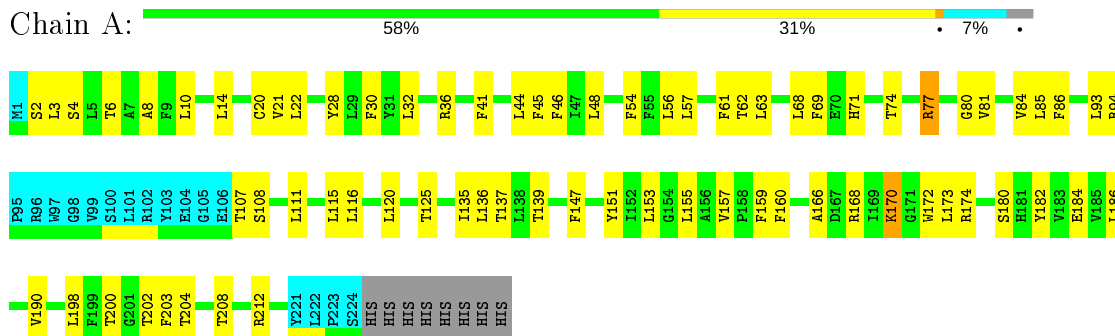
#### 4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Cytochrome c-type biogenesis protein CcdA



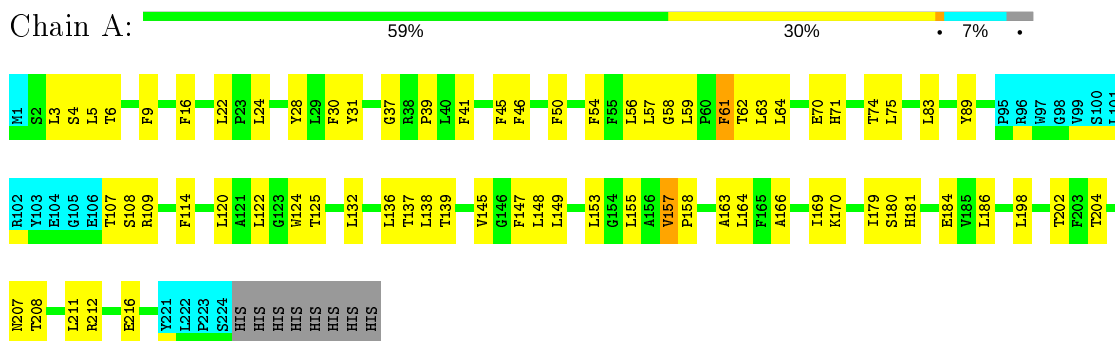
### 4.2.2 Score per residue for model 2

- Molecule 1: Cytochrome c-type biogenesis protein CcdA



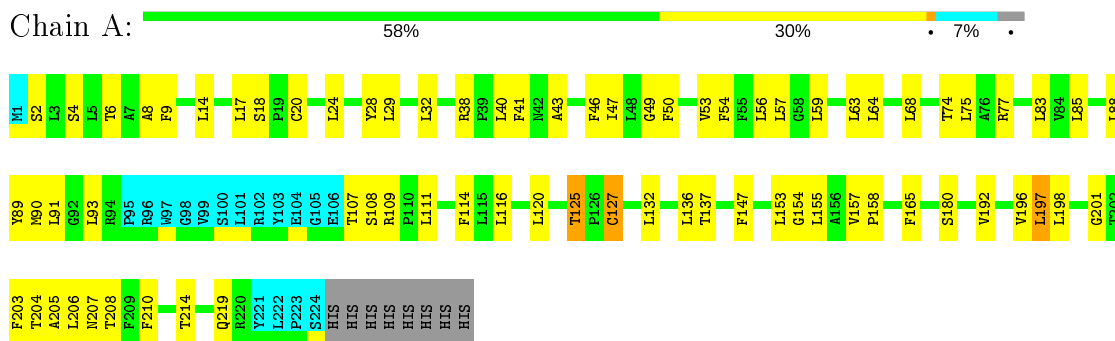
### 4.2.3 Score per residue for model 3

- Molecule 1: Cytochrome c-type biogenesis protein CcdA



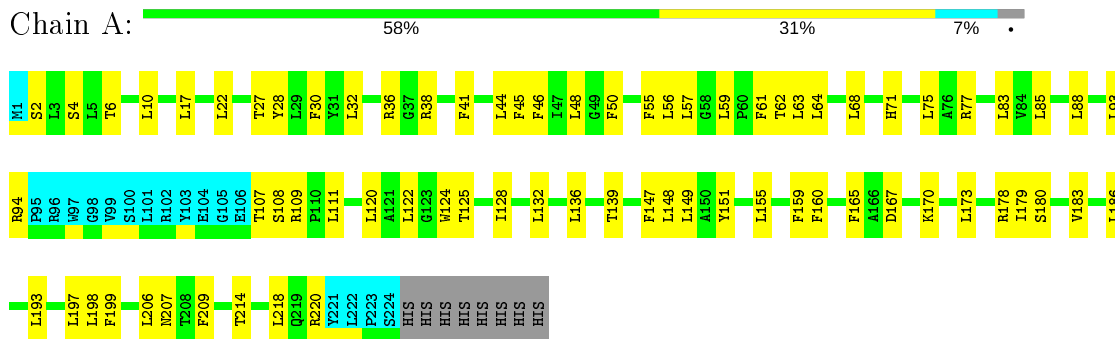
### 4.2.4 Score per residue for model 4

- Molecule 1: Cytochrome c-type biogenesis protein CcdA



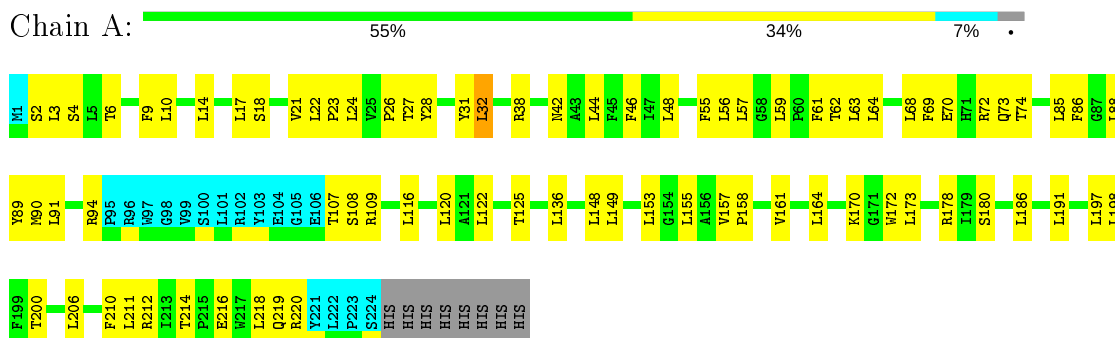
### 4.2.5 Score per residue for model 5

- Molecule 1: Cytochrome c-type biogenesis protein CcdA



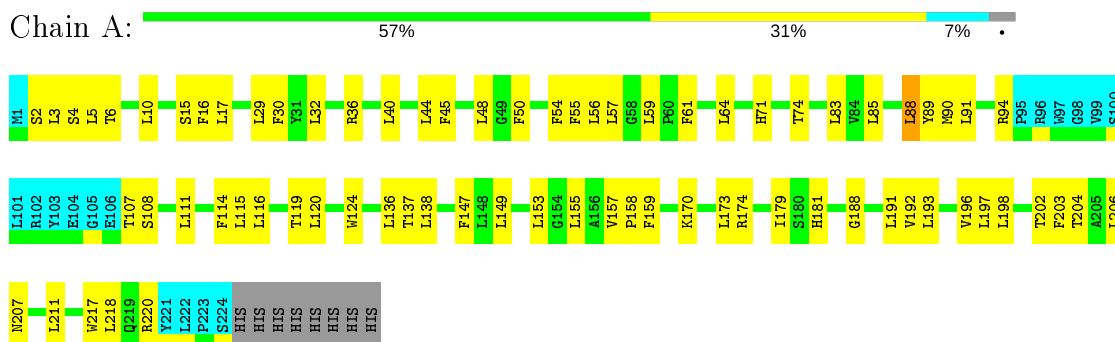
### 4.2.6 Score per residue for model 6

- Molecule 1: Cytochrome c-type biogenesis protein CcdA



### 4.2.7 Score per residue for model 7

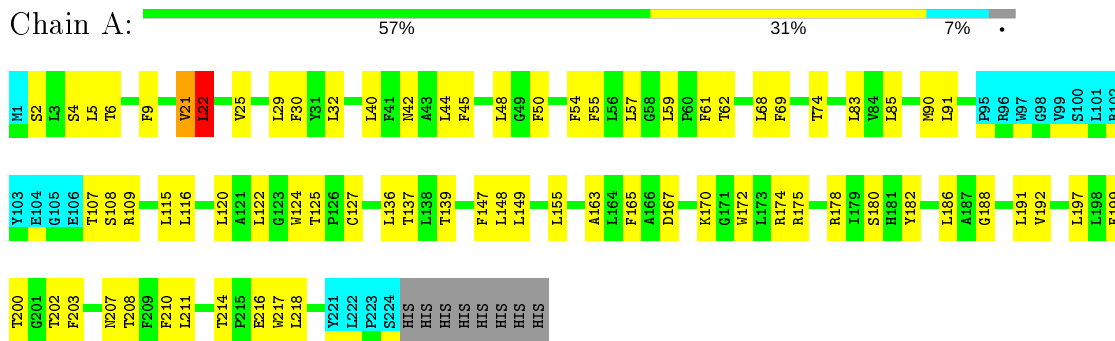
- Molecule 1: Cytochrome c-type biogenesis protein CcdA



### 4.2.8 Score per residue for model 8

- Molecule 1: Cytochrome c-type biogenesis protein CcdA

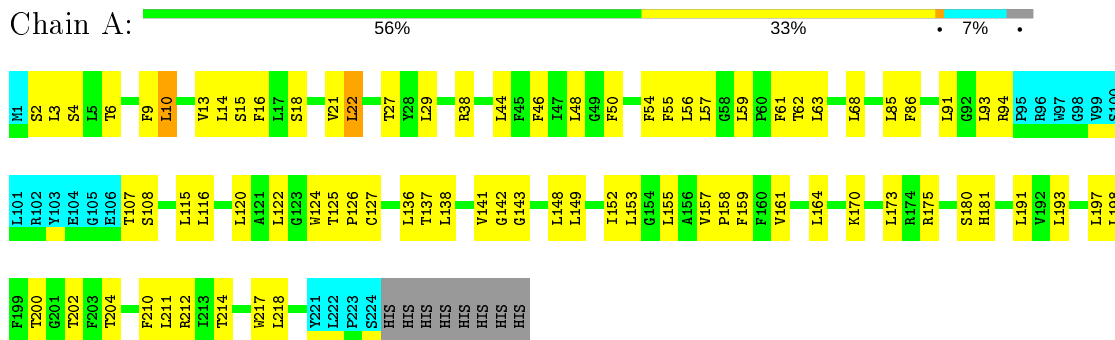
Chain A:



### 4.2.9 Score per residue for model 9

- Molecule 1: Cytochrome c-type biogenesis protein CcdA

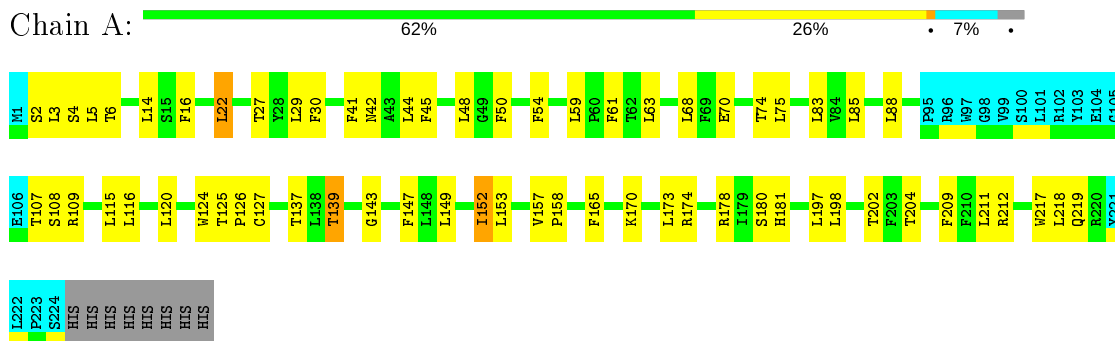
Chain A:



### 4.2.10 Score per residue for model 10

- Molecule 1: Cytochrome c-type biogenesis protein CcdA

Chain A:





## 5 Refinement protocol and experimental data overview

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

Of the 200 calculated structures, 10 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
XPLOR-NIH	refinement	2.39
X-PLOR NIH	structure calculation	2.39

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)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	881
Number of shifts mapped to atoms	881
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	33%

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.

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

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	1574	1676	1676	5±2
All	All	15740	16760	16760	49

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:157:VAL:HG13	1:A:158:PRO:HD3	0.59	1.74	3	1
1:A:15:SER:HB2	1:A:155:LEU:HD13	0.58	1.76	9	1
1:A:88:LEU:HA	1:A:91:LEU:HD12	0.57	1.75	7	1
1:A:204:THR:HG23	1:A:206:LEU:H	0.50	1.66	1	1
1:A:23:PRO:O	1:A:26:PRO:HD2	0.50	2.07	6	1
1:A:201:GLY:HA2	1:A:205:ALA:HB3	0.48	1.84	4	1
1:A:49:GLY:O	1:A:53:VAL:HG23	0.48	2.09	1	2
1:A:31:TYR:O	1:A:37:GLY:HA3	0.48	2.09	3	1
1:A:39:PRO:HB3	1:A:163:ALA:HB1	0.47	1.86	3	1
1:A:139:THR:HA	1:A:143:GLY:O	0.47	2.09	10	1
1:A:166:ALA:O	1:A:170:LYS:HB2	0.47	2.09	2	1
1:A:125:THR:HB	1:A:127:CYS:SG	0.46	2.50	4	1
1:A:8:ALA:HB1	1:A:151:TYR:HA	0.46	1.87	2	1
1:A:166:ALA:HA	1:A:169:ILE:HG22	0.46	1.87	3	1

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:10:LEU:HG	1:A:218:LEU:HD21	0.46	1.88	9	1
1:A:43:ALA:O	1:A:47:ILE:HG12	0.46	2.11	4	1
1:A:197:LEU:O	1:A:201:GLY:HA3	0.45	2.11	4	1
1:A:163:ALA:O	1:A:167:ASP:HB2	0.45	2.11	8	1
1:A:124:TRP:O	1:A:126:PRO:HD3	0.45	2.11	9	1
1:A:157:VAL:HB	1:A:158:PRO:CD	0.45	2.41	9	5
1:A:198:LEU:HD23	1:A:206:LEU:HD11	0.44	1.87	4	1
1:A:29:LEU:HD21	1:A:116:LEU:HD22	0.44	1.90	8	1
1:A:179:ILE:O	1:A:183:VAL:HG23	0.44	2.12	5	1
1:A:216:GLU:HA	1:A:219:GLN:OE1	0.44	2.12	6	1
1:A:186:LEU:O	1:A:190:VAL:HG23	0.43	2.12	2	1
1:A:59:LEU:O	1:A:62:THR:HG22	0.43	2.13	9	1
1:A:157:VAL:O	1:A:161:VAL:HG23	0.43	2.14	6	3
1:A:80:GLY:O	1:A:84:VAL:HG23	0.42	2.14	2	1
1:A:192:VAL:O	1:A:196:VAL:HG23	0.42	2.15	7	2
1:A:155:LEU:C	1:A:158:PRO:HD2	0.42	2.34	3	1
1:A:77:ARG:O	1:A:81:VAL:HG23	0.42	2.14	2	1
1:A:58:GLY:O	1:A:61:PHE:HB2	0.41	2.14	3	1
1:A:32:LEU:HG	1:A:173:LEU:HD21	0.41	1.92	6	1
1:A:21:VAL:HG12	1:A:22:LEU:H	0.41	1.75	8	1
1:A:149:LEU:O	1:A:152:ILE:HG13	0.41	2.15	10	1
1:A:125:THR:O	1:A:128:ILE:HG22	0.40	2.15	5	1
1:A:138:LEU:O	1:A:141:VAL:HG12	0.40	2.17	9	1
1:A:8:ALA:HB1	1:A:154:GLY:HA3	0.40	1.94	4	1
1:A:157:VAL:HB	1:A:158:PRO:HD3	0.40	1.93	7	1
1:A:188:GLY:O	1:A:192:VAL:HG23	0.40	2.17	7	2

## 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	207/232 (89%)	198±2 (95±1%)	8±2 (4±1%)	1±1 (1±1%)	26	73
All	All	2070/2320 (89%)	1975 (95%)	81 (4%)	14 (1%)	26	73

All 9 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	22	LEU	3
1	A	203	PHE	2
1	A	142	GLY	2
1	A	21	VAL	2
1	A	38	ARG	1
1	A	144	GLY	1
1	A	126	PRO	1
1	A	143	GLY	1
1	A	124	TRP	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	160/183 (87%)	95±5 (59±3%)	65±5 (41±3%)	<b>0</b> <b>4</b>
All	All	1600/1830 (87%)	947 (59%)	653 (41%)	<b>0</b> <b>4</b>

All 130 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	108	SER	10
1	A	107	THR	10
1	A	4	SER	10
1	A	6	THR	10
1	A	120	LEU	9
1	A	180	SER	9
1	A	170	LYS	9
1	A	85	LEU	9
1	A	57	LEU	9
1	A	2	SER	9
1	A	136	LEU	9
1	A	74	THR	8
1	A	147	PHE	8
1	A	197	LEU	8
1	A	61	PHE	8

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	50	PHE	8
1	A	44	LEU	8
1	A	153	LEU	8
1	A	125	THR	8
1	A	48	LEU	8
1	A	59	LEU	8
1	A	137	THR	8
1	A	56	LEU	8
1	A	116	LEU	7
1	A	68	LEU	7
1	A	83	LEU	7
1	A	155	LEU	7
1	A	198	LEU	7
1	A	54	PHE	7
1	A	202	THR	7
1	A	45	PHE	7
1	A	22	LEU	7
1	A	63	LEU	7
1	A	46	PHE	7
1	A	30	PHE	7
1	A	3	LEU	7
1	A	109	ARG	7
1	A	64	LEU	6
1	A	212	ARG	6
1	A	218	LEU	6
1	A	214	THR	6
1	A	9	PHE	6
1	A	211	LEU	6
1	A	204	THR	6
1	A	115	LEU	6
1	A	149	LEU	6
1	A	32	LEU	6
1	A	88	LEU	6
1	A	148	LEU	6
1	A	28	TYR	6
1	A	173	LEU	6
1	A	94	ARG	6
1	A	210	PHE	5
1	A	91	LEU	5
1	A	16	PHE	5
1	A	111	LEU	5
1	A	93	LEU	5

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	207	ASN	5
1	A	10	LEU	5
1	A	122	LEU	5
1	A	5	LEU	5
1	A	55	PHE	5
1	A	165	PHE	5
1	A	139	THR	5
1	A	127	CYS	5
1	A	41	PHE	5
1	A	71	HIS	5
1	A	159	PHE	5
1	A	14	LEU	5
1	A	191	LEU	5
1	A	62	THR	5
1	A	208	THR	5
1	A	124	TRP	5
1	A	181	HIS	4
1	A	89	TYR	4
1	A	40	LEU	4
1	A	86	PHE	4
1	A	90	MET	4
1	A	29	LEU	4
1	A	132	LEU	4
1	A	36	ARG	4
1	A	178	ARG	4
1	A	27	THR	4
1	A	77	ARG	4
1	A	206	LEU	4
1	A	17	LEU	4
1	A	75	LEU	4
1	A	174	ARG	4
1	A	186	LEU	4
1	A	114	PHE	4
1	A	200	THR	4
1	A	217	TRP	4
1	A	38	ARG	4
1	A	164	LEU	4
1	A	209	PHE	3
1	A	203	PHE	3
1	A	175	ARG	3
1	A	24	LEU	3
1	A	220	ARG	3

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	70	GLU	3
1	A	138	LEU	3
1	A	42	ASN	3
1	A	69	PHE	3
1	A	172	TRP	3
1	A	219	GLN	3
1	A	18	SER	3
1	A	20	CYS	3
1	A	193	LEU	3
1	A	216	GLU	2
1	A	21	VAL	2
1	A	184	GLU	2
1	A	199	PHE	2
1	A	157	VAL	2
1	A	151	TYR	2
1	A	160	PHE	2
1	A	179	ILE	2
1	A	31	TYR	2
1	A	72	ARG	2
1	A	152	ILE	2
1	A	182	TYR	2
1	A	168	ARG	2
1	A	145	VAL	1
1	A	73	GLN	1
1	A	67	LEU	1
1	A	135	ILE	1
1	A	15	SER	1
1	A	167	ASP	1
1	A	25	VAL	1
1	A	119	THR	1
1	A	13	VAL	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](#)

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](#)

The completeness of assignment taking into account all chemical shift lists is 33% for the well-defined parts and 32% for the entire structure.

### 7.1 Chemical shift list 1

File name: input\_cs.cif

Chemical shift list name: *CcdA24\_1\_NMRStar.str.txt*

#### 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	881
Number of shifts mapped to atoms	881
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

#### 7.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	205	$-0.16 \pm 0.15$	None needed ( $< 0.5$ ppm)
$^{13}\text{C}_\beta$	114	$1.35 \pm 0.05$	Should be applied
$^{13}\text{C}'$	191	$0.01 \pm 0.07$	None needed ( $< 0.5$ ppm)
$^{15}\text{N}$	182	$1.07 \pm 0.22$	Should be applied

#### 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 33%, i.e. 821 atoms were assigned a chemical shift out of a possible 2506. 0 out of 67 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	$^1\text{H}$	$^{13}\text{C}$	$^{15}\text{N}$
Backbone	711/1015 (70%)	171/404 (42%)	370/414 (89%)	170/197 (86%)
Sidechain	106/1212 (9%)	0/699 (0%)	106/472 (22%)	0/41 (0%)

*Continued on next page...*

Continued from previous page...

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Aromatic	4/279 (1%)	2/151 (1%)	0/123 (0%)	2/5 (40%)
Overall	821/2506 (33%)	173/1254 (14%)	476/1009 (47%)	172/243 (71%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 32%, i.e. 881 atoms were assigned a chemical shift out of a possible 2730. 0 out of 70 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	761/1096 (69%)	183/436 (42%)	396/448 (88%)	182/212 (86%)
Sidechain	114/1327 (9%)	0/769 (0%)	114/511 (22%)	0/47 (0%)
Aromatic	6/307 (2%)	3/165 (2%)	0/136 (0%)	3/6 (50%)
Overall	881/2730 (32%)	186/1370 (14%)	510/1095 (47%)	185/265 (70%)

#### 7.1.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

#### 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:

