Full wwPDB NMR Structure Validation Report (i)

Jun 17, 2024 – 04:04 AM EDT

PDB ID	:	5IX9
BMRB ID	:	30040
Title	:	Cell surface anchoring domain
Authors	:	Guo, S.; Langelaan, D.
Deposited on	:	2016-03-23

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 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:

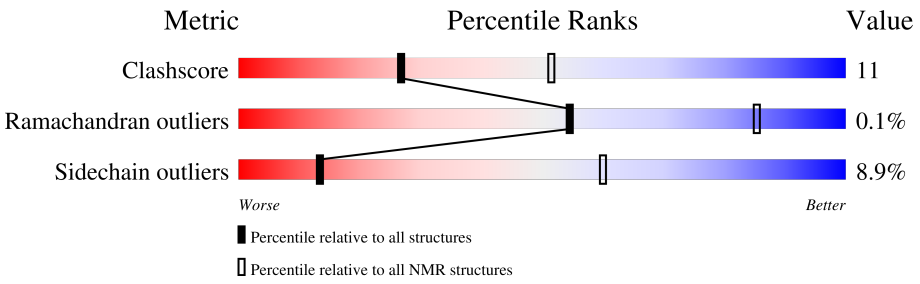
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 is 95%.

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	195	25% 9% • 5% 61%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *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:2-A:41, A:45-A:71 (67)	0.26	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 3 clusters and 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1084 atoms, of which 535 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Antifreeze protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	76	1084	343	535	87	118	1	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-10	MET	-	initiating methionine	UNP A1YIY2
A	-9	SER	-	expression tag	UNP A1YIY2
A	-8	ASP	-	expression tag	UNP A1YIY2
A	-7	ASN	-	expression tag	UNP A1YIY2
A	-6	GLN	-	expression tag	UNP A1YIY2
A	-5	PHE	-	expression tag	UNP A1YIY2
A	-4	PRO	-	expression tag	UNP A1YIY2
A	-3	PHE	-	expression tag	UNP A1YIY2
A	-2	ALA	-	expression tag	UNP A1YIY2
A	-1	THR	-	expression tag	UNP A1YIY2
A	0	LEU	-	expression tag	UNP A1YIY2
A	20	ASN	ASP	engineered mutation	UNP A1YIY2
A	177	LEU	-	expression tag	UNP A1YIY2
A	178	GLU	-	expression tag	UNP A1YIY2
A	179	HIS	-	expression tag	UNP A1YIY2
A	180	HIS	-	expression tag	UNP A1YIY2
A	181	HIS	-	expression tag	UNP A1YIY2
A	182	HIS	-	expression tag	UNP A1YIY2
A	183	HIS	-	expression tag	UNP A1YIY2
A	184	HIS	-	expression tag	UNP A1YIY2

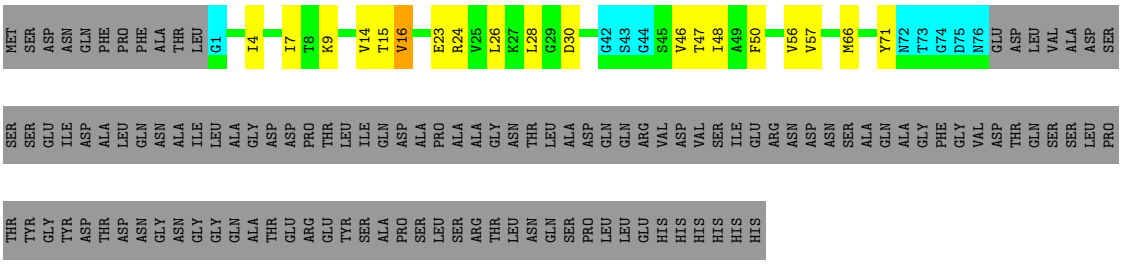
4 Residue-property plots (i)

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: Antifreeze protein

Chain A: 25% 9% • 5% 61%



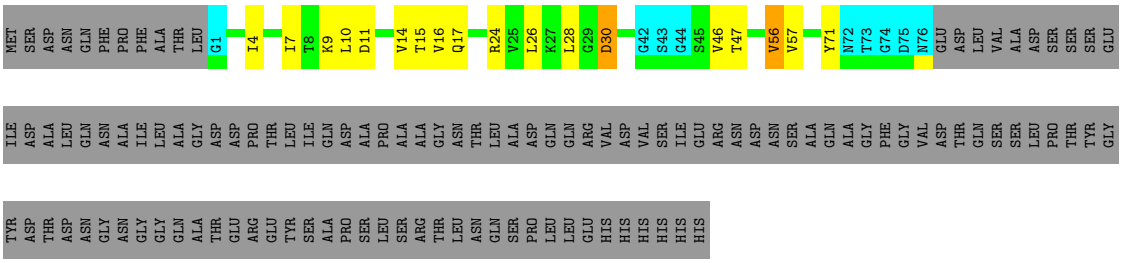
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Antifreeze protein

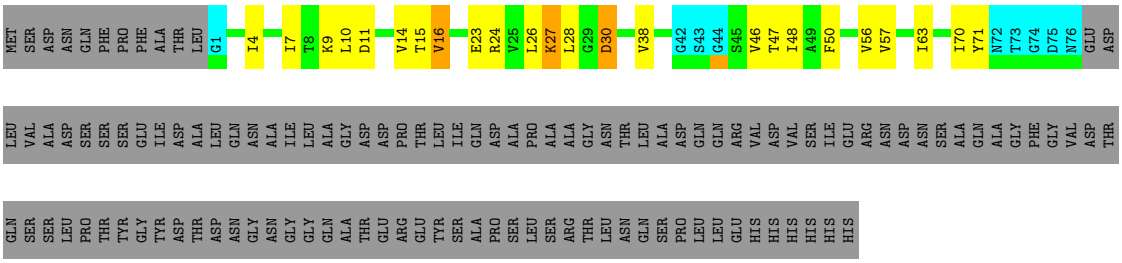
Chain A: 25% 8% • 5% 61%



4.2.2 Score per residue for model 2

- Molecule 1: Antifreeze protein

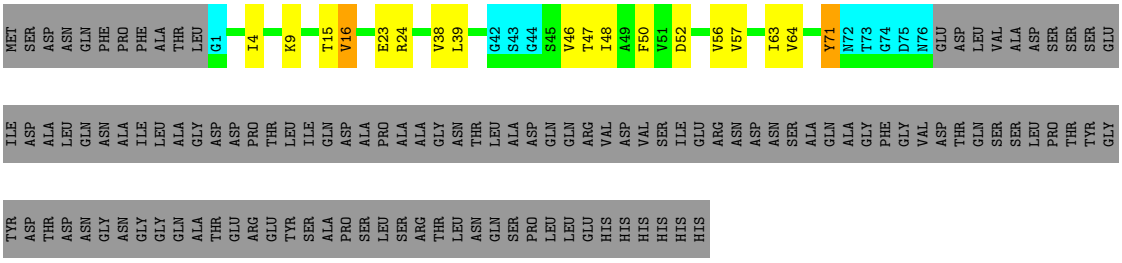
Chain A: 22% 11% • 5% 61%



4.2.3 Score per residue for model 3

- Molecule 1: Antifreeze protein

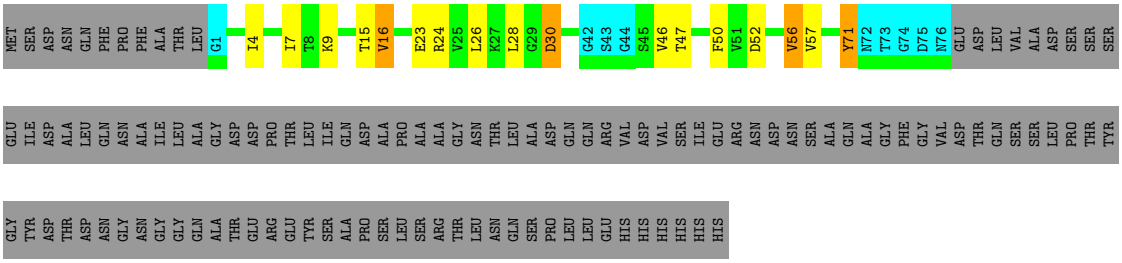
Chain A: 25% 8% • 5% 61%



4.2.4 Score per residue for model 4

- Molecule 1: Antifreeze protein

Chain A: 26% 7% • 5% 61%



4.2.5 Score per residue for model 5

- Molecule 1: Antifreeze protein

Chain A: 24% 9% • 5% 61%

MET SER ASP ASN GLN PHE PRO PHE ALA THR LEU G1 I4 I7 T15 V16 Q17 Q22 E23 R24 V25 L26 K27 L28 G29 D30 E36 G42 S43 G44 S45 F50 V56 V57 I58 G59 G60 M66 I70 Y71 N72 T73 G74 D75 N76 GLU ASP LEU VAL ALA

ASP SER SER SER GLU ILE ASP ALA LEU GLN ASN ALA ILE LEU ALA GLY ASP ASP PRO THR LEU ILE GLN ASP ALA PRO ALA ALA GLY ASN THR LEU ASP GLN GLN ARG VAL ASP VAL SER ILE GLU ARG ASN ASP ASN SER ALA GLN ALA GLY PHE GLY VAL ASP THR GLN SER SER

LEU PRO THR TYR GLY TYR ASP THR ASP ASN GLY ASN GLY GLY GLN ALA THR GLU ARG GLU TYR SER ALA PRO SER LEU SER ARG THR LEU ASN GLN SER PRO LEU LEU GLU HIS HIS HIS HIS HIS HIS HIS HIS

4.2.6 Score per residue for model 6

- Molecule 1: Antifreeze protein

Chain A: 26% 7% • 5% 61%

MET SER ASP ASN GLN PHE PRO PHE ALA THR LEU G1 I4 L10 D11 G12 S13 V14 T15 V16 E23 R24 L28 G42 S43 G44 I48 T54 D55 V56 M66 I70 Y71 N72 T73 G74 D75 N76 GLU ASP LEU VAL ALA ASP SER SER SER GLU ILE ASP ALA

LEU GLN ASN ALA ILE LEU ALA GLY ASP ASP PRO THR LEU ILE GLN ASP ALA PRO ALA ALA GLY ASN THR LEU ALA ASP GLN GLN ARG VAL ASP VAL SER ILE GLU ARG ASN ASP ASN SER ALA GLN ALA GLY PHE GLY VAL ASP THR GLN SER SER LEU PRO THR TYR GLY TYR ASP THR

ASP ASN GLY ASN GLY GLY GLN ALA THR GLU ARG GLU TYR SER ALA PRO SER LEU SER ARG THR LEU ASN GLN SER PRO LEU LEU GLU HIS HIS HIS HIS HIS

4.2.7 Score per residue for model 7

- Molecule 1: Antifreeze protein

Chain A: 26% 7% • 5% 61%

MET SER ASP ASN GLN PHE PRO PHE ALA THR LEU G1 I4 I7 T15 V16 E23 R24 V25 L26 K27 L28 G29 D30 V38 G42 S43 G44 S45 F50 V51 D52 I58 I63 Y71 N72 T73 G74 D75 N76 GLU ASP LEU VAL ALA ASP SER SER SER GLU

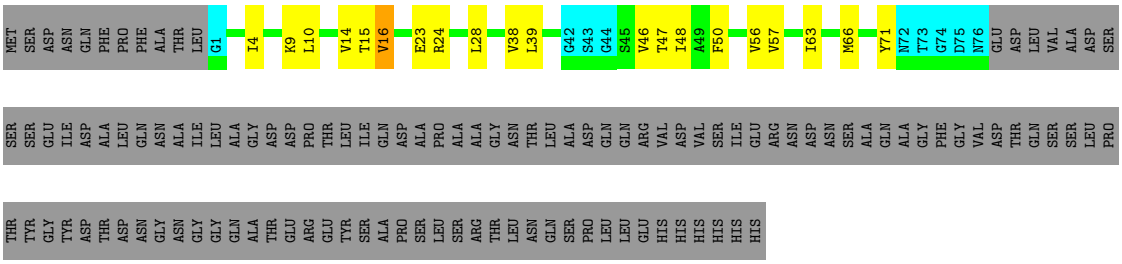
ILE ASP ALA LEU GLN ASN ALA ILE LEU ALA GLY ASP ASP PRO THR LEU ILE GLN ASP ALA PRO ALA ALA GLY ASN THR LEU ALA ASP GLN GLN ARG VAL ASP VAL SER ILE GLU ARG ASN ASP ASN SER ALA GLN ALA GLY PHE GLY VAL ASP THR GLN SER SER LEU PRO THR TYR GLY

TYR ASP THR ASP ASN GLY ASN GLY GLY GLN ALA THR GLU ARG GLU TYR SER ALA PRO SER LEU SER ARG THR LEU ASN GLN SER PRO LEU LEU GLU HIS HIS HIS HIS HIS

4.2.8 Score per residue for model 8 (medoid)

- Molecule 1: Antifreeze protein

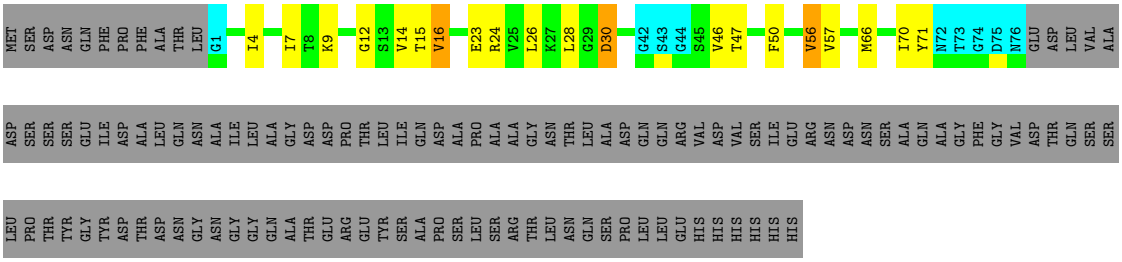
Chain A: 24% 10% • 5% 61%



4.2.9 Score per residue for model 9

- Molecule 1: Antifreeze protein

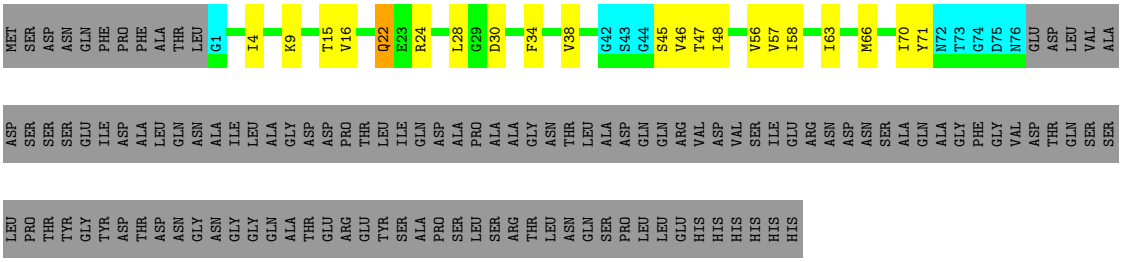
Chain A: 24% 9% • 5% 61%



4.2.10 Score per residue for model 10

- Molecule 1: Antifreeze protein

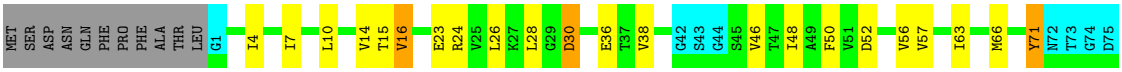
Chain A: 24% 10% • 5% 61%

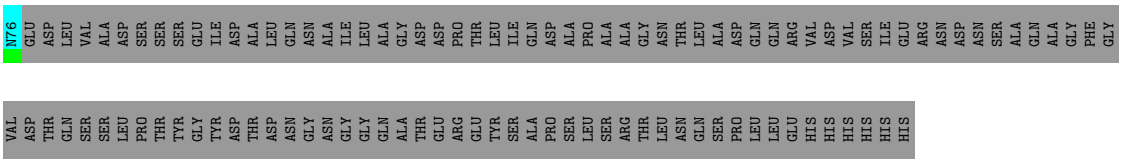


4.2.11 Score per residue for model 11

- Molecule 1: Antifreeze protein

Chain A: 23% 10% • 5% 61%

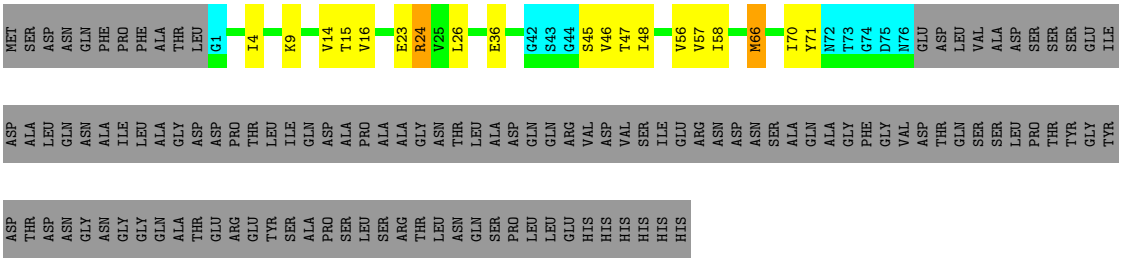




4.2.12 Score per residue for model 12

- Molecule 1: Antifreeze protein

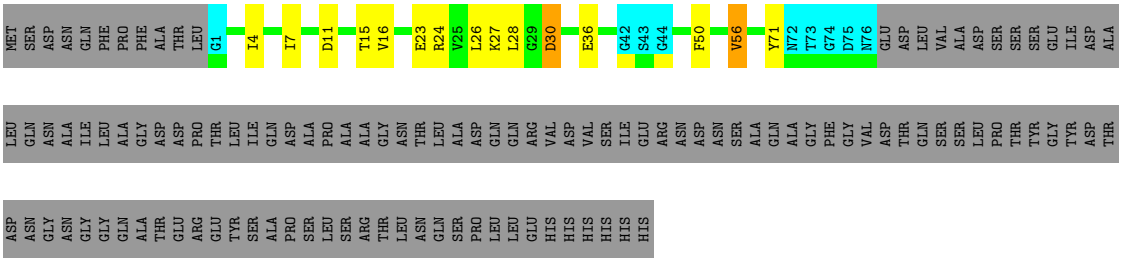
Chain A:



4.2.13 Score per residue for model 13

- Molecule 1: Antifreeze protein

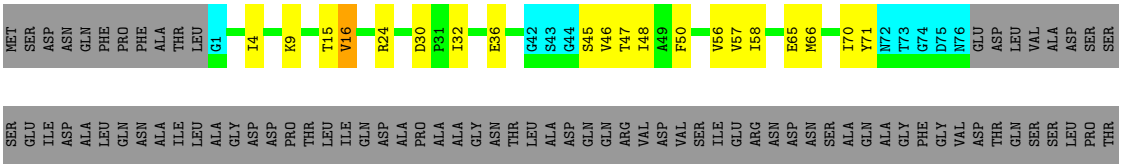
Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Antifreeze protein

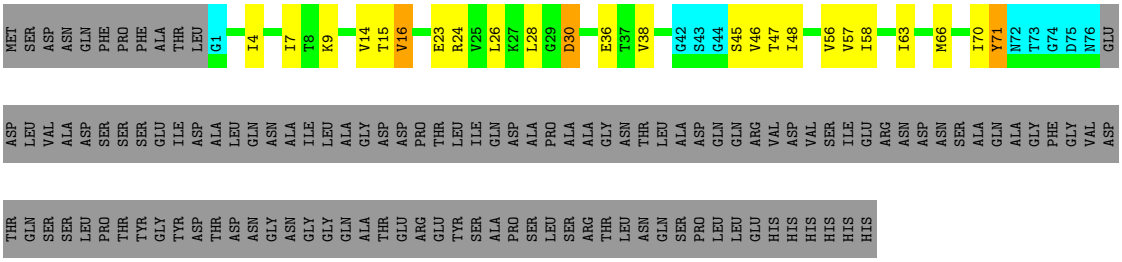
Chain A:



4.2.15 Score per residue for model 15

- Molecule 1: Antifreeze protein

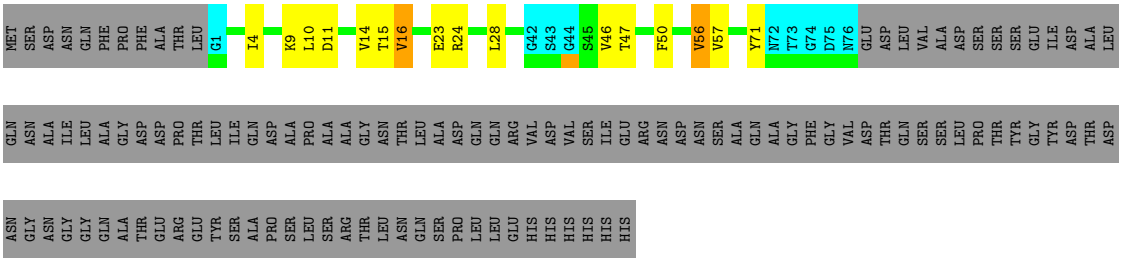
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: Antifreeze protein

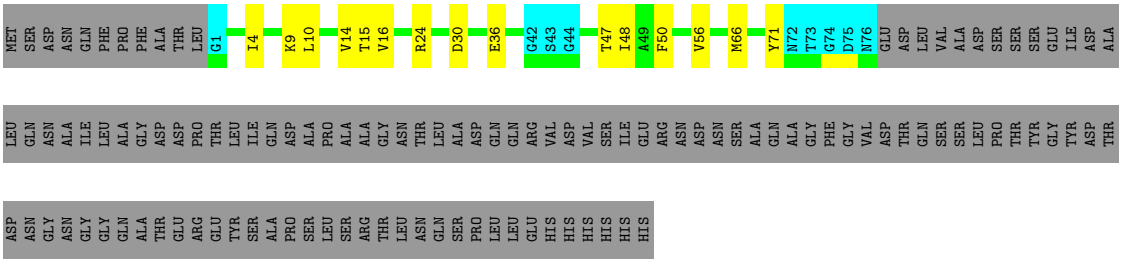
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Antifreeze protein

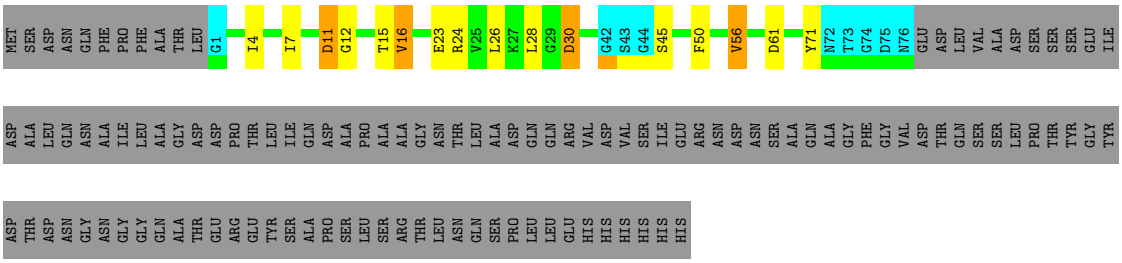
Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: Antifreeze protein

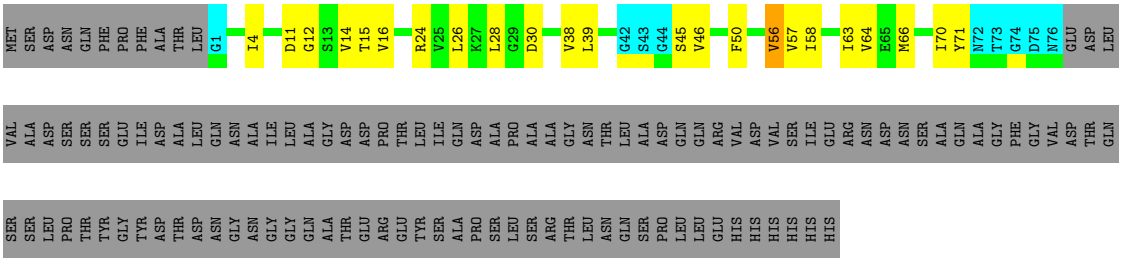
Chain A: 26% 6% • 5% 61%



4.2.19 Score per residue for model 19

- Molecule 1: Antifreeze protein

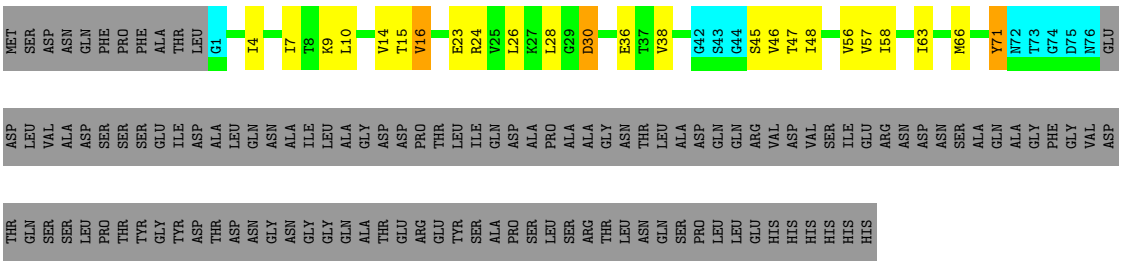
Chain A: 23% 11% • 5% 61%



4.2.20 Score per residue for model 20

- Molecule 1: Antifreeze protein

Chain A: 22% 11% • 5% 61%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *na*.

Of the 20 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
ARIA	refinement	
CNS	structure calculation	

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)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	889
Number of shifts mapped to atoms	888
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	95%

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	0.49±0.08	0±1/502 (0.1± 0.2%)	0.50±0.02	0±0/683 (0.0± 0.0%)
All	All	0.49	10/10040 (0.1%)	0.50	0/13660 (0.0%)

All unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	71	TYR	CE2-CZ	-8.73	1.27	1.38	7	6
1	A	71	TYR	CE1-CZ	7.42	1.48	1.38	7	4

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	496	495	495	11±2
All	All	9920	9900	9900	218

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 11.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:4:ILE:HB	1:A:71:TYR:CE2	0.73	2.19	13	20
1:A:4:ILE:HB	1:A:71:TYR:CD2	0.64	2.28	19	20
1:A:52:ASP:HB2	1:A:71:TYR:OH	0.63	1.93	7	4
1:A:45:SER:HA	1:A:58:ILE:O	0.59	1.98	7	8
1:A:16:VAL:O	1:A:23:GLU:HA	0.59	1.98	11	15
1:A:16:VAL:HG12	1:A:24:ARG:HG2	0.55	1.79	3	13
1:A:38:VAL:O	1:A:63:ILE:HA	0.55	2.00	10	9
1:A:66:MET:SD	1:A:70:ILE:HG21	0.53	2.43	12	4
1:A:9:LYS:HE2	1:A:47:THR:OG1	0.53	2.03	9	2
1:A:50:PHE:CE1	1:A:56:VAL:HB	0.52	2.40	19	7
1:A:46:VAL:O	1:A:57:VAL:HA	0.51	2.04	9	14
1:A:27:LYS:O	1:A:30:ASP:HB3	0.50	2.06	2	2
1:A:26:LEU:HD22	1:A:30:ASP:OD1	0.50	2.07	13	9
1:A:7:ILE:HG12	1:A:30:ASP:O	0.49	2.08	9	10
1:A:47:THR:HA	1:A:56:VAL:O	0.49	2.08	1	5
1:A:50:PHE:HB3	1:A:71:TYR:CE1	0.47	2.45	4	5
1:A:66:MET:HE3	1:A:70:ILE:HG21	0.47	1.86	6	2
1:A:10:LEU:HD11	1:A:14:VAL:HG21	0.46	1.87	2	8
1:A:9:LYS:HB3	1:A:47:THR:HB	0.46	1.87	16	10
1:A:66:MET:CE	1:A:70:ILE:HG21	0.46	2.41	5	3
1:A:48:ILE:HB	1:A:56:VAL:CG2	0.44	2.42	10	11
1:A:36:GLU:O	1:A:66:MET:HG2	0.44	2.12	11	5
1:A:26:LEU:HD22	1:A:30:ASP:CG	0.43	2.33	2	7
1:A:9:LYS:HB2	1:A:47:THR:HB	0.43	1.90	4	1
1:A:14:VAL:HG23	1:A:26:LEU:HB2	0.43	1.89	15	6
1:A:48:ILE:HD13	1:A:66:MET:HE2	0.43	1.91	8	2
1:A:50:PHE:CE1	1:A:56:VAL:CG2	0.43	3.02	8	6
1:A:7:ILE:HG13	1:A:30:ASP:OD1	0.43	2.14	1	1
1:A:16:VAL:HG22	1:A:36:GLU:HB3	0.42	1.91	14	3
1:A:32:ILE:CG2	1:A:66:MET:HG3	0.42	2.44	14	1
1:A:9:LYS:HB3	1:A:47:THR:CB	0.42	2.44	16	1
1:A:22:GLN:HA	1:A:22:GLN:HE21	0.41	1.75	10	1
1:A:4:ILE:HG21	1:A:34:PHE:HD2	0.41	1.75	10	1
1:A:17:GLN:HA	1:A:22:GLN:O	0.40	2.15	5	1
1:A:24:ARG:NH2	1:A:36:GLU:OE2	0.40	2.53	12	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	67/195 (34%)	62±1 (93±1%)	4±1 (7±1%)	0±0 (0±0%)	54 85
All	All	1340/3900 (34%)	1250 (93%)	89 (7%)	1 (0%)	54 85

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	66	MET	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	56/159 (35%)	51±2 (91±3%)	5±2 (9±3%)	13 60
All	All	1120/3180 (35%)	1020 (91%)	100 (9%)	13 60

All 16 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	15	THR	20
1	A	16	VAL	16
1	A	28	LEU	16
1	A	30	ASP	15
1	A	56	VAL	8
1	A	24	ARG	7
1	A	11	ASP	4
1	A	70	ILE	3
1	A	39	LEU	3
1	A	64	VAL	2
1	A	17	GLN	1
1	A	27	LYS	1
1	A	54	THR	1
1	A	22	GLN	1
1	A	65	GLU	1
1	A	45	SER	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)

The completeness of assignment taking into account all chemical shift lists is 95% for the well-defined parts and 94% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list*

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	889
Number of shifts mapped to atoms	888
Number of unparsed shifts	0
Number of shifts with mapping errors	1
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. All 1 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	1	GLY	H1	8.367	0.003	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	76	-0.21 ± 0.19	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	63	-0.13 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	75	0.12 ± 0.20	None needed (< 0.5 ppm)
^{15}N	75	0.27 ± 0.29	None needed (< 0.5 ppm)

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 95%, i.e. 818 atoms were assigned a chemical shift out of a possible 863. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	341/342 (100%)	142/142 (100%)	133/134 (99%)	66/66 (100%)
Sidechain	429/472 (91%)	288/311 (93%)	136/152 (89%)	5/9 (56%)
Aromatic	48/49 (98%)	24/24 (100%)	24/25 (96%)	0/0 (—%)
Overall	818/863 (95%)	454/477 (95%)	293/311 (94%)	71/75 (95%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 94%, i.e. 887 atoms were assigned a chemical shift out of a possible 939. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	386/391 (99%)	160/164 (98%)	151/152 (99%)	75/75 (100%)
Sidechain	453/499 (91%)	304/327 (93%)	142/161 (88%)	7/11 (64%)
Aromatic	48/49 (98%)	24/24 (100%)	24/25 (96%)	0/0 (—%)
Overall	887/939 (94%)	488/515 (95%)	317/338 (94%)	82/86 (95%)

7.1.4 Statistically unusual chemical shifts (i)

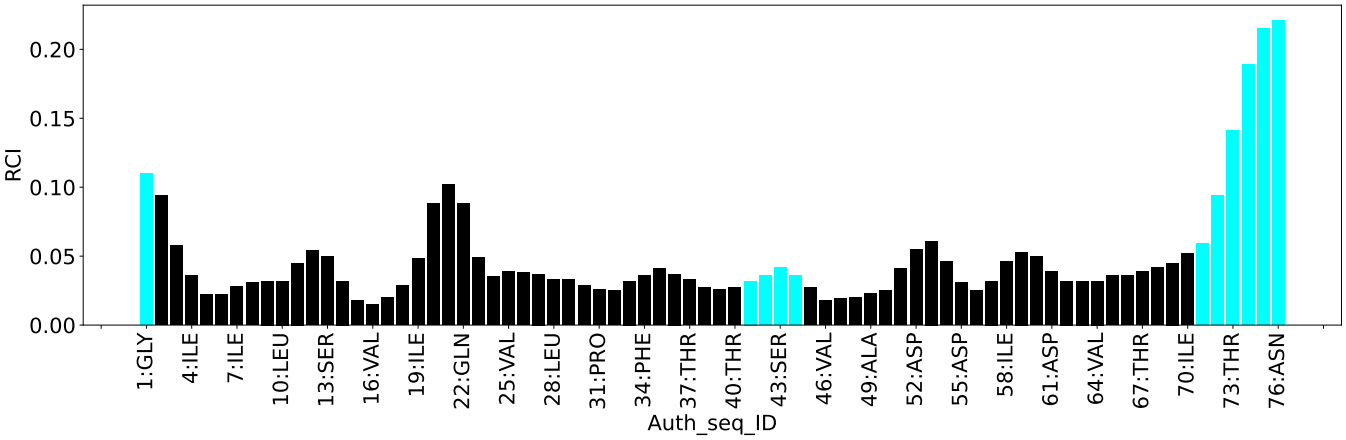
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	40	THR	HG1	6.19	0.08 – 2.19	23.9

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:


RCI
0.00
0.05
0.10
0.15
0.20
1:GLY
4:ILE
7:ILE
10:LEU
13:SER
16:VAL
19:ILE
22:GLN
25:VAL
28:LEU
31:PRO
34:PHE
37:THR
40:THR
43:SER
46:VAL
49:ALA
52:ASP
55:ASP
58:ILE
61:ASP
64:VAL
67:THR
70:ILE
73:THR
76:ASN
Auth_seq_ID