UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO
DEPARTAMENTO DE GENÉTICA

NILSON NICOLAU JUNIOR

**Diferenças estruturais e “docking” receptor-ligante
da proteína E7 do vírus do papiloma humano (HPV)
de alto e baixo riscos para o câncer cervical**

Ribeirão Preto - SP

2013

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Nicolau Junior, Nilson

Diferenças estruturais e “docking” receptor-ligante da proteína E7 do vírus do papiloma humano (HPV) de alto e baixo riscos para o câncer cervical / Nilson Nicolau Junior; Orientadora: Silvana Giuliatti, Ribeirão Preto, 2013.

125 p. : il.

Tese (Doutorado) – Universidade de São Paulo, 2013.

1. Modelagem Macromolecular. 2. Dinâmica Molecular. 3. Desordem Intrínseca. 4. Farmacóforos. 5. “Docking”. 6. HPV. 7. E7.

NILSON NICOLAU JUNIOR

**Diferenças estruturais e “docking” receptor-ligante da
proteína E7 do vírus do papiloma humano (HPV) de
alto e baixo riscos para o câncer cervical**

Tese apresentada à Faculdade de Medicina de
Ribeirão Preto da Universidade de São Paulo
para a obtenção do título de Doutor em
Ciências

Área de concentração: Genética

Orientadora: Prof.^a Dr.^a. Silvana Giuliatti

Ribeirão Preto

2013

FOLHA DE APROVAÇÃO

Nilson Nicolau Junior

Diferenças estruturais e “docking” receptor-ligante da proteína E7 do vírus do papiloma humano (HPV) de alto e baixo riscos para o câncer cervical.

Tese apresentada à Faculdade de
Medicina de Ribeirão Preto da
Universidade de São Paulo, para
obtenção do título de Doutor em
Ciências.

Área de concentração: Genética

Aprovada em:

Banca examinadora

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr. _____ Instituição: _____

Julgamento: _____ Assinatura: _____

À minha esposa, Mariana, pelo amor, dedicação e apoio.

Aos meus pais, Nilson e Marlene, pela vida, amor e incentivo.

Aos meus irmãos, André e Andréa, pela amizade e apoio.

Aos meus sobrinhos, Bianca, Camila, Julia, Lucas e Rafael,
pelos bons momentos juntos.

A todos aqueles que de uma forma ou de outra contribuíram
na minha tese de Doutorado e na minha vida pessoal.

AGRADECIMENTOS

Agradeço em especial a Professora Doutora **Silvana Giuliatti** pela orientação, compreensão, direcionamento e principalmente sabedoria.

A minha esposa, **Mariana Sant'Anna Pereira Nicolau**, pelo companheirismo e apoio incondicional.

Aos meus pais, **Nilson Nicolau** e **Marlene Aparecida Calçada Nicolau**, que sempre me deram amor e incentivo e sempre foram fundamentais para meu desenvolvimento pessoal e profissional.

Aos meus irmãos, **André** e **Andréa**, pelo apoio e amizade de sempre.

Aos meus sogros, **José Aparecido** e **Darcy**, pelo incentivo e apoio.

Ao **Wilson, Bianca, Rafael, Marcelo, Cláudia, Bada, Camila, Lucas e Júlia**, pelo carinho e companheirismo.

Aos amigos e companheiros de laboratório **Daniel, Saulo, André, Fernando, Fernanda, Luciano, Pablo, Iuliana, Beatriz e Gabriela**, pelo apoio e amizade.

Ao Professor Doutor **Ademilson Espencer Egea Soares** pelo apoio ao longo do desenvolvimento do meu projeto de doutorado.

À **Susie** pelo suporte técnico e acadêmico.

Ao Professor Doutor **Christoph A. Sotriffer**, pela possibilidade de realizar parte do meu projeto de doutorado no Instituto de Farmácia e Química de Alimentos da Universidade de Würzburg, Alemanha.

Aos colegas da Universidade de Würzburg, **Benjamin, Daniel, David, Manuel, Michael, Raphael, Sarah, Steffen, Thomas** e **Yogesh**, pelo apoio no desenvolvimento do meu projeto de Doutorado e pelos bons momentos que passei na Alemanha.

À **USP, à Faculdade de Medicina e Ribeirão Preto** e ao **Programa de Pós Graduação em Genética** pela oportunidade da realização do meu curso de Doutorado.

À **CAPES, CNPq e FAPESP**, pelo apoio financeiro durante o desenvolvimento do projeto.

Saber muito não lhe torna inteligente. A inteligência se traduz na forma que você recolhe,
julga, maneja e, sobretudo, onde e como aplica esta informação.

Carl Sagan

RESUMO

NICOLAU-JUNIOR, N. **Diferenças estruturais e “docking” receptor-ligante da proteína E7 do vírus do papiloma humano (HPV) de alto e baixo riscos para o câncer cervical.** 2013. 125 f. Tese (Doutorado) - Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, Ribeirão Preto, 2013.

O câncer cervical afeta milhões de mulheres em todo o mundo a cada ano. A maioria dos casos de câncer cervical é causada pelo vírus do papiloma humano (HPV) que é sexualmente transmissível. Cerca de 40 tipos de HPV infectam o colo do útero e estes são designados como sendo de alto ou de baixo risco com base no seu potencial para provocar lesões de alto grau e câncer. A oncoproteína E7 do HPV está diretamente envolvida no aparecimento de câncer de colo do útero. Esta se associada com a proteína pRb e outros alvos celulares que promovem a imortalização celular e carcinogênese. Apesar de muito progresso nos estudos sobre os HPVs de alto risco, ainda não existe uma terapêutica adequada para o tratamento das lesões e câncer causados por este vírus. Este trabalho teve como objetivo entender as diferenças estruturais entre E7 de alto e baixo risco e sugerir, através de análises de bioinformática, possíveis sítios de ligação e inibidores para a E7. Esta é a primeira descrição da modelagem e análise de dinâmica molecular de quatro estruturas tridimensionais completas da E7 dos tipos de alto risco (HPV tipos 16 e 18), de baixo risco (HPV tipo 11) e não relacionadas ao câncer cervical (HPV tipo 1A). Os modelos foram construídos por uma abordagem híbrida usando modelagem por homologia e *ab initio*. Os modelos foram usados em simulações de dinâmica molecular por 50 ns, sob condições normais de temperatura e pressão. A desordem intrínseca da sequência da proteína E7 foi avaliada com o uso de ferramentas *in silico*. Os domínios N-terminal de todas as E7 estudadas, mesmo as de alto risco, exibiram estruturas secundárias depois da modelagem. Nas análises da trajetória da dinâmica molecular, as E7s dos HPVs dos tipos 16 e 18 apresentaram maior instabilidade nos seus domínios N-terminais em relação aos do HPV dos tipos 11 e 01. No entanto, esta variação não afetou a conformação das estruturas secundárias durante a simulação. A análise com ANCHOR indicou que as regiões CR1 e CR2 regiões dos tipos de HPV 16 e 18 contêm possíveis alvos para a descoberta da droga. Já a região CR3 do domínio C-terminal indicou estabilidade nas análises *in silico* e, por isso, foi usada como alvo de busca de modelos farmacofóricos e “docking” macromolecular. A proteína usada como modelo foi a E7 do HPV tipo 45 resultante de análises de ressonância magnética nuclear (RMN) e depositada no banco

de dados de proteína (ID: 2F8B). Foram selecionados por análises sequenciais de busca farmacofórica, “docking” e “re-docking”, 19 compostos (extraídos de amplas bibliotecas de pequenos ligantes) com potencial para candidatos a inibidores da E7. Eles foram avaliados quanto a sua função de pontuação, mapas de interação receptor-ligante e toxicidade e os melhores foram indicados para estudos futuros.

Palavras-chave: Modelagem Macromolecular. Dinâmica Molecular. Desordem Intrínseca. Farmacóforos. “Docking”. HPV. E7.

ABSTRACT

NICOLAU-JUNIOR, N. **Structural differences and receptor-ligand docking of E7 protein from human papillomavirus (HPV) of high and low risk for cervical cancer.** 2013. 125 f. Tese (Doutorado) - Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, Ribeirão Preto, 2013.

Cervical cancer affects millions of women around the world each year. Most cases of cervical cancer are caused by human papilloma virus (HPV) which is sexually transmitted. About 40 types of HPV infect the cervix and these are designated as being at high or low risk based on their potential to cause high-grade lesions and cancer. The E7 oncoprotein from HPV is directly involved in the onset of cervical cancer. It associates with the pRb protein and other cellular targets that promote cell immortalization and carcinogenesis. Although the progress in studies with high-risk HPVs there is still no adequate therapy for the treatment of lesions and cancers caused by this virus. This study aimed to understand the structural differences between E7 of high and low risk and suggest, with the aid of bioinformatics analyzes, possible binding sites and inhibitors for the E7. This is the first description of the modeling and molecular dynamics analysis of four complete three-dimensional structures of E7 from high-risk types (HPV types 16 and 18), low risk (HPV type 11) and that not related to cervical cancer (HPV 01). The models were constructed by a hybrid approach using homology modeling and *ab initio*. The models were used in molecular dynamics simulations for 50 ns, under normal temperature and pressure. The intrinsic disorder of the E7 protein sequence was assessed using *in silico* tools. The N-terminal domains of all E7s, even the high-risks, showed secondary structures after modeling. In the trajectory analyzes of molecular dynamics, the E7s of HPV types 16 and 18 showed high instability in their N-terminal domains than those of HPV types 11 and 01, however, this variation did not affect the conformation of secondary structures during the simulation. The analysis with ANCHOR indicated that regions CR1 and CR2 regions of types of HPV 16 and 18 contain possible targets for drug discovery. The CR3 region of the C-terminal domain indicated stability by *in silico* analyzes and was therefore used as target to search for pharmacophoric models and "docking". The protein used as a model was the E7, from HPV type 45, constructed by analysis of nuclear magnetic resonance (NMR) and deposited in the protein data bank (ID: 2F8B). It was selected 19 compounds as potential candidates for E7 inhibitors (extracted from large libraries of small ligands) using sequential pharmacophore search, docking and re-docking analyzes. They were evaluated for

their scoring function, maps of receptor-ligand interactions and toxicity and the best suited were indicated for future studies.

Keywords: Macromolecular Modeling. Molecular Dynamics. Intrinsically Disorder. Pharmacophores. Docking. HPV. E7.

LISTA DE FIGURAS

Figura 1 - Fluxograma da metodologia resumida desenvolvida ao longo do projeto.....	32
Figura 2 - Análise da desordem intrínseca ao longo da cadeia de aminoácidos da proteína E7 dos HPVs dos tipos 1A, 11, 16, e 18 usando PONDR-FIT.....	42
Figura 3 - Resultado do gráfico de Ramachandran para o modelo 3 da E7 do HPV tipo 16, gerado por modelagem por homologia. Os quadrados negros representam os aminoácidos comuns e os triângulos negros representam o aminoácido glicina.	44
Figura 4 - Modelo 3 da E7 do HPV tipo 16 gerado por modelagem por homologia parcial, apenas do domínio C-terminal (azul). A região em vermelho representa o domínio N-terminal da E7 que não apresentou homólogos no banco de dados do PDB.	45
Figura 5 - Modelos completos, S_109 (A) e S_139 (B), da E7 do HPV tipo 16, gerados pela modelagem por homologia da região C-terminal (azul) e modelagem por <i>ab initio</i> da região N-terminal.	47
Figura 6 - Estruturas refinadas da E7 (α -hélices = verde, β -folhas = amarelo e “loops” = azul) com o átomo de zinco (vermelho). (a) HPV tipo 1A, (b) HPV tipo 11, (c) HPV tipo 16, e (d) HPV tipo 18. As figuras foram geradas usando CHIMERA (PETTERSEN et al., 2004).....	49
Figura 7 - Representação da proteína E7 do HPV tipo 16 inserida em uma caixa de água usada na simulação da dinâmica molecular.	50
Figura 8 - Desvio da média quadrática (RMSD) das coordenadas dos átomos C α em função do tempo de simulação, calculado para: (a) HPV tipo 16 vs. HPV tipo 1A, (b) HPV tipo 16 vs. HPV tipo 11 e (c) HPV tipo 16 vs. HPV tipo 18.	51

Figura 9 - Flutuações da média da raiz quadrática (RMSF) das coordenadas dos átomos C α para cada resíduo em função do tempo de simulação calculado para: (a) HPV tipo 16 vs. HPV tipo 1A, (b) HPV tipo 16 vs. HPV tipo 11 e (c) HPV tipo 16 vs. HPV tipo 18..... 53

Figura 10 - Estabilidade das estruturas secundárias durante a simulação, calculada pelo algoritmo STRIDE implementado no “plugin” Timeline do VMD. Cada cor representa um tipo de estrutura secundária: α -hélice = magenta, β -folha = amarelo, “turns” = ciano, “ loops” = branco, 3_{10} -hélice = azul, e π -hélice = vermelho. 55

Figura 11 - Análise da desordem intrínseca dos sítios de ligação da E7 ao longo da cadeia de aminoácidos dos tipos de HPV 1A, 11, 16 e 18, usando ANCHOR. 57

Figura 12 - Superfície do domínio C-terminal de um dímero da E7 do HPV tipo 45; as esferas brancas e vermelhas indicam o sítio 6 gerado pela análise feita no SiteFinder (MOE). 58

Figura 13 - Superfície do sítio de ligação 6 do domínio C-terminal da E7 do HPV tipo 45 com os “grids” gerados pela interação entre a sonda e a superfície; (a) sonda C.ar (laranja), (b) sonda C1 (laranja), (c) sonda N.ams (magenta) e (d) sonda N1 (magenta). 60

Figura 14 - Modelo farmacofórico 11, no sítio de ligação 6, gerado com base nas análises de “hot spots” e validada por busca farmacofórica em bibliotecas de pequenos compostos utilizando MOE. F1 e F3 são pontos de anotação Don e F2 Aro. 61

Figura 15 - Exemplo de um composto que convergiu em relação seu posicionamento tridimensional (a) após o “docking” e um composto que não convergiu (b). 63

Figura 16 - Complexo do composto ZINC03163900 com a proteína E7 junto com o farmacóforo 16. A imagem evidencia os pontos de anotação (esferas grandes) centralizados com os respectivos átomos e outros pontos de anotação que foram gerados automaticamente pelo MOE devido a presença da estrutura do composto (esferas pequenas)..... 64

Figura 17 - Representação do complexo E7-ligante obtidos por “docking” macromolecular. Os potenciais de cada átomo das interações são representados por esferas azuis e vermelhas (bons e maus contatos, respectivamente) resultantes de análise com DSX e Pymol. Os compostos são: (a) ZINC70441103; (b) ZINC03163900; (c) ZINC36377976; (d) ZINC38069241; (e) ZINC01513677 e (f) ZINC2519022..... 71

Figura 18 - Representação do complexo E7-ligante obtidos por “docking” macromolecular analisados pelo DSX. Os compostos são: (a) ZINC16400113; (b) EMC 001A1113; (c) ZINC5889902; (d) ZINC38069289; (e) ZINC04770631 e (f) ZINC59957745..... 72

Figura 19 - Representação do complexo E7-ligante obtidos por “docking” macromolecular analisados pelo DSX. Os compostos são: (a) ZINC06095504; (b) ZINC00519056; (c) ZINC37246053; (d) ZINC5898533; (e) ZINC4060682; (f) ZINC9458158 e (g) ZINC01594847..... 73

Figura 20 - Ligantes que apresentaram 3 ou mais pontes de hidrogênio com a proteína E7: (a) ZINC16400113, (b) EMC 001A1113, (c) ZINC5889902 e (d) ZINC9458158. Os círculos rosa indicam resíduos envolvidos em interações de pontes de hidrogênio ligações de carga ou polar; os círculos verdes indicam resíduos envolvidos em interações de van der Waals; a linha tracejada em verde representa interações de ponte de hidrogênio com a cadeia primária de aminoácidos; a linha tracejada em azul representa interações de ponte de hidrogênio com a cadeia lateral dos aminoácidos. 75

Figura 21 - Ligantes que não apresentaram pontes de hidrogênio com a proteína E7: (a) ZINC38069241, (b) ZINC38069289 e (c) ZINC4060682. Os círculos rosa indicam resíduos envolvidos em interações de pontes de hidrogênio ligações de carga ou polar; os círculos verdes indicam resíduos envolvidos em interações de van der Waals..... 76

LISTA DE TABELAS

Tabela 1- Análise do conteúdo de desordem da proteína E7 dos tipos de HPV1A, 11, 16, 18 usando DisCon.	42
Tabela 2 - Resultados da análise do gráfico de Ramachandran (PROCHECK) para os modelos incompletos, da E7 do HPV tipo 16, gerados por modelagem por homologia.	43
Tabela 3 - Seleção dos melhores modelos da E7 do HPV tipo 16 baseado em pontuações de MQAPs.....	46
Tabela 4 - Seleção dos modelos da E7 dos HPVs tipo 1A, 11 e 18 gerados por homologia com ajuda dos valores do gráfico de Ramachandran (PROCHECK).	48
Tabela 5 - Exemplo das pontuações, por GoldScore, que convergiram e não convergiram ao longo das 10 interações do “docking”	63
Tabela 6 - Pontuações da interação entre os compostos e a proteína E7 do HPV tipo 45, selecionados através da busca farmacofórica e docking macromolecular.	66
Tabela 7 - Os 19 melhores compostos selecionados a partir de funções de pontuação e suas respectivas estruturas e características físico-químicas.	67
Tabela 8 - Predição de toxicidade dos 19 compostos selecionados utilizando o programa DEREK.....	78

LISTA DE SIGLAS

CCI	Carcinoma Cervical Invasivo
DM	Dinâmica Molecular
IDP	Intrinsically Disordered Protein (Proteína Intrinsecamente Desordenada)
IUP	Intrinsically Unstructured Protein (Proteína Intrinsecamente Desestruturada)
MQAP	Model Quality Assessment Program (Programa de Avaliação de Qualidade de Modelo)
NIC	Neoplasia Intra-epitelial Cervical
NPT	Normal Temperature and Pressure (Temperatura e Pressão Normais)
PDB	Protein Data Bank (Banco de Dados de Proteínas)
PPI	Protein-Protein Interaction (Interação Proteína-Proteína)
pRb	Proteína do Retinoblastoma
RMN	Ressonância Magnética Nuclear
RMSD	Root Mean Square Deviation (Desvio da Média Quadrática)
RMSF	Root Mean Square Fluctuation (Flutuação da Média Quadrática)

LISTA DE SÍMBOLOS

Å	angstrom
atm	atmosfera padrão
Ca	carbono alfa
fs	femtosegundos
K	escala Kelvin
ns	nanosegundos
µs	microsegundos
ps	picosegundos
xlogP	método de cálculo de coeficiente de partição etanol/água

SUMÁRIO

1 INTRODUÇÃO	18
1.1 Câncer Cervical e HPV	18
1.2 Oncoproteína E7	19
1.3 Desordem Intrínseca	22
1.4 Análises de Bioinformática Estrutural	23
1.4.1 Modelagem Macromolecular	23
1.4.2 Modelagem por Homologia	24
1.4.3 Modelagem por <i>ab initio</i>	24
1.4.4 Dinâmica Molecular	26
1.4.5 Farmacóforos	26
1.4.6 “Docking” Macromolecular	27
1.5 Motivação	28
2 HIPÓTESES	29
3 OBJETIVOS	30
3.1 Objetivos Específicos	30
4 MATERIAIS E MÉTODOS	32
4.1 Análise <i>in silico</i> de desordem intrínseca	33
4.2 Modelagem Macromolecular	33
4.2.1 Seleção de Molde e Modelagem por Homologia da E7 do HPV tipo 16	33
4.2.2 Modelagem <i>ab initio</i>	33
4.2.3 Modelagem por Homologia da E7 do HPV dos Tipos 1A, 11 e 18	34
4.2.4 Modelagem de Rotâmetros e Inserção das Coordenadas do Zinco	34
4.2.5 Validação dos Modelos	35
4.3 Simulações de Dinâmica Molecular	36
4.4 Análises das Trajetórias da Dinâmica Molecular	36
4.5 Modelagem de Farmacóforos	37
4.5.1 Busca de sítios de ligação	37
4.5.2 Análise de “hot spots” e Modelagem Farmacofórica	38
4.5.3 Busca Farmacofórica	38
4.6 “Docking” Macromolecular	39

4.6.1 “Re-docking” e Avaliação	39
4.7 Análise de Toxicidade <i>in silico</i>	40
5 RESULTADOS E DISCUSSÃO	41
5.1 Avaliação da desordem intrínseca das E7s de HPVs de alto e baixo risco.....	41
5.2 Modelagem Molecular da proteína E7	43
5.3 Análise de Dinâmica Molecular.....	50
5.4 Análises de Sítios de Ligação da Região Desordenada	56
5.5 Busca de Sítios de Ligação no C-terminal da E7 de Alto Risco	57
5.6 “Hot spots” e Modelos Farmacofóricos da E7	59
5.7 “Docking” Macromolecular do C-terminal da E7 do HPV tipo 45	62
5.8 Análise de Toxicidade dos compostos selecionados	77
6 CONCLUSÕES	80
REFERÊNCIAS	82
APÊNDICES	91
APÊNDICE A	91
APÊNDICE B.....	92
APÊNDICE C.....	93
APÊNDICE D	101

1 INTRODUÇÃO

1.1 Câncer Cervical e HPV

O câncer cervical é o segundo tipo de câncer mais comum entre as mulheres (PARKIN; BRAY, 2006). O Carcinoma Cervical Invasivo (CCI) é precedido por lesões precursoras 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 em exames citológicos de células cervicais esfoliadas e confirmado em exame histológico do material. Estas alterações pré-malignas representam um espectro de anormalidades histológicas e variam de NIC1 (displasia leve), NIC2 (displasia moderada) e NIC3 (displasia grave/carcinoma *in situ*). Embora o tratamento de lesões seja terapeuticamente eficaz, ele é ineficiente, em termos procedimentais, pois os exames citológicos e histológicos não conseguem distinguir se as anormalidades encontradas vão se tornar um câncer invasivo ou se vão regredir como na maioria dos casos (WOODMAN; COLLINS; YOUNG, 2007).

Para o surgimento do câncer do colo do útero, na grande maioria dos casos, a condição necessária é a presença de infecção pelo vírus do papiloma humano (HPV). A infecção por HPV cervical é sexualmente transmissível e a maioria das mulheres é infectada logo após o início da sua vida sexual. Mais de 100 tipos de vírus do papiloma humano (HPV) foram identificados, dos quais 40 infectam o trato genital (DE VILLIERS et al., 2004). A maioria dos indivíduos permanece assintomático depois de adquirirem a infecção e apenas uma pequena porcentagem desenvolve lesões reconhecíveis clinicamente ou histologicamente durante o curso da doença (KANODIA; FAHEY; KAST, 2007). Estudos epidemiológicos e moleculares têm demonstrado que o papiloma vírus humano induz a transformação celular epitelial durante a infecção. Infecções causadas por tipos específicos de HPV parece ser o passo essencial para o desenvolvimento de lesões precursoras e o aparecimento do câncer cervical invasivo (ZUR HAUSEN et al., 1974). Os tipos de HPV que infectam o trato genital podem ser classificados em dois grupos, os de alto risco e de baixo risco, baseados no tipo do

risco relativo de lesões de alto-grau e à progressão do câncer cervical (KLETER et al., 1998). Os tipos de baixo risco induzem apenas verrugas genitais benignas e incluem HPV 6, 11 entre outros. O grupo de alto risco compreende os tipos 16, 18, 31, 33, 45, 56 entre outros, que estão associados ao desenvolvimento de cânceres anogenitais e podem ser detectados em 99% dos cânceres cervicais (KANODIA; FAHEY; KAST, 2007).

A infecção persistente com tipos de HPV de alto risco é um fator etiológico necessário para o desenvolvimento de câncer de colo do útero, e os tipos HPV 16 e HPV 18 são, de longe, os mais prevalentes tipos de alto risco e estão associados com, aproximadamente, 70% dos cânceres cervicais no mundo (SCHEFFNER et al., 1990; ZUR HAUSEN, 2002). As oncoproteínas E6 e E7 dos HPVs de alto risco são essenciais para a imortalização, transformação e carcinogênese do HPV (NARISAWA-SAITO; KIYONO, 2007), pois, estas agem interagindo com vários alvos celulares (BOULET et al., 2007). Os principais alvos da E6 e da E7 são importantes genes supressores de tumor, a p53 e a pRb, respectivamente. A E6 age mais especificamente como uma repressora da apoptose, enquanto a E7 promove a replicação e crescimento celular. Porém, ambas podem, independentemente, imortalizar as células humanas e juntas potencializam este efeito (ZUR HAUSEN, 2000).

1.2 Oncoproteína E7

A E7 está localizada primariamente no núcleo, onde ela se associa com o produto do gene do retinoblastoma (pRb) facilitando a progressão da fase S do ciclo celular (ZUR HAUSEN, 2002). Em células normais, a pRb é hipofosforilada no início da fase G1 e liga-se com fatores de transcrição E2F, formando complexos que funcionam como repressores transcricionais. Associando-se com a pRb hipofosforilada, a E7 impede a ligação com E2F, dessa forma promovendo a progressão do ciclo celular (BOULET et al., 2007). Além da pRb, as E7s de alto risco também apresentam outros alvos na célula. A E7 estimula os genes das ciclinas E e A da fase S do ciclo celular, interage com complexos ciclina-quinase e anula a atividade inibitória de inibidores dependentes de quinase (CKIs) como a p21 e a p27 levando a estimulação do crescimento celular nas células infectadas por HPV (ZUR HAUSEN, 2000).

Experimentos de equilíbrio de sedimentação mostraram que a E7 apresenta propriedades de oligomerização e pode ser encontrada na forma de monômeros, dímeros e

tetrâmeros, mas, sobre condições fisiológicas, esta é, primariamente dimérica (CLEMENTS et al., 2000). Entretanto, análises de estequiometria e de afinidade sub-nanomolar indicaram que a E7 liga-se à pRb como um monômero (CHEMES et al., 2010).

A proteína E7 apresenta três regiões conservadas, chamadas de CR, baseadas na sua homologia com a proteína E1A do adenovírus: CR1, CR2 e CR3 (PHELPS et al., 1988). As regiões CR1 e CR2 estão localizadas no domínio N-terminal da proteína (resíduos 1-40 no HPV tipo 16); este domínio apresenta propriedades de proteínas intrinsecamente desordenadas (IDPs), que conferem flexibilidade, transições conformacionais, e adaptabilidade estrutural para alvos da oncoproteína (GARCÍA-ALAI; ALONSO; DE PRAT-GAY, 2007). A região CR1 (resíduos 2-15) é conhecida por competir com os fatores de transcrição E2F e participar na desestabilização da pRb, embora ainda não esteja claro que aminoácidos específicos estão envolvidos nestas atividades (HELT; GALLOWAY, 2001). A região CR2 (resíduos 16-40) possui uma sequência específica de aminoácidos, o motivo LXCXE, que tem uma alta afinidade pela proteína pRb e é necessária para a associação entre a E7 e a pRb (DYSON et al., 1989; JONES et al., 1990, 1992).

A região CR3 está localizada no domínio C-terminal da proteína e tem dois motivos CXXC que se ligam por pontes dissulfeto a um átomo de zinco (BARBOSA; LOWY; SCHILLER, 1989), sendo estas características de motivos dedo de zinco. Os motivos CXXC funcionam como um domínio de dimerização da proteína (MCINTYRE et al., 1993). Porém, não existem evidências de que a E7 existe na forma de um dímero *in vivo* ou que a dimerização é necessária para as atividades biológicas da E7 (MCLAUGHLIN-DRUBIN; MÜNGER, 2009). A CR3 coopera com as regiões CR1 e CR2 na inibição da pRb e na transformação da célula. De acordo com Patrick, Oliff e Heimbroke (1994), a E7 completa, contendo a CR3, liga-se à pRb com afinidade, aproximadamente, 100 vezes maior do que fragmentos que não possuem a região CR3. Estudos com o domínio CR3 de uma E7 dimérica do HPV tipo 1A compreendia a análise de dois “patches” conservados na superfície da proteína; o “patch” 1 faz contatos de baixa afinidade com a região básica dentro do domínio C-terminal da pRb, enquanto o “patch” 2 faz contatos com o domínio conhecido como “marked-box” na região C-terminal da E2F (E2F-MB) (LIU et al., 2006). Liu et al. (2006) sugeriram que a ligação de pequenas moléculas aos “patches” 1 e 2 da região CR3 da E7 poderiam inibir a habilidade da E7 em perturbar o funcionamento da pRb e, assim, o desenvolvimento de novas moléculas poderiam abrir um novo campo de intervenções terapêuticas contra o HPV.

O recente desenvolvimento de vacinas para o HPV tem representado um grande avanço nos métodos preventivos do câncer humano. Porém, apesar das grandes expectativas os resultados promissores de testes clínicos, ainda não existem provas suficientes de uma vacina eficaz contra o câncer cervical, ou seja, o efeito global das vacinas ainda permanece desconhecido (HAUG, 2008). Além disso, a vacina é cara (completa em três doses, no Brasil custa entre R\$ 250 a R\$ 400 por dose), e será um grande desafio implementá-la onde ela é mais necessária e tem um impacto mais significativo, como nos países de baixa renda (HAMMOUD, 2008). O elevado custo das vacinas atuais tem levado pesquisadores a continuar buscando novos métodos de produção e entrega. A maioria dos métodos está em fase pré-clínica ou fases iniciais piloto. Além de estratégias preventivas, como as vacinas, estratégias terapêuticas permanecem de interesse para os milhões de mulheres que, atualmente, estão infectadas e/ou que tem doenças que exigem tratamento (MOSCICKI, 2008). Beutner e Ferenczy (1997) afirmaram, nos anos 90, que não existia tratamento médico para doenças induzidas por HPVs e que a única terapia era a destruição ou a retirada do tecido infectado por agentes citotóxicos ou cirurgia. Hoje, mais de 15 anos depois, a não ser pelos métodos preventivos como as vacinas recém criadas, nenhuma metodologia terapêutica foi criada de forma efetiva para o tratamento das pessoas infectadas com HPV ou que desenvolveram o câncer cervical devido a esta infecção.

Os genes E6 e E7 são colocados hoje na posição de alvos primários no desenvolvimento de terapias gênicas para infecções por HPV. Outra abordagem seria a busca de compostos antivirais que também podem ser úteis no tratamento da doença pré-existente. Um inibidor efetivo de HPV reduziria a incidência de displasia cervical e anal, e carcinomas associados que afligem milhões de mulheres e homens em todo mundo (BALEJA et al., 2006). Em relação a E7, estudos demonstraram que o domínio CR3 desta proteína é essencial para o ciclo de vida do HPV (LONGWORTH; LAIMINS, 2004) e este pode ser um excelente candidato para inativação por pequenos compostos moleculares (LIU et al., 2006). Para estudos estruturais com a E7, existe a disposição uma estrutura experimental incompleta, apenas o domínio C-terminal, obtida por ressonância magnética nuclear (RMN) na forma de monômero e dímero depositadas no Banco de Dados de Proteínas (PDB) (BERMAN et al., 2000), códigos PDB ID: 2EWL e 2F8B respectivamente (OHLENSCHLÄGER et al., 2006).

Devido à escassez de dados na literatura em relação à estrutura das proteínas E7 e também a ausência destas estruturas totalmente resolvidas, este projeto visou, em uma primeira etapa, gerar modelos estruturais completos e próximos ao nativo utilizando

ferramentas de modelagem molecular (métodos *ab initio* e por homologia). A seguir, já com as estruturas resolvidas E7, entender as diferenças estruturais entre as proteínas de baixo e alto risco para o câncer, e por fim, realizar uma busca por inibidores moleculares por meio do método de “docking” macromolecular para a E7 de alto risco. O conhecimento sobre a estrutura da E7 pode ser utilizada em estudos de interações proteína-proteína e proteína-ligante gerando novas possibilidades para o tratamento da infecção por HPV e/ou câncer cervical.

1.3 Desordem Intrínseca

Recentemente, muitas proteínas ou regiões específicas de proteínas vêm sendo identificadas por apresentarem ausência de uma estrutura tridimensional bem definida sob condições naturais e fisiológicas, contrastando com o paradigma de estruturas bem definidas encontradas nestas condições (TOMPA, 2002, 2005). Estas proteínas são chamadas de proteínas intrinsecamente desordenadas (IDPs) ou intrinsecamente desestruturadas (IUPs) e podem apresentar espaços conformacionais muito diferentes, seja estados de desordem completa (“coils” aleatórios) ou estados de compactação (“molten globule” – desnaturação parcial) com contatos secundários ou terciários transientes (UVERSKY; GILLESPIE; FINK, 2000; UVERSKY, 2002).

Dois atributos são vistos como a principal causa da desordem intrínseca: um baixo valor de hidrofobicidade, que impede a formação de um núcleo globular estável; e uma carga líquida (função das somas de cargas positivas e negativas) que favorecem os estados estendidos da proteína devido a repulsões eletrostáticas (TOMPA, 2009). Estas condições ocorrem devido a disposição de aminoácidos na proteína desordenada. Alguns aminoácidos são promotores de ordem (W, C, F, I, Y, V, e L), pois possuem baixo índice de flexibilidade devido a sua característica hidrofóbica; já alguns são promotores de desordem (A, R, G, Q, S, P, E, e K), pois, possuem altos índices de flexibilidade devido a suas características polares e carregadas (DUNKER et al., 2001).

A proteína E7 do HPV apresenta características de IDPs em seu domínio N-terminal associadas, principalmente, aos HPVs de alto risco. Resultados obtidos por Uversky *et al.* (2006) em análises *in silico*, usando preditores de detecção de desordem intrínseca para a E6 e

E7 do HPV, identificaram que o nível de desordem destas proteínas aumenta conforme o nível de risco para o câncer cervical. Assim como a E7, outras IDPs têm sido associadas a doenças humanas, incluindo câncer, doenças cardiovasculares, amiloidoses, doenças neuro-degenerativas e diabetes. A associação de IDPs com doenças sugere que estas podem surgir não apenas por mal dobramento de proteínas, como erros em identificação e sinalização de alvos, mas também com o surgimento de dobramentos não nativos ou não naturais (UVERSKY; OLDFIELD; DUNKER, 2008).

1.4 Análises de Bioinformática Estrutural

A bioinformática estrutural é uma sub-disciplina da bioinformática que foca na representação, estocagem, recuperação, análise e exibição de informação estrutural em escala atômica e sub-celular (ALTMAN; DUGAN, 2009). Um dos objetivos principais da bioinformática estrutural é o de modelar, analisar e identificar funções e entender como as estruturas de DNA, RNA e proteínas se relacionam. Hoje, diversas técnicas, algoritmos e programas estão disponíveis e muitos vêm sendo desenvolvidos, tendo como alvo a geração de modelos proteicos tridimensionais, assim como sua validação, a interação com outras moléculas e simulações de dinâmica molecular (DM) em condições fisiológicas.

1.4.1 Modelagem Macromolecular

Um modelo tridimensional de proteína pode ser obtido de forma experimental ou teórica. Dentre os métodos da metodologia experimental destacam-se a Cristalografia de Raios-X, a Ressonância Magnética Nuclear e a Microscopia Eletrônica. No entanto, limitações técnicas e a própria característica da proteína, muitas vezes, impossibilitam a determinação de sua estrutura. Adicionalmente, a produção de cristais adequados para a resolução de estrutura tridimensional de proteínas é de um para vinte (MAGGIO; RAMNARAYAN, 2001). Uma alternativa às técnicas experimentais seria a utilização de metodologias de modelagem teóricas ou *in silico*. Atualmente, existem três abordagens possíveis para a modelagem macromolecular *in silico*: a modelagem por *ab initio*, “threading”

e modelagem por homologia ou comparativa. Tais abordagens teóricas fazem uso de ferramentas da Bioinformática Estrutural que é uma área da bioinformática que estuda estruturas biológicas, dentre elas, as proteínas.

1.4.2 Modelagem por Homologia

Modelagem por homologia utiliza estruturas de proteínas determinadas experimentalmente (moldes) para prever a conformação de outra proteína que tem uma sequência de aminoácidos semelhante (alvo) (JOHNSON et al., 1994). Este procedimento de múltiplos passos envolve: a seleção de uma estrutura molde, alinhamento entre as sequências da estrutura molde e da proteína alvo, a modelagem, refinamento e avaliação dos modelos gerados (KOSINSKI et al., 2009). Esta abordagem só é possível porque pequenas mudanças entre as sequências das proteínas molde e alvo normalmente resulta também em pequenas mudanças nas suas estruturas 3D (LESK; CHOTHIA, 1986). Para a seleção da estrutura molde, devem ser realizadas buscas em bancos de dados por sequências similares. Com o molde selecionado é realizado o alinhamento das sequências. O alinhamento é feito utilizando-se a sequência da proteína alvo e a sequência da estrutura molde. Em seguida é realizada a modelagem propriamente dita e a sua avaliação com a utilização de ferramentas bioinformática estrutural. O modelo final deve ser avaliado para certificar-se de que as características estruturais deste são consistentes com os parâmetros físico-químicos esperados. Como a modelagem comparativa se baseia na semelhança entre sequências alvo com sequências de estruturas conhecidas, é desejável que se tenha similaridade acima de trinta por cento de identidade sequencial entre o molde e o alvo (SALI, 1998).

1.4.3 Modelagem por *ab initio*

Predição de estruturas de proteínas a partir de sequências de aminoácidos tem sido um dos desafios mais difíceis na bioinformática estrutural há muitos anos (BAKER; SALI, 2001). Diante da inexistência de uma estrutura homóloga, a proteína alvo deve ser construída sem informação da estrutura *a priori*. Este procedimento é denominado de várias formas,

modelagem *ab initio* (KLEPEIS et al., 2005), modelagem *de novo* (BRADLEY; MISURA; BAKER, 2005), modelagem baseada na Física (OLDZIEJ et al., 2005) ou modelagem livre de modelo (JAUCH et al., 2007). Tipicamente, a modelagem *ab initio* conduz a uma busca conformacional sob a orientação de uma função de energia gerando múltiplas conformações possíveis de onde os modelos finais são selecionados. No entanto, a modelagem por *ab initio* depende de três fatores: uma função de energia precisa com a qual a estrutura nativa de uma proteína corresponde ao estado mais estável termodinamicamente, comparados com todas as estruturas geradas; um eficiente método de busca que possa, rapidamente, identificar estados de baixa energia através de busca conformacional; e a seleção de modelos características nativas dentre as várias estruturas geradas (LEE; WU; ZHANG, 2009).

As funções de energia são classificadas em dois grupos: baseadas na Física e baseadas em conhecimento. No primeiro grupo, a modelagem *ab initio* é realizada com a ajuda de campos de força que simulam um número selecionado de tipos de átomos os quais possuem propriedades físicas similares com os parâmetros calculados para a formação de cristais ou para a teoria da mecânica quântica (HAGLER; HULER; LIFSON, 1974; WEINER et al., 1984). Estes potenciais contêm termos associados com comprimentos de ligação, ângulos, ângulos de torção, interações de van der Waals e eletrostáticas. O segundo grupo refere-se às condições de energia empíricas derivadas de estatísticas de estruturas resolvidas depositadas no banco de dados de proteína (PDB). Porém, o sucesso de modelagem *ab initio* depende da disponibilidade de um método de busca conformacional poderoso que possa eficientemente encontrar a estrutura de energia mínima global para uma dada função de energia. Historicamente, o mais popular método de busca conformacional é o de Monte Carlo (BINDER; HEERMAN, 1992), que tem sido usado para duas finalidades: estimar valores termodinâmicos sobre o espaço conformacional da proteína (HANSMANN, 1997) e na busca conformações de baixa energia, incluindo a sua estrutura nativa (ABAGYAN; TOTROV, 1994).

Por fim, torna-se necessário o uso de ferramentas que selecionem, dentre os modelos proteicos gerados, aqueles estruturalmente próximos ao estado nativo. Para a validação e seleção destes modelos tornou-se necessário a utilização de programas de validação, os chamados Programas de Avaliação de Qualidade de Modelos ou MQAPs (do inglês “Model Quality Assessment Programs”). Os MQAPs baseiam-se, geralmente, numa energia física efetiva por análises fundamentais de força de partículas ou energias pseudo-empíricas de estruturas de proteínas conhecidas (LAZARIDIS; KARPLUS, 2000).

1.4.4 Dinâmica Molecular

A dinâmica molecular representa a abordagem computacional para a mecânica estatística e é usada para estimar propriedades de equilíbrio e dinâmica de sistemas complexos que não podem ser calculados analiticamente (SHLICK, 2010). Entender os movimentos das proteínas tem sido um importante campo de pesquisa, haja visto que a dinâmica esta inexoravelmente ligada a sua função (SAMISH; GU; KLEIN, 2009). As simulações em dinâmica molecular, plausíveis para os atuais padrões de processamento computacional, representam escalas de tempo muito pequenas no universo macromolecular. Segundo Becker (2001), entre pico (ps) e femtosegundos (fs) apenas movimentos de átomos nas cadeias laterais, vibrações de ligações e flexibilidade de ligantes são percebidos, já de nano (ns) a microssegundos (μ s) é possível a visualização e análise de mudanças em conformações de sítios ativos, movimentos de dobramento, rotações de ligações de peptídeos, rotação de ligações e formação de “loops”. As simulações de dinâmica molecular de macromoléculas biológicas vêm gerando muita informação relativa a movimentos internos destes sistemas desde que a primeira proteína foi estudada (MCCAMMON; GELIN; KARPLUS, 1977). Devido aos contínuos avanços nas metodologias e na velocidade dos computadores, os estudos em DM estão sendo estendidos a sistemas e mudanças conformacionais cada vez maiores, e maiores escalas de tempo de simulação (KARPLUS; MCCAMMON, 2002).

1.4.5 Farmacóforos

Os farmacóforos são, por sua definição mais atual, um conjunto de características estéricas e eletrônicas, que são necessárias para garantir interações supramoleculares ótimas com um alvo biológico específico, ativando ou bloqueando sua resposta biológica (WERMUTH et al., 1998). O conceito central farmacóforo envolve a noção de que o reconhecimento molecular de um alvo biológico por um grupo de compostos pode ser atribuída a um pequeno conjunto de características comuns que interagem com regiões complementares no alvo biológico, sendo estas características bastante gerais: doadores de ponte de hidrogênio, aceptores de pontes de hidrogênio, grupos carregados positiva e negativamente e regiões hidrofóbicas (LEACH et al., 2010).

Existem basicamente dois métodos de modelagem de farmacóforos, o baseado na estrutura e o baseado no ligante. O baseado na estrutura usa as informações sobre as interações proteína-ligante ou apenas da proteína, obtidas a partir de uma estrutura 3D derivada experimentalmente, para gerar um modelo. Já o método baseado no ligante extrai do modelo informações apenas do ligante, seja na obtenção de características químicas comuns em conjuntos de compostos ativos ou apenas o uso de um único composto que serve como um molde para a elucidação do farmacóforo (MARKT; SCHUSTER; LANGER, 2011).

A modelagem de farmacóforos é uma técnica *in silico* bem estabelecida que gera muitos benefícios para pesquisas iniciais com desenhos de fármacos. Sua vantagens estão na visualização das características químicas supostamente importantes na interação proteína-ligante assim como a obtenção de compostos bioisotéricos estruturalmente diversos que não teriam sido descobertos por derivatização sistemática de compostos existentes (MARKT; SCHUSTER; LANGER, 2011). A crescente gama de aplicações dos farmacóforos, juntamente com casos de sucesso na descoberta de fármacos, vem permitindo o enriquecimento do conceito de farmacóforo promovendo o desenvolvimento e aplicação de novas abordagens (YANG, 2010).

1.4.6 “Docking” Macromolecular

O “docking” macromolecular envolve a busca de modos de interação entre duas moléculas, como por exemplo interações proteína-proteína, proteína-ligante ou proteína-DNA. O sistema de “docking” macromolecular proteína-ligante consiste de duas etapas: um algoritmo de busca conformacional que envolva todos os graus de liberdade do ligante (translacional, rotacional e conformacional) e uma função de pontuação que ranquee as prováveis posições espaciais para um composto ou vários provenientes de uma grande biblioteca (ROGNAN, 2011). O “docking” proteína-ligante foi uma das metodologias pioneiras, nos anos 80, nas análises em bioinformática (KUNTZ et al., 1982) e quando apenas a estrutura de um alvo e seu sítio de ligação estão disponíveis, o “docking” de alta performance é a primeira escolha como ferramenta de identificação de compostos candidatos a fármacos. Apesar do “docking” e das funções de pontuação basearem-se em aproximações, a aplicação desta técnicas durante a otimização de compostos-líder, em conjunto com outros

métodos computacionais, soma-se a abordagens tradicionais no desenho de fármacos baseado na estrutura (KITCHEN et al., 2004). Atualmente, a técnica dominante para a identificação de novos compostos líderes na descoberta de fármacos é a prospecção física de grandes bibliotecas de substâncias contra um alvo biológico. Uma abordagem alternativa, conhecida como “virtual screening” visa rastrear computacionalmente, grandes bibliotecas de compostos que tenham como alvo uma estrutura conhecida, e testar, experimentalmente, as que apresentaram interações boas (SHOICHET, 2004).

1.5 Motivação

A expressão da proteína E7 dos HPVs de alto risco é, junto com a proteína E6, um dos principais fatores de surgimento de lesões de alto grau e câncer cervical. O tratamento seja das lesões ou seja do câncer, parece ser ainda um dos grandes desafios contra infecções por HPV. Hoje a prevenção é um dos principais alvos nesta luta e a estratégia em destaque na atualidade é a utilização de vacinas com eficácia comprovada para alguns tipos do vírus de alto risco. Porém, as vacinas ainda são caras e não chegam para a população mais humilde, principalmente, em países pobres e em desenvolvimento. As perguntas que surgem com isso são: o que fazer com as milhares de mulheres já infectadas com o vírus e sem acesso a vacina? Como tratar de maneira eficaz lesões de alto grau e o câncer que mais mata mulheres no mundo? Este é um desafio abordado por várias frentes de pesquisa, seja com estudos sobre a integração do DNA viral no núcleo das células, na busca de RNAs de interferência contra E6 e E7 ou na busca de pequenos ligantes para estas proteínas. Esta última abordagem é sugerida por vários autores como uma das possíveis estratégias contra o HPV. Porém, pouco se produziu a respeito de aumentar o entendimento da estrutura da E6 ou E7 com o intuito de buscar sítios ativos que possam ser alvos de pequenos ligantes.

O desafio deste projeto foi: aumentar a compreensão acerca da estrutura da E7, utilizando como base, os dados da literatura disponíveis, assim como o uso das ferramentas mais modernas de bioinformática estrutural. O foco foi a produção de novas ideias acerca da estrutura, da função e da interação da E7 de alto risco com pequenos ligantes a fim de contribuir com estudos de desenho de fármacos contra o HPV e o câncer cervical.

2 HIPÓTESES

Este trabalho foi norteado por quatro hipóteses, acerca da estrutura da E7 e suas interações, que seguem:

- o modelo teórico tridimensional completo da estrutura da E7 pode ser obtido e validado, com o uso de ferramentas *in silico*, apesar da presença de regiões com desordem intrínseca;
- o nível de desordem intrínseca no domínio N-terminal da E7 tem relação direta com o tipo de HPV e o risco atribuído a este para o surgimento do câncer cervical;
- a região CR3 da E7 possui as características estruturais adequadas para a busca e definição de sítios de ligação;
- a busca de pequenos ligantes, *in silico*, para a E7 pode ser realizada a partir um sítio de ligação específico na região CR3.

3 OBJETIVOS

Os objetivos primários do projeto foram:

- obter as estruturas completas das proteínas E7 oriundas dos HPVs tipo 1A, 11, 16, 18;
- obter, por meio de dinâmica molecular, dados sobre a flexibilidade e desordem intrínseca atribuída a E7;
- Buscar por modelos farmacofóricos e “docking” macromolecular de compostos que interajam satisfatoriamente com a E7 de alto-risco para o câncer.

3.1 Objetivos Específicos

Predizer a estrutura macromolecular completa das proteínas E7 dos tipos de HPV 16 e 18 (alto risco) e HPV 01, 11 (baixo risco), ainda não resolvidas, por meio de um protocolo misto de modelagem por homologia e *ab initio*.

Validar os modelos selecionados por meio da análise de parâmetros esteroquímicos, de contatos e interações atômicas e ambientes químicos.

Entender as diferenças estruturais entre os vários tipos de HPV por meio de simulações de Dinâmica Molecular.

Selecionar regiões ou sítios de ligação específicos para a estrutura da oncoproteína E7 do HPV de alto risco.

Buscar regiões propensas a interações não covalentes através de análises de “hot spots” dos sítios de ligação.

Desenvolver modelos farmacofóricos capazes de auxiliarem na prospecção e filtragem de grandes bancos de dados de compostos.

Propor um modo de ligação de pequenos compostos à proteína E7 de alto risco utilizando a abordagem de “docking” flexível por meio de simulações de “screening” em bases de dados virtuais de pequenas moléculas, filtradas por farmacóforos.

Determinar os potenciais de interação molecular receptor-ligante para o resultado do “screening” virtual.

Determinar *in silico* os potenciais efeitos nocivos, ao ser humano, dos compostos selecionados.

4 MATERIAIS E MÉTODOS

A metodologia, desenvolvida neste trabalho, está resumida no fluxograma abaixo (Figura 1).

Figura 1 - Fluxograma da metodologia resumida desenvolvida ao longo do projeto.

4.1 Análise *in silico* de desordem intrínseca

A análise *in silico* de desordem intrínseca da proteína E7 dos HPV's tipos 11 (baixo risco), 16 e 18 (de alto risco) e 1A (não relacionado com a infecção cervical e câncer) foi realizada utilizando os programas PONDR-FIT (XUE et al., 2010) e Discon (MIZIANTY et al., 2011), que detectam, respectivamente, a contribuição de cada resíduo à desordem e o conteúdo de desordem de cada proteína. O programa ANCHOR (DOSZTÁNYI; MÉSZÁROS; SIMON, 2009) identifica segmentos que residem em regiões desordenadas que obtêm a estabilização de energia através da interação com proteínas globulares. Ele foi utilizado a fim de prever sítios de ligação na porção intrinsecamente desordenada da proteína E7.

4.2 Modelagem Macromolecular

4.2.1 Seleção de Molde e Modelagem por Homologia da E7 do HPV tipo 16

A sequência de aminoácidos da E7 do HPV tipo 16 (P03129) foi recuperada a partir do banco de dados de UniProt (<http://www.uniprot.org>). A proteína foi modelada, parcialmente, por homologia usando a estrutura de solução de RMN do domínio C-terminal (monômero) da E7 do HPV tipo 45 (PDB ID: 2EWL) extraído do banco de dados de proteína (PDB). A região C-terminal da proteína molde apresentou 65% de similaridade com E7 de HPV tipo 16, sendo viável a realização da modelagem por homologia. A modelagem de homologia foi realizada utilizando o programa MODELLER (SALI; BLUNDELL, 1993) implementado no servidor @TOME2 (<http://atome.cbs.cnrs.fr>) (PONS; LABESSE, 2009).

4.2.2 Modelagem *ab initio*

A região N-terminal do modelo da E7 do HPV tipo 16 possui características de proteínas intrinsecamente desordenadas e por isso não foi modelada pela metodologia de modelagem por homologia. Foi utilizada a técnica de modelagem por *ab initio* para a

obtenção da porção N-terminal da molécula e, assim, gerar um modelo tridimensional completo da E7 do HPV tipo 16. A modelagem por *ab initio* é usada quando há pouca informação inicial sobre a estrutura da molécula. O programa Rosetta 3.1 (ROHL et al., 2004), mais especificamente, o protocolo de modelagem usando o algoritmo “cyclic coordinate descent” (CCD) (CANUTESCU; DUNBRACK, 2003) foi utilizado para esta finalidade. Cerca de 1000 modelos foram gerados, um dos quais foi escolhido após avaliação e validação usando os MQAPs e inspeção visual. A proteína modelo da E7 do HPV tipo 16 foi utilizada como molde para a modelagem pela técnica de homologia das proteínas E7 dos tipos de HPV 1A (P06465), 11 (P04020) e 18 (P06788).

4.2.3 Modelagem por Homologia da E7 do HPVs do Tipos 1A, 11 e 18

O Programa MODELLER (v9.10) foi utilizado para gerar modelos por homologia da proteína E7 dos HPVs dos tipos 1A, 11 e 18 usando como molde a estrutura completa da E7 do HPV do tipo 16 previamente gerada. Vinte modelos foram gerados para cada proteína e avaliados utilizando o programa PROCHECK (LASKOWSKI et al., 1993).

4.2.4 Modelagem de Rotâmetros e Inserção das Coordenadas do Zinco

Com o intuito de se obter estruturas de alta qualidade foi realizada uma modelagem de rotâmetros das cadeias laterais de todos os modelos previamente gerados utilizando SCWRL4 (KRIVOV; SHAPOVALOV; DUNBRACK, 2009). As coordenadas do átomo de zinco do dedo de zinco da estrutura 2EWL (PDB) foram manualmente transferidas aos modelos homólogos usando o programa DS Visualizer 3.5 (Discovery Studio Visualizer, 2005).

4.2.5 Validação dos Modelos

A validação das estruturas proteicas e, mais particularmente, a escolha dos modelos próximo aos nativos exigem não só uma análise visual da estrutura, mas análises locais (cada resíduo de aminoácido) e validação global (proteína como um todo) da estrutura 3-dimensional da proteína. Vários MQAPs foram usados para selecionar os melhores modelos da proteína E7. No primeiro passo, as funções de pontuação ModFOLD (MCGUFFIN, 2008), dDFIRE (YANG; ZHOU, 2008) e QMEAN (BENKERT; TOSATTO; SCHOMBURG, 2008) foram utilizadas para classificar os 1.000 modelos da E7 do HPV tipo 16. O ModFOLD faz comparações de alinhamentos das estruturas produzindo previsões globais e locais acerca da qualidade dos modelos. O dDFIRE é uma função de pontuação que baseia-se nos ângulos de orientação envolvidos em interações dipolo-dipolo. O QMEAN é uma função de pontuação, composta por 6 termos, que é capaz de derivar estimativas do erro estrutural global (para toda a estrutura) e local (estimativas por resíduo) para a proteína estudada.

Desta forma, as 10 estruturas mais bem pontuadas para cada MQAP continuaram na análise. Não havia nenhuma redundância entre os 10 modelos selecionados por cada programa, assim, as 30 estruturas resultantes foram classificados de novo com base nos resultados do PROCHECK. A fim de escolher as estruturas de melhor qualidade, os 10 melhores modelos restantes desta última análise foram então submetidos a vários MQAPs: Preditor de Qualidade Proteica (ProQ) (WALNNER; ELOFSSON, 2003), ERRAT(COLOVOS, 1993), Verify 3D (BOWIE; LUTHY; EISENBERG, 1991; LUETHY; BOWIE; EISENBERG, 1992) e Avaliação do Volume Proteico (PROVE) (PONTIUS; RICHELLE; WODAK, 1996).

O ProQ é um método baseado em redes neurais para prever a qualidade do modelo proteico. Ele extrai características estruturais, tal como a frequência de contato átomo-átomo, e usa dois tipos de medidas de qualidade chamadas de LGscore (CRISTOBAL et al., 2001) e MaxSub (SIEW et al., 2000). O ERRAT é um método que diferencia entre determinadas regiões corretas e incorretas na estrutura proteica. Ele se baseia na análise de características das interações atômicas comparadas à resultados de estruturas altamente refinadas. Já o Verify 3D determina a compatibilidade do modelo atômico tridimensional com sua própria sequência de aminoácidos, atribuindo uma classe estrutural, para cada resíduo, baseada na localização e ambiente, e comparando os resultados com estruturas de boa qualidade. Por fim, o PROVE

calcula o volume dos átomos em macromoléculas usando um algoritmo que trata os átomos como esferas e calcula o desvio estatístico, Z-escore, do modelo em relação a estruturas altamente resolvidas (2,0 Å, ou menor) e refinadas (fator-R de 0,2 ou menor) depositadas no banco de dados do PDB.

4.3 Simulações de Dinâmica Molecular

A dinâmica foi realizada usando o campo de força CHARMM27 (MACKERELL et al., 1998) implementado no programa NAMD 2.8 (PHILLIPS et al., 2005). A simulação de MD foi realizada por 50 ns para cada proteína (E7, tipos 1A,11,16,18) somando um total de 0,20 μ s, utilizando temperatura constante a 310 K. Cada 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. As cargas do sistema foram balanceadas acrescentando-se íons de cloro e sódio. O método “Particle Mesh Ewald” (PMEM) foi usado para tratar a eletrostática de longa distância. Todas as ligações envolvendo hidrogênio foram fixadas pelo algoritmo SHAKE (VAN GUNSTEREN; BERENDSEN, 1977). O número de átomos para cada sistema foi de 17.551, 17.393, 17.331, 20.034 para os HPVs tipo 1A, 11, 16, 18, respectivamente. Antes da dinâmica, todo sistema foi minimizado por 5000 passos para reduzir interações desfavoráveis entre a proteína e solvente. Cada sistema foi lentamente aquecido até 310 K nos primeiros 52 ps em condições normais de temperatura e pressão (NPT). Cada sistema foi simulado com temperatura (310 K) e pressão constante (1atm). Os arquivos de topologia e parâmetros do CHARMM27 foram modificados com a inserção do resíduo cisteína apresentando a forma tiolato (no qual um átomo de metal substitui o hidrogênio ligado ao enxofre). Isso foi necessário para que houvesse interação entre o átomo de enxofre das cisteínas do dedo de zinco com o átomo do zinco (FOLOPPE et al., 2001).

4.4 Análises das Trajetórias da Dinâmica Molecular

As trajetórias obtidas na dinâmica molecular foram analisadas com o auxílio dos programas VMD (HUMPHREY; DALKE; SCHULTEN, 1996), eucb (TSOULOS; STAVRAKOUDIS, 2011) e carma (GLYKOS, 2006). O programa carma e VMD foram

utilizados para transformar os dados da trajetória completa (solvente, Cl, Na, Zn e proteína) em trajetórias que continham apenas as proteínas, para estas servirem de entrada para o programa eucb. Dois protocolos do eucb foram utilizados com o intuito de obter das trajetórias: o desvio da média quadrática (RMSD) dos átomos C α e a flutuação da média quadrática dos átomos C α usando a estrutura inicial como referência (RMSF). As variações na estrutura secundária ao longo da trajetória foram analisadas usando o algoritmo STRIDE (ANDERSEN et al., 2002) implementado no “plug-in” TIMELINE do VMD.

4.5 Modelagem de Farmacóforos

4.5.1 Busca de sítios de ligação

Para a modelagem da E7 do tipo 16 foi utilizado como molde à estrutura de solução de RMN do domínio C-terminal (monômero) da E7 do HPV tipo 45 (PDB ID: 2EWL). A mesma estrutura na forma de dímero, também obtida de solução de RMN do domínio C-terminal da E7 do HPV tipo 45 (PDB ID: 2F8B) foi utilizada nas etapas de busca de sítio de ligação, modelagem farmacofórica e “docking” macromolecular. A CR3 da proteína E7 do HPV tipo 45 (na forma de dímero, 2F8B) foi obtida e estudada por Ohlenschläger et al. (2006) quanto a suas prováveis regiões de ligação proteína-proteína, por meio de análises de modificações causadas por perturbações químicas. Estes resultados indicaram prováveis aminoácidos com potencial de interação que serviram de base para o desenvolvimento desta etapa do projeto. A opção pelo uso da E7 do HPV tipo 45 (2F8B) foi realizada devido a esta: possuir caráter experimental, possuir resultados experimentais relacionados a possíveis regiões de interação, ser de alto risco para o câncer cervical e apresentar alta similaridade, 50%, 57,4%, 59,3%, 89,6%, com as outras E7s estudadas: 1A, 11, 16, 18 respectivamente. A proteína 2F8B foi então protonada e a busca pelo sítio de ligação foi realizado utilizando um método automatizado chamado de SiteFinder, que usa a metodologia de esferas “alpha shapes” (EDELSBRUNNER et al., 1995), implementado no programa “Molecular Operating Environment” (MOE, 2010). Além disso, a inspeção visual e os resultados experimentais obtidos por Ohlenschläger et al. (2006) e Liu et al. (2006), ajudaram na escolha da região alvo para a busca de modelos farmacofóricos.

4.5.2 Análise de “hot spots” e Modelagem Farmacofórica

A análise de “hot spots” ou de campos de interação foi realizada a fim de encontrar as regiões ao redor do sítio de ligação que apresentavam interações favoráveis entre uma sonda química e a superfície da molécula. Esta análise foi realizada com a ajuda de dois programas, o HotspotsX 0.61 (NEUDERT; KLEBE, 2011) e o método “Interaction Potentials” implementado no MOE. Os seus resultados foram analisados por inspeção visual. Os mapas de interação (“grids”) criados foram utilizados como base na construção dos modelos farmacofóricos. Foram gerados 15 modelos farmacofóricos com três pontos de referência cada um, também chamados de pontos de anotação que foram criados com base na informação das sondas químicas que interagem com o sítio escolhido. Para a visualização dos “hot spots” foram utilizados os programas Pymol (The PyMOL Molecular Graphics System, 2010) e o MOE.

4.5.3 Busca Farmacofórica

Os modelos farmacofóricos gerados foram avaliados quanto a sua capacidade de encontrar ligantes que preenchessem os requisitos exigidos pelos três pontos de anotação, definidos para cada modelo. Além disso, foram realizadas inspeções visuais a fim de comprovar se as posições dos ligantes, aceitos para cada modelo, apresentavam encaixe adequado ao sítio proposto. Foram utilizadas duas bibliotecas de confôrmeros, que apresentam mais de uma conformação para a mesma molécula, para a busca de ligantes adequados aos modelos propostos: a biblioteca “druglike” ZINC 2007 (1.778.466 moléculas) (IRWIN; SHOICHET, 2005) e a “lead-like” MOE 2011 (653.214 moléculas). Para tanto, foi utilizado uma busca farmacofórica pelo programa MOE que busca em bibliotecas de compostos, conformações de estruturas 3D que correspondem as restrições impostas para um farmacóforo específico. As bibliotecas resultantes da busca farmacofórica foram então preparadas com a adição de hidrogênios e a deprotonação de ácidos fortes e a protonação de bases fortes, utilizando a opção “wash” no MOE.

4.6 “Docking” Macromolecular

As bibliotecas de compostos geradas a partir da busca pelos modelos farmacofóricos escolhidos foram então submetidas a técnica de “docking” macromolecular. Para tanto foi utilizado o programa GOLD (JONES; WILLETT; GLEN, 1995; JONES et al., 1997) que calcula o “docking” flexível entre as proteínas e os ligantes. Para tanto, ele utiliza parâmetros específicos para a busca ou “screening” virtual de compostos nas bases de dados selecionadas. O programa GOLD usa um algoritmo genético, que busca soluções por meio de “docking” propagando múltiplas cópias de modelos flexíveis do ligante no sítio ativo do receptor, recombinaando segmentos desta cópia aleatoriamente até um conjunto convergido de estruturas ser gerado. O sítio de ligação escolhido foi o mesmo utilizado na construção dos modelos farmacofóricos, tendo como átomo central o SG da cisteína 14 e um raio de 7Å ao redor deste foi definido para restringir o espaço amostral para o “docking”. A função de pontuação utilizada foi o GoldScore, que é uma função definida por quatro componentes; a energia da ponte de hidrogênio entre a proteína-ligante; a energia de van der Waals entre a proteína e o ligante; a energia de van der Waals interna do ligante e a energia de deformação de torção do ligante. O GOLD utiliza, para a busca das melhores soluções de “docking”, o algoritmo genético, definido neste experimento pela opção “Very Flexible”, que corresponde a uma eficiência de busca de 200%, enquanto os outros parâmetros foram utilizados nas opções pré-definidas pelo programa. Todos os ligantes selecionados para o “docking” foram submetidos a 10 interações cada, pelo algoritmo genético, porém apenas os três melhores compostos, baseados no GoldScore, foram selecionados.

4.6.1 “Re-docking” e Avaliação

Após o primeiro “docking”, a biblioteca, das soluções resultantes no número de 3 para cada composto, foi novamente submetida à busca farmacofórica. Esta etapa foi realizada, a fim de selecionar apenas os compostos que após o “docking” apresentavam coerência, no seu posicionamento tridimensional com os farmacóforos propostos. Desta forma, foi gerada uma biblioteca com um representante de cada composto, ou seja, aqueles com melhor GoldScore, que passaram pela busca farmacofórica. Essa nova biblioteca foi submetida ao “docking” com o GOLD utilizando os mesmos parâmetros anteriores. As representações posicionais da

interação ligante-proteína foram analisadas visualmente após o “docking” no intuito de selecionar apenas os ligantes que convergiram para uma posição similar. Desta forma foram selecionados apenas aqueles ligantes cujas posições tridimensionais para as 10 corridas do algoritmo genético eram próximas espacialmente (análise visual). Além disso, as pontuações deveriam convergir para um valor similar entre as 10 interações de cada composto. O ligantes que convergiram foram extraídos da análise e formaram então uma nova biblioteca que foi novamente testada contra os modelos farmacofóricos.

Os compostos que passaram na busca farmacofórica para cada farmacóforo estudado foram reunidos e a biblioteca resultante foi submetida novamente ao “docking” (“re-docking”) com o GOLD utilizando os mesmos parâmetros anteriores, porém foi adicionada uma nova função de pontuação a CHEMPLP. A CHEMPLP é a função padrão do GOLD e usa o chamado “piece-wise linear potential” (PLP) para interações hidrofóbicas e não complementares, além de termos da função de pontuação Chemscore para pontes de hidrogênio e energia interna. Além das funções de pontuação utilizada pelo GOLD, foram utilizadas também a função de pontuação do programa DSX 0.89 (NEUDERT; KLEBE, 2011) com o intuito de avaliar as interações ligante-proteína geradas no “re-docking”. O programa DSX contém uma função de pontuação baseada em conhecimento que é formada de potencial de distância dependentes dos pares, potencial de ângulos de torção e potencial de acessibilidade de solvente a superfície. O programa DS Visualizer 3.5 foi utilizado para gerar diagramas 2D da interação ligantes-E7, a fim de revelar quais tipos de interações foram resultantes do “docking”. Além da avaliação das funções de pontuação das interações, uma inspeção visual dos resultados foi realizada a fim de avaliar as representações posicionais geradas.

4.7 Análise de Toxicidade *in silico*

Por fim, os compostos selecionados foram analisados, *in silico*, quanto a sua toxicidade com o auxílio do programa DEREK (GREENE et al., 1999). O programa DEREK prediz se um dado químico é tóxico para humanos, mamíferos e bactérias. Esta etapa foi fundamental para comprovar ou não se os inibidores selecionados são potencialmente danosos quando administrados no organismo humano.

5 RESULTADOS E DISCUSSÃO

5.1 Avaliação da desordem intrínseca das E7s de HPVs de alto e baixo risco

As análises das sequências das E7s realizadas com o programa de DisCon (Tabela 1) mostram que as proteínas E7 do HPV dos tipos 16 e 18 apresentaram conteúdos aumentados de desordem, 28,57% e 10,47%, respectivamente, ao passo que um conteúdo muito baixo (~1,0%) para o tipo de baixo risco 11 e zero para o tipo 1A. De acordo com Uversky et al. (2006), a desordem intrínseca da proteína E7 aumenta com o risco de câncer de colo do útero. Portanto, as proteínas E7 dos tipos de baixo risco de HPV 06 e 11 são menos desordenadas, a proteína do HPV do tipo 16 é a mais desordenada, e a E7 do tipo 18 é intermediária. Curiosamente, as proteínas ligadas a outros tipos de câncer, tais como P53 (LEE et al., 2000) e BRCA1 (MARK et al., 2005), também têm propriedades de proteínas intrinsecamente desordenadas.

O PONDR-FIT foi utilizado para determinar quais os aminoácidos na sequência da proteína estão desordenados; pontuações superiores a 0,5 representam os aminoácidos com características intrinsecamente desordenados (Figura 2). Todas as proteínas E7 estudadas apresentaram alta desordem nos primeiros e últimos 10 aminoácidos. As proteínas E7 dos HPVs dos tipos 16 e 18 apresentaram maiores níveis de desordem, e a maior contribuição para esta desordem veio de seus domínios N-terminais, nas quais os resultados para quase todos os aminoácidos estavam acima do valor limite de 0,5. As pontuações PONDR-FIT para o HPV tipo 11 foram inferiores aos dos HPVs de alto risco, apesar da sua elevada similaridade de sequência (70%) com o tipo 16. Assim como os resultados obtidos usando DisCon, os valores PONDR-FIT para a desordem para o domínio N-terminal do HPV do tipo 1A foram abaixo do limiar. Como o HPV tipo 1A não está relacionado com a infecção do trato genital e câncer, este resultado indica que a presença de doença pode estar associada com os tipos de HPV alto risco. De acordo com García-Alai, Alonso e De Prat-Gay (2007), o domínio N-terminal da E7 do HPV tipo 16 pode ser descrito como intrinsecamente desordenado e tem um domínio plástico estrutural que provavelmente otimiza a velocidade de interações desta região e permite ligações específicas, mas pouco seletivas, a várias proteínas da célula hospedeira. As IDPs são classificadas com base nos seus modos de ação (TOMPA, 2002), e as proteínas E7 de HPV de alto risco podem ser classificados como IDPs efectoras

porque sua desordem modifica a atividade de um único alvo ou um conjunto de alvos moleculares (neste caso, a proteína pRb).

Tabela 1- Análise do conteúdo de desordem da proteína E7 dos tipos de HPV 1A, 11, 16, 18 usando DisCon.

Número de Acesso	Tipo de HPV	Conteúdo de Desordem (%)
P03129	16	28.57
P06788	18	10.47
P04020	11	1.02
P06465	1	0

Figura 2 - Análise da desordem intrínseca ao longo da cadeia de aminoácidos da proteína E7 dos HPVs dos tipos 1A, 11, 16, e 18 usando POND-R-FIT

5.2 Modelagem Molecular da proteína E7

O servidor @TOME2 guiou a primeira etapa da modelagem macromolecular por homologia. O @TOME2 realiza, de forma automatizada, a busca pela proteína molde (no banco de dados do PDB) que melhor se alinha com o alvo e gera modelos com ajuda do programa MODELLER. Neste processo, os modelos gerados apresentaram apenas a região CR3 da E7 do HPV tipo 16 modelada. Isto ocorreu, pois, a proteína molde que apresentou a maior similaridade com a E7 do HPV tipo 16 (65%) contém apenas o domínio C-terminal da E7 do monômero do HPV tipo 45 obtidos por RMN (2EWL). Desta forma, foram gerados 4 modelos e o melhor foi escolhido com base nos resultados do gráfico de Ramachandran, gerados pelo programa PROCHECK. O gráfico de Ramachandran indica: regiões favoráveis, adicionalmente permitidas, generosamente permitidas e não permitidas para os graus de torção dos ângulos Phi e Psi dos aminoácidos. Esta análise é realizada para todos os aminoácidos, exceto glicina e prolina que não apresentam torção destes ângulos devido a suas composições particulares de cadeia lateral. O mapa Ramachandran foi construído baseado em estruturas bem definidas presentes no banco de dados do PDB e serve de guia para escolha dentre um entre vários modelos gerados por homologia. Dos quatro modelos gerados, o modelo 3 foi o que obteve os melhores resultados no Ramachandran (Tabela 2 e Figura 3). A proteína modelo selecionada para a E7 do HPV tipo 16 apresentou um domínio modelado (C-terminal) e outro sem estrutura definida (N-terminal) (Figura 4).

Tabela 2 - Resultados da análise do gráfico de Ramachandran (PROCHECK) para os modelos incompletos, da E7 do HPV tipo 16, gerados por modelagem por homologia.

Modelo	Região favorável	Região adic. permitida	Região gener. permitida	não permitida
1	88,4%	10,5%	1,2%	0
2	80,2%	17,4%	2,3%	0
3	88,4%	11,6%	0	0
4	83,7%	14%	1,2%	1,2%

Estatística do Gráfico

Resíduos em regiões favoráveis [A,B,L]	76	88.4%
Resíduos em regiões adicionalmente permitidas [a,b,l,p]	10	11.6%
Resíduos em regiões generosamente permitidas [~a,~b,~l,~p]	0	0.0%
Resíduos em regiões não permitidas	0	0.0%
Número de resíduos não glicina e não prolina	86	100.0%

Figura 3 - Resultado do gráfico de Ramachandran para o modelo 3 da E7 do HPV tipo 16, gerado por modelagem por homologia. Os quadrados negros representam os aminoácidos comuns e os triângulos negros representam o aminoácido glicina.

Figura 4 - Modelo 3 da E7 do HPV tipo 16 gerado por modelagem por homologia parcial, apenas do domínio C-terminal (azul). A região em vermelho representa o domínio N-terminal da E7 que não apresentou homólogos no banco de dados do PDB.

O domínio N-terminal da E7 do HPV tipo 16, não modelado por modelagem por homologia, foi submetido à modelagem *ab initio* por meio do programa Rosetta. O domínio C-terminal, já modelado, da E7 foi mantido fixo ao longo das simulações *ab initio*. A abordagem de modelagem *ab initio* permitiu a geração de 1000 modelos teóricos completos da E7 do HPV tipo 16. Estes foram validados e o melhor foi selecionado, usando o resultado dos escores gerados pelos MQAPs (Tabela 3) e inspeção visual. Nas análises realizadas com PROCHECK foram utilizados apenas os gráficos de Ramachandran que, nos resultados apresentados pela Tabela 3, lê-se na sequência: resíduos em regiões favoráveis/ resíduos em regiões adicionalmente permitidas/ resíduos em regiões generosamente permitidas/ regiões não permitidas. O LGscore e o MaxSub são métricas que avaliam a qualidade do modelo e devem apresentar valores acima de 3 e 0,5, respectivamente, para considerar modelos proteicos como de boa qualidade. O programa ERRAT considera estruturas com boa qualidade apenas aquelas que apresentam valores iguais ou superiores a 95%. O Verify3D analisa a compatibilidade do modelo 3D com a sequência de aminoácidos e são considerados bons resultados aqueles que ficam acima de 80%. A média e o RMS do Z-score do programa

PROVE indicam como estruturas de boa qualidade pontuações próximas a 0 e 1,00, respectivamente.

A inspeção visual, também, foi essencial para selecionar entre estruturas com pontuações semelhantes para os MQAPs, como foi o caso dos modelos S_109 e S_139 (Tabela 3 e Figura 5). Embora os modelos S_109 e S_139 possuam boas pontuações nos MQAPs, a inspeção visual indica que o modelo S_139 é mais compacto e possui estruturas secundárias mais bem definidas no domínio N-terminal do que o modelo S_109.

Embora as proteínas E7 dos tipos de HPV de alto risco apresentem propriedades de IDPs, uma estrutura inicial da proteína E7 de HPV tipo 16 foi obtida com sucesso (Figura 5b) e utilizada como um molde para modelar as E7 de outros tipos (Figura 6). Estes modelos foram usados como estruturas iniciais para as simulações de dinâmica molecular. Os resultados indicam que a E7 do HPV tipo 16 é uma proteína alfa-beta que contém 5 hélices α , 3 das quais estão no domínio N-terminal, e 1 folha β antiparalela no domínio C-terminal (Figura 6c). O domínio C-terminal contém também duas sequências CXXC que juntas formam um motivo de dedo de zinco.

Tabela 3 - Seleção dos melhores modelos da E7 do HPV tipo 16 baseado em pontuações de MQAPs.

Modelo	PROCHECK (%)	LGscore	MaxSub	ERRAT (%)	Verify3D (%)	Z-Score (Média)	Z-Score (RMS)
S_0652	94,2 / 5,8 / 0/0 *	0,189	0,077	84,44	14,14	0,06*	1,33
S_0139	94,2 / 5,8 / 0/0 *	0,042	0,022	96,67 *	34,34	0,06*	1,3
S_0885	94,2 / 5,8 / 0/0 *	0,052	0,049	96,67 *	55,56	0,14	1,37
S_0911	93,0 / 7,0 / 0/0 *	0,034	0,057	89,65	41,41	0,14	1,45
S_0109	93,0 / 7,0 / 0/0 *	0,124	0,069	98,89 *	21,21	0,02*	1,29
S_0101	93,0 / 7,0 / 0/0 *	-0,288	-0,002	93,33	48,48	0,13	1,33
S_0382	93,0 / 7,0 / 0/0 *	0,471	0,11	94,44	77,78	-0,1	1,34
S_0994	91,9 / 8,1 / 0/0 *	0,41	0,048	82,22	22,22	0,14	1,33
S_0515	91,9 / 8,1 / 0/0 *	0,484	0,15	100	27,27	-0,11	1,11
S_0415	91,9 / 8,1 / 0/0 *	-0,061	-0,019	91,9	43,43	0,05 *	1,17

* Representa as pontuações que são esperadas para estruturas de boa qualidade.

Figura 5 - Modelos completos, S_109 (A) e S_139 (B), da E7 do HPV tipo 16 gerados pela modelagem por homologia da região C-terminal (azul) e modelagem por *ab initio* da região N-terminal.

A estrutura selecionada (modelo S_0139) foi usada como molde para a modelagem por homologia das proteínas E7 de HPV dos tipos 1A, 11, e 18, (Figura 6a, b, d) as quais apresentam 56,4%, 70%, e 61,7% de similaridade de sequência, respectivamente, com a proteína do HPV do tipo 16. A modelagem por homologia gerou 20 modelos para cada proteína, e os melhores modelos foram escolhidos usando o resultado do gráfico de Ramachandran do programa PROCHECK (Tabela 4). Além do gráfico de Ramachandran, foi usada em alguns casos de resultados semelhantes, outra análise do PROCHECK, o gráfico de ângulos de torção Chi1 e Chi2. Chi1 e Chi2 são ângulos de torção de cadeias laterais de alguns aminoácidos e foram usados, por exemplo, como critério para escolha do modelo 7 do HPV tipo 11 ao invés do modelo 20. O modelo 20 apresentou conformações desfavoráveis para os ângulos Chi1 e Chi2 para os aminoácidos histidina 24 e 51, e o modelo 7 apenas para a isoleucina 11.

Tabela 4 - Seleção dos modelos da E7 dos HPVs tipo 1A, 11 e 18 gerados por homologia com ajuda dos valores do gráfico de Ramachandran (PROCHECK).

HPV 1A modelo	PROCHECK (%)	HPV 11 modelo	PROCHECK (%)	HPV 18 modelo	PROCHECK (%)
1	91.6/6.0/1.2/1.2	1	91.8/8.2/0/0	1	86.3/9.5/4.2
2	92.8/6.0/0/1.2 *	2	91.8/8.2/0/0	2	89.5/6.3/3.2/1.1
3	88.0/7.2/1.2/3.6	3	96.5/3.5/0/0	3	87.4/8.4/2.1/2.1
4	86.7/9.6/1.2/2.4	4	91.8/7.1/1.2/0	4	85.3/12.6/2.1
5	91.6/7.2/0/1.2	5	92.9/5.9/1.2/0	5	88.4/7.4/2.1/2.1
6	89.2/9.6/0/1.2	6	95.3/4.7/0/0	6	89.5/7.4/1.1/2.1
7	92.8/4.8/1.2/1.2	7	96.5/3.5/0/0*	7	92.6/5.3/1.1/1.1
8	91.6/6.0/1.2/1.2	8	88.2/10.6/1.2/0	8	85.3/8.4/6.3/
9	89.2/9.6/0/1.2	9	90.6/8.2/1.2/0	9	90.5/4.2/2.1/3.2
10	88/9.6/1.2/1.2	10	92.9/7.1/0/0	10	88.4/7.4/4.2/0
11	89.2/7.2/1.2/2.4	11	88.2/10.6/1.2/0	11	86.3/10.5/2.1/1.1
12	91.6/7.2/0/1.2	12	89.4/9.4/1.2/0	12	89.5/4.2/3.2/3.2
13	92.8/4.8/0/2.4	13	90.6/8.2/1.2/0	13	89.5/8.4/2.1/0
14	92.8/3.6/2.4/1.2	14	91.8/7.1/1.2/0	14	92.6/4.2/3.2/0*
15	89.2/8.4/1.2/1.2	15	91.8/8.2/0/0	15	91.6/7.4/1.1/0
16	92.8/3.6/2.4/1.2	16	92.9/5.9/1.2/0	16	88.4/7.4/2.1/2.1
17	86.7/10.8/1.2/1.2	17	92.9/7.1/0/0	17	90.5/5.3/3.2/1.1
18	88/6/2.4/3.6	18	94.1/4.7/1.2/0	18	87.4/9.5/1.1/2.1
19	86.7/9.6/1.2/2.4	19	91.8/8.2/0/0	19	88.4/10.5/1.1/0
20	91.6/7.2/0/1.2	20	97.6/2.4/0/0	20	87.4/10.5/2.1/0

* indica os modelos selecionados

As proteínas E7 foram submetidas à modelagem de rotâmeros a fim de refinar as cadeias laterais dos modelos. A análise realizada com DaliLite (HOLM; PARK, 2000) mostrou que as proteínas dos tipos de HPV 1A, 11 e 18 exibiram distâncias estruturais de 1,9 Å, 0,5 Å e 0,8 Å, respectivamente, do modelo da proteína E7 do HPV tipo 16. Os átomos de zinco coordenados com as sequências CXXC foram adicionados a fim de preparar as estruturas para a dinâmica molecular (Figura 6). A visualização da região CR2 do N-terminal da E7 nos modelos iniciais das proteínas de todos os tipos de HPV revelaram que todos os modelos exceto o do tipo 1A, continham uma hélice α na região da sequência LXCXE, que é conhecida por ter afinidade de ligação com a proteína pRb (DYSON et al., 1989). Esta informação, assim como a semelhança de sequência (56,4%) e a distância estrutural (1,9 Å) do HPV tipo 1A com o HPV tipo 16 corroboram com a informação de que o HPV do tipo 1A não está relacionado com infecções cervicais e, conseqüentemente, para o aparecimento de

câncer cervical. Esta é outra indicação de que o conteúdo aumentado de desordem da E7 pode estar diretamente relacionado com o desenvolvimento de lesões de alto grau e câncer.

Figura 6 - Estruturas refinadas da E7 (α -hlices = verde, β -folhas = amarelo e “loops” = azul) com o átomo de zinco (vermelho). (a) HPV tipo 1A, (b) HPV tipo 11, (c) HPV tipo 16, e (d) HPV tipo 18. As figuras foram geradas usando CHIMERA (PETTERSEN et al., 2004).

5.3 Análise de Dinâmica Molecular

A fim de avaliar a estabilidade e refinar as estruturas moleculares das E7s, os modelos foram submetidos a simulações de 50 ns de dinâmica molecular (Figura 7). Durante as simulações de DM, as variações estruturais dos quatro sistemas foram medidas utilizando a análise RMSD. Os valores médios de RMSD das proteínas dos tipos 1A, 11, 16 e 18 foram $7,88 (\pm 0,9)$, $5,68 (\pm 0,75)$, $7,36 (\pm 1,53)$ e $8,26 (\pm 1,00)$ Å, respectivamente (Figura 8). Todas as proteínas E7 estudadas apresentaram alta distância média de suas estruturas iniciais, e todas as proteínas, exceto a E7 do tipo 16, estabilizaram consideravelmente após 5 ns de simulação. A E7 do HPV tipo 16 apresentou o maior desvio padrão e a ausência de uma curva plana no RMSD, é um indicativo de instabilidade. A E7 do HPV tipo 11 apresentou o menor desvio padrão da média do RMSD apesar das grandes semelhanças de sua sequência (70%) e de alinhamento estrutural (0,5 Å) com as do tipo 16. Os resultados do RMSD para os tipos de HPV 1A e 18 foram muito semelhantes, com altos valores de distância da estrutura inicial e desvios-padrão elevados. No entanto, as suas estabilidades ao longo da simulação, pareceram ser maiores do que a da E7 do HPV tipo 16.

Figura 7 - Representação da proteína E7 do HPV 16 inserida em uma caixa de água usada na simulação da dinâmica molecular.

Figura 8 - Desvio da média quadrática (RMSD) das coordenadas dos átomos C α em função do tempo de simulação, calculado para: (a) HPV tipo 16 vs. HPV tipo 1A, (b) HPV tipo 16 vs. HPV tipo 11 e (c) HPV tipo 16 vs. HPV tipo 18.

A RMSF dos átomos do carbono alfa ($C\alpha$) foi realizada ao longo do tempo de simulação, a fim de determinar quais os aminoácidos das proteínas são altamente flexíveis. Os resultados do RMSF (Figura 9) mostram que os resíduos 1-20 das proteínas E7 de todos os tipos de HPVs estudados tiveram alta variação estrutural em relação aos modelos iniciais, o que é consistente com os resultados obtidos utilizando PONDR-FIT (Figura 2). Outra semelhança entre o PONDR-FIT e os resultados RMSF foram as altas distâncias (Å) em relação aos modelos iniciais dos resíduos 20 a 50 na proteína E7 do HPV tipo 16. Os resíduos 16-40 constituem a região CR2 que contém a sequência LXCXE e altas oscilações estruturais nesta região parecem estar claramente relacionadas com a instabilidade da E7 do HPV do tipo 16, o que resulta de suas propriedades desordenadas. Os resíduos 35-60 da E7 do HPV tipo 18, que representam uma parte da região CR2 e o início do C-terminal, também apresentaram valores elevados de flutuação em relação à estrutura inicial. Chemes et al. (2010), determinaram que a contribuição do motivo LXCXE da E7 é de 90% do total de energia livre de ligação, fazendo com que este motivo seja o determinante principal da ligação da E7 à proteína pRb. Para os tipos de alto risco, HPV 16 e 18, a contribuição de CR2 aos resultados elevados de RMSD (Figura 8) parece estar diretamente relacionada com a instabilidade do N-terminal da E7. As simulações dos HPVs tipo 1A e 11 mostraram flutuações menores ao longo das suas cadeias de aminoácidos, que podem ser indicativos de estabilidade. Os resultados obtidos pela predição de desordem intrínseca e a análise de simulação de DM (RMSD e RMSF) corroboram com os dados obtidos por Uversky et al. (2006), mostrando que os HPVs oncogênicos de alto risco e suas respectivas proteínas são caracterizados por teores elevados de desordem intrínseca. As propriedades intrinsecamente desordenadas dos domínios N-terminais das E7 de alto risco também são responsáveis pela plasticidade estrutural e a seletividade, que lhes permitem ligar não só a pRb, mas a múltiplos alvos proteicos do hospedeiro (GARCÍA-ALAI; ALONSO; DE PRAT-GAY, 2007).

Figura 9 - Flutuações da média da raiz quadrática (RMSF) das coordenadas dos átomos C α para cada resíduo em função do tempo de simulação calculado para: (a) HPV tipo 16 vs. HPV tipo 1A, (b) HPV tipo 16 vs. HPV tipo 11 e (c) HPV tipo 16 vs. HPV tipo 18.

A análise dos elementos estruturais secundários durante a simulação (Figura 10) mostrou que as estruturas secundárias das E7 não apresentam grandes transformações estruturais ao longo do curso da simulação de DM, apesar da instabilidade das proteínas dos tipos de alto risco. As estruturas secundárias dos HPVs tipo 16 e 18 apresentaram padrões muito semelhantes durante a simulação e as suas estruturas diferem apenas um pouco em relação ao HPV tipo 11, que não possui a última α -hélice do domínio C-terminal, e ao HPV tipo 1A, que tem uma hélice-3₁₀, em vez de uma hélice- α na segunda hélice do domínio N-terminal. É interessante notar a ausência de uma hélice- α no motivo LXCXE da E7 do HPV tipo 1A, o que mostra que a estrutura secundária do motivo pode também estar relacionada com o tipo de infecção por HPV e/ou ligada ao grau de risco para oncogênese, porém mais dados experimentais são necessários para esclarecer essa questão.

Os resultados mostraram que a modelagem por *ab initio* produziu estruturas secundárias bem definidas nos domínios N-terminais das proteínas E7 e, apesar das suas propriedades intrinsecamente desordenadas, estes resultados foram confirmados por simulações de DM sob condições fisiológicas, com temperatura e pressão normais. As estruturas secundárias (α -hélices) no domínio N-terminal não se desenovelaram durante a simulação em nenhuma das proteínas E7 estudadas, o que difere dos resultados de experimentos de cromatografia de exclusão de tamanho (SEC) relatadas por García-Alai, Alonso e De Prat-Gay (2007), que concluíram que a região N-terminal da E7 é um domínio estrutural estendido com regiões de estrutura secundária residual dinâmica em solução.

Figura 10 - Estabilidade das estruturas secundárias durante a simulação, calculada pelo algoritmo STRIDE implementado no “plugin” Timeline do VMD. Cada cor representa um tipo de estrutura secundária: α -hélice = magenta, β -folha = amarelo, “turns” = ciano, “ loops” = branco, 3_{10} -hélice = azul, e π -hélice = vermelho.

As simulações de DM também foram utilizadas para refinar os modelos iniciais E7. As estruturas do último passo de cada uma das trajetórias da DM foram recuperadas e analisadas com PROCHECK. Os modelos da E7 dos HPVs tipos 1A, 11, 16 e 18 mostraram 86,7%, 92,9%, 86,0% e 92,6% de ocupação das posições mais favoráveis e 0,0%, 0,0%, 1,2% e 0,0% de ocupação das regiões não permitidas no gráfico de Ramachandran, respectivamente. Estes modelos apresentaram redução de qualidade em relação aos modelos iniciais (92,8%, 96,5%,

94,2% e 92,6% na mais favorecida e 1,2, 0,0%, 0,0% e 0,0% nas regiões não permitidas para tipos de HPV 1A, 11, 16, e 18, respectivamente), mostrando que o refinamento por DM das estruturas da E7 não foi útil, o que pode ter ocorrido devido a natureza desordenada da E7 mesmo para os tipos de baixo risco de HPV e devido alta distância de similaridade entre a E7 do HPV tipo 1A e a do tipo 16.

5.4 Análises de Sítios de Ligação da Região Desordenada

A previsão do sítio de ligação de IDPs utilizando ANCHOR (Figura 11) mostrou que as proteínas E7 têm uma elevada probabilidade de se ligar aos seus alvos, através dos seus domínios N-terminais, em particular os primeiros 30 aminoácidos que correspondem às regiões de CR1 e CR2. Também é notável que as E7s dos HPVs tipos 16 e 18 apresentaram valores mais elevados do que as proteínas do HPV de baixo risco, tipo 11 e o HPV tipo 1A. Os resultados para E7 do HPV tipo 16 corroboram com as previsões *in silico* (Figura 2), o RMSF das trajetórias (Figura 9) e os resultados de outros estudos experimentais (CHEMES et al., 2010) e teóricos (UVERSKY et al., 2006) que apontam que a sequência LXCXE (região CR2) é essencial para a ligação a pRb. Além disso, o conteúdo da desordem desta região é fator determinante para o grau de risco ao câncer cervical. Tanto o HPV tipo 16, aminoácidos 10-16 (CR1) e 20-26 (CR2), assim como HPV tipo 18, aminoácidos 8-14 (CR1) e 25-27 (CR2), mostraram valores de probabilidade acima do limiar estabelecido pelo ANCHOR, que é de 0,5. Estas regiões constituem a primeira e segunda α -hélices geradas pela modelagem *ab initio*.

O conceito de que a presença de uma hélice- α na região CR1 indica a sua provável capacidade de ligação foi proposto por Chemes *et al.* (2010), que demonstraram que a CR1 foi capaz de se ligar ao domínio RbAB da proteína pRb *in vitro* e deve ter se ligado realizando uma transição estrutural de “coil” para hélice. De acordo com García-Alai, Alonso e De Prat-Gay (2007), o domínio N-terminal da E7 do HPV tipo 16 é capaz de sofrer uma transição de α -hélice para β -folhas mostrando a elevada plasticidade funcional deste domínio. Muitas IDPs associadas ao câncer estão sendo estudadas atualmente como alvos potenciais de drogas, e pequenas moléculas antagonistas têm sido descritas para vários deles (SALMA; CHHATBAR; SESHADRI, 2009). A proteína E7 tem dois domínios distintos: o domínio C-terminal, um domínio estável com características de proteínas globulares que contém a região

CR3 e o domínio N-terminal, que contém as regiões de CR1 e CR2, sendo estas intrinsecamente desordenadas, em um nível que aumenta com o risco de câncer e varia de acordo com o tipo de HPV em questão. Com suporte da literatura e os resultados deste estudo, propõe-se que a região CR3 isolada e/ou a interação proteína-proteína, E7-pRb (CR1/CR2/CR3 e domínio RbAB), podem ser utilizada como alvo de descoberta de fármacos, a fim de encontrar novas abordagens para o tratamento de lesões cervicais e câncer.

Figura 11 - Análise da desordem intrínseca dos sítios de ligação da E7 ao longo da cadeia de aminoácidos dos tipos de HPV 1A, 11, 16 e 18, usando ANCHOR.

5.5 Busca de Sítios de Ligação no C-terminal da E7 de Alto Risco

Com o auxílio do método SiteFinder do programa MOE, foram encontrados seis possíveis sítios no domínio C-terminal da E7 do HPV tipo 45 sendo escolhido, o sítio de número 6 (Figura 12) que compreende os resíduos C64, V65, C66, C67, F91, L92, L95, S96, F97 da E7. A escolha do sítio 6 foi baseada em resultados obtidos por Ohlenschläger et al. (2006) que indicaram perturbações nos resíduos I62, L63, V65, C64, L75, S79, L87, Q88, Q89, L90 e F9 da região CR3, quando um peptídeo originário da proteína p21 foi adicionado no “buffer” da análise de ressonância magnética nuclear da E7 do HPV tipo 45. Esta região de ligação da p21 predita por Ohlenschläger et al. (2006) sobrepõe também à região indicada por

Liu et al. (2006) como uma região de baixa afinidade com a proteína pRb localizada no domínio C-terminal da mesma. Além disso, uma análise visual cuidadosa da superfície proteica apontou a região do sítio 6 como uma das poucas cavidades ao longo da estrutura da E7. É fato que as interfaces de interação proteína-proteína (PPI) tendem a ser grandes e relativamente planas e, geralmente, não evoluíram para ligar moléculas pequenas (SPERANDIO; MITEVA; VILLOUTREIX, 2011). Porém, apesar da cavidade pequena e plana evidenciada no sítio 6, ela representa o alvo mais indicado para a busca de modelos farmacofóricos e para a prospecção virtual com pequenos compostos, haja visto a ausência de outras regiões com características desejáveis na superfície das regiões da PPI. As PPIs são importantes alvos de estudo, pois, estão ligadas com a maioria dos processos biológicos e também com o surgimento de doenças. Assim, facilitar a busca de compostos que modulem as PPIs é de importância primária na busca de novas terapêuticas, e o uso de ferramentas *in silico* e de quimioinformática terão papel principal na busca deste objetivo (SPERANDIO et al., 2010).

Figura 12 - Superfície do domínio C-terminal de um dímero da E7 do HPV tipo 45; as esferas brancas e vermelhas indicam o sítio 6 gerado pela análise feita no SiteFinder (MOE).

5.6 “Hot spots” e Modelos Farmacofóricos da E7

Os chamados “hot spots” são pontos energéticos específicos na superfície proteica, que são os principais responsáveis pela energia de ligação, tanto na interação proteína-proteína quanto na interação entre proteínas e pequenos compostos (HAJDUK; HUTH; FESIK, 2005). A análise de “hot spots” da E7 do HPV tipo 45 foi realizada utilizando duas ferramentas, o HotspotsX 0.61 e o método “Interaction Potentials” do MOE, ambas apresentam diversas sondas químicas que foram usadas a fim de encontrar, no sítio de ligação selecionado, regiões favoráveis a interação ou campos de interação molecular. As principais características avaliadas pelas sondas identificavam os tipos de moléculas ou átomos que envolviam a interação, principalmente: se eram doadores ou aceptores de pontes de hidrogênio; se faziam parte de um composto aromático; ou se eram hidrofóbicas. Os resultados são apresentados na forma de mapas de interação na forma dos chamados “grids”, para cada sonda química utilizada. As sondas que resultaram em “grids” de interação na superfície do sítio 6 foram: C.ar (carbonosp2 em compostos aromáticos) e N.ams (carbono ou tioamida com dois átomos pesados ligados) no HotspotsX e C1= (grupo CH aromático) e N1 (grupo amida NH) no MOE. Ambas C.ar e C1= são sondas de carbonos pertencentes a compostos aromáticos e N.ams e N1 são doadoras de pontes de hidrogênio (Figura 13).

Figura 13 - Superfície do sítio de ligação 6 do domínio C-terminal da E7 do HPV tipo 45 com os “grids” gerados pela interação entre a sonda e a superfície; (a) sonda C.ar (laranja), (b) sonda C1 (laranja), (c) sonda N.ams (magenta) e (d) sonda N1 (magenta).

Os resultados das análises de “hots pots” da E7 indicaram regiões energeticamente favoráveis à ligação. Como esperado, as sondas similares, ou seja, com as mesmas características químicas (C.ar e C1= para compostos aromáticos e N.ams e N1 para doadores de ponte de hidrogênio) apresentaram “grids” em regiões semelhantes entre si para as duas análises. Os pontos de sobreposição entre elas foram usados para a geração de modelos farmacofóricos para o sítio 6 com ajuda do programa MOE. Esta etapa é fundamental, pois restringe o campo amostral de possíveis interações para apenas aquelas previstas pelos “hot spots”. Segundo Delano (2002) a identificação de resíduos envolvidos na ligação nas proteínas, e por fim os “hot spots”, podem fornecer informações para otimização de ligantes em química medicinal.

Baseado nos espaços preenchidos pelos “grids” foram gerados 15 modelos farmacofóricos, cada um contendo três pontos de anotação farmacofóricas e com posições variadas. Os pontos de anotação são caracterizados por esferas de tamanho variável, que são definidos pelo usuário. Quanto maior o tamanho das esferas menor são as restrições espaciais

de um ponto de anotação específico. Os três pontos de anotação possuem também classes e para estas análises foram utilizadas duas classes específicas: átomos e centróides. A classe átomo indica diretamente a posição, onde um átomo de uma determinada função na interação proteína-ligante deve estar, e a característica escolhida foi a de doador (Don) baseada na análise de “hot spots”. A classe centróide indica anotações que são localizadas no centro geométrico de um conjunto de átomos de uma molécula e a característica escolhida foi a de compostos aromáticos (Aro) baseada também na análise de “hot spots”. Os raios das esferas dos pontos de anotação são por padrão de 1 Å e também foram modificados para maior ou menor a fim de encontrar os melhores farmacóforos. Os modelos finais, constituídos dos três pontos de anotação, foram testados contra os bancos de dados de confôrmeros do ZINC e do MOE, a fim de se extrair uma biblioteca de ligantes que correspondessem às restrições impostas pelo farmacóforo. Os bancos de confôrmeros são essenciais na busca utilizando farmacóforos, pois o alinhamento dos compostos ao modelo não é flexível, ou seja, a flexibilização dos compostos é realizada numa etapa anterior na própria geração de confômeros espaciais para uma mesma molécula.

Os resultados para a busca indicaram que o farmacóforo 11 obteve o maior número de ligantes aceitos na biblioteca resultante das buscas com todos os farmacóforos estudados: 185 compostos. Isso indica que 185 compostos se adequaram as restrições impostas pelos pontos de anotação para este farmacóforo que foram: raios de 0.8 Å para os três pontos, sendo dois pontos Don e um Aro como pode ser visto na Figura 14.

Figura 14 - Modelo farmacofórico 11, no sítio de ligação 6, gerado com base nas análises de “hot spots” e validada por busca farmacofórica em bibliotecas de pequenos compostos utilizando MOE. F1 e F3 são pontos de anotação Don e F2 Aro.

5.7 “Docking” Macromolecular do C-terminal da E7 do HPV tipo 45

Dos milhões de compostos presentes nos bancos de dados ZINC e do MOE apenas uma pequena parcela passou pelo filtro do farmacóforo 11 e estes 185 compostos formaram uma nova e pequena biblioteca que foi utilizada na prospecção virtual por “docking” macromolecular. Segundo Walters; Stahl e Murcko (1998), as buscas farmacofóricas com milhões de confôrmeros podem ser feitas em simples estações de trabalho em apenas alguns minutos; e em casos que se possui a informação estrutural de uma proteína, torna-se útil a aplicação destes filtros farmacofóricos, pois, reduzem drasticamente a quantidade de “docking” requerido.

Na primeira etapa do “docking” macomolecular foram realizadas 10 simulações de “docking” para cada composto e as melhores posições da interação entre as moléculas e a proteína E7 foram pontuadas pelo GoldScore. As três melhores posições de cada molécula, baseada no GoldScore, foram novamente submetidas a busca farmacofórica com o farmacóforo 11. Esta segunda busca teve como intuito a seleção apenas de moléculas que apresentaram coerência com as restrições espaciais impostas pelo farmacóforo após o “docking”. Assim, foram obtidos apenas 17 moléculas (uma para cada composto) que passaram pelo filtro farmacofórico. Estas 17 moléculas foram novamente submetidas ao “docking” com os mesmos parâmetros anteriores. O complexo ligante-proteína de cada interação para cada composto foi analisado de duas formas: primeiro, visualmente com o intuito de ver se as 10 conformações de cada composto convergiam para uma mesma posição na interação com a E7 e segundo, se estes possuíam pontuações do GoldScore próximas entre si. Na Figura 15 apresenta-se um exemplo de um composto que convergiu (a) e um composto que não convergiu (b). A análise visual deixa claro que alguns compostos, após varias interações, convergem seu posicionamento para uma região determinada na formação do complexo e assim assinalando uma maior especificidade de interação. Compostos que não convergiram como o da Figura 15b, não são confiáveis para a continuidade das análises por isso foram excluídos da nova seleção.

Figura 15 - Exemplo de um composto que convergiu em relação seu posicionamento tridimensional (a) após o “docking” e um composto que não convergiu (b).

A análise visual foi também corroborada pelos resultados apresentados pela pontuação obtida no GoldScore para os complexos gerados (Tabela 5). O composto ZINC03163900 apresenta pontuações no GoldScore muito próximas para as 10 interações do Gold enquanto que o composto TimTec ST2007234 apresenta pontuações muito discrepantes entre si. As pontuações de várias simulações de “docking” em compostos que não convergem tendem a ficar distantes entre si, não demonstrando nenhuma tendência a especificidade de interações em regiões bem delimitadas do sítio ativo.

Tabela 5 - Exemplo das pontuações, por GoldScore, que convergiram e não convergiram ao longo das 10 interações do “docking”.

ID do Composto	Goldscore	ID do Composto	Goldscore
ZINC03163900	34.6	TimTec ST2007234	19.5
	34.5		28.5
	31.3		23.3
	34.3		22.9
	33.8		6.9
	34.6		28
	34.3		24.3
	33.6		16.7
	34.6		21.9
	33.7		5.6

Após a análise de convergência, apenas 10 compostos foram selecionados. Um dos compostos que apresentou a melhor convergência foi ZINC03163900, que através da análise visual evidenciou que seu posicionamento tridimensional em relação ao sítio de ligação era privilegiado e por isso mostrou-se, *a priori*, um composto promissor (Figura 15a). Desta forma, o composto ZINC03163900 foi utilizado de modelo para a geração de um novo farmacóforo, uma modificação do modelo 11 que foi chamado de modelo 16 (Figura 16).

Figura 16 - Complexo do composto ZINC03163900 com a proteína E7 junto com o farmacóforo 16. A imagem evidencia os pontos de anotação (esferas grandes) centralizados com os respectivos átomos e outros pontos de anotação que foram gerados automaticamente pelo MOE devido a presença da estrutura do composto (esferas pequenas).

O modelo 16 contém, basicamente, todas as características do modelo 11, exceto que seus pontos de anotação foram alinhados, com os átomos do composto ZINC03163900, e suas respectivas características determinadas pelos pontos foram mantidas, ou seja, dois doadores (F1 e F3 na Figura 16) e o do anel aromático (F2 na Figura 16). A busca com o farmacóforo 16 nas bibliotecas de conformeros do ZINC 2007 e do MOE 2011, resultou em 617 compostos que passaram no filtro das restrições espaciais determinadas pelo mesmo. Este farmacóforo encontrou cerca de três vezes mais compostos que o modelo 11 apesar de poucas diferenças entre eles. Estes compostos foram submetidas aos mesmos parâmetros da primeira etapa de “docking” realizada anteriormente com o farmacóforo 11 assim como uma nova

busca farmacofórica que resultou em 100 moléculas. Essas 100 moléculas foram novamente submetidas ao “docking” e analisadas posteriormente em relação à convergência utilizando os mesmos parâmetros utilizados anteriormente com o farmacóforo 11. Ao final, apenas 21 moléculas convergiram.

Devido ao seu posicionamento importante no sítio 6 da E7 o composto ZINC03163900 também serviu de base para buscas de análogos no banco de dados ZINC através da busca por estruturas similares. Foram utilizados, como parâmetros na busca de análogos para o ZINC03163900, compostos com 70% de similaridade a este. Além disso, os átomos de zinco da estrutura foram substituídos pelos de carbono a fim encontrar um número maior de análogos. Os 30 análogos encontrados foram submetidos ao “docking” utilizando os mesmos parâmetros utilizados anteriormente nos farmacóforos 11 e 16, porém, estas moléculas não passaram pela busca farmacofórica. Após o “docking” apenas 8 moléculas convergiram para mesma posição e apresentaram pontuações de GoldScore próximas.

Por fim, foi compilada uma biblioteca unindo as 10 moléculas resultantes da análise do farmacóforo 11, as 21 moléculas do farmacóforo 16 e os 8 análogos que convergiram, totalizando 39 moléculas. Esta biblioteca foi então submetida a uma nova etapa de “docking” (“re-docking”), com os mesmos parâmetros anteriores exceto pela adição da função de pontuação CHEMPLP. Por fim, as moléculas, uma representante para cada composto, que tiveram as melhores pontuações para a GoldScore e para a CHEMPLP foram analisadas com o programa DSX. A Tabela 6 apresenta as 39 moléculas e suas identificações assim como as pontuações para interação com o sítio 6 da E7. A Tabela 6 esta ordenada das maiores para os menores pontuações geradas pela função de pontuação DrugScore. As identificações dos compostos foram inseridas de acordo com número referente do banco de dados ZINC (Tabela 6). Alguns compostos não foram encontrados no ZINC, por isso, foram identificados apenas com os códigos do seu próprio banco de dados.

Tabela 6 - Pontuações da interação entre os compostos e a proteína E7 do HPV tipo 45, selecionados através da busca farmacofórica e docking macromolecular.

ID	GoldScore	CHEMPLP	DSX
ZINC70441103	*35.64	*43.21	-54.89
ZINC03163900	*34.55	36.88	-52.43
ZINC36377976	*32.49	39.37	*-59.19
ZINC38069241	*32.42	40.14	*-60.37
ZINC01513677	*31.78	29.25	-32.82
ZINC2519022	*31.70	*43.31	*-70.15
ZINC16400113	*31.30	*45.31	-48.88
EMC 001A1113	*29.67	*45.18	*-74.87
ZINC5889902	*29.50	37.50	-44.58
ZINC38069289	*29.25	*41.92	-53.36
ZINC04770631	29.24	41.38	*-69.43
ZINC00984434	29.12	35.68	-40.51
ZINC00518372	28.59	33.67	-41.98
ASDI 100077556	28.31	31.78	-45.20
ZINC00235955	27.98	35.31	-42.65
ZINC70994140	27.90	41.44	-47.37
ZINC70994100	27.89	36.70	-47.01
ZINC2164492	27.07	30.42	-46.23
ZINC306613	27.00	38.05	-41.30
ZINC03703596	26.96	40.11	-51.03
ZINC02525691	26.33	35.85	-40.05
ZINC59957745	25.97	*45.23	-56.64
Chemstar CHS 0158468	25.57	29.54	-30.35
ZINC06095504	25.46	41.32	*-64.31
ZINC04290867	25.11	33.24	-39.92
ZINC00519056	24.43	*41.90	*-58.88
ZINC39290	23.87	40.97	-55.40
ZINC37246053	22.45	*42.89	-58.36
ZINC5898533	20.55	40.91	*-65.34
ZINC4060682	20.10	40.35	*-69.03
ZINC01415049	19.90	34.12	-30.06
ZINC00381327	19.89	33.80	-32.41
ZINC03881558	18.55	41.10	-53.70
ZINC00006787	16.24	37.99	-35.97
ZINC16431666	15.85	38.99	-55.71
ZINC9458158	13.12	*53.70	*-95.56
MDPI 15475	10.85	36.64	-43.97
ZINC4844446	-4.73	39.64	-56.33
ZINC01594847	-16.07	*42.72	-53.69

* As 10 dez melhores moléculas de acordo com cada função de pontuação

As dez melhores moléculas de cada função de pontuação (indicada pelo símbolo * na Tabela 6) e suas interações com a E7 foram analisadas visualmente com ajuda do programa DSX e Pymol. Foram retiradas, para tanto, a redundância dos compostos que ficaram entre os dez melhores para duas ou mais funções de pontuação totalizando assim 19 compostos (Tabela 7).

Tabela 7 - Os 19 melhores compostos selecionados a partir de funções de pontuação e suas respectivas estruturas e características físico-químicas.

ID	Estrutura	xlogP	Don H	Accept H	Massa mol (g/mol)
ZINC70441103		4.54	2	1	225.34
ZINC03163900		2.86	4	2	246.36
ZINC36377976		2.85	4	2	226.32
ZINC38069241		5.04	2	1	239.36
ZINC01513677		-0.11	5	7	224.25

Continua

Continuação

ID	Estrutura	xlogP	Don H	Acept H	Massa mol (g/mol)
ZINC2519022		1.89	2	6	290.32
ZINC16400113		2.16	4	5	273.27
EMC 001A1113	-	-	-	-	-
ZINC5889902		-2.01	5	7	239.21
ZINC38069289		4.49	2	1	239.36
ZINC04770631		-0.06	4	7	312.28
ZINC59957745		5.61	4	2	344.5

Continua

Conclusão

ID	Estrutura	xlogP	Don H	Acept H	Massa mol (g/mol)
ZINC06095504		2.68	3	5	270.24
ZINC00519056	-	-	-	-	-
ZINC37246053		2.57	3	6	260.249
ZINC5898533		2.75	2	5	305.289
ZINC4060682		1.33	2	4	221.26
ZINC9458158		1.59	2	8	426.45
ZINC01594847		3.94	0	1	223.32

Todos os compostos selecionados, presentes na Tabela 7, apresentam características de compostos chamados “drug-like”, que são moléculas que têm propriedades que as tornam uma droga provavelmente ativa por via oral em humanos. Essa definição foi criada pela Regra dos Cinco de Christopher A. Lipinski (LIPINSKI et al., 1997), que propõe que compostos ativos oralmente, não violam mais que uma das seguintes regras: presença de não mais que cinco átomos doadores de ponte de hidrogênio; presença de não mais que 10 átomos aceptores de ponte de hidrogênio; massa molecular menor que 500 g/mol e o coeficiente de partição octanol-água $\log P$ (LEO; HANSCH; ELKINS, 1971) não superior a cinco. Apenas os compostos ZINC59957745 e ZINC38069241 apresentaram uma violação às regras ($x\log P$). Porém, estas moléculas continuam a ser admitidos como compostos “drug-like”. Os compostos EMC 001A1113 e ZINC00519056, e as informações referentes a eles, não foram encontradas no banco de dados do ZINC e por isso seus dados não aparecem na Tabela 7. Os compostos selecionados também foram analisados visualmente quanto a sua disposição em relação ao sítio de ligação e o potencial de interação dos átomos dos ligantes. A análise visual gerada pelo programa DSX (Figuras 17, 18 e 19) descreve que átomos que interagem favoravelmente são cercados por esferas azuis enquanto interações desfavoráveis são mostradas em vermelho.

Figura 17 - Representação do complexo E7-ligante obtidos por “docking” macromolecular. Os potenciais de cada átomo das interações são representados por esferas azuis e vermelhas (bons e maus contatos, respectivamente) resultantes de análise com DSX e Pymol. Os compostos são: (a) ZINC70441103; (b) ZINC03163900; (c) ZINC36377976; (d) ZINC38069241; (e) ZINC01513677 e (f) ZINC2519022.

Figura 18 - Representação do complexo E7-ligante obtidos por “docking” macromolecular analisados pelo DSX. Os compostos são: (a) ZINC16400113; (b) EMC 001A1113; (c) ZINC5889902; (d) ZINC38069289; (e) ZINC04770631 e (f) ZINC59957745.

Figura 19 - Representação do complexo E7-ligante obtidos por “docking” macromolecular analisados pelo DSX. Os compostos são: (a) ZINC06095504; (b) ZINC00519056; (c) ZINC37246053; (d) ZINC5898533; (e) ZINC4060682; (f) ZINC9458158 e (g) ZINC01594847.

Os tamanhos das esferas (Figuras 17, 18 e 19) correspondem aos valores da contribuição das pontuações de cada átomo. Esferas grandes e azuis indicam uma interação mais atraente do que esferas pequenas, grandes esferas vermelhas representam uma interação mais repulsiva do que esferas pequenas. Traços azuis ligando átomos indicam pontes de hidrogênio de distâncias adequadas, enquanto que traços vermelhos representam ponte de hidrogênio de distâncias inadequadas. A análise visual indicou que todos os ligantes apresentam-se dentro do espaço definido pelo sítio de ligação número 6, originalmente definido para o C-terminal da E7 do HPV tipo 45. Porém, a análise mostrou que, apesar da presença de maioria de átomos com potenciais de interação favoráveis (esferas azuis) em todas as moléculas, as pontes de hidrogênio apresentadas pelos traços entre átomos foram poucas e em sua maioria inadequadas (traços vermelhos). Apenas as moléculas ZINC03163900, EMC 001A1113 e ZINC59957745 apresentaram uma ponte de hidrogênio adequada (traços azuis).

Foi realizada outra análise visual das interações entre os ligantes e a E7 utilizando o programa DS Visualizer 3.5 na forma de diagramas 2D. Os resultados evidenciaram, assim como na análise do DSX, um baixo número de interações entre a E7 e seus ligantes. A grande maioria dos ligantes apresentou uma ou duas interações do tipo ponte de hidrogênio. Isto ocorre pela própria estrutura tridimensional do sítio 6. Por ser um sítio com concavidade pouco acentuada, as possibilidades de formação de pontes de hidrogênio são restritas pela sua própria conformação espacial. Isto limita a interação dos ligantes a poucos resíduos acessíveis na proteína. Apenas 4 compostos apresentaram três ou mais pontes de hidrogênio preditas: ZINC16400113, EMC 001A1113, ZINC5889902 e ZINC9458158 (Figura 20). Já os compostos ZINC38069241, ZINC38069289 e ZINC4060682 não apresentaram nenhuma ponte de hidrogênio predita (Figura 21).

Figura 20 - Ligantes que apresentaram 3 ou mais pontes de hidrogênio com a proteína E7: (a) ZINC16400113, (b) EMC 001A1113, (c) ZINC5889902 e (d) ZINC9458158. Os círculos rosa indicam resíduos envolvidos em interações de pontes de hidrogênio ligações de carga ou polar; os círculos verdes indicam resíduos envolvidos em interações de van der Waals; a linha tracejada em verde representa interações de ponte de hidrogênio com a cadeia primária de aminoácidos; a linha tracejada em azul representa interações de ponte de hidrogênio com a cadeia lateral dos aminoácidos.

Figura 21 - Ligantes que não apresentaram pontes de hidrogênio com a proteína E7: (a) ZINC38069241, (b) ZINC38069289 e (c) ZINC4060682. Os círculos rosa indicam resíduos envolvidos em interações de pontes de hidrogênio ligações de carga ou polar; os círculos verdes indicam resíduos envolvidos em interações de van der Waals.

Os aminoácidos indicados no diagrama não estão na numeração correta encontrada na natureza, pois este modelo da E7 representa apenas o domínio C-terminal da proteína. Dessa forma, para cada aminoácido do diagrama, deve-se acrescentar 50 unidades a sua numeração a fim de identificar-se a real posição de cada um. Assim, dentre os aminoácidos que aparecem circundando os compostos indicados na Figura 20, sete deles (I62, L63, C64, V65, L75, Q88

e F91; que correspondem no diagrama a ILE12, LEU13, CYS14, VAL15, LEU25, GLN38 e PHE41) estão presentes nos resultados obtidos por Ohlenschläger et al. (2006) que indicaram estes resíduos como prováveis receptores da proteína p21. Estes resultados também corroboram com a região indicada por Liu et al. (2006) como receptora de parte da interação da E7 com proteína pRb. Além disso, alguns destes aminoácidos-alvo como L63, C64, V65 e F91 (LEU13, CYS14, VAL15 e PHE41 no diagrama) apresentaram interações desejáveis através de pontes de hidrogênio com os fármacos (Figura 20). Estes resultados indicam, que o alvo biológico definido para este sítio ativo foi atingido pela interação com os compostos selecionados (exceto os da Figura 21) e que os modelos farmacofóricos tiveram êxito neste processo. Segundo Sanders et al. (2012), um dos maiores desafios dos farmacóforos baseados na estrutura é reduzir o alto número de características associadas à estrutura, para apenas aquelas relacionadas à atividade biológica

Destaca-se por fim, a análise com DSX e DS Visualizer do composto ZINC9458158, que é o maior dentre as moléculas selecionadas (426.45 g/mol) e apresentou um maior número de pontes de hidrogênio (Figura 19f e Figura 20d) com o sítio de ligação da E7. Além disso, este apresentou a melhor pontuação no DSX (-95.56) e uma análise visual indica que não houve muitos desvios de sua posição tridimensional em relação ao sítio 6 da E7.

5.8 Análise de Toxicidade dos compostos selecionados

Foram realizadas análises de toxicidade *in silico* com os 19 compostos selecionados através do programa DEREK (Tabela 8). Nestas análises foram levadas em conta, apenas predições em que a probabilidade da ocorrência dos efeitos fosse suportada por evidências científicas (indicadas pela variável PLAUSIBLE no DEREK). Os resultados apontam que apenas 3 compostos (ZINC2519022, ZINC4060682 e ZINC9458158) não apresentam efeitos tóxicos preditos para mamíferos. Seis apresentaram apenas efeitos menores, como a sensibilização da pele (ZINC70441103, ZINC16400113, ZINC59957745, ZINC06095504, ZINC00519056 e ZINC5898533). Alguns apresentaram efeitos altamente indesejáveis para candidatos a fármacos, como carcinogenicidade (ZINC03163900, ZINC36377976, ZINC38069241, ZINC38069289 e ZINC01594847), danos cromossomais (ZINC5889902 e ZINC04770631) e hepatotoxicidade (EMC 001A1113 e ZINC37246053).

Tabela 8 - Predição de toxicidade dos 19 compostos selecionados utilizando o programa DEREK.

ID	Efeito em Mamíferos	Causa
ZINC70441103	Sensibilização da pele	Amina aromática primária ou secundária
	Carcinogenicidade	Amina ou amida aromática
ZINC03163900	Sensibilização da pele	Amina aromática primária secundária Tiol ou agente de substituição de tiol
	Carcinogenicidade	Amina ou amida aromática
ZINC36377976	Sensibilização da pele	Amina aromática primária ou secundária
	Carcinogenicidade	Amina ou amida aromática
ZINC38069241	Sensibilização da pele	Amina aromática primária ou secundária
ZINC01513677	Toxicidade ocular	Base purina ou análoga
ZINC2519022	Ausência de efeitos	-
ZINC16400113	Sensibilização da pele	Fenol ou precursor Hidrazina ou precursor
EMC 001A1113	Hepatotoxicidade	Tiazole ou derivado
ZINC5889902	Dano cromossomal <i>in vitro</i>	Catecol
	Carcinogenicidade	Amina ou amida aromática
ZINC38069289	Sensibilização da pele	Amina aromática primária ou secundária
	Dano cromossomal <i>in vitro</i>	Cetona alfa, beta-insaturada Alquilfenol
ZINC04770631	Efeitos tipo cianeto	Nitrilo
	Sensibilização da pele	Fenol ou precursor
ZINC59957745	Sensibilização da pele	Amina aromática primária ou secundária
ZINC06095504	Sensibilização da pele	Resorcinol ou precursor
ZINC00519056	Sensibilização da pele	Fenol ou precursor
	Hepatotoxicidade	Composto nitro aromático
ZINC37246053	Sensibilização da pele	Fenol ou precursor
ZINC5898533	Sensibilização da pele	Fenol ou precursor
ZINC4060682	Ausência de efeitos	-
ZINC9458158	Ausência de efeitos	-
ZINC01594847	Carcinogenicidade	Amina ou amida aromática

A principal causa dos efeitos de sensibilização de pele indicados na Tabela 8, referem-se a fenóis e seus precursores e a presença de amina aromática primária ou secundária. Entre

os compostos que apresentaram apenas sensibilização de pele destaca-se o ZINC16400113 que obteve bons resultados no “docking” como evidenciado pelos digramas de interação (Figura 20a). Em relação aos efeitos de maior impacto na saúde humana os danos cromossomais e a hepatotoxicidade foram atribuídas a causas diversas. Já os efeitos de carcinogenicidade foram todos atribuídos a presença de amins e amidas aromáticas nos compostos. Amins e amidas aromáticas podem agir como agentes carginogênicos através de ativação metabólica de N-hidroxilação, O-esterificação e subsequente formação de íons nitrenium que podem ligar-se diretamente ao DNA e proteínas (ANDERSEN et al., 1964; KADLUBAR; BELAND, 1985; WILLIAMS; WEISBURGER, 1993). Interessantemente, o composto que apresenta o maior número de pontes de hidrogênio na interação com a E7 (ZINC9458158) não apresentou potenciais efeitos tóxicos ao ser humano. Já outro composto que se mostrava promissor, o ZINC03163900, que apresentou bons resultados visuais quanto a sua adequação ao sítio 6 e, por isso, foi usado também como molde para o farmacóforo 16, apresentou provável efeito carcinogênico e sensibilização da pele.

Apesar de um alto número de compostos com potencial tóxico, três que não apresentaram efeito, podem ser utilizados em análises posteriores e são indicados para etapas experimentais, com exceção do ZINC4060682, que não apresentou bons resultados na interação com a E7 (Figura 21c). Os resultados do DEREK para as 19 moléculas estudadas devem ser usados não a fim de deliberadamente eliminar os compostos deletérios, mas de apontar indicativos de toxicidade que guiem modificações químicas que anulem a toxicidade ou a busca de análogos que não apresentem tais características.

6 CONCLUSÕES

Este trabalho foi a primeira tentativa de obter modelos completos das proteínas E7 de HPVs de alto e baixo risco para o câncer cervical, apesar de suas propriedades de proteína intrinsecamente desordenada.

A modelagem *ab initio* previu que a região N-terminal desordenada da E7 contém 3 α -hélices, que foram avaliadas e validadas por MQAPs, inspeção visual, e por 50 ns de simulação de dinâmica molecular.

De acordo com a nossa modelagem e os resultados da DM, a sequência LXCXE da E7, que é determinante para a ligação a pRb, mostra uma estrutura secundária de α -hélices nas condições fisiológicas simuladas por HPV dos tipos 11, 16, e 18.

Os resultados também corroboram a literatura apontando que a desordem da E7 é maior para os tipos de HPV de alto risco do que para os tipos de baixo risco e que a desordem é determinada principalmente pelo domínio N-terminal.

Foi possível sugerir dois alvos potenciais para o desenvolvimento de pequenos compostos que possam ser utilizados para tratar a infecção pelo HPV. Em primeiro lugar, a sequência LXCXE presente nas regiões CR1 e CR2 poderia ser um possível alvo de compostos que perturbem as interações proteína-proteína, tal como na previsão *in silico* que mostra que a E7 do HPV tipo 16 tem uma elevada possibilidade de fazer contatos com outras moléculas através desta região, no intuito para formar uma estrutura mais estável.

Além disso, a estrutura de CR3 desempenha um papel importante na ligação a pRb, e os resultados da nossa modelagem e análise de DM mostrou que o domínio C-terminal é estável e pode ser utilizado como um alvo para estudos de interação ligante-proteína.

A análise de “hot spots” propiciou o desenvolvimento de modelos farmacofóricos no sítio de 6 na região CR3 do C-terminal da E7 do HPV tipo 45. Estes possibilitaram a filtragem de grandes bancos de dados selecionando apenas centenas de compostos.

O “docking” macromolecular assim como a avaliação dos potenciais de interações e a inspeção visual possibilitaram a seleção de 19 compostos candidatos a fármacos para a E7.

Dentre os 19 compostos selecionados apenas sete apresentaram pouca ou nenhuma toxicidade predita e são os candidatos para a continuação da pesquisa.

Os compostos preditos como tóxicos devem ser modificados ou a busca de análogos devem ser realizadas, a fim de dar continuidade a estudos experimentais com os mesmos

Antes de estudos experimentais, os compostos selecionados devem ser submetidos a análises futuras de dinâmica molecular em complexo com a E7 a fim de confirmar ou rejeitar uma interação coesa entre o receptor e os ligantes.

REFERÊNCIAS

- ABAGYAN, R.; TOTROV, M. Biased Probability Monte Carlo Conformational Searches and Electrostatic Calculations for Peptides and Proteins. **Journal of Molecular Biology**, v. 235, n. 3, p. 983-1002, 20 jan. 1994.
- ALTMAN, R.; DUGAN, J. Defining Bioinformatics and Structural Bioinformatics. In: **Structural Bioinformatics**. [S.l: s.n.]. p. 3-14.
- ANDERSEN, C. A. F. et al. Continuum secondary structure captures protein flexibility. **Structure**. v. 10, n. 02, p. 175-184, 2002.
- ANDERSEN, R. A. et al. Carcinogenesis and Inhibition of the Walker 256 Tumor in the Rat by trans-4-Acetylaminostilbene, Its N-Hydroxy Metabolite, and Related Compounds. **Cancer Research** , v. 24 , n. 1 Part 1 , p. 128-143, 1 jan. 1964.
- BAKER, D.; SALI, A. Protein structure prediction and structural genomics. **Science**, v. 294, n. 5540, p. 93-96, 2001.
- BALDWIN, P.; LASKEY, R.; COLEMAN, N. improving cervical screening. **Nature reviews. Cancer**, v. 3, n. March, p. 1-10, 2003.
- BALEJA, J. D. et al. Identification of inhibitors to papillomavirus type 16 E6 protein based on three-dimensional structures of interacting proteins. **Antiviral research**, v. 72, n. 1, p. 49-59, out. 2006.
- BARBOSA, M. S.; LOWY, D. R.; SCHILLER, J. T. Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. **Journal of virology**, v. 63, n. 3, p. 1404-7, mar. 1989.
- BECKER, O. **Computational Biochemistry and Biophysics**. New York: [s.n.]. p. 512
- 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.
- BERMAN, H. M. et al. The Protein Data Bank. **Nucleic Acids Research** , v. 28 , n. 1 , p. 235-242, 1 jan. 2000.
- BEUTNER, K. R.; FERENCZY, A. Therapeutic approaches to genital warts. **The American journal of medicine**, v. 102, n. 5a, p. 28-37, 5 maio. 1997.
- BINDER, K.; HEERMAN, D. **Monte Carlo Simulation in Statistical Physics**. 2nd. ed. Berlin: Springer, 1992.
- 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. **Science**, v. 253, p. 164-170, 1991.

BRADLEY, P.; MISURA, K.; BAKER, D. TOWARD HIGH-RESOLUTION DE NOVO STRUCTURE PREDICTION FOR SMALL PROTEINS. **Science**, v. 309, n. 5742, p. 1868-1871, 2005.

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.

CHEMES, L. B. et al. Targeting mechanism of the retinoblastoma tumor suppressor by a prototypical viral oncoprotein. Structural modularity, intrinsic disorder and phosphorylation of human papillomavirus E7. **The FEBS journal**, v. 277, n. 4, p. 973-88, fev. 2010.

CLEMENTS, A et al. Oligomerization properties of the viral oncoproteins adenovirus E1A and human papillomavirus E7 and their complexes with the retinoblastoma protein. **Biochemistry**, v. 39, n. 51, p. 16033-45, 26 dez. 2000.

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.

DELANO, W. L. Unraveling hot spots in binding interfaces: progress and challenges. **Current opinion in structural biology**, v. 12, n. 1, p. 14-20, fev. 2002.

Discovery Studio Visualizer. Accelrys Software Inc, , 2005.

DOSZTÁNYI, Z.; MÉSZÁROS, B.; SIMON, I. ANCHOR: web server for predicting protein binding regions in disordered proteins. **Bioinformatics (Oxford, England)**, v. 25, n. 20, p. 2745-6, 15 out. 2009.

DUNKER, A. K. et al. Intrinsically disordered protein. **Journal of Molecular Graphics and Modelling**, v. 19, n. 1, p. 26-59, fev. 2001.

DYSON, N. et al. The human papilloma virus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. **Science**, v. 17, p. 934-937, 1989.

EDELSBRUNNER, H. et al. Measuring proteins and voids in proteins. **System Sciences, 1995. Vol. V. Proceedings of the Twenty-Eighth Hawaii International Conference on**, v. 5, p. 256-264, 1995.

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, 6 jul. 2001.

GARCÍA-ALAI, M. M.; ALONSO, L. G.; DE PRAT-GAY, G. The N-Terminal Module of HPV16 E7 Is an Intrinsically Disordered Domain That Confers Conformational and

Recognition Plasticity to the Oncoprotein. **Biochemistry**, v. 46, n. 37, p. 10405-10412, 23 ago. 2007.

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

GREENE, N. et al. Knowledge-based expert systems for toxicity and metabolism prediction: DEREK, StAR and METEOR. **SAR and QSAR in Environmental Research**, v. 10, n. 2-3, p. 299-314, 1999.

HAGLER, A. T.; HULER, E.; LIFSON, S. Energy functions for peptides and proteins. I. Derivation of a consistent force field including the hydrogen bond from amide crystals. **Journal of the American Chemical Society**, v. 96, n. 17, p. 5319-5327, 1 ago. 1974.

HAJDUK, P. J.; HUTH, J. R.; FESIK, S. W. Druggability indices for protein targets derived from NMR-based screening data. **Journal of medicinal chemistry**, v. 48, n. 7, p. 2518-25, 7 abr. 2005.

HAMMOUD, M. M. HPV vaccine: Not immune to controversy. **International Journal of Gynecology & Obstetrics**, v. 101, n. 2, p. 2, maio. 2008.

HANSMANN, U. H. E. Parallel tempering algorithm for conformational studies of biological molecules. **Chemical Physics Letters**, v. 281, n. 1–3, p. 140-150, 19 dez. 1997.

HAUG, C. J. Human papillomavirus vaccination--reasons for caution. **The New England journal of medicine**, v. 359, n. 8, p. 861-2, ago. 2008.

HELT, A.; GALLOWAY, D. A. Destabilization of the Retinoblastoma Tumor Suppressor by Human Papillomavirus Type 16 E7 Is Not Sufficient To Overcome Cell Cycle Arrest in Human Keratinocytes Destabilization of the Retinoblastoma Tumor Suppressor by Human Papillomavirus Type 16 E7 Is N. **Society**, 2001.

HOLM, L.; PARK, J. DaliLite workbench for protein structure comparison. **Bioinformatics**, v. 16, n. 6, p. 566-567, 1 jun. 2000.

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.

IRWIN, J.; SHOICHET, B. ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. **Journal of Chemical Information and Modeling**, v. 45, n. 1, p. 177-182, 2005.

JAUCH, R. et al. Assessment of CASP7 structure predictions for template free targets. **Proteins: Structure, Function, and Bioinformatics**, v. 69, n. S8, p. 57-67, 2007.

JOHNSON, M. S. et al. Knowledge-Based Protein Modeling. **Critical Reviews in Biochemistry and Molecular Biology**, v. 29, n. 1, p. 1-68, 1 jan. 1994.

JONES, G. et al. Development and validation of a genetic algorithm for flexible docking. **Journal of molecular biology**, v. 267, n. 3, p. 727-48, 4 abr. 1997.

- JONES, G.; WILLETT, P.; GLEN, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. **Journal of molecular biology**, v. 245, n. 1, p. 43-53, 6 jan. 1995.
- JONES, R. E. et al. Identification of HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein. . **Journal of Biological Chemistry** , v. 265 , n. 22 , p. 12782-12785, 5 ago. 1990.
- JONES, R. E. et al. Specific N-methylations of HPV-16 E7 peptides alter binding to the retinoblastoma suppressor protein. **The Journal of biological chemistry**, v. 267, n. 2, p. 908-12, 15 jan. 1992.
- KADLUBAR, F. F.; BELAND, F. A. Chemical Properties of Ultimate Carcinogenic Metabolites of Arylamines and Arylamides. In: **Polycyclic Hydrocarbons and Carcinogenesis**. [S.l.] American Chemical Society, 1985. v. 283p. 14-341.
- 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.
- KARPLUS, M.; MCCAMMON, A. Molecular dynamics simulations of biomolecules. **Nature Structural Biology**, v. 9, n. 9, p. 646-652, 2002.
- KITCHEN, D. B. et al. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature reviews. Drug discovery**, v. 3, n. 11, p. 935-49, nov. 2004.
- KLEPEIS, J. L. et al. Ab initio prediction of the three-dimensional structure of a de novo designed protein: A double-blind case study. **Proteins: Structure, Function, and Bioinformatics**, v. 58, n. 3, p. 560-570, 2005.
- 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.
- KOSINSKI, J. et al. Template Based Prediction of Three-dimensional Protein Structures. In: **Prediction of Protein Structures, Functions, and Interactions**. [S.l.: s.n.]. p. 87-116.
- 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.
- KUNTZ, I. D. et al. A geometric approach to macromolecule-ligand interactions. **Journal of Molecular Biology**, v. 161, n. 2, p. 269-288, 25 out. 1982.
- LASKOWSKI, R. A. et al. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, 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. 139-145, 1 abr. 2000.
- LEACH, A. R. et al. Three-dimensional pharmacophore methods in drug discovery. **Journal of medicinal chemistry**, v. 53, n. 2, p. 539-58, 28 jan. 2010.

LEE, H. et al. Local structural elements in the mostly unstructured transcriptional activation domain of human p53. **The Journal of biological chemistry**, v. 275, n. 38, p. 29426-32, 22 set. 2000.

LEE, J.; WU, S.; ZHANG, Y. Ab Initio Protein Structure Prediction. In: **From Protein Structure to Function with Bioinformatics**. [S.l: s.n.]. p. 3-25.

LEO, A.; HANSCH, C.; ELKINS, D. Partition coefficients and their uses. **Chemical Reviews**, v. 71, n. 6, p. 525-616, 1 dez. 1971.

LESK, A. M.; CHOTHIA, C. H. The Response of Protein Structures to Amino-Acid Sequence Changes. **Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences**, v. 317, n. 1540, p. 345-356, 30 abr. 1986.

LIPINSKI, C. A. et al. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. **Advanced Drug Delivery Reviews**, v. 23, n. 1-3, p. 3-25, 15 jan. 1997.

LIU, X. et al. Structure of the human Papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. **The Journal of biological chemistry**, v. 281, n. 1, p. 578-86, jan. 2006.

LONGWORTH, M. S.; LAIMINS, L. A. Pathogenesis of Human Papillomaviruses in Differentiating Epithelia. **Society**, v. 68, n. 2, p. 362-372, 2004.

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. **The Journal of Physical Chemistry B**, v. 5647, n. 97, p. 3586-3616, 1998.

MAGGIO, E. T.; RAMNARAYAN, K. Recent developments in computational proteomics. **Drug Discovery Today**, v. 6, n. 19, p. 996-1004, 1 out. 2001.

MARK, W.-Y. et al. Characterization of segments from the central region of BRCA1: an intrinsically disordered scaffold for multiple protein-protein and protein-DNA interactions? **Journal of molecular biology**, v. 345, n. 2, p. 275-87, 14 jan. 2005.

MARKT, P.; SCHUSTER, D.; LANGER, T. Pharmacophore Models for Virtual Screening. In: **Virtual Screening**. [S.l: s.n.]. p. 115-152.

MCCAMMON, J. A.; GELIN, B. R.; KARPLUS, M. Dynamics of folded proteins. **Nature**, v. 267, n. 5612, p. 585-590, jun. 1977.

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.

MCINTYRE, M. C. et al. Human papillomavirus type 18 E7 protein requires intact Cys-X-X-Cys motifs for zinc binding, dimerization, and transformation but not for Rb binding. **Journal of virology**, v. 67, n. 6, p. 3142-50, jun. 1993.

MCLAUGHLIN-DRUBIN, M. E.; MÜNGER, K. The human papillomavirus E7 oncoprotein. **Virology**, v. 384, n. 2, p. 335-44, fev. 2009.

MIZIANTY, M. J. et al. In-silico prediction of disorder content using hybrid sequence representation. **BMC bioinformatics**, v. 12, n. 1, p. 245, jan. 2011.

MOE (The Molecular Operating Environment). Chemical Computing Group Inc., , 2010.

MOSCICKI, A.-B. HPV Vaccines: today and in the Future. **The Journal of adolescent health : official publication of the Society for Adolescent Medicine**, v. 43, n. 4 Suppl, p. S26-40, out. 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.

NEUDERT, G.; KLEBE, G. DSX: a knowledge-based scoring function for the assessment of protein-ligand complexes. **Journal of chemical information and modeling**, v. 51, n. 10, p. 2731-45, 24 out. 2011.

OHLENSCHLÄGER, O. et al. Solution structure of the partially folded high-risk human papilloma virus 45 oncoprotein E7. **Oncogene**, v. 25, n. 44, p. 5953-9, set. 2006.

OLDZIEJ, S. et al. Physics-based protein-structure prediction using a hierarchical protocol based on the UNRES force field: assessment in two blind tests. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 21, p. 7547-52, 24 maio. 2005.

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.

PATRICK, D. R.; OLIFF, A; HEIMBROOK, D. C. Identification of a novel retinoblastoma gene product binding site on human papillomavirus type 16 E7 protein. **The Journal of biological chemistry**, v. 269, n. 9, p. 6842-50, 4 mar. 1994.

PETTERSEN, E. F. et al. UCSF Chimera—A visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605-1612, 2004.

PHELPS, W. C. et al. The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to those of adenovirus E1A. **Cell**, v. 53, n. 4, p. 539-47, 20 maio. 1988.

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

PONS, J.-L.; LABESSE, G. @TOME-2: a new pipeline for comparative modeling of protein-ligand complexes. **Nucleic acids research**, v. 37, n. Web Server issue, p. W485-91, jul. 2009.

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.

ROGNAN, D. Docking Methods for Virtual Screening : Principles and Recent Advances. In: **Virtual Screening**. [S.l: s.n.]. p. 153-176.

ROHL, C. A et al. Protein structure prediction using Rosetta. **Methods in enzymology**, v. 383, n. 2003, p. 66-93, jan. 2004.

SALI, A. 100,000 Protein Structures for the Biologist. **Nature structural biology**, v. 5, n. 12, p. 1029-32, dez. 1998.

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

SALMA, P.; CHHATBAR, C.; SESHADRI, S. Intrinsically Unstructured Proteins: Potential Targets for Drug Discovery. **American Journal of Infectious Diseases**, v. 5, n. 2, p. 126-134, 2009.

SAMISH, I.; GU, J.; KLEIN, M. Protein Motion: Simulation. In: **Structural Bioinformatics**. 2nd. ed. [S.l.] Wiley-Blackwell, 2009. p. 907-936.

SANDERS, M. P. A. et al. From the protein's perspective: the benefits and challenges of protein structure-based pharmacophore modeling. **MedChemComm**, v. 3, n. 1, p. 28, 2012.

SCHEFFNER, M. et al. The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. **Cell**, v. 63, n. 6, p. 1129-1136, 21 dez. 1990.

SHLICK, T. Molecular Dynamics: Basics. In: **Molecular Modeling and Simulation: An Interdisciplinary Guide**. 2nd. ed. [S.l: s.n.]. p. 425-462.

SHOICHET, B. K. Virtual screening of chemical libraries. **Nature**, v. 432, n. 7019, p. 862-5, dez. 2004.

SIEW, N. et al. MaxSub: an automated measure for the assessment of protein structure prediction quality. **Bioinformatics**, 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.

SPERANDIO, O. et al. Rationalizing the chemical space of protein-protein interaction inhibitors. **Drug discovery today**, v. 15, n. 5-6, p. 220-9, mar. 2010.

SPERANDIO, O.; MITEVA, M. A.; VILLOUTREIX, B. O. Target-Based Virtual Screening to Address Protein–Protein Interfaces. In: **Virtual Screening**. [S.l.] Wiley-VCH Verlag GmbH & Co. KGaA, 2011. p. 435-465.

The PyMOL Molecular Graphics System. Schrödinger, LLC, , 2010.

TOMPA, P. Intrinsically unstructured proteins. **Trends in Biochemical Sciences**, v. 27, n. 10, p. 527-533, 1 out. 2002.

TOMPA, P. The interplay between structure and function in intrinsically unstructured proteins. **FEBS letters**, v. 579, n. 15, p. 3346-54, 13 jun. 2005.

TOMPA, P. Bioinformatics Approaches to the Structure and Function of Intrinsically Disordered Proteins. In: **From Protein Structure to Function with Bioinformatics**. [S.l.: s.n.]. p. 113-142.

TSOULOS, I. G.; STAVRAKOUDIS, A. Euch: A C++ program for molecular dynamics trajectory analysis. **Computer Physics Communications**, v. 182, n. 3, p. 8, mar. 2011.

UVERSKY, V. N. Natively unfolded proteins : A point where biology waits for physics. **Protein Science**, v. 11, n. 4, p. 739-756, 2002.

UVERSKY, V. N. et al. Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. **Journal of proteome research**, v. 5, n. 8, p. 1829-42, ago. 2006.

UVERSKY, V. N.; GILLESPIE, J. R.; FINK, A. L. Why are “natively unfolded” proteins unstructured under physiologic conditions? **Proteins: Structure, Function, and Bioinformatics**, v. 41, n. 3, p. 415-427, 2000.

UVERSKY, V. N.; OLDFIELD, C. J.; DUNKER, A. K. Intrinsically disordered proteins in human diseases: introducing the D2 concept. **Annual review of biophysics**, v. 37, p. 215-46, jan. 2008.

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

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

WALTERS, W. P.; STAHL, M. T.; MURCKO, M. A. Virtual screening—an overview. **Drug Discovery Today**, v. 3, n. 4, p. 160-178, abr. 1998.

WEINER, S. et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. **Journal of the American Chemical Society**, v. 106, n. 3, p. 765-784, 1 fev. 1984.

WERMUTH, C.-G. et al. Chapter 36. Glossary of Terms Used in Medicinal Chemistry (IUPAC Recommendations 1997. In: **CHEMISTRY, J. A. B. B. T.-A. R. IN M.** (Ed.). [S.l.] Academic Press, 1998. v. Volume 33p. 385-395.

WILLIAMS, G.; WEISBURGER, J. Chemical carcinogenesis. In: **Casarett and Doull's Toxicology: The Basic Science of Poisons**. 4th editio ed. New York: [s.n.]. p. 127-200.

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.

XUE, B. et al. PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. **Biochimica et biophysica acta**, v. 1804, n. 4, p. 996-1010, abr. 2010.

YANG, S.-Y. Pharmacophore modeling and applications in drug discovery: challenges and recent advances. **Drug discovery today**, v. 15, n. 11-12, p. 444-50, jun. 2010.

YANG, Y.; ZHOU, Y. Ab initio folding of terminal segments with secondary structures reveals the fine difference between two closely related all-atom statistical energy functions. **Protein science : a publication of the Protein Society**, v. 17, n. 7, p. 1212-9, jul. 2008.

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.

ZUR HAUSEN, H. Papillomaviruses and cancer: from basic studies to clinical application. **Nature reviews. Cancer**, v. 2, n. 5, p. 342-350, maio. 2002.

APÊNDICES

APÊNDICE A – Sequência de aminoácidos da proteína E7 dos HPVs 1A, 11, 16 e 18, no formato FASTA.

>HPV1A

MVGEMPALKDLVLQLEPSVLDLDLYCYEEVPPDDIEEELVSPQQPYAVVASCAYCEK
LVRLTVLADHSAIRQLEELLRLSLNIVCPLCTLQRQ

>HPV11

MHGRLVTLKDIVLDLQPPDPVGLHCYEQLSDSEDEVDKVDKQDAQPLTQHYQILTC
CCGCDSNVRLVVECTDGDRLQLDLLLLGTLNIVCPICAPKP

>HPV16

MHGDTPTLHEYMLDLQPETTDLYCYEQLNDSSEEEDEIDGPAGQAEPDRAHYNIVTF
CCKCDSTLRLCVQSTHVDIRTLEDLLMGTLGIVCPICSQKP

>HPV18

MHGPKATLQDIVLHLEPQNEIPVDLLCHEQLSDSEENDEIDGVNHQHLPARRAEPQR
HTMLCMCKCEARIKLVVESSADDLRAFQQLFLNTLSFVCPWCASQQ

APÊNDICE B – Comando utilizado para a modelagem do domínio N-terminl da E7 do HPV 16 por *ab initio* no Rosetta (modelagem de “loops”)

```
loopmodel.linuxgccrelease -in::file::fullatom -in::file::psipred_ss2 E7_16.psipred -
loops::input_pdb E7_16.pdb -loops::loop_file loop_file.txt -loops::frag_sizes 9 3 -
loops::frag_files E7_16_09_05.200_v1_3 E7_16_03_05.200_v1_3 -loops::ccd_closure -
loops::random_loop -loops::refine refine_ccd -loops::remodel quick_ccd_moves -loops::relax
fullrelax -out:pdb -out:nstruct 4000 -mute core.io.database --database
```

APÊNDICE C – Arquivos de configuração para a dinâmica molecular da E7 dos HPVs 1A, 11, 16, 18, respectivamente.

```
#####
## JOB DESCRIPTION                                     ##
#####

# Minimization and Equilibration of
# E701zinc in a Water Box

#####
## ADJUSTABLE PARAMETERS                               ##
#####

structure          E701mdz_wb_ion.psf
coordinates         E701mdz_wb_ion.pdb

set start_temperature 0
set temperature      310
set outputname       E701mdz_wb_eq_ion

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    58.0    0.    0.
cellBasisVector2     0.    55    0.
cellBasisVector3     0.     0    61
cellOrigin          -11    9.5   11.5
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 25000000 ;# 50ns
```

```
#####
## JOB DESCRIPTION                                ##
#####
```

```
# Minimization and Equilibration of
# E711zinc in a Water Box
```

```
#####
## ADJUSTABLE PARAMETERS                          ##
#####
```

```
structure      E711mdz_wb_ion.psf
coordinates    E711mdz_wb_ion.pdb
```

```
set start_temperature 0
set temperature      310
set outputname       E711mdz_wb_eq_ion
```

```
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
stepspercycle        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    53.0    0.    0.
cellBasisVector2     0.   54.5    0.
cellBasisVector3     0.    0   65.5
cellOrigin          -13.5    9.5   13.5
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 25000000 ;# 50ns
```

```
#####
## JOB DESCRIPTION                                     ##
#####
```

```
# Minimization and Equilibration of
# E716zinc in a Water Box
```

```
#####
## ADJUSTABLE PARAMETERS                               ##
#####
```

```
structure      E716mdz_wb_ion.psf
coordinates    E716mdz_wb_ion.pdb
```

```
set start_temperature 0
set temperature      310
set outputname       E716mdz_wb_eq_ion
```

```
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
stepspercycle        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     53.5 0. 0.
cellBasisVector2     0. 55.0 0.
```

```

cellBasisVector3      0.      0      66.0
cellOrigin             -12.5    9.5   13.5
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 25000000 ;# 50ns

```

```
#####
## JOB DESCRIPTION                                     ##
#####

# Minimization and Equilibration of
# E718zinc in a Water Box

#####
## ADJUSTABLE PARAMETERS                               ##
#####

structure          E718mdz_wb_ion.psf
coordinates         E718mdz_wb_ion.pdb

set start_temperature 0
set temperature      310
set outputname       E718mdz_wb_eq_ion

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
stepspercycle       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    54.0    0.    0.
cellBasisVector2    0.     58.5   0.
```

```

cellBasisVector3      0.    0   69.5
cellOrigin            -13.5   11   9.5
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 25000000 ;# 50ns

```

APÊNDICE D – Manuscrito do artigo submetido a revista Journal of Molecular Modeling
(não publicado)

Authors Nilson Nicolau-Junior and Silvana Giuliatti

Title Modelling and Molecular Dynamics of the Intrinsically Disordered E7 Proteins from High- and Low-risk
Types of Human Papillomavirus (HPV)

Affiliation Department of Genetics, School of Medicine of Ribeirão Preto, University of São Paulo

Address Av. Bandeirantes, 3900, Ribeirão Preto, SP, Brazil, zip code: 14049-900

E-mail silvana@rge.fmrp.usp.br **Phone** 55 (16) 3602-4503 **Fax** 55 (16) 3633-0069

Abstract Cervical cancer affects millions of women worldwide each year. Most cases of cervical cancer are caused by the sexually transmitted human papillomavirus (HPV). The approximately 40 HPV types that infect the cervix are designated high- or low-risk based on their potentials to lead to high-grade lesions and cancer. The HPV E7 oncoprotein is directly involved in the onset of cervical cancer and associates with the pRb protein and other cellular targets that promote cell immortalisation and carcinogenesis. This is the first description of the modelling and molecular dynamics analysis of complete 3-dimensional structures of high-risk (HPV types 16 and 18), low-risk (HPV type 11), and HPV type 01 E7 proteins. The models were constructed by a hybrid approach using homology (MODELLER) and *ab initio* (Rosetta) modelling, and the protein dynamics were simulated for 50 ns under normal pressure and temperature (NPT) conditions. The intrinsic disorder of the E7 protein sequence was assessed *in silico*. Complete models of E7 were obtained despite the predicted intrinsic disorder of the N-termini from the high-risk HPV types. The N-terminal domains of all of the E7 proteins studied, even those from high-risk strains, exhibited secondary structures after modelling. Trajectory analysis of E7 proteins from HPV types 16 and 18 showed higher instability in their N-terminal domains than in those of HPV types 11 and 01; however, this variation did not affect the secondary structure during the simulation. ANCHOR analysis indicated that the CR1 and CR2 regions of HPV types 16 and 18 contain possible targets for future drug-discovery studies.

Keywords Homology modelling; *Ab initio* modelling; Human papillomavirus; E7; Molecular dynamics

Introduction

Cervical cancer is the second-most common cancer among women [1]. In most cases, infection with human papillomavirus (HPV) is necessary to trigger the onset of cervical cancer. Cervical HPV infection is sexually transmitted, and most women are infected soon after becoming sexually active. More than 100 types of HPV have been identified, of which 40 infect the genital tract [2]. Most individuals remain asymptomatic after acquiring HPV infection, and only a small percentage develop clinically or histologically identifiable lesions during the course of the disease [3]. Epidemiological and molecular studies have shown that human papillomavirus induces epithelial cell transformation during infection. Specific types of HPV appear to be essential for the development of premalignant lesions and the onset of invasive cervical cancer [4]. The HPV types that infect the genital tract can be classified into 2 groups, high-risk and low-risk, based on their relative risks of high-grade lesions and cervical cancer progression [5]. The low-risk types, the most common of which are HPV types 6 and 11, induce only benign genital warts. The high-risk group, comprising types 16, 18, 31, 33, 45, and 56, is associated with the development of anogenital cancers and may be detected in 99% of cervical cancers [3].

Persistent infection with high-risk HPVs is a necessary etiological factor for developing cervical cancer. The most prevalent high-risk types, 16 and 18, are associated with approximately 70% of cervical cancers worldwide [6,7]. The high-risk HPV oncoproteins E6 and E7 are essential for HPV-induced cellular immortalisation, transformation, and carcinogenesis [8] and act by interacting with various cellular targets [9]. The E7 protein is located primarily in the nucleus, where it associates with the retinoblastoma gene product (pRb) to facilitate the progression of the S phase of the cell cycle [6]. In normal cells, pRb is hypophosphorylated at the beginning of the G1 phase and binds to E2F transcription factors to form complexes that act as transcriptional repressors. The association of the E7 protein with the hypophosphorylated pRb prevents its binding to E2F, thereby promoting cell cycle progression [9]. Sedimentation equilibrium experiments have shown that E7 has oligomerisation properties and can be found as monomers, dimers, and tetramers, but under physiological conditions it is primarily dimeric [10]. However, stoichiometry and sub-nanomolar affinity analyses indicate that E7 binds to pRb as a monomer [11]. E7 has 3 conserved regions (CR)

based on its homology with the adenovirus E1A protein: CR1, CR2, and CR3 [12]. The CR1 and CR2 regions are located in the N-terminal domain of the protein (amino acids 1–40 for HPV type 16); this domain has properties of intrinsically disordered proteins (IDPs), which confer flexibility, conformational transitions, and structural adaptability to targets of the oncoprotein [13]. The CR1 region (residues 2–15) is known to compete with E2F transcription factors and to participate in the destabilisation of pRb, although it is not clear which specific amino acids are involved in these activities [14]. The CR2 region (residues 16–40) has a specific amino acid sequence, the LXCXE motif, that has a high binding affinity for the pRb protein and is necessary for the association between E7 and pRb [15, 16, 17]. The CR3 region is located in the C-terminal domain of the protein and has 2 CXXC amino acid motifs that bind to a zinc atom [18], giving CR3 characteristics of a zinc-finger motif. The CR3 region also cooperates with the CR1 and CR2 regions to inhibit pRb and transform the cell. According to Patrick et al. [19], the full E7 protein containing the CR3 domains binds to pRb with an affinity approximately 100-fold higher than those of E7 fragments that lack the CR3 region. Studies of dimeric HPV type 1A E7-CR3 implicated 2 conserved surface patches as contact sites; patch 1 makes low-affinity contacts with the basic region within the C domain of pRb, while patch 2 makes contacts with the marked-box domain in the C-terminal region of E2F (E2F-MB) [20]. Liu et al. [20] suggested that the binding of hypothetical small molecules to patch 1 and/or 2 of the E7 CR3 region might disrupt the ability of E7 to perturb pRb function, and the development of such molecules could open a new avenue of HPV therapeutics.

The recent development of vaccines against HPV represents a major advance in human cancer prevention methods. Despite great expectations and the promising results of clinical trials, there is insufficient evidence of an effective vaccine against cervical cancer, and the overall effect of these vaccines remains unknown [21]. Moreover, the vaccine is expensive and will be a great challenge to implement where it is most needed and will have the most significant impact, as in low-income countries [22]. In addition to preventive strategies such as vaccines, therapeutic strategies remain of interest to the millions of women who are currently infected [23]. No effective treatment methodology currently exists for individuals who are already infected with HPV or have developed cervical cancer due to the infection. This study aimed to model and analyse the dynamics of the structure of E7 from high- and low-risk types of HPV. The resulting knowledge about the E7 structure may be used in further studies of protein-protein and protein-ligand interactions to create new possibilities for the treatment of HPV infection and/or cervical cancer.

Methods

In silico Intrinsic Disorder Analysis

A wide range of methods and algorithms for the detection of intrinsic disorder are currently available. *In silico* protein disorder analysis of the E7 proteins from HPV types 11 (low risk), 16 and 18 (high risk), and 01 (not related to cervical infection and cancer) was performed using PONDR-FIT [24] and DisCon [25], which detect both the contribution of each residue to the disorder and the disorder content of the protein. ANCHOR [26] analysis, which seeks to identify segments that reside in disordered regions and gain stabilising energy by interacting with globular proteins, was performed on the E7 proteins in order to predict disordered binding regions.

Macromolecular Modelling

Template Selection and Homology Modelling of E7 from HPV type 16

The E7 sequence from HPV type 16 (P03129) was retrieved from the UniProt database (<http://www.uniprot.org>) and modelled partially by homology using the NMR solution structure from the Protein Data Bank (PDB) of the C-terminal domain (monomer) of the HPV type 45 oncoprotein E7 (PDB ID: 2EWL), which has 65% similarity with E7 from HPV type 16, as a template. The homology modelling was performed using MODELLER [27] implemented on the @TOME2 server (<http://atome.cbs.cnrs.fr>) [28].

Ab Initio Modelling

The N-terminal region of the model of E7 from HPV type 16 appeared unstructured due to its intrinsically disordered nature. The *ab initio* modelling technique was used to generate a complete 3-dimensional protein molecule. *Ab initio* modelling is used when there is little initial structural information about the molecule. Rosetta 3.1 [29], more specifically the protocol of loop modelling using the cyclic coordinate descent (CCD) algorithm [30], was used for this purpose. About 1000 models were generated, one of which was chosen after assessment and validation using Model Quality Assessment Predictors (MQAPs) and

visual inspection. The HPV type 16 E7 model was used as a template for homology modelling of the E7 proteins from HPV types 01 (P06465), 11 (P04020), and 18 (P06788).

Homology Modelling of E7 from HPV types 01, 11, and 18

MODELLER (v9.10) was used to generate homology models of the E7 proteins from HPV types 01, 11, and 18 using the previously generated structure of the HPV type 16 E7 protein as a template. Twenty models of each protein were generated and then assessed using PROCHECK [31].

Rotamer Modelling and Zinc Atom Coordination

In order to obtain high-quality structures, the side chains (rotamers) of all protein models were re-modelled using SCWRL4 [32]. The coordinates of the zinc atoms of the zinc finger in the structure 2EWL (PDB) were manually transferred to the homology models using *Accelrys* Discovery Studio Visualizer v2.0 (<http://accelrys.com/>).

Structure Validation

The structure validation and more particularly the choice of the near-native models demand not only visual analysis of the structure but local (each amino acid residue) and global (whole protein) validation of the 3-dimensional protein structure. Several MQAPs were used to select the most accurate E7 protein models. In the first step, the ModFOLD [33], dDFIRE [34], and QMEAN [35] scoring functions were used to rank the 1000 decoys from the model of E7 from HPV type 16. The 10 top-scoring structures of each MQAP continued in the analysis. There was no redundancy in the top 10 models selected by each program, so the resultant 30 structures were ranked again based on the PROCHECK results. In order to choose the best-quality structures, the top 10 remaining models were then submitted to the following MQAPs: Protein Quality Predictor (ProQ) [36], which uses quality measures called LGscore [37] and MaxSub [38], ERRAT [39], Verify 3D [40,41], and Protein Volume Evaluation (PROVE) [42].

Simulation of Molecular Dynamics

The molecular dynamics (MD) of the E7 proteins were determined using the CHARMM27 force field [43] implemented in the NAMD 2.8 software [44]. The MD simulation was performed over 50 ns for each protein (E7 from HPV types 01, 11, 16, and 18) for a total of 0.20 μ s using a constant temperature (310 K) and pressure (1 atm). Each system was solvated in a cubic box of TIP3P water with a minimum distance of 10 Å imposed between the solute atoms and the edge of the box. The system charges were balanced by adding chloride and sodium ions. The Particle Mesh Ewald Method (PMEM) was used to determine long-distance electrostatic interactions. All bonds involving hydrogen were fixed using the SHAKE algorithm [45]. The number of atoms in each system was 17,551, 17,081, 17,331, and 20,034 for HPV types 01, 11, 16, and 18, respectively. Before the MD simulation, the entire system was minimised for 5000 steps to reduce any unfavourable interactions between the protein and solvent. Each system was slowly heated up to 310 K over the first 52 ps under NPT conditions.

Trajectory Analysis

The trajectory obtained in the molecular dynamics simulation was analysed using VMD [46], eucb [47], and carma [48]. Carma and VMD were used to transform the complete trajectory data (solvent, Cl, Na, Zn, and protein) into a trajectory including only the protein structure; this then served as the input for the eucb program. Two eucb protocols were used to analyse the trajectories: the root mean-square deviation (RMSD) and root mean-square fluctuation (RMSF) of the α -C atoms using the initial structure as the reference. The variations in the secondary structures over time were analysed using the STRIDE [49] algorithm implemented in the TIMELINE plug-in from VMD.

Results and Discussion

The E7 proteins from high- and low-risk HPV types are intrinsically disordered

Analyses of the E7 sequences performed with the program DisCon (Table 1) showed that the E7 proteins from HPV types 16 and 18 had greater contents of disorder, 28.57% and 10.47%, respectively, whereas this content was very low (~ 1.0) for the low-risk type 11 and zero for type 01. According to Uversky et al. [50], the intrinsic disorder of the E7 protein increases the risk for cervical cancer; therefore, the E7 proteins of the low-risk HPV types 06 and 11 are less disordered, that of HPV type 16 is the most disordered, and that of HPV type 18 is intermediate. Interestingly, proteins linked to other types of cancer, such as P53 [51] and BRCA1 [52], also have properties of intrinsically disordered proteins.

PONDR-FIT is used to determine which amino acids in the protein backbone are disordered; scores above 0.5 represent amino acids with intrinsically disordered characteristics (Fig. 1). All of the E7 proteins studied showed increased disorder in the first and last 10 amino acids. The E7 proteins from HPV types 16 and 18 showed higher levels of disorder, and the greatest contribution to this disorder came from their N-terminal domains, in which the scores for almost all amino acids were above the threshold value of 0.5. The PONDR-FIT scores of HPV type 11 were lower than those of the high-risk HPVs despite its high sequence similarity (70%) with type 16. As for the results obtained using DisCon, the PONDR-FIT values of disorder for the HPV type 01 N-terminal domain were below the threshold, indicating the absence of disorder. As HPV type 01 is not related to genital tract infection and cancer, this result indicates that the presence of disorder may be linked to the types of HPV that lead to the onset of cancer. According to Alaí-García et al. [13], the N-terminus of HPV type 16 E7 can be described as intrinsically disordered and has a plastic structural domain that probably optimises the speed of interactions and allows less-specific but selective connections to multiple host-cell proteins. The IDPs have been classified based on their modes of action [53], and the E7 proteins from high-risk HPVs could be classified as *effector* IDPs because their disorder modifies the activity of a single partner or assembly (in this case, the prB protein).

Table 1 Analysis of the disorder contents of the E7 proteins from HPV types 01, 11, 16, 18 using DisCon.

Acession Number	HPV type	Disorder Content (%)
P03129	16	28.57
P06788	18	10.47
P04020	11	1.02
P06465	1	0

Fig. 1 Analysis of the intrinsic disorder along the amino acids chains of the E7 proteins from HPV types 01, 11, 16, and 18 using POND-R-FIT.

Molecular Modelling of E7

The HPV type 16 E7 models generated by the combination of homology modelling and *ab initio* techniques were validated, and MQAP scores (Table 2) and visualisation was used to select the best model.

Although the E7 proteins from high-risk HPV types showed IDP properties, an initial structure of the E7 protein of HPV type 16 was successfully obtained and used as a template to model the E7 proteins of the other types. These models were used as starting structures for the MD simulations. The HPV type 16 E7 protein is an alpha-beta protein containing 5 α -helices, 3 of which are in the N-terminal domain, and 1 antiparallel β -sheet in the C-terminal domain (Fig. 2c). The C-terminal domain also contains 2 CXXC sequences that together form a zinc-finger motif.

The structure selected, initially called S_0139, was used as the template for homology modelling of the E7 proteins of HPV types, 01, 11, and 18, which exhibit 56.4%, 70%, and 61.7% sequence similarity, respectively, with that of HPV 16. Homology modelling generated 20 models for each protein, and the best models were chosen using PROCHECK.

Table 2 Selection of best model of HPV type 16 E7 based on the MQAP scores

Model	PROCHECK (%)	Lg score	MaxSub	ERRAT (%)	Verify3D (%)	Z-Score(Mean)	Z-Score (RMS)
S_0652	94.2 / 5.8 / 0/0 *	0.189	0.077	84.44	14.14	0.06*	1.33
S_0139	94.2 / 5.8 / 0/0 *	0.042	0.022	96.67 *	34.34	0.06*	1.3
S_0885	94.2 / 5.8 / 0/0 *	0.052	0.049	96.67 *	55.56	0.14	1.37
S_0911	93.0 / 7.0 / 0/0 *	0.034	0.057	89.65	41.41	0.14	1.45
S_0109	93.0 / 7.0 / 0/0 *	0.124	0.069	98.89 *	21.21	0.02*	1.29
S_0101	93.0 / 7.0 / 0/0 *	-0.288	-0.002	93.33	48.48	0.13	1.33
S_0382	93.0 / 7.0 / 0/0 *	0.471	0.11	94.44	77.78	-0.1	1.34
S_0994	91.9 / 8.1 / 0/0 *	0.41	0.048	82.22	22.22	0.14	1.33
S_0515	91.9 / 8.1 / 0/0 *	0.484	0.15	100	27.27	-0.11	1.11
S_0415	91.9 / 8.1 / 0/0 *	-0.061	-0.019	91.9	43.43	0.05 *	1.17

* Represents scores that are expected to indicate good structural quality

The E7 proteins were submitted to rotamer modelling in order to refine the side chains of the protein models. Analysis performed with Dalilite [54] showed that the HPV type 01, 11, and 18 proteins exhibited distances of 1.9 Å, 0.5 Å, and 0.8 Å, respectively, from the template E7 protein from HPV type 16. Zinc atoms coordinated with the CXXC sequences were added in order to prepare the structures for the MD step (Fig. 2). Visualisation of the CR2 region from the E7 N-terminus in the initial models of the proteins from all HPV types showed that all of the models except that of HPV type 01 E7 contained an alpha helix spanning the LXCXE sequence, which is known to have binding affinity for the pRb protein [15]. This information, as well as the sequence similarity (56.4%) and structural distance (1.9 Å) of HPV type 01 from HPV type 16, corroborate that HPV type 01 is not related to cervical infections or, consequently, to the onset of cervical cancer. This is another indication that the increased content of disorder of E7 may be directly related to the development of high-grade lesions and cancer.

Fig. 2 Refined structures of E7 (alpha helix = green, beta sheets = yellow, and loops = blue) with the zinc atom (red). (a) HPV type 01, (b) HPV type 11, (c) HPV type 16, and (d) HPV type 18. The figures were generated using CHIMERA [55].

Molecular Dynamics Analysis

In order to assess the stability of and refine the E7 molecular structures, the models were subjected to 50-ns molecular dynamics simulations. During the MD simulations, the structural variations of the 4 systems were measured using RMSD analysis. The mean RMSD values of the HPV type 01, 11, 16, and 18 proteins were $7.88 (\pm 0.9)$, $5.68 (\pm 0.75)$, $7.36 (\pm 1.53)$ and $8.26 (\pm 1.00)$ Å, respectively (Fig. 3). All of the E7 proteins studied showed high mean distances from their initial structures, and all of the proteins except for the type 16 E7 stabilised considerably after approximately 5 ns. The E7 from HPV type 16 showed the highest standard deviation, and the absence of a flat RMSD curve may indicate instability. The E7 from HPV type 11 exhibited the lowest mean RMSD and standard deviation despite the great similarities of its sequence (70%) and structural alignment (0.5 Å) with those of type 16. The results for HPV types 01 and 18 were very similar, showing high values of distance from the initial structure and high standard deviations; however, their stabilities over the course of the simulation seemed to be higher than those of the E7 from HPV type 16.

Fig. 3 Root mean-square deviations (RMSD) of the E7 C α atom coordinates as a function of the simulation time, calculated for: (a) HPV type 16 vs. HPV type 01, (b) HPV type 16 vs. HPV type 11, and (c) HPV type 16 vs. HPV type 18.

A RMSF of the α -C atoms was performed during the simulation time in order to determine which amino acids of the proteins are highly flexible. The RMSF results (Fig. 4) showed that residues 1–20 of the E7 proteins of all of the HPVs studied fluctuated greatly from the initial models, which is consistent with the results obtained using PONDR-FIT (Fig. 1). Another similarity between the PONDR-FIT and RMSF results was the high distances ($\text{\AA}$) from the initial models for residues 20–50 of the HPV type 16 E7. Residues 16–40 make up the CR2 region that contains the LXCXE sequence, and the high fluctuations in this region may be clearly related to the instability of the HPV type 16 E7 that results from its disordered properties. Residues 35–60 of E7 from HPV type 18, which represent part of the CR2 region and the beginning of the C-terminus, also showed high values of fluctuation from the initial structure. Chemes et al. [11] determined the contribution of the LXCXE motif of E7 to be 90% of the total binding free energy, making this motif the main determinant of the binding of E7 to the pRb protein. For the high-risk HPV types 16 and 18, the contribution of CR2 to the high RMSD results (Fig. 3) seems to be directly linked to the N-terminal instability of E7. The HPV types 01 and 11 simulations showed minor fluctuations along their amino acid chains, which could be indicative of stability. The results obtained by intrinsic disorder prediction and MD simulation analysis (RMSD and RMSF) corroborate the data gathered by Uversky et al. [50] showing that the high-risk oncogenic HPVs and their respective proteins are characterised by increased contents of intrinsic disorder. The intrinsically disordered properties of the N-terminal domains of high-risk E7 proteins are also responsible for the structural plasticity and selectivity that allow them to bind not only pRb but multiple other host proteins as well [13].

Fig. 4 Root mean-square fluctuations (RMSF) of the averaged E7 C α atom coordinates for each residue as a function of the simulation time calculated for: (a) HPV type 16 vs. HPV type 01, (b) HPV type 16 vs. HPV type 11, and (c) HPV type 16 vs. HPV type 18.

Analysis of the secondary structural elements during the simulation (Fig. 5) showed that the E7 secondary structures were not subject to great changes over the course of the MD simulation despite the instability of the proteins from high-risk types. The secondary structures of HPV types 16 and 18 showed very similar patterns throughout the simulation, and their structures differed only a little from those of HPV type 11, which lacks the last α -helix of the C-terminal domain, and HPV type 01, which has a 3_{10} -helix instead of an α -helix as the second helix of the N-terminal domain. Interestingly, the absence of an α -helix spanning the LXCXE motif of HPV type 01 E7 shows that the secondary structure of that motif could be related to the type of HPV infection or/and linked with the grade of risk for cancer formation, but further research will be required to clarify that question.

Our results have already shown that ab initio modelling predicted well-defined secondary structures in the N-terminal domains of the E7 proteins despite their intrinsically disordered properties, and these results were confirmed by MD simulations under physiological conditions with normal temperature and pressure. The secondary structures (α -helices) in the N-terminal domain did not unfold during the simulation for any of the E7 proteins studied; this differs from the results of size-exclusion chromatography (SEC) experiments reported by Garcia-Alai et al. [13], who concluded that the E7 N-terminal region is an extended structural domain with regions of dynamic residual secondary structure in solution.

Fig. 5 Stability of the secondary structures during the simulation calculated by the STRIDE algorithm implemented in the Timeline plug-in of VMD. Each colour represents a type of secondary structure: α -helix = magenta, β -sheet = yellow, turns = cyan, loops = white, 3_{10} -helix = blue, and π -helix = red.

The MD simulations were also used to refine the initial E7 models. The structures of the last frame from each of the MD trajectories were retrieved and assessed with PROCHECK. The models of E7 from HPV types 01, 11, 16, and 18 showed 86.7%, 92.9%, 86.0%, and 92.6% occupancies of the most-favoured positions and 0.0%, 0.0%, 1.2%, and 0.0% occupancies of the disallowed regions of the Ramachandran plot, respectively. These models were considerably reduced in quality from the initial models (92.8%, 96.5%, 94.2% and 92.6% in the most-favoured and 1.2%, 0.0%, 0.0%, and 0.0% in the disallowed regions for HPV types 01, 11, 16, and 18,

respectively), showing that MD refinement of the E7 structures was not helpful; this may be due to the disordered natures of the E7 proteins from even the low-risk types of HPV and the high similarity distance between the E7 protein from HPV type 01 to type 16.

Binding Sites Analysis

The IDP binding site prediction using ANCHOR (Fig. 6) showed that the E7 proteins have a high probability of binding to their targets via their N-terminal domains, especially the first 30 amino acids that correspond to the CR1 and CR2 regions. It is also remarkable that the E7 proteins from HPV types 16 and 18 showed higher values than those from the low-risk HPV type 11 and HPV type 01. The results for E7 from HPV type 16 corroborate the *in silico* predictions (Fig. 1), the RMSF of the trajectories (Fig. 4), and the findings of other studies [11,50] that point out that the LXCXE sequence (the CR2 region) is essential for binding to pRb. In addition, the content of disorder of this region is a determinant of the level of risk for cervical cancer. Both HPV type 16 amino acids 10–16 (CR1) and 20–26 (CR2), as well as HPV type 18 amino acids 8–14 (CR1) and 25–27 (CR2), showed probability values above the Anchor threshold of 0.5; these regions constitute the first and second α -helices from the results of *ab initio* modelling. The concept that the presence of an α -helix in the CR1 region indicates its probable binding capacity has already been proposed by Chemes et al. [11], who showed that CR1 was able to bind to the RbAB domain of the pRb protein *in vitro* and may have done so by undergoing a coil-to-helix transition. According to Garcia-Alai et al. [13], the N-terminal domain of E7 from HPV type 16 is capable of undergoing an α -helix to β -sheet transition, showing the high functional plasticity of this domain. Many cancer-associated IDPs are currently being studied as potential drug targets, and small molecule antagonists of several have been described [56]. The E7 protein has 2 distinct domains: the C-terminal domain, a stable domain with protein globularity features that contains the CR3 region, and the N-terminal domain, which contains the CR1 and CR2 regions and is intrinsically disordered at a level that increases with the risk for cancer posed by the type of HPV in question. From the previously published literature and the results of our studies, we propose that the E7 CR3 region alone and/or the E7-pRb protein-protein interaction (CR1/CR2/CR3-RbAB domain) could be used as targets of drug-design programs in order to find new approaches to the treatment of cervical lesions and cancer.

Fig. 6 Intrinsic disorder analysis using ANCHOR of the binding sites along the E7 protein amino acid chains of HPV types 01, 11, 16, and 18.

Conclusions

This work was the first attempt to obtain complete models of the E7 proteins from high- and low-risk types of HPV despite their intrinsically disordered features. The *ab initio* modelling predicted that the disordered C-terminus of E7 contains 3 α -helices, which were assessed by MQAPs, visual inspection, and a 50-ns molecular dynamics simulation. According to our modelling and MD results, the LXCXE sequence of E7, which is a determinant of pRb binding, shows an α -helical structure under the physiological conditions simulated for HPV types 01, 11, 16, and 18. Our results also corroborate the literature showing that the disorder of E7 is greater for high-risk HPV types than for low-risk types and that disorder is mainly determined by the N-terminal domain. It was possible to suggest 2 potential targets for the development of small compounds that could be used to treat HPV infection. First, the LXCXE sequence of the CR1 and CR2 regions could be a possible target for disrupting protein-protein interactions, as *in silico* prediction shows that the E7 of HPV type 16 has a high possibility of making contacts with other molecules via this region in order to form a more stable structure. In addition, the CR3 structure plays an important role in the binding to pRb, and the

results of our protein modelling and MD analysis showed that the C-terminal domain is stable and could be used as a target for ligand-interaction studies. The validation of the complete model of the high-risk E7 may be a step towards the rational design of drugs against HPV infections and cervical cancer.

Acknowledgements

This work was supported by CAPES Fellowship.

References

1. Parkin DM, Bray F (2006) Chapter 2: the burden of HPV-related cancers. *Vaccine* 24:11–25
2. De Villiers EM, Fauquet C, Broker TR, Bernard HU, Zur Hausen H (2004) Classification of papillomaviruses. *Virology* 324:17–27
3. Kanodia S, Fahey LM, Kast WM (2007) Mechanisms used by human papillomaviruses to escape the host immune response. *Curr Cancer Drug Targets* 7:79–89
4. Zur Hausen H, Meinhof W, Scheiber W, Bornkamm GW (1974) Attempts to detect virus-specific DNA sequences in human tumors: I. Nucleic acid hybridizations with complementary RNA of human wart virus. *Int J Cancer* 13:650–656
5. Kleter B, Van Doorn LJ, Ter Schegget J, Schrauwen L, Van Krimpen K, Burger M, Ter Harmsel B, Quint W (1998) Novel short-fragment PCR assay for highly sensitive broad-spectrum detection of anogenital human papillomaviruses. *Am J Pathol* 153:1731–1739
6. Zur Hausen H (2002) Papillomaviruses and cancer: from basic studies to clinical application. *Nat Rev Cancer* 2:342–350
7. Scheffner M, Werness BA, Huibregtse JM, et al. (1990) The E6 oncoprotein encoded by human papillomavirus types 16 and 18 promotes the degradation of p53. *Cell* 63:1129–1136
8. Narisawa-Saito M, Kiyono T (2007) Basic mechanisms of high-risk human papillomavirus-induced carcinogenesis: roles of E6 and E7 proteins. *Cancer Sci* 98:1505–1511

9. Boulet G, Horvath C, Vanden BD, Sahebali S, Bogers J (2007) Human papillomavirus: E6 and E7 oncogenes. *Int J Biochem Cell Biol* 39:2006–2011
10. Clements A, Johnston K, Mazzarelli JM, Ricciardi RP, Marmorstein R (2000) Oligomerization properties of the adenovirus E1A viral oncoproteins and human papillomavirus E7 and their complexes with the protein retinoblastoma. *Biochemistry* 39:16033–16045
11. Chemes LB, Sanchez IE, Smal C, Prat-Gay G (2010) Targeting mechanism of the retinoblastoma tumor suppressor by the prototypical viral oncoprotein. *FEBS J* 277: 973–988
12. Phelps WC, Yee CL, Munger K, Howley PM (1988) The human papillomavirus type 16 E7 gene encodes transactivation and transformation functions similar to adenovirus E1A. *Cell* 53:539–547
13. Garcia-Alai MM, Alonso LG, de Prat-Gay G (2007) The N-terminal module of HPV16 E7 is an intrinsically disordered domain that confers conformational and recognition plasticity to the oncoprotein. *Biochemistry* 46:10405–10412
14. Helt AM, Galloway DA (2001) Destabilization of the retinoblastoma tumor suppressor by human papillomavirus type 16 e7 is not sufficient to overcome cell cycle arrest in human keratinocytes. *J Virol* 75(15):6737–6747
15. Dyson N, Howley PM, Munger K, Harlow E (1989) The human papillomavirus-16 E7 oncoprotein is able to bind to the retinoblastoma gene product. *Science* 243:934–937
16. Jones RE, Wegrzyn RJ, Patrick DR, Balishin NL, Vuocolo GA, Riemen MW, Defeo-Jones D, Garsky VM, Heimbrook DC, Oliff A (1990) Identification of HPV-16 E7 peptides that are potent antagonists of E7 binding to the retinoblastoma suppressor protein. *J Biol Chem* 265:12782–12785
17. Jones RE, Heimbrook DC, Huber HE, Wegrzyn RJ, Rotberg NS, Stauffer KJ, Lumma PK, et al. (1992) Specific N-methylations of HPV-16 E7 peptides alter binding to the retinoblastoma suppressor protein. *J Biol Chem* 267:908–912
18. Barbosa MS, Lowy DR, Schiller JT (1989) Papillomavirus polypeptides E6 and E7 are zinc-binding proteins. *J Virol* 63:1404–1407

19. Patrick DR, Oliff A, Heimbrook DC (1994) Identification of a novel retinoblastoma gene product binding site on human papillomavirus type 16 E7 protein. *J Biol Chem* 269:6842–6850
20. Liu X, Clements A, Zhao K, Marmorstein R (2006) Structure of the human papillomavirus E7 oncoprotein and its mechanism for inactivation of the retinoblastoma tumor suppressor. *J Biol Chem* 281:578–586
21. Haug CJ (2008) Human papillomavirus vaccination – reasons for caution. *N Engl J Med* 359:861–862
22. Hammound M (2008) HPV vaccine: not immune to controversy. *Int J Gynaecol Obstet* 101:123–124
23. Moscicki A (2008) HPV vaccines: today and in the future, *J Adolescent Health* 43:26–40
24. Xue B, Dunbrack RL, Williams RW, Dunker AK, Uversky VN (2010) PONDR-FIT: a meta-predictor of intrinsically disordered amino acids. *Biochim Biophys Acta* 1804:996–1010
25. Mizianty MJ, Zhang T, Xue B, Zhou Y, Dunker AK, Uversky VN, Kurgan L (2011) In-silico prediction of disorder content using hybrid sequence representation. *BMC Bioinformatics* 12(1):245
26. Dosztányi Z, Mészáros B, Simon I (2009) ANCHOR: web server for predicting protein binding regions in disordered proteins. *Bioinformatics* 25(20):2745–2746
27. Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815
28. Pons JL, Labesse G (2009) @TOME-2: A new pipeline for comparative modeling of protein-ligand complexes. *Nucleic Acids Res* 37:485–491
29. Rohl CA, Strauss CE, Misura KMS, Baker D (2004) Protein structure prediction using Rosetta. *Methods Enzymol* 383:66–93
30. Canutescu A, Dunbrack RL Jr (2003) Cyclic coordinate descent: a robotics algorithm for protein loop closure. *Protein Sci* 12:963–72
31. Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst* 26:283–291

32. Krivov GG, Shapovalov MV, Dunbrack RL Jr (2009) Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* 77:778–795
33. McGuffin LJ (2008) The ModFOLD server for the quality assessment of protein structural models. *Bioinformatics* 24:586–587
34. Yang Y, Zhou Y (2008) Specific interactions for ab initio folding of protein terminal regions with secondary structures. *Proteins* 72:793–803
35. Benkert P, Tosatto SCE, Schomburg D (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins* 71(1):261–277
36. Wallner B, Elofsson A (2003) Can correct protein models be identified? *Protein Sci* 12:1073–1086
37. Cristobal S, Zemla A, Fischer D, Rychlewski L, Elofsson A (2001) A study of quality measures for protein threading models. *BMC Bioinformatics* 2:5–20
38. Siew N, Elofsson A, Rychlewski L, Fischer D (2000) Maxsub: An automated measure to assess the quality of protein structure predictions. *Bioinformatics* 16:776–785
39. Colovos C, Yeates TO (1993) Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci* 2:1511–1519
40. Bowie JU, Luethy RE, Eisenberg D (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* 253:164–170
41. Luethy R, Bowie JU, Eisenberg D (1992) Assessment of protein models with three-dimensional profiles. *Nature* 356:83–85
42. Pontius J, Richelle J, Wodak SJ (1996) Deviations from standard atomic volumes as a quality measure of protein crystal structures. *J Mol Biol* 264:121–136
43. MacKerell AD Jr, Bashford D, Bellott M, Dunbrack RL Jr, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WR III, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiórkiewicz-Kuczera J, Yin D, Karplus M

- (1998) All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J Phys Chem B* 102:3586–3616
44. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K (2005) Scalable molecular dynamics with NAMD. *J Comp Chem* 26:1781-1802
45. Van Gunsteren WF, Berendsen HJC (1977) Algorithms for macromolecular dynamics and constraint dynamics. *Mol Phys* 34:1311–1327
46. Humphrey W, Dalke A, Schulten K (1996) VMD – Visual Molecular Dynamics. *J Mol Graph* 14:33–38
47. Tsoulos IG, Stavrakoudis A (2011) Eucb: a C++ program for molecular dynamics trajectory analysis. *Comp Phys Commun* 182:834–841
48. Glykos, NM (2006) Carma: a molecular dynamics analysis program. *J Comp Chem* 27:1765–1768
49. Andersen CAF, Palmer AG, Brunak S, Rost B (2002) Continuum secondary structure captures protein flexibility. *Structure* 10:175–185
50. Uversky VN, Roman A, Oldfield CJ, Dunker AK (2006) Protein intrinsic disorder and human papillomaviruses: increased amount of disorder in E6 and E7 oncoproteins from high risk HPVs. *J Proteome Res* 5:1829–1842
51. Lee H, Mok KH, Muhandiram R, Park JE, Suk KH, et al. (2000) Local structural elements in mostly the unstructured transcriptional activation domain of human p53. *J Biol Chem* 275:29426–29432
52. Mark WY, Liao JC, Lu Y, Ayed A, Laister R, et al. (2005) Characterization of segments from the central region of BRCA1: an intrinsically disordered multiple scaffold for protein-protein and protein-DNA interactions. *J Biol* 345:275–287
53. Tompa, P (2002) Intrinsically unstructured proteins. *Trends Biochem Sci* 27:527–533
54. Holm L, Park J (2000) DaliLite workbench for protein structure comparison. *Bioinformatics* 16:566–567
55. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) UCSF Chimera: a visualization system for exploratory research and analysis. *J Comput Chem* 25(13):1605–1612

56. Salma P, Chhatbar C, Seshadri S (2009) Intrinsically unstructured proteins: potential targets for drug discovery. *Am J Infect Dis* 5:126–134