Virology 525 (2018) 117–131

Contents lists available at ScienceDirect

Virology

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

Interplay between sequence, structure and linear motifs in the adenovirus E1A hub protein

Juliana Glavina[a], Ernesto A. Román[b], Rocío Espada[a], Gonzalo de Prat-Gay[c], Lucía B. Chemes[d,e,*], Ignacio E. Sánchez[a,**]

[a] Universidad de Buenos Aires. Consejo Nacional de Investigaciones Científicas y Técnicas. Instituto de Química Biológica de la Facultad de Ciencias Exactas y Naturales (IQUIBICEN). Facultad de Ciencias Exactas y Naturales. Laboratorio de Fisiología de Proteínas. Buenos Aires, Argentina
[b] Instituto de Química y Físico-Química Biológicas, Universidad de Buenos Aires, Junín 956, 1113AAD, Buenos Aires, Argentina
[c] Protein Structure-Function and Engineering Laboratory, Fundación Instituto Leloir and IIBBA-CONICET, Buenos Aires, Argentina
[d] Consejo Nacional de Investigaciones Científicas y Técnicas. Instituto de Investigaciones Biotecnológicas IIB-INTECH, Universidad Nacional de San Martín, San Martín, Buenos Aires, Argentina
[e] Departamento de Fisiología y Biología Molecular y Celular (DFBMC), Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, Buenos Aires, Argentina

ARTICLE INFO

Keywords:
Adenovirus
E1A
Intrinsic disorder
Linear motifs
Motif repertoire
Sequence conservation
Co-evolution
Random polymer

ABSTRACT

E1A is the main transforming protein in mastadenoviruses. This work uses bioinformatics to extrapolate experimental knowledge from Human adenovirus serotype 5 and 12 E1A proteins to all known serotypes. A conserved domain architecture with a high degree of intrinsic disorder acts as a scaffold for multiple linear motifs with variable occurrence mediating the interaction with over fifty host proteins. While linear motifs contribute strongly to sequence conservation within intrinsically disordered E1A regions, motif repertoires can deviate significantly from those found in prototypical serotypes. Close to one hundred predicted residue-residue contacts suggest the presence of stable structure in the CR3 domain and of specific conformational ensembles involving both short- and long-range intramolecular interactions. Our computational results suggest that E1A sequence conservation and co-evolution reflect the evolutionary pressure to maintain a mainly disordered, yet non-random conformation harboring a high number of binding motifs that mediate viral hijacking of the cell machinery.

1. Introduction

The *Adenoviridae* family groups ubiquitous small, non-enveloped DNA viruses with an icosahedral capsid Reddy et al., 2010). Over 250 characterized adenovirus serotypes from five genera infect a wide range of vertebrate hosts (Harrach et al., 2011), with all serotypes infecting humans belonging to the *Mastadenovirus* genus. Adenoviruses infect mucoepithelial cells and persist in the lymphatic system, occasionally infiltrating the cerebrospinal fluid and brain tissue within the central nervous system (Khanal et al., 2018). Adenovirus infection can lead to acute respiratory diseases, pneumonia, gastroenteritis, keratoconjunctivitis and acute hemorrhagic cystitis (Khanal et al., 2018). Because of their oncogenic properties, adenoviruses are often classified together with papillomaviruses and polyomaviruses as small dsDNA tumor viruses. All human adenoviruses can transform baby rat kidney cells *in vitro*, while Human adenovirus 12 (HAdV12) but not Human adenovirus 5 (HAdV5) is also able to cause tumors in immunocompetent rodents (Graham et al., 1977; Williams et al., 2004).

The adenovirus genome is a double stranded linear DNA molecule coding for 20–50 proteins (Davison et al., 2003). The genes transcribed early in the viral reproductive cycle code for proteins involved in virus-host interactions that enable viral genome replication and transcription (King et al., 2018). Among these, the E1A gene is unique to the genus *Mastadenovirus*, which infects a variety of mammals including humans (Davison et al., 2003). The E1A protein is essential for a productive viral infection, deregulating the host cell cycle and transcriptional machinery in favor of conditions suitable for viral replication (Pelka et al., 2008). In cooperation with the E1B protein, E1A transforms rodent cells *in vitro* and is a main oncogenicity determinant (Williams et al., 2004). The biological activity of the E1A protein involves the

* Corresponding author at:.Consejo Nacional de Investigaciones Científicas y Técnicas. Instituto de Investigaciones Biotecnológicas IIB-INTECH, Universidad Nacional de San Martín, San Martín, Buenos Aires, Argentina.
** Corresponding author.
E-mail addresses: lchemes@iib.unsam.edu.ar (L.B. Chemes), isanchez@qb.fcen.uba.ar (I.E. Sánchez).

https://doi.org/10.1016/j.virol.2018.08.012
Received 22 May 2018; Received in revised form 13 August 2018; Accepted 14 August 2018

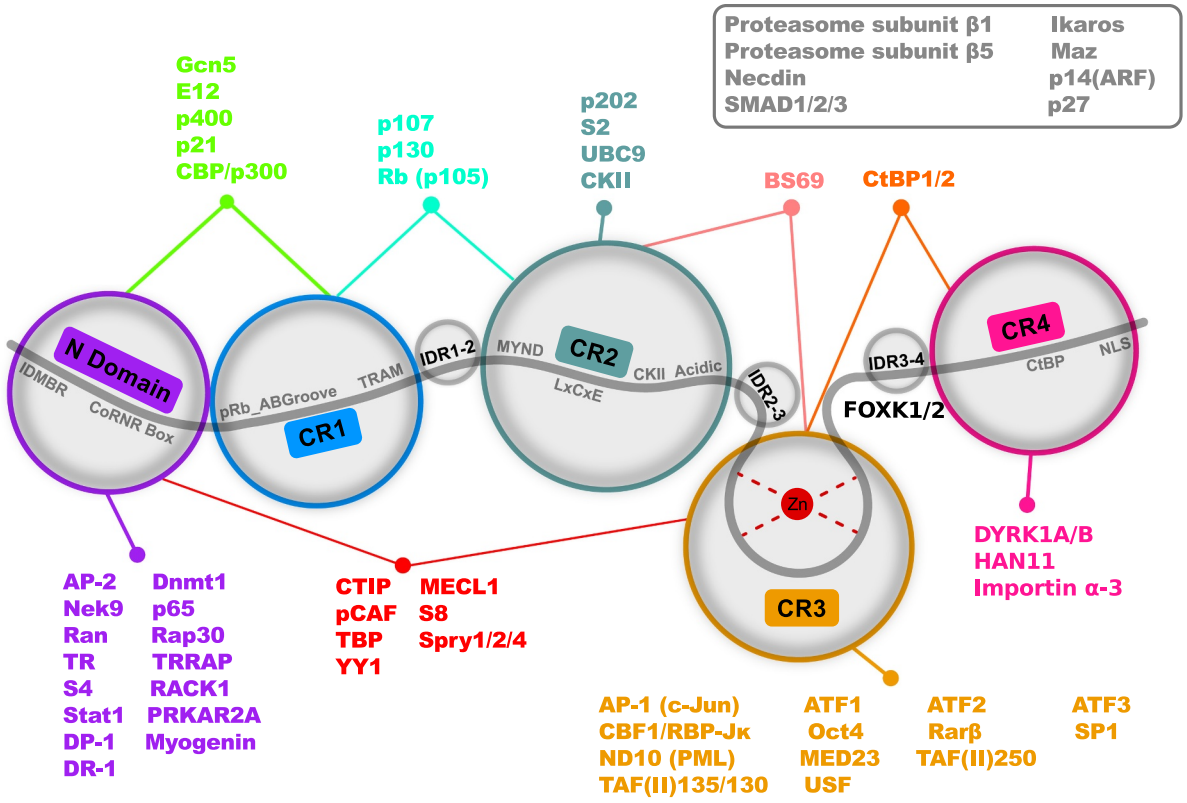

Fig. 1. Schematic representation of E1A oncoprotein domains and interactions. The approximate location of each conserved domain is indicated by colored circles (N domain, purple; CR1, blue; CR2, green; CR3, yellow; CR4, pink). The CR3-associated zinc atom is represented as a red sphere and the approximate location of the linear motifs defined in Table 1 is also shown, namely IDMBR, CoRNR Box, pRB_ABGroove, TRAM, MYND, LxCxE, CKII, Acidic Stretch, CtBP, NLS. The Inter Domain Regions IDR12, IDR23 and IDR34 described in the results sections are indicated by gray circles. The size of each circle is not related to the size of the domain or region. The targets for each domain of the E1A oncoprotein have been compiled from literature (File S1) and are shown grouped according to their target domain/regions, with the unmapped targets are shown in gray.

formation of protein-protein interactions with more than 50 host target molecules (Fig. 1), making it a crucial molecular hub for infection. Well-studied E1A protein targets include the Retinoblastoma (pRb) tumor suppressor protein, p300 and CBP (Boyd et al., 2002; Whyte et al., 1988a; Ferreon et al., 2009), all of which are shared with small DNA tumor virus oncoproteins such as papillomavirus E7 and the polyomavirus Large T antigen. Among other effects, these interactions lead to relocalization and functional inactivation of the retinoblastoma protein family members (pRb, p130, and p107) and of the p300/CBP regulators of histone acetylation (Ferrari et al., 2008). Through this global perturbation in the cell transcriptional network (Ferrari et al., 2008), the E1A protein subverts the cell cycle and contributes to immune system evasion (Strath and Blair, 2006).

The sequence features of the 280-residue long E1A oncoprotein are related to its large number of molecular interactions (Pelka et al., 2008; King et al., 2018). The HAdV5 E1A gene encodes five protein isoforms (Radko et al., 2015). The two largest isoforms differ in an internal sequence of 46 residues (Perricaudet et al., 1979). The largest E1A consists of 5 domains, which are referred to in the literature as follows: N domain, CR1, CR2, CR3 and CR4 (Kimelman et al., 1985; Avvakumov et al., 2002). The CR3 domain is absent in the second largest isoform (Perricaudet et al., 1979). The CR1 and CR2 domains present similarity to sequence stretches within papillomavirus E7 and the polyomavirus Large T antigen. In addition to the canonical domains, a “spacer” was identified as an oncogenic determinant in HAdV12 E1A (Telling and Williams, 1994), and auxiliary regions 1 and 2 were characterized in HAdV5 E1A as co-regulators of viral early gene transcription (Bondesson et al., 1992). However, it is yet unknown whether these regions are present in other E1A proteins.

The structural features of E1A and its regions are still poorly understood. Nuclear magnetic resonance experiments indicated that the CR1, CR2 and CR4 domains of HAdV5 E1A are intrinsically disordered, while the N domain shows partial order and the CR3 domain folds into a poorly understood alpha-helical structure (Pelka et al., 2008; Hošek et al., 2016), but no X-ray diffraction structures have been reported for E1A or its domains. The CR3 domain contains a double CxxC motif involved in structural zinc binding (Culp et al., 1988) reminiscent of the globular domain of papillomavirus E7, although there is no discernible sequence similarity outside of the double CxxC motif. Structural studies using fragments from the HAdV5 and/or HAdV12 E1A proteins further confirmed that the CR1 (Ferreon et al., 2009), CR4 (Molloy et al., 2000) and the N-CR1 domains (Haberz et al., 2016) are disordered when isolated from the rest of the protein. E1A is intrinsically disordered yet able to undergo disorder-to-order transitions. For example, in the presence of trifluoroethanol the binding sites for the CtBP and TBP proteins adopt beta turn and alpha conformations, respectively (Molloy et al., 1998, 1999, 2000). Additionally, both the CR1 domain and the N domain-CR1 constructs undergo local disorder-to-order transitions upon binding of a target protein (Ferreon et al., 2009; Haberz et al., 2016). This folding-upon-binding phenomenon may regulate the simultaneous formation of multiple protein interactions (Ferreon et al., 2013).

Binding of E1A to many of its host cellular targets can be rationalized in terms of multiple linear motifs (Davey et al., 2011), which are short sequence elements of 5–15 residues often found within intrinsically disordered domains (Davey et al., 2012) that mediate many protein-protein interactions and are often key to host-pathogen interactions and molecular mimicry (Davey et al., 2011). Biochemical studies performed mainly on HAdV12 E1A or HAdV5 E1A showed that several E1A domains are densely packed with linear motifs mediating binding to host proteins (Fig. 1, Table 1). The N domain presents an