

Zika virus phosphoproteome through the computational looking-glass and what we found there?

Computational phospho-mining and related post translational modification analyses using *in silico* methods is debateable. For some wet lab scientists, it is similar to the twilight zone, where nothing might be quite what it seems. We certainly agree that the wet lab is critical for validation but computational assays rapidly generate preliminary data for assessment and aid wet lab. Zika virus is wrecking a havoc and it is the need of the hour to rapidly assess various phenomenon associated with its pathogenicity and disseminate it to the scientific community for assessment. Phosphorylation assists in viral replication and its association with host receptors. Here, we present a vignette of the phosphorylation potential of Zika virus proteins.

Zika virus phosphoproteome through the computational looking-glass and what we found there?

Zarrin Basharat*, Tabeer Khan, Azra Yasmin

Microbiology & Biotechnology Research Lab, Department of Environmental Sciences,
Fatima Jinnah Women University, 46000, Pakistan.

*Corresponding author.

Email: zarrin.iiui@gmail.com

In wake of the recent outbreaks of Zika virus which resulted in massive increase in microcephaly in the affected areas, Zika virus infection has recently been declared as a global health emergency by WHO [1]. Rapid analysis of mechanisms underlying the pathogenesis of this virus is required for speedy development of therapeutics. Computational screening using webservers trained on experimental phosphoprotein datasets is a handy method for quick scrutinization of phosphosites in pathogenic proteins.

Virus protein phosphorylation aids in replication and disrupts the normal host-cell functions, leading to pathogenesis [2]. Phosphorylation and dephosphorylation of protein residues may also be involved in cell regulation/ signalling processes [3]. Phosphorylation information is thus, critical for understanding virus protein functionality and in drug design. Prediction of this post translational modification for viral proteins has been attempted previously for Ebola virus nucleoprotein [4], envelope 2 protein of Hepatitis C virus genotype 1a [5], envelope proteins of Hepatitis C genotype 5a [6], Hepatitis G virus envelope glycoprotein E2 [7], whole proteome of rotavirus [8], using computational webservers.

We have used support vector machine based computational method, ViralPhos (<http://csb.cse.yzu.edu.tw/ViralPhos/>) [9], for phosphorylation potential prediction of the Zika virus proteome. ViralPhos is specifically tailored for phosphorylation analysis of viral proteins and probes virus substrate site motifs for identification of potential phosphorylation sites on the viral proteins.

We identified a plethora of phosphosites in the whole proteome of Zika virus (Table 1). We anticipate that this information will be handy in further experimental studies focused on phosphorylation of Zika virus proteins. This computational analysis could provide rapid baseline information and pave way for further elucidation of the phosphorylation in Zika virus pathogenesis until the availability of experimentally determined phosphoproteome of this virus.

References

- [1] Heymann DL, Hodgson A, Sall AA, Freedman DO, Staples JE, Althabe F, Baruah K, Mahmud G, Kandun N, Vasconcelos PF, Bino S. Zika virus and microcephaly: why is this situation a PHEIC?. *Lancet* 2016; **387**(10020):719.

- [2] Schwartz D, Church GM. Collection and motif-based prediction of phosphorylation sites in human viruses. *Sci Signal* 2010; **3**(137):rs2.
- [3] Basharat Z, Yasmin, A. (2016). *In silico* assessment of phosphorylation and O- β -GlcNAcylation in human NPC1 protein critical for Ebola virus entry. *Infect Genet Evol.* **34**: 326-338.
- [4] Yasri S, Wiwanitkit V. Phosphorylation sites within Ebola virus nucleoprotein. *Asian Pac J Trop Biomed* 2015; **5**(7): 598-9.
- [5] Afzal S, Idrees M, Ali M, Ilyas M, Hussain A, Akram M, Butt S, Saleem S, Rehman I, Ali L, Shahid M. Envelope 2 protein phosphorylation sites S75 & 277 of hepatitis C virus genotype 1a and interferon resistance: A sequence alignment approach. *Virology* 2011; **8**(1):1.
- [6] Gededzha MP, Mphahlele MJ, Selabe SG. Characterization of HCV Genotype 5a Envelope Proteins: Implications for Vaccine Development and Therapeutic Entry Target. *Hepat Mon* 2014; **14**(11).
- [7] Ranjbar MM, Ghorban K, Alavian SM, Keyvani H, Dadmanesh M, Ardakany AR, Motedayen MH, Sazmand A. GB virus C/hepatitis G virus envelope glycoprotein E2: computational molecular features and immunoinformatics study. *Hepat Mon* 2013; **13**(12).
- [8] Chattopadhyay S, Bagchi P, Dutta D, Mukherjee A, Kobayashi N, Chawlasarkar M. Computational identification of posttranslational modification sites and functional families reveal possible moonlighting role of rotaviral proteins. *Bioinformatics* 2010; **4**(10):448.
- [9] Huang K, Lu C, Bretaña NA, Lee T, Chang T. ViralPhos: incorporating a recursively statistical method to predict phosphorylation sites on virus proteins. *BMC Bioinformatics* 2013; **14**(Suppl 16):S10.

Table 1. Identification of phosphorylated sites on Zika virus proteome consisting of proteins. Reference genome sequence with Accession no: YP_009227198.1 was used as an information source of Zika virus proteins. Residue in the phosphorylated fragment is underlined.

Protein	Location	Length (amino acids)	Phosphorylated residue	Fragment containing phosphorylated residue
Glycoprotein E	291-790	500	322	HGGCV <u>T</u> VMAQD
			337	DIELV <u>T</u> TTVSN
			338	IELV <u>T</u> TTVSNM
			339	ELV <u>T</u> TTVSNMA
			366	DSRC <u>P</u> TQGEAY
			371	TQGEA <u>Y</u> LDKQS
			405	KGSLV <u>T</u> CAKFT
			416	CSKKM <u>T</u> GKSIQ
			436	LSVHG <u>S</u> QHSGM
			446	MIGYE <u>T</u> DEDRA
			456	AKVEV <u>T</u> PNSPR
			459	EVT <u>P</u> NSPRAEA
			465	PRAEA <u>T</u> LGGFG
			471	LGGFG <u>S</u> LGLDC
			480	DCEP <u>R</u> TGLDFS
			517	HAGAD <u>T</u> GTPHW
540	HAKRQ <u>T</u> VVVVLG			
553	EGAVH <u>T</u> ALAGA			

			590	RLKGV <u>S</u> YSLCT
			595	SYSLCTAAFTF
			601	AAFTF <u>T</u> KVPAE
			611	ETLHG <u>T</u> VTVEV
			613	LHGTV <u>T</u> VEVQY
			618	TVEVQ <u>Y</u> AGTDG
			639	DMQTL <u>T</u> PVGRL
			646	VGRLI <u>T</u> ANPVI
			654	PVITE <u>S</u> TENSK
			655	VITE <u>S</u> TENSKM
			672	PFGD <u>S</u> YIVIGV
			683	GDKKI <u>T</u> HHWHR
			691	WHRSG <u>S</u> TIGKA
			700	KAFEAT <u>T</u> VRGAK
			713	AVLGD <u>T</u> AWDFG
			719	AWDFG <u>S</u> VGGVF
			769	NTKNG <u>S</u> ISLTC
			771	KNGSIS <u>L</u> TCLA
			773	KNGSIS <u>L</u> TCLA
Capsid protein C	12-122	111	71	INRWG <u>S</u> VGKKE
DEXDc	1679-1824	146	1688	KKKQL <u>T</u> VLDLH
			1699	PGAGK <u>T</u> RRVLP
			1717	KKRLR <u>T</u> VILAP
			1723	VILAP <u>T</u> RVVAA
			1743	PVRYM <u>T</u> TAVNV

			1749	TAVNV <u>T</u> HSGTE
			1753	VTHSGTE <u>I</u> VDL
			1763	LMCHA <u>T</u> FTSRL
			1788	DEAHF <u>T</u> DPSSI
			1791	HFTDP <u>S</u> SIAAR
			1801	RGYIS <u>T</u> RVEMG
			1816	IFMTA <u>T</u> PPGTR
NS2A	1154- 1368	215	1171	LKKRM <u>T</u> TKIIM
			1172	KKRM <u>T</u> TKIIMS
			1178	KIIM <u>S</u> TSM AVL
			1212	FAEMN <u>T</u> GGDVA
			1242	FRANW <u>T</u> PRESM
			1246	WTPRE <u>S</u> MLLAL
			1258	SCLLQ <u>T</u> AISAL
			1307	PLARG <u>T</u> LLVAW
			1318	RAGLA <u>T</u> CGGIM
			1332	LKGK <u>G</u> SVKKNL
			1363	GLLL <u>L</u> TRSGKR
NS2B	1372- 1498	127	1377	PSEVL <u>T</u> AVGLI
			1413	VSYV <u>V</u> SGKSVD
			1420	KSVD <u>M</u> YIERAG
			1436	KDAEV <u>T</u> GN SPR
			1439	EVTGN <u>S</u> PRLDV
			1449	VALDE <u>S</u> GDFSL
			1453	ESGDF <u>S</u> LVEED
NS3	1499-	617	1517	KKGET <u>T</u> DGVYR

	2115		1521	TTDGV <u>Y</u> RVMT <u>R</u>
			1531	RRL <u>L</u> G <u>S</u> TQVG <u>V</u>
			1532	RLLG <u>S</u> TQVG <u>V</u> G
			1551	TMWHV <u>T</u> KGA <u>A</u> L
			1576	KQDLV <u>S</u> YCGP <u>W</u>
			1616	PGIFK <u>T</u> KDG <u>D</u> I
			1635	AGTSG <u>S</u> PIL <u>D</u> K
			1658	VIKNG <u>S</u> YV <u>S</u> AI
			1672	KREEE <u>T</u> PVE <u>C</u> F
			1688	KKKQ <u>L</u> T <u>V</u> LDL <u>H</u>
			1699	PGAGK <u>T</u> RRV <u>L</u> P
			1717	KKRLR <u>T</u> VIL <u>A</u> P
			1723	VILAP <u>T</u> RVV <u>A</u> A
			1743	PVRYM <u>T</u> TAV <u>N</u> V
			1749	TAVNV <u>T</u> HSG <u>T</u> E
			1753	VTHSG <u>T</u> EIV <u>D</u> L
			1763	LMCHA <u>T</u> F <u>T</u> SRL
			1788	DEAHF <u>T</u> DP <u>S</u> SI
			1791	HFTDP <u>S</u> SIA <u>A</u> R
			1801	RGYIS <u>T</u> RVEM <u>G</u>
			1816	IFMTA <u>T</u> PPG <u>T</u> R
			1820	ATPPG <u>T</u> RD <u>A</u> FP
			1829	FPDSN <u>S</u> PIM <u>D</u> T
			1834	SPIMD <u>T</u> EVE <u>V</u> P
			1844	PERAW <u>S</u> SGFD <u>W</u>
			1863	VWVFV <u>P</u> S <u>V</u> VRNG <u>N</u>

			1875	IAACLTKAGKR
			1888	QLSRKTFETEF
			1947	GPMPVTHASAA
			1978	GGCAETDEGHA
			2024	EFKLRTEQRKT
			2029	TEQRKTFVELM
			2068	GTTNNTIMEDS
			2073	TIMEDSVPAEV
			2080	PAEVWTKYGEK
			2106	HAALKSFKEFA
NS1	792- 1144	353	803	FSKKETRCGTG
			807	ETRCGTGVFIY
			812	TGVFIYNDVEA
			828	KYHPDSPRRLA
			877	NGVQLTVVVGS
			919	VRAAKTNNSFV
			922	AKTNN\$FVVDG
			942	HRAWNSFLVED
			955	FGVFHTSVWLK
			966	VREDYSLECDP
			976	PAVIGTAVKGR
			1012	LIEMKTCEWPK
			1020	WPKSHTLWTDG
			1023	SHTLWTDGVEE
			1046	LSHHNTREGYR
			1052	REGYRTQVKGP

			1073	EECPG <u>T</u> KVYVE
			1080	VYVEE <u>T</u> CGTRG
			1083	EETCG <u>T</u> RGPLS
			1087	GTRG <u>P</u> SLRSTT
			1090	G <u>P</u> SLR <u>S</u> TTASG
			1091	PSLR <u>S</u> TTASGR
			1092	SLR <u>S</u> TTASGRV
			1107	CCRECT <u>M</u> PPLS
NS4B	2266- 2516	251	2290	REEGAT <u>M</u> GFMSM
			2308	SAWAI <u>Y</u> AALTT
			2312	IYAAL <u>T</u> TLITP
			2316	LTTLIT <u>P</u> AVQH
			2324	VQHAV <u>T</u> TSYNN
			2325	QHAV <u>T</u> TSYNNY
			2326	HAV <u>T</u> TSYNNYS
			2327	AV <u>T</u> TSYNNYSL
			2404	AAQKR <u>T</u> AAGIM
			2424	VTDID <u>T</u> MTIDP
			2465	AGALI <u>T</u> AATST
			2475	TLWEG <u>S</u> PNKYW
			2483	KYWNS <u>S</u> STATSL
			2505	ASLIY <u>T</u> VTRNA
NS4A	2119- 2263	145	2127	MEALG <u>T</u> LPGHM
			2133	LPGHM <u>T</u> ERFQE
			2153	RAETG <u>S</u> RPYKA
			2156	TGSR <u>P</u> YKAAAA

			2166	AQLPET <u>LE</u> TIM
			2243	PEKQRSPQDNQ
FtsJ	2571- 2742	172	2586	LEERGY <u>LQ</u> PYG
			2605	RGGWS <u>Y</u> YAATI
			2620	EVRGY <u>T</u> KGGPG
			2667	DIGESS <u>SS</u> PEV
			2668	IGESS <u>SS</u> PEVE
			2669	GESS <u>SS</u> PEVEE
			2675	PEVEE <u>T</u> RTRLV
			2704	VLCPY <u>T</u> STMME
			2706	CPYT <u>S</u> TMMETM
			2710	STMM <u>E</u> TMERLQ
			2731	PLCRN <u>S</u> THEMY
			2732	LCRN <u>S</u> THEMYW
			2736	STHEM <u>Y</u> WVSGA
HELICc	1855- 1971	117	1863	VWVFP <u>S</u> VRNGN
			1875	IAACL <u>T</u> KAGKR
			1888	QLSRK <u>T</u> FETEF
			1947	GPMPV <u>T</u> HASAA
NS2B	1369- 1498	130	1377	PSEVL <u>T</u> AVGLI
			1413	VSYVV <u>S</u> GKSVD
			1420	KSVD <u>M</u> YIERAG
			1436	KDAE <u>V</u> TGNspr
			1439	EVTGN <u>S</u> PRLDV
			1449	VALDE <u>S</u> GDFSL
			1453	ESGD <u>F</u> SLVEED

S7	1516-1666	151	1521	TTDGV <u>Y</u> RVMT <u>R</u>
			1531	RRL <u>L</u> G <u>S</u> TQVG <u>V</u>
			1532	RLLG <u>S</u> TQVG <u>V</u> G
			1551	TMWHV <u>T</u> KGA <u>A</u> L
			1576	KQDLV <u>S</u> YCGP <u>W</u>
			1616	PGIFK <u>T</u> KDG <u>D</u> I
			1635	AGTSG <u>S</u> PIL <u>D</u> K
			1658	VIKNG <u>S</u> YV <u>S</u> AI
E stem	694-790	97	700	KA <u>F</u> EAT <u>V</u> RG <u>A</u> K
			713	AVLGD <u>T</u> AWDF <u>G</u>
			719	AWDFG <u>S</u> VGG <u>V</u> F
			769	NTKNG <u>S</u> ISL <u>T</u> C
			771	
			773	
				KNG <u>S</u> ISL <u>T</u> CLA
				GSISL <u>T</u> CL <u>A</u> LG
Glycoprotein M	216-290	75	224	LPSHSTR <u>K</u> LQ <u>T</u>
			233	QTRSQ <u>T</u> WLES <u>R</u>
			267	AWLLG <u>S</u> S <u>T</u> SQK
			271	WLLG <u>S</u> S <u>T</u> SQK <u>V</u>
			272	LLG <u>S</u> S <u>T</u> SQK <u>V</u> I
			273	LGS <u>S</u> S <u>T</u> SQK <u>V</u> IY