



# Full wwPDB NMR Structure Validation Report ⓘ

Feb 12, 2017 – 10:34 pm GMT

PDB ID : 2HV4  
Title : NMR solution structure refinement of yeast iso-1-ferrocytochrome c  
Authors : Assfalg, M.; Bertini, I.; Del Conte, R.; Turano, P.  
Deposited on : 2006-07-27

This is a Full wwPDB NMR Structure Validation Report for a publicly released PDB entry.

We welcome your comments at [validation@mail.wwpdb.org](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/NMRValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.2 (RC1), CSD as538be (2017)
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	trunk28760
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

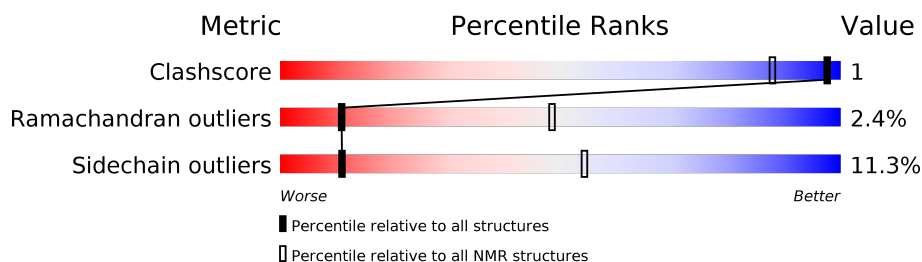
# 1 Overall quality at a glance

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	125131	11601
Ramachandran outliers	121729	10391
Sidechain outliers	121581	10367

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  $\geq 3$ , 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	108	

## 2 Ensemble composition and analysis

This entry contains 35 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:1-A:21, A:25-A:42, A:46-A:103 (97)	0.30	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 4 clusters and 10 single-model clusters were found.

Cluster number	Models
1	1, 4, 6, 7, 12, 18, 21, 28, 32, 34
2	2, 5, 8, 9, 13, 14, 22, 30, 33
3	19, 20, 24, 27
4	3, 16
Single-model clusters	10; 11; 15; 17; 23; 25; 26; 29; 31; 35

### 3 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 1774 atoms, of which 883 are hydrogens and 0 are deuteriums.

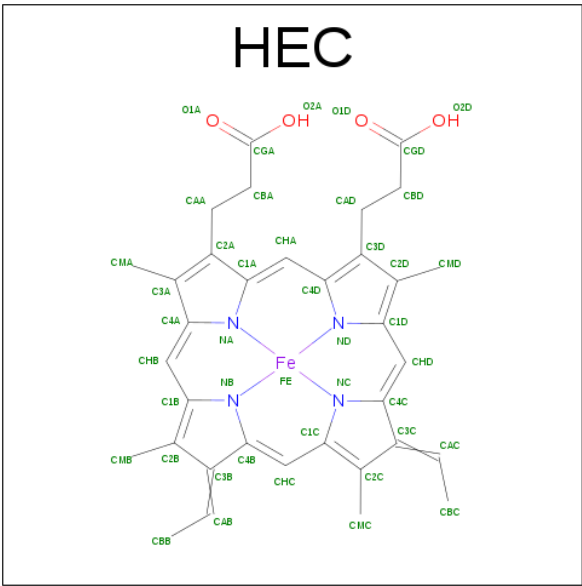
- Molecule 1 is a protein called Cytochrome c iso-1.

Mol	Chain	Residues	Atoms						Trace
1	A	108	Total	C	H	N	O	S	0
			1699	535	851	151	158	4	

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	102	THR	CYS	ENGINEERED	UNP P00044

- Molecule 2 is HEME C (three-letter code: HEC) (formula:  $C_{34}H_{34}FeN_4O_4$ ).



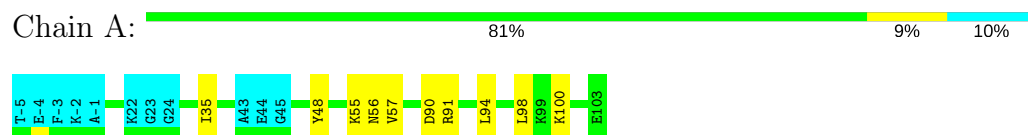
Mol	Chain	Residues	Atoms					
2	A	1	Total	C	Fe	H	N	O
			75	34	1	32	4	4

## 4 Residue-property plots

### 4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: Cytochrome c iso-1

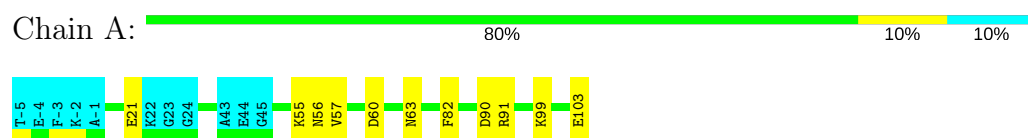


### 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

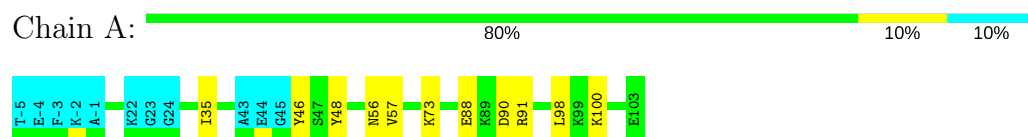
#### 4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Cytochrome c iso-1



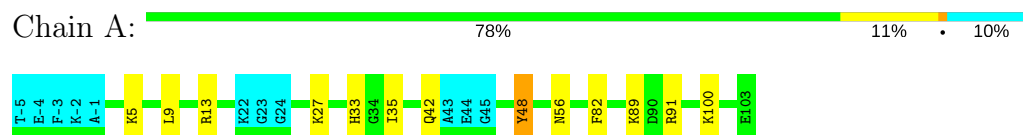
#### 4.2.2 Score per residue for model 2

- Molecule 1: Cytochrome c iso-1



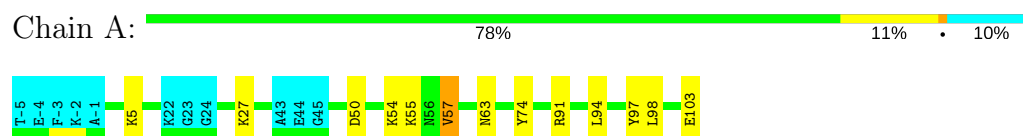
### 4.2.3 Score per residue for model 3

- Molecule 1: Cytochrome c iso-1



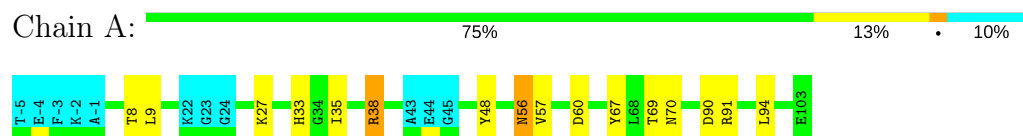
### 4.2.4 Score per residue for model 4

- Molecule 1: Cytochrome c iso-1



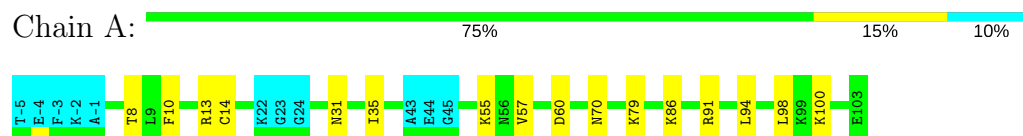
### 4.2.5 Score per residue for model 5

- Molecule 1: Cytochrome c iso-1



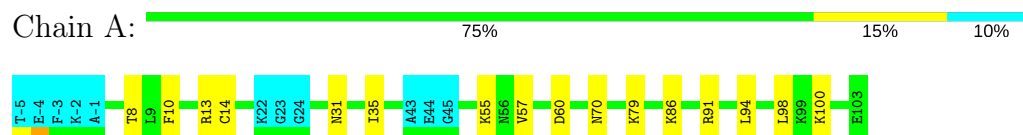
### 4.2.6 Score per residue for model 6

- Molecule 1: Cytochrome c iso-1



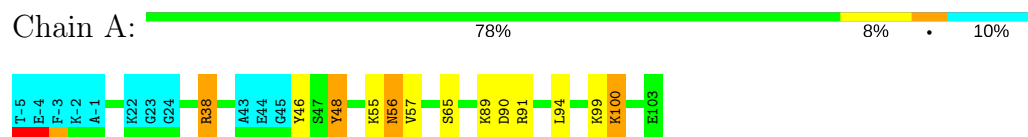
### 4.2.7 Score per residue for model 7

- Molecule 1: Cytochrome c iso-1



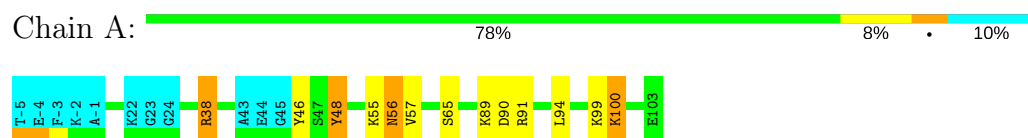
### 4.2.8 Score per residue for model 8

- Molecule 1: Cytochrome c iso-1



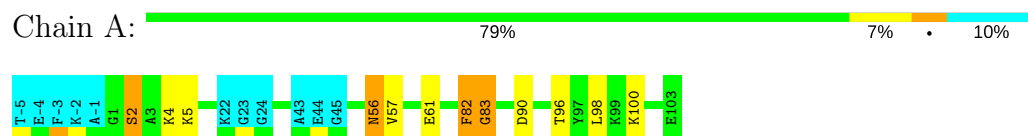
### 4.2.9 Score per residue for model 9

- Molecule 1: Cytochrome c iso-1



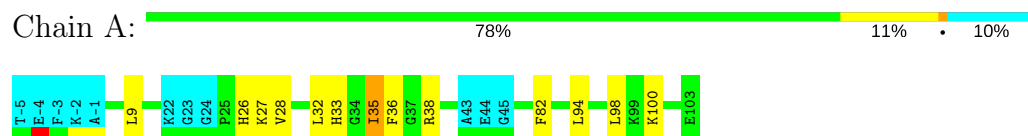
### 4.2.10 Score per residue for model 10

- Molecule 1: Cytochrome c iso-1



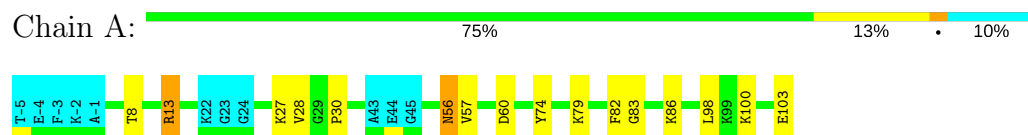
### 4.2.11 Score per residue for model 11

- Molecule 1: Cytochrome c iso-1



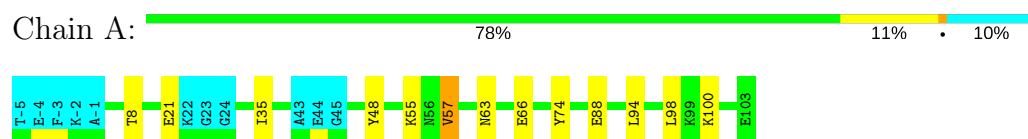
### 4.2.12 Score per residue for model 12

- Molecule 1: Cytochrome c iso-1



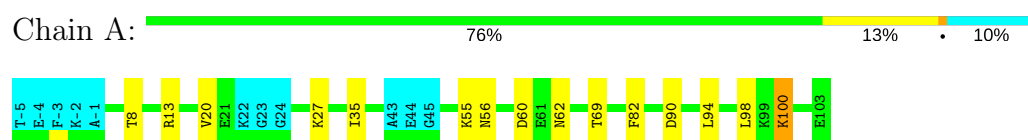
### 4.2.13 Score per residue for model 13

- Molecule 1: Cytochrome c iso-1



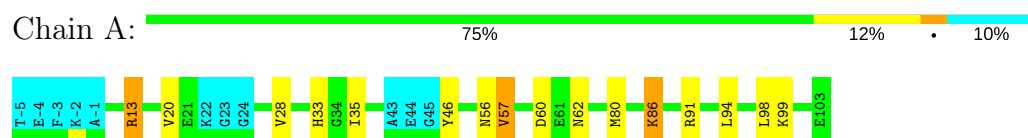
### 4.2.14 Score per residue for model 14

- Molecule 1: Cytochrome c iso-1



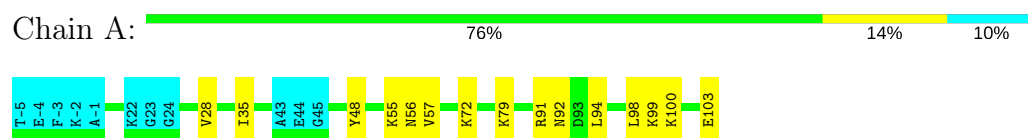
### 4.2.15 Score per residue for model 15

- Molecule 1: Cytochrome c iso-1



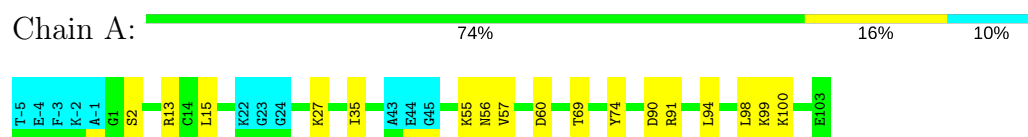
### 4.2.16 Score per residue for model 16

- Molecule 1: Cytochrome c iso-1



### 4.2.17 Score per residue for model 17

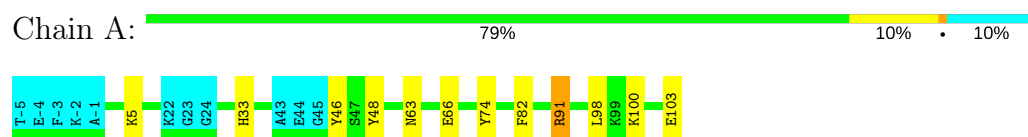
- Molecule 1: Cytochrome c iso-1





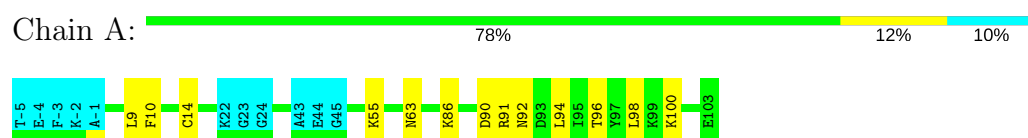
#### 4.2.18 Score per residue for model 18

- Molecule 1: Cytochrome c iso-1



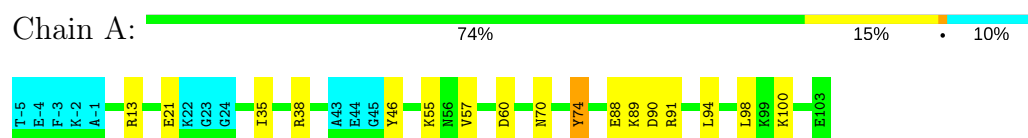
#### 4.2.19 Score per residue for model 19

- Molecule 1: Cytochrome c iso-1



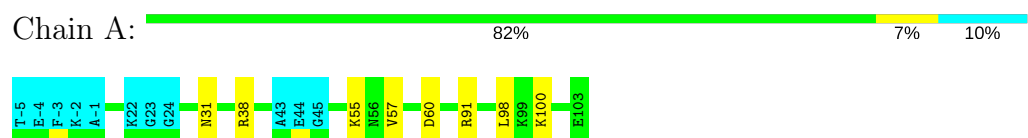
#### 4.2.20 Score per residue for model 20

- Molecule 1: Cytochrome c iso-1



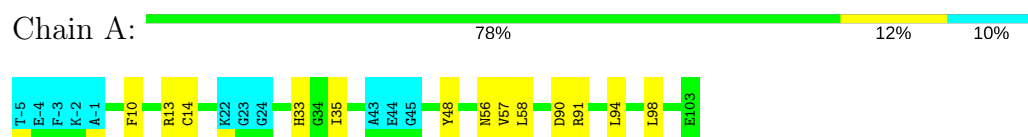
#### 4.2.21 Score per residue for model 21

- Molecule 1: Cytochrome c iso-1



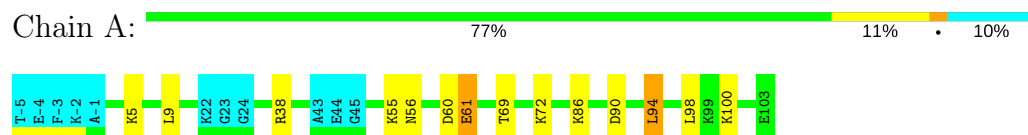
#### 4.2.22 Score per residue for model 22

- Molecule 1: Cytochrome c iso-1



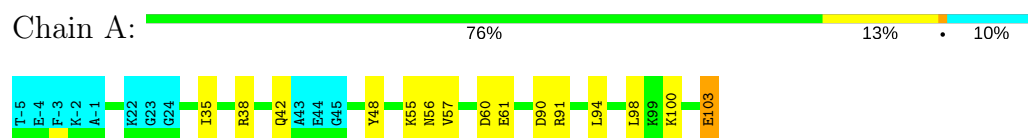
#### 4.2.23 Score per residue for model 23

- Molecule 1: Cytochrome c iso-1



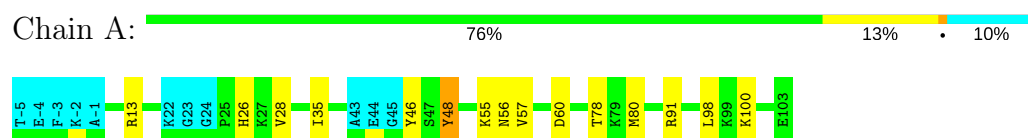
#### 4.2.24 Score per residue for model 24

- Molecule 1: Cytochrome c iso-1



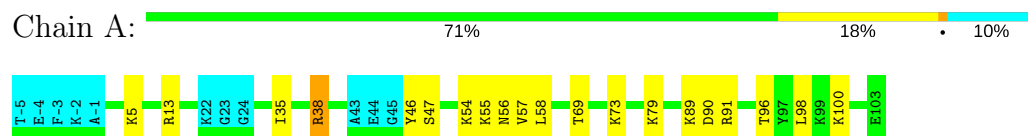
#### 4.2.25 Score per residue for model 25

- Molecule 1: Cytochrome c iso-1



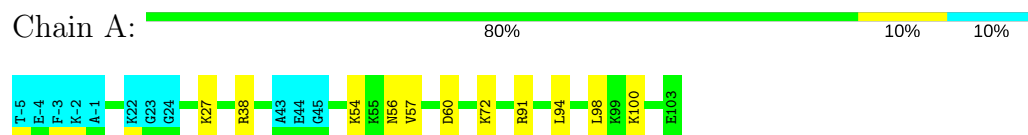
#### 4.2.26 Score per residue for model 26

- Molecule 1: Cytochrome c iso-1



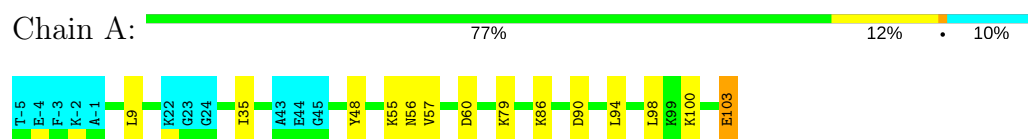
#### 4.2.27 Score per residue for model 27

- Molecule 1: Cytochrome c iso-1



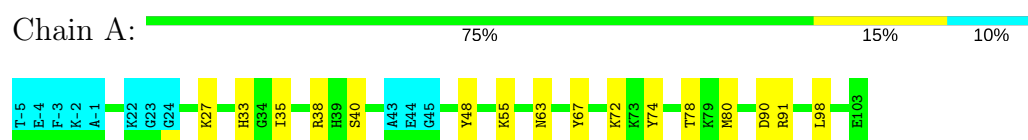
#### 4.2.28 Score per residue for model 28

- Molecule 1: Cytochrome c iso-1



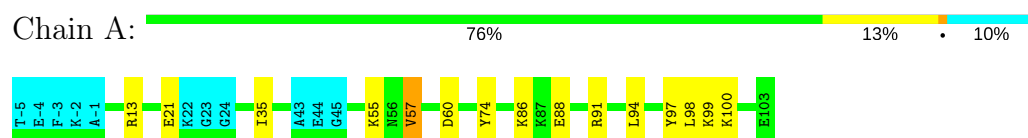
#### 4.2.29 Score per residue for model 29

- Molecule 1: Cytochrome c iso-1



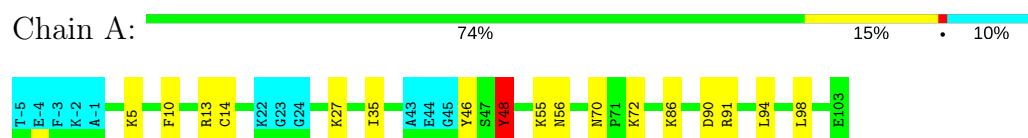
#### 4.2.30 Score per residue for model 30

- Molecule 1: Cytochrome c iso-1



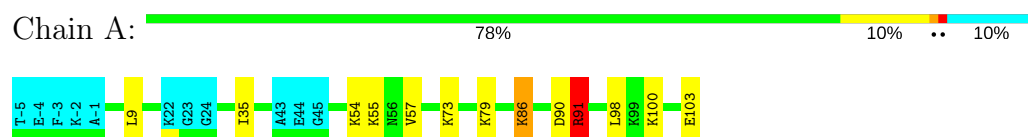
#### 4.2.31 Score per residue for model 31

- Molecule 1: Cytochrome c iso-1



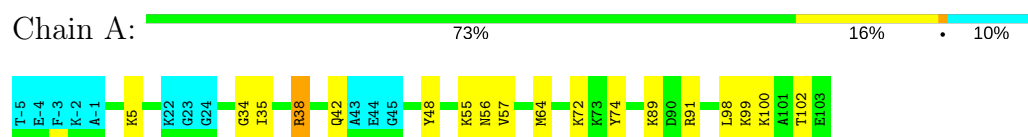
#### 4.2.32 Score per residue for model 32

- Molecule 1: Cytochrome c iso-1



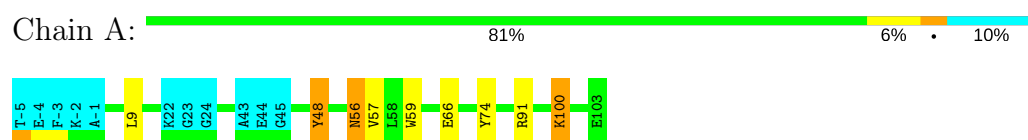
### 4.2.33 Score per residue for model 33

- Molecule 1: Cytochrome c iso-1



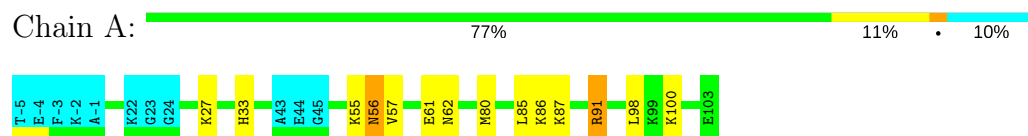
### 4.2.34 Score per residue for model 34

- Molecule 1: Cytochrome c iso-1



### 4.2.35 Score per residue for model 35

- Molecule 1: Cytochrome c iso-1



## 5 Refinement protocol and experimental data overview

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

Of the 500 calculated structures, 35 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
AMBER	refinement	8.0

No chemical shift data was provided. No validations of the models with respect to experimental NMR restraints is performed at this time.

## 6 Model quality i

### 6.1 Standard geometry i

Bond lengths and bond angles in the following residue types are not validated in this section: HEC

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	0.70±0.01	0±0/789 (0.0±0.0%)	1.06±0.03	2±1/1056 (0.1±0.1%)
All	All	0.70	0/27615 (0.0%)	1.06	53/36960 (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	1.1±1.0
All	All	0	37

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	91	ARG	NE-CZ-NH1	10.03	125.31	120.30	2	18
1	A	38	ARG	NE-CZ-NH1	7.07	123.84	120.30	11	8
1	A	13	ARG	NE-CZ-NH1	6.90	123.75	120.30	6	10
1	A	91	ARG	NE-CZ-NH2	6.65	123.62	120.30	27	2
1	A	100	LYS	CB-CG-CD	5.79	126.65	111.60	14	3
1	A	91	ARG	CD-NE-CZ	5.76	131.67	123.60	32	2
1	A	46	TYR	CB-CG-CD1	-5.59	117.65	121.00	26	1
1	A	48	TYR	CB-CG-CD2	-5.52	117.69	121.00	31	1
1	A	57	VAL	CA-CB-CG1	5.38	118.97	110.90	15	2
1	A	103	GLU	OE1-CD-OE2	-5.30	116.94	123.30	28	2
1	A	38	ARG	NE-CZ-NH2	-5.14	117.73	120.30	8	2
1	A	13	ARG	CD-NE-CZ	5.12	130.78	123.60	20	1

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	100	LYS	CA-CB-CG	-5.02	102.35	113.40	14	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	48	TYR	Sidechain	11
1	A	46	TYR	Sidechain	7
1	A	13	ARG	Sidechain	4
1	A	91	ARG	Sidechain	4
1	A	74	TYR	Sidechain	3
1	A	67	TYR	Sidechain	2
1	A	97	TYR	Sidechain	2
1	A	42	GLN	Peptide	1
1	A	21	GLU	Peptide	1
1	A	100	LYS	Peptide	1
1	A	38	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	772	776	776	1±1
2	A	43	32	30	0±1
All	All	28525	28280	28210	33

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
2:A:104:HEC:HMD1	2:A:104:HEC:CGD	0.67	2.19	3	1
2:A:104:HEC:HMD1	2:A:104:HEC:HBD1	0.63	1.71	28	2
2:A:104:HEC:HBD1	2:A:104:HEC:HMD1	0.58	1.75	11	1

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:28:VAL:O	2:A:104:HEC:HMD3	0.53	2.03	16	5
2:A:104:HEC:CGD	2:A:104:HEC:HMD1	0.52	2.35	24	1
1:A:57:VAL:HG22	1:A:63:ASN:HD21	0.49	1.67	13	2
1:A:85:LEU:H	1:A:91:ARG:HH12	0.48	1.51	35	1
1:A:91:ARG:HA	1:A:91:ARG:HH11	0.47	1.69	32	1
1:A:55:LYS:HE2	1:A:74:TYR:CE2	0.47	2.45	30	2
1:A:10:PHE:CE2	1:A:14:CYS:HB2	0.46	2.45	7	4
1:A:99:LYS:HE2	1:A:103:GLU:CD	0.45	2.32	16	1
1:A:10:PHE:CD2	1:A:14:CYS:HB2	0.44	2.46	7	2
1:A:50:ASP:OD2	1:A:54:LYS:HE2	0.44	2.12	4	1
1:A:9:LEU:HD22	1:A:94:LEU:HD23	0.42	1.90	23	1
1:A:55:LYS:HE2	1:A:74:TYR:CE1	0.42	2.50	4	2
1:A:82:PHE:CG	1:A:83:GLY:N	0.41	2.88	10	1
1:A:10:PHE:CD1	1:A:14:CYS:HB2	0.41	2.51	22	1
1:A:79:LYS:HZ3	2:A:104:HEC:CGD	0.41	2.28	28	1
1:A:100:LYS:O	1:A:100:LYS:HE3	0.40	2.16	9	1
1:A:100:LYS:HE3	1:A:100:LYS:O	0.40	2.16	8	2

## 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	96/108 (89%)	81±3 (85±3%)	12±2 (13±2%)	2±2 (2±2%)	11	48
All	All	3360/3780 (89%)	2842 (85%)	436 (13%)	82 (2%)	11	48

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	56	ASN	20
1	A	48	TYR	11
1	A	86	LYS	9
1	A	27	LYS	7
1	A	82	PHE	7

Continued on next page...



*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	33	HIS	6
1	A	80	MET	4
1	A	54	LYS	2
1	A	26	HIS	2
1	A	2	SER	2
1	A	83	GLY	2
1	A	87	LYS	1
1	A	36	PHE	1
1	A	59	TRP	1
1	A	40	SER	1
1	A	20	VAL	1
1	A	35	ILE	1
1	A	4	LYS	1
1	A	30	PRO	1
1	A	34	GLY	1
1	A	47	SER	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	83/89 (93%)	74±2 (89±3%)	9±2 (11±3%)	11	54
All	All	2905/3115 (93%)	2577 (89%)	328 (11%)	11	54

All 48 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	98	LEU	29
1	A	57	VAL	27
1	A	100	LYS	25
1	A	35	ILE	22
1	A	94	LEU	21
1	A	55	LYS	20
1	A	90	ASP	18
1	A	60	ASP	16
1	A	56	ASN	10

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Models (Total)
1	A	5	LYS	8
1	A	38	ARG	8
1	A	103	GLU	7
1	A	91	ARG	7
1	A	99	LYS	7
1	A	9	LEU	7
1	A	89	LYS	6
1	A	72	LYS	6
1	A	79	LYS	6
1	A	8	THR	6
1	A	69	THR	5
1	A	70	ASN	5
1	A	74	TYR	4
1	A	27	LYS	4
1	A	61	GLU	4
1	A	88	GLU	4
1	A	86	LYS	4
1	A	63	ASN	4
1	A	21	GLU	3
1	A	31	ASN	3
1	A	62	ASN	3
1	A	66	GLU	3
1	A	96	THR	3
1	A	73	LYS	3
1	A	78	THR	2
1	A	65	SER	2
1	A	92	ASN	2
1	A	58	LEU	2
1	A	42	GLN	2
1	A	13	ARG	1
1	A	102	THR	1
1	A	32	LEU	1
1	A	54	LYS	1
1	A	15	LEU	1
1	A	33	HIS	1
1	A	64	MET	1
1	A	46	TYR	1
1	A	2	SER	1
1	A	20	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 carbohydrates in this entry.

## 6.6 Ligand geometry ⓘ

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	HEC	A	104	1	28,50,50	0.73±0.02	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with  $|Z| > 2$  is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	HEC	A	104	1	16,82,82	1.48±0.21	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means

no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	HEC	A	104	1	-	0±0,6,54,54	0±0,0,8,8

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

## 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