



**Pontifícia Universidade Católica do Rio Grande do Sul**  
**Faculdade de Biociências**  
**Programa de Pós-Graduação em Biologia Celular e Molecular**



**MODELAGEM POR HOMOLOGIA E DINÂMICA MOLECULAR DA  
ESTRUTURA SELVAGEM COMPLETA DE E6 DE HPV 16**

**Camila Mundt Ilgenfritz**

**Orientado por: Virgínia Minghelli Schmitt**  
**Co-Orientado por: Osmar Norberto de Souza**

DISSERTAÇÃO SUBMETIDA AO PROGRAMA DE  
PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E  
MOLECULAR – PUCRS COMO REQUISITO PARA  
A OBTENÇÃO DE TÍTULO DE MESTRE EM  
BIOLOGIA CELULAR E MOLECULAR.

**Porto Alegre**

**2009**

O presente trabalho foi realizado no período de Janeiro a Julho de 2007 e Agosto de 2008 a Maio de 2009 no Laboratório de Bioinformática, Modelagem e Simulação de Biosistemas (LABIO/FACIN) da PUCRS, sob orientação dos Professores Dra. Virgínia Minghelli Schmitt e Dr. Osmar Norberto de Souza.

## **Agradecimentos**

**Ao. Dr. Osmar Norberto de Souza.**

**À Dra. Virgínia Minghelli Schmitt.**

**À minha mãe e meu pai.**

**Ao Cristiano.**

**Aos amigos.**

**... cada um sabe porquê.**

## Resumo

O câncer cervical é um problema que vem afetando cada vez mais mulheres de todo o mundo e, se não detectado em tempo, pode resultar em alta taxa de letalidade. Virtualmente todos os cânceres cervicais têm como agente causal essencial o Papilomavírus Humano (HPV). A Organização Mundial da Saúde (OMS) estima que cerca de 660 milhões de pessoas são infectadas todos os anos, além de ser o vírus que causa Doenças Sexualmente Transmissíveis (DST) mais comum do trato genital. Os HPVs são subdivididos em várias categorias, sendo classificados por tipo e subtipo, grau de risco (alto e baixo risco) e pelo sítio preferencial de infecção (cutâneo, genital, etc).

Os HPV genitais de alto risco (HR-HPV) dos tipos 16 e 18 são responsáveis por aproximadamente 70% dos casos de câncer cervical no mundo, com alguma diferença de prevalência dependendo da região. Estudos têm aumentado muito o conhecimento acerca destes vírus nos últimos anos, mas algumas questões ainda ficam em aberto devido a dificuldades técnicas e experimentais. Um foco recente importante dos estudos sobre estes vírus tem sido buscar o entendimento sobre a estrutura e funções das oncoproteínas E6 e E7 e seus alvos moleculares, com intuito de buscar novos alvos ou metodologias para o desenvolvimento de vacinas ou fármacos contra o HPV.

Neste trabalho, buscamos complementar os trabalhos já desenvolvidos na área de modelagem estrutural da proteína E6 de HPV 16, que tem como principal função a condução de p53, proteína celular supressora de tumores, à degradação. O trabalho foi desenvolvido através de modelagem por homologia utilizando como molde o modelo experimental parcial mutado publicado por Nominé et al (2006). O modelo obtido apresenta dois domínios parcialmente independentes

(N- e C-terminal) ligadas por um conector central (*linker*) sem estrutura secundária definida. Apesar de a proteína apresentar grande tendência à desordem intrínseca, os domínios mantiveram a maior parte de suas topologias ao longo da simulação em meio aquoso, indicando boa qualidade do modelo sugerido.

## **Abstract**

Cervix cancer is a worldwide problem, increasingly affecting women all over the globe which could lead to high mortality rates when not diagnosed in early stages. Virtually all cervical cancers has Human Papillomavirus (HPV) as a necessary cause. World Health Organization (WHO) estimates around 660 million people infected each year by the virus, besides being the most common viral Sexually Transmitted Disease (DST). HPVs are subdivided by type and subtype, carcinogenic risk (high or low risk) and by infection site (cutaneous, genital, etc).

The high-risk HPV (HR-HPV) types 16 and 18 are responsible together for approximately 70% of the cervical cancer cases in the world, showing some prevalence differences among regions. Studies have provided increasing knowledge on HPV infection and oncogenic capacity, but some aspects are still to be elucidated, mainly due to technical and experimental difficulties. An important recent focus of HPV research has been the search for defining the structure and function of E6 and E7 viral oncoproteins and their molecular targets, aiming to find new targets and methodologies for the development of vaccines or drugs against HPV.

In this work, our goal is to complement the available research on the tridimensional structure modeling of HPV 16 E6 protein, which interacts with tumor suppressor p53 as a main target, among others, leading to its degradation. The model on this work was developed by homology modeling, using the partial mutated experimental C-terminal model published by Nominé et al (2006) as template. The model we generated is a protein with two partially independent domains (N- and C-terminal) connected by a central linker that lacks a secondary structure. In spite the high tendency the protein has to intrinsic disorder, each domain of the final model

maintained most of its topology throughout the aqueous solution simulation, indicating the good quality of our suggested model.

# Índice

<b>Resumo</b>	4
<b>Abstract</b>	6
<b>Lista de Figuras</b>	10
<b>Lista de Tabelas</b>	12
<b>Lista de Abreviações</b>	13
<b>1 Apresentação do tema</b>	14
<b>1.1 Papilomavirus Humano</b>	14
<b>1.2 Características e funções das oncoproteínas E6 e E7</b>	17
<b>1.3 A estrutura da E6 e sua interação com outras proteínas</b>	23
<i>1.3.1 Estudos mutacionais e de deleção</i>	24
<i>1.3.2 Estrutura e modelagem</i>	26
<b>2 Objetivos e Justificativa</b>	37
<b>3 Materiais e Métodos</b>	38
<b>3.1 Material de origem</b>	38
<b>3.2 Modelagem Comparativa</b>	39
<i>3.2.1 E6C selvagem</i>	39
<i>3.2.2 E6N</i>	40
<i>3.2.3 Modelagem e montagem da estrutura quimera E6N/E6C ...</i>	41
<b>3.3 Dinâmica molecular: simulação em meio aquoso</b>	43
<b>3.4 Outras Análises da estrutura</b>	44
<b>4 Resultados</b>	46
<b>4.1 A modelagem E6 selvagem</b>	46
<i>4.1.1 E6Cwt</i>	46
<i>4.1.2 E6Nwt</i>	48
<i>4.1.3 E6wt</i>	52
<b>4.2 Análise da estrutura</b>	58
<b>4.3 Dinâmica molecular</b>	61
<i>4.3.1 Proteína completa</i>	61
<b>5 Discussão</b>	69
<b>6 Artigo Científico</b>	83
<b>7 Considerações Finais e Perspectivas</b>	130
<b>8 Referências Bibliográficas</b>	131



<b>Apêndice I – Arquivos de entrada para modelagem por homologia</b>	<b>141</b>
<b>Apêndice II – Arquivo de Saída do PROSCAN/PBIL</b>	<b>144</b>
<b>Apêndice III – Resumo da revisão sobre estudos mutacionais e de deleção em E6 de HPV 16</b>	<b>152</b>
<b>Apêndice IV – TBP e E6: existem semelhanças?</b>	<b>155</b>

## Lista de Figuras

Figura 1: Esquema do genoma do HPV 16. _____	16
Figura 2: Alvos celulares conhecidos de E6 que participam das funções de E6. _____	20
Figura 3: Esquema da interação de E6 com E6AP e p53. _____	22
Figura 4: Esquema de alguns dos resíduos conservados ao longo da cadeia de E6 de HPV 16. ____	27
Figura 5: Esquema comparativo das topologias encontradas por Ullman (1996) e Nominé (2006). _	28
Figura 6: Representação dos tipos <i>ribbons</i> e palitos de diferentes domínios PDZ interagindo com a extremidade C-terminal de E6. _____	29
Figura 7: Representação do tipo <i>ribbons</i> do modelo experimental obtido por NMR do domínio C-terminal da proteína E6 de HPV contendo 4 mutações C/S não conservadas. _____	33
Figura 8: Representação do tipo <i>ribbons</i> , demonstrando a interação entre os domínios N- (azul) e C-terminal (amarelo) conforme sugerido por Nominé et al (2006). _____	33
Figura 9: Representação do tipo <i>ribbons</i> do Modelo 3D de E6 6C/6S, indicando seqüências de resíduos que não foram inclusos. _____	34
Figura 10: Alinhamento utilizado para a geração do modelo da extremidade C terminal de E6. ____	40
Figura 11: Alinhamento que deu origem à estrutura do domínio N terminal de E6. _____	40
Figura 12: Alinhamento entre a seqüência estrutura molde (intermediária) e a seqüência completa de E6. _____	42
Figura 13: Estrutura da porção C-terminal E6 de HPV 16 modelada por homologia a partir do molde de 2FK4, modelo 3. _____	47
Figura 14: Distâncias Cys-Zn de 6E6C_15. _____	48
Figura 15: Ângulos Cys-Zn-Cys de 6E6C_15. _____	48
Figura 16: Modelo da estrutura E6N (4E6N_10) elaborado por modelagem por homologia. _____	51
Figura 17: Representação do modelo 3D do tipo <i>ribbons</i> de 4E6N_10 de HPV 16. _____	52
Figura 18: Representação do tipo <i>ribbons</i> da estrutura 3D da proteína E6 do HPV 16 obtida por modelagem por homologia. _____	53
Figura 19: Detalhe de um dos ângulos do modelo final da estrutura de E6 de HPV16. _____	54
Figura 20: Representação do tipo palitos da face de interação entre os dois domínios (N- e C-terminal) de E6 de HPV 16. _____	54
Figura 21: Representação do tipo palito com coloração CPK das distâncias Cys-ZN nos sítios de coordenação do zinco no modelo da estrutura completa de E6 de HPV 16 obtida por modelagem por homologia. _____	57
Figura 22: Resultado da análise da ferramenta PPI-Pred, representado com janela de intervalo de 3 posições. _____	59
Figura 23: A-D: Gráficos indicativos do valor do RMSD em Angstroms, de diferentes porções da proteína E6 completa (A: 9-71: N-term; B: 80-144: C-term; C: 1-151: proteína completa) em relação às suas porções equivalentes da estrutura do modelo inicial, ao longo do tempo (em picosegundos). E: Médias das distâncias Cys-Zn ao longo de toda a simulação. _____	67
Figura 26: Gráfico representando a melhora do desempenho do modelo ao longo da simulação antes e depois da minimização de energia _____	68
Figura 27: Galeria de interação entre oncoproteínas virais e proteínas supressoras de tumor. ____	70
Figura 28: Comparação entre a predição de desordem intrínseca e a estrutura secundária de E6 de HPV16. _____	79

Figura 29: Gráfico representativo do resultado da análise de *Fold Index* da proteína E6 de HPV 16 com janela de 26 resíduos. \_\_\_\_\_ 80

Figura 30: Representações do tipo *ribbons* dos modelos 3D de HPV16 E6 (A) e porção C-terminal (155-339) de hTBP (B). \_\_\_\_\_ 156

Figura 31: Comparação dos resultados de desordem intrínseca e FoldIndex entre E6 de HPV 16 e TBP. \_\_\_\_\_ 159

## Lista de Tabelas

Tabela 1: Resultado da análise dos modelos individuais em 2FK4 para a escolha do template para o processo de modelagem. _____	39
Tabela 2: Alças 1 a 3 das estrutura completa, refinados individualmente. _____	42
Tabela 3: Dados dos sistemas das utilizados nas dinâmicas moleculares de simulação em meio aquoso. _____	44
Tabela 4: Tabela comparativa dos resultados para o modelo obtido de E6 C-terminal selvagem (6E6C_15) e seu molde E6C 4S/4C (2FK4_3). _____	47
Tabela 5: Resultados das análises dos modelos da estrutura de E6N. _____	50
Tabela 6: Resumo dos resultados do PROSITE com corte de 70%. _____	60
Tabela 7: Valores de RMSD para as porções de E6 de HPV 16 ao longo da trajetória da dinâmica molecular, em relação ao modelo inicial 5CHI_03. _____	63
Tabela 8: Tabela mostrando resultados comparativos dos modelos da estrutura de E6C que sofreram a simulação. _____	66
Tabela 9: Tabela resumindo o levantamento de resultados obtidos por diversos autores em estudos mutacionais e de deleção utilizando a proteína E6 de HPV 16. _____	152

## Lista de abreviações

3D – tridimensional

AcMo – Anticorpo Monoclonal

Bak – proteína pró-apoptótica

C-term – Carboxi-terminal

CBP/300 – Proteína ligante de CREB

E6AP – Proteína Associada à E6

HPV – Papillomavirus humano

HR – alto risco (High risk)

hTERT – subunidade catalítica da telomerase humana

hSCRIB – Proteína humana SCRIB homóloga à SCRIBBLE de *Drosophila*

IRF – Fator Regulador do Interferon

MAGI-1 – Proteína Guanilato Quinase Associada à Membrana (*Membrane associated guanylate kinase*)

N-term – Amino-terminal

MAGUK – Família de Proteínas Guanilato Quinase Associadas à Membrana

MUPP – MPDZ (*Multiple PDZ domain protein*)

NLS – Sinal de Localização Nuclear (*Nuclear Localization Sign*)

NMR – Ressonância Magnética Nuclear (*Nuclear Magnetic Resonance*)

p21 – WAF-1/CIP-1 – proteína inibidora de proteína da família das ciclínas/CDF

P27 – proteína inibidora de CDK

p53 – proteína codificada pelo gene TP53

PDZ (domínio PDZ) – domínio estrutural comum de 80 a 90 aa, que auxiliam a ancoragem de proteínas transmembrana ao citoesqueleto e mantêm juntos complexos de sinalização

PHD (domínio PHD) – Homeodomínio de planta (*Plant homeodomain*)

pRB – proteína do Retinoblastoma

TBP – TATA Box *Binding Protein*

TNF – Fator de Necrose Tumoral

wt – selvagem (*wild type*)

ZB – *zinc binding* (... de ligação a zinco)

## 1 Apresentação do tema

O câncer cervical tem como agente causador em sua maioria o Papilomavírus Humano (HPV). Os HPVs são classificados como de baixo e alto risco de acordo com sua prevalência em cânceres cervicais. Os HPVs dos tipos 16 e 18 são classificados como de alto risco e juntos são responsáveis por aproximadamente 70% dos casos de câncer cervical.

### 1.1 Papilomavírus Humano

O Papilomavírus Humano (HPV) é um agente que está presente em, virtualmente, todos os cânceres cervicais, em uma quantidade significativa de cânceres da área orofaríngea e em alguns cânceres de pele e retais. Os HPVs podem ser classificados em dois grandes grupos, os de alto e os de baixo risco, de acordo com sua propensão à transformação celular. Dentre os de alto risco estão os tipos 16, 18, 31, 33 e 45. O HPV 16 é responsável, sozinho, por mais de 50% dos casos de câncer cervical em todo mundo, e junto com o tipo 18, somam aproximadamente 70% dos casos (Bosch et al, 2002).

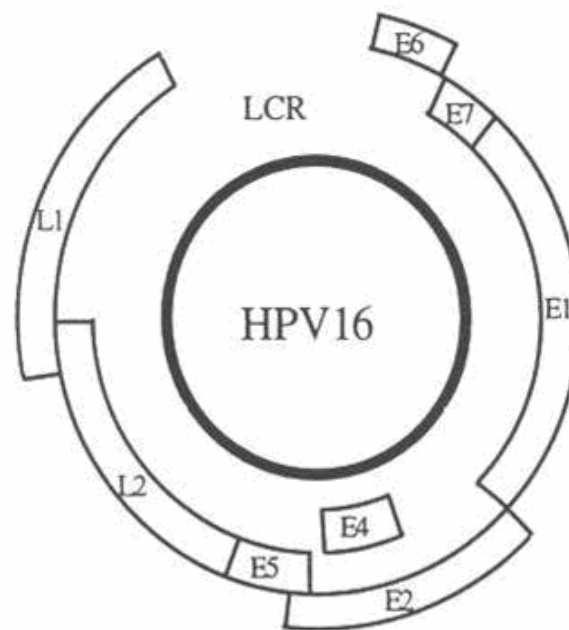
O genoma do HPV é uma dupla fita de DNA circular de aproximadamente 8 kb, com pequenas variações de tamanho entre os tipos e subtipos. Todos os genes são transcritos a partir da mesma fita do DNA e 3 regiões podem ser identificadas: a primeira região é chamada de *Early* (precoce) e contém as regiões de leitura aberta, ORFs (do inglês *Open Reading Frames*), que codificam as proteínas E1<sup>E7</sup>; a segunda região é chamada *Late* (tardia) e contém as ORFs L1 e L2, que codificam apenas as proteínas estruturais que formam o capsídeo viral; a terceira região é a LCR (*Long Control Region*), de aproximadamente 1 kb, onde se localizam o promotor principal e outras seqüências reguladoras (Howley, 2001;

Schmitt, 2006).

As proteínas expressas no início da fase precoce são E1, E2 e E5. A ação destas proteínas promove a replicação do genoma viral, aumentando a quantidade de cópias para 50 a 100 por célula. Ao chegar a este patamar, a replicação se mantém em baixos níveis apenas para assegurar uma infecção latente das células. As proteínas E1 e E2 são as responsáveis pelo início da transcrição, alongamento do DNA e recrutamento de fatores celulares de replicação. A passagem da fase precoce para a fase tardia do ciclo vira está, de alguma forma, vinculada à diferenciação da célula hospedeira e à regulação por E5 (Fehrmann et al, 2003). A proteína E5 é uma oncoproteína cujas funções ainda não foram totalmente elucidadas. Esta proteína apresenta muitas características conservadas entre os vários tipos de HPV e é capaz de formar complexos com diversas proteínas celulares. Na célula infectada, E5 encontra-se associada com a membrana do Complexo de Golgi, na região perinuclear e na membrana citoplasmática. Vários estudos já demonstraram que E5 é expressa desde o início da infecção até a fase pré-maligna de lesões em células escamosas cervicais (Howley, 2001). Um esquema do genoma do HPV 16 pode ser visto na Figura 1.

O HPV replica-se de forma episossomal multicópia, eventualmente integrando-se no genoma da célula hospedeira. A replicação é dependente da diferenciação celular por mecanismos ainda a serem elucidados. O vírus infecta, inicialmente, as células da camada basal do epitélio, onde há proliferação celular. A infecção produtiva, momento em que há a liberação de novas partículas virais, ocorre nas camadas mais externas com a expressão de L1 e L2, e é regulada por genes como E1 e E2 – estes genes impedem e regulam o ciclo e a expressão dos genes virais assegurando a infecção latente e a formação de novas partículas virais.

Foi relatado que em tumores causados por HPV não há infecção produtiva (zur Hausen, 1996), provavelmente devido à integração que ocorre, na maioria das vezes, previamente à malignização. Na integração, acontece uma quebra do genoma viral na ORF E2. A perda de E2 leva a um desequilíbrio da expressão e/ou função de E6 e E7, levando a célula a uma proliferação descontrolada, que pode contribuir para a malignização da lesão (Gammoh et al, 2006).



**Figura 1:** Esquema do genoma do HPV 16.

A classificação de tipos, subtipos e variantes de HPV é feita com base na similaridade das seqüências das ORFs E6, E7 e L1. Para ser classificado como mesmo tipo, a seqüência destas ORFs deve ter até 90% de semelhança com o protótipo, entre 90 e 98% para um mesmo subtipo e mais de 98% para mesma variante (De Villiers, 1989; Van Ranst et al, 1993).

Estas diferenças entre as seqüências das proteínas dos HPVs, dificulta a busca por drogas e vacinas eficientes contra todos tipos, subtipos ou variantes de HPV e até mesmo por um sistema diagnóstico para detectar todos



HPVs que possam estar envolvidos em neoplasias (Stewart, 1996; Termine et al, 2008).

## **1.2 Características e funções das oncoproteínas E6 e E7**

As proteínas codificadas pelas ORFs E6 e E7, por sua função oncogênica, são as proteínas de HPV melhor estudadas até o momento. Ambas as proteínas são multifuncionais. A principal função de E7 é inativar pRb (proteína do Retinoblastoma). A de E6 por sua vez, sabe-se que é diferente entre os grupos de HPV. Nos HPV de alto risco de mucosas, uma das principais funções de E6 é marcar p53, proteína supressora de tumores, para degradação pela subunidade 26S do proteossoma (Scheffner et al, 1990; Werness et al, 1990; Huibregtse et al, 1991 e 1993). Em HPVs cutâneos, E6 não interage com p53 nem induz à sua degradação (Elbel et al, 1997), mas induz a proteína pró-apoptótica Bak à degradação (Jackson et al, 2000). Até o presente momento, não foi descoberto nenhum alvo de degradação para E6 de papilomavírus bovinos.

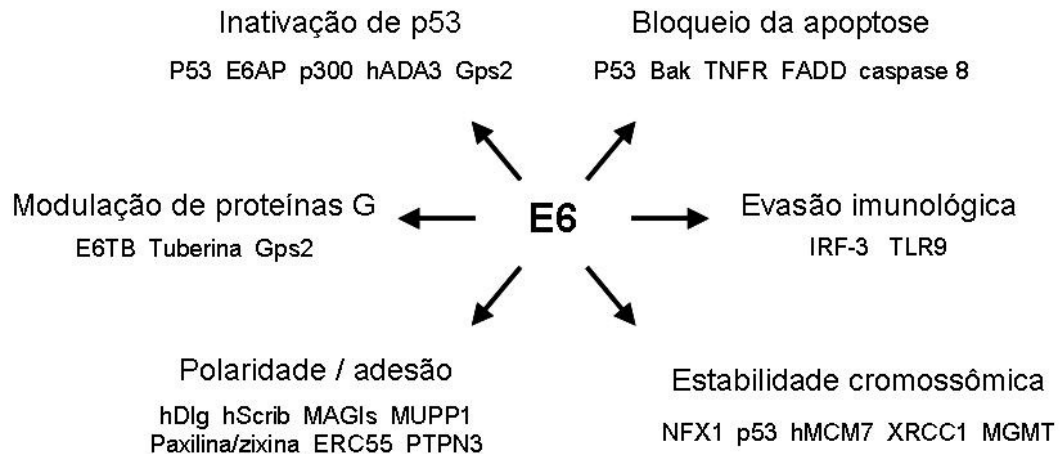
Apesar da inativação destes alvos, mesmo separadamente, ser suficiente para a imortalização da célula, é, no entanto insuficiente para a transformação celular (Münger et al, 1989). Park e Kang (2000) mostraram que ambas as proteínas E6 e E7 dos HPVs de alto risco são responsáveis por gerar grande instabilidade genética. Estes autores demonstraram que E6 e E7 dos HPVs de alto risco, mas não dos de baixo risco, combinadas ou separadamente, aumentam as taxas de mutações espontâneas em presença de carcinógenos como o derivado do tabaco, MNNG (N-Metil-N'-nitro-N-nitrosoguanidina). Outros estudos também demonstraram formas de instabilidade genética, como atrasos no sistema de reparo por excisão de nucleotídeos, após incidência de radiação ultravioleta em

células imortalizadas por HPVs dos tipos 16 e 18 (Rey et al, 1999). E7 se liga à forma hipofosforilada (ativa) da proteína pRb (Howley, 2001), desestabilizando-a e promovendo sua proteólise por via independente de ubiquitinação (Gonzalez et al, 2001; Helt e Galloway, 2003). Uma vez formado o complexo E7/pRb, o fator de transcrição E2F, que se mantinha inativo em associação com pRb, é liberado e atua como ativador transcricional, levando a um avanço do ciclo celular. E7 também interage com diversos outros fatores celulares, como a TBP (proteína que se liga à TATA box) (Massimi et al, 1997), a S4 (subunidade do proteossoma) (Berezutskaya e Bagchi, 1997), a p21 (Helt et al, 2002), a p27, além de induzir a duplicação anormal do centrôssomo (Duensing e Münger, 2003), levar as células rapidamente à aneuploidia e à instabilidade genômica (Duensing e Münger, 2002) e reduzir a expressão de fibronectina (Lee e Park, 2000).

A p53 é a principal proteína responsável por pausar o ciclo celular entre G1 e S em resposta a fatores de estresse, como a radiação ultravioleta ou outros agentes que causam danos no DNA e ativar uma série de sinais que levam ao início de um programa de reparo de DNA. No caso de danos irreparáveis, à indução de apoptose. A proteína E6 dos HPVs de alto risco liga-se à p53 e, na presença de E6AP (*E6 Associated Protein*), promove a ubiquitinação e degradação da p53. Como consequência, o ciclo celular progride mesmo quando há danos no DNA. A E6AP é uma proteína citoplasmática celular de aproximadamente 100 kDa que forma um complexo estável com E6 e p53, quando há excesso de E6 em relação à E6AP (Huibregtse, 1993). A proteína E6AP faz parte da família E3 de ubiquitina-ligases responsáveis pela especificidade protéica da ubiquitinação, que é realizada por sua vez, por membros da família E2 de ubiquitina-ligases (Finley e Chau, 1991). Huibregtse et al (1993) também demonstraram que E6 ou E6AP,

separadamente, não são capazes de se ligar de forma estável à p53, sendo estável apenas o complexo ternário.

E6 interfere em outras funções importantes para a integridade celular, por exemplo, ativando a transcrição de hTERT - a subunidade catalítica da telomerase (Veldman et al, 2001); inibindo a indução de interferon  $\beta$  através da interação com IRF-3 (Ronco et al, 1998); inibindo a atividade transcricional de p53, interagindo com CBP/p300 (Zimmermann et al, 1999; Patel et al, 1999); inibindo a apoptose via Bak através de interação com a própria Bak (Thomas e Banks, 1999; Jackson et al, 2000) e via TNF, ligando-se ao receptor de TNF, TNF R1 (Filippova et al, 2002). E6 também interage com hScrib e com proteínas das famílias MAGI-1 e MUPP-1, enviando-os para degradação mediado pelo proteossoma por um mecanismo ainda não elucidado (Nakagawa e Huijbregtse, 2000; Glaunsinger et al, 2000; Thomas et al, 2002; Lee et al, 2000). E6 apresenta sítios de ligação para seqüências conservadas em alguns de seus diversos alvos celulares, semelhantes ao motivo  $\alpha$ -hélice presente em E6-AP, necessários para a função de ubiquitinização de p53 (Nguyen et al, 2002). Therrien et al (1999) compararam o efeito de E6 ou E7 sobre o reparo de DNA em uma linhagem de fibroblastos humanos e demonstraram que ambas proteínas influenciam negativamente na eficiência dos sistemas de reparo, sendo o efeito da E6 mais intenso do que o de E7. Uma revisão das atividades conhecidas de E6 foram sumarizadas na revisão de Howie et al (2009) e estão esquematizadas na Figura 2.



**Figura 2:** Alvos celulares conhecidos de E6 que participam das funções de E6. E6 altera várias cascatas celulares através da ligação com numerosas outras proteínas. Modificado de Howie et al (2009).

Li et al (1996) conseguiram realizar uma co-immunoprecipitação de E6 e p53 a partir de um sistema *in vitro*. No mesmo trabalho, eles descrevem a diferença de afinidade das E6 dos HPVs de alto e baixo risco pela p53. As proteínas E6 em geral ligam-se à região carboxi-terminal de p53, mas apenas as dos HPVs de alto risco conseguem se ligar à região central (*core*) de p53 e promover sua ubiquitinação. Estudos *in vivo* demonstraram que mesmo sendo alta a eficiência de degradação de p53 ( $88,0 \pm 2,0\%$ ), ainda há quantidades detectáveis da proteína, provavelmente devido à regulação por mecanismos compensatórios, como a estabilização do mRNA de p53 e elevação da atividade transcricional por vias auto-reguladas (Gardiol e Banks, 1998; Kawamata et al, 2002). Desaintes et al (1997), Butz et al (1995) e Scheffner et al (1990) demonstraram que, sob certas condições, várias linhagens celulares positivas para HPV apresentam atividade de p53, apesar de também já ser consenso que a presença de E6 na célula diminui a meia-vida da p53 de algumas horas para cerca de 20 minutos (Howley, 2001).

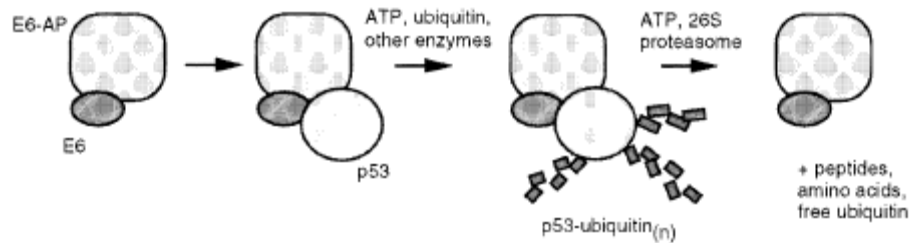
A forma de ligação de p53 à E6 ainda está sendo muito estudada. Freedman e Levine (1998) demonstraram que para a degradação de p53 é necessária a translocação da p53 para o citoplasma, onde será degradada. Já foi

identificado o provável sítio de ligação de E6 na proteína p53, que estaria entre os resíduos 100 e 150. A E6 de HPV 16 é mais eficiente em degradar p53, em comparação a E6 de HPV 18 (Scheffner et al, 1990). No entanto, há evidências de que HPV 18 induz tumores mais agressivos do que HPV 16. Esta diferença pode ser devida à força da interação da proteína E6 com as proteínas MAGUK (*Membrane Associated Guanylate Kinase homologues*) que é maior para HPV 18 do que para HPV 16. As proteínas MAGUK são proteínas multifuncionais que atuam mediando uma série de funções associadas à membrana e que possuem um número variável de domínios PDZ. HPV 18 possui a sequência consenso de ligação à PDZ perfeita, enquanto HPV 16 possui a sequência degenerada (Gardioli et al, 1999; Thomas et al, 2001; Zhang et al, 1995), como será descrito mais adiante. No entanto, é interessante ressaltar que apenas HPVs de alto-risco mucosa-específicos são capazes de realizar esta função (Pim et al, 2002).

Vaeteewoottecharn et al (2005) demonstrou que E6 colocaliza-se com E6AP principalmente no citoplasma, mas E6AP também apresenta localização nuclear e perinuclear. No entanto, Hengstermann et al (2001) e Daniels et al (1998) afirmam que a localização de E6AP é predominantemente citoplasmática e, em menor extensão, no núcleo. Além disso, Tao et al (2003) mostraram que existem três sinais de localização nuclear em E6 (aa Arg8-Lys11, Lys65-Lys72 e Lys115-Arg124) que, ao sofrerem mutações pontuais, levam a perdas de função de E6 (p. ex., a degradação de p53), indicando que a presença de E6 no núcleo é necessária. Da mesma forma, Masson et al (2003) demonstraram a localização preferencial de E6 de HPV 16 no núcleo de várias linhagens de carcinoma cervical.

A interação de E6 com E6AP, por sua vez, é essencial para a degradação de p53 (Figura 3), mas também é responsável por um amplo espectro

de funções resultantes da expressão de E6, como demonstrado pelo estudo de Kelley et al (2005) onde E6AP media virtualmente todas as funções que causam impacto no perfil transcricional de linhagens HPV positivas.



**Figura 3:** Esquema da interação de E6 com E6AP e p53. Ordem de eventos na degradação de p53, dependente de E6. E6 e E6AP formam um complexo que precede a associação com p53; p53 se junta ao complexo e é multi-ubiquitinizado e então degradado pelo proteossoma 26S. (Fonte: Huibregtse e Beaudenon, 1996)

Algumas variações na seqüência de p53 que gerariam resistência à degradação mediada por E6 já foram descritas. A maioria destas variantes são alterações pontuais que, sugere-se, interferem no sítio de ligação entre as duas proteínas. Storey et al (1998) demonstraram que a variação alélica Pro72Arg está relacionada com maior resistência da variante prolina à degradação do que a variante arginina e que a p53 de indivíduos heterozigotos seria mais resistente que a de homozigotos para arginina. Outros autores pesquisaram a variante arginina como fator de risco para o desenvolvimento de cânceres ocasionados por HPV e encontraram uma relação entre variante e incidência em diferentes etnias (Agorastos et al, 2000; Zehbe et al, 2001; Makni et al, 1999, Mitra et al, 2005). No entanto, essa constatação ainda é controversa, pois outros autores não confirmaram estes achados (Brady et al, 1999; Kawamata et al, 2002; Rosenthal et al, 1998; Helland et al, 1998). Em 2002, Calhoun et al avaliaram, além do códon 72, outros polimorfismos de p53 sem, no entanto, encontrar relação com a progressão da lesão. Alguns autores sugerem que existe, na realidade, uma perda seletiva de heterozigidade, ou seja, em indivíduos 72ProArg que apresentam transcritos de HPV, ocorre a perda

da expressão do alelo 72Arg, sem no entanto haver uma relação ou maior risco nestes indivíduos ou em homozigotos 72ArgArg de incidência de cânceres de cabeça e pescoço (Scheckenbach et al, 2004).

Rodriguez et al (2000) identificaram um mutante sintético de p53 que seria mais resistente a degradação por E6/E6-AP. Este mutante tem 6 resíduos de lisina substituídos por arginina ao longo da porção C-terminal de sua cadeia, nas posições 370, 372, 373, 381, 382, 386 e apresenta distribuição subcelular similar ao p53 *wt*. No entanto, mutantes como  $\Delta$ 370-393 ou Lys381Ile/Lys382Ile/Lys386Ile são eficientemente degradados. Outros mutantes ainda, foram isolados a partir de câncer de pulmão também apresentam resistência: Cys135Tyr, Ala154Val, Gly266Val e Cys277Phe - estes mutantes, no entanto, provavelmente alteram a conformação da proteína e não necessariamente estão relacionados com os sítios de interação com E6 ou E6AP (Scheffner et al, 1992).

### **1.3 A estrutura da E6 e sua interação com outras proteínas**

O desenvolvimento de novos fármacos e vacinas pode ser amplamente auxiliado com o conhecimento mais extenso das características físico-químicas e estruturais das proteínas virais oncogênicas. Estas características, uma vez conhecidas, tornam as proteínas foco de estudos que fazem delas alvos terapêuticos para vacinas (terapêuticas ou preventivas) e tratamentos (utilizando abordagens como docagem de moléculas inibidoras, sintéticas ou não). Para tanto, o conhecimento acerca da estrutura tridimensional da proteína E6 dos papilomavirus humanos é de suma importância. O foco deste trabalho é a proteína E6 do HPV 16, e por este motivo, nos ateremos mais a ela.

### 1.3.1 Estudos mutacionais e de deleção

O foco da maioria dos trabalhos que buscam entender as características funcionais de E6, até o momento, foi entender a relação da estrutura de pequenas porções da proteína, separadamente ou com alguns de seus ligantes, através de estudos utilizando mutações pontuais sítio dirigidas ou deleções de pequenas porções.

As proteínas E6 de HPV 16 e 18 apresentam alto grau de conservação de funções entre si, o que permitiu a identificação de domínios funcionais nas moléculas tanto para ligação a p53 quanto para a estimulação da degradação da mesma *in vitro*. Até 1994, estudos mutacionais haviam sugerido que a região C-terminal de E6 de HPV 16 e ambos os domínios de E6 de HPV 18 seriam necessários para ligar a p53, e também que a região N-terminal é a que deteria a função de degradação de p53 (Pim et al, 1994 e Crook et al, 1991). Park e Androphy (2002) mostraram que a deleção dos aminoácidos His118 a Lys122 de E6 impede a ativação da telomerase, reduz a degradação de p53 e baixa a afinidade por E6AP *in vitro*, mas mantém boa eficiência de degradação *in vivo*. Gardiol e Banks (1998) identificaram resíduos Thr21 a Asp25, Gln36 a Arg39, Phe47 a Asp49 e Thr86 a Gln91 como resíduos provavelmente envolvidos na degradação de p53 *in vivo* através de deleções destas pequenas sequências e da comparação da taxa de degradação de p53 com a proteína completa.

Lagrange et al (2005) demonstraram que anticorpos monoclonais contra o segundo domínio de ligação ao zinco de E6 (aminoácidos Lys94 a Arg144) bloqueiam a ligação de E6 a p53 e E6AP, sendo que 3 deles, afetam a ligação a p53 independente de E6AP (3B8: aa Gln116-Asp120; 1F5: aa Asp120-Phe125; 3F8: aa His126-Arg131). Foster et al (1994), testaram uma série de deleções e mutações



pontuais na proteína E6 e verificaram que a maioria inativa a proteína de forma integral ou parcialmente. Como até então, e mesmo agora, não havia sido encontradas alterações de sequência que distinguíssem com convicção HPVs de alto e baixo risco, foi sugerido que as características que levam a esta diferenciação e conseqüentemente levam a proteína a realizar funções específicas de E6 de alto-risco como a degradação de p53, são de ordem estrutural e/ou conformacional. Desta forma, a integridade da porção N-terminal e dos domínios de ligação ao zinco, bem como a região central seriam necessários para manter tal estrutura/conformação. No entanto, alguns dos resultados obtidos pelo autor foram parcialmente contraditórios aos obtidos para os mesmos mutantes por outros grupos de pesquisa (Crook et al, 1991 e Lechner e Laimins, 1992). Por exemplo, Crook et al detectaram ligação do mutante E6  $\Delta$ Pro9-Pro13 a p53 sem levar a degradação, enquanto Foster et al não detectaram ligação alguma.

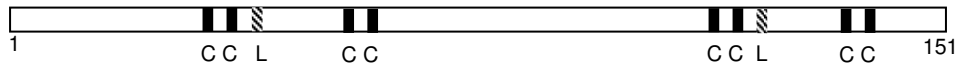
Um fator que implica em maior complexidade no entendimento dos estudos de mutantes é o fato de a interação com E6AP ser necessária para estabilizar o complexo E6-p53. Algumas mutações podem interferir na interação com E6AP enquanto outras na interação direta com p53 de qualquer maneira, a função de degradação de p53 estará comprometida. A proteína E6AP interage com E6 através de uma hélice na região central da E6AP localizada entre os resíduos 391 e 408 (necessária e suficiente para a ocorrência da ligação; Huibregtse, 1993). Além de E6, E6AP e ATP, são necessários mais dois componentes para realizar a ubiquitinação de p53: E1 (enzima ativadora de ubiquitina) e E2 (uma de várias enzimas conjugadoras de ubiquitina). E6AP, por sua vez, parece catalisar a formação de uma ligação isopeptídica entre a ubiquitina e o substrato. Desta forma, a ativação da ubiquitina seria realizada por E1, na forma de tioéster, e seria então,

transferida para E2 e depois para E3 (E6AP), que catalisa a formação da ligação da ubiquitina à extremidade C-terminal do substrato. E6AP pode ser reativada por E2 sem dissociar-se do substrato, e o resultado é um substrato de alto peso molecular multi-ubiquitinizado, que será então degradado de forma ATP dependente pelo proteossoma 26S (revisado por Huibregtse e Beaudenon, 1996).

### 1.3.2 Estrutura e modelagem

A proteína E6 é uma proteína básica e pequena, de 151 aminoácidos, e possui dois motivos  $C_4$  de ligação a zinco. Ullman et al (1996), que foram os primeiros a estudar a estrutura secundária de E6 e E7, basearam-se em similaridades de sequências e na localização de resíduos-chave em comparação com motivos de ligação ao zinco (ZB) de hormônios nucleares e fatores de transcrição. Eles chegaram à conclusão que os motivos ZB de E6 provavelmente possuem topologia  $\alpha\beta\alpha\beta$ .

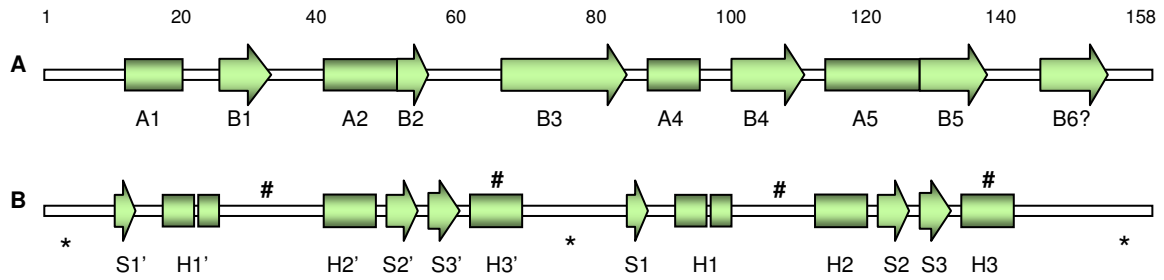
Paralelamente, Lipari et al (2001) demonstrou resultados que são consistentes com modelo onde E6 é um monômero que possui dois domínios ZB funcionalmente críticos. A proteína E6 tem 8 resíduos de cisteína conservados ao longo de sua cadeia (Figura 4) dispostas em dois grupos Cys-X-X-Cys-X<sub>29</sub>-Cys-X-X-Cys compondo as duas regiões de ligação a zinco, sendo uma em cada domínio (C- e N-terminal). Outros resíduos conservados também estão presentes nas mesmas posições relativas nas duas porções da proteína, como por exemplo, Lys38 e Lys111. As similaridades entre os dois domínios, tanto no modelo de Lipari quanto no de Ullman, sugerem que eles apresentam a mesma estrutura secundária com seqüência de  $\alpha\beta\alpha\beta$ .



**Figura 4:** Esquema de alguns dos resíduos conservados ao longo da cadeia de E6 de HPV 16, neste caso cisteínas (C) e leucinas (L) (adaptado de Ullman, 1996).

Em 2004, Swaminathan, R; Malik, BK e Balaji, S, depositaram no PDB um modelo de estrutura (número de acesso: 1VZN) identificada como E6 de HPV 16. O trabalho que deveria descrever tal estrutura nunca chegou a ser publicado (consta no arquivo do PDB “*to be published*”). Sua estrutura diverge totalmente das outras estruturas secundárias ou terciárias já obtidas até hoje. É formada de um único domínio com predominância de hélices alfa e não apresenta qualquer estrutura de ligação a zinco. Não há qualquer publicação (artigo) destes autores com relação a HPV ou a proteína E6 (busca realizada no Highwire, Pubmed e RCBS Protein Data Bank). Por estes motivos, não vamos ater-nos a este modelo.

Posteriormente, os trabalhos de Nominé et al (2003, 2005, 2006) apresentaram uma proposta de modelo em contraposição à estrutura secundária proposta por Ullman et al, como pode ser visto na Figura 5. Nominé et al (2005) realizaram um estudo de seu mutante E6C 4C/4S e demonstrou a presença de um íon de zinco nesta extremidade e identificaram as cisteínas conservadas como sendo os resíduos envolvidos nesta interação. Nominé et al, em artigo publicado em 2001, já previram a posição provável do íon de zinco no domínio C-terminal (como pode ser visto na Figura 5) interagindo com a alça grande L2 (entre H1 e H2) e com a hélice H3. Acredita-se que a extremidade N-terminal também comporte um íon de zinco com características físicas e estruturais semelhantes.

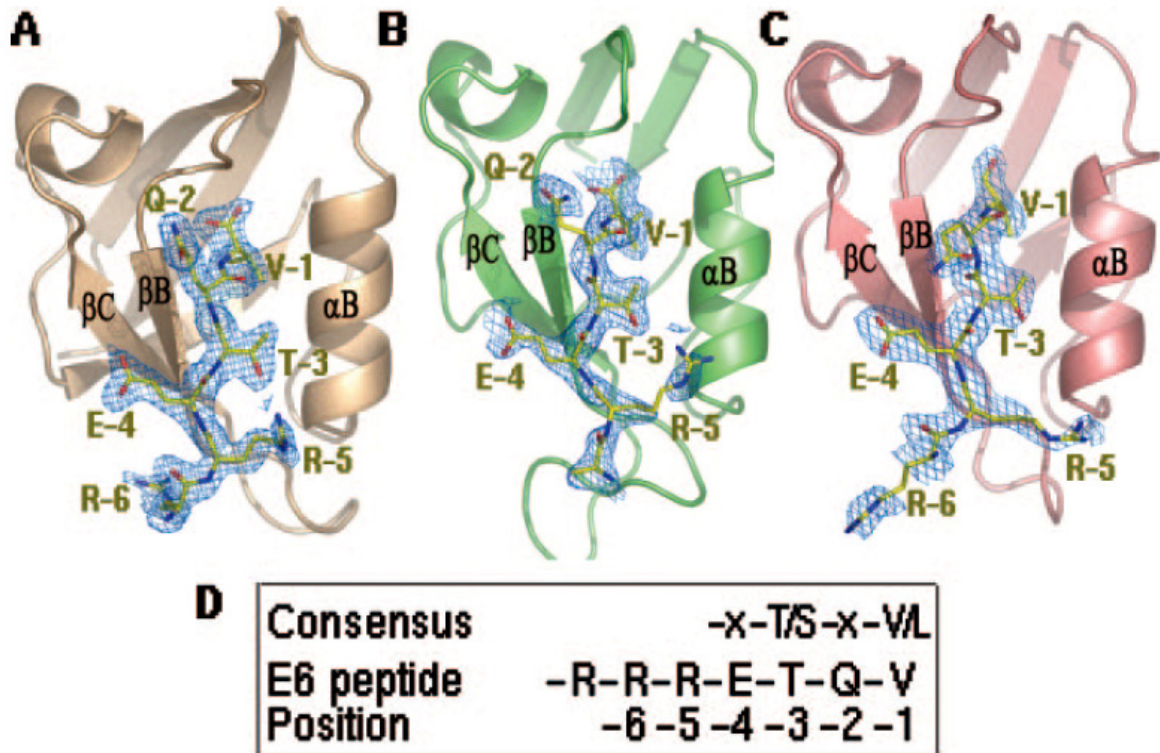


**Figura 5:** Esquema comparativo das topologias encontradas por Ullman (1996) e Nominé (2006). Topologias referentes às estruturas secundárias encontradas pelos autores (A) Ullman et al, 1996 e (B) Nominé et al, 2006. Os números acima da figura indicam o número dos resíduos na sequência de aminoácidos; (\*) indica a posição das seqüências que foram excluídas no modelo de Nominé; (#) indica a provável posição de ligação dos íons zinco na proteína.

Kiyono et al (1997) identificaram a sequência consenso ETQL de ligação ao domínio PDZ em hDLG (*human homologue Drosophila discs large tumor suppressor protein*) na extremidade C-terminal da proteína E6 de HPV 16 e ETQV em E6 de HPV 18 (Kiyono et al, 1997; Pim et al, 2002). Outros autores já haviam sugerido, através de resultados de experimentos cristalográficos, que a sequência consenso XS/TXV/L é responsável pela ligação a diferentes domínios PDZ de classe I e que a especificidade pelos domínios PDZ é dada pelo primeiro resíduo X da sequência (Doyle et al, 1996; Songyang et al, 1997; Hung e Sheng, 2002).

Zhang et al (2007) determinaram a estrutura do polipeptídeo terminal RRRETQV, através de cristalografia por difração de raio X, interagindo com o domínio PDZ 1 de MAGI-I e PDZ 2 e 3 de SAP97/Dlg9 (estas últimas, responsáveis pela junções aderentes e junções compactas, respectivamente). O peptídeo utilizado do HPV 18 contém a sequência consenso de ligação a domínios PDZ, como citado anteriormente (ver Figura 6) nas posições 154-158. A estrutura do peptídeo ligado ao domínio PDZ apresenta as seguintes características: 1) forma uma folha  $\beta$  adicional, anti-paralela à folha  $\beta$ B de PDZ; 2) o grupo carboxil da cadeia principal do peptídeo de E6 está ancorado aos nitrogênios da cadeia principal do motivo GLGF de PDZ; 3) dois grupos metil do resíduo V de HPV 18 E6 interagem com um bolso hidrofóbico nos 3 domínios PDZ testados (Figura 6). Esta estrutura apresentada

estaria intimamente relacionada com a interação com o ligante, no caso os domínios PDZ, pois em contato com o ligante adquire conformação de fita beta.



**Figura 6:** Representação dos tipos *ribbons* e palitos de diferentes domínios PDZ interagindo com a extremidade C-terminal de E6. A – complexo peptídico MAGI-1 PDZ1-E6 – PDB ID nº 2I04; B – complexo peptídico SAP97/Dlg PDZ2-E6 – PDB ID nº 2I0L; C – complexo peptídico SAP97/Dlg PDZ3-E6 – PDB ID nº 2I0I; D – consenso do ligante de PDZ e a sequência de E6 usada na publicação. A posição (-1, -2, -3, etc.) está assinalada abaixo de cada aminoácido. Os modelos das estruturas dos domínios PDZ estão representadas por *ribbons* e a do consenso de ligação de E6 por palitos e VDW. A sequência de HPV16 E6 compreende apenas uma alteração em relação à utilizada no trabalho de Zhang que corresponde à substituição do resíduo Valina terminal pela Leucina (L), mantendo, ainda, a conformidade com a sequência consenso da interação. (Figura retirada de Zhang, 2007).

O tratamento em condições de baixa salinidade (que em condições limitantes clivam preferencialmente entre domínios e nas extremidades não estruturadas da sequência) sugere que E6 seja composto de dois domínios ZB dobrados individualmente (N-term resíduos Met1-Arg77 e C-term, resíduos Tyr81-Leu151 – Nomine et al, 2003), corroborando com Lipari et al (2001). Alguns experimentos de dicroísmo circular realizados por Nomine et al (2003) demonstraram que as estruturas obtidas para as porções N- e C-terminais são

similares e corroboram com resultados encontrados anteriormente de que o conteúdo de estruturas secundárias entre as duas porções também são equivalentes.

Alguns autores sugeriram que a proteína E6 é, também, uma proteína de ligação ao DNA que liga a “Junções de DNA com 4 braços” (*four way DNA junctions* ou Junção de Holliday) (Cole e Danos, 1987; Ristriani et al, 2000, 2001). Esta função foi identificada ao notar-se certa similaridade de sequência com a endonuclease VII do bacteriófago T4 (White et al, 1997), sendo que a semelhança com a endonuclease VII é maior com a E6 dos HPV de alto-risco, do que com os de baixo-risco. Além disto, também se observou que não houve ligação de E6 de HPV de baixo-risco com as Junções de Holliday, corroborando a suposição de que a ligação ao DNA depende da presença destas características em comum entre a E6 dos HPV de alto-risco e a endonuclease VII. A similaridade entre as proteínas é encontrada na porção de ligação ao zinco na região N-terminal da endonuclease VII e a porção C-terminal de E6 (Ristriani et al, 2000).

Um dos principais problemas encontrados até hoje para desenvolver o modelo tridimensional de E6 completo pelos métodos experimentais tradicionais (NMR e cristalografia) é a dificuldade de obter a proteína recombinante purificada estável, solúvel e na sua forma nativa. Uma das razões prováveis para isto acontecer é a quantidade particularmente grande de cisteínas, que implica na necessidade de controle extremo de condições redutoras. Nominé et al (2003) desenvolveu um mutante da porção C-terminal de E6 (aa 80-151) que possui 4 cisteínas não conservadas substituídas por serinas (Cys80Ser, Cys97Ser, Cys111Ser e Cys140Ser), chamado E6C (4C/4S). Este polipeptídeo mantém a função de ligar ao DNA e é estável em solução. A partir dele, Nominé e seus colegas

iniciaram os estudos da estrutura 3D desta proteína. Nominé demonstrou que a proteína E6, quando super-expressa em bactérias, formava corpos de inclusão (uma massa de seqüências peptídicas não enoveladas). No entanto, com a E6 fusionada com a MBP (*maltose binding protein*), os autores obtiveram agregados não enovelados, porém solúveis. Em 2007, o mesmo grupo obteve a proteína fusionada com MBP, porém desta vez, formando agregados esféricos, de grande peso molecular e bem definidos. Estas estruturas possuem dois íons de zinco por monômero, o que sugere que estes agregados apresentam dobramento do tipo nativo, conservando a coordenação do centro de ligação a zinco (Zanier et al, 2007).

Nominé et al (2005, 2006) assumiram que esta proteína recombinante, apesar de conter 4 mutações, se assemelha suficientemente à proteína selvagem pelos seguintes motivos: 1) as cisteínas mutadas não são conservadas nos alinhamentos; 2) a mutação Cys-Ser é relativamente isostérica; 3) as mutações permitem ao domínio adotar uma conformação estável; 4) todas as posições mutadas ficam expostas na superfície da estrutura calculada; 5) essas mutações, quando presentes na estrutura completa de E6 (E6 6C/6S) mantêm as propriedades de degradação *in vivo* e *in vitro* de p53 em níveis iguais aos de E6 selvagem. Os autores assumiram, por consequência, que a estrutura também não seria alterada em função das modificações. Em 2001, o grupo de Nominé realizou experimentos de dicroísmo circular que indicam as mesmas estruturas  $\alpha/\beta$  em E6N, E6C e E6 completa.

Cole e Danos (1987) já sugeriram que a origem evolutiva desta proteína é a duplicação de uma proteína com um único domínio e, como consequência, os domínios adotariam dobramento similar. Supondo, que as duas metades das proteínas tenham de fato, estruturas similares, Nominé et al (2006)

modelaram a porção N-terminal (E6N) de E6 utilizando a C-terminal (E6C) como modelo. Conforme sugerido pelo autor, o modelo obtido para E6N a partir de E6C mostra algumas similaridades e algumas diferenças esperadas que geram certa especulação:

1) Quase todos os resíduos internos correspondem a posições hidrofóbicas conservadas;

2) Parte dos resíduos são conservados em todas as proteínas E6 de HPV (modelos simplificados experimentais foram propostos para outros tipos de HPV). Sugere-se que estes resíduos (p.ex. E41, R48, E89 e E114) podem ter funções estruturais ou genéricas compartilhadas entre as espécies de HPV;

3) Existem resíduos expostos conservados nos dois domínios, que formam uma região hidrofóbica, o que sugere que estas regiões fazem contato entre si.

A presença destas regiões hidrofóbicas expostas que existem em E6C e E6N deram margem para que Nominé e colegas tentassem realizar o encaixe entre as peças deste quebra-cabeça. Este encaixe deve acontecer de forma a minimizar choques estéricos e esconder as regiões hidrofóbicas do contato com a água. A estrutura experimental de E6C 4C/4S está demonstrada na Figura 7. A forma de interação dos dois domínios (N- e C-terminal) sugerida pelos autores pode ser vista nas Figuras 8 e 9.



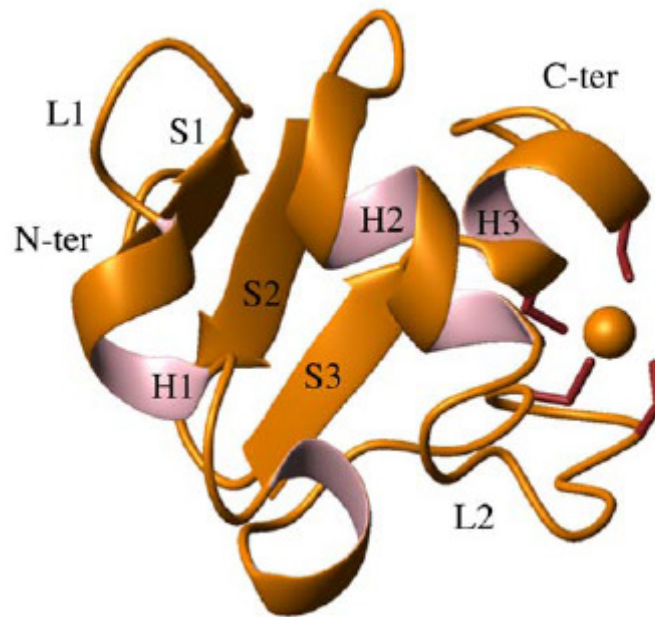


Figura 7: Representação do tipo *ribbons* do modelo experimental obtido por NMR do domínio C-terminal da proteína E6 de HPV contendo 4 mutações C/S não conservadas (Nordine et al, 2006). A esfera representa o íon  $Zn^{2+}$ , e os resíduos de cisteína que coordenam o zinco estão representados por palitos. A representação mostra a seqüência de estruturas secundárias: S1-L1-H1-L2-H2-S2-S3-H3. A estrutura foi publicada no banco de dados PDB sob o código de acesso 2FK4 (retirado de Nordiné et al, 2006)

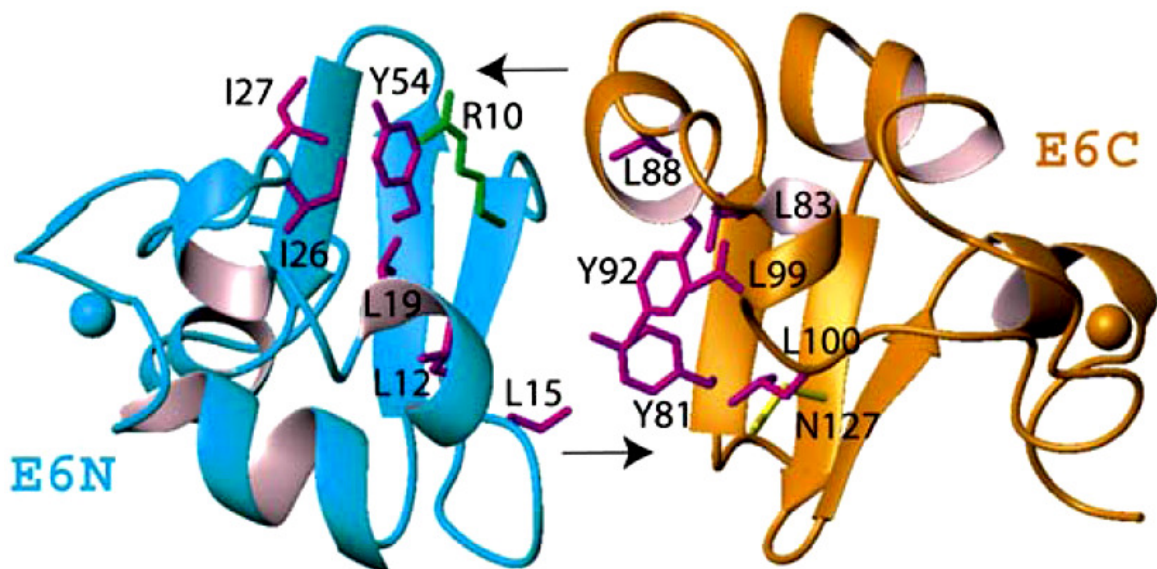


Figura 8: Representação do tipo *ribbons*, demonstrando a interação entre os domínios N- (azul) e C-terminal (amarelo) conforme sugerido por Nordiné et al (2006). Os resíduos responsáveis pela interação, conforme este modelo, estão representados como palitos. (retirado de Nordiné et al, 2006)

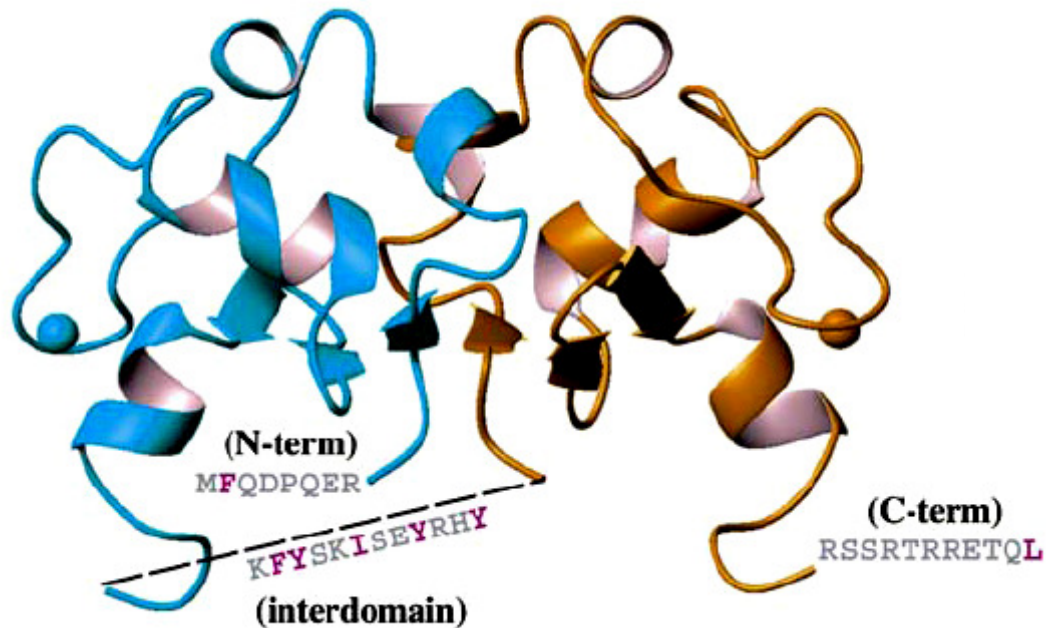


Figura 9: Representação do tipo *ribbons* do Modelo 3D de E6 6C/6S, indicando seqüências de resíduos que não foram inclusos. Esferas indicam posição dos íons de zinco. Em azul, domínio N-terminal; em amarelo, C-terminal. (retirado de: Nominé, 2006)

Esta estrutura experimental (2FK4) obtida por NMR é, no entanto, a melhor e mais completa realizada até hoje.

Em correspondência virtual (correio eletrônico) que trocamos com um dos autores a respeito da estrutura publicada 2FK4 e o modelo sugerido para a proteína inteira no artigo publicado em 2006 (Nominé et al, 2006), Dr. Gilles Travé (autor de correspondência do artigo) comenta que o modelo completo é hipotético e apenas a estrutura do domínio carboxi de E6C é experimental, como segue abaixo:

*“The representation of the full length structure that we showed in our Mol Cell 2006 [Nominé, 2006] paper was only a model of how the two subunits may be arranged in E6. Not even a classical structural model of the protein, just a proposition of how things may be arranged. Quite hypothetical, so that it has to be taken with caution [...]. This model is useful for thinking about the role of residues, to see which ones you can mutate, but it is not experimental, and it is too hypothetical to be released. For this reason my policy is not to distribute the model, but only the experimental data corresponding to the E6C domain. [...] If you would have liked to perform calculations, then as I told you before the current pseudodimeric “model” is not accurate enough for such purposes.” Dr. Gilles Travé*

Em função desta comunicação recebida, passamos a buscar a estrutura completa mais acurada de E6, que compreendesse também as porções negligenciadas pelos autores. Outra questão crítica evidenciada no trabalho de Nominé et al foi a simetria absoluta observada entre os dois domínios da proteína, que aparenta ser nada mais que a duplicação da estrutura experimental, tendo sido usada apenas para fins ilustrativos. Desta forma, na impossibilidade de realizar experimentos *in vitro* para obter o modelo da estrutura de E6 de HPV 16, por motivos já descritos, realizamos a predição do modelo da estrutura tridimensional por modelagem por homologia.

A modelagem por homologia, que é considerada a forma mais acurada de prever a estrutura tridimensional de uma proteína, consiste em criar um modelo de uma proteína alvo a partir das coordenadas molde de proteínas homólogas, tendo como pressuposto que a estrutura é conservada entre elas (Flores et al, 1993; Sali e Blundel, 1993). Este método costuma fornecer resultados bastante promissores, mas está atrelado a algumas restrições. Já é bem estabelecido que a modelagem comparativa pode produzir modelos inadequados se a identidade e a similaridade entre as sequências alvo e molde forem muito baixas. Também deve ser levado em consideração que a possibilidade de erro é tanto maior quanto maior for a disparidade entre as sequências. Esta baixa identidade e similaridade entre as sequências podem levar à inserção de numerosas lacunas e/ou lacunas muito extensas.

A base para um bom resultado na modelagem por homologia consiste em: 1) alinhamento adequado das sequências; 2) checagem da qualidade estereoquímica do modelo; 3) comparação das características da estrutura secundária dentro da estrutura tridimensional predita; e, 4) comparação entre as

características funcionais e estruturais já descritas com o modelo encontrado.

No desenvolvimento do nosso trabalho, observamos o cumprimento dos aspectos citados acima, no intuito de desenvolver um trabalho com o máximo de cuidado e a máxima acurácia que pudéssemos alcançar.

## 2 Objetivos e Justificativa

### Justificativa:

Este trabalho foi realizado com o intuito de contribuir para um maior conhecimento acerca da estrutura tridimensional da proteína E6 de HPV 16. O HPV 16 e o 18 são os tipos de HPV mais comumente envolvidos na malignização de lesões cervicais que levam ao câncer, sendo conjuntamente responsáveis por mais de 70% dos casos. O maior entendimento em relação às características estruturais e outras características físico-químicas pode favorecer o desenvolvimento e descobrimento de novos fármacos e vacinas para tratamento e prevenção da infecção por este vírus, responsável por tantas mortes em todo o mundo.

### Objetivo geral:

Agregar novas informações ao trabalho de Nominé et al (2006) e outros autores em relação à estrutura tridimensional de E6 de HPV 16.

### Objetivos específicos:

- 1 Sugerir a estrutura tridimensional da proteína completa de E6 de HPV 16 mutantes através de ferramentas de modelagem por homologia, utilizando como molde o domínio carboxi-terminal de E6 já conhecido – 2FK4.
- 2 Realizar análise da estrutura encontrada utilizando diversas ferramentas como Procheck, Topmatch, Prosa-Web e Verify3D, dando ênfase a sítios relevantes encontrados em outros trabalhos.
- 3 Realizar a dinâmica molecular em meio aquoso da estrutura modelada, de forma a verificar a qualidade da estrutura em condições fisiológicas aproximadas.

## 3 Materiais e Métodos

### 3.1 Material de origem:

A sequência da proteína selvagem de E6 de HPV 16 foi obtida no GeneBank sob o número de acesso nº NP\_041325.1. A proteína pode apresentar 158 ou 151 aminoácidos, como resultado de processamento alternativo. Para a obtenção dos modelos, foi utilizada a proteína de 151 aminoácidos. A estrutura E6C 4C/4S obtida a partir de NMR foi extraída do Protein Data Bank (PDB), número de acesso 2FK4.

A estrutura de origem 2FK4, de Nominé (2006), é composta de 10 modelos gerados por NMR. Para utilizá-la como molde, foi gerada uma estrutura média (resultado da média dos valores de coordenadas de todos os modelos originais), utilizando o software PROCHECK\_NMR (Laskowski et al, 1993; Morris et al, 1992). A partir deste modelo médio, escolhemos o modelo individual com base no resultado de cada um em relação a este modelo médio, utilizando as ferramentas Verify3D (Bowie et al, 1991; Lüthy et al, 1992), PROCHECK, PROSA-WEB (Sippl, 1993; Wiederstein e Sippl, 2007) e TOPMATCH (Sippl, 1993; Wiederstein e Sippl, 2007), como pode ser visto na Tabela 1.

**Tabela 1:** Resultado da análise dos modelos individuais em 2FK4 para a escolha do template para o processo de modelagem. LI: Limite inferior, LS: Limite superior, C $\alpha$ -RMS: *Root Mean Square dos carbonos alfa*; NP: não permitido, GP: generosamente permitido, AP: adicionalmente permitido. Em cinza: modelo 3 (2FK4\_03) escolhido como molde para as modelagens.

Modelo	Verify3D		Topmatch	Prosa	Ramachandran (% permitido)	Procheck			G-factor overall (Fator G)
	LI	LS	C $\alpha$ -RMS	Z-score		NP	GP	AP	
1	0,14	0,49	1,1	-4,01	94,9	0	1	2	0,09
2	-0,06	0,46	1,2	-4,06	84,7	1	0	8	0,1
3	0,02	0,5	0,7	-4,34	96,6	0	1	1	0,06
4	0,13	0,48	1,1	-3,49	88,1	1	0	6	0,08
5	-0,07	0,31	1,1	-3,12	91,5	3	0	2	0,11
6	0,09	0,45	0,8	-3,34	89,8	1	1	4	0,06
7	0,04	0,37	1	-3,4	86,4	2	0	6	0,09
8	0,04	0,36	1	-3,7	88,1	1	0	6	0,05
9	-0,22	0,28	1,2	-4,1	93,2	0	0	4	0,09
10	0,04	0,41	1,3	-3,97	88,1	1	1	5	0,08

### 3.2 Modelagem Comparativa:

Todas as modelagens por homologia deste trabalho foram realizadas utilizando a ferramenta MODELLER 9 versões 4 e 5 (Eswar et al, 2000; Martí-Renom et al, 2000). O refinamento das estruturas foi realizado manualmente com Swiss PDB Viewer (SPDBv – Guex e Peitsch, 1997) e posteriormente com a simulação em meio aquoso através da ferramenta LEAP contida em AMBER 9 (Case et al, 2008 – ver tópico específico). Os melhores modelos foram escolhidos entre os resultados apresentados pelas ferramentas PROCHECK, Verify3D, PROSA-WEB com base nos seguintes parâmetros: diagrama de Ramachandram e G-factor (PROCHECK), Z-score e distribuição (PROSA), Desvio quadrático médio – RMSD – em relação ao molde, distâncias e ângulos alcançados entre o Zinco e as Cisteínas - 2,3 Å para a distância Cys-Zn e de 110° para o ângulo Cys-ZN-Cys - (SPDBv) e o escore do Verify3D.

#### 3.2.1 E6C selvagem:

Foram gerados 20 modelos da extremidade E6C selvagem (wt) a partir de E6C 4C/4S (PDB ID 2FK4), modelo 3. O alinhamento utilizado para esta

modelagem é mostrado na Figura 10 e foi criado com a matriz Gonnet de alinhamento.

```

                10      20      30      40      50      60
                |      |      |      |      |      |
E6C4C4S    AMSYSLYGTTLLEQQYNKPLSDLLIRCINCQKPLSPEEKQRHLDKKQRFHNIIRGRWTGRCM3CSRSSRX
E6Cwt      --CYSLYGTTLLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDKKQRFHNIIRGRWTGRCM5CCRSSRX
           .*****.*****.*****.*****.*****.*****
Prim.cons. AM2YSLYGTTLLEQQYNKPL2DLLIRCINCQKPL2PEEKQRHLDKKQRFHNIIRGRWTGRCM5C2R5SRX

```

**Figura 10:** Alinhamento utilizado para a geração do modelo da extremidade C terminal de E6. O alinhamento compreende os resíduos de 80 a 144. X corresponde ao íon de zinco.

### 3.2.2 E6N:

A partir do mesmo modelo mutado E6C 4C/4S, foram gerados 30 modelos da extremidade N-terminal (chamados 4E6N), que como discutido anteriormente, apresenta alto grau de simetria com a extremidade E6C. O alinhamento utilizado como dado de entrada para esta modelagem pode ser visto na Figura 11. Foi utilizada a matriz Gonnet. Este alinhamento apresenta 15,4% de identidade, 18,5% de similaridade forte, e 15,4% de similaridade fraca. Nota-se que as cisteínas que se ligam ao Zinco ficam alinhadas, bem como os demais resíduos considerados chave trabalho de Nominé et al (2006).

```

                10      20      30      40      50      60
                |      |      |      |      |      |
E6N          PRKLP--QLCTELQTTIHDIILECVYCKQQLLRREVDYDFAFRDLCIVYRDGNPYAVCDKCLKFYXS
E6C          CYSLYGTTLLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDKKQRFHNIIRGRWTGRCM5CCRSSRX
           . *  :  :...: *:::.*: *:: *  .*  .  :.  :  *  .  *  .*  :  *
Prim.cons.   222L2GT2L222222222D2222C22C222L222E2222222222222222G22222C22C22222X

```

**Figura 11:** Alinhamento que deu origem à estrutura do domínio N terminal de E6. O alinhamento compreende os resíduos de 9 a 71 da porção N-terminal e 80 a 144 da C terminal. X corresponde ao íon de zinco.



### 3.2.3 Modelagem e montagem da estrutura quimera E6N/E6C:

Num primeiro momento as extremidades E6N e E6C obtidas das modelagens anteriores foram orientadas entre si, partindo das figuras disponíveis em Nominé et al (2006) como base e utilizando a ferramenta SPDBv. A posição relativa das duas extremidades levou em conta o ângulo de inclinação em relação ao eixo das fitas S1 de ambas as extremidades, assim como a distância entre as mesmas de forma que fosse possível a interação entre as fitas criando uma folha beta anti-paralela. A distância entre as fitas deve ser menor que 3,3 Ångstrons. A confirmação da posição se deu pela observação de ausência de incompatibilidades ou sobreposições das duas estruturas e por comparação visual com a estrutura hipotética obtida por Nominé.

Neste molde intermediário ainda estão ausentes as extremidades N- e C-terminal, com 8 e 7 resíduos, respectivamente, e a região de conexão entre os dois domínios, que contém 8 resíduos. Uma vez identificada a posição correta, foi criado um arquivo único contendo os dois domínios identificados apenas como camadas diferentes. A numeração dos resíduos foi ajustada para igualar à numeração dos resíduos na sequência completa original de E6 de HPV 16.

A quimera foi modelada utilizando as duas metades posicionadas corretamente, que chamamos de modelo intermediário, como molde, e a sequência completa (incluindo as três porções excluídas do modelo original). Estas porções foram, posteriormente, refinadas individualmente através da ferramenta de refinamento de *Loops* (Alças) do Modeller (Fiser et al, 2000). Os refinamentos compreenderam mais alguns resíduos adjacentes, não se limitando àqueles faltantes, para facilitar a adaptação da estrutura inteira, uma vez que o modelo pode ter sofrido certa distorção em função da estrutura terciária resultante da modelagem

inicial não otimizada. A obtenção deste modelo quimérico intermediário se deu com base no alinhamento da Figura 12 e foi chamado de 5CHI.

```

          10      20      30      40      50      60      70
          |      |      |      |      |      |      |
interm  -----PRKLPQLCTELQTTIHDIIILECVYCKQQLRREVYDFAFRDLCIVYRDGNPYAVCDKCLKFY-----
chim    MFQDPQERPRKLPQLCTELQTTIHDIIILECVYCKQQLRREVYDFAFRDLCIVYRDGNPYAVCDKCLKFYISKISEYRHY
          *****
          MFQDPQERPRKLPQLCTELQTTIHDIIILECVYCKQQLRREVYDFAFRDLCIVYRDGNPYAVCDKCLKFYISKISEYRHY

          89      99      109     119     129     139     149
          |      |      |      |      |      |      |
interm  CYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDDKKQRFHNIIRGRWTGRCMSCCRSSR-----XX
chim    CYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDDKKQRFHNIIRGRWTGRCMSCCRSSRTRRETQLXX
          *****
          CYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDDKKQRFHNIIRGRWTGRCMSCCRSSRTRRETQLXX

```

**Figura 12:** Alinhamento entre a sequência estrutura molde (intermediária - interm) e a sequência completa de E6 (chim).

Cada refinamento de Alça (as porções foram denominadas Alça 1, 2 e 3, respectivamente – vide Tabela 2 -, e continham as sequências ausentes no alinhamento da Figura 12, adicionados dos resíduos que as flanqueavam) foi realizado no modo *slow\_refinement* e gerou 20 modelos. O *score* de cada modelagem foi verificado nas ferramentas PROSA, Verify3D e Procheck. O melhor modelo em cada passo foi, então, levado para o refinamento da alça seguinte.

**Tabela 2:** Alças 1 a 3 das estrutura completa, refinados individualmente.

Alça	Resíduos	Sequência
1	1-10	MFQDPERPR
2	67-81	LKFYSKIEYRHYCY
3	140-151	CRSSRTRRETQL

O ajuste da posição dos resíduos Cys de ligação ao zinco foi realizado manualmente através do programa SPDBv. Os valores considerados para este ajuste foram de 2,3 Å para a distância Cys-Zn e de 110° para o ângulo Cys-ZN-Cys, como citado anteriormente.

### 3.3 Dinâmica molecular: simulação em meio aquoso

Para testar o comportamento e estabilidade dos modelos escolhidos, foi realizada a dinâmica molecular onde foram simuladas condições em meio aquoso utilizando o módulo SANDER (*Simulated Annealing with NMR-Derived Energy Restraints*) do AMBER 9 (Case et al, 2008). Foram testados 3 protocolos e algumas variações de parâmetros de campos de força a partir dos padrões disponibilizados por Weiner et al (1984; 1986) ou Cornell et al (1995), dois com plasma de carga residual e outro com íons de Cl<sup>-</sup> para neutralizar a carga positiva. Outros parâmetros testados foram o raio do ponto de equilíbrio da interação por forças de Van de Waals (de 1,1 a 2,2 Å) e a energia  $\epsilon$  (Épsilon – de 0,0125 a 0,025) resultante desta interação. Os dados de entrada para o AMBER foram construídos com o módulo LEaP do AMBER (Schafmeister, 1995). Para tratar os íons de zinco foi utilizado o modelo *non-bonded*, com uma carga formal de +2. Os parâmetros Lennard-Jones do Zn<sup>2+</sup> foram adaptados de Stote e Karplus (1995).

Os dados gerais de cada sistema podem ser vistos na Tabela 3. Para cada um dos sistemas de simulação os primeiros 500 ps foram de minimização e equilíbrio da molécula e não estão contabilizados no tempo final da simulação. Neste ponto são adicionadas as moléculas de água tipo TIP3P (Jorgensen et al, 1983) e delimitadas as dimensões da caixa ortorrômbica. Durante o processo de equilíbrio a densidade molecular no sistema passa de aproximadamente 0,75 para em torno de 1,01 g/cm<sup>3</sup>.

Os testes foram realizados em pacotes de 500 ps, sendo os primeiros 200 ps de simulação propriamente dita, reservada para o aquecimento e estabilização do sistema. A simulação foi realizada em um *ensemble* NTP, a um temperatura de 298,16 K (25°C) (aumentando 50 K a cada 20 ps), sob pressão de 1

atm, pH 6,8, com tempo de integração de 2 fs (femtosegundo). A escolha do protocolo foi realizada testando os procedimentos com o modelo experimental inicial 2FK4\_03. O protocolo escolhido foi utilizado posteriormente no modelo completo desenvolvido neste trabalho e utilizou como parâmetros: raio do ponto de equilíbrio da interação por forças de Van de Waals = 1,5 Å e a energia  $\epsilon = 0,025$ .

**Tabela 3:** Dados dos sistemas das utilizados nas dinâmicas moleculares de simulação em meio aquoso.

	Nome do modelo		
	2FK4_03	6E6C_15	5CHI_03
Nome da simulação	SIM001	SIM002	SIMNC001
Número de átomos (-H <sub>2</sub> O)	1016	1086	2550
Número de moléculas H <sub>2</sub> O	4611	4619	7635
Dimensões da caixa (Å <sup>3</sup> )	63,4 x 55,5 x 56,0	60,9 x 54,2 x 59,0	70,8 x 67,0 x 68,6
Forma de neutralização da carga residual	plasma	plasma	6 Cl <sup>-</sup>
Duração da simulação	5,0 ns	10,5 ns	9,242 ns

As análises realizadas após a obtenção da trajetória da proteína foram realizadas utilizando AMBER, VMD (Humphrey et al, 1996) e Procheck. Para a análise com Procheck foram retiradas da simulação 11 tempos (0, 1001, 2001, 3001, 4001, 5001, 6001, 7001, 8001, 9001 e 9242 ps) e tais estruturas foram minimizadas para diminuir incompatibilidades estéricas e distorções da estrutura.

Todas as figuras das estruturas modeladas apresentadas nesta dissertação foram criadas com auxílio das ferramentas SPDBv, PyMOL ou VMD.

### 3.4 Outras análises da estrutura:

A análise de similaridade de estrutura/motivos foi realizada com a ferramenta PROSCAN (PROSITE SCAN - Combet et al 2000; Hulo et al, 2008), utilizando *cut-offs* de 100%, 70% e 50%.

Foi realizada também a análise de predição de pontos de interação

proteína-proteína (Protein-Protein Interface prediction) através da ferramenta PPI-Pred (Bredford e Westhead, 2005) utilizando como dado de entrada a sequência da proteína E6 selvagem.

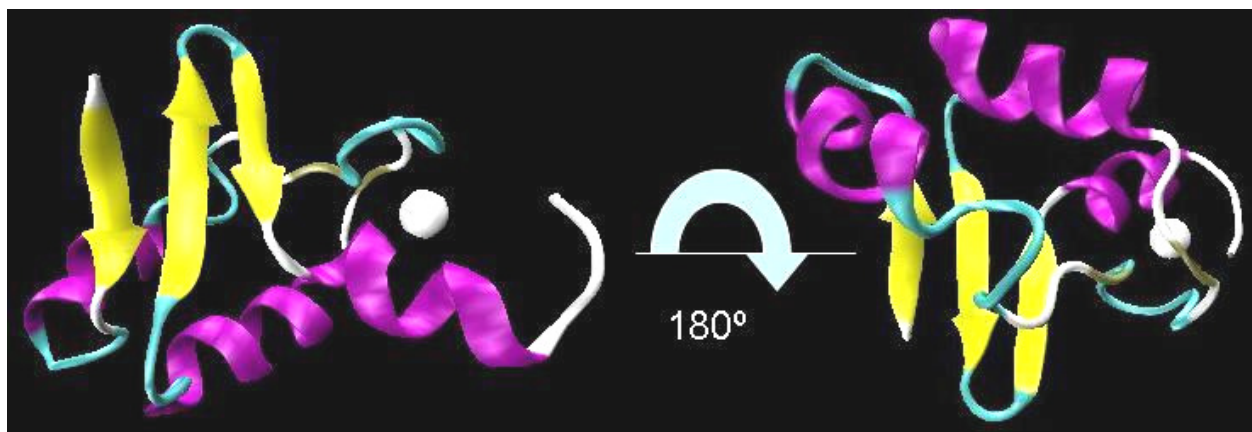
## 4 Resultados

### 4.1 A modelagem E6 selvagem:

#### 4.1.1 E6Cwt

Optou-se por gerar 20 modelos para a estrutura do domínio E6C selvagem de HPV16. O modelo escolhido foi o número 15 (E6C6\_15) com base em vários parâmetros. A representação do modelo pode ser vista na Figura 13. A estrutura modelada manteve a estrutura terciária de seu molde. Apenas algumas poucas alterações podem ser notadas. Por exemplo, a hélice H1 no molde (2FK4\_03) possui 2 resíduos a mais que o modelo obtido (Thr86 e Thr87). A análise com TOPMATCH indicou que em ambas as estruturas, a alça que contém 2 dos resíduos de cisteína (entre H1 e H2) pode ser reconhecida como uma folha beta composta de duas fitas beta (em ambos molde e modelo). Na realidade, esta característica se repete para os outros modelos obtidos ao longo do trabalho e depende do software utilizado para visualizar os modelos das estruturas (dado não mostrado – esta característica será melhor descrita mais adiante). Como esperado, a cauda C-terminal apresentou grande variação em relação ao molde, apesar da pequena porção que se apresenta “livre” de estruturas secundárias.

Na Tabela 4, podemos verificar a qualidade do modelo obtido da porção E6C com relação a diversas ferramentas. Como pode ser observado, o *G-factor*, o principal parâmetro de conformidade da estrutura, teve seu valor final (geral) inalterado, permanecendo em 0,06. O número de contatos ruins diminuiu para 7, enquanto no molde eram 18. Os demais parâmetros não apresentaram resultados tão bons quanto os do modelo, mas ainda estão dentro da faixa considerada aceitável.

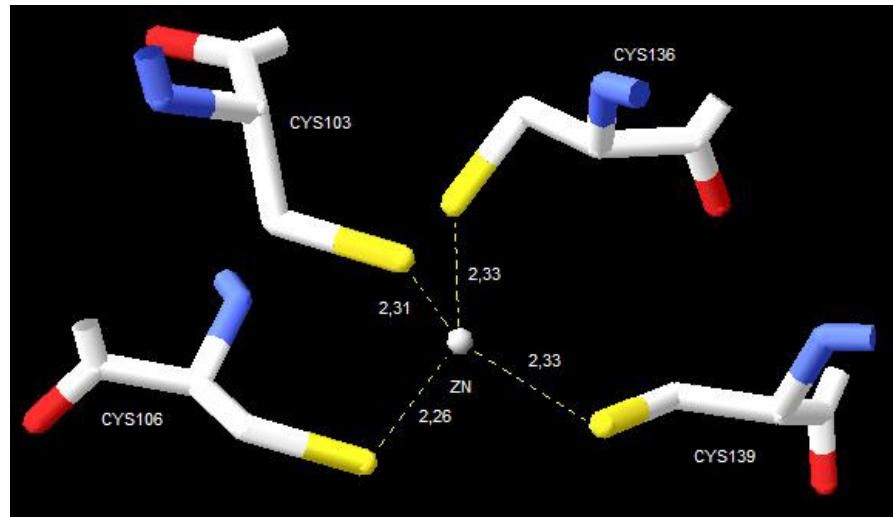


**Figura 13:** Estrutura da porção C-terminal E6 de HPV 16 modelada por homologia a partir do molde de 2FK4, modelo 3. Esquema de cores das unidades de estrutura secundária: fitas de folha  $\beta$  amarelo), volta (branco), hélices  $\alpha$  (violeta), alças (ciano) e voltas- $\beta$  (ocre). A estrutura da direita representa a rotação de 180° graus sobre o eixo X da estrutura da esquerda.

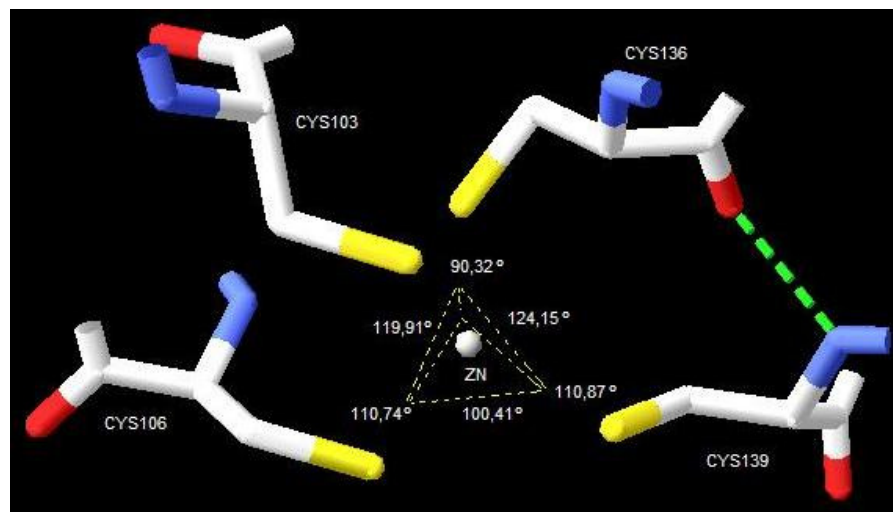
**Tabela 4:** Tabela comparativa dos resultados para o modelo obtido de E6 C-terminal selvagem (6E6C\_15) e seu molde E6C 4S/4C (2FK4\_3). Onde, LI – Limite Inferior; LS – Limite Superior; %P – porcentagem de resíduos em região permitida; NP – número de resíduos em regiões não permitidas; GP – número de resíduos em regiões generosamente permitidas; AP – Número de resíduos em regiões adicionalmente permitidas.

Modelo	Verify3D		PROSA	Ramachandran				Procheck			Nº de contatos ruins
	LI	LS	Z-score	% P	NP	GP	AP	<i>G-factor</i>		Nº de contatos ruins	
								Ângulos diédricos	Ligações cadeia principal		
6E6C_15	0,01	0,4	-3,79	94,7	0	1	2	0,19	-0,17	0,06	7
2FK4_03	0,02	0,5	-4,34	96,6	0	1	1	0,17	-0,09	0,06	18

Em relação aos resíduos de cisteína que realizam a coordenação do íon  $Zn^{2+}$ , obteve-se valores razoáveis de distância entre o átomo de enxofre e o zinco e de ângulo, para formar o ângulo tetraédrico (em forma de pirâmide) esperado para tal estrutura. As distâncias de Cys-Zn ficaram dentro do esperado (em torno de 2,3 Å) – Figura 14 - e os ângulos Cys-Zn-Cys tiveram alguma variação, mantendo ainda a média em 109,4° (Figura 15).



**Figura 14:** Distâncias Cys-Zn de 6E6C\_15. Representação em modelo palito mostrando as distâncias entre os átomos de enxofre dos resíduos de cisteína e o íon de zinco na estrutura modelada por homologia da porção C-terminal de E6 selvagem, que utilizou como molde o modelo 3 de 2FK4.



**Figura 15:** Ângulos Cys-Zn-Cys de 6E6C\_15. Representação em modelo palito mostrando os ângulos CYS-ZN-CYS na estrutura modelada por homologia da porção C-terminal de E6 selvagem, que utilizou como molde o modelo 3 de 2FK4.

#### 4.1.2 E6Nwt

A estrutura de E6N foi escolhida pelo resultado geral nos vários parâmetros. O modelo escolhido foi o número 10 (chamaremos de 4E6N\_10). Os resultados podem ser vistos na Tabela 5.

O modelo da extremidade N-terminal obtido neste passo da modelagem por homologia pode ser visto na Figura 16. A configuração utilizada no

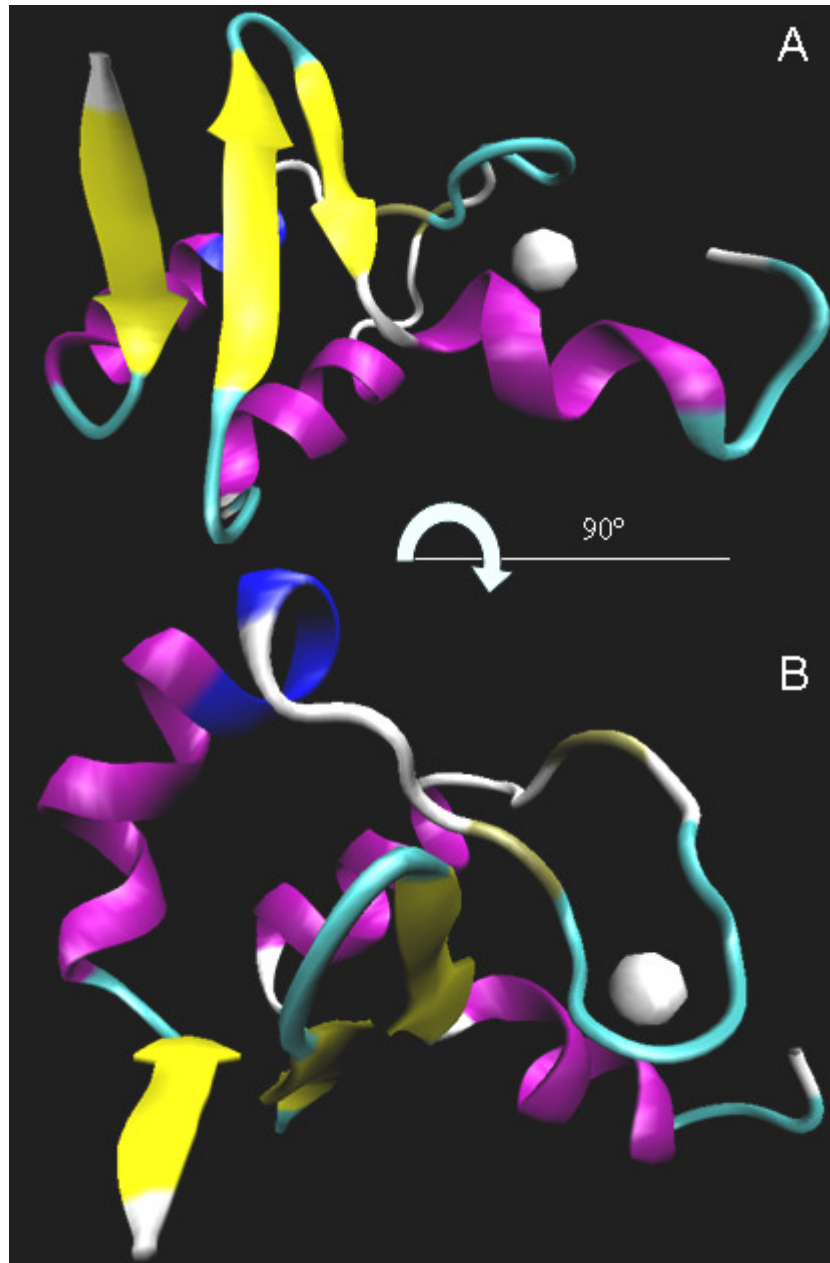


programa VMD de visualização demonstra uma coloração diferenciada em dois pontos do modelo (melhor observado nas Figuras 17 B ou 16). Na grande alça que contém as cisteínas envolvidas na ligação com o zinco (resíduos 26 a 39) há duas porções coloridas em tom ócreo, uma de cada lado das cisteínas, que indica uma tendência à formação de estrutura secundária. No Procheck, esta estrutura aparece contendo resíduos de volta que realizam ligações de folha beta (*residue isolated beta bridge / hydrogen-bonded turn*). Na ponta N-terminal desta alça, onde se encontra a hélice H1, os resíduos Thr22 a Asp25 aparecem em cor azul, indicando uma hélice estendida, igualmente identificada pelo Procheck.

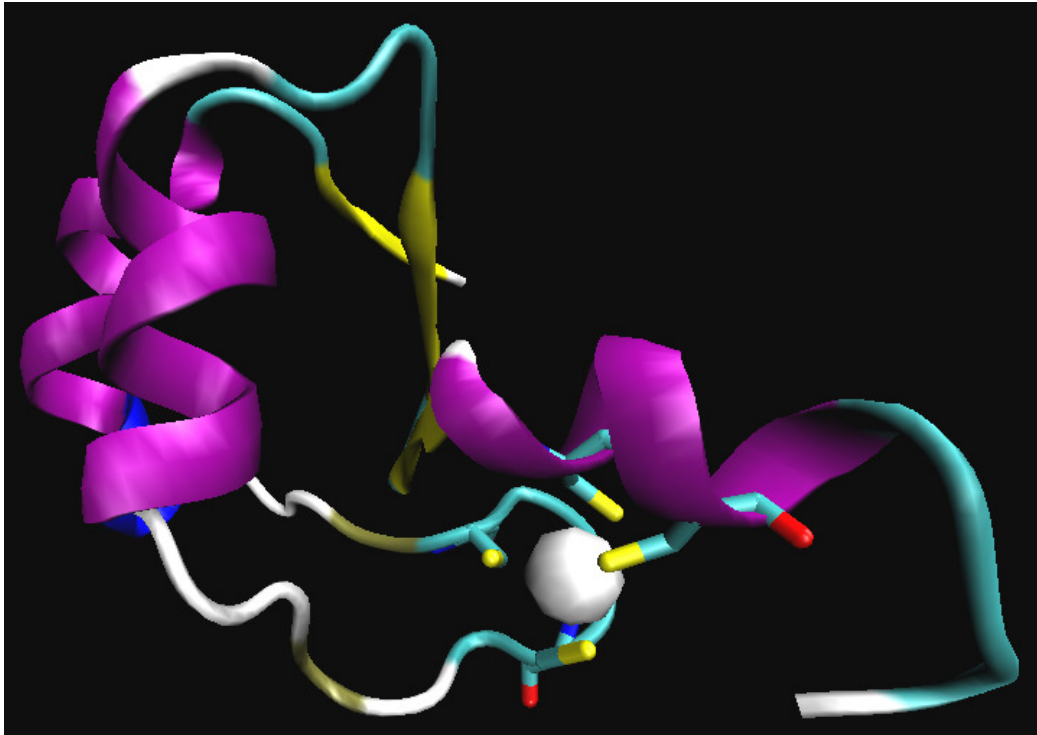
As distâncias do zinco em relação ao átomo S dos resíduos Cys ficaram entre 2,29 e 2,31 Å, com apenas o resíduo de cisteína Cys63 com 2,46 Å. Os ângulos Cys-ZN-Cys obtiveram uma média de  $109,2^\circ \pm 8,5$ .

**Tabela 5:** Resultados das análises dos modelos da estrutura de E6N. A primeira linha indica As ferramentas utilizadas e a segunda os principais parâmetros analisados. LI – Limite inferior; LS – Limite superior; NP – Não permitido; GP – Generosamente permitido; AP – Adicionalmente permitido. Em cinza: Modelo 4E6N\_10, escolhido para fazer parte do molde da proteína completa.

4E6N	Verify 3D		Prosa	Topmatch	Procheck				
Modelo	LI	SL	Z-score	RMS do molde (C-alfa)	Ramachandran (%) permitido	NP	GP	AP	G-factor (geral)
1	0,03	0,36	-3,74	0,8	91,4	0	1	4	-0,12
2	-0,03	0,38	-3,52	0,7	91,4	0	1	4	-0,19
3	0,04	0,34	-3,74	0,6	91,4	1	0	4	-0,21
4	0,06	0,36	-3,9	0,8	93,1	0	1	3	-0,09
5	0,03	0,46	-3,69	0,8	91,4	1	1	3	-0,18
6	-0,09	0,33	-3,9	0,8	93,1	1	0	3	-0,04
7	-0,06	0,39	-4,06	0,8	91,4	1	0	4	-0,12
8	-0,10	0,37	-3,87	0,6	94,8	0	1	2	-0,11
9	-0,01	0,40	-3,84	0,7	91,4	0	1	4	-0,13
10	0,02	0,44	-3,84	0,6	93,1	0	1	3	-0,17
11	0,08	0,37	-3,94	0,8	89,7	1	1	4	-0,17
12	0,11	0,39	-3,7	0,8	87,9	1	1	5	-0,25
13	-0,01	0,39	-3,81	0,7	91,4	1	0	4	-0,12
14	-0,03	0,34	-3,88	0,8	91,4	1	0	4	-0,17
15	-0,04	0,31	-3,6	0,8	89,7	0	1	5	-0,15
16	-0,03	0,32	-3,76	0,8	91,4	1	0	4	-0,18
17	0,00	0,26	-3,5	0,9	86,2	0	3	5	-0,15
18	-0,04	0,40	-3,86	0,6	91,4	1	1	3	-0,14
19	-0,02	0,35	-3,78	0,9	91,4	1	0	4	-0,17
20	0,03	0,35	-3,63	0,8	93,1	1	0	3	-0,13
21	-0,07	0,41	-3,91	0,9	91,4	1	1	3	-0,17
22	-0,05	0,43	-3,64	0,9	93,1	0	1	3	-0,09
23	-0,03	0,41	-3,78	1,0	93,1	0	1	3	-0,20
24	-0,09	0,31	-3,98	0,9	93,1	0	1	3	-0,16
25	-0,05	0,37	-3,75	0,8	91,4	0	1	4	-0,16
26	0,03	0,46	-3,56	0,9	89,7	0	1	5	-0,13
27	-0,05	0,38	-3,88	0,9	87,9	1	3	3	-0,24
28	-0,12	0,41	-3,59	0,7	87,9	0	1	6	-0,15
29	-0,05	0,38	-3,55	0,9	91,4	1	0	4	-0,10
30	0,06	0,46	-3,73	0,8	94,8	1	0	2	-0,08



**Figura 16:** Modelo da estrutura E6N (4E6N\_10) elaborado por modelagem por homologia. A Figura A e B são representações do mesmo modelo com 90° de rotação em relação ao eixo X. Em B, pode-se ver com mais detalhe as estruturas com coloração diferenciada. A esfera branca representa o íon Zinco.



**Figura 17:** Representação do modelo 3D do tipo *ribbons* de 4E6N\_10 de HPV 16. Detalhe da região de ligação ao zinco em representação tipo palitos. As estruturas em palito de cor amarela representam os átomos de enxofre dos resíduos de cisteína envolvidos na coordenação do zinco. A esfera branca representa o íon de zinco. O esquema de cores da representação tipo *ribbons* está explicitada na figura 13, em azul, hélice estendida.

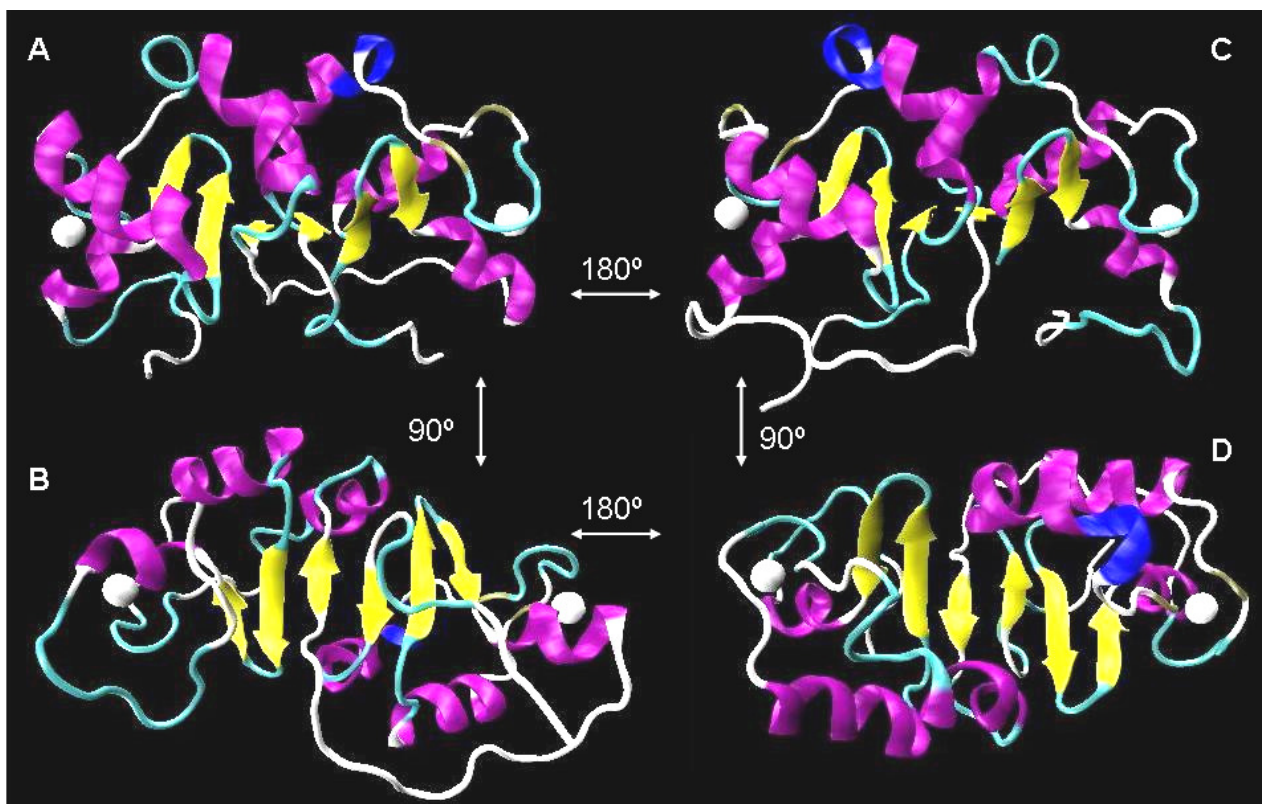
#### 4.1.3 E6wt

A estrutura final de E6 obtida neste trabalho é apresentada nas Figuras 18 A a D e 19. Os dois domínios de ligação ao zinco se localizam diametralmente opostos, como previamente sugerido por outros autores (Nominé et al 2005, 2006). Os resíduos de cisteínas responsáveis pela ligação estão em distâncias menores que 2,5 Å (sendo apenas 2 dos 8 resíduos, Cys66 e Cys63, maiores que 2,33 Å).

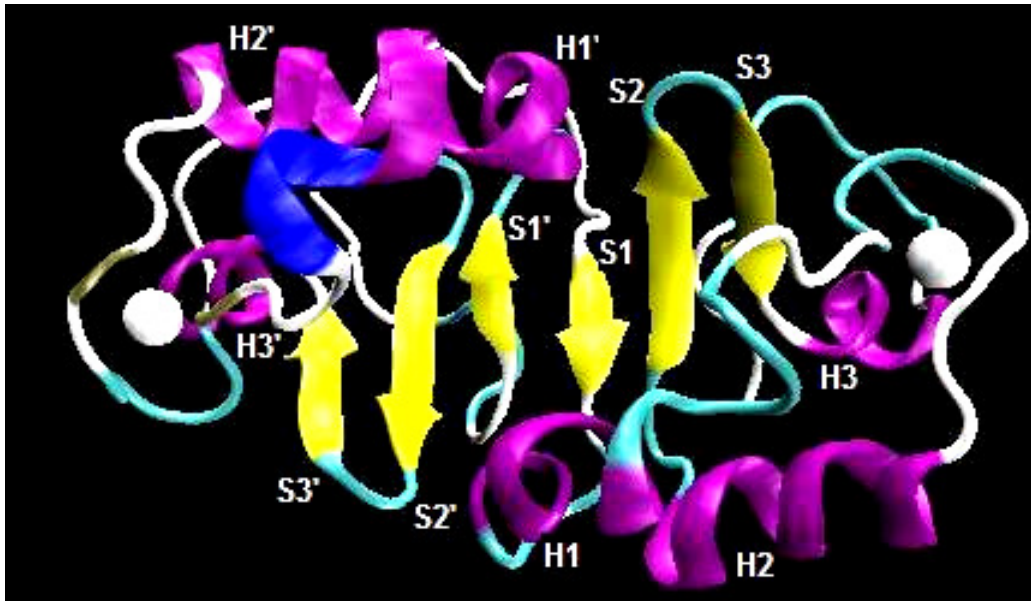
A extremidade N-terminal é fortemente hidrofílica, com exceção dos resíduos Met1 e Phe2, que são hidrofóbicos. Há, possivelmente, grande flexibilidade nesta região de 8 aminoácidos, anteriormente excluída. Os dois resíduos hidrofóbicos Met1 e Phe2 podem interagir de alguma maneira com a alça Lys68-Cys80, de forma a satisfazer o caráter apolar.

Os dois domínios são conectados por uma alça de 13 resíduos que passa transversalmente ao eixo das fitas beta, na mesma linha de orientação que as fitas S2' e S3'. A alça parte da hélice H3 e chega à fita S1, que é central na folha beta juntamente com a S1'. A alça é composta em grande parte de resíduos hidrofílicos e neutros, o que provavelmente permite mais flexibilidade nesta região. Boa parte das cadeias laterais dos resíduos da alça está voltada em direção à proteína.

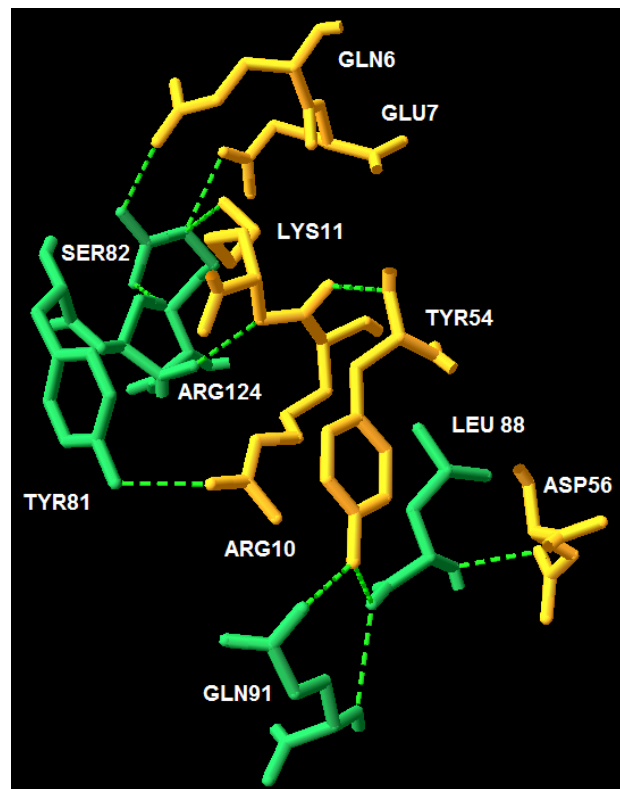
A interação através de pontes de hidrogênio entre os domínios permite a manutenção parcial da posição relativa deles. A listagem dos pares de resíduos que interagem para esta estabilização pode ser vista na legenda da Figura 20.



**Figura 18:** Representação do tipo *ribbons* da estrutura 3D da proteína E6 do HPV 16 obtida por modelagem por homologia (ver seção Materiais e Métodos). Os íons de zinco (cinza) estão representados por suas esferas de van Der Waals. Figura produzida pelo VMD. As figuras B, C e D representam visualizações da figura A sob diferentes ângulos.



**Figura 19:** Detalhe de um dos ângulos do modelo final da estrutura de E6 de HPV16. Modelo representado *ribbons*, coloração explicada na figura 13. Esfera branca representa os íons de zinco. S1' a S3': fitas do domínio N-term; H1' a H3': hélices do domínio N-term; S1 a S3: fitas do domínio C-term; e, H1 a H3: hélices do domínio C-term.



**Figura 20:** Representação do tipo palitos da face de interação entre os dois domínios (N- e C-terminal) de E6 de HPV 16. Os resíduos em amarelo pertencem ao domínio N-terminal e em verde ao C-terminal. As linhas pontilhadas representam a interação por pontes de hidrogênio entre os átomos.

A extremidade C-terminal que foi adicionada nesta modelagem é

composta de 7 aminoácidos – TRRETQL (posições 145 a 151). Corroborando os resultados de Nominé, esta porção fica aparentemente 'solta' e totalmente exposta. De acordo com resultados de Zhang et al (2007) e Charbonier et al (2008) esta porção deve ficar exposta pois ganha conformação ao entrar em contato com o alvo celular (domínios PDZ) formando uma fita beta.

Além das porções que haviam sido excluídas anteriormente neste trabalho e no trabalho de Nominé (Alças 1, 2 e 3), a proteína ainda apresenta outras duas alças extensas (de 14 e 13 aminoácidos, respectivamente resíduos Ile26-Arg39 e Leu99-Cys111). No centro de cada uma destas alças se encontram dois dos resíduos de cisteína que coordenam o íon de zinco. Novamente, como ocorreu na modelagem da porção N-terminal, conforme o software utilizado para produzir as imagens do modelo da estrutura, a coloração indica a presença de (ou tendência à) uma estruturação diferente de alça. Aparentemente a tendência é a formação a uma folha beta anti-paralela resultante da interação de cada metade da alça – a coloração ócrea diferenciada pode ser novamente observada nas Figuras 18 e 19.

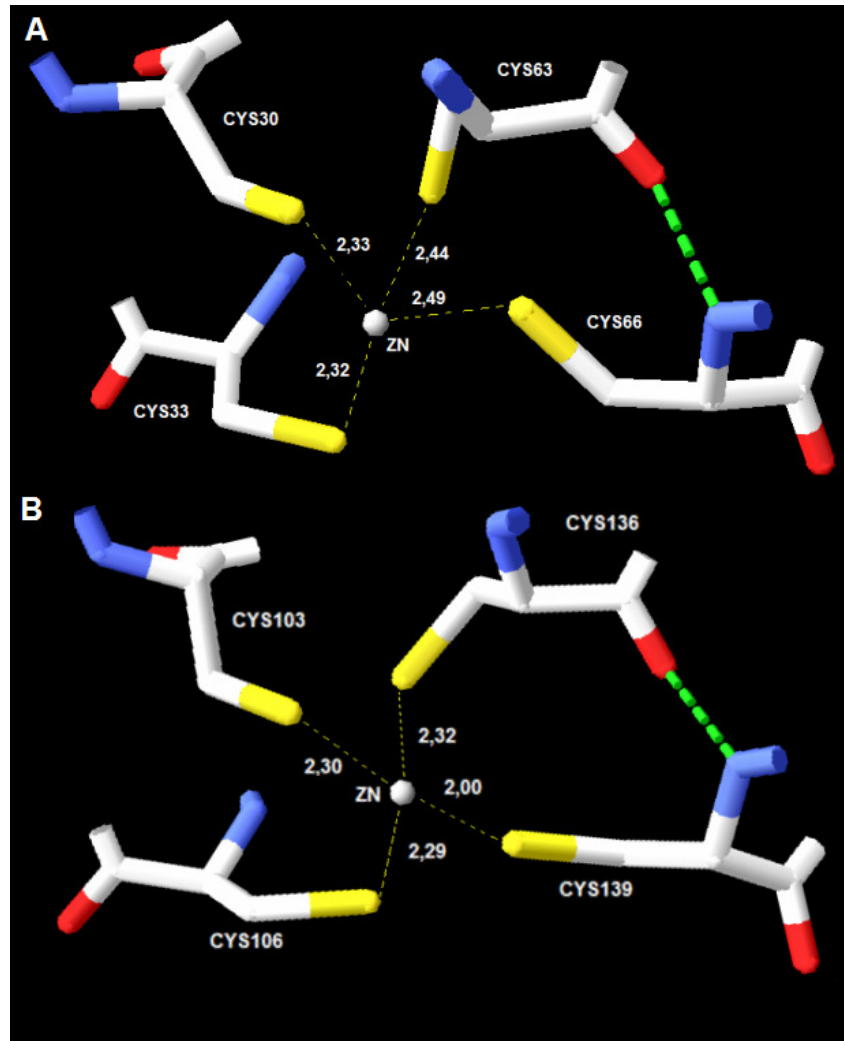
A proteína contém uma folha beta composta de 6 fitas anti-paralelas, sendo 3 fitas pertencentes a cada domínio ZB. Esta estrutura apresenta uma leve curvatura das extremidades opostas (onde estão posicionados os resíduos Gly57 na porção N-terminal e Arg129 na porção C-terminal) em direção à face da proteína onde estão as hélices. A proteína apresenta também 6 hélices (H1' a H3' – N-terminal - e H1 a H3 – C-terminal) todas localizadas na mesma face. A parte final da hélice H1 é mais estendida que as hélices alfa convencionais, sendo aparentemente uma hélice PI.

As hélices alfa H1', H2', H1 e H2 são praticamente paralelas no eixo ao longo da proteína, mas perpendiculares se vistas de lado em relação ao plano da

folha beta. A hélice H3 se distancia a partir da metade de cada domínio em direção à ponta da proteína, onde se aproxima da alça que conecta as hélices H1 e H2. O mesmo acontece com a hélice H3' na outra extremidade da estrutura. Nestas pequenas hélices, que compreendem os resíduos Cys63-Leu67 (N-terminal) e Cys136-Cys139 (C-terminal), se encontram dois dos resíduos de cisteína responsáveis pela coordenação do zinco em cada domínio.

As porções inicialmente removidas da estrutura (alças L1 a L3) foram adicionadas no momento de produzir o modelo final da estrutura. As alças L1 e L3, nos diversos modelos resultantes do *script* do Modeller apresentaram bastante diversidade de estruturas possíveis, mais ou menos adequadas considerando seu score nas ferramentas de análise, mas ainda assim passíveis de adquirir conformação adequada com poucos ajustes. Por outro lado, a alça L2 produziu apenas uma forma razoável de disposição dos seus resíduos, que consiste em ela passar por fora da estrutura ao lado e ao longo da folha beta.





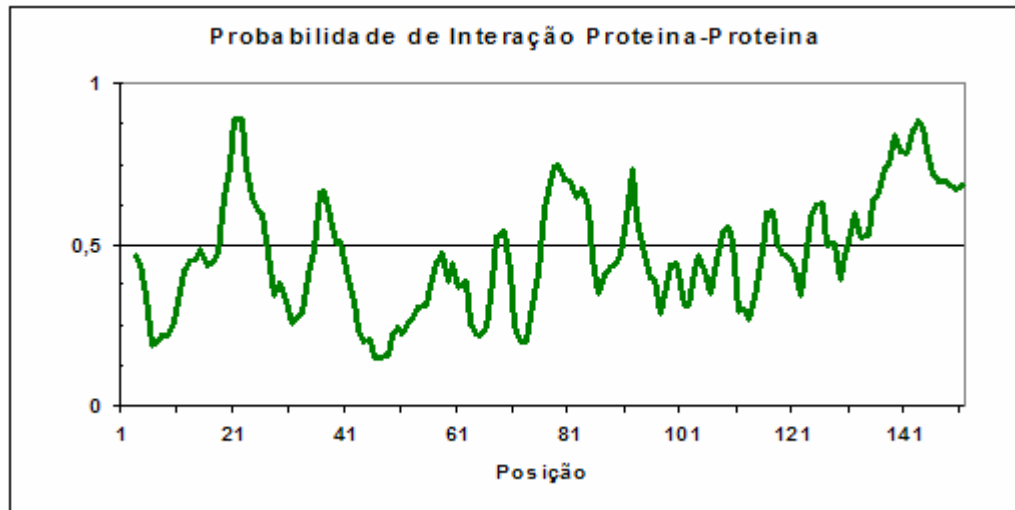
**Figura 21:** Representação do tipo palito com coloração CPK das distâncias Cys-ZN nos sítios de coordenação do zinco no modelo da estrutura completa de E6 de HPV 16 obtida por modelagem por homologia. A: domínio N-terminal; B: domínio C-terminal. As esferas brancas no centro de cada figura representam os íons de zinco.

## 4.2 Análise da estrutura:

A análise de similaridade de motivos com PROSCAN, PBIL, indicou uma certa semelhança de assinatura com o motivo dedos de zinco PHD (plant homologue domain) entre 70 e 75%. Foi realizado o alinhamento com o motivo PHD de algumas proteínas que o continham, tais como AIRE (nº acesso Genbank: CAA08759.1) e ASH1 (nº acesso Genbank: CAI13211.2) (ver arquivo de saída no Apêndice II), com as porções equivalentes de E6. A similaridade de sequência entre elas ficou em torno de 10% (alinhamento e dados não mostrados).

Uma série de outras assinaturas foram retornadas por esta ferramenta, sendo no total 16 padrões/assinaturas diferentes encontrados para o corte de 70%. Muitos deles são padrões e sítios de fosforilação de diferentes enzimas ou de miristoilação, com alto grau de semelhança, o que pode significar novas funções e atividades ainda não conhecidas de E6. O resumo destes resultados ser visto na Tabela 6.

A análise utilizando a ferramenta PPI-Pred demonstrou vários pontos da proteína com propensão a realizar interações, como pode ser visto na Figura 22. Os resultados são dados por probabilidade de 0 a 1 (0% a 100%) com base nas características da sequência de aminoácidos da proteína. As regiões com maior probabilidade de interação estão localizadas entre os resíduos 18 e 23, 35 a 40, 75 a 85, 90 a 93 e 135 a 151.



**Figura 22:** Resultado da análise da ferramenta PPI-Pred, representado com janela de intervalo de 3 posições. O eixo Y representa a probabilidade de interação, sendo 1 a maior probabilidade.

**Tabela 6:** Resumo dos resultados do PROSITE com corte de 70%.

Assinatura	Padrão	Porção de E6	Similaridade	
Sítio de fosforilação de cAMP e cGMP dependente de proteína quinase	[ST]-x-[RK]	146 a 149	100%	
Sítio de fosforilação da proteína quinase C	[ST]-x-[RK]	133 a 135	100%	
		142 a 144	100%	
		145 a 147	100%	
Sítio de fosforilação da Caseína quinase II	[ST]-x(2)-[DE]	22 a 25	100%	
		86 a 89	100%	
		145 a 148	100%	
Sítio de Fosforilação da tirosina quinase	[RK]-x(2,3)-[DE]-x(2,3)-Y	26 a 32	71%	
		48 a 54	72%	
		53 a 60	71%	
		68 a 76	72%	
		72 a 79	100%	
		77 a 84	72%	
Sítio de N-miristoilação	G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}	86 a 92	71%	
		57 a 62	100%	
		85 a 90	74%	
		130 a 135	88%	
Assinatura da família araC , proteína bacteriana regulatória	[KRQ]-[LIVMA]-x(2)-[GSTALIV]-{FYWPGDN}-x(2)-[LIVMSA]-x(4,9)-[LIVMF]-x-{PLH}-[LIVMSTA]-[GSTACIL]-{GPK}-{F}-x-[GANQRF]-[LIVMFY]-x(4,5)-[LFY]-x(3)-[FYIVA]-{FYWHCM}-{PGVI}-x(2)-[GSADENQKR]-x-[NSTAPKL]-[PARL]	134 a 139	88%	
		11 to 55	86%	
		14 to 55	76%	
Sítio ativo da família da tioredoxina	[LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C-[GATPLVE]-[PHYWSTA]-C-{I}-x-{A}-x(3)-[LIVMFYWT]	66 to 109	80%	
Assinatura da Lipocalina	[DENG]-{A}-[DENQGSTARK]-x(0,2)-[DENQARK]-[LIVFY]-{CP}-G-{C}-W-[FYWLRH]-{D}-[LIVMTA]	128 a 146	70%	
Assinatura 1 da cadeia beta de hormônios glicoprotéicos	C-[STAGM]-G-[HFYL]-C-x-[ST]	122 a 135	72%	
Assinatura da família de inibidores da protease serina Bowman-Birk	C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDEKS]-[DEKRHSTA]-C	136 a 142	72%	
Assinatura de proteínas transferidoras de lipídio de plantas	C-x(1,2)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDEKS]-[DEKRHSTA]-C	51 a 66	76%	
Assinatura do domínio de lectina tipo .	C-[LIVMFYATG]-x(5,12)-[WL]-{T}-[DNSR]-{C}-{LI}-C-x(5,6)-[FYWLIVSTA]-[LIVMSTA]-C	[LIVM]-[PA]-x(2)-C-x(1,2)-[LIVM]-x(1,2)-[LIVMST]-x-[LIVMFY]-x(1,2)-[LIVMF]-[STRD]-x(3)-[DN]-C-x(2)-[LIVM]	12 a 36	70%
		30 a 51	74%	
Domínio associado a metais pesados	[LIVNS]-x-{L}-[LIVMFA]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]-{LV}-x(2)-[LIV]-x(9,11)-[IVA]-x-[LVFYS]	80 a 106	74%	
		25 a 54	77%	
Assinatura da proteína ribossomal L31	[DES]-[IVT]-x(4)-H-[PT]-[FAVY]-[FYW]-[TISN]-x(9,13)-[GN]-[KRHNQ]	25 a 55	70%	
Assinatura da proteína laminina-type, domínio tipo EGF	C-x(1,2)-C-x(5)-G-x(2)-C-x(2)-C-x(3,4)-[FYW]-x(3,15)-C	72 a 94	70%	
Assinatura tipo dedos de zico tipo PHD	C-x(1,2)-C-x(5,45)-[VMFLWIE]-x-C-x(1,4)-C-x(1,4)-[WYFVQHLT]-H-x(2)-C-x(5,45)-[WFLYI]-x-C-x(2)-C	51 a 80	73%	
		30 a 106	75%	
		63 a 139	75%	

### 4.3 Dinâmica molecular

Vários parâmetros foram testados em relação ao zinco, às forças e à carga da proteína. Os resultados dos testes não serão todos mostrados neste trabalho. O teste protótipo realizado com 2FK4\_03, que foi escolhido para ser utilizado na estrutura completa obtida neste trabalho, manteve ao longo dos 5 ns, a topologia da proteína. As duas porções que mais foram perturbadas ao longo da dinâmica foram a fita S1 (que supostamente realiza a interação com o outro domínio) e a hélice H3. As demais características do peptídeo se mantêm significativamente inalteradas, incluindo o sítio de ligação ao zinco.

Através destes testes estipulamos que o modelo que deve ser utilizado para a ligação ao zinco deve ser o modelo *non-bonded*. Como esperado, a RMSD aumenta ao longo do tempo. Em torno dos 700 ps houve um salto no valor do RMSD, que passou de 2,5 para aproximadamente 3,2 Å. Este salto coincide com o momento em que a fita S1 se desprende da folha beta, passando a gerar maior flutuação da porção amino-terminal da estrutura.

O protocolo utilizado para 2FK4\_03 também foi utilizado para o modelo obtido por homologia de E6C e para a proteína completa, com pequenas alterações de alguns parâmetros do zinco. Para E6C, a topologia geral também foi mantida, a estrutura relaxou levemente e o íon de zinco se manteve seguro pela proteína durante todo o processo. Diferentemente do modelo experimental, a fita S1 não se desprende da folha beta e foi capaz de manter a estrutura íntegra ao longo de toda a simulação.

#### 4.3.1 Proteína completa

A dinâmica molecular quando aplicada à proteína completa, produziu resultados interessantes, como pode ser observado na Figura 24. Ao longo da

dinâmica, entre 6 e 7 ns, a proteína separou seus dois domínios, abrindo a estrutura, porém mantendo todo o restante da topologia relativamente intacta. O movimento relativo entre os domínios envolveu não só o afastamento dos dois domínios, como também uma considerável rotação sobre o eixo longitudinal. A estrutura em grampo da folha beta em cada um dos domínios permanece intacta ao longo de todo o processo, mas a fita que realiza a interface entre os domínios apresenta, como esperado, maior propensão à perturbação. As hélices H1 e H2 de ambos os domínios permanecem aparentes, enquanto a hélice H3, novamente, é bastante perturbada.

No modelo inicial da estrutura nenhum dos domínios apresentava o sítio de ligação ao zinco com geometria tetraédrica. Ao contrário do que se esperava, a dinâmica não corrigiu a posição tetraédrica das cisteínas em relação ao zinco. Isso pode ter se dado, principalmente, por persistir certa imprecisão dos parâmetros utilizados nestes protocolos, o que não invalida os demais resultados obtidos. De qualquer forma, como pode ser visto ao fim da dinâmica molecular, o íon de zinco se mantém coordenado pelos mesmos 4 resíduos de cisteína de cada domínio da proteína. Calculamos a média das distâncias Cys-Zn ao longo de toda a trajetória para ambos os domínios e o domínio N-terminal mostrou menor distância média ( $2,34 \pm 0,06 \text{ \AA}$ ) que o domínio C-terminal ( $2,37 \pm 0,07 \text{ \AA}$ ). Na Figura 23 E, podemos ver a comparação da distância Cys-Zn entre todas as cisteínas envolvidas com a coordenação de zinco.

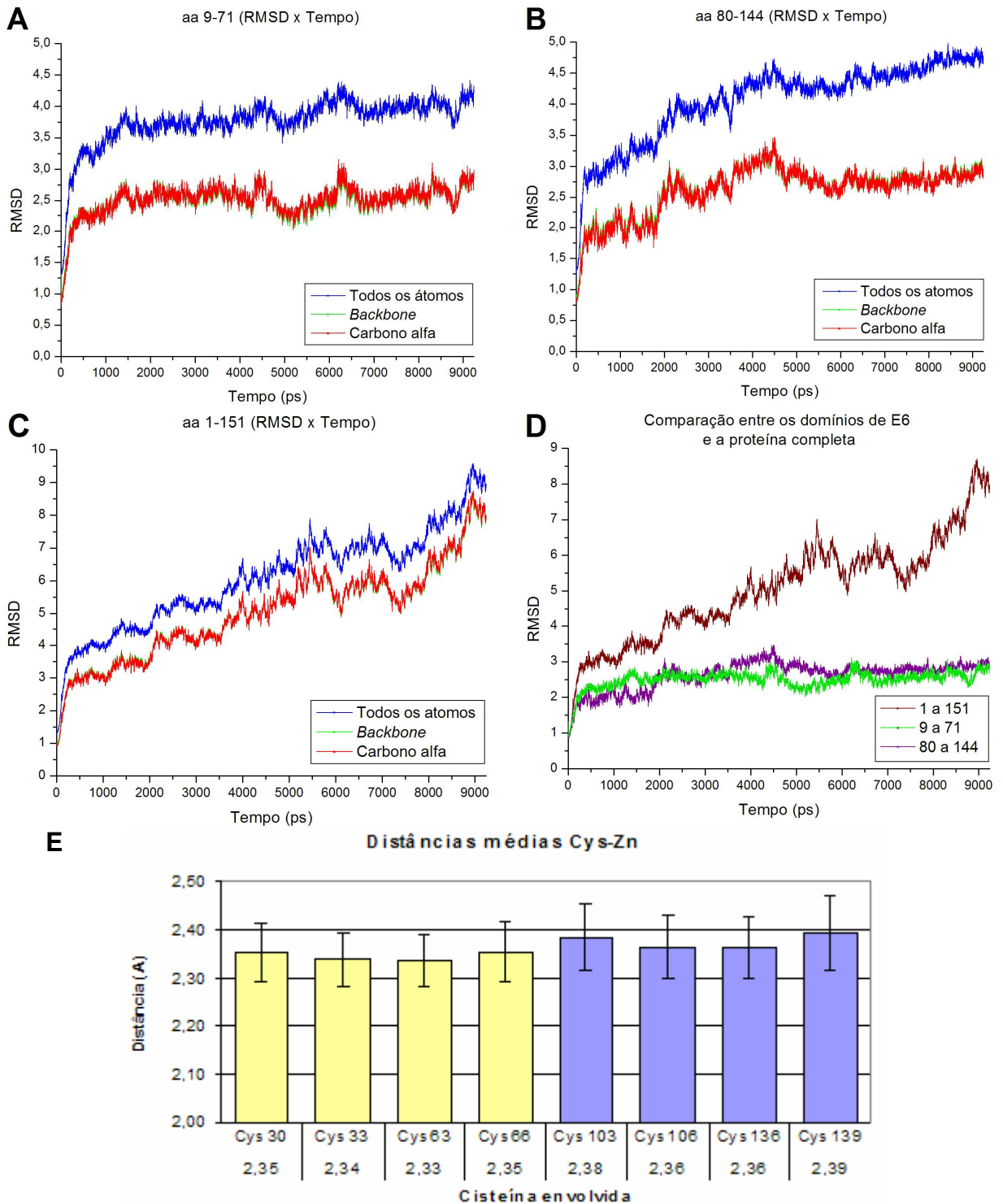
O RMSD, para a estrutura completa, se mostrou bastante alto. Os valores de RMSD para todos os parâmetros aumentam de forma gradativa e quase constante até 4,5 ns, onde chegam a um patamar e se mantêm relativamente estável até 8,0 ns (esqueleto/*backbone*:  $5,74 \pm 0,36 \text{ \AA}$ ). No último nanosegundo da

simulação o RMSD variou em torno de 2 Å, aumentando até em torno de 8,8 ns e se mantendo bastante estável no restante do tempo, indicando possível estabilização da estrutura.

Ao analisarmos, porém, individualmente as porções da proteína, podemos ver que há influência bastante grande das regiões inseridas sem molde, sabidamente mais flexíveis, no resultado final do RMSD da proteína inteira. Uma visão geral destes resultados pode ser vista na Tabela 7 e na Figura 23 A a D. De uma forma objetiva podemos observar que a porção C-term, do resíduo 80 ao 144 obteve um RMSD máximo de 3,46 Å, enquanto a porção N-term, removendo os 8 aminoácidos iniciais (apenas os aminoácidos 9 a 71) apresenta RMSD em torno de 3,1 Å. Os gráficos que demonstram o distanciamento dos domínios ao longo da trajetória dinâmica podem ser vistos na Figura 24 A a E.

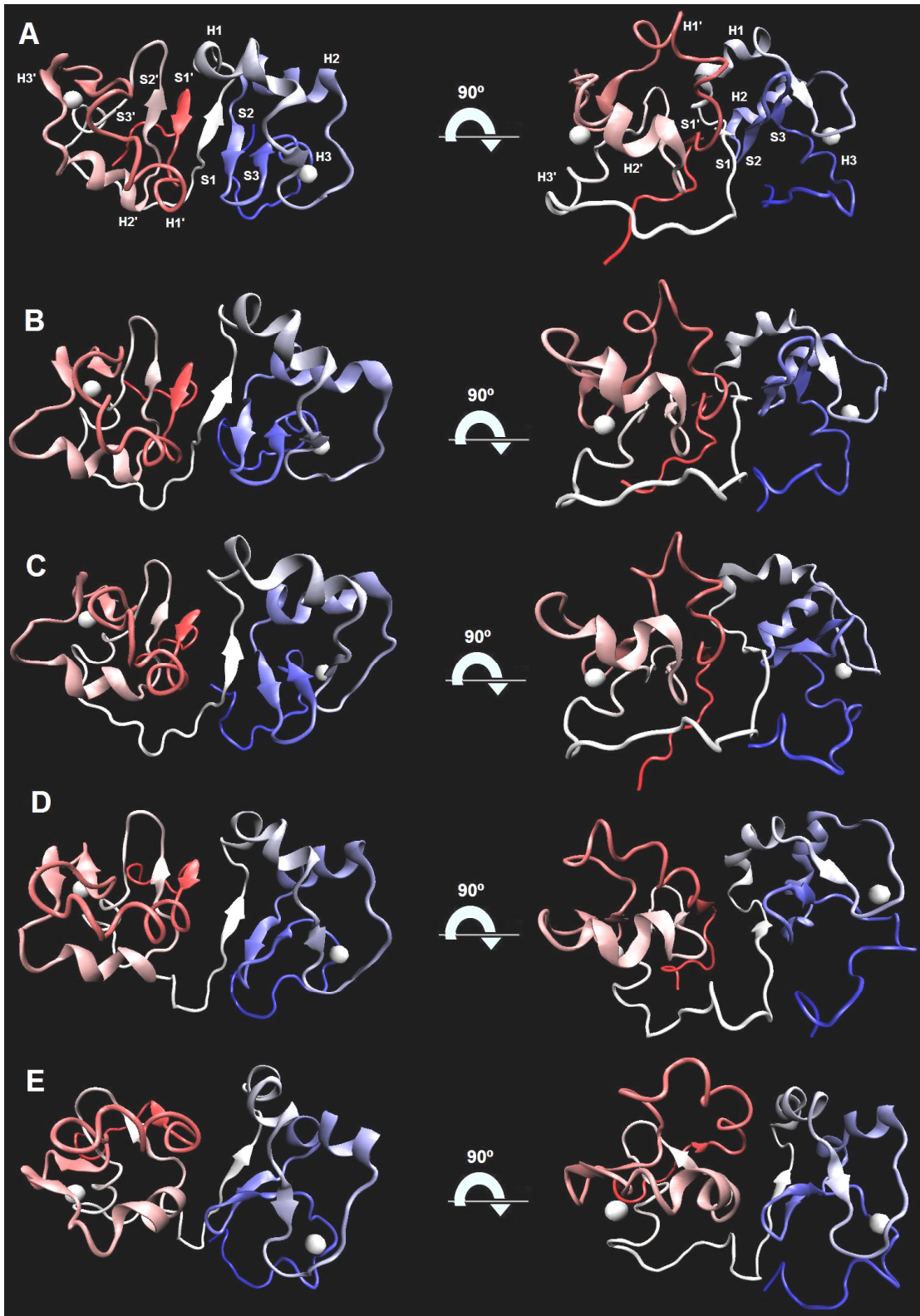
**Tabela 7:** Valores de RMSD para as porções de E6 de HPV 16 ao longo da trajetória da dinâmica molecular, em relação ao modelo inicial 5CHI\_03. \* valor em Ångstron; \*\* A trajetória mantém uma tendência contínua de aumento, não sendo viável o cálculo de uma média de RMSD.

Porção da E6 (aa)	RMSD final (9,242 ns)*			RMSD <i>backbone</i> máximo*	RMSD <i>backbone</i> *
	Todos átomos	<i>Backbone</i>	Carbono alfa		média±DP
1 a 151	8,94901	7,95393	7,99958	8,70579	N/A**
9 a 71	4,21171	2,83663	2,85201	3,1002	2,55±0,16
80 a 144	4,68700	2,90283	2,88311	3,46476	2,79±0,17



**Figura 23: A-D:** Gráficos indicativos do valor do RMSD em Angstroms, de diferentes porções da proteína E6 completa (A: 9-71: N-term; B: 80-144: C-term; C: 1-151: proteína completa) em relação às suas porções equivalentes da estrutura do modelo inicial, ao longo do tempo (em picosegundos). Em vermelho, o RMSD do Carbono alfa; em verde, do esqueleto (átomos N-CA-C-O); em azul, de todos os átomos, incluindo cadeias laterais. D: Comparação do RMSD dos átomos do esqueleto de todas as porções da E6 (N-term – verde; C-term – violeta; proteína completa - marrom). **E:** Médias das distâncias Cys-Zn ao longo de toda a simulação. Em amarelo: N-term; em azul: C-term.





**Figura 24:** Representação do tipo *New Cartoon* de pontos da trajetória dinâmica da proteína completa selvagem E6 de HPV16. A: tempo 0 ps; B: 2146 ps; C: 4556 ps; D: 7226 ps; E: 9242 ps. Observa-se claramente o distanciamento dos domínios N- (vermelho) e C-terminal (azul) ao longo da trajetória.

As simulações realizadas no modelo sugerido de E6C (6E6C\_15) e da proteína completa (5CHI\_03) mostraram resultados melhores que os apresentados por 2FK4 (Tabela 8). Calculamos a média dos valores do RMSD a partir de 2 ns, para excluirmos completamente a interferência da estabilização inicial do complexo. Os resultados mostram que ambos os modelos obtidos por modelagem por homologia neste trabalho, apresentaram maior estabilidade ao longo da simulação, quando comparados ao molde experimental 2FK4.

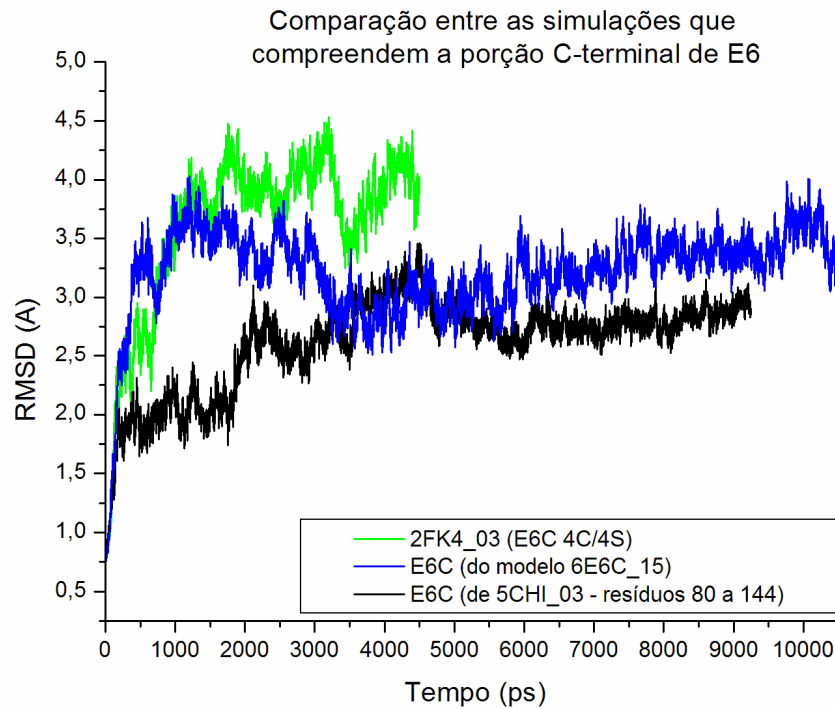
**Tabela 8:** Tabela mostrando resultados comparativos dos modelos da estrutura de E6C que sofreram a simulação. O RMSD está expresso em Ångstrons e a média foi calculado para tempo maior que 2,0 ns. 5CHI\_03 (80-144) corresponde à porção do modelo da proteína completa que compreende a região equivalente à 6E6C e 2FK4. Quanto menor o valor de RMSD, menor a divergência da estrutura em relação ao molde. Em todos os casos, o molde utilizado para estas regiões foi o mesmo: 2FK4\_03.

<b>Porção da E6 (aa)</b>	<b>Simulação</b>	<b>RMSD médio (&gt;2,0 ns)</b>	<b>RMSD máximo</b>
2FK4_03	SIM001	3,93±0,23	4,53
6E6C_15	SIM002	3,23±0,26	4,03
5CHI_03 (80-144)	SIMNC001	2,79±0,17	3,46

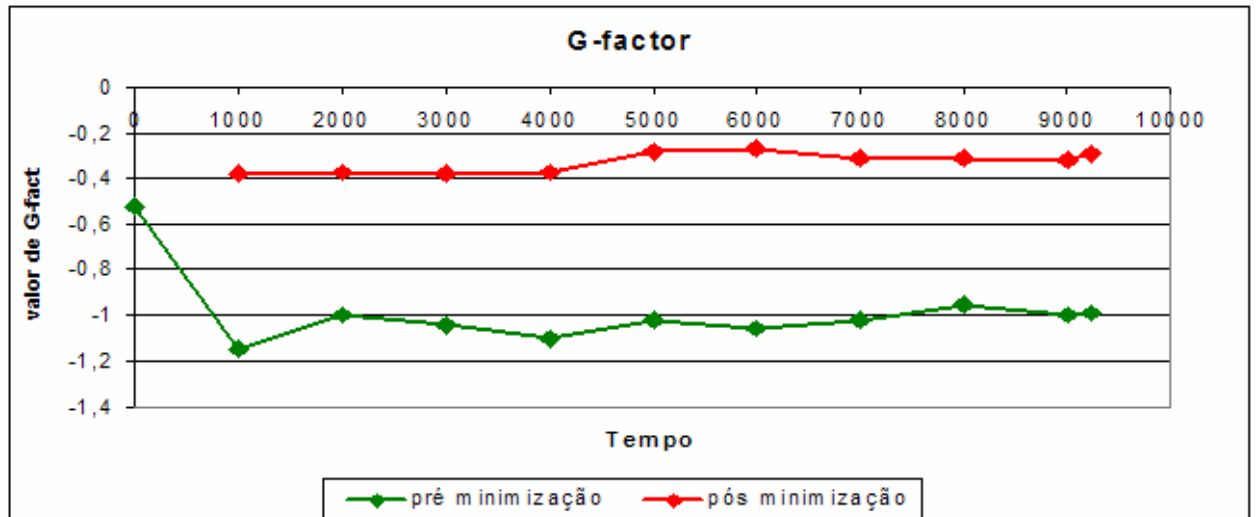
Na figura 25 também fica ressaltada a observação de que no modelo da estrutura completa, a porção C-terminal quando comparada individualmente, se mostra ainda mais estável que nos dois outros modelos, apesar de, como comentado anteriormente, haver a separação dos dois domínios.

Para análise com Procheck do modelo final após simulação, foram extraídos da trajetória 11 estruturas representando 11 tempos da simulação (0, 1001, 2001, 3001, 4001, 5001, 6001, 7001, 8001, 9001 e 9242 ps), os arquivos tiveram sua energia minimizada para diminuirmos distorções e incompatibilidades estéricas. Em todos os modelos, antes ou após a minimização, os gráficos de Ramachandran ficaram aceitáveis, apresentando mais de 98% dos seus resíduos em áreas Permitidas e Adicionalmente Permitidas, sendo que o último modelo (referente ao tempo 9,242 ns) apresentou a seguinte distribuição: 68,1% em Região

Permitida; 31,9% em Região Adicionalmente Permitida e 0% nas demais regiões. Os valores do *g-factor* obtidos para o modelo final de E6 mostraram considerável melhora (Figura 26) após minimização. Para os arquivos minimizados, o *G-factor* apresentou valores sempre maiores que -0,5, o que representa bons resultados.



**Figura 25:** Gráfico comparativo entre as trajetórias das simulações de 2FK4\_03 (modelo experimental com 4 mutações Cys/Ser não conservadas), 6E6C\_15 (modelo por homologia da estrutura do domínio C-terminal selvagem) e 5CHI\_03 (modelo por homologia da estrutura completa selvagem de E6).

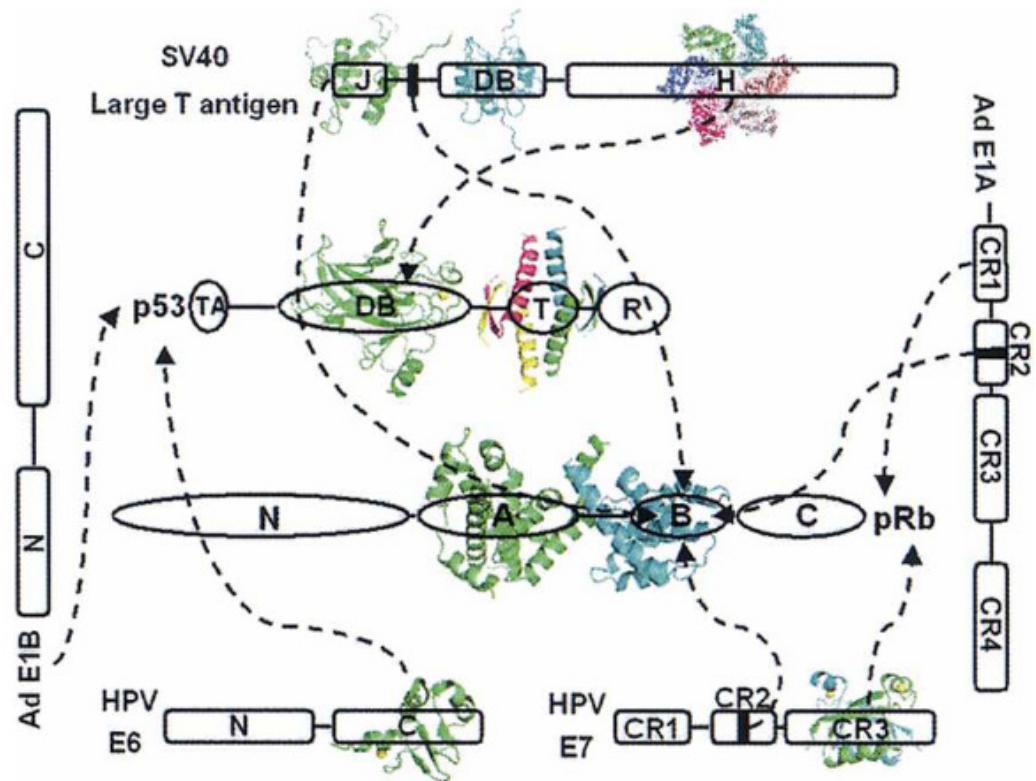


**Figura 26:** Gráfico representando a melhora do desempenho do modelo ao longo da simulação antes e depois da minimização de energia. O ponto 0 no eixo Y representa a estrutura inicial de onde partiu a simulação. A linha verde representa pontos amostrais antes da minimização e a linha vermelha após.

## 5 Discussão

A interação das proteínas E6 de HPV com proteínas celulares ainda é um campo de pesquisa muito aberto para discussão. O pleiotropismo da E6 e suas variedades de funções, nem sempre compartilhadas entre os mais de 100 tipos de Papillomavirus humanos já descritos, torna esta família um quebra-cabeça científico. Nominé et al (2006) propuseram um modelo experimental para a estrutura 3D do domínio C-terminal da proteína E6 (E6C) contendo 4 mutações pontuais em cisteínas não conservadas - Cys80Ser, Cys97Ser, Cys111Ser e Cys140Ser (E6C 4C/4S) - que contém o segundo domínio de ligação ao zinco, compreendendo os resíduos 80 a 144 da proteína. Esta foi a primeira tentativa de sucesso de purificar uma porção da proteína de forma estável e em quantidade suficiente para realizar experimentos com fim de caracterizar biofisicamente a proteína. Nesta dissertação, dedicamo-nos a agregar mais informações, investigando sobre as relações sequência / estrutura / dinâmica / função da proteína E6 completa.

Liu e Marmorstein (2006) resumem o conhecimento até então, acerca da interação física de algumas oncoproteínas virais com seus alvos celulares (Figura 27). O que se sabe com solidez a respeito de E6 é bastante basal e resume-se a uma gama grande de possíveis resíduos de E6, passíveis de realizar ligação de E6 à p53 e duas regiões de p53 que foram confirmadas como sítio de ligação, além da interação com E6AP – a qual se sabe que interage com E6 através de uma sequência rica em leucina que tem conformação de hélice (Cui et al, 2003). Nada se sabe até o momento sobre os resíduos de E6 realmente envolvidos em tais interações. Menos ainda se sabe a respeito da interação física com outros alvos, principalmente quando se trata da proteína como um todo.



**Figura 27:** Galeria de interação entre oncoproteínas virais e proteínas supressoras de tumor. Os supressores de tumor p53, pRb e as oncoproteínas LTag (SV40), E7 e E6 (HPV) e E1A e E1B (Ad) dos vírus de DNA pequenos são mostrados como estruturas compostas de domínios (adaptado da definição de domínios do Pfam e desenhado fora de escala). Representações do tipo *cartoon* do dobramento de domínios independentes destas proteínas também são mostrados. Linhas pontilhadas apontado para os domínios das prteínas indicam a localização de interações estruturais confirmadas entre as proteínas virais e as supressoras de tumor. Linhas pontilhadas apontando apenas para o nome das proteínas indicam interações apenas bioquimicamente definidas. As barras pretas sólidas dentro de SV40 LTag, HPV E7 e Ad E1A indicam o motivo LxCxE. (Extraído de Liu and Marmorstein, 2006)

Os resultados que obtivemos mostram uma proteína com considerável propensão à desordem e que provavelmente tende à estabilidade do dobramento ao interagir com seus ligantes e alvos celulares.

A estrutura completa de E6 é composta de 6 hélices e de uma folha beta de 6 fitas, sendo que cada um dos dois domínios é composto de 3 hélices e 3 fitas. Os domínios ZB aparentam ser relativamente independentes entre si, e apresentam uma flexibilidade na região da interface, o que permitiria as flutuações de orientação relativa entre eles. Esta característica é reforçada pela observação de que, durante a realização da dinâmica molecular em meio aquoso, os *hairpins*

formados pelas duplas de fitas betas mais próximas à extremidade (S2, S3 e S2' e S3') se mantêm com características de formação de folha beta, mas as duas fitas centrais por vezes perdem a característica da estrutura secundária. Isto pode se dar devido ao fato de que estas duas fitas estão envolvidas diretamente na interface entre os dois domínios, estando, desta forma, mais sujeitas ao estresse da dinâmica molecular.

Nominé et al (2005) sugeriram que duas das alças (a primeira entre S1' e H1' e a segunda entre H1 e S1 – ver Figura 19) estão envolvidos na interface entre os domínios. A idéia é reforçada pela presença de dois resíduos hidrofóbicos (Tyr54 e Leu88) posicionados nestas alças e conservados entre vários tipos de HPV. De fato, o que encontramos, foi que estes resíduos participam da interação entre os domínios. A cadeia lateral de Tyr54 interage com as cadeias dos resíduos Leu88 e Gln91 e do resíduo de Leu88, o qual por sua vez, também interage com a cadeia lateral de Asp56.

Conforme havia sido sugerido por Tao et al (2003), a proteína E6 de HPV 16 apresenta 3 sinais de localização nuclear (NLS), sendo o primeiro localizado entre os resíduos Phe2 a Glu7, o segundo entre Lys65 a Lys72 e o terceiro entre os resíduos Lys115 a Arg124. O resultado que obtivemos na modelagem por homologia demonstrou que todas estas porções onde se sugere que estejam os NLS, são regiões acessíveis. A primeira faz parte da Alça 1 (anterior à primeira fita – S1) que não foi considerado no modelo experimental de Nominé et al (2006). A segunda compreende parte da hélice H3' e da alça que conecta o domínio N-terminal ao C-terminal. O terceiro NLS se localiza entre H2 e L2. Todas estas porções estão em regiões expostas e se encontram na mesma face da proteína. Em relação à estrutura secundária das NLS, com exceção de 4 resíduos, todos demais

pertencem a alças (regiões não estruturadas).

Mutações sítio dirigidas e deleções nestas regiões levam à inibição e/ou degradação de p53 ou até mesmo à inibição da ligação entre as proteínas p53 e E6, p.ex.  $\Delta$ Val2-Leu12 (Foster et al, 1994) ou Phe2Val (Liu et al, 1999). Masson et al (2003) demonstraram que um anticorpo monoclonal cujo epítipo são os resíduos de aminoácidos Gln3 a Arg8, ligado ao seu antígeno, inibe a degradação de p53. Foster et al (1994), Park et al (2002) e Dalal et al (1996) demonstraram uma diminuição significativa da ligação de E6 com p53 e E6AP e da degradação de p53 pelo mutante  $\Delta$ His118-Lys122, entre outros mutantes. Na Tabela 9 (ver Apêndice III), pode-se ver que a quase totalidade dos ensaios mutacionais e de deleções realizados, ou nestas porções (NLSs) ou englobando-as, levam a perda parcial ou total da capacidade de ligar e/ou degradar p53.

Liu et al (2009) realizaram experimentos para a determinação do conteúdo de zinco e obtiveram resultados que indicam uma menor habilidade da porção N-terminal de ligar ao zinco, quando comparada com a porção C-terminal. Uma deleção na alça localizada na porção central da proteína ( $\Delta$ Tyr76-Tyr79), que faz a conexão entre os dois domínios, causa a perda quase total da capacidade de ligação a zinco do domínio N-terminal. Os resultados do grupo de Liu demonstraram que tal mutação não apenas perde a capacidade de ligar o zinco em sua porção N-terminal, como também acaba por reduzir em 40% suas funcionalidades biológicas, como por exemplo, a capacidade de ligação ao peptídeo E6apn1, que mimetiza a porção de E6AP que se liga à E6. Conforme o autor sugere (corroborando a hipótese de outros trabalhos – Nominé et al, 2006) de que esta região de conexão auxiliaria a definir a orientação entre os dois domínios e o encurtamento desta alça afetaria a estrutura neste sentido. O modelo sugerido neste trabalho aponta uma



explicação alternativa: isto também poderia ocorrer devido ao fato que o encurtamento da alça levaria ao afastamento da pequena hélice H3 que contém dois dos resíduos de cisteína (posições 63 e 66) que em conjunto com os resíduos Cys30 e Cys33, realizam a coordenação do íon  $Zn^{2+}$ .

Liu et al (2009) também demonstraram que há diferença entre os domínios C- e N-terminal na estabilidade térmica da estrutura secundária, através de estudos de desnaturação por temperatura, onde é realizado o monitoramento da estrutura secundária por dicroísmo circular enquanto a temperatura é elevada a partir de 0°C (nesta temperatura proteína em solução estaria totalmente estruturada). O domínio C-terminal apresentou o equilíbrio 1:1 de estrutura íntegra e desnaturada (onde metade da proteína na solução está desnaturada e a outra metade está em sua forma nativa) em 58,0°C, enquanto a porção N-terminal foi em 52°C. A estrutura completa da E6 obteve este equilíbrio em 55,4°C. Como esperado, a temperatura de equilíbrio obtida para o mutante  $\Delta Tyr76-Tyr79$  foi de 52,5°C fica abaixo da obtida pela estrutura completa, levando mais uma vez à conclusão de que a mutação leva a uma desestruturação da proteína, diminuindo sua estabilidade.

Apesar destas observações, nos experimentos computacionais de dinâmica molecular em meio aquoso, a porção N-terminal se mostrou mais estável que a C-terminal em relação à distorções da estrutura terciária (Figura 23). De uma forma geral, ambas se distanciaram da estrutura inicial, mas a estrutura terciária do domínio C-terminal ficou um pouco mais comprometida. A estrutura N-terminal pode ser mais estável, como observado na dinâmica molecular, como uma forma compensatória pela menor afinidade com pelo íon  $Zn^{2+}$ , uma vez que a presença do íon é sugerida como fator limitante que permite à estrutura se manter íntegra (Lazarczyk e Lavré, 2008).

Liu et al (2009) ainda demonstraram que E6N tem maior afinidade com por E6apc2 do que E6C, sugerindo que E6N contribui mais para ligação com E6AP. O resultado obtido pelo autor mostra que a porção N-terminal, mas não a C-terminal, perderia seu íon de zinco na presença de agente quelante (lembrando que os kits de expressão *in vitro* – reticulócitos de coelho - possuem certa quantidade de EGTA), mas não perderia a capacidade de ligar a E6AP ou de degradar p53. Estes dados, juntamente com nossos resultados de que o domínio N-terminal demonstra maior estabilidade em simulação computacional, nos levam a concluir que o zinco na porção N-terminal de fato não é indispensável para as funções e para a estabilidade da proteína. Conseqüentemente, o domínio E6N deve conter outras características que permitam a manutenção da estrutura geral – e função – que não existam no domínio C-terminal, como por exemplo, a diferença de sequência de aminoácidos.

Nossos resultados de dinâmica molecular também sugerem que existe bastante flexibilidade intra- e inter-domínios, já que foi demonstrada a abertura da molécula em dois domínios separados, presos apenas pelo conector. Corroborando outros autores, a separação dos domínios parece implicar em funções/estruturas independentes (Zhang et al, 2007; Ristriani et al, 2000, 2001; Lagrange et al, 2005). Não existe uma forma de definir qual foi o domínio que se deslocou, sendo, aparentemente, afastamento recíproco. O fato de ao longo de toda a modelagem o zinco permanecer preso ao modelo pelos mesmos resíduos de cisteína apesar de não ser uma dinâmica com protocolo de ligação a zinco modelo de ligação covalente, demonstra que o modelo é bom, mas outras falhas, como a geometria não-tetraédrica das ligações com o Zinco, lembram que este modelo ainda necessita de aprimoramentos.

Um problema que foi encontrado com freqüência nas dinâmicas

moleculares que realizamos, envolve o oxigênio da carbonila da Arg102 que no decorrer da dinâmica acabava se aproximando demais do zinco, levando à ocorrência da interação entre o oxigênio do resíduo e o zinco e conseqüentemente ao afastamento da cisteína 103 (dados não mostrados). Dependendo dos parâmetros utilizados, essa interação era transitória ou permanente. O fato foi corrigido ajustando os dados de entrada da dinâmica, conforme citado anteriormente.

Uversky et al (2006) demonstraram por ferramentas de bioinformática como PONDR (*Predictor of Naturally Disordered Regions*) que E6 de HPV 16 apresenta em torno de 44% de seus resíduos em desordem intrínseca. Isso poderia explicar que o desenvolvimento de um modelo completo da proteína, principalmente no que diz respeito às duas extremidades (N- e C-terminal – 8 e 7 resíduos, respectivamente) e à região central que conecta os dois domínios, seja difícil. Também explicaria, em parte, porque apesar das diversas tentativas que foram feitas no sentido de se purificar a proteína para realizar experimentos de cristalografia ou ressonância magnética, foi apenas inserindo mutações e gerando uma proteína truncada que o procedimento foi capaz de gerar um produto suficientemente estável para experimentos de NMR.

Curiosamente, o trabalho de Uversky et al (2006) sugere que, independente da ferramenta de análise utilizada, tanto a proteína E6 como E7 seriam, em sua forma nativa, majoritariamente desestruturadas e que as proteínas E6 dos HPV de alto risco apresentam tendência a uma maior desordem que os de baixo risco, principalmente nos 7 últimos resíduos. A grande gama de alvos celulares para estas duas proteínas, característica chamada de pleiotropia, é explicada em parte por esta característica de desordem intrínseca.

A extremidade C-terminal da E6 (principalmente os 7 resíduos terminais), de fato apresenta características de desordem intrínseca. Nos experimentos de NMR, desenvolvidos por Nominé e seu grupo (2006) mostra a extremidade C-terminal da estrutura modelada de forma bastante irregular. Todos os 10 modelos que foram publicados para esta estrutura apresentam orientações distintas nestas posições. Em nosso modelo, a quantidade de possíveis conformações aceitáveis (visualmente e pelas análises gráficas – os dados específicos desta região não são mostrados) corrobora os resultados dos autores supracitados, confirmando a desordem intrínseca da região.

Zhang et al (2007) utilizou justamente esta porção para modelar domínios PDZ de algumas proteínas da família MAGUK (como SAP97 e MAGI-1) com o peptídeo (aa 144-151) em seu sítio de ligação. Na estrutura que Zhang obteve (Figura 6), pode-se ver que a extremidade da porção C-terminal, ao encaixar no domínio PDZ, assume a conformação de uma fita antiparalela com a fita  $\beta$ B de PDZ, resultando em uma folha  $\beta$ . Adquirir conformação e estabilidade apenas em contato com seu alvo é característica bastante comum em muitas outras proteínas (Uversky et al, 2006; Xie, 2007).

Vários autores já sugerem uma grande associação entre proteínas relacionadas com câncer e outras que atuam como fatores de transcrição com grande incidência de desordem intrínseca. Iakoucheva et al (2002), por exemplo, demonstraram por ferramentas de bioinformática que mais de 70% das proteínas associadas à câncer e à sinalização celular, apresentam estrutura em sua maior parte desordenada.

Através da identificação de similaridades entre os domínios de ligação à zinco do E6 e da Endonuclease VII de T4 (descrita por White et al, 1997),

Ristriani et al (2000) analisou a ligação de E6 ao DNA e verificou que apenas o domínio E6C se ligaria, de forma independente de E6N, ao DNA (Junções de Holiday), podendo inclusive ligar 2 domínios C-terminal à estrutura de DNA sem realizar contatos proteína-proteína, sem apresentar tendência à dimerização e sem deformar a estrutura cruciforme (Ristriani et al, 2001). O reconhecimento reportado é estrutura-dependente (fato demonstrado por ensaios que utilizaram a mesma sequência para gerar estruturas com 3 pontas ou dupla-fita, além da utilização de estruturas cruciformes com sequências totalmente diferentes da inicialmente usadas). Outra característica interessante é que, ao contrário de tantas outras proteínas que se ligam ao mesmo tipo de estrutura, E6C se liga ao DNA em forma de monômeros, e não em dímeros.

Até hoje não se sabe a função real da ligação de E6 ao DNA, assim como não se conhece nenhuma atividade enzimática para E6. Com base nas várias regiões de E6N com probabilidade de realizar interação proteína-proteína e no fato de que os domínios de E6 podem atuar fisicamente separados, associados apenas pelo conector, sugerimos que uma explicação de porque apenas a porção C-terminal se ligaria ao DNA. A solução pode estar baseada na possibilidade de a porção N-terminal realizar interação com outras proteínas, ainda potencialmente desconhecidas, que agiriam sobre o complexo E6C-DNA. Desta forma, o domínio N-terminal poderia servir como um agente de recrutamento para a formação de um complexo protéico/enzimático pra realizar alguma função sobre o DNA. A maior instabilidade / desordem mostrada por E6C, em comparação à E6N é compatível com sua função de ligação ao DNA (revisado por Xie et al, 2007)

Apenas HPVs de alto-risco reconhecem essas junções de DNA (fato que corrobora com o menor grau de similaridade da E6 dos HPV de baixo risco com

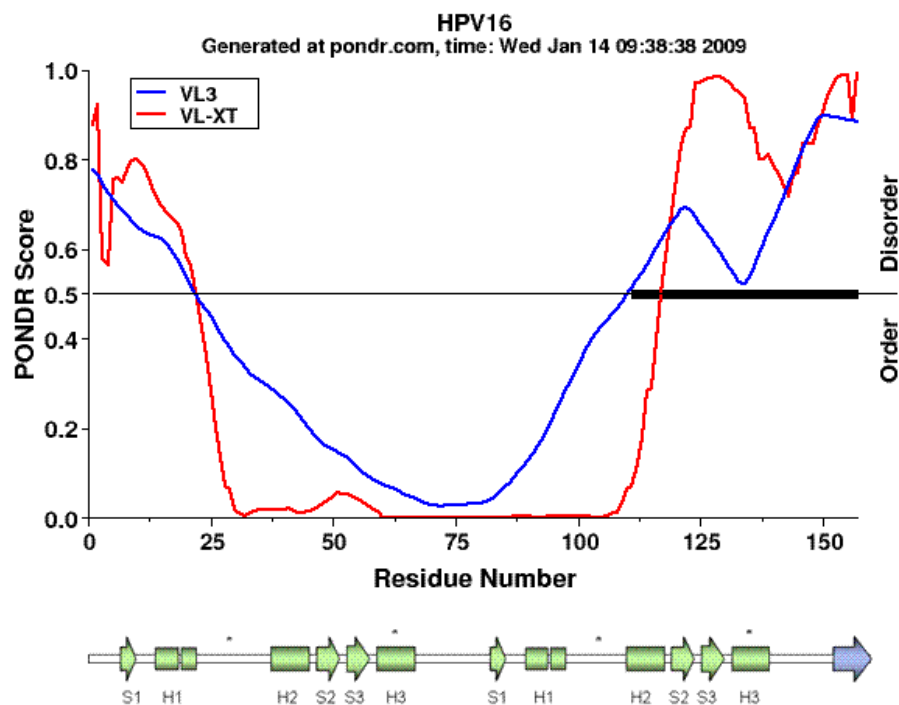
a endonuclease VII). Foram analisados dois conjuntos de mutações pontuais: R109A/Q114A/K115A (ZD2-a) e K128A/K129A/Q130A/R131A (ZD2-b) (a numeração dos resíduos no trabalho de Ristriani e colegas considerou a sequência completa como tendo 158 aminoácidos). ZD2-a compreende os resíduos que se encontram flanqueando a dupla de cisteínas envolvidas na ligação com o zinco, enquanto o segundo, ZD2-b, compreende os resíduos que se encontram na volta entre as fitas S2 e S3. Todos os resíduos mutados são polares e ficam voltados para fora e para o mesmo lado da proteína. Os resultados mostrados pelo grupo indicam que estas regiões estariam envolvidas na ligação com o DNA, pois em ambos os casos a presença das mutações abole a formação do complexo E6-DNA. Uma análise superficial destes mutantes através de modelagem por homologia mostra que as mutações não alterariam significativamente a estrutura terciária da proteína, não sendo este o motivo pelo qual a função é perdida (dado não mostrado).

Nas Figuras 28 e 29 seguem os resultados da FoldIndex (Prilusky, 2005) e PONDR, como já realizado por (Uversky et al, 2006), para fins ilustrativos. Em E6 a porção C-terminal apresenta maior desordem, com alguns pontos da N-terminal com alguma tendência à desordem. No entanto, como já citado, é a porção C-terminal de E6 que liga ao DNA, de forma independente da N-terminal.

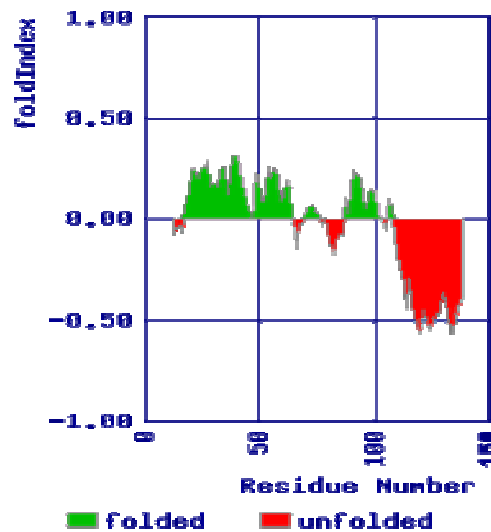
VL3 é um algoritmo que combina a predição de 30 redes neurais para proteínas inteiras e, em adição a outros parâmetros ainda caracteriza a proteína por dicroísmo circular, proteólise limitada e outras abordagens físicas ou químicas (Radivojak et al, 2003). VL-XT é uma junção de duas ferramentas de predição de desordem intrínseca (VL1 e XT), onde VL1 é *Variously characterized Long* versão 1, que prediz regiões de desordem em resíduos não terminais e XT é *X-ray characterized N- and C-terminal predictors* que prediz se os resíduos das regiões

N- e C-terminais apresentam desordem intrínseca. A predição de VL1 começa e termina a 11 resíduos das extremidades e o XT prediz os 14 resíduos iniciais e finais. Uma média simples é obtida das regiões de sobreposição. Uma janela de 9 aminoácidos é utilizada para atenuar os valores ao longo da sequência. As 4 posições iniciais e finais não são atenuadas.

Nosso trabalho, corroborando o de Nominé et al (2006) e Liu et al (2009), sugere que a estrutura de E6 é globular e possui o final da extremidade C-terminal desestruturado. Uversky et al (2006) obtiveram também, através das análises de probabilidade de desordem intrínseca, o mesmo resultado. Na Figura 28, também temos uma comparação entre os resultados do PONDR VL-XT e VL3 com a topologia obtida de E6 de HPV 16.



**Figura 28:** Comparação entre a predição de desordem intrínseca e a estrutura secundária de E6 de HPV16. Os pontos acima do esquema da estrutura secundária indicam os pontos de ligação com o zinco. A seta azul indica a extremidade C-terminal que ganha a conformação de fita beta ao interagir com o ligante PDZ.



**Figura 29:** Gráfico representativo do resultado da análise de *Fold Index* da proteína E6 de HPV 16 com janela de 26 resíduos.

Como podemos ver em ambas figuras, os finais das extremidades N- e C-terminal apresentam grande propensão à desordem, assim como sugerimos com base nos resultados que obtivemos neste trabalho. Ambas as extremidades, Alças 1 e 3, ficam expostas e devem interagir com seu(s) alvo(s) celulare(s), quando apenas então assumirão alguma conformação secundária. No caso da Alça 3, esta característica já foi descrita em relação à ligação com os domínios tipo PDZ (Zhang et al, 2007).

Em relação à Alça 2, que é a região de conexão entre os dois domínios, para obtermos a estrutura final (fixa) de E6 selvagem completa, apenas encontramos uma possível conformação para esta alça. Este resultado é sustentado pelos resultados VL3 e VL-XT (Figura 28) que mostra baixa propensão à desordem nesta região da proteína. Por outro lado, ao realizarmos a simulação por dinâmica molecular no modelo da estrutura, vimos que ela permite a separação dos dois domínios, que são mantidos conectados apenas pela Alça 2. Neste caso, o resultado do Fold Index (Figura 29) vai mais de encontro às nossas observações de que nesta região central também necessita haver flexibilidade, de forma a permitir as funções



independentes dos domínios de E6.

Várias proteínas que possuem motivos de ligação ao zinco estruturam-se como domínios globulares  $\beta\beta\alpha$  enovelados independentemente e que são contêm por regiões conectoras flexíveis. Essa região conectora é um elemento importante que ajuda a controlar o espaço entre os domínios ZB e o DNA e dessa forma contribui para a coordenação da atividade do fator de transcrição (Liu et al, 2006). A dinâmica molecular realizada com a proteína completa (5CHI\_03) mostrou que a proteína pode se abrir em dois domínios independentes unidas por um conector, da mesma forma que os fatores de transcrição comentados no artigo de Liu et al (2006).

Ristriani et al (2000, 2001) demonstraram a ligação de E6 de HPVs de alto-risco a Junções de Holiday através de seu domínio C-terminal. Nominé et al (2006) realizaram ensaios de co-purificação que sugerem que há um sítio de ligação a RNA que pode se sobrepor ao sítio do DNA. Masson et al (2003) demonstraram a co-localização de E6 com super-estruturas ribonucleoprotéicas e com áreas ribossomais no citoplasma. Isso abre um novo campo de suposições em relação às funções de E6 relacionadas ao DNA e RNA.

Os resultados que obtivemos com a ferramenta PPI-Pred mostram muitos possíveis pontos de interação com outras proteínas. A porção C-terminal obteve os maiores valores de probabilidade de interação concordando com outros resultados até agora obtidos por nós e outros grupos de que esta região está envolvida diretamente com diversas funções de E6, como ligação ao DNA (Ristriani et al, 2001) ligação aos domínios PDZ (Zhang et al, 2007) entre outros.

Outras possíveis regiões de interação abrem portas para a discussão de outras funções de E6. Por exemplo, o mutante  $\Delta\text{Glu18-Asp25}$

estudado por Elbel et al (1997) apresentou ausência completa de degradação a p53 e ligação tanto a p53 quanto à E6AP, além de reprimir a totalmente transformação celular e parcialmente transativação de p53. O mutante Leu37Ser (também contido em região com alta probabilidade de interação com outras proteínas) inibe a ligação à E6AP e ativa hTERT (Liu et al, 1999; Sekaric et al, 2008). A terceira região com grande probabilidade de interação compreende parte da Alça 2 e está adjacente ao segundo Sinal de Localização Nuclear (Tao et al, 2003).

O resultado do Prosite, que analisa semelhanças entre motivos e domínios conhecidos e a sequência de interesse, retornou semelhanças com o motivo dedos de zinco Cys<sub>4</sub>-HisCys<sub>3</sub>, também conhecido como PHD (Plant Homeodomain). PHD é um motivo presente em alguns fatores de transcrição de animais e plantas que possui dois sítios de ligação a zinco sobrepostos e que se ligam a dois íons Zn<sup>2+</sup> de forma semelhante à dos domínios RING e FYVE (citado por Pascual et al, 2000). Em função dos sítios de ligação ao zinco se sobreporem nestes motivos PHD, a estrutura toda forma um único domínio globular. Apesar dos domínios PHD não apresentarem semelhança tridimensional com o modelo sugerido da estrutura de E6, não podemos deixar de pensar na possibilidade de esta semelhança representar alguma função homóloga ainda por ser descoberta.

## 6 Artigo Científico

- Artigo: “**Homology modeling and molecular dynamics of the complete wild type E6 structure from Human Papillomavirus 16**” – submetido ao periódico *Journal of Molecular Graphics and Modelling*, ISSN 1093-3263, classificação Qualis B1 (área CB-1 da CAPES - © ISI *Journal Citation Report*), de circulação internacional, como um dos pré-requisitos para obtenção do título de mestre pelo Programa de Pós-Graduação em Biologia Celular e Molecular da PUCRS.

## **Homology modeling and molecular dynamics of the complete wild type E6 structure from Human Papillomavirus 16**

Camila Mundt Ilgenfritz<sup>1,2</sup>, Osmar Norberto de Souza<sup>1\*</sup> and Virgínia Minghelli Schmitt<sup>2</sup>

<sup>1</sup> Laboratório de Bioinformática

Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS

Av. Ipiranga, 6681 - Prédio 12

90619-900 - Porto Alegre - RS - Brasil

Fone: 55-51-3320-8608

\*corresponding author E-mail address: [osmar.norberto@pucrs.br](mailto:osmar.norberto@pucrs.br)

<sup>2</sup> Departamento de Análises Clínicas e Toxicológicas

Pontifícia Universidade Católica do Rio Grande do Sul - PUCRS

Av. Ipiranga, 6681 - Prédio 12

90619-900 - Porto Alegre - RS - Brasil

Fone: 55-51- 3320 3512

## **Abstract**

Human Papillomavirus (HPV) is a necessary cause of cervical cancer. There has been no success so far in modeling, experimentally or not, the tridimensional structure of the complete wild type HPV E6 protein, which in association with E7 are the most studied HPVs oncoproteins. In this work, we propose a complete structure of the E6 protein of HPV 16 (most frequent high risk HPV) through homology modeling using as template the experimental NMR model reported by Nominé et al (2006), which comprehends the C-terminal domain with four non conserved C/S mutations (E6C 4C/4S). The quality analysis of the putative structure was carried out by 3D structure assessment tools, comparison with the experimental NMR data and by molecular dynamics in aqueous environment.

Keywords:

Human Papillomavirus; HPV 16; E6 oncogene; homology modeling; molecular dynamics

## Introduction

Human Papillomavirus (HPV) are small DNA virus considered a necessary cause for cervical cancer, associated with virtually all cases. HPV type 16 is responsible alone for over 50% of the cervical cancer cases around the world. (Bosch et al, 2002). The genome is approximately 8 Kb long, with small variations among types and subtypes, with 3 identified regions: (Early – E1<sup>^</sup>E7); Late (L1 and L2); and the long controlling region (LCR) (Howley, 2001; Schmitt, 2006).

E6 is a small pleiotropic oncoprotein with 151 amino acids. It has several cellular targets, amongst which is the tumor suppressor protein p53 (Huibregtse et al, 1993), responsible of regulating cellular cycle progression as a sentinel of the process conformity. Other target proteins are, for example, the MAGUK proteins (Membrane Associated Guanylate Kinase homologues – Zhang, 2007), CBP/p300 (Zimmermann et al, 1999; Patel et al, 1999), hTERT (Veldman et al, 2001) and Bak (Thomas and Banks, 1999; Jackson et al, 2000), among many others. Interaction with proteins harboring the PDZ domain, such as the MAGUK proteins, happens through their PDZ domain itself and the extreme C-terminal portion of E6 (which has the consensus sequence –x-T/S-x-V/L). One of the major pathways of E6 to deregulate cell cycle and lead to uncontrolled growth is by targeting key proteins, like p53, for ubiquitination and degradation, which half-life decreases from several hours to approximately 20 minutes in the presence of E6 (Howley, 2001). The interaction E6-p53 involves a cellular protein, E6AP (E6 Associated Protein). In fact, the participation of E6AP seems to play an important role by mediating many of the transcriptional changes in

HPV positive lineages (Kelley et al, 2005). Major functions and ligand partners of E6 are reviewed in Howie et al (2009).

Time has come when knowledge on structural organization of HPV oncoproteins, such as E6, is crucial, once it will provide data for fully understanding the actual molecular role of E6 in cellular transformation and carcinogenesis, contributing for the design of new drugs and vaccines. Tridimensional full structure of E6 protein has not been published so far. Many research groups have made several attempts to purify the wild type protein with partial success. Major problems found were: 1) the stability; 2) the formation of clusters and agglomerates of unfolded or misfolded proteins; and 3) the lack of a native fold for the poorly purified mass (Imai, et al 1989; Lechner and Laimins, 1994; Kukimoto et al, 1998; Nominé et al, 2001).

Cole and Danos (1987) have suggested that this protein is in fact the duplication of a protein with only one domain, and as a consequence, both domains of E6 would adopt a similar folding.

Nominé et al (2001, 2006) reported that the developed mutants of the E6 protein (E6 6C/6S and E6C 4C/4S) which were stable and biologically active, which was considered a great success, although, they were just able to purify the C-terminal domain (E6C 4C/4S) in a folded monodispersed manner. They experimentally developed a structural model of this E6 4C/4S by NMR (Nuclear Magnetic Resonance) and published on PDB the 2FK4 entry containing their 10 best models (2006). Only in 2009, Liu et al published their research where they were able to purify enough quantities for biophysical characterization of the the whole protein, as will be discussed later.

For complementing the work of Nominé and colleagues, we developed a complete model made by homology modeling of the wild type E6 protein of HPV16 based on the following presumptions: 1) the protein has two zinc binding domains, each of them binds to one zinc ion; 2) the domains are, at least, roughly symmetrical and complements in a way that the hydrophobic clusters from each domain face each other; 3) the two domains are connected by a flexible linker, and 4) the non conserve mutations C/S used by Nominé and his colleagues, doesn't imply in major conformational changes.



## Results

### *E6Cwt*

We obtained 20 models for the wild type structure of the HPV16 E6 C-terminal domain. Overall, the tertiary structure was maintained, with only some minor disturbances. For example, the H1 helix has two extra residues on the N-terminal portion (Thr86 e Thr87). The analysis made with TOPMATCH showed that, in both structures (model and template) the big loop that contains 2 of the cystein residues involved in zinc coordination, can be recognized as a 2 strand beta sheet (this feature, actually appears more than once, depending on the viewing software). As expected, the extreme C-terminal portion presented a lot of variation due to its exposition and lack of interaction with the rest of the protein.

After the analysis, the 15<sup>th</sup> model was chosen (6E6C\_15). The comparison between the model and the template results can be seen on Table 1. The g-factor, which is the main structure conformity parameter, had its final value unaltered (0,06), with the g-factor of the angles and main-chain length fluctuating around the initial individual value. The bad contacts parameter had better results in the model than in the template, decreasing from 18 to 7. For the other parameters, values were not as good on the model than on the template, but all of them kept in an acceptable range.

Concerning the cystein residues that coordinate the Zn<sup>2+</sup> ion, the distances obtained between the sulfur atom from cysteins ranged around 2,3 Å. The tetrahedral angle had wider variation but maintained a mean of 109,4°.

### *E6Nwt*

The model of E6N was chosen based on various parameters (see Material and Methods – Structure analysis). On Table 1 we can see a summary of the 4E6N\_10 results. The model obtained on this step can be seen in Figure 1. The VMD color set used in these figures, shows two small other colored portions between residues 26 and 29. The Procheck tool defines these regions as isolated beta bridges or hydrogen-bonded turn residues. Other interesting feature that appears on this model and again on the complete structure, is that the Thr22 to Asp25 residues from H1' at the N-terminal edge of this big turn/coil appears in blue, identified as extended helix.

The distances between the sulfur atoms from Cys residues and the zinc ions were kept between 2,29 Å and 2,31 Å, with just the Cys63 with 2,46 Å. The angles Cys-Zn-Cys showed a mean value of  $109,2^\circ \pm 8,5^\circ$ .

### *E6wt*

The final structure of E6 wild type from HPV16 is shown in Figures 2 and 3. The zinc binding domains are diametrically opposed and the cystein residues that coordinate the zinc ion all have Cys-S distances below 2,5 Å, being only Cys63 and Cys66 wider than 2,33 Å.

The N-terminal extreme is heavily hydrophilic, with exception of Met1 and Phe2 residues. This portion of 8 amino acids, excluded on previous models, is probably very flexible, and these 2 hydrophobic initial residues most likely interact with Loop2 (residues Lys68 to Cys80) in some way to satisfy the hydrophobic force.

The two domains are connected by a 13 residue long linker that passes transversally to the beta sheet axis, oriented the same way as the S2 and S3 strands. This loop is mainly composed by neutral and hydrophilic residues, which probably makes it more flexible in this region. Most of the sidechains of the loop residues are facing towards the protein. The interaction between the domains occurs by hydrogen bonds that help maintaining their relative position. Residues involved in this interaction are shown on Figure 4.

The extreme C-terminal is composed of 7 amino acids – TRRETQL (positions 145 to 151). Agreeing with Nominé et al (2006), this portion apparently presents random conformation and is fully exposed. As suggested by Charbonier et al (2008) and Zhang et al (2007), this portion should stay exposed to be able to connect its target and only then it will gain secondary conformation, forming a beta sheet with the PDZ domain of the target.

Besides the portions formerly excluded from the initial modeling in this research and in Nominé's work, the E6 protein also has two other 14 and 13 long turns, respectively Ile26 to Arg39 and Leu99 to Cys111. At the middle of each one of these turns there are two of the four Cys residues that binds to each  $Zn^{2+}$  ion. As appears in the E6N model, depending on the tools used for viewing the structure, these turns appear other colored, meaning they tend to form a secondary/tertiary beta-like structure.

The protein contains a beta sheet made of 6 anti-parallel beta strands, being 3 from each domain. This supersecondary structure is slightly curved reaching the edges (where residues Gly57 and Arg129 are located on N- and C-terminal portions, respectively), back towards the protein directed to where

the helices are, and increasing the curvatures towards the H1/H2 and H1'/H2' helices.

The H1, H2, H1' and H2' are parallel to the axis of the protein, but are perpendicular if seen from the edge of the protein taking the beta sheet as reference. The H3 helix grows apart from the core of the protein towards the edge, where it reaches the loop that connects the H1 to the H2 helix. The same event is observed for the H3' helix located at the opposite extreme of the protein. In each of these small domain terminal helices (H3' comprehends residues 63 to 67 and the H3 residues 136 to 139), we find two cystein residues involved in zinc coordination in each domain.

The loops cited above, contains the other two zinc coordinating cystein residues. As we chose the software to produce the images of the structure's model, a different color set indicates the presence (or tendency) of a different secondary structure. Apparently, it tends to the formation of an antiparallel beta sheet made by two beta strands, each of them made by half of the loop. This observation can be seen in Figures 1 to 3, defined by the other colored loops.

Loop1 and Loop3, which were removed from the first homology modelings and were added back to produce the final E6 wild type model, showed a great diversity of folding possibilities, more or less adequate considering its score on the analysis tools, but mainly compatible. On the other hand, Loop2, also primarily removed, produced only one reasonable folding form alongside the beta sheet.

## **Struture analysis**

### *Motifs and patterns*

Similarity analysis of motifs indicated 70 to 75% similarity between the E6 protein zinc binding domains and the PHD (Plant Homologue Domain) zinc finger motif. We made the alignment of some PHD containing proteins, such as AIRE (Genbank access number: CAA08759.1) and ASH1 (Genbank access number: CAI13211.2) among others and the equivalent parts of E6 (data not shown) and similarity of sequence was approximately 10%.

A total of 16 different patterns and signatures resulted from the 70% cutoff signature similarity search. Many of those are phosphorylation sites of different enzymes with high degree of similarity, which can mean novel function or activity yet unknown for the E6 protein. The signatures returned can be seen on Table 2.

#### *Protein-Protein Interaction Predictor*

The PPI-Pred results returned many regions with propensity to form protein-protein interactions, as shown in Figure 5. The results are given from a 0 to 1 scale – 0% to 100% interaction probability – based on sequence characteristics. The regions with higher probability of interaction are residues 18 to 23, 35 to 40, 75 to 85, 90 to 93 and 135 to 151.

#### **Molecular Dynamics**

A number of parameters were tested (the zinc coordination, the strength of the ligation and the residual charge of the protein). Prototype tests were realized with 2FK4\_03 and kept the topology of the protein and the zinc coordination site relatively intact until the end of the 5.000 ps simulation. The two most perturbed portions throughout the dynamics were the S1 strand

(which supposedly interacts with the other domain) and the H3 helix. The remaining portions of the peptide are kept mainly unaltered, including the zinc binding domain. The definition of the best prototype, as cited above, defined the non-bonded as the best zinc binding model, and that the residual charge of the protein should be neutralized by Cl<sup>-</sup> ions. The mean RMSD value for the SIM001 simulation were  $3,93 \pm 0,23$  Å with a peak value of 4,53 Å.

The protocol chosen was used for simulating both E6C and E6wt. The results obtained for E6C (SIM002), were similar to those obtained for 2FK4\_03. The main topology was maintained, the whole structure slightly relaxed and the zinc ion was kept in place during all process. The mean RMSD value from the template, were lower than for 2FK4\_03 simulation (SIM002 RMSD mean value:  $3,23 \pm 0,26$  Å, peak: 4,03 Å) indicating higher stability on the simulated conditions.

The molecular dynamics, applied to the whole protein, showed very interesting results, as can be observed in Figure 6. During the process, between 6 and 7 ns, the protein separated its two domains, 'opening' the structure but keeping all the topology intact. This movement between the two domains is not only a diversion but also a rotation along the axis of the protein. The hairpins of the beta sheet on each domain are kept during all the process, as are the helices H1, H2, H1' and H2'. On the other hand, both the S1 and S1' and the H3, H3' are, once more, very perturbed.

On the initial homology model of the structure, none of the domains showed the zinc binding domain with tetrahedral geometry. We expected the dynamics would correct the tetrahedral geometry of the zinc binding cysteins, but that did not happen. It could be due to the imprecision of the parameters used in the

protocol, such as the van der Waals radius, energy and others. But in spite of that, at the end of the simulation, the zinc ion is still coordinated by the same four Cys residues on each domain. We calculated the mean Cys-Zn distance for each one of the zinc coordinating Cysteins and the results showed a slight difference of forces of the N- and C-terminal domains in coordinating each zinc (general N-term mean  $2,34 \pm 0,06$  Å; general C-term mean  $2,37 \pm 0,07$  Å).

The RMSD value for the entire structure was high. At the last 242 ps (last data pack), the variation range of the RMSD was 1 Å, increasing more at the beginning and showing a tendency of stabilization at the end. As we analyze the individual portions of the protein, we can see there is a great influence of Loops 1, 2 and 3 (modelled without template and known to have a higher flexibility) on the final RMSD results of the entire protein. A general view of the results is shown in Table 3. The C-term domain, from residue 80 to 144, got a maximum RMSD value of 3,46 Å, while the N-term domain, without the 8 initial residues (from residues 9 to 71) had a maximum RMSD value of 3,1. A representation the protein through the simulation with the initial, final and 3 intermediate times of the trajectory is shown in Figure 6. The RMSD from time zero until the end of the simulation of the backbone of each portion and the entire protein can be seen in Figure 7.

The Procheck analysis from the energy minimized steps extracted from the E6wt (SIMNC001 – 5CHI\_03) simulation, returned 100% of the residues on Allowed (68,1%) and Additionally Allowed (39,1%) regions at the end of the 9,242 ns long simulation. The overall G-factor values of all steps were above -0,5, with -0,29 on the last step.

When we compared SIM001 (2FK4\_03), SIM002 (6E6C\_15) and C-terminal domain of SIMNC001 (5CHI\_03 - residues 80-144), an increasing stability on simulated conditions shown up as the resemblance with the wild type complete structure increased. Can be observed on Figure 8, that the C-terminal domain shows stabilization at an earlier point for the truncated polypeptides (2FK4 and 6E6C) compared to E6wt, since both SIM001 (2FK4\_03) and SIM002 (6E6C\_15) show faster increase and subsequent stabilization (approximately at 1,5 to 2 ns) than the model of the complete structure (around 4 to 4,5 ns). Albeit the SIMNC001 showed this delay on the stabilization, it is noticeable that its RMSD value at any point, surpasses the SIM001 RMSD.

## **Material and Methods**

### *Input data*

The sequence of the wild type E6 protein from HPV16 was obtained from GenBank, access number NP\_041325.1. The protein can be 151 or 158 amino acids long, depending on which Methionin is picked as the start residue. To do the modeling procedures in this work, the 151 aminoacid long protein was used.

The E6C 4C/4S structure used as a template for several procedures, was obtained at PDB, access number 2FK4. To choose which of the 10 NMR models of 2FK4 to use, PROCHECK\_NMR were used to calculate the mean structure. The NMR model that was more similar to the mean structure and that had better Procheck scores was used.

### *Homology modeling*



All the homology modeling of this work was made using MODELLER 9 v 4. Some manual modeling to adjust a few sidechains positions were made with Swiss PDB Viewer and later with aqueous solution simulation with LEAP tool in AMBER 9 (see Molecular Dynamics topic). The best models were chosen analyzing the scores from different sets of tools, depending what was best suited for each modeling (see Analysis Structure).

#### E6Cwt and E6Nwt

We used the Modeller 9 version 4 to do the homology modeling of both E6C and E6N wild type. Both were made using the experimental NMR model 3 of the 2FK4 E6C 4C/4S as template (2FK4\_03). The zinc ion was included as an HETATM group to keep the Cysteins residues involved in zinc binding into place. The alignments used to do the modeling can be seen in Figures 9 (A and B). All the excluded parts on Nominé work remained excluded at this point. The sequence identity between the E6C 4C/4S and the E6Nwt is 15,4% and the similarity (strong + weak) is 35,4%.

#### E6 wild type modeling

As the best E6N and E6C models were chosen, the two structures were put together to form a chimera structure. They were oriented to form an antiparallel beta sheet connected by S1 and S1' of each domain and to avoid stearical incompatibilities between them, as was proposed by Nominé et al. The modeling included, now, the formerly missing portions. Again, the two zinc ions were included as HETATM groups. The alignment used for this modeling can be seen if Figure 9 C and was made through ClustalW (PBIL) using the Gonnet

matrix with gap opening penalty of 10 and extension penalty of 0,1. The best model was chosen by its RMS deviation from the template (chimera structure) and the newly added loops (Loop1: residues 1 to 8; Loop2: 72 to 79 and Loop3: 145 to 151) were refined with Modeller, one at a time, as each step was evaluated by Prosa-Web and Procheck.

### *Structure analysis*

All the structure analysis were made using a set of the following tools: 1) Procheck (Ramachandran Plot, G-factor, bad contacts and others); 2) Verify 3D (upper and lower score); 3) PROSA-Web (Z-score and distribution); 4) Topmatch (RMSD from the template, alignment).

Other sequence analysis were carried out by Protein-protein interaction predictor (PPI-Pred) and PROSITE (with a 70% cut-off).

### *Molecular Dynamics*

With the intention to test the chosen structures in more realistic conditions, we have challenged the models with simulated aqueous environment using the SANDER module of AMBER 9. Three protocols and a number of parameter changes were tested by using the default available by Weiner et al (1984; 1986) or Cornell et al (1995). Two of the protocols acted with residual plasma and the other with Cl<sup>-</sup> ions to neutralize the positive charge. To treat the zinc ions, the non-bonded model was used with a formal charge of +2. The Lennard-Jones parameters of the Zn<sup>2+</sup> ions were adapted from Stote and Karplus (1995), such as energy  $\epsilon$  (*Épsilon* – from 0,0125 to

0,025) and the radius of the equilibrium point of VDW forces (from 1,1 to 2,2 Ångstroms).

On Table 4 there is a summary of the simulation systems. For all the systems, the 500 ps were used for minimization and equilibration of the molecule and are not included on the final time count of the simulation. At this point, TIP3P water molecules were added (Jorgensen et al, 1983) and the orthorhombic box dimensions determined.

The simulations were made in an NTP ensemble, in 500 ps packs being the first 200 reserved for heating and stabilization of the complex. The physical /chemical conditions of the simulations were 298,16 K (25°C), at 1 atm, pH 6,8 and with a timepass of 2 fs (femtoseconds). The protocol used in the homology modeled structures was chosen based on the 2FK4\_03 molecular dynamic results. The simulation of the complete E6 structure lasted 9,242 ns. The final protocol used with E6C and E6wt used radius = 1,5 Å e  $\epsilon = 0,025$ . For the E6wt simulation 6 Cl<sup>-</sup> ions were used to neutralize the residual charge.

The analysis made on the simulation output was done using VMD, AMBER and PROCHECK. For the Procheck analysis 11 timesteps were extracted from the trajectory (0, 1001, 2001, 3001, 4001, 5001, 6001, 7001, 8001, 9001 and 9242 ps) and later minimized to reduce incompatibilities and distortions.

## **Discussion**

The interaction between HPVs E6 and cellular proteins is still a very wide and complex research field. Its promiscuity and variety of functions, not always conserved between the more than 100 type of human Papillomavirus already

described, makes this protein family a scientific puzzle. Nominé et al (2006) proposed an experimental model for the 3D structure of the C-terminal domain. It was the first successful attempt to achieve a model of this protein. Here we aggregate information on the 3D structure HPV 16 E6 protein, proposing a model for the complete E6 wild type protein (see Figures 1 and 2).

Although many mutational studies have been made in order to understand and try to localize the portions of E6 that interacts with a number of its ligands (see Table 4), little is known about the residues actually involved in such interactions.

On our work (as also suggested by Nominé et al, 2006) the model shows the complete protein composed by 6 helices and a 6 strand beta sheet, 3 of each from the C-terminal and 3 from the N-terminal domain. The two domains appear to be independent from each other and show a flexible interface region, which allow fluctuations on the orientation between them. This characteristic is reinforced by the fact that the outer hairpins from the beta sheet kept their structure during all the molecular dynamics but the two inner strands (S1 and S1') were not so stable.

Nominé et al (2005) suggested that along with the two inner beta strands, two loops (the first between S1' and H1' and the other between S1 and H1 – see Figure 7) would be involved in the interface between the two domains. The conserved residues Tyr54 and Leu88 are believed to participate on the interface (Nominé et al, 2005). On our model, we have found that the sidechain of Tyr54 interacts with the sidechains of Leu88 and Gln91, and the backbone of Leu88 interacts with the sidechain of Asp56. This happens on the homology modeling procedure, but we suggest that this interaction is also transient, as shown in the

molecular dynamic simulation. Probably depending on environment conditions (salinity, pH) this interaction would form or break in order to allow the protein to separate the domains which are expected to have some independent functional abilities.

Tao et al (2003) described 3 Nuclear Localization Signals (NLS) – from residues Phe2 to Glu7, from Lys65 to Lys72 and from Lys115 to Arg124. The first is part of the Loop1 (excluded on the template), the second is located at the end of the H3' helix and the Loop2 (also excluded on the template) and the last one is located between H2 and L2. Our model shows that all these portions are all on the same side of the protein in accessible areas. In exception of 4 residues, all others are located in loops (unstructured regions). Point mutations or deletions on these regions leads to inhibition of p53 binding and degradation, eg.  $\Delta$ Val2-Leu12 (Foster et al, 1994) or Phe2Val (Liu et al, 1999). Masson et al (2003) have shown that a monoclonal antibody that binds to Gln3-Arg8 epitope also inhibits p53 degradation. Foster et al (1994), Park et al (2002) and Dalal et al (1996) demonstrated a significant decrease of E6 binding to p53 and E6AP and the degradation by mutant  $\Delta$ His118-Lys122. On the table of Appendix A it is shown a review of analysis of mutants (point mutations and deletions), there we can see that most of the mutants that have the NLS regions affected lose the p53 binding or degradation capacity.

The biggest difficulty on determining which residues are involved on each function may probably rise from the fact that this protein shows big propensity to disorder. There are many candidate residues to be involved p53 binding and degradation. Oncoproteins and signaling proteins that share this disorder characteristic, often becomes ordered only when bound the target. The extreme

8 residue long extreme of the E6 protein from HPV 18 is a good example of that, since it is a highly flexible turn on the protein model but when bound to the PDZ domain of MAGI or SAP97/Dlg it acquires a beta strand structure. If a deletion is made, for example, it may affect the rest of the protein. As a mainly disordered protein, the “rearrangements” on the tertiary structure may become infeasible for the protein to bind its target, even indirectly.

We have challenged this structure with a simulation in aqueous solution and found that the N-terminal domain is slightly more stable than the C-terminal domain, being the final and the maximum RMSD for the backbone 0,07 Å and 0,36 Å wider in the C-terminal domain, respectively. The N-terminal domain also stabilizes faster on all parameters (All atoms, backbone, C-alpha – data not shown). In contrast, Liu et al (2009) have shown results of thermal stability experiments monitored by circular dichroism where the C-terminal domain has shown higher thermal stability. They also made zinc content experiments where the N-terminal domains showed lower ability to bind zinc ions in comparison to the C-terminal domain with and without chelating agents. Compiling the published results and the simulation we made, we suggest that in spite the C-terminal may have higher binding ability to  $Zn^{2+}$  and high thermal stability, N-terminal domain is more stable under normal conditions (aqueous simulation) due to compensation characteristics, such as its residue content (it contains higher amount of apolar residues).

Our simulation results also suggest a lot of flexibility inter and intra-domain, since it has clearly shown the domain separation between N- and C-term, which are kept together just by the linker. Agreeing with previous works, the domains appear to have independent functions (Zhang et al, 2007; Ristriani et al, 2000,

2001; Lagrange et al, 2005). Besides all the wide diversion of the structure model at the end of the simulation, the zinc ion is coordinated by the cysteins sidechains with compatible distances, showing that the non bonded is a good zinc model for the E6 protein simulation.

The difference of stability in computational simulated conditions found between 2FK4\_03 (NMR model used as template) and both 6E6C\_15 and the C-terminal domain of 5CHI\_03 (homology modeled structures) shows that: 1) the homology modeling and molecular dynamics protocols are accurate enough for this first attempt of modeling the complete structure of E6; 2) the wild type truncated protein seems to be more stable than the template and the complete model of the wild type structure is more stable than both previous cited ones, hence, a model containing the complete sequence of the protein must be more accurate than a model that contains just a part of the protein or just one of its domains.

Ristriani et al described a novel DNA binding function for E6 HR proteins after finding some similarities between them and Phage T4 Endonuclease VII. He found that only the C-terminal domain binds to Holiday Junctions in a monomeric way (Ristriani et al, 2000/2001). As the actual purpose of DNA binding by E6 is still unknown and we have shown that in our predicted model and simulation, the E6 protein is able to separate their N-and C-terminal domains to act independently, we suggest that E6N acts as a recruiting agent, forming a bigger protein/enzymatic complex. We have also results that show regions with high probability of interacting with other proteins on E6N domain that could be somehow involved in this new feature to be studied. Also soothing

to be noticed is that the higher instability / disorder of E6C than E6N is compatible with E6C DNA binding ability, as reviewed by Xie et al (2007).



## **Conclusion**

Our model is based on an experimental published polypeptide, consisting of the E6C terminal domain with 4 non conserved Cys to Ser mutations. The consistence of our predicted model with previous works shows that this is good and can be used to help finding new drugs and vaccines to fight HPV. In addition new aqueous dynamic simulation results we show in this work brings new features for the studies of the functions of the E6 proteins.

The work represents some steps towards a reliable 3D structure model of the complete E6 protein from HPVs. However, many improvements need to be made on the modeling and on the molecular dynamics process for this goal to be achieved with higher accuracy. Even though the structure suggested by this work could help enlighten some of the many pending questions so far.

## References

1. Bosch, F.X., Lorincz, A., Meijer, C.J.L.M. and Shah, K.V. The causal relation between human papillomavirus and cervical cancer. *J. Clin. Pathol.* 2002, 55, 244-265.
2. Howley, P. *Papillomavirinae: the viruses and their replication*. In: *Fields Virology*, Fields, B. and Knipe, D., Eds., 4th ed., Raven Press, New York, EUA, 2001, digital version.
3. Schmitt, V.M. Papilomavírus Humano. In: *Doenças infecciosas: diagnóstico molecular*, Rosseti, M. L., da Silva, C.M.D., and Rodrigues, J.J.S., Eds., Rio de Janeiro, Guababara-Koogan, 2006, p.135-145.
4. Hubregtse, J.M., Scheffner, M. and Howley, P.M. Localization of the E6 AP regions that direct HPV E6 binding, association with p53 and ubiquitination of associated proteins. *Mol. Cell. Biol.*, 1993, 13(8), 4918-4927.
5. Zimmermann, H., Degenkolbe, R., Bernard, H.-U. and O'connor, M.J. The HPB 16 E6 oncoprotein can downregulate p53 activity by targeting the transcriptional coactivator CBP/p300. *J. Virol.*, 1999, 73(8), 6209-6219.
6. Patel, D., Huang, S.M., Baglia, L.A. and McCance, D.J. The E6 protein of HPV 16 binds to and inhibits co-activation by CBP/p300. *EMBO J.*, 1999, 18(18), 5061-5072.
7. Veldman, T., Horikawa, I., Barrett, J.C. and Schlegel, R. Transcriptional Activation of the Telomerase hTERT Gene by Human Papillomavirus Type 16 E6 Oncoprotein. *J. Virol.*, 2001, 75, 4467-4472.
8. Thomas, M. and Banks, L. Human papillomavirus (HPV) E6 interactions with Bak are conserved amongst E6 proteins from high and low risk HPV types. *J. Gen. Virol.*, 1999, 80, 1513–1517.
9. Jackson, S., Harwood, C., Thomas, M., Banks, L. and Storey, A. Role of Bak in UV-induced apoptosis in skin cancer and abrogation by HPV E6 proteins. *Genes & Dev.*, 2000, 14, 3065-3073.
10. Kelley, M.L., Keiger, K.E., Lee, C.J. and Huihregtse, J.M. The global transcriptional effects of the HPV E6 protein in cervical carcinoma cell lines are mediated by E6AP ubiquitin ligase. *J. Virol.*, 2005, 79 (6), 3737-3747.
11. Howie, H.L., Katwenellenbogen, R.A. and Galloway, D.A. Papillomavirus E6 proteins. *Virology*, 2009, 384, 324-334.

12. Imai, Y., Tsunokawa, Y., Sugimura, T. and Terada, M. Purification and DNA-binding properties of human papillomavirus type 16 E6 protein expressed in *Escherichia coli*. *Biochem. Biophys. Res. Commun.*, 1989, 164(3), 1402-1410.
13. Lechner, M.S. and Laimins, L.A. Inhibition of p53 DNA binding by human papillomavirus E6 proteins. *J. Virol.*, 1994, 68, 4262-4273.
14. Kukimoto, I., Aihara, S., Yoshiike, K. and Kanda, T. Human papillomavirus oncoprotein E6 binds to the C-terminal region of human minichromosome maintenance 7 protein. *Biochem. Biophys. Res. Commun.*, 1998; 249(1), 258-262.
15. Nominé, Y., Ristriani, T., Laurent, C., Lefèvre, J.-F., Weiss, E. and Travé, G. A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein. *Protein Eng.*, 2001, 14(4), 297-305.
16. Cole, S.T. and Danos, O. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. *J. Mol. Biol.*, 1987, 193, 599-608.
17. Nominé, Y., Masson, M., Charbonnier, S., Zanier, K., Ristriani, T., Deryckere, F., Sibler, A.P., Desplancq, D., Atkinson, R.A., Weiss, E., Orfanoudakis, G., Kieffer, B. and Travé, G. Structural and functional analysis of E6 oncoprotein: Insights in the molecular pathways of human Papillomavirus-mediated pathogenesis. *Mol. Cell.*, 2006, 21, 665-678.
18. Liu, Y., Cherry, J.J., Dineen, J.V., Androphy, E.J. and Baleja, J.D. Determinants of Stability for the E6 Protein of Papillomavirus Type 16. *J. Mol. Biol.*, 2009, 386, 1123-1137.
19. Charbonnier, S., Stier, G., Orfanoudakis, G., Kieffer, B., Atkinson, R.A., and Trave, G. Defining the minimal interacting regions of the tight junction protein MAGI-1 and HPV16 E6 oncoprotein for solution structure studies. *Protein Expr. Purif.*, 2008, 60(1), 64-73.
20. Zhang, Y., Dasgupta, J., Ma, R.Z., Banks, L., Thomas, M. and Chen, X.S. Structures of a Human Papillomavirus (HPV) E6 polypeptide bound to MAGUK

- proteins: mechanisms of targeting tumor suppressors by a high-risk HPV protein. *J. Virol.*, 2007, 81(7), 3618-3626.
21. CLUSTAL W: Thompson J.D., et al. *Nucleic Acids Res.*, 1994, 22(22), 4673-4680.
  22. MODELLER: Eswar, N. et al. *Current Protocols in Bioinformatics*, John Wiley & Sons, Inc., Supplement 15, 5.6.1-5.6.30, 2000.
  23. MODELLER: Marti-Renom, M.A. et al. *Annu. Rev. Biophys. Biomol. Struct.*, 2000, 29, 291-325.
  24. Loop Modeling (MODELLER): Fiser, A. et al. *Protein Science*, 2000, 9, 1753-1773.
  25. SPDBv: Guex, N. and Peitsch, M.C. *Electrophoresis*, 1997, 18, 2714-2723.
  26. PROCHECK: Laskowski, R.A. et al. *J Appl Cryst*, 1993, 26, 283-291.
  27. PROCHECK: Morris, A.L. et al. *Proteins*, 1992, 12, 345-364.
  28. Verify3D: Bowie, J.U., et al. *Science*, 1991, 253(5016), 164-170.
  29. Verify3D: Lüthy, R. et al. *Nature*. 1992, 356(6364), 83-85.
  30. PROSA-WEB/TOPMATCH: Sippl, M.J. *Proteins*, 1993, 17, 355-362.
  31. PROSA-WEB: Wiederstein, M. e Sippl, M.J. *Nucleic Acids Res.*, 2007, 35, W407-W410.
  32. PROSITE: Hulo, N. et al. *Nucleic Acids Res.*, 2008, 36, D245–D249.
  33. PPI-Pred: Bradford, J.R. and Westhead, D.R. *Bioinformatics*, 2005, 21, 1487-1494.
  34. AMBER: Case, D.A. et al. University of California, San Francisco, 2008.
  35. VMD: Humphrey, W., Dalke, A. and Schulten, K. 'VMD – Visual Molecular Dynamics', *J Molec Graphics* 1996 14.1, 33-38.
  36. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C. Alagona, G., Profeta Jr.S. and Weiner, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. *J. Am. Chem. Soc.*, 1984, 106, 765-784.
  37. Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. *J. Comp. Chem.*, 1986, 7,230-252.

38. Cornell, W.D. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 1995, 117, 5179–5197.
39. Stote, R.H. e Karplus, M. Zinc Binding in Proteins and Solution: A Simple but Accurate Nonbonded Representation. *Proteins: Struct. Func. Gen.*, 1995, 23: 12-31.
40. Jorgensen, W., Chandrasekar, J., Madura, J., Impey, R. e Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 1983, 79, 926-935.
41. Nominé, Y., Charbonnier, S., Miguet, L., Potier, N., Van Dorsselaer, A., Atkinson, R.A., Trave, G. and Kieffer, B.  $^1\text{H}$  and  $^{15}\text{N}$  resonance assignment, secondary structure and dynamic behavior of the C-terminal domain of human papillomavirus oncoprotein E6. *J. Biomol. NMR*, 2005, 31, 129-141.
42. Tao, M., Kruhlak, M., Xia, S., Androphy, E. and Zheng, Z.-M. Signals that dictate nuclear localization of HPV 16 oncoprotein E6 in living cells. *J. Virol.*, 2003, 77(24), 13232-13247.
43. Foster, S.A., Demers, G.W., Etscheid, B.G. and Galloway, D.A. The ability of HPV 16 E6 proteins to target p53 for degradation *in vivo* correlates with their ability to abrogate actinomycin D-induced growth-arrest. *J. Virol.*, 1994, 68(9), 5698-5705.
44. Liu, Y., Chen, J.J., Gao, Q., Dalal, S., Hong, Y., Mansur, L.P., Band, V. and Androphy, E.J. Multiple Functions of Human Papillomavirus Type 16 E6 Contribute to the Immortalization of Mammary Epithelial Cells *J. Virol.*, 1999, 73(9), 7297–7307.
45. Masson, M., Hindelang, C., Sibling, A.-P., Schwalbach, G., Travé, G. and Weiss, E. Preferential nuclear localization of the human papillomavirus type 16 E6 oncoprotein in cervical carcinoma cells. *J. Gen. Virol.*, 2003, 84, 2099–2104.
46. Park, R B E Androphy, E J. Genetic Analysis of High-Risk E6 in Episomal Maintenance of Human Papillomavirus Genomes in Primary Human Keratinocytes. *J Virol*, v.76, p.11359-11364, 2002.
47. Dalal, S., Gao, Q., Androphy, E.J. and Band, V. Mutational analysis of human papillomavirus type 16 E6 demonstrates that p53 degradation is

- necessary for mammary epithelial cell immortalization. *J. Virol.*, 1996, 70, 683–688.
48. Ristriani, T., Masson, M., Nomine, Y., Laurent, C., Lefevre, J.F., Weiss, E. and Trave, G. HPV oncoprotein E6 is a structure-dependent DNA-binding protein that recognizes four-way junctions. *J. Mol. Biol.*, 2000, 296, 1189-1203.
  49. Lagrange, M., Charbonnier, S., Orfanoudakis, G., Robinson, P., Zanier, K., Masson, M., Lutz, Y., Trave, G., Weiss, E. and Deryckere, F. Binding of human papillomavirus 16 E6 to p53 and E6AP is impaired by monoclonal antibodies directed against the second zinc-binding domain of E6. *J. Gen. Virol.*, 2005, 86, 1001-1007.
  50. Ristriani, T., Nomine, Y., Masson, M., Weiss, E. and Trave, G. Specific recognition of four-way DNA junctions by the C-terminal zinc-binding domain of HPV oncoprotein E6. *J. Mol. Biol.*, 2001, 305(4), 729-739.
  51. Xie, H., Vucetic, S., Iakoucheva, L.M., Oldfield, C.J., Dunker, A.K., Uversky, V.N. and Obradovic, Z. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. *J. Proteome. Res.*, 2007, 6(5), 1882-1898.
  52. Gewin, L., Myers, H., Kiyono, T. and Galloway, D.A. Identification of a novel telomerase repressor that interacts with the human papillomavirus type-16 E6/E6-AP complex. *Genes & Dev.*, 2004, 18, 2269-2282.
  53. Gewin, L. and Galloway, D.A. E Box-Dependent Activation of Telomerase by Human Papillomavirus Type 16 E6 Does Not Require Induction of c-myc. *J. Virol.*, 2001, 75, 7198-7201.
  54. Klingelhutz, A.J., Foster, S.A. and McDougall, J.K. Telomerase activation by the E6 gene product of human papillomavirus type 16. *Nature*, 1996, 380(6569), 79-82.
  55. Elbel, M., Carl, S., Spaderna, S. and Iftner, T. A comparative analysis of the interactions of the E6 proteins from cutaneous and genital papillomaviruses with p53 and E6AP in correlation to their transforming potential. *Virology*, 1997, 239(1), 132-149.
  56. Sekaric, P., Cherry, J.J. and Androphy, E.J. Binding of Human Papillomavirus Type 16 E6 to E6AP Is Not Required for Activation of hTERT. *J. Virol.*, 2008, 82, 71-76.

57. Zanier, K., Charbonnier, S., Baltzinger, M., Nomine, Y., Altschuh, D. and Trave, G. Kinetic analysis of the interactions of human papillomavirus E6 oncoproteins with the ubiquitin ligase E6AP using surface plasmon resonance. *J. Mol. Biol.*, 2005, 349(2), 401-412.
58. Ristriani, T., Fournane, S., Orfanoudakis, G., Trave, G. and Masson, M. A single-codon mutation converts HPV16 E6 oncoprotein into a potential tumor suppressor, which induces p53-dependent senescence of HPV-positive HeLa cervical cancer cells. *Oncogene*, 2009, 28(5), 762-772.
59. Shamanin, V.A., Sekaric, P. and Androphy, E.J. hAda3 Degradation by Papillomavirus Type 16 E6 Correlates with Abrogation of the p14ARF-p53 Pathway and Efficient immortalization of Human Mammary Epithelial Cells. *J. Virol.*, 82, 3912-3920.
60. Kiyono, T., Hiraiwa, A., Fujita, M., Hayashi, A. Binding of high-risk human papillomavirus E6 oncoproteins to the human homologue discs large tumor suppressor protein. *PNAS USA*, 94, 11612-11616.
61. Le Roux, L.G. and Moroianu, J. Nuclear Entry of High-Risk Human Papillomavirus Type 16 E6 Oncoprotein Occurs via Several Pathways. *J. Virol.*, 2003, 77, 2330-2337.

Figures

Fig1.tif

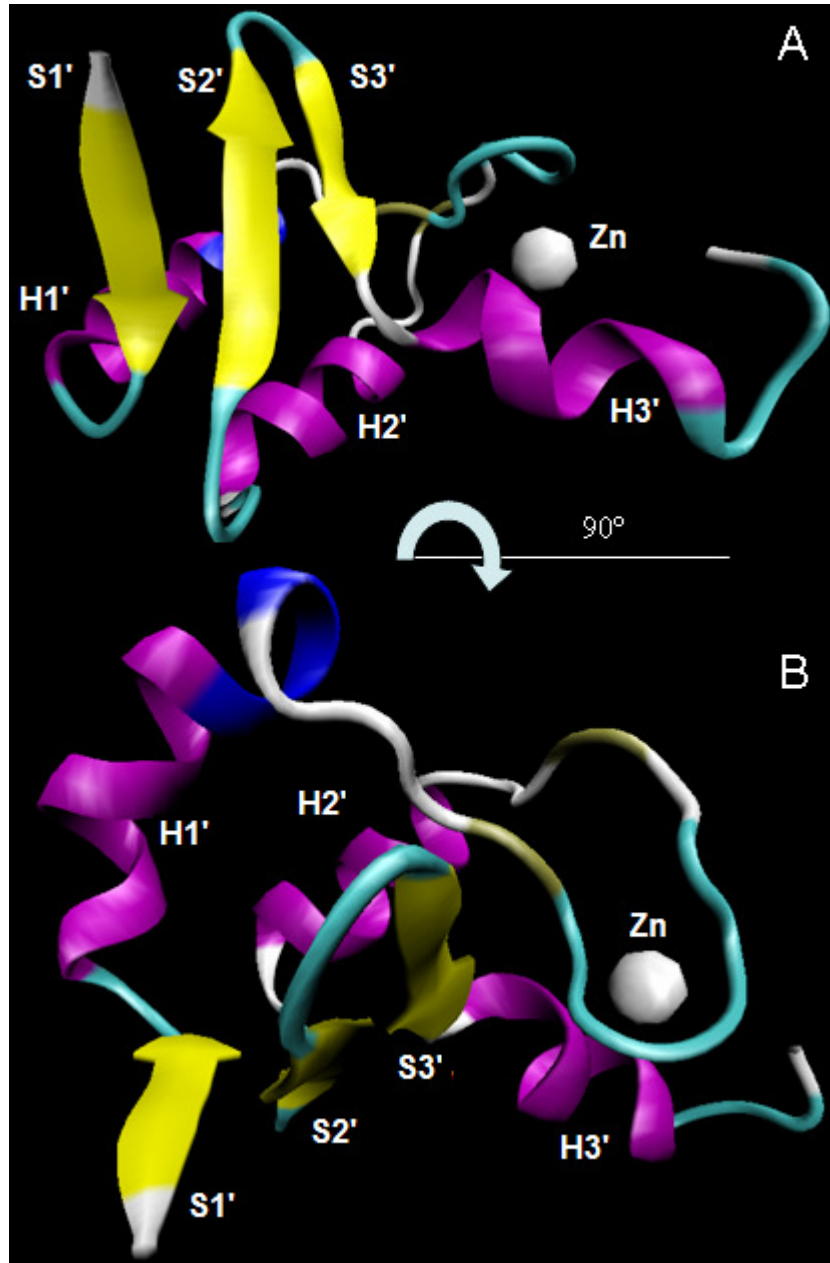




Fig2.tif

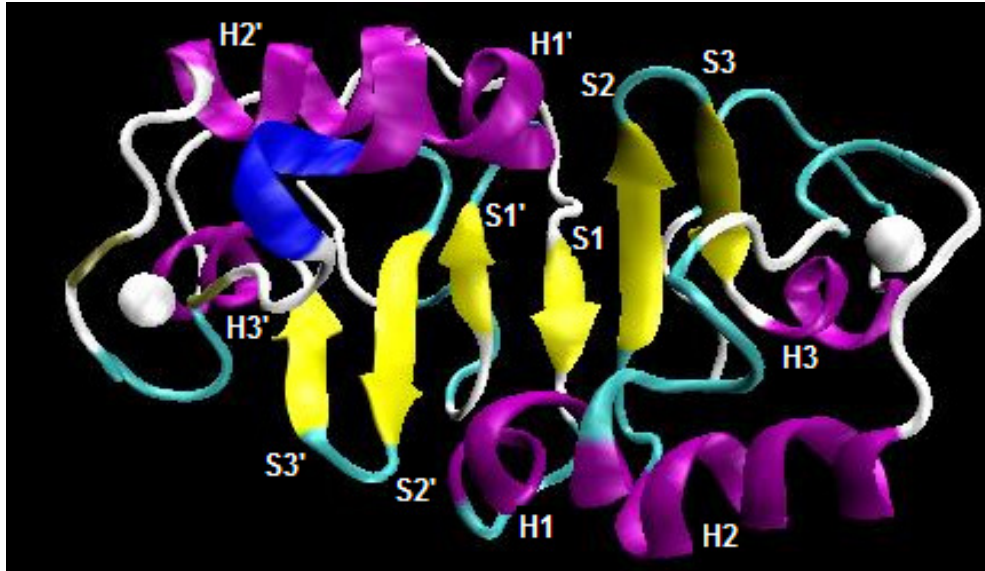


Fig3.tif

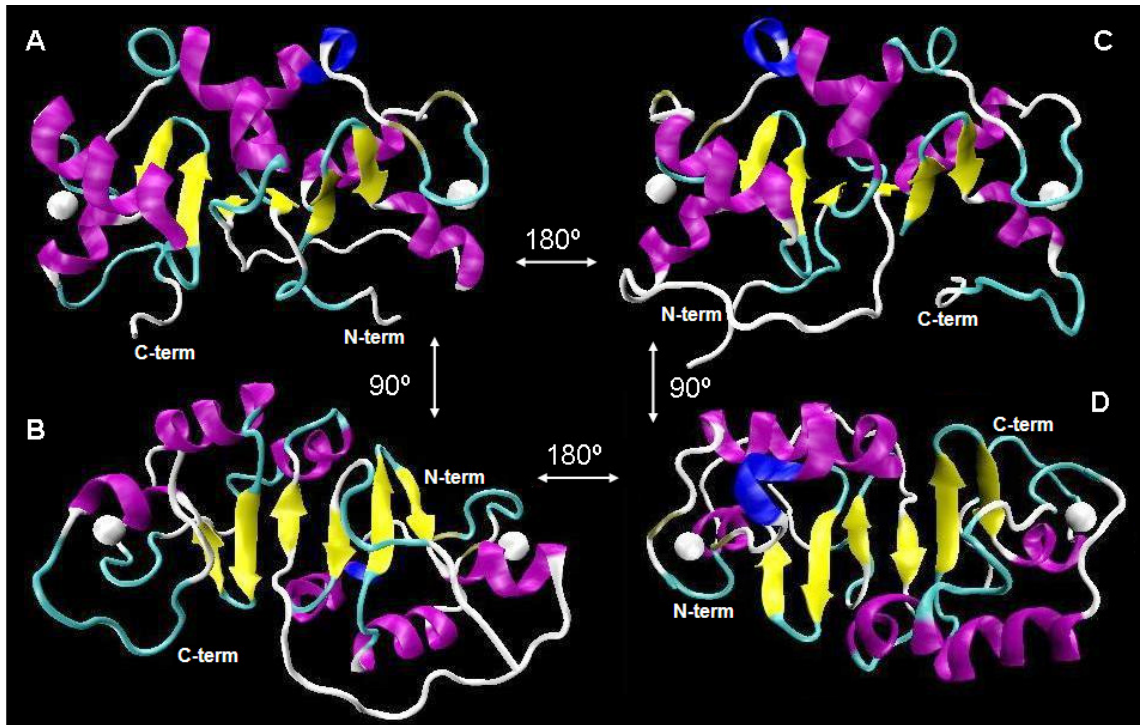


Fig4.tif

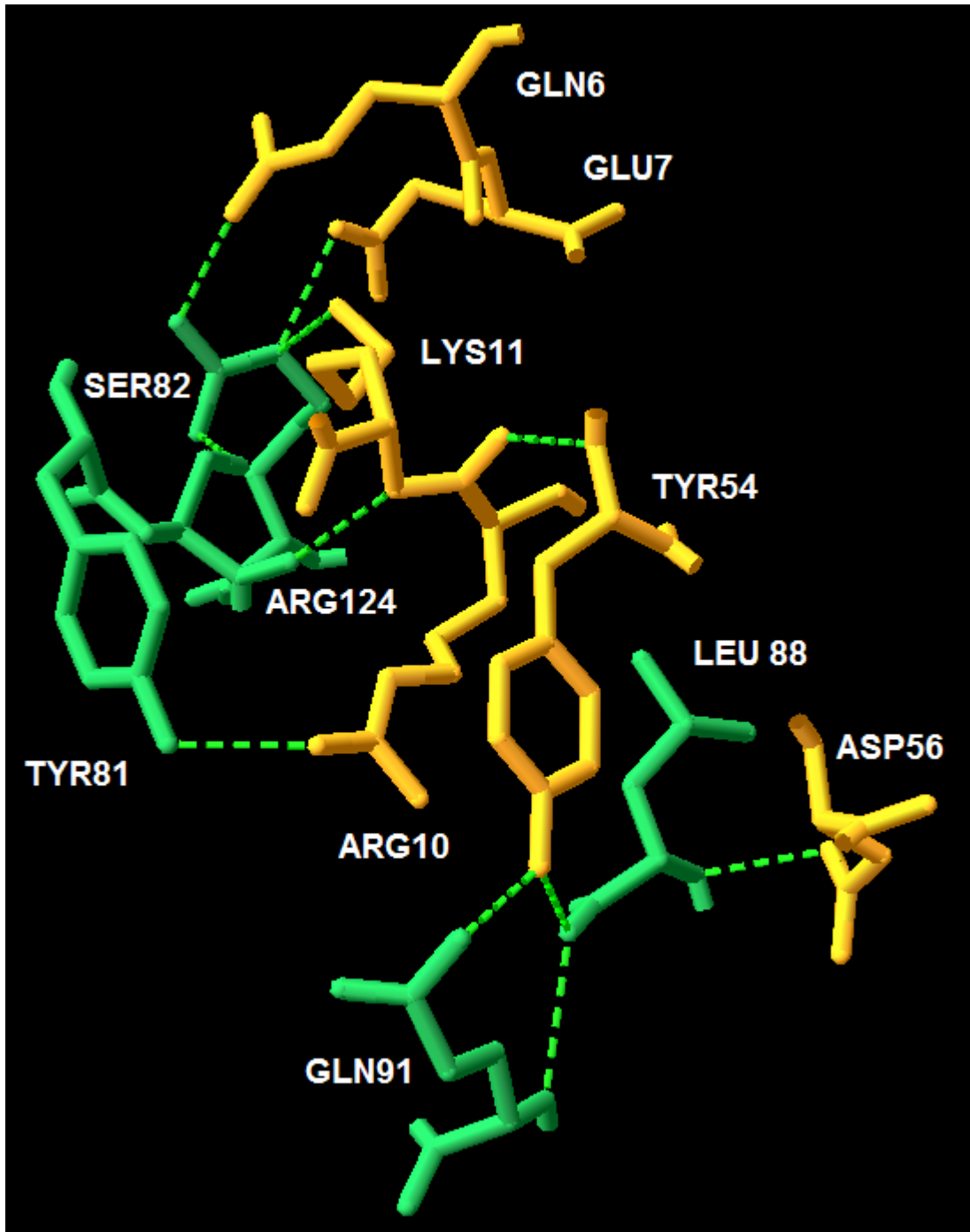


Fig5.tif

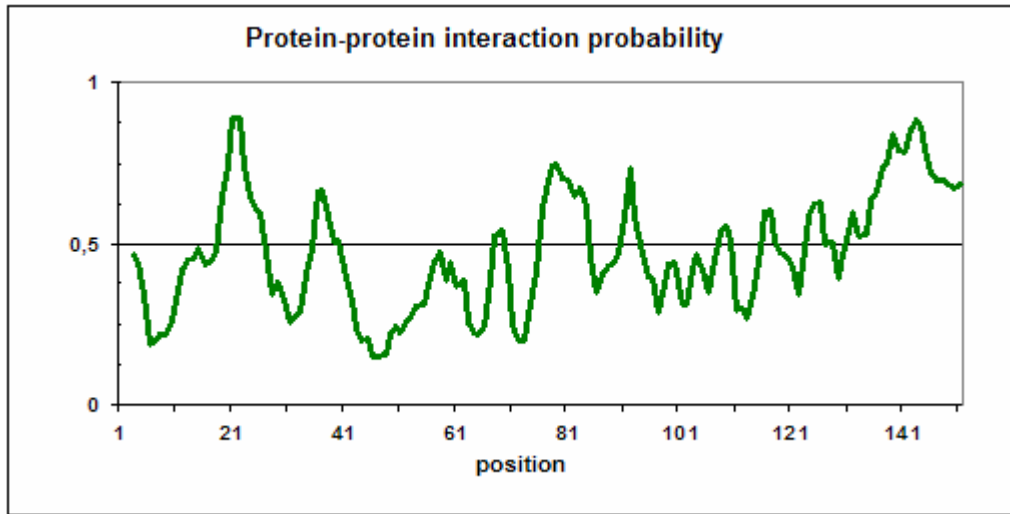


Fig6.tif

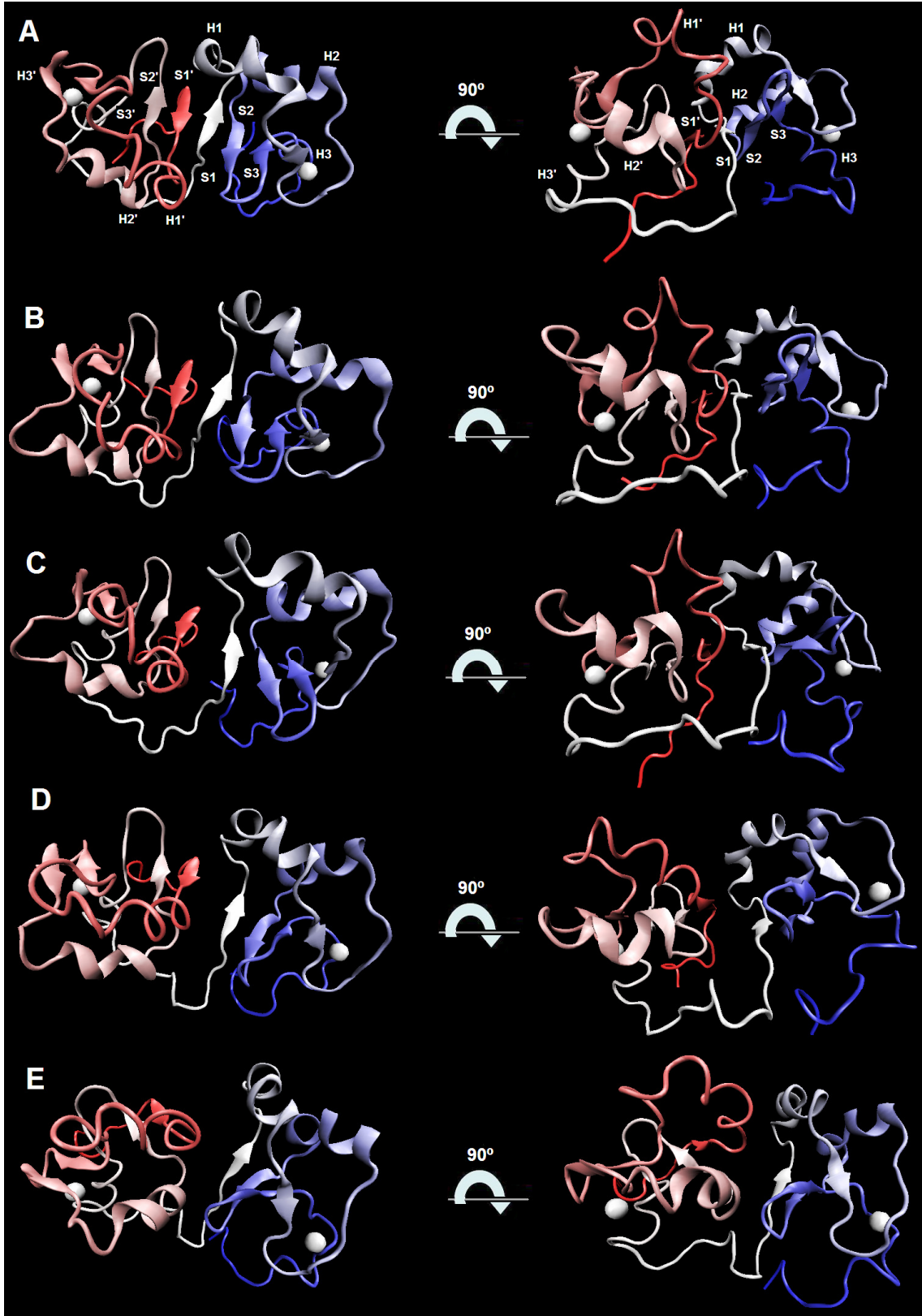


Fig7.tif

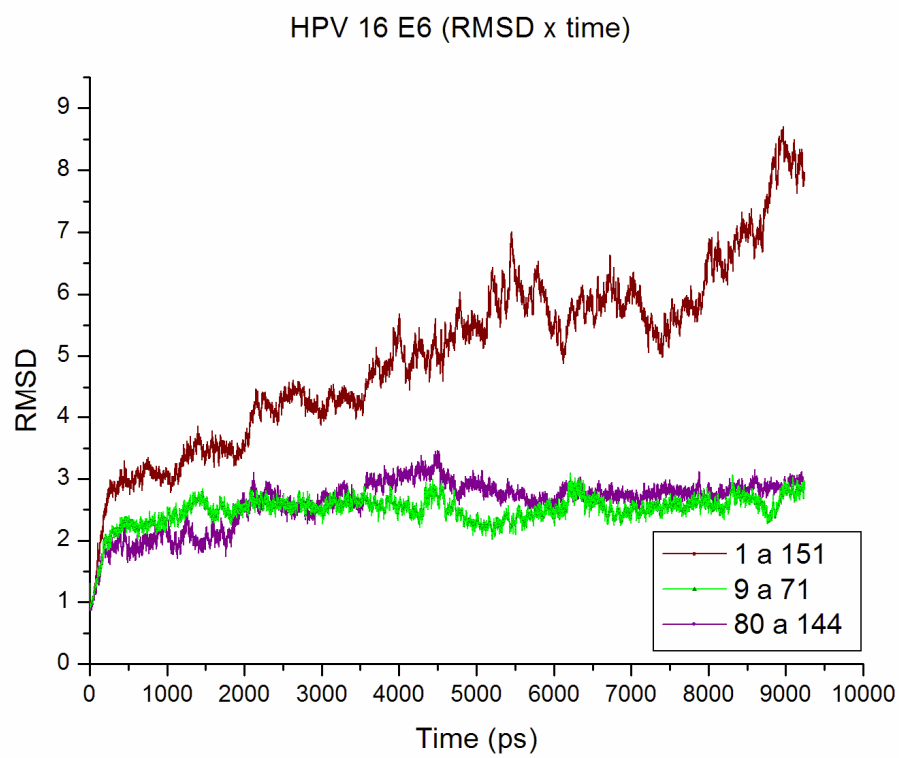
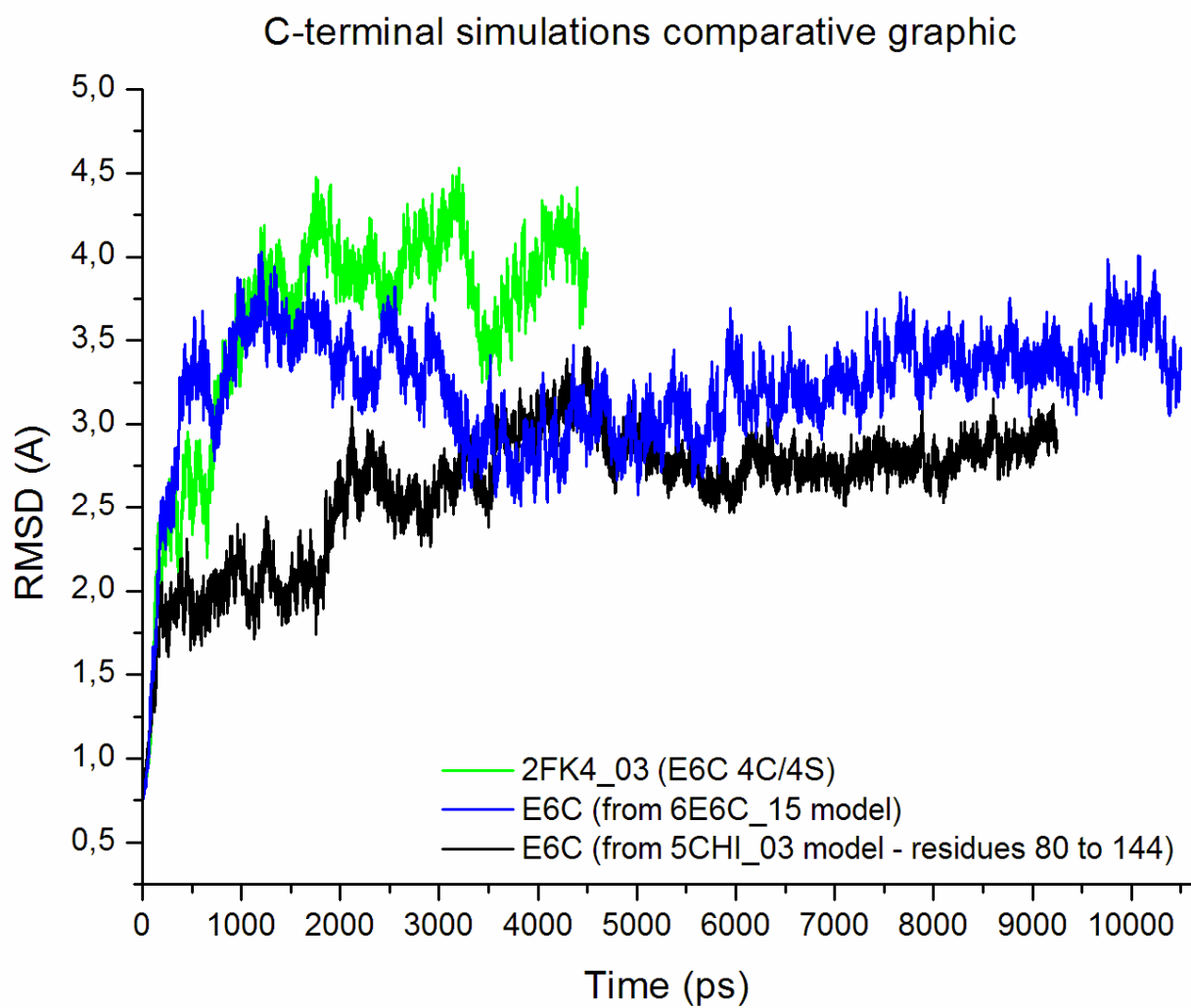


Fig8.tif







## Legends

### Fig1.tif

Figure 1: Ribbon representation of the E6N structure model made by homology modeling. Figures A and B represents two different angles of the same model with 90° rotation on X axis. Yellow: beta strands; purple: alpha helix; white and cyan: loops and turns; blue: extended helix; ocher: residue isolated beta bridge / hydrogen-bonded turn. The second figure (B) shows with better details the ocher and blue coloured regions. The white sphere represents the zinc ion.

### Fig2.tif

Figure 2: Ribbon representation of the model of the complete E6 wild type protein structure. H1' to H3' and S1' to S3' are on the N-terminal domain and S1 to S3 and H1 to H3 on C-terminal.

### Fig3.tif

Figure 3: Different views of the E6wt complete structure model. Each view indicates the position of the N- and C-terminal ending. The arrows in the middle indicate the angle and axis rotation. The white spheres represents the zinc ion.

### Fig4.tif

Figure 4: Sticks representation of the residues that forms the interface between the N- (yellow) and C-terminal (green) domains of HPV 16 E6. the dotted lines represent the hydrogen bonds.

### Fig5.tif

Figure 5: PPI-Pred results represented with a 3 aminoacid window. The Y axis is the interaction probability, being 1 the most probable and 0 the least.

### Fig6.tif

Figure 6: New cartoon representation of 5 points of the dynamic trajectory. A: 0 ps; B: 2146 ps; C: 4556 ps; D: 7226 ps; E: 9242 ps. The figures are colored by

index being red N-terminal and blue C-terminal. The white spheres represent the zinc ion. The second column is a 90° rotated view of the first column.

Fig7.tif

Figure 7: Backbone RMSD values (in Å) of the entire protein (1 to 151 – maroon), the N-terminal domain (9 to 71 – green) and the C-terminal domain (80 to 144 – purple).

Fig8.tif

Figure 8: Comparative graphic between the 2FK4\_03 (E6C 4C/4S model 3), 6E6C\_15 (homology model of the C-terminal wild type structure of E6) and 5CHI\_03 (homology model of the complete structure of the E6wt) simulation trajectories.

Fig9.tif

Figure 9: Alignments used on the input files on the modeling protocols.

## Tables

**Table 1:** Comparative table comparing some of the results for the models (6E6C\_15 and 4E6N\_10) and the template (2FK4\_03) analysis. LL – Lower Limit; UL – Upper Limit; %F – % residues in most favored regions; AA – residues in additional allowed regions; GA – residues in generously allowed regions; NA – residues in disallowed regions.

Model	Verify3D		PROSA	Procheck							
	LL	UL	Z-score	Ramachandran				<i>G-factor</i>			Bad contacts
				% F	NA	GA	AA	dihedral angles	mainchain bonds lengths	Overall	
2FK4_03	0,02	0,5	-4,34	96,6	0	1	1	0,17	-0,09	0,06	18
6E6C_15	0,01	0,4	-3,79	94,7	0	1	2	0,19	-0,17	0,06	7
4E6N_10	0,02	0,44	-3,84	93,1	0	1	3	-0,05	-0,39	-0,17	4

**Table 2:** Resume of the PROSITE results.

Signature	Pattern	E6 Portion	Similarity
cAMP- and cGMP-dependent protein kinase phosphorylation site.	[ST]-x-[RK]	146 a 149	100%
Protein kinase C phosphorylation site	[ST]-x-[RK]	133 a 135	100%
		142 a 144	100%
		145 a 147	100%
Casein kinase II phosphorylation site	[ST]-x(2)-[DE]	22 a 25	100%
		86 a 89	100%
		145 a 148	100%
Tyrosine kinase phosphorylation site	[RK]-x(2,3)-[DE]-x(2,3)-Y	26 a 32	71%
		48 a 54	72%
		53 a 60	71%
		68 a 76	72%
		72 a 79	100%
		77 a 84	72%
N-myristoylation site.	G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}	86 a 92	71%
		57 a 62	100%
		85 a 90	74%
		130 a 135	88%
Bacterial regulatory proteins, araC family signature.	[KRQ]-[LIVMA]-x(2)-[GSTALIV]-{FYWPGDN}-x(2)-[LIVMSA]-x(4,9)-[LIVMF]-x-{PLH}-[LIVMSTA]-[GSTACIL]-{GPK}-{F}-x-[GANQRF]-[LIVMFY]-x(4,5)-[LFY]-x(3)-[FYIVA]-{FYWHCM}-{PGVI}-x(2)-[GSADENQKR]-x-[NSTAPKL]-[PARL]	134 a 139	88%
		11 to 55	86%
		14 to 55	76%
Thioredoxin family active site.	[LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C-[GATPLVE]-[PHYWSTA]-C-{I}-x-{A}-x(3)-[LIVMFYWT]	66 to 109	80%
Lipocalin signature.	[DENG]-{A}-[DENQGSTARK]-x(0,2)-[DENQARK]-[LIVFY]-{CP}-G-{C}-W-[FYWLRH]-{D}-[LIVMTA]	128 a 146	70%
Glycoprotein hormones beta chain signature 1.	C-[STAGM]-G-[HFYL]-C-x-[ST]	122 a 135	72%
Bowman-Birk serine protease inhibitors family signature.	C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDEKS]-[DEKRHSTA]-C	136 a 142	72%
Plant lipid transfer proteins signature	[LIVM]-[PA]-x(2)-C-x(1,2)-[LIVM]-x(1,2)-[LIVMST]-x-[LIVMFY]-x(1,2)-[LIVMF]-[STRD]-x(3)-[DN]-C-x(2)-[LIVM]	51 a 66	76%
C-type lectin domain signature.	C-[LIVMFYATG]-x(5,12)-[WL]-{T}-[DNSRI]-{C}-{LI}-C-x(5,6)-[FYWLIVSTA]-[LIVMSTA]-C	12 a 36	70%
		30 a 51	74%
Heavy-metal-associated domain	[LIVNS]-x-{L}-[LIVMFA]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]-{LV}-x(2)-[LIV]-x(9,11)-[IVA]-x-[LVFYS]	80 a 106	74%
		25 a 54	77%
Ribosomal protein L31 signature.	[DES]-[IVT]-x(4)-H-[PT]-[FAVY]-[FYW]-[TISN]-x(9,13)-[GN]-[KRHNQ]	25 a 55	70%
Laminin-type EGF-like (LE) domain signature	C-x(1,2)-C-x(5)-G-x(2)-C-x(2)-C-x(3,4)-[FYW]-x(3,15)-C	72 a 94	70%
Zinc finger PHD-type signature	C-x(1,2)-C-x(5,45)-[VMFLWIE]-x-C-x(1,4)-C-x(1,4)-[WYFVQHLT]-H-x(2)-C-x(5,45)-[WFLYI]-x-C-x(2)-C	51 a 80	73%
		30 a 106	75%
		63 a 139	75%

**Table 3:** RMSD values for different portions of E6 thru the molecular dynamics trajectory, with the initial structure as reference.

E6 portion (aa)	RMSD in Å (9,242 ns)			RMSD maximum (backbone)
	All Atoms	Backbone	Alpha carbon	
1 a 151	8,94901	7,95393	7,99958	8,70579
9 a 71	4,21171	2,83663	2,85201	3,1002
80 a 144	4,68700	2,90283	2,88311	3,46476

**Table 4: Summary of the simulation systems used on the molecular dynamics.**

	<b>Model name</b>		
	<b>2FK4_03</b>	<b>6E6C_15</b>	<b>5CHI_03</b>
Simulation ID	SIM001	SIM002	SIMNC001
Number of atoms (- H <sub>2</sub> O)	1016	1086	2550
Number of H <sub>2</sub> O molecules	4611	4619	7635
Box dimensions (Å <sup>3</sup> )	63,4 x 55,5 x 56,0	60,9 x 54,2 x 59,0	70,8 x 67,0 x 68,6
Residual charge neutralization	plasma	plasma	6 Cl <sup>-</sup>
Simulation lenght	5,0 ns	10,5 ns	9,242 ns

**Table 5:** References of the mutational analysis with percentage of p53 binding and degradation and E6AP binding when tested.

Region: Mutation/Deletion	Binds p53	Degrades p53	Binds E6AP	Comments	References
Phe2Leu	NT	50%	Sim	<ul style="list-style-type: none"> <li>• Stable replication</li> </ul>	Park et al, 2002.
Phe2Val	NT	Não	Sim	<ul style="list-style-type: none"> <li>• hTERT: increases activity</li> <li>• Reduces half-life of NFX-1 by degradation/ubiquitination</li> <li>• No stable replication</li> </ul>	Liu et al, 1999; Gewin et al, 2004 et al; Park et al, 2002.
Gln3-Arg8	NT	Não	NT	<ul style="list-style-type: none"> <li>• MoAb (6F4) against this region inhibits p53 degradation.</li> </ul>	Masson et al, 2003.
ΔPhe2-Leu12	Não	NT	NT		Foster et al, 1994.
Arg8Ser/Pro9Ala/Arg10Thr	Não	não	30%	<ul style="list-style-type: none"> <li>• hTERT: increases activity</li> <li>• Reduces half-life of NFX-1 by degradation/ubiquitination</li> </ul>	Foster et al, 1994; Gewin and Galloway, 2001; Klingelutz et al, 1996; Gewin et al, 2004; Park et al, 2002.
Arg8Gly/Arg10Gly/Lys11Glu *truncated	NT	NT	NT	<ul style="list-style-type: none"> <li>• E*1 e *II (41 and 48 residues)</li> <li>• Breaks the NLS</li> </ul>	Tao et al, 2003
ΔPro9-Pro13	Não	Não	Não	<ul style="list-style-type: none"> <li>• hTERT: no activation</li> </ul>	Foster et al, 1994; Gewin and Galloway, 2001
ΔGlu18-Asp25	0%	0%	0%	<ul style="list-style-type: none"> <li>• Lack of cellular transformation</li> <li>• ~40% p53 transactivation repression.</li> </ul>	Elbel et al, 1997
Leu19Ala	NT	Sim	NT		Nominé et al, 2006.
Ile26Ala	NT	Sim	NT		Nominé et al, 2006.
ΔGln35-Asp56	0%	NT	0%	<ul style="list-style-type: none"> <li>• Medium cellular transformation</li> <li>• ~40% p53 transactivation repression.</li> </ul>	Elbel et al, 1997
Leu37Ser	NT	NT	Não	<ul style="list-style-type: none"> <li>• hTERT: activation</li> <li>• MEC immortalization</li> </ul>	Liu et al, 1999; Sekaric et al, 2008
Phe45Val	29%	25%	NT		Foster et al, 1994.
Phe47Arg	0%	0%	Sim	<ul style="list-style-type: none"> <li>• Modify H1' helix</li> <li>• Increases E6AP affinity</li> <li>• Acts as negative dominant with the wt.</li> </ul>	Nominé et al, 2006; Zanier et al, 2005; Ristriani et al, 2009
Phe45Tyr/Phe47Tyr/Asp49His	não	Não	NT		Foster et al, 1994.
Leu50Gly	NT	Não	Não	<ul style="list-style-type: none"> <li>• Lack of hTERT activation</li> <li>• Lack of MEC immortalization</li> </ul>	Zimmermann et al, 1999; Sekaric et al, 2008; Shamanin et al, 2008.

**Table 5:** [cont.]

Region: Mutation/Deletion	Binds p53	Degrades p53	Binds E6AP	Comments	References
Asp51Ala, Phe54Ala, Arg55Ala, Asp56Ala (158aa)	NT	Não	NT	• DNA: No effect on Holiday Junction binding	Ristriani et al, 2000
Tyr54His	NT	Não	Sim	• hTERT: increases activity • No stable replication	Liu et al, 1999; Park et al, 2002.
Cys63Arg	NT	30%	50%	• Stable replication. • Conserved Cys	Park et al, 2002.
Cys66Gly	Não	Não	NT	• Conserved Cys	Foster et al, 1994
ΔSer71-Gln90	0%	NT	0%	• Medium cellular transformation • p53: no effect on p53 transactivation	Elbel et al, 1997
ΔIle73-Arg77	Não	Não	NT		Foster et al, 1994.
ΔTyr76-Tyr79	NT	40%	NT	• Decreases thermal stability • Decreases below 50% zinc binding capacity	Liu et al, 2009.
ΔAsn105-Leu110	Fraco	NT	NT		Crook et al, 1991; Lechner et al, 1994.
Leu110Gln	NT	0%	0%	• No stable replication • MEC immortalization • hTERT activation	Park et al, 2002; Liu et al, 1999; Sekaric et al, 2008.
ΔLys108-His126	0%	NT	0%	• Lack of cellular transformation • p53: no effect on p53 transactivation	Elbel et al, 1997.
ΔCys111-Lys115, Asn127Lys	Não	Não	NT		Foster et al, 1994.
Arg109Ala, Gln114Ala, Lys115Ala (158aa)	NT	NT	NT	• DNA: lack of Holiday Junction binding	Ristriani et al, 2000
ΔHis118-Lys122	Não	2-30%	30%	• hTERT: low activation	Foster et al, 1994; Dalal et al, 1996; Kiyono et al, 1998; Park et al, 2002.
ΔGln123-Asn127	Não 30% (AcMo)	Não	Não (80% AcMo)	• hTERT: no activation • MoAb (3B8) against this region inhibits 70% p53 binding, partially degrades p53, and inhibits 20% of E6AP binding.	Foster et al, 1994; Klingelutz et al, 1996; Liu et al, 1999, Lagrange et al, 2005.
Asn127-Trp132	22% (AcMo)	NT	63% (AcMo)	• MoAb (1F5) against this region inhibits 78% p53 binding.	Lagrange et al, 2005.
Ile128Ala	NT	Sim	NT	• p53: just in vivo degradation, not in vitro	Nominé et al, 2006.



**Table 5:** [cont.]

Region: Mutation/Deletion	Binds p53	Degrades p53	Binds E6AP	Comments	References
ΔIle128-Trp132	Não	Não	NT		Foster et al, 1994
Gly130Val	NT	NT	Não	<ul style="list-style-type: none"> <li>No MEC immortalization</li> <li>Lack of hTERT activation</li> </ul>	Sekaric et al, 2008; Shamanin et al, 2008
Thr133-Ser138	79% (AcMo)	não	7,3%(AcMo)	<ul style="list-style-type: none"> <li>MoAb (3F8) against this region inhibits 79% p53 binding.</li> </ul>	Lagrange et al, 2005.
ΔCys140-Leu151	35%	25%	NT		Foster et al, 1994.
ΔSer142-Thr149	~60%	~83%	100%	<ul style="list-style-type: none"> <li>Lack of cellular transformation</li> <li>p53: inhibits p53 transactivation</li> </ul>	Elbel et al, 1997
ΔArg146-Leu151	33%	Sim (67%)	NT	<ul style="list-style-type: none"> <li>hTERT: increases activity</li> </ul>	Foster et al, 1994; Klingelutz et al, 1996; Kiyono et al, 1998, Nominé et al, 2006
6C/6S Lys121Ala / Lys122Ala / Gln123Ala / Arg124Ala	NT	Sim	NT	<ul style="list-style-type: none"> <li>Alters the NLS</li> <li>DNA: Lack of Holiday Junction binding</li> <li>p53: degradação apenas <i>in vivo</i>, não <i>in vitro</i></li> </ul>	Nominé et al, 2006; Le Roux and Moroianu, 2003, Ristriani et al, 2000.
Apenas porção C terminal E6C	NT	NT	NT	<ul style="list-style-type: none"> <li>DNA: no effect on Holiday Junction binding</li> </ul>	Ristriani et al, 2000/2001
ΔLeu151	NT	Sim	NT	<ul style="list-style-type: none"> <li>Lack of cellular transformation</li> <li>PDZ: Lack of hDLG binding</li> </ul>	Kiyono et al, 1997.
Leu151Val	NT	Sim	NT	<ul style="list-style-type: none"> <li>(C-term = HPV 18)</li> <li>Shows cellular transformation</li> <li>PDZ: no effect on hDLG binding</li> </ul>	Kiyono et al, 1997.
Leu151Ile	NT	Sim	NT	<ul style="list-style-type: none"> <li>Shows cellular transformation</li> <li>PDZ: no effect on hDLG binding</li> </ul>	Kiyono et al, 1997.
Leu151Phe	NT	Sim	NT	<ul style="list-style-type: none"> <li>Lack of cellular transformation</li> <li>PDZ: inhibits hDLG binding</li> </ul>	Kiyono et al, 1997.
Leu151Phe-Ile	NT	Sim	NT	<ul style="list-style-type: none"> <li>Lack of cellular transformation</li> <li>PDZ: inhibits hDLG binding</li> </ul>	Kiyono et al, 1997.
Leu151Glu-Thr-Gln-Val	NT	Sim	NT	<ul style="list-style-type: none"> <li>Shows cellular transformation</li> <li>PDZ: no effect on hDLG binding</li> </ul>	Kiyono et al, 1997.

## 7 Considerações Finais e Perspectivas

A modelagem da estrutura da proteína E6 do HPV 16 realizada neste trabalho mostrou-se consistente com resultados de outros grupos e avançou as possibilidades de pesquisa e utilização de um modelo confiável da estrutura tridimensional da proteína E6 de HPV 16.

Ao mesmo tempo, é possível apontar várias possibilidades de melhorias no processo de modelagem e dinâmica, que aumentariam a acurácia do conhecimento sobre a estrutura real desta proteína.

Dentre os projetos futuros e perspectivas abertas por este trabalho podemos relacionar:

1. Otimizar o protocolo de dinâmica molecular para a proteína completa E6 de HPV16;
2. Realizar a modelagem por homologia para prever a estrutura tridimensional de proteínas E6 de diferentes tipos de HPV como, por exemplo HPV 18, 31 e 33;
3. Realizar dinâmica molecular das novas proteínas modeladas;
4. Realizar modelagem de mutantes descritos e identificar problemas no dobramento, e conseqüentemente na função, da estrutura resultante.

## 8 Referências Bibliográficas

1. AGORASTOS, T; LAMBROPOULOS, A F; CONSTANTINIDIS, T C; KOTSIS, A e BONTIS, J N. p53 codon 72 polymorphism and risk of intra-epithelial and invasive cervical neoplasia in Greek women. **Eur J Cancer Prev**, v.9, n.2, p.113-118, 2000.
2. BEREZUTSKAYA, E; BAGCHI, S. The Human Papillomavirus E7 oncoprotein functionally interacts with the S4 subunit of the 26 S proteasome. **J Biol Chem**,v.272, p.30135, 1997.
3. BRADY, C S; DUGGAN-KEEN, M F; DAVIDSON, J A; VARLEY, J M e STERN, P L. Human papillomavirus type 16 E6 variants in cervical carcinoma: relationship to host genetic factors and clinical parameters. **J. Gen. Virol**, v.80, p.3233, 1999.
- 1.4. BOSCH, F X; LORINCZ, A; MEIJER, C J L M e SHAH, K V. The causal relation between human papillomavirus and cervical cancer. **J Clin Pathol**, v.55, p.244-265, 2002.
5. BOWIE, J U; LÜTHY, R; EISENBERG, D. A method to identify protein sequences that fold into a known three-dimensional structure. **Science**, v.253, n.5016, p.164-170, 1991.
6. BRADFORD, J R e WESTHEAD, D R. Improved prediction of protein-protein binding sites using a support vector machines approach. **Bioinformatics**, v.21: p.1487-1494, 2005.
7. BUTZ, K; SHAHABEDDIN, L; GEISEN, C; SPITKOVSKY, D; ULLMANN, A e HOPPE-SEYLER, F. Functional p53 protein in HPV positive cells. **Oncogene**, v.10, n.5, p.927-936, 1995.
8. CALHOUN, E S; MCGOVERN, R M; JANNEY, CAROL, A; CERHAN, J R; ITURRIA, S J; SMITH, D I; GOSTOUT, B S e PERSING, D H. Host genetic polymorphism analysis in cervical cancer. **Clin Chem**, v.48, n.8, p.1218-1224, 2002.
9. CASE, D A; DARDEN, T A; CHEATHAM, T E; SIMMERLING, C L; WANG, J; DUKE, R E; LUO, R; CROWLEY, M; WALKER, R C; ZHANG, W; MERZ, K M; WANG, B; HAYIK, S; ROITBERG, A; SEABRA, G; KOLOSSVÁRY, I; WONG, K F; PAESANI, F; VANICEK, J; WU, X; BROZELL, S R; STEINBRECHER, T; GOHLKE, H; YANG, L; TAN, C; MONGAN, J; HORNÁK, V; CUI, G; MATHEWS, D H; SEETIN, M G; SAGUI, C; BABIN, V e KOLLMAN, P A. AMBER 10, University of California, San Francisco, 2008.
10. CHARBONNIER, S; STIER, G; ORFANOUDAKIS, G; KIEFFER, B; ATKINSON, R A e TRAVE, G. Defining the minimal interacting regions of the tight junction protein MAGI-1 and HPV16 E6 oncoprotein for solution structure studies. **Protein Expr Purif**, v.60, n.1, p.64-73, 2008.
11. COLE, S T e DANOS, O. Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. **J Mol Biol**, v.193, p599-608, 1987.
12. COMBET, C; BLANCHET, C; GEOURJON, C e DELÉAGE, G. NPS@: Network Protein Sequence Analysis. **TIBS**, v.25, n.3, p147-150, 2000.
13. CORNELL, W D *et al.* A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. **J Am Chem Soc**, v.117, p.5179–5197, 1995.
14. CROOK, T; TIDY, J A e VOUSDEN, K H. Degradation of p53 can be targeted by HPV E6 sequences distinct from those required for p53 binding and trans-activation. **Cell**, v.67, n.3, p.547-556, 1991.

15. CUI, B; SHEN, M-Y; FREED, K F. Folding and misfolding of the Papillomavirus E6 interacting peptide E6ap. **PNAS**, v.100, n.12 p.7087-7092, 2003.
16. DALAL, S; GAO, Q; ANDROPHY, E J e BAND,V. Mutational analysis of human papillomavirus type 16 E6 demonstrates that p53 degradation is necessary for mammary epithelial cell immortalization. **J Virol**, v.70, p.683–688, 1996.
17. DANIELS, P R; SANDERS, C M; MAITLAND, N J. Characterization of the interactions of human Papillomavirus type 16 E6 with p53 and E6-associated protein in insect and human cells. **J Gen Virol**, v.79, p.489–499, 1998.
18. DESAINTEs, C; DEMERET, C; GOYAT, S; YANIV, M e THIERRY, F. Expression of papillomavirus E2 protein in HeLa cells leads to apoptosis. **EMBO J**, v.16, n.3, p.504-514, 1997.
19. DE VILLIERS, E M. Heterogeneity of the Human Papillomavirus group. **J Virol**, v.63, p. 4898-4903, 1989.
20. DOYLE, D A; LEE, A; LEWIS, J; KIM, E; SHENG, M e MACKINNON, R. Crystal structures of a complexed and peptide-free membrane protein-binding domain: molecular basis of peptide recognition by PDZ. **Cell**, v.85, p.1067-1076, 1996.
21. DUENSING, S; MÜNGER, K. The Human Papillomavirus Type 16 E6 and E7 oncoproteins independently induce numerical and structural chromosome instability. **Cancer Res**, v.62, p.7075-7082, 2002.
22. DUENSING, S; MÜNGER, K. Human Papillomavirus type 16 E7 oncoprotein can induce abnormal centrosome duplication through a mechanism independent of inactivation of retinoblastoma protein family members. **J Virol**, v.77, p.12331-12335, 2003.
23. ELBEL, M; CARL, S; SPADERNA, S e IFTNER, T. A comparative analysis of the interaction of the E6 from cutaneous and genital papillomaviruses with p53 and E6AP in correlation with transforming potential. **Virol**, v.239, p.132-149, 1997.
24. ESWAR, N *et al.* Comparative Protein Structure Modeling With MODELLER. **Current Protocols in Bioinformatics**, John Wiley & Sons, Inc., Suplemento 15, 5.6.1-5.6.30, 2000.
25. FEHRMANN, F; KLUMPP, D J e LAIMINS, L A. Human Papillomavirus Type 31 E5 protein supports cell cycle progression and activates late viral functions upon epithelial differentiation. **J Virol**, v.77, n.5, p.2819-2831, 2003.
26. FILIPPOVA, M; SONG, H; CONNOLLY, J L; DERMODY, T S; e DUERKSEN-HUGHES P J. The Human Papillomavirus 16 E6 protein binds to tumor necrosis Factor (TNF) R1 and protects cells from TNF-induced apoptosis. **J Biol Chem**. v.277, n.24, p. 21730–21739, 2002.
27. FINLEY, D; CHAU, V. Ubiquitination. **Annual Rev Cell Biol**, v.7, p.25-69, 1991.
28. FISER, A; DO, R K e SALI, A. Modeling of loops in protein structures. **Protein Science**, v.9, p.1753-1773, 2000.
29. FLORES, T P; ORENGO, C A; MOSS, D S e THORNTON, J M. Comparison of conformational characteristics in structurally similar protein pairs. **Protein Sci**, v.2, n.11; p.1811-1826, 1993.

30. FOSTER, S A; DEMERS, G W; ETSCHIED, B G e GALLOWAY, D A. The ability of HPV 16 E6 proteins to target p53 for degradation *in vivo* correlates with their ability to abrogate actinomycin D-induced growth-arrest. **J Virol**, v.68, n.9, p.5698-5705, 1994.
31. FREEDMAN, D A e LEVINE, A J. Nuclear export is required for degradation of endogenous p53 by MDM2 and HPV E6. **Mol Cell Biol**, v.18, n.12, p.7288-7293, 1998.
32. GAMMOH, N; GRM, H S; MASSIMI, P e BANKS, L. Regulation of Human Papillomavirus Type 16 E7 activity through direct interaction with the E2 transcriptional activator. **J Virol**, v.80, n.4, p.1787-1797, 2006.
33. GARDIOL, D; BANKS, L. Comparisson of HPV 18 E6-mediated degradation of p53 *in vitro* and *in vivo* reveals significant differences based on p53 structure and cell type but little difference with respect to mutants of HPV 18 E6. **J Gen Virol**, v.79, p.1963-1970, 1998.
34. GARDIOL, D; KUHNE, C; GLAUNSINGER, B; LEE, S S; JAVIER, R e BANKS, L. Oncogenic human papillomavirus E6 proteins target the discs large tumour suppressor for proteasome-mediated degradation. **Oncogene**, v.18, p.5487-5496, 1999.
35. GEWIN, L e GALLOWAY, D A. E Box-Dependent Activation of Telomerase by Human Papillomavirus Type 16 E6 Does Not Require Induction of c-myc. **J Virol**, v.75, p.7198-7201, 2001.
36. GEWIN, L; MYERS, H; KIYONO, T e GALLOWAY, D A. Identification of a novel telomerase repressor that interacts with the human papillomavirus type-16 E6/E6-AP complex. **Genes & Dev**, v.18, p.2269-2282, 2004.
37. GLAUNSINGER. B A; LEE, S S; THOMAS, M; BANKS, L e JAVIER, R. Interactions of the PDZ-protein MAGI-1 with adenovirus E4-ORF1 and high-risk papillomavirus E6 oncoproteins. **Oncogene**, v.19, p.5270-5280, 2000.
38. GONZALEZ, S L; STREMLAU, M; HE, X; BASILE, J R, e MÜNGER, K. Degradation of the Retinoblastoma tumor suppressor by the Human Papillomavirus Type 16 E7 oncoprotein is important for functional inactivation ad is separable from proteosomal degradation of E7. **J Virol**, v.75, n.16, p.7583-7591, 2001.
39. GUEX, N e PEITSCH, M C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. **Electrophoresis**, v.18, p.2714-2723, 1997.
40. HELLAND, A; LANGEROD, A; JOHNSEN, H; OLSEN, A O; SKOVLUND, E e BORRESEN-DALE, A L. p53 polymorphism and risk of cervical cancer. **Nature**, v.396, n.6711, p.530-531, réplica do autor 532, 1998.
41. HELT, A-M; FUNK, J O; GALLOWAY D A. Inactivation of both the retinoblastoma tumor suppressor and p21 by the Human Papillomavirus type 16 E7 oncoprotein is necessary to inhibit cell cycle arrest in human epithelial cells. **J Virol**, v.76, p.10559-10568, 2002.
42. HELT, A M; GALLOWAY, D A. Mechanisms by which DNA tumor virus oncoproteins target the Rb family of pocket proteins. **Carcinogenesis**, v.24, n.2, p.159-169, 2003
43. HENGSTERMANN, A; LINARES, L K; CIECHANOVER, A; WHITAKER, N J e SCHEFFNER, M. Complete switch from Mdm2 to human papillomavirus E6-mediated degradation of p53 in cervical cancer cells. **PNAS USA**, v.98, n.3, p.1218-1223, 2001.
44. HOWIE, H L; KATWENELLENBOGEN, R A e GALLOWAY, D A. Papillomavirus E6 proteins. **Virology**, v.384, p.324-334, 2009.

45. HOWLEY, P. *Papillomavirinae: the viruses and their replication*. In **B. Fields and D. Knipe (ed.)**, **Fields virology**, 4th ed., versão digital, New York, EUA: Raven Press, 2001.
46. HUIBREGTSE, J M, SCHEFFNER M, e HOWLEY PM. A cellular protein mediates association of p53 with the E6 oncoprotein of human papillomavirus types 16 or 18. **EMBO J**, v.10, n. 13. p. 4129-4135, 1991.
47. HUIBREGTSE, J M, SCHEFFNER M, e HOWLEY PM. Localization of the E6 AP regions that direct HPV E6 binding, association with p53 and ubiquitination of associated proteins. **Mol Cell Biol**, v.13, n.8, p.4918-4927, 1993.
48. HUIBREGTSE, J M e BEAUDENON, S L. Mechanism of HPV E6 proteins in cellular transformation. **Semin Cancer Biol**, v.7, n.6, p.317-326, 1996.
49. HULO, N; BAIROCH, A; BULLIARD, V; CERUTTI, L; CUCHE, B A; DE CASTRO, E; LACHAIZE, C; LANGENDIJK-GENEVAUX, P S e SIGRIST, C J. The 20 years of PROSITE. **Nucleic Acids Res**, v.36, p.D245–D249, 2008.
50. HUMPHREY, W; DALKE, A e SCHULTEN, K. VMD – Visual Molecular Dinamics. **J Molec Graphics**, v.14.1, p.33-38, 1996.
51. HUNG, A Y e SHENG, M. PDZ domains: structural modules for protein complex assembly. **J Biol Chem**, v.277, n.8, p. 5699-5702, 2002.
52. IAKOUCHEVA, L M; BROWN, C J; LAWSON, J D; OBRADOVIC, Z e DUNKER, A K. Intrinsic disorder in cell-signaling and cancer-associated proteins. **J Mol Biol**, v.323, p.573-584, 2002.
53. JACKSON, S; HARWOOD, C; THOMAS, M; BANKS, L e STOREY, A. Role of Bak in UV-induced apoptosis in skin cancer and abrogation by HPV E6 proteins. **Genes & Dev**, v.14, p.3065-3073, 2000.
54. JORGENSEN, W; CHANDRASEKAR, J; MADURA, J; IMPEY, R e KLEIN, M. Comparison of simple potential functions for simulating liquid water. **J Chem Phys**, v.79, p.926-935, 1983.
55. KAWAMATA, Y; MITSUHASHI, A; UNNO, Y; KADO, S; SHINO, Y; UESUGI, K; EGUCHI, O; ISHII, J; SEKI, K; SEKIYA, S e SHIRASAWA, H. HPV E6-mediated degradation is compensated by upregulation of p53 gene expression in normal cervical keratinocytes. **Int J Onc**, v.21, p.561-567, 2002.
56. KELLEY, M L; KEIGER, K E; LEE, C J e HUIBREGTSE, J M. The global transcriptional effects of the HPV E6 protein in cervical carcinoma cell lines are mediated by E6AP ubiquitin ligase. **J Virol**. v.79, n.6, p.3737-3747, 2005.
57. KIYONO, T; HIRAIWA, A; FUJITA, M; HAYASHI, Y; AKIYAMA, T e ISHIBASHI, M. Binding of high-risk human Papillomavirus E6 oncoproteins to the human homologue of the *Drosophila* discs large tumor suppressor protein. **PNAS**, v.94, p.11612-11616, 1997.
58. KLINGELHUTZ, A J; FOSTER, S A. e MCDOUGALL, J K. Telomerase activation by the E6 gene product of human papillomavirus type 16. **Nature**, v.380, n.6569, p.79-82, 1996.
59. LAGRANGE, M; CHARBONNIER, S; ORFANOUDAKIS, G; ROBINSON, P; ZANIER, K; MASSON, M; LUTZ, Y; TRAVE, G; WEISS, E e DERYCKERE, F. Binding of human papillomavirus 16 E6 to p53 and E6AP is impaired by monoclonal antibodies directed against the second zinc binding domain of E6. **J Gen Virol**, v.86, p.1001-1007, 2005.

60. LASKOWSKI, R A; MACCARTHUR, M W; MOSS, D S e THORNTON, J M. PROCHECK: a program to check the stereochemical quality of protein structures. **J Appl Cryst**, v.26; p.283-291, 1993.
61. LAZARCZYK, M e FAVRÉ, M. Role of Zn<sup>2+</sup> Ions in Host-Virus Interactions. **J Virol**, v.82, n.23, p.11486-11494, 2008.
62. LECHNER, M S e LAIMINS, L A. Inhibition of p53 DNA binding by Human Papillomavirus E6 proteins. **J Virol**, v.68, n.7, p.4262-4273, 1992.
63. LEE, S; PARK, N-H. HPV E7 oncoprotein inhibits the expression of fibronectin, Livro de resumos do 18<sup>th</sup> **International Papillomavirus Conference**, resumo 640, p.473, 2000.
64. LE ROUX, L G e MOROIANU, J. Nuclear Entry of High-Risk Human Papillomavirus Type 16 E6 Oncoprotein Occurs via Several Pathways. **J Virol**, v.77, p.2330-2337, 2003.
65. LI, X; COFFINO, P. High-risk HPV protein has two distinct sites within p53, of which only one determines degradation. **J Virol**, v.70, n.7, p.4509-4516, 1996.
66. LIPARI, F; McGIBBON, G A; WARDROP, E e CORDINGLEY, M G. Purification and biophysical characterization of a minimal function domain and of an N-Terminal Zn<sup>2+</sup>-binding fragment of the human Papillomavirus E6 protein. **Biochemistry**, v.40, n.5, p.1196-1204, 2001.
67. LIU, Y; CHEN, J J; GAO, Q; DALAL, S; HONG, Y; MANSUR, C P; BAND, V e ANDROPHY, E J. Multiple Functions of Human Papillomavirus Type 16 E6 Contribute to the Immortalization of Mammary Epithelial Cells **J Virol**, v.73, p.7297-7307, 1999.
68. LIU, X; ROMERO, P; RANI, M; DUNKER, A K e OBRADOVIC, Z. Predicting Protein Disorder for N-, C-, and Internal Regions. **Genome Informatics**, v.10, p.30-40, 1999.
69. LIU, X e MARMORSTEIN, R. When viral oncoprotein meets tumor suppressor: a structural view. **Genes & Dev**, v.20, p.2332-2337, 2006
70. LIU, J; PERUMAL, N B; OLDFIELD, C J; SU, E W; UVERSKY, V N e DUNKER, K A. Intrinsic disorder in transcription factors. **Biochemistry**, v.45, n.22, p.6873-6888, 2006.
71. LIU, Y; CHERRY, J J; DINEEN, J V; ANDROPHY, E J e BALEJA, J D. Determinants of Stability for the E6 Protein of Papillomavirus Type 16. **J Mol Biol**, v.386, p.1123-1137, 2009.
72. LÜTHY, R; BOWIE, J U; EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**. v.356, n.6364, p.83-85, 1992.
73. MAKNI, H; FRANCO, E L; KAIANO, J; VILLA, L L; LABRECQUE, S; DUDLEY, R; STOREY, A e MATLASHEWSKI, G. p53 polymorphism in codon 72 and risk of human papillomavirus-induced cervical cancer: effect of inter-laboratory variation. **Int J Cancer**, v.87, n.4, p.528-533, 2000.
74. MARTI-RENOM, M A; STUART, A C; FISER, A; SANCHEZ, R; MELO, F e SALI, A. Comparative protein structure modeling of genes and genomes. **Annu Rev Biophys Biomol Struct**, v.29, p.291-325, 2000.
75. MASSIMI, P; PIM, D; BANKS, L. Human Papillomavirus type 16 E7 binds to the conserved carboxy-terminal region of the TATA box binding protein and this contributes to E7 transforming activity. **J. Gen. Virol**, v.78, p.2607, 1997.

76. MASSON, M; HINDELANG, C; SIBLER, A-P; SCHWALBACH, G; TRAVÉ, G e WEISS, E. Preferential nuclear localization of the human papillomavirus type 16 E6 oncoprotein in cervical carcinoma cells. **J Gen Virol**, v.84, p.2099–2104, 2003.
77. MITRA, S; MISRA, C; SINGH, R K; PANDA, C K e ROYCHOUDHURY, S. Association of specific genotype and haplotype of p53 gene with cervical cancer in India. **J Clin Pathol**, v.58, p.26–31, 2005.
78. MORRIS, A L; MACARTHUR, M W; HUTCHINSON, E G e THORNTON, J M. Stereochemical quality of protein structure coordinates. **Proteins**, 12, 345-364, 1992.
79. MÜNGER, K; PHELPS, W C; BUBB, V; HOWLEY, P M e SCHLEGEL, R. The E6 and E7 genes of the human papillomavirus type 16 together are necessary and sufficient for transformation of primary human. **J Virol**, v.63, n.10, p.4417-4412, 1989.
80. NAKAGAWA, S e HUIBREGTSE, J M. Human Scribble (Vartul) Is Targeted for Ubiquitin-Mediated Degradation by the High-Risk Papillomavirus E6 Proteins and the E6AP Ubiquitin-Protein Ligase. **Mol Cell Biol**, v.20, p.8244 – 8253, 2000.
81. NGUYEN, M; SONG, S; LIEM, A; ANDROPHY, E; LIU, Y e LAMBERT, P F. A mutant of HPV 16 E6 deficient in binding  $\alpha$ -helix displayed reduced oncogenic potential in vivo. **J Virol**, v.76, n.24, p.13039-13048, 2002.
82. NOMINÉ, Y; RISTRANI, T; LAURENT, C; LEFÈVRE, J-F; WEISS, E e TRAVÉ, G. A strategy for optimizing the monodispersity of fusion proteins: application to purification of recombinant HPV E6 oncoprotein. **Protein Eng**, v.14, n.4, p.297-305, 2001.
83. NOMINÉ, Y; CHARBONNIER, S; RISTRANI, T; STIER, G; MASSON, M; CAVUSOGLU, N; VAN DORSSELAER, A; WEISS, E; KIEFFER, B e TRAVE, G. Domain substructure of HPV oncoprotein: biophysical characterization of the E6 terminal binding domain. **Biochem**, v.42, n.17, 4909-4917, 2003.
84. NOMINÉ, Y; CHARBONNIER, S; MIGUET, L; POTIER, N; VAN DORSSELAER, A; ATKINSON, R A; TRAVE, G e KIEFFER, B.  $^1\text{H}$  and  $^{15}\text{N}$  resonance assignment, secondary structure and dynamic behavior of the C-terminal domain of human papillomavirus oncoprotein E6. **J Biomol NMR**, v.31, p.129-141, 2005.
85. NOMINÉ, Y; MASSON, M; CHARBONNIER, S; ZANIER, K; RISTRANI, T; DERYCKERE, F; SIBLER, A P; DESPLANCQ, D; ATKINSON, R A; WEISS, E; ORFANOUDAKIS, G; KIEFFER, B e TRAVE, G. Structural and functional analysis of E6 oncoprotein: Insights in the molecular pathways of human Papillomavirus-mediated pathogenesis. **Mol Cell**, v.21, p.665-678, 2006.
86. PARK, N-H; KANG, M K. Genetic instability and oral cancer. **EJB Eletronic J Biotech**, v.3, n.1, p.66-71, 2000.
87. PARK, R B e ANDROPHY, E J. Genetic Analysis of High-Risk E6 in Episomal Maintenance of Human Papillomavirus Genomes in Primary Human Keratinocytes. **J Virol**, v.76, p.11359-11364, 2002.
88. PASCUAL, J; MARTINEZ-YAMOUT, M; DYSON, H J E WRIGHT, P E. Structure of the PHD zinc finger from human Williams-Beuren syndrome transcription factor. **J Mol Biol**, v.304, n.5, p.723-729, 2000.
89. PATEL, D; HUANG, S M; BAGLIA, L A e McCANCE, D J. The E6 protein of HPV 16 binds to and inhibits co-activation by CBP/p300. **EMBO J**, v.18, n.18, p.5061-5072, 1999.



90. PIM, D; STOREY, A; THOMAS, M; MASSIMI, P e BANKS, L. Mutational analysis of HPV-18 E6 identifies domains required for p53 degradation in vitro, abolition of p53 transactivation in vivo and immortalisation of primary BMK cells. **Oncogene**, v.9, n.7, p. 1869-1876, 1994.
91. PIM, D; THOMAS, M e BANKS, L. Chimaeric HPV E6 proteins allow dissection of the proteolytic pathways regulating different E6 cellular target proteins. **Oncogene**, v.21, p.8140-8148, 2002.
92. PRILUSKY, J; FELDER, C E; ZEEV-BEN-MORDEHAI, T; RYDBERG, E; MAN, O; BECKMANN, J S; SILMAN, I e SUSSMAN, J L. Fold Index: FoldIndex©: a simple tool to predict whether a given protein sequence is intrinsically unfolded. **Bioinformatics**, v.21, n.16, p.3435-3438.
93. RADIVOJAC, P; OBRADOVIC, Z; BROWN, C J e DUNKER, A K. Prediction of boundaries between intrinsically ordered and disordered protein regions. **Pac Symp Biocomput**, v.8, p.216-227, 2003.
94. REY, O; LEE, S; PARK, N H. Impaired nucleotide excision repair in UV-irradiated human oral keratinocytes immortalized with type 16 human papillomavirus genome. **Oncogene**, v.18, n.50, p.6997-7001. 1999.
95. RISTRIANI, T; MASSON, M; NOMINE, Y; LAURENT, C; LEFEVRE, J F; WEISS, E e TRAVE, G. HPV oncoprotein E6 is a structure-dependent DNA-binding protein that recognizes four-way junctions. **J Mol Biol**, v.296. p.1189-1203, 2000.
96. RISTRIANI, T; NOMINE, Y; MASSON, M; WEISS, E e TRAVE, G. Specific recognition of four-way DNA junctions by the C-terminal zinc-binding domain of HPV oncoprotein E6. **J Mol Biol**, v.305, n.4, p.729-739, 2001.
97. RODRIGUEZ, M S; DESTERRO, J M P; LAIN, S; LANE, D P e HAY, R T. Multiple C-terminal Lysine residues target p53 for ubiquitin-proteasome-mediated degradation. **Mol Cell Biol**, v.20, n.22, p.8458-8467, 2000.
98. ROMERO, P; OBRADOVIC, Z; LI, X; GARNER, E; BROWN, C e DUNKER, A K. Sequence complexity of disordered protein. **Proteins: Struct Funct Gen**, v.42, p.38-48, 2001.
99. RONCO, L V; KARPOVA, A Y; VIDAL, M e HOWLEY, P M. Human papillomavirus 16 E6 oncoprotein binds to interferon regulatory factor-3 and inhibits its transcriptional activity. **Genes and Dev**, v.12, n.13, p.2061-2072, 1998.
100. ROSENTHAL, A N; RYAN, A; AL-JEHANI, R M; STOREY, A; HARWOOD, C A e JACOBS, I J. p53 codon 72 polymorphism and risk of cervical cancer in UK. **Lancet**, v.352, n.9131, p.871-872, 1998.
101. SALI, A e BLUNDELL, T L. Comparative protein modelling by satisfaction of spatial restraints. **J Mol Biol**, v.234; n.3; p.779-815, 1993.
102. SCHAFMEISTER, C E A F; ROSS, W S e ROMANOVSKI, V. LEaP, University of California, San Francisco. 1995.
103. SCHECKENBACH, K; LIEVEN, O; GÖTTE, K; BOCKMÜHL, U; ZOTZ, R; BIER, H e BALZ, V. p53 codon 72 polymorphic variants, loss of allele-specific transcription, and human papillomavirus 16 and/or 18 E6 messenger RNA expression in squamous cell carcinomas of the head and neck. **Cancer Epidemiol Biomarkers Prev**, v.13, p.1805-1809, 2004.

104. SCHEFFNER, M; WERNESS, B A; HUIBREGTSE, J M; LEVINE, J M e HOWLEY, P M. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53.. **Cell**, v.63, n.6, p.1129-1136, 1990.
105. SCHEFFNER, M; TAKAHASHI, T; HUIBREGTSE, J M; MINNA, J D e HOWLEY, P M. Interaction of the Human Papillomavirus type 16 E6 oncoprotein with wild type and mutant human p53 proteins. **J Virol**, v.66, n.8, p.5100-5105, 1992.
106. SCHMITT, V M. Papilomavírus Humano. In: Rosseti, M L. **Doenças infecciosas: diagnóstico molecular**. Rio de Janeiro: Guababara-Koogan, 2006.
107. SEKARIC, P; CHERRY, J J e ANDROPHY, E J. Binding of Human Papillomavirus Type 16 E6 to E6AP Is Not Required for Activation of hTERT. **J Virol**, v.82, n.1, p.71–76, 2008.
108. SHAMANIN, V A; SEKARIC, P e ANDROPHY, E J. hAda3 Degradation by Papillomavirus Type 16 E6 Correlates with Abrogation of the p14ARF-p53 Pathway and Efficient Immortalization of Human Mammary Epithelial Cells. **J Virol**, v.82, p.3912-3920, 2008.
109. SIPPL, M J. Recognition of Errors in Three-Dimensional Structures of Proteins. **Proteins** v.17, p.355-362, 1993.
110. SONGYANG Z; FANNING, A S; FU, C; XU, J; MARFATIA, S M; CHISHTI, A H; CROMPTON, A; CHAN, A C; ANDERSON, J M e CANTLEY, L C. Recognition of Unique Carboxyl-Terminal Motifs by Distinct PDZ Domains. **Science**, v.275, p.73-77, 1997.
111. STEWART, A C. Intratype Variation in 12 Human Papillomavirus types: a worldwide perspective. **J Virol**, v.70, n.5, p.3127-3136, 1996.
112. STOREY, A; THOMAS, M; KALITA, A; HARWOOD, C; GARDIOL, D; MANTOVANI, F; BREUER, J; LEIGH, I M; MATLASHEWSKI, G e BANKS, L. Role of a p53 polymorphism in the development of human papillomavirus-associated cancer. **Nature**, v.393, n.6682, p.229-234, 1998.
113. STOTE, R H e KARPLUS, M. Zinc Binding in Proteins and Solution: A Simple but Accurate Nonbonded Representation. **Proteins: Struct Func Gen**, v.23: p.12-31, 1995.
114. TAO, M; KRUHLAK, M; XIA, S; ANDROPHY, E e ZHENG, Z-M. Signals that dictate nuclear localization of HPV 16 oncoprotein E6 in living cells. **J Virol**, v.77, n.24, p.13232-13247, 2003.
115. TERMINE, N; PANZARELLA, V; FALASCHINI, S; RUSSO, A; MATRANGA, D; MUZIO, L L e CAMPISI, G. HPV in oral squamous cell carcinoma vs head and neck squamous cell carcinoma biopsies: a meta-analysis (1988–2007). **An Oncol**, v.19, p.1681–1690, 2008.
116. THERRIEN, J-P; DROUIN, R; BARIL, C e DROBETSKY, E A. Human cells compromised for p53 function exhibit defective global and transcription coupled nucleotide excision, whereas cells compromised for pRb function are defective only for global repair. **PNAS USA**, v.96, p.15038-15043, 1999.
117. THOMAS, M; BANKS, L. Human papillomavirus (HPV) E6 interactions with Bak are conserved amongst E6 proteins from high and low risk HPV types. **J Gen Virol**, v.80, p.1513–1517, 1999.
118. THOMAS, M; GLAUNSINGER, B; PIM, D; JAVIER, R e BANKS, L. HPV E6 and MAGUK protein interactions: determination of the molecular basis for specific recognition and degradation. **Oncogene**, v.20, p.5431-5439, 2001.

119. THOMAS, M; LAURA, R; HEPNER, K; GUCCIONE, E; SAWYERS, C; LASKY, L e BANKS, L. Oncogenic human papillomavirus E6 proteins target the MAGI-2 and MAGI-3 proteins for degradation. **Oncogene**, v.21, n.33, p.5088-5096, 2002.
120. ULLMAN, C G; HARIS, P I; GALLOWAY, D A; EMERY, V C e PERKINS, S J. Predicted  $\alpha$ -helix/ $\beta$ -sheet secondary structures for the zinc-binding motifs of human Papillomavirus E6 and E7 proteins by consensus prediction averaging and spectroscopic studies of E7. **Biochem J**, v.319(Pt.1), p. 229-239, 1996.
121. UVERSKY, VN; ROMAN, A; OLDFIELD, C J e DUNKER A K. Protein Intrinsic Disorder and Human Papillomaviruses: Increased Amount of Disorder in E6 and E7 Oncoproteins from High Risk HPVs. **J Proteome Res**, v.5, n.8, p.1829-1842, 2006.
122. VAETEEWOOTTACHARN, K; CHAMUTPONG, S; PONGLIKITMONGKOL, M e ANGELETTI, P C. Differential localization of HPV 16 E6 splice products with E6AP. **Viol J**, v.2, p.50, 2005.
123. VAN RANST, M A; TACHEZY R; DELIUS H e BURK RD. Taxonomy of the Human Papillomavirus. **Papillomavirus Rep**, v4, p.61-65, 1993.
124. VELDMAN, T ; HORIKAWA, I; BARRETT, J C e SCHLEGEL, R . Transcriptional Activation of the Telomerase hTERT Gene by Human Papillomavirus Type 16 E6 Oncoprotein. **J Virol**, n.75, p.4467-4472, 2001.
125. WEINER, S J; KOLLMAN, P A; CASE, D A; SINGH, U C; GHIO, C; ALAGONA, G; PROFETA JR. S e WEINER, P. A New Force Field for Molecular Mechanical Simulation of Nucleic Acids and Proteins. **J Am Chem Soc**, v.106, p.765-784, 1984.
126. WEINER, S J; KOLLMAN, P A; NGUYEN, D T e CASE, D A. An All Atom Force Field for Simulations of Proteins and Nucleic Acids. **J Comp Chem**, v.7, p.230-252, 1986.
127. WERNESS, B A; LEVINE A J e HOWLEY, P M. Association of Human Papillomavirus types 16 and 18 E6 proteins with p53. **Science**, v.248, p. 76-79, 1990.
128. WHITE, M F; GIRAUD-PANIS, M J; POHLER, JR e LILLEY, D M. Recognition and manipulation of branched DNA structure by junction-resolving enzymes. **J Mol Biol**, 1997; v.269, n.5, p.647-664.
129. WIEDERSTEIN, M e SIPPL, M J. Prosa-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids Research**, v.35, p.W407-W410, 2007.
130. XIE, H; VUCETIC, S; IAKOUCHEVA, L M; OLDFIELD, C J; DUNKER, A K; UVERSKY, V N e OBRADOVIC, Z. Functional anthology of intrinsic disorder. 1. Biological processes and functions of proteins with long disordered regions. **J Proteome Res**, v.6, n.5, p.1882-1898, 2007.
131. ZANIER, K; NOMINE, Y; CHARBONNIER, S; RUHLMANN, C; SCHULTZ, P; SCHWEIZER, J; e TRAVE, G. Formation of well-defined soluble aggregates upon fusion to MBP is a generic property of E6 proteins from various human papillomavirus species. **Prot Expr Purif**, v.51, n.1, p.59-70, 2007.
132. ZEHBE, I; VOGLINO, G; WILANDER, E; DELIUS, H; MARONGIU, A; EDLER, L; KLIMEK, F; ANDERSSON, S e TOMMASINO, M. p53 Codon 72 Polymorphism and Various Human Papillomavirus 16 E6 Genotypes Are Risk Factors for Cervical Cancer Development. **Cancer Res**, v.61 p.608-61, 2001.

133. ZHANG, J; ROSE, B R; THOMPSON, C H; JARRETT, C; RUSSELL, P; HOUGHTON, R S e COSSART, Y E. Associations between oncogenic human papillomaviruses and local invasive patterns in cervical cancer. **Gynecol Oncol**, v.57, p.170-177, 1995.
134. ZHANG, Y; DASGUPTA, J; MA, R Z; BANKS, L; THOMAS, M e CHEN, X S. Structures of a Human Papillomavirus (HPV) E6 polypeptide bound to MAGUK proteins: mechanisms of targeting tumor suppressors by a high-risk HPV protein. **J Virol**, v.81, n.7, p.3618-3626, 2007.
135. ZIMMERMANN, H; DEGENKOLBE, R; BERNARD, H-U; e O'CONNOR, M J. The HPB 16 E6 oncoprotein can downregulate p53 activity by targeting the transcriptional coactivator CBP/p300. **J Virol**, v.73, n.8, p.6209-6219, 1999.
136. zur HAUSEN, H. Papillomavirus infections - a major cause of human cancers. **Biochim Biophys Acta**, v.1288, n.2, p.F55-F78, 1996.

## Apêndice I – Arquivos de entrada para modelagem por homologia

### Arquivo de alinhamento para modelagem por homologia de E6Cwt

```
#
C; The template is a NMR ensemble. (X for X-ray and N for NMR structures
and M for Model)

>P1;2FK4_03
structureN: 2FK4_03 :3 :A :68 :A :E6C 4C4S :HPV type 16:...
SYSLYGTTLEQQYNKPLSDLLIRCINCQKPLSPEEKQRHLDKKQRFHNIIRGRWTGRCMSCSRSSRz*

>P1;6E6C
sequence: 6E6C :1 :A :66 :A :E6C WT :HPV type 16:...
CYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDKKQRFHNIIRGRWTGRCMSCCRSSRz*
#
```

### Arquivo de entrada da modelagem por homologia de E6C wt

```
#
# Addition of restraints to the default ones
from modeller import *
from modeller.automodel import * # Load the automodel class

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = ['. ', '../atom_files']

# Read in HETATM records from template PDBs
env.io.hetatm = True

a = automodel(env,
              alnfile = 'alignment_6E6C.ali', # alignment filename
              knowns = '2FK4_03', # codes of the
templates
              sequence = '6E6C') # code of the target
a.starting_model= 1 # index of the first model
a.ending_model = 20 # index of the last model
# (determines how many models to
calculate)
a.make() # do the actual homology modeling
#
```

### Arquivo de alinhamento para modelagem por homologia de E6Nwt

```
#
C; The template is a NMR ensemble. (X for X-ray and N for NMR structures
and M for Model)

>P1;2FK4_03
structureN:2FK4_03 :3 :A :68 :A :E6C 4C4S :HPV type 16:...
SYSLYGTTLEQQYNKPLSDLLIRCINCQKPLSPEEKQRHLDKKQRFHNIIRGRWTGRCMSCSRSSR.*

>P1;4E6N
sequence: 4E6N :1 :A :66 :A :E6N wt :HPV type 16:...
PRKLP--QLCTELQTTIHDIILECVYCKQQLLRREVVYDFAFRDLCIVYRDGNPYAVCDKCLKFYS.*
```

#

Arquivo de entrada da modelagem por homologia de E6N wt

```
#
# Addition of restraints to the default ones
from modeller import *
from modeller.automodel import *      # Load the automodel class

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = ['. ', '../atom_files']

# Read in HETATM records from template PDBs
env.io.hetatm = True

a = automodel(env,
              alnfile = 'alignment_4E6N.ali',      # alignment filename
              knowns = '2FK4_03',                # codes of the templates
              sequence = '4E6N')                 # code of the target
a.starting_model= 1                             # index of the first model
a.ending_model  = 30                             # index of the last model
                                                    # (determines how many models to
calculate)
a.make()                                         # do the actual homology modeling
#
```

Arquivo de alinhamento para modelagem por homologia de E6 completa selvagem

```
#
C; Pegando o Model orientado
>P1;4NC_orient
structureM:4NC_orient:9: :146: :E6_orient      :HPV type 16 C/N
orientado:..
-----PRKLPQLCTELQTTIHDIILECVYCKQQLLRREYDFAFRDLCIVYRDGNPYAVCDKCLKFYS----
----CYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDKKQRFHNIIRGRWTGRCMSSCRSSR-----
-..*

>P1;5CHI
sequence:5CHI:1      : :153 : :E6NC           :H. papilomavirus type 16:..
MFQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREYDFAFRDLCIVYRDGNPYAVCDKCLKFYSKISE
YRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDKKQRFHNIIRGRWTGRCMSSCRSSRTRRETQ
L..*
#
```

Arquivo de entrada da modelagem por homologia de E6 completa selvagem

```
#
# Homology modeling by the automodel class
from modeller import *                  # Load standard Modeller classes
from modeller.automodel import *       # Load the automodel class

log.verbose()      # request verbose output
env = environ()    # create a new MODELLER environment to build this model in

# directories for input atom files
env.io.atom_files_directory = ['. ', '../atom_files']
```

```

# Read in HETATM records from template PDBs
env.io.hetatm = True

a = automodel(env,
               alnfile = 'alignment_5NC.ali',      # alignment filename
               knowns   = ('4NC_orient'),         # codes of the templates
               sequence = '5CHI')                 # code of the target
a.starting_model= 1          # index of the first model
a.ending_model  = 20        # index of the last model
                                # (determines how many models to
calculate)
a.md_level = None           # No refinement of model

a.make()                    # do the actual homology modeling
#

```

### Exemplo dos arquivos de entrada do refinamento das Alças (*Loops*) de E6 completa selvagem

#### Alça 1

```

#
# Loop refinement of an existing model
from modeller import *
from modeller.automodel import *

log.verbose()
env = environ()

# directories for input atom files
env.io.atom_files_directory = ['.', '../atom_files']

# Read in HETATM records from template PDBs
env.io.hetatm = True

# Create a new class based on 'loopmodel' so that we can redefine
# select_loop_atoms (necessary)
class MyLoop(loopmodel):
    # This routine picks the residues to be refined by loop modeling
    def select_loop_atoms(self):
        # One loop from residue 1 to 10 inclusive
        return selection(self.residue_range('1:', '10:'))

m = MyLoop(env,
            inimodel='5CHI_03.pdb',      # initial model of the target
            sequence='5CHI')            # code of the target

m.loop.starting_model= 1          # index of the first loop model
m.loop.ending_model  = 20        # index of the last loop model
m.loop.md_level = refine.slow     # loop refinement method

m.make()
#

```

## Apêndice II – Arquivo de Saída do PROSCAN/PBIL

<<http://pbil.ibcp.fr/>>

NPS@ <<http://npsa-pbil.ibcp.fr>> is the IBCP <<http://www.ibcp.fr>> contribution to PBIL <<http://pbil.univ-lyon1.fr>> in Lyon, France

```
-----
[HOME <http://pbil.ibcp.fr/>] [NPS@
<http://npsa-pbil.ibcp.fr/cgi-
bin/npsa\_automat.pl?page=/NPSA/npsa\_server.html>]
[SRS <http://srs-pbil.ibcp.fr/>] [HELP
<http://npsa-pbil.ibcp.fr/cgi-
bin/npsa\_automat.pl?page=/NPSAHLN/npsahlp\_npsaindex.html>]
[REFERENCES
<http://npsa-pbil.ibcp.fr/cgi-
bin/npsa\_automat.pl?page=/NPSA/npsa\_references.html>]
[NEWS
<http://npsa-pbil.ibcp.fr/cgi-
bin/npsa\_automat.pl?page=/NPSA/npsa\_news.html>]
[MPSA <http://mps-pbil.ibcp.fr/>] [ANTHEPROT
<http://antheprot-pbil.ibcp.fr/>] [Geno3D <http://geno3d-pbil.ibcp.fr/>]
[SuMo <http://sumo-pbil.ibcp.fr/>] [Positions
<http://pbil.ibcp.fr/html/pbil\_ibcp\_Jobs.html>] [PBIL
<http://pbil.univ-lyon1.fr/>]
-----
```

Wednesday, August 27th 2008: Computer upgrade for increase speed of computation (see news

<[http://npsa-pbil.ibcp.fr/cgi-
bin/npsa\\_automat.pl?page=/NPSA/npsa\\_news.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_news.html)>)

```
-----
      Job *PROSCAN* (ID: 37082d74df58) is running on *NPS@* server (started
on 20090323-012118).
```

Results will be shown below. \*Please wait and don't go back.\*

```
-----
*In your publication cite : *
```

NPS@: Network Protein Sequence Analysis

TIBS 2000 March Vol. 25, No 3 [291]:147-150

Combet C., Blanchet C., Geourjon C. and Deléage G.

```
-----
```



PROSCAN result for : UNK\_185150

PROSITE </NPSA/npsa\_references.html#prosite> : Bairoch A., Bucher P. and Hofmann K. The PROSITE database, its status in 1997 Nucleic Acids Res. (1997) Jan 1;25(1):217-221

PROSCAN : PROSCAN has been developed at IBCP.

Protein UNK\_185150 using PROSITE.BASE as reference site file

MFQDPQERPRKLPQLCTELQTTIHDIIILECVYCKQQLLRREYDFAFRDLCIVYRDGNPYAVCDKCLKFYISKISE  
YRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQRHLDKKQRFHNIIRGRWTGRCMSSCRSSRTRRETQ  
L

Similarity percentage 70

-----  
cAMP- and cGMP-dependent protein kinase phosphorylation site.

Prosite access number: PS00004 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00004>>

Prosite documentation access number: PDOC00004 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00004>>

[RK] (2)-x-[ST].

Randomized probability: 1.572e-03 .

Site : 146 to 149 RRET. Identity.

-----  
Protein kinase C phosphorylation site.

Prosite access number: PS00005 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00005>>

Prosite documentation access number: PDOC00005 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00005>>

[ST]-x-[RK].

Randomized probability: 1.423e-02 .

Site : 133 to 135 TGR. Identity.

Site : 142 to 144 SSR. Identity.

Site : 145 to 147 TRR. Identity.

-----  
Casein kinase II phosphorylation site.

Prosite access number: PS00006 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00006>>

Prosite documentation access number: PDOC00006 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00006>>

[ST]-x(2)-[DE].

Randomized probability: 1.482e-02 .

Site : 22 to 25 TIHD. Identity.

Site : 86 to 89 TTLE. Identity.

Site : 145 to 148 TRRE. Identity.

-----  
Tyrosine kinase phosphorylation site.

Prosite access number: PS00007 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00007>>

Prosite documentation access number: PDOC00007 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00007>>

[RK]-x(2,3)-[DE]-x(2,3)-Y.

Randomized probability: min = 4.074e-04 max = 4.083e-04

Site : 26 to 32 iILECVY. 71% similarity found.

Site : 48 to 54 RDLcIVY. 72% similarity found.

Site : 86 to 92 tTLEQQY. 71% similarity found.

Site : 53 to 60 vYRDGNPY. 71% similarity found.

Site : 72 to 79 KISEYRHY. Identity.

Site : 77 to 84 RHYcYSLY. 72% similarity found.

Site : 25 to 32 dIILECVY. 71% similarity found.

Site : 72 to 79 KISEyRHY. 72% similarity found.

Site : 77 to 84 RHYCySLY. 72% similarity found.

Site : 85 to 92 gTTLEQQY. 71% similarity found.

Site : 52 to 60 iVYRDGNPY. 71% similarity found.

Site : 68 to 76 KFYSkISEY. 72% similarity found.

Site : 71 to 79 sKISEYRHY. 71% similarity found.  
-----

N-myristoylation site.

Prosite access number: PS00008 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00008>>

Prosite documentation access number: PDOC00008 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00008>>

G-{EDRKHPFYW}-x(2)-[STAGCN]-{P}.

Randomized probability: 1.397e-02 .

Site : 57 to 62 GNPYAV. Identity.

Site : 85 to 90 GTTLeQ. 74% similarity found.

Site : 130 to 135 GrWTGR. 88% similarity found.

Site : 134 to 139 GrCMSC. 88% similarity found.

-----  
 Bacterial regulatory proteins, araC family signature.

Prosite access number: PS00041 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00041>>

Prosite documentation access number: PDOC00040 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00040>>

[KRQ]-[LIVMA]-x(2)-[GSTALIV]-{FYWPGDN}-x(2)-[LIVMSA]-x(4,9)-[LIVMF]-x-  
 {PLH}-[LIV

MSTA)-[GSTACIL]-{GPK}-{F}-x-[GANQRF]-[LIVMFY]-x(4,5)-[LFY]-x(3)-[FYIVA]-  
 {FYWHCM}

-{PGVI}-x(2)-[GSADENQKR]-x-[NSTAPKL]-[PARL].

Randomized probability: min = 3.882e-08 max = 3.905e-08

Site : 14 to 55 QLCTeLQTtIHDIILECVyCKQQLLRREVDFAFRDLCiVyR. 76%  
 similarity found.

Site : 66 to 109 cLKfYskISEYRHycYsLYGTTLEQQYnkPLcDLLIRCINcQKP. 70%  
 similarity found.

Site : 11 to 55 KLPQLCTELQTTIHDIILECVyCKQQLLRREVDFAFRDLCiVyR. 86%  
 similarity found.

Site : 11 to 55 KLPQLCTELQTTIHDIILeCVyCkQQLLRREVDFAFRDLCiVyR. 71%  
 similarity found.

Site : 65 to 109 KcLKfySKISEYRHycYsLYGTTLEQQYnkPLcDLLIRCINcQKP. 72%  
 similarity found.

Site : 10 to 55 RkLPqLCTeLQTTIHDIILECVyCKQQLLRREVDFAFRDLCiVyR. 70%  
 similarity found.

-----  
 Thioredoxin family active site.

Prosite access number: PS00194 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00194>>

Prosite documentation access number: PDOC00172 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00172>>

[LIVMF]-[LIVMSTA]-x-[LIVMFYC]-[FYWSTHE]-x(2)-[FYWGTN]-C-[GATPLVE]-  
 [PHYWSTA]-C-{I}-x-{A}-x(3)-[LIVMFYWT].

Randomized probability: 4.976e-08 .

Site : 128 to 146 IrGrWTGrCmSCCRSSRTr. 70% similarity found.

-----  
 Lipocalin signature.

Prosite access number: PS00213 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00213>>

Prosite documentation access number: PDOC00187 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00187>>

[DENG]-{A}-[DENQGSTARK]-x(0,2)-[DENQARK]-[LIVFY]-{CP}-G-{C}-W-[FYWLRH]-{D}-  
 [LIVMTA].

Randomized probability: min = 1.027e-06 max = 1.029e-06

Site : 122 to 135 kQRFHNIRGRWtGr. 72% similarity found.

-----  
 Glycoprotein hormones beta chain signature 1.

Prosites access number: PS00261 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00261>>

Prosites documentation access number: PDOC00234 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00234>>

C-[STAGM]-G-[HFYL]-C-x-[ST].

Randomized probability: 1.634e-07 .

Site : 136 to 142 CMscCRS. 72% similarity found.

-----  
 Bowman-Birk serine protease inhibitors family signature.

Prosites access number: PS00281 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00281>>

Prosites documentation access number: PDOC00253 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00253>>

C-x(5,6)-[DENQKRHSTA]-C-[PASTDH]-[PASTDK]-[ASTDV]-C-[NDEKS]-[DEKRHSTA]-C.

Randomized probability: min = 2.824e-10 max = 2.827e-10

Site : 51 to 66 CIVYRDGNpyAVCDKC. 76% similarity found.

-----  
 Plant lipid transfer proteins signature.

Prosites access number: PS00597 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00597>>

Prosites documentation access number: PDOC00516 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00516>>

[LIVM]-[PA]-x(2)-C-x(1,2)-[LIVM]-x(1,2)-[LIVMST]-x-[LIVMFY]-x(1,2)-[LIVMF]-[STRD]-x(3)-[DN]-C-x(2)-[LIVM].

Randomized probability: min = 3.707e-10 max = 3.718e-10

Site : 12 to 36 LPQLCTELQTTIhDIIIECVyCKQq. 70% similarity found.

-----  
 C-type lectin domain signature.

Prosites access number: PS00615 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS00615>>

Prosites documentation access number: PDOC00537 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00537>>

C-[LIVMFYATG]-x(5,12)-[WL]-{T}-[DNSR]-{C}-{LI}-C-x(5,6)-[FYWLIVSTA]-[LIVMSTA]-C.

Randomized probability: min = 1.169e-08 max = 1.178e-08

Site : 30 to 51 CVYCKQQLLRREvYDFAFRdLC. 74% similarity found.

Site : 30 to 51 CVYCKQQLLRREvYDFAFRdLC. 74% similarity found.  
 Site : 80 to 106 CYSlyGTTLEQQyNkPlCDLLIRcInC. 74% similarity found.

-----  
 Heavy-metal-associated domain.

Prosite access number: PS01047 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS01047>>

Prosite documentation access number: PDOC00804 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00804>>

[LIVNS]-x-{L}-[LIVMFA]-x-C-x-[STAGCDNH]-C-x(3)-[LIVFG]-{LV}-x(2)-[LIV]-x(9,11)-[IVA]-x-[LVFYS].

Randomized probability: min = 4.650e-08 max = 4.660e-08

Site : 25 to 54 dIILECVyCKQQLlRReVYDFAFRDLCIVY. 77% similarity found.  
 Site : 25 to 55 dIILECVyCKQQLlRReVYDFAFRDLCIVYr. 70% similarity found.

-----  
 Ribosomal protein L31 signature.

Prosite access number: PS01143 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS01143>>

Prosite documentation access number: PDOC00880 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00880>>

[DES]-[IVT]-x(4)-H-[PT]-[FAVY]-[FYW]-[TISN]-x(9,13)-[GN]-[KRHNQ].

Randomized probability: min = 8.339e-09 max = 8.373e-09

Site : 72 to 94 kISEYRHycYSlyGTTLEQQYnk. 70% similarity found.

-----  
 Laminin-type EGF-like (LE) domain signature.

Prosite access number: PS01248 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS01248>>

Prosite documentation access number: PDOC00961 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC00961>>

C-x(1,2)-C-x(5)-G-x(2)-C-x(2)-C-x(3,4)-[FYW]-x(3,15)-C.

Randomized probability: min = 1.076e-11 max = 1.091e-11

Site : 51 to 80 CIVyRDGNPyAVCDKCLKFYSKISEYRHYC. 73% similarity found.

-----  
 Zinc finger PHD-type signature.

Prosite access number: PS01359 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PS01359>>

Prosite documentation access number: PDOC50016 <<http://www.expasy.ch/cgi-bin/prosite-search-ac?PDOC50016>>

C-x(1,2)-C-x(5,45)-[VMFLWIE]-x-C-x(1,4)-C-x(1,4)-[WYFVQHLT]-H-x(2)-C-x(5,45)-[WFLYI]-x-C-x(2)-C.

Randomized probability: min = 3.537e-16 max = 3.858e-16

Site : 30 to 106  
 CVYCKQQLLRREYDFAFRdLCIVYrDGNpyAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCI  
 NC. 72% similarity found.

Site : 30 to 106  
 CVYCKQQLLRREYDFAFRdLCIVYrDGNpyAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCI  
 NC. 72% similarity found.

Site : 30 to 106  
 CVYCKQQLLRREYDFAFRdLCIVYrDGNpyAVCDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCI  
 NC. 72% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEqQyNKPLCDLlIRCINCQKPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 70% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQqYnKPLCDLlIRCINCQKPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 70% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQyNkPLCDLlIRCINCQKPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 70% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYnKpLCDLlIRCINCQKPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 70% similarity found.

Site : 30 to 106  
 CVYCKQQLLRREYDFAFRDLCIVYRDGNPYaVCDKCLkfySkISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCI  
 NC. 72% similarity found.

Site : 30 to 106  
 CVYCKQQLLRREYDFAFRDLCIVYRDGNPYaVCDKCLKFySkIseYRHYCYSLYGTTLEQQYNKPLCDLLIRCI  
 NC. 75% similarity found.

Site : 30 to 106  
 CVYCKQQLLRREYDFAFRDLCIVYRDGNPYaVCDKCLKFYSkIseYRHYCYSLYGTTLEQQYNKPLCDLLIRCI  
 NC. 75% similarity found.

Site : 30 to 106  
 CVYCKQQLLRREYDFAFRDLCIVYRDGNPYaVCDKCLKFYSkIseYRHYCYSLYGTTLEQQYNKPLCDLLIRCI  
 NC. 72% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKpLCDLlIrCINCQkPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 70% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCdLLIRCINCQkPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 73% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCdLlIRCINCQkPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 70% similarity found.

Site : 30 to 106  
 CVYCKQQLLRREYDFAFRDLCIVYRDGNPYAVCdKCLKFYSKISEYrHYCYSLYGTTLEQQYNKPLCDLLIRCI  
 NC. 75% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCdLlIRCINCQkPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 70% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLlIRCINCQkPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 73% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCInCQkPLCPEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 73% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQkpLCpEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 71% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKpLCPeEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 71% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLcPEeEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 73% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLcpEEKQRHLDDKKQRFHNIRGRWTgRCM  
 SC. 71% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPeeKQrHLDDKKQRFHNIRGRWTgRCM  
 SC. 71% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEekQRhLDDKKQRFHNIRGRWTgRCM  
 SC. 71% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEekqRHLDKKQRFHNIRGRWTgRCM  
 SC. 71% similarity found.

Site : 63 to 139  
 CDKCLKFYSKISEYRHYCYSLYGTTLEQQYNKPLCDLLIRCINCQKPLCPEEKQrHLdKKQRFHNIRGRWTgRCM  
 SC. 73% similarity found.

-----  
 16 different patterns found.

PROSITE result file (text): [PROSITE </tmp/37082d74df58.PROSCAN>]

-----  
 /\*User :\* public@201.37.84.223. \*Last modification time :\* Mon Mar 23  
 01:21:28 2009. \*Current time :\* Mon Mar 23 01:21:28 2009/ / This service  
 is supported by Ministere de la recherche <<http://www.recherche.gouv.fr>>  
 (ACI IMPBio, ACC-SV13), CNRS <<http://www.cnrs.fr>> (IMABIO, COMI, GENOME)  
 and Région Rhône-Alpes <<http://www.cr-rhone-alpes.fr>> (Programme  
 EMERGENCE) . Comments <<mailto:pbiladm@ibcp.fr>>. /

### Apêndice III – Resumo da revisão sobre estudos mutacionais e de deleção em E6 de HPV 16.

**Tabela 9:** Tabela resumindo o levantamento de resultados obtidos por diversos autores em estudos mutacionais e de deleção utilizando a proteína E6 de HPV 16.

Região: Mutação/Deleção	Liga p53	Degrada p53	Liga a E6AP	Outros comentários	Referência(s)
Phe2Leu	NT	50%	Sim	• Replicação estável	Park et al, 2002.
Phe2Val	NT	Não	Sim	• hTERT: aumenta atividade • reduz meia vida de NFX-1 por degradação/ubiquitinação. • Sem replicação estável.	Liu et al, 1999; Gewin et al, 2004; Park et al, 2002.
Gln3-Arg8	NT	Não	NT	• AcMo (6F4) contra esta região inibe degradação de p53.	Masson et al, 2003.
ΔPhe2-Leu12	Não	NT	NT		Foster et al, 1994.
Arg8Ser/Pro9Ala/Arg10Thr	Não	não	30%	• hTERT: ativa atividade • reduz meia vida de NFX-1 por degradação/ubiquit.	Foster et al, 1994; Gewin and Galloway, 2001; Klingelhutz et al, 1996; Gewin et al, 2004; Park et al, 2002.
Arg8Gly/Arg10Gly/Lys11Glu *truncado	NT	NT	NT	• E*I e *II (41 e 48 aa apenas, respectivamente) • Rompe o NLS	Tao et al, 2003.
ΔPro9-Pro13	Não	Não	Não	• hTERT: Sem ativação	Foster et al, 1994; Gewin and Galloway, 2001.
ΔGlu18-Asp25	0%	0%	0%	• Ausência de transformação celular • Repressão ~40% da transativação de p53.	Elbel et al, 1997.
Leu19Ala	NT	Sim	NT		Nominé et al, 2006.
Ile26Ala	NT	Sim	NT		Nominé et al, 2006.
ΔGln35-Asp56	0%	NT	0%	• Transformação celular mediana • Repressão ~40% da transativação de p53.	Elbel et al, 1997.
Leu37Ser	NT	NT	Não	• Ativação de hTERT • Imortalização de MEC	Liu et al, 1999; Sekaric et al, 2008.
Phe45Val	29%	25%	NT		Foster et al, 1994.
Phe47Arg	0%	0%	Sim	• Modifica hélice H1' • Aumenta a afinidade por E6AP • Atua como dominante negativa com wt.	Nominé et al, 2006; Zanier et al, 2005; Ristriani et al, 2009.
Phe45Tyr/Phe47Tyr/Asp49His	não	Não	NT		Foster et al, 1994.



**Tabela 9: continuação.**

Região: Mutação/Deleção	Liga p53	Degrada p53	Liga a E6AP	Outros comentários	Referência(s)
Leu50Gly	NT	Não	Não	<ul style="list-style-type: none"> <li>• Não ativa hTERT</li> <li>• Sem imortalização de MEC</li> </ul>	Zimmermann et al, 1999; Sekaric et al, 2008; Shamanin et al, 2008.
Asp51Ala, Phe54Ala, Arg55Ala, Asp56Ala (158aa)	NT	Não	NT	<ul style="list-style-type: none"> <li>• DNA: Não afeta ligação à Junção de Holiday</li> </ul>	Ristriani et al, 2000.
Tyr54His	NT	Não	Sim	<ul style="list-style-type: none"> <li>• hTERT: aumenta atividade</li> <li>• Sem replicação estável.</li> <li>• Com replicação estável.</li> </ul>	Liu et al, 1999; Park et al, 2002.
Cys63Arg	NT	30%	50%	<ul style="list-style-type: none"> <li>• Mutação da Cys conservada – de ligação ao Zn</li> </ul>	Park et al, 2002.
Cys66Gly	Não	Não	NT	<ul style="list-style-type: none"> <li>• Mutação da Cys conservada – de ligação ao Zn</li> </ul>	Foster et al, 1994.
$\Delta$ Ser71-Gln90	0%	NT	0%	<ul style="list-style-type: none"> <li>• Transformação celular mediana</li> <li>• P53: Sem efeito sobre a transativação de p53</li> </ul>	Elbel et al, 1997.
$\Delta$ Ile73-Arg77	Não	Não	NT	<ul style="list-style-type: none"> <li>• Diminuição da estabilidade térmica da proteína.</li> <li>• Diminuição &gt; 50% na capacidade de ligação ao zinco</li> </ul>	Foster et al, 1994.
$\Delta$ Tyr76-Tyr79	NT	40%	NT		Liu et al, 2009.
$\Delta$ Asn105-Leu110	Fraco	NT	NT		Crook et al, 1991; Lechner et al, 1994.
Leu110Gln	NT	0%	0%	<ul style="list-style-type: none"> <li>• Sem replicação estável.</li> <li>• Imortalização de MEC.</li> <li>• Ativação de hTERT</li> </ul>	Park et al, 2002; Liu et al, 1999; Sekaric et al, 2008.
$\Delta$ Lys108-His126	0%	NT	0%	<ul style="list-style-type: none"> <li>• Ausência de transformação celular</li> <li>• P53: Sem efeito sobre a transativação de p53</li> </ul>	Elbel et al, 1997.
$\Delta$ Cys111-Lys115, Asn127Lys	Não	Não	NT		Foster et al, 1994.
Arg109Ala, Gln114Ala, Lys115Ala (158aa)	NT	NT	NT	<ul style="list-style-type: none"> <li>• DNA: ausência da ligação à Junção de Holiday</li> </ul>	Ristriani et al, 2000.
$\Delta$ His118-Lys122	Não	2-30%	30%	<ul style="list-style-type: none"> <li>• hTERT: Baixa ativação</li> </ul>	Foster et al, 1994; Dalal et al, 1996; Kiyono et al, 1998; Park et al, 2002.
$\Delta$ Gln123-Asn127	Não 30% (AcMo)	Não	Não (80% AcMo)	<ul style="list-style-type: none"> <li>• hTERT: Sem ativação</li> <li>• AcMo (3B8) contra esta região inibe em 70% a ligação com p53, mas degrada parcialmente p53.</li> <li>• AcMo contra esta região inibe em 20% a ligação com E6AP.</li> </ul>	Foster et al, 1994; Klingelutz et al, 1996; Liu et al, 1999 et al, Lagrange et al, 2005.
Asn127-Trp132	22% (AcMo)	NT	63% (AcMo)	<ul style="list-style-type: none"> <li>• AcMo (1F5) contra esta região inibe em 78% a ligação com p53.</li> </ul>	Lagrange et al, 2005.

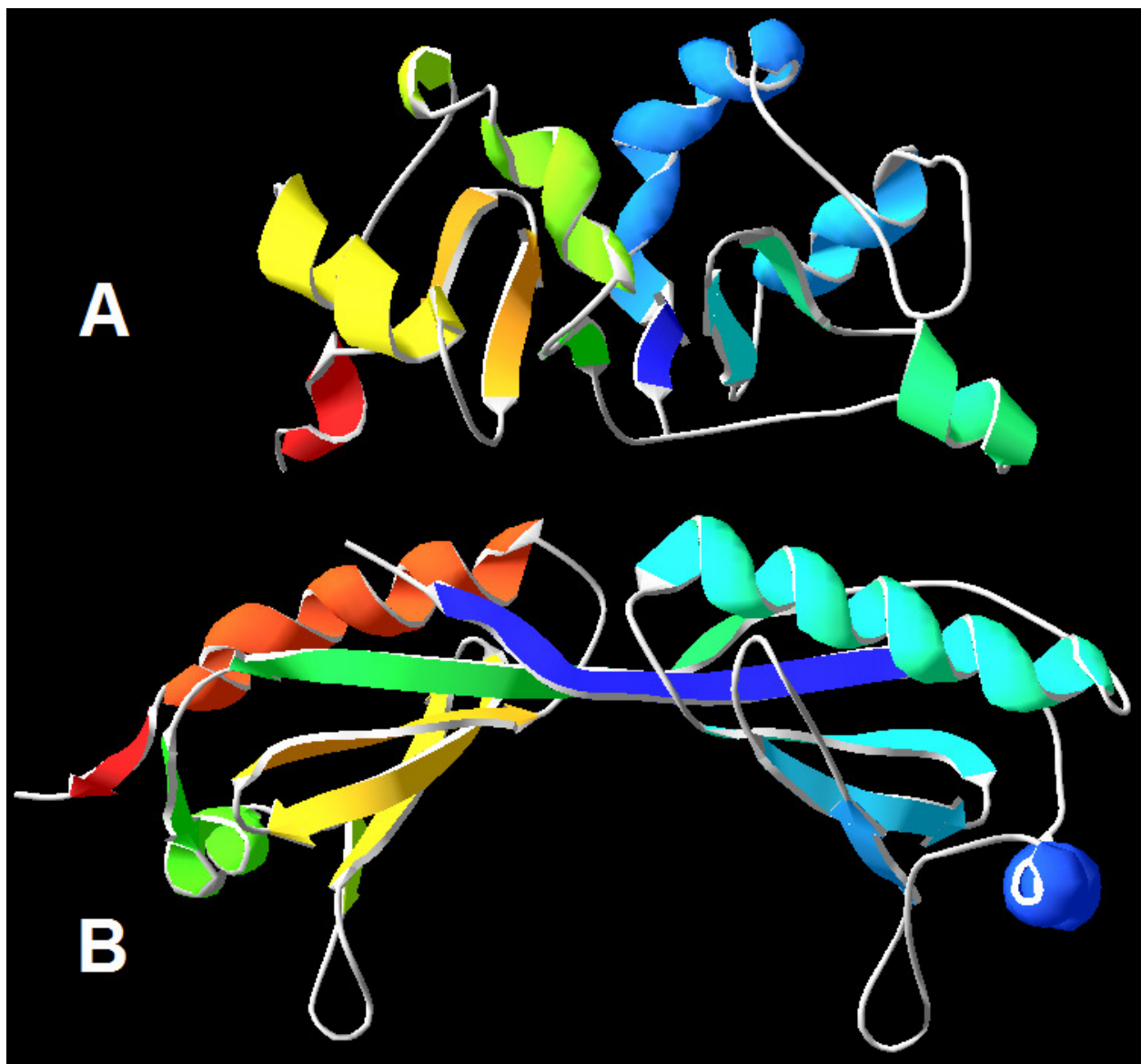
**Tabela 9: continuação.**

Região: Mutação/Deleção	Liga p53	Degrada p53	Liga a E6AP	Outros comentários	Referência(s)
Ile128Ala	NT	Sim	NT	• p53: degradação apenas <i>in vivo</i> , não <i>in vitro</i>	Nominé et al, 2006.
ΔIle128-Trp132	Não	Não	NT		Foster et al, 1994.
Gly130Val	NT	NT	Não	• Sem imortalização de MEC • Não ativa hTERT	Sekaric et al, 2008; Shamanin et al, 2008.
Thr133-Ser138	79% (AcMo)	não	7,3%(AcMo)	• AcMo (3F8) contra esta região inibe em 79% a ligação com p53.	Lagrange et al, 2005.
ΔCys140-Leu151	35%	25%	NT		Foster et al, 1994.
ΔSer142-Thr149	~60%	~83%	100%	• Ausência de transformação celular. • P53: Inibe da transativação de p53.	Elbel et al, 1997
ΔArg146-Leu151	33%	Sim (67%)	NT	• hTERT: ativa atividade	Foster et al, 1994; Klingelutz et al, 1996; Kiyono et al, 1998; Nominé et al, 2006
6C/6S Lys121Ala / Lys122Ala / Gln123Ala / Arg124Ala	NT	Sim	NT	• NLS – padrão de localização alterado • DNA: ausência de ligação à Junção de Holliday • P53: degradação apenas <i>in vivo</i> , não <i>in vitro</i>	Nominé et al, 2006; Le Roux e Moroianu, 2003; Ristriani et al, 2000.
Apenas porção C terminal E6C	NT	NT	NT	• DNA: Mantém atividade de ligação à junções de Holiday	Ristriani et al, 2000/2001.
ΔLeu151	NT	Sim	NT	• Ausência de transformação celular • PDZ: Ausência de ligação à hDLG • (C-term = HPV 18)	Kiyono et al, 1997.
Leu151Val	NT	Sim	NT	• Presença de transformação celular • PDZ: há ligação a hDLG	Kiyono et al, 1997.
Leu151Ile	NT	Sim	NT	• Presença de transformação celular • PDZ: há ligação a hDLG	Kiyono et al, 1997.
Leu151Phe	NT	Sim	NT	• Ausência de Transformação celular • PDZ: Ausência de ligação à hDLG	Kiyono et al, 1997.
Leu151Phe-Ile	NT	Sim	NT	• Ausência de Transformação celular • PDZ: Ausência de ligação à hDLG	Kiyono et al, 1997.
Leu151Glu-Thr-Gln-Val	NT	Sim	NT	• Presença de transformação celular • PDZ: há ligação a hDLG	Kiyono et al, 1997.

#### **Apêndice IV – TBP e E6: existem semelhanças?**

Durante a execução deste trabalho, nos deparamos com a proteína de reconhecimento do TATA Box (TBP – TATA Box Binding Protein) que parece ter similaridades estruturais com a E6. Apesar de as duas proteínas serem bastante diferentes tanto em relação à topologia quanto ao número de resíduos (151 para HPV16 E6 e 339 para hTBP) elas lembram bastante uma à outra. As duas estruturas podem ser vistas da Figura 30 A a F. A similaridade de topologia encontrada entre TBP (*TATA Box Binding Protein*) (código de acesso PDB: 1TGH) e E6 nos levou também a buscar outras características semelhantes entre as proteínas. Fizemos uma comparação entre as características de desordem intrínseca estrutural das duas proteínas ao longo de toda a sequência.

Ambas as proteínas são compostas por dois domínios complementares e apresentam a característica de apresentarem uma única folha beta torcida e inclinada que é formada de todas as suas fitas beta; terem todas as hélices alfa no mesmo lado da proteína formando o que parece ser um sulco e que tem como base a folha beta; e, além disto, os domínios carboxi- e amino- terminal são conectados por uma alça que se desprende de uma hélice no final domínio N-terminal que se encontra em uma das extremidades e se liga a uma fita no início do domínio C-terminal que está localizado no meio das proteínas. Uversky et al (2006), como já citado anteriormente, demonstrou que HPVs de alto risco tendem a maior desordenamento que os de baixo risco em vários dos parâmetros testados. Resolvemos comparar as estruturas E6 e TBP utilizando os parâmetros utilizados por Uversky, no intuito de identificar algum padrão na semelhança encontrada.



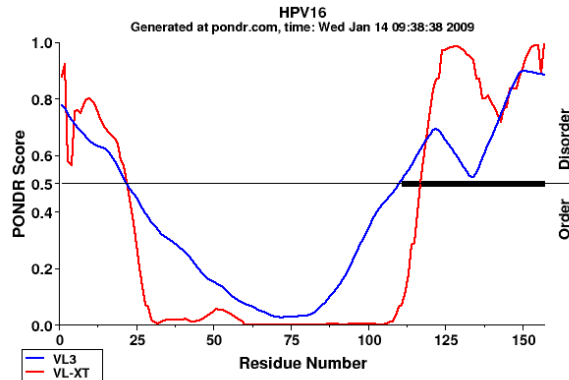
**Figura 30:** Representações do tipo *ribbons* dos modelos 3D de HPV16 E6 (A) e porção C-terminal (155-339) de hTBP (B). Nesta figura, nota-se as similaridades entre os modelos, como citado no texto.

Como pode ser visto na 30 A a H, apesar da similaridade visual encontrada entre as proteínas, as características de desordem intrínseca não são de todo semelhantes. No entanto, se considerarmos que falamos de duas proteínas compostas de 2 domínios espelhados e complementares, enxergamos certa simetria na presença de porções desordenadas. Nas Figuras 31 G e H estão os resultados da FoldIndex (Prilusky et al, 2005). Novamente, neste indicador de desordem

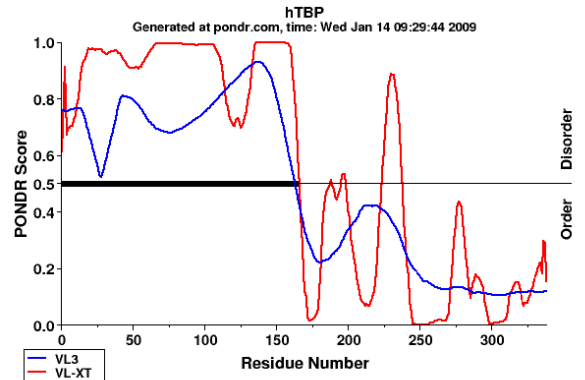
intrínseca baseado na sequência da proteína podemos ver uma certa simetria 'espelhada' entre as duas proteínas. Em E6 a porção C-terminal apresenta maior desordem, com alguns pontos da N-terminal com alguma tendência à desordem. Por sua vez, a hTBP apresenta a porção N-terminal com maior desordem e, da mesma forma, alguns pontos da C-terminal com alguma tendência à desordem. No entanto, como já citado, é a porção C-terminal de E6 que liga ao DNA, de forma independente da N-terminal, e a na TBP também é a porção C terminal que se liga ao DNA. Desta forma, a "simetria" não representa homologia de funcionalidade, mas as similaridades da estrutura terciária de E6 com a porção C-terminal de hTBP ainda pode fornecer alguma informação relevante.

Estas similaridades entre as duas proteínas poderiam indicar funcionalidades homólogas. Como já foi comentado antes, a E6 de HPV 16 se liga ao DNA com o domínio C terminal, independente do N terminal; e a TBP, como indica o nome, se liga ao DNA na região específica do TATA Box e assume uma conformação característica contorcida ao realizar a interação. Não existe nenhum estudo indicando o mesmo tipo de contorção da folha beta de E6 de HPV quando em contato com o DNA, nem indicando qual região da porção C-terminal que participa da interação.

A

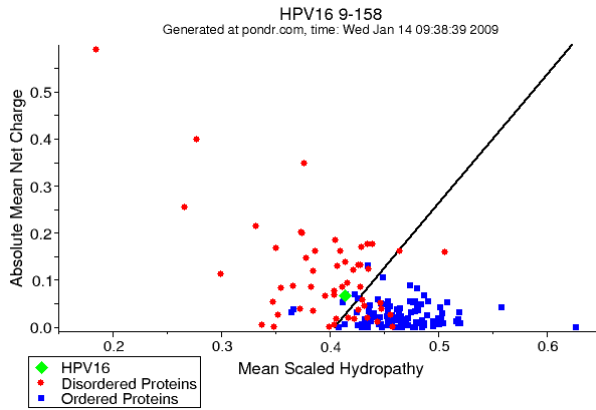


B



Predição de desordem intrínseca através dos parâmetros VL3 e VL-XT.

C



D

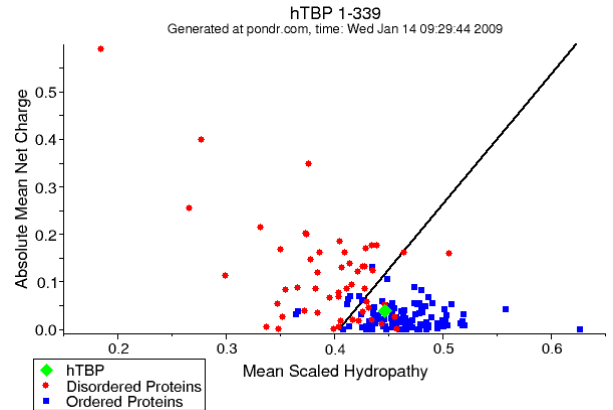
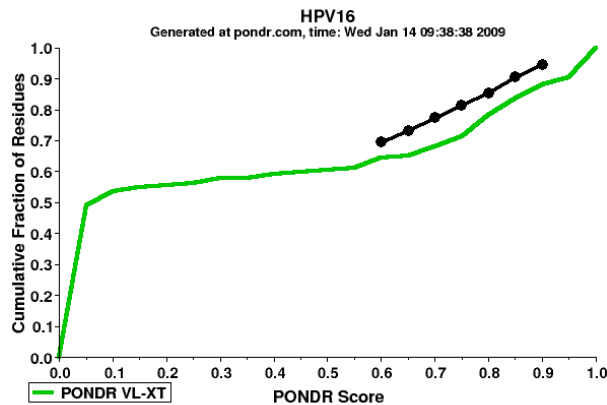
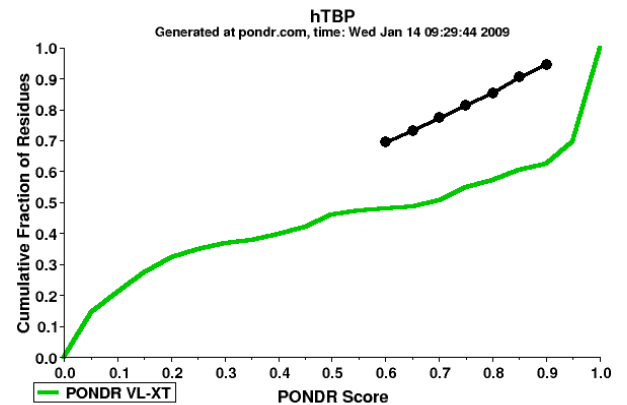


Gráfico de dispersão baseado no parâmetro de predição de desordem hidropatia média por carga absoluta média.

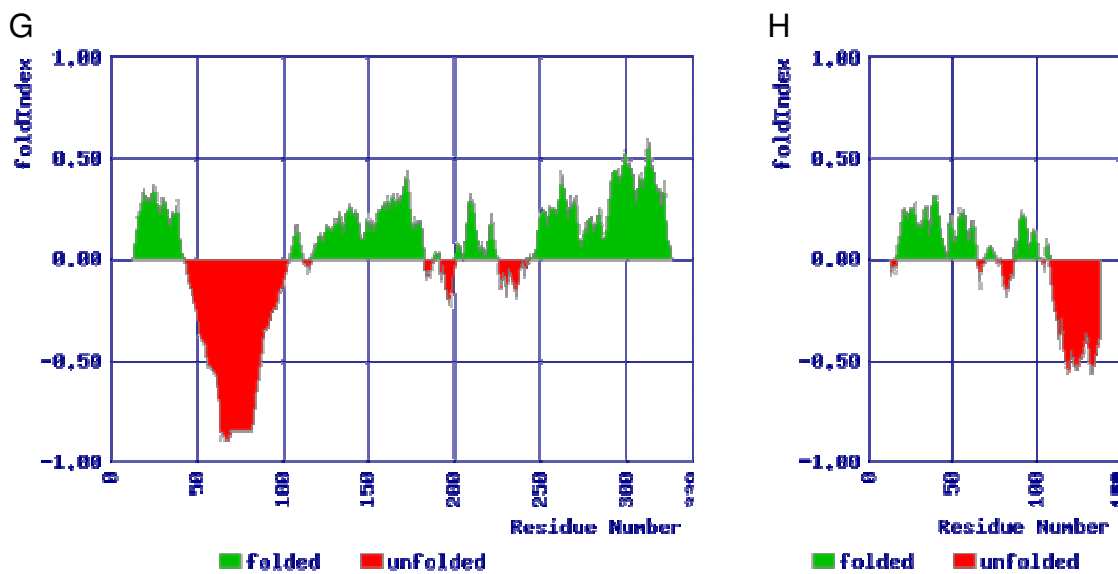
E



F



Análise da probabilidade de desordem intrínseca pelo Score de VL-XT pela contagem cumulativa de resíduos.



Análise de índice de dobramento da estrutura secundária FoldIndex com janela de 26 aminoácidos.

**Figura 31:** Comparação dos resultados de desordem intrínseca e FoldIndex entre E6 de HPV 16 e TBP. Cada comparação está acompanhada por uma legenda explicativa logo abaixo dos gráficos.