

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE GENÉTICA E BIOQUÍMICA
Curso de Biotecnologia

**MODELAGEM E DINÂMICA MOLECULAR DA ONCOPROTEÍNA E6 DO VÍRUS
DO PAPILOMA HUMANO (HPV) TIPO 18**

NATHÁLIA DE REZENDE COSTA NAGIB
NILSON NICOLAU JUNIOR

UBERLÂNDIA
2017

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE GENÉTICA E BIOQUÍMICA
CURSO BIOTECNOLOGIA

**MODELAGEM E DINÂMICA MOLECULAR DA ONCOPROTEÍNA E6 DO VÍRUS
DO PAPILOMA HUMANO (HPV) TIPO 18**

NATHÁLIA DE REZENDE COSTA NAGIB

Monografia apresentada à Coordenação do
Curso de Biotecnologia, da Universidade
Federal de Uberlândia, para obtenção do grau
de Bacharel em Biotecnologia

UBERLÂNDIA

2017

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE GENÉTICA E BIOQUÍMICA
CURSO DE BIOTECNOLOGIA

MODELAGEM E DINÂMICA MOLECULAR DA ONCOPROTEÍNA E6 DO VÍRUS
DO PAPILOMA HUMANO (HPV) TIPO 18

NATHÁLIA DE REZENDE COSTA NAGIB

NILSON NICOLAU JUNIOR
INSTITUTO DE GENÉTICA E BIOQUÍMICA

Homologado pela coordenação do
Curso de Biotecnologia em

__/__/__

EDGAR SILVEIRA CAMPOS

UBERLÂNDIA – MG

2017

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE GENÉTICA E BIOQUÍMICA
CURSO DE BIOTECNOLOGIA

MODELAGEM E DINÂMICA MOLECULAR DA ONCOPROTEÍNA E6 DO VÍRUS
DO PAPILOMA HUMANO (HPV) TIPO 18

NATHÁLIA DE REZENDE COSTA NAGIB

Aprovado pela Banca Examinadora em: __/__/__

Nota: ____

NILSON NICOLAU JUNIOR

UBERLÂNDIA ____ DE _____ DE 2017

SUMÁRIO

1. Introdução.....	1
1.1. Papiloma Vírus Humano e Câncer Cervical	1
1.2. Oncoproteína E6.....	3
1.3. Desordem Intrínseca	4
1.4. Bioinformática e Dinâmica molecular	5
2. Objetivo.....	6
3. Material e Métodos.....	6
3.1. Molde e Modelagem por homologia da Oncoproteína E6 do HPV tipo 18	6
3.2. Modelagem em <i>ab initio</i>	6
3.3. Validação dos Modelos	7
3.4. Refinamento da proteína E6	8
3.4.1. Modelagem de Rotâmetros e Inserção das Coordenadas do Zinco	8
3.4.2. Simulações de Dinâmica Molecular (DM).....	8
3.4.3. Análises das Trajetórias da Dinâmica Molecular.....	8
4. Resultados e Discussão.....	9
4.1. Molde e Modelagem por homologia da Oncoproteína E6 do HPV tipo 18	9
4.2. Modelagem em <i>ab initio</i>	11
4.3. Validação dos Modelos	12
4.4. Refinamento da proteína E6	15
4.4.1. Modelagem de Rotâmetros e Inserção das Coordenadas do Zinco	15
4.4.2. Simulações de Dinâmica Molecular (DM).....	15
4.4.3. Análises das Trajetórias da Dinâmica Molecular.....	16
9. Conclusão.....	19
10. Referências.....	20
APÊNDICES.....	24
APÊNDICE A:	24
APÊNDICE B	25
APÊNDICE C:	26

Lista de Abreviações

HPV	-----	Vírus do Papiloma Humano
OMS	-----	Organização Mundial da Saúde
NIC	-----	Neoplasia Intra-epitelial
NIC1	-----	Displasia Leve
NIC2	-----	Displasia Moderada
NIC3	-----	Displasia Grave/Carcinoma <i>In Situ</i>
DNA	-----	Ácido Desoxirribonucleico
P53	-----	Proteína de tumor 53
Rb	-----	Proteína Retino blastoma
E6	-----	Oncoproteína E6
E6AP	-----	Proteína associada E6 p53
PDB	-----	<i>Protein Data Bank</i>
IDP	-----	Proteínas Intrinsecamente Desordenada
DM	-----	Dinâmica Molecular
.FASTA	-----	Formato baseado em texto
CCD	-----	Algoritmo <i>Cyclic Coordinate Descent</i>
MQAPs	-----	Programas de Avaliação de Qualidade de Modelos
K	-----	Kelvins
ps	-----	pico segundos

LISTA DE FIGURAS

Figura 1: Interação E6 e E7 com a célula somática	3
Figura 2: Visualização por superfície do modelo 20 da modelagem por homologia da proteína E6 do HPV18.....	10
Figura 3: Gráfico de Ramachandran do modelo escolhido (Modelo 20) eixo X – ângulos PHI, eixo Y – ângulos PSI. Azul claro – Áreas Permitidas. Azul Ciano – Áreas Favorecidas. Branco áreas não permitidas.	11
Figura 4: Visualização por superfície do modelo obtido (modelo 427) pela modelagem <i>ab initio</i>.	12
Figura 5: Ramachandran do melhor modelo (427) eixo X – ângulos PHI, eixo Y – ângulos PSI. Azul claro – Áreas Permitidas. Azul Ciano – Áreas Favorecidas. Branco áreas não permitidas.	15
Figura 6: Visualização da proteína E6 do HPV 18 dentro da caixa de água TIP produzida para a Dinâmica Molecular	16
Figura 7: Gráfico de resultado do RMSD do modelo da E6 do HPV18.	17
Figura 8: Visualização da superfície da E6 do HPV18 após da dinâmica molecular representando o modelo mais próximo do modelo biológico	18
Figura 9: Visualização em Cartoon da proteína E6 da HPV18, representando as estruturas secundárias, após da dinâmica molecular representando o modelo mais próximo do modelo biológico	18

LISTA DE TABELAS

Tabela 1: Porcentagem de prevalência de HPV 16/18 em mulheres infectadas no Brasil.....	1
Tabela 3: Comparação das porcentagens de área favorecida e de área não favoráveis dos 20 modelos produzidos pelo modeller.	10
Tabela 4: Resultado dos Gráficos de Ramachandan das Validações dos 1000 modelos de acordo com os softwares de validação.....	12
Tabela 5: Resultados da validação dos 10 melhores modelos pelo software Errat.....	14
Tabela 6: Resultados da validação dos modelos a partir do software ProQ.....	14
Tabela 7: Comparação do modelo inicial com os frames do modelo refinado	19

Resumo

NAGIB, N.R.C.- Modelagem e Dinâmica molecular da oncoproteína E6 do vírus do papiloma humano (HPV) do tipo 18.

Um dos tipos de câncer que mais afeta mulheres no mundo é o câncer de colo de útero. Só no Brasil, existem 84,5 milhões de mulheres com risco de desenvolver o câncer cervical, isso pois a maioria dos cânceres cervicais desenvolvidos dependem do vírus do papiloma humano (HPV). Existem mais de 40 tipos de HPV que causam infecções cervicais podendo ser HPV de alto risco para câncer cervical e de baixo risco. A oncoproteína E6 é uma das proteínas essenciais para a formação do câncer de colo de útero, esta proteína é responsável por impedir a apoptose celular ligando-se no supressor de tumor p53, promovendo a imortalização celular e então a carcinogênese. Parte da modelagem da E6 foi feita por homologia baseada na proteína 4GIZ já a região N-terminal foi modelada por *ab initio*, visando refinar o modelo da E6 gerado. A região N-terminal é onde encontra-se a desordem intrínseca. As simulações de dinâmica molecular foram realizadas em 10ns em 5000 passos. Este trabalho tem como intuito a modelagem molecular baseada homologia e em *ab initio* da proteína E6 além da dinâmica molecular desta proteína relacionada ao HPV do tipo 18, um dos tipos ligados ao alto risco de desenvolver o câncer cervical. Como futuro esse trabalho serve como base para a busca de modelos farmacofóricos na busca de novos fármacos visando tratamento do câncer cervical pela HPV18

Palavras Chave: Modelagem Molecular, Modelagem em *ab initio*, Dinâmica Molecular, HPV18, E6, Desordem Intrínseca

1. Introdução

1.1. Papiloma Vírus Humano e Câncer Cervical

O HPV (papiloma vírus humano) é o nome dado a mais de cem tipos de vírus que afetam as mulheres de todo o mundo, sendo que quarenta destes, afetam o trato genital (DE VILLIERS et al., 2004). O HPV é o causador de 20-70% dos cânceres da orofaríngea, 5-30% dos cânceres de pescoço (LITWIN et al 2017), 88% dos cânceres anais (HARTWIG et al, 2015), 48% dos cânceres de pênis (BACKES et al 2009), 19% dos cânceres da vulva (HARTWIG et al, 2015) e 71% dos cânceres vaginais (HARTWIG et al, 2015).

Segundo a Organização Mundial da Saúde (OMS) de acordo com as informações do *HPV Information Centre* de Julho de 2017, no Brasil existem 84,5 milhões de mulheres com risco de câncer cervical, 18.503 casos anuais de câncer cervical tendo 8.414 casos de óbito (BRUNI et al, 2017). Os tipos de HPV que mais infectam mulheres no Brasil são os tipos 16 e 18, que estão no grupo de alto risco de câncer cervical (BRUNI et al, 2017). Dos casos de Câncer Cervical, 68,2% são causados por infecção por HPV dos tipos 16 e 18 (Tabela 1) e dos casos de alto número de lesões, 56,8% são causados por tipos 16 e 18.

O vírus HPV é sexualmente transmissível com maior incidência em mulheres do que em homens (BURCHELL et al, 2006). Por técnicas de biologia molecular foi detectado que a maioria das infecções de HPV são transitórias e não são detectadas por mais de 1-2 anos (TOTTIER; FRANCO, 2006). A probabilidade de transmissão é em média de 60% para o tipo 16 e 40% para os outros tipos do HPV, incluindo o HPV de tipo 18 (BURCHELL; RICHARDSON et al, 2006).

Tabela 1: Porcentagem de prevalência de HPV 16/18 em mulheres infectadas no Brasil

Tipo de Citologia	Prevalência HPV16/18 % (Média)
Citologia Normal	5,7
Baixo número de lesões	30,6
Alto número de lesões	56,8
Câncer Cervical	68,2

Fonte: OMS, 2017

O Câncer Cervical é um dos resultados da infecção propagada por HPV e é o segundo tipo de câncer mais comum entre as mulheres (PARKIN; BRAY, 2006). O Carcinoma Cervical Invasivo (CCI) tem como precursores lesões caracterizadas pelo distúrbio da maturação,

estratificação e atipia celular, que podem ser classificadas histologicamente e citologicamente como Neoplasia Intra-epitelial Cervical (NIC) ou como Lesões Intra-epiteliais Escamosas (LIE) segundo a terminologia de Bethesda (BALDWIN; LASKEY; COLEMAN, 2003; SOLOMON et al., 2002). A fase pré-maligna do câncer cervical pode ser detectada via exames como o Papanicolau e outros exames de células cervicais esfoliadas e pode ser confirmada via exames histológicos do material coletado. As alterações pré-malignas podem ser divididas em três espectros de anormalidade histológica de NIC1 (displasia leve), NIC2 (displasia moderada) e NIC3 (displasia grave/carcinoma *in situ*) e embora haja tratamento para esse tipo de alterações clínicas, não é eficiente, uma vez que os exames histológicos e citológicos não são capazes de diferenciar se a anormalidade tem a capacidade de se regredir, como ocorre na maioria dos casos, ou se tem capacidade de evoluir para um câncer cervical.

Existem as formas histológicas e citológicas de diagnóstico do HPV como o teste de Papanicolau, porém são testes falhos, sendo que, dependem do modo de coleta da amostra (ROCHA; MELO, 2010) e não detecta o HPV em si, mas sim as alterações citológicas causadas pela infecção viral (PIAS et al, 2009). Pode-se utilizar detecção baseada nos ácidos nucleicos como Southern Blot, Dot Blot, Microarray, PCR, Real-Time PCR e testes já patenteados como Cervista HPV, Captura Híbrida 2, Linear Array, COBAS, INNO-LiPA, CLART HPV2 e MCHA (ABREU et al, 2012)

Para o surgimento do câncer cervical, em grande parte dos casos diagnosticados, a mulher primeiro foi infectada pelo vírus HPV, apenas uma porcentagem da população infectada desenvolve lesões reconhecíveis clinicamente no curso da doença (KANODIA; FAHEY; KAST, 2007). A infecção causada por tipos deste vírus são essenciais para o aparecimento de lesões precursoras e do câncer cervical (ZUR HAUSEN et al., 1974), uma vez que, tem-se demonstrado que o HPV causa transformações das células epiteliais cervicais durante o processo infeccioso.

Os 40 tipos de HPV que infectam o trato genital podem ser divididos em dois grupos, os de alto e de baixo risco, baseados na progressão do câncer cervical (KLETER et al., 1998). Os de alto risco são relacionados a cânceres anogenitais, neste grupo estão os vírus dos tipos 16, 18, 31, 33, 45, 56 e são detectados em grande maioria dos cânceres cervicais e cerca de 80% destes cânceres são causados pelos tipos 16, 18, 45 e 31, sendo o tipo 18 responsável por 15,8% dos casos (CASTELLSAGUÉ, 2008).

1.2. Oncoproteína E6

A imortalização celular, transformação e carcinogênese dos vírus HPV de alto risco, devem-se as oncoproteínas E6 e E7 (NARISAWA-SAITO; KIYONO, 2007), uma vez que, estas estão em interação com uma grande quantidade de alvos celulares (BOULET et al., 2007) e têm como principais alvos os genes supressores de tumores p53 e pRb, sendo a E6 responsável por suprimir a apoptose celular enquanto a E7 promove a replicação celular.

A mais importante função da oncoproteína E6 é a ligação que faz com a proteína supressora de tumor p53, uma proteína reparadora do DNA que é responsável pela apoptose (figura1) (BOULET et al., 2007). A proteína associada a E6 (E6-AP), recruta o complexo ubiquitina de enzimas, ubiquitinando lisinas na p53 e iniciando sua proteólise. A degradação da p53 ignora os sinais de crescimento normal sendo a primeira causa de instabilidade cromossomal, com consequências mutacionais para células HPV positivas aumentando a integração de DNA exógeno no genoma da célula hospedeira (THOMAS; PIM; BANKS, 1999).

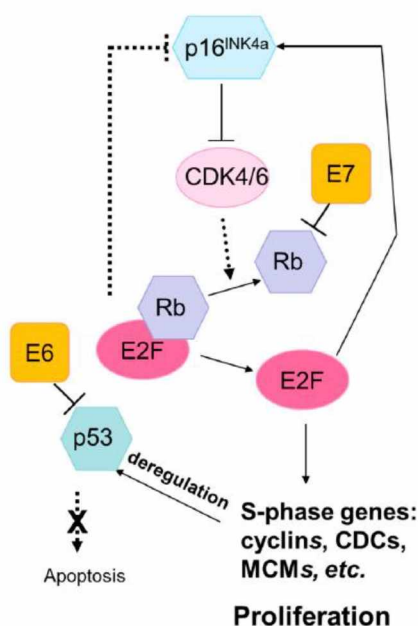


Figura 1: Interação E6 e E7 com a célula somática

Fonte: Litwin et al, 2017

A expressão das oncoproteínas E6 e E7 são as grandes causadoras da evolução da infecção pelo HPV para a formação do câncer cervical. O tratamento ainda é um grande desafio para a medicina moderna. Logo, a prevenção da infecção é a maior tática utilizada para diminuição dos casos e na atualidade a prevenção mais utilizada é a vacina contra os tipos de

alto risco deste vírus, como os 18 e 16. Porém, esta vacina é de alto custo e países como o Brasil decidiram vacinar mulheres de 9 a 13 anos a partir de 2014, a partir de 2016 foi permitido meninos entre 9 e 15 anos e meninas de 14 e 15 anos de se vacinar e em 2017 foi permitido jovens de até 26 anos de se vacinar. (OMS, 2017)

Um dos grandes desafios acerca do câncer cervical é como tratar a doença de maneira mais eficaz do que se tem feito e evitar a morte. Este desafio tem sido abordado por várias frentes de pesquisa sendo levado em consideração tanto a genômica quanto a proteômica como a busca de ligantes das oncoproteínas E6 e E7 e é indicado por muitos estudiosos como uma possível estratégia contra o HPV. Foram apresentados por Cherry et al. (2013) avanços na busca de inibidores da função da proteína E6 do HPV tipo 16, mas ainda se têm pouco em relação a inibição das funções das proteínas E6 e E7 do HPV tipo 18 com o intuito de tratamento de pacientes com câncer cervical.

1.3. Desordem Intrínseca

Durante muitos anos, o consenso científico era de que as proteínas tinham estruturas bem definidas, sendo que, isso permitia saber sua função devido a suas estruturas, durante esse período uma das funções mais bem descritas foi a catalise, do qual Fischer em 1929 criou o método “Chave e Fechadura” do qual depende da exata estrutura da proteína enzimática para que esta se ligue no substrato de forma eficiente. (HABCHI et al, 2014)

No fim do século XX, foi criado o *Protein Data Bank (PDB)* que confirmou pela visualização 3D que as proteínas eram estáticas, porém as proteínas depositadas apresentavam uma falha na densidade eletrônica e várias apresentam comportamento diferente do esperado em purificação. Essas particularidades fizeram que pesquisadores estudassem sobre a desordem intrínseca das proteínas. (HABCHI et al, 2014)

Em 1998, Romero et al. descobriu que 15,000 proteínas armazenadas no *SwissProt* apresentavam a desordem intrínseca em regiões de até 40 aminoácidos. Já em 1999 Wright e Dyson revelaram que grande parte das sequencias genicas são desordenadas em condições naturais e funcionais.

As proteínas intrinsecamente desordenadas (IDPs) ou intrinsecamente desnaturadas (IUPs) possuem certos atributos específicos para que a desordem aconteça: a baixa hidrofobicidade impedindo a formação de um núcleo globular estável e carga líquida, que cria repulsões eletrostáticas favorecendo os estados estendidos. (DUNKER et al., 2001).

Uma das características mais importantes da oncoproteína E6 do HPV18 é a desordem intrínseca da proteína na cauda N-Terminal em HPV de alto risco. Essa desordem acontece com

mais intensidade quando maior for o risco da formação do câncer de colo de útero. (Uversky *et al.* 2006). A formação das IDPs são ocorre a de mal dobramentos das proteínas, a partir dos erros de identificação e sinalização do alvo e a partir de surgimento de dobramentos não nativos ou não naturais. (UVERSKY *et al.*, 2008).

1.4. Bioinformática e Dinâmica molecular

A bioinformática é uma ferramenta relativamente nova da biotecnologia, esta surgiu com o intuito de analisar proteínas e genes de forma singular e única, permitindo observar as estruturas tridimensionais. No geral, a bioinformática é a utilização de ferramentas *in silico* para estudo e solução de problemas biológicos com aplicações relacionadas principalmente à saúde humana, como por exemplo, a criação de novos fármacos. Existem dois tipos principais de Bioinformática: a tradicional, que se relaciona a predição e resolução de problemas relacionados aos aminoácidos e nucleotídeos; e a estrutural, que se relaciona principalmente a modelagem molecular.

As informações necessárias para a bioinformática são normalmente indexadas em banco de dados, que são bibliotecas *in silico* contendo todas as informações relacionadas ao assunto. No caso da bioinformática as *databases* mais utilizadas são o *GeneBank* para nucleotídeos e o *Protein Data Bank* utilizado principalmente para estruturas proteicas.

A dinâmica molecular (DM) é uma das técnicas de bioinformática utilizadas neste trabalho. A Dinâmica consiste em descrever a variação da conformação molecular pelo tempo. Durante a dinâmica se varia seu conteúdo de estrutura secundária, a orientação das cadeias laterais, a conformação de alças e a energia de interação entre as diferentes moléculas.

A Dinâmica oferece um modelo mais próximo do modelo biológico real, pois utiliza características como a flexibilidade molecular em relação ao tempo e sendo que, a maioria dos processos biológicos depende dessa flexibilidade molecular. Para realizar a simulação por dinâmica molecular é necessário estabelecer algumas características como: o campo de força, o modelo da caixa de água e seus parâmetros, as condições periódicas de contorno, o tempo de integração e o cálculo de interações de átomos não ligados.

A oncoproteína E6 do HPV tipo 18 foi modelada em partes, ou seja, cada domínio da proteína foi modelada *in silico* a partir de modelos diferentes, sendo assim, esta proteína não apresenta um modelo refinado e próximo da realidade biológica, a dinâmica é utilizada então para aproximar este modelo ao modelo biológico real.

2. Objetivo

Este trabalho teve por objetivo realizar a modelagem *in silico* da estrutura tridimensional da oncoproteína E6 do HPV do tipo 18 por metodologias baseadas em homologia e *ab initio*, fazer a validação e visualização da proteína E6 gerada, por fim, refinar a estrutura modelada por meio de modelagem de rotâmetros e dinâmica molecular.

3. Material e Métodos

3.1. Molde e Modelagem por homologia da Oncoproteína E6 do HPV tipo 18

A modelagem da oncoproteína E6 foi feita por partes, cada domínio foi modelado de forma diferente, utilizando diferentes moldes. Isso ocorreu porque não se conhecia um molde que abrange proteína por inteiro, com isso os domínios foram modelados e unidos formando o modelo conhecido da proteína E6.

Para a seleção de moldes para a Oncoproteína é necessário a recuperação da mesma via o banco de dados proteicos UniProt (www.uniprot.org). A proteína será então parcialmente modelada no domínio C-Terminal por meio do método de modelagem por homologia usando como molde a estrutura obtida por cristalografia de raio-X da oncoproteína E6 do HPV tipo 16 em complexo com o peptídeo LXXLL presente na proteína E6AP (PDB ID: 4GIZ) (ZANIER et al., 2013) extraído do banco de dados de proteínas (PDB) a modelagem então foi realizada por meio do programa MODELLER (v9.21) (SALI; BLUNDELL, 1993).

Foram gerados 20 modelos a partir da modelagem em homologia do domínio C-terminal da proteína E6 a partir da 4GIZ, estes modelos foram então analisados por meio do gráfico de Ramachandran no software RAMPAGE (KUMAR et al, 2011). A partir da análise das regiões permitidas e não permitidas no gráfico de Ramachandran, foi escolhido o modelo com as melhores porcentagens.

3.2. Modelagem em *ab initio*

A modelagem do domínio n-terminal da E6 do HPV tipo 18 foi realizada por meio da modelagem *ab initio* via algoritmo “*cyclic coordinate descent*” (CCD) (CANUTESCU; DUNBRACK, 2003) por meio do programa Rosetta 3.1 (ROHL et al., 2004), foram gerados 1000 modelos, do qual apenas um será escolhido para avaliação por programas específicos e avaliação visual.

Para a realização da modelagem *ab initio* é criada uma linha de comando (apêndice B) citando o local do modelo, o tamanho dos fragmentos e o arquivo dos loops, do qual é inserido o arquivo de entrada, ou seja, onde se encontra o modelo a ser modelado, a quantidade de

modelos gerados e qual domínio foi modelado e o arquivo de saída, no qual se detalha onde os modelos foram armazenados.

3.3. Validação dos Modelos

Após a escolha do modelo mais apropriado é necessário realizar a validação e avaliação do mesmo. Para isso é necessário avaliar a qualidade dos modelos visualmente e também análises locais da sequência de aminoácidos uma validação global da estrutura 3-dimensional da proteína gerada. Utiliza-se então programas baseados em biofísica, levando em consideração a análise de forças intermoleculares e intramoleculares das partículas formadoras das moléculas proteicas conhecido como MQAP's (Programas de Avaliação de Qualidade de Modelos). Primeiro, serão utilizados ModFOLD (MCGUFFIN, 2008), dDFIRE (YANG; ZHOU, 2008) e QMEAN (BENKERT; TOSATTO; SCHOMBURG, 2008) para selecionar o melhor modelo fornecido dos 1000 modelos gerados anteriormente por meio das técnicas de modelagem por homologia e *ab initio*; sendo que o ModFOLD faz comparações dos alinhamentos das estruturas predizendo local e globalmente a qualidade dos modelos. O dDFIRE baseia-se na orientação angular das interações dipolo-dipolo, trata cada átomo polar como um dipolo e a orientação do dipolo é definida pelos vetores de ligação que conectam o átomo polar com outros átomos pesados. E QMEAN faz estimativas de erro da estrutura global e local da proteína E6. Todos estes MQAPs são programas desenvolvidos usando diferentes funções de energia potencial.

Logo, as 10 estruturas mais bem pontuadas nos MQAPs continuarão em análise. O programa PROCHECK (LASKOWSKI et al., 1993) cria o gráfico Ramachandran que classifica os ângulos PHI e PSI em áreas permitidas, áreas favoráveis e áreas não favoráveis. Assim, serão utilizados outros MQAPs para escolher o modelo de melhor qualidade a partir dos 10 modelos já escolhidos. O primeiro dos programas agora utilizado é o Preditor de Qualidade Proteica (ProQ) (WALNNER; ELOFSSON, 2003) que utiliza dois medidores de qualidade LGscore (CRISTOBAL et al., 2001) e MaxSub (SIEW et al., 2000), que extrai características estruturais, baseado em redes neurais. ERRAT (COLOVOS, 1993) é o segundo programa a ser utilizado baseado nas análises entre estruturas altamente refinadas e as interações atômicas entre átomos não ligados comparados do modelo, ele diferencia determinadas regiões corretas e incorretas das estruturas proteicas.

3.4. Refinamento da proteína E6

3.4.1. Modelagem de Rotâmetros e Inserção das Coordenadas do Zinco

A modo de se obter uma modelagem de alta qualidade, é realizada a modelagem de Rotâmetros por meio do SCWRL4 (KRIVOV; SHAPOVALOV; DUNBRACK, 2009), as coordenadas dos átomos de zinco, encontrados primeiramente na 4GIZ (PDB) foram transferidas manualmente para o modelo usando o programa DS Visualizer 3.5 (Discovery Studio Visualizer, 2005).

3.4.2. Simulações de Dinâmica Molecular (DM)

A Dinâmica molecular foi realizada por meio do campo de força CHARMM27 (MACKERELL et al., 1998) implementado no programa NAMD 2.8 (PHILLIPS et al., 2005) e a simulação da proteína E6 tipo18 foi realizada durante 10 ns utilizando a temperatura constante a 310 K. O sistema foi solvatado numa caixa cúbica de água TIP3P com no mínimo de 10 Å de distância entre a borda e qualquer átomo da proteína e suas cargas foram balanceadas acrescentando ions sódio e cloro. A eletrostática de longa distância foi tratada por meio do método “Particle Mesh Ewald” (PMEM) e todas as ligações que envolvem átomos de hidrogênio foram fixadas por meio do algoritmo SHAKE (VAN GUNSTEREN; BERENDSEN, 1977).

Para reduzir interações intermoleculares desfavoráveis entre o solvente e a proteína, todo o sistema foi minimizado por 5000 passos, antes de se rodar a dinâmica. O sistema então foi gradativamente aquecido a temperatura de 310K nos primeiros 52 ps em condições normais de temperatura e pressão constante a 1atm. Os arquivos de topologia e parâmetros do CHARMM27 foram modificados com a inserção de resíduo cisteína na forma qual um átomo de metal substitui o hidrogênio ligado ao enxofre, chamada Tiolato, para que haja interação entre o enxofre da cisteína com o átomo de zinco (FOLOPPE et al., 2001). Todo esse processo foi realizado a fim de avaliar a convergência estrutural e estabilidade da estrutura da proteína E6 do HPV tipo 18.

3.4.3. Análises das Trajetórias da Dinâmica Molecular

A trajetória obtida na dinâmica molecular foi analisado pelo VegaZZ(PEDRETTI et al, 2004) do qual é um software que engloba todas as técnicas de modelagem, dinâmica molecular e visualização. Neste programa foi utilizado o protocolo para obter as trajetórias e avaliar a estabilidade da dinâmica molecular, o desvio da média quadrática (RMSD) dos átomos Cα.

4. Resultados e Discussão

4.1. Molde e Modelagem por homologia da Oncoproteína E6 do HPV tipo 18

Primeiramente, a sequência da proteína obtida no banco de dados de sequência de aminoácidos UNIPROT e então foram utilizadas durante a modelagem.

Foi utilizado o programa Modeller local pelo sistema operacional UBUNTU/LINUX (Alva et al., 2016) para a realização da modelagem por homologia. Para isso foi utilizado a sequência .fasta retirada do UNIPROT (PDB ID: VE6_HP18) e a estrutura molde obtida por cristalografia de raio-X da oncoproteína E6 do HPV tipo 16 em complexo com o peptídeo LXXLL presente na proteína E6AP (PDB ID: 4GIZ). Foi realizada uma busca por sequências similares no PDB, sendo que essa similaridade foi encontrada somente no domínio N terminal.

Após completar a modelagem por homologia, foi feita a validação dos 20 modelos criados a partir da criação de Gráficos de Ramachandran a partir do programa Procheck , que são gráficos que apresentam uma representação visual dos ângulos PSI e PHI e suas posições preferenciais e das estruturas alfa hélice e folha beta da proteína. Os ângulos PHI e PSI são os ângulos do carbono alfa de cada molécula da proteína.

Dos 20 modelos gerados pelo Modeller, foi realizado a análise a partir de uma tabela comparando a porcentagem dos resíduos nas regiões mais favoráveis e nas regiões não permitidas dos ângulos PHI e PSI (tabela 01). A partir desta comparação foi escolhido o modelo número 20 para seguir para a próxima modelagem:

A partir da análise dos gráficos de Ramachandran o modelo escolhido foi o modelo 20 (figura 2), do qual apresentou a melhor porcentagem de áreas favorecidas, ou seja, o modelo 20 apresentou mais resíduos com ângulos PHI e PSI mais energeticamente favoráveis para a formação da estrutura (figura 3), por mais que apresente uma porcentagem de erro, o modelo ainda foi o melhor para passar para as próximas etapas

Tabela 3: Comparação das porcentagens de área favorecida e de área não favoráveis dos 20 modelos produzidos pelo modeller.

Gráfico de Ramachandran		
Modelo	% Áreas Favorecidas	% Áreas Não Permiridas
1	90,4%	0,7%
2	88,4%	0,7%
3	91,1%	0,0%
4	88,4%	0,0%
5	87,0%	0,0%
6	89,7%	0,7%
7	89,7%	0,0%
8	89,7%	0,7%
9	88,4%	0,0%
10	89,0%	0,0%
11	89,0%	0,7%
12	87,7%	0,0%
13	88,4%	0,7%
14	87,7%	0,0%
15	89,0%	0,0%
16	88,4%	0,0%
17	89,0%	0,7%
18	90,4%	0,0%
19	89,0%	0,7%
20	96.80%	0,6%

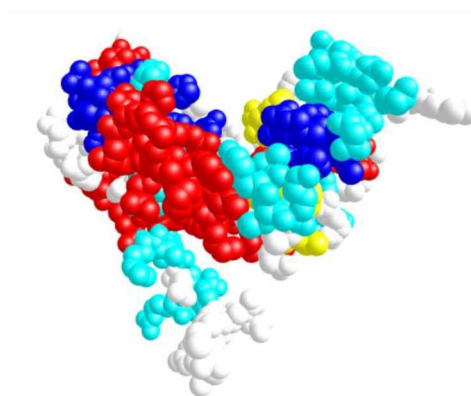


Figura 2: Visualização por superfície do modelo 20 da modelagem por homologia da proteína E6 do HPV18

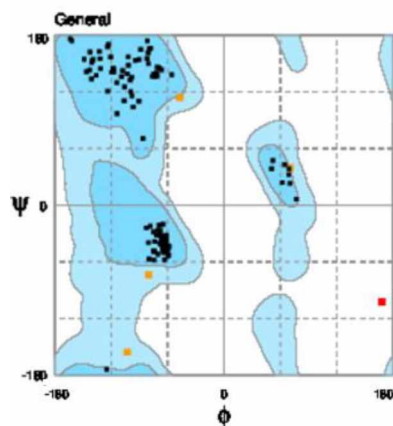


Figura 3: Gráfico de Ramachandran do modelo escolhido (Modelo 20) eixo X – ângulos PHI, eixo Y – ângulos PSI. Azul claro – Áreas Permitidas. Azul Ciano – Áreas Favorecidas. Branco áreas não permitidas.

4.2. Modelagem em *ab initio*

A porção N-terminal da E6 foi modelada por *ab initio* já que esta não possui molde conhecido no banco de dados PDB.

A Modelagem em *ab initio* foi feita a partir do programa Rosetta 3.1 (ROHL et al., 2004) no sistema operacional LINUX-UBUNTU utilizando o modelo escolhido com melhor Ramachandran das validações do modeller, do qual foi escolhido o modelo 20.

Embora a oncoproteína E6 do HPV18 tenha desordem intrínseca, os modelos obtidos a partir da modelagem em *ab initio* apresentaram cadeias secundárias e estruturas bem definidas a partir de visualizações por meio dos softwares VMD (HUMPHREY, DALKE, SCHULTEN, 1996) e Pymol (SHRODINGER, 2014).

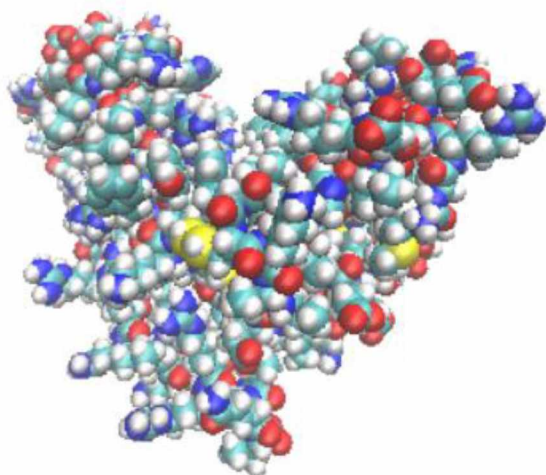


Figura 4: Visualização por superfície do modelo obtido (modelo 427) pela modelagem *ab initio*.

4.3. Validação dos Modelos

Para a validação dos 1000 modelos produzidos pela modelagem em *ab initio* pelo programa Rosetta, foram utilizados três programas: ModFOLD (MCGUFFIN, 2008), dDFIRE (YANG; ZHOU, 2008) e QMEAN (BENKERT; TOSATTO; SCHOMBURG, 2008). Para cada programa, foram escolhidos os dez melhores modelos, sendo estes analisados também por gráficos de ramachandran a partir do servidor RAMPAGE. Os dez melhores modelos obtidos dos gráficos de Ramachandran foram então analisados usando os softwares conhecidos como MQAPs: ERRAT e ProQ (Tabelas 5 e 6).

Tabela 4: Resultado dos Gráficos de Ramachandran das Validações dos 1000 modelos de acordo com os softwares de validação.

Gráfico de Ramachandran			
Modfold	% Áreas Favorecidas	% Áreas Permitidas	% Áreas Não Permitidas
427	99.4	0.6	0
594	98.7	1.3	0
682	98.7	1.3	0
361	98.1	1.9	0
360	98.1	1.9	0
739	98.7	1.3	0
48	98.7	1.3	0
708	98.7	1.3	0
287	98.1	1.9	0
850	98.7	1.3	0
dDFire	Score	allowed	error
215	98.7	1.3	0
899	96.8	3.2	0
185	98.7	1.3	0
241	99.4	0.6	0
633	98.1	1.9	0
774	96.8	3.2	0
52	97.4	2.6	0
64	98.7	1.3	0
168	98.7	1.3	0
177	96.2	3.8	0
Qmean	score	allowed	error
129	98.7	1.3	0
158	98.1	1.9	0
34	98.1	1.9	0
718	98.1	1.9	0
748	98.7	1.3	0
698	97.4	2.6	0
677	98.1	1.9	0
864	98.1	1.9	0
802	98.7	1.3	0
937	99.4	0.6	0

Dos modelos escolhidos para a realização do gráfico de Ramachandran foram escolhidos os dez modelos com melhores porcentagens de ângulos PHI e PSI na área favorecida. Os modelos escolhidos foram: 48, 64, 168, 241, 427, 682, 739, 748, 802, 937. Estes modelos foram então avaliados pelos softwares Errat e ProQ.

Tabela 5: Resultados da validação dos 10 melhores modelos pelo software Errat.

Errat	
Modelo	score
48	100
64	92.667
427	100
682	96.667
739	100
241	96.667
168	99.329
748	100
802	99.333
937	98.658

Tabela 6: Resultados da validação dos modelos a partir do software ProQ.

ProQ		
Modelo	Lgscore	MaxSub
48	-0.321	0.001
427	-0.312	0
682	-0.32	0.001
739	-0.363	-0.007
241	-0.29	-0.004
168	-0.315	-0.003
748	-0.356	0.003
802	-0.377	0.002
937	-0.369	-0.002
64	-0.27	0.003

De acordo com os melhores modelos apresentados por cada software de validação, foi escolhido o modelo de número 427 para seguir com a dinâmica, levando em consideração que este modelo além de obter a maior porcentagem de resíduos nas áreas favorecidas do gráfico de ramachandran (Figura 5), também obteve a melhor pontuação no software Errat (Tabela 5) e na pontuação do ProQ chamada MaxSub (Tabela 6).

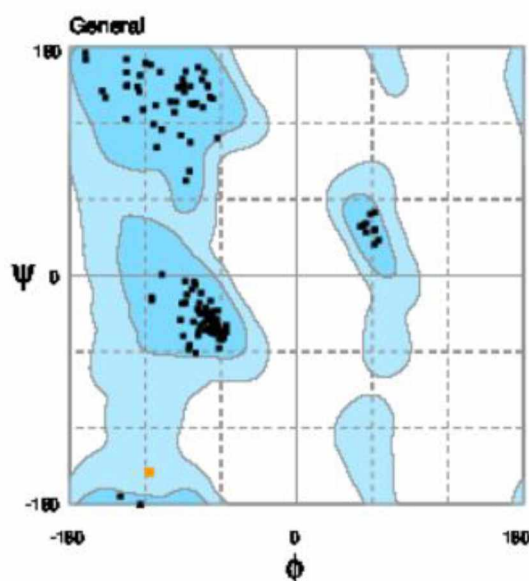


Figura 5: Ramachandran do melhor modelo (427) eixo X – ângulos PHI, eixo Y – ângulos PSI. Azul claro – Áreas Permitidas. Azul Ciano – Áreas Favorecidas. Branco áreas não permitidas.

4.4. Refinamento da proteína E6

4.4.1. Modelagem de Rotâmetros e Inserção das Coordenadas do Zinco

A modelagem de rotâmetros das cadeias laterais do modelo escolhido (modelo 427) foi realizada pelo software SCWRL4 (KRIVOV; SHAPOVALOV; DUNBRACK, 2009) visando refinar o modelo. Foi utilizado a estrutura 4GIZ (PDB) para a transferência das coordenadas dos átomos de zinco do dedo de zinco. A transferência foi realizada manualmente por homologia utilizando o programa DS Visualizer 3,5 (Discovery Studio Visualizer, 2005).

4.4.2. Simulações de Dinâmica Molecular (DM)

Para a Dinâmica Molecular foi utilizado o campo de força CHARMM27 (MACKERELL et al., 1998) incluído no programa NAMD (PHILLIPS et al., 2005). A simulação foi feita para 10ns, utilizando a temperatura constante de 310K e pressão de 1atm. A proteína E6 (modelo 427) foi colocado em uma caixa cúbica de água TIP3P com no mínimo 10 Å de distância entre a borda e os átomos da proteína. Foi adicionado íons de sódio e cloro para balanceamento de cargas. Para tratamento da eletrostática de longa distância foi usado o método PMEM (*Particle Mesh Ewald*). As Ligações que possuem hidrogênio foram fixadas com o software SHAKE (VAN GUNSTEREN; BERENDSEN, 1977). O sistema foi minimizado a 5000 passos para reduzir as interações desfavoráveis entre proteína e solvente.

A imagem abaixo representa a proteína E6 do HPV18 dentro da caixa de água produzida na dinâmica molecular a partir do programa VMD usando os algoritmos de NAMD, a imagem está dentro do esperado, ou seja, que a proteína se encontre dentro da caixa de água, criando assim, a partir das interações moleculares entre proteína e solvente, um modelo mais próximo da realidade biológica.

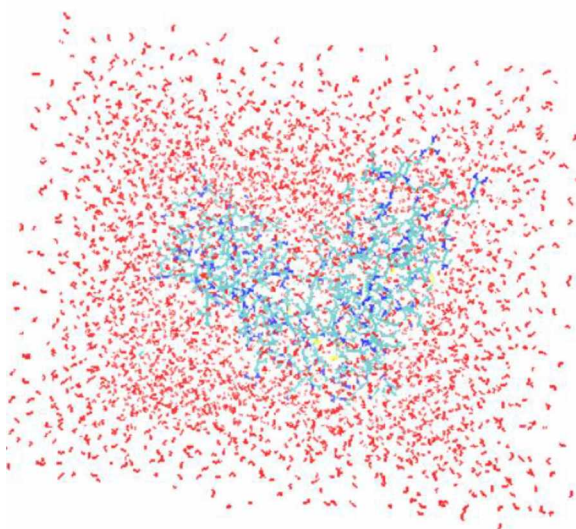


Figura 6: Visualização da proteína E6 do HPV 18 dentro da caixa de água TIP produzida para a Dinâmica Molecular

4.4.3. Análises das Trajetórias da Dinâmica Molecular

A análise das trajetórias foi realizada a partir do VegaZZ (PEDRETTI et al 2004) que foi utilizado para visualização da dinâmica, e também para o método de o desvio de média quadrática (RMSD), criando o gráfico abaixo.

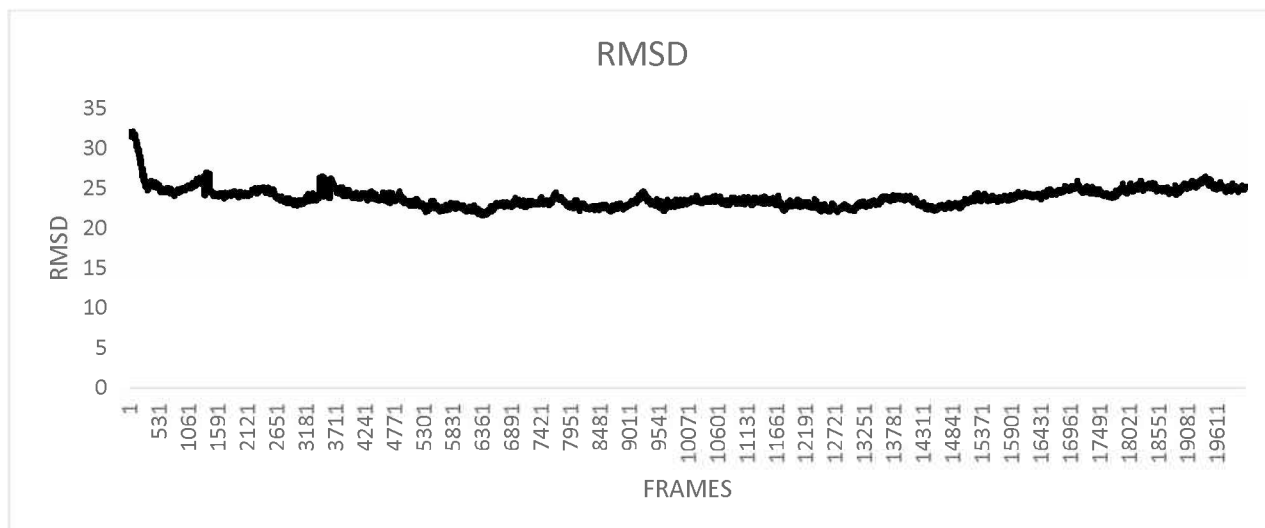


Figura 7: Gráfico de resultado do RMSD do modelo da E6 do HPV18.

O gráfico acima é produzido a partir do RMSD, que calcula a distância entre dois átomos da proteína, gerando o gráfico a partir destas distâncias, quando o gráfico atinge o platô as distâncias permanecem as mesmas assim o sistema chega na estabilidade. Pode-se observar que os primeiros 400 frames há uma queda no gráfico, essa queda representa a minimização do sistema, que permite a análise dos outros frames. Entre os primeiros 11000 frames, há diferenças de distância entre os átomos, causando então instabilidade no sistema, isso ocorre principalmente pelas áreas intrinsecamente desordenadas da proteína E6. Já depois dos 11000 primeiros frames pode-se observar o início do platô, com isso pode-se afirmar que o sistema é estável.

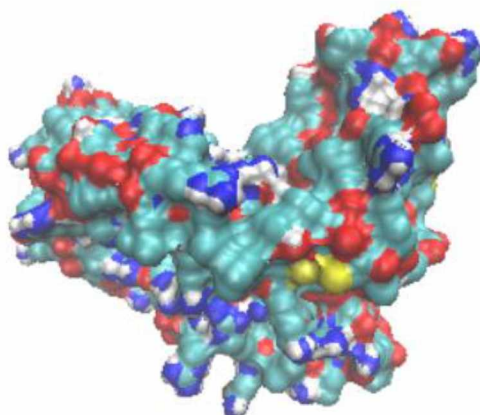


Figura 8: Visualização da superfície da E6 do HPV18 após da dinâmica molecular representando o modelo mais próximo do modelo biológico

A imagem acima representa a proteína E6 do HPV18 na visualização de superfície, do qual permite a visualização do sítio ativo da proteína indicando assim a área acessível para o solvente e macro e micromoléculas.

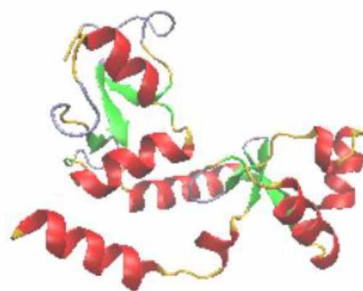


Figura 9: Visualização em Cartoon da proteína E6 da HPV18, representando as estruturas secundárias, após da dinâmica molecular representando o modelo mais próximo do modelo biológico

A imagem a seguir representa a visualização por cartoon, do qual as partes em vermelho representam as cadeias alfa hélice e as partes em verde representam as cadeias beta pareadas. O modelo de visualização em cartoon permite a análise das cadeias secundárias da estrutura.

Tabela 7: Comparação do modelo inicial com os frames do modelo refinado

Frame HPV18E6	Ramachandran			Errat	Verify 3D
	Áreas Favorecidas	Áreas Permitidas	Áreas de Erro		
Modelo 427	99.40%	0.60%	0.00%	96.667%	74.05%
Frame 10064	98.50%	1.50%	0.00%	87.826%	78.15%
Frame 20124	93.30%	6.00%	0.07%	96.330%	73.51%

Foi realizada uma comparação entre o modelo inicial, o frame central e o ultimo frame do modelo gerido pela dinâmica. O resultado do refinamento da proteína E6 não foi o esperado, sendo que, as melhores porcentagens foram obtidas pelo modelo inicial e não pelos refinados, com isso para realização de experimentos futuros deve ser utilizado o modelo sem o refinamento.

9. Conclusão

Neste estudo foi proposto a criação de modelos por meio da E6 do HPV18 mais estruturados a partir da modelagem e dinâmica molecular da proteína, com o intuito principal de modelar as áreas de desordem intrínseca na cauda N-terminal da proteína. Levando em consideração, o estudo teve sucesso em realizar a modelagem e na escolha de uma estrutura de qualidade e bem definida (modelo 427). Apesar do resultado no refinamento da proteína E6 do HPV18 foi inesperado, o estudo teve sucesso em promover a estabilidade do modelo após a dinâmica.

A continuidade deste trabalho é a realização da busca e validação de modelos farmacofóricos *in silico* para o desenvolvimento de novos fármacos para tratamento do câncer cervical.

10. Referências

- ABREU ALP, SOUZA RP, GIMENES F, CONSONARO MEL. A review of methods for detect human Papillomavirus infection. *Virol J.* 2012 Jan;9:262–71.
- ANDERSEN, C. A. F. et al. Continuum secondary structure captures protein flexibility. v. 10, n. 02, p. 175–184, 2002.
- BACKES, D.M.; KURMAN, R.J.; PIMENTA, J.M.; SMITH, J.S. Systematic review of human papillomavirus prevalence in invasive penile cancer. *Cancer Causes Control*, 2009.
- Litwin, T.R.; Clarke, M.A.; Dean, M.; Wentzensen, N. Somatic Host Cell Alterations in HPV Carcinogenesis. *Viruses* **2017**.
- BALDWIN, P.; LASKEY, R.; COLEMAN, N. improving cervical screening. **Nature reviews. Cancer**, v. 3, n. March, p. 1–10, 2003.
- BENKERT, P.; TOSATTO, S. C. E.; SCHOMBURG, D. QMEAN: A comprehensive scoring function for model quality assessment. **Proteins**, v. 71, n. 1, p. 261–77, abr. 2008.
- BOULET, G. et al. Human papillomavirus: E6 and E7 oncogenes. **The international journal of biochemistry & cell biology**, v. 39, n. 11, p. 2006–11, jan. 2007.
- BOWIE, J. U.; LUTHY, R.; EISENBERG, D. A method to identify protein sequences that fold into a known three-dimensional structure. v. 253, p. 164–170, 1991.
- BURCHELL NA, WINER RL, SANJOSÉ S, FRANCO EL. Chapter 6: Epidemiology and transmission dynamics of genital HPV infection. *Vaccine*. 2006.
- CANUTESCU, A. A.; DUNBRACK, R. L. Cyclic coordinate descent: A robotics algorithm for protein loop closure. **Protein Science**, v. 12, n. 5, p. 963–972, 2003.
- CASTELLSAGUÉ, X. Natural history and epidemiology of HPV infection and cervical cancer. **Gynecologic oncology**, v. 110, n. 3 Suppl 2, p. S4–7, set. 2008.
- CHERRY, J. J. et al. Structure Based Identification and Characterization of Flavonoids That Disrupt Human Papillomavirus-16 E6 Function. v.8, n. 12, p. 1–21, 2013.
- COLOVOS, C. Verification of protein structures: Patterns of nonbonded atomic interactions. **Molecular Biology**, v. 2, p. 1511–1519, 1993.
- CRISTOBAL, S. et al. A study of quality measures for protein threading models. **BMC Bioinformatics**, v. 2, n. 1, p. 5, 2001.
- DE VILLIERS, E.-M. et al. Classification of papillomaviruses. **Virology**, v. 324, n. 1, p. 17–27, 20 jun. 2004.
- FOLOPPE, N. et al. Structure, dynamics and electrostatics of the active site of glutaredoxin 3 from Escherichia coli: comparison with functionally related proteins. **Journal of Molecular Biology**, v. 310, n. 2, p. 449–470, jul. 2001.

GLYKOS, N. M. Software News and Updates Carma : A Molecular Dynamics Analysis Program. v. 27, p. 1765–1768, 2006.

HABCHI J, TOMPA P, LONGHI S, UVERSKY VN. Introducing Protein Intrinsic Disorder. *Chem Rev.* 9 de julho de 2014;114(13):6561–88.

HARTWIG, S.; BALDAULF, J.-J.; DOMINIAK-FELDEN, G.; SIMONDON, F.; ALEMANY, L.; DE SANJOSE, S.; CASTELLSAGUÉ, X. Estimation of the epidemiological burden of HPV-related anogenital cancers, precancerous lesions, and genital warts in women and men in Europe: Potential additional benefit of a nine-valent second generation HPV vaccine compared to first generation HPV vaccines. *Papillomavirus Res*, n. 1, p. 90–100, 2015.

HUMPHREY, W.; DALKE, A; SCHULTEN, K. VMD: visual molecular dynamics. **Journal of molecular graphics**, v. 14, n. 1, p. 33–8, 27–8, fev. 1996.

KANODIA, S.; FAHEY, L. M.; KAST, W. M. Mechanisms used by human papillomaviruses to escape the host immune response. **Current cancer drug targets**, v. 7, n. 1, p. 79–89, fev. 2007.

KLETER, B. et al. Novel Short-Fragment PCR Assay for Highly Sensitive Broad-Spectrum Detection of Anogenital. **American Journal of Pathology**, v. 153, n. 6, p. 1731–1739, 1998.

KRIVOV, G. G.; SHAPOVALOV, M. V; DUNBRACK, R. L. Improved prediction of protein side-chain conformations with SCWRL4. **Proteins**, v. 77, n. 4, p. 778–95, dez. 2009.

LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **J. Appl. Cryst.**, v. 26, p. 283–291, 1993.

LAZARIDIS, T.; KARPLUS, M. Effective energy functions for protein structure prediction. **Current opinion in structural biology**, v. 10, n. 2, p. 139–45, abr. 2000.

LITWIN, T.R.; CLARKE, M.A.; DEAN, M.; WENTZESEN, N. Somatic Host Cell Alterations in HPV Carcinogenesis. *Viruses* **2017**, 9, 206.

LOVELL S.C., DAVIS I.W., ARENDALL III W.B., BAKKER P.I.W., WORD J.M., PRISANT M.G., RICHARDSON J.S. e RICHARDSON D.C Structure validation by Calpha geometry: phi,psi and Cbeta deviation. *Proteins: Structure, Function & Genetics*. N. **50**, p. 437-450, 2002.

LUETHY, R.; BOWIE, J. U.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**, v. 356, p. 83–85, 1992.

MACKERELL, A. D. et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins †. v. 5647, n. 97, p. 3586–3616, 1998.

MCGUFFIN, L. J. The ModFOLD server for the quality assessment of protein structural models. **Bioinformatics (Oxford, England)**, v. 24, n. 4, p. 586–7, fev. 2008.

NARISAWA-SAITO, M.; KIYONO, T. Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: roles of E6 and E7 proteins. **Cancer science**, v. 98, n. 10, p. 1505–11, out. 2007.

PARKIN, D. M.; BRAY, F. Chapter 2: The burden of HPV-related cancers. **Vaccine**, v. 24 Suppl 3, p. S3/11–25, 31 ago. 2006.

PEDRETTI A., VILLA L., VISTOLI G. "VEGA - AN OPEN PLATFORM TO DEVELOP CHEMO-BIO-INFORMATICS APPLICATIONS, USING PLUG-IN ARCHITECTURE AND SCRIPT PROGRAMMING" [J.C.A.M.D.](#), Vol. 18, 167-173 (2004)

PIAS ADA, REGINA V, VARGAS A. Avaliação dos exames citológicos de papanicolaou com células epiteliais atípicas e respectivos exames colposcópicos com relação aos exames histopatológicos. *Rev Bras Análises Clínicas*. 2009;41(2):155–60.

PHILLIPS, J. C. et al. Scalable molecular dynamics with NAMD. **Journal of computational chemistry**, v. 26, n. 16, p. 1781–802, dez. 2005.

PONTIUS, J.; RICHELLE, J.; WODAK, S. J. Deviations from standard atomic volumes as a quality measure for protein crystal structures. **Journal of molecular biology**, v. 264, n. 1, p. 121–36, nov. 1996.

ROCHA GA, MELO VH. Biologia molecular no rastreamento das neoplasias cervicais uterinas. *Feminina*. 2010;38:1–6.

ROMERO, P.; OBRADOVIC, Z.; KISSINGER, C. R.; VILLAFRANCA, J. E.; GARNER, E.; GUILLIOT, S.; DUNKER, A. K. Pacific Symposium on Biocomputing Chemical Reviews Review 6583 [dx.doi.org/10.1021/cr400514h](https://doi.org/10.1021/cr400514h) | Chem. Rev. 2014, 114, 6561–6588 98; Altman, R. B., Dunker, A. K., Hunter, L., Klein, T. E., Eds.; World Scientific Publishing Co.: London, 1998; p 437.

SALI, A.; BLUNDELL, T. Comparative protein modelling by satisfaction of spatial restraints. **Journal of molecular biology**, v. 234, p. 779–815, 1993.

SCHRODINGER, LLC The PyMOL Molecular Graphics System, Version 2. 2014

SIEW, N. et al. MaxSub: an automated measure for the assessment of protein structure prediction quality. **Bioinformatics (Oxford, England)**, v. 16, n. 9, p. 776–85, set. 2000.

SOLOMON, D. et al. The 2001 Bethesda System: terminology for reporting results of cervical cytology. **JAMA : the journal of the American Medical Association**, v. 287, n. 16, p. 2114–9, 24 abr. 2002.

THOMAS, M.; PIM, D.; BANKS, L. The role of the E6-p53 interaction in the molecular pathogenesis of HPV. **Oncogene**, v. 18, n. 53, p. 7690–700, 13 dez. 1999.

TSOULOS, I. G.; STAVRAKLOUDIS, A. Eucb: A C++ program for molecular dynamics trajectory analysis. **Comput Phys Commun**, v. 182, n. 3, p. 8, mar. 2011.

VAN GUNSTEREN, W. F.; BERENDSEN, H. J. C. Algorithms for macromolecular dynamics and constraint dynamics. **Molecular Physics**, v. 34, p. 1311–1327, 1977.

VIKRAM A., SEUNG-ZIN N., SOLDING J., LUNPAS A. N.; The MPI bioinformatics Toolkit as an integrative platform for advanced protein sequence and structure analysis, *Nucleic Acids Research*, V. 44, n. 1, p. 410–415, 2016.

WALNNER, B.; ELOFSSON, A. Can correct protein models be identified? ". **Society**, p. 1073–1086, 2003.

WHIGHT, P. E.; DYSON, H. J. J. Mol. Biol. 1999, 293, 321.

WOODMAN, C. B. J.; COLLINS, S. I.; YOUNG, L. S. The natural history of cervical HPV infection: unresolved issues. **Nature reviews. Cancer**, v. 7, n. 1, p. 11–22, jan. 2007.

YANG, Y.; ZHOU, Y. Specific interactions for *ab initio* folding of protein terminal regions with secondary structures. **Proteins**, v. 72, n. 2, p. 793–803, ago. 2008.

ZANIER, K. et al. Structural basis for hijacking of cellular LxxLL motifs by papillomavirus E6 oncoproteins. **Science (New York, N.Y.)**, v. 339, n. 6120, p. 694–8, 8 fev. 2013.

ZUR HAUSEN, H. et al. Attempts to detect virus-specific DNA in human tumors. I. Nucleic acid hybridizations with complementary RNA of human wart virus. **International Journal of Cancer**, v. 13, n. 5, p. 650–656, 1974.

ZUR HAUSEN, H. Papillomaviruses Causing Cancer : Evasion From. **Journal of the National Cancer Institute**, v. 92, n. 9, p. 690–698, 2000.

APÊNDICES

APÊNDICE A:

Sequência .FASTA da proteína E6 do HPV18 e da 4giz utilizada como molde para a modelagem molecular

>E6 HPV18

MARFEDPTRRPYKLPDLCTELNTSLQDIEITCVYCKTVLELTEVFEFKDLFVVYRDSIPHAA
CHKCIDFYSRIRELRHYSDSVYGDLEKLTNTGLYNLLIRCLRCQKPLNPAEKLRLNEKRRF
HNIAGHYRGQCHSCCNRARQERLQRRRETQV*

>4GIZ

FQDPQERPRKLPQLCTELQTTIHDIILECVYCKQQLLRREVYDFARRDLCIVYRDGNPYAVCD
KCLKFYISKISEYRHYSYSLYGTTLQYQYNKPLSDLLIRCINCQKPLSPEEKQRHLDKKQRFHNI
RGRWTGRCMSCSRS-----*

APÊNDICE B

Linha de comando utilizada para a modelagem em *ab initio* da cauda N-terminal da proteína E6 HPV18 realizado pelo software Rosetta

```
./loopmodel.linuxccdrelease/in:file:s  
c://Users/Nathy/Desktop/estágio/modeller/HPV18.B99990020.pdb/loops:remodel'quick_ccd'/loops:ref  
ine'refine_kic'/loops:loop_file t000_.dat/loops:vall_file/loops:frag_sizes 9,3
```

APÊNDICE C:

Arquivo de configuração para a dinâmica molecular da proteína E6 do HPV18

```
#####
## JOB DESCRIPTION                                ##
#####
# Minimization and Equilibration of
# E718zinc in a Water Box
#####
## ADJUSTABLE PARAMETERS                        ##
#####
structure      E6HPV18_wb.psf
coordinates     E6HPV18_wb.pdb
set start_temperature 0
set temperature  310
set outputname   E618mdz_wb_eq
firsttimestep    0
#####
## SIMULATION PARAMETERS                        ##
#####
# Input
paraTypeCharmm      on
parameters          par_all27_prot_lipid_CYM.inp
temperature          $temperature

# Force-Field Parameters
exclude             scaled1-4
1-4scaling          1.0
cutoff              12.0
switching           on
switchdist          10.0
pairlistdist        13.5

# Integrator Parameters
timestep            2.0 ;# 2fs/step
rigidBonds          all ;# needed for 2fs steps
nonbondedFreq       1
```

```

fullElectFrequency 2
stepspcycle      10

# Constant Temperature Control
langevin         on ;# do langevin dynamics
langevinDamping  5 ;# damping coefficient (gamma) of 5/ps
langevinTemp      $temperature
langevinHydrogen  off ;# don't couple langevin bath to hydrogens

# Periodic Boundary Conditions
cellBasisVector1  52.6  0.  0.
cellBasisVector2  0.  50.2  0.
cellBasisVector3  0.  0  51.8
cellOrigin         0  0  0
margin            1.0

wrapAll           on

# PME (for full-system periodic electrostatics)
PME               yes
PMEGridSpacing    1.0

#manual grid definition
#PMEGridSizeX     54
#PMEGridSizeY     60
#PMEGridSizeZ     72

# Constant Pressure Control (variable volume)
useGroupPressure  yes ;# needed for rigidBonds
useFlexibleCell   no
useConstantArea   no
langevinPiston    on
langevinPistonTarget 1.01325 ;# in bar -> 1 atm
langevinPistonPeriod 100.0
langevinPistonDecay 50.0
langevinPistonTemp $temperature

```

```

# Output
outputName      $outputname
restartfreq      500    ;# 500steps = every 1ps
dcdfreq          250
xstFreq          250
outputEnergies   100
outputPressure   100
#####

## EXTRA PARAMETERS                                ##
#####
#####

## EXECUTION SCRIPT                                ##
#####

# Minimization
minimize          5000

#Heat
for { set TEMP $start_temperature } { $TEMP < $temperature } {
incr TEMP 6 } {
    reassignTemp $TEMP
    langevinTemp $TEMP
    run 500 }
run 5000000 ;# 10ns

```