



wwPDB NMR Structure Validation Summary Report ⓘ

May 31, 2020 – 12:15 pm BST

PDB ID : 6PSI
Title : Structural Basis for Client Recognition and Activity of Hsp40 Chaperones
Authors : Jiang, Y.; Rossi, P.; Kalodimos, C.G.
Deposited on : 2019-07-12

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 ⓘ symbol.

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

Cyrange : Kirchner and Güntert (2011)
NmrClust : Kelley et al. (1996)
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 : 2.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.11

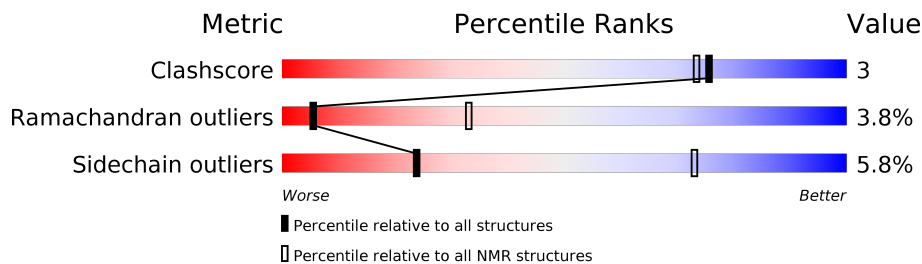
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 22%.

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 $\leq 5\%$.

Mol	Chain	Length	Quality of chain
1	A	280	87% 10% .
1	C	280	86% 10% .
2	B	471	5% . 94%

2 Ensemble composition and analysis

This entry contains 20 models. Model 4 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:4-A:108, A:115-A:280, B:1-B:11, B:179-B:185, B:238-B:245, C:3-C:105, C:116-C:280 (565)	1.61	4

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. No single-model clusters were found.

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

3 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 15729 atoms, of which 7873 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Chaperone protein DnaJ 2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	280	4399	1398	2207	393	398	3	0
1	C	280	4399	1398	2207	393	398	3	0

- Molecule 2 is a protein called Alkaline phosphatase.

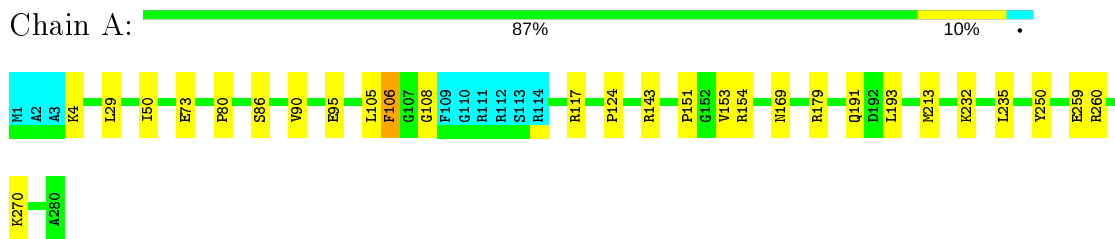
Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	471	6931	2154	3459	609	696	13	0

4 Residue-property plots [i](#)

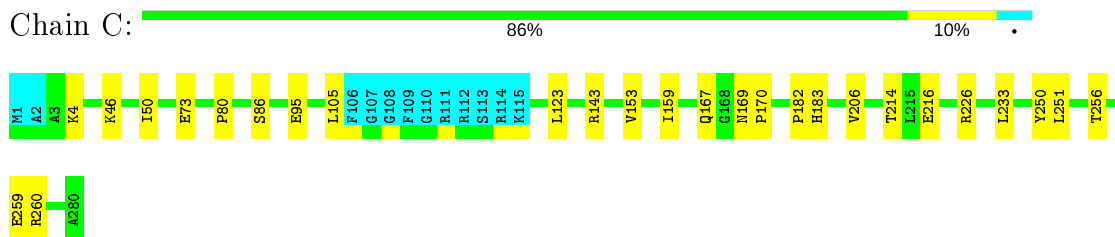
4.1 Average score per residue in the NMR ensemble

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.

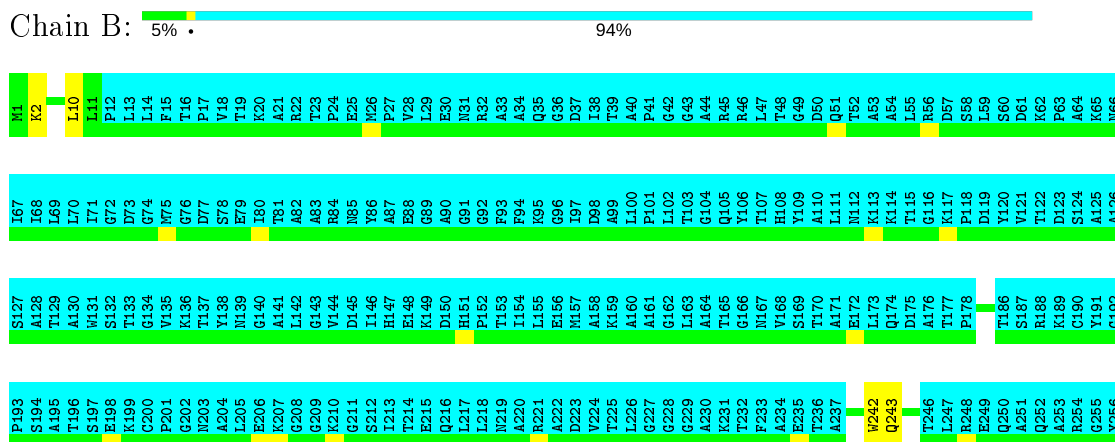
- Molecule 1: Chaperone protein DnaJ 2



- Molecule 1: Chaperone protein DnaJ 2



- Molecule 2: Alkaline phosphatase

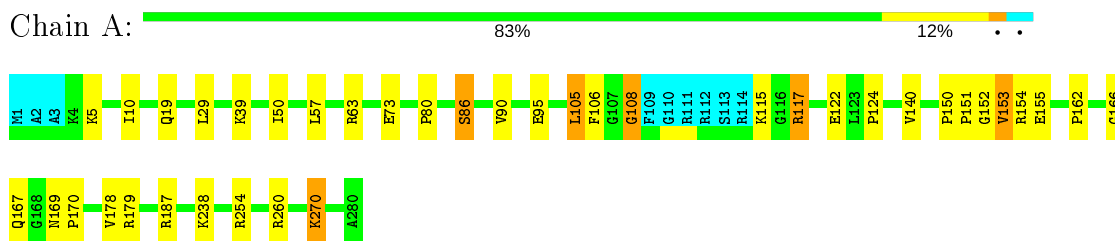




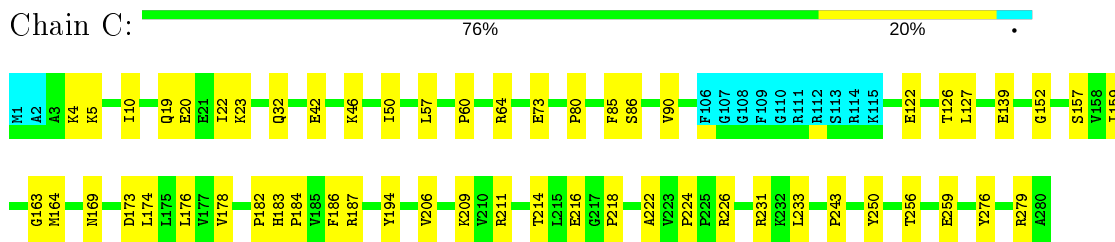
4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 4. Colouring as in section 4.1 above.

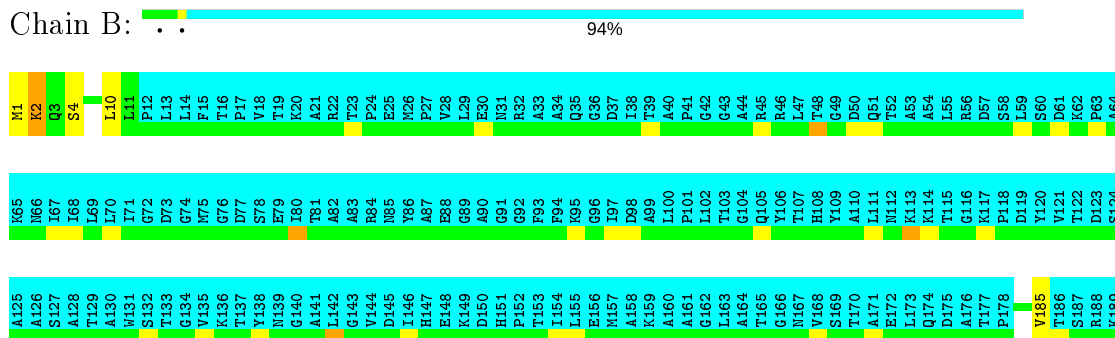
- Molecule 1: Chaperone protein DnaJ 2



- Molecule 1: Chaperone protein DnaJ 2



- Molecule 2: Alkaline phosphatase



Q482	Q372	P312	Q352	C190
E433	R373	Q313	A283	Y191
H434	A374	R314	R284	G192
T435	L375	N315	G285	P193
G436	E376	D316	Y286	S194
S437	F377	S317	Q287	A195
Q438	A378	N318	L288	T196
L439	K379	F319	V289	S197
R440	K380	T320	S290	E198
I441	E381	L321	D291	K199
A442	G382	A322	A292	C200
A443	N383	Q323	A293	P201
Y444	T384	N324	S294	G202
G445	L385	T325	L295	N203
P446	V386	D326	M296	A204
H447	I387	K327	S297	L205
A448	V388	A328	V298	E206
A449	T389	I329	T299	K207
M450	A390	R330	E270	G208
V451	D391	L331	A271	G209
V452	R392	L332	M272	K210
G453	A393	S333	Q273	G211
L454	R394	K334	Q274	S212
T455	A395	N335	K275	I213
D456	S396	R336	P276	T214
Q457	Q397	K337	L277	E215
T458	I398	G338	G278	Q216
D459	V399	F339	L279	L217
F461	A400	L341	L280	L218
Y462	D402	Q342	F281	M219
T463	T403	R343	A282	A220
M464	K404	E344	D283	R221
K465	A405	G345	A222	A222
A466	P406	A346	M285	D223
A467	G407	S347	M286	Y224
L468	L408	I348	P287	T225
G469	T409	R349	V288	L226
L470	Q410	K350	R289	G227
R471	A411	Q351	M290	G228
	L412	D352	L291	G229
	N413	H353	G292	A230
	T414	A354	K294	K231
	K415	A355	A295	F232
	D416	N356	T296	F233
	G417	P357	Y297	A234
	A418	C358	H298	E235
	V419	G359	G299	T236
	M420	Q360	M300	A237
	M421	I361	I301	T238
	M422	G362	D302	A239
	S423	E363	K303	N242
	Y424	T364	P304	Q243
	G425	V365	A305	
	M426	D366	V306	T246
	S427	L367	T307	L247
	E428	D368	C308	K248
	E429	E369	T309	E249
	D430	A370	P310	Q250
	S431	V371	N311	A251

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *molecular dynamics*.

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *target function*.

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

Software name	Classification	Version
CNS	refinement	
CYANA	structure calculation	
TALOS	geometry optimization	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 6 of this report.

Chemical shift file(s)	input_cs.cif
Number of chemical shift lists	2
Total number of shifts	2726
Number of shifts mapped to atoms	2726
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	22%

No validations of the models with respect to experimental NMR restraints is performed at this time.

COVALENT-GEOMETRY INFOmissingINFO

5.1 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	2120	2130	2127	14±3
1	C	2097	2107	2104	15±4
2	B	189	205	204	2±1
All	All	88120	88840	88700	556

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 3.

5 of 322 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:C:118:ASP:HA	1:C:166:GLY:HA2	0.70	1.62	19	1
1:A:127:LEU:HB3	1:A:215:LEU:HD22	0.68	1.64	18	6
1:C:22:ILE:HD13	1:C:57:LEU:HB3	0.68	1.64	8	7
1:A:166:GLY:HA3	1:A:172:GLY:HA3	0.63	1.69	6	2
1:A:108:GLY:HA3	1:A:167:GLN:HB3	0.62	1.71	20	4

5.2 Torsion angles [i](#)

5.2.1 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	270/280 (96%)	233±4 (86±2%)	29±4 (11±1%)	9±2 (3±1%)	6	37
1	C	267/280 (95%)	235±3 (88±1%)	24±3 (9±1%)	9±2 (3±1%)	7	38
2	B	25/471 (5%)	14±3 (56±11%)	7±3 (28±11%)	4±2 (16±7%)	0	4
All	All	11240/20620 (55%)	9630 (86%)	1183 (11%)	427 (4%)	5	33

5 of 102 unique Ramachandran outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Models (Total)
1	C	182	PRO	19
1	C	259	GLU	19
1	A	259	GLU	18
1	C	80	PRO	16
1	A	86	SER	16

5.2.2 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	218/224 (97%)	207±3 (95±1%)	12±3 (5±1%)	26	75
1	C	216/224 (96%)	204±3 (94±2%)	12±3 (6±2%)	24	73
2	B	18/359 (5%)	16±1 (88±7%)	2±1 (12±7%)	8	51
All	All	9040/16140 (56%)	8520 (94%)	520 (6%)	24	73

5 of 125 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	C	214	THR	20
1	A	73	GLU	20
1	C	73	GLU	19
1	A	260	ARG	17
1	A	95	GLU	17

5.2.3 RNA [i](#)

There are no RNA molecules in this entry.

5.3 Non-standard residues in protein, DNA, RNA chains [i](#)

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

5.4 Carbohydrates [i](#)

There are no carbohydrates in this entry.

5.5 Ligand geometry [i](#)

There are no ligands in this entry.

5.6 Other polymers [i](#)

There are no such molecules in this entry.

5.7 Polymer linkage issues [i](#)

There are no chain breaks in this entry.

6 Chemical shift validation [i](#)

The completeness of assignment taking into account all chemical shift lists is 22% for the well-defined parts and 19% for the entire structure.

6.1 Chemical shift list 1

File name: input_cs.cif

Chemical shift list name: *assigned_chemical_shifts_1*

6.1.1 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	1621
Number of shifts mapped to atoms	1621
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

6.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	225	-0.12 ± 0.08	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	194	0.21 ± 0.08	None needed (< 0.5 ppm)
$^{13}\text{C}'$	86	-0.16 ± 0.07	None needed (< 0.5 ppm)
^{15}N	218	0.01 ± 0.26	None needed (< 0.5 ppm)

6.1.3 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 19%, i.e. 1327 atoms were assigned a chemical shift out of a possible 7017. 0 out of 102 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	807/2713 (30%)	292/1074 (27%)	301/1130 (27%)	214/509 (42%)
Sidechain	520/3831 (14%)	181/2263 (8%)	339/1371 (25%)	0/197 (0%)

Continued on next page...

Continued from previous page...

	Total	¹ H	¹³ C	¹⁵ N
Aromatic	0/473 (0%)	0/247 (0%)	0/209 (0%)	0/17 (0%)
Overall	1327/7017 (19%)	473/3584 (13%)	640/2710 (24%)	214/723 (30%)

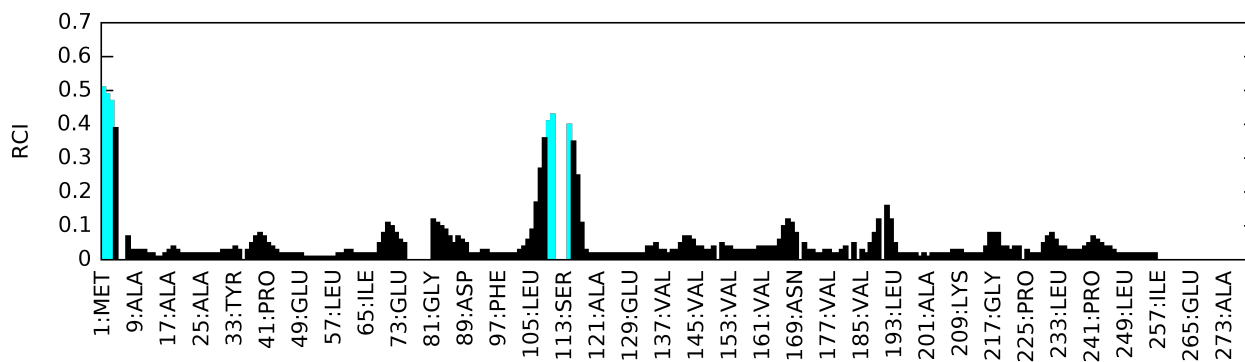
6.1.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

6.1.5 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.

Random coil index (RCI) for chain A:



6.2 Chemical shift list 2

File name: input_cs.cif

Chemical shift list name: *assigned_chemical_shifts_2*

6.2.1 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	1105
Number of shifts mapped to atoms	1105
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

6.2.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, <i>ppm</i>	Suggested action
$^{13}\text{C}_\alpha$	31	-0.19 ± 0.16	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	32	0.18 ± 0.41	None needed (< 0.5 ppm)
$^{13}\text{C}'$	23	—	None (insufficient data)
^{15}N	379	-0.81 ± 0.14	Should be applied

6.2.3 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 3%, i.e. 187 atoms were assigned a chemical shift out of a possible 7017. 0 out of 102 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	81/2713 (3%)	35/1074 (3%)	26/1130 (2%)	20/509 (4%)
Sidechain	94/3831 (2%)	54/2263 (2%)	39/1371 (3%)	1/197 (1%)
Aromatic	12/473 (3%)	6/247 (2%)	5/209 (2%)	1/17 (6%)
Overall	187/7017 (3%)	95/3584 (3%)	70/2710 (3%)	22/723 (3%)

6.2.4 Statistically unusual chemical shifts [i](#)

There are no statistically unusual chemical shifts.

6.2.5 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.

Random coil index (RCI) for chain B:

