wwPDB NMR Structure Validation Summary Report (i)

Jun 11, 2024 – 06:34 PM EDT

PDB ID	:	1QW1
Title	:	Solution Structure of the C-Terminal Domain of DtxR residues 110-226
Authors	:	Wylie, G.P.; Rangachari, V.; Bienkiewicz, E.A.; Love, J.F.; Murphy, J.R.; Logan, T.M.
Deposited on	:	2003-08-29

This is a wwPDB NMR Structure Validation Summary 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:

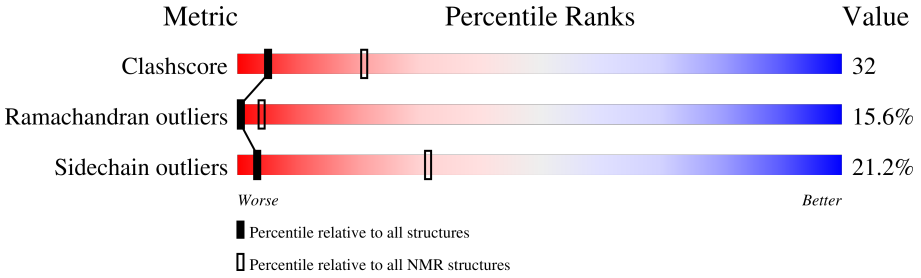
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

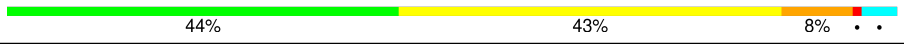
The overall completeness of chemical shifts assignment was not calculated.

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	121	

2 Ensemble composition and analysis (i)

This entry contains 14 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:111-A:226 (116)	1.86	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. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1874 atoms, of which 938 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Diphtheria toxin repressor.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	121	1874	577	938	168	189	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	106	GLY	-	cloning artifact	UNP P33120
A	107	SER	-	cloning artifact	UNP P33120
A	108	HIS	-	cloning artifact	UNP P33120
A	109	MET	-	cloning artifact	UNP P33120

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: Diphtheria toxin repressor

Chain A: 44% 43% 8%

G106 S107 H108 M109 D110 E111 V112 E113 L120 N130 P131 I132 P133 G134 L135 D136 E137 L138 G139 V140 A146 A147 P148 G149 T150 R151 V152 I153 D154 A155 A156 T157 S158 V163 R164 I165 V166 Q167 I168 N169 E170 I171 F172 Q173 V174 L183 D184 A185 D186 I187 R188 V189 V193 E194

I195 V196 I202 T203 L204 S205 H206 N207 G208 K209 D210 V211 E212 L213 L214 D215 D216 L217 A218 H219 T220 I221 R222 I223 E224 E225 L226

4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: Diphtheria toxin repressor

Chain A: 45% 34% 16%

G106 S107 H108 M109 D110 E111 V112 E113 V119 L120 V123 F128 G129 N130 P131 I132 P133 G134 L135 D136 E137 L138 G139 V140 G141 N142 S143 D144 A147 P148 G149 T150 R151 V152 I153 D154 A155 A156 T157 S158 M159 P160 I165 V166 Q167 I168 N169 E170 I171 F172 Q173 V174 L183 D184

A185 D186 I187 R188 V189 G190 V193 E194 I195 V196 I202 H206 N207 G208 K209 D210 V211 D216 L217 A218 H219 T220 I221 R222 I223 E224 E225 L226

5 Refinement protocol and experimental data overview (i)

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

Of the 60 calculated structures, 14 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
CNS	refinement	1.0

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	900	907	904	59±12
All	All	12600	12698	12656	820

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 32.

5 of 569 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:152:VAL:HG11	1:A:218:ALA:HB1	1.07	1.26	10	2
1:A:163:VAL:HG13	1:A:195:ILE:HD11	0.93	1.38	12	1
1:A:147:ALA:HB1	1:A:148:PRO:HD2	0.83	1.50	10	4
1:A:165:ILE:HD11	1:A:168:ILE:HG12	0.83	1.48	12	1
1:A:152:VAL:HG22	1:A:223:ILE:CD1	0.83	2.03	7	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	115/121 (95%)	68±4 (59±4%)	29±4 (25±4%)	18±3 (16±3%)	0 4
All	All	1610/1694 (95%)	957 (59%)	402 (25%)	251 (16%)	0 4

5 of 66 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	174	VAL	14
1	A	187	ILE	13
1	A	132	ILE	11
1	A	154	ASP	11
1	A	207	ASN	10

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	101/105 (96%)	80±3 (79±3%)	21±3 (21±3%)	3 31
All	All	1414/1470 (96%)	1114 (79%)	300 (21%)	3 31

5 of 76 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	132	ILE	12
1	A	184	ASP	12
1	A	210	ASP	11
1	A	165	ILE	10
1	A	171	ILE	10

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