

**UNIVERSIDADE DE SÃO PAULO**  
**FACULDADE DE MEDICINA DE RIBEIRÃO PRETO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM IMUNOLOGIA BÁSICA E APLICADA**

**THAIS CRISTINE ARNS**

**Modelagem molecular e comportamento dinâmico da molécula imunomoduladora HLA-G e isoformas solúveis**

**RIBEIRÃO PRETO**

**2018**

**THAIS CRISTINE ARNS**

**Modelagem molecular e comportamento dinâmico da molécula imunomoduladora HLA-G e isoformas solúveis**

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

**Área de Concentração: Imunologia.**

**Orientador: Prof. Dr. Eduardo Antônio Donadi.**

**RIBEIRÃO PRETO**

**2018**

**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.**

### **FICHA CATALOGRÁFICA**

Arns, Thais Cristine

Modelagem molecular e comportamento dinâmico da molécula imunomoduladora HLA-G e isoformas solúveis.

Ribeirão Preto, 2018.  
245p.

Tese de Doutorado apresentada à Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

Área de concentração: Imunologia.

Orientador: Donadi, Eduardo Antônio.

1.HLA-G. 2.Modelagem molecular. 3.Dinâmica molecular. 4. *Docking* molecular. 5. Bioinformática.

## FOLHA DE APROVAÇÃO

ARNS, T. C. **Modelagem molecular e comportamento dinâmico da molécula imunomoduladora HLA-G e isoformas solúveis.** 2018. 245 f. Tese de Doutorado em Imunologia – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2018.

**Aprovado em:**

Banca Examinadora

Prof. Dr. \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

*Dedico este trabalho à minha mãe Elizete, meu exemplo de vida.*

## AGRADECIMENTOS

Aos meus pais, Edson Arns e Elizete Fronza, pelo esforço, dedicação e compreensão, em todos os momentos desta jornada. Sem vocês, nada disso seria possível.

Ao meu irmão Augusto, meu melhor amigo, pelo incentivo e confiança.

Aos meus avós paternos e maternos, Adalberto (*In memoriam*) e Diva, Faustino (*In memoriam*) e Divanir, por todo amor e ensinamentos.

Ao amigo Carl Janer, por todo apoio e incentivo.

Aos meus amigos Adam Martins e Cássio Gondro, que continuaram sempre presentes em minha vida, apesar da distância.

Ao Prof. Dr. Eduardo Antônio Donadi pela orientação e dedicação durante o desenvolvimento deste trabalho. Agradeço a oportunidade, os ensinamentos e a confiança. Muito obrigada, Professor!

À Prof<sup>ª</sup>. Dr<sup>ª</sup>. Silvana Giuliatti por me receber de braços abertos em seu laboratório e me apresentar o mundo da bioinformática estrutural. Agradeço as conversas, conselhos e a orientação. Muito obrigada, Professora!

À Prof<sup>ª</sup>. Dr<sup>ª</sup>. Lydia Kavradi pela oportunidade de estágio e parceria imprescindível para realização do trabalho. Ao Dr. Dinler Antunes pela ajuda constante e ensinamentos valiosos, parte essencial deste trabalho.

Aos colegas de laboratório, em especial à minha amiga Elvira Tamarozzi.

Ao Programa de Pós-graduação em Imunologia Básica e Aplicada da Faculdade de Medicina de Ribeirão Preto – USP. Agradeço aos professores pela contribuição em minha formação científica e à Secretaria do Departamento, em especial à secretária Ana Cristine Ferreira, por todo apoio durante essa jornada.

À CAPES pela bolsa de Doutorado concedida e ao Santander pela bolsa de estágio na Rice University em Houston, Texas.

Aos membros da banca examinadora pelas contribuições, críticas e sugestões.

A todos aqueles que, mesmo não tendo sido mencionados, contribuíram para a realização deste trabalho.

Muito obrigada.  
Thais Cristine Arns.

*"Lembre-se de olhar para as estrelas e não para baixo, para os seus pés. Tente entender o que você vê e se pergunte sobre o que faz o Universo existir. Seja curioso. E por mais difícil que a vida possa parecer, sempre existirá algo que você pode fazer e ter sucesso. O importante é não desistir."*

*Stephen Hawking*

## RESUMO

ARNS, T. C. **Modelagem molecular e comportamento dinâmico da molécula imunomoduladora HLA-G e isoformas solúveis**. 2018. 245 f. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2018.

HLA-G é uma molécula de histocompatibilidade de classe Ib, considerada tolerogênica, que possui um importante papel na supressão da resposta imune. Existem 58 alelos descritos para o gene *HLA-G* que codificam 18 proteínas e ainda 2 alelos nulos. O RNA mensageiro que codifica a proteína HLA-G pode dar origem a pelo menos 7 isoformas distintas por mecanismos de edição alternativa (*splicing* alternativo), sendo que as funções dessas isoformas são pouco conhecidas, mas poderiam estar envolvidas na modulação da atividade de leucócitos. Considerando que todos os segmentos da molécula HLA-G, bem como suas isoformas possam contribuir para a capacidade supressora e interação com receptores do sistema imune, o conhecimento da estrutura terciária completa dessas proteínas é necessário. O intuito do trabalho foi a obtenção da estrutura completa da proteína HLA-G, incluindo a região transmembrana e cauda citoplasmática, além das estruturas completas das isoformas solúveis HLA-G5, HLA-G6 e HLA-G7, por meio da combinação de metodologias de modelagem molecular por homologia, modelagem molecular *ab initio* e *docking* molecular. Nosso estudo demonstrou a estabilidade estrutural e o padrão de movimentação da molécula HLA-G completa inserida em uma bicamada lipídica. As simulações referentes à isoforma HLA-G5 demonstraram a grande instabilidade das estruturas que não possuem o peptídeo e  $\beta$ 2-microglobulina acoplada. As conformações monoméricas testadas para isoforma HLA-G6 não foram estáveis quando submetidas à dinâmica molecular e talvez HLA-G6 não possa existir como monômero solúvel e precise estar em uma conformação dimérica para ser expressa de forma estável. Para as possibilidades estruturais avaliadas para o monômero da isoforma HLA-G7, foi possível observar que o monômero contendo peptídeo demonstrou ser muito estável e de forma geral, mais estável do que a mesma construção na ausência do peptídeo. Em relação aos dímeros de HLA-G7, um dímero *MHC-like*, contendo o peptídeo, foi a estrutura mais estável gerada. Tais modelos completos poderão ser utilizados para modelagem *in silico* dos demais alelos e isoformas produzidas, além da realização de *screenings* virtuais para busca de possíveis receptores do sistema imune e moléculas que ainda não possuem interação descrita com HLA-G e suas isoformas.

**PALAVRAS-CHAVE:** HLA-G, modelagem molecular, dinâmica molecular, *docking* molecular, bioinformática.

## ABSTRACT

ARNS, T. C. **Molecular modeling and dynamics of immunomodulatory molecule HLA-G and soluble isoforms.** 2018. 245 f. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2018.

HLA-G is a class Ib histocompatibility molecule, considered tolerogenic, that plays an important role in suppressing the immune response. There are 58 alleles described for the *HLA-G* gene encoding 18 proteins and 2 null alleles. The messenger RNA encoding the HLA-G protein may give rise to at least 7 distinct isoforms by alternative splicing mechanisms, and little is known regarding the functions of these isoforms, but they could be involved in the modulation of leukocyte activity. Considering that all segments of the HLA-G molecule, as well as its isoforms may contribute to the suppressive capacity and interaction with receptors of the immune system, knowledge of the complete tertiary structure of these proteins is necessary. The aim of this work was to obtain the complete structure of the HLA-G protein, including the transmembrane and cytoplasmic tail region, as well as the complete structures of the soluble isoforms HLA-G5, HLA-G6 and HLA-G7, by combining homology and *ab initio* molecular modeling and molecular docking. Our study demonstrated the structural stability and movement patterns of the complete HLA-G molecule inserted in a lipid bilayer. The simulations associated to the HLA-G5 isoform demonstrated the great instability of structures lacking the peptide and coupled  $\beta$ 2-microglobulin. The monomeric conformations tested for the HLA-G6 isoform were not stable when subjected to molecular dynamics and perhaps HLA-G6 may not exist as a soluble monomer, requiring a dimeric conformation for stable expression. For the HLA-G7 structural possibilities evaluated, it was possible to observe that the peptide-containing monomer was stable and generally more stable than the same construct in the absence of the peptide. Regarding the HLA-G7 dimers, the MHC-like dimer containing the peptide was the most stable structure generated. Such complete models can be used for *in silico* modeling of the other alleles and isoforms produced, as well as virtual screenings to search for possible receptors of the immune system and molecules that do not yet have an interaction described with HLA-G and its isoforms.

**KEYWORDS:** HLA-G, molecular modeling, molecular dynamics, molecular docking, bioinformatics.

## LISTA DE FIGURAS

<b>Figura 1.</b> Mapa simplificado do MHC humano, conhecido como sistema HLA.....	36
<b>Figura 2.</b> Representação esquemática do receptor de células T.....	37
<b>Figura 3.</b> Estrutura esquemática da molécula HLA de classe I.....	39
<b>Figura 4.</b> Biossíntese de HLA de classe I e ligação de peptídeo antigênico no ER.....	43
<b>Figura 5.</b> Estrutura do gene HLA-G e suas isoformas produzidas por edição alternativa a partir do transcrito primário.....	54
<b>Figura 6.</b> Expressão de HLA-G.....	55
<b>Figura 7.</b> Fluxograma geral das estruturas modeladas para a molécula de HLA-G contidas na tese. ....	70
<b>Figura 8.</b> Fluxograma detalhado das ferramentas utilizadas para obtenção da estrutura completa de HLA-G inserida na bicamada lipídica. Caixas com bordas pretas mostram os processos realizados, caixas com bordas azuis pontilhadas mostram as ferramentas computacionais utilizadas em todas as etapas das análises.....	71
<b>Figura 9.</b> Fluxograma detalhado das ferramentas utilizadas para obtenção da estrutura do dímero de HLA-G1. Caixas com bordas pretas mostram os processos realizados, caixas com bordas azuis pontilhadas mostram as ferramentas computacionais utilizadas em todas as etapas das análises. ....	72
<b>Figura 10.</b> Fluxograma detalhado das ferramentas utilizadas para obtenção das estruturas completas das isoformas HLA-G5, HLA-G6 e HLA-G7. Caixas com bordas pretas mostram os processos realizados, caixas com bordas azuis pontilhadas mostram as ferramentas computacionais utilizadas em todas as etapas das análises. ....	73
<b>Figura 11.</b> (A) Sequência primária completa de HLA-G no formato FASTA. As cores são referentes às porções da proteína compostas por tais aminoácidos, segundo esquema na porção inferior da figura (cinza – peptídeo sinal, azul – domínio $\alpha$ 1, verde - domínio $\alpha$ 2, vermelho -	

domínio  $\alpha 3$ , roxo - região conectora, preto – região transmembrana e amarelo – cauda citoplasmática). (B) Sequência primária da proteína HLA-G madura no formato FASTA.....85

**Figura 12.** Perfil de hidrofobicidade dos resíduos que compõem o peptídeo sinal da proteína HLA-G (SIMOSSIS *et al.*, 2005) .....86

**Figura 13.** Alinhamento múltiplo da sequência primária da proteína codificada pelo alelo *HLA-G\*01:01:01:01* completa e sequências referentes às estruturas cristalográficas PDB\_ID:1YDP (CLEMENTS *et al.*, 2005), PDB\_ID:3KYN e PDB\_ID:3KYO (WALPOLE *et al.*, 2010).....87

**Figura 14.** Alinhamento local da sequência primária madura da proteína codificada pelo alelo *HLA-G\*01:01:01:01* e a sequência primária da estrutura cristalográfica PDB\_ID:1YDP (CLEMENTS *et al.*, 2005). O resíduo em azul (Cys42) indica a localização da cisteína que sofreu mutação na estrutura PDB\_ID:1YDP, enquanto os resíduos em vermelho são provenientes da diferença de cadeia primária do alelo utilizado na estrutura PDB\_ID:1YDP. Os resíduos alinhados foram modelados por homologia.....88

**Figura 15.** Alinhamento global da sequência primária madura da proteína codificada pelo alelo *HLA-G\*01:01:01:01* e a sequência primária da estrutura cristalográfica PDB\_ID:1YDP (CLEMENTS *et al.*, 2005). Os resíduos em vermelho e as porções alinhadas foram modelados por homologia, enquanto a porção terminal em verde (que corresponde ao peptídeo conector localizado no final do domínio  $\alpha 3$ , porção transmembrana e cauda citoplasmática) foi modelada pela combinação de técnicas de modelagem *ab initio* e *threading*.....89

**Figura 16.** Resultados da análise preconizada por Ramachandran para a estrutura da proteína HLA-G. Área em vermelho representa regiões muito favoráveis, em amarelo regiões favoráveis, em amarelo claro regiões pouco favoráveis e em branco regiões desfavoráveis. Os quadrados pretos representam o posicionamento dos aminoácidos no gráfico e os triângulos pretos representam as glicinas. ....92

**Figura 17.** Melhor modelo (modelo 1) para proteína codificada pelo alelo *HLA-G\*01:01:01:01*, gerado por homologia. Visão frontal da molécula (A), lateral (B) e superior do sulco que acomoda o peptídeo a ser apresentado pelo *HLA-G\*01:01:01:01* (C).....93

**Figura 18.** Melhor modelo (modelo 496) da proteína codificada pelo alelo *HLA-G\*01:01:01:01*, gerado por modelagem *ab initio* da porção referente ao peptídeo conector. Visão frontal da molécula (A), lateral (B) e superior do sulco que acomoda o peptídeo a ser

apresentado pela proteína codificada pelo alelo *HLA-G\*01:01:01:01* (C). A porção em amarelo corresponde aos resíduos modelados utilizando metodologia *ab initio*. .....94

**Figura 19.** Modelo da proteína codificada pelo alelo *HLA-G\*01:01:01:01*, gerado por *threading* da porção transmembrana. Visão frontal da molécula (A), lateral (B). A porção em amarelo corresponde aos resíduos que compõem a porção transmembrana. ....95

**Figura 20.** Modelo da proteína codificada pelo alelo *HLA-G\*01:01:01:01*, gerado por modelagem *ab initio* da porção referente à cauda citoplasmática. Visão frontal da molécula (A) e aproximação da porção transmembrana (demonstração da localização do resíduo inicial M - metionina e final W - triptofano) e cauda citoplasmática (demonstração da localização do resíduo inicial R – arginina, e final D - ácido aspártico) (B). A porção em amarelo corresponde aos resíduos modelados utilizando metodologia *ab initio*. ....97

**Figura 21.** Sequência FASTA da proteína  $\beta$ 2-microglobulina. ....98

**Figura 22.** Modelo da proteína codificada pelo alelo *HLA-G\*01:01:01:01* (roxo) acoplada à proteína  $\beta$ 2-microglobulina (rosa). Visão frontal da molécula (A), lateral (B) e aproximação da cadeia composta pela  $\beta$ 2-microglobulina (C) .....99

**Figura 23.** Estrutura tridimensional do nonapeptídeo RIIPRHLQL (cadeia P) contendo os códigos que representam os resíduos (R, arginina; I, isoleucina; P, prolina; H, histidina; L, leucina; Q, glutamina). ....100

**Figura 24.** Visão frontal da proteína codificada pelo alelo *HLA-G\*01:01:01:01* (roxo) completa, acoplada à proteína  $\beta$ 2-microglobulina (rosa) e apresentando no sulco o peptídeo RIIPRHLQL (verde) .....101

**Figura 25.** Sulco da molécula HLA-G e peptídeo RIIPRHLQL acoplado, demonstrando as interações dos resíduos da estrutura HLA-G (rosa claro) e o peptídeo, indicando os resíduos âncora 2 - Isoleucina e 9 - Leucina (pink) .....102

**Figura 26.** Sulco da molécula HLA-G e peptídeo RIIPRHLQL acoplado, demonstrando as interações entre os resíduos da estrutura HLA-G (azul claro) e o peptídeo, indicando os resíduos Y (Tyr, tirosina, posição 84), S (Ser, serina, posição 143), K (Lys, lisina, posição 146), T (Thr, treonina, posição 73), N (Asn, asparagina, posição 77) e T (Thr, treonina, posição 80).....103

<b>Figura 27.</b> Sulco da molécula HLA-G e peptídeo RIIPRHLQL acoplado, demonstrando cadeias laterais (representação em <i>sticks</i> ) dos resíduos P (prolina, posição 4), R (arginina, posição 5) e Q (glutamina, posição 8) .....	103
<b>Figura 28.</b> Em vermelho os resíduos Cys42 (cisteína, posição 42) e CYS147 (cisteína, posição 147), ampliados para melhor visualização da área.....	104
<b>Figura 29.</b> Em vermelho os resíduos M (metionina, posição 76) e Q (glutamina, posição 79), ampliados para melhor visualização da área.....	104
<b>Figura 30.</b> Em vermelho os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197); em azul os resíduos D (ácido aspártico, posição 223), Q (glutamina, posição 224), T (treonina, posição 225), Q (glutamina, posição 226), D (ácido aspártico, posição 227), V (valina, posição 228) e E (ácido glutâmico, posição 229), ampliados para melhor visualização da área.....	105
<b>Figura 31.</b> Complexo da proteína HLA-G completa (roxo), proteína $\beta$ 2-microglobulina (rosa), peptídeo RIIPRHLQL acoplado na fenda (verde) e bicamada lipídica composta de unidades de fosfolipídio DLPA (verde claro). Recorte na bicamada lipídica para expor a porção transmembrana.....	107
<b>Figura 32.</b> Sistema pronto para dinâmica molecular contendo água (azul claro), íons $\text{Na}^+$ e $\text{Cl}^-$ (esferas douradas), proteína HLA-G completa (roxo), proteína $\beta$ 2-microglobulina (rosa), peptídeo RIIPRHLQL acoplado na fenda (verde escuro) e bicamada lipídica composta de unidades de fosfolipídio DLPA (verde claro) .....	108
<b>Figura 33.</b> Visão detalhada da região transmembrana do sistema pronto para dinâmica molecular contendo água (azul claro), íons $\text{Na}^+$ e $\text{Cl}^-$ (esferas douradas), proteína HLA-G completa (roxo) e bicamada lipídica composta de unidades de fosfolipídio DLPA (verde claro).....	109
<b>Figura 34.</b> Gráfico de RMSD do sistema contendo a proteína HLA-G completa, bicamada lipídica, água e íons.....	111
<b>Figura 35.</b> Variação estrutural por resíduo da replicata r3 da proteína HLA-G. Valores de RMSF das coordenadas dos átomos $\text{C}\alpha$ em relação à estrutura inicial.....	112

**Figura 36.** Proteína codificada pelo alelo *HLA-G\*01:01:01:01*, juntamente com a proteína  $\beta$ 2-microglobulina e o peptídeo RIIPRHLQL nos momentos iniciais (branco) e finais (vermelho) da simulação de 100ns. (A) Visão frontal da proteína HLA-G. (B) Visão superior da proteína HLA-G, mostrando em primeiro plano o sulco e o peptídeo RIIPRHLQL inserido no mesmo. (C) Aproximação da porção transmembrana e pequena cauda citoplasmática da proteína HLA-G.....113

**Figura 37.** Proteína codificada pelo alelo *HLA-G\*01:01:01:01*, juntamente com a proteína  $\beta$ 2-microglobulina e o peptídeo RIIPRHLQL nos momentos iniciais (branco), finais (vermelho) e intervalos de 10ns (cinza claro) da simulação de 100ns. (A) Visão frontal da proteína HLA-G. (B) Aproximação da porção transmembrana e pequena cauda citoplasmática da proteína HLA-G, demonstrando o deslocamento máximo de 22,5 Å da estrutura na bicamada lipídica.....113

**Figura 38.** Alinhamento múltiplo das sequências primárias do dímero formado pela proteína codificada pelo alelo *HLA-G\*01:01:01:01* (denominada aqui Dímero\_completo) e estruturas cristalográficas PDB\_ID:1YDP (CLEMENTS *et al.*, 2005) e PDB\_ID: 2D31 (SHIROISHI *et al.*, 2006). Os resíduos em vermelho demarcam resíduos distintos entre a sequência completa do dímero (Dímero\_completo), sua estrutura resolvida (PDB\_ID: 2D31) e o molde referente ao monômero de HLA-G (PDB\_ID:1YDP). Os resíduos não resolvidos na estrutura PDB\_ID:2D31 demarcados em azul e o restante dos resíduos alinhados foram modelados por homologia com a estrutura do monômero de HLA-G (PDB\_ID:1YDP) .....115

**Figura 39.** Alinhamento estrutural do modelo 3 referente ao dímero da proteína codificada pelo alelo *HLA-G\*01:01:01:01* (cinza) e a estrutura experimental PDB\_ID: 2D31 (cinza) (SHIROISHI *et al.*, 2006). As regiões em vermelho demonstram os aminoácidos que foram modelados por homologia com a estrutura experimental PDB\_ID: 1YDP (CLEMENTS *et al.*, 2005), com o intuito de preencher os resíduos faltantes (*gaps*) existentes na cristalografia....116

**Figura 40.** Resultados da análise preconizada por Ramachandran para a estrutura do dímero solúvel de HLA-G1. Área em vermelho representa regiões muito favoráveis, em amarelo regiões favoráveis, em amarelo claro regiões pouco favoráveis e em branco regiões desfavoráveis. Os quadrados pretos representam o posicionamento dos aminoácidos no gráfico e os triângulos pretos representam as glicinas.....117

<b>Figura 41.</b> Em vermelho os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197) relacionados à interação com receptores ILT2 e ILT4; em azul os resíduos D (ácido aspártico, posição 223), Q (glutamina, posição 224), T (treonina, posição 225), Q (glutamina, posição 226), D (ácido aspártico, posição 227), V (valina, posição 228) e E (ácido glutâmico, posição 229), relacionados à interação com CD8. Em amarelo a ponte dissulfeto formada pelas cisteínas localizadas na posição 42 de cada monômero.....	118
<b>Figura 42.</b> Aproximação da ponte dissulfeto (em amarelo) formada pelas cisteínas na posição 42 (Cys42-Cys42), de cada monômero do dímero de HLA-G1 (em vermelho) .....	119
<b>Figura 43.</b> Gráfico de RMSD do dímero solúvel de HLA-G1.....	120
<b>Figura 44.</b> Variação estrutural por resíduo da simulação r1 do dímero de HLA-G1. Valores de RMSF das coordenadas dos átomos C $\alpha$ em relação à estrutura inicial.....	121
<b>Figura 45.</b> Dímero solúvel de HLA-G1 nos momentos iniciais (branco) e finais (vermelho) da simulação de 200ns. Em azul os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197) .....	122
<b>Figura 46.</b> Aproximação da região dos sulcos que acomodam o peptídeo no dímero solúvel de HLA-G1, nos momentos iniciais (branco) e finais (vermelho) da simulação de 200ns.....	122
<b>Figura 47.</b> Dímero solúvel de HLA-G1 nos momentos iniciais (azul), finais (vermelho) e intervalos de 20ns (cinza claro) da simulação de 200ns. Em preto os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197) .....	123
<b>Figura 48.</b> (A) Sequência primária da isoforma HLA-G5 no formato FASTA. As cores são referentes à porção da proteína compostas por tais aminoácidos, segundo esquema na porção inferior da figura (cinza – peptídeo sinal, azul – domínio $\alpha$ 1, verde - domínio $\alpha$ 2, vermelho - domínio $\alpha$ 3, roxo - porção proveniente da tradução do íntron 4). (B) Sequência primária da isoforma HLA-G5 madura no formato FASTA.....	124
<b>Figura 49.</b> Perfil de hidrofobicidade dos resíduos que compõem cauda proveniente da tradução do íntron 4, presente nas isoformas HLA-G5 e HLA-G6 (SIMOSSIS <i>et al.</i> , 2005).....	125
<b>Figura 50.</b> Alinhamento múltiplo das sequências primárias da proteína completa HLA-G (denominada HLA-G_completo), a sequência primária da isoforma HLA-G5 (denominada	

HLA-G5) e a sequência primária do molde PDB\_ID:1YDP (CLEMENTS *et al.*, 2005). Os resíduos em vermelho e os resíduos restantes alinhados foram modelados por homologia, enquanto os resíduos em azul foram modelados por metodologia *ab initio*.....126

**Figura 51.** Melhor modelo (modelo 1) para isoforma HLA-G5, gerada por homologia. Visão frontal da molécula (A), lateral (B) e superior do sulco que acomoda o peptídeo a ser apresentado (C) .....127

**Figura 52.** Melhor modelo obtido para isoforma HLA-G5, gerada pela combinação de modelagem por homologia e *ab initio*. Visão frontal (A) e lateral (B) da molécula. A porção em preto corresponde aos resíduos modelados utilizando metodologia *ab initio*.....128

**Figura 53.** Gráfico de RMSD da isoforma HLA-G5.....129

**Figura 54.** Variação estrutural por resíduo da simulação r1 da isoforma HLA-G5. Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.....130

**Figura 55.** Isoforma HLA-G5 nos momentos iniciais (em branco) e finais (em vermelho) dos 400ns da simulação de dinâmica molecular. (A) Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$ . (B) Visão frontal da isoforma HLA-G5, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.....131

**Figura 56.** Gráfico de RMSD da dinâmica molecular de 300ns referente à isoforma HLA-G5 contendo peptídeo no sulco formado pelas hélices do domínio  $\alpha 1$  e  $\alpha 2$ .....133

**Figura 57.** Variação estrutural por resíduo da simulação r2 da simulação de dinâmica molecular referente à isoforma HLA-G5 que possui o nonapeptídeo RIIPRHLQL acoplado ao sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$ . Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.....134

**Figura 58.** Isoforma HLA-G5 contendo peptídeo no sulco formado pelas hélices do domínio  $\alpha 1$  e  $\alpha 2$  nos momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular. (A) Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  e o peptídeo. (B) Visão frontal da isoforma HLA-G5, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.....135

**Figura 59.** Isoforma HLA-G5 contendo peptídeo no sulco formado pelas hélices do domínio  $\alpha 1$  e  $\alpha 2$  nos momentos iniciais (A, em branco) e finais (B, em vermelho) dos 300ns da simulação de dinâmica molecular. (A) Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  e o peptídeo, demonstrando as interações dos resíduos do peptídeo com o sulco durante os momentos iniciais da simulação de dinâmica molecular. Os resíduos únicos para os momentos iniciais estão numerados. (B) Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  e o peptídeo, demonstrando as interações dos resíduos do peptídeo com o sulco durante os momentos finais da simulação de dinâmica molecular. Os resíduos únicos para os momentos finais estão numerados. (C) Comparação das interações dos resíduos do peptídeo com o sulco nos momentos iniciais (estrutura em branco) e finais (estrutura em vermelho), sem o peptídeo na fenda para melhor observação do assoalho formado por folhas- $\beta$ . .....136

**Figura 60.** Isoforma HLA-G5 contendo peptídeo no sulco formado pelas hélices do domínio  $\alpha 1$  e  $\alpha 2$  nos momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular. Visualização frontal (A, C) e lateral (B) da isoforma HLA-G5, mostrando momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular, onde pode ser observado o peptídeo fora do sulco (em vermelho) .....137

**Figura 61.** Melhor modelo obtido para isoforma HLA-G5 (rosa) com o nonapeptídeo RIIPRHLQL (verde) inserido no sulco e acoplada à proteína  $\beta 2$ -microglobulina (lilás). Visão frontal (A), lateral (B) e aproximação da cadeia composta pela  $\beta 2$ -microglobulina (C) .....139

**Figura 62.** Visualização do melhor modelo obtido para isoforma HLA-G5 (rosa) com o nonapeptídeo RIIPRHLQL (verde) inserido no sulco e acoplada à proteína  $\beta 2$ -microglobulina (lilás). Representação de superfície; visão frontal (A), traseira (B) e sulco do peptídeo (C). Representação de superfície com transparência (60%), possibilitando a visualização da estrutura secundária (em formato *cartoon*) da isoforma HLA-G5 (rosa) com o nonapeptídeo RIIPRHLQL (verde) inserido no sulco e acoplada à proteína  $\beta 2$ -microglobulina (lilás). Visão frontal (D) e sulco do peptídeo (E) .....140

**Figura 63.** Gráfico de RMSD da dinâmica molecular de 200ns referente à isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta 2$ -microglobulina. ....141

<b>Figura 64.</b> Variação estrutural por resíduo da simulação r3 de dinâmica molecular referente à isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína $\beta$ 2-microglobulina. Valores de RMSF das coordenadas dos átomos C $\alpha$ em relação à estrutura inicial.....	142
<b>Figura 65.</b> Isoforma HLA-G5, peptídeo RIIPRHLQL acoplado ao sulco e proteína $\beta$ 2-microglobulina nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular. (A) Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio $\alpha$ 1 e $\alpha$ 2 e o peptídeo. (B) Visão frontal da isoforma HLA-G5, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.....	142
<b>Figura 66.</b> Gráfico de RMSD comparativo das simulações envolvendo a isoforma HLA-G5 em suas 3 alternativas estruturais (HLA-G5 – monômero; HLA-G5 e peptídeo RIIPRHLQL acoplado ao sulco e HLA-G5, peptídeo RIIPRHLQL acoplado ao sulco e proteína $\beta$ 2-microglobulina) .....	143
<b>Figura 67.</b> (A) Sequência primária da isoforma HLA-G6 no formato FASTA. As cores são referentes à porção da proteína compostas por tais aminoácidos, segundo esquema na porção inferior da figura (cinza – peptídeo sinal, azul – domínio $\alpha$ 1, vermelho - domínio $\alpha$ 3, roxo - porção proveniente da tradução do íntron 4). (B) Sequência primária da isoforma madura HLA-G6 no formato FASTA.....	144
<b>Figura 68.</b> Alinhamento múltiplo das sequências primárias da proteína completa HLA-G (denominada HLA-G_completo), a sequência primária da isoforma HLA-G6 (denominada HLA-G6) e a sequência primária do molde PDB_ID:1YDP (CLEMENTS <i>et al.</i> , 2005). Os resíduos em vermelho e resíduos alinhados foram modelados por homologia, enquanto os resíduos em azul foram modelados por metodologia <i>ab initio</i> .....	145
<b>Figura 69.</b> Melhor modelo (modelo 1) para isoforma HLA-G6, gerado por homologia. Visão frontal da molécula (A), lateral (B) e superior do domínio $\alpha$ 1 (C) .....	146
<b>Figura 70.</b> Melhor modelo obtido para isoforma HLA-G6, gerado pela combinação de modelagem por homologia e <i>ab initio</i> . Visão frontal (A) e lateral (B) da molécula. A porção em preto corresponde aos resíduos modelados utilizando metodologia <i>ab initio</i> .....	148

<b>Figura 71.</b> Gráfico de RMSD da dinâmica molecular de 300ns referente ao monômero da isoforma HLA-G6.....	149
<b>Figura 72.</b> Variação estrutural por resíduo da simulação r1 da isoforma HLA-G6. Valores de RMSF das coordenadas dos átomos C $\alpha$ em relação à estrutura inicial.....	150
<b>Figura 73.</b> Isoforma HLA-G6 nos momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular. (A) Visão superior da isoforma HLA-G6, mostrando em primeiro plano o domínio $\alpha$ 1. (B) Visão frontal da isoforma HLA-G6, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.....	151
<b>Figura 74.</b> Melhor modelo obtido da isoforma HLA-G6 (vermelho) com o nonapeptídeo RIIPRHLQL (verde) acoplado ao domínio $\alpha$ 1. Visão frontal (A), lateral (B) e aproximação do domínio $\alpha$ 1 (C) .....	152
<b>Figura 75.</b> Gráfico de RMSD da dinâmica molecular de 300ns referente à isoforma HLA-G6 contendo peptídeo acoplado ao domínio $\alpha$ 1.....	153
<b>Figura 76.</b> Variação estrutural por resíduo da simulação r1 da simulação de dinâmica molecular referente à isoforma HLA-G6 que possui o nonapeptídeo RIIPRHLQL acoplado ao domínio $\alpha$ 1. Valores de RMSF das coordenadas dos átomos C $\alpha$ em relação à estrutura inicial.....	154
<b>Figura 77.</b> Isoforma HLA-G6 contendo peptídeo acoplado ao domínio $\alpha$ 1 momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular. (A) Visão superior da isoforma HLA-G6, mostrando em primeiro plano o domínio $\alpha$ 1 e o peptídeo. (B) Visão frontal da isoforma HLA-G6, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.....	155
<b>Figura 78.</b> Resíduos em vermelho serviram para o direcionamento do <i>docking</i> da proteína $\beta$ 2-microglobulina aos domínios $\alpha$ 1 e $\alpha$ 3 da isoforma HLA-G6 no servidor <i>online</i> HADDOCK.....	157
<b>Figura 79.</b> Melhor modelo obtido para isoforma HLA-G6 (vermelho) contendo nonapeptídeo RIIPRHLQL (verde) e acoplada à proteína $\beta$ 2-microglobulina (azul). Visão frontal (A), lateral (B) e aproximação da cadeia composta pela $\beta$ 2-microglobulina (C) .....	158

**Figura 80.** Demonstração da localização dos resíduos envolvidos na interação HLA-G5 e  $\beta$ 2-microglobulina e HLA-G6 e  $\beta$ 2-microglobulina. (A) Visualização frontal das regiões de interação entre HLA-G6 e  $\beta$ 2-microglobulina. (B) Visualização lateral das regiões de interação entre HLA-G6 e  $\beta$ 2-microglobulina. (C) Visualização lateral das regiões de interação entre HLA-G5 e  $\beta$ 2-microglobulina. (D) Visualização frontal das regiões de interação entre HLA-G5 e  $\beta$ 2-microglobulina. (E) Sobreposição das regiões de interação entre HLA-G5 e  $\beta$ 2-microglobulina (rosa) e HLA-G6 e  $\beta$ 2-microglobulina (verde) .....158

**Figura 81.** Gráfico de RMSD da dinâmica molecular de 60ns referente à isoforma HLA-G6 contendo nonapeptídeo RIIPRHLQL e acoplada à proteína  $\beta$ 2-microglobulina.....160

**Figura 82.** Isoforma HLA-G6 contendo nonapeptídeo RIIPRHLQL e acoplada à proteína  $\beta$ 2-microglobulina nos momentos iniciais (em branco) e finais (em vermelho) dos 60ns da simulação de dinâmica molecular. (A) Visão superior da isoforma HLA-G6, mostrando em primeiro plano o domínio  $\alpha$ 1 e o peptídeo. (B) Visão frontal da isoforma HLA-G6, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.....161

**Figura 83.** Gráfico de RMSD comparativo das simulações envolvendo a isoforma HLA-G6 em suas 3 alternativas estruturais (HLA-G6 – monômero; HLA-G6 e nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha$ 1; HLA-G6 com o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha$ 1 e acoplada à proteína  $\beta$ 2-microglobulina) .....162

**Figura 84.** (A) Sequência primária da isoforma HLA-G7 no formato FASTA. As cores são referentes à porção da proteína compostas por tais aminoácidos, segundo esquema na porção inferior da figura (cinza – peptídeo sinal, azul – domínio  $\alpha$ 1, roxo – 2 resíduos provenientes da tradução do íntron 2). (B) Sequência primária da isoforma madura HLA-G7 no formato FASTA.....163

**Figura 85.** Alinhamento múltiplo das sequências primárias da proteína completa HLA-G (denominada HLA-G\_completo), a sequência primária da isoforma HLA-G7 (denominada G7) e a sequência primária do molde PDB\_ID:1YDP (CLEMENTS *et al.*, 2005). Os resíduos em vermelho e resíduos alinhados foram modelados por homologia, enquanto os resíduos em azul foram modelados por metodologia *ab initio*.....164

<b>Figura 86.</b> Modelo obtido para isoforma HLA-G7, gerado pela combinação de modelagem por homologia e <i>ab initio</i> . Visão superior (A), frontal (B) e lateral (C) da molécula. A porção em amarelo corresponde aos resíduos modelados utilizando metodologia <i>ab initio</i> .....	166
<b>Figura 87.</b> Gráfico de RMSD da isoforma HLA-G7.....	167
<b>Figura 88.</b> Variação estrutural por resíduo da simulação r1 da isoforma HLA-G7. Valores de RMSF das coordenadas dos átomos C $\alpha$ em relação à estrutura inicial.....	168
<b>Figura 89.</b> Isoforma HLA-G7 nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular. (A) Visualização superior da isoforma HLA-G7, mostrando em primeiro plano o domínio $\alpha$ 1. (B) Visualização frontal da isoforma HLA-G7. (C) Visualização lateral da isoforma HLA-G7 demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.....	169
<b>Figura 90.</b> Visualização do melhor modelo obtido para HLA-G7 (azul) contendo peptídeo (verde) acoplado ao domínio $\alpha$ 1. Representação em formato <i>cartoon</i> ; visualização superior (A), frontal (B) e lateral (C) da molécula. Representação de superfície (transparência 60%), possibilitando a visualização da estrutura secundária (em formato <i>cartoon</i> ) da isoforma HLA-G7 (azul) contendo peptídeo (verde) acoplado ao domínio $\alpha$ 1; visualização superior (D) e frontal (E) .....	170
<b>Figura 91.</b> Gráfico de RMSD da dinâmica molecular de 250ns referente à isoforma HLA-G7 contendo peptídeo acoplado ao domínio $\alpha$ 1.....	171
<b>Figura 92.</b> Variação estrutural por resíduo da simulação r1 de dinâmica molecular referente à isoforma HLA-G7 que possui o nonapeptídeo RIIPRHLQL acoplado ao domínio $\alpha$ 1. Valores de RMSF das coordenadas dos átomos C $\alpha$ em relação à estrutura inicial.....	172
<b>Figura 93.</b> Isoforma HLA-G7 e nonapeptídeo RIIPRHLQL acoplado ao domínio $\alpha$ 1 nos momentos iniciais (em branco) e finais (em vermelho) dos 250ns da simulação de dinâmica molecular. (A) Visão superior da isoforma HLA-G7, mostrando em primeiro plano o domínio $\alpha$ 1 e o peptídeo. (B) Visão lateral da isoforma HLA-G7, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.....	173
<b>Figura 94.</b> Isoforma HLA-G7 contendo peptídeo acoplado ao domínio $\alpha$ 1 nos momentos iniciais (A, em branco) e finais (B, em vermelho) dos 250ns da simulação de dinâmica	

molecular. (A) Visão superior da isoforma HLA-G7, mostrando em primeiro plano o domínio  $\alpha 1$  e o peptídeo, demonstrando as interações entre os resíduos do peptídeo e isoforma HLA-G7 durante os momentos iniciais da simulação de dinâmica molecular. Resíduos numerados são únicos para momentos iniciais. (B) Visão superior da isoforma HLA-G7, mostrando em primeiro plano o domínio  $\alpha 1$  e o peptídeo, demonstrando as interações entre os resíduos do peptídeo e isoforma HLA-G7 durante os momentos finais da simulação de dinâmica molecular. Resíduos numerados são únicos para momentos finais. (C) Comparação das interações nos momentos iniciais (estrutura em branco) e finais (estrutura em vermelho) .....174

**Figura 95.** Estrutura do dímero proposto G7-Cys42, após a formação da ponte dissulfeto (amarelo) entre as cisteínas localizadas na posição 42 de cada monômero, posicionados pelo servidor *online* HADDOCK.....176

**Figura 96.** Gráfico de RMSD da dinâmica molecular de 200ns referente ao dímero G7-Cys42.....177

**Figura 97.** Variação estrutural por resíduo da simulação r2 de dinâmica molecular referente ao dímero G7-Cys42. Valores de RMSF das coordenadas dos átomos Ca em relação à estrutura inicial.....178

**Figura 98.** Dímero G7-Cys42 nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular (A) e (B). Ponte dissulfeto em amarelo.....179

**Figura 99.** Demonstração da orientação do dímero solúvel de HLA-G completo (cinza, demonstrado anteriormente) e monômeros da isoforma HLA-G7 contendo peptídeo, formando a estrutura G7 *fit* 2D31 (rosa) .....180

**Figura 100.** Estrutura do dímero alternativo de HLA-G7, G7 *fit* 2D31. A orientação dos monômeros que compõem o dímero encontram-se na orientação oblíqua semelhante ao dímero solúvel de HLA-G, demonstrado anteriormente.....180

**Figura 101.** Estrutura do dímero alternativo de HLA-G7, G7 *fit* 2D31, após a formação da ponte dissulfeto (amarelo) .....181

<b>Figura 102.</b> Gráfico de RMSD da dinâmica molecular de 200ns referente ao dímero G7 <i>fit</i> 2D31 .....	182
<b>Figura 103.</b> Dímero G7 <i>fit</i> 2D31 nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular. Ponte dissulfeto em amarelo.....	183
<b>Figura 104.</b> Variação estrutural por resíduo da simulação r2 da simulação de dinâmica molecular referente ao dímero G7 <i>fit</i> 2D31. Valores de RMSF das coordenadas dos átomos C $\alpha$ em relação à estrutura inicial.....	183
<b>Figura 105.</b> Estrutura do dímero proposto MHC- <i>like</i> , formado utilizando o servidor <i>online</i> ClusPro. (A) Visão superior, (B) lateral e (C) frontal do mesmo.....	185
<b>Figura 106.</b> Estrutura do dímero proposto MHC- <i>like</i> . (A) Visão superior e (B) frontal do mesmo. Tracejado em amarelo representa a distância de 37.9 Å entre as cisteínas na posição 42 (Cys42) .....	185
<b>Figura 107.</b> Gráfico de RMSD da dinâmica molecular de 200ns referente ao dímero MHC- <i>like</i> .....	187
<b>Figura 108.</b> Variação estrutural por resíduo da simulação r1 de dinâmica molecular referente ao dímero MHC- <i>like</i> . Valores de RMSF das coordenadas dos átomos C $\alpha$ em relação à estrutura inicial.....	188
<b>Figura 109.</b> Dímero MHC- <i>like</i> nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular (A, B e C) .....	189
<b>Figura 110.</b> Estrutura do dímero proposto MHC- <i>like</i> contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado. (A) Visão superior, (B) frontal e (C) lateral do mesmo.....	190
<b>Figura 111.</b> Gráfico de RMSD da dinâmica molecular de 300ns referente ao dímero MHC- <i>like</i> contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado.....	191

<b>Figura 112.</b> Variação estrutural por resíduo da simulação r1 de dinâmica molecular referente ao dímero <i>MHC-like</i> contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado. Valores de RMSF das coordenadas dos átomos C $\alpha$ em relação à estrutura inicial.....	192
<b>Figura 113.</b> Dímero <i>MHC-like</i> contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado nos momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular (A, B e C) .....	193
<b>Figura 114.</b> Gráfico de RMSD comparativo das simulações envolvendo os dímeros de HLA-G7 em suas 4 alternativas estruturais (dímero G7-Cys42; dímero G7 <i>fit</i> 2D31; dímero <i>MHC-like</i> e dímero <i>MHC-like</i> contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado).....	194

## LISTA DE TABELAS

<b>Tabela 1.</b> Possíveis moldes para modelagem por homologia da proteína HLA-G.....	86
<b>Tabela 2.</b> Melhores modelos provenientes da metodologia de modelagem ab initio para a isoforma HLA-G5. ....	128
<b>Tabela 3.</b> Melhores modelos provenientes da metodologia de modelagem ab initio para a isoforma HLA-G6. ....	147
<b>Tabela 4.</b> Melhores modelos provenientes da metodologia de modelagem ab initio para a isoforma HLA-G7. ....	165

## SUMÁRIO

1. INTRODUÇÃO .....	35
1.1 Características estruturais das moléculas de histocompatibilidade .....	39
1.2 Processamento de antígenos para apresentação por moléculas de HLA de classe I..	41
1.3 Proteólise e introdução de antígenos no citosol.....	44
1.4 Arquitetura do TCR $\alpha\beta$ .....	45
1.5 HLA-G: molécula de histocompatibilidade de classe I não clássica .....	47
1.5.1 Funções imunomodulatórias da molécula HLA-G.....	48
1.5.1.1 Inibição de células efectoras .....	49
1.5.1.2 Modulação de células reguladoras/supressoras.....	51
1.5.1.3. Outras funções imunológicas de HLA-G.....	52
1.6 Edição alternativa do gene HLA-G.....	52
1.7 Papel da molécula HLA-G em situações fisiológicas e na doença.....	54
1.7.1 HLA-G e gravidez .....	56
1.7.2 HLA-G e tumores.....	57
1.7.3 HLA-G e transplantes.....	59
1.7.4 HLA-G e doenças autoimunes .....	60
1.7.5 HLA-G e doenças infecciosas .....	62
1.8 Obtenção experimental de estruturas e bioinformática estrutural.....	63
2 HIPÓTESE .....	68
3. OBJETIVOS.....	69
3.1 Objetivo geral .....	69
3.2 Objetivos específicos .....	69
4 FLUXOGRAMAS DAS ANÁLISES <i>IN SILICO</i> .....	70
5 MATERIAL E MÉTODOS.....	74
5.1 Obtenção das sequências .....	74
5.2 Identificação da sequência referente ao peptídeo sinal.....	74
5.3 Perfil de hidrofobicidade do peptídeo sinal de HLA-G .....	74
5.4 Alinhamento de sequências proteicas .....	75
5.5 Modelagem <i>in silico</i> por homologia .....	75
5.6 Avaliação da qualidade dos modelos gerados .....	76
5.7 Modelagem <i>in silico</i> por metodologia <i>ab initio</i> .....	77

5.8	Modelagem <i>in silico</i> por metodologia de <i>threading</i> .....	78
5.9	Modelagem <i>in silico</i> por metodologia <i>ab initio</i> das isoformas HLA-G5, HLA-G6 e HLA-G7 .....	78
5.10	Seleção dos modelos gerados por modelagem <i>ab initio</i> .....	79
5.11	Inserção da estrutura completa de HLA-G na bicamada lipídica .....	79
5.12	<i>Docking</i> do peptídeo RIIPRHLQL no sulco da proteína HLA-G e isoformas HLA-G5, HLA-G6 e HLA-G7 .....	80
5.13	<i>Docking</i> proteína-proteína para formação dos dímeros de HLA-G7 .....	80
5.14	Ajuste induzido da isoforma HLA-G7 .....	81
5.15	Simulações de Dinâmica Molecular .....	81
5.16	Resíduos presentes na interface de interação .....	82
5.17	Visualização dos Modelos .....	83
6.	RESULTADOS .....	84
6.1	Molécula de histocompatibilidade de classe I não clássica HLA-G .....	84
6.1.1	Análise da sequência primária da proteína HLA-G .....	84
6.1.2	Modelagem molecular da proteína HLA-G .....	86
6.1.2.1	Escolha do molde .....	86
6.1.3	Modelagem molecular por homologia .....	89
6.1.3.1	Avaliação e seleção dos modelos .....	90
6.1.4	Modelagem molecular <i>ab initio</i> .....	93
6.1.4.1	Seleção dos modelos .....	94
6.1.5	Modelagem <i>in silico</i> por metodologia de <i>threading</i> .....	95
6.1.6	Modelagem molecular <i>ab initio</i> da cauda citoplasmática .....	96
6.1.6.1	Seleção dos modelos .....	96
6.1.7	Modelagem da proteína $\beta$ 2-microglobulina acoplada à proteína HLA-G completa 97	
6.1.7.1	Avaliação e seleção dos modelos .....	98
6.1.8	<i>Docking</i> do peptídeo ao sulco da proteína HLA-G completa .....	99
6.1.9	Inserção da estrutura completa de HLA-G em bicamada lipídica .....	106
6.1.10	Dinâmica molecular da proteína HLA-G completa .....	109
6.2	Conformação do dímero HLA-G1 solúvel .....	114
6.2.1	Modelagem por homologia dos resíduos faltantes do dímero de HLA-G1 .....	114
6.2.1.1	Avaliação dos modelos .....	115
6.2.2	Formação das pontes dissulfeto .....	118

6.2.3	Dinâmica Molecular.....	119
6.3	Isoforma solúvel HLA-G5 .....	123
6.3.1	Obtenção da sequência.....	123
6.3.2	Modelagem homologia do monômero da isoforma HLA-G5.....	125
6.3.2.1	Avaliação dos modelos gerados por homologia.....	126
6.3.3	Modelagem <i>ab initio</i> .....	127
6.3.3.1	Seleção dos modelos .....	128
6.3.4	Dinâmica molecular .....	129
6.4	Isoforma solúvel HLA-G5 contendo peptídeo acoplado no sulco.....	131
6.4.1	Seleção do <i>docking</i> .....	131
6.4.2	Dinâmica Molecular.....	132
6.5	Isoforma solúvel HLA-G5 contendo peptídeo acoplado ao sulco e acoplada à $\beta$ 2-microglobulina .....	137
6.5.1	Modelagem da proteína $\beta$ 2-microglobulina acoplada à isoforma HLA-G5 .....	137
6.5.1.1	Avaliação e seleção dos modelos .....	138
6.5.2	Dinâmica Molecular.....	140
6.6	Isoforma solúvel HLA-G6 .....	143
6.6.1	Obtenção da sequência.....	143
6.6.2	Modelagem homologia do monômero da isoforma HLA-G6.....	145
6.6.2.1	Avaliação dos modelos gerados por homologia.....	145
6.6.3	Modelagem <i>ab initio</i> .....	147
6.6.3.1	Seleção dos modelos .....	147
6.6.4	Dinâmica molecular .....	148
6.7	Isoforma solúvel HLA-G6 contendo peptídeo acoplado ao domínio $\alpha$ 1 .....	151
6.7.1	Seleção do <i>docking</i> .....	151
6.7.2	Dinâmica Molecular.....	152
6.8	Isoforma solúvel HLA-G6 contendo peptídeo acoplado ao domínio $\alpha$ 1 e acoplada à $\beta$ 2-microglobulina.....	155
6.8.1	<i>Docking</i> da proteína $\beta$ 2-microglobulina à isoforma HLA-G6 .....	155
6.8.2	Dinâmica Molecular.....	159
6.9	Isoforma solúvel HLA-G7 .....	162
6.9.1	Obtenção da sequência.....	162
6.9.2	Modelagem por homologia do monômero da isoforma HLA-G7 .....	163
6.9.2.1	Avaliação dos modelos gerados por homologia.....	164

6.9.3	Modelagem <i>ab initio</i> .....	165
6.9.3.1	Seleção dos modelos .....	165
6.9.4	Dinâmica molecular .....	166
6.10	Isoforma solúvel HLA-G7 contendo peptídeo acoplado ao domínio $\alpha 1$ .....	169
6.10.1	Seleção do <i>docking</i> .....	169
6.10.2	Dinâmica Molecular .....	170
6.11	Dímero da isoforma solúvel HLA-G7 (G7-Cys42) .....	174
6.11.1	Construção do dímero .....	174
6.11.2	Dinâmica Molecular .....	175
6.12	Dímero da isoforma solúvel HLA-G7 (G7 fit 2D31).....	179
6.12.1	Construção do dímero .....	179
6.12.2	Dinâmica Molecular .....	181
6.13	Isoforma solúvel HLA-G7 (G7 MHC-like) .....	184
6.13.1	Formação do dímero.....	184
6.13.2	Dinâmica Molecular .....	186
6.14	Isoforma solúvel HLA-G7 (G7 MHC-like contendo nonapeptídeo RIIPRHLQL) .	189
6.14.1	Seleção do <i>docking</i> .....	189
6.14.2	Dinâmica Molecular .....	190
7	DISCUSSÃO .....	195
7.1	Molécula de histocompatibilidade de classe I não clássica HLA-G.....	195
7.2	Dímero solúvel da isoforma HLA-G1 .....	202
7.3	Isoforma solúvel HLA-G5 .....	206
7.3.1	Monômero de HLA-G5 .....	208
7.3.2	HLA-G5 e peptídeo RIIPRHLQL acoplado ao sulco .....	209
7.3.3	HLA-G5 contendo peptídeo RIIPRHLQL acoplado ao sulco e acoplada à proteína $\beta 2$ -microglobulina.....	210
7.4	Isoforma solúvel HLA-G6 .....	212
7.4.1	Monômero de HLA-G6 .....	213
7.4.2	Isoforma solúvel HLA-G6 contendo peptídeo acoplado ao domínio $\alpha 1$ .....	214
7.4.3	Isoforma solúvel HLA-G6 contendo peptídeo acoplado ao domínio $\alpha 1$ e acoplada à $\beta 2$ -microglobulina .....	215
7.5	Isoforma solúvel HLA-G7 .....	217
7.5.1	Monômero de HLA-G7 .....	217
7.5.2	Monômero de HLA-G7 acoplado ao peptídeo RIIPRHLQL .....	218

7.6	Isoforma solúvel HLA-G7: possíveis dímeros formados .....	219
7.6.1	Dímero da isoforma solúvel HLA-G7 (G7-Cys42).....	220
7.6.2	Dímero alternativo da isoforma solúvel HLA-G7 (G7 <i>fit</i> 2D31).....	221
7.6.3	Dímero da isoforma solúvel HLA-G7 (G7 MHC- <i>like</i> ).....	222
7.6.4	Dímero da isoforma solúvel HLA-G7 (G7 MHC- <i>like</i> contendo nonapeptídeo RIIPRHLQL) .....	223
8	RESUMO DOS RESULTADOS .....	225
9	CONCLUSÃO.....	227
10	REFERÊNCIAS BIBLIOGRÁFICAS .....	228
11	ANEXOS .....	243

## LISTA DE SÍMBOLOS

*APC: Antigen-presenting cell*

*ATP: Adenosine triphosphate*

*CCD: Cyclic coordinate descent*

*CD160: Cluster of differentiation 160*

*CDRs: Complementarity-determining regions*

*CNX: Calnexin*

*CRT: Calreticulin*

*CTL: Cytotoxic T cell*

*CYS: Cisteína*

*DC: Dendritic cells*

*DLPA: 1,2-Dilauroyl-sn-glycero-3-phosphate*

*DM: Dinâmica molecular*

*DNA: Deoxyribonucleic acid*

*DRiP: Defective ribosomal products*

*ELISA: Enzyme-linked immunosorbent assay*

*ER: Endoplasmic reticulum*

*ERAAP: Endoplasmic reticulum aminopeptidase 1*

*ERp57: Protein disulfide-isomerase like*

*Fc: Fragment crystallizable region*

*FIV: Fertilização in vitro*

*HIV: Human immunodeficiency virus*

*HLA: Human leukocyte antigen*

*HLA-G: Human leukocyte antigen, G*

*HLA-G5: Isoforma HLA-G5*

*HLA-G6: Isoforma HLA-G6*

*HLA-G7: Isoforma HLA-G7*

*sHLA-G: Shed HLA-G*

*HLA-E: Human leukocyte antigen, E*

*IFN- $\gamma$ : Interferon, gamma*

*Ig: Immunoglobulin*

*IL-10: Interleukin 10*

*ILT2: Leukocyte immunoglobulin like receptor B1*

*ILT4: Leukocyte immunoglobulin like receptor B2*

*IMGT/HLA: Immunogenetics / Human leukocyte antigen database*

*ITIM: Immunoreceptor tyrosine-based inhibition motif*

*K: graus Kelvin*

*KIR2DL4: Killer cell immunoglobulin like receptor, two Ig domains and long cytoplasmic tail*

*4*

*MHC: Major histocompatibility complex*

*mRNA: RNA mensageiro*

*NBD: Nucleotide binding domain*

*NK: Natural killer*

*NKG2: Killer cell lectin like receptor C1*

*NKp46: Natural cytotoxicity triggering receptor 1*

*NOTS: Night owls time-sharing service*

*NT2-N: Neuron-like cells*

*PBMCs: Peripheral blood mononuclear cells*

*PDB: Protein data bank*

*PLC: Peptide loading complex.*

*RMN: ressonância magnética nuclear*

*RMSD: Root mean square deviation (Desvio da media quadrática)*

*RMSF: Root mean square fluctuations (Flutuação da media quadrática)*

*RNA: Ribonucleic acid*

*SHP-1: Src homology 2 domain containing phosphatase 1*

*TAP: Transporter associated with antigen processing*

*TCR: T cell receptor*

*TNF: Tumor necrosis factor*

*Treg: Regulatory T cells*

*Tr1: Type 1 regulatory T cells*

*Å: Ångström*

*ns: Nanosegundo*

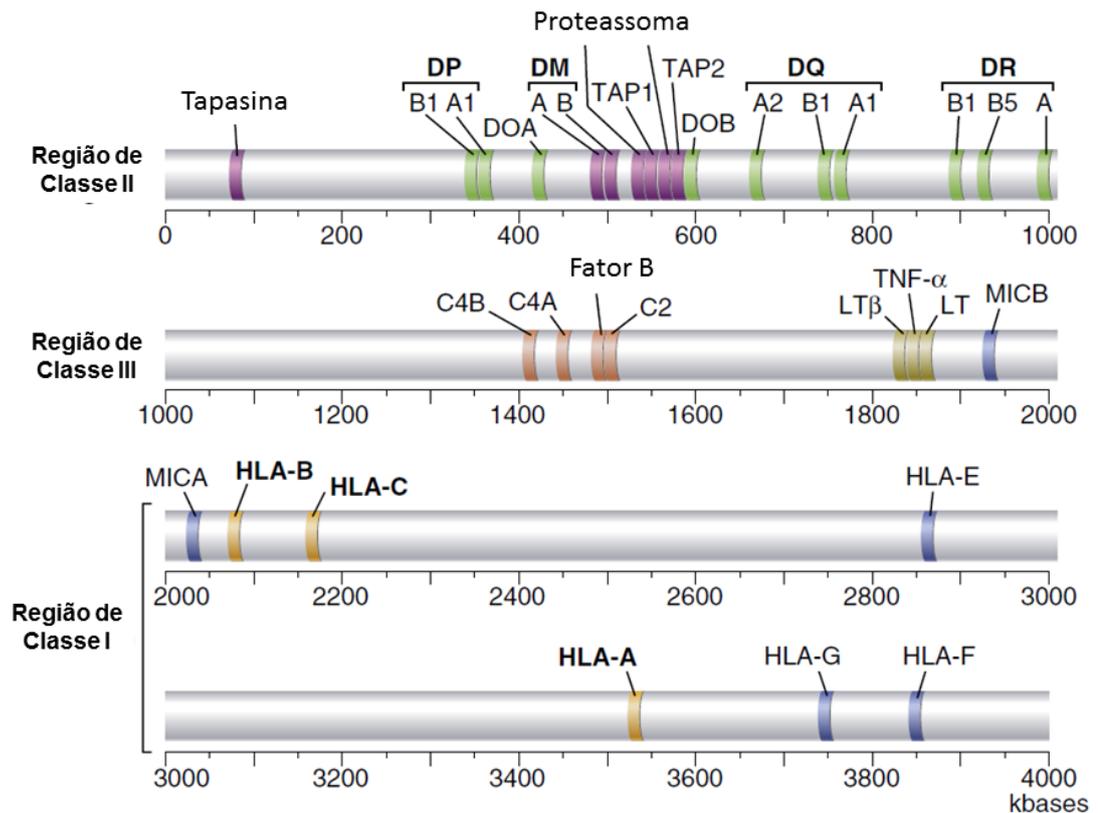
*ps: Picosegundo*

## 1. INTRODUÇÃO

O sistema imune está envolvido em duas principais linhas de defesa utilizadas por organismos superiores para o combate aos patógenos: a imunidade humoral e a imunidade celular. Enquanto os anticorpos provenientes da imunidade humoral reconhecem antígenos intactos, as células T que compõem a imunidade celular distinguem moléculas e estruturas próprias de substâncias exógenas por meio da apresentação de fragmentos de antígeno pelas moléculas do complexo de principal de histocompatibilidade (MHC) presentes nas superfícies celulares (RUDOLPH *et al.* 2006, P J BJORKMAN; PARHAM 1990).

Todos os vertebrados possuem moléculas de MHC, codificadas em uma grande região cromossômica com muitos genes conservados, sendo que alguns desses codificam produtos essenciais para a função das moléculas MHC de classe I (MHC-I) e MHC de classe II (MHC-II) (COHEN *et al.* 2009).

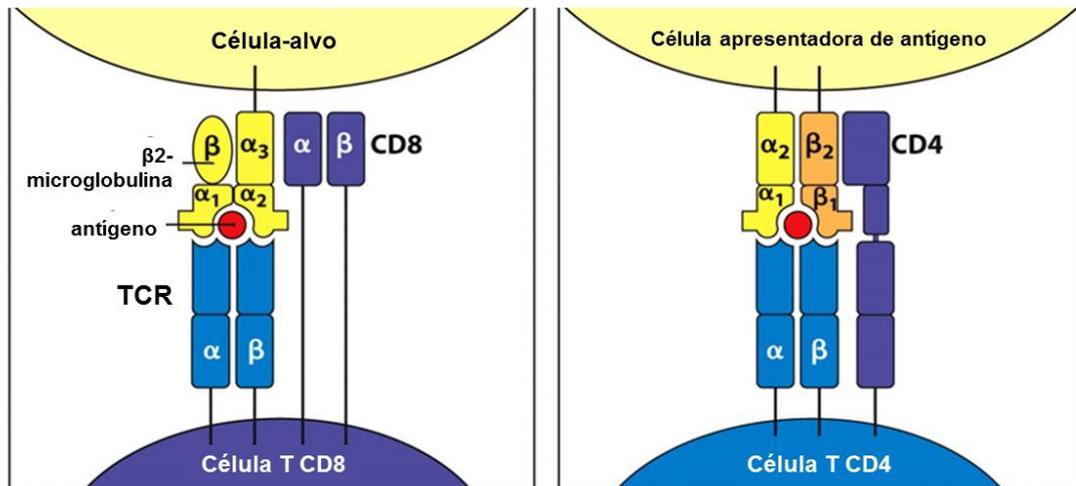
Os genes que codificam as moléculas de MHC classe I e classe II clássicas, de evolução mais recente, estão localizados dentro da região do MHC no cromossomo 6 em humanos e no cromossomo 17 em camundongos. Essas proteínas altamente polimórficas de classe I e classe II possuem a função de apresentação de peptídeos de tamanho variável para células T (P J BJORKMAN; PARHAM 1990).



**Figura 1. Mapa simplificado do MHC humano, conhecido como sistema HLA.** Além dos genes clássicos de HLA de classe I (*HLA-A*, *-B* e *-C*) e de classe II (*HLA-DR*, *DQ* e *DP*) que têm a função de apresentação de antígenos, os genes não clássicos de classe I (*HLA-E*, *HLA-F*, *HLA-G*) e os genes *MIC* codificam moléculas que modulam a resposta imune. Outros genes importantes para resposta imune incluem genes que codificam algumas proteínas do complemento (*C4*, *C2* e *Fator B*), citocinas (*LT $\alpha$* , *LT $\beta$*  e *TNF*), tapasina, *HLA-DM*, *HLA-DO*, *TAP* e proteínas do proteassoma. Fonte: Modificado de ABBAS, A. 2015.

A interação central na imunidade adaptativa mediada por células é entre o receptor de antígeno localizado na superfície de células T (TCR) e a molécula apresentadora de antígeno (MHC) carregada com um determinado antígeno (peptídeo). A maioria (aproximadamente 95%) das células T são células T  $\alpha\beta$  (alfa/beta) com TCR composto por duas cadeias de glicoproteínas chamadas cadeias  $\alpha$  e  $\beta$ . O TCR $\alpha\beta$  reconhece o complexo antígeno-MHC, e dada a diversidade inerente desse sistema, muitos eventos de reconhecimento do TCR $\alpha\beta$  controlam a imunidade das células T. Após o reconhecimento produtivo da molécula TCR $\alpha\beta$  e do complexo antígeno-MHC, o TCR $\alpha\beta$  transmite sinais para o complexo CD3, que desencadeia, posteriormente, cascatas de sinalização intracelular que levam à ativação da resposta imune (van der MERWE; DUSHEK 2011). O evento de reconhecimento do TCR $\alpha\beta$  é central para a imunidade protetora contra patógenos, mas também protege contra a reatividade aberrante das células T (BHARADWAJ *et al.* 2012, YIN *et al.* 2012).

A resposta imune adaptativa de células T evoluiu para reconhecer os produtos da proteólise intracelular parcial, com as células T CD8<sup>+</sup> reconhecendo os peptídeos derivados de proteínas em associação com as moléculas MHC-I, enquanto as células T CD4<sup>+</sup> reconhecem os peptídeos ligados às moléculas MHC-II (COHEN *et al.* 2009).



**Figura 2. Representação esquemática do receptor de células T.** O receptor de antígenos presente na maioria das células T, o receptor de células T  $\alpha\beta$  é composto de duas cadeias de proteínas, TCR $\alpha$  e TCR $\beta$ . Os receptores das células T são sempre ligados à membrana. Os receptores de células T  $\alpha\beta$  não reconhecem o antígeno no seu estado nativo, mas reconhecem um ligante composto de antígeno peptídico ligado à molécula de MHC. Cada molécula de MHC liga-se a grande variedade de peptídeos diferentes, mas as diferentes variantes reconhecem preferencialmente conjuntos de peptídeos com sequência e características físicas específicas. O antígeno peptídico é gerado intracelularmente e é ligado de forma estável na fenda de ligação de peptídeo na superfície da molécula de MHC. Existem duas classes de moléculas de MHC, e essas estão ligadas em seus domínios não polimórficos por moléculas CD8 e CD4 que distinguem duas classes funcionais diferentes de células T  $\alpha\beta$ , CD4 e CD8. CD8 liga moléculas MHC de classe I e pode se ligar simultaneamente ao mesmo complexo peptídeo:MHC-I que é reconhecido por receptor de células T, assim atuando como correceptor e aumentando a resposta das células T; CD4 liga moléculas de MHC de classe II e atua como correceptor para receptores de células T que reconhecem peptídeo:MHC-II. O receptor de células T interage diretamente com o peptídeo antigênico e com a molécula do MHC que o exhibe, e essa especificidade dual está ligada à restrição do MHC e respostas de célula T. Fonte: Modificado de MURPHY, K. 2012.

Os seres humanos possuem três genes que codificam moléculas MHC-I clássicas dentro do complexo antígeno leucocitário humano (HLA), denominados *HLA-A*, *HLA-B* e *HLA-C*, e existem três moléculas MHC-II clássicas, chamadas *HLA-DR*, *HLA-DQ* e *HLA-DP*. Existem outros genes de classe I presentes no MHC de camundongos e humanos, conhecidos como genes de classe Ib não clássicos (*HLA-G*, *HLA-E* e *HLA-F*) (ADAMS; LUOMA 2013). Como o foco do presente trabalho é a molécula de classe I não clássica *HLA-G*, os temas tratados na introdução serão relacionados e aprofundados para as moléculas de HLA de classe I.

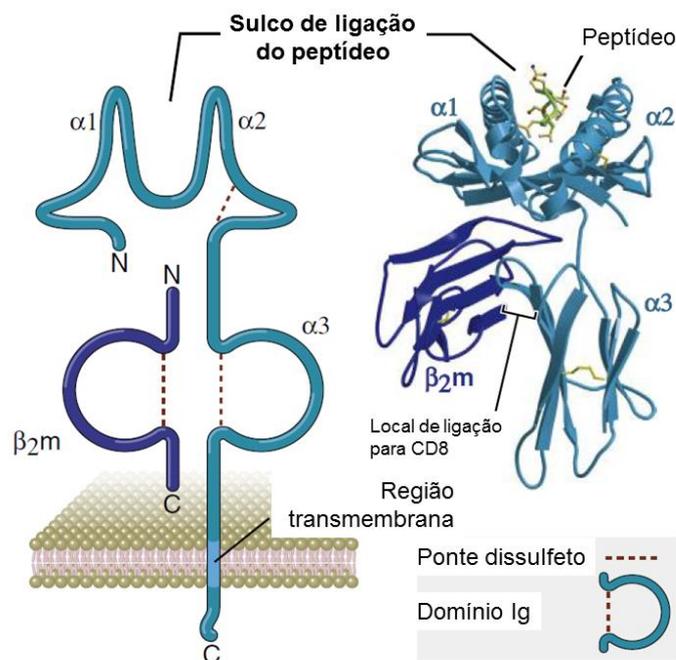
As regiões codificadoras dos genes *HLA-A/B/C* são bastante polimórficas, codificando milhares de variantes proteicas, caracterizando o alto grau de variabilidade genética dessas regiões, no entanto, as regiões controladoras desses genes não apresentam variabilidade proporcionalmente tão alta. Por outro lado, os genes que codificam as moléculas não clássicas *HLA-E/F/G* apresentam restrita variabilidade genética na região codificadora, no entanto, apresentam grande variabilidade genética nas regiões controladoras desses genes (TOMASEC *et al.* 2000; DONADI *et al.* 2011). Levando em conta as principais funções dessas moléculas, podemos considerar que as forças de seleção agiram a favor da variabilidade genética nas regiões codificadoras dos genes clássicos, permitindo grande diversidade de apresentação de antígenos. Por outro lado, considerando que as moléculas não clássicas atuam pobremente como apresentadoras de antígenos, mas robustamente na modulação da resposta imune, forças de seleção agiram contra a variabilidade genética nas regiões codificadoras e a favor da variabilidade nas regiões controladoras dos genes (CASTELLI *et al.*, 2011).

As moléculas de classe I clássicas (HLA-A, HLA-B e HLA-C) e não clássicas (HLA-G, HLA-E e HLA-F) apresentam homologia molecular e estrutural, no entanto, suas funções, padrões de expressão e níveis de polimorfismos são distintos. Essas moléculas são heterodímeros formados por uma cadeia pesada polimórfica  $\alpha$ , codificada por gene do MHC, e uma cadeia leve, não polimórfica ( $\beta_2$ -microglobulina), codificada por gene localizado no cromossomo 15. Todas essas moléculas apresentam sulco que acomoda peptídeos, mas as clássicas têm função bem reconhecida de apresentação de peptídeos aos linfócitos TCD8<sup>+</sup>, enquanto que as não clássicas têm função reconhecida de modulação da resposta imune. As moléculas não clássicas apresentam expressão restrita a alguns tipos celulares, enquanto as clássicas se expressam constitutivamente na membrana celular de todas as células nucleadas (FISCHER *et al.*, 2001). Os representantes pouco polimórficos, denominados "HLAs não clássicos", para aqueles codificados por genes localizados na região do MHC e moléculas "MHC-like" (semelhantes às moléculas MHC) para aqueles que estão fora da região do MHC, existem para proteínas de classe I e classe II, sendo as mais diversas relacionadas às moléculas de classe I. A maioria dessas proteínas não clássicas de classe I possuem funções diferentes da apresentação de peptídeos, sendo essas funções imunes ou não relacionadas à imunidade. Embora as proteínas não clássicas sejam codificadas dentro do MHC, os representantes "MHC-like" são distribuídos entre os cromossomos 1, 2, 7 e 19 em humanos e 3, 5, 7, 10 e 20 em camundongos. A evolução dos genes que codificam essas proteínas tem sido dinâmica, em relação aos cromossomos e *loci*, com diversidade proteica estrutural adaptada aos muitos papéis

que essas proteínas desempenham no organismo hospedeiro (P J BJORKMAN; PARHAM, 1990).

### 1.1 Características estruturais das moléculas de histocompatibilidade

Para entender como os TCRs podem reconhecer antígenos, primeiro é necessário determinar as estruturas das moléculas apresentadoras de antígenos. Essas moléculas de histocompatibilidade ligam peptídeos, sendo a fenda de ligação de antígenos das moléculas de HLA clássicas de classe I mais restrita nas porções N e C-terminal em comparação com as moléculas de classe II, que possuem as extremidades abertas (ADAMS; LUOMA 2013).



**Figura 3. Estrutura esquemática da molécula HLA de classe I.** À esquerda, as diferentes regiões da molécula HLA (não desenhado em escala). As moléculas de classe I são compostas por uma cadeia polimórfica  $\alpha$  ligada à  $\beta 2$ -microglobulina de forma não covalente. O diagrama (à direita) mostra a resolução estrutural por cristalografia da porção extracelular da molécula HLA de classe I com um peptídeo ligado na fenda. Fonte: Modificado de ABBAS, A. 2015.

As moléculas HLA de classe I clássicas apresentam fragmentos peptídicos de 8-10 aminoácidos (ADAMS; LUOMA 2013). O polimorfismo dessas moléculas é concentrado em torno da fenda de ligação de antígenos e determina como e quais peptídeos se ligam a qualquer molécula de HLA. Dentro dos complexos peptídeo:HLA, as regiões do peptídeo que estão

orientadas para fora e expostas ao microambiente podem entrar em contato direto com o TCR, enquanto que os resíduos enterrados na fenda podem afetar indiretamente a ligação com o TCR (THEODOSSIS, *et al.*, 2010). As moléculas não clássicas de HLA de classe I apresentam repertório mais restrito de peptídeos e desempenham papel fundamental na imunidade inata, embora possam mediar as respostas das células T (RODGERS *et al.*, 2005).

Os genes *HLA* de classe I clássicos exibem enorme polimorfismo alélico, e a variação da sequência de aminoácidos está fortemente concentrada na porção da estrutura que interage com os peptídeos, permitindo que diferentes alelos se liguem a uma variedade diferente de peptídeos. O sulco de ligação ao peptídeo é formado por duas hélices  $\alpha$  antiparalelas e o assoalho é formado por 8 folhas  $\beta$ . No caso do HLA de classe I, o sulco corresponde à sequência de aminoácidos contínua formada pela cadeia pesada e a região proximal da membrana é constituída por dois domínios conservados que são homólogos aos domínios da região constante de imunoglobulina (Ig). Para o HLA de classe I, uma porção é fornecida pela cadeia pesada e a outra é uma proteína separada, a  $\beta$ 2-microglobulina, produto de um gene não ligado ao HLA. A cadeia pesada para HLA de classe I é caracterizada como glicoproteína transmembrana com domínio citoplasmático curto. A molécula HLA de classe I possui uma plataforma estruturalmente homóloga capaz de unir peptídeos com afinidade muito alta que podem interagir com o receptor de células T, sendo o peptídeo confinado no sulco por ligações nos terminais N e C (RUDENSKY *et al.*, 1990; FALK, *et al.*, 1991; VAN BLEEK *et al.*, 1991; BLUM *et al.*, 2013). A estrutura das moléculas clássicas de HLA de classe I está bem adaptada para apresentar peptídeos; esses pequenos fragmentos de proteínas (próprias e exógenas) são parte integrante das proteínas HLA clássicas estáveis. Cada molécula de HLA pode formar complexos estáveis com centenas ou milhares de peptídeos diferentes, compartilhando apenas alguns resíduos-âncora conservados que são acomodados no sulco do HLA. A composição química e estrutural do sulco e de suas bolsas difere para cada molécula de HLA e determina o repertório de peptídeos que cada molécula pode ligar. A superfície formada pela molécula de HLA e peptídeo é analisada por células T  $\alpha\beta$  através do seu TCR. Complexos formados por moléculas de HLA contendo peptídeos não-próprios podem iniciar respostas de células T que levam à rejeição de transplantes e eliminação de células tumorais, respostas antivirais, além do recrutamento e modulação de outras células imunes (ADAMS; LUOMA 2013).

## 1.2 Processamento de antígenos para apresentação por moléculas de HLA de classe I

Os peptídeos apresentados pelas moléculas de HLA são os produtos da proteólise intracelular, e existem dois sistemas proteolíticos principais operando dentro da célula que contribuem para o reconhecimento de células T dependentes de HLA. No citosol, a maior parte da proteólise é mediada pelo proteassoma, que é uma estrutura em forma de barril que consiste em quatro anéis empilhados de sete subunidades cada. Os anéis externos são compostos por subunidades  $\alpha$  e os do meio são duas subunidades  $\beta$ , três das quais,  $\beta 1$ ,  $\beta 2$  e  $\beta 5$ , constituem os componentes proteolíticos ativos. O núcleo (20S) é amparado em cada extremidade por um complexo multisubunidade adicional que reconhece as proteínas conjugadas com ubiquitina marcadas para degradação. O componente 19S possui atividade proteolítica que permite o desmonte das estruturas proteicas, facilitando a entrada no canal no centro do barril onde os sítios ativos da subunidade  $\beta$  residem. A função de desmonte, em particular, exige que a proteólise pelo proteassoma seja dependente de ATP. Os produtos finais de proteólise pelo proteassoma formam a fonte dominante de peptídeos para ligação na fenda de HLA de classe I (ADAMS; LUOMA 2013).

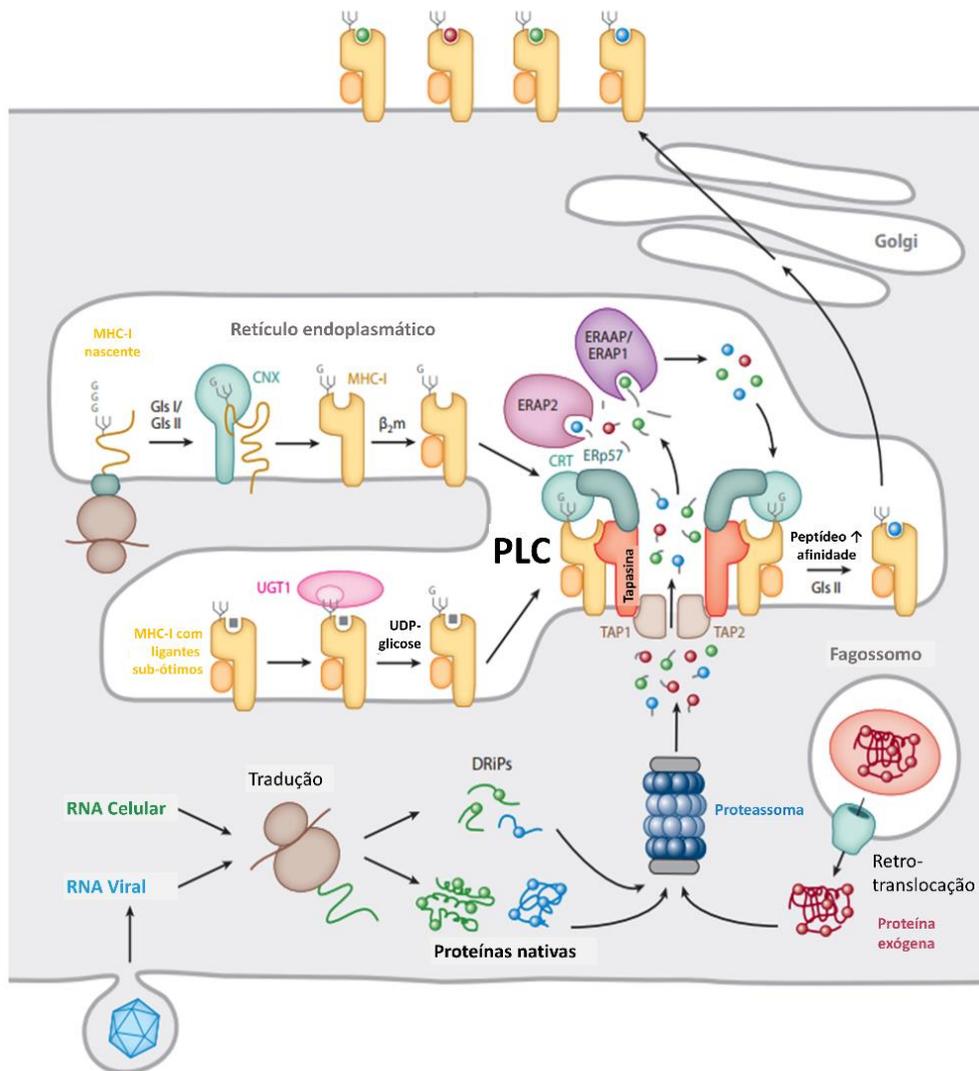
As proteínas provenientes de fontes exógenas que são internalizadas por uma célula são degradadas por proteólise lisossômica. Em resumo, as proteínas endocitadas entram em uma via vesicular que consiste de compartimentos progressivamente mais ácidos e proteoliticamente ativos, classicamente referidos como endossomos iniciais, endossomos tardios e lisossomos (HUOTARI *et al.*, 2011). As partículas internalizadas por fagocitose seguem um caminho semelhante, culminando em fagolisossomos formados pela fusão de fagossomos e lisossomos. Os lisossomos e os fagolisossomos têm pH de 4 a 4,5 e contêm uma série de proteases ácidas comumente chamadas de catepsinas (WATTS *et al.*, 2004).

Em geral, as moléculas HLA de classe I acoplam os peptídeos gerados por proteólise proteossômica e o encaixe no sulco de apresentação de peptídeos é realizado no retículo endoplasmático (ER), após os peptídeos serem translocados do citosol. A ligação do peptídeo pelo HLA de classe I está integrada na via de montagem que estabelece a união da cadeia pesada com a  $\beta 2$ -microglobulina. Além disso, HLA de classe I pode ligar peptídeos derivados de proteínas exógenas internalizadas pela endocitose ou fagocitose, um fenômeno chamado apresentação cruzada. Subconjuntos específicos de células dendríticas (DCs) são

particularmente eficientes na mediação deste processo, que é criticamente importante para o início de uma resposta primária por células T CD8<sup>+</sup> *naïve* (ADAMS; LUOMA 2013).

Os peptídeos gerados no citosol são translocados para o ER pelo transportador associado ao processamento de antígenos (TAP), membro da família ABC de transportadores dependentes de ATP (OANCEA *et al.*, 2009). TAP é uma proteína heterodimérica, e as subunidades TAP1 e TAP2 são codificadas por genes presentes no HLA, amplamente distribuídos em procariotos e eucariotos, responsáveis pela transferência de uma variedade de moléculas através das membranas. Evidências bioquímicas combinadas com modelagem molecular sugerem que cada subunidade TAP consiste em um domínio central de seis  $\alpha$ -hélices transmembranares, que constituem o canal, que está localizado na porção N-terminal do domínio de ligação a nucleotídeos (NBD). Os *loops* citosólicos nos domínios centrais que são proximais às NBDs constituem o local de reconhecimento de peptídeos, e a hidrólise de ATP vai mediar o evento de translocação. Ambas as subunidades possuem domínios adicionais N-terminais (N-domínios), que compreendem quatro segmentos transmembranares para TAP1 e três para TAP2 (HINZ *et al.*, 2012).

O heterodímero TAP associa-se a várias outras proteínas para formar o complexo de carregamento de peptídeos (PLC). A glicoproteína transmembrana tapasina, que é codificada por gene presente no HLA (ORTMANN *et al.*, 1997), interage dentro da membrana com os N-domínios (TAN *et al.*, 2002; LEONHARDT *et al.*, 2005). A proteína tapasina tem uma função de ponte, recrutando o complexo HLA de classe I: $\beta$ 2m e a proteína calreticulina (CRT) para o PLC. Experimentos recentes confirmaram que existem duas moléculas de tapasina no PLC, uma associada a cada subunidade TAP. A tapasina por sua vez é ligada de forma estável por meio de uma ponte dissulfeto a uma segunda molécula, a proteína dissulfeto isomerase ERp57. O ERp57 auxilia na dobra de glicoproteínas recém-sintetizadas no ER mediando a isomerização de pontes dissulfeto (PANTER *et al.*, 2012).



**Figura 4. Biossíntese de HLA de classe I e ligação de peptídeo antigênico no ER.** As enzimas glucosidases I e II (Gls I / Gls II) aparam o glicano até um único resíduo de glicose terminal (G), que permite a interação da cadeia pesada de HLA de classe I com chaperonas semelhantes à lectina em vários estágios durante o enovelamento (*folding*) e a montagem da estrutura. Os eventos de *folding* inicial envolvem a calnexina (CNX) e permitem o acoplamento subsequente com  $\beta$ 2-microglobulina ( $\beta$ 2m). O heterodímero vazio, que é intrinsecamente instável, é então recrutado por calreticulina (CRT) através do glicano monoglucosilado ao complexo de carregamento de peptídeos (PLC). A associação dos heterodímeros HLA de classe I: $\beta$ 2m com o PLC estabiliza a molécula de HLA de classe I vazia e mantém o sulco de ligação em uma conformação que favorece o carregamento de peptídeos de alta afinidade. Essas funções são mediadas por interações diretas entre a cadeia pesada de HLA de classe I e tapasina, e são suportadas pela coordenação de interações com CRT e ERp57 no PLC. As moléculas de HLA de classe I com peptídeos subótimos são substratos para UGT1, que faz a reglicosilação do glicano de cadeia pesada, permitindo a reentrada do HLA de classe I no PLC e a troca por peptídeos de alta afinidade. Os peptídeos translocados para o ER associados a TAP originam-se principalmente da degradação proteossomal de proteínas endógenas ou DRiPs. Estas proteínas podem surgir a partir da tradução de RNA próprio ou exógeno (viral, por exemplo) ou, no caso de apresentação cruzada, por translocação para o citosol a partir de endossomos ou fagossomos. Muitos dos peptídeos que são levados ao ER são maiores do que 8-10 resíduos (tamanhos preferidos pelas moléculas de HLA de classe I) e sofrem cortes por aminopeptidases presentes no ER, conhecidas como ERAAP / ERAP1 e ERAP2. Finalmente, os peptídeos de alta afinidade se ligam preferencialmente às moléculas de HLA de classe I no PLC por processo de edição mediado por tapasina; os complexos peptídeo:HLA de classe I são liberados e então transitam para a superfície celular para o reconhecimento de células T por células T CD8<sup>+</sup>. Fonte: Modificado de Blum *et al.* (2013).

### 1.3 Proteólise e introdução de antígenos no citosol

Os antígenos proteicos são convencionalmente introduzidos no citosol pela síntese proteica resultante da maquinaria celular. Quando um vírus infecta uma célula, os genes virais são transcritos em RNAs mensageiros (mRNAs) e estes são traduzidos em ribossomos hospedeiros para gerar proteínas virais, sendo os antígenos citosólicos a principal fonte de peptídeos associados ao HLA de classe I. Este processo não é específico para proteínas virais; as proteínas hospedeiras são igualmente degradadas e geram peptídeos que se ligam ao HLA de classe I. De fato, no caso de autoimunidade ou imunidade tumoral, os peptídeos derivados de proteínas hospedeiras associadas ao HLA de classe I podem ser reconhecidas pelas células T CD8<sup>+</sup> (HAHN *et al.*, 1991; BOON *et al.*, 2006). No entanto, a geração destes complexos de peptídeos:HLA de classe I é praticamente sempre dependente de TAP e proteassoma. Isso implica que, apesar da presença de uma sequência sinal (também conhecido como peptídeo sinal) e do potencial de translocação para o ER, os mecanismos de processamento não são diferentes dos envolvidos na geração de peptídeos de antígenos exclusivamente citosólicos. Essas observações contribuíram para a hipótese de que proteínas citosólicas intactas e com a conformação estrutural correta não são a principal fonte de peptídeos que se ligam ao HLA de classe I. Em vez disso, as fontes são proteínas incompletas, talvez por causa de terminação prematura na sua geração, ou de conformação estrutural incorreta, pois as chaperonas citosólicas não são 100% efetivas na mediação do *folding* de proteínas recém-sintetizadas (ADAMS; LUOMA 2013).

Em células de mamíferos, aproximadamente 30% das proteínas totais são degradadas rapidamente após a síntese (SCHUBERT *et al.*, 2000). Yewdell e colaboradores são fortes defensores da hipótese de que este *pool* rapidamente degradado é a principal fonte de peptídeos associados ao HLA de classe I, cunhando a DRiP (produto ribossomal com defeito) para descrevê-los. Experimentos mostraram que a expressão de proteínas truncadas, que são instáveis, gera complexos de peptídeo:HLA de classe I tão efetivamente quanto proteínas completas (YEWDELL *et al.*, 2011). Embora os DRiPs sejam fonte significativa e talvez a principal, os peptídeos associados ao HLA de classe I podem ser derivados de proteínas intactas (VILLANUEVA *et al.*, 1995).

A extraordinária sensibilidade do reconhecimento de células T está bem estabelecida. São necessários pouquíssimos complexos de peptídeo:HLA de classe I para ativar uma célula

T, até mesmo apenas um complexo podendo ser suficiente (SYKULEV *et al.*, 1996). Possivelmente devido a isso, alguns peptídeos associados a HLA de classe I têm origens que não dependem da tradução convencional, sendo que existem exemplos de peptídeos antigênicos que são derivados de sequências incorporadas em íntrons (VIGNERON *et al.*, 2011).

#### 1.4 Arquitetura do TCR $\alpha\beta$

Durante o desenvolvimento das células T no timo, a geração de TCRs  $\alpha\beta$  funcionais requer:

- Rearranjo aleatório de segmentos de genes variáveis (V), de diversidade (D) e de junção (J);
- Rearranjo aleatório de segmentos gênicos V e J a partir do *loci* do gene TCR $\beta$  e TCR $\alpha$ , respectivamente.

Dentro do TCR, existem seis regiões de determinação de complementaridade hipervariáveis (CDRs) que realizam o reconhecimento. As regiões CDR1 e CDR2 são codificadas pelos genes *TRAV* e *TRBV* (DAVIS *et al.*, 1988), e as regiões CDR3 são formadas na junção de diferentes rearranjos de genes V (D) J (GARCIA *et al.*, 1996). O estabelecimento do repertório de TCR diversificado depende da diversidade combinatória e juncional. A diversidade combinatória surge através das muitas permutações e combinações diferentes de segmentos de genes V, D e J, dos quais há aproximadamente 43-45 *TRAV*, 50 *TRAJ*, 40-48 *TRBV*, 2 *TRBD* e 12-13 segmentos de genes funcionais *TRBJ*. A diversidade juncional expande drasticamente o repertório TCR pela falta de precisão durante o rearranjo dos genes V (D) J e pela inclusão de nucleotídeos (N) nas junções V (D) J. Além disso, diversidade combinatória surge do pareamento aleatório de diferentes cadeias  $\alpha$  e cadeia  $\beta$ . Do número relativamente pequeno de genes TCR, há, após a seleção tímica, aproximadamente  $2 \times 10^7$  TCRs em cada ser humano (ARSTILA *et al.*, 1999).

Como as moléculas de HLA de classe I geralmente ligam peptídeos de 8-10 resíduos (em média 9-mers, chamados nonapeptídeos) em uma conformação prolongada e os chamados resíduos-âncora enterrados em bolsões que diferem de alelo para alelo, o peptídeo ligado (P1-

P9) é mais acessível para inspeção pelo TCR no HLA de classe I clássico devido à sua protuberância para fora do sulco (FREMONT *et al.*, 1992; MADDEN *et al.*, 1993).

Além de exibir peptídeos para moléculas de TCRs, as moléculas de HLA em geral acumularam muitas outras funções durante a evolução, que afetam a organização e a flexibilidade do seu domínio, bem como sua especificidade de substrato. Por exemplo, na molécula de HLA não clássica CD1, o sulco de ligação ao peptídeo é mais profundo, mais estreito e mais hidrofóbico do que nos HLA de classe I, de modo que as caudas lipídicas de glicolípidos e lipopeptídeos estão ligadas ao sulco e suas porções polares são apresentadas às células T (ZAJONC *et al.*, 2003; ZAJONC *et al.*, 2005).

Geralmente, o heterodímero TCR é orientado diagonalmente em relação ao sulco de ligação ao peptídeo na molécula de HLA (GARBOCZI *et al.*, 1996; KJER-NIELSEN). Os contatos peptídicos são feitos principalmente através dos *loops* CDR3, que apresentam o maior grau de variabilidade genética. As primeiras estruturas cristalográficas demonstrando os complexos formados por TCRs e HLA de classe I levaram a propostas de que a orientação TCR é aproximadamente diagonal com uma média de cerca de 35° (RUDOLPH *et al.*, 2002). Em contraste, nos complexos de classe II, a orientação foi descrita como sendo mais próxima à perpendicular (REISER *et al.*, 2000; HENNECKE *et al.*, 2000), sugerindo um modo de ligação diferente entre as classes de HLA (WANG *et al.*, 2002).

O banco de dados atual das estruturas cristalográficas contendo complexos TCR:peptídeo:HLA apresenta um cenário em que o TCR se aproxima do peptídeo:HLA de forma diagonal, e, em seguida, usa a plasticidade conformacional inerente aos *loops* CDR para moldar e aumentar o contato com o peptídeo:HLA, que então determina o resultado final do encaixe (SIM *et al.*, 1997). Como o ângulo de ancoragem ditado pela interface TCR:peptídeo:HLA influencia a disposição dos domínios de TCR em relação a outros componentes do complexo de sinalização TCR, como CD4 ou CD8, é provável que também influencie a sinalização das células T.

As primeiras estruturas cristalográficas do complexo TCR:peptídeo:HLA de classe I proporcionaram uma riqueza de informações relativas a este evento de reconhecimento de células T (GARCIA *et al.*, 1998). O TCR encaixa-se acima da fenda de ligação do peptídeo:HLA de classe I em uma orientação aproximadamente diagonal, com os *loops* CDR3 localizados acima do peptídeo, enquanto que os *loops* CDR1 e CDR2 vão principalmente mediar os contatos do HLA (BRIDGEMAN *et al.*, 2012).

Muitos estudos estruturais ajudaram a entender os principais conceitos imunológicos no contexto do reconhecimento de TCR, incluindo a restrição do HLA, o polimorfismo do HLA de classe I e a autotolerância ao HLA de classe I (GRAS *et al.*, 2009); seleção de TCR (TURNER *et al.*, 2006); reatividade cruzada; aloreatividade; autoimunidade mediada por células T; e alergias (GRAS *et al.*, 2012; BIRNBAUM *et al.*, 2014).

### 1.5 HLA-G: molécula de histocompatibilidade de classe I não clássica

Com base em todas as informações anteriores, vamos agora caracterizar a molécula de HLA de classe I não clássica HLA-G, foco do presente trabalho devido às suas funções imunomodulatórias.

O nosso grupo tem sido pioneiro nos estudos dos diversos segmentos dos genes não clássicos, particularmente *HLA-G* e *HLA-E*, em estudos populacionais e em associação a doenças (CASTELLI *et al.*, 2009; CASTELLI *et al.*, 2010; VEIGA-CASTELLI *et al.*, 2012; VEIGA-CASTELLI *et al.*, 2013).

O gene *HLA-G* foi primeiramente descrito por Geraghty e colaboradores em 1987, sendo a sua expressão inicialmente descrita na interface materno-fetal no citotrofoblasto, contribuindo para o processo de tolerância imunológica do feto pela mãe. Na placenta, o repertório de peptídeos para HLA-G é extremamente restrito, com um único peptídeo representando 15% de todos os ligantes recuperados (ISHITANI *et al.*, 2003). Além da gravidez, *HLA-G* também pode ser constitutivamente expresso em outros tecidos, por exemplo, timo, córnea, matriz proximal da unha, pâncreas e células tronco hematopoiéticas (DONADI *et al.*, 2011).

De forma geral, uma função bem reconhecida da molécula HLA-G é a inibição da citotoxicidade das células *natural killer* (NK) e de linfócitos TCD8<sup>+</sup>, por interação com receptores leucocitários presentes na superfície dessas células, incluindo ILT2, ILT4, CD160 e KIR2DL4 (CAROSELLA *et al.*, 2008). ILT2 e ILT4 também interagem com várias moléculas HLA de classe I clássicas, mas têm maior afinidade para HLA-G (SHIROISHI *et al.*, 2003), sendo KIR2DL4 receptor específico para HLA-G (BOYSON *et al.*, 2002; McINTIRE *et al.*, 2005). A ligação de HLA-G com esses receptores modula a atividade de células T citotóxicas

e NK, induzindo a expressão de receptores inibitórios contendo motivos de tirosina (ITIM) (CAROSELLA *et al.*, 2003). Discutiremos a interação da molécula HLA-G e seus receptores com mais detalhes adiante.

Características biológicas bem conhecidas da molécula não clássica HLA-G que diferem de outras moléculas clássicas HLA de classe I (-A, -B e -C) incluem: (1) limitada variabilidade proteica, (2) presença de diversas isoformas solúveis e ligadas à membrana, geradas por edição alternativa do transcrito primário, (3) estrutura molecular única, que apresenta cauda citoplasmática reduzida, (4) modulação da resposta imune, e (5) expressão tecidual restrita. Sítios polimórficos presentes na região codificante e não codificante do gene *HLA-G* podem potencialmente afetar todas essas características biológicas. Variabilidade de nucleotídeos na região promotora ou na região 3' não traduzida (3' NT) pode influenciar os níveis de expressão de *HLA-G*, modificando a afinidade de sequências gênicas específicas para fatores de transcrição ou pós-transcricionais, respectivamente. Da mesma forma, a variabilidade de nucleotídeos na região codificadora pode produzir alterações conformacionais na molécula de HLA-G, o que pode modificar as suas funções principais, isto é, a interação com receptores celulares, produção de determinadas isoformas, características de polimerização (formação de dímeros), capacidade para acoplar e apresentar peptídeos, além de afetar a modulação da resposta imune (DONADI *et al.*, 2011).

A região codificadora do gene *HLA-G* mostra semelhança com os genes HLA de classe I clássicos, mas com menor variabilidade, apresentando 58 alelos que codificam apenas 18 proteínas (IMGT/HLA. Disponível em <http://www.ebi.ac.uk/cgi-bin/ipd/imgt/hla/align.cgi> banco de dados versão 3.22; ROBINSON *et al.*, 2013).

### **1.5.1 Funções imunomodulatórias da molécula HLA-G**

Como citado anteriormente, a molécula de HLA-G e as outras moléculas não clássicas atuam pobremente como apresentadoras de antígenos, mas robustamente na modulação da resposta imune através da sua interação com receptores imunes (CASTELLI *et al.*, 2011). Os mecanismos de inibição imunológica dos receptores de HLA-G podem ser organizados em três grupos: (i) funções imunoinibitórias diretas através do bloqueio de células efetoras; (ii) funções

imunoinibitórias indiretas através da geração de células reguladoras; e (iii) outras funções de HLA-G que têm consequências imunoinibitórias.

### 1.5.1.1 Inibição de células efectoras

Os citotrofoblastos presentes na interface materno-fetal não expressam outras moléculas HLA de classe I, além de HLA-C em níveis baixos e HLA-G. Neste contexto fisiológico, HLA-G inibe a função citolítica das células NK uterinas e periféricas (ROUAS-FREISS *et al.*, 1997; ROUAS-FREISS *et al.*, 1997). Os trofoblastos são protegidos da citólise por células NK decíduais e periféricas de origem semialogênica ou alogênica por moléculas de HLA-G. HLA-G pode interagir com receptores inibitórios na superfície das células NK, levando à sua inibição funcional. Além disso, mostrou-se que, através da interação direta com os receptores inibidores de ILT2 ou ILT4, HLA-G inibiu células T CD8<sup>+</sup> citotóxicas (CTLs).

Os receptores ILTs são codificados por uma família de genes de receptores imunes localizados em humanos no cromossomo 19q13.4. Esse *locus* é chamado de complexo de receptor leucocitário e inclui, além dos genes de ILTs, os genes que codificam receptores KIRs, NKp46 e receptor Fc (SHIROISHI *et al.*, 2003). Ambos ILT2 e ILT4 possuem 4 domínios semelhantes à imunoglobulina - Ig (do inglês *Ig-like*) extracelulares, além de 4 e 3 *motifs* inibitórios baseados em tirosina (ITIM), respectivamente, em suas caudas citoplasmáticas. Esses *motifs* ITIM recrutam a proteína tirosina fosfatase SHP-1 (do inglês *Src homology 2 domain containing phosphatase 1*), os quais acredita-se que possam inibir eventos de sinalização iniciais desencadeados por receptores estimuladores (DIETRICH *et al.*, 2001).

Os receptores ILT2 são expressos na superfície de diversos leucócitos, tais como monócitos, células B, células dendríticas, NK, subgrupos de células T e células linfomononucleares, enquanto receptores ILT4 são primariamente expressos na superfície de células da linhagem mielomonocítica, incluindo monócitos e células dendríticas. Tanto ILT2 quanto ILT4 possuem diversos *motifs* de tirosina inibitórios (ITIM) em suas caudas citoplasmáticas, inibindo a sinalização proveniente da ativação de receptores estimuladores. ILT2 e ILT4 se ligam à molécula de HLA-G com afinidade 3 a 4 vezes maior do que à HLA's clássicos, sugerindo que o reconhecimento HLA-G:ILT deve possuir papel dominante na

regulação da ativação de células NK, T e mielomonocíticas (SHIROISHI *et al.*, 2003; DONADI *et al.*, 2011).

ILT2 e ILT4 se ligam ao domínio  $\alpha 3$  e à  $\beta 2$ -microglobulina presente na molécula HLA-G, mas ILT4 se liga com maior afinidade do que ILT2. ILT4 possui maior ligação com o domínio  $\alpha 3$  do que com a porção da  $\beta 2$ -microglobulina e os sítios de ligação desses receptores são distintos (SHIROISHI *et al.*, 2003). O fato de que ILT2 e ILT4 se ligam com maior afinidade à molécula de HLA-G do que a moléculas clássicas de HLA de classe I levanta a possibilidade de que ambos os ILTs possam contribuir para interações funcionais entre leucócitos expressando ILTs (células T, NK e mielomonocíticas) e células expressando HLA-G, como por exemplo células epiteliais tímicas e células infectadas por citomegalovírus humano. ILT2 pode influenciar o desenvolvimento de timócitos através da interação com HLA-G e HLA de classe I clássicos em células epiteliais tímicas, modulando dessa forma o limiar de ativação do TCR. É importante notar que a afinidade de ligação entre ILT2:HLA-G é relativamente alta, consistente com observações prévias de que efeitos de inibição mediados por ILT2 em células NK do sangue periférico podem ser em grande parte atribuídas ao reconhecimento de HLA-G. De forma semelhante, a interação ILT2:HLA-G pode também afetar o reconhecimento pelas células NK de trofoblastos e células tumorais expressando HLA-G, contribuindo para tolerância materno-fetal e escape de células tumorais. Embora a afinidade da ligação de ILT2 ao HLA-G seja levemente menor do que ILT4, ILT2 demonstra preferência muito maior pelo HLA-G *versus* HLA de classe I clássicos, o que sugere que a interação ILT2:HLA-G pode ser de considerável importância para a regulação da maturação e/ou função celular (SHIROISHI *et al.*, 2003).

HLA-G apresenta papel inibitório em células citotóxicas exibindo CD8 em sua superfície. A molécula CD8 interage com o domínio  $\alpha 3$  de moléculas clássicas e não clássicas, incluindo HLA-G e HLA-E, embora exibam diferentes afinidades. CD8 $\alpha/\alpha$  liga-se a HLA-G com uma afinidade comparável à da interação com *HLA-A\*02*, contrastando com a menor afinidade para HLA-E. A ligação direta de CD8 $\alpha/\alpha$  a alelos clássicos de HLA foi inicialmente demonstrada usando ensaios celulares de adesão. Os dados de mutagênese sugeriram que CD8 e TCR se ligam a sítios separados nas moléculas de HLA. Recentemente, as estruturas cristalográficas de complexos CD8 $\alpha/\alpha$ :HLA humanos e murinos mostraram que CD8 $\alpha/\alpha$  se liga a um região extensa em moléculas clássicas de HLA, que inclui resíduos não polimórficos do domínio  $\alpha 2$  e  $\alpha 3$ , bem como  $\beta 2$ -microglobulina (GAO *et al.*, 2000; DONADI *et al.*, 2011).

### 1.5.1.2 Modulação de células reguladoras/supressoras

O HLA-G também possui funções tolerogênicas "de longo prazo", através da geração de células reguladoras/supressoras. As células reguladoras são células efetoras imunes que são criticamente envolvidas na promoção e manutenção da tolerância imune porque podem inibir a reatividade de outros efetores. A capacidade de HLA-G para induzir células T reguladoras foi investigada, e foi demonstrado que as células T CD4<sup>+</sup> e CD8<sup>+</sup> que foram estimuladas na presença de HLA-G não só perderam a capacidade de responder a estimulação antigênica, mas também foram diferenciadas em células T reguladoras capazes de inibir a reatividade de outras células T (LEMAOULT *et al.*, 2004).

As células T reguladoras induzidas por HLA-G necessitaram de HLA-G para sua geração, mas não para sua função e, quando geradas, as células reguladoras induzidas pelo HLA-G não exigem mais HLA-G para que sejam tolerogênicas (LEMAOULT *et al.*, 2007).

As células NK e T ativadas - e, em menor medida, os monócitos - também podem comportar-se temporariamente como células supressoras reguladoras e inibir as respostas imunes através do HLA-G depois de terem adquirido membranas contendo HLA-G a partir das células nas suas proximidades pelo mecanismo da trogocitose. O fenômeno da trogocitose é a transferência de membrana plasmática e proteínas ancoradas durante o contato célula-célula. Um estudo sobre a estrutura de HLA-G em exsudatos mostrou que HLA-G pode formar complexos de alto peso molecular por meio de pontes de dissulfeto, que estão presentes em exsudatos de pacientes com doenças inflamatórias ou câncer (CAUMARTIN *et al.*, 2007).

Foram identificadas DCs humanas tolerogênicas que caracterizam um novo subconjunto, caracterizados pela produção de interleucina-10 (IL-10) (GREGORI *et al.*, 2010). Chamadas de DC-10, estas células estão presentes *in vivo* em sangue periférico e órgãos linfoides secundários, são induzíveis *in vitro* a partir de monócitos na presença de IL-10 e são caracterizados pela expressão de níveis elevados de HLA-G ligado a membrana. Eles são potentes indutores de células T reguladoras específicas de tipo 1 (Tr1), um subconjunto induzível de Tregs que desempenham papel fundamental na promoção e manutenção da tolerância, cuja geração depende criticamente da expressão de HLA-G ligado à membrana, interação com ILT4 e IL-10 produzida por células DC-10 (ROSSETTI *et al.*, 2010).

### 1.5.1.3. Outras funções imunológicas de HLA-G

Existem diversas outras funções imunomodulatórias de HLA-G que foram descritas. Uma delas é a capacidade de HLA-G solúvel (sHLA-G) de induzir apoptose de células NK e células T citotóxicas (FOURNEL *et al.*, 2000), outra a capacidade de sHLA-G e ligado à membrana de influenciar a expressão e a liberação de IFN- $\gamma$  por células NK (RAJAGOPALAN *et al.*, 2006; VAN DER MEER *et al.*, 2004; VAN DER MEER *et al.*, 2007), a capacidade de sHLA-G e ligado à membrana para induzir a regulação positiva de receptores inibitórios, incluindo o seu próprio (LEMAOULT *et al.*, 2005), e a função antiangiogênica de sHLA-G através da ligação de CD160 (FONS *et al.*, 2006).

Ao gerar um nonapeptídeo derivado de peptídeo sinal que pode ligar e estabilizar o HLA-E, HLA-G também pode aumentar a expressão na superfície celular de HLA-E e, portanto, modificar a função efetora das células NK através da interação com complexos CD94:NKG2 de caráter ativador ou inibitório (PENDE *et al.*, 1997).

## 1.6 Edição alternativa do gene HLA-G

Considerando que a variabilidade presente nas moléculas de HLA forma a base para a resposta do sistema imune adaptativo, devemos levar em consideração os eventos de edição alternativa (do inglês *alternative splicing*) aos quais tais genes são submetidos. O processo de edição alternativa tem papel fundamental na geração de complexidade biológica e a sua desregulação está frequentemente associada a doenças humanas. Os transcritos de cerca de 95% dos genes humanos que possuem multiéxons sofrem edição alternativa em mais de uma maneira e, na maioria dos casos, os transcritos resultantes são variavelmente expressos entre diferentes células e tipos de tecidos. Esse processo de edição alternativa evidencia como a informação genética controla numerosos processos celulares críticos, e estima-se que 15% a 50% das mutações que ocorrem em doenças humanas afetam a edição alternativa desses genes (BARASH *et al.*, 2010).

O processo de edição alternativa ocorre de forma que mRNAs eucarióticos são transcritos como precursores contendo íntrons, que são subsequentemente removidos. A

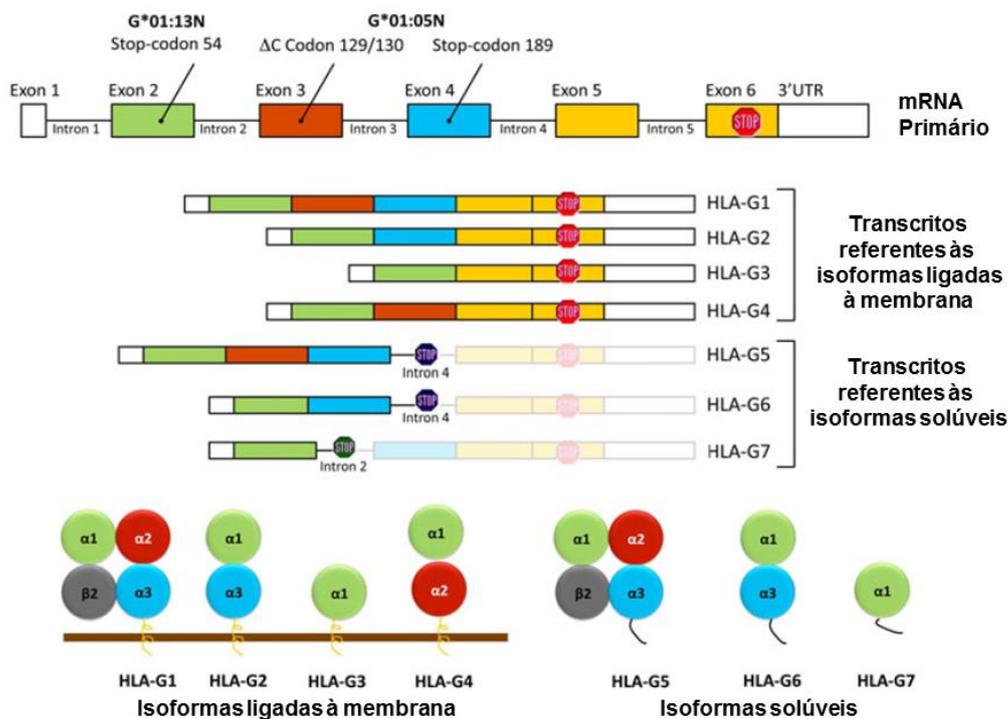
decisão sobre quais éxons serão removidos e quais serão incluídos envolve sequências de RNA específicas e proteínas reguladoras (KRÁLOVICOVÁ *et al.*, 2006).

No contexto de diversidade de isoformas dos genes do HLA, o *HLA-G* tem sido o mais estudado. O gene *HLA-G* apresenta 7 íntrons e 8 éxons, codificando uma cadeia pesada associada não covalentemente à  $\beta$ 2-microglobulina. O éxon 1 codifica o peptídeo sinal, os éxons 2, 3 e 4 codificam, respectivamente, os domínios extracelulares  $\alpha$ 1,  $\alpha$ 2 e  $\alpha$ 3, o éxon 5 codifica o domínio transmembrana e os outros éxons o domínio citoplasmático da cadeia pesada. No entanto, a presença do códon de parada no início do éxon 6 do gene *HLA-G* reduz a seis aminoácidos o domínio citoplasmático da proteína (HODGKINSON *et al.*, 2000).

Considerando o processo de edição alternativa, a molécula HLA-G apresenta pelo menos sete isoformas já identificadas HLA-G1 a HLA-G7, que são formadas a partir do transcrito primário. Quatro dessas isoformas são de membrana: HLA-G1 é a forma mais completa com estrutura semelhante à de uma molécula HLA de classe I clássica associada à  $\beta$ 2-microglobulina e ligada a um peptídeo de sequência consenso XI/LPXXXXXL; HLA-G2 é desprovida do domínio  $\alpha$ 2 codificado pelo éxon 3; HLA-G3 não possui os domínios  $\alpha$ 2 e  $\alpha$ 3 codificados pelos éxons 3 e 4; HLA-G4 perdeu o domínio  $\alpha$ 3 codificado pelo éxon 4. As outras três isoformas são solúveis: HLA-G5 e HLA-G6 possuem os mesmos domínios extra globulares que HLA-G1 e HLA-G2, respectivamente e são geradas por transcritos que conservam o íntron 4, que bloqueia a tradução do domínio transmembrana (éxon 5). A região 5' do íntron, em fase de leitura com o éxon 4, é traduzida até um códon de parada e gera as isoformas HLA-G5 e HLA-G6, contendo uma cauda de 21 aminoácidos específicos envolvidos na sua solubilidade. A isoforma solúvel HLA-G7 limita-se ao domínio  $\alpha$ 1 e possui dois aminoácidos específicos do íntron 2. Os transcritos alternativos são todos desprovidos do éxon 7. A **Figura 5** resume e ilustra as diversas isoformas de HLA-G (CAROSELLA *et al.*, 2008; DONADI *et al.*, 2011).

Igualmente às formas solúveis produzidas por edição alternativa, HLA-G1 de superfície pode ser clivada por metaloproteases que geram a forma solúvel HLA-G1s, também conhecido como sHLA-G. Alguns trabalhos se referem às isoformas solúveis de HLA-G como sHLA-G de forma indiscriminada, com o intuito de demonstrar que as funções imunomodulatórias de HLA-G estão sendo executadas por isoformas solúveis e não isoformas ligadas à membrana, sem especificar exatamente qual das isoformas solúveis está sendo analisada. Além disso, dependendo do tipo celular e da situação algumas isoformas podem ser expressas, e outras não (CAROSELLA *et al.*, 2008; DONADI *et al.*, 2011).

A capacidade de a proteína HLA-G formar dímeros é uma de suas principais características, sendo que tais dímeros são capazes de se ligar aos receptores de HLA-G, demonstrando maior afinidade e menores taxas de dissociação, quando comparados com monômeros. A dimerização é realizada por pontes dissulfeto, devido à presença de 2 resíduos únicos de cisteína localizados na posição 42 (Cys42) do domínio  $\alpha 1$  e na posição 147 do domínio  $\alpha 2$  (Cys147) (SHIROISHI *et al.*, 2006; HOVANGWYIN *et al.*, 2012). De fato, uma vez que todas as isoformas HLA-G possuem o resíduo Cys42, todas elas deveriam ser capazes de dimerizar (BOYSON *et al.*, 2002).



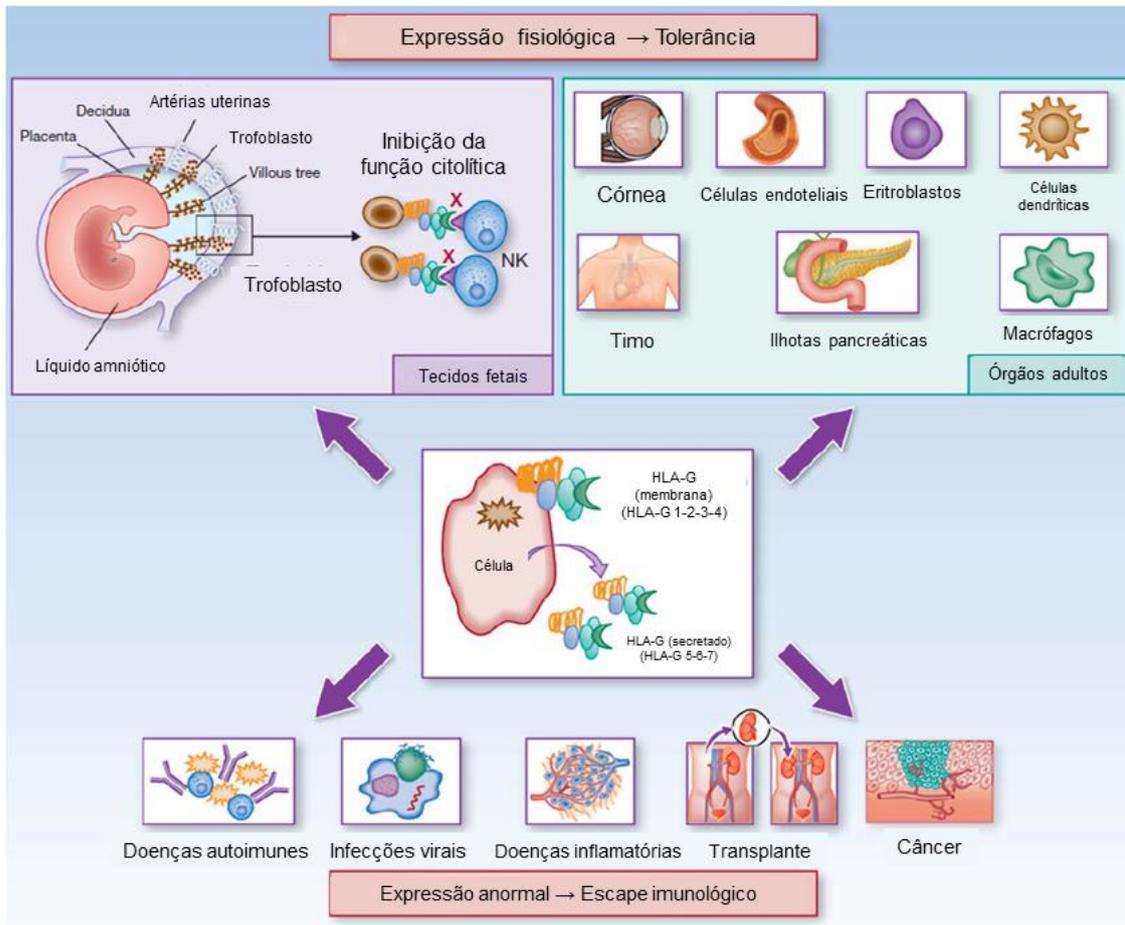
**Figura 5.** Estrutura do gene HLA-G e suas isoformas produzidas por edição alternativa a partir do transcrito primário. Fonte: Modificado de Donadi *et al.* (2011).

## 1.7 Papel da molécula HLA-G em situações fisiológicas e na doença

A principal expressão fisiológica da molécula HLA-G ocorre na placenta, onde participa da tolerância materna em relação ao feto. Sua expressão foi encontrada em tecidos embrionários, em órgãos imunoprivilegiados adultos e em células da linhagem hematopoiética. É expresso em vários tipos de tumores sólidos primários (melanoma, cabeça e pescoço, pulmão,

urogenital, gastrointestinal e câncer de mama), malignidades hematológicas (leucemia aguda, linfomas) e metástases. Além disso, está relacionado ao aceite ou rejeição de transplantes pelo hospedeiro, doenças autoimunes e doenças infecciosas.

A seguir abordaremos as funções molécula HLA-G em situações fisiológicas e relacionadas às doenças.



**Figura 6. Expressão de HLA-G.** Fisiologicamente, HLA-G é encontrado em células derivadas de placenta e, em adultos, em células de órgãos imunoprivilegiados, como córnea, timo, ilhotas pancreáticas, eritroblastos, macrófagos e precursores de células endoteliais. A expressão de HLA-G pode ser induzida em várias doenças, incluindo doenças autoimunes e inflamatórias, infecções virais, transplante e câncer. Fonte: Modificado de Curigliano *et al.* (2013).

### 1.7.1 HLA-G e gravidez

HLA-G interage com os receptores inibidores das células NK uterinas. Essas interações previnem o ataque do feto pelo sistema imunológico materno e também regulam a produção de citocinas e quimiocinas. A ligação de HLA-G ao receptor KIR2DL4 presente nas células NK uterinas aumenta a secreção de citocinas e fatores angiogênicos, o que favorece a implantação, vascularização e desenvolvimento da placenta (RAJAGOPALAN *et al.*, 2006). Para equilibrar esses fatores pró-angiogênicos, os efeitos antiangiogênicos foram atribuídos à molécula HLA-G secretada pelos trofoblastos (FONS *et al.*, 2006). Resultados publicados por Menier e colaboradores mostram que não só a isoforma HLA-G1, mas também as outras isoformas truncadas HLA-G, podem inibir a citólise por células NK e, portanto, contribuir para o privilégio imune e proteção do feto (MENIER *et al.*, 2000).

O papel do HLA-G como molécula tolerogênica chave na tolerância materno-fetal é altamente apoiado pelas correlações estabelecidas entre a diminuição da expressão de HLA-G e distúrbios da gravidez associados à perda fetal, como abortos espontâneos recorrentes e pré-eclâmpsia. A relevância do HLA-G na reprodução humana é ainda enfatizada no contexto da fertilização *in vitro* (FIV), uma vez que foi demonstrado que a expressão de moléculas de sHLA-G por embriões foi fundamental para a implantação uterina. Isto fica claro pela correlação dos níveis de sHLA-G em sobrenadantes de cultura de embriões após FIV com implantação bem-sucedida (RIZZO *et al.*, 2007). Hoje existem testes ELISA disponíveis comercialmente especificamente para a detecção de HLA-G solúvel em sobrenadantes de embriões pré-implantação. Esses testes para selecionar embriões para transferência estão agora disponíveis em clínicas que oferecem fertilização *in vitro* (APPS *et al.*, 2008).

É possível que a produção de HLA-G por embriões possa estar envolvida nos mecanismos que levam à sua capacidade de estabelecer a comunicação adequada com o endométrio e alcançar a implantação embrionária. A possibilidade de identificar embriões produtores de HLA-G abre novas perspectivas para a melhoria da fertilização *in vitro*, no que se refere à seleção de embriões. Além disso, o sHLA-G de origens alternativas que é detectável no sangue materno, mesmo antes da ovulação, provavelmente também influencia a implantação do embrião humano, presumivelmente promovendo a receptividade endometrial (COMISKEY *et al.*, 2007).

### 1.7.2 HLA-G e tumores

Desde a primeira descrição da expressão de HLA-G por células tumorais, em 1998 (PAUL *et al.*, 1998) vários estudos foram realizados em mais de mil lesões malignas, mostrando que, embora não expresso nas áreas normais circundantes, a transcrição do gene HLA-G e a expressão da proteína são ativadas em várias lesões tumorais (ROUAS-FREISS *et al.*, 2005).

Além da expressão de HLA-G *in situ*, moléculas de sHLA-G foram encontradas circulando em altas concentrações em alguns pacientes com câncer (UGUREL *et al.*, 2001; SINGER *et al.*, 2003; LELEU *et al.*, 2005).

À luz das propriedades tolerogênicas conhecidas do HLA-G, seu papel no câncer foi assim especulado como mecanismo de escape imune dirigido por tumores. Todos os estudos realizados até à data suportam essa hipótese. De fato, a expressão de HLA-G por células tumorais os protege da citólise mediada por NK e CTL; assim, a depuração imune de células tumorais pode ser reforçada através do bloqueio de HLA-G (ROUAS-FREISS *et al.*, 2005). Conseqüentemente, o efeito funcional das moléculas de sHLA-G purificadas a partir do plasma de pacientes que sofrem de distúrbios linfoproliferativos foi demonstrada pela sua forte capacidade de inibir a proliferação de células T em concentrações atualmente observadas durante esses distúrbios (SEBTI *et al.*, 2007).

Além disso, níveis elevados de HLA-G no plasma foram observados recentemente em pacientes com neuroblastoma e correlacionados com recidivas (MORANDI *et al.*, 2007). Aqui novamente, o sHLA-G a partir do plasma patológico foi funcional, como mostrado pela inibição da lise de células de neuroblastoma mediada por NK e CTL. Neste último estudo, o HLA-G foi secretado por monócitos ativados por tumor e não pelas próprias células tumorais (MENIER *et al.*, 2004; PISTOIA *et al.*, 2007).

Glioblastoma é um tumor altamente maligno do sistema nervoso central que se acredita ser originário de astrócitos ou células precursoras astrocíticas. Apesar de receberem tratamento agressivo, os pacientes com glioblastoma têm tempo médio de sobrevivência de 12 meses. Em pacientes com glioblastoma, os níveis séricos de sHLA-G não diferiram dos encontrados em controles saudáveis, mas os pacientes com níveis altos de sHLA-G tiveram sobrevida significativamente menor do que aqueles com níveis de HLA-G baixos (PISTOIA *et al.*, 2007).

Nas células tumorais, o HLA-G pode ser expresso na superfície celular, segregado ou incorporado em exossomos derivados de tumores (ROUAS-FREISS *et al.*, 2005). A trogocitose constitui mecanismo de inibição imune e de proteção para células tumorais HLA-G que estão em a proximidade de células tumorais HLA-G-positivas (LEMAOULT *et al.*, 2007; CAUMARTIN *et al.*, 2007). De fato, após a aquisição de membranas contendo HLA-G1 a partir de células tumorais, as células NK efetoras param de proliferar, deixam de ser citotóxicas em relação a alvos legítimos e se comportam como células reguladoras capazes de inibir as funções citotóxicas de outras células NK. Esta inversão funcional imediata de uma célula efetora a uma célula reguladora é diretamente decorrente da HLA-G1 de superfície celular adquirida.

Estudos funcionais usando células NK, bem como linfócitos T citotóxicos CD8<sup>+</sup> específicos demonstraram que a expressão de HLA-G evita a lise das células de carcinoma renal por estas diferentes células efetoras imunes. Em contraste, as células renais normais HLA-G-negativas, bem como as células de carcinoma renal HLA-G-negativas não foram reconhecidas pelas células NK e T. Assim, o HLA-G representa mecanismo importante de escape imune do carcinoma renal humano e isso tem impacto no desenvolvimento de imunoterapias baseadas em células T e NK para essa doença (SELIGER *et al.*, 2007)

A relevância clínica do HLA-G em câncer é suportada pelas seguintes observações. Primeiro, a expressão de HLA-G está associada a transformação maligna e usualmente não é detectada em áreas circundantes normais (ROUAS-FREISS *et al.*, 2005). Por exemplo, o HLA-G pode ser usado para identificar a natureza e avaliar a agressividade dos tumores ovarianos (SHIH *et al.*, 2007). Além disso, HLA-G mostrou ser útil como ferramenta de diagnóstico diferencial para identificar tumores de trofoblasto *versus* tumores de origem não trofoblástica (SHIH *et al.*, 2007). Em segundo lugar, o papel do HLA-G no escape imune é enfatizado pelo envolvimento na resistência às terapias baseadas em intérféron observadas em alguns pacientes com melanoma (WAGNER *et al.*, 2000). Em terceiro lugar, a correlação entre resultados clínicos pobres e expressão de HLA-G em tumores gástricos e colorretais (YE *et al.*, 2007) ou na leucemia linfocítica crônica de células B (NÜCKEL *et al.*, 2005) mostra que HLA-G pode ser usado como marcador de prognóstico.

Com base em todos esses dados publicados, a expressão de HLA-G pode ser alvo promissor para a terapia antitumoral. Pode-se parar a transcrição de HLA-G, atuar sobre o *splicing* alternativo de HLA-G ou bloquear a função de células supressoras que expressam HLA-G (ROUAS-FREISS *et al.*, 2005).

### 1.7.3 HLA-G e transplantes

Até hoje, a expressão de HLA-G foi estudada em mais de mil amostras de plasma de pacientes após transplante de coração (LILA *et al.*, 2007), rins (QIU *et al.*, 2006), e fígado (NAJI *et al.*, 2007).

Os resultados obtidos mostram que os pacientes que expressam HLA-G no enxerto exibem significativamente melhor aceitação do mesmo (PISTOIA *et al.*, 2007). Além de determinar o status relacionado à expressão de HLA-G nos tecidos dos pacientes através da coloração de biópsias obtidas a partir de transplantes de órgãos, medir os níveis plasmáticos de sHLA-G em receptores de transplante representa uma estratégia para monitorar o grau de rejeição de aloenxertos. A este respeito, o aumento dos níveis plasmáticos de HLAG foi detectado em pacientes com incidência reduzida de rejeição aguda e crônica após transplante cardíaco e transplantes renais e hepáticos combinados (LE ROND *et al.*, 2006; NAJI *et al.*, 2007).

Além disso, pacientes que receberam transplante renal que expressavam HLA-G apresentaram menor produção de anticorpos aloreativos, principalmente envolvidos na rejeição aguda e crônica do enxerto (QIU *et al.*, 2006). Esses dados sugerem fortemente que a tolerância induzida por HLA-G também é alcançada atuando sobre a resposta das células B, mais especificamente por meio da produção de anticorpos contra os aloantígenos de HLA.

Em relação aos fatores que regulam a expressão de HLA-G em pacientes, os níveis plasmáticos de sHLA-G aumentaram logo após a administração dos fármacos imunossupressores ciclosporina e tacrolimus em alguns pacientes com transplante de coração (LUQUE *et al.*, 2006).

PBMCs de pacientes transplantados que expressam HLA-G possuem propriedades imunossupressoras e contêm proporção aumentada de células T CD3<sup>+</sup>CD4<sup>low</sup> e CD3<sup>+</sup>CD8<sup>low</sup> supressoras. Notavelmente, foi obtida correlação clara entre a sobrevivência do enxerto e os parâmetros imunológicos relacionados ao HLA-G, como as células T supressoras produtoras de IL-10, CD4 ou CD8 solúvel, CD3<sup>+</sup>CD4<sup>low</sup> ou CD3<sup>+</sup>CD8<sup>low</sup>, que podem contribuir para a tolerância em pacientes que expressam HLA-G transplantados (NAJI *et al.*, 2007).

Em conjunto, esses achados têm importantes implicações clínicas. Primeiro, a avaliação dos níveis plasmáticos de HLA-G pode ser usada para monitorar pacientes transplantados em que se espera melhor aceitação de seu enxerto. Em segundo lugar, a quantidade de fármacos imunossupressores poderia ser reduzida nestes pacientes HLA-G-positivos, reduzindo assim os efeitos secundários deletérios. Em terceiro lugar, o tratamento com HLA-G, seus derivados ou indutores de HLA-G pode constituir uma nova e eficiente terapia de rejeição anti-enxerto (CAROSELLA *et al.*, 2008).

Além disso, células tronco mesenquimais humanas que expressam HLA-G constituem uma nova abordagem terapêutica destinada a prevenir a rejeição após transplante de tecido sólido e líquido. Essas implicações são reforçadas por estudos no sistema murino que mostram que a injeção de pérolas revestidas com tetrâmero de HLA-G em camundongos receptores antes do enxerto de pele alogênico induziu células T supressoras e resultou no aumento significativo da sobrevivência do aloenxerto cutâneo (LIANG *et al.*, 2006). Estes dados mostram pela primeira vez que a administração *in vivo* de HLA-G é eficaz e melhora a sobrevivência do aloenxerto.

#### 1.7.4 HLA-G e doenças autoimunes

A expressão de HLA-G foi descrita em tecidos tais como a pele (CAROSELLA *et al.*, 2003), músculo (WIENDL *et al.*, 2003), pâncreas (CIRULLI *et al.*, 2006), trato digestivo (TORRES *et al.*, 2004), e sistema nervoso central (FEGER *et al.*, 2007) em processos inflamatórios que muitas vezes têm etiologia autoimune.

De fato, HLA-G foi detectado na psoríase e dermatite atópica (CAROSELLA *et al.*, 2003), bem como na doença celíaca (TORRES *et al.*, 2007) e na colite ulcerativa, mas não na doença de Crohn (RIZZO *et al.*, 2008).

A expressão de HLA-G foi detectada também em biópsias intestinais e no soro de pacientes com doença celíaca, uma doença autoimune caracterizada por resposta imune ao glúten ingerido. Esse achado pode sugerir que a expressão HLA-G reflete a tentativa de restaurar a tolerância ao glúten e compensar a inflamação (PISTOIA *et al.*, 2007).

Os níveis plasmáticos de sHLA-G reduzidos ou a expressão de *HLA-G* mais baixa por monócitos foram encontrados em pacientes que sofrem de artrite reumatoide (VERBRUGGEN *et al.*, 2006), asma (RIZZO *et al.*, 2005), e esclerose múltipla (FAINARDI *et al.*, 2006) em comparação com controles.

Essa baixa expressão de *HLA-G* correlaciona-se positivamente com os parâmetros da atividade da doença, sugerindo fortemente que o HLA-G não protege mais de forma eficiente os tecidos danificados das células NK e T. Curiosamente, a expressão de *HLA-G* por monócitos de pacientes com esclerose múltipla pode ser regulada positivamente pelo interferon- $\beta$  (IFN- $\beta$ ), que é comumente empregado no tratamento desta doença (MITSDOERFFER *et al.*, 2005).

Os efeitos benéficos da terapia com interferon podem ser devidos à restauração da expressão de HLA-G por monócitos que, portanto, são capazes de exercer função tolerogênica nesses pacientes. De acordo com esses dados, a ativação pós-parto da esclerose múltipla foi recentemente associada à diminuição da expressão de *HLA-G* (AIRAS *et al.*, 2007). Além disso, as células Treg que expressam HLA-G, definidas como um pequeno subconjunto de células T do sangue periférico em doadores saudáveis, estão em maior quantidade no líquido cefalorraquidiano de pacientes com esclerose múltipla, apoiando ainda a função importante do HLA-G na atividade de modulação da doença *in vivo* (FEGER *et al.*, 2007).

Em crianças asmáticas, os níveis globais de sHLA-G sérico não foram diferentes dos detectados nos controles. No entanto, quando a análise foi restrita aos asmáticos atópicos, os níveis de sHLA-G foram significativamente maiores nos pacientes do que nos controles saudáveis (TAHAN *et al.*, 2006; PISTOIA *et al.*, 2007).

Notavelmente, o HLA-G foi detectado em um subconjunto de grânulos contendo insulina que atingem a superfície celular das ilhotas pancreáticas que a insulina está sendo secretada. Uma vez que muitos autoantígenos na imunidade das ilhotas são componentes de grânulos de secreção e ativação de células T autoreativas, foi proposto que a ativação deletéria de células T autoreativas nos locais da exocitose de insulina pode ser evitada pelo agrupamento local de moléculas de HLA-G. Esses achados têm implicações potencialmente importantes para a prevenção de diabetes, bem como para o estabelecimento da tolerância ao transplante para ilhotas pancreáticas através da utilização do HLA-G (CIRULLI *et al.*, 2006).

### 1.7.5 HLA-G e doenças infecciosas

A indução da expressão de HLA-G por células infectadas por vírus pode ser mecanismo adicional que ajuda os vírus a subverter as defesas do hospedeiro. Foram realizadas investigações sobre o papel do HLA-G após infecções pelo vírus da imunodeficiência humana (HIV), vírus neurotrópicos, entre outros. Em pacientes infectados pelo HIV, a expressão de HLA-G foi regulada positivamente em células T CD8<sup>+</sup> e monócitos (LOZANO *et al.*, 2002). Considerando que os pacientes progressores de longo prazo expressaram altos níveis plasmáticos de HLA-G, enquanto os pacientes não progressores de longo prazo não o fizeram. Isso sugere um novo mecanismo empregado pelo HIV para evadir a citotoxicidade das células imunes (DONAGHY *et al.*, 2007). O papel do HLA-G na infecção pelo HIV foi ainda mais apoiado pela associação entre variantes genéticas HLA-G com susceptibilidade ao HIV (FAVIER *et al.*, 2007).

Um estudo mostrou aumento dos níveis plasmáticos de HLA-G5 em pacientes com choque séptico. Tais altos níveis de HLA-G permitem a previsão de sobrevivência melhor do que a detecção de HLA-DR em monócitos. Estes dados sugerem fortemente que a medição de HLA-G no plasma pode substituir o HLA-DR usado atualmente como marcador prognóstico após sepse (MONNERET *et al.*, 2007).

Laaribi e colaboradores descreveram níveis séricos de sHLA-G aumentados significativamente em pacientes com hepatite B crônica em comparação com pacientes que resolveram a infecção espontaneamente. Curiosamente, foram observados aumentos nos níveis de sHLA-G1 em pacientes com hepatite B crônica quando comparados com pacientes que resolveram a infecção espontaneamente e controles saudáveis. Além disso, a expressão de HLA-G5 parece ser maior no soro de pacientes com hepatite B crônica quando comparados com pacientes que resolveram a infecção espontaneamente. A análise dos dímeros de HLA-G mostrou a presença de homodímeros em 93% dos pacientes com hepatite B crônica, 67% em pacientes que resolveram a infecção espontaneamente e 60% em controles saudáveis. Esses resultados fornecem evidências de que sHLA-G pode ter papel crucial no resultado da infecção crônica de hepatite B e pode ser proposto como biomarcador para o resultado da infecção. Com base na sua função tolerogênica, o HLA-G pode ser considerado uma nova abordagem imunoterapêutica promissora para tratar essa infecção crônica (LAARIBI *et al.*, 2017).

A análise da expressão de *HLA-G* em infecções virais neurotrópicas mostrou que o vírus do vírus *herpes simplex 1* e o vírus da raiva regulam a expressão neuronal das isoformas de *HLA-G* tanto nas células infectadas como nas células vizinhas não infectadas (LAFON *et al.*, 2005). De fato, a regulação positiva da expressão do gene *HLA-G* foi observada em uma linha neuronal humana (células NT2-N) após infecção *in vitro* por uma cepa altamente patogênica do vírus da raiva ou uma cepa do vírus do herpes induzindo a latência dos neurônios. A expressão de *HLA-G* na superfície celular foi restrita aos neurônios infectados pelo vírus da raiva (MÉGRET *et al.*, 2007).

Esses achados mostram que esses vírus neurotrópicos regulam diferencialmente a expressão de *HLA-G* em neurônios humanos e que o *HLA-G* pode estar envolvido no seu escape da resposta imune no sistema nervoso.

## 1.8 Obtenção experimental de estruturas e bioinformática estrutural

A determinação da estrutura tridimensional de biomoléculas, com destaque para as proteínas, deu origem à cristalografia de proteínas, caracterizada como processo complexo que engloba uma variedade de estratégias e métodos tradicionais e modernos, integrando especialidades como a física, química, biologia, bioquímica e computação. Embora a estrutura tridimensional de macromoléculas possa ser obtida por meio de diversos métodos experimentais, tais como a ressonância magnética nuclear (RMN) e a criomicroscopia eletrônica, a cristalografia de raios-X ocupa papel de destaque. Isto pode ser evidenciado, por exemplo, no fato de que em janeiro de 2014 o banco de dados PDB (do inglês *Protein Data Bank*) apresentava aproximadamente 97.000 estruturas de macromoléculas depositadas (incluindo proteínas, ácidos nucleicos, complexos macromoleculares e polissacarídeos), dentre as quais aproximadamente 90% tiveram sua estrutura tridimensional determinada pelo método de cristalografia de raios-X. As estruturas de proteínas determinadas por métodos cristalográficos são frequentemente o ponto de partida para a construção de modelos moleculares que visam elucidar a estrutura e função de proteínas homólogas ou o planejamento de novas moléculas bioativas (VERLI, H. 2014).

A cristalografia da proteína *HLA-G*, contendo o peptídeo RIIPRHLQL na fenda, acoplada à  $\beta$ 2-microglobulina foi realizada por Clements e colaboradores, que desvendaram grande parte de sua porção extracelular, mas devido às próprias limitações da técnica, não foi

possível resolver a porção terminal do domínio  $\alpha 3$  que conecta a proteína à sua porção transmembrana e pequena cauda citoplasmática (CLEMENTS *et al.*, 2005).

O presente trabalho utilizou as estruturas cristalográficas disponíveis e técnicas de bioinformática estrutural com o intuito de contribuir para a resposta de questões relacionadas à estrutura tridimensional da molécula HLA-G, suas isoformas solúveis e possíveis dímeros que podem ser formados na natureza.

A bioinformática estrutural é o ramo da bioinformática que está relacionado com a análise, predição e modelagem da estrutura tridimensional de macromoléculas biológicas, tais como proteínas, RNA e DNA.

A modelagem molecular de proteínas tem amplos objetivos e o interesse pelo uso da técnica aumentou muito em função do crescente desempenho proporcionado pelos computadores nas últimas décadas e do aumento do número de métodos disponíveis. O processamento da informação estrutural, em especial quando combinado a dados experimentais, funcionais e evolucionários, é um procedimento aprimorado que pode fornecer várias informações relevantes para o entendimento de como determinada proteína funciona e sua estrutura tridimensional (PETREY *et al.*, 2015; XIANG *et al.*, 2006; SCHWEDE *et al.*, 2013).

Existem diversos métodos que podem ser aplicados para a modelagem molecular, e a seguinte classificação é ampla e atual quando se deseja avaliar e comparar os métodos de modelagem objetivamente:

i) Métodos baseados em estruturas molde (também chamados de *template based*). Incluem *threading* e modelagem comparativa.

ii) Métodos independentes de estruturas molde (também chamados de métodos *template free*). Inclui a predição *ab initio*.

No método de modelagem comparativa, também chamada de modelagem por homologia, a proteína de interesse (o alvo) terá sua estrutura tridimensional predita usando como referência a estrutura tridimensional de outra proteína similar (também chamada de molde, e na maioria das vezes evolutivamente relacionada). A modelagem comparativa é o método empregado mais frequentemente, e seu limite de predição está intrinsecamente relacionado com o grau de similaridade entre as estruturas alvo e molde. Essa proteína similar

deve possuir estrutura tridimensional resolvida experimentalmente, e as coordenadas cartesianas de seus átomos devem estar depositadas em banco de dados de estruturas (VERLI, H. 2014).

O método de predição do enovelamento ou *threading* parte da ideia de observações de que a estrutura tridimensional é mais conservada que a sequência, de forma que mesmo sequências com pouca similaridade podem possuir estruturas muito semelhantes, o que limita o número de enovelamentos que proteínas podem assumir. Esse método é assim menos dependente da proximidade evolutiva entre a sequência de aminoácidos da proteína de interesse e seus possíveis moldes, ou seja, as sequências podem apresentar baixa identidade. As limitações do método são percebidas em situações onde a identidade entre a sequência alvo e as proteínas utilizadas na construção do banco de enovelamentos é muito baixa, sendo possível que o enovelamento da sequência simplesmente não esteja representado no banco (VERLI, H. 2014).

A metodologia de predição *ab initio* difere-se das demais por não usar informações de estruturas conhecidas, relacionadas com a sequência alvo, e por usar funções de energia contendo somente termos de significado físico. Essa estratégia é baseada em dois pressupostos: todas as informações necessárias sobre a estrutura de uma proteína estão contidas em sua sequência de aminoácidos, e acredita-se que as proteínas se enovelam para um estado nativo, ou um conjunto de estados nativos, que se encontra no (ou próximo ao) mínimo global de energia livre. A metodologia *ab initio* pode ser aplicada para correção ou refinamento de estruturas modeladas por outras metodologias ou mesmo na predição de proteínas desordenadas. O sucesso dos métodos *ab initio* depende, principalmente, de uma função de energia acurada, na qual o estado nativo da proteína corresponda ao estado termodinamicamente mais estável, e de algoritmo eficiente capaz de varrer a superfície de energia (ou seja, gerar diversas novas conformações) (VERLI, H. 2014; RANGWALA; KARYPIS. 2010).

Além da modelagem molecular, existe a técnica de ancoragem (*docking*), cujo objetivo principal é a predição do acoplamento de estruturas tridimensionais moleculares, produzindo conformações estruturais candidatas plausíveis. Estas estruturas são então classificadas utilizando métodos com funções de pontuação para identificar as estruturas mais prováveis de ocorrer na natureza (MOREIRA *et al.*, 2010; VAKSER *et al.*, 2014). O reconhecimento molecular proteína-ligante nas simulações de *docking* está baseado na complementaridade de características físico-químicas e estruturais das moléculas do complexo. As características físico-químicas definem o grau de afinidade e de especificidade do ligante pela proteína, e estão

relacionadas com as interações intermoleculares existentes no complexo. As características estruturais, por sua vez, estão associadas aos arranjos espaciais moleculares, dados por variações na orientação, posicionamento espacial e rotações de ligações químicas das moléculas interagentes (VERLI, H. 2014).

O termo *docking* se originou no final de 1970, com significado mais restrito, que na época significava um modelo de estrutura complexa, otimizando a separação entre os interagentes, mas mantendo as suas orientações fixas, sendo reconhecido como *docking* rígido. Com o aumento do poder computacional, tornou-se possível trabalhar com alterações na geometria interna dos componentes do complexo, simulando alterações que podem ocorrer quando o complexo é formado em uma situação biológica. Este tipo de modelagem é conhecido como *docking* flexível. O problema do *docking* proteína-proteína é um dos pontos principais da atualidade em biologia e biofísica computacional, pois a estrutura tridimensional do complexo proteína-proteína, em geral, é mais difícil de ser determinada experimentalmente do que a estrutura de uma proteína individual (MOREIRA *et al.*, 2010; VAKSER *et al.*, 2014).

Após a obtenção dos modelos desejados *in silico*, tanto por meios experimentais quanto por modelagem e técnicas de *docking*, estes são submetidos à dinâmica molecular (DM), que tem como objetivo observar se tais modelos são capazes de suportar as condições que a mesma proteína estaria sujeita em um sistema biológico. A dinâmica molecular é a base do que poderia ser definido como um "laboratório virtual", usado para estudar processos complexos em nível molecular, incluindo a difícil tarefa de criar e destruir ligações químicas, por meio de métodos teóricos (KUKOL, A. 2008).

A lista a seguir inclui a variedade de maneiras em que a DM é utilizada:

- Fluidos complexos: líquidos moleculares, água pura e soluções aquosas, cristais líquidos, líquidos iônicos, interfaces fluidas, filmes e monocamadas.
- Polímeros: cadeias, anéis e moléculas ramificadas, conformação de equilíbrio, processos de relaxamento e transporte.
- Sólidos: fratura, transformações estruturais, danos por radiação, propriedades mecânicas elásticas e plásticas, fricção, ondas de choque, cristais moleculares.
- Biomoléculas: estrutura e dinâmica de proteínas, dobramento (*fold*ing) de proteínas, micelas, membranas, *docking* de moléculas (KUKOL, A. 2008).

Simulações por dinâmica molecular são ferramentas importantes para a compreensão da base física da estrutura e função de macromoléculas biológicas. A visão antiga de proteínas como estruturas relativamente rígidas tem sido substituída por modelo dinâmico em que os movimentos internos e mudanças conformacionais resultantes desempenham papel essencial na sua função. Métodos de simulação de dinâmica molecular são amplamente utilizados para obter informações sobre a evolução temporal das conformações de proteínas e outras macromoléculas biológicas, além da cinética e informação termodinâmica. Essa técnica é utilizada para quantificar as propriedades do sistema com uma precisão e em uma escala de tempo que é de outro modo inacessível, e é, portanto, ferramenta valiosa para nossa compreensão e validação computacional dos modelos de um sistema (KARPLUS *et al.*, 2002; ADCOCK *et al.*, 2006).

Com base em todas as informações expostas acima, pode-se perceber como a bioinformática e suas metodologias de modelagem estrutural, *docking* e dinâmica molecular são extremamente importantes na elucidação de problemas biológicos. Tais metodologias foram aplicadas no presente trabalho para a resolução da conformação tridimensional e estabilidade dinâmica da proteína HLA-G e suas formas solúveis, auxiliando a compreensão da apresentação de antígenos por tais molécula, além da interação dessas moléculas com receptores do sistema imune.

## 2 HIPÓTESE

Considerando que *i)* todos os segmentos da molécula de HLA-G e suas isoformas solúveis podem contribuir para sua capacidade supressora e interação com receptores do sistema imune, *ii)* só existem estruturas cristalográficas incompletas para HLA-G, *iii)* não existem estruturas resolvidas ou modelos tridimensionais propostos para as isoformas de HLA-G; neste estudo a determinação de suas estruturas tridimensionais completas e estabilidade estrutural dinâmica *in silico* pode elucidar características que influenciam na apresentação de antígenos por tais molécula, além da interação dessa molécula com receptores do sistema imune.

### 3. OBJETIVOS

#### 3.1 Objetivo geral

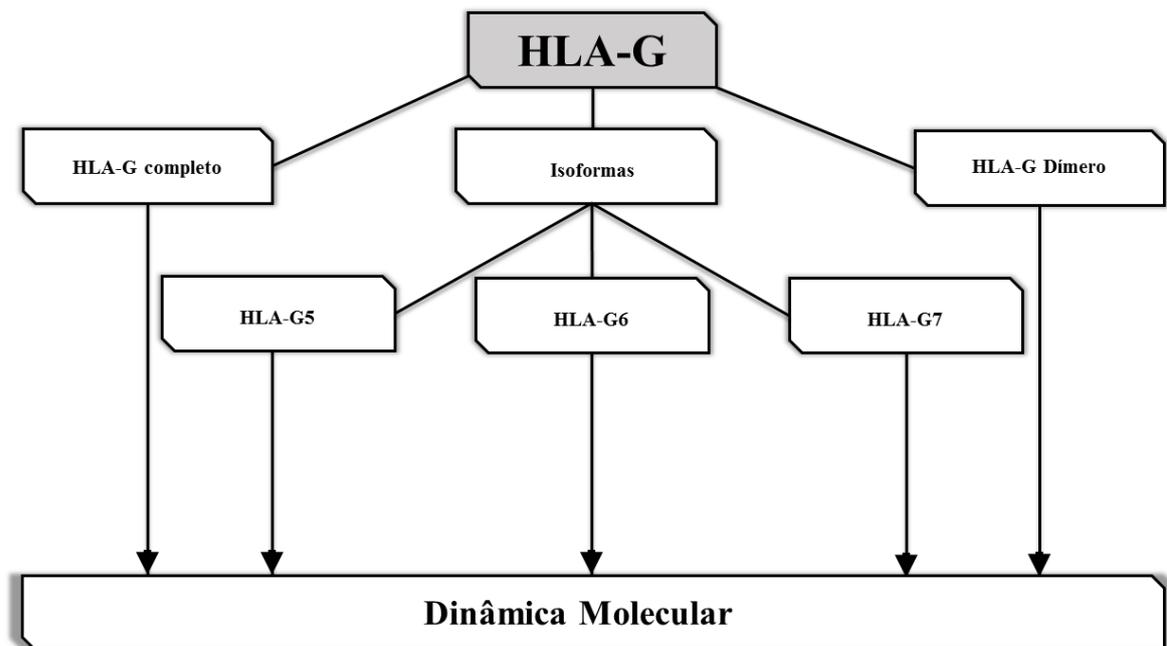
Determinar os modelos completos e estabilidade dinâmica *in silico* da proteína HLA-G e suas isoformas solúveis em diversas conformações estruturais, que poderão ser utilizados como molde para modelagem *in silico* dos demais alelos e isoformas produzidas, além da realização de *screenings* virtuais para interação com possíveis receptores do sistema imune e moléculas relacionadas ao HLA-G ainda não descritas.

#### 3.2 Objetivos específicos

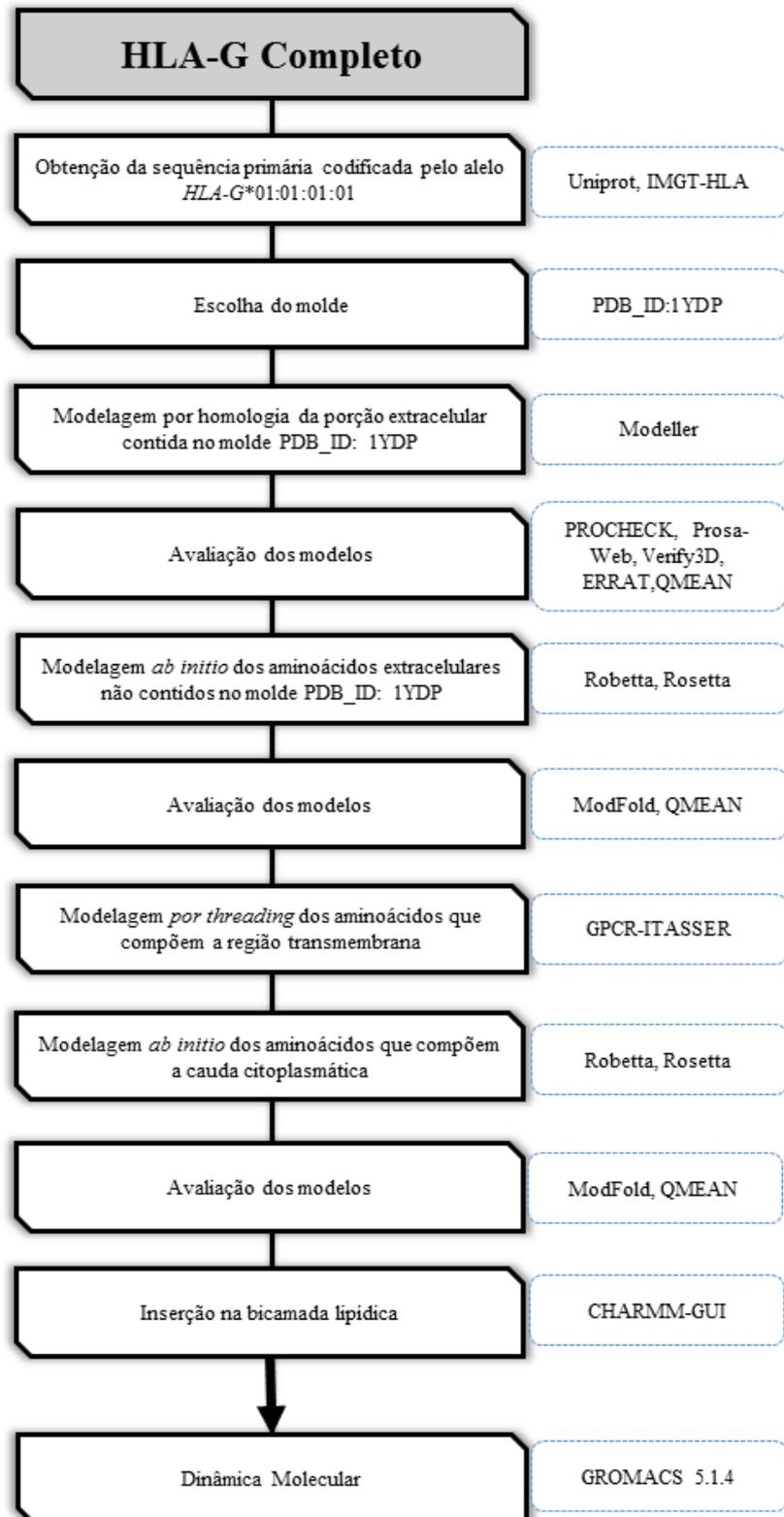
1. Por meio de metodologias *in silico*, avaliar as estruturas tridimensionais da proteína HLA-G completa, seu dímero solúvel e isoformas HLA-G5, HLA-G6 e HLA-G7 geradas por meio de modelagem por homologia e *ab initio*.
2. Avaliar possíveis estruturas diméricas formadas pela proteína HLA-G7.
3. Avaliar a estabilidade de todos os modelos produzidos utilizando simulações de dinâmica molecular.

#### 4 FLUXOGRAMAS DAS ANÁLISES *IN SILICO*

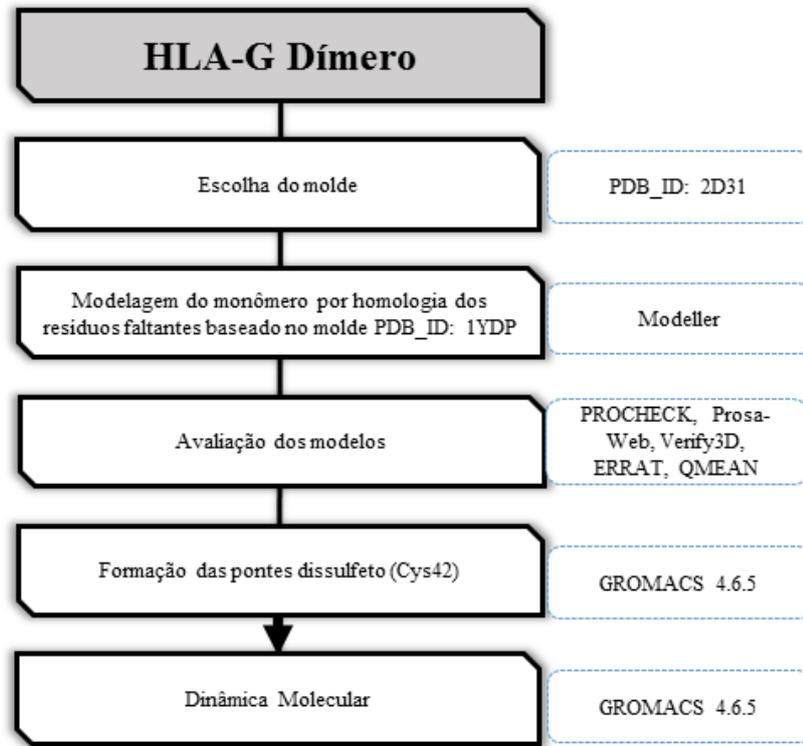
As figuras, a seguir, representam os fluxogramas contendo as análises *in silico* que foram empregadas neste trabalho. A **Figura 7** demonstra o fluxograma geral das estruturas estudadas no presente trabalho. A **Figura 8** demonstra as etapas necessárias para obtenção do modelo da proteína codificada pelo alelo *HLA-G\*01:01:01:01*, inserção na bicamada lipídica e simulação de dinâmica molecular, enquanto a **Figura 9** ilustra as etapas desenvolvidas para obtenção do dímero completo de HLA-G1 e simulação de dinâmica molecular. A **Figura 10** ilustra a obtenção dos modelos e análises desenvolvidas para estudo das isoformas HLAG5, HLA-G6 e HLA-G7.



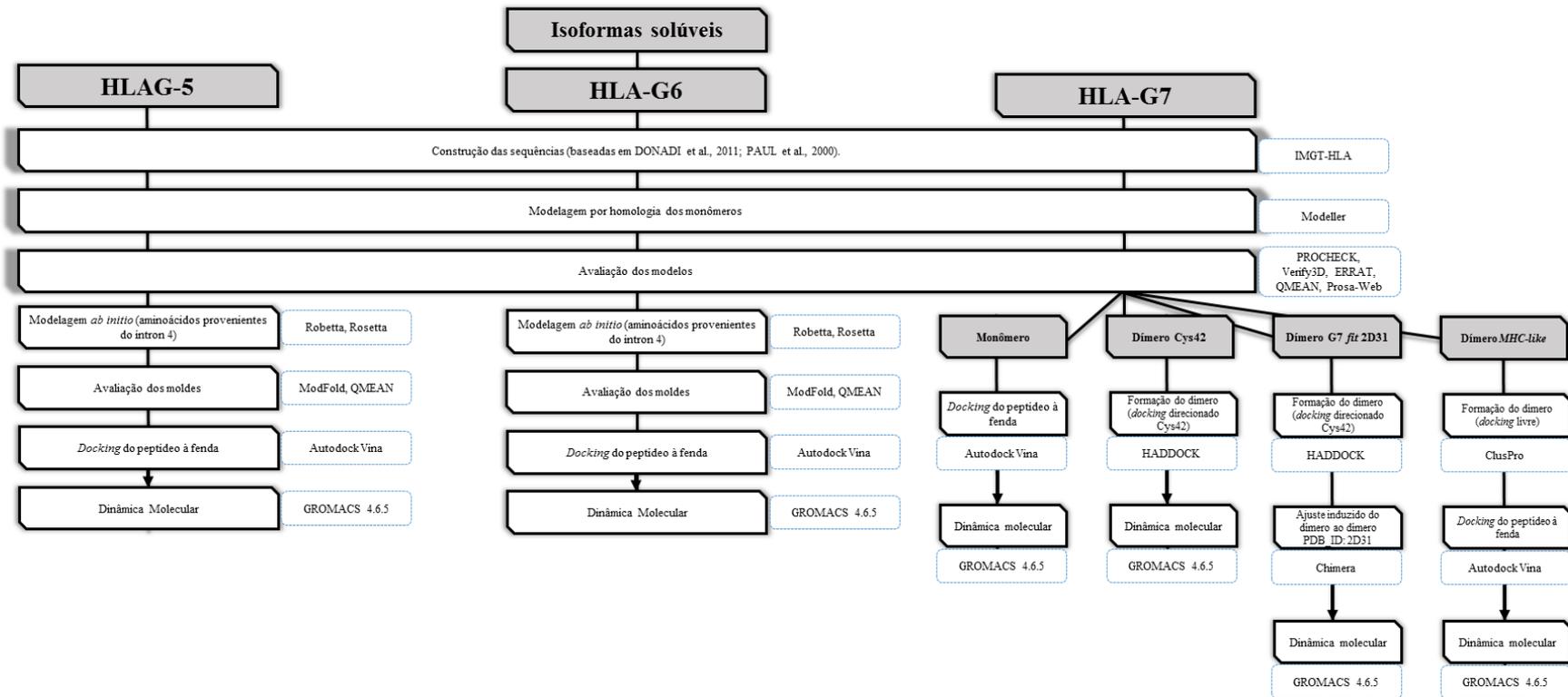
**Figura 7.** Fluxograma geral das estruturas modeladas para a molécula de HLA-G contidas na tese.



**Figura 8.** Fluxograma detalhado das ferramentas utilizadas para obtenção da estrutura completa de HLA-G inserida na bicamada lipídica. Caixas com bordas pretas mostram os processos realizados, caixas com bordas azuis pontilhadas mostram as ferramentas computacionais utilizadas em todas as etapas das análises.



**Figura 9.** Fluxograma detalhado das ferramentas utilizadas para obtenção da estrutura do dímero de HLA-G1. Caixas com bordas pretas mostram os processos realizados, caixas com bordas azuis pontilhadas mostram as ferramentas computacionais utilizadas em todas as etapas das análises.



**Figura 10.** Fluxograma detalhado das ferramentas utilizadas para obtenção das estruturas completas das isoformas HLA-G5, HLA-G6 e HLA-G7. Caixas com bordas pretas mostram os processos realizados, caixas com bordas azuis pontilhadas mostram as ferramentas computacionais utilizadas em todas as etapas das análises.

## 5 MATERIAL E MÉTODOS

### 5.1 Obtenção das sequências

A sequência de aminoácidos, DNA e RNA referente ao alelo *HLA-G\*01:01:01:01* utilizada para o desenvolvimento do presente trabalho foi obtida do banco de dados IMGT/HLA (IMGT/HLA. Disponível em <http://www.ebi.ac.uk/cgi-bin/ipd/imgt/hla/align.cgi> banco de dados versão 3.22; ROBINSON *et al.*, 2013). A montagem das sequências das isoformas solúveis foram baseadas em artigos publicados na literatura (DONADI *et al.*, 2011; PAUL *et al.*, 2000), baseadas na sequência do alelo *HLA-G\*01:01:01:01* e confirmada com o depósito das mesmas no banco de dados UniProt (THE UNIPROT CONSORTIUM, 2017).

### 5.2 Identificação da sequência referente ao peptídeo sinal

A sequência primária da proteína codificada pelo alelo *HLA-G\*01:01:01:01* utilizada para o desenvolvimento do presente trabalho foi investigada utilizando os servidores *online* Phobius (KÄLL *et al.*, 2007), PrediSi (HILLER *et al.*, 2004) e SignalIP (NIELSEN *et al.*, 2017), para predição da localização de peptídeo sinal, porção transmembrana e porção extracelular que compõem a proteína HLA-G.

### 5.3 Perfil de hidrofobicidade do peptídeo sinal de HLA-G

O servidor *online* PRALINE (SIMOSSIS *et al.*, 2005) facilita a visualização abrangente de alinhamentos gerados por meio de cinco esquemas de cores padrão com base em: tipo de resíduo, conservação de posição, confiabilidade de posição, hidrofobicidade de resíduos e estrutura secundária, dependendo das opções definidas. No presente trabalho foi utilizado para determinar o perfil de hidrofobicidade dos resíduos que compõem a sequência do peptídeo sinal de HLA-G.

#### 5.4 Alinhamento de sequências proteicas

O alinhamento múltiplo das sequências utilizadas no presente trabalho foi realizado utilizando a ferramenta CLUSTAL OMEGA (SIEVERS *et al.*, 2011). Alinhamentos locais foram realizados utilizando a ferramenta EMBOSS WATER, enquanto alinhamentos globais foram realizados utilizando a ferramenta EMBOSS NEEDLE (LI *et al.*, 2015). Nas imagens de alinhamento os seguintes significados são atribuídos aos símbolos:

- \* indica posições que possuem um único resíduo conservado;
- : indica conservação entre grupos de propriedades fortemente similares;
- . indica conservação entre grupos de propriedades fracamente similares.

#### 5.5 Modelagem *in silico* por homologia

Para modelagem comparativa da proteína HLA-G, porções do seu dímero solúvel e das isoformas solúveis foi utilizado o *software* Modeller 9.15 (WEBB *et al.*, 2002), no qual o usuário fornece o alinhamento de uma sequência a ser modelada com estruturas relacionadas conhecidas e o *software* calcula automaticamente um modelo que contém todos os átomos que não sejam o hidrogênio.

Modeller utiliza a linguagem de programação Python (disponível em <https://www.python.org/>) e implementa modelagem estrutural comparativa pela satisfação de restrições espaciais e pode executar muitas tarefas adicionais, incluindo modelagem *de novo* de *loops*, otimização dos diferentes modelos de estrutura da proteína em relação a uma função definida de forma flexível, alinhamento múltiplo de sequências de proteínas e/ou estruturas, *clustering*, pesquisa de bases de dados de sequência, comparação de estruturas de proteínas, entre outras funções.

## 5.6 Avaliação da qualidade dos modelos gerados

Após a geração dos modelos pelo *software* Modeller 9.15, cinco ferramentas de validação foram utilizadas. Cada ferramenta foi responsável por avaliar diferentes propriedades: Verify 3D (LUTHY *et al.*, 1992), ERRAT (COLOVOS e YATES, 1993), PROCHECK (LASKOWSKI *et al.*, 1993), QMEAN (BENKERT *et al.*, 2008) e Prosa-Web (WIEDERSTEIN e SIPPL, 2007).

Verify 3D avalia os ambientes químicos dos resíduos para conferir a compatibilidade entre a sequência de aminoácidos e o modelo gerado. A ferramenta analisa a qualidade e confiança do enovelamento do modelo, definindo os ambientes químicos de cada resíduo do modelo para atribuir *scores* com referência a uma matriz estabelecida a partir da análise estatística de estruturas de proteínas provenientes do PDB. O valor de *score* determina se o modelo gerado é compatível com sua sequência primária de aminoácidos e deve ter mais de 80% de compatibilidade entre modelo e sequência (LUTHY *et al.*, 1992).

ERRAT foi, inicialmente, desenvolvido para a avaliação de estruturas de proteínas obtidas pelo método experimental de difração de raios-X, mas tem sido amplamente utilizado para a avaliação de estruturas tridimensionais obtidas por meio de diversas metodologias experimentais e *in silico*. A ferramenta avalia a melhoria do refinamento proteico através da análise de interação entre átomos não ligados e determina a probabilidade de erro em relação à disposição do aminoácido, por meio de cálculos que compararam o modelo avaliado com valores estatísticos de estruturas altamente refinadas e confiáveis (COLOVOS e YATES, 1993).

O servidor *online* PDBSum (DE BEER *et al.*, 2014) foi utilizado para ferramenta PROCHECK (LASKOWSKI *et al.*, 1993), que avalia a qualidade estereoquímica dos modelos por meio de avaliação dos comprimentos de ligação, ângulos planos, planaridade dos anéis de cadeias laterais, planaridade das ligações peptídicas, ângulos torcionais da cadeia principal e das cadeias laterais, quiralidade, conformações das cadeias laterais, impedimento estérico entre pares de átomos não-ligados e a qualidade do gráfico de Ramachandran. A análise dos resultados apresentados pelo gráfico de Ramachandran gerado pelo *software* PROCHECK determina se os aminoácidos se encontram em regiões energeticamente favoráveis ou desfavoráveis, orientando a avaliação da qualidade dos modelos gerados. O *score* desejável para que um modelo seja positivamente avaliado é que >90% dos aminoácidos que compõem

a proteína estejam localizados em regiões favoráveis do gráfico (LASKOWSKI *et al.*, 1993). No gráfico de Ramachandran é possível visualizar os dois ângulos de torção  $\phi$  (phi) e  $\psi$  (psi) dos aminoácidos, que determinam a conformação da estrutura tridimensional das proteínas (RAMACHANDRAN *et al.*, 1963).

ProSA-Web calcula a pontuação local e global de qualidade para uma estrutura tridimensional, realizando cálculos com potenciais de força média da estrutura a ser avaliada com base no conhecimento de estruturas de proteínas depositadas no PDB obtidas com boa taxa de resolução por métodos experimentais. Os valores referentes à avaliação local são expressos em um gráfico com padrões de energia e os valores da pontuação global são apresentados em um gráfico com a pontuação das estruturas provenientes do PDB. Essa pontuação é apresentada como *Z-score*, cujo valor negativo representa que o modelo avaliado tem boa qualidade e pouca probabilidade de erro; entretanto, valores positivos representam modelos de baixa qualidade (WIEDERSTEIN e SIPPL, 2007).

A ferramenta QMEAN consiste em uma função de pontuação composta tanto para a estimativa da qualidade global do modelo quanto para a análise de resíduos locais em diferentes regiões dentro do modelo (BENKERT *et al.*, 2008).

## 5.7 Modelagem *in silico* por metodologia *ab initio*

Para a região extracelular não resolvida na estrutura da proteína HLA-G foi utilizada a metodologia de modelagem *ab initio*, utilizando predição de estruturas proteicas a partir da sequência de aminoácidos. Tais regiões foram modeladas utilizando o algoritmo *cyclic coordinate descent* (CCD). O enovelamento da proteína é ditado por forças físicas agindo nos átomos que a compõem e esse método é considerado de maior acurácia (HUMPFREY *et al.*, 1996). As conformações que possuem a menor energia livre são consideradas as mais prováveis de serem adotadas pela proteína em condições nativas. O servidor *online* Robetta foi utilizado para geração da biblioteca de fragmentos e o protocolo Rosetta (PETTERSEN *et al.*, 2004) (servidor *online* Rosetta) para construção dos modelos *ab initio*.

## 5.8 Modelagem *in silico* por metodologia de *threading*

Para modelagem da porção transmembrana foi utilizado o servidor *online* GPCR-ITASSER (ZHANG *et al.*, 2015), método computacional projetado para a predição de estrutura 3D de receptores de membrana. A sequência de destino é primeiro indexada através da biblioteca do PDB para busca por modelos putativos. Os modelos homólogos identificados são utilizados no procedimento de montagem de fragmentos para construção de modelos completos. Caso não existam modelos homólogos disponíveis, é realizada modelagem *ab initio* de dobra para hélices a partir do zero, gerando 5 modelos finais.

As pontuações utilizadas para mensurar a qualidade dos modelos gerados são *TM-score* e *C-score*. *TM-score* é uma escala recentemente proposta para medir a semelhança estrutural entre duas estruturas (ZHANG, SKOLNICK *et al.*, 2004). O objetivo é resolver o problema do RMSD que é sensível ao erro local. Como o RMSD é uma distância média de todos os pares de resíduos em duas estruturas, um erro local (por exemplo, uma orientação incorreta da cauda) aumentará o valor de RMSD de forma expressiva, embora a topologia global esteja correta. Utilizando o valor *TM-score*, no entanto, a pontuação se torna insensível ao erro de modelagem local. Um *TM-score* >0.5 indica um modelo de topologia correta.

*C-score* é o *score* de confiança para estimar a qualidade dos modelos preditos pelo servidor GPCR-I-TASSER. É calculado com base na significância do alinhamento dos modelos utilizados para *threading* e nos parâmetros de convergência das simulações de montagem da estrutura. Valores de *C-score* geralmente estão na faixa de -5.0 a 2.0, onde uma pontuação de maior valor significa um modelo com alta confiança, e vice-versa.

## 5.9 Modelagem *in silico* por metodologia *ab initio* das isoformas HLA-G5, HLA-G6 e HLA-G7

Para as caudas de 21 aminoácidos provenientes da tradução do íntron 4 presentes nas isoformas HLA-G5 e HLA-G6, e os 2 aminoácidos terminais provenientes da tradução do íntron 2 na isoforma HLA-G7, foi utilizada a metodologia de modelagem *ab initio*, utilizando predição de estruturas proteicas a partir da sequência de aminoácidos. Tais regiões foram modeladas

utilizando o algoritmo *cyclic coordinate descent* (CCD). O enovelamento da proteína é ditado por forças físicas agindo nos átomos que a compõem e esse método é considerado de maior acurácia (HUMPFREY *et al.*, 1996). As conformações que possuem a menor energia livre são consideradas as mais prováveis de serem adotadas pela proteína em condições nativas. O servidor *online* Robetta foi utilizado para geração da biblioteca de fragmentos e o protocolo Rosetta (PETTERSEN *et al.*, 2004) (servidor *online* Rosetta) para construção dos modelos *ab initio*.

### 5.10 Seleção dos modelos gerados por modelagem *ab initio*

Foram gerados 2000 modelos em cada etapa de modelagem *ab initio*. Os arquivos .pdb gerados foram avaliados utilizando as ferramentas QMEAN (BENKERT, *et al.*, 2008) e MODFOLD (MCGUFFIN *et al.*, 2008). O valor utilizado para verificar a qualidade do modelo na plataforma QMEAN foi o *QMEANnorm\_score* e os valores utilizados para verificar a qualidade do modelo utilizando MODFOLD foram *Global Model Quality score* e *p-value*. A escolha do melhor modelo foi baseada nos melhores resultados das ferramentas QMEAN e MODFOLD.

### 5.11 Inserção da estrutura completa de HLA-G na bicamada lipídica

A plataforma *online* CHARMM-GUI *Membrane Builder* auxilia o usuário a gerar uma série de arquivos de entrada necessários para construção de um complexo proteína:membrana para simulações de dinâmica molecular, facilitando e padronizando o uso de técnicas de simulação comuns e avançadas. O ambiente fornece uma plataforma ideal para construir e validar um modelo molecular de forma interativa. A bicamada utilizada no trabalho foi composta por fosfolipídios (JO *et al.*, 2008).

### 5.12 *Docking* do peptídeo RIIPRHLQL no sulco da proteína HLA-G e isoformas HLA-G5, HLA-G6 e HLA-G7

O protocolo de *docking* molecular do nonapeptídeo RIIPRHLQL ao sulco de apresentação de antígeno da proteína HLA-G completa e das isoformas HLA-G5, HLA-G6 e HLA-G7 foi realizado utilizando o *software* Autodock Vina, utilizado para *docking* de pequenas moléculas e fármacos (TROTT *et al.*, 2010). Cada experimento *in silico* de *docking* foi executado 10 vezes, produzindo em cada rodada 10 possíveis conformações para o encaixe do peptídeo no sulco. Todas as simulações de *docking* foram executadas com exaustividade= 1000.

### 5.13 *Docking* proteína-proteína para formação dos dímeros de HLA-G7

O protocolo de *docking* molecular utilizando monômeros da isoforma HLA-G7 para formação dos diversos dímeros apresentados neste trabalho foram realizados utilizando os servidores *online* HADDOCK (*High Ambiguity Driven protein-protein DOCKing*) (van ZUNDERT *et al.*, 2016) e ClusPro (KOZAKOV *et al.*, 2017). O servidor HADDOCK utiliza abordagem de *docking* flexível baseada em informações já conhecidas para formação complexos biomoleculares. HADDOCK distingue-se dos métodos de *docking ab initio* no fato de que utiliza informações de interfaces de proteínas identificadas ou previstas para direcionar o processo de *docking*, podendo lidar com uma grande variedade de tipos de *docking*, incluindo proteína-proteína, proteína-ácidos nucleicos e complexos proteína-ligante. A avaliação dos resultados é baseada no HADDOCK *score* e *z-score*. Quanto menores os valores obtidos para o *z-score*, melhor o resultado do *docking*. Foi utilizado de forma direcionada com a seleção dos resíduos envolvidos na interação, para formação do dímero Cys42 e *docking* da proteína  $\beta$ 2-microglobulina à isoforma HLA-G6.

O servidor *online* ClusPro utiliza metodologia de *docking* rígido, mantendo o receptor rígido e o ligante móvel, podendo realizar *docking* sem informações *a priori* do complexo de interesse. O *ranking* dos resultados é baseado na população dos *clusters* formados, não no *score* de energia. Foi utilizado de forma exploratória para os dímeros semelhantes à fenda do HLA. Tais protocolos de *docking* proteína-proteína permitem diversas possibilidades com relação à

flexibilidade da ancoragem proteica, além de ferramentas para análise dos resultados obtidos com o *docking*.

#### 5.14 Ajuste induzido da isoforma HLA-G7

A manipulação da orientação espacial dos monômeros da isoforma HLA-G7 para obtenção do dímero alternativo da isoforma HLA-G7 formado por ponte dissulfeto entre as cisteínas na posição 42 (Cys42) de cada monômero (identificado como *G7 fit* 2D31), foi realizada utilizando o *software* Chimera (PETTERSEN *et al.*, 2004). Tal estrutura foi obtida através do ajuste induzido do dímero formado pela isoforma HLA-G7 ao dímero solúvel de HLA-G, baseado na estrutura cristalográfica PDB\_ID: 2D31 publicada por (SHIROISHI *et al.*, 2006) (PDB\_ID: 2D31).

#### 5.15 Simulações de Dinâmica Molecular

A simulação de dinâmica molecular (DM) clássica utiliza a descrição newtoniana das forças envolvidas entre os elementos que constituem o sistema a ser simulado. Essa metodologia calcula a posição atômica de cada átomo que compõem o sistema em diferentes e sucessivos momentos ao longo do tempo, separados por intervalos temporais previamente determinados. Este conjunto de coordenadas atômicas do sistema e sua evolução temporal descrevem a trajetória do sistema, proteína-solvente, a ser analisada.

A dinâmica molecular consistiu em uma simulação para cada modelo gerado, de no mínimo 100ns. Foi utilizado o *software* GROMACS (HESS *et al.*, 2008), nas versões 4.6.5 (para modelos unicamente compostos de estruturas proteicas) e 5.1.4 (simulação da proteína HLA-G inserida na bicamada lipídica), que implementam o campo de força G54a7 (para modelos unicamente compostos de estruturas proteicas) e CHARMM36m (simulação da proteína HLA-G inserida na bicamada lipídica) (HUANG *et al.*, 2017).

A DM simulou o sistema proteína-solvente em condições físico-químicas controladas, com temperatura máxima de 310 K (ou seja 36.85 °C) e pressão constante de 1 atm. A estrutura

tridimensional de cada proteína foi solvatada em uma caixa cúbica contendo modelo de água SPC (*simple point charge*) com espessura de 20 Å a mais em cada direção nos eixos  $x$ ,  $y$  e  $z$  partindo das coordenadas do átomo mais distante em cada direção. Foram adicionados íons  $\text{Na}^+$  e  $\text{Cl}^-$  suficientes para o equilíbrio do sistema, com concentração final de 0,15 mol / L.

Para todas as simulações de DM, o estágio de produção foi precedido por (i) três etapas de Minimização de Energia e (ii) oito etapas de Equilíbrio. O estágio de Equilíbrio iniciou com restrições de posição e temperatura de 310 K, por um período de 300 ps, para permitir a formação de camadas de solvatação. A temperatura foi então reduzida para 280 K e as restrições de posição foram gradualmente reduzidas. Este processo foi seguido por um aumento gradual da temperatura, até 300 K. Juntos, esses passos de Equilíbrio representam os primeiros 500 ps de cada simulação. Durante o estágio de produção, o sistema foi mantido a temperatura constante (300 K), sem restrições.

Para cada simulação foram calculados os valores de RMSD (*Root Mean Square Deviation*) dos átomos  $\text{Ca}$  e a flutuação da média quadrática RMSF (*Root Mean Square Fluctuation*) dos átomos  $\text{Ca}$  utilizando a estrutura inicial como referência.

As trajetórias obtidas na dinâmica molecular foram analisadas com o auxílio dos softwares GROMACS (HESS *et al.*, 2008) e VMD (HUMPHREY *et al.*, 1996), utilizados para interpretar os dados das trajetórias completas.

As simulações de dinâmica molecular foram realizadas em colaboração com a Prof<sup>a</sup>. Lydia Kavradi e Dr. Dinler Antunes, utilizando o servidor NOTS (*Night Owls Time-Sharing Service*) da Rice University em Houston, Texas, Estados Unidos.

## 5.16 Resíduos presentes na interface de interação

Com o intuito de verificar os resíduos que fazem parte das interações proteína-proteína ou proteína-peptídeo, foi utilizado o pacote BioPython (COCK *et al.*, 2009).

### 5.17 Visualização dos Modelos

Os modelos obtidos foram visualizados através dos *softwares* Chimera (PETTERSEN *et al.*, 2004) e PyMol (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC).

## 6. RESULTADOS

### 6.1 Molécula de histocompatibilidade de classe I não clássica HLA-G

#### 6.1.1 Análise da sequência primária da proteína HLA-G

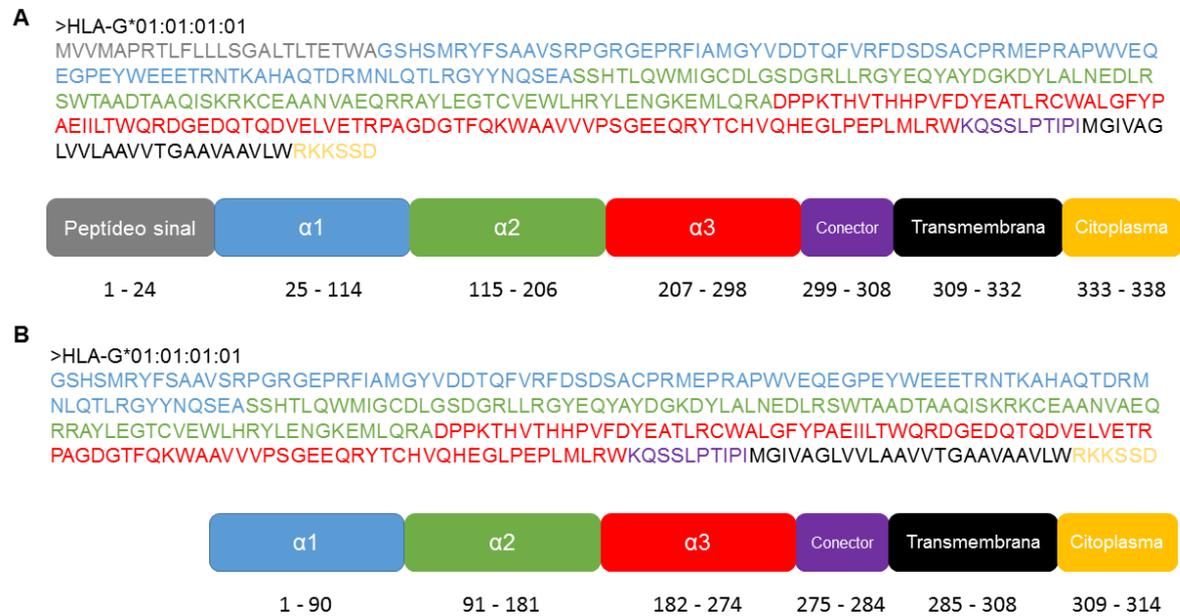
Para que fosse possível a modelagem completa da proteína de histocompatibilidade de classe I não clássica HLA-G, o primeiro passo executado foi o estudo minucioso de sua sequência primária, obtida utilizando o banco de dados IMGT-HLA (IMGT/HLA. Disponível em <http://www.ebi.ac.uk/cgi-bin/ipd/imgt/hla/align.cgi> banco de dados versão 3.31.0; ROBINSON *et al.*, 2013). Dentre os alelos existentes para *HLA-G*, a sequência primária codificada pelo alelo *HLA-G\*01:01:01:01* foi selecionada para todas as modelagens estruturais, cuja sequência primária completa no formato FASTA<sup>1</sup> engloba 338 resíduos. Esse alelo e sequência foram utilizados como referência para todas as modelagens da proteína HLA-G completa e para construção das sequências das isoformas solúveis HLA-G5, HLA-G6 e HLA-G7, que serão discutidas mais adiante.

A análise minuciosa da sequência da proteína HLA-G, utilizando servidores *online* Phobius (KÄLL *et al.*, 2007), PrediSi (HILLER *et al.*, 2004) e SignalIP (NIELSEN *et al.*, 2017) revelou que essa estrutura possui um peptídeo sinal (resíduos 1 - 24), porção extracelular (25 - 308), porção transmembrana (309 - 332) e pequena cauda citoplasmática (333 - 338).

Além disso, informações encontradas no banco de dados UniProt (THE UNIPROT CONSORTIUM, 2017) indicam os resíduos que compõem os 3 domínios extracelulares da proteína HLA-G ( $\alpha$ 1, 25 – 114;  $\alpha$ 2, 115 – 206;  $\alpha$ 3, 207 – 298). Podemos observar essas informações citadas acima de forma esquemática na **Figura 11**.

---

<sup>1</sup> <http://www.ncbi.nlm.nih.gov/blast/blastcgihelp.shtml>

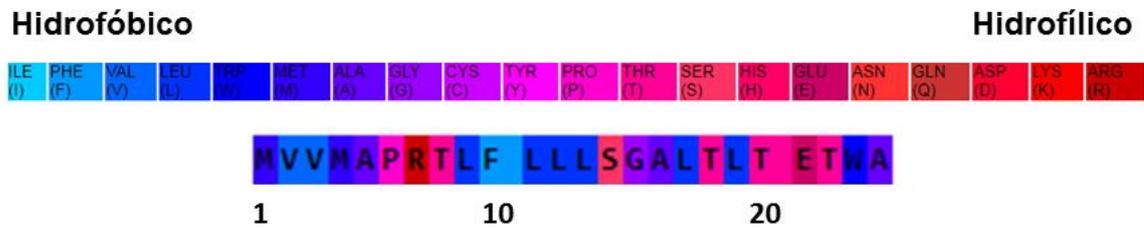


**Figura 11.** (A) Sequência primária completa de HLA-G no formato FASTA<sup>1</sup>. As cores são referentes às porções da proteína compostas por tais aminoácidos, segundo esquema na porção inferior da figura (cinza – peptídeo sinal, azul – domínio  $\alpha 1$ , verde - domínio  $\alpha 2$ , vermelho - domínio  $\alpha 3$ , roxo - região conectora, preto – região transmembrana e amarelo – cauda citoplasmática). (B) Sequência primária da proteína HLA-G madura no formato FASTA<sup>1</sup>.

A determinação das regiões que compõem a proteína HLA-G e sua sequência foi de extrema importância para as próximas etapas de modelagem molecular que foram desenvolvidas.

Após a identificação da região relacionada ao peptídeo sinal e sua remoção, a sequência primária da proteína HLA-G completa possui 314 aminoácidos ( $\alpha 1$ , 1 – 90;  $\alpha 2$ , 91 – 181;  $\alpha 3$ , 182 – 274; peptídeo conector 275 – 284; região transmembrana 285 - 308; cauda citoplasmática 309 - 314). A partir deste ponto, utilizaremos a sequência e numeração dos resíduos referentes à proteína madura, sem o peptídeo sinal.

O servidor *online* PRALINE (SIMOSSIS *et al.*, 2005) facilita a visualização abrangente de alinhamentos gerados por meio de cinco esquemas de cores padrão com base em: tipo de resíduo, conservação de posição, confiabilidade de posição, hidrofobicidade de resíduos e estrutura secundária, dependendo das opções definidas. No presente trabalho, esse servidor foi utilizado para determinar o perfil de hidrofobicidade dos resíduos que compõem a sequência do peptídeo sinal de HLA-G.



**Figura 12.** Perfil de hidrofobicidade dos resíduos que compõem o peptídeo sinal da proteína HLA-G (SIMOSSIS *et al.*, 2005).

Os resíduos 1, 2, 3, 4, 5, 9, 10, 11, 12, 13, 17, 19 e 23 apresentam características hidrofóbicas.

## 6.1.2 Modelagem molecular da proteína HLA-G

### 6.1.2.1 Escolha do molde

De posse da sequência primária da proteína HLA-G após a remoção da sequência referente ao peptídeo sinal, foi realizada busca por possíveis estruturas que pudessem servir como molde para sua modelagem por homologia, utilizando a sequência primária demonstrada anteriormente e o banco de dados PDB (BERMAN *et al.*, 2002).

As seguintes estruturas foram encontradas:

**Tabela 1.** Possíveis moldes para modelagem por homologia da proteína HLA-G.

PDB_ID	Alelo	Peptídeo ligado	Técnica experimental	Resolução (Å)	Resíduos	Referência
1YDP	HLA-G*01:04	RIIPRHLQL	Difração de raio-X	1.9	275	CLEMENTS <i>et al.</i> , 2005.
3KYN	HLA-G*01:04	KGPPAALTL	Difração de raio-X	2.4	275	WALPOLE <i>et al.</i> , 2010.
3KYO	HLA-G*01:04	SKLPAQFYIL	Difração de raio-X	1.7	273	WALPOLE <i>et al.</i> , 2010.

As sequências das estruturas disponíveis foram alinhadas com a sequência primária de HLA-G, para escolha do melhor molde (**Figura 13**).

```

HLA-G*01:01:01:01      MVVMAPRTLFLLLSGALTLTETWAGSHSMRYFSAAVSRPGRGEPFRFIAMGYVDDTQFVRF
3KYN                   -----MSHSMRYFSAAVSRPGRGEPFRFIAMGYVDDTQFVRF
1YDP                   -----SHSMRYFSAAVSRPGRGEPFRFIAMGYVDDTQFVRF
3KYO                   -----SHSMRYFSAAVSRPGRGEPFRFIAMGYVDDTQFVRF
                        *****

HLA-G*01:01:01:01      DSDSACPRMEPRAPWVEQEGPEYWEETRNKKAHAQTD RMNLQTLRGYYNQSEASSHTLQ
3KYN                   DSDSASPRMEPRAPWVEQEGPEYWEETRNKKAHAQTD RMNLQTLRGYYNQSEASSHTLQ
1YDP                   DSDSASPRMEPRAPWVEQEGPEYWEETRNKKAHAQTD RMNLQTLRGYYNQSEASSHTLQ
3KYO                   DSDSASPRMEPRAPWVEQEGPEYWEETRNKKAHAQTD RMNLQTLRGYYNQSEASSHTLQ
                        *****

HLA-G*01:01:01:01      WMIGCDLGS DGRLLRGYEQYAYDGKDYALNEDLRSHWTAADTAAQISKRCKEAAANVAEQR
3KYN                   WMIGCDLGS DGRLLRGYERYAYDGKDYALNEDLRSHWTAADTAAQISKRCKEAAANVAEQR
1YDP                   WMIGCDLGS DGRLLRGYERYAYDGKDYALNEDLRSHWTAADTAAQISKRCKEAAANVAEQR
3KYO                   WMIGCDLGS DGRLLRGYERYAYDGKDYALNEDLRSHWTAADTAAQISKRCKEAAANVAEQR
                        *****

HLA-G*01:01:01:01      RAYLEGTCEWLHRYLENGKEMLQRADPPKTHVTHHPVFDYEATLR C WALGFYPAEIIIT
3KYN                   RAYLEGTCEWLHRYLENGKEMLQRADPPKTHVTHHPVFDYEATLR C WALGFYPAEIIIT
1YDP                   RAYLEGTCEWLHRYLENGKEMLQRADPPKTHVTHHPVFDYEATLR C WALGFYPAEIIIT
3KYO                   RAYLEGTCEWLHRYLENGKEMLQRADPPKTHVTHHPVFDYEATLR C WALGFYPAEIIIT
                        *****

HLA-G*01:01:01:01      WQRDGEDQTQDVELVETRPAGDGT FQKWAAVVPSGEEQRYTCHVQHEGLPEPLMLRWKQ
3KYN                   WQRDGEDQTQDVELVETRPAGDGT FQKWAAVVPSGEEQRYTCHVQHEGLPEPLMLRWKQ
1YDP                   WQRDGEDQTQDVELVETRPAGDGT FQKWAAVVPSGEEQRYTCHVQHEGLPEPLMLRWKQ
3KYO                   WQRDGEDQTQDVELVETRPAGDGT FQKWAAVVPSGEEQRYTCHVQHEGLPEPLMLRWKQ
                        *****

HLA-G*01:01:01:01      SSLPTIPIMGIVAGLVVLA AVVTGA AVAVLWRKSSD
3KYN                   -----
1YDP                   -----
3KYO                   -----

```

**Figura 13.** Alinhamento múltiplo da sequência primária da proteína codificada pelo alelo *HLA-G\*01:01:01:01* completa e sequências referentes às estruturas cristalográficas PDB\_ID:1YDP (CLEMMENTS *et al.*, 2005), PDB\_ID:3KYN e PDB\_ID:3KYO (WALPOLE *et al.*, 2010).

Levando em conta o número de resíduos resolvidos nas cristalografias e as resoluções das mesmas, a estrutura selecionada como molde foi a PDB\_ID: 1YDP. Tal estrutura é proveniente da cristalografia por difração de raio-X da proteína codificada pelo grupo de alelos *HLA-G\*01:04*, possui 275 resíduos resolvidos (aqui denominada cadeia A) e resolução de 1.9Å. Além disso, contém o nonapeptídeo RIIPRHLQL (aqui denominada cadeia P) localizado no sulco e está acoplada à  $\beta$ 2-microglobulina (aqui denominada cadeia B), que possui 99 resíduos. Essa estrutura foi resolvida por Clements e colaboradores e possui uma mutação na cisteína localizada na posição 42 (resíduo Cys42) (CLEMMENTS *et al.*, 2005).

O alinhamento da sequência primária madura de HLA-G com a sequência dos resíduos presentes na estrutura PDB\_ID:1YDP foi comparado utilizando as ferramentas de alinhamento local EMBOSS WATER, e a ferramenta de alinhamento global EMBOSS NEEDLE, para definição exata das porções que não foram resolvidas na cristalografia e foram modeladas por metodologias *in silico* (LI *et al.*, 2015).

Para o alinhamento local temos uma sequência de 275 resíduos, que possui 98.9% de identidade e 99.6% de similaridade entre a sequência primária de HLA-G e a sequência primária da estrutura cristalográfica PDB\_ID:1YDP.

<i>HLA-G*01:01:01:01</i>	2	SHSMRYFSAAVSRPGRGEPRFIAMGYVDDTQFVRFSDSACPRMEPRAPW	51
1YDP	1	SHSMRYFSAAVSRPGRGEPRFIAMGYVDDTQFVRFSDSASPRMEPRAPW	50
<i>HLA-G*01:01:01:01</i>	52	VEQEGPEYWEEETRNTKAHAQTDRMNLQTLRGYYNQSEASSHTLQWMIGC	101
1YDP	51	VEQEGPEYWEEETRNTKAHAQTDRMNLQTLRGYYNQSEASSHTLQWMIGC	100
<i>HLA-G*01:01:01:01</i>	102	DLGSDGRLLRGYEQYAYDGKDYLALNEDLRSWTAADTAAQISKRKCEAAN	151
1YDP	101	DLGSDGRLLRGYERYAYDGKDYLALNEDLRSWTAADTAAQISKRKCEAAN	150
<i>HLA-G*01:01:01:01</i>	152	VAEQRRAYLEGTCEVWLHRYLENGKEMLRADPPKTHVTHHPVFDYEATL	201
1YDP	151	VAEQRRAYLEGTCEVWLHRYLENGKEMLRADPPKTHVTHHPVFDYEATL	200
<i>HLA-G*01:01:01:01</i>	202	RCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGTQKWAAVVWVPS	251
1YDP	201	RCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGTQKWAAVVWVPS	250
<i>HLA-G*01:01:01:01</i>	252	GEEQRYTCHVQHEGLPEPLMLRWKQ	276
1YDP	251	GEEQRYTCHVQHEGLPEPLMLRWKQ	275

**Figura 14.** Alinhamento local da sequência primária madura da proteína codificada pelo alelo *HLA-G\*01:01:01:01* e a sequência primária da estrutura cristalográfica PDB\_ID:1YDP (CLEMENTS *et al.*, 2005). O resíduo em azul (Cys42) indica a localização da cisteína que sofreu mutação na estrutura PDB\_ID:1YDP, enquanto os resíduos em vermelho são provenientes da diferença de cadeia primária do alelo utilizado na estrutura PDB\_ID:1YDP. Os resíduos alinhados foram modelados por homologia.

Para o alinhamento global temos uma sequência de 314 resíduos, que possui 86.6% de identidade e 87.3% de similaridade entre a sequência primária madura de HLA-G e a sequência primária da estrutura cristalográfica PDB\_ID:1YDP. Foi possível identificar com os alinhamentos locais e globais para sequência primária codificada pelo alelo *HLA-G\*01:01:01:01* e do molde PDB\_ID: 1YDP 2 resíduos distintos nessas sequências:

- L, leucina, posição 110 e Q, glutamina, posição 115 - *HLA-G\*01:01:01:01*,
- I, isoleucina, posição 110 e R, arginina, posição 115 – PDB\_ID: 1YDP.

<i>HLA-G*01:01:01:01</i>	1	GS <sup>1</sup> HSMRYFSAAVSRPGRGEPRIAMGYVDDTQFVRFDSDSA <sup>2</sup> CP <sup>3</sup> PRMEPRAP	50
1YDP	1	-SHSMRYFSAAVSRPGRGEPRIAMGYVDDTQFVRFDSDSA <sup>2</sup> S <sup>3</sup> PRMEPRAP	49
<i>HLA-G*01:01:01:01</i>	51	WVEQEGPEYWEEETRNTKAHAQTDRMNLQTLRGYYNQSEASSHTLQWMIG	100
1YDP	50	WVEQEGPEYWEEETRNTKAHAQTDRMNLQTLRGYYNQSEASSHTLQWMIG	99
<i>HLA-G*01:01:01:01</i>	101	CDLGS <sup>1</sup> DGRLL <sup>2</sup> RGYER <sup>3</sup> AYD <sup>4</sup> GKDYALNEDLR <sup>5</sup> SWTAADTAAQISK <sup>6</sup> RKCEAA	150
1YDP	100	CDLGS <sup>1</sup> DGRLL <sup>2</sup> RGYER <sup>3</sup> AYD <sup>4</sup> GKDYALNEDLR <sup>5</sup> SWTAADTAAQISK <sup>6</sup> RKCEAA	149
<i>HLA-G*01:01:01:01</i>	151	NVAEQRRAYLEGTCVEWLHRYLENGKEMLQRADPPKTHVTHHPVFDYEAT	200
1YDP	150	NVAEQRRAYLEGTCVEWLHRYLENGKEMLQRADPPKTHVTHHPVFDYEAT	199
<i>HLA-G*01:01:01:01</i>	201	LRCWALGFYPAEIILTWQRDGEDQTQDVELVETRPAGDGT <sup>1</sup> FQK <sup>2</sup> WAAV <sup>3</sup> VVP	250
1YDP	200	LRCWALGFYPAEIILTWQRDGEDQTQDVELVETRPAGDGT <sup>1</sup> FQK <sup>2</sup> WAAV <sup>3</sup> VVP	249
<i>HLA-G*01:01:01:01</i>	251	SGEEQRYTCHVQHEGLPEPLMLRWK <sup>1</sup> SSLPTIPI <sup>2</sup> MGIVAGLV <sup>3</sup> VLA <sup>4</sup> AV <sup>5</sup> VTG	300
1YDP	250	SGEEQRYTCHVQHEGLPEPLMLRWK <sup>1</sup> -----	275
<i>HLA-G*01:01:01:01</i>	301	AAVA <sup>1</sup> AVL <sup>2</sup> WRK <sup>3</sup> KSSD	314
1YDP	276	-----	275

**Figura 15.** Alinhamento global da sequência primária madura da proteína codificada pelo alelo *HLA-G\*01:01:01:01* e a sequência primária da estrutura cristalográfica PDB\_ID:1YDP (CLEMMENTS *et al.*, 2005). Os resíduos em vermelho e as porções alinhadas foram modelados por homologia, enquanto a porção terminal em verde (que corresponde ao peptídeo conector localizado no final do domínio  $\alpha_3$ , porção transmembrana e cauda citoplasmática) foi modelada pela combinação de técnicas de modelagem *ab initio* e *threading*.

### 6.1.3 Modelagem molecular por homologia

Para a modelagem por homologia da cadeia A da proteína HLA-G foi utilizado o *software* Modeller 9.15 (WEBB *et al.*, 2002), no qual o usuário fornece um alinhamento de uma sequência a ser modelada com estruturas relacionadas conhecidas e o *software* calcula automaticamente um modelo que contém todos os átomos que não sejam o hidrogênio. Modeller utiliza a linguagem de programação Python (disponível em <https://www.python.org/>) e implementa modelagem estrutural comparativa pela satisfação de restrições espaciais.

Além dessas funções, o *software* pode executar muitas tarefas adicionais, incluindo modelagem *de novo* de *loops*, otimização dos diferentes modelos de estrutura da proteína em relação a uma função definida de forma flexível, alinhamento múltiplo de sequências de proteínas e/ou estruturas, *clustering*, pesquisa de bases de dados de sequência, comparação de estruturas de proteínas, entre outras funções (WEBB *et al.*, 2002).

Foram gerados 05 modelos da proteína madura codificada pelo alelo *HLA-G\*01:01:01:01* contendo 276 resíduos. Essa estrutura não contém os 8 resíduos pertencentes ao peptídeo conector (resíduos SSLPTIPI), além de toda porção transmembrana (MGIVAGLVVLAADVTTG AAVAAVLW) e citoplasmática (RKKSSD), que não se encontram na cristalografia PDB\_ID:1YDP e foram modelados por metodologias *ab initio* e *threading*.

Como o resíduo glicina (G, posição 01) não foi capturado na estrutura experimental do nosso molde PDB\_ID:1YDP, para modelagem específica desse resíduo foi utilizada a função de moldes múltiplos (*multiple templates*) do *software* Modeller 9.15 (WEBB *et al.*, 2002), utilizando o molde PDB\_ID:1YDP e o molde PDB\_ID:2D31 (SHIROISHI *et al.*, 2006). O molde PDB\_ID: 2D31 referente ao dímero de HLA-G foi utilizado somente para o preenchimento específico do resíduo da posição 01, enquanto todos os outros resíduos foram modelados baseados na estrutura PDB\_ID: 1YDP.

O molde PDB\_ID: 2D31 e o dímero solúvel de HLA-G serão discutidos com mais detalhes posteriormente.

### 6.1.3.1 Avaliação e seleção dos modelos

Cada um dos cinco modelos gerados para a proteína HLA-G foi avaliado pelos *softwares* PROCHECK (LASKOWSKI *et al.*, 1993), Verify 3D (LUTHY *et al.*, 1992), ERRAT (COLOVOS e YATES, 1993), QMEAN (BENKERT *et al.*, 2008) e ProSA-Web (WIEDERSTEIN *et al.*, 2007), para que o melhor fosse selecionado.

A compatibilidade dos modelos tridimensionais com suas sequências primárias foi realizada por meio do *software* Verify 3D, que considera modelos com boa qualidade aqueles com pontuação acima de 80%. Dentre nossos modelos, o modelo 1 atingiu maior pontuação nesse quesito, de 94.57%.

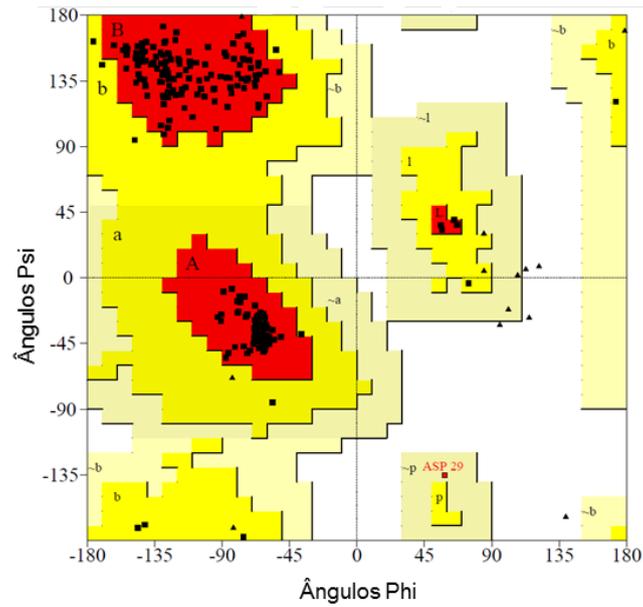
O *software* ERRAT analisou, estatisticamente, as interações dos átomos não ligados, considerando a pontuação acima de 95% como ideal para bons modelos preditos experimentalmente. Dentre nossos modelos, o modelo 1 atingiu maior pontuação também nesse quesito, de 85.44%.

O *software* ProSA-Web avaliou a qualidade do modelo gerado comparada com a qualidade de estruturas proteicas obtidas por meio de métodos experimentais. Novamente, o modelo 1 foi o melhor, com *z-score* de -8.56.

A ferramenta QMEAN pontuou 0.789 para o modelo 1, baseado na função de pontuação composta tanto para a estimativa da qualidade global de todo o modelo quanto para a análise de resíduos locais em diferentes regiões dentro de um modelo.

A avaliação dos modelos tridimensionais pelos *softwares* ERRAT, Verify 3D e Prosa-Web demonstrou que o melhor modelo gerado foi o modelo 1. Esse modelo foi utilizado para as análises e modelagens subsequentes e pode ser observado na **Figura 17**.

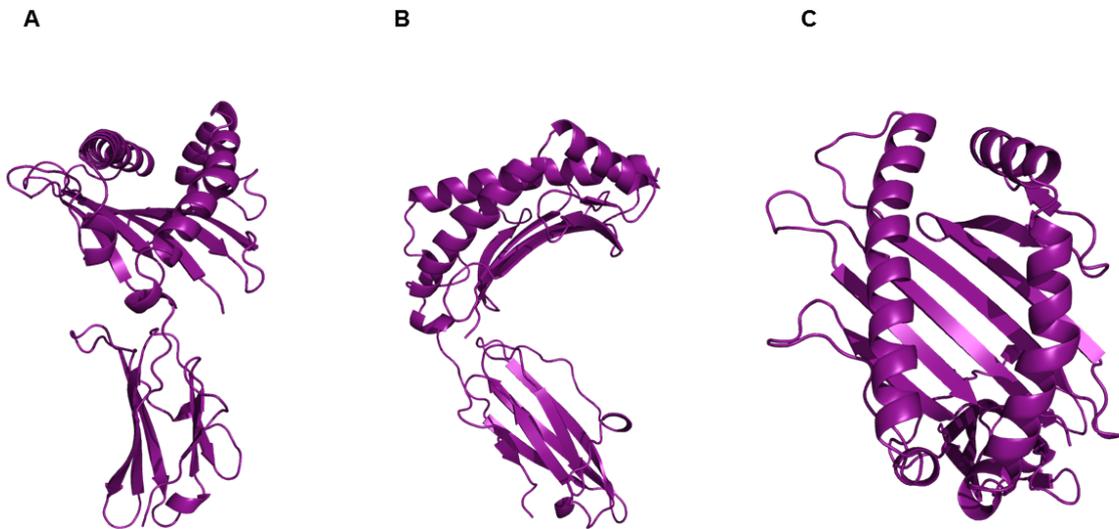
A **Figura 16** representa o resultado do gráfico de Ramachandran para o modelo 1. A análise dos resultados apresentados pelo gráfico de Ramachandran gerado pelo *software* PROCHECK determina se os aminoácidos se encontram em regiões energeticamente favoráveis ou desfavoráveis, orientando a avaliação da qualidade dos modelos gerados. O *score* desejável para que um modelo seja positivamente avaliado é que >90% dos aminoácidos que compõem a proteína estejam localizados em regiões favoráveis do gráfico (LASKOWSKI *et al.*, 1993). No gráfico de Ramachandran é possível visualizar os dois ângulos de torção  $\phi$  (phi) e  $\psi$  (psi) dos aminoácidos que determinam a conformação da estrutura tridimensional das proteínas (RAMACHANDRAN *et al.*, 1963).



#### Estadística do gráfico

Aminoácidos em regiões favoráveis [A,B,L]	231 - 95,5%
Aminoácidos em regiões fadionalmente permitidas [a,b,l,p]	10 - 4,1%
Aminoácidos em regiões generosamente permitidas [~a,~b,~l,~p]	1 - 0,4%
Aminoácidos em regiões não permitidas	0 - 0,0%
Resíduos terminais (não Glicina/Prolina)	1
Resíduos (glicina)	19
Resíduos (prolina)	14

**Figura 16.** Resultados da análise preconizada por Ramachandran para a estrutura da proteína HLA-G. Área em vermelho representa regiões muito favoráveis, em amarelo regiões favoráveis, em amarelo claro regiões pouco favoráveis e em branco regiões desfavoráveis. Os quadrados pretos representam o posicionamento dos aminoácidos no gráfico e os triângulos pretos representam as glicinas.



**Figura 17.** Melhor modelo (modelo 1) para proteína codificada pelo alelo *HLA-G\*01:01:01:01*, gerado por homologia. Visão frontal da molécula (A), lateral (B) e superior do sulco que acomoda o peptídeo a ser apresentado pelo *HLA-G\*01:01:01:01* (C).

#### 6.1.4 Modelagem molecular *ab initio*

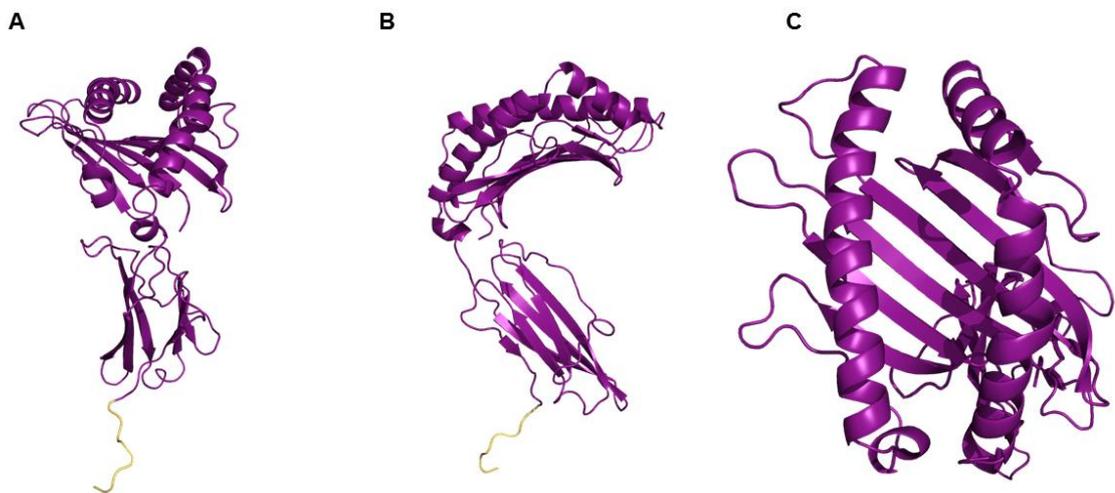
Após a obtenção do modelo por homologia, a próxima etapa realizada foi a modelagem da porção ocupada pelo peptídeo conector na cadeia A, que não foi resolvida na estrutura PDB\_ID: 1YDP (CLEMENTS, *et al.*, 2005) e se estende da posição 277 a 284, contendo os resíduos SSLPTIPI.

Para essa região extracelular não resolvida experimentalmente, foi utilizada a metodologia de modelagem *ab initio*, utilizando predição de estruturas proteicas a partir da sequência de aminoácidos. Tal região foi utilizada para geração da biblioteca de fragmentos com a ferramenta Robetta, sendo tais fragmentos utilizados como parte dos arquivos de entrada para construção dos modelos *ab initio*. O protocolo Rosetta (PETTERSEN *et al.*, 2004) foi utilizado para a modelagem de 2000 estruturas *ab initio* contendo os resíduos selecionados, empregando o algoritmo *cyclic coordinate descent* (CCD). A porção já modelada por homologia foi mantida fixa ao longo das simulações *ab initio*. O enovelamento da proteína é ditado por forças físicas agindo nos átomos que a compõem e esse método é considerado de

maior acurácia (HUMPFREY *et al.*, 1996). As conformações que possuem a menor energia livre são consideradas as mais prováveis de serem adotadas pela proteína em condições nativas.

#### 6.1.4.1 Seleção dos modelos

A escolha do melhor modelo foi baseada nos melhores resultados das ferramentas QMEAN (BENKERT, *et al.*, 2008) e MODFOLD (MCGUFFIN *et al.*, 2008). Dentre os 2000 modelos gerados, o modelo número 496 foi o melhor. Apresentou para a ferramenta MODFOLD o *score* de 0.8121 para Qualidade Global do Modelo (*Global Model Quality*), além de *p-value* de  $1.946 \times 10^{-4}$ . Na ferramenta QMEAN o mesmo modelo apresentou *score* 0.735 (*QMEANnorm\_score*).



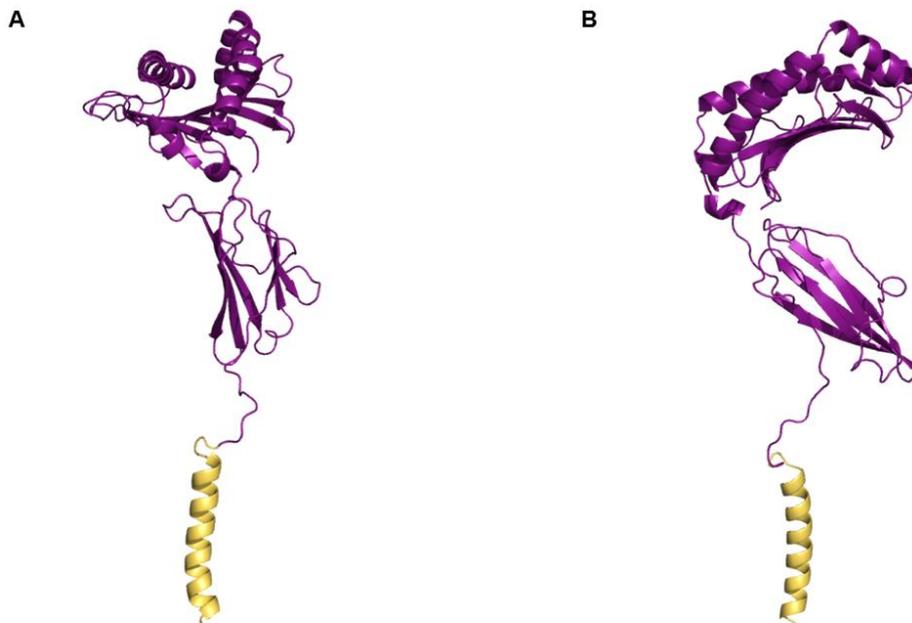
**Figura 18.** Melhor modelo (modelo 496) da proteína codificada pelo alelo *HLA-G\*01:01:01:01*, gerado por modelagem *ab initio* da porção referente ao peptídeo conector. Visão frontal da molécula (A), lateral (B) e superior do sulco que acomoda o peptídeo a ser apresentado pela proteína codificada pelo alelo *HLA-G\*01:01:01:01* (C). A porção em amarelo corresponde aos resíduos modelados utilizando metodologia *ab initio*.

### 6.1.5 Modelagem *in silico* por metodologia de *threading*

Para realizar a modelagem da porção transmembrana encontrada na cadeia A (resíduos MGIVAGLVVLAADVVTGAAVA AVLW, posição 285 a 308), foi utilizado o servidor *online* GPCR-ITASSER (ZHANG *et al.*, 2015), que emprega método computacional projetado para a predição de estrutura 3D de receptores de membrana GPCR. A sequência de destino é primeiro indexada através da biblioteca do PDB para busca por modelos putativos. Os modelos homólogos identificados são utilizados no procedimento de montagem de fragmentos para construção de modelos completos. Caso não existam modelos homólogos disponíveis, é realizada modelagem *ab initio* para hélices a partir do zero. Ao todo foram gerados 5 modelos.

O melhor modelo apresentou os valores de *TM-score* e *C-score* de  $0.68 \pm 0.15$  e  $-1.02$ , respectivamente.

A porção transmembrana gerada por *threading* foi incorporada ao modelo anterior utilizando a função de moldes múltiplos (*multiple templates*) do *software* Modeller 9.15, gerando um modelo único que contém a porção extracelular (domínios  $\alpha 1$ ,  $\alpha 2$  e  $\alpha 3$ ; peptídeo conector) e a porção transmembrana.



**Figura 19.** Modelo da proteína codificada pelo alelo *HLA-G\*01:01:01:01*, gerado por *threading* da porção transmembrana. Visão frontal da molécula (A), lateral (B). A porção em amarelo corresponde aos resíduos que compõem a porção transmembrana.

### 6.1.6 Modelagem molecular *ab initio* da cauda citoplasmática

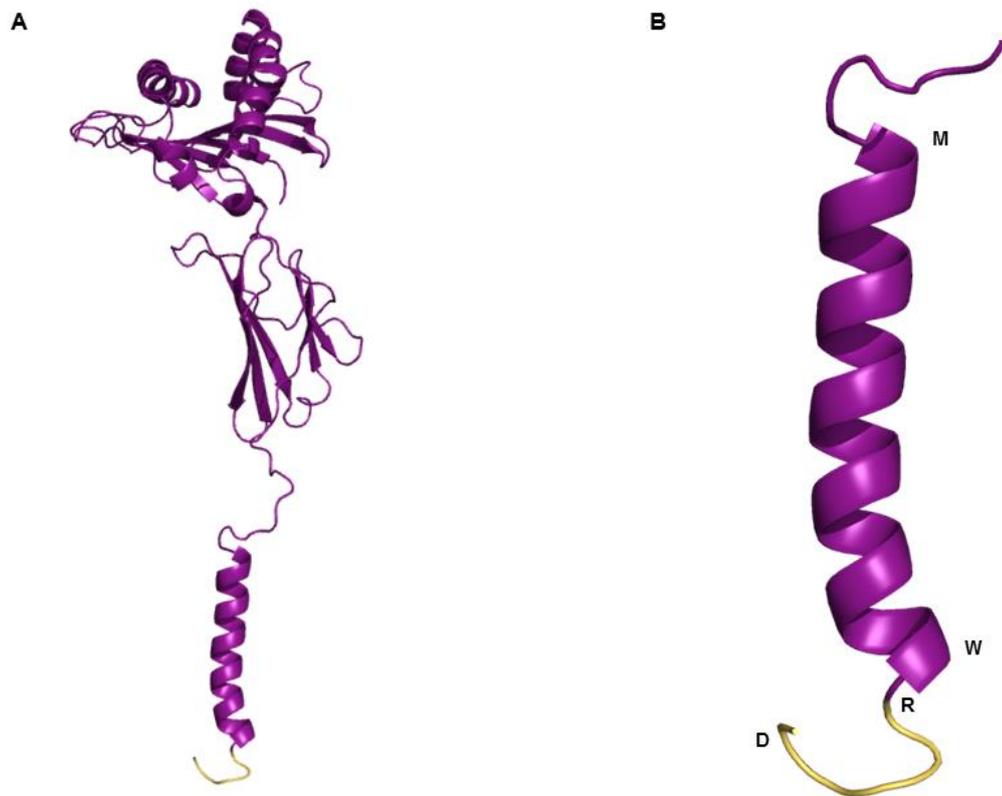
Com o intuito de finalizar a modelagem da cadeia A que compõem a estrutura da proteína HLA-G, foi realizada a modelagem da pequena cauda citoplasmática não resolvida na estrutura PDB\_ID: 1YDP (CLEMENTS *et al.*, 2005), que se estende da posição 309 a 314 (resíduos RKKSSD).

Novamente, foi utilizada a metodologia de modelagem *ab initio*, da mesma forma empregada anteriormente, utilizando as ferramentas Robetta e Rosetta (PETTERSEN *et al.*, 2004). Foram geradas 2000 estruturas *ab initio* contendo os resíduos selecionados, empregando o algoritmo *cyclic coordinate descent* (CCD). Toda a porção já modelada previamente e descrita nos tópicos anteriores foi mantida fixa ao longo das simulações *ab initio*.

#### 6.1.6.1 Seleção dos modelos

A escolha do melhor modelo foi baseada nos melhores resultados das ferramentas QMEAN (BENKERT *et al.*, 2008) e MODFOLD (MCGUFFIN *et al.*, 2008).

O melhor modelo gerado foi o modelo número 547, que apresentou para a ferramenta MODFOLD o *score* de 0.8263 para Qualidade Global do Modelo (*Global Model Quality*), além de *p-value* de  $3.67 \times 10^{-7}$ . Na ferramenta QMEAN o mesmo modelo apresentou *score* 1.19 (*QMEANnorm\_score*).



**Figura 20.** Modelo da proteína codificada pelo alelo *HLA-G\*01:01:01:01*, gerado por modelagem *ab initio* da porção referente à cauda citoplasmática. Visão frontal da molécula (A) e aproximação da porção transmembrana (demonstração da localização do resíduo inicial M - metionina e final W - triptofano) e cauda citoplasmática (demonstração da localização do resíduo inicial R – arginina, e final D - ácido aspártico) (B). A porção em amarelo corresponde aos resíduos modelados utilizando metodologia *ab initio*.

### 6.1.7 Modelagem da proteína $\beta$ 2-microglobulina acoplada à proteína HLA-G completa

Para modelagem por homologia da proteína  $\beta$ 2-microglobulina acoplada à proteína HLA-G, foi utilizado o *software* Modeller 9.15 (WEBB *et al.*, 2002), especificamente a função de modelagem de cadeias múltiplas (*multiple chains*).

Como exposto anteriormente, a estrutura cristalográfica PDB\_ID:1YDP (CLEMENTS *et al.*, 2005) que foi utilizada como molde desde a primeira modelagem por homologia para a proteína HLA-G contém também a estrutura da proteína  $\beta$ 2-microglobulina resolvida (cadeia

B), podendo servir de molde para a modelagem da mesma. A **Figura 21** mostra a sequência primária da proteína  $\beta$ 2-microglobulina, no formato FASTA.

```
> $\beta$ 2-microglobulina
MIQRTPKIQVYSRHPAENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWS
FYLLYYTEFTPTKDEYA CRVNHVTL SQPKIVKWRDM
```

**Figura 21.** Sequência FASTA da proteína  $\beta$ 2-microglobulina.

Foram gerados 05 modelos da proteína codificada pelo alelo *HLA-G\*01:01:01:01* contendo 308 resíduos referentes à estrutura da proteína HLA-G e 99 aminoácidos referentes à  $\beta$ 2-microglobulina.

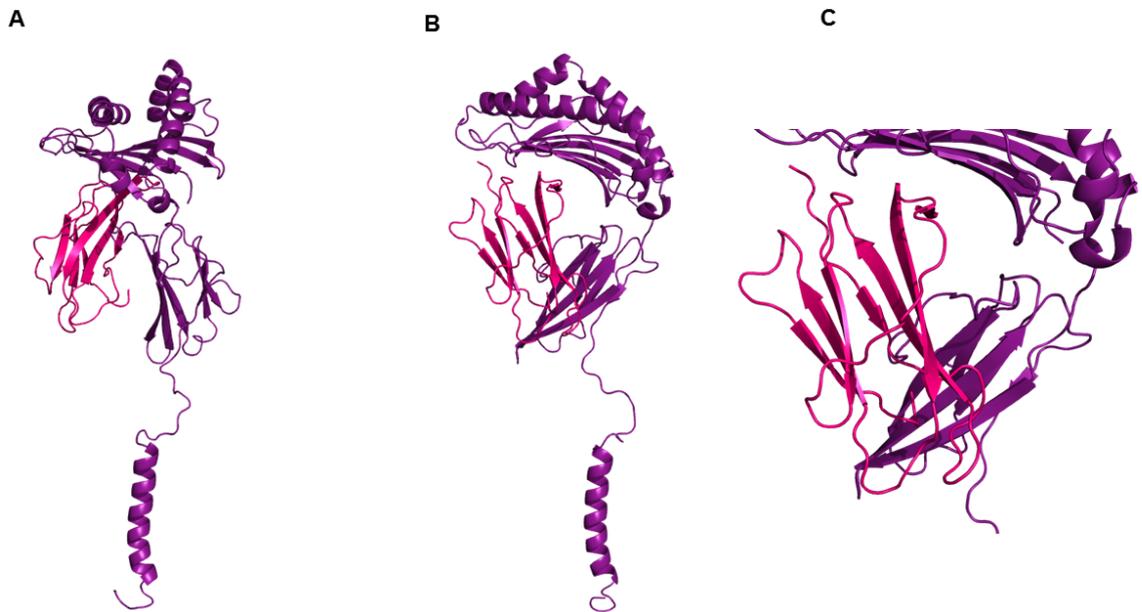
#### 6.1.7.1 Avaliação e seleção dos modelos

Cada um dos cinco modelos gerados foi avaliado pelos *softwares* PROCHECK (LASKOWSKI *et al.*, 1993), Verify 3D (LUTHY *et al.*, 1992), ERRAT (COLOVOS e YATES, 1993), QMEAN (BENKERT *et al.*, 2008), e ProSA-Web (WIEDERSTEIN *et al.*, 2007), para que o melhor fosse selecionado.

Dentre os modelos gerados, o modelo 1 atingiu as seguintes pontuações, sendo escolhido como melhor modelo gerado:

- Verify 3D: 97.98%.
- ERRAT: 82.81%.
- ProSA-Web: *z-score* de -7.43.
- QMEAN: 0.756.

Esse modelo foi utilizado para as análises e modelagens subsequentes e pode ser observado na **Figura 22**.

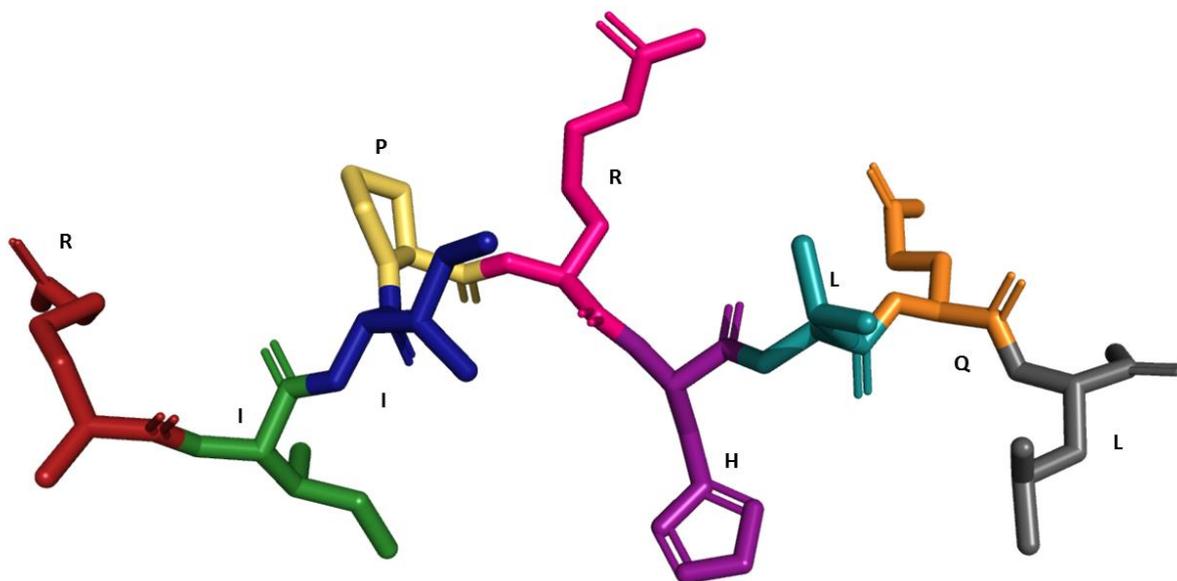


**Figura 22.** Modelo da proteína codificada pelo alelo *HLA-G\*01:01:01:01* (roxo) acoplada à proteína  $\beta$ 2-microglobulina (rosa). Visão frontal da molécula (A), lateral (B) e aproximação da cadeia composta pela  $\beta$ 2-microglobulina (C).

### 6.1.8 *Docking* do peptídeo ao sulco da proteína HLA-G completa

O peptídeo foi inserido no sulco, utilizando o *software* Autodock Vina (TROTT *et al.*, 2010), capaz de realizar o *docking* de pequenas moléculas e fármacos.

A estrutura do nonapeptídeo RIIPRHLQL (cadeia P) foi obtida a partir da cristalografia PDB\_ID:1YDP (CLEMENTS *et al.*, 2005). Utilizando o *software* Pymol foi possível salvá-la em um arquivo de coordenadas estruturais de extensão *.pdb*.



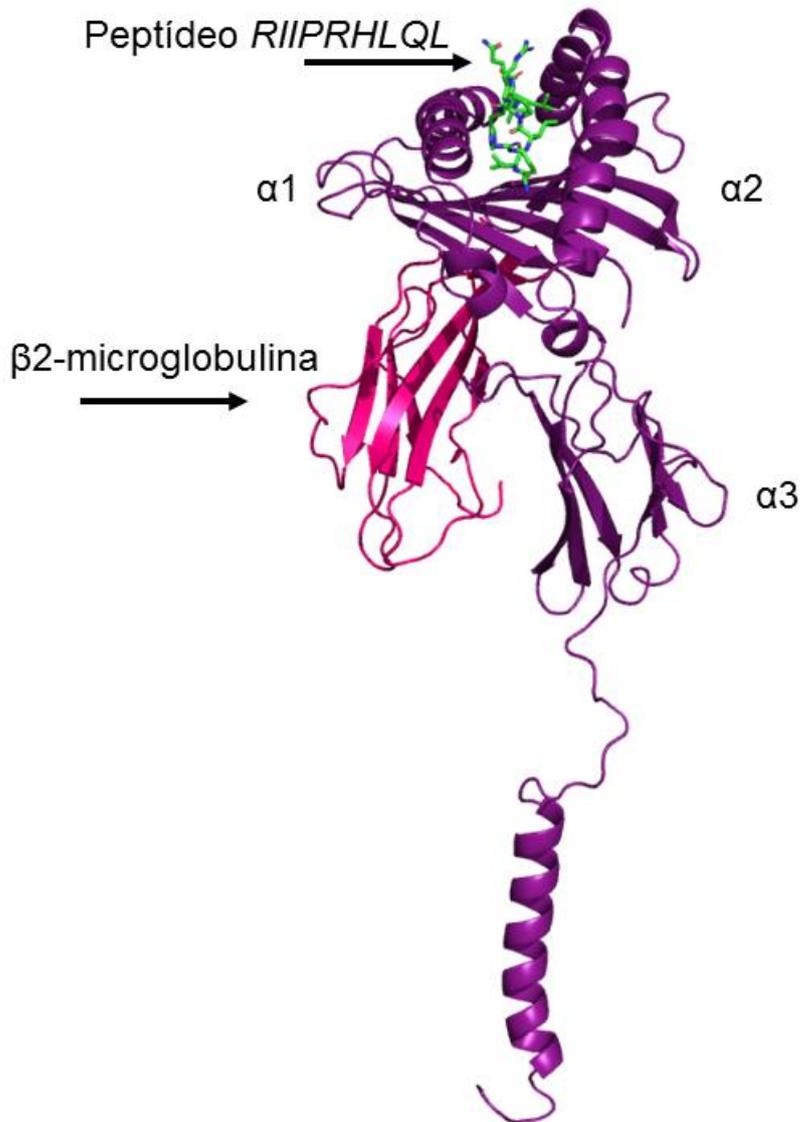
**Figura 23.** Estrutura tridimensional do nonapeptídeo RIIPRHLQL (cadeia P) contendo os códigos que representam os resíduos (R, arginina; I, isoleucina; P, prolina; H, histidina; L, leucina; Q, glutamina).

O arquivo *.pdb* que contém a estrutura da proteína codificada pelo alelo *HLA-G\*01:01:01:01* completa, acoplada à proteína  $\beta$ 2-microglobulina foi denominado receptor para esse experimento de *docking*.

Cada experimento *in silico* de *docking* foi repetido 10 vezes, produzindo em cada rodada 10 possíveis conformações para o encaixe do peptídeo no sulco. Todas as simulações foram executadas com alto nível de exaustividade (exaustividade = 1000). A cadeia principal do peptídeo foi mantida rígida.

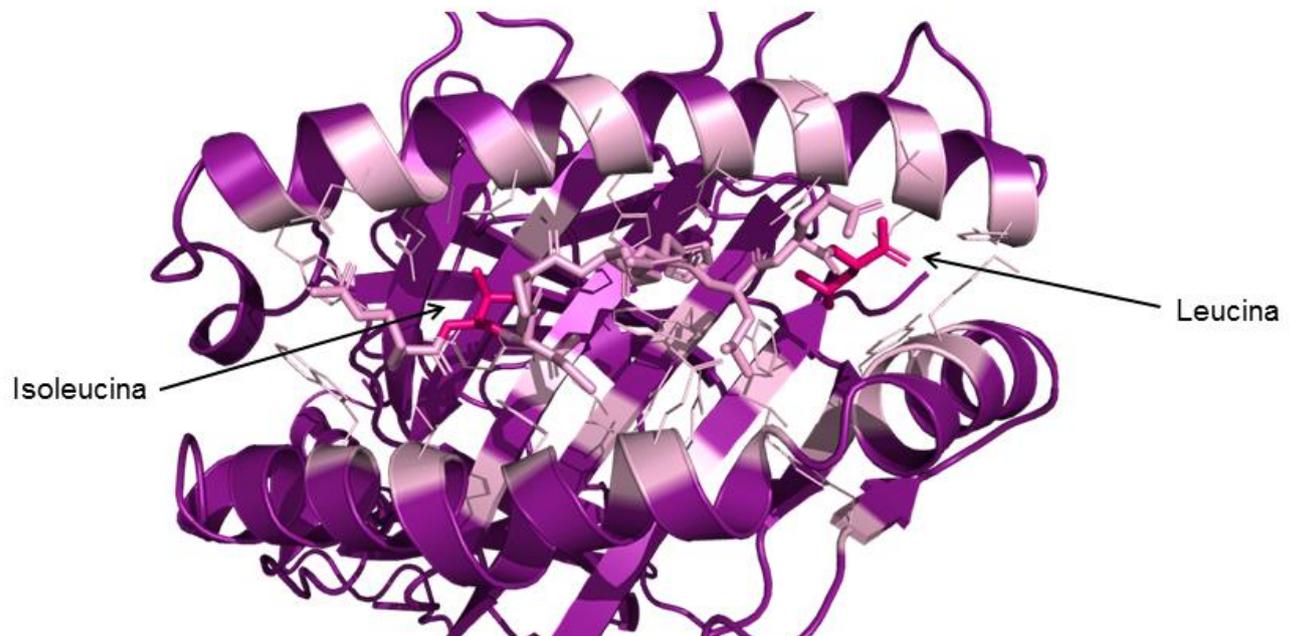
A melhor conformação do peptídeo RIIPRHLQL após o *docking* ao sulco da proteína codificada pelo alelo *HLA-G\*01:01:01:01* apresentou energia de ligação de -12,13 kcal/mol.

Além disso, foi mensurado o desvio da média quadrática (RMSD, *Root Mean Square Deviation*) entre a estrutura proveniente da cristalografia PDB\_ID:1YDP (CLEMENTS *et al.*, 2005) e a estrutura modelada durante a execução do presente trabalho (contendo as cadeias A, B e P), cujo resultado obtido foi de 0,850 Å.



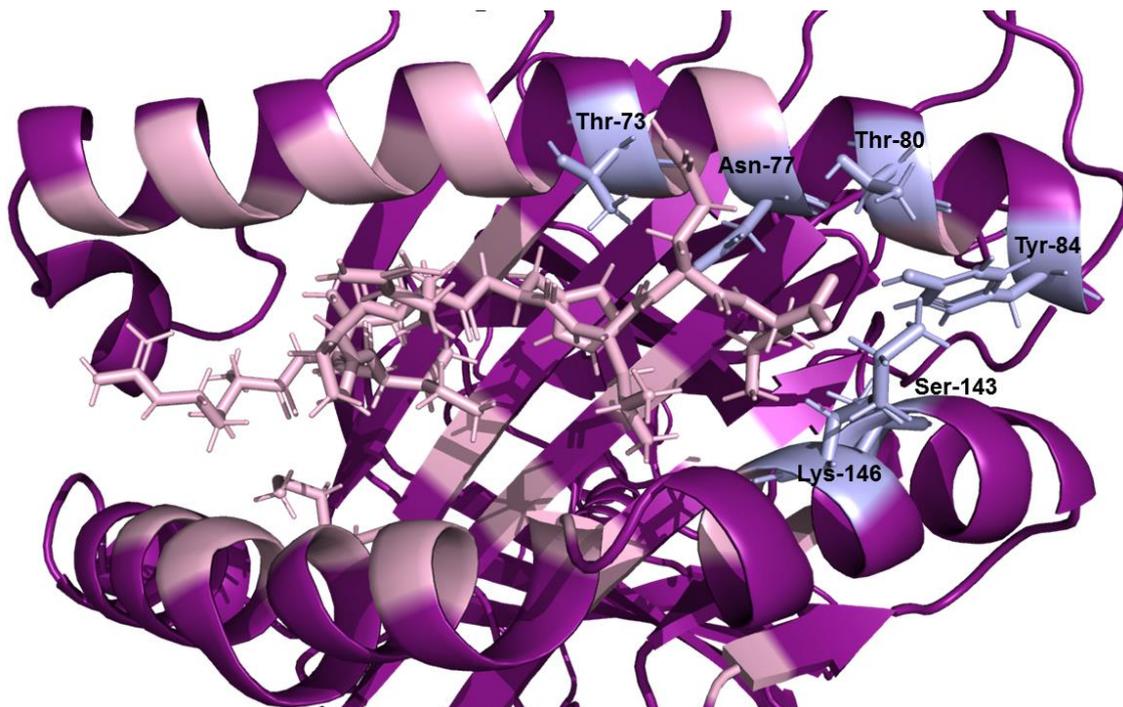
**Figura 24.** Visão frontal da proteína codificada pelo alelo *HLA-G\*01:01:01:01* (roxo) completa, acoplada à proteína  $\beta 2$ -microglobulina (rosa) e apresentando no sulco o peptídeo RIIPRHLQL (verde).

O pacote BioPython (COCK *et al.*, 2009) foi utilizado para verificar os resíduos que fazem parte das interações proteína-peptídeo. Nas **Figura 25** pode-se observar em rosa claro os resíduos da proteína HLA-G que interagem com o peptídeo RIIPRHLQL, sendo que os resíduos âncora estão em rosa *pink* (resíduos 2 - Isoleucina, 9 - Leucina).



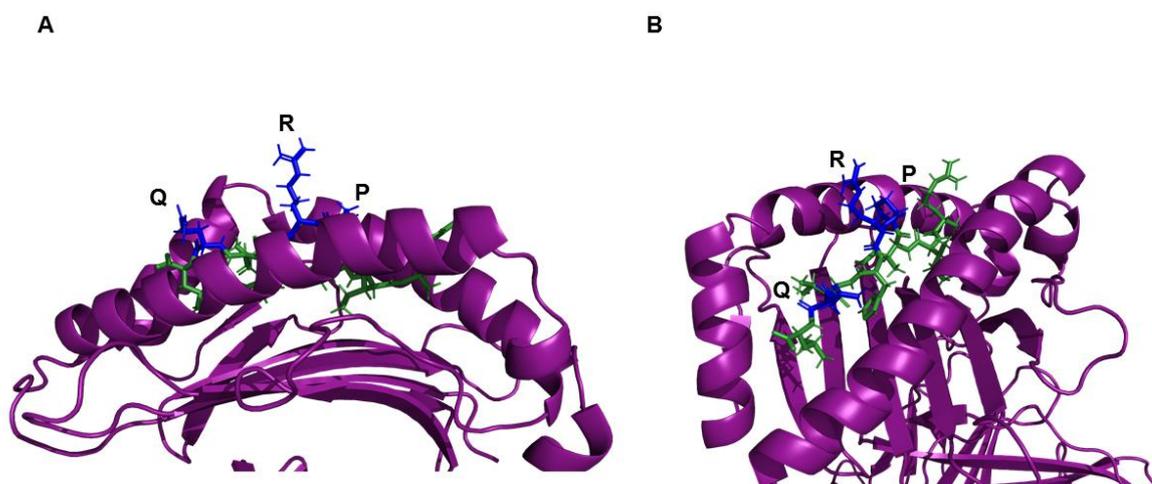
**Figura 25.** Sulco da molécula HLA-G e peptídeo RIIPRHLQL acoplado, demonstrando as interações dos resíduos da estrutura HLA-G (rosa claro) e o peptídeo, indicando os resíduos âncora 2 - Isoleucina e 9 - Leucina (pink).

Na **Figura 26** pode-se observar em azul claro os resíduos do sulco da proteína HLA-G que interagem com o peptídeo.



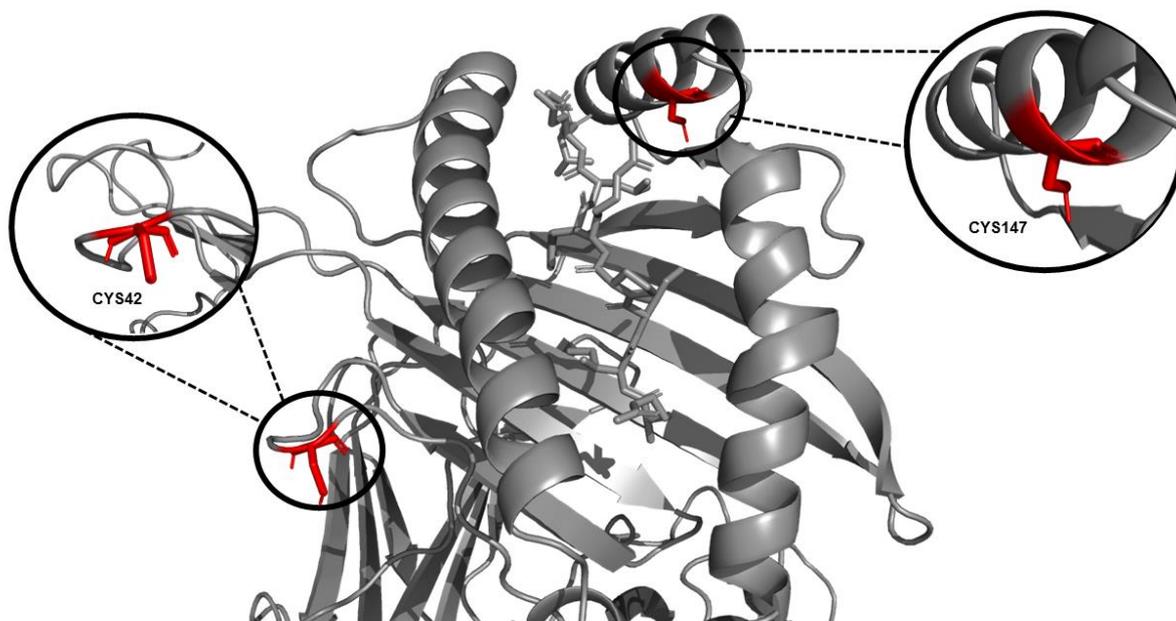
**Figura 26.** Sulco da molécula HLA-G e peptídeo RIIPRHLQL acoplado, demonstrando as interações entre os resíduos da estrutura HLA-G (azul claro) e o peptídeo, indicando os resíduos Y (Tyr, tirosina, posição 84), S (Ser, serina, posição 143), K (Lys, lisina, posição 146), T (Thr, treonina, posição 73), N (Asn, asparagina, posição 77) e T (Thr, treonina, posição 80).

Na **Figura 27** pode-se observar em verde o peptídeo no sulco da proteína HLA-G e em azul os resíduos que se projetam para fora do sulco da proteína HLA-G.

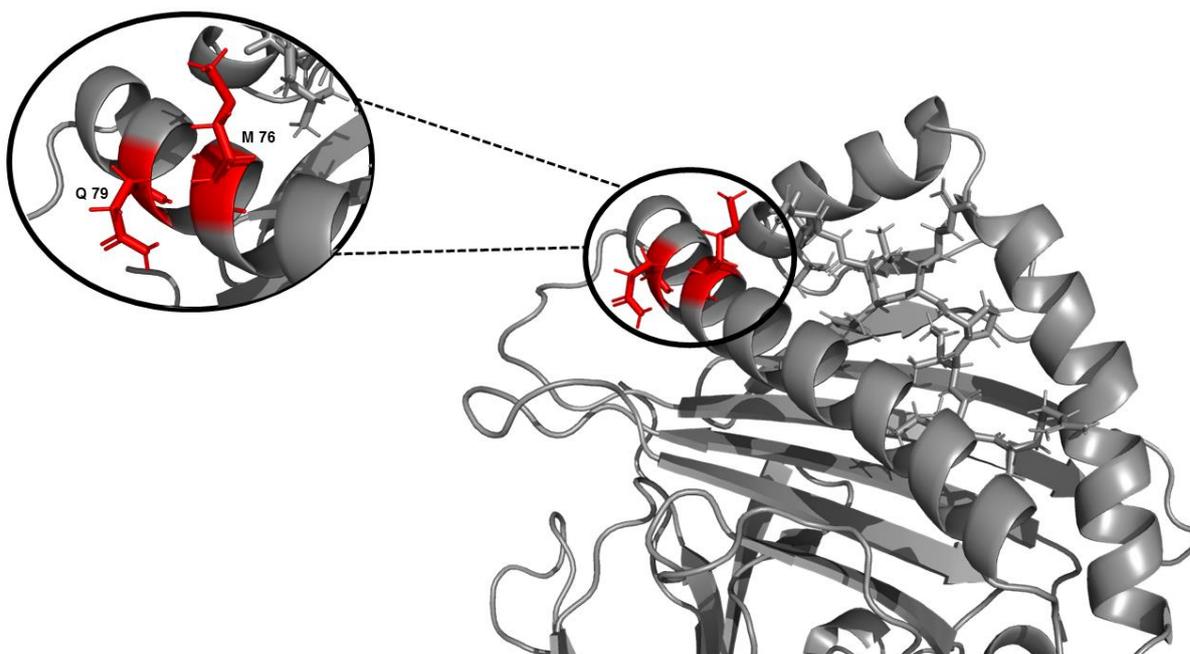


**Figura 27.** (A e B) Sulco da molécula HLA-G e peptídeo RIIPRHLQL acoplado, demonstrando cadeias laterais (representação em *sticks*) dos resíduos P (prolina, posição 4), R (arginina, posição 5) e Q (glutamina, posição 8).

Na **Figura 28** estão representados os resíduos de cisteína nas posições 42 e 147; enquanto na **Figura 29**, os resíduos M (metionina, posição 76) e Q (glutamina, posição 79).

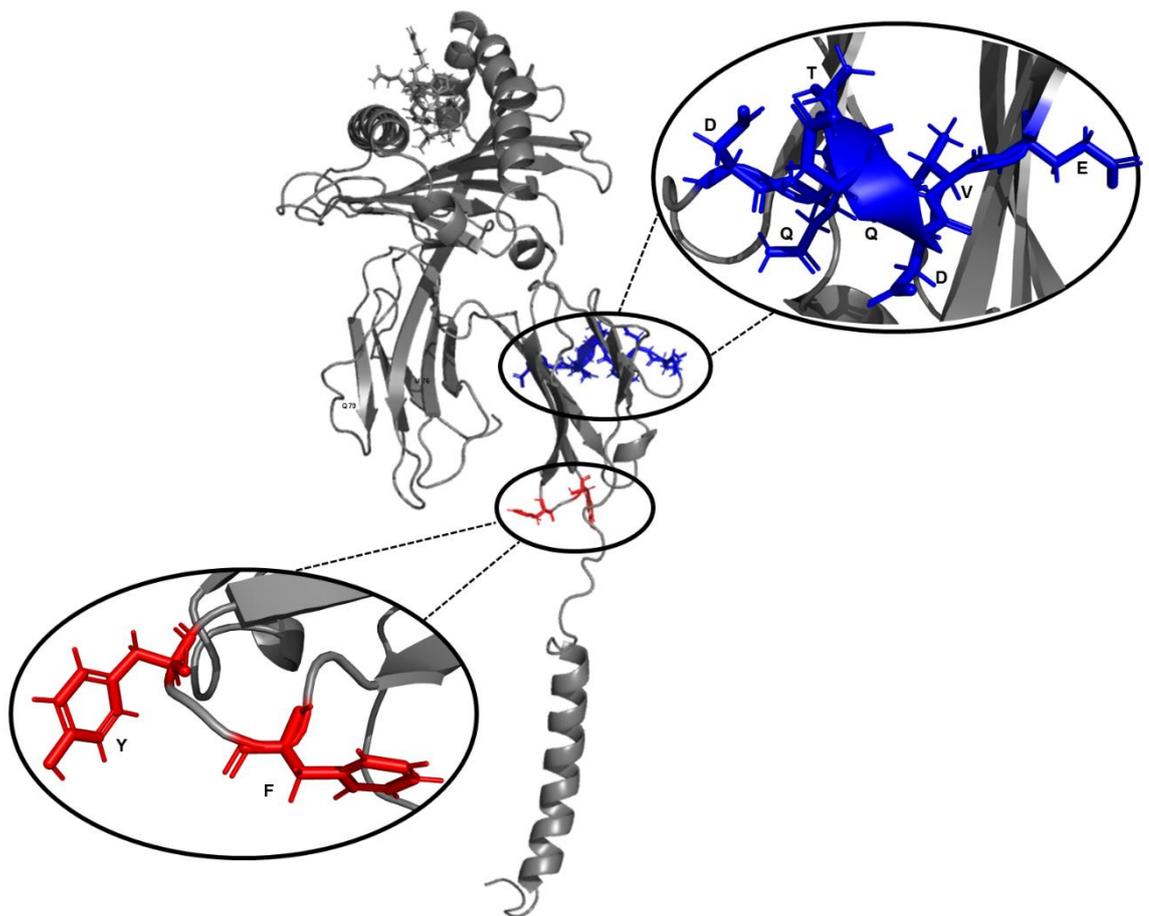


**Figura 28.** Em vermelho os resíduos Cys42 (cisteína, posição 42) e Cys147 (cisteína, posição 147), ampliados para melhor visualização da área.



**Figura 29.** Em vermelho os resíduos M (metionina, posição 76) e Q (glutamina, posição 79), ampliados para melhor visualização da área.

Na **Figura 30** estão representados os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197), além dos resíduos D (ácido aspártico, posição 223), Q (glutamina, posição 224), T (treonina, posição 225), Q (glutamina, posição 226), D (ácido aspártico, posição 227), V (valina, posição 228) e E (ácido glutâmico, posição 229).

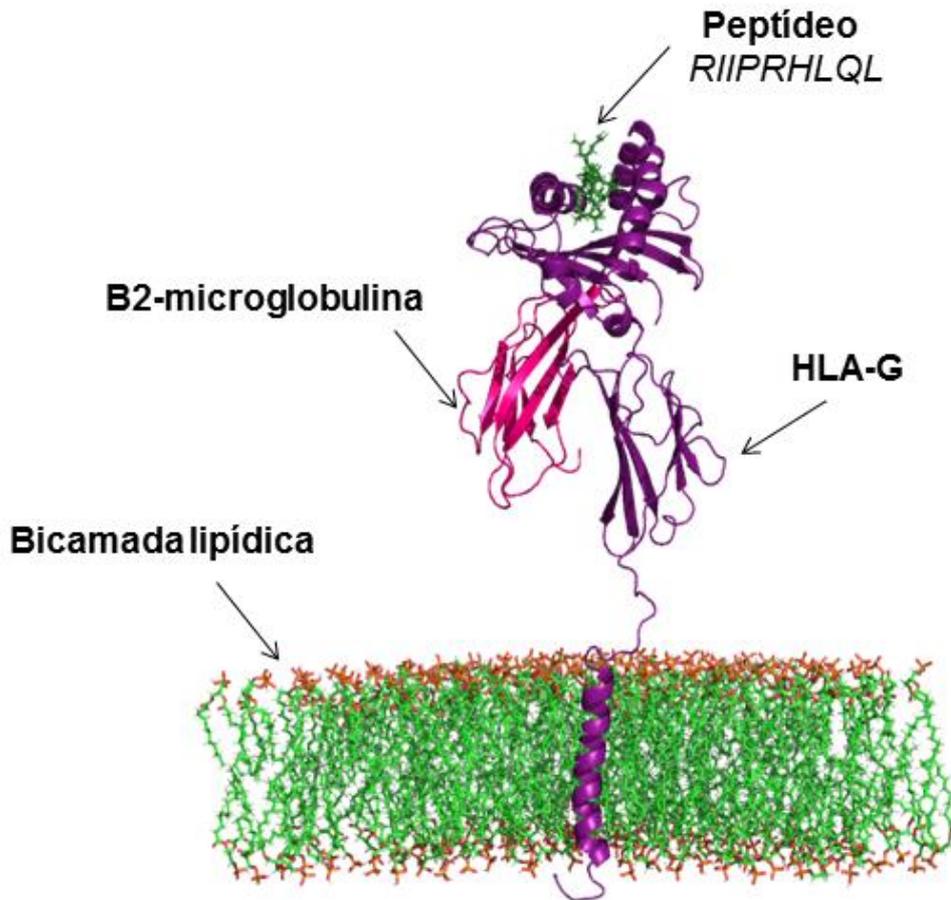


**Figura 30.** Em vermelho os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197); em azul os resíduos D (ácido aspártico, posição 223), Q (glutamina, posição 224), T (treonina, posição 225), Q (glutamina, posição 226), D (ácido aspártico, posição 227), V (valina, posição 228) e E (ácido glutâmico, posição 229), ampliados para melhor visualização da área.

### 6.1.9 Inserção da estrutura completa de HLA-G em bicamada lipídica

Com o intuito de mimetizar a disposição da proteína HLA-G na superfície celular, o molde completo descrito anteriormente foi inserido em uma bicamada lipídica, composta por moléculas de DLPA (fosfolipídio, do inglês *1,2-Dilauroyl-sn-glycero-3-phosphate*), utilizando o servidor *online* CHARMM-GUI, especificamente a ferramenta para construção de bicamadas lipídicas (*Bilayer Builder*).

Os resíduos selecionados para inserção e posicionamento correto da proteína HLA-G na bicamada lipídica foram resíduos M (metionina, posição 285, início da porção transmembrana) e W (triptofano, final da porção transmembrana, posição 308).

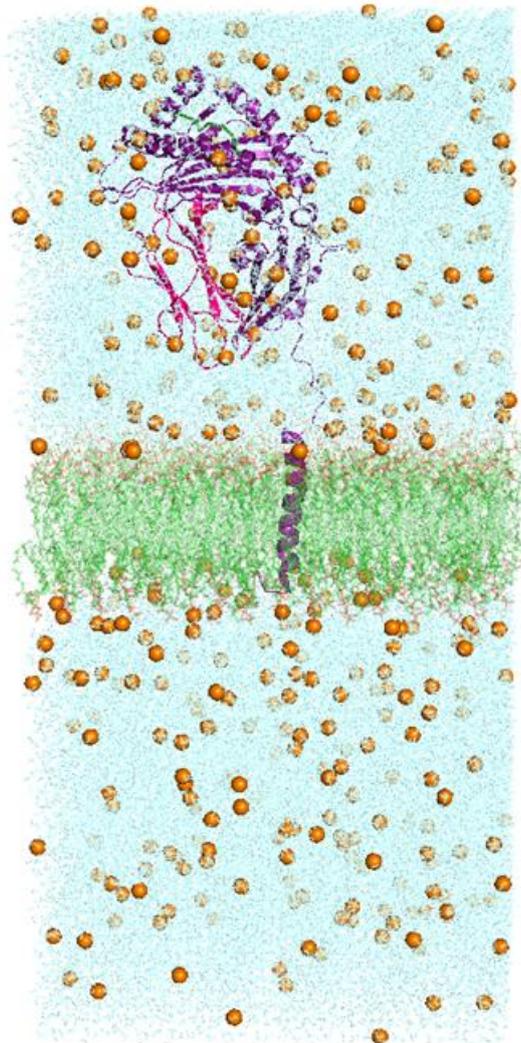


**Figura 31.** Complexo da proteína HLA-G completa (roxo), proteína  $\beta$ 2-microglobulina (rosa), peptídeo RIIPRHLQL acoplado na fenda (verde) e bicamada lipídica composta de unidades de fosfolípido DLPA (verde claro). Recorte na bicamada lipídica para expor a porção transmembrana.

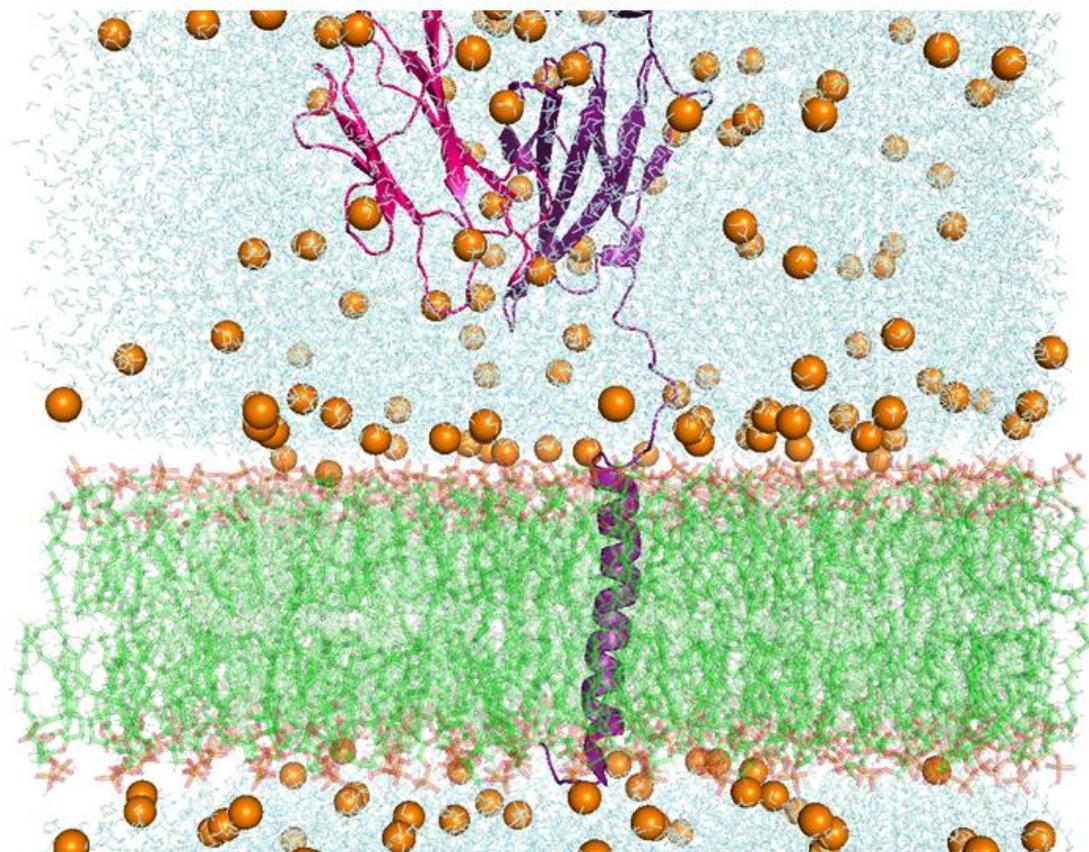
Além da inserção da proteína HLA-G em uma bicamada lipídica, o servidor CHARMM-GUI realiza a montagem de um sistema para simulação de dinâmica molecular, que contém além da proteína HLA-G, proteína  $\beta$ 2-microglobulina e o peptídeo RIIPRHLQL:

- 32560 moléculas de fosfolípidios (DLPA),
- 184197 moléculas de água,
- 380 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

O sistema completo encontra-se na **Figura 32**.



**Figura 32.** Sistema pronto para dinâmica molecular contendo água (azul claro), íons  $\text{Na}^+$  e  $\text{Cl}^-$  (esferas douradas), proteína HLA-G completa (roxo), proteína  $\beta 2$ -microglobulina (rosa), peptídeo RIIPRHLQL acoplado na fenda (verde escuro) e bicamada lipídica composta de unidades de fosfolípido DLPA (verde claro).



**Figura 33.** Visão detalhada da região transmembrana do sistema pronto para dinâmica molecular contendo água (azul claro), íons  $\text{Na}^+$  e  $\text{Cl}^-$  (esferas douradas), proteína HLA-G completa (roxo) e bicamada lipídica composta de unidades de fosfolípido DLPA (verde claro).

#### 6.1.10 Dinâmica molecular da proteína HLA-G completa

A simulação de dinâmica molecular (DM) clássica utiliza a descrição newtoniana das forças envolvidas entre os elementos que constituem o sistema a ser simulado. Essa metodologia calcula a posição atômica de cada átomo que compõem o sistema em diferentes e sucessivos momentos ao longo do tempo, separados por intervalos temporais previamente determinados. Esse conjunto de coordenadas atômicas do sistema e sua evolução temporal descrevem a trajetória do sistema, proteína-solvente, a ser analisado.

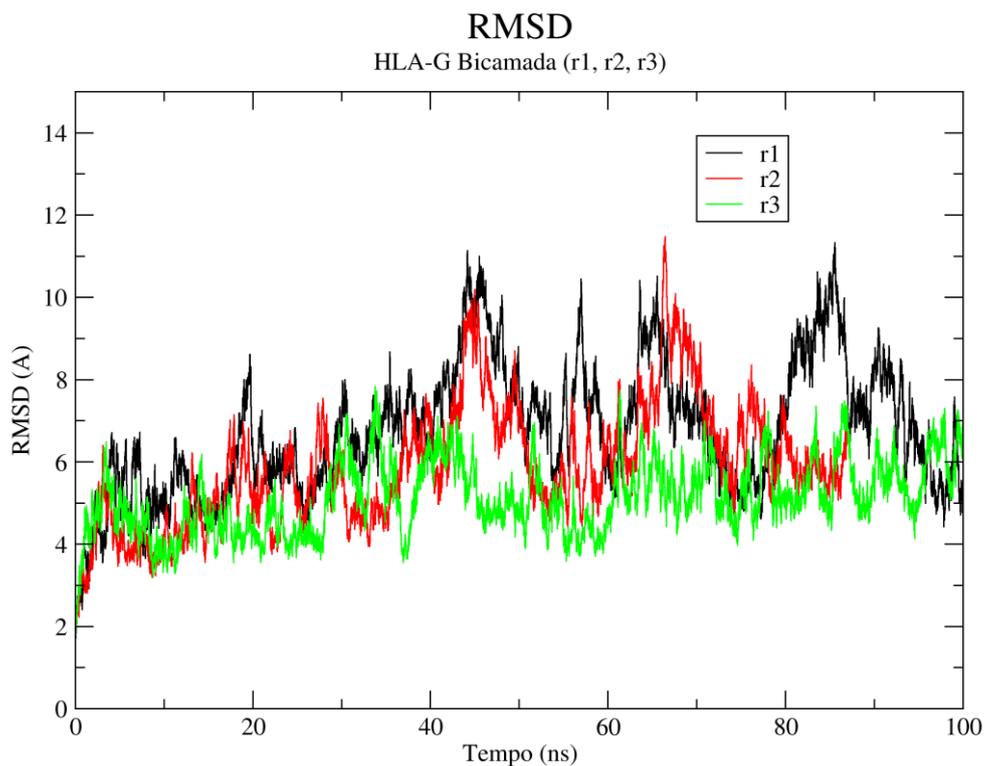
A dinâmica molecular consistiu em uma triplicata (r1, r2, r3) das simulações para o modelo gerado. Foi utilizado o *software* GROMACS 5.1.4 (HESS *et al.*, 2008), que implementou o campo de força CHARMM36m (HUANG *et al.*, 2017).

A DM simulou o sistema proteína-solvente em condições físico-químicas controladas, com temperatura máxima de 310 K (ou seja 36,85 °C) e pressão constante de 1 atm. A estrutura tridimensional da proteína foi solvatada em uma caixa cúbica contendo modelo de água SPC (*simple point charge*) com espessura suficiente em cada direção nos eixos *x*, *y* e *z*. Foram adicionados íons Na<sup>+</sup> Cl<sup>-</sup> suficientes para o equilíbrio do sistema, com concentração final de 0,15 mol / L.

Para todas as simulações de DM, o estágio de produção foi precedido por (i) três etapas de Minimização de Energia e (ii) oito etapas de Equilíbrio. O estágio de Equilíbrio começou com restrições de posição e temperatura de 310 K, por um período de 300 ps, para permitir a formação de camadas de solvatação. A temperatura foi então reduzida para 280 K e as restrições de posição foram gradualmente reduzidas. Este processo foi seguido por um aumento gradual da temperatura (até 300 K). Juntos, esses passos representam os primeiros 500 ps de cada simulação. Durante o estágio de produção, o sistema foi mantido à temperatura constante (300 K) sem restrições.

As simulações de dinâmica molecular para o sistema que contém a proteína HLA-G atingiram o tempo total de 100ns.

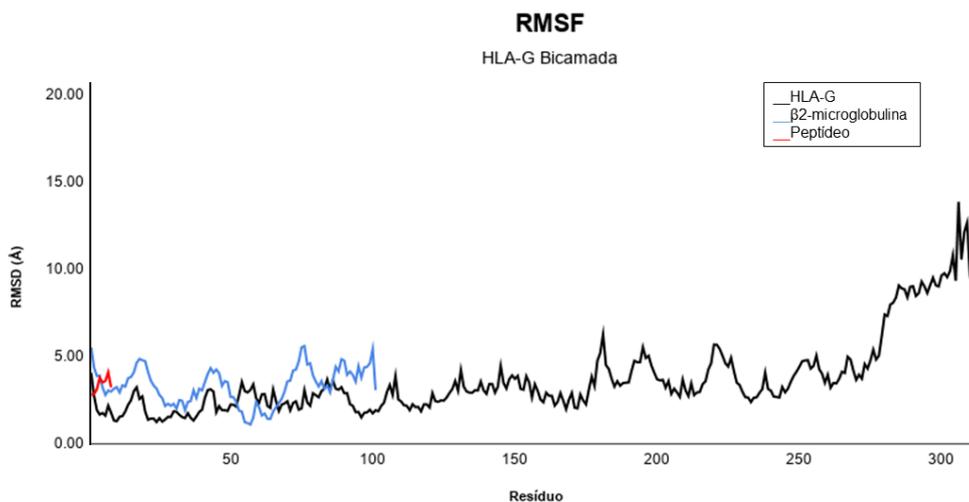
O gráfico de RMSD, apresentado na **Figura 34**, demonstra as triplicatas da dinâmica molecular do complexo HLA-G e bicamada lipídica. Na simulação de dinâmica molecular os valores de RMSD são utilizados como medida da estabilidade do modelo, sendo mensurado a cada momento o desvio padrão da distância entre os átomos de um modelo ao longo do tempo. Quanto mais estável uma estrutura, menor seu RMSD. Podemos observar que as triplicatas do complexo HLA-G apresentaram valores de RMSD que não ultrapassaram 11,46 Å, mantendo-se majoritariamente na faixa de 8 a 10 Å durante a maior parte da trajetória.



**Figura 34.** Gráfico de RMSD do sistema contendo a proteína HLA-G completa, bicamada lipídica, água e íons.

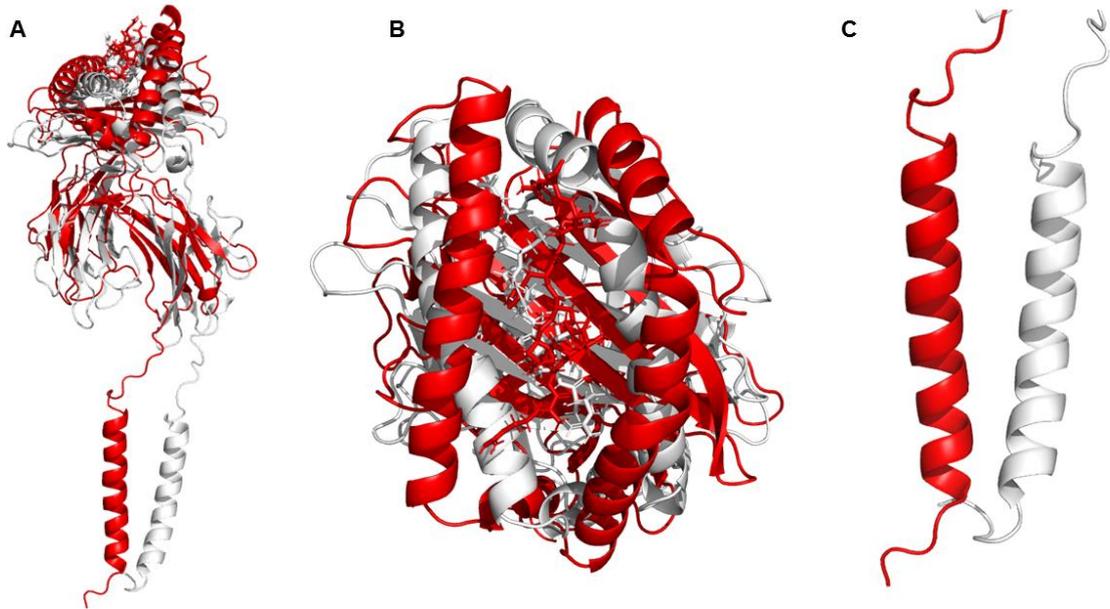
Os valores de RMSF foram obtidos para as coordenadas dos átomos C $\alpha$  da estrutura da proteína ao longo do tempo de simulação usando como referência sua estrutura inicial, a fim de analisar quais aminoácidos fornecem características mais flexíveis às proteínas. RMSF calcula o desvio médio quadrático da flutuação das posições dos átomos das estruturas das proteínas em relação a uma estrutura de referência (FUGLEBAKK *et al.*, 2012).

O gráfico apresentando o resultado de RMSF para a simulação mais estável da proteína HLA-G no sistema de bicamada lipídica demonstrou que a flutuação está mais presente na porção terminal da proteína HLA-G, em torno dos aminoácidos que compõem a porção transmembrana (resíduos 285 a 308).

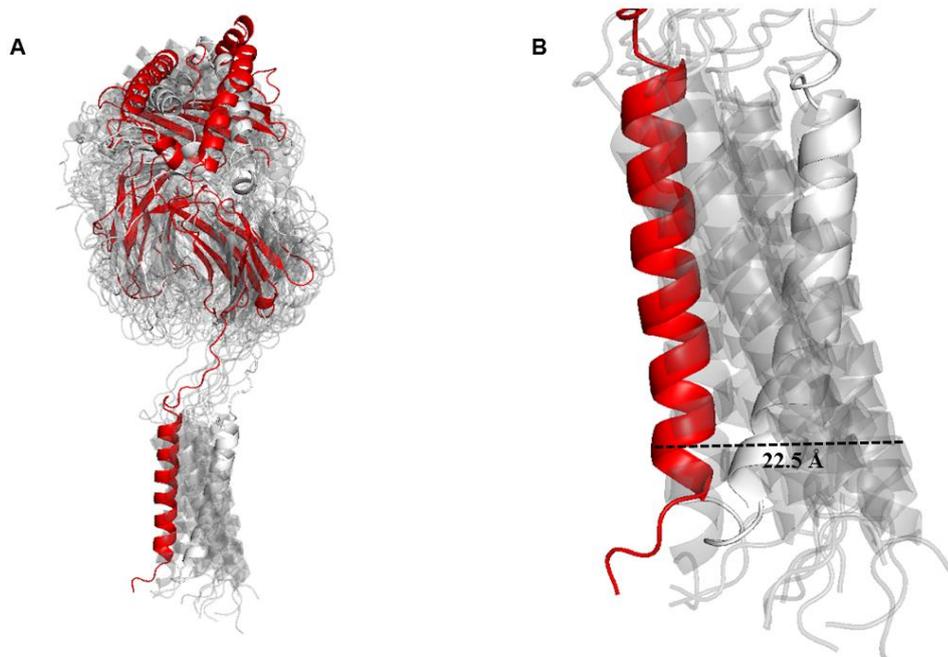


**Figura 35.** Variação estrutural por resíduo da replicata r3 da proteína HLA-G. Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.

A **Figura 36** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 100ns da dinâmica da proteína HLA-G. Estão representadas nas figuras somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 3,138 Å.



**Figura 36.** Proteína codificada pelo alelo *HLA-G\*01:01:01:01*, juntamente com a proteína  $\beta 2$ -microglobulina e o peptídeo RIIPRHLQL nos momentos iniciais (branco) e finais (vermelho) da simulação de 100ns. **(A)** Visão frontal da proteína HLA-G. **(B)** Visão superior da proteína HLA-G, mostrando em primeiro plano o sulco e o peptídeo RIIPRHLQL inserido no mesmo. **(C)** Aproximação da porção transmembrana e pequena cauda citoplasmática da proteína HLA-G.



**Figura 37.** Proteína codificada pelo alelo *HLA-G\*01:01:01:01*, juntamente com a proteína  $\beta 2$ -microglobulina e o peptídeo RIIPRHLQL nos momentos iniciais (branco), finais (vermelho) e intervalos de 10ns (cinza claro) da simulação de 100ns. **(A)** Visão frontal da proteína HLA-G. **(B)** Aproximação da porção transmembrana e pequena cauda citoplasmática da proteína HLA-G.

cauda citoplasmática da proteína HLA-G, demonstrando o deslocamento máximo de 22,5 Å da estrutura na bicamada lipídica.

As trajetórias obtidas na dinâmica molecular foram analisadas com o auxílio dos *softwares* GROMACS e VMD (HUMPHREY *et al.*, 1996), utilizados para interpretar os dados das trajetórias completas.

## 6.2 Conformação do dímero HLA-G1 solúvel

A presença de uma cisteína na posição 42 do monômero de HLA-G possibilita a formação de um homodímero solúvel, contendo 2 monômeros, que contém 3 cadeias cada um, respectivamente:

- Cadeia A: 275 resíduos, *HLA-G\*01:01:01:01*;
- Cadeia P: 9 resíduos, nonapeptídeo RIIPRHLQL;
- Cadeia B: 99 resíduos,  $\beta$ 2-microglobulina.

Tal estrutura foi resolvida por cristalografia de difração de raio-X por Shiroishi e colaboradores (SHIROISHI *et al.*, 2006) e possui resolução de 3.2 Å.

### 6.2.1 Modelagem por homologia dos resíduos faltantes do dímero de HLA-G1

Os seguintes resíduos não foram capturados na estrutura experimental, em nenhum dos monômeros que compõem o dímero (**Figura 38**):

- D: ácido aspártico, posição 196,
- Y: tirosina, posição 197,
- E: ácido glutâmico, posição 198,
- P: prolina, posição 266,
- E: ácido glutâmico, posição 267,

- M: metionina, posição 01 da  $\beta$ 2-microglobulina.

Para obtenção a estrutura completa, os mesmos foram modelados por homologia utilizando a função de moldes múltiplos (*multiple templates*) do *software* Modeller 9.15 (WEBB *et al.*, 2002). O molde escolhido para preenchimento das posições faltantes foi o PDB\_ID:1YDP, referente ao monômero de HLA-G, utilizado previamente (CLEMENTS *et al.*, 2005).

```

2D31      GSHSMRYFSAAVSRPGRGEPRIAMGYVDDTQFVRFSDSAACPRMEPRAPWVEQEGPEYW
Dímero_completo  GSHSMRYFSAAVSRPGRGEPRIAMGYVDDTQFVRFSDSAACPRMEPRAPWVEQEGPEYW
1YDP      -SHSMRYFSAAVSRPGRGEPRIAMGYVDDTQFVRFSDSAASPRMEPRAPWVEQEGPEYW
*****

2D31      EEETRNTKAHAQTDRMNLQTLRGYINQSEASSHTLQWIMIGCDLGS DGRLLRGYEQYAYDG
Dímero_completo  EEETRNTKAHAQTDRMNLQTLRGYINQSEASSHTLQWIMIGCDLGS DGRLLRGYEQYAYDG
1YDP      EEETRNTKAHAQTDRMNLQTLRGYINQSEASSHTLQWIMIGCDLGS DGRLLRGYERYAYDG
*****

2D31      KDYLALNEDLRSWTAADTAAQISKRKCEANVAEQRRAYLEGTCVEWLHRYLENGKEMLQ
Dímero_completo  KDYLALNEDLRSWTAADTAAQISKRKCEANVAEQRRAYLEGTCVEWLHRYLENGKEMLQ
1YDP      KDYLALNEDLRSWTAADTAAQISKRKCEANVAEQRRAYLEGTCVEWLHRYLENGKEMLQ
*****

2D31      RADPPKTHVTHHPVFDYEATLRCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGT
Dímero_completo  RADPPKTHVTHHPVFDYEATLRCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGT
1YDP      RADPPKTHVTHHPVFDYEATLRCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGT
*****

2D31      FQKWAAVVPSGEEQRYTCHVQHEGEPFLMLRW--
Dímero_completo  FQKWAAVVPSGEEQRYTCHVQHEGEPFLMLRW--
1YDP      FQKWAAVVPSGEEQRYTCHVQHEGEPFLMLRWKQ
*****

```

**Figura 38.** Alinhamento múltiplo das sequências primárias do dímero formado pela proteína codificada pelo alelo *HLA-G\*01:01:01:01* (denominada aqui Dímero\_completo) e estruturas cristalográficas PDB\_ID:1YDP (CLEMENTS *et al.*, 2005) e PDB\_ID: 2D31 (SHIROISHI *et al.*, 2006). Os resíduos em vermelho demarcam resíduos distintos entre a sequência completa do dímero (Dímero\_completo), sua estrutura resolvida (PDB\_ID: 2D31) e o molde referente ao monômero de HLA-G (PDB\_ID:1YDP). Os resíduos não resolvidos na estrutura PDB\_ID:2D31 demarcados em azul e o restante dos resíduos alinhados foram modelados por homologia com a estrutura do monômero de HLA-G (PDB\_ID:1YDP).

### 6.2.1.1 Avaliação dos modelos

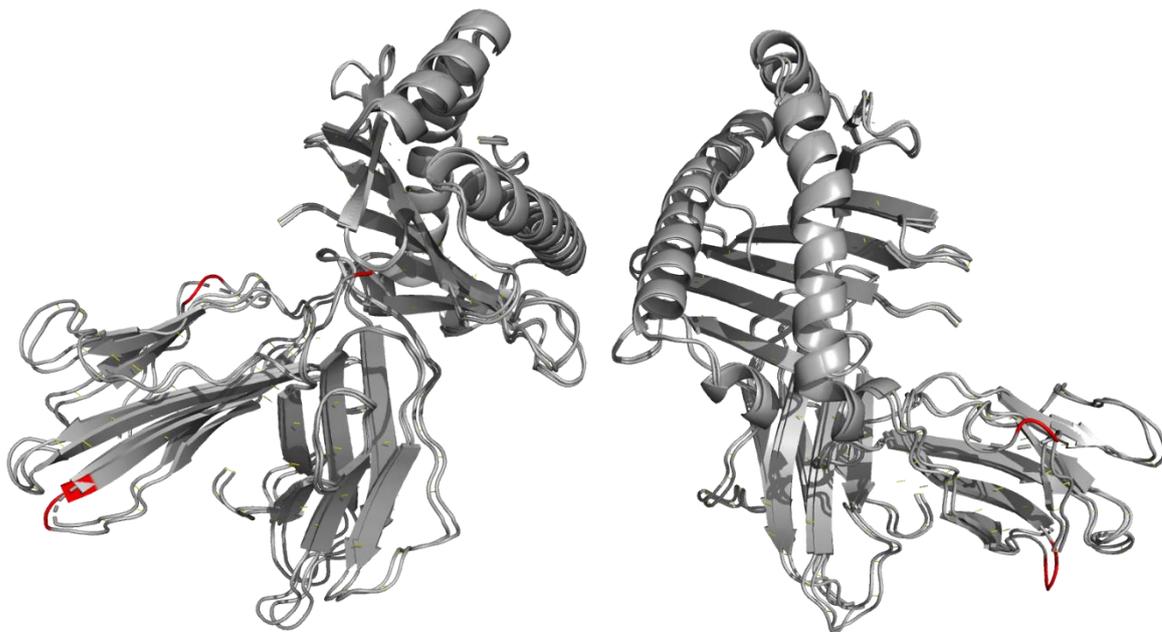
Cada um dos cinco modelos gerados para a estrutura completa do dímero de HLA-G1 foi avaliado pelos *softwares* PROCHECK (LASKOWSKI *et al.*, 1993), ERRAT (COLOVOS

e YATES, 1993), Verify 3D (LUTHY *et al.*, 1992), QMEAN (BENKERT *et al.*, 2008) e ProSA-Web (WIEDERSTEIN *et al.*, 2007), para que o melhor fosse selecionado.

Dentre os modelos gerados, o modelo 3 atingiu as seguintes pontuações, sendo escolhido como melhor modelo gerado:

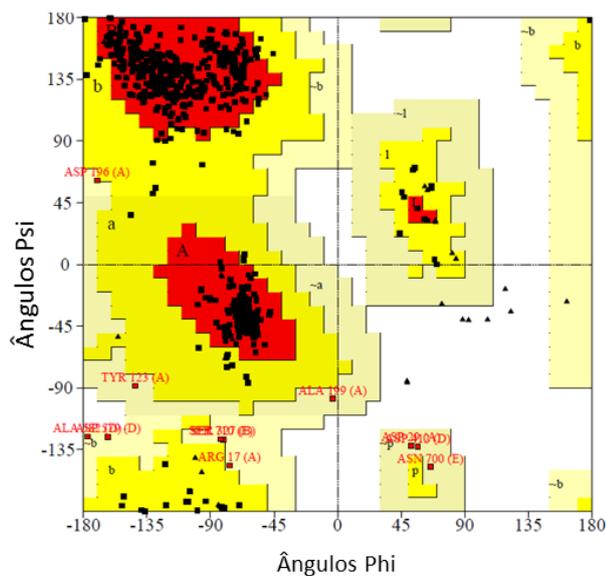
- Verify 3D: 91.1%.
- ERRAT: 84.56%.
- ProSA-Web: *z-score* de -12.9.
- QMEAN: 0.619.

Esse modelo foi utilizado para as análises e modelagens subsequentes e pode ser observado na **Figura 39**.



**Figura 39.** Alinhamento estrutural do modelo 3 referente ao dímero da proteína codificada pelo alelo *HLA-G\*01:01:01:01* (cinza) e a estrutura experimental PDB\_ID: 2D31 (cinza) (SHIROISHI *et al.*, 2006). As regiões em vermelho demonstram os aminoácidos que foram modelados por homologia com a estrutura experimental PDB\_ID: 1YDP (CLEMENTS *et al.*, 2005), com o intuito de preencher os resíduos faltantes (*gaps*) existentes na cristalografia.

A **Figura 40** representa o resultado do gráfico de Ramachandran para o modelo 3.



#### Estadística do gráfico

Aminoácidos em regiões favoráveis [A,B,L]	604 – 89.9%
Aminoácidos em regiões fadicionalmente permitidas [a,b,l,p]	57 – 8.4%
Aminoácidos em regiões generosamente permitidas [-a,-b,-l,-p]	11 – 1.6%
Aminoácidos em regiões não permitidas	0 – 0.0%
Resíduos terminais (não Glicina/Prolina)	10
Resíduos (glicina)	44
Resíduos (prolina)	40

**Figura 40.** Resultados da análise preconizada por Ramachandran para a estrutura do dímero solúvel de HLA-G1. Área em vermelho representa regiões muito favoráveis, em amarelo regiões favoráveis, em amarelo claro regiões pouco favoráveis e em branco regiões desfavoráveis. Os quadrados pretos representam o posicionamento dos aminoácidos no gráfico e os triângulos pretos representam as glicinas.

Os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197) estão demonstrados na **Figura 41**, além dos resíduos D (ácido aspártico, posição 223), Q (glutamina, posição 224), T (treonina, posição 225), Q (glutamina, posição 226), D (ácido aspártico, posição 227), V (valina, posição 228) e E (ácido glutâmico, posição 229).



**Figura 41.** Em vermelho os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197) relacionados à interação com receptores ILT2 e ILT4; em azul os resíduos D (ácido aspártico, posição 223), Q (glutamina, posição 224), T (treonina, posição 225), Q (glutamina, posição 226), D (ácido aspártico, posição 227), V (valina, posição 228) e E (ácido glutâmico, posição 229), relacionados à interação com CD8. Em amarelo a ponte dissulfeto formada pelas cisteínas localizadas na posição 42 de cada monômero.

### 6.2.2 Formação das pontes dissulfeto

Após a escolha do melhor modelo, foi formada a ponte dissulfeto intercadeia na posição Cys42 (cadeia A) e Cys42 (cadeia D), utilizando a função *pdb2gmx* do *software* GROMACS 4.6.5 (HESS *et al.*, 2008). A distância entre os 2 resíduos de cisteína foi de 1,952 Å.



**Figura 42.** Aproximação da ponte dissulfeto (em amarelo) formada pelas cisteínas na posição 42 (Cys42-Cys42), de cada monômero do dímero de HLA-G1 (em vermelho).

### 6.2.3 Dinâmica Molecular

A dinâmica molecular consistiu em 2 simulações (r1 e r2) para o modelo gerado. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

O sistema para simulação de dinâmica molecular, que contém além do dímero solúvel da proteína HLA-G, cada monômero acoplado à proteína  $\beta$ 2-microglobulina e ao peptídeo RIIPRHLQL:

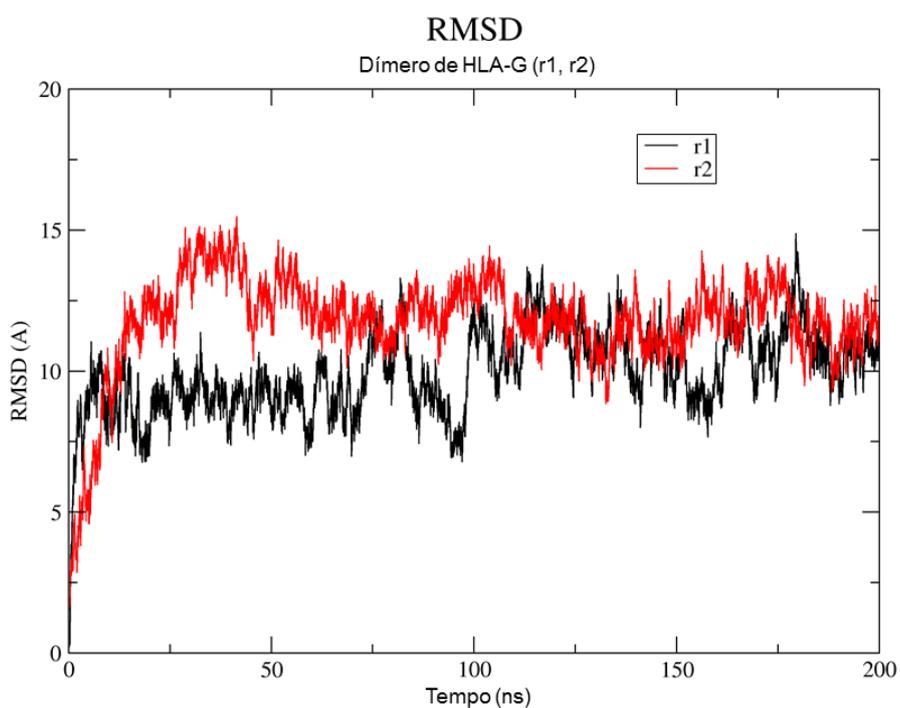
- 383325 moléculas de água,
- 470 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

A DM simulou o sistema proteína-solvente em condições físico-químicas controladas, com temperatura máxima de 310 k (ou seja 36.85 °C) e pressão constante de 1 atm. A estrutura tridimensional da proteína foi solvatada em uma caixa cúbica contendo modelo de água SPC (*simple point charge*) com espessura suficiente para cada direção nos eixos  $x$ ,  $y$  e  $z$ . Foram

adicionados íons  $\text{Na}^+$   $\text{Cl}^-$  suficientes para o equilíbrio do sistema, com concentração final de  $0,15 \text{ mol / L}$ .

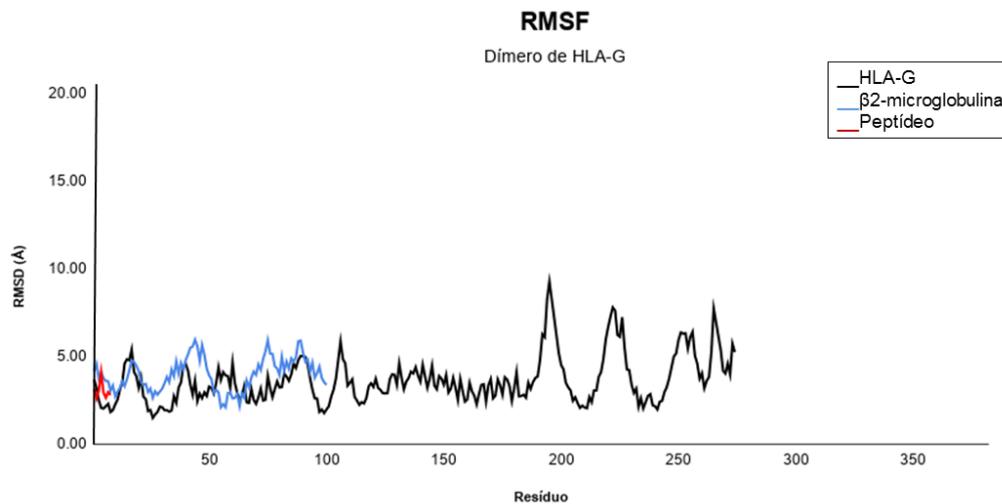
As simulações de dinâmica molecular para o sistema que contém o dímero de HLA-G1 atingiram o tempo total de 200ns.

O gráfico de RMSD, apresentado na **Figura 43**, demonstra as 2 simulações (r1 e r2) da dinâmica molecular do dímero de HLA-G1, cujos valores de RMSD não ultrapassaram  $15,5 \text{ \AA}$ , mantendo-se majoritariamente na faixa de  $9 \text{ a } 12 \text{ \AA}$ , durante a maior parte da trajetória.



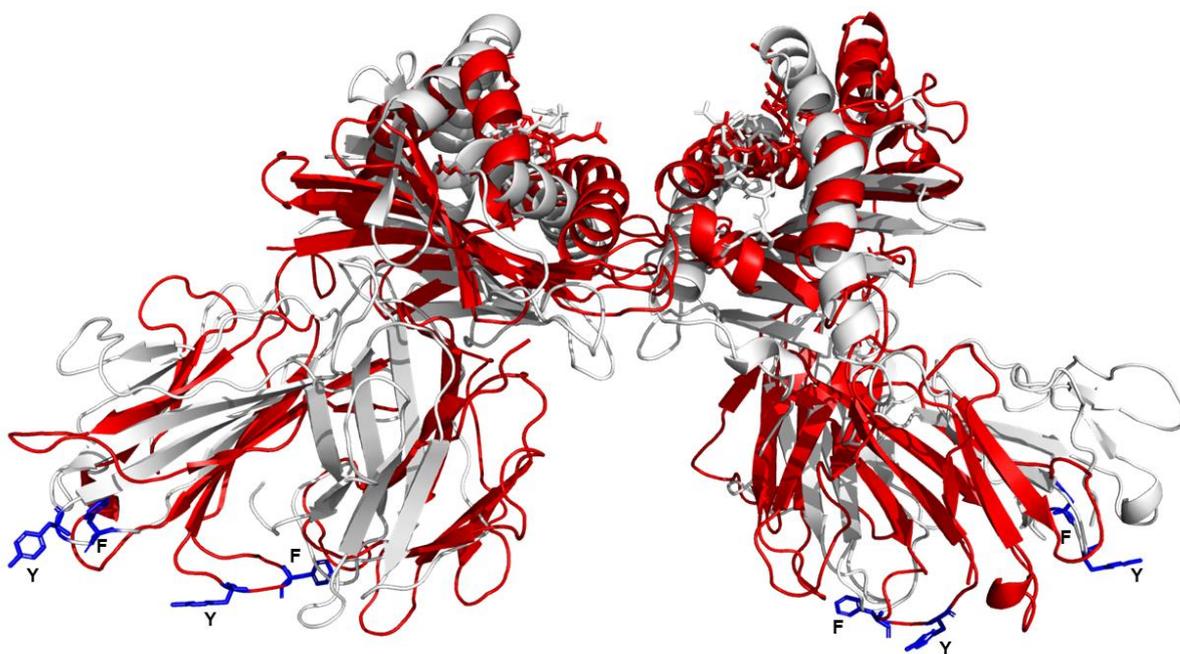
**Figura 43.** Gráfico de RMSD do dímero solúvel de HLA-G1.

O gráfico apresentando o resultado de RMSF para a simulação mais estável do dímero de HLA-G1 demonstrou que a flutuação está mais presente em torno dos aminoácidos que compõem o domínio  $\alpha 3$  (resíduos 200, 225, 250 e 275).

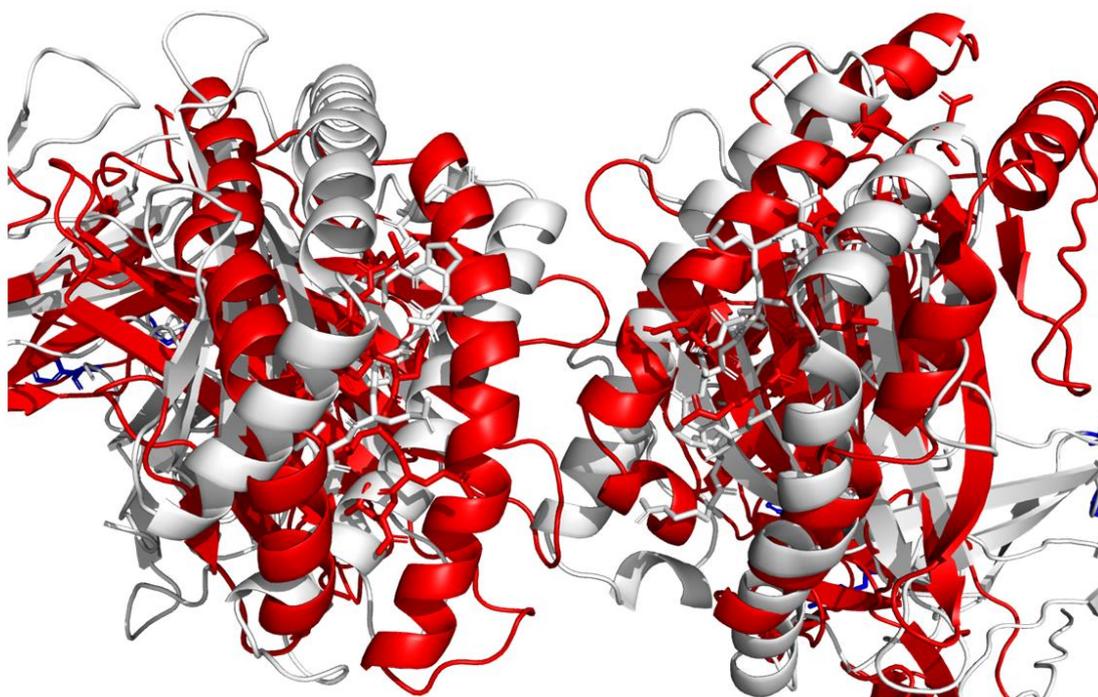


**Figura 44.** Variação estrutural por resíduo da simulação r1 do dímero de HLA-G1. Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.

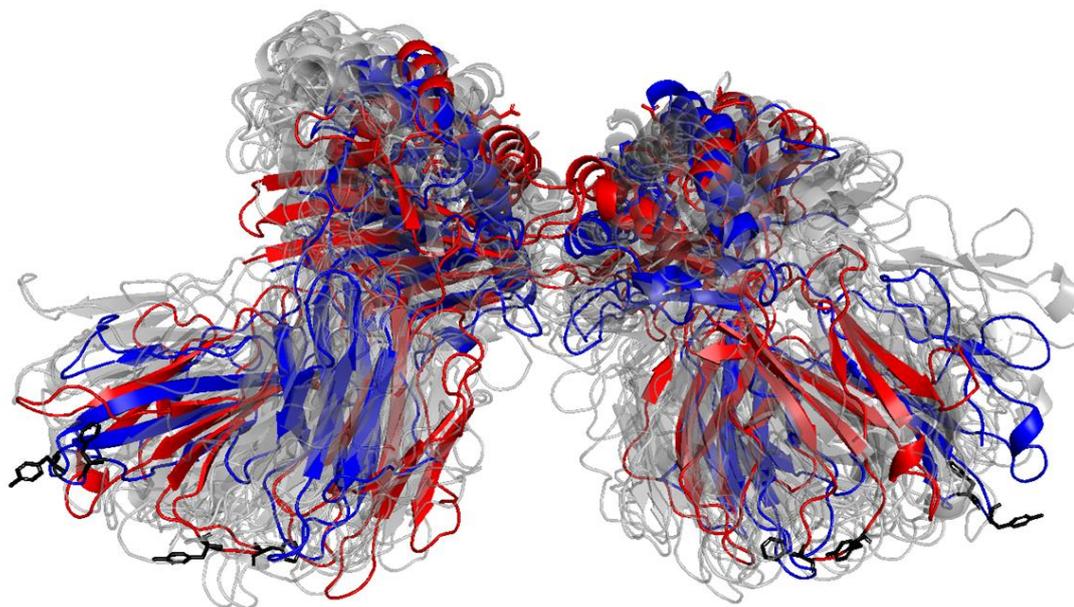
A **Figura 45** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 200ns da dinâmica para o dímero de HLA-G1. Estão representadas nas figuras somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo.



**Figura 45.** Dímero solúvel de HLA-G1 nos momentos iniciais (branco) e finais (vermelho) da simulação de 200ns. Em azul os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197).



**Figura 46.** Aproximação da região dos sulcos que acomodam o peptídeo no dímero solúvel de HLA-G1, nos momentos iniciais (branco) e finais (vermelho) da simulação de 200ns.



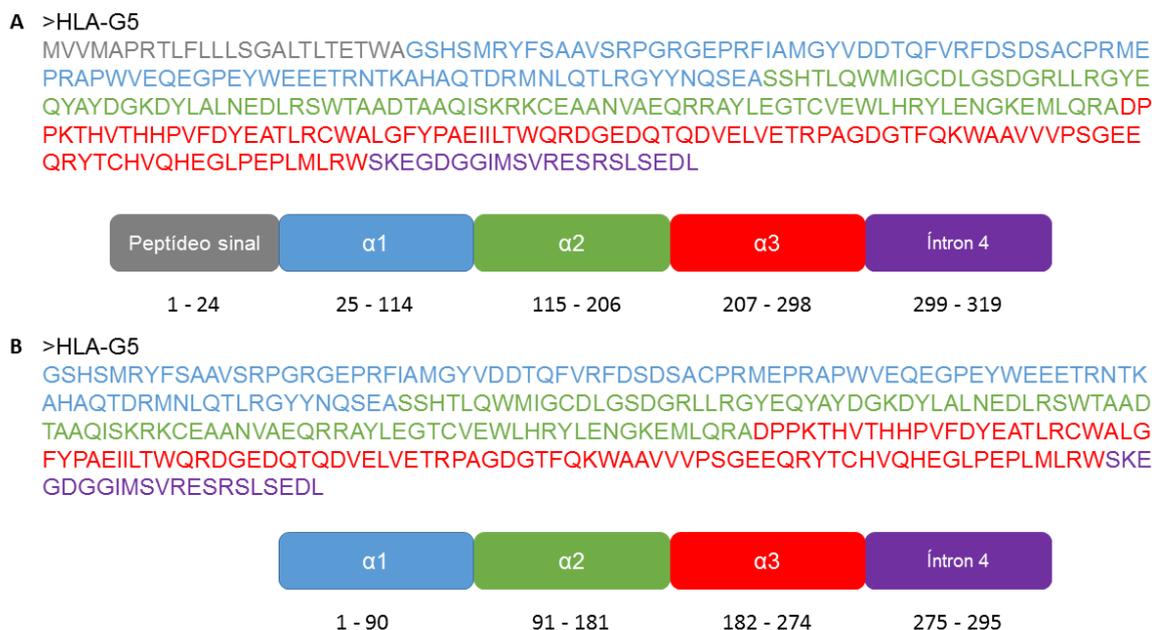
**Figura 47.** Dímero solúvel de HLA-G1 nos momentos iniciais (azul), finais (vermelho) e intervalos de 20ns (cinza claro) da simulação de 200ns. Em preto os resíduos F (fenilalanina, posição 195) e Y (tirosina, posição 197).

### 6.3 Isoforma solúvel HLA-G5

#### 6.3.1 Obtenção da sequência

A montagem da sequência primária das isoformas solúveis de HLA-G foi baseada em artigos publicados na literatura (DONADI *et al.*, 2011; PAUL *et al.*, 2000), na sequência codificada pelo alelo *HLA-G\*01:01:01:01* e confirmada com o depósito das mesmas no banco de dados UniProt (THE UNIPROT CONSORTIUM, 2017).

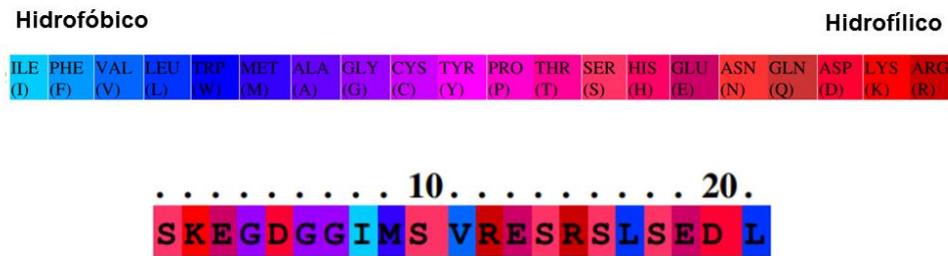
Da mesma forma que a proteína HLA-G, podemos determinar as regiões que compõem a isoforma HLA-G5 com base nas informações já discutidas para a proteína HLA-G completa. Essa isoforma é composta pelos domínios  $\alpha 1$  (25 – 114),  $\alpha 2$  (115 – 206) e  $\alpha 3$  (207 – 298) e pela porção proveniente da tradução do íntron 4. Podemos observar essas informações citadas acima de forma esquemática na **Figura 48**.



**Figura 48.** (A) Sequência primária da isoforma HLA-G5 no formato FASTA. As cores são referentes à porção da proteína compostas por tais aminoácidos, segundo esquema na porção inferior da figura (cinza – peptídeo sinal, azul – domínio  $\alpha 1$ , verde - domínio  $\alpha 2$ , vermelho - domínio  $\alpha 3$ , roxo - porção proveniente da tradução do íntron 4). (B) Sequência primária da isoforma HLA-G5 madura no formato FASTA.

Semelhante ao que ocorre com a proteína HLA-G, a isoforma HLA-G5 após a identificação da região relacionada ao peptídeo sinal e sua remoção, possui 295 aminoácidos:  $\alpha 1$  (1 – 90),  $\alpha 2$ (91 – 181),  $\alpha 3$  (182 – 274) e porção proveniente da tradução do íntron 4 (275 - 295).

O servidor *online* PRALINE (SIMOSSIS *et al.*, 2005) foi utilizado para determinar o perfil de hidrofobicidade dos resíduos que compõem a cauda proveniente da tradução do íntron 4.



**Figura 49.** Perfil de hidrofobicidade dos resíduos que compõem cauda proveniente da tradução do íntron 4, presente nas isoformas HLA-G5 e HLA-G6 (SIMOSSIS *et al.*, 2005).

Os resíduos 1, 2, 3, 5, 10, 11, 12, 13, 14, 15, 18, 19 e 20 apresentam características hidrofílicas.

Por possuir 3 domínios iguais à proteína HLA-G, será utilizada estratégia de modelagem por homologia nessa porção; enquanto a região proveniente do íntron 4 será modelada por metodologia *ab initio*.

### 6.3.2 Modelagem homologia do monômero da isoforma HLA-G5

Para modelagem por homologia foi utilizado o mesmo molde utilizado na modelagem por homologia da proteína HLA-G completa, a cristalografia PDB\_ID:1YDP (CLEMENTS *et al.*, 2005).

Na **Figura 50** abaixo pode-se observar o alinhamento múltiplo entre a sequência primária completa de HLA-G (denominada HLA-G\_completo), a sequência primária da isoforma HLA-G5 e a sequência primária do molde PDB\_ID:1YDP.

```

HLA-G5          GSHSMRYFSAAVSRPGRGEPRFIAMGYVDDTQFVRFDSDSACPRMEPRAPWVEQEGPEYW
HLA-G_completo GSHSMRYFSAAVSRPGRGEPRFIAMGYVDDTQFVRFDSDSACPRMEPRAPWVEQEGPEYW
1YDP           -S HSMRYFSAAVSRPGRGEPRFIAMGYVDDTQFVRFDSDSASPRMEPRAPWVEQEGPEYW
*****

HLA-G5          EEETRNTKAHAQTDRMNLQTLRGYYNQSEASSHTLQWMIIGCDLGS DGRLLRGYEQYAYDG
HLA-G_completo EEETRNTKAHAQTDRMNLQTLRGYYNQSEASSHTLQWMIIGCDLGS DGRLLRGYEQYAYDG
1YDP           EEETRNTKAHAQTDRMNLQTLRGYYNQSEASSHTLQWMIIGCDLGS DGRLLRGYERYAYDG
*****

HLA-G5          KDYLALNEDLRSWTAADTAAQISKRKCEAANVAEQRRAYLEGTCVEWLHRYLENGKEMLQ
HLA-G_completo KDYLALNEDLRSWTAADTAAQISKRKCEAANVAEQRRAYLEGTCVEWLHRYLENGKEMLQ
1YDP           KDYLALNEDLRSWTAADTAAQISKRKCEAANVAEQRRAYLEGTCVEWLHRYLENGKEMLQ
*****

HLA-G5          RADPPKTHVTHHPVFDYEATLRCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGT
HLA-G_completo RADPPKTHVTHHPVFDYEATLRCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGT
1YDP           RADPPKTHVTHHPVFDYEATLRCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGT
*****

HLA-G5          FQKWAAVVWVPSGEEQRYTCHVQHEGLPEPLMLRWKSKEGDGGIMSVRESRSLSE--EDL---
HLA-G_completo FQKWAAVVWVPSGEEQRYTCHVQHEGLPEPLMLRWKQSSLPITIPIMGIVAGLVVLAAVVTG
1YDP           FQKWAAVVWVPSGEEQRYTCHVQHEGLPEPLMLRWKQ-----
*****

HLA-G5          -----
HLA-G_completo AAVA AVLWRKKSSD
1YDP           -----

```

**Figura 50.** Alinhamento múltiplo das seqüências primárias da proteína completa HLA-G (denominada HLA-G\_completo), a seqüência primária da isoforma HLA-G5 (denominada HLA-G5) e a seqüência primária do molde PDB\_ID:1YDP (CLEMENTS *et al.*, 2005). Os resíduos em vermelho e os resíduos restantes foram modelados por homologia, enquanto os resíduos em azul foram modelados por metodologia *ab initio*.

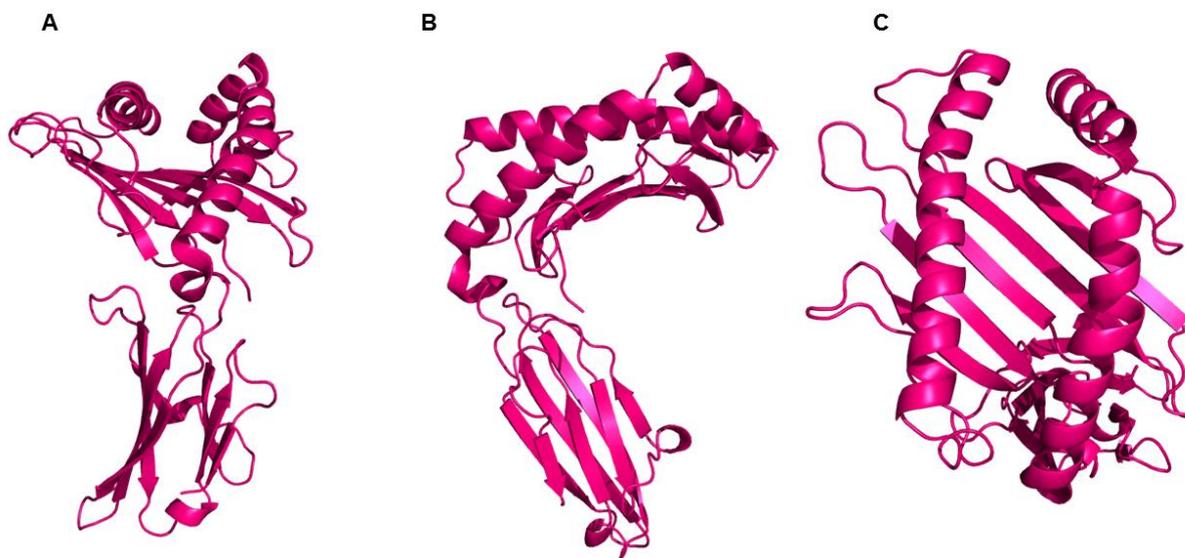
### 6.3.2.1 Avaliação dos modelos gerados por homologia

Foram gerados 05 modelos para a isoforma HLA-G5. Cada um dos cinco modelos gerados para a isoforma HLA-G5 foi avaliado pelos *softwares* PROCHECK (LASKOWSKI *et al.*, 1993), Verify 3D (LUTHY *et al.*, 1992), ERRAT (COLOVOS e YATES, 1993), QMEAN (BENKERT *et al.*, 2008) e ProSA-Web (WIEDERSTEIN *et al.*, 2007), para que o melhor fosse selecionado.

Dentre os modelos gerados, o modelo 1 atingiu as seguintes pontuações, sendo escolhido como melhor modelo gerado:

- Verify 3D: 93.07%.
- ERRAT: 93.07%.
- ProSA-Web: *z-score* de -8.9.
- QMEAN: 0.711.

Esse modelo foi utilizado para as análises e modelagens subsequentes e pode ser observado na **Figura 51**.



**Figura 51.** Melhor modelo (modelo 1) para isoforma HLA-G5, gerada por homologia. Visão frontal da molécula (A), lateral (B) e superior do sulco que acomoda o peptídeo a ser apresentado (C).

### 6.3.3 Modelagem *ab initio*

Novamente, foi utilizada a metodologia de modelagem *ab initio*, de forma semelhante ao protocolo já exposto em seções anteriores, utilizando as ferramentas Robetta e Rosetta (PETTERSEN *et al.*, 2004). Foram modeladas 2000 estruturas *ab initio* contendo os resíduos selecionados, empregando o algoritmo *cyclic coordinate descent* (CCD). Toda a porção já modelada previamente por homologia e descrita nos tópicos anteriores foi mantida fixa ao longo das simulações *ab initio*.

A cauda de 21 resíduos proveniente do íntron 4 foi modelada por metodologia *ab initio* em 3 etapas de modelagem sucessivas, abrangendo os resíduos 274 a 295.

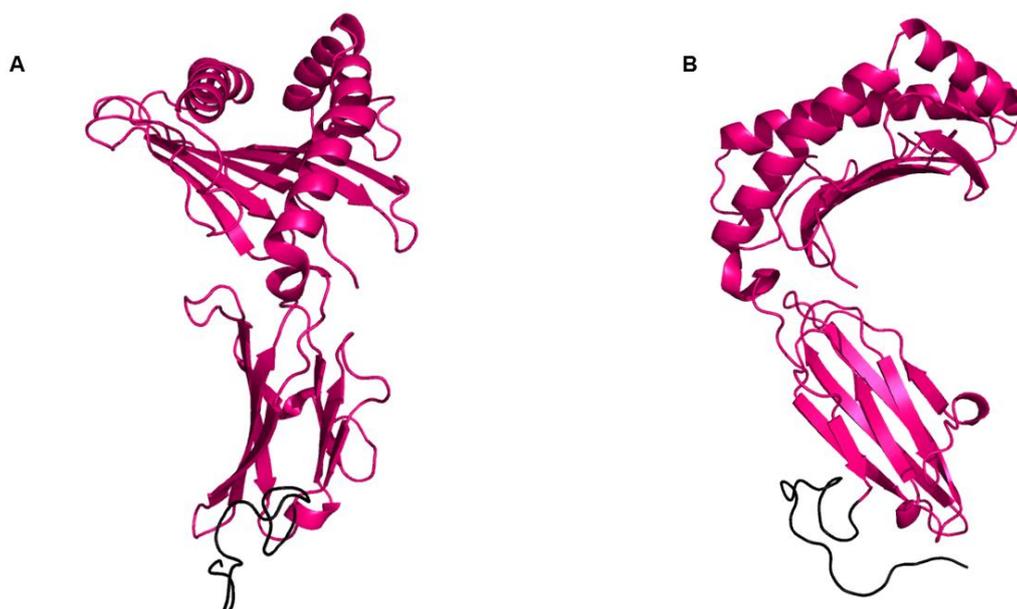
### 6.3.3.1 Seleção dos modelos

A escolha do melhor modelo foi baseada nos melhores resultados das ferramentas QMEAN (BENKERT *et al.*, 2008) e MODFOLD (MCGUFFIN *et al.*, 2008).

O melhor modelo gerado para cada etapa foi selecionado e mantido fixo durante a modelagem *ab initio* das porções seguintes. Os *scores* utilizados para escolha do melhor modelo encontram-se na **Tabela 2**.

**Tabela 2.** Melhores modelos provenientes da metodologia de modelagem *ab initio* para a isoforma HLA-G5.

HLA-G5: <i>ab initio</i>					
Etapa	Resíduos	Modelo	ModFold ( <i>p</i> )	ModFold ( <i>score</i> )	QMEAN
1	274-284	297	$3.46 \times 10^{-8}$	0.781	0.81
2	285-291	502	$4.97 \times 10^{-8}$	0.7726	0.954
3	292 - 295	163	$6.84 \times 10^{-8}$	0.654	0.89



**Figura 52.** Melhor modelo obtido para isoforma HLA-G5, gerada pela combinação de modelagem por homologia e *ab initio*. Visão frontal (A) e lateral (B) da molécula. A porção em preto corresponde aos resíduos modelados utilizando metodologia *ab initio*.

### 6.3.4 Dinâmica molecular

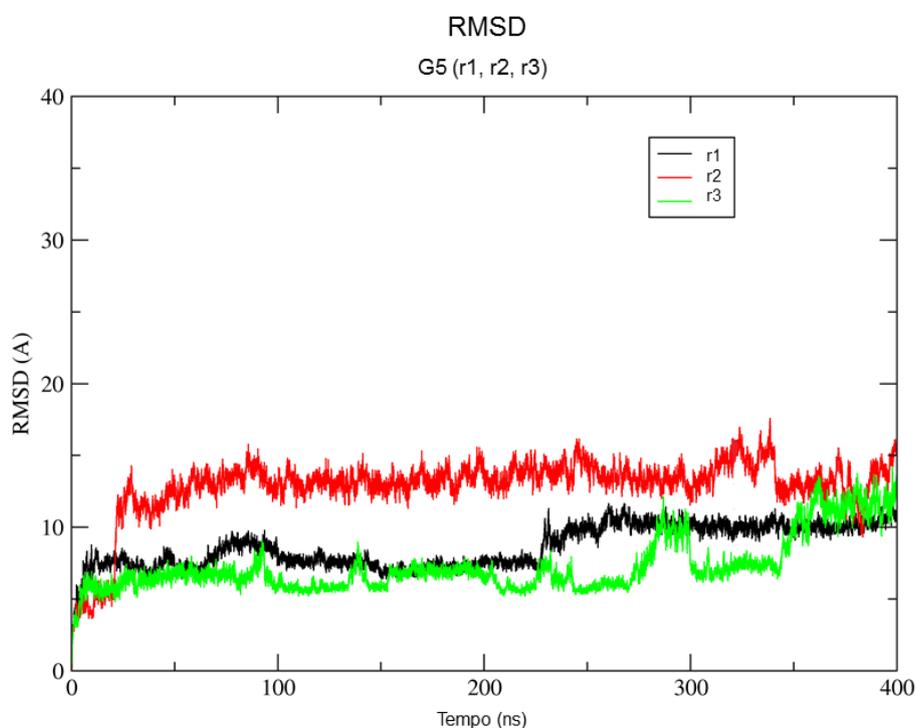
A dinâmica molecular consistiu em triplicata (r1, r2, r3) das simulações para o modelo gerado para o monômero da isoforma HLA-G5. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

O sistema para simulação de dinâmica molecular, além da isoforma HLA-G5 contém:

- 90375 moléculas de água,
- 187 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

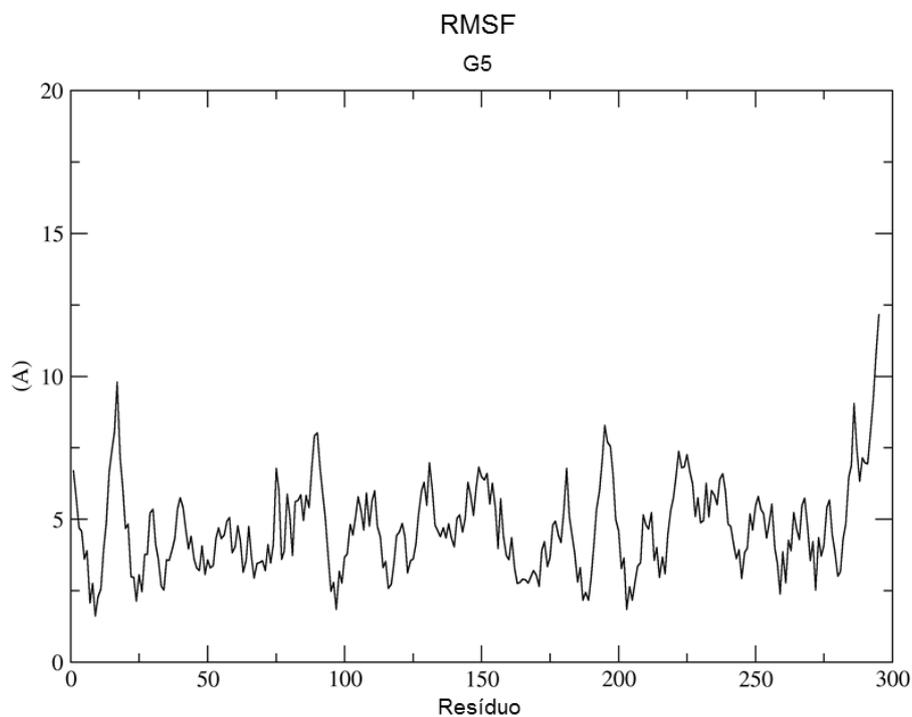
As simulações de dinâmica molecular para o sistema que contém a isoforma HLA-G5 atingiram o tempo total de 400ns.

O gráfico de RMSD apresentado na **Figura 53**, demonstra as triplicatas da dinâmica molecular da isoforma HLA-G5. Podemos observar que as triplicatas da isoforma HLA-G5 apresentam valores de RMSD que chegam a 15,62 Å, enquanto a replicata r1 se mantém na faixa de 8 a 11 Å.



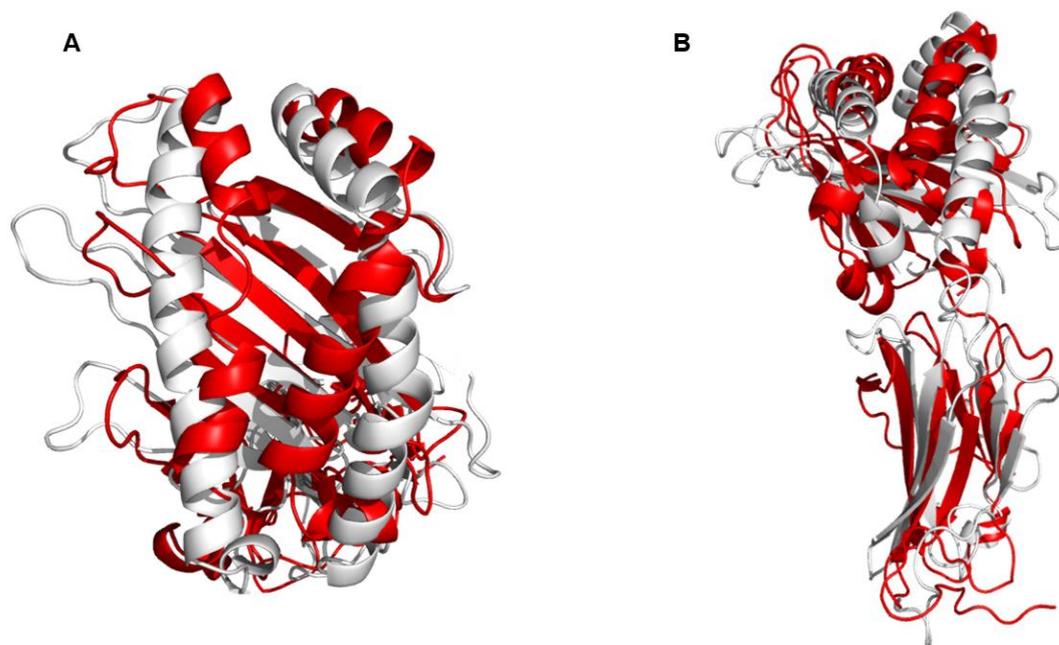
**Figura 53.** Gráfico de RMSD da isoforma HLA-G5.

A **Figura 54** apresenta o gráfico de RMSF para a replicata mais estável obtida durante a simulação do monômero da isoforma HLA-G5.



**Figura 54.** Variação estrutural por resíduo da simulação r1 da isoforma HLA-G5. Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.

A **Figura 55** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 400ns da dinâmica do monômero da isoforma HLA-G5. Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 5.140 Å.



**Figura 55.** Isoforma HLA-G5 nos momentos iniciais (em branco) e finais (em vermelho) dos 400ns da simulação de dinâmica molecular. **(A)** Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$ . **(B)** Visão frontal da isoforma HLA-G5, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.

## 6.4 Isoforma solúvel HLA-G5 contendo peptídeo acoplado no sulco

Para examinar o comportamento dinâmico da estrutura modelada HLA-G5, contendo o peptídeo acoplado ao sulco, foi realizado o *docking* do nonapeptídeo RIIPRHLQL no sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  do monômero da isoforma HLA-G5 demonstrada anteriormente, utilizando o *software* Autodock Vina (TROTT *et al.*, 2010).

### 6.4.1 Seleção do *docking*

A estrutura do nonapeptídeo RIIPRHLQL demonstrada anteriormente foi utilizada para o ensaio *in silico* de *docking* no sulco da isoforma HLA-G5, denominada receptor para esse experimento de *docking*. O experimento *in silico* de *docking* foi executado da mesma forma já

descrita anteriormente para o nonapeptídeo RIIPRHLQL acoplado à fenda da proteína HLA-G completa.

A melhor conformação do peptídeo RIIPRHLQL após o *docking* com a isoforma HLA-G5 apresentou energia de ligação de -7,9 kcal/mol.

O cálculo de RMSD foi executado entre a estrutura proveniente da cristalografia PDB\_ID:1YDP (CLEMENTS *et al.*, 2005) e da isoforma HLA-G5 modelada durante a execução do presente trabalho (contendo as cadeias A e P), sendo o melhor resultado obtido de 1,057 Å.

#### 6.4.2 Dinâmica Molecular

A dinâmica molecular da isoforma HLA-G5 contendo o peptídeo RIIPRHLQL no sulco consistiu em triplicata (r1, r2, r3) das simulações para o modelo gerado. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

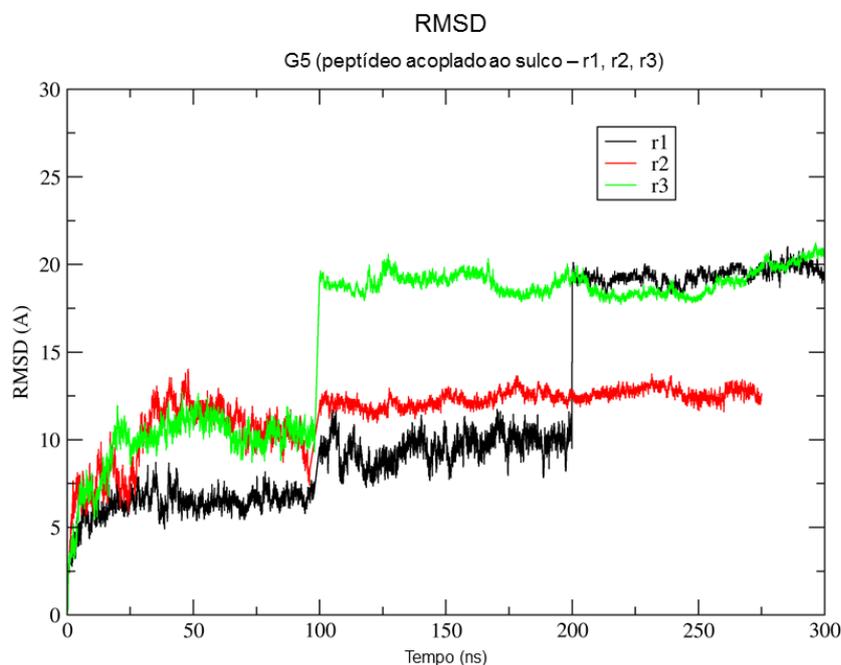
O sistema para simulação de dinâmica molecular, além da isoforma HLA-G5 e o peptídeo RIIPRHLQL contém:

- 90135 moléculas de água,
- 185 moléculas de íons (Na<sup>+</sup> e Cl<sup>-</sup>).

As simulações de dinâmica molecular para o sistema que contém a isoforma HLA-G5 contendo o peptídeo RIIPRHLQL no sulco atingiram o tempo total de 300ns.

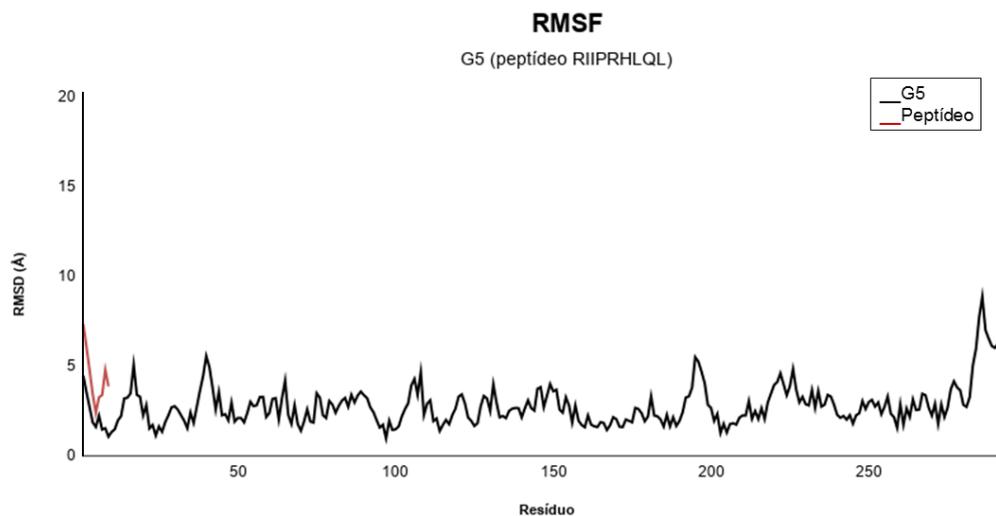
O gráfico de RMSD apresentado na **Figura 56**, demonstra as triplicatas da dinâmica molecular da isoforma HLA-G5 com peptídeo acoplado ao sulco. Podemos observar que as triplicatas da isoforma HLA-G5 com peptídeo acoplado ao sulco apresentam valores de RMSD que chegam a 21,22 Å.

As replicatas r1 e r2 apresentam um aumento brusco nos valores de RMSD em torno de 200ns (r1) e 100ns (r3).



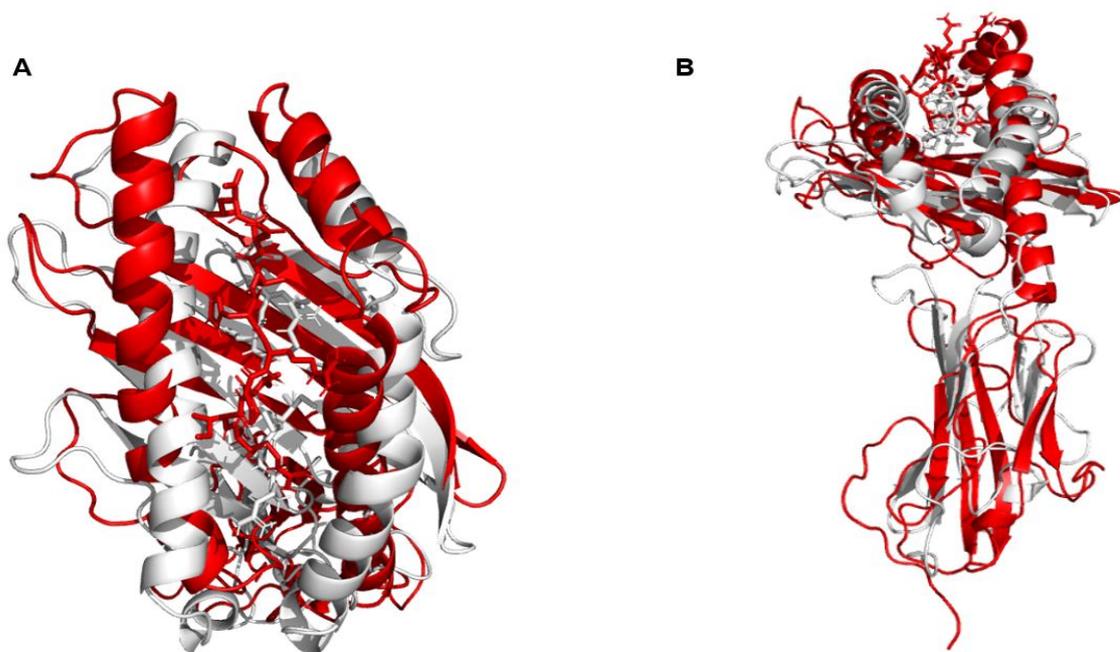
**Figura 56.** Gráfico de RMSD da dinâmica molecular de 300ns referente à isoforma HLA-G5 contendo peptídeo no sulco formado pelas hélices do domínio  $\alpha 1$  e  $\alpha 2$ .

O gráfico apresentando o resultado de RMSF para a simulação mais estável do monômero da isoforma HLA-G5 contendo o peptídeo RIIPRHLQL acoplado ao sulco demonstrou flutuação localizada no peptídeo RIIPRHLQL acoplado, concentrando os maiores valores na porção terminal da estrutura, na porção terminal referente à tradução do íntron 4 (resíduos 275 a 300).



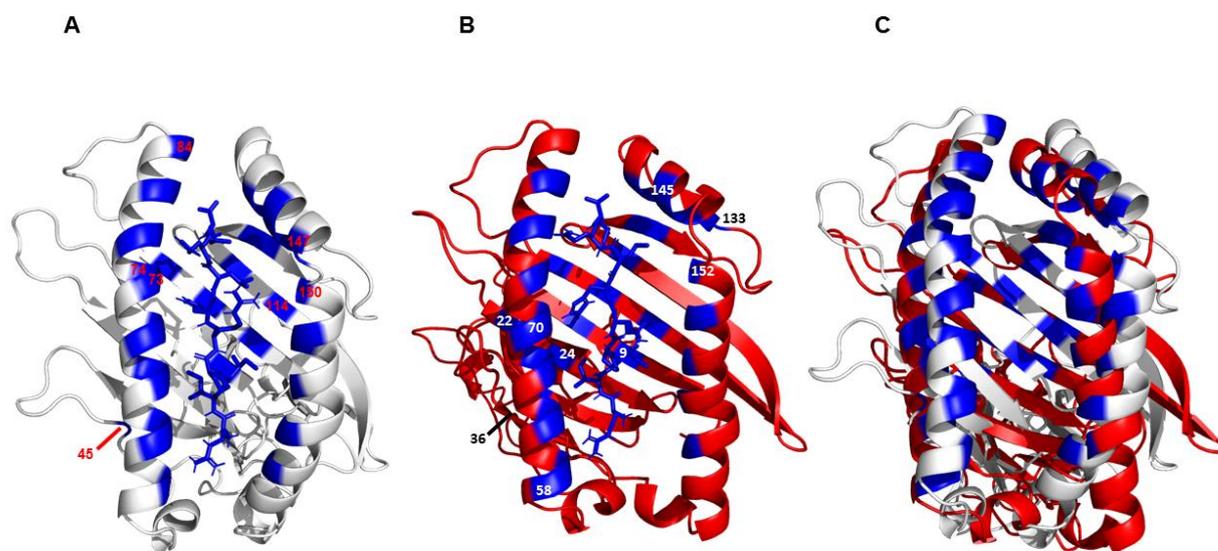
**Figura 57.** Variação estrutural por resíduo da simulação r2 da simulação de dinâmica molecular referente à isoforma HLA-G5 que possui o nonapeptídeo RIIPRHLQL acoplado ao sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$ . Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.

A **Figura 58** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 300ns da dinâmica da isoforma HLA-G5 contendo o peptídeo RIIPRHLQL acoplado ao sulco. Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 6,244 Å.



**Figura 58.** Isoforma HLA-G5 contendo peptídeo no sulco formado pelas hélices do domínio  $\alpha 1$  e  $\alpha 2$  nos momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular. **(A)** Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  e o peptídeo. **(B)** Visão frontal da isoforma HLA-G5, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.

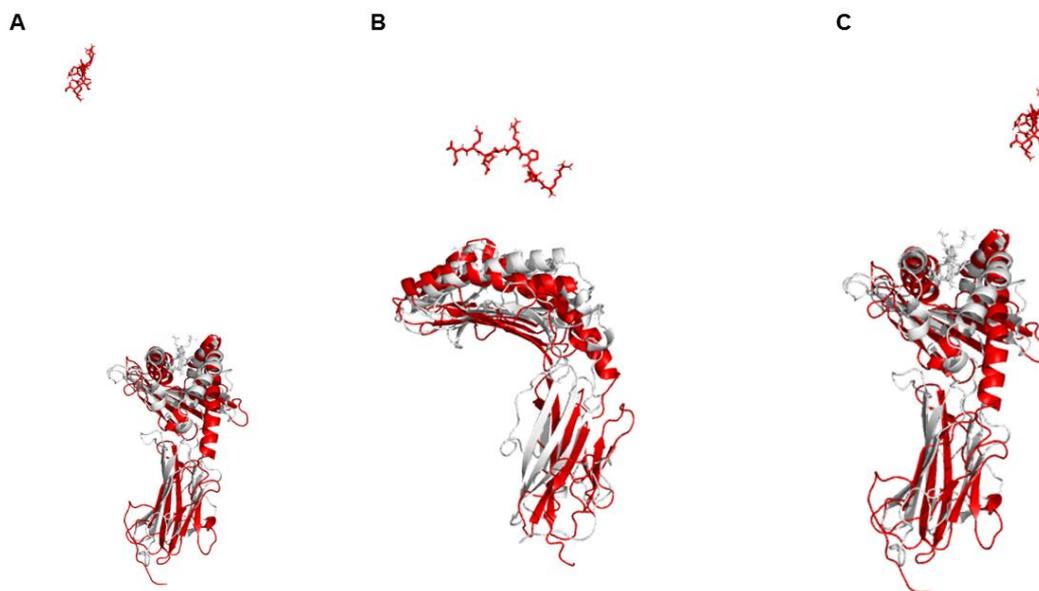
O pacote BioPython (COCK *et al.*, 2009) foi utilizado para verificar os resíduos que fazem parte das interações proteína-peptídeo nos momentos iniciais (branco) e finais (vermelho) da simulação referente à isoforma HLA-G5 contendo o peptídeo RIIPRHLQL acoplado ao sulco. Na **Figura 59** pode-se observar em azul os resíduos da isoforma HLA-G5 que interagem com o peptídeo RIIPRHLQL, nos momentos iniciais e finais da simulação de dinâmica molecular.



**Figura 59.** Isoforma HLA-G5 contendo peptídeo no sulco formado pelas hélices do domínio  $\alpha 1$  e  $\alpha 2$  nos momentos iniciais (A, em branco) e finais (B, em vermelho) dos 300ns da simulação de dinâmica molecular. **(A)** Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  e o peptídeo, demonstrando as interações dos resíduos do peptídeo com o sulco durante os momentos iniciais da simulação de dinâmica molecular. Os resíduos únicos para os momentos iniciais estão numerados. **(B)** Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  e o peptídeo, demonstrando as interações dos resíduos do peptídeo com o sulco durante os momentos finais da simulação de dinâmica molecular. Os resíduos únicos para os momentos finais estão numerados. **(C)** Comparação das interações dos resíduos do peptídeo com o sulco nos momentos iniciais (estrutura em branco) e finais (estrutura em vermelho), sem o peptídeo na fenda para melhor observação do assoalho formado por folhas- $\beta$ .

Resíduos envolvidos interação com o peptídeo nos momentos iniciais da simulação de dinâmica molecular r2: 7, **45**, 59, 62, 63, 66, 67, 69, 70, **73, 74**, 76, 77, 80, 81, **84**, 95, 97, 99, **114**, 116, 123, 124, 143, 146, **147, 150**, 155, 156, 159, 163, 167 (em negrito os resíduos únicos para momentos iniciais).

Resíduos envolvidos interação com o peptídeo nos momentos finais da simulação de dinâmica molecular r2: 7, **9, 22, 24, 36, 58**, 59, 62, 63, 66, 67, 69, **70**, 76, 77, 80, 81, 95, 97, 99, 116, 123, 124, **133**, 143, **145**, 146, **152**, 155, 156, 159, 163, 167 (em negrito os resíduos únicos para momentos finais).



**Figura 60.** Isoforma HLA-G5 contendo peptídeo no sulco formado pelas hélices do domínio  $\alpha 1$  e  $\alpha 2$  nos momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular. Visualização frontal (**A**, **C**) e lateral (**B**) da isoforma HLA-G5, mostrando momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular, onde pode ser observado o peptídeo fora do sulco (em vermelho).

## 6.5 Isoforma solúvel HLA-G5 contendo peptídeo acoplado ao sulco e acoplada à $\beta 2$ -microglobulina

A estrutura gerada anteriormente contendo a isoforma HLA-G5 e peptídeo RIIPRHLQL localizado no sulco foi acoplada à  $\beta 2$ -microglobulina.

### 6.5.1 Modelagem da proteína $\beta 2$ -microglobulina acoplada à isoforma HLA-G5

A estrutura utilizada como molde para a modelagem por homologia foi a isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL acoplado ao sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$ , anteriormente demonstrada.

Para obter a modelagem por homologia da isoforma HLA-G5, contendo o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina, foi utilizado o *software* Modeller 9.15 (WEBB *et al.*, 2002), especificamente a função de modelagem de múltiplas cadeias (*multiple chains*).

Como exposto anteriormente, a estrutura cristalográfica PDB\_ID:1YDP (CLEMENTS *et al.*, 2005) que foi utilizada como molde desde a primeira modelagem por homologia para a proteína HLA-G contém também a estrutura da proteína  $\beta$ 2-microglobulina resolvida (cadeia B), podendo servir de molde para a modelagem isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina.

Foram gerados 05 modelos da isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina. Tal estrutura gerada contém 295 resíduos referentes à estrutura da isoforma HLA-G5, 9 resíduos referentes ao peptídeo e 99 aminoácidos referentes à  $\beta$ 2-microglobulina.

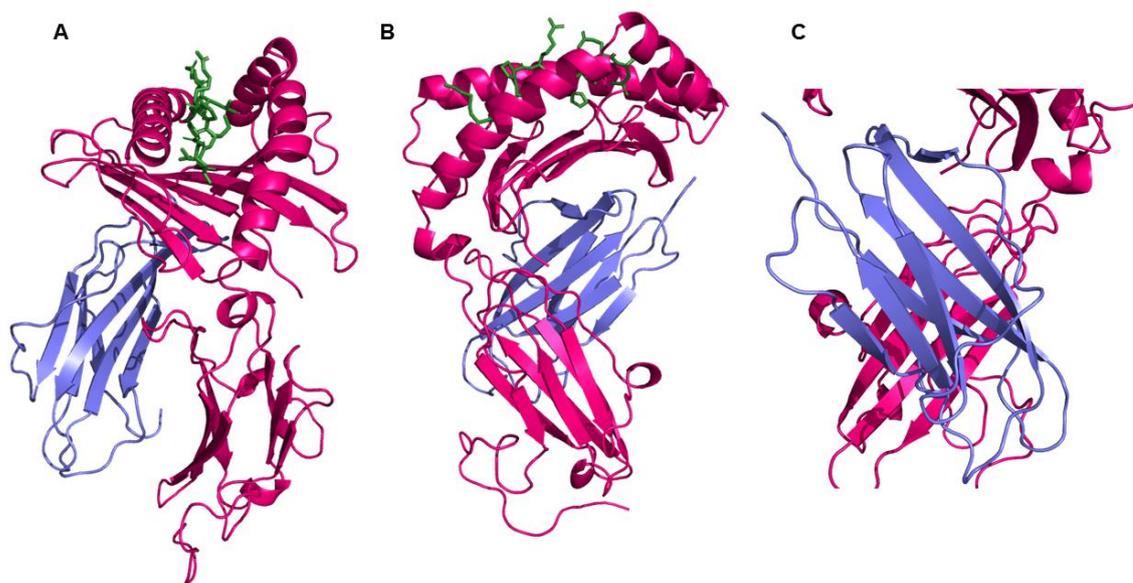
#### 6.5.1.1 Avaliação e seleção dos modelos

Cada um dos cinco modelos gerados foi avaliado pelos *softwares* PROCHECK (LASKOWSKI *et al.*, 1993), Verify 3D (LUTHY *et al.*, 1992), ERRAT (COLOVOS e YATES, 1993), QMEAN (BENKERT *et al.*, 2008), e ProSA-Web (WIEDERSTEIN *et al.*, 2007), para que o melhor fosse selecionado.

Dentre os modelos gerados, o modelo 2 atingiu as seguintes pontuações, sendo escolhido como melhor modelo gerado:

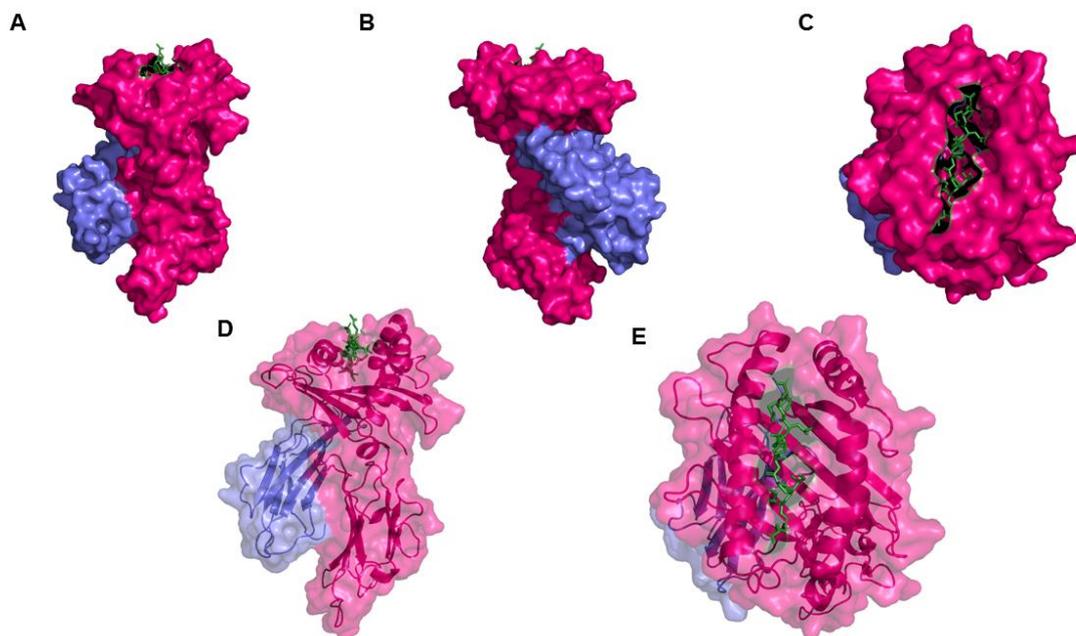
- Verify 3D: 89.3%.
- ERRAT: 91.84%.
- ProSA-Web: *z-score* de -7.41.
- QMEAN: 0.724.

Esse modelo foi utilizado para as análises subsequentes e pode ser observado na **Figura 61**.



**Figura 61.** Melhor modelo obtido para isoforma HLA-G5 (rosa) com o nonapeptídeo RIIPRHLQL (verde) inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina (lilás). Visão frontal (A), lateral (B) e aproximação da cadeia composta pela  $\beta$ 2-microglobulina (C).

A **Figura 62** apresenta a visualização da superfície da isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina, além da visualização da transparência da mesma estrutura, possibilitando a visão da estrutura secundária (em formato *cartoon*).



**Figura 62.** Visualização do melhor modelo obtido para isoforma HLA-G5 (rosa) com o nonapeptídeo RIIPRHLQL (verde) inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina (lilás). Representação de superfície; visão frontal (A), traseira (B) e sulco do peptídeo (C). Representação de superfície com transparência (60%), possibilitando a visualização da estrutura secundária (em formato *cartoon*) da isoforma HLA-G5 (rosa) com o nonapeptídeo RIIPRHLQL (verde) inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina (lilás). Visão frontal (D) e sulco do peptídeo (E).

### 6.5.2 Dinâmica Molecular

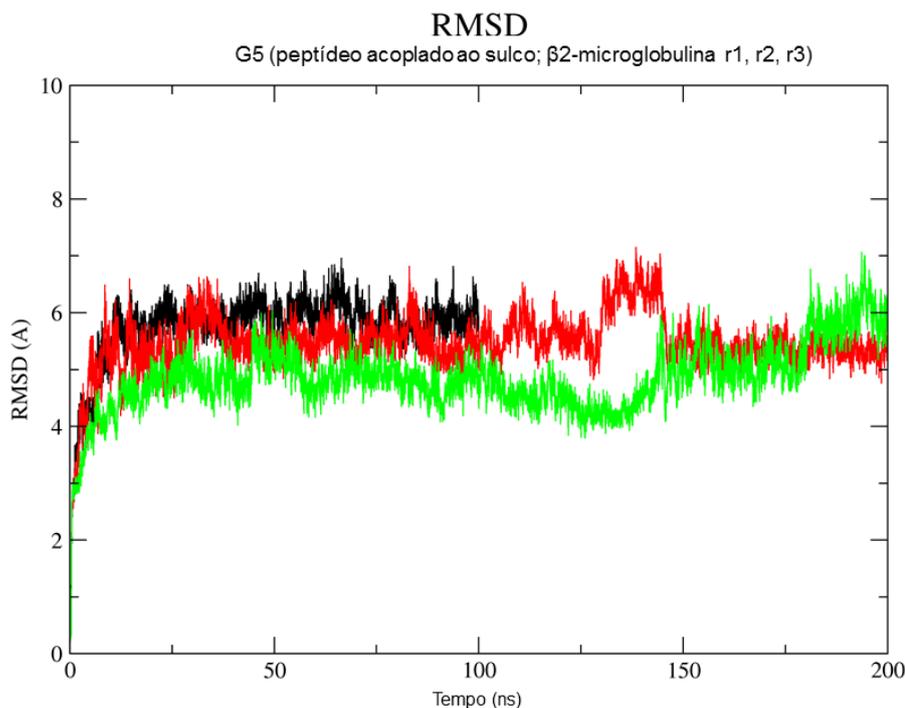
A dinâmica molecular para isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina consistiu em triplicata (r1, r2, r3) das simulações para o modelo gerado, utilizando o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

O sistema para simulação de dinâmica molecular, além da isoforma HLA-G5, o peptídeo RIIPRHLQL e a proteína  $\beta$ 2-microglobulina contém:

- 83676 moléculas de água,
- 179 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

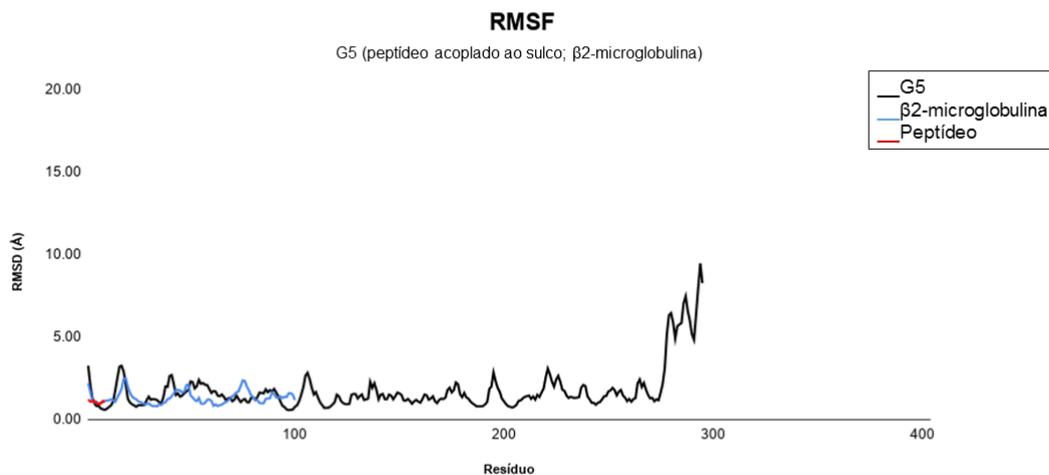
As simulações de dinâmica molecular para o sistema que contém a isoforma HLA-G5 atingiram o tempo total de 200ns.

O gráfico de RMSD, apresentado na **Figura 63**, demonstra as triplicatas da dinâmica molecular isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina. Podemos observar que as triplicatas da isoforma HLA-G5 com peptídeo acoplado ao sulco apresentam valores máximos de RMSD de 7,19 Å, mantendo-se na faixa de 4 a 6 Å durante a maior parte da trajetória.



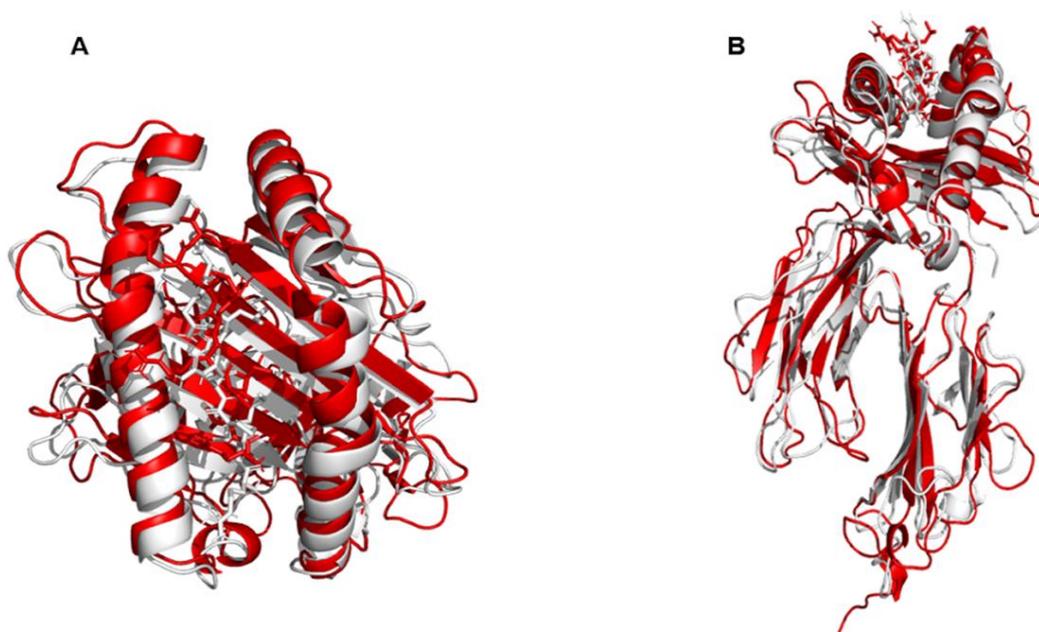
**Figura 63.** Gráfico de RMSD da dinâmica molecular de 200ns referente à isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina.

O gráfico apresentando o resultado de RMSF para a simulação mais estável da isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina demonstrou maior flutuação localizada na porção terminal da estrutura, referente à tradução do íntron 4 (resíduos 275 a 300).



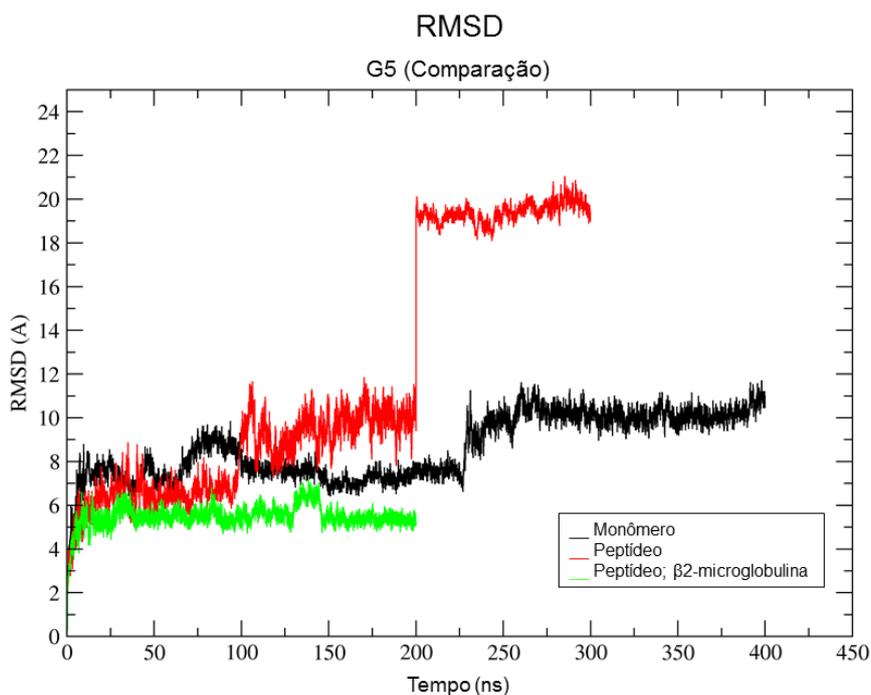
**Figura 64.** Variação estrutural por resíduo da simulação r3 de dinâmica molecular referente à isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL inserido no sulco e acoplada à proteína  $\beta$ 2-microglobulina. Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.

O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 2,361 Å.



**Figura 65.** Isoforma HLA-G5, peptídeo RIIPRHLQL acoplado ao sulco e proteína  $\beta$ 2-microglobulina nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular. **(A)** Visão superior da isoforma HLA-G5, mostrando em primeiro plano o sulco formado entre o domínio  $\alpha$ 1 e  $\alpha$ 2 e o peptídeo. **(B)** Visão frontal da isoforma HLA-G5, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.

Para possibilitar a comparação entre os valores de RMSD das simulações envolvendo a isoforma HLA-G5 em suas 3 alternativas estruturais (HLA-G5 – monômero; HLA-G5 e peptídeo RIIPRHLQL acoplado ao sulco e HLA-G5, peptídeo RIIPRHLQL acoplado ao sulco e proteína  $\beta$ 2-microglobulina) foi gerado um gráfico comparativo (**Figura 66**).



**Figura 66.** Gráfico de RMSD comparativo das simulações envolvendo a isoforma HLA-G5 em suas 3 alternativas estruturais (HLA-G5 – monômero; HLA-G5 e peptídeo RIIPRHLQL acoplado ao sulco e HLA-G5, peptídeo RIIPRHLQL acoplado ao sulco e proteína  $\beta$ 2-microglobulina).

## 6.6 Isoforma solúvel HLA-G6

### 6.6.1 Obtenção da sequência

A montagem da sequência primária da isoforma solúvel HLA-G6 foi realizada de forma semelhante à isoforma HLA-G5, descrita anteriormente.

Da mesma forma que a isoforma HLA-G5, podemos determinar as regiões que compõem a isoforma HLA-G6 com base nas informações já discutidas para a proteína HLA-G

completa. Essa isoforma é composta pelos domínios  $\alpha 1$  (25 – 114),  $\alpha 3$  (207 – 298) e pela porção proveniente da tradução do íntron 4. Podemos observar essas informações citadas acima de forma esquemática na **Figura 67**.



**Figura 67.** (A) Sequência primária da isoforma HLA-G6 no formato FASTA. As cores são referentes à porção da proteína compostas por tais aminoácidos, segundo esquema na porção inferior da figura (cinza – peptídeo sinal, azul – domínio  $\alpha 1$ , vermelho - domínio  $\alpha 3$ , roxo - porção proveniente da tradução do íntron 4). (B) Sequência primária da isoforma madura HLA-G6 no formato FASTA.

Após a identificação da região relacionada ao peptídeo sinal e sua remoção, a isoforma HLA-G6 possui 203 aminoácidos:  $\alpha 1$  (1 – 90),  $\alpha 3$  (91 – 181) e porção proveniente da tradução do íntron 4 (182 - 203).

Por possuir 2 domínios iguais à proteína HLA-G completa, foi utilizada estratégia de modelagem por homologia nessa porção; enquanto a região proveniente do íntron 4 foi modelada por metodologia *ab initio*.

## 6.6.2 Modelagem homologia do monômero da isoforma HLA-G6

Para modelagem por homologia foi utilizado o mesmo molde utilizado na modelagem por homologia da proteína HLA-G completa, a cristalografia PDB\_ID:1YDP (CLEMETS *et al.*, 2005).

Na **Figura 68** abaixo pode-se observar o alinhamento múltiplo entre a sequência primária completa de HLA-G (denominada HLA-G\_completo), a sequência primária da isoforma HLA-G6 e a sequência primária do molde PDB\_ID:1YDP.

```

HLA-G6          GSHSMRYFSAAVSRPGRGEPRFIAMGYVDDTQFVRFSDSACPRMEPRAPWVEQEGPEYW
HLA-G_completo GSHSMRYFSAAVSRPGRGEPRFIAMGYVDDTQFVRFSDSACPRMEPRAPWVEQEGPEYW
1YDP           -SHSMRYFSAAVSRPGRGEPRFIAMGYVDDTQFVRFSDSASPRMEPRAPWVEQEGPEYW
*****

HLA-G6          EEETRNTKAHAQTDRMNLQTLRGYYNQSEA-----
HLA-G_completo EEETRNTKAHAQTDRMNLQTLRGYYNQSEASSHTLQWMIGCDLGS DGRLLRGYEQYAYDG
1YDP           EEETRNTKAHAQTDRMNLQTLRGYYNQSEASSHTLQWMIGCDLGS DGRLLRGYERYAYDG
*****

HLA-G6          -----
HLA-G_completo KDYLALNEDLRSWTAADTAAQISKRKCEANVAEQRRAYLEGTCVEWLHRYLENGKEMLQ
1YDP           KDYLALNEDLRSWTAADTAAQISKRKCEANVAEQRRAYLEGTCVEWLHRYLENGKEMLQ

HLA-G6          --DPPKTHVTHHPVFDYEATLRCWALGFYP AEIILTWQRDGEDQTQDVELVETRPAGDGT
HLA-G_completo RADPPKTHVTHHPVFDYEATLRCWALGFYP AEIILTWQRDGEDQTQDVELVETRPAGDGT
1YDP           RADPPKTHVTHHPVFDYEATLRCWALGFYP AEIILTWQRDGEDQTQDVELVETRPAGDGT
*****

HLA-G6          FQKWA AVVPSGEEQRYTCHVQHEGLPEPLMLRWSKEG DGGIMSVRESRSL S--EDL ---
HLA-G_completo FQKWA AVVPSGEEQRYTCHVQHEGLPEPLMLRWSKSSLPTIPIMGIVAGLVVLA AVVTG
1YDP           FQKWA AVVPSGEEQRYTCHVQHEGLPEPLMLRWSKQ-----
*****
. :

HLA-G6          -----
HLA-G_completo AAVA AVLWRKKSSD
1YDP           -----

```

**Figura 68.** Alinhamento múltiplo das sequências primárias da proteína completa HLA-G (denominada HLA-G\_completo), a sequência primária da isoforma HLA-G6 (denominada HLA-G6) e a sequência primária do molde PDB\_ID:1YDP (CLEMETS *et al.*, 2005). Os resíduos em vermelho e resíduos alinhados foram modelados por homologia, enquanto os resíduos em azul foram modelados por metodologia *ab initio*.

### 6.6.2.1 Avaliação dos modelos gerados por homologia

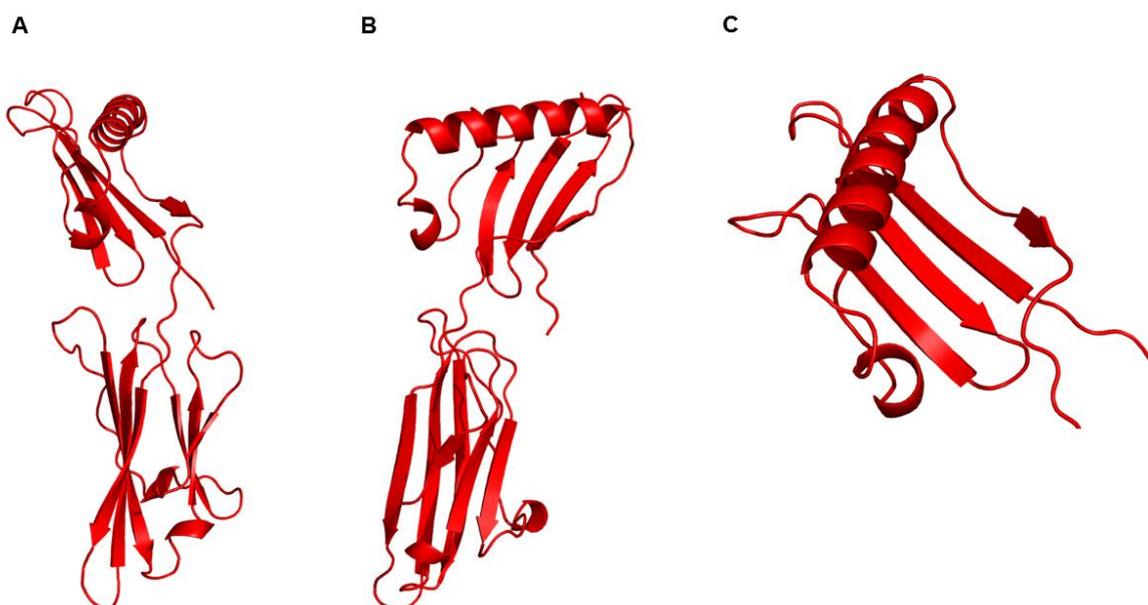
Foram gerados 05 modelos para a isoforma HLA-G6. Cada um dos cinco modelos gerados para a isoforma HLA-G6 foi avaliado pelos *softwares* PROCHECK (LASKOWSKI *et*

*al.*, 1993), Verify 3D (LUTHY *et al.*, 1992), ERRAT (COLOVOS e YATES, 1993), QMEAN (BENKERT *et al.*, 2008) e ProSA-Web (WIEDERSTEIN *et al.*, 2007), para que o melhor fosse selecionado.

Dentre os modelos gerados, o modelo 1 atingiu as seguintes pontuações, sendo escolhido como melhor modelo gerado:

- Verify 3D: 81.87%.
- ERRAT: 87.6%.
- ProSA-Web: *z-score* de -6.93.
- QMEAN: 0.644.

Esse modelo foi utilizado para as análises e modelagens subsequentes e pode ser observado na **Figura 69**.



**Figura 69.** Melhor modelo (modelo 1) para isoforma HLA-G6, gerado por homologia. Visão frontal da molécula (A), lateral (B) e superior do domínio  $\alpha 1$  (C).

### 6.6.3 Modelagem *ab initio*

Novamente, foi utilizada a metodologia de modelagem *ab initio*, de forma semelhante ao protocolo já exposto para isoforma HLA-G5, utilizando as ferramentas Robetta e Rosetta (PETTERSEN *et al.*, 2004). A cauda de 21 resíduos proveniente do íntron 4 foi modelada por metodologia *ab initio* em 3 etapas de modelagem sucessivas, abrangendo os resíduos 182 a 203.

#### 6.6.3.1 Seleção dos modelos

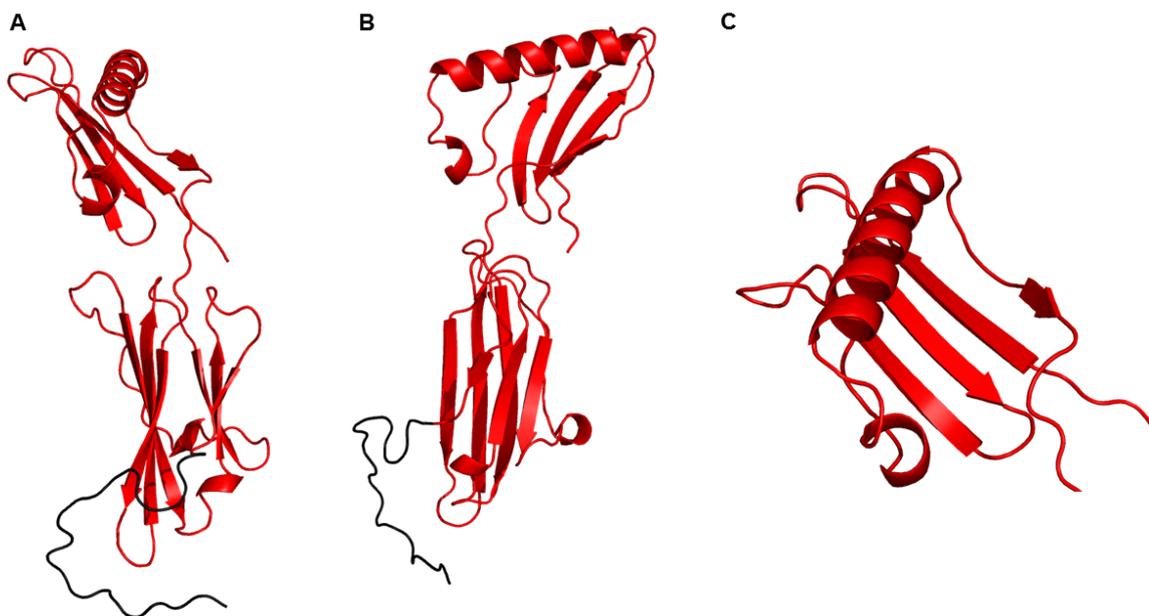
A escolha do melhor modelo foi baseada nos melhores resultados das ferramentas QMEAN (BENKERT *et al.*, 2008) e MODFOLD (MCGUFFIN *et al.*, 2008).

O melhor modelo gerado para cada etapa foi selecionado e mantido fixo durante a modelagem *ab initio* das porções seguintes. Os *scores* utilizados para escolha do melhor modelo encontram-se na **Tabela 3**.

**Tabela 3.** Melhores modelos provenientes da metodologia de modelagem *ab initio* para a isoforma HLA-G6.

#### HLA-G6: *ab initio*

Etapa	Resíduos	Estrutura	ModFold (p)	ModFold (score)	QMEAN
<b>1</b>	182 - 190	511	$9.63 \times 10^{-6}$	0.69	0.81
<b>2</b>	191 - 197	57	$7.98 \times 10^{-7}$	0.754	0.713
<b>3</b>	198 - 203	146	$5.20 \times 10^{-7}$	0.761	0.89



**Figura 70.** Melhor modelo obtido para isoforma HLA-G6, gerado pela combinação de modelagem por homologia e *ab initio*. Visão frontal (A) e lateral (B) da molécula. A porção em preto corresponde aos resíduos modelados utilizando metodologia *ab initio*.

#### 6.6.4 Dinâmica molecular

A dinâmica molecular consistiu em triplicata (r1, r2, r3) das simulações para o modelo gerado para o monômero da isoforma HLA-G6. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

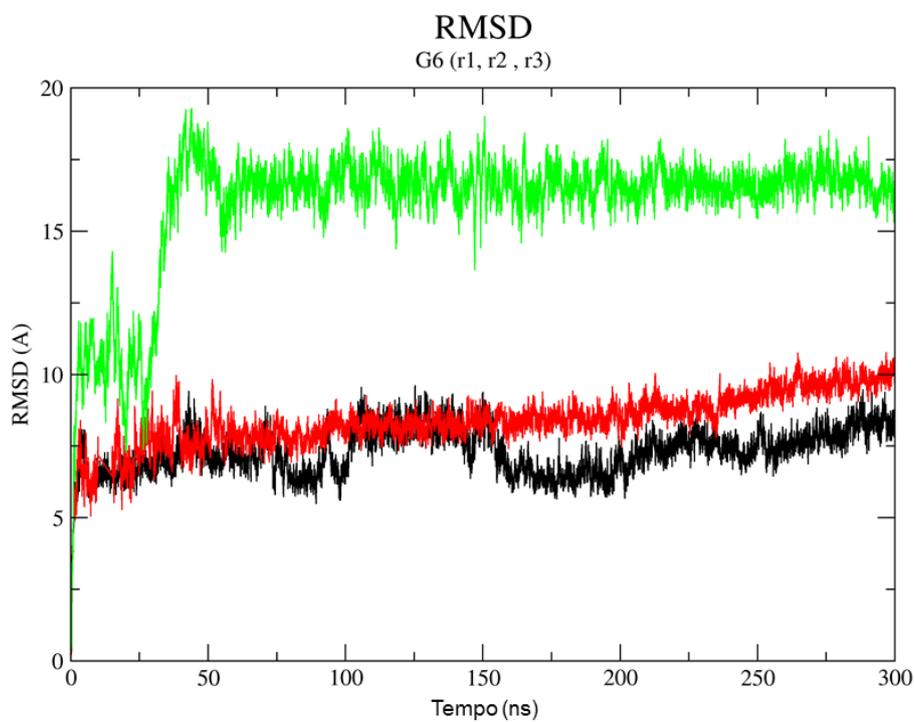
O sistema para simulação de dinâmica molecular, além da isoforma HLA-G6 contém:

- 125931 moléculas de água,
- 246 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

As simulações de dinâmica molecular para o sistema que contém a isoforma HLA-G6 atingiram o tempo total de 300ns.

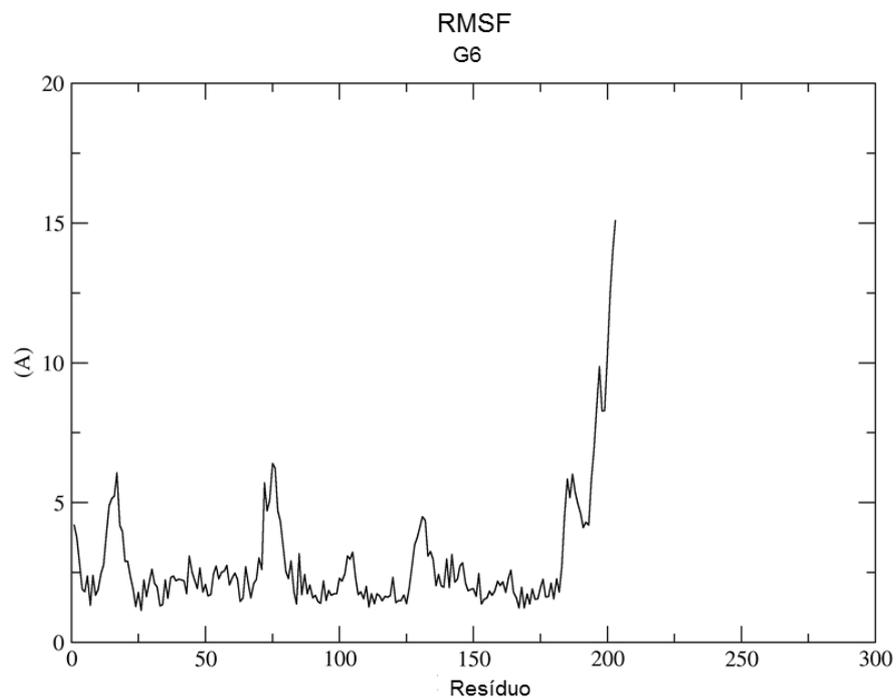
O gráfico de RMSD, apresentado na **Figura 71**, demonstra as triplicatas da dinâmica molecular da isoforma HLA-G6. Podemos observar que as triplicatas da isoforma HLA-G6 apresentam valores de RMSD máximos de 18,8 Å, sendo que as simulações r1 e r2 mantêm-se

na faixa de 7 a 10 Å durante a trajetória, enquanto r3 mantém-se na faixa de 17 Å após 50ns de simulação.



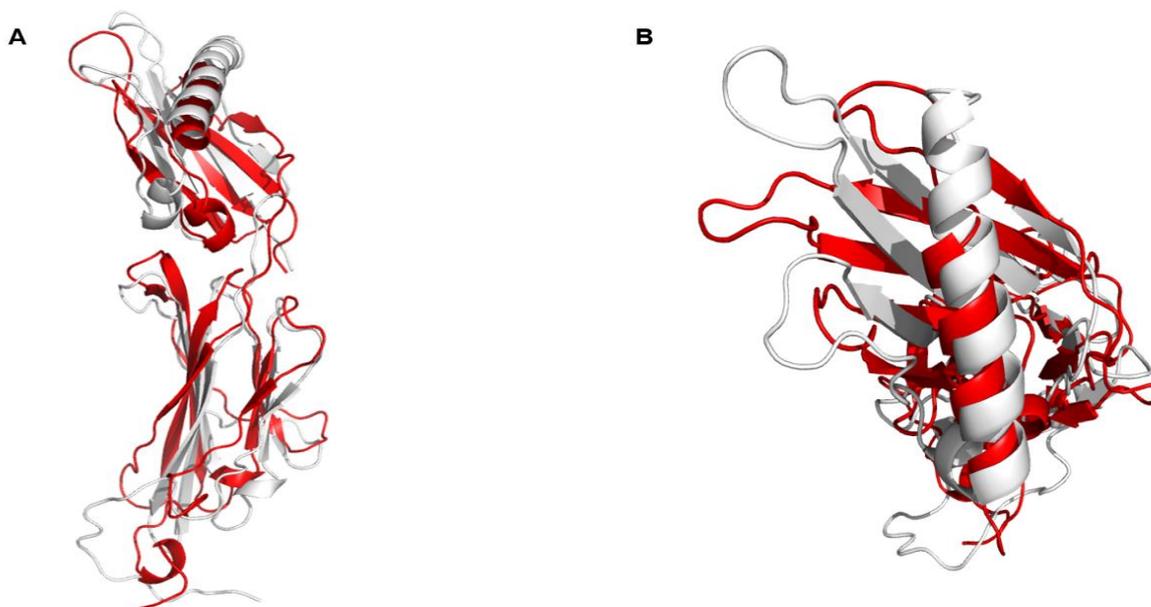
**Figura 71.** Gráfico de RMSD da dinâmica molecular de 300ns referente ao monômero da isoforma HLA-G6.

O gráfico apresentando o resultado de RMSF para a simulação mais estável do monômero da isoforma HLA-G6 demonstrou que a flutuação atingiu os maiores níveis nas porções dos resíduos 15 a 20, 70 a 85 (domínio  $\alpha 1$ ) e porção terminal referente à tradução do íntron 4.



**Figura 72.** Variação estrutural por resíduo da simulação r1 da isoforma HLA-G6. Valores de RMSF das coordenadas dos átomos  $C\alpha$  em relação à estrutura inicial.

A **Figura 73** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 300ns da dinâmica do monômero da isoforma HLA-G6. Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 4,819 Å.



**Figura 73.** Isoforma HLA-G6 nos momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular. **(A)** Visão superior da isoforma HLA-G6, mostrando em primeiro plano o domínio  $\alpha 1$ . **(B)** Visão frontal da isoforma HLA-G6, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.

## 6.7 Isoforma solúvel HLA-G6 contendo peptídeo acoplado ao domínio $\alpha 1$

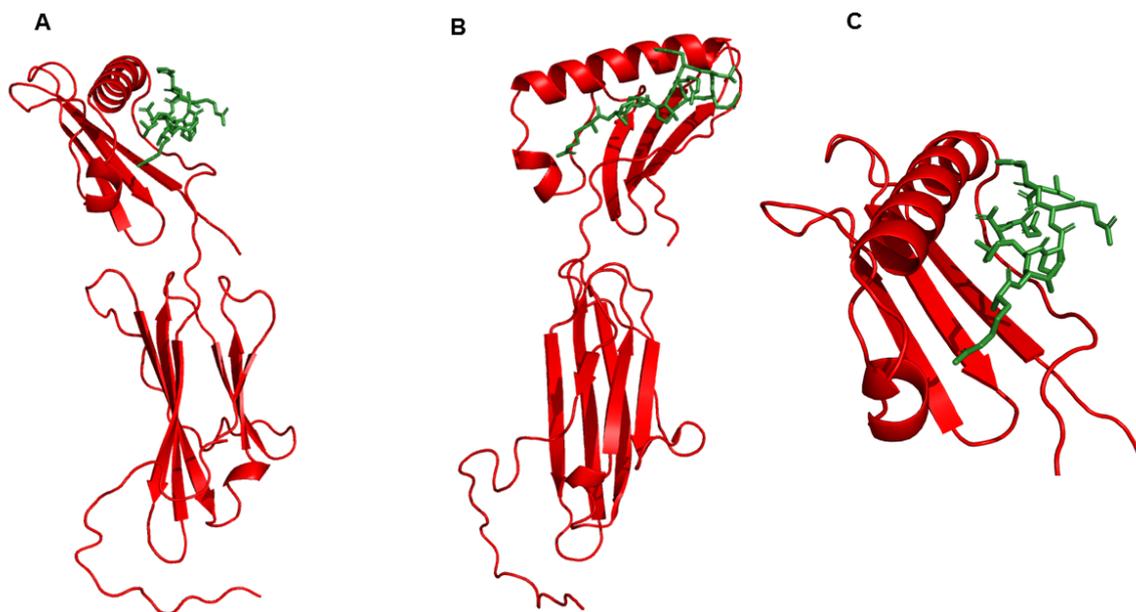
Para examinar o comportamento dinâmico da estrutura modelada HLA-G6 contendo o peptídeo acoplado ao domínio  $\alpha 1$ , foi realizado o *docking* do nonapeptídeo RIIPRHLQL ao domínio  $\alpha 1$  do monômero da isoforma HLA-G6 demonstrada anteriormente, utilizando o *software* Autodock Vina (TROTT *et al.*, 2010).

### 6.7.1 Seleção do *docking*

A estrutura do nonapeptídeo RIIPRHLQL demonstrada anteriormente foi utilizada para o ensaio *in silico* de *docking* ao domínio  $\alpha 1$  do monômero da isoforma HLA-G6, denominada receptor para esse experimento de *docking*.

Os experimentos *in silico* de *docking* foram executados como descrito anteriormente.

A melhor conformação do peptídeo RIIPRHLQL após o *docking* com a isoforma HLA-G6 apresentou energia de ligação de -6,8 kcal/mol.



**Figura 74.** Melhor modelo obtido da isoforma HLA-G6 (vermelho) com o nonapeptídeo RIIPRHLQL (verde) acoplado ao domínio  $\alpha 1$ . Visão frontal (A), lateral (B) e aproximação do domínio  $\alpha 1$  (C).

### 6.7.2 Dinâmica Molecular

A dinâmica molecular da isoforma HLA-G6 contendo o peptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  consistiu em triplicata (r1, r2, r3) das simulações para o modelo gerado. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

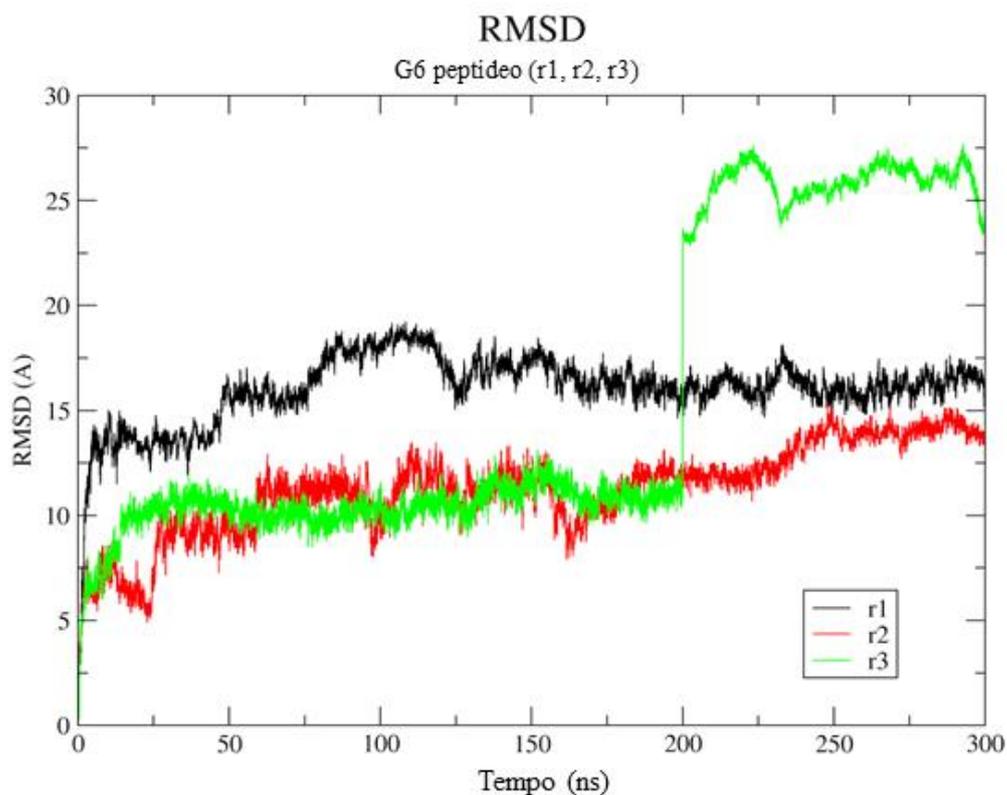
O sistema para simulação de dinâmica molecular, além da isoforma HLA-G6 e o peptídeo RIIPRHLQL contém:

- 125826 moléculas de água,
- 244 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

As simulações de dinâmica molecular para o sistema que contém a isoforma HLA-G6 contendo o peptídeo RIIPRHLQL localizado no domínio  $\alpha 1$  atingiram o tempo total de 300ns.

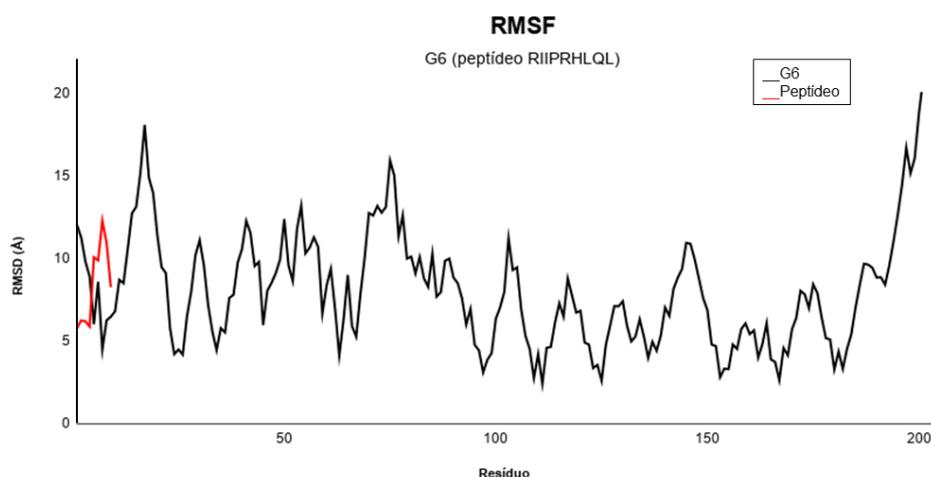
O gráfico de RMSD, apresentado na **Figura 75**, demonstra as triplicatas da dinâmica molecular da isoforma HLA-G6 contendo o peptídeo RIIPRHLQL localizado no domínio  $\alpha 1$ . Podemos observar que a replicata r3 da isoforma HLA-G6 com peptídeo acoplado ao sulco apresenta valores máximos de RMSD de 27,7 Å, enquanto as outras 2 simulações se mantêm no intervalo de 10 a 16 Å.

A replicata r3 apresenta um aumento brusco nos valores de RMSD em torno de 200ns, momento em que é possível perceber o escape do peptídeo que estava acoplado ao domínio  $\alpha 1$ , o mesmo ficando livre durante o restante dessa simulação.



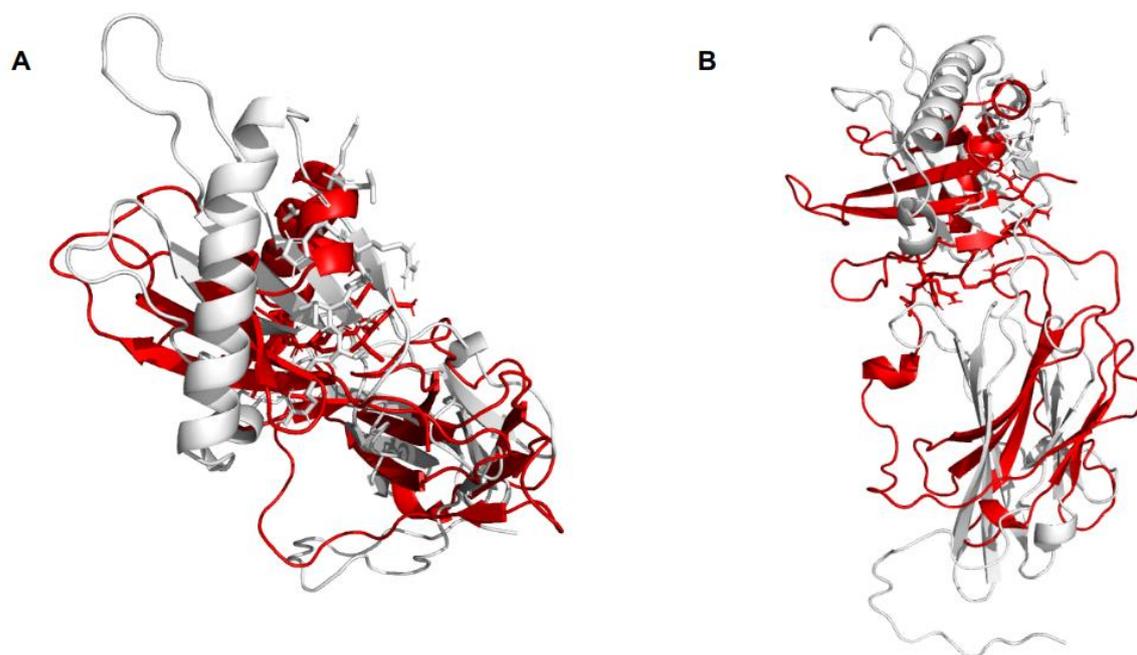
**Figura 75.** Gráfico de RMSD da dinâmica molecular de 300ns referente à isoforma HLA-G6 contendo peptídeo acoplado ao domínio  $\alpha 1$ .

O gráfico apresentando o melhor resultado de RMSF para a simulação da isoforma HLA-G6 contendo o peptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  demonstrou flutuação disseminada na isoforma HLA-G6 e peptídeo acoplado.



**Figura 76.** Variação estrutural por resíduo da simulação r1 da simulação de dinâmica molecular referente à isoforma HLA-G6 que possui o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$ . Valores de RMSF das coordenadas dos átomos  $C\alpha$  em relação à estrutura inicial.

A **Figura 77** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 300ns da dinâmica da isoforma HLA-G contendo o peptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$ . Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 14,5 Å.



**Figura 77.** Isoforma HLA-G6 contendo peptídeo acoplado ao domínio  $\alpha 1$  momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular. **(A)** Visão superior da isoforma HLA-G6, mostrando em primeiro plano o domínio  $\alpha 1$  e o peptídeo. **(B)** Visão frontal da isoforma HLA-G6, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.

## 6.8 Isoforma solúvel HLA-G6 contendo peptídeo acoplado ao domínio $\alpha 1$ e acoplada à $\beta 2$ -microglobulina

Para verificar a possibilidade da existência da isoforma HLA-G6 contendo peptídeo acoplado ao domínio  $\alpha 1$  e acoplada à  $\beta 2$ -microglobulina, foi realizada modelagem molecular e estudo da sua estabilidade dinâmica.

### 6.8.1 *Docking* da proteína $\beta 2$ -microglobulina à isoforma HLA-G6

A estrutura utilizada como molde para a modelagem por homologia foi a isoforma HLA-G6 com o nonapeptídeo RIIPRHLQL ao domínio  $\alpha 1$ , anteriormente demonstrada.

Para obter a modelagem por homologia da isoforma HLA-G6 com o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  e acoplada à proteína  $\beta 2$ -microglobulina, foi utilizado o

*software* Modeller 9.15 (WEBB *et al.*, 2002), especificamente a função de modelagem de múltiplas cadeias (*multiple chains*). Não foi possível obter resultados com essa metodologia, pois o *software* Modeller não foi capaz de fazer o encaixe da proteína  $\beta$ 2-microglobulina na ausência do domínio  $\alpha$ 2.

Para tentar contornar esse problema, foi realizado um experimento *in silico* de *docking* utilizando o servidor *online* HADDOCK (*High Ambiguity Driven protein-protein DOCKing*) (van ZUNDERT *et al.*, 2015). O servidor HADDOCK é uma abordagem de *docking* flexível baseada em informações já conhecidas para formação complexos biomoleculares. HADDOCK distingue-se dos métodos de *docking ab initio* no fato de que utiliza informações de interfaces de proteínas identificadas ou previstas para direcionar o processo de *docking*, podendo lidar com uma grande variedade de tipos de *docking*, incluindo proteína-proteína, proteína-ácidos nucleicos e complexos proteína-ligante. A avaliação dos resultados é baseada no HADDOCK *score* e *z-score*. Quanto menores os valores obtidos para o *z-score*, melhor o resultado do *docking*.

A obtenção das posições referentes aos resíduos envolvidos na interação entre a proteína  $\beta$ 2-microglobulina, domínio  $\alpha$ 1 e  $\alpha$ 3 foi realizada utilizando a interação da isoforma HLA-G5 e a proteína  $\beta$ 2-microglobulina, para definição dos resíduos envolvidos na interação, com o uso do pacote BioPython (COCK *et al.*, 2009). Os seguintes resíduos estão envolvidos na interação HLA-G5 e  $\beta$ 2-microglobulina: 3, 8, 10, 12, 21, 23, 25, 27, 32, 35, 48, 87, 91, 92, 93, 94, 96, 97, 98, 115, 117, 119, 120, 121, 122, 188, 192, 194, 200, 202, 206, 229, 230, 231, 232, 233, 234, 235, 236, 242 e 243.

Os resíduos envolvidos na interação serviram como informação prévia utilizada no arquivo de entrada para o direcionamento do *docking* da proteína  $\beta$ 2-microglobulina ao domínio  $\alpha$ 1 e  $\alpha$ 3 da isoforma HLA-G6 no servidor *online* HADDOCK.

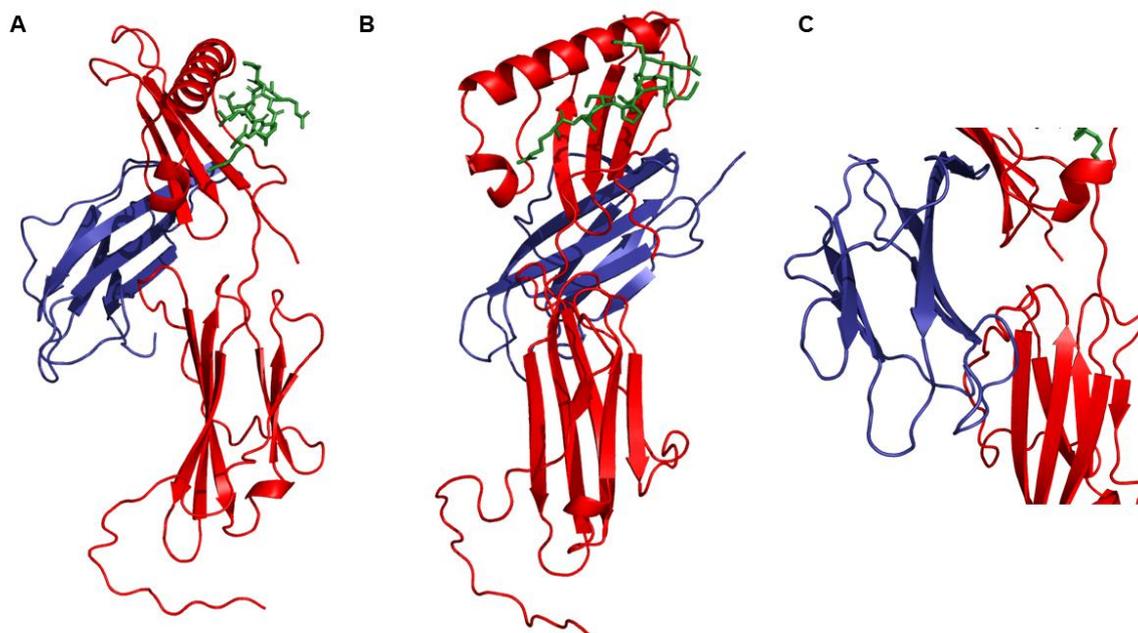
>HLA-G6  
 GSHSMRYFSAAVSRPGRGEPRFIAMGYVDDTQFVRFDSDSACPRMEPRAPWVEQEGPEYWEEETRNTK  
 AHAQTDRMNLQTLRGYYNQSEADPPKTHVTHHPVFDYEATLRCWALGFYPAEIIITWQRDGEDQTQDVEL  
VETRPAGDGTFQKWAAVVVPSGEEQRYTCHVQHEGLPEPLMLRWSKEGDGGIMSVRESRSLSEDL



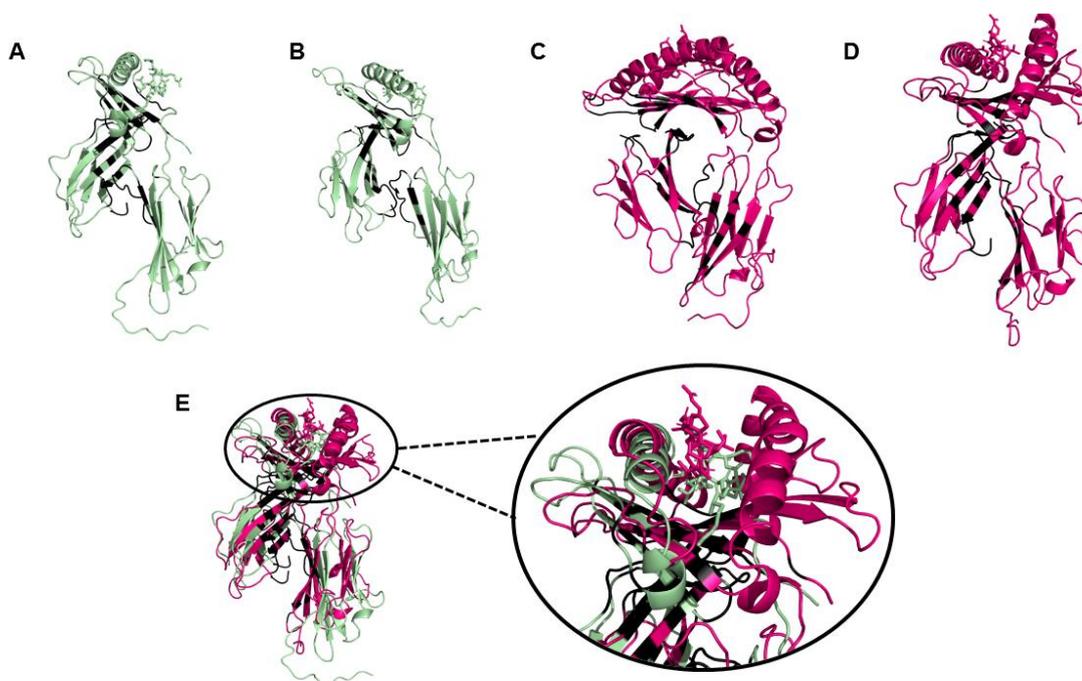
>β2-microglobulina  
 MIQRTPKIQYSRHPAEENGKSNFLNCYVSGFHPSDIEVDLLKNGERIEKVEHSDLSFSKDWSFYLLYYTEFT  
 PTEKDEYA CRVNHVTLSPKIVKWDRDM

**Figura 78.** Resíduos em vermelho serviram para o direcionamento do *docking* da proteína β2-microglobulina aos domínios α1 e α3 da isoforma HLA-G6 no servidor *online* HADDOCK.

O melhor resultado proveniente do *docking* da proteína β2-microglobulina ao domínio α1 e α3 da isoforma HLA-G6 no servidor *online* HADDOCK foi uma estrutura presente no cluster 2, que apresentou *z-score* de -1.74. Além disso, tal estrutura obteve o score de qualidade 0.854 na ferramenta ModFold, além de  $p=6.89 \times 10^{-5}$ . A interação HLA-G6 e β2-microglobulina após o *docking* envolve os resíduos: 4, 5, 6, 8, 9, 10, 21, 23, 25, 27, 29, 30, 32, 35, 46, 48, 79, 80, 81, 82, 83, 119, 139, 140, 141, 142, 143, 144, 145, 146, 147 e 150.



**Figura 79.** Melhor modelo obtido para isoforma HLA-G6 (vermelho) contendo nonapeptídeo RIIPRHLQL (verde) e acoplada à proteína  $\beta$ 2-microglobulina (azul). Visão frontal (A), lateral (B) e aproximação da cadeia composta pela  $\beta$ 2-microglobulina (C).



**Figura 80.** Demonstração da localização dos resíduos envolvidos na interação HLA-G5 e  $\beta$ 2-microglobulina e HLA-G6 e  $\beta$ 2-microglobulina. (A). Visualização frontal das regiões de interação entre HLA-G6 e  $\beta$ 2-

microglobulina. **(B)** Visualização lateral das regiões de interação entre HLA-G6 e  $\beta$ 2-microglobulina. **(C)** Visualização lateral das regiões de interação entre HLA-G5 e  $\beta$ 2-microglobulina. **(D)** Visualização frontal das regiões de interação entre HLA-G5 e  $\beta$ 2-microglobulina. **(E)** Sobreposição das regiões de interação entre HLA-G5 e  $\beta$ 2-microglobulina (rosa) e HLA-G6 e  $\beta$ 2-microglobulina (verde).

## 6.8.2 Dinâmica Molecular

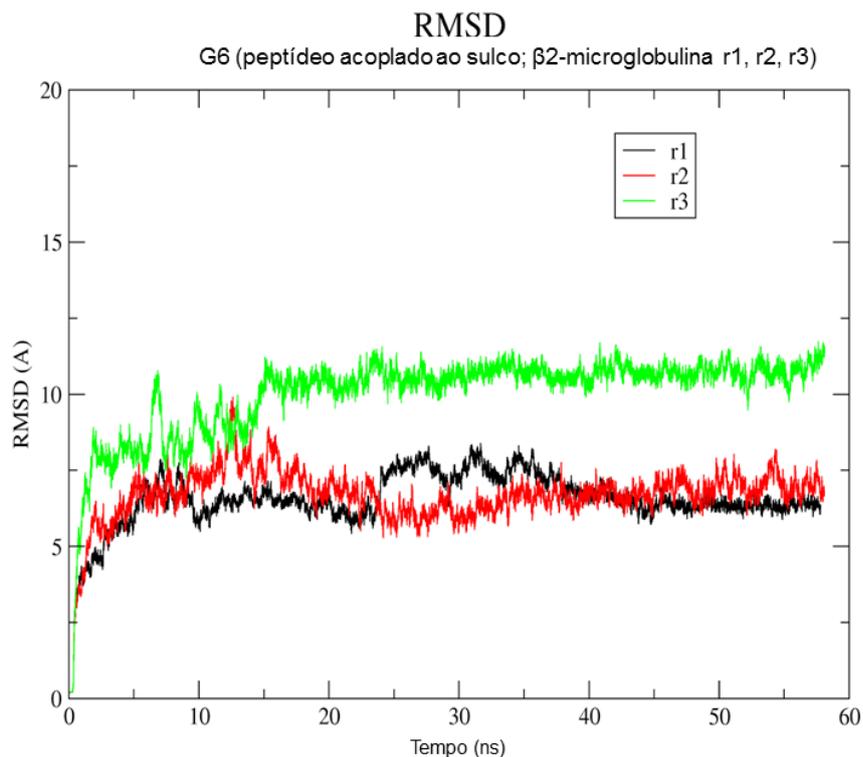
A dinâmica molecular para isoforma HLA-G6 com o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha$ 1 e acoplada à proteína  $\beta$ 2-microglobulina consistiu em triplicata (r1, r2, r3) das simulações para o modelo gerado, utilizando o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

O sistema para simulação de dinâmica molecular, além da isoforma HLA-G6 com o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha$ 1 e acoplada à proteína  $\beta$ 2-microglobulina contém:

- 125617 moléculas de água,
- 240 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

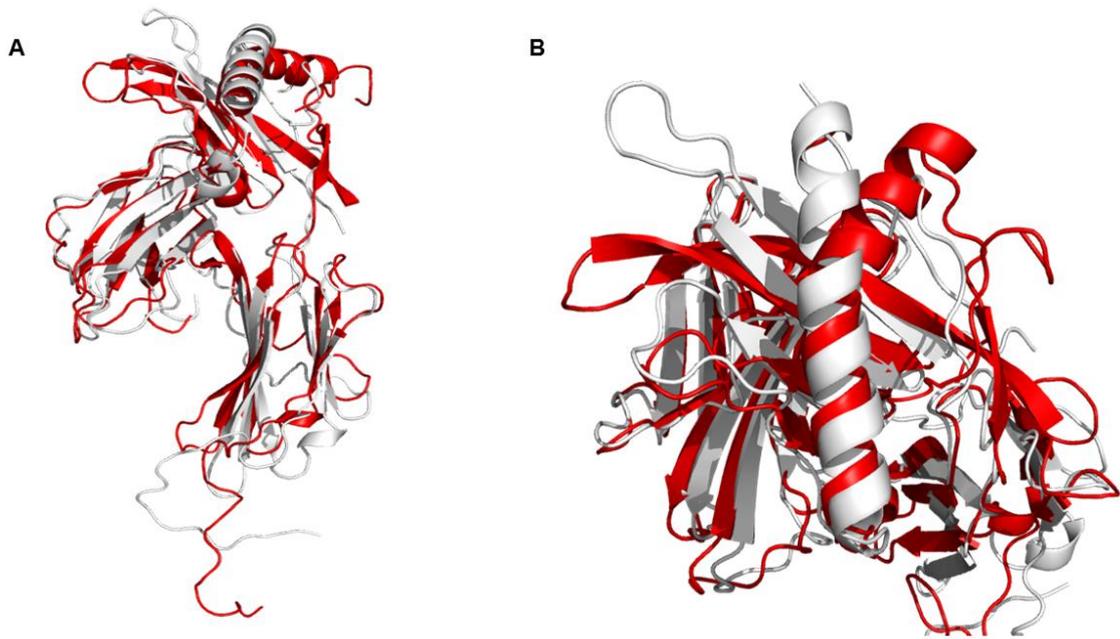
As simulações de dinâmica molecular para o sistema que contém a HLA-G6 com o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha$ 1 e acoplada à proteína  $\beta$ 2-microglobulina atingiram o tempo total de aproximadamente 60ns. As simulações foram terminadas abruptamente perto do 60ns, pois houve um aviso do *software* GROMACS 4.6.5 de que o sistema se encontrava em desequilíbrio extremo e não seria possível dar continuidade à simulação.

O gráfico de RMSD, apresentado na **Figura 81**, demonstra as triplicatas da dinâmica molecular HLA-G6 com o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha$ 1 e acoplada à proteína  $\beta$ 2-microglobulina. Podemos observar que as triplicatas da isoforma HLA-G6 com peptídeo acoplado ao sulco apresentam valores máximos de RMSD de 11,73 Å. As replicatas r1 e r2 foram mantidas na faixa de 6 a 7 Å e a replicata r3, 10 Å.



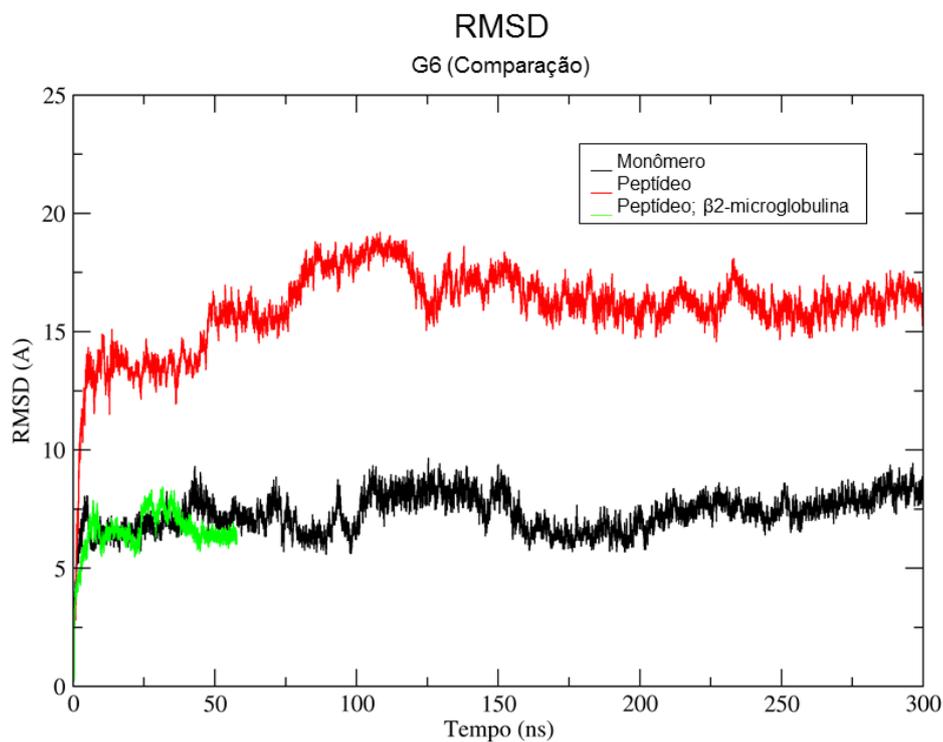
**Figura 81.** Gráfico de RMSD da dinâmica molecular de 60ns referente à isoforma HLA-G6 contendo nonapeptídeo RIIPRHLQL e acoplada à proteína β2-microglobulina.

A **Figura 82** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 60ns da dinâmica da isoforma HLA-G6 contendo nonapeptídeo RIIPRHLQL e acoplada à proteína β2-microglobulina. Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 4,05 Å.



**Figura 82.** Isoforma HLA-G6 contendo nonapeptídeo RIIPRHLQL e acoplada à proteína  $\beta$ 2-microglobulina nos momentos iniciais (em branco) e finais (em vermelho) dos 60ns da simulação de dinâmica molecular. **(A)** Visão superior da isoforma HLA-G6, mostrando em primeiro plano o domínio  $\alpha$ 1 e o peptídeo. **(B)** Visão frontal da isoforma HLA-G6, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.

Para possibilitar a comparação entre os valores de RMSD das simulações envolvendo a isoforma HLA-G6 em suas 3 alternativas estruturais (HLA-G6 – monômero; HLA-G6 e nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha$ 1; HLA-G6 com o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha$ 1 e acoplada à proteína  $\beta$ 2-microglobulina) foi gerado um gráfico comparativo (**Figura 83**).

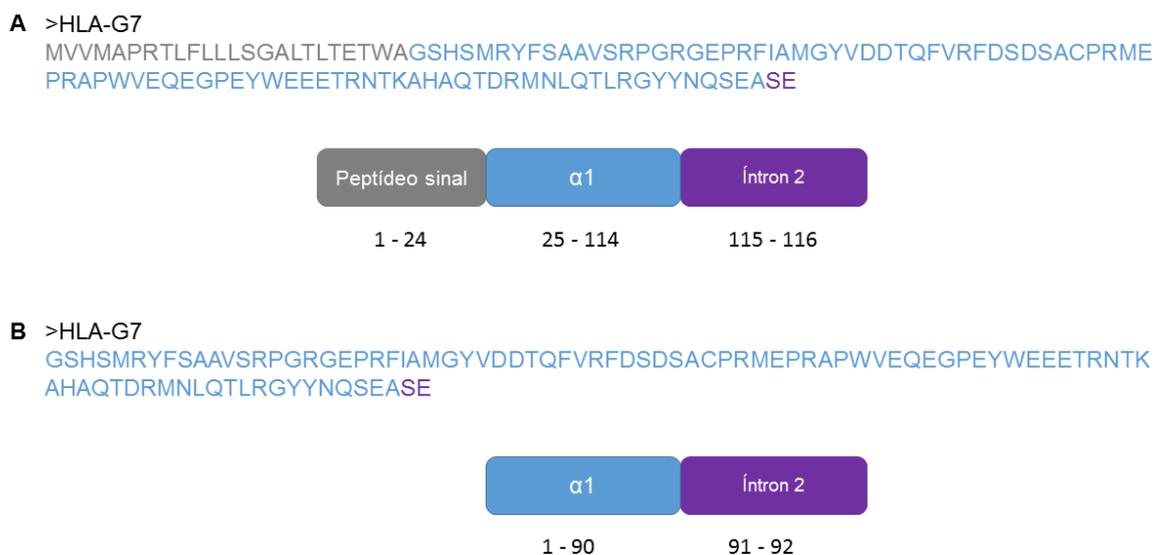


**Figura 83.** Gráfico de RMSD comparativo das simulações envolvendo a isoforma HLA-G6 em suas 3 alternativas estruturais (HLA-G6 – monômero; HLA-G6 e nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$ ; HLA-G6 com o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  e acoplada à proteína  $\beta 2$ -microglobulina).

## 6.9 Isoforma solúvel HLA-G7

### 6.9.1 Obtenção da sequência

A montagem da sequência primária da isoforma HLA-G7 foi realizada da mesma forma descrita para as isoformas HLA-G5 e HLA-G6. Podemos observar essas informações citadas acima de forma esquemática na **Figura 84**.



**Figura 84.** (A) Sequência primária da isoforma HLA-G7 no formato FASTA. As cores são referentes à porção da proteína compostas por tais aminoácidos, segundo esquema na porção inferior da figura (cinza – peptídeo sinal, azul – domínio  $\alpha 1$ , roxo – 2 resíduos provenientes da tradução do íntron 2). (B) Sequência primária da isoforma madura HLA-G7 no formato FASTA.

Semelhante ao que ocorre com a proteína HLA-G, a isoforma HLA-G7 após a identificação da região relacionada ao peptídeo sinal e sua remoção, possui 92 aminoácidos  $\alpha 1$  (1 – 90) e porção proveniente da tradução do íntron 2 (91 - 92).

Por possuir 1 domínio igual à proteína HLA-G completa, foi utilizada estratégia de modelagem por homologia nessa porção, enquanto a região proveniente do íntron 2 foi modelada por metodologia *ab initio*.

### 6.9.2 Modelagem por homologia do monômero da isoforma HLA-G7

Para modelagem por homologia foi utilizado o mesmo molde utilizado na modelagem por homologia da proteína HLA-G completa, a cristalografia PDB\_ID:1YDP (CLEMENTS *et al.*, 2005).

Na **Figura 85** abaixo pode-se observar o alinhamento múltiplo entre a sequência primária completa de HLA-G (denominada HLA-G\_completo), a sequência primária da isoforma HLA-G7 e a sequência primária do molde PDB\_ID:1YDP.

```

1YDP          - SHSMRYFSAAVSRPGRGEPFIAMGYVDDTQFVRFSDSASPRMEPRAPWVEQEGPEYW
HLA-G_completo G SHSMRYFSAAVSRPGRGEPFIAMGYVDDTQFVRFSDSACPRMEPRAPWVEQEGPEYW
HLA-G7        G SHSMRYFSAAVSRPGRGEPFIAMGYVDDTQFVRFSDSACPRMEPRAPWVEQEGPEYW
*****

1YDP          EEETRNTKAHAQTRMNLQTLRGYYNQSEASSHTLQWMIGCDLGS DGRLIRGYERYAYDG
HLA-G_completo EEETRNTKAHAQTRMNLQTLRGYYNQSEASSHTLQWMIGCDLGS DGRLIRGYERYAYDG
HLA-G7        EEETRNTKAHAQTRMNLQTLRGYYNQSEASE-----
*****

1YDP          KDYLALNEDLRSWTAADTAAQISKRKCEANVAEQRRAYLEGT CVEWLHRYLENGKEMLQ
HLA-G_completo KDYLALNEDLRSWTAADTAAQISKRKCEANVAEQRRAYLEGT CVEWLHRYLENGKEMLQ
HLA-G7        -----

1YDP          RADPPKTHVTHHPVFDYEATLRCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGT
HLA-G_completo RADPPKTHVTHHPVFDYEATLRCWALGFYPAEIIITWQRDGEDQTQDVELVETRPAGDGT
HLA-G7        -----

1YDP          FQKWAAVVPSGEEQRYTCHVQHEGLPEPLMLRWKQ-----
HLA-G_completo FQKWAAVVPSGEEQRYTCHVQHEGLPEPLMLRWKQSSSLPTIPIMGVAGLVVLAAVVTG
HLA-G7        -----

1YDP          -----
HLA-G_completo AAVAAVLWRKKSSD
HLA-G7        -----

```

**Figura 85.** Alinhamento múltiplo das sequências primárias da proteína completa HLA-G (denominada HLA-G\_completo), a sequência primária da isoforma HLA-G7 (denominada G7) e a sequência primária do molde PDB\_ID:1YDP (CLEMENTS *et al.*, 2005). Os resíduos em vermelho e resíduos alinhados foram modelados por homologia, enquanto os resíduos em azul foram modelados por metodologia *ab initio*.

### 6.9.2.1 Avaliação dos modelos gerados por homologia

Foram gerados 05 modelos para a isoforma HLA-G7. Cada um dos cinco modelos gerados para a isoforma HLA-G7 foi avaliado pelos *softwares* PROCHECK (LASKOWSKI *et al.*, 1993), Verify 3D (LUTHY *et al.*, 1992), ERRAT (COLOVOS e YATES, 1993), QMEAN (BENKERT *et al.*, 2008) e ProSA-Web (WIEDERSTEIN *et al.*, 2007), para que o melhor fosse selecionado.

Dentre os modelos gerados, o modelo 1 atingiu as seguintes pontuações, sendo escolhido como melhor modelo gerado:

- Verify 3D: 95.65%.

- ERRAT: 90.47%.
- ProSA-Web: *z-score* de -7.21.
- QMEAN: 0.714.

Esse modelo será utilizado para as análises e modelagens subsequentes.

### 6.9.3 Modelagem *ab initio*

Novamente, foi utilizada a metodologia de modelagem *ab initio*, de forma semelhante ao protocolo já exposto em seções anteriores, utilizando as ferramentas Robetta e Rosetta (PETTERSEN *et al.*, 2004). Os dois resíduos proveniente do íntron 2 foram modelados por metodologia *ab initio* em 1 etapa de modelagem, abrangendo os resíduos 91 e 92.

#### 6.9.3.1 Seleção dos modelos

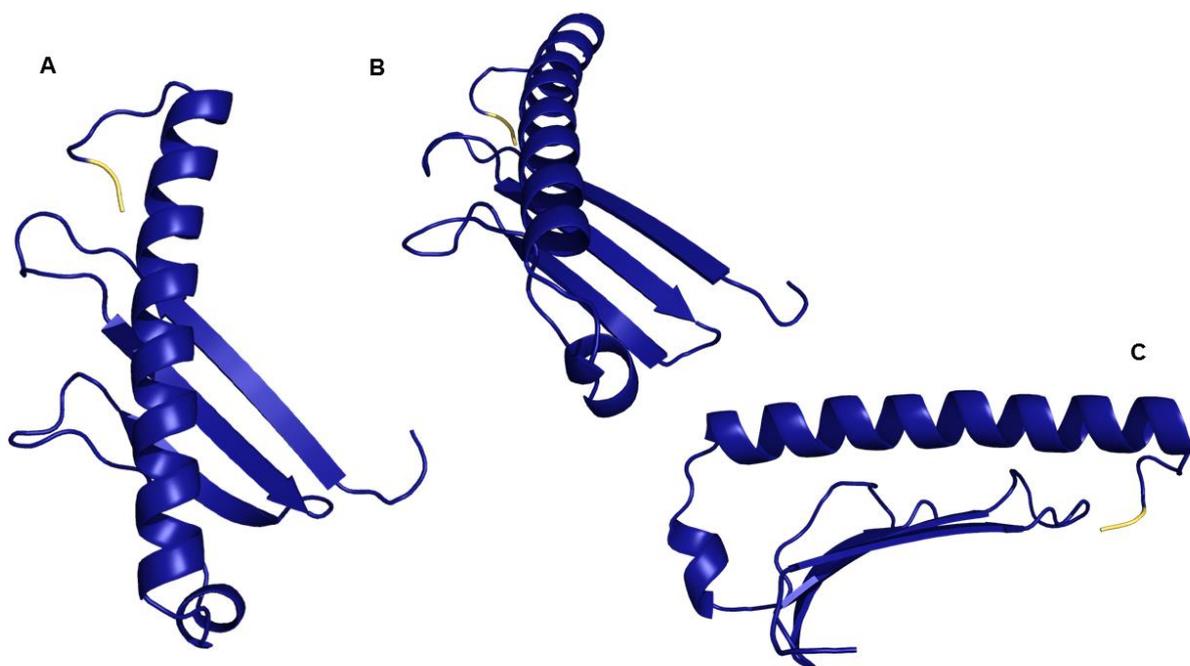
A escolha do melhor modelo foi baseada nos melhores resultados das ferramentas QMEAN (BENKERT *et al.*, 2008) e MODFOLD (MCGUFFIN *et al.*, 2008).

O melhor modelo gerado foi selecionado e o *score* utilizados para escolha do melhor modelo encontram-se na **Tabela 4**.

**Tabela 4.** Melhores modelos provenientes da metodologia de modelagem *ab initio* para a isoforma HLA-G7.

#### HLA-G7: *ab initio*

Etapa	Resíduos	Estrutura	ModFold (p)	ModFold (score)	QMEAN
1	91 - 92	723	$5.97 \times 10^{-8}$	0.802	0.89



**Figura 86.** Modelo obtido para isoforma HLA-G7, gerado pela combinação de modelagem por homologia e *ab initio*. Visão superior (A), frontal (B) e lateral (C) da molécula. A porção em amarelo corresponde aos resíduos modelados utilizando metodologia *ab initio*.

#### 6.9.4 Dinâmica molecular

A dinâmica molecular consistiu em triplicata (r1, r2, r3) das simulações para o modelo gerado para o monômero da isoforma HLA-G7. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

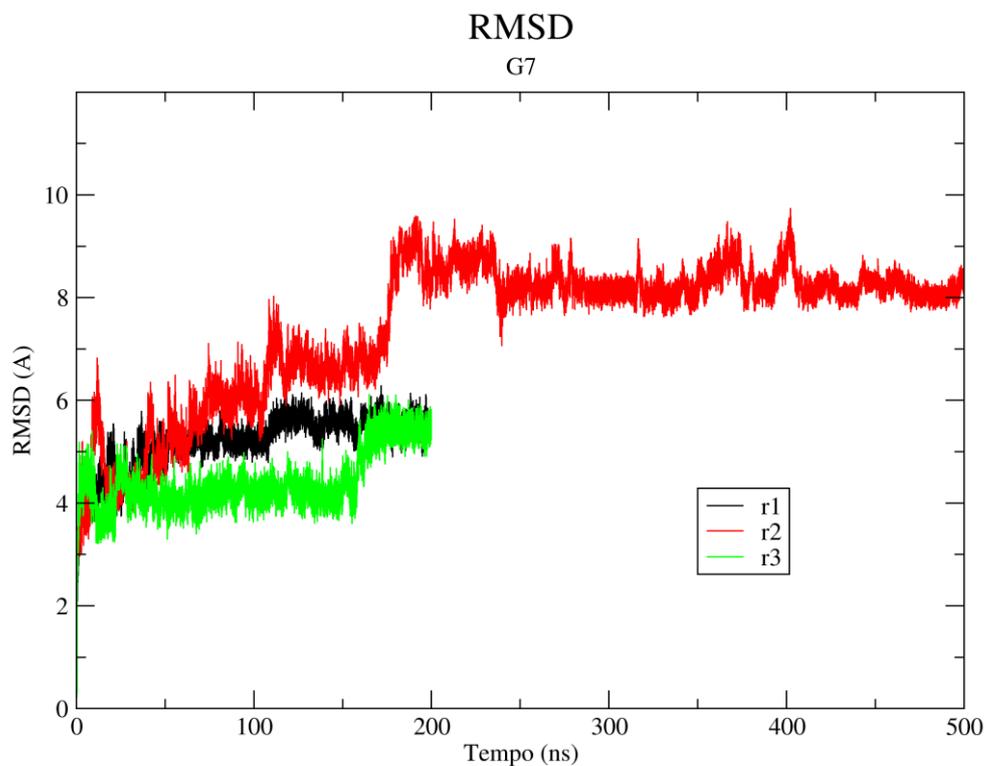
O sistema para simulação de dinâmica molecular, além da isoforma HLA-G7 contém:

- 32628 moléculas de água,
- 66 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

As simulações de dinâmica molecular para o sistema que contém a isoforma HLA-G7 atingiram o tempo total de 500ns.

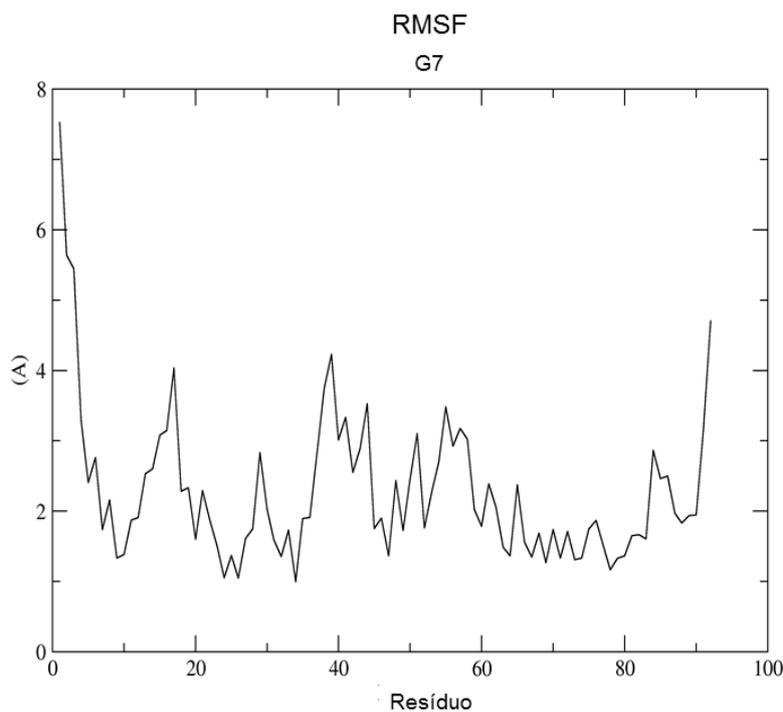
O gráfico de RMSD, apresentado na **Figura 87**, demonstra as triplicatas da dinâmica molecular da isoforma HLA-G7. Podemos observar que as triplicatas da isoforma HLA-G7

apresentam valores de RMSD que chegam a 9,75 Å, mas as replicatas r1 e r3 mantêm-se na faixa de 4 a 5,5 Å.



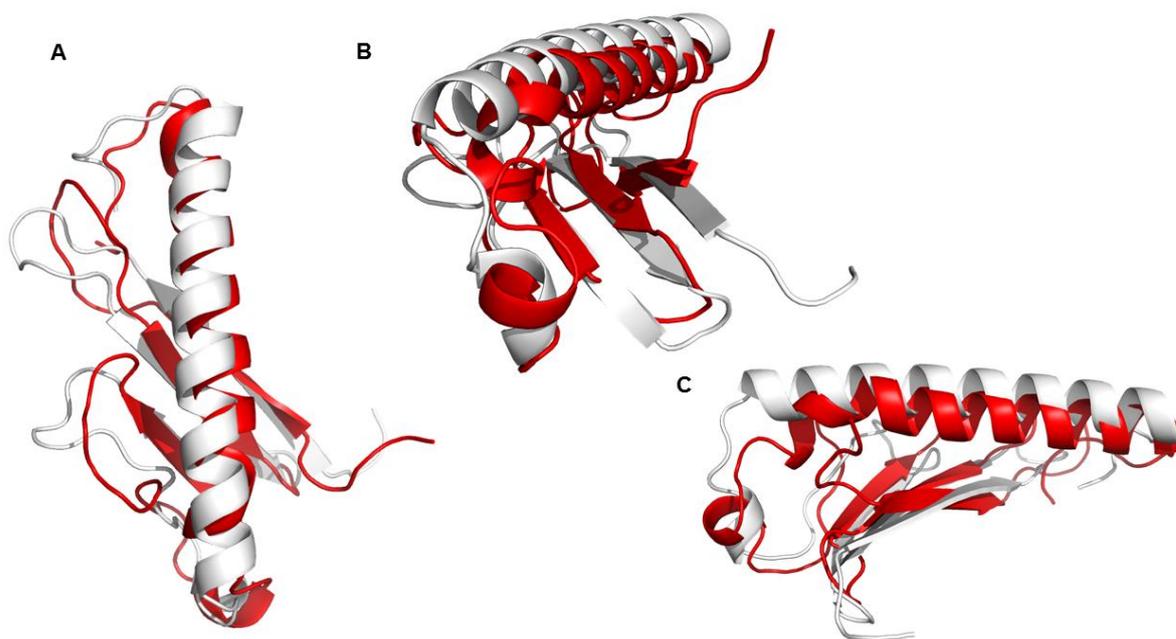
**Figura 87.** Gráfico de RMSD da isoforma HLA-G7.

O gráfico apresentando o resultado de RMSF (**Figura 88**) para a simulação mais estável do monômero da isoforma HLA-G7 demonstrou que a flutuação está distribuída por toda a estrutura, mas atinge os maiores níveis na porção inicial da molécula (resíduos 1 a 10).



**Figura 88.** Variação estrutural por resíduo da simulação r1 da isoforma HLA-G7. Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.

A **Figura 89** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 200ns da dinâmica do monômero da isoforma HLA-G7. Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 1,054 Å.



**Figura 89.** Isoforma HLA-G7 nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular. **(A)** Visualização superior da isoforma HLA-G7, mostrando em primeiro plano o domínio  $\alpha 1$ . **(B)** Visualização frontal da isoforma HLA-G7. **(C)** Visualização lateral da isoforma HLA-G7 demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.

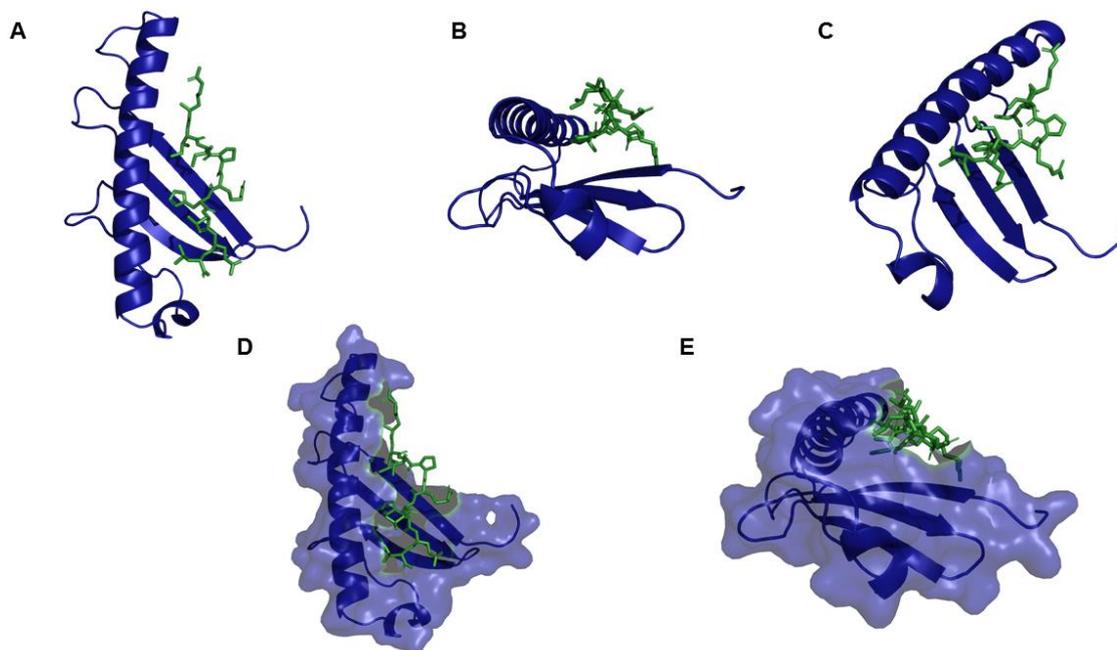
## 6.10 Isoforma solúvel HLA-G7 contendo peptídeo acoplado ao domínio $\alpha 1$

Para examinar o comportamento dinâmico da estrutura modelada HLA-G7 contendo o peptídeo acoplado ao domínio  $\alpha 1$ , foi realizado o *docking* do nonapeptídeo RIIPRHLQL ao monômero da isoforma HLA-G7 demonstrada anteriormente, utilizando o *software* Autodock Vina (TROTT *et al.*, 2010).

### 6.10.1 Seleção do *docking*

A estrutura do nonapeptídeo RIIPRHLQL demonstrada anteriormente foi utilizada para o ensaio *in silico* de *docking* no sulco da isoforma HLA-G7, denominada receptor para esse experimento de *docking*. O experimento *in silico* de *docking* foi executado como descrito anteriormente.

A melhor conformação do peptídeo RIIPRHLQL após o *docking* com a isoforma HLA-G7 apresentou energia de ligação de -9,2 kcal/mol.



**Figura 90.** Visualização do melhor modelo obtido para HLA-G7 (azul) contendo peptídeo (verde) acoplado ao domínio  $\alpha 1$ . Representação em formato *cartoon*; visualização superiora (A), frontal (B) e lateral (C) da molécula. Representação de superfície (transparência 60%), possibilitando a visualização da estrutura secundária (em formato *cartoon*) da isoforma HLA-G7 (azul) contendo peptídeo (verde) acoplado ao domínio  $\alpha 1$ ; visualização superiora (D) e frontal (E).

### 6.10.2 Dinâmica Molecular

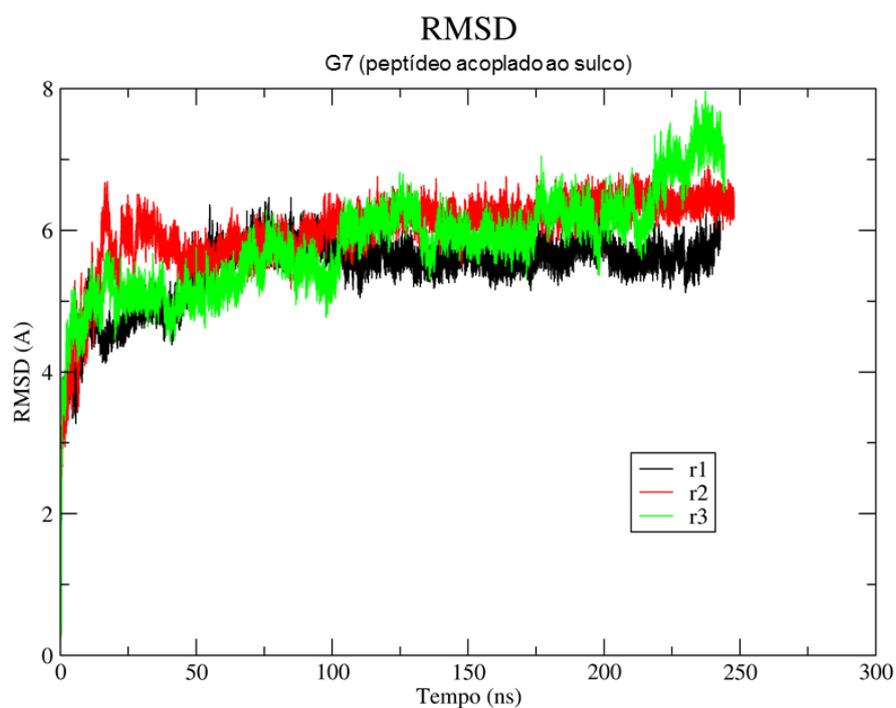
A dinâmica molecular da isoforma HLA-G7 contendo o peptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  foi realizada em triplicata (r1, r2, r3) das simulações para o modelo gerado. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

O sistema para simulação de dinâmica molecular, além da isoforma HLA-G7 e o peptídeo RIIPRHLQL acoplado, contém:

- 32415 moléculas de água,
- 64 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

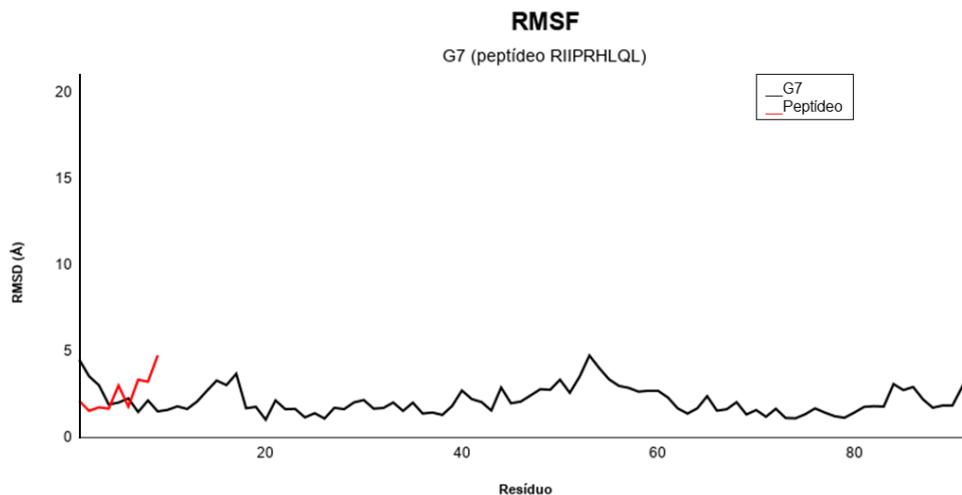
As simulações de dinâmica molecular para o sistema que contém a isoforma HLA-G7 contendo o peptídeo acoplado ao domínio  $\alpha 1$  atingiram o tempo total de 250ns.

O gráfico de RMSD, apresentado na **Figura 91**, demonstra as triplicatas da dinâmica molecular da isoforma HLA-G7 o peptídeo acoplado ao domínio  $\alpha 1$ . Podemos observar que as triplicatas da isoforma HLA-G7 com peptídeo acoplado ao sulco apresentam valores de RMSD que chegam a 7,96 Å, mantendo-se no intervalo de 5 a 6 Å durante a trajetória completa.



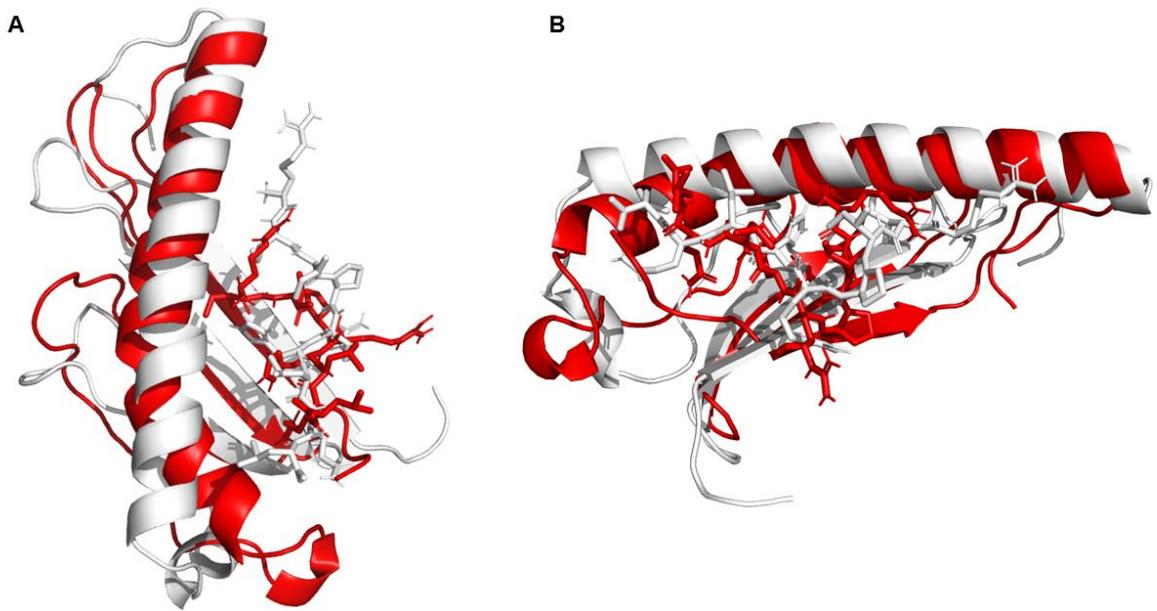
**Figura 91.** Gráfico de RMSD da dinâmica molecular de 250ns referente à isoforma HLA-G7 contendo peptídeo acoplado ao domínio  $\alpha 1$ .

O gráfico apresentando o resultado de RMSF para a simulação mais estável do monômero da isoforma HLA-G7 contendo o peptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  encontra-se na **Figura 92**.



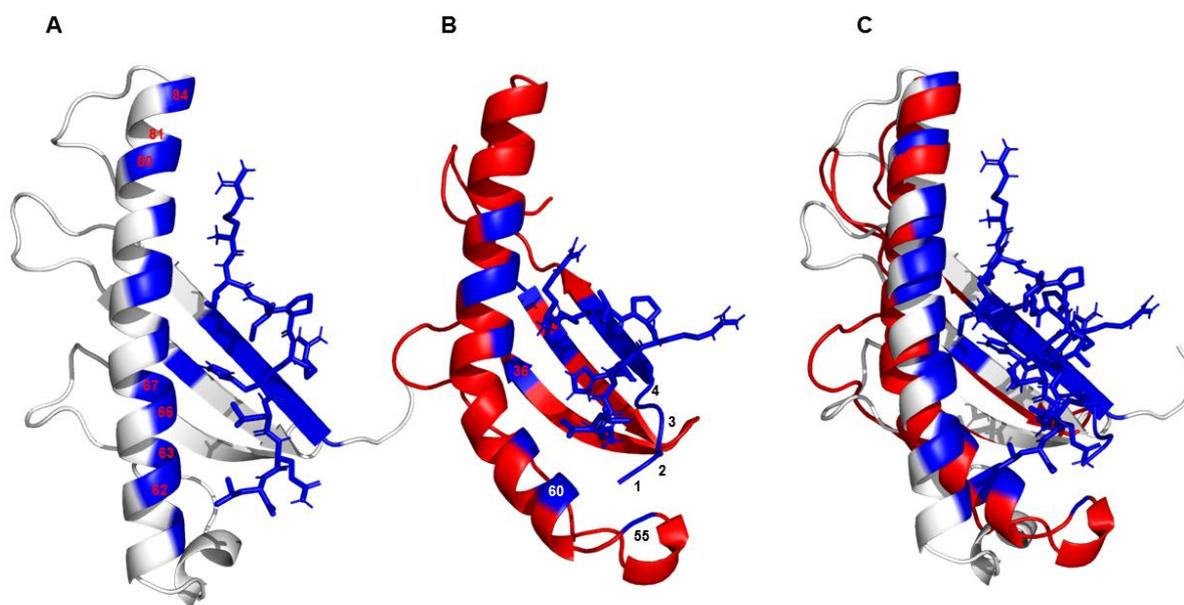
**Figura 92.** Variação estrutural por resíduo da simulação r1 de dinâmica molecular referente à isoforma HLA-G7 que possui o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$ . Valores de RMSF das coordenadas dos átomos  $C\alpha$  em relação à estrutura inicial.

A **Figura 93** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 250ns da dinâmica da isoforma HLA-G7 contendo nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$ . Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 3,02 Å.



**Figura 93.** Isoforma HLA-G7 e nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  nos momentos iniciais (em branco) e finais (em vermelho) dos 250ns da simulação de dinâmica molecular. **(A)** Visão superior da isoforma HLA-G7, mostrando em primeiro plano o domínio  $\alpha 1$  e o peptídeo. **(B)** Visão lateral da isoforma HLA-G7, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular.

O pacote BioPython (COCK *et al.*, 2009) foi utilizado para verificar os resíduos que fazem parte das interações proteína-peptídeo nos momentos iniciais (branco) e finais (vermelho) da simulação referente à isoforma HLA-G7 contendo nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$ , apresentados na **Figura 94**.



**Figura 94.** Isoforma HLA-G7 contendo peptídeo acoplado ao domínio  $\alpha 1$  nos momentos iniciais (A, em branco) e finais (B, em vermelho) dos 250ns da simulação de dinâmica molecular. (A) Visão superior da isoforma HLA-G7, mostrando em primeiro plano o domínio  $\alpha 1$  e o peptídeo, demonstrando as interações entre os resíduos do peptídeo e isoforma HLA-G7 durante os momentos iniciais da simulação de dinâmica molecular. Resíduos numerados são únicos para momentos iniciais. (B) Visão superior da isoforma HLA-G7, mostrando em primeiro plano o domínio  $\alpha 1$  e o peptídeo, demonstrando as interações entre os resíduos do peptídeo e isoforma HLA-G7 durante os momentos finais da simulação de dinâmica molecular. Resíduos numerados são únicos para momentos finais. (C) Comparação das interações nos momentos iniciais (estrutura em branco) e finais (estrutura em vermelho).

Durante os momentos iniciais da simulação de dinâmica molecular os seguintes resíduos interagem com o peptídeo: 5, 6, 7, 8, 9, 22, 24, 59, **62, 63, 66, 67, 70, 73, 74, 77, 80, 81 e 84** (em negrito os resíduos únicos para momentos iniciais). Nos momentos finais da simulação, os seguintes resíduos interagem com o peptídeo: **1, 2, 3, 4, 5, 6, 7, 8, 9, 22, 24, 36, 55, 59, 60, 70, 73, 74 e 77** (em negrito os resíduos únicos para momentos finais).

## 6.11 Dímero da isoforma solúvel HLA-G7 (G7-Cys42)

### 6.11.1 Construção do dímero

Para construção do dímero identificado no trabalho como G7-Cys42, foi realizado um experimento *in silico* de *docking* utilizando o servidor *online* HADDOCK (*High Ambiguity*

*Driven protein-protein DOCKing*) (van ZUNDERT *et al.*, 2015), da mesma forma como exposto para estruturas já apresentadas anteriormente.

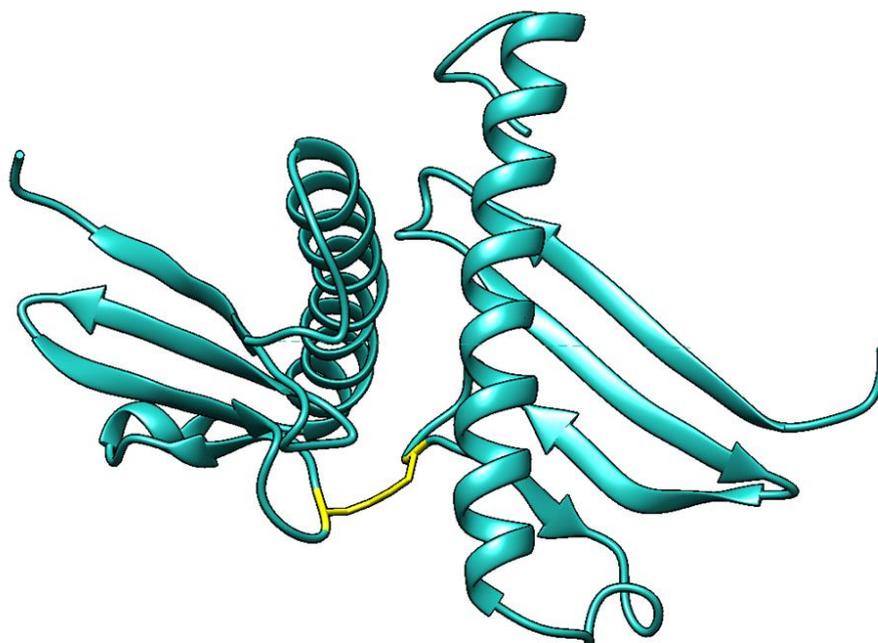
O resíduo de cisteína localizado na posição 42 de cada um dos monômeros da isoforma HLA-G7 que compõem essa proposta de dímero serviram como informação prévia utilizada no arquivo de entrada para o direcionamento do *docking* no servidor *online* HADDOCK.

O melhor resultado proveniente do *docking* para o dímero G7-Cys42 no servidor *online* HADDOCK foi uma estrutura presente no cluster 3, que apresentou *z-score* de -2.0 Além disso, tal estrutura obteve o score de qualidade 0.795 na ferramenta ModFold, além de  $p= 5.15 \times 10^{-5}$ .

A distância entre os dois resíduos de cisteína no melhor modelo foi de 1,993 Å.

### 6.11.2 Dinâmica Molecular

Para formação da ponte dissulfeto entre a estrutura dimérica proveniente do ensaio *in silico* de *docking* (**Figura 95**) foi utilizado o *software* GROMACS (HESS *et al.*, 2008), assim como para verificar a estabilidade da estrutura dimérica proposta por meio de dinâmica molecular.



**Figura 95.** Estrutura do dímero proposto G7-Cys42, após a formação da ponte dissulfeto (amarelo) entre as cisteínas localizadas na posição 42 de cada monômero, posicionados pelo servidor *online* HADDOCK.

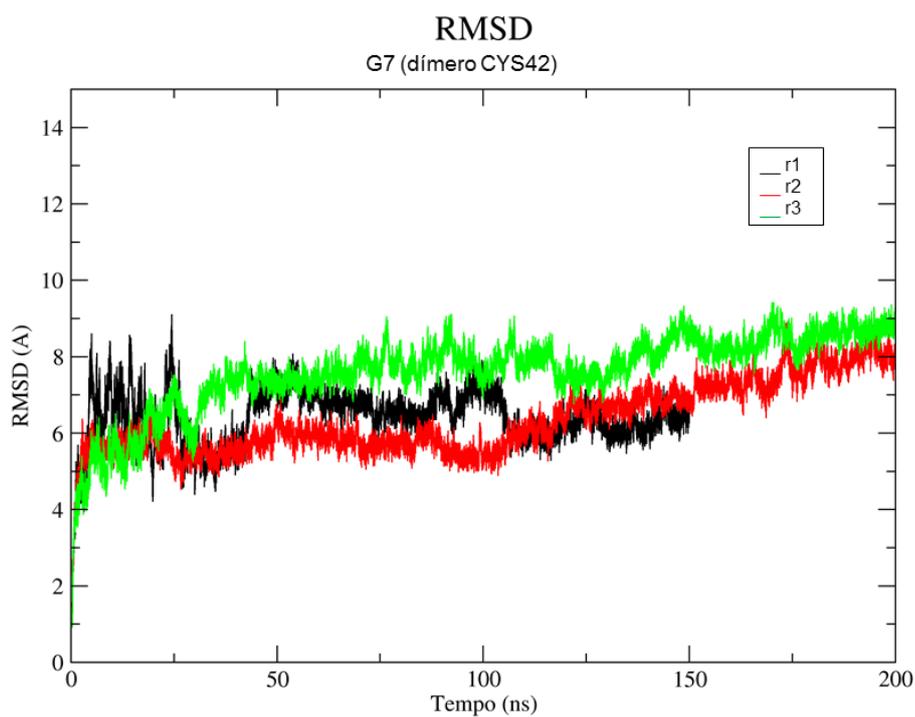
A dinâmica molecular do dímero G7-Cys42 foi realizada em triplicata (r1, r2, r3) para as simulações do modelo gerado. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

O sistema para simulação de dinâmica molecular, além dímero G7-Cys42 contém:

- 68538 moléculas de água,
- 138 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

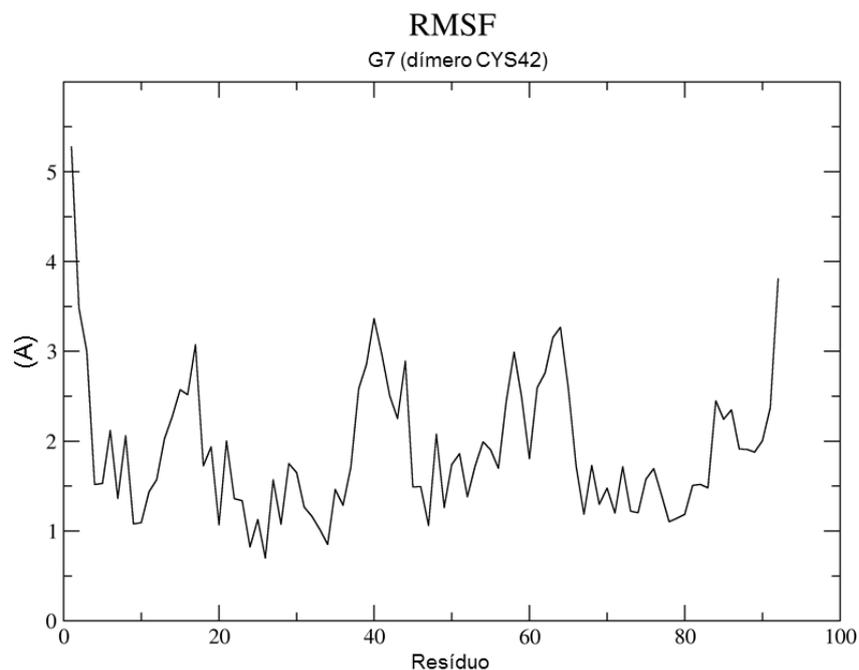
As simulações de dinâmica molecular para o sistema que contém dímero G7-Cys42 atingiram o tempo total de 100ns.

O gráfico de RMSD, apresentado na **Figura 96**, demonstra as triplicatas da dinâmica molecular do dímero G7-Cys42. Podemos observar que as triplicatas da simulação apresentam valores de RMSD que chegam a 9,44 Å, mantendo-se na faixa de 6 a 7,5 durante toda a trajetória.



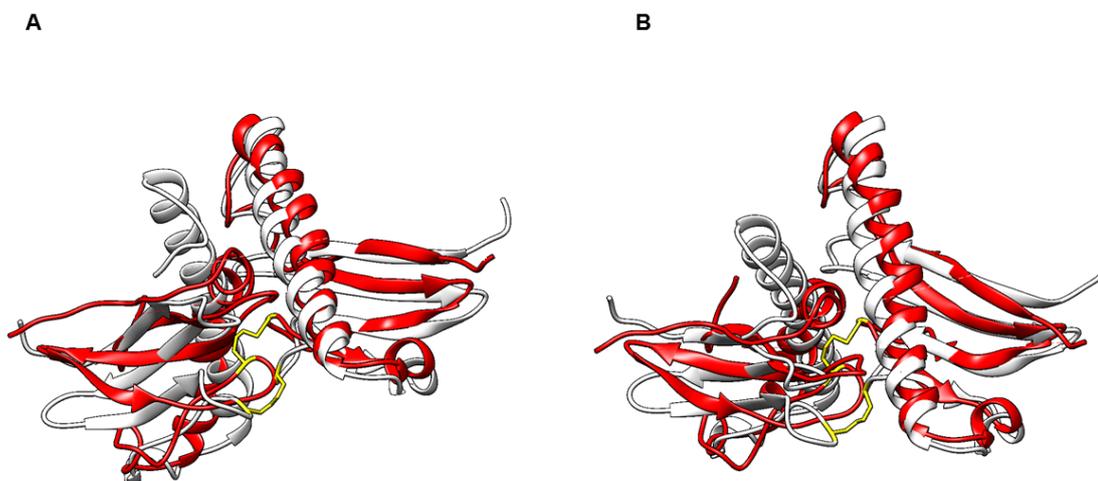
**Figura 96.** Gráfico de RMSD da dinâmica molecular de 200ns referente ao dímero G7-Cys42.

O gráfico apresentando o resultado de RMSF para a simulação mais estável do dímero G7-Cys42. Foi observada flutuação de forma homogênea em toda extensão do dímero, com exceção dos resíduos 1 a 5, que apresentaram flutuação mais acentuada.



**Figura 97.** Variação estrutural por resíduo da simulação r2 de dinâmica molecular referente ao dímero G7-Cys42. Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.

A **Figura 98** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 200ns da dinâmica do dímero G7-Cys42. Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 1,260 Å.



**Figura 98.** Dímero G7-Cys42 nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular (A) e (B). Ponte dissulfeto em amarelo.

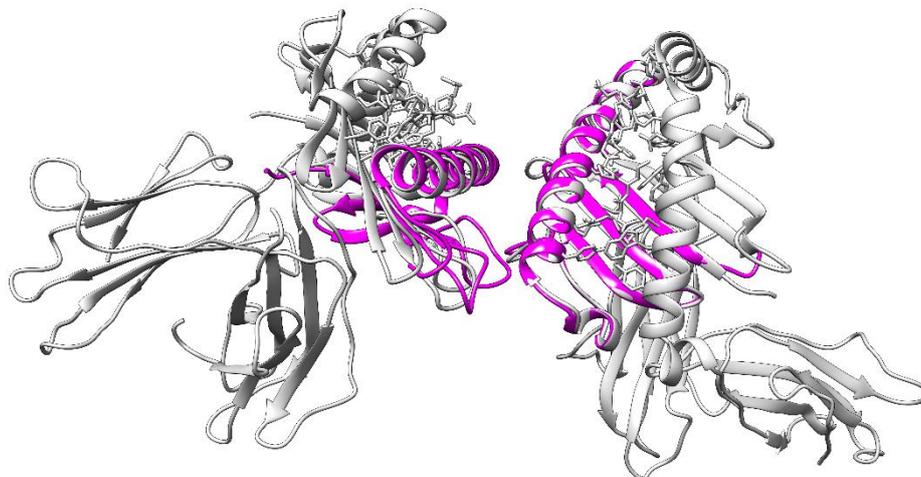
## 6.12 Dímero da isoforma solúvel HLA-G7 (G7 fit 2D31)

### 6.12.1 Construção do dímero

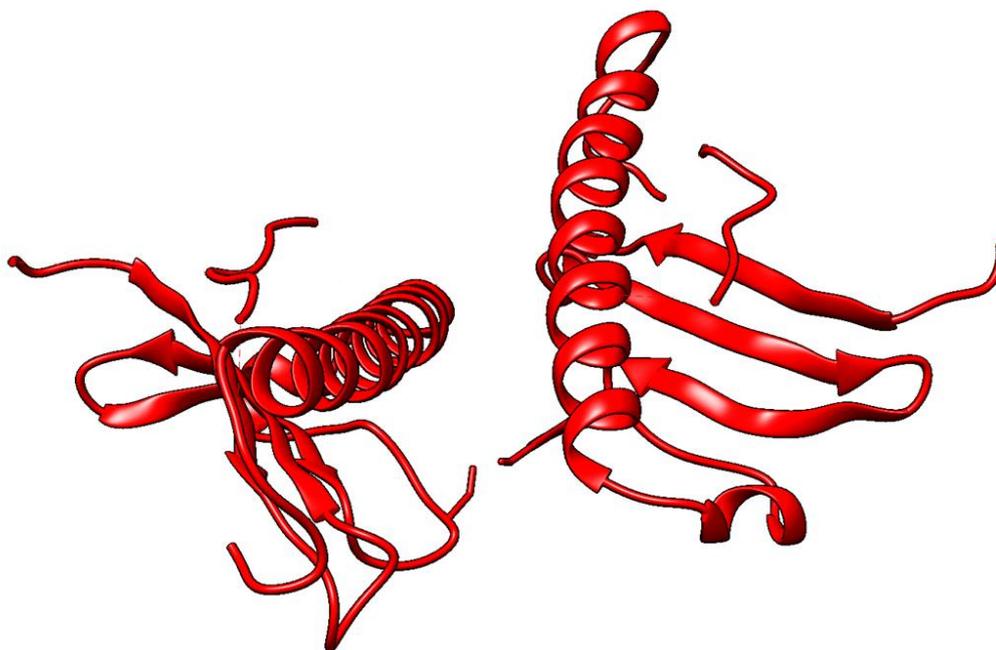
O dímero alternativo foi formado por monômeros da isoforma HLA-G7 contendo peptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  (demonstrado anteriormente) ligados por ponte dissulfeto entre as cisteínas da posição 42 (Cys42) de cada monômero, orientados de forma a manterem a conformação apresentada pelo dímero de HLA-G solúvel (PDB\_ID: 2D31, SHIROISHI *et al.*, 2006) e modelo completo do dímero solúvel de HLA-G gerado durante o trabalho, demonstrado anteriormente).

Para construção desse dímero alternativo o *software* Chimera (PETTERSEN *et al.*, 2004) foi utilizado, para que fosse possível replicar a orientação oblíqua de cada monômero que compõem o dímero solúvel de HLA-G.

A distância entre os dois resíduos de cisteína no modelo foi de 1,989 Å.



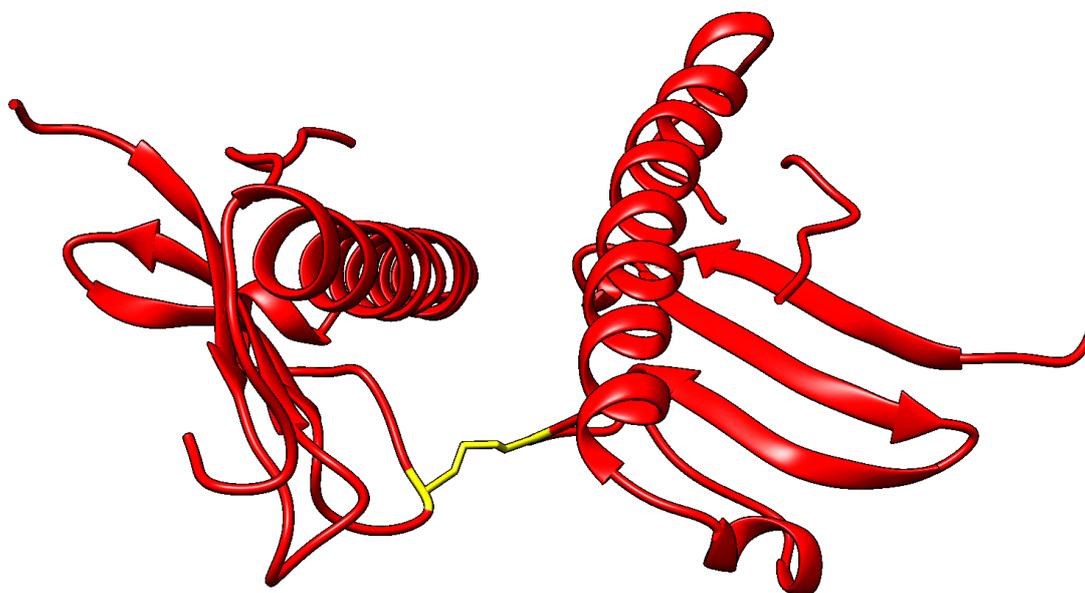
**Figura 99.** Demonstração da orientação do dímero solúvel de HLA-G completo (cinza, demonstrado anteriormente) e monômeros da isoforma HLA-G7 contendo peptídeo, formando a estrutura *G7 fit 2D31* (rosa).



**Figura 100.** Estrutura do dímero alternativo de HLA-G7, *G7 fit 2D31*. A orientação dos monômeros que compõem o dímero encontram-se na orientação oblíqua semelhante ao dímero solúvel de HLA-G, demonstrado anteriormente.

### 6.12.2 Dinâmica Molecular

Para formação da ponte dissulfeto entre o dímero alternativo G7 *fit* 2D31 foi utilizado o *software* GROMACS (HESS *et al.*, 2008), assim como para verificar a estabilidade da estrutura dimérica proposta por meio de dinâmica molecular.



**Figura 101.** Estrutura do dímero alternativo de HLA-G7, G7 *fit* 2D31, após a formação da ponte dissulfeto (amarelo).

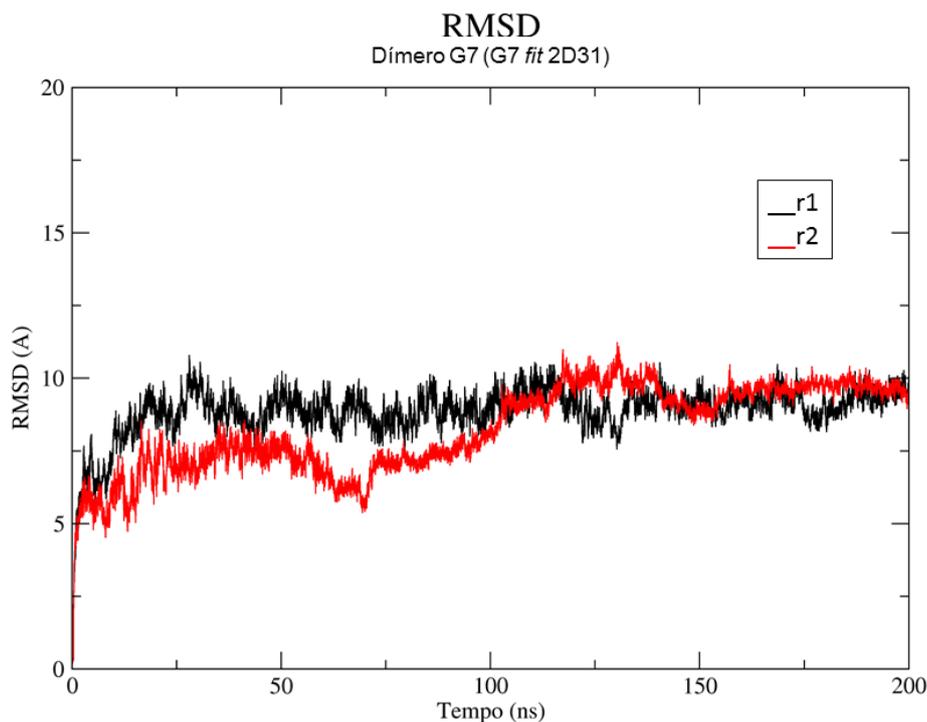
A dinâmica molecular do dímero alternativo G7 *fit* 2D31 foi realizada em duplicata (r1 e r2) das simulações para o modelo gerado. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

O sistema para simulação de dinâmica molecular, além dímero alternativo G7 *fit* 2D31 contém:

- 102441 moléculas de água,
- 196 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

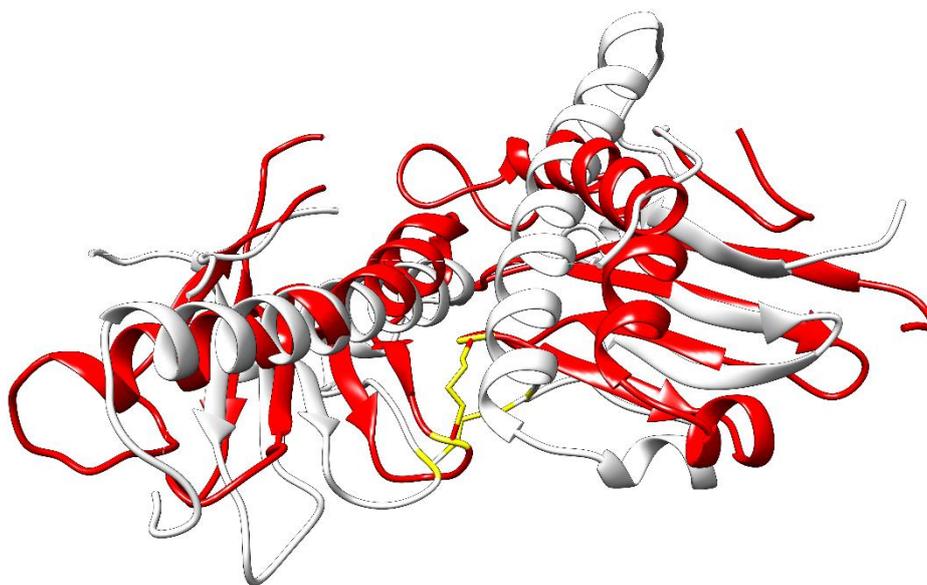
As simulações de dinâmica molecular para o sistema que contém dímero alternativo G7 *fit* 2D31 atingiram o tempo total de 200ns.

O gráfico de RMSD, apresentado na **Figura 102**, demonstra as duplicatas da dinâmica molecular do dímero alternativo *G7 fit 2D31*. Podemos observar que as simulações apresentam valores de RMSD que chegam a 11,24 Å, mas durante toda a trajetória estiveram localizadas na faixa de 8 a 9 Å.



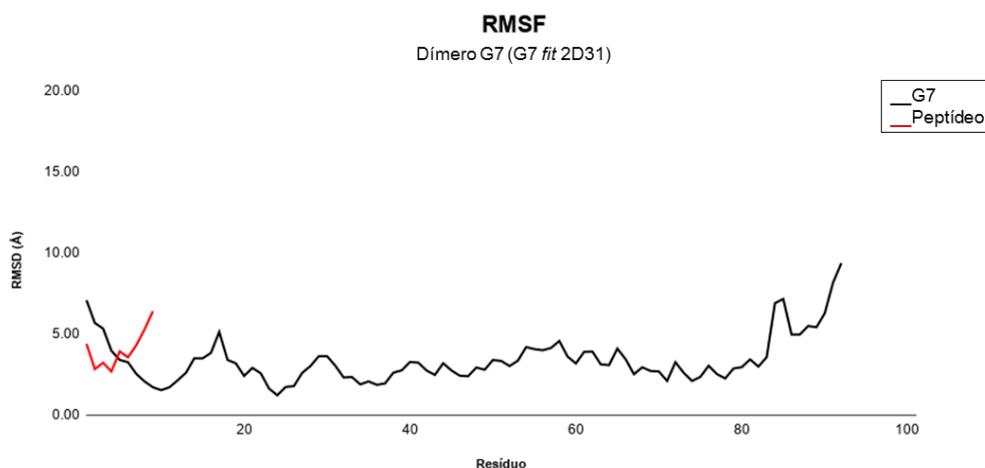
**Figura 102.** Gráfico de RMSD da dinâmica molecular de 200ns referente ao dímero *G7 fit 2D31*.

A **Figura 103** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 200ns da dinâmica do dímero alternativo *G7 fit 2D31*. Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 5,075 Å.



**Figura 103.** Dímero G7 *fit* 2D31 nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular. Ponte dissulfeto em amarelo.

O gráfico apresentando o resultado de RMSF (**Figura 104**) para a simulação mais estável do dímero G7 *fit* 2D31. Foi observada flutuação mais acentuada na porção inicial (resíduos 1 a 5), porção final (resíduos 80 a 92) e resíduos referente ao peptídeo acoplado.



**Figura 104.** Variação estrutural por resíduo da simulação r2 da simulação de dinâmica molecular referente ao dímero G7 *fit* 2D31. Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.

## 6.13 Isoforma solúvel HLA-G7 (G7 MHC-like)

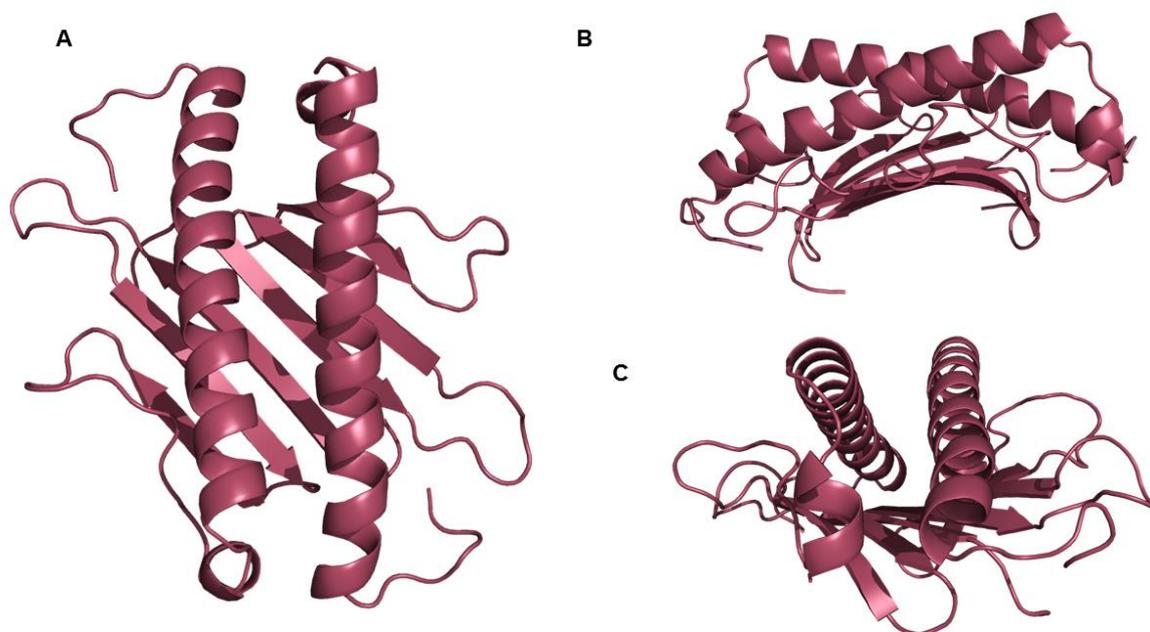
### 6.13.1 Formação do dímero

Com a utilização do servidor *online* ClusPro (KOZAKOV *et al.*, 2017) foi possível observar a formação de outro dímero, que se assemelha muito à fenda que acopla o peptídeo na molécula de HLA-G, o qual foi denominado de MHC-like.

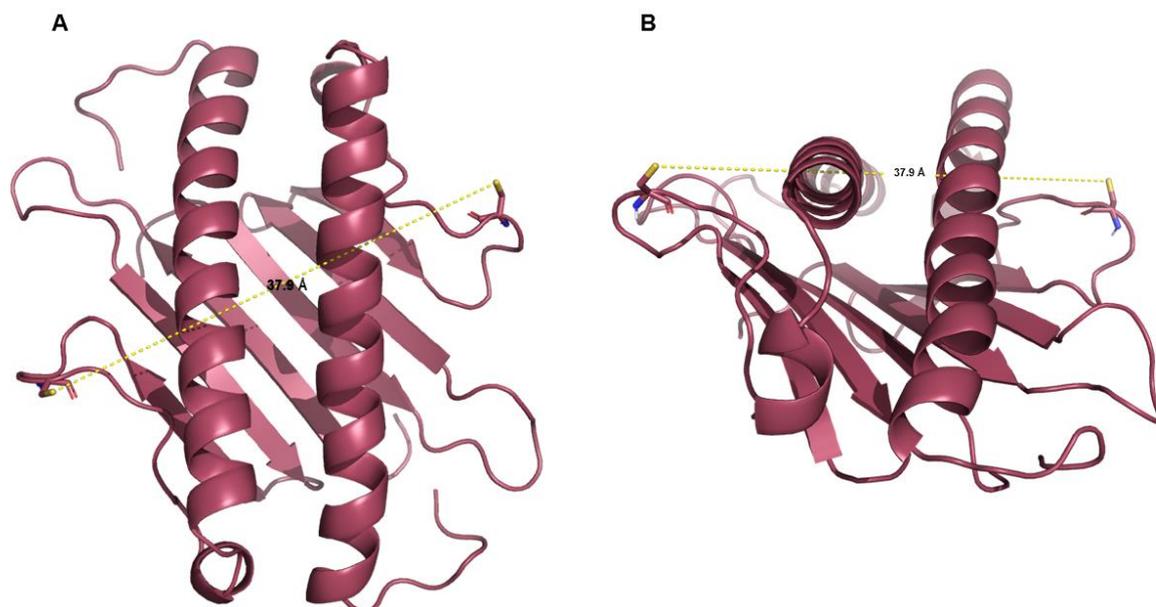
O servidor *online* ClusPro utiliza metodologia de *docking* rígido, mantendo o receptor rígido e o ligante móvel, podendo realizar *docking* sem informações *a priori* do complexo de interesse. O *ranking* dos resultados é baseado na população dos *clusters* formados, não no *score* de energia. Tais protocolos de *docking* proteína-proteína permitem diversas possibilidades com relação à flexibilidade da ancoragem proteica, além de ferramentas para análise dos resultados obtidos com o *docking*.

O *docking* realizado utilizando o servidor ClusPro foi utilizado de forma exploratória e o dímero que obteve a melhor pontuação no algoritmo da ferramenta foi o dímero MHC-like. O modelo selecionado pertencia ao *cluster* 0 e possuía *score* de -1009.

Como comparativo com os dímeros anteriores analisados para a isoforma HLA-G7, o dímero MHC-like possui a distância de 37,9 Å entre as cisteínas na posição 42.



**Figura 105.** Estrutura do dímero proposto MHC-like, formado utilizando o servidor *online* ClusPro. (A) Visão superior, (B) lateral e (C) frontal do mesmo.



**Figura 106.** Estrutura do dímero proposto MHC-like. (A) Visão superior e (B) frontal do mesmo. Tracejado em amarelo representa a distância de 37,9 Å entre as cisteínas na posição 42 (Cys42).

### 6.13.2 Dinâmica Molecular

Para verificar a estabilidade da estrutura dimérica proposta, foi utilizado o *software* GROMACS (HESS *et al.*, 2008).

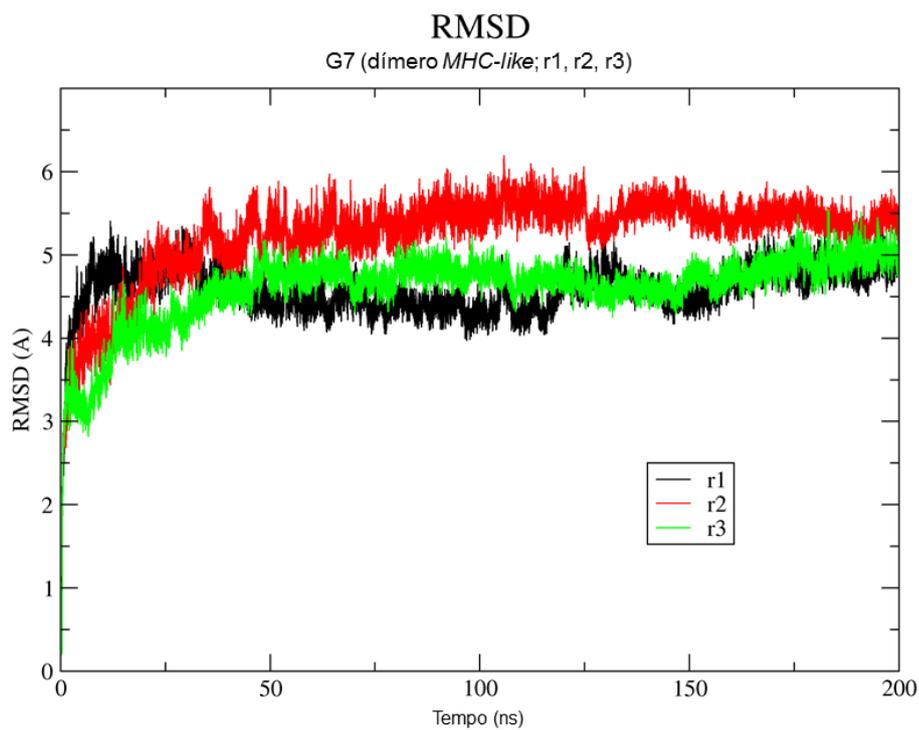
A dinâmica molecular do dímero *MHC-like* foi realizada em triplicata (r1, r2, r3) das simulações para o modelo gerado. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

O sistema para simulação de dinâmica molecular, além dímero *MHC-like* contém:

- 34941 moléculas de água,
- 78 moléculas de íons (Na<sup>+</sup> e Cl<sup>-</sup>).

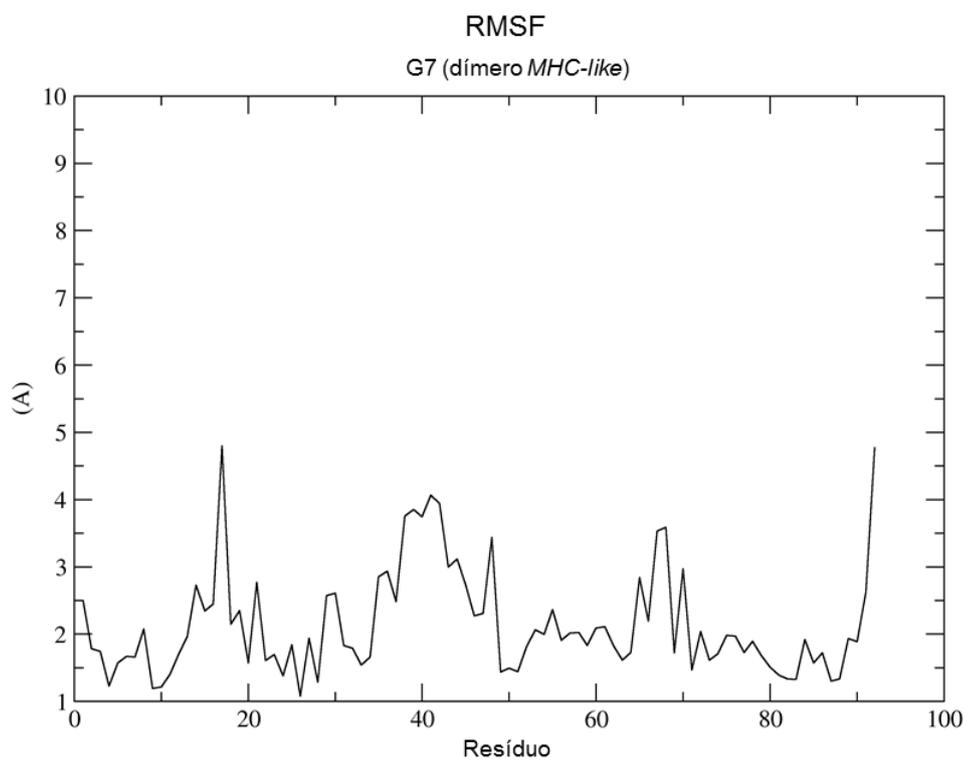
As simulações de dinâmica molecular para o sistema que contém dímero *MHC-like* atingiram o tempo total de 200ns.

O gráfico de RMSD, apresentado na **Figura 107**, demonstra as triplicatas da dinâmica molecular do dímero *MHC-like*. Podemos observar que as triplicatas da simulação apresentaram valores de RMSD que chegaram a 6,19 Å, mas mantiveram-se na faixa de 4,5 a 5,5 durante a trajetória.



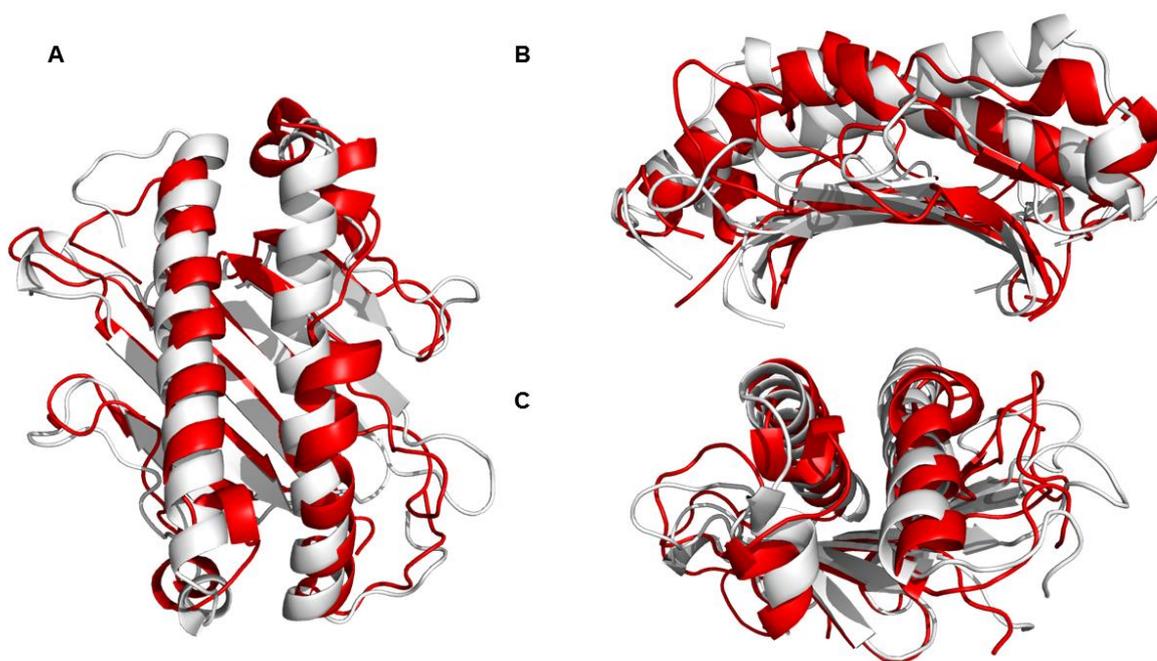
**Figura 107.** Gráfico de RMSD da dinâmica molecular de 200ns referente ao dímero *MHC-like*.

O gráfico apresentando o resultado de RMSF (**Figura 108**) para a simulação mais estável do dímero *MHC-like*, que apresenta flutuação em diversos pontos da estrutura, principalmente nos resíduos 15 a 18, 35 a 45, 65 a 70, 65 a 70 e 90 a 92.



**Figura 108.** Variação estrutural por resíduo da simulação r1 de dinâmica molecular referente ao dímero *MHC-like*. Valores de RMSF das coordenadas dos átomos Ca em relação à estrutura inicial.

A **Figura 109** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 200ns da dinâmica dímero *MHC-like*. Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 2,543 Å.



**Figura 109.** Dímero *MHC-like* nos momentos iniciais (em branco) e finais (em vermelho) dos 200ns da simulação de dinâmica molecular, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular (A, B e C).

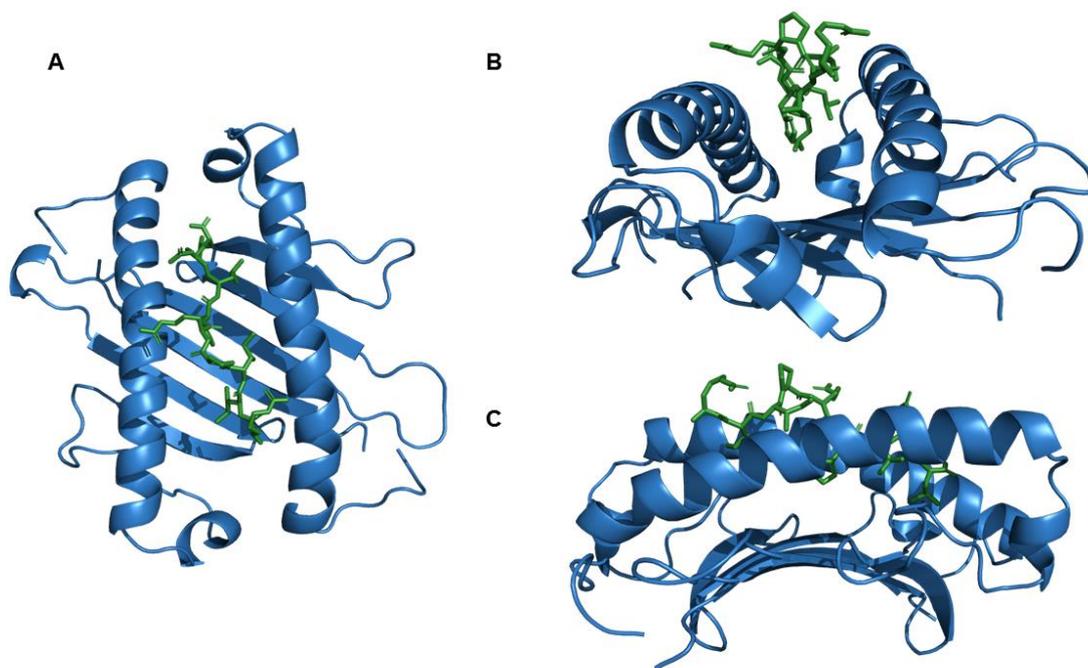
#### 6.14 Isoforma solúvel HLA-G7 (G7 *MHC-like* contendo nonapeptídeo RIIPRHLQL)

Para examinar o comportamento dinâmico da estrutura do dímero *MHC-like*, contendo o peptídeo acoplado ao sulco formado, foi realizado o *docking* do nonapeptídeo RIIPRHLQL, utilizando o *software* Autodock Vina (TROTT *et al.*, 2010).

##### 6.14.1 Seleção do *docking*

A estrutura do nonapeptídeo RIIPRHLQL demonstrada anteriormente foi utilizada para o ensaio *in silico* de *docking* acoplado ao sulco formado no dímero *MHC-like*, denominada receptor para esse experimento de *docking*. O experimento *in silico* de *docking* foi executado da mesma forma já descrita anteriormente.

A melhor conformação de *docking* com dímero *MHC-like* apresentou energia de ligação de -10,5 kcal/mol.



**Figura 110.** Estrutura do dímero proposto *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado. (A) Visão superior, (B) frontal e (C) lateral do mesmo.

### 6.14.2 Dinâmica Molecular

Para verificar a estabilidade da estrutura dimérica proposta, foi utilizado o *software* GROMACS (HESS *et al.*, 2008).

A dinâmica molecular do dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado foi realizada em triplicata (r1, r2, r3) das simulações para o modelo gerado. Foi utilizado o *software* GROMACS 4.6.5 (HESS *et al.*, 2008), que implementa o campo de força G54a7.

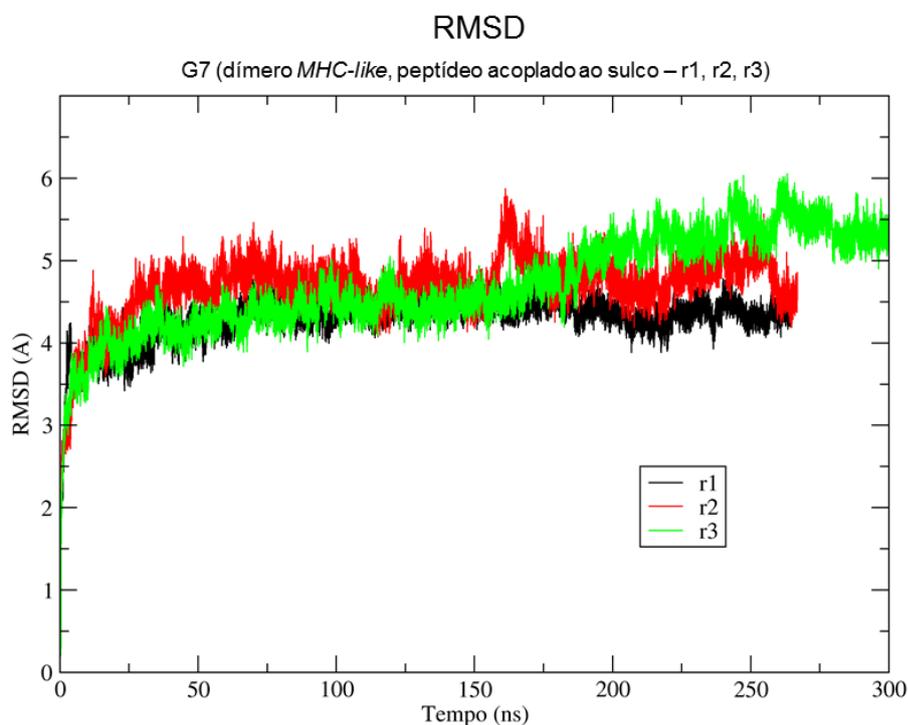
O sistema para simulação de dinâmica molecular, além do dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL contém:

- 38052 moléculas de água,

- 80 moléculas de íons ( $\text{Na}^+$  e  $\text{Cl}^-$ ).

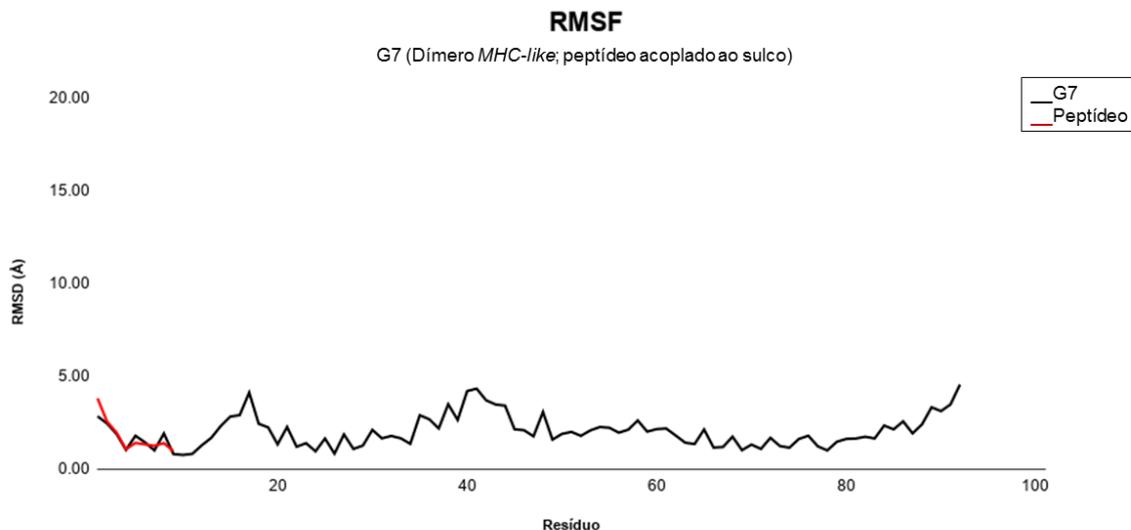
As simulações de dinâmica molecular para o sistema que contém dímero *MHC-like* atingiram o tempo total de 300ns.

O gráfico de RMSD, apresentado na **Figura 111**, demonstra as triplicatas da dinâmica molecular do dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado. Podemos observar que as triplicatas da simulação apresentam valores de RMSD que chegam a 6,05 Å, mas mantiveram-se na faixa de 4 Å durante toda trajetória.



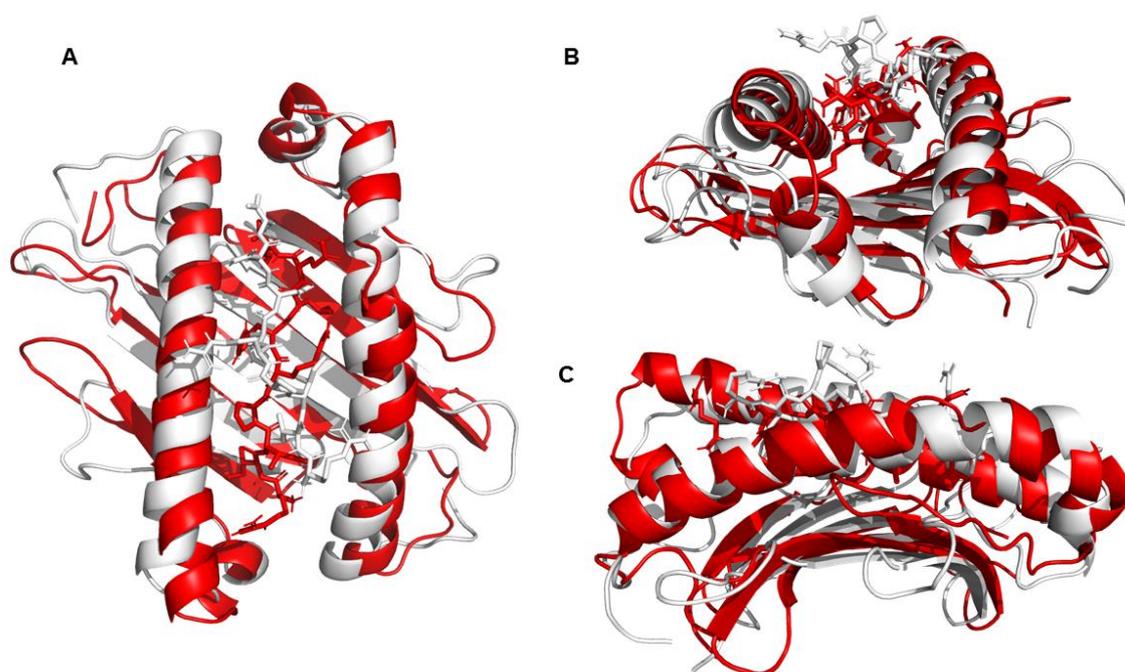
**Figura 111.** Gráfico de RMSD da dinâmica molecular de 300ns referente ao dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado.

O gráfico apresentando o resultado de RMSF (**Figura 112**) para a simulação mais estável do dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado. A flutuação dos resíduos foi homogênea em toda extensão da estrutura e peptídeo acoplado.



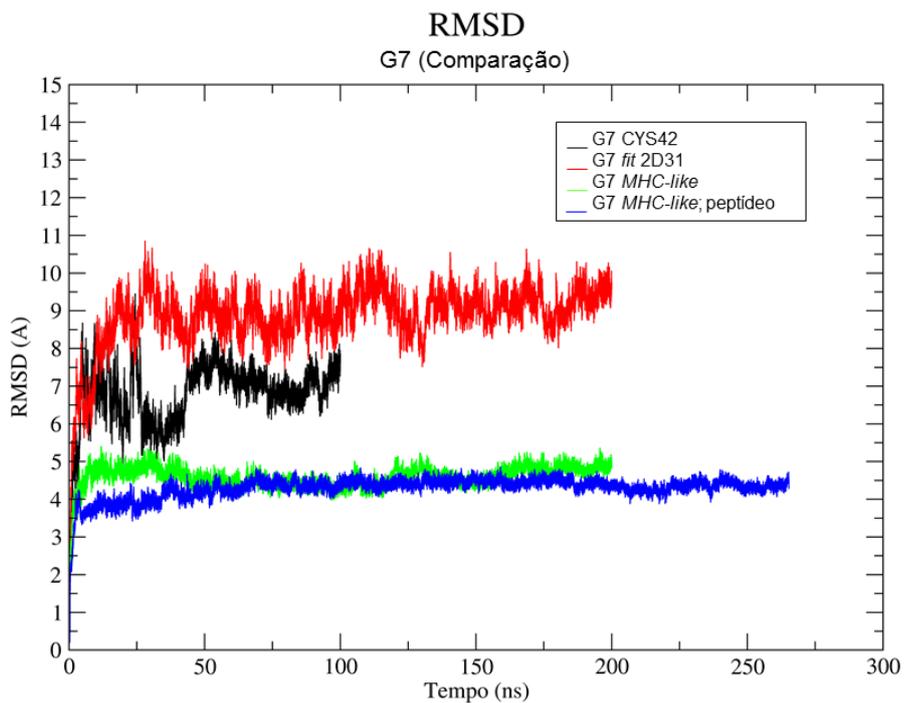
**Figura 112.** Variação estrutural por resíduo da simulação r1 de dinâmica molecular referente ao dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado. Valores de RMSF das coordenadas dos átomos C $\alpha$  em relação à estrutura inicial.

A **Figura 113** ilustra as conformações iniciais (em branco) e finais (em vermelho) dos 300ns da dinâmica dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado. Estão representadas na imagem somente as estruturas proteicas, para facilitar a visualização de quaisquer mudanças conformacionais das proteínas ao longo do tempo. O valor de RMSD calculado entre a conformação inicial (em branco) e final (em vermelho) foi de 3,015 Å.



**Figura 113.** Dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado nos momentos iniciais (em branco) e finais (em vermelho) dos 300ns da simulação de dinâmica molecular, demonstrando as alterações conformacionais sofridas pelo modelo durante a simulação de dinâmica molecular (A, B e C).

Para possibilitar a comparação entre os valores de RMSD das simulações envolvendo os dímeros de HLA-G7 em suas 4 alternativas estruturais (dímero G7-Cys42; dímero G7 *fit* 2D31; dímero *MHC-like* e dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado) foi gerado um gráfico comparativo dessas estruturas (**Figura 114**).



**Figura 114.** Gráfico de RMSD comparativo das simulações envolvendo os dímeros de HLA-G7 em suas 4 alternativas estruturais (dímero G7-Cys42; dímero G7 *fit* 2D31; dímero *MHC-like* e dímero *MHC-like* contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado).

## 7 DISCUSSÃO

### 7.1 Molécula de histocompatibilidade de classe I não clássica HLA-G

HLA-G possui importante papel na supressão da resposta imune e essa função pode ser exercida por moléculas ligadas às superfícies celulares e também em fluidos corporais, na configuração de isoformas solúveis. Não existem estruturas tridimensionais demonstrando a molécula HLA-G de forma completa, então nosso primeiro objetivo foi a obtenção de tal estrutura. A estrutura completa contém os domínios  $\alpha 1$ ,  $\alpha 2$  e  $\alpha 3$ , peptídeo conector, porção transmembrana e cauda citoplasmática. Nas formas membranares e solúveis, HLA-G está usualmente acoplada à proteína  $\beta 2$ -microglobulina, que junto com o peptídeo acoplado ao sulco formado pelos domínios  $\alpha 1$  e  $\alpha 2$  confere estabilidade à molécula.

Com relação à sequência de mRNA, *HLA-G* contém no éxon 1 a sequência referente ao peptídeo sinal. Como para outras moléculas de HLA, a sequência referente ao peptídeo sinal é traduzida juntamente com a molécula HLA-G, estando envolvido no endereçamento e transporte de proteínas na célula. O destino final de tais moléculas depende da sua sequência de aminoácidos, que pode direcionar o transporte do citosol para o núcleo, ER, mitocôndrias, entre outros locais. A maioria das sequências de peptídeos sinal contém de 15-60 resíduos de comprimento e embora suas sequências primárias possam variar muito, suas propriedades físicas, como a hidrofobicidade, são mais importantes no processo de reconhecimento de sinal do que a sequência de aminoácidos exata (ALBERTS, B. 2014; BLUM *et al.*, 2013; VOSS *et al.*, 2013). Usualmente o peptídeo sinal para as moléculas de histocompatibilidade contém cerca de 24 resíduos.

As moléculas HLA-A, -B, -C e -G possuem peptídeo sinal que as direcionam para a via secretora. O peptídeo sinal dessas moléculas contém um segmento altamente conservado que eventualmente é apresentado na superfície celular pela molécula não clássica HLA-E. Na superfície celular, os complexos HLA-E:peptídeo podem se ligar a receptores CD94 / NKG2A em células NK e inibir a lise mediada por células NK. Durante a translocação de proteínas através da membrana do retículo endoplasmático, as sequências referentes ao peptídeo sinal geralmente são clivadas da pré-proteína pela enzima sinalpeptidase, ou seja, a proteína HLA-G madura localizada na superfície celular e suas isoformas solúveis não possuem a sequência referente ao peptídeo sinal (LEMBERG *et al.*, 2001).

Após a clivagem do peptídeo sinal temos a proteína de HLA-G em sua forma madura. Assim, neste estudo, cada uma das porções da molécula foi analisada para obtenção da estrutura completa utilizando metodologias *in silico*, abrangendo a modelagem molecular por homologia, *threading* e *ab initio*. Tais metodologias trazem alternativas à cristalização de proteínas, um processo extremamente complexo, demorado e que para a obtenção do cristal muitas vezes é necessário a modificação de determinados resíduos da molécula. De fato, a obtenção experimental de estruturas solúveis é um processo bem documentado, que segue leis físico-químicas bem compreendidas e envolve a supersaturação da proteína de interesse para a obtenção das moléculas em um cristal tridimensional. A grande dificuldade começa quando tentamos considerar quais condições estimularão as proteínas a se organizarem e produzirem cristais de qualidade. As etapas de modelagem molecular por homologia do presente trabalho utilizaram como molde a cristalografia PDB\_ID: 1YDP, resolvida por Clements e colaboradores para a proteína codificada pelo alelo *HLA-G\*01:04:01:01*. Tal estrutura molde compreende os domínios  $\alpha 1$ ,  $\alpha 2$  e  $\alpha 3$ , peptídeo RIIPRHLQL e proteína  $\beta 2$ -microglobulina acoplada. Dois resíduos são distintos entre o molde e a sequência primária codificada pelo alelo *HLA-G\*01:01:01:01* e o molde PDB\_ID: 1YDP (CLEMENTS *et al.*, 2005).

A resolução estrutural das proteínas de membrana é de maior complexidade, pois para fazer a extração das mesmas da membrana utilizam-se detergentes, o que desestabiliza essas proteínas. As proteínas da membrana representam alvos farmacêuticos importantes e temas interessantes de estudo em relação à biologia celular e bioquímica de proteínas, mas, ainda representam alvos desafiadores para resolução estrutural. Dados publicados em 2016 demonstraram que proteínas contendo a porção transmembrana representam menos de 1% das estruturas experimentais conhecidas determinadas por métodos físicos, reforçando o uso de metodologias *in silico* para obtenção de tais estruturas (CARPENTER *et al.*, 2008; PARKER *et al.*, 2016).

Neste estudo, a metodologia *ab initio* foi utilizada para modelagem do peptídeo conector extracelular e os 6 resíduos (RKKSSD) que compõem a cauda citoplasmática da molécula HLA-G, responsável por: *i*) reter a proteína HLA-G por mais tempo no retículo endoplasmático, *ii*) prolongar a meia-vida na superfície celular, devido à falta de um motivo de endocitose, permitindo interações mais duradouras com células do sistema imunológico (PARK *et al.*, 2001), *iii*) influenciar significativamente a mobilidade da molécula HLA-G, particularmente aquelas com cauda citoplasmática curta, afetando o movimento da molécula na membrana (CAPPS *et al.*, 2004).

Considerando que, a proteína  $\beta$ 2-microglobulina: *i*) é parte integrante da proteína HLA-G completa, contendo 99 aminoácidos e sete cadeias  $\beta$  antiparalelas, *ii*) está vinculada de forma não covalente à molécula de HLA-G, fazendo contatos extensivos com os três domínios da cadeia  $\alpha$ , importantes para manutenção do sulco, *iii*) existem relatos de moléculas de HLA-G que não possuem a  $\beta$ 2-microglobulina (GONEN-GROSS *et al.*, 2005), neste trabalho, a modelagem do HLA-G de membrana e isoformas solúveis foram modeladas por homologia com ou sem a  $\beta$ 2-microglobulina, utilizando o molde PDB\_ID: 1YDP. Além da utilização da  $\beta$ 2-microglobulina, as moléculas HLA-G membranares ou solúveis foram acopladas ao peptídeo RIIPRHLQL, proveniente da histona H2A. Esse peptídeo é um dos mais abundantes em células transfectadas com HLA-G, fazendo extensos contatos polares e apolares, incluindo 17 pontes de hidrogênio (H), 16 pontes de hidrogênio mediadas por água e grande número de interações de van der Waals. O peptídeo ligado adota uma conformação alongada, ancorada nas extremidades N- e C-terminal, com uma região abaulada, localizada no aminoácido P (Pro, prolina, posição 4), que se projeta para fora do sulco. Clements e colaboradores descreveram a ancoragem do peptídeo ao sulco por uma rede de pontes de hidrogênio bem conservadas em moléculas HLA de classe I, envolvendo Y (Tyr, tirosina, posição 84), S (Ser, serina, posição 143), K (Lys, lisina, posição 146), e interações mediadas por água com T (Thr, treonina, posição 73), N (Asn, asparagina, posição 77) e T (Thr, treonina, posição 80) (CLEMENTS *et al.*, 2005; WALPOLE *et al.*, 2010). Considerando: *i*) a importância da disposição do peptídeo dentro do sulco, *ii*) a protrusão mínima do peptídeo na fenda, atribuída aos contatos das cadeias laterais do peptídeo ao sulco, predominantemente hidrofóbicas, dependentes dos resíduos R (arginina, posição 1), I (isoleucina, posição 2), I (isoleucina, posição 3), H (histidina, posição 6), L (leucina, posição 7), *iii*) a alta exposição ao microambiente, incluindo P (prolina, posição 4), R (arginina, posição 5) e Q (glutamina, posição 8), neste estudo, essas características foram analisadas para inferir a semelhança entre o modelo gerado e o molde utilizado (CLEMENTS *et al.*, 2005).

O peptídeo RIIPRHLQL é acomodado no sulco de ligação ao peptídeo por resíduos-âncora conservados na posição 2 (I, isoleucina) e 9 (L, leucina) (ISHITANI *et al.*, 2003; LEE *et al.*, 1995; DIEHL *et al.*, 1996). De acordo com trabalho publicado por Walpole e colaboradores, o modo de ligação do peptídeo RIIPRHLQL no sulco de HLA-G é semelhante ao HLA-E. Embora o repertório peptídico de HLA-G não seja tão restrito quanto o de HLA-E, o grupo determinou a estrutura da proteína HLA-G em complexo com dois peptídeos próprios não relacionados, KGPPAALTL e KLPAQFYIL, para avaliar simultaneamente as características conservadas da ligação peptídica pelo HLA-G e fornecer uma base para a adaptabilidade de HLA-G ao acomodar

sequências de peptídeos distintos. Esse grupo demonstrou que uma extensa rede de contatos ao longo do comprimento do peptídeo está presente em todos os complexos peptídeo:HLA-G, mas, no entanto, a superposição dos três complexos peptídeo: HLA-G mostrou que, embora houvesse pouca diferença nos terminais N e C, a região central do peptídeo demonstrou um maior grau de flexibilidade conformacional. Esta maleabilidade no peptídeo também resultou em um grau de variabilidade nos resíduos de contato de HLA-G, o que, em última instância, afeta a região da hélice do domínio  $\alpha 2$ , onde foi possível observar o aumento do sulco no complexo HLA-G:KLPAQFYIL. A natureza do peptídeo também afetou a estabilidade térmica do complexo peptídeo:HLA-G. O peptídeo RIIPRHLQL teve o maior efeito estabilizador, possivelmente devido ao grande número de interações polares que faz com HLA-G, enquanto HLA-G:RLPKDFRIL foi o menos estável (WALPOLE *et al.*, 2010).

A estrutura cristalográfica de HLA-G (PDB\_ID: 1YDP) utilizada como molde para modelagem por homologia no presente trabalho possui uma mutação localizada na posição 42, uma troca de cisteína (resíduo Cys42) para serina, a fim de evitar a formação de dímeros e possibilitar a obtenção do cristal do monômero de HLA-G. Podemos mencionar outros resíduos em posições de importância biológica para a molécula de HLA-G, como por exemplo a cisteína localizada na posição 147 no domínio  $\alpha 2$ . Alguns autores acreditam que essa cisteína poderia levar à formação de estruturas diméricas, enquanto outros argumentam que a pouca acessibilidade a esse resíduo dificultaria a formação de dímeros (REBMANN *et al.*, 2016; BOYSON *et al.*, 2002).

Os domínios extracelulares da molécula HLA-G interagem com receptores de leucócitos, incluindo CD8, ILT2, ILT4 e o receptor de células NK, KIR2DL4 (DONADI *et al.*, 2011). Genes *KIR* podem codificar moléculas que exibem dois (KIR2D) ou três (KIR3D) domínios de tipo Ig extracelular, que podem conter motivos ITIM (L - cauda longa, receptores inibitórios) ou não (S - cauda curta, receptores ativadores) no domínio citoplasmático. O gene *KIR2DL4* codifica um receptor que contém uma arginina no domínio transmembranar e um único motivo ITIM no domínio citoplasmático. Uma série de receptores ativadores e inibidores é expressa na superfície da maioria das células NK e macrófagos, e a função efetora final dos mesmos depende do equilíbrio entre receptores ativadores e inibidores. Devido a essas características, o papel da interação de KIR2DL4 com HLA-G na modulação da resposta imune tem sido alvo de debate. Uma vez que os resíduos M (metionina, posição 76) e Q (glutamina, posição 79) são únicos para HLA-G, e uma vez que a mutação desses resíduos para A (alanina, posição 76 / 79) afeta a afinidade de ligação entre KIR2DL4 e HLA-G, esses resíduos podem ser candidatos relevantes para essa interação (HSU *et al.*, 2002; APPS *et al.*, 2008; DONADI *et al.*, 2011).

Resíduos localizados nas posições 195 e 197 do domínio  $\alpha 3$  da proteína HLA-G são ocupados por fenilalanina e tirosina, respectivamente, posições já descritas como sítio de ligação para o receptor inibitório ILT4. Outras moléculas de HLA de classe I possuem serina e histidina nessas mesmas posições e acredita-se que essa alteração estrutural e maior hidrofobicidade decorrente dessas substituições seja a base para a maior afinidade de HLA-G para ILT4 em comparação com outras moléculas HLA de classe I. Os resíduos Y (tirosina, posição 36) e R (arginina, posição 38) de ILT4 se ligam ao anel 195-197 do domínio  $\alpha 3$  de HLA-G, enquanto que os resíduos Y (tirosina, posição 38) e Y (tirosina, posição 76) de ILT2 se ligam ao resíduo F (fenilalanina, posição 195) de HLA-G (CLEMENTS *et al.*, 2005; SHIROISHI *et al.*, 2006).

CD8  $\alpha/\alpha$  faz contato com o domínio  $\alpha 3$  de HLA-G nos resíduos 223 a 229. Este modo de interação evita a ligação simultânea de uma segunda molécula de HLA a CD8  $\alpha/\alpha$ . HLA-G é capaz de ligar CD8  $\alpha/\alpha$  com afinidade mediana, quando comparado com afinidades clássicas de HLA:CD8  $\alpha/\alpha$ . Essas medidas são consistentes com experimentos que indicam que HLA-G é capaz de mediar adesão célula-célula através da interação com CD8  $\alpha/\alpha$ . Portanto, é possível que a interação HLA-G:CD8 possa mediar a seleção positiva e negativa de CTLs restritas à classe I e facilite o reconhecimento de antígenos restritos a HLA-G por essas células, de forma semelhante à CTLs convencionais restritas à moléculas de classe I clássica (GAO *et al.*, 2000; ALEGRE *et al.*, 2014).

É interessante observar que o domínio  $\alpha 3$  de HLA-G é o mesmo local de interação com CD8  $\alpha/\alpha$  e ILTs. Embora os sítios de ligação de CD8  $\alpha/\alpha$  e ILTs se sobreponham, os ILTs inibem a ligação de moléculas de CD8  $\alpha/\alpha$  a HLA, indicando que a função inibidora de ILTs pode ser dupla, deslocando CD8  $\alpha/\alpha$  e ativando ITIMs (DONADI *et al.*, 2011).

De forma semelhante à outras proteínas conectadas à membrana celular, a porção transmembrana da molécula HLA-G incorporada na bicamada lipídica é mantida por interações hidrofóbicas com o interior lipídico da bicamada e também por interações iônicas com os grupos de cabeça polar dos fosfolipídios (LODISH, Molecular Cell Biology, 2000).

As moléculas lipídicas constituem cerca de 50% da massa da maioria das membranas celulares animais, o restante sendo proteínas. Existem aproximadamente  $5 \times 10^6$  moléculas lipídicas em uma área de  $1 \mu\text{m} \times 1 \mu\text{m}$  de bicamada lipídica, ou cerca de  $10^9$  moléculas lipídicas na membrana plasmática de uma célula animal pequena. Todas as moléculas lipídicas nas membranas celulares são anfifílicas, ou seja, têm uma extremidade hidrofílica (polar) e uma extremidade hidrofóbica (apolar). Os lipídios da membrana mais abundantes são os fosfolipídios. Estes têm um

grupo de cabeça polar contendo um grupo de fosfato e duas caudas de hidrocarbonetos hidrofóbicos. As diferenças no comprimento e saturação das caudas dos ácidos graxos influenciam a forma como as moléculas de fosfolipídios se organizam umas contra as outras, afetando assim a fluidez da membrana (ALBERTS, B. 2014).

Com relação ao peptídeo sinal, este trabalho mostra pela primeira vez o modelo completo da proteína HLA-G, especificamente aquela codificada pelo alelo *HLA-G\*01:01:01:01*. A análise dos 58 alelos HLA-G descritos até o momento indica que o peptídeo sinal não varia, (IMGT/HLA. Disponível em <http://www.ebi.ac.uk/cgi-bin/ipd/imgt/hla/align.cgi> banco de dados versão 3.31.0; ROBINSON *et al.*, 2013), sendo composto por uma maioria de resíduos de caráter hidrofóbico, localizados em toda sua extensão, condizente com as propriedades físico-químicas esperadas de tal sequência sinalizadora. Após a retirada da sequência referente ao peptídeo sinal e definição da sequência primária correta para a modelagem molecular foi possível a obtenção da estrutura completa da molécula HLA-G.

O molde escolhido para a modelagem por homologia apresentou 86.6% de identidade e 87.3% de similaridade para o alinhamento global (**Figura 15**) entre a sequência primária madura codificada pelo alelo *HLA-G\*01:01:01:01* e a sequência primária da estrutura cristalográfica PDB\_ID:1YDP. Para o alinhamento local (**Figura 14**) temos uma sequência de 275 resíduos, que possui 98.9% de identidade e 99.6% de similaridade entre a sequência primária de HLA-G e a sequência primária da estrutura cristalográfica PDB\_ID:1YDP. Tais valores de identidade e similaridade obtidos para os alinhamentos locais e globais demonstraram que o molde escolhido era indicado para a modelagem por homologia. Como o intuito foi a modelagem da proteína completa mais parecida com a forma nativa, neste trabalho a estrutura gerada possui a cisteína na posição 42, mutada na estrutura cristalográfica.

A etapa de modelagem por homologia da porção extracelular de HLA-G contendo os domínios  $\alpha 1$ ,  $\alpha 2$  e  $\alpha 3$  foi executada com sucesso e a estrutura obtida foi bem pontuada em todos os *softwares* utilizados para análise da qualidade dos modelos, como descrito na seção de Resultados. A avaliação da qualidade nos garantiu que os modelos gerados possuem compatibilidade entre a sequência de aminoácidos e o modelo gerado; grande confiança do enovelamento do modelo; melhoria do refinamento proteico; qualidade estereoquímica por meio de avaliação dos comprimentos de ligação, ângulos planos, planaridade dos anéis de cadeias laterais, planaridade das ligações peptídicas, ângulos torcionais da cadeia principal e das cadeias laterais; quiralidade; conformações das cadeias laterais e impedimento estérico entre pares de

átomos não-ligados. Além disso, o gráfico de Ramachandran (**Figura 16**) demonstrou que os aminoácidos que compõem o modelo se encontram em regiões energeticamente favoráveis e nenhum dos resíduos do modelo encontra-se em regiões não permitidas. A avaliação visual do modelo gerado durante essa etapa confirmou a qualidade do modelo gerado.

As etapas de modelagem *ab initio* das porções referentes ao peptídeo conector (**Figura 18**) e cauda citoplasmática (**Figura 20**) também seguiram padrões rigorosos para escolha do melhor modelo, levando em conta o enovelamento da proteína e a conformação que apresentava menor energia livre, e conseqüentemente, estaria mais próxima de seu enovelamento nativo. Essa porção referente ao peptídeo conector é formada por uma alça e serve para a ligação do domínio  $\alpha 3$  à região transmembrana. Não existe nenhum relato na literatura de resíduos com funções relevantes localizados nessa porção da estrutura, mas nosso modelo é o primeiro a conter tais resíduos resolvidos. A cauda citoplasmática de HLA-G é formada por uma alça de 6 resíduos.

A porção transmembrana da molécula HLA-G (**Figura 19**) foi resolvida *in silico* com sucesso utilizando a metodologia de *threading*, pois assim como o peptídeo conector, essa porção não foi resolvida na cristalografia PDB\_ID: 1YDP. A molécula HLA-G pode ser classificada como proteína de membrana de tipo I, de passagem única pela membrana plasmática, com sua porção transmembrana organizada como  $\alpha$ -hélice. O comprimento desta  $\alpha$ -hélice é apenas suficiente para atravessar a bicamada de fosfolipídios, contendo 24 resíduos (MGIVAGLVVLAADVVTGAAVAVLW) de caráter hidrofóbico.

Após a obtenção da cadeia pesada de HLA-G completa (**Figura 20**), a etapa de modelagem por homologia da proteína  $\beta 2$ -microglobulina foi executada com sucesso e a estrutura obtida foi bem pontuada em todos os *softwares* utilizados para análise da qualidade dos modelos, como descrito na seção de Resultados (**Figura 22**). A avaliação da qualidade nos garantiu que os modelos gerados possuem compatibilidade entre a sequência de aminoácidos e o modelo gerado; grande confiança do enovelamento do modelo; melhoria do refinamento proteico; qualidade estereoquímica por meio de avaliação dos comprimentos de ligação, ângulos planos, planaridade dos anéis de cadeias laterais, planaridade das ligações peptídicas, ângulos torcionais da cadeia principal e das cadeias laterais; quiralidade; conformações das cadeias laterais e impedimento estérico entre pares de átomos não-ligados. A avaliação visual do modelo gerado durante essa etapa confirmou a qualidade do modelo gerado.

O modelo apresentado neste trabalho para a molécula de HLA-G completa (**Figura 24**) contém o peptídeo de forma muito semelhante ao demonstrado na estrutura cristalográfica

PDB\_ID: 1YDP, fato confirmado pelo baixo valor de RMSD obtido quando feita a comparação entre o sulco da estrutura resolvida experimentalmente e nosso modelo *in silico*.

O modelo demonstrado neste trabalho foi submetido à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo contendo a molécula de HLA-G completa, peptídeo RIIPRHLQL,  $\beta$ 2-microglobulina, bicamada lipídica, íons e água (**Figura 32**). Os valores de RMSD (**Figura 34**) demonstram a estabilidade do modelo gerado como um todo, indicando que as variações nos valores de RMSD das triplicatas se deve principalmente à movimentação da porção transmembrana na bicamada lipídica, além da movimentação natural observada pela interação de todas as moléculas do sistema. O valor de RMSD calculado para comparação entre as posições iniciais e finais da dinâmica (**Figura 36**), além das imagens retiradas nos momentos iniciais e finais da trajetória apoiam essas conclusões, pois foi possível observar que todos os componentes do modelo são mantidos enovelados durante a simulação de forma muito parecida ao estágio inicial da simulação, o peptídeo foi mantido inserido no sulco sem que fossem perdidas as ligações com os resíduos que compõem o sulco. Tais observações são corroboradas pelos resultados obtidos para os valores de RMSF (**Figura 35**) da simulação mais estável, que indica maior flutuação dos átomos referente às porções terminais da molécula gerada, que compreendem a porção transmembrana e cauda citoplasmática. A flutuação concentrada nessas regiões se deve à movimentação da porção transmembrana na bicamada lipídica, que possui certa fluidez e permite tais movimentações (**Figura 37**). A simulação confirmou a qualidade da modelagem realizada, pois as estruturas modeladas foram conservadas durante toda a simulação, além de demonstrar pela primeira vez o comportamento dinâmico da proteína HLA-G inserida em uma bicamada lipídica.

## 7.2 Dímero solúvel da isoforma HLA-G1

O modelo completo para o dímero solúvel de HLA-G1 foi gerado e analisado dinamicamente. Como citado anteriormente para o monômero de HLA-G, a presença de um resíduo de cisteína na posição 42 no domínio  $\alpha$ 1 da cadeia pesada é uma característica única da molécula HLA-G, que permite a formação de dímeros por meio de uma ligação dissulfeto intermolecular Cys42-Cys42. Os dímeros podem ser formados por monômeros de HLA-G ligados à membrana ou moléculas solúveis e até mesmo por construções de HLA-G livres de  $\beta$ 2-

microglobulina. Uma vez formado, o dímero tem uma orientação oblíqua, expondo os locais de ligação para CD8 e receptores ILT2 e ILT4, permitindo que uma molécula dimérica de HLA-G possa interagir com duas moléculas ILTs ou CD8. Além de proporcionar maior afinidade para os ILTs em relação aos monômeros, os dímeros de HLA-G exibem melhor sinalização intracelular, mediada por ILTs (DONADI *et al.*, 2011).

Nenhum dos alelos de HLA-G já descritos apresenta variação de nucleotídeos no códon 42 (TGT) que codifica o resíduo de cisteína. Portanto, aparentemente, todos os alelos que codificam moléculas HLA-G não truncadas podem formar dímeros. Uma vez que a interface do dímero pode ser estabilizada por interações eletrostáticas e pontes de hidrogênio, é possível que os resíduos polimórficos próximos à posição 42 possam influenciar a estabilidade do dímero ou alterar a flexibilidade da ponte dissulfeto (DONADI *et al.*, 2011).

Da mesma forma como demonstrado para a modelagem da estrutura completa da proteína HLA-G inserida na bicamada lipídica, todas as etapas de modelagem por homologia do dímero solúvel de HLA-G foram executadas com sucesso e as estruturas obtidas foram bem pontuadas em todos os *softwares* utilizados para análise da qualidade dos modelos. Além disso, o gráfico de Ramachandran demonstra que nenhum dos aminoácidos que compõe o modelo se encontra em regiões não permitidas (RAMACHANDRAN *et al.*, 1963; LUTHY *et al.*, 1992; WIEDERSTEIN e SIPPL, 2007; BENKERT *et al.*, 2008).

A cristalografia utilizada como molde (PDB\_ID:2D31) não possuía os resíduos das posições 195, 196, 197, 266 e 267 resolvidos, todos localizados no domínio  $\alpha 3$ , região já relacionada a interações com receptores do sistema imune. Da mesma forma que observado para o monômero, as posições 195 e 197 são ocupadas por fenilalanina e tirosina, respectivamente, na molécula de HLA-G e serina e histidina em outras moléculas de HLA de classe I. O aumento da hidrofobicidade desta região de HLA-G, em relação a outras moléculas de HLA de classe I, provavelmente contribui para a maior afinidade de HLA-G para ILTs, de modo que possam competir com CD8 para ligação de HLA de classe I (CLEMENTS *et al.*, 2005).

A capacidade de HLA-G para formar homodímeros ligados por pontes dissulfeto pode teoricamente aumentar a avidéz de HLA-G para os receptores ILT2 e ILT4. Tanto na estrutura cristalográfica PDB\_ID: 2D31 quanto na modelagem apresentada neste trabalho, o homodímero de HLA-G ligado por pontes dissulfeto apresenta configuração oblíqua, expondo dois locais de ligação facilmente acessíveis para receptores ILTs e CD8, fornecendo modelos plausíveis para formação de complexos 1:2, ou seja, 1 molécula de dímero de HLA-G:2 receptores

(ILT2/ILT4/CD8). Resultados publicados por outros grupos de pesquisa indicaram que o dímero de HLA-G conferiu maior avidéz em uma orientação estrutural adequada para induzir a sinalização eficiente por ILTs, resultando em efeitos imunossupressores dominantes. Portanto, as interações HLA-G: ILTs e HLA-G: CD8 podem induzir uma ampla gama de tolerância imunológica (SHIROISHI *et al.*, 2006).

A comparação da estrutura gerada neste trabalho para a molécula de HLA-G completa e o dímero solúvel de HLA-G1 nos permite observar que os resíduos implicados na interação com receptores ILT2 e ILT4 ficam bem próximos da porção transmembrana na molécula HLA-G completa de superfície, enquanto no dímero solúvel de HLA-G1, devido à orientação oblíqua do dímero, existe maior acesso à região desses resíduos. Esse poderia ser um dos motivos para maior afinidade do dímero de HLA-G1 solúvel aos receptores ILT2 e ILT4.

Tanto monômeros quanto dímeros de HLA-G podem existir na superfície celular e em fluidos biológicos, no entanto, no caso do dímero HLA-G1, os sulcos que comportam o peptídeo em cada monômero do dímero de HLA-G1 parecem muito próximos para serem totalmente acessíveis para receptores de células T. Portanto, o dímero de HLA-G pode não funcionar como molécula apresentadora de antígeno para respostas de células T, pois além de os resíduos ficarem pouco acessíveis ao microambiente, a própria orientação da molécula dificulta a apresentação do peptídeo. No entanto, os resíduos que fazem contato com os receptores ILTs e CD8 ficam livres para interação. Essas características podem explicar em parte a indução menos eficiente de células T citotóxicas. Além disso, existe a possibilidade de que dímeros que não possuem  $\beta$ 2-microglobulina possam participar da regulação de função das células T (BOYSON *et al.*, 2002; CLEMENTS *et al.*, 2005). No entanto dímeros sem a  $\beta$ 2-microglobulina não foram gerados no presente trabalho.

A dimerização de HLA-G parece não ocorrer até que a molécula tenha passado através do aparelho de Golgi, uma vez que a forma dimérica é completamente resistente à endoglicosidase H (CAMPBELL *et al.*, 2012). A endoglicosidase H é amplamente utilizada para relatar a extensão do processamento de oligossacarídeos que uma proteína sofreu (ALBERTS, B. 2014). Estudos estruturais recentes mostraram que o homodímero de HLA-G1 associado a  $\beta$ 2-microglobulina é a principal fração funcional expressa por células de trofoblasto (APPS *et al.*, 2007), enquanto os dímeros HLA-G1 ligados por pontes dissulfeto isentos de  $\beta$ 2-microglobulina são produzidos por células de citotrofoblasto viloso (MORALES *et al.*, 2007). Esses resultados fornecem fortes

evidências de que a oligomerização de HLA-G desempenha papel imunossupressor importante na interface materno-fetal.

Estudo publicado por Ezeakile e colaboradores mostrou que aumento da expressão de HLA-G ligado à membrana de monócitos e altos níveis de dímeros de HLA-G1 no plasma estão associados ao aumento da sobrevivência do aloenxerto renal. Do mesmo modo, determinaram que a presença de dímeros HLA-G1 solúveis está relacionada aos níveis mais baixos de citocinas pró-inflamatórias encontrado nos pacientes, relacionando aos dímeros um papel no controle do estado inflamatório que acompanha o enxerto renal (EZEAKILE *et al.*, 2014).

Neste trabalho, mostramos pela primeira vez o modelo completo do dímero de HLA-G1 solúvel (**Figura 41**), contendo os resíduos referentes aos sítios de ligação para receptores inibitórios ILT2 e ILT4. A dinâmica molecular do sistema contendo o dímero solúvel de HLA-G1 completo, peptídeo RIIPRHLQL,  $\beta$ 2-microglobulina, íons e água apresentou altos valores de RMSD (**Figura 43**) e a análise da trajetória demonstrou grande movimentação da orientação oblíqua do dímero, indicando que os altos valores de RMSD das duplicatas se devem principalmente à movimentação de aproximação e distanciamento de um monômero em relação ao outro durante a simulação, ao invés de estarem relacionados ao desenovelamento da estrutura. A ponte dissulfeto foi mantida intacta durante toda a trajetória, assim como os peptídeos localizados na fenda de cada monômero que compõem o dímero, sem que fossem perdidas as ligações com os resíduos que compõem os sulcos.

Durante a simulação, foi possível observar que dependendo do momento analisado, a conformação oblíqua do dímero poderia favorecer a ligação dos receptores ILTs ou CD8 ou não, dependendo do ângulo da orientação que o dímero de HLA-G1 apresentava. Tais observações são corroboradas pelos resultados obtidos para os valores de RMSF (**Figura 44**) da simulação mais estável, que indica maior flutuação dos resíduos referentes às porções terminais do dímero solúvel gerado, que compreendem a porção referente ao domínio  $\alpha$ 3, sítio de ligação para receptores inibitórios ILT2 e ILT4, além de CD8.

A simulação confirmou a qualidade da modelagem realizada, pois as estruturas modeladas foram conservadas durante toda a trajetória dinâmica. Ainda, a orientação oblíqua do dímero HLA-G1 durante a maior parte da simulação permitiu maior exposição do domínio  $\alpha$ 3, onde estão localizados os sítios de ligação aos receptores ILTs.

### 7.3 Isoforma solúvel HLA-G5

A isoforma solúvel HLA-G5 é composta dos éxons 1, 2, 3 e 4 do gene *HLA-G*, apresentando os mesmos domínios extra globulares da proteína HLA-G, além de conservar o íntron 4, que bloqueia a tradução do domínio transmembrana (éxon 5). A região 5' do íntron 4 é traduzida até a geração de um *stop* códon, conferindo à isoforma HLA-G5 uma cauda de 21 aminoácidos relacionados com sua solubilidade (FUJII *et al.*, 1994). O perfil majoritariamente hidrofílico da cauda proveniente da tradução do íntron 4 foi confirmado neste trabalho pela análise do perfil de hidrofobicidade.

A estrutura extracelular da isoforma HLA-G5 é idêntica às estruturas de moléculas de HLA de classe I clássicas, composta por 3 domínios globulares ligados de forma não covalente à  $\beta$ 2-microglobulina, um nonapeptídeo e a cauda de 21 aminoácidos resultante da tradução do íntron. As demais isoformas constituem estruturas mais simples, com somente um ou dois domínios globulares, podendo ou não apresentar ligação com a  $\beta$ 2-microglobulina e peptídeos (KURIMOTO *et al.*, 2013; CAROSELLA *et al.*, 2008).

Isoformas que possuem o domínio  $\alpha$ 1 teoricamente podem ser reconhecidas pelo receptor KIR2DL4, enquanto as que possuem o domínio  $\alpha$ 3 expresso, podem interagir com CD8, ILT2 ou ILT4. Como acontece com a ligação de HLA-G à  $\beta$ 2-microglobulina, a presença de estruturas associadas a  $\beta$ 2-microglobulina direciona o reconhecimento pelo receptor ILT2, enquanto as moléculas livres de  $\beta$ 2-microglobulina servem como ligantes para ILT4 (NARDI *et al.*, 2016). Existem relatos da interação de ILT2 com isoformas HLA-G5 que possuem a  $\beta$ 2-microglobulina associada, enquanto ILT4 é capaz de reconhecer isoformas que não possuem  $\beta$ 2-microglobulina associada (HOWANGYIN *et al.*, 2012). O receptor KIR2DL4 também poderia ligar a isoforma HLA-G5 em teoria, pois a mesma possui o domínio  $\alpha$ 1 (CLEMENTS *et al.*, 2005; SHIROISHI *et al.*, 2006).

A isoforma HLA-G5 foi detectada como cadeia pesada livre de  $\beta$ 2-microglobulina em sobrenadantes de cultura de células que expressam HLA-G (JUCH *et al.*, 2005). Já foi demonstrado que a expressão de HLA-G5 solúvel poderia inibir a citotoxicidade das células NK, e que o grau de inibição da citólise das células NK por HLA-G5 era mais evidente do que o HLA-G ligado à membrana. No entanto, a inibição da citólise de NK provocada por HLA-G ou HLA-G5 pode variar de acordo com o nível de expressão das mesmas. Mais importante ainda, dados

mostraram que a combinação de HLA-G e HLA-G5 tem uma supressão significativamente maior quando comparada aos efeitos de apenas HLA-G ou HLA-G5, separadamente. Tal achado pode ser importante, pois tanto a expressão de HLA-G quanto a HLA-G5 foram induzidas na maioria dos pacientes portadores de tumores e infectados por vírus. O efeito aditivo de HLA-G e HLA-G5 nas células NK pode fornecer mecanismos mais poderosos para que as células alvo escapem das respostas imunes do hospedeiro. Considerando que HLA-G ou HLA-G5 também têm efeitos de inibição nas funções de outras células imunes, como células T, DC, monócitos e células B, o significado do efeito combinado de HLA-G e HLA-G5 sobre essas populações ainda permanece inexplorado (ZHANG *et al.*, 2014).

Além da correlação clínica, o envolvimento direto de HLA-G na indução de aceitação de enxerto *in vivo* após transplante humano foi fornecido pelas seguintes observações: *i*) as células mononucleares do sangue periférico (PBMCs) de pacientes transplantados que foram expostos a altos níveis de HLA-G circulante não respondem ao estímulo alogênico; *ii*) HLA-G5 purificado a partir de plasma de pacientes transplantados HLA-G-positivos suprime a aloproliferação de células T *in vitro* (LE ROND *et al.*, 2006).

Além disso, mesmo que a estrutura de HLA-G associada a  $\beta$ 2-microglobulina seja conhecida (CLEMENTS *et al.*, 2005) e os parâmetros estruturais de sua associação com LILRB1 e LILRB2 estejam descritos (SHIROISHI *et al.*, 2006), nada é certo quanto à estrutura das isoformas truncadas ou aos receptores aos quais se ligam.

Não existem estruturas tridimensionais publicadas para a isoforma HLA-G5 e com base em evidências presentes na literatura científica de que a isoforma HLA-G5 pode ser encontrada na forma de cadeia pesada livre ou complexada com a proteína  $\beta$ 2-microglobulina (JUCH *et al.*, 2005; MORALES *et al.*, 2007), 3 alternativas estruturais para essa isoforma foram modeladas e seu comportamento dinâmico analisado:

- HLA-G5 (monômero),
- HLA-G5 e peptídeo RIIPRHLQL acoplado ao sulco,
- HLA-G5, peptídeo RIIPRHLQL acoplado ao sulco e proteína  $\beta$ 2-microglobulina.

### 7.3.1 Monômero de HLA-G5

A modelagem do monômero de HLA-G5 foi realizada utilizando técnicas de modelagem molecular por homologia para os domínios extra globulares  $\alpha 1$ ,  $\alpha 2$  e  $\alpha 3$ , utilizando como molde a estrutura PDB\_ID: 1YDP, citada anteriormente (CLEMENTS *et al.*, 2005). O modelo gerado foi selecionado baseado em valores rigorosos de controle de qualidade, utilizando os mesmos *softwares* e parâmetros já citados anteriormente para os modelos apresentados modelados por homologia.

Sabe-se que além dos domínios extra globulares  $\alpha 1$ ,  $\alpha 2$  e  $\alpha 3$ , a isoforma HLA-G5 conserva o íntron 4, conferindo à isoforma uma cauda de 21 aminoácidos. A cauda proveniente do íntron foi modelada utilizando técnicas de modelagem *ab initio* e ao final de 3 etapas de modelagem o melhor modelo foi selecionado, de acordo com os *scores* resultantes da utilização de *softwares* para avaliação de qualidade. Essa cauda de 21 resíduos possui conformação de alça.

O trabalho apresentado mostra pela primeira vez a estrutura do monômero da isoforma HLA-G5 (**Figura 52**), modelada de acordo com a sequência do alelo *HLA-G\*01:01:01:01*. Essa estrutura foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

Os altos valores de RMSD (**Figura 53**) observados no decorrer da simulação de dinâmica molecular demonstram a instabilidade do modelo gerado e a análise das estruturas no decorrer da trajetória indicam que as variações nos valores de RMSD das triplicatas se deve principalmente à movimentação da porção referente à cauda proveniente do íntron 4. Após a análise de imagens provenientes da trajetória em intervalos definidos foi possível perceber que a alça proveniente do íntron 4 possui movimentação intensa, interagindo com os domínios  $\alpha 1$  e  $\alpha 2$  que formam o sulco. O alto valor de RMSD calculado para comparação entre as posições iniciais e finais da dinâmica (**Figura 55**), além das imagens retiradas nos momentos iniciais e finais da trajetória apoiam essas conclusões, pois foi possível observar que o sulco formado entre os domínios  $\alpha 1$  e  $\alpha 2$  sofreu desenovelamento de algumas porções das  $\alpha$ -hélices que o compõem, devido às interações com a cauda proveniente do íntron 4.

Tais observações são corroboradas pelos resultados obtidos para os valores de RMSF (**Figura 54**) da simulação mais estável, pois existe grande flutuação dos átomos em toda extensão da isoforma, atingindo os maiores valores e conseqüentemente maior movimentação na porção referente à cauda proveniente do íntron 4.

### 7.3.2 HLA-G5 e peptídeo RIIPRHLQL acoplado ao sulco

A mesma estrutura do monômero da isoforma HLA-G5 foi utilizada para o *docking* do nonapeptídeo RIIPRHLQL no sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  do monômero da isoforma HLA-G5, com intuito de avaliar se a presença do peptídeo no sulco seria capaz de conferir maior estabilidade à estrutura.

A estrutura selecionada a partir do *docking* demonstrou forte energia de ligação do nonapeptídeo RIIPRHLQL ao sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  da isoforma HLA-G5. A comparação da conformação da estrutura da isoforma com o peptídeo acoplado ao sulco e a estrutura cristalográfica PDB\_ID: 1YDP comprovou a similaridade da disposição do peptídeo em ambas estruturas, baseado no valor reduzido de RMSD.

O trabalho apresentado mostra pela primeira vez a estrutura do monômero da isoforma HLA-G5 com o nonapeptídeo RIIPRHLQL acoplado ao sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$ . Essa estrutura foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

O início das triplicatas das trajetórias de dinâmica molecular para a isoforma HLA-G5 acoplada ao nonapeptídeo RIIPRHLQL demonstraram valores de RMSD (**Figura 56**) semelhantes aos observados para a dinâmica molecular do monômero da isoforma HLA-G5, discutida anteriormente. Aos 200ns (referente à r1) e 100ns (referente à r3) foi possível observar um aumento abrupto nos valores de RMSD para essas simulações, e a investigação da trajetória nos levou a concluir que exatamente nesses momentos o peptídeo se desprende do sulco da isoforma HLA-G5, ficando livre durante o restante dessas 2 simulações. Uma explicação plausível para esse fenômeno é a grande movimentação da cauda proveniente do íntron 4, como demonstrado anteriormente. Tal movimentação extrema teria interagido com os domínios  $\alpha 1$  e  $\alpha 2$  que formam o sulco e levado à deformação do mesmo. Essa conclusão foi baseada na análise dos resíduos do sulco que fazem contato com o peptídeo no início e final das simulações, indicando a perda da interação peptídeo: sulco nessa simulação. Foi possível observar alterações nas ligações realizadas entre o peptídeo e o sulco formado pelos domínios  $\alpha 1$  e  $\alpha 2$ , com algumas interações presentes nos momentos iniciais sendo desfeitas no decorrer da dinâmica, enquanto outras interações surgiram durante o decorrer da mesma.

Como descrito para a simulação anterior referente ao monômero de HLA-G5, a instabilidade do modelo gerado e a análise das estruturas no decorrer da trajetória indicam que as variações nos valores de RMSD das triplicatas se deve principalmente à movimentação da porção referente à cauda proveniente do íntron 4. Curioso observar que essa estrutura demonstrou maior deformação do sulco, mas seus valores de RMSF (**Figura 57**) são menores quando comparados ao monômero de HLA-G5. Após a análise das imagens provenientes da trajetória em intervalos definidos foi possível perceber a grande deformação sofrida na região do sulco, com a  $\alpha$ -hélice do domínio  $\alpha 2$  apresentando regiões desenoveladas nos momentos finais da simulação.

O alto valor de RMSD calculado para comparação entre as posições iniciais e finais da dinâmica, além das imagens retiradas nos momentos iniciais e finais da trajetória (**Figura 58**) apoiam essas conclusões, sendo possível observar as deformações sofridas pelo sulco que acomoda o peptídeo.

### 7.3.3 HLA-G5 contendo peptídeo RIIPRHLQL acoplado ao sulco e acoplada à proteína $\beta 2$ -microglobulina

A isoforma HLA-G5 contendo peptídeo RIIPRHLQL e associada à  $\beta 2$ -microglobulina já teve seu papel imunomodulatório descrito em diversas publicações (ZHANG *et al.*, 2014; REBMANN *et al.*, 2007).

O modelo da isoforma HLA-G5 analisado anteriormente foi acoplado à  $\beta 2$ -microglobulina, resultando em um modelo de excelente qualidade estrutural.

O trabalho apresentado mostra pela primeira vez a estrutura do monômero da isoforma HLA-G5 contendo o nonapeptídeo RIIPRHLQL acoplado ao sulco formado entre o domínio  $\alpha 1$  e  $\alpha 2$  (**Figura 61**). Essa estrutura foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

Ao contrário do observado para as outras duas possibilidades estruturais da isoforma HLA-G5, as triplicatas das trajetórias de dinâmica molecular para a isoforma HLA-G5 contendo nonapeptídeo RIIPRHLQL e  $\beta 2$ -microglobulina demonstram valores de RMSD (**Figura 63**) muito estáveis durante toda a simulação, em todas as triplicatas. Os valores de RMSF (**Figura 64**) se encontram muito mais estáveis do que para as estruturas anteriores, mas ainda demonstrando a

maior flutuação localizada na porção terminal da estrutura, referente à cauda proveniente da tradução do íntron 4.

O valor de RMSD calculado entre as estruturas retiradas da trajetória nos momentos iniciais e finais (**Figura 65**) é muito menor para essa estrutura do que para as outras alternativas estruturais da isoforma HLA-G5, demonstrando uma maior similaridade estrutural entre o início e final da simulação.

Existe um conflito entre nossos achados e alguns publicados na literatura, como por exemplo relatos de que a isoforma HLA-G5 poderia existir sem a  $\beta$ 2-microglobulina (JUCH *et al.*, 2005), pois em nossas simulações foi demonstrada a grande instabilidade das possibilidades estruturais da isoforma HLA-G5 que não possuem o peptídeo e  $\beta$ 2-microglobulina acoplada. Uma possibilidade que poderia unificar nossos achados e esses achados contrários seria a existência de dímeros de HLA-G5 sem a  $\beta$ 2-microglobulina. Talvez nesses dímeros, mesmo sem a presença da  $\beta$ 2-microglobulina para conferir maior estabilidade, as caudas provenientes dos íntrons de cada monômero poderiam interagir entre si, não causando o efeito de desenovelamento do sulco observado em nossos ensaios sem a  $\beta$ 2-microglobulina.

Tais estruturas diméricas para HLA-G5 sem a  $\beta$ 2-microglobulina poderiam ser semelhantes aos dímeros compostos pelos domínios  $\alpha$ 1 –  $\alpha$ 3:  $\alpha$ 1 –  $\alpha$ 3, como no trabalho publicado por Kuroki e colaboradores. Nesse trabalho, os pesquisadores descobriram inesperadamente que HLA-G2 ( $\alpha$ 1 –  $\alpha$ 3 ligada à membrana) forma naturalmente um homodímero livre de  $\beta$ 2-microglobulina e que não possui pontes dissulfeto unindo os monômeros do dímero. Análises de microscopia eletrônica revelaram que a estrutura geral e a organização do domínio do homodímero HLA-G2 se assemelham aos dos heterodímeros de HLA de classe II ( $\alpha$ 1 –  $\alpha$ 2:  $\beta$ 1 –  $\beta$ 2) (KUROKI *et al.*, 2017).

Um trabalho recente (JONES *et al.*, 2011) descreve a ligação de formas de HLA associadas a  $\beta$ 2-microglobulina e livre de  $\beta$ 2-microglobulina aos membros da família ILTs e demonstra que, além de ILT4, as estruturas livres de  $\beta$ 2-microglobulina são reconhecidos por vários membros da família ILTs. Na verdade, os membros "ativadores" da família ILTs mostram preferência por essas estruturas. Portanto, é possível que seja também o caso do HLA-G. Isso apoiaria a noção de que as variações estruturais de HLA-G podem ser relevantes em modulações de função biológica (HOWANGYIN *et al.*, 2012).

Também é intrigante considerar que, de forma semelhante às moléculas clássicas HLA de classe I, HLA-G pode possuir receptores ativadores. Na verdade, é possível que existam outros receptores para HLA-G, específicos para isoformas ou não, e o estudo das estruturas de HLA-G diferentes de HLA-G1 e HLA-G5 associadas a  $\beta$ 2-microglobulina pode nos permitir identificá-los (HOWANGYIN *et al.*, 2012).

Em nossa comparação de valores de RMSD (**Figura 66**) atingidos durante a trajetória, foi possível observar de forma muito clara que dentre essas possibilidades estruturais para a isoforma HLA-G5, definitivamente a construção contendo o peptídeo localizado no sulco e a  $\beta$ 2-microglobulina acoplada foi a mais estável.

#### 7.4 Isoforma solúvel HLA-G6

A isoforma solúvel HLA-G6 é composta dos éxons 1 e 3 do gene *HLA-G*, além de conservar o íntron 4, que bloqueia a tradução do domínio transmembrana (éxon 5). A região 5' do íntron 4 é traduzida até a geração de um *stop* códon, conferindo à isoforma HLA-G6 a mesma cauda de 21 aminoácidos presente na isoforma HLA-G5 (DONADI *et al.*, 2011).

Sabe-se que HLA-G6 pode: *i*) interagir com o receptor ILT4, *ii*) modular a resposta imune de células dendríticas, macrófagos e monócitos, *iii*) prolongar a sobrevivência de aloenxertos de pele em camundongos B6 e em modelo de camundongos transgênicos para ILT4, demonstrando pela primeira vez que as isoformas de HLA-G truncadas podem ser usadas como agentes terapêuticos (HOWANGYIN *et al.*, 2012).

O fato de que HLA-G6 atua através de ILT4 é de particular interesse, pois esse receptor é expresso apenas por células da linhagem mieloide, isto é, monócitos, células dendríticas e macrófagos. Assim, ao contrário das isoformas HLA-G1 e HLA-G5 associadas à  $\beta$ 2-microglobulina, HLA-G2 e HLA-G6 atuam diretamente sobre os efetores linfocíticos, mas diretamente em APCs (HOWANGYIN *et al.*, 2012).

Existem dados relacionando a expressão da isoforma HLA-G6 e tolerância materna ao feto, e os resultados mostraram a diminuição significativa no nível de expressão da isoforma HLA-G6 em gestantes que apresentavam gravidez de risco em comparação com os controles.

Portanto, conclui-se que HLA-G6 é uma molécula marcadora preditiva de riscos relacionados à gravidez em estágios iniciais (FARZAD; ABEDIANKENARI, 2013).

Existem relatos da expressão de HLA-G5 e HLA-G6 em lesões de câncer de ovário, onde essas isoformas foram expressas em 79,7% das amostras. Os dados mostraram que a expressão de HLA-G5 e HLA-G6 na lesão não estava relacionada aos parâmetros clínicos, incluindo tipo histológico, idade e sobrevida do paciente. A diferença na expressão de HLA-G total, HLA-G5 e HLA-G6 dosados em pacientes com câncer de pulmão também foi observada, pois ao contrário da expressão total de HLA-G da lesão, que estava fortemente correlacionada ao estágio clínico e ao prognóstico ruim em pacientes com câncer de pulmão, a expressão de HLA-G5 e HLA-G6 na lesão não estava associada aos parâmetros clínicos (estágio, acometimento de vasos linfáticos e sobrevida do paciente) (ZHANG *et al.*, 2016).

Outro estudo mostrou que as isoformas HLA-G5 e HLA-G6 estavam predominantemente expressas em adenocarcinoma pulmonar, podendo representar biomarcador útil para discriminar adenocarcinoma de carcinoma de células escamosas em pacientes (YAN *et al.*, 2015).

Não existem estruturas tridimensionais publicada para a isoforma HLA-G6 e para avaliar o comportamento dinâmico de algumas das possibilidades conformacionais, foram modeladas 3 alternativas estruturais para essa isoforma:

- HLA-G6 (monômero),
- HLA-G6 e peptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$ ,
- HLA-G6 e peptídeo RIIPRHLQ, acoplados à proteína  $\beta 2$ -microglobulina.

#### 7.4.1 Monômero de HLA-G6

A modelagem do monômero de HLA-G6 foi realizada da mesma forma como exposto anteriormente para isoforma HLA-G5, utilizando técnicas de modelagem molecular por homologia para os domínios extraglobulares  $\alpha 1$  e  $\alpha 3$ , utilizando como molde a estrutura PDB\_ID: 1YDP, citada anteriormente (CLEMENTS *et al.*, 2005). O modelo gerado foi selecionado baseado em valores rigorosos de controle de qualidade, utilizando os mesmos *softwares* e parâmetros já citados anteriormente para os modelos apresentados modelados por homologia.

Da mesma maneira que a isoforma HLA-G5, HLA-G6 conserva o íntron 4 e possui a cauda de 21 aminoácidos proveniente do mesmo, que assume conformação de alça, modelada utilizando técnicas de modelagem *ab initio*. Ao final de 3 etapas de modelagem o melhor modelo foi selecionado, de acordo com os *scores* resultantes da utilização de *softwares* para avaliação de qualidade (**Figura 70**).

O trabalho apresentado mostra pela primeira vez a estrutura do monômero da isoforma HLA-G6, modelada de acordo com a sequência do alelo *HLA-G\*01:01:01:01*. Essa estrutura foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

No decorrer da simulação de dinâmica molecular 2 replicatas (r1 e r2) assumem comportamento relativamente estável, quando comparadas à replicata r3, que após 50ns de simulação atinge níveis altíssimos de RMSD (**Figura 71**). Novamente, a alça proveniente da tradução do íntron é a grande responsável pela desestabilização do modelo, interagindo com os domínios  $\alpha 1$  e  $\alpha 3$ , desestabilizando sua estrutura como um todo. Os dados de RMSF (**Figura 72**) demonstraram que a flutuação estrutural esteve mais concentrada no domínio  $\alpha 1$  e na alça proveniente da tradução do íntron.

As imagens retiradas nos momentos iniciais e finais da simulação (**Figura 73**) demonstram que a interação da alça proveniente da tradução do íntron com o domínio  $\alpha 1$  foi capaz de levar ao desenovelamento parcial da  $\alpha$ -hélice encontrada nessa região.

O valor de RMSD calculado entre as estruturas retiradas da trajetória nos momentos iniciais e finais é semelhante aos valores obtidos anteriormente para estruturas instáveis, como demonstrado para o monômero da isoforma HLA-G5, contendo ou não o peptídeo acoplado ao sulco.

#### **7.4.2 Isoforma solúvel HLA-G6 contendo peptídeo acoplado ao domínio $\alpha 1$**

A mesma estrutura do monômero da isoforma HLA-G6 foi utilizada para o *docking* do nonapeptídeo RIIPRHLQL ao domínio  $\alpha 1$  do monômero da isoforma HLA-G6, com intuito de avaliar se a presença do peptídeo no sulco seria capaz de conferir maior estabilidade à estrutura.

A estrutura selecionada a partir do *docking* demonstrou forte energia de ligação do nonapeptídeo RIIPRHLQL ao domínio  $\alpha 1$ .

O trabalho apresentado mostra pela primeira vez a estrutura do monômero da isoforma HLA-G6 com o nonapeptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  (**Figura 74**). Essa estrutura foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

Foi possível observar que mesmo durante as simulações mais estáveis (r1 e r2), os valores de RMSD atingidos foram altos (**Figura 75**), o que se correlaciona com a instabilidade da estrutura. A estrutura referente à replicata r3 perde os contatos necessários com o peptídeo em 200ns da simulação, momento em que o mesmo escapa do domínio  $\alpha 1$  e leva ao aumento brusco nos valores de RMSD. Além disso, os valores de RMSF (**Figura 76**) demonstraram variação estrutural muito grande em toda extensão da isoforma HLA-G6 e do peptídeo, ao contrário do que foi observado para o monômero de HLA-G6, onde a flutuação dos resíduos estava mais concentrada nos resíduos que compõem o domínio  $\alpha 1$ .

A comparação das imagens iniciais e finais da trajetória demonstra extensas deformações em toda isoforma HLA-G6 (**Figura 77**). Interessante notar que existe uma inversão da  $\alpha$ -hélice e folhas- $\beta$  que compõem o domínio  $\alpha 1$  na imagem da replicata demonstrada na seção de Resultados. Tais alterações estruturais no decorrer da simulação estão refletidas no altíssimo valor de RMSD calculado entre a conformação inicial e final da trajetória.

### **7.4.3 Isoforma solúvel HLA-G6 contendo peptídeo acoplado ao domínio $\alpha 1$ e acoplada à $\beta 2$ -microglobulina**

Existem relatos na literatura que indicam a falta do domínio  $\alpha 2$  como razão para impossibilidade da presença da  $\beta 2$ -microglobulina nas isoformas compostas pelos domínios  $\alpha 1$  -  $\alpha 3$ , como a isoforma HLA-G6 (TYSOE-CALNON *et al.*, 1991). Mas novamente, nenhuma estrutura existe para essa ou qualquer uma das outras isoformas, e com base em nossos resultados anteriores e alta instabilidade das construções de HLA-G6 sem a  $\beta 2$ -microglobulina, foi realizada a montagem de uma estrutura de HLA-G6 contendo peptídeo e  $\beta 2$ -microglobulina acoplada.

A geração da isoforma HLA-G6, contendo peptídeo acoplado ao domínio  $\alpha 1$  e  $\beta 2$ -microglobulina não foi possível por meio de modelagem molecular, pois o *software* Modeller não

foi capaz de fazer o encaixe da proteína  $\beta$ 2-microglobulina na ausência do domínio  $\alpha$ 2. Acreditamos que essa impossibilidade de realizar o encaixe se deve à inclinação do domínio  $\alpha$ 1, que aumenta a distância entre o domínio  $\alpha$ 3 e  $\beta$ 2-microglobulina.

Tomando como base a estrutura da isoforma HLA-G5 acoplada à  $\beta$ 2-microglobulina e as posições de interação encontradas nessas estruturas, foi possível obter o *docking* da isoforma HLA-G6 contendo peptídeo acoplado e  $\beta$ 2-microglobulina. A comparação dos resíduos que fazem parte da interação HLA-G6 e  $\beta$ 2-microglobulina demonstra um número menor de resíduos do que os encontrados na interface de interação de HLA-G5 e  $\beta$ 2-microglobulina, mesmo levando em conta a ausência do domínio  $\alpha$ 2 na isoforma HLA-G6 (**Figura 78**).

As simulações de dinâmica molecular para o sistema que contém a HLA-G6 com o nonapeptídeo RIIPRHLQL e acoplada à proteína  $\beta$ 2-microglobulina atingiram o tempo total de aproximadamente 60ns (**Figura 81**). Ao atingir esse ponto da trajetória o *software* utilizado para as simulações de dinâmica indicava a grande instabilidade do sistema e a impossibilidade de dar continuidade ao ensaio. Todas as replicatas das simulações de HLA-G6 atingiram esse ponto de instabilidade e foram terminadas abruptamente.

Os valores de RMSD para as replicatas r1 e r2 foram relativamente estáveis se comparados com as simulações anteriores para as outras possibilidades estruturais de HLA-G6. A visualização dos momentos iniciais e finais da trajetória (**Figura 82**) demonstrou torção da  $\alpha$ -hélice que compõem o domínio  $\alpha$ 1, mas as outras alterações estruturais foram bem menos agressivas do que as alterações vistas nas simulações anteriores para as outras possibilidades estruturais de HLA-G6.

Acreditamos que a impossibilidade de dar prosseguimento às simulações de dinâmica molecular se deve ao posicionamento da  $\beta$ 2-microglobulina, seu distanciamento do domínio  $\alpha$ 3 e conseqüentemente à falta de interações imprescindíveis entre os domínios  $\alpha$ 1 -  $\alpha$ 3 e  $\beta$ 2-microglobulina. Em nossa comparação de valores de RMSD (**Figura 83**) atingidos durante as trajetórias de HLA-G6 foi possível observar que dentre as possibilidades estruturais para essa isoforma, a construção mais estável foi o monômero de HLA-G6, pois mesmo a isoforma HLA-G6 contendo peptídeo e  $\beta$ 2-microglobulina atingindo valores próximos de RMSD, o sistema de dinâmica molecular da mesma se mostrou tão instável que não foi possível dar continuidade às simulações.

Da mesma forma como discutido para as isoformas HLA-G5 sem a  $\beta$ 2-microglobulina, poderíamos discutir nossos resultados para os monômeros de HLA-G6. Como nenhuma de nossas possibilidades estruturais apresentadas foram estáveis quando submetidas à dinâmica molecular, uma possível conformação estrutural estável para essas estruturas solúveis HLA-G6 que possuem os domínios  $\alpha$ 1 e  $\alpha$ 3 poderiam ser dímeros compostos por  $\alpha$ 1 –  $\alpha$ 3:  $\alpha$ 1 –  $\alpha$ 3, como no trabalho publicado por Kuroki e colaboradores (KUROKI *et al.*, 2017), onde HLA-G2 ( $\alpha$ 1 –  $\alpha$ 3 de membrana) forma naturalmente um homodímero livre de  $\beta$ 2-microglobulina e que não possui pontes dissulfeto unindo os monômeros do dímero, se assemelhando à disposição dos heterodímeros de HLA de classe II ( $\alpha$ 1 –  $\alpha$ 2:  $\beta$ 1 –  $\beta$ 2), ideia proposta por Ishitani e colaboradores em 1992 (ISHITANI *et al.*, 1992). Pode-se pensar que talvez a isoforma HLA-G6 não possa existir como monômero solúvel e precise estar em uma conformação dimérica para ser expressa de forma estável.

## 7.5 Isoforma solúvel HLA-G7

A isoforma solúvel HLA-G7 é composta dos éxons 1 e 2 do gene *HLA-G*, que codificam na isoforma madura apenas o domínio  $\alpha$ 1 ligado a 2 aminoácidos provenientes da tradução do íntron 2, que é retido na transcrição pelo processo de edição alternativa (*splicing* alternativo).

Paul e colaboradores identificaram a isoforma HLA-G7 e transcritos de HLA-G7 são detectados em tecidos humanos e sua distribuição corresponde ao de outros transcritos de HLA-G, com altos níveis sendo expressos em células de trofoblasto, placenta, e tecidos amnióticos (PAUL *et al.*, 2000). Praticamente não existem relatos sobre essa isoforma na literatura, mas um trabalho correlacionou a expressão de HLA-G5 e HLA-G7 em mulheres grávidas e complicações durante a gravidez (ABEDIANKENARI *et al.*, 2015).

### 7.5.1 Monômero de HLA-G7

A modelagem do monômero de HLA-G7 foi realizada da mesma forma como exposto anteriormente para as isoformas HLA-G5 e HLA-G6, utilizando técnicas de modelagem molecular por homologia para o domínio extra globulares  $\alpha$ 1, utilizando como molde a estrutura PDB\_ID:

1YDP, citada anteriormente (CLEMENTS *et al.*, 2005). O modelo gerado foi selecionado baseado em valores rigorosos de controle de qualidade, utilizando os mesmos *softwares* e parâmetros já citados anteriormente para os modelos apresentados modelados por homologia.

A pequena porção formada por 2 resíduos provenientes do íntron 2 foi modelada utilizando técnicas de modelagem *ab initio*. Apenas 1 etapa de modelagem foi suficiente e o melhor modelo foi selecionado, de acordo com os *scores* resultantes da utilização de *softwares* para avaliação de qualidade.

O trabalho apresentado mostra pela primeira vez a estrutura do monômero da isoforma HLA-G7 (**Figura 86**), modelada de acordo com a sequência do alelo *HLA-G\*01:01:01:01*. Essa estrutura foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

No decorrer da simulação de dinâmica molecular 2 replicatas (r1 e r3) assumem comportamento estável durante os 200ns de simulação (**Figura 87**). A replicata r3 apresentou um aumento gradual nos valores de RMSD, o que levou à extensão do tempo de simulação até 500ns. Foi possível observar que após 200ns de simulação, a mesma manteve-se estável. Acredita-se que os maiores valores de RMSD se devem à oscilação dos resíduos iniciais e finais que compõem a estrutura.

Os dados de RMSF (**Figura 88**) demonstraram que a flutuação estrutural esteve mais concentrada nos resíduos iniciais que compõem o domínio  $\alpha 1$  e nos 2 resíduos finais provenientes da tradução do íntron 2. As imagens retiradas nos momentos iniciais e finais da simulação demonstram essa é uma estrutura extremamente estável, mantendo todos os componentes de estrutura secundária enovelados.

O valor de RMSD calculado entre as estruturas retiradas da trajetória nos momentos iniciais e finais (**Figura 89**) é condizente com o comportamento estável observado pelo monômero da isoforma HLA-G7 durante toda trajetória.

### 7.5.2 Monômero de HLA-G7 acoplado ao peptídeo RIIPRHLQL

Da mesma forma como idealizado para os experimentos das isoformas HLA-G5 e HLA-G6, foi realizada a construção da estrutura da isoforma HLA-G7 contendo peptídeo RIIPRHLQL

acoplado ao domínio  $\alpha 1$ . A estrutura selecionada a partir do *docking* demonstrou forte energia de ligação do nonapeptídeo RIIPRHLQL ao domínio  $\alpha 1$ .

O trabalho apresentado mostra pela primeira vez a estrutura do monômero da isoforma HLA-G7, modelada de acordo com a sequência do alelo *HLA-G\*01:01:01:01* e contendo peptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  (**Figura 90**). A estrutura foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

No decorrer das simulações de dinâmica molecular (**Figura 91**) foi possível observar que essa construção da isoforma HLA-G7 contendo o peptídeo demonstrou ser muito estável e de forma geral, mais estável do que a mesma na ausência do peptídeo. Os dados de RMSF (**Figura 92**) corroboram as afirmações realizadas para os valores de RMSD, pois as variações demonstraram que a flutuação estrutural não esteve localizada em um ponto específico da estrutura, estando distribuída de forma homogênea por toda extensão da mesma.

As imagens retiradas nos momentos iniciais e finais da simulação (**Figura 93**) demonstram essa é uma estrutura extremamente estável, mantendo todos os componentes de estrutura secundária enovelados, mesmo que seja possível observar grande flexibilidade da estrutura de forma a manter o peptídeo acoplado no sulco. Foi possível observar alterações nas ligações realizadas entre o peptídeo e o sulco, com algumas interações presentes nos momentos iniciais sendo desfeitas no decorrer da dinâmica, enquanto outras interações surgiram durante o decorrer da mesma.

O valor de RMSD calculado entre as estruturas retiradas da trajetória nos momentos iniciais e finais é condizente com o comportamento estável observado pelo monômero da isoforma HLA-G7 contendo o peptídeo durante toda trajetória.

## 7.6 Isoforma solúvel HLA-G7: possíveis dímeros formados

Como citado anteriormente, a capacidade da proteína HLA-G de formar dímeros é uma de suas principais características, sendo que os dímeros demonstrados até o momento são capazes de se ligar aos receptores de HLA-G, demonstrando maior afinidade e menores taxas de dissociação, quando comparados com monômeros.

A forma de dimerização mais estudada até o momento é realizada por pontes dissulfeto, devido à presença do resíduo de cisteína localizado na posição 42 (Cys42) do domínio  $\alpha 1$  (SHIROISHI *et al.*, 2006; HOVANGWYIN *et al.*, 2012). De fato, uma vez que todas as isoformas HLA-G possuem o resíduo Cys42, todas elas deveriam ser capazes de dimerizar, como foi mostrado para HLA-G (BOYSON *et al.*, 2002).

Estudos com enfoque na dimerização das demais isoformas de HLA-G são raros, mas Howangyin e colaboradores demonstraram a dimerização das isoformas HLA-G1 a HLA-G6, utilizando células M8 transfectadas com essas isoformas. Os dímeros de isoformas HLA-G foram facilmente observados para HLA-G1 e HLA-G5 (domínios  $\alpha 1$ - $\alpha 2$ - $\alpha 3$ ) e HLA-G4 (domínios  $\alpha 1$ - $\alpha 2$ ). Esta foi a primeira descrição dos homodímeros HLA-G4 (HOWANGYIN *et al.*, 2012).

Além disso, os dados de Howangyin e colaboradores mostram claramente que as isoformas de HLA-G apresentam variações estruturais importantes, que podem corresponder a diferentes funções biológicas. Como mencionado anteriormente, HLA-G1 e HLA-G5 ( $\alpha 1$ - $\alpha 2$ - $\alpha 3$  associados a  $\beta 2$ -microglobulina) são as isoformas mais extensamente estudadas, provavelmente por causa de sua maior abundância alegada, sua disponibilidade para experimentos *in vitro* e a existência de anticorpos específicos dirigidos contra eles. Assim, acredita-se que essas isoformas sejam as principais isoformas responsáveis pela regulação imune *in vivo*.

No entanto, não sabemos em que medida isso é verdade. Isoformas e estruturas HLA-G sem  $\beta 2$ -microglobulina também podem desempenhar um papel específico. Não saber a estrutura precisa das isoformas truncadas HLA-G implica que também não sabemos a quais receptores se ligam (HOWANGYIN *et al.*, 2012).

### 7.6.1 Dímero da isoforma solúvel HLA-G7 (G7-Cys42)

Partindo do princípio de que todas as 7 isoformas codificadas pelo gene HLA-G possuem a cisteína na posição 42 o domínio  $\alpha 1$ , foi realizada a construção de um dímero formado pelo monômero da isoforma HLA-G7 e ligado por ponte dissulfeto utilizando a cisteína na posição 42. Essa construção dimérica foi designada G7-Cys42 (**Figura 95**).

A estrutura foi selecionada com base nos valores excelentes obtidos a partir do ensaio de *docking*, para todas as *softwares* utilizados na verificação da qualidade. Os monômeros que

compõem o dímero G7-Cys42 apresentam orientação perpendicular, com os domínios  $\alpha 1$  de cada monômero expostos em uma direção.

O trabalho apresentado mostra pela primeira vez uma possibilidade conformacional dimérica para o monômero da isoforma HLA-G7, modelada de acordo com a sequência do alelo *HLA-G\*01:01:01:01*. A estrutura obtida foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

No decorrer da simulação de dinâmica molecular foi possível observar que dímero G7-Cys42 de forma geral, demonstrou ser estável, mesmo que os valores de RMSD obtidos fossem mais elevados do que quando comparados aos valores das estruturas monoméricas de HLA-G7 (**Figura 96**). Acreditamos que esses valores de RMSD foram mais elevados devido ao movimento de rotação dos monômeros do dímero, que estão dispostos perpendicularmente, pois não foi possível observar desenovelamento em nenhuma porção do dímero. Os dados de RMSF (**Figura 97**) indicam que as variações estruturais estiveram distribuídas por toda extensão da mesma, com exceção dos resíduos iniciais que apresentaram flutuação mais acentuada.

As imagens retiradas nos momentos iniciais e finais da simulação (**Figura 98**) demonstraram novamente que essa é uma estrutura estável, mesmo que seja possível observar rotação dos monômeros que compõem o dímero.

O valor de RMSD calculado entre as estruturas retiradas da trajetória nos momentos iniciais e finais é condizente com o comportamento estável observado pelo dímero G7-Cys42 durante toda trajetória.

### 7.6.2 Dímero alternativo da isoforma solúvel HLA-G7 (*G7 fit 2D31*)

Foi realizada uma construção alternativa para o dímero de HLA-G7 ligado por ponte dissulfeto utilizando as cisteínas da posição 42 contendo o peptídeo RIIPRHLQL acoplado ao domínio  $\alpha 1$  de cada monômero, denominado *G7 fit 2D31*. Esse dímero alternativo teve os monômeros que foram o dímero orientados de forma a manterem a conformação apresentada pelo dímero de HLA-G solúvel (PDB\_ID: 2D31, SHIROISHI *et al.*, 2006) e modelo completo do dímero solúvel de HLA-G gerado durante o trabalho, demonstrado anteriormente.

Nossos resultados mostraram pela primeira vez uma possibilidade conformacional dimérica alternativa para o dímero formado pela isoforma HLA-G7, modelada de acordo com a sequência do alelo *HLA-G\*01:01:01:01* (**Figura 100**). A estrutura obtida foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

No decorrer da simulação de dinâmica molecular foi possível observar que o dímero G7 *fit* 2D31 de forma geral, demonstrou ser ligeiramente mais instável do que o dímero G7-Cys42, apresentando valores de RMSD mais elevados (**Figura 102**). Acreditamos que esses valores de RMSD foram mais elevados devido à presença do peptídeo na fenda, que nesse caso parece ter sido um fator de desestabilização para essa construção alternativa. Os dados de RMSF (**Figura 104**) indicam que as variações estruturais estiveram distribuídas por toda extensão da mesma, com exceção dos resíduos iniciais da estrutura e dos resíduos referentes ao peptídeo, que apresentaram flutuação mais acentuada.

As imagens retiradas nos momentos iniciais e finais da simulação (**Figura 103**) demonstraram deformação em uma das  $\alpha$ -hélices do dímero, além de grande movimentação dos peptídeos acoplados a esse dímero. A movimentação dos peptídeos já era esperada, como visto anteriormente para a estrutura do monômero de HLA-G7 contendo peptídeo acoplado ao domínio  $\alpha 1$ .

O valor de RMSD calculado entre as estruturas retiradas da trajetória nos momentos iniciais e finais reflete essas diferenças estruturais citadas entre os momentos iniciais e finais da simulação.

### 7.6.3 Dímero da isoforma solúvel HLA-G7 (G7 MHC-like)

Baseados nos dados já citados de que HLA-G2 forma naturalmente um homodímero livre de  $\beta 2$ -microglobulina e esse não possui pontes dissulfeto para união dos monômeros do dímero (KUROKI *et al.*, 2017), o mesmo raciocínio foi aplicado a uma possível conformação dimérica de HLA-G7 sem a participação de pontes dissulfeto (**Figura 105**).

Nessa estrutura, temos as cisteínas da posição 42 a 37,9Å de distância e a estrutura obtida se assemelha ao sulco de ligação ao peptídeo formado pelos domínios  $\alpha 1$  e  $\alpha 2$  de uma molécula de HLA de classe I, então surgiu a denominação dímero MHC-like (**Figura 106**).

Mostramos aqui pela primeira vez a estrutura de um possível dímero de HLA-G7, chamado MHC-*like*, formado pela união das folhas- $\beta$  que formam o assoalho do sulco de moléculas de HLA de classe I. Esse dímero é resultante da união de 2 monômeros da isoforma HLA-G7 e sua estrutura foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

No decorrer da simulação de dinâmica molecular foi possível observar que essa construção do dímero MHC-*like* demonstrou ser muito estável, com valores baixos de RMSD similares em todas as 3 simulações (**Figura 107**). Os dados de RMSF (**Figura 108**) demonstram que apesar de valores de RMSD baixos, algumas variações pontuais foram observadas, como movimentação das  $\alpha$ -hélices que formam o sulco e pequeno desenovelamento na porção central de uma das  $\alpha$ -hélices.

As imagens retiradas nos momentos iniciais e finais da simulação (**Figura 109**) demonstraram que essa é uma estrutura estável, que mantém praticamente todos os componentes de estrutura secundária enovelados, com exceção da porção central de uma das  $\alpha$ -hélices, como dito anteriormente.

O valor de RMSD calculado entre as estruturas retiradas da trajetória nos momentos iniciais e finais é condizente com o comportamento estável observado pelo dímero MHC-*like* durante toda trajetória.

#### **7.6.4 Dímero da isoforma solúvel HLA-G7 (G7 MHC-*like* contendo nonapeptídeo RIIPRHLQL)**

A estrutura discutida anteriormente para o dímero MHC-*like* recebeu o peptídeo RIIPRHLQL em seu sulco por meio de ensaio *in silico* de *docking* molecular. A estrutura selecionada após o *docking* possuía a melhor energia de ligação entre o peptídeo e o sulco do dímero MHC-*like*.

Novamente, demonstramos aqui pela primeira vez a estrutura do dímero MHC-*like*, formado pela união das folhas- $\beta$  que formam o assoalho do sulco de moléculas de HLA de classe I com peptídeo RIIPRHLQL em seu sulco (**Figura 110**). Sua estrutura foi submetida à simulação de dinâmica molecular, afim de examinar a estabilidade do modelo.

No decorrer da simulação de dinâmica molecular foi possível observar que essa construção do dímero *MHC-like* demonstrou-se muito estável, com valores baixos de RMSD similares em todas as 3 simulações (**Figura 111**). Os dados de RMSF (**Figura 112**) demonstraram flutuações muito pequenas disseminadas por toda estrutura do dímero e peptídeo acoplado, corroborando as afirmações feita sobre a estabilidade dessa estrutura. A comparação dos valores de RMSD e RMSF entre essa construção dimérica contendo peptídeo e a demonstrada anteriormente sem o peptídeo demonstra que nesse caso a inserção do peptídeo no sulco trouxe maior estabilidade para a estrutura.

As imagens retiradas nos momentos iniciais e finais da simulação (**Figura 113**) demonstram essa é uma estrutura extremamente estável, que mantém todos os componentes de estrutura secundária enovelados, inclusive a porção central de uma das  $\alpha$ -hélices que tinha sofrido pequeno desenovelamento durante a simulação de dinâmica molecular desse dímero sem o peptídeo acoplado ao sulco.

O valor de RMSD calculado entre as estruturas retiradas da trajetória nos momentos iniciais e finais é condizente com o comportamento estável observado pelo dímero *MHC-like* durante toda trajetória, sendo possível perceber movimentação no peptídeo acoplado ao sulco.

Em nossa comparação de valores de RMSD atingidos durante as trajetórias, referentes aos dímeros formados pelos monômeros de HLA-G7 (**Figura 114**), foi possível observar de forma muito clara que dentre essas possibilidades estruturais definitivamente a construção *MHC-like* contendo o peptídeo localizado no sulco foi a mais estável.

Com todas as estruturas e informações expostas, nos parece óbvio que a extensão total da diversidade estrutural de HLA-G continua desconhecida e que a variação estrutural de HLA-G e suas isoformas pode desempenhar papel fundamental na sua função biológica e regulação do sistema imune.

## 8 RESUMO DOS RESULTADOS

O trabalho mostra de forma inédita até o momento os modelos completos para a molécula codificada pelo alelo *HLA-G\*01:01:01:01*, dímero de HLA-G solúvel, isoformas solúveis HLA-G5, HLA-G6, HLA-G7 e possíveis dímeros de HLA-G7 que foram obtidos por meio de modelagem molecular por homologia e modelagem molecular *ab initio*.

As simulações de dinâmica molecular mostraram a estabilidade da molécula HLA-G completa, contendo peptídeo no sulco e acoplada à  $\beta$ 2-microglobulina inserida em uma bicamada lipídica, onde as estruturas modeladas foram conservadas durante toda a simulação. Tal comportamento dinâmico confirmou a qualidade da modelagem realizada e demonstrou a movimentação da molécula na superfície da bicamada lipídica.

A avaliação do comportamento dinâmico do dímero de HLA-G demonstrou que a conformação oblíqua do dímero possui extensa movimentação e poderia favorecer a ligação dos receptores ILTs, dependendo da orientação que a estrutura apresentava no momento. As observações são corroboradas pelos resultados obtidos para análises de RMSD e RMSF da simulação, que indicaram maior flutuação dos resíduos referente às porções terminais que compreendem o sítio de ligação para receptores inibitórios ILT2 e ILT4. Novamente, a simulação confirmou a qualidade da modelagem realizada.

As simulações referentes à isoforma HLA-G5 e suas alternativas estruturais (HLA-G5 (monômero), HLA-G5 e peptídeo RIIPRHLQL acoplado ao sulco e HLA-G5, peptídeo RIIPRHLQL acoplado ao sulco e proteína  $\beta$ 2-microglobulina) foram analisadas utilizando simulações de dinâmica molecular e foi demonstrada a grande instabilidade das estruturas que não possuem o peptídeo e  $\beta$ 2-microglobulina acoplada. Com base em nossos achados, tais estruturas livres de  $\beta$ 2-microglobulina são instáveis na forma de monômeros e estudos estruturais mais abrangentes seriam necessários para verificação de sua existência em outras conformações estruturais.

Como nenhuma das possibilidades estruturais para a isoforma HLA-G6 apresentadas foram estáveis quando submetidas à dinâmica molecular, talvez a isoforma HLA-G6 não possa existir como monômero solúvel e precise estar em uma conformação dimérica para ser expressa de forma estável. Os dados da literatura que indicavam a impossibilidade da isoforma HLA-G6 estar acoplada à  $\beta$ 2-microglobulina foram confirmados em nossos estudos.

Para as duas possibilidades estruturais avaliadas do monômero da isoforma HLA-G7 foi possível observar que o monômero contendo peptídeo demonstrou ser muito estável e de forma geral, mais estável do que a mesma construção na ausência do peptídeo.

Como todas as 7 isoformas codificadas pelo gene HLA-G possuem a cisteína na posição 42 o domínio  $\alpha 1$ , foi realizada a construção de um dímero formado pelo monômero da isoforma HLA-G7 e ligado por ponte dissulfeto utilizando a cisteína na posição 42, designado G7-Cys42. As simulações de dinâmica molecular demonstraram a estabilidade dessa estrutura, mesmo que os valores de RMSD obtidos fossem mais elevados do que quando comparados aos valores das estruturas monoméricas de HLA-G7. Acreditamos que esses valores de RMSD foram mais elevados devido ao movimento de rotação dos monômeros do dímero, que estão dispostos perpendicularmente, pois não foi possível observar desenovelamento ou qualquer outra alteração estrutural em nenhuma porção do dímero.

A construção alternativa para o dímero de HLA-G7 ligado por ponte dissulfeto denominado G7 *fit* 2D31 demonstrou ser ligeiramente mais instável do que o dímero G7-Cys42, apresentando valores de RMSD mais elevados. Acreditamos que esses valores de RMSD foram mais elevados devido à presença do peptídeo na fenda, que nesse caso parece ter sido um fator de desestabilização para essa construção alternativa.

A conformação dimérica de HLA-G7 sem a participação de pontes dissulfeto obtida se assemelha ao sulco de ligação do peptídeo formado pelos domínios  $\alpha 1$  e  $\alpha 2$  de uma molécula de HLA de classe I, denominada dímero MHC-*like*. Tal estrutura foi avaliada por simulações de dinâmica molecular, contendo ou não o peptídeo acoplado ao sulco, sendo possível constatar que a estrutura que contém o peptídeo é extremamente estável, sendo considerada a possibilidade dimérica de HLA-G7 mais estável apresentada no trabalho.

## 9 CONCLUSÃO

Este trabalho contém os primeiros modelos tridimensionais completos para a molécula codificada pelo alelo *HLA-G\*01:01:01:01*, dímero de HLA-G solúvel, isoformas solúveis HLA-G5, HLA-G6, HLA-G7 e possíveis dímeros de HLA-G7, obtidos pela combinação de metodologias de modelagem molecular por homologia, modelagem molecular *ab initio* e *docking* molecular. A estabilidade dos modelos foi avaliada por simulações *in silico* de dinâmica molecular, confirmando a qualidade dos modelos gerados.

Os modelos completos gerados poderão ser utilizados como molde para modelagem *in silico* dos demais alelos e isoformas produzidas, além da realização de *screenings* virtuais para busca de possíveis receptores do sistema imune e moléculas que ainda não possuem interação com HLA-G descrita.

## 10 REFERÊNCIAS BIBLIOGRÁFICAS

ABBAS, A.; LICHTMAN, A.; PILLAI, S. Cellular and Molecular Immunology. 8ª Edição. Filadélfia: Editora Elsevier Saunders, 2015.

ABEDIANKENARI, S.; FARZAD, F.; RAHMANI, Z.; HASHEMI-SOTEH, M. B. HLA-G5 and G7 Isoforms in Pregnant Women. **Iranian Journal of Allergy, Asthma, and Immunology**, v. 14, n. 2, p. 217–221, 2015.

ADAMS, E. J.; LUOMA, A. M. The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules. **Annual Review of Immunology**, v. 31, p. 529–561, 2013.

ADCOCK, S. A.; MCCAMMON, J. A. Molecular dynamics: survey of methods for simulating the activity of proteins. **Chemical Reviews**, v. 106, n. 5, p. 1589–1615, 2006.

AIRAS, L.; NIKULA, T.; HUANG, Y.-H.; LAHESMAA, R.; WIENDL, H. Postpartum-activation of multiple sclerosis is associated with down-regulation of tolerogenic HLA-G. **Journal of Neuroimmunology**, v. 187, n. 1–2, p. 205–211, 2007.

ALBERTS, B.; JOHNSON, A.; LEWIS, J.; *et al.* Molecular Biology of the Cell. 6ª Edição. Editora Garland Science, 2014.

ALEGRE, E.; RIZZO, R.; BORTOLOTTI, D.; *et al.* Some basic aspects of HLA-G biology. **Journal of Immunology Research**, v. 2014, p. 657625, 2014.

APPS, R.; GARDNER, L.; MOFFETT, A. A critical look at HLA-G. **Trends in Immunology**, v. 29, n. 7, p. 313–321, 2008.

ARSTILA, T. P.; CASROUGE, A.; BARON, V.; *et al.* A direct estimate of the human alphabeta T cell receptor diversity. **Science (New York, N.Y.)**, v. 286, n. 5441, p. 958–961, 1999.

BARASH, Yoseph; CALARCO, John A; GAO, Weijun; *et al.* Deciphering the splicing code. **Nature**, v. 465, n. 7294, p. 53–59, 2010.

BENKERT, P.; TOSATTO, S. C. E.; SCHOMBURG, D. QMEAN: A comprehensive scoring function for model quality assessment. **Proteins**, v. 71, n. 1, p. 261–277, 2008.

BERMAN, H. M.; BATTISTUZ, T.; BHAT, T. N.; *et al.* The Protein Data Bank. **Acta Crystallographica. Section D, Biological Crystallography**, v. 58, n. Pt 6 No 1, p. 899–907, 2002.

BHARADWAJ, M.; ILLING, P.; THEODOSSIS, A.; *et al.* Drug hypersensitivity and human leukocyte antigens of the major histocompatibility complex. **Annual Review of Pharmacology and Toxicology**, v. 52, p. 401–431, 2012.

BIRNBAUM, M. E.; MENDOZA, J. L.; SETHI, D. K.; *et al.* Deconstructing the peptide-MHC specificity of T cell recognition. **Cell**, v. 157, n. 5, p. 1073–1087, 2014.

BJORKMAN, P. J.; PARHAM, AND P. Structure, Function, and Diversity of Class I Major Histocompatibility Complex Molecules. **Annual Review of Biochemistry**, v. 59, n. 1, p. 253–288, 1990.

BLUM, J. S.; WEARSCH, P. A.; CRESSWELL, P. Pathways of antigen processing. **Annual Review of Immunology**, v. 31, p. 443–473, 2013.

BOON, T.; COULIE, P. G.; VAN DEN EYNDE, B. J.; VAN DER BRUGGEN, P. Human T cell responses against melanoma. **Annual Review of Immunology**, v. 24, p. 175–208, 2006.

BOYSON, J. E.; ERSKINE, R.; WHITMAN, M. C.; *et al.* Disulfide bond-mediated dimerization of HLA-G on the cell surface. **Proceedings of the National Academy of Sciences of the United States of America**, v. 99, n. 25, p. 16180–16185, 2002.

BRIDGEMAN, J. S.; SEWELL, A. K.; MILES, J. J.; PRICE, D. A.; COLE, D. K. Structural and biophysical determinants of  $\alpha\beta$  T-cell antigen recognition. **Immunology**, v. 135, n. 1, p. 9–18, 2012.

CAMPBELL, E. C.; ANTONIOU, A. N.; POWIS, S. J. The multi-faceted nature of HLA class I dimer molecules. **Immunology**, v. 136, n. 4, p. 380–384, 2012.

CAPPS, G. G.; PINE, S.; EDIDIN, M.; ZÚÑIGA, M. C. Short Class I Major Histocompatibility Complex Cytoplasmic Tails Differing in Charge Detect Arbiters of Lateral Diffusion in the Plasma Membrane. **Biophysical Journal**, v. 86, n. 5, p. 2896–2909, 2004.

CAROSELLA, E. D.; MOREAU, P.; LEMAULT, J.; ROUAS-FREISS, N. HLA-G: from biology to clinical benefits. **Trends in Immunology**, v. 29, n. 3, p. 125–132, 2008.

CAROSELLA, Edgardo D; MOREAU, Philippe; LE MAOULT, Joël; *et al.* HLA-G molecules: from maternal-fetal tolerance to tissue acceptance. **Advances in immunology**, v. 81, p. 199–252, 2003.

CARPENTER, E. P.; BEIS, K.; CAMERON, A. D.; IWATA, S. Overcoming the challenges of membrane protein crystallography. **Current Opinion in Structural Biology**, v. 18, n. 5, p. 581–586, 2008.

CASTELLI, E C; MENDES-JUNIOR, C T; DEGHAIDE, N H S; *et al.* The genetic structure of 3' untranslated region of the HLA-G gene: polymorphisms and haplotypes. **Genes and immunity**, v. 11, n. 2, p. 134–141, 2010.

CASTELLI, E C; MOREAU, P; OYA E CHIROMATZO, A; *et al.* In silico analysis of microRNAs targeting the HLA-G 3' untranslated region alleles and haplotypes. **Human immunology**, v. 70, n. 12, p. 1020–1025, 2009.

CASTELLI, E. C.; MENDES-JUNIOR, C. T.; VEIGA-CASTELLI, L. C.; *et al.* A comprehensive study of polymorphic sites along the HLA-G gene: implication for gene regulation and evolution. **Molecular Biology and Evolution**, v. 28, n. 11, p. 3069–3086, 2011.

CAUMARTIN, J.; FAVIER, B.; DAOUYA, M.; *et al.* Trogocytosis-based generation of suppressive NK cells. **The EMBO journal**, v. 26, n. 5, p. 1423–1433, 2007.

CIRULLI, V.; ZALATAN, J.; MCMASTER, M.; *et al.* The class I HLA repertoire of pancreatic islets comprises the nonclassical class Ib antigen HLA-G. **Diabetes**, v. 55, n. 5, p. 1214–1222, 2006.

CLEMENTS, C. S.; KJER-NIELSEN, L.; KOSTENKO, L.; *et al.* Crystal structure of HLA-G: a nonclassical MHC class I molecule expressed at the fetal-maternal interface. **Proceedings of the National Academy of Sciences of the United States of America**, v. 102, n. 9, p. 3360–3365, 2005.

COCK, P. J. A.; ANTAO, T.; CHANG, J. T.; *et al.* Biopython: freely available Python tools for computational molecular biology and bioinformatics. **Bioinformatics (Oxford, England)**, v. 25, n. 11, p. 1422–1423, 2009.

COHEN, N. R.; GARG, S.; BRENNER, M. B. Antigen Presentation by CD1 Lipids, T Cells, and NKT Cells in Microbial Immunity. **Advances in Immunology**, v. 102, p. 1–94, 2009.

COLOVOS, C.; YEATES, T. O. Verification of protein structures: patterns of nonbonded atomic interactions. **Protein Science: A Publication of the Protein Society**, v. 2, n. 9, p. 1511–1519, 1993.

COMISKEY, M.; DOMINO, K. E.; WARNER, C. M. HLA-G is found in lipid rafts and can act as a signaling molecule. **Human Immunology**, v. 68, n. 1, p. 1–11, 2007.

CURIGLIANO, G.; CRISCITIELLO, C.; GELAO, L.; GOLDBIRSCHE, A. Molecular Pathways: Human Leukocyte Antigen G (HLA-G). **Clinical Cancer Research**, v. 19, n. 20, p. 5564–5571, 2013.

DAVIS, M. M.; BJORKMAN, P. J. T-cell antigen receptor genes and T-cell recognition. **Nature**, v. 334, n. 6181, p. 395–402, 1988.

DE BEER, T. A. P.; BERKA, K.; THORNTON, J. M.; LASKOWSKI, R. A. PDBsum additions. **Nucleic Acids Research**, v. 42, n. Database issue, p. D292-296, 2014.

DIEHL, M.; MÜNZ, C.; KEILHOLZ, W.; *et al.* Nonclassical HLA-G molecules are classical peptide presenters. **Current biology: CB**, v. 6, n. 3, p. 305–314, 1996.

DIETRICH, J.; CELLA, M.; COLONNA, M. Ig-like transcript 2 (ILT2)/leukocyte Ig-like receptor 1 (LIR1) inhibits TCR signaling and actin cytoskeleton reorganization. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 166, n. 4, p. 2514–2521, 2001.

DONADI, E. A.; CASTELLI, E. C.; ARNAIZ-VILLENA, A.; *et al.* Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. **Cellular and molecular life sciences: CMLS**, v. 68, n. 3, p. 369–395, 2011.

DONAGHY, L.; GROS, F.; AMIOT, L.; *et al.* Elevated levels of soluble non-classical major histocompatibility class I molecule human leucocyte antigen (HLA)-G in the blood of HIV-infected patients with or without visceral leishmaniasis. **Clinical and Experimental Immunology**, v. 147, n. 2, p. 236–240, 2007.

EZEAKILE, M.; PORTIK-DOBOS, V.; WU, J.; *et al.* HLA-G dimers in the prolongation of kidney allograft survival. **Journal of Immunology Research**, v. 2014, p. 153981, 2014.

FAINARDI, E.; RIZZO, R.; MELCHIORRI, L.; *et al.* Intrathecal synthesis of soluble HLA-G and HLA-I molecules are reciprocally associated to clinical and MRI activity in patients with multiple sclerosis. **Multiple Sclerosis (Houndmills, Basingstoke, England)**, v. 12, n. 1, p. 2–12, 2006.

FALK, K.; RÖTZSCHKE, O.; STEVANOVIĆ, S.; JUNG, G.; RAMMENSEE, H. G. Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. **Nature**, v. 351, n. 6324, p. 290–296, 1991.

FARZAD, F.; ABEDIANKENARI, S. HLA-G6 and reproductive immunology. **Front. Immunol. Conference Abstract: 15th International Congress of Immunology (ICI)**. 2013.

FAVIER, B.; LEMAOULT, J.; ROUAS-FREISS, N.; *et al.* Research on HLA-G: an update. **Tissue Antigens**, v. 69, n. 3, p. 207–211, 2007.

FEGER, U.; TOLOSA, E.; HUANG, Y.-H.; *et al.* HLA-G expression defines a novel regulatory T-cell subset present in human peripheral blood and sites of inflammation. **Blood**, v. 110, n. 2, p. 568–577, 2007.

FISCHER, G. F.; MAYR, W. R. Molecular genetics of the HLA complex. **Wiener Klinische Wochenschrift**, v. 113, n. 20–21, p. 814–824, 2001.

FONS, P.; CHABOT, S.; CARTWRIGHT, J. E.; *et al.* Soluble HLA-G1 inhibits angiogenesis through an apoptotic pathway and by direct binding to CD160 receptor expressed by endothelial cells. **Blood**, v. 108, n. 8, p. 2608–2615, 2006.

FOURNEL, S.; AGUERRE-GIRR, M.; HUC, X.; *et al.* Cutting edge: soluble HLA-G1 triggers CD95/CD95 ligand-mediated apoptosis in activated CD8+ cells by interacting with CD8. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 164, n. 12, p. 6100–6104, 2000.

FREMONT, D. H.; MATSUMURA, M.; STURA, E. A.; PETERSON, P. A.; WILSON, I. A. Crystal structures of two viral peptides in complex with murine MHC class I H-2Kb. **Science (New York, N.Y.)**, v. 257, n. 5072, p. 919–927, 1992.

FUGLEBAKK, E.; ECHAVE, J.; REUTER, N. Measuring and comparing structural fluctuation patterns in large protein datasets. **Bioinformatics (Oxford, England)**, v. 28, n. 19, p. 2431–2440, 2012.

FUJII, T.; ISHITANI, A.; GERAGHTY, D. E. A soluble form of the HLA-G antigen is encoded by a messenger ribonucleic acid containing intron 4. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 153, n. 12, p. 5516–5524, 1994.

GAO, G. F.; WILLCOX, B. E.; WYER, J. R.; *et al.* Classical and nonclassical class I major histocompatibility complex molecules exhibit subtle conformational differences that affect binding to CD8 $\alpha$  $\alpha$ . **The Journal of Biological Chemistry**, v. 275, n. 20, p. 15232–15238, 2000.

GARBOCZI, D. N.; GHOSH, P.; UTZ, U.; *et al.* Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. **Nature**, v. 384, n. 6605, p. 134–141, 1996.

GARCIA, K. C.; DEGANNO, M.; PEASE, L. R.; *et al.* Structural basis of plasticity in T cell receptor recognition of a self peptide-MHC antigen. **Science (New York, N.Y.)**, v. 279, n. 5354, p. 1166–1172, 1998.

GARCIA, K. C.; DEGANNO, M.; STANFIELD, R. L.; *et al.* An alphabeta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. **Science (New York, N.Y.)**, v. 274, n. 5285, p. 209–219, 1996.

GONEN-GROSS, T.; ACHDOUT, H.; ARNON, T. I.; *et al.* The CD85J/leukocyte inhibitory receptor-1 distinguishes between conformed and beta 2-microglobulin-free HLA-G molecules. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 175, n. 8, p. 4866–4874, 2005.

GRAS, S.; BURROWS, S. R.; KJER-NIELSEN, L.; *et al.* The shaping of T cell receptor recognition by self-tolerance. **Immunity**, v. 30, n. 2, p. 193–203, 2009.

GRAS, S.; BURROWS, S. R.; TURNER, S. J.; *et al.* A structural voyage toward an understanding of the MHC-I-restricted immune response: lessons learned and much to be learned. **Immunological Reviews**, v. 250, n. 1, p. 61–81, 2012.

GREGORI, S.; TOMASONI, D.; PACCIANI, V.; *et al.* Differentiation of type 1 T regulatory cells (Tr1) by tolerogenic DC-10 requires the IL-10-dependent ILT4/HLA-G pathway. **Blood**, v. 116, n. 6, p. 935–944, 2010.

HAHN, Y. S.; BRACIALE, V. L.; BRACIALE, T. J. Presentation of viral antigen to class I major histocompatibility complex-restricted cytotoxic T lymphocyte. Recognition of an immunodominant influenza hemagglutinin site by cytotoxic T lymphocyte is independent of the position of the site in the hemagglutinin translation product. **The Journal of Experimental Medicine**, v. 174, n. 3, p. 733–736, 1991.

HENNECKE, J.; CARFI, A.; WILEY, D. C. Structure of a covalently stabilized complex of a human alphabeta T-cell receptor, influenza HA peptide and MHC class II molecule, HLA-DR1. **The EMBO journal**, v. 19, n. 21, p. 5611–5624, 2000.

HESS, B.; KUTZNER, C.; VAN DER SPOEL, D.; LINDAHL, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. **Journal of Chemical Theory and Computation**, v. 4, n. 3, p. 435–447, 2008.

HILLER, K.; GROTE, A.; SCHEER, M.; MÜNCH, R.; JAHN, D. PrediSi: prediction of signal peptides and their cleavage positions. **Nucleic Acids Research**, v. 32, n. Web Server issue, p. W375-379, 2004.

HINZ, A.; TAMPÉ, R. ABC transporters and immunity: mechanism of self-defense. **Biochemistry**, v. 51, n. 25, p. 4981–4989, 2012.

HODGKINSON, A D; MILLWARD, B A; DEMAINE, A G. The HLA-E locus is associated with age at onset and susceptibility to type 1 diabetes mellitus. **Human immunology**, v. 61, n. 3, p. 290–295, 2000.

HOWANGYIN, K.-Y.; LOUSTAU, M.; WU, J.; *et al.* Multimeric structures of HLA-G isoforms function through differential binding to LILRB receptors. **Cellular and molecular life sciences : CMLS**, v. 69, n. 23, p. 4041–4049, 2012.

HSU, K. C.; CHIDA, S.; GERAGHTY, D. E.; DUPONT, B. The killer cell immunoglobulin-like receptor (KIR) genomic region: gene-order, haplotypes and allelic polymorphism. **Immunological Reviews**, v. 190, p. 40–52, 2002.

HUANG, J.; RAUSCHER, S.; NAWROCKI, G.; *et al.* CHARMM36m: an improved force field for folded and intrinsically disordered proteins. **Nature Methods**, v. 14, n. 1, p. 71–73, 2017

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: visual molecular dynamics. **Journal of Molecular Graphics**, v. 14, n. 1, p. 33–38, 27–28, 1996.

HUOTARI, J.; HELENIUS, A. Endosome maturation. **The EMBO journal**, v. 30, n. 17, p. 3481–3500, 2011.

ISHITANI, A.; SAGESHIMA, N.; LEE, N.; *et al.* Protein expression and peptide binding suggest unique and interacting functional roles for HLA-E, F, and G in maternal-placental immune recognition. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 171, n. 3, p. 1376–1384, 2003.

JO, S.; KIM, T.; IYER, V. G.; IM, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. **Journal of Computational Chemistry**, v. 29, n. 11, p. 1859–1865, 2008.

JONES, D. C.; KOSMOLIAPTISIS, V.; APPS, R.; *et al.* HLA class I allelic sequence and conformation regulate leukocyte Ig-like receptor binding. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 186, n. 5, p. 2990–2997, 2011.

JUCH, H.; BLASCHITZ, A.; DAXBÖCK, C.; *et al.* A novel sandwich ELISA for alpha1 domain based detection of soluble HLA-G heavy chains. **Journal of Immunological Methods**, v. 307, n. 1–2, p. 96–106, 2005.

KÄLL, L.; KROGH, A.; SONNHAMMER, E. L. L. Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. **Nucleic Acids Research**, v. 35, n. Web Server issue, p. W429-432, 2007.

KARPLUS, M.; MCCAMMON, J. A. Molecular dynamics simulations of biomolecules. **Nature Structural Biology**, v. 9, n. 9, p. 646–652, 2002.

KJER-NIELSEN, L.; CLEMENTS, C. S.; BROOKS, A. G.; *et al.* The 1.5 Å crystal structure of a highly selected antiviral T cell receptor provides evidence for a structural basis of immunodominance. **Structure (London, England: 1993)**, v. 10, n. 11, p. 1521–1532, 2002.

KOZAKOV, D.; HALL, D. R.; XIA, B.; *et al.* The ClusPro web server for protein-protein docking. **Nature Protocols**, v. 12, n. 2, p. 255–278, 2017.

KRÁLOVICOVÁ, J.; LEI, H.; VORECHOVSKÝ, I. Phenotypic consequences of branch point substitutions. **Human Mutation**, v. 27, n. 8, p. 803–813, 2006.

KUKOL, A. Molecular modeling of proteins. 1<sup>a</sup> edição. Hatfield, Hertfordshire, UK. Humana Press, 2008. 383 páginas.

KURIMOTO, E.; KUROKI, K.; YAMAGUCHI, Y.; *et al.* Structural and functional mosaic nature of HLA class I molecules in their peptide-free form. **Molecular Immunology**, v. 55, n. 3–4, p. 393–399, 2013.

KUROKI, K.; MIO, K.; TAKAHASHI, A.; *et al.* Cutting Edge: Class II-like Structural Features and Strong Receptor Binding of the Nonclassical HLA-G2 Isoform Homodimer. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 198, n. 9, p. 3399–3403, 2017.

LAARIBI, A. B.; BORTOLOTTI, D.; HANNACHI, N.; *et al.* Increased levels of soluble HLA-G molecules in Tunisian patients with chronic hepatitis B infection. **Journal of Viral Hepatitis**, v. 24, n. 11, p. 1016–1022, 2017.

LAFON, M.; PREHAUD, C.; MEGRET, F.; *et al.* Modulation of HLA-G expression in human neural cells after neurotropic viral infections. **Journal of Virology**, v. 79, n. 24, p. 15226–15237, 2005.

LASKOWSKI, R. A.; RULLMANN, J. A.; MACARTHUR, M. W.; KAPTEIN, R.; THORNTON, J. M. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. **Journal of biomolecular NMR**, v. 8, n. 4, p. 477–486, 1996.

LE ROND, S.; AZÉMA, C.; KRAWICE-RADANNE, I.; *et al.* Evidence to support the role of HLA-G5 in allograft acceptance through induction of immunosuppressive/ regulatory T cells. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 176, n. 5, p. 3266–3276, 2006.

LEE, N.; MALACKO, A. R.; ISHITANI, A.; *et al.* The membrane-bound and soluble forms of HLA-G bind identical sets of endogenous peptides but differ with respect to TAP association. **Immunity**, v. 3, n. 5, p. 591–600, 1995.

LELEU, X.; LE FRIEC, G.; FACON, T.; *et al.* Total soluble HLA class I and soluble HLA-G in multiple myeloma and monoclonal gammopathy of undetermined significance. **Clinical Cancer Research: An Official Journal of the American Association for Cancer Research**, v. 11, n. 20, p. 7297–7303, 2005.

LEMAOULT, J.; CAUMARTIN, J.; DAOUYA, M.; *et al.* Immune regulation by pretenders: cell-to-cell transfers of HLA-G make effector T cells act as regulatory cells. **Blood**, v. 109, n. 5, p. 2040–2048, 2007.

LEMAOULT, J.; KRAWICE-RADANNE, I.; DAUSSET, J.; CAROSELLA, E. D. HLA-G1-expressing antigen-presenting cells induce immunosuppressive CD4<sup>+</sup> T cells. **Proceedings of the National Academy of Sciences of the United States of America**, v. 101, n. 18, p. 7064–7069, 2004.

LEMAOULT, J.; ZAFARANLOO, K.; LE DANFF, C.; CAROSELLA, E. D. HLA-G up-regulates ILT2, ILT3, ILT4, and KIR2DL4 in antigen presenting cells, NK cells, and T cells. **FASEB journal: official publication of the Federation of American Societies for Experimental Biology**, v. 19, n. 6, p. 662–664, 2005.

LEMBERG, M. K.; BLAND, F. A.; WEIHOFEN, A.; BRAUD, V. M.; MARTOGLIO, B. Intramembrane proteolysis of signal peptides: an essential step in the generation of HLA-E epitopes. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 167, n. 11, p. 6441–6446, 2001.

LEONHARDT, R. M.; KEUSEKOTTEN, K.; BEKPEN, C.; KNITTLER, M. R. Critical role for the tapasin-docking site of TAP2 in the functional integrity of the MHC class I-peptide-loading complex. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 175, n. 8, p. 5104–5114, 2005.

LI, W.; COWLEY, A.; ULUDAG, M.; *et al.* The EMBL-EBI bioinformatics web and programmatic tools framework. **Nucleic Acids Research**, v. 43, n. W1, p. W580-584, 2015.

LIANG, S.; ZHANG, W.; HORUZSKO, A. Human ILT2 receptor associates with murine MHC class I molecules in vivo and impairs T cell function. **European Journal of Immunology**, v. 36, n. 9, p. 2457–2471, 2006.

LILA, N.; AMREIN, C.; GUILLEMAIN, R.; *et al.* Soluble human leukocyte antigen-G: a new strategy for monitoring acute and chronic rejections after heart transplantation. **The Journal of Heart and Lung Transplantation: The Official Publication of the International Society for Heart Transplantation**, v. 26, n. 4, p. 421–422, 2007.

LODISH, H.; BERK, A.; ZIPURSKY, S.L.. *Molecular Cell Biology*. 4<sup>a</sup> Edição. Nova York: Editora W. H. Freeman, 2000.

LOZANO, J. M.; GONZÁLEZ, R.; KINDELÁN, J. M.; *et al.* Monocytes and T lymphocytes in HIV-1-positive patients express HLA-G molecule. **AIDS (London, England)**, v. 16, n. 3, p. 347–351, 2002.

LUQUE, J.; TORRES, M. I.; AUMENTE, M. D.; *et al.* Soluble HLA-G in heart transplantation: their relationship to rejection episodes and immunosuppressive therapy. **Human Immunology**, v. 67, n. 4–5, p. 257–263, 2006.

LÜTHY, R.; BOWIE, J. U.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**, v. 356, n. 6364, p. 83–85, 1992.

MADDEN, D. R.; GARBOCZI, D. N.; WILEY, D. C. The antigenic identity of peptide-MHC complexes: a comparison of the conformations of five viral peptides presented by HLA-A2. **Cell**, v. 75, n. 4, p. 693–708, 1993.

MCGUFFIN, L. J. The ModFOLD server for the quality assessment of protein structural models. **Bioinformatics (Oxford, England)**, v. 24, n. 4, p. 586–587, 2008.

MCINTIRE, R H; HUNT, J S. Antigen presenting cells and HLA-G--a review. **Placenta**, v. 26 Suppl A, p. S104–109, 2005.

MÉGRET, F.; PREHAUD, C.; LAFAGE, M.; *et al.* Modulation of HLA-G and HLA-E expression in human neuronal cells after rabies virus or herpes virus simplex type 1 infections. **Human Immunology**, v. 68, n. 4, p. 294–302, 2007.

MENIER, C.; RABREAU, M.; CHALLIER, J.-C.; *et al.* Erythroblasts secrete the nonclassical HLA-G molecule from primitive to definitive hematopoiesis. **Blood**, v. 104, n. 10, p. 3153–3160, 2004.

MENIER, C.; RITEAU, B.; DAUSSET, J.; CAROSELLA, E. D.; ROUAS-FREISS, N. HLA-G truncated isoforms can substitute for HLA-G1 in fetal survival. **Human Immunology**, v. 61, n. 11, p. 1118–1125, 2000.

MITSDOERFFER, M.; SCHREINER, B.; KIESEIER, B. C.; *et al.* Monocyte-derived HLA-G acts as a strong inhibitor of autologous CD4 T cell activation and is upregulated by interferon-beta in vitro and in vivo: rationale for the therapy of multiple sclerosis. **Journal of Neuroimmunology**, v. 159, n. 1–2, p. 155–164, 2005.

MONNERET, G.; VOIRIN, N.; KRAWICE-RADANNE, I.; *et al.* Soluble human leukocyte antigen-G5 in septic shock: marked and persisting elevation as a predictor of survival. **Critical Care Medicine**, v. 35, n. 8, p. 1942–1947, 2007.

MORALES, P. J.; PACE, J. L.; PLATT, J. S.; LANGAT, D. K.; HUNT, J. S. Synthesis of beta(2)-microglobulin-free, disulphide-linked HLA-G5 homodimers in human placental villous cytotrophoblast cells. **Immunology**, v. 122, n. 2, p. 179–188, 2007.

MORANDI, F.; LEVRERI, I.; BOCCA, P.; *et al.* Human neuroblastoma cells trigger an immunosuppressive program in monocytes by stimulating soluble HLA-G release. **Cancer Research**, v. 67, n. 13, p. 6433–6441, 2007.

MOREIRA, I. S.; FERNANDES, P. A.; RAMOS, M. J. Protein-protein docking dealing with the unknown. **Journal of Computational Chemistry**, v. 31, n. 2, p. 317–342, 2010.

MURPHY, K. Janeway's Immunobiology. 8<sup>a</sup> Edição. Nova York: Editora Garland Science, 2012.

NAJI, A.; LE ROND, S.; DURRBACH, A.; *et al.* CD3+CD4<sup>low</sup> and CD3+CD8<sup>low</sup> are induced by HLA-G: novel human peripheral blood suppressor T-cell subsets involved in transplant acceptance. **Blood**, v. 110, n. 12, p. 3936–3948, 2007.

NARDI, F. DA S.; KÖNIG, L.; WAGNER, B.; *et al.* Soluble monomers, dimers and HLA-G-expressing extracellular vesicles: the three dimensions of structural complexity to use HLA-G as a clinical biomarker. **HLA**, v. 88, n. 3, p. 77–86, 2016.

NIELSEN, H. Predicting Secretory Proteins with SignalP. **Methods in Molecular Biology (Clifton, N.J.)**, v. 1611, p. 59–73, 2017.

NÜCKEL, H.; REBMANN, V.; DÜRIG, J.; DÜHRSEN, U.; GROSSE-WILDE, H. HLA-G expression is associated with an unfavorable outcome and immunodeficiency in chronic lymphocytic leukemia. **Blood**, v. 105, n. 4, p. 1694–1698, 2005.

OANCEA, G.; O'MARA, M. L.; BENNETT, W. F. D.; *et al.* Structural arrangement of the transmission interface in the antigen ABC transport complex TAP. **Proceedings of the National Academy of Sciences of the United States of America**, v. 106, n. 14, p. 5551–5556, 2009.

ORTMANN, B.; COPEMAN, J.; LEHNER, P. J.; *et al.* A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. **Science (New York, N.Y.)**, v. 277, n. 5330, p. 1306–1309, 1997.

PANTER, M. S.; JAIN, A.; LEONHARDT, R. M.; HA, T.; CRESSWELL, P. Dynamics of major histocompatibility complex class I association with the human peptide-loading complex. **The Journal of Biological Chemistry**, v. 287, n. 37, p. 31172–31184, 2012.

PARK, B.; SPOONER, E.; HOUSER, B. L.; STROMINGER, J. L.; PLOEGH, H. L. The HCMV membrane glycoprotein US10 selectively targets HLA-G for degradation. **The Journal of Experimental Medicine**, v. 207, n. 9, p. 2033–2041, 2010.

PARKER, J. L.; NEWSTEAD, S. Membrane Protein Crystallisation: Current Trends and Future Perspectives. **Advances in Experimental Medicine and Biology**, v. 922, p. 61–72, 2016.

PAUL, P.; CABESTRE, F. A.; IBRAHIM, E. C.; *et al.* Identification of HLA-G7 as a new splice variant of the HLA-G mRNA and expression of soluble HLA-G5, -G6, and -G7 transcripts in human transfected cells. **Human Immunology**, v. 61, n. 11, p. 1138–1149, 2000.

PAUL, P.; ROUAS-FREISS, N.; KHALIL-DAHER, I.; *et al.* HLA-G expression in melanoma: a way for tumor cells to escape from immunosurveillance. **Proceedings of the National Academy of Sciences of the United States of America**, v. 95, n. 8, p. 4510–4515, 1998.

PENDE, D.; SIVORI, S.; ACCAME, L.; *et al.* HLA-G recognition by human natural killer cells. Involvement of CD94 both as inhibitory and as activating receptor complex. **European Journal of Immunology**, v. 27, n. 8, p. 1875–1880, 1997.

PETREY, D.; CHEN, T. S.; DENG, L.; *et al.* Template-based prediction of protein function. **Current Opinion in Structural Biology**, v. 32, p. 33–38, 2015.

PETTERSEN, E. F.; GODDARD, T. D.; HUANG, C. C.; *et al.* UCSF Chimera--a visualization system for exploratory research and analysis. **Journal of Computational Chemistry**, v. 25, n. 13, p. 1605–1612, 2004.

PISTOIA, V.; MORANDI, F.; WANG, X.; FERRONE, S. Soluble HLA-G: Are they clinically relevant? **Seminars in Cancer Biology**, v. 17, n. 6, p. 469–479, 2007.

QIU, J.; TERASAKI, P. I.; MILLER, J.; *et al.* Soluble HLA-G expression and renal graft acceptance. **American Journal of Transplantation: Official Journal of the American Society of Transplantation and the American Society of Transplant Surgeons**, v. 6, n. 9, p. 2152–2156, 2006.

RAJAGOPALAN, S.; BRYCESON, Y. T.; KUPPUSAMY, S. P.; *et al.* Activation of NK cells by an endocytosed receptor for soluble HLA-G. **PLoS biology**, v. 4, n. 1, p. e9, 2006.

RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. **Journal of Molecular Biology**, v. 7, p. 95–99, 1963.

RANGWALA, H.; KARYPIS, G. Introduction to Protein Structure Prediction: Methods and Algorithms. 1<sup>a</sup> Edição. Editora Wiley. 2010.

REBMANN, V.; KÖNIG, L.; NARDI, F. DA S.; *et al.* The Potential of HLA-G-Bearing Extracellular Vesicles as a Future Element in HLA-G Immune Biology. **Frontiers in Immunology**, v. 7, p. 173, 2016.

REISER, J. B.; DARNAULT, C.; GUIMEZANES, A.; *et al.* Crystal structure of a T cell receptor bound to an allogeneic MHC molecule. **Nature Immunology**, v. 1, n. 4, p. 291–297, 2000.

RIZZO, R.; MAPP, C. E.; MELCHIORRI, L.; *et al.* Defective production of soluble HLA-G molecules by peripheral blood monocytes in patients with asthma. **The Journal of Allergy and Clinical Immunology**, v. 115, n. 3, p. 508–513, 2005.

RIZZO, R.; MELCHIORRI, L.; SIMONE, L.; *et al.* Different production of soluble HLA-G antigens by peripheral blood mononuclear cells in ulcerative colitis and Crohn's disease: a noninvasive diagnostic tool? **Inflammatory Bowel Diseases**, v. 14, n. 1, p. 100–105, 2008.

RIZZO, R.; MELCHIORRI, L.; STIGNANI, M.; BARICORDI, O. R. HLA-G expression is a fundamental prerequisite to pregnancy. **Human Immunology**, v. 68, n. 4, p. 244–250, 2007.

ROBINSON, J.; HALLIWELL, J. A.; MCWILLIAM, H.; *et al.* The IMGT/HLA database. **Nucleic Acids Research**, v. 41, n. Database issue, p. D1222-1227, 2013.

RODGERS, J. R.; COOK, R. G. MHC class Ib molecules bridge innate and acquired immunity. **Nature Reviews. Immunology**, v. 5, n. 6, p. 459–471, 2005.

ROSSETTI, M.; GREGORI, S.; RONCAROLO, M. G. Granulocyte-colony stimulating factor drives the in vitro differentiation of human dendritic cells that induce anergy in naïve T cells. **European Journal of Immunology**, v. 40, n. 11, p. 3097–3106, 2010.

ROUAS-FREISS, N.; BRUEL, S.; MENIER, C.; *et al.* Switch of HLA-G alternative splicing in a melanoma cell line causes loss of HLA-G1 expression and sensitivity to NK lysis. **International Journal of Cancer**, v. 117, n. 1, p. 114–122, 2005.

ROUAS-FREISS, N.; GONÇALVES, R. M.; MENIER, C.; DAUSSET, J.; CAROSELLA, E. D. Direct evidence to support the role of HLA-G in protecting the fetus from maternal uterine natural killer cytotoxicity. **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 21, p. 11520–11525, 1997.

ROUAS-FREISS, N.; MARCHAL, R. E.; KIRSZENBAUM, M.; DAUSSET, J.; CAROSELLA, E. D. The alpha1 domain of HLA-G1 and HLA-G2 inhibits cytotoxicity induced by natural killer cells: is HLA-G the public ligand for natural killer cell inhibitory receptors? **Proceedings of the National Academy of Sciences of the United States of America**, v. 94, n. 10, p. 5249–5254, 1997.

ROUAS-FREISS, N.; MOREAU, P.; FERRONE, S.; CAROSELLA, E. D. HLA-G proteins in cancer: do they provide tumor cells with an escape mechanism? **Cancer Research**, v. 65, n. 22, p. 10139–10144, 2005.

RUDENSKY, A. Y.; MAZEL, S. M.; YURIN, V. L. Presentation of endogenous immunoglobulin determinant to immunoglobulin-recognizing T cell clones by the thymic cells. **European Journal of Immunology**, v. 20, n. 10, p. 2235–2239, 1990.

RUDOLPH, M. G.; STANFIELD, R. L.; WILSON, I. A. How TCRs bind MHCs, peptides, and coreceptors. **Annual Review of Immunology**, v. 24, p. 419–466, 2006.

RUDOLPH, M. G.; WILSON, I. A. The specificity of TCR/pMHC interaction. **Current Opinion in Immunology**, v. 14, n. 1, p. 52–65, 2002.

SCHUBERT, U.; ANTÓN, L. C.; GIBBS, J.; *et al.* Rapid degradation of a large fraction of newly synthesized proteins by proteasomes. **Nature**, v. 404, n. 6779, p. 770–774, 2000.

SCHWEDE, T. Protein modeling: what happened to the “protein structure gap”? **Structure (London, England: 1993)**, v. 21, n. 9, p. 1531–1540, 2013.

SEBTI, Y.; LE MAUX, A.; GROS, F.; *et al.* Expression of functional soluble human leucocyte antigen-G molecules in lymphoproliferative disorders. **British Journal of Haematology**, v. 138, n. 2, p. 202–212, 2007.

SELIGER, B.; SCHLAF, G. Structure, expression and function of HLA-G in renal cell carcinoma. **Seminars in Cancer Biology**, v. 17, n. 6, p. 444–450, 2007.

SHIH, I.-M. Application of human leukocyte antigen-G expression in the diagnosis of human cancer. **Human Immunology**, v. 68, n. 4, p. 272–276, 2007.

SHIROISHI, M.; KUROKI, K.; OSE, T.; *et al.* Efficient leukocyte Ig-like receptor signaling and crystal structure of disulfide-linked HLA-G dimer. **The Journal of Biological Chemistry**, v. 281, n. 15, p. 10439–10447, 2006.

SHIROISHI, M.; KUROKI, K.; RASUBALA, L.; *et al.* Structural basis for recognition of the nonclassical MHC molecule HLA-G by the leukocyte Ig-like receptor B2 (LILRB2/LIR2/ILT4/CD85d). **Proceedings of the National Academy of Sciences of the United States of America**, v. 103, n. 44, p. 16412–16417, 2006.

SHIROISHI, M.; TSUMOTO, K.; AMANO, K.; *et al.* Human inhibitory receptors Ig-like transcript 2 (ILT2) and ILT4 compete with CD8 for MHC class I binding and bind preferentially to HLA-G. **Proceedings of the National Academy of Sciences of the United States of America**, v. 100, n. 15, p. 8856–8861, 2003.

SIEVERS, F.; WILM, A.; DINEEN, D.; *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular Systems Biology**, v. 7, p. 539, 2011.

SIM, B. C.; TRAVERS, P. J.; GASCOIGNE, N. R. V alpha 3.2 selection in MHC class I mutant mice: evidence for an alternate orientation of TCR-MHC class I interaction. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 159, n. 7, p. 3322–3329, 1997.

SIMOSSIS, V. A.; HERINGA, J. PRALINE: a multiple sequence alignment toolbox that integrates homology-extended and secondary structure information. **Nucleic Acids Research**, v. 33, n. Web Server issue, p. W289-294, 2005

SINGER, G.; REBMANN, V.; CHEN, Y.-C.; *et al.* HLA-G is a potential tumor marker in malignant ascites. **Clinical Cancer Research: An Official Journal of the American Association for Cancer Research**, v. 9, n. 12, p. 4460–4464, 2003.

SYKULEV, Y.; JOO, M.; VTURINA, I.; TSOMIDES, T. J.; EISEN, H. N. Evidence that a single peptide-MHC complex on a target cell can elicit a cytolytic T cell response. **Immunity**, v. 4, n. 6, p. 565–571, 1996.

TAHAN, F.; PATIROGLU, T. Plasma soluble human leukocyte antigen G levels in asthmatic children. **International Archives of Allergy and Immunology**, v. 141, n. 3, p. 213–216, 2006.

TAN, P.; KROPSHOFER, H.; MANDELBOIM, O.; *et al.* Recruitment of MHC class I molecules by tapasin into the transporter associated with antigen processing-associated complex is essential for optimal peptide loading. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 168, n. 4, p. 1950–1960, 2002.

THE UNIPROT CONSORTIUM. UniProt: the universal protein knowledgebase. **Nucleic Acids Research**, v. 45, n. D1, p. D158–D169, 2017.

THEODOSSIS, A.; GUILLONNEAU, C.; WELLAND, A.; *et al.* Constraints within major histocompatibility complex class I restricted peptides: presentation and consequences for T-cell recognition. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 12, p. 5534–5539, 2010.

TOMASEC, P.; BRAUD, V. M.; RICKARDS, C.; *et al.* Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. **Science (New York, N.Y.)**, v. 287, n. 5455, p. 1031, 2000.

TORRES, M. I.; LE DISCORDE, M.; LORITE, P.; *et al.* Expression of HLA-G in inflammatory bowel disease provides a potential way to distinguish between ulcerative colitis and Crohn's disease. **International Immunology**, v. 16, n. 4, p. 579–583, 2004.

TORRES, M. I.; LÓPEZ CASADO, M. A.; RÍOS, A. New aspects in celiac disease. **World Journal of Gastroenterology**, v. 13, n. 8, p. 1156–1161, 2007.

TROTT, O.; OLSON, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. **Journal of Computational Chemistry**, v. 31, n. 2, p. 455–461, 2010.

TURNER, S. J.; DOHERTY, P. C.; MCCLUSKEY, J.; ROSSJOHN, J. Structural determinants of T-cell receptor bias in immunity. **Nature Reviews. Immunology**, v. 6, n. 12, p. 883–894, 2006.

TYSOE-CALNON, V. A.; GRUNDY, J. E.; PERKINS, S. J. Molecular comparisons of the beta 2-microglobulin-binding site in class I major-histocompatibility-complex alpha-chains and proteins of related sequences. **Biochemical Journal**, v. 277, n. Pt 2, p. 359–369, 1991.

UGUREL, S.; REBMANN, V.; FERRONE, S.; *et al.* Soluble human leukocyte antigen--G serum level is elevated in melanoma patients and is further increased by interferon-alpha immunotherapy. **Cancer**, v. 92, n. 2, p. 369–376, 2001.

VAKSER, I. A. Protein-protein docking: from interaction to interactome. **Biophysical Journal**, v. 107, n. 8, p. 1785–1793, 2014.

VAN BLEEK, G. M.; NATHENSON, S. G. The structure of the antigen-binding groove of major histocompatibility complex class I molecules determines specific selection of self-peptides. **Proceedings of the National Academy of Sciences of the United States of America**, v. 88, n. 24, p. 11032–11036, 1991.

VAN DER MEER, A.; LUKASSEN, H. G. M.; VAN CRANENBROEK, B.; *et al.* Soluble HLA-G promotes Th1-type cytokine production by cytokine-activated uterine and peripheral natural killer cells. **Molecular Human Reproduction**, v. 13, n. 2, p. 123–133, 2007.

VAN DER MEER, A.; LUKASSEN, H. G. M.; VAN LIEROP, M. J. C.; *et al.* Membrane-bound HLA-G activates proliferation and interferon-gamma production by uterine natural killer cells. **Molecular Human Reproduction**, v. 10, n. 3, p. 189–195, 2004.

VAN ZUNDERT, G. C. P.; RODRIGUES, J. P. G. L. M.; TRELLET, M.; *et al.* The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. **Journal of Molecular Biology**, v. 428, n. 4, p. 720–725, 2016

VEIGA-CASTELLI, L C; CASTELLI, E C; MENDES, C T, Jr; *et al.* Non-classical HLA-E gene variability in Brazilians: a nearly invariable locus surrounded by the most variable genes in the human genome. **Tissue antigens**, v. 79, n. 1, p. 15–24, 2012.

VEIGA-CASTELLI, L C; DE PAULA CRUZ, A S; INÁCIO, M M; *et al.* Lack of association between HLA-E polymorphisms and transitional cell carcinoma of the bladder. **Tissue antigens**, v. 82, n. 3, p. 197–200, 2013.

VERBRUGGEN, L. A.; REBMANN, V.; DEMANET, C.; DE COCK, S.; GROSSE-WILDE, H. Soluble HLA-G in rheumatoid arthritis. **Human Immunology**, v. 67, n. 8, p. 561–567, 2006.

VERLI, H. *Bioinformática: da Biologia à Flexibilidade Molecular*. 1ª Edição. São Paulo: Sociedade Brasileira de Bioquímica e Biologia Molecular SBBq, 2014.

VIGNERON, N.; VAN DEN EYNDE, B. J. Insights into the processing of MHC class I ligands gained from the study of human tumor epitopes. **Cellular and molecular life sciences: CMLS**, v. 68, n. 9, p. 1503–1520, 2011.

VILLANUEVA, M. S.; SIJTS, A. J.; PAMER, E. G. Listeriolysin is processed efficiently into an MHC class I-associated epitope in *Listeria monocytogenes*-infected cells. **Journal of Immunology (Baltimore, Md.: 1950)**, v. 155, n. 11, p. 5227–5233, 1995.

VOSS, M.; SCHRÖDER, B.; FLUHRER, R. Mechanism, specificity, and physiology of signal peptide peptidase (SPP) and SPP-like proteases. **Biochimica Et Biophysica Acta**, v. 1828, n. 12, p. 2828–2839, 2013.

WAGNER, S. N.; REBMANN, V.; WILLERS, C. P.; GROSSE-WILDE, H.; GOOS, M. Expression analysis of classic and non-classic HLA molecules before interferon alfa-2b treatment of melanoma. **Lancet (London, England)**, v. 356, n. 9225, p. 220–221, 2000.

WALPOLE, N. G.; KJER-NIELSEN, L.; KOSTENKO, L.; *et al.* The structure and stability of the monomorphic HLA-G are influenced by the nature of the bound peptide. **Journal of Molecular Biology**, v. 397, n. 2, p. 467–480, 2010.

WANG, J.; REINHERZ, E. L. Structural basis of T cell recognition of peptides bound to MHC molecules. **Molecular Immunology**, v. 38, n. 14, p. 1039–1049, 2002.

WATTS, C. The exogenous pathway for antigen presentation on major histocompatibility complex class II and CD1 molecules. **Nature Immunology**, v. 5, n. 7, p. 685–692, 2004.

WEBB, B.; SALI, A. Comparative Protein Structure Modeling Using MODELLER. *Current Protocols in Bioinformatics*, 2002.

WIEDERSTEIN, M.; SIPPL, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids Research**, v. 35, n. Web Server issue, p. W407–410, 2007.

WIENDL, H.; MITSDOERFFER, M.; HOFMEISTER, V.; *et al.* The non-classical MHC molecule HLA-G protects human muscle cells from immune-mediated lysis: implications for myoblast transplantation and gene therapy. **Brain: A Journal of Neurology**, v. 126, n. Pt 1, p. 176–185, 2003.

XIANG, Z. Advances in homology protein structure modeling. **Current Protein & Peptide Science**, v. 7, n. 3, p. 217–227, 2006.

YAN, W.-H.; LIU, D.; LU, H.-Y.; *et al.* Significance of tumour cell HLA-G5/-G6 isoform expression in discrimination for adenocarcinoma from squamous cell carcinoma in lung cancer patients. **Journal of Cellular and Molecular Medicine**, v. 19, n. 4, p. 778–785, 2015.

YE, S.-R.; YANG, H.; LI, K.; *et al.* Human leukocyte antigen G expression: as a significant prognostic indicator for patients with colorectal cancer. **Modern Pathology: An Official Journal of the United States and Canadian Academy of Pathology, Inc**, v. 20, n. 3, p. 375–383, 2007.

YEWDELL, J. W. DRiPs solidify: progress in understanding endogenous MHC class I antigen processing. **Trends in Immunology**, v. 32, n. 11, p. 548–558, 2011.

YIN, Y.; LI, Y.; MARIUZZA, R. A. Structural basis for self-recognition by autoimmune T-cell receptors. **Immunological Reviews**, v. 250, n. 1, p. 32–48, 2012.

ZAJONC, D. M.; CRISPIN, M. D. M.; BOWDEN, T. A.; *et al.* Molecular mechanism of lipopeptide presentation by CD1a. **Immunity**, v. 22, n. 2, p. 209–219, 2005.

ZAJONC, D. M.; ELSLIGER, M. A.; TEYTON, L.; WILSON, I. A. Crystal structure of CD1a in complex with a sulfatide self antigen at a resolution of 2.15 Å. **Nature Immunology**, v. 4, n. 8, p. 808–815, 2003.

ZHANG, J.; YANG, J.; JANG, R.; ZHANG, Y. GPCR-I-TASSER: A Hybrid Approach to G Protein-Coupled Receptor Structure Modeling and the Application to the Human Genome. **Structure (London, England: 1993)**, v. 23, n. 8, p. 1538–1549, 2015.

ZHANG, W.-Q.; XU, D.-P.; LIU, D.; *et al.* HLA-G1 and HLA-G5 isoforms have an additive effect on NK cytotoxicity. **Human Immunology**, v. 75, n. 2, p. 182–189, 2014.

ZHANG, X.; HAN, Q.-Y.; LI, J.-B.; *et al.* Lesion HLA-G5/-G6 isoforms expression in patients with ovarian cancer. **Human Immunology**, v. 77, n. 9, p. 780–784, 2016.

ZHANG, Y.; SKOLNICK, J. Scoring function for automated assessment of protein structure template quality. **Proteins**, v. 57, n. 4, p. 702–710, 2004.

## 11 ANEXOS

## ANEXO A – Comitê de ética

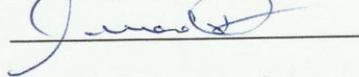
Ribeirão Preto, 15 de outubro de 2015.

Quanto à aprovação do comitê de ética,

Eu, **Professor Doutor Eduardo Antônio Donadi**, declaro que a aluna de Doutorado **Thaís Cristine Arns**, matriculada no programa de pós-graduação em Imunologia Básica e Aplicada da Faculdade de Medicina de Ribeirão Preto, n° **USP 7364378** e **RG 6.522.487-9**, desenvolve um projeto de doutorado que não necessita a aprovação do comitê de ética, pois não são utilizados animais ou amostras provenientes de pacientes. O trabalho consiste da predição de estruturas gênicas e modelagem *in silico* de proteínas utilizando ferramentas computacionais. Todo o trabalho é realizado utilizando ferramentas de bioinformática e dados provenientes de bancos de dados públicos disponíveis na internet.

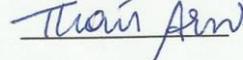
Atenciosamente,

Prof. Dr. Eduardo Antônio Donadi



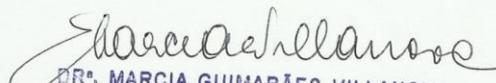
Eduardo A. Donadi  
Imunologia Clínica  
CRM 28795

Thaís Cristine Arns




Não há necessidade  
de assinatura ao CEP

20/10/15



DRª. MARCIA GUIMARÃES VILLANOVA  
Coordenadora do Comitê de Ética em Pesquisa  
do HCFMRP-USP e da FMRP-USP

**ANEXO B – Sequência de aminoácidos da proteína codificada pelo alelo *HLA-G\*01:01:01:01***

>HLA-G\*01:01:01:01

MVVMAPRTLFLLLSGALTLTETWAGSHSMRYFSAAVSRPGRGEPRIAMGYVDDTQ  
 FVRFSDSACPRMEPRAPWVEQEGPEYWEEETRNTKAHAQTDRMNLQTLRGYYNQ  
 SEASSHTLQWMIGCDLGS DGRLLRGYEQYAYDGKDYLALNEDLRSWTAADTAAQIS  
 KRKCEAANVAEQRRAYLEGTCVEWLHRYLENGKEMLQRADPPKTHVTHHPVFDYE  
 ATLRCWALGFYP AEIILTWQRDGEDQTQDVELVETRPAGDGT FQKWA AVVVPSGEE  
 QRYTCHVQHEGLPEPLMLRWKQSSLPTIPIMGIVAGLVVLA AVVTGA AVAVLWRK  
 KSSD

**ANEXO C – Sequência de aminoácidos das isoformas *HLA-G5*, *HLA-G6* e *HLA-G7***

>HLA-G5

MVVMAPRTLFLLLSGALTLTETWAGSHSMRYFSAAVSRPGRGEPRIAMGYVDDTQ  
 FVRFSDSACPRMEPRAPWVEQEGPEYWEEETRNTKAHAQTDRMNLQTLRGYYNQ  
 SEASSHTLQWMIGCDLGS DGRLLRGYEQYAYDGKDYLALNEDLRSWTAADTAAQIS  
 KRKCEAANVAEQRRAYLEGTCVEWLHRYLENGKEMLQRADPPKTHVTHHPVFDYE  
 ATLRCWALGFYP AEIILTWQRDGEDQTQDVELVETRPAGDGT FQKWA AVVVPSGEE  
 QRYTCHVQHEGLPEPLMLRWSKEGDGGIMSVRESRSLSEDL

>HLA-G6

MVVMAPRTLFLLLSGALTLTETWAGSHSMRYFSAAVSRPGRGEPRIAMGYVDDTQ  
 FVRFSDSACPRMEPRAPWVEQEGPEYWEEETRNTKAHAQTDRMNLQTLRGYYNQ  
 SEADPPKTHVTHHPVFDYEATLRCWALGFYP AEIILTWQRDGEDQTQDVELVETRPA  
 GDGT FQKWA AVVVPSGEEQRYTCHVQHEGLPEPLMLRWSKEGDGGIMSVRESRSL  
 EDL

>HLA-G7

MVVMAPRTLFLLLSGALTLTETWAGSHSMRYFSAAVSRPGRGEPRIAMGYVDDTQ  
 FVRFSDSACPRMEPRAPWVEQEGPEYWEEETRNTKAHAQTDRMNLQTLRGYYNQ  
 SEASE

**ANEXO D – Manuscrito em preparação para submissão**

**Authors** Thais Arns<sup>1</sup>, Dinler Antunes<sup>2</sup>, Elvira Tamarozzi<sup>3</sup>, Lydia Kavraki<sup>2</sup>, Silvana Giuliatti<sup>3</sup>, Eduardo Antônio Donadi<sup>1</sup>.

**Title** Unveiling the molecular structure and dynamics of lipid-anchored immunomodulatory molecule HLA-G and its soluble isoforms

**Affiliation**

<sup>1</sup> Department of Basic and Applied Immunology, Ribeirão Preto Medical School, University of São Paulo, Brazil.

<sup>2</sup> Department of Computer Science, Rice University, Houston, TX 77005 USA

<sup>3</sup> Department of Genetics, Ribeirão Preto Medical School, University of São Paulo, Brazil.

**Address** Av. Bandeirantes, 3900 - Monte Alegre - CEP: 14049-900 Ribeirão Preto/SP.

**E-mail** thaisarns@usp.br

**Abstract** HLA-G is a histocompatibility molecule, considered tolerogenic that plays an important role in suppressing the immune response. The messenger RNA encoding the HLA-G protein may give rise to at least 7 distinct isoforms by alternative splicing, and the functions of these isoforms are unknown but could be involved in the modulation of leukocyte activity. Since all segments of the HLA-G molecule and its isoforms may contribute to the suppressive capacity and interaction with immune system receptors, the aim of this work was to obtain the complete structure of HLA-G protein inserted in a lipid bilayer, as well as the complete structures of the HLA-G5, HLA-G6 and HLA-G7 soluble isoforms. By combining structural bioinformatics methodologies, our study demonstrated the structural stability and movement patterns of the complete HLA-G molecule inserted in a lipid bilayer. The simulations associated to HLA-G5 demonstrated the great instability of structures lacking the peptide and coupled  $\beta$ 2-microglobulin. The monomeric conformations tested for the HLA-G6 isoform were not stable when subjected to molecular dynamics and perhaps HLA-G6 may not exist as a soluble monomer, requiring a dimeric conformation for its stability. It was possible to observe that HLA-G7 peptide-containing monomer was more stable than the same construct in the absence of the peptide. Such complete models can be used for *in silico* modeling of the other alleles and isoforms produced, as well as virtual screenings in the search for possible immune system receptors and molecules that do not yet have an interaction described with HLA-G and its isoforms.

**Keywords** HLA-G, isoforms, molecular dynamics, molecular docking, bioinformatics.