Full wwPDB NMR Structure Validation Report (i)

Mar 5, 2022 – 02:59 PM EST

PDB ID	:	2K6V
Title	:	Solution structures of apo Sco1 protein from Thermus Thermophilus
Authors	:	Abriata, L.A.; Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Gkazonis, P.; Spyroulias, G.A.; Vila, A.J.; Wang, S.
Deposited on	:	2008-07-28

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:

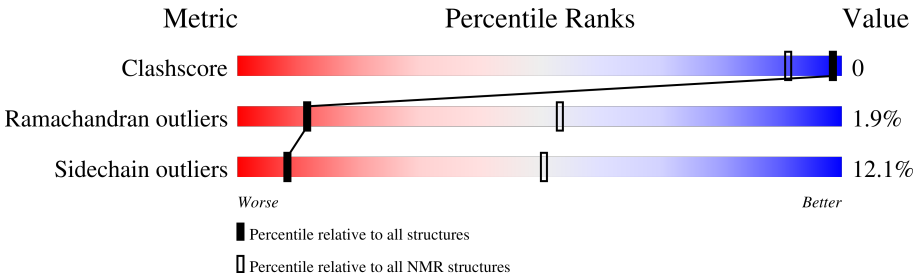
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)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

1 Overall quality at a glance

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	172	83% 8% 9%

2 Ensemble composition and analysis

This entry contains 31 models. Model 1 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:124, A:135-A:172 (156)	0.47	1

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 2 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3, 4, 5, 6, 8, 9, 10, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 30, 31
2	7, 11, 12, 24, 28, 29

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2714 atoms, of which 1359 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Putative cytochrome c oxidase assembly protein.

Mol	Chain	Residues	Atoms						Trace
1	A	172	Total	C	H	N	O	S	0
			2714	875	1359	230	247	3	

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	expression tag	UNP Q5SGY8
A	2	ALA	-	expression tag	UNP Q5SGY8
A	3	MET	-	expression tag	UNP Q5SGY8

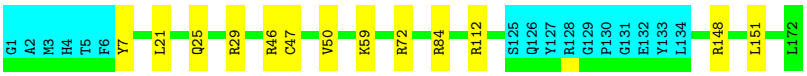
4 Residue-property plots

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: Putative cytochrome c oxidase assembly protein

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 (medoid)

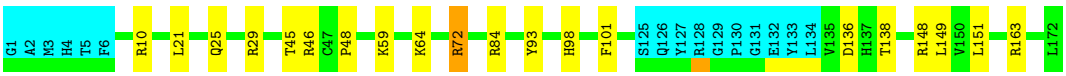
• Molecule 1: Putative cytochrome c oxidase assembly protein

Chain A:

4.2.2 Score per residue for model 2

• Molecule 1: Putative cytochrome c oxidase assembly protein

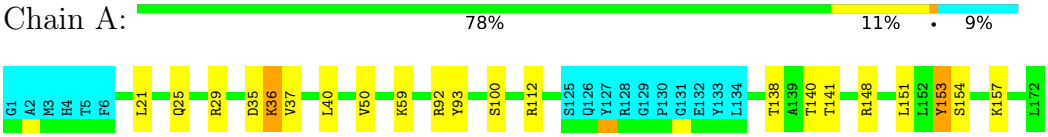
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Putative cytochrome c oxidase assembly protein

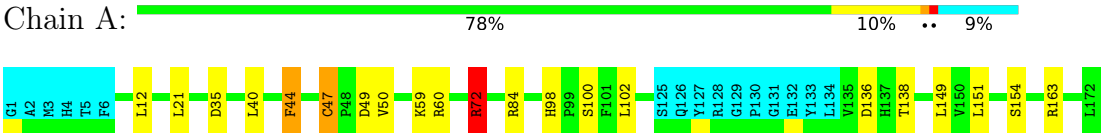
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Putative cytochrome c oxidase assembly protein

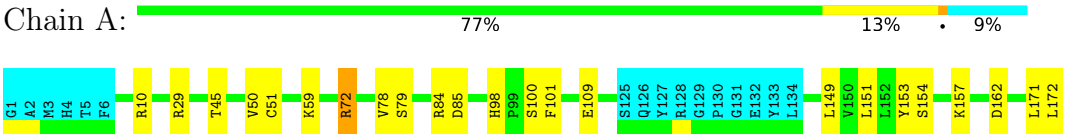
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Putative cytochrome c oxidase assembly protein

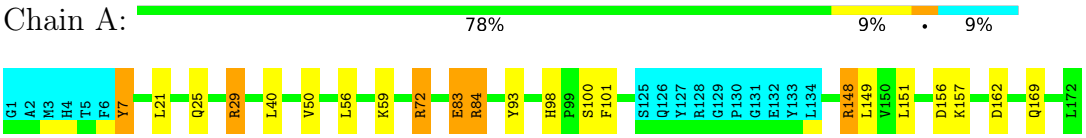
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Putative cytochrome c oxidase assembly protein

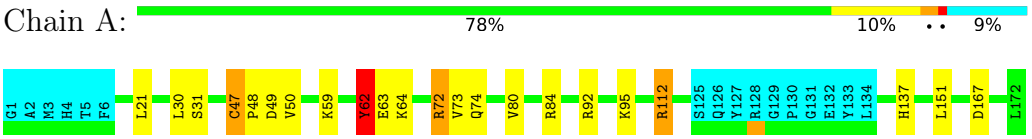
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Putative cytochrome c oxidase assembly protein

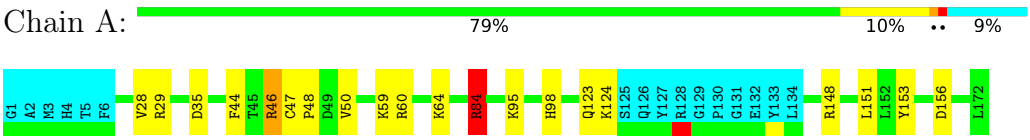
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Putative cytochrome c oxidase assembly protein

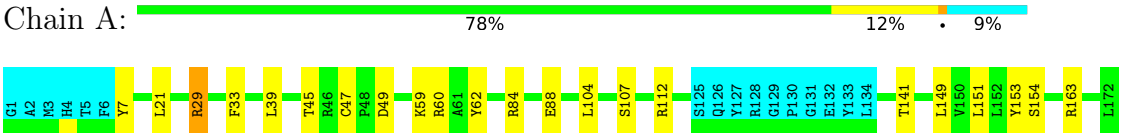
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: Putative cytochrome c oxidase assembly protein

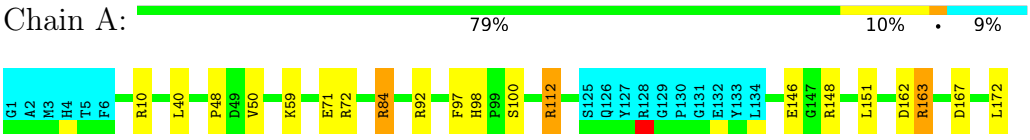
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: Putative cytochrome c oxidase assembly protein

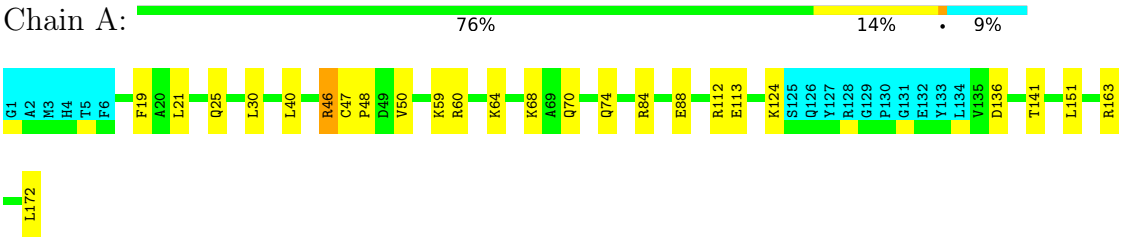
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: Putative cytochrome c oxidase assembly protein

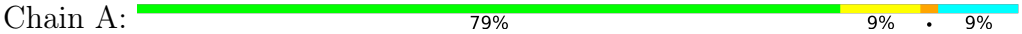
Chain A:



4.2.12 Score per residue for model 12

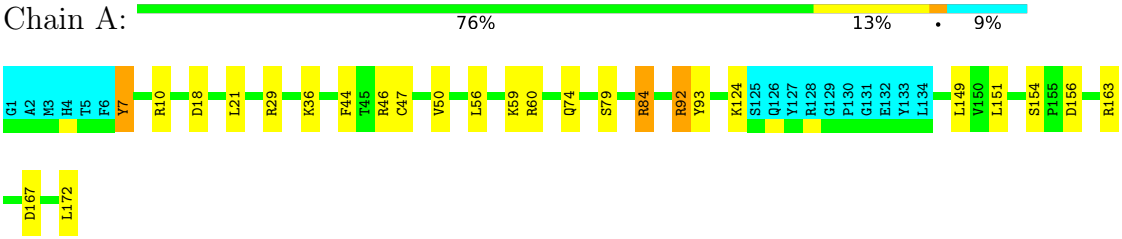
• Molecule 1: Putative cytochrome c oxidase assembly protein

Chain A:



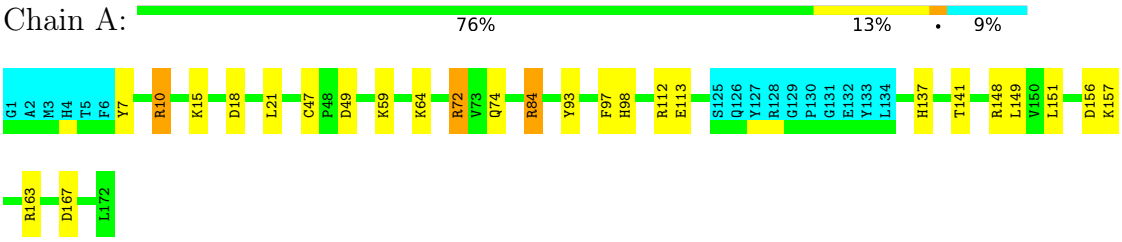
4.2.13 Score per residue for model 13

• Molecule 1: Putative cytochrome c oxidase assembly protein



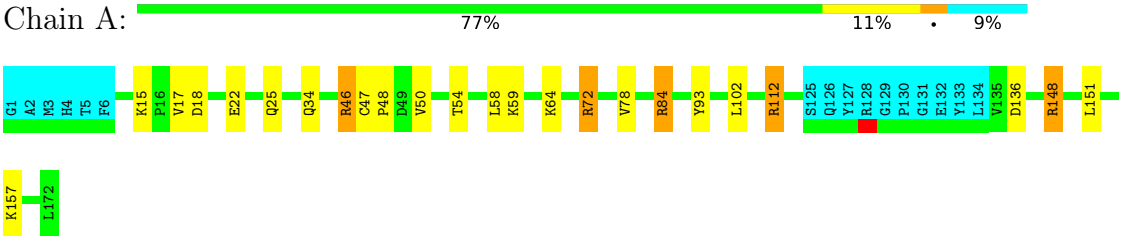
4.2.14 Score per residue for model 14

• Molecule 1: Putative cytochrome c oxidase assembly protein



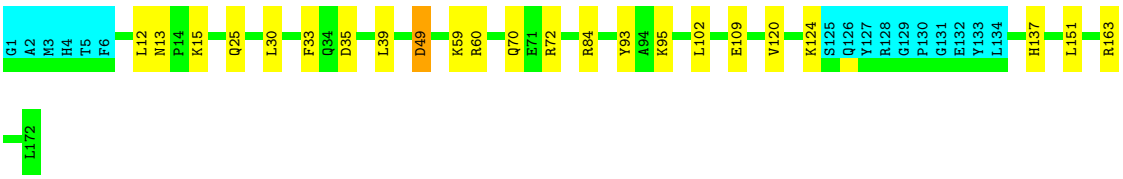
4.2.15 Score per residue for model 15

• Molecule 1: Putative cytochrome c oxidase assembly protein



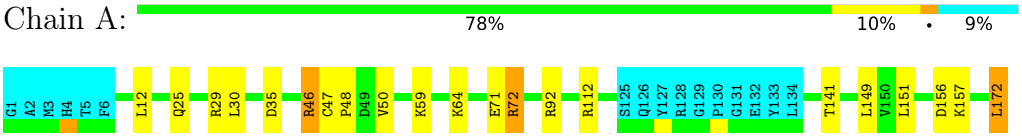
4.2.16 Score per residue for model 16

• Molecule 1: Putative cytochrome c oxidase assembly protein



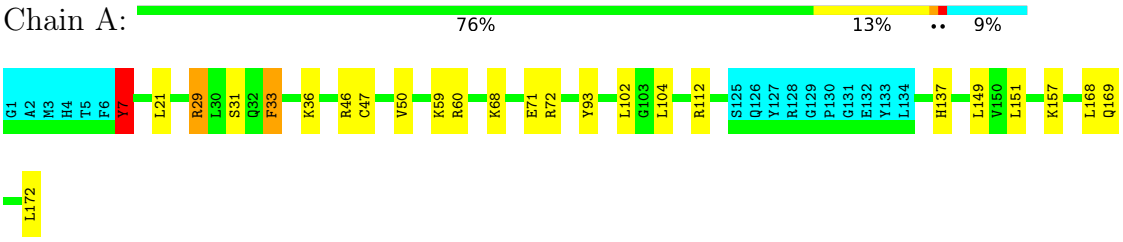
4.2.17 Score per residue for model 17

• Molecule 1: Putative cytochrome c oxidase assembly protein



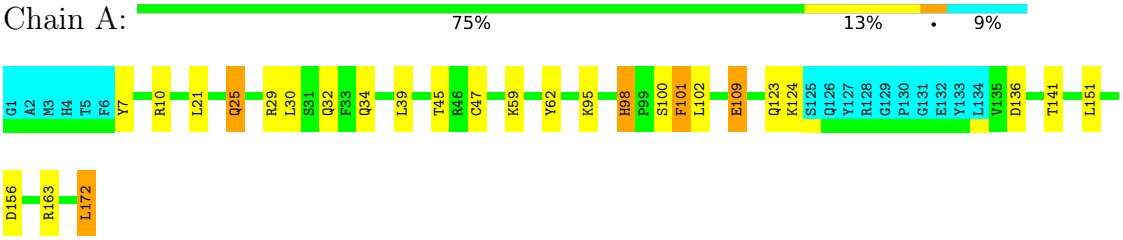
4.2.18 Score per residue for model 18

• Molecule 1: Putative cytochrome c oxidase assembly protein



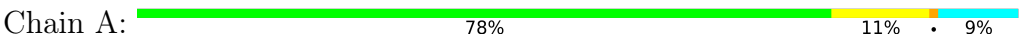
4.2.19 Score per residue for model 19

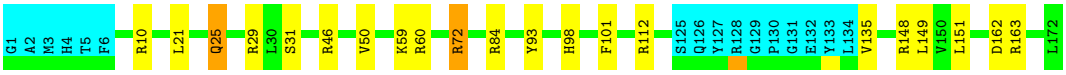
• Molecule 1: Putative cytochrome c oxidase assembly protein



4.2.20 Score per residue for model 20

• Molecule 1: Putative cytochrome c oxidase assembly protein

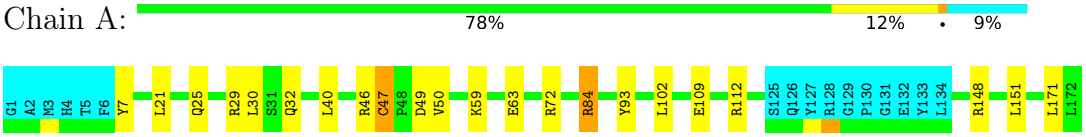




4.2.21 Score per residue for model 21

• Molecule 1: Putative cytochrome c oxidase assembly protein

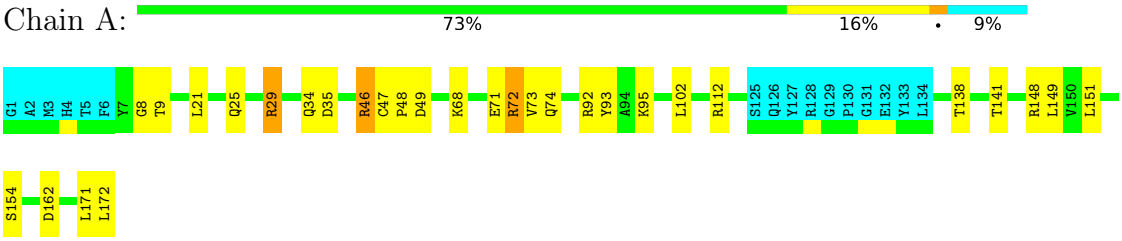
Chain A:



4.2.22 Score per residue for model 22

• Molecule 1: Putative cytochrome c oxidase assembly protein

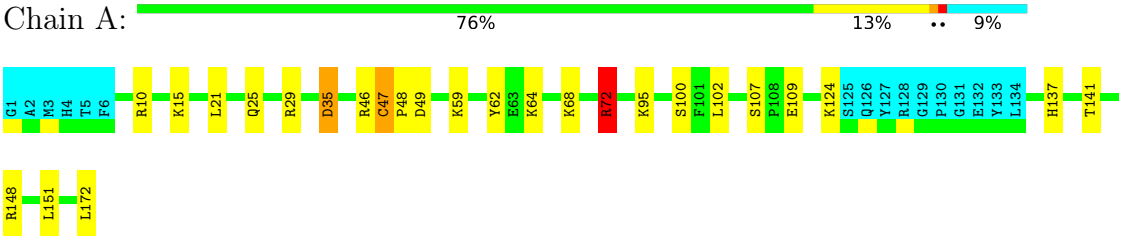
Chain A:



4.2.23 Score per residue for model 23

• Molecule 1: Putative cytochrome c oxidase assembly protein

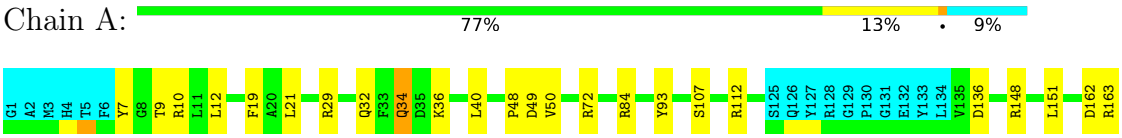
Chain A:



4.2.24 Score per residue for model 24

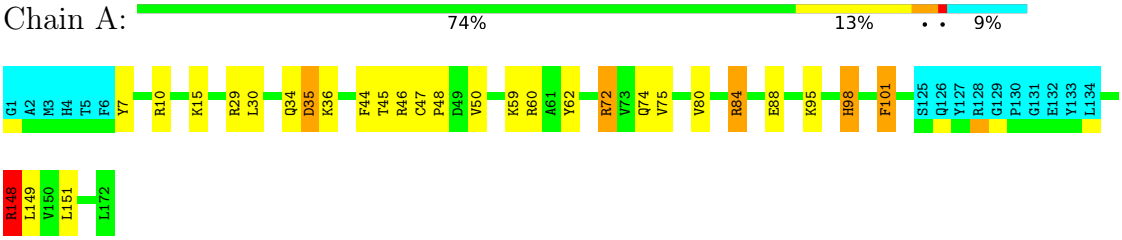
• Molecule 1: Putative cytochrome c oxidase assembly protein

Chain A:



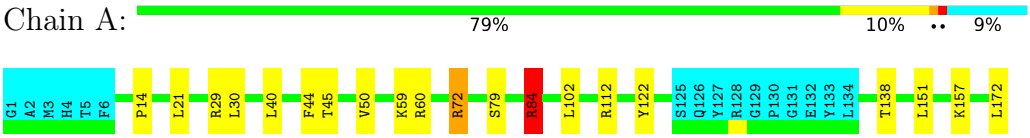
4.2.25 Score per residue for model 25

• Molecule 1: Putative cytochrome c oxidase assembly protein



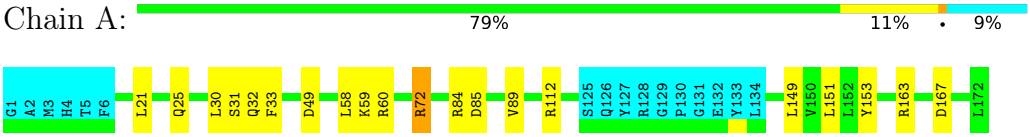
4.2.26 Score per residue for model 26

• Molecule 1: Putative cytochrome c oxidase assembly protein



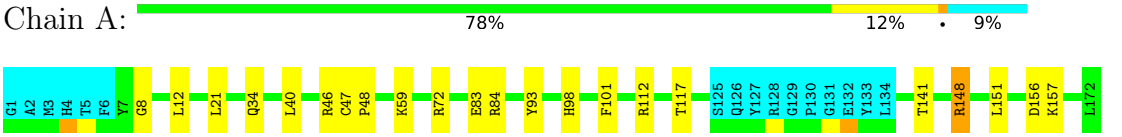
4.2.27 Score per residue for model 27

• Molecule 1: Putative cytochrome c oxidase assembly protein



4.2.28 Score per residue for model 28

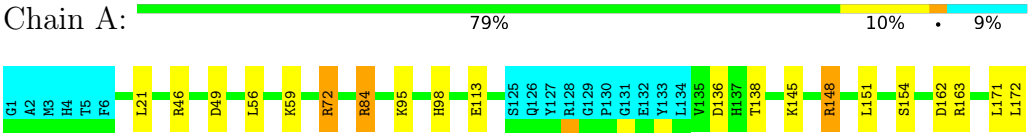
• Molecule 1: Putative cytochrome c oxidase assembly protein



4.2.29 Score per residue for model 29

• Molecule 1: Putative cytochrome c oxidase assembly protein

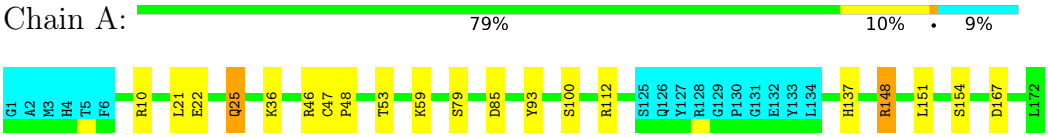
Chain A:



4.2.30 Score per residue for model 30

• Molecule 1: Putative cytochrome c oxidase assembly protein

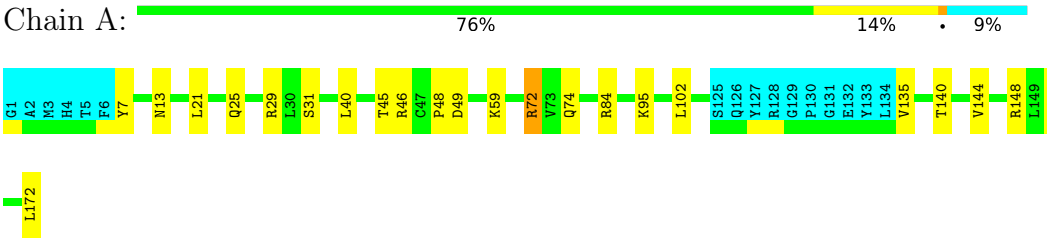
Chain A:



4.2.31 Score per residue for model 31

• Molecule 1: Putative cytochrome c oxidase assembly protein

Chain A:



5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics*.

Of the 31 calculated structures, 31 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
CYANA	structure solution	
Amber	refinement	

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	0.69±0.01	0±0/1259 (0.0± 0.0%)	1.10±0.03	5±2/1713 (0.3± 0.1%)
All	All	0.69	0/39029 (0.0%)	1.10	167/53103 (0.3%)

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	1.9±1.3
All	All	0	58

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	112	ARG	NE-CZ-NH1	13.81	127.20	120.30	10	14
1	A	72	ARG	NE-CZ-NH2	-11.46	114.57	120.30	14	15
1	A	84	ARG	NE-CZ-NH1	10.81	125.71	120.30	13	15
1	A	72	ARG	NE-CZ-NH1	10.43	125.51	120.30	10	18
1	A	84	ARG	NE-CZ-NH2	-9.58	115.51	120.30	20	4
1	A	148	ARG	NE-CZ-NH1	8.75	124.67	120.30	21	11
1	A	46	ARG	NE-CZ-NH1	8.19	124.40	120.30	13	11
1	A	29	ARG	NE-CZ-NH1	7.96	124.28	120.30	9	16
1	A	35	ASP	CB-CG-OD2	-7.81	111.27	118.30	3	1
1	A	163	ARG	NE-CZ-NH1	7.64	124.12	120.30	27	12
1	A	92	ARG	NE-CZ-NH1	7.46	124.03	120.30	7	6
1	A	112	ARG	NE-CZ-NH2	-7.13	116.73	120.30	10	7
1	A	112	ARG	CD-NE-CZ	6.86	133.21	123.60	10	1
1	A	10	ARG	NE-CZ-NH1	6.86	123.73	120.30	12	9
1	A	60	ARG	NE-CZ-NH1	6.72	123.66	120.30	4	10

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	60	ARG	NE-CZ-NH2	-6.54	117.03	120.30	13	4
1	A	62	TYR	CB-CG-CD2	-6.21	117.27	121.00	7	1
1	A	46	ARG	NE-CZ-NH2	-5.84	117.38	120.30	8	1
1	A	148	ARG	CD-NE-CZ	5.71	131.59	123.60	3	1
1	A	148	ARG	NE-CZ-NH2	-5.51	117.54	120.30	15	3
1	A	72	ARG	CD-NE-CZ	5.42	131.19	123.60	5	2
1	A	63	GLU	OE1-CD-OE2	-5.29	116.95	123.30	21	1
1	A	153	TYR	CB-CG-CD2	-5.26	117.85	121.00	3	1
1	A	172	LEU	CB-CA-C	5.25	120.17	110.20	17	1
1	A	101	PHE	CB-CG-CD2	-5.08	117.24	120.80	28	1
1	A	101	PHE	CB-CG-CD1	5.07	124.35	120.80	28	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	47	CYS	Peptide	8
1	A	148	ARG	Sidechain	7
1	A	93	TYR	Sidechain	5
1	A	153	TYR	Sidechain,Peptide	3
1	A	72	ARG	Sidechain	3
1	A	84	ARG	Sidechain	3
1	A	171	LEU	Peptide	3
1	A	97	PHE	Sidechain	3
1	A	124	LYS	Peptide	3
1	A	44	PHE	Peptide	2
1	A	163	ARG	Sidechain	2
1	A	7	TYR	Sidechain	2
1	A	62	TYR	Sidechain	2
1	A	36	LYS	Peptide	1
1	A	83	GLU	Peptide	1
1	A	46	ARG	Peptide	1
1	A	60	ARG	Sidechain	1
1	A	17	VAL	Peptide	1
1	A	29	ARG	Sidechain	1
1	A	35	ASP	Peptide	1
1	A	80	VAL	Peptide	1
1	A	14	PRO	Peptide	1
1	A	122	TYR	Sidechain	1
1	A	137	HIS	Sidechain	1

6.2 Too-close contacts

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	1228	1242	1240	0±1
All	All	38068	38502	38440	8

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:98:HIS:CE1	1:A:101:PHE:CD2	0.53	2.97	25	2
1:A:80:VAL:HG21	1:A:137:HIS:CD2	0.50	2.40	7	1
1:A:172:LEU:C	1:A:172:LEU:HD12	0.49	2.27	19	1
1:A:62:TYR:CE2	1:A:73:VAL:HB	0.45	2.46	7	1
1:A:75:VAL:HG12	1:A:101:PHE:CE1	0.43	2.48	25	1
1:A:171:LEU:C	1:A:171:LEU:HD12	0.42	2.36	21	1
1:A:109:GLU:H	1:A:109:GLU:CD	0.41	2.18	19	1

6.3 Torsion angles

6.3.1 Protein backbone

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	155/172 (90%)	140±3 (90±2%)	12±3 (8±2%)	3±1 (2±1%)	11	53
All	All	4805/5332 (90%)	4339 (90%)	373 (8%)	93 (2%)	11	53

All 21 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	48	PRO	15

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	25	GLN	12
1	A	50	VAL	11
1	A	49	ASP	10
1	A	47	CYS	6
1	A	7	TYR	6
1	A	45	THR	5
1	A	46	ARG	4
1	A	154	SER	3
1	A	35	ASP	3
1	A	84	ARG	3
1	A	138	THR	2
1	A	44	PHE	2
1	A	34	GLN	2
1	A	31	SER	2
1	A	8	GLY	2
1	A	12	LEU	1
1	A	83	GLU	1
1	A	156	ASP	1
1	A	157	LYS	1
1	A	33	PHE	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	133/145 (92%)	117±3 (88±2%)	16±3 (12±2%)	8 51
All	All	4123/4495 (92%)	3625 (88%)	498 (12%)	8 51

All 92 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	151	LEU	31
1	A	59	LYS	29
1	A	21	LEU	24
1	A	72	ARG	20
1	A	84	ARG	14

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	149	LEU	14
1	A	40	LEU	12
1	A	98	HIS	12
1	A	172	LEU	12
1	A	95	LYS	11
1	A	141	THR	10
1	A	157	LYS	10
1	A	102	LEU	10
1	A	100	SER	9
1	A	93	TYR	9
1	A	30	LEU	9
1	A	47	CYS	8
1	A	148	ARG	8
1	A	64	LYS	8
1	A	7	TYR	8
1	A	50	VAL	8
1	A	136	ASP	7
1	A	25	GLN	7
1	A	162	ASP	7
1	A	74	GLN	7
1	A	46	ARG	6
1	A	154	SER	6
1	A	10	ARG	6
1	A	101	PHE	6
1	A	36	LYS	6
1	A	156	ASP	6
1	A	167	ASP	6
1	A	79	SER	5
1	A	109	GLU	5
1	A	29	ARG	5
1	A	35	ASP	5
1	A	15	LYS	5
1	A	34	GLN	5
1	A	138	THR	4
1	A	33	PHE	4
1	A	49	ASP	4
1	A	71	GLU	4
1	A	68	LYS	4
1	A	113	GLU	4
1	A	92	ARG	4
1	A	137	HIS	4
1	A	12	LEU	4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	32	GLN	4
1	A	22	GLU	3
1	A	85	ASP	3
1	A	56	LEU	3
1	A	31	SER	3
1	A	62	TYR	3
1	A	112	ARG	3
1	A	124	LYS	3
1	A	39	LEU	3
1	A	88	GLU	3
1	A	107	SER	3
1	A	18	ASP	3
1	A	60	ARG	2
1	A	45	THR	2
1	A	140	THR	2
1	A	78	VAL	2
1	A	153	TYR	2
1	A	169	GLN	2
1	A	44	PHE	2
1	A	123	GLN	2
1	A	104	LEU	2
1	A	19	PHE	2
1	A	70	GLN	2
1	A	144	VAL	2
1	A	58	LEU	2
1	A	13	ASN	2
1	A	135	VAL	2
1	A	9	THR	2
1	A	37	VAL	1
1	A	51	CYS	1
1	A	63	GLU	1
1	A	28	VAL	1
1	A	146	GLU	1
1	A	163	ARG	1
1	A	159	GLU	1
1	A	54	THR	1
1	A	120	VAL	1
1	A	168	LEU	1
1	A	73	VAL	1
1	A	89	VAL	1
1	A	83	GLU	1
1	A	117	THR	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	145	LYS	1
1	A	53	THR	1
1	A	150	VAL	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

There are no monosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

6.7 Other polymers

There are no such molecules in this entry.

6.8 Polymer linkage issues

There are no chain breaks in this entry.

7 Chemical shift validation

No chemical shift data were provided