W O R L D W I D E
wwPDB
P R O T E I N   D A T A   B A N K

Full wwPDB NMR Structure Validation Report (i)

Nov 11, 2020 – 02:10 PM JST

PDB ID : 6KH8
Title : Solution structure of Zn free Bovine Pancreatic Insulin in 20% acetic acid-d4 (pH 1.9)
Authors : Bhunia, A.; Ratha, B.N.; Kar, R.K.; Brender, J.R.
Deposited on : 2019-07-14

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

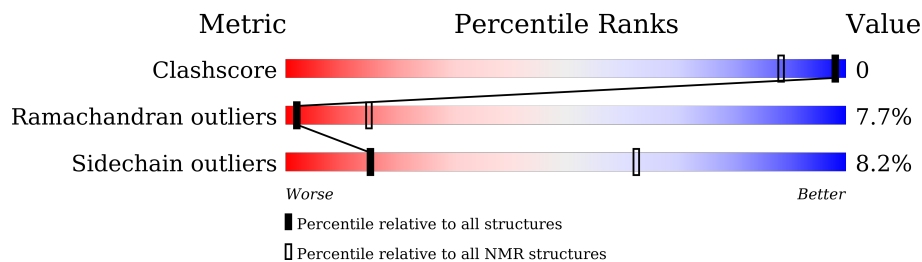|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.14.6
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.14.6

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 43%.

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		0
Ramachandran outliers		7.7%
Sidechain outliers		8.2%

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	A	21	86% 14%
2	B	30	80% 17% •

2 Ensemble composition and analysis ⓘ

This entry contains 10 models. Model 10 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:1-A:21, B:1-B:29 (50)	0.53	10

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

NmrClust was unable to cluster the ensemble.

Error message: Inconsistent models

3 Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 775 atoms, of which 375 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Insulin A Chain.

Mol	Chain	Residues	Atoms						Trace
1	A	21	Total	C	H	N	O	S	0
			305	97	145	25	34	4	

- Molecule 2 is a protein called Insulin B chain.

| Mol	Chain	Residues	Atoms						Trace
2	B	30	Total	C	H	N	O	S	0
			470	157	230	40	41	2	

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: Insulin A Chain
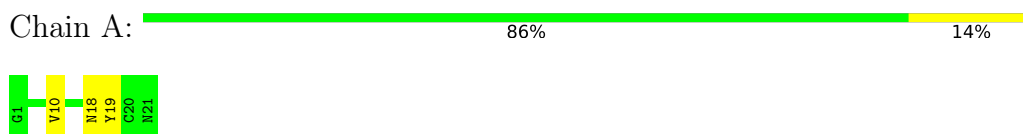
Chain A:

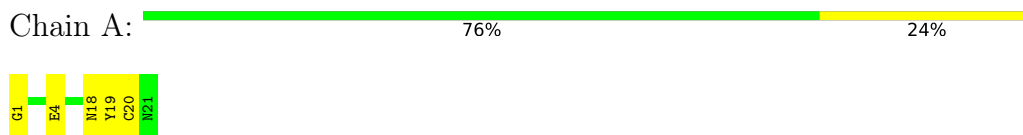• Molecule 2: Insulin B chain
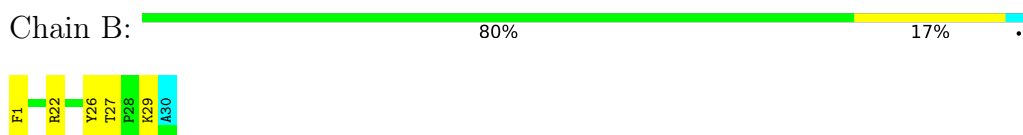
Chain B:

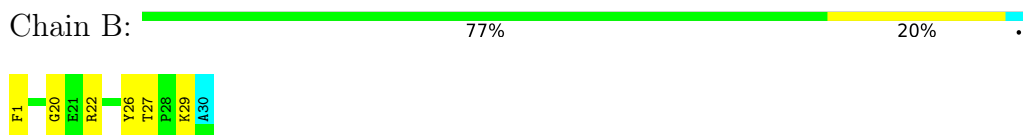## 4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1
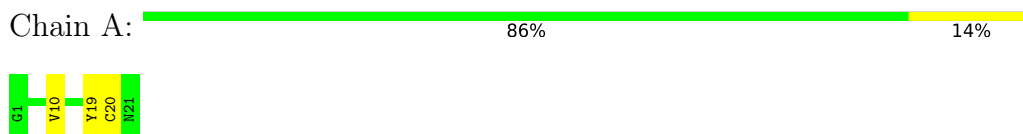
• Molecule 1: Insulin A Chain
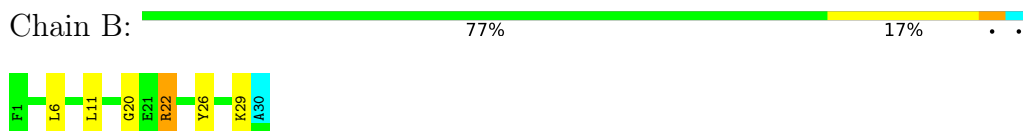
Chain A:

• Molecule 2: Insulin B chain

Chain B:

4.2.2 Score per residue for model 2

• Molecule 1: Insulin A Chain

Chain A:

86%   14%

G1   V10   Y19 C20 N21

• Molecule 2: Insulin B chain

Chain B:

77%   17%

F1   L6   L11   G20 E21 R22   Y26   K29 A30

4.2.3 Score per residue for model 3

• Molecule 1: Insulin A Chain

Chain A:

67%   33%

G1   E4   A8 S9 V10 C11   Y14   N18 Y19 C20 N21

• Molecule 2: Insulin B chain

Chain B:

83%   10%

F1   Y26 T27 P28 K29 A30
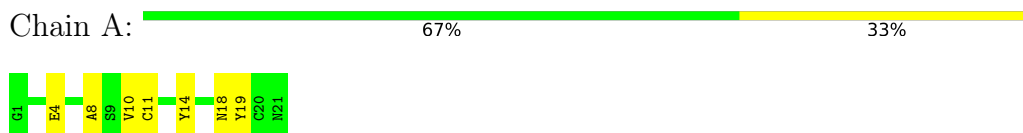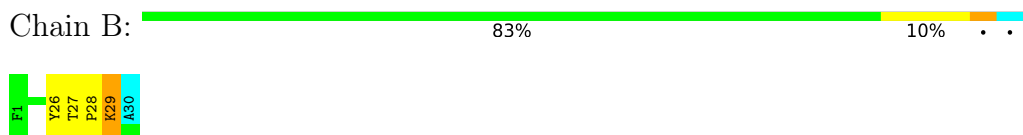
4.2.4 Score per residue for model 4

• Molecule 1: Insulin A Chain

Chain A:

81%   19%

G1   E4   S9   N18   N21

• Molecule 2: Insulin B chain

Chain B:

83%   7%

F1   Y26 T27 P28 K29 A30
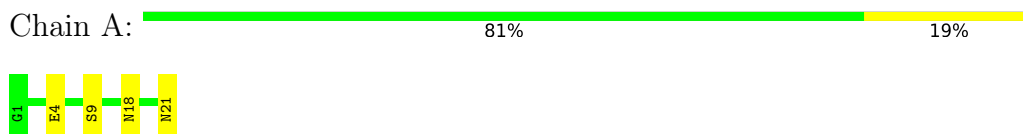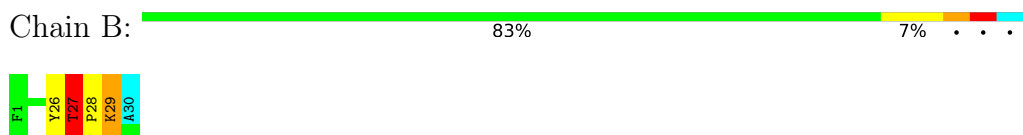
4.2.5 Score per residue for model 5

• Molecule 1: Insulin A Chain

Chain A:

• Molecule 2: Insulin B chain

Chain B:

4.2.6 Score per residue for model 6

• Molecule 1: Insulin A Chain

Chain A:

• Molecule 2: Insulin B chain

Chain B:

4.2.7 Score per residue for model 7

• Molecule 1: Insulin A Chain

Chain A:

• Molecule 2: Insulin B chain

Chain B:

4.2.8 Score per residue for model 8

• Molecule 1: Insulin A Chain

Chain A:

• Molecule 2: Insulin B chain

Chain B:

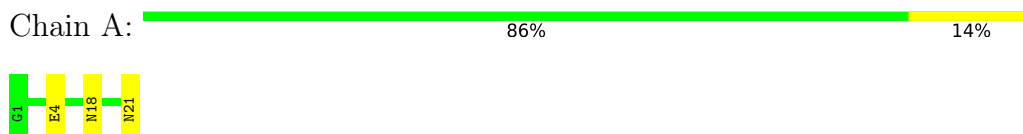### 4.2.9 Score per residue for model 9

• Molecule 1: Insulin A Chain
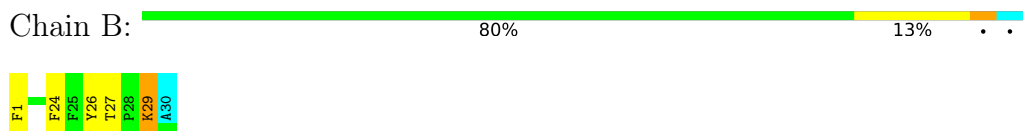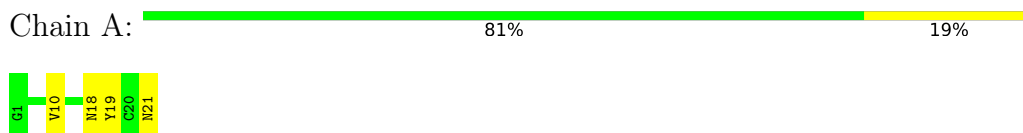
Chain A:

• Molecule 2: Insulin B chain
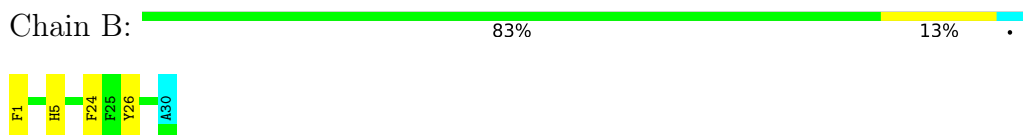
Chain B:

4.2.10 Score per residue for model 10 (medoid)

• Molecule 1: Insulin A Chain
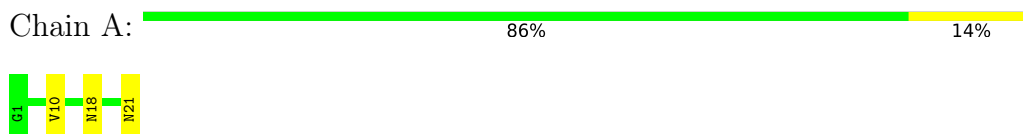
Chain A:

• Molecule 2: Insulin B chain

Chain B:
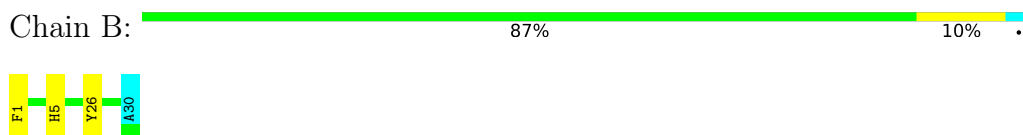
5 Refinement protocol and experimental data overview ⓘ

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

Of the 100 calculated structures, 10 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
Amber	structure calculation	
Amber	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	1
Total number of shifts	314
Number of shifts mapped to atoms	314
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	43%

No validations of the models with respect to experimental NMR restraints is performed at this time.

6   Model quality ⓘ

6.1   Standard geometry ⓘ

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	0.74±0.02	0±0/161 (0.0± 0.0%)	1.05±0.07	0±0/216 (0.0± 0.0%)
2	B	0.81±0.01	0±0/242 (0.0± 0.0%)	1.09±0.06	0±0/328 (0.1± 0.1%)
All	All	0.78	0/4030 (0.0%)	1.07	4/5440 (0.1%)

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	0.3±0.6
2	B	0.0±0.0	1.6±0.9
All	All	0	19

There are no bond-length outliers.

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models Worst	Models Total
2	B	22	ARG	NE-CZ-NH1	6.97	123.78	120.30	2	4

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)
2	B	29	LYS	Peptide	6
2	B	1	PHE	Peptide	5
2	B	27	THR	Peptide	3
1	A	1	GLY	Peptide	2
1	A	19	TYR	Sidechain	1
2	B	21	GLU	Peptide	1

Continued on next page...

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

Continued from previous page...

Mol	Chain	Res	Type	Group	Models (Total)
2	B	26	TYR	Peptide	1

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	160	145	145	0±0
2	B	234	225	225	0±0
All	All	3940	3700	3700	2

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.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Models Total
2:B:27:THR:O	2:B:29:LYS:HE2	0.46	2.11	4	1
1:A:1:GLY:HA2	1:A:19:TYR:CE1	0.40	2.51	1	1

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	19/21 (90%)	15±2 (78±8%)	3±2 (16±8%)	1±1 (6±4%)	3	21
2	B	28/30 (93%)	23±2 (80±7%)	3±1 (11±5%)	3±1 (9±4%)	1	12
All	All	470/510 (92%)	373 (79%)	61 (13%)	36 (8%)	2	15

All 10 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	26	TYR	9
2	B	27	THR	7
1	A	10	VAL	5
2	B	28	PRO	4
2	B	29	LYS	3
1	A	2	ILE	3
2	B	20	GLY	2
1	A	9	SER	1
1	A	11	CYS	1
1	A	8	ALA	1

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	19/19 (100%)	17±1 (89±7%)	2±1 (11±7%)	9	53
2	B	25/25 (100%)	24±1 (94±4%)	2±1 (6±4%)	23	72
All	All	440/440 (100%)	404 (92%)	36 (8%)	15	62

All 14 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	18	ASN	6
2	B	22	ARG	4
1	A	4	GLU	4
1	A	21	ASN	4
1	A	19	TYR	3
2	B	29	LYS	2
1	A	20	CYS	2
2	B	24	PHE	2
2	B	2	VAL	2
2	B	11	LEU	2
2	B	5	HIS	2
1	A	14	TYR	1
2	B	6	LEU	1
1	A	11	CYS	1

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 ⓘ

There are no chain breaks in this entry.

7 Chemical shift validation ⓘ

The completeness of assignment taking into account all chemical shift lists is 43% for the well-defined parts and 43% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *starch_output*

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	314
Number of shifts mapped to atoms	314
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

7.1.2 Chemical shift referencing ⓘ

No chemical shift referencing corrections were calculated (not enough data).

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 43%, i.e. 258 atoms were assigned a chemical shift out of a possible 595. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	94/248 (38%)	94/99 (95%)	0/100 (0%)	0/49 (0%)
Sidechain	132/274 (48%)	132/161 (82%)	0/103 (0%)	0/10 (0%)
Aromatic	32/73 (44%)	32/39 (82%)	0/32 (0%)	0/2 (0%)
Overall	258/595 (43%)	258/299 (86%)	0/235 (0%)	0/61 (0%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 43%, i.e. 258 atoms were assigned a chemical shift out of a possible 602. 0 out of 11 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	94/253 (37%)	94/101 (93%)	0/102 (0%)	0/50 (0%)
Sidechain	132/276 (48%)	132/162 (81%)	0/104 (0%)	0/10 (0%)
Aromatic	32/73 (44%)	32/39 (82%)	0/32 (0%)	0/2 (0%)
Overall	258/602 (43%)	258/302 (85%)	0/238 (0%)	0/62 (0%)

7.1.4   Statistically unusual chemical shifts ⓘ

There are no statistically unusual chemical shifts.

7.1.5   Random Coil Index (RCI) plots ⓘ

The images below report *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.

Random coil index (RCI) for chain A:

Random coil index (RCI) for chain B: