wwPDB NMR Structure Validation Summary Report (i)

Feb 8, 2022 – 02:21 PM EST

PDB ID	:	1CF4
Title	:	CDC42/ACK GTPASE-BINDING DOMAIN COMPLEX
Authors	:	Mott, H.R.; Owen, D.; Nietlispach, D.; Lowe, P.N.; Lim, L.; Laue, E.D.
Deposited on	:	1999-03-23

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 (i) symbol.

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

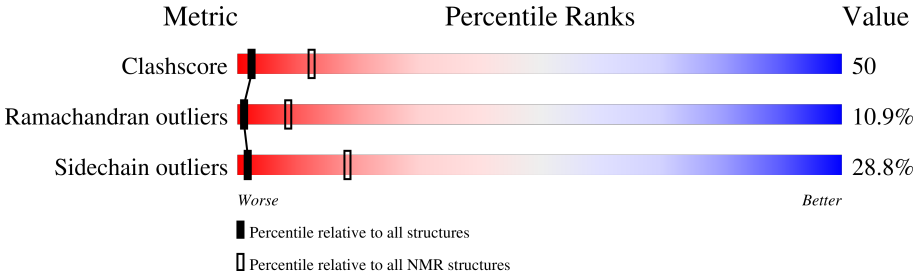
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	2.26
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.26

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 was not calculated.

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	184	28% 58% 13% •
2	B	44	23% 68% 9%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 19 is the overall representative, medoid model (most similar to other models).

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:184, B:502-B:545 (228)	0.85	19

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

3 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 3602 atoms, of which 1789 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called PROTEIN (CDC42 HOMOLOG).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	184	2899	926	1463	228	275	7	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	61	LEU	GLN	conflict	UNP P60953
A	?	-	LYS	deletion	UNP P60953
A	?	-	LYS	deletion	UNP P60953
A	?	-	SER	deletion	UNP P60953
A	?	-	ARG	deletion	UNP P60953
A	?	-	ARG	deletion	UNP P60953
A	?	-	CYS	deletion	UNP P60953
A	?	-	VAL	deletion	UNP P60953

- Molecule 2 is a protein called PROTEIN (ACTIVATED P21CDC42HS KINASE).

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	44	651	210	309	62	69	1	0

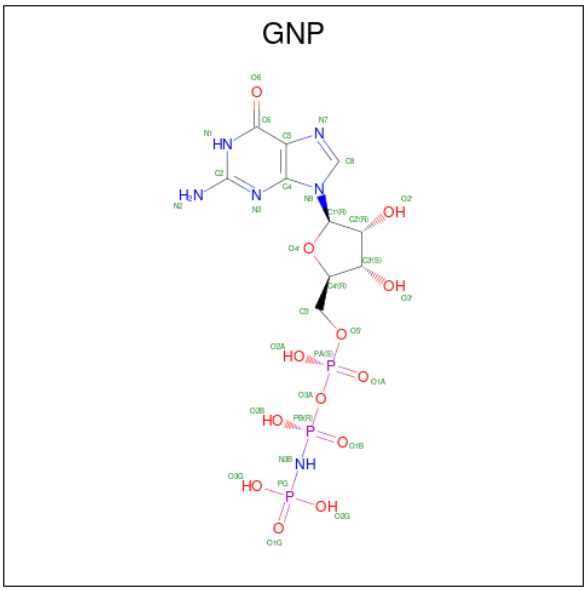
There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	502	GLY	VAL	conflict	UNP Q07912
B	503	SER	ALA	conflict	UNP Q07912

- 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 PHOSPHOAMINOPHOSPHONIC ACID-GUANYLATE ESTER (three-letter code: GNP) (formula: $C_{10}H_{17}N_6O_{13}P_3$).



Mol	Chain	Residues	Atoms					
			Total	C	H	N	O	P
4	A	1	45	10	13	6	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, 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: PROTEIN (CDC42 HOMOLOG)

Chain A: 28% 58% 13%

M1 Q2 T3 I4 K5 C6 V7 V8 D11 G12 A13 V14 G15 K16 T17 C18 L19 L20 I21 T24 T25 N26 K27 F28 P29 S30 E31 Y32 V33 P34 T35 V36 F37 D38 N39 Y40 A41 V42 T43 V44 M45 I46 G47 G48 E49 P50 Y51 T52 L53 G54 L55 F56 D57 T58 A59 G60 L61 E62

D63 Y64 D65 R66 L67 R68 P69 L70 S71 Y72 T75 D76 V77 F78 L79 V80 C81 F82 S83 V84 V85 S86 P87 S88 S89 F90 E91 N92 V93 K94 E95 K96 W97 V98 P99 E100 I101 C105 P106 K107 T108 P109 F110 L111 L112 V113 G114 T115 Q116 I117 D118 L119 R120 T125 T126 E127 K128

L129 A130 K131 N132 K133 Q134 K135 P136 I137 A142 L145 L149 K150 A151 V152 K153 Y154 V155 E156 C157 S158 A159 L160 T161 Q162 K163 G164 L165 K166 N167 V168 F169 D170 E171 A172 I173 L174 A175 A176 L177 E178 K184

• Molecule 2: PROTEIN (ACTIVATED P21CDC42HS KINASE)

Chain B: 23% 68% 9%

G502 S503 G504 L505 S506 A507 Q508 D509 I510 S511 Q512 P513 L514 Q515 N516 S517 F518 I519 H520 T521 G522 H523 G524 D525 S526 D527 P528 R529 H530 C531 W532 D536 R537 I538 D539 E540 L541 Y542 L543 G544 N545

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

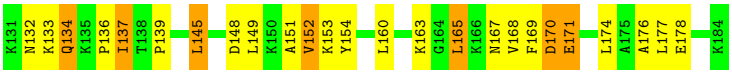
The representative model is number 19. Colouring as in section 4.1 above.

• Molecule 1: PROTEIN (CDC42 HOMOLOG)

Chain A: 33% 49% 16%

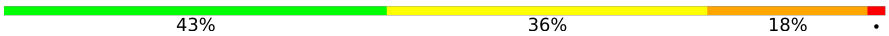
M1 Q2 T3 I4 K5 C6 V7 V8 V9 V14 G15 K16 T17 C18 L19 L20 I21 S22 T25 N26 K27 F28 P29 S30 E31 Y32 V33 P34 T35 V36 F37 D38 N39 Y40 A41 V42 T43 V44 M45 I46 G47 G48 E49 P50 Y51 T52 L53 G54 L55 F56 D57 T58 A59 G60 L61 E62 D63 Y64

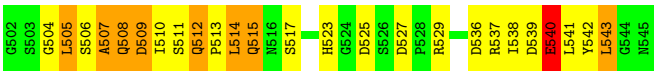
D65 R66 L67 R68 P69 L70 S71 Y72 P73 Q74 T75 F78 L79 V80 C81 F82 S83 V84 V85 S86 P87 S88 S89 F90 E91 N92 V93 K94 E95 K96 W97 V98 P99 E100 I101 K107 T108 P109 F110 L111 L112 V113 G114 T115 Q116 I117 D118 L119 R120 P123 S124 T125 I126 E127 K128 L129 A130



• Molecule 2: PROTEIN (ACTIVATED P21CDC42HS KINASE)

Chain B:





5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *DISTANCE GEOMETRY, SIMULATED ANNEALING*.

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

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

Software name	Classification	Version
X-PLOR	refinement	3.851
ANSIG	structure calculation	
XPLOR	structure calculation	3.851

No chemical shift data was provided.

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: MG, GNP

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	1436	1463	1461	166±12
2	B	342	309	303	38±8
4	A	32	13	13	12±3
5	A	2	4	0	0±0
All	All	36260	35780	35540	3580

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 50.

5 of 1814 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:44:VAL:HG13	1:A:52:THR:HG21	1.03	1.26	12	1
1:A:113:VAL:HG13	1:A:168:VAL:HG21	1.01	1.33	20	12
1:A:8:VAL:HG11	1:A:14:VAL:HG13	1.01	1.20	3	1
1:A:44:VAL:HG22	1:A:52:THR:HG23	1.00	1.30	19	2
1:A:14:VAL:HG23	1:A:19:LEU:HD21	1.00	1.27	4	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	182/184 (99%)	127±4 (70±2%)	40±5 (22±3%)	16±3 (9±2%)	2 13
2	B	42/44 (95%)	19±2 (45±5%)	14±2 (35±6%)	9±2 (21±5%)	0 2
All	All	4480/4560 (98%)	2907 (65%)	1086 (24%)	487 (11%)	1 8

5 of 96 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	152	VAL	20
1	A	47	GLY	18
1	A	134	GLN	18
1	A	33	VAL	17
1	A	50	PRO	17

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	164/164 (100%)	118±5 (72±3%)	46±5 (28±3%)	2 19
2	B	37/37 (100%)	25±3 (68±9%)	12±3 (32±9%)	1 13
All	All	4020/4020 (100%)	2861 (71%)	1159 (29%)	2 18

5 of 168 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	117	ILE	20
1	A	126	ILE	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	70	LEU	19
1	A	66	ARG	17
1	A	88	SER	17

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 monosaccharides 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	GNP	A	185	3	28,34,34	2.34±0.02	7±0 (26±1%)

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	GNP	A	185	3	30,54,54	2.26±0.01	11±0 (36±1%)

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	GNP	A	185	3	-	0±0,17,38,38	0±0,3,3,3

5 of 8 unique bond 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(Å)	Ideal(Å)	Models	
								Worst	Total
4	A	185	GNP	C4-N9	7.27	1.38	1.47	5	20
4	A	185	GNP	C5-C6	5.36	1.43	1.52	8	20
4	A	185	GNP	C6-N1	4.81	1.41	1.33	14	20
4	A	185	GNP	PG-O2G	3.22	1.48	1.56	17	20
4	A	185	GNP	PB-O2B	3.18	1.48	1.56	3	20

5 of 11 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
4	A	185	GNP	C4-C5-N7	6.16	110.62	102.46	19	20
4	A	185	GNP	C5-C6-N1	4.99	112.04	118.19	20	20
4	A	185	GNP	O6-C6-C5	4.10	128.22	119.86	19	20
4	A	185	GNP	O2B-PB-O1B	3.89	118.09	109.92	7	20
4	A	185	GNP	O1B-PB-N3B	3.35	106.83	111.77	16	20

There are no chirality outliers.

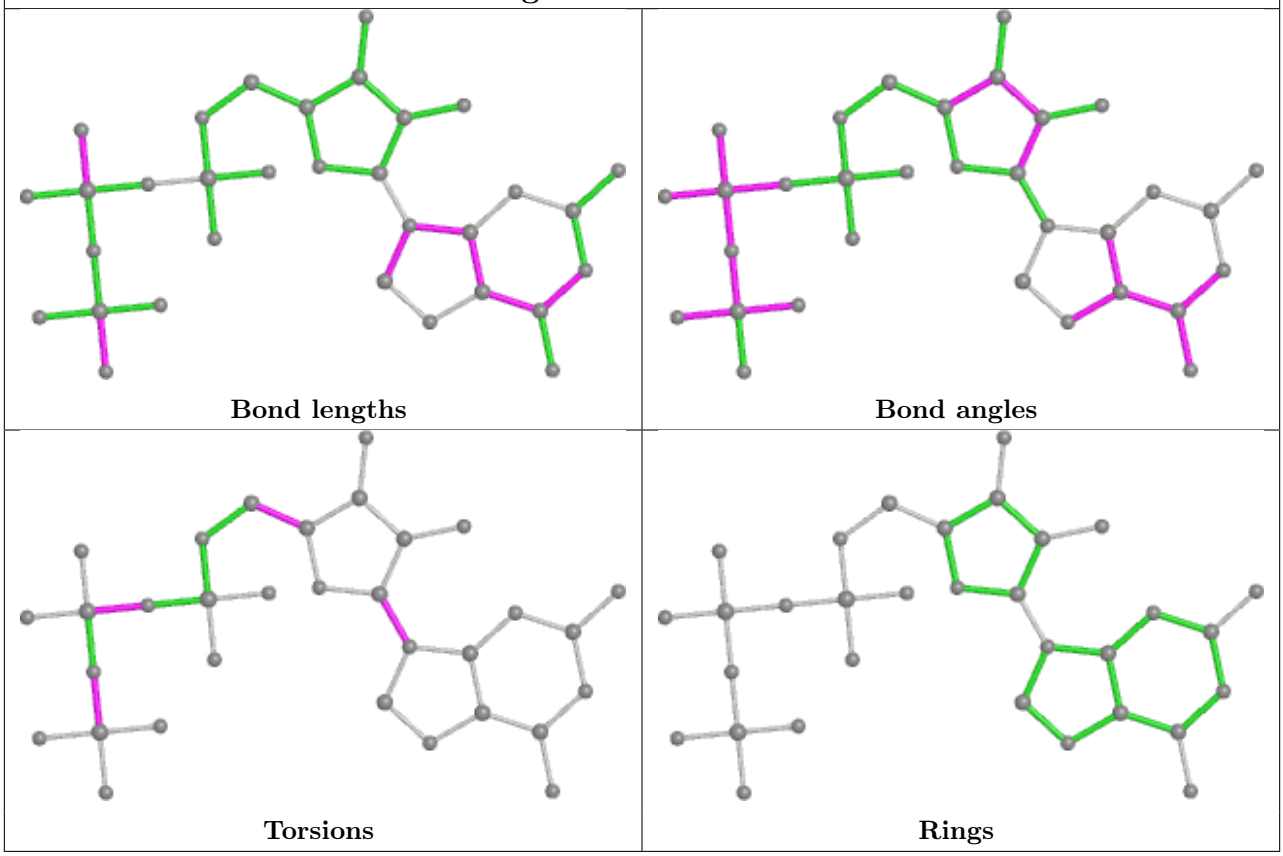
All unique torsion outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Atoms	Models (Total)
4	A	185	GNP	PB-N3B-PG-O1G	3
4	A	185	GNP	PG-N3B-PB-O1B	1

There are no ring outliers.

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

Ligand GNP A 185



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)

No chemical shift data were provided