Full wwPDB NMR Structure Validation Report (i)

Mar 6, 2022 – 08:50 PM EST

PDB ID	:	2OQ3
Title	:	Solution Structure of the mannitol- specific cryptic phosphotransferase enzyme IIA CmtB from Escherichia coli
Authors	:	Jin, C.; Yu, C.
Deposited on	:	2007-01-31

This is a Full wwPDB NMR Structure Validation 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 following versions of software and data (see references (i)) were used in the production of this report:

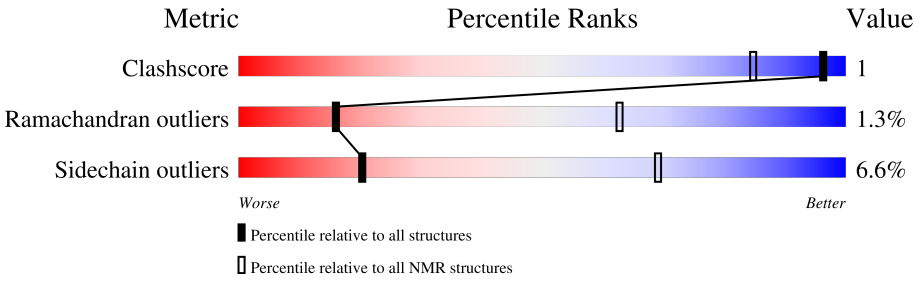
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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	150	85% 6% 7% •

2 Ensemble composition and analysis (i)

This entry contains 21 models. Model 21 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *minimized average structure*.

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:1-A:50, A:57-A:90, A:96-A:147 (136)	0.22	21

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 10 single-model clusters were found.

Cluster number	Models
1	1, 2, 6, 7, 14, 17, 21
2	9, 15
3	5, 11
Single-model clusters	3; 4; 8; 10; 12; 13; 16; 18; 19; 20

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2240 atoms, of which 1114 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Mannitol-specific cryptic phosphotransferase enzyme IIA component.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	147	2240	713	1114	180	227	6	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-2	GLY	-	expression tag	UNP P69824
A	-1	SER	-	expression tag	UNP P69824
A	0	HIS	-	expression tag	UNP P69824

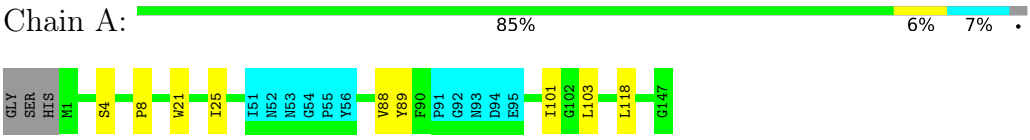
4 Residue-property plots (i)

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: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A:



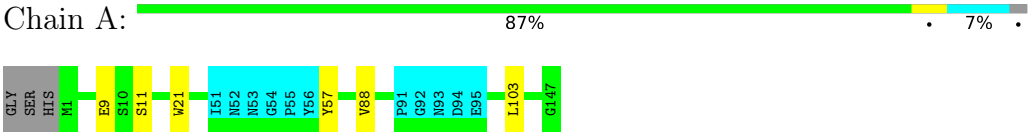
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

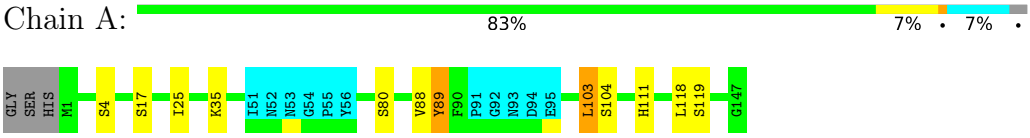
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

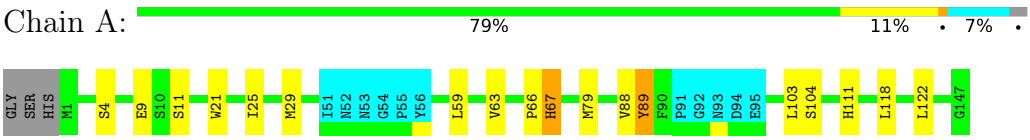
Chain A:



4.2.3 Score per residue for model 3

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

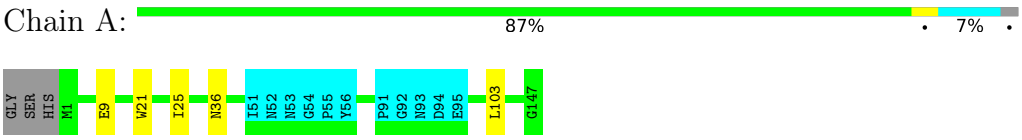
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

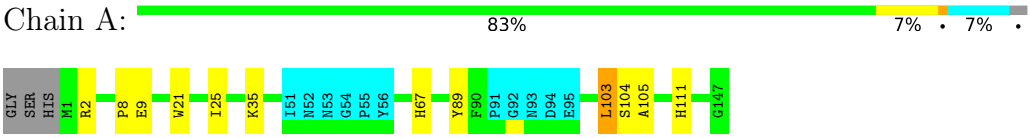
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

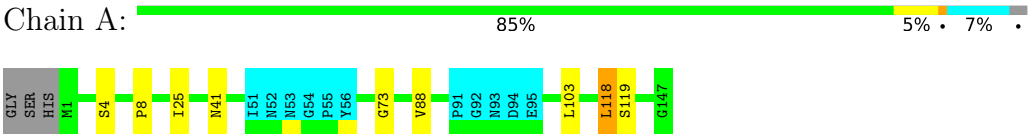
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

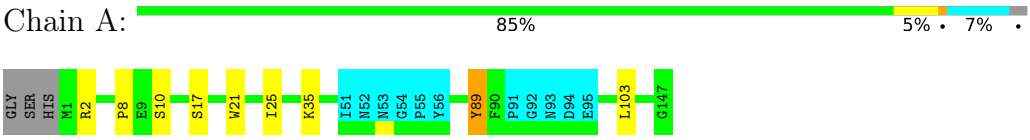
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

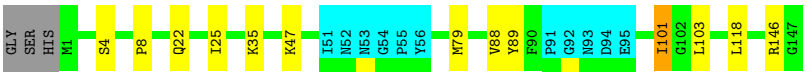
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

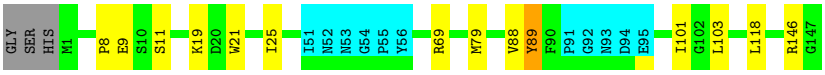
Chain A: 82% 8% • 7% •



4.2.9 Score per residue for model 9

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A: 81% 9% • 7% •



4.2.10 Score per residue for model 10

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A: 83% 7% • 7% •

4.2.11 Score per residue for model 11

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A: 79% 11% • 7% •

4.2.12 Score per residue for model 12

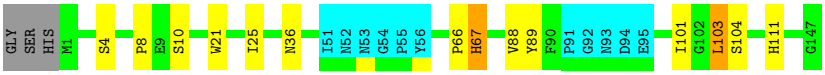
• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A: 85% 5% • 7% •

4.2.13 Score per residue for model 13

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

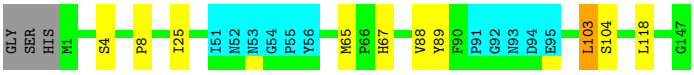
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

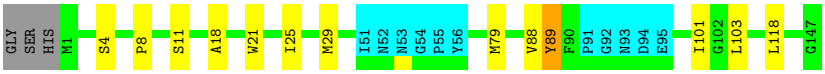
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

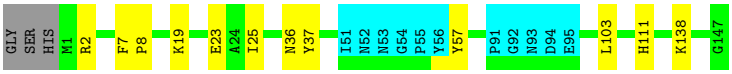
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

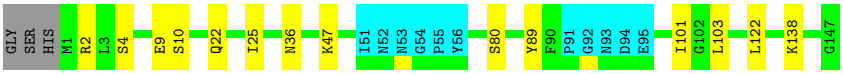
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A: 80% 10% • 7% •

GLY SER HIS M1 S4 P8 E9 S10 S17 I25 K35 I51 N52 N53 G54 P55 Y56 H67 M79 P91 G92 N93 D94 E95 I101 G102 L103 S104 A105 H111 L118 G147

4.2.19 Score per residue for model 19

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A: 79% 10% • 7% •

GLY SER HIS M1 R2 L3 S4 P8 E9 S10 S11 I25 I51 N52 N53 G54 P55 Y56 P66 H67 M79 V88 Y89 F90 P91 G92 N93 D94 E95 I101 G102 L103 L118 R146 G147

4.2.20 Score per residue for model 20

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A: 81% 10% 7% •

GLY SER HIS M1 R2 F7 P8 E9 S10 Q22 I25 N36 K47 I51 N52 N53 G54 P55 Y56 M79 V88 P91 G92 N93 D94 E95 I101 G102 L103 H111 L118 A134 G147

4.2.21 Score per residue for model 21 (medoid)

• Molecule 1: Mannitol-specific cryptic phosphotransferase enzyme IIA component

Chain A: 82% 8% • 7% •

GLY SER HIS M1 S4 P8 E9 S10 S11 K19 D20 W21 I25 N36 I51 N52 N53 G54 P55 Y56 V88 P91 G92 N93 D94 E95 L103 S104 A105 H111 G147

5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 21 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	1.0.6
Amber	refinement	7.0

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	0.59±0.00	0±0/1060 (0.0± 0.0%)	0.92±0.03	1±1/1434 (0.0± 0.1%)
All	All	0.59	0/22260 (0.0%)	0.92	14/30114 (0.0%)

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	0.0±0.0	0.2±0.5
All	All	0	5

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	2	ARG	NE-CZ-NH2	-7.11	116.75	120.30	7	1
1	A	146	ARG	NE-CZ-NH1	6.81	123.70	120.30	19	1
1	A	89	TYR	CB-CG-CD2	-6.55	117.07	121.00	3	8
1	A	29	MET	CA-CB-CG	6.01	123.52	113.30	15	1
1	A	69	ARG	NE-CZ-NH2	-5.80	117.40	120.30	9	2
1	A	57	TYR	CB-CG-CD2	-5.25	117.85	121.00	1	1

There are no chirality outliers.

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

Mol	Chain	Res	Type	Group	Models (Total)
1	A	146	ARG	Sidechain	2
1	A	6	TYR	Sidechain	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
1	A	7	PHE	Sidechain	1
1	A	37	TYR	Sidechain	1

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	1043	1046	1048	2±1
All	All	21903	21966	22008	36

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:79:MET:SD	1:A:118:LEU:HD21	0.63	2.34	20	7
1:A:101:ILE:HG21	1:A:118:LEU:HD22	0.56	1.77	20	6
1:A:105:ALA:HB3	1:A:111:HIS:CD2	0.55	2.37	18	4
1:A:101:ILE:CG2	1:A:118:LEU:HD22	0.49	2.38	20	1
1:A:103:LEU:HD23	1:A:104:SER:N	0.47	2.25	5	8
1:A:101:ILE:HB	1:A:122:LEU:HD11	0.46	1.87	17	1
1:A:118:LEU:HD12	1:A:119:SER:N	0.45	2.27	2	2
1:A:18:ALA:HB3	1:A:88:VAL:HG21	0.44	1.88	15	1
1:A:65:MET:CE	1:A:118:LEU:HD11	0.44	2.42	14	2
1:A:22:GLN:O	1:A:47:LYS:HE2	0.41	2.15	17	3
1:A:59:LEU:HD12	1:A:63:VAL:HG12	0.40	1.94	3	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	134/150 (89%)	126±2 (94±1%)	7±2 (5±1%)	2±1 (1±1%)	16	63
All	All	2814/3150 (89%)	2641 (94%)	137 (5%)	36 (1%)	16	63

All 8 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	8	PRO	16
1	A	9	GLU	10
1	A	66	PRO	3
1	A	67	HIS	3
1	A	73	GLY	1
1	A	57	TYR	1
1	A	10	SER	1
1	A	134	ALA	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	115/126 (91%)	107±2 (93±2%)	8±2 (7±2%)	20	69
All	All	2415/2646 (91%)	2256 (93%)	159 (7%)	20	69

All 26 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	25	ILE	20
1	A	103	LEU	19
1	A	88	VAL	14
1	A	21	TRP	11
1	A	4	SER	11
1	A	89	TYR	11
1	A	11	SER	9
1	A	36	ASN	7
1	A	35	LYS	6
1	A	67	HIS	6

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	10	SER	6
1	A	111	HIS	5
1	A	2	ARG	5
1	A	101	ILE	5
1	A	17	SER	4
1	A	19	LYS	4
1	A	7	PHE	3
1	A	1	MET	3
1	A	80	SER	2
1	A	138	LYS	2
1	A	29	MET	1
1	A	122	LEU	1
1	A	41	ASN	1
1	A	118	LEU	1
1	A	99	LEU	1
1	A	23	GLU	1

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