wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 11:20 PM BST

PDB ID	:	2KKT
Title	:	Solution structure of the SCA7 domain of human Ataxin-7-L3 protein
Authors	:	Wang, Y.; Atkinson, A.R.; Bonnet, J.; Romier, C.; Kieffer, B.
Deposited on	:	2009-06-29

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:

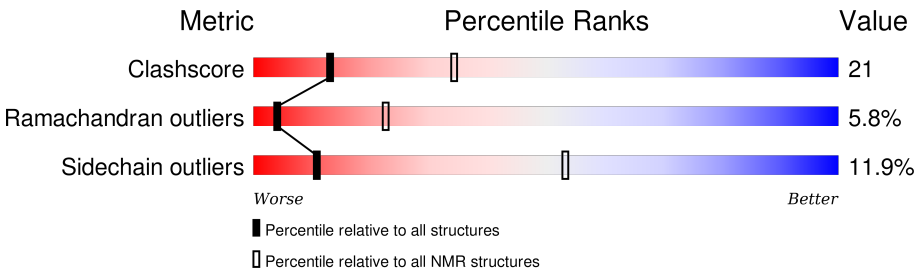
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	84	30% 13% 5% • 13% 38%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 15 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:198-A:212, A:218-A:243 (41)	0.26	15

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 4 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 839 atoms, of which 423 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ataxin-7-like protein 3.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	52	838	256	423	79	76	4	0

There are 5 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	193	GLY	-	EXPRESSION TAG	UNP Q14CW9
A	194	SER	-	EXPRESSION TAG	UNP Q14CW9
A	195	HIS	-	EXPRESSION TAG	UNP Q14CW9
A	196	MET	-	EXPRESSION TAG	UNP Q14CW9
A	239	VAL	ILE	ENGINEERED MUTATION	UNP Q14CW9

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
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: Ataxin-7-like protein 3

Chain A: 30% 13% 5% • 13% 38%

GLY SER HIS MET GLY P198 L201 C209 G210 V211 I212 S213 E214 H215 T216 K217 K218 R222 S223 L224 R225 C226 P227 Q228 H229 T230 Q233 R234 V237 R238 V239 L242 G243 P244 S245 A246 V247 L248 P249 GLU VAL GLU SER SER LEU ASP ASN ASP SER PHE ASP MET THR ASP

SER GLN ALA LEU ILE SER ARG LEU GLN TRP ASP GLY

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

The representative model is number 15. Colouring as in section 4.1 above.

- Molecule 1: Ataxin-7-like protein 3

Chain A: 33% 10% • • 13% 38%

GLY SER HIS MET GLY P198 L201 C209 G210 V211 I212 S213 E214 H215 T216 K217 R222 S223 L224 R225 C226 P227 Q228 H229 T230 Q233 R234 L242 G243 P244 S245 A246 V247 L248 P249 GLU VAL GLU SER SER LEU ASP ASN ASP SER PHE ASP MET THR ASP SER GLN ALA LEU ILE

SER ARG LEU GLN TRP ASP GLY

5 Refinement protocol and experimental data overview (i)

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

Of the 64 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	structure solution	
X-PLOR NIH	refinement	

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)

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

The Z score for a bond length (or angle) is the number of standard deviations the observed value is removed from the expected value. A bond length (or angle) with $|Z| > 5$ is considered an outlier worth inspection. RMSZ is the (average) root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Chain	Bond lengths		Bond angles	
		RMSZ	#Z>5	RMSZ	#Z>5
1	A	1.14±0.05	2±1/339 (0.5±0.4%)	0.96±0.05	0±0/457 (0.1±0.1%)
All	All	1.14	33/6780 (0.5%)	0.96	9/9140 (0.1%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	0.8±0.5
All	All	0	17

5 of 6 unique bond outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)	Models	
								Worst	Total
1	A	234	ARG	NE-CZ	-10.37	1.19	1.33	2	1
1	A	229	HIS	CG-ND1	7.60	1.55	1.38	9	13
1	A	227	PRO	N-CD	-6.69	1.38	1.47	17	9
1	A	229	HIS	CB-CG	6.11	1.61	1.50	2	5
1	A	220	CYS	CB-SG	-5.59	1.72	1.81	17	2

All unique angle outliers are listed below. They are sorted according to the Z-score of the worst occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	227	PRO	N-CA-CB	-8.20	93.46	103.30	9	8
1	A	234	ARG	NE-CZ-NH2	-5.01	117.79	120.30	4	1

There are no chirality outliers.

All unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	234	ARG	Sidechain	15
1	A	235	ARG	Sidechain	2

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	334	341	341	14±2
All	All	6700	6820	6820	290

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 21.

5 of 60 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:224:LEU:HB3	1:A:234:ARG:HD3	0.77	1.56	3	8
1:A:234:ARG:CZ	1:A:234:ARG:HA	0.73	2.13	7	6
1:A:229:HIS:HB3	1:A:234:ARG:HH21	0.71	1.46	5	8
1:A:229:HIS:HB3	1:A:234:ARG:NH2	0.67	2.04	14	8
1:A:209:CYS:HB2	1:A:222:ARG:O	0.64	1.93	12	9

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	40/84 (48%)	32±1 (81±2%)	5±1 (13±3%)	2±1 (6±2%)	4 23

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	800/1680 (48%)	647 (81%)	107 (13%)	46 (6%)	4 22

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	227	PRO	20
1	A	209	CYS	12
1	A	226	CYS	6
1	A	218	LYS	5
1	A	222	ARG	3

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	39/77 (51%)	34±1 (88±2%)	5±1 (12±2%)	11 53
All	All	780/1540 (51%)	687 (88%)	93 (12%)	11 53

5 of 12 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	226	CYS	20
1	A	201	LEU	17
1	A	209	CYS	17
1	A	242	LEU	17
1	A	234	ARG	8

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)

No chemical shift data were provided