wwPDB NMR Structure Validation Summary Report (i)

Apr 20, 2024 – 10:03 AM EDT

PDB ID	:	2N8A
BMRB ID	:	25888
Title	:	1H, 13C and 15N chemical shift assignments and solution structure for PARP-1 F1F2 domains in complex with a DNA single-strand break
Authors	:	Neuhaus, D.; Eustermann, S.; Yang, J.; Wu, W.
Deposited on	:	2015-10-08

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
https://www.wwpdb.org/validation/2017/NMRValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

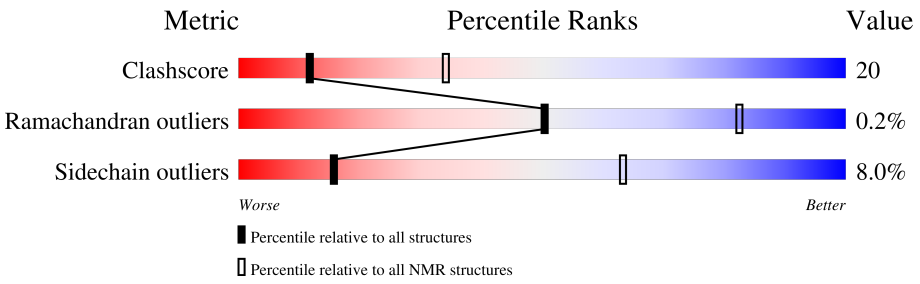
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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 30%.

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

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	214	63% 21% • 15%
2	B	45	31% 69%

2 Ensemble composition and analysis (i)

This entry contains 78 models. Model 36 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:5-A:91 (87)	0.09	65
2	A:108-A:201 (94)	0.10	36

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 11 clusters and 11 single-model clusters were found.

Cluster number	Models
1	4, 5, 6, 7, 11, 12, 13, 14, 15, 16, 17, 19, 20, 22, 23, 24, 26, 27, 28, 29, 30, 34, 39, 46, 47, 51, 52, 74
2	31, 32, 41, 42, 43, 45, 54, 56, 57, 58
3	1, 10, 18, 44, 49, 50, 53, 66
4	2, 9, 25, 60
5	8, 21, 61
6	55, 63, 73
7	40, 62, 70
8	75, 78
9	3, 37
10	64, 65
11	38, 67
Single-model clusters	33; 35; 36; 48; 59; 68; 69; 71; 72; 76; 77

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 4788 atoms, of which 2176 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Poly [ADP-ribose] polymerase 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	214	3356	1055	1670	299	319	13	0

- Molecule 2 is a DNA chain called DNA (45-MER).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	45	1430	436	506	170	273	45	0

- Molecule 3 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms	
			Total	Zn
3	A	2	2	2

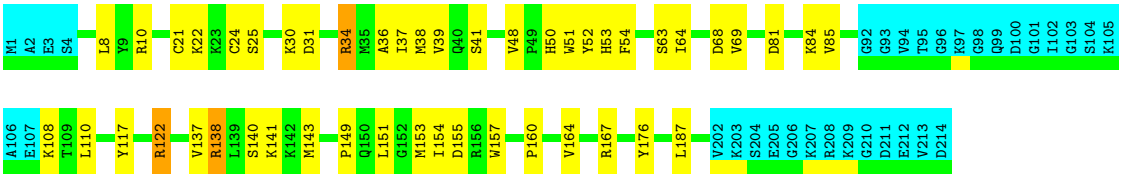
4 Residue-property plots (i)

4.1 Average score per residue in the NMR ensemble

These plots are provided for all protein, RNA, 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.

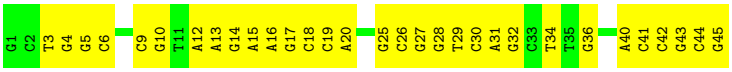
• Molecule 1: Poly [ADP-ribose] polymerase 1

Chain A: 63% 21% • 15%



• Molecule 2: DNA (45-MER)

Chain B: 31% 69%

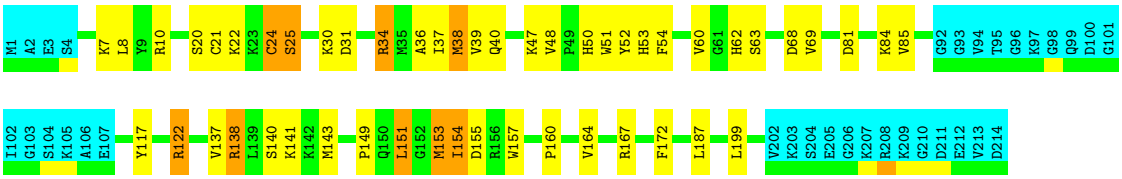


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 36. Colouring as in section 4.1 above.

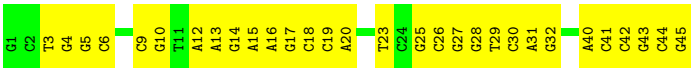
• Molecule 1: Poly [ADP-ribose] polymerase 1

Chain A: 61% 19% • 15%



• Molecule 2: DNA (45-MER)

Chain B: 33% 67%


G1
C2
T3
G4
G5
C6
C9
G10
T11
A12
A13
G14
A15
A16
G17
C18
C19
A20
T23
C24
G25
C26
G27
G28
T29
C30
A31
G32
A40
C41
C42
G43
C44
G45

5 Refinement protocol and experimental data overview (i)

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

Of the 78 calculated structures, 78 were deposited, based on the following criterion: *Total, Tensor and NOE xplor energies simultaneously below thresholds (6000, 1500 and 2 kcal.mol-1 respectively)*.

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

Software name	Classification	Version
X-PLOR NIH	structure solution	2.28
X-PLOR NIH	refinement	

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)	working_cs.cif
Number of chemical shift lists	3
Total number of shifts	1564
Number of shifts mapped to atoms	1564
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	30%

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: ZN

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	1.07±0.02	0±0/1492 (0.0± 0.0%)	1.30±0.01	0±1/1999 (0.0± 0.0%)
2	B	0.34±0.09	0±1/1035 (0.0± 0.1%)	0.75±0.03	0±1/1594 (0.0± 0.0%)
All	All	0.85	20/197106 (0.0%)	1.09	39/280254 (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
2	B	0.0±0.0	0.1±0.4
All	All	0	10

5 of 17 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	22	LYS	CA-CB	7.84	1.71	1.53	15	2
1	A	20	SER	CA-CB	-7.50	1.41	1.52	36	1
1	A	27	SER	CA-CB	-7.01	1.42	1.52	75	2
1	A	25	SER	CA-CB	-6.82	1.42	1.52	72	1
2	B	21	DG	C2'-C1'	6.47	1.58	1.52	75	1

5 of 29 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	20	SER	N-CA-CB	-10.99	94.02	110.50	42	3
1	A	25	SER	N-CA-CB	-10.20	95.20	110.50	72	1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})	Models	
								Worst	Total
1	A	24	CYS	N-CA-CB	-9.93	92.72	110.60	15	2
2	B	41	DC	O5'-P-OP1	7.17	119.31	110.70	67	1
1	A	20	SER	CB-CA-C	7.05	123.50	110.10	42	1

There are no chirality outliers.

5 of 6 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	B	24	DC	Sidechain	4
2	B	1	DG	Sidechain	2
2	B	2	DC	Sidechain	1
2	B	3	DT	Sidechain	1
2	B	13	DA	Sidechain	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	A	1456	1443	1443	46±5
2	B	924	506	505	46±4
All	All	185796	152022	151942	6667

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 20.

5 of 422 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:3:DT:N3	2:B:4:DG:C5	1.23	2.07	2	78
2:B:3:DT:C2	2:B:4:DG:C5	1.10	2.39	2	78
2:B:3:DT:C2	2:B:4:DG:C8	1.07	2.43	2	3
1:A:24:CYS:O	1:A:25:SER:OG	1.03	1.75	72	2
2:B:3:DT:C2	2:B:4:DG:N7	1.00	2.29	2	5

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	181/214 (85%)	168±1 (93±1%)	13±1 (7±1%)	0±1 (0±0%)	50	82
All	All	14118/16692 (85%)	13112 (93%)	976 (7%)	30 (0%)	50	82

All 5 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	91	ALA	18
1	A	108	LYS	7
1	A	109	THR	2
1	A	153	MET	2
1	A	5	SER	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	
1	A	158/181 (87%)	145±3 (92±2%)	13±3 (8±2%)	16	63
All	All	12324/14118 (87%)	11337 (92%)	987 (8%)	16	63

5 of 35 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	10	ARG	78
1	A	34	ARG	78
1	A	122	ARG	78
1	A	138	ARG	78
1	A	167	ARG	78

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 monosaccharides in this entry.

6.6 Ligand geometry (i)

Of 2 ligands modelled in this entry, 2 are monoatomic - leaving 0 for Mogul analysis.

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 30% for the well-defined parts and 31% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *CS_bound_F1F2*

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	877
Number of shifts mapped to atoms	877
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
$^{13}C_{\alpha}$	210	-0.04 ± 0.14	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	0	—	None (insufficient data)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	200	0.98 ± 0.23	Should be applied

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 22%, i.e. 758 atoms were assigned a chemical shift out of a possible 3383. 0 out of 22 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	523/899 (58%)	174/365 (48%)	179/362 (49%)	170/172 (99%)
Sidechain	231/1370 (17%)	176/880 (20%)	55/428 (13%)	0/62 (0%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	4/217 (2%)	4/108 (4%)	0/100 (0%)	0/9 (0%)
Sugar	0/540 (0%)	0/315 (0%)	0/225 (0%)	0/0 (—%)
Base	0/357 (0%)	0/222 (0%)	0/75 (0%)	0/60 (0%)
Overall	758/3383 (22%)	354/1890 (19%)	234/1190 (20%)	170/303 (56%)

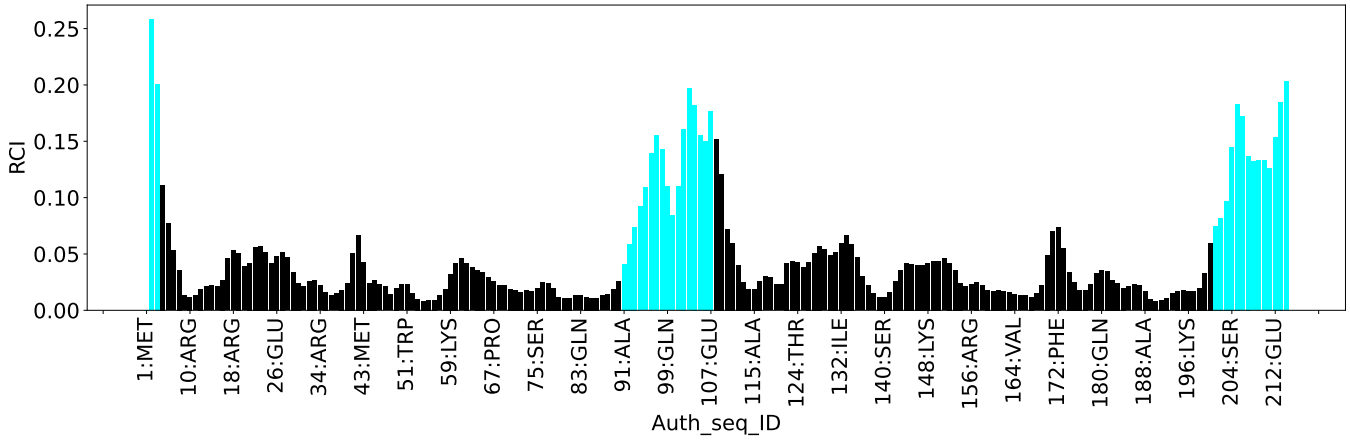
7.1.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.1.5 Random Coil Index (RCI) plots (i)

The image below reports *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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:



7.2 Chemical shift list 2

File name: working_cs.cif

Chemical shift list name: *CS_bound_F1F2_high_salt*

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	392
Number of shifts mapped to atoms	392
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

7.2.2 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	—	None (insufficient data)
$^{13}C_{\beta}$	0	—	None (insufficient data)
$^{13}C'$	0	—	None (insufficient data)
^{15}N	194	0.85 ± 0.28	Should be applied

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 10%, i.e. 340 atoms were assigned a chemical shift out of a possible 3383. 0 out of 22 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	336/899 (37%)	168/365 (46%)	0/362 (0%)	168/172 (98%)
Sidechain	0/1370 (0%)	0/880 (0%)	0/428 (0%)	0/62 (0%)
Aromatic	4/217 (2%)	2/108 (2%)	0/100 (0%)	2/9 (22%)
Sugar	0/540 (0%)	0/315 (0%)	0/225 (0%)	0/0 (—%)
Base	0/357 (0%)	0/222 (0%)	0/75 (0%)	0/60 (0%)
Overall	340/3383 (10%)	170/1890 (9%)	0/1190 (0%)	170/303 (56%)

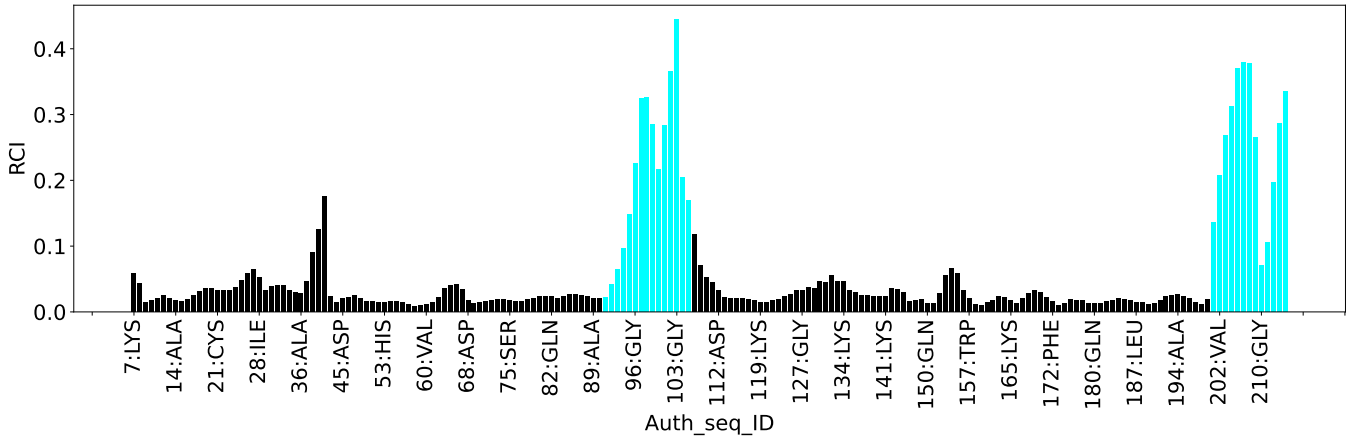
7.2.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.2.5 Random Coil Index (RCI) plots (i)

The image below reports *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.If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:



7.3 Chemical shift list 3

File name: working_cs.cif

Chemical shift list name: *bound_DNA*

7.3.1 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	295
Number of shifts mapped to atoms	295
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

7.3.2 Chemical shift referencing (i)

No chemical shift referencing corrections were calculated (not enough data).

7.3.3 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 8%, i.e. 268 atoms were assigned a chemical shift out of a possible 3383. 0 out of 22 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	**^{1}H**	**^{13}C**	**^{15}N**
Backbone	0/899 (0%)	0/365 (0%)	0/362 (0%)	0/172 (0%)
Sidechain	0/1370 (0%)	0/880 (0%)	0/428 (0%)	0/62 (0%)
Aromatic	0/217 (0%)	0/108 (0%)	0/100 (0%)	0/9 (0%)
Sugar	173/540 (32%)	173/315 (55%)	0/225 (0%)	0/0 (—%)
Base	95/357 (27%)	95/222 (43%)	0/75 (0%)	0/60 (0%)
Overall	268/3383 (8%)	268/1890 (14%)	0/1190 (0%)	0/303 (0%)

7.3.4 Statistically unusual chemical shifts (i)

There are no statistically unusual chemical shifts.

7.3.5 Random Coil Index (RCI) plots (i)

No *random coil index*(RCI) plot could be generated from the current chemical shift list. RCI is only applicable to proteins