W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 06:08 PM BST

PDB ID	:	1RCS
Title	:	NMR STUDY OF TRP REPRESSOR-OPERATOR DNA COMPLEX
Authors	:	Zhao, D.; Zheng, Z.
Deposited on	:	1995-05-12

The following versions of software and data (see references (i)) were used in the production of this report:

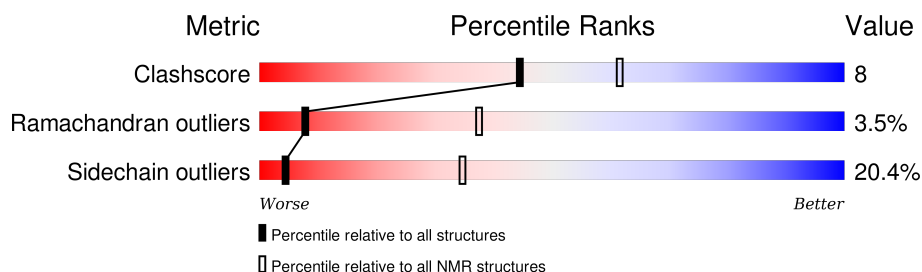|		
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
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	Percentile Ranks	Value
Clashscore		8
Ramachandran outliers		3.5%
Sidechain outliers		20.4%

Worse   *Better*

▮ Percentile relative to all structures
▯ Percentile relative to all NMR structures

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	E	20	70% 30%
1	F	20	75% 20% 5%
2	A	105	50% 29% 5% 15% •
2	B	105	49% 31% • 15% •

2 Ensemble composition and analysis (i)

This entry contains 15 models. Model 1 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:18-A:105, B:518-B:605 (176)	0.40	1

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 and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4704 atoms, of which 2188 are hydrogens and 0 are deuteriums.

- Molecule 1 is a DNA chain called DNA (5'-D(*CP*GP*TP*AP*CP*TP*AP*GP*TP*TP* AP*AP*CP*TP*AP*GP*TP*AP*CP*G)-3').

Mol	Chain	Residues	Atoms						Trace
1	E	20	Total	C	H	N	O	P	0
			635	196	228	74	118	19	
1	F	20	Total	C	H	N	O	P	0
			635	196	228	74	118	19	

- Molecule 2 is a protein called TRP REPRESSOR.

Mol	Chain	Residues	Atoms						Trace
2	A	104	Total	C	H	N	O	S	0
			1690	524	854	152	157	3	
2	B	104	Total	C	H	N	O	S	0
			1690	524	854	152	157	3	

- Molecule 3 is TRYPTOPHAN (three-letter code: TRP) (formula: $C_{11}H_{12}N_2O_2$).
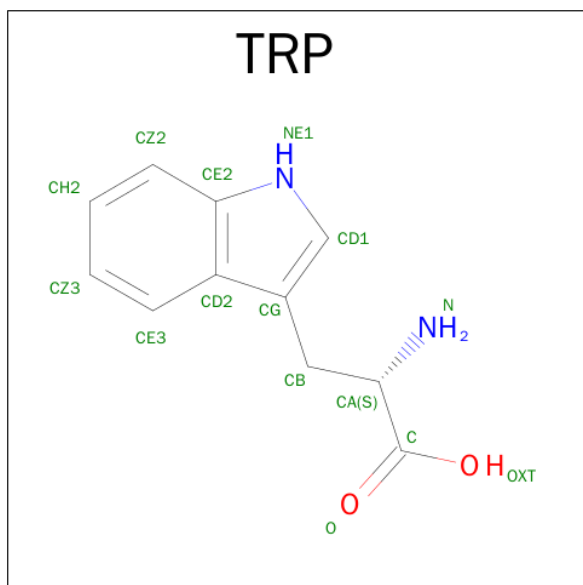
| Mol	Chain	Residues	Atoms				
3	A	1	Total	C	H	N	O
			27	11	12	2	2
3	B	1	Total	C	H	N	O
			27	11	12	2	2

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: DNA (5'-D(*CP*GP*TP*AP*CP*TP*AP*GP*TP*TP*AP*AP*CP*TP*AP*GP*TP*AP*CP*G)-3')
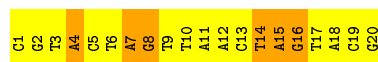
Chain E:

• Molecule 1: DNA (5'-D(*CP*GP*TP*AP*CP*TP*AP*GP*TP*TP*AP*AP*CP*TP*AP*GP*TP*AP*CP*G)-3')
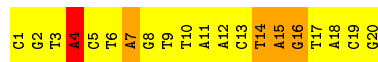
Chain F:

• Molecule 2: TRP REPRESSOR

Chain A:
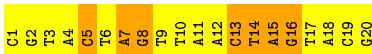

• Molecule 2: TRP REPRESSOR

Chain B:

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

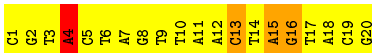The representative model is number 1. Colouring as in section 4.1 above.

• Molecule 1: DNA (5'-D(*CP*GP*TP*AP*CP*TP*AP*GP*TP*TP*AP*AP*CP*TP*AP*GP *TP*AP*CP*G)-3')

Chain E:

• Molecule 1: DNA (5'-D(*CP*GP*TP*AP*CP*TP*AP*GP*TP*TP*AP*AP*CP*TP*AP*GP *TP*AP*CP*G)-3')

Chain F:

• Molecule 2: TRP REPRESSOR

Chain A:

• Molecule 2: TRP REPRESSOR

Chain B:

5 Refinement protocol and experimental data overview (i)

Of the ? calculated structures, 15 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 ⓘ

6.1 Standard geometry ⓘ

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	E	4.33±0.03	115±2/456 (25.3±0.5%)	4.90±0.03	109±3/702 (15.6±0.5%)
1	F	4.33±0.03	115±1/456 (25.2±0.3%)	4.89±0.03	107±3/702 (15.2±0.5%)
2	A	1.80±0.01	3±1/720 (0.4±0.1%)	1.07±0.02	5±1/975 (0.5±0.1%)
2	B	1.80±0.02	2±1/720 (0.3±0.1%)	1.08±0.02	5±2/975 (0.5±0.2%)
All	All	3.04	3537/35280 (10.0%)	3.27	3388/50310 (6.7%)

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	E	1.3±0.5	1.9±0.6
1	F	1.1±0.2	2.3±0.6
2	A	0.0±0.0	2.2±1.3
2	B	0.0±0.0	1.9±0.9
All	All	36	126

5 of 273 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	F	7	DA	C4'-C3'	-11.91	1.40	1.52	12	15
1	F	7	DA	C3'-C2'	-11.36	1.38	1.52	14	15
1	F	13	DC	C3'-O3'	11.26	1.58	1.44	6	15
1	E	13	DC	C3'-O3'	11.06	1.58	1.44	5	15
1	E	9	DT	C3'-O3'	11.05	1.58	1.44	10	15

5 of 308 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	F	13	DC	O4'-C1'-N1	23.09	124.17	108.00	15	15
1	E	13	DC	O4'-C1'-N1	22.40	123.68	108.00	2	15
1	E	11	DA	O4'-C1'-N9	19.82	121.88	108.00	13	15
1	F	11	DA	O4'-C1'-N9	19.33	121.53	108.00	9	15
1	F	4	DA	O4'-C1'-N9	18.49	120.94	108.00	15	15

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

Mol	Chain	Res	Type	Atoms	Models (Total)
1	E	7	DA	C1'	15
1	F	7	DA	C1'	15
1	E	13	DC	C1'	5
1	F	13	DC	C1'	1

5 of 23 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	F	16	DG	Sidechain	15
1	E	16	DG	Sidechain	15
2	A	69	ARG	Sidechain	11
2	B	569	ARG	Sidechain	10
1	F	4	DA	Sidechain	9

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	E	407	228	228	6$\pm$1
1	F	407	228	228	5$\pm$1
2	A	710	734	734	19$\pm$2
2	B	710	734	734	22$\pm$3
3	B	15	12	9	0$\pm$1
3	A	15	12	9	0$\pm$1
All	All	33960	29220	29130	522

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 8.

5 of 117 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models Worst	Total
2:A:19:TRP:CE3	2:B:551:LEU:HD22	0.82	2.09	6	7
1:E:15:DA:H62	2:A:80:ALA:HB2	0.80	1.37	5	15
1:F:15:DA:H62	2:B:580:ALA:HB2	0.79	1.38	12	15
1:E:4:DA:H62	2:B:579:ILE:HG21	0.78	1.38	15	2
1:F:14:DT:H72	2:B:580:ALA:HB3	0.74	1.60	2	13

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	
2	A	88/105 (84%)	73±3 (83±3%)	12±3 (14±3%)	3±1 (3±1%)	8	39
2	B	88/105 (84%)	73±1 (83±2%)	12±2 (13±2%)	3±1 (4±1%)	7	36
All	All	2640/3150 (84%)	2189 (83%)	359 (14%)	92 (3%)	8	37

5 of 18 unique Ramachandran outliers are listed below. They are sorted by the frequency of
occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	B	533	ASP	15
2	A	33	ASP	15
2	A	44	THR	14
2	B	544	THR	10
2	A	78	GLY	9

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	
2	A	77/91 (85%)	61±2 (79±3%)	16±2 (21±3%)	4	33
2	B	77/91 (85%)	62±2 (80±3%)	15±2 (20±3%)	5	36
All	All	2310/2730 (85%)	1838 (80%)	472 (20%)	4	35

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)
2	A	72	LYS	15
2	B	557	ILE	15
2	B	597	ARG	15
2	A	58	VAL	15
2	A	57	ILE	15

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)

2 ligands 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
3	TRP	A	898	-	11,16,16	0.75±0.03	0±0 (0±0%)

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
3	TRP	B	998	-	11,16,16	0.67±0.04	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
3	TRP	A	898	-	9,22,22	1.75±0.24	1±0 (9±3%)
3	TRP	B	998	-	9,22,22	1.66±0.22	0±0 (5±5%)

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
3	TRP	A	898	-	-	0±0,3,8,8	0±0,2,2,2
3	TRP	B	998	-	-	0±0,3,8,8	0±0,2,2,2

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
3	B	998	TRP	CB-CG-CD1	5.85	120.92	128.01	13	7
3	A	898	TRP	CB-CG-CD1	5.72	121.08	128.01	8	13

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

6.7 Other polymers ⓘ

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