wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 05:45 AM BST

PDB ID	:	5HV8
Title	:	Solution structure of an octanoyl- loaded acyl carrier protein domain from module MLSA2 of the mycolactone polyketide synthase
Authors	:	Vance, S.; Tkachenko, O.; Thomas, B.; Bassuni, M.; Hong, H.; Nietlispach, D.; Broadhurst, R.W.
Deposited on	:	2016-01-28

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 (i) symbol.

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

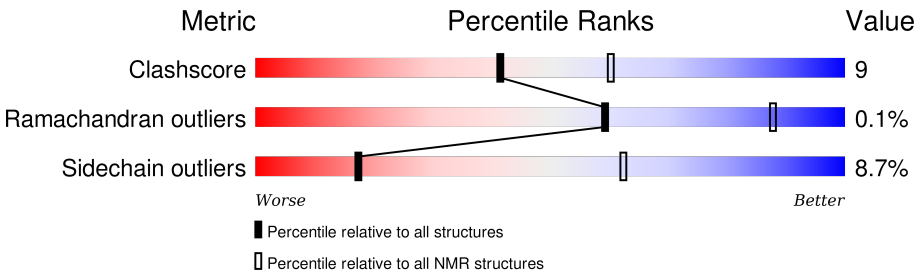
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 80%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	95	57% 20% • 22%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 6 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:2059-A:2075, A:2079-A:2135 (74)	0.32	6

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 3 clusters and 6 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1511 atoms, of which 760 are hydrogens and 0 are deuteriums.

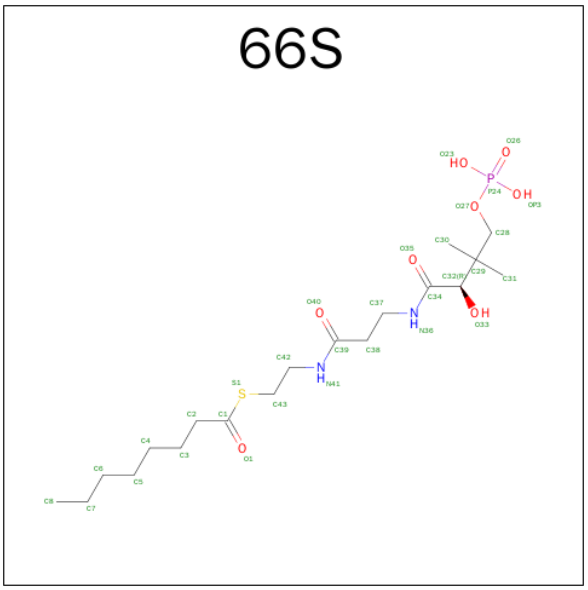
- Molecule 1 is a protein called Type I modular polyketide synthase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	95	1446	448	725	133	139	1	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2046	GLY	-	expression tag	UNP Q6MZA5
A	2047	SER	-	expression tag	UNP Q6MZA5
A	2048	HIS	-	expression tag	UNP Q6MZA5
A	2049	MET	-	expression tag	UNP Q6MZA5

- Molecule 2 is S-[2-({N-[(2R)-2-hydroxy-3,3-dimethyl-4-(phosphonooxy)butanoyl]-beta-alanyl}amino)ethyl] octanethioate (three-letter code: 66S) (formula: $C_{19}H_{37}N_2O_8PS$).



Mol	Chain	Residues	Atoms						
			Total	C	H	N	O	P	S
2	A	1	65	19	35	2	7	1	1

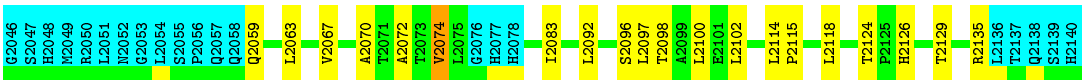
4 Residue-property plots (i)

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: Type I modular polyketide synthase

Chain A:

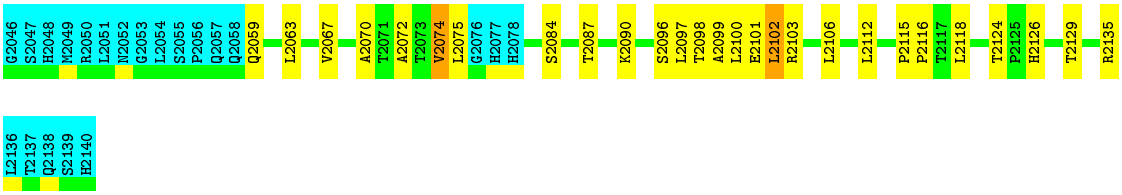


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 6. Colouring as in section 4.1 above.

- Molecule 1: Type I modular polyketide synthase

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 50 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations*.

The authors did not provide any information on software used for structure solution, optimization or refinement.

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)	5hv8_cs.str
Number of chemical shift lists	1
Total number of shifts	1082
Number of shifts mapped to atoms	1082
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	80%

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: 66S

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	0.34±0.02	0±0/571 (0.0±0.0%)	0.48±0.01	0±0/789 (0.0±0.0%)
All	All	0.34	1/11420 (0.0%)	0.48	0/15780 (0.0%)

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	2120	PHE	CE1-CZ	5.19	1.47	1.37	19	1

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	558	568	563	10±2
2	A	30	35	0	0±1
All	All	11760	12060	11260	206

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 9.

5 of 68 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:2097:LEU:O	1:A:2100:LEU:HG	0.69	1.87	12	18
1:A:2131:HIS:O	1:A:2135:ARG:HG2	0.64	1.93	12	8
1:A:2063:LEU:O	1:A:2067:VAL:HG23	0.64	1.93	18	1
1:A:2074:VAL:HG12	1:A:2098:THR:HG22	0.63	1.71	16	2
1:A:2059:GLN:O	1:A:2063:LEU:HG	0.63	1.94	1	2

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	74/95 (78%)	68±1 (91±2%)	6±1 (9±2%)	0±0 (0±0%)	59	88
All	All	1480/1900 (78%)	1351 (91%)	128 (9%)	1 (0%)	59	88

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	2079	THR	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	63/81 (78%)	58±1 (91±2%)	5±1 (9±2%)	17	62
All	All	1260/1620 (78%)	1151 (91%)	109 (9%)	17	62

5 of 15 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	2124	THR	18

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	2092	LEU	15
1	A	2074	VAL	15
1	A	2096	SER	15
1	A	2098	THR	8

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 carbohydrates in this entry.

6.6 Ligand geometry (i)

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	66S	A	2201	1	22,29,30	0.51±0.05	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	66S	A	2201	1	27,36,39	0.74±0.03	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	66S	A	2201	1	-	2±0 ,34,36,37	0±0,0,0,0

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique torsion outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	2201	66S	O1-C1-S1-C43	10
2	A	2201	66S	C2-C1-S1-C43	10
2	A	2201	66S	C38-C39-N41-C42	3
2	A	2201	66S	C32-C34-N36-C37	3

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 (i)

The completeness of assignment taking into account all chemical shift lists is 80% for the well-defined parts and 78% for the entire structure.

7.1 Chemical shift list 1

File name: 5hv8_cs.str

Chemical shift list name: *oct_160127_ourshifts_b.txt*

7.1.1 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	1082
Number of shifts mapped to atoms	1082
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	87	-0.88 ± 0.12	Should be applied
$^{13}C_{\beta}$	86	0.05 ± 0.17	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	82	-0.33 ± 0.16	None needed (< 0.5 ppm)

7.1.3 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 80%, i.e. 688 atoms were assigned a chemical shift out of a possible 863. 0 out of 15 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	278/358 (78%)	142/142 (100%)	68/148 (46%)	68/68 (100%)
Sidechain	375/447 (84%)	225/257 (88%)	144/177 (81%)	6/13 (46%)

Continued on next page...

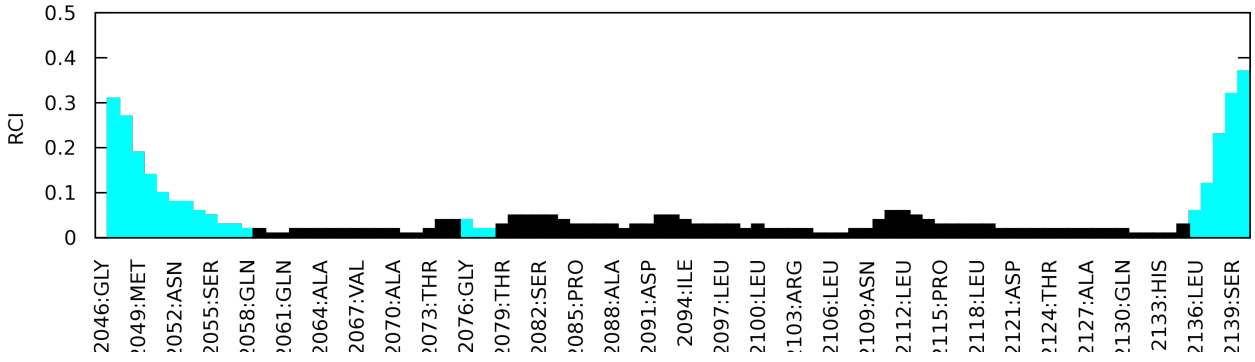
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	35/58 (60%)	19/30 (63%)	16/18 (89%)	0/10 (0%)
Overall	688/863 (80%)	386/429 (90%)	228/343 (66%)	74/91 (81%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

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: