Full wwPDB NMR Structure Validation Report (i)

Apr 26, 2016 – 03:18 PM BST

PDB ID	:	1ICO
Title	:	SOLUTION STRUCTURE OF DESIGNED BETA-SHEET MINI-PROTEIN TH10BOX
Authors	:	Ottesen, J.J.; Imperiali, B.
Deposited on	:	2001-04-02

This is a Full wwPDB NMR Structure Validation 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:

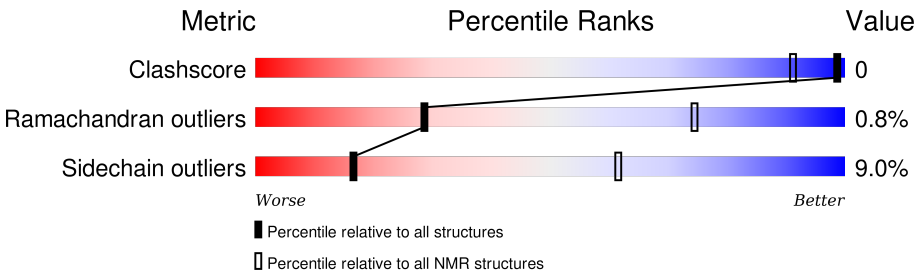
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 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	29	76% 14% • 7%

2 Ensemble composition and analysis (i)

This entry contains 42 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: *closest to the average*.

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:1-A:7, A:9-A:22, A:24-A:29 (27)	0.19	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 7 clusters and 11 single-model clusters were found.

Cluster number	Models
1	2, 3, 7, 8, 12, 14, 15, 18, 24, 25
2	1, 4, 11, 32, 33, 38
3	5, 6, 19, 36, 39
4	16, 22, 29, 42
5	27, 28
6	21, 34
7	30, 35
Single-model clusters	9; 10; 13; 17; 20; 23; 26; 31; 37; 40; 41

3 Entry composition (i)

There is only 1 type of molecule in this entry. The entry contains 456 atoms, of which 227 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called TH10BOX.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	29	456	147	227	37	43	2	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: TH10BOX

Chain A:

4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

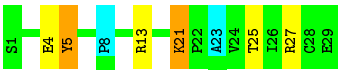
- Molecule 1: TH10BOX

Chain A:

4.2.2 Score per residue for model 2

- Molecule 1: TH10BOX

Chain A:



4.2.3 Score per residue for model 3

- Molecule 1: TH10BOX

Chain A:

4.2.4 Score per residue for model 4

- Molecule 1: TH10BOX

Chain A:

4.2.5 Score per residue for model 5

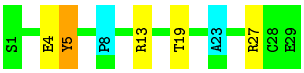
- Molecule 1: TH10BOX

Chain A:

4.2.6 Score per residue for model 6

- Molecule 1: TH10BOX

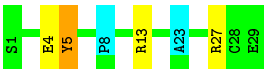
Chain A:



4.2.7 Score per residue for model 7

- Molecule 1: TH10BOX

Chain A:



4.2.8 Score per residue for model 8

- Molecule 1: TH10BOX

Chain A: 79% 10% • 7%

S1 E4 Y5 P8 R13 A23 R27 C28 E29

4.2.9 Score per residue for model 9

- Molecule 1: TH10BOX

Chain A: 72% 14% • • 7%

S1 E4 Y5 T6 I7 P8 R13 K21 P22 A23 R27 C28 E29

4.2.10 Score per residue for model 10

- Molecule 1: TH10BOX

Chain A: 72% 14% 7% 7%

S1 E4 Y5 P8 R13 T19 V20 K21 P22 A23 R27 C28 E29

4.2.11 Score per residue for model 11

- Molecule 1: TH10BOX

Chain A: 72% 14% 7% 7%

S1 E4 Y5 P8 R13 T19 V20 K21 P22 A23 R27 C28 E29

4.2.12 Score per residue for model 12

- Molecule 1: TH10BOX

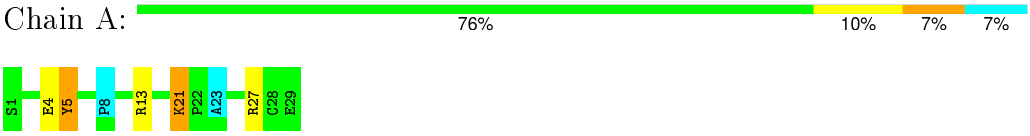
Chain A: 72% 17% • 7%

S1 E4 Y5 P8 R13 T19 A23 V24 T25 I26 R27 C28 E29

4.2.13 Score per residue for model 13

- Molecule 1: TH10BOX

Chain A:



4.2.14 Score per residue for model 14

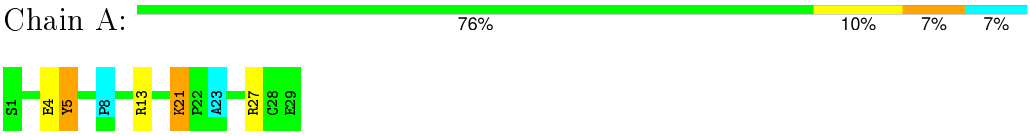
- Molecule 1: TH10BOX

Chain A:

4.2.15 Score per residue for model 15 (medoid)

- Molecule 1: TH10BOX

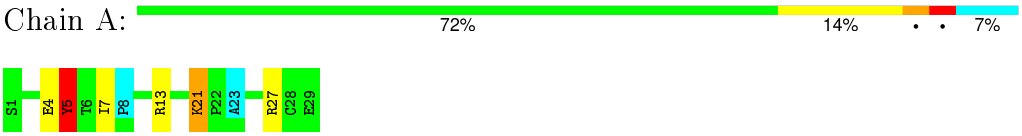
Chain A:



4.2.16 Score per residue for model 16

- Molecule 1: TH10BOX

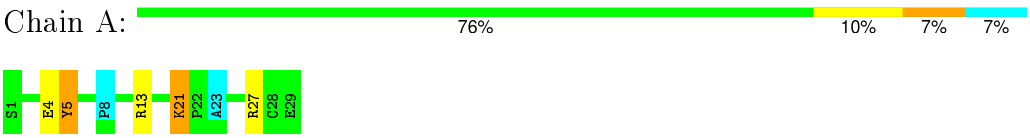
Chain A:



4.2.17 Score per residue for model 17

- Molecule 1: TH10BOX

Chain A:



4.2.18 Score per residue for model 18

- Molecule 1: TH10BOX

Chain A: 76% 10% 7% 7%

S1 Y5 P8 R13 T19 V20 K21 P22 A23 R27 C28 E29

4.2.19 Score per residue for model 19

- Molecule 1: TH10BOX

Chain A: 72% 14% 7% 7%

S1 E4 Y5 P8 R13 T19 V20 K21 P22 A23 R27 C28 E29

4.2.20 Score per residue for model 20

- Molecule 1: TH10BOX

Chain A: 79% 10% • 7%

S1 E4 Y5 P8 R13 A23 R27 C28 E29

4.2.21 Score per residue for model 21

- Molecule 1: TH10BOX

Chain A: 79% 10% • 7%

S1 E4 Y5 P8 R13 A23 R27 C28 E29

4.2.22 Score per residue for model 22

- Molecule 1: TH10BOX

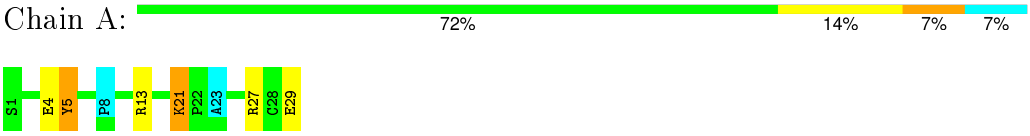
Chain A: 76% 7% 10% 7%

S1 E4 Y5 P8 R13 K21 P22 A23 R27 C28 E29

4.2.23 Score per residue for model 23

- Molecule 1: TH10BOX

Chain A:



4.2.24 Score per residue for model 24

- Molecule 1: TH10BOX

Chain A:

4.2.25 Score per residue for model 25

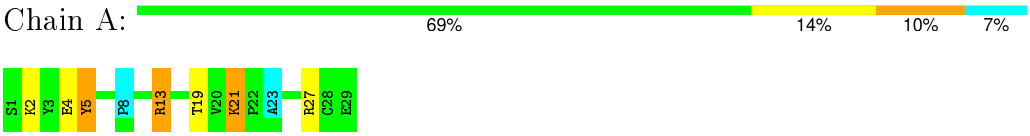
- Molecule 1: TH10BOX

Chain A:

4.2.26 Score per residue for model 26

- Molecule 1: TH10BOX

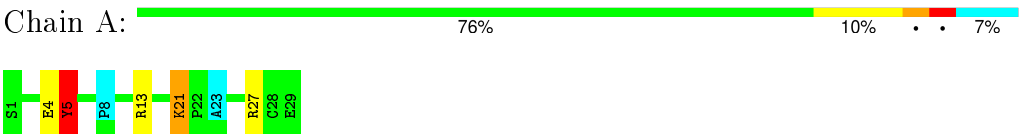
Chain A:



4.2.27 Score per residue for model 27

- Molecule 1: TH10BOX

Chain A:



4.2.28 Score per residue for model 28

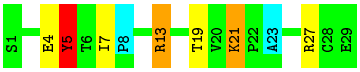
- Molecule 1: TH10BOX

Chain A:

4.2.29 Score per residue for model 29

- Molecule 1: TH10BOX

Chain A:



4.2.30 Score per residue for model 30

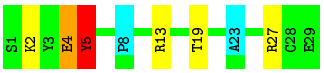
- Molecule 1: TH10BOX

Chain A:

4.2.31 Score per residue for model 31

- Molecule 1: TH10BOX

Chain A:



4.2.32 Score per residue for model 32

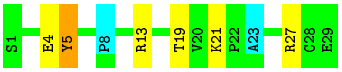
- Molecule 1: TH10BOX

Chain A:

4.2.33 Score per residue for model 33

- Molecule 1: TH10BOX

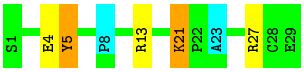
Chain A:



4.2.34 Score per residue for model 34

- Molecule 1: TH10BOX

Chain A:



4.2.35 Score per residue for model 35

- Molecule 1: TH10BOX

Chain A:

4.2.36 Score per residue for model 36

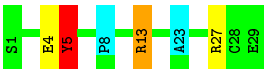
- Molecule 1: TH10BOX

Chain A:

4.2.37 Score per residue for model 37

- Molecule 1: TH10BOX

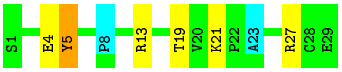
Chain A:



4.2.38 Score per residue for model 38

- Molecule 1: TH10BOX

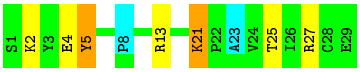
Chain A:



4.2.39 Score per residue for model 39

- Molecule 1: TH10BOX

Chain A:



4.2.40 Score per residue for model 40

- Molecule 1: TH10BOX

Chain A:

4.2.41 Score per residue for model 41

- Molecule 1: TH10BOX

Chain A:

4.2.42 Score per residue for model 42

- Molecule 1: TH10BOX

Chain A:

5 Refinement protocol and experimental data overview (i)

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

Of the 50 calculated structures, 42 were deposited, based on the following criterion: *structures with the least restraint violations, structures with the lowest energy*.

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

Software name	Classification	Version
DISCOVER	structure solution	95
DISCOVER	refinement	95

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: DAL, DPR

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	1.23±0.01	0±0/221 (0.0±0.0%)	1.95±0.03	7±1/295 (2.4±0.3%)
All	All	1.23	0/9282 (0.0%)	1.95	301/12390 (2.4%)

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	0.2±0.4
All	All	0	10

There are no bond-length outliers.

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(°)	Ideal(°)	Models	
								Worst	Total
1	A	13	ARG	NE-CZ-NH1	11.87	126.23	120.30	32	42
1	A	27	ARG	NE-CZ-NH1	11.84	126.22	120.30	24	42
1	A	5	TYR	CB-CG-CD2	-7.30	116.62	121.00	20	1
1	A	5	TYR	CA-CB-CG	7.23	127.14	113.40	23	42
1	A	10	TYR	CB-CG-CD2	-7.10	116.74	121.00	25	1
1	A	13	ARG	NE-CZ-NH2	-6.89	116.86	120.30	36	42
1	A	27	ARG	NE-CZ-NH2	-6.86	116.87	120.30	2	42
1	A	25	THR	CA-CB-CG2	6.46	121.45	112.40	39	5
1	A	10	TYR	CB-CG-CD1	6.43	124.86	121.00	25	1
1	A	4	GLU	CB-CA-C	6.38	123.15	110.40	14	41
1	A	21	LYS	CA-CB-CG	5.90	126.38	113.40	26	29
1	A	5	TYR	CB-CG-CD1	5.58	124.35	121.00	20	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)	Models	
								Worst	Total
1	A	10	TYR	CA-CB-CG	5.42	123.71	113.40	25	1
1	A	29	GLU	N-CA-CB	5.42	120.36	110.60	23	1
1	A	19	THR	CA-CB-CG2	5.36	119.90	112.40	40	9
1	A	27	ARG	N-CA-C	-5.04	97.40	111.00	23	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	5	TYR	Sidechain	8
1	A	12	PHE	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	217	215	215	0±0
All	All	9114	9030	9030	3

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 0.

All unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:5:TYR:CD2	1:A:7:ILE:HD11	0.41	2.50	29	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	25/29 (86%)	22±1 (90±4%)	2±1 (10±3%)	0±0 (1±2%)	29 74
All	All	1050/1218 (86%)	941 (90%)	101 (10%)	8 (1%)	29 74

All 1 unique Ramachandran outliers are listed below.

Mol	Chain	Res	Type	Models (Total)
1	A	21	LYS	8

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	25/25 (100%)	23±1 (91±4%)	2±1 (9±4%)	16 61
All	All	1050/1050 (100%)	955 (91%)	95 (9%)	16 61

All 10 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	5	TYR	42
1	A	21	LYS	21
1	A	19	THR	11
1	A	13	ARG	6
1	A	26	ILE	3
1	A	3	TYR	3
1	A	2	LYS	3
1	A	10	TYR	3
1	A	4	GLU	2
1	A	28	CYS	1

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

2 non-standard protein/DNA/RNA residues are 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
1	DAL	A	23	1	1,4,5	1.00±0.02	0±0 (0±0%)
1	DPR	A	8	1	5,7,8	0.57±0.02	0±0 (0±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
1	DAL	A	23	1	1,4,6	0.84±0.10	0±0 (0±0%)
1	DPR	A	8	1	7,8,10	1.39±0.07	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
1	DAL	A	23	1	-	0±0,0,2,4	0±0,0,0,0
1	DPR	A	8	1	-	0±0,0,9,11	0±0,1,1,1

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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