wwPDB NMR Structure Validation Summary Report (i)

Apr 27, 2016 – 05:50 AM BST

PDB ID	:	2YU7
Title	:	Solution structure of the SHP-1 C-terminal SH2 domain complexed with a tyrosine-phosphorylated peptide from NKG2A
Authors	:	Kasai, T.; Koshiba, S.; Inoue, M.; Kigawa, T.; Yokoyama, S.; RIKEN Structural Genomics/Proteomics Initiative (RSGI)
Deposited on	:	2007-04-05

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:

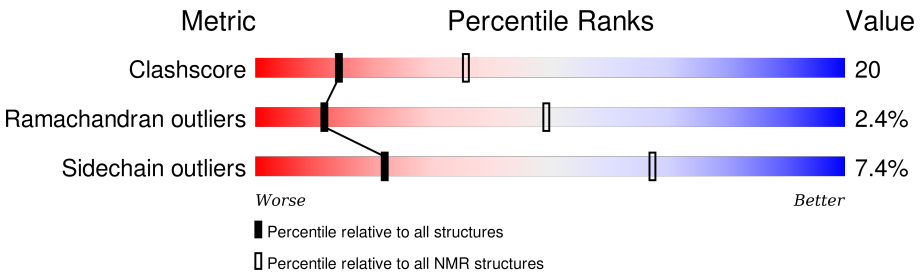
Cyrange	:	Kirchner and Güntert (2011)
NmrClust	:	Kelley et al. (1996)
MolProbity	:	4.02b-467
Mogul	:	1.7.1 (RC1), CSD as537be (2016)
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
ShiftChecker	:	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	118	48% 29% • 21%
2	B	15	13% 13% 73%

2 Ensemble composition and analysis (i)

This entry contains 20 models. Model 3 is the overall representative, medoid model (most similar to other models). The authors have identified model 1 as representative, based on the following criterion: *lowest energy*.

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:8-A:36, A:42-A:53, A:58-A:69, A:73-A:112, B:9-B:12 (97)	0.24	3

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 and 1 single-model cluster was found.

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

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1985 atoms, of which 971 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Tyrosine-protein phosphatase non-receptor type 6.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	118	1741	561	853	150	174	3	0

There are 13 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	GLY	-	EXPRESSION TAG	UNP P29350
A	2	SER	-	EXPRESSION TAG	UNP P29350
A	3	SER	-	EXPRESSION TAG	UNP P29350
A	4	GLY	-	EXPRESSION TAG	UNP P29350
A	5	SER	-	EXPRESSION TAG	UNP P29350
A	6	SER	-	EXPRESSION TAG	UNP P29350
A	7	GLY	-	EXPRESSION TAG	UNP P29350
A	113	SER	-	EXPRESSION TAG	UNP P29350
A	114	GLY	-	EXPRESSION TAG	UNP P29350
A	115	PRO	-	EXPRESSION TAG	UNP P29350
A	116	SER	-	EXPRESSION TAG	UNP P29350
A	117	SER	-	EXPRESSION TAG	UNP P29350
A	118	GLY	-	EXPRESSION TAG	UNP P29350

- Molecule 2 is a protein called natural killer group 2A.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	P	
2	B	15	244	76	118	19	30	1	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: Tyrosine-protein phosphatase non-receptor type 6

Chain A: 48% 29% • 21%

G1 S2 S3 G4 S5 S6 G7 W8 G11 S14 L21 A24 P28 F31 L32 V33 R34 L37 S38 Q39 P40 G41 D42 F43 V44 L45 S46 V47 D50 A54 G55 P56 G57 V62 T63 H64 I65 K66 E70 G71 G72 R73 Y74 T75 V76 G77 G78 L79 E80

T81 F82 L85 L88 V89 F92 G96 V105 Y106 L107 R108 Q109 P110 Y111 Y112 S113 G114 P115 S116 S117 G118

- Molecule 2: natural killer group 2A

Chain B: 13% 13% 73%

A1 T2 E3 Q4 E5 I6 T7 Y8 A9 E10 L11 N12 L13 Q14 K15

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

The representative model is number 3. Colouring as in section 4.1 above.

- Molecule 1: Tyrosine-protein phosphatase non-receptor type 6

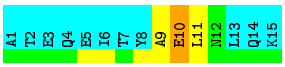
Chain A: 47% 31% • 21%

G1 S2 S3 G4 S5 S6 G7 H12 L21 A24 P28 W29 T30 F31 L32 V33 R34 E35 S36 L37 S38 Q39 P40 G41 D42 F43 V44 L45 S46 V47 D50 A54 G55 P56 G57 V62 T63 H64 I65 K66 E70 G71 G72 V76 L79 E80 T81 F82 T86

D87 L88 V89 E90 H91 F92 G96 A100 V105 Y106 L107 R108 Q109 P110 Y111 Y112 S113 G114 P115 S116 S117 G118

- Molecule 2: natural killer group 2A

Chain B: 7% 13% 7% 73%


A1
T2
E3
Q4
E5
I6
T7
Y8
A9
E10
L11
N12
L13
Q14
K15

5 Refinement protocol and experimental data overview (i)

The models were refined using the following method: *torsion angle dynamics, restrainted molecular dynamics.*

Of the 100 calculated structures, 20 were deposited, based on the following criterion: *structures with the least restraint violations, structures with the lowest energy, target function.*

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

Software name	Classification	Version
CYANA	structure solution	2.0.17
CYANA	refinement	2.0.17

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	748	732	728	31±6
2	B	30	28	28	3±2
All	All	15560	15200	15120	626

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 20.

5 of 201 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:22:LEU:HD11	1:A:32:LEU:HD23	0.91	1.40	14	1
1:A:100:ALA:HB2	2:B:10:GLU:O	0.90	1.65	5	5
1:A:31:PHE:CB	1:A:107:LEU:HD23	0.86	2.01	3	14
1:A:34:ARG:NH2	1:A:44:VAL:HG11	0.81	1.91	9	11
1:A:65:ILE:HG21	2:B:11:LEU:HD11	0.79	1.53	15	10

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	93/118 (79%)	81±2 (87±2%)	10±2 (11±3%)	2±1 (2±2%)	13	53
2	B	4/15 (27%)	3±1 (82±20%)	0±0 (8±11%)	0±0 (10±12%)	2	10
All	All	1940/2660 (73%)	1690 (87%)	203 (10%)	47 (2%)	12	49

5 of 11 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	80	GLU	17
2	B	11	LEU	5
1	A	11	GLY	5
1	A	103	ALA	4
1	A	42	ASP	4

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	81/95 (85%)	75±2 (93±2%)	6±2 (7±2%)	22	67
2	B	3/12 (25%)	3±0 (93±13%)	0±0 (7±13%)	25	70
All	All	1680/2140 (79%)	1556 (93%)	124 (7%)	22	67

5 of 25 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	32	LEU	20
1	A	107	LEU	20

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Models (Total)
1	A	45	LEU	18
1	A	50	ASP	11
1	A	60	LEU	7

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	B	8	2	13,16,17	0.85±0.00	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	B	8	2	19,22,24	0.67±0.00	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	B	8	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