wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 12:24 AM BST

PDB ID	:	2L23
Title	:	NMR structure of the ACID (ACtivator Interacting Domain) of the human mediator Med25 protein
Authors	:	Bontems, F.; Monte, D.; Dewitte, F.; Villeret, V.
Deposited on	:	2010-08-10

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:

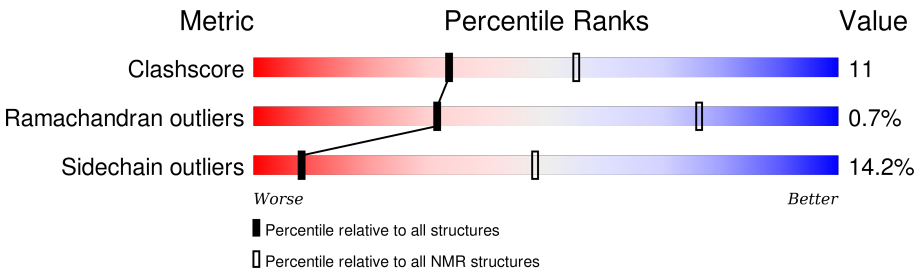
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 was not calculated.

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	54% 19% • 20% •

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative.

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:8-A:21, A:34-A:111, A:120-A:154 (127)	0.32	11

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 1 clusters. No single-model clusters were found.

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

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 2609 atoms, of which 1319 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Mediator of RNA polymerase II transcription subunit 25.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	161	2609	828	1319	227	226	9	0

There are 9 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	INITIATING METHIONINE	UNP B9TX31
A	160	LEU	-	EXPRESSION TAG	UNP B9TX31
A	161	GLU	-	EXPRESSION TAG	UNP B9TX31
A	162	HIS	-	EXPRESSION TAG	UNP B9TX31
A	163	HIS	-	EXPRESSION TAG	UNP B9TX31
A	164	HIS	-	EXPRESSION TAG	UNP B9TX31
A	165	HIS	-	EXPRESSION TAG	UNP B9TX31
A	166	HIS	-	EXPRESSION TAG	UNP B9TX31
A	167	HIS	-	EXPRESSION TAG	UNP B9TX31

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: Mediator of RNA polymerase II transcription subunit 25

Chain A: 54% 19% • 20% •

M1 G2 Q3 Q4 S5 V6 S7 N8 K9 L10 L11 A12 W13 L17 E18 W19 K22 P23 K24 P25 A26 S27 V28 D29 A30 N31 T32 K33 L34 T35 L38 V42 Y43 V44 T52 W55 L59 I60 M61 Q62 L63 Q66 L69 T70 T71 L72 L75 F76 R77

R80 F84 N88 L97 I100 P112 H113 T114 A115 P116 C117 E118 V119 M123 L124 L125 Y126 S127 K130 Q141 I148 R149 Q150 V151 H155 K156 Q157 V158 Q159 L160 E161 HIS HIS HIS HIS HIS HIS

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

The representative model is number 11. Colouring as in section 4.1 above.

- Molecule 1: Mediator of RNA polymerase II transcription subunit 25

Chain A: 51% 19% 5% 20% •

M1 G2 Q3 Q4 S5 V6 S7 N8 K9 L10 L11 A12 W13 L17 E18 W19 K22 P23 K24 P25 A26 S27 V28 D29 A30 N31 T32 K33 L34 T35 V42 Y43 V44 N45 H46 L50 K51 W55 K58 L59 I60 M61 Q66 L69 T70 T71 L72 L75 F76 R77

R80 N88 K89 D90 L94 L97 P112 H113 T114 A115 P116 C117 E118 V119 L122 L125 Y126 S127 K130 I137 R149 Q150 V151 I152 T153 N154 H155 K156 Q157 V158 Q159 L160 E161 HIS HIS HIS HIS HIS HIS

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *TORSION ANGLE DYNAMICS*.

Of the 20 calculated structures, 20 were deposited, based on the following criterion: *ALL CALCULATED STRUCTURES SUBMITTED*.

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

Software name	Classification	Version
CYANA	refinement	2.1
CYANA	structure solution	2.1
CCPNMR	structure solution	

No chemical shift data was provided. 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	1035	1063	1063	23±4
All	All	20700	21260	21260	458

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 11.

5 of 122 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:63:LEU:HD13	1:A:100:ILE:HG21	0.78	1.52	9	2
1:A:97:LEU:HD12	1:A:100:ILE:HD12	0.72	1.60	6	3
1:A:122:LEU:HD12	1:A:137:ILE:HD11	0.70	1.62	17	5
1:A:147:GLY:O	1:A:151:VAL:HG23	0.67	1.89	4	2
1:A:19:TRP:CH2	1:A:97:LEU:HD12	0.66	2.25	3	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	127/167 (76%)	114±2 (89±1%)	12±2 (10±2%)	1±1 (1±1%)	31 76
All	All	2540/3340 (76%)	2273 (89%)	248 (10%)	19 (1%)	31 76

All 5 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	130	LYS	8
1	A	34	LEU	5
1	A	104	GLY	4
1	A	88	ASN	1
1	A	87	THR	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	115/151 (76%)	99±3 (86±2%)	16±3 (14±2%)	8 48
All	All	2300/3020 (76%)	1973 (86%)	327 (14%)	8 48

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	69	LEU	20
1	A	55	TRP	20
1	A	11	LEU	20
1	A	127	SER	20
1	A	77	ARG	19

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)

No chemical shift data were provided