Full wwPDB NMR Structure Validation Report (i)

Feb 26, 2022 – 03:36 PM EST

PDB ID	:	2AD9
Title	:	Solution structure of Polypyrimidine Tract Binding protein RBD1 complexed with CUCUCU RNA
Authors	:	Oberstrass, F.C.; Auweter, S.D.; Erat, M.; Hargous, Y.; Henning, A.; Wenter, P.; Reymond, L.; Pitsch, S.; Black, D.L.; Allain, F.H.T.
Deposited on	:	2005-07-20

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
https://www.wwpdb.org/validation/2017/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:

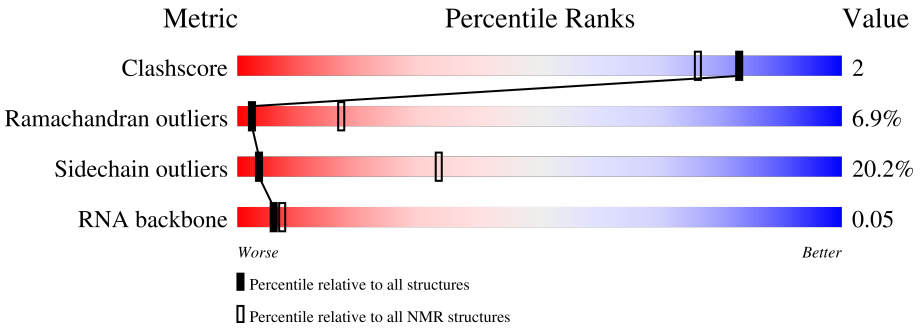
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.27
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.27

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	158937	12864
Ramachandran outliers	154571	11451
Sidechain outliers	154315	11428
RNA backbone	4643	676

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	B	6	33% 67%
2	A	119	49% 15% • 15% 18%

2 Ensemble composition and analysis

This entry contains 20 models. 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:58-A:137 (80)	0.41	17

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1717 atoms, of which 839 are hydrogens and 0 are deuteriums.

- Molecule 1 is a RNA chain called 5'-R(*CP*UP*CP*UP*CP*U)-3'.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	B	6	182	54	65	15	43	5	0

- Molecule 2 is a protein called Polypyrimidine tract-binding protein 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	A	98	1535	478	774	134	146	3	0

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	28	MET	-	expression tag	UNP P26599
A	29	GLY	-	expression tag	UNP P26599
A	30	SER	-	expression tag	UNP P26599
A	31	SER	-	expression tag	UNP P26599
A	32	HIS	-	expression tag	UNP P26599
A	33	HIS	-	expression tag	UNP P26599
A	34	HIS	-	expression tag	UNP P26599
A	35	HIS	-	expression tag	UNP P26599
A	36	HIS	-	expression tag	UNP P26599
A	37	HIS	-	expression tag	UNP P26599
A	38	SER	-	expression tag	UNP P26599
A	39	SER	-	expression tag	UNP P26599
A	40	GLY	-	expression tag	UNP P26599
A	41	LEU	-	expression tag	UNP P26599
A	42	VAL	-	expression tag	UNP P26599
A	43	PRO	-	expression tag	UNP P26599
A	44	ARG	-	expression tag	UNP P26599
A	45	GLY	-	expression tag	UNP P26599
A	46	SER	-	expression tag	UNP P26599
A	47	HIS	-	expression tag	UNP P26599
A	48	MET	-	expression tag	UNP P26599

4 Residue-property plots

4.1 Average score per residue in the NMR ensemble

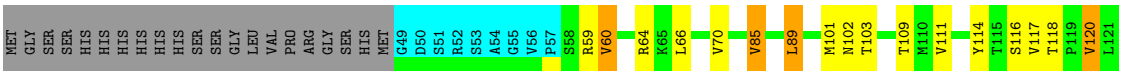
These plots are provided for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for 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.

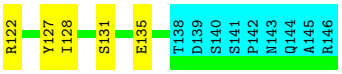
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:

• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:





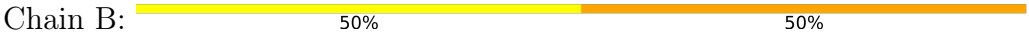
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

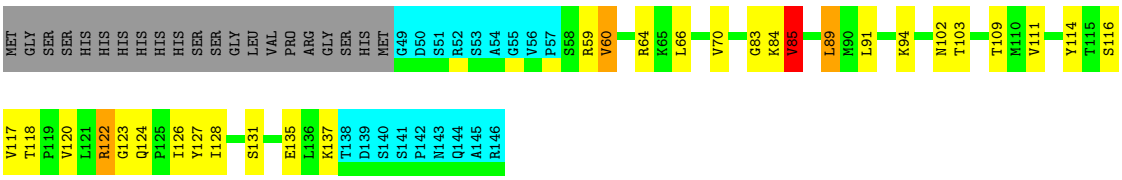
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:



• Molecule 2: Polypyrimidine tract-binding protein 1

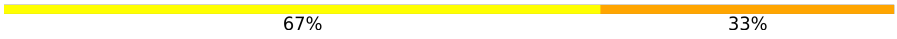
Chain A:



4.2.2 Score per residue for model 2

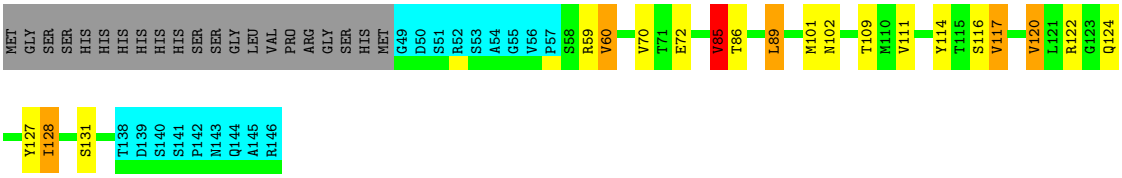
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:



• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:



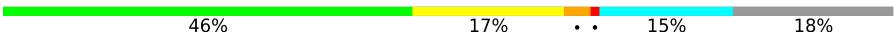
4.2.3 Score per residue for model 3

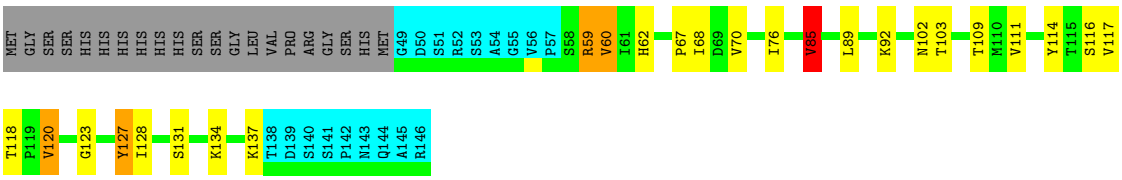
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:

• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:





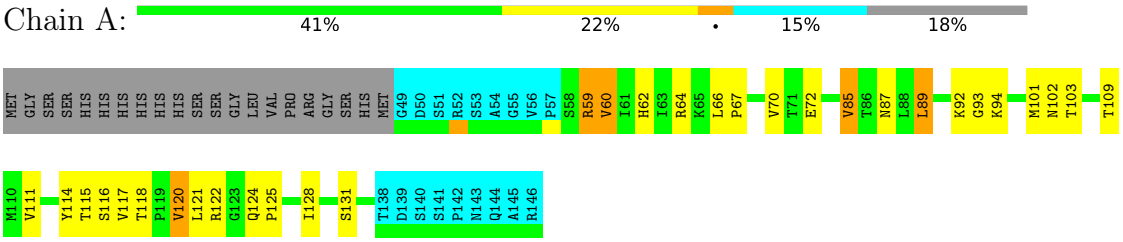
4.2.4 Score per residue for model 4

• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:

• Molecule 2: Polypyrimidine tract-binding protein 1

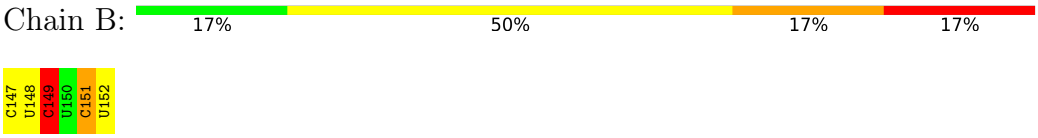
Chain A:



4.2.5 Score per residue for model 5

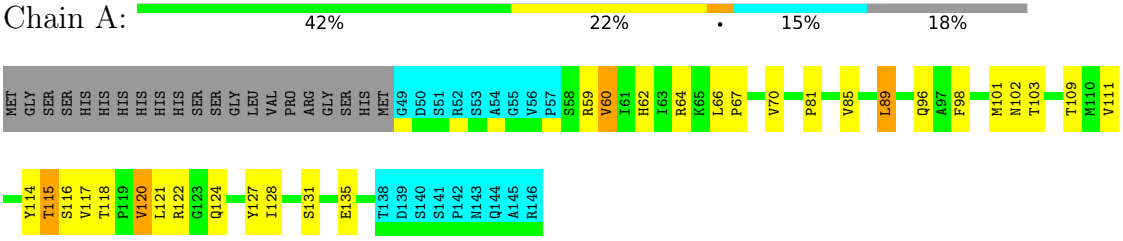
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:



• Molecule 2: Polypyrimidine tract-binding protein 1

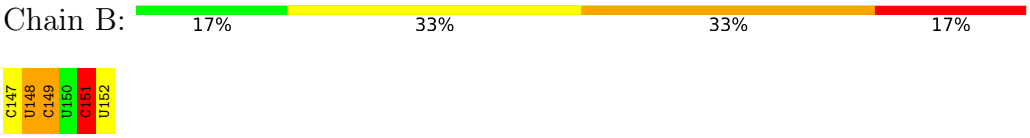
Chain A:



4.2.6 Score per residue for model 6

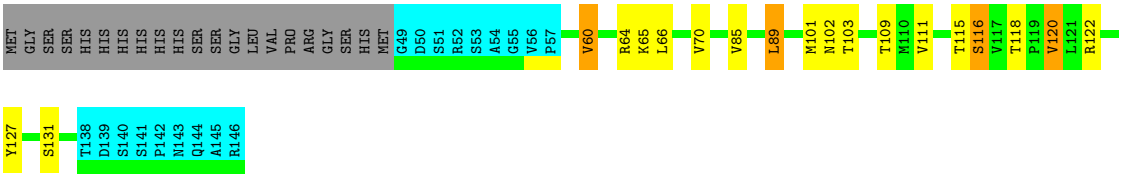
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:



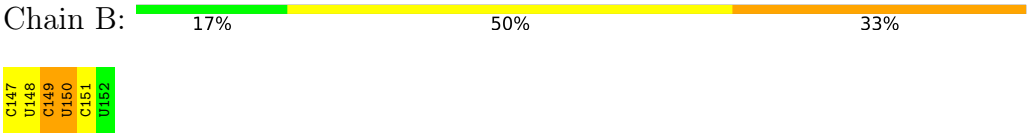
• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:

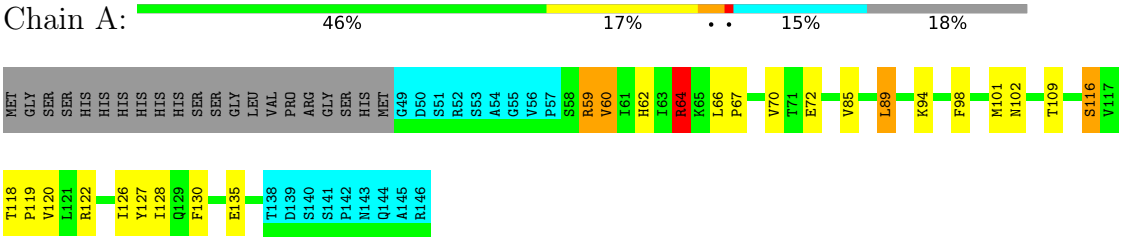


4.2.7 Score per residue for model 7

• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

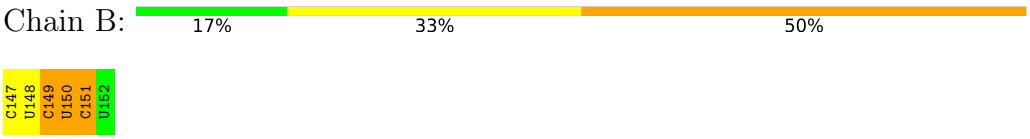


• Molecule 2: Polypyrimidine tract-binding protein 1

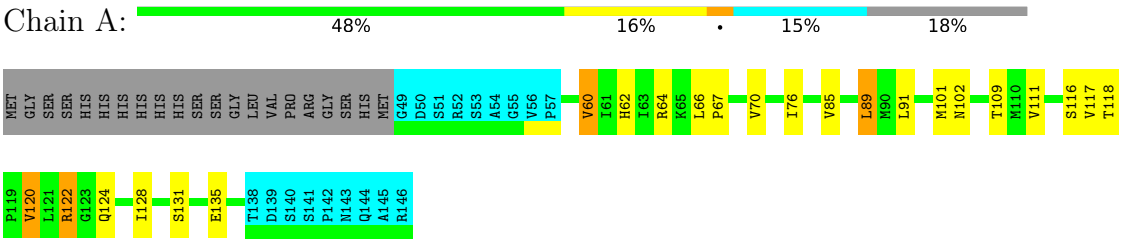


4.2.8 Score per residue for model 8

• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'



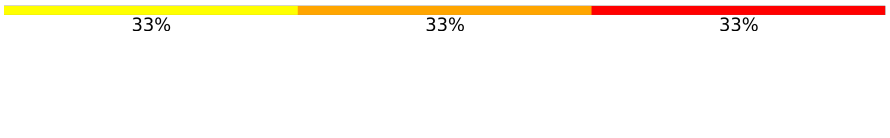
• Molecule 2: Polypyrimidine tract-binding protein 1



4.2.9 Score per residue for model 9

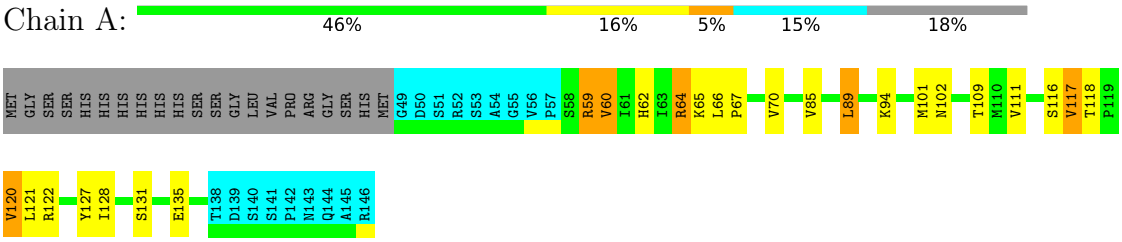
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:



• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:



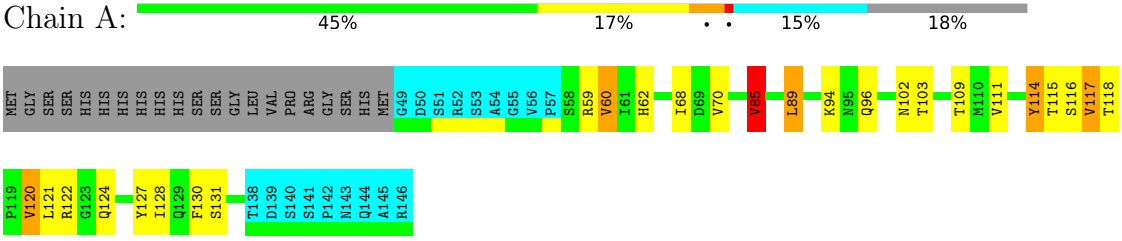
4.2.10 Score per residue for model 10

• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:

• Molecule 2: Polypyrimidine tract-binding protein 1

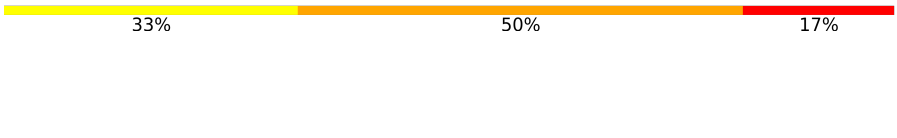
Chain A:



4.2.11 Score per residue for model 11

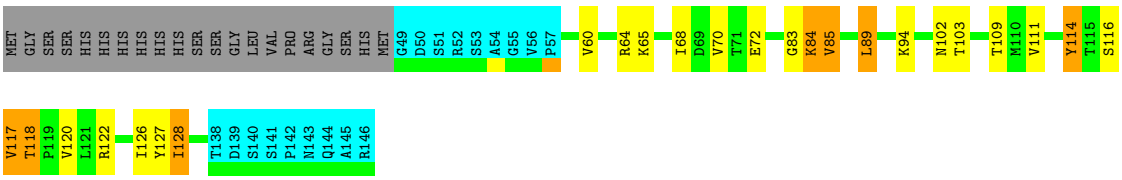
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:



• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:



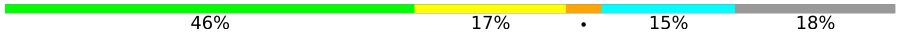
4.2.12 Score per residue for model 12

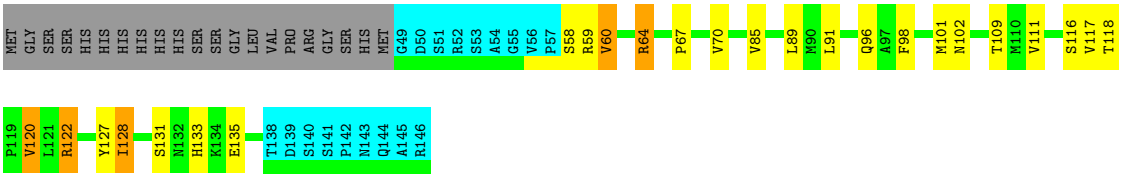
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:

• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:





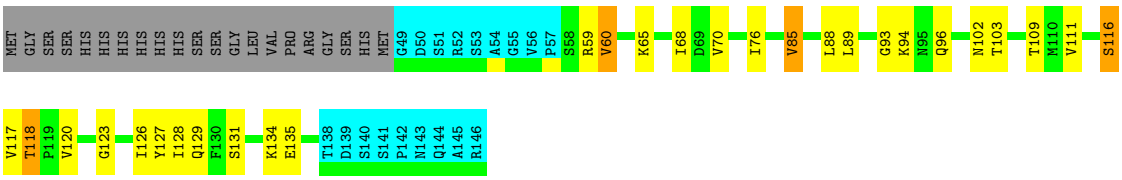
4.2.13 Score per residue for model 13

• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:

• Molecule 2: Polypyrimidine tract-binding protein 1

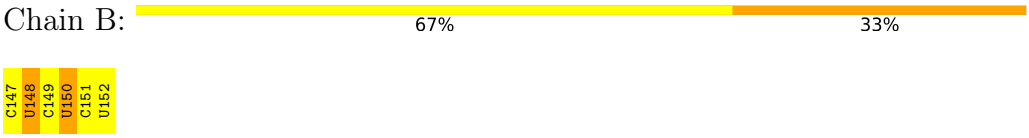
Chain A:



4.2.14 Score per residue for model 14

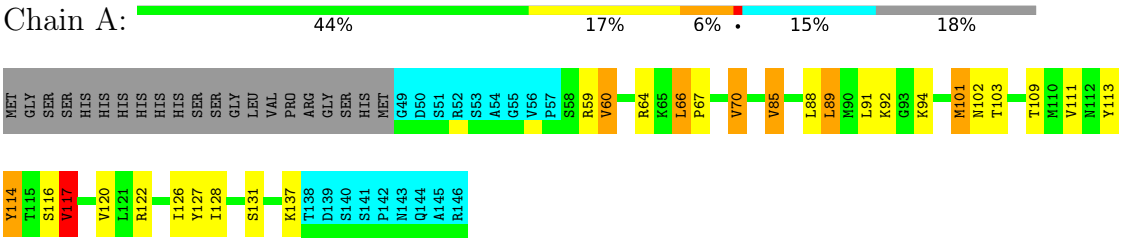
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:



• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:



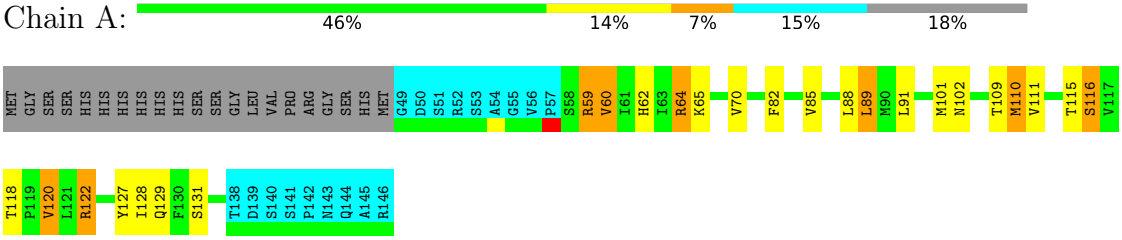
4.2.15 Score per residue for model 15

• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:

• Molecule 2: Polypyrimidine tract-binding protein 1

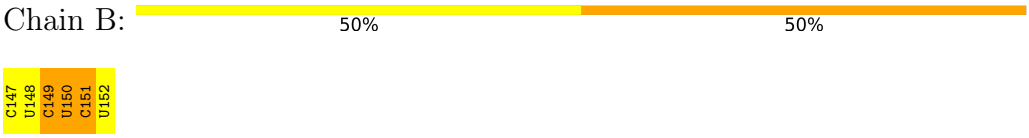
Chain A:



4.2.16 Score per residue for model 16

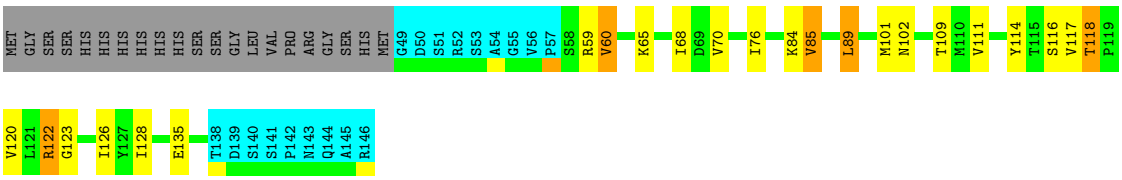
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:



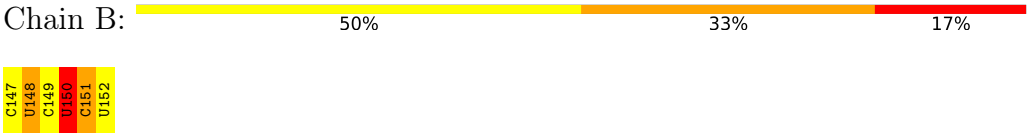
• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:

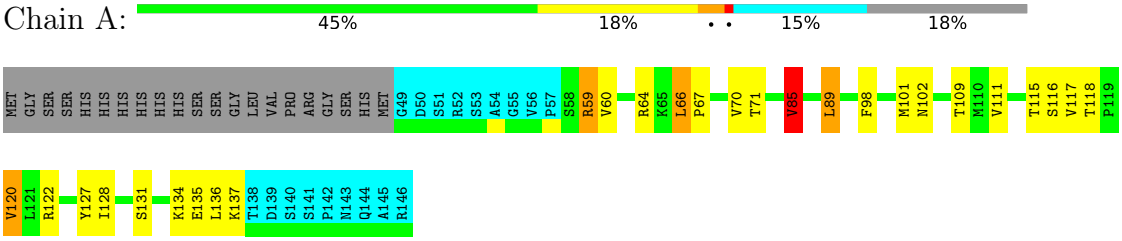


4.2.17 Score per residue for model 17 (medoid)

• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'



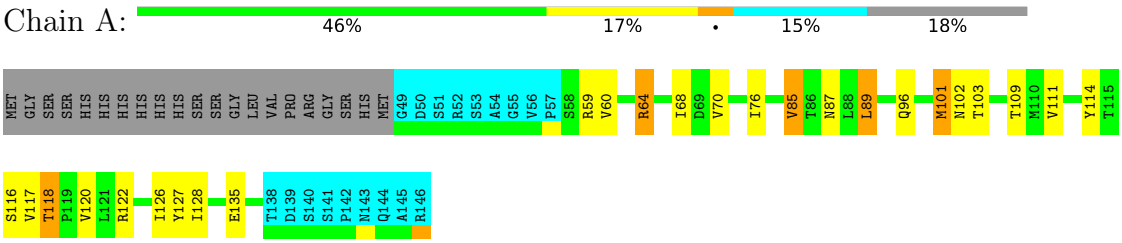
• Molecule 2: Polypyrimidine tract-binding protein 1



4.2.18 Score per residue for model 18

• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

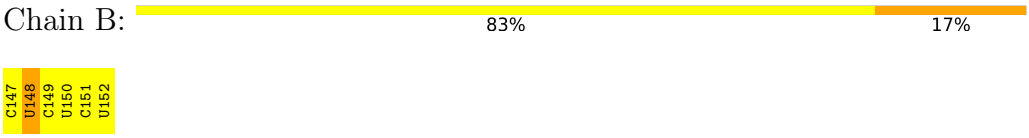
• Molecule 2: Polypyrimidine tract-binding protein 1



4.2.19 Score per residue for model 19

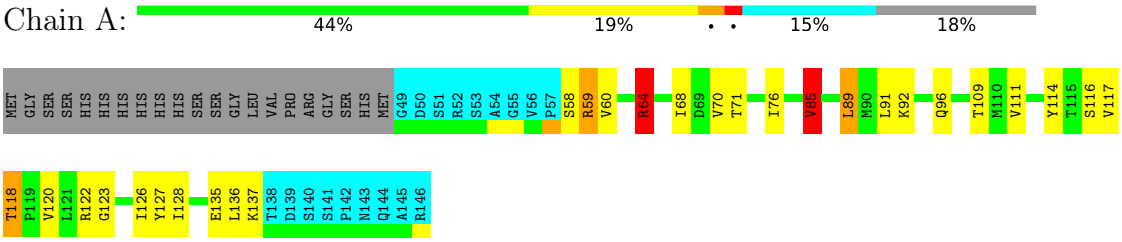
• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:



• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:



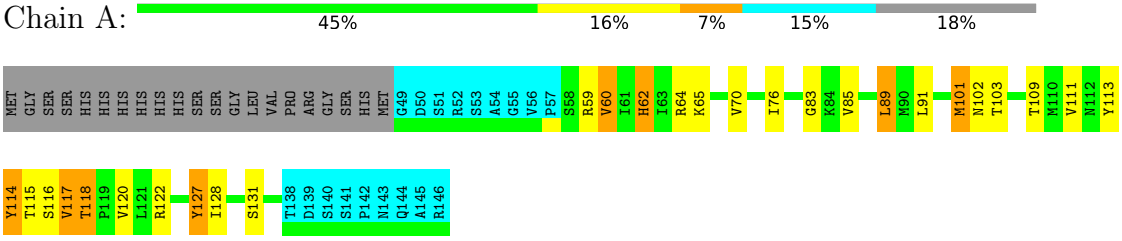
4.2.20 Score per residue for model 20

• Molecule 1: 5'-R(*CP*UP*CP*UP*CP*U)-3'

Chain B:

• Molecule 2: Polypyrimidine tract-binding protein 1

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 40 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
DYANA	structure solution	3.02
Amber	refinement	7

No chemical shift data was provided.

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	B	1.34±0.03	0±0/128 (0.0± 0.0%)	2.15±0.06	8±1/196 (4.0± 0.7%)
2	A	0.67±0.01	0±0/647 (0.0± 0.0%)	1.27±0.03	4±1/876 (0.5± 0.1%)
All	All	0.82	0/15500 (0.0%)	1.47	242/21440 (1.1%)

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	B	0.0±0.0	0.8±0.7
2	A	0.0±0.0	3.0±0.8
All	All	0	75

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(^{o})	Ideal(^{o})	Models	
								Worst	Total
2	A	85	VAL	CA-CB-CG1	10.51	126.66	110.90	1	7
2	A	127	TYR	CB-CG-CD2	-8.61	115.83	121.00	20	15
1	B	150	U	O4'-C1'-N1	8.35	114.88	108.20	9	17
1	B	149	C	N3-C2-O2	-7.85	116.41	121.90	1	20
2	A	59	ARG	NE-CZ-NH1	7.38	123.99	120.30	19	14
2	A	60	VAL	CA-CB-CG1	7.36	121.95	110.90	11	19
1	B	151	C	N3-C2-O2	-7.34	116.76	121.90	6	20
1	B	152	U	O4'-C1'-N1	7.21	113.97	108.20	4	8
1	B	147	C	N3-C2-O2	-6.76	117.17	121.90	1	20
2	A	64	ARG	NE-CZ-NH1	6.76	123.68	120.30	19	9
2	A	85	VAL	CG1-CB-CG2	6.64	121.53	110.90	19	13
1	B	151	C	O4'-C1'-N1	6.45	113.36	108.20	13	6
1	B	149	C	N1-C2-O2	6.35	122.71	118.90	16	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	B	148	U	O4'-C1'-N1	6.34	113.27	108.20	9	3
1	B	149	C	P-O3'-C3'	6.19	127.13	119.70	1	1
1	B	152	U	N3-C2-O2	-6.04	117.97	122.20	11	7
2	A	122	ARG	NE-CZ-NH1	5.97	123.28	120.30	15	2
2	A	121	LEU	CB-CA-C	5.67	120.97	110.20	4	1
1	B	151	C	N1-C2-O2	5.66	122.30	118.90	6	15
1	B	147	C	N1-C2-O2	5.50	122.20	118.90	11	13
1	B	148	U	C5-C6-N1	-5.50	119.95	122.70	11	1
2	A	117	VAL	CA-CB-CG1	5.38	118.96	110.90	2	1
1	B	150	U	C3'-C2'-C1'	5.28	105.72	101.50	9	4
1	B	149	C	O4'-C1'-N1	5.18	112.34	108.20	4	1
2	A	110	MET	CA-CB-CG	-5.13	104.57	113.30	15	1
1	B	148	U	N3-C2-O2	-5.11	118.62	122.20	13	1
2	A	127	TYR	CB-CG-CD1	5.07	124.04	121.00	20	2
1	B	151	C	N3-C4-N4	-5.01	114.49	118.00	2	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)
2	A	117	VAL	Peptide	14
2	A	114	TYR	Sidechain	12
1	B	148	U	Sidechain	12
2	A	120	VAL	Peptide	11
2	A	59	ARG	Sidechain	6
2	A	116	SER	Peptide	4
2	A	128	ILE	Peptide	3
2	A	84	LYS	Peptide	2
2	A	64	ARG	Sidechain	2
2	A	66	LEU	Peptide	2
1	B	149	C	Sidechain	2
2	A	122	ARG	Sidechain,Peptide	2
1	B	151	C	Sidechain	1
2	A	121	LEU	Peptide	1
2	A	133	HIS	Peptide	1

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	B	117	65	65	0±0
2	A	635	660	660	4±2
All	All	15040	14500	14500	72

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 2.

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

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:A:60:VAL:HG13	2:A:131:SER:HB3	0.58	1.75	13	15
2:A:85:VAL:HG22	2:A:101:MET:CE	0.57	2.29	6	3
2:A:85:VAL:HG22	2:A:101:MET:HE2	0.52	1.81	6	5
2:A:60:VAL:HG13	2:A:131:SER:CB	0.50	2.36	13	6
2:A:85:VAL:HG13	2:A:101:MET:HB2	0.50	1.83	4	1
2:A:60:VAL:HG12	2:A:135:GLU:N	0.48	2.23	13	1
2:A:114:TYR:HA	2:A:117:VAL:HG12	0.48	1.86	20	4
2:A:66:LEU:HD22	2:A:70:VAL:HG11	0.48	1.86	14	1
2:A:89:LEU:C	2:A:89:LEU:HD13	0.47	2.30	17	5
1:B:149:C:C6	2:A:62:HIS:CD2	0.47	3.03	7	9
2:A:89:LEU:HD23	2:A:89:LEU:C	0.46	2.30	16	11
2:A:66:LEU:CD2	2:A:121:LEU:HD13	0.45	2.42	5	1
2:A:82:PHE:CD2	2:A:110:MET:HG3	0.43	2.48	15	1
2:A:60:VAL:HG21	2:A:98:PHE:CD2	0.43	2.48	5	4
2:A:98:PHE:CD2	2:A:136:LEU:HD13	0.43	2.49	17	1
2:A:85:VAL:CG2	2:A:101:MET:CE	0.41	2.98	16	1
1:B:150:U:C4	2:A:136:LEU:HD12	0.41	2.51	17	1
2:A:89:LEU:CD1	2:A:91:LEU:HD22	0.41	2.46	14	1
2:A:89:LEU:C	2:A:89:LEU:HD23	0.40	2.36	4	1

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	80/119 (67%)	65±3 (81±3%)	10±2 (12±2%)	6±2 (7±3%)	2 17
All	All	1600/2380 (67%)	1296 (81%)	194 (12%)	110 (7%)	2 17

All 19 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
2	A	120	VAL	20
2	A	118	THR	18
2	A	70	VAL	12
2	A	135	GLU	10
2	A	67	PRO	9
2	A	85	VAL	8
2	A	68	ILE	7
2	A	123	GLY	5
2	A	122	ARG	4
2	A	83	GLY	3
2	A	127	TYR	3
2	A	93	GLY	2
2	A	58	SER	2
2	A	134	LYS	2
2	A	92	LYS	1
2	A	125	PRO	1
2	A	81	PRO	1
2	A	59	ARG	1
2	A	136	LEU	1

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	72/104 (69%)	57±2 (80±3%)	15±2 (20±3%)	3 33
All	All	1440/2080 (69%)	1149 (80%)	291 (20%)	3 33

All 41 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	89	LEU	20
2	A	109	THR	20
2	A	116	SER	20
2	A	102	ASN	19
2	A	111	VAL	19
2	A	128	ILE	19
2	A	122	ARG	15
2	A	64	ARG	12
2	A	103	THR	11
2	A	101	MET	11
2	A	70	VAL	8
2	A	94	LYS	8
2	A	126	ILE	8
2	A	66	LEU	7
2	A	85	VAL	7
2	A	76	ILE	7
2	A	115	THR	7
2	A	65	LYS	7
2	A	91	LEU	6
2	A	124	GLN	6
2	A	96	GLN	6
2	A	118	THR	6
2	A	137	LYS	5
2	A	117	VAL	5
2	A	72	GLU	4
2	A	59	ARG	3
2	A	92	LYS	3
2	A	88	LEU	3
2	A	87	ASN	2
2	A	130	PHE	2
2	A	84	LYS	2
2	A	129	GLN	2
2	A	113	TYR	2
2	A	71	THR	2
2	A	86	THR	1
2	A	134	LYS	1
2	A	119	PRO	1
2	A	121	LEU	1
2	A	60	VAL	1
2	A	62	HIS	1
2	A	127	TYR	1

6.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers	Suiteness
1	B	5/6 (83%)	4±1 (71±21%)	0±1 (7±13%)	0.05±0.07
All	All	100/120 (83%)	71 (71%)	7 (7%)	0.05

The overall RNA backbone suiteness is 0.05.

All unique RNA backbone outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	150	U	18
1	B	148	U	15
1	B	152	U	14
1	B	151	C	13
1	B	149	C	11

All unique RNA pucker outliers are listed below:

Mol	Chain	Res	Type	Models (Total)
1	B	150	U	3
1	B	151	C	2
1	B	149	C	1
1	B	148	U	1

6.4 Non-standard residues in protein, DNA, RNA chains

There are no non-standard protein/DNA/RNA residues in this entry.

6.5 Carbohydrates

There are no monosaccharides in this entry.

6.6 Ligand geometry

There are no ligands in this entry.

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

No chemical shift data were provided