A 3D molecular model of a protein-DNA complex. The protein is represented by a large, blue, semi-transparent surface composed of many small spheres, showing a complex, multi-domain structure. The DNA is represented by a double helix structure at the bottom, colored in orange and yellow. The background is a light blue gradient.

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

CINTHIA CAROLINE ALVES

Caracterização estrutural do polimorfismo -964G>A na interação do fator de transcrição HIF1 com seu elemento regulatório na região promotora do gene *HLA-G*.

Ribeirão Preto
2021

UNIVERSIDADE DE SÃO PAULO
FACULDADE DE MEDICINA DE RIBEIRÃO PRETO

CINTHIA CAROLINE ALVES

Caracterização estrutural do polimorfismo -964G>A na interação do fator de transcrição HIF1 com seu elemento regulatório na região promotora do gene *HLA-G*.

Ribeirão Preto
2021

CINTHIA CAROLINE ALVES

Caracterização estrutural do polimorfismo -964G>A na interação do fator de transcrição HIF1 com seu elemento regulatório na região promotora do gene *HLA-G*.

Versão corrigida. A versão original encontra-se disponível tanto na Biblioteca da Unidade que aloja o Programa, quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD).

Tese apresentada ao programa de Pós-Graduação em Genética da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo para obtenção do título de doutor(a) em Ciências.

Área de concentração: Genética

Orientador: Profa. Dra. Silvana Giuliatti

Ribeirão Preto
2021

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.

Catálogo de publicação preparada pela biblioteca do serviço de biblioteca de documentação da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo.

Alves, Cinthia Caroline

Caracterização estrutural do polimorfismo -964G>A na interação do fator de transcrição HIF1 com seu elemento regulatório na região promotora do gene *HLA-G*. Ribeirão Preto, 2021.

168f. il.; 30 cm

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

Área de concentração: Genética.

Orientador (a): Giuliatti, Silvana.

1. Polimorfismo 2. MHC de classe I 3. Ligações intermoleculares
4. Proteína-DNA. 5. Hipóxia.

Nome: Cinthia Caroline Alves

Título: Caracterização estrutural do polimorfismo -964G>A na interação do fator de transcrição HIF1 com seu elemento regulatório na região promotora do gene *HLA-G*.

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

Aprovada em: 10 de agosto de 2021.

Banca examinadora:

Profa. Dra. Silvana Giuliatti

Instituição: Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo

Prof. Dr. Marco Antônio Alves da Silva

Instituição: Faculdade de Ciências Farmacêuticas de Ribeirão Preto da Universidade de São Paulo

Prof. Dr. Celso Teixeira Mendes Junior

Instituição: Faculdade de Ciências, Filosofia e Letras de Ribeirão Preto da Universidade de São Paulo

Prof. Dr. Andrei Leitão

Instituição: Instituto de Química de São Carlos da Universidade de São Paulo

Dedico

À Maria,

Mãe, rainha e serva. Minha luz hoje e sempre.

*Aos meus pais, Suzimara B. Alves e José Flávio Alves, e à
minha irmã, Ciane C. Alves,*

Minha família é meu salva-vidas no mar agitado da vida.

Ao Theodoro,

Meu anjinho de quatro patas.

AGRADECIMENTO

A Deus, por me amparar nos momentos difíceis e sustentar minha força em meio as dificuldades; por me mostrar os caminhos nas horas incertas e me suprir em todas as minhas necessidades. Obrigada por ser minha luz nos momentos de escuridão. Obrigada, Senhor!

À Profa. Dra. Silvana Giuliatti pela orientação, dedicação e paciência. Por compartilhar seu conhecimento e experiência e, assim, contribuir para minha formação como docente e pesquisadora. Muito obrigada, Professora!

Ao Prof. Dr. Eduardo Antônio Donadi, pelo apoio incondicional para meu crescimento pessoal e profissional. Por ser uma constante fonte de motivação e incentivo ao longo de todo o projeto, como me incentivar a nunca desistir dos meus objetivos por mais difíceis que sejam! Obrigada por não me fazer desistir de mim, Professor!

Aos amigos e colegas do Grupo de Bioinformática (GBi) do Departamento de Genética da FMRP-USP agradeço pelo companheirismo e ajuda! Em especial agradeço aos amigos Levy, Elvira, Rauni, Marcelo e Daniel por tantas risadas, conhecimentos, experiências e momentos compartilhados.

Aos meus pais e minha irmã, por sempre apoiarem minhas decisões, principalmente pelo auxílio emocional frente as dificuldades e momentos de incerteza, onde o afeto e o amor que sempre me proporcionaram foram o combustível para enfrentar as adversidades da vida. O amparo e o amor de vocês nos momentos de reclusão e dificuldade foram, e são, essenciais em minha vida. Eu amo vocês. Obrigada por tudo.

A todos os meus familiares, em especial, meus avós, Maria e Wilson, meus tios Maria Cristina e Adriano, e meu primo, Otávio, pelo carinho, paciência e incentivo, e a todos os outros não destacados aqui, mas que estavam comigo em muitos momentos desta jornada.

À banca examinadora por aceitar julgar meu trabalho e contribuir para a melhoria do mesmo.

Aos meus amigos, pelos momentos de descontração e amizade. A todos os colegas, professores e funcionários da Pós-Graduação em Genética da FMRP-USP, em especial, a secretária do programa, Susie, pela assistência, apoio e amizade. E as funcionárias, Maria e Gislane, pela amizade.

À Superintendência da Tecnologia e Informação da USP pelos Recursos Computacionais de *High Performance Computing* (HPC) usados para o desenvolvimento deste trabalho.

Às agências de fomento, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, e Conselho Nacional de Desenvolvimento Científico e Tecnológico – Brasil (CNPq), processo #141224/2017-7, pelo auxílio financeiro na realização do presente trabalho.

“Não permita que as adversidades da vida acabem com seus sonhos mais ternos. Perdoe o passado, viva o hoje e tenha esperança no amanhã”.

- ALVES, Cinthia C. (2021)

RESUMO

ALVES, Cinthia Caroline. Caracterização estrutural do polimorfismo -964G>A na interação do fator de transcrição HIF1 com seu elemento regulatório na região promotora do gene *HLA-G*. Tese (Doutorado) – Faculdade de Medicina de Ribeirão, Universidade de São Paulo, Ribeirão Preto, 2021.

O gene do Antígeno leucocitário humano-G (HLA-G, do inglês, *Human Antigen Leukocyte-G*) codifica uma molécula reguladora do sistema imunológico com expressão tecidual restrita em condições fisiológicas; entretanto, o gene pode ser induzido em condições de hipóxia pela interação com o fator induzível por hipóxia 1 (do inglês, *Hypoxia Inducible Factor 1*, HIF1). Os elementos responsivos à hipóxia (do inglês, *Hypoxia Responsive Element*, HRE) localizados na região promotora (posição -966 a -962) e no éxon 2 são os principais elementos regulatórios alvos do HIF1. O alelo G no sítio polimórfico -964 G>A induz maior expressão de *HLA-G* quando comparado ao alelo A em situações de baixo teor de oxigênio. Uma vez que a variabilidade genômica do *HLA-G* pode influenciar a formação do complexo HIF1-HRE, neste estudo foi analisado a interação física do complexo a nível de pares de átomos considerando ambos os alelos -964G e -964A, usando abordagens computacionais. O cristal do heterodímero HIF2 de camundongo (*Protein Data Bank* ID: 4ZPK, 3,6Å-<https://www.rcsb.org/>) que pertence à mesma família de proteínas que o HIF1, foi usado como molde para realizar a modelagem molecular por homologia da estrutura quaternária do HIF1 humano. A estrutura primária do HIF1 humano que abrange os resíduos 15 a 349 para a subunidade HIF1a, e 87 a 470 para a subunidade ARNT (*UniProt* ID: Q16665 e P27540, respectivamente-<https://www.uniprot.org/>) foi recuperada do banco de dados UniProt. Definidos o molde e o alvo, o *software* MODELLER v9.24 foi usado para a modelagem molecular por homologia. O modelo mais semelhante estruturalmente ao molde foi avaliado de acordo com a qualidade físico-química e estereoquímica da proteína. Duas estruturas 3D de DNA foram construídas a partir da sequência 5'GCRTG'3 do HRE contendo o alelo -964G ou -964A na posição "R" usando o *software* x3DNA. O *docking* molecular proteína-DNA foi realizado usando o servidor HADDOCK v2.4, e as interações intermoleculares foram calculadas pelo servidor DNAProDB. A simulação de dinâmica molecular foi realizada para a proteína livre e os complexos proteína-DNA por 200 ns usando o *software* Gromacs v.2019. A partir das coordenadas atômicas da trajetória foi calculado os parâmetros *inter-base pair steps* do DNA pelo programa Curves v.2.6. O modelo do HIF1 mais semelhante ao molde gerado por modelagem comparativa obteve o RMSD mais baixo (0,396Å) após o alinhamento das estruturas em comparação com outros modelos, e apresentou 87,2% dos resíduos na região central dos ângulos de torção *phi-psi*, enquanto o molde apresentou 80,10%. A ligação do HIF1 com o HRE contendo o alelo -964 G resultou na formação de mais ligações de hidrogênio e contatos de Van der Waals do que HRE com alelo -964 A. A análise das trajetórias do complexo RMSD proteína-DNA revelou que o complexo HIF1-HRE-964G é mais estável. Em conclusão, o HIF1 se liga de maneira mais estável e específica no HRE com o alelo G.

Palavras-chaves: proteína-DNA, MHC de classe I, *docking* molecular

ABSTRACT

ALVES, Cinthia Caroline. *Structural characterization of the -964G>A polymorphism in the interaction of HIF1 transcription factor with its regulatory element in the promoter region of the HLA-G gene*. Thesis (PhD) – Ribeirão Preto Medical School, University of São Paulo, 2021.

Human Antigen Leukocyte-G (HLA-G) gene encodes an immune checkpoint molecule that has restricted tissue expression in physiological conditions; however, the gene may be induced in hypoxic conditions by the interaction with the hypoxia inducible factor-1 (HIF1). Hypoxia responsive elements (HRE) located into HLA-G promoter region (-966 and -962 nucleotides) and exon 2 are the major HIF1 target sites. The G allele of the -964 G>A transversion induces higher HLA-G expression when compared to the A allele in hypoxic conditions. Since HLA-G variability may influence the HIF1-HRE binding, we analyzed this complex interaction at the atomistic level considering both -964G and -964A alleles, using computational approaches. Mouse HIF2 dimer crystal (Protein Data Bank ID: 4ZPK, 3.6Å-<https://www.rcsb.org/>) that belongs to the same HIF1 protein family was used as template to perform homology modelling of human HIF1 quaternary structure, which encompasses residues 15-349 to HIF1a subunit and 87-470 to ARNT subunit (UniProt ID: Q16665 and P27540, respectively-<https://www.uniprot.org/>), using MODELLER v.9.24. The most similar template structural model was evaluated according to protein physical-chemical and stereochemical quality. Two 3D DNA structures were built from 5'GCRTG'3 HRE sequence containing the -964G or -964A allele at "R" position using x3DNA v.2.4. Protein-DNA docking was performed using HADDOCK v2.4, and complex intermolecular interactions were computed by DNAproDB server. Molecular dynamic simulation was carried out to protein and protein-DNA complexes per 200 ns using Gromacs v.2019. Inter bp step from DNA atomic coordinate trajectories were calculated by Curves v.2.6 program. HIF1 most similar template model generated by homology modeling showed the lowest RMSD of 0.396Å after template alignment in comparison to other models, and it presented 87.2% of the residues in the core region of phi-psi torsion angles, while template presented 80.10%. HIF1 binding in the HRE containing -964 G allele results in more hydrogen bonds and Van der Waals contacts formation than HRE with -964 A allele. RMSD protein-DNA complex trajectories analysis reveal that HIF1-HRE-964G complex is more stable than HIF1-HRE complex with A allele. In conclusion, HIF1 binds in a more stable and specific manner in the HRE with G allele.

Key-words: protein-DNA, MHC class I, molecular docking.

LISTA DE FIGURAS

Figura 1. Representação esquemática dos domínios proteicos que compõem as subunidades HIF1 α e ARNT.....	4
Figura 2. Mecanismos de ativação e inibição do HIF1 α	5
Figura 3. Isoformas da subunidade HIF α humano.....	7
Figura 4. A sequência consenso RCGTG (onde R é A ou G) do HRE e nucleotídeos de flaqueamento..	8
Figura 5. HIF1 α ativa a expressão de HLA-G através de HREs localizados no locus <i>HLA-G</i>	11
Figura 6. Representação esquemática do controle transcricional mediado por fatores de transcrição específicos.....	13
Figura 7. Número total de estruturas tridimensionais de complexos proteína-ácidos nucleicos depositadas no PDB por ano.....	14
Figura 8. Estrutura da dupla-hélice de DNA	15
Figura 9. Mecanismos de reconhecimento das bases nitrogenadas pelos aminoácidos do DNA.	17
Figura 10. Parâmetros conformacionais empregados para descrever a geometria da dupla-hélice de DNA.	18
Figura 11. Relação entre as abordagens computacionais para predição de estruturas 3D de proteínas e o uso de informações provenientes de estruturas resolvidas experimentalmente depositadas no PDB.....	21
Figura 12. Representação de uma proteína dentro de uma caixa cúbica com água.	27
Figura 13. Representação das condições periódicas de contorno em uma caixa cúbica.....	28
Figura 14. Fluxograma do processo de modelagem e docking dos complexos moleculares ..	35
Figura 15. Domínios proteicos selecionados para modelagem do fator de transcrição HIF1.	36
Figura 16. Alinhamento global feito por meio do software EMBOSS Needle entre as sequências primárias das subunidades β (ARNT) e α (HIF1/2 α) dos fatores de transcrição HIF1 (humano) e HIF2 (camundongo).	51
Figura 17. Predição de estrutura secundária do complexo HIF2 de camundongo.	52
Figura 18. Representação da estrutura tridimensional completa do HIF2 de camundongo, referente ao cristal identificado pelo PDB ID: 4ZPK.....	53

Figura 19. Alinhamento das estruturas tridimensionais do modelo 8 gerado por homologia do HIF1 e o cristal do HIF2 (PDB ID: 4ZPK).	54
Figura 20. Avaliação da estrutura quaternária do HIF1 humano obtida por modelagem molecular por homologia.	57
Figura 21. Gráfico gerado pelo ERRAT com os valores de erro por resíduo com os aminoácidos de regiões de loops selecionados para otimização para a subunidade α , o HIF1a, e para a subunidade β , o ARNT do complexo HIF1	58
Figura 22. Gráfico gerado pelo ERRAT com os valores de erro por resíduo do modelo otimizado da estrutura quaternária do HIF1	59
Figura 23. Representação tridimensional do modelo final para a estrutura quaternária do complexo HIF1 (HIF1a-ARNT) humano	61
Figura 24. Comparação do domínio de ligação ao DNA, bHLH, entre as proteínas que se ligam ao ARNT.	62
Figura 25. Representação tridimensional do fator de transcrição HIF1 humano formado pelas proteínas HIF1a e ARNT	63
Figura 26. Análise das características físico-químicas do domínio de ligação do fator de transcrição HIF1.	64
Figura 27. Valores do Desvio Quadrático Médio (RMSD) e do Raio de giro (RoG) ao longo de 200 ns para o fator de transcrição HIF1 (HIF1a-ARNT) modelado por homologia após o refinamento da estrutura 3D.	66
Figura 28. Valores de RMSF para o fator de transcrição HIF1. Resíduos numerados de 1 a 381 equivalem a subunidade ARNT e os resíduos de 382 a 715 equivalem a subunidade HIF1a.	68
Figura 29. Representação tridimensional do HIF1 com os resíduos de maior flutuação. Os aminoácidos com $RMSF > 3,0\text{\AA}$ estão coloridos	68
Figura 30. Visualização das estruturas capturadas em tempos específicos da trajetória para o modelo refinado do HIF1 e sobreposição das estruturas nos tempos 0, 50, 150 e 200ns	69
Figura 31. Representação da dupla fita de DNA para o HRE considerando o polimorfismo -964 G/A do gene <i>HLA-G</i> .	71
Figura 32. Comparação entre a estrutura tridimensional do complexo referência, o HIF2a-ARNT-HRE (PDB ID: 4ZPK), e as poses agrupadas no cluster 20, referente ao complexo HIF1 (HIF1a-ARNT) -HRE-964G, e no cluster 2, referente ao complexo HIF1 (HIF1a-ARNT) -HRE-964A.	74

Figura 33. Interações intermoleculares entre átomos do HRE e do HIF1 enfatizando o polimorfismo -964G>A do gene <i>HLA-G</i> .	77
Figura 34. Valores do Desvio Quadrático Médio (RMSD) e do Raio de giro (RoG) ao longo de 200 ns para o fator de transcrição HIF1 (HIF1a-ARNT) ligado ao HRE em função do polimorfismo -964G>A do gene <i>HLA-G</i> .	79
Figura 35: Valores do Desvio Quadrático Médio (RMSD) e do Raio de giro (RoG) ao longo de 200 ns para a cadeia de DNA dos complexos HIF1-HRE-964G>A.	81
Figura 36. Valores de RMSF para o fator de transcrição HIF1 ligado ao HRE considerando o polimorfismo -964G>A do <i>HLA-G</i> em comparação com o HIF1 livre e para a cadeia de DNA.	82
Figura 37. Visualização das estruturas capturadas em tempos específicos da trajetória das estruturas nos tempos 0 e 200ns para os complexos HIF1-HRE-964G e HIF1-HRE-964A....	84
Figura 38. Comparação dos valores médios (Å) do shift, slide, e rise calculados para o HRE a partir das coordenadas atômicas extraídas da trajetória dos complexos HIF1-HRE-964G>A a cada 10 ns.	86
Figura 39. Comparação dos valores médios (°) do roll, tilt, e twist calculados para o HRE a partir das coordenadas atômicas extraídas da trajetória dos complexos HIF1-HRE-964G>A a cada 10 ns.	87

LISTA DE TABELAS

Tabela 1. Mecanismos regulatórios pós transcricionais do HIF1a.	6
Tabela 2. Resultado da avaliação de qualidade do modelo 8 do HIF1 gerado pelo MODELLER e seu molde para os seguintes MQAPs.....	55
Tabela 3. Resultado da avaliação de qualidade do modelo inicial HIF1 em comparação com o modelo refinado final e seu molde.	59
Tabela 4. Comparação dos valores para os termos de energia de ligação entre os complexos HIF1-HRE-964G e HIF1-HRE-964A gerados pelo HADDOCK.	75

LISTA DE ABREVIATURAS

Sigla	Nome
3'UTR	<i>3'Untranslated Region</i>
5'URR	<i>5'Untranslated Regulatory Region</i>
AIRs	<i>Ambiguous Interaction Restraints</i>
AMBER	<i>Assisted Model Building with Energy Refinement</i>
ARNT	<i>Aryl Hydrocarbon Receptor Nuclear Translocator</i>
bHLH	<i>basic Helix-Loop-Helix</i>
BSA	<i>Buried surface area</i>
CAP	<i>Catabolic gene Activator Proteins</i>
CBP	<i>CREB-binding protein</i>
CKI	<i>Casein kinase 1</i>
CSP	<i>Chemical shift perturbation</i>
C-TAD	<i>Transcriptional activation domain-C</i>
DFX	Desferrioxamina
DM	Simulação de dinâmica molecular
DNA	<i>Deoxyribonucleic acid</i>
Eair	Energia de violação das AIRs
Edesol	Energia intermolecular de dessolvatação
Eelec	Energia intermolecular eletrostática
EMT	Transição epitelial-mesenquimal
EPAS1	<i>Endothelial PAS domain-containing protein 1</i>
EVdW	Energia intermolecular de Van der Waals
FIH	<i>Factor inhibiting HIF</i>
GROMOS	<i>Groningen Molecular Simulation</i>
GSK3 β	<i>Glycogen synthase kinase 3β</i>
HADDOCK	<i>High Ambiguity Driven DOCKing</i>
HIF1	<i>Hypoxia Inducible Factor 1</i>
HIF1AN	<i>HIF1α subunit inhibitor</i>
HLA-G	<i>Human Antigen Leukocyte-G</i>
HRE	<i>Hypoxia Responsive Element</i>
HS	HADDOCK <i>score</i>

ID	<i>Inhibition domain</i>
IPAS	<i>Inhibitory PAS domain protein</i>
kDa	Quilodalton
LSD1	<i>Lysine-specific demethylase-1</i>
MAPK	<i>Mitogen-activated protein kinase</i>
MHC	<i>Major Histocompatibility Complex</i>
MQAP	<i>Model Quality Assesment Program</i>
mRNA	RNA mensageiro
NK	<i>Natural Killers</i>
NPT	Número constante de partículas
N-TAD	<i>Transcriptional activation domain-N</i>
NVT	Número constante de partículas, volume e temperatura
ODD	<i>Oxygen-dependent degradation</i>
OPLS	<i>Optimized Potentials for Liquid Simulations</i>
PAC	<i>PAS associated C-terminal</i>
PAS	PER-ARNT-SIM
PBC	<i>Periodic boundary condition</i>
PCAF	<i>p300/CBP-associated factor</i>
PDB	<i>Protein Data Bank</i>
PHD	Prolil-hidroxilases
PI3K	<i>Phosphatidylinositol-3-kinase</i>
PLK3	<i>Polo-like kinase 3</i>
PSIPRED	<i>PSI-blast secondary structure PREdiction</i>
QMEAN	<i>Model Energy Analysis</i>
RMSD	<i>Root mean square deviation</i>
RMSF	<i>Root Mean Square Fluctuation</i>
RoG	<i>Ratio of Gyration</i>
ROS	<i>Reactive oxygen species</i>
RSUME	<i>RWD-containing SUMOylation enhancer</i>
SEN1	<i>Sentrin/SUMO-specific protease</i>
SIRT1	<i>Sirtuin 1</i>
SIRT2	<i>Sirtuin 2</i>
SNP	<i>Single nucleotide polymorphism</i>

SPC	<i>Simple point Charge</i>
SPC/E	<i>Extended Simple point Charge</i>
TAD	<i>Temperature-accelerated dynamic</i>
TBP	<i>TATA box binding protein</i>
TF	<i>Transcription factor</i>
TIP4P	<i>For-point transferable intermolecular potentials</i>
VdW	Van der Waals
VHP	Von Hippel-Lindau

SUMÁRIO

Assunto	Página
1. Introdução	1
1.1. O Fator Induzível por Hipóxia-1	2
1.2. <i>HLA-G</i> : gene, estrutura e associação a hipóxia.....	8
1.3. Interação proteína-DNA	12
1.4. Bioinformática estrutural.....	19
2. Hipótese e Justificativa	30
3. Objetivos.....	32
3.1. Objetivo Geral	33
3.2. Objetivos Específicos	33
4. Metodologia.....	34
4.1. Obtenção dos Dados	36
4.2. Predição da estrutura secundária das regiões não resolvidas experimentalmente no molde	37
4.3. Modelagem das regiões não resolvidas experimentalmente no molde.	37
4.4. Modelagem molecular da estrutura quaternária do HIF1	38
4.5. Validação dos modelos	38
4.6. Otimização dos <i>loops</i>	40
4.7. Análise do potencial eletrostático e perfil de hidrofobicidade da superfície.....	41
4.8. Modelagem molecular do HRE localizado na região promotora do gene <i>HLA-G</i> em função do polimorfismo -964G>A	41
4.9. <i>Docking</i> proteína-DNA	42
4.11. Simulação de dinâmica molecular do fator de transcrição HIF1 livre e complexo ao HRE enfatizando o polimorfismo -964G>A do <i>HLA-G</i>	45
4.11. Análise dos parâmetros conformacionais do DNA a partir da trajetória.	47
5. Resultados e Discussão.....	49
5.1. Modelagem da estrutura quaternária do fator de transcrição HIF1	50
5.2. Análise do potencial eletrostático e perfil de hidrofobicidade	60
5.3. Simulação de dinâmica molecular do fator de transcrição HIF1	65
5.4. Modelagem molecular da dupla hélice de DNA.	70
5.5. <i>Docking</i> molecular dos complexos HIF1-HRE-964G/A.....	71

5.6. Análise do padrão de ligações intermoleculares dos complexos HIF1-HRE-964G/A.....	75
5.7. Simulação de dinâmica molecular dos complexos HIF1-HRE-964G/A.....	79
6. Conclusão.....	90
7. Referências...	92
8. Anexo	110
9. Apêndices	145

1 Introdução

1.1. O Fator Induzível por Hipóxia-1

Um suprimento constante de oxigênio deve ser fornecido a todas as células e tecidos do organismo de metazoários para sua sobrevivência contínua. Portanto, é crucial que os níveis de oxigênio sejam mantidos dentro dos limites normais ao longo da vida, visto que situações em que há alteração ou redução na disponibilidade de oxigênio podem resultar em disfunção ou morte celular pelo acúmulo de radicais livres na célula (SAMANTA; PRABHAKAR; SEMENZA, 2017). Um estado de baixo nível de oxigênio celular e tecidual é chamado de hipóxia, cuja ocorrência é desencadeada por diversos fatores, tais como alterações nos mecanismos de transporte de oxigênio, obstrução física do fluxo sanguíneo de uma área determinada ou redução da pressão atmosférica (SEMENZA, 2012).

Tal condição desempenha um papel importante em processos fisiológicos, como na embriogênese e no desenvolvimento placentário (DUNWOODIE, 2009; SHAO; ZHAO, 2014), mas, também, em processos patológicos, principalmente no desenvolvimento de tumores sólidos malignos, como o câncer de mama, cólon e útero (GILKES; SEMENZA, 2013; SEMENZA, 2003; ZHONG et al., 1999). De fato, o nível médio de oxigênio em tecidos normais varia entre 40 a 50 mmHg, enquanto que em neoplasias tais níveis atingem de 5 a 10 mmHg (SEMENZA, 2012). A hipóxia desencadeia respostas celulares a fim de melhorar a oxigenação e sobrevivência da célula tais como aumentar a captação de glicose para produção de energia (BENKERT; KÜNZLI; SCHWEDE, 2009), aumentar o fornecimento de oxigênio ao promover a angiogênese e a eritropoiese (FORSYTHE et al., 1996; HAASE, 2013), e alterar o metabolismo de fosforilação oxidativa para a glicólise anaeróbica para reduzir o consumo de oxigênio (ARAGONÉS et al., 2009).

A expressão de diversos genes é alterada em células expostas à hipóxia, cuja regulação é mediada principalmente pelo fator de transcrição chamado fator induzível por hipóxia 1 (do inglês, *Hypoxia Inducible Factor 1*, HIF1) (NAKAYAMA; KATAOKA, 2019). Este fator de transcrição é considerado um regulador mestre da homeostase do oxigênio e como tal também contribui para o crescimento de tumores, onde nestes casos a maior expressão do HIF1 tem sido relacionada com mau prognóstico em pacientes (SEMENZA, 2010; ZHONG et al., 1999). Dessa forma, ele tem demonstrado seu valor como alvo terapêutico na compreensão da regulação gênica em situações de hipóxia.

O HIF1 é uma proteína de ligação ao DNA do tipo hélice-volta-hélice (em inglês, *basic Helix-loop-helix*, bHLH) da família PER-ARNT-SIM (bHLH-PAS), o qual funciona como um

heterodímero composto por uma subunidade α (HIF1a) regulada por oxigênio e uma subunidade β (HIF1b, também conhecida como translocador nuclear do receptor de hidrocarboneto de arilo, do inglês, *Aryl hydrocarbon receptor nuclear translocator*, ARNT) expressa de forma estável. Em resumo, quando há situações de hipóxia, o HIF1a dimeriza com a subunidade ARNT para estimular a transcrição de seus genes alvos (DENGLER; GALBRAITH; ESPINOSA, 2014).

O gene *Hif1a* humano que codifica a subunidade α do HIF1 está localizado na região 2 do braço longo do cromossomo 14 (14q23.2), e produz uma proteína com 826 aminoácidos, com massa molecular de 92 kDa. Já o gene *Arnt*, o qual codifica a subunidade β , está localizado na região 2 do braço longo do cromossomo 1 (1q21.3) e produz uma proteína com 789 aminoácidos com 86 kDa de peso molecular. Cada subunidade apresenta um domínio amino-terminal (N-terminal), o bHLH, essencial para a ligação com o DNA e os domínios PAS (-A, -B e associado ao C-terminal, o PAC) necessários para interações proteína-proteína e dimerização (TIAN; MCKNIGHT; RUSSELL, 1997). Ambas as proteínas apresentam domínios de ativação transcricionais (do inglês, *transcriptional activation domain-N* e *-C*, N-TAD e C-TAD) os quais são importantes para ativação de genes alvos: o HIF1a apresenta dois domínios de transativação, um N-terminal (N-TAD) e outro na região C-terminal (C-TAD) conectados por um domínio inibitório (do inglês, *inhibition domain*, ID), enquanto que o ARNT possui apenas um C-TAD (JIANG et al., 1996, 1997). Além destes domínios, a subunidade α do HIF1 contém o domínio de degradação dependente de oxigênio (do inglês, *oxygen-dependent degradation*, ODD) localizado sobreposto ao domínio N-TAD, responsável pelo controle da estabilidade e da atividade do HIF1a em resposta a tensão de oxigênio presente (HUANG et al., 1998). A Figura 1 mostra a organização dos domínios proteicos das subunidades HIF1a e ARNT do fator de transcrição HIF1 humano.

Figura 1. Representação esquemática dos domínios proteicos que compõem as subunidades HIF1a e ARNT. As subunidades são compostas por vários domínios conservados que estão envolvidos na ligação ao DNA (bHLH), interações proteína-proteína e dimerização (PER-ARNT-SIM -A, -B e domínio C-terminal associado ao PAS, o PAC), degradação dependente de oxigênio (ODD) e ativação transcricional (N-TAD, C-TAD).

Fonte: O autor (2021).

As subunidades HIF1a e ARNT são constitutivamente transcritas e traduzidas independentemente do nível de oxigênio presente na célula. Porém, o ARNT é expresso de forma estável enquanto a estabilidade do HIF1a é dependente de oxigênio (Figura 2): em baixa tensão de oxigênio a subunidade α se acumula na célula e em níveis suficientes de oxigênio celular a proteína é instável e rapidamente degradada. Em condições de normóxia, quando as células estão bem oxigenadas, esta rápida degradação é mediada via hidroxilação dos resíduos de prolina 405 e 564 no interior do domínio ODD do HIF1a pelas enzimas prolil-hidroxilases (PHD) na presença de ferro, oxigênio e 2-oxoglutarato (APPELHOFF et al., 2004; BERRA, 2003; HUANG et al., 1998; SALCEDA; CARO, 1997). A hidroxilação dos resíduos de prolina permite a ligação da proteína Von Hippel-Lindau (VHP), o qual é um substrato de reconhecimento para o complexo E3 ubiquitina ligase, tornando o HIF1a alvo para degradação proteossomal por meio da poli-ubiquitinação da lisina 532 (ILIOPOULOS et al., 1996; OHH et al., 2000).

O HIF1a está sujeito a uma hidroxilação regulatória adicional em normóxia. Uma enzima conhecida como fator inibidor de HIF (do inglês, *Factor Inhibiting HIF*, FIH), ou inibidor da subunidade HIF1a (do inglês, *HIF1 α subunit inhibitor*, HIF1AN), hidroxila a asparagina 803 dentro do domínio C-TAD da subunidade α , interrompendo o potencial de transativação do HIF1 por interferir na ligação de co-ativador CBP/p300 ao C-TAD, o qual forma um grande complexo proteico com o HIF1 que auxilia na ativação de genes alvos (LANDO, 2002; MAHON, 2001).

Figura 2. Mecanismos de ativação e inibição do HIF1 α . Em normóxia, ocorre a hidroxilação dos resíduos de prolina 405 e 564 por enzimas prolil hidroxilases (PHD), possibilitando a ligação da proteína Von Hippel-Lindau (VHL) e posterior ubiquitinação de HIF1 α , o que resulta em degradação proteossomal. A hidroxilase FIH (fator inibidor de HIF1) também desestabiliza o HIF1 α . Em hipóxia, a produção elevada de espécies reativas ao O₂ (ROS) inibe a atividade das PHDs, estabilizando HIF1 α . O HIF1 α estabilizado se associa ao ARNT, que se liga a elementos responsivos à hipóxia (HREs) em genes alvo. Além disso, a ativação transcricional e atividade do HIF1 α está associada às vias MAPK (proteína quinase ativada por mitógeno) e fosfoinositídeo-3-quinase (PI3K).

Fonte: O autor (2021).

Similar a PHD, o FIH é um sensor de oxigênio que também requer 2-oxoglutarato e oxigênio para desempenhar sua atividade, porém, o FIH tem afinidade mais baixa pelo oxigênio do que os PHDs, o que significa que é capaz de manter sua atividade em tensões de oxigênio mais baixas. Dessa forma, quando a hipóxia é moderada, mesmo após a estabilização, o HIF1 α está sujeito à regulação dependente do oxigênio pelo FIH (KOIVUNEN et al., 2004).

Em condições de hipóxia, quando as células são expostas a níveis reduzidos de oxigênio, a atividade de PHD é reduzida levando à estabilização e rápido acúmulo do HIF1 α na célula

ocorrendo, dessa forma, a dimerização da subunidade α com o ARNT, formando o fator de transcrição HIF1. A dimerização leva a translocação para o núcleo e à ligação aos elementos responsivos à hipóxia (HRE) dentro de regiões promotoras de genes alvos para transcrição pelo HIF1 e, também, ao recrutamento de cofatores que auxiliam na transcrição gênica, como o CBP/p300 (SCHÖDEL et al., 2011; SEMENZA et al., 1996).

Além da regulação negativa da atividade do HIF1 por hidroxilases dependentes de oxigênio (PHDs e FIH), o HIF1a também é modulado por mecanismos pós-transcricionais (Figura 2). Em resposta a fatores de crescimento e citocinas (interleucina-1 β) que regulam o HIF1a sob condições normais de oxigênio na célula, as vias de sinalização da MAPK (do inglês, *mitogen-activated protein kinase*, MAPK) e da fosfatidilinositol-3-quinase (do inglês, *phosphoinositide 3-kinase*, PI3K) são ativadas e seus produtos fosforilam resíduos de serina C-terminais no HIF1a aumentando sua atividade (KIETZMANN; MENNERICH; DIMOVA, 2016). Em situações de hipóxia, quando a via de degradação do HIF1a está inibida, a adição de um fosfato nas serinas 641 e 643 pela ERK1 (MAPK) leva ao acúmulo de HIF1 no núcleo e aumento da atividade transcricional de genes alvos (CARRERA et al., 2014; MURUGAN, 2010). Outros sinais de fosforilação regulam negativamente o HIF1 ao diminuir sua atividade e estabilidade, por inibir sua associação com o ARNT e, também, suprimir a via PI3K pela ação das cinases CKI (do inglês, *casein kinase 1*, CKI) e GSK3 β (do inglês, *glycogen synthase kinase 3 β* , GSK3 β) (KALOUSHI et al., 2010; MOTTET et al., 2003). Além destes, outros mecanismos pós transcricionais regulam a atividade e estabilidade do HIF1a (KOYASU et al., 2018), sendo alguns deles descritos na Tabela 1.

Tabela 1. Mecanismos regulatórios pós transcricionais do HIF1a.

Proteína	Regulação	Atividade	Mecanismo
PLK3	Negativa	Estabilidade	Degradação proteossomal independente de PHDs-VHL por fosforilar as serinas 576 e 657 do HIF1a (XU et al., 2010);
SIRT1	Negativa	Transativação	Inibição da interação CBP/p300-HIF1a por desacetilar a lisina 647 do HIF1a (LIM et al., 2010);
PCAF	Positiva	Transativação	Promoção da interação CBP/p300-HIF1a por acetilar a lisina 647 do HIF1a (LIM et al., 2010);
SIRT2	Negativa	Estabilidade	Promoção da interação do HIF1a com PHDs por desacetilar a lisina 709 do HIF1a (SEO et al., 2015);
LSD1	Positiva	Estabilidade	Supressão da degradação proteossomal independente de PHD-VHL por demetilar a lisina 32 do HIF1a (BAEK; KIM, 2016);
RSUME	Positiva	Estabilidade	SUMOilação das lisinas 391 e 477 (BAE et al., 2004; CARBIA-NAGASHIMA et al., 2007);
SENP1	Negativa	Estabilidade	DeSUMOilação das lisinas 391 e 477 promove degradação proteossomal independente do PHDs-VHL (CHENG et al., 2007);

PLK3: *polo-like kinase 3*, SIRT1: *sirtuin 1*, PCAF: *p300/CBP-associated factor*, SIRT2: *sirtuin 2*, LSD1: *lysine-specific demethylase-1*, RSUME: *RWD-containing SUMOylation enhancer*, SENP1: *senptrin/SUMO-specific protease*; PHD: *prolyl-hydroxylase*, VHL: *von Hippel-Lindau*.

Quando há baixo teor de oxigênio celular, outras duas subunidades α podem ser induzidas em mamíferos, além do HIF1 α , e interagir com o ARNT para formação de fatores de transcrição membros da família bHLH-PAS: HIF2 α /EPAS1 (do inglês, *Endothelial PAS domain-containing protein 1*, EPAS1) e HIF3 α (LEE et al., 2004). O HIF1 α e HIF2 α são mais semelhantes estruturalmente dentre as três isoformas (Figura 2), como os domínios PAS e bHLH que apresentam mais de 70% de identidade com regiões específicas de ligação ao DNA conservadas (WU et al., 2015). O HIF3 α tem múltiplas variantes resultantes de *splicing* alternativo, as quais possuem apenas um domínio de transativação N-terminal (N-TAD) e têm se mostrado baixos ativadores transcricionais de genes alvos em comparação com as outras subunidades HIF α (HEIKKILÄ et al., 2011; HUANG et al., 2013; MAKINO et al., 2002). A sua variante mais bem caracterizada, a HIF3 α 4 (também chamada de IPAS, do inglês, *Inhibitory PAS domain protein*, IPAS) não possui nenhum domínio de transativação e atua como um regulador negativo do HIF1 α (MAYNARD et al., 2005; TANAKA et al., 2009).

Figura 3. Isoformas da subunidade HIF α humano. bHLH, domínio de ligação ao DNA (*Basic-helix-loop-helix*); PAS/PAC, domínios PER-ARNT-SIM -A, -B e domínio C-terminal associado ao PAS, o PAC); ODD, domínio de degradação dependente de oxigênio; ID, região inibitória; N-TAD e C-TAD, domínios de ativação transcricional

Fonte: Adaptado de LEE et al. (2004).

Após a translocação do HIF1 para o núcleo, resíduos ligantes ao DNA do domínio bHLH se ligam a sequência *cis* regulatória 5'RCGTG'3 do HRE (onde R é uma Adenina ou Guanina) presente em genes alvos do HIF1. A Figura 4 resume a frequência com que nucleotídeos específicos ocorrem em uma determinada posição nos HRE, baseado na análise de HRE de genes cujo a transcrição é dependente de fatores de transcrição induzidos por hipóxia, como o HIF1. De acordo com tal análise, a sequência consenso conservada no HRE é CGTG (destacado em uma caixa pontilhada vermelha na Figura 4), com preferência para o nucleotídeo A em relação ao G na posição "R" e nucleotídeos específicos ocorrendo com frequência não aleatória posições circundantes ao HRE nas bases de flaqueamento 5', os quais podem interagir com fatores de transcrição (Destacado por uma caixa azul na figura 4) e nucleotídeos com frequência

aleatória, principalmente nas bases de flaqueamento 3' (WENGER; STIEHL; CAMENISCH, 2005). Os HREs normalmente estão localizados em regiões ativadoras (*enhancers*) na região promotora, na região codificadora e na região 3'UTR (do inglês, 3' *Untranslated Region*, 3'UTR) dos genes alvos, os quais atuam em sinergia para que ocorra a transcrição gênica (FORDEL et al, 2004; PESCADOR et al, 2005; YAGHI et al, 2016; ORLANDO et al, 2020). Por fim, uma vez no núcleo, o HIF1 em associação ao CBP/p300 regula positivamente a transcrição de vários genes responsivos à hipóxia envolvidos na angiogênese, transição epitelial-mesenquimal (EMT), reprogramação metabólica e regulação do sistema imune (LI et al., 2018).

Figura 4. A sequência consenso RCGTG (onde R é A ou G) do HRE e nucleotídeos de flaqueamento. A ocorrência relativa das distribuições de nucleotídeos em 180 sequências consenso do HRE é indicada como porcentagem do total. As caixas sombreadas (azul claro) indicam composições nucleotídicas distribuídas não aleatoriamente, que ocorrem principalmente em bases de flaqueamento 5' onde fatores de transcrição induzidos por hipóxia entram em contato com o DNA. A caixa em vermelho indica a sequência consenso RCGTG do HRE.

Fonte: Adaptado de WENGER et al. (2005).

1.2. *HLA-G*: gene, estrutura e associação a hipóxia.

Durante a hipóxia, o HIF1 induz a transcrição de vários genes que codificam proteínas envolvidas na regulação do sistema imunológico (SCHITO; SEMENZA, 2016). O antígeno leucocitário humano-G (do inglês, *Human Antigen Leukocyte-G*, *HLA-G*), o qual codifica uma molécula de HLA não clássica de classe I do complexo principal de histocompatibilidade (do inglês, *Major Histocompatibility Complex*, MHC) (KLEIN; SATO, 2000a) demonstrou ser modulado em condições de hipóxia por meio do HIF1 (MOUILLOT et al., 2007; YAGHI et al., 2016).

De acordo com a anotação funcional do banco de dados do NCBI (do inglês, *National Center for Biotechnology Database*, NCBI), a organização gênica do *HLA-G* é composta por sete íntrons e oito éxons. Porém, a porção final do éxon 7 e o éxon 8 não são traduzidos e, por

esta razão, são considerados a região 3' UTR do RNA mensageiro (mRNA) maduro. O éxon 1 corresponde a parte inicial da região 5'UTR (do inglês, *5' Untranslated Region*, 5'UTR) que se mantém no mRNA maduro; o éxon 2 codifica a porção final da 5'UTR e contém o ponto de início da tradução (no início da região codificadora); os éxons 3, 4 e 5 codificam, respectivamente, os domínios $\alpha 1$, $\alpha 2$ e $\alpha 3$; o éxon 6 codifica o domínio transmembrana e o éxon 7 a cauda citoplasmática. A região 3'UTR é formada pelo segmento que se estende do éxon 7 ao éxon 8. Já no banco de dados IGMT/HLA (do inglês, *International Immunogenetics Database/Human Leukocyte Antigen*, IGMT/HLA), o éxon 1 é anotado funcionalmente como codificador do peptídeo sinal, o éxons 2, 3 e 4 codificam os domínios $\alpha 1$, $\alpha 2$ e $\alpha 3$, respectivamente, o éxon 5 codifica a porção transmembrana e o éxon 6 codifica a cauda citoplasmática (CASTELLI et al., 2014a).

Os transcritos primários do *HLA-G* sofrem edição alternativa, responsável pela produção de sete isoformas de proteínas distintas, sendo quatro ligadas à membrana (HLA-G1 a HLA-G4) e três são proteínas solúveis (HLA-G5 a HLA-G7). Tais isoformas regulam a resposta imune por meio da interação com receptores inibitórios específicos, presentes em diversos leucócitos, como linfócitos T (CD4 e CD8), células apresentadoras de antígenos (macrófagos e células dendríticas), células NK (do inglês, *Natural Killer*; NK), dentre outros. A interação do HLA-G com estes receptores inibitórios podem resultar em dois efeitos: (i) Inibição a curto prazo de células efectoras e da resposta imune: células NK (LE GAL et al., 1999), linfócitos T citotóxicos (CTL) (LE GAL et al., 1999), células dendríticas (HORUZSKO et al., 2001) e células B (HORUZSKO, LE GAL, RITEAU, et al., 1999, NAJI, MENIER, et al., 2014); ou (ii) indução a longo prazo de células supressoras da resposta imune, as células T reguladoras (Treg) (GREGORI; MAGNANI; RONCAROLO, 2009).

Situações de hipóxia que ocorrem tanto em condições fisiológicas como patogênicas estão associadas a uma maior expressão da molécula HLA-G. No início da gestação, a hipóxia é um importante microambiente fisiológico para o desenvolvimento da placenta e formação da interface materno-fetal, onde o aumento da expressão de HLA-G em células endoteliais fetais, citotrofoblasto e no fluido amniótico contribui para o processo de tolerância imunológica do feto pela mãe (BLASCHITZ et al., 2011; HAMMER; HUTTER; DOHR, 1997; KILBURN et al., 2000; KLEIN; SATO, 2000a, 2000b) conferindo proteção imunológica (BLASCHITZ et al., 2011). Em condições não fisiológicas, situações de baixo teor de oxigênio celular e tecidual estão relacionados ao aumento da transcrição e produção de *HLA-G* em diversas linhagens tumorais *HLA-G*-negativas, como melanoma e glioma (CHANG; FERRONE, 2003;

MOUILLOT et al., 2007; WASTOWSKI et al., 2013; YAGHI et al., 2016), dos quais as células tumorais podem usufruir dos mecanismos imunomodulatórios do HLA-G para escapar do sistema imunológico (CURIGLIANO et al., 2013).

O gene *HLA-G* possui regiões regulatórias (5'UTR e 3'UTR) altamente polimórficas que modulam a expressão da molécula através de mecanismos transcricionais e pós-transcricionais (CASTELLI et al., 2014b; DONADI et al., 2011) e tal variabilidade tem sido associada a produção da molécula. A regulação a nível transcricional pode ocorrer através da ligação de fatores de transcrição a elementos regulatórios presentes na região promotora (engloba a região 5'UTR) do gene reprimindo ou estimulando a transcrição. Análises *in silico* da região de 1,4 kb que antecede o ponto de início da tradução ATG identificaram a presença de dois elementos regulatórios responsivos ao HIF1 (HREs) nas posições -242 a -238 pares de bases (do inglês, *base pair*, bp) e -966 a -962 bp na região promotora do *HLA-G*, sendo que o primeiro HRE é não funcional e nenhuma ligação ao HIF1 foi detectada experimentalmente devido a metilação do DNA que ocorre neste HRE (CASTELLI et al., 2014b; CHANG; FERRONE, 2003; YAGHI et al., 2016), e o segundo HRE contém um polimorfismo G>A na posição -964 (rs1632947) (TAN; SHON; OBER, 2005). Denomina-se por polimorfismo variações nucleotídicas no genoma que são encontradas em uma frequência maior que 1% na população.

Alterações nos nucleotídeos dos elementos regulatórios aos quais os fatores de transcrição se ligam poderiam alterar a afinidade de ligação da proteína por um sítio específico. Consequentemente, a alteração na afinidade de ligação de um fator de transcrição por sua sequência reguladora devido à variabilidade na região promotora pode influenciar os níveis de expressão da molécula HLA-G. O *HLA-G* é o mais polimórfico dentre os HLAs de classe I não clássicos: 35 sítios polimórficos foram descritos para a região promotora, compreendendo os nucleotídeos -1305 a -56, dos quais 26 sítios polimórficos (73,4%) são SNPs (do inglês, *Single Nucleotide Polymorphism*, SNP) e a presença de um polimorfismo trialélico na posição -725 (G/C/T) (rs1233334). Para a região promotora, foram inferidos dez haplótipos diferentes com frequência superior a 1% nas populações do projeto 1000Genomes (CASTELLI et al., 2014a).

Em relação ao polimorfismo -964G>A da região promotora do gene *HLA-G*, tanto na população mundial quanto brasileira, o alelo alternativo -964A (0,5014 e 0,5330, respectivamente) é um pouco mais frequente do que o alelo -964G (0,4986 e 0,5060 respectivamente). Apesar disso, dos 23 haplótipos da região promotora inferidos para a

população brasileira, cerca de 30% apresentaram o alelo alternativo na posição -964 na região promotora do gene (CASTELLI et al., 2014a, 2017).

O impacto da variabilidade populacional da região promotora do *HLA-G* em influenciar a ligação do HIF1 em seu elemento regulatório e, portanto, na modulação da expressão do *HLA-G* em condições de hipóxia, foi explorado pela primeira vez por YAGHI et al (2016). Os autores observaram em uma linhagem de células tumorais de glioma (não expressam *HLA-G*) tratadas com desferrioxamina (DFX), um agente mimetizador de hipóxia, que o SNP -964 (G/A) localizado dentro do HRE pode influenciar na resposta a condições semelhantes à hipóxia em sinergia com o HRE presente no éxon 2 (posições +281 a +285 dentro da região codificadora do domínio $\alpha 1$ de acordo com o IGMT/HLA) do gene *HLA-G* pela ligação do HIF1 (Figura 5).

Figura 5. HIF1 α ativa a expressão de *HLA-G* através de HREs localizados no *locus HLA-G*. (A) Representação esquemática da localização dos HREs nos fragmentos de 1,4 kb-5'URR e 1,4 kb-Exon2 do *HLA-G* clonados no vetor pGL3 com o gene da enzima luciferase. Descrição das sequências dos HREs: o HRE -966 contém o polimorfismo natural -964G>A indicado como -966 (G) e -966 (A); o HRE +281 contém um elemento regulatório duplo que foi clonado em suas formas de tipo selvagem (WT) ou com mutação dupla (mutEx2). (B) Ensaio da luciferase em células U251MG transfectadas com vetor contendo qualquer um dos fragmentos, tratada ou não com DFX (400 μ M) por 24h. Os resultados foram normalizados para a atividade da luciferase em células U251MG transfectadas com o vetor do promotor pGL3, e comparados com a atividade da luciferase em células não tratadas (atribuído um valor de 1). Os dados são apresentados como média $\pm$ DP de 3 experimentos independentes em duplicata, e a análise estatística foi realizada usando teste de Wilcoxon (* indica valor de $P<0,05$, ** indica valor de $P<0,01$ e *** indica valor de $P<0,001$). RLU: unidades relativas de luz.

Fonte: YAGHI et al. (2016).

Como descrito por YAGHI *et al.* (2016), apesar do HRE -966 não atuar individualmente na expressão do *HLA-G* em condições de hipóxia pelo HIF1, o polimorfismo existente em seu interior influencia significativamente na produção da molécula quando o HRE -966 atua em conjunto com os HREs localizados no éxon 2, com a atividade observada para células tratadas com DFX apresentando um aumento médio de 5 vezes na presença de -964A e um aumento médio de 10 vezes na presença do alelo -964G. Mediante tais dados, a descrição das interações moleculares existentes entre o HIF1 e o HRE -966 considerando o alelo existente no sítio polimórfico é importante para a compreensão dos mecanismos regulatórios a nível transcricional do gene *HLA-G* pelo HIF1. Porém, a metodologia utilizada por YAGHI *et al.* (2016) não verifica estruturalmente a interação do fator de transcrição com a sequência reguladora na região promotora do gene, observando, por exemplo, a disposição dos átomos da proteína em relação a cadeia de DNA na sequência regulatória em função do polimorfismo existente.

1.3. Interação proteína-DNA

A homeostase nos organismos vivos depende da interação existente entre diversas moléculas, o que torna possível a realização de vários processos biológicos necessários para sobrevivência e funcionamento celular. Neste contexto, as interações proteína-DNA são de extrema importância, visto que processos celulares vitais, como o controle da expressão gênica, representam a interação constante entre proteínas funcionais e sequências de ácidos nucleicos (SI; ZHAO; WU, 2015).

O controle da expressão gênica em eucariotos acontece em diferentes níveis pré e pós transcricionais. Porém, a transcrição de genes é o principal mecanismo de regulação da expressão gênica, onde a informação genética herdada no DNA é transcrita em mRNA (RNA mensageiro) e, posteriormente, traduzido em proteína. Em resumo, uma vez que o gene esteja ativado pela remodelagem da cromatina (KLEMM; SHIPONY; GREENLEAF, 2019), ocorre a ligação do complexo de transcrição basal no promotor localizado na região *TATA box* do gene, sendo este complexo formado pela RNA polimerase II, a proteína de ligação ao *TATA box* (do inglês, *TATA box binding protein*, TBP) e enzimas chamadas fatores gerais de transcrição (TFIIB, TFIID, TFIIE, TFIIH) (BURLEY, 1996; HAHN, 2004). Além do promotor central onde a RNA polimerase II se liga (Figura 6), outras sequências *cis*-regulatórias atuantes de 5 a 10 pares de base (pb) (chamados também de elementos regulatórios) podem ser alvos de fatores

de transcrição específicos (proteínas regulatórias) que controlam a eficiência do início da transcrição. Às vezes um único elemento regulatório não é suficiente para induzir a transcrição de um gene alvo individualmente, sendo necessário outros elementos regulatórios para completar a ativação da transcrição gênica, os quais podem estar localizados em centenas de pb do promotor central, na região 5'UTR ou/e em éxons (NEZNANOV; UMEZAWA; OSHIMA, 1997; ROLFS et al., 1997; SIM; HUI, 1994; YANG; REMMERS; MARCU, 1986).

Figura 6. Representação esquemática do controle transcricional mediado por fatores de transcrição específicos, os quais atuam sobre o complexo de transcrição basal por meio das proteínas mediadoras. Fatores de transcrição (F_{Te}) se ligam aos elementos regulatórios da cadeia de DNA, ativando (F_{Te} ativadores) ou reprimindo (F_{Te} repressores) a transcrição gênica. Posteriormente os F_{Te}s se ligam a proteínas mediadoras (cofatores) que por sua vez atuam sobre o complexo de transcrição basal, induzindo sua alteração para a conformação ativa capaz de reconhecer a região promotora do gene (a proteína TBP liga-se à porção TATA box do promotor) iniciando o processo de transcrição. A linha interrompida indica que o *loop* de DNA formado tem comprimento variável.

Fonte: Adaptado de ALBERTS et al. (2017).

A interação destes fatores de transcrição específicos com proteínas mediadoras (cofatores) induzem a formação de um *loop* na cadeia de DNA permitindo sua aproximação com o complexo de transcrição basal o que resulta na interação entre estes a fim de estimular (ativadores) ou reprimir (repressores) a transcrição gênica (Figura 6) (KOLOVOS et al., 2012; WEAKE; WORKMAN, 2010). As proteínas mediadoras não se ligam diretamente ao DNA, elas se ligam a um ou mais fatores de transcrição específicos e, assim, funcionam como ponte de conexão destes fatores de transcrição com o ponto de início da transcrição a partir da sua interação com componentes do complexo de transcrição basal (AMOUTZIAS et al., 2008). Por

sua vez, um mesmo fator de transcrição pode induzir a ativação transcricional de genes diferentes caso eles tenham o mesmo elemento regulatório, como ocorre com diversos genes induzidos pelo HIF1 em situações de hipóxia. Fatores de transcrição ativadores transcricionais também recrutam enzimas modificadoras de histonas e complexos de remodelagem da cromatina resultando em alterações locais na estrutura da cromatina para permitir um maior acesso ao DNA, o que facilita a montagem dos fatores de transcrição gerais no promotor (ALBERTS et al., 2017).

A interação entre fatores de transcrição e ácidos nucleicos foi observada pela primeira vez na década de 80, onde a interação entre a proteína ativadora CAP (do inglês, *Catabolic gene Activator Proteins*, CAP), a proteína repressora Cro-434 e o repressor λ de bactéria com a molécula de DNA foram os primeiros complexos proteína-DNA detectados por cristalografia de raio-X (ANDERSON et al., 1981; MATTHEWS, 1988; MCKAY; STEITZ, 1981). A disponibilidade de estruturas de complexos proteína-DNA no banco de dados de estruturas tridimensionais resolvidas experimentalmente, o *Protein Data Bank* (PDB), nos anos seguintes aumentou consideravelmente e, atualmente, há 9.062 estruturas 3D de complexos proteína-DNA depositadas (Figura 7). Com isso, pesquisadores começaram a estudar as características importantes na interação proteína-DNA a partir das estruturas depositadas no PDB, a fim de compreender os mecanismos de reconhecimento da dupla hélice de DNA pelas proteínas.

Figura 7. Número total de estruturas tridimensionais de complexos proteína-ácidos nucleicos depositadas no PDB por ano.

Fonte: RCSB (2021).

O DNA é um longo polímero o qual é formado por uma fita dupla (dupla hélice) complementar (Figura 8), composta por unidades repetidas chamadas nucleotídeos. Por sua vez, os nucleotídeos são formados por uma das 4 bases nitrogenadas existentes (Adenina, “A”, guanina, “G”, citosina, “C” ou timina, “T”), um açúcar de cinco carbonos (desoxirribose) e um grupo fosfato, sendo que o açúcar e o fosfato se unem formando o esqueleto de açúcar-fosfato. As bases nitrogenadas são moléculas planares divididas em dois grupos: purinas (A e G, representadas por Y), as quais são formadas por dois anéis aromáticos, e pirimidinas (T e C, representadas por R) as quais são formadas por um anel aromático. De acordo com Watson e Crick, as duas fitas que constituem o DNA se enrolam sobre a outra em espiral e unem-se através de pontes de hidrogênio que se formam entre as bases nitrogenadas. Tal pareamento das fitas ocorre de maneira complementar, ou seja, sempre ocorrerá da união de uma base pirimidina e uma base purina: adenina pareia apenas com timina e guanina apenas com citosina (Figura 8). Os nucleotídeos dentro de cada fita são conectados pelo esqueleto-açúcar-fosfato por meio de ligações chamadas fosfodiéster. A conformação das espirais unidas resulta em assimetria da dupla hélice de DNA em sulco maior e menor (Figura 8), o qual é essencial para processos biológicos, como o controle transcricional de um gene, pois estes dependem do reconhecimento de sequências específicas do DNA por proteínas para ocorrer, e tal reconhecimento é facilitado pela maior exposição das bases no sulco maior da cadeia (ALBERTS et al., 2017).

Figura 8. Estrutura da dupla-hélice de DNA. A estrutura das quatro bases nitrogenadas e o pareamento das bases segundo Watson-Crick são mostradas à esquerda. À direita é mostrada a dupla-hélice de DNA em duas representações diferentes mostrando os sulcos maior e menor na parte externa da dupla-hélice. A primeira é a representação do desenho da dupla-hélice e a segunda é a representação dos átomos coloridos de acordo com cada tipo de átomo, com os átomos de carbono em verde, nitrogênio em azul, oxigênio em vermelho, fósforo em laranja.

Fonte: O autor (2021).

Este reconhecimento tem sido visto por duas perspectivas principais: (i) o contato direto entre pares de átomos de aminoácidos e bases nitrogenadas, que é responsável por interações intermoleculares e é denominado como mecanismo de leitura direta ou reconhecimento intermolecular, e (ii) o contato indireto através de mudanças conformacionais ou alterações na flexibilidade do DNA, que é denominado como mecanismo de leitura indireta ou reconhecimento intramolecular.

O mecanismo de leitura direto, refere-se à interação direta entre os aminoácidos e as bases nitrogenadas no sulco maior e sulco menor da cadeia de DNA, onde a porção exterior da dupla-hélice de DNA apresenta informação suficiente para o reconhecimento da sequência nucleotídica pelos fatores de transcrição: a extremidade de cada par de bases possui um padrão específico de doadores de ligações de hidrogênio e aceptores de ligações de hidrogênio em ambos os sulcos maior e menor reconhecidos por aminoácidos polares (Figura 9A) e, também, pela formação de interações de Van der Waals (VdWs) entre pares de nucleotídeos e aminoácidos (MURUGAN, 2010). As interações que ocorrem entre as bases nitrogenadas e as cadeias laterais dos aminoácidos são responsáveis pela especificidade da ligação, mas os fatores de transcrição também formam pontes de hidrogênio com átomos do esqueleto açúcar-fosfato da cadeia de DNA a fim de estabilizar e orientar o complexo molecular (Figura 9B). Embora não haja um código padrão de reconhecimento entre as cadeias laterais dos aminoácidos e as bases do DNA, alguns pares aminoácidos-bases ocorrem com mais frequência, como arginina com guanina, e asparagina e glutamina com adenina (LUSCOMBE, Nicholas M, THORNTON, 2002, MATTHEWS, 1988, PABO, SAUER, 1983).

A presença de aminoácidos carregados positivamente nos domínios proteicos dos fatores de transcrição que interagem com o DNA favorece a ocorrência de interações eletrostáticas devido ao esqueleto açúcar-fosfato, que confere ao DNA carga global negativa. Apesar desse tipo de interação não ser específica, elas desempenham papel crucial na orientação da proteína para formação de interações intermoleculares com o DNA, as quais são responsáveis pela especificidade de ligação do complexo molecular por meio dos mecanismos de leitura de base (direto) (SOUSA; CRUZ; QUEIROZ, 2010; SVINGEN; TONISSEN, 2006).

O mecanismo de leitura direto não é o único responsável pela especificidade na interação proteína-DNA, as mudanças conformacionais do DNA também são importantes no reconhecimento do complexo. A arquitetura de uma dupla hélice de DNA pode ser descrita por parâmetros derivados da localização espacial das bases, enquanto que a estrutura do esqueleto açúcar-fosfato não é levada em consideração. Estes parâmetros foram definidos na convenção

de Cambridge (DICKERSON, 1989) e compreendem ângulos e distâncias entre as bases de um par e entre pares de bases adjacentes na dupla hélice (chamado *base pair step*) oriundos de movimentos de translação e rotação (OLSON et al., 1998).

Figura 9. Mecanismos de reconhecimento das bases nitrogenadas pelos aminoácidos. (A) As quatro possíveis combinações de pares de bases, com os possíveis doadores de ligações de hidrogênio indicados em azul, os possíveis aceptores de ligações de hidrogênio indicados em vermelho, e as ligações de hidrogênio ilustradas como uma série de pequenas linhas paralelas vermelhas, os átomos de hidrogênio que estão ligados a carbonos e, portanto, não estão disponíveis para formar ligações de hidrogênio estão em branco. (B) Ligação de um fator de transcrição a um elemento regulatório específico. À esquerda, um único contato é mostrado entre a proteína e o DNA, o qual permite à proteína reconhecer a sequência do DNA. À direita, o conjunto completo de contatos entre o fator de transcrição e sua sequência regulatória cis-atuante. Ainda que as interações no sulco maior sejam as mais importantes, a proteína também estabelece contatos com o sulco menor e com os fosfatos da cadeia principal de açúcar-fosfato do DNA.

Fonte: Adaptado de ALBERTS et al. (2017).

Os parâmetros conformacionais da estrutura do DNA são definidos em um sistema de coordenadas local anexado a cada par de bases individuais, ou em relação a um eixo curvilíneo, dado por dois pares de bases adjacentes (dinucleotídeos). Os parâmetros que descrevem a orientação e posição dos pares de bases individualmente são chamados de intra-pares de base (em inglês, *intra base pair*) definidos por três parâmetros de translação (*shear*, *stretch* e *stagger*)

e três parâmetros de rotação (*buckle*, *propeller* e *opening*) (Figura 10) (DICKERSON, 1989; LAVERY; SKLENAR, 1989):

- *Shear*, *stretch* e *stagger* são deslocamentos de uma base em relação a outra, como cisalhamento ao longo do eixo x (*shear*), como uma distensão ao longo do eixo y (*stretch*), ou um escalonamento abaixo ou acima do plano xy (*stagger*);
- *Buckle*, *propeller* e *opening*: movimentos de rotação de uma base em relação a outra em torno dos eixos x , y e z , respectivamente.

Figura 10. Parâmetros conformacionais empregados para descrever a geometria da dupla-hélice de DNA. À esquerda são mostrados os parâmetros intra-pares de base e a direita os parâmetros inter-pares de bases. No topo da Figura está o sistema de coordenadas referência. As orientações e posições dos referentes parâmetros são referentes a valores positivos.

Fonte: Adaptado de VERLI (2014).

Os parâmetros que descrevem a orientação e posição de dinucleotídeos, ou seja, dois pares de bases empilhados em sequência (também chamado *step*) são chamados de inter-pares de base (em inglês, *inter base pair step*). Assim, os graus de liberdade de um par de bases, o qual é tratado como um corpo rígido, é caracterizado por 6 coordenadas relativas ao par de bases anterior em referência ao seu vizinho, também definidos por três parâmetros de translação

(*shift*, *slide* e *rise*) e três parâmetros de rotação (*tilt*, *roll*, e *twist*) (Figura 10) (DICKERSON, 1989; LAVERY; SKLENAR, 1989):

- *Shift*, *slide* e *rise*: são deslocamentos de um par de base em relação a outro par de base, como deslizamento em direção ao eixo x (*shift*), como um deslizamento ao longo do eixo y (*slide*), ou uma elevação em direção ao eixo z (*rise*);
- *Tilt*, *roll* e *twist*: movimentos de rotação de um par de base em relação ao seu par de base vizinho em torno dos eixos *x*, *y* e *z*, respectivamente.

Em resumo, estes parâmetros são um conjunto completo de descritores geométricos da dupla-hélice. Desvios nestes parâmetros ocorrem principalmente nos seis parâmetros *base-pair step shift*, *slide*, *rise*, *tilt*, *roll* e *twist*, após a formação do complexo proteína-DNA. Além disso, a ligação da proteína ao DNA também resulta em mudanças na geometria global da dupla hélice de DNA formando distorções/curvaturas na cadeia de DNA ao longo de vários pares de bases (DICKERSON, 1989; DICKERSON; CHIU, 1997; LAVERY; SKLENAR, 1989).

1.4. Bioinformática estrutural

A bioinformática estrutural é uma das duas grandes vertentes da bioinformática que aborda questões biológicas a partir de um ponto de vista tridimensional. A bioinformática clássica está relacionada principalmente com sequências de nucleotídeos e aminoácidos, com aplicações em análises individuais ou de populações a partir do estudo do genoma. Já a bioinformática estrutural está relacionada com a estrutura, a fim de prever a estrutura tridimensional das moléculas e sua relação com a atividade funcional da mesma, além de analisar como as diferentes moléculas (proteínas, DNA e RNA) se relacionam (CHOU, 2012). Nas seguintes seções serão abordados algumas das técnicas utilizadas neste trabalho.

1.4.1. Modelagem molecular de proteínas

A forma estrutural que uma proteína adquire é muito importante para sua função. Assim, para entender como uma proteína obtém sua forma ou conformação final, é necessário conhecer os quatro níveis de estrutura distintas que ela apresenta (SUN; FOSTER; BOYINGTON, 2004):

- Estrutura primária: A estrutura primária de uma proteína é a sequência linear de aminoácidos que forma a molécula, a qual é a base da identidade de uma proteína, pois a modificação de apenas um único aminoácido, altera a estrutura primária e

cria uma proteína diferente. Esta proteína diferente pode ser inativa ou ter outra função biológica.

- Estrutura secundária: A estrutura secundária de uma proteína é gerada através de ligações de hidrogênio entre os aminoácidos da cadeia principal da proteína formando organizações espaciais características que podem ser principalmente hélices- α ou folhas- β . Em uma hélice- α , o padrão de ligação entre os aminoácidos leva a cadeia polipeptídica a uma estrutura helicoidal que se assemelha a uma fita enrolada, sendo que cada volta da hélice pode conter até 10 aminoácidos. Já em uma folha- β , dois ou mais segmentos de uma cadeia polipeptídica se alinham uns próximos aos outros, formando uma estrutura parecida com uma folha dobrada, a qual é mantida por ligações de hidrogênio.
- Estrutura terciária: A partir do enovelamento das estruturas secundárias, temos a formação da estrutura terciária da proteína, que nada mais é que a ligação entre os elementos secundários formando a estrutura tridimensional final da proteína.
- Estrutura quaternária: A estrutura quaternária é a união de várias estruturas terciárias que assumem formas espaciais bem definidas.

A estrutura tridimensional de uma proteína pode ser obtida de forma experimental ou teórica. Dentre os métodos experimentais utilizados para resolução de estruturas 3D estão a Cristalografia de Raio-X, a ressonância magnética e microscopia crio-eletrônica. No entanto, a determinação da estrutura 3D por meio de técnicas experimentais é um processo longo e caro, em que a macromolécula alvo deve ser produzida em quantidades relativamente grandes e purificada em altas concentrações. Por exemplo, na cristalografia de raios-X, que é de longe a abordagem mais utilizada para a determinação da estrutura de macromoléculas, o alvo deve ser cristalizado e o cristal resultante deve difratar para resolução suficiente. Não surpreendentemente, apenas uma pequena porcentagem das tentativas iniciais de determinação da estrutura é bem-sucedida. Uma falha em qualquer uma das muitas etapas que vão desde a seleção de um alvo até a determinação de sua estrutura inevitavelmente atrasa ou até mesmo interrompe o processo. Esses fatores aumentam o custo geral da determinação da estrutura, tanto em materiais quanto em tempo (SLABINSKI et al., 2007). Dessa forma, uma alternativa as técnicas experimentais é a utilização de técnicas com metodologia teórica, ou também chamadas *in silico*.

Três abordagens computacionais podem ser empregadas para prever a estrutura tridimensional *in silico* de uma proteína de interesse (alvo): modelagem por homologia

(comparativa), *threading* e *ab initio*. Tais métodos diferem quanto a utilização das informações disponíveis nos bancos de dados de proteínas com estruturas tridimensionais resolvidas por métodos experimentais, como o PDB, sendo a modelagem por homologia mais dependente dessas informações, enquanto que a modelagem por *ab initio* é totalmente independente (Figura 11) (KUHLMAN; BRADLEY, 2019). A modelagem comparativa utiliza métodos de compatibilidade sequência-sequência, assim, parte do princípio de que a semelhança entre as sequências de aminoácidos da proteína e de proteínas homólogas com estruturas tridimensionais resolvidas implica em similaridade estrutural entre elas (GUPTA; MOHANTY; BHATNAGAR, 2016; SANTOS FILHO; ALENCASTRO, 2003). A modelagem por *threading* busca identificar proteínas estruturalmente similares, sem detectar similaridade entre as sequências, isto é, utiliza métodos de compatibilidade sequência-estrutura, onde é feito a comparação da sequência da proteína alvo contra um banco de dados com proteínas de estruturas 3D resolvidas, e o melhor enovelamento é escolhido por meio de um potencial empírico derivado de uma tabela de contatos de resíduos observados em proteínas com estruturas resolvidas experimentalmente (ROST; SANDER, 1996). As abordagens de modelagem *ab initio* podem ser aplicadas a proteínas sem similaridade estrutural com proteínas do PDB. Na falta de um molde, esse método requer uma estratégia de amostragem conformacional para gerar modelos candidatos, bem como um critério de classificação pelo qual os modelos podem ser selecionados (XU; ZHANG, 2012).

Figura 11. Relação entre as abordagens computacionais para predição de estruturas 3D de proteínas e o uso de informações provenientes de estruturas resolvidas experimentalmente depositadas no PDB. A aplicabilidade de cada técnica depende da taxa de identidade entre as sequências de aminoácidos da proteína-alvo e de proteínas com estrutura resolvida experimentalmente.

Fonte: O autor (2021).

A técnica de modelagem molecular comparativa foi a escolhida para a predição da estrutura tridimensional do fator de transcrição HIF1. Para a geração do modelo, a modelagem molecular por homologia é composta por quatro etapas: busca e seleção dos moldes, alinhamento do molde com o alvo, construção e avaliação dos modelos. Primeiro, é necessário realizar a busca por moldes em um banco de dados de estruturas de proteínas com estrutura

tridimensional resolvida, como o PDB, utilizando como parâmetro de entrada a estrutura primária da proteína alvo cuja a estrutura tridimensional não é conhecida. Assim, a sequência de aminoácidos da proteína alvo será alinhada com possíveis sequências homólogas de estruturas conhecidas depositadas no PDB (moldes). Dessa forma, se obtém uma lista com potenciais moldes usando métodos de busca para a seleção de bons moldes para a modelagem molecular em questão. Geralmente, os moldes que possuem identidade mais elevada, ou seja, porcentagem mais alta de aminoácidos idênticos e menor número de *gaps* no alinhamento são mais apropriados. Porém, para a construção de complexos moleculares, como complexos proteína-DNA, a escolha do molde ligado a uma molécula de DNA é provavelmente mais importante que a resolução do modelo (GUPTA; MOHANTY; BHATNAGAR, 2016). Com a seleção do molde, é feito um alinhamento do molde com a proteína alvo. A partir do alinhamento são extraídos as restrições espaciais para geração dos modelos na modelagem comparativa, assim, é crucial que o alinhamento entre o alvo e o molde esteja correto (ŠALI et al., 1993). É necessário que as estruturas primárias das proteínas tenham identidade superior a 30% para que o alinhamento seja preciso, pois alinhamentos abaixo deste valor apresentam muitas falhas com grandes extensões de *gaps* e erros nos alinhamentos.

O modelo final deve ser avaliado para checar se os parâmetros físico-químicos e estereoquímicos estão dentro do esperado. A avaliação da qualidade do modelo pode ser feita como um todo ou em regiões individuais da proteína baseado na similaridade entre as sequências do molde e do alvo, observando resíduos importantes em regiões específicas da proteína como sítios de ligação e sua conservação. Um requerimento básico para um modelo é ter uma boa qualidade estereoquímica, o qual é avaliada por meio do gráfico de Ramachandran. Neste gráfico, todas as combinações possíveis para os ângulos de torção ϕ (phi) e ψ (psi) nos aminoácidos da cadeia polipeptídica estão mapeados. Estes ângulos são importantes na definição do enovelamento da proteína, sendo o ângulo ϕ localizado anterior ao $C\alpha$, e ângulo ψ localizado após o $C\alpha$ no aminoácido. Assim, em função das cadeias laterais de cada aminoácido, algumas combinações de ângulos *phi-psi* seriam favorecidas, enquanto outras proibidas, sendo que é esperado que um bom modelo apresente 90% do total de seus aminoácidos em regiões favoráveis para tais ângulos (LASKOWSKI et al., 1993; RAMACHANDRAN; RAMAKRISHNAN; SASISEKHARAN, 1963). Além deste, existe métodos para testar modelos 3D que implicitamente apresentam características espaciais compiladas de estruturas 3D provenientes de proteínas de alta resolução. Estes métodos se baseiam nos perfis 3D e potenciais estatísticos derivados de campos de força a fim de avaliar o ambiente químico de

cada resíduo em um modelo a respeito do ambiente químico esperado como aqueles identificados em estruturas geradas por cristalografia de raios-X de alta resolução (COLOVOS; YEATES, 1993; LÜTHY; BOWIE; EISENBERG, 1992; WIEDERSTEIN; SIPPL, 2007).

1.4.2. *Docking* molecular proteína-DNA

A abordagem de *docking* molecular pode ser usada para modelar a interação entre uma pequena molécula e uma proteína em nível atômico a fim de caracterizar o comportamento de pequenas moléculas no sítio de ligação de proteínas alvo, bem como elucidar processos bioquímicos fundamentais (MENG, X. Y., ZHANG, H. X., MEZEI, M., & CUI, 2011). A dificuldade neste processo varia no tratamento das moléculas parceiras, ou seja, se ambas as moléculas são mantidas rígidas, ou apenas a proteína é mantida rígida, ou se ambas são consideradas flexíveis. Na inclusão de flexibilidade na molécula, o aumento dos graus de liberdade dos ângulos diedrais internos são considerados, o que aumenta o custo computacional para otimização dos átomos em questão e o cálculo dos termos de energia de ligação intermolecular das moléculas. Apesar de aumentar a complexidade do processo de *docking*, a adição da flexibilidade é importante pois simula de maneira mais realística a formação do complexo molecular (SADJAD; ZSOLDOS, 2011).

No caso do *docking* de complexos proteína-DNA, a flexibilidade é importante devido a molécula de DNA ser naturalmente flexível. Dessa forma, o DNA frequentemente exhibe grandes mudanças conformacionais ao se ligar a uma proteína, o que pode alterar muito a forma da superfície de interação entre as duas moléculas. Com isso, o espaço conformacional total que precisa ser pesquisado para encontrar conformações favoráveis torna-se ainda maior (PABO; NEKLUDOVA, 2000). A flexibilidade no DNA é resultante de uma combinação de mudanças conformacionais nos pares de bases e na estrutura do esqueleto açúcar-fosfato (TRAVERS, 2004). No entanto, permitir a flexibilidade no DNA durante o processo de *docking* de complexos proteína-DNA enquanto ocorre a formação do complexo é um grande desafio.

Felizmente, Bonvin e colaboradores conseguiram driblar este problema em seu *software* HADDOCK (High Ambiguity Driven DOCKing), o qual é gratuito e amplamente utilizado pela comunidade acadêmico-científica. (DOMINGUEZ; BOELEN; BONVIN, 2003b; VAN DIJK et al., 2006b). O HADDOCK faz uso de dados experimentais disponíveis na literatura acerca da superfície de interação das moléculas parceiras para conduzir o processo de *docking*, tais dados podem ser derivados de experimentos de mutagênese, CSP (do inglês, *chemical shift*

perturbation, CSP), dentre outros, e também, na ausência de dados experimentais, pode ser utilizado abordagens de predição *in silico* para identificação de resíduos importantes na interação. Estes dados brutos são traduzidos em restrições de comprimento, chamadas de restrições de interação ambíguas (do inglês, *Ambiguous Interaction Restraints*, AIRs) que são incorporadas na função de energia usada nos cálculos. Tais AIRs são definidas por meio de uma lista de resíduos descritos como resíduos ativos e passivos, sendo que os resíduos ativos são aqueles importantes para que a formação do complexo molecular ocorra, enquanto que os resíduos passivos geralmente são os resíduos vizinhos aos ativos que potencialmente fazem contato dentro do complexo. Esta abordagem reduz a busca necessária através do espaço de interação e aumenta a fração de poses únicas (DE VRIES; VAN DIJK; BONVIN, 2010; NILGES et al., 1988).

O protocolo de *docking* do HADDOCK consiste em três estágios: (i) *docking* de corpo rígido e minimização de energia (it0); (ii) refinamento semi flexível dos ângulos de torção (it1) e (iii) refinamento em solvente explícito (W) (DE VRIES; VAN DIJK; BONVIN, 2010; DOMINGUEZ; BOELEN; BONVIN, 2003; VAN ZUNDERT et al., 2016). A adição de flexibilidade no DNA é feita por meio da amostragem de conformações do esqueleto açúcar-fosfato e os pares de bases do DNA durante o estágio de refinamento semi flexível (it1) e, assim, o *docking* é feito a partir de uma biblioteca de estruturas de DNA recém geradas representando vários graus de flexibilidade conformacional (VAN DIJK et al., 2006b). A flexibilidade na proteína também é introduzida neste estágio, primeiro ao longo das cadeias laterais na interface e, em seguida, para a cadeia principal e as cadeias laterais neste mesmo estágio (DOMINGUEZ; BOELEN; BONVIN, 2003b).

No estágio inicial (it0), as moléculas parceiras são tratadas como corpos rígidos, o que significa que todos os parâmetros estruturais, como comprimentos de ligação, ângulos de ligação e ângulos diédricos, estão estáticos. As moléculas são separadas no espaço com permissão para movimentos de rotação, seguido por uma etapa de minimização de energia de corpo rígido, onde os movimentos de translação e rotação são permitidos a fim de otimizar a superfície interação. O papel dos AIRs nesta fase é de particular importância. Uma vez que eles estão incluídos na função de energia que está sendo minimizada, os complexos resultantes serão viesados em direção a eles. Por exemplo, definir um conjunto muito restrito de AIRs leva a uma amostragem muito limitante do espaço conformacional, o que significa que as poses geradas serão muito semelhantes. Por outro lado, restrições muito dispersas (por exemplo, toda a superfície de um parceiro) resultarão em soluções muito diferentes, exibindo maior

variabilidade na região de ligação (DE VRIES; VAN DIJK; BONVIN, 2010; DOMINGUEZ; BOELEN; BONVIN, 2003).

O segundo estágio do protocolo do *docking* (it1) introduz flexibilidade para os parceiros de interação por meio de um refinamento baseado em dinâmica molecular de três etapas, a fim de otimizar a interface de interação. Na primeira etapa, as moléculas são mantidas rígidas e apenas suas orientações são otimizadas. A flexibilidade é então introduzida na interface, que é automaticamente definida com base em uma análise de contatos intermoleculares dentro de um corte de 5 Å. Isso permite que diferentes poses de ligação provenientes de it0 tenham diferentes regiões flexíveis definidas. No estágio final (W), o complexo é submerso em uma concha de solvente explícito para melhorar a energia da interação. Neste refinamento de solvente explícito curto, os modelos são submetidos a uma simulação de dinâmica molecular curta em 300K de temperatura, com restrições de posição nos átomos não pesados da interface, os quais são relaxados posteriormente para otimização das moléculas (DE VRIES; VAN DIJK; BONVIN, 2010; DOMINGUEZ; BOELEN; BONVIN, 2003).

O número de poses geradas diminui a cada etapa do protocolo de *docking* como se fosse um filtro. Tais poses são classificadas de acordo seu HADDOCK score (HS), uma função de *score* que consiste em uma combinação linear de vários termos de energias, como energia eletrostática, energia de VdWs, dentre outros, a qual pode apresentar peso diferente para estes termos de energia em cada uma das etapas do *docking* (DOMINGUEZ; BOELEN; BONVIN, 2003b). Esses termos de energia são usados para delinear as poses corretas de poses incorretas geradas no *docking* através da estimativa da afinidade de ligação entre as moléculas ordenando as diferentes poses de acordo com suas afinidades (MENG et al., 2011), sendo as melhores poses geradas em cada etapa do processo de *docking* são as que apresentam HS mais negativo (DE VRIES; VAN DIJK; BONVIN, 2010; KITCHEN et al., 2004)(DE VRIES; VAN DIJK; BONVIN, 2010a; KITCHEN et al., 2004).

1.4.3. Simulação de dinâmica molecular clássica

A simulação de dinâmica molecular (DM) é muito utilizada para o estudo da dinâmica de macromoléculas biológicas. Tal metodologia fornece informações sobre o comportamento dinâmico microscópico dos átomos que compõem o sistema ao longo do tempo. Dadas as posições de todos os átomos em um sistema biomolecular, pode-se calcular a força exercida em cada átomo por todos os outros átomos utilizando as leis de movimento de Newton para prever

a posição espacial de cada átomo em função do tempo. Por sua vez, os cálculos das forças que atuam sobre cada átomo são feitos calculando-se a função de energia potencial, obtida a partir de um campo de força, em relação as posições desses átomos. O campo de força é uma função de energia potencial empírica que inclui tipos de interações não-ligadas, como forças eletrostáticas e Lennard-Jones, mas também vários outros potenciais harmônicos resultantes de distribuições estatísticas derivadas de estruturas proteicas de resolução ótima (1 a 2 Å de resolução, por exemplo) para preservar a integridade estrutural do sistema biomolecular, como termos relacionados a comprimento da ligação, ângulo da ligação, alguns ângulos diédricos e ângulos diédricos impróprios (FRENKEL; SMIT, 2002; MONTICELLI; TIELEMAN, 2013).

Ao integrar as equações de movimento se obtém a aceleração da partícula (átomo) e, conseqüentemente, a mudança da posição do átomo. Assim, ao calcular repetidamente as forças em cada átomo e, em seguida, usando essas forças para atualizar a posição e a velocidade de cada átomo é possível avançar no tempo e obter o que se denomina de “trajetória”, que nada mais é do que o conjunto de posições e velocidades de cada átomo ao longo do tempo. A trajetória é obtida na última etapa da dinâmica molecular (chamada de produção) e nela estão contidas todas as coordenadas atômicas do sistema durante toda a simulação, as quais são utilizadas para gerar diversas análises estruturais e físico-químicas. Se trata de um filme que descreve a configuração em nível atômico do sistema em cada ponto durante o intervalo de tempo simulado. (HOLLINGSWORTH; DROR, 2018; NAMBA; SILVA; SILVA, 2008)

Existem muitos tipos de campos de força desenvolvidos por diferentes grupos de pesquisa em todo mundo. Para ilustrar, no pacote de simulação GROMACS, por exemplo, existem vários campos de força disponíveis. Entre os mais conhecidos destacam-se *Groningen Molecular Simulation* (GROMOS) (GUSTEREN VAN; BERENDSEN, 1987), *Assisted Model Building with Energy Refinement* (AMBER) (WEINER; KOLLMAN, 1981), *Optimized Potentials for Liquid Simulations* (OPLS) (JORGESSEN; TIRADO-RIVES, 1988) com parâmetros apropriados para proteínas (HORNAK et al., 2006; MONTICELLI; TIELEMAN, 2013), lipídios (LYUBARTSEV; RABINOVICH, 2016) e diversas outras moléculas orgânicas, como pequenos ligantes e ácidos nucleicos (IVANI et al., 2016; LIN; MACKERELL, 2019; MINHAS et al., 2020), especialmente no papel de solventes no estado líquido.

Assim, em resumo, o sistema a ser simulado consiste em uma coleção de partículas interagentes representadas como átomos, descrevendo tanto o soluto (molécula biológica) quanto o solvente, colocados dentro de uma caixa de simulação (Figura 12) suficientemente grande, onde seus movimentos são descritos pelas leis dos movimentos de Newton. Muitos

modelos de água são usados como solventes: SPC (do inglês, *Simple point Charge*, SPC) (BERENDSEN et al., 1981), TIP4P (do inglês, *for-point transferable intermolecular potentials*, TIP4P) (JORGENSEN; MADURA, 1983) e SPC/E (do inglês, *Extended Simple point Charge*, SPC/E) (BERENDSEN; GRIGERA; STRAATSMA, 1987). No presente trabalho, o modelo adotado foi o solvente explícito SPC/E, o qual é considerado um dos modelos mais bem-sucedidos por sua simplicidade e eficiência computacional e descreve uma molécula de água com três sítios de interação, ou seja, três pontos de carga (um no oxigênio e dois nos hidrogênios).

Figura 12. Representação de uma proteína dentro de uma caixa cúbica com água.

Fonte: BRAY (2021).

As equações de movimento em função do tempo são integradas a partir de algoritmos baseados nos métodos das diferenças finitas, como o algoritmo de *Verlet* (VERLET, 1968) e *leap-frog* (HOCKNEY, 1970), sendo que todos estes algoritmos assumem uma expansão de séries de Taylor utilizando as posições e acelerações dos átomos em função do tempo (BROOKS, 1989; NAMBA; SILVA; SILVA, 2008).

As simulações são geralmente conduzidas em condições de temperatura e pressão constantes, por processos conhecidos como o termostato de Berendsen (BERENDSEN et al., 1984) e barostato de Parrinello-Rahman (PARRINELLO; RAHMAN, 1981), respectivamente, os quais foram utilizados no presente trabalho: (i) para controle da temperatura, o sistema é adicionado a um banho térmico com temperatura fixa, onde as velocidades são re-escaladas a cada passo da integração das equações de movimento durante a simulação, a fim de ajustar a

energia cinética do sistema até a temperatura desejada, enquanto que (ii) para o controle da pressão, as variáveis relacionadas a temperatura são substituídas pelas variáveis relacionadas a pressão e as velocidades dos átomos substituídos pelas coordenadas atômicas (BERENDSEN et al., 1984; PARRINELLO; RAHMAN, 1981).

Para simular um sistema molecular muito grande, várias técnicas para estender artificialmente o seu tamanho foram desenvolvidas. Pois se considerarmos que o sistema seja simulado numa caixa de tamanho finito, ao integrar as equações de movimento, as partículas do sistema interagem podendo bater na parede da caixa de simulação e, até, sair da caixa alterando o número de partículas e, conseqüentemente, a densidade dentro do sistema em questão. Assim, para minimizar tais efeitos no sistema a ser simulado, devem ser usadas condições periódicas de contorno (do inglês, *periodic boundary condition*, PBC), a fim de que o sistema seja simulado como se estivesse numa caixa de tamanho infinito (HOLLINGSWORTH; DROR, 2018; SALO-AHEN et al., 2020). Nesta técnica (Figura 13), os átomos do sistema são colocados em uma caixa (normalmente cúbica), e a caixa original (central) é replicada em todas as direções do espaço em número infinito. Durante as simulações, quando um átomo se move na caixa original, sua imagem periódica, em uma das caixas réplica, move-se da mesma maneira. Se um átomo sai da caixa original, sua imagem entra pela face oposta da caixa réplica, com a mesma velocidade. Assim, o número total de átomos na caixa central e no sistema como um todo é conservado. Isso é importante principalmente na inclusão do efeito de solvatação do sistema a ser simulado (NAMBA; SILVA; SILVA, 2008).

Figura 13. Representação das condições periódicas de contorno em uma caixa cúbica. Apenas a caixa central é simulada (destacada em azul), enquanto que as réplicas garantem a continuidade do sistema, isso é, ausência de contato com o vácuo.

Fonte: Adaptado de Verli et al. (2014).

Após a configuração inicial do sistema, este é minimizado para garantir a inexistência de choques estéricos entre átomos. Além disso, a minimização otimiza a geometria das moléculas, que visa encontrar um conjunto de coordenadas atômicas que diminua a energia potencial do sistema de interesse em direção a um mínimo de energia. Isso é feito por meio de ajustes nas posições atômicas, ângulos de ligações e, também, nos contatos de VdWs (NAMBA; SILVA; SILVA, 2008). O algoritmo de minimização utilizado neste trabalho foi o *steep descent* e é usado para minimizar inicialmente uma estrutura que está longe do ponto de mínimo de energia, fazendo deste um método de minimização robusto (LEACH, 2001).

A minimização garante a obtenção de uma estrutura inicial razoável, em termos de geometria e orientação do solvente. Para começar a dinâmica real, devemos equilibrar o solvente e os íons em torno da proteína. Se tentarmos uma dinâmica desenfreada neste ponto, o sistema pode entrar em colapso. A razão é que o solvente é otimizado principalmente dentro de si mesmo, e não necessariamente com o soluto. Ele precisa ser levado à temperatura que desejamos simular e estabelecer a orientação adequada sobre o soluto (molécula biológica). Depois de chegar à temperatura correta (com base nas energias cinéticas), aplicaremos pressão no sistema até que ele atinja a densidade adequada. O equilíbrio é freqüentemente conduzido em duas fases: a primeira fase é conduzida sob um conjunto NVT (número constante de partículas, volume e temperatura), também chamado de canônico, e a segunda fase é conduzida por um conjunto NPT (número constante de partículas, pressão e temperatura), também chamado de isotérmico-isobárico (BROOKS, 1989). A primeira visa equilibrar a temperatura do sistema, enquanto que a segunda busca estabilizar a pressão e, conseqüentemente, a densidade do sistema. Com o sistema bem equilibrado, a etapa final é chamada de produção, onde a simulação de dinâmica molecular ocorre de fato, e será gerado as trajetórias do sistema para posterior análise (LEMKUL, 2019).

2 Hipótese e Justificativa

Visto que: (i) a interação de proteínas com o DNA é crucial para muitos processos celulares como a transcrição, e (ii) há evidências *in vitro* mostrando que alterações no elemento regulatório responsivo a hipóxia (do inglês, *Hypoxia Inducible Element*, HRE) do fator de transcrição HIF1 (do inglês, *Hypoxia Inducible Factor 1*, HIF1) resulta em diminuição da transativação e transcrição do gene *HLA-G*, a hipótese deste trabalho é que o polimorfismo -964G>A localizado no interior do HRE na região promotora do gene *HLA-G* pode afetar a ligação com o HIF1. Assim, a análise da interação deste fator de transcrição com ação na região promotora do gene considerando a variabilidade populacional do *HLA-G* nos permite uma melhor compreensão dos mecanismos de regulação a nível transcricional e a sua relação com a expressão da molécula, através da observação estrutural da interação física entre os átomos do fator de transcrição e o polimorfismo presente no elemento regulatório.

3 Objetivos

3.1. Objetivo Geral

Por meio de ferramentas de bioinformática estrutural, modelar o fator de transcrição (proteína) de interesse e realizar o *docking* do complexo fator de transcrição e DNA enfatizando a variabilidade populacional da região promotora do gene *HLA-G*.

3.2. Objetivos Específicos

Como objetivos específicos deste projeto destacam-se os seguintes itens:

3.2.1. Análise estática e dinâmica da estrutura quaternária do fator de transcrição HIF1;

3.2.2. Análise estática e dinâmica do complexo proteína-DNA considerando o polimorfismo -964G>A existente no interior do elemento regulatório alvo do fator de transcrição HIF1 na região promotora do *HLA-G*.

4 Metodologia

Na Figura 14 está representado o fluxograma das etapas da metodologia e *softwares* que foram empregados nesse projeto. Todas as etapas foram realizadas por técnicas *in silico*.

Figura 14. Fluxograma das etapas utilizadas na metodologia. A caixa em verde representa os servidores e a caixa em azul representa os programas usados em cada etapa. As etapas correspondentes a análise estática dos objetivos específicos estão destacadas por *, e as etapas correspondentes a análise dinâmica dos objetivos específicos estão destacadas em **. FT: fator de transcrição, PDB *Protein Data Bank*, MQPAs: *Model Quality Assessment Programs*; HS: *Haddock score*.

Fonte: O autor (2021).

4.1. Obtenção dos Dados

A estrutura primária da proteína-alvo foi obtida no banco de dados do UniProt¹ na extensão *fasta*. Visto que o HIF1 é um fator de transcrição formado pela ligação de duas proteínas, a subunidade α , HIF1, e a subunidade β , ARNT, a sequência primária para modelagem destas proteínas foram obtidas a partir do arquivo *fasta* referente ao código Q16665 do UniProt, contendo os resíduos de aminoácidos 15 a 349 para o HIF1a humano (Figura 15A), e P27540 contendo os resíduos 87 a 470, para o ARNT humano (Figura 15B). Uma vez que a metodologia empregada foi a modelagem comparativa, foi feito a busca por moldes no PDB², por meio da ferramenta Mmseqs2³ (STEINEGGER; SÖDING, 2017) por complexos proteína-DNA semelhantes ao fator de transcrição alvo. Dessa forma, a estrutura tridimensional escolhido como melhor molde para o HIF1 foi o fator de transcrição HIF2 de camundongo no formato pdb referente ao código 4ZPK (3,6Å de resolução) (WU et al., 2015), sendo a cadeia A referente ao ARNT, e a cadeia B ao HIF2.

Figura 15. Domínios proteicos selecionados para modelagem do fator de transcrição HIF1. (A) Subunidade α , o HIF1a. (B) Subunidade β (ARNT). Motivos estruturais destacados: domínio de ligação ao DNA *basic Helix-loop-helix* (bHLH) em laranja, o domínio Per-Arnt-Sim S1(PAS1) em rosa e S2 (PAS2) em verde, e o motivo PAC em azul, para as duas proteínas.

Fonte: O autor (2021).

¹ <https://uniprot.org>

² *Protein Data Bank* - <https://www.rcsb.org/>

³ <https://www.rcsb.org/search/advanced/sequence/>

4.2. Predição da estrutura secundária das regiões não resolvidas experimentalmente no molde.

A sequência primária das cadeias A (ARNT) e B (HIF2a) do cristal molde, com PDB ID: 4ZPK, foram extraídas do banco de dados PDB⁴ na extensão *fasta* e submetidas ao servidor PSIPRED v.4.0⁵ (*PSI-blast secondary structure PRediction*) (BUCHAN; JONES, 2019) para predição da estrutura secundária das regiões ausentes do cristal molde para a modelagem molecular por homologia do HIF1. O método corresponde a uma ferramenta de alta acurácia implementada a partir de redes neurais para a predição da estrutura secundária de uma proteína baseado em matrizes de pontuação de posição específicas geradas pelo PSI-BLAST (JONES, 1999). Como arquivo de saída, é utilizado gráficos para a representação dos resíduos de aminoácidos da proteína e a predição da estrutura secundária acompanhado por uma escala de confiança da predição da estrutura.

4.3. Modelagem das regiões não resolvidas experimentalmente no molde.

Antes de continuar com a modelagem molecular, o molde HIF2 (PDB ID: 4ZPK), teve as regiões não resolvidas no cristal preenchidas pela ferramenta *PDB Reader* do servidor CHARMM-GUI (JO et al., 2008). (JO, KIM, et al., 2008). O *PDB Reader* usa o algoritmo implementado no *GalaxyFill* para fornecer predições de *loop* de alta acurácia por meio de um refinamento global na estrutura da proteína que emprega uma função de energia híbrida. Esta função de energia combina componentes baseados em física pelos termos correspondentes na função potencial do CHARMM-22 (MACKERREL et al, 1998), e componentes baseados em conhecimento de potenciais estatísticos extraídos da própria estrutura. Assim, os termos de energia baseados em física ajudam a localizar estruturas precisas perto da estrutura nativa enquanto que os termos baseados em conhecimento tendem a suavizar a superfície da energia livre para que a imprecisão do ambiente estrutural possa ser tolerada. Durante a etapa de otimização, a biblioteca de fragmentos é gerada com o algoritmo de fechamento de *loop* triaxial (TLC), o qual é utilizado para manter a integridade estrutural do *loop*. No fim, a estrutura que

⁴ <https://www.rcsb.org/>

⁵ <https://bioinf.cs.ucl.ac.uk/psipred/>

apresentar menor energia livre é seleccionada como modelo final (COUTSIAS, SEOK, et al., 2004, LEE, LEE, et al., 2010).

4.4. Modelagem molecular da estrutura quaternária do HIF1

A modelagem molecular foi feita por homologia por meio do *software* MODELLER 9.24 (WEBB; SALI, 2016). Tal programa implementa a modelagem comparativa de estruturas tridimensionais por meio da satisfação das restrições espaciais (SALI; BLUNDELL, 1993), além disso, ele pode construir modelos de proteínas de várias cadeias a partir de restrições de comprimento existentes no heterodímero molde homólogo ao complexo alvo. As sequências obtidas no UniProt referentes as subunidades α (HIF1a) e β (ARNT) foram alinhadas as sequências primárias das cadeias do cristal 4ZPK (Cadeia A referente ao ARNT, e cadeia referente ao HIF2a) por meio do *software Clustal X* (THOMPSON et al., 1997). O arquivo de alinhamento resultante foi convertido manualmente para o formato PIR e um caracter de quebra de cadeia ('/') foi adicionado para indicar o fim de uma subunidade e início de outra. Os aminoácidos pertencentes a região da hélice- $\alpha 1$ (Resíduos 18 a 43 para o HIF1a, e 88 a 114 para o ARNT) foram tratados como corpos rígidos durante a etapa de otimização dos modelos, ou seja, tais aminoácidos foram movidos juntos durante o refinamento feito por dinâmica molecular a fim de que suas orientações iniciais não mudem. No final, 10 modelos representando a estrutura quaternária do HIF1 foram gerados e avaliados, posteriormente.

4.5. Validação dos modelos

Os modelos gerados por homologia foram validados primeiramente quanto a sua similaridade estrutural com a estrutura nativa, ou seja, o cristal, e depois quanto sua qualidade físico-química e estereoquímica por meio dos MQAPs (do inglês, *Model Quality Assessment Programs*, MQPA): PROCHECK (LASKOWSKI et al., 1993), Verify 3D (LÜTHY; BOWIE; EISENBERG, 1992), ERRAT (COLOVOS & YEATES, 1993), QMEAN (do inglês, *Qualitative Model Energy ANalysis*, QMEAN) (BENKERT; KÜNZLI; SCHWEDE, 2009) e Prosa-Web (WIEDERSTEIN; SIPLL, 2007).

Em primeiro, os modelos foram avaliados quanto à similaridade estrutural com o molde através do cálculo do *root mean square deviation* (RMSD) obtido pelo alinhamento dos átomos da cadeia principal das estruturas tridimensionais alvos pelo *software* PyMol v.2.4 (SCHRÖDINGER, 2020). O RMSD é um método matemático que calcula o desvio médio

quadrático para todos os átomos de uma estrutura em relação a outra, ou seja, é a medida da distância média entre as posições dos átomos das proteínas sobrepostas. Quanto maior o valor do RMSD, maior é a diferença estrutural entre elas.

Posteriormente, o modelo predito com menor RMSD dentre os demais foi submetido a avaliação da qualidade pelos *softwares* descritos anteriormente. A avaliação da qualidade estereoquímica foi feita por meio do PROCHECK (LASKOWSKI et al., 1993). Os resultados gerados por este *software* são apresentados pelo gráfico de Ramachandran a fim de verificar se os aminoácidos estão em posições energeticamente favoráveis ou desfavoráveis para os ângulos de torção *phi-psi*, sendo que um modelo positivamente avaliado possui >90% dos aminoácidos que compõem a proteína localizados em regiões favoráveis do gráfico (LASKOWSKI et al., 1993). O gráfico de Ramachandran foi construído baseado em estruturas resolvidas experimentalmente depositadas no PDB com resolução de até 2,0 Å e serve como validação de modelos gerados por modelagem comparativa.

O Verify 3D determina a compatibilidade do modelo gerado com a própria sequência de aminoácidos a partir da avaliação dos ambientes químicos dos resíduos. O *software* analisa a qualidade e confiança do enovelamento do modelo gerado através da definição dos ambientes químicos de cada resíduo do modelo e, dessa forma, atribui *scores* referentes a uma matriz previamente estabelecida com base na análise estatística de estruturas de proteínas provenientes do PDB (LÜTHY ; BOWIE; EISENBERG, 1992). Um modelo bem avaliado apresenta compatibilidade maior que 80% entre sua estrutura 3D e a sequência de aminoácidos (1D), um modelo avaliado satisfatoriamente apresenta compatibilidade entre 65% e 80%, e um modelo de baixa qualidade apresenta compatibilidade menor que 65%.

O ERRAT avalia a estrutura proteica baseado na análise de interação entre os átomos não ligados. A ferramenta também determina a probabilidade do valor de erro em relação à disposição do aminoácido por meio de cálculos da comparação de valores estatísticos de estruturas altamente refinadas com o modelo avaliado (COLOVOS & YEATES, 1993). Dessa forma, é possível distinguir regiões da proteína que necessitam de melhoria, sendo que modelos que se assemelham a estruturas tridimensionais de alta resolução apresentam valores de qualidade geral médio acima de 95%, enquanto que estruturas com resolução baixa (entre 2,5 e 3,0 Å) apresentam um entre 91% e 94%.

O *software* ProSA-Web (WIEDERSTEIN; SIPPL, 2007) atribui um *z-score* para a estrutura 3D avaliada, o qual indica a qualidade geral do modelo e quantifica o desvio de energia total da estrutura em relação a uma distribuição de energia derivada de conformações aleatórias

provenientes de estruturas de proteínas depositadas no PDB e resolvidas experimentalmente (WIEDERSTEIN; SIPPL, 2007).

O servidor QMEAN compara o *score* obtido para a estrutura 3D em relação a *scores* de estruturas com proteínas resolvidas experimentalmente que possuem tamanho similar a proteína alvo. O QMEAN é uma função de pontuação capaz de derivar estimativas de qualidade absoluta globais (para toda a estrutura) e locais (por resíduo) com base em um único modelo. A função de pontuação se baseia na análise estatística de características físicas e químicas de um conjunto de proteínas não redundantes de alta resolução (1,8 Å), tais como os ângulos de torção, distância entre pares de átomos, acessibilidade ao solvente dentre outros. O *score* bruto (S) resultante está em um intervalo [0,1], sendo que a estrutura apresenta boa qualidade quando próximo de 1 (BENKERT; KÜNZLI; SCHWEDE, 2009). Para permitir a comparação do modelo com estruturas tridimensionais de alta resolução com tamanho similar, um *z-score* é calculado a partir da normalização do *score* bruto do modelo (BENKERT; BIASINI; SCHWEDE, 2011). Um modelo apresentando *z-score* próximo a zero se assemelha as características físicas e químicas de uma proteína de alta resolução de mesmo tamanho e possui um estado energético favorável, enquanto que um *z-score* de -4,0 ou menor apresenta baixa qualidade.

4.6. Otimização dos *loops*

Após a avaliação de qualidade, o modelo predito por modelagem molecular por homologia foi otimizado quanto a sua estrutura por meio do refinamento dos *loops* pelo MODELLER v.9.24 (FERNANDEZ-FUENTES; FISER, 2010). Este método otimiza as posições de todos os átomos de hidrogênio do *loop* em um ambiente fixo em relação a uma função de pseudo-energia. Tal função contém termos de um campo de força da mecânica molecular, bem como restrições baseadas em distribuições estatísticas derivadas de estruturas proteicas de resolução ótima, ou seja, comprimento da ligação, ângulo da ligação, alguns ângulos diédricos e ângulos diédricos impróprios são restringidos pelos termos correspondentes na função potencial do CHARMM-22 (MACKERREL et al, 1998), enquanto que os ângulos diédricos das cadeias principal e lateral, assim como pares de átomos não ligados, são restringidos por potenciais estatísticas extraídos de muitas estruturas proteicas conhecidas. Por fim, a função de energia é otimizada com o método de gradientes conjugados combinados com dinâmica molecular (FERNANDEZ-FUENTES; FISER, 2010; FISER; SALI, 2003).

4.7. Análise do potencial eletrostático e perfil de hidrofobicidade da superfície

Para predição da distribuição do potencial eletrostático da superfície da proteína foi utilizado o software PyMol v.2.4 (SCHRÖDINGER, 2020). O software calcula o potencial eletrostático da superfície das proteínas com base na equação de Poisson-Boltzmann (FOGOLARI; BRIGO; MOLINARI, 2002), representando um gradiente de potencial negativo para positivo por meio do plugin APBS *Eletrostatics* no PyMol v.2.4 (SCHRÖDINGER, 2020). Os valores referentes ao potencial eletrostático são representados por cores, sendo que as regiões em azul são carregadas positivamente, as regiões em vermelho são negativamente carregadas e regiões em branco representam carga neutra.

Para a predição do perfil de hidrofobicidade foi utilizado valores atribuídos individualmente aos aminoácidos de acordo com a escala de Eisenberg (EISENBERG et al., 1984). Os valores referentes ao potencial de hidrofobicidade são representados em cores, sendo que quanto maior a intensidade da cor vermelha na região, mais hidrofóbica serão os resíduos ali presentes, enquanto que, quanto mais próximo de branco, mais hidrofílico.

4.8. Modelagem molecular do HRE localizado na região promotora do gene *HLA-G* em função do polimorfismo -964G>A

O HRE presente na região promotora do gene *HLA-G* foi modelado pelo *software* x3DNA v.2.4 (LU; OLSON, 2008), cobrindo a sequência -980 a -947 *upstream* da região promotora do gene *HLA-G* (Posição: hg19_Ch6: 29794642-29794674), a qual contém o HRE entre as posições -966 a -962 e o polimorfismo bialélico -964 G>A (rs1632947) em seu interior (5'TAAAAACAGGCAGTGCRTGAGCACTAGTGAGGG'3, onde o "R" é referente ao sítio polimórfico -964G>A) (CASTELLI et al., 2014a, CASTELLI et al., 2014b). Tal sequência foi utilizada pelo x3DNA para construção de duas moléculas de DNA tipo B (B-DNA) dupla-fita com 33 pares de bases (do inglês, *base-pair*, pb) de acordo com o alelo presente no sítio polimórfico -964G>A, sendo a fita complementar construída a partir da sequência de entrada. O algoritmo se baseia em dados provenientes de estruturas resolvidas por difração de raios X de fibras de DNA em condições de 92% de umidade para a modelagem do B-DNA (CHANDRASEKARAN; ARNOTT, 1996).

4.9. Docking proteína-DNA

O *docking* dos complexos HIF1-HRE-964G>A foi feito pelo servidor HADDOCK v.2.4 (do inglês, *High Ambiguity Driven DOCKing*, HADDOCK) (VAN DIJK et al., 2006b; VAN ZUNDERT et al., 2016). O HADDOCK faz uso de dados da interação bioquímica e biofísica entre as moléculas do complexo molecular, os quais são introduzidos como restrições de interação ambíguas (do inglês, *Ambiguous Interaction Restrains*, AIRs) para orientar o processo de *docking* por meio da seleção de resíduos ativos e passivos. Dessa forma, os resíduos ativos e passivos para o direcionamento do processo de *docking* foram definidos: os resíduos conservados entre as proteínas da família bHLH-PAS de fatores de transcrição foram selecionados como resíduos ativos para o HIF1 (HIF1a-ARNT) S22, R23, A25, A26, R29 e R30 para a subunidade α (HIF1a), e H94, E98, R101 e R102 para a subunidade β (ARNT), enquanto que os resíduos ativos para as moléculas de DNA (HRE-964G/A), foram as bases da sequência do HRE 5'GCRTG'3 compreendendo os nucleotídeos nas posições -966 a -962 da região promotora do gene *HLA-G* considerando os alelos do sítio polimórfico -964G>A (CASTELLI et al., 2014b); os resíduos passivos foram automaticamente selecionados em torno dos resíduos ativos pelo servidor.

O algoritmo implementado para *docking* molecular no HADDOCK segue três etapas: (i) minimização de energia e *docking* do corpo rígido (it0), (ii) refinamento semi-flexível dos ângulos de torção (it1) e (iii) refinamento em solvente explícito (água) (W). Por fim, os parâmetros de entrada para o servidor HADDOCK permaneceram no modo padrão, com exceção dos seguintes parâmetros: (i) a constante dielétrica (*epsilon*) que foi definido para 78 para it0 e it1 devido à alta flexibilidade da fita dupla de DNA; (ii) a temperatura inicial para a terceira etapa de resfriamento do TAD (em inglês, *temperature-accelerated dynamic*, TAD) com interface totalmente flexível definido para 300, o fator para o intervalo de tempo no TAD definido para 4, e o número de passos para a fase de 300K (uma das etapas de aquecimento) definidos para 750; (iii) os parâmetros de amostragem foram alterados para: 10.000 /400/400 para os estágios it0/it1/W, respectivamente (VAN DIJK; BONVIN, 2010).

A redução da temperatura tanto na etapa de resfriamento quanto na etapa de aquecimento, e a diminuição do número e do tamanho dos intervalos de tempo para o estágio it1 ajudam na manutenção da estrutura helicoidal do DNA. Visto que temperaturas muito altas ou baixas demais podem desnaturar a dupla fita de DNA, os valores destes parâmetros foram reduzidos pela metade para complexos proteína-DNA baseado nos valores estabelecidos no

protocolo para *docking* proteína-proteína (HONORATO; ROEL-TOURIS; BONVIN, 2019; VAN DIJK et al., 2006b). No estágio final da última etapa, 400 modelos refinados em água foram obtidos no processo de *docking*, os quais foram agrupados usando um *cut-off* de 0,75 de frações de contatos comuns (FCC) (RODRIGUES et al., 2012) e um tamanho de *cluster* mínimo de 4 estruturas como critério.

Como arquivo de saída, o servidor disponibiliza uma página indicando quantas estruturas de modelos refinados em solvente explícito podem ser agrupados e os grupos (*clusters*) são ordenados de acordo com seu HADDOCK *score* (HS). A função de pontuação do HADDOCK, o HS, é a soma ponderada de vários termos de energias intermoleculares de Van der Waals (EVdW), eletrostática (Eelec), dessolvatação (Edesol), AIR (Eair) e, também, da área de superfície enterrada (BSA do inglês, *Buried surface area*, BSA). O HS se difere para cada etapa do *docking* com a configuração padrão da função de pontuação do HADDOCK para os seguintes pesos:

$$HS - it0 = 0,01Evdw + 1,0Eelec + 1,0Edesol + 0,01Eair - 0,01BSA \quad (2)$$

$$HS - it1 = 1,0Evdw + 1,0Eelec + 1,0Edesol + 0,1Eair - 0,01BSA \quad (3)$$

$$HS - W = 1,0Evdw + 0,2Eelec + 1,0Edesol + 0,1Eair \quad (4)$$

onde, a equação 2 se refere ao HS da etapa de minimização de energia e *docking* de corpo rígido (*HS-it0*); a equação 3 se refere ao HS do refinamento semi-flexível dos ângulos de torção (*HS-it1*); e a equação 5 se refere ao HS ao refinamento em solvente explícito (água) (*HS-W*).

A energia de dessolvatação é calculada usando os parâmetros de solvatação atômica empírica descrito por Fernandez-Recio (FERNÁNDEZ-RECIO; TOTROV; ABAGYAN, 2004), enquanto que as demais funções de energia de interação intermolecular são advindas do campo de força OPLS-X, o qual é calculado para todos os resíduos que fazem pelo menos um contato de até 5,0 Å dentro do conjunto de resíduos analisado. Já a área de superfície enterrada é calculada considerando a diferença entre a soma da área de superfície acessível ao solvente para cada molécula separadamente e a área acessível ao solvente do complexo e, assim, usada para quantificar a superfície da proteína que não está exposta à água. Valores mais altos de BSA permitem uma grande proximidade e interação entre as moléculas parceiras, assim, quanto maior o valor de BSA, maior a interface de interação entre as moléculas (DE VRIES; VAN DIJK; BONVIN, 2010).

Para cada *cluster* gerado pelo HADDOCK, há um resumo com estatísticas detalhadas para o HS, para as funções de energias padrão (van der Waals, etc) e para o *z-score*, os quais representam os valores médios calculados sobre as 4 estruturas com melhor *score* de cada grupo. O *z-score* representa quantos desvios padrão o HS de um determinado *cluster* é separado da média de todos os *clusters*. A numeração do *cluster* reflete a quantidade de estruturas agrupadas, com o *cluster* 1 sendo o *cluster* com maior número de poses que os demais (DEVRIES; VAN DIJK; BONVIN, 2010b; DOMINGUEZ; BOELEN; BONVIN, 2003b). As estruturas finais representativas dos *dockings* proteína-DNA realizados para o complexo HIF1-HRE em função do SNP -964G>A do *HLA-G* foram nomeadas de acordo com o alelo presente no sítio -964 do elemento regulatório, ou seja, complexos HIF1-HRE-964G e HIF1-HRE-964A.

4.10. Análise de interação intermolecular dos complexos HIF1-*HLA-G*-HRE-964G/A

A análise da interface molecular foi feita para verificar o padrão de ligação entre os átomos do complexo proteína-DNA por meio do servidor DNAProDB⁶ (SAGENDORF et al., 2019). O servidor extrai informações do complexo a partir do arquivo de coordenadas atômicas submetido como entrada, o qual é utilizado por programas e bibliotecas incorporados no servidor para o cálculo das interações intermoleculares. Em primeiro, o servidor adiciona átomos de hidrogênio em todas as estruturas do complexo-molecular com o módulo *Reduce* do servidor *MolProbity* (CHEN et al., 2010).

O cálculo das pontes de hidrogênio foi feito pelo HBPLUS (MCDONALD; THORNTON, 1994). Para identificar ligações de hidrogênio, o programa encontra todos os pares de átomos doadores proximais (D) e aceptores (A) que satisfazem os critérios geométricos especificados para a formação de ligações. As posições teóricas dos átomos de hidrogênio (H) são então calculadas para os átomos doadores que se enquadram nos critérios e as ligações são calculadas entre o hidrogênio e os átomos aceptores. Os critérios usados são: a distância H-A <2,7 Å, a distância D-A <3,35 Å, o ângulo D-H-A >90° e o ângulo H-A-AA >90°, onde AA é o átomo ligado ao aceptor. Visto que o HBPLUS utiliza uma lista de contatos entre átomos pré-definidos, o programa pode ser incapaz de calcular corretamente todas as ligações de hidrogênio

⁶ <https://dnaprodb.usc.edu/>

que o complexo faz, assim, o programa SNAP (do inglês, *Structures of Nucleic Acid-Protein*, SNAP), um componente do programa 3DNA, também é usado para calcular as ligações de hidrogênio existentes no complexo molecular (LU; OLSON, 2008).

Para o cálculo das forças de Van der Waals (VdWs), todos os pares de átomos pesados a uma distância máxima de 3,9Å que não formam pontes de hidrogênio foram considerados interagindo pelo DNAproDB por meio do módulo KDTTree do Biophyton (COCK et al., 2009). As pontes de hidrogênio formadas entre o elemento regulatório e o domínio de ligação ao DNA do HIF1 (bHLH) identificadas pelo DNAproDB foram visualizadas a partir do *software* PyMol v.2.4 (SCHRÖDINGER, 2020).

4.11. Simulação de dinâmica molecular do fator de transcrição HIF1 livre e complexado ao HRE enfatizando o polimorfismo -964G>A do *HLA-G*

O HIF1 na sua forma livre (não complexada) e os dois complexos proteína-DNA, o HIF1-HRE-964G e HIF1-HRE-964A, foram submetidos a simulações de dinâmica molecular (DM) no pacote GROMACS v.2019. As simulações para a proteína livre foram realizadas usando o campo de força AMBER-99SB (HORNACK et al., 2006), enquanto que, para os complexos proteína-DNA, foi usado o mesmo campo de força com os parâmetros referentes ao DNA modificados de acordo com o campo de força AMBER-Parmbsc1 (IVANI et al., 2016). Os sistemas foram inicialmente inseridos em um sistema construído em uma caixa cúbica (GALINDO-MURILLO; BERGONZO; CHEATHAM, 2014), os quais foram posteriormente solvatados através da adição de moléculas de solvente explícito do tipo SPC/E com no mínimo de 14Å de distância entre a borda e qualquer átomo da proteína ou complexos moleculares (BERENDSEN; GRIGERA; STRAATSMAN, 1987). Depois, a carga geral dos sistemas foi neutralizada com a adição de íons Cl⁻ e K⁺ a fim de atingir a concentração total de KCl para 150 mM.

Antes da etapa de produção, o sistema foi submetido a uma minimização de energia por 0,1 nanossegundos (ns) para eliminar possíveis choques estéricos entre os átomos. Com a etapa de minimização é possível obter uma estrutura inicial razoável, em termos de geometria e orientação do solvente. Porém, para iniciar a dinâmica real, precisamos equilibrar o sistema, ou seja, o solvente e os íons ao redor da proteína. Assim, após a minimização, duas etapas de equilíbrio foram feitas: na primeira, o número de partículas, volume e temperatura foram propriedades mantidas constantes no sistema durante a integração das equações newtonianas

(chamado *Ensemble* NVT) e, na segunda, o número de partículas, pressão e temperatura foram mantidas constantes (chamado *Ensemble* NPT) (NAMBA; SILVA; SILVA, 2008).

Posteriormente, as simulações de DM foram realizadas com condições periódicas de contorno à pressão e temperatura constantes (1 atm, 300 K), usando um termostato modificado de Berendsen (BERENDSEN et al., 1984) com constante de acoplamento de 0,1 os, e um barostato de ParrinelloRahman (PARRINELLO; RAHMAN, 1981) com constante de acoplamento de 0,1 ps, e 2 fentossegundo (fs) de *time step*. Todas as ligações envolvendo átomos de hidrogênio foram restritas com o algoritmo LINCS (HESS et al., 1997) e a lista de pares de átomos de contatos não ligados foi atualizada a cada 20 fs por meio do algoritmo de Verlet (PÁLL; HESS, 2013), enquanto que as forças eletrostáticas foram avaliadas usando o algoritmo PME (do inglês, *Particle Mesh Ewald*, PME) (DARDEN; YORK; PEDERSEN, 1993). O movimento do centro de massa foi removido a cada 0,2 ps para eliminar a energia cinética translacional (HARVEY; TAN; CHEATHAM, 1998). Por fim, os sistemas provenientes do fator de transcrição livre e dos complexos proteína-DNA, os HIF1-HRE-964G/A, foram submetidos a uma DM de 200ns para gerar o arquivo de trajetória.

4.12. Análise das trajetórias resultantes da DM

A partir das coordenadas atômicas presentes na trajetória dos sistemas simulados, o HIF1 livre e os complexos HIF1-HRE-964G/A, foi calculado o RMSD, a flutuação quadrática média (do inglês, *Root Mean Square Fluctuation*, RMSF) e o raio de giro (do inglês, *Ratio of Gyration*, RoG) (NAMBA; SILVA; SILVA, 2008) para o fator de transcrição na sua forma livre e complexada ao HRE, e para o DNA ligado a proteína.

O RMSD na DM fornece a medida do desvio da estrutura de uma dada molécula ao longo do tempo em relação a uma estrutura de referência. No cálculo, a cada intervalo de tempo, é realizado sobreposições da estrutura atual com a estrutura de referência para obter o menor valor de RMSD entre as estruturas. O RMSD pode ser calculado segundo a equação (5):

$$RMSD_{(t)} = \sqrt{\frac{1}{M} \sum_{i=1}^N m_i |r_i(t) - r_i^{ref}|^2} \quad (5)$$

onde $M = \sum_{i=1}^N m_i$ é a massa total em uma proteína com N átomos e $r_i(t)$ é a posição (coordenadas) do átomo i no tempo t . A estrutura inicial ($t=0$) é usada como referência para o

cálculo. O RMSF é calculado para os átomos do C α de cada resíduo e é simplesmente a raiz quadrada da variação da flutuação em torno da posição média equação (6).

$$RMSF_{(i)} = \left[\frac{1}{T} \sum_{t_j=1}^T |r_i(t_i) - r_i^{ref}|^2 \right]^{1/2} \quad (6)$$

onde T é o tempo durante o qual se deseja calcular a média e r_i^{ref} é a posição de referência da partícula i . Esta posição de referência será a posição média da mesma partícula i . A diferença entre RMSD e RMSF é que o último é calculado pela média ao longo do tempo, fornecendo um valor para cada partícula i , enquanto que para o RMSD a média é feita sobre as partículas, fornecendo valores específicos de tempo. Para explicar melhor, pode-se usar o exemplo de uma cadeia lateral de aminoácido numa simulação com a proteína. Caso as estruturas ao longo da simulação sejam diferentes da estrutura de partida, tem-se um alto valor de RMSD, porém se estas estruturas foram semelhantes, não haverá flutuação estrutural, logo, o RMSF é quase nulo (FRENKEL; SMIT, 2002).

O cálculo do RoG evidencia o processo de compactação ou expansão, ou seja, dobramento e desdobramento da cadeia polipeptídica ou DNA, indicando mudança conformacional na molécula em questão. O cálculo é dado pela raiz quadrática média da distância entre o centro gravitacional e as suas extremidades, conforme a equação (7):

$$R_g = \sqrt{\frac{\sum_i |r_i|^2 m_i}{\sum_i m_i}} \quad (7)$$

onde m_i é a massa do átomo i e r_i é a posição do átomo i em relação ao centro de massa da molécula.

4.11. Análise dos parâmetros conformacionais do DNA a partir da trajetória.

Para analisar as diferenças nos parâmetros conformacionais da estrutura 3D da dupla hélice de DNA durante a simulação de dinâmica molecular dos complexos HIF1-HRE em função do polimorfismo -964G>A do gene *HLA-G*, as coordenadas atômicas dos complexos proteína-DNA foram extraídas da trajetória a cada 10 ns, iniciando no tempo $t=0$ ns e, assim, um total de 21 estruturas foram analisadas por meio do programa Curves v.2.6 (LAVERY et al.,

2009), o qual ajusta o eixo helicoidal da dupla hélice de DNA para um sistema de linha curvilínea na direção da propagação da hélice, considerando a posição e orientação das bases nitrogenadas dentro de um par de bases, para calcular os parâmetros conformacionais. Assim, o movimento entre dois pares de bases é descrito por três distâncias de translação (*shift*, *slide* e *rise*) e três ângulos de rotação (*roll*, *tilt* e *twist*) (DICKERSON, 1989; OLSON et al., 2001):

Shift: movimento de translação em torno do eixo X;

Slide: movimento de translação em torno do eixo Y;

Rise: movimento de translação em torno do eixo Z;

Roll: Movimento de rotação em torno do eixo Y;

Tilt: Movimento de rotação em torno do eixo X;

Twist: movimento de rotação em torno do eixo Z.

Este conjunto de parâmetros foi calculado para cada coordenada atômica dos complexos HIF1-HRE-964G/A e os resultados foram apresentados como a média e o desvio padrão a partir das 21 estruturas extraídas da trajetória de cada complexo molecular.

5 Resultados e Discussão

5.1. Modelagem da estrutura quaternária do fator de transcrição HIF1

Para realização do *docking* molecular proteína-DNA, é necessário que a estrutura tridimensional das moléculas parceiras tenha sido resolvida por cristalografia de raio X, ressonância magnética ou métodos computacionais. Não há estrutura tridimensional resolvida para o domínio de ligação ao DNA do HIF1 humano, apenas para os domínios PAS-A/B-PAC da subunidade α (HIF1a humano), o qual foi resolvido por cristalografia de raio-X com uma resolução de 1,52Å e está depositada no PDB identificada pelo PDB ID: 4H6J (CARDOSO et al., 2012). Assim, será necessário realizar a modelagem molecular *in silico* do complexo HIF1 humano, sendo as regiões alvos para modelagem: o domínio de ligação ao DNA (bHLH) e os domínios PAS-A/B-PAC necessários para dimerização das subunidades α e β para formação do fator de transcrição em questão.

Para resolver a estrutura tridimensional do heterodímero HIF1 foi utilizado a técnica de modelagem molecular por homologia com uso do *software* MODELLER v9.24 (WEBB; SALI, 2016). Visto que tanto proteínas quanto moléculas de DNA diferem da sua geometria na forma livre e complexada (SVOZIL et al., 2008), o molde escolhido para a modelagem deve ser um complexo proteína-DNA em que a proteína apresente ao menos 30% de identidade ao HIF1. O molde escolhido para a modelagem molecular foi o cristal do fator de transcrição HIF2 de camundongo, identificado no PDB como 4ZPK, com 3,60Å de resolução, a qual é uma proteína pertencente à mesma família de proteínas que o HIF1 e está ligado ao HRE (DENGLER; GALBRAITH; ESPINOSA, 2014), também chamado de complexo HIF2a-ARNT-HRE. Neste cristal, a cadeia A é referente a subunidade β , o ARNT, e a cadeia B é referente a subunidade α , o HIF2a. Por sua vez, estes dois heterodímeros apresentaram similaridade acima de 80%, eles têm arquiteturas intimamente relacionadas, pois mutações pontuais localizadas nos sítios de ligação que desestabilizam o complexo HIF2 também desestabilizaram o HIF1 de camundongo (WU et al., 2015), o que torna o HIF2 uma boa escolha como molde para a modelagem molecular. Além disso, o alinhamento global entre as subunidades α e β dos HIF1 (humano) e HIF2 (camundongo) revelou que as subunidades β compartilham 97,1% de identidade e 99% de similaridade, e as subunidades α , 64,8% de identidade e 81,4% de similaridade entre os fatores de transcrição em questão (Figura 16A e B, respectivamente).

Figura 16. Alinhamento global feito por meio do *software* EMBOSS *Needle* entre as sequências primárias das (A) subunidades β (ARNT) e (B) α (HIF1/2a) dos fatores de transcrição HIF1 (humano) e HIF2 (camundongo).

Fonte: O autor (2021).

O cristal do complexo proteico HIF2 apresenta longas sequências de aminoácidos ausentes (13 a 31 resíduos) na estrutura terciária das duas subunidades, ARNT e HIF2a. Então, antes de proceder com a modelagem molecular do HIF1, as sequências ausentes do cristal do complexo HIF2 de camundongo serão preenchidas para, assim, ele ser utilizado como molde para a modelagem molecular do complexo HIF1 humano.

Primeiro, a sequência primária das cadeias do cristal 4ZPK foram submetidas ao servidor PSIPRED para predição da estrutura secundária das regiões ausentes na proteína resolvida experimentalmente (Figura 17). Os resultados mostram que as sequências ausentes no cristal, tanto para a subunidade α (HIF2a, cadeia B) quanto para a subunidade β (ARNT, cadeia A) do fator de transcrição HIF2 de camundongo, são na sua maioria regiões de *loop*, com exceção para algumas pequenas regiões ausentes preditas como hélices- α ou folhas- β , porém, no caso das folhas- β , algumas apresentam baixa confiabilidade na predição (Regiões identificadas com um “*” na Figura 17A).

Figura 17. Predição de estrutura secundária do complexo HIF2 de camundongo. (A) Subunidade β do HIF2, o ARNT (PDB ID: 4ZPK, cadeia A); (B) Subunidade α do HIF2, o HIF2a (PDB ID: 4ZPK, cadeia B). Sequências de resíduos destacadas por retângulos coloridos são regiões ausentes no cristal. Folhas- β previstas com baixa confiabilidade identificadas com “*”.

Fonte: Adaptado do servidor PSIPRED.

Depois, o molde HIF2 teve as regiões não resolvidas no cristal preenchidas pela ferramenta *PDB Reader* do servidor CHARMM-GUI (JO et al., 2008). Na Figura 18, são mostradas as regiões ausentes modeladas em cada uma das subunidades formadoras do complexo HIF2, as quais estão destacadas em diferentes cores em concordância com as cores das mesmas regiões destacadas previamente na Figura 17.

Figura 18. Representação da estrutura tridimensional completa do HIF2 de camundongo, referente ao cristal identificado pelo PDB ID: 4ZPK. Na Figura, observamos a representação da estrutura tridimensional incompleta para cada uma das subunidades do HIF2, o ARNT e o HIF2a, com as regiões ausentes na estrutura terciária, e a representação da estrutura tridimensional completa após a modelagem dos resíduos perdidos, com as sequências ausentes já preenchidas e destacadas. A cadeia A referente ao ARNT, está ilustrado em azul e a cadeia B, referente ao HIF2a, em amarelo.

Fonte: O autor (2021).

Uma vez obtido a estrutura quaternária completa do HIF2 de camundongo, foi possível prosseguir com a modelagem molecular do fator de transcrição humano. Para a modelagem estrutural, as estruturas primárias das duas subunidades proteicas do HIF1, o HIF1a e o ARNT humanos, foram recuperadas no banco de dados do UniProt correspondentes aos domínios de bHLH, PASA/B e PAC (Aminoácidos 15 a 349 para o HIF1a e aminoácidos 87 a 470 para o ARNT), as quais foram, posteriormente, alinhadas a sequência primária das cadeias do cristal 4ZPK para modelagem (Ver metodologia). Por meio do MODELLER v.9.24, foram gerados dez modelos diferentes com a região da hélice- $\alpha 1$, a qual é a principal região de ligação ao

DNA, do HIF1a e ARNT tratadas como corpos rígidos para minimizar a geração de conformações muito distintas do domínio de ligação ao DNA do molde. Tais modelos foram posteriormente analisados quanto à similaridade estrutural dos modelos preditos com o cristal 4ZPK através do cálculo do RMSD (do inglês, *root mean square deviation*, RMSD) pelo alinhamento das estruturas tridimensionais no *software* PyMol v.2.4 (SCHRÖDINGER, 2020) considerando as posições dos C α da cadeia polipeptídica.

O modelo que apresentasse dobramento mais parecido com o cristal seria escolhido para as próximas etapas. Em geral, a organização conformacional dos modelos foi semelhante ao cristal do HIF2, assim, aquele que obtivesse menor valor de RMSD seria escolhido para prosseguir com a validação baseado nas análises de qualidade. Dessa forma, o modelo 8 foi o que apresentou o menor valor de RMSD no alinhamento entre o modelo e o molde (RMSD: 0,396 Å) dentre os demais e o resultado deste alinhamento é mostrado na Figura 19.

Figura 19. Alinhamento das estruturas tridimensionais do modelo 8 gerado por homologia do HIF1 e o cristal do HIF2 (PDB ID: 4ZPK). A estrutura em vermelho corresponde ao cristal e em verde-água o modelo 8 (RMSD: 0,396Å). As regiões destacadas são referentes às hélices- α 1 das subunidades HIF1a/HIF2a, α -H1 (a), e ARNT, α -H1 (b).

Fonte: O autor (2021).

Posteriormente, o modelo escolhido foi avaliado quanto à qualidade química, física e estereoquímica pelos MQAPs (do inglês, *Model Quality Assessment Program*, MQPAs): PROCHECK (LASKOWSKI et al., 1993), ERRAT (COLOVOS; YEATES, 1993), Verify 3D (LÜTHY; BOWIE; EISENBERG, 1992), QMEAN (BENKERT; KÜNZLI; SCHWEDE, 2009) e DFIRE (ZHANG, 2004; ZHOU; ZHOU, 2009). Os resultados obtidos por estes *softwares* para o melhor modelo gerado por homologia para o fator de transcrição HIF1 e de seu molde são mostrados na Tabela 2.

Tabela 2. Resultado da avaliação de qualidade do modelo 8 do HIF1 gerado pelo MODELLER e seu molde para os seguintes MQAPs.

MQAPs	4ZPKi	4ZPKc	Modelo 8
PROCHECK – Gráfico de Ramachandran (1)			
Regiões mais favoráveis	83,10	80,10	86,60
Regiões adicionalmente favoráveis	15,70	17,50	11,10
Regiões adicionalmente favoráveis	0,80	1,20	1,40
Regiões desfavoráveis	0,40	1,20	0,90
Verify-3D (2)	41,77	58,90	73,01
QMEAN score bruto (Z-score) (3)	0,52 (-2,44)	0,66 (-5,82)	0,62 (-3,49)
ProSA-Web (4)*	-5,16/-6,37	-7,38/-7,14	-7,19/-7,76
ERRAT (5)*	66,67/58,43	76,42/75,16	67,97/63,27

Valores de referência: (1) >90% dos aminoácidos em regiões favoráveis; (2) Bom: ≥80%, Satisfatório: <80%, Ruim: <65% de compatibilidade 3D-1D >0,2; (3) *score bruto*: valores entre 0 e 1, sendo 1 o ideal; (4) *z-score* dentro do intervalo de distribuição de proteínas obtidas experimentalmente com tamanho similar; (5) Estruturas de alta resolução geralmente produzem valores em torno de 95% ou mais, enquanto estruturas em resoluções mais baixas (2,5-3Å), a qualidade geral média é de cerca de 91%.

Número total de aminoácidos: ZPKi: 610; 4ZPc: 715; Modelo 8: 715; MQAP: *Model Quality Assessment Program*; 4ZPKi: cristal do HIF2 com estrutura incompleta (regiões ausentes); 4ZPKc: cristal do HIF2 com estrutura tridimensional completa.

* Valores atribuídos as subunidades individualmente: ARNT/HIFa.

O gráfico de Ramachandran construído através do PROCHECK indica as regiões mais favoráveis, adicionalmente favoráveis, generosamente favoráveis e desfavoráveis para os ângulos de torção *phi* e *psi*. Tal análise é realizada para todos os aminoácidos, com exceção dos aminoácidos glicina e prolina por não apresentarem torção destes ângulos devido a composição de suas cadeias laterais. Na Figura 20A, está ilustrado a disposição dos ângulos de torção *phi* e *psi* para o modelo 8. Segundo o gráfico de Ramachandran, o modelo 8 apresentou 86,60% de seus aminoácidos em regiões mais favoráveis, e 0,90% de aminoácidos em regiões não permitidas. Apesar de o modelo não apresentar a quantidade ideal de aminoácidos em regiões favoráveis e desfavoráveis (acima de 90% e 0%, respectivamente), este apresentou maior qualidade estereoquímica em comparação com o cristal.

De acordo com os resultados do Verify 3D, o modelo mais similar ao molde apresentou qualidade satisfatória, com compatibilidade de 73,01% entre sua estrutura 3D e a sequência de aminoácidos, enquanto que o cristal foi avaliado com qualidade baixa. Quanto ao ProSA-Web, é esperado que o valor do *z-score* para a estrutura analisada esteja dentro do intervalo de *z-scores* de estruturas resolvidas experimentalmente depositadas no PDB com tamanho similar, atribuindo ao modelo uma boa qualidade por apresentar pouca probabilidade de erro em relação a energia conformacional da estrutura avaliada (WIEDERSTEIN; SIPPL, 2007). Dessa forma, de acordo com a Figura 20 (B), o modelo gerado para o HIF1 obteve boa qualidade para os parâmetros estabelecidos para o *software* ProSA-Web por apresentar um *z-score* dentro do intervalo de distribuição de *z-scores* tipicamente encontrados para proteínas obtidas por métodos experimentais, assim como para o cristal do fator de transcrição HIF2 de camundongo.

O QMEAN atribui um *score* global de qualidade para o modelo avaliado que varia de [0] a [1], onde 1 é considerado ótimo. Tal *score* global, chamado aqui de *score* bruto, também é expresso como *Z-score* para permitir que este valor seja comparado com valores provenientes de estruturas proteicas tridimensionais depositadas no PDB. Os valores de *score* bruto e *Z-score* obtidos para o modelo 8 do HIF1 é de 0,62 e -3,49, respectivamente (Tabela 02). O resultado do *Z-score* para o HIF1 está dentro do limite da faixa de valores de *Z-score* de estruturas obtidas por métodos experimentais e tamanho similar (Figura 20C). Apesar de não ilustrar no gráfico os valores de *Z-score* para o molde, o cristal com a estrutura completa (4ZPKc) apresentou valor inferior a -4,0 (-5,82) indicando que este tem baixa qualidade de acordo com o *software* QMEAN.

Figura 20. Avaliação da estrutura quaternária do HIF1 humano obtida por modelagem molecular por homologia. (A) Gráfico de Ramachandran representativo do modelo melhor avaliado (modelo 8) para o fator de transcrição HIF1. Regiões mais favoráveis para os ângulos torcionais estão ilustradas em vermelho; regiões adicionalmente favoráveis em amarelo; regiões generosamente favoráveis em bege; e regiões desfavoráveis/não permitidas em branco. Os quadrados pretos representam o posicionamento dos aminoácidos no gráfico, os triângulos pretos representam resíduos de glicinas e prolinas. Os aminoácidos localizados em regiões generosamente favoráveis ou desfavoráveis estão identificados em vermelho. (B) Resultados do *z-score* obtido pelo ProSA-Web. O *z-score* da estrutura analisada deve estar dentro do intervalo de distribuição de *z-scores* encontrados para proteínas resolvidas experimentalmente com mesmo número de aminoácidos. O ponto em azul indica a subunidade ARNT e o ponto rosa indica a subunidade HIF1a do HIF1. (C) Resultados do *z-score* do QMEAN. Valores de *z-score* próximos a 0 indicam que o modelo avaliado tem qualidade compatível com estruturas de alta resolução de tamanho similar e valores inferiores a -4,0 indicam que o modelo avaliado tem baixa qualidade. A estrela vermelha no gráfico indica a posição da estrutura avaliada na mesma faixa de valores *z-scores* que estruturas preditas experimentalmente com mesmo tamanho.

Fonte: O autor (2021).

Por meio dos resultados obtidos pelo ERRAT, é possível identificar regiões na proteína com valores de erro acima de 99% indicando a necessidade de melhoria para estes resíduos, como observado na Figura 21. Tal melhoria pode ser alcançada através do refinamento de *loops*,

a fim de corrigir possíveis erros locais nos modelos gerados por modelagem molecular por homologia, como comprimento de ligações incomuns e contatos irregulares, principalmente quando o cristal usado como molde não apresenta tão boa resolução (LANCE; DEANE; WOOD, 2010). A seleção das regiões de *loops* para o refinamento foi feita de acordo com a avaliação do ERRAT e os resíduos selecionados para otimização destacados por um quadrado vermelho na Figura 21.

Figura 21. Gráfico gerado pelo ERRAT com os valores de erro por resíduo com os aminoácidos de regiões de *loops* selecionados para otimização (em vermelho) para a subunidade α , o (A) HIF1a, e para a subunidade β , o (B) ARNT do complexo HIF1. O eixo Y representa o valor do erro e o eixo X representa os resíduos de aminoácidos do modelo proteico. Um valor de erro superior a 99% indica uma região que necessita de melhoria (destacado em preto).

Fonte: O autor (2021).

A otimização dos *loops* foi feita por um algoritmo implementado no *software* MODELLER v.9.24 (FERNANDEZ-FUENTES; FISER, 2010; FISER; SALI, 2003). A Figura 22 mostra os valores de erro para cada resíduo no modelo da estrutura quaternária do HIF1 após o refinamento de acordo com o ERRAT. O modelo otimizado para os *loops* do HIF1 apresentou um valor de qualidade global de 70,86% e 76,16% para o HIF1a e o ARNT, respectivamente, o que é maior que o valores referentes ao modelo inicial de 63,27% para o HIF1a e 67,97% para o ARNT, demonstrando que as regiões de *loops* que apresentavam erro foram, em sua maioria, otimizadas.

Figura 22. Gráfico gerado pelo ERRAT com os valores de erro por resíduo do modelo otimizado da estrutura quaternária do HIF1. (A) Subunidade α , o HIF1 α . (B) Subunidade β , o ARNT. O eixo Y representa o valor do erro e o eixo X representa os resíduos de aminoácidos do modelo proteico.

Fonte: O autor (2021).

O modelo otimizado foi reavaliado também pelos MQAPs anteriores a fim de realizar uma análise comparativa entre os resultados de avaliação do cristal e o modelo antes e após o refinamento dos *loops*. Os resultados obtidos por estes *softwares* para o modelo refinado do fator de transcrição HIF1 são mostrados na Tabela 3 em comparação com o modelo inicial e o cristal.

Tabela 3. Resultado da avaliação de qualidade do modelo inicial HIF1 em comparação com o modelo refinado final e seu molde.

MQAPs	4ZPKi	4ZPKc	Modelo 8	Modelo 8 Ref
PROCHECK - Ramachandran Plot (1)				
Regiões mais favoráveis	83,10	80,10	86,60	87,20
Regiões adicionalmente favoráveis	15,70	17,50	11,10	11,10
Regiões adicionalmente favoráveis	0,80	1,20	1,40	0,90
Regiões desfavoráveis	0,40	1,20	0,90	0,80
Verify-3D (2)	41,77	58,90	73,01	68,87
QMEAN score bruto (Z-score) (3)	0,52 (-2,44)	0,66 (-5,82)	0,62 (-3,49)	0,65 (-2,77)
ProSA-Web (4)*	-5,16/-6,37	-7,38/-7,14	-7,19/-7,76	-7,75/-8,00
ERRAT (5)*	66,67/58,43	76,42/75,16	67,97/63,27	76,16/70,86

Valores de referência: (1) >90% dos aminoácidos em regiões favoráveis; (2) Bom: $\geq 80\%$, Satisfatório: $< 80\%$, Ruim: $< 65\%$ de compatibilidade 3D-1D $> 0,2$; (3) *score bruto*: valores entre 0 e 1, sendo 1 o ideal; (4) *z-score* dentro do intervalo de distribuição de proteínas obtidas experimentalmente com tamanho similar; (5) Estruturas de alta resolução geralmente produzem valores em torno de 95% ou mais, enquanto estruturas em resoluções mais baixas (2,5-3Å), a qualidade geral média é de cerca de 91%; Número total de aminoácidos: ZPKi: 610; 4ZPKc: 715; Modelo 8: 715; MQAP: *Model Quality Assessment Program*; 4ZPKi: cristal do HIF2 com estrutura incompleta (regiões ausentes); 4ZPKc: cristal do HIF2 com estrutura tridimensional completa.

* Valores atribuídos as subunidades individualmente: ARNT/HIF α

Como observado na Tabela 3, a etapa de refinamento dos *loops* resultou em melhoria da qualidade da proteína quando comparado a estrutura inicial para alguns *softwares*, porém, não atingiu a qualidade físico-química e estereoquímica observada em proteínas de alta resolução ou de ótima qualidade. Contudo, nenhum dos resíduos envolvidos na interação intermolecular com o elemento regulatório presente na região promotora do gene *HLA-G* apresentaram distribuição dos ângulos torcionais em regiões generosamente favoráveis e desfavoráveis. Apesar do valor atribuído pelo Verify 3D para o modelo otimizado ser inferior em comparação com modelo anterior ao refinamento, a qualidade ainda é satisfatória de acordo com os parâmetros estabelecidos pelo *software*. E os resultados obtidos de *Z-score* do ProSA-Web e QMEAN para o modelo otimizado continuam dentro do limite do intervalo de valores de *Z-score* de estruturas obtidas por métodos experimentais e com mesmo número de aminoácidos. Assim, visto que (i) o presente trabalho tem como objetivo analisar a interação estrutural do DNA com o dímero em questão, e (ii) os resíduos ligantes ao HRE (DNA) não apresentaram má qualidade quanto à disposição dos ângulos de torção *phi-psi* que possa afetar a ligação do complexo proteico com o DNA, o modelo otimizado para os *loops* representará a estrutura tridimensional final do HIF1 para a realização do *docking*. Por fim, na Figura 23, o modelo refinado para os *loops* foi escolhido como modelo final da estrutura quaternária do fator de transcrição HIF1.

5.2. Análise do potencial eletrostático e perfil de hidrofobicidade

A composição das superfícies moleculares dos fatores de transcrição e de outras proteínas é determinante para a afinidade e especificidade das interações entre as moléculas, principalmente para o reconhecimento de seus alvos genômicos pelas cadeias laterais dos aminoácidos (CHIU et al., 2017). Baseado nisso, foi feita a análise da superfície das subunidades do fator de transcrição HIF1 quanto à distribuição do potencial eletrostático e perfil de hidrofobicidade a fim de identificar a carga e a hidrofobicidade dos aminoácidos que interagem com a dupla-hélice de DNA.

Na literatura, muitos dos domínios de ligação ao DNA foram estudados por métodos experimentais como cristalografia de raio X de complexos proteína-DNA. Isso permite a identificação *in vitro* de aminoácidos no domínio de ligação ao DNA envolvidos diretamente na ligação intermolecular com as bases nitrogenadas ou com o esqueleto açúcar-fosfato da dupla hélice.

Figura 23. Representação tridimensional do modelo final para a estrutura quaternária do complexo HIF1 (HIF1a-ARNT) humano. Em azul está destacado a cadeia A em referência a subunidade β , o ARNT, e em rosa está destacado a cadeia B em referência a subunidade α , o HIF1a. Nos quadros estão a estrutura de cada cadeia separada com os domínios proteicos ilustrados.

Fonte: O autor (2021).

O HIF1 se liga ao HRE principalmente por meio das hélices- $\alpha 1$ N-terminais presentes nas suas subunidades α (HIF1a) e β (ARNT). Logo, a identificação dos resíduos de tais hélices- $\alpha 1$ importantes para a interação com o DNA foi feita com base no estudo de estruturas cristalográficas de proteínas pertencentes à mesma família bHLH-PAS de fatores de transcrição que o HIF1. Assim como o HIF1, tais proteínas são formadas por duas subunidades as quais interagem entre si formando heterodímeros e, posteriormente, se ligam ao elemento regulatório presente em diversos genes alvos (GRAHAM; PRESNELL, 2017).

A estrutura cristalográfica de heterodímeros que compartilham a subunidade ARNT apresentam arquitetura globalmente similar entre eles quando ligado ao elemento regulatório para os cristais provenientes de camundongo dos complexos ARNT: AHR-ARNT-DNA que codifica o DRE (PDB ID: 5V0L), (SCHULTE et al., 2017), HIF1a/HIF2a-ARNT ligado ao

HRE (PDB ID: 4ZPR para o HIF1a e 4ZPK para o HIF2a (WU et al., 2015) e NPAS3-ARNT ligado ao HRE (PDB ID: 5SY7) (WU et al., 2016). Na Figura 24, foi feita uma comparação entre as estruturas primárias do bHLH das proteínas que se ligam ao ARNT formando dímeros citados anteriormente com destaque aos resíduos conservados entre eles que se ligam ao DNA. Assim, os resíduos de ligação ao DNA para o fator de transcrição HIF1 são: S22, R23, A25, A26, R29 e R30 para a subunidade α (HIF1a); e H94, E98, R101 e R102 para a subunidade β (ARNT).

Figura 24. Comparação do domínio de ligação ao DNA, bHLH, entre as proteínas que se ligam ao ARNT. Alinhamento múltiplo entre as subunidades α que interagem com o ARNT (subunidade β) para formação de heterodímeros que se ligam ao DNA realizado pelo servidor *Clustal Omega* v.1.2 e editado manualmente. Na Figura, está ilustrado a estrutura secundária do domínio bHLH para as subunidades α e o ARNT (subunidade β) com os resíduos ligantes ao DNA destacados em vermelho.

Fonte: O autor (2021).

Por fim, na Figura 25 está ilustrado a representação tridimensional do fator de transcrição HIF1 humano modelado anteriormente com os resíduos que interagem com o DNA em destaque, baseado na análise dos resíduos ligantes ao DNA conservados entre as proteínas da família bHLH-PAS. A análise do potencial eletrostático e perfil de hidrofobicidade da superfície foi feito considerando tais propriedades nestes resíduos por meio do software PyMol v.2.4 (SCHRÖDINGER, 2020).

Os valores referentes ao potencial eletrostático são representados por cores, sendo que as regiões em azul são carregadas positivamente, as regiões em vermelho são negativamente carregadas e as regiões em branco representam carga neutra (FOGOLARI; BRIGO; MOLINARI, 2002). Para a predição do perfil de hidrofobicidade foi utilizado valores atribuídos individualmente aos aminoácidos de acordo com a escala de Eisenberg (EISENBERG et al., 1984), os quais são representados em cores, sendo que quanto maior a intensidade da cor vermelha na região, mais hidrofóbica serão os resíduos ali presentes, enquanto que, quanto mais

próximo de branco, mais hidrofílico. Na Figura 26 é mostrado a análise do potencial eletrostático e do perfil de hidrofobicidade da superfície para o HIF1.

Figura 25. Representação tridimensional do fator de transcrição HIF1 humano formado pelas proteínas HIF1a (rosa) e ARNT (azul). As hélices- $\alpha 1$ do ARNT, α -H1(b), e do HIF1a, α -H1(a), em destaque. Os resíduos que interagem com o DNA estão destacados em vermelho em diferentes ângulos.

Fonte: O autor (2021).

De acordo com a Figura 26, as regiões proteicas importantes para a interação com o DNA, em especial os aminoácidos ligantes ao HRE conservados entre os domínios bHLH que se ligam ao ARNT, apresentaram potencial eletrostático positivo ou levemente positivo. Tal resultado é esperado, visto que o grupo fosfato presente nos nucleotídeos confere à superfície da cadeia de DNA densidade global de carga negativa, geralmente.

Em relação ao perfil de hidrofobicidade dos aminoácidos ligantes ao DNA conservados do HIF1, 80% são hidrofílicos (HIS94, GLU98, ARG101 e ARG102 para o ARNT; e SER22, ARG23, ARG29 e ARG30 para o HIF1a) enquanto apenas dois aminoácidos são hidrofóbicos ALA25 e ALA26, ambos presentes na subunidade α do HIF1, o HIF1a. Essa prevalência de aminoácidos ligantes ao DNA hidrofílicos pode ser devido ao tipo de DNA ao qual a proteína está interagindo, pois quando há transcrição, os fatores de transcrição se ligam a DNAs no

formato B-DNA (formato do DNA ativo) e os nucleotídeos neste tipo de DNA interagem preferencialmente com aminoácidos com perfil hidrofílico.

Figura 26. Análise das características físico-químicas do domínio de ligação do fator de transcrição HIF1. (A) Representação tridimensional da superfície com destaque aos resíduos ligantes ao HRE conservados entre a família de proteínas bHLH. (B) Representação tridimensional do potencial eletrostático da molécula. Regiões em vermelho são negativamente carregadas, em azul são positivamente carregadas e em branco representam carga neutra. (C) Representação tridimensional do perfil de hidrofobicidade da molécula. Regiões em branco representam aminoácidos hidrofílicos enquanto que regiões em vermelho representam aminoácidos hidrofóbicos, sendo que quanto maior a intensidade da cor vermelha, mais hidrofóbico o resíduo em questão. α -H1(a) se refere a hélice- α 1 do ARNT e o α -H1(b) a hélice- α 1 do HIF1a.

Fonte: O autor (2021).

5.3. Simulação de dinâmica molecular do fator de transcrição HIF1

Com a finalização da análise da superfície da estrutura quaternária modelada para o HIF1, a análise estática deste fator de transcrição foi concluída e, posteriormente, submetido à simulação de 200 ns de simulação de dinâmica molecular (DM). Enquanto que a análise estática do modelo é feito com a proteína sem movimento algum, a DM tem como objetivo simular a proteína em um ambiente fisiológico com seus átomos em movimento e, assim, observar se a estrutura tridimensional do modelo se mantém estável durante sua movimentação em função do tempo.

As variações conformacionais do HIF1 foram medidas pela análise do RMSD e do Raio de giro (do inglês, *Radius of gyration*, RoG) durante o tempo de simulação. A Figura 27 (A) apresenta os resultados obtidos do RMSD dos átomos do *backbone* (N, Ca e C) das estruturas em relação à estrutura inicial para o modelo gerado por homologia para o fator de transcrição, enquanto que a Figura 27 (B) apresenta os resultados do RoG para o mesmo grupo de átomos da proteína, sendo que quanto menor o RoG, maior o grau de compactação (enovelamento) da proteína. Segundo o gráfico de RMSD, o HIF1 tende a se equilibrar próximo a 50 ns e permanece sem grandes movimentações pelos próximos 100 ns, com ponto máximo de 9,19Å e média de $7,44 \pm 0,46$ Å, e se mantém equilibrado até o fim da trajetória com ponto máximo de 9,39Å e média de $8,79 \pm 0,22$ Å. Em relação ao RoG, o modelo apresentou baixa compactação estrutural durante toda a trajetória, com ponto máximo de 31,50Å e média de $29,15 \pm 0,49$ Å, adquirindo um formato levemente enovelado no tempo final em relação a estrutura inicial.

É comum que haja um aumento progressivo no RMSD de uma proteína, partindo de 0, até um patamar, o que pode indicar a estabilização do sistema, sendo que este patamar pode variar em função das características da proteína, onde valores menores que 3 ou 4 Å tendem a indicar sistemas mais semelhantes a estrutura de referência ou nativa enquanto que valores muito altos de RMSD podem sugerir movimentos maiores dos *loops*, perda de estrutura secundária e distanciamento da conformação nativa da proteína. Além disso, os valores observados para o RMSD sofrem influências de diferentes regiões da proteína de acordo com o tipo de estrutura secundária presente nesta região. Por exemplo, normalmente conjuntos de hélices- α se modificam menos durante uma simulação do que regiões de *loops*. Quanto aos *loops*, é esperado que regiões da proteína com *loops* apresentem maior flutuação que os demais resíduos, visto eles são reconhecidos como regiões de alto coeficiente cinético (VERLI, 2014).

Figura 27. Valores do (A) Desvio Quadrático Médio (RMSD) e do (B) Raio de giro (RoG) ao longo de 200 ns para o fator de transcrição HIF1 (HIF1 α -ARNT) modelado por homologia após o refinamento da estrutura 3D.

Fonte: O autor (2021).

Os valores de RoG sofrem influência do tipo de estrutura secundária e quantidade de aminoácidos que a proteína possui. De acordo com a composição das estruturas secundárias, as proteínas podem ser classificadas em: proteínas α , cuja estrutura secundária é composta principalmente por hélices- α ; proteínas β , cuja estrutura secundária é composta principalmente por folhas- β ; proteínas α/β , as quais são estruturalmente compostas por hélices- α e folhas- β alternadas, nas quais as folhas- β são paralelas entre si; e proteínas $\alpha+\beta$, que possuem estrutura

secundária composta por hélices- α e folhas- β que ocorrem separadamente ao longo de toda a cadeia, assim, as folhas- β são antiparalelas entre si. As proteínas α têm o maior RoG, sugerindo menor enovelamento em comparação com proteínas β e $\alpha+\beta$, sendo que valores menores são característicos de proteínas α/β . Além disso, independente da composição da estrutura secundária da proteína, quanto maior o tamanho da cadeia polipeptídica, maior o RoG da proteína em questão, visto que proteínas com cerca de 350 aminoácidos possuem valores de aproximadamente 20Å para proteínas α/β (LOBANOV; BOGATYREVA; GALZITSKAYA, 2008).

Os altos valores de RoG observados para o HIF1 (Figura 27B) no tempo inicial podem ser resultantes da composição da estrutura secundária e tamanho da proteína, uma vez que as duas subunidades do complexo HIF1 são constituídas principalmente por hélices- α e folhas- β antiparalelas, classificando-as como proteínas $\alpha+\beta$, além de apresentarem tamanho considerável, com um total de 335 aminoácidos para o HIF1a, e 381 aminoácidos para o ARNT, totalizando 716 resíduos para o HIF1.

Outra grandeza utilizada na análise da trajetória é o RMSF, que descreve a variação da posição dos resíduos durante a simulação, indicando a flexibilidade do sistema. Na Figura 28 está ilustrado o RMSF para cada resíduo do HIF1. A representação da estrutura tridimensional com os resíduos apresentando RMSF acima de 3,0Å está ilustrada na Figura 29. Na maior parte, as regiões as quais se apresentaram altamente flexíveis são regiões de *loops* e os domínios de ligação ao DNA, principalmente as hélices- $\alpha 1$ de ambas subunidades do HIF1 (região da hélice- $\alpha 1$ do HIF1a indicada por seta rosa, e a região da hélice- $\alpha 1$ do ARNT indicada por uma seta azul no gráfico de RMSF) com um RMSF acima de 3,0Å. O fato de a proteína não estar ligada ao DNA pode ter resultado na alta flutuação observada para os resíduos ligantes ao DNA, visto que a proteína deve estar complexada ao seu elemento regulatório na região promotora do gene para exercer sua função.

Figura 28. Valores de RMSF para o fator de transcrição HIF1. Resíduos numerados de 1 a 381 equivalem a subunidade ARNT e os resíduos de 382 a 715 equivalem a subunidade HIF1a. A linha vermelha pontilhada demarca os aminoácidos com $\text{RMSF} > 3,0 \text{ \AA}$. A seta azul indica as hélice- $\alpha 1$ do ARNT (representado pelos resíduos 1 a 18) e a seta rosa indica a hélice- $\alpha 1$ do HIF1a (representados pelos resíduos 382 a 400).

Fonte: O autor (2021).

Figura 29. Representação tridimensional do HIF1 com os resíduos de maior flutuação. Os aminoácidos com $\text{RMSF} > 3,0 \text{ \AA}$ estão coloridos. A seta azul e rosa indicam a região da hélice- $\alpha 1$ das subunidades do ARNT e HIF1a, respectivamente.

Fonte: O autor (2021).

Por fim, as mudanças conformacionais ocorridas ao longo da trajetória são visualizadas na inspeção visual da estrutura do HIF1 em tempos específicos sobrepondo-as com a estrutura de referência ($t=0$ ns). Na Figura 30 são apresentadas as estruturas capturadas em tempos determinados na trajetória para o HIF1.

Em geral, o modelo refinado para o HIF1 teve suas estruturas secundárias preservadas até o final da trajetória, com exceção para o domínio bHLH. Nos primeiros 50 ns de simulação, as hélices- $\alpha 1$ N-terminais das subunidades α e β sofrem uma dobra diferindo de sua posição inicial, e desnovelam ao final dos 200 ns corridos, com a hélice- $\alpha 1$ do ARNT parcialmente desnovelada e com a hélice- $\alpha 1$ do HIF1a totalmente desnovelada. A desnaturação dessa região do domínio de ligação ao DNA justifica os altos valores encontrados para RMSD ao longo da trajetória e a alta flutuação observada para os resíduos deste domínio observados pelo RMSF anteriormente

Figura 30. Visualização das estruturas capturadas em tempos específicos da trajetória para o modelo refinado do HIF1 (A) e sobreposição das estruturas nos tempos 0, 50, 150 e 200ns (B). A seta azul e rosa indicam a região da hélice- $\alpha 1$ das subunidades do ARNT e HIF1a.

Fonte: O autor (2021).

Em condições de hipóxia, o HIF1 α se estabiliza e se liga ao ARNT para formação de um fator de transcrição funcional, onde se ligará a elementos regulatórios específicos (HRE) para ativar a transcrição de genes alvos. Dessa forma, o desnovelamento do domínio bHLH observado durante a simulação de dinâmica molecular pode ser devido à ausência do DNA ligado ao HRE, visto que a interação com o HRE é importante para exercer sua função. Apesar disso, na literatura há relatos que a subunidade α do HIF1 contribui no controle transcricional durante baixa tensão de oxigênio celular sem a necessidade de se ligar ao seu elemento regulatório, por meio da interação do bHLH com outras proteínas ligantes ao DNA, como o domínio ICD (do inglês, *intracellular domain*, ICD) da proteína *Notch* (GUSTAFSSON et al., 2005; PISTOLLATO et al., 2010). Assim, em meio fisiológico, o domínio de ligação ao DNA estará associado ao HRE ou a outras proteínas para sua estabilização e ação na transcrição gênica.

5.4. Modelagem molecular da dupla hélice de DNA.

Feito a estrutura tridimensional do complexo proteico HIF1, seguiu-se com a modelagem da dupla fita de DNA e consequentemente o *docking* proteína-DNA. O HRE presente na região promotora do gene *HLA-G* foi modelado pelo *software* x3DNA v.2.4 (LU; OLSON, 2008), compreendendo a sequência -980 a -947 *upstream* da região promotora do gene *HLA-G*, a qual contém o polimorfismo bialélico -964 G/A (rs1632947) (CASTELLI et al., 2014b, 2014a) e o HRE entre as posições -966 a -962 (CASTELLI et al., 2014a) (Figura 31). A sequência 5'TAAAAACAGGCAGTGCRTGAG CACTAGTGAGGG'3 foi utilizada para construir um DNA do tipo B (B-DNA) dupla fita, sendo a segunda fita complementar construída a partir da sequência de entrada.

As duas moléculas de DNA geradas, uma para o alelo -964G e outra para o alelo -964A, são idênticas quanto a sua conformação estrutural, com exceção para sua sequência de nucleotídeos diferindo na posição -964 e seu par complementar na outra fita. Além da estrutura do DNA tipo B, duas outras conformações principais são mais comumente descritas chamados de A-DNA e Z-DNA (FRANK-KAMENETSKII, 1995; USSERY, 2002), porém, o B-DNA foi a conformação escolhida para a modelagem do HRE da região promotora do gene *HLA-G* porque equivale a forma biologicamente predominante do DNA no organismo.

Figura 31. Representação da dupla fita de DNA para o HRE considerando o polimorfismo -964 G/A do gene *HLA-G*. (A) Sequência linear da região promotora -980 a -947 do gene *HLA-G* com o polimorfismo -964 G/A em vermelho e o HRE em amarelo. O **R** é referência ao código IUPAC, onde sinaliza que nesta posição (-964) pode haver uma guanina ou adenina, enquanto que o * refere-se ao par complementar do SNP em questão. (B) Representação tridimensional da dupla fita de DNA modelada pelo x3DNA para a região promotora -980 a -947 do gene *HLA-G* com o HRE em amarelo e o polimorfismo -964 G/A em vermelho.

Fonte: O autor (2021).

Sua estrutura consiste em uma dupla hélice espiralada para a direita com as duas hélices antiparalelas enrolando-se sobre a superfície sendo tais hélices complementares entre si obedecendo o pareamento de bases de Watson-Crick. A principal característica dessa forma de DNA é a orientação quase perpendicular das bases em relação ao eixo da hélice e o sulco maior e sulco menor entre as cadeias de açúcar-fosfato: o sulco menor é relativamente pequeno enquanto que o sulco maior é grande e profundo com acesso facilitado aos pares de bases por moléculas regulatórias, como os fatores de transcrição. A molécula observada em ambientes fisiológicos não corresponde a uma molécula de B-DNA perfeita, uma vez que a formação de complexos proteína-DNA afeta a sua estrutura (HO, P. S. AND CARTER, 2011.). Apesar disso, foi observado que a família bHLH de fatores de transcrição geralmente não introduz distorções (como curvatura acentuadas) na estrutura do DNA (ELLENBERGER et al., 1994; FERRÉ-D'AMARÉ et al., 1994; SCHULTE et al., 2017; SHIMIZU, 1997; WU et al., 2015, 2016).

5.5. Docking molecular dos complexos HIF1-HRE-964G/A

O *docking* molecular é um método muito utilizado na biologia estrutural para explorar a interação de resíduos entre duas moléculas. Com as estruturas tridimensionais do HIF1 e do DNA (HRE-964G/A) modeladas, é possível realizar o *docking* proteína-DNA por meio do

servidor HADDOCK v.2.4 (DE VRIES; VAN DIJK; BONVIN, 2010). O HIF1 terá seu domínio de ligação ao DNA ancorado ao seu elemento regulatório na região promotora do gene *HLA-G* enfatizando o polimorfismo existente dentro do HRE. Assim, dois complexos proteína-DNA serão formados por meio da técnica de *docking* molecular utilizando o mesmo protocolo: o complexo HIF1-HRE-964G e HIF1-HRE-964A, os quais foram nomeados de acordo com o alelo presente no sítio polimórfico -964 no interior do HRE da região promotora do gene *HLA-G*.

Em resumo, o HADDOCK se baseia em informações da interface de interação do complexo molecular preditas *in vitro* ou *in silico*, as quais são introduzidos como restrições de interação ambíguas (do inglês, *Ambiguous Interaction Restrains*, AIRs) para direcionar o *docking*. Tais AIRs são definidas através de uma lista de resíduos que se enquadram em duas categorias: ativa e passiva (DE VRIES; VAN DIJK; BONVIN, 2010), sendo que uma seleção cuidadosa de quais resíduos são ativos e quais são passivos é fundamental *para* o sucesso do *docking*.

Os resíduos ligantes ao DNA conservados entre a família bHLH de fatores de transcrição descritos anteriormente (Figura 25) foram selecionados como resíduos ativos para o HIF1 (HIF1a-ARNT) para direcionar o processo de *docking* de ambos os complexos moleculares, HIF1-HRE-964G e HIF1-HRE-964A, enquanto que os resíduos ativos para molécula de DNA foram selecionados baseado na sequência central do HRE 5'GCRTG'3 (WENGER; STIEHL; CAMENISCH, 2005), compreendendo os nucleotídeos nas posições -966 a -962 da região promotora do gene *HLA-G* (CASTELLI et al., 2014b), onde na posição “R” pode haver o alelo G (complexo HIF1-HRE-964G) ou alelo A (HIF1-HRE-964A). Os resíduos passivos foram automaticamente selecionados em torno dos resíduos ativos pelo servidor.

Para o complexo HIF1-HRE-964G, o HADDOCK agrupou 313 estruturas em 23 *clusters*, o que representa 78% dentre os 400 modelos refinados em água que o HADDOCK gerou, enquanto que, para o complexo HIF1-HRE-964A, o HADDOCK agrupou 326 estruturas em 25 *clusters*, representando 81% dos 400 modelos refinados em água, sendo tais *clusters* classificados de acordo com o seu HADDOCK *score* (HS) (quanto mais negativo melhor) segundo o servidor.

Em ambos os *dockings* realizados, as quatro melhores *poses* agrupadas nos 10 primeiros *clusters* ordenados de acordo com o HADDOCK *score* (Apêndices 1 e 2) foram analisadas quanto a sua similaridade estrutural com um determinado complexo referência pertencente à família bHLH-PAS de fatores de transcrição, o complexo HIF2a-ARNT-HRE de camundongo

(PDB ID: 4ZPK), por meio do cálculo do RMSD pela sobreposição das estruturas tridimensionais no *software* PyMol v.2.4 (SCHRÖDINGER, 2020), a fim de identificar uma pose com estrutura semelhante à forma clássica dos complexos dessa família de proteínas. Como resultado, a maioria dos *clusters* apresentaram poses das quais a orientação do HRE ligado ao HIF1 diferia do esperado, de acordo com as estruturas resolvidas experimentalmente para as proteínas da mesma família do HIF1a que interagem com ARNT. No entanto, o segundo *cluster* melhor classificado para o HIF1-HRE-964G, o *cluster* 20, e o quarto *cluster* com melhor classificação para o HIF1-HRE-964A, o *cluster* 2, apresentaram poses mais similares ao complexo referência dentre os demais *clusters*. Nas Figuras 32A e B são mostrados os resultados dos alinhamentos das 4 poses dos *clusters* 20 e 2 com o HIF2-HRE, respectivamente. As poses de ambos os *clusters* apresentaram RMSD menor que o nível de tolerância de 2,0 Å quando alinhados ao cristal demonstrando boa similaridade estrutural.

As características físico-químicas das moléculas parceiras definem o grau de afinidade entre elas, e estão diretamente relacionadas com as interações intermoleculares existentes no complexo. Por isso as funções de energia derivadas de campos de forças e empíricas são usadas para delinear as poses corretas de poses incorretas geradas no *docking* através da estimativa da afinidade de ligação entre as moléculas ordenando as diferentes poses de acordo com suas afinidades (MENG et al., 2011). Por sua vez, a afinidade de ligação entre as moléculas está relacionada com a energia livre de Gibbs (ΔG). Em analogia a qualquer processo espontâneo, a ligação entre duas moléculas apenas ocorre quando a mudança da energia livre de Gibbs (ΔG) do sistema é negativa indicando que o sistema atingiu um estado de equilíbrio com pressão e temperatura constantes (KITCHEN et al., 2004). Assim, as melhores poses geradas no processo de *docking* são aqueles com valores de energia de interação intermoleculares negativos, por indicar provável interação biomolecular (DE VRIES; VAN DIJK; BONVIN, 2010; KITCHEN et al., 2004).

Figura 32. Comparação entre a estrutura tridimensional do complexo referência, o HIF2a-ARNT-HRE (PDB ID: 4ZPK), e as poses agrupadas no (A) cluster 20, referente ao complexo HIF1 (HIF1a-ARNT) -HRE-964G, e no (B) cluster 2, referente ao complexo HIF1 (HIF1a-ARNT) -HRE-964A. A sequência cis regulatória conservada entre o HREs está destacada em verde nas poses e em amarelo na estrutura referência.

Fonte: O autor (2021).

Em especial, o HADDOCK ordena as melhores poses baseado em uma função de *score* que consiste na combinação linear de vários termos de energia, o HADDOCK *score* (HS), sendo que a pose com menor HS é considerada a estrutura mais energeticamente favorável. Assim, as poses melhores classificadas para o HS nos *clusters* 20 e 2 foram escolhidas como estrutura final dos complexos HIF1-HRE-964G e HIF1-HRE-964A, respectivamente. Na Tabela 4 é

mostrado os valores para os termos de energia de ligação para as estruturas finais dos complexos HIF1-HRE-964G e HIF1-HRE-964A gerados pelo HADDOCK.

Tabela 4. Comparação dos valores para os termos de energia de ligação entre os complexos HIF1-HRE-964G e HIF1-HRE-964A gerados pelo HADDOCK.

Complexo	HIF1-HRE-964G	HIF1-HRE-964A
HS (a.u.)	-163,53 (-152,70±6,60)	-147,90 (-140,60± 4,60)
Evw (kcal/mol)	-84,98 (-87,70±3,90)	-84,37 (-81,00± 3,90)
Eelec (kcal/mol)	-536,02 (-478,20±38,90)	-435,78 (-425,80± 23,10)
Eair (kcal/mol)	3,68 (12,10±13,00)	4,70 (7,80±3,00)
Edesol(kcal/mol)	28,29 (29,40±2,40)	23,15 (24,80±1,50)
BSA (Å)	2195,65 (2241,2±122,40)	2050,90 (2065,30± 86,90)

HS: HADDOCK *score*; EVdW: Energia de Van der Waals; Eelec: Energia Eletrostática; Edesol: Energia de dessolvatação; Eair: Energia de violação das restrições; BSA: Área de superfície enterrada (do inglês, Buried Surface Area, BSA; a.u.: unidades arbitrárias (do inglês, *arbitrary units*, a.u.); A média e o desvio padrão foram calculados a partir dos 4 principais modelos dos *clusters* 20, para o HIF1-HRE-964G, e 2, para o HIF1-HRE-964A.

De acordo com a Tabela acima, o valor do HS para o complexo HIF1 ligado ao HRE com o alelo -964G (HS: -163,53) foi menor em comparação com o HIF1 ligado ao HRE contendo o alelo A (HS: -147,90). Essa diferença sugere que o complexo HIF1-HRE-964G é mais energeticamente favorável do que o complexo HIF1-HRE-964A. Além disso, o complexo HIF1-HRE-964G apresentou maior valor para a área de superfície enterrada em relação ao fator de transcrição ligado ao HRE com o alelo -964A presente em seu interior, indicando maior interface de interação entre as moléculas parceiras no primeiro complexo. Tais dados sugerem que a presença do alelo A no sítio -964 do HRE na região promotora do *HLA-G* implica em diminuição da região de interação entre o HIF1 e o HRE.

5.6. Análise do padrão de ligações intermoleculares dos complexos HIF1-HRE-964G/A

Nas últimas décadas, o aumento do número de estruturas de alta resolução de complexos proteína-DNA depositadas no PDB e o avanço das tecnologias para explorar elementos regulatórios no DNA, como ChIP-seq, permitiu maior expansão do conhecimento sobre a especificidade de ligação proteína-DNA (ROHS et al., 2010; SLATTERY et al., 2014). Por sua vez, a especificidade de ligação de complexos proteína-DNA pode ser compreendida por interações atômicas entre aminoácidos e nucleotídeos, a qual pode ser alcançada por meio de ligações de hidrogênio e contatos de Van der Waals (VdW) (MANDEL-GUTFREUND; SCHUELER; MARGALIT, 1995). Tais interações ocorrem entre as cadeias laterais das

proteínas com o esqueleto açúcar-fosfato e com as bases do DNA nos sulcos maiores e menores, sendo que o primeiro tipo de interação é importante para estabilizar o complexo molecular e o segundo tipo de interação, também chamado de leitura de base, é responsável pela especificidade (LUSCOMBE; THORNTON, 2002; MATTHEWS, 1988; PABO; SAUER, 1984). Visto que variações genéticas em elementos regulatórios podem alterar a especificidade da ligação de complexos proteína-DNA (IZUMI, 2016), a análise do padrão de ligação intermolecular do HIF1 ligado ao HRE em função do polimorfismo -964 G/A na região promotora do *HLA-G* é importante para compreensão da regulação da transcrição do gene.

A análise das ligações intermoleculares foi feita por meio do servidor DNAProDB (SAGENDORF et al., 2019) e visualizadas no PyMol v.2.4 (SCHRÖDINGER, 2020). Ambos os complexos se ligam a seus elementos regulatórios por meio da formação de ligações de hidrogênio e interações de VdWs entre os resíduos da hélice- α 1 das subunidades do HIF1 com as bases nitrogenadas no sulco maior do DNA e com o esqueleto açúcar-fosfato. Quando o HIF1 se liga ao HRE contendo o alelo alternativo -964A no sítio polimórfico -964G>A do *HLA-G*, 11 ligações de hidrogênio e 87 contatos de VdWs estabilizam como um todo o complexo, enquanto que na presença do alelo G, 16 pontes de hidrogênio 81 contatos de VdWs são observadas.

Como esperado, os contatos de VdWs são mais abundantes em complexos proteína-DNA do que as ligações de hidrogênio e estabilizam o complexo molecular como um todo, porém as pontes de hidrogênio são as mais fortes dentre as ligações não covalentes (LUSCOMBE et al., 2000; LUSCOMBE, 2001; LUSCOMBE; THORNTON, 2002; MANDEL-GUTFREUND; SCHUELER; MARGALIT, 1995). Dessa forma, a análise e discussão dos resultados será focada principalmente nos contatos aminoácido-DNA formados pelas pontes de hidrogênio. Das ligações de hidrogênio existentes entre os complexos moleculares estudados, oito e seis se formam compreendendo a região do HRE para os complexos HIF1-HRE-964G e HIF1-HRE-964A, respectivamente, como observado na Figura 33.

Em relação ao complexo HIF1-HRE-964G, cinco resíduos se ligam ao HRE por átomos do esqueleto açúcar-fosfato (R91 e R102 para o ARNT, e K19, R27 e R29 para o HIF1 α), enquanto que apenas o resíduo GLU98 se liga diretamente a base nitrogenada. Os aminoácidos K19 e R27 formam duas pontes de hidrogênio com os nucleotídeos -966rC e -964rC,

respectivamente. Para o complexo HIF1-HRE-964A, cinco ligações de hidrogênio são formadas entre pares de átomos dos aminoácidos e o esqueleto açúcar-fosfato do HRE, as quais

Figura 33. Interações intermoleculares entre átomos do HRE e do HIF1 enfatizando o polimorfismo -964G>A do gene *HLA-G*. (A) Representação tridimensional das pontes de hidrogênios formadas nos complexos HIF1-HRE-964G/A em relação ao seu elemento regulatório. A caixa central a esquerda representa a região do HRE (destacada em vermelho na dupla hélice de DNA, e sua sequência complementar destacada em areia) LH do HIF1, e a direita são mostrados detalhes das ligações de hidrogênio formadas entre os átomos dos resíduos do ARNT (azul) e HIF1a (rosa) com os átomos dos nucleotídeos. As ligações de hidrogênio (distância 2,5-3,2Å) estão indicadas por linha pontilhada amarela. As caixas vermelhas indicam nucleotídeos do HRE (no sentido 5'-3') e caixas na cor areia indicam sua sequência complementar (no sentido 3'-5'). Os átomos de hidrogênio estão coloridos de branco; fosfato em laranja; oxigênio em vermelho; nitrogênio em azul; e os átomos de carbono têm cores diferentes entre as moléculas parceiras: carbono pertencente ao DNA em marrom, carbono pertencente ao HIF1a em rosa e carbono pertencente ao ARNT em azul claro. (B) Diagrama de interação 2D com as pontes de hidrogênio existentes nos complexos HIF1-HRE-964G>A em relação ao HRE. As pontes de hidrogênio estão indicadas por linhas pontilhadas amarelas; a sequência central do HRE ilustrada em caixa vermelha e sua sequência complementar em caixa de coloração areia, aminoácidos pertencentes ao ARNT e ao HIF1a sublinhados em azul e rosa, respectivamente.

Fonte: O autor (2021).

são compartilhadas entre os dois complexos moleculares, com exceção da ligação formada pelo R30 do HIF1a com a base nitrogenada do nucleotídeo -964rT, que é observado apenas no complexo HIF1-HRE contendo o alelo -964A. Nenhum resíduo HIF1 formou pontes de hidrogênio diretamente com o sítio de variação G/A, apenas as forças VdW atuam na interação

com tais nucleotídeos (R102-964G para o HIF1-HRE-964G e R102-964A, R99-964A, E98-964A para o HIF1-HRE-964A). Em contraste, ligações de hidrogênio são formadas com as bases complementares dos alelos -964G/A na fita *reverse* (direção 3'-5') através do radical fosfato da cadeia de DNA (interação R27-964rT/C na Figura 33), e esta ligação ocorre independentemente do alelo presente no sítio polimórfico. Além dos contatos de VdW que se formam entre aminoácidos e os alelos -964G e A, este tipo de interação também é observada nos complexos HIF1-HRE entre os resíduos-nucleotídeos K128-966G, E98-963T, E98-962rC, S95-963T, H94-962G e R91-962G no ARNT, e R30-965C, R30-964rC, R30-963rA e R23-965rG no HIF1a, para o complexo com alelo G, e H94-963T e R91-963T no ARNT, e R30-965C, R30-963rA, R23-965rG e S22-966rC no HIF1a, para o complexo com alelo A (dados não mostrados).

A presença do alelo A no sítio -964 no interior do elemento regulatório do complexo HIF1-HRE-964A alterou a geometria do DNA, resultando em perda de interação entre moléculas parceiras, como as ligações de hidrogênio entre K19-966rC e E98-962rC (Figura 33) observado para o complexo com o alelo G. O número de pontes de hidrogênio que um aminoácido forma com o nucleotídeo é diretamente proporcional a força de ligação do complexo (LUSCOMBE, 2001), assim, o complexo HIF1-HRE-964G é mais estável do que o HIF1-HRE-964A por apresentar dois aminoácidos formando dupla ligação com os nucleotídeos. O aminoácido GLU98 é altamente conservado entre os fatores de transcrição da família bHLH-PAS (SCHULTE et al., 2017; WU et al., 2015, 2016), os quais compartilham padrão de ligação semelhante em relação ao GLU98. Além disso, estudos prévios demonstraram que o aminoácido GLU98 é essencial para a interação com o HRE de forma adequada (BACSI; HANKINSON, 1996). Dessa forma, é possível inferir que o alelo -964A diminui a especificidade de ligação do HIF1 pelo HRE, assim como a estabilidade do complexo molecular.

Além dos resíduos mostrados na Figura 33, outros foram identificados formando pontes de hidrogênios com nucleotídeos adjacentes ao HRE interagindo com átomos das bases nitrogenadas (H94-961rT, S22-968G e S22-967T para o HIF1-HRE-964G; S22-967T e S22-968G para o HIF1-HRE-964A) e átomos do esqueleto açúcar-fosfato (R101-961rT, K21-969A e R29-967T para o HIF1-HRE-964G; K21-969A e R29-967T para o HIF1-HRE-964A).

A partir da análise do padrão de ligações de hidrogênio nos complexos HIF1-HRE em função do polimorfismo -964G>A, observa-se diminuição na interação do HIF1 com o DNA

na presença do alelo A no interior do HRE em comparação com o alelo G. Assim, conclui-se que o HIF1 se liga de forma mais estável e específica com o HRE contendo o alelo -964G.

5.7. Simulação de dinâmica molecular dos complexos HIF1-HRE-964G>A

Após a análise do padrão de ligação intermolecular dos complexos, prosseguiu com a realização da simulação de dinâmica molecular para verificar o comportamento dinâmico dos átomos dos complexos proteína-DNA em função do polimorfismo existente dentro do elemento regulatório responsivo a hipóxia (HRE) e, assim, avaliar a estabilidade dos complexos ao longo do tempo. Tal como para a proteína livre, as variações conformacionais do HIF1-HRE964G/A foram medidas pela análise do RMSD, RoG e RMSF durante 200 ns de simulação. A Figura 34 apresenta os resultados obtidos para o RMSD (A) e RoG (B) dos átomos do *backbone* (N, Ca e C) das estruturas em relação à estrutura inicial do fator de transcrição ligado ao complexo molecular.

Figura 34. Valores do (A) Desvio Quadrático Médio (RMSD) e do (B) Raio de giro (RoG) ao longo de 200 ns para o fator de transcrição HIF1 (HIF1a-ARNT) ligado ao HRE em função do polimorfismo -964G>A do gene *HLA-G*.

Continuação Figura 34.

(B)

Fonte: O autor (2021).

Em relação ao gráfico de RMSD, na presença do alelo G no sítio -964 do HRE na região promotora do gene *HLA-G*, o HIF1 atinge um patamar de equilíbrio nos primeiros 10 ns de simulação e permanece sem grandes movimentações até o fim da trajetória com ponto máximo de 6,95Å e média de $5,27 \pm 0,35$ Å, enquanto que, na presença do alelo -964A, o HIF1 se estabiliza em 60 ns com ponto máximo de 8,51Å e média de $7,19 \pm 0,35$ Å até o final dos 200 ns. O fator de transcrição apresentou baixa compactação estrutural durante toda a trajetória independente do alelo presente no HRE como observado nos resultados referentes ao RoG (Figura 34B), com ponto máximo de 32,21Å e média de $31,08 \pm 0,46$ Å para o HIF1-HRE-964G e ponto máximo de 32,00Å e média de $31,18 \pm 0,17$ Å para o HIF1-HRE-964A. O HIF1 tem maior estabilidade estrutural quando ligado ao HRE contendo o alelo -964G quando comparado ao complexo proteína-DNA com o alelo -964A (Figura 34A). Além disso, o fator de transcrição ligado ao HRE com o alelo A no sítio polimórfico -964G>A atinge valores de RMSD semelhantes aos da proteína livre, reforçando a hipótese de que a ligação do HIF1 ao HRE com o alelo G é mais estável. O enovelamento da proteína não diferiu entre os dois complexos moleculares, o qual atingiu valores de RoG similares ao da proteína livre no tempo inicial, porém, a proteína livre apresentou uma leve compactação ao decorrer da simulação, enquanto que os complexos permaneceram sem alterações de acordo com o gráfico. Também foi analisado o RMSD e RoG para o DNA (Figura 35A e B, respectivamente), e uma alta flexibilidade foi observada nos dois complexos moleculares.

Figura 35: Valores do (A) Desvio Quadrático Médio (RMSD) e do (B) Raio de giro (RoG) ao longo de 200 ns para a cadeia de DNA dos complexos HIF1-HRE-964G>A.

Fonte: O autor (2021).

O RMSF oferece informações sobre a flexibilidade de áreas específicas das moléculas. Nas Figuras 36 (A) e (B), estão ilustrados os resultados do RMSF para cada resíduo do HIF1 ligado ao HRE-964G>A e para os nucleotídeos da cadeia de DNA. O HIF1 demonstrou um padrão de flutuação muito similar entre os dois complexos proteína-DNA, diferindo apenas para algumas regiões de *loops*, onde o fator de transcrição ligado ao HRE com o alelo A obteve maior flutuação nestes resíduos do que o HRE com o alelo G. A ligação do elemento regulatório

no domínio bHLH estabilizou as regiões da hélice- $\alpha 1$, região a qual estão localizados os resíduos de ligação ao DNA nas subunidades ARNT e HIF1 α (destacados pelas setas azul e rosa na Figura 34, respectivamente) do heterodímero em ambos os complexos HIF1-HRE-964G>A, devido a menor flutuação nestas regiões em comparação a proteína livre.

Figura 36. Valores de RMSF para (A) o fator de transcrição HIF1 ligado ao HRE considerando o polimorfismo -964G>A do *HLA-G* em comparação com o HIF1 livre e para (B) a cadeia de DNA. (A) Resíduos numerados de 1 a 381 equivalem a subunidade ARNT e os resíduos de 382 a 715 equivalem a subunidade HIF1 α . A linha vermelha pontilhada demarca os aminoácidos com RMSF>3,0Å. A seta azul e rosa indicam as hélices- $\alpha 1$ do ARNT (Aminoácidos 1 a 18) e HIF1 α (Aminoácidos 382 a 400), respectivamente. (B) Resíduos numerados de 15 a 19 equivalem ao elemento responsivo a hipóxia, e estes nucleotídeos bem como sua sequência complementar (fcHRE) estão destacados pelo caixa verde no gráfico. Cada fita de DNA contém 33 bp. fc: Fita complementar.

Fonte: O autor (2021).

Quanto ao ácido nucleico, flutuações foram observadas principalmente nos nucleotídeos terminais da dupla-hélice de DNA em ambos os complexos moleculares por meio da análise do

RMSF, enquanto que o local de ligação da proteína foi considerado estável (Figura 34). A alta flutuação dos nucleotídeos terminais em ambos os complexos proteína-DNA justifica a alta flexibilidade conformacional observada para os valores de RMSD e RoG para o DNA.

A inspeção visual dos complexos proteína-DNA foi feita para visualizar as mudanças conformacionais ocorridas ao decorrer da etapa de produção da simulação de dinâmica molecular. Foram extraídos as coordenadas atômicas dos complexos moleculares em tempos específicos da trajetória e sobrepostas a estrutura inicial ($t=0$ ns) por meio do *software* PyMol v.2.4 (SCHRÖDINGER, 2020). Na Figura 37 são apresentadas as estruturas capturadas em tempos determinados na trajetória para os complexos HIF1-HRE em função do polimorfismo -964G>A do gene *HLA-G*. O HIF1 se manteve ligado ao DNA em ambos os complexos ao longo do tempo e as estruturas secundárias do fator de transcrição mantidas. Como observado na Figura 37 (C) e (D), as regiões de *loop* do HIF1 e as extremidades da cadeia de DNA são regiões com maior cinética nos dois complexos moleculares analisados, corroborando os altos valores de RMSF observados para tais regiões das moléculas.

Estudos prévios demonstraram que o domínio bHLH do HIF1 se liga de forma estável ao HRE durante uma simulação de DM de 10ns, porém, os valores de RMSD para o fator de transcrição são considerados elevados, sendo que os autores justificam que a ausência dos domínios PAS-A/B e PAC no seu modelo estrutural teórico pode ter sido responsável por tais valores, uma vez que estes domínios são responsáveis por estabilizar o heterodímero (SOKKAR; SATHIS; RAMACHANDRAN, 2012). Em primeiro, para concluir quanto a estabilidade de um dado complexo molecular é desejável que o tempo de simulação seja maior do que 10ns e, em segundo, nos complexos HIF1-HRE do nosso estudo, além do domínio de ligação ao DNA (bHLH), os domínios PAS-A/B e PAC estão presentes, contudo, os valores de RMSD continuam altos. De fato, dentre os complexos HIF1-HRE-964G/A, o fator de transcrição ligado ao HRE com o alelo G no sítio polimórfico -964G>A apresentou maior estabilidade em relação ao complexo HIF1-HRE contendo o alelo A, mas os valores de RMSD e RoG ainda são considerados elevados. Isso pode ocorrer devido as subunidades do HIF1, o ARNT e o HIF1 α , apresentarem regiões intrinsecamente desordenadas (do inglês, *intrinsically disordered regions*, IDRs) situadas nas regiões C-terminais da proteína e N-terminais do domínio de ligação ao DNA (KOLONKO; GREB-MARKIEWICZ, 2019; TARCZEWSKA; GREB-MARKIEWICZ, 2019). Neste contexto, proteínas intrinsecamente desordenadas (do inglês, *intrinsically disordered proteins*, IDPs) e IDRs não apresentam uma conformação estável e única, podendo adotar diferentes conformações de acordo com o parceiro molecular

que interage para exercer sua função biológica (ANFENSEN, 1973). As propriedades das IDRs são proporcionadas pela sua composição de aminoácidos e podem apresentar diferentes comprimentos de resíduos acompanhados por domínios bem estruturados (DYSON; WRIGHT, 2005). Assim, a aquisição de estabilidade proteica ocorre quando ligado ao seu parceiro molecular (OLDFIELD et al., 2019).

Figura 37. Visualização das estruturas capturadas em tempos específicos da trajetória das estruturas nos tempos 0 (A) e 200ns (B) para os complexos HIF1-HRE-964G e HIF1-HRE-964A. Na Figura (C) está representado a sobreposição das estruturas em tempos específicos (0, 50, 150 e 200 ns) da trajetória para os complexos HIF1-HRE964G/A, onde a estrutura no tempo inicial está ilustrado de amarelo e a estrutura no tempo final está ilustrado em vermelho, as demais Figuras estão ilustradas em cinza. A seta azul e rosa indicam a região da hélice- $\alpha 1$ das subunidades do ARNT e HIF1a, respectivamente. O HRE está destacado em verde na estrutura 3D do DNA.

Fonte: O autor (2021).

Estudos *in vitro* demonstraram que, em condições semelhantes a hipóxia, o HIF1 ligado ao HRE localizado na região promotora do *HLA-G* (-966 HRE) não induz a transativação do gene e, consequentemente, a produção da molécula individualmente, mas apenas em ação conjunta com outros HREs localizados no éxon 2 (posições +281 *sense* e +291 *antisense*) (YAGHI et al., 2016). A atuação conjunta destes dois HREs no *HLA-G* para regular positivamente a transcrição do gene a fim de resultar na produção da molécula é possível por meio de proteínas mediadoras chamadas cofatores, como o co-ativador CBP/p300 (SCHÖDEL et al., 2011; SEMENZA et al., 1996). Por sua vez, a formação de um *loop* no DNA possibilita a conexão dos fatores de transcrição com elementos regulatórios alvos localizados a longas distâncias do promotor ao complexo de transcrição basal por meio das proteínas mediadoras, resultando no início do processo de transcrição gênica (KOLOVOS et al., 2012; WEAKE; WORKMAN, 2010). Assim, por meio da formação de um *loop* no DNA, os HREs localizados na região promotora e no éxon 2 do gene *HLA-G* se tornam acessíveis a ligação do HIF1 em situações de hipóxia (GARZIERA et al., 2017).

A ausência dos HREs no éxon 2 do gene *HLA-G* e dos cofatores em nosso sistema, os complexos HIF1-HRE-964G>A, bem como a falta do domínio proteico N-TAD ao qual o mediador interage no HIF1, pode ter influenciado nos altos valores de RMSD e RoG, pois estes auxiliam na estabilidade do fator de transcrição. Por outro lado, o complexo contendo o HRE com alelo G se apresentou mais estável que o outro nos resultados da dinâmica molecular, corroborando dados experimentais prévios que, apesar do HRE na região promotora do gene *HLA-G* não conseguir induzir a produção da proteína sozinho *in vitro*, o alelo G apresentou maior afinidade ao HIF1 do que alelo A (YAGHI et al., 2016).

Devido à alta flexibilidade existente no DNA, a ligação de proteínas, como os fatores de transcrição, induz mudanças na geometria da dupla hélice de DNA e a deformação causada pela proteína na estrutura do ácido nucleico é essencial para o reconhecimento molecular, o qual é chamado mecanismo de leitura indireta (LU; OLSON, 1999; MICHAEL GROMIHA et al., 2004; SARAI et al., 1989). A fim de analisar as diferenças nos parâmetros conformacionais da estrutura 3D da dupla hélice de DNA durante a simulação de dinâmica molecular dos complexos HIF1-HRE em função do polimorfismo -964G>A do gene *HLA-G*, as coordenadas atômicas dos complexos proteína-DNA foram extraídas da trajetória a cada 10 ns, iniciando no tempo $t=0$ ns e, assim, um total de 21 estruturas foram analisadas para cada complexo molecular pelo programa Curves v.2.6 (LAVEY et al., 2009). Por meio deste *software* foi feito a análise dos parâmetros conformacionais locais na dupla hélice a nível de pares de bases inter cadeias a

cada *step/bp-step* (dinucleotídeos subjacentes), chamado de *inter-bp step* (do inglês, *inter-base pair step* (DICKERSON, 1989; LU; OLSON, 1999). A média e o desvio padrão foram calculados a partir das 21 estruturas extraídas da trajetória de cada complexo molecular para os parâmetros conformacionais analisados e os resultados são mostrados nas Figuras 38 e 39 para o HRE dos complexos moleculares analisados considerando o polimorfismo -964G>A do *HLA-G*.

Os parâmetros *inter-bp step* são dependentes da sequência de nucleotídeos da cadeia de DNA e são analisados para cada dinucleotídeo dentro das três classes químicas observadas para o HRE: pirimidina-purina (*YR step*), purina-purina (*RR step*) e purina-pirimidina (*RY step*). Os parâmetros *inter-bp step* compreende três distancias translacionais, *rise*, *shift* e *slide* (Figura 38), e três ângulos rotacionais, *roll*, *twist* e *tilt* (Figura 39). Em relação as distâncias translacionais, de acordo com a Figura 38, os complexos moleculares apresentaram valores muito similares para a maioria dos dinucleotídeos do HRE, independente do alelo presente no sítio polimórfico -964G>A.

Figura 38. Comparação dos valores médios (Å) do shift, slide, e rise calculados para o HRE a partir das coordenadas atômicas extraídas da trajetória dos complexos HIF1-HRE-964G>A a cada 10 ns.

Fonte: O autor (2021).

Quanto aos ângulos rotacionais ilustrados na Figura 39, a presença do polimorfismo gênico causou leves deformações na estrutura da dupla hélice referente ao HRE: (i) o *roll* variou cerca de 4 graus no primeiro e terceiro *bp step* (RY e YR *steps*, respectivamente) do HRE entre os complexos moleculares analisados; (ii) o ângulo *tilt* deslocou 4 graus no segundo *bp step*, ou seja, nos dinucleotídeos dCG e dCA para os complexos contendo o HRE com o alelo G e A, respectivamente; (iii) o ângulo *twist* apresentou uma variação de aproximadamente 3 graus no segundo e quarto *bp step* (dTG), sendo que no segundo *bp step*, o complexo contendo o HRE com o alelo A tem um valor maior para o *twist* em relação ao complexo com o HRE-964G e, então, no quarto *bp step* ocorre uma conversão de valores para o *twist* entre os complexos moleculares.

Figura 39. Comparação dos valores médios (°) do roll, tilt, e twist calculados para o HRE a partir das coordenadas atômicas extraídas da trajetória dos complexos HIF1-HRE-964G>A a cada 10 ns.

Fonte: O autor (2021).

Geralmente, tais parâmetros estruturais podem ou não se combinar para descrever as deformações na dupla hélice de DNA. Por exemplo, a curvatura nos dinucleotídeos ocorre

preferencialmente por meio do ângulo *roll* e pode ou não estar acompanhada por movimentos translacionais pelo *slide* (DJURANOVIC; HARTMANN, 2003; OLSON et al., 1998; TOLSTORUKOV et al., 2007). Os parâmetros *roll* e *slide* apresentam uma correlação negativa, onde valores positivos para *roll* e negativos para *slide* tendem a provocar uma inclinação nos dois nucleotídeos empilhados em direção ao sulco maior, ou seja, compreendendo o sulco maior, enquanto que valores negativos para *roll* e positivos para *slide* causam uma inclinação da dupla hélice de DNA em direção ao sulco menor (TOLSTORUKOV et al., 2007). Porém, tal combinação não é observada entre os parâmetros *roll* e *slide* para os *steps* do HRE de nossos complexos moleculares. Em especial, um padrão de ângulos *roll* positivos e negativos lado a lado nos dinucleotídeos, como é observado na Figura 39 para a maioria dos *steps* do HRE, resulta na produção de uma conformação local linear na dupla-hélice (DICKERSON; CHIU, 1997), inferindo que o HIF1 ao se ligar no HRE não causa uma curvatura/dobramento na dupla hélice como observado em estudos prévios (ELLENBERGER et al., 1994; FERRÉ-D'AMARÉ et al., 1994; SCHULTE et al., 2017; SHIMIZU, 1997; WU et al., 2015, 2016).

Os ângulos rotacionais *roll* e *twist* apresentam maior flexibilidade em comparação com o *tilt*, devido a isso contribuem para a intensidade da curvatura do ácido nucleico (OLSON et al., 1998; SARAI et al., 1989). Dentre as três classes químicas observadas para o HRE, em geral os YR *steps* (segundo e quarto *bp*-step) apresentaram valores mais altos para o *roll* e *twist* em comparação com os demais corroborando dados da literatura (DICKERSON; CHIU, 1997; GU et al., 2015), com exceção para o dCG do elemento regulatório contendo o alelo G para o *twist*. De fato, as extremidades 5' e 3' do HRE com o alelo -964A são mais flexíveis do que o meio da sequência regulatória, enquanto que para o HRE com o alelo -964G, maior flexibilidade é observada para a extremidade 3' dada pelo quarto *step* em relação ao ângulo *twist*. Essa diferença na flexibilidade dos complexos para o YR *steps* para o *twist* dos complexos analisados pode ser devido a presença do polimorfismo -964G>A da região promotora do gene *HLA-G* no interior do HRE, visto que o ângulo *twist* é mais fortemente afetado pelas bases subjacentes do que os outros parâmetros conformacionais (GU et al., 2015) promovendo um mecanismo de leitura diferente para a interação HIF1-HRE em função do polimorfismo presente.

O ângulo rotacional *tilt* não contribui substancialmente para a curvatura do DNA, e deformações na dupla-hélice resultantes deste ângulo é mais energeticamente custoso em relação aos demais (CURUKSU et al., 2009; OLSON et al., 1998). Valores positivos de *tilt* provocam a abertura dos pares de bases empilhados na fita, onde valores altos causam choques estéricos entre estes pares de bases resultando em penalidade para os termos de energia da

estrutura (LAVERY; SKLENAR, 1989; SARAI et al., 1989; ZHURKIN et al., 1982). Assim, as deformações causadas nos *bp-steps* pelo *tilt* gera conformações locais que podem afetar a interação com outras moléculas, como os fatores de transcrição. Novamente, os YR *steps* se demonstraram mais flexíveis para o *tilt* em relação aos demais *steps* do HRE, como observado na Figura 39, contudo o segundo *step* dCA do HRE com o alelo A apresentou valores mais altos em comparação com o dCG do elemento regulatório com o alelo G, dessa forma, a presença do alelo A no HRE gera uma conformação na dupla hélice de DNA menos energeticamente favorável. Para suprir esta penalidade energética causada pelo ângulo *tilt*, é necessário um aumento do parâmetro *rise* nos *bp-steps* em questão (SARAI et al., 1989), porém, nenhuma diferença é observada para *rise* no dCA do complexo HIF1-HRE-964A e o custo energético causado pelo ângulo *tilt* não é suprido neste elemento regulatório. Esta deformação na dupla hélice de DNA produzida pelo *tilt* no HRE-964A pode ter influenciado na sua interação com o HIF1, diminuindo a estabilidade do complexo molecular formado como foi observado anteriormente para os resultados da dinâmica molecular para o complexo HIF1-HRE-964A.

De fato, o presente estudo demonstrou que o HIF1 quando ligado ao HRE contendo o alelo G no sítio polimórfico -964G>A da região promotora do gene HLA-G, resulta na formação de um complexo mais estável e energeticamente favorável em relação ao complexo HIF1-HRE com o alelo A neste sítio polimórfico, de acordo com os resultados obtidos por DM. Porém, a ausência de componentes importantes nos nossos modelos teóricos, como o elemento regulatório responsivo a hipóxia localizado no éxon 2 do HLA-G e cofatores importantes para a transativação gênica pelo HIF1, não permite a total compreensão do papel do HIF1 na regulação da transcrição do gene HLA-G sob condições de hipóxia à nível atômico. Assim, neste contexto, outros estudos serão necessários para elucidar como ocorre a regulação transcricional do HLA-G pelo HIF1. Apesar disso, este é o primeiro estudo realizado para o gene HLA-G que visa analisar estruturalmente por meio de abordagem computacional, a interação de fatores de transcrição e o DNA em função de polimorfismos gênicos presentes na região promotora do gene, o que torna este estudo uma referência para futuros trabalhos acerca do gene HLA-G e o impacto da sua variabilidade populacional na interação com proteínas.

6 Conclusão

Em conclusão, o HIF1 interage de maneira mais estável e favorável com o HRE contendo o alelo G no sítio polimórfico -964 G>A da região promotora do gene *HLA-G* do que o HRE com o alelo A. Porém, a limitação para entender o papel do HIF1 na regulação da transcrição do gene *HLA-G* por meio de técnicas de modelagem molecular computacional é a presença de outros elementos regulatórios responsivos a hipóxia HRE (HRE no éxon 2), além daquele na região promotora, exigindo análises adicionais. Não obstante, foi identificado comportamento dinâmico diferencial entre os alelos -964G/A na interação do HRE com o fator de transcrição HIF1 bem como diferenças na geometria da dupla hélice de DNA que podem auxiliar na compreensão da produção diferencial da molécula *HLA-G* em condições de hipóxia.

7 Referências

ALBERTS, B. et al. **Biologia Molecular da Célula**. 6. ed. Porto Alegre: Artmed, 2017. 1464 p.

AMOUTZIAS, G. D.; ROBERTSON, D. L.; VAN DE PEER, Y.; OLIVER, S. G. Choose your partners: dimerization in eukaryotic transcription factors. **Trends in Biochemical Sciences**, [S. l.], v. 33, n. 5, p. 220–229, 2008.

ANDERSON, W. F.; OHLENDORF, D. H.; TAKEDA, Y.; MATTHEWS, B. W. Structure of the cro repressor from bacteriophage λ and its interaction with DNA. **Nature**, [S. l.], v. 290, n. 5809, p. 754–758, 1981.

ANFINSEN, C. B. Principles that Govern the Folding of Protein Chains. **Science**, [S. l.], v. 181, n. 4096, p. 223–230, 1973.

APPELHOFF, R. J.; TIAN, Y.; RAVAL, R. R.; TURLEY, H.; HARRIS, A. L.; PUGH, C. W.; RATCLIFFE, P. J.; GLEADLE, J. M. Differential Function of the Prolyl Hydroxylases PHD1, PHD2, and PHD3 in the Regulation of Hypoxia-inducible Factor. **Journal of Biological Chemistry**, [S. l.], v. 279, n. 37, p. 38458–38465, 2004

ARAGONÉS, J.; FRAISL, P.; BAES, M.; CARMELIET, P. Oxygen Sensors at the Crossroad of Metabolism. **Cell Metabolism**, [S. l.], v. 9, n. 1, p. 11–22, 2009.

BACSI, S.G.; HANKINSON, O. Functional Characterization of DNA-binding Domains of the Subunits of the Heterodimeric Aryl Hydrocarbon Receptor Complex Imputing Novel and Canonical Basic Helix-Loop-Helix Protein-DNA Interactions. **Journal of Biological Chemistry**, [S. l.], v. 271, n. 15, p. 8843–8850, 1996.

BAE, S.; JEONG, J.; PARK, J. A.; KIM, S.; BAE, M.; CHOI, S.; KIM, K. Sumoylation increases HIF-1 α stability and its transcriptional activity. **Biochemical and Biophysical Research Communications**, [S. l.], v. 324, n. 1, p. 394–400, 2004.

BAEK, S.; KIM, K. II. Regulation of HIF-1 α stability by lysine methylation. **BMB Reports**, [S. l.], v. 49, n. 5, p. 245–246, 2016.

BENKERT, P.; BIASINI, M; SCHWEDE, Torsten. Toward the estimation of the absolute quality of individual protein structure models. [S. l.], v. 27, n. 3, p. 343–350, 2011.

BENKERT, P; KÜNZLI, Ml; SCHWEDE, T. QMEAN server for protein model quality estimation. **Nucleic Acids Research**, [S. l.], v.37, p. W510-4, 2009.

BERENDSEN, H. J. C.; GRIGERA, J. R.; STRAATSMA, T. P. The missing term in effective pair potentials. **The Journal of Physical Chemistry**, [S. l.], v. 91, n. 24, p. 6269–6271, 1987.

BERENDSEN, H. J. C.; POSTMA, J. P. M.; VAN GUNSTEREN, W. F.; DINOLA, A.; HAAK, J. R. Molecular dynamics with coupling to an external bath. **The Journal of Chemical Physics**, [S. l.], v. 81, n. 8, p. 3684–3690, 1984.

BERENDSEN, H. J. C.; POSTMA, J. P. M.; VAN GUNSTEREN, W. F.; HERMANS, J. Interaction Models for Water in Relation to Protein Hydration. In: PULLMAN, B. (org.).

Intermolecular Forces. 14. ed. Dordrecht: Springer, 1981. p. 331–342.

BERRA, E. HIF prolyl-hydroxylase 2 is the key oxygen sensor setting low steady-state levels of HIF-1 in normoxia. **The EMBO Journal**, [S. l.], v. 22, n. 16, p. 4082–4090, 2003.

BLASCHITZ, A. et al. Vascular Endothelial Expression of Indoleamine 2,3-Dioxygenase 1 Forms a Positive Gradient towards the Feto-Maternal Interface. **PLoS ONE**, [S. l.], v. 6, n. 7, p. e21774, 2011.

BRAY, S. **Running molecular dynamics simulations using GROMACS**. 2021. Disponível em: <https://training.galaxyproject.org/training-material/topics/computational-chemistry/tutorials/md-simulation-gromacs/tutorial.html>. Acesso em: 1 maio. 2021.

BROOKS, C. L. Computer simulation of liquids. **Journal of Solution Chemistry**, [S. l.], v. 18, n. 1, p. 99–99, 1989.

BUCHAN, D.W. A.; JONES, David T. The PSIPRED Protein Analysis Workbench: 20 years on. **Nucleic Acids Research**, [S. l.], v. 47, n. W1, p. W402–W407, 2019.

BURLEY, S. K. The TATA box binding protein. **Current Opinion in Structural Biology**, [S. l.], v. 6, n. 1, p. 69–75, 1996.

CARBIA-NAGASHIMA, A.; GEREZ, J.; PEREZ-CASTRO, C.; PAEZ-PEREDA, M.; SILBERSTEIN, S.; STALLA, G. K.; HOLSBOER, F.; ARZT, E. RSUME, a Small RWD-Containing Protein, Enhances SUMO Conjugation and Stabilizes HIF-1 α during Hypoxia. **Cell**, [S. l.], v. 131, n. 2, p. 309–323, 2007.

CARRERA, S.; SENRA, J.; ACOSTA, M. I.; ALTHUBITI, M.; HAMMOND, E. M.; DE VERDIER, P. J.; MACIP, S. The Role of the HIF-1 α Transcription Factor in Increased Cell Division at Physiological Oxygen Tensions. **PLoS ONE**, [S. l.], v. 9, n. 5, p. e97938, 2014.

CARDOSO, Rosa et al. Identification of Cys255 in HIF-1 α as a novel site for development of covalent inhibitors of HIF-1 α /ARNT PasB domain protein-protein interaction. **Protein Science**, [S. l.], v. 21, n. 12, p. 1885–1896, 2012.

CASTELLI, E. C. et al. HLA-G variability and haplotypes detected by massively parallel sequencing procedures in the geographically distinct population samples of Brazil and Cyprus. **Molecular Immunology**, [S. l.], v. 83, p. 115–126, 2017.

CASTELLI, E. C.; RAMALHO, J.; PORTO, I. O. P.; LIMA, T. H. A.; FELÍCIO, L. P.; SABBAGH, A.; DONADI, E. A.; MENDES-JUNIOR, C. T. Insights into HLA-G Genetics Provided by Worldwide Haplotype Diversity. **Frontiers in Immunology**, [S. l.], v. 5, 2014. a.

CASTELLI, E. C.; VEIGA-CASTELLI, L. C.; YAGHI, L.; MOREAU, P.; DONADI, E. A. Transcriptional and Posttranscriptional Regulations of the HLA-G Gene. **Journal of Immunology Research**, [S. l.], v. 2014, p. 1–15, 2014. b.

CHANDRASEKARAN, R.; ARNOTT, S. The Structure of B-DNA in Oriented Fibers. **Journal of Biomolecular Structure and Dynamics**, [S. l.], v. 13, n. 6, p. 1015–1027, 1996.

CHANG, Ch.; FERRONE, S. HLA-G in melanoma: can the current controversies be solved? **Seminars in Cancer Biology**, [S. l.], v. 13, n. 5, p. 361–369, 2003.

CHEN, V. B.; ARENDALL, W. B.; HEADD, J. J.; KEEDY, D. A.; IMMORMINO, R. M.; KAPRAL, G. J.; MURRAY, L. W.; RICHARDSON, J. S.; RICHARDSON, D. C. MolProbity: all-atom structure validation for macromolecular crystallography. **Acta Crystallographica Section D Biological Crystallography**, [S. l.], v. 66, n. 1, p. 12–21, 2010.

CHENG, J.; KANG, X.; ZHANG, S.; YEH, E. T. H. SUMO-Specific Protease 1 Is Essential for Stabilization of HIF1 α during Hypoxia. **Cell**, [S. l.], v. 131, n. 3, p. 584–595, 2007.

CHIU, T. P.; RAO, S.; MANN, R. S.; HONIG, B.; ROHS, R. Genome-wide prediction of minor-groove electrostatic potential enables biophysical modeling of protein-DNA binding. **Nucleic Acids Research**, [S. l.], v. 45, n. 21, p. 12565–12576, 2017.

CHOU, K. Structural Bioinformatics and its Impact to Biomedical Science and Drug Discovery. *In: Frontiers in Medicinal Chemistry - (Volume 3)*. [s.l.] : BENTHAM SCIENCE PUBLISHERS, 2012. p. 455–502.

COCK, P. J. A. et al. Biopython: freely available Python tools for computational molecular biology and bioinformatics. **Bioinformatics**, [S. l.], v. 25, n. 11, p. 1422–1423, 2009.

COLOVOS, C.; YEATES, T. O. Verification of protein structures: Patterns of nonbonded atomic interactions. **Protein Science**, [S. l.], v. 2, n. 9, p. 1511–1519, 1993.

COUTSIAS, E. A.; SEOK, C.; JACOBSON, M. P.; DILL, K. A. A Kinematic View of Loop Closure. **Journal of Computational Chemistry**, [S. l.], v. 25, n. 4, p. 510–528, 2004.

CURIGLIANO, G.; CRISCITIELLO, C.; GELAO, L.; GOLDBIRSCHE, A. Molecular pathways: human leukocyte antigen G (HLA-G). **Clinical cancer research: an official journal of the American Association for Cancer Research**, [S. l.], v. 19, n. 20, p. 5564–71, 2013.

CURUKSU, J.; ZACHARIAS, M.; LAVERY, R.; ZAKRZEWSKA, K. Local and global effects of strong DNA bending induced during molecular dynamics simulations. **Nucleic Acids Research**, [S. l.], v. 37, n. 11, p. 3766–3773, 2009.

DARDEN, T.; YORK, D.; PEDERSEN, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. **The Journal of Chemical Physics**, [S. l.], v. 98, n. 12, p. 10089–10092, 1993.

DE VRIES, S. J.; VAN DIJK, M.; BONVIN, A. M. J. J. The HADDOCK web server for data-driven biomolecular docking. **Nature Protocols**, [S. l.], v. 5, n. 5, p. 883–897, 2010.

DENGLER, V. L.; GALBRAITH, M. D. ; ESPINOSA, J. M. Transcriptional regulation by hypoxia inducible factors. **Critical Reviews in Biochemistry and Molecular Biology**, [S. l.], v. 49, n. 1, p. 1–5, 2014.

DICKERSON, R. E. Definitions and nomenclature of nucleic acid structure components.

Nucleic Acids Research, [S. l.], v. 17, n. 5, p. 1797–1803, 1989.

DICKERSON, R. E.; CHIU, T. K. Helix Bending as a Factor in Protein/DNA Recognition. **Biopolymers - Biospectroscopy Section**, [S. l.], v. 44, n. 4, p. 361–403, 1997.

DJURANOVIC, D.; HARTMANN, B. Conformational characteristics and correlations in crystal structures of nucleic acid oligonucleotides: Evidence for sub-states. **Journal of Biomolecular Structure and Dynamics**, [S. l.], v. 20, n. 6, p. 771–788, 2003.

DOMINGUEZ, C.; BOELEN, R.; BONVIN, A. M. J. J. HADDOCK: A Protein–Protein Docking Approach Based on Biochemical or Biophysical Information. **Journal of the American Chemical Society**, [S. l.], v. 125, n. 7, p. 1731–1737, 2003.

DONADI, E. A.; CASTELLI, E. C.; ARNAIZ-VILLENA, A.; ROGER, M.I; REY, D.; MOREAU, P. Implications of the polymorphism of HLA-G on its function, regulation, evolution and disease association. **Cellular and Molecular Life Sciences**, [S. l.], v. 68, n. 3, p. 369–395, 2011.

DUNWOODIE, S. L. The Role of Hypoxia in Development of the Mammalian Embryo. **Developmental Cell**, [S. l.], v. 17, n. 6, p. 755–773, 2009.

DYSON, H. J.; WRIGHT, P. E. Intrinsically unstructured proteins and their functions. **Nature Reviews Molecular Cell Biology**, [S. l.], v. 6, n. 3, p. 197–208, 2005.

EISENBERG, D.; SCHWARZ, E.; KOMAROMY, M.; WALL, R. Analysis of membrane and surface protein sequences with the hydrophobic moment plot. **Journal of Molecular Biology**, [S. l.], v. 179, n. 1, p. 125–142, 1984.

ELLENBERGER, T.; FASS, D.; ARNAUD, M.; HARRISON, S. C. Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. **Genes & Development**, [S. l.], v. 8, n. 8, p. 970–980, 1994.

FERNANDEZ-FUENTES, N.; FISER, A. Modeling Loops in Protein Structures. *In: Introduction to Protein Structure Prediction*. Hoboken, NJ, USA: John Wiley & Sons, Inc., 2010. p. 279–298.

FERNÁNDEZ-RECIO, J.; TOTROV, M.; ABAGYAN, R. Identification of Protein–Protein Interaction Sites from Docking Energy Landscapes. **Journal of Molecular Biology**, [S. l.], v. 335, n. 3, p. 843–865, 2004.

FERRÉ-D'AMARÉ, A. R.; POGNONEC, P.; ROEDER, R. G.; BURLEY, S. K. Structure and function of the b/HLH/Z domain of USF. **The EMBO Journal**, [S. l.], v. 13, n. 1, p. 180–189, 1994.

FISER, A.; SALI, A. ModLoop: automated modeling of loops in protein structures. **Bioinformatics**, [S. l.], v. 19, n. 18, p. 2500–2501, 2003.

FOGOLARI, F.; BRIGO, A.; MOLINARI, H. The Poisson-Boltzmann equation for biomolecular electrostatics: a tool for structural biology. **Journal of Molecular Recognition**,

[*S. l.*], v. 15, n. 6, p. 377–392, 2002.

FORDEL, E.; GEUENS, E.; DEWILDE, S.; DE COEN, W.; MOENS, L. Hypoxia/ischemia and the regulation of neuroglobin and cytoglobin expression. **IUBMB Life**, [*S. l.*], v.56, p.681–687, 2004.

FORSYTHE, J. A.; JIANG, B. H.; IYER, N. V; AGANI, F.; LEUNG, S. W.; KOOS, R. D.; SEMENZA, G. L. Activation of vascular endothelial growth factor gene transcription by hypoxia-inducible factor 1. **Molecular and Cellular Biology**, [*S. l.*], v. 16, n. 9, p. 4604–4613, 1996.

FRANK-KAMENETSKII, M. D. DNA topology. **Journal of Molecular Structure: THEOCHEM**, [*S. l.*], v. 336, n. 2–3, p. 235–243, 1995.

FRENKEL, D.; SMIT, B. **Understanding Molecular Simulation**. 1. ed. [s.l.] : Elsevier, 2002.

GALINDO-MURILLO, R.; BERGONZO, C.; CHEATHAM, T. E. Molecular Modeling of Nucleic Acid Structure: Setup and Analysis. **Current Protocols in Nucleic Acid Chemistry**, [*S. l.*], v. 56, n. 1, 2014.

GARZIERA, M., SCARABEL, L., TOFFOLI, G. Hypoxic Modulation of HLA-G expression through the Metabolic Sensor HIF-1 in Human Cancer Cells. **Journal of Immunology Research**, [*S. l.*], v.2017, p.4587520, 2017.

GILKES, D. M.; SEMENZA, G. L. Role of hypoxia-inducible factors in breast cancer metastasis. **Future Oncology**, [*S. l.*], v. 9, n. 11, p. 1623–1636, 2013.

GRAHAM, A. M.; PRESNELL, J. S. Hypoxia Inducible Factor (HIF) transcription factor family expansion, diversification, divergence and selection in eukaryotes. **PLoS ONE**, [*S. l.*], 2017.

GREGORI, S.; MAGNANI, C. F.; RONCAROLO, M. G. Role of human leukocyte antigen-G in the induction of adaptive type 1 regulatory T cells. **Human Immunology**, [*S. l.*], 2009.

GU, C. et al. DNA Structural Correlation in Short and Long Ranges. **Journal of Physical Chemistry B**, [*S. l.*], v. 119, n. 44, p. 13980–13990, 2015.

GUPTA, A.; MOHANTY, P.; BHATNAGAR, S. Protein Structure Prediction Using Homology Modeling. **Pharmaceutical Sciences**, [*S. l.*], n. January, p. 877–897, 2016.

GUSTAFSSON, M. V. et al. Hypoxia Requires Notch Signaling to Maintain the Undifferentiated Cell State. **Developmental Cell**, [*S. l.*], v. 9, n. 5, p. 617–628, 2005.

HAASE, Volker H. Regulation of erythropoiesis by hypoxia-inducible factors. **Blood Reviews**, [*S. l.*], v. 27, n. 1, p. 41–53, 2013.

HAHN, S. Structure and mechanism of the RNA polymerase II transcription machinery. **Nature Structural & Molecular Biology**, [*S. l.*], v. 11, n. 5, p. 394–403, 2004.

HALPERIN, I.I.; MA, B.; WOLFSON, H.; NUSSINOV, R.. Principles of docking: An overview of search algorithms and a guide to scoring functions. **Proteins: Structure, Function and Genetics**, [*S. l.*], v. 47, n. 4, p. 409–443, 2002.

HAMMER, A.; HUTTER, H.; DOHR, G.. HLA class I expression on the materno-fetal interface. **American Journal of Reproductive Immunology**, [*S. l.*], v. 38, n. 3, p. 150–157, 1997.

HARVEY, S. C.; TAN, R. t K. Z.; CHEATHAM, T. E. The flying ice cube: Velocity rescaling in molecular dynamics leads to violation of energy equipartition. **Journal of Computational Chemistry**, [*S. l.*], v. 19, n. 7, p. 726–740, 1998.

HEIKKILÄ, M.; PASANEN, A.; KIVIRIKKO, K. I.; MYLLYHARJU, J. Roles of the human hypoxia-inducible factor (HIF)-3 α variants in the hypoxia response. **Cellular and Molecular Life Sciences**, [*S. l.*], v. 68, n. 23, p. 3885–3901, 2011.

HESS, B.; BEKKER, H.; BERENDSEN, H.J. C.; FRAAIJE, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, [*S. l.*], v. 18, n. 12, p. 1463–1472, 1997.

HOCKNEY, R. W. Potential calculation and some applications. **Methos Comput. Phys**, [*S. l.*], v. 9, p. 135–211, 1970.

HOLLINGSWORTH, Scott A.; DROR, Ron O. Molecular Dynamics Simulation for All. **Neuron**, [*S. l.*], v. 99, n. 6, p. 1129–1143, 2018.

HONORATO, R. V.; ROEL-TOURIS, J.; BONVIN, A. M. J. J. MARTINI-Based Protein-DNA Coarse-Grained HADDOCKing. **Frontiers in Molecular Biosciences**, [*S. l.*], v. 6, 2019.

HORNAK, V.; ABEL, R.; OKUR, A.; STROCKBINE, B.; ROITBERG, A.; SIMMERLING, C. Comparison of multiple Amber force fields and development of improved protein backbone parameters. **Proteins: Structure, Function, and Bioinformatics**, [*S. l.*], v. 65, n. 3, p. 712–725, 2006.

HORUZSKO, A.; LENFANT, F.; MUNN, D. H.; MELLOR, A. L. Maturation of antigen-presenting cells is compromised in HLA-G transgenic mice. **International Immunology**, [*S. l.*], v. 13, n. 3, p. 385–394, 2001.

HUANG, L. E.; GU, J.; SCHAU, M.; BUNN, H. F. Regulation of hypoxia-inducible factor 1 is mediated by an O₂-dependent degradation domain via the ubiquitin-proteasome pathway. **Proceedings of the National Academy of Sciences**, [*S. l.*], v. 95, n. 14, p. 7987–7992, 1998.

HUANG, Y. et al. Hypoxia Inducible Factor 3 α Plays a Critical Role in Alveolarization and Distal Epithelial Cell Differentiation during Mouse Lung Development. **PLoS ONE**, [*S. l.*], v. 8, n. 2, p. e57695, 2013.

ILIOPOULOS, O.; LEVY, A. P.; JIANG, C.; KAELEN, W. G.; GOLDBERG, M. A. Negative regulation of hypoxia-inducible genes by the von Hippel-Lindau protein. **Proceedings of the National Academy of Sciences**, [*S. l.*], v. 93, n. 20, p. 10595–10599, 1996.

IVANI, I. et al. Parmbsc1: a refined force field for DNA simulations. **Nature Methods**, [S. l.], v. 13, n. 1, p. 55–58, 2016.

IZUMI, K.. Disorders of Transcriptional Regulation: An Emerging Category of Multiple Malformation Syndromes. **Molecular Syndromology**, [S. l.], v. 7, n. 5, p. 262–273, 2016.

JIANG, B.; RUE, E.; WANG, G. L.; ROE, R.; SEMENZA, G. L. Dimerization, DNA Binding, and Transactivation Properties of Hypoxia-inducible Factor 1. **Journal of Biological Chemistry**, [S. l.], v. 271, n. 30, p. 17771–17778, 1996.

JIANG, B.; ZHENG, J. Z.; LEUNG, S. W.; ROE, R.; SEMENZA, G. L. Transactivation and Inhibitory Domains of Hypoxia-inducible Factor 1 α . **Journal of Biological Chemistry**, [S. l.], v. 272, n. 31, p. 19253–19260, 1997.

JO, S.; KIM, T.; IYER, V. G.; IM, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. **Journal of Computational Chemistry**, [S. l.], v. 29, n. 11, p. 1859–1865, 2008.

JONES, D. T. Protein secondary structure prediction based on position-specific scoring matrices 1 Edited by G. Von Heijne. **Journal of Molecular Biology**, [S. l.], v. 292, n. 2, p. 195–202, 1999.

JORGENSEN, W. L.; MADURA, J. D. Quantum and statistical mechanical studies of liquids. 25. Solvation and conformation of methanol in water. **Journal of the American Chemical Society**, [S. l.], v. 105, n. 6, p. 1407–1413, 1983.

JORGENSEN; W. L.; TIRADO-RIVES, J. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. **Journal of the American Chemical Society**. v. 110, n. 6, p. 1657–1666, 1988

KALOUSHI, A.I.; MYLONIS, I.; POLITOU, A.S.; CHACHAMI, G.; PARASKEVA, E.; SIMOS, G. Casein kinase 1 regulates human hypoxia-inducible factor HIF-1. **Journal of Cell Science**, [S. l.], v. 123, n. 17, p. 2976–2986, 2010.

KIETZMANN, T; MENNERICH, D; DIMOVA, E. Y. Hypoxia-Inducible Factors (HIFs) and Phosphorylation: Impact on Stability, Localization, and Transactivity. **Frontiers in Cell and Developmental Biology**, [S. l.], v. 4, 2016.

KILBURN, B. A.; WANG, J.; DUNIEC-DMUCHKOWSKI, Z. M.; LEACH, R. E.; ROMERO, R.; ARMANT, D. R. Extracellular Matrix Composition and Hypoxia Regulate the Expression of HLA-G and Integrins in a Human Trophoblast Cell Line1. **Biology of Reproduction**, [S. l.], v. 62, n. 3, p. 739–747, 2000.

KITCHEN, D. B.; DECORNEZ, H.; FURR, J.R.; BAJORATH, J. Docking and scoring in virtual screening for drug discovery: methods and applications. **Nature Reviews Drug Discovery**, [S. l.], v. 3, n. 11, p. 935–949, 2004.

KLEIN, J.; SATO, A. The HLA system: first of two parts. **The New England Journal of Medicine Review**, [S. l.], v. 343, n. 10, p. 702–709, 2000. a.

KLEIN, J.; SATO, A. The HLA system: Second of Two Parts. **The New England Journal of Medicine**, [S. l.], v. 343, n. 11, p. 782–786, 2000. b.

KLEMM, S. L.; SHIPONY, Zohar; GREENLEAF, William J. Chromatin accessibility and the regulatory epigenome. **Nature Reviews Genetics**, [S. l.], v. 20, n. 4, p. 207–220, 2019.

KOIVUNEN, P.i; HIRSILÄ, M.; GÜNZLER, V.; KIVIRIKKO, K. I.; MYLLYHARJU, J. Catalytic Properties of the Asparaginyl Hydroxylase (FIH) in the Oxygen Sensing Pathway Are Distinct from Those of Its Prolyl 4-Hydroxylases. **Journal of Biological Chemistry**, [S. l.], v. 279, n. 11, p. 9899–9904, 2004.

KOLONKO, M.; GREB-MARKIEWICZ, B.. BHLH–PAS proteins: Their structure and intrinsic disorder. **International Journal of Molecular Sciences**, [S. l.], v. 20, n. 15, 2019.

KOLOVOS, P.; KNOCH, T. A.; GROSVELD, F. G.; COOK, P. R.; PAPANTONIS, A. Enhancers and silencers: an integrated and simple model for their function. **Epigenetics & Chromatin**, [S. l.], v. 5, n. 1, p. 1, 2012.

KOYASU, S.; KOBAYASHI, M.; GOTO, Y.; HIRAOKA, M.; HARADA, H. Regulatory mechanisms of hypoxia-inducible factor 1 activity: Two decades of knowledge. **Cancer Science**, [S. l.], v. 109, n. 3, p. 560–571, 2018.

KUHLMAN, B.; BRADLEY, P. Advances in protein structure prediction and design. **Nature Reviews Molecular Cell Biology**, [S. l.], v. 20, n. 11, p. 681–697, 2019.

LANCE, Br. K.; DEANE, C. M.; WOOD, G. R. Exploring the potential of template-based modelling. **Bioinformatics**, [S. l.], v. 26, n. 15, p. 1849–1856, 2010.

LANDO, D. FIH-1 is an asparaginyl hydroxylase enzyme that regulates the transcriptional activity of hypoxia-inducible factor. **Genes & Development**, [S. l.], v. 16, n. 12, p. 1466–1471, 2002.

LASKOWSKI, R. A.; MACARTHUR, M. W.; MOSS, D. S.; THORNTON, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. **Journal of Applied Crystallography**, [S. l.], v. 26, n. 2, p. 283–291, 1993.

LAVERY, R.; MOAKHER, M.; MADDOCKS, J. H.; PETKEVICIUTE, D.; ZAKRZEWSKA, K. Conformational analysis of nucleic acids revisited: Curves+. **Nucleic Acids Research**, [S. l.], v. 37, n. 17, p. 5917–5929, 2009.

LAVERY, R.; SKLENAR, H. Defining the structure of irregular nucleic acids: Conventions and principles. **Journal of Biomolecular Structure and Dynamics**, [S. l.], v. 6, n. 4, p. 655–667, 1989.

LE GAL, F.; RITEAU, B.; SEDLIK, C.; KHALIL-DAHER, I.; MENIER, C.; DAUSSET, J.; GUILLET, J.; CAROSELLA, E. D.; ROUAS-FREISS, N. HLA-G-mediated inhibition of

antigen-specific cytotoxic T lymphocytes. **International Immunology**, [S. l.], v. 11, n. 8, p. 1351–1356, 1999.

LEACH, A. R. **Molecular Modelling: Principles and Applications**. 2. ed. New York, NY: Prentice Hall, 2001.

LEE, J.; BAE, S.; JEONG, J.; KIM, S.; KIM, K. Hypoxia-inducible factor (HIF-1) α : its protein stability and biological functions. **Experimental & Molecular Medicine**, [S. l.], v. 36, n. 1, p. 1–12, 2004.

LEE, J.; LEE, D.; PARK, H.; COUTSIAS, E. A.; SEOK, C. Protein loop modeling by using fragment assembly and analytical loop closure. **Proteins: Structure, Function, and Bioinformatics**, [S. l.], v. 78, n. 16, p. 3428–3436, 2010.

LEMKUL, J. From Proteins to Perturbed Hamiltonians: A Suite of Tutorials for the GROMACS-2018 Molecular Simulation Package [Article v1.0]. **Living Journal of Computational Molecular Science**, [S. l.], v. 1, n. 1, p. 1–53, 2019.

LI, Y.; PATEL, S. P.; ROSZIK, J.; QIN, Y. Hypoxia-Driven Immunosuppressive Metabolites in the Tumor Microenvironment: New Approaches for Combinational Immunotherapy. **Frontiers in Immunology**, [S. l.], v. 9, 2018.

LIM, J.; LEE, Y.; CHUN, Y.; CHEN, J.; KIM, J.; PARK, J. Sirtuin 1 Modulates Cellular Responses to Hypoxia by Deacetylating Hypoxia-Inducible Factor 1 α . **Molecular Cell**, [S. l.], v. 38, n. 6, p. 864–878, 2010.

LIN, F. Y.; MACKERELL, A. D. Force Fields for Small Molecules. *In*: BONOMI, M.; CAMILLONI, C. (org.). **Methods in Molecular Biology**. 2022. ed. New York, NY: Humana, 2019. v. 2022p. 21–54.

LOBANOV, M. Yu; BOGATYREVA, N. S.; GALZITSKAYA, O. V. Radius of gyration as an indicator of protein structure compactness. **Molecular Biology**, [S. l.], v. 42, n. 4, p. 623–628, 2008.

LU, X.; OLSON, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. **Nature Protocols**, [S. l.], v. 3, n. 7, p. 1213–1227, 2008.

LU, X.; OLSON, W. K. Resolving the discrepancies among nucleic acid conformational analyses. **Journal of Molecular Biology**, [S. l.], v. 285, n. 4, p. 1563–1575, 1999.

LUSCOMBE, N. M.; AUSTIN, S. E.; BERMAN, H. M.; THORNTON, J. M. An overview of the structures of protein-DNA complexes. **Genome biology**, [S. l.], v. 1, n. 1, p. 1–37, 2000.

LUSCOMBE, N. M. Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. **Nucleic Acids Research**, [S. l.], v. 29, n. 13, p. 2860–2874, 2001.

LUSCOMBE, N. M.; THORNTON, J. M. Protein – DNA Interactions: Amino Acid

Conservation and the Effects of Mutations on Binding Specificity. *[S. l.]*, v. 2836, n. 02, p. 991–1009, 2002.

LÜTHY, R.; BOWIE, J. U.; EISENBERG, D. Assessment of protein models with three-dimensional profiles. **Nature**, *[S. l.]*, v. 356, n. 6364, p. 83–85, 1992.

LYUBARTSEV, Alexander P.; RABINOVICH, Alexander L. Force Field Development for Lipid Membrane Simulations. **Biochimica et Biophysica Acta - Biomembranes**, *[S. l.]*, v. 1858, n. 10, p. 2483–2497, 2016.

MAHON, P. C. FIH-1: a novel protein that interacts with HIF-1alpha and VHL to mediate repression of HIF-1 transcriptional activity. **Genes & Development**, *[S. l.]*, v. 15, n. 20, p. 2675–2686, 2001.

MAKINO, Y.; KANOPKA, A.; WILSON, W. J.; TANAKA, H.; POELLINGER, L. Inhibitory PAS Domain Protein (IPAS) Is a Hypoxia-inducible Splicing Variant of the Hypoxia-inducible Factor-3 α Locus. **Journal of Biological Chemistry**, *[S. l.]*, v. 277, n. 36, p. 32405–32408, 2002.

MACKERELL, A.D. et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. **Journal of Physical Chemistry B**, *[S. l.]*, v. 102, p. 3586–3616, 1998.

MANDEL-GUTFREUND, Y.; SCHUELER, O.; MARGALIT, H. Comprehensive Analysis of Hydrogen Bonds in Regulatory Protein DNA-Complexes: In Search of Common Principles. **Journal of Molecular Biology**, *[S. l.]*, v. 253, n. 2, p. 370–382, 1995.

MATTHEWS, B. W. Protein-DNA interaction. No code for recognition. **Nature**, *[S. l.]*, v. 335, n. 6188, p. 294–295, 1988.

MAYNARD, M. A.; EVANS, A. J.; HOSOMI, T.; HARA, S.; JEWETT, M. A. S.; OHH, M. Human HIF-3 α 4 is a dominant-negative regulator of HIF-1 and is down-regulated in renal cell carcinoma. **The FASEB Journal**, *[S. l.]*, v. 19, n. 11, p. 1396–1406, 2005.

MCDONALD, I. K.; THORNTON, J. M. Satisfying hydrogen bonding potential in proteins. **Journal of Molecular Biology**, *[S. l.]*, 1994. DOI: 10.1006/jmbi.1994.1334.

MCKAY, David B.; STEITZ, Thomas A. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. **Nature**, *[S. l.]*, v. 290, n. 5809, p. 744–749, 1981.

MENG, X. Y., ZHANG, H. X., MEZEI, M., & CUI, M. Molecular docking: a powerful approach for structure-based drug discovery. Current computer-aided drug design. **Current Computer Aided Drug Design**, *[S. l.]*, v. 7, n. 2, p. 146–157, 2011.

MENG, X.; ZHANG, H.; MEZEI, M.; CUI, M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. **Current Computer Aided-Drug Design**, *[S. l.]*, v. 7, n. 2, p. 146–157, 2011.

MICHAEL GROMIHA, M.; SIEBERS, J. G.; SELVARAJ, S.; KONO, Hi.; SARAI, A. Intermolecular and intramolecular readout mechanisms in protein-DNA recognition. **Journal**

of **Molecular Biology**, [S. l.], v. 337, n. 2, p. 285–294, 2004.

MINHAS, V.; SUN, T.; MIRZOEV, A.; KOROLEV, N.; LYUBARTSEV, A. P.; NORDENSKIÖLD, L. Modeling DNA Flexibility: Comparison of Force Fields from Atomistic to Multiscale Levels. **Journal of Physical Chemistry B**, [S. l.], v. 124, n. 1, p. 38–49, 2020.

MONTICELLI, L.; TIELEMAN, D. P. Force Fields for Classical Molecular Dynamics. In: MONTICELLI, L.; SALONEN, E. (org.). **Biomolecular Simulations. Methods in Molecular Biology (Methods and Protocols)**. 924. ed. Totowa, NJ: Humana Press, 2013. b. v. 924p. 197–213.

MOTTET, D.; DUMONT, V.; DECCACHE, Y.; DEMAZY, C.; NINANE, N.; RAES, M.; MICHIELS, C. Regulation of Hypoxia-inducible Factor-1 α Protein Level during Hypoxic Conditions by the Phosphatidylinositol 3-Kinase/Akt/Glycogen Synthase Kinase 3 β Pathway in HepG2 Cells. **Journal of Biological Chemistry**, [S. l.], v. 278, n. 33, p. 31277–31285, 2003.

MOUILLOT, G.; MARCOU, C.; ZIDI, I.; GUILLARD, C.; SANGROUBER, D.; CAROSELLA, E. D.; MOREAU, P. Hypoxia Modulates HLA-G Gene Expression in Tumor Cells. **Human Immunology**, [S. l.], v. 68, n. 4, p. 277–285, 2007.

MURUGAN, R. Theory of Site-Specific DNA-Protein Interactions in the Presence of Conformational Fluctuations of DNA Binding Domains. **Biophysical Journal**, [S. l.], v. 99, n. 2, p. 353–359, 2010.

NAJI, A. et al. Binding of HLA-G to ITIM-Bearing Ig-like Transcript 2 Receptor Suppresses B Cell Responses. **The Journal of Immunology**, [S. l.], v. 192, n. 4, p. 1536–1546, 2014.

NAKAYAMA, K.; KATAOKA, N. Regulation of gene expression under hypoxic conditions. **International Journal of Molecular Sciences**, [S. l.], v. 20, n. 13, 2019.

NAMBA, A. M.; SILVA, V. B. Da; SILVA, C. H. T. P. Da. Dinâmica molecular: teoria e aplicações em planejamento de fármacos. **Eclética Química**, [S. l.], v. 33, n. 4, p. 13–24, 2008.

NEZNANOV, N.; UMEZAWA, A.; OSHIMA, R. G. A Regulatory Element within a Coding Exon Modulates Keratin 18 Gene Expression in Transgenic Mice. **Journal of Biological Chemistry**, [S. l.], v. 272, n. 44, p. 27549–27557, 1997.

NILGES, M.; GRONENBORN, A. M.; BRÜNGER, A.T.; CLORE, G. M. Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. Application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2. **“Protein Engineering, Design and Selection”**, [S. l.], v. 2, n. 1, p. 27–38, 1988.

OHH, M.; PARK, C. W.; IVAN, M.; HOFFMAN, M. A.; KIM, T.; HUANG, L. E.; PAVLETICH, N.; CHAU, Vi.; KAELIN, W. G. Ubiquitination of hypoxia-inducible factor requires direct binding to the β -domain of the von Hippel–Lindau protein. **Nature Cell Biology**, [S. l.], v. 2, n. 7, p. 423–427, 2000.

ORLANDO, I.M.C. et al. Distal and proximal hypoxia response elements cooperate to regulate

organ-specific erythropoietin gene expression. **Haematologica**, [S.I.], v.105, n.12, p. 2020, 2020.

OLDFIELD, C. J.; UVERSKY, V. N.; DUNKER, A. K.; KURGAN, L. Introduction to intrinsically disordered proteins and regions. In: **Intrinsically Disordered Proteins**. [s.l.] : Elsevier, 2019. p. 1–34.

OLSON, W. K. et al. A standard reference frame for the description of nucleic acid base-pair geometry. **Journal of Molecular Biology**, [S. l.], v. 313, n. 1, p. 229–237, 2001.

OLSON, W. K.; GORIN, A. A.; LU, X. J.; HOCK, L. M.; ZHURKIN, V. B. DNA sequence-dependent deformability deduced from protein-DNA crystal complexes. **Proceedings of the National Academy of Sciences**, [S. l.], v. 95, n. 19, p. 11163–11168, 1998.

PABO, C. O.; SAUER, R. T. Protein-DNA recognition. **Annual Review of Biochemistry**, [S. l.], v. 53, n. 1, p. 293–321, 1983.

PABO, C. O.; NEKLUDOVA, L. Geometric analysis and comparison of protein-DNA interfaces: Why is there no simple code for recognition? **Journal of Molecular Biology**, [S. l.], v. 301, n. 3, p. 597–624, 2000.

PÁLL, S.; HESS, B. A flexible algorithm for calculating pair interactions on SIMD architectures. **Computer Physics Communications**, [S. l.], v.184, n.12, p.2641–2650, 2013.

PARRINELLO, M.; RAHMAN, A. Polymorphic transitions in single crystals: A new molecular dynamics method. **Journal of Applied Physics**, [S. l.], v. 52, n. 12, p. 7182–7190, 1981.

PESCADOR, N.; CUEVAS, Y.; NARANJO, S.; VILLAR, D.; LANDÁZURI, M. O.; PESO, L. D. Identification of a functional hypoxia-responsive element that regulates the expression of the egl nine homologue 3 (egln3/phd3) gene. **Biochemical Journal**, [S.I.], v. 390, n. Pt 1, p.189–197, 2005.

PISTOLLATO, F. et al. Interaction of Hypoxia-Inducible Factor-1 α and Notch Signaling Regulates Medulloblastoma Precursor Proliferation and Fate. **STEM CELLS**, [S. l.], v. 28, n. 11, p. 1918–1929, 2010.

RAMACHANDRAN, G. N.; RAMAKRISHNAN, C.; SASISEKHARAN, V. Stereochemistry of polypeptide chain configurations. **Journal of molecular biology**, [S. l.], v. 7, n. 1, p. 95–99, 1963.

RCSB. **PDB Statistics: Protein-Nucleic Acid Complexes Released Per Year**. 2021. Disponível em: <https://www.rcsb.org/stats/growth/growth-protein-na-complex>. Acesso em: 2 fev. 2021.

RODRIGUES, J. P. G. L. M.; TRELLET, M.; SCHMITZ, C.; KASTRITIS, P.; KARACA, E.; MELQUIOND, A. S. J.; BONVIN, A. e M. J. J. Clustering biomolecular complexes by residue contacts similarity. **Proteins: Structure, Function, and Bioinformatics**, [S. l.], v.80, n.7, p.1810–7, 2012.

ROHS, R.; JIN, X.; WEST, S. M.; JOSHI, R.; HONIG, B.; MANN, R. S. Origins of Specificity in Protein-DNA Recognition. **Annual Review of Biochemistry**, [S. l.], v. 79, n. 1, p. 233–269, 2010.

ROLFS, A.; KVIETIKOVA, I.; GASSMANN, M.; WENGER, R. H. Oxygen-regulated Transferrin Expression Is Mediated by Hypoxia-inducible Factor-1. **Journal of Biological Chemistry**, [S. l.], v. 272, n. 32, p. 20055–20062, 1997.

ROST, B.; SANDER, C. Sequence-Structure Gap By Structure Predictions. **Annual review of biophysics and biomolecular structure**, [S. l.], v. 25, n. 1, p. 113–136, 1996.

SADJAD, B.; ZSOLDOS, Z. Toward a robust search method for the protein-drug docking problem. **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, [S. l.], v. 8, n. 4, p. 1120–1133, 2011.

SAGENDORF, J. M.; MARKARIAN, N.; BERMAN, H. M.; ROHS, R.. DNAProDB: An expanded database and web-based tool for structural analysis of DNA-protein complexes. **Nucleic Acids Research**, [S. l.], 2019.

SALCEDA, S.; CARO, J.. Hypoxia-inducible Factor 1 α (HIF-1 α) Protein Is Rapidly Degraded by the Ubiquitin-Proteasome System under Normoxic Conditions. **Journal of Biological Chemistry**, [S. l.], v. 272, n. 36, p. 22642–22647, 1997.

ŠALI, A.; BLUNDELL, T. L.; SALI, A.; BLUNDELL, T. L. Comparative protein modelling by satisfaction of spatial restraints. **Journal of Molecular Biology**, [S. l.], v. 234, n. 3, p. 779–815, 1993.

SALO-AHEN, O. M. H. et al. Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development. **Processes**, [S. l.], v. 9, n. 1, p. 71, 2020.

SAMANTA, D.; PRABHAKAR, N. R.; SEMENZA, G. L. Systems biology of oxygen homeostasis. **Wiley Interdisciplinary Reviews: Systems Biology and Medicine**, [S. l.], v. 9, n. 4, p. e1382, 2017.

SANTOS FILHO, O. A.; ALENCASTRO, R. B.. Modelagem de proteínas por homologia. **Química Nova**, [S. l.], v. 26, n. 2, p. 253–259, 2003.

SARAI, A.; MAZUR, J.; NUSSINOV, R.; JERNIGAN, R. L. Sequence dependence of DNA conformational flexibility. **Biochemistry**, [S. l.], v. 28, n. 19, p. 7842–7849, 1989.

SCHITO, L.; SEMENZA, G. L. Hypoxia-Inducible Factors: Master Regulators of Cancer Progression. **Trends in Cancer**, [S. l.], v. 2, n. 12, p. 758–770, 2016.

SCHÖDEL, J.; OIKONOMOPOULOS, S.; RAGOISSIS, J.; PUGH, C. W.; RATCLIFFE, P. J.; MOLE, D. R. High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. **Blood**, [S. l.], v. 117, n.23, p.e207-17, 2011.

SCHRÖDINGER. The PyMOL Molecular Graphics System, Version 2.4 Schrödinger, LLC.

SCHULTE, K. W.; GREEN, E.; WILZ, A.; PLATTEN, M.; DAUMKE, O. Structural Basis for Aryl Hydrocarbon Receptor-Mediated Gene Activation. **Structure**, [S. l.], v. 25, n. 7, p. 1025-1033.e3, 2017.

SEMENZA, G. L. Defining the role of hypoxia-inducible factor 1 in cancer biology and therapeutics. **Oncogene**, [S. l.], v. 29, n. 5, p. 625–634, 2010.

SEMENZA, G. L. Targeting HIF-1 for cancer therapy. **Nature Reviews Cancer**, [S. l.], v. 3, n. 10, p. 721–732, 2003.

SEMENZA, G. L. Hypoxia-inducible factors in physiology and medicine. **Cell**, [S. l.], v. 148, n. 3, p. 399-408, 2012.

SEMENZA, G. L.; JIANG, B.; LEUNG, S. W.; PASSANTINO, R.; CONCORDET, J.I.; MAIRE, P.; GIALONGO, A. Hypoxia Response Elements in the Aldolase A, Enolase 1, and Lactate Dehydrogenase A Gene Promoters Contain Essential Binding Sites for Hypoxia-inducible Factor 1. **Journal of Biological Chemistry**, [S. l.], v. 271, n. 51, p. 32529–32537, 1996.

SEO, K. S. et al. SIRT2 regulates tumour hypoxia response by promoting HIF-1 α hydroxylation. **Oncogene**, [S. l.], v. 34, n. 11, p. 1354–1362, 2015.

SHAO, Y.; ZHAO, F. Emerging evidence of the physiological role of hypoxia in mammary development and lactation. **Journal of Animal Science and Biotechnology**, [S. l.], v. 5, n. 1, p. 9, 2014.

SHIMIZU, T. Crystal structure of PHO4 bHLH domain-DNA complex: flanking base recognition. **The EMBO Journal**, [S. l.], v. 16, n. 15, p. 4689–4697, 1997.

SHING, P.; CARTER, M. DNA Structure: Alphabet Soup for the Cellular Soul. In: SELIGMANN, H. (org.). **DNA Replication-Current Advances**. [s.l.] : InTech, 2011.

SI, J.; ZHAO, R.; WU, R. An overview of the prediction of protein DNA-binding sites. **International Journal of Molecular Sciences**, [S. l.], v. 16, n. 3, p. 5194–5215, 2015.

SIM, B. C.; HUI, K. M. A HLA class I cis-regulatory element whose activity can be modulated by hormones. **International Journal of Cancer**, [S. l.], v. 59, n. 5, p. 646–654, 1994.

SLABINSKI, L.; JAROSZEWSKI, L.; RODRIGUES, A. P. C.; RYCHLEWSKI, L.; WILSON, I. A.; LESLEY, S. A.; GODZIK, A. The challenge of protein structure determination-lessons from structural genomics. **Protein Science**, [S. l.], v. 16, n. 11, p. 2472–2482, 2007.

SLATTERY, M.; ZHOU, T.; YANG, L.; DANTAS MACHADO, A. C.; GORDÂN, R.; ROHS, R.. Absence of a simple code: how transcription factors read the genome. **Trends in Biochemical Sciences**, [S. l.], v. 39, n. 9, p. 381–399, 2014.

SOKKAR, P.; SATHIS, V.; RAMACHANDRAN, M. Computational modeling on the recognition of the HRE motif by HIF-1: molecular docking and molecular dynamic studies. **J of Molecular Modeling**, [S. l.], v.18, n.5, p.1691-700, 2012.

SOUSA, F.; CRUZ, C.; QUEIROZ, J. A. Amino acids-nucleotides biomolecular recognition: from biological occurrence to affinity chromatography. **Journal of Molecular Recognition**, [S. l.], v. 23, n. 6, p. 505–518, 2010.

STEINEGGER, M.; SÖDING, J.. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. **Nature Biotechnology**, [S. l.], v. 35, n. 11, p. 1026–1028, 2017.

SUN, P. D.; FOSTER, C. E.; BOYINGTON, J. C. Overview of Protein Structural and Function Folds. **Curr Protoc Protein Sci**, [S. l.], v. 35, n. 1, p. 1711–171189, 2004.

SVINGEN, T.; TONISSEN, K. F. Hox transcription factors and their elusive mammalian gene targets targets. **Heredity**, v. 97, p. 88–96, 2006.

SVOZIL, Daniel; KALINA, Jan; OMELKA, Marek; SCHNEIDER, Bohdan. DNA conformations and their sequence preferences. **Nucleic Acids Research**, [S. l.], v. 36, n. 11, p. 3690–3706, 2008.

TAN, Z.; SHON, A. M.; OBER, C.. Evidence of balancing selection at the HLA-G promoter region. **Human Molecular Genetics**, [S. l.], v. 14, n. 23, p. 3619–3628, 2005.

TANAKA, T.; WIESENER, M.; BERNHARDT, W.; ECKARDT, K.; WARNECKE, C.. The human HIF (hypoxia-inducible factor)-3 α gene is a HIF-1 target gene and may modulate hypoxic gene induction. **Biochemical Journal**, [S. l.], v. 424, n. 1, p. 143–151, 2009.

TARCZEWSKA, A.; GREB-MARKIEWICZ, B. The significance of the intrinsically disordered regions for the functions of the bHLH transcription factors. **International Journal of Molecular Sciences**, [S. l.], v. 20, n. 21, 2019.

THOMPSON, J. D.; GIBSON, T. J.; PLEWNIAK, F.; JEANMOUGIN, F.; HIGGINS, D. G. The CLUSTAL X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. **Nucleic Acids Research**, [S. l.], v. 25, n. 24, p. 4876–4882, 1997.

TIAN, H.; MCKNIGHT, S. L.; RUSSELL, D. W. Endothelial PAS domain protein 1 (EPAS1), a transcription factor selectively expressed in endothelial cells. **Genes & Development**, [S. l.], v. 11, n. 1, p. 72–82, 1997.

TOLSTORUKOV, M. Y., COLASANTI, A. V., MCCANDLISH, D., OLSON, W. K., ZHURKIN, C. B. A Novel “Roll-and-Slide” mechanism of DNA Folding in Chromatin. Implications for nucleosome positioning. **Jornal of Molecular Biology**, [S. l.], v. 371, n. 3, p. 725–738, 2007.

TRAVERS, A. A. The structural basis of DNA flexibility. **Philosophical Transactions of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences**, [S. l.], v. 362, n. 1820, p. 1423–1438, 2004.

USSERY, David W. DNA Structure: A-, B- and Z-DNA Helix Families. In: **Encyclopedia of Life Sciences**. Chichester, UK: John Wiley & Sons, Ltd, 2002.

VAN DIJK, M.; BONVIN, A. M. J. J. Pushing the limits of what is achievable in protein–DNA docking: benchmarking HADDOCK’s performance. **Nucleic Acids Research**, [S. l.], v. 38, n. 17, p. 5634–5647, 2010.

VAN DIJK, M.; HSU, V.; ROLF, B.; BONVIN, A. M. J. J. Information-driven protein-DNA docking using HADDOCK: It is a matter of flexibility. **Nucleic Acids Research**, [S. l.], v. 34, n. 11, p. 3317–3325, 2006.

Van Gunsteren, W. F.; BERENDSEN, H. J. C. Groningen Molecular Simulation (GROMOS) Library Manual, **BIOMOS** b.v., Groningen, 1987.

VAN ZUNDERT, G. C. P. et al. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. **Journal of Molecular Biology**, [S. l.], v. 428, n. 4, p. 720–725, 2016.

VERLET, LOUP. Computer “Experiments” on Classical Fluids. II. Equilibrium Correlation Functions. **Physical Review**, [S. l.], v. 165, n. 1, p. 201–214, 1968.

VERLI, H. **Bioinformática da Biologia à Flexibilidade Molecular**. 1. ed. São Paulo: Sociedade Brasileira de Bioquímica e Biologia Molecular, 2014.

WASTOWSKI, I. J. et al. Human Leukocyte Antigen-G Is Frequently Expressed in Glioblastoma and May Be Induced in Vitro by Combined 5-Aza-2'-Deoxycytidine and Interferon- γ Treatments. **The American Journal of Pathology**, [S. l.], v. 182, n. 2, p. 540–552, 2013.

WEAKE, V. M.; WORKMAN, J. L. Inducible gene expression: diverse regulatory mechanisms. **Nature Reviews Genetics**, [S. l.], v. 11, n. 6, p. 426–437, 2010.

WEBB, B.; SALI, A. Comparative Protein Structure Modeling Using MODELLER. **Current Protocols in Bioinformatics**, [S. l.], v. 54, n. 1, 2016.

WEINER, P. K.; KOLLMAN, P.A. AMBER: Assisted model building with energy refinement. A general program for modeling molecules and their interactions **Journal of Computational Chemistry**. v. 2, n. 3, p. 287-303, 1981

WENGER, R. H.; STIEHL, D. P.; CAMENISCH, G. Integration of Oxygen Signaling at the Consensus HRE. **Science Signaling**, [S. l.], v. 2005, n. 306, p. re12–re12, 2005.

WIEDERSTEIN, M.; SIPPL, M. J. ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. **Nucleic Acids Research**, [S. l.], v. 35, n. Web Server, p. W407–W410, 2007.

WU, D.; POTLURI, N.; LU, J.; KIM, Y.; RASTINEJAD, F. Structural integration in hypoxia-inducible factors. **Nature**, [S. l.], v. 524, n. 7565, p. 303–308, 2015.

WU, D.; SU, X.; POTLURI, N.; KIM, Y.; RASTINEJAD, F. NPAS1-ARNT and NPAS3-

ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. **eLife**, [S. l.], v. 5, 2016.

XU, D. YAO, Y.; LU, L.; COSTA, M.; DAI, W. Plk3 Functions as an Essential Component of the Hypoxia Regulatory Pathway by Direct Phosphorylation of HIF-1 α . **Journal of Biological Chemistry**, [S. l.], v. 285, n. 50, p. 38944–38950, 2010.

XU, D.; ZHANG, Y. Ab initio Protein Structure Assembly Using continuous structure fragments and optimized knowledge-based force field. **Proteins**, [S. l.], v. 80, n. 7, p. 1715–1735, 2012.

YAGHI, L.; PORAS, I.; SIMOES, R. T.; DONADI, E. A.; TOST, J.; DAUNAY, A.; DE ALMEIDA, B. S.; CAROSELLA, E. D.; MOREAU, P. Hypoxia inducible factor-1 mediates the expression of the immune checkpoint HLA-G in glioma cells through hypoxia response element located in exon 2. **Oncotarget**, [S. l.], v. 7, n. 39, p. 63690–63707, 2016.

YANG, J. Q.; REMMERS, E. F.; MARCU, K. B. The first exon of the c-myc proto-oncogene contains a novel positive control element. **The EMBO Journal**, [S. l.], v. 5, n. 13, p. 3553–3562, 1986.

ZHANG, C. Accurate and efficient loop selections by the DFIRE-based all-atom statistical potential. **Protein Science**, [S. l.], v. 13, n. 2, p. 391–399, 2004.

ZHONG, H. et al. Overexpression of hypoxia-inducible factor 1 α in common human cancers and their metastases. **Cancer research**, [S. l.], v. 59, n. 22, p. 5830–5, 1999.

ZHOU, H.; ZHOU, Y.. Distance-scaled, finite ideal-gas reference state improves structure-derived potentials of mean force for structure selection and stability prediction. **Protein Science**, [S. l.], v.11, n. 11, p. 2714-2726, 2009.

ZHURKIN, V. B.; LYSOV, Y. P.; FLORENTIEV, V. L.; IVANOV, V. I. Torsional flexibility of DNA as revealed by conformational analysis. **Nucleic Acids Research**, [S. l.], v. 10, n. 5, p. 1811–1830, 1982.

8 Anexo

*Manuscript***Title**

**STRUCTURAL CHARACTERIZATION OF THE INTERACTION OF HYPOXIA
INDUCIBLE FACTOR-1 WITH ITS HYPOXIA RESPONSIVE ELEMENT AT THE -
964G>A VARIATION SITE OF THE *HLA-G* PROMOTER REGION**

Alves, Cinthia Caroline¹, Donadi, Eduardo Antônio², Giuliatti, Silvana¹

¹Department of Genetic, School of Medicine, University of São Paulo, Ribeirão Preto, Brazil;

² Department of Medicine, Ribeirão Preto Medical School, University of São Paulo, Ribeirão
Preto, São Paulo, Brazil.

ABSTRACT

Human Antigen Leukocyte-G (*HLA-G*) gene encodes an immune checkpoint molecule that has restricted tissue expression in physiological conditions; however, the gene may be induced in hypoxic conditions by the interaction with the hypoxia inducible factor-1 (HIF1). Hypoxia regulatory elements (HRE) located at *HLA-G* promoter region (-966 and -962 nucleotides) and at exon 2 are the major HIF1 target sites. The G allele of the -964 G>A transversion induces higher *HLA-G* expression when compared to the A allele in hypoxic conditions. Since *HLA-G* variability may influence the HIF1-HRE binding, we analyzed this complex interaction at the atomistic level considering both -964G and -964A alleles, using computational approaches. Mouse HIF2 dimer crystal (Protein Data Bank ID: 4ZPK, 3.6Å-<https://www.rcsb.org/>) that belongs to the same HIF1 protein family was used as template to perform homology modelling of human HIF1 quaternary structure, which encompasses residues 15-349 to HIF1a subunit and 87-470 to HIF1b subunit (UniProt ID: Q16665 and P27540, respectively-<https://www.uniprot.org/>), using MODELLERv9.24. The most similar template structural model was evaluated according to protein stereochemical quality (PROCHECK). Two 3D DNA structures were built from 5'GCRTG'3 HRE sequence containing the -964G or -964A allele at "R" position using x3DNA. Protein-DNA docking was performed using HADDOCKv2.4, and complex intermolecular interactions were computed by DNAProDB. Molecular dynamic simulation was carried out to protein and protein-DNA complexes per 200 ns, using Gromacs v.2019. Inter bp step from DNA atomic coordinate trajectories were calculated by Curves

program V.2.6. HIF1 most template similar model generated by homology modeling showed the lowest RMSD of 0.396Å after template alignment in comparison to other models, and it presented 87.2% of the residues in the core region of phi-psi torsion angles, while template presented 80.10%. HIF1 binding in the HRE containing -964 G allele results in more hydrogen bonds and Van der Waals contact formation than HRE with -964 A allele. RMSD protein-DNA complex trajectory analysis revealed that HIF1-HRE-964G complex is more stable. In conclusion, HIF1 binds in a more stable and specific manner at the HRE with G allele.

Key words: hypoxia induced factor-1, hypoxia regulatory element, HLA-G, protein-DNA interaction, molecular docking.

INTRODUCTION

Hypoxia plays an important role in physiological processes, such as embryogenesis, and placental development ^{1,2}, but also in pathological conditions, mainly in solid malignant tumors ³⁻⁶. Several studies have shown that hypoxia is strongly associated with lower overall survival and disease-free survival of various tumor types ^{3,4,6}. A decreased oxygen concentration in the tumor micro-environment leads to an adaptive response with modulation of gene expression involved in glycolysis, erythropoiesis, angiogenesis, and in many immune system regulators, facilitating cell survival and immune system evasion ⁶⁻⁹.

The master regulator of gene expression control in hypoxia is the hypoxia inducible factor 1 (HIF1), a heterodimeric protein that belongs to the basic helix-loop-helix PER-ARTN-SIM (bHLH-PAS) family ¹⁰. As a major regulator of oxygen homeostasis, HIF1 also contributes to the growth of tumors and its increased expression has been associated with poor prognosis. Therefore, the understanding of gene regulation by HIF1 has been considered as a potential therapeutic target ⁶. The HIF1 heterodimer is formed by an α subunit (HIF1a), whose stability is oxygen-regulated, and a stable β subunit (ARNT). In both subunits, the N-terminal domain (bHLH-PAS-A/B-PAC) is responsible for dimerization and DNA binding, whereas the C-terminal domains are required for protein stabilization and transactivation ¹¹⁻¹⁴. In the presence of cellular oxygen, HIF1a is hydroxylated by prolyl-hydroxylases (PHD) and by a factor Inhibitor of HIF1 (FIH), resulting in proteasomal degradation ¹⁴⁻¹⁹ and coactivator inhibition, such as CBP/p300, which is necessary to target gene transcription activation ^{19,20}. In contrast, under hypoxia conditions, the hydroxylases (PHD and FIH) become nonfunctional and the

stabilized HIF1 α can translocate into the nucleus to bind ARNT¹². Then, the formed heterodimeric HIF1 transcription factor recognizes and binds to the hypoxia responsive elements (HREs) at the 5'RCGTG'3 sequence, which are found in hypoxia responsive genes, activating the transcription process in association with CBP/p300²¹.

During hypoxia, HIF1 induces transcription of several genes that encode proteins involved in the regulation of the immune system⁶. The immune checkpoint human leukocyte antigen-G (*HLA-G*) gene is modulated in hypoxic conditions through HIF1^{22,23}. The membrane-bound HLA-G and the soluble isoforms inhibit the function of several leukocytes, including antigen presenting cells (macrophages and dendritic cells), T (CD4 and CD8), B, and Natural Killer (NK) cells, among others, by the interaction with the specific inhibitory receptors present on the leukocyte surface²⁴⁻²⁷. Physiologically, HLA-G plays a key role at the fetal-maternal interface keeping immune tolerance due the inhibition of the cytotoxic activity of the maternal TCD8+ and NK cells²⁸. However, HLA-G may be induced in numerous malignant tumors, such as glioblastoma and melanoma tumors, contributing to tumor immune escape²⁹.

Since microenvironment hypoxia is an important physiological mechanism for placentation and maternal fetal interface formation, hypoxia has also been associated with increased *HLA-G* expression³⁰. Additionally, hypoxia induces *HLA-G* transcription and protein production in a series of HLA-G-negative tumor lineages^{22,31,32}. *In silico* analysis of the 1.4kb region *upstream* the ATG translation initiation site revealed the presence of two HREs at positions -242 to -238 bp (nonfunctional) and at positions -966 to -962 positions^{23,31,33}, which contains a G>A polymorphism at position -964 (rs1632947)³⁴. Yaghi et al described for the first time the impact of the *HLA-G* gene variability at the HIF1-HRE interaction sites in association with the *HLA-G* expression under hypoxic conditions. The authors observed that a glioma tumor cell line that does not express HLA-G, when treated with the hypoxia-mimicking agent desferrioxamine (DFX), expressed high amounts of HLA-G mRNA and protein. The interaction of HIF1 with HRE at exon 2 presented synergy with the HRE at the promoter (-966 to -962) region, and the expression of HLA-G was increased in presence of the -964G allele when compared to the -964A allele²³.

Considering that: (i) the interaction of proteins with DNA is a crucial event for many cellular processes, such as transcription, and (ii) the promoter region -964G>A polymorphism affects the transcription of the *HLA-G* gene in hypoxic conditions²³, we evaluated the influence of the HIF1 binding at the *HLA-G* -964 G>A polymorphic site at an atom-pair interaction level³⁵. This strategy improved the understanding of the differential *HLA-G* regulation by HF1

by observing the physical interaction between the transcription factor and the polymorphism present at the *HLA-G* promoter region regulatory element.

METHODOLOGY

HIF1 dimer 3D structure modelling

For HIF1 quaternary 3D structure modelling, the crystal of mouse HIF2 dimer (PDB ID: 4ZPK. Chain A represents subunit β and chain B represents subunit α of HIF2 complex) with 3.6 Å was chosen as template to perform homology modelling. The HIF2 is also a transcription factor that belongs to the bHLH-PAS family and presents more than 70% of similarity with HIF1³⁶. Missing regions in the crystal were filled by PDB Reader tool on the CHARMM-GUI server³⁷. Primary structures of the HIF1 subunits, HIF1a (α), and ARNT (β) were retrieved from UniProt in *fasta* format with UniProt IDs Q16665 for human HIF1a residues 15-349, and P27540 for human ARNT residues 87-470. These primary sequences were aligned against crystal chains according to their correspondents α and β subunits, using the Clustal X software³⁸ and, then, the HIF1 dimer was modelled and refined using MODELLER v.9.24^{39,40}. The resulted models were evaluated for their template structural similarity by root mean square deviation (RMSD) of C α atoms, and after, the model with the lowest RMSD value was evaluated according to the protein stereochemical and physical-chemical properties, using the following Model Quality Assessment Programs (MQAPs): PROCHECK^{41,42}, Verify 3D⁴³, ERRAT⁴⁴ and Qualitative Model Energy Analysis (QMEAN)⁴⁶.

DNA modelling

We generated 3D structural models of DNA from the HRE sequence 5'GCGTG'3 present in the *HLA-G* promoter region at positions -966 and -962 (hg19_Chr6 position reference: 29794642-29794674), containing the G or A alleles of the -964G>A variation site (rs1632947)³⁴, along with 14 flanking nucleotides (5'TAAAAACAGGCAGTGC(G/A)TGAGCACTAGTGAGGG'3, referring to the -980 to -948 nucleotides), using the x3DNA v.2.4 software (LU, OLSON, 2008).

HIF1-HRE-964G/A molecular docking

The High Ambiguity Driven DOCKing HADDOCK webserver v.2.4 was used to drive protein-DNA dockings^{47,48}. HADDOCK uses biochemical and biophysical interaction data of partner molecules, which are introduced as ambiguous interaction restrictions to guide the docking process by active and passive residue selection⁴⁹. HIF1 was directly docked onto HRE consensus sequence containing the -964G or -964A allele. Active residues selected to HIF1 (HIF1a-ARNT) were S22, R23, A26, A27, R29 and R30 for the α subunit (HIF1a) and H94, E98, R101 and R102 for the β subunit (ARNT), which were conserved among the members of the bHLH-PAS protein family^{36,50,51}, and to DNA, -966G, -965C, -964G/A, -963T and -962G9, which correspond to the HRE sequence. Passive residues were automatically selected around active residues by the server.

Input parameters for the HADDOCK server remained in the standard mode, with the exception of the following parameters: (i) the dielectric constant (epsilon), which was set to 78 for it0 and it1 due to the high DNA charge; (ii) the initial temperature for the third cooling step of the temperature-accelerated dynamic (TAD) with fully flexible interface set to 300, (iii) the factor for the time interval in the TAD set to 4, (iv) the number of heating steps for the 300K phase set to 750, and (v) the sampling parameters were changed to: 10,000/400/400 for rigid body (it0), simulated annealing (it1) and water refinement (W) stages, respectively^{52,53}. In each docking stage, poses were classified according HADDOCK score (HS)⁴⁹. Protein-DNA interactions were analyzed with the aid of DNAProDB server⁵⁴.

Molecular dynamic simulations

Molecular dynamic simulations (DM) were performed to the free HIF1 and protein-DNA complexes using the GROMACS v.2019 package. The simulations for the free protein were performed using the all atom AMBER-99SB force field⁵⁵, whereas, for protein-DNA complexes, the same force field was used with the parameters referring to the DNA modified according to the AMBER-Parmbsc1⁵⁶. Both systems were built in a cubic box solvated with SPC/E water with a minimum of 14Å of distance between the edge and any atom of the protein or molecular complexes. General charge of the systems was neutralized with the addition of Cl⁻ and K⁺ ions to reach the total concentration of KCl to 150 mM. An energy minimization for 0.1 nanoseconds (ns) was carried out to eliminate bad contacts between the atoms, followed by an equilibration step with number of particles, volume, and temperature constants (NVT) and

another with number of particles, pressure, and temperature constant (NPT). MD simulations were carried out at constant pressure and temperature (1 atm and 300 K) per 200 ns ⁵⁷. Structural analyses deviation such as RMSD, root mean square fluctuations (RMSF) and raddi of gyration (RoG) were done from MD trajectories.

Conformational parameter differences of the DNA double helix structure during the molecular dynamic simulation of the HIF1-HRE-964G/A were assessed with Curves software ⁵⁸. All the helical parameters, such as the translational distances (rise, shift, slide), and the rotational angles (roll, tilt, twist) of the DNA were calculated from protein-DNA trajectory atomic coordinates every 10 ns, starting at time t=0 ns (n=21). Results are presented as mean and standard deviation.

3D structure visual inspection

The 3D structures of free protein as well protein-DNA complexes and their intermolecular interactions were visualized using PyMol Molecular Graphics System v.2.4 ⁵⁹.

RESULTS AND DISCUSSION

HIF1 heterodimer 3D structure modelling

We investigated the differential DNA-binding mechanism of HIF-1 into HRE at the *HLA-G* promoter region, considering the -964G>A variation site. Since proteins and DNA molecules differ from their geometry in unbound and bound forms ⁶⁰, the template chosen for modelling must be a protein-DNA complex, in which the protein has at least 30% of identity to HIF1. Thus, the template chosen for molecular modelling was the crystal of the mouse HIF2 (HIF2a-ARNT-HRE complex) transcription factor, which is a protein belonging to the same protein family as HIF1, which also interacts with HRE ¹⁰. In this crystal, chain A refers to the β subunit (ARNT), while chain B refers to the α subunit (HIF2a). In addition, HIF1 and HIF2 heterodimers have closely related architectures, as point mutations located at the binding sites that destabilize the HIF2 complex also destabilized HIF1 ³⁶, which makes HIF2 a good choice as a template for homology molecular modelling. This crystal has long sequences of missing amino acids (13 to 31 residues) in the tertiary structure of the two subunits, ARNT and HIF2a, which were filled before HIF1 modelling procedure.

Once the complete tertiary structure of the mouse HIF2 was obtained, it was possible to proceed with the molecular modeling of the human HIF1 transcription factor. For structural modeling, the primary structures of the two HIF1 protein subunits, human HIF1a and ARNT, were recovered from the UniProt database (<http://uniprot.org>) along with the residues corresponding to the bHLH domains, PAS-A/B and PAC (Residues 15-349 for HIF1a and residues 87-470 for ARNT), which were subsequently aligned with the primary sequence of the 4ZPK crystal chains for modeling, using MODELLER v.9.24^{39,61}. Ten different models were generated and analyzed for their structural similarity with crystal template by calculating the RMSD of protein C α atom superposition in the PyMol software. The model with most similar fold to the crystal was that one presenting the lowest RMSD value (RMSD: 0.396 Å). Then, this model was refined for loops with MODELLER v.9.24⁴⁰ and evaluated regarding chemical, physical, and stereochemical quality by MQAPs. Results obtained from the quality assessment for the transcription factor HIF1 in comparison with its template are shown in Table 1.

Table 1: Quality assessment of the HIF1 model compared to template.

MQAPs	4ZPKi	4ZPKc	HIF1
PROCHECK - Ramachandran Plot (1)			
Most favorable regions (A,B,L)	83.10	80.10	87.20
Additional allowed regions (a,b,l,p)	15.70	17.50	11.10
Generously allowed regions (~a,~b,~l,~p)	0.80	1.20	0.90
Disallowed regions	0.40	1.20	0.80
Verify-3D (2)	41.77	58.90	68.87
QMEAN global score (Z-score) (3)	0.52 (-2.44)	0.66 (-5.82)	0.65 (-2.77)
ERRAT (4)*	66.67/58.43	76.42/75.16	76.16/70.86

Reference values: (1) >90% of the amino acids in most favorable regions; (2) Good: $\geq 80\%$, Regular: <80%, Worse: <65% of 3D-1D sequence compatibility; (3) global score: values between 0 and 1, i which 1 is better; Z-scores around zero indicate good agreement between the model structure and experimental structures of similar size. Scores of -4.0 or below are an indication of models with low quality; (4) Good high resolution structures generally produce values around 95% or higher while lower resolutions (2.5-3Å) the average overall quality is around 91%; Total residues number: ZPKi: 610 ; 4ZPKc: 715; HIF1: 715; MQAP: Model Quality Assessment Program; i4ZPK: HIF2 crystal incomplete structure (missing regions); c4ZPK: HIF2 crystal complete structure; HIF1: HIF1 refined homology model; aa: amino acids.; * Values denominated as HIF subunits individually: ARNT/HIF a.

In general, the quality assessment shows that the model generated by homology for HIF1 showed good physical and chemical quality for most MQAPs in relation to the crystal (4ZPKi). In particular, Ramachandran plot indicates the most favorable, additionally allowed, generously allowed, and disallowed regions for the *phi* and *psi* torsion angles. Such analysis is performed for all amino acids, except for glycine and proline because they do not present these angles due to their side chain composition. HIF1 model presented 87.20% of its amino acids in

most favorable regions, and 0.80% of amino acids in disallowed regions. Although this model has not an ideal number of amino acids in most favorable and unfavorable regions (above 90% and 0%, respectively), it showed greater stereochemical quality compared to the crystal. However, none of the residues involved in intermolecular interaction with HRE showed torsion angles distribution in generously allowed, and unfavorable regions. Since: (i) the present work aims to analyze the structural interaction of DNA with the HIF1 dimer, and (ii) residues that bind to the HRE (DNA) have not presented poor quality regarding *phi-psi* torsion angles arrangement that may affect the binding of the protein complex with DNA, the model generated by homology modelling represented the HIF1 3D structure for carrying out the docking, as presented in Figure 1.

Figure 1: 3D representation of the human HIF1 quaternary structure modelling (HIF1a-ARNT complex). Chain A is highlighted in reference to β subunit (ARNT) in blue and chain B is highlighted in reference to α subunit (HIF1a) in pink. Boxes are the structures of each separate chain with illustrated protein domains.

DNA modelling

The HRE consensus sequence (5'GCGTG'3) at the *HLA-G* promoter region, along with fourteen 5' and 3' extremity flanking bases, was selected to model the 3D structure considering the -964G>A polymorphism (rs1632947) into HRE.^{33,35} The 5'TAAAAACAGGCAGTGC(G/A)TGAGCACTAGTGAGGG'3 (-980 to -948) sequence was used to construct a double-stranded type B (B-DNA) DNA, with the second complementary strand being constructed from the input sequence (named as -980r to -948r as a mention of reverse sequence). Both DNA molecules generated, one for the -964G allele and other for the -964A allele, are identical in terms of their structural conformation, except for the nucleotide sequence differing at the -964 position. The B-DNA of DNA conformation was chosen for modeling of the HRE at the *HLA-G* promoter region because it is equivalent to the biologically predominant form of DNA in the body^{62,63}.

HIF1-HRE-964G/A docking

Molecular docking is a method widely used in structural biology to explore the interaction of residues between two molecules. Once obtained, 3D structures of HIF1 and DNA (HRE-964G/A) by molecular modelling, it was possible to proceed with protein-DNA docking using the HADDOCK v.2.4 server^{48,49}. HIF1 has its DNA binding domain (DBD), bHLH, docked to its HRE in the promoter region of the *HLA-G* gene, emphasizing the existing polymorphism within the HRE. Thus, two protein-DNA complexes may be formed with the molecular docking technique using the same protocol: the HIF1-HRE-964G and HIF1-HRE-964A complex, which were named according to the allele present at the -964 polymorphic site within the HRE of the promoter region of the *HLA-G* gene.

In summary, HADDOCK is based on information from the interaction interface of the molecular complex predicted by experimental or computational approaches, which are introduced as ambiguous interaction restrictions (AIRs) to the direct docking. Such AIRs are defined through a list of residues that fall into two categories: active and passive⁴⁹, with a careful selection of which residues are active and which are passive, a step that is fundamental for the docking success.

Several DBDs have been studied in the literature by experimental methods, such as X-ray crystallography of protein-DNA complexes. This allows the identification of the DBD amino acids directly involved in the intermolecular bond with DNA bases or with sugar-phosphate backbone. HIF1 binds to HRE mainly through the N-terminal α -helices 1 present in

its α (HIF1a) and β (ARNT) subunits. Important DNA binding residues within α -helices 1 for DNA interaction were introduced based on crystallographic structures belonging to the same bHLH-PAS family of HIF1 transcription factors. These proteins are formed by two subunits that interact with each other forming heterodimers that bind to the regulatory element present in several target genes, such as the HIF1 transcription factor ⁶⁴. The crystallographic structures of heterodimers that share the ARNT subunit have a globally similar architecture between them when linked to the regulatory element, as observed for the mouse crystals from the AHR-ARNT-DNA ⁵¹, HIF1a/HIF2a-ARNT-DNA, ³⁶ and NPAS3-ARNT-DNA ⁵⁰ complexes. These proteins present DNA-binding residues conserved among the bHLH family of transcription factors, such as S22, R23, A26, A27, R29 and R30 for the α subunit (HIF1a); and H94, E98, R101 and R102 for the β subunit (ARNT), which were selected as active residues for HIF1 (HIF1a-ARNT) to drive docking process of both molecular complexes, HIF1-HRE-964G and HIF1-HRE-964A. Active residues for DNA molecule were selected based on the consensus sequence of HRE 5'GCRTG'3 ⁶⁵, in which the “R” position could be “G” (HIF1-HRE-964G complex) or “A” (HIF1-HRE-964A) allele. Passive residues were automatically selected by the server.

For the HIF1-HRE-964G complex, HADDOCK grouped 313 structures in 23 clusters, representing 78% of the 400 water-refined models that HADDOCK generated, whereas, for the HIF1-HRE-964A complex, HADDOCK grouped 326 structures in 25 clusters, representing 81% of the 400 models refined in water. These clusters were classified according to their HADDOCK score (HS), being the more negative the better.

Top ranked poses which were clustered in the ten top clusters by HADDOCK server were analyzed for their structural similarity against a protein-HRE reference complex that belongs to the bHLH-PAS family of transcription factors: HIF2a-ARNT-HRE complex (PDB ID: 4ZPK). Their 3D structures were superimposed in PyMol to RMSD calculation (data not shown) to identify a pose with similar structure to the classic conformation of the bHLH-PAS family ⁶⁴. Most of clustering poses presented HRE orientation at the HIF1 DBD different from expected; however, the second-best classified cluster for HIF1-HRE-964G (cluster 20) and the fourth-best classified cluster for HIF1-HRE-964A (cluster 2) presented the most similar poses to the reference complex. Poses of both clusters were similar with it one (Figure 2) and showed RMSD lower than the 2.0Å tolerance level when aligned to the crystal, showing good structural similarity.

Figure 2: 3D representation of similar docking pose obtained to HIF1-HRE molecular complexes generated by HADDOCK server. HRE in DNA double helix is highlight in red. Chain A is highlighted in reference to β subunit (ARNT) in blue and chain B is highlighted in reference to α subunit (HIF1a) in pink.

Physical-chemical characteristics from partner molecules define the affinity degree between them and are causally related to the intermolecular interactions in the complex. Then, energy functions derived from force fields and empirical terms are used to delineate correct and incorrect poses generated in the docking procedure by binding affinity estimation between molecules⁶⁶. In turn, binding affinity between molecules is related to Gibbs free energy (ΔG). In analogy to any spontaneous process, the connection between two molecules only occurs when changes in the free Gibbs energy (ΔG) of the system is negative, indicating that the system has reached an equilibrium state with constant pressure and temperature⁶⁷. Thus, the best poses generated in the docking process are those with negative intermolecular interaction energy values, as they indicate a probable biomolecular interaction^{49,67}. HADDOCK orders the best poses based on their HADDOCK score (HS), which is a score function that consists of the linear combination of various energy terms. Then, a pose with the lowest HS is considered the most energetically favorable structure. Best ranked poses for HS in clusters 20 and 2 were chosen as the final structure of the HIF1-HRE-964G and HIF1-HRE-964A complexes, respectively. Table 2 shows the values of the binding energy terms for the final structures of the HIF1-HRE-964G and HIF1-HRE-964A complexes generated by HADDOCK.

Table 2: Binding energy terms values comparison between the HIF1-HRE-964G and HIF1-HRE-964A complexes generated by HADDOCK.

Complex	HIF1-HRE-964G	HIF1-HRE-964A
HS (a.u.)	-163.53 (-152.70±6.60)	-147.90 (-140.60± 4.60)
Evw (kcal/mol)	-84.98 (-87.70±3.90)	-84.37 (-81.00± 3.90)
Eelec (kcal/mol)	-536.02 (-478.20±38.90)	-435.78 (-425.80± 23.10)
Eair (kcal/mol)	3.68 (12.10±13.00)	4.70 (7.80±3.00)
Edesol(kcal/mol)	28.29 (29.40±2.40)	23.15 (24.80±1.50)
BSA (Å)	2195.65 (2241.2±122.40)	2050.90 (2065.30± 86.90)

HS: HADDOCK *score*; EVdW: Van der Walls (VdW) energy; Eelec: Eletrostatic energy; Edesol: desolvation energy; Eair: restrains violation energy; BSA: Buried Surface Area; a.u.: arbitrary units. Statistics were calculated from the four best poses HS ranked for cluster 20 to the HIF1-HRE-964G. and for cluster 2 to the HIF1-HRE-964A.

As illustrated in Table 2. the HS value for the HIF1-HRE complex presenting the -964G allele (HS: -163.53 kcal/Mol) was lower compared to the HIF1-HRE complex containing the -964A allele (HS: -147.90 kcal/Mol). This difference suggests that the HIF1-HRE-964G complex is more energetically favorable than the HIF1-HRE-964A complex. In addition. HIF1-HRE964G complex showed a higher value for the buried surface area in relation to the transcription factor bound to the HRE of the -964A allele. indicating a higher interaction interface between the partner molecules in the first complex. These results show that the presence of the A allele at the -964 polymorphic site in the HRE at *HLA-G* promoter region decreases the interaction between HIF1 and HRE.

In the last decades. the increased number of protein-DNA high-resolution structures deposited in the PDB along with the availability of advanced technologies to explore regulatory elements in DNA. such as ChIP-seq. have allowed a greater expansion of knowledge about the protein-DNA binding specificity^{68,69}. Additionally. protein-DNA complex binding specificity can be understood by atomic interactions between amino acids and nucleotides including hydrogen bonds and Van der Walls (VdW) contacts⁷⁰. These interactions occur between protein side chains with sugar-phosphate backbone and with nitrogen bases in the major and minor grooves. in which the first interaction type is important for molecular complex stabilization and the second interaction type is responsible for specificity⁷¹⁻⁷³. Since genetic variations at the regulatory elements may alter the specificity of protein-DNA complex binding.⁷⁴ the analysis of HIF1-HRE intermolecular binding pattern considering -964G>A SNP at the *HLA-G* promoter region is important to the understanding of the gene transcription regulation.

Intermolecular interaction analysis was performed using the DNAProDB server⁵⁴ and visualized in PyMol. Both complexes are bound to their regulatory elements through hydrogen

bonds and VdWs contacts formation between α -helix 1 residues of the HIF1 subunits with the nitrogen bases in the major DNA groove and with the sugar-phosphate skeleton. When HIF1 binds to HRE containing the alternative -964A allele at the -964G>A variation site. 11 hydrogen bonds and 87 VdW contacts stabilize the complex. while in the presence of the -964G allele. 16 hydrogen bridges and 81 VdW contacts were identified.

As expected. VdW contacts are more abundant in protein-DNA complexes than hydrogen bonds and stabilize the molecular complex as a whole. but hydrogen bonds are stronger non-covalent bonds ^{70.75.76}. Then. the result analysis and discussion will be focused mainly on the amino acid-DNA contacts formed by the hydrogen bonds. In this context. eight hydrogen bonds are formed between HIF1 residues with HRE region in HIF1-HRE-964G. while six hydrogen bonds were observed in HIF1-HRE-964A. as shown in Figure 3.

Figure 3: Intermolecular contacts of HRE (DNA) and HIF1 (protein) atoms, emphasizing the *HLA-G* -964G>A polymorphism. (A) 3D representation showing hydrogen bonds at the HIF1-HRE-964G/A complexes regarding the regulatory element. The central box is a 3D illustration of the HRE (highlight in red at the DNA structure, and its complementary sequence in sand) recognition by HIF1 bHLH domains and on the right are illustrated the details regarding hydrogen bonds formed between residue atoms of the ARNT (blue) and HIF1a (pink) with nucleotide atoms of the -964G/A alleles. Hydrogen bonds (distance between 2.5-3.2Å) are indicated by yellow dotted lines. The red boxes indicate HRE nucleotides (5'-3') and sand boxes indicate their complementary bases. Hydrogen atoms are colored in white; phosphate in orange; oxygen in red; nitrogen in blue; and carbon are different colored between partner molecules: DNA carbon in brown. HIF1a carbon in pink. and ARNT carbon in light blue. (B) 2D interaction diagram showing intermolecular contacts at the HIF1-HRE-964G/A complexes regarding the regulatory element. The red boxes indicate the *HLA-G* -964G/A alleles at the HRE. considering the base pair.

Hydrogen bonds are indicated by yellow dotted lines. The red boxes indicate HRE nucleotides (5'-3') and sand boxes indicate their complementary bases. Despite hydrogen bonds, VdW contacts at the HIF1-HRE complex also occur between residues-nucleotides (K128-966G, R102-964G, E98-963T, E98-962rC, S95-963T, H94-962G, R91-962G, in ARNT and R30-965C, R30-964rC, R30-963rA, R23-965rG, in HIF1a for the complex with G allele, and R102-964A, R99-964A, E98-964A, H94-963T and R91-963T, in ARNT and R30-965C, R30-963rA, R23-965rG and S22-966rC, in HIF1a for the complex with A allele, data not shown).

Regarding the HIF1-HRE-964G complex, five residues are bound to HRE by sugar-phosphate skeleton atoms (R91 and R102 for ARNT, and K19, R27 and R29 for HIF1a), while only the E98 residue binding directly to the bases. The K19 and R27 amino acids form two hydrogen bonds with -966rC and -964rC nucleotides, respectively. For the HIF1-HRE-964A complex, five bonds are formed between amino acid atom pairs and HRE-sugar-phosphate skeleton, which are shared between the two protein-DNA complexes, except for the hydrogen bond formed by the R30 of HIF1a with the nucleotide -964rT, which is observed only in the HIF1-HRE complex containing the -964A allele. Noteworthy, no HIF1 residue formed hydrogen bridges directly with the G/A variation site, only VdW forces act at the interaction of these bases (R102-964G for the HIF1-HRE-964G, and R102-964A, R99-964A, E98-964A for the HIF1-HRE-964A). In contrast, hydrogen bonds are formed with complementary bases of the -964G/A alleles on the reverse strand (3'-5' direction) through the phosphate radical of the DNA chain (R27-964rT/C interaction in Figure 3), and this bond occurs independently of the -964G or A alleles.

The -964A allele altered HRE geometry in the complex HIF1-HRE-964A, resulting in interaction loss between partner molecules, such as the hydrogen bonds between K19-966rC and E98-962rC (Figure 3) observed to the complex with G allele. The number of hydrogen bonds formed between an amino acid and a nucleotide is directly proportional to the stability of the complex ⁷⁶, thus, the HIF1-HRE-964G complex is more stable than HIF1-HRE-964A. The E98 residue is highly conserved and shares a similar binding pattern among the bHLH-PAS family of transcription factors ^{36,50,51}. In addition, previous studies have shown that E98 is essential for the HRE interaction ⁷⁷. Thus, it is possible to infer that the -964A allele decreases HIF1-HRE binding specificity, as well as the stability of the molecular complex.

In addition to the residues shown in figure 3, hydrogen bonds between amino acids and nucleotides neighboring HRE were identified interacting with nucleotide base atoms (H94-961rT, S22-968G, S22-967T for the HIF1-HRE-964G complex; S22-967T, S22-968G for the HIF1-HRE-964A complex) and with sugar-phosphate backbone atoms (R101-961rT, K21-969A, R29-967T for the HIF1-HRE-964G complex; K21-969A, R29-967T for the HIF1-HRE-964A complex). Overall, hydrogen bond analysis showed a decreased interaction of HIF1 with

HRE in the presence of the *HLA-G* -964A allele when compared to the G allele. indicating that HIF1 binds in a more stable and specific manner with HRE in the presence of the -964G allele.

Protein-DNA molecular dynamic simulation

To verify the dynamic behavior of the protein-DNA complexes. we proceeded with the molecular dynamic simulation that evaluates complexes stability over time. Conformational variations of HIF1 complexed or not to HRE-964G/A were measured by RMSD, RoG, and RMSF during 200 ns of simulation. Figure 4 shows the results obtained for RMSD (A), and RoG (B) for backbone atoms (N, C α and C) from HIF1 structures in relation to the initial structure of the transcription factor. According to the RMSD results (Figure 4A), unbound HIF1 tends to equilibrate close to 50 ns, remains without major movements for the next 100 ns ($7.44 \pm 0.46 \text{ \AA}$) and remains equilibrated until the end of the trajectory ($8.79 \pm 0.22 \text{ \AA}$). HIF1 bound to HRE with G allele reaches a level of equilibrium in the first 10 ns of simulation and remains without movements until the end of the trajectory ($5.27 \pm 0.35 \text{ \AA}$). while in the presence of the -964A allele HIF1 stabilizes at 60 ns ($7.19 \pm 0.35 \text{ \AA}$) until the end of the 200 ns. RoG shows the compaction or expansion processes related to the polypeptide chain or DNA folding, indicating conformational change in the molecule. in which higher degree of compactness means lower values of RoG. According to RoG, free HIF1 showed low structural compaction throughout its trajectory ($29.15 \pm 0.49 \text{ \AA}$), acquiring a slightly folding shape at the end in relation to the initial structure. HIF1 bound to HRE showed low structural compaction throughout the

trajectory. independently of the HRE allele. is observed in Figure 4B (RoG average of $31.08 \pm 0.46 \text{ \AA}$ for HIF1-HRE-964G. and $31.18 \pm 0.17 \text{ \AA}$ for HIF1-HRE-964A).

Figure 4: Values of (A) Root Mean Square Deviation (RMSD) and (B) Radii of Gyration (RoG) over 200 ns for the HIF1 transcription factor (HIF1a-ARNT) bound to the HRE according to the *HLA-G* -964G>A polymorphism in comparison with the free HIF1.

HIF1 has higher structural stability when bound to the HRE containing the -964G allele when compared to the protein-DNA complex exhibiting the -964A allele. In addition, the HIF1-HRE with the A allele hits RMSD values like those of the free HIF1, reinforcing the hypothesis that HIF1 binding to HRE with the G allele is more stable. HIF1 folding did not differ between the two molecular complexes, which reached RoG values like that of the free transcription

factor in the initial time; however, the free protein showed a slight compaction during the simulation, while the complexes remained unchanged. accordingly, RoG results. RMSD and RoG for DNA were also analyzed (Figure 5A and B, respectively) showing a high flexibility for the two molecular complexes.

Figure 5: Values of (A) Root Mean Square Deviation (RMSD) and (B) Radii of Gyration (RoG) over 200 ns for the DNA double helix of HIF1-HRE complexes, according to the *HLA-G* -964G>A polymorphism.

RMSF describes residue position variation during the simulation, indicating the system flexibility. Figures 6A and 6B show the results of the RMSF for each HIF1 residue bound to HRE-964G/A in comparison to free HIF1, and to DNA chain nucleotides, respectively. HIF1

demonstrated a remarkably similar fluctuation pattern between the two protein-DNA complexes, differing only for some loop regions (regions above red line in figure 6A), where the transcription factor linked to the HRE with the A allele presented higher fluctuation in these residues than the HRE with the G allele. The HRE binding to bHLH domain of the HIF1 has stabilized the $\alpha 1$ helix regions in the ARNT and HIF1 α subunits (highlighted in the blue and pink, respectively, as shown in Figure 6) of the dimer for both HIF1-HRE-964G/A complexes, due to less fluctuation in these regions compared to the free protein. In fact, HIF1 must be complexed to its regulatory element at the promoter region of the gene to play its function, then, the absence of DNA anchorage to the HIF1 (free HIF1) results in high fluctuations at α -helix 1 of the DBD. In relation to the nucleic acid, fluctuations were observed mainly in the terminal nucleotides of the DNA double helix in both molecular complexes, while the protein binding site was considered stable (Figure 6B). The high fluctuation of the terminal nucleotides in both

protein-DNA complexes justifies the high conformational flexibility. as observed for the RMSD. and RoG values for DNA.

Figure 6: (A) RMSF values for the HIF1-HRE complexes according to the *HLA-G* -964G>A polymorphism. Residues 1 to 381 correspond to the ARNT subunit and residues 382 to 715 correspond to the HIF1a subunit. The dotted red line marks the amino acids with $\text{RMSF} > 3.0 \text{ \AA}$. The blue and pink arrows indicate the α -helices 1 of the ARNT (residues 1 to 18) and of the HIF1a (residues 382 to 400), respectively. (B) Comparisons of the DNA of the complex according to the *HLA-G* -964G>A polymorphism. Residues 15 to 19 correspond to the -966 to 962 hypoxia-responsive element (HRE) sequence and residues 48 to 52 correspond to the -966r to 962r complementary nucleotide sequence to the HRE. Both are highlighted in the beige boxes (the forward: HRE and the nucleotide sequence complementary to the HRE: csHRE). Each strand of DNA contains 33 bp.

Visual inspection of the protein-DNA complexes was performed to visualize conformational changes during molecular dynamic simulation. as shown in Figure 7. Atomic coordinates from molecular complexes were extracted at specific times in the trajectory and were superimposed with initial structure ($t=0\text{ns}$) using the PyMol software. Figure 6 shows

snapshots captured at specific times in the trajectory for the HIF1-HRE-964G/A complexes. HIF1 remained bound to DNA in both complexes over the time and their secondary structures were preserved. HIF1 loop regions and the 5'-3' terminal strand of the DNA double helix are regions with greater kinetics in the two molecular complexes analyzed, corroborating with the high RMSF values observed for such regions of the molecules.

Figure 7: Visualization of trajectory snapshots at times 0 (A) and 200ns (B) for the HIF1-HRE-964G and HIF1-HRE-964A complexes. In the Figure (C) is illustrated the structure superposition at specific times (0, 50, 150 and 200 ns) of the trajectory for the HIF1-HRE964G/A complexes, in which the initial structure is illustrated in yellow, the time final structure is illustrated in red, and other structures are illustrated in gray. The blue and pink arrows indicate the α -helix 1 region of the ARNT and HIF1 α subunits, respectively. The HRE is highlighted in green in the 3D DNA structure.

Although the HIF1-HRE complex with the G allele presents more stability in comparison to the HIF1-HRE complex containing the A allele, RMSD and RoG values are still high. This may be due to the subunits of HIF1 (ARNT and HIF1 α) that have intrinsically disordered regions (IDRs) located in the protein C-terminal regions and DBD N-terminal^{78,79}. In this context, intrinsically disordered proteins (IDPs) and IDRs do not present a stable and unique conformation, which are able to adopt different conformations according to the

molecular partner that interacts to play its biological function⁸⁰. IDRs properties are provided by their amino acid composition and can present different residue lengths accompanied by well-structured domains⁸¹. Thus, the acquisition of protein stability occurs when complexed to its molecular partner⁸².

In vitro studies have shown that HIF1 bound to HRE located at the *HLA-G* promoter region (-966 HRE) does not induce gene transactivation and protein production under hypoxic conditions, but it acts in synergy with another HRE located at exon 2 (positions +281 sense and +291 antisense)²³. The joint action of the two HREs positively regulates *HLA-G* gene transcription and protein production with the aid of cofactors, such as the CBP/p300 co-activator^{21,83}. In turn, a DNA loop formation connects transcription factors and their target regulatory elements located away from the basal transcription complex, propitiating that the transcriptional process occurs independently of physical distance between HREs^{84,85}. Thus, through the formation of a DNA loop, the HREs located at the promoter region and at exon 2 of the *HLA-G* gene become accessible to HIF1 binding in hypoxic conditions.

The absence of *HLA-G* exon 2 HRE and cofactors in our protein-DNA systems as well as the lack of the N-terminal transactivation domain (N-TAD) of the HIF1^{12,13}, in which the cofactor CBP/p300 interacts with HIF1, may have influenced the higher values of RMSD and RoG, since these components support the transcription factor stability. On the other hand, HIF1-HRE complex containing the G allele was more stable than the other HIF1-HRE complex with A allele according molecular dynamics results, corroborating previous experimental data; i. e., the HRE at the *HLA-G* promoter region alone is not able to induce the protein production *in vitro*; however, the G allele showed higher affinity for HIF1 than the A allele²³.

Protein-DNA interaction induces changes in DNA double helix geometry due DNA flexibility and the deformation caused by the protein in the nucleic acid structure is essential for molecular recognition, which is called indirect readout mechanism⁸⁶⁻⁸⁸. To analyze the differences in the conformational parameters of the 3D structure of the DNA double helix during the simulation of molecular dynamics of the HIF1-HRE complexes, inter base pair (bp) step parameters from protein-D trajectory atomic coordinates were calculated using Curves program⁵⁸. Results are shown in Figures 8 and 9 for the HRE in the analyzed molecular complexes.

Inter-bp step parameters are dependent of the DNA nucleotide sequence and are analyzed for each dinucleotide within the three chemical classes observed for the HRE: pyrimidine-purine (YR step), purine-purine (RR step) and purine- pyrimidine (RY step). These

parameters comprise three translational distances (rise, shift, and slide) and three rotational angles (roll, twist, and tilt). Regarding the translational distances, molecular complexes showed remarkably similar values for most of the HRE dinucleotides in an independent manner of the -964G/A polymorphism (Figure 8).

Figure 8: Average values (Å) comparisons regarding the shift, slide, and rise calculated for the HRE from HIF1-HRE-964G/A complexes trajectory atomic coordinates every 10 ns.

In contrast, the presence of the polymorphism in the HRE caused slight deformations in the DNA double helix structure (Figure 9): (i) roll angle varied about 4 degrees in the first and third bp steps (RY and YR steps, respectively) of the HRE among the molecular complexes analyzed; (ii) tilt angle shifted 4 degrees in the second bp step, that is, in the dCG and dCA dinucleotides for the complexes containing the HRE with the G and A allele, respectively; (iii) twist angle showed a variation of approximately 3 degrees in the second and fourth bp step (TG dinucleotide, TGd): in relation to the second bp step, the HRE with A allele has a higher value

for the twist in relation to the complex with the HRE-964G and then. in the fourth bp step. there is a conversion of values for the twist between these molecular complexes.

Figure 9: Average values (°) comparison of the roll, tilt and twist calculated for the HRE from HIF1-HRE-964G/A complexes trajectory atomic coordinates every 10 ns.

Generally, a structural parameter combination describes the deformations in the DNA double helix. For instance, dinucleotide curvature occurs preferably through the roll angle and may or may not be accompanied by translational movements on the slide⁸⁹⁻⁹¹. Roll and slide show a negative correlation, where positive values for roll and negative for slide tend to cause a bending in the two stacked nucleotides towards the major groove, while negative values for roll and positive for slides cause a bend in the minor groove⁹¹. However, such combination is not observed between roll and slide parameters for the HRE steps of our molecular complexes. In particular, patterns of side by side dinucleotides presenting positive and negative roll angles result in the production of a linear conformation of the double helix⁹², inferring that HIF1 when bound to the HRE does not cause a curvature in the double helix, as shown in Figure 9 for most HRE steps.

Roll and twist rotational angles present higher flexibility and contribute to the curvature magnitude of the nucleic acid then tilt angle^{88,90}. Among the three chemical classes observed for the HRE, YR steps (second and fourth bp-step in Figure 9) presented higher values for the

roll and twist compared to the other bp steps, corroborating previous studies^{92,93}, except for the CGd of the regulatory element containing the G allele for the twist. In fact, the 5' and 3' ends of the HRE with the -964A allele are more flexible than the middle of the regulatory sequence, whereas for the HRE with the -964G allele, higher flexibility is observed for the 3' end given by the fourth step in relation to the twist angle. This difference in the complex flexibility for the YR steps referring to the twist angle between analyzed complexes may be due to the presence of the -964G/A *HLA-G* SNP into the HRE, since the twist angle is more strongly affected by neighboring bases than the other conformational parameters⁹³. And then, the HIF1-HRE interaction may present a different indirect read out mechanism, irrespective of the *HLA-G* -964G>A polymorphism.

Tilt rotational angle does not substantially contribute to the DNA curvature and deformations in the double helix resulting from this angle are more energetically cost than other inter bp parameters^{90,94}. Positive tilt values cause an opening of base pairs stacked on the strand, which high values cause steric clashes between these base pairs resulting in a penalty for the structure energy terms^{88,95,96}. Thus, deformations caused by tilt angle in the bp-steps generate local conformations that may affect DNA interaction with biological molecules, such as transcription factors. Again, YR steps are more flexible for tilt compared to the other HRE steps, as shown in Figure 9; however, the second CAd step of the HRE with A allele showed higher values compared to the CGd of the HRE-964G. Then, the presence of the A allele at the HRE generates a DNA double helix conformation less energetically favorable. To supply this energy penalty caused by the tilt angle, an increase in the rise parameter is required for the bp-steps⁸⁸; however, no difference is observed for rise in the CAd step of the HIF1-HRE-964A complex and the energy cost is not fixed. Furthermore, tilt DNA double helix deformation in the HRE-964A may have influenced its interaction with HIF1, decreasing the stability of this molecular complex, as previously mentioned regarding the results of the HIF1-HRE-964A molecular dynamic simulations.

In conclusion, HIF1 interacts in a more stable and favorable manner with the HRE containing the *HLA-G* -964G allele when compared to the HRE containing the -964A allele. A limitation to understand the role of HIF1 on the transcription regulation of the *HLA-G* gene using computational molecular modelling techniques is the presence of other HRE motifs (exon 2 HRE), beside the one at the promoter region²³, demanding further analyses. Notwithstanding, the identification of the differential behavior of *HLA-G* -964G/A alleles at the atomistic level may explain the differential production of the immune checkpoint HLA-G molecule under

hypoxic conditions. The understanding of the *HLA-G* gene regulation under hypoxic conditions may be particularly useful in conditions, in which the expression of the molecule may be advantageous like in pregnancy and in transplantation, as opposed to conditions associated with harmful effects of the HLA-G expression, such as in cancer disorders.

Acknowledgements

We acknowledge the Brazilian Research Foundations CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico, processes # 141224/2017-7, and 302060/2019-7) and CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, Finance code 001) for the financial support and STI/USP for providing HPC resources for this research.

Author contributions statement

Silvana Giuliatti and Eduardo Antônio Donadi conceived the idea of the project; Silvana Giuliatti, devised and implement the scientific project; Cinthia Caroline Alves, performed computational techniques and analysis. Cinthia Caroline Alves also wrote the manuscript in consultation with Silvana Giuliatti and Eduardo Antônio Donadi, who provided critical feedback to construct the article.

Conflicts of interest

The authors declare absence conflict of interest regarding this paper publication.

REFERENCES

- (1) Dunwoodie, S. L. The Role of Hypoxia in Development of the Mammalian Embryo. *Dev. Cell* **2009**, 17 (6), 755–773. <https://doi.org/10.1016/j.devcel.2009.11.008>.
- (2) Shao, Y.; Zhao, F.-Q. Emerging Evidence of the Physiological Role of Hypoxia in Mammary Development and Lactation. *J. Anim. Sci. Biotechnol.* **2014**, 5 (1), 9. <https://doi.org/10.1186/2049-1891-5-9>.
- (3) Gilkes, D. M.; Semenza, G. L. Role of Hypoxia-Inducible Factors in Breast Cancer Metastasis. *Futur. Oncol.* **2013**, 9 (11), 1623–1636. <https://doi.org/10.2217/fon.13.92>.
- (4) Zhong, H.; De Marzo, A. M.; Laughner, E.; Lim, M.; Hilton, D. A.; Zagzag, D.; Buechler, P.; Isaacs, W. B.; Semenza, G. L.; Simons, J. W. Overexpression of Hypoxia-Inducible Factor 1alpha in Common Human Cancers and Their Metastases. *Cancer Res.* **1999**, 59 (22), 5830–5835.

- (5) Semenza. G. L. Defining the Role of Hypoxia-Inducible Factor 1 in Cancer Biology and Therapeutics. *Oncogene* **2010**. 29 (5). 625–634. <https://doi.org/10.1038/onc.2009.441>.
- (6) Schito. L.; Semenza. G. L. Hypoxia-Inducible Factors: Master Regulators of Cancer Progression. *Trends in Cancer* **2016**. 2 (12). 758–770. <https://doi.org/10.1016/j.trecan.2016.10.016>.
- (7) Benkert. P.; Künzli. M.; Schwede. T. QMEAN Server for Protein Model Quality Estimation. *Nucleic Acids Res.* **2009**. 37 (suppl_2). W510–W514. <https://doi.org/10.1093/nar/gkp322>.
- (8) Forsythe. J. A.; Jiang. B. H.; Iyer. N. V; Agani. F.; Leung. S. W.; Koos. R. D.; Semenza. G. L. Activation of Vascular Endothelial Growth Factor Gene Transcription by Hypoxia-Inducible Factor 1. *Mol. Cell. Biol.* **1996**. 16 (9). 4604–4613. <https://doi.org/10.1128/MCB.16.9.4604>.
- (9) Haase. V. H. Regulation of Erythropoiesis by Hypoxia-Inducible Factors. *Blood Rev.* **2013**. 27 (1). 41–53. <https://doi.org/10.1016/j.blre.2012.12.003>.
- (10) Dengler. V. L.; Galbraith. M. D. M.; Espinosa. J. M. J. M. Transcriptional Regulation by Hypoxia Inducible Factors. *Crit. Rev. Biochem. Mol. Biol.* **2014**. 49 (1). 1–5. <https://doi.org/10.3109/10409238.2013.838205>.
- (11) Tian. H.; McKnight. S. L.; Russell. D. W. Endothelial PAS Domain Protein 1 (EPAS1). a Transcription Factor Selectively Expressed in Endothelial Cells. *Genes Dev.* **1997**. 11 (1). 72–82. <https://doi.org/10.1101/gad.11.1.72>.
- (12) Jiang. B.-H.; Rue. E.; Wang. G. L.; Roe. R.; Semenza. G. L. Dimerization. DNA Binding. and Transactivation Properties of Hypoxia-Inducible Factor 1. *J. Biol. Chem.* **1996**. 271 (30). 17771–17778. <https://doi.org/10.1074/jbc.271.30.17771>.
- (13) Jiang. B.-H.; Zheng. J. Z.; Leung. S. W.; Roe. R.; Semenza. G. L. Transactivation and Inhibitory Domains of Hypoxia-Inducible Factor 1 α . *J. Biol. Chem.* **1997**. 272 (31). 19253–19260. <https://doi.org/10.1074/jbc.272.31.19253>.
- (14) Huang. L. E.; Gu. J.; Schau. M.; Bunn. H. F. Regulation of Hypoxia-Inducible Factor 1 Is Mediated by an O₂-Dependent Degradation Domain via the Ubiquitin-Proteasome Pathway. *Proc. Natl. Acad. Sci.* **1998**. 95 (14). 7987–7992. <https://doi.org/10.1073/pnas.95.14.7987>.
- (15) Appelhoff. R. J.; Tian. Y.-M.; Raval. R. R.; Turley. H.; Harris. A. L.; Pugh. C. W.; Ratcliffe. P. J.; Gleadle. J. M. Differential Function of the Prolyl Hydroxylases PHD1. PHD2. and PHD3 in the Regulation of Hypoxia-Inducible Factor. *J. Biol. Chem.* **2004**.

- 279 (37). 38458–38465. <https://doi.org/10.1074/jbc.M406026200>.
- (16) Berra. E. HIF Prolyl-Hydroxylase 2 Is the Key Oxygen Sensor Setting Low Steady-State Levels of HIF-1 in Normoxia. *EMBO J.* **2003**. 22 (16). 4082–4090. <https://doi.org/10.1093/emboj/cdg392>.
 - (17) Salceda. S.; Caro. J. Hypoxia-Inducible Factor 1 α (HIF-1 α) Protein Is Rapidly Degraded by the Ubiquitin-Proteasome System under Normoxic Conditions. *J. Biol. Chem.* **1997**. 272 (36). 22642–22647. <https://doi.org/10.1074/jbc.272.36.22642>.
 - (18) Iliopoulos. O.; Levy. A. P.; Jiang. C.; Kaelin. W. G.; Goldberg. M. A. Negative Regulation of Hypoxia-Inducible Genes by the von Hippel-Lindau Protein. *Proc. Natl. Acad. Sci.* **1996**. 93 (20). 10595–10599. <https://doi.org/10.1073/pnas.93.20.10595>.
 - (19) Lando. D. FIH-1 Is an Asparaginyl Hydroxylase Enzyme That Regulates the Transcriptional Activity of Hypoxia-Inducible Factor. *Genes Dev.* **2002**. 16 (12). 1466–1471. <https://doi.org/10.1101/gad.991402>.
 - (20) Mahon. P. C. FIH-1: A Novel Protein That Interacts with HIF-1 α and VHL to Mediate Repression of HIF-1 Transcriptional Activity. *Genes Dev.* **2001**. 15 (20). 2675–2686. <https://doi.org/10.1101/gad.924501>.
 - (21) Schödel. J.; Oikonomopoulos. S.; Ragoussis. J.; Pugh. C. W.; Ratcliffe. P. J.; Mole. D. R. High-Resolution Genome-Wide Mapping of HIF-Binding Sites by ChIP-Seq. *Blood* **2011**. <https://doi.org/10.1182/blood-2010-10-314427>.
 - (22) Mouillot. G.; Marcou. C.; Zidi. I.; Guillard. C.; Sangrouber. D.; Carosella. E. D.; Moreau. P. Hypoxia Modulates HLA-G Gene Expression in Tumor Cells. *Hum. Immunol.* **2007**. 68 (4). 277–285. <https://doi.org/10.1016/j.humimm.2006.10.016>.
 - (23) Yaghi. L.; Poras. I.; Simoes. R. T.; Donadi. E. A.; Tost. J.; Daunay. A.; de Almeida. B. S.; Carosella. E. D.; Moreau. P. Hypoxia Inducible Factor-1 Mediates the Expression of the Immune Checkpoint HLA-G in Glioma Cells through Hypoxia Response Element Located in Exon 2. *Oncotarget* **2016**. 7 (39). 63690–63707. <https://doi.org/10.18632/oncotarget.11628>.
 - (24) Le Gal. F.-A.; Riteau. B.; Sedlik. C.; Khalil-Daher. I.; Menier. C.; Dausset. J.; Guillet. J.-G.; Carosella. E. D.; Rouas-Freiss. N. HLA-G-Mediated Inhibition of Antigen-Specific Cytotoxic T Lymphocytes. *Int. Immunol.* **1999**. 11 (8). 1351–1356. <https://doi.org/10.1093/intimm/11.8.1351>.
 - (25) Horuzsko. A.; Lenfant. F.; Munn. D. H.; Mellor. A. L. Maturation of Antigen-Presenting Cells Is Compromised in HLA-G Transgenic Mice. *Int. Immunol.* **2001**. 13 (3). 385–394.

- <https://doi.org/10.1093/intimm/13.3.385>.
- (26) Naji. A.; Menier. C.; Morandi. F.; Agaugué. S.; Maki. G.; Ferretti. E.; Bruel. S.; Pistoia. V.; Carosella. E. D.; Rouas-Freiss. N. Binding of HLA-G to ITIM-Bearing Ig-like Transcript 2 Receptor Suppresses B Cell Responses. *J. Immunol.* **2014**. 192 (4). 1536–1546. <https://doi.org/10.4049/jimmunol.1300438>.
 - (27) Gregori. S.; Magnani. C. F.; Roncarolo. M. G. Role of Human Leukocyte Antigen-G in the Induction of Adaptive Type 1 Regulatory T Cells. *Hum. Immunol.* **2009**. <https://doi.org/10.1016/j.humimm.2009.07.022>.
 - (28) Klein. J.; Sato. A. The HLA System: First of Two Parts. *New Engl. J. Med. Rev.* **2000**. 343 (10). 702–709. <https://doi.org/10.1186/s40246-017-0107-5>.
 - (29) Lin. A.; Yan. W. H. Heterogeneity of HLA-G Expression in Cancers: Facing the Challenges. *Frontiers in immunology*. 2018. <https://doi.org/10.3389/fimmu.2018.02164>.
 - (30) Kilburn. B. A.; Wang. J.; Duniec-Dmuchowski. Z. M.; Leach. R. E.; Romero. R.; Armant. D. R. Extracellular Matrix Composition and Hypoxia Regulate the Expression of HLA-G and Integrins in a Human Trophoblast Cell Line1. *Biol. Reprod.* **2000**. 62 (3). 739–747. <https://doi.org/10.1095/biolreprod62.3.739>.
 - (31) Chang. C.-C.; Ferrone. S. HLA-G in Melanoma: Can the Current Controversies Be Solved? *Semin. Cancer Biol.* **2003**. 13 (5). 361–369. [https://doi.org/10.1016/S1044-579X\(03\)00027-0](https://doi.org/10.1016/S1044-579X(03)00027-0).
 - (32) Wastowski. I. J.; Simões. R. T.; Yaghi. L.; Donadi. E. A.; Pancoto. J. T.; Poras. I.; Lechapt-Zalcman. E.; Bernaudin. M.; Valable. S.; Carlotti. C. G.; Flajollet. S.; Jensen. S. S.; Ferrone. S.; Carosella. E. D.; Kristensen. B. W.; Moreau. P. Human Leukocyte Antigen-G Is Frequently Expressed in Glioblastoma and May Be Induced in Vitro by Combined 5-Aza-2'-Deoxycytidine and Interferon- γ Treatments. *Am. J. Pathol.* **2013**. 182 (2). 540–552. <https://doi.org/10.1016/j.ajpath.2012.10.021>.
 - (33) Castelli. E. C.; Veiga-Castelli. L. C.; Yaghi. L.; Moreau. P.; Donadi. E. A. Transcriptional and Posttranscriptional Regulations of the HLA-G Gene. *J. Immunol. Res.* **2014**. 2014. 1–15. <https://doi.org/10.1155/2014/734068>.
 - (34) Tan. Z.; Shon. A. M.; Ober. C. Evidence of Balancing Selection at the HLA-G Promoter Region. *Hum. Mol. Genet.* **2005**. 14 (23). 3619–3628. <https://doi.org/10.1093/hmg/ddi389>.
 - (35) Castelli. E. C.; Ramalho. J.; Porto. I. O. P.; Lima. T. H. A.; FelÃ-cio. L. P.; Sabbagh. A.; Donadi. E. A.; Mendes-Junior. C. T. Insights into HLA-G Genetics Provided by

- Worldwide Haplotype Diversity. *Front. Immunol.* **2014**. 5. <https://doi.org/10.3389/fimmu.2014.00476>.
- (36) Wu. D.; Potluri. N.; Lu. J.; Kim. Y.; Rastinejad. F. Structural Integration in Hypoxia-Inducible Factors. *Nature* **2015**. 524 (7565). 303–308. <https://doi.org/10.1038/nature14883>.
- (37) Jo. S.; Kim. T.; Iyer. V. G.; Im. W. CHARMM-GUI: A Web-Based Graphical User Interface for CHARMM. *J. Comput. Chem.* **2008**. 29 (11). 1859–1865. <https://doi.org/10.1002/jcc.20945>.
- (38) Thompson. J. D.; Gibson. T. J.; Plewniak. F.; Jeanmougin. F.; Higgins. D. G. The CLUSTAL X Windows Interface: Flexible Strategies for Multiple Sequence Alignment Aided by Quality Analysis Tools. *Nucleic Acids Res.* **1997**. 25 (24). 4876–4882. <https://doi.org/10.1093/nar/25.24.4876>.
- (39) Webb. B.; Sali. A. Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinforma.* **2016**. 54 (1). <https://doi.org/10.1002/cpbi.3>.
- (40) Fiser. A.; Do. R. K. G.; Šali. A. Modeling of Loops in Protein Structures. *Protein Sci.* **2000**. <https://doi.org/10.1002/9780470882207.ch13>.
- (41) Ramachandran. G. N.; Ramakrishnan. C.; Sasisekharan. V. Stereochemistry of Polypeptide Chain ConFigurations. *J. Mol. Biol.* **1963**. 7 (1). 95–99. [https://doi.org/10.1016/S0022-2836\(63\)80023-6](https://doi.org/10.1016/S0022-2836(63)80023-6).
- (42) Laskowski. R. a.; MacArthur. M. W.; Moss. D. S.; Thornton. J. M. PROCHECK: A Program to Check the Stereochemical Quality of Protein Structures. *J. Appl. Crystallogr.* **1993**. 26 (2). 283–291. <https://doi.org/10.1107/S0021889892009944>.
- (43) Eisenberg. D.; Lüthy. R.; Bowie. J. U. [20] VERIFY3D: Assessment of Protein Models with Three-Dimensional Profiles. In *Methods in Enzymology*; 1997; pp 396–404. [https://doi.org/10.1016/S0076-6879\(97\)77022-8](https://doi.org/10.1016/S0076-6879(97)77022-8).
- (44) Colovos. C.; Yeates. T. O. Verification of Protein Structures: Patterns of Nonbonded Atomic Interactions. *Protein Sci.* **1993**. 2 (9). 1511–1519. <https://doi.org/10.1002/pro.5560020916>.
- (45) Zhou. H.; Zhou. Y. Distance-Scaled. Finite Ideal-Gas Reference State Improves Structure-Derived Potentials of Mean Force for Structure Selection and Stability Prediction. *Protein Sci.* **2009**. 11 (11). 2714–2726. <https://doi.org/10.1110/ps.0217002>.
- (46) Benkert. P.; Künzli. M.; Schwede. T. QMEAN Server for Protein Model Quality Estimation. *Nucleic Acids Res.* **2009**. <https://doi.org/10.1093/nar/gkp322>.

- (47) Dominguez. C.; Boelens. R.; Bonvin. A. M. J. J. HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* **2003**. <https://doi.org/10.1021/ja026939x>.
- (48) Van Zundert. G. C. P.; Rodrigues. J. P. G. L. M.; Trellet. M.; Schmitz. C.; Kastitis. P. L.; Karaca. E.; Melquiond. A. S. J.; Van Dijk. M.; De Vries. S. J.; Bonvin. A. M. J. J. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**. 428 (4). 720–725. <https://doi.org/10.1016/j.jmb.2015.09.014>.
- (49) De Vries. S. J.; Van Dijk. M.; Bonvin. A. M. J. J. The HADDOCK Web Server for Data-Driven Biomolecular Docking. *Nat. Protoc.* **2010**. 5 (5). 883–897. <https://doi.org/10.1038/nprot.2010.32>.
- (50) Wu. D.; Su. X.; Potluri. N.; Kim. Y.; Rastinejad. F. NPAS1-ARNT and NPAS3-ARNT Crystal Structures Implicate the BHLH-PAS Family as Multi-Ligand Binding Transcription Factors. *Elife* **2016**. 5. <https://doi.org/10.7554/eLife.18790>.
- (51) Schulte. K. W.; Green. E.; Wilz. A.; Platten. M.; Daumke. O. Structural Basis for Aryl Hydrocarbon Receptor-Mediated Gene Activation. *Structure* **2017**. 25 (7). 1025–1033.e3. <https://doi.org/10.1016/j.str.2017.05.008>.
- (52) Honorato. R. V.; Roel-Touris. J.; Bonvin. A. M. J. J. MARTINI-Based Protein-DNA Coarse-Grained HADDOCKing. *Front. Mol. Biosci.* **2019**. 6. <https://doi.org/10.3389/fmolb.2019.00102>.
- (53) Van Dijk. M.; Van Dijk. A. D. J.; Hsu. V.; Rolf. B.; Bonvin. A. M. J. J. Information-Driven Protein-DNA Docking Using HADDOCK: It Is a Matter of Flexibility. *Nucleic Acids Res.* **2006**. <https://doi.org/10.1093/nar/gkl412>.
- (54) Sagendorf. J. M.; Markarian. N.; Berman. H. M.; Rohs. R. DNAProDB: An Expanded Database and Web-Based Tool for Structural Analysis of DNA–Protein Complexes. *Nucleic Acids Res.* **2019**. <https://doi.org/10.1093/nar/gkz889>.
- (55) Hornak. V.; Abel. R.; Okur. A.; Strockbine. B.; Roitberg. A.; Simmerling. C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins Struct. Funct. Bioinforma.* **2006**. 65 (3). 712–725. <https://doi.org/10.1002/prot.21123>.
- (56) Ivani. I.; Dans. P. D.; Noy. A.; Pérez. A.; Faustino. I.; Hospital. A.; Walther. J.; Andrio. P.; Goñi. R.; Balaceanu. A.; Portella. G.; Battistini. F.; Gelpí. J. L.; González. C.; Vendruscolo. M.; Laughton. C. A.; Harris. S. A.; Case. D. A.; Orozco. M. Parmbsc1: A

- Refined Force Field for DNA Simulations. *Nat. Methods* **2016**. 13 (1). 55–58. <https://doi.org/10.1038/nmeth.3658>.
- (57) Maciaszczyk-Dziubinska. E.; Reymer. A.; Kumar. N. V.; Białek. W.; Mizio. K.; Tamás. M. J.; Wysocki. R. The Ancillary N-Terminal Region of the Yeast AP-1 Transcription Factor Yap8 Contributes to Its DNA Binding Specificity. *Nucleic Acids Res.* **2020**. <https://doi.org/10.1093/nar/gkaa316>.
- (58) Lavery. R.; Moakher. M.; Maddocks. J. H.; Petkeviciute. D.; Zakrzewska. K. Conformational Analysis of Nucleic Acids Revisited: Curves+. *Nucleic Acids Res.* **2009**. 37 (17). 5917–5929. <https://doi.org/10.1093/nar/gkp608>.
- (59) DeLano. W. L. References. *Hypertens. Res.* **2014**. 37 (4). 362–387. <https://doi.org/10.1038/hr.2014.17>.
- (60) Svozil. D.; Kalina. J.; Omelka. M.; Schneider. B. DNA Conformations and Their Sequence Preferences. *Nucleic Acids Res.* **2008**. 36 (11). 3690–3706. <https://doi.org/10.1093/nar/gkn260>.
- (61) Šali. A.; Blundell. T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**. 234 (3). 779–815. <https://doi.org/10.1006/jmbi.1993.1626>.
- (62) Frank-Kamenetskii. M. D. DNA Topology. *J. Mol. Struct. THEOCHEM* **1995**. 336 (2–3). 235–243. [https://doi.org/10.1016/0166-1280\(94\)04078-7](https://doi.org/10.1016/0166-1280(94)04078-7).
- (63) Ussery. D. W. DNA Structure: A-, B- and Z-DNA Helix Families. In *Encyclopedia of Life Sciences*; John Wiley & Sons. Ltd: Chichester. UK. 2002. <https://doi.org/10.1038/npg.els.0003122>.
- (64) Graham. A. M.; Presnell. J. S. Hypoxia Inducible Factor (HIF) Transcription Factor Family Expansion. Diversification. Divergence and Selection in Eukaryotes. *PLoS One* **2017**. <https://doi.org/10.1371/journal.pone.0179545>.
- (65) Wenger. R. H.; Stiehl. D. P.; Camenisch. G. Integration of Oxygen Signaling at the Consensus HRE. *Sci. Signal.* **2005**. 2005 (306). re12–re12. <https://doi.org/10.1126/stke.3062005re12>.
- (66) Meng. X.-Y.; Zhang. H.-X.; Mezei. M.; Cui. M. Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Curr. Comput. Aided-Drug Des.* **2011**. 7 (2). 146–157. <https://doi.org/10.2174/157340911795677602>.
- (67) Kitchen. D. B.; Decornez. H.; Furr. J. R.; Bajorath. J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nat. Rev. Drug Discov.* **2004**.

- 3 (11). 935–949. <https://doi.org/10.1038/nrd1549>.
- (68) Rohs. R.; Jin. X.; West. S. M.; Joshi. R.; Honig. B.; Mann. R. S. Origins of Specificity in Protein-DNA Recognition. *Annu. Rev. Biochem.* **2010**. 79 (1). 233–269. <https://doi.org/10.1146/annurev-biochem-060408-091030>.
 - (69) Slattery. M.; Zhou. T.; Yang. L.; Dantas Machado. A. C.; Gordân. R.; Rohs. R. Absence of a Simple Code: How Transcription Factors Read the Genome. *Trends Biochem. Sci.* **2014**. 39 (9). 381–399. <https://doi.org/10.1016/j.tibs.2014.07.002>.
 - (70) Mandel-Gutfreund. Y.; Schueler. O.; Margalit. H. Comprehensive Analysis of Hydrogen Bonds in Regulatory Protein DNA-Complexes: In Search of Common Principles. *J. Mol. Biol.* **1995**. 253 (2). 370–382. <https://doi.org/10.1006/jmbi.1995.0559>.
 - (71) Luscombe. N. M.; Thornton. J. M. Protein – DNA Interactions: Amino Acid Conservation and the Effects of Mutations on Binding Specificity. **2002**. 2836 (02). 991–1009. [https://doi.org/10.1016/S0022-2836\(02\)00571-5](https://doi.org/10.1016/S0022-2836(02)00571-5).
 - (72) Matthews. B. W. Protein-DNA Interaction. No Code for Recognition. *Nature* **1988**. 335 (6188). 294–295. <https://doi.org/10.1038/335294a0>.
 - (73) Pabo. C. O.; Sauer. R. T. Protein-DNA Recognition. *Annu. Rev. Biochem.* **1984**. 53 (1). 293–321. <https://doi.org/10.1146/annurev.bi.53.070184.001453>.
 - (74) Izumi. K. Disorders of Transcriptional Regulation: An Emerging Category of Multiple Malformation Syndromes. *Mol. Syndromol.* **2016**. 7 (5). 262–273. <https://doi.org/10.1159/000448747>.
 - (75) Luscombe. N. M.; Austin. S. E.; Berman. H. M.; Thornton. J. M. An Overview of the Structures of Protein-DNA Complexes. *Genome Biol.* **2000**. 1 (1). 1–37. <https://doi.org/10.1186/gb-2000-1-1-reviews001>.
 - (76) Luscombe. N. M. Amino Acid-Base Interactions: A Three-Dimensional Analysis of Protein-DNA Interactions at an Atomic Level. *Nucleic Acids Res.* **2001**. 29 (13). 2860–2874. <https://doi.org/10.1093/nar/29.13.2860>.
 - (77) Bacsí. S. G.; Hankinson. O. Functional Characterization of DNA-Binding Domains of the Subunits of the Heterodimeric Aryl Hydrocarbon Receptor Complex Imputing Novel and Canonical Basic Helix-Loop-Helix Protein-DNA Interactions. *J. Biol. Chem.* **1996**. 271 (15). 8843–8850. <https://doi.org/10.1074/jbc.271.15.8843>.
 - (78) Kolonko. M.; Greb-Markiewicz. B. BHLH–PAS Proteins: Their Structure and Intrinsic Disorder. *Int. J. Mol. Sci.* **2019**. 20 (15). <https://doi.org/10.3390/ijms20153653>.
 - (79) Tarczewska. A.; Greb-Markiewicz. B. The Significance of the Intrinsically Disordered

- Regions for the Functions of the BHLH Transcription Factors. *Int. J. Mol. Sci.* **2019**, *20* (21). <https://doi.org/10.3390/ijms20215306>.
- (80) Anfinsen. C. B. Principles That Govern the Folding of Protein Chains. *Science* (80-.). **1973**, *181* (4096), 223–230. <https://doi.org/10.1126/science.181.4096.223>.
- (81) Dyson. H. J.; Wright. P. E. Intrinsically Unstructured Proteins and Their Functions. *Nat. Rev. Mol. Cell Biol.* **2005**, *6* (3), 197–208. <https://doi.org/10.1038/nrm1589>.
- (82) Oldfield. C. J.; Uversky. V. N.; Dunker. A. K.; Kurgan. L. Introduction to Intrinsically Disordered Proteins and Regions. In *Intrinsically Disordered Proteins*; Elsevier. 2019; pp 1–34. <https://doi.org/10.1016/B978-0-12-816348-1.00001-6>.
- (83) Semenza. G. L.; Jiang. B.-H.; Leung. S. W.; Passantino. R.; Concordet. J.-P.; Maire. P.; Giallongo. A. Hypoxia Response Elements in the Aldolase A. Enolase 1. and Lactate Dehydrogenase A Gene Promoters Contain Essential Binding Sites for Hypoxia-Inducible Factor 1. *J. Biol. Chem.* **1996**, *271* (51), 32529–32537. <https://doi.org/10.1074/jbc.271.51.32529>.
- (84) Kolovos. P.; Knoch. T. A.; Grosveld. F. G.; Cook. P. R.; Papantonis. A. Enhancers and Silencers: An Integrated and Simple Model for Their Function. *Epigenetics Chromatin* **2012**, *5* (1), 1. <https://doi.org/10.1186/1756-8935-5-1>.
- (85) Weake. V. M.; Workman. J. L. Inducible Gene Expression: Diverse Regulatory Mechanisms. *Nat. Rev. Genet.* **2010**, *11* (6), 426–437. <https://doi.org/10.1038/nrg2781>.
- (86) Lu. X. J.; Olson. W. K. Resolving the Discrepancies among Nucleic Acid Conformational Analyses. *J. Mol. Biol.* **1999**, *285* (4), 1563–1575. <https://doi.org/10.1006/jmbi.1998.2390>.
- (87) Michael Gromiha. M.; Siebers. J. G.; Selvaraj. S.; Kono. H.; Sarai. A. Intermolecular and Intramolecular Readout Mechanisms in Protein-DNA Recognition. *J. Mol. Biol.* **2004**, *337* (2), 285–294. <https://doi.org/10.1016/j.jmb.2004.01.033>.
- (88) Sarai. A.; Mazur. J.; Nussinov. R.; Jernigan. R. L. Sequence Dependence of DNA Conformational Flexibility. *Biochemistry* **1989**, *28* (19), 7842–7849. <https://doi.org/10.1021/bi00445a046>.
- (89) Djuranovic. D.; Hartmann. B. Conformational Characteristics and Correlations in Crystal Structures of Nucleic Acid Oligonucleotides: Evidence for Sub-States. *J. Biomol. Struct. Dyn.* **2003**, *20* (6), 771–788. <https://doi.org/10.1080/07391102.2003.10506894>.
- (90) Olson. W. K.; Gorin. A. A.; Lu. X.-J.; Hock. L. M.; Zhurkin. V. B. DNA Sequence-Dependent Deformability Deduced from Protein-DNA Crystal Complexes. *Proc. Natl.*

- Acad. Sci.* **1998**. 95 (19). 11163–11168. <https://doi.org/10.1073/pnas.95.19.11163>.
- (91) Tolstorukov. Michael Y.. Colasanti. Andrew V.. McCandlish. D.. Olson. Wilma K.. Zhurkin. C. B. A Novel “Roll-and-Slide” Mechanism of DNA Folding in Chromatin. Implications for Nucleosome Positioning. *J. Mol. Biol.* **2007**. 371 (3). 725–738.
- (92) Dickerson. R. E.; Chiu. T. K. Helix Bending as a Factor in Protein/DNA Recognition. *Biopolym. - Biospectroscopy Sect.* **1997**. 44 (4). 361–403. [https://doi.org/10.1002/\(sici\)1097-0282\(1997\)44:4<361::aid-bip4>3.0.co;2-x](https://doi.org/10.1002/(sici)1097-0282(1997)44:4<361::aid-bip4>3.0.co;2-x).
- (93) Gu. C.; Zhang. J.; Yang. Y. I.; Chen. X.; Ge. H.; Sun. Y.; Su. X.; Yang. L.; Xie. S.; Gao. Y. Q. DNA Structural Correlation in Short and Long Ranges. *J. Phys. Chem. B* **2015**. 119 (44). 13980–13990. <https://doi.org/10.1021/acs.jpcb.5b06217>.
- (94) Curuksu. J.; Zacharias. M.; Lavery. R.; Zakrzewska. K. Local and Global Effects of Strong DNA Bending Induced during Molecular Dynamics Simulations. *Nucleic Acids Res.* **2009**. 37 (11). 3766–3773. <https://doi.org/10.1093/nar/gkp234>.
- (95) Lavery. R.; Sklenar. H. Defining the Structure of Irregular Nucleic Acids: Conventions and Principles. *J. Biomol. Struct. Dyn.* **1989**. 6 (4). 655–667. <https://doi.org/10.1080/07391102.1989.10507728>.
- (96) Zhurkin. V. B.; Lysov. Y. P.; Florentiev. V. L.; Ivanov. V. I. Torsional Flexibility of DNA as Revealed by Conformational Analysis. *Nucleic Acids Res.* **1982**. 10 (5). 1811–1830. <https://doi.org/10.1093/nar/10.5.1811>.

9 Apêndices

APÊNDICE 1

Tabela 4: Resumo das estatísticas para os termos de energia de ligação para os 10 melhores clusters gerados para o complexo HIF1-HRE-964G. As estatísticas foram calculadas a partir dos 4 principais modelos de cada cluster.

<i>Cluster</i>	HS	Tamanho do cluster	RMSD	EVdW	Eelec	Edesol	Eair	BSA	z-score
1	-158,4±4,8	46	1,2±1,6	-96,6±8,6	-459,4±50,9	29,4±1,3	6,6±2,2	2336,5±88,2	-1,3
20	-152,7±6,6	5	25,5±0,1	-87,7±3,9	-478,2±38,9	29,4±2,4	12,1±13,0	2241,2±122,4	-1,1
5	-148,4±2,6	17	25,7±0,1	-82,6±1,7	-463,7±15,6	26,4±1,9	4,8±0,6	2031,8±108,3	-1
7	-145,7±2,6	14	3,7±0,2	-81,4±5,8	-452,8±40,6	25,7±1,7	5,3±0,9	2106,0±113,5	-0,9
11	-124,6±17,1	12	4,6±1,1	-86,5±8,6	-327,3±40,8	25,1±3,6	23,1±11,0	2090,5±188,1	-0,1
12	-121,3±13,7	11	1,9±1,4	-75,7±5,1	-375,8±56,1	26,7±2,6	28,7±6,1	1870,6±226,0	0
14	-101,4±18,9	7	25,5±0,2	-73,3±13,4	-279,7±41,2	24,2±1,8	37,3±4,2	1771,0±299,3	0,8
6	-95,9±3,2	15	23,6±0,1	-74,3±3,7	-232,5±6,3	17,1±0,8	77,8±12,7	1720,8±12,9	1
10	-88,6±4,8	12	18,1±0,1	-76,3±2,0	-164,6±32,3	11,3±0,8	93,5±14,9	1877,8±75,4	1,2
15	-82,4±16,8	6	8,1±1,2	-59,0±6,2	-225,3±75,7	18,3±5,8	33,5±13,3	1444,1±239,9	1,5

HS: *score* HADDOCK (a.u.);

RMSD: Desvio Médio Quadrático (do inglês, *Root Mean Square Deviation*) da estrutura geral de menor energia em Å;

EVdW: Energia de Van Der Walls (kcal/Mols);

Eelec: Energia Eletrostática (kcal/mol);

Edesol: Energia de dessolvatação (kcal/mol);

Eair: Energia de violação das restrições (kcal/Mol);

BSA: Superfície de área enterrada (do inglês, *Buried Surface Area*, BSA),

APÊNDICE 2

Tabela 5: Resumo das estatísticas para os termos de energia de ligação para os 10 melhores clusters gerados para o complexo HIF1-HRE-964A, As estatísticas foram calculadas a partir dos 4 principais modelos de cada cluster,

<i>Cluster</i>	HS	Tamanho do cluster	RMSD	EVdW	Eelec	Edesol	Eair	BSA	Z-score
15	-150,7±8,5	6	1,1±0,8	-93,1±6,4	-450,8±22,7	29,4±1,4	31,8±18,3	2368,0±117,7	-1,2
4	-143,8±6,6	26	25,2±0,1	-86,8±3,4	-430,8±39,1	28,6±3,4	5,3±0,6	2167,7±95,6	-0,9
11	-142,0±10,0	8	4,8±0,2	-86,4±9,8	-413,1±16,5	26,4±1,4	6,0±1,6	2177,0±122,8	-0,8
2	-140,6±4,6	36	2,9±0,8	-81,0±3,9	-425,8±23,1	24,8±1,5	7,8±3,0	2065,3±86,9	-0,8
9	-136,5±7,5	11	25,4±0,0	-83,2±8,3	-395,7±28,6	25,2±4,0	6,1±1,9	2058,1±186,9	-0,6
3	-129,6±6,9	36	25,3±0,1	-88,7±5,0	-348,4±27,3	27,9±3,1	9,3±5,6	2173,5±42,0	-0,3
6	-102,5±1,8	22	6,3±0,1	-75,7±3,0	-252,9±15,3	16,8±1,6	70,0±5,1	1745,8±53,0	0,8
21	-100,0±14,2	5	7,1±0,1	-71,7±12,3	-285,0±32,4	24,9±4,9	38,1±1,6	1690,1±261,6	0,9
20	-90,0±31,4	5	24,7±0,5	-69,2±17,6	-220,6±98,0	20,1±7,1	31,3±14,4	1569,6±405,2	1,3
12	-80,7±7,3	7	17,6±0,0	-73,6±5,8	-140,5±23,0	11,9±2,0	92,0±3,6	1837,3±65,2	1,7

HS: *score* HADDOCK (a.u.);

RMSD: Desvio Médio Quadrático (do inglês, *Root Mean Square Deviation*) da estrutura geral de menor energia em Å;

EVdW: Energia de Van Der Walls (kcal/Mols);

Eelec: Energia Eletrostática (kcal/mol);

Edesol: Energia de dessolvatação (kcal/mol);

Eair: Energia de violação das restrições (kcal/Mol);

BSA: Superfície de área enterrada (do inglês, *Buried Surface Area*, BSA),

[PÁGINA EM BRANCO]