wwPDB NMR Structure Validation Summary Report (i)

Jun 17, 2021 – 12:08 PM BST

PDB ID	:	2C34
Title	:	Leishmania mexicana ICP
Authors	:	Smith, B.O.; Picken, N.C.; Bromek, K.; Westrop, G.D.; Mottram, J.C.; Coombs, G.H.
Deposited on	:	2005-10-04

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:

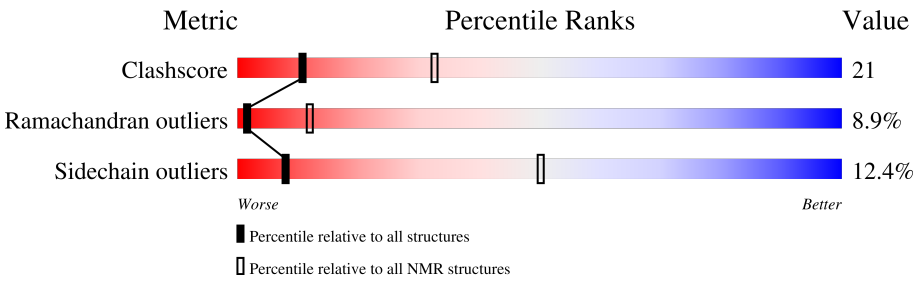
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.20
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.20

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	116	36% 33% • 27%

2 Ensemble composition and analysis (i)

This entry contains 38 models. Model 1 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:10-A:29, A:34-A:43, A:47-A:60, A:73-A:113 (85)	0.29	1

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 9 single-model clusters were found.

Cluster number	Models
1	1, 5, 6, 8, 10, 11, 12, 13, 14, 15, 16, 17, 21, 24, 25, 27, 29, 37, 38
2	4, 9, 26
3	19, 31, 33
4	18, 35
5	2, 34
Single-model clusters	3; 7; 20; 22; 23; 28; 30; 32; 36

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1855 atoms, of which 938 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called INHIBITOR OF CYSTEINE PEPTIDASES.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	116	1855	587	938	159	167	4	0

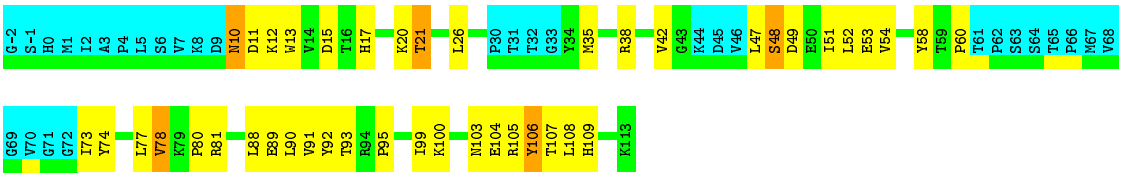
4 Residue-property plots (i)

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: INHIBITOR OF CYSTEINE PEPTIDASES

Chain A: 36% 33% • 27%

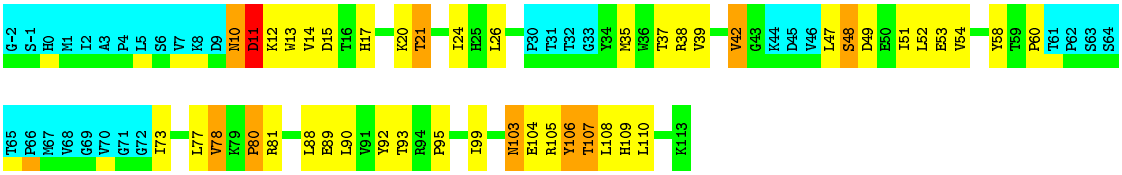


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

- Molecule 1: INHIBITOR OF CYSTEINE PEPTIDASES

Chain A: 34% 30% 8% • 27%



5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *ARIA-ALIKE*.

Of the 50 calculated structures, 38 were deposited, based on the following criterion: *LOWEST RESTRAINT ENERGY*.

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

Software name	Classification	Version
CNS	refinement	1.1
CcpNmr Analysis	structure solution	ANALYSIS

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	1456
Number of shifts mapped to atoms	1382
Number of unparsed shifts	0
Number of shifts with mapping errors	74
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	90%

6 Model quality (i)

6.1 Standard geometry (i)

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

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no planarity outliers.

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	708	722	717	30±5
All	All	26904	27436	27246	1126

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 21.

5 of 230 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:42:VAL:HG21	1:A:93:THR:HG21	0.80	1.53	4	22
1:A:38:ARG:NH2	1:A:47:LEU:HD12	0.76	1.95	28	1
1:A:10:ASN:HB2	1:A:106:TYR:HA	0.69	1.62	7	23
1:A:93:THR:HB	1:A:99:ILE:HG23	0.66	1.66	36	32
1:A:42:VAL:HG11	1:A:93:THR:HG23	0.65	1.68	32	13

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	84/116 (72%)	58±3 (70±3%)	18±3 (22±4%)	7±2 (9±2%)	1 12
All	All	3192/4408 (72%)	2220 (70%)	689 (22%)	283 (9%)	1 12

5 of 21 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	48	SER	37
1	A	21	THR	34
1	A	80	PRO	29
1	A	20	LYS	29
1	A	11	ASP	27

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	79/104 (76%)	69±2 (88±3%)	10±2 (12±3%)	8 50
All	All	3002/3952 (76%)	2631 (88%)	371 (12%)	8 50

5 of 39 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	78	VAL	38
1	A	58	TYR	35
1	A	10	ASN	32
1	A	53	GLU	31
1	A	103	ASN	24

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 85% 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	1456
Number of shifts mapped to atoms	1382
Number of unparsed shifts	0
Number of shifts with mapping errors	74
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	7

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Chain not found in structure. First 5 (of 74) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
UNMAPPED	1	GLY	HA2	3.879	0.003	1
UNMAPPED	1	GLY	HA3	3.879	0.003	1
UNMAPPED	1	GLY	C	177.851	0.005	1
UNMAPPED	1	GLY	CA	43.393	0.028	1
UNMAPPED	2	SER	H	8.671	0.003	1

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	116	-0.06 ± 0.11	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	105	0.16 ± 0.14	None needed (< 0.5 ppm)
$^{13}C'$	114	1.90 ± 0.40	Should be applied

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
^{15}N	106	-0.69 ± 0.36	None needed (imprecise)

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. 1006 atoms were assigned a chemical shift out of a possible 1120. 0 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	417/417 (100%)	166/166 (100%)	170/170 (100%)	81/81 (100%)
Sidechain	504/581 (87%)	308/339 (91%)	187/215 (87%)	9/27 (33%)
Aromatic	85/122 (70%)	45/62 (73%)	38/48 (79%)	2/12 (17%)
Overall	1006/1120 (90%)	519/567 (92%)	395/433 (91%)	92/120 (77%)

7.1.4 Statistically unusual chemical shifts (i)

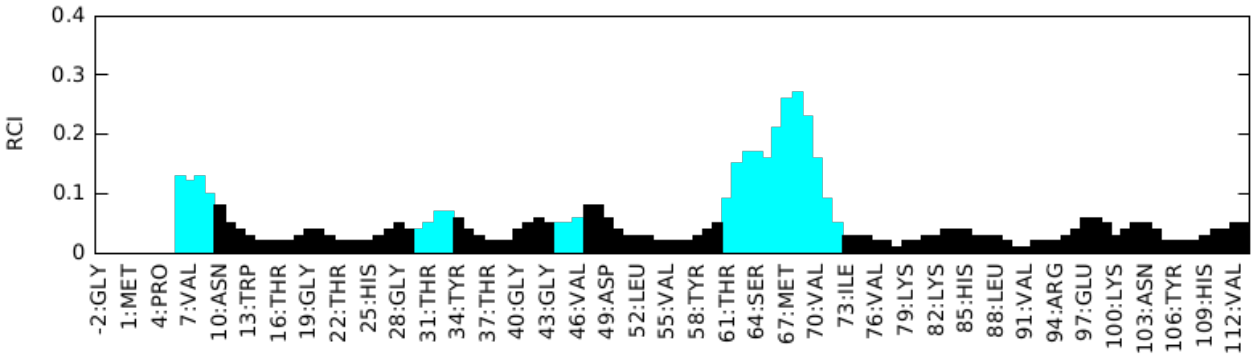
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	92	TYR	HB3	-0.19	4.75 – 0.95	-8.0
1	A	26	LEU	HD11	-1.03	2.16 – -0.64	-6.4
1	A	26	LEU	HD12	-1.03	2.16 – -0.64	-6.4
1	A	26	LEU	HD13	-1.03	2.16 – -0.64	-6.4
1	A	34	TYR	HB3	0.60	4.75 – 0.95	-5.9
1	A	30	PRO	C	168.43	184.42 – 169.02	-5.4
1	A	94	ARG	HG2	0.16	2.92 – 0.22	-5.2

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.

Random coil index (RCI) for chain A:


RCI
0.4
0.3
0.2
0.1
0
-2:GLY
1:MET
4:PRO
7:VAL
10:ASN
13:TRP
16:THR
19:GLY
22:THR
25:HIS
28:GLY
31:THR
34:TYR
37:THR
40:GLY
43:GLY
46:VAL
49:ASP
52:LEU
55:VAL
58:TYR
61:THR
64:SER
67:MET
70:VAL
73:ILE
76:VAL
79:LYS
82:LYS
85:HIS
88:LEU
91:VAL
94:ARG
97:GLU
100:LYS
103:ASN
106:TYR
109:HIS
112:VAL