wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 01:43 PM BST

PDB ID	:	1BBX
Title	:	NON-SPECIFIC PROTEIN-DNA INTERACTIONS IN THE SSO7D-DNA COMPLEX, NMR, 1 STRUCTURE
Authors	:	Agback, P.; Baumann, H.; Knapp, S.; Ladenstein, R.; Hard, T.
Deposited on	:	1998-04-24

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:

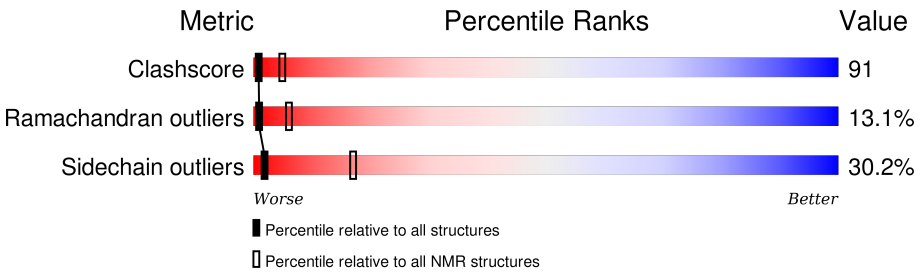
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 is 13%.

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	12	33% 67%
1	B	12	33% 67%
2	C	63	14% 60% 22% .
2	D	63	14% 60% 22% .

2 Ensemble composition and analysis (i)

This entry contains 1 models. Identification of well-defined residues and clustering analysis are not possible.

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2834 atoms, of which 1344 are hydrogens and 0 are deuteriums.

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

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
1	A	12	379	116	136	46	70	11	0
1	B	12	379	116	136	46	70	11	0

- Molecule 2 is a protein called DNA-BINDING PROTEIN 7D.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	C	63	1038	320	536	87	93	2	0
2	D	63	1038	320	536	87	93	2	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	13	GLN	GLU	CONFLICT	UNP P39476
D	13	GLN	GLU	CONFLICT	UNP P39476

4 Residue-property plots (i)

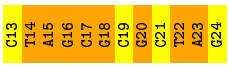
These plots are provided for all protein, RNA and DNA chains in the entry. The first graphic is the same as shown in the summary in section 1 of this report. The second graphic shows the sequence where residues are colour-coded according to the number of geometric quality criteria for which they contain 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*TP*AP*GP*CP*GP*CP*GP*CP*TP*AP*G)-3')

Chain A:

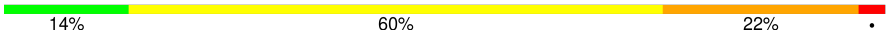
- Molecule 1: DNA (5'-D(*CP*TP*AP*GP*CP*GP*CP*GP*CP*TP*AP*G)-3')

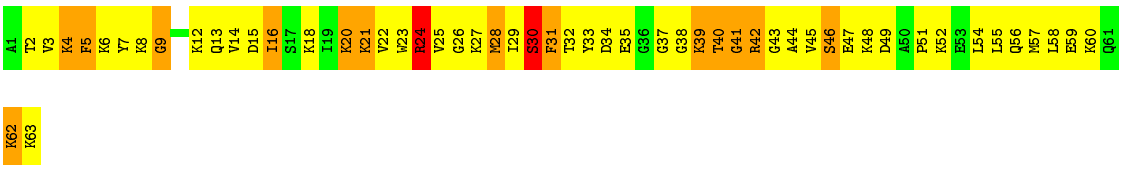
Chain B:



- Molecule 2: DNA-BINDING PROTEIN 7D

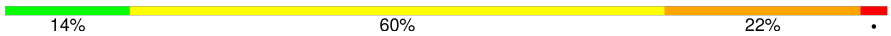
Chain C:

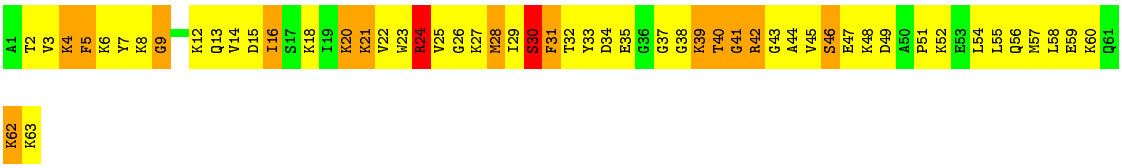




- Molecule 2: DNA-BINDING PROTEIN 7D

Chain D:





5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *SIMULATED ANNEALING*.

Of the 50 calculated structures, 1 were deposited, based on the following criterion: *LOWEST RESTRAINT ENERGY*.

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

Software name	Classification	Version
X-PLOR	refinement	3.1
VNMR	structure solution	
ANSIG	structure solution	
XPLOR	structure solution	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	BMRB entry 5909, BMRB entry 5910
Number of chemical shift lists	2
Total number of shifts	516
Number of shifts mapped to atoms	512
Number of unparsed shifts	0
Number of shifts with mapping errors	4
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	13%

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	1.33	1/272 (0.4%)	2.18	19/418 (4.5%)
1	B	1.33	1/272 (0.4%)	2.19	20/418 (4.8%)
2	C	1.04	0/508 (0.0%)	1.31	0/669 (0.0%)
2	D	1.04	0/508 (0.0%)	1.31	0/669 (0.0%)
All	All	1.15	2/1560 (0.1%)	1.70	39/2174 (1.8%)

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
2	C	0	2
2	D	0	2
All	All	0	4

All bond outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	6	DG	N9-C8	-5.11	1.34	1.37
1	B	18	DG	N9-C8	-5.08	1.34	1.37

5 of 39 angle outliers are listed below. They are sorted according to the Z-score.

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	24	DG	N7-C8-N9	9.37	117.79	113.10
1	A	6	DG	N7-C8-N9	9.32	117.76	113.10
1	A	12	DG	N7-C8-N9	9.30	117.75	113.10
1	B	18	DG	N7-C8-N9	9.28	117.74	113.10
1	A	4	DG	N7-C8-N9	8.95	117.58	113.10

There are no chirality outliers.

All planar outliers are listed below.

Mol	Chain	Res	Type	Group
2	C	42	ARG	Sidechain
2	D	24	ARG	Sidechain
2	C	24	ARG	Sidechain
2	D	42	ARG	Sidechain

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	243	136	136	36
1	B	243	136	136	36
2	C	502	536	536	104
2	D	502	536	536	108
All	All	1490	1344	1344	259

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 91.

5 of 259 clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)
2:D:14:VAL:HG11	2:D:33:TYR:CB	0.97	1.89
2:C:14:VAL:HG11	2:C:33:TYR:CB	0.97	1.89
1:A:10:DT:H5"	1:A:11:DA:OP2	0.97	1.60
1:B:22:DT:H5"	1:B:23:DA:OP2	0.96	1.60
2:D:28:MET:O	2:D:29:ILE:HD13	0.95	1.62

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	C	61/63 (97%)	36 (59%)	17 (28%)	8 (13%)	1 6
2	D	61/63 (97%)	36 (59%)	17 (28%)	8 (13%)	1 6
All	All	122/126 (97%)	72 (59%)	34 (28%)	16 (13%)	1 6

5 of 16 Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
2	D	30	SER
2	C	30	SER
2	C	20	LYS
2	D	20	LYS
2	D	38	GLY

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	C	53/53 (100%)	37 (70%)	16 (30%)	2 17
2	D	53/53 (100%)	37 (70%)	16 (30%)	2 17
All	All	106/106 (100%)	74 (70%)	32 (30%)	2 17

5 of 32 residues with a non-rotameric sidechain are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type
2	C	46	SER
2	D	46	SER
2	C	20	LYS
2	D	20	LYS
2	C	24	ARG

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)

The completeness of assignment taking into account all chemical shift lists is 13% for the well-defined parts and 13% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 5909

Chemical shift list name: *assigned_ chem_ shift_ list_ 1*

7.1.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	268
Number of shifts mapped to atoms	268
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	—
$^{13}C_{\beta}$	0	—	—
$^{13}C'$	0	—	—
^{15}N	122	0.21 ± 0.69	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 13%, i.e. 270 atoms were assigned a chemical shift out of a possible 2104. 0 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	244/626 (39%)	122/250 (49%)	0/252 (0%)	122/124 (98%)
Sidechain	20/906 (2%)	0/536 (0%)	20/324 (6%)	0/46 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	6/92 (7%)	4/48 (8%)	0/42 (0%)	2/2 (100%)
Overall	270/2104 (13%)	126/1122 (11%)	20/778 (3%)	124/204 (61%)

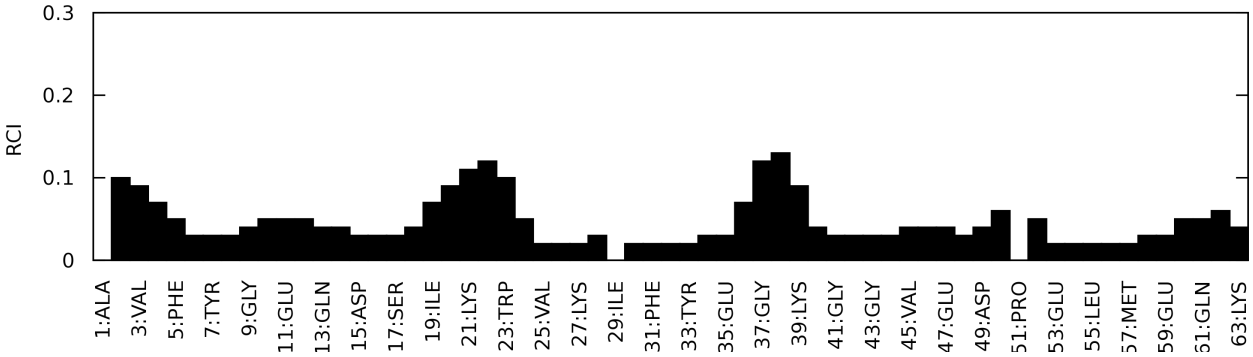
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

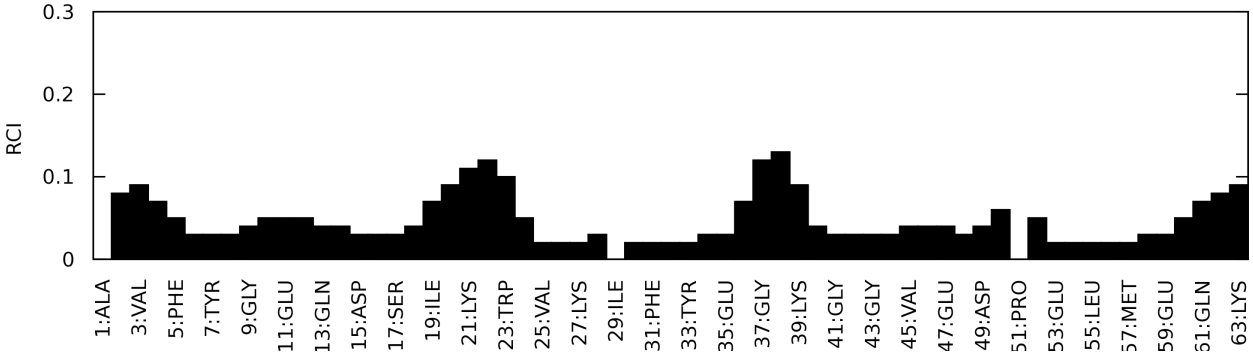
7.1.5 Random Coil Index (RCI) plots (i)

The images below report *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain C:



Random coil index (RCI) for chain D:



7.2 Chemical shift list 2

File name: BMRB entry 5910

Chemical shift list name: *assigned_chem_shift_list_1*

7.2.1 Bookkeeping (i)

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	248
Number of shifts mapped to atoms	244
Number of unparsed shifts	0
Number of shifts with mapping errors	4
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	0

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- Residue not found in structure. All 4 occurences are reported below.

Chain	Res	Type	Atom	Shift Data		
				Value	Uncertainty	Ambiguity
C	30	VAL	N	124.296	0.05	1
D	30	VAL	N	124.296	0.05	1
D	30	VAL	H	9.316	0.01	1
C	30	VAL	H	9.316	0.01	1

7.2.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	0	—	—
$^{13}C_{\beta}$	0	—	—
$^{13}C'$	0	—	—
^{15}N	122	0.30 ± 0.33	None needed (< 0.5 ppm)

7.2.3 Completeness of resonance assignments (i)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 12%, i.e. 246 atoms were assigned a chemical shift out of a possible 2104. 0 out of 16 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	240/626 (38%)	120/250 (48%)	0/252 (0%)	120/124 (97%)
Sidechain	0/906 (0%)	0/536 (0%)	0/324 (0%)	0/46 (0%)
Aromatic	6/92 (7%)	4/48 (8%)	0/42 (0%)	2/2 (100%)
Overall	246/2104 (12%)	124/1122 (11%)	0/778 (0%)	122/204 (60%)

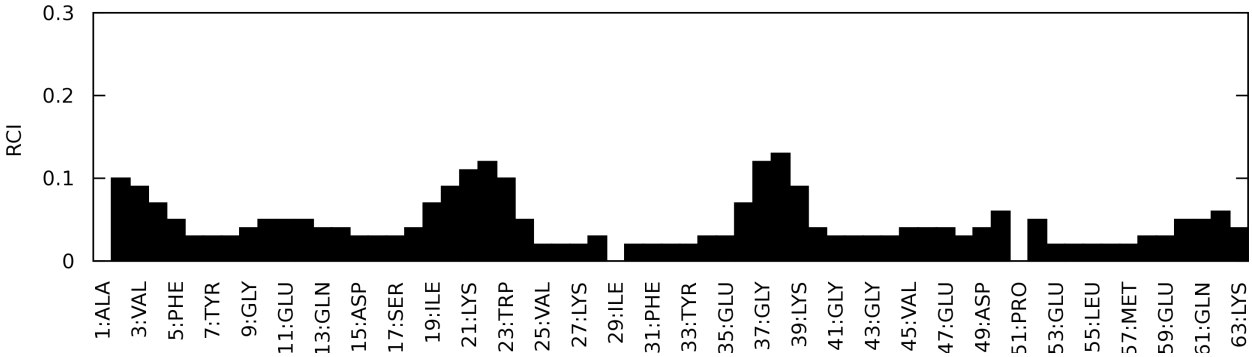
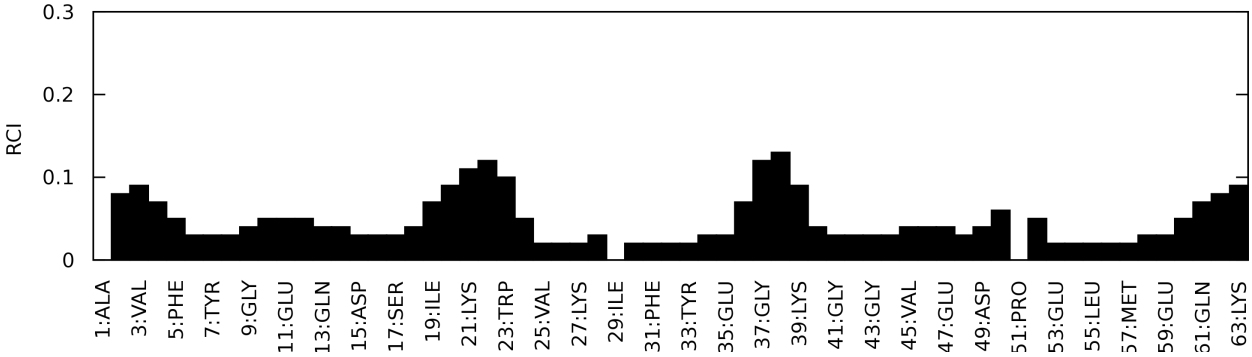
7.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.2.5 Random Coil Index (RCI) plots (i)

The images below report *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition.

Random coil index (RCI) for chain C:



Random coil index (RCI) for chain D: