

Viral Protein	PDB Structure	Chain	Total BLAST Hits	BLAST Hits Used	Unique Sequences with Difference ¹		
					10%	5%	2%
Dengue Protease-Helicase	2VBC	A	149	30	0 / 0%	7 / 23%	7 / 23%
		B	7	7	2 / 29%	3 / 43%	3 / 43%
West Nile Protease	2FP7	A	21	12	2 / 17%	2 / 7%	5 / 42%
		B	77	15	2 / 13%	6 / 40%	6 / 40%
HIV – 1 Reverse Transcriptase	2HMI	B	354	322	2 / 0.6%	2 / 0.6%	3 / 0.9%
Influenza Nucleoprotein	3TG6	A	18	5	2 / 40%	2 / 40%	3 / 60%
Marburg RNA binding domain	4GH9	A	44	32	3 / 9%	3 / 9%	3 / 9%
Hepatitis C Protease	4AEX	A	353	222	4 / 2%	5 / 2%	10 / 5%
Japanese Encephalitis Helicase/Nucleoside	2Z83	A	119	31	6 / 19%	7 / 23%	7 / 23%
Crimean-Congo Hemorrhagic Fever Nucleocapsid	4AKL	A	16	6	1 / 17%	1 / 17%	3 / 50%
Rift Valley Fever Virus Nucleoprotein	3OV9	A	145	73	1 / 1%	1 / 1%	3 / 4%
Hemagglutinin Precursor	1RD8	A	365	313	14 / 4%	19 / 6%	27 / 9%
		B	343	309	8 / 3%	9 / 3%	14 / 45%

¹Results presented as count/percent of BLAST hits used.

Table 1. Unique sequence found through BLAST in viral protein PDB structures. Numbers in BLAST hits used column are the sequences that pass $\geq 40\%$ sequence identity and $\geq 90\%$ alignment length criteria from the total found. Generally the number of unique sequences found is rather low compared to BLAST hits used; except in those cases with low number of BLAST hits used. The highest percentage of unique sequences found is mostly in those PDB structures with lowest number of used BLAST hits.