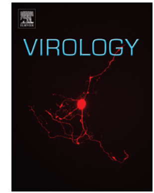
Virology 454-455 (2014) 34–39

Contents lists available at ScienceDirect

Virology

journal homepage: www.elsevier.com/locate/yviro



Constraints from protein structure and intra-molecular coevolution influence the fitness of HIV-1 recombinants

Jeongmin Woo, David L. Robertson *, Simon C. Lovell **

Computational and Evolutionary Biology, Faculty of Life Sciences, University of Manchester, Oxford Road, Manchester M13 9PT, UK

ARTICLE INFO

Article history:
Received 17 December 2013
Returned to author for revisions
10 January 2014
Accepted 29 January 2014
Available online 22 February 2014

Keywords:
HIV-1
Evolution
Protein structure
Coevolution

ABSTRACT

A major challenge for developing effective treatments for HIV-1 is the viruses' ability to generate new variants. Inter-strain recombination is a major contributor to this high evolutionary rate, since at least 20% of viruses are observed to be recombinant. However, the patterns of recombination vary across the viral genome. A number of factors influence recombination, including sequence identity and secondary RNA structure. In addition the recombinant genome must code for a functional virus, and expressed proteins must fold to stable and functional structures. Any intragenic recombination that disrupts internal residue contacts may therefore produce an unfolded protein. Here we find that contact maps based on protein structures predict recombination breakpoints observed in the HIV-1 pandemic. Moreover, many pairs of contacting residues that are unlikely to be disrupted by recombination are coevolving. We conclude that purifying selection arising from protein structure and intramolecular coevolutionary changes shapes the observed patterns of recombination in HIV-1.

© 2014 Elsevier Inc. All rights reserved.

Introduction

HIV-1, the causative agent of AIDS, is well known for its high genetic variability both within infected individuals and across the pandemic as a whole (Rambaut et al., 2004). The high rate of evolution in HIV-1 is due to the error-prone nature of its reverse transcriptase (Jetzt et al., 2000), high rate of viral turnover (Ho et al., 1995) and diversifying selection caused by active immune response (Choisy et al., 2004; Wolinsky et al., 1996; Yang et al., 2003). In addition, recombination occurs readily within a viral population and if an infected individual experiences multiple infections with viruses from, for example, different subtypes, recombination can produce an inter-subtype recombinant (Robertson et al., 1995; Fang et al., 2004). In this way, genetic recombination reassorts existing mutations, either bringing together mutations or separating them, and so contributes significantly to HIV-1 evolution (Negroni and Buc, 2001).

Cocirculation of different subtypes is increasingly found in large areas of the world, and so inter-subtype recombinants are frequently identified (Robertson et al., 1995; Takeb et al., 2004). If the same inter-subtype recombinant is transmitted to multiple individuals it is termed a circulating recombinant form (CRF) (Robertson et al., 2000). To date 61 CRF lineages have been identified and some (CRF01 and CRF02 in particular) account for a large proportion of HIV infections worldwide (Ramirez et al., 2008), whereas those recombinants that have only been identified once are termed unique recombinant forms (URFs). It is estimated that in total intersubtype recombinants comprise up to 20% of the global HIV infections (Hemelaar et al., 2006). Recombination therefore provides a substantial contribution to genetic variation in HIV-1 populations.

Selective advantage from recombination is inferred from the successful establishment of strains of the same recombinant origin. Such selective advantage may arise because recombination has the potential to generate a novel and highly divergent variant in a short time, as compared to variants generated solely by single point mutations. This increase in genetic variation can be beneficial in the presence of selection pressure arising from the immune response and drug treatment (Rouzine and Coffin, 2005). Moreover, since the recombination process incorporates mutations that can produce a functional protein, it is likely to produce a viable virus (Drummond et al., 2005; Simon-Loriere et al., 2009). For example, multidrug-resistant HIV-1 has been generated from reassortment between different drug-resistance mutations (Geretti, 2006).

Interestingly, restricted rates of inter-subtype recombination have been observed in HIV's *pol* (Galli et al., 2010) and *env* (Archer

* Corresponding author. Tel.: +44 161 2755089; fax: +44 161 275 5082.
** Corresponding author. Tel.: +44 161 2755748; fax: +44 0161 275 5082.
E-mail addresses: david.robertson@manchester.ac.uk (D.L. Robertson), simon.lovell@manchester.ac.uk (S.C. Lovell).

http://dx.doi.org/10.1016/j.virol.2014.01.029
0042-6822 © 2014 Elsevier Inc. All rights reserved.

et al., 2008) gene regions. Strand switching during reverse transcription is the mechanism that generates HIV recombinants, and there are several sequence features that can affect the probability of this process occurring. These factors include sequence identity between the co-packaged ‘parental’ viral RNAs (Archer et al., 2008; Zhang and Temin, 1994; Baird et al., 2006), the presence of homopolymeric runs (Klarmann et al., 1993) and secondary RNA structure (Moumen et al., 2003; Galetto et al., 2004; Simon-Loriere et al., 2010). In particular, the low frequency of recombination in hypervariable regions of envelope protein gp120 can be explained by the lower levels of sequence identity between strains in this region of the viral genome (Baird et al., 2006).

The requirement to form folded, functional proteins may also influence patterns of recombination (Voigt et al., 2002) due to the action of selection once the recombinant has been generated. As a consequence recombination locations or ‘breakpoint’ positions that produce functional chimeric proteins should be observed at higher frequencies than those that do not (Voigt et al., 2002). Several studies have been performed to explain the location of breakpoint “hot” and “cold” regions in other systems, including β-lactamase (Voigt et al., 2002; Meyer et al., 2003), ssDNA viruses including circovirus, microvirus, parovovirus, mastrevirus and begomoviruses (Lefeuvre et al., 2007, 2009) and in gp120 of HIV-1 (Simon-Loriere et al., 2009). These analyses use the SCHEMA algorithm (Voigt et al., 2002), which predicts the degree of fold disruption in chimeric proteins. In addition, the degree to which recombination disrupts coevolved intra-protein amino acid interaction could be an important factor in determining functionality of a recombinant protein, as compensatory interactions occurring between co-adapted sequence substitutions might also be required for maintaining fitness (Travers et al., 2007).

Here, we hypothesize that protein structure is an additional and important factor in shaping the observed recombination patterns in the HIV-1 pandemic. We find that residues that are in contact and display correlated evolution are unlikely to be disrupted by recombination because when they are disrupted the recombinant will probably be non-functional. This will be because sets of coevolved residues will be coadapted to each other, and so are likely to evolve as a single structural or functional unit. We have analyzed the relationship between breakpoint positions and both amino-acid contacts formed by the protein structure, and the positions of coevolving residues of CRFs and URFs in the gp120 protein. We have also compared the pattern of constraints in gp120 with those of other HIV proteins. Our results suggest that the maintenance of protein structure and coevolving amino acid units is a significant constraint shaping observed patterns of recombination in HIV-1.

Results

Constraint on viral recombination from protein structure

In order to function viral proteins must form folded three-dimensional structures. To determine whether this requirement constrains the success of recombination events, *i.e.*, the fitness of a recombinant, we assume that all residue contacts in the native structure are favourable and that disrupting these contacts by recombination may be deleterious. If the observed pattern of recombination follows that predicted by this model we conclude that these simplifying assumptions may be correct.

To do this, we identified all residue contacts, and decomposed them into pairs of contacting residues. We assess all positions in the protein sequence as potential recombination breakpoints, and determined how many pairs of contacting residues would be disrupted if there was a recombination at each position. As a null distribution we used the same number of randomly selected pairs of residues, and determined whether potential breakpoints fell between them (Fig. 1).

There are 9 recombination events observed in gp120 arising from 6 CRFs; 4 recombination breakpoints occurred in regions where N_c/N_b values were significantly lower than the random distribution; these are in the C2 region (residues 197 to 295). For URF breakpoints, 8 out of 16 breakpoints in 12 URFs were located within these areas. Conversely, there are few breakpoints (2 CRF, 1 URF), in areas where large numbers of interactions are disrupted by the potential breakpoint. Several breakpoints (3 from CRFs and 3 from URFs) are clustered in the V1/V2 loop regions. For these we are unable to calculate an exact value for the number of contacts disrupted as this region of the protein (residues 129 to 195) are omitted from the crystal structure. Nevertheless, evidence from a different crystal form (Huang et al., 2005) indicates that these regions form loops with low contact order.

Interactions of amino acid substitutions in protein structure are not disrupted by recombination breakpoints in CRFs

It is likely that specific pairs of residue contacts vary in their importance with regard to inhibiting the success of a recombinant. We hypothesise that coevolved residue pairs that are in contact in the three-dimensional structure will be coadapted, and so it will be advantageous if they are present as a pair. Should recombination occur in the linear sequence between a coevolving pair, this advantageous interaction will be disrupted, potentially leading to a decrease in fitness. Our hypothesis predicts, therefore, that the success of recombination breakpoints will be systematically suppressed in the linear sequence between a pair of coevolving residues that are close in the three-dimensional structure.

Coevolution was defined in two ways. In the first definition, termed the “covarying contact” model we determined where changes from the consensus sequence mapped to the three-dimensional protein structure. The mismatches between two consensus sequences of parental subtypes were regarded as substitutions. If the two mismatch sites are in van der Waals contact with each other on the three dimensional structure, as defined by Probe (Word et al., 1999) we hypothesize that the pairs of mismatch sites are co-varying. In the second model we used mutual information (MI); amino acid pairs were counted as coevolving if MI values were >0.25 and were in contact in the three-dimensional protein structure. Both methods of indicating coevolution are simple to compute; although false positives are possible (Marks et al., 2011), their frequency is limited by the condition that residues are in contact in the crystal structures.

For the covarying contact model we mapped sequence substitution sites and breakpoints of 6 known CRFs in the gp120 structure (Kwong et al., 1998). We find that the co-varying substitution pairs are mostly located within regions of the protein that originated from a single subtype (Fig. 2). Similar results are seen for the p66 domain of pol, with few covarying pairs interacting across recombination boundaries (5 out of 28 CRFs, Supplementary Fig. 1).

In all the cases of CRFs with recombination breakpoints within gp120, the number of coevolving pairs in contact across breakpoints was smaller than the number of pairs predicted from the null hypothesis (Fig. 3A; paired permutation test $p=0.015$). A smaller but significant difference was seen for all HIV proteins.

Similar results were seen for the mutual information model. For most of the CRFs in the gp120 region, few co-evolving pairs were segregated by the breakpoints (6 of the subtypes are shown in Supplementary Fig. 2). Furthermore, fewer breakpoints were

Download English Version:

https://daneshyari.com/en/article/6140397

Download Persian Version:

https://daneshyari.com/article/6140397

Daneshyari.com