wwPDB NMR Structure Validation Summary Report

Jun 22, 2021 – 04:02 PM EDT

PDB ID	:	1O2F
Title	:	COMPLEX OF ENZYME IIAGLC AND IIBGLC PHOSPHOCARRIER PROTEIN HPR FROM ESCHERICHIA COLI NMR, RESTRAINED REGULARIZED MEAN STRUCTURE
Authors	:	Clore, G.M.; Cai, M.; Williams, D.C.
Deposited on	:	2003-03-11

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 ⓘ symbol.

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

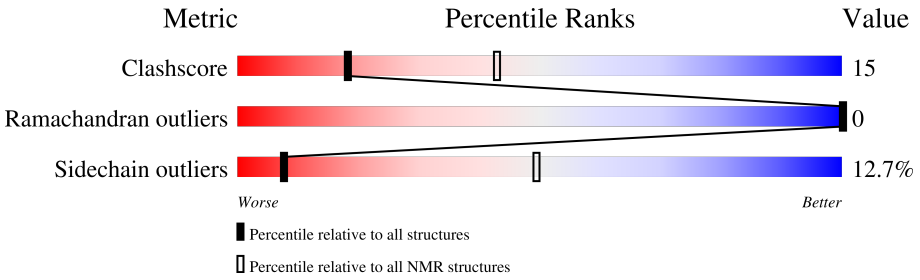
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.20
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.20

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	168	53% 31% 5% 11%
2	B	90	64% 20% • 14%

2 Ensemble composition and analysis

This entry contains 3 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 2 is the overall representative, medoid model (most similar to other models).

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:19-A:168, B:314-B:390 (227)	0.03	2

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 1 clusters. No single-model clusters were found.

Cluster number	Models
1	1, 2, 3

3 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 3439 atoms, of which 1742 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PTS system, glucose-specific IIA component.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	150	2305	729	1169	181	224	2	0

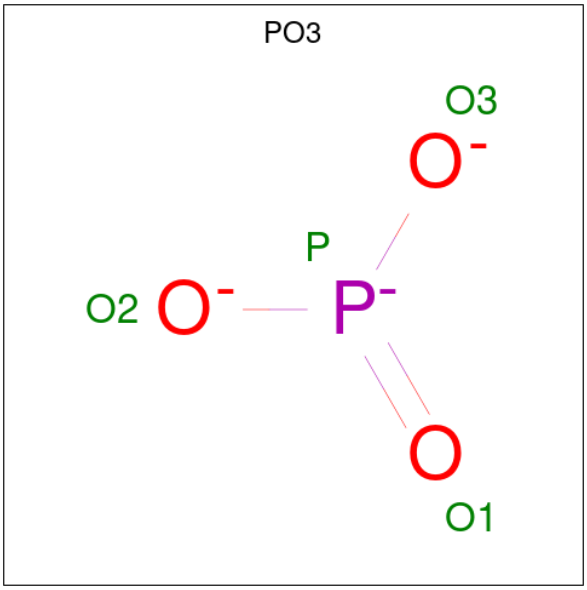
- Molecule 2 is a protein called PTS system, glucose-specific IIBC component.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	77	1130	345	573	98	111	3	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	317	ALA	PRO	SEE REMARK 999	UNP P69786

- Molecule 3 is PHOSPHITE ION (three-letter code: PO3) (formula: O_3P).



Mol	Chain	Residues	Atoms		
			Total	O	P
3	B	1	4	3	1

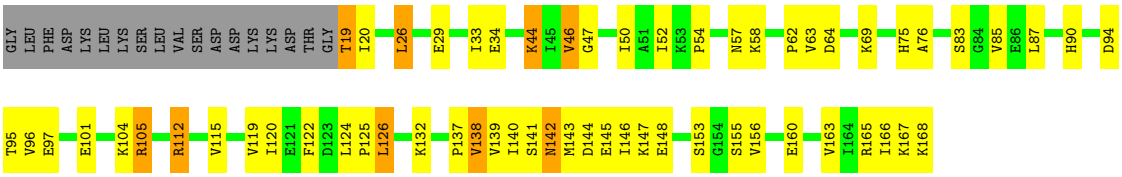
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, 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.

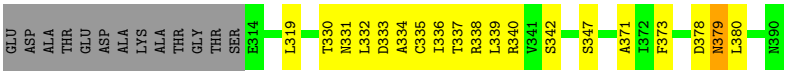
• Molecule 1: PTS system, glucose-specific IIA component

Chain A:



• Molecule 2: PTS system, glucose-specific IIBC component

Chain B:

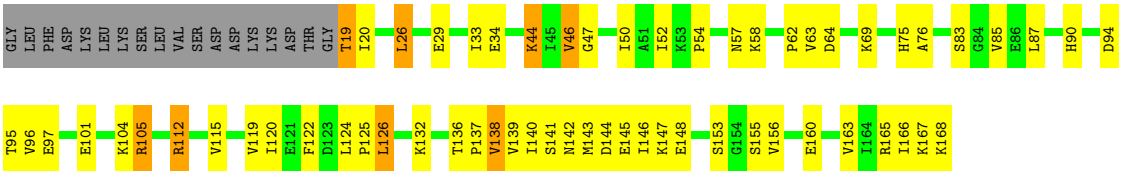


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 2. Colouring as in section 4.1 above.

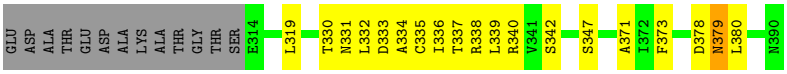
• Molecule 1: PTS system, glucose-specific IIA component

Chain A:



• Molecule 2: PTS system, glucose-specific IIBC component

Chain B:


GLU
ASP
ALA
THR
GLU
ASP
ALA
LYS
ALA
THR
GLY
THR
SER
E314
L319
T330
N331
L332
D333
A334
C335
I336
T337
R338
L339
R340
V341
S342
S347
A371
I372
F373
D378
N379
L380
N390

5 Refinement protocol and experimental data overview

The models were refined using the following method: *CONJOINED RIGID BODY/TORSION ANGLE DYNAMICS*.

Of the 60 calculated structures, 3 were deposited, based on the following criterion: *REGULARIZED MEAN STRUCTURES*.

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

Software name	Classification	Version
X-PLOR NIH	refinement	(HTTP://NMR.CIT.NIH.GOV/XPLOR_NIH)

No chemical shift data was provided.

6 Model quality

6.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: PO3

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	1.10±0.01	0±0/1151 (0.0± 0.0%)	0.97±0.04	0±0/1557 (0.0± 0.0%)
2	B	1.06±0.01	0±0/559 (0.0± 0.0%)	0.89±0.01	0±0/750 (0.0± 0.0%)
All	All	1.09	0/5130 (0.0%)	0.95	1/6921 (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	3.0±0.0
All	All	0	9

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	142	ASN	N-CA-CB	-5.19	101.25	110.60	3	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	105	ARG	Sidechain	3
1	A	112	ARG	Sidechain	3
1	A	165	ARG	Sidechain	3

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	1136	1169	1178	40±5
2	B	557	573	573	19±1
3	B	0	0	0	3±2
All	All	5087	5226	5252	152

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 15.

5 of 76 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:90:HIS:NE2	3:B:200:PO3:P	1.26	2.09	1	2
1:A:126:LEU:HD12	1:A:126:LEU:O	1.00	1.55	1	3
2:B:335:CYS:SG	3:B:200:PO3:P	0.91	2.68	2	1
1:A:64:ASP:OD1	1:A:115:VAL:HG23	0.79	1.78	3	3
1:A:138:VAL:CG2	1:A:138:VAL:O	0.77	2.33	1	2

6.3 Torsion angles

6.3.1 Protein backbone

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	148/168 (88%)	140±0 (95±0%)	8±0 (5±0%)	0±0 (0±0%)	100	100
2	B	75/90 (83%)	73±1 (97±1%)	2±1 (3±1%)	0±0 (0±0%)	100	100
All	All	669/774 (86%)	640 (96%)	29 (4%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	129/145 (89%)	107±2 (83±1%)	22±2 (17±1%)	5 40
2	B	57/66 (86%)	55±0 (96±0%)	2±0 (4±0%)	39 86
All	All	558/633 (88%)	487 (87%)	71 (13%)	8 49

5 of 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	19	THR	3
1	A	26	LEU	3
1	A	29	GLU	3
1	A	34	GLU	3
1	A	44	LYS	3

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 monosaccharides 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
3	PO3	B	200	-	0,3,3	0.00±0.00	-

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
3	PO3	B	200	-	0,3,3	0.00±0.00	-

There are no bond-length outliers.

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 (i)

No chemical shift data were provided