wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 07:54 PM BST

PDB ID	:	2E4J
Title	:	Solution Structure of mouse Lipocalin-type Prostaglandin D Synthase
Authors	:	Shimamoto, S.; Ohkubo, T.
Deposited on	:	2006-12-11

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:

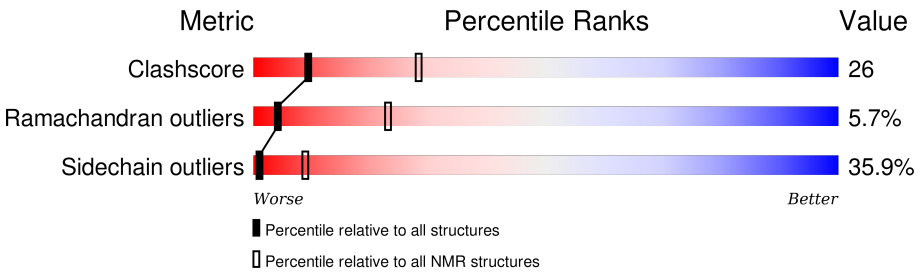
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 54%.

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	167	28% 47% 10% 15%

2 Ensemble composition and analysis (i)

This entry contains 15 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:30-A:84, A:88-A:108, A:114-A:137, A:142-A:183 (142)	0.53	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 3 clusters and 5 single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2592 atoms, of which 1282 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Prostaglandin-H2 D-isomerase.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	167	2592	828	1282	223	255	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	23	GLY	-	EXPRESSION TAG	UNP O09114
A	24	SER	-	EXPRESSION TAG	UNP O09114
A	89	ALA	CYS	ENGINEERED	UNP O09114
A	186	ALA	CYS	ENGINEERED	UNP O09114

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: Prostaglandin-H2 D-isomerase

Chain A: 28% 47% 10% 15%

G23 S24 Q25 G26 H27 D28 T29 V30 Q31 F34 D37 L40 G41 R42 W43 Y44 S45 A46 G47 L48 A49 S52 S53 W54 F55 R56 K59 L62 Y63 M64 C65 K66 T67 V68 V69 S72 T73 E74 G75 G76 L77 N78 L79 T80 S81 T82 F83 L84 R85 K86 N87 Q88 A89

E90 N94 V95 L96 Q97 P98 A99 G100 A101 P102 G103 H104 Y105 S108 S109 P110 H111 S112 G113 S114 I115 S119 V120 Y121 E122 E127 Y128 A129 L130 L131 F132 S133 R134 G135 T136 K137 G138 P139 G140 Q141 D142 F143 R144 M145 A146 T147 L148 Y149 S150 R151 T152 Q153 T154 L155 K156

L159 K160 E161 K162 F163 T164 T165 F166 S167 K168 A169 Q170 G171 L172 T173 E174 E175 D176 I177 V178 F179 L180 P181 Q182 P183 D184 K185 A186 I187 Q188 E189

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: Prostaglandin-H2 D-isomerase

Chain A: 34% 37% 14% 15%

G23 S24 Q25 G26 H27 D28 T29 V30 Q31 Q36 D37 K38 F39 L40 G41 R42 W43 Y44 S45 A46 G47 L48 A49 S50 N51 S52 K59 L62 Y63 M64 C65 K66 T67 V68 V69 S72 T73 E74 G75 G76 L77 N78 L79 T80 S81 T82 F83 L84 R85 K86 N87 Q88 A89 E90 L96

Q97 P98 A99 P102 G103 H104 Y105 S109 P110 H111 S112 G113 S114 I115 S119 V120 E127 Y128 L131 F132 S133 R134 K137 G138 P139 G140 Q141 D142 F143 R144 M145 A146 T147 L148 Y149 S150 R151 T152 Q153 T154 L155 K156 D157 E158 L159 K160 E161 K162 F163 T164 T165 F166 S167 K168

A169 Q170 D176 I177 V178 F179 L180 P181 Q182 P183 D184 K185 A186 I187 Q188 E189

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DISTANCE GEOMETRY, SIMULATED ANNEALING, TORSION ANGLE DYNAMICS*.

Of the 30 calculated structures, 15 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
CNS	refinement	1.1

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)	BMRB entry 10137
Number of chemical shift lists	1
Total number of shifts	1132
Number of shifts mapped to atoms	1120
Number of unparsed shifts	0
Number of shifts with mapping errors	12
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	54%

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	1128	1114	1112	59±4
All	All	16920	16710	16680	888

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 26.

5 of 316 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:148:LEU:HD21	1:A:159:LEU:HD13	1.09	1.20	5	13
1:A:148:LEU:HD11	1:A:159:LEU:HD22	0.94	1.36	12	8
1:A:159:LEU:HD12	1:A:160:LYS:N	0.93	1.79	13	10
1:A:84:LEU:HD13	1:A:89:ALA:HB2	0.85	1.48	15	3
1:A:164:THR:HG22	1:A:177:ILE:HD13	0.84	1.49	13	1

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	142/167 (85%)	115±3 (81±2%)	18±3 (13±2%)	8±2 (6±1%)	4 23
All	All	2130/2505 (85%)	1732 (81%)	276 (13%)	122 (6%)	4 23

5 of 24 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	102	PRO	15
1	A	46	ALA	15
1	A	183	PRO	11
1	A	128	TYR	10
1	A	75	GLY	9

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	123/142 (87%)	79±5 (64±4%)	44±5 (36±4%)	1 9
All	All	1845/2130 (87%)	1183 (64%)	662 (36%)	1 9

5 of 94 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	160	LYS	15
1	A	168	LYS	15
1	A	63	TYR	15
1	A	163	PHE	15
1	A	72	SER	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)

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 54% for the well-defined parts and 51% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 10137

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	1132
Number of shifts mapped to atoms	1120
Number of unparsed shifts	0
Number of shifts with mapping errors	12
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

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

- Residue not found in structure. First 5 (of 12) occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
A	67	CYS	N	125.995	-1.0	1
A	164	CYS	HB2	1.356	-1.0	1
A	67	CYS	HA	4.758	-1.0	1
A	164	CYS	H	8.218	-1.0	1
A	67	CYS	CA	52.272	-1.0	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
$^{13}C_{\alpha}$	119	-0.03 ± 0.13	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	110	0.01 ± 0.26	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—

Continued on next page...

Continued from previous page...

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
^{15}N	107	-0.64 ± 0.52	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 54%, i.e. 936 atoms were assigned a chemical shift out of a possible 1747. 10 out of 23 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	399/698 (57%)	199/278 (72%)	105/284 (37%)	95/136 (70%)
Sidechain	509/872 (58%)	309/513 (60%)	200/322 (62%)	0/37 (0%)
Aromatic	28/177 (16%)	28/93 (30%)	0/78 (0%)	0/6 (0%)
Overall	936/1747 (54%)	536/884 (61%)	305/684 (45%)	95/179 (53%)

7.1.4 Statistically unusual chemical shifts (i)

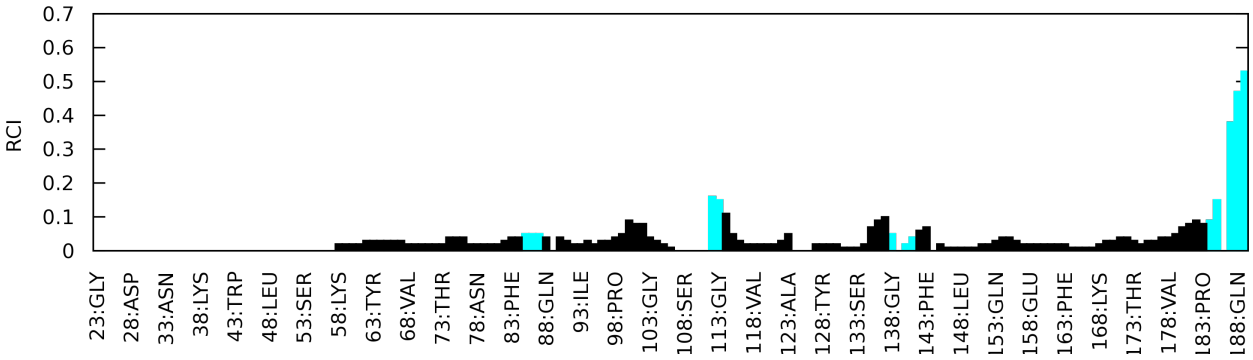
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	152	THR	CG2	36.73	27.15 – 15.95	13.5
1	A	88	GLN	CG	27.40	39.38 – 28.18	-5.7
1	A	140	GLY	CA	38.46	51.81 – 38.91	-5.4

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
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
23:GLY
28:ASP
33:ASN
38:LYS
43:TRP
48:LEU
53:SER
58:LYS
63:TYR
68:VAL
73:THR
78:ASN
83:PHE
88:GLN
93:ILE
98:PRO
103:GLY
108:SER
113:GLY
118:VAL
123:ALA
128:TYR
133:SER
138:GLY
143:PHE
148:LEU
153:GLN
158:GLU
163:PHE
168:LYS
173:THR
178:VAL
183:PRO
188:GLN