wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 06:33 PM BST

PDB ID	:	1YUA
Title	:	C-TERMINAL DOMAIN OF ESCHERICHIA COLI TOPOISOMERASE I
Authors	:	Yu, L.; Zhu, C.-X.; Tse-Dinh, Y.-C.; Fesik, S.W.
Deposited on	:	1995-03-02

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:

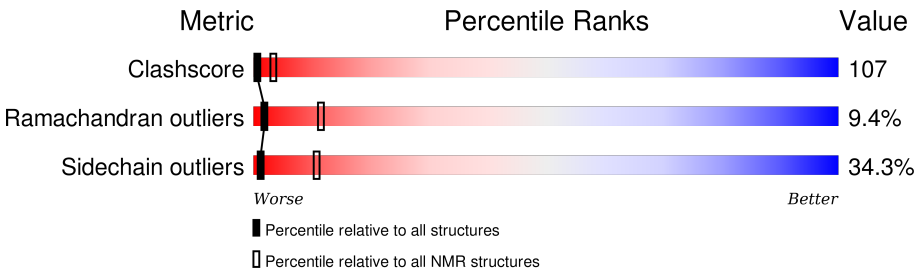
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	122	10% 52% 25% • 11%

2 Ensemble composition and analysis (i)

This entry contains 26 models. Model 23 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:12-A:120 (109)	0.49	23

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 2 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26
2	4, 5
Single-model clusters	14

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1943 atoms, of which 966 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called TOPOISOMERASE I.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	122	1943	620	966	170	184	3	0

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: TOPOISOMERASE I

Chain A: 10% 52% 25% • 11%

M1 N2 G3 E4 V5 A6 P7 P8 K9 E10 D11 P12 V13 P14 L15 P16 E17 L18 P19 C20 E21 K22 S23 D24 A25 Y26 F27 V28 L29 R30 D31 G32 A33 A34 G35 V36 F37 L38 A39 A40 N41 T42 F43 P44 K45 S46 R47 E48 T49 R50 A51 P52 L53 V54 E55 E56 L57 Y58 R59 F60

R61 D62 R63 L64 P65 E66 K67 L68 R69 Y70 L71 A72 D73 A74 P75 Q76 Q77 D78 P79 E80 G81 N82 K83 T84 M85 V86 R87 F88 S89 R90 K91 T92 K93 Q94 Q95 Y96 V97 S98 S99 E100 K101 D102 G103 K104 A105 T106 G107 W108 S109 A110 F111 Y112 V113 D114 G115 K116 W117 V118 E119 G120

K121 K122

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

The representative model is number 23. Colouring as in section 4.1 above.

- Molecule 1: TOPOISOMERASE I

Chain A: 14% 45% 27% • 11%

M1 N2 G3 E4 V5 A6 P7 P8 K9 E10 D11 P12 V13 P14 L15 P16 E17 L18 P19 C20 E21 K22 S23 D24 A25 Y26 F27 V28 L29 R30 D31 A34 G35 V36 F37 L38 A39 A40 N41 T42 F43 P44 K45 S46 R47 E48 T49 R50 A51 P52 L53 V54 E55 E56 L57 Y58 R59 F60 R61

D62 R63 L64 P65 E66 K67 L68 R69 Y70 L71 A72 D73 A74 P75 Q76 Q77 D78 P79 E80 G81 N82 K83 T84 M85 V86 R87 F88 S89 R90 K91 T92 K93 Q94 Q95 Y96 V97 S98 S99 E100 K101 D102 G103 K104 A105 T106 G107 W108 S109 A110 F111 Y112 V113 D114 G115 K116 W117 V118 E119 G120 K121

K122

5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 26 were deposited, based on the following criterion: *?*.

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

Software name	Classification	Version
X-PLOR	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)

There are no covalent bond-length or bond-angle outliers.

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	8.7±0.6
All	All	0	225

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

5 of 9 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	69	ARG	Sidechain	26
1	A	87	ARG	Sidechain	26
1	A	47	ARG	Sidechain	26
1	A	63	ARG	Sidechain	26
1	A	30	ARG	Sidechain	26

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	877	863	863	186±12
All	All	22802	22438	22438	4830

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 107.

5 of 1111 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:57:LEU:HD22	1:A:72:ALA:HB2	1.05	1.11	4	4
1:A:20:CYS:SG	1:A:25:ALA:HB3	1.05	1.92	6	3
1:A:57:LEU:HD12	1:A:72:ALA:HB2	1.05	1.29	21	13
1:A:29:LEU:HD11	1:A:68:LEU:HD21	1.02	1.29	22	2
1:A:74:ALA:HB2	1:A:117:TRP:CE2	1.02	1.89	25	25

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	109/122 (89%)	74±3 (68±3%)	25±2 (23±2%)	10±2 (9±2%)	2	11
All	All	2834/3172 (89%)	1920 (68%)	647 (23%)	267 (9%)	2	11

5 of 36 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	77	GLN	26
1	A	113	VAL	26
1	A	12	PRO	22
1	A	83	LYS	20
1	A	66	GLU	18

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	92/103 (89%)	60±4 (66±4%)	32±4 (34±4%)	1	11
All	All	2392/2678 (89%)	1571 (66%)	821 (34%)	1	11

5 of 72 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	101	LYS	26
1	A	53	LEU	26
1	A	118	VAL	26
1	A	18	LEU	25
1	A	91	LYS	25

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