



Full wwPDB NMR Structure Validation Report ⓘ

Mar 8, 2026 – 05:26 PM UTC

PDB ID : 2KNQ / pdb_00002knq
Title : Solution structure of E.Coli GspH
Authors : Zhang, Y.; Jin, C.
Deposited on : 2009-08-31

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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity : 4-5-2 with Phenix2.0
Percentile statistics : 20250101.v01 (using entries in the PDB archive January 1st 2025)
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) : 2.49

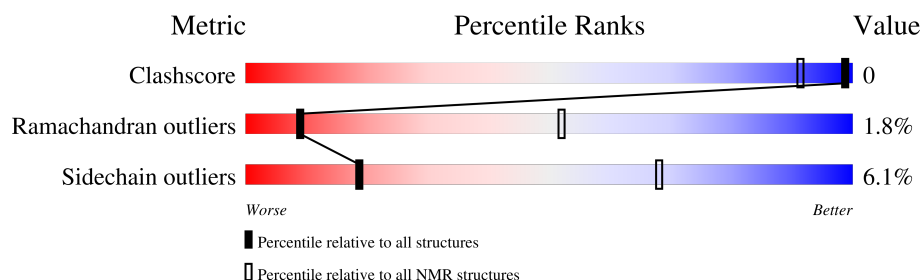
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

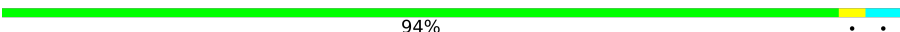
The overall completeness of chemical shifts assignment was not calculated.

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	229148	14424
Ramachandran outliers	224038	12848
Sidechain outliers	223484	12823

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	142	 94%

2 Ensemble composition and analysis

This entry contains 20 models. Model 2 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *structure with lowest energy*.

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:142 (137)	1.15	2

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 1 single-model cluster was found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 2158 atoms, of which 1059 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called General secretion pathway protein H.

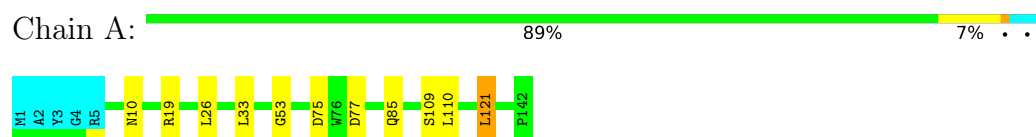
Mol	Chain	Residues	Atoms						Trace
1	A	142	Total	C	H	N	O	S	0
			2158	687	1059	192	216	4	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP B1IPX0
A	2	ALA	-	expression tag	UNP B1IPX0

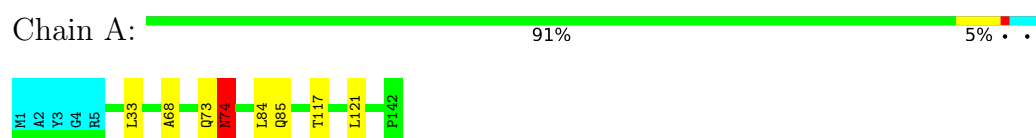
4.2.3 Score per residue for model 3

- Molecule 1: General secretion pathway protein H



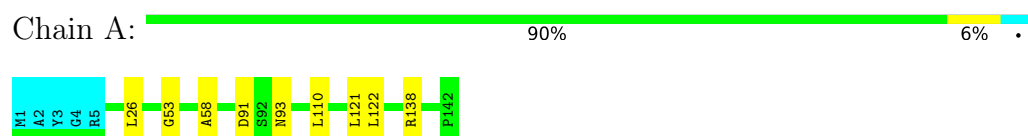
4.2.4 Score per residue for model 4

- Molecule 1: General secretion pathway protein H



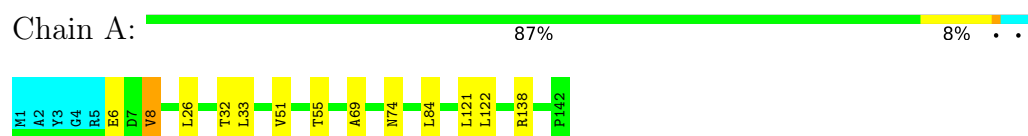
4.2.5 Score per residue for model 5

- Molecule 1: General secretion pathway protein H



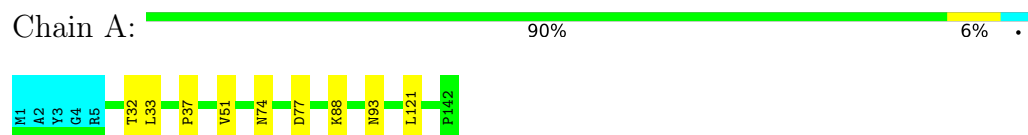
4.2.6 Score per residue for model 6

- Molecule 1: General secretion pathway protein H



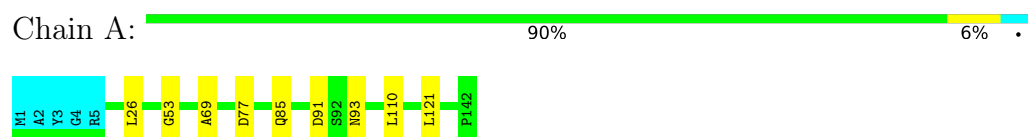
4.2.7 Score per residue for model 7

- Molecule 1: General secretion pathway protein H



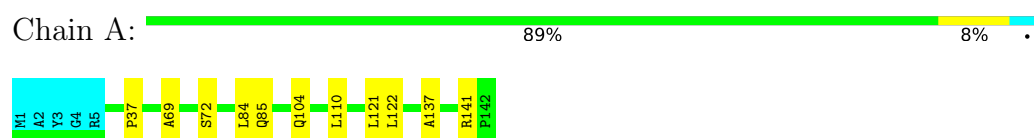
4.2.8 Score per residue for model 8

- Molecule 1: General secretion pathway protein H



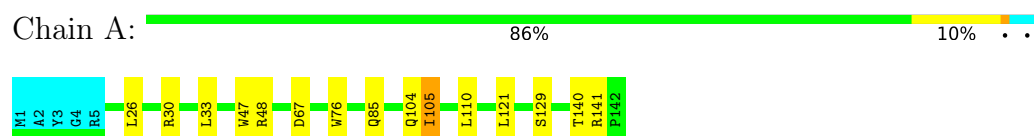
4.2.9 Score per residue for model 9

- Molecule 1: General secretion pathway protein H



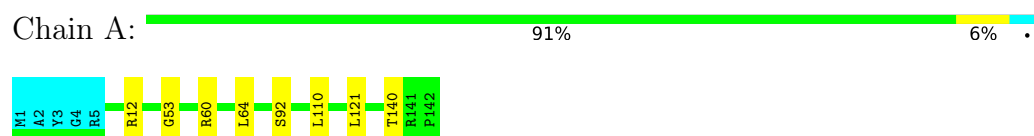
4.2.10 Score per residue for model 10

- Molecule 1: General secretion pathway protein H



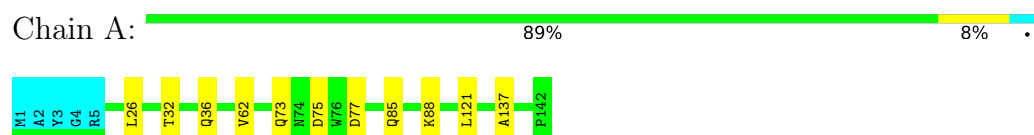
4.2.11 Score per residue for model 11

- Molecule 1: General secretion pathway protein H



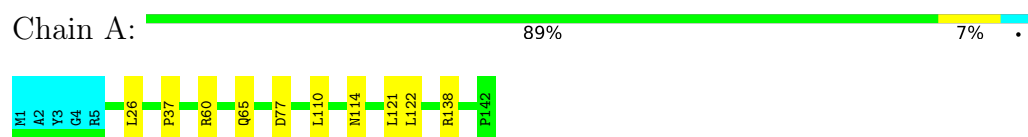
4.2.12 Score per residue for model 12

- Molecule 1: General secretion pathway protein H



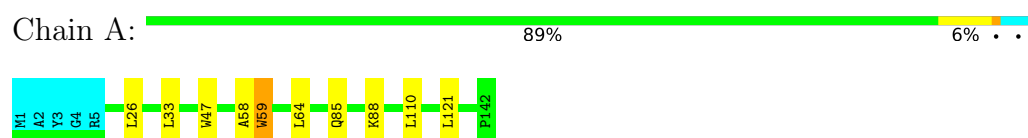
4.2.13 Score per residue for model 13

- Molecule 1: General secretion pathway protein H



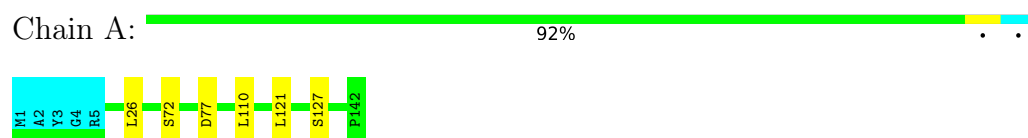
4.2.14 Score per residue for model 14

- Molecule 1: General secretion pathway protein H



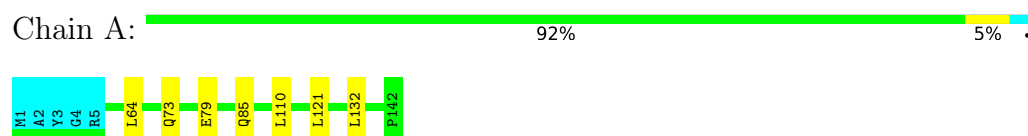
4.2.15 Score per residue for model 15

- Molecule 1: General secretion pathway protein H



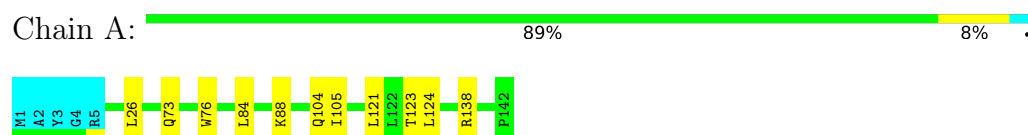
4.2.16 Score per residue for model 16

- Molecule 1: General secretion pathway protein H



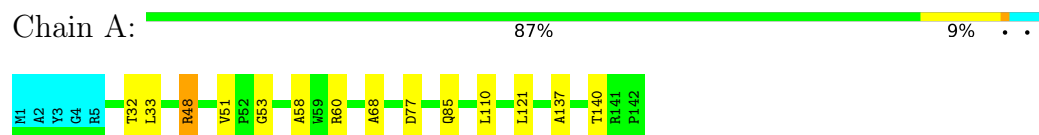
4.2.17 Score per residue for model 17

- Molecule 1: General secretion pathway protein H



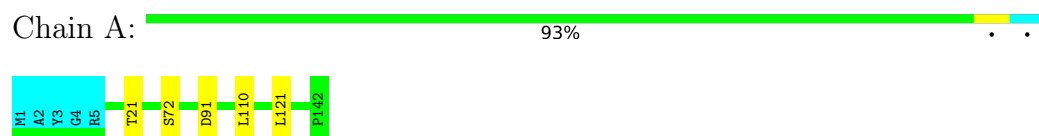
4.2.18 Score per residue for model 18

- Molecule 1: General secretion pathway protein H



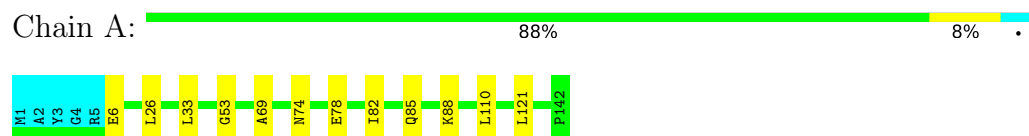
4.2.19 Score per residue for model 19

- Molecule 1: General secretion pathway protein H



4.2.20 Score per residue for model 20

- Molecule 1: General secretion pathway protein H



5 Refinement protocol and experimental data overview

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

Of the 100 calculated structures, 20 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
Amber	refinement	7
CYANA	structure solution	2.0

No chemical shift data was provided.

6 Model quality

6.1 Standard geometry

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	0.88±0.01	0±0/1085 (0.0± 0.0%)	1.46±0.02	1±1/1485 (0.0± 0.1%)
All	All	0.88	0/21700 (0.0%)	1.46	13/29700 (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	0.2±0.4
All	All	0	4

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	114	ASN	CA-CB-CG	5.91	118.51	112.60	13	1
1	A	51	VAL	N-CA-CB	-5.67	105.64	111.87	6	1
1	A	73	GLN	CA-C-N	5.46	131.97	121.54	1	2
1	A	73	GLN	C-N-CA	5.46	131.97	121.54	1	2
1	A	129	SER	CA-C-N	5.29	127.74	120.39	10	1
1	A	129	SER	C-N-CA	5.29	127.74	120.39	10	1
1	A	58	ALA	CA-C-N	5.25	131.57	121.54	14	1
1	A	58	ALA	C-N-CA	5.25	131.57	121.54	14	1
1	A	74	ASN	N-CA-C	5.11	121.68	110.80	4	1
1	A	88	LYS	N-CA-C	5.08	116.28	109.83	12	1
1	A	85	GLN	OE1-CD-NE2	-5.03	117.57	122.60	12	1

There are no chirality outliers.

All 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	19	ARG	Sidechain	3
1	A	48	ARG	Sidechain	1

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	1059	1018	1018	0±0
All	All	21180	20360	20360	4

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 0.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:104:GLN:C	1:A:105:ILE:HD12	0.50	2.31	10	2
1:A:84:LEU:HD23	1:A:85:GLN:N	0.41	2.31	9	1
1:A:10:ASN:HA	1:A:121:LEU:HD11	0.41	1.93	3	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	136/142 (96%)	120±2 (88±2%)	14±2 (10±2%)	2±1 (2±1%)	9	52
All	All	2720/2840 (96%)	2399 (88%)	273 (10%)	48 (2%)	9	52

All 19 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	53	GLY	7
1	A	74	ASN	5
1	A	137	ALA	4
1	A	69	ALA	4
1	A	68	ALA	3
1	A	91	ASP	3
1	A	37	PRO	3
1	A	60	ARG	3
1	A	64	LEU	3
1	A	75	ASP	2
1	A	58	ALA	2
1	A	72	SER	2
1	A	8	VAL	1
1	A	67	ASP	1
1	A	92	SER	1
1	A	65	GLN	1
1	A	59	TRP	1
1	A	127	SER	1
1	A	76	TRP	1

6.3.2 Protein sidechains ⓘ

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	114/117 (97%)	107±2 (94±2%)	7±2 (6±2%)	19 68
All	All	2280/2340 (97%)	2140 (94%)	140 (6%)	19 68

All 40 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	121	LEU	20
1	A	110	LEU	15
1	A	26	LEU	13
1	A	85	GLN	10
1	A	77	ASP	9
1	A	33	LEU	8
1	A	88	LYS	6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	32	THR	5
1	A	140	THR	4
1	A	122	LEU	4
1	A	138	ARG	4
1	A	141	ARG	3
1	A	6	GLU	3
1	A	84	LEU	3
1	A	93	ASN	3
1	A	73	GLN	3
1	A	51	VAL	2
1	A	47	TRP	2
1	A	48	ARG	2
1	A	109	SER	1
1	A	74	ASN	1
1	A	117	THR	1
1	A	8	VAL	1
1	A	55	THR	1
1	A	104	GLN	1
1	A	30	ARG	1
1	A	76	TRP	1
1	A	105	ILE	1
1	A	12	ARG	1
1	A	36	GLN	1
1	A	62	VAL	1
1	A	59	TRP	1
1	A	72	SER	1
1	A	79	GLU	1
1	A	132	LEU	1
1	A	123	THR	1
1	A	124	LEU	1
1	A	21	THR	1
1	A	78	GLU	1
1	A	82	ILE	1

6.3.3 RNA ⓘ

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates [i](#)

There are no oligosaccharides 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

No chemical shift data were provided