



wwPDB NMR Structure Validation Summary Report ⓘ

Jun 23, 2024 – 12:08 AM EDT

PDB ID : 6EQY
BMRB ID : 34184
Title : 4th KOW domain of human hSpt5
Authors : Zuber, P.K.; Schweimer, K.; Roesch, P.; Woehrl, B.M.
Deposited on : 2017-10-16

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 ⓘ 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 ⓘ](#)) 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) : 2.37.1

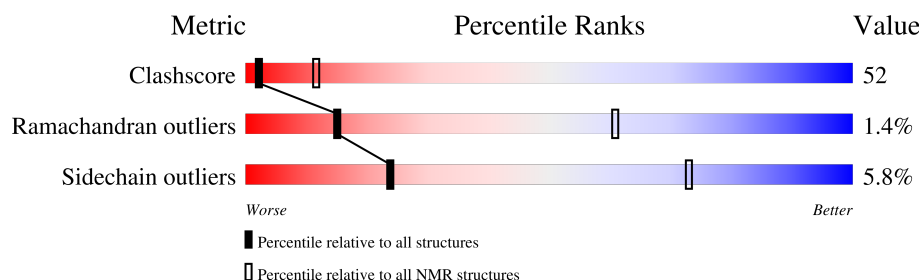
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 87%.

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	129	

2 Ensemble composition and analysis

This entry contains 14 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:17-A:60, A:67-A:130 (108)	0.90	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 6 single-model clusters were found.

Cluster number	Models
1	2, 3, 4, 8, 9, 10
2	5, 7
Single-model clusters	1; 6; 11; 12; 13; 14

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 1870 atoms, of which 942 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Transcription elongation factor SPT5.

Mol	Chain	Residues	Atoms						Trace
1	A	115	Total	C	H	N	O	S	0
			1870	586	942	176	162	4	

There are 3 discrepancies between the modelled and reference sequences:

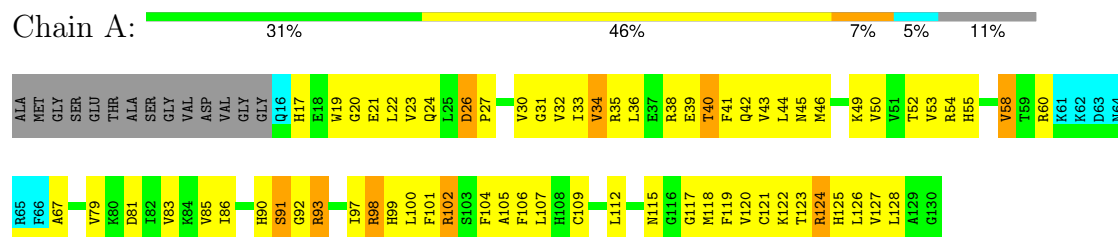
Chain	Residue	Modelled	Actual	Comment	Reference
A	2	ALA	-	expression tag	UNP O00267
A	3	MET	-	expression tag	UNP O00267
A	4	GLY	-	expression tag	UNP O00267

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.

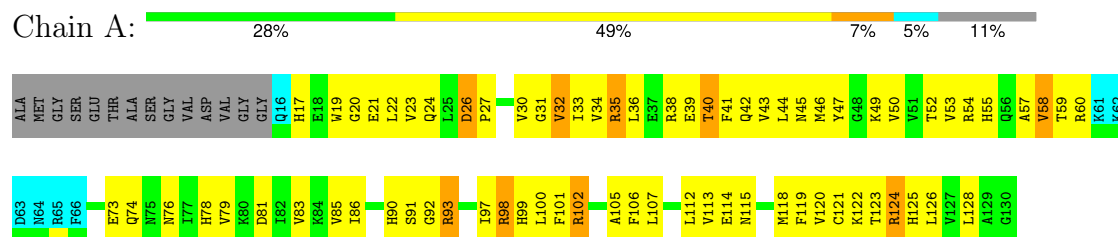
- Molecule 1: Transcription elongation factor SPT5



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: Transcription elongation factor SPT5



5 Refinement protocol and experimental data overview

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

Of the 120 calculated structures, 14 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
Xplor-NIH	structure calculation	

The following table shows chemical shift validation statistics as aggregates over all chemical shift files. Detailed validation can be found in section 7 of this report.

Chemical shift file(s)	working_cs.cif
Number of chemical shift lists	1
Total number of shifts	1492
Number of shifts mapped to atoms	1373
Number of unparsed shifts	0
Number of shifts with mapping errors	119
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	87%

6 Model quality [i](#)

6.1 Standard geometry [i](#)

There are no covalent bond-length or bond-angle outliers.

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	Chirality	Planarity
1	A	0.0±0.0	7.8±0.4
All	All	0	109

There are no bond-length outliers.

There are no bond-angle outliers.

There are no chirality outliers.

5 of 8 unique planar outliers are listed below. They are sorted by the frequency of occurrence in the ensemble.

Mol	Chain	Res	Type	Group	Models (Total)
1	A	35	ARG	Sidechain	14
1	A	54	ARG	Sidechain	14
1	A	60	ARG	Sidechain	14
1	A	93	ARG	Sidechain	14
1	A	98	ARG	Sidechain	14

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	863	876	876	90±9
All	All	12082	12264	12264	1256

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 52.

5 of 475 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:53:VAL:HG21	1:A:58:VAL:HG13	0.98	1.34	7	3
1:A:22:LEU:HD13	1:A:23:VAL:N	0.96	1.74	4	5
1:A:23:VAL:HG12	1:A:59:THR:O	0.96	1.61	7	4
1:A:22:LEU:HD22	1:A:101:PHE:CE2	0.95	1.95	2	1
1:A:30:VAL:HG11	1:A:101:PHE:CD1	0.94	1.98	9	6

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	107/129 (83%)	98±2 (91±2%)	8±2 (7±1%)	2±1 (1±1%)	15	61
All	All	1498/1806 (83%)	1367 (91%)	110 (7%)	21 (1%)	15	61

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	67	ALA	6
1	A	91	SER	4
1	A	102	ARG	4
1	A	46	MET	2
1	A	112	LEU	2

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	94/109 (86%)	89±1 (94±1%)	5±1 (6±1%)	24	73
All	All	1316/1526 (86%)	1240 (94%)	76 (6%)	24	73

5 of 16 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	26	ASP	14
1	A	40	THR	14
1	A	34	VAL	10
1	A	109	CYS	8
1	A	58	VAL	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

The completeness of assignment taking into account all chemical shift lists is 87% for the well-defined parts and 83% for the entire structure.

7.1 Chemical shift list 1

File name: working_cs.cif

Chemical shift list name: *kow4.bmr*

7.1.1 Bookkeeping

The following table shows the results of parsing the chemical shift list and reports the number of nuclei with statistically unusual chemical shifts.

Total number of shifts	1492
Number of shifts mapped to atoms	1373
Number of unparsed shifts	0
Number of shifts with mapping errors	119
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	3

The following assigned chemical shifts were not mapped to the molecules present in the coordinate file.

- No matching atom found in the structure. First 5 (of 119) occurrences are reported below.

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	3	MET	HA	4.5	0.03	1
1	A	3	MET	HB2	2.12	0.03	2
1	A	3	MET	HB3	2.05	0.03	2
1	A	3	MET	HE1	2.1	0.03	1
1	A	3	MET	HE2	2.1	0.03	1
1	A	3	MET	HE3	2.1	0.03	1
1	A	3	MET	CA	55.68	0.20	1
1	A	3	MET	CB	32.12	0.20	1
1	A	3	MET	CG	32.12	0.20	1
1	A	3	MET	CE	16.89	0.20	1
1	A	4	GLY	H	8.53	0.03	1
1	A	4	GLY	HA2	4.03	0.03	2
1	A	4	GLY	HA3	4.03	0.03	2
1	A	4	GLY	C	174.33	0.20	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	4	GLY	CA	45.4	0.20	1
1	A	4	GLY	N	110.88	0.20	1
1	A	5	SER	H	8.25	0.03	1
1	A	5	SER	HA	3.89	0.03	1
1	A	5	SER	HB2	4.48	0.03	2
1	A	5	SER	HB3	4.48	0.03	2
1	A	5	SER	C	174.86	0.20	1
1	A	5	SER	CA	58.49	0.20	1
1	A	5	SER	CB	63.99	0.20	1
1	A	5	SER	N	115.7	0.20	1
1	A	6	GLU	H	8.7	0.03	1
1	A	6	GLU	HA	4.37	0.03	1
1	A	6	GLU	HB2	2.11	0.03	2
1	A	6	GLU	HB3	1.98	0.03	2
1	A	6	GLU	HG2	2.32	0.03	2
1	A	6	GLU	HG3	2.28	0.03	2
1	A	6	GLU	C	176.85	0.20	1
1	A	6	GLU	CA	57.03	0.20	1
1	A	6	GLU	CB	30.1	0.20	1
1	A	6	GLU	CG	36.28	0.20	1
1	A	6	GLU	N	122.96	0.20	1
1	A	7	THR	H	8.17	0.03	1
1	A	7	THR	HA	4.33	0.03	1
1	A	7	THR	HB	4.22	0.03	1
1	A	7	THR	HG21	1.21	0.03	1
1	A	7	THR	HG22	1.21	0.03	1
1	A	7	THR	HG23	1.21	0.03	1
1	A	7	THR	C	174.46	0.20	1
1	A	7	THR	CA	61.91	0.20	1
1	A	7	THR	CB	69.88	0.20	1
1	A	7	THR	CG2	21.63	0.20	1
1	A	7	THR	N	114.86	0.20	1
1	A	8	ALA	H	8.34	0.03	1
1	A	8	ALA	HA	4.39	0.03	1
1	A	8	ALA	HB1	1.42	0.03	1
1	A	8	ALA	HB2	1.42	0.03	1
1	A	8	ALA	HB3	1.42	0.03	1
1	A	8	ALA	C	177.74	0.20	1
1	A	8	ALA	CA	52.63	0.20	1
1	A	8	ALA	CB	19.26	0.20	1
1	A	8	ALA	N	126.83	0.20	1

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	9	SER	H	8.35	0.03	1
1	A	9	SER	HA	4.46	0.03	1
1	A	9	SER	HB2	3.89	0.03	2
1	A	9	SER	HB3	3.89	0.03	2
1	A	9	SER	C	175.15	0.20	1
1	A	9	SER	CA	58.53	0.20	1
1	A	9	SER	CB	64.0	0.20	1
1	A	9	SER	N	115.25	0.20	1
1	A	10	GLY	H	8.4	0.03	1
1	A	10	GLY	HA2	4.02	0.03	2
1	A	10	GLY	HA3	4.02	0.03	2
1	A	10	GLY	C	174.11	0.20	1
1	A	10	GLY	CA	45.44	0.20	1
1	A	10	GLY	N	110.93	0.20	1
1	A	11	VAL	H	7.93	0.03	1
1	A	11	VAL	HA	4.13	0.03	1
1	A	11	VAL	HB	2.07	0.03	1
1	A	11	VAL	HG11	0.9	0.03	2
1	A	11	VAL	HG12	0.9	0.03	2
1	A	11	VAL	HG13	0.9	0.03	2
1	A	11	VAL	HG21	0.9	0.03	2
1	A	11	VAL	HG22	0.9	0.03	2
1	A	11	VAL	HG23	0.9	0.03	2
1	A	11	VAL	C	175.79	0.20	1
1	A	11	VAL	CA	62.2	0.20	1
1	A	11	VAL	CB	32.91	0.20	1
1	A	11	VAL	CG1	20.92	0.20	2
1	A	11	VAL	CG2	20.92	0.20	2
1	A	11	VAL	N	118.71	0.20	1
1	A	12	ASP	H	8.48	0.03	1
1	A	12	ASP	HA	4.67	0.03	1
1	A	12	ASP	HB2	2.76	0.03	2
1	A	12	ASP	HB3	2.61	0.03	2
1	A	12	ASP	C	176.53	0.20	1
1	A	12	ASP	CA	54.2	0.20	1
1	A	12	ASP	CB	41.25	0.20	1
1	A	12	ASP	N	124.13	0.20	1
1	A	13	VAL	H	8.17	0.03	1
1	A	13	VAL	HA	4.15	0.03	1
1	A	13	VAL	HB	2.18	0.03	1
1	A	13	VAL	HG11	0.92	0.03	2

Continued on next page...

Continued from previous page...

List ID	Chain	Res	Type	Atom	Shift Data		
					Value	Uncertainty	Ambiguity
1	A	13	VAL	HG12	0.92	0.03	2
1	A	13	VAL	HG13	0.92	0.03	2
1	A	13	VAL	HG21	0.92	0.03	2
1	A	13	VAL	HG22	0.92	0.03	2
1	A	13	VAL	HG23	0.92	0.03	2
1	A	13	VAL	C	177.01	0.20	1
1	A	13	VAL	CA	62.55	0.20	1
1	A	13	VAL	CB	32.36	0.20	1
1	A	13	VAL	CG1	20.14	0.20	2
1	A	13	VAL	CG2	21.32	0.20	2
1	A	13	VAL	N	120.38	0.20	1
1	A	14	GLY	H	8.52	0.03	1
1	A	14	GLY	HA2	3.94	0.03	2
1	A	14	GLY	HA3	3.94	0.03	2
1	A	14	GLY	C	174.8	0.20	1
1	A	14	GLY	CA	45.58	0.20	1
1	A	14	GLY	N	111.6	0.20	1
1	A	15	GLY	H	8.21	0.03	1
1	A	15	GLY	HA2	3.9	0.03	2
1	A	15	GLY	HA3	3.9	0.03	2
1	A	15	GLY	C	174.06	0.20	1
1	A	15	GLY	CA	45.3	0.20	1
1	A	15	GLY	N	108.41	0.20	1

7.1.2 Chemical shift referencing [i](#)

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	127	-0.22 ± 0.14	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	112	0.15 ± 0.22	None needed (< 0.5 ppm)
$^{13}\text{C}'$	113	0.08 ± 0.14	None needed (< 0.5 ppm)
^{15}N	121	1.26 ± 0.24	Should be applied

7.1.3 Completeness of resonance assignments [i](#)

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 87%, i.e. 1330 atoms were assigned a chemical shift out of a possible 1537. 0 out of 26 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	¹ H	¹³ C	¹⁵ N
Backbone	529/545 (97%)	219/223 (98%)	205/216 (95%)	105/106 (99%)
Sidechain	735/872 (84%)	506/568 (89%)	225/265 (85%)	4/39 (10%)
Aromatic	66/120 (55%)	36/63 (57%)	29/49 (59%)	1/8 (12%)
Overall	1330/1537 (87%)	761/854 (89%)	459/530 (87%)	110/153 (72%)

7.1.4 Statistically unusual chemical shifts [i](#)

The following table lists the statistically unusual chemical shifts. These are statistical measures, and large deviations from the mean do not necessarily imply incorrect assignments. Molecules containing paramagnetic centres or hemes are expected to give rise to anomalous chemical shifts.

List Id	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	22	LEU	HB2	-0.43	-0.07 – 3.30	-6.1
1	A	46	MET	HG2	0.46	0.65 – 4.19	-5.5
1	A	114	GLU	HG2	3.30	1.24 – 3.30	5.0

7.1.5 Random Coil Index (RCI) plots [i](#)

The image below reports *random coil index* values for the protein chains in the structure. The height of each bar gives a probability of a given residue to be disordered, as predicted from the available chemical shifts and the amino acid sequence. A value above 0.2 is an indication of significant predicted disorder. The colour of the bar shows whether the residue is in the well-defined core (black) or in the ill-defined residue ranges (cyan), as described in section 2 on ensemble composition. If well-defined core and ill-defined regions are not identified then it is shown as gray bars.

Random coil index (RCI) for chain A:

