



wwPDB NMR Structure Validation Summary Report ⓘ

May 29, 2020 – 01:33 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

<https://www.wwpdb.org/validation/2017/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
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
RCI : v_1n_11_5_13_A (Berjanski et al., 2005)
PANAV : Wang et al. (2010)
ShiftChecker : 2.11
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.11

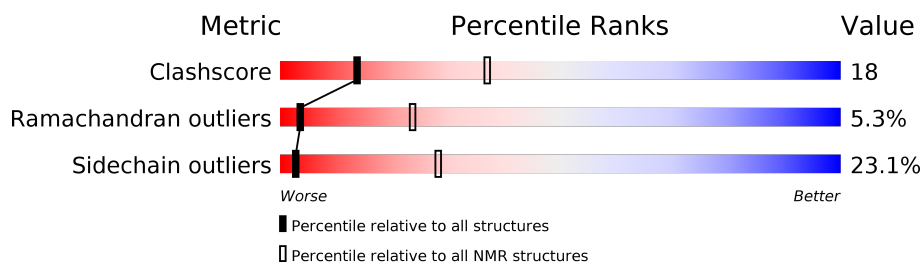
1 Overall quality at a glance

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	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	403	

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 2 clusters and 7 single-model clusters were found.

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

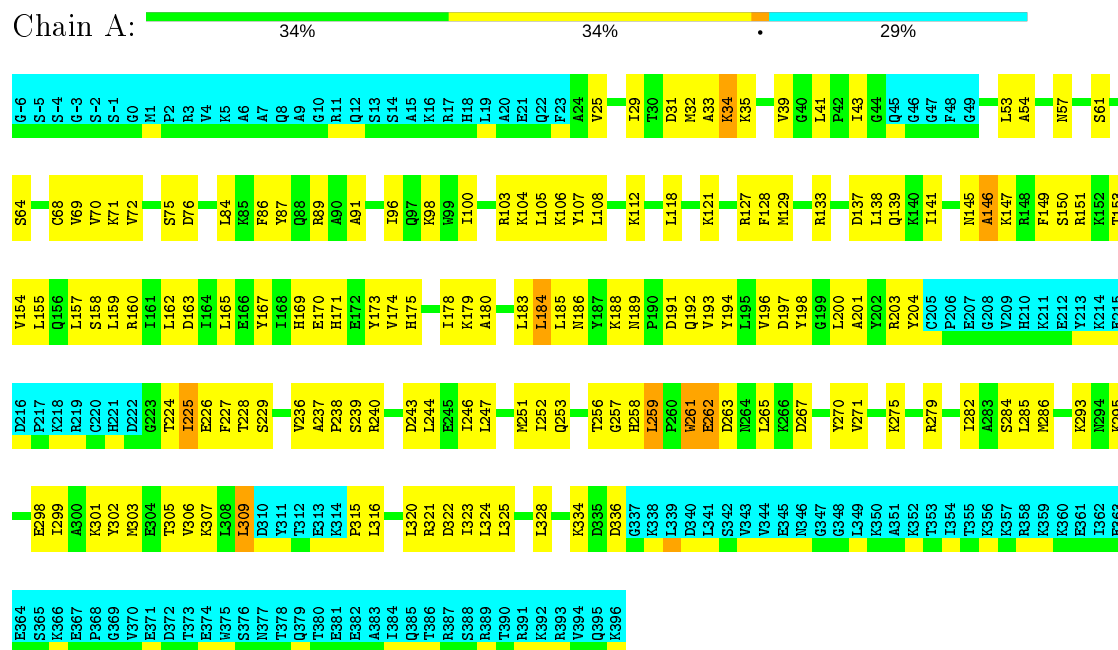
Chain	Residue	Modelled	Actual	Comment	Reference
A	-6	GLY	-	EXPRESSION TAG	UNP Q99986
A	-5	SER	-	EXPRESSION TAG	UNP Q99986
A	-4	SER	-	EXPRESSION TAG	UNP Q99986
A	-3	GLY	-	EXPRESSION TAG	UNP Q99986
A	-2	SER	-	EXPRESSION TAG	UNP Q99986
A	-1	SER	-	EXPRESSION TAG	UNP Q99986
A	0	GLY	-	EXPRESSION TAG	UNP Q99986

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.

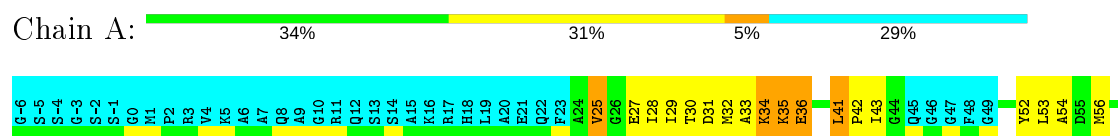
- Molecule 1: Serine/threonine-protein kinase VRK1



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	T305	G369
O68	L155	D222	V306	V370
V69			K307	E371
V70	S158	I225	L308	D372
K71	L159	E226	L309	T373
V72	R160	F227	D310	E374
E73		T228	Y311	K375
P74	D163	S229	T312	S376
S75	I164		E313	N377
D76	L165	N234	K314	T378
	E166	G235	P315	Q379
E83	Y167	Y236	L316	T380
L84	I168	A237		E381
R85	H169	F238	L320	E382
F86	E170	S239	R321	A383
Y87	H171		D322	I384
O88	E172	E245	I323	Q385
R89	Y173	T246	L324	T386
A90	V174	L247	L325	R387
A91	H175	G248		S388
	G176		L328	R389
K98	D177	K251	I331	T390
Y99	I178	I252	G332	R391
I100	K179	Q253	S333	K392
			K334	R393
K104	L183	G257	D335	V394
L105	L184	H258	L336	K395
K106	L185	L259	G337	
Y107	N186	P260	K338	
L108	Y187	E261	L339	
G109	K188	E262	D340	
		D263	L341	
K112	D191	L264	S342	
	Q192	K265	V343	
L118	V193	K266	V344	
	Y194	D267	E345	
			K346	
R127	D197	Y270	G347	
F128	Y198	V271	G348	
M129		R272	L349	
I130	A201		K350	
		Y278	A351	
R133	Y202	R279	K352	
	R203		T353	
S136	Y204	T282	I354	
D137	C205	L283	T355	
L138	P206	S284	K356	
Q139	E207	L285	K357	
G140	G208	V286	R358	
K141	V209	D287	K359	
I141	H210		K360	
Y142	K211	P291	E361	
E143	E212	E292	I362	
A144	Y213	K293	E363	
N145	K214	N294	E364	
A146	E215		S365	
K147	D216	K301	K366	
R148	F149	Y302	E367	
S150	K218	R303		
	R219			

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)	input_cs.cif
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 567 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%)	3	23
All	All	5130/7254 (71%)	4259 (83%)	600 (12%)	271 (5%)	3	23

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	28
All	All	4500/6282 (72%)	3459 (77%)	1041 (23%)	3	28

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: input_cs.cif

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	0.11 ± 0.10	None needed (< 0.5 ppm)
$^{13}\text{C}_\beta$	325	1.09 ± 0.10	Should be applied
$^{13}\text{C}'$	363	-0.01 ± 0.07	None needed (< 0.5 ppm)
^{15}N	343	-0.14 ± 0.16	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	¹ H	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 ⓘ

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 ⓘ

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:

