Full wwPDB NMR Structure Validation Report (i)

Jun 17, 2024 – 10:06 PM EDT

PDB ID	:	5VX7
BMRB ID	:	30298
Title	:	Solution NMR structure of the BRCT domain of S. cerevisiae Rev1
Authors	:	Xu, C.; Cui, G.; Botuyan, M.V.; Mer, G.
Deposited on	:	2017-05-23

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 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:

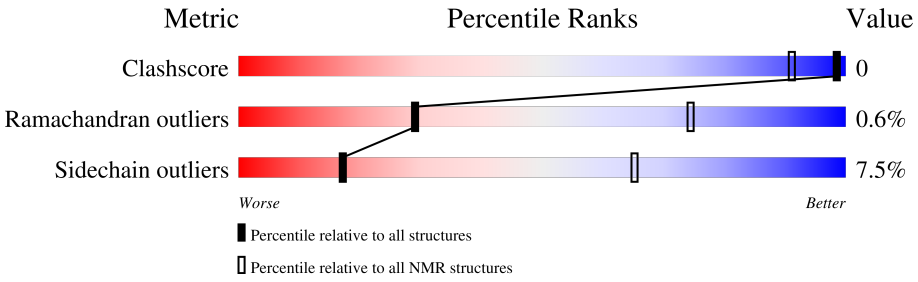
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
SOLUTION NMR

The overall completeness of chemical shifts assignment is 83%.

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	93	83% 5% 10% •

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 2 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:164-A:198, A:205-A:251 (82)	0.70	2

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 2 clusters and 1 single-model cluster was found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1512 atoms, of which 771 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called DNA repair protein REV1.

Mol	Chain	Residues	Atoms						Trace
1	A	91	Total	C	H	N	O	S	0
			1512	484	771	127	127	3	

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-3	GLY	-	expression tag	UNP P12689
A	-2	HIS	-	expression tag	UNP P12689
A	-1	MET	-	expression tag	UNP P12689

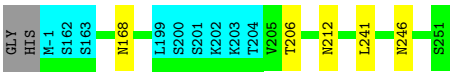
4 Residue-property plots

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: DNA repair protein REV1

Chain A:



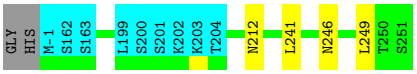
4.2 Scores per residue for each member of the ensemble

Colouring as in section 4.1 above.

4.2.1 Score per residue for model 1

• Molecule 1: DNA repair protein REV1

Chain A:



4.2.2 Score per residue for model 2 (medoid)

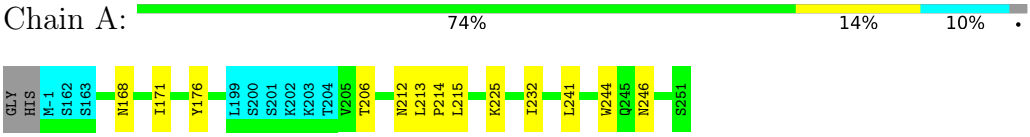
• Molecule 1: DNA repair protein REV1

Chain A:

4.2.3 Score per residue for model 3

• Molecule 1: DNA repair protein REV1

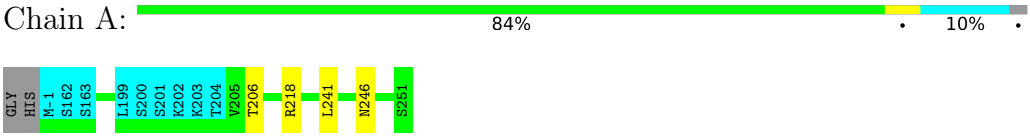
Chain A:



4.2.4 Score per residue for model 4

• Molecule 1: DNA repair protein REV1

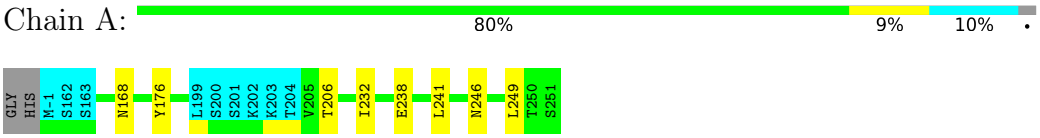
Chain A:



4.2.5 Score per residue for model 5

• Molecule 1: DNA repair protein REV1

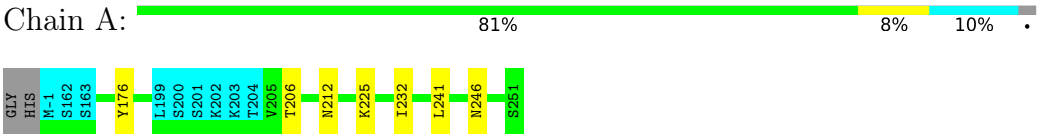
Chain A:



4.2.6 Score per residue for model 6

• Molecule 1: DNA repair protein REV1

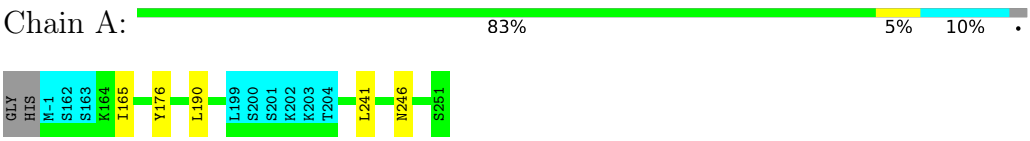
Chain A:



4.2.7 Score per residue for model 7

• Molecule 1: DNA repair protein REV1

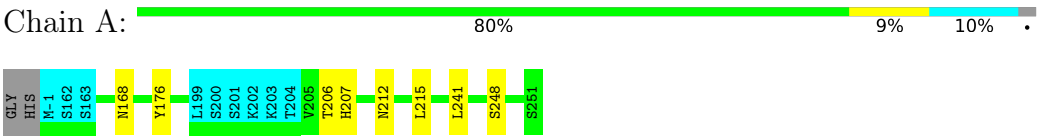
Chain A:



4.2.8 Score per residue for model 8

• Molecule 1: DNA repair protein REV1

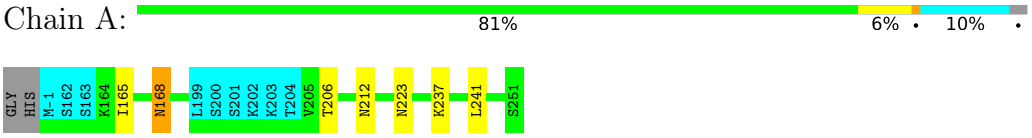
Chain A:



4.2.9 Score per residue for model 9

• Molecule 1: DNA repair protein REV1

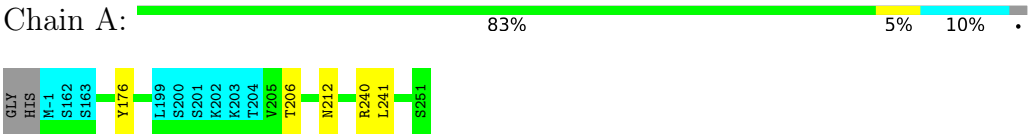
Chain A:



4.2.10 Score per residue for model 10

• Molecule 1: DNA repair protein REV1

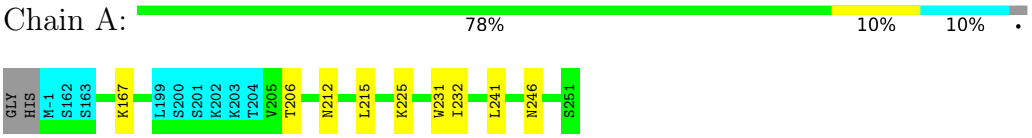
Chain A:



4.2.11 Score per residue for model 11

• Molecule 1: DNA repair protein REV1

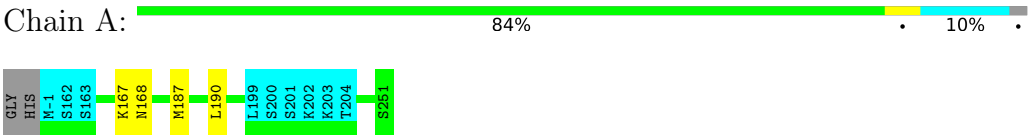
Chain A:



4.2.12 Score per residue for model 12

• Molecule 1: DNA repair protein REV1

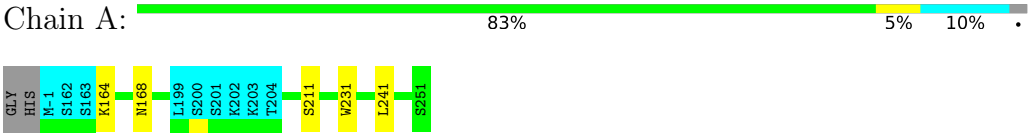
Chain A:



4.2.13 Score per residue for model 13

• Molecule 1: DNA repair protein REV1

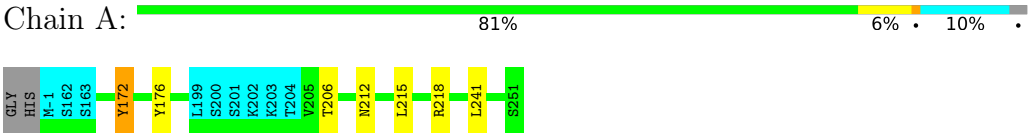
Chain A:



4.2.14 Score per residue for model 14

• Molecule 1: DNA repair protein REV1

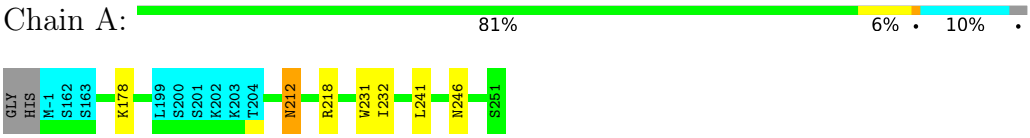
Chain A:



4.2.15 Score per residue for model 15

• Molecule 1: DNA repair protein REV1

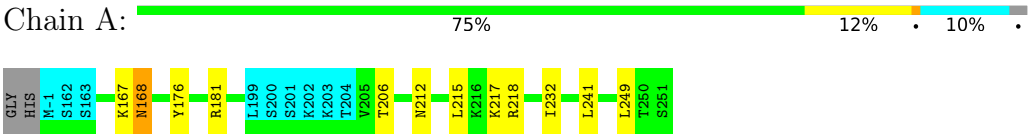
Chain A:



4.2.16 Score per residue for model 16

• Molecule 1: DNA repair protein REV1

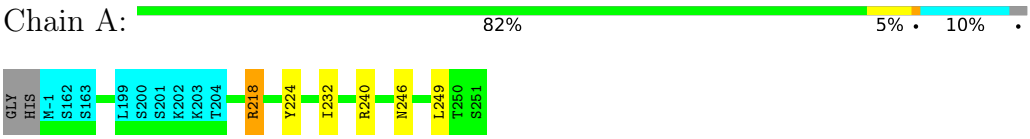
Chain A:



4.2.17 Score per residue for model 17

• Molecule 1: DNA repair protein REV1

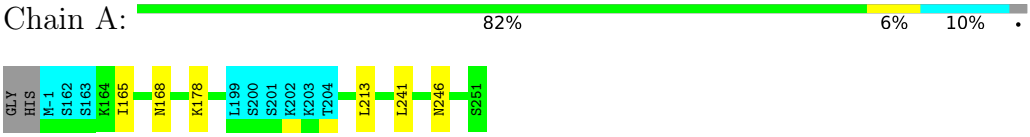
Chain A:



4.2.18 Score per residue for model 18

• Molecule 1: DNA repair protein REV1

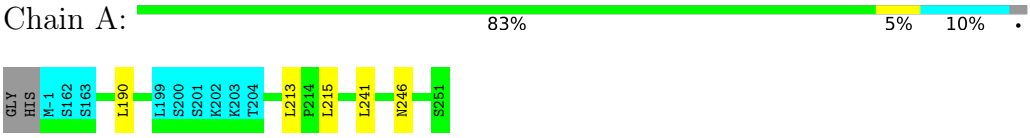
Chain A:



4.2.19 Score per residue for model 19

• Molecule 1: DNA repair protein REV1

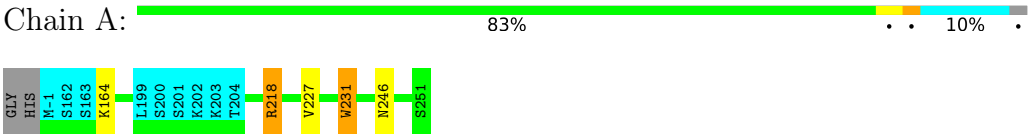
Chain A:



4.2.20 Score per residue for model 20

• Molecule 1: DNA repair protein REV1

Chain A:



5 Refinement protocol and experimental data overview

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

Of the 200 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
CYANA	structure calculation	
Amber	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	1
Total number of shifts	1078
Number of shifts mapped to atoms	1076
Number of unparsed shifts	0
Number of shifts with mapping errors	2
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	83%

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	0.67±0.00	0±0/695 (0.0± 0.0%)	0.98±0.02	0±1/941 (0.0± 0.1%)
All	All	0.67	0/13900 (0.0%)	0.98	4/18820 (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
1	A	0.0±0.0	0.5±0.5
All	All	0	9

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
1	A	172	TYR	CB-CG-CD1	-5.89	117.46	121.00	14	1
1	A	240	ARG	NE-CZ-NH2	-5.47	117.56	120.30	17	2
1	A	224	TYR	CB-CG-CD2	-5.01	117.99	121.00	17	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	176	TYR	Sidechain	7
1	A	218	ARG	Sidechain	2

6.2 Too-close contacts

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	676	698	698	0±0
All	All	13520	13960	13960	2

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:227:VAL:HG11	1:A:231:TRP:CE3	0.45	2.46	20	1
1:A:225:LYS:HE3	1:A:244:TRP:CZ2	0.44	2.47	3	1

6.3 Torsion angles

6.3.1 Protein backbone

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	81/93 (87%)	73±2 (90±3%)	8±3 (10±3%)	0±1 (1±1%)	29	74
All	All	1620/1860 (87%)	1454 (90%)	156 (10%)	10 (1%)	29	74

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	167	LYS	4
1	A	168	ASN	3
1	A	214	PRO	1
1	A	164	LYS	1
1	A	212	ASN	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	75/85 (88%)	69±2 (92±3%)	6±2 (8±3%)	17	65
All	All	1500/1700 (88%)	1387 (92%)	113 (8%)	17	65

All 28 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	241	LEU	17
1	A	246	ASN	12
1	A	206	THR	11
1	A	212	ASN	10
1	A	168	ASN	9
1	A	232	ILE	8
1	A	215	LEU	6
1	A	218	ARG	6
1	A	249	LEU	4
1	A	231	TRP	4
1	A	213	LEU	3
1	A	165	ILE	3
1	A	190	LEU	3
1	A	225	LYS	2
1	A	178	LYS	2
1	A	171	ILE	1
1	A	238	GLU	1
1	A	176	TYR	1
1	A	207	HIS	1
1	A	248	SER	1
1	A	223	ASN	1
1	A	237	LYS	1
1	A	187	MET	1
1	A	211	SER	1
1	A	172	TYR	1
1	A	181	ARG	1
1	A	217	LYS	1
1	A	164	LYS	1

6.3.3 RNA

There are no RNA molecules in this entry.

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

There are no chain breaks in this entry.

7 Chemical shift validation

The completeness of assignment taking into account all chemical shift lists is 83% for the well-defined parts and 82% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *Rev1-BRCT_star31*

7.1.1 Bookkeeping

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	1078
Number of shifts mapped to atoms	1076
Number of unparsed shifts	0
Number of shifts with mapping errors	2
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

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

- No matching atom found in the structure. All 2 occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	185	HIS	HD1	7.789	0.03	1
1	A	207	HIS	HD1	9.495	0.03	1

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	91	2.07 ± 0.11	Should be applied
$^{13}C_{\beta}$	87	2.33 ± 0.21	Should be applied
$^{13}C'$	83	2.57 ± 0.16	Should be applied
^{15}N	84	0.82 ± 0.60	None needed (imprecise)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 83%, i.e. 991 atoms were assigned a chemical shift out of a possible 1199. 0 out of 17 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	395/406 (97%)	161/164 (98%)	157/164 (96%)	77/78 (99%)
Sidechain	557/666 (84%)	379/437 (87%)	171/205 (83%)	7/24 (29%)
Aromatic	39/127 (31%)	37/63 (59%)	0/58 (0%)	2/6 (33%)
Overall	991/1199 (83%)	577/664 (87%)	328/427 (77%)	86/108 (80%)

The following table shows the completeness of the chemical shift assignments for the full structure. The overall completeness is 82%, i.e. 1076 atoms were assigned a chemical shift out of a possible 1311. 0 out of 18 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	435/451 (96%)	177/182 (97%)	174/182 (96%)	84/87 (97%)
Sidechain	602/733 (82%)	408/481 (85%)	187/226 (83%)	7/26 (27%)
Aromatic	39/127 (31%)	37/63 (59%)	0/58 (0%)	2/6 (33%)
Overall	1076/1311 (82%)	622/726 (86%)	361/466 (77%)	93/119 (78%)

7.1.4 Statistically unusual chemical shifts

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	177	THR	HG1	5.25	0.08 – 2.19	19.5
1	A	225	LYS	HD3	-0.53	0.54 – 2.65	-10.1
1	A	225	LYS	HG2	-0.56	0.13 – 2.61	-7.8
1	A	225	LYS	HD2	0.03	0.58 – 2.64	-7.7
1	A	245	GLN	HG3	0.39	0.91 – 3.68	-6.9
1	A	225	LYS	HB2	0.31	0.58 – 2.97	-6.1

7.1.5 Random Coil Index (RCI) plots

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: