wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 02:31 PM BST

PDB ID	:	1ENW
Title	:	ELONGATION FACTOR TFIIS DOMAIN II
Authors	:	Morin, P.E.; Awrey, D.E.; Edwards, A.M.; Arrowsmith, C.H.
Deposited on	:	2000-03-21

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:

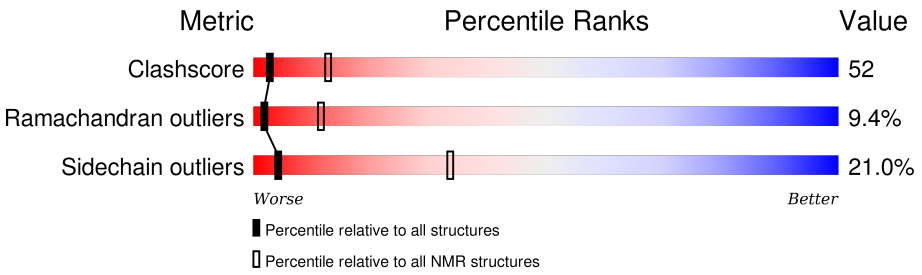
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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	114	23% 46% 6% • 22%

2 Ensemble composition and analysis (i)

This entry contains 20 models. The atoms present in the NMR models are not consistent. Some calculations may have failed as a result. All residues are included in the validation scores. Model 17 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:13-A:88 (76)	0.86	17
2	A:101-A:113 (13)	0.90	8

Ill-defined regions of proteins are excluded from the global statistics.

Ligands and non-protein polymers are included in the analysis.

NmrClust was unable to cluster the ensemble.

Error message: Inconsistent models in file

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 1776 atoms, of which 883 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called TRANSCRIPTION ELONGATION FACTOR S-II.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	114	1776	554	883	161	174	4	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION ARTIFACT	UNP P07273
A	2	SER	-	EXPRESSION ARTIFACT	UNP P07273
A	3	HIS	-	EXPRESSION ARTIFACT	UNP P07273
A	4	MET	-	EXPRESSION ARTIFACT	UNP P07273

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: TRANSCRIPTION ELONGATION FACTOR S-II

Chain A: 23% 46% 6% • 22%

G1 S2 H3 M4 P5 R6 N7 S8 K9 N10 D11 G12 V13 D14 T15 A16 I17 Y18 H19 H20 K21 L22 R23 D24 Q25 V26 L27 K28 A29 L30 Y31 D32 V33 L34 S38 E39 H40 P41 P42 Q43 S44 I45 L46 H47 T48 I52 E53 S54 E55 M56 N57 K58 V59 N60 N61 C62 D63 T64

N65 E66 Y69 R72 Y73 R74 I75 I76 Y77 I81 S82 K83 N84 N85 P86 D87 L88 K89 H90 K91 I92 A93 N94 G95 D96 I97 T98 P99 E100 F101 L102 A103 T104 C105 D106 A107 K108 D109 L110 A111 P112 A113 P114

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

The representative model is number 17. Colouring as in section 4.1 above.

- Molecule 1: TRANSCRIPTION ELONGATION FACTOR S-II

Chain A: 25% 39% 11% • 22%

G1 S2 H3 M4 P5 R6 N7 S8 K9 N10 D11 G12 V13 D14 T15 A16 I17 Y18 H19 H20 K21 L22 R23 L27 K28 A29 L30 Y31 D32 V33 L34 A35 K36 E37 S38 E39 H40 P41 P42 Q43 S44 I45 L46 H47 T48 I52 E53 S54 E55 M56 N57 K58 V59 N60 N61 C62 D63 T64

N65 E66 A67 A68 Y69 R72 Y73 R74 Y77 I81 S82 K83 N84 N85 P86 D87 L88 K89 H90 K91 I92 A93 N94 G95 D96 I97 T98 P99 E100 F101 C105 L110 A111 P112 A113 P114

5 Refinement protocol and experimental data overview (i)

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

Of the 50 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	structure solution	3.851
X-PLOR	refinement	3.851

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)

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	1.10±0.01	4±0/718 (0.6±0.0%)	1.11±0.02	0±0/977 (0.0±0.0%)
All	All	1.10	80/14360 (0.6%)	1.11	1/19540 (0.0%)

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	2.8±0.5
All	All	0	56

All 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	20	HIS	CG-ND1	-6.29	1.25	1.38	9	20
1	A	40	HIS	CG-ND1	-6.24	1.25	1.38	11	20
1	A	19	HIS	CG-ND1	-6.20	1.25	1.38	11	20
1	A	47	HIS	CG-ND1	-6.19	1.25	1.38	19	20

All unique angle outliers are listed below.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	20	HIS	N-CA-C	5.11	124.80	111.00	19	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	72	ARG	Sidechain	20
1	A	74	ARG	Sidechain	19
1	A	23	ARG	Sidechain	17

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	704	700	700	72±7
All	All	14080	13464	14000	1449

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 52.

5 of 561 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:17:ILE:HG23	1:A:18:TYR:H	0.97	1.16	17	11
1:A:88:LEU:H	1:A:88:LEU:HD22	0.95	1.16	3	1
1:A:17:ILE:HG23	1:A:18:TYR:N	0.84	1.88	19	10
1:A:26:VAL:HG12	1:A:30:LEU:HD11	0.83	1.47	6	8
1:A:102:LEU:HD22	1:A:102:LEU:H	0.81	1.34	10	2

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	89/114 (78%)	55±3 (62±4%)	26±3 (29±3%)	8±2 (9±2%)	2 11
All	All	1780/2280 (78%)	1097 (62%)	515 (29%)	168 (9%)	2 11

5 of 32 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	44	SER	17
1	A	19	HIS	16
1	A	62	CYS	16
1	A	17	ILE	15
1	A	20	HIS	11

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	77/98 (79%)	61±3 (79±3%)	16±3 (21±3%)	4	33
All	All	1540/1960 (79%)	1216 (79%)	324 (21%)	4	33

5 of 61 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	110	LEU	20
1	A	46	LEU	18
1	A	27	LEU	17
1	A	40	HIS	12
1	A	101	PHE	11

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