Virus Research 171 (2013) 263–277

Contents lists available at SciVerse ScienceDirect

Virus Research

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

# Aromatic residue mutations reveal direct correlation between HIV-1 nucleocapsid protein's nucleic acid chaperone activity and retroviral replication

Hao Wu[a], Mithun Mitra[b], Micah J. McCauley[a], James A. Thomas[c], Ioulia Rouzina[d], Karin Musier-Forsyth[b], Mark C. Williams[a,*], Robert J. Gorelick[c,**]

[a] Northeastern University, Department of Physics, Boston, MA 02115, USA
[b] The Ohio State University, Department of Chemistry and Biochemistry, Center for Retrovirus Research, Center for RNA Biology, Columbus, OH 43210, USA
[c] AIDS and Cancer Virus Program, SAIC-Frederick, Inc., Frederick National Laboratory for Cancer Research, Frederick, MD 21702, USA
[d] University of Minnesota, Department of Biochemistry, Molecular Biology, and Biophysics, Minneapolis, MN 55455, USA

ARTICLE INFO

*Article history:*
Available online 16 July 2012

*Keywords:*
HIV-1
Infectivity
Nucleic acid chaperone
Nucleocapsid
Single-molecule DNA stretching
Zinc finger

ABSTRACT

The human immunodeficiency virus type 1 (HIV-1) nucleocapsid (NC) protein plays an essential role in several stages of HIV-1 replication. One important function of HIV-1 NC is to act as a nucleic acid chaperone, in which the protein facilitates nucleic acid rearrangements important for reverse transcription and recombination. NC contains only 55 amino acids, with 15 basic residues and two zinc fingers, each having a single aromatic residue (Phe16 and Trp37). Despite its simple structure, HIV-1 NC appears to have optimal chaperone activity, including the ability to strongly aggregate nucleic acids, destabilize nucleic acid secondary structure, and facilitate rapid nucleic acid annealing. Here we combine single molecule DNA stretching experiments with ensemble solution studies of protein-nucleic acid binding affinity, oligonucleotide annealing, and nucleic acid aggregation to measure the characteristics of wild-type (WT) and aromatic residue mutants of HIV-1 NC that are important for nucleic acid chaperone activity. These in vitro results are compared to in vivo HIV-1 replication for viruses containing the same mutations. This work allows us to directly relate HIV-1 NC structure with its function as a nucleic acid chaperone in vitro and in vivo. We show that replacement of either aromatic residue with another aromatic residue results in a protein that strongly resembles WT NC. In contrast, single amino acid substitutions of either Phe16Ala or Trp37Ala significantly slow down NC's DNA interaction kinetics, while retaining some helix-destabilization capability. A double Phe16Ala/Trp37Ala substitution further reduces the latter activity. Surprisingly, the ensemble nucleic acid binding, annealing, and aggregation properties are not significantly altered for any mutant except the double aromatic substitution with Ala. Thus, elimination of a single aromatic residue from either zinc finger strongly reduces NC's chaperone activity as determined by single molecule DNA stretching experiments without significantly altering its ensemble-averaged biochemical properties. Importantly, the substitution of aromatic residues with Ala progressively decreases NC's nucleic acid chaperone activity while also progressively inhibiting viral replication. Taken together, these data support the critical role of HIV-1 NC's aromatic residues, and establish a direct and statistically significant correlation between nucleic acid chaperone activity and viral replication.



*Abbreviations:* 5′UTR, 5′ untranslated region between the LTR and *gag*; A, Ala; BCFU, blue cell focus units; CA, capsid; cTAR, complementary trans-activation response DNA element; dsDNA, double-stranded DNA; ERT, endogenous reverse transcription; F16, Phe16; FA, fluorescence anisotropy; FAM, carboxyfluorescein; FJC, freely jointed chain model for ssDNA elasticity; Gag, group specific antigen; HIV-1, human immunodeficiency virus type 1; LTR, long terminal repeat; NC, nucleocapsid; PAGE, polyacrylamide gel electrophoresis; PCR, polymerase chain reaction; PR, protease; qRT-PCR, quantitative reverse transcriptase PCR; R, repeat region; RT, reverse transcriptase; ssDNA, single stranded DNA; SU, surface glycoprotein; TAR, trans-activation response RNA element; TCIU, tissue culture infectious units; U3, 3′ untranslated region of LTR; U5, 5′ untranslated region of LTR; W37, Trp37; WLC, wormlike chain model for dsDNA elasticity; WT, wild type.

* Corresponding author. Tel.: +1 617 373 7323; fax: +1 617 373 2943.
** Corresponding author. Tel.: +1 301 846 5980; fax: +1 301 846 7119.
*E-mail addresses:* mark@neu.edu (M.C. Williams), gorelicr@mail.nih.gov (R.J. Gorelick).

0168-1702/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.virusres.2012.07.008

## 1. Introduction

The nucleocapsid protein (NC) of HIV-1 performs functions in many of the steps of viral replication (Thomas and Gorelick, 2008). As a domain of $Pr55^{Gag}$ it binds to and directs packaging of genomic RNA (D'Souza and Summers, 2005), as well as placement of the primer $tRNA^{Lys,3}$ (Hargittai et al., 2004) during viral assembly. When the viral protease (PR) cleaves Gag during maturation, the evolving precursor and mature forms of the NC protein (NCp15, NCp9, and NCp7) facilitate formation of a stable dimer of the genomic RNA (Feng et al., 1996; Fu et al., 1994; Fu and Rein, 1993), which is essential for reverse transcription and subsequent infection processes (Kafaie et al., 2008).

The NC protein of HIV-1 is a small, highly basic nucleic acid binding protein that contains 15 basic residues out of a total of 55 amino acids and two strictly conserved zinc-binding motifs of the form –Cys–$X_2$–Cys–$X_4$–His–$X_4$–Cys–. Its amino-terminal tail contains a region of high charge density, which is known to contribute significantly to its ability to aggregate nucleic acids (Darlix et al., 1995; Lapadat-Tapolsky et al., 1993, 1995; Le Cam et al., 1998; Stoylov et al., 1997). Each zinc finger contains a CCHC motif, each of which coordinates one zinc ion. Additionally, the other amino acid residues in these motifs are almost completely conserved as well (Leitner et al., 2003; Thomas and Gorelick, 2008). This sequence conservation indicates important roles in viral replication. Of particular interest for the current report is the aromatic residues present in the first loop of each motif, a Phe at position 16 in the amino-terminal zinc binding motif and a Trp at position 37 in the carboxy-terminal $Zn^{2+}$-finger (Fig. 1A). It is important to note that in virtually every other orthoretrovirus, the second amino acid of the amino-terminal zinc-binding motif is aromatic, although the aromatic residue is less well conserved in the carboxy-terminal motif. Thus, due to their high conservation, it is presumed that these aromatic residues play a critical role in NC function, which we investigate in this work.

Numerous NMR studies have been performed on HIV-1 NC, both with the free- and nucleic acid bound-protein forms (Amarasinghe et al., 2000, 2001; Bazzi et al., 2011; de Guzman et al., 1998; Lee et al., 1998; Morellet et al., 1994, 1998, 1992; South and Summers, 1993; Summers et al., 1990; Tisne et al., 2001). Based on these studies, three key observations have been made: first, the protein becomes more structured when bound to nucleic acids; second, the Trp residue in the carboxy-terminal zinc finger and the Phe residue in the amino-terminal zinc finger form hydrophobic platforms and interact with nucleobases through both stacking and hydrogen bonding interactions; third, the two zinc fingers are in close spatial proximity. Intercalation of the aromatic residue at position 37 with nucleobases is also supported by Trp-fluorescence quenching studies (Mely et al., 1994). Based on additional biophysical characterizations, the aromatic residues play a non-electrostatic role in the binding of NC to nucleic acids (Avilov et al., 2008; Bourbigot et al., 2008; Fisher et al., 2006; Godet and Mely, 2010; Lam et al., 1994; Mely et al., 1993), so that eliminating these residues virtually eliminates this component of binding (Fisher et al., 2006; Remy et al., 1998; Urbaneja et al., 1999). Intriguingly, these experiments have demonstrated that elimination of these aromatic residues can be just as detrimental to interactions between NC and nucleic acids as elimination of the zinc-binding residues.

HIV-1 NC functions as a nucleic acid chaperone, facilitating the rearrangement of nucleic acids into the structure of lowest energy, which should contain the maximum number of base pairs (Levin et al., 2005; Rein et al., 1998; Tsuchihashi and Brown, 1994). These nucleic acid rearrangements are essential for many viral replication processes, including reverse transcription and recombination (Anderson et al., 1998; Bampi et al., 2004; Mark-Danieli et al., 2005; Negroni and Buc, 1999, 2001). They are especially important given the recently characterized complex structures formed by HIV-1 genomic RNA in vivo (Watts et al., 2009). There are a number of steps that require significant rearrangement of nucleic acid secondary structure during retroviral reverse transcription: tRNA primer annealing, genome dimer formation, minus-strand transfer, plus-strand transfer, and whenever recombination occurs during reverse transcription. For example, HIV-1 NC can accelerate tRNA annealing (Hargittai et al., 2004) and minus-strand transfer by over three orders of magnitude in vitro (You and McHenry, 1994).

We have shown that HIV-1 NC's nucleic acid chaperone activity is composed of three major components: nucleic acid aggregation, nucleic acid destabilization, and rapid protein-nucleic acid interaction kinetics (Cruceanu et al., 2006a,b; Levin et al., 2005; Vo et al., 2006; Williams et al., 2001, 2002a,b). We have also shown that NC proteins from other retroviruses exhibit only some of these activities, suggesting that HIV-1 NC has optimal chaperone activity relative to that of other NC proteins (Stewart-Maynard et al.,

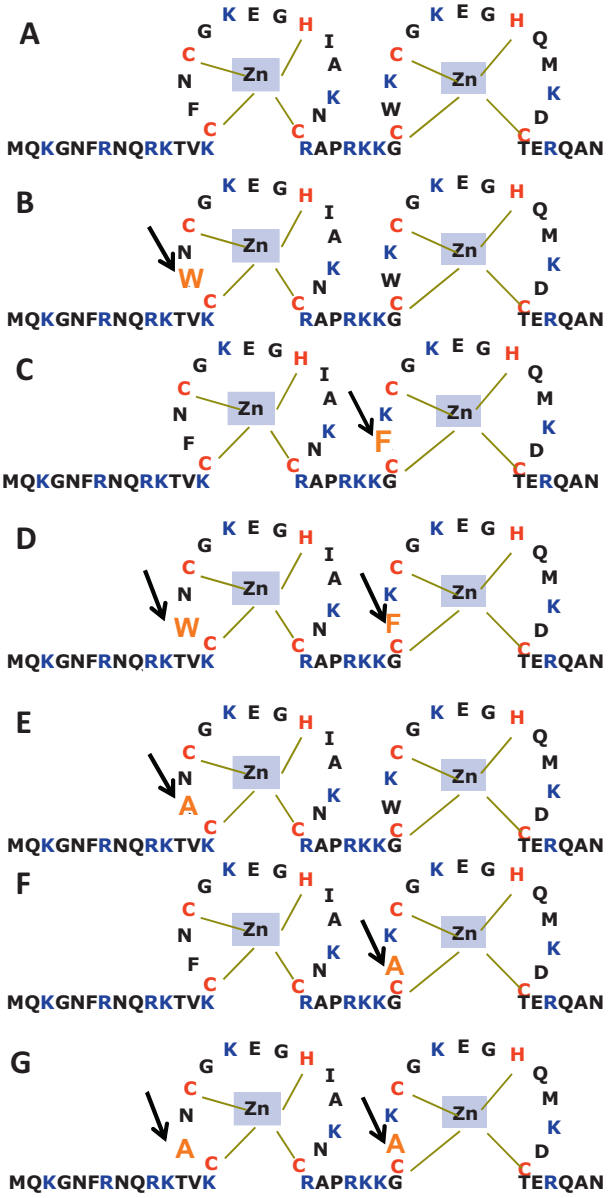


**Fig. 1.** Sequences of wild type (A), F16W (B) W37F (C), F16W/W37F (D), F16A (E), W37A (F), and F16A/F37A (G) HIV-1 NC proteins. Substitutions are denoted by orange text and highlighted with an arrow, while basic residues are presented in blue and zinc coordinating residues in red.