



wwPDB NMR Structure Validation Summary Report ⓘ

Apr 27, 2016 – 05:17 AM BST

PDB ID : 2RSV
Title : Solution structure of human full-length vaccinia related kinase 1 (VRK1)
Authors : Koshiba, S.; Tochio, N.; Yokoyama, J.; Kigawa, T.
Deposited on : 2012-07-12

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 ⓘ symbol.

The following versions of software and data (see [references ⓘ](#)) 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

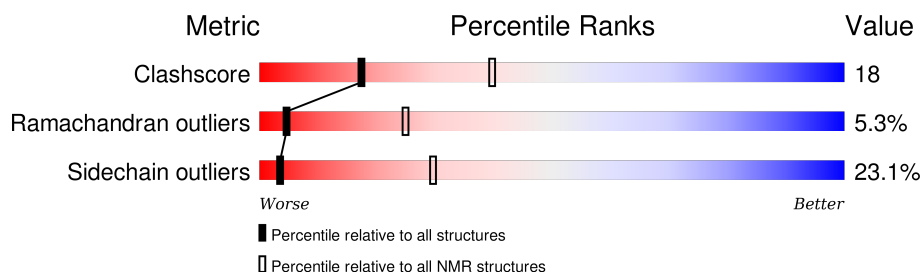
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

SOLUTION NMR

The overall completeness of chemical shifts assignment is 50%.

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 $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	403	<div> <div>34%</div> <div>34%</div> <div>.</div> <div>29%</div> </div>

2 Ensemble composition and analysis

This entry contains 18 models. Model 13 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:24-A:44, A:50-A:204, A:223-A:309, A:315-A:336 (285)	0.86	13

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 3 clusters and 5 single-model clusters were found.

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

3 Entry composition

There is only 1 type of molecule in this entry. The entry contains 6497 atoms, of which 3260 are hydrogens and 0 are deuteriums.

- Molecule 1 is a protein called Serine/threonine-protein kinase VRK1.

Mol	Chain	Residues	Atoms						Trace
1	A	403	Total	C	H	N	O	S	0
			6497	2043	3260	571	609	14	

There are 7 discrepancies between the modelled and reference sequences:

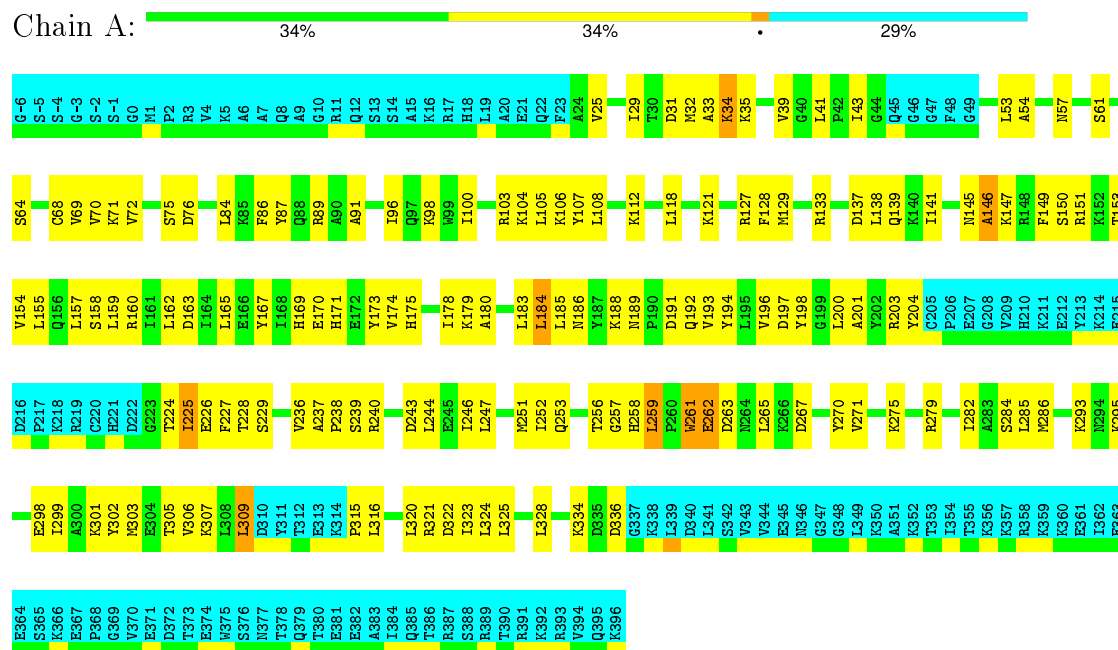
Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	GLY	-	EXPRESSION TAG	UNP Q99986
A	-5	SER	-	EXPRESSION TAG	UNP Q99986
A	-4	SER	-	EXPRESSION TAG	UNP Q99986
A	-3	GLY	-	EXPRESSION TAG	UNP Q99986
A	-2	SER	-	EXPRESSION TAG	UNP Q99986
A	-1	SER	-	EXPRESSION TAG	UNP Q99986
A	0	GLY	-	EXPRESSION TAG	UNP Q99986

4 Residue-property plots

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.

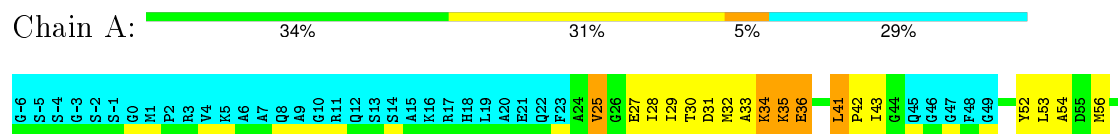
- Molecule 1: Serine/threonine-protein kinase VRK1



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

The representative model is number 13. Colouring as in section 4.1 above.

- Molecule 1: Serine/threonine-protein kinase VRK1



S61	T153	G220	E304	P368
V154	H221	H225	T305	G369
L155	D222	I226	V306	V370
V69	S158	E226	K307	E371
V70	L159	F227	L308	D372
K71	R160	T228	I309	T373
V72	D163	S229	D310	E374
E73	I164	N234	Y311	K375
P74	L165	G235	T312	S376
S75	E166	V236	E313	N377
D76	Y167	A237	K314	T378
E83	I168	P238	P315	Q379
L84	H169	S239	L316	T380
K85	E170	E245	I320	E381
F86	H171	T246	R321	E382
Y87	E172	L247	D322	A383
O88	Y173	G248	I323	I384
R89	H174	K251	L324	Q385
A90	H175	T252	L325	T386
A91	G176	Q253	L326	R387
K98	D177	G257	L328	S388
N99	I178	H258	I331	R389
I100	K179	E259	G332	T390
K104	L183	L260	S333	R391
L105	L184	E261	K334	G392
K106	L185	D262	D335	R393
Y107	N186	E263	L336	V394
L108	Y187	D264	G337	G395
G109	K188	D265	K338	K396
K112	D191	L266	L339	
L118	Q192	K267	D340	
L133	V193	Y270	L341	
R127	F128	V271	S342	
F128	M129	R272	V343	
I130	A201	Y278	V344	
R133	Y202	R279	E345	
S136	R203	T282	K346	
D137	Y204	L283	G347	
L138	P206	S284	G348	
Q139	E207	L285	L349	
G140	G208	V286	K350	
I141	V209	D287	A351	
Y142	H210	P291	K352	
E143	K211	E292	T353	
A144	E212	K293	I354	
N145	Y213	N294	T355	
A146	K214	K301	K356	
K147	E215	Y302	K357	
R148	D216	R303	R358	
F149	P217		K359	
S150	K218		K360	
	R219		E361	
			I362	
			E363	
			E364	
			S365	
			K366	
			E367	

5 Refinement protocol and experimental data overview

The models were refined using the following method: *torsion angle dynamics, simulated annealing*.

Of the 200 calculated structures, 18 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	structure solution	
CNS	refinement	

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)	2rsv_cs.str
Number of chemical shift lists	1
Total number of shifts	2601
Number of shifts mapped to atoms	2601
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Assignment completeness (well-defined parts)	50%

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	2313	2336	2330	83±7
All	All	41634	42048	41940	1486

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 18.

5 of 568 unique clashes are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	Clash(Å)	Distance(Å)	Models	
				Worst	Total
1:A:309:LEU:HD21	1:A:316:LEU:HD22	0.93	1.40	7	1
1:A:141:ILE:HD11	1:A:185:LEU:HD11	0.91	1.37	1	10
1:A:259:LEU:HD22	1:A:261:TRP:CH2	0.89	2.02	2	18
1:A:278:TYR:CD1	1:A:285:LEU:HD12	0.85	2.06	16	3
1:A:29:ILE:HD12	1:A:128:PHE:CZ	0.84	2.06	9	14

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	285/403 (71%)	237±3 (83±1%)	33±3 (12±1%)	15±3 (5±1%)	4	25
All	All	5130/7254 (71%)	4259 (83%)	600 (12%)	271 (5%)	4	25

5 of 66 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	261	TRP	18
1	A	146	ALA	18
1	A	197	ASP	14
1	A	257	GLY	13
1	A	191	ASP	12

6.3.2 Protein sidechains ⓘ

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	250/349 (72%)	192±6 (77±2%)	58±6 (23±2%)	3	30
All	All	4500/6282 (72%)	3459 (77%)	1041 (23%)	3	30

5 of 169 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	259	LEU	18
1	A	251	MET	15
1	A	184	LEU	14
1	A	309	LEU	13
1	A	284	SER	12

6.3.3 RNA ⓘ

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

The completeness of assignment taking into account all chemical shift lists is 50% for the well-defined parts and 44% for the entire structure.

7.1 Chemical shift list 1

File name: 2rsv_cs.str

Chemical shift list name: *assigned_chem_shift_list_1*

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	2601
Number of shifts mapped to atoms	2601
Number of unparsed shifts	0
Number of shifts with mapping errors	0
Number of shifts with mapping warnings	0
Number of shift outliers (ShiftChecker)	6

7.1.2 Chemical shift referencing

The following table shows the suggested chemical shift referencing corrections.

Nucleus	# values	Correction \pm precision, ppm	Suggested action
$^{13}\text{C}_\alpha$	363	0.12 ± 0.08	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	325	1.09 ± 0.06	Should be applied
$^{13}\text{C}'$	363	-0.01 ± 0.09	None needed (< 0.5 ppm)
^{15}N	343	-0.14 ± 0.15	None needed (< 0.5 ppm)

7.1.3 Completeness of resonance assignments

The following table shows the completeness of the chemical shift assignments for the well-defined regions of the structure. The overall completeness is 50%, i.e. 1827 atoms were assigned a chemical shift out of a possible 3660. 0 out of 45 assigned methyl groups (LEU and VAL) were assigned stereospecifically.

	Total	^1H	^{13}C	^{15}N
Backbone	1092/1399 (78%)	266/557 (48%)	560/570 (98%)	266/272 (98%)
Sidechain	609/1946 (31%)	187/1147 (16%)	422/707 (60%)	0/92 (0%)

Continued on next page...

Continued from previous page...

	Total	^1H	^{13}C	^{15}N
Aromatic	126/315 (40%)	63/161 (39%)	58/137 (42%)	5/17 (29%)
Overall	1827/3660 (50%)	516/1865 (28%)	1040/1414 (74%)	271/381 (71%)

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.

Mol	Chain	Res	Type	Atom	Shift, ppm	Expected range, ppm	Z-score
1	A	29	ILE	HG23	-0.64	2.13 – -0.57	-5.3
1	A	29	ILE	HG21	-0.64	2.13 – -0.57	-5.3
1	A	29	ILE	HG22	-0.64	2.13 – -0.57	-5.3
1	A	29	ILE	HD11	-0.81	2.13 – -0.77	-5.1
1	A	29	ILE	HD12	-0.81	2.13 – -0.77	-5.1
1	A	29	ILE	HD13	-0.81	2.13 – -0.77	-5.1

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.

Random coil index (RCI) for chain A:

