wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 03:28 AM BST

PDB ID	:	2MGY
Title	:	Solution structure of the mitochondrial translocator protein (TSPO) in complex with its high-affinity ligand PK11195
Authors	:	Jaremko, M.; Jaremko, L.; Giller, K.; Becker, S.; Zweckstetter, M.
Deposited on	:	2013-11-11

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:

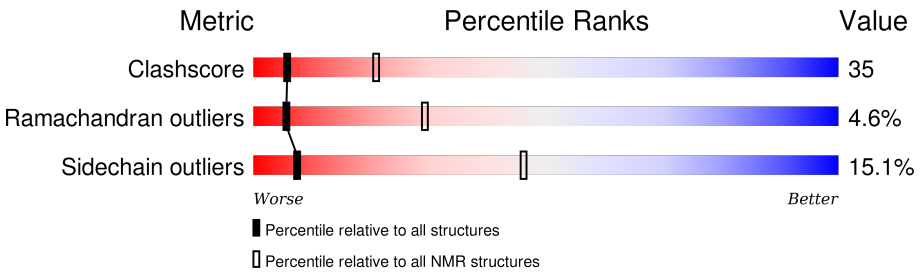
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 15%.

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	169	40% 40% 7% • 14%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 8 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:5-A:40, A:49-A:158 (146)	0.29	8

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 4 clusters and 1 single-model cluster was found.

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

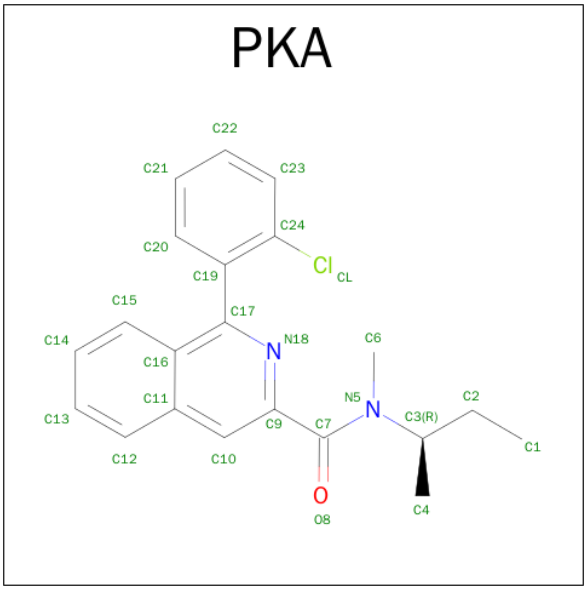
3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2691 atoms, of which 1330 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Translocator protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	169	2645	888	1309	223	220	5	0

- Molecule 2 is N-[(2R)-BUTAN-2-YL]-1-(2-CHLOROPHENYL)-N-METHYLISOQUINOLINE-3-CARBOXAMIDE (three-letter code: PKA) (formula: $C_{21}H_{21}ClN_2O$).



Mol	Chain	Residues	Atoms					
			Total	C	Cl	H	N	O
2	A	1	46	21	1	21	2	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: Translocator protein

Chain A: 40% 40% 7% • 14%

M1 P2 E3 S4 W5 V6 P7 A8 V9 T12 L13 S16 A23 V26 R27 W33 L37 Q38 K39 P40 S41 W42 H43 P44 P45 R46 W47 T48 L49 A50 P51 I52 W53 Y57 M60 S64 V67 W68 K69 E70 L71 G72 G73 F74 T75 E76 P81 T86

L89 A90 L91 N92 W93 A94 W95 P96 P97 I98 F99 F100 G101 A102 R103 Q104 M105 G106 W107 A108 D111 L112 V115 S116 G117 V118 A119 T120 A121 T122 T123 W126 H127 S130 P131 P132 L137 Y138 P139 Y140 L141 A142 W143 L144 A145 V149 L150 N151 W155 R156 D157 N158

S159 G160 R161 R162 G163 G164 S165 R166 L167 A168 E169

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

The representative model is number 8. Colouring as in section 4.1 above.

- Molecule 1: Translocator protein

Chain A: 42% 36% 8% • 14%

M1 P2 E3 S4 W5 V6 P7 A8 V9 G10 L11 T12 A23 V26 R27 W33 G36 L37 Q38 K39 P40 S41 W42 H43 P44 P45 R46 W47 T48 L49 A50 P51 I52 W53 G54 T55 L56 Y57 M60 S64 V67 W68 L71 G72 G73 F74 T75 E76 L84 Q88

L89 A90 L91 N92 W93 A94 W95 P96 P97 I98 F99 F100 G101 A102 R103 Q104 M105 G106 W107 A108 D111 L112 V115 S116 G117 V118 A119 T120 A121 T122 T123 W126 H127 P131 P132 A133 A134 L137 Y138 P139 Y140 L141 L144 A145 V149 L150 N151 N158 S159 G160 R161 R162

G163 G164 S165 R166 L167 A168 E169

5 Refinement protocol and experimental data overview (i)

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

Of the 105 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
X-PLOR NIH	refinement	
X-PLOR NIH	structure solution	

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)	2mgy_cs.str
Number of chemical shift lists	1
Total number of shifts	316
Number of shifts mapped to atoms	316
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	15%

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: PKA

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	1150	1135	1134	79±6
2	A	25	21	21	3±1
All	All	23500	23120	23100	1623

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 35.

5 of 362 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:49:LEU:O	1:A:52:ILE:HG23	0.84	1.71	3	20
1:A:90:ALA:O	1:A:94:ALA:HB2	0.76	1.80	11	19
1:A:140:TYR:CE2	1:A:144:LEU:HD11	0.75	2.16	10	13
1:A:37:LEU:C	1:A:37:LEU:HD12	0.72	2.05	9	4
1:A:95:TRP:N	1:A:96:PRO:CD	0.71	2.54	2	20

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	146/169 (86%)	120±3 (82±2%)	19±3 (13±2%)	7±2 (5±1%)	5 29
All	All	2920/3380 (86%)	2399 (82%)	386 (13%)	135 (5%)	5 29

5 of 10 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	94	ALA	20
1	A	51	PRO	20
1	A	73	GLY	18
1	A	7	PRO	17
1	A	107	TRP	16

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	110/129 (85%)	93±3 (85±3%)	17±3 (15±3%)	7 46
All	All	2200/2580 (85%)	1867 (85%)	333 (15%)	7 46

5 of 48 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	123	THR	20
1	A	107	TRP	20
1	A	6	VAL	19
1	A	68	TRP	17
1	A	37	LEU	17

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)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	PKA	A	201	-	27,27,27	1.97±0.03	1±0 (3±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	PKA	A	201	-	37,38,38	1.37±0.02	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means

no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	PKA	A	201	-	-	0±0,18,18,18	0±0,3,3,3

All unique bond outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
2	A	201	PKA	C17-N18	8.12	1.40	1.32	18	20

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 15% for the well-defined parts and 15% for the entire structure.

7.1 Chemical shift list 1

File name: 2mgy_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	316
Number of shifts mapped to atoms	316
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	168	-1.68 ± 0.16	Should be applied
$^{13}C_{\beta}$	148	0.31 ± 0.08	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	0	—	—

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 15%, i.e. 273 atoms were assigned a chemical shift out of a possible 1769. 0 out of 33 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	145/710 (20%)	0/282 (0%)	145/292 (50%)	0/136 (0%)
Sidechain	128/797 (16%)	0/467 (0%)	128/304 (42%)	0/26 (0%)

Continued on next page...

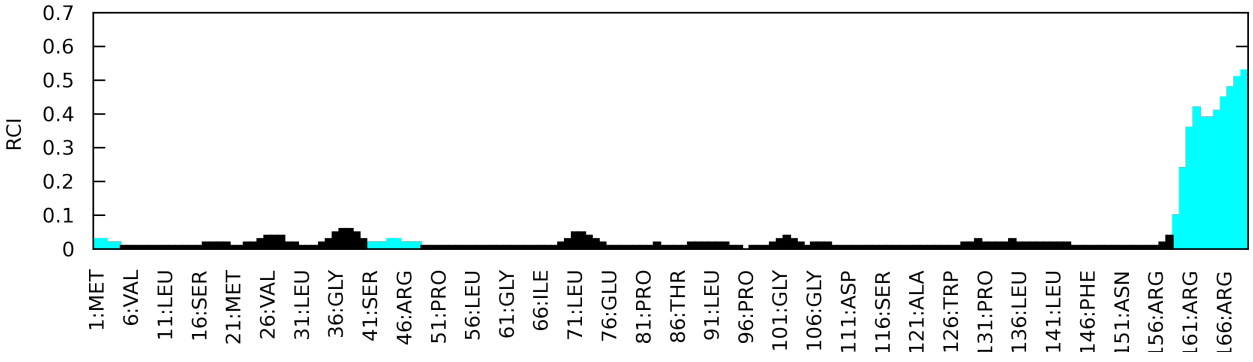
Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	0/262 (0%)	0/134 (0%)	0/116 (0%)	0/12 (0%)
Overall	273/1769 (15%)	0/883 (0%)	273/712 (38%)	0/174 (0%)

7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

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: