FEBS Letters 582 (2008) 3619–3624

Characterization of the papillomavirus α_1E2 peptide unfolded to folded transition upon DNA binding

Guilherme Menegon Giesel[a], Luís Maurício T.R. Lima[c], Joana Faber-Barata[c,d], Jorge Almeida Guimarães[a], Hugo Verli[a,b,*]

[a] Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves 9500, CP 15005, Porto Alegre 91500-970, RS, Brazil
[b] Faculdade de Farmácia, Universidade Federal do Rio Grande do Sul, Av Ipiranga 2752, Porto Alegre 90610-000, RS, Brazil
[c] Faculdade de Farmácia, Universidade Federal do Rio de Janeiro, Ilha do Fundão, 21941-590 Rio de Janeiro, RJ, Brazil
[d] Instituto de Bioquímica Médica, Universidade Federal do Rio de Janeiro, Ilha do Fundão, 21941-590 Rio de Janeiro, RJ, Brazil

Received 26 May 2008; revised 11 September 2008; accepted 22 September 2008

Available online 1 October 2008

Edited by Robert B. Russell

Abstract Transcriptional regulation depends on sequence-specific binding of regulatory proteins to their responsive elements in viral DNA. The papillomavirus E2 protein binds to DNA through the consensus sequence ACCG-NNNN-CGGT, activating or inhibiting viral replication. Through molecular dynamics simulations we were able to characterize the role of the DNA molecule on E2 binding region (named α_1E2) conformation, acquiring structural insights for previous works suggesting an unfolded to folded transition upon α_1E2 complexation to DNA. Moreover, the results indicate sites to guide the design of α_1E2 synthetic derivatives to inhibit the HPV infection.

Keywords: Molecular dynamics; Papillomavirus; E2 protein; DNA structure; Induced fit

1. Introduction

The E2 protein is the major regulatory protein from papillomavirus. It acts as both recognition and regulator factors of the early transcription in infected cells [1,2]. The E2 protein is composed by an N-terminal trans-activation domain and a C-terminal DNA binding domain (DBD), separated by a very flexible proline rich linker [2]. While the N-terminal domain is responsible for the interaction with viral protein E1 and other cellular factors like Sp1, TFIIB, and AMF-1 [3–5], the C-terminal domain (E2CT) is responsible for homo- and heterodimerization and DNA binding. The E2CT recognizes and specifically binds to double-stranded DNA comprising the palindromic consensus sequence 5′ACCG-NNNN-CGGT3′, where N can be any nucleotide. The three-dimensional structure of the E2CT has been determined by both X-ray [6–13] and NMR method [14–16] for several viral strains. These structures revealed the E2CT as a dimeric β-barrel composed of four anti-parallel β-sheets with two α-helices (α-helix-1 and α-helix-2) aligned perpendicular to the β-barrel [2]. Most of the amino acids involved in DNA binding are located in α-helix-1, which is described as the recognition helix [1,2], whereas recent data indicate that such protein–DNA recognition is additionally tuned by mechanical and dynamical properties of the whole protein scaffold [17]. The α-helix-1, comprised between residues Arg296 and Arg307 for HPV-18 E2 protein [6], added by residues in β2–β3 loop [17], participate in a hydrogen bond network with the DNA binding site, allowing the occurrence of sequence-specific contacts with the DNA major groove.

Recent data also demonstrate that a synthetic peptide derived from HPV-16 recognition helix (named α_1E2), fully capable of ACCG recognition, is indeed capable to bind and specifically recognize DNA sequences, even outside the context of the full protein [18]. The authors pointed that this peptide is intrinsically unfolded in solution, acquiring some degree of secondary structure only upon complex formation [18], suggesting that binding is independent of a pre-existent helical conformation and leads to a folding transition.

This recognition helix was previously observed by NMR, in the absence of its DBD, to present fast amide exchange rates, suggesting this helix as a particularly flexible region of E2 protein [14], a phenomenon already identified for a number of other DNA binding proteins, as for GCN4 [19], *trp* repressor [20] and *ets* of Fli-1 [21], so probably constituting a common theme in protein–DNA recognition. It was latter demonstrated, both by NMR [22] and molecular dynamics (MD) simulation [23], that this same helix display limited flexibility when bounded to DNA. However, while such unfolded to folded transition was already demonstrated to occur in other biological systems [24], the structural evidence supporting such phenomenon on α_1E2 is still absent.

Recent works had described the recognition between E2 proteins and its binding to DNA, as for HPV-16 and BPV-1 [17,23,25], pointing to a dynamic picture of DBD-E2 recognition where both DNA conformational changes and protein intrinsic deformability appears to cooperate in the recognition process. Nevertheless, such behavior do not explain the unfolded to folded transition of α_1E2 upon binding to DBD [18], a process suggesting an induction of conformational changes in E2 by DNA, i.e. a mutual induced fit process. In this context the current work intends to evaluate the dynamical aspects of the mutual recognition between α_1E2 and its specific DNA binding site through MD of both unbounded and bounded forms of these two molecules. The obtained results allowed the explaining of an apparent contradiction between the

*Corresponding author. Address: Centro de Biotecnologia, Universidade Federal do Rio Grande do Sul, Av Bento Gonçalves 9500, CP 15005, Porto Alegre 91500-970, RS, Brazil. Fax: +55 51 3316 7309.
E-mail address: hverli@cbiot.ufrgs.br (H. Verli).

0014-5793/$34.00
doi:10.1016/j.febslet.2008.09.041

structural information currently available for E2CT, i.e. the absence of secondary structure on α_1E2 in solution and its folded state when bounded to the DBD. Additionally, it suggest potential targets, i.e. nucleotides, and references, i.e. amino acids, for the development of synthetic compounds able to inhibit the E2 protein binding to DNA.

2. Methods

2.1. Nomenclature and software

The recommendations and symbols of nomenclature as proposed by IUPAC [26] were used. The manipulation of structures was performed with VMD [27] and PyMol [28], while the homology modeling was performed with the Swiss-PDB Viewer [29]. All the MD calculations and analysis were performed using the GROMACS package [30] and AMBER99 force field [31]. Analyses of DNA structural parameters were carried out using CURVES [32,33], while PROCHECK [34] was employed to evaluate the content of peptide secondary structure.

2.2. Structure refinement and molecular simulations

The structure of the E2 protein complexed to DNA was retrieved from Protein Data Bank under code 1JJ4 [6]. Based on this structure, four different systems were simulated for 30.0 ns: (1) the isolated DNA sequence containing E2 DBD; (2) the isolated α-helix 1, from E2 protein (α_1E2 peptide, residues 294-311); (3) the complex formed by DNA and α_1E2 peptide and; (4) the complex formed by a mutated DNA sequence and α_1E2 peptide. These structures were solvated in a rectangular box using TIP3 P water model [35] by a layer of at least 10 Å from the solute (α_1E2, DNA or α_1E2-DNA complex) atoms. Counter ions (Na^+ or Cl^-) were added to neutralize the systems charges. The MD protocol employed was based on previous MD studies [36]. The Lincs and Settle methods [37,38] were applied to constrain covalent bond lengths, allowing an integration step of 2 fs after an initial energy minimization using steepest descents algorithm under periodic boundary conditions. The so obtained systems, composed by α_1E2–solvent–ions, DNA–solvent–ions, and α_1E2–DNA–solvent–ions (in its $G_{22}C_{21}C_{20}A_{19}$ or $A_{22}A_{21}A_{20}A_{19}$ containing sequences) were heated slowly from 50 K to 310 K, in steps of 5 ps, in which the reference temperature was increased by 50 K. Temperature and pressure were kept constant during all simulations by coupling the protein, nucleic acid, ions, and solvent to external temperature and pressure baths with coupling constants of $\tau = 0.1$ and 0.5 ps [39], respectively. No restraints were applied after the thermalization phase. The electrostatic interactions were evaluated by the particle–mesh Ewald method [40] with a charge grid spacing of 1.2 Å, while Coulomb and Lennard–Jones interactions were evaluated using a 9.0 Å atom-based cutoff [41]. The analyses were performed in all trajectory length.

The amino acids of α_1E2 peptide were numbered following the crystallographic sequence (from N- to C-terminal)

Gly294-Asp295-Arg296-Asn297-Ser298-Leu299-Lys300-Cys301-Leu302-Arg303-Tyr304-Arg305-Leu306-Arg307-Lys308-His309-Ser310-Asp311, and the numbering of the simulated native DNA sequence is

$5'$-C_1 A_2 A_3 C_4 C_5 G_6 A_7 A_8 T_9 $T_{10}C_{11}G_{12}G_{13}T_{14}T_{15}G_{16}$-$3'$

$3'$-$G_{32}T_{31}T_{30}G_{29}G_{28}C_{27}T_{26}T_{25}A_{24}A_{23}G_{22}C_{21}C_{20}A_{19}A_{18}C_{17}$-$5'$,

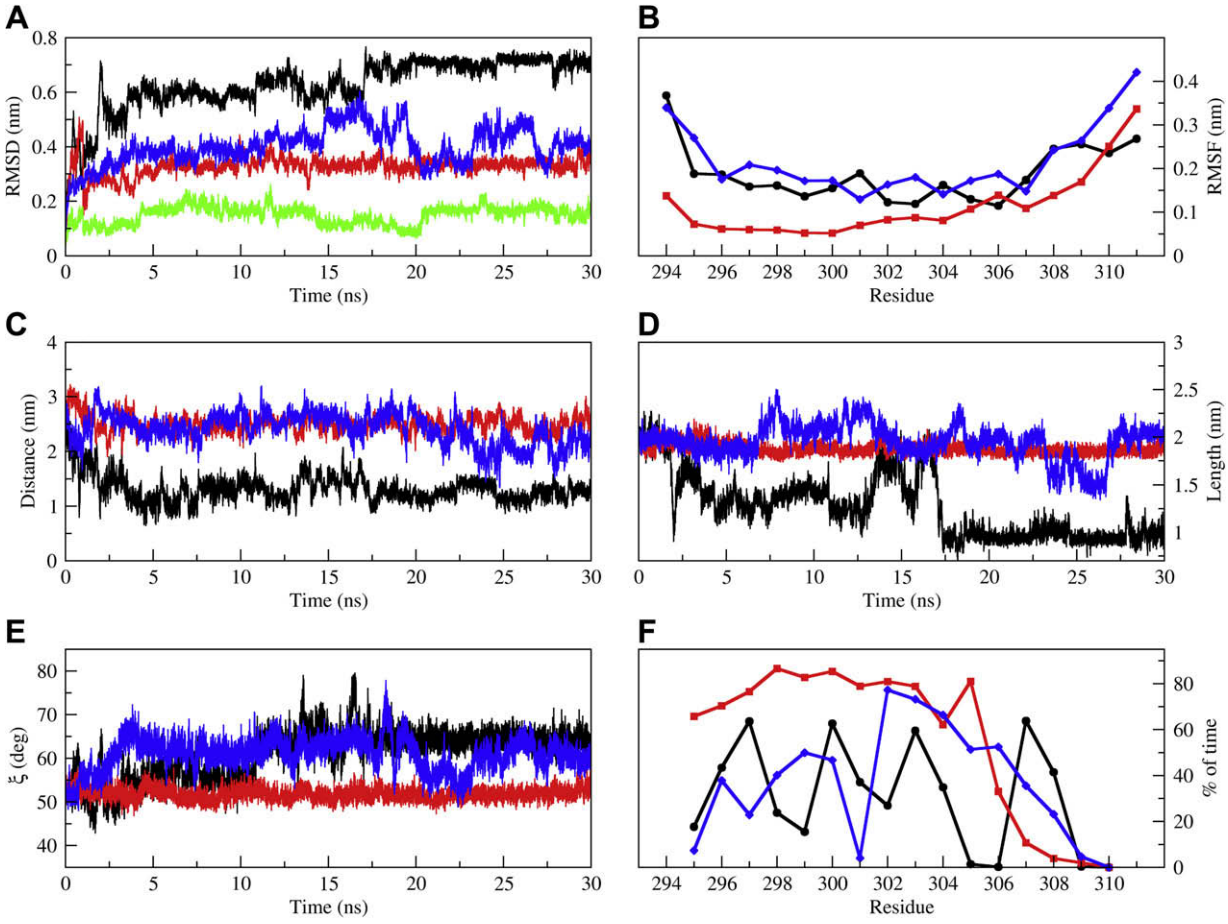

Fig. 1. MD analyses as a function of time or residue number for α_1E2 in its uncomplexed form for all amino acid residues (black) and for residues 296–303 (green), when complexed with the non-mutated DBD (red) and with the mutated DBD (blue) for the following properties: (A) all-atom RMSD from crystal structure; (B) all-atom average root mean square fluctuation (RMSF) over the 30.0 ns simulations; (C) distance between N- and C-terminal residues; (D) α-helix length; (E) average pseudo dihedral ξ (i.e. the dihedral formed by 4 consecutive α-carbon atoms Cαi, $C\alpha i_{+1}$, $C\alpha i_{+2}$, and $C\alpha i_{+3}$) [47] over the peptide chain, and F) ellipticity at 222 nm (percentage of time in which the residues presents a helix character) [48].