



# wwPDB NMR Structure Validation Summary Report ⓘ

May 28, 2020 – 10:48 pm BST

PDB ID : 2L0Y  
Title : Complex hMia40-hCox17  
Authors : Bertini, I.; Ciofi-Baffoni, S.; Gallo, A.  
Deposited on : 2010-07-20

This is a wwPDB NMR Structure Validation Summary 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

<https://www.wwpdb.org/validation/2017/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  
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 : 2.11  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : 2.11

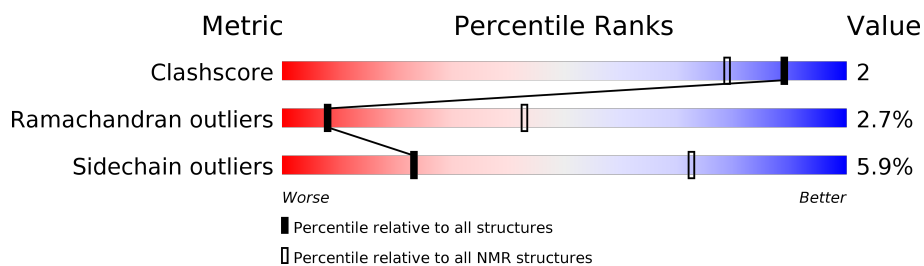
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*SOLUTION NMR*

The overall completeness of chemical shifts assignment is 68%.

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  $\geq 3$ , 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	146	
2	B	67	

## 2 Ensemble composition and analysis i

This entry contains 20 models. Model 14 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:3-A:7, A:11-A:53, B:66-B:77 (60)	0.39	14

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 3 single-model clusters were found.

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

### 3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 1214 atoms, of which 580 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Mitochondrial intermembrane space import and assembly protein 40.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	60	881	289	415	77	92	8	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-48	GLY	-	expression tag	UNP Q8N4Q1
A	-47	SER	-	expression tag	UNP Q8N4Q1
A	-46	PHE	-	expression tag	UNP Q8N4Q1
A	-45	THR	-	expression tag	UNP Q8N4Q1
A	8	SER	CYS	engineered mutation	UNP Q8N4Q1

- Molecule 2 is a protein called COX17 cytochrome c oxidase assembly homolog (*S. cerevisiae*) pseudogene (COX17).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	21	333	105	165	32	29	2	0

There are 5 discrepancies between the modelled and reference sequences:

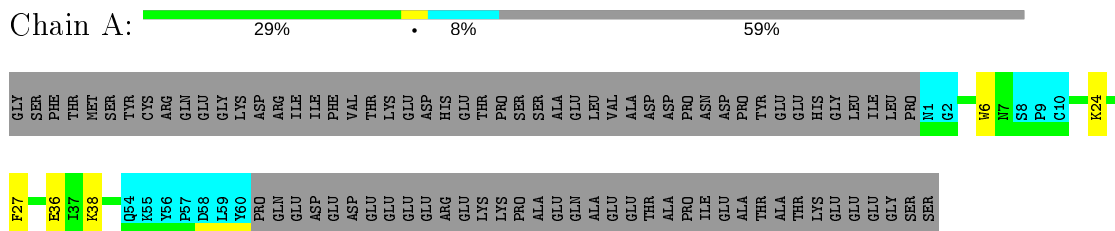
Chain	Residue	Modelled	Actual	Comment	Reference
B	15	GLY	-	expression tag	UNP Q5W0Q5
B	16	SER	-	expression tag	UNP Q5W0Q5
B	17	PHE	-	expression tag	UNP Q5W0Q5
B	18	THR	-	expression tag	UNP Q5W0Q5
B	73	SER	CYS	engineered mutation	UNP Q5W0Q5

## 4 Residue-property plots [i](#)

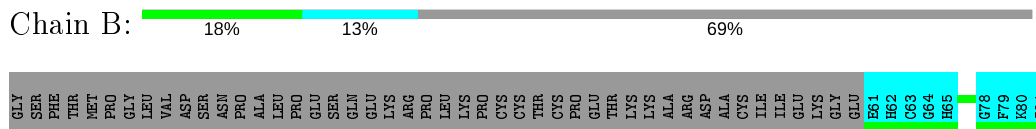
### 4.1 Average score per residue in the NMR ensemble

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: Mitochondrial intermembrane space import and assembly protein 40



- Molecule 2: COX17 cytochrome c oxidase assembly homolog (*S. cerevisiae*) pseudogene (COX17)



### 4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 14. Colouring as in section 4.1 above.

- Molecule 1: Mitochondrial intermembrane space import and assembly protein 40



- Molecule 2: COX17 cytochrome c oxidase assembly homolog (*S. cerevisiae*) pseudogene (COX17)



GLY	GLY
SER	H61
PHE	H62
THR	G63
MET	G64
PRO	H65
GLY	L66
LEU	I67
VAL	R75
ASP	G78
SER	F79
ASN	R80
PRO	I81
ALA	
LEU	
PRO	
GLU	
SER	
GLN	
GLU	
LYS	
ARG	
PRO	
LEU	
LYS	
PRO	
CYS	
THR	
CYS	
PRO	
GLU	
THR	
LYS	
LYS	
ALA	
ARG	
ASP	
ALA	
CYS	
ILE	
ILE	
GLU	
LYS	
GLY	
GLU	

## 5 Refinement protocol and experimental data overview (i)

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

Of the 400 calculated structures, 20 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
Amber	refinement	10.0
CYANA	structure solution	2.1

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	1
Total number of shifts	1135
Number of shifts mapped to atoms	1115
Number of unparsed shifts	0
Number of shifts with mapping errors	20
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	68%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

### 5.1 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	370	329	329	2±1
2	B	96	102	102	0±0
All	All	9320	8620	8620	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 2.

5 of 8 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:66:LEU:H	2:B:66:LEU:HD23	0.55	1.60	2	1
1:A:11:LEU:HD23	1:A:23:PHE:CE2	0.53	2.39	19	2
1:A:6:TRP:CH2	1:A:24:LYS:HA	0.52	2.40	3	13
1:A:4:ILE:HG21	1:A:11:LEU:HD22	0.49	1.85	19	1
1:A:15:ALA:HB1	1:A:24:LYS:CE	0.48	2.38	6	2

## 5.2 Torsion angles [i](#)

### 5.2.1 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	48/146 (33%)	42±1 (87±3%)	5±1 (10±3%)	2±1 (3±1%)	6	37
2	B	12/67 (18%)	12±0 (99±2%)	0±0 (1±3%)	0±0 (0±0%)	100	100
All	All	1200/4260 (28%)	1070 (89%)	98 (8%)	32 (3%)	8	43

All 5 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	36	GLU	15
1	A	38	LYS	12
1	A	5	ASN	3
1	A	3	ASN	1
1	A	35	GLU	1

### 5.2.2 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	40/126 (32%)	38±1 (96±3%)	2±1 (4±3%)	31 80
2	B	10/57 (18%)	9±1 (88±9%)	1±1 (11±9%)	9 52
All	All	1000/3660 (27%)	941 (94%)	59 (6%)	23 72

5 of 18 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	27	PHE	17
2	B	67	ILE	8
2	B	74	MET	6
1	A	38	LYS	5
1	A	51	GLU	3

### 5.2.3 RNA [i](#)

There are no RNA molecules in this entry.

### 5.3 Non-standard residues in protein, DNA, RNA chains [i](#)

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

### 5.4 Carbohydrates [i](#)

There are no carbohydrates in this entry.

### 5.5 Ligand geometry [i](#)

There are no ligands in this entry.

### 5.6 Other polymers [i](#)

There are no such molecules in this entry.

### 5.7 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

## 6 Chemical shift validation i

The completeness of assignment taking into account all chemical shift lists is 68% for the well-defined parts and 64% for the entire structure.

### 6.1 Chemical shift list 1

File name: input\_cs.cif

Chemical shift list name: *assigned\_chem\_shift\_list\_1*

#### 6.1.1 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	1135
Number of shifts mapped to atoms	1115
Number of unparsed shifts	0
Number of shifts with mapping errors	20
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	13

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

- Residue not found in structure. First 5 (of 20) occurrences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
B	22	LYS	C	171.6	0.3	1
B	30	SER	CB	60.6	0.3	1
B	30	SER	N	116.7	0.3	1
B	22	LYS	CB	29.7	0.3	1
B	40	SER	CB	60.9	0.3	1

#### 6.1.2 Chemical shift referencing i

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	155	$2.36 \pm 0.30$	Should be applied
$^{13}\text{C}_\beta$	144	$3.24 \pm 0.08$	Should be applied
$^{13}\text{C}'$	148	$2.33 \pm 0.30$	Should be applied

*Continued on next page...*

Continued from previous page...

Nucleus	# values	Correction $\pm$ precision, ppm	Suggested action
<sup>15</sup> N	148	0.33 $\pm$ 0.26	None needed (< 0.5 ppm)

### 6.1.3 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 68%, i.e. 487 atoms were assigned a chemical shift out of a possible 713. 2 out of 4 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	<sup>1</sup> H	<sup>13</sup> C	<sup>15</sup> N
Backbone	285/298 (96%)	107/119 (90%)	119/120 (99%)	59/59 (100%)
Sidechain	189/345 (55%)	101/206 (49%)	88/124 (71%)	0/15 (0%)
Aromatic	13/70 (19%)	13/38 (34%)	0/29 (0%)	0/3 (0%)
Overall	487/713 (68%)	221/363 (61%)	207/273 (76%)	59/77 (77%)

### 6.1.4 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.

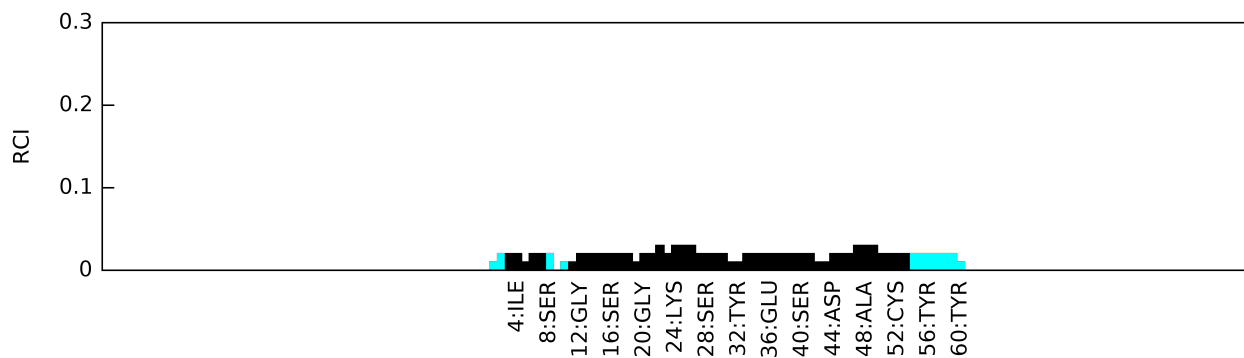
Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	53	MET	CE	47.90	26.97 – 7.37	15.7
1	A	49	MET	CE	39.30	26.97 – 7.37	11.3
1	A	11	LEU	CD2	6.80	32.60 – 15.60	-10.2
1	A	85	ILE	CG1	14.70	36.54 – 18.94	-7.4
1	A	91	LYS	HE2	1.55	3.87 – 1.97	-7.2
1	A	82	THR	HB	1.88	5.82 – 2.52	-6.9
1	A	11	LEU	CD1	13.90	32.77 – 16.57	-6.6
1	A	85	ILE	CD1	24.50	21.91 – 5.01	6.5
1	A	59	LEU	CD2	14.10	32.60 – 15.60	-5.9
1	A	8	SER	HB3	2.27	5.25 – 2.45	-5.6
1	A	91	LYS	HD3	2.87	2.75 – 0.45	5.5
1	A	36	GLU	CG	29.40	42.24 – 29.94	-5.4
1	A	85	ILE	CG2	10.10	24.63 – 10.43	-5.2

### 6.1.5 Random Coil Index (RCI) plots [i](#)

The images below report *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.

Random coil index (RCI) for chain A:



Random coil index (RCI) for chain B:

