Journal of Molecular Evolution (2020) 88:473–476
https://doi.org/10.1007/s00239-020-09953-1

LETTER TO THE EDITOR

Protein Evolution in the Flaviviruses

Miguel Arenas[1,2,3]

Received: 7 May 2020 / Accepted: 15 May 2020 / Published online: 25 May 2020

Abstract

Proteins are commonly used as molecular targets against pathogens such as viruses and bacteria. However, pathogens can evolve rapidly permitting their populations to increase in protein diversity over time and thus escape to the activity of a molecular therapy. Subsequently, in order to design more durable and robust therapies as well as to understand viral evolution in a host and subsequent transmission, it is central to understand the evolution of pathogen proteins. This understanding can enable the detection of protein regions that can be potential targets for therapies and predict the emergence of molecular resistance against therapies. In this direction, two articles published recently in the *Journal of Molecular Evolution* investigated the evolution of proteomes of diverse flaviviruses, including Zika virus, Dengue virus and West Nile virus. Here I discuss the importance of considering the evolution of viral proteins, with the use of as realistic as possible models and methods that mimic protein evolution, to improve the design of antiviral therapies.

Keywords Protein evolution · Virus evolution · Flavivirus · Molecular adaptation · Substitution process · Antiviral therapy

Introduction

The rapid evolution of multiple viruses, through mutation and recombination processes followed by selection (Arenas et al. 2018), has been associated with the alteration of transmission vector specificities, increase in virulence and pathogenesis, evasion of host immunity and resistance against drug therapies (including vaccine escape), among others (e.g., Lemey et al. 2006; Woodford and Ellington 2007; Arenas and Posada 2010; Voskarides 2020). At the molecular level, viral proteins can show the acquisition of these capabilities. Note that proteins perform a variety of functions within organisms, including catalysis of chemical reactions, building of cellular structures, and molecular recognition, among others. Because of their central role in these organisms, certain proteins have been selected as molecular targets for multiple treatments against pathogens such as viruses. A relevant case is flavivirus, which includes a variety of highly pathogenic viruses for humans such as dengue virus (DENV), West Nile virus (WNV), yellow fever virus (YFV), Zika virus (ZIKV), Japanese encephalitis virus (JEV) and tick-borne encephalitis virus (TBEV), causing millions of human infections every year (Gubler 2002; Gould and Solomon 2008). Flaviviruses are small enveloped viruses presenting a positive-sense single-stranded RNA genome with 9500–12,500 nucleotides that encodes three structural proteins (nucleocapsid assembly *C*, precursor to membrane protein *prM*, and envelope *E*; all of them mainly related with the capsid) and seven non-structural proteins (binding control protein *NS1*, formation of the viral replication complex *NS2A*, cofactor for protease *NS2B*, helicase *NS3*, formation of the viral replication complex *NS4A*, formation of the viral replication complex together with the membrane protein *NS4B* and, RNA polymerase *NS5*; all of them mainly related with virus replication). Vaccines have been successfully developed for some flaviviruses, but not for others despite much effort (Ishikawa et al. 2014). Indeed, new drugs are being developed to effectively inhibit different flavivirus proteins, especially the envelope, protease, helicase and polymerase (e.g., Luo et al. 2015; de Wispelaere et al. 2018). Still resistance against therapies can be

Handling editor: **David Liberles**.

✉ Miguel Arenas
marenas@uvigo.es

[1] Department of Biochemistry, Genetics and Immunology, University of Vigo, 36310 Vigo, Spain

[2] Biomedical Research Center (CINBIO), University of Vigo, 36310 Vigo, Spain

[3] Galicia Sur Health Research Institute (IIS Galicia Sur), 36310 Vigo, Spain

observed as a consequence of certain evolutionary events occurring in proteins of these viruses (e.g., Wang et al. 2017). Therefore, investigating the evolution of these viral proteins is crucial to design durable and effective antiviral treatments. In this concern, two articles recently published in the *Journal of Molecular Evolution* interestingly investigate evolutionary processes of flavivirus proteins. In particular, Le and Vinh (2020) estimated relative substitution rates among amino acids using the proteome of WNV, DENV and ZIKV, which could be useful to perform more realistic phylogenetic inference of flavivirus proteins. The other study by Nunez-Castilla et al. (2020) identified regions of the proteomes of ZIKV, DENV and other flaviviruses presenting evolutionary constraints that could be used as candidates for broadly neutralizing antiviral drugs targets across flaviviruses. The importance, goals and limits of both studies are discussed in the following section.

Evolutionary Dynamics of Flavivirus Proteins

Protein variants emerge from evolutionary mechanisms (i.e., mutation and recombination) upon which selection operates. The action of selection can be observed in the relative substitution rates among amino acids (i.e., some amino acids could be favored over others to maintain the protein folding and activity) and in the level of conservation in certain regions (i.e., evolutionary constrains also to maintain the protein activity). These aspects of viral protein evolution are discussed below with focus on the particular case of flavivirus, although the discussion could also be extended to other viruses.

The Need for an Empirical Substitution Model for Flavivirus

Traditional phylogenetic inferences based on probabilistic approaches (which currently are the most accurate approaches in phylogenetics) apply a substitution model of molecular evolution (Arenas 2015). At the protein level, a substitution model consists of a 20×20 matrix of relative exchangeability rates among amino acids (hereafter, exchangeability matrix Q) and 20 amino acid frequencies at the equilibrium. These parameters are usually estimated from large empirical protein data [i.e., mitochondrial proteins of Arthropoda (Abascal et al. 2007)]. The success of empirical substitution models of protein evolution in phylogenetic inferences is caused by the availability of the models (the exchangeability matrix and the amino acid frequencies of many empirical models have been already developed and implemented in user-friendly computational frameworks that are ready for use without further requirements) and technical simplicity (they assume site-independent evolution that allows straightforward incorporation into likelihood functions). Currently, around 50 empirical substitution models of protein evolution are available. Most of them are based on general nuclear or mitochondrial protein data and some of them were developed from proteins of certain organisms, including viruses like human immunodeficiency virus (HIV) (Nickle et al. 2007) or influenza virus (Dang et al. 2010). Despite the selection and use of a best-fitting substitution model of protein evolution being traditionally considered in phylogenetics as a mandatory procedure to obtain accurate inferences [i.e., topology and branch lengths in the inferred phylogenetic tree (Lemmon and Moriarty 2004)], recently there is some controversy in the field with authors against (Spielman and Kosakovsky Pond 2018; Abadi et al. 2019) and for (Kaehler et al. 2017; Gerth 2019) the need for substitution model selection. This relates to the more general question of if the best fit model is phenomenological or if it relates more directly to the lineage- and gene-specific processes of evolution (Liberles et al. 2013). In my opinion much work is required to assess this issue (i.e., evaluating data with variable molecular diversity and exploring other independent metrics different from the traditionally used sequence similarity). In any case, the need for new empirical substitution models of protein evolution seems to be real due to the lack of representation of many data in the currently available empirical substitution models. For example, using software for substitution model selection Keane et al. (2006) found that the best-fitting empirical substitution model for large proteobacteria and archaea protein datasets was a model inferred from retroviral *Pol* proteins that is not expected to properly describe their evolutionary processes.

In a recent issue of *Journal of Molecular Evolution*, Le and Vinh (2020) present a new empirical substitution model of protein evolution based on proteomes of the flavivirus WNV, DENV and ZIKV. Viruses, including flaviviruses, usually present high evolutionary rates and can be subjected to evolutionary constraints (i.e., caused by transmission and co-evolution with the host) different from those occurring in other organisms. Hence, an empirical substitution model based on the proteome of a specific family of viruses can better mimic the evolution of proteins belonging to that family than other empirical substitution models (e.g., see for proteins of the influenza virus Dang et al. 2010). This outcome was also found by Le and Vinh (2020) for the new empirical substitution model of flavivirus proteins where the model better fit (in terms of maximum likelihood) test flavivirus protein datasets than other empirical substitution models, including substitution models based on proteins from other viruses such as HIV and influenza virus. Consequently, this new empirical substitution model of protein evolution can be useful to obtain accurate phylogenetic inference from protein sequences of flaviviruses. Still future work to properly model protein evolution is needed. In general,

empirical models can be improved with the incorporation of protein sequences from new studies, a clear description of underlying modeling error, and implementation in software for substitution model selection and phylogenetic inference. Being more ambitious, new substitution models of protein evolution could increase in realism by avoiding technical assumptions, like substitution reversibility, that are made for mathematical simplicity. Indeed, analyses based on empirical substitution models for protein evolution ignore that different protein sites can evolve under different evolutionary patterns with different effects on the protein stability and activity (Echave et al. 2016; Jiménez-Santos et al. 2018), suggesting the use of more complex substitution models of protein evolution (e.g., Wilke 2012; Bordner and Mittelmann 2013; Echave and Wilke 2017; Bastolla and Arenas 2019; Arenas and Bastolla 2020).

The Need for Identifying Regions with Evolutionary Constraints Along the Proteomes of Flaviviruses

Genomic regions are often subjected to different strengths of selection on molecular stability and activity indicating that evolutionary patterns usually vary across genome sequences (Arbiza et al. 2011; Jiménez-Santos et al. 2018; Del Amparo et al. 2020). Therefore, in order to design an antiviral therapy one should investigate, among other aspects, which regions of the proteome could act as potential therapy targets (i.e., evaluating their biological function) and also under which patterns such regions evolve (i.e., evaluating their capacity to acquire diversity and potential drug resistance during a given period of time). Viral proteomes include active proteins with and without unique 3D structures (i.e., note that flaviviruses include both type of proteins, as discussed above). Interestingly, a large fraction of proteomes consists of intrinsically disordered proteins that are biologically active (Xue et al. 2012). Moreover, it is known that the conformational flexibility of structurally disordered protein regions allow them to acquire a new function while maintaining the original one (i.e., affecting antibody binding in envelope proteins) and, actually, disordered proteins of flaviviruses have shown rapid evolutionary dynamics of structural disorder favoring functional change (Ortiz et al. 2013).

In a recent issue of *Journal of Molecular Evolution*, Nunez-Castilla et al. (2020) identified highly conserved regions (in both sequence and structure) in the proteomes of ZIKV, DENV and other flaviviruses that could be used as candidates for broadly neutralizing antiviral drugs targets against flaviviruses. The study also investigates regions related to viral transmission (vector specificity) by analyzing their evolutionary rates. An important consideration made in this study is that, following from previous work (Chong et al. 2018), intrinsically disordered proteins can present structural features conserved across the conformational ensemble that could be used as potential drug targets. Clearly, highly conserved regions (in protein sequence and structure) are often catalytic residues (Ribeiro et al. 2020) that should be considered in the design of antiviral drugs, but also other more variable regions (i.e., stabilizing residues near a binding pocket) can present important roles in protein activity and should not be ignored when designing antiviral drugs (i.e., note that the drug must properly fit within the protein binding pockets). Interestingly, the study by Nunez-Castilla et al. (2020) discusses the possible application of some inhibitors, which have been already used against other viruses like Hepatitis C virus (HCV), to flaviviruses. This opens the door towards future research involving molecular docking between those possible inhibitors and the identified regions in the proteomes of flaviviruses with potential as drug targets. Clearly, this study presents progress in the field and suggests directions for future research to improve our current set of therapies against flaviviruses.

Concluding Thoughts

The evolutionary patterns observed in the proteomes of flaviviruses can be summarized by an empirical substitution model of evolution that can be useful to obtain more accurate phylogenetic inference than those performed under other empirical substitution models that are currently available. This new empirical substitution model of protein evolution should help to clarify our knowledge about the origin and evolution of flaviviruses. However, one should be aware about the assumptions made by the empirical substitution models of protein evolution and, in this concern, the establishment and use of more realistic models, like those that directly consider stability and functional constraints, should be encouraged.

The heterogeneous strength of selection throughout the proteomes of flaviviruses (i.e., caused by constraints from molecular stability and function) results in proteome regions presenting variable levels of molecular diversity. In this direction, some regions present large degrees of conservation in sequence and structure, suggesting that they play an important role in the activity of the virus and alterations can be fatal for its life cycle. Consequently, these regions can be used as molecular targets of antiviral therapies. However, other less conserved regions could also be used for this purpose despite their identification being more complex (i.e., they could not be recognized by analyzing only genetic diversity). In any case, it is clear that identifying flaviviruses proteome regions that can potentially be used as molecular targets of antiviral therapies is important progress in the field. A crucial subsequent step will be a computational and experimental evaluation of those potential candidate regions with current and new molecular therapies.

Acknowledgments I thank the Grants "RYC-2015-18241" (MA) from the Spanish Ministry of Economy and Competitivity and "ED431F 2018/08" (MA) from the Xunta de Galicia that support my research.

Compliance with Ethical Standards

Conflicts of interest The author declares no conflict of interest.

References

Abadi S, Azouri D, Pupko T, Mayrose I (2019) Model selection may not be a mandatory step for phylogeny reconstruction. Nat Commun 10:934

Abascal F, Posada D, Zardoya R (2007) MtArt: a new model of amino acid replacement for Arthropoda. Mol Biol Evol 24:1–5

Arbiza L, Patricio M, Dopazo H, Posada D (2011) Genome-wide heterogeneity of nucleotide substitution model fit. Genome Biol Evol 3:896–908

Arenas M (2015) Trends in substitution models of molecular evolution. Front Genet 6:319

Arenas M, Bastolla U (2020) ProtASR2: ancestral reconstruction of protein sequences accounting for folding stability. Methods Ecol Evol 11:248–257

Arenas M, Posada D (2010) Computational design of centralized HIV-1 genes. Curr HIV Res 8:613–621

Arenas M, Araujo NM, Branco C, Castelhano N, Castro-Nallar E, Perez-Losada M (2018) Mutation and recombination in pathogen evolution: relevance, methods and controversies. Infect Genet Evol 63:295–306

Bastolla U, Arenas M (2019) The influence of protein stability on sequence evolution: applications to phylogenetic inference. In: Sikosek T (ed) Computational methods in protein evolution. Springer, New York, pp 215–231

Bordner AJ, Mittelmann HD (2013) A new formulation of protein evolutionary models that account for structural constraints. Mol Biol Evol 31:736–749

Chong B, Li M, Li T, Yu M, Zhang Y, Liu Z (2018) Conservation of potentially druggable cavities in intrinsically disordered proteins. ACS Omega 3:15643–15652

Dang CC, Le QS, Gascuel O, Le VS (2010) FLU, an amino acid substitution model for influenza proteins. BMC Evol Biol 10:99

de Wispelaere M, Lian W, Potisopon S, Li PC, Jang J, Ficarro SB, Clark MJ, Zhu X, Kaplan JB, Pitts JD et al (2018) Inhibition of flaviviruses by targeting a conserved pocket on the viral envelope protein. Cell Chem Biol 25(1006–1016):e1008

Del Amparo R, Vicens A, Arenas M (2020) The influence of heterogeneous codon frequencies along sequences on the estimation of molecular adaptation. Bioinformatics 36:430–436

Echave J, Wilke CO (2017) Biophysical models of protein evolution: understanding the patterns of evolutionary sequence divergence. Annu Rev Biophys 46:85–103

Echave J, Spielman SJ, Wilke CO (2016) Causes of evolutionary rate variation among protein sites. Nat Rev Genet 17:109–121

Gerth M (2019) Neglecting model selection alters phylogenetic inference. bioRxiv. https://doi.org/10.1101/849018v1.abstract

Gould EA, Solomon T (2008) Pathogenic flaviviruses. Lancet 371:500–509

Gubler DJ (2002) The global emergence/resurgence of arboviral diseases as public health problems. Arch Med Res 33:330–342

Ishikawa T, Yamanaka A, Konishi E (2014) A review of successful flavivirus vaccines and the problems with those flaviviruses for which vaccines are not yet available. Vaccine 32:1326–1337

Jiménez-Santos MJ, Arenas M, Bastolla U (2018) Influence of mutation bias and hydrophobicity on the substitution rates and sequence entropies of protein evolution. PeerJ 6:e5549

Kaehler BD, Yap VB, Huttley GA (2017) Standard codon substitution models overestimate purifying selection for nonstationary data. Genome Biol Evol 9:134–149

Keane TM, Creevey CJ, Pentony MM, Naughton TJ, McLnerney JO (2006) Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol Biol 6:29

Le TK, Vinh LS (2020) FLAVI: an amino acid substitution model for flaviviruses. J Mol Evol. In press

Lemey P, Rambaut A, Pybus OG (2006) HIV evolutionary dynamics within and among hosts. AIDS Rev 8:125–140

Lemmon AR, Moriarty EC (2004) The importance of proper model assumption in bayesian phylogenetics. Syst Biol 53:265–277

Liberles DA, Teufel AI, Liu L, Stadler T (2013) On the need for mechanistic models in computational genomics and metagenomics. Genome Biol Evol 5:2008–2018

Luo D, Vasudevan SG, Lescar J (2015) The flavivirus NS2B-NS3 protease-helicase as a target for antiviral drug development. Antiviral Res 118:148–158

Nickle DC, Heath L, Jensen MA, Gilbert PB, Mullins JI, Kosakovsky Pond SL (2007) HIV-specific probabilistic models of protein evolution. PLoS ONE 2:e503

Nunez-Castilla J, Rahaman J, Ahrens JB, Balbin CA, Siltberg-Liberles J (2020) Exploring evolutionary constraints in the proteomes of Zika, Dengue and other flaviviruses to find fitness-critical sites. J Mol Evol 88:399–414

Ortiz JF, MacDonald ML, Masterson P, Uversky VN, Siltberg-Liberles J (2013) Rapid evolutionary dynamics of structural disorder as a potential driving force for biological divergence in flaviviruses. Genome Biol Evol 5:504–513

Ribeiro AJM, Tyzack JD, Borkakoti N, Holliday GL, Thornton JM (2020) A global analysis of function and conservation of catalytic residues in enzymes. J Biol Chem 295:314–324

Spielman SJ, Kosakovsky Pond SL (2018) Relative evolutionary rates in proteins are largely insensitive to the substitution model. Mol Biol Evol 35:2307–2317

Voskarides K (2020) Animal-to-human viral transitions: is SARS-CoV-2 an evolutionary successful one? J Mol Evol. In press

Wang S, Liu Y, Guo J, Wang P, Zhang L, Xiao G, Wang W (2017) Screening of FDA-approved drugs for inhibitors of Japanese encephalitis virus infection. J Virol 91:e01055-17

Wilke CO (2012) Bringing molecules back into molecular evolution. PLoS Comput Biol 8:e1002572

Woodford N, Ellington MJ (2007) The emergence of antibiotic resistance by mutation. Clin Microbiol Infect 13:5–18

Xue B, Dunker AK, Uversky VN (2012) Orderly order in protein intrinsic disorder distribution: disorder in 3500 proteomes from viruses and the three domains of life. J Biomol Struct Dyn 30:137–149