wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 07:16 PM BST

PDB ID	:	2BJC
Title	:	NMR STRUCTURE OF A PROTEIN-DNA COMPLEX OF AN ALTERED SPECIFICITY MUTANT OF THE LAC REPRESSOR HEADPIECE THAT MIMICS THE GAL REPRESSOR
Authors	:	Salinas, R.K.; Folkers, G.E.; Bonvin, A.M.J.J.; Das, D.; Boelens, R.; Kaptein, R.
Deposited on	:	2005-02-01

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:

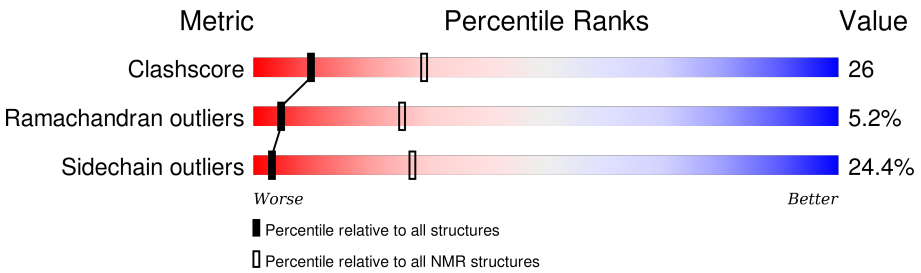
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 72%.

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	62	58% 32% • 8%
1	B	62	58% 32% • 8%
2	C	22	18% 82%
2	D	22	9% 91%

2 Ensemble composition and analysis (i)

This entry contains 16 models. Model 1 is the overall representative, medoid model (most similar to other models). The authors have identified model 3 as representative.

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:3-A:59, B:103-B:159 (114)	0.26	1

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. No single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 3088 atoms, of which 1256 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called LACTOSE OPERON REPRESSOR.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	62	845	290	377	84	91	3	0
1	B	62	845	290	377	84	91	3	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	17	VAL	TYR	ENGINEERED MUTATION	UNP P03023
A	18	ALA	GLN	ENGINEERED MUTATION	UNP P03023
A	52	CYS	VAL	ENGINEERED MUTATION	UNP P03023
B	117	VAL	TYR	ENGINEERED MUTATION	UNP P03023
B	118	ALA	GLN	ENGINEERED MUTATION	UNP P03023
B	152	CYS	VAL	ENGINEERED MUTATION	UNP P03023

- Molecule 2 is a DNA chain called 5'-D(*GP*AP*AP*TP*TP*GP*TP*AP*AP*GP *CP*GP*CP*TP*TP*AP*CP*AP*AP*TP*TP*C)-3'.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	C	22	699	216	251	81	130	21	0
2	D	22	699	216	251	81	130	21	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: LACTOSE OPERON REPRESSOR

Chain A:

• Molecule 1: LACTOSE OPERON REPRESSOR

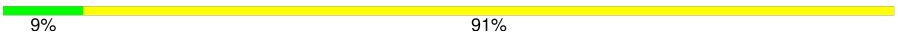
Chain B:

• Molecule 2: 5'-D(*GP*AP*AP*TP*TP*GP*TP*AP*AP*GP *CP*GP*CP*TP*TP*AP*CP*AP*AP*TP*TP*C)-3'

Chain C:

• Molecule 2: 5'-D(*GP*AP*AP*TP*TP*GP*TP*AP*AP*GP *CP*GP*CP*TP*TP*AP*CP*AP*AP*TP*TP*C)-3'

Chain D:

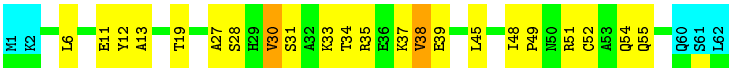


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 1. Colouring as in section 4.1 above.

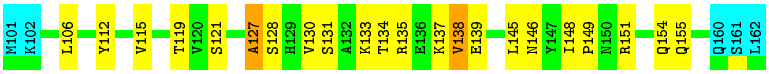
• Molecule 1: LACTOSE OPERON REPRESSOR

Chain A: 56% 32% • 8%



• Molecule 1: LACTOSE OPERON REPRESSOR

Chain B: 56% 32% • 8%



• Molecule 2: 5'-D(*GP*AP*AP*TP*TP*GP*TP*AP*AP*GP *CP*GP*CP*TP*TP*AP*CP*AP*AP*TP*TP*C)-3'

Chain C: 14% 86%

• Molecule 2: 5'-D(*GP*AP*AP*TP*TP*GP*TP*AP*AP*GP *CP*GP*CP*TP*TP*AP*CP*AP*AP*TP*TP*C)-3'

Chain D: 9% 91%

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *HADDOCK*.

Of the 50 calculated structures, 16 were deposited, based on the following criterion: *RMSD TO AVERAGE STRUCTURE AND LOWEST ENERGY*.

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

Software name	Classification	Version
CNS	refinement	
CNS	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 7354
Number of chemical shift lists	1
Total number of shifts	1752
Number of shifts mapped to atoms	1752
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	72%

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)

There are no covalent bond-length or bond-angle outliers.

There are no bond-length outliers.

There are no bond-angle outliers.

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	427	337	374	10±2
1	B	427	337	374	9±3
2	C	448	251	251	29±5
2	D	448	251	251	31±5
All	All	28000	18816	20000	1229

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 26.

5 of 232 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:D:115:DT:H2"	2:D:116:DA:O5'	0.83	1.74	11	6
2:C:15:DT:H2"	2:C:16:DA:O5'	0.79	1.76	11	6
2:C:15:DT:H2"	2:C:16:DA:N7	0.77	1.95	4	3
2:D:115:DT:H1'	2:D:116:DA:N7	0.76	1.96	14	3
2:C:15:DT:H1'	2:C:16:DA:N7	0.74	1.98	14	3

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	63/62 (102%)	55±1 (88±2%)	5±1 (8±2%)	3±1 (5±2%)	5 28
1	B	63/62 (102%)	55±1 (87±2%)	4±2 (7±3%)	4±1 (6±2%)	4 23
All	All	2016/1984 (102%)	1766 (88%)	146 (7%)	104 (5%)	5 25

5 of 20 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	27	ALA	15
1	B	127	ALA	14
1	B	128	SER	11
1	A	28	SER	10
1	B	146[1]	ASN	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	51/50 (102%)	39±2 (75±4%)	13±2 (25±4%)	3 27
1	B	51/50 (102%)	39±2 (76±5%)	12±2 (24±5%)	3 27
All	All	1632/1600 (102%)	1233 (76%)	399 (24%)	3 27

5 of 46 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	B	135[2]	ARG	16
1	B	106	LEU	16

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	6	LEU	16
1	B	135[1]	ARG	16
1	B	119	THR	16

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 72% for the well-defined parts and 71% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 7354

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	1752
Number of shifts mapped to atoms	1752
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	120	-0.47 ± 0.16	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	116	-0.06 ± 0.15	None needed (< 0.5 ppm)
$^{13}C'$	114	-0.28 ± 0.07	None needed (< 0.5 ppm)
^{15}N	118	0.62 ± 0.29	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 72%, i.e. 1569 atoms were assigned a chemical shift out of a possible 2188. 24 out of 24 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	558/562 (99%)	224/224 (100%)	224/228 (98%)	110/110 (100%)
Sidechain	616/702 (88%)	370/406 (91%)	224/260 (86%)	22/36 (61%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	58/64 (91%)	30/32 (94%)	28/28 (100%)	0/4 (0%)
Overall	1569/2188 (72%)	961/1170 (82%)	476/816 (58%)	132/202 (65%)

7.1.4 Statistically unusual chemical shifts (i)

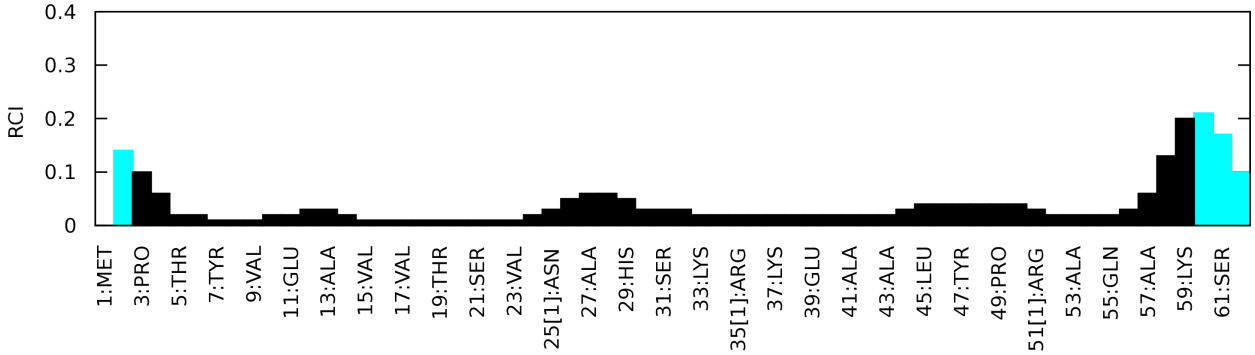
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	12	TYR	CG	175.12	141.74 – 116.84	18.4
1	B	112	TYR	CG	175.12	141.74 – 116.84	18.4
1	B	135	ARG	NH1	113.70	94.37 – 52.87	9.7
1	A	35	ARG	NH1	113.70	94.37 – 52.87	9.7
1	B	122	ARG	CB	40.08	39.81 – 21.51	5.1
1	A	22	ARG	CB	40.08	39.81 – 21.51	5.1

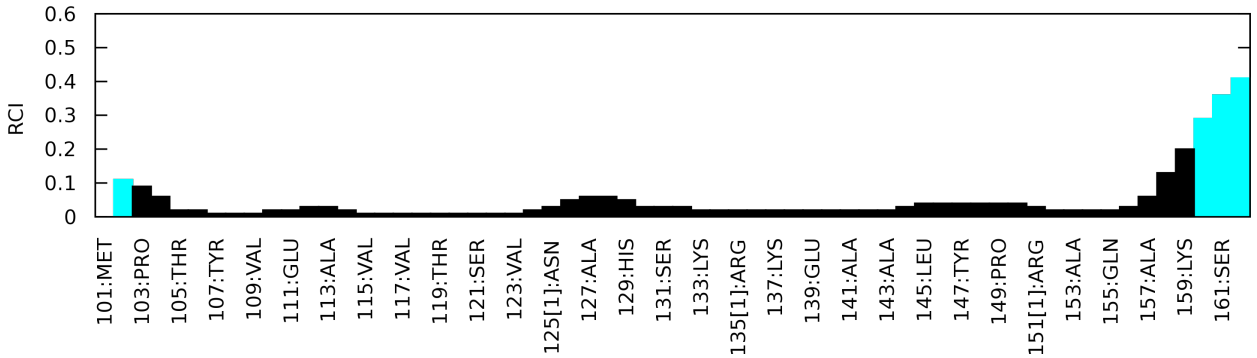
7.1.5 Random Coil Index (RCI) plots (i)

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 A:



Random coil index (RCI) for chain B:


RCI
0.6
0.5
0.4
0.3
0.2
0.1
0
101:MET
103:PRO
105:THR
107:TYR
109:VAL
111:GLU
113:ALA
115:VAL
117:VAL
119:THR
121:SER
123:VAL
125[1]:ASN
127:ALA
129:HIS
131:SER
133:LYS
135[1]:ARG
137:LYS
139:GLU
141:ALA
143:ALA
145:LEU
147:TYR
149:PRO
151[1]:ARG
153:ALA
155:GLN
157:ALA
159:LYS
161:SER