wwPDB NMR Structure Validation Summary Report (i)

Feb 17, 2022 – 10:38 AM EST

PDB ID	:	1OVQ
Title	:	Solution structure of the hypothetical protein YqgF from Escherichia coli
Authors	:	Liu, D.; Wang, Y.S.; Wyss, D.F.
Deposited on	:	2003-03-27

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

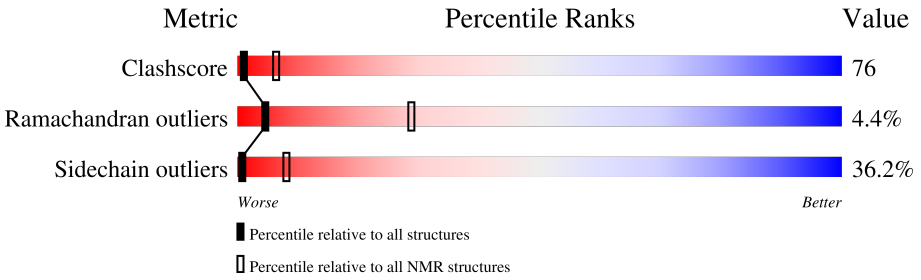
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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	138	13% 39% 12% • 34%

2 Ensemble composition and analysis (i)

This entry contains 25 models. Model 18 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:3-A:61, A:76-A:94, A:123-A:135 (91)	0.35	18

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 5 clusters and 2 single-model clusters were found.

Cluster number	Models
1	3, 7, 10, 12, 14, 15, 16, 18, 23
2	4, 5, 9, 11
3	6, 13, 17
4	21, 22, 24, 25
5	2, 19, 20
Single-model clusters	1; 8

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2151 atoms, of which 1079 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hypothetical protein yqgF.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	138	2151	678	1079	191	201	2	0

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: Hypothetical protein yqgF

Chain A: 13% 39% 12% • 34%

M1 S2 G3 T4 L5 L6 A7 F8 D9 F10 G11 T12 K13 S14 I15 G16 V17 A18 V19 G20 Q21 R22 I23 T24 G25 T26 A27 R28 P29 L30 P31 A32 I33 K34 A35 Q36 D37 G38 T39 P40 D41 W42 N43 I44 I45 E46 R47 L48 L49 Q53 P54 D55 E56 I57 I58 V59 G60 L61 P62

L63 N64 M65 D66 G67 T68 E69 Q70 P71 L72 T73 A74 R75 A76 R77 K78 F79 R82 I83 H84 G85 R86 F87 G88 V89 E90 V91 K92 L93 H94 D95 E96 R97 L98 S99 T100 V101 E102 A103 R104 S105 G106 L107 F108 E109 Q110 G111 G112 Y113 R114 A115 L116 N117 K118 G119 K120 V121 D122 S123

A124 S125 A126 V127 I128 I129 L130 E131 S132 Y133 F134 E135 Q136 G137 Y138

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

The representative model is number 18. Colouring as in section 4.1 above.

• Molecule 1: Hypothetical protein yqgF

Chain A: 21% 30% 14% • 34%

M1 S2 G3 T4 L5 L6 A7 F8 D9 F10 G11 S14 I15 G16 V17 A18 V19 G20 Q21 R22 I23 T26 A27 R28 P29 L30 P31 A32 I33 K34 A35 Q36 T39 P40 D41 W42 N43 I44 I45 E46 R47 L48 L49 K50 E51 W52 Q53 P54 D55 E56 I57 I58 V59 G60 L61 P62 L63

N64 M65 D66 G67 T68 E69 Q70 P71 L72 T73 A74 R75 A76 R77 N81 R82 I83 H84 G85 R86 V89 E90 V91 K92 D95 E96 R97 L98 S99 T100 V101 E102 A103 R104 S105 G106 L107 F108 E109 Q110 G111 G112 Y113 R114 A115 L116 N117 K118 G119 K120 V121 D122 A126 V127 I128 I129

L130 E131 S132 Y133 F134 E135 Q136 G137 Y138

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 25 were deposited, based on the following criterion: *back calculated data agree with experimental NOESY spectrum*.

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

Software name	Classification	Version
X-PLOR	structure solution	98
X-PLOR	refinement	98

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	1.05±0.00	0±0/725 (0.0± 0.0%)	1.28±0.01	0±0/984 (0.0± 0.0%)
All	All	1.05	0/18125 (0.0%)	1.28	2/24600 (0.0%)

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	6.0±0.0
All	All	0	150

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	14	SER	N-CA-CB	-5.47	102.29	110.50	22	2

There are no chirality outliers.

5 of 6 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	22	ARG	Sidechain	25
1	A	28	ARG	Sidechain	25
1	A	47	ARG	Sidechain	25
1	A	77	ARG	Sidechain	25
1	A	82	ARG	Sidechain	25

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	709	726	726	109±8
All	All	17725	18150	18150	2717

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 76.

5 of 471 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:6:LEU:HD21	1:A:48:LEU:HD13	1.12	1.20	19	15
1:A:6:LEU:HD11	1:A:19:VAL:HG13	1.07	1.25	14	16
1:A:5:LEU:HD21	1:A:58:ILE:HD11	1.04	1.20	22	16
1:A:5:LEU:HD22	1:A:133:TYR:CE2	0.97	1.95	13	7
1:A:83:ILE:HD11	1:A:91:VAL:HG21	0.96	1.33	24	15

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	91/138 (66%)	80±2 (88±2%)	7±1 (8±1%)	4±1 (4±1%)	4 29
All	All	2275/3450 (66%)	1996 (88%)	180 (8%)	99 (4%)	4 29

5 of 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	31	PRO	25
1	A	33	ILE	25
1	A	53	GLN	17

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	3	GLY	15
1	A	12	THR	6

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	73/111 (66%)	47±3 (64±4%)	26±3 (36±4%)	1 8
All	All	1825/2775 (66%)	1165 (64%)	660 (36%)	1 8

5 of 56 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	5	LEU	25
1	A	23	ILE	25
1	A	26	THR	25
1	A	33	ILE	25
1	A	58	ILE	25

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