Review article

Codon Usage Bias Study Reveals the Function of the Genomes in Human Immuno-Deficiency Virus

Sushanta Kumar Barik, [1] Akshata Sewalikar, [1] Rohit Ghuge, [1] Sanjay N Harke [1], Annasaheb S Khemnar[1], Archana Panche [1], Mulay Manjushree Vijay [2], Smita Sitaram Kulkarni,[2] Ashish S. Deshmukh [3], Prapti Deshmukh,[4] Sanjay J Dhoble[5], Keshar Kunja Mohanty[6], Srikanth Prasad Tripathy[7]

1. Department of Biotechnology, Institute of Biosciences and Technology, MGM University, Chhatrapati Sambhaji Nagar, Maharashtra, India-431001.
2. Department of Microbiology, MGM Medical College and Hospital, Chhatrapati Sambhaji Nagar, Maharashtra, India-431001.
3. Department of Respiratory Medicine, MGM Medical College and Hospital, Chhatrapati Sambhaji Nagar, Maharashtra, India-431001
4. Dean, Basic and Applied Sciences, MGM University, India and Principal, MGM Dr G. Y. Pathrikar College of Computer Science and IT, Chhatrapati Sambhajinagar, Maharashtra, India-431001
5. Department of Physics, R.T.M. Nagpur University, Nagpur, Maharashtra, India-440033
6. ICMR-National JALMA Institute for Leprosy and Other Mycobacterial Diseases, Agra, India-282001
7. Dr. D Y Patil Medical College, Hospital and Research Centre, Dr. D Y Patil Vidyapeeth (Deemed to be University), Pune, India-411018

Author for Correspondence: Dr Sushanta Kumar Barik (Assistant Professor, IBT-MGM University, Chhatrapati Sambhaji Nagar, Maharashtra, India), Email ID: sbarik@mgmu.ac.in, sushantakumarbarik82@gmail.com

Abstract

Codon usage effects the efficient expression of Human Immuno-deficient Virus (HIV) genes. Poorly expressed proteins expression could be modified using codon re-engineering in the host. The bias nucleotide composition of the human immunodeficiency virus-1 genes shown the replication efficiency in the human host. The evolution of the nucleotide sequence of HIV-1 contributes the replication capacity, evolvability, and pathogenesis in the human host. Using the synonymous codon pairs, the codons of the HIV-1 genes could be optimized and deoptimized for the expression study of the specific gene in the human host. Changes in the codon pair bias of the gene of HIV-1 effects the virus replication in cell culture. Synonymous nucleotides of the HIV-1 genome impact the life cycle within the human host. The new therapeutic target could possible to develop through recoding the synonymous codons of the HIV-1. The relative synonymous codon usage pattern (RSCU) of the HIV-1 genome reflects the genetic variation among various subtypes and the useful bioinformatics approach with phylogenetic methods predicting the evolutionary patterns of pandemic viruses. Codon usage bias study of the HIV-1 reveals codon frequencies of the isolates from antiretroviral therapy experienced persons as well as treatment naïve individuals.

Key Words: Codon Usage Bias, HIV, AIDS, PR, RT

Introduction

HIV-1 proteins are difficult to produce efficiently due to certain RNA elements that slow their exit from the nucleus and reduce translation. A significant issue is the use of less favorable codons in the viral genome, which impairs protein production. By substituting these less common codons with ones more frequently used in highly expressed human genes, protein expression can be greatly improved. This technique has successfully enhanced the production of HIV-1 proteins as well as other proteins like Thy-1 and GFP. This highlights how optimizing codon usage can help overcome barriers to protein expression in mammalian cells (Haas and Seed., 1996). Interestingly, modifying the codon pairs in HIV-1's genetic code can significantly affect the virus's replication. Rearranging these codon pairs to less favorable configurations greatly reduces the virus's replication ability, while pairing codons in more favorable ways has less impact. This demonstrates that altering genetic encoding can influence viral behavior and presents a potential avenue for developing new treatments (Martrus et al., 2013).

HIV-1 uses a high number of adenine (A) nucleotides in its RNA, but certain regions, known as ‘noA’ sites, avoid having A nucleotides. These sites are crucial for maintaining RNA structure, particularly in areas involved in RNA folding and splicing. Disrupting these noA sites by introducing A nucleotides affects the RNA's structure, indicating their importance in the virus’s replication and function. This balance between accumulating A nucleotides and preserving vital RNA structures highlights the complex evolutionary pressures HIV faces (Berkhout and Hemert.,2022). Synonymous mutations in HIV-1, which do not change the protein sequence, can still impact how the virus behaves. By introducing these mutations into the HIV-1 genome, researchers have gained insights into immune system responses and the effects of codon usage on viral gene expression. Advances in DNA synthesis techniques have made studying virus adaptation and infection easier, revealing new therapeutic targets. However, optimizing codon usage in mammalian cells remains challenging as these mutations can still influence virus function and replication (Jordan-Paiz et al., 2021).

Analysis of HIV-1 sequences shows significant variations in synonymous codon usage across different regions. African isolates display broad distributions, while Asian isolates divide into two distinct groups, with southern Asia showing notable patterns. European isolates have narrower distributions, and American isolates exhibit some variability. Among the non-CRF subtypes, subtype C stands out with the most distinct codon usage patterns. These variations reflect genetic diversity and may provide insights into HIV-1’s evolutionary patterns (Ahn and Son., 2006). Despite using less adapted codons, HIV-1 replicates efficiently in human cells by packaging specific transfer RNAs (tRNAs) that decode codons commonly used by HIV-1 but less so by human genes, particularly A-ending codons. Early HIV-1 genes use codons similar to those of human genes, while late HIV-1 genes use more A-ending codons due to changes in the tRNA pool. This adaptation helps HIV-1 maintain efficient protein translation throughout its lifecycle and could reveal potential drug targets, especially in disrupting the translation of late-stage structural proteins (van Weringh et al., 2011).

Codon usage bias, which affects translation efficiency and immune response, varies among organisms. Studies have found that antiretroviral therapy alters HIV-1 codon frequencies in a subtype-specific manner. Treatment-experienced patients showed increased frequencies of certain codons in the pol and gag genes, suggesting that drug pressure affects codon usage, though not all changes align with resistance mutations. These findings indicate other influencing factors and highlight the need for further research with larger datasets (Palanisamy et al., 2017). Schlafen14 (SLFN14) plays a key role in inhibiting HIV-1 replication by targeting

viral mRNAs with rare codon usage for degradation. This protein interacts with ribosomes to degrade these mRNAs, reducing the virus's ability to replicate. SLFN14's endoribonuclease activity is crucial for this process, and it may work alongside other Schlafen proteins to enhance its antiviral effects. This suggests that SLFN14 could be a promising target for limiting viral replication and provides a potential therapeutic pathway for treating viruses that exploit rare codon translation (Valenzuela et al., 2024).

Effects of codon patterns in genomes of HIV

HIV-1 has a unique way of using codons, the building blocks of its genes, which affects how well it produces its envelope proteins like gp120 and gp41. The virus tends to use certain codons more frequently and avoid others, making it harder to express these proteins. Studies have shown that if we adjust the codons in HIV-1's genes to match those commonly used by human genes, we can dramatically increase the production of these proteins. For example, when researchers optimized the codons for the gp120 protein, they saw its expression go up by several hundred times compared to the natural version. On the other hand, changing the codons in a commonly expressed protein, like Thy-1, to those preferred by HIV-1 led to a sharp drop in its expression (Haas and Seed., 1996). HIV-1's codon patterns provide valuable insights into how the virus evolves. The virus tweaks its genetic code to better fit with the host's preferred codons, which helps it survive and adapt. By looking at HIV-1 genomes from different subtypes over time, we see that the virus's codon usage isn't random but is an evolutionary strategy to enhance its survival in the host (Pandit and Sinha., 2011). Researchers have also used a method called SAVE (Synthetic Attenuated Virus Engineering) to modify HIV-1 by adjusting codon-pair biases in key parts of its genome. This resulted in versions of the virus that replicated more slowly in the lab. However, some of these modified viruses eventually regained their original replication abilities, driven by reverting changes and new mutations. This suggests that while codon-pair changes can weaken the virus, more research is needed to understand about the stable these modifications in the genome (Martrus et al., 2013).

Synonymous or silent mutations in HIV-1, once thought to be harmless, can actually have significant effects. These mutations can alter RNA structures and disrupt important regulatory motifs, showing that some codon regions are crucial beyond just their amino acid sequence. Analysis found 21 such regions, with some linked to known functions and others still unknown. Despite this, experiments didn't show noticeable differences in viral replication, possibly because single-cycle infections may not fully capture the virus's behavior in real infections (Mayrose et al., 2013). Codon usage bias in HIV genomes can result from either mutational bias or selection for efficient translation. Research

suggests that nucleotide composition constraints mainly shape HIV codon usage. While structural genes follow predictions based on GC composition, regulatory genes like tat and rev show more varied codon usage, which might help their expression in early infection stages (Phakaratsakul et al., 2018).

HIV-2, in contrast to HIV-1, has lower viral loads and slower disease progression. Analysis showed that HIV-2 has fewer mutations and a less pronounced G-to-A transition bias compared to HIV-1. This difference might explain HIV-2's slower disease progression, although differences in replication fidelity are not likely a major factor (Rawson et al., 2015). A study on HIV-1 progression revealed that the rate of evolution in the env gene correlates with disease progression. Long-term non-progressors (LTNP) experience more intense functional constraints compared to rapid progressors (RP). Patterns of amino acid usage and glycosylation sites also differ among these categories, showing how host factors influence HIV evolution and disease progression (Roy et al., 2017). HIV-2's tat gene has a higher effective number of codons (ENC) compared to HIV-1's tat gene, suggesting lower early-stage expression and potentially contributing to HIV-2's reduced pathogenicity. The GC3 composition of HIV-2's regulatory genes indicate strong evolutionary adaptation, which might explain its clinical differences from HIV-1 (Vidyavijayan et al., 2017).

Finally, understanding how well viral genes are expressed in different cells involves more than just looking at surface receptors. It also depends on codon adaptation, which affects how efficiently viral mRNAs are translated. A study using the Codon Adaptation Index (CAI) found that while CAI generally correlates with translation efficiency, there are exceptions. HIV-1 and SARS-CoV-2 affect cellular translation differently, with HIV-1 disrupting the CAI-translation efficiency relationship and SARS-CoV-2 showing increased CAI in infected cells. This makes CAI a useful tool for assessing viral susceptibility but suggests it should be used alongside other methods for a complete picture (Zhou et al., 2023).

Importance of codon usage bias study in genome analysis

Codon usage bias (CUB) is a crucial aspect of genome analysis, revealing evolutionary relationships, gene expression, and organism adaptation. For instance, in the study of cotton species, CUB showcases how historical events like polyploidization and hybridization have influenced evolution. Genes with high CUB are often highly expressed, indicating improved protein synthesis efficiency. Understanding environmental pressures on CUB helps researchers grasp adaptive mechanisms in organisms, aiding biotechnology and breeding efforts for improved agricultural practices. This

knowledge underpins future research on genomic diversity, evolution, and adaptation (Wang et.al.,2018).

Studying codon usage bias (CUB) is crucial for genome analysis, especially in understanding the genetic and evolutionary traits of organisms like viruses such as SARS-CoV-2. By examining codon preferences, researchers can deduce evolutionary pressures and relationships between different strains, shedding light on how organisms adapt to their surroundings and host-specific adaptations in viruses, impacting viral pathogenesis and host interactions. CUB research also aids in grasping viral transmission dynamics and can hint at host-switching events, zoonotic transmission potential, and aid in therapeutic and vaccine design. Analyzing CUB helps in creating evolutionary trees, understanding mutational biases, and identifying characteristics influencing pathogenicity and transmissibility. Overall, studying CUB provides insights into genome integrity, evolutionary processes, pathogenic traits, and potential therapeutic targets of viruses and other organisms, facilitating better understanding of outbreaks and future trends (Parvathy et al., 2022). Thus, the research on CUB of Zika virus (ZIKV) is important for explaining evolutionary dynamics and host interactions and mutational pressures on this virus. A detailed examination of CUB brings to the fore how geographic origins affect evolutionary relationships between ZIKV strains as the virus primarily adapts to its mosquito vector Aedes aegypti. This analysis shows that the codon preferences of ZIKV are determined by the availability of tRNA isoacceptors in host cells, which can be used to devise strategies to disrupt viral protein synthesis. The study also points out the significant role of mutational pressure in determining ZIKV codon usage, providing insights into the mechanisms driving viral evolution. This knowledge is important for developing targeted interventions and therapeutic strategies to address ZIKV infections effectively (Rahman et.al.,2017).

In-depth analysis of the genome composition and codon usage patterns of SARS-CoV-2 and other Beta-coronaviruses (βCoVs). It highlights a significant genomic and codon usage bias in SARS-CoV-2, with a preference for A- and U-ending codons, likely driven by mutational pressure and host adaptation. Differences in codon usage patterns compared to human cells suggest challenges in viral translation efficiency. The results reveal a closer evolutionary relationship between SARS-CoV-2 and bat coronaviruses compared to other βCoVs. This analysis underscores the role of codon usage and nucleotide composition in shaping viral evolution and host adaptation mechanisms (Tort et.al.,2020). This research analyzes the codon usage bias (CUB) across four cotton species: *Gossypium arboreum*, *G. raimondii*, *G. hirsutum*, and *G. barbadense*. The study found that these genomes exhibit weak CUB with a preference for pyrimidine-rich codons (T and A). Translational selection was identified as the dominant factor shaping CUB over mutation pressure. The analysis

determined 19-26 optimal codons per species, most ending in A or T. These findings highlight that codon usage patterns are influenced by evolutionary pressures and provide insights into gene expression and functional genomic adaptations in cotton (Wang et.al., 2018).

That codon usage bias plays a crucial role in regulating gene transcription in the filamentous fungus *Neurospora crassa*. Codon optimality was found to correlate with transcription efficiency, independent of translation. Using RNA sequencing, 18 regulatory factors were identified, primarily chromatin remodelers and transcription factors, that influence this process. Among these, SET-2 (a histone methyltransferase) and FKH1 (a fork head transcription factor) were found to play significant roles in codon-dependent transcription regulation. The study highlights the intricate relationship between codon usage, chromatin modifications, and transcription, providing valuable insights into its role as a regulatory mechanism and evolutionary driver (Zhao et.al., 2021).

The phenomenon of codon usage bias (CUB), where specific codons are preferred over others during gene translation. It identifies CUB as a result of a balance between mutation, translational selection, and genetic drift. Factors influencing CUB include GC content, gene expression levels, recombination rates, and environmental stress. Advances in genome sequencing have allowed for genome-wide analyses, revealing how CUB varies among species and contributes to genome evolution. Highlighting its practical applications, such as optimizing recombinant gene expression and developing attenuated viral vaccines, the paper underscores the biological significance and evolutionary implications of CUB, with particular emphasis on insect genomes. This comparative study investigates codon usage bias (CUB) and codon context patterns across 22 insect species, including 15 dipterans (e.g., *Drosophila* and mosquitoes) and 7 hymenopterans (e.g., ants, bees, and wasps). The findings reveal distinct patterns: dipterans predominantly use GC-rich codons, while hymenopterans favor AT-rich ones. Codon usage orders vary between species, correlating with genome structure and evolutionary pressures. The study also highlights differences in amino acid composition and codon context patterns that align with species phylogeny. These results offer valuable insights into translational selection and genome evolution in insects (Behura et.al., 2013).

Conclusion

Codon usage bias study of HIV reveals various factors governing virus evolution. The codon usage bias in between HIV genes is low. The strong correlation between the base and dinucleotide composition of the HIV genome suggested the mutational pressure rather than the natural selection is the most important determinant of the codon usage bias.

Acknowledgement: Thanks to Governor Gretchen Whitmer and Kathy Hochul, USA. Also, thanks to the Obama Foundation, USA (www.obama.org).

References

1. Haas, J., Park, E.C. and Seed, B., 1996. Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Current Biology*, *6*(3), pp.315-324.
2. Martrus, G., Nevot, M., Andres, C., Clotet, B. and Martinez, M.A., 2013. Changes in codon-pair bias of human immunodeficiency virus type 1 have profound effects on virus replication in cell culture. *Retrovirology*, *10*, pp.1-12.
3. Berkhout, B. and van Hemert, F.J., 2022. Silent codon positions in the A-rich HIV RNA genome that do not easily become A: Restrictions imposed by the RNA sequence and structure. *Virus evolution*, *8*(2), p.veac072.
4. Jordan-Paiz, A., Franco, S. and Martínez, M.A., 2021. Impact of synonymous genome recoding on the HIV life cycle. *Frontiers in Microbiology*, *12*, p.606087.
5. Ahn, I. and Son, H.S., 2006. Epidemiological comparisons of codon usage patterns among HIV-1 isolates from Asia, Europe, Africa and the Americas. *Experimental & Molecular Medicine*, *38*(6), pp.643-651.
6. van Weringh, A., Ragonnet-Cronin, M., Pranckeviciene, E., Pavon-Eternod, M., Kleiman, L. and Xia, X., 2011. HIV-1 modulates the tRNA pool to improve translation efficiency. *Molecular biology and evolution*, *28*(6), pp.1827-1834.
7. Palanisamy, N., Osman, N., Ohnona, F., Xu, H. T., Brenner, B., Mesplède, T., & Wainberg, M. A. 2017. Does antiretroviral treatment change HIV-1 codon usage patterns in its genes: A preliminary bioinformatics study. *AIDS Research and Therapy*, *14*(1).
8. Valenzuela, C., Saucedo, S. and Llano, M., 2024. Schlafen14 Impairs HIV-1 Expression in a Codon Usage-Dependent Manner. *Viruses*, *16*(4), p.502.
9. Haas, J., Park, E.C. and Seed, B., 1996. Codon usage limitation in the expression of HIV-1 envelope glycoprotein. *Current Biology*, *6*(3), pp.315-324.
10. Pandit, A. and Sinha, S., 2011. Differential trends in the codon usage patterns in HIV-1 genes. *PLoS One*, *6*(12), p.e28889.
11. Mayrose, I., Stern, A., Burdelova, E.O., Sabo, Y., Laham-Karam, N., Zamostiano, R., Bacharach, E. and Pupko, T., 2013. Synonymous site conservation in the HIV-1 genome. *BMC evolutionary biology*, *13*, pp.1-11.

12. Phakaratsakul, S., Sirihongthong, T., Boonarkart, C., Suptawiwat, O. and Auewarakul, P., 2018. Codon usage of HIV regulatory genes is not determined by nucleotide composition. *Archives of virology*, *163*, pp.337-348.
13. Rawson, J.M., Landman, S.R., Reilly, C.S. and Mansky, L.M., 2015. HIV-1 and HIV-2 exhibit similar mutation frequencies and spectra in the absence of G-to-A hypermutation. *Retrovirology*, *12*, pp.1-17.
14. Roy, A., Banerjee, R. and Basak, S., 2017. HIV progression depends on codon and amino acid usage profile of envelope protein and associated host-genetic influence. *Frontiers in Microbiology*, *8*, p.1083.
15. Vidyavijayan, K.K., Hassan, S., Precilla, L.K., Ashokkumar, M., Chandrasekeran, P., Swaminathan, S. and Hanna, L.E., 2017. Biased nucleotide composition and differential codon usage pattern in HIV-1 and HIV-2. *AIDS Research and Human Retroviruses*, *33*(3), pp.298-307.
16. Zhou, H., Ren, R. and Yau, S.S.T., 2023. Utilizing the codon adaptation index to evaluate the susceptibility to HIV-1 and SARS-CoV-2 related coronaviruses in possible target cells in humans. *Frontiers in cellular and infection microbiology*, *12*, p.1085397.
17. Wang, L., Xing, H., Yuan, Y., Wang, X., Saeed, M., Tao, J., Feng, W., Zhang, G., Song, X. and Sun, X., 2018. Genome-wide analysis of codon usage bias in four sequenced cotton species. *PloS one*, *13*(3), p.e0194372.
18. Parvathy, S.T., Udayasuriyan, V. and Bhadana, V., 2022. Codon usage bias. *Molecular biology reports*, *49*(1), pp.539-565.
19. Ur Rahman, S., Mao, Y. and Tao, S., 2017. Codon usage bias and evolutionary analyses of Zika virus genomes. *Genes & Genomics*, *39*, pp.855-866.
20. Tort, F.L., Castells, M. and Cristina, J., 2020. A comprehensive analysis of genome composition and codon usage patterns of emerging coronaviruses. *Virus Research*, *283*, p.197976.
21. Zhao, F., Zhou, Z., Dang, Y., Na, H., Adam, C., Lipzen, A., Ng, V., Grigoriev, I.V. and Liu, Y., 2021. Genome-wide role of codon usage on transcription and identification of potential regulators. *Proceedings of the National Academy of Sciences*, *118*(6), p.e2022590118.
22. Behura, S.K. and Severson, D.W., 2013. Codon usage bias: causative factors, quantification methods and genome-wide patterns: with emphasis on insect genomes. *Biological Reviews*, *88*(1), pp.49-61.