

Insilico Comprehension of Stop Codon Readthrough in Human Viruses

Arockiyajainmary M¹, Balaji S², Sivashankari Selvarajan³

¹Research Scholar, Bharathiar University, Coimbatore, Tamil Nadu, India

²Associate Professor, Department of Biotechnology, Manipal University, Manipal, Karnataka, India

³Assistant Professor, Nirmala College for Women, Coimbatore, Tamil Nadu, India

ABSTRACT

Readthrough is an event in which stop codon is misread, resulting in elongation of polypeptides. Stop codon suppression or termination codon readthrough, is a mechanism of expression of many disorder proteins. Many important cellular functions are carried out by way of the Readthrough process. This could alter the gene function which thereon produces either destructive or constructive effects. Hence, this study aims to diagnose this recoding mechanism in certain selected humans infecting pathogenic viruses through insilico approach. For this target, the 3'UnTranslated Regions of the selected viruses were retrieved from the Genbank database. Each of these 3'UTRs were translated into all their reading frames. Motif search using Interproscan in each of the frames, followed by homology search using BLASTX, and were achieved to identify stop codon readthrough candidates in each of the selected viruses. Finally, the secondary structure of RNA was predicted using RNAFold web server to ensure the stability of the RNA. The 3'UTRs from Aichi Virus 1, Cosa Virus A, Dengue Virus 1, Duvenhage Lyssavirus, Enterovirus A, HepatitisGB Virus B, Human Cosavirus A, Human Pegivirus 2, Langat Virus, Parechovirus A, WestNile Virus and Zika Virus were retrieved. A total of 48 motifs were identified in different reading frames of 3'UTR of the selected viruses. BlastX search recognized 9 homologs in the reading frames of 3'UTR. The secondary structure analysis and search of motifs and homologs resulted in the confirmation of 5 candidates with strong evidence for the readthrough event. These candidates showed homology with proteins of prime importance such as Imidazole glycerol Phosphate synthase protein, 50S ribosomal protein L27, DNA replication, and repair protein, replication origin-binding protein, and adenosine deaminase. Hence, we proved that the 3'untranslated regions would undergo translation. This strongly suggests that many such readthrough events are to be determined to exactly unravel the pathogenicity behind Viruses. To design anti-viral drugs to impede this viral machinery, it is essential to analyse their 3'UTR regions.

How to cite this paper: Arockiyajainmary M | Balaji S | Sivashankari Selvarajan "Insilico Comprehension of Stop Codon Readthrough in Human Viruses" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-4, June 2020, pp.1711-1719,

URL: www.ijtsrd.com/papers/ijtsrd31550.pdf



IJTSRD31550

Copyright © 2020 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



KEYWORDS: Stop Codon read-through, Viral Genome, 3' UnTranslated Region, Motif, RNA Secondary Structure

INTRODUCTION

Viruses have become an incredible hazard to all the living life forms viz., plants, animals, humans [1]. Millions of people across the globe are experiencing havocing viral sickness such as AIDS/HIV, Ebola, Zika, Polio, Rabies, Dengue fever, Malaria and so forth., Fatality rates due to pandemic viral invasions are still increasing around the globe annually. Complications from multiple infections eventually overwhelm the body and death follows [2], [3]. Recently, a novel corona Virus, nCoV-2019, outbreak in Wuhan, China nearly affects a huge number of individuals over the world. Many of these diseases causing pathogenic viral strains become resistant to the available chemo-therapeutic drugs [4]. We are in peak time to guard ourselves and abate these life-threatening pathogenic viruses.

Viruses are sub-microscopic infective particles comprising of two significant parts namely, an envelope called capsid and a core made up of nucleic acid. 'Virus' signifies 'poison' (La.).

In 1892, Dimitri Ivanovsky was the first to discover virus, a non-bacterial pathogen in contaminated tobacco plants [5]. Virus either have DNA or RNA as their hereditary substance. Viral genes rarely interrupted by introns [6]. World Health Organization (WHO) and the Centre for Disease Control and prevention (CDC) reported that hepatitis B and Dengue fever were one among the most widely recognized viral illness. As indicated by current insights, Hepatitis-B affects around 2 billion individuals every year. Dengue commonly occurs in Africa and Asia. *Aedes aegypti* is responsible for the transmission of dengue virus. Even though this viral infection affects some 50 million people annually, unfortunately there is no specific drug to treat dengue fever. In 2016, WHO reported Zika virus was first identified in Uganda in 1947 in monkeys. In India, 2017, the Ministry of Health and Family Welfare-Government of India (MoHFW) reported three laboratory-confirmed cases of Zika virus disease in Bapunagar area, Ahmedabad District, Gujarat

States [7].

Virus multiplies utilizing the host's genome. Reverse transcriptase permits the viral nucleic acid to integrate with the host genome. When they enter into a cell, it takes over its protein synthesis machinery, assembling several viral particles. Some virulent viruses harm the cells by causing lysis, exhibiting a lytic life cycle. Some viruses persist in the cell for a long time by the integration of its genome with the host genetic material. This process is known as lysogenic life cycle. Thereby they invade living organisms and become a great threat. In order to broaden their protein synthesis, they undergo stop codon readthrough process. Translation of transcript into a polypeptide is high accurate process. Termination of protein synthesis is not 100% effectual [8]. The readthrough mechanism enables the ribosome to pass through the termination codon in mRNA and continues translation to the next stop codon in the reading frame. The translational process proceeds, and peptide chain grows further yielding a nascent protein product with a modified structure which lacks normal functionality. Such extended polypeptides interfere normal cellular processes. The translational read-through arose as a chief regulatory mechanism influencing hundreds of genes. The release factors eRF1 and eRF3 mediate translational termination [9]. The codons UAA, UGA and UAG do not code any amino acids, but acts as translation termination signals. The efficiency of translational termination depends on competition between detection of stop codon by class I release factors and decoding by a near-cognate tRNAs. Jungreis *et al.*, (2016) [10] reported the abundance of stop codon readthrough via evolutionary signatures. The translational readthrough relies on several natural factors, including the nature of termination codon, surrounding mRNA sequence, and presence of stimulating compounds. Arribere *et al.*, (2016) [11] interpreted that translation mechanism would fail to terminate while reaching a stop codon, producing nascent proteins.

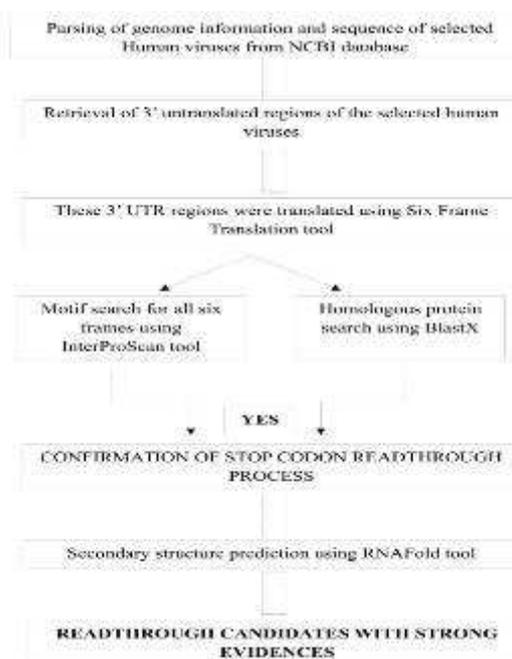
Despite the fact, the known mechanisms of translational surveillance were inadequate to guard cells from potential predominant consequences. Jungreis *et al.*, (2011) [12], employed a comparative genomic study and verified the existence of abundant readthrough event in *Drosophila melanogaster* and suggested that they are functionally significant. Steneberg and Samakovlis (2001) [13] has investigated that translational regulation is the effective means to control the production of polypeptides in *headcase (hdc)* mRNA in *Drosophila*. Such readthrough is required for the function of *hdc* as a branching inhibitor in tracheal development. Prior investigations demonstrated that stop codons are usually suppressed at a frequency of 0.001%–0.1% [14]–[16]. The suppression of stop codon takes place when a near-related aminoacyl-tRNA pair with a stop codon forming codon–anticodon complex. This permits its amino acid to be erroneously incorporated into the peptide sequence and the subsequent extension of translation beyond the termination signal. Several studies showed that a cytidine 3'-adjacent to the stop codon stimulate readthrough process in prokaryotes and eukaryotes [17], [18]. Firth and Brierley (2012) [19] have reviewed over the readthrough mechanism is well-known in viral decoding, particularly RNA viruses and employs extensively to develop their gene expression.

Viral genome consists of 5' UTR or leader sequence, start codon, exon, a stop codon either UAA, UGA or UAG followed by 3' UTR region or trailer sequence. Usually, 3' UTR ranges from 50-250 nucleotides long. The extended translation beyond stop codon thereon produces longer polypeptides with altered functions. Babu *et al.*, (2011) [20] proved that intrinsically disordered proteins (IDPs) are enriched in regulatory functions since they allow interaction with several other proteins and responsible for many diseases. The newer proteins produced were utilized for viral proliferation. 3' end-regions of their stop codon and 3' structural elements are well known inducers of functionally utilized readthrough. In order to block the viral replicating machinery, it is indispensable to analyze their 3'UTR regions. Translational readthrough is similar to alternative splicing [21].

Thus, stop codon readthrough provide an alternative way for organisms to tune their gene expression and functions of their protein products, throughout the lifetime of an individual, which also leads to an evolution of a species. In this study, we employed a computational strategy comprising finding of protein motifs and homologs in 3'UnTranslated Region of virus genes. They further confirmed the read through process by predicting the secondary structure for the RNA *via* computing its stability. The ultimate way of eradicating these viruses is to analyze their genome and to re-design the drug appropriately. The main motive of this work is to predict the stop codon readthrough in 3'untranslated regions of human viral genes through *insilico* approach. And so, we attempted to unravel this readthrough mechanism in human viral genes.

Materials and methods

About 22 human infecting viruses were chosen for the study. The complete genomic data of the viruses were retrieved from NCBI genome database and the 3' Untranslated sequences were extracted. Sequence length of 3'UTR region varies with species. The end of 3'UTR is recognized by a sequence of polyA tail. They were submitted in Six frame translation tool and translated to all reading frames. Then the translated proteins were analyzed to know their functionality. The proteins obtained from all frames were submitted in InterProScan database for the search of motifs. The methodology is depicted in the following flowchart:



Finn *et al.*, (2017) [22] has developed InterPro, a freely available database, the underlying software allows both protein and nucleic acid sequences to be compared against InterPro's predictive models, which are provided by its member databases. The 3' untranslated sequences of the viruses were submitted to NCBI's BlastX tool. The nucleotide query is translated, and the subsequent protein was searched against the non-redundant database, reference protein sequence database and UniProt databases. Resulting homologous proteins predict the structure and functionality of the translated protein. Under appropriate conditions, the RNA folds to form a secondary structure and becomes functional. The functional potential of the protein is determined by its structural stability. Therefore, the stability of the 3'UTR region can be determined using RNAfold webserver.

Tools and Databases

GenBank constitutes an annotated assemblage of all publicly available genome sequences. Altschul *et al.*, (1990) [23] have developed Basic local alignment search tool (BLAST), an approach for rapid sequence comparison, approximates local similarity, the maximal segment pair (MSP) score. InterPro is a web-resource that classifies proteins into families and identifies domains and functional sites. By uniting the member databases, it emerges as a powerful diagnostic tool and integrated resource for functional annotation. Zuker and Stiegler (1981) [24] have developed a tool for folding RNA molecule which finds a minimum free energy confirmation using published stacking values and destabilizing energies, based on dynamic programming algorithm, which is more efficient, faster, and can fold larger molecules. Up to the present, RNA secondary structures have been predicted by applying various topological and thermodynamic rules to find energetically most favorable structure for a given sequence. Hofacker (2003) [25] developed Vienna RNA secondary structure server providing for the analysis of RNA secondary structures. This RNAfold web server predicts secondary structures of single stranded RNA or DNA sequences. The Functional translational readthrough (FTR) creates functional extensions to proteins by continuing translation of the mRNA downstream of the stop codon [26]. Loughran *et al.*, (2018) [27] experimentally investigated the readthrough of Vitamin D receptor (VDR) mRNA in mammalian genes. Efficiency of eukaryotic translational termination is influenced by the nucleotides of Ribosomal mRNA [28]. Therefore, unraveling the disordered complement of proteomes and understanding their function can extend the structure-function paradigm to herald new breakthrough in drug development.

RESULTS AND DISCUSSION

A. Parsing of genomic information of selected human infecting viruses

In accordance with the Baltimore classification of viruses, human viruses belonging to Group I, Group IV and V viruses were taken for the study. 22 human infecting viruses were selected. The complete genomic information of the selected human viruses was retrieved from NCBI genome database and depicted in table-1. The pathology of the selected viruses is gathered from literatures and clinical case reports and were presented in table -2.

B. Retrieval of 3' UTRs of selected human virus

Viral genome consists of genes encoding proteins which are responsible for replication and structural/Non-structural

components. A total of 12 viruses including Aichi virus 1, Cosa virus A, Dengue virus 1, Duvenhage lyssavirus, Enterovirus A, Hepatitis GB virus B, Human cosavirus, Human pegivirus 2, Langat virus, Parechovirus, WestNile virus and Zika virus were analyzed for a stop codon read through. The genomic data of these 12 viruses *viz.*, number of genes, their coding context, 3' untranslated regions, gene ID, protein product and their protein ID were retrieved and depicted in table - 3. The length of the 3'UTR ranges between 52 and 630. Longest 3' UTR is seen in WestNile virus having 630 bases. The nucleoprotein gene of duvenhage lyssavirus has the smallest 3'UTR of 52 bases.

C. Identification of stop codon read through candidates

Initially, the 3' untranslated regions of the chosen human viruses were translated into all reading frames. The corresponding protein sequences obtained were submitted to Interproscan tool for the search of motifs. Table - 4 shows the results obtained from InterProScan. The obtained proteins were found to be intrinsically disordered protein, Cytoplasmic domain, Non-cytoplasmic domain, TM helix and signal peptides. Disorder protein is one which does not have a defined 3D Structure [29], [30], yet it is involved in signaling and regulatory functions [31]. These are highly conserved over species [32]. Most of the 3'UTR of viral genes were predicted to contain cytoplasmic domains which are reported to engage in multiple roles such as viral replication and cell-cell spread in Herpes simplex virus [33]. Similarly, disordered proteins identified in Zika virus were found to perform particle formation and replication [34]. These studies serve as strong evidence for proteins which explored in viruses have occurred due to readthrough process. To further confirm the readthrough, homologous protein search for the genes encoded by the virus was performed. The readthrough process leads to error in translational process [35]. Along with the tRNAs, release factors are also tangled in readthrough events [36]. Stop codon is the key determinant of translation termination in both prokaryotes and eukaryotes [37]. Intrinsic disorder-focused investigation of viral proteome is significant for the development of disorder-based drugs. Petroczy *et al.*, (2017) [38] reported that the protein flexibility ranges from simple hinge movements to functional disorder. Tompa *et al.*, (2014) [39] has analyzed that majority of the bio-molecular interactions in all cellular processes is mediated by compact segments, referred to as motifs. Such motifs are typically less than ten residues in length, occur intrinsically disordered regions, and are post-translational modified by structured domains of the interacting partner. They enable both high functional diversity and functional density to polypeptide domains containing them.

D. Homologous protein search

The results of homologous protein search through BLASTX are presented in table - 5. Through BlastX 9 homologous proteins to the stop codon readthrough candidates were identified.

The +3 frame of the 3'UTR of the Polyprotein of Aichi virus 1 share 46% identity with Imidazole glycerol phosphate synthase subunit HisF [IGPS] of *Acidovorax citrulli with a query coverage of 30%*. The Imidazole glycerol phosphate synthase subunit HisF has imidazoleglycerol-phosphate synthase activity and lyase activity. The -3 frame of the *Duvenhage lyssavirus* matrix protein gene has a homologous

protein of Nitrate ABC transporter substrate-binding protein in *Corynebacterium stationis* with an identity of 41% with the highest query coverage of 72%. Its +1 frame of the glycoprotein gene has a homologous protein of 50S ribosomal protein L27 in *Helicobacter pylori* with an identity of 41%. It is a ribonucleo protein which aids in translation process. The -1 frame of the 3'UTR of Dengue virus 1 has a homologous protein of 3-isopropylmalate dehydratase large subunit 2 from *Deinococcus radiodurans* with an identity of 29%. They have 3-isopropylmalate dehydratase activity, metal ion binding and lyase activity. The +1 frame of the 3'UTR of the polyprotein from Enterovirus A has a homologous protein of Putative fimbrial assembly protein FimD, serogroup D of *Dichelobacter nodosus* with an identity of 57%. It is involved in fimbrium biogenesis. The -1 frame of the 3'UTR of the polyprotein from the Human pegivirus 2 has a homologous protein, DNA replication and repair protein RecF in *Rickettsia massiliae* with an identity of 35%. It is involved in DNA replication and repair mechanism and has single-stranded DNA binding ability. The -2 frame of the 3'UTR of the polyprotein from Langat virus has a homologous protein, UDP-glucuronosyltransferase 2B7 precursor, putative of *Pediculus humanus corporis* with an identity of 31%. It is involved in metabolic process and has glucuronosyltransferase activity. The +3 frame of the 3'UTR of flavivirus polyprotein from West Nile virus has a homologous protein, Replication origin-binding protein in *Equine herpesvirus 1* with an identity of 45%. This protein functions as a docking protein to recruit essential components of the viral replication machinery to viral DNA origins. The +1 frame of the 3'UTR from Zika virus has a homologous protein, Adenosine deaminase of *Arthrobacter sp. H14* with an identity of 37% and query coverage of 45%. It has the Adenosine deaminase activity. Thus, we found that the translated 3' UTR were probable to produce these similar proteins.

E. Prediction of secondary structure

Hofacker and Stadler (2008) [40] have reported that the RNA folding can be regarded as a hierarchical process in which secondary structure forms before tertiary structure. Secondary structures are highly conserved in evolution for many classes of RNA molecules. *The understanding of structure of a biomolecule provides a principle way to know about its function and so the secondary structure of the RNA was predicted using RNAfold tool. The selected 12 human viruses have single-stranded RNA as their genetic material. From the above analysis, it was proven that the viruses undergo stop codon translational readthrough beyond the 3' UTR regions and encodes some proteins in certain reading frames. To know its functional ability, the stability of the structure is predicted. The stable biomolecule could perform various cellular functions. These 3' non-coding regions of the chosen 12 human virus's genome forms the secondary structural components such as stem, pseudoknots, bulge loops, hairpin loops, multi-loops with branches and internal loops. The energy values of the predicted RNA structure are presented in table - 6. The stability of the structure was calculated using the free energy calculation of thermodynamic ensemble. It could be found from the table that the RNA predicted from the 3'UTR of Polyprotein gene from Langat virus has the lowest free energy of -221.30. The RNA from 3'UTR of flavivirus polyprotein from West Nile virus (-203.90 kcal/mol), flavivirus polyprotein from Zika virus (-171.90 kcal/mol), flavivirus polyprotein from Dengue virus (-152.40*

kcal/mol), polyprotein from Human pegivirus 2 (-144.80 kcal/mol), flavivirus polyprotein from Hepatitis GB virus B (-136.10 kcal/mol), glycoprotein from Duvenhage lyssavirus (-117.50 kcal/mol) and polyprotein from Aichi virus 1 (-70.70 kcal/mol) follow the order. The predicted RNA structures are displayed in Figure below:

F. Strong Read through candidates

The viruses may undergo translation afar 3' untranslated regions and encode certain proteins that play inevitable role. The existence of motifs and homologs for the 3'UTR along with a stable RNA structure is strong evidence that read through process has occurred. In this way, we identified 5 strong readthrough candidates in the 12 selected human viruses. In Aichi virus 1, the +3 frame of the 3'UTR of polyprotein gene had a disorder protein motif and was homologous to Imidazole glycerol Phosphate synthase protein. The free energy of its predicted RNA was found to be -70.70 kcal/mol confirming readthrough process. In Duvenhage lyssavirus, the +1 frame of the 3' UTR region of the glycoprotein gene has motifs such as cytoplasmic domain, non-cytoplasmic domain, TM helix and transmembrane region. The BlastX homologous search found a hit sequence, 50S ribosomal protein L27 of *Helicobacter pylori* in the same frame. *Its predicted RNA structure has a free energy of -117.50 kcal/mol and thus a strong evidence for readthrough process. The -1 frame of the 3'UTR of polyprotein gene from Human pegivirus 2 has a motif and a homolog DNA replication and repair protein. Its RNA secondary structure has a free energy of -144.80 kcal/mol which is a stable structure. Hence, polyprotein 3'UTR from Human pegivirus is a strong read through candidate. The +3 frame of the 3'UTR of flavivirus polyprotein from West Nile virus has a motif and a homolog, replication origin-binding protein. They have a stable RNA secondary structure has a free energy of -203.90 kcal/mol. Hence, flavivirus polyprotein 3'UTR from Human pegivirus is a strong read through candidate. The motifs predicted in the Zika virus are non-cytoplasmic domain, signal peptide C-region, signal peptide H-region, signal peptide N-region, signalP-TM, signalP- noTM and mobiDB-lite (disorder prediction) in +1 frame. The homolog found in this frame is adenosine deaminase of *Arthrobacter sp. H14*. Its RNA secondary Structure has a free energy of -171.90 kcal/mol and is a stable structure. Thus, 3'UTR of the flavivirus polyprotein from Zika virus is a strong read through candidate.*

CONCLUSIONS

It is evident that the viruses suppress the termination signals and extends their translation process along the 3' non-coding region. The stop codon readthrough candidates were confirmed in all the above investigation. The newer proteins synthesized were utilized for the viral replication and virulence potential. The present work has resulted in the identification of 5 readthrough candidates in selected Human viruses the method had three levels of investigation namely motif, homolog and RNA secondary structure for the occurrence of stop codon readthrough process. More importantly, the homology of these candidates to other species indicates that the products are functional. Further, the readthrough candidates can be evaluated by *in vivo* experiments that prove the presence of C-terminus proteins that correspond to first and second stop codons. This method could identify readthrough candidates for only the known motifs. As the number of known motifs increase,

there will be an increase in number of readthrough candidates. Similarly, the readthrough candidates have occurred in the different reading frame from that of the first stop codon which provides a strong evidence for readthrough mechanism. Viruses employs stop codon readthrough as an alternative translational strategy. The analysis of conserved protein coding signatures that extend beyond annotated stop codon could predict readthrough genes, all of which could be validated experimentally. Finally, the functional importance of the readthrough candidates gives more clues to prove the expression of these readthrough events. In future anti-viral drugs could be designed to suppress this readthrough mechanism and against their synthesized abnormal virulent proteins could serves as an eternal solution to eradicate viral invasions. the method had three levels of evidence namely motif, homolog and RNA secondary structure for the occurrence of stop

codon readthrough process. More importantly, the homology of these candidates to other species indicates that the products are functional. Further, the readthrough candidates can be evaluated by *in vivo* experiments that prove the presence of C-terminus proteins that correspond to first and second stop codons.

This method could identify readthrough candidates for only the known motifs. As the number of known motifs increase, there will be a increase in the number of readthrough candidates. Similarly, the readthrough candidates have occurred in the different reading frame from that of the first stop codon which provides a strong evidence for frameshift mechanism. Finally, the functional importance of the readthrough candidates gives more clues to prove experimentally the expression of these readthrough events.

SUPPLEMENTARY TABLES:

Table – 1: Human infecting viruses

| S. No | Organism | Definition | Accession No | Classification (Family) | Type of Genome | Genome Length (base pairs) | No of Genes |
|-------|--|---|--------------|-------------------------|----------------|----------------------------|-------------|
| 1. | Aichi virus 1 | Aichi virus genomic RNA, complete genome | AB040749.1 | Picornaviridae | (+) ssRNA | 8280 | 1 |
| 2. | Cosa virus A | Cosavirus A strain HCoV-A1 polyprotein gene, complete cds | NC_012800.1 | Picornaviridae | (+)ssRNA | 7632 | 1 |
| 3. | Dengue virus 1 | Dengue virus 1, complete genome | NC_001477.1 | Flaviviridae | (+)ssRNA | 10735 | 1 |
| 4. | Duvenhage lyssavirus | Duvenhage virus isolate 86132SA, complete genome | NC_020810.1 | Rhabdoviridae | (-)ssRNA | 11976 | 5 |
| 5. | Enterovirus A | Human enterovirus A, complete genome | NC_001612.1 | Picornaviridae | (+)ssRNA | 7413 | 1 |
| 6. | Hepatitis GB virus B | Hepatitis GB virus B, complete genome | NC_001655.1 | Flaviviridae | (+) ssRNA | 9399 | 1 |
| 7. | Human cosavirus | Human cosavirus isolate Cosavirus Amsterdam 1994, complete genome | NC_023984.1 | Picornaviridae | (+)ssRNA | 7802 | 1 |
| 8. | Human pegivirus 2 | Human pegivirus 2 isolate UC0125.US | NC_027998.2 | Flaviviridae | (+)ssRNA | 9867 | 1 |
| 9. | Langat virus | Langat virus, complete genome | NC_003690.1 | Flaviviridae | (+)ssRNA | 10943 | 1 |
| 10. | Mayaro virus | Mayaro virus, complete genome | NC_003417.1 | Togaviridae | (+)ssRNA | 11411 | 4 |
| 11. | Middle East respiratory syndrome-related coronavirus | Middle East respiratory syndrome corona virus, complete genome | NC_019843.3 | Coronaviridae | (+)ssRNA | 30119 | 10 |
| 12. | Murray Valley encephalitis virus | Murray Valley encephalitis virus, complete genome | NC_000943.1 | Flaviviridae | (+)ssRNA | 11014 | 2 |
| 13. | Molluscum contagiosum virus subtype 1 | Molluscum contagiosum virus subtype 1, complete genome | NC_001731.1 | Poxviridae | dsDNA | 190289 | 163 |
| 14. | Ockelbo virus | Ockelbo virus strain Edsbyn, complete genome | M69205.1 | Togaviridae | (+)ssRNA | 11708 | 2 |
| 15. | Parechovirus A | Human parechovirus, genome | NC_001897.1 | Picornaviridae | (+)ssRNA | 7348 | 1 |
| 16. | Primate norovirus | Primate norovirus, complete genome | NC_031324.1 | Caliciviridae | (+)ssRNA | 7753 | 3 |
| 17. | Ross River virus | Ross River virus, complete genome | NC_001544.1 | Togaviridae | (+)ssRNA | 11657 | 3 |
| 18. | Sagiyama virus | Sagiyama virus genomic RNA, complete genome | AB032553.1 | Togaviridae | (+)ssRNA | 11698 | 3 |
| 19. | Sapovirus C12 | Sapovirus C12 strain C12 | NC_006554.1 | Caliciviridae | (+)ssRNA | 7476 | 2 |
| 20. | Sputnik virophage | Sputnik virophage, complete genome | NC_011132.1 | Lavidaviridae | DNA | 18343 | 21 |
| 21. | West Nile virus | West Nile virus lineage 1, complete genome | NC_009942.1 | Flaviviridae | (+)ssRNA | 11029 | 1 |
| 22. | Zika virus | Zika virus isolate ZIKV/Monkey/Uganda/MR766/1947, complete genome | NC_012532.1 | Flaviviridae | (+)ssRNA | 10794 | 1 |

Table – 2: Pathology of human infecting viruses

| S. No | Organism | Disease | Transmission | No of proteins |
|-------|--|---|------------------------------|----------------|
| 1. | Aichi virus 1 | Gastroenteritis | Fecal-oral | 1 |
| 2. | Cosa virus A | Gastroenteritis, Co-infection with HIV | Fecal-oral | 1 |
| 3. | Duvenhage lyssavirus | Rabies-like encephalitic illness | Insectivorous Bat bite | 5 |
| 4. | Dengu virus 1 | Hemorrhagic fever, flu-like illness | Mosquito-borne | 1 |
| 5. | Enterovirus A | Diarrhea, neurological disorder | Fecal-oral | 1 |
| 6. | Hepatitis GB virus B | Hepatitis | Sexual contact, Blood | 1 |
| 7. | Human cosavirus | Gastroenteritis | Fecal-oral | 1 |
| 8. | Human pegivirus 2 | Hepatitis, Co-infection with HIV | Blood borne | 1 |
| 9. | Langat virus | Encephalitis | Zoonosis, arthropod borne | 1 |
| 10. | Mayaro virus | Fever, joint pain | Zoonosis, arthropod borne | 4 |
| 11. | Middle East respiratory syndrome related-coronavirus | Respiratory | Zoonosis | 10 |
| 12. | Molluscum contagiosum virus sugtype 1 | Skin lesions | Contact | 163 |
| 13. | Murray valley encephalitis virus | Encephalitis | Zoonosis, arthropod bite | 2 |
| 14. | Ockelbo virus | Exanthema, Arthralgia | Mosquito bite | 2 |
| 15. | Parechovirus A | Respiratory illness, myocarditis, encephalitis | Zoonosis | 1 |
| 16. | Primate norovirus | Gastroenteritis | Fecal-oral | 3 |
| 17. | Ross river virus | Fever, joint pain | Zoonosis, arthropod bite | 3 |
| 18. | Sagiyama virus | Fever, joint pain | Zoonosis, arthropod bite | 3 |
| 19. | Sapo virus C12 | Gastroenteritis | Fecal-oral | 2 |
| 20. | Sputnik virophage | Pneumonia like illness | Fecal-oral | 21 |
| 21. | West Nile virus | Hemorrhagic fever, encephalitis | Zoonosis, arthropod bite | 1 |
| 22. | Zika virus | Fever, skin rash, conjunctivitis, muscle and joint pain, malaise or headache. Microcephaly and Guillain-Barré syndrome | Mosquito-borne | 1 |

Table – 3: Genomic data on 3'UTR of selected Human viruses

| AICHI VIRUS 1 | | | | | | |
|------------------------------|-------------|--------------|--------------|-----------------|---------------------------|--------------------|
| No of genes | Regions | | 3'UTR Length | GENE ID | PROTEIN PRODUCT | PROTEIN ID |
| | CDS | 3' UTR | | | | |
| 1 | 745..8043 | 8044..8280 | 236 | AB04074 9.1 | Polyprotein | BAB62889.1 |
| COSA VIRUS A | | | | | | |
| 1 | 1165..7539 | 7540..7632 | 92 | 7986820 | Polyprotein | YP_00295607 4.1 |
| DENGUE VIRUS 1 | | | | | | |
| 1 | 95..10273 | 10274..10735 | 461 | 5075725 | Flavivirus polyprotein | NP_059433.1 |
| DUVENHAGE LYSSAVIRUS | | | | | | |
| 5 | 71..1426 | 1427..1479 | 52 | 14857942 | Nucleoprotein | YP_00764140 2.1 |
| | 1517..2413 | 2414..2470 | 56 | 14857938 | Phosphoprotein | YP_00764140 3.1 |
| | 2497..3105 | 3106..3259 | 153 | 14857939 | Matrix protein | YP_00704140 4.1 |
| | 3297..4898 | 4899..5386 | 487 | 14857940 | Glycoprotein | YP_00764140 5.1 |
| | 5462..11845 | 11846..11900 | 54 | 14857941 | Polymerase | YP_00764140 6.1 |
| Genome 3' UTR (11907..11976) | | | | | | |
| ENTEROVIRUS A | | | | | | |
| 1 | 751..7332 | 7333..7413 | 80 | 1461111 | polyprotein | NP_042242.1 |
| HEPATITIS GB VIRUS B | | | | | | |
| 1 | 446..9040 | 9041..9399 | 358 | NC_001655 .1 | Flavivirus polyprotein | NP_056931.1 |
| HUMAN COSAVIRUS A | | | | | | |
| 1 | 1362..7727 | 7728..7802 | 74 | 1403460 | Polyprotein | YP_0090263 76.1 |
| HUMAN PEGIVIRUS 2 | | | | | | |
| 1 | 328..9501 | 9502..9867 | 365 | 26044861 | Polyprotein | YP_0092272 95.1 |
| LANGAT VIRUS | | | | | | |
| 1 | 131..10375 | 10376..10943 | 567 | 940444 | Polyprotein | NP_620108.1 |
| PARECHOVIRUS A | | | | | | |
| 1 | 703..7242 | 7243..7329 | 86 | 1403455 | Polyprotein | NP_0468804. 1 |
| WEST NILE VIRUS | | | | | | |
| 1 | 97..10398 | 10399..11029 | 630 | 5714902 | Flavivirus polyprotein | YP_00152787 7.1 |
| ZIKA VIRUS | | | | | | |
| 1 | 107..10366 | 10367..10794 | 427 | 7751225 | Flavivirus polyprotein | YP_00279088 1.1 |

Table - 4: Results of InterProScan

| S. No | Organism Name | Protein Name (Gene ID) | 3' UTR region | Frame | MOTIF NAME |
|-------|----------------------|--------------------------------|---------------|-------|---|
| 1. | Aichi virus 1 | Polyprotein AB040749.1 | 8044..8280 | +3 | Disorder protein |
| | | | | -1 | Disorder protein |
| | | | | -2 | Disorder protein |
| | | | | -3 | Disorder protein |
| 2. | Cosa virus A | Polyprotein 7986820 | 7540..7632 | +1 | Cytoplasmic domain, Non-cytoplasmic domain, TMhelix, Transmembrane region |
| | | | | +2 | Cytoplasmic domain, Non-cytoplasmic domain, TMhelix, Transmembrane region |
| | | | | +3 | Cytoplasmic domain, Non-cytoplasmic domain, TMhelix, Transmembrane region |
| | | | | -1 | Disorder protein |
| 3. | Dengue virus 1 | Flavivirus polyprotein 5075725 | 10274..10735 | +1 | Disorder protein |
| | | | | +2 | Disorder protein |
| | | | | +3 | Disorder protein |
| | | | | -2 | Cytoplasmic domain, Non-cytoplasmic domain, TMhelix, Transmembrane region |
| | | | | -3 | Cytoplasmic domain, Non-cytoplasmic domain, TMhelix, Transmembrane region |
| 4. | Duvenhage lyssavirus | Nucleoprotein 14857942 | 1427..1479 | - | - |
| | | Phosphoprotein 14857938 | 2414..2470 | - | - |
| | | Matrix protein 14857939 | 3106..3259 | -1 | Cytoplasmic domain, Non-cytoplasmic domain, TMhelix, Transmembrane region |
| | | | | -2 | Cytoplasmic domain, Non-cytoplasmic domain, Transmembrane region. |
| | | Glycoprotein 14857940 | 4899..5386 | +1 | Cytoplasmic domain, Non-cytoplasmic domain, TMhelix, Transmembrane region |
| | | | | +3 | Cytoplasmic domain, Non-cytoplasmic domain, Transmembrane region |
| | | | | -1 | SignalP-no TM |
| | | | | -2 | Cytoplasmic domain, Non-cytoplasmic domain, TMhelix, Transmembrane region |
| | | Polymerase 14857941 | 11846..11900 | - | - |
| | | Genome 3' UTR | 11907..11976 | -1 | Disorder protein |
| -2 | Disorder protein | | | | |

Table -5: BlastX results

| S. No | Organism | Protein Gene ID | Homolog | Frame | Query Coverage | E-value | Identity |
|--------------------------|-----------------------|------------------------------------|---|-------|----------------|---------|----------|
| 1. | Aichi virus 1 | Polyprotein AB040749.1 | Imidazole glycerol phosphate synthase subunit HisF [<i>Acidovorax citrullii</i>] AITL06.1 | +3 | 30% | 0.72 | 46% |
| 2. | Cosa virus A | Polyprotein 7986820 | - | - | - | - | - |
| 3. | Dengue virus 1 | Flavivirus polyprotein 5075725 | 3-isopropylmalate dehydratase large subunit 2 [<i>Deinococcus radiodurans</i>] Q9RT16.1 | -1 | 33% | 0.27 | 29% |
| 4. | Duvenhage lyssa virus | Nucleoprotein 14857942 | - | - | - | - | - |
| | | Phosphoprotein 14857938 | - | - | - | - | |
| | | Matrix protein 14857939 | Nitrate ABC transporter substrate-binding protein [<i>Corynebacterium stationis</i>] OAH30059.1 | -3 | 72% | 2.8 | 41% |
| | | Glycoprotein 14857940 | 50S ribosomal protein L27 [<i>Helicobacter pylori</i>] Q9ZMD8.1 | +1 | 24% | 8.9 | 41% |
| | | Polymerase 14857941 | - | - | - | - | - |
| Genome 3'UTR NC_020810.1 | - | - | - | - | - | - | |
| 5. | Enterovirus A | Polyprotein 1461111 | Putative fimbrial assembly protein FimD, serogroup D [<i>Dichelobacter nodosus</i>] P17420.1 | +1 | 51% | 8.5 | 57% |
| 6. | Hepatitis GB virus B | Flavivirus polyprotein NC_001655.1 | - | - | - | - | - |
| 7. | Human cosavirus | Polyprotein 1403460 | - | - | - | - | - |
| 8. | Human pegivirus 2 | Polyprotein 26044861 | DNA replication and repair protein RecF [<i>Rickettsia massiliae</i>] A8F0D4.1 | -1 | 30% | 2.8 | 35% |
| 9. | Langat virus | Polyprotein 940444 | UDP-glucuronosyl transferase 2B7 precursor, putative [<i>Pediculus humanus corporis</i>] XP_002432886.1 | -2 | 45% | 1.7 | 31% |
| 10. | Parechovirus | Polyprotein 1403455 | - | - | - | - | - |
| 11. | West Nile virus | Flavivirus polyprotein 5714902 | Replication origin-binding protein [<i>Equine herpesvirus 1</i>] P28947.1 | +3 | 13% | 6.8 | 45% |
| 12. | Zika virus | Flavivirus polyprotein 7751225 | Adenosine deaminase [Arthrobacter sp. H14] WP_026534345.1 | +1 | 45% | 0.31 | 37% |

Table - 6: Thermodynamic Free energies of predicted RNA structures

| S. No | Organism | Protein (Gene ID) | Free energy (kcal/mol) | | |
|-------|----------------------|------------------------------------|------------------------|----------|--------------------|
| | | | MFE structure | Ensemble | Centroid structure |
| 1. | Aichi virus 1 | Polyprotein AB040749.1 | -70.70 | -73.96 | -62.20 |
| 2. | Cosa virus A | Polyprotein 7986820 | -14.00 | -16.17 | -9.60 |
| 3. | Dengue virus 1 | Flavivirus Polyprotein 5075725 | -152.40 | -157.42 | -113.30 |
| 4. | Duvenhage lyssavirus | Matrix Protein 14857939 | -25.90 | -28.13 | -17.80 |
| | | Glycoprotein 14857940 | -117.50 | -125.90 | -99.10 |
| | | Genome 3'UTR | -3.10 | -5.25 | -1.70 |
| 5. | Enterovirus | Polyprotein 1461111 | -22.20 | -23.82 | -22.20 |
| 6. | Hepatitis GB virus B | Flavivirus Polyprotein NC_001655.1 | -136.10 | -142.29 | -116.20 |
| 7. | Human cosavirus | Polyprotein 1403460 | -12.10 | -14.07 | -12.10 |
| 8. | Human pegivirus 2 | Polyprotein 26044861 | -144.80 | -149.80 | -106.60 |
| 9. | Langat virus | Polyprotein 940444 | -221.30 | -227.37 | -205.34 |
| 10. | Parechovirus | Polyprotein 1403455 | -13.00 | -14.79 | -13.00 |
| 11. | West Nile virus | Flavivirus Polyprotein 5714902 | -203.90 | -213.56 | -176.10 |
| 12. | Zika virus | Flavivirus Polyprotein 7751225 | -171.90 | -175.31 | -168.60 |

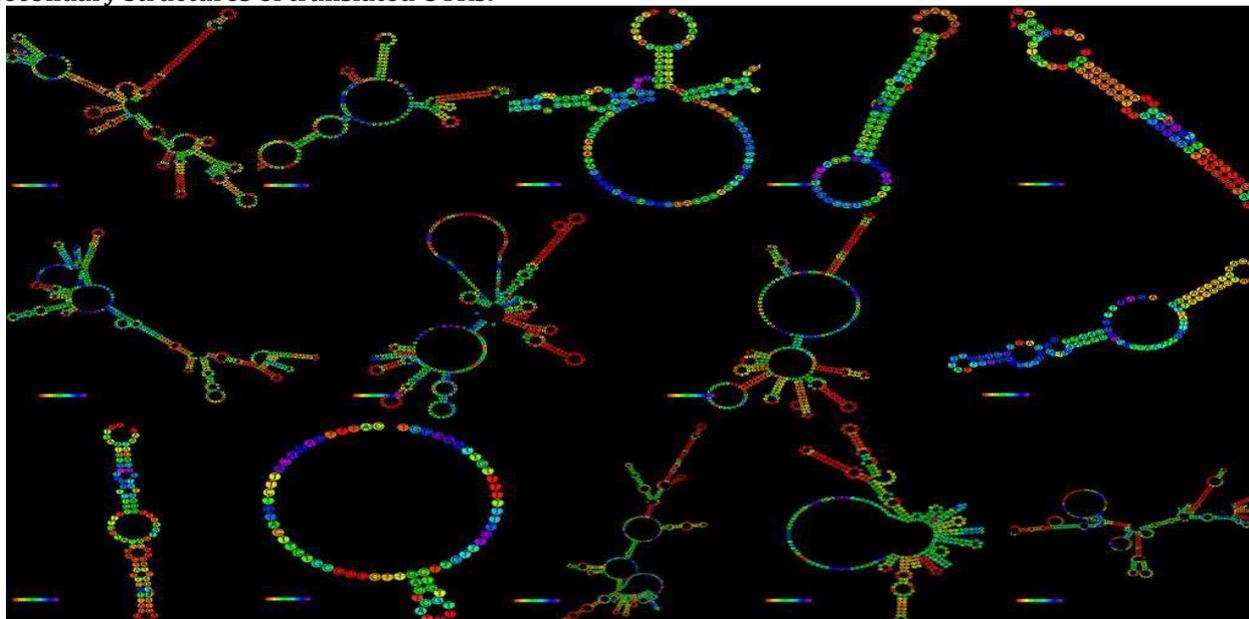
RNA secondary structures of translated UTRs:

Figure (labelled along left to right) – RNA secondary structure of 3,UTR of flavivirus polyprotein gene from Zika virus, Aichi virus, Matrix protein gene from Duvenhage lyssa virus, Popyprotein gene from Enterovirus A, Polypeptide gene from human Cosavirus, Glycoprotein gene from Duvenhage Lyssa virus, Flavivirus polyprotein gene from Dengue virus 1, Flavivirus polyprotein gene from Hepatitis GB virus B, Polypeptide gene from Cosavirus, Polypeptide gene from Parechovirus, Duvenhage Lyssa viral genome, flavivirus polyprotein gene from HepatitisGB virus B, polypeptide gene from Human pegivirus, polypeptide gene from Langat virus.

FUNDING: This research received no external funding.

CONFLICTS OF INTEREST: The authors declare no conflict of interest.

REFERENCES:

- [1] Institute of Medicine (US) Forum on Microbial Threats. Microbial Evolution and Co-Adaptation: A Tribute to the Life and Scientific Legacies of Joshua Lederberg: Workshop Summary. Washington (DC): National Academies Press (US); 2009. 5, Infectious Disease Emergence: Past, Present, and Future. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK45714/>
- [2] Rothberg, Michael B., and Sarah D. Haessler. 2010. "Complications of Seasonal and Pandemic Influenza." *Critical Care Medicine* 38 (SUPPL. 4): 91–97. <https://doi.org/10.1097/CCM.0b013e3181c92eeb>.
- [3] Holmes KK, Bertozzi S, Bloom BR, et al., editors. Major 329 Infectious Diseases. 3rd edition. Washington (DC): The International Bank for Reconstruction and Development / The World Bank; 2017 Nov 3. Chapter 1. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK525197/> doi: 10.1596/978-1-4648-0524-0/ch1
- [4] Tacconelli, E., Carrara, E., Savoldi, A., Harbarth, S., Mendelson, M., Monnet, D. L., Zorzet, A. (2018). Discovery, research, and development of new antibiotics: the WHO priority list of antibiotic-resistant bacteria and tuberculosis. *The Lancet Infectious Diseases*, 18(3), 318–327. doi: 10.1016/S1473-3099(17)30753-3
- [5] Gelderblom, H. R. (1996). Structure and Classification of Viruses. *Medical Microbiology*, (January 1996). Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/21413309>
- [6] Viral Genomes - Molecular Structure, Diversity, Gene Expression Mechanisms and Host-Virus Interactions. (2012). In *Viral Genomes - Molecular Structure, Diversity, Gene Expression Mechanisms and Host-Virus Interactions*. doi: 10.5772/1346
- [7] Mourya, Devendra T et al. "Emerging/re-emerging viral diseases & new viruses on the Indian horizon." *The Indian journal of medical research* vol. 149, 4 (2019): 447-467. doi:10.4103/ijmr.IJMR_1239_18
- [8] Dabrowski M, Bukowy-Bieryllo Z and Zietkiewicz E (2015), Translational readthrough potential of natural termination codons in eucaryotes – The impact of RNA sequence, *RNA Biology* 12:9, Sep 2015, P. 950—958.
- [9] Salas-Marco, J., & Bedwell, D. M. (2004). GTP hydrolysis by eRF3 facilitates stop codon decoding during eukaryotic translation termination. *Molecular and cellular biology*, 24(17), 7769–7778. doi:10.1128/MCB.24.17.7769-7778.2004.
- [10] Jungreis L, S. Chan C, M. Waterhouse 1,2,3,4 Gabrie Robert, 5 Michael F. Lin,6, and kellis. M (2016), Evolutionary Dynamics of Abundant Stop Codon Readthrough, September 7, 2016, Evolutionary Dynamics of Readthrough, P. 3108–3132.
- [11] A. Arribere, J, S. Cenik, E, Jain, N, T. Hess, G, H. Lee, C, C. Bassik, M, and Z. Fire, A (2016), Translation Readthrough Mitigation, June 30, 2016, doi:10.1038/nature18308, P. 719–723.
- [12] Jungreis, I, Lin, M. F., Spokony, R., Chan, C. S., Negre, N., Victorsen, A., White, K. P. and Kellis, M. (2011) Evidence of abundant stop codon readthrough in *Drosophila* and other metazoa. *Genome Res.*, 21, 2096–2113.
- [13] Par Steneberg and Christos Samakovlis (2001), a Novel stop codon readthrough mechanism produces functional *Headcase* protein in *Drosophila trachea*, *EMBO reports* Vol.2/no.7, Apr 2001, P. 593-597.

- [14] Loftfield, R. B. and Vanderjagt, D. 1972. The frequency of errors in protein biosynthesis. *Biochem. J.* 128: 1353–1356.
- [15] Mori, N., Funatsu, Y., Hiruta, K., and Goto, S. 1985. Analysis of translational fidelity of ribosomes with protamine messenger RNA as a template. *Biochemistry* 24: 1231–1239
- [16] Stansfield, I., Jones, K. M., Herbert, P., Lewendon, A., Shaw, W.V., and Tuite, M. F. 1998. Missense translation errors in *Saccharomyces cerevisiae*. *J. Mol. Biol.* 282: 13–24.
- [17] Brown, C. M., Stockwell, P. A., Trotman, C. N. and Tate, W. P. (1990a) Sequence analysis suggests that tetra-nucleotides signal the termination of protein synthesis in eukaryotes. *Nucleic Acids Res.*, 18, 6339–45.
- [18] Brown, C. M., Stockwell, P. A., Trotman, C. N. and Tate, W. P. (1990b) The signal for the termination of protein synthesis in prokaryotes. *Nucleic Acids Res.*, 18, 2079–86
- [19] Firth, A. E. and Brierley, I. (2012) Non-canonical translation in RNA viruses. *J. Gen. Virol.*, 93, 1385–409.
- [20] Babu M, Lee R, Groot N and Gsponer J (2011), Intrinsically disordered proteins: regulation and disease, *Current Opinion in Structural Biology* volume 21, Issue 3, June 2011, Pages 432-440.
- [21] Pancsa, R., Macossay-Castillo, M., Kosol, S., & Tompa, P. (2016). Computational analysis of translational readthrough proteins in *Drosophila* and yeast reveals parallels to alternative splicing. *Scientific reports*, 6, 32142. doi:10.1038/srep32142.
- [22] Finn R, Teresa K. Attwood, Patricia C. Babbitt, Alex Bateman, Peer Bork, Alan J. Bridge, Hsin-Yu Chang, Zsuzsanna Dosztanyi, Sara El-Gebali, Matthew Fraser, Julian Gough, David Haft, Gemma L. Holliday, Hongzhan Huang, Xiaosong Huang, Ivica Letunic, Rodrigo Lopez, Shennan Lu, Aron Marchler-Bauer, Huaiyu Mi, Jaina Mistry, Darren A Natale, Marco Necci, Gift Nuka, Christine A. Orengo, Youngmi Park, Sebastien Pesseat, Damiano Piovesan, Simon C. Potter, Neil D. Rawlings, Nicole Redaschi, Lorna Richardson, Catherine Rivoire, Amaia Sangrador-Vegas, Christian Sigrist, Ian Sillitoe, Ben Smithers, Silvano Squizzato, Granger Sutton, Narmada Thanki, Paul D Thomas, Silvio C. E. Tosatto, Cathy H. Wu, Ioannis Xenarios, Lai-Su Yeh, Siew-Yit Young and Alex L. Mitchell, (2016), InterPro in 2017—beyond protein family and domain annotations, *Nucleic Acids Research*, 2017, Vol. 45, Database issue, D190–D199 Published online 28 November 2016.
- [23] Altschul F, Gish W, Miller W, Myers E and Lipman D, (1990), Basic local alignment search tool, *Journal of Molecular Biology* volume 21, Issue 3, 5 October 1990, Pages 403-410.
- [24] Zuker M and Stiegler P, (1981), Optimal computer folding of lare RNA sequences using thermodynamics and auxiliary information, *Nucleic Acids Research* Volume 9 Number 1, 1981, November 1980.
- [25] Hofacker I (2003), Vienna RNA secondary structure server, *Nucleic Acids Research*, 2003, Vol. 31, No. 13, DOI: 10.1093/nar/gkg599, P. 3429–3431
- [26] Schueren F and Thoms S, (2016), Functional Translational Readthrough: A Systems Biology Perspective, *PLOS Genetics* 12(11):e1006434, Published: August 4, 2016.
- [27] Loughran G, Jungreis I, Tzani I, Power M, Dmitriev R, Ivanov I, Kellis M and Atkins, (2018), Stop codon readthrough generates a C-terminally extended variant of the human vitamin D receptor with reduced calcitriol response, *Journal of Biological chemistry*, Published on January 31, 2018 as Manuscript M117.818526.
- [28] Cridge, A. G., Crowe-Mcauliffe, C., Mathew, S. F. and Tate, W. P. (2018) Eukaryotic translational termination efficiency is influenced by the 3 nucleotides within the ribosomal mRNA channel. *Nucleic Acids Res.*
- [29] Uverskya N Vladimir and Dunkera A Keith, (2010), Understanding Protein Non-Folding, *Biochim Biophys Acta*, 2010 June; 1804(6): 1231–1264. doi:10.1016/j.bbapap.2010.01.017.
- [30] Tompa P and Fersht A (2009), *Structure and Function of Intrinsically Disordered Proteins*, ISBN 9781420078923, November 18, 2009.
- [31] Babu M, Lee R, Groot N and Gsponer J (2011), Intrinsically disordered proteins: regulation and disease, *Current Opinion in Structural Biology* volume 21, Issue 3, June 2011, Pages 432-440.
- [32] Dyson, H. J. and Wright, P. E (2005), Intrinsically unstructured 416 proteins and their functions, *Nature Reviews Molecular Cell Biology* 6, April 2005, P. 197-208.
- [33] Arii J, Shindo K, Koyanagi N, Kato A and Kawaguchi Y, (2016), Multiple Roles of the Cytoplasmic Domain of Herpes Simplex Virus 1 Envelope Glycoprotein D in Infected Cells, *J. Virol*, November 2016 vol. 90 no. 22 10170-10181.
- [34] Giri R, Kumar D, Sharma N and Uversky V, (2016), Intrinsically Disordered Side of the Zika Virus Proteome, *Front. Cell. Infect. Microbiol.*, 04 November 2016.
- [35] Namy, O., Hatin, I. and Rousset, J. P. (2001) Impact of the six nucleotides downstream of the stop codon on translation termination. *EMBO Rep.*, 2, 787–793.
- [36] Drugeon, G., Jean-Jean, O., Frolova, L., Le Goff, X., Philippe, M., Kisselev, L. and Haenni, A. L. (1997) Eukaryotic release factor 1 (eRF1) abolishes readthrough and competes with suppressor tRNAs at all three termination codons in messenger RNA. *Nucleic Acids Res.*, 25, 2254–2258.
- [37] Poole, E. S., Brown, C. M. and Tate, W. P. (1995) The identity of the base following the stop codon determines the efficiency of in vivo translational termination in *Escherichia coli*. *Embo J.*, 14, 151–158.
- [38] Petroczy. A, Palmedo. P, Ingraham. J, A. Hopf. T, Berger. B, Sander. C, S. Marks. D (2017), Structured States of Disordered Proteins from Genomic Sequences, September 22, 2016, *Cell* 167, page 158–170.
- [39] Tompa P, Norman E. Davey, Toby J. Gibson and M. Madan Babu, (2014), A Million Peptide Motifs for the Molecular Biologist, *Molecular Cell* 55, July 17, 2014, Elsevier Inc. 161-169.
- [40] Hofacker I and Stadler P (2008), *RNA Secondary Structures, Bioinformatics- From Genomes to Therapies*, Published Online: 5 FEB 2008.