wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 05:05 AM BST

PDB ID	:	2RMK
Title	:	Rac1/PRK1 Complex
Authors	:	Modha, R.; Campbell, L.J.; Nietlispach, D.; Buhecha, H.R.; Owen, D.; Mott, H.R.
Deposited on	:	2007-10-25

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:

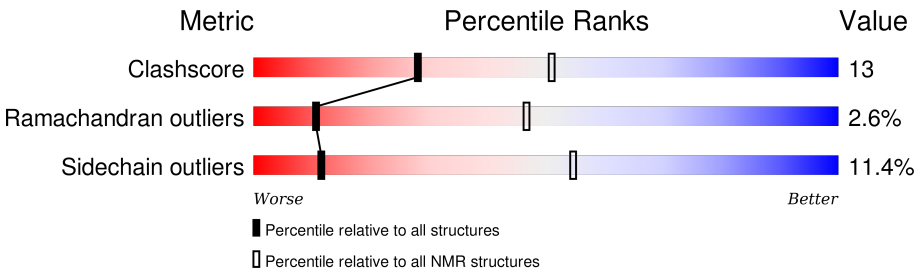
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 82%.

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	192	59% 28% 5% 8%
2	B	81	54% 17% 28%

2 Ensemble composition and analysis (i)

This entry contains 25 models. Model 14 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *closest to the average*.

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:1-A:29, A:35-A:36, A:40-A:184, B:127-B:153, B:159-B:189 (234)	0.51	14

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 and 1 single-model cluster was found.

Cluster number	Models
1	1, 2, 3, 4, 12, 13, 14, 16, 17, 18, 20
2	6, 7, 8, 10, 11, 15, 19, 22, 23, 24, 25
3	5, 9
Single-model clusters	21

3 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 4417 atoms, of which 2248 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Ras-related C3 botulinum toxin substrate 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	192	3066	970	1561	252	273	10	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	61	LEU	GLN	ENGINEERED	UNP P63000

- Molecule 2 is a protein called Serine/threonine-protein kinase N1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	81	1298	387	669	117	122	3	0

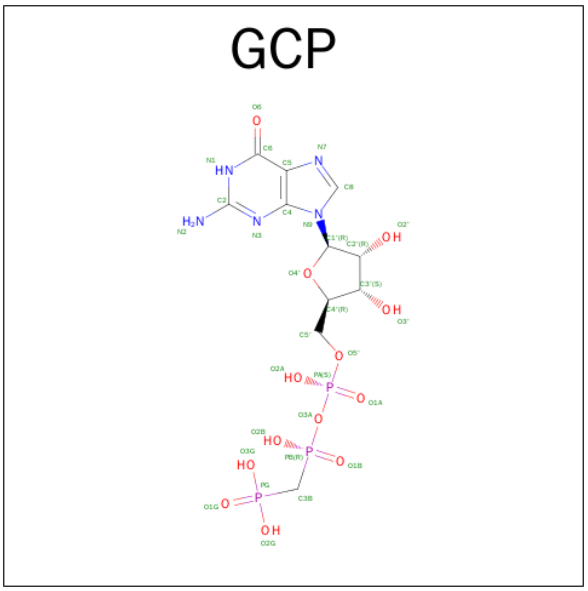
There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	119	GLY	-	EXPRESSION TAG	UNP Q16512
B	120	ILE	-	EXPRESSION TAG	UNP Q16512
B	121	PRO	-	EXPRESSION TAG	UNP Q16512

- Molecule 3 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms	
			Total	Mg
3	A	1	1	1

- Molecule 4 is PHOSPHOMETHYLPHOSPHONIC ACID GUANYLATE ESTER (three-letter code: GCP) (formula: $C_{11}H_{18}N_5O_{13}P_3$).



Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	P
4	A	1	46	11	14	5	13	3

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		
			Total	H	O
5	A	2	6	4	2

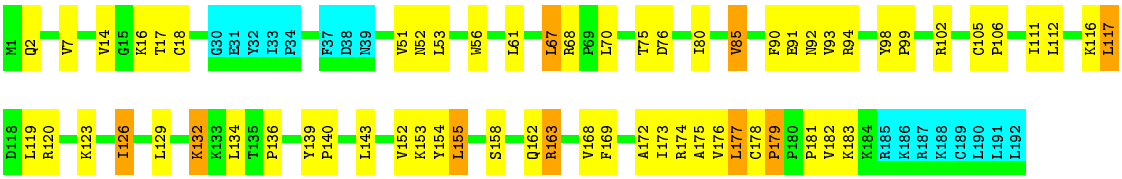
4 Residue-property plots (i)

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.

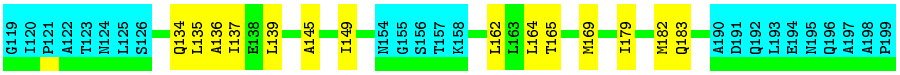
• Molecule 1: Ras-related C3 botulinum toxin substrate 1

Chain A: 59% 28% 5% 8%



• Molecule 2: Serine/threonine-protein kinase N1

Chain B: 54% 17% 28%

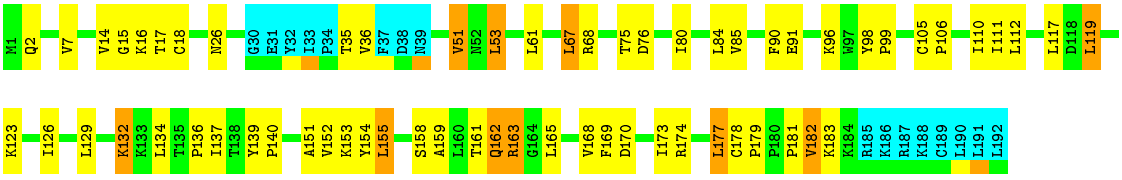


4.2 Residue scores for the representative (medoid) model from the NMR ensemble

The representative model is number 14. Colouring as in section 4.1 above.

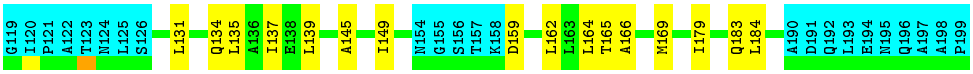
• Molecule 1: Ras-related C3 botulinum toxin substrate 1

Chain A: 59% 28% 5% 8%



• Molecule 2: Serine/threonine-protein kinase N1

Chain B: 52% 20% 28%


G119
I120
P121
A122
T123
N124
L125
S126
L131
Q134
L135
A136
I137
E138
L139
A145
I149
N154
G155
S156
T157
K158
D159
L162
L163
L164
T165
A166
M169
I179
Q183
L184
A190
D191
Q192
L193
E194
N195
Q196
A197
A198
P199

5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 25 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
CNS	structure solution	1.1
CNS	refinement	1.1

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 11010
Number of chemical shift lists	1
Total number of shifts	2881
Number of shifts mapped to atoms	2881
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	82%

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: GCP, MG

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	1367	1416	1414	41±5
2	B	468	513	513	9±2
4	A	32	14	14	3±1
All	All	46750	48675	48525	1196

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 13.

5 of 288 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:67:LEU:HD22	2:B:169:MET:SD	0.74	2.23	22	14
1:A:153:LYS:HZ3	1:A:155:LEU:HD22	0.73	1.42	21	17
1:A:153:LYS:HD2	1:A:154:TYR:N	0.70	2.01	24	1
1:A:4:ILE:HD12	1:A:176:VAL:HG21	0.70	1.61	3	4
1:A:70:LEU:HD22	2:B:169:MET:HE1	0.67	1.66	21	1

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	175/192 (91%)	153±3 (87±1%)	16±2 (9±1%)	6±2 (3±1%)	8 38
2	B	58/81 (72%)	57±1 (99±1%)	0±1 (1±1%)	0±0 (0±0%)	59 88
All	All	5825/6825 (85%)	5262 (90%)	412 (7%)	151 (3%)	11 47

5 of 24 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	181	PRO	19
1	A	68	ARG	16
1	A	183	LYS	15
1	A	179	PRO	14
1	A	178	CYS	12

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	153/168 (91%)	132±2 (86±2%)	21±2 (14±2%)	8 48
2	B	51/68 (75%)	49±1 (96±2%)	2±1 (4±2%)	45 87
All	All	5100/5900 (86%)	4519 (89%)	581 (11%)	11 54

5 of 79 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	177	LEU	25
1	A	123	LYS	25

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	132	LYS	25
1	A	152	VAL	25
1	A	126	ILE	25

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)

Of 2 ligands modelled in this entry, 1 is monoatomic - leaving 1 for Mogul analysis.

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
4	GCP	A	193	3	29,34,34	1.27±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
4	GCP	A	193	3	29,54,54	0.80±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
4	GCP	A	193	3	-	0±0,15,38,38	0±0,3,3,3

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 82% for the well-defined parts and 78% for the entire structure.

7.1 Chemical shift list 1

File name: BMRB entry 11010

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	2881
Number of shifts mapped to atoms	2881
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	5

7.1.2 Chemical shift referencing (i)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction ± precision, *ppm*	Suggested action
$^{13}C_{\alpha}$	250	-0.37 ± 0.08	None needed (< 0.5 ppm)
$^{13}C_{\beta}$	229	0.14 ± 0.07	None needed (< 0.5 ppm)
$^{13}C'$	137	0.02 ± 0.19	None needed (< 0.5 ppm)
^{15}N	224	0.49 ± 0.22	None needed (< 0.5 ppm)

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 82%, i.e. 2390 atoms were assigned a chemical shift out of a possible 2930. 42 out of 47 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^{1}H	^{13}C	^{15}N
Backbone	979/1144 (86%)	417/455 (92%)	360/468 (77%)	202/221 (91%)
Sidechain	1294/1637 (79%)	796/957 (83%)	480/605 (79%)	18/75 (24%)

Continued on next page...

Continued from previous page...

	Total	^{1}H	^{13}C	^{15}N
Aromatic	117/149 (79%)	60/77 (78%)	55/66 (83%)	2/6 (33%)
Overall	2390/2930 (82%)	1273/1489 (85%)	895/1139 (79%)	222/302 (74%)

7.1.4 Statistically unusual chemical shifts (i)

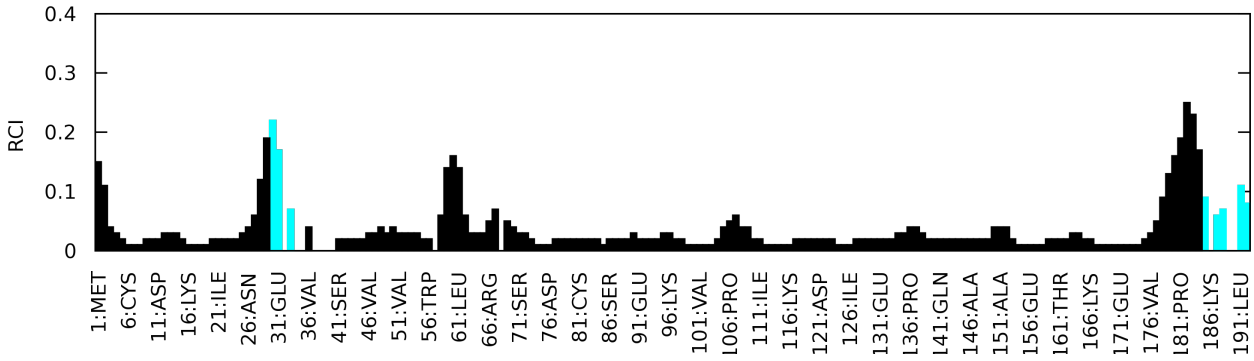
The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

Mol	Chain	Res	Type	Atom	Shift, *ppm*	Expected range, *ppm*	Z-score
1	A	92	ASN	HD22	10.27	9.59 – 4.69	6.4
1	A	96	LYS	HG2	-0.28	2.67 – 0.07	-6.3
1	A	75	THR	HG21	-0.03	2.29 – -0.01	-5.1
1	A	75	THR	HG23	-0.03	2.29 – -0.01	-5.1
1	A	75	THR	HG22	-0.03	2.29 – -0.01	-5.1

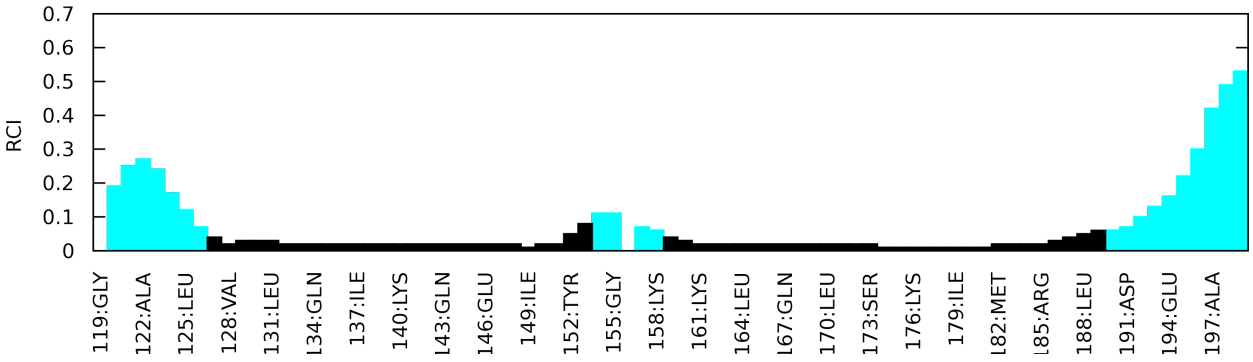
7.1.5 Random Coil Index (RCI) plots (i)

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
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
119:GLY
122:ALA
125:LEU
128:VAL
131:LEU
134:GLN
137:ILE
140:LYS
143:GLN
146:GLU
149:ILE
152:TYR
155:GLY
158:LYS
161:LYS
164:LEU
167:GLN
170:LEU
173:SER
176:LYS
179:ILE
182:MET
185:ARG
188:LEU
191:ASP
194:GLU
197:ALA