wwPDB NMR Structure Validation Summary Report (i)

Apr 26, 2016 – 03:37 PM BST

PDB ID	:	1K2M
Title	:	Solution Structure of the FHA2 Domain of Rad53 Complexed with a Phosphotyrosyl Peptide Derived from Rad9
Authors	:	Byeon, I.-J.L.; Yongkiettrakul, S.; Tsai, M.-D.
Deposited on	:	2001-09-28

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:

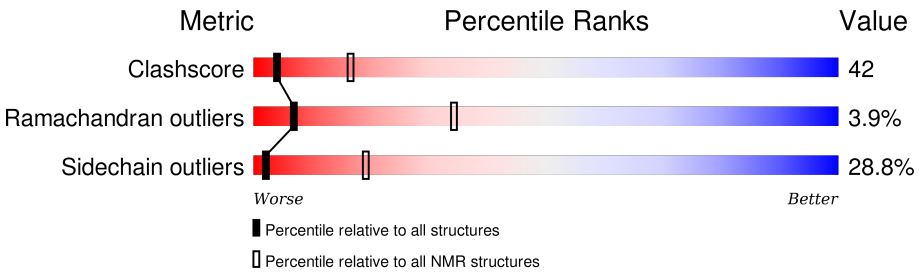
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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 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	158	24% 50% 11% • 15%
2	P	7	100%

2 Ensemble composition and analysis (i)

This entry contains 22 models. Model 11 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:575-A:630, A:643-A:705, A:715-A:730 (135)	0.26	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 5 clusters and 2 single-model clusters were found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 2676 atoms, of which 1332 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Protein Kinase SPK1.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	158	2551	806	1277	222	239	7	0

- Molecule 2 is a protein called DNA repair protein Rad9.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	P	7	125	43	55	7	19	1	0

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
P	829	PTR	TYR	MODIFIED RESIDUE	UNP P14737

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: Protein Kinase SPK1

Chain A: 24% 50% 11% • 15%

G573 N574 G575 R576 F577 L578 T579 L580 K581 P582 L583 P584 D585 S586 I587 I588 Q589 E590 S591 L592 E593 I594 Q595 Q596 P600 F601 F602 I603 G604 R605 S606 N610 C611 K612 I613 E614 D615 N616 R617 L618 S619 R620 V621 H622 C623 F624 I625 F626 K627 K628 R629 H630 A631 V632 G633 K634 S635 M636

Y637 E638 S639 P640 A641 Q642 G643 L644 D645 D646 I647 W648 Y649 C650 H651 T652 G653 T654 N655 V656 S657 Y658 L659 N660 N661 N662 R663 M664 I665 Q666 F670 L671 L672 Q673 D674 G675 D676 E677 I678 K679 I680 I681 W682 D683 K687 F688 V689 I690 G691 F692 K693 V694 E695 I696 N697 D698 T699

L702 F703 N704 E705 G706 L707 G708 N709 L710 Q711 E712 Q713 R714 V715 V716 L717 K718 E723 K724 V727 K728 K729 L730

- Molecule 2: DNA repair protein Rad9

Chain P: 100%

E826 D827 I828 Y829 Y830 L831 D832

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: Protein Kinase SPK1

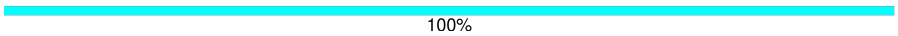
Chain A: 30% 42% 13% • 15%

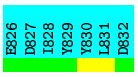
G573 N574 F577 L578 T579 L580 K581 P582 L583 I587 I588 Q589 E590 S591 L592 E593 I594 Q595 Q596 G597 V598 N599 P600 F601 F602 I603 S606 N610 C611 K612 I613 E614 D615 N616 R617 L618 S619 R620 V621 H622 C623 F624 I625 F626 K627 K628 R629 H630 A631 V632 G633 K634 S635 M636 Y637 E638

S639 P640 A641 Q642 G643 L644 D645 D646 I647 W648 Y649 C650 H651 T652 G653 T654 N655 V656 S657 Y658 L659 N660 R663 M664 I665 Q666 K669 F670 L671 L672 Q673 D674 G675 D676 E677 I678 K679 I680 I681 W682 D683 N686 K687 F688 V689 I690 G691 F692 K693 V694 E695 I696 N697 D698 L702 F703

- Molecule 2: DNA repair protein Rad9

Chain P:





5 Refinement protocol and experimental data overview (i)

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

Of the 100 calculated structures, 22 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
X-PLOR	structure solution	3.851
X-PLOR	refinement	3.851

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)

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

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	1105	1112	1109	94±8
2	P	0	0	0	0±0
All	All	24310	24464	24398	2068

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 42.

5 of 610 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:580:LEU:HD22	1:A:694:VAL:HG22	1.04	1.25	20	1
1:A:618:LEU:HD12	1:A:681:ILE:HD12	1.02	1.30	5	9
1:A:592:LEU:HD21	1:A:613:ILE:HD11	1.01	1.32	19	7
1:A:578:LEU:HD12	1:A:647:ILE:HD13	1.00	1.32	9	12
1:A:588:ILE:HD12	1:A:690:ILE:HD11	1.00	1.31	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	134/158 (85%)	117±1 (87±1%)	12±2 (9±1%)	5±1 (4±1%)	7 34
2	P	0	-	-	-	-
All	All	2948/3630 (81%)	2564 (87%)	268 (9%)	116 (4%)	7 34

5 of 12 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	660	ASN	22
1	A	656	VAL	22
1	A	582	PRO	21
1	A	705	GLU	18
1	A	704	ASN	8

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	125/142 (88%)	89±4 (71±3%)	36±4 (29±3%)	2 19
2	P	0	-	-	-
All	All	2750/3256 (84%)	1958 (71%)	792 (29%)	2 19

5 of 98 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	628	LYS	22
1	A	659	LEU	22
1	A	596	GLN	22

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	589	GLN	22
1	A	583	LEU	22

6.3.3 RNA (i)

There are no RNA molecules in this entry.

6.4 Non-standard residues in protein, DNA, RNA chains (i)

1 non-standard protein/DNA/RNA residue 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	PTR	P	829	2	13,16,17	0.61±0.01	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	PTR	P	829	2	19,22,24	0.90±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	PTR	P	829	2	-	0±0,9,11,13	0±0,1,1,1

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.

6.5 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