wwPDB NMR Structure Validation Summary Report (i)

Oct 3, 2021 – 09:29 AM EDT

PDB ID	:	2KQQ
Title	:	NMR structure of human insulin mutant gly-b8-d-ala, his-b10-asp, pro-b28-lys, lys-b29-pro, 20 structures
Authors	:	Hua, Q.X.; Wan, Z.L.; Zhao, M.; Jia, W.; Huang, K.; Weiss, M.A.
Deposited on	:	2009-11-13

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

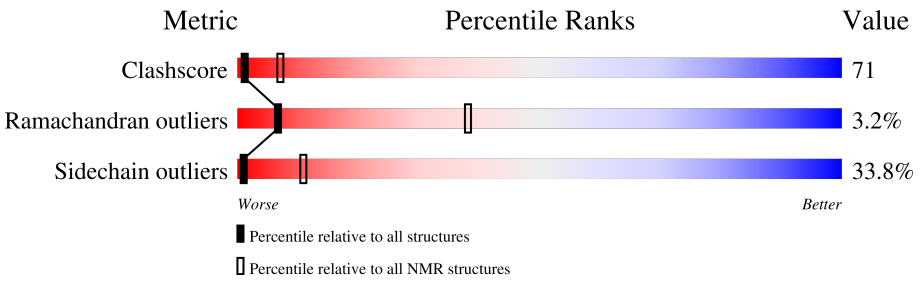
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.23.2

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 is 47%.

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	21	10% 57% 24% 10%
2	B	30	7% 60% 10% 23%

2 Ensemble composition and analysis (i)

This entry contains 20 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: *closest to the average*.

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:2-A:20, B:4-B:7, B:9-B:27 (42)	0.22	10

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 6 clusters and 4 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 784 atoms, of which 380 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Insulin B chain.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	21	312	99	149	25	35	4	0

- Molecule 2 is a protein called Insulin A chain.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	30	472	157	231	38	44	2	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	8	DAL	GLY	engineered mutation	UNP P01308
B	10	ASP	HIS	engineered mutation	UNP P01308
B	28	LYS	PRO	engineered mutation	UNP P01308
B	29	PRO	LYS	engineered mutation	UNP P01308

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 B chain

Chain A: 10% 57% 24% 10%

G1 I2 V3 E4 Q5 C6 C7 T8 S9 I10 C11 S12 L13 Y14 Q15 L16 E17 N18 Y19 C20 N21

• Molecule 2: Insulin A chain

Chain B: 7% 60% 10% 23%

F1 V2 N3 Q4 H5 L6 C7 A8 S9 D10 L11 V12 E13 A14 L15 Y16 L17 V18 C19 G20 E21 R22 G23 F24 F25 Y26 T27 K28 P29 T30

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

The representative model is number 10. Colouring as in section 4.1 above.

• Molecule 1: Insulin B chain

Chain A: 10% 57% 24% 10%

G1 I2 V3 E4 Q5 C6 C7 T8 S9 I10 C11 S12 L13 L16 E17 N18 Y19 C20 N21

• Molecule 2: Insulin A chain

Chain B: 10% 50% 17% 23%

F1 V2 N3 Q4 H5 L6 C7 A8 S9 D10 L11 V12 E13 A14 L15 Y16 L17 V18 C19 G20 E21 R22 G23 F24 F25 Y26 T27 K28 P29 T30

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DISTANCE GEOMETRY AND SIMULATED.*

Of the 40 calculated structures, 20 were deposited, based on the following criterion: *STRUCTURES WITH THE LOWEST ENERG.*

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

Software name	**Classification**	**Version**
VNMR	structure solution	3.5
VNMR	refinement	3.5

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	343
Number of shifts mapped to atoms	343
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	47%

6 Model quality (i)

6.1 Standard geometry (i)

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

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
2	B	0.0±0.0	0.9±0.2
All	All	0	19

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)
2	B	22	ARG	Sidechain	19

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	150	138	138	22±3
2	B	186	173	173	31±3
All	All	6720	6220	6220	917

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 71.

5 of 241 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:2:ILE:HD12	1:A:19:TYR:CE1	1.01	1.90	17	7
2:B:6:LEU:HD21	2:B:14:ALA:CB	0.99	1.87	3	13
1:A:4:GLU:O	1:A:8:THR:HG22	0.94	1.63	17	13
2:B:4:GLN:OE1	2:B:6:LEU:HD11	0.94	1.62	6	2
2:B:14:ALA:O	2:B:18:VAL:HG23	0.92	1.65	1	11

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	19/21 (90%)	15±1 (79±7%)	3±1 (14±8%)	1±1 (7±5%)	2 17
2	B	23/30 (77%)	18±1 (80±5%)	4±1 (19±5%)	0±0 (0±1%)	50 82
All	All	840/1020 (82%)	671 (80%)	142 (17%)	27 (3%)	7 38

5 of 6 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	20	CYS	7
1	A	9	SER	7
1	A	8	THR	5
1	A	4	GLU	4
1	A	5	GLN	3

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	19/20 (95%)	13±2 (71±10%)	6±2 (29±10%)	1 17
2	B	20/26 (77%)	12±2 (62±9%)	8±2 (38±9%)	0 6

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	780/920 (85%)	516 (66%)	264 (34%)	1 11

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	20	CYS	20
2	B	10	ASP	20
2	B	11	LEU	20
2	B	27	THR	19
2	B	17	LEU	15

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue is modelled in this entry.

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.

6.5 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)

The completeness of assignment taking into account all chemical shift lists is 47% for the well-defined parts and 46% 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 (i)

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	343
Number of shifts mapped to atoms	343
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 (i)

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

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 47%, i.e. 235 atoms were assigned a chemical shift out of a possible 502. 0 out of 9 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}**H**	^{13}**C**	^{15}**N**
Backbone	83/210 (40%)	83/84 (99%)	0/84 (0%)	0/42 (0%)
Sidechain	126/235 (54%)	126/137 (92%)	0/91 (0%)	0/7 (0%)
Aromatic	26/57 (46%)	26/30 (87%)	0/26 (0%)	0/1 (0%)
Overall	235/502 (47%)	235/251 (94%)	0/201 (0%)	0/50 (0%)

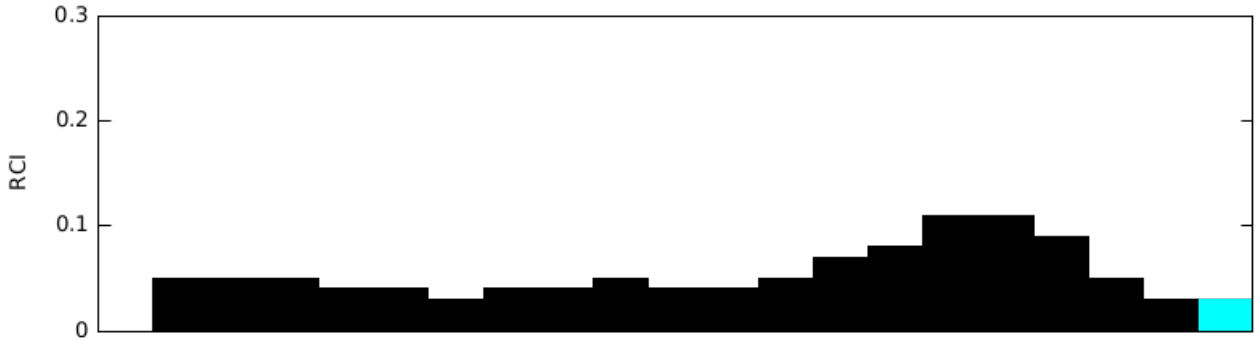
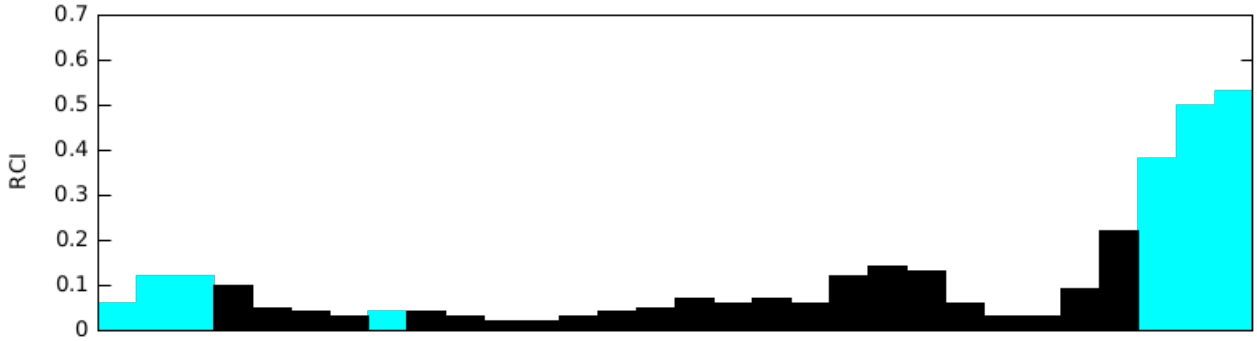
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

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: