

Computational identification of the post-translational modification sites and the functional family prediction reveals possible moonlighting role of rotaviral proteins

Shiladitya Chattopadhyay¹, Parikshit Bagchi¹, Dipanjan Dutta¹, Anupam Mukherjee¹, Nobumichi Kobayashi², Mamta Chawla-Sarkar^{1*}

¹Division of Virology, National Institute of Cholera and Enteric Diseases, P-33, C.I.T. Road Scheme- XM, Beliaghata, Kolkata-700010, India; ²Department of Hygiene, Sapporo Medical University, S-1 W-17, Chuo-ku, Sapporo 060-8556, Japan. Mamta Chawla Sarkar – Email- chawlam70@gmail.com. Tel. +91-33-2353-7470. Fax. +91-33-2370-5066, *corresponding author

Received February 02, 2010 ; accepted March 02, 2010 ; published April 30, 2010

Abstract:

Rotavirus (RV) diarrhoea causes huge number deaths in children less than 5 years of age. In spite of available vaccines, it has been difficult to combat RV due to large number of antigenically distinct genotypes, high mutation rates, generation of reassortant viruses due to segmented genome. RV is an eukaryotic virus which utilizes host cell machinery for its propagation. Since RV only encodes 12 proteins, post-translational modification (PTM) is important mechanism for modification, which consequently alters their function. A single protein exhibiting different functions in different locations or in different subcellular sites, are known to be 'moonlighting'. So there is a possibility that viral proteins moonlight in separate location and in different time to exhibit diverse cellular effects. Based on the primary sequence, the putative behaviour of proteins in cellular environment can be predicted, which helps to classify them into different functional families with high reliability score. In this study, sites for phosphorylation, glycosylation and SUMOylation of the six RV structural proteins (VP1, VP2, VP3, VP4, VP6 & VP7) & five non-structural proteins (NSP1, NSP2, NSP3, NSP4 & NSP5) and the functional families were predicted. As NSP6 is a very small protein and not required for virus growth & replication, it was not included in the study. Classification of RV proteins revealed multiple putative functions of each structural protein and varied number of PTM sites, indicating that RV proteins may also moonlight depending on requirements during viral life cycle. Targeting the crucial PTM sites on RV structural proteins may have implications in developing future anti-rotaviral strategies.

Background:

Rotavirus, member of genus Reoviridae is one of the major diarrheal agents causing huge number of deaths worldwide especially among children and dairy animals. An estimated 527,000 children aged <5 years die from rotavirus diarrhea each year, with >85% of these deaths occurring in low-income countries of Africa and Asia [1]. There are seven groups of rotavirus, designated as A, B, C, D, E, F and G. Humans are primarily infected by group A, B and C, however group A is most pathogenic. The viruses consist of eleven double stranded RNA genome encoding twelve proteins. Among the twelve proteins NSP1-NSP6 are nonstructural and remaining six VP1-VP4, VP6 & VP7 are viral RNA polymerase and others structural components of the virus [2]. Rotaviruses primarily infect the enterocytes of the small intestine though there are few recent reports of extra-intestinal infections. From 1975 through 2000 there were 12 case reports where rotavirus infection was found in CNS complications [3].

Viruses are the major cause of several newly emerged diseases throughout the world. In spite of their small genomes, viruses are capable of capturing host machinery for replication and propagation. Viruses in general express multifunctional proteins, with various sites for post-translational modification (PTM). This flexibility gives them chance to participate in cellular signaling events, either to increase their pathogenicity or to escape the robust host immune system. The modification of this kind is not probably mediated by host factors. Some modifications of the rotavirus proteins are known, such as the hyperphosphorylation of the NSP5 protein and glycosylation of the VP7 protein. But overall post-translational modification status of the proteome of rotavirus is yet to be elucidated. As rotavirus is a eukaryotic virus and utilizes the host machinery to replicate and propagate, it can be hypothesized that rotaviral proteins are also modified in host cells.

The phenomenon by which a single protein exhibits different roles in different environment or in different subcellular location is known as 'moonlighting' [4]. As previously stated, viruses with their small repertoire of proteins are capable to capture host machineries, so it is possible that viral proteins moonlight in the host environment either by interaction with cellular multifunctional protein complexes or with the help of post-translational modification. Examples of multifunctional viral proteins are, NS1 protein of influenza virus which inhibits host immune responses by

limiting both interferon (IFN) production and the downstream effects of IFN-induced proteins and also acts directly to modulate other important aspects of the virus replication cycle, including viral RNA replication, viral protein synthesis, and general host-cell physiology [5]. Nucleocapsid (N) protein of Hantaviruses is associated with viral RNA but it also interacts with the virus polymerase and as one of the glycoproteins, interferes with important regulatory pathways in the infected cells [6]. Similarly hepatitis B virus (HBV) protein HBx is a multifunctional viral protein that modulates transcription, cell responses to genotoxic stress, protein degradation, and signaling pathways [7] and NS5A of hepatitis C virus (HCV) has role in viral replication, modulation of the cellular signaling apparatus and cell cycle-regulatory kinases [8]. Functional families of JEV proteins have been assigned previously using a Support Vector Machine based software (SVMProt) which categorizes every viral protein into different families [9].

Examples of major modifications of the proteins which might help the virus during the life cycles are phosphorylation of PB1-F2 protein which results in promotion of apoptosis in influenza [10], phosphorylation of HCV NS5A to modulate host interferon-induced antiviral response [11]. Similarly glycosylated WNV envelope protein plays an important role in replication and maturation processes, glycosylation of E1 protein in HCV caused the translocation of the protein to the cell surface whereas glycosylation decreases the virulence of influenza virus H3N2 [12]. SUMO modification of the adenoviral oncoprotein E1B-55kDa results in localization of the protein from cytoplasm to nucleus [13] and SUMOylated EBV protein BZLF1 disrupts PML Bodies [14]. Methylation is mostly studied in the case of histones, which are associated with DNA and can activate or repress gene expression [15]. Nuclear proteins and various cytoplasmic regulators are subject to lysine acetylation. Lysine acetylation can result in diverse functional role of proteins [16]. In proteins with one or few sites, it generally acts as an on-off switch or it may interact with other PTMs to exert their effects. Thus understanding the functional role of viral proteins is a very important aspect in studying the virus mediated pathogenesis.

Methodology:

Sequence retrieval:

Rotavirus protein sequences were downloaded from NCBI database (<http://www.ncbi.nlm.nih.gov/>). Sequences of human Rotavirus strain PA169 were considered for the study. As NSP6 protein is not indispensable for the Rotavirus life-cycle in the cell, it is not included in the study. Accession numbers for the viral sequences are as follows: NSP1 (EF554132), NSP2 (EF554133), NSP3 (EF554134), NSP4 (EF554135), NSP5 (EF554136), VP1 (EF554126), VP2 (EF554127), VP3 (EF554128), VP4 (EF554129), VP6 (EF554130) and VP7 (EF554131).

Prediction of protein functional family by web-based Support Vector Machine software SVMProt:

SVMProt (<http://jing.cz3.nus.edu.sg/cgi-bin/svmprot.cgi>) classifies a protein into different functional families from its primary sequence information. A slightly modified reliability score, R-value, is used in SVMProt. R-value is a scoring function for estimating the accuracy of support vector machine classification. P-value is the expected classification accuracy (probability of correct classification). It is derived from the statistical relationship between the R-value and actual classification accuracy based on the analysis of 9,932 positive and 45,999 negative samples of proteins [17].

Prediction of phosphorylation sites in proteins using NetPhos 2.0 server:

Serine, threonine and tyrosine phosphorylation status of the Rotavirus proteins were predicted using the online open access server NetPhos 2.0 (<http://www.cbs.dtu.dk/services/NetPhos/>). The NetPhos 2.0 server predicts serine, threonine and tyrosine phosphorylation sites in eukaryotic proteins based on artificial neural network. The threshold value for the prediction is 0.5 [18].

Prediction of N-Linked and O-Linked Glycosylation profile for the rotavirus proteins:

NetNGlyc 1.0 server was used to find the N-glycosylated proteins of the virus. NetNGlyc 1.0 server (<http://www.cbs.dtu.dk/services/NetNGlyc/>) predicts N-Glycosylation sites in human proteins using artificial neural networks that examine the sequence context of Asn-Xaa-Ser/Thr sequences. Sequences having N-glycosylation potential >0.5 are considered as cut-off value [19].

NetOGlyc 3.1 server (<http://www.cbs.dtu.dk/services/NetOGlyc/>) was used to find the O-glycosylation of the viral proteins during infection. The NetOglyc server produces neural network based predictions of mucin type GalNAc O-glycosylation sites in mammalian proteins. The G-score is the score from the best general predictor; the I-score is the score from the best isolated site predictor. If the G-score is >0.5 the residue is predicted as glycosylated; the higher the score the more confident the prediction. For threonines an additional score is used [20].

Prediction of potential SUMOylation sites in viral proteins:

SUMOylation prediction sites were analyzed using SUMOsp2.0 server (<http://sumosp.biocuckoo.org/online.php>). Low threshold was chosen for the prediction of the modification. GPS and MotifX, two powerful prediction strategies previously used for phosphorylation prediction were used to create this algorithm [21].

Prediction of few other post-translational modifications in AutoMotif2.0 Server:

The AutoMotif 2.0 Server (<http://ams2.bioinfo.pl/>) allows the identification of post-translational modification (PTM) sites in proteins based only on local sequence information. The Support Vector Machine (SVM) search type was used for the screening. The performance of SVM models is described by recall R and the precision P. The recall R value measures the percentage of correct predictions, whereas precision P gives the percentage of observed positives that are correctly predicted [22]. AutoMotif 2.0 search was done taking Methylation and Acetylation in consideration.

Discussion:

NSP1, the viral IRF3 antagonist contains 26 phosphorylation sites. It also has total 5 glycosylation sites, 4 of them are N-linked in nature. Though SUMOylation reveals 3 type II sites, no type I site

was found. It does not have a methylation site but one acetylation site is present. Prediction of functional family of proteins using the SVMProt classifies NSP1 into 6 functional families with highest P-value for the Zinc-binding family. Next to come according to the P-value are RNA-binding, Metal-binding, Nuclear receptor, DNA repair and Magnesium-binding. The results are consistent with previous wet-lab experiments of having a Zinc-binding domain and RNA-binding activity.

NSP2 is a smaller protein than NSP1 and contain 12 phosphorylation sites. It has only one N-linked glycosylation site. NSP2 has 5 type II and type I SUMOylation sites. It has only one methylation site but is devoid of any acetylation sites. NSP2 may function as Zinc binding, Actin binding, Calcium binding, Type II secretory pathway and Pore-forming toxins.

NSP3 has 23 phosphorylation sites; it has no predicted glycosylation and acetylation sites. NSP3 has 4 SUMOylation sites and one methylation site. NSP3 belongs to seven families which according to P-value are Zinc binding, all lipid-binding, metal binding chlorophyll biosynthesis, Actin biosynthesis, DNA repair and DNA condensation.

NSP4, the putative viral enterotoxin has 9 phosphorylation sites. NSP4 sequence contains 3 glycosylation and one type II SUMOylation sites. It also possesses one methylation site. NSP4 is predicted to be belonged to following three families Transmembrane, ABC family and Magnesium-binding.

NSP5 as previously reported was found to be hyperphosphorylated in 31 sites. Though VP1 & VP4 contains more phosphorylation sites; considering the size of NSP5 the extent of phosphorylation is much higher in NSP5. As NSP5 is involved in replication of the virus so phosphorylation might have a role during it. Glycosylation status prediction has shown total 9 sites but contrary to other proteins, 8 out of them are O-glycosylated. It has 5 SUMOylation sites and 1 acetylation site but no methylation sites. According to SVMProt, NSP5 belonged to the Coat protein family which was not known previously.

VP1, the largest protein of the virus is the viral polymerase contains 64 phosphorylation sites, high compared to other viral proteins. VP1 also has 5 glycosylation sites, 3 SUMOylation sites, 2 methylation sites and 3 acetylation sites. VP1 belongs to following seven families; Transferase, Structural, DNA replication, DNA-binding, Zinc-binding, Photosystem I and Pore forming toxins.

VP2 has 26 phosphorylation sites and 4 glycosylation sites. The VP2 protein has 13 SUMOylation sites and may play a big role in the virus life-cycle. VP2 protein is found to have 4 methylation sites, which is more than any other rotaviral proteins but it lacks the acetylation site. Previously it is known that VP2 remain associated with viral RNA and acts as a scaffolding protein. So there is a possibility that it may control the expression of viral RNAs after infection. VP2 may function as RNA-binding, Oxidoreductases, Type II secretory pathway and Magnesium-binding proteins.

VP3 has 46 phosphorylation sites. It is also one of the most glycosylated proteins and has 9 N-linked & 1- O-linked glycosylation sites. It is also highly SUMOylated as it contains 5, type I and 5, type II sites. VP3 has only 2 methylation and 1 acetylation sites. VP3 classifies as with Zinc-binding and Pore-forming toxins family of proteins.

VP4 also contains 64 phosphorylation sites same as VP1. It is the most glycosylated protein and contains 13 sites. Rotavirus VP4 is the surface protein so there is a possibility that number of glycosylation sites modulate the virulence of the virus which can only be proved by *in vivo* studies. It has 3 SUMOylation sites and 1 methylation site. VP4 is predicted to be member of two families, Coat protein and Structural protein.

VP6 has 18 phosphorylation and 4 glycosylation sites. It contains 1 SUMOylation and 1 methylation sites but no acetylation site. VP6

belong to coat protein family which is same with respect to its actual role in virus.

VP7 contains 13 phosphorylation, 1 glycosylation and 2 SUMOylation sites. VP7 has neither methylation nor acetylation sites. According to SVMProt, VP7 belong to transmembrane, coat protein and all lipid-binding proteins.

Conclusion:

Based on the sequence analysis, it is found that rotavirus proteins have many potential sites which may undergo PTM. On the other hand, study of “functional family of proteins” has categorized each of them into different groups. These two aspects of the present study have raised the possibility of moonlighting behavior of rotavirus proteins. Protein phosphorylation results in either modulation of protein activity or effect replication or changes their specificity for interaction with other proteins. Phosphorylation prediction of rotavirus reveals moderate to many phosphorylation sites on different proteins. Further experimental studies on analyzing the role of phosphorylation in different viral proteins will uncover function of this kind of modification. Glycosylation is generally correlated with virulence of viruses. So glycosylated proteins of rotavirus may have some possible roles in their virulence or cellular effects. Similarly SUMOylation, Acetylation and methylation of the proteins may have different roles which depends on the context of its action or interaction with others cellular factors.

Functional family analysis of rotavirus proteins revealed different predicted roles, some of which were known previously, while some are completely novel; proper investigation of these functions in wet laboratory may open a new window for better understanding of rotavirus pathogenesis. Understanding of the role of post-translational modifications in biology of viral infection is one step towards developing successful treatment strategies. Importance of glycosylation, phosphorylation or SUMOylation in viral protein function & virus replication is well established. This knowledge should be exploited in future while designing novel therapeutics, which specifically target one and more crucial PTMs in virus life cycle.

Acknowledgement:

The study was supported by financial assistance from Indian Council of Medical Research (ICMR), New Delhi and the Program of funding research centers for emerging and reemerging infectious diseases (Okayama University-NICED, India) from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

Reference:

- [1] Centers for Disease Control and Prevention (CDC). MMWR Morb Mortal Wkly Rep. 2008 21 : 57: 1255-7 [PMID: 19023263]
- [2] JW Graff *et al. J Gen Virol* 88 : 613-620 (2007) [PMID: 17251580]
- [3] M Lynch *et al. Clinical Infectious Diseases* (2001) Clinical Infectious Diseases 2001; 33: 932-8 [PMID: 11528562]
- [4] CJ Jeffery, Trends Biochem Sci. 1999 Jan; 24(1): 8-11 [PMID: 10087914]
- [5] BG Hale *et al. J Gen Virol*. 2008 Oct; 89: 2359-76 [PMID: 18796704]
- [6] P Kaukinen *et al. Arch Virol*. 2005 Sep; 150:1693-713 [PMID: 15931462]
- [7] J Seishi Murakami, *Gastroenterol*. 2001; 36: 651-660 [PMID: 11686474]
- [8] R Gregory, Reyes, *J. Biomed. Sci.* 2002; 9: 187-197, [PMID: 12065893]
- [9] GC Sahoo *et al. Bioinformatics* 2008; 3: 1-7 [PMID: 19052658]
- [10] D Mitzner *et al. Cell Microbiol*. 2009 Oct; 11: 1502-16. [PMID: 19523156]
- [11] Y Huang *et al. Virology*. 2007 Jul 20; 364: 1-9. [PMID: 17400273]
- [12] DJ Vigerust *et al. Trends Microbiol*. 2007 May; 15: 211-8 [PMID: 17398101]
- [13] C Endter *et al. Proc Natl Acad Sci U S A*. 2001 Sep 25; 98: 11312-7 [PMID: 11553772]
- [14] AL Adamson *et al. J Virol*. 2001 Mar; 75: 2388-99 [PMID: 11160742]
- [15] T Kouzarides, *Cell*. 2007 Feb 23; 128: 693-705 [PMID: 17320507]
- [16] Anton Eberharter *et al. EMBO reports* 3, 3 : 224-229 (2002) [PMID: 11882541]
- [17] CZ Cai *et al. Nucleic Acids Res.* 31: 3692-3697(2003) [PMID: 12824396]
- [18] N Blo *et al. J. Mol. Biol.* 294: 1351-1362, (1999)[PMID: 10600390]
- [19] <http://www.cbs.dtu.dk/services/NetNGlyc/>
- [20] K Julenius *et al. Glycobiology*, 15:153-164, 2005 [PMID: 15385431]
- [21] J Ren *et al. Proteomics.*; 9:3409-3412(2009)[PMID: 19504496]
- [22] D Plewczynski *et al. Bioinformatics*. 15; 21(10):2525-7 (2005) [PMID: 15728119]

Edited by P. Kanguane

Citation: Chattopadhyay *et al. Bioinformatics* 4(10): 448-451 (2010)

License statement: This is an open-access article, which permits unrestricted use, distribution, and reproduction in any medium, for non-commercial purposes, provided the original author and source are credited.

Supplementary material:

Table 1: Types & numbers of PTM and functional classification of Rotavirus proteins

Protein	No & type of Phosphorylation sites	No & type of Glycosylation sites	No & type of SUMOylation sites	No. & type of Methylation sites	No. & type of Acetylation sites	Protein Functional Family (only P-value ≥ 80% are taken)
NSP1 (59 kDa)	Ser= 10, Thr=8 & Tyr= 8	N-linked= 4 O-linked= 1	Type I=0 TypeII=3		N6-acetyl lysine=1	Zinc-binding RNA-binding Metal binding Zinc-binding
NSP2 (35 kDa)	Ser= 7, Thr=2 & Tyr=3	N-linked= 1 O-linked= 0	Type I=1 TypeII=5	Assymmetric Dimethyl arginine=1		Zinc-binding
NSP3 (37 kDa)	Ser= 15, Thr=3 & Tyr= 5	N-linked= 0 O-linked= 0	Type I=1 TypeII=3	Assymmetric Dimethyl arginine=1		Zinc-binding
NSP4 (20 kDa)	Ser= 3, Thr= 5 & Tyr= 1	N-linked= 2 O-linked= 1	Type I=0 TypeII=1	Assymmetric Dimethyl arginine=1		Transmembrane
NSP5 (22 kDa)	Ser= 25, Thr=1 & Tyr= 5	N-linked= 1 O-linked=8	Type I=1 TypeII=4		N6-acetyl lysine=1	Coat protein
VP1 (125 kDa)	Ser= 33, Thr= 10 & Tyr= 21	N-linked= 5 O-linked=0	Type I=2 TypeII=1	Assymmetric Dimethyl arginine=2	N6-acetyl lysine=3	Transferase Structural protein DNA replication All DNA-binding RNA-binding
VP2 (102 kDa)	Ser= 12, Thr= 6 & Tyr= 8	N-linked= 3 O-linked= 1	Type I=1 TypeII=12	Assymmetric Dimethyl arginine=4		Zinc-binding
VP3 (88 kDa)	Ser= 15, Thr=10 & Tyr= 21	N-linked= 9 O-linked= 1	Type I=5 TypeII=5	Assymmetric Dimethyl arginine=2	N6-acetyl lysine=1	Zinc-binding
VP4 (87 kDa)	Ser= 34, Thr= 18 & Tyr= 12	N-linked= 12 O-linked= 1	Type I=2 TypeII=1	Assymmetric Dimethyl arginine=1		Coat protein
VP6 (45 kDa)	Ser= 9, Thr= 6 & Tyr= 3	N-linked= 3 O-linked= 1	Type I=0 TypeII=1	Assymmetric Dimethyl arginine=1		Coat protein
VP7 (34/38 kDa)	Ser= 7, Thr= 3 & Tyr= 3	N-linked= 1 O-linked= 0	Type I=0 TypeII= 2			Transmembrane Coat protein