

**FUNDAÇÃO OSWALDO CRUZ
CENTRO DE PESQUISAS AGGEU MAGALHÃES
MESTRADO EM BIOCÊNCIAS E BIOTECNOLOGIA EM SAÚDE**

MAÍRA DE ARRUDA LIMA

**MUTAÇÕES PONTUAIS NÃO SINÔNIMAS E SUA INFLUÊNCIA NA ESTRUTURA
TRIDIMENSIONAL DA PROTEÍNA L1 DO PAPILOMAVÍRUS HUMANO (HPV)**

**Recife
2017**

MAÍRA DE ARRUDA LIMA

MUTAÇÕES PONTUAIS NÃO SINÔNIMAS E SUA INFLUÊNCIA NA ESTRUTURA TRIDIMENSIONAL DA PROTEÍNA L1 DO PAPILOMAVÍRUS HUMANO (HPV)

Dissertação apresentada ao Curso de Mestrado em Biociências e Biotecnologia em Saúde do Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, como requisito para a obtenção do grau de Mestre em Ciências.

Orientadora:

Dra. Norma Lucena Cavalcanti Licinio da Silva

Recife

2017

Catálogo na fonte: Biblioteca do Centro de Pesquisas Aggeu Magalhães

L732m Lima, Maíra de Arruda.
Mutações pontuais não sinônimas e sua influência na estrutura tridimensional da proteína L1 do papilomavírus humano (HPV) / Maíra de Arruda Lima. - Recife: [s. n.], 2017.
51 p. : il., graf., tab.

Dissertação (Mestrado em Biociências e Biotecnologia em Saúde) - Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2017.

Orientadora: Norma Lucena Cavalcanti Licínio da Silva.

1. Papilomavírus Humano. 2. Proteína do capsídeo viral. 3. Homologia de Sequência. 4. Homologia Estrutural de Proteína. I. Silva, Norma Lucena Cavalcanti Licínio da. II. Título.

MAÍRA DE ARRUDA LIMA

MUTAÇÕES PONTUAIS NÃO SINÔNIMAS E SUA INFLUÊNCIA NA ESTRUTURA TRIDIMENSIONAL DA PROTEÍNA L1 DO PAPILOMAVÍRUS HUMANO (HPV)

Dissertação apresentada ao Curso de Mestrado em Biociências e Biotecnologia em saúde do Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, para a obtenção do grau de Mestre em Ciências.

Aprovada em 20/02/2017

Banca Examinadora

Dra. Norma Lucena Cavalcanti Licinio da Silva (Orientadora)
Centro de Pesquisas Aggeu Magalhães – FIOCRUZ/PE

Dr. Roberto Dias Lins Neto
Centro de Pesquisas Aggeu Magalhães – FIOCRUZ/PE

Dr. Albert Eduardo Silva Martins
Faculdades Integradas de Pernambuco – FACIPE/PE

LIMA, Maíra de Arruda. **Mutações pontuais não sinônimas e sua influência na estrutura tridimensional da proteína L1 do papilomavírus humano (HPV)**. 2016. Dissertação (Mestrado em Biociências e Biotecnologia em saúde) – Centro de Pesquisas Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2017.

RESUMO

O papilomavírus humano (HPV) é amplamente estudado devido a sua associação com um amplo espectro de manifestações clínicas incluindo câncer. Atualmente, mais de 200 tipos de HPV foram identificados e classificados em alto e baixo risco de acordo com seu potencial oncogênico. A tipagem viral é feita de acordo com a variabilidade genética observada na sequência nucleotídica do gene L1, além disso, a proteína codificada por este gene é responsável pela interação inicial entre o vírus e o hospedeiro. Portanto, substituições de aminoácidos podem alterar a infectividade e a antigenicidade viral ou afetar a resposta às vacinas atualmente comercializadas. Diante disso, o objetivo deste estudo é avaliar o possível impacto de mutações não sinônimas na interação vírus-hospedeiro através do estudo da estrutura tridimensional da proteína L1. Foram analisadas seqüências provenientes de amostras clínicas de mulheres HIV-HPV positivas atendidas em três hospitais públicos do Recife. As seqüências foram agrupadas por tipo viral e analisadas em relação à sua variabilidade genética através de alinhamentos múltiplos de seqüências. Em seguida, foram selecionadas seqüências com mutações não sinônimas para realização da modelagem comparativa de suas estruturas por meio dos programas: *SwissModel*, *Phyre2*, *ITasser*, *3DRefine*, *ModRefiner*, *Rampage* e *TM-Align*. Os resultados mostram que mutações não sinônimas são menos frequentes em HPVs de alto risco em comparação a baixo risco; algumas destas mutações definem a linhagem C de HPV58 na população estudada e sugerem que a coinfeção HIV-HPV seja fator de pressão seletiva para esta linhagem; além disso, as avaliações estruturais mostram maior divergência em relação à proteína selvagem das seqüências com a mutação G378D presente em 4 das 6 seqüências do gene L1 de HPV58 com mutações não sinônimas.

Palavras chaves: Papilomavírus Humano. Proteína do capsídeo viral. Homologia de Sequência. Homologia Estrutural de Proteína.

LIMA, Maíra de Arruda. **Non synonymous point mutations and its influence on the three-dimensional structure of human papillomavirus (HPV) L1 protein.** 2016. Dissertation (Master's in Biosciences and Biotechnology in Health) – Aggeu Magalhães Research Center, Oswaldo Cruz Foundation, Recife, 2017.

ABSTRACT

Human papillomavirus (HPV) are extensively studied due to its association with a wide spectrum of clinical manifestations including cancer. Currently, more than 200 HPV types were identified and classified into high and low risk according to their oncogenic potential. The viral typing is made according to the genetic variability observed on the nucleotide sequence of the L1 gene; furthermore, the protein encoded by this gene is responsible for the initial interaction between the virus and the host. So amino acid replacements may affect viral infectivity and antigenicity, or even the response to vaccines currently available. Thus, this study aims to evaluate the possible impact of non-synonymous mutations in virus-host interaction through the study of three-dimensional structure of the L1 protein. The analyzed sequences were obtained from clinical samples of HIV-HPV positive women treated at three public hospitals in Recife. The sequences were grouped by viral type and analyzed with regard to their genetic variability through multiple sequence alignment. Then sequences with non-synonymous mutations were chosen to the homology modeling of their structures using the softwares *SwissModel*, *Phyre2*, *ITasser*, *3DRefine*, *ModRefiner*, *Rampage* and *TM-Align*. The results shows that not synonymous mutations are less frequent in high-risk HPVs comparing to low risk types; some of these mutations defines the C lineage of HPV58 in the studied population and suggests the HIV-HPV coinfection to be a positive selective pressure factor for this lineage; furthermore, the structural alignment analysis showed a greater divergence from the wild-type protein of sequences with G378D mutation observed in 4 out of 6 HPV58 L1 gene sequences with non-synonymous mutations.

Keywords: Human Papillomavirus. Viral Capsid Proteins. Sequence Homology. Protein Structural Homology.

LISTA DE ILUSTRAÇÕES

| | |
|--|----|
| Figura 1 - Organização do genoma do papilomavírus humano tipo 16 em cinza e representação tridimensional do capsídeo viral no centro. | 11 |
| Figura 2 - Classificação molecular dos papilomavírus baseado na sequência da ORF L1..... | 13 |
| Quadro 1 - Números de acesso das sequências de cada tipo viral utilizadas como referência nos alinhamentos múltiplos realizados..... | 20 |
| Figura 3 - Sequências de aminoácidos construídas para a ORF completa da proteína L1 de HPV58 das variantes observadas. | 21 |
| Quadro 2 - Substituição de aminoácidos resultante de mutação não sinônima nos tipos virais de HPV de baixo risco..... | 27 |
| Quadro 3 - Substituição de aminoácidos resultante de mutação não sinônima nos tipos virais de HPV de alto risco..... | 28 |
| Quadro 4 - Substituição de aminoácidos resultante de mutações não sinônimas no tipo viral de alto risco HPV58..... | 29 |
| Figura 4 - Relações evolutivas entre as sequências parciais do gene L1 do tipo 58 de HPV e as sequências de referência das linhagens e sublinhagens descritas para este tipo viral..... | 30 |
| Quadro 5 - Sequências utilizadas na construção das estruturas tridimensionais e seus respectivos sítios com mutação..... | 31 |
| Quadro 6 - Estruturas utilizadas como molde por cada um dos programas usados na predição das estruturas tridimensionais da proteína L1 de HPV58..... | 32 |
| Figura 5 - Representação gráfica da estrutura tridimensional predita para a proteína L1 de HPV58..... | 40 |

LISTA DE TABELAS

| | |
|--|----|
| Tabela 1 - Distribuição da variabilidade genética das sequências analisadas de acordo com o grau de oncogenicidade..... | 25 |
| Tabela 2 - Distribuição da variabilidade genética das sequências analisadas de acordo com os tipos virais..... | 26 |
| Tabela 3 - Resultados obtidos para a predição das estruturas tridimensionais da proteína L1 de HPV58 de acordo com os programas utilizados e seus respectivos parâmetros de avaliação de qualidade..... | 33 |
| Tabela 4 - Resultado do refinamento, com o <i>software 3DRefine</i> , das estruturas tridimensionais preditas..... | 34 |
| Tabela 5 - Resultado do refinamento, com o <i>software ModRefiner</i> , das estruturas tridimensionais preditas..... | 35 |
| Tabela 6 - Resultados gerados pelo programa <i>Rampage</i> na construção do gráfico de <i>Ramachadran</i> para as estruturas tridimensionais preditas e refinadas para a proteína L1 de HPV58..... | 36 |
| Tabela 7 - Resultados gerados pelo programa <i>TM-Align</i> no alinhamento estrutural realizando entre as várias estruturas tridimensionais preditas para a proteína L1 de HPV58..... | 36 |
| Tabela 8 - Resultados obtidos na predição e no refinamento das estruturas tridimensionais da proteína L1 do HPV58, a partir de um único molde, pelos programas <i>ITasser</i> e <i>ModRefiner</i> | 38 |
| Tabela 9 - Resultados gerados pelo programa <i>Rampage</i> na construção do gráfico de <i>Ramachadran</i> para as estruturas tridimensionais da proteína L1 de HPV 58 preditas a partir de um único molde..... | 38 |
| Tabela 10 - Resultados gerados pelo programa <i>TM-Align</i> no alinhamento estrutural realizando entre as várias estruturas tridimensionais preditas para a proteína L1 de HPV58 a partir de um único molde..... | 39 |

SUMÁRIO

| | |
|---|-----------|
| 1 REFERENCIAL TEÓRICO..... | 10 |
| 1.1 O PAPILOMAVÍRUS HUMANO | 10 |
| 1.2 A SEQUÊNCIA DA ORF L1 E A CLASSIFICAÇÃO DO TIPO DE HPV | 12 |
| 1.3 PAPEL DA PROTEÍNA L1 NA PATOGÊNESE DO CÂNCER CERVICAL..... | 13 |
| 1.4 A PROTEÍNA L1 E AS VACINAS ANTI-HPV | 14 |
| 1.5 FERRAMENTAS DE BIOINFORMÁTICA NA PREDIÇÃO DE ESTRUTURAS PROTEICAS..... | 15 |
| 2 OBJETIVO | 17 |
| 3 METODOLOGIA | 18 |
| 3.1 ORIGEM DAS SEQUÊNCIAS | 18 |
| 3.2 DEFINIÇÃO DO TIPO VIRAL..... | 18 |
| 3.3 AVALIAÇÃO DE QUALIDADE DO SEQUENCIAMENTO | 19 |
| 3.4 VARIABILIDADE GENÉTICA DAS SEQUÊNCIAS | 19 |
| 3.5 ÁRVORE FILOGENÉTICA | 20 |
| 3.6 CONSTRUÇÃO DA SEQUÊNCIA COMPLETA DA ORF L1 PARA OS FRAGMENTOS SEQUENCIADOS | 20 |
| 3.7 MODELAGEM MOLECULAR COMPARATIVA..... | 21 |
| 3.7.1 Predição das estruturas tridimensionais | 21 |
| 3.7.2 Refinamento das estruturas | 22 |
| 3.7.3 Gráfico de Ramachandran..... | 22 |
| 3.7.4 Variações nas estruturas previstas | 23 |
| 4 CONSIDERAÇÕES ÉTICAS..... | 24 |
| 5 RESULTADOS..... | 25 |
| 5.1 GRAU DE ONCOGENICIDADE E VARIABILIDADE GENÉTICA | 25 |
| 5.2 TIPOS VIRAIS E VARIABILIDADE GENÉTICA | 25 |
| 5.3 CARACTERÍSTICAS BIOQUÍMICAS DOS AMINOÁCIDOS SUBSTITUÍDOS ... | 27 |
| 5.4 PADRÃO DE SUBSTITUIÇÃO DOS AMINOÁCIDOS RESULTANTE DE MUTAÇÃO NÃO SINÔNIMA..... | 29 |
| 5.5 ÁRVORE FILOGENÉTICA | 30 |
| 5.6 MODELAGEM MOLECULAR COMPARATIVA..... | 31 |

| | |
|--|-----------|
| 5.6.1 Modelagem comparativa utilizando os parâmetros padrão dos programas | 32 |
| | |
| 5.6.1.1 <i>Predição das estruturas</i> | 32 |
| 5.6.1.2 <i>Refinamento das estruturas</i> | 33 |
| 5.6.1.3 <i>Gráfico de Ramachandran</i> | 35 |
| 5.6.1.4 <i>Alinhamento estrutural</i> | 36 |
| 5.6.2 Modelagem comparativa utilizando estrutura molde definida manualmente | 37 |
| | |
| 5.6.2.1 <i>Predição e refinamento das estruturas</i> | 37 |
| 5.6.2.2 <i>Gráfico de Ramachandran</i> | 38 |
| 5.6.2.3 <i>Alinhamento estrutural</i> | 39 |
| 6 DISCUSSÃO | 41 |
| 6.1 VARIABILIDADE GENÉTICA E ÁRVORE FILOGENÉTICA | 41 |
| 6.2 ESTRUTURAS TRIDIMENSIONAIS | 43 |
| 7 CONCLUSÕES | 45 |
| REFERÊNCIAS | 46 |

1 REFERENCIAL TEÓRICO

1.1 O PAPILOMAVÍRUS HUMANO

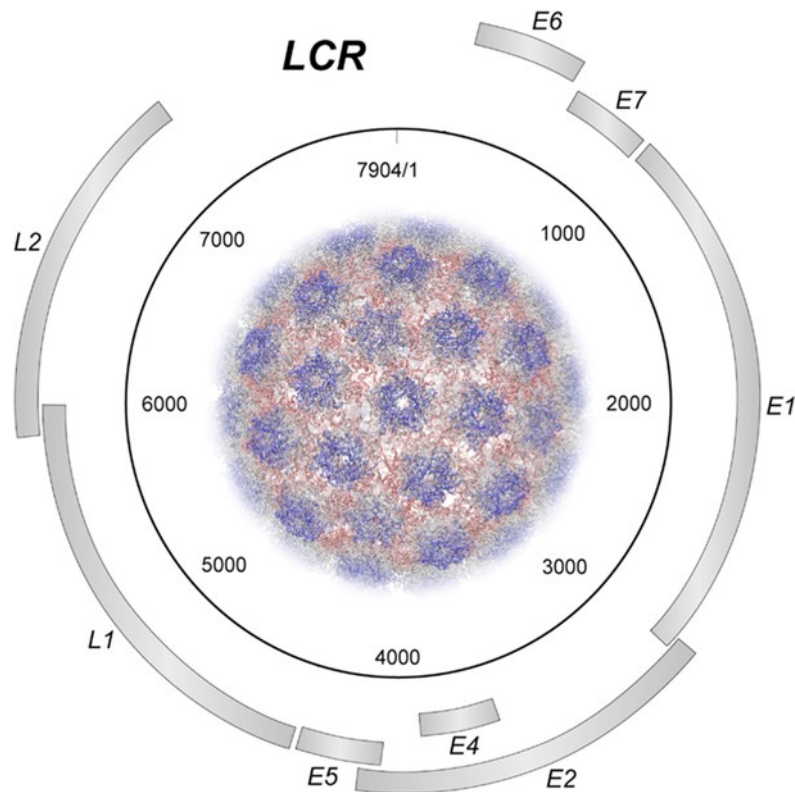
Os papilomavírus pertencem à família *Papillomaviridae* e classificam-se em cinco grandes gêneros (*alpha-*, *beta-*, *gamma-*, *mu-* e *nu-papillomavirus*) compostos por 49 espécies e mais de 200 tipos virais (BZHALAVA; EKLUND; DILLNER, 2015). Estes vírus infectam células epiteliais de vertebrados podendo causar neoplasias ou persistir de forma assintomática (BERNARD et al., 2010; BZHALAVA; EKLUND; DILLNER, 2015).

O papilomavírus humano (HPV) é reconhecido agente causal do câncer de colo uterino (ZUR HAUSEN, 2009) e está envolvido em outros tipos de câncer como os anogenitais e orais (MOODY; LAIMINS, 2010).

Seus tipos virais oncogênicos pertencem ao gênero *Alpha-papillomavirus* (DE VILLIERS et al., 2004) e apresentam características de tropismo tecidual e patogenicidade distintos. Ao infectar queratinócitos das camadas basais do epitélio escamoso estratificado do colo uterino, esses vírus são replicados e montados no núcleo destas células e tem sua expressão gênica orquestrada de forma intimamente relacionada aos diferentes estágios de diferenciação celular que ocorrem ao longo do tecido epitelial (DOORBAR, 2005; GRAHAM, 2012; HAUSEN, 1996; ZHENG; BAKER, 2006).

Os HPVs são vírus não envelopados, de DNA dupla hélice, cujo genoma contém oito quadros abertos de leitura ou ORFs, do inglês *Open Reading Frames*, classificados em precoce (E, do inglês *Early*) e tardio (L, do inglês *Late*) e a região longa de controle (LCR, do inglês *Long Control Region*), totalizando aproximadamente 8.000 pares de base (Figura 1) (BURD, 2003; FAVRE, 1975).

Figura 1 - Organização do genoma do papilomavírus humano tipo 16 em cinza e representação tridimensional do capsídeo viral no centro.



Fonte: Modificado de Kajitani et al. (2012)

A região LCR corresponde a 10% do genoma viral e nela localizam-se elementos de regulação gênica, como promotores e elementos reguladores de transcrição (HAUSEN, 1996). As ORFs precoces são seis (E6, E7, E1, E2, E4 e E5) e codificam proteínas reguladoras do ciclo viral sendo E6 e E7 suas principais oncoproteínas (SCHIFFMAN et al., 2007), por outro lado as ORFs tardias são duas (L1 e L2) e codificam as proteínas estruturais do capsídeo viral (ZHENG; BAKER, 2006).

As proteínas precoces são expressas em células nos estágios iniciais de diferenciação celular e atuam da seguinte forma: E1 está envolvida na replicação do genoma viral; E2 participa da regulação da expressão gênica, ciclo celular e apoptose da célula hospedeira; E4 interfere na formação da rede de queratina pelos queratinócitos; E5 está envolvida na modulação imunológica, e no controle de crescimento e diferenciação celular, E6 e E7 estimulam a proliferação celular, além de induzirem a imortalização e transformação celular, evento indispensável à malignização (DOORBAR, 2005; ZUR HAUSEN, 2002).

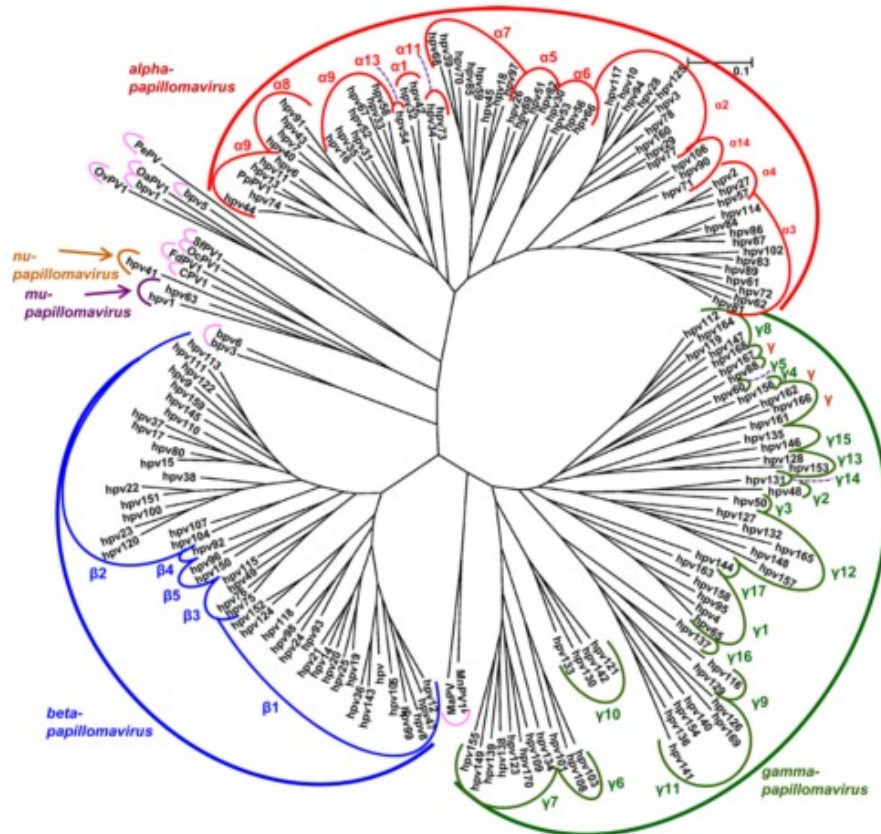
Em complemento, as proteínas tardias são expressas em células nos estágios finais de diferenciação quando ocorre a montagem e dispersão das partículas virais, dessa forma, L1 corresponde à proteína estrutural e L2 aos componentes estruturais adicionais necessários para montagem do capsídeo viral (DOORBAR, 2005; GRAHAM, 2012).

1.2 A SEQUÊNCIA DA ORF L1 E A CLASSIFICAÇÃO DO TIPO DE HPV

A classificação taxonômica dos HPVs, em vigor desde 1995 (HARARI; CHEN; BURK, 2014), é regulamentada pelo Comitê Internacional de Taxonomia de Vírus (ICTV) e se baseia na identidade entre as sequências nucleotídicas de L1 e sua posição topológica dentro da árvore filogenética estabelecida para a família *Papillomaviridae* (Figura 2) (CHEN; DE FREITAS; BURK, 2015). A ORF de L1 é uma região relativamente conservada dentro do genoma dos papilomavírus, mas possui diversidade suficiente para permitir a classificação e identificação dos diversos tipos virais (CHAN et al., 2012).

Para ser reconhecido como um novo tipo viral a sequência do isolado em questão deve diferir em mais de 10% da sequência do tipo viral mais próximo já conhecido. Por outro lado, diferenças de similaridade de 2 a 10% entre um novo isolado e todos os outros tipos de HPV conhecidos, definem subtipos, de menos de 2% de diferença, definem variantes (DE VILLIERS et al., 2004; KANESHIMA et al., 2005).

Figura 2 - Classificação molecular dos papilomavírus baseado na sequência da ORF L1



Fonte: De Villiers (2013)

Os diversos tipos virais são ainda classificados como tipos de alto e baixo risco de acordo com seu potencial oncogênico. Os tipos virais de alto risco estão envolvidos nos casos de câncer de colo uterino e em outros cânceres anogenitais e orais enquanto os tipos virais de baixo risco estão envolvidos em lesões benignas como verrugas e condilomas (DE VILLIERS et al., 2004; HAUSEN, 1996; MOODY; LAIMINS, 2010).

De forma geral os tipos virais de alto risco mais frequentes são HPV16,18, 31, 33, 45, 58, e de baixo risco os tipos HPV 6, 11 e 61, no entanto a distribuição dessa frequência varia de acordo com a área geográfica e comorbidades como, por exemplo, a coinfeção pelo HIV (MARTINS et al., 2014a, 2014b).

1.3 PAPEL DA PROTEÍNA L1 NA PATOGÊNESE DO CÂNCER CERVICAL

A interação inicial entre o vírus e a célula hospedeira é atribuída à ligação entre a proteína L1 do capsídeo viral e proteoglicanas de heparan sulfato

presentes na superfície celular (GIROGLOU et al., 2001; SCHILLER; DAY; KINES, 2010). Essa interação desencadeia alterações conformacionais no capsídeo viral permitindo sua ligação com proteínas de adesão da membrana celular, o que dispara o gatilho de sinalização celular para endocitose que resulta na internalização da partícula viral pela célula (RAFF et al., 2013).

Ao longo de sua história evolutiva, paralela à de seus hospedeiros, vários tipos de HPV desenvolveram diversos mecanismos para se evadir das defesas do sistema imune do hospedeiro e estabelecer uma infecção bem sucedida (GARNETT; DUERKSEN-HUGHES, 2006).

Já foi demonstrado, por exemplo, que variações específicas na região L1 do genoma resultando em substituição de aminoácidos, especialmente nas alças onde se localizam os epítomos imunodominantes, podem afetar de forma relevante a eficiência de infecção, antigenicidade e patogenicidade viral. As taxas em que essas substituições ocorrem parecem ser resultado da pressão seletiva exercida pela resposta imune do hospedeiro (FRATI et al., 2011; HILDESHEIM et al., 2007).

1.4 A PROTEÍNA L1 E AS VACINAS ANTI-HPV

A vacinação contra tipos virais de alto risco de HPV tem se mostrado uma opção possivelmente eficiente e logisticamente factível para a prevenção do câncer cervical (BOSCH et al., 2002). Atualmente existem três vacinas disponíveis no mercado: uma bivalente para os tipos HPV16 e 18 (Cervarix[®], GlaxoSmithKline, Middlesex, UK), uma quadrivalente (qHPV) para os tipos HPV16, 18, 6 e 11 (Gardasil[®], Merck, Rahway, NJ, USA) e uma nonavalente (Gardasil[®] 9, Merck, Rahway, NJ, USA) para os tipos HPV31, 33, 45, 52 e 58 além daqueles contemplados pela qHPV (JOURA et al., 2015; KIM; PARK; KO, 2014; PERES, 2011).

Essas vacinas se baseiam na capacidade da proteína L1 de se auto-montar em partículas parecidas com vírus (virus-like particles ou VLPs) (SCHILLER; DAY; KINES, 2010) que consistem somente no capsídeo viral montado sem nenhum material genético em seu interior. Esses VLPs são altamente imunogênicas (von BUBNOFF, 2012) atuando como antígenos para elicitar a resposta imune do organismo em resposta à vacinação.

Diante disso mutações em L1 e/ou em proteínas que interagem com L1 na montagem do capsídeo, podem afetar a resposta e a efetividade das vacinas que se baseiam nestas moléculas do capsídeo viral (BUCK; DAY; TRUS, 2013; CHEN et al., 2011).

1.5 FERRAMENTAS DE BIOINFORMÁTICA NA PREDIÇÃO DE ESTRUTURAS PROTEICAS

Apesar dos grandes avanços obtidos no campo de pesquisa de resolução experimental de estruturas proteicas, este processo ainda se mostra demorado e oneroso, além de não oferecer garantia de sucesso ao final (SCHWEDE et al., 2003). Por isso, observa-se uma grande lacuna entre a quantidade de informações geradas pelas ferramentas de sequenciamento de nova geração e a geração de informações sobre suas estruturas correspondentes.

Buscando preencher essa lacuna, diversas técnicas de modelagem comparativa vêm sendo aplicadas em diversos campos da ciência a fim de prover informações sobre as estruturas tridimensionais quando não há estruturas experimentais disponíveis (MARIANI et al., 2011).

Os métodos computacionais para predição de estruturas proteicas vêm sendo desenvolvidos e refinados ao longo dos últimos 30 anos (KELLEY et al., 2015) e atualmente, algoritmos totalmente automatizados permitem a aplicação da modelagem comparativa em estudos de escala proteômica (ZHANG, 2008).

Por gerar resultados acurados e rápidos a técnica de modelagem comparativa é a mais amplamente utilizada no campo da predição de estruturas proteicas (BHATTACHARYA; CHENG, 2013) e suscita novas perguntas de forma precisa e eficiente.

Nesse contexto, o HPV é um vírus amplamente disseminado ao redor do mundo, e por muitos de seus tipos possuírem grande potencial oncogênico se mostra como um importante alvo de estudo. Especificamente, o tipo viral HPV58 é de grande importância na região nordeste devido à sua maior frequência em relação à outras regiões do Brasil e seu já descrito envolvimento com lesões cervicais de alto grau e cancer cervical

A ORF L1 é adequada ao estudo de mutações genéticas intra e intertípicas importantes por se tratar de uma região com a taxa de variabilidade

suficientemente alta para esse tipo de observação e flanqueada por regiões altamente conservadas dentro do genomas dos papilomavírus, o que permitiu a construção do conjunto de primers MY09/11(MANOS, 1989). Dessa forma, o fragmento gerado pela reação de amplificação com estes primers permite a identificação do tipo viral de HPV envolvido na infecção cervical.

A partir da tipagem viral é possível associar dados de frequência, prevalência e infectividade à tipos virais específicos, bem como buscar características moleculares associadas aos sinais e sintomas da infecção cervical. Para isso a modelagem molecular apresenta-se como uma poderosa ferramenta de análise permitindo possíveis descobertas que auxiliem no desenvolvimento de vacinas mais eficazes na prevenção da infecção por esses vírus e em diagnósticos mais precisos das doenças decorrentes dessa infecção.

2 OBJETIVO

Identificar mutações não sinônimas em fragmentos de sequências nucleotídicas de diferentes tipos de HPV e avaliar sua influência nas estruturas tridimensionais da proteína L1 preditas por meio de modelagem molecular comparativa.

3 METODOLOGIA

3.1 ORIGEM DAS SEQUÊNCIAS

As 135 sequências utilizadas neste trabalho são provenientes de um estudo anteriormente realizado pelo grupo de pesquisa para avaliação de prevalência do Papilomavírus humano na cidade do Recife. Neste estudo, foram coletadas amostras biológicas de uma coorte de pacientes coinfectadas por HIV-HPV atendidas em três serviços de referência da cidade de Recife (Hospital Correia Picanço, Hospital das Clínicas e Centro Integrado de Saúde Amaury de Medeiros - CISAM) entre maio de 2007 e agosto de 2011. Este estudo foi aprovado pelo Comitê de Ética em Pesquisa do Centro de Saúde Amaury de Medeiros (CEP-CISAM#0011.0.250.000-05) e todas as mulheres convidadas a participar do estudo assinaram o termo de consentimento livre e esclarecido.

No âmbito deste projeto as amostras cervicais coletadas foram submetidas à extração de DNA genômico e diagnóstico molecular da infecção pelo HPV através de Reação em Cadeia da Polimerase. A reação foi realizada com o par de *primers* MY09 e MY11 (MANOS et al., 1989), que amplificam uma região de aproximadamente 450 pares de base da ORF L1 viral, e posterior sequenciamento deste fragmento para identificação do tipo de HPV envolvido na infecção cervical.

As sequências de nucleotídeos foram determinadas pelo equipamento *Genetic Analyzer ABI 3100 (Applied Biosystem)*, e depositadas no banco de dados do Núcleo de Plataformas Tecnológicas do Centro de Pesquisas Aggeu Magalhães (NPT - CPqAM) integrante da Rede Fiocruz de Plataformas Tecnológicas, de onde foram extraídas para realização das análises desta pesquisa.

Os detalhes sobre a coorte de mulheres coinfectadas por HIV-HPV em Pernambuco, incluindo dados de prevalência e distribuição dos tipos de HPV, já foram publicados (MARTINS et al., 2014a, 2014b).

3.2 DEFINIÇÃO DO TIPO VIRAL

Os cromatogramas resultantes do sequenciamento foram inicialmente visualizados através do programa MEGA5 (TAMURA et al., 2011) para avaliação da qualidade das sequências. As que apresentaram picos definidos e pouco ruído no

cromatograma foram submetidas a alinhamento múltiplo de sequências através da ferramenta BLAST (MADDEN, 2013) contra o banco de dados do *National Center for Biotechnology Information* (NCBI) a fim de identificar o tipo viral baseado em similaridade de sequências.

3.3 AVALIAÇÃO DE QUALIDADE DO SEQUENCIAMENTO

Após a separação das sequências de acordo com o seu tipo viral, foi realizada nova triagem utilizando o programa *Pregap4* através do algoritmo *Phred* (EWING; GREEN, 1998) que realiza a leitura dos cromatogramas gerado pelo sequenciador automático atribuindo uma base a cada pico de fluorescência com uma pontuação de qualidade correspondente; permaneceram no estudo, as sequências que obtiveram valores de *Phred* acima de 30, que significa a probabilidade de erro em 1 a cada 1000 bases atribuídas (99,9% de acurácia).

Em seguida, se fez a montagem de sequência consenso para cada amostra utilizando o programa *Gap4*. A sequência consenso corresponde à fusão das sequências geradas pelo sequenciamento *forward* e *reverse* de cada amostra, e suas bases nucleotídicas são as que apresentam maior valor de *Phred* para cada sítio. Os programas acima citados compõem o pacote *Staden* (STADEN, 1996).

3.4 VARIABILIDADE GENÉTICA DAS SEQUÊNCIAS

Sequências de nucleotídeos da ORF L1 completa de cada tipo viral foram obtidas no *GenBank* (BENSON et al., 2013), base de dados do NCBI, cujos números de acesso encontram-se descritos no Quadro 1. Cada sequência referência foi utilizada nos alinhamentos múltiplos entre os fragmentos sequenciados do mesmo tipo viral da sequência referência, para identificação de mutações sinônimas e não sinônimas presentes nos fragmentos amplificados a partir das amostras clínicas. O alinhamento foi realizado através do algoritmo *ClustalW* (THOMPSON; HIGGINS; GIBSON, 1994) implementado no programa MEGA5 (TAMURA et al., 2011).

Quadro 1 - Números de acesso das sequências de cada tipo viral utilizadas como referência nos alinhamentos múltiplos realizados.

| Números de acesso GenBank | | | |
|---------------------------|------------|-----------------|------------|
| HPV alto risco | | HPV baixo risco | |
| 16 | FJ006723.1 | 6 | AF092932.1 |
| 18 | GQ180791.1 | 11 | JQ773412.1 |
| 31 | HQ537668.1 | 54 | U37488.1 |
| 33 | HQ537694.1 | 61 | U31793.1 |
| 45 | KC470260.1 | 62 | AY395706.1 |
| 53 | EF546473.1 | 70 | U21941.1 |
| 56 | EF177180.1 | 71 | AY330623.1 |
| 58 | D90400.1 | 83 | AF151983.1 |
| 59 | X77858.1 | 84 | AF293960.1 |
| 66 | U31794.1 | | |
| 82 | AB027021.1 | | |
| 85 | AF131950.1 | | |

Fonte: A autora

3.5 ÁRVORE FILOGENÉTICA

A avaliação das relações evolutivas entre as sequências foi realizada mediante a construção de árvore filogenética, utilizando o programa MEGA6 (TAMURA et al., 2013), com base no algoritmo de Agrupamento de Vizinhos (*Neighbour-Joining*) e no modelo de evolução molecular de Kimura-2-Parâmetro, com teste de confiança de *bootstrap* com 1.000 réplicas (FELSENSTEIN, 1985; KIMURA, 1980; SAITOU; NEI, 1987) seguindo as mesmas condições de Gurgel et al. (2015) para fins de comparação.

3.6 CONSTRUÇÃO DA SEQUÊNCIA COMPLETA DA ORF L1 PARA OS FRAGMENTOS SEQUENCIADOS

Para a realização da etapa de modelagem comparativa, fazia-se necessário possuir a sequência completa da ORF L1, que tem em média 1.570 pares de base, e como foi citado anteriormente as sequências obtidas a partir das amostras clínicas correspondiam a fragmentos de apenas 450 pares de base.

Inicialmente todas as sequências de nucleotídeos foram convertidas em sequências de aminoácidos utilizando o programa MEGA5 (TAMURA et al.,

A análise consiste na utilização de ferramentas de alinhamento múltiplo de sequências proteicas, utilizando como referência estruturas biológicas resolvidas experimentalmente contidas no banco de dados *Protein Data Bank* (PDB); desta forma, a partir da sequência de aminoácidos fornecida são selecionadas estruturas que sejam adequadas a servir como molde para a predição computacional da proteína L1. Com base nas informações contidas no modelo experimental é predita a estrutura tridimensional da proteína de interesse, de forma comparativa e virtual.

3.7.2 Refinamento das estruturas

Uma vez que a proteína funcional será a termodinamicamente mais estável, o refinamento da estrutura tridimensional do modelo predito foi realizado para obtenção do modelo de menor energia possível (HABER; ANFINSEN, 1962).

Para isto, fornecendo o arquivo do modelo gerado na etapa anterior, foram utilizados os programas *3DRefine* (BHATTACHARYA; CHENG, 2013) e *ModRefiner* (XU; ZHANG, 2011), que virtualmente testam as diversas possibilidades de posicionamento dos átomos na estrutura e calculam a energia livre do sistema,

Nessa etapa, o programa *ModRefiner* gerou um modelo refinado para cada modelo bruto construído a partir de cada um dos três servidores utilizados, enquanto o *3DRefine* gerou cinco opções de modelos refinados para cada modelo bruto fornecido.

3.7.3 Gráfico de Ramachandran

O gráfico de Ramachandran representa todas as combinações possíveis dos ângulos diedros (*phi* e *psi*) de cada aminoácido em uma proteína; as posições dos ângulos na proteína são responsáveis por sua conformação tridimensional. Neste sentido, o gráfico permite avaliar a distribuição dos aminoácidos da estrutura predita e refinada nas etapas anteriores, observando se há posicionamento incompatível destes aminoácidos, que causariam choques estéricos entre eles, impossibilitando a estabilidade da estrutura proteica.

Para essa etapa foi utilizado o programa *Rampage* (LOVELL et al., 2003), que posiciona os aminoácidos no gráfico e fornece a porcentagem dos que se

encontram em regiões permitidas, favoráveis e não permitidas, a partir dos modelos refinados fornecidos pelo usuário.

De acordo com essas análises, os modelos escolhidos para serem submetidos à etapa seguinte foram os que apresentaram maior porcentagem de resíduos localizados em regiões permitidas.

3.7.4 Variações nas estruturas preditas

Com os modelos construídos, refinados e avaliados quanto a permissividade estérica, seguiu-se para a etapa de identificar e quantificar possíveis variações entre as estruturas tridimensionais. Para isto, cada uma das estruturas proteicas com mutação foi submetida ao alinhamento estrutural com a estrutura da proteína selvagem construída a partir da sequência de referência do mesmo tipo viral.

Essa análise foi realizada com o programa *TM-Align* (ZHANG; SKOLNICK, 2005), que sobrepõe as estruturas tridimensionais fornecidas e calcula a distância média entre seus átomos indicando similaridade ou diferença entre elas, por meio do parâmetro *TM-score* que é expresso como um número que varia entre 0 e 1 onde maiores valores indicam maior correspondência estrutural entre as estruturas, sendo acima de 0.5 considerados satisfatórios.

4 CONSIDERAÇÕES ÉTICAS

Este estudo utilizou dados secundários de banco de dados, sem acesso a nenhuma das informações pessoais das pacientes, não preenchendo, portanto, os requisitos para submissão em Comitê de Ética em Pesquisa.

5 RESULTADOS

5.1 GRAU DE ONCOGENICIDADE E VARIABILIDADE GENÉTICA

Foram analisadas 135 sequências da ORF L1 de HPV das quais 53 (39,3%) foram de tipos virais de baixo risco e 82 (60,7%) de tipos virais de alto risco.

No grupo dos tipos virais de baixo risco, 45 das 53 sequências nucleotídicas (84,9%) apresentaram sítios com mutações em relação às sequências gênicas de referência; nessas 45 sequências foram observados 93 diferentes sítios mutados, sendo 37 (39,8%) com substituição de seus aminoácidos.

No que diz respeito às sequências de HPV de alto risco, 54 das 82 sequências (65,9%) apresentaram mutação em relação as suas respectivas sequências de referência, nas quais foram identificados 65 sítios com variação nucleotídica, sendo apenas 18 (27,7%) sítios com mutação não sinônima (Tabela 1).

A menor variabilidade nucleotídica e menor proporção de mutações não sinônimas nas sequências de HPV de alto risco sugere maior pressão de seleção na ORF L1 do capsídeo de tipos oncogênicos de HPV.

Tabela 1 - Distribuição da variabilidade genética das sequências analisadas de acordo com o grau de oncogenicidade

| Grau de oncogenicidade | N(n) | % n | n(m) | %m | V(v) | % v |
|------------------------|--------|------|--------|------|--------|------|
| Baixo | 53(45) | 84,9 | 45(20) | 44,4 | 93(37) | 39,8 |
| Alto | 82(54) | 65,9 | 54(20) | 24,4 | 65(18) | 27,7 |

Fonte: A autora

Legenda: N – número total de sequências; n – número de sequências com mutação; m – número de sequências com mutação não sinônima; V – número total de sítios com mutação; v – número de sítios com mutação não sinônima.

5.2 TIPOS VIRAIS E VARIABILIDADE GENÉTICA

Dentre as sequências dos 9 tipos de HPV de baixo risco analisados o HPV6 (18 de 53 sequências) e o HPV11 (11 de 53 sequências) foram os mais frequentes, correspondendo a 54,8% do total de sequências do grupo. Em relação aos tipos de HPV de alto risco, foram analisadas sequências de 12 tipos distintos dos quais o HPV53 (23 de 82 sequências) e o HPV58 (21 de 82 sequências) foram os mais frequentes, representando 53,7% de todas as sequências do grupo.

No que diz respeito aos sítios com mutação nessas sequências, os tipos virais de baixo risco, HPV6 (27 de 93 sítios), HPV11 (25 de 93 sítios) e HPV54 (23 de 93 sítios) somaram 80,6% de toda a variação observada no grupo. Já para os tipos virais de alto risco, o HPV58 (12 de 65 sítios), o HPV18 (10 de 65 sítios) e o HPV31 (10 de 65 sítios) corresponderam a 49,3% do total de variação do grupo (Tabela 2).

Tabela 2 - Distribuição da variabilidade genética das sequências analisadas de acordo com os tipos virais

| Grau de oncogenicidade | Frequência | | | | | | |
|------------------------|------------|-------------|------------|-------------|-----------|-------------|------------|
| | Baixo | N(n) | % n | n(m) | %m | V(v) | % v |
| HPV 6 | | 18(13) | 72,2 | 13(8) | 61,5 | 27(10) | 37,0 |
| HPV 11 | | 11(9) | 81,8 | 9(4) | 44,4 | 25(12) | 48,0 |
| HPV 54 | | 1(1) | 100,0 | 1(1) | 100,0 | 23(8) | 34,8 |
| HPV 61 | | 4(4) | 100,0 | 4(4) | 100,0 | 6(6) | 100,0 |
| HPV 62 | | 3(3) | 100,0 | 3(0) | 0,0 | 1(0) | 0,0 |
| HPV 70 | | 3(3) | 100,0 | 3(0) | 0,0 | 1(0) | 0,0 |
| HPV 71 | | 2(2) | 100,0 | 2(0) | 0,0 | 2(0) | 0,0 |
| HPV 83 | | 6(6) | 100,0 | 6(0) | 0,0 | 6(0) | 0,0 |
| HPV 84 | | 5(4) | 80,0 | 4(3) | 75,0 | 2(1) | 50,0 |
| Alto | | N(n) | % n | n(m) | %m | V(v) | % v |
| HPV 16 | | 6(2) | 33,3 | 2(2) | 100,0 | 2(2) | 100,0 |
| HPV 18 | | 5(4) | 80,0 | 4(4) | 100,0 | 10(3) | 30,0 |
| HPV 31 | | 10(3) | 30,0 | 3(0) | 0,0 | 10(0) | 0,0 |
| HPV 33 | | 4(2) | 50,0 | 2(2) | 100,0 | 5(4) | 80,0 |
| HPV 45 | | 1(1) | 100,0 | 1(1) | 100,0 | 6(2) | 33,3 |
| HPV 53 | | 23(17) | 73,9 | 17(4) | 23,5 | 9(1) | 11,1 |
| HPV 56 | | 1(1) | 100,0 | 1(0) | 0,0 | 1(0) | 0,0 |
| HPV 58 | | 21(20) | 95,2 | 20(6) | 30,0 | 12(5) | 41,7 |
| HPV 59 | | 1(1) | 100,0 | 1(0) | 0,0 | 1(0) | 0,0 |
| HPV 66 | | 7(1) | 14,3 | 1(0) | 0,0 | 5(0) | 0,0 |
| HPV 82 | | 1(1) | 100,0 | 1(1) | 100,0 | 3(1) | 33,3 |
| HPV 85 | | 2(1) | 50,0 | 1(0) | 0,0 | 1(0) | 0,0 |

Fonte: A autora

Legenda: N – número total de sequências; n – número de sequências com mutação; m – número de sequências com mutação não sinônima; V – número total de sítios com mutação; v – número de sítios com mutação não sinônima

Dentre os tipos de baixo risco, chama atenção a única sequência de HPV54 com 23 sítios mutados, sendo 8 mutações não sinônimas; e o tipo HPV61 que apresentou mutações não sinônimas em todos os seus 6 sítios variantes nas 4 sequências analisadas, porém sem um padrão definido na troca desses aminoácidos.

Dentre os tipos de HPV de alto risco o HPV31 teve uma das maiores frequências de sítios com mutação (10 sítios em 3 sequências), mas nenhuma das mutações foi não sinônima; e o HPV53 apresentou 9 sítios com mutação em 17 sequências, sendo apenas 1 delas não sinônima. Por fim, 20 das 21 sequências analisadas do tipo viral de alto risco HPV58 apresentaram 12 sítios com mutação, das quais 5 foram não sinônimas e apresentaram nítido padrão de troca em seus aminoácidos.

5.3 CARACTERÍSTICAS BIOQUÍMICAS DOS AMINOÁCIDOS SUBSTITUÍDOS

A estrutura tridimensional e a função das proteínas estão relacionadas às características bioquímicas dos aminoácidos que as compõem, classificados de acordo com sua carga em neutros (carga nula), ácidos (negativa) ou básicos (positiva). Nesse estudo, observou-se que a substituição de aminoácidos neutros por ácidos ou básicos e vice-versa foi mais frequente do que a troca de um aminoácido ácido por básico e vice-versa.

Dentre as substituições ocorridas em tipos virais de HPV de baixo risco, 9 das 35 substituições resultaram em aminoácido com características diferentes do original, dentre elas 4 foram de neutro para ácido ou vice-versa, 3 de neutro para básico ou vice-versa, e 2 de básico para ácido (Quadro 2).

Quadro 2 - Substituição de aminoácidos resultante de mutação não sinônima nos tipos virais de HPV de baixo risco

| (continua) | | | |
|-------------|-------|----------------------------|----------------------------|
| Baixo Risco | | AA (x) | aa (x) |
| Tipo viral | Sítio | | |
| HPV 6 | 353 | Serina (N) | Cisteína (N) |
| | 359 | Metionina (N) | Valina (N) |
| | 384 | Alanina (N) | Ácido aspártico (A) |
| | 416 | Valina (N) | Glicina (N) |
| | 417 | Glutamina (N) | Prolina (N) |
| | 429 | Ácido glutâmico (A) | Ácido aspártico (A) |
| | 440 | Serina (N) | Arginina (B) |
| | 445 | Asparagina (N) | Isoleucina (N) |
| | 451 | Serina (N) | Treonina (N) |
| | 453 | Ácido glutâmico (A) | Ácido aspártico (A) |
| HPV 11 | 316 | Histidina (B) | Prolina (N) |
| | 317 | Asparagina (N) | Tirosina (N) |
| | 353 | Asparagina (N) | Treonina (N) |
| | 357 | Lisina (B) | Ácido glutâmico (A) |
| | 358 | Ácido glutâmico (A) | Glicina (N) |
| | 360 | Metionina (N) | Leucina (N) |
| | 361 | Arginina (B) | Prolina (N) |
| | 363 | Valina (N) | Alanina (N) |
| | 366 | Fenilalanina (N) | Tirosina (N) |
| | 372 | Fenilalanina (N) | Leucina (N) |
| | 379 | Leucina (N) | Fenilalanina (N) |
| | 434 | Glutamina (N) | Ácido glutâmico (A) |

Quadro 2 - Substituição de aminoácidos resultante de mutação não sinônima nos tipos virais de HPV de baixo risco

| (conclusão) | | | |
|-------------|-------|----------------------------|----------------------------|
| Baixo Risco | | AA (x) | aa (x) |
| Tipo viral | Sítio | | |
| HPV 54 | 344 | Treonina (N) | Serina (N) |
| | 347 | Serina (N) | Treonina (N) |
| | 351 | Serina (N) | Alanina (N) |
| | 352 | Ácido aspártico (A) | Asparagina (N) |
| | 354 | Arginina (B) | Lisina (B) |
| | 432 | Lisina (B) | Ácido Glutâmico (A) |
| | 438 | Fenilalanina (N) | Tirosina (N) |
| HPV 61 | 439 | Asparagina (N) | Alanina (N) |
| | 327 | Fenilalanina (N) | Tirosina (N) |
| | 362 | Arginina (B) | Lisina (B) |
| | 464 | Valina (N) | Fenilalanina (N) |
| | 465 | Serina (N) | Prolina (N) |
| HPV 84 | 466 | Fenilalanina (N) | Prolina (N) |
| | 441 | Ácido aspártico (A) | Ácido glutâmico (A) |

Fonte: A autora

Legenda: AA – aminoácido na sequência de referência; aa – aminoácido substituído na sequência com mutação; x – classificação do aminoácido quanto à sua carga.

Já entre os tipos de HPV de alto risco, 6 das 18 substituições resultaram em aminoácidos com características diferentes dos aminoácidos originais, sendo 1 troca de aminoácido básico para neutro, 1 de ácido para neutro e 3 de neutro para ácido (Quadro 3).

Mais uma vez, o tipo HPV58 se destacou, dentre os demais, por apresentar em 4 de suas 5 substituições, aminoácidos com carga diferente do aminoácido original.

Quadro 3 - Substituição de aminoácidos resultante de mutação não sinônima nos tipos virais de HPV de alto risco

| Alto Risco | | AA (x) | aa (x) |
|------------|-------|----------------------------|----------------------------|
| Tipo viral | Sítio | | |
| HPV 16 | 350 | Cisteína (N) | Triptofano (N) |
| | 379 | Treonina (N) | Alanina (N) |
| HPV 18 | 384 | Valina (N) | Isoleucina (N) |
| | 429 | Valina (N) | Alanina (N) |
| | 466 | Fenilalanina (N) | Leucina (N) |
| HPV 33 | 349 | Treonina (N) | Alanina (N) |
| | 370 | Ácido aspártico (A) | Ácido glutâmico (A) |
| | 390 | Isoleucina (N) | Valina (N) |
| | 443 | Tirosina (N) | Ácido aspártico (A) |
| HPV 45 | 383 | Asparagina (N) | Serina (N) |
| | 392 | Histidina (B) | Glutamina (N) |
| HPV 53 | 430 | Prolina (N) | Serina (N) |
| HPV 58 | 378 | Glicina (N) | Ácido aspártico (A) |
| | 383 | Ácido aspártico (A) | Asparagina (N) |
| | 412 | Isoleucina (N) | Valina (N) |
| | 420 | Ácido aspártico (A) | Asparagina (N) |
| | 422 | Asparagina (N) | Ácido aspártico (A) |
| HPV 82 | 350 | Prolina (N) | Glutamina (N) |

Fonte: A autora

Legenda: AA – aminoácido na sequência de referência; aa – aminoácido substituído na sequência com mutação; x – classificação do aminoácido quanto à sua carga.

Em resumo, a mudança das características bioquímicas nos aminoácidos substituídos por mutação não sinônima não foi frequente, tendo ocorrido em apenas 8 das 35 substituições (22,9%) dentre os tipos de HPV de baixo risco, e 6 das 18 substituições (33,3%) dentre os tipos de HPV de alto risco.

5.4 PADRÃO DE SUBSTITUIÇÃO DOS AMINOÁCIDOS RESULTANTE DE MUTAÇÃO NÃO SINÔNIMA

Dentre as 40 seqüências com mutações não sinônimas foi possível observar que, dentro de um mesmo tipo viral, diferentes seqüências apresentaram substituições de aminoácidos semelhantes. No entanto, apenas o HPV58 apresentou um padrão claro em relação ao aminoácido substituído e posição.

Nesse tipo viral, 6 das 21 seqüências apresentaram a mesma troca de aminoácidos em 5 sítios específicos e de forma espelhada (Quadro 4). Da mesma forma, as características bioquímicas também variaram de forma espelhada como mostrado no Quadro 3.

Quadro 4 - Substituição de aminoácidos resultante de mutações não sinônimas no tipo viral de alto risco HPV58

| Seqüências | Códon | | | | | | | | | | | |
|------------|------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| | 344 | 359 | 376 | 378 | 383 | 394 | 412 | 420 | 421 | 422 | 484 | 487 |
| | Aminoácido | | | | | | | | | | | |
| | G | V | K | G | D | E | I | D | S | N | L | F |
| PREV48 | * | * | | | | * | | | | | | |
| PREV75 | | * | | | | * | | | | | | |
| PREV93 | | * | | | | * | | | | | | |
| PREV264 | | | * | D | N | * | V | | | | | |
| PREV335 | | | | | | | | | | | | |
| PREV357 | * | * | | | | * | | | | | | |
| PREV359 | | | | | | * | | | * | | | |
| PREV397 | | * | | | | * | | | | | | |
| PREV419 | | * | | | | * | | | | | | |
| PREV436 | | * | | | | * | | | | | | |
| PREV439 | | | | | | * | | | | | | |
| PREV502 | | * | | | | * | | | | | | |
| PREV524 | | | | | | | V | N | * | D | * | |
| PREV557 | | | | | | * | | | | | | |
| PREV568 | | * | | | | * | | | | | | |
| PREV569 | | | * | D | N | * | V | N | * | D | * | * |
| PREV586 | | * | | | | * | | | | | | * |
| PREV603 | | | * | D | N | * | V | N | * | D | | |
| PREV616 | | | | | | * | V | | | | | |
| PREV633 | | * | | | | * | | | | | | |
| PREV637 | | | * | D | N | * | V | N | * | D | * | * |

Fonte: A autora

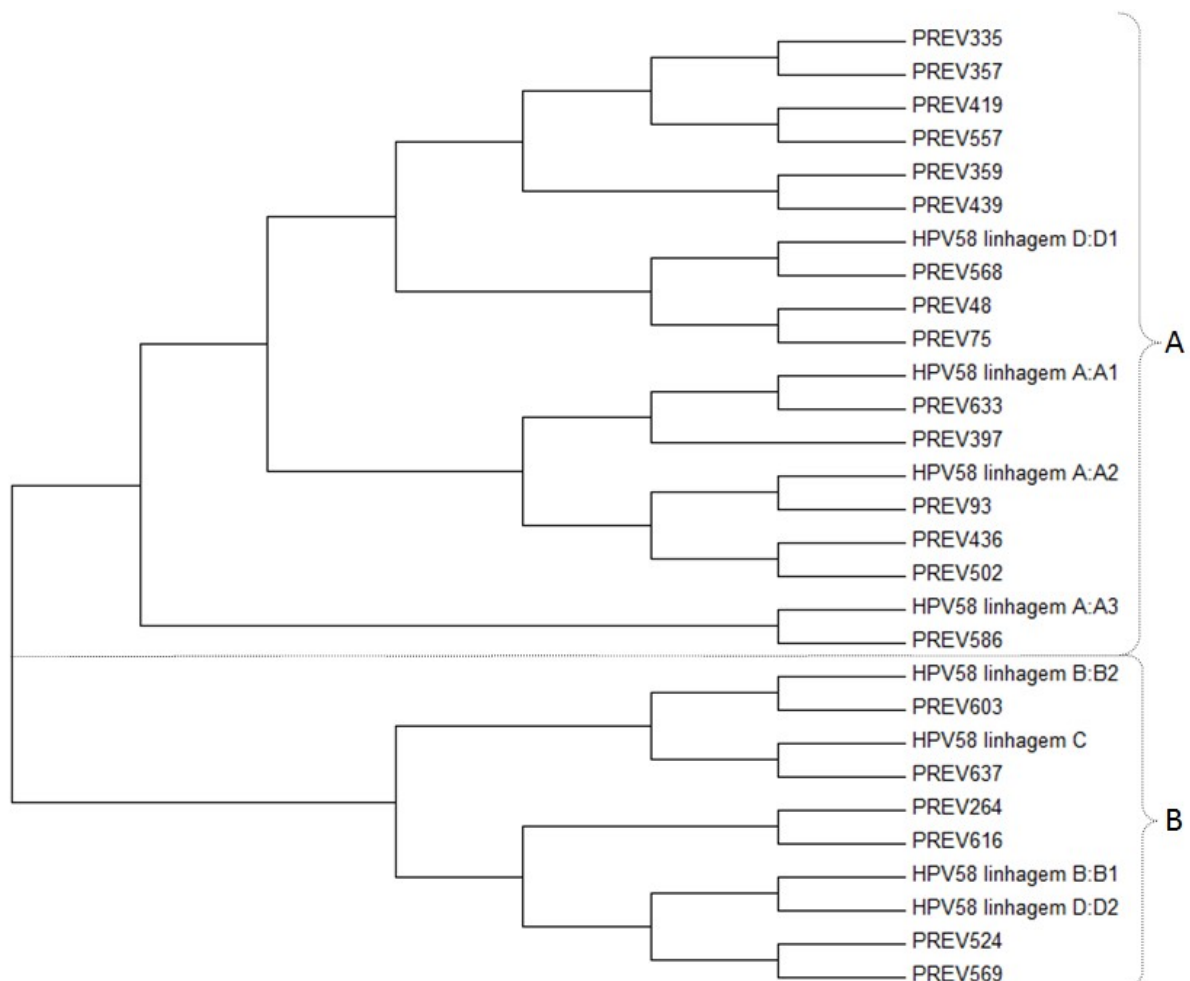
Legenda: * - mutações sinônimas

Diante da singularidade dos resultados observados no tipo HPV58 suas seqüências polipeptídicas variantes (PREV264, 524, 569 e 616) foram utilizadas para realização da modelagem comparativa de suas estruturas tridimensionais.

5.5 ÁRVORE FILOGENÉTICA

O cladograma construído com as 21 seqüências de fragmentos amplificados da ORF L1 das amostras positivas para HPV58 e as seqüências de referência da Linhagem A, sublinhagem A1 (D90400), A2 (HQ537752) e A3 (HQ537758); Linhagem B, sublinhagem B1 (HQ537762) e B2 (HQ537764), Linhagem C (HQ537774) e Linhagem D, sublinhagem D1 (HQ537768) e D2 (HQ537770) de HPV58 está mostrado na Figura 4.

Figura 4 - Relações evolutivas entre as seqüências parciais do gene L1 do tipo 58 de HPV e as seqüências de referência das linhagens e sublinhagens descritas para este tipo viral



A análise filogenética revelou dois grupos distintos, um grupo designado pela letra B, formado pelas sequências de fragmentos que apresentaram mutações não sinônimas juntamente com as sequências de referência da linhagem C e das sublinhagens B1, B2 e D2; e o outro grupo designado pela letra A com as demais sequências do tipo 58, incluindo as duas sequências que correspondem a lesões com diagnóstico de neoplasia cervical intraepitelial grau II (NIC2), agrupadas com as sublinhagens A1, A2, A3 e D1.

5.6 MODELAGEM MOLECULAR COMPARATIVA

Inicialmente, a modelagem molecular comparativa para a construção das estruturas das sequências com mutações não sinônimas e da sequência referência da ORF L1 de HPV58 foi realizada utilizando todos os parâmetros padrão dos programas.

No segundo momento, a estrutura experimental utilizada como molde foi a estrutura cristalográfica da cadeia A do pentâmero da proteína L1 de HPV16 (PDB Id:30AE), escolhida devido a sua maior resolução e indicada manualmente aos programas no momento de sua execução, todas as estruturas das sequências com mutações não sinônimas e da sequência referência foram refeitas nestas condições.

Desta forma, no total, foram construídas sete estruturas tridimensionais para a proteína L1 do tipo viral de alto risco HPV58 descritas no Quadro 5 abaixo.

Quadro 5 - Sequências utilizadas na construção das estruturas tridimensionais e seus respectivos sítios com mutação

| Sequências | Sítios com mutação não sinônima | | | | |
|------------|---------------------------------|-------|-------|-------|-------|
| | G378D | D383N | I412V | D420N | N422D |
| HPV58wt | | | | | |
| PREV616 | | | ■ | | |
| PREV264 | ■ | ■ | ■ | | |
| PREV524 | | | ■ | ■ | ■ |
| PREV569 | ■ | ■ | ■ | ■ | ■ |

Fonte: A autora

5.6.1 Modelagem comparativa utilizando os parâmetros padrão dos programas

5.6.1.1 Predição das estruturas

Cada um dos *softwares* utilizados na construção das estruturas tridimensionais da proteína L1 do HPV58 possuem parâmetros específicos para avaliar a qualidade dos modelos construídos e para selecionar as estruturas a serem utilizadas como molde.

Ao serem utilizados os parâmetros padrão de configuração destes programas, assume-se que a estrutura molde a ser selecionada será a que apresentar o melhor alinhamento com a sequência de aminoácidos fornecida pelo usuário de acordo com a homologia e a similaridade entre as sequências.

No entanto, apesar de se basearem no mesmo princípio de escolha do molde, o *SwissModel* utiliza apenas uma estrutura, a de melhor alinhamento como um todo, enquanto o *ITasser* e o *Phyre2* podem utilizar várias de forma simultânea, pois consideram regiões de maior similaridade ao invés da sequência como um todo.

As estruturas selecionadas como molde por cada um dos programas encontram-se listadas no Quadro 6.

Quadro 6 - Estruturas utilizadas como molde por cada um dos programas usados na predição das estruturas tridimensionais da proteína L1 de HPV58

| Programa | PDB id | Estrutura molde | Referência |
|-------------------|--------|-------------------------------|-----------------------------|
| <i>SwissModel</i> | 3j6r | L1 do HPV16 | CARDONE; MOYER; CHENG, 2014 |
| | 3iyj | L1 do BPV1 | WOLF et al., 2010 |
| <i>Itasser</i> | 3ofL | L1 do HPV18 ligado a heparina | DASGUPTA et al., 2011 |
| | 3oae | L1 do HPV16 ligado a heparina | |
| | 1dzL | L1 do HPV16 | CHEN et al., 2000 |
| <i>Phyre2</i> | 3j6r | L1 do HPV16 | CARDONE; MOYER; CHENG, 2014 |
| | 1dzL | L1 do HPV16 | CHEN et al., 2000 |
| | 2r5k | L1 do HPV11 | BISHOP et al., 2007 |

Fonte: A autora

Em relação à qualidade dos modelos construídos, cada programa possui um valor resultante de um cálculo matemático que permitem avaliar se as estruturas estão adequadas para serem utilizadas.

Para o *SwissModel* se chama Estimativa Global de Qualidade do Modelo (GMQE, do inglês *Global Model Quality Estimation*), expresso como um

número entre 0 e 1 em que, valores maiores que 0,5 indicam correta topologia do modelo gerado em relação ao seu molde. Para o *ITasser*, *C-score* se baseia na significância dos alinhamentos realizados e na convergência dos parâmetros utilizados ao longo das simulações de montagem da estrutura feitos pelo algoritmo de análise. Esse resultado é expresso como número entre -5 e 2, sendo a confiabilidade do modelo construído diretamente proporcional ao valor encontrado, neste estudo foi considerado como adequados valores maiores que 0. Para o *Phyre2*, a avaliação da qualidade dos modelos construídos é realizada através da porcentagem de confiabilidade dos resíduos modelados. Foram considerados satisfatórios os resíduos modelados com mais de 90% de confiabilidade.

Os valores obtidos para cada uma das estruturas encontram-se descritos na Tabela 3 mostrando que as estruturas construídas estavam aptas para a etapa seguinte de análises.

Tabela 3 - Resultados obtidos para a predição das estruturas tridimensionais da proteína L1 de HPV58 de acordo com os programas utilizados e seus respectivos parâmetros de avaliação de qualidade

| Sequência | SwissModel | ITasser | Phyre2 |
|-----------|------------|---------|--------|
| | GMQE | C-Score | R (%) |
| PREV616 | 0.78 | 0.34 | > 90 |
| PREV264 | 0.78 | 0.30 | > 90 |
| PREV524 | 0.78 | 0.37 | > 90 |
| PREV569 | 0.78 | 0.25 | > 90 |
| HPV58wt | 0.86 | 1.38 | > 90 |

Fonte: A autora

O fato dos três *softwares* utilizarem parâmetros e metodologias diferentes e apresentarem resultados convergentes para a escolha das estruturas a serem utilizadas como molde, com valores considerados satisfatórios para os parâmetros analisados, mostra consistência dos resultados.

5.6.1.2 Refinamento das estruturas

As cinco estruturas tridimensionais construídas na etapa anterior foram inicialmente submetidas aos programas *3DRefine*, que gerou cinco modelos refinados para cada modelo fornecido. O parâmetro *3DScore* expressa a energia da

estrutura refinada pelo programa *3DRefine*, sendo a estrutura com menor valor a adequada.

Diante disso, para as análises realizadas com o *3DRefine* foram selecionados um dos modelos refinados a partir da estrutura construída pelo programa *Phyre2* para as sequências PREV616, PREV264 e HPV58wt; enquanto para as sequências PREV524 e PREV569 o modelo refinado selecionado foi um dos elaborados a partir da estrutura construída pelo programa *ITasser* como descrito na Tabela 4.

Tabela 4 - Resultado do refinamento, com o software *3DRefine*, das estruturas tridimensionais preditas

| 3DRefine | | | |
|--------------------------|----------------|----------------|-------------------|
| Modelos refinados | 3DScore | | |
| PREV616 | Itasser | Phyre2 | SwissModel |
| 1 | 38379.0 | 35997.2 | 46526.4 |
| 2 | 32424.4 | 30879.0 | 36181.2 |
| 3 | 31348.2 | 30078.9 | 32733.9 |
| 4 | 30817.3 | 29580.2 | 31450.6 |
| 5 | 30442.8 | 29283.8 | 30589.7 |
| PREV264 | Itasser | Phyre2 | SwissModel |
| 1 | 38328.9 | 35324.4 | 46338.8 |
| 2 | 32720.9 | 30391.8 | 36021.6 |
| 3 | 31653.6 | 29534.5 | 32699.6 |
| 4 | 31014.9 | 29041.8 | 31419.7 |
| 5 | 30612.9 | 28705.0 | 30601.6 |
| PREV524 | Itasser | Phyre2 | SwissModel |
| 1 | 37159.8 | 36707.1 | 46570.0 |
| 2 | 31683.3 | 31608.5 | 36101.4 |
| 3 | 30588.8 | 30731.8 | 32719.2 |
| 4 | 29765.1 | 30253.4 | 31389.5 |
| 5 | 29385.3 | 29907.1 | 30514.2 |
| PREV569 | Itasser | Phyre2 | SwissModel |
| 1 | 37612.7 | 36647.4 | 46515.5 |
| 2 | 31775.2 | 31625.2 | 36112.1 |
| 3 | 30619.0 | 30711.5 | 32657.4 |
| 4 | 29773.4 | 30185.7 | 31328.7 |
| 5 | 29288.4 | 29845.2 | 30781.8 |
| HPV58wt | Itasser | Phyre2 | SwissModel |
| 1 | 36520.3 | 30998.9 | 43927.2 |
| 2 | 31803.0 | 26485.3 | 34132.8 |
| 3 | 30799.2 | 25775.4 | 30991.5 |
| 4 | 30201.3 | 25275.4 | 29831.6 |
| 5 | 29821.4 | 24938.9 | 29071.4 |

Fonte: A autora

As estruturas construídas também foram submetidas ao programa *ModRefiner*, que gerou um modelo refinado para cada modelo fornecido. O

parâmetro *TM-score* mede de similaridade estrutural entre o modelo refinado pelo *ModRefiner* e a estrutura que foi utilizada como molde. É expresso por um número entre 0 e 1 onde maiores valores indicam melhor correspondência entre as estruturas, sendo acima de 0,5 considerados satisfatórios.

De forma semelhante, para as análises realizadas com o programa *ModRefiner* foram escolhidos, para a sequência PREV264 o modelo refinado a partir da estrutura construída pelo programa *ITasser* e para as demais sequências, os modelos refinados a partir das estruturas construídas pelo programa *Phyre2* como mostra a Tabela 5. As estruturas selecionadas foram então submetidas à etapa seguinte.

Tabela 5 - Resultado do refinamento, com o software *ModRefiner*, das estruturas tridimensionais preditas

| Sequências | ModRefiner | | |
|------------|---------------|---------------|------------|
| | TM-score | | |
| | Itasser | Phyre2 | SwissModel |
| PREV616 | 0.9865 | 0.9914 | 0.9807 |
| PREV264 | 0.9890 | 0.9888 | 0.9817 |
| PREV524 | 0.9807 | 0.9923 | 0.9811 |
| PREV569 | 0.9912 | 0.9930 | 0.9751 |
| HPV58wt | 0.9903 | 0.9936 | 0.9833 |

Fonte: A autora

5.6.1.3 Gráfico de Ramachandran

Nos resultados dos gráficos de Ramachandran é ideal que a porcentagem de resíduos localizados na região *outlier* seja igual a zero, diante disso, os modelos refinados pelo programa *ModRefiner* obtiveram os melhores resultados pois apresentaram menores porcentagens de resíduos na região não permitida para todos os modelos construídos e refinados (Tabela 6).

Tabela 6 - Resultados gerados pelo programa *Rampage* na construção do gráfico de *Ramachadran* para as estruturas tridimensionais preditas e refinadas para a proteína L1 de HPV58

| Rampage | | | |
|-------------------|----------------------------------|------------------|----------------|
| ModRefiner | Localização espacial N(n) | | |
| | Favorável | Permitido | Outlier |
| PREV616 | 492 (94.3%) | 25 (4.8%) | 5 (1.0%) |
| PREV264 | 493 (94.4%) | 24 (4.6%) | 5 (1.0%) |
| PREV524 | 495 (94.8%) | 24 (4.6%) | 3 (0.6%) |
| PREV569 | 493 (94.4%) | 23 (4.4%) | 6 (1.1%) |
| HPV58wt | 438 (96.9%) | 10 (2.2%) | 4 (0.9%) |
| 3DRefine | | | |
| 3DRefine | Localização espacial N(n) | | |
| | Favorável | Permitido | Outlier |
| PREV616 | 483 (92.5%) | 29 (5.6%) | 10 (1.9%) |
| PREV264 | 481 (92.1%) | 30 (5.7%) | 11 (2.1%) |
| PREV524 | 467 (89.5%) | 32 (6.1%) | 23 (4.4%) |
| PREV569 | 464 (88.9%) | 42 (8.0%) | 16 (3.1%) |
| HPV58wt | 427 (94.5%) | 20 (4.4%) | 5 (1.1%) |

Fonte: A autora

5.6.1.4 Alinhamento estrutural

Por fim, o resultado do alinhamento estrutural entre os modelos foi avaliado através do parâmetro *TM-Score*, considerando que, quanto maiores fossem os valores maior a semelhança entre as estruturas alinhadas, e vice-versa (Tabela 7).

Tabela 7 - Resultados gerados pelo programa *TM-Align* no alinhamento estrutural realizando entre as várias estruturas tridimensionais preditas para a proteína L1 de HPV58

| TM-Align | | | |
|-----------------|----------------|-------------------|----------|
| E (L) | e (L) | Parâmetros | |
| | | TM-score | S |
| HPV58wt (454) | HPV58wt (454) | 1.0 | 454 |
| HPV58wt (454) | HPV11PDB (419) | 0.84763 | 392 |
| HPV58wt (454) | HPV16PDB (455) | 0.91924 | 428 |
| HPV58wt (454) | PREV616 (524) | 0.94074 | 439 |
| HPV58wt (454) | PREV264 (524) | 0.93530 | 439 |
| HPV58wt (454) | PREV524 (524) | 0.93375 | 439 |
| HPV58wt (454) | PREV569 (524) | 0.94334 | 439 |

Fonte: A autora

Legenda: E – estrutura de referência; e – estrutura com mutação;
L – tamanho da sequência; S – número de resíduos alinhados.

A comparação entre a estrutura da proteína selvagem com ela mesma serviu para testar a fidelidade do método pela completa sobreposição estrutural. A comparação desta mesma estrutura com as estruturas do HPV11 e do HPV16

mostrou maior semelhança entre as estruturas da proteína L1 do HPV16 e do HPV58, ambos os tipos virais de alto risco, como era esperado.

Na comparação entre a estrutura da proteína selvagem e as estruturas construídas neste estudo, os valores de *TM-Score* mostram maior semelhança da proteína selvagem (HPV58wt) com a estrutura contendo as cinco mutações (PREV569; *TM-score*=0,943) e com a estrutura contendo apenas uma mutação (PREV616; *TM-score*=0,941).

Por outro lado, o alinhamento estrutural da proteína selvagem (HPV58wt) com a estrutura contendo as três primeiras (PREV264; *TM-score* = 0,935) ou com as três últimas mutações (PREV524; *TM-score* = 0,934) foram muito semelhantes entre si e as mais divergentes da proteína de referência.

5.6.2 Modelagem comparativa utilizando estrutura molde definida manualmente

5.6.2.1 Predição e refinamento das estruturas

A resolução da estrutura tridimensional diz respeito à precisão do posicionamento de seus átomos e, portanto da acurácia das estruturas modeladas a partir dela. Desta forma, foi selecionada, para ser utilizada como molde, a estrutura cristalográfica da cadeia A do pentâmero da proteína L1 de HPV16 (PDB id: 3oae) (DASGUPTA et al., 2011) com resolução de 2.8 Angstroms; a melhor dentre todas as estruturas resolvidas experimentalmente, para a proteína L1, contidas no PDB.

Com base na etapa anterior foram utilizados os programas que apresentaram os melhores resultados em cada uma das etapas: *ITasser*, *ModRefiner*, *Rampage* e *TM-Align*, com os mesmos parâmetros de avaliação.

A predição da estrutura foi feita com o programa *ITasser* e sua qualidade avaliada através do parâmetro *C-score* e em seguida as estruturas construídas foram refinadas no programa *ModRefiner* e seu resultado avaliado a partir do valor de *TM-score* (Tabela 8).

Tabela 8 - Resultados obtidos na predição e no refinamento das estruturas tridimensionais da proteína L1 do HPV58, a partir de um único molde, pelos programas *ITasser* e *ModRefiner*.

| Sequência | ITasser | ModRefiner |
|-----------|----------------|-------------------|
| | C-Score | TM-Score |
| PREV616 | -0.83 | 0.9659 |
| PREV264 | -1.07 | 0.9749 |
| PREV524 | -0.73 | 0.9806 |
| PREV569 | -0.86 | 0.9702 |
| HPV58wt | -0.71 | 0.9771 |

Fonte: A autora

Observou-se que os valores de *C-Score* foram menores em relação à análise com múltiplos moldes o que pode ser explicado pelo fato de que múltiplos moldes proporcionam mais chances de que toda a sequência a ser modelada tenha contrapartida nas estruturas experimentais. Já em relação aos valores de *TM-Score*, estes se mostraram maiores que os seus correspondentes na análise com múltiplos moldes, o que indica que foi necessária uma quantidade menor de modificações no modelo gerado para que ele alcançasse seu estado de mínima energia.

5.6.2.2 Gráfico de Ramachandran

As estruturas refinadas tiveram seus gráficos de Ramachandran construídos para avaliação do posicionamento de seus aminoácidos (Tabela 9).

Tabela 9 - Resultados gerados pelo programa *Rampage* na construção do gráfico de *Ramachandran* para as estruturas tridimensionais da proteína L1 de HPV 58 preditas a partir de um único molde

| Sequências | Rampage | | |
|------------|-----------------------------|------------------|----------------|
| | Localização espacial | | |
| | Favorável | Permitido | Outlier |
| PREV616 | 473 (90.6%) | 30 (5.7%) | 19 (3.6%) |
| PREV264 | 467 (89.5%) | 37 (7.1%) | 18 (3.4%) |
| PREV524 | 475 (91.0%) | 35 (6.7%) | 12 (2.3%) |
| PREV569 | 471 (90.2%) | 35 (6.7%) | 16 (3.1%) |
| HPV58wt | 464 (88.9%) | 40 (7.7%) | 18 (3.4%) |

Fonte: A autora

As estruturas preditas a partir de um só molde apresentaram maior número de aminoácidos em regiões não permitidas do gráfico em relação as construídas com múltiplos moldes.

5.6.2.3 Alinhamento estrutural

Por fim, cada uma destas estruturas refinadas foi estruturalmente alinhada com a estrutura da proteína selvagem para avaliação das possíveis variações (Tabela 10).

Tabela 10 - Resultados gerados pelo programa TM-Align no alinhamento estrutural realizando entre as várias estruturas tridimensionais preditas para a proteína L1 de HPV58 a partir de um único molde

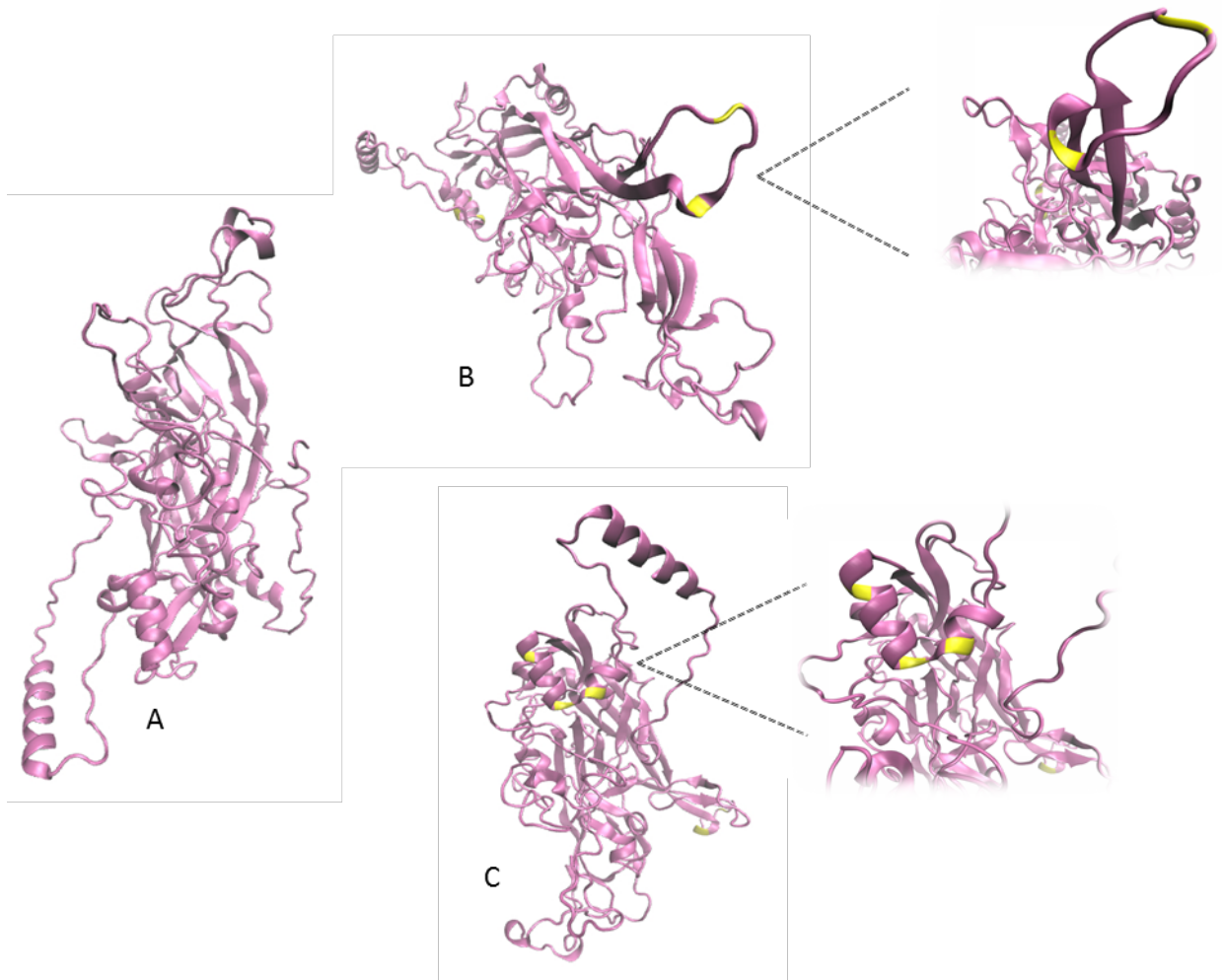
| TM-Align | | Parâmetros | |
|-----------------|----------------|-------------------|----------|
| E (L) | e (L) | TM-score | S |
| HPV58wt (524) | HPV58wt (524) | 1.0 | 524 |
| HPV58wt (524) | HPV11PDB (419) | 0.78883 | 418 |
| HPV58wt (524) | HPV16PDB (455) | 0.85760 | 453 |
| HPV58wt (524) | PREV616 (524) | 0.93650 | 515 |
| HPV58wt (524) | PREV264 (524) | 0.90183 | 493 |
| HPV58wt (524) | PREV524 (524) | 0.95048 | 518 |
| HPV58wt (524) | PREV569 (524) | 0.91091 | 496 |

Fonte: A autora

Legenda: E – estrutura de referência; e – estrutura com mutação;
L – tamanho da sequência; S – número de resíduos alinhados.

Os valores de *TM-Score* mostram maior semelhança entre as estruturas do modelo selvagem (HPV58wt) com o modelo com as três últimas mutações (PREV524; *TM-score* = 0,950) e com o modelo com apenas a mutação central (PREV616; *TM-score* = 0,936). Por outro lado, o alinhamento estrutural do modelo selvagem (HPV58wt) com o modelo com as cinco mutações (PREV569; *TM-score* = 0,910) e com o modelo com as três primeiras mutações (PREV264; *TM-score* = 0,901) mostram maior divergência entre eles. Esse resultado sugere uma maior variação estrutural resultante da primeira mutação, uma vez que os modelos mais divergentes a possuem e os mais semelhantes ao modelo selvagem não; e isto pode estar relacionado às características do aminoácido substituído bem como à localização do sítio de mutação (Figura 5).

Figura 5 - Representação gráfica da estrutura tridimensional predita para a proteína L1 de HPV58



Legenda: (A) estrutura da proteína selvagem; (B) localização dos sítios de mutação G378D e D383N; (C) localização dos sítios de mutação I412V, D420N e N422D.

6 DISCUSSÃO

6.1 VARIABILIDADE GENÉTICA E ÁRVORE FILOGENÉTICA

Neste estudo, foram analisadas 135 sequências de fragmentos da ORF L1 de diversos tipos de HPV, sendo 82 destas sequências (60,7%) pertencentes a tipos virais de alto risco. Dentre eles, destacaram-se, o tipo HPV53 com 23 sequências (28%) e o HPV58 com 21 sequências (25,6%). Estes dois tipos virais foram, portanto os mais frequentes dentre todos os tipos observados, tanto de alto quanto de baixo risco.

De forma geral, segundo dados da literatura, os tipos virais de alto risco mais frequentes são HPV16 e HPV18, e a distribuição de frequência dos tipos não-16/18 varia de forma consistente ao redor do globo (CALLEJA-MACIAS et al., 2005; CANCHE et al., 2010). A população da qual estas sequências foram provenientes foi de mulheres coinfectadas HIV-HPV, e sabe-se que mulheres HIV-positivas são mais susceptíveis à infecção por tipos virais de alto risco menos frequentes, no entanto à medida que os estudos avançam neste sentido, observa-se que a maior frequência de tipos virais como HPV53 e HPV58 parece estar associada também a fatores geográficos e de origem étnica das populações (CAMARA et al., 2003; MARTINS et al., 2014b). As mais altas prevalências do tipo HPV58, por exemplo, vêm sendo observadas em países da África, Américas Central e do Sul e Ásia (DE SANJOSÉ et al., 2007; LIU et al., 2012), de forma similar à observada neste estudo.

Além disso, estudos têm mostrado grande frequência de DNA viral de HPV58 detectado em amostras de pacientes com citologia normal (SANJOSÉ et al., 2007), bem como em lesões precursoras (CINs) de câncer cervical (CAMARA et al., 2003; FERNANDES et al., 2013) como foi observado neste estudo, em duas amostras com diagnóstico de CIN2 indicando que a persistência da infecção por este tipo viral deve ser acompanhada de perto frente à sua importância no possível desenvolvimento do câncer cervical.

Neste sentido, um ponto importante a se considerar é a presença de variantes e linhagens dentro de um mesmo tipo viral. O tipo HPV58, especificamente, se classifica em quatro linhagens (A,B,C e D) cujo padrão de distribuição sugere a existência de preferências etnogeográficas por estas variantes (CHAN et al., 2013). A saber, há uma maior frequência mundial da linhagem A, especificamente de sua

sublinhagem A2, seguida de A3 em todos os continentes, exceto na África onde a segunda linhagem mais frequente é a C (CHAN et al., 2011).

Corroborando estes dados, observou-se neste estudo que as 6 sequências do tipo HPV58 que apresentaram um padrão diferenciado de mutações não sinônimas correspondiam a diferentes linhagens virais. Destas, 2 não puderam ser classificadas com exatidão pois, as mutações presentes no fragmento analisado estão presentes tanto na linhagem B quanto na linhagem D, as outras 4 no entanto, foram classificadas pertencentes à linhagem C por possuírem mutações características desta linhagem (G378D e D383N). A maior frequência desta última linhagem dentre as sequências do tipo HPV58 analisadas (35%) pode ser associada à grande miscigenação afro-brasileira da região e do país como um todo.

Da mesma forma, a árvore filogenética construída com as sequências de referência de cada uma das linhagens e sublinhagens de HPV58 juntamente com as sequências provenientes das amostras corroboraram os dados descritos anteriormente, nos quais as sequências que apresentaram as mutações não sinônimas agruparam-se com a sequência de referência da linhagem C.

Por outro lado, as sequências que corresponderam à lesões com diagnóstico de NIC2, agruparam-se com as sequências de referência da linhagem A e suas sublinhagens, o que também é compatível com os primeiros resultados dos estudos que buscam associar as variantes ao risco de desenvolvimento de câncer (BURK; HARARI; CHEN, 2013).

Interessantemente, outro estudo realizado na mesma região em condições semelhantes, porém com sequências provenientes de amostras de mulheres apenas HPV-positivas (GURGEL et al., 2015) não observaram os dois sítios de mutação específicos da linhagem C (G378D e D383N) e estes se localizam em regiões de epítomos de células T, afetadas diretamente pela infecção por HIV o que poderia sugerir que a coinfeção com HIV atue como um fator de pressão positiva seletiva para esta linhagem bem como reforça a origem étnica da mesma.

Diante deste cenário, fica ressaltada a importância de estudos como este para elucidar questões etnogeográficas específicas a fim de permitir que as futuras gerações de vacinas profiláticas para HPV incluam outros tipos virais, bem como variantes específicas a fim de garantir maior eficácia de proteção.

A vacina profilática para HPV atualmente disponível é a tetravalente que confere proteção contra os tipos HPV16, 18, 6 e 11, no entanto, observa-se que

características etnogeográficas devem ser levadas em consideração a fim de garantir melhor cobertura às diversas populações (FERNANDES et al., 2013; MEJÍA et al., 2016; WU et al., 2009).

6.2 ESTRUTURAS TRIDIMENSIONAIS

A escolha da estrutura utilizada como molde com maior resolução do que as estruturas escolhidas pelos programas computacionais, nas condições padrão, modifica os resultados. Se, por um lado, fazer as análises levando em conta apenas a homologia e a similaridade das sequências, como é o padrão dos programas, permite que um número maior de estruturas sejam utilizadas como molde, aumentando a chance de que toda a sequência fornecida tenha contrapartida no modelo experimental; por outro, deve-se questionar a acurácia da informação obtida através de modelos com resoluções muito baixas, que nesse caso indicam menor capacidade de visualização das posições dos átomos e, portanto, menor acurácia no posicionamento dos átomos da estrutura construída a partir desta (OLIVELLA et al., 2013).

As estruturas menos semelhantes a selvagem apresentam a mutação G378D, isso se deve em parte às diferenças bioquímicas decorrentes da troca de aminoácidos. A glicina é um aminoácido neutro, pequeno, que possui apenas um hidrogênio em sua cadeia lateral, enquanto o ácido aspártico, que o substitui nessa mutação, possui carga negativa e um grupo carboxílico em sua cadeia lateral.

A localização dos sítios G378D e D383N, numa região de alça da proteína sugere maior disponibilidade para interagir com outras moléculas. Os sítios I412V, D420N e N422D, por sua vez, encontram-se localizados dentro de uma estrutura de α -hélice e, portanto, estão mais envolvidos na manutenção da estrutura de forma estável do que na interação desta proteína com outras moléculas (SADOWSKI; TAYLOR, 2010).

A proteína L1 de HPV interage com dois grupos de moléculas: os heparan sulfato da superfície celular dos queratinócitos do hospedeiro, o que permite a internalização da partícula viral pela célula; e com linfócitos B e T do sistema imune, através de epítomos virais, que são reconhecidas por estas células elicitando a resposta imunológica do organismo (CHAN et al., 2010; GIROGLOU et al., 2001; GURGEL et al., 2015).

Os sítios mutados G378D e D383N característicos da linhagem C de HPV58, não coincidem com os sítios de ligação descritos na literatura para a interação entre a proteína L1 de HPV16 e 18 e as moléculas de heparan sulfato da superfície celular (DASGUPTA et al., 2011).

Por outro lado, estão presentes em uma região descrita como epítipo reconhecido por células T linfocitárias como descrito em estudo realizado com o objetivo de caracterizar a resposta de células T a peptídeos das proteínas L1, E6 e E7 de HPV58 em mulheres com infecções debruladas, lesões neoplásicas e câncer cervical (CHAN et al., 2010).

Considerando que os casos de lesão cervical intraepitelial grau II (NIC2) foram em mulheres com HPV58 da linhagem A, é possível que o HPV58 de linhagem C apresente uma evolução mais favorável. Adicionalmente, a infecção pelo HIV parece atuar como fator de pressão seletiva para a linhagem C do HPV58, com base na taxa das mutações G378D e D383N em 4 das 6 sequências com mutações não sinônimas de HPV58.

Estudos de variabilidade do genoma de HPV contribuem na descrição da distribuição de linhagens virais nas populações e seguimento dos programas de vacinação contra HPV, e em associação com estudos de modelagem computacional de proteínas virais permitem elucidar mecanismos de interação e invasão de células humanas. Nesse aspecto, diante dos dados apresentados nesse estudo, é importante avaliar a variabilidade genética e alterações conformacionais em proteínas do HPV em mulheres imunocompetentes.

7 CONCLUSÕES

- a) Mutações não sinônimas são menos frequente em HPVs de alto risco em comparação a baixo risco;
- b) A população de mulheres coinfectadas por HIV-HPV apresentou mutações não sinônimas com padrão definido nas sequências analisadas de fragmentos do gene L1 de HPV58;
- c) As mutações G378D e D383N definem a linhagem C em 4 das 6 sequências com mutações não sinônimas de HPV58
- d) As estruturas tridimensionais das sequências com a mutação G378D apresentam maior divergência estrutural em relação à estrutura da proteína selvagem;
- e) Considerando os sítios G378 e D383 envolvidos no reconhecimento por receptores de linfócitos T, é possível que as mutações G378D e D383N alterem a interação entre esse epítipo e os receptores de células T;
- f) É possível que a coinfecção pelo HIV seja fator de pressão seletiva para a linhagem C de HPV58.

REFERÊNCIAS

- BENSON, D. A. et al. GenBank. **Nucleic Acids Research**, London, v. 41, n. D1, p. 36–42, 2013.
- BERNARD, H. U. et al. Classification of papillomaviruses (PVs) based on 189 PV types and proposal of taxonomic amendments. **Virology**, New York, v. 401, n. 1, p. 70–79, 2010.
- BHATTACHARYA, D.; CHENG, J. 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization. **Proteins: Structure, Function and Bioinformatics**, New York, v. 81, May, p. 119–131, 2013.
- BIASINI, M. et al. SWISS-MODEL: Modelling protein tertiary and quaternary structure using evolutionary information. **Nucleic Acids Research**, London, v. 42, n. W1, p. 252–258, 2014.
- BOSCH, F. X. et al. The causal relation between human papillomavirus and cervical cancer. **Journal of Clinical Pathology**, London, v. 55, n. 4, p. 244–265, 2002.
- BUCK, C. B.; DAY, P. M.; TRUS, B. L. The papillomavirus major capsid protein L1. **Virology**, New York, v. 445, n. 1–2, p. 169–174, 2013.
- BURD, E. Human papillomavirus and cervical cancer. **Clinical Microbiology Reviews**, Washington, v. 16, n. 1, p. 1–17, 2003.
- BURK, R. D.; HARARI, A.; CHEN, Z. Human papillomavirus genome variants. **Virology**, New York, v. 445, n. 1–2, p. 232–243, 2013.
- BZHALAVA, D.; EKLUND, C.; DILLNER, J. International standardization and classification of human papillomavirus types. **Virology**, New York, v. 476, p. 341–344, 2015.
- CALLEJA-MACIAS, I. E. et al. Worldwide Genomic Diversity of the High-Risk Human Papillomavirus Types 31 , 35 , 52 , and 58 , Four Close Relatives of Human Papillomavirus Type 16. **Journal of Virology**, Washington, v. 79, n. 21, p. 13630–13640, 2005.
- CAMARA, G. N. L. et al. Prevalence of Human Papillomavirus Types in Women with Pre-neoplastic and Neoplastic Cervical Lesions in the Federal District of Brazil. **Memorias do Instituto Oswaldo Cruz**, Rio de Janeiro, v. 98, n. 7, p. 879–883, 2003.
- CANCHE, J. C. et al. High prevalence and low E6 genetic variability of human papillomavirus 58 in women with cervical cancer and precursor lesions in Southeast Mexico. **Memorias do Instituto Oswaldo Cruz**, Rio de Janeiro, v. 105, n. 2, p. 144–148, 2010.
- CHAN, P. K. S. et al. T-cell response to human papillomavirus type 58 L1, E6, and E7 peptides in women with cleared infection, cervical intraepithelial neoplasia, or invasive cancer. **Clinical and Vaccine Immunology**, Washington, v. 17, n. 9, p. 1315–1321, 2010.

- CHAN, P. K. S. et al. Identification of human papillomavirus type 58 lineages and the distribution worldwide. **Journal of Infectious Diseases**, Chicago, v. 203, n. 11, p. 1565–1573, 2011.
- CHAN, P. K. S. et al. Laboratory and clinical aspects of human papillomavirus testing. **Critical Reviews in Clinical Laboratory Sciences**, Boca Raton, v. 49, n. 4, p. 117–136, 2012.
- CHAN, P. K. S. et al. Geographical distribution and oncogenic risk association of human papillomavirus type 58 E6 and E7 sequence variations. **International Journal of Cancer**, New York, v. 132, n. 11, p. 2528–2536, 2013.
- CHEN, H. S. et al. Papillomavirus capsid proteins mutually impact structure. **Virology**, New York, v. 412, n. 2, p. 378–383, 2011.
- CHEN, Z.; DE FREITAS, L. B.; BURK, R. D. Evolution and Classification of Oncogenic Human Papillomavirus Types and Variants Associated with Cervical Cancer. In: KEPLER, D.; LIN, A.W. **Cervical Cancer: Methods and Protocols**. 1 ed. New York. Humana Press, v. 1249, p. 3–26, 2015.
- DASGUPTA, J. et al. Structural basis of oligosaccharide receptor recognition by human papillomavirus. **Journal of Biological Chemistry**, Baltimore, v. 286, n. 4, p. 2617–2624, 2011.
- DE SANJOSÉ, S. et al. Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: a meta-analysis. **Lancet Infectious Diseases**, New York, v. 7, n. 7, p. 453-459, 2007.
- DE VILLIERS, E. M. et al. Classification of papillomaviruses. **Virology**, New York, v. 324, n. 1, p. 17–27, 2004.
- DE VILLIERS, E. M. Cross-roads in the classification of papillomaviruses. **Virology**, New York, v. 445, n. 1–2, p. 2–10, 2013.
- DOORBAR, J. The papillomavirus life cycle. **Journal of Clinical Virology**, Amsterdam, v. 32, p. 7–15, 2005.
- EWING, B.; GREEN, P. Base-calling of automated sequencer traces using phred. II. Error probabilities. **Genome Research**, New York, v. 8, n. 3, p. 186–194, 1998.
- FAVRE, M. Structural polypeptides of rabbit, bovine, and human papillomaviruses. **Journal of virology**, Washington, v. 15, n. 5, p. 1239–1247, 1975.
- FELSENSTEIN, J. Confidence limits on phylogenies: an approach using the bootstrap. **Evolution**, Lancaster, v. 39, n. 4, p. 783-791, 1985.
- FERNANDES, J. et al. Prevalence of human papillomavirus type 58 in women with or without cervical lesions in northeast Brazil. **Annals of Medical and Health Sciences Research**, Mumbai, v. 3, n. 4, p. 504–510, 2013.
- FRATI, E. et al. Genetic variability in the major capsid L1 protein of human papillomavirus type 16 (HPV-16) and 18 (HPV-18). **Infection, Genetics and Evolution**, Amsterdam, v. 11, n. 8, p. 2119–2124, 2011.

GARNETT, T. O.; DUERKSEN-HUGHES, P. J. Modulation of apoptosis by human papillomavirus (HPV) oncoproteins. **Archives of Virology**, Wien, v. 151, n. 12, p. 2321–2335, 2006.

GIROGLOU, T. et al. Human Papillomavirus Infection Requires Cell Surface Heparan Sulfate. **Journal of Virology**, Washington, v. 75, n. 3, p. 1565–1570, 2001.

GRAHAM, S. V. Human papillomavirus: gene expression, regulation and prospects for novel diagnostic methods and antiviral therapies. **Future Microbiology**, London, v. 5, n. 10, p. 1493–1506, 2012.

GURGEL, A. P. A. D. et al. Prevalence of human papillomavirus variants and genetic diversity in the L1 gene and long control region of HPV16, HPV31, and HPV58 found in north-east Brazil. **BioMed Research International**, New York, v. 2015, 2015.

HARARI, A.; CHEN, Z.; BURK, R. D. Human Papillomavirus Genomics: Past, Present and Future. **Current Problems in Dermatology**, Basel, v. 45, p. 1–18, 2014.

HAUSEN, ZUR H. Papillomavirus infections — a major cause of human cancers. **Biochimica et Biophysica Acta (BBA) - Reviews on Cancer**, Amsterdam, v. 1288, n. 2, p. F55–F78, 1996.

HILDESHEIM, A. et al. Effect of human papillomavirus 16/18 L1 viruslike particle vaccine among young women with preexisting infection: a randomized trial. **JAMA: the journal of the American Medical Association**, Chicago, v. 298, n. 7, p. 743–753, 2007.

JOURA, E. A. et al. A 9-Valent HPV Vaccine against Infection and Intraepithelial Neoplasia in Women. **The New England journal of medicine**, Boston, v. 372, n. 8, p. 711–723, 2015.

KANESHIMA, E. N. et al. Importância da aplicação de critérios morfológicos não-clássicos para o diagnóstico citopatológico de Papillomavirus humano (HPV) previamente detectado. **Acta Bioquímica Clínica Latinoamericana**, Buenos Aires, v. 39, n. 1, p. 61–68, 2005.

KELLEY, L. A. et al. The Phyre2 web portal for protein modeling, prediction and analysis. **Nature Protocols**, London, v. 10, n. 6, p. 845–858, 2015.

KIM, K.; PARK, S.; KO, K. Current status of human papillomavirus vaccines. **Clinical and experimental vaccine research**, Seoul, v. 3, n. 2, p. 168–175, 2014.

KIMURA, M. A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. **Journal of Molecular Evolution**, New York, v. 16, n. 2, p. 111–120, 1980.

LIU, J. H. et al. Variations of human papillomavirus type 58 E6, E7, L1 genes and long control region in strains from women with cervical lesions in Liaoning province, China. **Infection, Genetics and Evolution**, Amsterdam, v. 12, n. 7, p. 1466–1472, 2012.

LOVELL, S. C. et al. Structure validation by C α geometry: phi, psi and C β deviation. **Proteins**, New York, v. 50, n. 3, p. 437–450, 2003.

MADDEN, T. The BLAST sequence analysis tool. In: HOEPPNER, M.; OSTELL, J. **The NCBI Handbook**. 2. ed. Bethesda: NCBI, 2013. p. 1–17.

MANOS, M. M. et al. The use of polymerase chain reaction amplification for the detection of genital human papillomaviruses. **Cancer Cells**, New York, v. 7, p. 209–214, 1989.

MARIANI, V. et al. Assessment of template based protein structure predictions in CASP9. **Proteins: Structure, Function and Bioinformatics**, New York, v. 79, n. Suppl. 10, p. 37–58, 2011.

MARTINS, A. E. S. et al. Prognostic evaluation of DNA index in HIV-HPV co-infected women cervical samples attending in reference centers for HIV-AIDS in Recife. **PLoS One**, San Francisco, v. 9, n. 8, p. e104801, 2014a.

MARTINS, A. E. S. et al. Prevalence of human papillomavirus infection, distribution of viral types and risk factors in cervical samples from human immunodeficiency virus-positive women attending three human immunodeficiency virus-acquired immune deficiency syndrome reference cent. **Memórias do Instituto Oswaldo Cruz**, Rio de Janeiro, v. 109, n. 6, p. 738–747, 2014b.

MEJÍA, L. et al. Prevalence of human papillomavirus types in cervical cancerous and precancerous lesions of Ecuadorian women. **Journal of Medical Virology**, New York, v. 88, n. 1, p. 144–152, 2016.

MOODY, C. A.; LAIMINS, L. A. Human papillomavirus oncoproteins: pathways to transformation. **Nature reviews. Cancer**, London, v. 10, n. 8, p. 550–560, 2010.

OLIVELLA, M. et al. Relation between sequence and structure in membrane proteins. **Bioinformatics**, Oxford, v. 29, n. 13, p. 1589–1592, 2013.

PERES, J. For cancers caused by HPV, two vaccines were just the beginning. **Journal of the National Cancer Institute**, Cary, v. 103, n. 5, p. 360–362, 2011.

RAFF, A. B. et al. The evolving field of human papillomavirus receptor research: a review of binding and entry. **Journal of virology**, Washington, v. 87, n. 11, p. 6062–72, 2013.

SADOWSKI, M. I.; TAYLOR, W. R. Protein structures, folds and fold spaces. **Journal of physics. Condensed matter**, Bristol, v. 22, n. 3, p. 33103, 2010.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular biology and evolution**, Chicago, v. 4, n. 4, p. 406–25, 1987.

SANJOSÉ, S. et al. Worldwide prevalence and genotype distribution of cervical HPV in women with normal cytology. **Lancet Infectious Diseases**, New York, v. 7, July, p. 453–459, 2007.

SCHIFFMAN, M. et al. Human papillomavirus and cervical cancer. **Lancet**, London, v. 370, n. 9590, p. 890–907, 2007.

SCHILLER, J. T.; DAY, P. M.; KINES, R. C. Current understanding of the mechanism of HPV infection. **Gynecologic Oncology**, New York, v. 118, n. 1, Suppl. 1, p. S12–

S17, 2010.

SCHWEDE, T. et al. SWISS-MODEL: An automated protein homology-modeling server. **Nucleic Acids Research**, London, v. 31, n. 13, p. 3381–3385, 2003.

STADEN, R. The Staden sequence analysis package. **Molecular Biotechnology**, Totowa, v. 5, p. 233–241, 1996.

TAMURA, K. et al. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. **Molecular biology and evolution**, Chicago, v. 28, n. 10, p. 2731–2739, 2011.

TAMURA, K. et al. MEGA6: Molecular evolutionary genetics analysis version 6.0. **Molecular Biology and Evolution**, Chicago, v. 30, n. 12, p. 2725–2729, 2013.

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. **Nucleic acids research**, London, v. 22, n. 22, p. 4673–4680, 1994.

VON BUBNOFF, A. Taking the gritty approach. **IAVI Report: newsletter on international AIDS vaccine research**, New York, v. 16, p. 4-9, 2012.

WU, E. Q. et al. Profile of physical status and gene variation of human papillomavirus 58 genome in cervical cancer. **Journal of General Virology**, London, v. 90, n. 5, p. 1229–1237, 2009.

XU, D.; ZHANG, Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. **Biophysical journal**, New York, v. 101, n. 10, p. 2525–2534, 2011.

ZHANG, Y. I-TASSER server for protein 3D structure prediction. **BMC bioinformatics**, London, v. 9, 2008.

ZHANG, Y.; SKOLNICK, J. TM-align: a protein structure alignment algorithm based on the TM-score. **Nucleic acids research**, London, v. 33, n. 7, p. 2302–2309, 2005.

ZHENG, Z.-M.; BAKER, C. C. Papillomavirus genome structure, expression, and post-transcriptional regulation. **Frontiers in bioscience**, Tampa, v. 11, p. 2286–2302, 2006.

ZUR HAUSEN, H. Papillomaviruses and Cancer: From Basic Studies to Clinical Application. **Nature**, London, v. 2, p. 342–350, May 2002.

ZUR HAUSEN, H. Papillomaviruses in the causation of human cancers - a brief historical account. **Virology**, New York, v. 384, n. 2, p. 260–265, 2009.