wwPDB NMR Structure Validation Summary Report (i)

Oct 23, 2021 – 04:55 PM EDT

PDB ID	:	1SJU
Title	:	MINI-PROINSULIN, SINGLE CHAIN INSULIN ANALOG MUTANT: DES B30, HIS(B 10)ASP, PRO(B 28)ASP AND PEPTIDE BOND BETWEEN LYS B 29 AND GLY A 1, NMR, 20 STRUCTURES
Authors	:	Hua, Q.X.; Hu, S.Q.; Jia, W.H.; Chu, Y.C.; Burke, G.T.; Wang, S.H.; Katsoyannis, P.G.; Weiss, M.A.
Deposited on	:	1997-10-09

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:

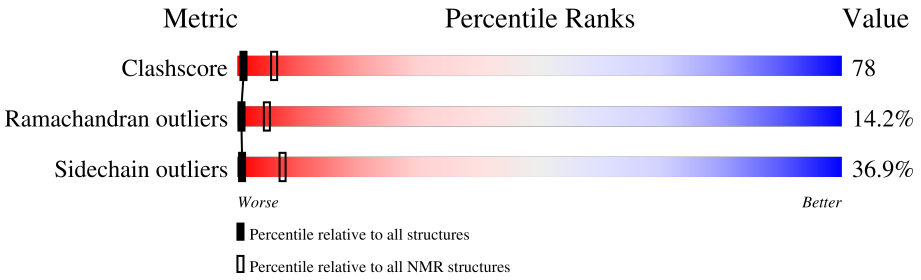
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.23.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al.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

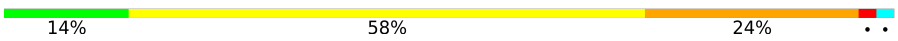
The overall completeness of chemical shifts assignment was not calculated.

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	50	

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:2-A:50 (49)	0.74	11

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 4 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 762 atoms, of which 366 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PROINSULIN.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	50	762	250	366	62	78	6	0

There are 38 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	10	ASP	HIS	engineered mutation	UNP P01308
A	28	ASP	PRO	engineered mutation	UNP P01308
A	?	-	THR	deletion	UNP P01308
A	?	-	ARG	deletion	UNP P01308
A	?	-	ARG	deletion	UNP P01308
A	?	-	GLU	deletion	UNP P01308
A	?	-	ALA	deletion	UNP P01308
A	?	-	GLU	deletion	UNP P01308
A	?	-	ASP	deletion	UNP P01308
A	?	-	LEU	deletion	UNP P01308
A	?	-	GLN	deletion	UNP P01308
A	?	-	VAL	deletion	UNP P01308
A	?	-	GLY	deletion	UNP P01308
A	?	-	GLN	deletion	UNP P01308
A	?	-	VAL	deletion	UNP P01308
A	?	-	GLU	deletion	UNP P01308
A	?	-	LEU	deletion	UNP P01308
A	?	-	GLY	deletion	UNP P01308
A	?	-	GLY	deletion	UNP P01308
A	?	-	GLY	deletion	UNP P01308
A	?	-	PRO	deletion	UNP P01308
A	?	-	GLY	deletion	UNP P01308
A	?	-	ALA	deletion	UNP P01308
A	?	-	GLY	deletion	UNP P01308
A	?	-	SER	deletion	UNP P01308
A	?	-	LEU	deletion	UNP P01308
A	?	-	GLN	deletion	UNP P01308
A	?	-	PRO	deletion	UNP P01308
A	?	-	LEU	deletion	UNP P01308
A	?	-	ALA	deletion	UNP P01308
A	?	-	LEU	deletion	UNP P01308

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	?	-	GLU	deletion	UNP P01308
A	?	-	GLY	deletion	UNP P01308
A	?	-	SER	deletion	UNP P01308
A	?	-	LEU	deletion	UNP P01308
A	?	-	GLN	deletion	UNP P01308
A	?	-	LYS	deletion	UNP P01308
A	?	-	ARG	deletion	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: PROINSULIN

Chain A: 14% 58% 24%

F1 V2 N3 Q4 H5 L6 C7 G8 S9 D10 L11 V12 E13 A14 L15 Y16 L17 V18 C19 G20 E21 R22 G23 F24 F25 Y26 T27 D28 K29 G30 I31 V32 E33 Q34 C35 S38 I39 C40 S41 L42 Y43 Q44 L45 E46 N47 Y48 C49 N50

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

The representative model is number 11. Colouring as in section 4.1 above.

• Molecule 1: PROINSULIN

Chain A: 26% 52% 14% 6%

F1 V2 L6 C7 G8 S9 D10 L11 V12 E13 A14 L15 Y16 L17 V18 C19 G20 E21 R22 G23 F24 F25 Y26 T27 D28 V32 E33 Q34 C35 C36 T37 S38 I39 C40 S41 L42 Y43 Q44 L45 E46 C49 N50

5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 20 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
X-PLOR	refinement	

No chemical shift data was provided.

6 Model quality (i)

6.1 Standard geometry (i)

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	A	0.0±0.0	1.0±0.0
All	All	0	20

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	A	22	ARG	Sidechain	20

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	385	355	355	58±9
All	All	7700	7100	7100	1151

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 78.

5 of 527 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:42:LEU:HD13	1:A:43:TYR:N	1.03	1.69	19	3
1:A:12:VAL:HG12	1:A:26:TYR:CE2	1.02	1.90	5	6

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:12:VAL:HG23	1:A:26:TYR:CE2	0.98	1.93	19	3
1:A:15:LEU:CD2	1:A:45:LEU:HD13	0.97	1.89	19	1
1:A:12:VAL:HG12	1:A:26:TYR:CZ	0.97	1.95	13	8

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	48/50 (96%)	28±3 (59±6%)	13±3 (27±6%)	7±2 (14±4%)	1 5
All	All	960/1000 (96%)	564 (59%)	260 (27%)	136 (14%)	1 5

5 of 23 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	7	CYS	18
1	A	42	LEU	15
1	A	8	GLY	13
1	A	9	SER	11
1	A	31	ILE	11

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	44/45 (98%)	28±3 (63±7%)	16±3 (37±7%)	1 7
All	All	880/900 (98%)	555 (63%)	325 (37%)	1 7

5 of 41 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	16	TYR	19
1	A	7	CYS	16
1	A	26	TYR	15
1	A	31	ILE	13
1	A	42	LEU	13

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