



wwPDB NMR Structure Validation Summary Report ⓘ

Sep 17, 2017 – 06:57 PM EDT

PDB ID : 5UHU
Title : Solution conformation of cytochrome P450 MycG with mycinamicin IV bound
Authors : Pochapsky, T.C.; Tietz, D.R.
Deposited on : unknown

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

We welcome your comments at 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 : rb-20029824
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : rb-20029824

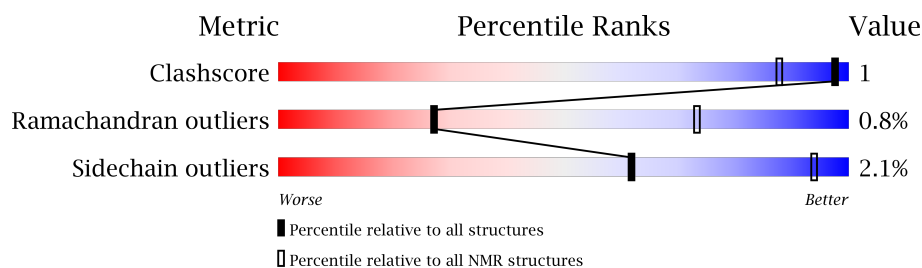
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 23%.

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 ≥ 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	397	<div>71%24%..</div>

2 Ensemble composition and analysis ⓘ

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

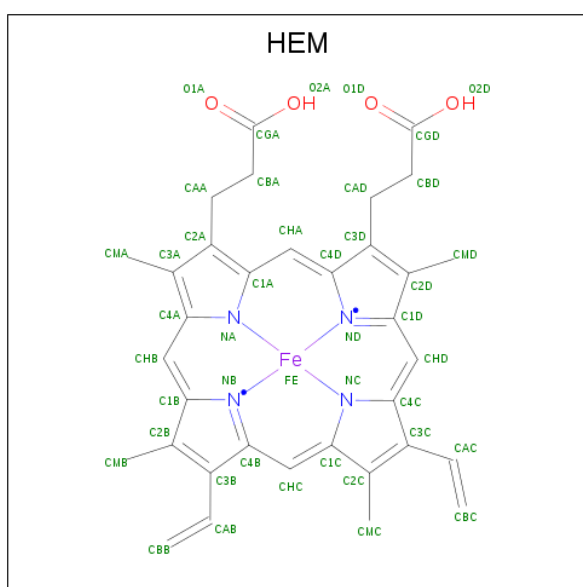
3 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 6366 atoms, of which 3179 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Mycinamicin IV hydroxylase/epoxidase.

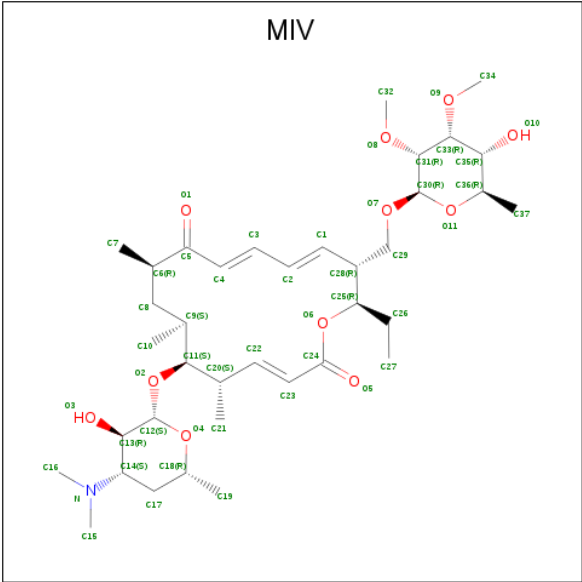
Mol	Chain	Residues	Atoms						Trace
1	A	393	Total	C	H	N	O	S	0
			6180	1939	3086	567	577	11	

- Molecule 2 is PROTOPORPHYRIN IX CONTAINING FE (three-letter code: HEM) (formula: $C_{34}H_{32}FeN_4O_4$).



Mol	Chain	Residues	Atoms					
2	A	1	Total	C	Fe	H	N	O
			73	34	1	30	4	4

- Molecule 3 is MYCINAMICIN IV (three-letter code: MIV) (formula: $C_{37}H_{61}NO_{11}$).



Mol	Chain	Residues	Atoms				
			Total	C	H	N	O
3	A	1	110	37	61	1	11

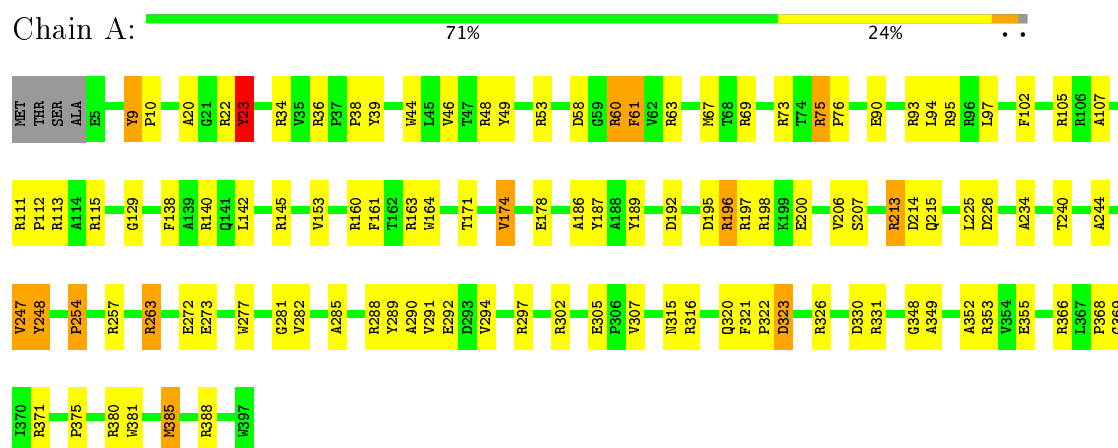
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		
			Total	H	O
4	A	1	3	2	1

4 Residue-property plots

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: Mycinamicin IV hydroxylase/epoxidase



5 Refinement protocol and experimental data overview

Of the 1200 calculated structures, 1 were deposited, based on the following criterion: *structures with the least restraint violations*.

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

Software name	Classification	Version
AMBER	refinement	14

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)	5uhu_cs.cif
Number of chemical shift lists	1
Total number of shifts	1135
Number of shifts mapped to atoms	1135
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	23%

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: HEM, MIV

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	1.61	22/3158 (0.7%)	2.24	124/4294 (2.9%)
All	All	1.61	22/3158 (0.7%)	2.24	124/4294 (2.9%)

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	14
All	All	0	14

5 of 22 bond outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	23	TYR	CE2-CZ	9.15	1.50	1.38
1	A	292	GLU	CD-OE1	7.29	1.33	1.25
1	A	22	ARG	CZ-NH2	-6.81	1.24	1.33
1	A	272	GLU	CD-OE1	6.09	1.32	1.25
1	A	381	TRP	CD2-CE3	6.05	1.49	1.40

5 of 124 angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	140	ARG	NE-CZ-NH1	20.39	130.50	120.30
1	A	326	ARG	NE-CZ-NH1	-17.76	111.42	120.30
1	A	326	ARG	NE-CZ-NH2	17.69	129.15	120.30
1	A	48	ARG	NE-CZ-NH1	16.03	128.32	120.30
1	A	198	ARG	NE-CZ-NH1	15.99	128.29	120.30

There are no chirality outliers.

5 of 13 planar outliers are listed below.

Mol	Chain	Res	Type	Group
1	A	388	ARG	Sidechain
1	A	60	ARG	Sidechain
1	A	369	GLY	Mainchain
1	A	23	TYR	Sidechain
1	A	248	TYR	Sidechain,Mainchain

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	3094	3086	3085	4
2	A	43	30	30	1
All	All	3187	3179	3176	4

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 clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
1:A:9:TYR:CG	1:A:10:PRO:HA	0.48	2.44
1:A:61:PHE:CE2	1:A:290:ALA:HB2	0.44	2.47
1:A:348:GLY:HA3	2:A:401:HEM:C2C	0.43	2.48
1:A:174:VAL:CG1	1:A:178:GLU:HG2	0.41	2.45

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	391/397 (98%)	357 (91%)	31 (8%)	3 (1%)	27	73
All	All	391/397 (98%)	357 (91%)	31 (8%)	3 (1%)	27	73

All 3 Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	A	215	GLN
1	A	254	PRO
1	A	281	GLY

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	326/329 (99%)	319 (98%)	7 (2%)	62	94
All	All	326/329 (99%)	319 (98%)	7 (2%)	62	94

5 of 7 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
1	A	94	LEU
1	A	23	TYR
1	A	67	MET
1	A	171	THR
1	A	93	ARG

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

2 ligands are 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	HEM	A	401	1,4	28,50,50	2.46	3 (10%)
3	MIV	A	402	-	50,51,51	1.67	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	HEM	A	401	1,4	17,82,82	1.69	0 (0%)
3	MIV	A	402	-	59,71,71	1.86	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	HEM	A	401	1,4	-	0,6,54,54	0,0,8,8
3	MIV	A	402	-	-	0,55,91,91	0,2,3,3

All bond outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	401	HEM	C3B-C2B	6.14	1.32	1.40
2	A	401	HEM	C1C-NC	5.79	1.43	1.36
2	A	401	HEM	C1B-NB	5.64	1.43	1.36

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

The completeness of assignment taking into account all chemical shift lists is 23% for the well-defined parts and 23% for the entire structure.

7.1 Chemical shift list 1

File name: 5uhu_cs.cif

Chemical shift list name: *data_complete.str*

7.1.1 Bookkeeping

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	1135
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	313	0.30 ± 0.10	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	249	1.26 ± 0.08	Should be applied
$^{13}\text{C}'$	0	—	None (insufficient data)
^{15}N	280	0.31 ± 0.25	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 23%, i.e. 1125 atoms were assigned a chemical shift out of a possible 4907. 0 out of 83 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	866/1909 (45%)	278/758 (37%)	310/786 (39%)	278/365 (76%)
Sidechain	255/2720 (9%)	6/1592 (0%)	246/975 (25%)	3/153 (2%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	4/278 (1%)	2/148 (1%)	0/117 (0%)	2/13 (15%)
Overall	1125/4907 (23%)	286/2498 (11%)	556/1878 (30%)	283/531 (53%)

7.1.4 Statistically unusual chemical shifts ⓘ

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	347	LEU	H	2.69	11.47 – 4.97	-8.5
1	A	158	HIS	CB	42.20	40.69 – 19.69	5.7

7.1.5 Random Coil Index (RCI) plots ⓘ

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.

Random coil index (RCI) for chain A:

