Full wwPDB NMR Structure Validation Report (i)

Jun 25, 2024 – 12:01 PM EDT

PDB ID	:	5XF0
BMRB ID	:	27064
Title	:	Solution structure of the IgI domain of CD147
Authors	:	Jin, S.J.; Xia, B.
Deposited on	:	2017-04-06

This is a Full wwPDB NMR Structure Validation 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

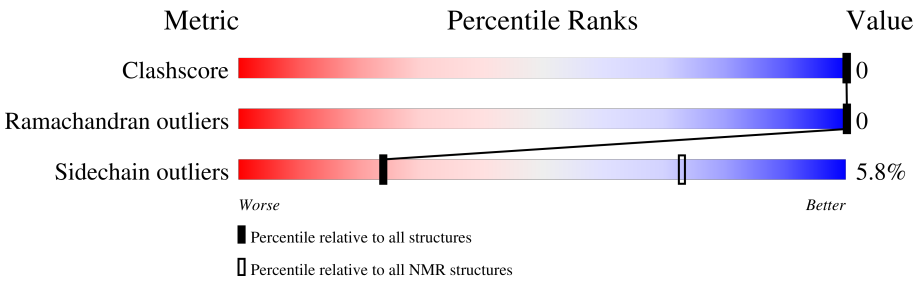
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 90%.

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	108	77% 5% • 18%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 16 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:105-A:128, A:133-A:149, A:158-A:205 (89)	0.26	16

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 1 single-model cluster was found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1656 atoms, of which 817 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Basigin.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	108	1656	514	817	151	168	6	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	98	MET	-	initiating methionine	UNP P35613

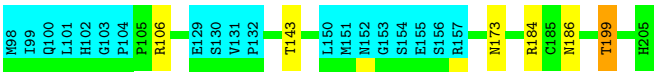
4 Residue-property plots (i)

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: Basigin

Chain A:



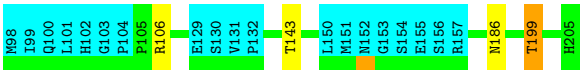
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: Basigin

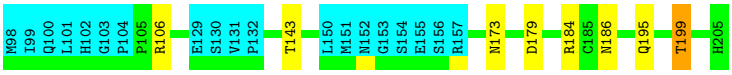
Chain A:



4.2.2 Score per residue for model 2

• Molecule 1: Basigin

Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: Basigin

Chain A: 75% 6% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 I116 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 N173 L174 R184 C185 N186 T199 H205

4.2.4 Score per residue for model 4

- Molecule 1: Basigin

Chain A: 75% 6% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 I116 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 R166 N173 L174 R184 T199 H205

4.2.5 Score per residue for model 5

- Molecule 1: Basigin

Chain A: 75% 6% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 I116 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 N173 L174 R184 C185 N186 T199 H205

4.2.6 Score per residue for model 6

- Molecule 1: Basigin

Chain A: 75% 6% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 R166 N173 R184 C185 N186 T199 R203 S204 H205

4.2.7 Score per residue for model 7

- Molecule 1: Basigin

Chain A: 77% 5% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 N173 R184 C185 N186 T199 H205

4.2.8 Score per residue for model 8

- Molecule 1: Basigin

Chain A: 74% 7% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 I116 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 R166 L174 R184 C185 N186 T199 R203 S204 H205

4.2.9 Score per residue for model 9

- Molecule 1: Basigin

Chain A: 75% 6% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 H115 I116 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 R166 R184 C185 N186 T199 H205

4.2.10 Score per residue for model 10

- Molecule 1: Basigin

Chain A: 76% 5% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 K127 S128 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 R166 R184 C185 N186 T199 H205

4.2.11 Score per residue for model 11

- Molecule 1: Basigin

Chain A: 79% • • 18%

M98 I99 Q100 L101 H102 G103 P104 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 R184 C185 N186 T199 H205

4.2.12 Score per residue for model 12

- Molecule 1: Basigin

Chain A: 76% 6% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 N173 D179 R184 C185 N186 T199 H205

4.2.13 Score per residue for model 13

- Molecule 1: Basigin

Chain A: 76% 6% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 I116 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 L174 R184 C185 N186 T199 H205

4.2.14 Score per residue for model 14

- Molecule 1: Basigin

Chain A: 75% 7% 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 I116 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 N173 R184 C185 N186 T199 R203 S204 H205

4.2.15 Score per residue for model 15

- Molecule 1: Basigin

Chain A: 76% 6% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 V160 N173 R184 C185 N186 T199 H205

4.2.16 Score per residue for model 16 (medoid)

- Molecule 1: Basigin

Chain A: 77% 5% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 N173 R184 C185 N186 T199 H205

4.2.17 Score per residue for model 17

- Molecule 1: Basigin

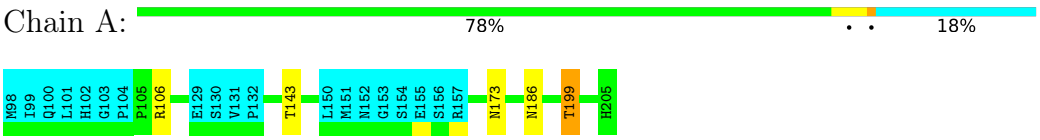
Chain A: 76% 6% • 18%

M98 I99 Q100 L101 H102 G103 P104 P105 R106 E129 S130 V131 P132 T143 L150 M151 N152 G153 S154 E155 S156 R157 R166 N173 R184 C185 N186 T199 H205

4.2.18 Score per residue for model 18

- Molecule 1: Basigin

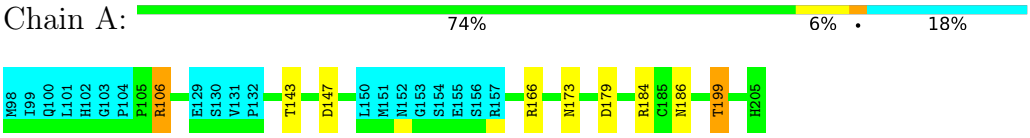
Chain A:



4.2.19 Score per residue for model 19

- Molecule 1: Basigin

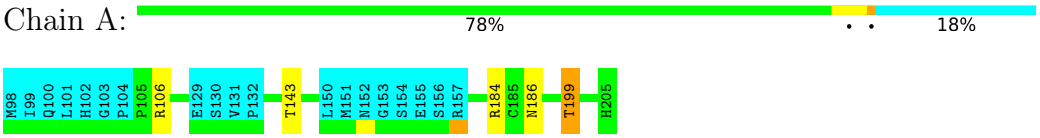
Chain A:



4.2.20 Score per residue for model 20

- Molecule 1: Basigin

Chain A:



5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 20 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	refinement	
DYANA	structure calculation	

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	1241
Number of shifts mapped to atoms	1241
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

6 Model quality (i)

6.1 Standard geometry (i)

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	0.68±0.00	0±0/709 (0.0± 0.0%)	1.13±0.02	3±1/953 (0.4± 0.1%)
All	All	0.68	0/14180 (0.0%)	1.13	69/19060 (0.4%)

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	0.1±0.3
All	All	0	2

There are no bond-length outliers.

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	184	ARG	NE-CZ-NH1	7.62	124.11	120.30	17	18
1	A	166	ARG	NE-CZ-NH1	7.22	123.91	120.30	10	5
1	A	106	ARG	NE-CZ-NH1	6.47	123.53	120.30	17	19
1	A	184	ARG	NE-CZ-NH2	-5.85	117.37	120.30	19	1
1	A	203	ARG	NE-CZ-NH1	5.85	123.23	120.30	14	3
1	A	199	THR	CA-CB-CG2	-5.76	104.33	112.40	5	19
1	A	179	ASP	CB-CG-OD2	5.53	123.28	118.30	19	2
1	A	106	ARG	NE-CZ-NH2	-5.06	117.77	120.30	4	1
1	A	179	ASP	CB-CG-OD1	5.05	122.84	118.30	12	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	115	HIS	Sidechain	1
1	A	166	ARG	Sidechain	1

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
All	All	13920	13460	13400	-

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 -.

There are no clashes.

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	88/108 (81%)	83±1 (94±1%)	5±1 (6±1%)	0±0 (0±0%)	100	100
All	All	1760/2160 (81%)	1663 (94%)	97 (6%)	0 (0%)	100	100

There are no Ramachandran outliers.

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	78/95 (82%)	73±1 (94±1%)	5±1 (6±1%)	24 73
All	All	1560/1900 (82%)	1469 (94%)	91 (6%)	24 73

All 12 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	143	THR	20
1	A	199	THR	20
1	A	186	ASN	19
1	A	173	ASN	13
1	A	116	ILE	7
1	A	174	LEU	5
1	A	166	ARG	2
1	A	195	GLN	1
1	A	127	LYS	1
1	A	160	VAL	1
1	A	106	ARG	1
1	A	147	ASP	1

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)

The completeness of assignment taking into account all chemical shift lists is 90% for the well-defined parts and 87% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *ppm_ THREE_ 23.str*

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	1241
Number of shifts mapped to atoms	1241
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	106	0.09 ± 0.10	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	99	-0.37 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	96	0.36 ± 0.14	None needed (< 0.5 ppm)
^{15}N	100	-1.62 ± 0.32	Should be applied

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 90%, i.e. 1056 atoms were assigned a chemical shift out of a possible 1172. 0 out of 10 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	440/444 (99%)	180/180 (100%)	174/178 (98%)	86/86 (100%)
Sidechain	574/642 (89%)	390/414 (94%)	176/200 (88%)	8/28 (29%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	42/86 (49%)	21/42 (50%)	19/36 (53%)	2/8 (25%)
Overall	1056/1172 (90%)	591/636 (93%)	369/414 (89%)	96/122 (79%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 87%, i.e. 1241 atoms were assigned a chemical shift out of a possible 1421. 0 out of 13 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	515/537 (96%)	213/218 (98%)	202/216 (94%)	100/103 (97%)
Sidechain	684/790 (87%)	464/511 (91%)	209/246 (85%)	11/33 (33%)
Aromatic	42/94 (45%)	21/46 (46%)	19/38 (50%)	2/10 (20%)
Overall	1241/1421 (87%)	698/775 (90%)	430/500 (86%)	113/146 (77%)

7.1.4 Statistically unusual chemical shifts (i)

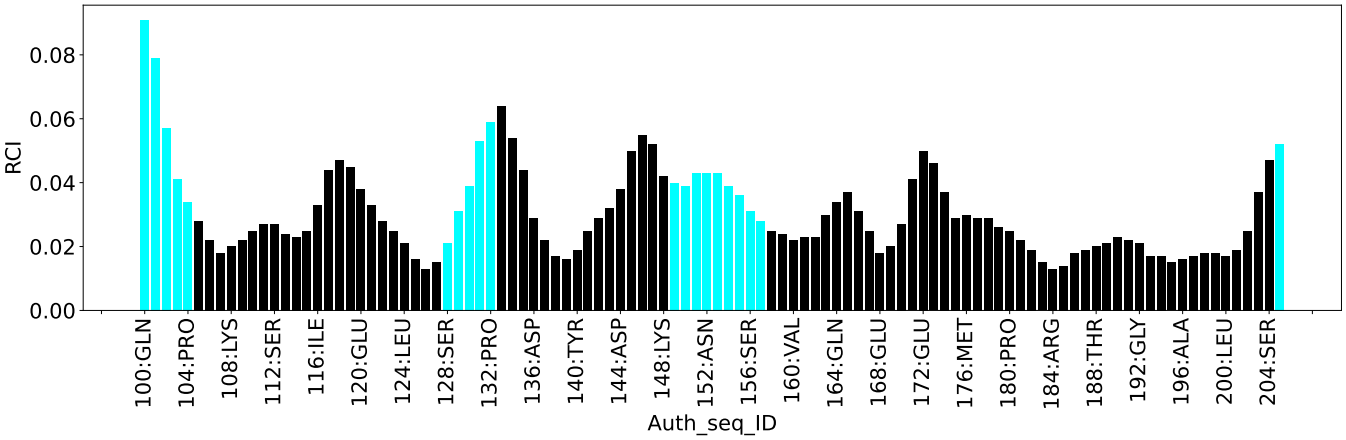
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	170	HIS	CE1	114.18	126.08 – 149.12	-10.2
1	A	169	LEU	HB3	-1.23	-0.26 – 3.31	-7.7
1	A	167	SER	HB3	2.00	2.49 – 5.20	-6.8
1	A	167	SER	HB2	2.21	2.61 – 5.13	-6.6
1	A	187	GLY	HA3	2.06	2.08 – 5.71	-5.0

7.1.5 Random Coil Index (RCI) plots (i)

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:


RCI
0.00
0.02
0.04
0.06
0.08
100:GLN
104:PRO
108:LYS
112:SER
116:ILE
120:GLU
124:LEU
128:SER
132:PRO
136:ASP
140:TYR
144:ASP
148:LYS
152:ASN
156:SER
160:VAL
164:GLN
168:GLU
172:GLU
176:MET
180:PRO
184:ARG
188:THR
192:GLY
196:ALA
200:LEU
204:SER
Auth_seq_ID