



UNIVERSIDADE ESTADUAL PAULISTA
“JÚLIO DE MESQUITA FILHO”
Câmpus de São José do Rio Preto

Wan Suk Augusto Mazucato

**Investigação computacional da interação M2-1/Fosfoproteína do Vírus
Sincicial Respiratório humano: explorando peptídeos miméticos como
potenciais agentes antivirais**

São José do Rio Preto
2024

Wan Suk Augusto Mazucato

**Investigação computacional da interação M2-1/Fosfoproteína do Vírus
Sincicial Respiratório humano: explorando peptídeos miméticos como
potenciais agentes antivirais**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao Programa de Pós-Graduação em Ciências Farmacológicas e Biomoleculares, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES e PROPG.

Orientador: Prof. Dr. Ícaro Putinhon Caruso

São José do Rio Preto
2024

Wan Suk Augusto Mazucato

**Investigação computacional da interação M2-1/Fosfoproteína do Vírus
Sincicial Respiratório humano: explorando peptídeos miméticos como
potenciais agentes antivirais**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao Programa de Pós-Graduação em Ciências Farmacológicas e Biomoleculares, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Comissão Examinadora

Prof. Dr. Ícaro Putinhon Caruso
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof. Dr. Alexandre Suman de Araújo
UNESP – Câmpus de São José do Rio Preto

Prof. Dr. Flávio Augusto Vicente Seixas
UEM- Universidade Estadual de Maringá-PR

São José do Rio Preto
04 de março de 2024

M476i

Mazucato, Wan Suk Augusto

Investigação computacional da interação M2-1/Fosfoproteína do
Vírus Sincicial Respiratório humano: explorando peptídeos miméticos
como potenciais agentes antivirais / Wan Suk Augusto Mazucato. -São
José do Rio Preto, 2024

121 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP),
Instituto de Biociências Letras e Ciências Exatas, São José do Rio
Preto

Orientador: Ícaro Putinhon Caruso

1. Biofísica Molecular. 2. Dinâmica Molecular. 3. hRSV. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Universidade
Estadual Paulista (UNESP), Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto.

Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Dedico este trabalho a minha mãe Valeria Regina de Cillo Mazucato,
por acreditar em mim, mesmo quando eu não conseguia.

AGRADECIMENTOS

Em primeiro lugar agradeço ao meu orientador Professor Dr. Ícaro Putinhon Caruso pela confiança depositada em mim, por me ensinar e iluminar o meu caminho. Obrigado pela oportunidade de experimentar a vida acadêmica. Obrigado por ser um exemplo de pessoa, que levarei comigo sempre. Obrigado por tudo.

Agradeço à professora Dra. Fátima Pereira de Souza pelos valiosos conselhos, principalmente na época de iniciação científica. Obrigado por me ajudar a desenvolver resiliência necessária na carreira acadêmica e em diversas profissões.

Agradeço à toda minha família, que me deu suporte para que eu pudesse me dedicar ao trabalho acadêmico. Em especial à minha esposa Ohk Sun dos Santos Durães Mazucato, pelo amor e dedicação para comigo; Aos meus pais Juvenal Mazucato e Valeria Regina de Cillo Mazucato pelo suporte familiar e financeiro; E a minha irmã Valéria Sun Hwa Mazucato Galvão, pela amizade e carinho.

Agradeço a todos os meus amigos. Em especial à Tereza Pereira de Souza, pela amizade e sinceridade.

Agradeço à minha psicóloga e amiga Gizeli Cristina Codonho Vilches de Camargos, por me apoiar nos melhores e piores momentos.

Agradeço aos alunos de pós-graduação Dr. Raphael Vinicius Rodrigues Dias e Ms. João Victor Piloto pelas valiosas discussões e contribuições sobre o presente trabalho, elevando em muito o nível do trabalho realizado.

Agradeço ao técnico Bruno Rogério Fernandes que sempre se mostrou solícito em relação às dificuldades encontradas no trabalho computacional, mostrando uma forma elegante de superar os problemas encontrados.

Agradeço aos integrantes da banca de qualificação: Prof. Dr. Alexandre Suman de Araújo, Dra. Ingrid Bernardes Santana Martins e Prof. Dr. Ícaro Putinhon Caruso.

Agradeço aos integrantes da banca de defesa: Prof. Dr. Alexandre Suman de Araújo, Prof. Dr. Flávio Augusto Vicente Seixas e Prof. Dr. Ícaro Putinhon Caruso.

Agradeço a todos os colegas e professores integrantes do CMIB e do departamento de Física, que de alguma forma contribuíram para o meu aprendizado.

Agradeço as instituições de pesquisa Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e da Pró-Reitoria de Pós-Graduação (PROPG da UNESP) pelo financiamento de minha bolsa de estudos para a realização do presente trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

RESUMO

O Vírus Sincicial Respiratório humano (hRSV) é um patógeno que causa infecções agudas do trato respiratório, como bronquiolite e pneumonia. Esse vírus apresenta alto potencial de contágio, afetando principalmente idosos e crianças até 2 anos de idade. Dentre as diversas proteínas envolvidas no processo replicativo do hRSV, destacamos a fosfoproteína P e a proteína M2-1, que interagem e facilitam a transcrição de todas as proteínas virais. Atualmente, estudos na literatura estabelecem que a interação entre essas proteínas é essencial para o processo de transcrição do hRSV, no qual a M2-1 atua como um fator de antiterminação da polimerase viral. Quando a interação M2-1/P é interrompida, o ciclo replicativo do vírus hRSV fica significativamente comprometido. O principal objetivo do presente estudo foi avaliar computacionalmente a ligação do domínio globular da proteína M2-1 (dgM2-1) à região helicoidal dos resíduos 90-110 da fosfoproteína P (peptídeo P90-110) utilizando simulações de dinâmica molecular e cálculos de energia livre de ligação. Além disso, buscou-se identificar as principais contribuições energéticas de resíduos do P90-110 envolvidos na interação com o dgM2-1 e propor mutações específicas no P90-110 que tinham potencial de aumentar a afinidade desse peptídeo mimético com o dgM2-1. Simulações de dinâmica molecular foram empregadas para avaliar a estabilidade da ligação dos quatro complexos cristalinos dgM2-1/P90-110 e, em seguida, foi empregada uma análise de clusterização para determinar uma estrutura representativa do complexo. A partir dessa estrutura representativa foi realizado um *scan* de alanina nos resíduos da região de alfa-hélice do peptídeo P90-110 para avaliar a contribuição energética para a ligação usando os métodos MM-PBSA, MM-GBSA e Integração Termodinâmica de Crescimento Rápido (ITCR). Através dos cálculos de energia de ligação via MM-PBSA, os resíduos carregados (Lys100, Lys103, Glu104 e Glu107) foram destacados com valores energéticos demasiados para a interação, o que se revelou com um viés do método. Para o método MM-GBSA, resíduos essenciais para a interação dgM2-1/P90-110 como Leu101 foram destacados, mas a incerteza estatística na determinação dos valores de energia mostrou-se como impeditivo para esse método. Já o método de ITCR apontou resíduos do P90-110 fundamentais para ligação ao dgM2-1, Leu101, Thr105, Thr108, Tyr102 e Ile106, que corroboram com resultados experimentais publicados na literatura e destacou os resíduos Phe98, Ser99, e Phe109 como pontos suscetíveis a mutações que poderiam dar origem a peptídeos miméticos do P90-110 com maior afinidade para o dgM2-1. A partir desses pontos suscetíveis de mutação foi proposto o peptídeo mimético FSF98/99/109YEEY que apresentou uma afinidade ($\Delta\Delta G = -17,08$ kJ/mol) e estabilidade conformacional (mais contatos moleculares e ligações de hidrogênio) significativamente maiores para a interação com o dgM2-1 em comparação com o peptídeo P90-110 selvagem. Vale destacar que o maior número de ligações de hidrogênio realizadas pelo peptídeo mimético FSF98/99/109YEEY está diretamente relacionada com a contribuição dos resíduos mutados, o que demonstra uma eficiência da estratégia computacional utilizada empregando o método ITCR. Com esses resultados foi proposto o peptídeo mimético FSF98/99/109YEEY do P90-110 como um possível inibidor/bloqueador da interação M2-1/P, que pode ser testado experimentalmente. Em conclusão, o conhecimento gerado no presente trabalho pode ser útil para futuros estudos sobre as proteínas M2-1 e P e contribuir na busca por antivirais capazes de combater a infecção causada pelo hRSV atuando na interação M2-1/P que é importante para o processo do ciclo de replicação viral.

Palavras-chave: hRSV, M2-1, fosfoproteína P, Dinâmica Molecular, Energia Livre de Ligação.

ABSTRACT

The human Respiratory Syncytial Virus (hRSV) is a pathogen that causes acute respiratory tract infections such as bronchiolitis and pneumonia. This virus has a high potential for contagion, mainly affecting the elderly and children up to 2 years of age. Among the various proteins involved in the replication process of hRSV, we highlight the phosphoprotein P and the M2-1 protein, which interact and facilitate the transcription of all viral proteins. Currently, studies in the literature establish that the interaction of these proteins is essential for the hRSV transcription process, in which M2-1 acts as a viral polymerase antitermination factor. When the M2-1/P interaction is disrupted, the replication cycle of the hRSV virus is significantly compromised. The main objective of the present study is to computationally evaluate the binding of the globular domain of the M2-1 protein (gdM2-1) to the helical region of residues 90-110 of phosphoprotein P (peptide P90-110) using molecular dynamics simulations and free energy calculations. Furthermore, it aims to identify the main energetic contributions of P90-110 residues involved in the interaction with gdM2-1 and propose specific mutations in P90-110 that may increase the affinity of this mimetic peptide with gdM2-1. Molecular dynamics simulations were used to evaluate the binding stability of gdM2-1/P90-110 crystalline complexes and then a clustering analysis was used to determine a representative structure of the complex. From this representative structure, an alanine scan was performed on the P90-110 peptide to evaluate the energetic contribution to binding using MM-PBSA, MM-GBSA and Fast Growth Thermodynamic Integration. Through binding energy calculations via MM-PBSA, the charged residues (Lys100, Lys103, Glu104 and Glu107) were highlighted with energetic values too high for the interaction, which revealed a method bias. For the MM-GBSA method, essential residues for the dgM2-1/P90-110 interaction such as Leu101 were highlighted, but the statistical uncertainty in determining energy values proved to be an impediment to this method. The Thermodynamic Integration method pointed out P90-110 residues that are fundamental for binding to dgM2-1, Leu101, Thr105, Thr108, Tyr102 and Ile106, which corroborate experimental results published in the literature and highlighted residues Phe98, Ser99, and Phe109 as susceptible points to mutations that could give rise to P90-110 mimetic peptides with greater affinity for dgM2-1. From these points susceptible to mutation, the mimetic peptide FSF98/99/109YEEY was proposed, which presented significantly greater affinity ($\Delta\Delta G = -17,08$ kJ/mol) and conformational stability (more molecular contacts and hydrogen bonds) for interaction with dgM2-1 compared to the wild-type P90-110 peptide. It is possible to highlight that the largest number of hydrogen bonds made by the mimetic peptide FSF98/99/109YEEY is directly related to the contribution of mutated residues, which demonstrates the efficiency used using the Thermodynamic Integration method. With these results, the P90-110 mimetic peptide FSF98/99/109YEEY was proposed as a possible inhibitor/blocker of the M2-1/P interaction, which can be tested experimentally. In conclusion, the knowledge generated in the present work can be useful for future studies on the M2-1 and P proteins and contribute to the search for infection caused by hRSV by acting on the M2-1/P interaction, which is important for the process of the viral replication cycle.

Keywords: hRSV, M2-1, phosphoprotein P, Molecular Dynamics, Binding Free Energy.

LISTA DE ILUSTRAÇÕES

Figura 1. Esquema estrutural da partícula viral do hRSV, com a localização das proteínas estruturais desse vírus.....	17
Figura 2. Proteínas e Genoma do hRSV.....	17
Figura 3. Domínio responsável pela oligomerização da fosfoproteína P. A fosfoproteína é representada em representação de <i>Ribbons</i> com cada monômero apresentado de uma cor diferente.	19
Figura 4. Forma monomérica da proteína M2-1. É possível observar os domínios apresentados por cores diferentes, sendo o domínio dedo de zinco de laranja, domínio de oligomerização em magenta, o domínio globular em azul e o domínio C-terminal desestruturado em cinza.....	20
Figura 5. Domínio globular da proteína M2-1. As cores dos números das alfa-hélices correspondem às cores apresentadas na estrutura da proteína.....	21
Figura 6. Modelo hipotético de reconhecimento co-transcricional de RNA mensageiro viral pela M2-1 em associação com o complexo transcricional viral. O nucleocapsídeo (N-RNA) é representado em azul, o RNA em preto e a proteína P em verde. A proteína P liga-se a polimerase viral L (laranja) e a M2-1 (vermelho).	23
Figura 7. Estrutura de cristal do complexo M2-1/P90-110. O monômero da M2-1 no complexo é mostrado a esquerda. O domínio de ligação ao zinco N-terminal é mostrado com um íon de zinco coordenado (esfera vermelha), a hélice de oligomerização é mostrada em rosa, o domínio globular é mostrado em azul e o peptídeo P90-110 é mostrado em laranja. O complexo M2-1/P90-110 com M2-1 em seu estado tetrâmico é apresentado a direita.	24
Figura 8. Ambiente de interação dgM2-1/P90-110 encontrados por Selvaraj e colaboradores.	25
Figura 9. Esquema da resolução das derivadas do Hamiltoniano pelo lambda para equações fora do equilíbrio. Esta forma de encontrar a energia proporciona uma maneira de resolver o ΔG para uma transformação química de um estado A para B. O ponto de intersecção das curvas do histograma formado nos mostra o ΔG obtido.	36
Figura 10. Representação esquemática de um ciclo termodinâmico. Para determinar a alteração na afinidade de ligação após uma mutação peptídica ou modificação do ligante, a diferença entre a energia livre de ligação do ligante ou peptídeo no estado A (selvagem, ΔG_3) e a do estado B (mutante, ΔG_4) precisa ser determinada. Porém, os valores ΔG_1 e ΔG_2 são mais acessíveis aos métodos de perturbação de energia livre e integração termodinâmica de crescimento lento e rápido. Usando a conservação de energia neste ciclo fechado, podemos derivar $\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_4 - \Delta G_3$	39
Figura 11. Configuração de sistema duplo em uma única caixa (<i>double-system/single-box</i>). Em a) tem-se dois tipos de sistema de um ciclo termodinâmico colocados em uma caixa de simulação. As diferentes caixas no esquema são indicadas pelas linhas tracejadas e pontilhadas. Em b) tem-se um exemplo de configuração de simulação para o cálculo da energia livre de ligação do ligante. Durante uma transição, o ligante ligado à enzima está sendo transformado do estado A para B, enquanto a molécula solvatada sofre uma transição reversa do estado B para A. A carga líquida da caixa de simulação permanece zero.....	45
Figura 12. Estrutura da proteína M2-1 complexada com o peptídeo P90-110 da fosfoproteína. A) Estrutura tetramérica de M2-1 (PDB 6G0Y), colorida em verde está a proteína M2-1, com uma região central formado por 4 hélices de oligomerização, uma de cada monômero, em vermelho o fragmento da fosfoproteína e azul o átomo de Zinco. B) As estruturas dos quatro complexos dgM2-1/P90-110, com o domínio colorido em verde e o fragmento P90-110 da	

fosfoproteína em vermelho. Tanto a proteína quanto os peptídeos são representados utilizando o modelo de cartoon e o átomo de zinco denotado pelo modelo de esfera.	46
Figura 13. Conjunto de figuras de RMSD das cadeias em A da interação M2-1/P90-110 e figura de RMSD médio em B. Em azul estão os valores de RMSD para o fragmento todo da fosfoproteína P90-110. Em ciano estão representados os valores de RMSD do domínio globular da M2-1(58-177). Em preto estão os valores de RMSD da porção estruturada em hélices-alfas da M2-1 (75-171) e em vermelho o RMSD para o complexo proteína-peptídeo com foco nas regiões estruturadas da proteína M2-1 e da proteína P.	48
Figura 14. Ligações de hidrogênio para as dinâmicas moleculares entre a M2-1 e a proteína P (90-110) das estruturas das 4 cadeias do tetrâmero. No eixo y tem-se os valores de ligações de hidrogênio e no eixo x os valores do tempo de simulação em nanosegundos. Em A tem-se os resultados para as 4 dinâmicas das estruturas das cadeias da proteína M2-1 e em B os valores médios de ligações de hidrogênio.....	50
Figura 15. Contatos intermoleculares para a interação M2-1/P90-110. Em A temos os valores de contatos intermoleculares no eixo y e o tempo de simulação no eixo x das 4 dinâmicas da proteína M2-1. Em B temos os valores médios de contatos intermoleculares.	51
Figura 16. Valores de energia de ligação (KJ/mol) calculados pelo método PBSA para a interação M2-1/P90-110. Em azul estão os resíduos com contribuição desfavorável à interação e em vermelho os resíduos com contribuição favorável à interação. O gráfico inferior representa a média dos resultados das dinâmicas dos monômeros da interação M2-1/P90-110.	53
Figura 17. Resíduos de aminoácidos em destaque pelo método PBSA para a interação M2-1/P90-110.	55
Figura 18. Valores de energia de ligação (kJ/mol) calculados pelo método GBSA para a interação M2-1/P90-110. Em azul estão representados os resíduos com contribuição desfavorável à interação e em vermelho os resíduos com contribuição favorável à interação. As figuras em A) são os resultados para as réplicas da dinâmica molecular do complexo M2-1/P90-110, já a figura em B) é a média dos resultados encontrados em A).	56
Figura 19. Resíduos destaques do cálculo de energia de ligação por GBSA. Em cinza é mostrado o domínio globular da proteína M2-1, em ciano o fragmento da proteína P90-110. Em azul são mostrados os resíduos que contribuem desfavoravelmente à interação e em vermelho os resíduos que contribuem de forma favorável à interação.	57
Figura 20. Figura da relação da derivada do Hamiltoniano em relação a lambda e o tempo de simulação para o frame 100 da dinâmica molecular do complexo mutado no aminoácido F98. No eixo y temos os valores para a derivada do Hamiltoniano pelo lambda e no eixo x temos o tempo de simulação, somando 50 picosegundos.	62
Figura 21. Demonstração dos passos de dinâmica molecular necessários para resolver o ΔG da mutação do resíduo F98 para alanina.	63
Figura 22. Ciclo termodinâmico da complexação da proteína M2-1 com a P90-110.	64
Figura 23. Resolução do ΔG_1 para o complexo F98 dgM2-1/P90-110 e sua projeção em forma de histograma. O valor de ΔG_1 está destacado como uma linha horizontal tracejada. .	65
Figura 24. Resolução do ΔG_2 para o peptídeo P90-110 e sua projeção em forma de histograma. O valor de ΔG_2 está destacado como uma linha horizontal tracejada.	65
Figura 25. Resíduos favoráveis à interação dgM2-1/P90-110. Em cinza é representado o domínio globular da proteína M2-1. Em ciano é representado a Fosfoproteína P, onde os resíduos favoráveis são representados na figura.	68
Figura 26. Resíduos suscetíveis a mutações planejadas. Em cinza é representado o domínio globular da proteína M2-1. Em ciano é representado a Fosfoproteína P. Os resíduos da proteína P que se destacam são representados em ciano e os resíduos da proteína M2-1 de	

interesse são representados em cinza. A linha preta tracejada denota um contato polar entre Lys 103 e Glu107 do peptídeo P90-110. 70

Figura 27. Conjunto de figuras de RMSD, Contatos moleculares e ligações de hidrogênio. À esquerda estão os dados da dinâmica com o P90-110 selvagem e a direita o do peptídeo mimético FSF98/99/109YEEY. Nas figuras A e B, temos em azul estão os valores de RMSD para o fragmento todo da fosfoproteína P90-110, e em vermelho a região estruturada em alfa-hélice do peptídeo P90-110. Na figura C e D temos os contatos intermoleculares para a interação M2-1/P90-110 do selvagem e do peptídeo mimético FSF98/99/109YEEY. Em E e F são apresentadas as ligações de hidrogênio para as dinâmicas moleculares entre a M2-1 e P90-110 na forma selvagem e mutada FSF98/99/109YEEY. 73

Figura 28. Interação entre o peptídeo mimético FSF98/99/109YEEY e o domínio globular da proteína M2-1. Em ciano é representado o FSF98/99/109YEEY e em cinza o domínio globular da M2-1. A Figura A e B referem-se ao mesmo complexo M2-1/P, apenas mudando o ângulo de observação. 76

LISTA DE TABELAS

Tabela 1. Resultados dos cálculos de energia livre de ligação via MM-PBSA para a interação dgM2-1/peptídeos com mutações dos resíduos de aminoácidos para alanina.....	59
Tabela 2. Resultados dos cálculos de energia livre de ligação via MM-GBSA para a interação dgM2-1/peptídeos com mutações dos resíduos de aminoácidos para alanina.....	60
Tabela 3. Resultados encontrados para o scan de alaninas através do método Integração Termodinâmica de Crescimento Rápido.	66
Tabela 4. Efeitos da mutação pontual da proteína P e seu efeito na interação com a proteína M2-1 encontrados por Mason et al, 2003.	68
Tabela 5. Resultados do cálculo de ITCR para as mutações direcionadas no peptídeo P90-110.	71
Tabela 6. Resultados do cálculo de ligações de hidrogênio entre o domínio globular da proteína M2-1 e o peptídeo P90-110 modificado com a tripla mutação. As ligações de hidrogênio envolvendo os resíduos modificados estão destacadas em negrito, e os nomes destacados com um asterisco indicam um resíduo do P90-110.....	75

LISTA DE ABREVIATURAS E SIGLAS

CFT	Crooks Fluctuation Theorem
dgM2-1	Domínio globular da proteína M2-1
DNA	Desoxyribonucleic Acid
F	Proteína de Fusão
FEP	Técnica de Perturbação da Energia Livre – Free Energy Perturbation
G	Glicoproteína
hRSV	Vírus Sincicial Respiratório humano (hRSV)
ICTV	International Committee on Taxonomy of Viruses
L	Principal subunidade da polimerase
M	Proteína de Matriz
M2-1	Proteína M2-1
M2-2	Proteína M2-2
MM-GBSA	Mecânica molecular de área de superfície generalizada
MM-PBSA	Mecânica molecular de área de superfície Poisson-Boltzmann
N	Nucleoproteína
P	Fosfoproteína
PDB	Banco de dados de proteínas – Protein Data Bank
RNA	Ribonucleic Acid
RNA_m	RNA mensageiro
RMSD	Root Mean Square Deviation - desvio quadrático médio das posições atômicas.
SH	Pequena proteína hidrofóbica

LISTA DE SÍMBOLOS

ΔG	Variação na Energia de Gibbs
$\Delta\Delta G$	Variação do Delta G
λ	Lambda
nm	Nanômetro
ns	Nanosegundo
kJ/mol	Energia por quantidade de substância no sistema
T	Temperatura
K	Kelvin
ps	Picosegundo
μs	Microsegundo

SUMÁRIO

1 INTRODUÇÃO	13
1.1 Epidemiologia.....	14
1.2 Vírus	16
1.3 Características e estrutura do Vírus Respiratório Sincicial Humano (hRSV)	17
1.4 Domínio globular da proteína M2-1	19
1.5 Interação da M2-1 com a proteína P.....	22
1.6 Estratégias de combate ao hRSV	26
1.7 A teoria dos métodos computacionais	28
1.7.1 Dinâmica molecular	28
1.7.2 Cálculo de variação de energia livre de ligação por MM-PBSA e MM-GBSA	31
1.7.3 Cálculo de diferença de energia livre via método de integração termodinâmica de crescimento rápido	33
2. JUSTIFICATIVA	39
3. OBJETIVOS	40
3.1 Objetivo geral	40
3.2 Objetivos específicos	40
4. MÉTODOS.....	41
4.1 Construção do sistema molecular	41
4.2 Dinâmica Molecular	41
4.3 Cálculos de energia livre de ligação via MM-PBSA e MM-GBSA.....	43
4.4 Cálculos de integração termodinâmica de crescimento rápido	43
5. RESULTADOS E DISCUSSÕES	45
5.1 Modelagem	45
5.2 Interação M2-1/P 90-110.....	47
5.3 Análise de MM-PBSA e MM-GBSA.....	52
5.4 Scan de Alaninas por MM-PBSA e MM-GBSA.....	57
5.5 Scan de Alanina por Integração Termodinâmica de Crescimento Rápido	61
5.6 Sugestões de pontos de mutação com base nos cálculos de energia	67
5.7 Mutações direcionadas na região de alfa-hélice do P90-110	70
6 CONCLUSÃO E PERSPECTIVAS	77
REFERÊNCIAS	79
ANEXO A.....	89
ANEXO B.....	101

1 INTRODUÇÃO

Desde os tempos mais remotos da história, a humanidade enfrenta e desenvolve métodos em busca de soluções de problemas, como infecções que afetam o equilíbrio das funções do organismo.

De nômades a desbravadores de horizontes, criando novos ambientes cada vez mais favoráveis para sua sobrevivência, passando a viver em sociedade, as populações humanas cresceram em número e densidade e o homem busca conhecer os limites do seu organismo e como sobreviver a diversidade da natureza.

Como resultado desta conquista, vários contatos interespécies aconteceram e dessa forma diversos tipos de patógenos foram transmitidos através das populações humanas, como bactérias, fungos e vírus. Dentre os agentes causadores de infecções se destacam os vírus, pois sua capacidade metabólica possibilita adaptações genéticas com mais frequência do que acontece em outros organismos, tornando-os em muitos casos, muito bem adaptados ao hospedeiro (SANTOS et al, 2015).

A história da virologia está interligada à dos seres humanos, embora tenha passado despercebido até o século 19, quando novas tecnologias surgiram e foi possível a detecção dos vírus. Os seres vivos podem ser infectados por ao menos algum tipo de vírus, e em muitos casos o ambiente também pode ser impactado diretamente pelos vírus presentes. Graças aos desafios relacionados aos vírus, muitos dos conhecimentos da biologia molecular e biologia celular foram desenvolvidos utilizando-se de estudos relacionados aos vírus (KNIPE et al., 2013).

Atualmente muito se conhece sobre as características moleculares dos vírus, mas devido ao seu tamanho, variando de 20 a 300 nanômetros, somente foram identificados, aproximadamente, 200 anos depois das bactérias (BROCK, 2016).

Ao estudar a doença do mosaico do tabaco, Adolf Mayer se deparou com organismos muito pequenos, que eram indetectáveis nos microscópios da época. A doença podia ser transferida de uma planta para outra através da seiva quando esta era esfregada em uma planta diferente. Contudo, por não conseguir detectar qual microrganismo era responsável, formulou a hipótese de que se tratava de uma bactéria muito pequena (SANTOS, 2023).

Posteriormente, Ivanowski e Martinus Beijerinck, trabalhando separadamente, trouxeram luz aos vírus. Ao tentar filtrar amostras de tabaco para confirmar as hipóteses de Mayer, Ivanowsky não obteve sucesso, chegando à conclusão de que poderiam ser bactérias menores ou que a doença era causada por alguma toxina. Já Beijerinck chegou em uma conclusão contrária à de Ivanowsky, observando que a causa da doença do tabaco não se

multiplicava em meios artificiais, ou seja, em meio de cultura, propondo assim que o causador da doença era uma partícula menor e mais simples. Dessa forma veio a aparecer o conceito do que veio a ser a ideia de vírus (SANTOS, 2023).

Diversas doenças podem causar impactos significativos em humanos, dentre elas as doenças respiratórias virais mostram grandes impactos nas taxas de mortalidade e perdas econômicas, como acontece em pandemias, por exemplo (LEUNG, 2021; BEHZADI, 2019). Os vírus mostram grande adaptação às diferentes formas de contágio, sendo as mais comuns o contato físico e gotículas contaminadas com o vírus, mostrando seu potencial de contaminação global (LEUNG, 2021). Dentre os vírus causadores de doenças respiratórias, os mais comuns são o vírus influenza, vírus sincicial respiratório, vírus parainfluenza, metapneumovírus, rinovírus, bocavírus, adenovírus e coronavírus (BONCRISTIANI et al., 2009). O estudo da biologia destes vírus é essencial para o conhecimento da sua biologia molecular e da sua estrutura, em prol do desenvolvimento de vacinas e medicamentos capazes de combater as doenças por eles causadas.

1.1 Epidemiologia

O Vírus Sincicial Respiratório humano (hRSV) consiste em um patógeno de alto poder de contágio. O vírus se propaga em grande parte pelo ar e, bem como outros vírus respiratórios, está sujeito a sazonalidade dependendo da região em que se apresenta. A depender se a região tem clima tropical ou temperado, fatores como umidade e temperatura podem afetar diretamente a transmissão do vírus, identificando-se os maiores números de casos de contágios nos meses do inverno (YANG et al., 2015; PAYNTER, 2015). Ele é transmitido através do ar na forma de partículas de fluidos do hospedeiro, ou de forma direta pela saliva ou de forma indireta por fluidos depositados em objetos e posteriormente levados à boca ou olhos (COLLINS e GRAHAM, 2008).

Segundo Collins et al. (2008), o hRSV invade o novo hospedeiro onde a defesa é reduzida além de possuir habilidades que atenuam a resposta do hospedeiro, fornecendo evasão da imunidade. Após o contágio, a infecção que começa no trato respiratório superior se espalha para o trato respiratório inferior, onde é mais eficiente (OTHUMPANGAT et al., 2009). Pode acontecer o aumento da quantidade e viscosidade do muco secretado devido a danos causados por necrose e edema da submucosa, resultando no bloqueio parcial do ar nas vias respiratórias (AHERNE et al., 1970). A infecção aguda do trato respiratório inferior pode acontecer com

diversos sintomas diferentes, como pneumonia e bronquiolite, muitas vezes acompanhados de desconforto respiratório, chiado ou tosse (CANEDO-MARROQUIN et al., 2017).

Os grupos mais afetados pelo hRSV são os de idosos e crianças de até 2 anos de idade, sendo os recém-nascidos mais suscetíveis devido ao seu sistema imunológico pouco desenvolvido (MUFSON et al., 1973; DOWELL et al., 1996). Neste grupo de crianças pequenas e idosos, o hRSV é um dos principais responsáveis pela infecção aguda do trato respiratório inferior (ZHANG et al., 2014; OHNO et al., 2013; SHAPIRO et al., 2014) e aproximadamente 2% dos casos requerem hospitalização (NAIR et al., 2010).

Outros fatores também podem ser responsáveis pela ocorrência da doença com sintomas graves, principalmente fatores que interfiram na resistência do organismo a infecções do trato respiratório, como imunodeficiência ou imunossupressão. Nascimento prematuro, ou com baixo peso ao nascer, displasia broncopulmonar, doença cardíaca congênita e fragilidade durante a velhice são alguns outros fatores de risco para a gravidade da doença (COLLINS e GRAHAM, 2008; DEVINCENZO et al., 2010). Acrescenta-se ainda que o nível de replicação do vírus pode estar relacionado à ocorrência de sintomas mais severos, enquanto a replicação do vírus na camada superficial do epitélio respiratório pode retardar e reduzir a exposição do antígeno viral ao sistema imunológico.

Mesmo após a recuperação de uma infecção por hRSV, a imunização não é garantida, sendo possível a reinfecção (HALL et al., 1991; WALSH et al., 2004), isso porque o hRSV tem a capacidade de infectar o mesmo indivíduo diversas vezes durante a vida. É possível que 36% dos indivíduos possam ser reinfetados em uma mesma estação (BOHMWALD et al., 2016). Durante o primeiro ano de vida, a maioria das crianças já foi infectada com o hRSV, já aos 2 anos de idade todas as crianças já devem ter se contaminado com o vírus (COLLINS e MELERO, 2011).

Estudos laboratoriais realizados até o momento, classificam o hRSV em dois grupos antigênicos (A e B), que apresentam diferença no comportamento na proteína G, cuja caracterização da evolução molecular indica uma separação dos grupos A e B cerca de 350 anos atrás a partir de um ancestral comum (JOHNSON et al., 1987; ZLATEVA et al., 2005). Cane (2001) explica que vários genótipos podem coexistir dentro de uma comunidade, com um ou dois genótipos dominantes em anos consecutivos, contribuindo para que as reinfecções ocorram.

1.2 Vírus

Dentre as principais estruturas virais estão o capsídeo, envelope e ácido nucleico. São essas estruturas que formam a partícula viral, que através da invasão a uma célula hospedeira, poderá completar seu ciclo replicativo.

O arranjo de unidades proteicas similares em um arranjo simétrico, e unidos por ligações não covalentes, consiste no capsídeo viral. Essa estrutura mantém a estabilidade do vírus fora da célula hospedeira e contém o material genético viral em seu interior. O capsídeo junto ao material genético viral se denomina nucleocapsídeo (SANTOS et al, 2015).

O envelope é originado das membranas celulares das quais o vírus realizou brotamento, parte da replicação viral na qual acontece a formação do vírus e posteriormente este é ejetado da célula hospedeira.

O material genético dos vírus pode se apresentar em uma única cópia (haploide) ou em duas cópias (diploide). Os genomas virais podem ser constituídos de apenas um segmento ou mais do que um segmento (segmentado) e podem ser de fita dupla, fita simples, circular ou linear. As moléculas de ácidos nucleicos dos vírus podem ser do tipo DNA de fita dupla ou fita simples, ou do tipo RNA de fita dupla ou RNA de fita simples, podendo ser de polaridade positiva ou negativa. As moléculas de RNA com polaridade positiva são sequências genômicas que correspondem ao RNA mensageiro, já as moléculas de RNA com polaridade negativa são complementares ao RNA mensageiro (SANTOS et al, 2015).

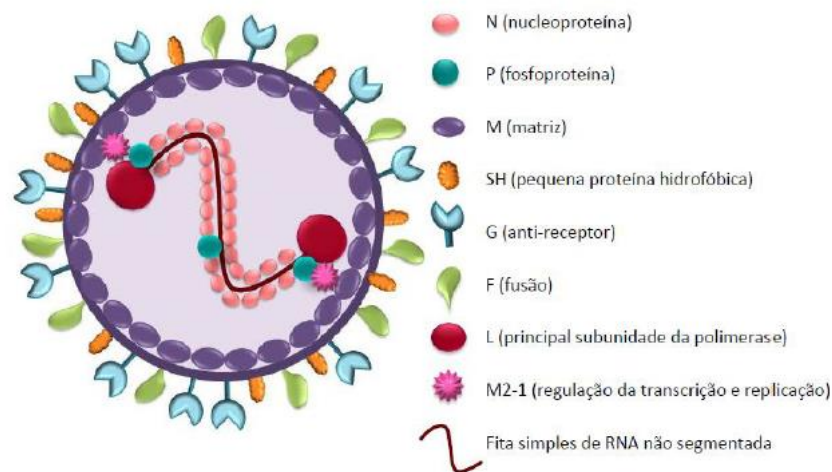
O processo replicativo inicialmente acontece na célula hospedeira, com o agrupamento do genoma viral e das proteínas estruturais em uma partícula nova, que é liberada para fora da célula pelo rompimento da parede celular. O contato desta partícula com uma nova célula, inicia o processo de adesão e fusão que culminará no processo de replicação do genoma viral na nova célula hospedeira (KNIPE et al., 2013).

Após a adesão do vírus na célula alvo, por ancoragem da glicoproteína (G), e da fusão entre o vírus e a célula alvo, pela proteína de fusão (F), é iniciado processo de replicação do genoma viral pela transcrição dos RNA mensageiros (RNAm) virais (SANTOS et al, 2015). Ao todo, o processo pode variar bastante dependendo do vírus em questão, mas envolve o contato entre o vírus e a célula hospedeira, a adsorção, a penetração na célula hospedeira, que pode ocorrer pelo englobamento do vírus ou pela injeção do material genético dentro da célula hospedeira, o processo de transcrição e replicação, o processo de montagem ou morfogênese, e o processo de liberação, no qual os novos vírus possivelmente infectaram novas células no organismo (SANTOS 2023).

1.3 Características e estrutura do Vírus Respiratório Sincicial Humano (hRSV)

O Vírus Respiratório Sincicial humano (hRSV) é um patógeno respiratório que causa infecções agudas do trato respiratório. Pelas normas do *International Committee on Taxonomy of Viruses* (ICTV), o hRSV, com esquema representado na Figura 1, se encontra dentro da ordem *Mononegavirales*, família *Pneumoviridae*, gênero *Orthopneumovirus* (BUKREYEV et al., 2016), caracterizando-se como uma molécula envelopada, cujo genoma consiste em uma fita simples de RNA não segmentada e com polaridade negativa (COLLINS e MELERO, 2011).

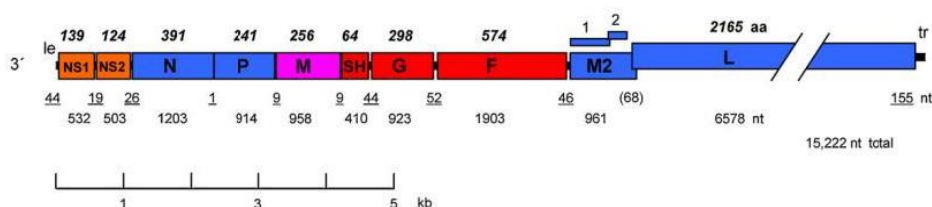
Figura 1. Esquema estrutural da partícula viral do hRSV, com a localização das proteínas estruturais desse vírus.



Fonte: DIAS (2017).

O genoma do hRSV compreende cerca de 15 kb, no qual existem instruções de fabricação de 11 proteínas através de 10 genes, como apresentado pela Figura 2.

Figura 2. Proteínas e Genoma do hRSV.



Fonte: COLLINS E MELERO (2011).

Das 11 proteínas, há as proteínas não estruturais (NS1 e NS2) que são responsáveis pela evasão da resposta imunológica. Existem também as proteínas responsáveis pela replicação viral, sendo elas N (nucleoproteína), P (fosfoproteína) e L (principal subunidade da polimerase), e as proteínas de regulação da transcrição e replicação, M2-1 e M2-2. Existem também as

proteínas relacionadas à membrana viral, que são SH (pequena proteína hidrofóbica), M (matriz), G (anti-receptor ou adesão) e F (fusão) (COLLINS et al., 2001).

As proteínas G e F são glicoproteínas e estão relacionadas com a fusão e adesão do vírus à célula hospedeira (PETERS e CROWE, 2008). A proteína G é responsável pela ancoragem do vírus à célula hospedeira (BOHMWALD et al., 2016) e a proteína F tem como papel a penetração viral e a fusão entre células infectadas para a formação de sincícios, que são característicos em sua infecção (COLLINS e MELERO, 2011). A proteína SH, por sua vez, pertence à família de proteínas virais pequenas e são altamente hidrofóbicas. Elas são capazes de formar canais iônicos em membranas celulares (GAN et al., 2008) e possuem efeito anti-apoptótico que promove a replicação viral (FUENTES et al., 2007).

A proteína de matriz (M) possui um papel estrutural no hRSV, e importante papel durante a montagem do vírus, se posicionando entre o nucleocapsídeo viral e o envelope viral (MELERO, 2007; RODRÍGUEZ et al., 2004). É a proteína de matriz que interrompe a replicação viral ao final do processo de transcrição, no qual acontece o brotamento (MITRA et al., 2012). Entre a proteína M e o complexo ribonucleoprotéico encontram-se as proteínas M2-1 e M2-2, provenientes do mesmo gene (KISS, G. et al., 2014). Já o complexo ribonucleoprotéico é feito da união do RNA com a nucleoproteína N, e está ligado à proteína P, à proteína M2-1 e à proteína L. Este conjunto de biomoléculas é o responsável pela replicação do vírus (SHAPIRO et al., 2014).

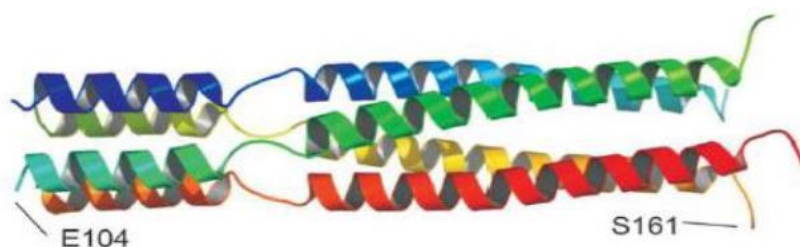
Quando ocorre a invasão do vírus na célula hospedeira, durante a replicação viral, aparecem estruturas que previamente não se encontravam na célula, chamadas de corpúsculos de inclusão, onde pode-se concluir por evidências encontradas, que as proteínas N, P, M2-1 e L estão presentes, participando de processos de transcrição e replicação (GARCIA et al., 1993; MCDONALD et al., 2004; BROWN et al., 2005; CARROMEU, 2007). Nesse processo, a função de polimerase é realizada pelas proteínas P e L.

Dentre as fosfoproteínas do gênero paramixovírus, a proveniente do hRSV é a mais curta, contendo 241 aminoácidos, e apresenta diversos pontos de fosforilações em serinas na sua cadeia (NAVARRO et al., 1991). Ao interagir com diversas proteínas responsáveis pela replicação do gene viral, a proteína P impede que a proteína N se agregue ou se conecte a moléculas de RNAs da célula hospedeira, atuando como chaperona e garantindo a especificidade do conjunto de replicação do RNA viral (CURRAN et al., 1995; SPEHNER et al., 1997).

A proteína P atua de forma a conectar as proteínas L e M2-1 ao nucleocapsídeo, ajudando a polimerase a se ligar no RNA (FIELDS, 2001). Ela é encontrada na forma de

tetrâmero onde é possível encontrá-la em sua forma habitualmente desordenada, com algumas estruturas estruturadas como a hélice de oligomerização (KARLIN et al., 2003). Além disso, a proteína P do hRSV contém uma região estruturada em alfa-hélice entre os aminoácidos 104 a 161, sendo esse o domínio responsável pela oligomerização, o qual está representado na Figura 3 (LLORENTE et al., 2008).

Figura 3. Domínio responsável pela oligomerização da fosfoproteína P. A fosfoproteína é representada em representação de *Ribons* com cada monômero apresentado de uma cor diferente.



Fonte: DIAS (2017).

O desenvolvimento de compostos antivirais específicos requer o entendimento o entendimento molecular das proteínas virais. A fosfoproteína P se apresenta como uma proteína importante, conectando a proteína L ao nucleocapsídeo (COLLINS et al., 2001), assim como realiza interações com a proteína M2-1, onde os resíduos L101, Y102 e F109 são necessários para a interação M2-1/P acontecer (MASON et al., 2003).

A proteína P é fosforilada em diversos pontos específicos, e essas fosforilações podem ter um papel crucial nas interações da proteína P no contexto da replicação do vírus hRSV (COLLINS et al., 2001). A fosforilação da Thr108 possivelmente controla a interação da proteína P com a M2-1, modulando dessa forma a atividade desse fator de antiparada transcricional na replicação viral (CARTEE et al., 2003). Segundo Asenjo e colaboradores (2006), os resíduos Thr105 e Thr108 da fosfoproteína estabelece contato com a proteína M2-1, sendo que a fosforilação da Thr108 atua como um regulador dessa interação. Dessa forma, o estado desfosforilado da Thr108 na proteína P é necessário para a atividade da proteína M2-1 no processo de transcrição viral.

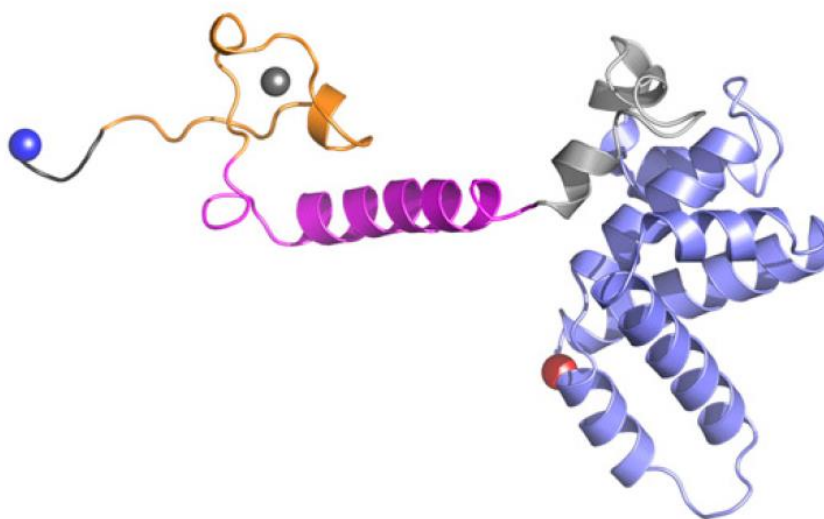
1.4 Domínio globular da proteína M2-1

A proteína M2-1 se apresenta no contexto da replicação viral como uma proteína bem versátil por apresentar a capacidade de se ligar tanto no RNA viral como na fosfoproteína P, atuando também como um fator de antiterminação (HARDY et al., 2000), proporcionando ao

RNA viral produzido a forma adequada, como também como uma proteína estrutural no virion (KISS et al., 2014; LILJEROOS et al., 2013). Uma de suas atividades de destaque é sua atividade de antiterminação ou antiparada no processo de transcrição, para o qual a proteína M2-1 atua como um fator de processividade da polimerase viral, evitando que essa se dissocie ao chegar ao códon de parada e, portanto, aumentando a eficiência de transcrição dos genes próximos à extremidade 5'. Na ausência da proteína M2-1 a atividade de replicação do RNA viral é significativamente reduzida para o hRSV (BLONDOT et al, 2012).

A proteína M2-1 forma um tetrâmero que apresenta 194 resíduos de aminoácidos em sua estrutura monomérica (Figura 4). É possível identificar quatro domínios principais em sua estrutura terciária: o domínio dedo de zinco N-terminal, que possui uma região de ligação ao zinco; o domínio de oligomerização em alfa-hélice, responsável pelas interações e estruturação da forma tetramérica; o domínio globular capaz de realizar interações com a proteína P e com o RNA; e o domínio C-terminal não estruturado (TANG et al., 2001; TANNER et al., 2014).

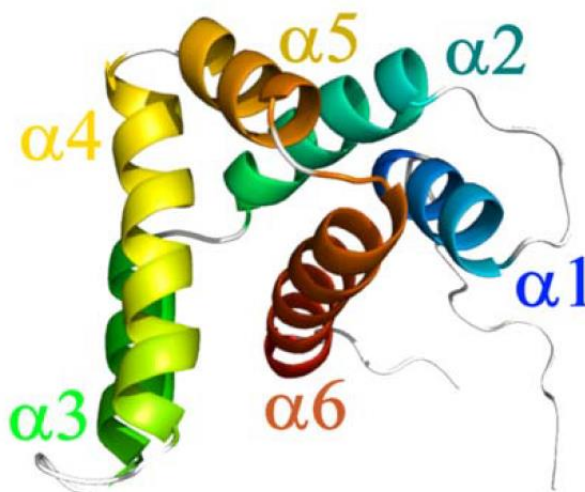
Figura 4. Forma monomérica da proteína M2-1. É possível observar os domínios apresentados por cores diferentes, sendo o domínio dedo de zinco de laranja, domínio de oligomerização em magenta, o domínio globular em azul e o domínio C-terminal desestruturado em cinza.



Fonte: Adaptado de TANNER et al., 2014.

O domínio globular na proteína M2-1 abrange os resíduos G75-I171 (Figura 5), que são estruturados em seis alfa-hélices: $\alpha 1$ (G75-G85), $\alpha 2$ (K92-E105), $\alpha 3$ (S108-D117), $\alpha 4$ (K124-K140), $\alpha 5$ (K143-R151) e $\alpha 6$ (D155-I171), sendo por isso também chamado por alguns autores de domínio α -helicoidal (BLONDOT et al, 2012).

Figura 5. Domínio globular da proteína M2-1. As cores dos números das alfa-hélices correspondem às cores apresentadas na estrutura da proteína.



Fonte: Adaptado de BLONDOT et al, 2012.

O domínio globular da proteína M2-1 (dgM2-1) está relacionado às interações intermoleculares presentes no complexo de replicação de RNA, encontrando-se evidências de interação da proteína M2-1 com RNA e com a fosfoproteína através desse domínio, no qual foram identificados resíduos que contribuem para duas superfícies de contato adjacentes, e parcialmente sobrepostas, com a proteína P e com o RNA. Isso significa que a fosfoproteína liga-se a uma região proximal à superfície de ancoragem do RNA e sua ligação específica parece estar envolvida com as hélices $\alpha 4$ e $\alpha 2$, assim como com as hélices $\alpha 5$ e $\alpha 6$, que contribuem nas interações (BLONDOT et al, 2012).

Mutações em resíduos responsáveis pelas interações intermoleculares no domínio globular da M2-1, podem interromper especificamente as interações do domínio com o RNA e com a proteína P e têm um efeito negativo na co-localização intracelular de M2-1 com a proteína P, bem como na sua função como cofator de transcrição. O dgM2-1 apresenta grande afinidade por fitas de RNA senso negativo de polaridade negativa, rica em uracilas e por RNA senso positivo, rico em purinas, especialmente a adenina, sendo que trechos de RNA ricos em adenina aumentam mais a afinidade da fita com o domínio do que trechos ricos em uracila. (BLONDOT et al.,2012; TANNER et al., 2014).

A importância do dgM2-1 para a atuação da proteína M2-1 o torna alvo de estudos nos quais há a busca de moléculas que tenham a capacidade de interferir nas interações intermoleculares que ele desempenha com seus parceiros biológicos, competindo pelos sítios

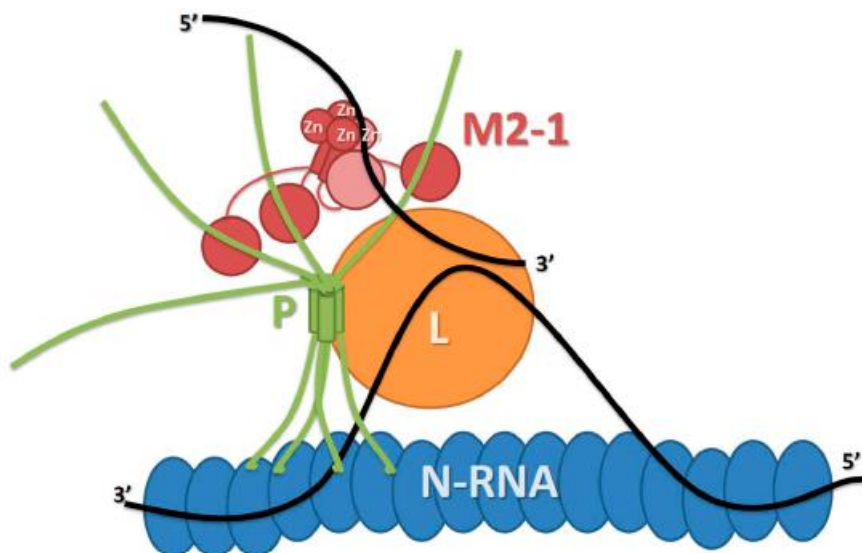
de ligação em detrimento da proteína P e do RNA, para afetar o complexo de transcrição do RNA viral.

1.5 Interação da M2-1 com a proteína P

A proteína M2-1 está presente no complexo de replicação do vírus, sendo um cofator essencial atuando no processamento da transcrição, nesse caso atuando em conjunto com a proteína P e o RNA viral (BLONDOT, M-L. et al., 2012). A proteína M2-1 atua em conjunto com a proteína P no complexo RNP, impedindo que o processo de transcrição se interrompa. Por meio da atividade da proteína M2-1, o processo de replicação viral pode acontecer sem interrupção, pois ela atua como um fator de anti-terminação da transcrição (TANG et al., 2001). Vale lembrar que a fosfoproteína P é um fator determinante para o sucesso da replicação do vírus, pois viabiliza a interação entre a proteína L (polimerase viral) e o complexo entre o RNA e a nucleoproteína N (FIZ et al., 2011).

A proteína M2-1 forma um tetrâmero possuindo 194 aminoácidos em sua estrutura monomérica. É possível identificar 4 domínios principais em sua estrutura: O domínio N-terminal que possui uma região de ligação ao zinco, um domínio de oligomerização em alfa-hélice, responsável pelas interações e formação da forma tetramérica, um domínio globular capaz de realizar interações com a proteína P e com o RNA, e um domínio C-terminal não estruturado (TANG et al., 2001; TANNER et al., 2014). A proteína M2-1 realiza interações com o complexo de replicação viral por meio da interação com o RNA e interage com a proteína P quando o resíduo Thr108 está desfosforilado (ASENJO et al., 2006; ASENJO et al., 2012).

Figura 6. Modelo hipotético de reconhecimento co-transcricional de RNA mensageiro viral pela M2-1 em associação com o complexo transcricional viral. O nucleocapsídeo (N-RNA) é representado em azul, o RNA em preto e a proteína P em verde. A proteína P liga-se a polimerase viral L (laranja) e a M2-1 (vermelho).



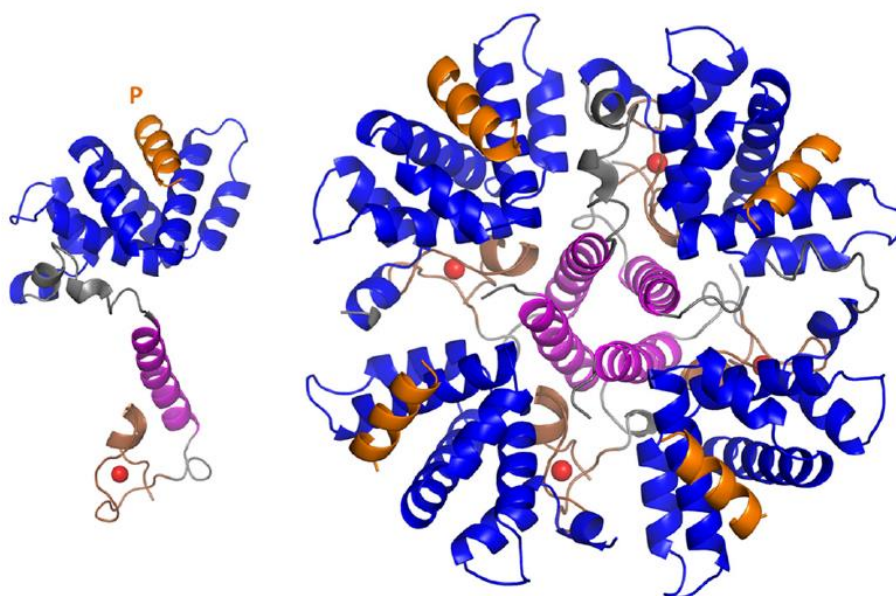
Fonte: adaptado de LEYRAT et al., 2014.

Na Figura 6 podemos ver um modelo hipotético do complexo de transcrição viral para o hRSV proposto por Leyrat e colaboradores (2014). Nesse modelo, a proteína P está ligada à proteína M2-1 e atua como um elo, conectando o nucleocapsídeo (nucleoproteína mais RNA) à proteína L e a M2-1. O modelo proposto apresenta a proteína M2-1 associada a proteína P em um complexo tetrâmero/tetrâmero não globular (ESPERANTE et al., 2012). A proteína P se posiciona de forma a conectar o nucleocapsídeo à proteína L e à M2-1, se apresentando dessa forma como uma molécula chave para as interações intermoleculares na replicação viral. O RNA mensageiro nascente é estabilizado pela interação com a proteína M2-1, prevenindo que ele se reibridize com a fita molde.

Como mencionado antes, o domínio globular da M2-1 está relacionado às interações intermoleculares presentes no complexo de replicação de RNA, sendo encontradas evidências de sua interação com RNA e com a proteína P (BLONDOT et al, 2012). Recentemente, foi determinada a estrutura de cristal da proteína M2-1 complexada com o domínio mínimo de interação da proteína P, resíduos de aminoácidos de 90 a 110 (P90-110). Identificou-se uma interação intermolecular relativamente fraca para o complexo M2-1/P quando comparada com a interação que a M2-1 pode estabelecer com moléculas de RNAs (SELVARAJ et al., 2018).

A Figura 7 mostra que o peptídeo P90-110 forma uma estrutura em α -hélice que se liga ao longo de uma fenda entre três α -hélices no dgM2-1. O peptídeo P90-110 se liga em uma posição na qual encontram-se resíduos da M2-1 previamente identificados como importantes para a interação com a P e o RNA. Um total de 15 dos 16 resíduos da M2-1 são conservados e interagem com a P90-110 através de interações de ligação iônica, hidrofóbica e ligação de hidrogênio (SELVARAJ et al., 2018).

Figura 7. Estrutura de cristal do complexo M2-1/P90-110. O monômero da M2-1 no complexo é mostrado a esquerda. O domínio de ligação ao zinco N-terminal é mostrado com um íon de zinco coordenado (esfera vermelha), a hélice de oligomerização é mostrada em rosa, o domínio globular é mostrado em azul e o peptídeo P90-110 é mostrado em laranja. O complexo M2-1/P90-110 com M2-1 em seu estado tetrâmero é apresentado a direita.



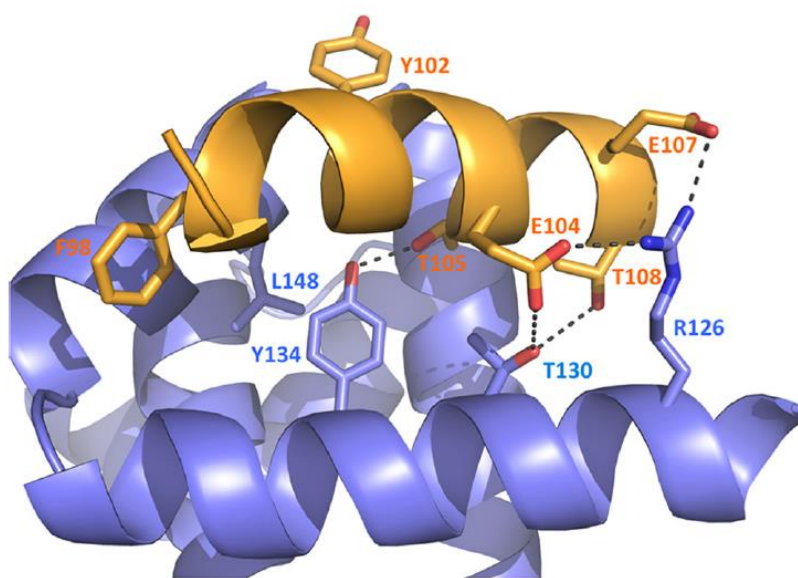
Fonte: Adaptado de SELVARAJ et al., 2018.

O reconhecimento molecular do RNA e P90-110 pelo dgM2-1 é fundamental para o ciclo de replicação do hRSV. Os pontos específicos da interação dgM2-1/RNA e dgM2-1/P90-110 podem ser considerados como potenciais sítios de ligação de antivirais, uma vez que a oclusão desses sítios poderia promover um bloqueio da atividade desempenhada pela M2-1 no contexto do complexo co-transcricional viral. Um exemplo dessa estratégia antiviral foi reportado por Bailly et al. (2016) que demonstrou a atuação do alcaloide esteroide ciclopamina como um inibidor potente e seletivo da replicação *in vivo* e *in vitro* do hRSV. Nesse estudo, os autores propuseram um efeito específico da ciclopamina na ligação do dgM2-1 ao RNA e a P, o que resultaria na exclusão do fator de antiterminação M2-1 do complexo co-transcricional viral (BAILLY et al., 2016). Portanto, pode-se destacar que o dgM2-1 é um promissor alvo para

o desenvolvimento de antivirais que objetivam inibir o ciclo de replicação do hRSV. Nesse contexto, a busca por ligantes como potenciais candidatos antivirais para o dgM2-1 poderia ser realizada a partir de uma abordagem sistematizada como o *design* racional de miméticos de peptídeos.

A região de interação M2-1/P apresenta um ambiente químico complexo, apresentando interações iônicas, hidrofóbicas e ligações de hidrogênio. Os resultados apresentados por Selvaraj e colaboradores, apresentam o fragmento P90-110 posicionado de forma que os resíduos de aminoácidos hidrofóbicos se encontram voltados para a fenda no domínio globular da proteína M2-1, como os resíduos P97, F98, L101 e I106. Enquanto os resíduos de aminoácidos hidrofílicos ficam mais expostos ao solvente, como os resíduos K100, K103, E104 e E107. Na Figura 8 é possível observar o fragmento de fosfoproteína P90-110 onde estão destacados os resíduos estudados por Selvaraj e colaboradores. As diversas ligações de hidrogênio são mostradas entre os resíduos destacados e aminoácidos da proteína M2-1 próximos a eles, como os aminoácidos E104 e E107 da fosfoproteína P fazendo interações iônicas com uma arginina da proteína M2-1.

Figura 8. Ambiente de interação dgM2-1/P90-110 encontrados por Selvaraj e colaboradores.



Fonte: Adaptado de SELVARAJ et al., 2018.

1.6 Estratégias de combate ao hRSV

O vírus hRSV afeta milhares de pessoas todos os anos, promovendo inúmeras hospitalizações e mortes. Uma das causas é a capacidade do vírus de reinfectar o hospedeiro mesmo com a presença ou desenvolvimento de imunidade contra ele. Isso pode ser agravado ainda mais nos casos em que há falha no desenvolvimento de resposta efetiva pelo sistema imunológico, o que pode resultar em sucessivas reinfecções (GLEZEN et al., 1986).

Proteínas não estruturais do vírus hRSV (NS1 e NS2) têm potencial de inibir a indução e sinalização de interferon, interferindo na resposta imune inata e atuam no retardo do processo de apoptose de células infectadas, propiciando que uma maior carga viral seja liberada posteriormente (BITKO et al. 2007; Spann et al. 2005, SEDEYN et al. 2019).

As glicoproteínas de superfície G e F são os principais alvos dos anticorpos neutralizantes, sendo os principais alvos para a pesquisa e desenvolvimento de vacinas e fármacos. No entanto, as mutações nessas proteínas prejudicam a especificidade deles. A alta variabilidade, diversidade genética e a falta de estabelecimento de critérios para a definição de genótipo viral, dificultam o entendimento da evolução do vírus, a atuação do sistema imunológico e consequentemente o desenvolvimento de vacinas (GRAHAM et al. 2015).

Na década de 60, baseando-se em testes para outros vírus, houve a primeira tentativa no desenvolvimento de uma vacina para hRSV, a qual falhou nos testes clínicos em humanos. Apesar de se apresentar imunogênica, quando os indivíduos imunizados contraíam o vírus posteriormente apresentavam um quadro muito mais grave da doença, em comparação com indivíduos não imunizados (KIM et al., 1969).

Em 1996 foi licenciada uma imunoglobulina policlonal hiperimune obtida de doadores que apresentavam grande taxa de anticorpos neutralizantes para hRSV. No entanto foi retirada do mercado oito anos depois, pois sua utilização resultava na sobrecarga de líquidos devido a necessidade de sessões de infusão intravenosa muito longas, e trazia desvantagens, como o adiamento da vacinação com a tríple viral e de outras vacinas baseadas em vírus atenuados (ISPAS et al., 2015).

Atualmente, o anticorpo monoclonal palivizumabe, que atua na proteína F inibindo a fusão do vírus à célula do hospedeiro, é utilizado como profilaxia em diversos países no mundo. Aprovado em 1998 pelo *Food and Drug Administration* (FDA) sua incorporação ao Sistema Único de Saúde (SUS) brasileiro foi aprovada em 2012 com protocolo de uso publicado em maio de 2013. A utilização deste medicamento é realizada nos períodos de sazonalidade do hRSV como prevenção de infecção em crianças em risco. (BRASIL, 2022).

Já para o tratamento é utilizado a ribavirina, o único medicamento licenciado para o tratamento de infecções por hRSV, sendo um agente viral de amplo espectro que atua na inibição da replicação viral pela inibição da síntese de proteínas (BORCHERS et al. 2013). Por causa de seu potencial tóxico, efeitos teratogênicos e alto custo terapêutico, devido a necessidade de uso prolongado, é geralmente utilizado em pacientes com evolução mais grave da doença, como os imunocomprometidos (MACHADO et al., 2009).

Não há um medicamento específico contra o hRSV e o cenário de desenvolvimento de vacinas é restrito. Recentemente, a agência federal de Administração de Alimentos e Medicamentos dos EUA (*USA Food and Drug Administration* - FDA) aprovou a vacina da GSK para pessoas com 60 anos ou mais contra o hRSV, que ainda não se encontra disponível no Brasil. Há também o desenvolvimento avançado de uma vacina contra o hRSV pela empresa de biotecnologia Novavax, a ResVax. Essa vacina, composta por nanopartículas recombinantes da proteína F e tem como adjuvante fosfato de alumínio, visa a proteção de bebês por meio da imunização materna. Em 2019 foram apresentados os resultados da fase três do ensaio clínico para a ResVax, o qual foi realizado em 4600 mulheres grávidas. (NOVAVAX, 2019).

A busca de compostos naturais que se ligam à proteína M2-1 do hRSV, podendo atuar como potenciais inibidores de sua atuação como fator de antiterminação transcricional, é uma estratégia promissora que nosso grupo de pesquisa desenvolve desde 2012 (GUIMARÃES et al., 2018; PIVA et al., 2020; TEIXEIRA et al., 2017). No entanto, atualmente estudamos novas estratégias para identificar ligantes para a M2-1, mais especificamente para seu domínio globular, que poderiam atuar como potenciais antivirais inibindo sua atividade no processo de transcrição do hRSV.

A estratégia apresentada neste trabalho é o desenho racional de miméticos de peptídeos baseados em interações proteína-proteína (IPP) que desempenham papel fundamental em eventos moleculares de sistemas biológicos (WANG et al., 2021). O desenho de miméticos de peptídeos de interfaces de IPP pode ser classificado como uma abordagem computacional de *design* de fármaco baseado em estrutura (PELAY-GIMENO et al., 2015), que é usualmente combinada com uma abordagem experimental envolvendo técnicas biofísicas para caracterização da interação proteína alvo-peptídeo. No geral, esse *design* racional começa com a seleção de resíduos de aminoácido chave da IPP como modelos/referências. Em seguida, por meio de modificações racionais dos resíduos, as estruturas iniciais dos modelos são otimizadas computacionalmente para garantir que os miméticos criados sejam capazes de competir (maior afinidade de ligação) pela proteína alvo com o peptídeo originalmente reconhecido na interface de IPP (SHIN et al., 2017). Ao final, os miméticos são avaliados experimentalmente quanto à

atividade de inibição e/ou ligação à proteína alvo e podem sofrer otimizações adicionais. O *design* de miméticos de peptídeos de interfaces de IPP é considerado uma abordagem eminentemente prática e promissora para descobrir inibidores (WANG et al., 2021). Um exemplo da aplicação da estratégia mencionada acima é o Nelfinavir, um medicamento para o tratamento de doenças infecciosas do vírus da imunodeficiência humana (WLODAWER, 2002). Nesse sentido, a busca por peptídeos como potenciais candidatos antivirais para o dgM2-1 do hRSV poderia ser realizada a partir da abordagem sistematizada de *design* racional de miméticos do peptídeo P90–110 baseada na interação M2-1/P. A identificação de peptídeos como ligantes ao dgM2-1 poderia fomentar as bases moleculares para o desenvolvimento de uma nova estratégia de combate as infecções causadas pelo hRSV, tendo como via a inibição/bloqueio do processo de replicação viral.

1.7 A teoria dos métodos computacionais

1.7.1 Dinâmica molecular

As simulações computacionais consistem em aproveitar a capacidade de processamento de computadores para solucionar problemas relacionados ao trabalho ou pesquisas científicas. O uso de simulações para elucidar o comportamento químico tem sido muito utilizado para entender as propriedades moleculares dos sistemas biológicos (MANOEL, 2016).

O planejamento de fármacos assistidos por métodos computacionais já é uma realidade em muitas indústrias, e a tendência é que sua participação na descoberta de fármacos aumente cada vez mais (VIVO et al., 2016; WARR, 2017). Nesse contexto, a busca por prever a energia de ligação de uma molécula envolvida na interação com um alvo biomolecular é de especial interesse. Uma vez selecionado o alvo biológico, pequenas moléculas orgânicas começam a ser selecionadas para testes e começa em geral uma campanha de triagem virtual e experimental a fim de encontrar o melhor protótipo de ligante possível (BROWN, 2007).

Os sistemas biológicos podem ser estudados com a utilização de técnicas computacionais como a simulação de dinâmica molecular, que é bem eficiente em descrever sistemas moleculares. Para utilizar a técnica de dinâmica molecular é preciso saber como as partículas vão interagir umas com as outras e as equações que descrevem seu movimento através do tempo. Pelo potencial de interação e das posições iniciais é possível encontrar as posições atômicas ao decorrer do tempo, encontrando assim informações como propriedades físico-químicas do sistema (COUTINHO et al., 2007).

Nas simulações de dinâmica molecular os átomos são tratados como esferas pontuais, obedecendo às leis clássicas da mecânica newtoniana, como as forças elásticas e harmônicas. Juntando todas as forças que atuam sobre essas esferas temos o campo de força, que basicamente descreve como uma esfera interage com esferas vizinhas (BROOKS et al., 1998.; GUNSTEREN et al., 1992).

Na busca de representar os experimentos científicos por meio de modelos computacionais, que possivelmente descrevem propriedades físicas das moléculas e compostos químicos presentes no meio modelado, são usadas as leis da física clássica. Aplicando as leis clássicas e a análise temporal das moléculas, torna-se possível a determinação no espaço experimental das moléculas do sistema.

A segunda lei de Newton é a base da Dinâmica Molecular, a qual descreve as partículas baseado em suas coordenadas espaciais e um campo de força, diferentemente do que acontece nos métodos baseados em física quântica (VERLI, 2014; BERENDSEN et al., 1995). Dessa forma a equação da segunda lei de Newton relaciona a força que atua em uma partícula com sua aceleração, ou seja, velocidade e sua mudança no decorrer do tempo (LOURENÇO, 2019):

$$F_{x_i} = \frac{d^2_{x_i}}{dt^2} m_i = \frac{dv_i}{dt} m_i = a_i m_i \quad (1)$$

sendo que F_{x_i} é a força da partícula i de massa m_i na coordenada x_i , enquanto v_i e a_i é a velocidade de aceleração dessa partícula, respectivamente.

Pelo cálculo de dinâmica molecular os comportamentos moleculares são simulados. Isso é possível pois um conjunto de funções e parâmetros são usados para interpretar como as moléculas interagem entre si. O campo de força é a forma como as ligações entre os átomos são interpretadas, normalmente como molas elásticas para reproduzir as características experimentais (VERLI, 2014).

A determinação dos potenciais de interação constitui uma importante etapa na construção da simulação computacional. É importante resolver o potencial total, que é a soma dos potenciais de interação intramolecular (interações ligadas) e intermolecular (interações não-ligadas). Após a resolução de cada parte da equação de potencial temos a equação a seguir (COUTINHO et al., 2007):

$$V_{total} = V_{intra} + V_{inter} = \sum_{ligações} \frac{1}{2} K_r (r - r_0)^2 + \sum_{ângulos} \frac{1}{2} K_\theta (\theta - \theta_0)^2 + \sum_{n,diedros} \frac{1}{2} V_n [1 + \cos(n\phi - \delta)] + \sum_{i,j} 4\varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i,j} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2)$$

Essa equação é a soma das partes das interações potenciais intramolecular e intermolecular. Na primeira parte temos a soma dos termos dos potenciais de interação ligada, sendo o potencial de distância de ligação, de ângulo de ligação e de ângulo de diedro. Em seguida são mostrados os termos referentes ao potencial de interação não ligada, sendo o potencial de Lennard-Jones e o de Coulomb. Cada um dos termos será apresentado a seguir.

Para entender as forças que atuam no sistema, em suas partículas, é necessário entender a escolha dos potenciais de interação intramolecular (V_{intra}) e intermolecular (V_{inter}). Nos sistemas moleculares, as interações são descritas como uma soma dos potenciais das interações que atuam nas partículas individualmente (PILOTO, 2022).

Deformações moleculares nos informam o grau de mobilidade interna de moléculas no sistema, sendo representadas pelas interações intramoleculares.

$$V_{intra} = \sum V_{ligação} + \sum V_{angular} + \sum V_{torção} \quad (3)$$

sendo que o $V_{ligação}$ nos informa o estiramento de uma ligação química e $V_{angular}$ e $V_{torção}$ descrevem as deformações angulares e deformações dos ângulos diedros, respectivamente. Por meio de potenciais harmônicos temos a definição de estiramento e deformação angular, como observado nas fórmulas a seguir:

$$V_{ligação}(r) = \frac{1}{2} K_r (r - r_0)^2 \quad (4)$$

$$V_{angular}(\theta) = \frac{1}{2} K_\theta (\theta - \theta_0)^2 \quad (5)$$

sendo que r_0 e θ_0 descrevem o comprimento e ângulo de equilíbrio da ligação, e os valores de K_r e K_θ dizem respeito às constantes de força (PILOTO, 2022). O potencial de torção é representado pela seguinte equação:

$$V_{torção}(\phi) = \sum_n \frac{1}{2} V_n [1 + \cos(n\phi - \delta)] \quad (6)$$

sendo que V_n é o potencial de energia para a torção e n indica o número máximo/mínimo de energia em uma torção completa. Os ângulos ϕ e δ correspondem ao ângulo de diedro e de fase, respectivamente.

A soma das interações não ligadas de pares de átomos a e b nas moléculas dado pelo potencial de Lennard-Jones somado ao termo de interações eletrostáticas do potencial de Coulomb forma um modelo de potencial intermolecular entre moléculas i e j .

$$V_{inter} = \sum_{a \in i} \sum_{b \in j} V_{ab}(r_{ab}) \quad (7)$$

$$V_{ab}(r_{ab}) = 4\varepsilon_{ab} \left[\left(\frac{\sigma_{ab}}{r_{ab}} \right)^{12} - \left(\frac{\sigma_{ab}}{r_{ab}} \right)^6 \right] + \frac{q_a q_b}{4\pi\epsilon_0 r_{ab}} \quad (8)$$

sendo que ε_{ab} indica a profundidade entre a barreira atrativa e repulsiva do potencial, σ_{ab} é a distância em que o potencial da interpartícula é igual a zero, conhecido como diâmetro de Lennard-Jones, e r_{ab} é a distância entre os núcleos atômicos. Os parâmetros q_a e q_b descrevem as cargas parciais. O primeiro termo r_{ab}^{-12} determina a repulsão a curtas distâncias devido a sobreposição dos raios de van der Waals, o segundo termo r_{ab}^{-6} descreve a contribuição dispersiva de London, e o último termo r_{ab}^{-1} se refere às interações eletrostáticas (PILOTO, 2022).

1.7.2 Cálculo de variação de energia livre de ligação por MM-PBSA e MM-GBSA

Em um sistema biológico simulado através de métodos computacionais, é possível estimar um valor para a energia livre de ligação entre uma proteína e ligante, e entre duas proteínas também (ALVAREZ, 2017). Conhecer interações intermoleculares contribuem para o entendimento dos sítios ligação e catalíticos, e métodos computacionais como o de perturbação da energia livre (do inglês FEP, *Free Energy Perturbation*), o de integração termodinâmica (do inglês TI, *Thermodynamic integration*), a mecânica molecular de área de superfície Poisson-Boltzmann (do inglês MM-PBSA, *Molecular Mechanics Poisson-Boltzmann Surface Area*) e a mecânica molecular de área de superfície generalizada de Born (do inglês MM-GBSA, *Molecular Mechanics Generalized Born Surface Area*), o quais são usados para quantificar a energia livre de ligação (KUMARI, 2014; HOMEYER, 2012).

Com custo computacional elevado, metodologias do tipo FEP atuam transformando uma molécula (alquimia) de um ponto inicial até um estado final por um caminho reversível através de uma simulação de dinâmica molecular fora de equilíbrio. É um método muito aplicado em pequenas modificações estruturais, já que o custo computacional é alto e o cálculo é realizado por pequenas mudanças estruturais na trajetória de transição (KUMARI, 2014; MEIROVITCH, 2007).

Para sistemas moleculares maiores, métodos como MM-PBSA e MM-GBSA tornam-se mais eficazes, pois obtêm o valor de variação de energia livre de ligação usando as estruturas iniciais e finais (FOLOPPE, 2006). Com as teorias dos métodos MM-GBSA e MM-PBSA, pode-se encontrar a variação de energia livre de ligação $\Delta G_{\text{ligação}}$ por meio de alguns parâmetros físico-químicos (SUN et al., 2018). Os cálculos da energia livre de ligação podem ser escritos em função dos componentes a seguir (KOLLMAN et al., 2000; GILSON e HONIG, 1988):

$$\Delta G_{\text{ligação}} = \langle G_{\text{complexo}} \rangle - (\langle G_{\text{proteína}} \rangle + \langle G_{\text{ligante}} \rangle) \quad (9)$$

sendo que o termo G_{complexo} refere-se à energia livre total do sistema proteína/ligante, o termo $G_{\text{proteína}}$ representa a energia livre total da proteína isolada e o termo G_{ligante} representa a energia livre total do ligante isolado. Cada termo de energia livre (G_x) pode ser calculado da seguinte forma (KUHN e KOLLMAN, 2000):

$$\langle G_x \rangle = \langle E_{MM} \rangle + \langle G_{\text{solvatação}} \rangle - \langle TS \rangle \quad (10)$$

sendo que E_{MM} é a energia potencial média de mecânica molecular no vácuo, incluindo a energia de ambas as interações ligadas e não ligadas (interação eletrostática e van der Waals modeladas por potencial de Coulomb e Lennard-Jones), que são calculadas com base nos parâmetros do campo de força da dinâmica molecular. $G_{\text{solvatação}}$ é a energia livre de solvatação que é composta por um termo de contribuição polar e outro de não polar. A contribuição polar para a energia de solvatação pode ser obtida resolvendo a equação de Poisson-Boltzmann (abordagem MM-PBSA) ou usando o modelo generalizado de Born (abordagem MM/GBSA), enquanto a contribuição não polar pode ser estimada a partir de uma relação linear com a área de superfície acessível ao solvente. A soma dos termos E_{MM} e $G_{\text{solvatação}}$ corresponde a contribuição entálpica da energia livre. O termo $-TS$, sendo T a temperatura e S a entropia, representa a contribuição

entrópica para a energia livre no vácuo. Quando o termo entrópico é descartado, o valor calculado é a energia livre efetiva, que geralmente é suficiente para comparar energias livres de ligação relativa de ligantes relacionados (VALDÉS-TRESANCO et al., 2021). Vale mencionar que o fator entropia de ligação entre os ligantes não é o foco do cálculo, e é ignorado devido ao custo computacional envolvido na sua obtenção (HOU et al., 2011).

Para o cálculo de ΔG de interação nas proteínas, buscar alcançar uma igualdade entre o valor obtido experimentalmente e o valor teórico se torna inviável devido ao termo entrópico. Porém é possível calcular a variação de ΔG , dessa forma o erro associado ao valor teórico é similar ao encontrado experimentalmente.

1.7.3 Cálculo de diferença de energia livre via método de integração termodinâmica de crescimento rápido

Abordagens rigorosas de cálculos de energia livre são baseadas nos princípios da mecânica estatística e usam simulações moleculares para explorar o espaço conformacional. Esses métodos normalmente fornecem previsões de diferenças de variação de energia livre (ΔG) mais precisas, em comparação com os não rigorosos. Uma razão para isto é que eles consideram inerentemente a flexibilidade conformacional das proteínas e, portanto, a contribuição entrópica (Patel, Patel, and Ytreberg 2021).

Ao longo dos anos, diferentes abordagens para calcular energias livres foram desenvolvidas, avaliadas e aplicadas, muitas vezes com foco em sistemas menores. Para sistemas maiores, como proteínas, o problema de amostragem torna-se cada vez mais grave, principalmente devido à sobreposição insuficiente das densidades termodinâmicas de espaço de fase envolvidas. Este problema também dificulta a avaliação confiável da precisão e da convergência, simplesmente devido à dificuldade de obter um valor de referência confiável. Os métodos para calcular as variações de energia livre entre dois estados A e B dividem-se em duas classes, métodos alquímicos de equilíbrio e de não-equilíbrio (Goette and Grubmüller 2009).

A perturbação de energia livre (*Free Energy Perturbation* – FEP) é um método de equilíbrio que tem sido amplamente utilizado para calcular diferenças de energia livre para substituições de resíduos de aminoácidos em proteínas ou modificações químicas na estrutura de um ligante. No entanto, esse método sofre particularmente do problema de amostragem. Devido ao crescimento exponencial da incerteza estatística com a diminuição da sobreposição da densidade do espaço de fase, o método de FEP requer amostragem excessiva, especialmente

se o tamanho da perturbação no sistema (mutação ou modificação química) for significativamente grande (Goette and Grubmüller 2009).

A Integração Termodinâmica (*Thermodynamic Integration* – TI) é um método que se baseia na força generalizada $\delta H / \delta \lambda$, sendo H a Hamiltoniana do sistema e λ um parâmetro de acoplamento. Nesse caso, o parâmetro λ varia continuamente (crescimento lento) ou, semelhantemente ao FEP, é amostrado em valores discretos. Uma consequência do deslocamento contínuo de λ no TI de crescimento lento (*Slow-Growth Thermodynamic Integration*) é que o sistema nunca está em equilíbrio, como é assumido na construção do método. Portanto, resultados precisos são esperados apenas para sistemas pequenos ou com tempos de simulação muito longos (Goette and Grubmüller 2009).

O método de TI com crescimento lento é baseado em simulações em equilíbrio e por outro lado sua versão fora de equilíbrio é conhecida como TI de rápido crescimento (*Fast-Growth Thermodynamic Integration*). Esse método tem suas raízes na igualdade de Jarzynski (mais tarde reformulada por Cuendet para um sistema NPT), que relaciona a variação de energia livre (ΔG_{AB}) dos estados A e B ao trabalho (W) realizado no sistema para transformá-lo do estado A para o estado B (Goette and Grubmüller 2009; Gapsys et al. 2015):

$$\exp\left(-\frac{\Delta G_{AB}}{k_B T}\right) = \langle \exp\left(-\frac{W}{k_B T}\right) \rangle \quad (11)$$

sendo que k_B é a constante de Boltzmann e T é a temperatura absoluta. A notação $\langle \rangle$ denota uma média sobre um conjunto de N trajetórias, que foram iniciadas a partir de um conjunto equilibrado.

A igualdade de Jarzynski exige que as transições sejam iniciadas a partir de um conjunto de equilíbrio. Os valores do trabalho podem ser obtidos a partir de uma relação semelhante à da integração termodinâmica (Goette and Grubmüller 2009; Gapsys et al. 2015):

$$W = \int_0^1 \frac{\delta H}{\delta \lambda} d\lambda \quad (12)$$

para um processo de conversão alquímica $A \rightarrow B$ em um tempo de simulação fora de equilíbrio que pode ser muito curto (50 a 200 ps). Na equação 12, λ é o parâmetro de acoplamento, que transforma o sistema durante uma simulação fora de equilíbrio do estado A para o estado B,

definido pelos Hamiltonianos H_A e H_B , respectivamente, por meio da seguinte relação (Goette and Grubmüller 2009):

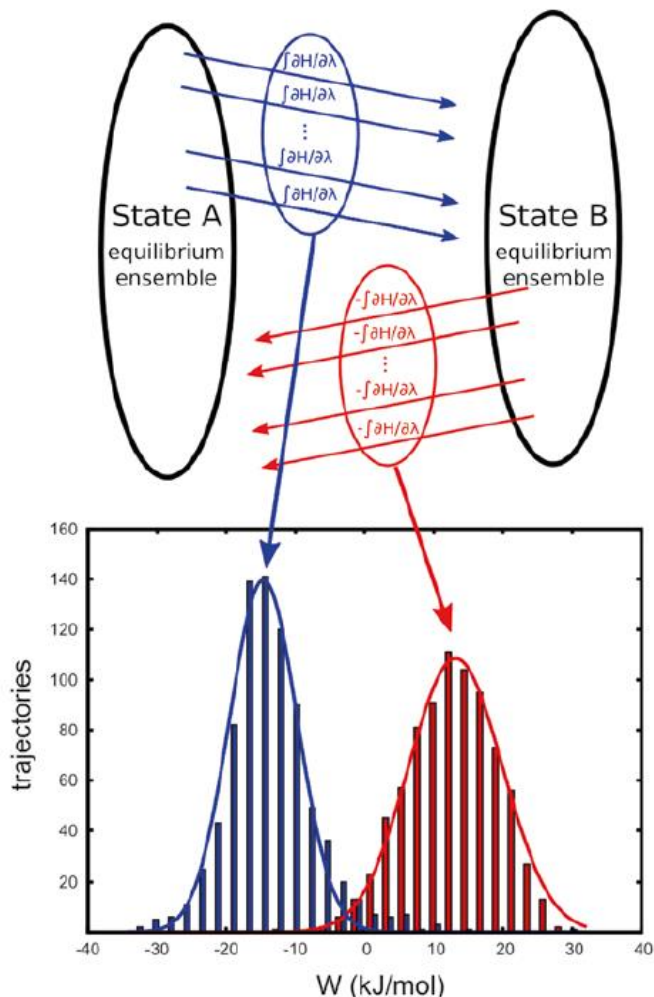
$$H = (1 - \lambda)H_A + \lambda H_B \quad (13)$$

O método de crescimento rápido baseia-se na equação de Crooks do Teorema de Flutuação de Crooks, que relaciona a variação de energia livre ΔG_{AB} com o trabalho direto (W) e reverso ($-W$) realizado em um sistema durante a transformação fora do equilíbrio (Gapsys et al. 2015; Silvestri et al. 2023):

$$\frac{P_d(W)}{P_r(-W)} = \exp \left[-\frac{(W - \Delta G_{AB})}{k_B T} \right] \quad (14)$$

sendo que $P_d(W)$ e $P_r(-W)$ corresponde a distribuição de trabalho durante o processo direto ($A \rightarrow B$) e reverso ($B \rightarrow A$), respectivamente. O parâmetro de acoplamento λ muda de 0 para 1 de modo contínuo e rápido para transformar o sistema do estado A para o estado B, a fim de obter o trabalho direto (W). O mesmo acontece para calcular o trabalho reverso ($-W$), mudando λ de 1 para 0. Tanto W quanto $-W$ são calculados como uma média sobre um conjunto de N trajetórias que são iniciadas a partir de um conjunto equilibrado para ambos os estados A e B. Em outras palavras, muitas transições muito rápidas (centenas de transições) são realizadas de A para B e vice-versa (Figura 9).

Figura 9. Esquema da resolução das derivadas do Hamiltoniano pelo λ para equações fora do equilíbrio. Esta forma de encontrar a energia proporciona uma maneira de resolver o ΔG para uma transformação química de um estado A para B. O ponto de intersecção das curvas do histograma formado nos mostra o ΔG obtido.



Fonte: GAPSYS et al., 2015.

Para cada trajetória o trabalho mecânico para o processo de comutação é calculado através da equação 12. A partir da equação 14 pode-se calcular que ΔG_{AB} é igual ao trabalho W para o qual $P_d(W) = P_r(-W)$, ou seja, o ponto de intersecção das duas distribuições de trabalho (Figura 9). A determinação direta do ponto de intersecção, como por exemplo a partir de histogramas, está sujeita a grandes erros estatísticos uma vez que são somente usados os valores de trabalho que caem no *bin* (“fatia”) que contém o ponto de intersecção. Particularmente se as distribuições diretas e reversas exibirem apenas uma pequena sobreposição, esse número pode ser muito pequeno ou até zero (Goette and Grubmüller 2009; Gapsys et al. 2015).

A partir do método de Intersecção Gaussiana de Crooks (*Crooks Gaussian Intersection* – CGI) utiliza-se aproximações Gaussianas para as distribuições de trabalho (W e $-W$) e o ponto de intersecção é dado por (Goette and Grubmüller 2009; Gapsys et al. 2015):

$$\Delta G_{AB} = \frac{\frac{W_d}{\sigma_d^2} - \frac{W_r}{\sigma_r^2} \pm \sqrt{\frac{1}{\sigma_d^2 \sigma_r^2} (W_d + W_r)^2 + 2 \left(\frac{1}{\sigma_d^2} - \frac{1}{\sigma_r^2} \right) \ln \frac{\sigma_r}{\sigma_d}}}{\frac{1}{\sigma_d^2} - \frac{1}{\sigma_r^2}} \quad (15)$$

sendo que W_d (trabalho direto) e $-W_r$ (trabalho reverso) são as médias e σ_d e σ_r são os desvios padrão das funções Gaussiana que representam as distribuições de trabalho direto e reverso, respectivamente. Quando $\sigma_d \neq \sigma_r$, dois pontos de intersecção geralmente são encontrados, lembrando que ΔG_{AB} é definido de forma única. Somente para $\sigma_d^2 = \sigma_r^2$ obtém-se uma solução única, $\Delta G_{AB} = (W_d - W_r)/2$. Tipicamente, um ponto de intersecção é localizado entre W_d e $-W_r$, e é próximo ao ponto $(W_d - W_r)/2$. O segundo ponto de intersecção bem na região da cauda e neste caso é um artefato da aproximação Gaussiana, que muitas vezes falha na região da cauda. Assim, escolhe-se o valor de intersecção mais próximo de $(W_d - W_r)/2$ como estimativa para ΔG_{AB} via método CGI.

A energia livre absoluta de um sistema é impossível de determinar, mas felizmente a maioria dos problemas pode ser formulada em termos de energias livres relativas. A variação de energia livre são mais acessíveis e contêm informações importantes sobre o sistema. A mudança na energia livre devido à ligação ($\Delta G_{\text{ligação}}$) pode ser determinada experimentalmente (por exemplo, por meio de calorimetria). Embora as afinidades de ligação absolutas possam ser calculadas utilizando, por exemplo, *Umbrella Sampling*, tais cálculos são geralmente complicados, uma vez que todo o processo de ligação/dissociação precisa ser levado em conta, enquanto o seu caminho é desconhecido na maioria dos casos. Portanto, investigar o efeito de uma transformação alquímica no sistema (mutação de resíduos de aminoácido ou uma modificação química do ligante) na ligação é geralmente mais viável. Para isso, é calculada a diferença de variação de energia livre ($\Delta \Delta G_{\text{mutação, ligação}}$) (Gapsys et al. 2015).

Os métodos alquímicos calculam o trabalho necessário para “mover” um sistema de um estado para outro ($A \rightarrow B$ ou $B \rightarrow A$) através de caminhos não físicos. Como a energia livre de um sistema é função do seu estado, a variação de energia livre (ΔG_{AB}) encontrada entre os estados é independente do caminho percorrido entre eles. Amostragem suficiente é de importância crítica em todos os métodos de energia livre, e a dificuldade computacional de alcançar a convergência aumenta dramaticamente com a magnitude da perturbação. As energias livres de ligação geralmente envolvem uma grande perturbação – em um estado, ambos os parceiros de ligação estão livres em solução, no outro estão em um complexo. A sobreposição

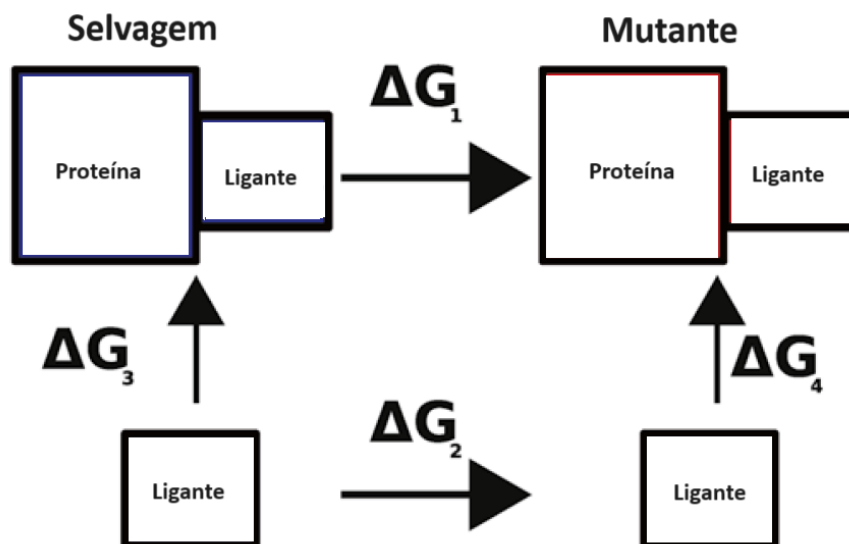
do espaço de fase entre esses dois estados é frequentemente pequena, resultando em uma convergência lenta. Uma transição alquímica do estado A (uma proteína de tipo selvagem ou um ligante) para o estado B (uma molécula mutada ou modificada), embora fisicamente impossível, requer uma perturbação muito menor. Como o interesse é a diferença nas variações de energia livre entre os estados A e B, podemos aproveitar o facto de que a energia livre de um estado não depende do caminho percorrido para alcançá-lo. Consequentemente, as variações de energia livre ao longo de um ciclo fechado de reações (Figura 10) sempre somarão zero. Este recurso permite o cálculo das diferenças duplas na energia livre ($\Delta\Delta G$) de ligação, termoestabilidade e partição em diferentes solventes entre dois estados de um sistema a partir da seguinte expressão (Gapsys et al. 2016):

$$\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_4 - \Delta G_3 \quad (16)$$

sendo que ΔG_2 e ΔG_1 são as variações de energia livre alquímica ΔG_{AB} da alteração do estado A para o estado B, que no contexto de uma interação proteína-ligante ou proteína-peptídeo (o caso do presente trabalho) corresponde a uma transformação (mutação ou modificação química) do ligante ou peptídeo no estado livre e complexado com a proteína, respectivamente. ΔG_4 e ΔG_3 são as variações de energia de ligação $\Delta G_{\text{ligação}}$ entre o estado livre e complexo do ligante/peptídeo para sua condição mutado/modificado e selvagem, respectivamente. Neste trabalho $\Delta\Delta G$ é calculado a partir $\Delta G_1 - \Delta G_2$, ou seja, $\Delta G_{AB}^{\text{complexo}} - \Delta G_{AB}^{\text{ligante}}$, lembrando que cada termo é determinado pelo método de TI de crescimento rápido e a partir da análise de Intersecção Gaussiana de Crooks.

Vale destacar que o $\Delta\Delta G$ resultante é frequentemente mais preciso do que os valores de ΔG usados para calculá-lo, pois erros sistemáticos (como aqueles causados pela presença de partículas fictícias e amostragem limitada) são cancelados quando as diferenças de energia livre são subtraídas. Observe-se que dessa forma, as afinidades de ligação relativas são obtidas sem estudar o evento real de ligação/dissociação.

Figura 10. Representação esquemática de um ciclo termodinâmico. Para determinar a alteração na afinidade de ligação após uma mutação peptídica ou modificação do ligante, a diferença entre a energia livre de ligação do ligante ou peptídeo no estado A (selvagem, ΔG_3) e a do estado B (mutante, ΔG_4) precisa ser determinada. Porém, os valores ΔG_1 e ΔG_2 são mais acessíveis aos métodos de perturbação de energia livre e integração termodinâmica de crescimento lento e rápido. Usando a conservação de energia neste ciclo fechado, podemos derivar $\Delta\Delta G = \Delta G_1 - \Delta G_2 = \Delta G_4 - \Delta G_3$.



Fonte: Adaptado de GAPSYS et al., 2015.

2. JUSTIFICATIVA

O Vírus Sincicial Respiratório humano é uma das maiores causas de infecções respiratórias, principalmente em crianças e idosos, e é altamente contagioso e associado a doenças como bronquiolite e pneumonia (NAIR et al., 2010). Como ainda não existem vacinas licenciadas no Brasil e somente uma para idosos no Estado Unidos, a busca por possíveis antivirais que tem como alvo molecular a interação da proteína M2-1 do hRSV com a fosfoproteína viral (mimetizado aqui pelo P90-110) pode se apresentar como uma estratégia promissora, visto que a interação M2-1/P desempenha um papel crucial na replicação e na regulação da expressão gênica desse vírus.

A interação M2-1/P é chave para o processo replicativo do hRSV, tornando-a um interessante objeto de estudo para que se possa realizar a inativação da M2-1, evitando assim a replicação do hRSV (BLONDOT et al, 2012). O design racional de miméticos do peptídeo P90-110, região responsável pela ligação a M2-1, pode ser realizada a fim de encontrar ligantes competitivos à interação M2-1/P e, para essa finalidade, ferramentas computacionais como cálculos de dinâmica molecular, cálculos de energia de ligação e métodos de integração

termodinâmica podem ser utilizados para investigar tanto as interações moleculares das proteínas selvagens como mutantes planejados.

Sabe-se que futuros medicamento, específicos contra o hRSV, podem ajudar os que sofrem com esse vírus e, portanto, o presente trabalho pode fornecer informações moleculares relevantes para obter conhecimento científico necessário para esse fim.

3. OBJETIVOS

3.1 Objetivo geral

A interação da fosfoproteína com a proteína M2-1 é fundamental para o processo de replicação do hRSV. Compreender quais resíduos de aminoácidos são essenciais para que essa interação ocorra é crucial e nos possibilita sugerir modificações na estrutura do peptídeo P90-110 que possam aprimorar essa interação, sugerindo assim seus miméticos como potenciais antivirais. Nesse contexto, o objetivo geral do presente trabalho é realizar a modelagem da interação do domínio globular da proteína M2-1 (dgM2-1) com a porção da fosfoproteína P90-110, clusterizar as estruturas mais representativas e, utilizando cálculos de dinâmica molecular, fazer um *scan* de alanina nesse peptídeo para verificar os principais resíduos envolvidos na interação e propor peptídeos miméticos (mutações) que modulem a capacidade de interação proteína-peptídeo.

3.2 Objetivos específicos

- 1- Realizar as simulações de dinâmica molecular da interação de cada monômero do complexo M2-1/P90-110. Os dados a serem modelados foram encontrados no PDB 6G0Y.
- 2- Realizar os cálculos de energia livre de ligação via métodos MM-PBSA e MM-GBSA, cálculos de RMSD, quantificação das ligações de hidrogênio e contatos moleculares entre a proteína M2-1 e fosfoproteína peptídeo selvagem P90-110.
- 3- Realizar a concatenação das trajetórias dos quatro monômeros do complexo M2-1/P90-110 e, em seguida, a análise de clusterização desse complexo, selecionando a estrutura representativa para as próximas etapas.
- 4- Fazer um *scan* de alanina, avaliando as mudanças energéticas de interação via cálculos de energia livre de ligação por MM-PBSA e MM-GBSA, e realizar cálculos de diferença de energia livre de ligação via método de integração termodinâmica de crescimento rápido.

5- Propor mutações no peptídeo P90-110 da fosfoproteína utilizando das informações moleculares advinda dos cálculos computacionais, que possam melhorar a energia de ligação entre a proteína e o peptídeo.

6- Realizar cálculos de energia livre ligação via integração termodinâmica de crescimento rápido para o complexo formado entre o peptídeo P90-110 mutado (miméticos) e o dgM2-1.

4. MÉTODOS

4.1 Construção do sistema molecular

A estrutura de cristal da proteína M2-1 do hRSV complexada com o fragmento da fosfoproteína viral do resíduo 90 ao 110, nomeado peptídeo P90-110, foi obtido do Banco de Dados de Proteína PDB a partir do código de acesso 6G0Y (Selvaraj et al. 2018). Nessa estrutura de cristal, a proteína M2-1 se apresenta como um tetrâmero, tendo ligado a cada um de seus domínios globular o peptídeo P90-110 da fosfoproteína viral. O programa PyMOL (Delano 2002) foi utilizado para a construção dos sistemas moleculares formados do domínio globular da M2-1 (dgM2-1, resíduos de 56 a 177) ligado ao peptídeo P90-110. Para isso, foram removidos os domínios de oligomerização e de dedo de zinco, assim como o C-terminal intrinsecamente desordenado da proteína. Além da remoção de partes da proteína, as porções N-terminais (resíduos de 90 a 96) dos quatro peptídeo P90-110 foram reconstruídas a partir do programa PyMOL, aplicando a função *sculpt* para evitar que essa reconstrução fique completamente esticada. Por fim, quatro estruturas do complexo dgM2-1/P90-110 foram geradas como ponto de partida para as simulações de dinâmica molecular.

4.2 Dinâmica Molecular

Os cálculos de dinâmica molecular foram realizados com o pacote de programas GROMACS 5.0.7 (Abraham et al. 2015) usando o Supercomputador Santos Dumont do Laboratório Nacional de Computação Científica (LNCC). O campo de força utilizado foi o AMBER14SB (Maier et al. 2015) e o modelo de água utilizado foi o TIP3P (Jorgensen et al. 1983). A estrutura dos quatro complexos dgM2-1/P90-110 foram colocadas em caixas cúbicas de simulação e em seguida solvatada com moléculas de água. As cargas foram neutralizadas utilizando íons de cloro (Cl^-) e uma concentração salina de 150 mM de NaCl foi adicionada. O servidor PDB2PQR (OLSSON et al, 2011) foi utilizado para determinar o estado de

protonação dos resíduos ionizáveis, definidas como não protonadas e com o hidrogênio na posição ND1, considerando pH 7,0. As condições de contorno utilizadas, como Temperatura e pressão, foram o *ensemble* NPT, mantendo o sistema em 298 K (25 °C) e 1,0 bar, usando o termostato de Nosé-Hoover ($\tau_T = 2 \text{ ps}$) and Parrinello-Rahman barostat ($\tau_T = 2 \text{ ps}$ e compressibilidade de $4,5 \times 10^{-5} \text{ bar}^{-1}$). Foi utilizado um raio de corte de 12 Å para os potenciais Lennard-Jones e Coulomb. As interações eletrostáticas de longo alcance foram calculadas usando o algoritmo de Ewald de malha de partículas. O sistema dgM2-1/P90-110 com a caixa solvatada e os íons já adicionados foram submetidos a um processo de minimização empregando o algoritmo de gradiente conjugado utilizado para relaxar a superposição de átomos gerados no processo de construção da caixa de simulação. Minimizações de energia foram realizadas com o integrador de descida íngreme (*steepest descent*) e o algoritmo de gradiente conjugado, usando $1000 \text{ kJ}\cdot\text{mol}^{-1}\text{nm}^{-1}$ como critério de força máxima. Um total de 500 mil passos de 2 fs de simulação de dinâmica molecular foram realizados para equilíbrio de temperatura e pressão, que foram definidas em 298 K (25 °C) e 1,0 bar, empregando constantes de força de $1000 \text{ kJ}\cdot\text{mol}^{-1}\text{nm}^{-2}$ para todos os átomos pesados do sistema moleculares. No final da preparação dos sistemas moleculares, simulações de dinâmica molecular de 250 ns dos quatro complexos dgM2-1/P90-110 foram realizadas, totalizando 1 μs quando somadas as 4 dinâmicas dos monômeros do complexo. Vale destacar que essas simulações começaram a partir de conjuntos de coordenadas diferentes para cada complexo.

Após a geração das trajetórias de dinâmica molecular, elas foram alinhadas com relação aos átomos da cadeia principal e analisadas de acordo com o RMSD, número de ligações de hidrogênio e número de contatos. Os valores de RMSD foram calculados para os átomos da cadeia principal do dgM2-1 dos resíduos de 56 a 177 e de 71 a 177 e para os átomos da cadeia principal do P90-110 dos resíduos de 90 a 110 e de 97 a 110. O número de ligações de hidrogênio entre dgM2-1 e P90-110 foram calculadas usando um critério de distância de 3,5 Å e ângulo máximo de 30°. O número de contatos para distâncias menores do que 6 Å foram quantificados entre os pares de átomos da proteína e do peptídeo. Na sequência das análises individuais de cada complexo dgM2-1/P90-110, as trajetórias foram concatenadas em um único arquivo e a análise de clusterização foi realizada utilizando ferramenta *g_cluster* do GROMACS com algoritmo *gromos* com critério de corte de 4 Å para a determinação de uma estrutura representativa dgM2-1/P90-110 empregada nas simulações dos mutantes de alanina.

Para o processo de *scan* de alanina, os resíduos de 97 a 109, região de alfa-hélice do peptídeo P90-110 da estrutura representativa obtida da análise de clusterização, foram mutados para alanina usando a ferramenta de mutagêneses do programa PyMOL (Delano 2002). Após a

realização de cada mutação, um total de 12, os sistemas moleculares de complexo proteína-peptídeo e peptídeo livre foram submetidos a simulações de dinâmica molecular seguindo o protocolo e utilizando os parâmetros descritos anteriormente. Vale mencionar que simulações de referência do complexo dgM2-1/P90-110 e peptídeo P90-110 selvagens, a partir da estrutura representativa vinda da clusterização, também foram realizadas. Todas essas trajetórias de dinâmica molecular foram empregadas nos cálculos de energia livre de ligação via MM-PBSA e MM-GBSA e de Integração Termodinâmica de Crescimento Rápido.

4.3 Cálculos de energia livre de ligação via MM-PBSA e MM-GBSA

Para a realização dos cálculos de energia livre de ligação via MM-PBSA e MM-GBSA, as trajetórias dos complexos mutados e selvagem foram preparados extraíndo 100 frames espaçados igualmente dos últimos 200 ns, excluindo os primeiros 50 ns. Para ambos os métodos cálculo de energia, uma concentração salina de 150 mM de NaCl e constante dielétrica do solvente de 80 e da proteína de 2 foram empregadas. Os cálculos via MM-PBSA foram realizados utilizando o programa *g_mmpbsa* desenvolvido para o pacote GROMACS juntamente com os *scripts* *MnPbSaStat.py* e *MmPbSaDecomp.py* (Baker et al. 2001; Kumari, Kumar, and Lynn 2014). Já os cálculos de MM-GBSA foram feitos usando o programa *gmx_MMPBSA* (Valdés-Tresanco et al. 2021) desenvolvido para pacote GROMACS.

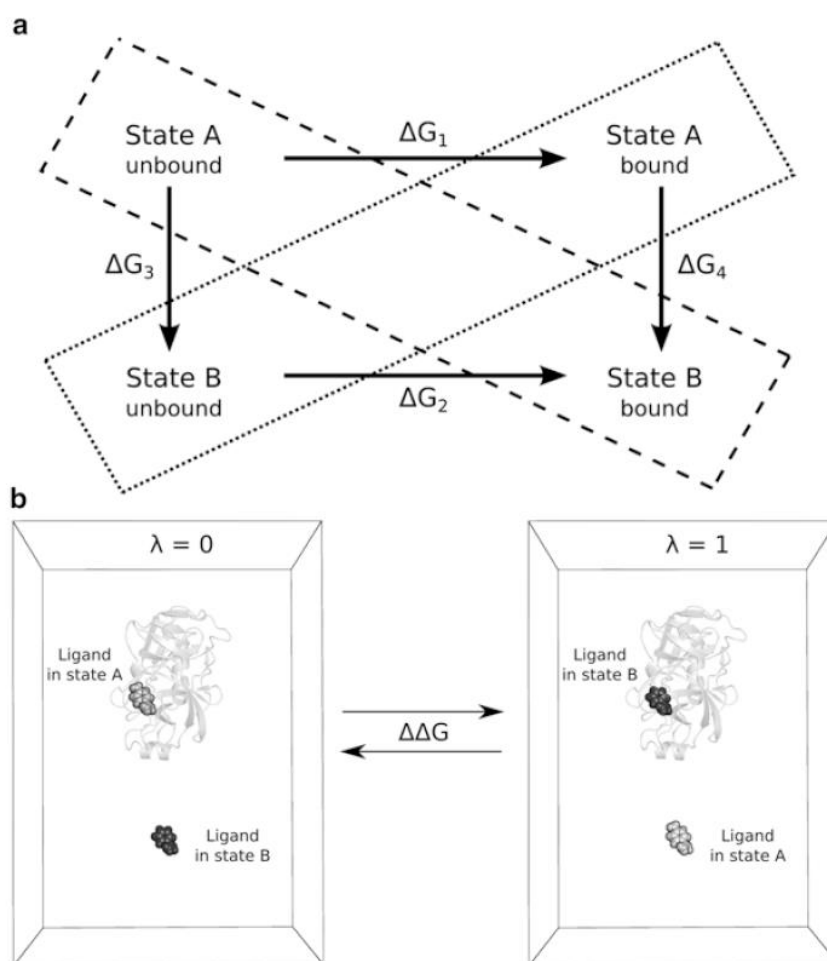
4.4 Cálculos de integração termodinâmica de crescimento rápido

Para a realização dos cálculos de integração termodinâmica de uma única mutação no processo de *scan* de alanina para a região de alfa-hélice de Phe98 a Phe109 do P90-110, quatro trajetórias de dinâmica molecular são necessárias, sendo elas: complexo dgM2-1/P90-110 com o peptídeo selvagem e mutado e peptídeo P90-110 livre no estado selvagem e mutado. A partir dessas trajetórias, 100 *frames* foram extraídos de cada uma ao longo dos 250 ns com espaçamentos iguais. Para cada um dos frames vindos dos quatro sistemas moleculares, foram realizadas simulações de dinâmica molecular fora de equilíbrio de 50 ps usando o programa GROMACS, seguindo o protocolo de preparo, minimização, equilibração de temperatura e pressão e produção da simulação como descrito na seção 4.2, com modificações quanto ao uso do termostato V-rescale e barostato Berendsen. O programa PMX (PMX: GENERATE HYBRID PROTEIN STRUCTURE AND TOPOLOGY, 2020) foi utilizado para gerar a topologia híbrida para a transformação alquímica aplicando as ferramentas

generate_hybrid_topology.py e *mutate.py* e para analisar (Intersecção Gaussiana de Crooks) as trajetórias das simulações fora de equilíbrio empregando a ferramenta *analyze_dhdl.py*. O valor do parâmetro de acoplamento λ variou de 0 a 1 para a simulação fora de equilíbrio na transformação alquímica direta e de 1 a 0 na transformação reversa, ambas com incremento de 4×10^{-5} .

Para os resíduos carregados, foi necessário preparar um sistema duplo em uma única caixa (do inglês *double-system/single-box*), na qual a caixa de simulação é preparada contendo o complexo dgM2-1/P90-110 com uma distância de afastamento maior do que 30 Å do peptídeo P90-110 isolado. Dessa forma é possível preparar uma caixa de simulação contendo um complexo com peptídeo P90-110 carregado e o peptídeo P90-110 isolado com um resíduo não carregado, e através da mudança alquímica, a carga se iguala durante a variação de λ de 0 a 1, como pode ser observado na Figura 11.

Figura 11. Configuração de sistema duplo em uma única caixa (*double-system/single-box*). Em a) tem-se dois tipos de sistema de um ciclo termodinâmico colocados em uma caixa de simulação. As diferentes caixas no esquema são indicadas pelas linhas tracejadas e pontilhadas. Em b) tem-se um exemplo de configuração de simulação para o cálculo da energia livre de ligação do ligante. Durante uma transição, o ligante ligado à enzima está sendo transformado do estado A para B, enquanto a molécula solvatada sofre uma transição reversa do estado B para A. A carga líquida da caixa de simulação permanece zero.



Fonte: GAPSYS et al., 2015.

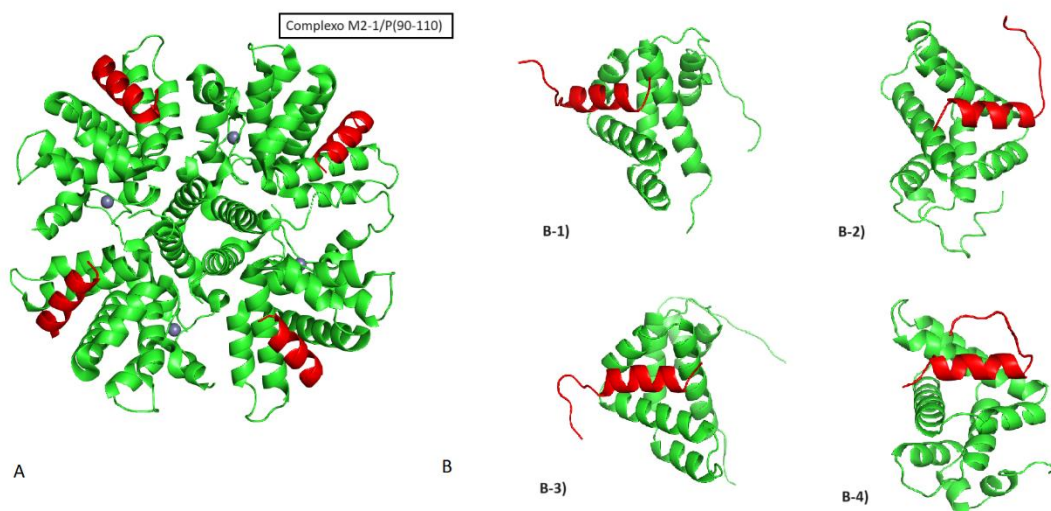
5. RESULTADOS E DISCUSSÕES

5.1 Modelagem

Inicialmente, obteve-se a estrutura do complexo da proteína M2-1 com o fragmento estrutural da fosfoproteína P contendo os resíduos de aminoácidos de 90 a 110 (nomeado peptídeo P90-110) a partir do banco de dados de proteínas PDB através do código de acesso 6G0Y. Estruturalmente, a M2-1 é um tetrâmero em seu estado nativo com cada uma de suas cadeias interagindo com o fragmento da fosfoproteína, peptídeo P90-110. Essa interação M2-1/P90-110 ocorre com o domínio globular da M2-1 (dgM2-1) e, portanto, na estrutura do PDB

6G0Y, há quatro complexos dgM2-1/P90-110 para cada cadeia monomérica. Nesta etapa, foi definido o trabalho apenas com uma unidade do complexo M2-1/P90-110, devido ao tamanho da estrutura ser menor, viabilizando os cálculos computacionais, já que a unidade monomérica do complexo M2-1/P90-110 representa as características da interação no tetrâmero M2-1/P. Utilizando o programa PyMOL, foi possível separar cada uma de suas cadeias juntamente com o peptídeo P0-110 da fosfoproteína, ficando no total o conjunto de quatro complexos, que tiveram os domínios de oligomerização e de dedo de zinco removidos posteriormente para restar somente o domínio globular da M2-1 (Figura 12). Vale mencionar que as regiões flexíveis do P90-110 (principalmente o N-terminal não helicoidal), que não estavam presentes na estrutura de cristal do peptídeo, foram reconstruídas utilizando o programa PyMOL.

Figura 12. Estrutura da proteína M2-1 complexada com o peptídeo P90-110 da fosfoproteína. A) Estrutura tetramérica de M2-1 (PDB 6G0Y), colorida em verde está a proteína M2-1, com uma região central formado por 4 hélices de oligomerização, uma de cada monômero, em vermelho o fragmento da fosfoproteína e azul o átomo de Zinco. B) As estruturas dos quatro complexos dgM2-1/P90-110, com o domínio colorido em verde e o fragmento P90-110 da fosfoproteína em vermelho. Tanto a proteína quanto os peptídeos são representados utilizando o modelo de cartoon e o átomo de zinco denotado pelo modelo de esfera.



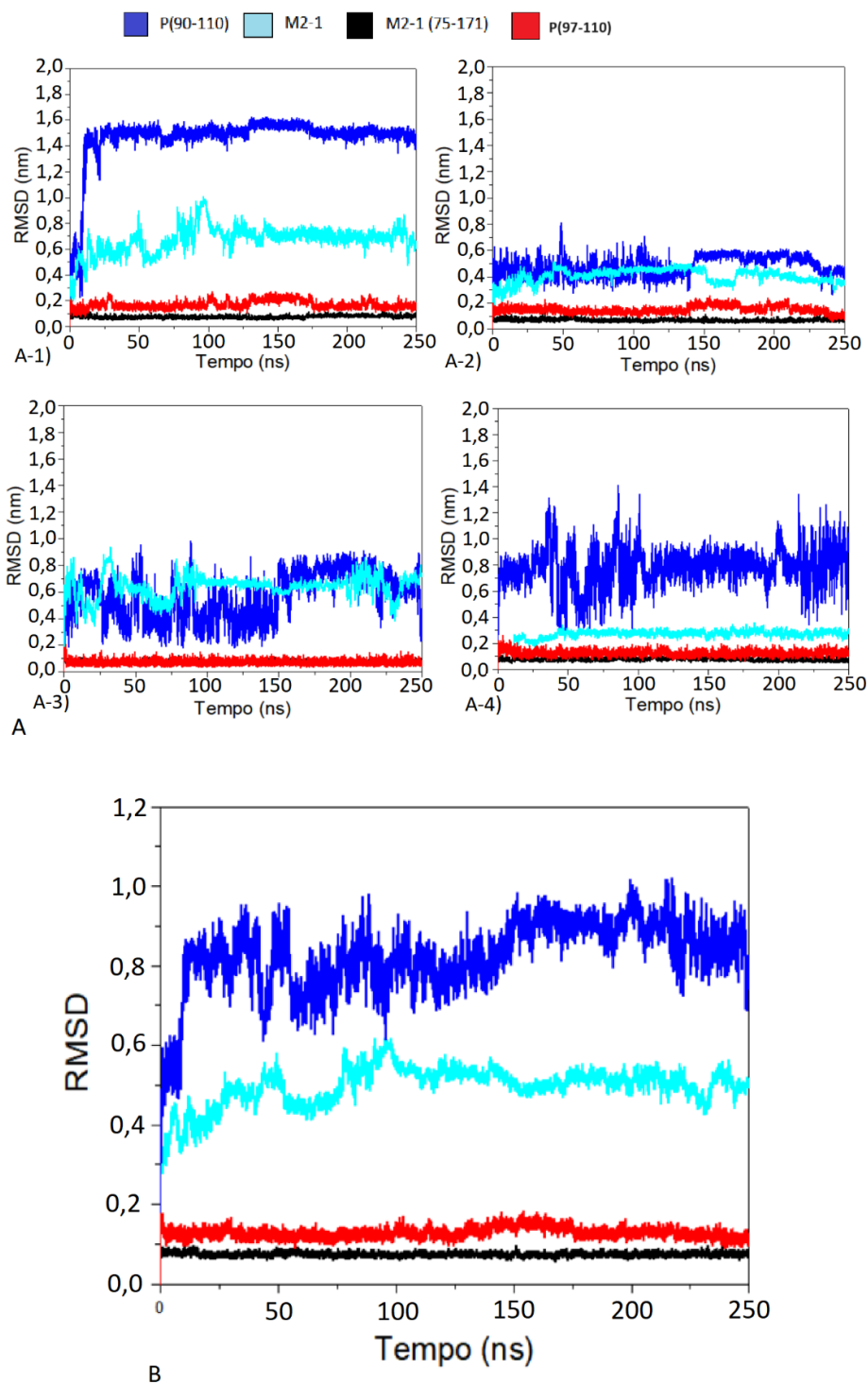
Fonte: Próprio autor.

Na Figura 11 pode-se observar a proteína M2-1 como um tetrâmero complexada com a porção 90-110 da proteína P. As Figuras B-1, B-2, B-3 e B-4 são as estruturas do dgM2-1 provenientes do tetrâmero, os quais estão complexados com o peptídeo P90-110 na conformação de alfa-hélice, o qual encontra-se posicionado em uma fenda formada pelas hélices $\alpha 4$, $\alpha 5$ e $\alpha 6$ na estrutura do M2-1.

5.2 Interação M2-1/P 90-110

As quatro estruturas dos complexos dgM2-1/P90-110 apresentadas na Figura 11 foram simuladas via dinâmica molecular por 250 ns. A partir das trajetórias de dinâmica molecular desses quatro complexos, calculou-se os valores de RMSD para os componentes do sistema molecular, separando por cadeia, como mostrado na Figura 13. Em azul é representado os RMSDs para o P90-110, que devido ao seu tamanho e sua cauda N-terminal, com características intrinsecamente desordenadas, apresentam significativas flutuações de seus valores. Em ciano é representado os RMSDs do domínio globular da proteína M2-1 (resíduos 58-177), para os quais também é possível observar variações significativas de seus valores devido a porção N-terminal que adota uma conformação instável ao longo das dinâmicas moleculares. Uma vez que a proteína e o peptídeo têm regiões flexíveis e que refletem nos valores de RMSD, recalculou-se esses valores considerando apenas as regiões que apresentaram estruturas secundárias definidas em alfa-hélices, o que corresponde aos resíduos 75-171 para o dgM2-1 e 97-110 para o peptídeo P90-110, os quais são vistos nas cores preta e vermelha, respectivamente. Aqui nota-se uma maior estabilidade dos valores de RMSD quando comparado àqueles do dgM2-1 e P90-110 completos, evidenciando que as flutuações observadas em azul (P90-110) e ciano (dgM2-1) são provenientes das porções N-terminais não estruturadas presentes na proteína e no peptídeo. A Figura 13B mostra os valores médios de RMSD das quatro dinâmicas moleculares, para os quais pode-se observar de uma maneira mais ampla, e com menor ruído, as estabilidades conformacionais da proteína e do peptídeo nos complexos.

Figura 13. Conjunto de figuras de RMSD das cadeias em A da interação M2-1/P90-110 e figura de RMSD médio em B. Em azul estão os valores de RMSD para o fragmento todo da fosfoproteína P90-110. Em ciano estão representados os valores de RMSD do domínio globular da M2-1(58-177). Em preto estão os valores de RMSD da porção estruturada em hélices-alfas da M2-1 (75-171) e em vermelho o RMSD para o complexo proteína-peptídeo com foco nas regiões estruturadas da proteína M2-1 e da proteína P.



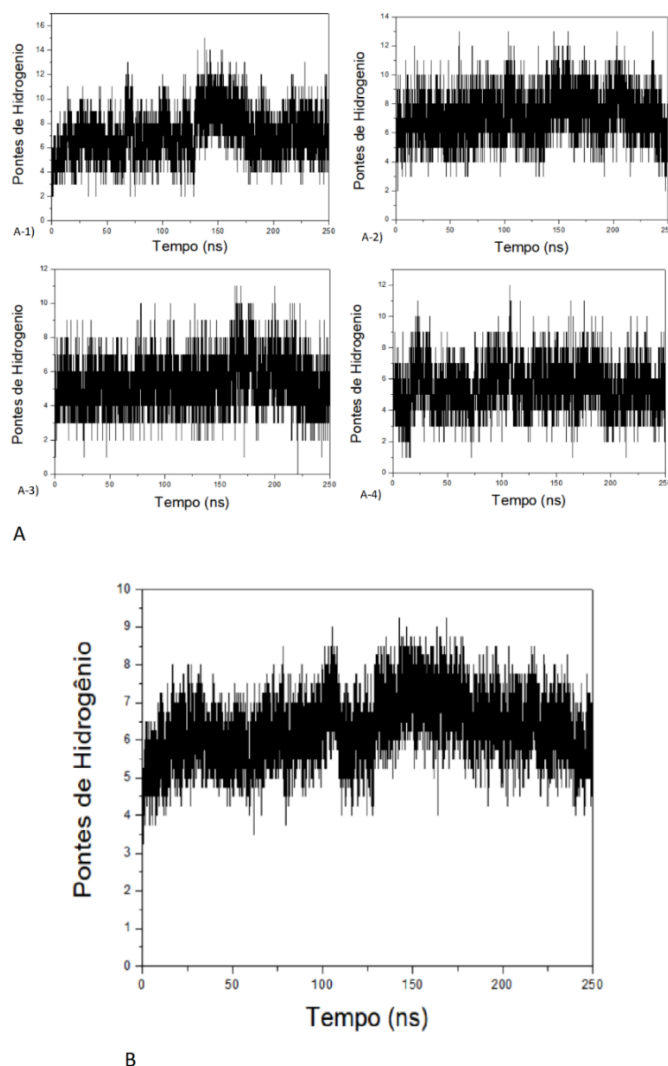
Fonte: Próprio autor.

Pode-se observar nos resultados das quatro simulações de dinâmica molecular que tanto a proteína, quanto o peptídeo, alcançam patamares de estabilidade dos valores de RMSD, mesmo com flutuações em torno de um valor médio como no caso dos valores para as sequências completas do dgM2-1 e P90-110. Para as sequências completas, em alguns casos os patamares de RMSD são alcançados após 10 ns de simulação e, de modo geral, indicam que mesmo com a flexibilidade conformacional dos N-terminais do dgM2-1 e da P90-110 é possível que essas regiões adotem conformações médias representativas. Quando retiramos essas regiões N-terminais são retiradas da sequência analisada nos cálculos de RMSD, os valores diminuem e estabilizam consideravelmente desde o início das simulações, mostrando uma estabilidade conformacional dos quatro complexos. Toda essa discussão pode ser mais facilmente analisada e observada a partir dos valores médios de RMSD apresentados na Figura 13B.

Pode-se também observar também que na média o P90-110 apresentou maiores valores de RMSD em comparação com aqueles do dgM2-1. Esse resultado corrobora com aqueles obtidos nas simulações das dinâmicas moleculares individuais, indicando que na estrutura dos quatro complexos o peptídeo P90-110 apresenta uma variação maior de sua conformação do que o dgM2-1 com relação ao ponto de partida (estrutura de cristal).

Os cálculos das ligações de hidrogênio entre o domínio dgM2-1 e o peptídeo P90-110 são mostrados na Figura 14. As ligações de hidrogênio são interações intermoleculares consideradas significativamente fortes energeticamente e favorecem a formação e principalmente a estabilização de complexos moleculares. Assim, quanto mais ligações de hidrogênio, possivelmente maior a afinidade de interação entre o dgM2-1 e o P90-110. No sistema adotado no presente estudo, as simulações de dinâmica molecular de cada complexo dgM2-1/P90-110 se mostraram bem semelhantes em relação ao número de ligações de hidrogênio formadas, o qual na média fica em torno de seis ligações (Figura 14B).

Figura 14. Ligações de hidrogênio para as dinâmicas moleculares entre a M2-1 e a proteína P (90-110) das estruturas das 4 cadeias do tetrâmero. No eixo y tem-se os valores de ligações de hidrogênio e no eixo x os valores do tempo de simulação em nanossegundos. Em A tem-se os resultados para as 4 dinâmicas das estruturas das cadeias da proteína M2-1 e em B os valores médios de ligações de hidrogênio.

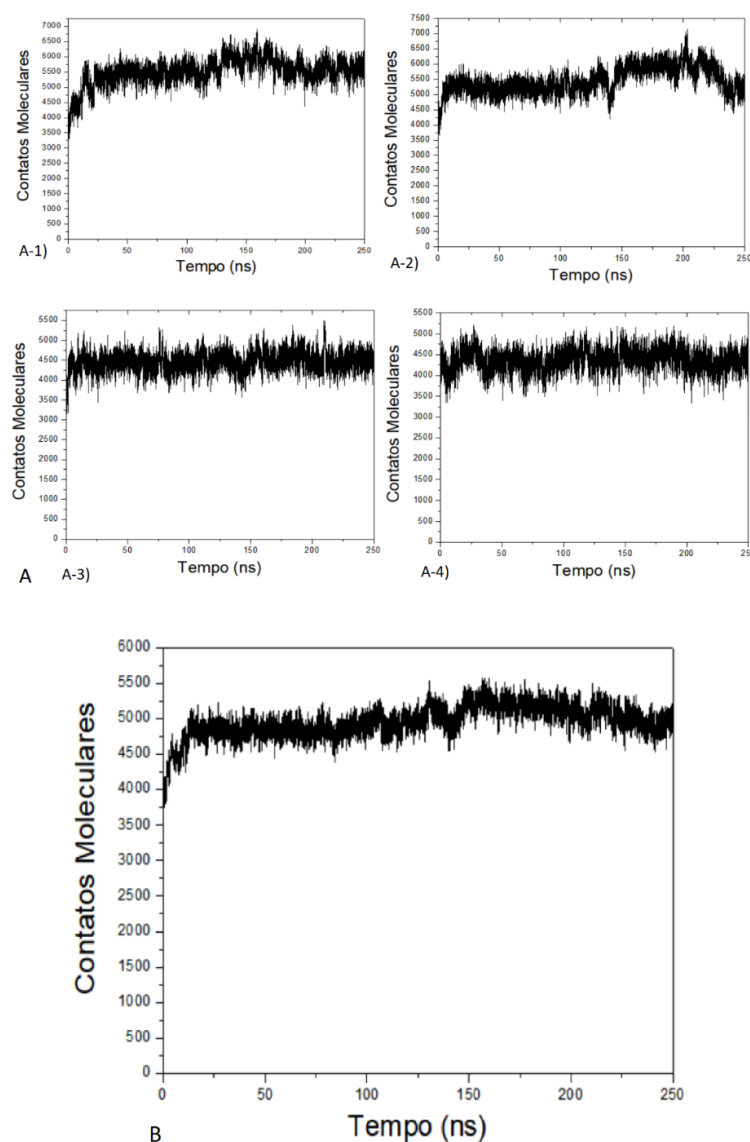


Fonte: Próprio autor.

Os contatos moleculares menores que 0,6 nanômetros mostram o quão próximo ou conectadas estão as estruturas do complexo analisados durante uma simulação de dinâmica molecular. Na Figura 15, pode-se observar os valores de contatos moleculares menores que 0,6 nm entre qualquer átomo do dgM2-1 e do P90-110. Observa-se que os valores apresentados se estabilizam depois de aproximadamente 25 ns e permanecem em uma faixa entre 4500 a 5500 contatos moleculares durante toda a simulação, indicando que em nenhum momento o dgM2-1 e o peptídeo se separam na caixa de simulação. Na Figura 15B, observa-se a média dos contatos

moleculares para a interação M2-1/P90-110, para o qual é possível ver um patamar de estabilidade alcançado por volta de aproximadamente 5000 contatos.

Figura 15. Contatos intermoleculares para a interação M2-1/P90-110. Em A temos os valores de contatos intermoleculares no eixo y e o tempo de simulação no eixo x das 4 dinâmicas da proteína M2-1. Em B temos os valores médios de contatos intermoleculares.



Fonte: Próprio autor.

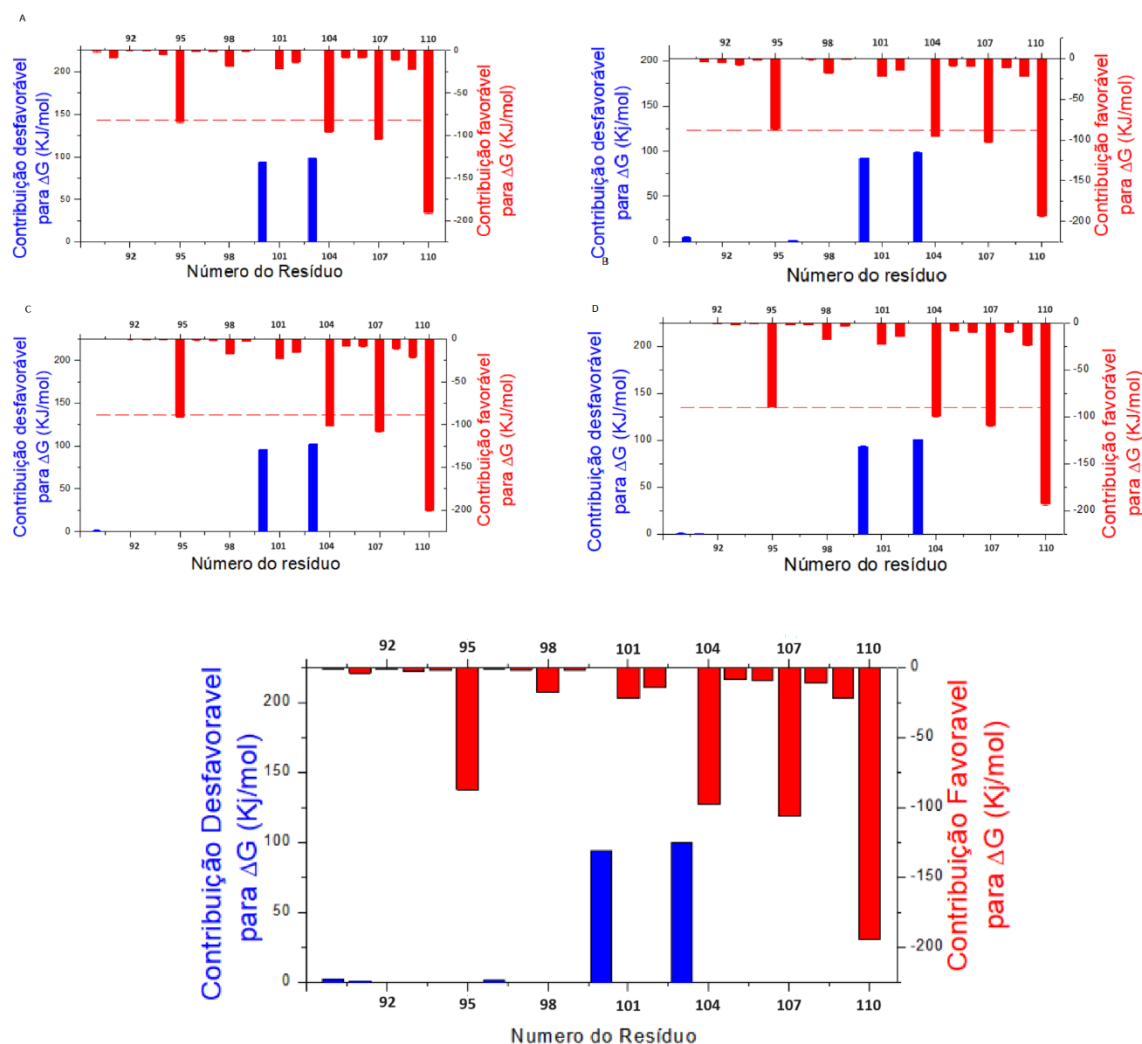
As quatro simulações de dinâmica molecular realizadas fornecem uma boa amostragem do sistema dgM2-1/P90-110. Os resultados de RMSD, na média, mostram valores considerados estáveis e indicam quais partes da proteína e do peptídeo possuem maior flexibilidade, com a parte estruturada em alfa-hélice mais estável conformacionalmente. Juntando esses resultados de RMSD com os dados de número de hidrogênio (~6 em média) e de contatos moleculares

(por volta de 4500 a 5000), pode-se concluir que o sistema molecular dgM2-1/P90-110 se apresentou bem-comportado estruturalmente, com as moléculas (proteína e peptídeo) em contato e interagindo ao longo de todas as simulações.

5.3 Análise de MM-PBSA e MM-GBSA

Utilizando cálculos de energia livre de ligação via os métodos de MM-PBSA e MM-GBSA foram determinadas as contribuições de cada resíduo de aminoácido do P90-110 para a interação dgM2-1/peptídeo. Na Figura 16, pode-se observar a contribuição de cada resíduo do P90-110 para a interação com a proteína M2-1 a partir do método MM-PBSA, ser destacando uma contribuição favorável para os resíduos em vermelho ou uma não favorável para os resíduos em azul. Para destacar os resíduos de aminoácidos mais relevantes foi demarcada nos gráficos da Figura 16 uma linha tracejada que representa a média dos valores das energias de ligação para as contribuições favoráveis.

Figura 16. Valores de energia de ligação (KJ/mol) calculados pelo método PBSA para a interação M2-1/P90-110. Em azul estão os resíduos com contribuição desfavorável à interação e em vermelho os resíduos com contribuição favorável à interação. O gráfico inferior representa a média dos resultados das dinâmicas dos monômeros da interação M2-1/P90-110.

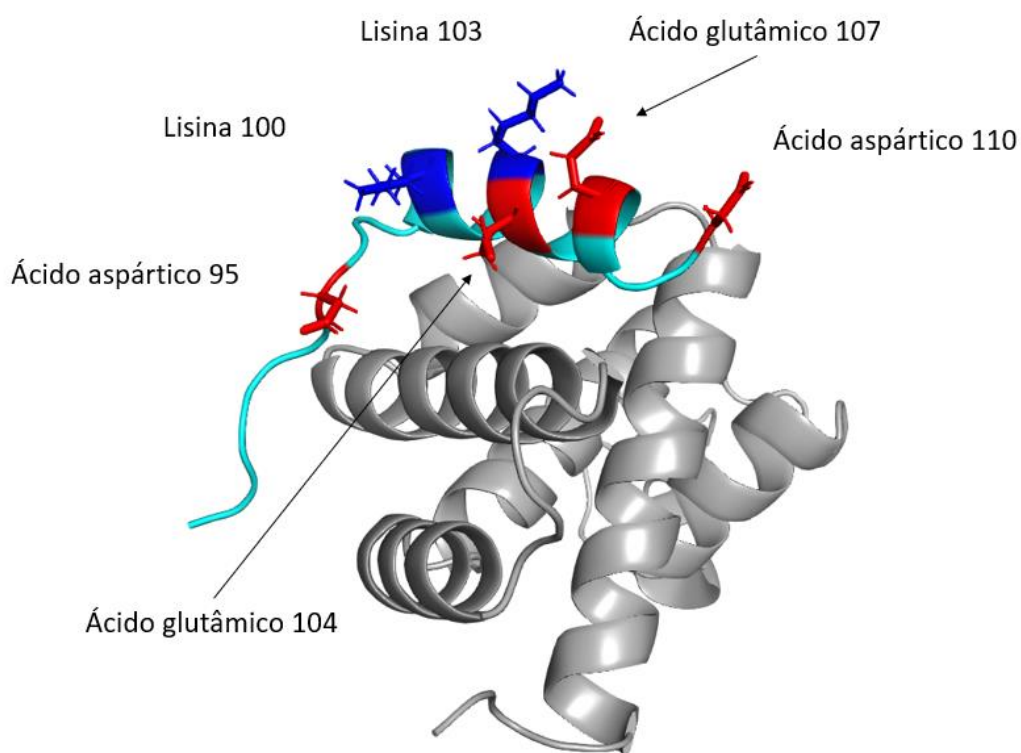


Fonte: Próprio autor.

Os resultados do cálculo de energia livre de ligação via MM-PBSA mostram que os resíduos Asp95, Glu104, Glu107 e Asp110 do peptídeo foram os mais energeticamente favoráveis para a interação, enquanto os resíduos de aminoácidos Lys100 e Lys103 foram os mais desfavoráveis. É evidente a contribuição favorável para a energia livre de ligação dos resíduos carregados negativamente nos cálculos de MM-PBSA. Para os resultados da média dos valores de contribuição de energia dos resíduos do P90-110 para cada dinâmica, observou-se o mesmo resultado, corroborando a contribuição favorável para a energia livre de ligação dos resíduos carregados negativamente a partir dos cálculos de MM-PBSA. A Figura 17 mostra detalhadamente as estruturas dos quatro complexos dgM-21/P90-110 com o peptídeo colorido

em escala de cores do azul para o vermelho, passando pelo branco, correspondendo as contribuições energeticamente desfavoráveis até as favoráveis, respectivamente. É possível observar pela Figura 17 que os resíduos mais favoráveis e desfavoráveis energeticamente do P90-110 indicado pelo método MM-PBSA encontram-se na região de conformação em alfa-hélice do peptídeo (exceto Asp95) e são expostos ao solvente, ou seja, não estão voltados para a interface de interação com o dgM2-1. Vale mencionar que a carga líquida do dgM2-1 é positiva e do P90-110 é negativa e, portanto, o método MM-PBSA está refletindo uma interação de caráter puramente eletrostático destacando mais evidentemente os resíduos de aminoácidos carregados do peptídeo. Entretanto, sabe-se por resultados experimentais publicados na literatura que resíduos hidrofóbicos voltados para a interface de interação dgM2-1/P90-110 desempenham um papel significativo para a formação e estabilidade do complexo (MASON et al, 2003). Esses resíduos hidrofóbicos são discutidos na sessão 5.6 à frente no texto.

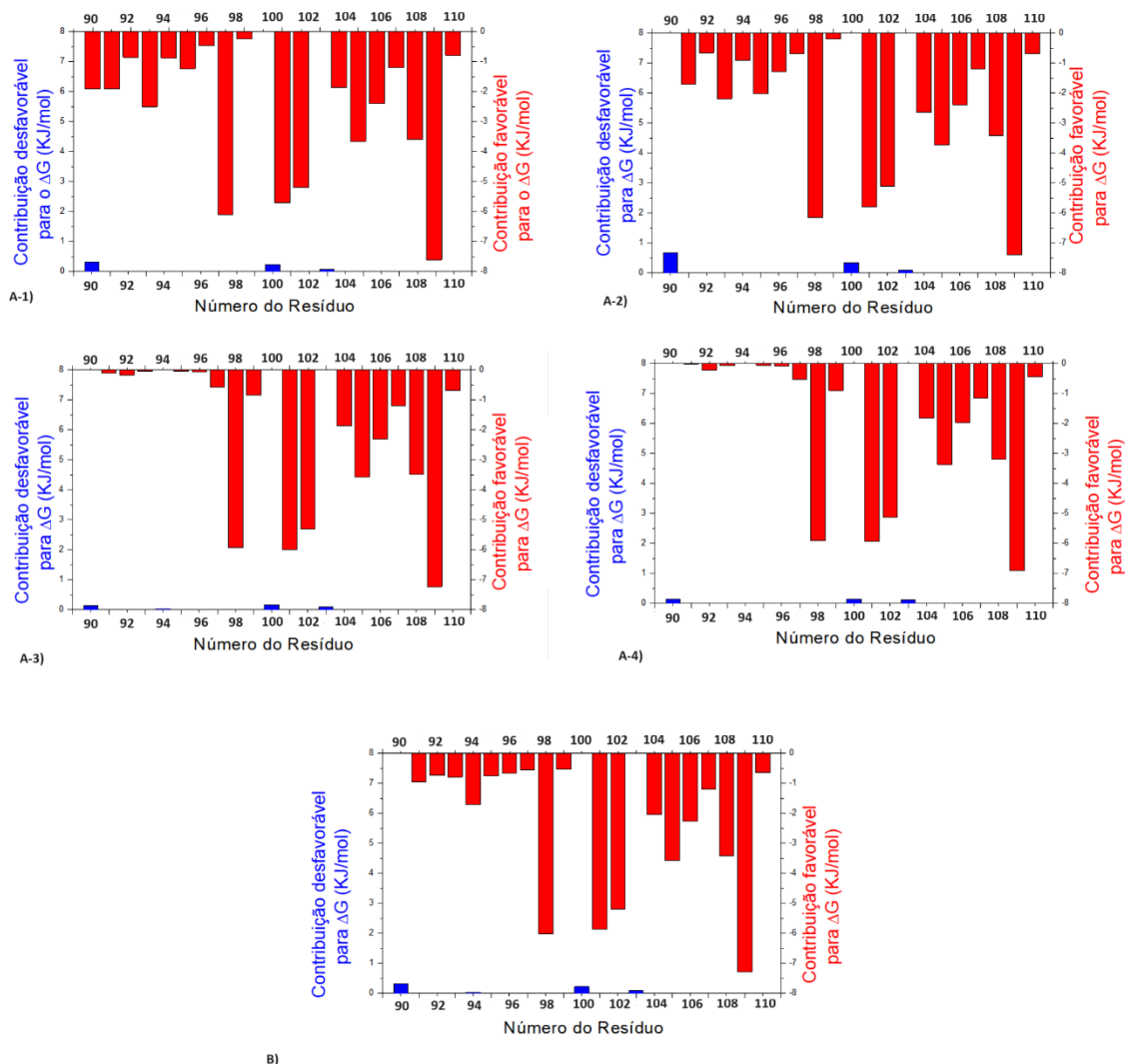
Figura 17. Resíduos de aminoácidos em destaque pelo método PBSA para a interação M2-1/P90-110.



Fonte: Próprio autor

Com o método MM-GBSA para cálculo de energia livre de ligação foram encontrados os valores de contribuição energética de cada resíduo de aminoácido do P90-110 para a interação. Na Figura 18 é apresentada a contribuição de cada resíduo do fragmento da proteína P90-110 para a interação com o dgM2-1, com as contribuições energeticamente favoráveis denotadas em vermelho e as desfavoráveis em azul.

Figura 18. Valores de energia de ligação (kJ/mol) calculados pelo método GBSA para a interação M2-1/P90-110. Em azul estão representados os resíduos com contribuição desfavorável à interação e em vermelho os resíduos com contribuição favorável à interação. As figuras em A) são os resultados para as réplicas da dinâmica molecular do complexo M2-1/P90-110, já a figura em B) é a média dos resultados encontrados em A).

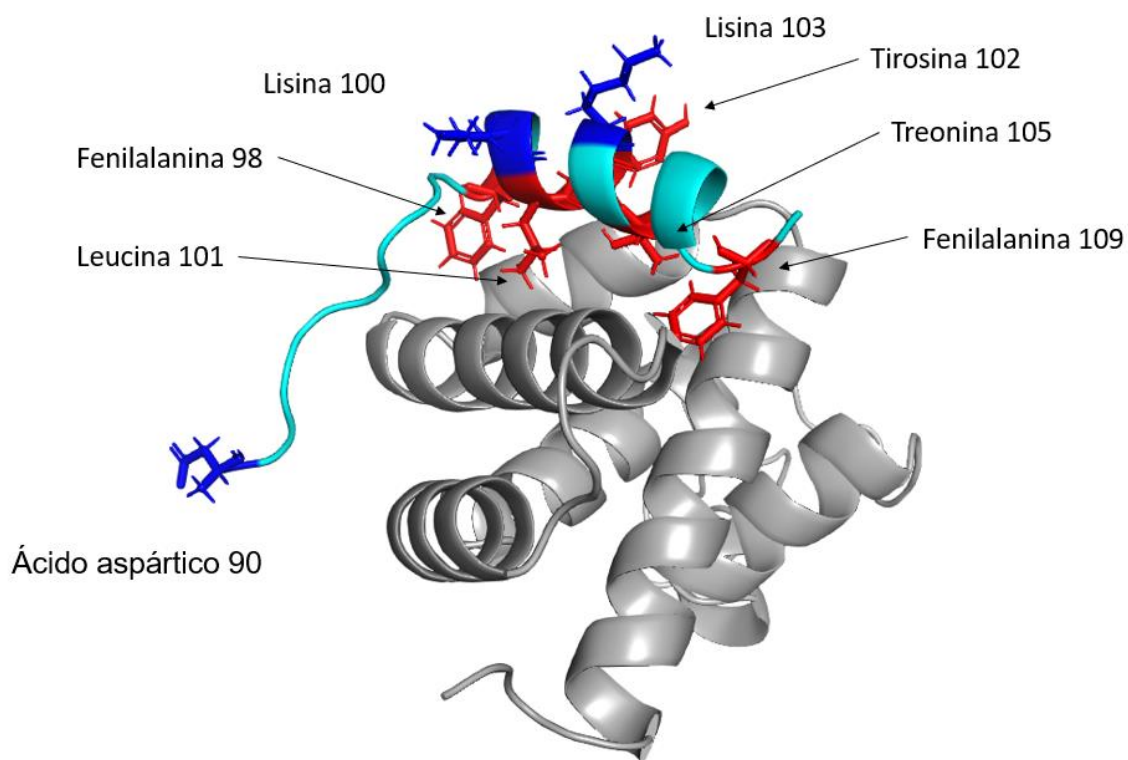


Fonte: Próprio autor.

Os resultados dos cálculos de energia livre por MM-GBSA mostram um perfil diferente dos cálculos por MM-PBSA. Para os cálculos via MM-PBSA, não há somente resíduos polares carregados contribuindo (favorável ou desfavorável) significativamente para a energia livre de ligação total como era o caso para os cálculos com MM-PBSA, nos quais o caráter da interação eletrostática foi muito evidenciado. Os resíduos que se destacaram de forma favorável energeticamente, com valores acima da média (-2,4594 kJ/mol), foram os resíduos Phe98, Leu101, Tyr102, Thr105 e Phe109. Nota-se que o valor absoluto dos resíduos de contribuição favorável é significativamente maior que para os resíduos de contribuição desfavorável para a interação, como é o caso dos resíduos Asp90, Lys100 e Lys103. Ao se observar o

posicionamento dos resíduos de contribuição favorável para a energia livre de ligação, nota-se que para o cálculo de MM-GBSA esses resíduos encontram-se na interface de interação com o dgM2-1, voltados para a fenda formada pelas hélices $\alpha 4$, $\alpha 5$ e $\alpha 6$ (Figura 19), enquanto os resíduos de contribuição desfavorável estão expostos para o solvente, ou seja, o lado oposto da face de interação (Figura 19).

Figura 19. Resíduos destaques do cálculo de energia de ligação por GBSA. Em cinza é mostrado o domínio globular da proteína M2-1, em ciano o fragmento da proteína P90-110. Em azul são mostrados os resíduos que contribuem desfavoravelmente à interação e em vermelho os resíduos que contribuem de forma favorável à interação.



Fonte: Próprio autor.

5.4 Scan de Alaninas por MM-PBSA e MM-GBSA

Como parte dos métodos computacionais, pode-se realizar um método de avaliação de resíduos de aminoácidos conhecido como varredura de alaninas (do inglês *scan* de alaninas). O *scan* de alanina é um método de substituição sistemática de alaninas, e tem sido particularmente útil para a identificação de epítomos funcionais. A substituição dos resíduos de aminoácidos por alanina remove todos os átomos da cadeia lateral após o carbono β , dessa forma o papel dos

grupos funcionais da cadeia lateral pode ser avaliado a partir de mutações para alanina (MORRISON e WEISS, 2001).

Após o estudo das simulações de dinâmica molecular dos quatro complexos dgM2-1/P90-110 (quatro réplicas), suas trajetórias foram concatenadas e foi realizado uma análise de clusterização (corte de 4 Å), da qual obteve-se uma estrutura representativa advinda de um único *cluster* que em seguida foi empregada nas subseqüentes simulações computacionais de dinâmica molecular e cálculos de energia livre para forma selvagem e mutada (alanina) do P90-110. A varredura de alanina (*scan* de alanina) para o P90-110 complexado ao dgM2-1 da estrutura representativa foi realizada substituindo os resíduos de aminoácidos da alfa-hélice do peptídeo para alanina, um por vez, a fim de verificar qual mutação de resíduo impacta mais significativamente nos resultados de cálculo de energia livre de ligação via os métodos de MM-PBSA e MM-GBSA. Cada estrutura representativa do complexo dgM2-1/P90-110 com uma mutação de alanina entre os resíduos de Pro97 e Phe109 da região de alfa-hélice do peptídeo foi submetida a simulações de dinâmica molecular de 250 ns e a partir dessas trajetórias os cálculos de variação de energia livre de ligação $\Delta G_{\text{ligação}}$ foram reavaliados pelos métodos de MM-PBSA e MM-GBSA. A partir dos de $\Delta G_{\text{ligação}}^{\text{PBSA}}$ e $\Delta G_{\text{ligação}}^{\text{GBSA}}$ obtidos, os valores de $\Delta \Delta G_{\text{ligação}}$ foram calculados de acordo com a seguinte equação (PEREIRA, 2012):

$$\Delta \Delta G_{\text{ligação}} = \Delta G_{\text{ligação}}^{\text{Mutado}} - \Delta G_{\text{ligação}}^{\text{Selvagem}} \quad (17)$$

sendo que $\Delta G_{\text{ligação}}^{\text{Mutado}}$ e $\Delta G_{\text{ligação}}^{\text{Selvagem}}$ é a variação de energia livre de ligação calculada pelo método MM-PBSA ou MM-GBSA para a trajetória do complexo dgM2-1/P90-110 com o peptídeo mutado e selvagem, respectivamente. Para valores de $\Delta \Delta G > 0$, interpreta-se que a mutação (por alanina) foi energeticamente desfavorável, enquanto valores de $\Delta \Delta G < 0$ caracteriza-se como uma mutação energeticamente favorável. Dessa forma, é possível identificar, a partir de um *scan* de alaninas realizado de modo computacional, os resíduos de aminoácidos do P90-110 mais revelantes energeticamente para a interação com o dgM2-1, assim como os resíduos propícios a outras mutações diferentes de alanina que poderiam melhorar a ligação do domínio ao peptídeo mimético do P90-110.

A Tabela 1 apresenta os resultados para os cálculos de variação de energia livre de ligação do *scan* de alanina realizados pelo método MM-PBSA. Os resíduos polares negativos

Glu104 e Glu107 do P90-110 se destacam com valores significativos de $\Delta\Delta G > 0$ para a mutação por alanina, enquanto valores predominantes de $\Delta\Delta G < 0$ são evidentes para os resíduos polares positivos Lys100 e Lys103. Assim como visto anteriormente para as contribuições energéticas favoráveis e desfavoráveis para os resíduos do peptídeo na seção anterior, o método de MM-PBSA resalta uma importância significativa para os resíduos carregados refletindo um caráter de contribuição eletrostática para o complexo dgM2-1/P90-110. Dessa forma, a mudança para alanina dos resíduos básicos Lys100 e Lys103 favorecem a interação dgM2-1/P90-110, enquanto dos resíduos ácidos Glu104 e Glu107 desfavorecem a ligação proteína/peptídeo. Esses resultados via MM-PBSA indicam que peptídeos miméticos do P90-110 com mutação das lisinas por resíduos não carregados, como a alanina, poderiam apresentar uma afinidade de interação maior com a proteína em comparação com o peptídeo selvagem.

Tabela 1. Resultados dos cálculos de energia livre de ligação via MM-PBSA para a interação dgM2-1/peptídeos com mutações dos resíduos de aminoácidos para alanina.

resíduo mutado	$\Delta G_{\text{ligação}}$ (kJ/mol)	$\pm$ Erro (kJ/mol)	$\Delta\Delta G$ (kJ/mol)	$\pm$ Erro (kJ/mol)
nenhum	-807,201	5,242		
P97A	-807,962	6,146	-0,761	8,077
F98A	-780,869	6,145	26,332	8,077
S99A	-806,544	5,188	0,657	7,375
K100A	-984,334	5,965	-177,133	7,941
L101A	-747,164	5,167	60,037	7,360
Y102A	-758,357	5,773	48,844	7,797
K103A	-1003,053	6,195	-195,852	8,115
E104A	-625,811	4,896	181,39	7,172
T105A	-793,661	5,449	13,54	7,561
I106A	-772,876	5,294	34,325	7,450
E107A	-601,842	5,801	205,359	7,818
T108A	-760,924	5,472	46,277	7,577
F109A	-777,455	4,848	29,746	7,140

Fonte: Próprio autor

A Tabela 2 apresenta os resultados para os cálculos de variação de energia livre de ligação do *scan* de alanina realizados pelo método MM-GBSA. É possível observar a partir da Tabela 2 que os valores de $\Delta G_{\text{ligação}}$ para os complexos dgM2-1/P90-110 com os peptídeos mutados para alanina são relativamente próximos aos valores de energia para o sistema com o peptídeo selvagem, ao passo que somente as mutações nos resíduos Leu101 e Phe109 apresentam relevância significativa para se destacarem. Vale mencionar que a Lys103 quase se

aproximou desse patamar de destaque considerando uma significância estatística. Ambas Leu101 e Phe109 do peptídeo foram caracterizados com um valor de $\Delta\Delta G > 0$ quando mutados para alanina, o que indica a importância desses resíduos para a interação dgM2-1/P90-110. Lembrando que Leu101 e Phe109 encontram-se na interface de interação com a proteína, com suas cadeias laterais voltadas para o interior da fenda formada pelas hélices $\alpha 4$, $\alpha 5$ e $\alpha 6$ (Figura 19). Por outro lado, uma possível mutação da Lys103 por um resíduo não carregado, como a alanina, poderia promover uma maior afinidade de ligação do peptídeo com relação a proteína quando se compara com a situação do complexo selvagem, uma vez que a mutação K103A sugere um valor de $\Delta\Delta G < 0$. Lembrando que o resíduo de Lys103 encontra-se com sua cadeia lateral exposta para o solvente no complexo dgM2-1/P90-110 (Figura 19). Esse resultado para a Lys103 corrobora com os dados determinados pelo método de MM-PBSA.

Com base nos resultados dos cálculos de energia livre de ligação para o método de MM-GBSA, sugere-se que exceto L101 e Phe109, os demais resíduos poderiam ser alvos de mutações para se alcançar peptídeos miméticos mais eficientes na ligação ao dgM2-1, principalmente a Lys103. Entretanto, a incerteza estatística nos valores de $\Delta\Delta G$ torna a tomada de decisão quanto uma mutação efetiva e eficiente uma tarefa de tentativa e erro.

Tabela 2. Resultados dos cálculos de energia livre de ligação via MM-GBSA para a interação dgM2-1/peptídeos com mutações dos resíduos de aminoácidos para alanina.

resíduo mutado	$\Delta G_{\text{ligação}}$ (kJ/mol)	$\pm$ Erro (kJ/mol)	$\Delta\Delta G$ (kJ/mol)	$\pm$ Erro (kJ/mol)
nenhum	-96,82	6,02		
P97A	-91,25	6,48	5,57	8,513
F98A	-92,42	10,95	4,40	8,844
S99A	-94,93	9,87	1,89	12,495
K100A	-92,08	5,72	4,74	11,561
L101A	-78,59	5,42	18,23	8,304
Y102A	-92,09	7,42	4,73	8,100
K103A	-105,60	5,97	-8,78	9,554
E104A	-97,43	5,83	-0,61	8,478
T105A	-95,94	7,60	0,88	8,380
I106A	-91,93	8,47	4,89	9,695
E107A	-100,47	7,16	-3,65	10,391
T108A	-95,07	7,88	1,75	9,354
F109A	-85,38	5,51	11,44	9,916

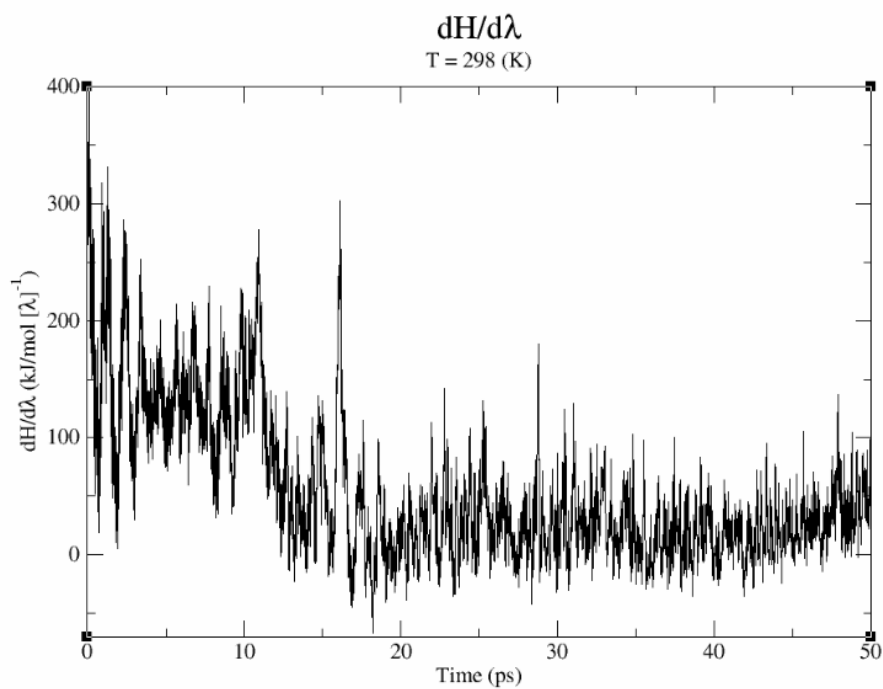
Fonte: Próprio autor.

5.5 Scan de Alanina por Integração Termodinâmica de Crescimento Rápido

Além dos cálculos de variação de energia livre de ligação $\Delta G_{\text{ligação}}$ calculados diretamente pelos métodos MM-PBSA e MM-GBSA para posterior determinação dos valores de $\Delta \Delta G$, também foram realizados cálculos de Integração Termodinâmica de Crescimento Rápido para estimar os valores de $\Delta \Delta G$ só que dessa vez a partir de uma variação de energia livre alquímica do processo de transformação/mutação dos resíduos Phe98-Phe109 do P90-110 através de simulações de dinâmica molecular fora de equilíbrio. Para a aplicação dessa metodologia de obtenção de $\Delta \Delta G$, foi necessário realizar também as simulações de dinâmica molecular do estado selvagem e do estado mutado do peptídeo P90-110 na ausência do dgM2-1. Dessa forma, para a realização de cada mutação do *scan* de alanina foi necessário obter quatro trajetórias, as duas já utilizadas nos métodos de MM-PBSA e MM-GBSA, que são referentes aos complexos dgM2-1/P90-110 com o peptídeo selvagem e mutado por alanina, e agora duas do peptídeo livre com e sem as mesmas mutações.

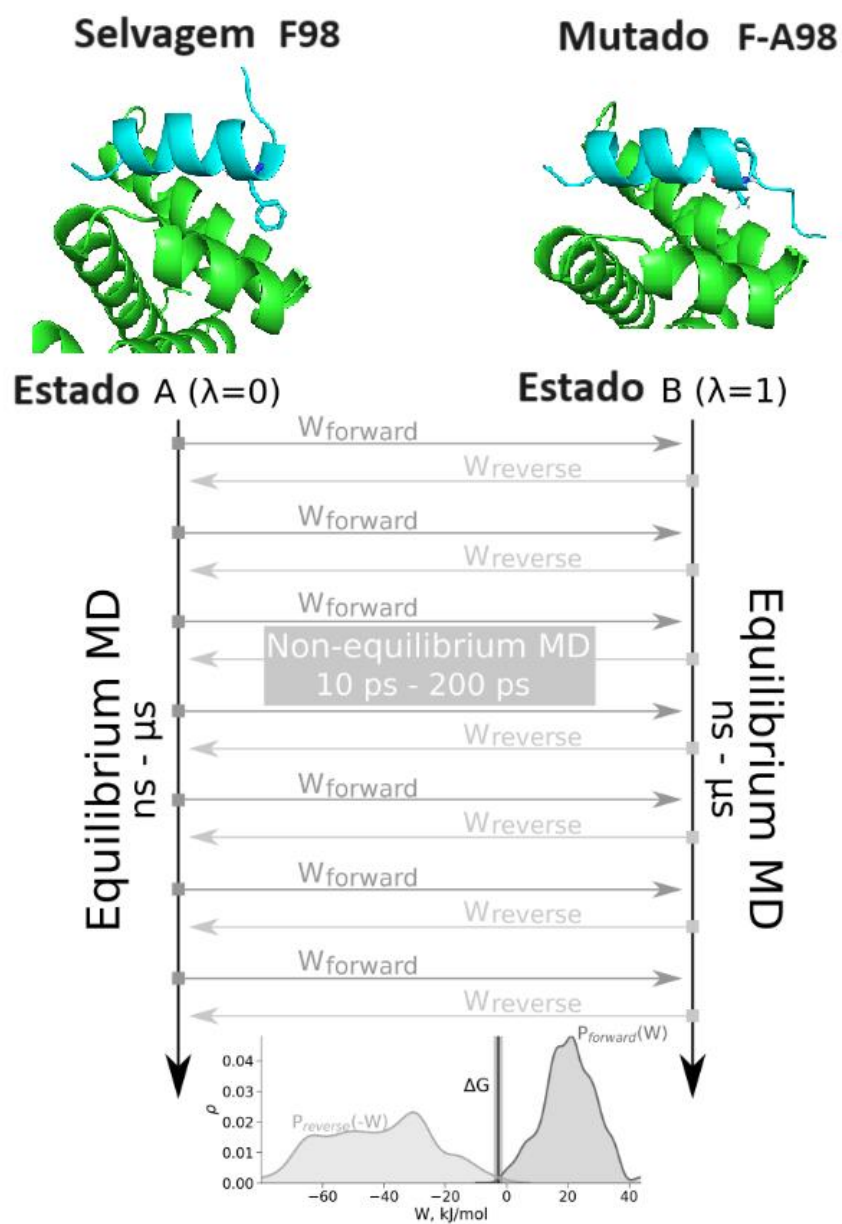
A partir das trajetórias das dinâmicas moleculares dos quatro sistemas, foram extraídos 100 frames ao longo dos 250 ns espaçados igualmente, os quais são a base das simulações de dinâmica molecular fora do equilíbrio e necessários para se calcular o valor de $\Delta \Delta G$ (Figura 20). Dos 100 *frames* extraídos do complexo dgM2-1/P90-110 com o peptídeo selvagem e mutado, 100 simulações fora de equilíbrio de 50 ps foram realizadas para cada estado (selvagem ou mutado) transformando um determinado resíduo em alanina e de alanina para o resíduo selvagem, variando o parâmetro de acoplamento λ da Hamiltoniana (Equação 13) de 0 a 1 e de 1 a 0, respectivamente. Dessa forma o trabalho realizado na transformação alquímica para cada um dos 100 *frames* é calculado através da Equação 12, como exemplificado na Figura 21 para o resíduo Phe98 do P90-110, e a distribuição dos seus valores para o processo direto e reverso proporciona um valor de variação de energia livre alquímica calculado a partir do Teorema de Flutuação de Crooks (Equação 14) e a análise Intersecção Gaussiana de Crooks (Figura 21). Esse conjunto de simulações foram utilizadas para se calcular variação de energia livre alquímica para a transformação do peptídeo complexo do dgM2-1, correspondendo ao ΔG_1 no ciclo termodinâmico da Figura 22. O valor de ΔG_2 corresponde ao mesmo processo mencionado anteriormente só que realizado para as simulações do P90-110 livre, não complexado com o dgM2-1.

Figura 20. Figura da relação da derivada do Hamiltoniano em relação a λ e o tempo de simulação para o frame 100 da dinâmica molecular do complexo mutado no aminoácido F98. No eixo y temos os valores para a derivada do Hamiltoniano pelo λ e no eixo x temos o tempo de simulação, somando 50 picosegundos.



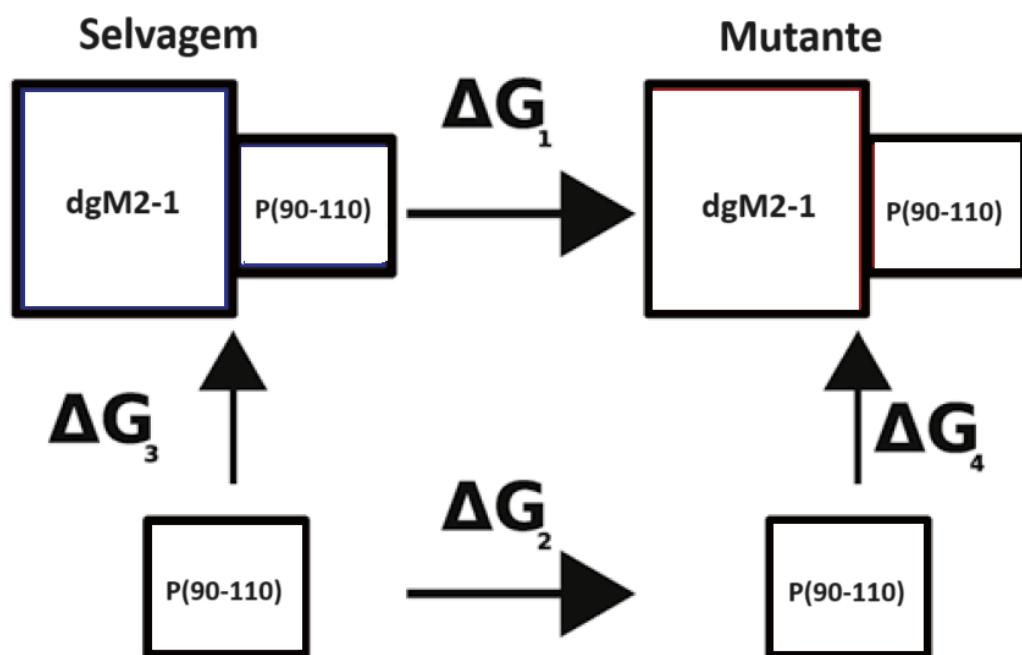
Fonte: Próprio autor.

Figura 21. Demonstração dos passos de dinâmica molecular necessários para resolver o ΔG da mutação do resíduo F98 para alanina.



Fonte: Adaptado de http://pmx.mpibpc.mpg.de/summerSchool2020_tutorial1/index.html#prep

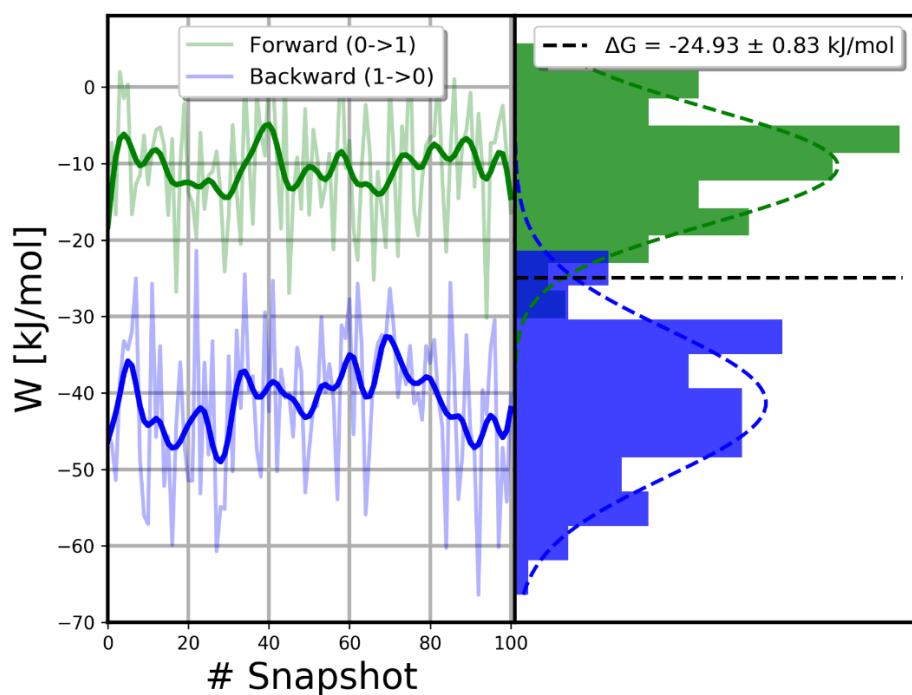
Figura 22. Ciclo termodinâmico da complexação da proteína M2-1 com a P90-110.



Fonte: Adaptado de GAPSYS et al., 2015.

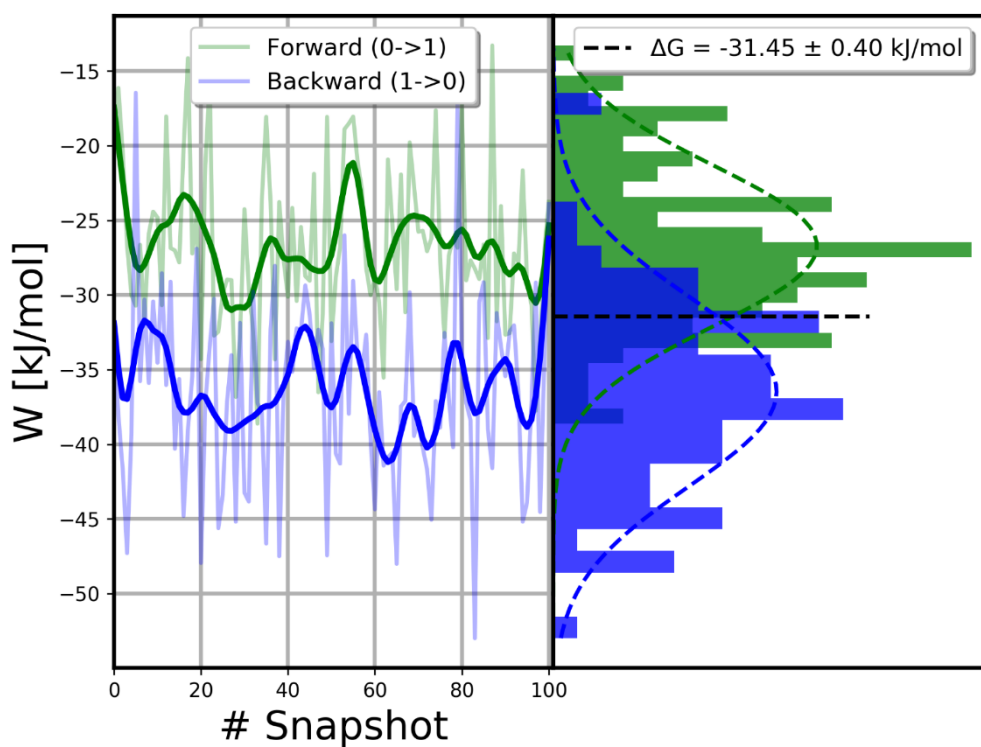
As Figuras 23 e 24 exemplificam um conjunto de dados de valores de trabalho e suas distribuições analisadas a partir da Intersecção Gaussiana de Crooks para a mutação F98A para o complexo dgM2-1/P90-110 e para o peptídeo livre, respectivamente. Os dados coloridos em verde para nas Figuras 23 e 24 representam as simulações fora de equilíbrio realizadas no sentido direto, mutando a Phe98 pela alanina (λ de 0 a 1), enquanto as cores azuis denotam o processo reverso, de A98 para fenilalanina novamente (λ de 1 a 0). No caso dessa mutação F98A, os valores de ΔG_1 e ΔG_2 foram $-24,93 \pm 0,83$ e $-31,45 \pm 0,40$ kJ/mol e, de acordo com o ciclo termodinâmico na Figura 21, $\Delta \Delta G = 6,52 \pm 0,92$ kJ/mol, indicando que a mutação para alanina não foi energeticamente favorável para a interação do P90-110 com o dgM2-1.

Figura 23. Resolução do $\Delta G1$ para o complexo F98 dgM2-1/P90-110 e sua projeção em forma de histograma. O valor de $\Delta G1$ está destacado como uma linha horizontal tracejada.



Fonte: Próprio autor.

Figura 24. Resolução do $\Delta G2$ para o peptídeo P90-110 e sua projeção em forma de histograma. O valor de $\Delta G2$ está destacado como uma linha horizontal tracejada.



Fonte: Próprio autor.

O processo de cálculo $\Delta\Delta G$ pelo método Integração Termodinâmica de Crescimento Rápido foi aplicado aos demais resíduos de aminoácidos de Phe98 a Phe109 da alfa-hélice do P90-110 responsável pela interação com o dgM2-1 e os todos gráficos de distribuição do trabalho direto e reverso juntamente com as respectivas análises de Intersecção Gaussiana de Crooks são apresentados no Anexo A. Os resultados de $\Delta\Delta G$ para o *scan* de alanina usando a Integração Termodinâmica são mostrados na Tabela 3. Vale mencionar que para mutações envolvendo resíduos carregados (positivo ou negativa) foi empregada a configuração de sistema duplo/caixa única para manter a carga total constante ao longo do processo de transformação alquímica e, dessa forma, uma única análise de Intersecção Gaussiana de Crooks foi realizada.

Tabela 3. Resultados encontrados para o scan de alaninas através do método Integração Termodinâmica de Crescimento Rápido.

Mutações	$\Delta G_{Complexo}$ (kJ/mol)	$\Delta G_{Peptídeo}$ (kJ/mol)	$\Delta\Delta G$ (kJ/mol)
F98A	$-24,93 \pm 0,83$	$-31,45 \pm 0,40$	$6,52 \pm 0,92$
S99A	$39,34 \pm 0,22$	$39,05 \pm 0,26$	$0,29 \pm 0,34$
K100A			$-0,23 \pm 2,03$
L101A	$84,19 \pm 1,86$	$57,33 \pm 0,98$	$26,86 \pm 2,10$
Y102A	$66,45 \pm 1,08$	$57,74 \pm 0,59$	$8,71 \pm 1,23$
K103A			$0,08 \pm 1,39$
E104A			$5,23 \pm 3,57$
T105A	$133,58 \pm 0,57$	$121,33 \pm 0,35$	$12,25 \pm 0,66$
I106A	$-28,81 \pm 0,50$	$-38,72 \pm 0,40$	$9,91 \pm 0,64$
E107A			$4,65 \pm 1,04$
T108A	$131,49 \pm 0,74$	$119,10 \pm 0,45$	$11,39 \pm 0,86$
F109A	$-29,15 \pm 2,97$	$-30,00 \pm 0,57$	$0,85 \pm 2,63$

Fonte: Próprio autor.

Ao analisar a Tabela 3 pode-se identificar alguns valores de $\Delta\Delta G$ que se destacam, apresentando relevância nos resultados encontrados. Os maiores valores de $\Delta\Delta G$ encontrados são dos resíduos Leu101, Thr105 e Thr108, seguido de um segundo conjunto representado pelos resíduos de Tyr102 e Ile106. Esses resíduos do peptídeo P90-110 podem ser classificados como essenciais para a interação dgM2-1, pois a mudança deles para alanina uma diferença de variação de energia livre positiva, ou seja, desfavorável.

Por outro lado, dentre os resíduos que chamaram a atenção por um valor de $\Delta\Delta G$ significativamente reduzido, destacam-se a Ser99, Lys100, Lys103 e Phe109. Por esses resíduos apresentarem valores de $\Delta\Delta G \approx 0$, indicando que a mutação para alanina não apresentou uma mudança apreciável para a interação do P90-110 com o dgM2-1. Dessa forma,

esses resíduos do peptídeo podem ser considerados como pontos suscetíveis para mutações dirigidas.

5.6 Sugestões de pontos de mutação com base nos cálculos de energia

A partir dos resultados de *scan* de alanina para os métodos de MM-PBSA, MM-GBSA e Integração Termodinâmica é possível sugerir pontos de mutação na estrutura do P90-110, com base em seu estado complexado com o dgM2-1, que poderiam aumentar a afinidade de interação entre a proteína e o peptídeo. Para embasar essas sugestões de mutações nos resíduos do P90-110, podemos comparar os resultados obtidos no *scan* de alanina do presente trabalho com resultados experimentais publicados na literatura. Mason e colaboradores (2003) avaliaram experimentalmente o efeito de mutações por alanina na fosfoproteína P na interação com a M2-1. Esses resultados de *scan* de alanina experimental são mostrados na Tabela 4 (MASON et al, 2003). Foram encontradas informações interessantes como a redução da afinidade de interação entre as proteínas M2-1 e P quando mutações únicas por alanina foram realizadas nos resíduos Leu101, Tyr102, Glu107, Thr108 e Phe109. Dentre essas mutações as mais relevantes para interação M2-1/P foram nos resíduos Leu101, Tyr102 e Phe109. Mason e coautores também realizaram duplas mutações por alaninas, as quais apresentaram resultados de diminuição da afinidade M2-1/P, como para as duplas Leu101-Tyr102 e Thr108-Phe109, e resultados interessante de aumento da afinidade para mutações nas duplas de resíduos Lys103-Glu104 e Thr105-Ile106. Comparando os resultados computacionais obtidos para os métodos aplicados no presente trabalho, pode-se notar que o MM-GBSA (Leu101 e Phe109, Tabela 2) e a Integração Termodinâmica de Crescimento Rápido (Leu101, Tyr102, Thr105, Ile106 e Thr108, Tabela 3) apresentam resultados mais coerentes com os achados experimentais quanto a relevância de resíduos da P para a interação com a M2-1.

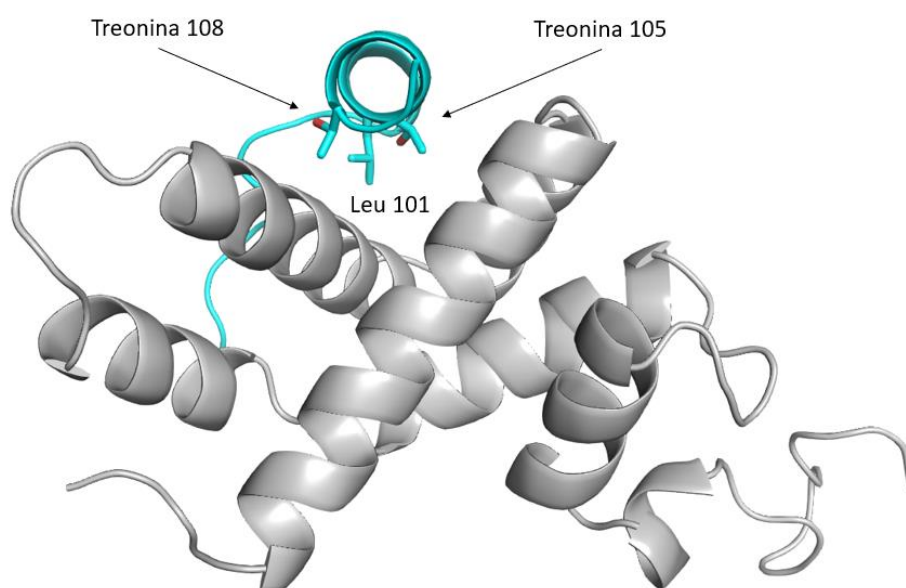
Tabela 4. Efeitos da mutação pontual da proteína P e seu efeito na interação com a proteína M2-1 encontrados por Mason et al, 2003.

Mutações na proteína P	Interação com a proteína M2-1 (%)
LY101/102AA	3
KE103/104AA	263
TI105/106AA	147
E107A	84
TF108/109AA	11
L101A	4
Y102A	4
T108A	77
F109A	3

Fonte: Adaptado de MASON et al, 2003.

Dos resíduos identificados computacionalmente e que corroboram com os resultados experimentais, vale destacar a Leu101, Thr105 e Thr108 do peptídeo P90-110 que se encontram inseridos na fenda formada pelas hélices $\alpha 4$, $\alpha 5$ e $\alpha 6$ na estrutura do dgM2-1, como pode ser visto na Figura 25. O posicionamento e orientação das cadeias laterais desse resíduos, juntamente com os seus valores de $\Delta\Delta G > 0$ no scan de alanina, fazem deles essenciais para a interação dgM2-1/P90-110 e, portanto, possivelmente pontos improváveis para mutação que aumentariam a afinidade de ligação do peptídeo com a proteína.

Figura 25. Resíduos favoráveis à interação dgM2-1/P90-110. Em cinza é representado o domínio globular da proteína M2-1. Em ciano é representado a Fosfoproteína P, onde os resíduos favoráveis são representados na figura.

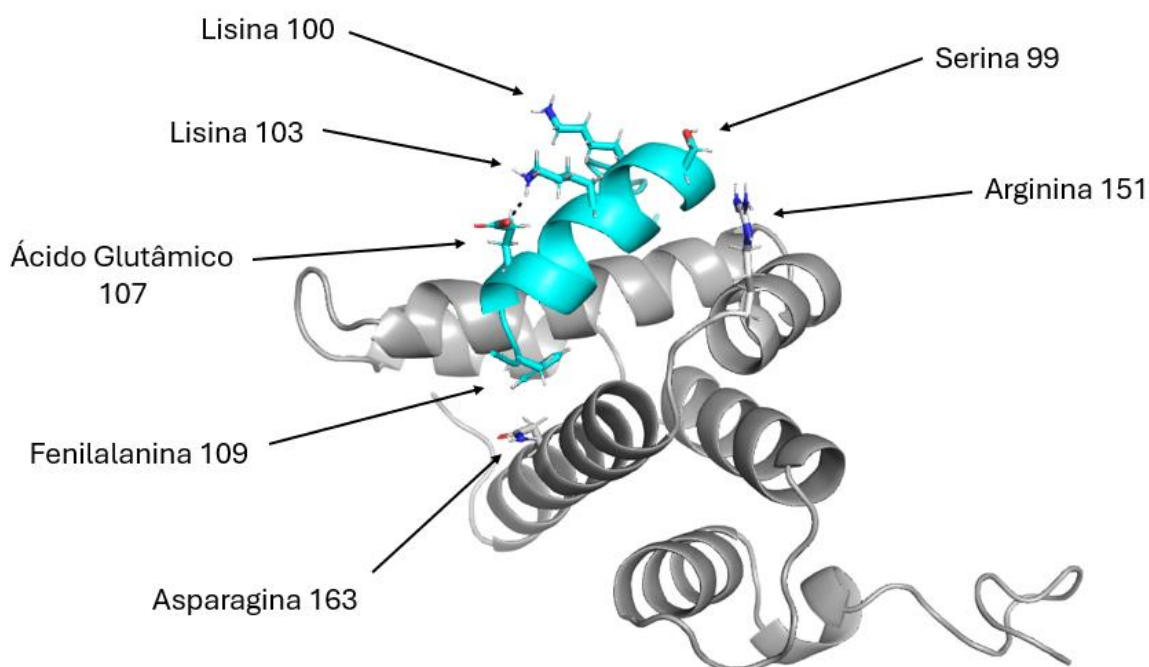


Fonte: Próprio autor.

Apesar dos resultados via MM-GBSA indicarem resíduos essenciais para a interação dgM2-1/P90-110 de modo coerente com os achados experimentais, a utilização desses resultados para nortear a proposta de possíveis pontos de mutação com finalidade de aumentar a afinidade de interação proteína-peptídeo é comprometida, uma vez que para a maioria dos valores de $\Delta\Delta G$ apresentam desvio padrão associado maiores do que o valor absoluto da média. Por outro lado, os dados provenientes do método de Integração Termodinâmica, além de se mostrarem mais coerentes com os resultados experimentais de Mason e colaboradores (2003), proporcionam informações mais concretas e úteis para uma tomada de decisão quanto a escolha de possíveis pontos de mutação que poderiam resultar em valores de $\Delta\Delta G < 0$, ou seja, uma interação para favorável para o peptídeo mimético do P90-110. Dentre esses pontos podemos destacar mutações nos resíduos Ser99, Lys100, Lys103 e Phe109. Entretanto, observando esses resíduos na estrutura representativa do complexo dgM2-1/P90-110, é possível notar que Lys103 participa de contatos polares com Glu107 do peptídeo, sugerindo a ocorrência de uma ponte salina importante para o enovelamento em α -hélice do P90-110 (Figura 26). Dessa forma, descartando Lys103 como um ponto de mutação, pode-se considerar os resíduos Ser99, Lys100 e Phe109 do peptídeo como pontos suscetíveis para mutações dirigidas.

Ao observar o microambiente ao redor do resíduo de Ser99 do P90-110, é possível identificar nas proximidades a Arg151 do dgM2-1 (Figura 26), o que sugere a mutação da Ser99 por resíduo polar ácido, como o ácido aspártico ou glutâmico, que na estrutura do peptídeo mimético poderia realizar ligações de hidrogênio e/ou ponte salina com a proteína. Já para a Lys100, o qual apresenta sua cadeia lateral voltada para o solvente e não participando de contato molecular com resíduos do dgM2-1 (Figura 26), poderia ser substituído por um resíduo polar neutro como asparagina ou glutamina, uma vez que manteriam o caráter hidrofílico dessa região do peptídeo e promoveriam um aumento em sua carga líquida negativa que favoreceria a interação com a domínio positivamente carregado. Apesar do dado computacional para a Phe109 contradizer os resultados experimentais (Tabela 3 e 4), uma mutação conservadora para um resíduo aromático como a tirosina poderia preservar os contatos hidrofóbicos do anel benzênico e incrementar a possibilidade de uma ligação de hidrogênio com resíduos do dgM2-1, como a Asn163 (Figura 26).

Figura 26. Resíduos suscetíveis a mutações planejadas. Em cinza é representado o domínio globular da proteína M2-1. Em ciano é representado a Fosfoproteína P. Os resíduos da proteína P que se destacam são representados em ciano e os resíduos da proteína M2-1 de interesse são representados em cinza. A linha preta tracejada denota um contato polar entre Lys 103 e Glu107 do peptídeo P90-110.



Fonte: Próprio autor.

5.7 Mutações direcionadas na região de alfa-hélice do P90-110

A partir dos pontos de mutação discutidos na seção anterior, foram realizadas mutações direcionadas na região de alfa-hélice do peptídeo P90-110 envolvendo os resíduos Ser99, Lys100 e Phe109, as quais foram avaliadas por cálculos de Integração Termodinâmica de Crescimento Rápido. Vale destacar que também foram realizadas mutações direcionadas por serinas nos resíduos Thr105 e Thr108 do peptídeo, como forma de controle da estratégia aplicada no presente trabalho, e por tirosina no resíduo Phe98, como caráter que uma mutação conservadora que poderia proporcionar a ocorrência de ligações de hidrogênio. Os valores de $\Delta\Delta G$ para essas mutações direcionadas a partir do método ITCR são apresentados na Tabela 5.

Tabela 5. Resultados do cálculo de ITCR para as mutações direcionadas no peptídeo P90-110.

Mutações	$\Delta\Delta G$ (kJ/mol)
F98Y	$-1,70 \pm 0,37$
S99E	$-9,96 \pm 2,65$
S99D	$-3,19 \pm 1,45$
K100Q	$6,40 \pm 1,70$
T105S	$2,53 \pm 0,38$
T108S	$4,60 \pm 0,71$
F109Y	$-3,72 \pm 0,54$
F109W	$5,80 \pm 0,93$
FSF98/99/109YFY	$-17,08 \pm 2,57$

Fonte: Próprio autor.

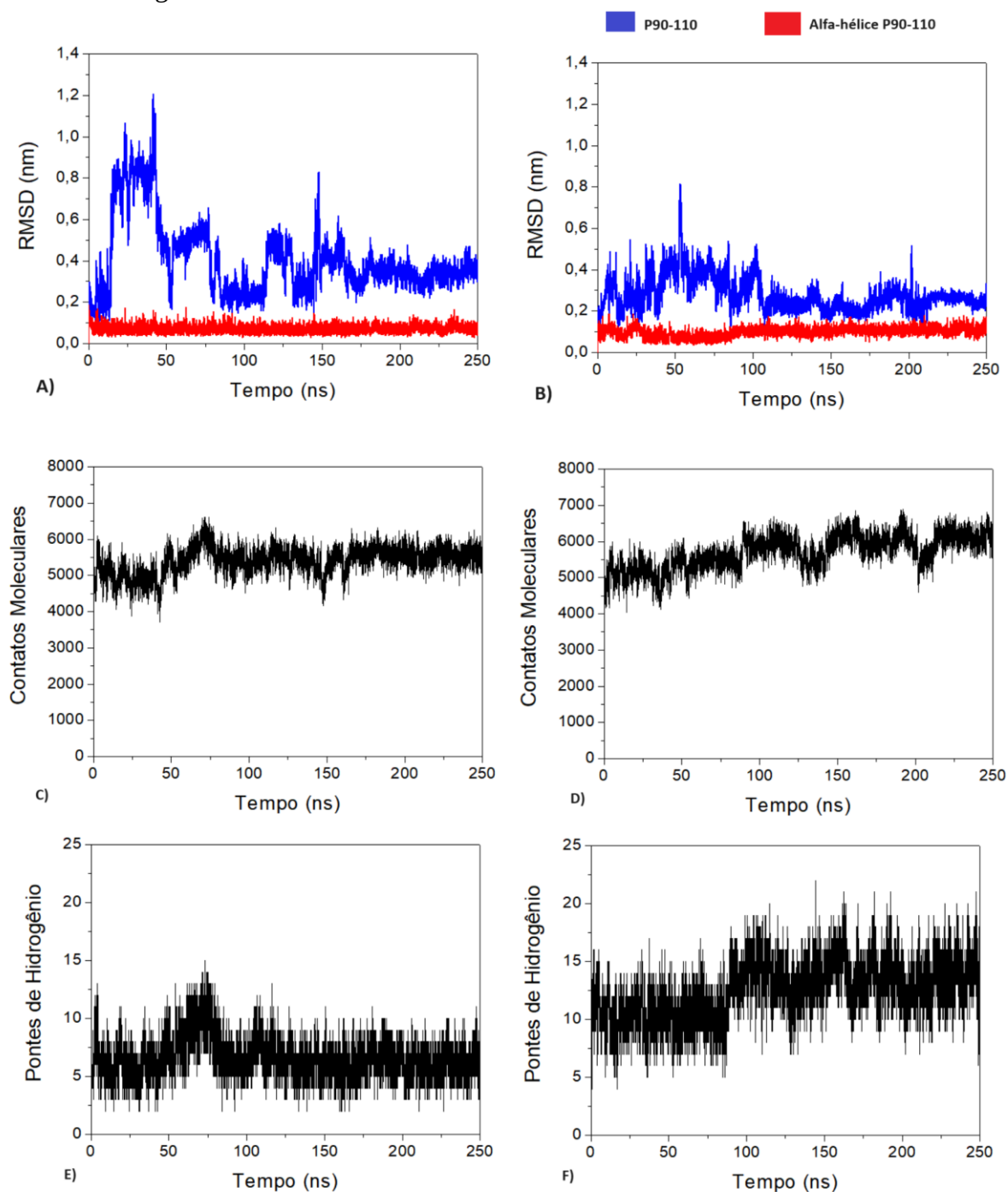
A partir dos resultados na Tabela 5, é possível observar que as mutações direcionadas K100Q, T105S, T108S e F109W do P90-110 apresentaram valores de $\Delta\Delta G > 0$, indicando que as mutações realizadas proporcionaram uma diminuição da afinidade de interação do dgM2-1 com os peptídeos miméticos do P90-110. As mutações de Thr105 e Thr108 por serinas confirmam a importância desses resíduos de treonina para a estabilização do complexo dgM2-1/P90-110. A proposta de substituir a Lys100 por um resíduo polar não carregado, tornando a carga líquida do peptídeo mimético mais negativa, não resultou no efeito esperado de aumento da afinidade de interação entre esse peptídeo mimético do P90-110 com o dgM2-1. Já a tentativa de substituir o anel aromático da Phe109 pelo do triptofano provavelmente resultou em um valor de $\Delta\Delta G > 0$, pois a cadeia lateral desse último é volumosa demais para ser acomodada na posição onde se encontrava a Phe109. Também é possível observar na Tabela 5 valores de $\Delta\Delta G < 0$ para os peptídeos miméticos do P90-110 com as mutações F98Y, S99E, S99D e F109Y, sugerindo um aumento da afinidade entre esses miméticos e o dgM2-1. As mutações das Phe98 e Phe109 por resíduos de tirosinas pode ter proporcionado a formação de ligações de hidrogênio com a hidroxila da cadeia lateral, o que seria uma possível explicação para o aumento de afinidade desses peptídeos miméticos. Já as propostas de substituição da Ser99 por resíduos de ácido glutâmico e aspártico, além de manter o caráter polar com possibilidade de formação de ligações de hidrogênio, proporcionaram cadeias laterais mais comprida o suficiente para estabelecer interações não covalentes mais estáveis (ligações de hidrogênio e ponte salina) com a Arg151 do dgM2-1, como foi idealizado na seção anterior (Figura 26).

Com base no conjunto de peptídeos miméticos do P90-110 com mutações que apresentaram valores de $\Delta\Delta G < 0$ (F98Y, S99E, S99D e F109Y), foi projetado um mimético

com uma tripla mutação FSF98/99/109YEEY que também foi avaliado pelo método ITCR. A tabela 5 mostra o resultado para esse peptídeo mimético, o qual apresenta o valor de $\Delta\Delta G$ mais negativo dentre todos testados, $-17,08$ kJ/mol. É possível notar que uma combinação linear dos valores de $\Delta\Delta G$ para as mutações individuais F98Y, S99E e F109Y resultaria em um total de $-15,38$ kJ/mol, faltando $-1,7$ kJ/mol para completar o valor de $\Delta\Delta G$ determinado para a tripla mutação no cálculo de ITCR. Tal resultado poderia sugerir um tipo de efeito sinérgico entre as três mutações na interação do peptídeo mimético FSF98/99/109YEEY do P90-110 com o dgM2-1.

Para melhor entender o valor negativo expressivo de $\Delta\Delta G$ do peptídeo mimético FSF98/99/109YEEY, a trajetória da dinâmica molecular em equilíbrio do complexo dgM2-1/FSF98/99/109YEEY foi avaliada com base nas análises de RMSD, contatos moleculares e ligações de hidrogênio em comparação com a referência para o complexo com peptídeo P90-110 selvagem. A partir da Figura 27A e 27B é possível notar que os valores de RMSD para a cadeia principal do peptídeo mimético FSF98/99/109YEEY são menores e com menos flutuações do que os observados para o P90-110 selvagem, o que indica um aumento da estabilidade conformacional do mimético na ligação com o dgM2-1, especialmente com relação aos resíduos N-terminais (resíduos 90–97) da região não estrutural em alfa-hélice do peptídeo. Quando os valores de RMSD são computados somente para a cadeia principal da região estrutura em alfa-hélice (resíduos 98–110), pode-se notar valores de RMSD significativamente estáveis, reduzidos e bastante semelhantes, por volta de 1 Å. Os resultados de número de contatos $< 0,6$ nm entre átomos dos peptídeos (selvagem e FSF98/99/109YEEY) e dgM2-1 (Figuras 27C e 27D) mostra que o complexo com o peptídeo mimético apresenta uma maior contagem de contatos na média (~ 5710) quando comparado com o complexo formado com o peptídeo selvagem (~ 5409). O mesmo achado é observado para a contagem do número de ligações de hidrogênio entre os peptídeos e o dgM2-1 (Figura 27E e 27F), que para o mimético FSF98/99/109YEEY apresenta um valor médio de ~ 12 enquanto para o selvagem é de ~ 7 . A quantificação do número de contatos moleculares e ligações de hidrogênio corroboram com a análise de RMSD, sugerindo que a ligação do peptídeo mimético FSF98/99/109YEEY ao dgM2-1 é mais estável estruturalmente quando comparada com a interação do peptídeo P90-110 selvagem.

Figura 27. Conjunto de figuras de RMSD, Contatos moleculares e ligações de hidrogênio. À esquerda estão os dados da dinâmica com o P90-110 selvagem e a direita o do peptídeo mimético FSF98/99/109Y EY. Nas figuras A e B, temos em azul estão os valores de RMSD para o fragmento todo da fosfoproteína P90-110, e em vermelho a região estruturada em alfa-hélice do peptídeo P90-110. Na figura C e D temos os contatos intermoleculares para a interação M2-1/P90-110 do selvagem e do peptídeo mimético FSF98/99/109Y EY. Em E e F são apresentadas as ligações de hidrogênio para as dinâmicas moleculares entre a M2-1 e P90-110 na forma selvagem e mutada FSF98/99/109Y EY.



Fonte: Próprio autor.

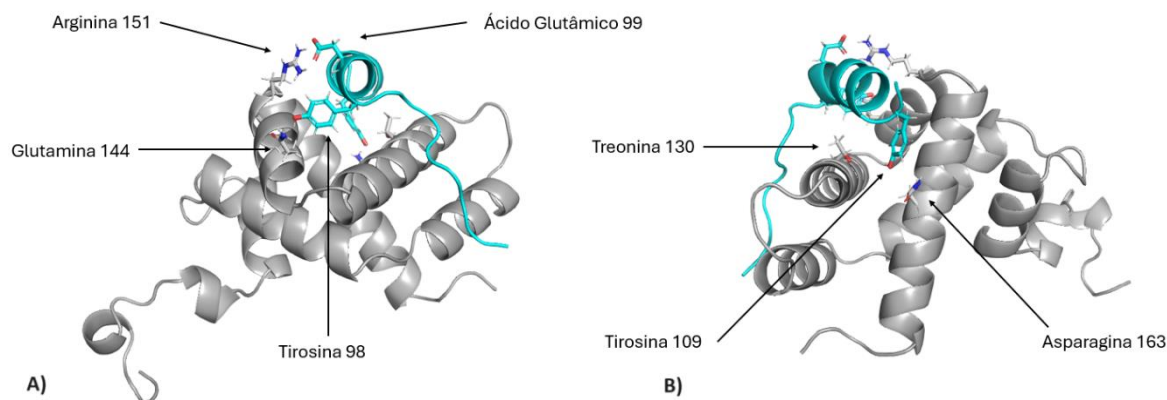
A porcentagem da persistência das ligações de hidrogênio para o complexo dgM2-1/FSF98/99/109YEEY foi avaliada com a finalidade de verificar que as mutações introduzidas de fato proporcionaram a formação de ligações de hidrogênio específicas com os resíduos substituídos. A Tabela 6 mostra a persistência das ligações de hidrogênio para com complexo dgM2-1/FSF98/99/109YEEY, na qual é possível observar que a cadeia lateral do resíduo Glu99 do peptídeo mimético estabelece uma significativa rede de ligações de hidrogênio com a Arg151 do dgM2-1 com valores de porcentagem acima de 50%. Na estrutura representativa do complexo dgM2-1/FSF98/99/109YEEY, a qual foi obtida a partir de uma análise de clusterização da trajetória, é possível notar a proximidade espacial desses dois resíduos, Glu99 do peptídeo mimético e Arg151 do dgM2-1 (Figura 28), de acordo com o que foi inicialmente projetado. As cadeias laterais dos resíduos de tirosinas mutados Tyr98 e Tyr109 do peptídeo mimético também estabelecem ligações de hidrogênio com porcentagem de persistências significativas com Gln144 (cadeia principal) e Thr130 e Asn163 (cadeias laterais de ambas), respectivamente. A proximidade espacial entre esses resíduos também é possível de ser observada na estrutura representativa do complexo (Figura28).

Tabela 6. Resultados do cálculo de ligações de hidrogênio entre o domínio globular da proteína M2-1 e o peptídeo P90-110 modificado com a tripla mutação. As ligações de hidrogênio envolvendo os resíduos modificados estão destacadas em negrito, e os nomes destacados com um asterisco indicam um resíduo do P90-110.

Doador	Aceitador	Persistência (%)
Arg126-side-NE	Glu104-side-OE2 *	47,44
Arg126-side-NH1	Glu104-side-OE1*	77,62
Lys159-side-NZ	Asp110-side-OC1*	17,45
Lys159-side-NZ	Asp110-side-OC2*	16,11
Tyr134-side-OH	Thr105-side-OG1*	98,7
Asn96-side-ND2*	Ser133-side-OG	67,91
Arg126-side-NH1	Glu104-side-OE2*	86,66
Asp95-Main-N*	Ser133-Side-OG	50,9
Asn96-Main-N*	Ser133-side-OG	28,32
Arg126-Side-NE	Glu104-Side-OE1*	64,32
Asn163-Side-ND2	Tyr109-Side-OH *	17,28
Arg151-Side-NH2	Glu99-Side-OE2 *	57,36
Arg151-Side-NH1	Glu99-Side-OE2 *	71,8
Arg151-Side-NH1	Glu99-Side-OE1 *	70,24
Arg151-Side-NH2	Glu99-Side-OE1 *	59,69
Lys159-Side-NZ	Tyr109-Main-O *	11,5
Arg126-Side-NH1	Glu107-Side-OE1*	23,96
Arg126-Side-NH1	Glu107-Side-OE2*	24,39
Tyr98-Side-OH *	Gln144-Main-O	66,09
Arg115-Side-NH1	Asp95-Side-OD2*	22,15
Arg115-Side-NH1	Asp95-Side-OD1*	22,44
Tyr109-Side-OH *	THR130-Side-OG1	11,99
Thr92-Side-OG1*	Asp109-Side-OD1	15,08
Thr92-Main-N*	Asp109-Side-OD2	14,1
Thr92-Side-OG1*	Asp109-Side-OD2	16,96
Thr92-Main-N*	Asp109-Side-OD1	13,36
Thr92-Main-N*	Ser108-Side-OG	11,7
Thr92-Side-OG1*	Ser108-Side-OG	15,22
Asp90-Main-N*	Asn107-Side-OD1	13,21

Fonte: Próprio autor; Software VMD, através do programa de cálculo de ligações de hidrogênio.

Figura 28. Interação entre o peptídeo mimético FSF98/99/109YEEY e o domínio globular da proteína M2-1. Em ciano é representado o FSF98/99/109YEEY e em cinza o domínio globular da M2-1. A Figura A e B referem-se ao mesmo complexo M2-1/P, apenas mudando o ângulo de observação.



Os resultados da análise da trajetória do complexo dgM2-1/FSF98/99/109YEEY em relação a do complexo selvagem mostram um aumento da estabilidade conformacional na ligação do peptídeo mimético com o dgM2-1 em comparação com aquela desenvolvida pelo peptídeo selvagem, o que corrobora com o valor de $\Delta\Delta G$ expressivamente negativo calculado pelo método de ITCR. Além disso, é possível observar uma participação significativa dos resíduos de aminoácidos mutados do peptídeo mimético na formação de ligações de hidrogênio que contribuem para a formação e estabilização com complexo. Isso demonstra uma eficiência na estratégia empregada no presente trabalho para o desenho racional desse peptídeo mimético FSF98/99/109YEEY do P90-110, que apresenta computacionalmente uma maior afinidade de ligação ao domínio globular da proteína M2-1 do hRSV.

6 Conclusão e Perspectivas

Compreender como ocorre a interação M2-1/P é um ponto importante para entender a replicação do hRSV. No presente trabalho, foi realizado um estudo da interação do domínio globular da M2-1 (dgM2-1) com um fragmento da fosfoproteína P (resíduos 90 a 110, nomeado P90-110) por métodos computacionais. A modelagem do ambiente de interação foi possível pela obtenção do complexo dgM2-1 do banco de dados de proteínas PDB (código de acesso 6G0Y). O complexo foi modelado para os quatro dgM2-1 das cadeias da proteína M2-1 e simulações de dinâmica molecular foram realizadas por 250 ns para cada um. Em seguida, as trajetórias foram concatenadas e foi realizada uma análise de clusterização que forneceu o cluster de maior representatividade para os próximos passos. A estratégia de *scan* de alanina foi feita a partir de simulações de dinâmica molecular em equilíbrio de 250 ns para cada mutação dos resíduos da alfa-hélice do peptídeo P90-110, com o objetivo de realizar uma análise energética pelos métodos de MM-PBSA e MM-GBSA. Por fim, a técnica de Integração Termodinâmica de Crescimento Rápido foi empregada para a realização do *scan* de alanina, com a realização de dinâmicas moleculares fora do equilíbrio para avaliar energeticamente a mudança alquímica.

Para o método de MM-PBSA, os resíduos polares negativos Glu104 e Glu107 do P90-110 se destacaram com valores significativos de $\Delta G > 0$ para a mutação por alanina, enquanto valores predominantes de $\Delta G < 0$ foram evidentes para os resíduos polares positivos Lys100 e Lys103. Dessa forma, pode-se observar que o método MM-PBSA ressalta uma importância exacerbada para os resíduos carregados refletindo um caráter de contribuição eletrostática para o complexo dgM2-1/P90-110, o que pode levar a uma tomada de decisão enviesada com relação a proposta de mutações. Já para o método MM-GBSA, os valores de ΔG_{lig} para os complexos dgM2-1/P90-110 com os peptídeos mutados para alanina são relativamente próximos aos valores de energia para o sistema com o peptídeo selvagem, dessa forma a incerteza estatística nos valores de $\Delta \Delta G$ dificultou a tomada de decisão quanto a mutações a serem realizadas. Somente as mutações em Leu101 e Phe109 apresentam relevância significativa para se destacarem, caracterizando-se por valores de $\Delta \Delta G > 0$ quando mutados para alanina, o que indica a importância desses resíduos para a interação dgM2-1/P90-110. Por fim, para o método de ITCR, pode-se identificar que os maiores valores de $\Delta \Delta G$ encontrados são dos resíduos Leu101, Thr105 e Thr108, seguido de um segundo conjunto representado pelos resíduos de Tyr102 e Ile106. Esses resíduos do P90-110 podem ser classificados como essenciais para a interação com o dgM2-1, pois a mudança deles para alanina promove $\Delta \Delta G > 0$, ou seja,

desfavorável. Esse resultado corrobora com os dados experimentais reportados por Mason e colaboradores, que investigaram a influência de mutações por alanina na interação M2-1/P (MASON et al, 2003). Por outro lado, dentre os resíduos que chamaram a atenção por um valor de $\Delta\Delta G$ significativamente reduzido, destacam-se a Ser99, Lys100 e Phe109 do peptídeo como pontos suscetíveis para mutações dirigidas.

A partir dos pontos suscetíveis para mutações dirigidas indicados pelo método ITCR, peptídeos miméticos do P90-110 foram gerados e avaliados pelos cálculos de dinâmica molecular fora de equilíbrio. Dentre esses peptídeos miméticos se destacaram as mutantes F98Y, S99E, S99D e F109Y do P90-110, os quais juntos abriram caminho para a proposta do mutante FSF98/99/109YEEY que mais eficientemente se ligou ao dgM2-1. O peptídeo mimético FSF98/99/109YEEY do P90-110 apresentou um valor significativo de $\Delta\Delta G$ de -17,08 kJ/mol de acordo com o método ITCR, assim como as análises de RMSD, contatos moleculares e ligações de hidrogênio da trajetória de seu complexo com o dgM2-1 reportaram uma estabilidade conformacional maior em comparação com o peptídeo P90-110 selvagem. A análise de persistência das ligações de hidrogênio mostrou os resíduos mutados no peptídeo FSF98/99/109YEEY como protagonistas na interação dgM2-1/peptídeo mimético, evidenciando que a estratégia computacional adotada no presente trabalho para o desenho racional de um peptídeo mimético do P90-110 foi eficiente.

Os dados gerados pelo presente estudo apontam resíduos de importância na interação dgM2-1/P90-110 e em alguns casos corroboram com dados já obtidos experimentalmente. Com os resultados obtidos propõem-se desenvolver um peptídeo mimético capazes de competir no sítio de ligação M2-1/P. Através de mutações planejadas no peptídeo P90-110 encontrou-se um peptídeo mimético FSF98/99/109YEEY que apresenta uma interação mais eficiente do que o selvagem, como também apresenta um comportamento conformacional mais estável com relação ao peptídeo P90-110. Com esses resultados propõem-se o peptídeo mimético FSF98/99/109YEEY como um possível inibidor/bloqueador da interação M2-1/P, na tentativa de interromper o ciclo viral do hRSV, que pode ser testado experimentalmente. Além disso, o conhecimento gerado no presente trabalho pode ser útil para futuros estudos sobre as proteínas M2-1 e P e contribuir na busca por antivirais capazes de combater a infecção causada pelo hRSV.

REFERÊNCIAS

- AHERNE W, BIRD T, COURT SD, et al., Pathological changes in virus infections of the lower respiratory tract in children. **J Clin Pathol.**, 1970, 23, 7–18.
- ABRAHAM, Mark James, Teemu Murtola, Roland Schulz, Szilárd Páll, Jeremy C. Smith, Berk Hess, and Erik Lindah. 2015. “Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers.” *SoftwareX* 1–2 (September). **Elsevier** B.V.: 19–25. doi:10.1016/j.softx.2015.06.001.
- ALMEIDA, J. R. G. S., LIMA, J. T., NUNES, X. P., CERQUEIRA, G. S., FIGUEIREDO, C. S. Flavonóides da dieta e sua importância para a saúde humana. **Revista de Ciências da Saúde Nova Esperança**, v.2, p.65-76, 2004. antibody against respiratory syncytial virus fusion protein. *PloS one*. v. 14, n. 2, 2019.
- ALVAREZ, L. H. **Identification and Characterization of Cruzain Allosteric Inhibitors: A Computer-Aided Approach**. Dissertação (Mestrado em Biofísica Molecular) - UNESP-IBILCE, São José do Rio Preto, 2017.
- ASENJO, A. et al. Phosphorylation of the human respiratory syncytial virus N protein provokes a decrease in viral RNA synthesis. **Virus research**, v. 163, n. 1, p. 396- 400, 2012.
- ASENJO, A.; CALVO, E.; VILLANUEVA, N. Phosphorylation of human respiratory syncytial virus P protein at threonine 108 controls its interaction with the M2-1 protein in the viral RNA polymerase complex. **Journal of general virology**, v. 87, n. 12, p. 3637-3642, 2006.
- BAILLY, B. et al. Targeting human respiratory syncytial virus transcription antitermination factor M2-1 to inhibit in vivo viral replication. **Scientific reports**, v. 6, n. 1, p. 1-11, 2016.
- BAKER, Nathan A., David Sept, Simpson Joseph, Michael J. Holst, and J. Andrew McCammon. 2001. “Electrostatics of Nanosystems: Application to Microtubules and the Ribosome.” *Proceedings of the National Academy of Sciences of the United States of America* 98 (18). **National Academy of Sciences**: 10037–41. doi:10.1073/pnas.181342398.
- BERENDSEN, H.J., VAN DER SPOEL, D., VAN DRUNEN, R. GROMACS: a message-passing parallel molecular dynamics implementation. **Computer physics communications**, 91(1-3), pp.43-56. 1995.
- BITKO, Vira et al. Nonstructural proteins of respiratory syncytial virus suppress premature apoptosis by an NF- κ B-dependent, interferon-independent mechanism and facilitate virus growth. **Journal of virology**, v. 81, n. 4, p. 1786-1795, 2007.
- BURKREYEV, A. et al. Taxonomy of the order Mononegavirales: update 2016. **Virology Division News**, p. 2351-2360, 2016.

BLONDOT, M-L. et al. Struture and Functional Analysis of the RNA- and Viral Phosphoprotein-Binding Domain of Respiratory Syncytial Virus M2-1 Protein. **PLoS Pathog.** V. 8, n. 5, e1002734, doi:10.1371/journal.ppat.1002734, 2012.

BOHMWALD, K. et al. **Human Respiratory Syncytial Virus:** Infection and Pathology.

BORCHERS, Andrea T. et al. Respiratory syncytial virus—a comprehensive review. **Clinical reviews in allergy & immunology**, v. 45, p. 331-379, 2013.

BRASIL. Ministério da Saúde. Secretaria de Ciência, Tecnologia, Inovação e Insumos Estratégicos em Saúde. Departamento de Assistência Farmacêutica e Insumos Estratégicos. **Uso do anticorpo monoclonal Palivizumabe durante a sazonalidade do Vírus Sincicial Respiratório - VSR.** Brasília : Ministério da Saúde, 2022. 11 p. : il. ISBN 978-65-5993-197-2.

BROWN, D. 1.07 - Overview of Sources of New Drugs. In: TAYLOR, J. B.; TRIGGLE, D. J. (Eds.). **Comprehensive Medicinal Chemistry II.** Oxford: Elsevier, 2007. p. 321–353.

BROWN, G. et al. Evidence for na association between heat Shock protein 70 and the respiratory syncytial vírus polymerase complex within lipid-raft membranes during vírus infection. **Virology**, v. 338, n. 1, p. 69-80, 2005.

BROOKS, C. L. III; KARPLUS, M.; PETTITT, B. M., Proteins: A theoretical perspective of dynamics structure and Thermodynamics, Advances in chemical physics, **John Wiley & Sons**, New York, 1988.

BEHZADI, Mohammad Amin; LEYVA-GRADO, Victor H. Overview of current therapeutics and novel candidates against influenza, respiratory syncytial virus, and Middle East respiratory syndrome coronavirus infections. **Frontiers in microbiology**, v. 10, p. 1327, 2019.

CARROMEU, C. **Estudos de Caracterização Estrutural, Localização Intracelular e Imunogenicidade da RNA Polimerase do Vírus Respiratório Sincicial Humano.** 128 f. Tese (Doutorado em Biotecnologia) – Interunidades em Biotecnologia – Universidade de São Paulo / Instituto Butantan / Instituto de Pesquisas Tecnológicas, São Paulo, 2007.

CARTEE, T. L., MEGAW, A. G., OOMENS, A. G. P. & WERTZ, G. W. Identification of a single amino acid change in the human respiratory syncytial virus L protein that affects transcriptional termination. **Journal of virology**, v.77, p. 7352–7360, 2003.

COLLINS, P. L.; GRAHAM, B. S. Viral and host factors in human respiratory syncytial virus pathogenesis. **Journal of virology**, v. 82, n. 5, p. 2040-2055, 2008.

COLLINS, P. L.; MELERO J. A. Progress in Understanding and Controlling Respiratory Syncytial Virus: Still Crazy After All These Years. **Virus Research**, v. 162, p. 80-99, 2011.

COLLINS, P. L.; Chanock, R. M.; Murphy, B. R.; Respiratory syncytial virus. In **“Fields Virology”** (D.M.Knipe, and P.M.Howley, Eds.), 4th ed., pp. 1443–1485. Lippincott, Williams, and Wilkins, Philadelphia, PA. 2001

CURRAN, J.; MARQ, J. B.; KOLAKOFSKY, D. An N-terminal domain of the Sendai paramyxovirus P protein acts as a chaperone for the NP protein during the nascent chain assembly step of genome replication. **Journal of Virology**, v. 69, p. 849-855, 1995.

DEVINCENZO, J.; LAMBKIN-WILLIAMS, R.; WILKINSON, T.; CEHELISKY, J.; NOCHUR, S.; WALSH, E.; MEYERS, R.; GOLLOB, J.; VAISHNAW, A.; A randomized, double-blind, placebo-controlled study of an RNAi-based therapy directed against respiratory syncytial virus. **Proc Natl Acad Sci U. S. A.** 2010; 107(19):8800-5.

DELANO, Warren L. 2002. “The PyMOL Molecular Graphics System.” DeLano Scientific San Carlos (CA): USA.

DIAS, T. D. **Interação entre tropomiosina e as proteínas de matriz e fosfoproteína do Virus Respiratório Sincicial Humano.** 2017. 68 f. Dissertação (Mestrado em Microbiologia) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2017.

DOWELL, S. F. et al. Respiratory syncytial virus is an important cause of community-acquired lower respiratory infection among hospitalized adults. **Journal of Infectious Diseases**, v. 174, n. 3, p. 456-462, 1996.

FIELDS, B. N.; KNIPE, D. M.; HOWLEY, P. M.; GRIFFIN, D. E. **Fields virology.** Philadelphia, Lippincott Williams & Wilkins, 2001.

FIZ, J.; GALLOUX, M.; BLONDOT, M.; ELÉQUÊT, J-F. The Insertion of Fluorescent Proteins in a Variable Region of Respiratory Syncytial Virus L polymerase Results in Fluorescent and Functional Enzymes But with Reduced Activities. **The Open Virology Journal**, v. 5, p. 103-108, 2011.

FOLOPPE, N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods? **Curr. Med. Chem.** 2006, 13, 3583–3608.

FUENTES, Sandra et al. Function of the respiratory syncytial virus small hydrophobic protein. **Journal of virology**, v. 81, n. 15, p. 8361-8366, 2007.

GAN, S. W. (nomes) structure and ion channel activity of the human respiratory syncytial virus (hRSV) small hydrophobic protein transmembrane domain. **Protein Science**, v. 17, n. 5, p. 813-820, 2008.

GAPSYS, V.; MICHIELSENS, S.; PETERS, de Groot, B. L.; LEONOV, H.; Calculation of Binding Free Energies. **Molecular Modeling of proteins**: 2nd edition. Book Series: Methods in Molecular Biology 1215: 173-209 (2015).

GAPSYS, Vytautas, Servaas Michielssens, Jan Henning Peters, Bert L. de Groot, and Hadas Leonov. 2015. “Calculation of Binding Free Energies.” Methods in Molecular

Biology (Clifton, N.J.) 1215. *Methods Mol Biol*: 173–209. doi:10.1007/978-1-4939-1465-4_9.

GAPSYS, Vytutas, Servaas Michielssens, Daniel Seeliger, and Bert L. de Groot. 2016. “Accurate and Rigorous Prediction of the Changes in Protein Free Energies in a Large-Scale Mutation Scan.” *Angewandte Chemie International Edition* 55 (26). John Wiley & Sons, Ltd: 7364–68. doi:10.1002/ANIE.201510054.

GARCIA, J.; GARCIA-BARRENO, B.; VIVO, A.; MELERO, J. A. Cytoplasmic inclusions of respiratory syncytial virus-infected cells: formation of inclusion bodies in transfected cells that co express the nucleoprotein, the phosphoprotein, and the 22 k protein. **Virology**, v. 195, p. 243-247, 1993.

GILSON, M. K.; Honig, B. Calculation of the total electrostatic energy of a macromolecular system: solvation energies, binding energies, and conformational analysis. *Proteins* 1988, 4, 7–18.

GUIMARÃES, G. C. et al., *Int. J. Biol. Macromol.* 111, 33–38 (2018).

GLEZEN, W. Paul et al. Risk of primary infection and reinfection with respiratory syncytial virus. **American journal of diseases of children**, v. 140, n. 6, p. 543-546, 1986.

GRAHAM, Barney S.; MODJARRAD, Kayvon; MCLELLAN, Jason S. Novel antigens for RSV vaccines. **Current opinion in immunology**, v. 35, p. 30-38, 2015.

GOETTE, Maik, and Helmut Grubmüller. 2009. “Accuracy and Convergence of Free Energy Differences Calculated from Nonequilibrium Switching Processes.” *Journal of Computational Chemistry* 30 (3). John Wiley & Sons, Ltd: 447–56. doi:10.1002/JCC.21073.

BONCRISTIANI, H. F.; CRIADO, M. F.; ARRUDA, E. Respiratory Viruses. **Encycl. Microbiol.** 500–518. 2009.

HALL CB, WALSH EE, LONG CE, et al., Immunity to and frequency of reinfection with respiratory syncytial virus., **J. Infect. Dis.**, 163, 1991, 693–698.

HARDY, R. W.; WERTZ, G. W.; The Cys(3)-His(1) motif of the respiratory syncytial virus M2-1 protein is essential for protein function. **J Virol.** V. 74, n. 13, 2000. <https://doi.org/10.1128/JVI.74.13.5880-5885.2000>.

HOLTHOFF, T.E.; BURNS, S.T.; WOODLING, K.A. et al. The dietary antioxidant resveratrol is a functional scavenger of peroxynitrite in vitro. **Faseb Jornal.** v 24. p 321.1. doi: 10.1096/fasebj.24.1_supplement.321.1. 2010.

HOMEYER, N.; GOHLKE, H. Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method. *Mol. Inform.* 2012, 31, 114–122.

ISPAS, Gabriela et al. Antiviral activity of TMC353121, a respiratory syncytial virus (RSV) fusion inhibitor, in a non-human primate model. **PloS one**, v. 10, n. 5, p. e0126959, 2015.

HOU, Tingjun et al. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. **Journal of chemical information and modeling**, v. 51, n. 1, p. 69-82, 2011.

JOHNSON, JR PR; OLMSTED, R. A.; PRINCE, G. A.; MURPHY, B. R.; ALLING, D. W.; WALSH, E. E.; COLLINS, P. L.; Antigenic relatedness between glycoproteins of human respiratory syncytial virus subgroups A and B: evaluation of the contributions of F and G glycoproteins to immunity. **J Virol**. 1987; 61(10),3163-66.

JORGENSEN, William L., Jayaraman Chandrasekhar, Jeffry D. Madura, Roger W. Impey, and Michael L. Klein. 1983. "Comparison of Simple Potential Functions for Simulating Liquid Water." **The Journal of Chemical Physics** 79 (2). American Institute of PhysicsAIP: 926–35. doi:10.1063/1.445869.

KARLIN, D. et al. Structural disorder and modular organization in Paramyxovirinae N and P. **Journal of General Virology**, v. 84, n. 12, p. 3239-52, 2003.

KIM, Hyun Wha et al. Respiratory syncytial virus disease in infants despite prior administration of antigenic inactivated vaccine. **American journal of epidemiology**, v. 89, n. 4, p. 422-434, 1969.

Kiss G, Holl JM, Williams GM, Alonas E, Vanover D, Lifland AW, Gudheti M, Guerrero-Ferreira RC, Nair V, Yi H, Graham BS, Santangelo PJ, Wright ER. 2014. Structural analysis of respiratory syncytial virus reveals the position of M2-1 between the matrix protein and the ribonucleoprotein complex. **J Virol** 88:7602–7617. <https://doi.org/10.1128/JVI.00256-14>.

KISS, G. et al. Strutural Analysis of Respiratory Syncytial Virus Reveals the Position of M2-1 between the Matrix Protein and the Ribonucleoprotein Complex. **Journal of Virology**, v. 88, n. 13, p. 7602-7617, 2014.

KNIFE, D. et al. **Fields Virology, Volumes 1 and 2**. Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2013.

KOLLMAN, P. A.; MASSOVA, I.; REYES, C.; KUHN, B.; HUO, S.; CHONG, L.; Lee, M.; LEE, T.; DUAN, Y.; WANG, W.; DONINI, O.; CIEPLAK, P.; SRINIVASAN, J.; CASE, D. A.; CHEATHAM, T. E.; Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. **Accounts of chemical research**, v. 33, n. 12, p. 889-897, 2000.

KUHN, B.; KOLLMAN, P. A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. **J. Med. Chem.** 2000, 43, 3786–3791.

KUMARI, R.; KUMAR, R.; CONSORTIUM, O. S. D. D.; LYNN, A.; g_mmpbsa– A GROMACS tool for High-Throughput MM-PBSA Calculations. **J. Chem. Inf. Model.** 2014, 54, 1951–1962.

KUMARI, Rashmi, Rajendra Kumar, and Andrew Lynn. 2014. “G-Mmpbsa -A GROMACS Tool for High-Throughput MM-PBSA Calculations.” **Journal of Chemical Information and Modeling** 54 (7). American Chemical Society: 1951–62. doi:10.1021/ci500020m.

LEYRAT, C. et al. Drastic changes in conformational dynamics of the antiterminator M2-1 regulate transcription efficiency in Pneumovirinae. **Elife**, v. 3, p. e02674, 2014.

LILJEROOS, Lassi et al. Architecture of respiratory syncytial virus revealed by electron cryotomography. **Proceedings of the National Academy of Sciences**, v. 110, n. 27, p. 11133-11138, 2013.

LLORENTE, M. T. et al. Structural properties of the human respiratory syncytial virus P protein: Evidence for an elongated homotetrameric molecule that is the smallest orthologue within the family of paramyxovirus polymerase cofactors. **Proteins: Structure, Function and Genetics**, v. 72, n. 3, p. 946-958, 2008.

LOURENÇO, I. O. **Estudo da interação da proteína NS1 do vírus sincicial respiratório humano (hRSV) com flavonoides por espectroscopia de fluorescência e dinâmica molecular**. Dissertação (Mestrado em Biofísica Molecular) - UNESP-IBILCE, São José do Rio Preto, 2019.

MACHADO, C.M. Infecção pelo vírus sincicial respiratório e o diagnóstico das infecções virais pós-transplante de células-tronco hematopoéticas no Brasil. **Rev. Bras. Hematol. Hemoter.** São Paulo , v. 31, n. 6, p. 401-402, 2009.

MAIER, James A., Carmenza Martinez, Koushik Kasavajhala, Lauren Wickstrom, Kevin E. Hauser, and Carlos Simmerling. 2015. “Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB.” **Journal of Chemical Theory and Computation** 11 (8). American Chemical Society: 3696–3713. doi:10.1021/acs.jctc.5b00255.

MARTINS, Luan Carvalho; **PYAUTOFEP**: Uma ferramenta de automação para cálculos de FEP para o programa GROMACS integrando técnicas de amostragem estendida. Tese (Doutorado em Bioinformática) - Universidade Federal de Minas Gerais, 2021.

MASON, S. W.; ABERG, E.; LAWETZ, C.; DeLONG, R.; WHITEHEAD, P.; LIUZZI, M.; Interaction between Human Respiratory Syncytial Virus (RSV) M2-1 and P Proteins Is Required for Reconstitution of M2-1-Dependent RSV Minigenome Activity. **journal of virology**, Oct. 2003, p. 10670–10676.

MCDONALD, T. P. et al. Evidence that the Respiratory Syncytial Virus Polymerase Complex Associates with Lipid Rafts in Virus-infected Cells: a Proteomic Analysis. **Virology**, v. 330, p. 147-157, 2004.

MEIROVITCH, Hagai. Recent developments in methodologies for calculating the entropy and free energy of biological systems by computer simulation. **Current opinion in structural biology**, v. 17, n. 2, p. 181-186, 2007.

MELERO, J. A. Molecular Biology of Human Respiratory Virus. In: CANE P (editor). **Perspectives in Medical Virology: Respiratory Syncytial Virus**. Elsevier, p. 01-30, 2007.

MITRA, R. et al. The human respiratory syncytial virus matrix protein is required for maturation of viral filaments. **J Virol**, v. 86, p. 4432-4443, 2012.

MORRISON, K. L.; WEISS, G. A.; Combinatorial alanine-scanning. **Elsevier Science**, 2001.

MUFSON et al. Epidemiology of respiratory syncytial virus infection among infants and children in Chicago. **American Journal of Epidemiology**, v. 98, n. 2, p. 88-95, 1973.

LEUNG, Nancy HL. Transmissibility and transmission of respiratory viruses. **Nature Reviews Microbiology**, v. 19, n. 8, p. 528-545, 2021.

NAIR, H. et al. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. **The Lancet**, v. 375, n. 9725, p. 1545-1555, 2010.

NAVARRO, J.; LOPEZ-OTIN, C.; VILLANUEVA, N. Location of phosphorylated residues in human respiratory syncytial virus phosphoprotein. **Journal of General Virology**, v. 72, n. Pt 6, p. 1455-59, 1991.

Novavax. **Novos dados do ensaio Novavax Fase 3 Prepare™ do ResVax™ apresentados na reunião anual da IDSOG de 2019**. Disponível em: <https://ir.novavax.com/press-releases/2019-08-12-New-Data-from-Novavax-Phase-3-Prepare-TM-Trial-of-ResVax-TM-Presented-at-2019-IDSOG-Annual-Meeting>. Acesso em: 3 out 2023.

OHNO, A. et al. Characterization of Human Respiratory Syncytial Virus Detected in Hospitalized Children in Philipines from 2008 to 2012. **Journal of Clinical Virology**, v. 53, p. 59-65, 2013.

OTHUMPANGAT, Sreekumar et al. NGF is an essential survival factor for bronchial epithelial cells during respiratory syncytial virus infection. **PloS one**, v. 4, n. 7, p. e6444, 2009.

PATEL, Dharmeshkumar, Jagdish Suresh Patel, and F. Marty Ytreberg. 2021. "Implementing and Assessing an Alchemical Method for Calculating Protein–Protein Binding Free Energy." *Journal of Chemical Theory and Computation* 17 (4). NIH Public Access: 2457. doi:10.1021/ACS.JCTC.0C01045.

PAYNTER, S. Humidity and respiratory virus transmission in tropical and temperate settings. **Epidemiology & Infection**, v. 143, n. 6, p. 1110-1118, 2015.

PEREIRA, J. G. de C. **Caracterização dos aminoácidos da interface proteína-proteína com maior contribuição na energia de ligação e sua predição a partir dos dados estruturais**. Dissertação (Mestrado em Genética de Biologia Molecular) - Universidade Estadual de Campinas, Campinas, 2012.

PELAY-GIMENO, M.; Glas, A.; Koch, O.; Grossmann, T. N.; *Angew. Chem. Int. Ed. Engl.* 54, 8896–8927 (2015).

PETERS TR, CROWE JE JR., Respiratory syncytial virus. In: Long SS, editor. **Principles and practice of pediatric infectious diseases**, 3rd edition. Philadelphia, PA: Churchill Livingstone; 2008. 1112–1117.

PILOTO, J. V. **Simulações Moleculares da Interação da Região Conservada de Cisteínas do Ectodomínio da Proteína G do hRSV com o Receptor Celular CX3CR1 do Hospedeiro**. Dissertação (Mestrado em Biofísica Molecular) - UNESP-IBILCE, São José do Rio Preto, 2022.

PIVA, H. M. R. et al., *Int. J. Mol. Sci.* 21 (2020), doi:10.3390/ijms21062241.

PMX: GENERATE HYBRID PROTEIN STRUCTURE AND TOPOLOGY. 2020. Disponível em: http://pmx.mpibpc.mpg.de/summerSchool2020_tutorial1/index.html. Acesso em: 22 de abril de 2024.

RODRÍGUEZ, L.; CUESTA, I.; ASENJO, A.; VILLANUEVA, N. Human respiratory syncytial virus matrix protein is an RNA-binding protein: binding Properties, location and identity of the RNA contact residues. **Journal of General Virology**, v. 85, p. 709-719, 2004.

ESPERANTE, S. A.; PARIS, G.; DE PRAT-GAY, G. Modular unfolding and dissociation of the human respiratory syncytial virus phosphoprotein P and its interaction with the M2–1 antiterminator: a singular tetramer–tetramer interface arrangement. **Biochemistry**, v. 51, n. 41, p. 8100-8110, 2012.

SANTOS, N. S. O.; ROMANOS, M. T. V.; WIGG, M. D.; **Virologia Humana**. Editora Guanabara Koogan. Rio de Janeiro, 2015.

SANTOS, Vanessa Sardinha dos. "**Vírus**"; *Brasil Escola*. Disponível em: <https://brasilecola.uol.com.br/biologia/virus-2.htm>. Acesso em 30 de agosto de 2023.

SEDEYN, Koen; SCHEPENS, Bert; SAELENS, Xavier. Respiratory syncytial virus nonstructural proteins 1 and 2: Exceptional disrupters of innate immune responses. **PLoS pathogens**, v. 15, n. 10, p. e1007984, 2019.

SELVARAJ, M.; YEGAMBARAM, K.; TODD, E. J. A. A.; RICHARD, C. A.; DODS, R. L.; PANGRATIOU, G. M.; TRINH, C. H.; MOUL, S. L.; MURPHY, J. C.; MANKOURI, J.; ÉLÉOUET, J. F.; BARR, J. N.; EDWARDS, T. A.; The Structure of the Human Respiratory Syncytial Virus M2-1 Protein Bound to the Interaction Domain of the Phosphoprotein P Defines the Orientation of the Complex. **American Society for Microbiology**, Volume 9, 2018.

SILVESTRI, Giuseppe, Federica Arrigoni, Francesca Persico, Luca Bertini, Giuseppe Zampella, Luca De Gioia, and Jacopo Vertemara. 2023. "Assessing the Performance of Non-Equilibrium Thermodynamic Integration in Flavodoxin Redox Potential Estimation." *Molecules* 28 (16). Multidisciplinary Digital Publishing Institute (MDPI): 6016. doi:10.3390/MOLECULES28166016/S1.

SHIN, W. H.; Christoffer, C. W.; Kihara, D.; *Methods*. 131, 22–32 (2017).

BOHMWALD, Karen et al. Human respiratory syncytial virus: infection and pathology. In: **Seminars in Respiratory and Critical Care Medicine**. Thieme Medical Publishers, 2016. p. 522-537.

SHAPIRO, A. B. et al. Quantitative Investigation of the Affinity of the Human Respiratory Syncytial Virus Phosphoprotein C-terminus Binding to Nucleocapsid Protein. **Virology Journal**. V. 19, p. 01-12, 2014.

SPEHNER, D.; DRILLIEN, R.; HOWLEY, P. M. The assembly of the measles virus nucleoprotein into nucleocapsid-like particles is modulated by the phosphoprotein. **Virology**, v. 9, n. 232(2), p. 260-8, 1997.

SUN, Huiyong et al. Assessing the performance of MM/PBSA and MM/GBSA methods. 7. Entropy effects on the performance of end-point binding free energy calculation approaches. **Physical Chemistry Chemical Physics**, v. 20, n. 21, p. 14450-14460, 2018.

TANG, Roderick S. et al. Requirement of cysteines and length of the human respiratory syncytial virus M2-1 protein for protein function and virus viability. **Journal of virology**, v. 75, n. 23, p. 11328-11335, 2001.

TANNER, S. J. et al. Crystal structure of the essential transcription antiterminator M2-1 protein of human respiratory syncytial virus and implications of its phosphorylation. **PNAS**, v. 11, n. 4, p 1580-1585, 2014.

TEIXEIRA, T. S. P. et al., *Int. J. Biol. Macromol.* 95, 63–71 (2017).

VALDÉS-TRESANCO, M. S.; Valdés-Tresanco, M. E.; Valiente, P. A.; Moreno, E.; gmx_MMPBSA: A New Tool to Perform End-State Free Energy Calculations with GROMACS. *J. Chem. Theory Comput.* 2021, 17, 10, 6281–6291.
<https://doi.org/10.1021/acs.jctc.1c00645>

VERLI, H. *Bioinformática: da biologia à flexibilidade molecular*. 2014.

VIVO, M. D. et al. Role of Molecular Dynamics and Related Methods in Drug Discovery. **Journal of Medicinal Chemistry**, v. 59, n. 9, p. 4035–4061, 12 maio 2016.

VAN GUNSTEREN, Wilfred F.; BERENDSEN, Herman JC. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. **Angewandte Chemie International Edition in English**, v. 29, n. 9, p. 992-1023, 1990.

WALSH EE, PETERSON DR, FALSEY AR, Risk factors for severe respiratory syncytial virus infection in elderly persons. **J Infect Dis**, 189, 2004, 233–238.

WANG, X.; Ni, D.; Liu, Y.; Lu, S.; *Front. Chem.* 9, 294 (2021).

WARR, W. A. A CADD-alog of strategies in pharma. **Journal of Computer-Aided Molecular Design**, v. 31, n. 3, p. 245–247, 18 mar. 2017.

WLODAWER, A.; Annu. Rev. Med. 53, 595–614 (2002).

YANG, L. et al. Age-specific epidemic waves of influenza and respiratory syncytial virus in a subtropical city. **Scientific reports**, v. 5, p. 10390, 2015.

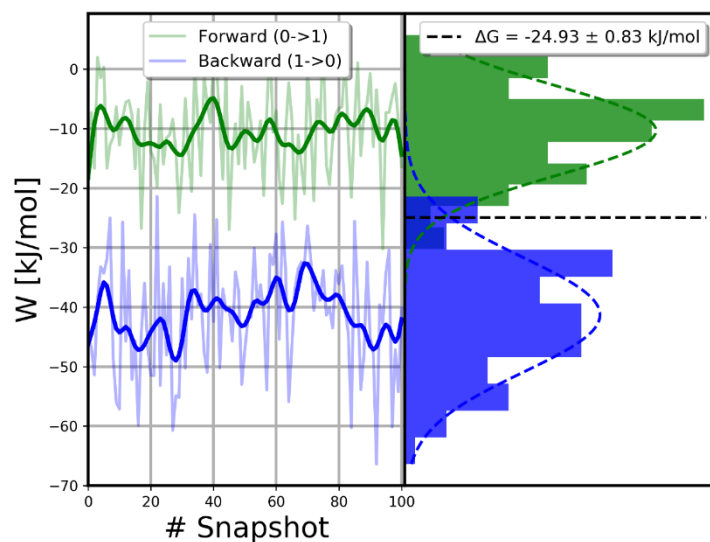
ZHANG, P. et al. Characterization of Recombinant Influenza A Virus as a Vector Expressing Respiratory Syncytial Virus Fusion Protein Epitopes. **Journal of General Virology**, v. 95, p. 1886-1891, 2014.

ZHANG, P. et al. Characterization of Recombinant Influenza A Virus as a Vector Expressing Respiratory Syncytial Virus Fusion Protein Epitopes. **Journal of General Virology**, v. 95, p. 1886-1891, 2014.

ZLATEVA, K. T.; LEMEY, P.; MOES, E.; VANDAMME, A. M.; VAN RANST, M.; Genetic variability and molecular evolution of the human respiratory syncytial virus subgroup B attachment G protein. **J Virol**. 2005; 79(14):9157-67.

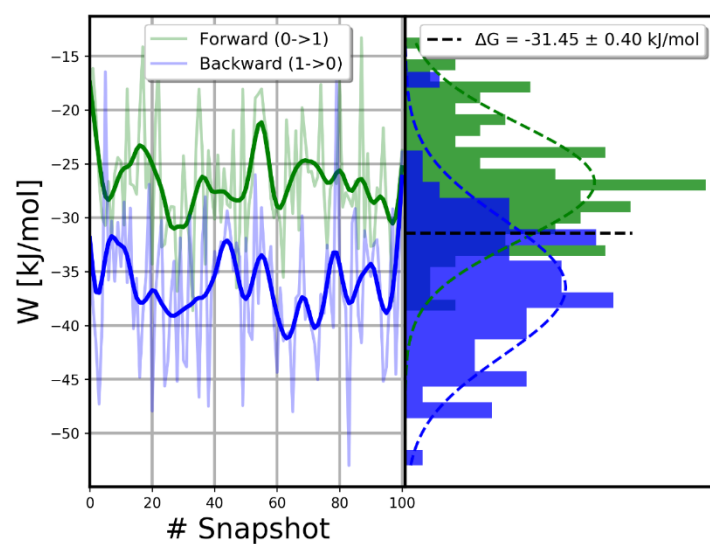
ANEXO A - Cálculo de ΔG por Integração Termodinâmica de Crescimento Rápido para o Scan de alaninas

Figura 1. Mutação F98A do P90-110 no contexto do complexo com dgM2-1.



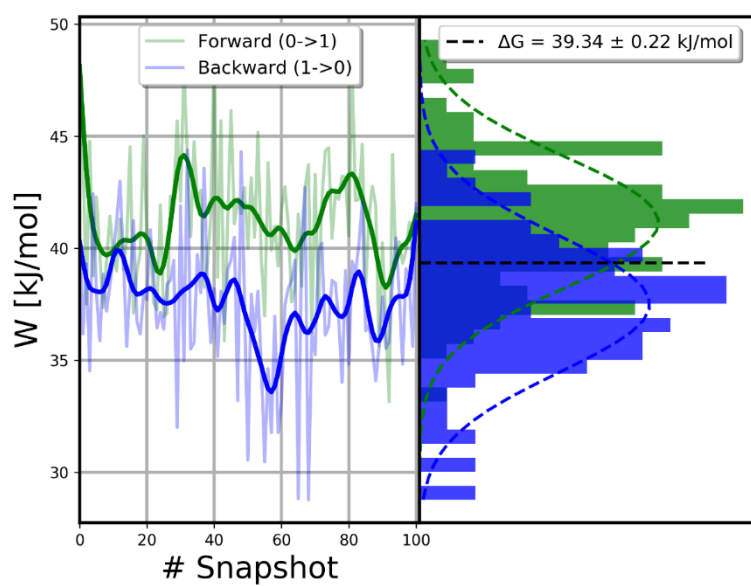
Fonte: Próprio autor

Figura 2. Mutação F98A do P90-110 na ausência do dgM2-1.



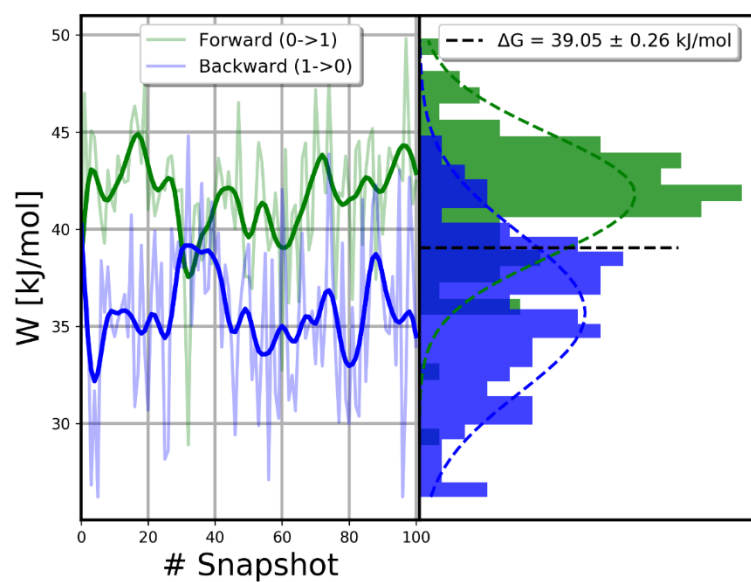
Fonte: Próprio autor

Figura 3. Muta  o S99A do P90-110 no contexto do complexo com dgM2-1.



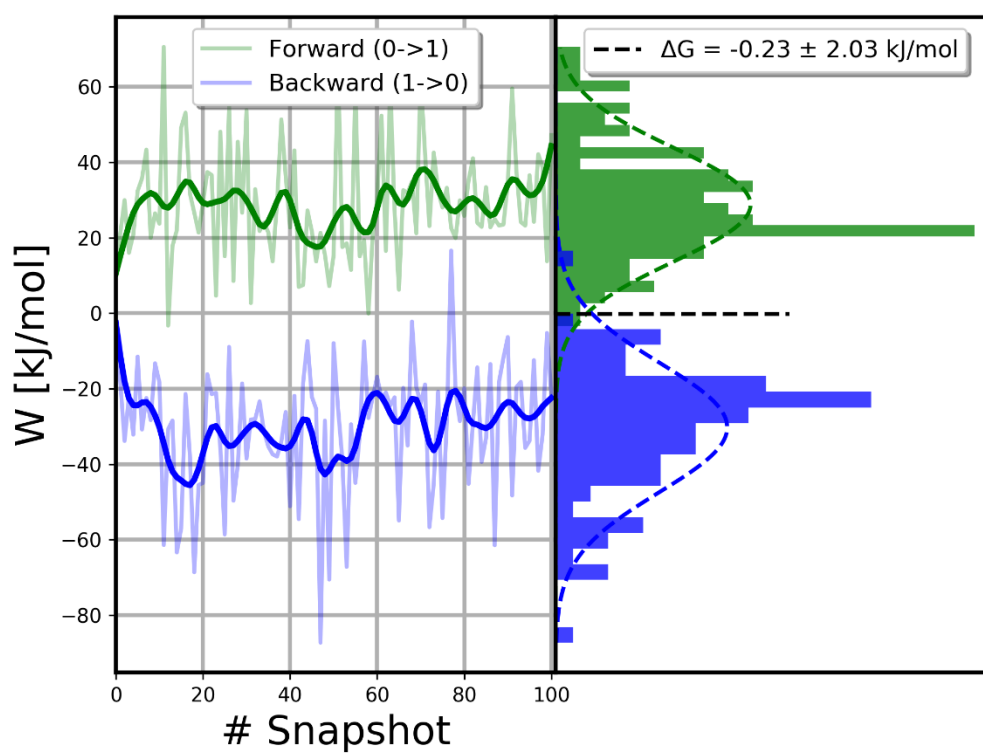
Fonte: Pr prio autor

Figura 4. Muta  o S99A do P90-110 na aus ncia do dgM2-1.



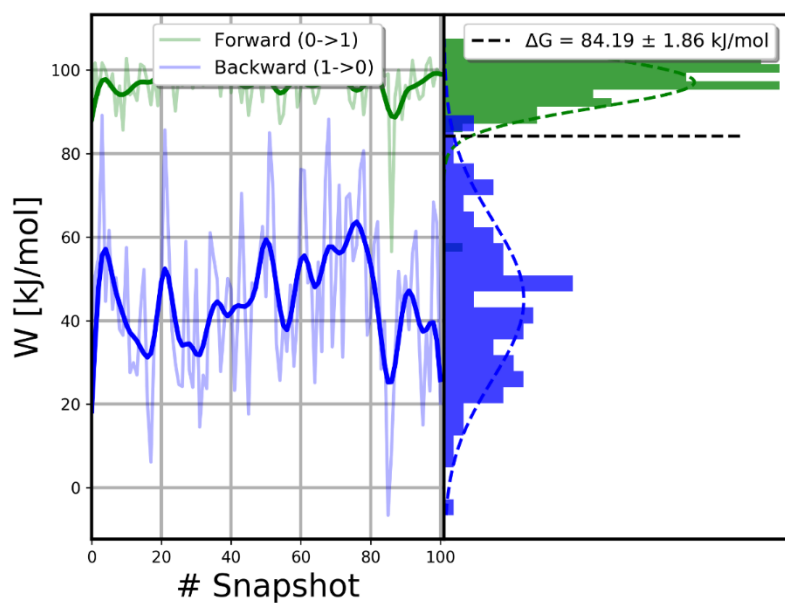
Fonte: Pr prio autor

Figura 5. Muta  o K100A do P90-110 usando a configura  o de sistema duplo/caixa  nica.



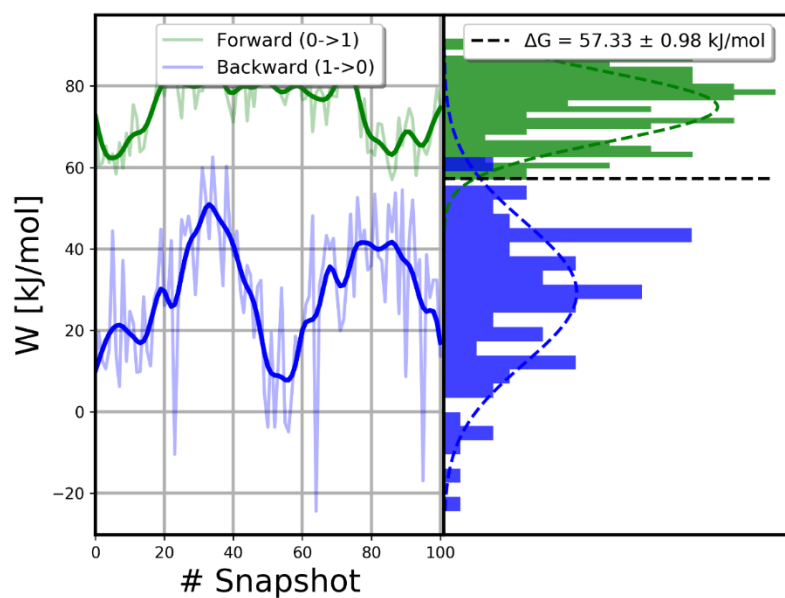
Fonte: Pr prio autor

Figura 6. Muta  o L101A do P90-110 no contexto do complexo com dgM2-1.



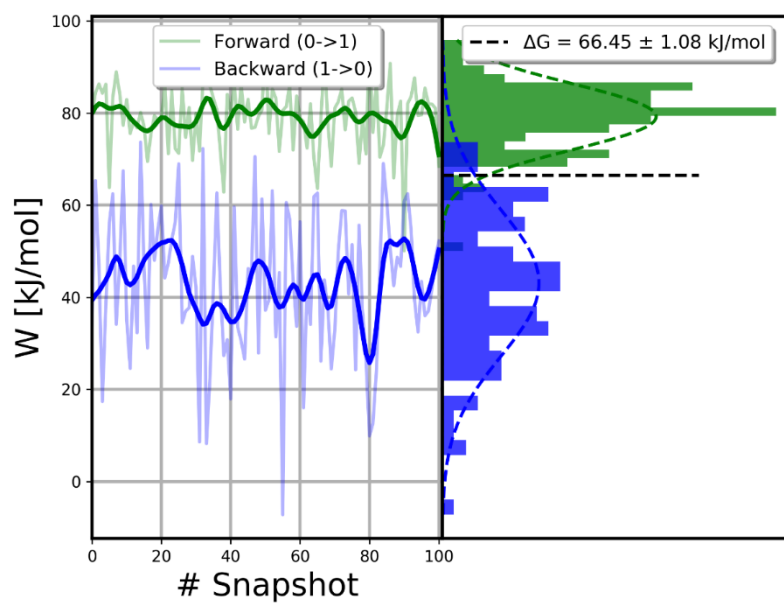
Fonte: Pr prio autor

Figura 7. Muta  o L101A do P90-110 na aus ncia do dgM2-1.



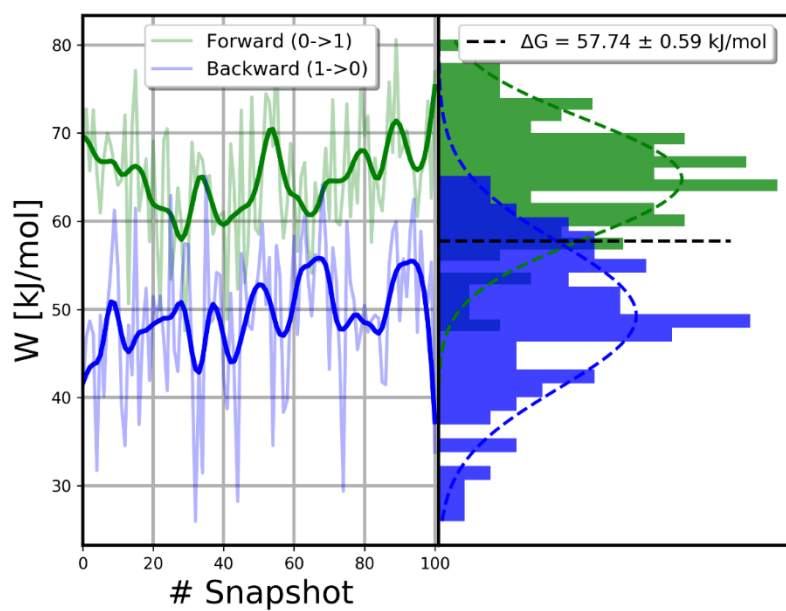
Fonte: Pr prio autor

Figura 8. Muta  o Y102A do P90-110 no contexto do complexo com dgM2-1.



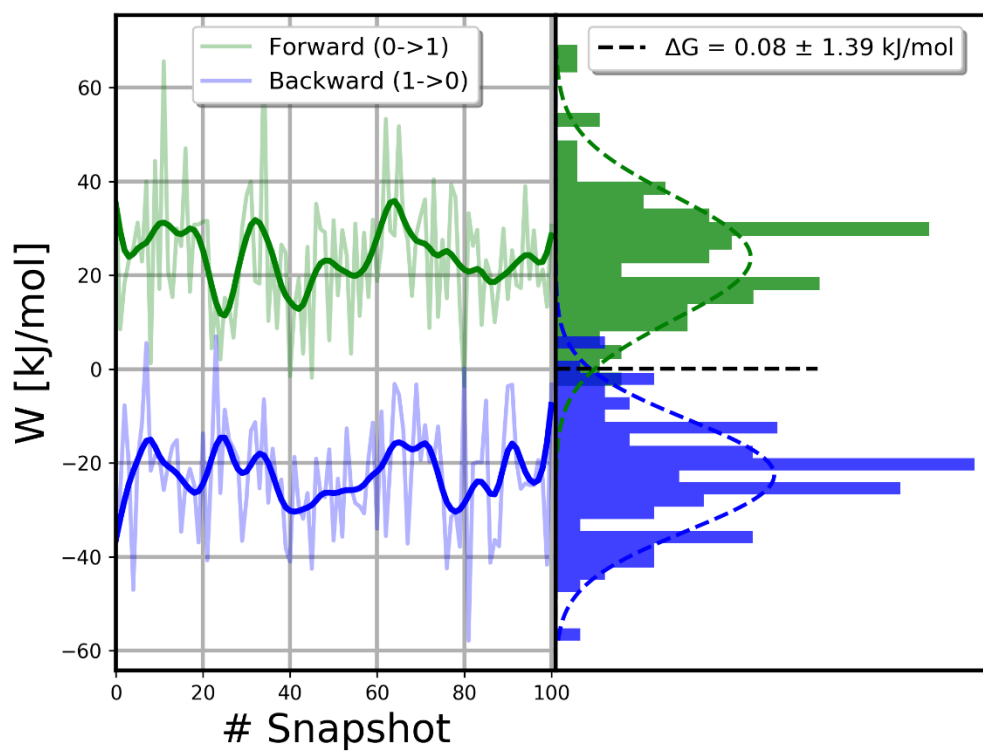
Fonte: Pr prio autor

Figura 9. Muta  o Y102A do P90-110 na aus ncia do dgM2-1.



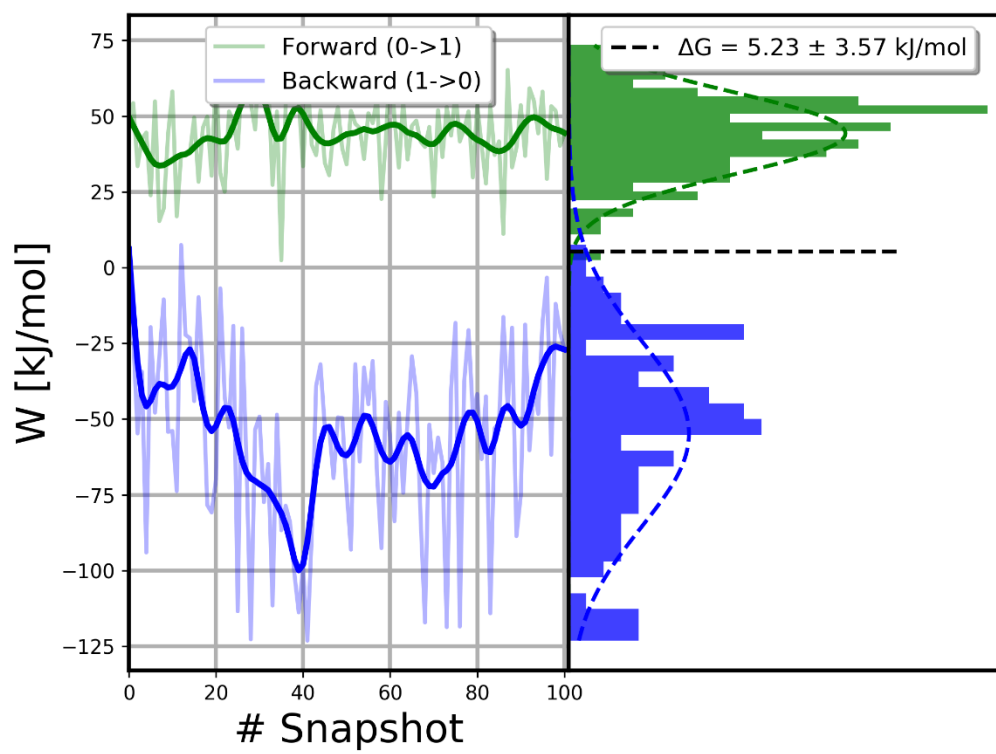
Fonte: Pr prio autor

Figura 10. Muta  o K103A do P90-110 usando a configura  o de sistema duplo/caixa  nica.



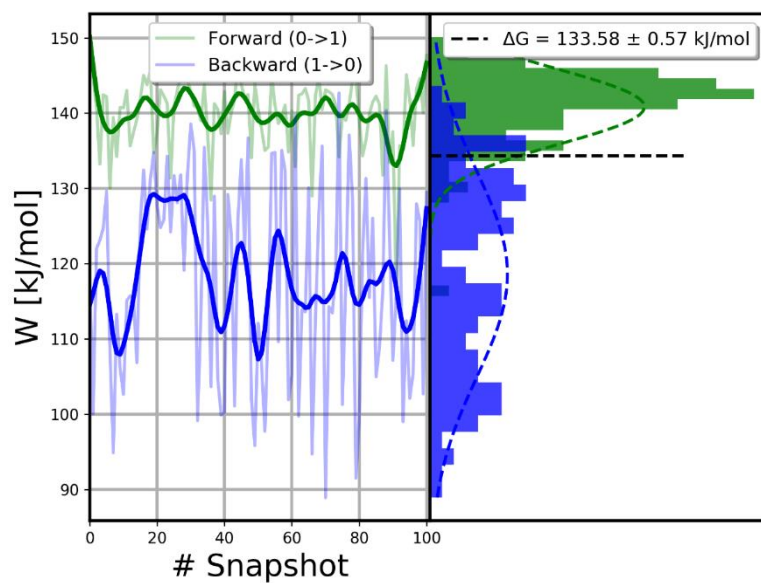
Fonte: Pr prio autor

Figura 11. Muta  o E104A do P90-110 usando a configura  o de sistema duplo/caixa  nica.



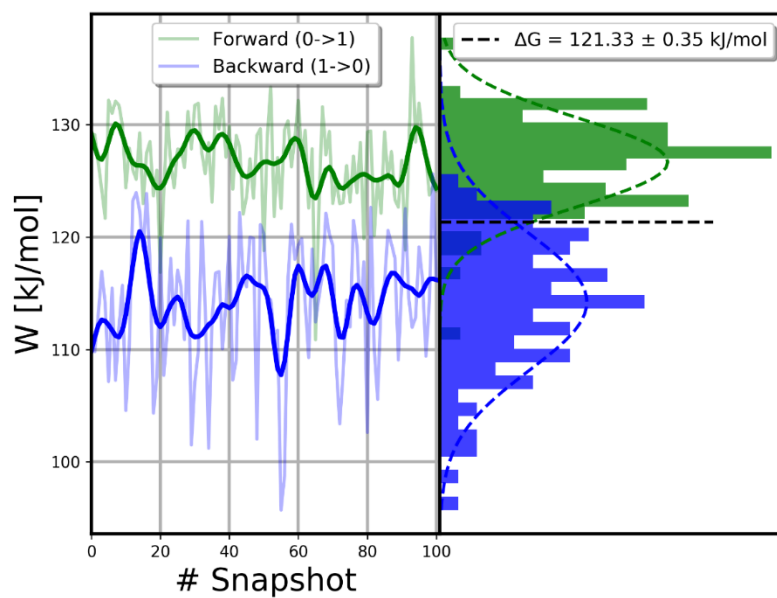
Fonte: Pr prio autor

Figura 12. Muta  o T105A do P90-110 no contexto do complexo com dgM2-1.



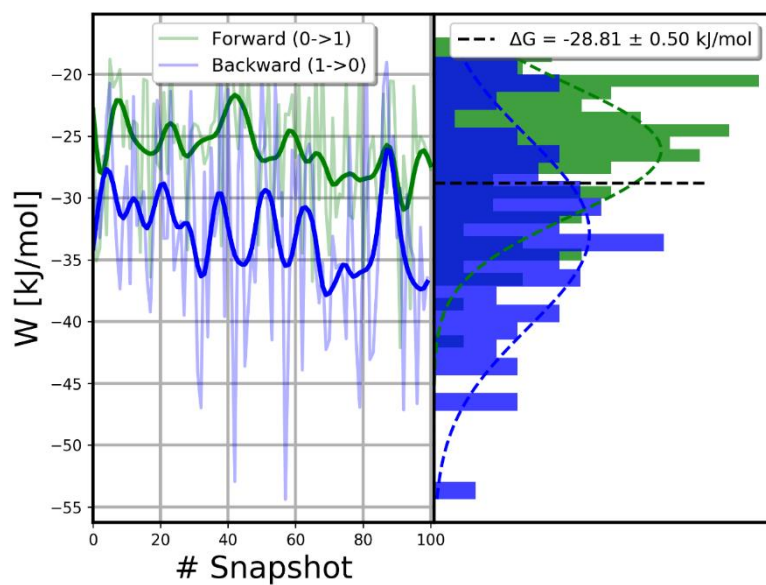
Fonte: Pr prio autor

Figura 13. Muta  o T105A do P90-110 na aus ncia do dgM2-1.



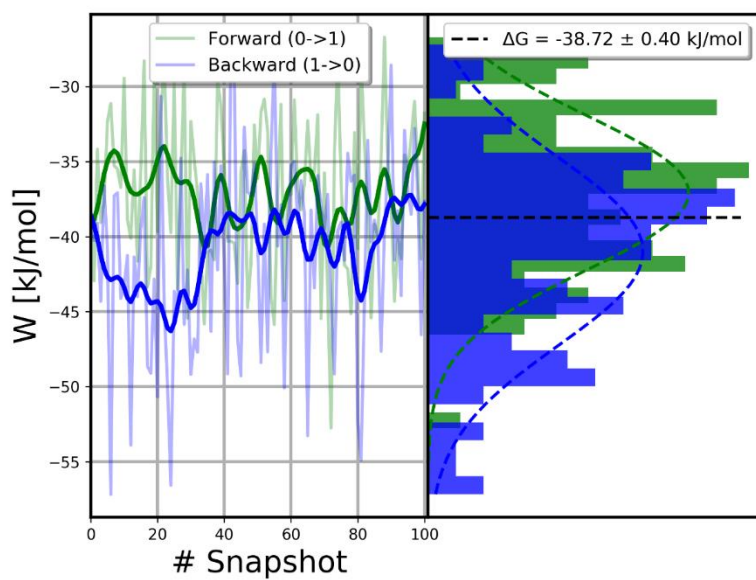
Fonte: Pr prio autor

Figura 14. Muta  o I106A do P90-110 no contexto do complexo com dgM2-1.



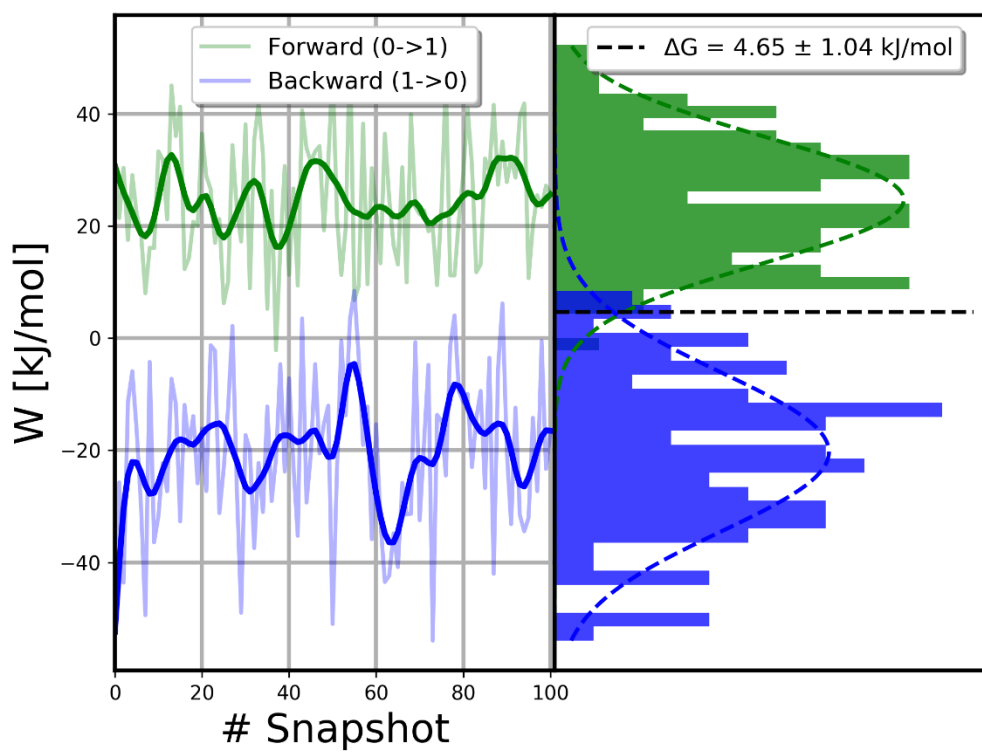
Fonte: Pr prio autor

Figura 15. Muta  o I106A do P90-110 na aus ncia do dgM2-1.



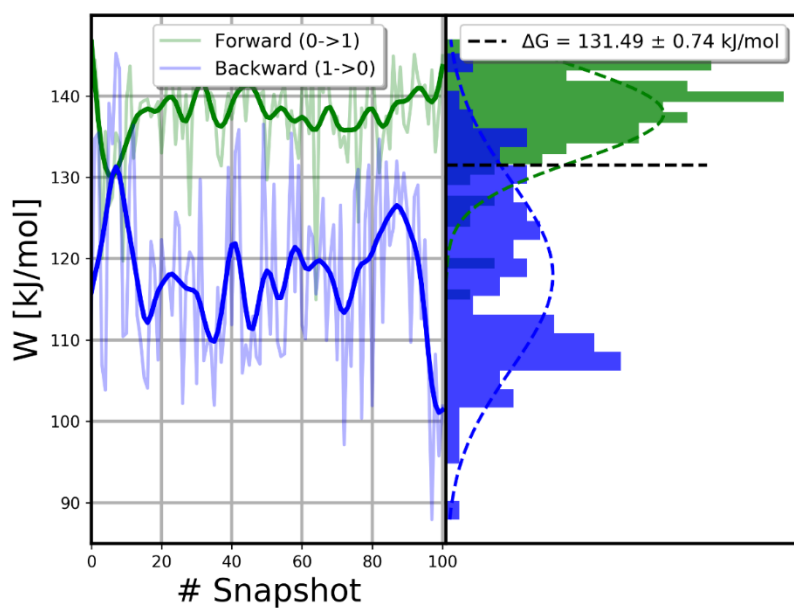
Fonte: Pr prio autor

Figura 16. Muta  o E107A do P90-110 usando a configura  o de sistema duplo/caixa  nica.



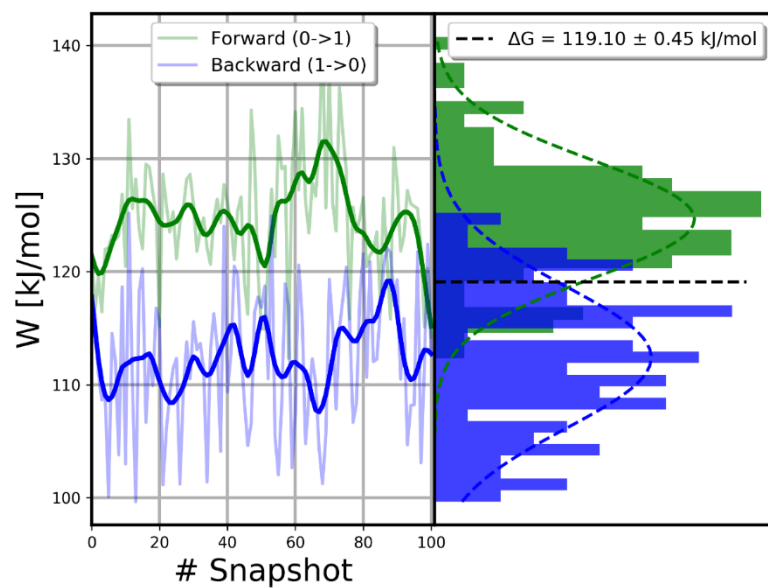
Fonte: Pr prio autor

Figura 17. Muta  o T108A do P90-110 no contexto do complexo com dgM2-1.



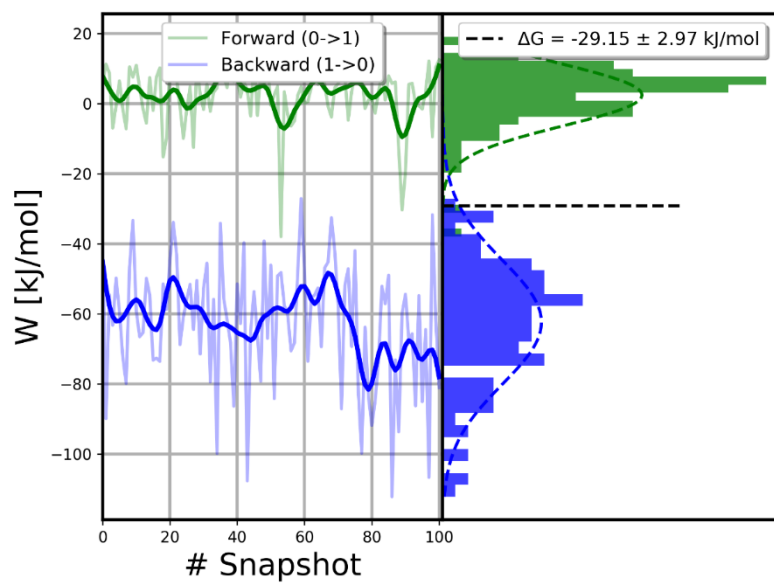
Fonte: Pr prio autor

Figura 18. Muta  o T108A do P90-110 na aus ncia do dgM2-1.



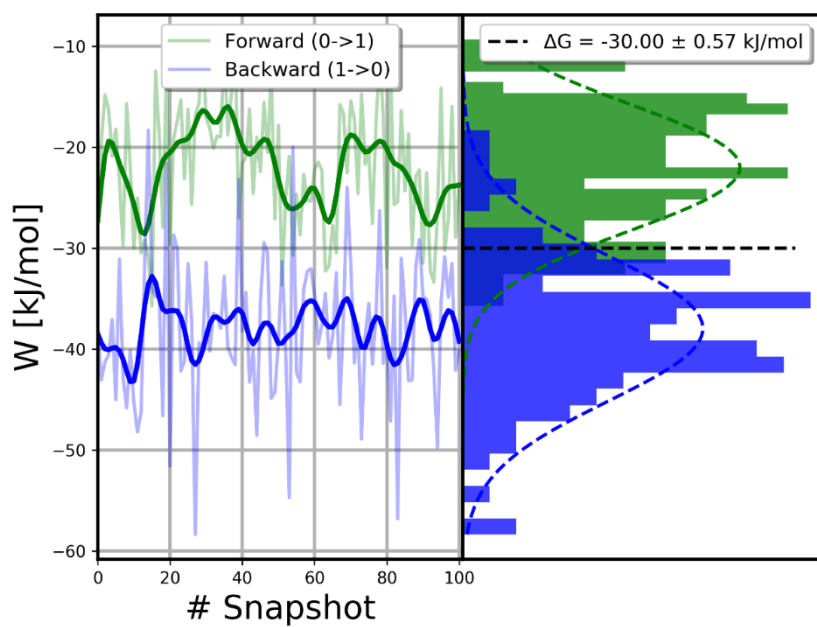
Fonte: Pr prio autor

Figura 19. Muta  o F109A do P90-110 no contexto do complexo com dgM2-1.



Fonte: Pr  prio autor

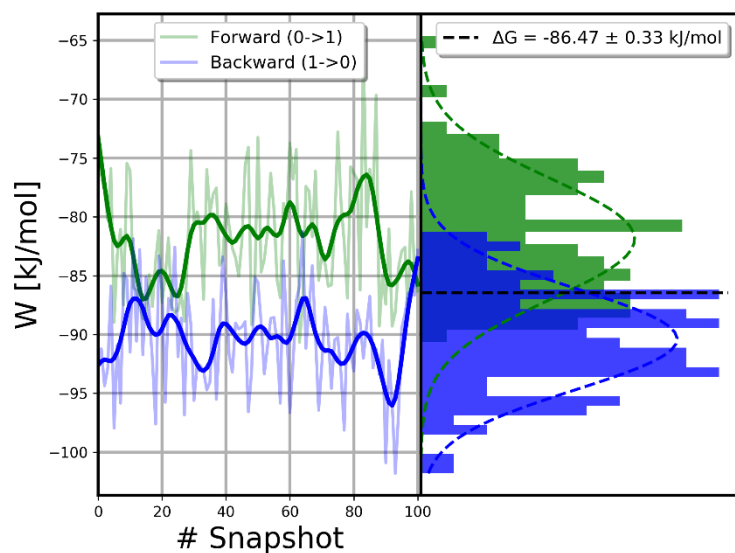
Figura 20. Muta  o F109A do P90-110 na aus  ncia do dgM2-1.



Fonte: Pr  prio autor

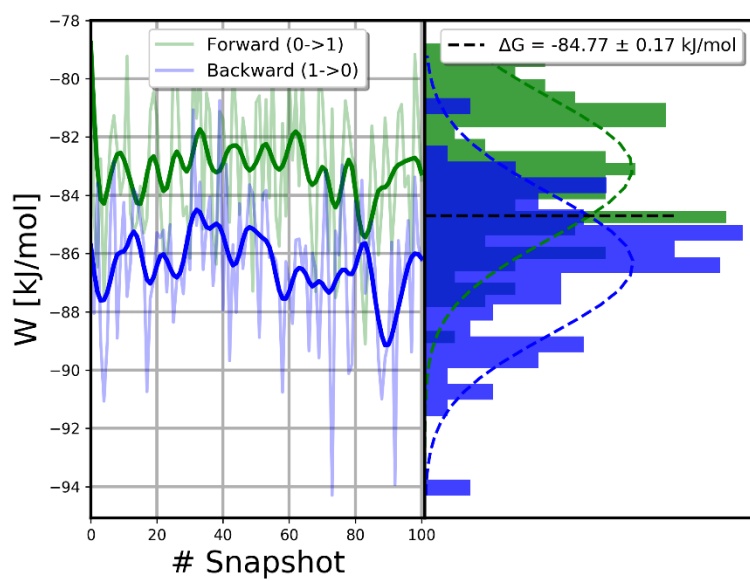
ANEXO B - Cálculo de ΔG por Integração Termodinâmica de Crescimento Rápido para as mutações direcionadas

Figura 21. Mutação F98Y do P90-110 no contexto do complexo com dgM2-1.



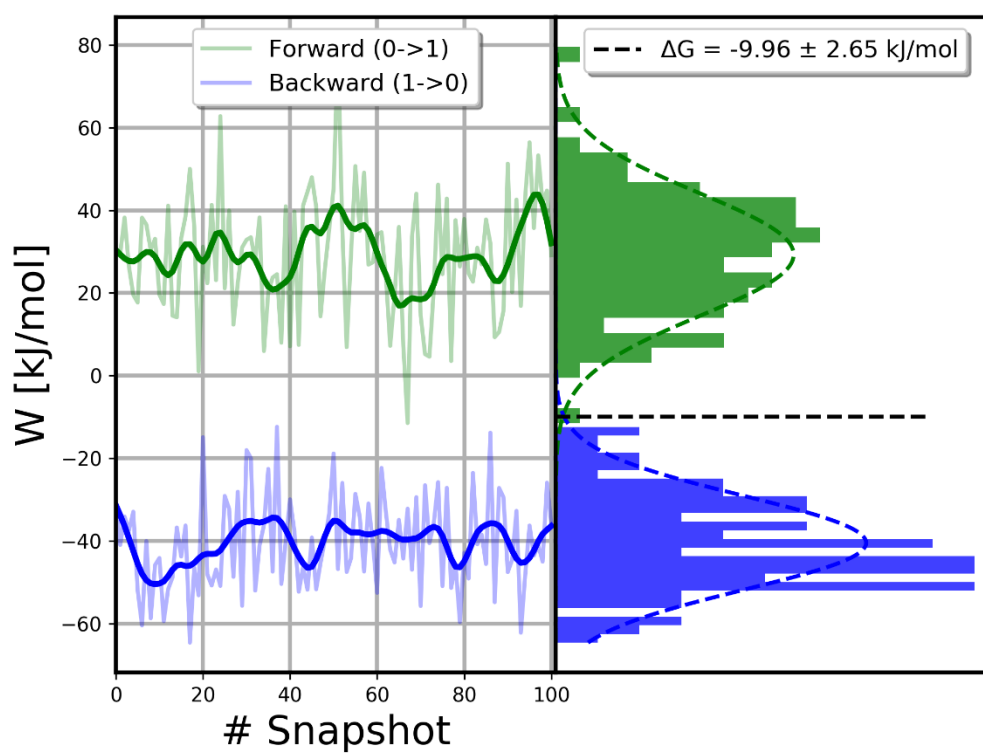
Fonte: Próprio autor

Figura 22. Mutação F98Y do P90-110 na ausência do dgM2-1.



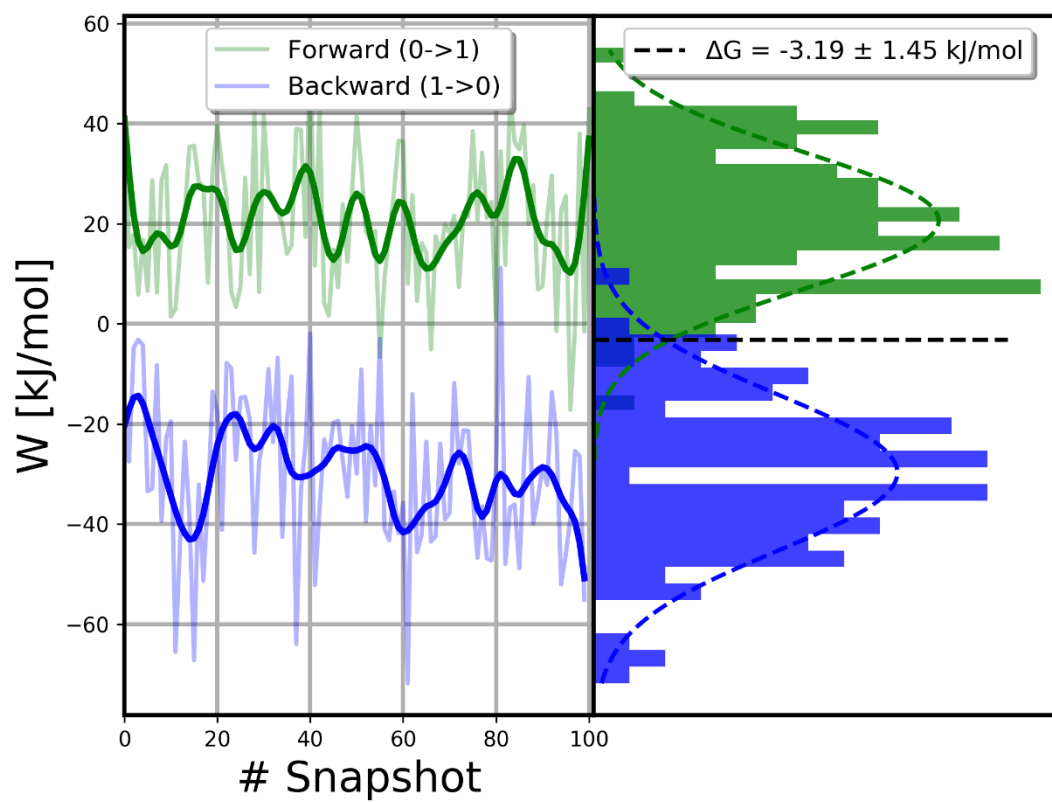
Fonte: Próprio autor

Figura 23. Muta  o S99E do P90-110 usando a configura  o de sistema duplo/caixa  nica.



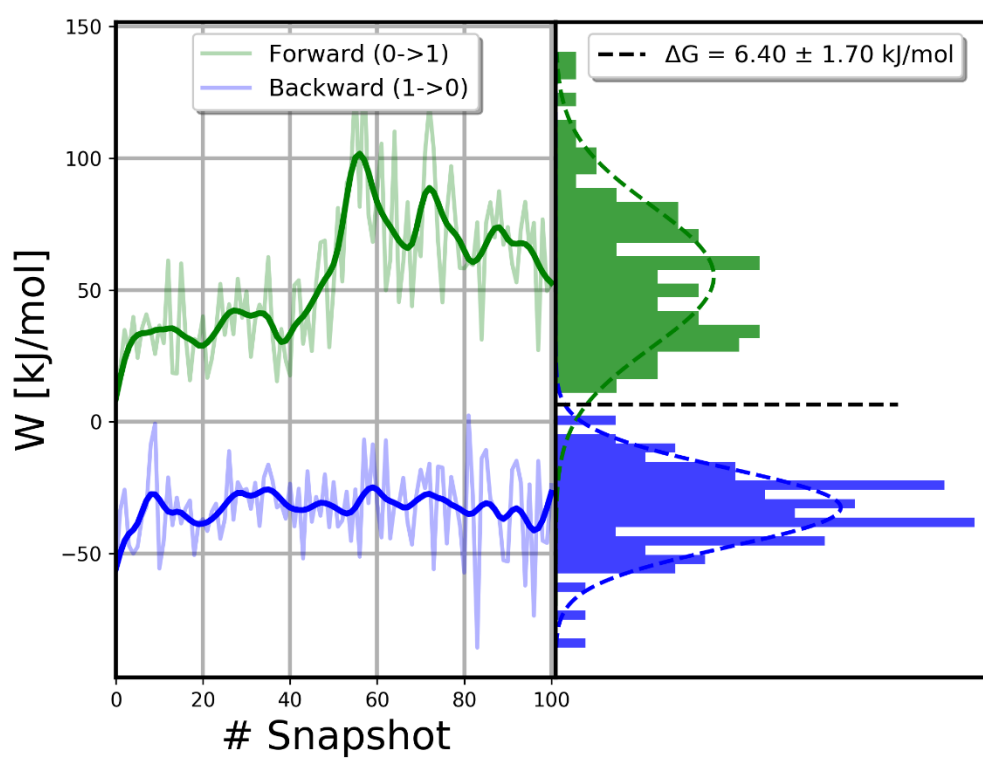
Fonte: Pr prio autor

Figura 24. Muta  o S99D do P90-110 usando a configura  o de sistema duplo/caixa  nica.



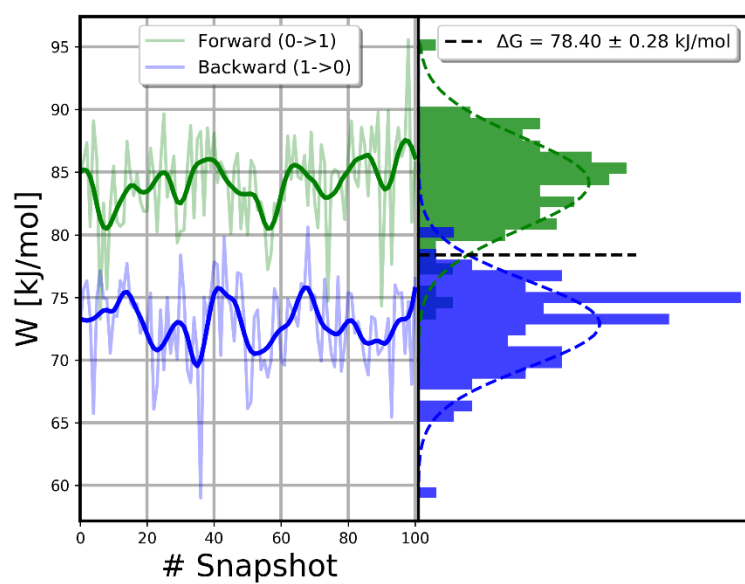
Fonte: Pr prio autor

Figura 25. Muta  o K100Q do P90-110 usando a configura  o de sistema duplo/caixa  nica.



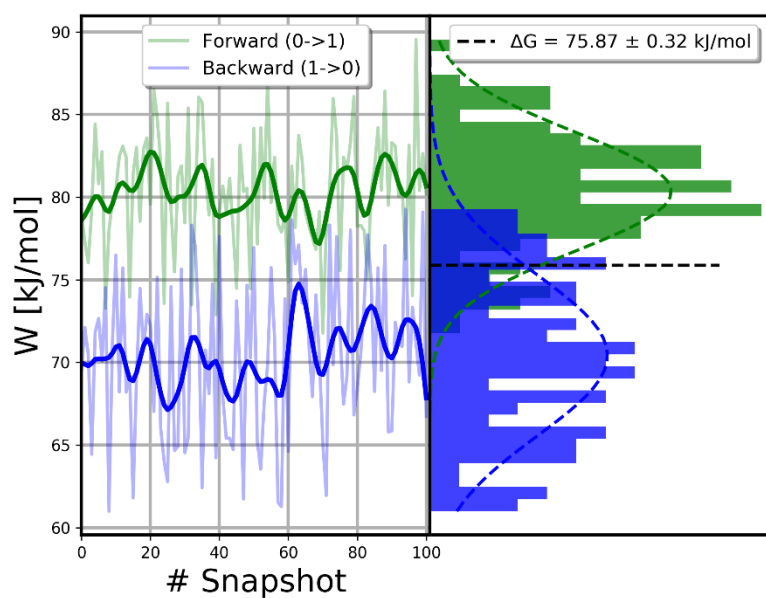
Fonte: Pr prio autor

Figura 26. Muta  o T105S do P90-110 no contexto do complexo com dgM2-1.



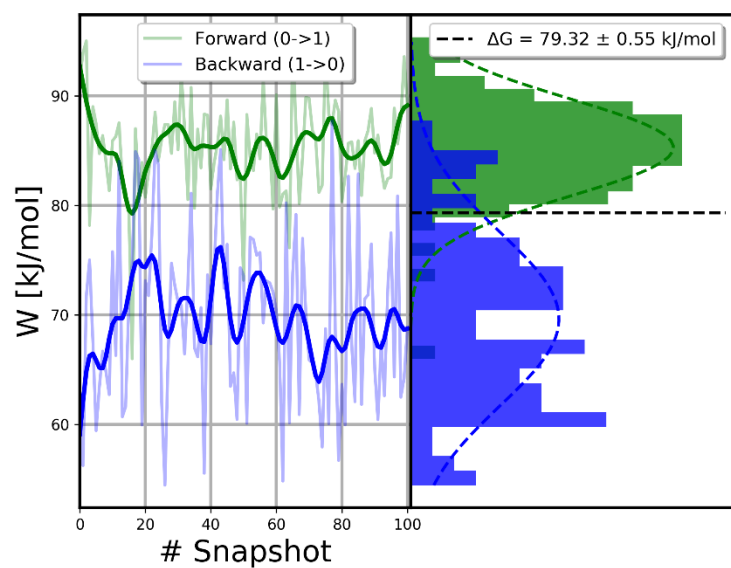
Fonte: Pr prio autor

Figura 27. Muta  o T105S do P90-110 na aus ncia do dgM2-1.



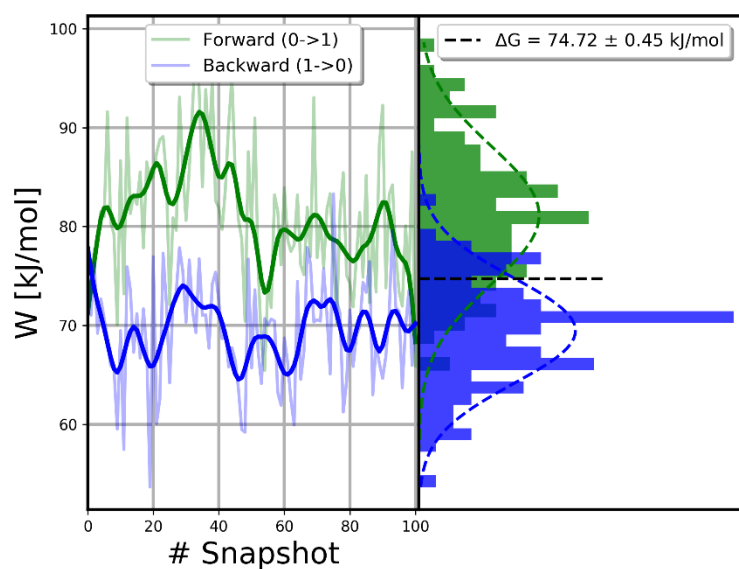
Fonte: Pr prio autor

Figura 28. Muta  o T108S do P90-110 no contexto do complexo com dgM2-1.



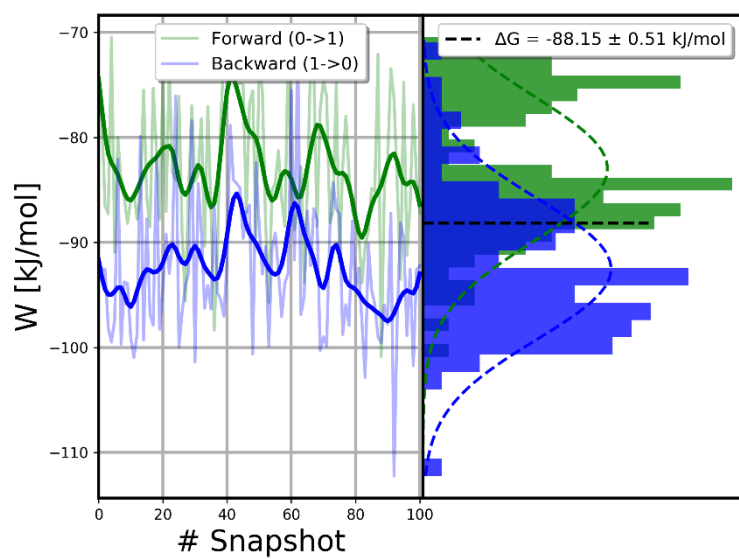
Fonte: Pr prio autor

Figura 29. Muta  o T108S do P90-110 na aus ncia do dgM2-1.



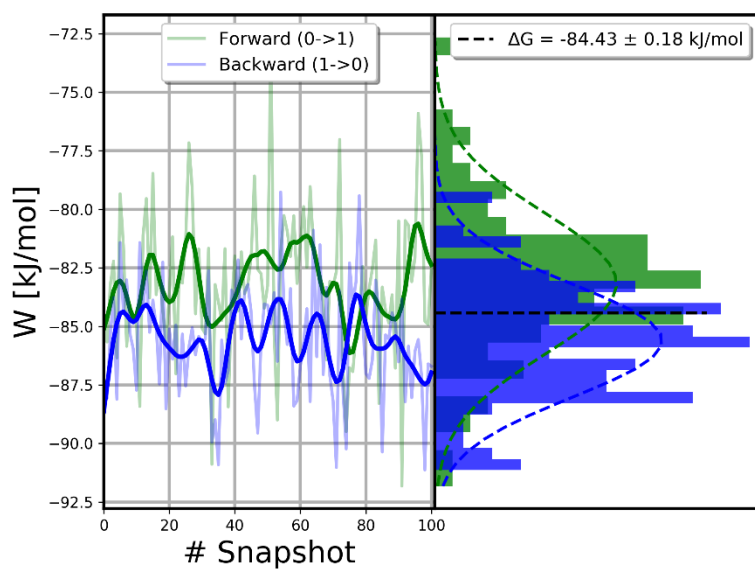
Fonte: Pr prio autor

Figura 30. Muta  o F109Y do P90-110 no contexto do complexo com dgM2-1.



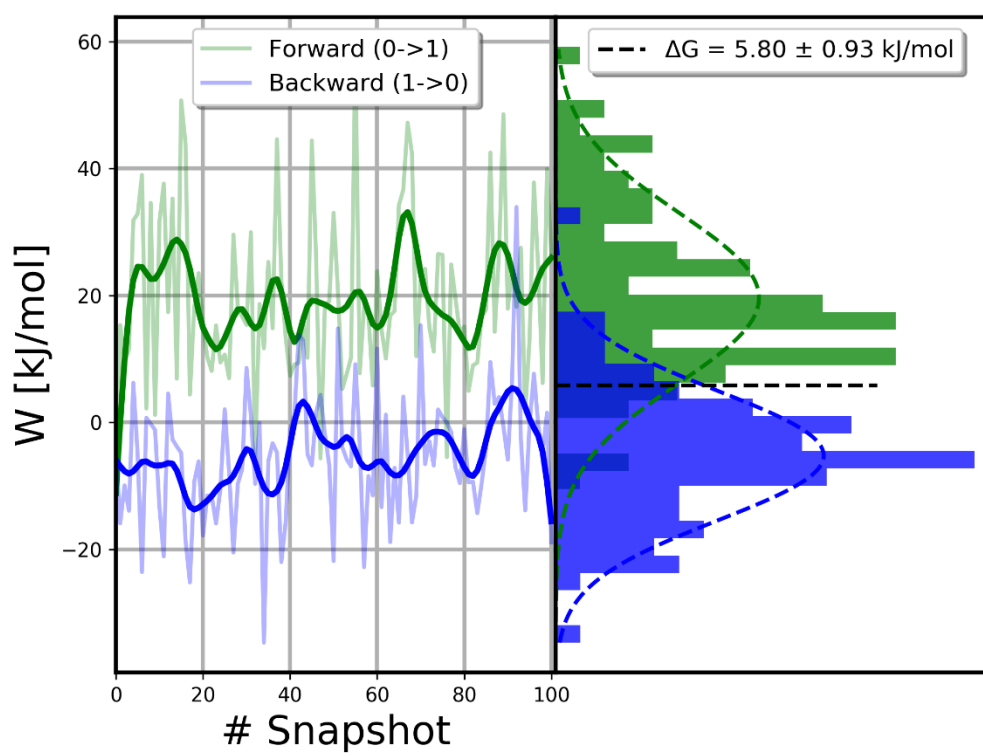
Fonte: Pr prio autor

Figura 31. Muta  o F109Y do P90-110 na aus ncia do dgM2-1.



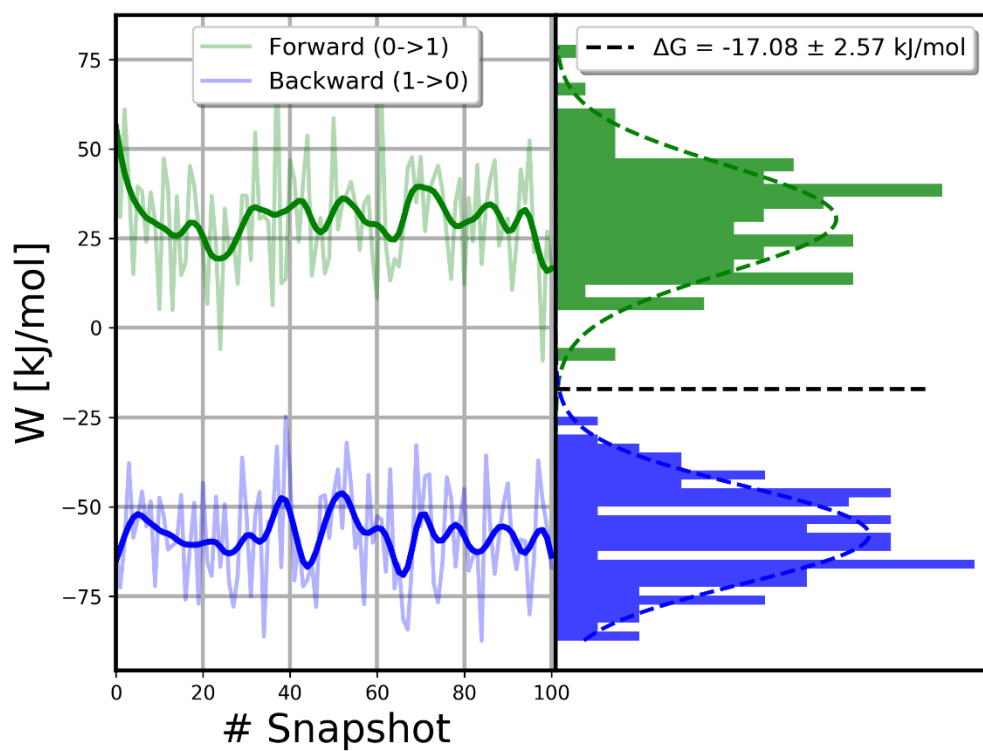
Fonte: Pr prio autor

Figura 32. Muta  o F109W do P90-110 usando a configura  o de sistema duplo/caixa  nica.



Fonte: Pr prio autor

Figura 33. Muta  o FSF98/99/109YEEY do P90-110 usando a configura  o de sistema duplo/caixa  nica.



Fonte: Pr prio autor