Full wwPDB NMR Structure Validation Report (i)

Apr 27, 2016 – 01:54 AM BST

PDB ID	:	2M0P
Title	:	Solution structure of the tenth complement type repeat of human megalin
Authors	:	Dagil, R.; Kragelund, B.
Deposited on	:	2012-11-01

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
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:

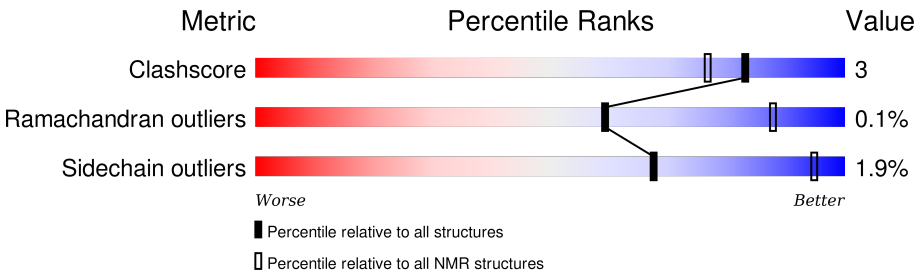
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 85%.

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	52	62% • 23% 12%

2 Ensemble composition and analysis (i)

This entry contains 20 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:1109-A:1142 (34)	0.34	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 4 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 636 atoms, of which 289 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Low-density lipoprotein receptor-related protein 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	46	635	198	289	60	82	6	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1149	HIS	-	EXPRESSION TAG	UNP P98164
A	1150	HIS	-	EXPRESSION TAG	UNP P98164
A	1151	HIS	-	EXPRESSION TAG	UNP P98164
A	1152	HIS	-	EXPRESSION TAG	UNP P98164
A	1153	HIS	-	EXPRESSION TAG	UNP P98164
A	1154	HIS	-	EXPRESSION TAG	UNP P98164

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	
			Total	Ca
2	A	1	1	1

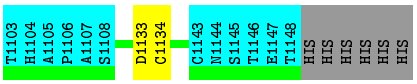
4 Residue-property plots (i)

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: Low-density lipoprotein receptor-related protein 2

Chain A:



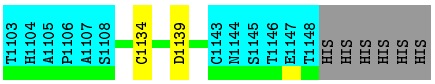
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1 (medoid)

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:



4.2.2 Score per residue for model 2

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

4.2.3 Score per residue for model 3

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

4.2.4 Score per residue for model 4

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

4.2.5 Score per residue for model 5

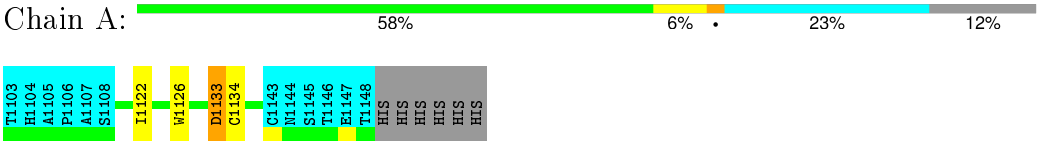
- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

4.2.6 Score per residue for model 6

- Molecule 1: Low-density lipoprotein receptor-related protein 2

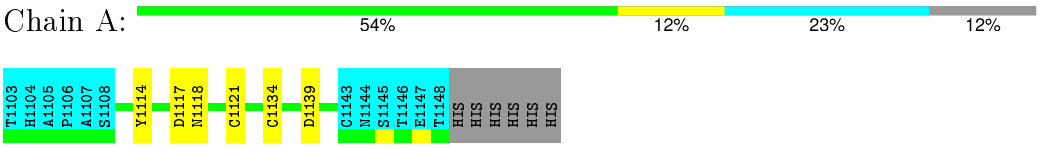
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A: 54% 10% • 23% 12%

T1103 H1104 A1105 P1106 A1107 S1108 Q1113 C1121 W1126 D1133 S1138 K1141 N1142 C1143 N1144 S1145 T1146 E1147 T1148 HIS HIS HIS HIS HIS HIS

4.2.9 Score per residue for model 9

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A: 52% 13% 23% 12%

T1103 H1104 A1105 P1106 A1107 S1108 Q1120 D1133 D1136 G1137 S1138 D1139 E1140 K1141 N1142 C1143 N1144 S1145 T1146 E1147 T1148 HIS HIS HIS HIS HIS HIS

4.2.10 Score per residue for model 10

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A: 58% 8% 23% 12%

T1103 H1104 A1105 P1106 A1107 S1108 T1112 Q1113 C1121 K1124 C1143 N1144 S1145 T1146 E1147 T1148 HIS HIS HIS HIS HIS HIS

4.2.11 Score per residue for model 11

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A: 62% • 23% 12%

T1103 H1104 A1105 P1106 A1107 S1108 C1134 D1139 C1143 N1144 S1145 T1146 E1147 T1148 HIS HIS HIS HIS HIS HIS

4.2.12 Score per residue for model 12

- Molecule 1: Low-density lipoprotein receptor-related protein 2

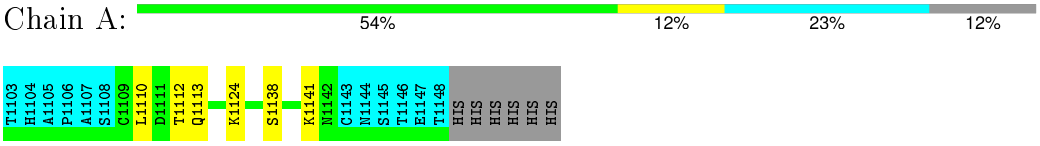
Chain A: 62% • 23% 12%

T1103 H1104 A1105 P1106 A1107 S1108 Q1120 C1134 C1143 N1144 S1145 T1146 E1147 T1148 HIS HIS HIS HIS HIS HIS

4.2.13 Score per residue for model 13

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:



4.2.14 Score per residue for model 14

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

4.2.15 Score per residue for model 15

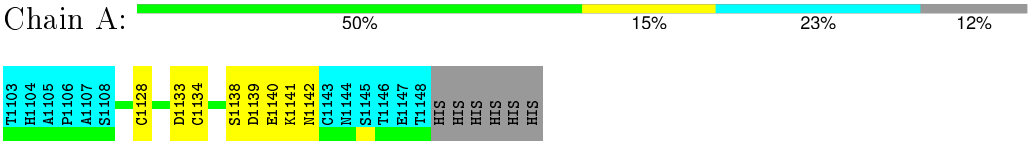
- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

4.2.16 Score per residue for model 16

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

4.2.18 Score per residue for model 18

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

4.2.19 Score per residue for model 19

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

4.2.20 Score per residue for model 20

- Molecule 1: Low-density lipoprotein receptor-related protein 2

Chain A:

5 Refinement protocol and experimental data overview (i)

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

Of the 200 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
ARIA	structure solution	
ARIA	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)	2m0p_cs.str
Number of chemical shift lists	1
Total number of shifts	451
Number of shifts mapped to atoms	451
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

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)

Bond lengths and bond angles in the following residue types are not validated in this section: CA

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	263	215	214	1±1
All	All	5280	4300	4280	29

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 3.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1134:CYS:SG	1:A:1139:ASP:HB3	0.64	2.33	17	8
1:A:1113:GLN:OE1	1:A:1121:CYS:HB3	0.61	1.96	10	2
1:A:1138:SER:O	1:A:1141:LYS:HG2	0.55	2.02	13	4
1:A:1126:TRP:HB3	1:A:1133:ASP:OD2	0.49	2.08	3	3
1:A:1114:TYR:O	1:A:1121:CYS:HA	0.47	2.09	7	1
1:A:1112:THR:O	1:A:1124:LYS:HG3	0.46	2.10	10	2
1:A:1133:ASP:O	1:A:1134:CYS:HB3	0.46	2.11	5	1
1:A:1122:ILE:HD13	1:A:1134:CYS:SG	0.44	2.52	6	1
1:A:1120:GLN:HB3	1:A:1134:CYS:SG	0.44	2.52	12	1
1:A:1116:CYS:SG	1:A:1120:GLN:HB3	0.43	2.53	19	1
1:A:1117:ASP:OD1	1:A:1118:ASN:N	0.43	2.52	7	1
1:A:1139:ASP:HA	1:A:1142:ASN:ND2	0.42	2.29	9	1

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1120:GLN:HA	1:A:1120:GLN:OE1	0.42	2.14	9	1
1:A:1110:LEU:CB	1:A:1113:GLN:HB2	0.41	2.45	13	1
1:A:1128:CYS:HA	1:A:1140:GLU:OE1	0.40	2.17	16	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	34/52 (65%)	31±2 (92±4%)	3±2 (8±5%)	0±0 (0±1%)	59	88
All	All	680/1040 (65%)	623 (92%)	56 (8%)	1 (0%)	59	88

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	1136	ASP	1

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	32/48 (67%)	31±1 (98±2%)	1±1 (2±2%)	67	95
All	All	640/960 (67%)	628 (98%)	12 (2%)	67	95

All 3 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	1133	ASP	10
1	A	1128	CYS	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	1142	ASN	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 carbohydrates in this entry.

6.6 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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 85% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: 2m0p_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	451
Number of shifts mapped to atoms	451
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	2

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	46	-0.27 ± 0.22	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	44	0.57 ± 0.17	Should be applied
$^{13}C'$	38	-0.56 ± 0.22	Should be applied
^{15}N	39	0.43 ± 0.64	None needed (< 0.5 ppm)

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 85%, i.e. 316 atoms were assigned a chemical shift out of a possible 373. 2 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	156/170 (92%)	63/68 (93%)	63/68 (93%)	30/34 (88%)
Sidechain	139/175 (79%)	86/101 (85%)	49/66 (74%)	4/8 (50%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	21/28 (75%)	11/14 (79%)	9/11 (82%)	1/3 (33%)
Overall	316/373 (85%)	160/183 (87%)	121/145 (83%)	35/45 (78%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 83%, i.e. 409 atoms were assigned a chemical shift out of a possible 490. 2 out of 2 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	207/228 (91%)	84/91 (92%)	84/92 (91%)	39/45 (87%)
Sidechain	179/226 (79%)	111/131 (85%)	64/86 (74%)	4/9 (44%)
Aromatic	23/36 (64%)	12/18 (67%)	10/13 (77%)	1/5 (20%)
Overall	409/490 (83%)	207/240 (86%)	158/191 (83%)	44/59 (75%)

7.1.4 Statistically unusual chemical shifts (i)

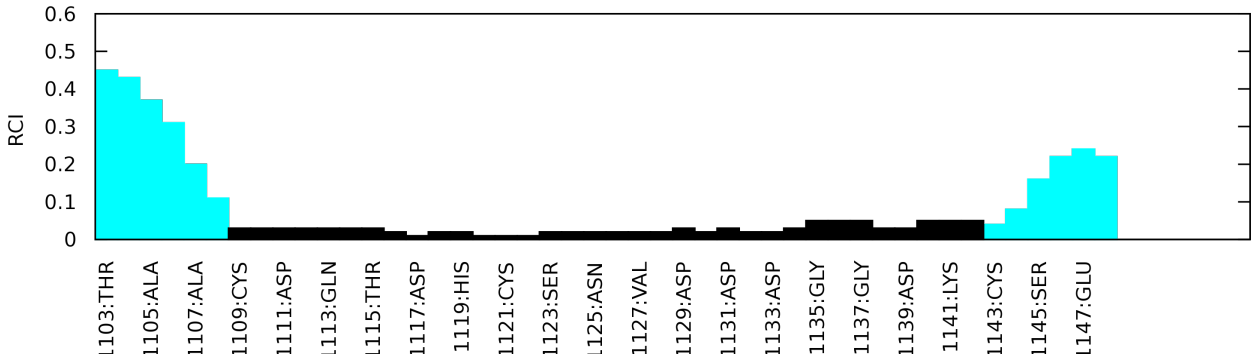
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	1123	SER	HB2	2.21	5.18 – 2.58	-6.4
1	A	1132	ASN	CB	30.05	47.13 – 30.23	-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:


RCI
0.6
0.5
0.4
0.3
0.2
0.1
0
1103:THR
1105:ALA
1107:ALA
1109:CYS
1111:ASP
1113:GLN
1115:THR
1117:ASP
1119:HIS
1121:CYS
1123:SER
1125:ASN
1127:VAL
1129:ASP
1131:ASP
1133:ASP
1135:GLY
1137:GLY
1139:ASP
1141:LYS
1143:CYS
1145:SER
1147:GLU