Full wwPDB NMR Structure Validation Report (i)

Mar 3, 2022 – 09:58 AM EST

PDB ID	:	2IGG
Title	:	DETERMINATION OF THE SOLUTION STRUCTURES OF DOMAINS II AND III OF PROTEIN G FROM STREPTOCOCCUS BY 1H NMR
Authors	:	Lian, L.-Y.; Derrick, J.P.; Sutcliffe, M.J.; Yang, J.C.; Roberts, G.C.K.
Deposited on	:	1992-08-26

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

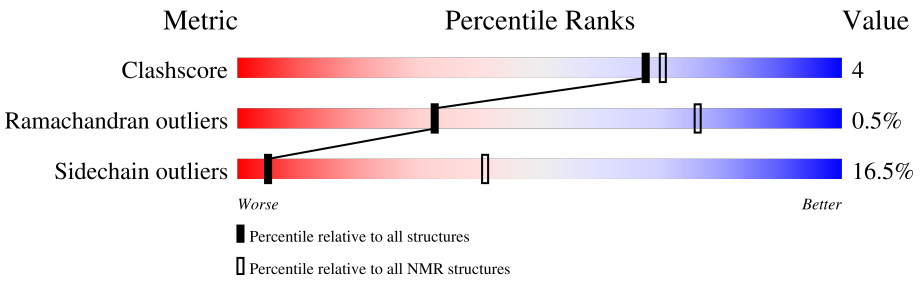
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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	64	59% 11% • 28%

2 Ensemble composition and analysis (i)

This entry contains 27 models. Model 8 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:6-A:13, A:19-A:51, A:56-A:60 (46)	0.30	8

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 4 clusters and 2 single-model clusters were found.

Cluster number	Models
1	1, 3, 4, 5, 7, 8, 11, 13, 14, 15, 18, 19, 20, 22, 23, 24, 25, 26
2	6, 10, 21
3	2, 27
4	12, 16
Single-model clusters	9; 17

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 973 atoms, of which 480 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PROTEIN G.

Mol	Chain	Residues	Atoms					Trace
			Total	C	H	N	O	
1	A	64	973	310	480	76	107	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: PROTEIN G

Chain A: 59% 11% • 28%

L1 T2 P3 A4 V5 L10 G14 K15 T16 L17 K18 T23 F35 W48 T49 Y50 D51 D52 A53 T54 K55 T56 F57 T58 V59 T60 E61 K62 P63 E64

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: PROTEIN G

Chain A: 48% 20% • • 28%

L1 T2 P3 A4 V5 V11 I12 N13 G14 K15 T16 L17 K18 G19 E20 T23 F35 N42 D45 G46 E47 W48 T49 Y50 D51 D52 A53 T54 K55 T56 F57 T58 V59 T60 E61 K62 P63 E64

4.2.2 Score per residue for model 2

• Molecule 1: PROTEIN G

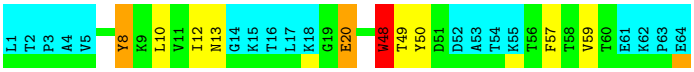
Chain A: 56% 12% • • 28%

L1 T2 P3 A4 V5 L10 G14 K15 T16 L17 K18 T23 E24 F35 W48 T49 Y50 D51 D52 A53 T54 K55 T56 F57 T58 V59 T60 E61 K62 P63 E64

4.2.3 Score per residue for model 3

- Molecule 1: PROTEIN G

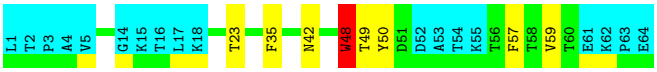
Chain A: 56% 11% 28%



4.2.4 Score per residue for model 4

- Molecule 1: PROTEIN G

Chain A: 59% 11% 28%



4.2.5 Score per residue for model 5

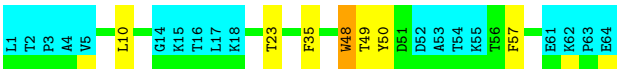
- Molecule 1: PROTEIN G

Chain A: 58% 12% 28%

4.2.6 Score per residue for model 6

- Molecule 1: PROTEIN G

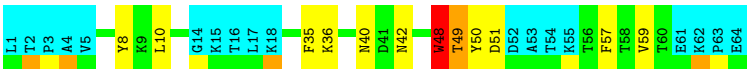
Chain A: 61% 9% 28%



4.2.7 Score per residue for model 7

- Molecule 1: PROTEIN G

Chain A: 53% 16% 28%



4.2.8 Score per residue for model 8 (medoid)

- Molecule 1: PROTEIN G

Chain A: 55% 14% • • 28%

L1 T2 P3 A4 V5 Y8 N13 G14 K15 T16 L17 K18 T23 F35 K36 Q37 Y38 W48 T49 Y50 D51 D52 A53 T54 K55 T56 F57 T58 V59 T60 E61 K62 P63 E64

4.2.9 Score per residue for model 9

- Molecule 1: PROTEIN G

Chain A: 58% 12% • 28%

L1 T2 P3 A4 V5 L10 G14 K15 T16 L17 K18 K36 D45 G46 E47 W48 T49 Y50 D51 D52 A53 T54 K55 T56 F57 T58 E61 K62 P63 E64

4.2.10 Score per residue for model 10

- Molecule 1: PROTEIN G

Chain A: 52% 16% • • 28%

L1 T2 P3 A4 V5 L10 V11 G14 K15 T16 L17 K18 G19 E20 T23 E24 F35 N40 W48 T49 Y50 D51 D52 A53 T54 K55 T56 F57 T58 V59 T60 E61 K62 P63 E64

4.2.11 Score per residue for model 11

- Molecule 1: PROTEIN G

Chain A: 56% 14% • 28%

L1 T2 P3 A4 V5 G14 K15 T16 L17 K18 T23 F35 D45 G46 E47 W48 T49 Y50 D51 D52 A53 T54 K55 T56 F57 T58 E61 K62 P63 E64

4.2.12 Score per residue for model 12

- Molecule 1: PROTEIN G

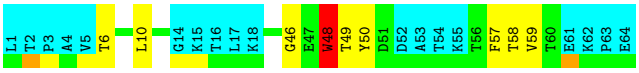
Chain A: 62% 8% • 28%

L1 T2 P3 A4 V5 G14 K15 T16 L17 K18 W48 T49 Y50 D51 D52 A53 T54 K55 T56 F57 T58 V59 T60 E61 K62 P63 E64

4.2.13 Score per residue for model 13

- Molecule 1: PROTEIN G

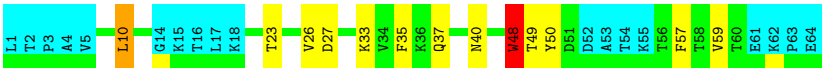
Chain A: 58% 12% • 28%



4.2.14 Score per residue for model 14

- Molecule 1: PROTEIN G

Chain A: 52% 17% • • 28%



4.2.15 Score per residue for model 15

- Molecule 1: PROTEIN G

Chain A: 56% 12% • 28%

4.2.16 Score per residue for model 16

- Molecule 1: PROTEIN G

Chain A: 56% 11% • • 28%

4.2.17 Score per residue for model 17

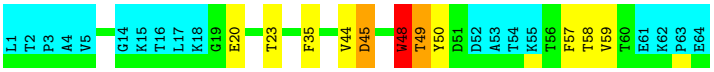
- Molecule 1: PROTEIN G

Chain A: 55% 16% • 28%

4.2.18 Score per residue for model 18

- Molecule 1: PROTEIN G

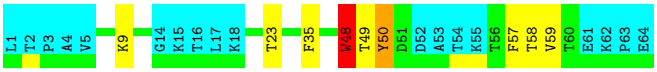
Chain A: 55% 12% • • 28%



4.2.19 Score per residue for model 19

- Molecule 1: PROTEIN G

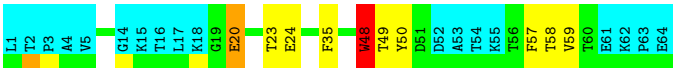
Chain A: 58% 11% • • 28%



4.2.20 Score per residue for model 20

- Molecule 1: PROTEIN G

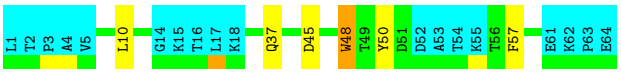
Chain A: 56% 12% • • 28%



4.2.21 Score per residue for model 21

- Molecule 1: PROTEIN G

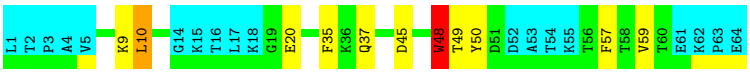
Chain A: 62% 8% • 28%



4.2.22 Score per residue for model 22

- Molecule 1: PROTEIN G

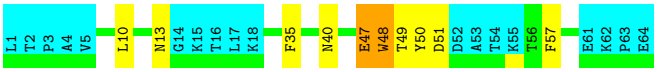
Chain A: 55% 14% • • 28%



4.2.23 Score per residue for model 23

- Molecule 1: PROTEIN G

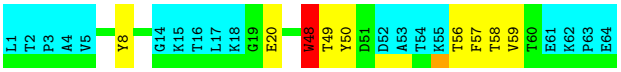
Chain A: 56% 12% • 28%



4.2.24 Score per residue for model 24

- Molecule 1: PROTEIN G

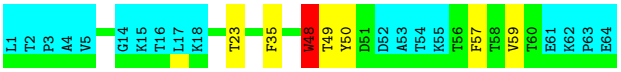
Chain A: 58% 12% • 28%



4.2.25 Score per residue for model 25

- Molecule 1: PROTEIN G

Chain A: 61% 9% • 28%



4.2.26 Score per residue for model 26

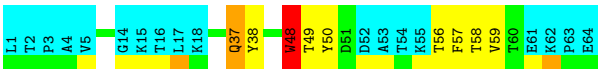
- Molecule 1: PROTEIN G

Chain A: 56% 14% • 28%

4.2.27 Score per residue for model 27

- Molecule 1: PROTEIN G

Chain A: 58% 11% • • 28%



5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 27 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
DIANA	refinement	
X-PLOR	refinement	

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	1.26±0.01	0±0/365 (0.0± 0.0%)	1.39±0.03	4±1/501 (0.9± 0.2%)
All	All	1.26	0/9855 (0.0%)	1.39	119/13527 (0.9%)

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	48	TRP	CD1-NE1-CE2	8.43	116.59	109.00	3	27
1	A	48	TRP	CB-CG-CD1	-7.79	116.88	127.00	3	27
1	A	8	TYR	CB-CG-CD1	-7.31	116.62	121.00	3	5
1	A	48	TRP	CG-CD1-NE1	-6.61	103.49	110.10	10	27
1	A	48	TRP	NE1-CE2-CZ2	6.01	137.02	130.40	5	23
1	A	8	TYR	CB-CG-CD2	5.80	124.48	121.00	3	1
1	A	50	TYR	CB-CG-CD2	-5.39	117.77	121.00	1	1
1	A	49	THR	CA-CB-CG2	-5.29	104.99	112.40	7	3
1	A	47	GLU	N-CA-C	-5.26	96.80	111.00	9	3
1	A	48	TRP	CB-CG-CD2	5.21	133.37	126.60	3	1
1	A	48	TRP	NE1-CE2-CD2	-5.04	102.27	107.30	20	1

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	358	331	331	3±1
All	All	9666	8937	8937	73

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 4.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:48:TRP:CZ3	1:A:59:VAL:HB	0.55	2.37	24	7
1:A:48:TRP:CE3	1:A:59:VAL:HB	0.54	2.38	24	20
1:A:10:LEU:HD23	1:A:10:LEU:C	0.51	2.25	14	1
1:A:23:THR:HB	1:A:35:PHE:CE1	0.49	2.42	14	6
1:A:23:THR:HG23	1:A:35:PHE:CE1	0.48	2.42	10	1
1:A:10:LEU:HB3	1:A:35:PHE:CE1	0.46	2.44	10	6
1:A:23:THR:HB	1:A:35:PHE:CZ	0.46	2.46	5	1
1:A:36:LYS:NZ	1:A:40:ASN:HD21	0.46	2.09	7	1
1:A:49:THR:HG22	1:A:50:TYR:H	0.45	1.71	1	1
1:A:23:THR:HG22	1:A:35:PHE:CZ	0.45	2.47	20	10
1:A:20:GLU:N	1:A:20:GLU:CD	0.45	2.71	20	4
1:A:10:LEU:HD22	1:A:11:VAL:N	0.44	2.28	10	2
1:A:37:GLN:NE2	1:A:38:TYR:N	0.44	2.66	8	2
1:A:50:TYR:HB2	1:A:57:PHE:CE1	0.43	2.48	19	1
1:A:26:VAL:CG1	1:A:27:ASP:N	0.43	2.82	14	1
1:A:10:LEU:CB	1:A:35:PHE:CE1	0.43	3.02	2	1
1:A:9:LYS:N	1:A:9:LYS:HD2	0.43	2.27	19	1
1:A:10:LEU:HD22	1:A:35:PHE:CD1	0.43	2.49	23	2
1:A:11:VAL:HG23	1:A:56:THR:HG21	0.42	1.91	1	1
1:A:12:ILE:HG12	1:A:59:VAL:CG1	0.42	2.45	3	2
1:A:23:THR:CG2	1:A:24:GLU:N	0.41	2.83	17	1
1:A:44:VAL:O	1:A:45:ASP:CB	0.40	2.69	18	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	46/64 (72%)	41±1 (89±3%)	5±1 (11±3%)	0±0 (0±1%)	32 76
All	All	1242/1728 (72%)	1104 (89%)	132 (11%)	6 (0%)	32 76

All 2 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	45	ASP	4
1	A	46	GLY	2

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	38/53 (72%)	32±1 (84±3%)	6±1 (16±3%)	5 41
All	All	1026/1431 (72%)	857 (84%)	169 (16%)	5 41

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	48	TRP	27
1	A	50	TYR	26
1	A	57	PHE	24
1	A	49	THR	23
1	A	58	THR	11
1	A	10	LEU	11
1	A	20	GLU	6
1	A	42	ASN	6
1	A	51	ASP	5
1	A	13	ASN	5
1	A	37	GLN	5
1	A	45	ASP	4
1	A	47	GLU	3
1	A	24	GLU	3
1	A	40	ASN	3
1	A	23	THR	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	8	TYR	1
1	A	36	LYS	1
1	A	6	THR	1
1	A	33	LYS	1
1	A	9	LYS	1
1	A	27	ASP	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