WORLDWIDE PDB PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report (i)

Apr 21, 2024 – 05:46 AM EDT

PDB ID	:	2LRK
BMRB ID	:	18379
Title	:	Solution Structures of the IIA(Chitobiose)-HPr complex of the N,N'-Diacetyl chitobiose
Authors	:	Cai, M.; Jung, Y.; Clore, M.
Deposited on	:	2012-04-06

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

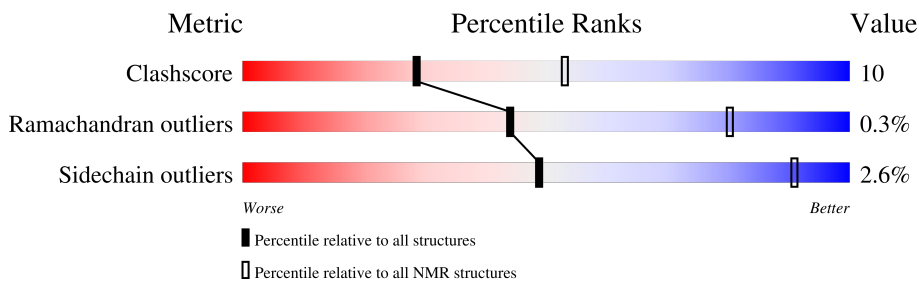|		
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson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

1 Overall quality at a glance ⓘ

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

The overall completeness of chemical shifts assignment is 36%.

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	Percentile Ranks	Value
Clashscore		10
Ramachandran outliers		0.3%
Sidechain outliers		2.6%

Worse   *Better*

▮ Percentile relative to all structures

▯ Percentile relative to all NMR structures

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	D	85	71% 27% •
2	A	103	73% 14% 14%
2	B	103	73% 14% 14%
2	C	103	70% 15% 16%

WORLDWIDE
wwPDB
PROTEIN DATA BANK

2   Ensemble composition and analysis ⓘ

This entry contains 21 models. Model 3 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	D:301-D:385, A:4-A:60, A:70-A:101, B:4-B:60, B:70-B:101, C:4-C:59, C:71-C:101 (350)	0.44	3

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 5 clusters and 1 single-model cluster was found.

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

3   Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 6100 atoms, of which 3114 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Phosphocarrier protein HPr.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	D	85	1294	401	654	107	130	2	0

- Molecule 2 is a protein called N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	A	103	1602	490	820	134	151	7	0
2	B	103	1602	490	820	134	151	7	0
2	C	103	1602	490	820	134	151	7	0

There are 6 discrepancies between the modelled and reference sequences:

| Chain	Residue	Modelled	Actual	Comment	Reference
A	76	GLU	HIS	engineered mutation	UNP P69791
A	79	LEU	ASP	engineered mutation	UNP P69791
B	76	GLU	HIS	engineered mutation	UNP P69791
B	79	LEU	ASP	engineered mutation	UNP P69791
C	76	GLU	HIS	engineered mutation	UNP P69791
C	79	LEU	ASP	engineered mutation	UNP P69791

4 Residue-property plots ⓘ

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.
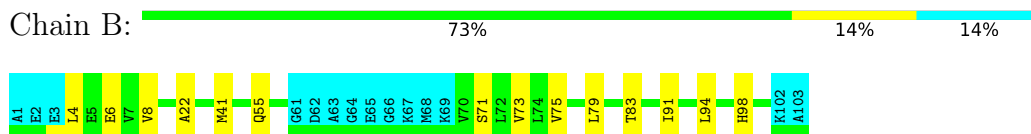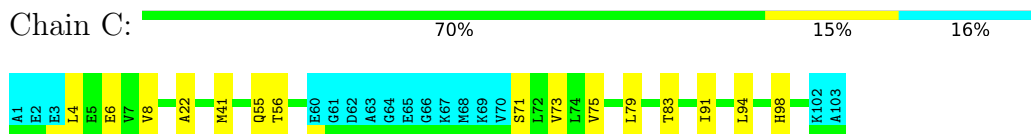
- Molecule 1: Phosphocarrier protein HPr
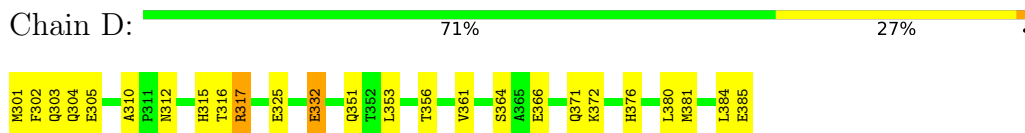
Chain D:

- Molecule 2: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component

Chain A:

- Molecule 2: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component

Chain B:

- Molecule 2: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component
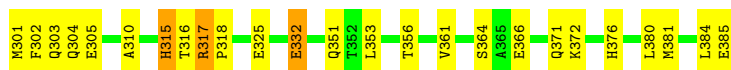
Chain C:
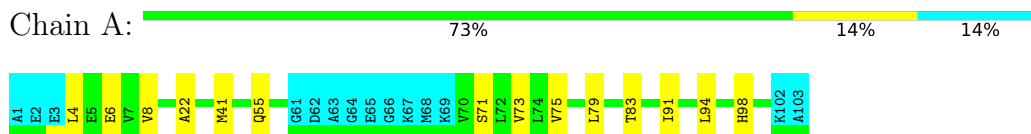
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

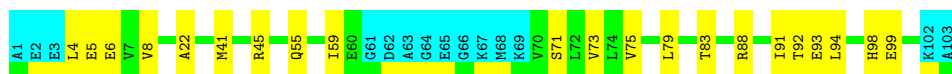The representative model is number 3. Colouring as in section 4.1 above.

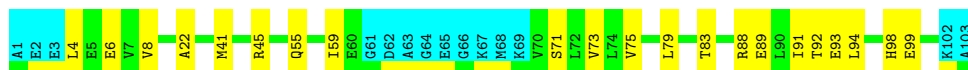- Molecule 1: Phosphocarrier protein HPr

Chain D:

71%   26%   •

M301 F302 Q303 Q304 E305 A310 H315 T316 R317 P318 E325 E332 Q351 T352 L353 T356 V361 S364 A365 E366 Q371 K372 H376 L380 M381 L384 E385

• Molecule 2: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component

Chain A:

66%   20%   14%

A1 E2 E3 L4 E5 E6 V7 V8 A22 M41 R45 Q55 I59 E60 G61 D62 A63 G64 E65 G66 K67 M68 K69 V70 S71 L72 V73 L74 V75 L79 T83 R88 I91 T92 E93 L94 H98 E99 K102 A103

• Molecule 2: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component

Chain B:

66%   20%   14%

A1 E2 E3 L4 E5 E6 V7 V8 A22 M41 R45 Q55 I59 E60 G61 D62 A63 G64 E65 G66 K67 M68 K69 V70 S71 L72 V73 L74 V75 L79 T83 R88 E89 L90 I91 T92 E93 L94 H98 E99 K102 A103

• Molecule 2: N,N'-diacetylchitobiose-specific phosphotransferase enzyme IIA component

Chain C:

64%   20%   16%

A1 E2 E3 L4 E5 E6 V7 V8 A22 Y23 M41 R45 Q55 T56 I59 E60 G61 D62 A63 G64 E65 G66 K67 M68 K69 V70 S71 L72 V73 L74 V75 L79 T83 R88 I91 T92 E93 L94 H98 E99 K102 A103

5   Refinement protocol and experimental data overview ⓘ

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

Of the 100 calculated structures, 21 were deposited, based on the following criterion: *structures with the least restraint violations.*

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

Software name	Classification	Version
X-PLOR NIH	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)	working_cs.cif
Number of chemical shift lists	2
Total number of shifts	1868
Number of shifts mapped to atoms	1868
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	36%

6   Model quality ⓘ

6.1   Standard geometry ⓘ

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

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	D	0.0±0.0	1.0±0.0
All	All	0	21

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

All unique planar outliers are listed below.

Mol	Chain	Res	Type	Group	Models (Total)
1	D	317	ARG	Sidechain	21

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	D	640	654	650	18±3
2	A	684	724	724	16±4
2	B	684	724	724	16±5
2	C	668	709	709	15±4
All	All	56196	59031	58947	1203

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 10.

5 of 377 unique clashes are listed below, sorted by their clash magnitude.

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
1:D:356:THR:HG22	2:C:56:THR:HG21	0.81	1.53	13	11
1:D:310:ALA:CB	1:D:385:GLU:OXT	0.79	2.31	13	21
1:D:310:ALA:HB1	1:D:385:GLU:OXT	0.79	1.78	14	21
1:D:372:LYS:HZ2	1:D:376:HIS:CE1	0.78	1.95	14	21
1:D:372:LYS:NZ	1:D:376:HIS:CE1	0.76	2.54	2	21

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	D	83/85 (98%)	80±1 (97±1%)	2±1 (3±1%)	0±0 (0±1%)	50	82
2	A	89/103 (86%)	88±0 (99±0%)	1±0 (1±0%)	0±0 (0±1%)	38	78
2	B	89/103 (86%)	88±0 (99±0%)	1±0 (1±0%)	0±0 (0±1%)	38	78
2	C	87/103 (84%)	86±0 (99±0%)	1±0 (1±0%)	0±0 (0±0%)	50	82
All	All	7308/8274 (88%)	7181 (98%)	106 (1%)	21 (0%)	44	80

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)
2	A	70	VAL	4
2	B	70	VAL	4
1	D	315	HIS	3
2	A	4	LEU	3
2	B	4	LEU	3

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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	D	70/70 (100%)	63±1 (91±1%)	7±1 (9±1%)	12	59
2	A	73/81 (90%)	73±0 (100±1%)	0±0 (0±1%)	89	97
2	B	73/81 (90%)	73±0 (100±1%)	0±0 (0±1%)	89	97
2	C	71/81 (88%)	71±0 (100±1%)	0±0 (0±1%)	89	97
All	All	6027/6573 (92%)	5869 (97%)	158 (3%)	49	91

5 of 27 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	D	305	GLU	21
1	D	332	GLU	21
1	D	364	SER	21
1	D	366	GLU	21
1	D	371	GLN	21

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 ⓘ

There are no ligands in this entry.

6.7 Other polymers ⓘ

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 ⓘ

The completeness of assignment taking into account all chemical shift lists is 36% for the well-defined parts and 35% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

7.1.1 Bookkeeping ⓘ

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	820
Number of shifts mapped to atoms	820
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	1

7.1.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	84	-0.29 ± 0.15	None needed (< 0.5 ppm)
$^{13}C_\beta$	78	0.13 ± 0.12	None needed (< 0.5 ppm)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	81	-0.87 ± 0.32	Should be applied

7.1.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 17%, i.e. 820 atoms were assigned a chemical shift out of a possible 4823. 0 out of 80 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	325/1761 (18%)	160/713 (22%)	84/700 (12%)	81/348 (23%)
Sidechain	469/2887 (16%)	292/1906 (15%)	177/897 (20%)	0/84 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	26/175 (15%)	10/91 (11%)	16/72 (22%)	0/12 (0%)
Overall	820/4823 (17%)	462/2710 (17%)	277/1669 (17%)	81/444 (18%)

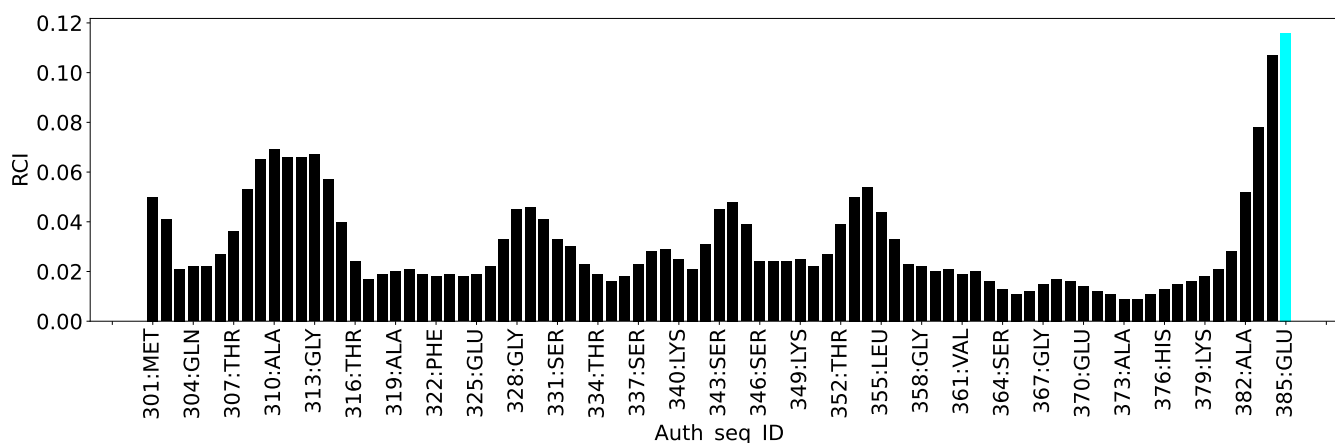### 7.1.4 Statistically unusual chemical shifts ⓘ

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	D	373	ALA	HA	1.91	2.13 − 6.34	-5.5

7.1.5 Random Coil Index (RCI) plots ⓘ

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain D:

7.2 Chemical shift list 2

File name: working_cs.cif

Chemical shift list name: *assigned_chem_shift_list_2*

7.2.1 Bookkeeping ⓘ

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	1048
Number of shifts mapped to atoms	1048
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

7.2.2 Chemical shift referencing ⓘ

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_\alpha$	101	-0.21 ± 0.05	None needed (< 0.5 ppm)
$^{13}C_\beta$	95	0.84 ± 0.12	Should be checked
$^{13}C'$	100	-0.51 ± 0.16	Should be applied
^{15}N	100	0.10 ± 0.20	None needed (< 0.5 ppm)

7.2.3 Completeness of resonance assignments ⓘ

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 19%, i.e. 931 atoms were assigned a chemical shift out of a possible 4823. 0 out of 80 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	430/1761 (24%)	166/713 (23%)	176/700 (25%)	88/348 (25%)
Sidechain	488/2887 (17%)	307/1906 (16%)	181/897 (20%)	0/84 (0%)
Aromatic	13/175 (7%)	5/91 (5%)	8/72 (11%)	0/12 (0%)
Overall	931/4823 (19%)	478/2710 (18%)	365/1669 (22%)	88/444 (20%)

7.2.4 Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

7.2.5 Random Coil Index (RCI) plots ⓘ

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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A: