Full wwPDB NMR Structure Validation Report

Feb 17, 2022 – 09:05 AM EST

PDB ID	:	1ON4
Title	:	Solution structure of soluble domain of Sco1 from Bacillus Subtilis
Authors	:	Balatri, E.; Banci, L.; Bertini, I.; Cantini, F.; Ciofi-Baffoni, S.; Structural Proteomics in Europe (SPINE)
Deposited on	:	2003-02-27

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:

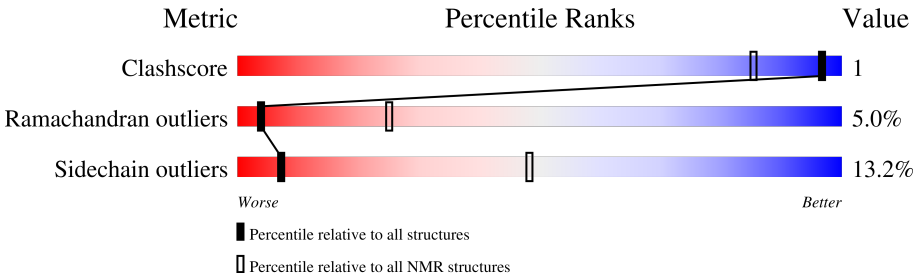
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)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

1 Overall quality at a glance

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	174	72% 7% • 18%

2 Ensemble composition and analysis

This entry contains 30 models. Model 28 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:7-A:43, A:49-A:123, A:138-A:153, A:159-A:172 (142)	0.46	28

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, 3, 4, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 25, 26, 27, 28, 29
2	8, 23, 24
3	6, 30
4	5, 19
Single-model clusters	2

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2765 atoms, of which 1366 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Sco1.

Mol	Chain	Residues	Atoms						Trace
1	A	174	Total	C	H	N	O	S	0
			2765	897	1366	220	277	5	

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	HIS	-	cloning artifact	UNP P54178
A	2	MET	-	cloning artifact	UNP P54178
A	3	LEU	-	cloning artifact	UNP P54178
A	4	GLU	-	cloning artifact	UNP P54178
A	66	GLU	GLY	SEE REMARK 999	UNP P54178

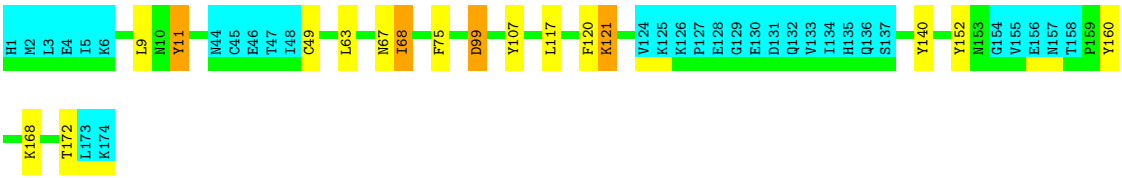
4 Residue-property plots (i)

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: Sco1

Chain A:



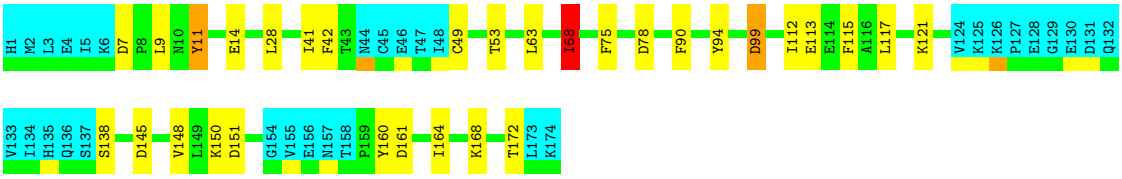
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Sco1

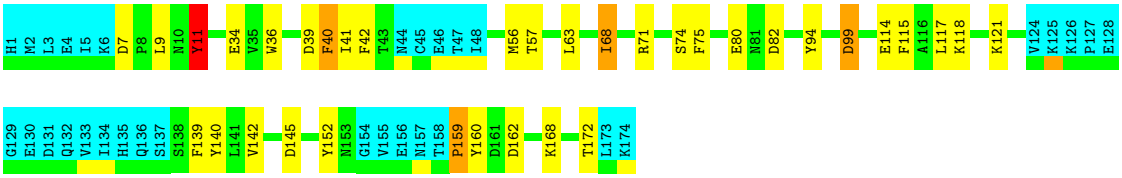
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Sco1

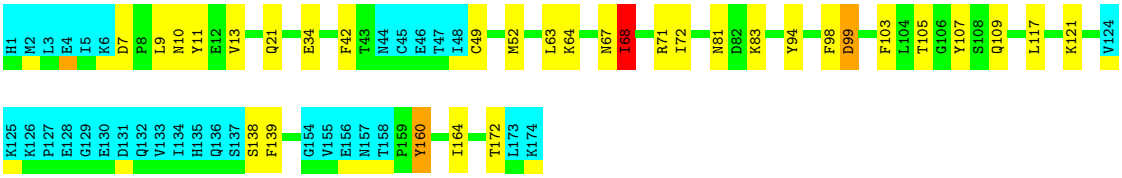
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Sco1

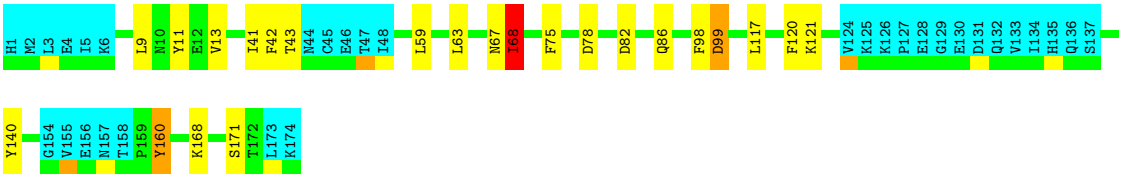
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Sco1

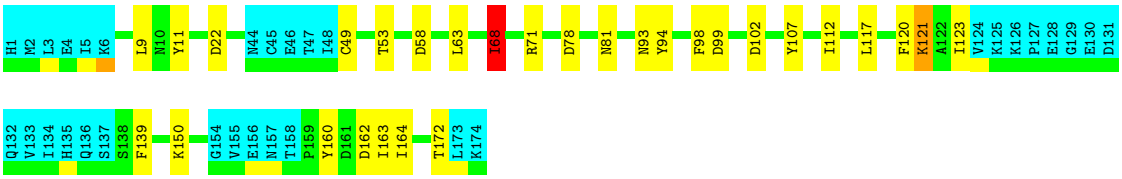
Chain A:



4.2.5 Score per residue for model 5

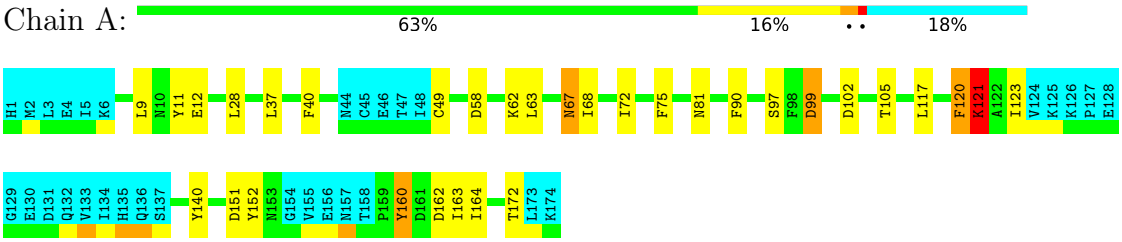
• Molecule 1: Sco1

Chain A:



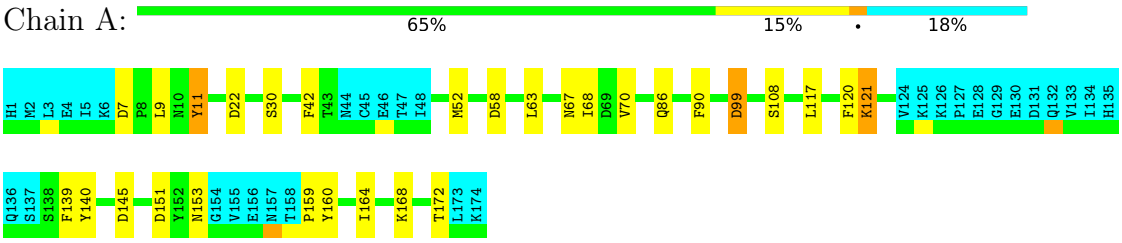
4.2.6 Score per residue for model 6

• Molecule 1: Sco1



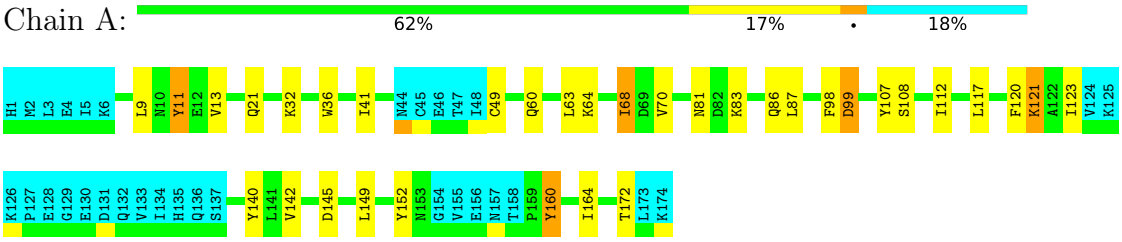
4.2.7 Score per residue for model 7

• Molecule 1: Sco1



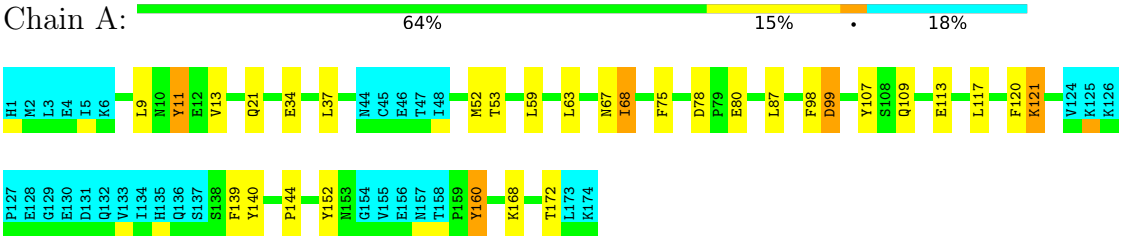
4.2.8 Score per residue for model 8

• Molecule 1: Sco1



4.2.9 Score per residue for model 9

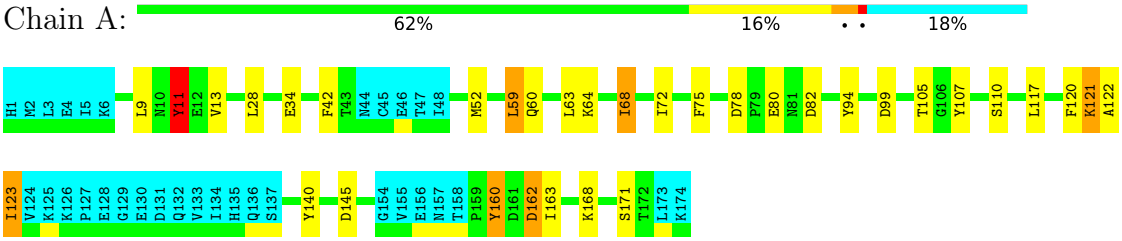
• Molecule 1: Sco1



4.2.10 Score per residue for model 10

• Molecule 1: Sco1

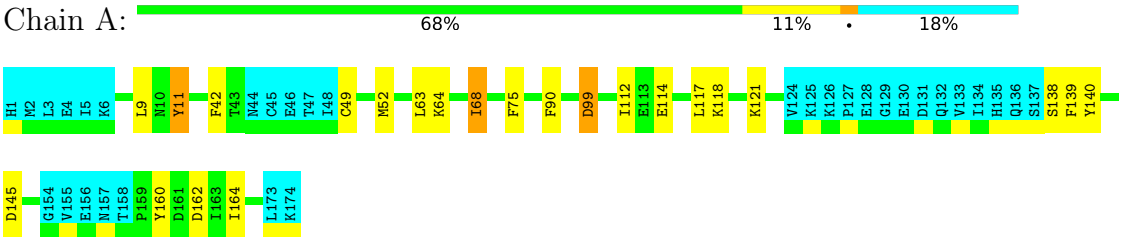
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Sco1

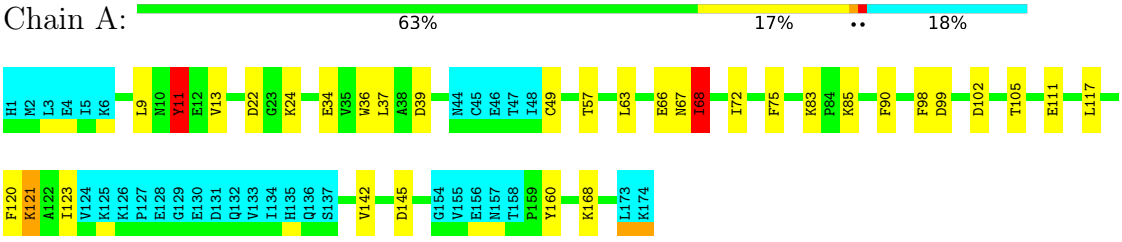
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: Sco1

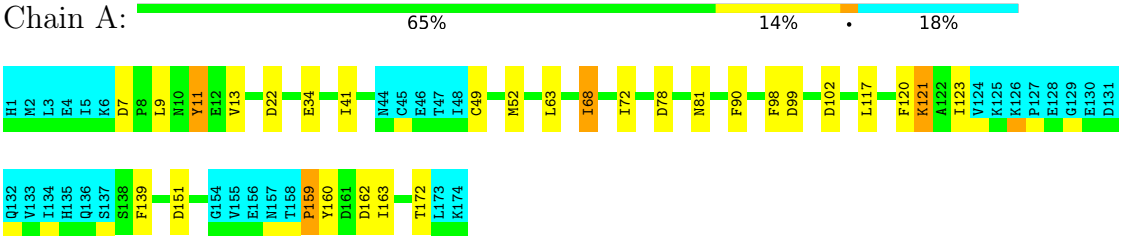
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: Sco1

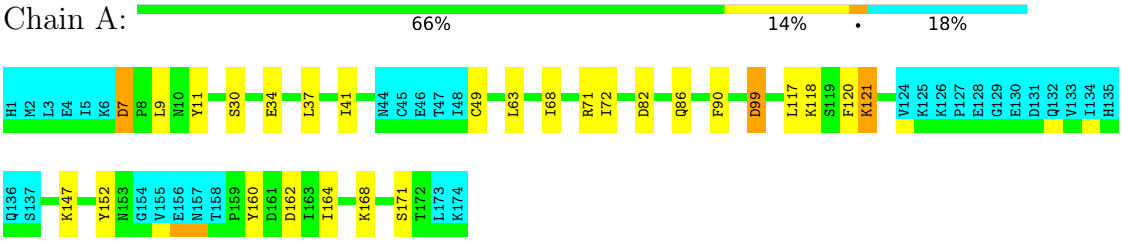
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Sco1

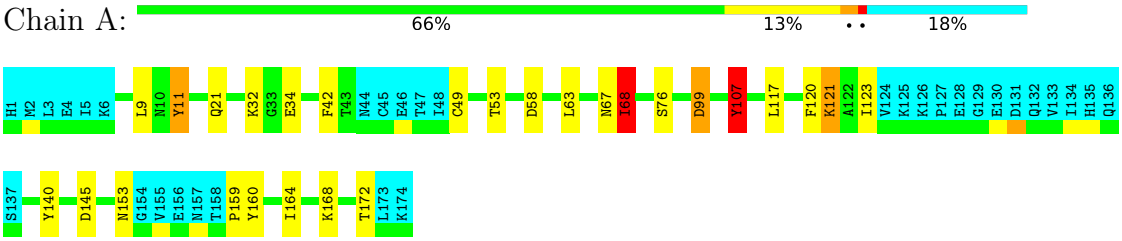
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Sco1

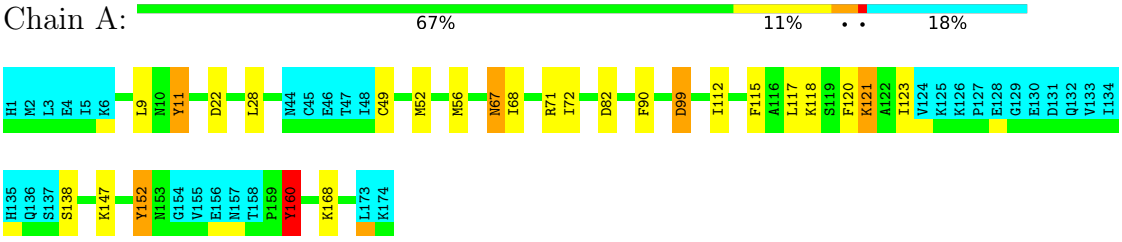
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Sco1

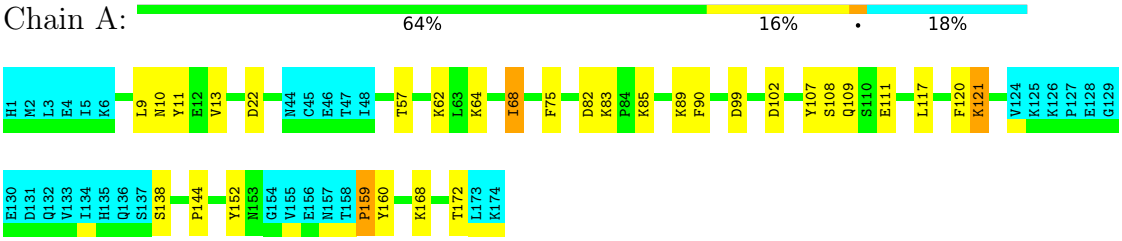
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Sco1

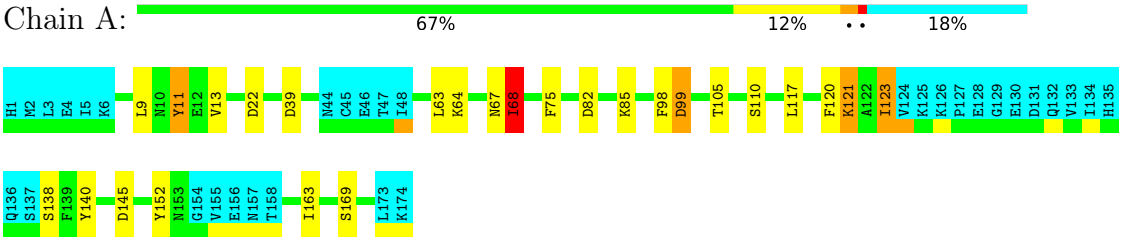
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Sco1

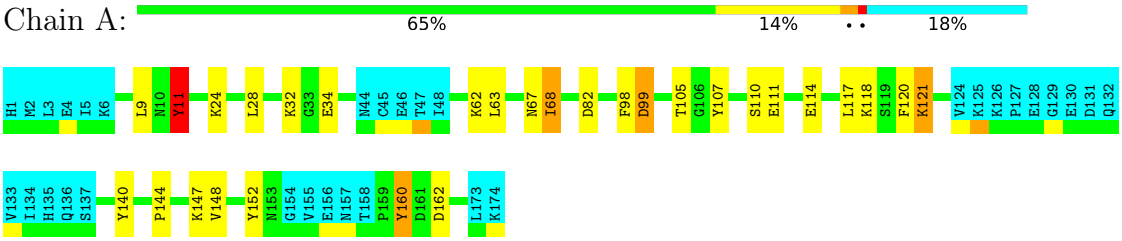
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: Sco1

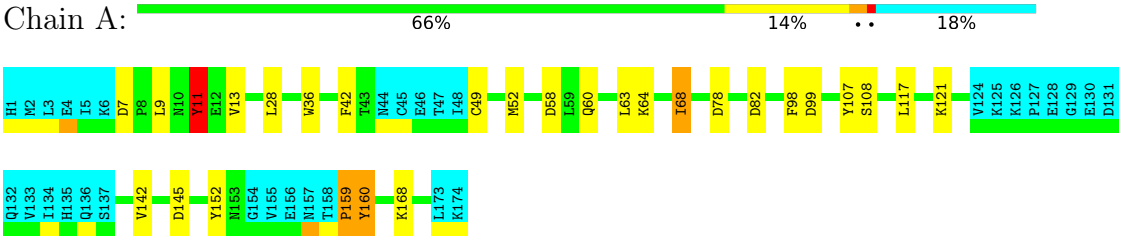
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: Sco1

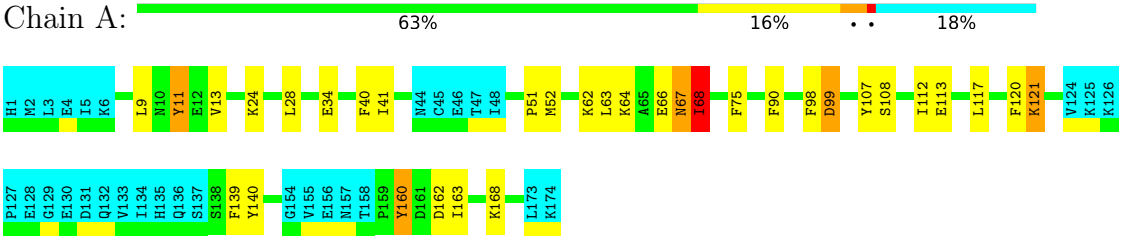
Chain A:



4.2.21 Score per residue for model 21

• Molecule 1: Sco1

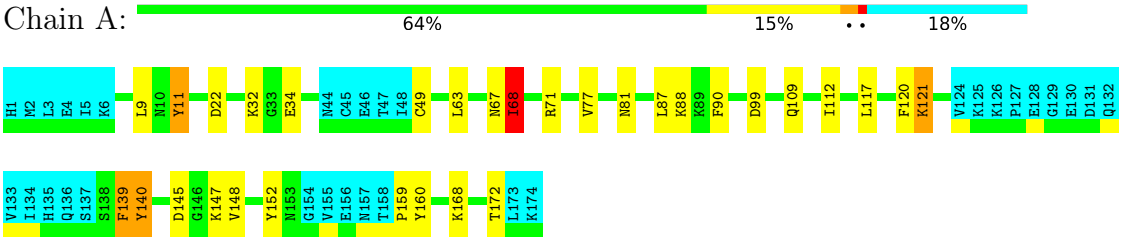
Chain A:



4.2.22 Score per residue for model 22

• Molecule 1: Sco1

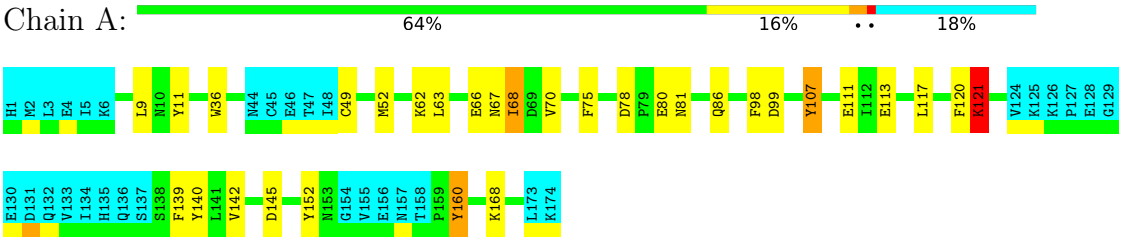
Chain A:



4.2.23 Score per residue for model 23

• Molecule 1: Sco1

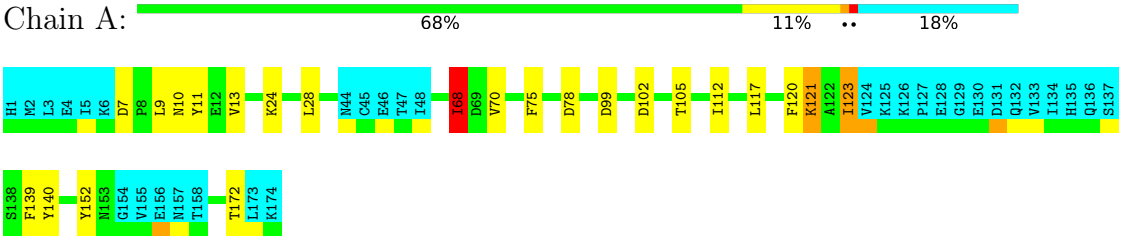
Chain A:



4.2.24 Score per residue for model 24

• Molecule 1: Sco1

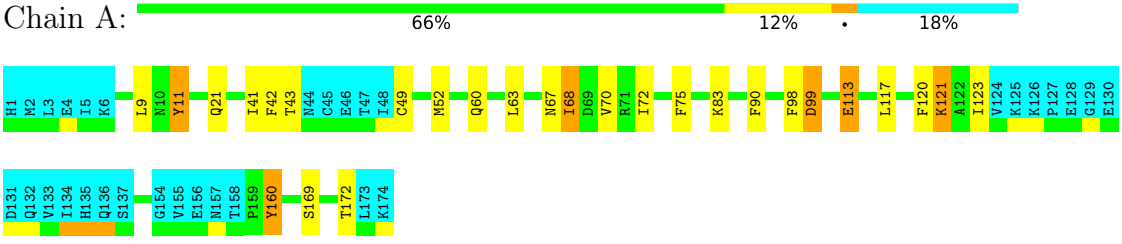
Chain A:



4.2.25 Score per residue for model 25

• Molecule 1: Sco1

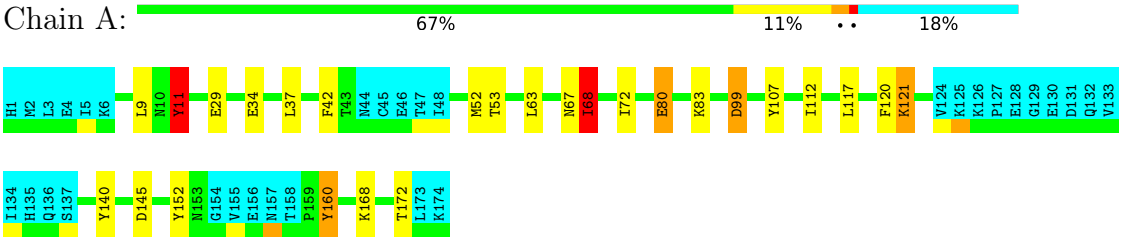
Chain A:



4.2.26 Score per residue for model 26

• Molecule 1: Sco1

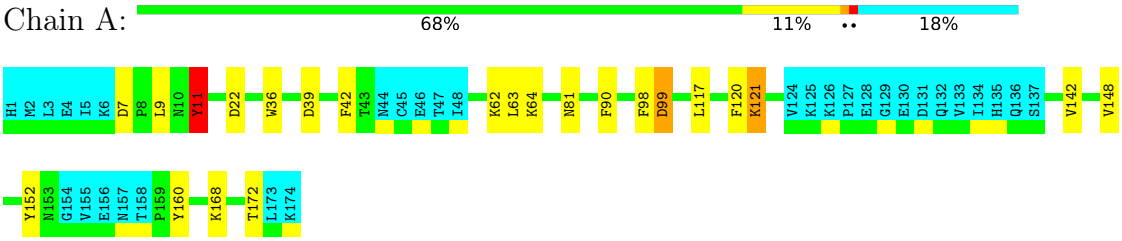
Chain A:



4.2.27 Score per residue for model 27

• Molecule 1: Sco1

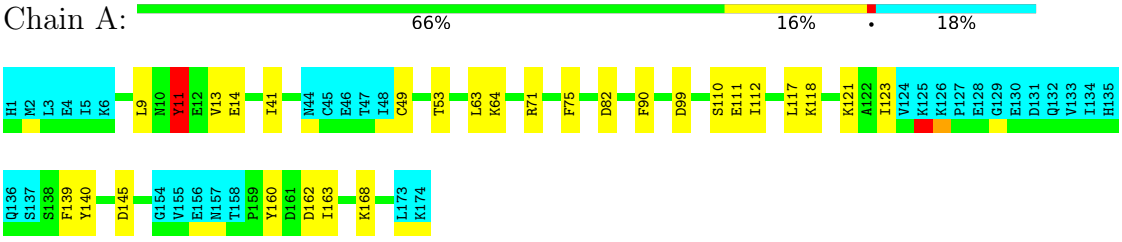
Chain A:



4.2.28 Score per residue for model 28 (medoid)

• Molecule 1: Sco1

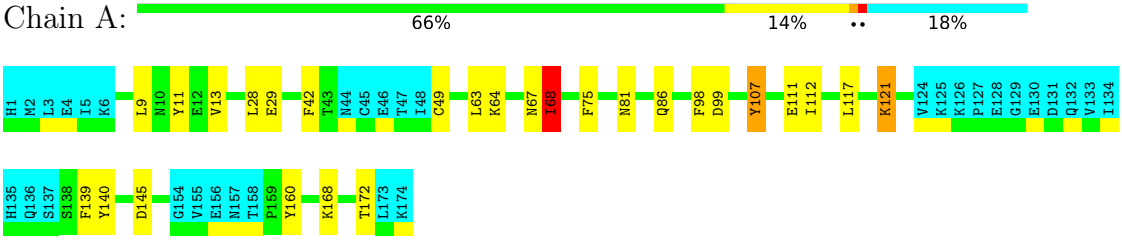
Chain A:



4.2.29 Score per residue for model 29

• Molecule 1: Sco1

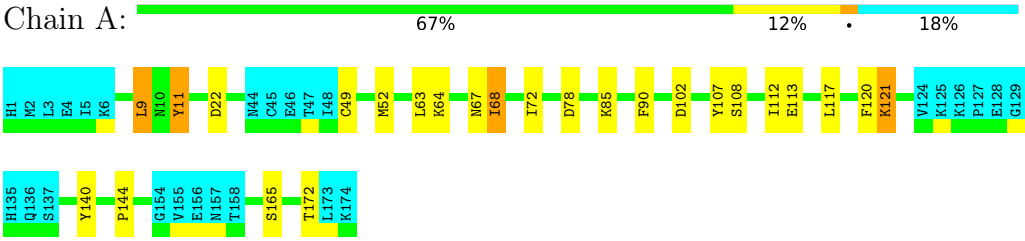
Chain A:



4.2.30 Score per residue for model 30

• Molecule 1: Sco1

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics coupled to simulated annealing followed by restrained energy minimization.*

Of the 300 calculated structures, 30 were deposited, based on the following criterion: *target function.*

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

Software name	Classification	Version
DYANA	structure solution	1.5
CYANA	structure solution	1.0
Amber	refinement	5

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	0.62±0.00	0±0/1176 (0.0± 0.0%)	1.01±0.02	1±1/1595 (0.1± 0.0%)
All	All	0.62	0/35280 (0.0%)	1.01	30/47850 (0.1%)

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	6.2±1.4
All	All	0	185

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	160	TYR	CB-CG-CD2	-9.14	115.52	121.00	10	10
1	A	11	TYR	CB-CG-CD2	-8.56	115.86	121.00	12	10
1	A	71	ARG	NE-CZ-NH2	-6.87	116.86	120.30	3	5
1	A	160	TYR	CB-CG-CD1	-5.91	117.45	121.00	9	3
1	A	107	TYR	CB-CG-CD2	-5.48	117.71	121.00	15	1
1	A	11	TYR	CA-CB-CG	5.21	123.30	113.40	27	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	68	ILE	Peptide	24
1	A	120	PHE	Peptide,Sidechain	23

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	160	TYR	Sidechain,Peptide	22
1	A	140	TYR	Sidechain	18
1	A	11	TYR	Sidechain	17
1	A	75	PHE	Sidechain	15
1	A	90	PHE	Sidechain	15
1	A	98	PHE	Sidechain	15
1	A	152	TYR	Sidechain	13
1	A	94	TYR	Sidechain	4
1	A	159	PRO	Peptide	4
1	A	115	PHE	Sidechain	3
1	A	40	PHE	Sidechain	3
1	A	139	PHE	Sidechain	2
1	A	103	PHE	Sidechain	1
1	A	66	GLU	Peptide	1
1	A	71	ARG	Sidechain	1
1	A	123	ILE	Peptide	1
1	A	107	TYR	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	1145	1108	1108	2±1
All	All	34350	33240	33240	45

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 1.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:11:TYR:CB	1:A:148:VAL:HG22	0.60	2.26	1	1
1:A:11:TYR:HB2	1:A:148:VAL:HG22	0.55	1.79	1	1
1:A:9:LEU:HD13	1:A:165:SER:CB	0.52	2.35	30	1
1:A:68:ILE:HD12	1:A:68:ILE:H	0.51	1.66	12	12
1:A:162:ASP:OD1	1:A:163:ILE:HD12	0.51	2.05	6	1
1:A:75:PHE:CE1	1:A:123:ILE:HD11	0.50	2.41	10	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:11:TYR:HB3	1:A:148:VAL:CG2	0.50	2.37	27	2
1:A:36:TRP:CZ2	1:A:142:VAL:HG11	0.49	2.42	23	6
1:A:75:PHE:HE1	1:A:123:ILE:HD11	0.45	1.72	10	1
1:A:120:PHE:O	1:A:121:LYS:HD2	0.45	2.12	6	1
1:A:114:GLU:OE2	1:A:118:LYS:NZ	0.44	2.48	11	1
1:A:109:GLN:CD	1:A:109:GLN:H	0.44	2.16	17	1
1:A:140:TYR:CD2	1:A:148:VAL:HG11	0.44	2.47	22	1
1:A:62:LYS:NZ	1:A:66:GLU:OE2	0.43	2.50	21	2
1:A:59:LEU:HD13	1:A:60:GLN:N	0.42	2.28	10	1
1:A:122:ALA:CB	1:A:123:ILE:HD12	0.42	2.45	10	1
1:A:11:TYR:CG	1:A:168:LYS:NZ	0.42	2.88	12	1
1:A:11:TYR:HB3	1:A:148:VAL:HG22	0.42	1.91	27	1
1:A:14:GLU:OE2	1:A:118:LYS:NZ	0.42	2.53	28	1
1:A:7:ASP:OD1	1:A:7:ASP:N	0.41	2.53	2	1
1:A:58:ASP:OD2	1:A:62:LYS:NZ	0.41	2.54	6	1
1:A:9:LEU:HD13	1:A:165:SER:HB3	0.41	1.92	30	1
1:A:59:LEU:HD13	1:A:59:LEU:C	0.41	2.36	10	1
1:A:114:GLU:OE1	1:A:118:LYS:NZ	0.40	2.54	19	1
1:A:68:ILE:H	1:A:68:ILE:HD12	0.40	1.75	29	1
1:A:150:LYS:NZ	1:A:161:ASP:OD2	0.40	2.54	1	1
1:A:122:ALA:HB3	1:A:123:ILE:HD12	0.40	1.92	10	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	142/174 (82%)	115±3 (81±2%)	20±3 (14±2%)	7±2 (5±1%)	4	25
All	All	4260/5220 (82%)	3447 (81%)	598 (14%)	215 (5%)	4	25

All 24 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	99	ASP	29

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	68	ILE	27
1	A	121	LYS	23
1	A	49	CYS	18
1	A	172	THR	18
1	A	52	MET	11
1	A	107	TYR	10
1	A	123	ILE	10
1	A	78	ASP	9
1	A	159	PRO	7
1	A	105	THR	7
1	A	163	ILE	6
1	A	67	ASN	6
1	A	113	GLU	5
1	A	162	ASP	5
1	A	83	LYS	5
1	A	7	ASP	4
1	A	144	PRO	4
1	A	160	TYR	3
1	A	171	SER	3
1	A	21	GLN	2
1	A	108	SER	1
1	A	145	ASP	1
1	A	51	PRO	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	129/159 (81%)	112±2 (87±2%)	17±2 (13±2%)	7 48
All	All	3870/4770 (81%)	3358 (87%)	512 (13%)	7 48

All 78 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	9	LEU	30
1	A	11	TYR	30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	117	LEU	30
1	A	121	LYS	30
1	A	63	LEU	27
1	A	99	ASP	18
1	A	168	LYS	18
1	A	145	ASP	14
1	A	13	VAL	14
1	A	67	ASN	14
1	A	42	PHE	13
1	A	112	ILE	12
1	A	34	GLU	12
1	A	139	PHE	12
1	A	64	LYS	12
1	A	82	ASP	10
1	A	72	ILE	10
1	A	22	ASP	10
1	A	28	LEU	9
1	A	41	ILE	9
1	A	164	ILE	9
1	A	81	ASN	9
1	A	102	ASP	7
1	A	53	THR	6
1	A	138	SER	6
1	A	86	GLN	6
1	A	107	TYR	6
1	A	111	GLU	6
1	A	7	ASP	5
1	A	37	LEU	5
1	A	70	VAL	5
1	A	108	SER	5
1	A	162	ASP	5
1	A	151	ASP	4
1	A	39	ASP	4
1	A	58	ASP	4
1	A	32	LYS	4
1	A	110	SER	4
1	A	24	LYS	4
1	A	85	LYS	4
1	A	147	LYS	4
1	A	160	TYR	3
1	A	57	THR	3
1	A	118	LYS	3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	10	ASN	3
1	A	21	GLN	3
1	A	109	GLN	3
1	A	59	LEU	3
1	A	152	TYR	3
1	A	60	GLN	3
1	A	87	LEU	3
1	A	62	LYS	3
1	A	56	MET	2
1	A	43	THR	2
1	A	30	SER	2
1	A	153	ASN	2
1	A	169	SER	2
1	A	52	MET	2
1	A	29	GLU	2
1	A	14	GLU	1
1	A	40	PHE	1
1	A	74	SER	1
1	A	114	GLU	1
1	A	93	ASN	1
1	A	94	TYR	1
1	A	150	LYS	1
1	A	12	GLU	1
1	A	97	SER	1
1	A	83	LYS	1
1	A	149	LEU	1
1	A	78	ASP	1
1	A	76	SER	1
1	A	71	ARG	1
1	A	89	LYS	1
1	A	88	LYS	1
1	A	123	ILE	1
1	A	80	GLU	1
1	A	140	TYR	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

No chemical shift data were provided