wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 02:25 PM EDT

PDB ID	:	2KC5
Title	:	Solution Structure of HybE from Escherichia coli
Authors	:	Jin, C.; Shao, X.; Lu, J.; Li, Y.
Deposited on	:	2008-12-17

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

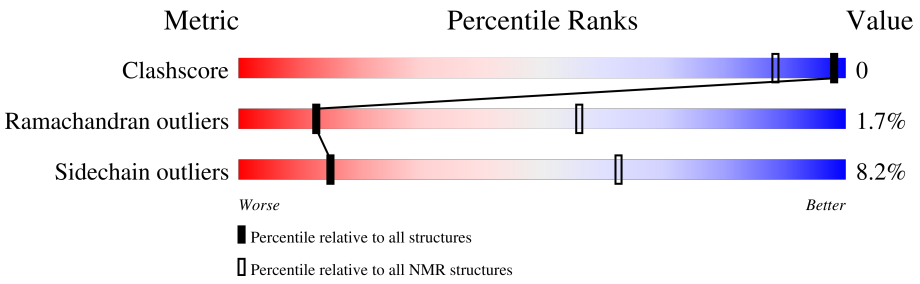
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 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	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 >=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	162	65% 5% 30%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 20 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:4-A:26, A:39-A:112, A:119-A:135 (114)	0.77	20

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 3 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2499 atoms, of which 1240 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Hydrogenase-2 operon protein hybE.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	162	2499	799	1240	213	236	11	0

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, DNA and oligosaccharide 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: Hydrogenase-2 operon protein hybE

Chain A: 65% 5% 30%

M1 T2 E3 S27 M28 H29 D30 L31 S32 F33 L34 H35 P36 S37 M38 T46 T53 M61 Q86 T93 V102 Y105 P113 L114 S115 H116 S117 M118 S119 I120 S136 L137 P138 V139 T140 N141 P142 D143 V144 P145 H146 A147 G148 R149 R150 A151 L152 L153 F154 G155

R156 R157 S158 G159 E160 N161 A162

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

The representative model is number 20. Colouring as in section 4.1 above.

• Molecule 1: Hydrogenase-2 operon protein hybE

Chain A: 66% • 30%

M1 T2 E3 S27 M28 H29 D30 L31 S32 F33 L34 H35 P36 S37 M38 T53 W74 Q86 T93 V102 Y105 P113 L114 S115 H116 S117 M118 S119 I120 S136 L137 P138 V139 T140 N141 P142 D143 V144 P145 H146 A147 G148 R149 R150 A151 L152 L153 F154 G155 R156 R157

S158 G159 E160 N161 A162

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the lowest energy*.

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

Software name	Classification	Version
CYANA	structure solution	2.0
Amber	refinement	9

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	0.65±0.00	0±0/912 (0.0± 0.0%)	0.99±0.02	0±1/1241 (0.0± 0.1%)
All	All	0.65	0/18240 (0.0%)	0.99	8/24820 (0.0%)

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(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	132	ARG	NE-CZ-NH2	-7.08	116.76	120.30	9	2
1	A	105	TYR	CB-CG-CD2	-6.22	117.27	121.00	9	2
1	A	125	ARG	NE-CZ-NH2	-5.59	117.50	120.30	18	2
1	A	26	ARG	NE-CZ-NH1	5.42	123.01	120.30	16	1
1	A	45	PHE	C-N-CA	5.08	134.41	121.70	19	1

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	890	880	880	0±1
All	All	17800	17600	17600	10

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 0.

5 of 6 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:45:PHE:CE1	1:A:131:ALA:HB1	0.69	2.23	18	2
1:A:120:ILE:HD12	1:A:120:ILE:H	0.50	1.66	19	1
1:A:45:PHE:CE1	1:A:131:ALA:CB	0.48	2.96	13	1
1:A:45:PHE:CE2	1:A:131:ALA:CB	0.48	2.96	4	4
1:A:40:VAL:HG13	1:A:58:THR:HG22	0.44	1.88	13	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	114/162 (70%)	108±2 (95±2%)	4±2 (4±2%)	2±1 (2±1%)	13	56
All	All	2280/3240 (70%)	2157 (95%)	84 (4%)	39 (2%)	13	56

5 of 7 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	46	THR	17
1	A	102	VAL	14
1	A	4	GLU	4
1	A	50	GLY	1
1	A	112	SER	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	98/140 (70%)	90±2 (92±2%)	8±2 (8±2%)	15	62

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	1960/2800 (70%)	1799 (92%)	161 (8%)	15 62

5 of 28 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	53	THR	20
1	A	93	THR	20
1	A	86	GLN	17
1	A	61	MET	16
1	A	105	TYR	16

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 monosaccharides in this entry.

6.6 Ligand geometry (i)

There are no ligands in this entry.

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)

No chemical shift data were provided