doi:10.1016/j.jmb.2009.03.013 J. Mol. Biol. (2009) 388, 327–344

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Indirect DNA Readout on the Protein Side: Coupling between Histidine Protonation, Global Structural Cooperativity, Dynamics, and DNA Binding of the Human Papillomavirus Type 16 E2C Domain

Tommaso Eliseo[1]†, Ignacio E. Sánchez[2]†, Alejandro D. Nadra[1,2], Mariano Dellarole[2], Maurizio Paci[1], Gonzalo de Prat Gay[2]* and Daniel O. Cicero[1]*

[1]*University of Rome "Tor Vergata" via della Ricerca Scientifica 1, 00133 Rome, Italy*

[2]*Fundación Instituto Leloir and IIBBA-CONICET, Patricias Argentinas 435 (C1405BWE), Buenos Aires, Argentina*

Received 24 November 2008; received in revised form 6 March 2009; accepted 6 March 2009
Available online 12 March 2009

DNA sequence recognition by the homodimeric C-terminal domain of the human papillomavirus type 16 E2 protein (E2C) is known to involve both direct readout and DNA-dependent indirect readout mechanisms, while protein-dependent indirect readout has been deduced but not directly observed. We have investigated coupling between specific DNA binding and the dynamics of the unusual E2C fold, using pH as an external variable. Nuclear magnetic resonance and isothermal titration calorimetry show that pH titration of His318 in the complex interface and His288 in the core of the domain is coupled to both binding and the dynamics of the β-barrel core of E2C, with a tradeoff between dimer stability and function. Specific DNA binding is, in turn, coupled to the slow dynamics and amide hydrogen exchange in the entire β-barrel, reaching residues far apart from the DNA recognition elements but not affecting the two helices of each monomer. The changes are largest in the dimerization interface, suggesting that the E2C β-barrel acts as a hinge that regulates the relative position of the DNA recognition helices. In conclusion, the cooperative dynamics of the human papillomavirus type 16 E2C β-barrel is coupled to sequence recognition in a protein-dependent indirect readout mechanism. The patterns of residue substitution in genital papillomaviruses support the importance of the protonation states of His288 and His318 and suggest that protein-dependent indirect readout and histidine pH titration may regulate DNA binding in the cell.

Keywords: papillomavirus E2 protein; DNA sequence recognition; indirect readout; histidine; protein dynamics

Edited by J. E. Ladbury

Introduction

The sequence specificity of globular DNA binding proteins arises from their three-dimensional (3D) structure in two ways. First, a number of residues contact the DNA directly by means of correct positioning in 3D space, a mechanism commonly called "direct readout."[1] Histidine side chains often take part in direct readout,[2–6] providing a plausible regulatory mechanism under physiological conditions due to their intrinsic pK_a of 6.3. Second, the dynamics of the globular domain[7–13] and, in particular, the target DNA sequence[1,14–18] modulate binding through indirect effects. This "indirect

*Corresponding authors. E-mail addresses: gpg@leloir.org.ar; cicero@scienze.uniroma2.it.

† T.E. and I.E.S. contributed equally to this work.

Present address: A. D. Nadra, EMBL-CRG Systems Biology Unit, Centre de Regulació Genòmica, Dr. Aiguader 88, 08003 Barcelona, Spain.

Abbreviations used: 3D, three-dimensional; HPV, human papillomavirus; ITC, isothermal titration calorimetry; Mes, 4-morpholineethanesulfonic acid; HSQC, heteronuclear single quantum coherence; NOE, nuclear Overhauser enhancement; HX, hydrogen–deuterium exchange.

readout" can couple DNA binding to protein dimerization and other molecular functions through allosterism.[7,8,12,19,20]

Human papillomaviruses (HPVs) are highly prevalent pathogens of epithelial tropism.[21,22] Mucosal HPV types are the etiological agents of cervical cancer, the second most common cancer in women, and of vaginal, anal, penile, and head and neck cancers.[22,23] Epidemiological studies have identified HPV type 16 as the type with the highest oncogenic potential.[21] The two available prophylactic vaccines[24] are not widely applied in developing countries and cannot cure the millions of people who are already infected. Understanding how HPV proteins function may lead to efficient antiviral compounds.[25] The circular double-stranded DNA HPV genome typically codes for only two late proteins and six early proteins.[26] The early protein E2[27] is a modular polypeptide consisting of an N-terminal transactivation domain, a flexible linker, and a C-terminal domain (E2C) responsible for dimerization and binding of the p53 protein.[28,29] E2C also targets four conserved DNA sites in the regulatory region of the HPV genome,[16,17] leading to the regulation of viral replication and transcription of the early proteins, including the two oncoproteins E6 and E7.[16–18] Fine discrimination between the DNA sequences of the four sites is therefore crucial for the regulation of the viral life cycle.[16–18] Unregulated transcription of E6 and E7 upon integration into the host genome and loss of the E2 gene is associated with cancer.[26]

The native state of the E2C domain is a homodimeric β-barrel consisting of two β–α–β motifs per monomer.[16,17,30–40] The β-strands of the motifs form the conserved eight-stranded β-barrel core of the structure, arranged in two β4–β1–β3–β2 sheets. The interface between the E2C monomers comprises the hydrophobic core of the barrel and of β4–β4′ and β2–β2′ hydrogen bonds. Both α-helices pack against the outside of the barrel, with most residues in helix 1 and Lys349 at the C-terminus of helix 2 being directly involved in DNA binding[16,17,33,35,38] and with most residues in helix 1 and Lys349 at the N-terminus of helix 2 being directly involved in p53 binding.[28,29] Helix 1 is best described as a strained bipartite helix, with a canonical α-helix N-terminal half and a 3_{10} helix C-terminal half.[41] The dimeric topology of E2C is unusual, with the DNA binding domain of the Epstein–Barr nuclear antigen 1 being the only known structural analog.[42] In spite of the structural similarity between these two viral DNA binding domains, their amino acid sequences and DNA binding specificities show no detectable homology. It is not known if this unusual topology is related to the p53 or DNA binding functions of E2C.

The E2C domain binds a pseudopalindromic DNA target sequence with the consensus ACCgaaaxcGGT, where capital letters represent strongly conserved bases and small letters represent weakly conserved bases.[16–18] Sequence recognition by E2C is mediated by direct protein–DNA contacts between the ACCg/cGGT hemisites and the highly conserved N294, K297, C298,Y301, and R302 residues from helix 1 in the first β–α–β element of each monomer.[16–18,33,35,38] Direct sequence readout constitutes the bottleneck of complex formation along the two-state kinetic binding route.[43] Direct E2C–DNA contacts also include nonspecific interactions between the DNA backbone of the ACCg/cGGT hemisites and N294, T295, K297, R300, Y301, R302, K304, and K305 in helix1; Ser315, Thr316, and Trp3177 in β_2-strand; and K349 in helix2.[16–18,33,35,38] Direct E2C–DNA contacts contribute most of the free energy for complex formation.[44] In addition, DNA flexibility can modulate the free energy of binding through indirect readout: the four-base A-rich spacer in the DNA, which does not establish direct contacts with the protein, strongly modulates binding through its ability to bend towards the minor groove.[14–18] We call this phenomenon DNA-dependent indirect readout.

The HPV16 E2C dimeric β-barrel has been proposed to play a role in indirect sequence readout in two ways, mainly from the comparison of the structures and binding specificities of homologous E2C domains. First, its flexibility may regulate the displacement of the DNA binding helices upon binding.[16,17,33,35,38] However, the β-barrel of the HPV16 E2C domain suffers only small changes in structure upon DNA binding,[30,31] in agreement with its overall rigidity in a molecular dynamics simulation relative to its bovine papillomavirus homolog.[45,46] Interestingly, we recently reported that an engineered monomeric version of the HPV16 E2C domain shows the same structure as the dimer, but altered DNA binding properties.[47] This suggested that the dynamics of the β-barrel, rather than the structure, influences DNA binding[47]—a mechanism that has not been investigated to date. Second, the electrostatic potential of the DNA binding side of the β-barrel may also play a role in indirect readout by favoring DNA bending.[16,17,33,35,38] His318 is located in the center of this surface and provides an ideal probe to validate this hypothesis, but its contribution to binding remains uncharacterized.

We present thermodynamic and nuclear magnetic resonance (NMR) evidence for coupling between DNA binding, protonation of His288 and His318, and the global conformational dynamics of E2C.[48] Our results reveal that the β-barrel of the domain is a highly cooperative structural unit that mediates indirect DNA sequence readout. This is also the first report of the E2C–DNA complex dynamics for the HPV type 16 E2 protein, a potential drug target from the most common strain of this oncogenic pathogen.[25]

Results

pH dependence of specific E2C–DNA binding

We have characterized the binding of HPV16 E2C to a cognate nucleotide (site 35; sequence GTAACC-GAAATCGGTTGA) using isothermal titration calorimetry (ITC) (Table 1 and Fig. 1a). We carried out the experiments at 25 °C in a three-component buffer