wwPDB NMR Structure Validation Summary Report (i)

Apr 20, 2024 – 09:30 PM EDT

PDB ID	:	2KXQ
Title	:	Solution Structure of Smurf2 WW2 and WW3 bound to Smad7 PY motif containing peptide
Authors	:	Chong, A.; Lin, H.; Wrana, J.; Forman-Kay, J.D.
Deposited on	:	2010-05-11

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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

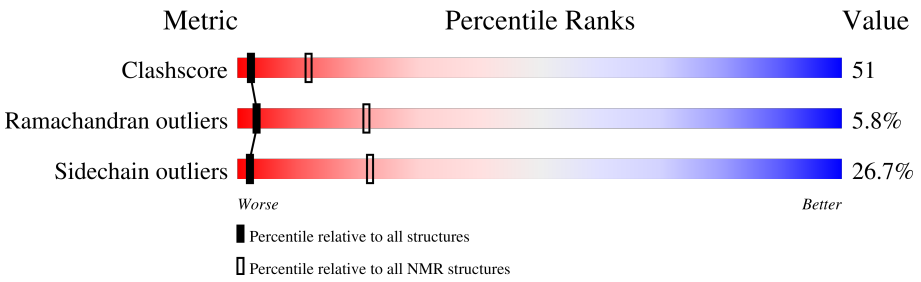
MolProbity	:	4.02b-467
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
wwPDB-RCI	:	v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV	:	Wang et al. (2010)
wwPDB-ShiftChecker	:	v1.2
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36.2

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	90	22% 51% 8% 19%
2	B	20	20% 15% 65%

2 Ensemble composition and analysis (i)

This entry contains 10 models. Model 5 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:255-A:285, A:290-A:331, B:206-B:212 (80)	0.45	5

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 5 single-model clusters were found.

Cluster number	Models
1	2, 5, 6
2	3, 4
Single-model clusters	1; 7; 8; 9; 10

3 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 1683 atoms, of which 814 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called E3 ubiquitin-protein ligase SMURF2.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
1	A	90	1386	443	671	131	140	1	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	244	GLY	-	expression tag	UNP Q9HAU4
A	245	PRO	-	expression tag	UNP Q9HAU4
A	246	LEU	-	expression tag	UNP Q9HAU4
A	247	GLY	-	expression tag	UNP Q9HAU4
A	248	GLY	-	expression tag	UNP Q9HAU4
A	249	SER	-	expression tag	UNP Q9HAU4

- Molecule 2 is a protein called Smad7 PY motif containing peptide.

Mol	Chain	Residues	Atoms						Trace
			Total	C	H	N	O	S	
2	B	20	297	98	143	23	32	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, 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: E3 ubiquitin-protein ligase SMURF2

Chain A: 22% 51% 8% 19%

G244 P245 L246 G247 G248 S249 P250 P251 D252 L253 P254 E255 G256 Y257 E258 Q259 R260 T261 T262 Q263 Q264 G265 Q266 V267 Y268 F269 L270 H271 G275 V276 S277 T278 W279 H280 D281 P282 R283 V284 P285 R286 D287 L288 S289 N290 I291 N292 C293 E294 E295 L296 L299 G302 W303 E304 I305 R306 N307

T308 A309 T310 G311 R312 V313 Y314 F315 V316 D317 N320 R321 T322 T323 Q324 F325 T326 D327 P328 R329 L330 S331 A332 N333

• Molecule 2: Smad7 PY motif containing peptide

Chain B: 20% 15% 65%

G198 P199 L200 G201 S202 E203 L204 E205 S206 P207 P208 P209 P210 Y211 S212 R213 Y214 P215 M216 D217

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

The representative model is number 5. Colouring as in section 4.1 above.

• Molecule 1: E3 ubiquitin-protein ligase SMURF2

Chain A: 28% 39% 14% 19%

G244 P245 L246 G247 G248 S249 P250 P251 D252 L253 P254 Y257 E258 Q259 R260 T261 T262 Q263 Q264 G265 Q266 V267 Y268 F269 L270 H271 T272 Q273 V276 S277 T278 W279 H280 D281 P282 R283 V284 P285 R286 D287 L288 S289 N290 I291 N292 C293 E294 E295 L296 L299 G302 W303 E304 I305 R306 N307

T308 A309 T310 G311 R312 V313 Y314 F315 V316 N320 R321 T322 T323 Q324 F325 T326 D327 P328 R329 L330 S331 A332 N333

• Molecule 2: Smad7 PY motif containing peptide

Chain B: 20% 15% 65%

G198
P199
L200
G201
S202
E203
L204
E205
S206
P207
P208
P209
P210
Y211
S212
R213
Y214
P215
M216
D217

5 Refinement protocol and experimental data overview (i)

Of the 200 calculated structures, 10 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	refinement	1.1

No chemical shift data was provided.

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	598	557	554	58±7
2	B	52	47	47	20±2
All	All	6500	6040	6010	640

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 51.

5 of 195 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:299:LEU:HD13	1:A:305:ILE:HD11	1.08	1.20	7	8
1:A:291:ILE:HG12	1:A:330:LEU:HD21	0.88	1.44	1	5
1:A:308:THR:HG23	1:A:312:ARG:O	0.85	1.72	9	6
1:A:262:THR:O	1:A:264:GLN:N	0.83	2.11	4	4
1:A:291:ILE:CG1	1:A:330:LEU:HD21	0.81	2.06	1	7

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	73/90 (81%)	65±1 (89±2%)	6±1 (9±1%)	2±1 (3±1%)	8 44
2	B	7/20 (35%)	4±0 (53±7%)	1±1 (9±9%)	3±0 (39±7%)	0 0
All	All	800/1100 (73%)	684 (86%)	70 (9%)	46 (6%)	3 21

5 of 7 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	263	GLN	10
2	B	207	PRO	10
2	B	210	PRO	10
2	B	212	SER	7
1	A	293	CYS	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	66/79 (84%)	48±3 (72±5%)	18±3 (28±5%)	2 19
2	B	7/18 (39%)	6±1 (86±11%)	1±1 (14±11%)	6 46
All	All	730/970 (75%)	535 (73%)	195 (27%)	2 22

5 of 42 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	271	HIS	10
1	A	310	THR	10
1	A	329	ARG	10
1	A	296	LEU	8
1	A	327	ASP	8

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)

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