



wwPDB NMR Structure Validation Summary Report ⓘ

Apr 26, 2016 – 10:51 PM BST

PDB ID : 2K5T
Title : Solution NMR Structure of Putative N-Acetyl Transferase YhhK from E. coli
Bound to Coenzyme A: Northeast Structural Genomics Consortium Target
ET106
Authors : Cort, J.R.; Yee, A.; Montelione, G.T.; Arrowsmith, C.H.; Kennedy, M.A.;
Northeast Structural Genomics Consortium (NESG)
Deposited on : 2008-06-30

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 : 1.7.1 (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

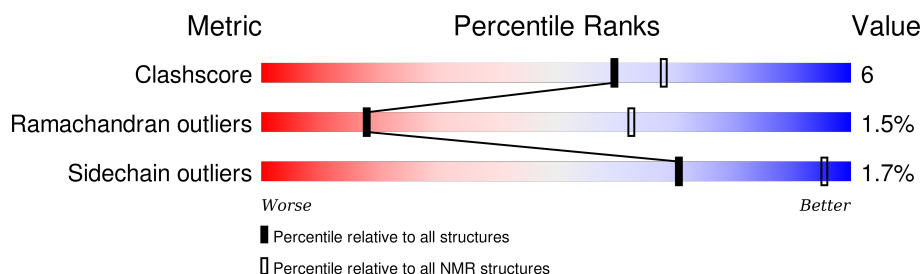
1 Overall quality at a glance

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	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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	128	 89% 5% 5%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA and RNA chains that are outliers for geometric criteria:

Mol	Chain	Compound	Res	Total models with violations	
				Chirality	Geometry
2	A	COA	999	13	-

2 Ensemble composition and analysis ⓘ

This entry contains 20 models. Model 4 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *no criterion*.

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:2-A:24, A:31-A:128 (121)	0.64	4

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 and 2 single-model clusters were found.

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

3 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 2121 atoms, of which 1049 are hydrogens and 0 are deuteriums.

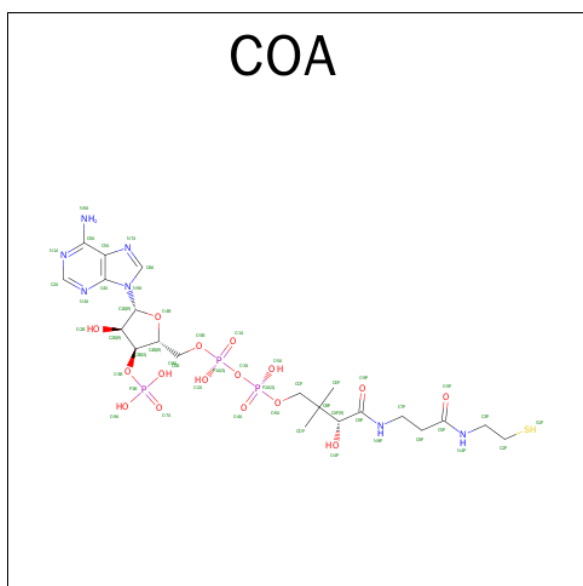
- Molecule 1 is a protein called Uncharacterized protein yhhK.

Mol	Chain	Residues	Atoms						Trace
1	A	128	Total	C	H	N	O	S	0
			2041	640	1017	188	190	6	

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	119	THR	ALA	ENGINEERED	UNP P37613
A	128	GLY	-	EXPRESSION TAG	UNP P37613

- Molecule 2 is COENZYME A (three-letter code: COA) (formula: $C_{21}H_{36}N_7O_{16}P_3S$).



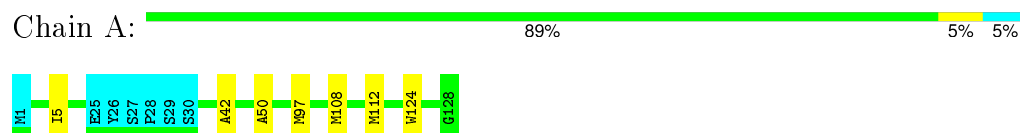
Mol	Chain	Residues	Atoms						
2	A	1	Total	C	H	N	O	P	S
			80	21	32	7	16	3	1

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.

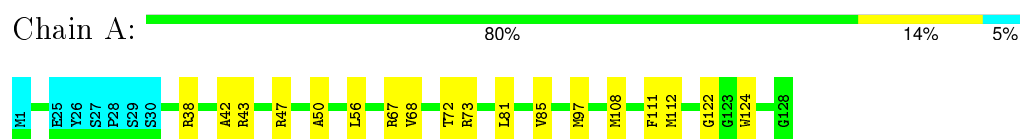
- Molecule 1: Uncharacterized protein yhhK



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

The representative model is number 4. Colouring as in section 4.1 above.

- Molecule 1: Uncharacterized protein yhhK



5 Refinement protocol and experimental data overview

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

Of the 30 calculated structures, 20 were deposited, based on the following criterion: *fewest restraints violations and energies, favorable backbone geometry, minimal clashes*.

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

Software name	Classification	Version
AutoStructure	structure solution	
AutoStructure	refinement	
CNS	structure solution	
CNS	refinement	
X-PLOR NIH	structure solution	
X-PLOR NIH	refinement	

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

6.1 Standard geometry

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

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	970	969	968	10±2
2	A	48	32	30	4±2
All	All	20360	20020	19980	233

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 6.

5 of 145 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:53:ARG:HB2	1:A:64:ASP:HB2	0.85	1.46	13	1
1:A:97:MET:HB2	1:A:124:TRP:HB2	0.81	1.50	18	11
1:A:73:ARG:HA	2:A:999:COA:O4A	0.77	1.80	3	7
1:A:95:TRP:HB2	1:A:126:LYS:HB3	0.76	1.58	6	1
1:A:67:ARG:HA	2:A:999:COA:H62	0.74	1.56	1	1

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	120/128 (94%)	109±3 (91±3%)	10±3 (8±2%)	2±1 (1±1%)	18	63
All	All	2400/2560 (94%)	2173 (91%)	192 (8%)	35 (1%)	18	63

5 of 14 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	116	GLY	6
1	A	90	PRO	5
1	A	65	SER	4
1	A	75	ARG	4
1	A	122	GLY	3

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	101/108 (94%)	99±1 (98±1%)	2±1 (2±1%)	71	95
All	All	2020/2160 (94%)	1985 (98%)	35 (2%)	71	95

5 of 22 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	6	ILE	5
1	A	72	THR	5
1	A	12	SER	2
1	A	15	ASP	2
1	A	118	THR	2

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	COA	A	999	-	41,50,50	1.37±0.07	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	COA	A	999	-	49,75,75	2.92±0.26	3±1 (6±1%)

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	COA	A	999	-	-	0±0,44,64,64	0±0,3,3,3

There are no bond-length outliers.

5 of 9 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
2	A	999	COA	C7P-C6P-C5P	14.15	87.96	112.22	7	10
2	A	999	COA	N3A-C2A-N1A	12.90	118.74	128.87	19	20
2	A	999	COA	C2P-C3P-N4P	11.79	89.08	112.44	16	6
2	A	999	COA	C6P-C7P-N8P	10.46	88.50	111.94	11	8
2	A	999	COA	O4B-C1B-N9A	9.22	125.53	108.11	19	6

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

Mol	Chain	Res	Type	Atoms	Models (Total)
2	A	999	COA	C1B	6
2	A	999	COA	CAP	5
2	A	999	COA	C3B	4
2	A	999	COA	C4B	3
2	A	999	COA	C2B	2

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	999	COA	CAP-C9P-N8P-C7P	7
2	A	999	COA	C6P-C5P-N4P-C3P	3
2	A	999	COA	P3B-O3B-C3B-C4B	2

There are no ring outliers.

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