Available online at www.sciencedirect.com

SCIENCE DIRECT

Gene 347 (2005) 237–246

GENE
SECTION
EVOLUTIONARY GENOMICS

www.elsevier.com/locate/gene

Protein evolution in viral quasispecies under selective pressure: A thermodynamic and phylogenetic analysis

Carlos Briones*, Ugo Bastolla

Molecular Evolution Laboratory, Centro de Astrobiología (CSIC-INTA), Carretera de Ajalvir, Km. 4, 28850 Torrejón de Ardoz, Madrid, Spain

Received 1 October 2004; accepted 10 December 2004

Abstract

The evolution of RNA viruses under antiviral pressure is characterized by high mutation rates and strong selective forces that induce extremely rapid changes of protein sequences. This makes the course of molecular evolution directly observable on time scales of months. Here we study the interplay between selection for drug resistance and selection for thermodynamic stability in the protease (PR) and the reverse transcriptase (RT) of human immunodeficiency virus type 1 (HIV-1) clones extracted from two patients with complex treatment histories. This analysis shows that folding thermodynamic properties may fluctuate very strongly in the course of quasispecies evolution under selective pressure. For the first case, our data suggest that folding efficiency of the RT is sacrificed at the advantage of drug resistance, while the corresponding PR seems to undergo selection for thermodynamic stability in the absence of substitutions associated to resistance. The PR of the second case is not submitted to antiviral pressure during the period analyzed and seems to initiate random fluctuations that lead to the accidental increase of its folding efficiency. In summary, joint consideration of sequence evolution and thermodynamic parameters can represent a more comprehensive approach for the study of the evolution of RNA viruses.

Keywords: RNA virus; HIV; Protein folding; Protease; Reverse transcriptase; Minority memory genomes

1. Introduction

Molecular evolution studies usually consider families of homologous macromolecules, either proteins or nucleic acids, diverged after millions of years of evolution. Statistical methods are then required in order to infer the evolutionary history of the macromolecules and the relative role of mutation and natural selection from the observed sequence patterns (Moritz and Hillis, 1990). Remarkably, there are certain situations in which molecular evolution happens on a time scale amenable to direct observation. This is the case of populations of viruses with RNA genome submitted to antiviral treatment. In them, not only the high mutation rate characteristic of RNA viruses, but also the strong selective pressure exerted by the antiviral drugs, make the evolutionary process very rapid, so that significant changes happen on the time scale of weeks to months.

Viruses with RNA genome are the most abundant group of human, animal and plant pathogens. They replicate with very high mutation rates, due to the absent or low proofreading activity of RNA-dependant RNA polymerases or DNA polymerases, the latter also called reverse transcriptases (RTs; Drake, 1993; Drake and Holland, 1999). As a consequence, RNA virus populations are highly heterogeneous and contain particles with closely related but non-identical genomes. Such a complex population structure is

Abbreviations: HIV, human immunodeficiency virus; HIV-1, human immunodeficiency virus type 1; FMDV, foot-and-mouth disease virus; RT, reverse transcriptase; PR, protease; CD4+, CD4+ T lymphocytes; CD4+ count, number of CD4+ T lymphocytes per L of plasma; HAART, highly active antiretroviral therapy; NRTI, nucleoside analog RT-inhibitor; PI, protease inhibitor; UPGMA, unweighted pair group method with arithmetic mean; NJ, neighbor joining; MP, maximum parsimony; PDB, protein data bank.

* Corresponding author. Tel.: +34 915206411; fax: +34 915201074.
E-mail address: brioneslc@inta.es (C. Briones).

0378-1119/$ - see front matter © 2004 Elsevier B.V. All rights reserved.
doi:10.1016/j.gene.2004.12.018

termed viral quasispecies (Eigen, 1971, 1996; Eigen and Schuster, 1979; Domingo et al., 2001). Quasispecies dynamics is characterized by a continuous process of mutant generation, competition and selection, that results in the dominance of one or several most-fit genomes surrounded by a mutant spectrum (Eigen and Biebricher, 1988). The behavior of viral quasispecies and their response to selective pressures is influenced by the ensemble of mutants that compose the evolving population. Experiments with foot-and-mouth disease virus (FMDV) in cell culture (Ruíz-Jarabo et al., 2000, 2002), and with HIV-1 in vivo (Briones et al., 2003) documented that viral quasispecies may possess a molecular memory of their past evolutionary history in the form of minority components (ranging from 0.1% to 10% of the total number of genomes) within the mutant spectra. Quasispecies memory reflects the genomes that were dominant at an earlier phase of the past evolutionary history, and such minority genomes can drive the evolution of the viral populations during chronic infections (Briones et al., submitted for publication).

Quasispecies structure and high population sizes allow RNA viruses to quickly react to selective pressures exerted either by the immune system of the host, or by the administration of antiviral drugs or monoclonal antibodies (Domingo and Holland, 1997; Domingo et al., 2001). This is the case of retroviruses, such as HIV-1, a family of RNA viruses that include two phases in their replication cycle. The first step involves the infection of the target cell (mainly CD4+ lymphocytes and macrophages in the case of HIV-1) by the virion, whose genomic RNA is copied into complementary DNA by the retrovirus-encoded RT (with an average mutation rate of 10^{-4} substitutions per nucleotide copied) and integrated in the host genome. As integrated proviruses, DNA copies of HIV-1 RNA replicate as cellular genes, with the high fidelity copying inherent to cellular DNA polymerases (typically, 10^{-9} substitutions per nucleotide copied). Upon expression from integrated sites in their host cells, viral RNAs are traduced to viral polyproteins that are in turn processed by means of a virus-encoded protease (PR) and certain cellular proteases. The process culminates in the formation and maturation of new retroviral particles capable of infecting new host cells (Meyerhans et al., 1989; Goff, 2001). Therefore RT and PR, coded by the HIV-1 *pol* gene, are the key proteins in the retroviral life cycle and have become the targets for antiretroviral therapy during the last two decades (Yeni et al., 2004).

HIV-1 RT is a heterodimeric enzyme that consists of two subunits of 66 and 51 kDa. The catalytic subunit (p66) is 560 amino acids long and the smaller subunit (p51) contains the first 440 amino acids of p66. The p66 subunit folds into five subdomains, known as “fingers”, “palm”, “thumb”, “connection” and “Rnase H” (Kohlstaedt et al., 1992; Jacobo-Molina et al., 1993). The first two RT subdomains span amino acids 1 to 234 and include the catalytic triad composed by Asp at positions 110, 185 and 186 (Jonckheere et al., 2000). HIV-1 PR is a homodimer with exact twofold rotational C_2 symmetry, composed of two identical 99-amino acid monomers (Navia et al., 1989). Each chain contains the characteristic Asp-Thr-Gly active site located in loops that approach the center of the dimer (Pearl and Taylor, 1987; Oroszlan and Luftig, 1990). Since the first description of a variant of the RT resistant to the drug zidovudine (Larder et al., 1989), a growing number of resistance mechanisms to RT and PR inhibitors have been reported. These involve amino acid replacements that confer either resistance to individual drugs or multidrug resistance, as well as insertions and deletions within the *pol* gene (Jonson et al., 2003; Yeni et al., 2004).

Here we study the evolution of the retroviral RT and PR, through a detailed thermodynamic and phylogenetic analysis of a large number of clones of these proteins. They were sequenced in two different HIV-1 evolving quasispecies that maintained minority memory genomes at certain phases of their complex histories of selective pressures. This analysis allows us to investigate the interplay between selection for drug resistance and selection for thermodynamic stability, and the role of population structure in protein evolution.

2. Materials and methods

2.1. Virus samples and sequences

Sequential viral samples were obtained from two patients infected by a B subtype HIV-1, who underwent highly active antiretroviral therapy (HAART) at Hospital Carlos III in Madrid, Spain. Treatment histories and evolution of their plasma viral load and CD4+ lymphocyte count are summarized in Fig. 1. Patient 1 (a man born in 1956, infected by homosexual intercourse some time before 1997) showed large fluctuations in viral load and CD4+ count associated with two periods of treatment interruption (TI; further details can be found in Briones et al., submitted for publication). Patient 2 (a haemophiliac man born in 1970, infected in 1982 by contaminated blood products) experimented a progressive CD4+ decrease during his treatment

Fig. 1. Evolution of treatment history, viral load (HIV-1 RNA copies/ml), CD4+ count (CD4+ lymphocytes/μl), and normalized energy gap for PR and RT from patients 1 (a) and 2 (b). Boxes delimit the periods of treatment with the different inhibitors, using the following abbreviations: NRTI, nucleoside reverse transcriptase inhibitors (3TC, lamivudine; ABC, abacavir; d4T, stavudine; ddC, zalcitabine; ddI, didanosine; ZDV, zidovudine); PI, protease inhibitors (IDV, indinavir; LPV, lopinavir; RTV, ritonavir). Sampling dates are written as day.month.year. Sample identification numbers of the seven (a) and six (b) samples subjected to sequence analysis (see details in the text) have been encircled. In (a) two treatment interruption periods of patient 1 are shadowed, and they are identified as TI_1 and TI_2, respectively. The normalized energy gap shown represents the average for all the clones sequenced in a given sample, with the exception of the RT extracted from patient 2, in which case it represents the gap for the consensus sequence. Panel (b) is adapted from Briones et al. (2000).