wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 10:40 PM BST

PDB ID	:	2K7F
Title	:	HADDOCK calculated model of the complex between the BRCT region of RFC p140 and dsDNA
Authors	:	Kobayashi, M.; Ab, E.; Bonvin, A.; Siegal, G.
Deposited on	:	2008-08-10

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:

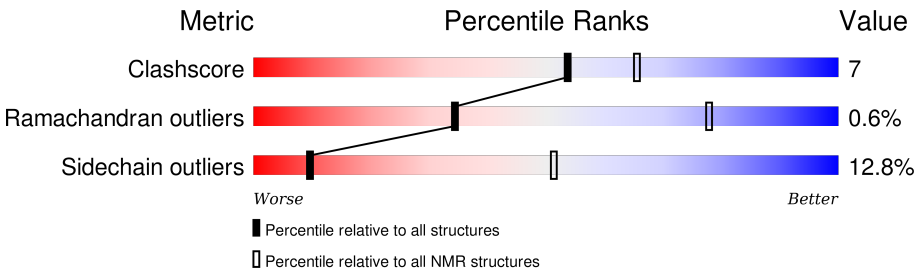
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	unknown
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	rb-20027457
Ideal 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 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	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	109	64% 17% 19%
2	B	14	29% 71%
3	C	10	50% 50%

2 Ensemble composition and analysis (i)

This entry contains 4 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: *fewest violations*.

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:391-A:393, A:399-A:483 (88)	0.29	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 1 clusters and 1 single-model cluster was found.

Cluster number	Models
1	1, 3, 4
Single-model clusters	2

3 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2471 atoms, of which 1138 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Replication factor C subunit 1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	109	1708	522	867	150	167	2	0

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	481	ASN	-	EXPRESSION TAG	UNP P35251
A	482	LEU	-	EXPRESSION TAG	UNP P35251
A	483	GLU	-	EXPRESSION TAG	UNP P35251

- Molecule 2 is a DNA chain called 5'-D(P*DCP*DGP*DAP*DCP*DCP*DTP*DCP*DGP*DAP*DGP*DAP*DTP*DCP*DA)-3'.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	14	442	135	157	54	82	14	0

- Molecule 3 is a DNA chain called 5'-D(P*DCP*DTP*DCP*DGP*DAP*DGP*DGP*DTP*DCP*DG)-3'.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
3	C	10	321	97	114	38	62	10	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: Replication factor C subunit 1

Chain A: 64% 17% 19%

K375 R376 T377 N378 Y379 Q380 A381 Y382 R383 S384 Y385 L386 N387 R388 E389 G390 L394 G395 S396 K397 E398 L407 L410 I411 F412 I421 E422 K427 S428 L429 V437 T438 S442 G455 Q456 S457 K458 S459 K468 L476 I479 E483

• Molecule 2: 5'-D(P*DCP*DGP*DAP*DCP*DCP*DTP*DCP*DGP*DAP*DGP*DAP*DTP*DCP*DA)-3'

Chain B: 29% 71%

C1 G2 A3 C4 C5 T6 C7 G8 A9 G10 A11 T12 C13 A14

• Molecule 3: 5'-D(P*DCP*DTP*DCP*DGP*DAP*DGP*DGP*DTP*DCP*DG)-3'

Chain C: 50% 50%

C19 T20 C21 G22 A23 G28

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.

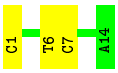
• Molecule 1: Replication factor C subunit 1

Chain A: 62% 17% • 19%

K375 R376 T377 N378 Y379 Q380 A381 Y382 R383 S384 Y385 L386 N387 R388 E389 G390 L394 G395 S396 K397 E398 L407 L410 I411 F412 L418 I421 E422 K427 S428 L429 V437 S442 V449 G455 Q456 S457 K458 S459 D460 K461 L465 L476 E483

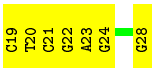
• Molecule 2: 5'-D(P*DCP*DGP*DAP*DCP*DCP*DTP*DCP*DGP*DAP*DGP*DAP*DTP*DCP*DA)-3'

Chain B:



• Molecule 3: 5'-D(P*DCP*DTP*DCP*DGP*DAP*DGP*DGP*DTP*DCP*DG)-3'

Chain C:



5 Refinement protocol and experimental data overview (i)

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

Of the 200 calculated structures, 4 were deposited, based on the following criterion: *back calculated data agree with experimental NOESY spectrum*.

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

Software name	Classification	Version
HADDOCK	structure solution	2.0
HADDOCK	refinement	2.0

No chemical shift data was provided. 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	663	691	691	4±2
2	B	285	157	157	6±3
3	C	207	114	113	6±1
All	All	4620	3848	3844	57

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 7.

5 of 49 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
2:B:1:DC:O2	3:C:28:DG:N2	0.98	1.97	1	4
3:C:19:DC:H2'	3:C:19:DC:OP3	0.71	1.84	1	1
1:A:427:LYS:HG3	1:A:437:VAL:HG11	0.66	1.67	3	2
3:C:19:DC:OP3	3:C:19:DC:C6	0.59	2.55	3	1
3:C:21:DC:H2"	3:C:22:DG:O5'	0.57	1.98	4	2

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	87/109 (80%)	77±1 (89±1%)	9±1 (11±2%)	1±1 (1±1%)	34 78
All	All	348/436 (80%)	309 (89%)	37 (11%)	2 (1%)	34 78

All 2 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	455	GLY	1
1	A	399	ILE	1

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	72/90 (80%)	63±1 (87±2%)	9±1 (13±2%)	10 51
All	All	288/360 (80%)	251 (87%)	37 (13%)	10 51

5 of 20 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	421	ILE	4
1	A	428	SER	3
1	A	468	LYS	3
1	A	457	SER	3
1	A	438	THR	2

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)

No chemical shift data were provided