wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 02:47 AM BST

PDB ID	:	2MIM
Title	:	NMR structure of the chicken CD3 epsilon delta/gamma heterodimer
Authors	:	Headey, S.; Berry, R.; Rossjohn, J.
Deposited on	:	2013-12-15

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
http://wwpdb.org/validation/2016/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:

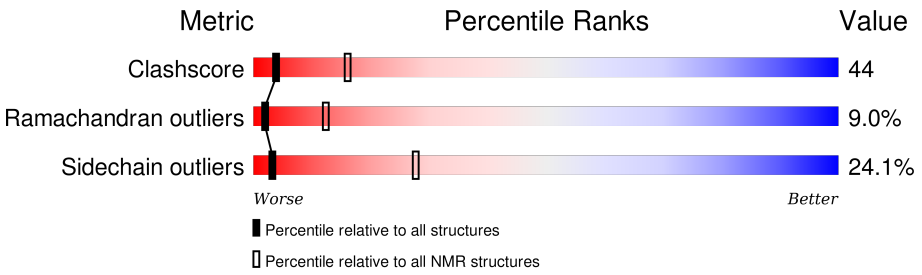
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20027457

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 73%.

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	114402	11133
Ramachandran outliers	111179	9975
Sidechain outliers	111093	9958

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	178	25% 37% 10% • 28%

2 Ensemble composition and analysis (i)

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:7-A:73, A:109-A:123, A:132-A:177 (128)	0.37	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 2 clusters and 2 single-model clusters were found.

Cluster number	Models
1	1, 3, 7, 9, 10
2	4, 6, 8
Single-model clusters	2; 5

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2687 atoms, of which 1320 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called CD3 epsilon protein, Linker, CD3 glycoprotein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	178	2687	837	1320	239	282	9	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP B3VMQ8
A	2	SER	-	EXPRESSION TAG	UNP B3VMQ8
A	3	HIS	-	EXPRESSION TAG	UNP B3VMQ8
A	4	MET	-	EXPRESSION TAG	UNP B3VMQ8

4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA and DNA 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: CD3 epsilon protein, Linker, CD3 glycoprotein

Chain A: 25% 37% 10% • 28%

G1 S2 H3 M4 G5 Q6 F9 A10 V11 E12 I13 S14 G15 T16 T17 V18 T19 I20 T21 S24 S25 G26 D27 D28 I29 K30 W31 K32 P33 D34 L37 G38 D39 N40 N41 K42 Y43 I44 I45 H48 D49 S50 S51 P52 L53 T54 V55 S56 C57 T58 A59 Q62 T65 M66

Y67 L68 N69 S74 A75 D76 D77 A78 K79 K80 D81 A82 A83 K84 K85 D86 D87 A88 K89 K90 D91 D92 A93 K94 K95 D96 G97 S98 V99 A100 K101 L102 G103 V104 H105 G106 L107 S108 M109 S110 V111 K112 E113 V114 S115 G116 K117 V118 F119 L120 E124 S125 K126 D127 L128 N129 T130 N131

Y132 L133 W134 K135 K136 E139 E140 L141 G142 N143 M144 L147 D148 L149 I152 Y153 D154 D155 P156 Y160 T161 C162 Q163 E166 N167 V168 N169 S170 T171 L172 H173 V174 H175 Y176 R177 M178

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: CD3 epsilon protein, Linker, CD3 glycoprotein

Chain A: 27% 35% 10% • 28%

G1 S2 H3 M4 G5 Q6 A10 V11 E12 I13 S14 G15 T16 T17 V18 T19 I20 S24 S25 G26 D27 D28 I29 K32 P33 D34 L37 N41 K42 Y43 I44 I45 H48 D49 S50 S51 P52 L53 T54 V55 S56 C57 T58 A59 G60 D61 Q62 T65 M66 Y67 L68 N69 A70

K71 S74 A75 D76 D77 A78 K79 K80 D81 A82 A83 K84 K85 D86 D87 A88 K89 K90 D91 D92 A93 K94 K95 D96 G97 S98 V99 A100 K101 L102 G103 V104 H105 G106 L107 S108 M109 S110 V111 K112 E113 V114 S115 G116 K117 V118 F119 L120 E124 S125 K126 D127 L128 N129 T130 N131 Y132 L133

W134 K135 K136 G137 K138 E139 E140 L141 G142 N143 M144 R145 Q146 L147 D148 L149 I152 Y153 D154 D155 P156 R157 Y160 T161 C162 Q163 E166 N167 V168 N169 S170 T171 L172 H173 V174 H175 M178

5 Refinement protocol and experimental data overview (i)

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 least restraint violations*.

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

Software name	Classification	Version
CNS	structure solution	
CNS	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)	2mim_cs.str
Number of chemical shift lists	1
Total number of shifts	1381
Number of shifts mapped to atoms	1381
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	73%

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

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	1004	961	961	86±7
All	All	10040	9610	9610	856

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 44.

5 of 257 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:20:ILE:HD11	1:A:45:ILE:HD11	1.10	1.18	3	10
1:A:147:LEU:HD22	1:A:149:LEU:HD21	1.01	1.31	7	7
1:A:147:LEU:HD22	1:A:149:LEU:HD11	0.96	1.35	6	6
1:A:67:TYR:CE1	1:A:111:VAL:HG13	0.92	1.99	3	10
1:A:19:THR:HG22	1:A:44:ILE:HD12	0.88	1.44	7	3

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	128/178 (72%)	99±3 (77±2%)	18±3 (14±2%)	12±1 (9±1%)	2 12
All	All	1280/1780 (72%)	989 (77%)	176 (14%)	115 (9%)	2 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	140	GLU	10
1	A	155	ASP	10
1	A	115	SER	10
1	A	10	ALA	10
1	A	168	VAL	10

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	113/151 (75%)	86±4 (76±3%)	27±4 (24±3%)	3 28
All	All	1130/1510 (75%)	858 (76%)	272 (24%)	3 28

5 of 64 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	144	MET	10
1	A	149	LEU	10
1	A	37	LEU	10
1	A	139	GLU	10
1	A	115	SER	10

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 carbohydrates 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 73% for the well-defined parts and 60% for the entire structure.

7.1 Chemical shift list 1

File name: 2mim_cs.str

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	1381
Number of shifts mapped to atoms	1381
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	4

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	147	-0.31 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	131	-0.56 ± 0.15	Should be applied
$^{13}C'$	0	—	—
^{15}N	135	0.71 ± 0.16	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 73%, i.e. 1110 atoms were assigned a chemical shift out of a possible 1531. 0 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	481/630 (76%)	239/251 (95%)	124/256 (48%)	118/123 (96%)
Sidechain	545/783 (70%)	335/458 (73%)	203/291 (70%)	7/34 (21%)

Continued on next page...

Continued from previous page...

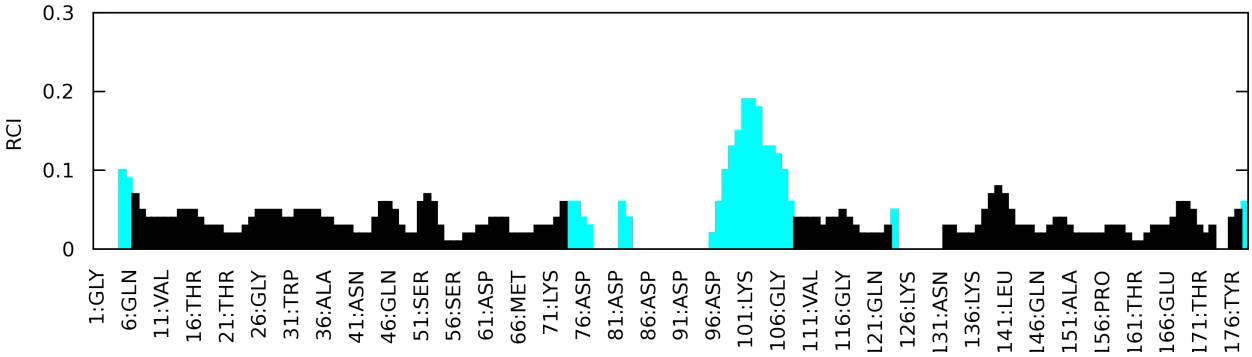
	Total	^{1}H	^{13}C	^{15}N
Aromatic	84/118 (71%)	45/62 (73%)	36/50 (72%)	3/6 (50%)
Overall	1110/1531 (73%)	619/771 (80%)	363/597 (61%)	128/163 (79%)

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	43	TYR	HB2	-0.49	4.76 – 1.06	-9.2
1	A	147	LEU	HB2	-0.68	3.32 – -0.08	-6.8
1	A	162	CYS	HB2	0.64	5.20 – 0.70	-5.1
1	A	23	PRO	HB2	0.29	3.82 – 0.32	-5.1

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: