wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 12:38 AM BST

PDB ID	:	2L85
Title	:	Solution NMR structures of CBP bromodomain with small molecule of HBS
Authors	:	Borah, J.C.; Mujtaba, S.; Karakikes, I.; Zeng, L.; Muller, M.; Patel, J.; Moshkina, N.; Morohashi, K.; Zhang, W.; Gerona-Navarro, G.; Hajjar, R.J.; Zhou, M.
Deposited on	:	2011-01-04

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:

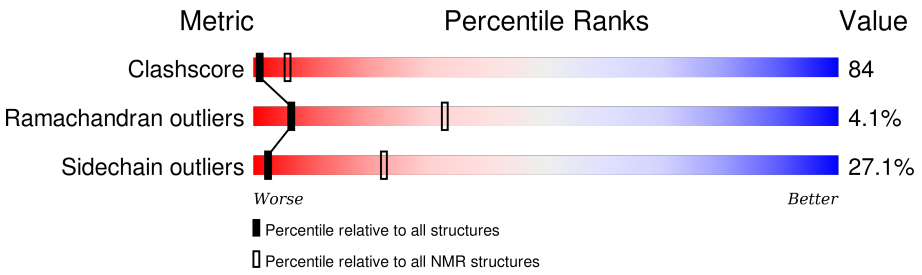
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 85%.

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	121	13% 54% 21% 12%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 11 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:1087-A:1168, A:1172-A:1195 (106)	0.18	11

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 4 clusters and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2051 atoms, of which 1017 are hydrogens and 0 are deuteriums.

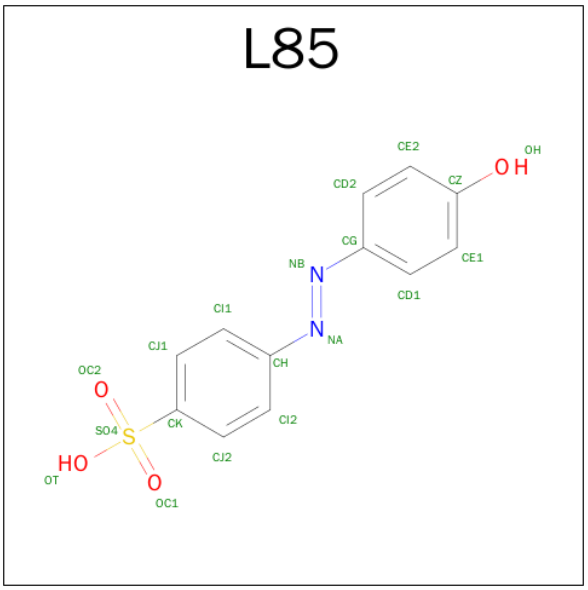
- Molecule 1 is a protein called CREB-binding protein.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	121	2022	655	1007	169	185	6	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1077	GLY	-	EXPRESSION TAG	UNP Q92793
A	1078	SER	-	EXPRESSION TAG	UNP Q92793
A	1079	HIS	-	EXPRESSION TAG	UNP Q92793
A	1080	MET	-	EXPRESSION TAG	UNP Q92793

- Molecule 2 is 4-[(E)-(4-HYDROXYPHENYL)DIAZENYL]BENZENESULFONIC ACID (three-letter code: L85) (formula: $C_{12}H_{10}N_2O_4S$).



Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	S
2	A	1	29	12	10	2	4	1

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: CREB-binding protein

Chain A: 13% 54% 21% 12%

G1077 S1078 H1079 M1080 R1081 K1082 K1083 I1084 F1085 K1086 E1089 L1090 R1091 Q1092 A1093 L1094 M1095 P1096 T1097 L1098 E1099 A1100 L1101 Y1102 R1103 Q1104 D1105 P1106 E1107 S1108 L1109 P1110 F1111 R1112 Q1113 P1114 V1115 D1116 P1117 Q1118 L1119 L1120 G1121 I1122 P1123 D1124 Y1125 F1126 D1127 I1128 V1129 P1132 M1133 D1134 L1135 I1138 K1139

R1140 K1141 L1142 D1143 T1144 G1145 Q1146 Y1147 Q1148 E1149 P1150 W1151 Q1152 Y1153 V1154 D1155 D1156 V1157 W1158 L1159 M1160 F1161 N1162 N1163 A1164 W1165 L1166 Y1167 N1168 R1169 K1170 T1171 S1172 R1173 V1174 Y1175 K1176 F1177 C1178 S1179 K1180 L1181 A1182 E1183 V1184 F1185 E1186 Q1187 E1188 I1189 D1190 P1191 V1192 M1193 Q1194 S1195 L1196 G1197

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

The representative model is number 11. Colouring as in section 4.1 above.

- Molecule 1: CREB-binding protein

Chain A: 16% 48% 21% • 12%

G1077 S1078 H1079 M1080 R1081 K1082 K1083 I1084 F1085 K1086 P1087 E1088 E1089 L1090 R1091 Q1092 A1093 L1094 M1095 P1096 T1097 L1098 E1099 A1100 L1101 Y1102 R1103 Q1104 D1105 P1106 E1107 S1108 L1109 P1110 F1111 R1112 Q1113 P1114 V1115 D1116 P1117 Q1118 L1119 L1120 G1121 I1122 P1123 D1124 Y1125 F1126 D1127 I1128 V1129 P1132 M1133 D1134 L1135 I1138

K1139 R1140 K1141 L1142 D1143 Q1146 Y1147 Q1148 E1149 P1150 W1151 Q1152 Y1153 V1154 D1155 D1156 V1157 W1158 L1159 M1160 F1161 N1162 N1163 A1164 W1165 L1166 Y1167 N1168 R1169 K1170 T1171 S1172 R1173 V1174 Y1175 K1176 F1177 C1178 S1179 K1180 L1181 A1182 E1183 V1184 F1185 E1186 Q1187 E1188 I1189 D1190 P1191 V1192 M1193 Q1194 S1195 L1196 G1197

5 Refinement protocol and experimental data overview (i)

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
ARIA	refinement	2.2
CNS	structure solution	1.2

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)	2l85_cs.str
Number of chemical shift lists	1
Total number of shifts	1378
Number of shifts mapped to atoms	1378
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	85%

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)

Bond lengths and bond angles in the following residue types are not validated in this section: L85

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	0.47±0.01	0±0/917 (0.0±0.0%)	0.65±0.01	0±0/1250 (0.0±0.0%)
All	All	0.47	0/18340 (0.0%)	0.65	4/25000 (0.0%)

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	1167	TYR	CB-CG-CD2	-5.41	117.76	121.00	15	3
1	A	1147	TYR	CB-CG-CD2	-5.09	117.95	121.00	16	1

There are no chirality outliers.

There are no planarity outliers.

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	890	861	861	150±7
2	A	19	10	9	3±1
All	All	18180	17420	17386	2997

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 84.

5 of 483 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:1093:ALA:HB1	1:A:1192:VAL:HG22	1.11	1.22	20	20
1:A:1101:LEU:HD13	1:A:1135:LEU:CD2	0.96	1.90	17	20
1:A:1094:LEU:HD12	1:A:1147:TYR:CE2	0.91	2.01	16	1
1:A:1094:LEU:HD11	1:A:1150:PRO:HA	0.84	1.48	3	20
1:A:1093:ALA:CB	1:A:1192:VAL:HG22	0.83	2.03	9	20

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	106/121 (88%)	75±3 (71±3%)	27±3 (25±3%)	4±1 (4±1%)	6	33
All	All	2120/2420 (88%)	1501 (71%)	532 (25%)	87 (4%)	6	33

5 of 10 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	1123	PRO	20
1	A	1105	ASP	20
1	A	1133	MET	18
1	A	1168	ASN	12
1	A	1110	PRO	6

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	100/113 (88%)	73±3 (73±3%)	27±3 (27±3%)	2	22

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
All	All	2000/2260 (88%)	1459 (73%)	541 (27%)	2 22

5 of 56 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	1159	LEU	20
1	A	1180	LYS	20
1	A	1167	TYR	20
1	A	1157	VAL	20
1	A	1109	LEU	20

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)

1 ligand is modelled in this entry.

In the following table, the Counts columns list the number of bonds for which Mogul statistics could be retrieved, the number of bonds that are observed in the model and the number of bonds that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond length is the number of standard deviations the observed value is removed from the expected value. A bond length with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond lengths.

Mol	Type	Chain	Res	Link	Bond lengths		
					Counts	RMSZ	#Z>2
2	L85	A	201	-	20,20,20	1.05±0.02	0±0 (0±0%)

In the following table, the Counts columns list the number of angles for which Mogul statistics could be retrieved, the number of angles that are observed in the model and the number of angles that are defined in the chemical component dictionary. The Link column lists molecule types, if any, to which the group is linked. The Z score for a bond angle is the number of standard deviations the observed value is removed from the expected value. A bond angle with $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the average root-mean-square of all Z scores of the bond angles.

Mol	Type	Chain	Res	Link	Bond angles		
					Counts	RMSZ	#Z>2
2	L85	A	201	-	28,28,28	0.62±0.01	0±0 (0±0%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	L85	A	201	-	-	0±0,11,11,11	0±0,2,2,2

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

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 85% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: 2l85_cs.str

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	1378
Number of shifts mapped to atoms	1378
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	8

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	117	-0.43 ± 0.15	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	114	0.34 ± 0.16	None needed (< 0.5 ppm)
$^{13}C'$	0	—	—
^{15}N	107	0.57 ± 0.17	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 85%, i.e. 1190 atoms were assigned a chemical shift out of a possible 1392. 0 out of 19 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	403/510 (79%)	202/202 (100%)	105/212 (50%)	96/96 (100%)
Sidechain	662/753 (88%)	415/446 (93%)	235/274 (86%)	12/33 (36%)

Continued on next page...

Continued from previous page...

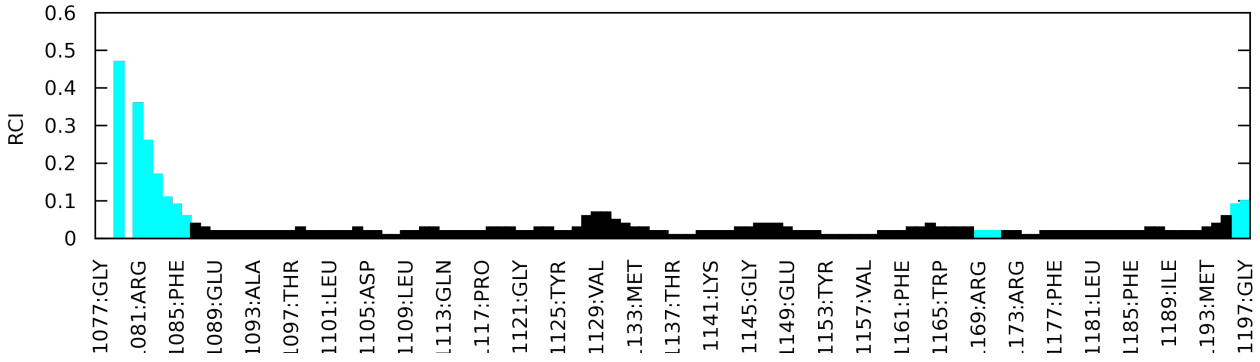
	Total	^{1}H	^{13}C	^{15}N
Aromatic	125/129 (97%)	65/67 (97%)	57/59 (97%)	3/3 (100%)
Overall	1190/1392 (85%)	682/715 (95%)	397/545 (73%)	111/132 (84%)

7.1.4 Statistically unusual chemical shifts (i)

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.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	1103	ARG	NE	123.77	92.63 – 76.73	24.6
1	A	1091	ARG	NE	123.61	92.63 – 76.73	24.5
1	A	1112	ARG	NE	121.92	92.63 – 76.73	23.4
1	A	1151	TRP	NE1	110.67	139.19 – 119.59	-9.6
1	A	1157	VAL	HG21	-1.51	2.20 – -0.60	-8.3
1	A	1157	VAL	HG22	-1.51	2.20 – -0.60	-8.3
1	A	1157	VAL	HG23	-1.51	2.20 – -0.60	-8.3
1	A	1165	TRP	HD1	5.02	8.95 – 5.35	-5.9

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.

Random coil index (RCI) for chain A: