

Gabriela Campos de Araujo

Estudo da Estrutura da Proteína SH do Vírus Sincicial Respiratório  
Humano: análise funcional da estrutura  
pentamérica por ferramentas de bioinformática.

São José do Rio Preto  
2013

Gabriela Campos de Araraju

Estudo da Estrutura da Proteína SH do Vírus Sincicial Respiratório  
Humano: análise funcional da estrutura  
pentamérica por ferramentas de bioinformática.

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

Orientador: Prof<sup>a</sup>. Dr<sup>a</sup>. Fátima Pereira de Souza

São José do Rio Preto  
2013

Araujo, Gabriela Campos de

Estudo da estrutura da proteína SH do vírus sincicial respiratório humano: análise funcional da estrutura pentamérica por ferramentas de bioinformática. / Gabriela Campos de Araujo. - São José do Rio Preto: [s.n.], 2013.  
80 f. : 43 il. ; 30 cm.

Orientador: Fátima Pereira de Souza

Co-orientador: Ronaldo Junio de Oliveira

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas

1. Proteína SH. 2. RSV. 3. Modelagem molecular. 4. Dinâmica molecular I. Souza, Fátima Pereira de. II Ronaldo Junio de Oliveira. III. Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências, Letras e Ciências Exatas. IV. Título.

CDU – 577.112

Gabriela Campos de Araujo

Estudo da Estrutura da Proteína SH do Vírus Sincicial Respiratório  
Humano: análise funcional da estrutura  
pentamérica por ferramentas de bioinformática.

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

Comissão Examinadora

Prof<sup>a</sup>. Dr<sup>a</sup>. Fátima Pereira de Souza  
UNESP – São José do Rio Preto  
Orientador

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

Prof. Dr. Luiz Paulo Barbour Scott  
UFABC – Santo André

São José do Rio Preto  
24 de maio de 2013

*A todas as pessoas que de alguma forma  
contribuíram para meu conhecimento.*

## **AGRADECIMENTOS**

Aos meus pais Adilson e Vera, aos meus irmãos Thalles e Túlio e a minha avó Amália, os quais amo muito, pela estrutura, carinho, paciência e incentivo para que eu pudesse realizar meus sonhos.

À Profa. Dra. Fátima Pereira de Souza, pela amizade, orientação, incentivo, crescimento profissional e científico que me proporcionou.

Ao Dr. Ronaldo Júnior de Oliveira, pela coorientação, atenção, receptividade, colaboração e conhecimento.

Ao Prof. Dr. Alexandre Suman de Araujo pela colaboração, atenção e incentivo.

Ao Prof. Dr. Marcelo Andrés Fossey, pela atenção, companheirismo e incentivo.

Aos Demais Professores do Departamento de Física que me mostraram o caminho da ciência.

Aos amigos Haroldo, Deriane, Danúbia, Thiago, Luciana, Karoline, Stellita, Gisele pela atenção, incentivo durante a realização desse trabalho, pela boa convivência e pelas horas de descontração e aprendizagem.

A TODOS os colegas e amigos do Departamento de Física pela colaboração.

Aos funcionários do Departamento de Física pela atenção e eficiência no atendimento.

A todos que direta ou indiretamente contribuíram para o andamento deste trabalho.

## LISTA DE FIGURAS

<b>Figura 1:</b> Esquema da organização genômica do RNA do RSV .....	16
<b>Figura 2:</b> Estrutura do Vírus Sincicial Respiratório .....	17
<b>Figura 3:</b> Ciclo de replicação do hRSV .....	18
<b>Figura 4:</b> Representação esquemática do genoma da proteína F do vírus hRSV .....	20
<b>Figura 5:</b> Representação esquemática da glicoproteína G do hRSV .....	21
<b>Figura 6:</b> Representação esquemática da estrutura secundária da SH do hRSV .....	22
<b>Figura 7:</b> Estrutura geral do pentâmero formado pelo proteína SH segundo Gan e seus colaboradores .....	24
<b>Figura 8:</b> Sequencia de aminoácidos representado por uma roda helicoidal .....	25
<b>Figura 9:</b> Estrutura do lipídio POPC (1-palmitoil-2-oleoilfosfatidilcolina).....	43
<b>Figura 10:</b> Passos da construção da bicamada hidratada .....	44
<b>Figura 11:</b> Estratégia utilizada para a inserção da proteína na bicamada lipídica .....	45
<b>Figura 12:</b> <i>Helical Wheel</i> dos resíduos da proteína SH RSV .....	47
<b>Figura 13:</b> Esquema da montagem do poro formado pela proteína SH com base na análise <i>Helical Wheel</i> .....	48
<b>Figura 14:</b> Estrutura terciária da proteína SH gerada pelo VMD utilizando a ferramenta <i>Molecule Factor</i> .....	49
<b>Figura 15:</b> Esquema do poro construído pelo VMD .....	49
<b>Figura 16:</b> Predição da estrutura secundária da proteína SH realizada pelo servidor PSIPred .....	51
<b>Figura 17:</b> Modelos da proteína SH gerados pelo servidor I-TASSER .....	53
<b>Figura 18:</b> Raio de giro dos modelos 1, 4 e 5 em água em função do tempo de simulação .....	54
<b>Figura 19:</b> Flutuação média da posição de cada resíduo dos modelos 1, 4 e 5 em água .....	55
<b>Figura 20:</b> Modelos da proteína SH após simulação em água a 310K .....	55
<b>Figura 21:</b> Flutuação média da posição de cada resíduo dos modelos 1, 4 e 5 com restrição .....	57
<b>Figura 22:</b> Modelos da proteína SH após simulação em água com restrições .....	57
<b>Figura 23:</b> Estrutura da proteína SH escolhida pelo modelo baseado em estrutura .....	59
<b>Figura 24:</b> Perfil de Densidade de Massa da bicamada de POPC hidratada .....	60
<b>Figura 25:</b> Parâmetro de ordem das cadeias Palmitoil e Oleil .....	61
<b>Figura 26:</b> Evolução temporal da área por fosfolipídio para o sistema bicamada POPC .....	62
<b>Figura 27:</b> Estrutura mais provável da proteína SH após 74 ns de simulação .....	64
<b>Figura 28:</b> Raio de giro do sistema proteína-bicamada em função do tempo à 310K .....	65

<b>Figura 29:</b> Parâmetro de ordem das cadeias Palmitoil e Oleil da bicamada POPC e do sistema proteína-bicamada .....	66
<b>Figura 30:</b> Área por lipídeo do sistema proteína SH – bicamada para os últimos 120 ns de simulação .....	67
<b>Figura 31:</b> Estrutura mais provável do pentâmero construído com o modelo da proteína SH gerado pelo I-TASSER .....	69
<b>Figura 32:</b> Parâmetro de ordem das cadeias Palmitoil e Oleiol da bicamada POPC e dos sistemas 1 e 2 .....	71
<b>Figura 33:</b> Área por lipídeo dos sistemas bicamada-poro e bicamada-poro (VMD) .....	72
<b>Figura 34:</b> Imagem obtida das simulações dos sistemas contendo a estrutura pentamérica da proteína SH .....	73
<b>Figura 35:</b> Perfil da densidade de água e energia livre da barreira de água ao longo do eixo normal da membrana para os dois sistemas simulados .....	75



## LISTA DE TABELAS

<b>Tabela 1:</b> Dados da simulação do sistema bicamada POPC .....	63
<b>Tabela 2:</b> Dados da simulação do sistema proteína SH - bicamada .....	66
<b>Tabela 3:</b> Dados da análise das pontes de hidrogênio entre os monômeros dos poros da proteína SH .....	70
<b>Tabela 4:</b> Dados da simulação do sistemas poror-bicamada 1 e 2 .....	76

## RESUMO

O Vírus Sincicial Respiratório Humano (hRSV) é o maior causador de infecções respiratórias agudas (IRAs) em recém-nascidos e crianças no mundo inteiro. Seu genoma codifica 11 proteínas entre as quais as proteínas de superfície F, G e SH que são responsáveis pela entrada e instalação do vírus na célula do hospedeiro. Entre as proteínas de superfície, pouco se sabe sobre a função da proteína SH. O objetivo do presente estudo foi realizar a modelagem e caracterização da proteína SH e analisar o seu comportamento estrutural em diferentes meios: água e bicamada fosfolipídica. O modelo da proteína SH foi gerado pelo servidor *I-Tasser*, e suas características funcionais e estruturais analisados pelo *PredictProtein* e *PsiPred*. Simulações de Dinâmica Molecular foram realizadas para análise da hidrofobicidade da região central da proteína, do seu comportamento em membrana lipídica e possível formação de oligômeros. Os resultados da predição do modelo da proteína SH resultou em uma estrutura linear com uma alfa-hélice compreendida entre os aminoácido 20-42 e as análises realizadas pelo PsiPred indicaram essa região como sendo uma região transmembrânica. As simulações de Dinâmica Molecular mostraram que, quando em solução, a proteína muda sua conformação linear para globular confirmando a hidrofobicidade do domínio central. A presença da proteína SH sozinha ou das cinco estruturas na bicamada resultou num decréscimo considerável da área por lipídeo, conferindo as cadeias menor mobilidade e um maior alinhamento. As simulações do pentâmero mostraram a passagem de moléculas de água pelo poro em ambiente onde os resíduos de histidinas H22 e H51 apresentaram-se na forma protonada, indicando a dependência desta atividade com o pH do meio. Com base nessas análises foi possível propor uma estrutura terciária e quaternária da proteína SH e, com as análises da formação pentamérica da proteína, analisar sua influência no

ambiente composto pela bicamada apresentando uma melhor compreensão da sua função durante a infectividade viral.

Palavras-chave: Proteína SH. RSV. Modelagem molecular. Dinâmica molecular

## **ABSTRACT**

*The human Respiratory Syncytial Virus (hRSV) is the major cause of lower respiratory tract illnesses in children and elderly people worldwide. Its genome encodes 11 proteins including the surface protein F, G and SH which are responsible for entry and distribution of virus in the host cell. Among the protein surface, little is known about the function of protein SH. Knowing their structure and function is fundamental to a better understanding of its mechanism. The aim of this study was modeling and characterization of the RSV SH protein and analysis of structural behavior in different environment : water and phospholipid bilayer for understanding and evaluating the formation of its pentameric structure. The SH protein model was generated by I-TASSER server, and its functional and structural characteristics was analyzed by PredictProtein and PsiPred. Molecular Dynamics Simulation were performed for analysis of hydrophobicity of protein central region, studies of the protein behavior on the membrane and pentamer formation. The SH protein model prediction resulted in a linear model with a helix-alpha between amino acid 20-42 and the analysis performed by PsiPred indicated this region as transmembrane region. Molecular Dynamics Simulation showed that, when in solution, the protein changes its linear conformation for globular conformation confirming the hydrophobicity of the central domain. The presence of the Sh protein itself or of the pentamer in bilayer resulted in a considerable decrease of the area per lipid,*

*giving the chains less mobility and greater alignment. The pentamer simulation showed passage of water molecules through the pore in an environment where histidine residues H22 and H51 are protonated, indicating the dependence of this activity with the pH of the medium. Based on this analysis, it was proposed the structure tertiary and quaternary of the SH protein and, with the analysis of the protein pentameric form, analyze its influence on the environment consisting of the bilayer for understanding its function in viral infectivity.*

*Keywords: SH Protein. RSV. Molecular Modeling. Molecular Dynamics.*

## Sumário

<b>1. INTRODUÇÃO</b>	14
Classificação, Estrutura e Genoma	16
O Ciclo de Replicação	17
As Proteínas Virais	18
<b>2. PROTEÍNA SH</b>	21
Topologia, Estrutura e Localização	21
Papel da Proteína SH	22
<b>3. OBJETIVO</b>	25
<b>4. MATERIAIS E MÉTODOS</b>	25
<b>Ferramentas de Bioinformática Utilizadas</b>	26
NCBI	26
BioEdit	26
PredictProtein	26
PSIPred	27
I-TASSER	27
<b>Simulação por Dinâmica Molecular</b>	28
Campo de Força	29
Minimização de Energia	31
Steepest Descent (declínio acentuado)	32
Conjugate Gradient (gradiente conjugado)	33
Métodos de Integração	33
<b>Modelo Baseado em Estrutura</b>	34
<b>5. ANÁLISES APLICADAS</b>	36
RMSD e RMSF	36
Raio de Giro	37
Perfil Densidade de Massa	38
Área por lipídeo	38
Parâmetro de Ordem	38
Helical Wheel (roda helicoidal)	39
<b>6. DETALHES COMPUTACIONAIS</b>	40
Seleção e Análise das sequências	40
Modelagem da proteína SH	40
Simulação da proteína SH em água	40
Construção do sistema Proteína SH - Membrana	42
Construção da Bicamada Fosfolipídica Hidratada	43
Proteína SH em Membrana POPC	45
Estrutura do Pentâmero	46

<b>7. RESULTADOS E DISCUSSÕES</b> .....	51
Seleção e Análise da estrutura primária da proteína SH.....	51
Modelagem da proteína SH .....	52
Simulação da proteína SH na água.....	53
Análise da Bicamada composta por POPC .....	59
Análises da simulação da Proteína SH na bicamada POPC .....	63
Análises das simulações do pentâmero inserido na bicamada de POPC.....	68
<b>8. CONCLUSÕES</b> .....	77
<b>9. REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	79
<b>10. ANEXO 1</b> .....	83

## 1. INTRODUÇÃO

As infecções respiratórias agudas (IRAs) são as maiores causadoras de morbidade e mortalidade entre pacientes pediátricos no mundo todo<sup>1</sup>, sendo responsáveis por aproximadamente 30% de todos os óbitos em crianças nos países em desenvolvimento<sup>2</sup>. Embora as IRAs causem um menor número de mortes em países desenvolvidos, elas produzem grandes custos diretos e indiretos com assistência à saúde.

Cerca de 50 a 90% das infecções do trato respiratório inferior em crianças menores que um ano de idade<sup>1</sup> são atribuídas aos vírus, principalmente: Vírus da Parainfluenza Humana (VPIH) I, II, e III, Influenza A e B, Adenovírus, mais recentemente o Metapneumovírus Humano (MPVH) e o vírus em estudo: o Vírus Sincicial Respiratório Humano (hRSV)<sup>3,4</sup>.

As doenças respiratórias provocadas pelo hRSV vão desde um resfriado comum em adultos, até quadros de bronquiolite em recém-nascidos e pneumonia e bronquite em crianças maiores. O vírus é responsável por 40% das bronquiolites e 25% de todas as pneumonias virais. O período de incubação do vírus é por volta de 4 a 5 dias, porém, a infecção viral poder ter um período de transmissão de até três semanas<sup>5</sup>.

O hRSV tem distribuição mundial e apresenta surtos principalmente nos países em que as estações são bem definidas. Em países de clima subtropical, o surto de hRSV ocorre nos meses mais chuvosos e frios<sup>5</sup>. No Brasil, a sazonalidade considerada pela Sociedade Brasileira de Pediatria (SBP) ocorre nos meses de maio a setembro.

Quase 100% das crianças em idade pré-escolar já foram infectadas e muitas delas sofrem mais de uma infecção pelo hRSV ao longo da vida. Desta forma, pode se verificar que a infecção pelo hRSV induz a uma memória imunológica fraca, pois sua recidiva é frequente. Apesar de possuírem anticorpos circulantes contra o vírus, pessoas idosas também se infectam principalmente em fases imunodeprimidas<sup>5</sup>.



Os isolados do RSV podem ser classificados em dois grupos, A e B. Esta classificação é baseada nas mutações das sequências de aminoácidos das proteínas virais, especialmente na proteína G. Os grupos A e B do vírus geralmente circulam simultaneamente, embora o vírus do grupo A seja mais predominante do que o vírus do grupo B. Estudos realizados indicam que as crianças infectadas pelo vírus do grupo A apresentam maior grau de severidade da doença em relação às infectadas pelo grupo B<sup>6</sup>. Acredita-se que o genoma viral e as diferenças funcionais das proteínas relacionadas à replicação do hRSV do tipo A sejam os principais fatores de gravidade para as infecções causadas por esse tipo viral<sup>7</sup>.

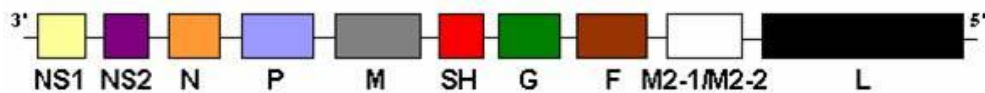
As principais dificuldades encontradas no desenvolvimento de um tratamento específico eficiente contra o hRSV é principalmente devido a variabilidade antigênica viral e as características do sistema imune do hospedeiro<sup>8</sup>.

O RSV é transmitido principalmente pelo contato, apresenta capacidade de sobrevivência, sendo viável por um período superior a 6 horas em uma grande variedade de locais. Este vírus apresenta alta virulência e patogenicidade quando localizado em mucosas nasais e orais, qualquer que seja a fonte de infecção. Segundo a Sociedade Brasileira de Pediatria, ele sobrevive bem em roupas e aventais (45 min), conta-gotas, brinquedos, objetos sobre a mesa, estetoscópios e grades de berço (6 horas), luvas cirúrgicas, pijamas, lenços de papel e pele.

O vírus penetra o hospedeiro através das membranas dos olhos, nariz e boca, atingindo a mucosa respiratória. Seu período de incubação varia de 2 a 8 dias. Após a incubação viral, a sintomatologia apresentada se caracteriza por secreção nasal de cor clara, tosse moderada e hipertermia baixa, podendo ocorrer sibilância em alguns casos (Sociedade Brasileira de Pediatria).

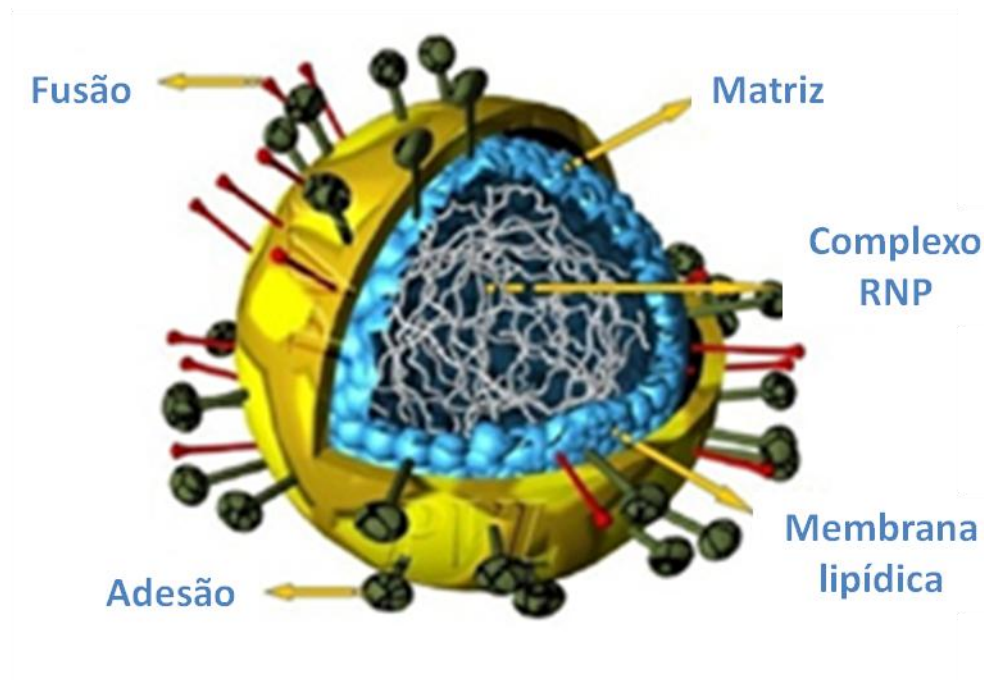
### Classificação, Estrutura e Genoma

O hRSV, pertencente ao gênero *Pneumovirus* da família *Paramyxoviridae*, é um vírus envelopado de tamanho médio de 120 a 300 nm e apresenta um genoma de RNA fita simples não segmentado de polaridade negativa. Este RNA é composto por aproximadamente 15.222 nucleotídeos que formam 10 genes os quais codificam 11 proteínas: NS1 e NS2, que correspondem às proteínas não estruturais; gene da nucleoproteína (N); da fosfoproteína (P); da proteína de matriz (M); a pequena proteína hidrofóbica (SH); o gene da glicoproteína de ligação (G); gene da proteína de fusão (F) e o gene da proteína polimerase (L)<sup>9</sup> (figura 1).



**Figura 1: Esquema da organização genômica do RNA do RSV.** Os genes NS1 e NS2 codificam proteínas não estruturais (NS1 e NS2); N, P, L codificam proteínas associadas ao nucleocapsídeo viral; M, M2-1 e M2-2 codificam proteínas de matriz; G, F e SH codificam proteínas transmembrânicas de superfície responsáveis pela ligação, fusão e adsorção viral à célula hospedeira, respectivamente. Fonte modificada a partir de [www.medscape.com/viewarticle/580494](http://www.medscape.com/viewarticle/580494).

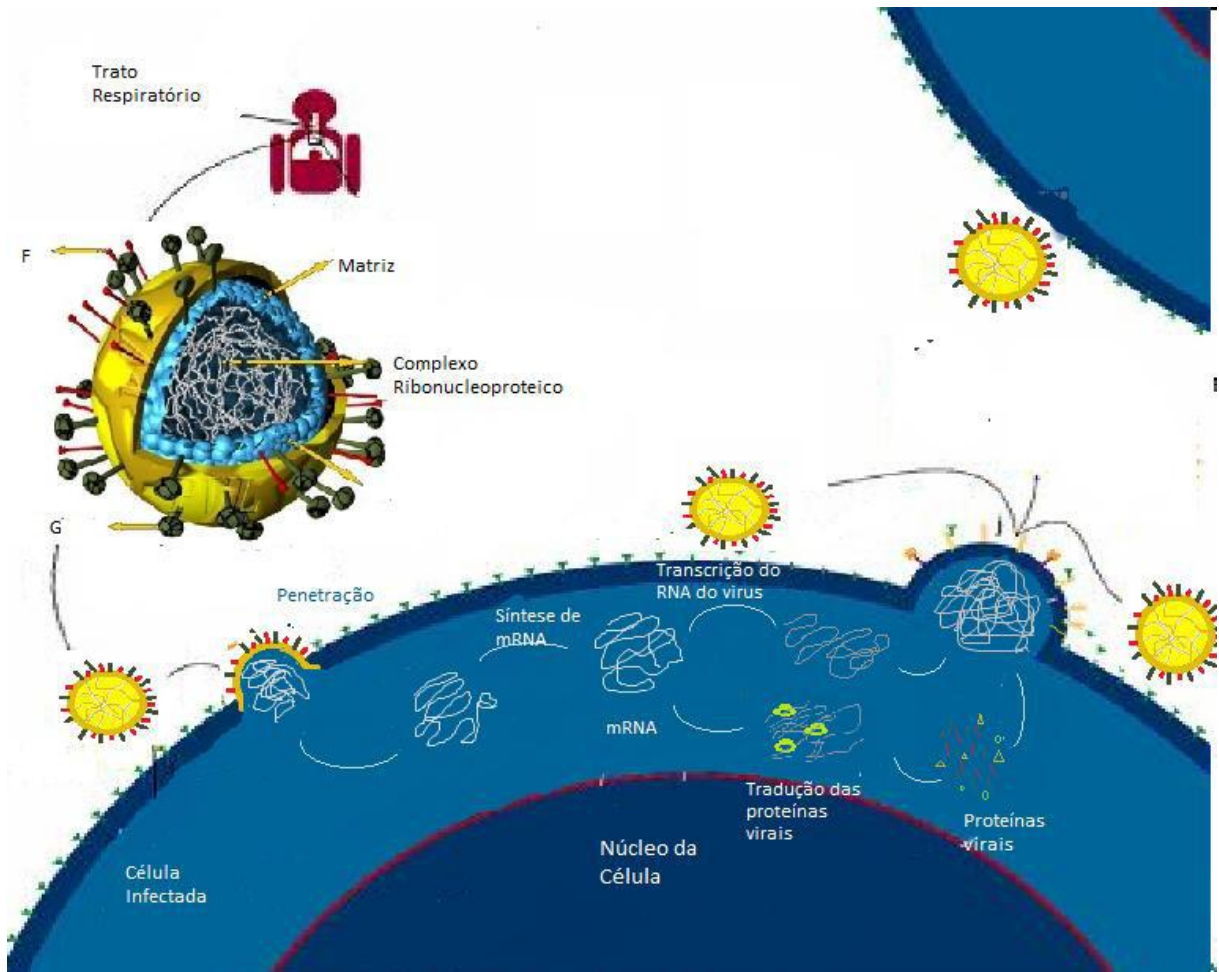
O envelope é uma bicamada lipídica derivada da membrana plasmática das células do hospedeiro e que contem em sua superfície a glicoproteína de adsorção (G), a proteína de fusão (F) e a pequena proteína hidrofóbica (SH). Essas glicoproteínas virais são montadas em "picos" espaçados na superfície e exerce funções de adesão do vírus a célula infectada<sup>9</sup> (figura 2).



**Figura 2: Estrutura do Vírus Sincicial Respiratório.** A principal proteína do nucleocapsídeo N, a fosfoproteína P, a grande subunidade da polimerase L e uma fita simples de RNA com polaridade negativa, compõem o complexo RNP. Duas das três glicoproteínas transmembrana de superfície estão presentes: a proteína de adesão G, a proteína de fusão F. A pequena proteína hidrofóbica SH não está representada na figura. A proteína de matriz M forma uma camada protetora sobre a camada interna do envelope de ligação da membrana viral com o complexo RNP. Fonte modificada a partir de [www.niehs.nih.gov/research/atniehs/labs/lrb/enviro-gen/studies/rsv/index.cfm](http://www.niehs.nih.gov/research/atniehs/labs/lrb/enviro-gen/studies/rsv/index.cfm).

### O Ciclo de Replicação

A penetração do vírus na célula inicia-se com a ligação do vírus à célula através da proteína G. O envelope viral se funde com a membrana plasmática da célula hospedeira por meio da proteína F, sendo este um passo necessário para a transferência da ribonucleoproteína viral para o interior da célula. Após a penetração, o nucleocapsídeo do vírus é liberado no citoplasma celular, onde inicia a transcrição de seu material genético. O RNA viral é usado como modelo para síntese do RNA mensageiro (mRNA). O mRNA, por sua vez, é usado como modelo para tradução das proteínas virais e transcrição do RNA do vírion<sup>10</sup>. Após a síntese de suas proteínas e montagem dos vírions, as partículas virais são liberadas das células infectadas por brotamento, causando a ruptura e fusão com as células adjacentes<sup>11</sup>. O ciclo de replicação do hRSV está esquematizado na **figura 3**.



**Figura 3: Esquema do ciclo de replicação do hRSV.** A proteína G reconhece receptores celulares desencadeando a fusão do envelope viral com a membrana do hospedeiro por meio da proteína F. Após a liberação do nucleocapsídeo, o RNA viral é transcrito em mRNA e este, por sua vez, é usado como modelo para tradução das proteínas virais e transcrição do RNA do vírion. Com os vírions montados as partículas virais são liberadas por brotamento para infecção de novas células. Adaptado de <http://www.profpc.com.br/Virus/Virus.htm>.

### As Proteínas Virais

A proteína de matriz M forma uma camada no lado interno do envelope viral. Ela desempenha um papel importante na montagem do vírus e brotamento pela mediação da associação entre a nucleoproteína viral e a membrana celular<sup>12</sup>.

As proteínas N, P e L são necessárias para a replicação do RNA. A principal proteína do complexo ribonucleoproteico é a nucleoproteína N. Ela se liga ao RNA genômico e antígenômico que confere resistência à RNase no complexo ribonucleoproteico. A fosfoproteína P é a principal espécie fosforilada. Ela funciona como um chaperonina e é

essencial, juntamente com a proteína N para a atividade de encapsidação. A proteína P é também um cofator da polimerase. Parece converter a polimerase por meio da sua fosforilação que é obrigatória para essa função<sup>13</sup>.

A proteína M2-1 é um fator de processamento de transcrição e é essencial para a replicação viral. Ela interage com a proteína N do nucleocapsídeo através da mediação do RNA<sup>14,15</sup> com a proteína P<sup>16</sup>. A fosforilação da M2-1 parece ser indispensável para a interação com a proteína P<sup>12,17</sup>. Os vírus recombinantes com a deleção dos genes da proteína M2-2 são atenuados *in vitro* e *in vivo* em relação ao tipo selvagem<sup>18</sup>.

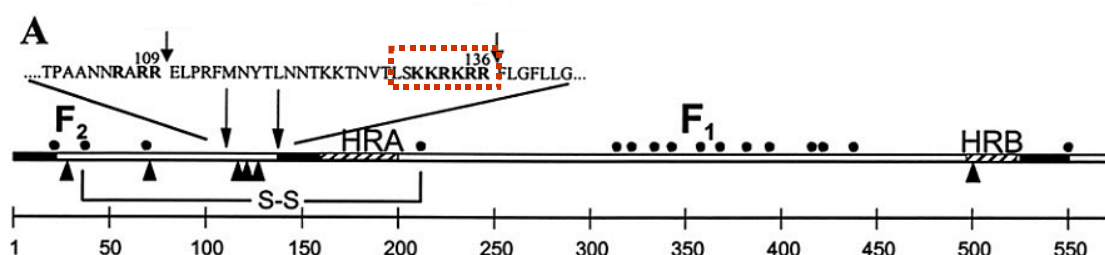
NS1 e NS2 são proteínas com uma massa estimada em cerca de 14-15 kDa. Elas são encontradas somente em *Pneumoviruses* e foram classificadas como proteínas não estruturais, uma vez que são encontrada apenas em traços nas partículas virais purificadas<sup>19</sup>.

As proteínas F, G e SH estão localizadas na superfície do envelope do vírion e estão diretamente relacionadas à adsorção, fusão e replicação viral<sup>20</sup>.

A proteína de fusão (F) é uma das proteínas de membrana que apresentam importante papel para replicação viral. Sua principal função é mediar o processo de adsorção e propagação viral, sendo também responsável pela formação do sincício, intervindo na fusão da membrana das células infectadas com a membrana das células adjacentes<sup>21</sup>.

Outra característica comum da proteína F do hRSV com a de outros vírus membros da sua família *Paramixovirus*, é a síntese da proteína como um precursor inativo, denominado F0 que, para desempenhar sua função, necessita ser clivado. Este precursor é sintetizado com 574 aminoácidos e peso molecular de 67 kDa. A clivagem é feita por proteases do tipo furina durante o transporte para a superfície da célula nas posições 109 e 136, resultando em dois polipeptídios, F1 e F2, ligados por uma ponte dissulfeto, sendo o polipeptídio F2 proveniente da extremidade N-terminal e o polipeptídio F1 proveniente da

extremidade C-terminal. Como um dos resultados da clivagem, tem-se um pequeno peptídeo composto de 27 aminoácidos (pep27), localizado entre os dois sítios de clivagem. Todos os três possuem cadeias de oligossacarídeos N-ligados<sup>21</sup>. Na **figura 4** está representado a composição da proteína F.

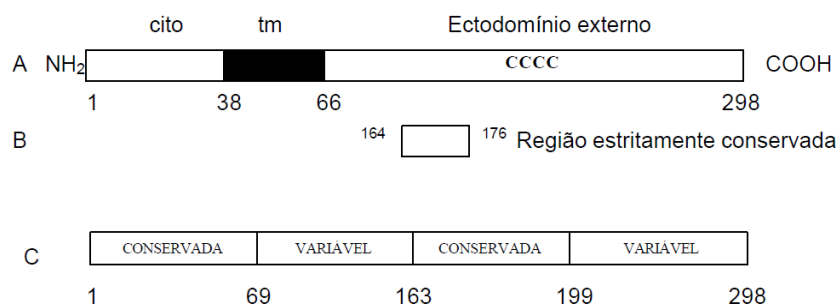


**Figura 4: Representação esquemática do genoma da proteína F do hRSV.** As setas indicam o ponto de clivagem da proteína F<sup>21</sup>.

A glicoproteína G é uma proteína transmembrana integral do tipo II (ou seja, está ancorada à membrana próximo à sua porção amino terminal por um domínio hidrofóbico que se estende do resíduo 38 ao 66), altamente glicosilada, com uma região localizada no citoplasma (N-terminal), uma região transmembrana e um grande ectodomínio (C-terminal)<sup>22</sup>. Esta proteína é sintetizada como um precursor de peso molecular de 36 kDa, sendo posteriormente glicosilada originando um complexo intermediário de 45 kDa. Estes açúcares são convertidos para tipos complexos produzindo uma forma madura de 90 kDa.

Sua variabilidade e glicosilação estão localizadas principalmente no ectodomínio, que contém dois segmentos hipervariáveis com alto teor de serina e treonina, entre os aminoácidos 69-164 e 207-298, com aproximadamente 56% de divergência entre as estirpes A e B<sup>23</sup>, separados por uma região altamente conservada entre os aminoácidos 164 e 199, e quatro resíduos de cisteínas (173, 176, 182, 186). Esta sequência motivo de 13 aminoácidos (173-186) é considerada como provável sítio de ligação ao receptor da célula hospedeira<sup>23,24,6</sup>. A proteína G também interage com o receptor celular CX3CR1, modulando

assim a resposta imunológica do hospedeiro e facilitando a infecção<sup>25</sup>. A representação esquemática da proteína G pode ser visualizada na **figura 5**.

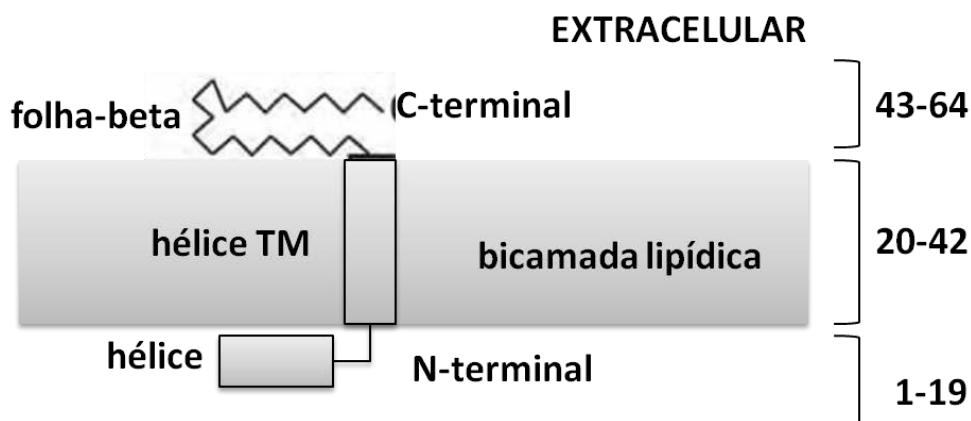


**Figura 5: Representação esquemática da glicoproteína G do hRSV.** A) Estrutura primária de aminoácidos sendo: TM = domínio transmembrana; cito = domínio interno à membrana citoplasmática; CCCC = quatro resíduos de cisteínas; B) Região conservada nas amostras de hRSV; C) Representação das duas regiões variáveis da proteína G, ricas em serina e treonina, compreendendo os aminoácidos 69-163 e 199-298 e região conservada compreendendo os aminoácidos 164-199. As regiões transmembrana e citoplasmática, compreendendo os aminoácidos 1-69, são conservadas. Extraído de SULLENDER, 2000.

## 2. PROTEÍNA SH

### Topologia, Estrutura e Localização

A proteína SH é a menor glicoproteína de superfície codificada pelo RSV<sup>26,27</sup>, com 64 ou 65 aminoácidos dependendo do tipo viral, A ou B, respectivamente<sup>28</sup>. Assim como a proteína G, estudos bioquímicos têm mostrado que a proteína SH também é uma proteína transmembrânica do tipo II com um único domínio transmembrânico (TM) (**figura 6**), onde a região C-terminal é orientada extracelularmente<sup>27</sup>. Experimentos utilizando um peptídeo sintético similar ao domínio TM da proteína SH (resíduo 18-43) indicou que quando inserida em bicamadas lipídicas essa região se apresentava em estruturas helicoidais<sup>29</sup>. Características da estrutura secundária da proteína SH também já foram investigadas por experimentos de infravermelho por Torres e seus colaboradores em 2006, indicando a região N-terminal composta por hélices e a região C-terminal por folhas<sup>30</sup>(**figura 6**).



**Figura 6: Representação esquemática da estrutura secundária da proteína SH do hRSV.** Região transmembrânica (TM) compreendida entre os aminoácidos 20-42 composta por uma alfa-hélice, a região N-terminal (1-19) compostas por hélices e a região C-terminal por folha-beta<sup>30</sup>.

Estudos realizados em células infectadas por RSV mostram que a maioria da proteína SH acumula-se em membranas do complexo de Golgi, mas também são encontradas nas membranas do retículo endoplasmático e no citoplasma<sup>31</sup>. Durante a infecção viral, várias formas da proteína SH estão presentes<sup>32</sup>: duas não glicosiladas (um na forma truncada com 4,5 KDa - SH<sub>t</sub>; outra completa com 7,5KDa - SH<sub>0</sub>); uma com uma glicosilação na região N-terminal - SH<sub>g</sub>; e uma forma modificada de polilactosaminoglicanos - SH<sub>p</sub>. Todas essas formas, exceto a forma truncada não glicosilada estão presentes na superfície das células infectadas, onde a forma não glicosilada SH<sub>0</sub> é a forma mais abundante<sup>27</sup>.

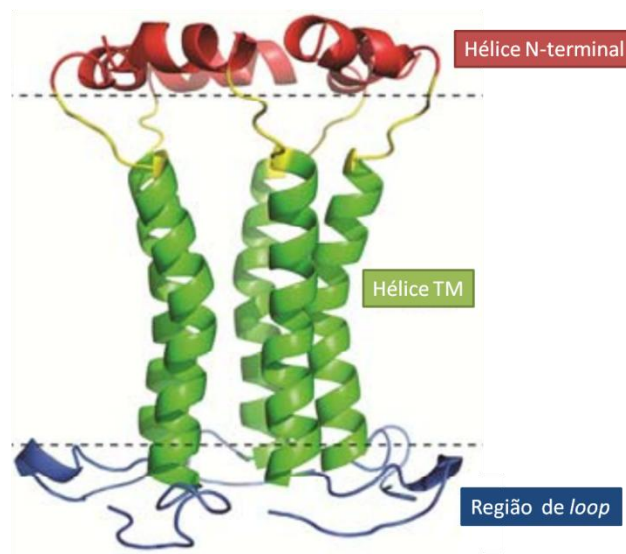
### Papel da Proteína SH

Ao contrário das proteínas F e G, pouco se sabe sobre a função desempenhada pela proteína SH na infecção e replicação do hRSV. Alguns estudos sugerem que esta proteína, juntamente com as outras proteína de superfície G e F, participa no processo de fusão celular mediada pelo vírus<sup>33</sup>. Recentemente, a proteína SH do hRSV, assim como outras proteínas de outros vírus da família *paramixoviridae* (Simian, Parainfluenza, Mump), têm mostrado inibir a apoptose celular em várias linhagens de células de mamíferos<sup>34,35,36</sup>,



essa inibição de apoptose em células do hospedeiro durante a infecção pode dar vantagens ao vírus durante sua replicação, pois aumenta o tempo de vida celular dando ambiente propício a replicação viral.

Perez e seus colaboradores, em 1997, mostraram que a proteína SH também aumenta a permeabilidade da membrana para compostos de baixo peso molecular quando expressa em *E. coli.*, sugerindo que esta proteína pertence à classe viroporina, um grupo pequeno de proteínas virais altamente hidrofóbicas que podem se oligomerizar e formarem poros<sup>29</sup>. A hipótese da formação do poro foi confirmada quando Gan e seus colaboradores em 2008, com um peptídeo sintético correspondente ao domínio TM da proteína SH (SH-TM) mostraram a forma pentamérica de um canal seletivo de íon formado pelo peptídeo SH-TM em um modelo planar de bicamadas lipídicas<sup>37</sup>. Em 2012, Gan e seus colaboradores examinaram a estrutura da proteína SH em micelas e bicamadas fosfolipídicas confirmando a região TM como sendo formada por alfa-hélice e utilizando análises por experimentos de RMN propôs a estrutura quaternária da proteína SH, confirmando sua formação do pentamérica<sup>38</sup> (figura 7).



**Figura 7: Estrutura geral do pentâmero formado pela proteína SH segundo Gan e seus colaboradores<sup>38</sup>. Região N-terminal e TM composta por hélices e C-terminal por loops.**

A atividade do canal iônico realizada pelo pentâmero da proteína SH pode levar a um vazamento iônico mudando o ambiente celular por dissipação do potencial da membrana e perturbações da homeostase (equilíbrio) celular, sendo essas modificações uma das formas apontadas como fator que possibilita aumento da estabilidade do vírus no ambiente celular ou ainda permite desempenhar um papel importante na apoptose celular que não está claramente descrito até o momento. Acredita-se ainda que estas atividades tenham como finalidade contribuir para a liberação de novas partículas virais na célula para aumentar a efetividade de células vizinhas, e ainda existe relatos que esta proteína tem funções importantes no recrutamento de células e mediadores do sistema imunológico do hospedeiro.

Estudos futuros sobre atividade vioporina da proteína SH pode ser de grande importância para melhor compreensão do ciclo viral do hRSV e seu papel durante a infecção pelo hRSV. A descrição desta proteína como vioporina e associação futura com as demais proteínas de superfície viral poderão apontar o mecanismo no processo de efetividade viral e a contribuição da atividade do canal para o sucesso ou fracasso da infecção viral. Deste modo entender como o poro está organizado na membrana, a disposição da estrutura proteica na bicamada lipídica e seu comportamento estrutural com a mudança do ambiente físico-químico contribuirá de forma substancial no entendimento do função da proteína SH.

### **3. OBJETIVOS**

Estudos envolvendo a proteína SH relatam que esta proteína está ligada efetivamente na infecção viral e sua interação com a membrana seja de grande relevância na elucidação no seu mecanismo de ação. Sendo assim, o presente trabalho tem por objetivo:

- I. Modelar e caracterizar a proteína SH;

- II. Analisar o comportamento estrutural da proteína em diferentes meios: água e bicamada fosfolipídica;
- III. Montar a estrutura pentamérica;
- IV. Correlacionar a função desta estrutura com a importância para atividade viral;

#### **4. MATERIAIS E MÉTODO**

Apesar das proteínas de superfícies serem as peças principais na infecção viral, tratá-las experimentalmente (expressão e purificação) ainda é um grande desafio devido às suas regiões altamente hidrofóbicas. Em geral, a expressão dessas proteínas ocorre em baixa escala, fazendo necessário o uso de artifícios (síntese de peptídeo e produção de proteínas recombinantes) que podem afetar fortemente os resultados finais. Devido a essas dificuldades, a utilização de ferramenta de bioinformática e simulações de Dinâmica molecular (DM) no estudo de sistemas envolvendo proteínas transmembrânicas associados aos dados experimentais existentes e modelos conceituais, tem proporcionado compreensão em nível molecular, além de interpretações e modelos para os dados experimentais.

##### **FERRAMENTAS DE BIOINFORMÁTICA UTILIZADAS**

###### **NCBI**

O NCBI, Centro Nacional de Informações Biotecnológica, é um banco de dados que oferece acesso a informações biomédicas e genômicas, além de informações para análise de estrutura e função de biomoléculas. Este banco de dados cria sistemas automatizados para armazenar e analisar o conhecimento sobre a biologia molecular, bioquímica e genética, facilitando o uso dessas informações para a pesquisa. No presente trabalho o NCBI foi utilizado para a procura de sequências através do banco de dados

*GenBank*<sup>39</sup> e referências da proteína SH. Este banco de dados pode ser facilmente acessado pelo site <http://www.ncbi.nlm.nih.gov/>.

### **BioEdit**

O BioEdit é um editor de alinhamento de sequências biológicas de livre uso e pode ser adquirido pelo site <http://www.mbio.ncsu.edu/bioedit/bioedit.html>. Este editor oferece uma variedade de recursos úteis automatizados entre eles o alinhamento ClustalW<sup>40</sup> e pesquisas BLAST<sup>41</sup>. No presente estudo foi utilizada a ferramenta ClustalW<sup>40</sup> oferecida pelo BioEdit para alinhamento das sequências da proteína SH retiradas do banco de dados NCBI. A pesquisa BLAST<sup>41</sup> também foi utilizada para verificar a existência de homologia da proteína SH com outras proteínas com estruturas já conhecidas. O alinhamento foi necessário para analisar a conservação das sequências da proteína SH e a pesquisa BLAST<sup>41</sup> para optar pelo melhor método de modelagem da proteína em estudo.

### **PredicProtein**

PredicProtein<sup>42</sup> é um servidor que baseado em pesquisas de banco de dados de proteínas, a partir de uma dada sequência de aminoácidos, faz previsões de estrutura secundária, acessibilidade do solvente, hélices transmembrânicas, sítios de ligação proteína-proteína e proteína-DNA, modificações pós-traducionais, presença de pontes dissulfeto, entre outras. Essas análises são feitas rapidamente pelo site <http://www.predictprotein.org/>. No presente trabalho, o PredictProtein<sup>42</sup> foi utilizado para análise estrutural e funcional da proteína SH.

### **PSIpred**

O PSIpred<sup>43</sup> é um servidor de análises de sequências de proteínas que agrega vários métodos de predição de estruturas em um único local. No site <http://bioinf.cs.ucl.ac.uk/psipred/> os usuários podem enviar uma sequência de uma determinada proteína, executar as predições de sua escolha (estrutura secundária, regiões

transmembrânicas, predição de hélice transmembrânica) e receber os resultados da predição via e-mail ou web. O PSIPred<sup>43</sup> foi utilizado para predição de estrutura secundária e regiões transmembrânicas da proteína SH.

### **I-TASSER**

O servidor I-TASSER<sup>44</sup> foi utilizado para modelagem da proteína SH, uma vez que, por se tratar de uma proteína transmembrânica, pouco são suas informações estruturais experimentais. O I-TASSER<sup>44</sup> utiliza o método de *Threading*. Esse método é baseado no princípio em que o número dos diferentes motivos proteicos é bem mais limitado que o número de sequências existentes depositadas no *Protein Data Bank*<sup>45</sup> (PDB). Desta forma, este método tem por objetivo encaixar a sequência alvo de uma proteína de estrutura não conhecida a uma biblioteca de estruturas com motivos conhecidos<sup>46</sup>. É uma ferramenta de fácil acesso, sendo necessário apenas um registro para utilizá-lo. O site onde o registro e as predições de estruturas podem ser realizadas é o <http://zhanglab.ccmb.med.umich.edu/I-TASSER/>.

### **SIMULAÇÃO POR DINÂMICA MOLECULAR**

As simulações de Dinâmica Molecular (DM) têm sido rotineiramente utilizadas para investigar a estrutura e função de biomoléculas<sup>47,48</sup>. Sua metodologia é fundamentada nos princípios da Mecânica Clássica e tem como ideia principal descrever como as posições, as velocidades e as orientações dos átomos que constituem o sistema em estudo variam em função do tempo.

As moléculas são tratadas como uma coleção de átomos pontuais ou esferas de Van der Waals que interagem entre si através das forças que atuam sobre cada partícula. Essas forças são calculadas através da primeira derivada do potencial em relação às posições e reavaliadas a cada passo da simulação. O potencial deve ser complexo o suficiente para

uma análise realística do sistema em estudo, no entanto, simples para uma maior eficiência dos cálculos.

A partir da estrutura tridimensional (3D) de uma proteína, por exemplo, um arquivo PDB (arquivo que contém as coordenadas dos átomos das estruturas de uma determinada proteína) sabe-se as posições de cada átomo. Com as coordenadas e os momentos de cada átomo podendo escrever a Hamiltoniana como sendo a soma das energias cinéticas e potencial do sistema:

$$H = \sum_{i=1}^N \frac{(\bar{p}_i)^2}{2m_i} + V\{\{\bar{x}_i\}\} \quad (1)$$

Sendo  $N$  e  $m_i$  o número total de átomos contidos e a massa de cada átomo respectivamente.

A partir da primeira derivada da equação 1 em relação às posições e momentos desses átomos é possível obter as equações de movimento de Newton (equações 2 e 3) que fornecem a evolução temporal de cada partícula do sistema:

$$\frac{\partial H}{\partial \bar{x}_i} = \frac{\partial V\{\{\bar{x}_i\}\}}{V\{\{\bar{x}_i\}\}} = \bar{F}_i = m_i \frac{d\bar{v}_i}{dt} \quad (2)$$

$$\frac{\partial H}{\partial \bar{p}_i} = \frac{\bar{p}_i}{m_i} = \bar{v}_i = \frac{d\bar{x}_i}{dt} \quad (3)$$

Onde  $F$  é a força sobre o átomo  $i$  e  $t$  é tempo.

As coordenadas de posições e velocidades das partículas são gravadas em intervalos de tempo durante as simulações por dinâmica molecular. Essas coordenadas em função do tempo são utilizadas posteriormente para análises do sistema.

### **Campo de Força**

Para descrever a energia potencial total do sistema são utilizadas funções potenciais clássicas, uma vez que um tratamento quântico seria bastante custoso computacionalmente. Desta forma, os parâmetros utilizados para descrever a energia potencial são ajustados a partir de cálculos quânticos e dados experimentais e a esse conjunto de parâmetros é denominado Campo de Forças.

De maneira geral podemos dividir os potenciais em duas categorias: potenciais ligados (torcional e harmônicos); e potenciais não-ligados (reapresentam os átomos separadas por três ou mais ligações químicas).

Os potenciais ligados descrevem as interações covalentes entre dois átomos, distorção do ângulo formado por três átomos e torção do ângulo diedro formado por quatro átomos.

Nas Interações covalentes entre dois átomos (equação 4) este potencial é descrito em termos das posições de equilíbrio de ligação covalente ( $r_0$ ) e da constante elástica de ligação ( $K_b$ ) que é obtida através de estudos espectroscópicos;

$$V_l = K_B(r - r_0)^2 \quad (4)$$

Em relação à distorção do ângulo formado por três átomos ligados covalentemente (equação 5) o potencial é descrito através do ângulo entre as ligações num certo instante ( $\theta$ ), o ângulo de equilíbrio entre as ligações ( $\theta_0$ ) e a constante de força da ligação ( $K_\theta$ );

$$V_\theta = K_\theta(\theta - \theta_0)^2 \quad (5)$$

Por fim, a descrição da torção do ângulo diedro formado por quatro átomos (equação 6) é feita por expansão em funções periódicas. Nesta função a periodicidade do potencial é dada pela função cosseno, sendo  $n$  o número de mínimos para torção de uma ligação química,  $\gamma$  a diferença de fase podendo ser  $0^\circ$  ou  $180^\circ$ ,  $\phi$  o ângulo diedral e  $K_\phi$  a constante que define a barreira de torção.

$$V_{die} = \sum_n K_\phi [1 + \cos(n\phi - \gamma)] \quad (6)$$

Os potenciais não-ligados representam as interações coulombianas (equação 7) entre as cargas parciais atribuídas aos átomos não ligados quimicamente e as interações de Van der Waals (equação 8) descrita pela função de Lennard-Jones. Os potenciais não ligados são apresentados nas equações 7 e 8:

$$V_{vdW} = 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (7)$$

$$V_{ele} = \frac{q_i q_j}{4\pi\epsilon_0} \quad (8)$$

Sendo  $i$  e  $j$  os dois átomos integrantes,  $\epsilon_{ij}$  e  $\sigma_{ij}$  os parâmetros da função de Lennard-Jones e  $q_i$  e  $q_j$  as cargas parciais atribuídas aos átomos  $i$  e  $j$ .

Desta forma, a energia potencial total do sistema,  $V(r)$ , calculada pelo campo de forças é dada pela equação 9:

$$V(r) = V_l + V_\theta + V_{die} + V_{vdW} + V_{ele} \quad (9)$$

Neste trabalho, foi utilizado o campo de forças GROMOS53a6<sup>49</sup> para descrever as interações da proteína SH. Este campo de forças trata os átomos de hidrogênio não-polar como parte de um átomo maior, ou seja, é um campo de forças átomo-unido resultando em 683 átomos para a proteína SH. Nas simulações contendo moléculas de lipídeos utilizou-se o campo de força Berger<sup>50</sup> que é baseado no campo de força OPLS. Este também é um campo de forças átomo-unido, sendo assim, o número de átomos em cada molécula de fosfolípido POPC é reduzido de 130 para 52 átomos. O campo de forças Berger<sup>50</sup> emprega os parâmetros de ligações químicas, ângulos, diedrais e diedrais impróprios para átomos da região da cabeça polar obtidos do GROMOS53a68. Os parâmetros que descrevem as interações entre os lipídeos e a proteína SH foram retirados do tutorial “Montagem de uma



Membrana POPC” fornecido pelo Prof. Dr. Hubert K. Stassen durante a VI Escola de Modelagem Molecular de Sistemas Biológicos realizada em Petrópolis em agosto de 2012.

### **Minimização de Energia**

A função energia potencial de um sistema molecular é muito complexa. Ela possui um ponto de mínimo global e vários outros pontos de mínimo local, onde todas as derivadas da função energia potencial são zero e as segundas derivadas são positivas.

O método de minimização é uma maneira de explorar esta superfície multidimensional de energia. Este método consiste basicamente em caminhar ao longo de toda superfície em estudo em direção da energia mais baixa, ou seja, do mínimo local.

A minimização de energia é um passo bastante importante em uma simulação de dinâmica molecular, uma vez que ela pode ajustar as posições atômicas relaxando os conflitos histórico entre os átomos.

A otimização do processo de minimização consiste na comparação do valor da função energia potencial em duas posições consecutivas  $x_t$  e  $x_{t+1}$  prevalecendo a de menor energia onde  $x_{t+1}$  é um vetor e dado pela equação 10.

$$\vec{x}_{t+1} = \vec{x}_t + \Delta\vec{x}_t \quad (10)$$

Sendo  $t$  o passo que está sendo analisado e  $\Delta\vec{x}_t$  uma perturbação aleatória nas coordenadas que estão sendo otimizadas.

Nas simulações realizadas neste trabalho, utilizou-se dois algoritmos de otimização o *Steepest Descent* (declínio acentuado) e o *Conjugate Gradient* (gradiente conjugado).

### **Steepest Descent (declínio acentuado)**

O *Steepest Descent*, por se tratar de um método de primeira derivada e que converge vagarosamente para o mínimo local, é um método mais eficaz em situações em que o vetor posição se encontra muito afastado do mínimo de energia.

Neste método, o incremento  $\Delta\vec{x}_t$  (equação 11) nas coordenadas de um átomo  $i$  é dado pela direção e sentido da força resultante sobre este mesmo átomo (equação 12).

$$\vec{x}_{t=1} = \vec{x}_t + \Delta\vec{x}_t \quad (11)$$

$$\Delta\vec{x}_t = K_n \left( \frac{\vec{F}_i}{|\vec{F}_i|} \right) \quad (12)$$

Sendo  $\left( \frac{\vec{F}_i}{|\vec{F}_i|} \right)$  um vetor unitário na direção e sentido da força resultante sobre o átomo  $i$  e  $K_n$  um parâmetro de ajuste do passo utilizado para acelerar a minimização de forma que se no passo  $n$  o potencial total dado pela equação 9 diminuir em relação ao potencial  $n-1$  do passo anterior a constante  $k_n$  é aumentada pela multiplicação de uma constante  $C$  positiva e maior que um (equação 13), caso contrário, ou seja, o potencial total  $n$  foi maior que o potencial  $n-1$ , a constante  $C$  será positiva porem menor que um.

$$K_{n+1} = K_n C \quad (13)$$

A cada passo é verificada a diferença entre os potenciais nos passos  $n$  e  $n-1$ . Se esta diferença for menor que um certo valor estabelecido o processo de minimização termina.

### **Conjugate Gradient (gradiente conjugado)**

O Gradiente Conjugado é um método de segunda ordem em que nas vizinhanças do mínimo local, ele converge mais rapidamente que o *Steepest Descent*. Este método utiliza para a determinação do passo seguinte o valor do gradiente no ponto atual e o valor do gradiente obtido no passo anterior (equação 15). Assim  $\Delta\vec{x}_t$  é representado pela equação 14.

$$\Delta \vec{x}_t = K_n \vec{\delta}_{i,n} \quad (14)$$

Onde:

$$\vec{\delta}_{i,n} = \left( \frac{\vec{F}_{i,n}}{|\vec{F}_{i,n}|} \right) + \vec{\delta}_{i,n-1} \left( \frac{|\vec{F}_{i,n}|^2}{|\vec{F}_{i,n-1}|^2} \right) \quad (15)$$

Os dois métodos descritos anteriormente podem ser usados em conjunto, como foi feito em algumas simulações do presente trabalho, sendo usado primeiramente o *Steepest Descent*, uma vez que quando se está muito longe do mínimo este método se faz mais eficaz, e posteriormente, já próximo de atingir o ponto de mínimo, o método de Gradiente conjugado, pois é mais rápido e preciso.

### Métodos de Integração

Muitos métodos de integração numérica estão disponíveis pra resolução das equações de movimento. Vários são rapidamente dispensáveis devido ao alto poder computacional exigido. Para integração numérica das equações de movimento em simulações de dinâmica molecular utilizando o pacote GROMACS<sup>51</sup> é utilizado o algoritmo de *Leapfrog*<sup>52</sup> (pula-cela em português). Este método é o método de integração mais simples usado em simulações de DM. Proposto por *Beeman* em 1976<sup>53</sup>, este método apresenta excelentes propriedades de conservação de energia e os requerimentos de informações sobre estados passados do sistema é mínimo.

O método de *Leapfrog* possui dois passos. Primeiro calcula-se a velocidade no instante de tempo  $t + \Delta t/2$  (equação 16). As posições, então, são calculadas no instante de tempo  $t + \Delta t$  (equação 17), a partir das velocidades calculadas.

$$v \left( t + \frac{\Delta t}{2} \right) = v \left( t - \frac{\Delta t}{2} \right) + \Delta t \cdot a(t) \quad (16)$$

$$r(t + \Delta t) = r(t) + \Delta t \cdot v \left( t + \frac{\Delta t}{2} \right) \quad (17)$$

O nome *leapfrog* vem do fato de que as posições e as velocidades não são calculadas simultaneamente, como consequência disto, as energias cinética e potencial não

podem ser calculadas no mesmo instante. No entanto, este fato pode ser contornado calculando-se a velocidade no instante  $t$  como média das velocidades nos instante  $t + \Delta t/2$  e  $t - \Delta t/2$  (equação 18).

$$v(t) = \frac{v\left(t+\frac{\Delta t}{2}\right)-v\left(t-\frac{\Delta t}{2}\right)}{2} \quad (18)$$

### MODELO BASEADO EM ESTRUTURA

O modelo baseado em estrutura é um modelo minimalista que descreve as propriedades do enovelamento de proteínas e permite explorar uma grande faixa de parâmetros dos sistemas em razoável custo computacional<sup>54,55</sup>.

Esta característica de ser um modelo minimalista e de pouco custo computacional se deve a duas simplificações: a primeira se refere ao potencial de interação entre os componentes do sistema; e a segunda se refere aos próprios componentes do sistema (todos os átomos ou somente os carbonos alfas). O escolha de qual modelo utilizar, modelo com todos os átomos ou modelo  $C_{\alpha}$ , é baseada nas informações que se deseja obter. Se as informações sobre as cadeias laterais são importantes para o entendimento do problema em estudo, como por exemplo, empacotamento e interação entre os resíduos de aminoácidos, aplica-se a modelo com todos os átomos. Caso contrário, o modelo  $C_{\alpha}$  alfa é o mais indicado, pois além de permitir uma economia de tempo computacional é o melhor modelo testado e utilizado. No presente estudo, foi utilizado o modelo com todos os átomos, uma vez que as informações sobre as cadeias laterais são importantes para o entendimento do problema em estudo.

O potencial do modelo baseado em estrutura é construído a partir da conformação nativa do monômero, ou seja, os parâmetros da expressão do potencial são alcançados por meio da estrutura nativa da proteína depositada no *Protein Data Bank* (PDB)

obtidas por técnicas experimentais como cristalografias de raio-X e ressonância magnética nuclear (RMN). Esses modelos cujos potenciais são construído a partir da estrutura nativa são conhecidos como modelo  $G\bar{o}$ <sup>56</sup>. O potencial de uma determinada configuração  $\Gamma$  de uma proteína, tendo  $\Gamma_0$  como sua configuração inicial, é dado pelo expressão:

$$\begin{aligned}
 V(\Gamma, \Gamma_0) = & \sum_{\text{ligações}} \varepsilon_r (r - r_0)^2 \\
 & + \sum_{\text{ângulos}} \varepsilon_\theta (\theta - \theta_0)^2 \\
 & + \sum_{\text{impróprios}} \varepsilon_\chi (\chi - \chi_0)^2 \\
 & + \sum_{\text{cadeia principal}} \varepsilon_{BB} F_D(\phi) \\
 & + \sum_{\text{cadeia lateral}} \varepsilon_{SC} F_D(\phi) \\
 & + \sum_{\text{ligados}} \varepsilon_c \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - 2 \left( \frac{\sigma_{ij}}{r} \right)^6 \right] \\
 & + \sum_{\text{não-ligados}} \varepsilon_{NN} \left[ \left( \frac{\sigma_{NN}}{r} \right)^{12} \right]
 \end{aligned}$$

Com

$$F(\phi) = [1 - \cos(\phi - \phi_0)] + \frac{1}{2} [1 - \cos(3(\phi - \phi_0))]$$

e  $\varepsilon_r = 100$ ,  $\varepsilon_\theta = 20$ ,  $\varepsilon_\chi = 40$  e  $\varepsilon_{NN} = 0,01$  em unidades de  $\varepsilon_c$ , a energia de interação entre os contatos (aproximadamente 1 Kcal/mol por ser um modelo reduzido). Nessa equação,  $r_0$  representa a distância entre dois átomos da estrutura nativa diretamente ligados entre si,  $\theta_0$  é o ângulo entre três átomos consecutivos da estrutura nativa e , analogamente,  $\phi_0$  é o ângulo diedral entre quatro átomos. O parâmetro  $\sigma_{ij}$  do termo de *van der Waals* é determinado através do mapa de contatos entre os átomos na estrutura nativa da proteína.

O mapa de contatos é obtido utilizando o algoritmo *Contact of Structural Units (CSU)*<sup>57</sup>. O parâmetro  $\sigma_{NN}$  serve para manter uma distância de máxima aproximação entre os átomos (sendo para este caso  $\sigma_{NN} = 2,5 \text{ \AA}$ ). Os arquivos de entrada podem ser obtidos facilmente utilizando a interface gráfica do servidor *SMOG@ctbp*<sup>58</sup>.

Neste trabalho, como a proteína SH não possui estrutura depositada no *Protein Data Bank*, utilizou-se as coordenadas do arquivo PDB geradas pelo I-TASSER como estrutura nativa, ou seja, como estrutura inicial.

## 5. ANÁLISES REALIZADAS

### RMSD e RMSF

O valor do desvio da raiz média quadrática (RMSD) é a medida da distância média entre os átomos de uma estrutura de uma molécula em relação a uma estrutura de referência, podendo esta ser a estrutura inicial. No estudo de conformações de proteína o RMSD é calculado através da sobreposição de uma estrutura qualquer com uma estrutura de referência. Esta sobreposição pode ser com todos os átomos ou somente o com os átomos da cadeia principal (N, C $\alpha$  e C). O RMSD pode ser calculado através da equação 19.

$$RMSD(t_1, t_2) = \left[ \frac{1}{M} \sum_{i=1}^N m_i \|r_i(t_1) - r_i(t_2)\|^2 \right]^{\frac{1}{2}} \quad (19)$$

Onde  $m = \sum_{i=1}^N m_i$  e  $r_i(t)$  é a posição do átomo  $i$  no tempo  $t$ .

Já a flutuação da raiz média quadrada (RMSF) calcula a flutuação das posições atômicas durante a trajetória. A partir do cálculo do RMSF é possível também obter um arquivo .pdb com valores de RMSF convertidos em valores do fator beta. O fator beta representa através de um espectro de cores qual região é mais estável ou menos estável durante a simulação, sendo as regiões mais estáveis representadas em azul e as regiões menos estáveis em vermelho.

### **Raio de Giro**

Para analisar a compactação de uma estrutura, pode-se calcular o raio de giro desta estrutura ao longo de sua trajetória. Esse raio de giro pode ser calculado através da equação 20.

$$Rg = \left( \frac{\sum_i \|r_i\|^2 m_i}{\sum_i m_i} \right)^{\frac{1}{2}} \quad (20)$$

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

No presente estudo o raio de giro foi utilizado para análise do comportamento estrutural da proteína SH em solução e escolha da estrutura mais linear para compor o sistema proteína-membrana.

### **Perfil Densidade de Massa**

Através de uma varredura ao longo do eixo  $z$  (eixo normal à membrana) é detectada a quantidade de massa de cada molécula presente em cada ponto que faz parte do sistema. Este mapeamento gera um perfil de densidade de massa PDM. Desta forma, pode-se determinar a região ocupada pelo solvente, a região ocupada pela membrana e a região ocupada pelos resíduos da proteína SH.

Sobrepondo esses perfis de densidade é possível observar a interface solvente-membrana e a altura da membrana, ou seja, sua espessura.

### **Área por lipídeo**

Os lipídeos possuem uma área superficial determinada principalmente pela cabeça polar. Em simulações de membranas, quando o sistema está equilibrado, o valor da área de uma monocamada da bicamada dividido pelo número de lipídeos em cada monocamada tende a permanecer constante. Desta forma, esse área é considerado um parâmetro de

equilíbrio. Para o cálculo da área por lipídeo para os sistemas contendo a proteína SH, foi utilizado o ferramenta de análises de membrana GridMAT-DM<sup>59</sup>.

### **Parâmetro de Ordem**

O Parâmetro de Ordem (P.O.) é um parâmetro normalizado que indica o grau de organização do seu sistema. No presente estudo este parâmetro foi calculado para as cadeias hidrofóbicas (Palmitoil e Oleoil) das moléculas de POPC para análise do seu alinhamento. Quanto mais próximo de zero for o P.O., menor o alinhamento das cadeias alifáticas, sendo um o valor para o estado ordenado absoluto<sup>60</sup>. A expressão usada para o cálculo do parâmetro de ordem está representada na equação 21.

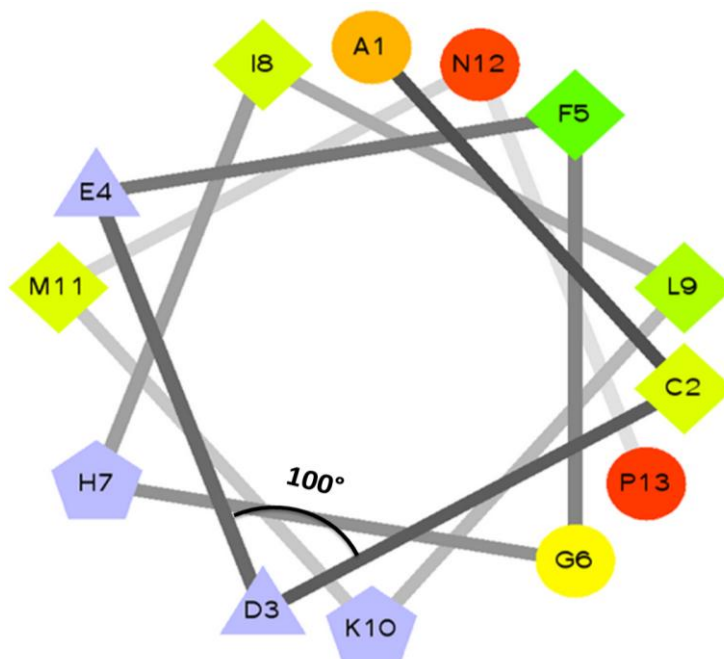
$$PO_i = \frac{1}{2} \langle 3\cos^2\theta_i - 1 \rangle \quad (21)$$

Sendo  $\theta_i$  o ângulo entre o eixo molecular  $i$  e a direção do vetor normal ( $\vec{N}$ ) à bicamada.

### **Helical Wheel (Roda Helicoidal)**

A roda helicoidal (*Helical Wheel*) é uma representação visual usada para ilustrar as propriedades das alfa-hélices de proteínas. A sequência de aminoácidos que forma a região hélice das proteínas são traçadas de forma que o ângulo de rotação entre os aminoácidos consecutivos é de 100° conforme podemos observar na **figura 8**.





**Figura 8:** Um exemplo de uma sequência de aminoácidos representada em uma roda helicoidal indicando o ângulo de rotação entre os aminoácidos consecutivos. Resíduos hidrofóbicos estão representados por diamantes, hidrofílicos por círculos, carregados negativamente por triângulos e carregados positivamente por pentágonos. O grau de hidrofobicidade está representados por cores sendo verde os resíduos mais hidrofóbicos e vermelho os hidrofílicos. O grau de hidrofobicidade e hidrofilicidade variam de acordo com a intensidade das cores verdes e vermelhas. Criado por Don Armstrong e Raphael Zidovetzki.

Desta forma, a representação revela os aminoácidos hidrofóbicos orientados em um dos lados da hélice e os aminoácidos polares ou hidrofílicos para o outro lado. Este arranjo é característico em proteínas globulares e que formam canais iônicos. No presente estudo a análise de roda helicoidal foi utilizada para a montagem da estrutura pentamérica formada pela proteína SH.

## 6. DETALHES COMPUTACIONAIS

### Seleção e Análise das sequências

Foram selecionadas 14 sequências da proteína SH do RSV tipo A a partir do banco de dados internacional GenBank<sup>39</sup>. O programa BioEdit<sup>53</sup> foi utilizado para a realização do alinhamento múltiplo através da ferramenta CLUSTAL-W<sup>40</sup> afim de verificar possíveis mutações entre as sequências.

Predições de estruturas secundárias e regiões transmembrânicas foram feitas usando os servidores PSIPred<sup>43</sup> e PredictProtein<sup>42</sup>. O servidor PredictProtein<sup>42</sup> também foi utilizado para determinar possíveis modificações pós-traducionais.

### **Modelagem da proteína SH**

As informações estruturais experimentais de proteínas de superfícies são poucas, uma vez que estas proteínas possuem uma região altamente hidrofóbica dificultando sua produção em laboratório. Desta forma, há uma necessidade de compreender seus mecanismos e conhecer sua estrutura e função. Sendo assim, a determinação da estrutura terciária da proteína transmembrânica SH foi realizada pela técnica de modelagem molecular usando o servidor I-TASSER<sup>44</sup>.

### **Simulações da proteína SH em água**

Com o objetivo de avaliar as características estruturais da proteína SH em solução e analisar o comportamento das regiões N-terminal e C-terminal em água, foram realizadas simulações das estruturas da proteína SH em solução. Todas as simulações em solução foram executadas com o pacote GROMACS versão 4.5.5<sup>51</sup> usando o campo de forças GROMOS53a6<sup>49</sup>, uma vez que as simulações desenvolvidas visaram sistemas líquidos contendo biomoléculas.

Para a simulação dos sistemas proteína SH em água, foram construídas 3 caixas cúbicas de lado 10 nm, uma para cada modelo, contendo em cada uma o modelo da estrutura da proteína SH, ~ 32800 moléculas de água e um íon Cl<sup>-</sup>.

O modelo de água utilizado foi o SPC "Simple Point Charge"<sup>61</sup>. A otimização dos sistemas foram realizadas com alteração de 5000 passos de *Steepest Descent* (declínio

acentuado) e *Conjugate Gradient* (gradiente conjugado). O N-terminal de cada modelo da proteína em estudo foi tratado como um grupo carregado positivamente ( $\text{NH}_3^+$ ) e o C-terminal carregado negativamente ( $\text{COO}^-$ ), uma vez que é desta forma que essas regiões se apresentam em pH 7.0. Primeiramente foi realizada para cada modelo uma simulação de 20 ps com restrição de posição dos átomos da estrutura da proteína para permitir a relaxação dos íons e dos solventes, em seguida uma simulação de 100 ns foi realizada para aquisição de dados. O passo de integração para as equações de movimento foi de 2 fs. A temperatura e pressão foram mantidas constantes a 1 bar e 310 K, respectivamente, utilizando o algoritmo de Berendsen<sup>62</sup> com constantes de tempo 1.0 ps e 0.1 ps. Condições Periódicas de contorno foram impostas nas simulações com raio de corte igual a 1,4 nm e a lista de vizinhos foi atualizada a cada 5 passos de integração. O algoritmo LINCS<sup>63</sup> foi usado para restringir ligações covalentes da proteína. As interações eletrostáticas de longa distância foram tratadas usando o método PME<sup>64</sup> (*particle mesh Ewald*) com raio de corte de 1 nm. Para o cálculo das contribuições eletrostáticas de curto alcance e das interações de *Van der Waals* empregou-se raio de corte igual a 1,4 nm.

Posteriormente, uma simulação com restrição de posição dos átomos localizados na região transmembrânica foi realizada para cada modelo com a finalidade de analisar o comportamento das regiões N-terminal e C-terminal na água. As condições para as simulações com restrição da região transmembrânica foram as mesma das simulações sem restrição.

### **Construção do sistema Proteína SH - Membrana**

Para estudar o comportamento da proteína SH inserida na membrana fosfolipídica, a estrutura do canal iônico e sua atividade foram necessários o ajuste do

modelo escolhido para uma estrutura mais linear e a construção da bicamada lipídica com dimensões específicas.

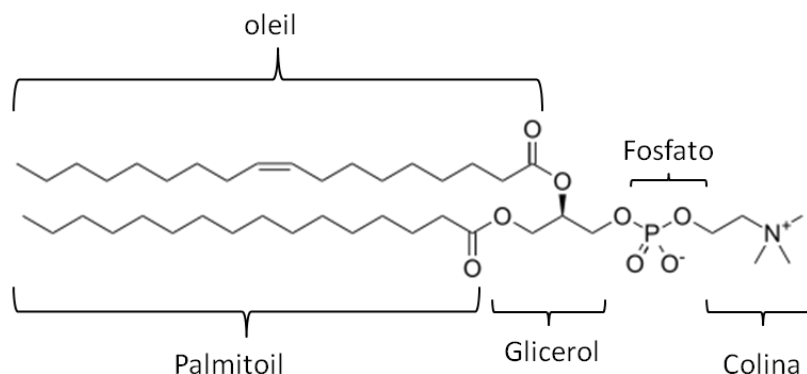
Para gerar a estrutura linear do sistema proteína-bicamada foi utilizado o modelo Baseado em Estrutura<sup>65</sup>.

Para esta simulação foi utilizado o modelo baseado em estrutura com todos os átomos<sup>55</sup> sendo o potencial inicial calculado com base no arquivo de coordenadas atômicas do modelo da proteína SH gerado pelo I-TASSER<sup>44</sup> que apresentou maior linearidade na sua estrutura (modelo 5). Da mesma forma que nas simulações anteriores, o N-terminal da proteína em estudo foi tratado como um grupo carregado positivamente ( $\text{NH}_3^+$ ) e o C-terminal carregado negativamente ( $\text{COO}^-$ ). A simulação foi realizada em temperatura elevada e com restrições de posição nos átomos localizados na região transmembrânica (20-42), uma vez que esta região teria alcançado a estrutura secundária esperada, alfa-hélice. Os arquivos iniciais de topologia (TOP) e estrutura inicial (GRO) para a simulação foram obtidos pela interface gráfica do servidor SMOG@ctbp<sup>58</sup>.

Uma simulação de 20ns foi realizada para gerar um modelo mais linear, ou seja com as regiões terminais mais abertas. A estrutura linear foi escolhida analisando-se o gráfico de raio de giro pelo tempo do sistema simulado. O modelo com maior raio de giro (o modelo mais aberto) foi escolhido para compor a estrutura inicial do sistema proteína-membrana e montagem do poro.

### **Construção da Bicamada fosfolipídica Hidratada**

Moléculas de 1-palmitoil-2-oleilfosfatidilcolina (POPC) foram escolhidas para compor a bicamada fosfolipídica (**figura 9**). Estas moléculas por serem fosfolipídeos zwinteriônicos, ou seja, carga resultante nula, melhor representam as membranas eucarióticas e são bastante utilizadas em experimentos biofísicos.



**Figura 9: Estrutura do lipídio POPC (1-palmitoil-2-oleilfosfatidilcolina).** As chaves indicando os grupos presentes na estrutura molecular do lipídeo: colina, fosfato, glicerol e caudas palmitoil e oleoil.

Para a montagem da bicamada fosfolipídica hidratada foram utilizadas as ferramentas do pacote GROMACS 4.5.5<sup>51</sup> Os parâmetros do lipídeo foram retirados do campo de Forças Berger<sup>50</sup> que possui compatibilidade com a campo de forças GROMOS53a6<sup>49</sup>.

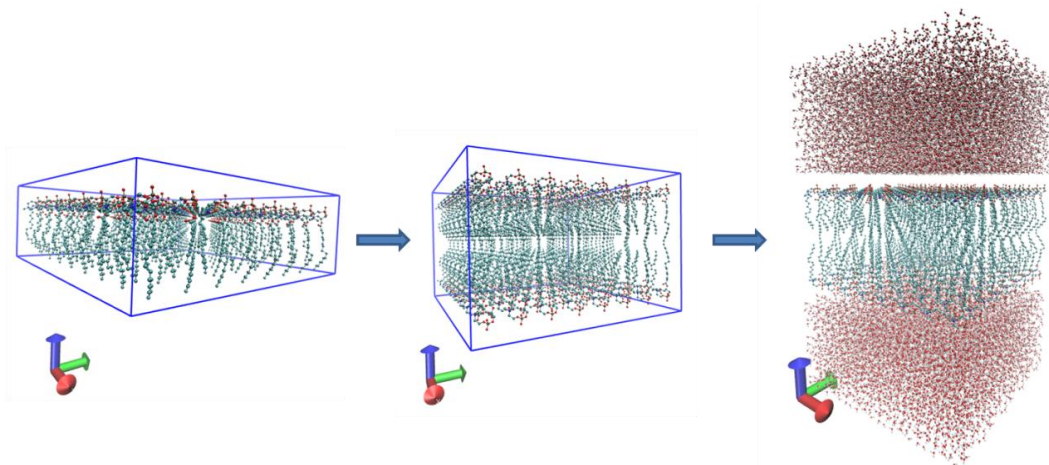
A construção da bicamada foi realizada em três passos:

- 1) Criou-se uma monocamada no plano xy com dimensões 15 nm em x e 15nm em y contendo 225 moléculas POPC;
- 2) Criou-se uma segunda monocamada a partir da inversão da primeira pelo eixo z;
- 3) Finalmente, as duas monocamadas foram acopladas produzindo uma bicamada POPC com 450 moléculas POPC.

Tendo em vista que a bicamada foi formada por configurações idênticas de moléculas POPC, uma minimização de 1000 passos foi realizada para permitir a relaxação das moléculas POPC.

Para hidratar a bicamada construída, foram criadas duas caixas de água com a mesma dimensão xy da bicamada após a minimização (~15 nm), uma para a parte superior da bicamada e outra para a parte inferior. As espessuras das caixas de água (eixo z = 6 nm) foram escolhida com base no comprimento da proteína neste mesmo eixo. O modelo de

água utilizado foi o SPC “Simple Point Charge”<sup>61</sup>. Uma representação esquemática da construção da bicamada fosfolipídica pode ser visualizada na **figura 10**.

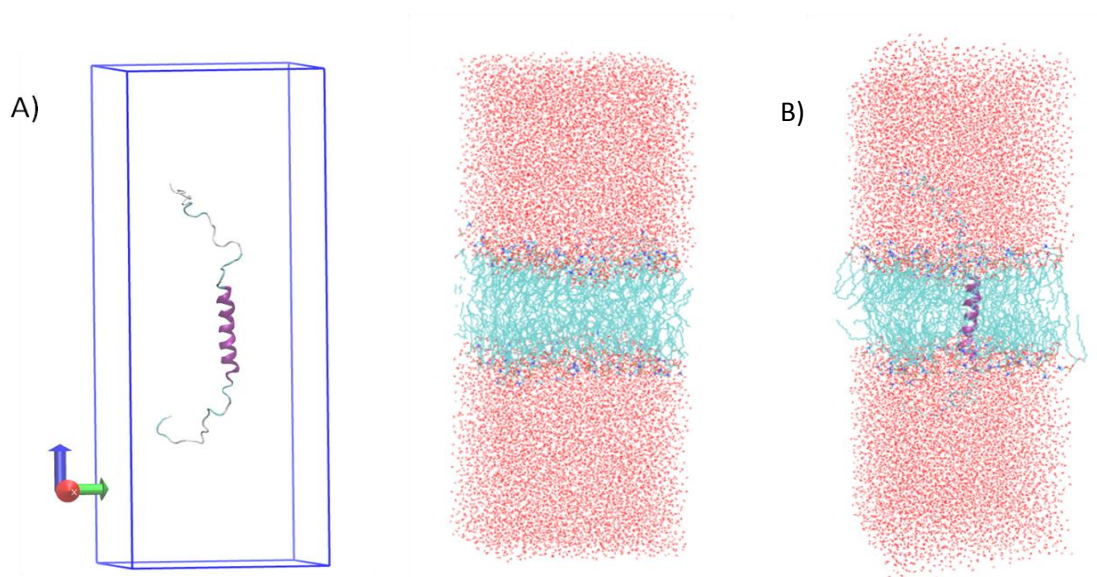


**Figura 10: Passos da construção da bicamada hidratada.** É criada uma monocamada no plano xy e uma segunda monocamada a partir da inversão da segunda. As duas monocamadas são “ligadas”. Duas caixas de água com as mesmas dimensões xy das monocamadas são criadas e “ligadas” uma a parte superior e outra a parte inferior da bicamada. Assim completando a construção da bicamada hidratada.

Após a montagem da bicama hidratada, uma minimização de 1000 passos foi realizada para rearranjo das moléculas e do sistema. Posteriormente, uma simulação de 10 ns foi realizada para aquisição de dados. O passo de integração para as equações de movimento foi de 2 fs. A temperatura e pressão foram mantidas constantes a 1 bar e 310 K, respectivamente, utilizando o algoritmo de Berendsen<sup>62</sup> com constantes de tempo 1.0 ps e 0.1 ps. Condições Periódicas de contorno foram impostas nas simulações com raio de corte igual a 1,4 nm e a lista de vizinhos foi atualizada a cada 10 passos de integração. O algoritmo LINCS<sup>63</sup> foi usado para restringir ligações covalentes da proteína. As interações eletrostáticas de longa distância foram tratadas usando o método PME<sup>64</sup> (*particle mesh Ewald*) com raio de corte de 1 nm. Para o cálculo das contribuições eletrostáticas de curto alcance e das interações de *Van der Waals* empregou-se raio de corte igual a 1,4 nm.

### Proteína SH em Membrana POPC

A estratégia para a inserção da proteína SH na bicamada lipídica consistiu da solvatação da proteína com a membrana. Para isto, as dimensões da caixa criada para a proteína foram as mesmas que as dimensões da caixa da membrana após simulação de 10 ns (11.8 nm, 11.8 nm, 23.5 nm), além de a proteína estar apropriadamente orientada ao longo do eixo z e centrada na caixa. A estratégia para inserção da proteína SH está esquematizada na **figura 11**.



**Figura 11: Estratégia utilizada para a inserção da proteína na bicamada lipídica. A)** Proteína centrada em caixa com as mesmas dimensões da membrana (11.8,11.8,23.5). **B)** Solvatação da proteína com a membrana.

Esse mesmo sistema foi neutralizado pela adição de um íon  $\text{Cl}^-$ . Um minimização de 5000 passos do sistema proteína SH - bicamada foi realizado para rearranjo das moléculas do sistema.

Após uma simulação de restrição de 5 ns do sistema proteína-membrana, observou-se que o comprimento da bicamada hidratada na direção z era bem maior que o comprimento da estrutura da proteína. Desta forma, para melhorar o desempenho das simulações e com a intenção de diminuir seu tempo, diminuimos o tamanho do sistema em z.

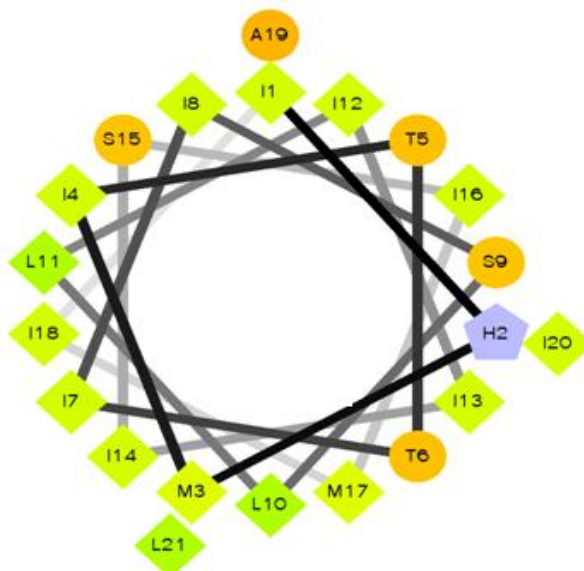
Utilizando o VMD<sup>66</sup> retirou-se todas as moléculas de água localizadas em  $z < 37,5 \text{ \AA}$  e  $z > 197,5 \text{ \AA}$ , resultando em um sistema final com dimensões  $x$ ,  $y$ , e  $z$  iguais a 11.8 nm, 11.8 nm e 16 nm respectivamente.

Posteriormente, a simulação do sistema proteína–membrana foi continuada, porém sem restrições, por mais 24 ns à uma temperatura de 310 K, a fim de analisar características estruturais da proteína SH inserida na membrana. No entanto, verificou-se que o tamanho do comprimento da bicamada hidratada na direção  $z$  ainda era bem maior que o comprimento da estrutura da proteína. Desta forma, o tamanho da caixa em  $z$  foi diminuído mais uma vez utilizando o VMD<sup>66</sup> para melhor desempenho das simulações. O sistema proteína–membrana final ficou com dimensões  $x$ ,  $y$ , e  $z$  iguais a 11.8 nm, 11.8 nm e 12 nm respectivamente. Os demais parâmetros da simulação foram similares aos da simulação anterior. A simulação do sistema proteína SH – membrana foi continuada por mais 120 ns à uma temperatura de 310 K.

### **Estrutura do Pentâmero**

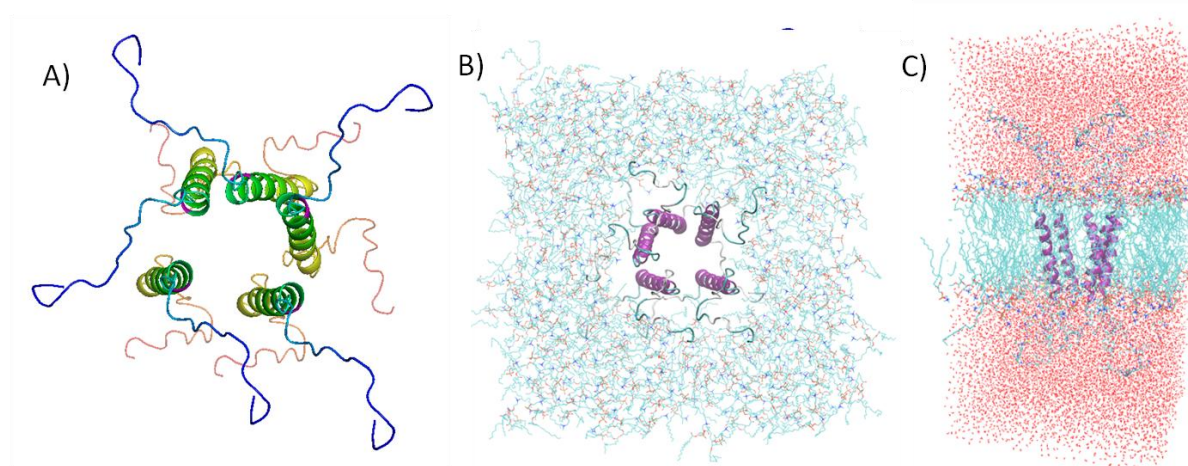
Para analisar a estrutura pentamérica da proteína SH, primeiramente foi realizada uma busca de como orientar cada oligômero dessa estrutura quaternária. A proteína SH possui 64 resíduos sendo os resíduos 20 ao 42 referentes a região transmembrânica e composta por alfa-hélice<sup>38</sup>. Assim, baseou-se na análise de *Helical Wheel* (<http://rzlab.ucr.edu/scripts/wheel/wheel.cgi?sequence=ABCDEFGHIJKLMNPO&submit=Submit>) para propor a primeira orientação: quais resíduos permaneceriam voltados para dentro do poro (resíduos de caráter mais hidrofílico) e quais estariam em contato com os lipídios (caráter hidrofóbico).





**Figura 12: Helical Wheel dos resíduos da proteína SH.** Os resíduos hidrofílicos como círculos, resíduos hidrofóbicos como diamantes, potencialmente carregados negativamente como triângulos, e potencialmente carregados positivamente como pentágonos. O resíduo mais hidrofóbico é verde, e a intensidade do verde diminui proporcionalmente a hidrofobicidade, com zero de hidrofobicidade codificado como amarelo. Resíduos hidrofílicos são codificados com vermelho e a intensidade do vermelho diminui proporcionalmente a hidrofobicidade.

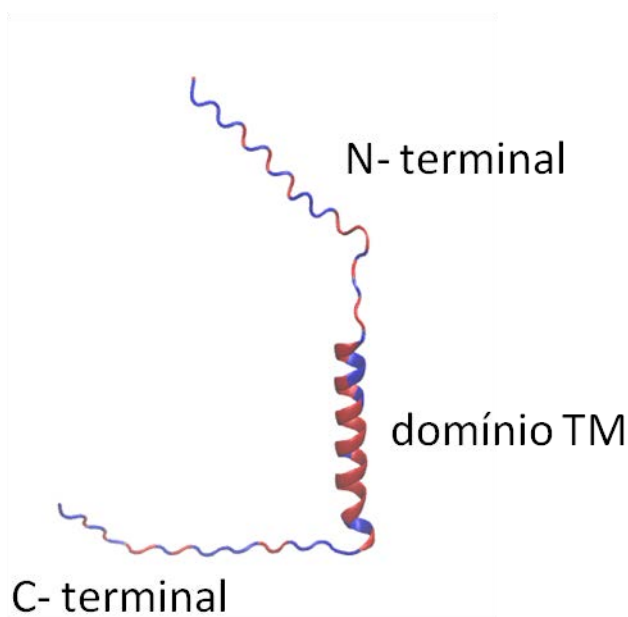
Analisando a **figura 12** e com base nos resultados experimentais<sup>38</sup>, uma conformação do poro foi montada, a poro-membrana, orientando os resíduos de histidinas para o interior do poro (**figura 13**).



**Figura 13: Esquema da montagem do poro formado pela proteína SH com base na análise Helical Wheel.** A) Visualização de cima do poro. B) Visualização de lado do poro. O espectro de cores varia do azul para o vermelho sendo azul a região N-terminal e vermelho a região C-terminal. Em rosa está destacada o resíduo de histidina na posição 22. C) Poro inserido na membrana em roxo o domínio transmembrânico TM da proteína SH.

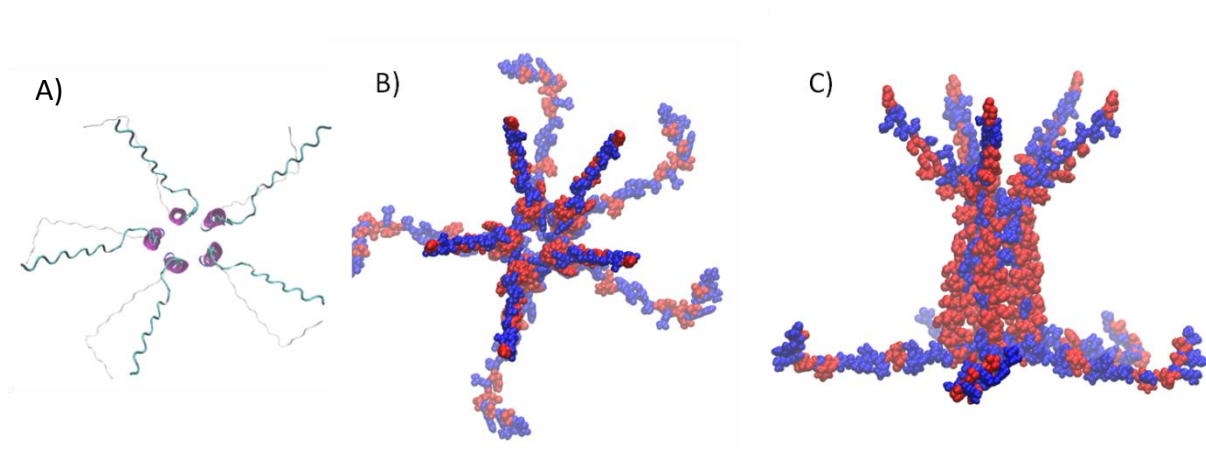
Conforme pode ser observado na **figura 13**, nem todos os resíduos de histidina foram orientados para o interior do poro. Isto ocorreu pois quando os monômeros eram rotacionados, afim de orientar os resíduos hidrofílicos para o interior do poro e os hidrofóbicos para o exterior, ocorria uma sobreposição das regiões terminais.

Sabendo-se que para o arranjo dos monômeros para a formação do pentâmero com as regiões hidrofílicas e hidrofóbicas devidamente orientadas levaria um tempo de simulação razoável devido ao tamanho do sistema, construiu-se uma nova estrutura terciária da proteína SH utilizando a ferramenta **Molecule Factor**(referencia) do VMD<sup>66</sup>, com base no modelo gerado pelo I-TASSER (regiões N e C-terminais sem estrutura e domínio transmembranico composto por hélices).



**Figura 14: Estrutura terciária da proteína SH gerada pelo VMD utilizando a ferramenta *Molecule Factor*.** Região em azul representa a região hidrofílica da proteína e a região em vermelho a hidrofóbica. O domínio transmembranico (TM) foi construído conforme topologia dos modelos terciários criados pelo I-TASSER: região compreendida entre os aminoácidos 20 ao 42 composta por alfa-hélice.

Com a nova estrutura terciária gerada pelo VMD<sup>66</sup> (**figura 14**) um novo poro foi montado com as regiões hidrofílicas e hidrofóbicas devidamente orientadas como podemos observar na **figura 15**.



**Figura 15: Esquema do poro construído pelo VMD. A)** Imagem vista de cima da estrutura secundária do poro construído pelo VMD, em roxo o domínio TM, em azul a região N-terminal. **B)** Imagem vista de cima da estrutura quaternária da proteína SH construída pelo VMD. **C)** Imagem da lateral da estrutura quaternária da proteína SH construída pelo VMD. Em azul estão representadas as regiões hidrofílicas e em vermelho as regiões hidrofóbicas da proteína SH.

Desta forma, foram realizadas duas simulações: uma simulação com o sistema poro-membrana (**figura 14**) com o modelo da proteína SH gerado pelo I-TASSER e , outra com o sistema poro-membrana (**figura 15**) onde a estrutura terciária da proteína SH foi construída pelo VMD<sup>66</sup> com base na topologia gerada pelo I-TASSER (**figura 15**).

Posteriormente, os sistemas foram neutralizados pela adição de cinco íons  $\text{Cl}^-$  cada. Uma minimização de 5000 passos para cada sistema poro-bicamada foi realizado para rearranjo das moléculas do sistema.

Inicialmente, desenvolveu-se uma simulação de 5 ns mantendo as coordenadas dos oligômeros restritas para relaxação do solvente e íons.

Posteriormente as simulações dos sistemas foram continuadas, porém sem restrições à 310 K e utilizando os mesmos parâmetros das simulações anteriores.

## 7. RESULTADOS E DISCUSSÃO

### Seleção e Análises da estrutura primária da proteína SH

A partir de 15 seqüências de aminoácidos da proteína SH, um alinhamento realizado pelo CLUSTAL-W<sup>40</sup> confirmou a característica conservada da proteína, resultando em uma identidade de 100%. A seqüência de código AEO45887.1, foi escolhida para as análises funcionais e estruturais.

Na **figura 16** são apresentadas os resultados das predições das estruturas secundárias da proteína realizadas pelo PSIPred<sup>36</sup>. Tanto o PSIPred<sup>43</sup> quanto o PredictProtein<sup>42</sup> indicaram a região compreendida entre os aminoácidos 20 ao 42 como sendo uma região transmembrânica composta por 1 alfa-hélice. Este resultado está de acordo com resultados experimentais realizados anteriormente, em que um peptídeo sintético da região 18 a 43 da proteína inserida em bicamada lipídica, apresentou a mesma conformação estrutural<sup>37</sup>.



**Figura 16: Predição da estrutura secundária da proteína SH realizada pelo servidor PSIPred.** As barras em azul na primeira linha representa a confiabilidade (Conf) da estrutura; A seta amarela indica uma folha beta na região N-terminal, o cilindro rosa representa alfa-hélice que compreende a região transmembrânica da proteína SH. Linha contínua representa regiões com estruturas não definidas. Na última linha da figura está a seqüência primária da proteína SH.

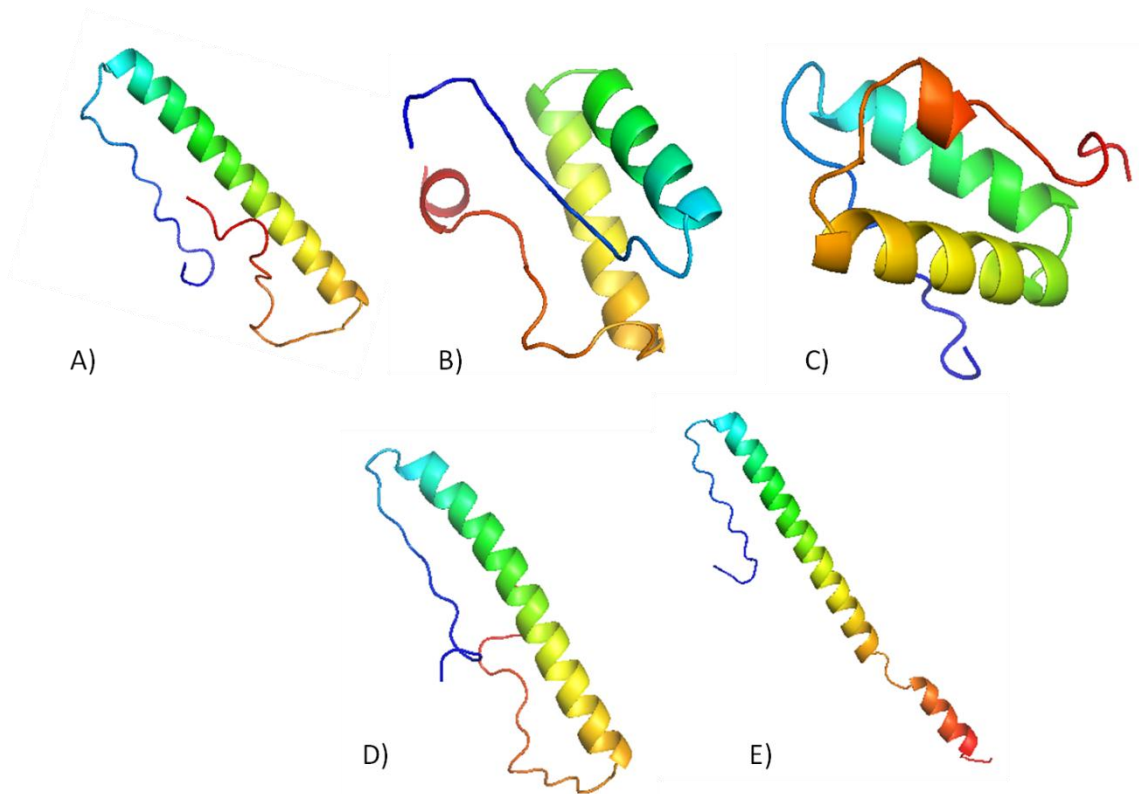
Os resultados preditos pelo PredictProtein<sup>42</sup> mostram ainda a presença de dois sítio de glicosilação nos resíduos N3 e N52 e um sítio de fosforilação no resíduo S11. Essas mudanças pós-traducionais são essenciais, pois determinam a forma como a proteína se apresenta durante a infecção viral podendo ela ser: duas formas não glicosiladas (uma

truncada SH<sub>t</sub> e outra completa SH<sub>0</sub>), uma forma N-glicosilada SH<sub>g</sub> e uma polilactosaminoglicano-modificada SH<sub>p</sub>. Além disso, essas modificações, por estarem localizadas nas regiões terminais, podem estar diretamente ligadas à sinalização para o funcionamento do canal iônico formado pela proteína.

Os sítios de interações com outras proteínas também foram indicados pelo PredictProtein<sup>42</sup>. Na região N-terminal os resíduos E2 e S12 e na região C-terminal N48, V49, F50, H51, K53, P58, R59, R61, V62 e N63. Estes dados concordam com estudos já realizados que mostram possíveis interações entre as proteínas de membrana do RSV: G, F e SH<sup>67,33,68</sup>, indicando que estas interações estejam diretamente ligadas a atividade de fusão do vírus<sup>33</sup>. Além das interações com proteínas virais, a proteína SH é necessária para inibição do fator de necrose tumoral alfa (TNF- $\alpha$ ) induzida por apoptose, sugerindo possíveis interações com proteínas do hospedeiro também<sup>36,69</sup>.

### **Modelagem da Proteína SH**

Os resultados da Modelagem com a sequência de código AEO45887.1, utilizando o servidor I-TASSER<sup>44</sup> gerou 5 modelos da proteína SH (**figura 17**). Desses cinco modelos, escolheu-se os modelos 1, 4 e 5 para prosseguir com as análises (**figura 17: A, D e E**), uma vez que suas estruturas secundárias foram bem similares às estruturas secundárias previstas anteriormente pelo PSIPred<sup>43</sup> e PredictProtein<sup>42</sup> e corroboram com a topologia estrutural dos resultados experimentais obtidos da literatura: região transmembrânica compreendida entre os aminoácidos 20-42, composta por uma alfa-hélice<sup>37</sup>.



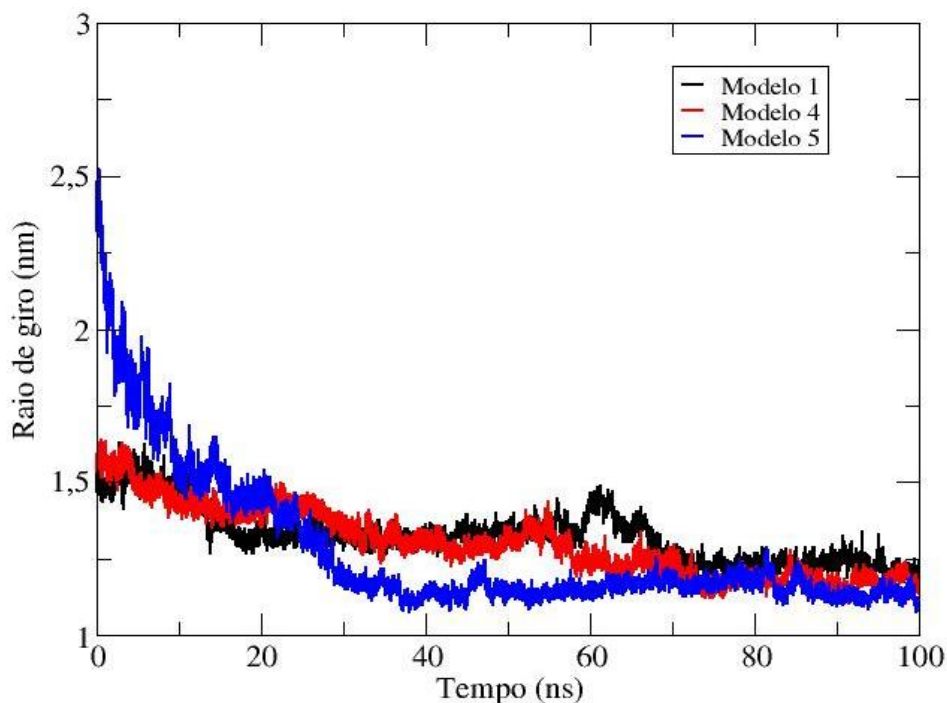
**Figura 17: Modelos da proteína SH gerados pelo servidor I-TASSER<sup>44</sup>. A) Modelo 1; B) Modelo 2; C) Modelo 3; D) Modelo 4; E) Modelo 5. O espectro de cores varia do azul para o vermelho sendo azul a região N-terminal e vermelho a região C-terminal.**

### Simulação da Proteína SH na Água

Com a finalidade de analisar o comportamento da proteína em meio totalmente hidrofílico, um sistema contendo proteína solvatada em água e neutralizada com um íon cloreto foi simulado para cada modelo (1, 4 e 5).

As **figuras 18 e 19** indicam respectivamente o raio de giro e a posição média dos átomos de cada resíduo que compõe cada modelo em função do tempo de simulação. Estes dados indicam um empacotamento da proteína, ou seja, uma diminuição do raio de giro e uma instabilidade principalmente dos resíduos terminais dos modelos que possuem maior liberdade de movimento quando comparados aos que estão no interior da estrutura. Este comportamento é característico de proteínas transmembrânicas que, quando expostas a

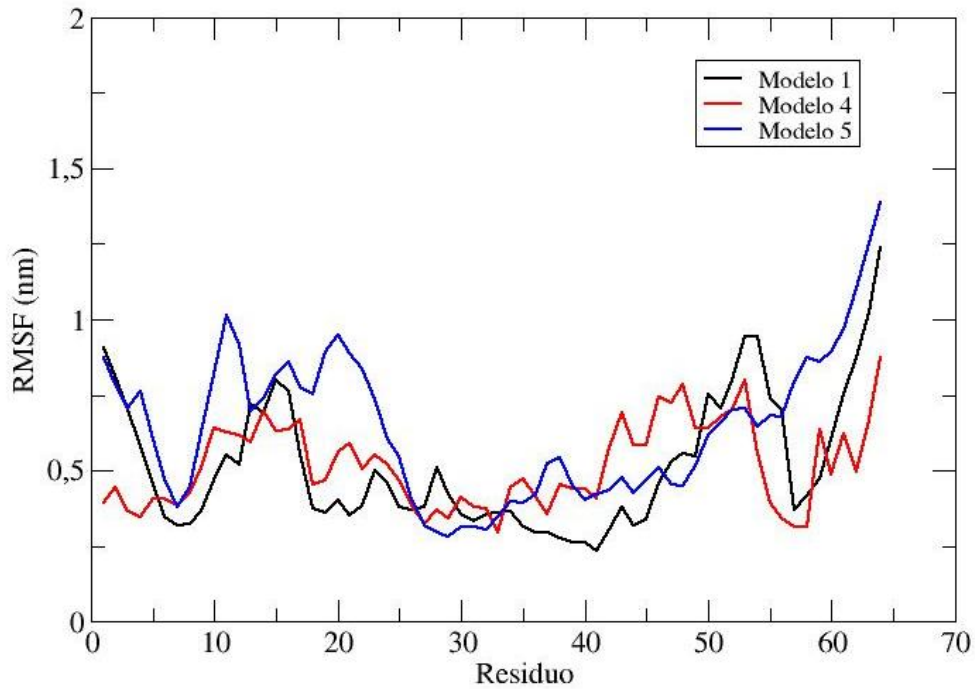
ambientes hidrofílicos, mudam sua conformação estrutural devido ao efeito hidrofóbico causado pelas moléculas de água explicando a dificuldade de tratá-las experimentalmente<sup>70</sup>.



**Figura 18: Raio de giro dos modelos 1, 4 e 5 em função do tempo de simulação.** Simulações em água à 310 K. A queda no valor do raio de giro indica a mudança estrutural dos modelos para uma conformação mais globular.

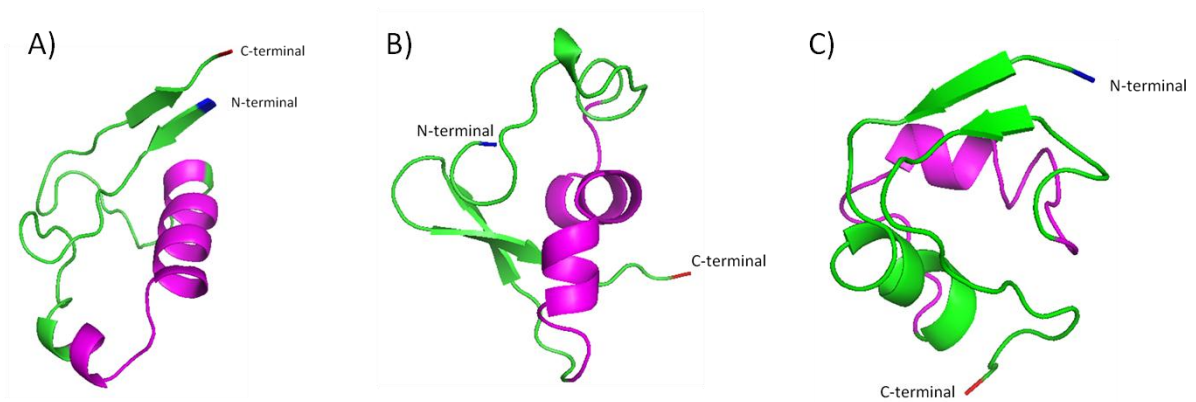
Os resultados da **figura 19** mostra também que esta alta variação da flutuação média dos resíduos terminais dos modelos 1, 4 e 5 comparados aos resíduos compreendido entre os aminoácidos 25 ao 40, reflete a instabilidade da proteína como um todo, indicando a água como sendo um ambiente não apropriado para desempenho de sua função.





**Figura 19: Flutuação média da posição de cada resíduo dos modelos 1, 4 e 5.** Simulações em água a 310 K. As flutuações são maiores nas regiões terminais indicando um maior grau de liberdade das extremidades.

Analisando as estruturas médias dos modelos após a simulação de 100 ns (**figura 20**), confirmamos a mudança estrutural dos três modelos (1, 4 e 5) de uma conformação mais linear para uma conformação mais compacta. Observa-se ainda que o domínio TM dos três modelos sofreu mudança estrutural, principalmente no modelo 5, onde essa região apresenta-se em sua maioria desenovelada.



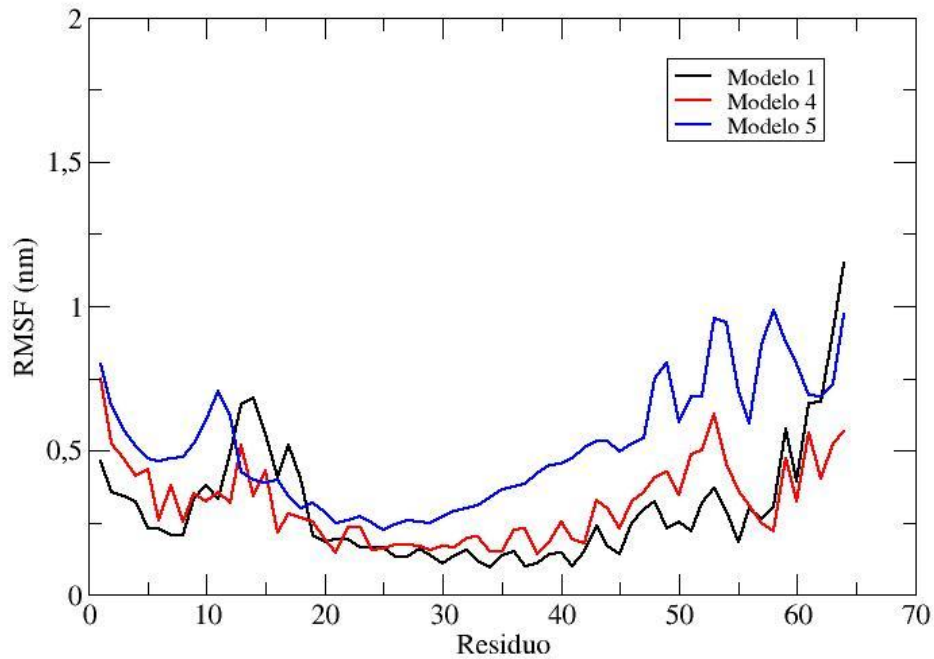
**Figura 20: Modelos da proteína SH após simulação em água a 310 K. A) Modelo 1; B) Modelo 4; C) Modelo 5.** Em rosa a região transmembrânica (TM), em azul a região N-terminal e em vermelho a região C-terminal.



Uma vez que os modelos em estudos já apresentavam a estrutura secundária esperada para o domínio TM, analisou-se o comportamento estrutural apenas das regiões N-terminal e C-terminal em solução.

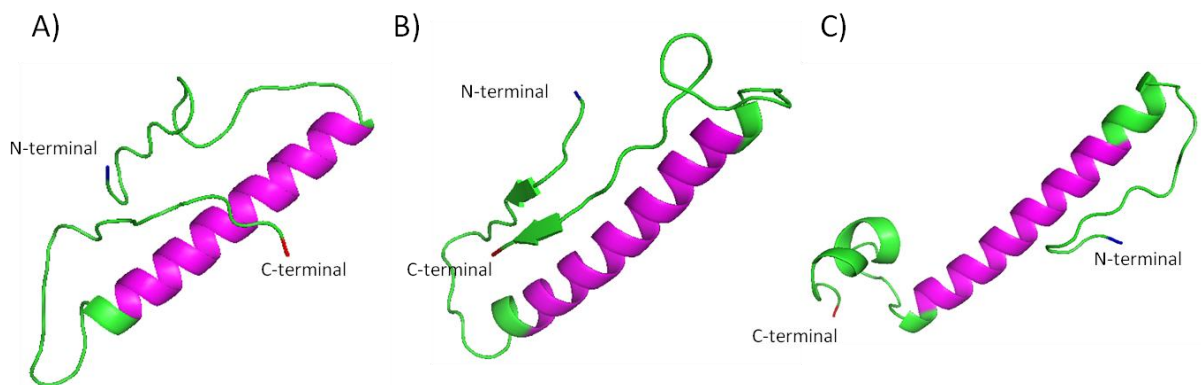
Um sistema contendo proteína solvatada em água, neutralizada com um íon cloreto e com restrição de posição aplicada à região transmembrânica (20-42) foi simulado para cada modelo da proteína SH durante 100 ns.

Comparando a posição média dos átomos de cada resíduo que compõe cada modelo sem restrição do domínio TM (**figura 19**) com a posição média dos átomos de cada resíduo que compõe cada modelo com restrição do domínio TM (**figura 21**), os resultados mostram que há uma maior flutuação nas regiões N-terminal e C-terminal dos modelos sem restrição. Este comportamento é devido a maior liberdade de movimento principalmente nas regiões terminais desses modelos. Devido as restrições dos átomos do domínio TM, as regiões transmembrânica desses modelos apresentam menor flutuação comparado aos modelos sem restrição.



**Figura 21:** Flutuação média da posição de cada resíduo dos modelos 1, 4 e 5. Simulações em água, com restrição da região transmembrânica (20-42) a 310 K. Nos três modelos, as flutuações dos resíduos são menores na região com restrição (20-42) e maiores nas regiões terminais.

Os resultado obtidos em simulação de 100 ns com restrição da região TM, nenhum dos modelos alcançou a estruturação das regiões terminais (**figura 22**), sugerindo que interações específicas entre a proteína e a bicamada ou interações proteína-proteína são importantes para a formação estrutural completa desta proteína.



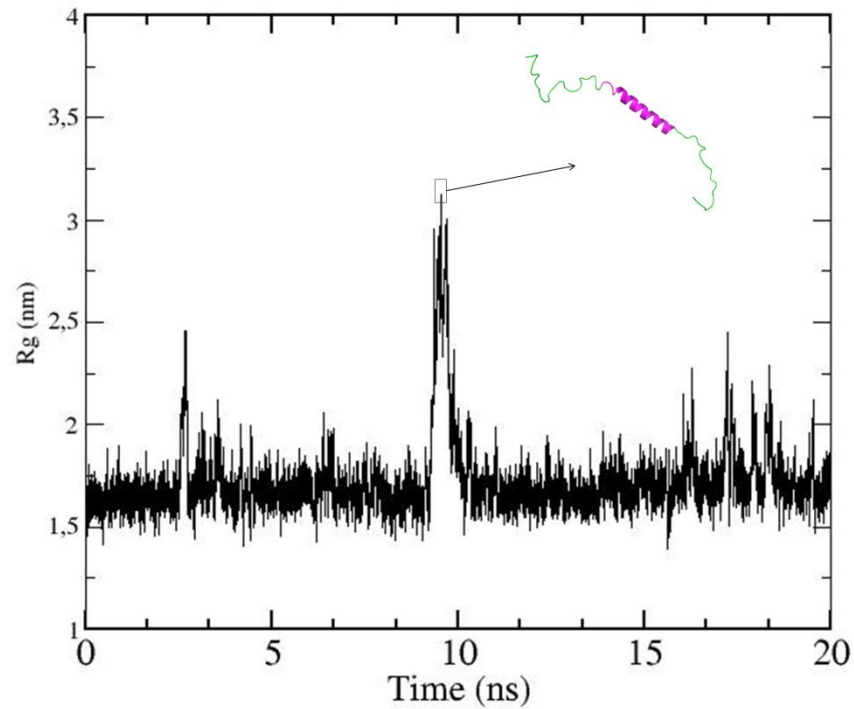
**Figura 22:** Modelos da proteína SH após simulação em água a 310 K com restrições de posição nos átomos localizados na região transmembrânica (20-42). A) Modelo 1; B) Modelo 4; C) Modelo 5. Em rosa a região transmembrânica (TM), em azul a região N-terminal e em vermelho a região C-terminal.

Desta forma, para modelagem completa da proteína SH, houve a necessidade de analisar o comportamento de cada modelo (1, 4 e 5) inserido na bicamada.

No entanto, para a montagem do sistema proteína-bicamada, verificou-se que as extremidades N-terminal e C-terminal estavam em posições não favoráveis, de modo que quando inseridas na membrana fosfolipídica suas extremidades também estariam no interior da membrana. Assim, simulações com restrição de posição aplicada à região transmembrânicas foram realizadas em altas temperaturas de modo a “desenovelar” as regiões terminais dos modelos da proteína.

Após a realização dessas simulações, verificou-se que elas não apresentaram modelos lineares o suficiente para serem inseridos na bicamada lipídica. Desta forma, considerando que ambos os modelos seriam equivalentes quando suas extremidades estivessem “linearmente desenoveladas”, escolheu-se o modelo 5 para gerar a estrutura linear do sistema proteína-bicamada utilizando o modelo baseado em estrutura<sup>55</sup>. O modelo 5 foi escolhido pois sua estrutura era mais linear comparada as estruturas dos modelos 1 e 4.

O modelo inicial escolhido para os sistema proteína-bicamada foi aquele que apresentou o maior raio de giro, em torno de 3,107 nm em 9.509 ps de simulação conforme pode ser observado na **figura 23**.

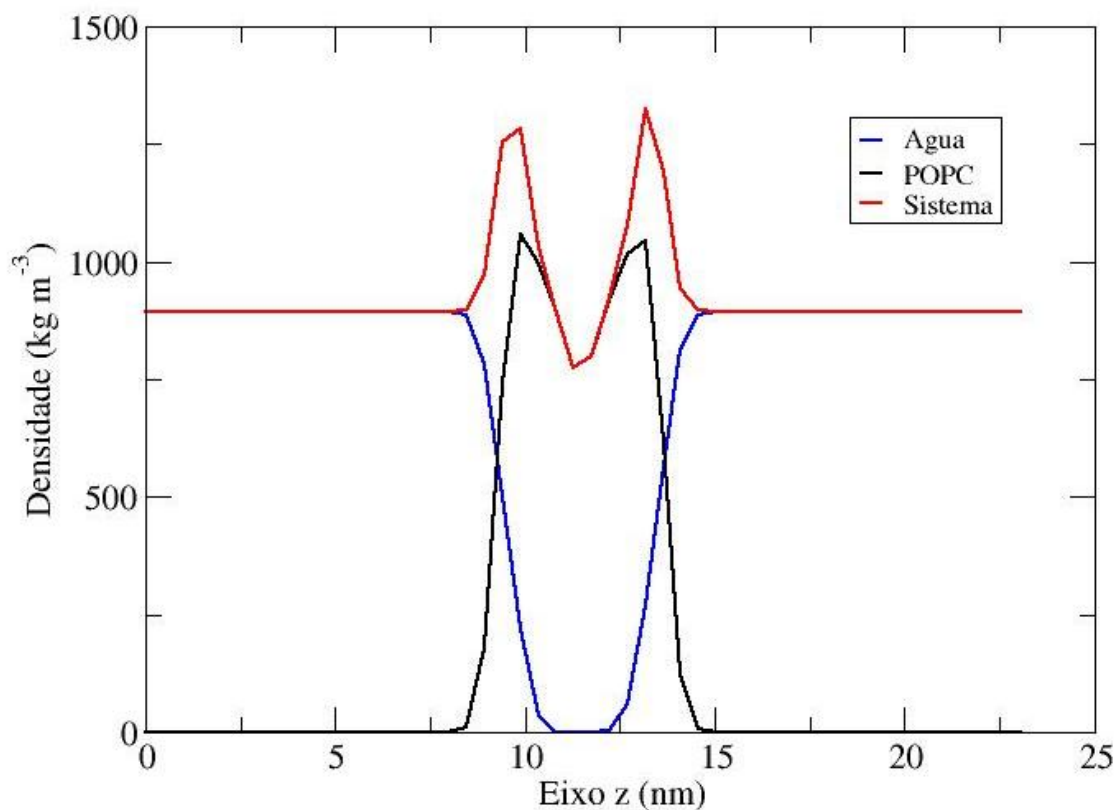


**Figura 23: Estrutura da proteína SH escolhida pelo modelo baseado em estrutura<sup>55</sup> a partir do raio de giro por tempo de simulação.** O modelo escolhido apresentou raio de giro em torno de 3,107 nm em 9.509 ps de simulação. Região rosa corresponde ao domínio transmembrânico (TM).

### **Análises da Bicamada composta por POPC**

O sistema hidratado composto por 225 moléculas de POPC por monocamada foi simulado por 10 ns.

Analisando o gráfico do Perfil de Densidade de Massa (PDM) do sistema bicamada hidratada podemos detectar, através de uma varredura ao longo do eixo z, a quantidade de massa de cada molécula (Água ou POPC) ao longo de cada ponto desse eixo.



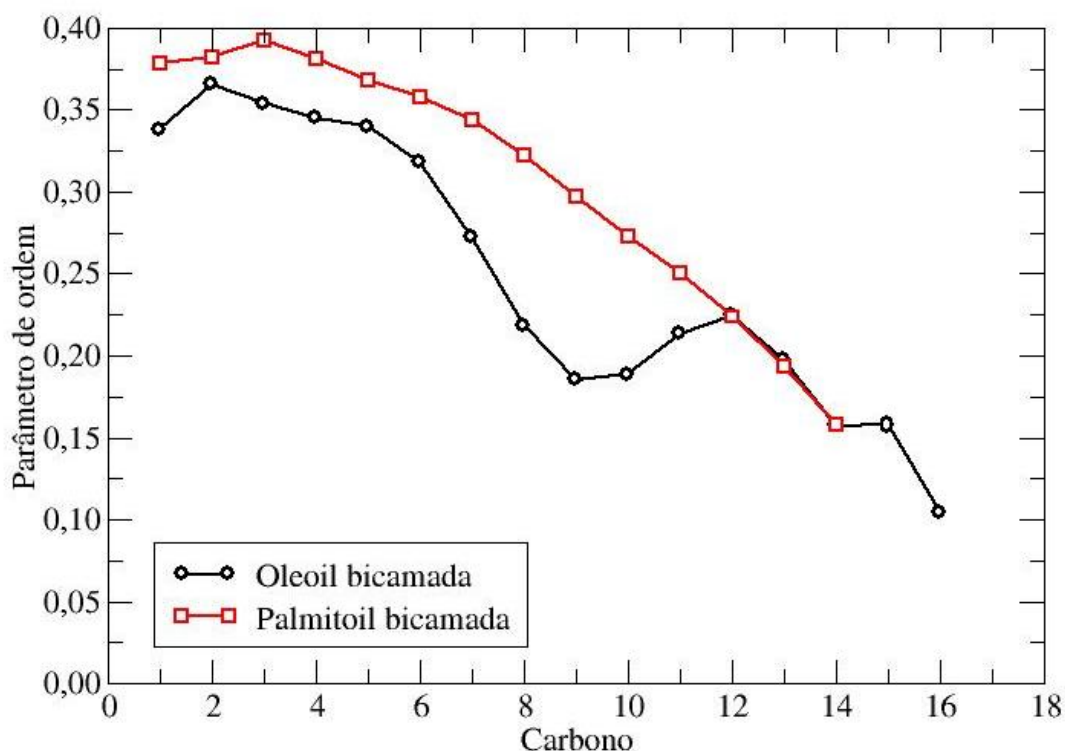
**Figura 24: Perfil de Densidade de Massa da bicamada de POPC hidratada.** Em vermelho a densidade do sistema proteína-bicamada, em azul a densidade de água e em preto a densidade de moléculas POPC.

O perfil apresentado na **figura 24** mostra que a densidade da membrana está entre aproximadamente 9,8 nm e 13,1 nm, essa distância entre os dois picos do perfil de densidade da molécula de POPC corresponde à espessura da bicamada fosfolipídica. Observa-se também que o perfil de moléculas de água é nulo neste intervalo, o que permite afirmar que não há presença de moléculas de água no interior da bicamada.

Na região das cabeças polares, onde o perfil de densidade da membrana é máximo, começam a aparecer moléculas de água na interface bicamada-água. Já para a região localizada ao centro, devido a grande mobilidade das terminações das cadeias alifáticas dos fosfolípidios, confere a essa região um caráter de menor densidade.

Para verificar o alinhamento das cadeias alifáticas (palmitoil e oleoil) foi utilizado o cálculo do parâmetro de ordem.

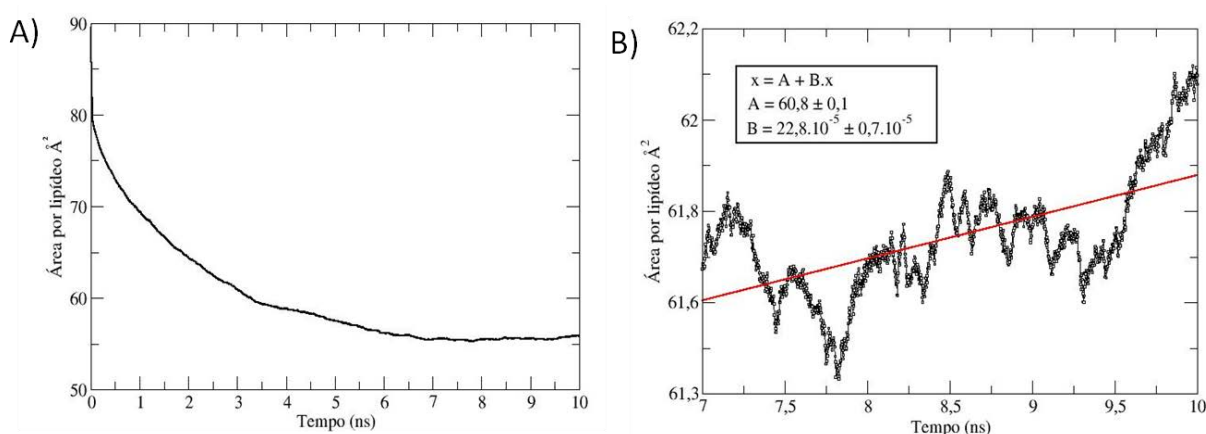
Os resultados dos parâmetros de ordem das cadeias Palmitoil (18 carbonos, saturada) e Oleil (16 carbonos, com insaturação no carbono 9) são mostrados na **figura 25**.



**Figura 25: Parâmetro de ordem das cadeias Palmitoil e Oleil.** Palmitoil em vermelho e Oleil em preto. Após 10 ns de simulação a cadeia saturada (Palmitoil) apresentou maior alinhamento que a cadeia insaturada (Oleil).

De acordo com a definição de P.O., quanto maior o alinhamento dos átomos de carbono na cadeia, maior o parâmetro de ordem. Analisando os dados, pode-se observar um maior alinhamento da cadeia saturada quando comparada a insaturada. Esta diminuição da ordem da cadeia insaturada se deve a presença da insaturação que dificulta o ajuste dessas cadeias. Verificou-se também que os átomos de carbono localizados no início da cadeia apresentam maior parâmetro de ordem do que aqueles das extremidades, este comportamento é devido a grande mobilidade da região terminal. Os valores médios dos parâmetros de ordem de moléculas de POPC calculados:  $0,28 \pm 0,1$  para oleoil e  $0,37 \pm 0,1$  para palmitoil estão de acordo com os valores médios encontrados na literatura,  $0,30 \pm 0,1$  para palmitoil e  $0,37 \pm 0,1$  para oleoil<sup>71</sup>.

A **figura 26-A** mostra o resultados da área calculada por lipídio durante toda a simulação. Observa-se que no período de equilíbrio ocorre uma queda dos valores da área do sistema e a partir do sétimo nano segundo ocorre a convergência deste valor para  $61,70 \pm 0,3 \text{ \AA}^2$ . O valor médio da área por lipídeo encontrado está próximo dos dados encontrados na literatura, onde simulações de membranas composta com o mesmo fosfolipídeo apresentou área por lipídeo média igual a  $63,5 \pm 0,2 \text{ \AA}^2$ <sup>71</sup>. Analisando a **figura 26-B**, verifica-se que o coeficiente linear  $B = 22,8 \cdot 10^{-5} \pm 0,7 \cdot 10^{-5}$  mostrou-se bem próximo de zero indicando uma variação de  $22,8 \cdot 10^{-5} \pm 0,7 \cdot 10^{-5} \text{ \AA}^2$  para cada 1 ns segundo de simulação. Esta análise mostra que o sistema apresenta-se equilibrado.



**Figura 26: Evolução temporal da área por fosfolipídeo para o sistema bicamada POPC: A)** Área por lipídeo por tempo de simulação. Após uma queda do valor da área por lipídeo, nota-se uma convergência após 7 ns de simulação para  $61,70 \pm 0,3 \text{ \AA}^2$ . **B)** Ajuste linear dos valores de área por lipídeo após o sétimo nano segundo. O valor baixo do coeficiente linear ( $B = 22,8 \cdot 10^{-5} \pm 0,7 \cdot 10^{-5}$ ) indica a estabilidade da área por lipídeo da bicamada constituída de POPC após 7 ns de simulação.

Os resultados da **tabela 1** mostram os dados obtidos da simulação da bicamada fosfolipídica constituída de moléculas de POPC.

**Tabela 1: Dados da simulação do sistema bicamada POPC.** (POPC): número de moléculas de lipídeos POPC; (SOL): número de moléculas de água; (x,y,z): dimensões x, y e z final da caixa de simulação em nm; Tempo da simulação de restrição em ns; Tempo da simulação para aquisição de dados em ns; (Ap): Área por lipídeo em Å<sup>2</sup> ; (P.O. palmitoil): parâmetro de ordem as cadeia palmitoil; (P.O. oleoil): Parâmetro de ordem da cadeia oleoil.

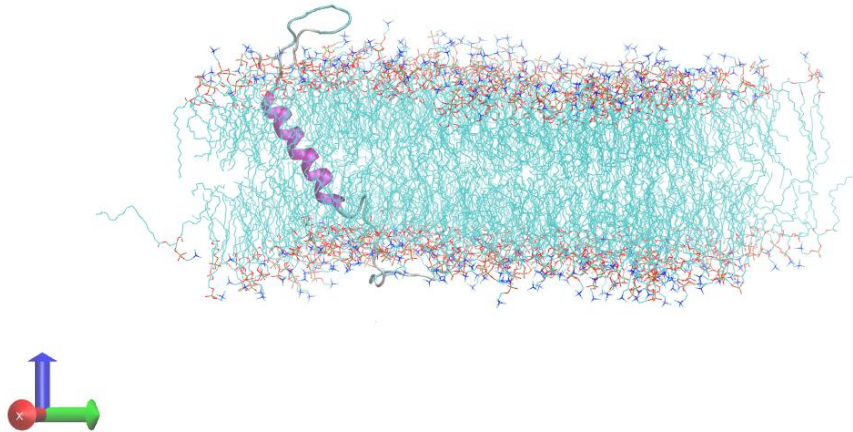
Dados da Simulação do Sistema Bicamada POPC		Dados da literatura
POPC	450	-
SOL	88704	-
x, y, z (nm)	(11.8, 11.8, 23.5)	-
Tempo da simulação de restrição (ns)	5	-
Tempo da simulação para aquisição de dados (ns)	10	-
Ap (Å <sup>2</sup> )	61,7±0,3	63,5±0,2 <sup>71</sup>
P.O. palmitoil	0,37±0,1	0.37±0.1 <sup>65</sup>
P. O. oleoil	0,28±0,1	0.30±0.1 <sup>71</sup>

#### Análises da simulação da proteína SH na bicamada de POPC

O sistema proteína-bicamada foi simulado durante 144 ns para análise estrutural da proteína SH inserida em uma bicamada fosfolipídica, sendo 24 ns em um sistema com dimensões x, y e z iniciais iguais a 11.8 nm, 11.8 nm, 16 nm e 120 ns em uma caixa com dimensões x, y, e z iniciais iguais a 11.8 nm, 11.8 nm, 12 nm.

A **figura 27** mostra a estrutura mais provável da proteína SH nos últimos 70 ns de simulação do sistema proteína-bicamada gerada pela ferramenta do GROMACS<sup>51</sup>, o g-cluster. Observa-se que o domínio TM da proteína foi conservado durante a simulação, enquanto as regiões N-terminal e C-terminal foram direcionadas para perto das cabeças hidrofílicas dos fosfolipídios. Essa aproximação das regiões terminais do modelo pode ser observada também no gráfico do raio de giro do sistema proteína-bicamada em função do tempo de simulação (**figura 28**).

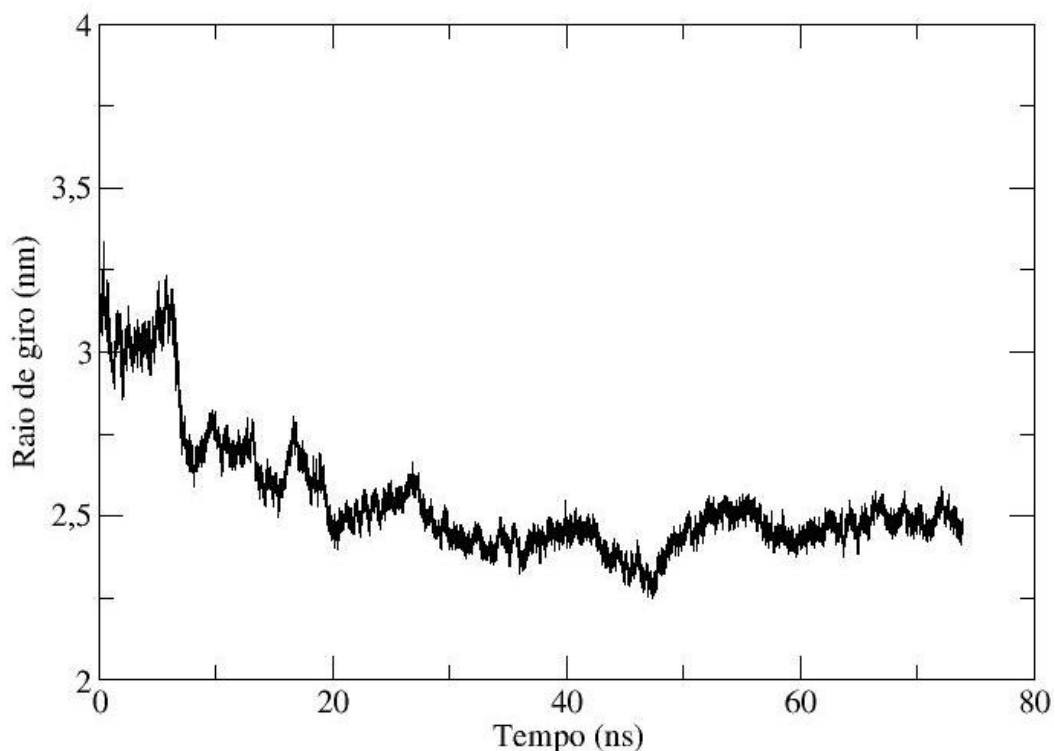




**Figura 27: Estrutura mais provável do sistema proteína-bicamada.** Em relação as estruturas secundárias, observa-se a conservação da hélice do domínio TM da proteína SH em roxo e regiões terminais desestruturadas. Verifica-se também uma inclinação da proteína em relação ao eixo normal da bicamada.

Os resultados da análise da estrutura da proteína SH (**figura 27**) mostra que as regiões terminais ainda não estão estruturadas. Estes resultados sugerem que a estabilidade das estruturas dessas regiões seja devido às interações entre seus oligômeros quando ela se apresenta na forma pentamérica. Isso foi verificado com as simulações do pentâmero inserida na bicamada fosfolipídica (poro-bicamada) que estão discutidas na próxima seção.

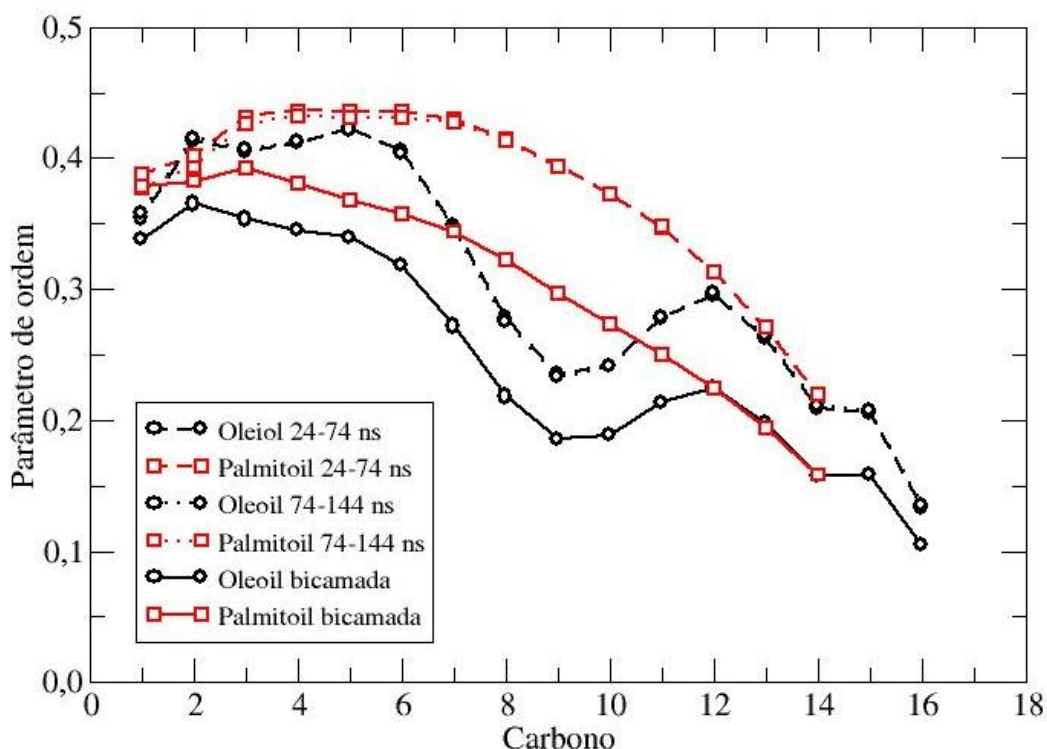
Na **figura 28**, verifica-se também a queda do valor do raio de giro por tempo simulado do sistema proteína SH – bicamada durante os vinte primeiros nano segundos de simulação, após este tempo não há variação considerável na sua compactação. Este comportamento indica uma preferência da regiões terminais por ambientes mais hidrofóbicos confirmando sua característica de proteína de membrana.



**Figura 28: Raio de giro do sistema proteína-bicamada em função do tempo à 310 K.** Nos primeiros nano segundos de simulação ocorre uma queda do valor do raio de giro, após 20 ns esse valor se mantém com pequenas variações até o fim da simulação.

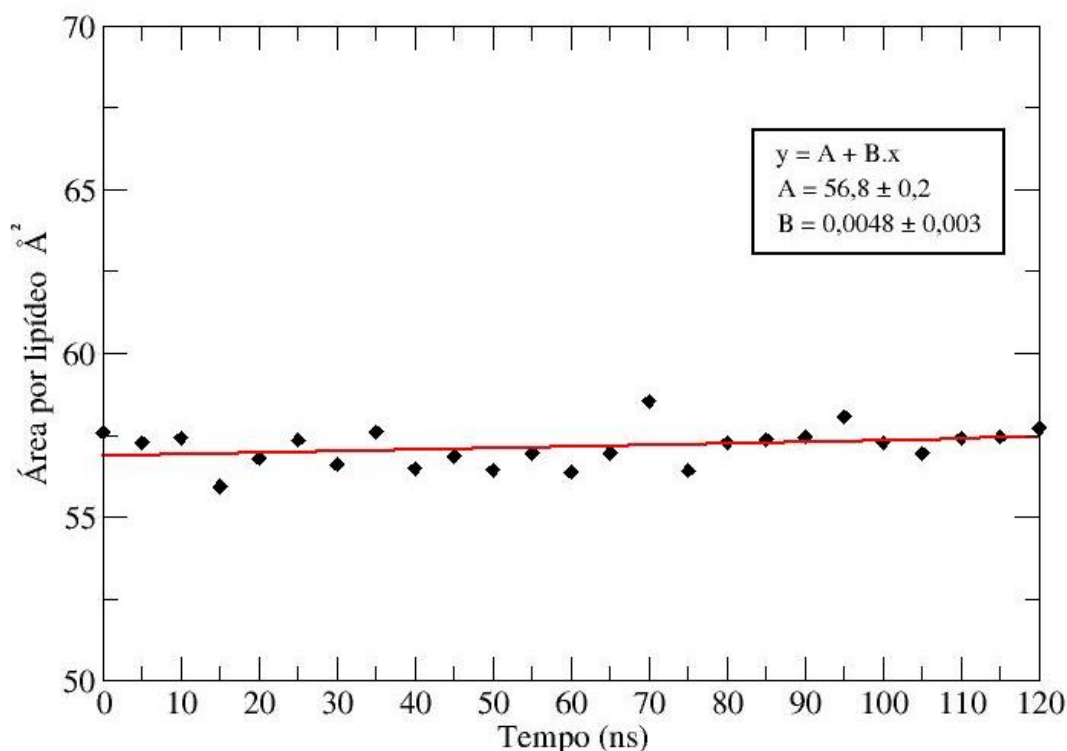
Para análise do parâmetro de ordem da bicamada e área por lipídeo, foi utilizado os últimos 120 ns de simulação do sistema.

A **figura 29** mostra o parâmetro de ordem das cadeias dos fosfolipídeos. Neste resultado observa-se que os parâmetros de ordem da bicamada na presença da proteína nos dois últimos intervalos de simulação (24-74 ns e 74-144 ns) são maiores quando comparados ao parâmetro de ordem da bicamada (**tabela 1**). Verifica-se também que os parâmetros de ordem das cadeias dos fosfolipídios nos dois últimos intervalos de simulação foram semelhantes, indicando uma estabilidade do sistema e sugerindo que a presença da proteína SH sozinha na bicamada não desestabiliza a estrutura da membrana.



**Figura 29: Parâmetro de ordem das cadeias Palmitoil e Oleoil da bicamada POPC e do sistema proteína-bicamada à 310 K.** Nota-se um aumento do PO para o sistema proteína-bicamada comparado ao PO da bicamada sozinha porém, para os dois últimos intervalos, os valores mostram-se similares indicando uma estabilidade do sistema.

Outra característica que foi analisada do sistema bicamada na presença da proteína SH foi a variação da área por lipídeo por tempo simulado. O resultado da **figura 30** mostra que, durante os últimos 120 ns de simulação, o valor da área por lipídeo oscilou em torno de um valor médio de  $57,3 \pm 0,6 \text{ \AA}^2$  (**tabela 2**) este valor se mostra menor que o valor médio obtido somente para a bicamada,  $61,7 \pm 0,3 \text{ \AA}^2$  (**tabela 1**). Analisando a regressão linear da área por lipídeo por tempo simulado, observa-se um valor baixo do coeficiente linear B ( $4 \cdot 10^{-3} \pm 3 \cdot 10^{-3} \text{ \AA}^2 \text{ ns}^{-1}$ ) indicando uma estabilidade da bicamada durante esses últimos 120 ns de simulação. Os resultados mostram que com a inserção da proteína na bicamada de POPC a área por lipídeo teve um decréscimo considerável, conferindo às cadeias menor mobilidade e um maior alinhamento, o que explica o aumento do parâmetro de ordem para esse sistema.



**Figura 30: Área por lipídeo do sistema proteína SH – bicamada para os últimos 120 ns de simulação.** O ajuste linear indica uma estabilidade do sistema para os últimos 120 ns de simulação, resultando em um valor médio da área por lipídeo de  $57,3 \pm 0,6 \text{ \AA}^2$ .

Os resultados da **tabela 2** mostram os dados obtidos da simulação do sistema proteína SH inserido na bicamada fosfolipídica constituída de moléculas de POPC.

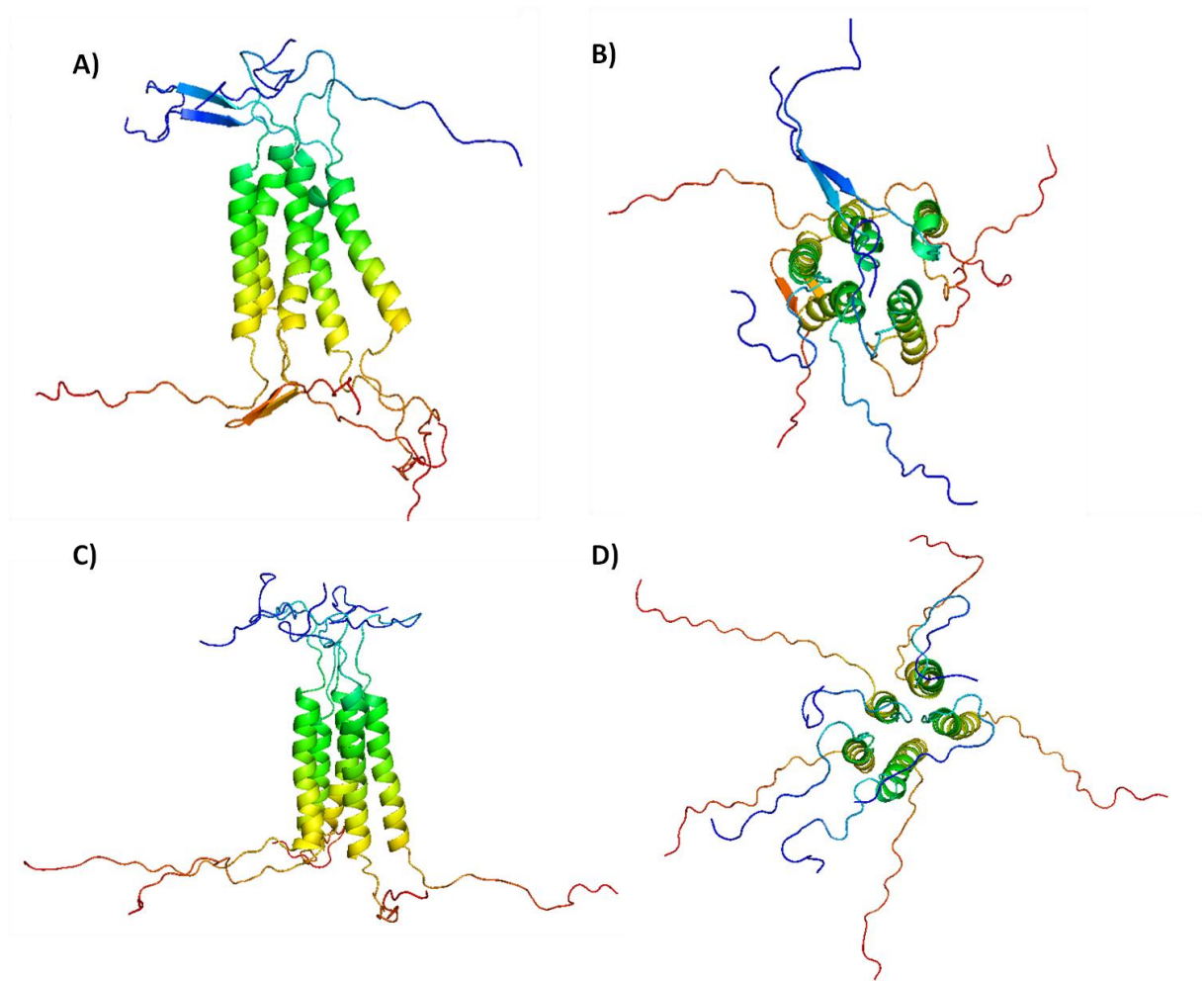
**Tabela 2: Dados da simulação do sistema proteína SH - bicamada.** Tempo (ns): intervalos de tempo da simulação para aquisição de dados em ns; (POPC): número de moléculas de lipídeos POPC; (SOL): número de moléculas de água; (Proteína): número de proteína; (CL-): número de íons cloreto (x,y,z): dimensões x, y e z final da caixa de simulação em nm; (Ap): Área por lipídeo em  $\text{Å}^2$ ; (P.O. palmitoil): parâmetro de ordem das cadeia palmitoil; (P.O. oleoil): parâmetro de ordem da cadeia oleoil.

<b>Dados da Simulação do Sistema proteína SH - bicamada</b>			
Tempo (ns)	0 – 24	24 – 74	74-144
POPC	387	387	387
SOL	48026	29849	29849
Proteína	1	1	1
CL-	1	1	1
x, y, z (nm)	(10.8, 10.8, 16.6)	(10.8, 10.8, 12.0)	(10.8, 10.8, 12.0)
Ap ( $\text{Å}^2$ )	-	$56,7 \pm 0,5$	$57,3 \pm 0,6$
P.O. palmitoil	-	$0,37 \pm 0,1$	$0,37 \pm 0,1$
P. O. oleoil	-	$0,30 \pm 0,1$	$0,30 \pm 0,1$

### **Análises das simulações do Pentâmero inserido na bicamada de POPC**

Para estudo da estrutura do pentâmero, dois modelos foram criados conforme consta na seção Detalhes Computacionais. Para melhor identificarmos os dois modelos chamaremos o sistema com o poro construído com o modelo da proteína SH gerado pelo I-TASSER<sup>44</sup> de sistema 1, e o sistema com o poro construído com o modelo da proteína SH gerado pelo VMD<sup>66</sup> de sistema 2.

Utilizando a ferramenta do GROMACS, o g-cluster, gerou-se a estrutura mais provável do poro construído com o modelo da proteína SH gerado pelo I-TASSER<sup>44</sup> (**figura 31- A e B**) e do poro construído com o modelo da proteína SH gerado pelo VMD<sup>66</sup> (**figura 31- C e B**). Para o primeiro sistema, foram gerados 21 grupos, sendo o mais populoso aquele que continha 3384 estruturas e o segundo mais populoso 1884. Para o segundo sistema, 126 grupos, sendo o mais populoso aquele que continha 289 estruturas e o segundo mais populoso 88. A **figura 31** mostra as estruturas representativas dos grupos com maior número de elementos dos dois sistemas.



**Figura 31: Estrutura mais provável da estrutura pentamérica construída com o modelo da proteína SH gerado pelo I-TASSER<sup>44</sup> e VMD<sup>66</sup>. A) Visualização de lateral da estrutura pentamérica (I-TASSER). B) Visualização de cima da estrutura pentamérica(I-TASSER). C) Visualização lateral da estrutura pentamérica (VMD). D) Visualização de cima da estrutura pentamérica (VMD). O espectro de cores varia do azul para o vermelho sendo azul a região N-terminal e vermelho a região C-terminal.**

Analisando a **figura 31**, verifica-se que as estruturas secundárias das regiões transmembrânicas do pentâmero permaneceram estruturadas nos dois sistemas simulados, porém houve uma maior reorientação dos monômeros do sistema 1 quando comparamos à sua estrutura inicial (**figura 14**). Já para o sistema 2 esta reorientação dos monômeros foi menor, uma vez que esta etapa já tinha sido realizado durante a construção da estrutura pentamérica. Pode ser observado também nas regiões terminais o aparecimento de estruturas secundárias no modelo do sistema 1, sugerindo que as interações entre os monômeros são essenciais para a estruturação C e N-terminais do pentâmero. Sendo assim, analisou-se a formação de pontes de hidrogênio entre os monômeros e verificou-se a

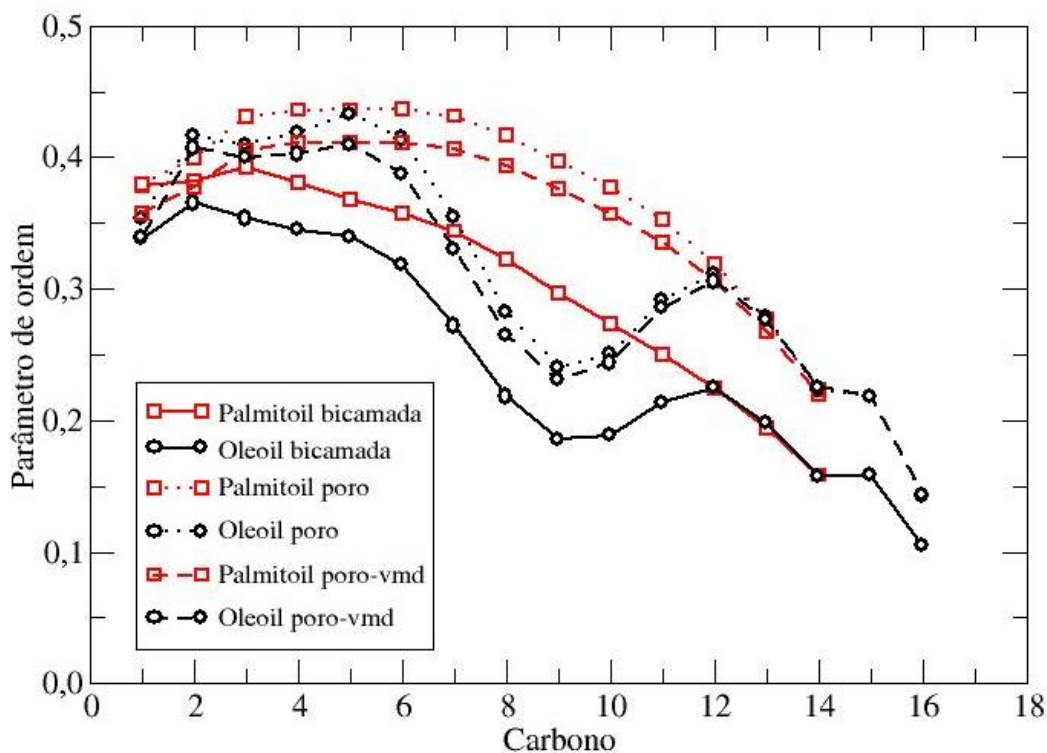
presença destas, principalmente, nas regiões terminais dos oligômeros. A **tabela 3** apresenta os resultados obtidos da análise de ligações de hidrogênio entre os resíduos de monômeros vizinhos (1-2, 1-5, 2-3, 3-4, 4-5,) e também interações entre os monômeros não vizinhos (1-3, 1-4, 2-4, 2-5, 3-5), todas as interações de pontes de hidrogênio identificadas estão localizadas em resíduos das regiões terminais de cada monômero.

**Tabela 3: Dados da análise de pontes de hidrogênio entre os monômeros dos poros da proteína SH.** Sistema 1 representa o sistema da bicamada que contém o poro construído com o modelo da proteína SH gerado pelo I-TASSER<sup>44</sup> e o Sistema 2 representa o sistema da bicamada que contém o poro construído com o modelo da proteína SH gerado pelo VMD<sup>66</sup>. Na tabela estão listados os monômeros que apresentam interações de hidrogênio, o número de pontes de hidrogênio e quais resíduos participam dessas interações.

Monômeros	SISTEMA 1		SISTEMA 2	
	Números de pontes de hidrogênio	Resíduos	Números de pontes de hidrogênio	Resíduos
1 e 2	5	15TRP – 10PHE 13LYS – 8ILE 11SER – 6ILE 10PHE – 13LYS 8ILE – 11SER	-	-
1 e 3	-	-	-	-
1 e 4	-	-	1	2GLU – 13LYS
1 e 5	2	47TYR – 46GLU 46GLU – 46GLU	-	-
2 e 3	2	61ARG – 52ASN 59ARG – 51HIS	2	19THR-17TYR
2 e 4	-	-	-	-
2 e 5	-	-	2	14PHE – 5SER 22HIS – 29SER
3 e 4	-	-	1	2GLU – 13LYS
3 e 5	1	1MET – 14PHE	-	-
4 e 5	-	-	5	19THR – 19THR 19THR – 17TYR 17TYR – 15TRP 19THR – 19THR 17TYR – 17TYR

Os resultados obtidos do gráfico do parâmetro de ordem da bicamada na presença tanto do poro1 como do poro2 (**figura 32**) teve o mesmo comportamento do gráfico do sistema proteína SH – bicamada (**figura 29**): a presença das cinco estruturas na bicamada aumentou o alinhamento das cadeias hidrofóbicas dos fosfolipídios. Para as

cadeias saturadas o valor obtido para o sistema 1 foi de  $0,40 \pm 0,1$  e para as cadeias insaturadas  $0,32 \pm 0,1$  e para o sistema 2 foi de  $0,38 \pm 0,1$  e para as cadeias insaturadas  $0,30 \pm 0,1$ , valores estes maiores aos calculados pelo sistema proteína SH – bicamada (tabela 2).



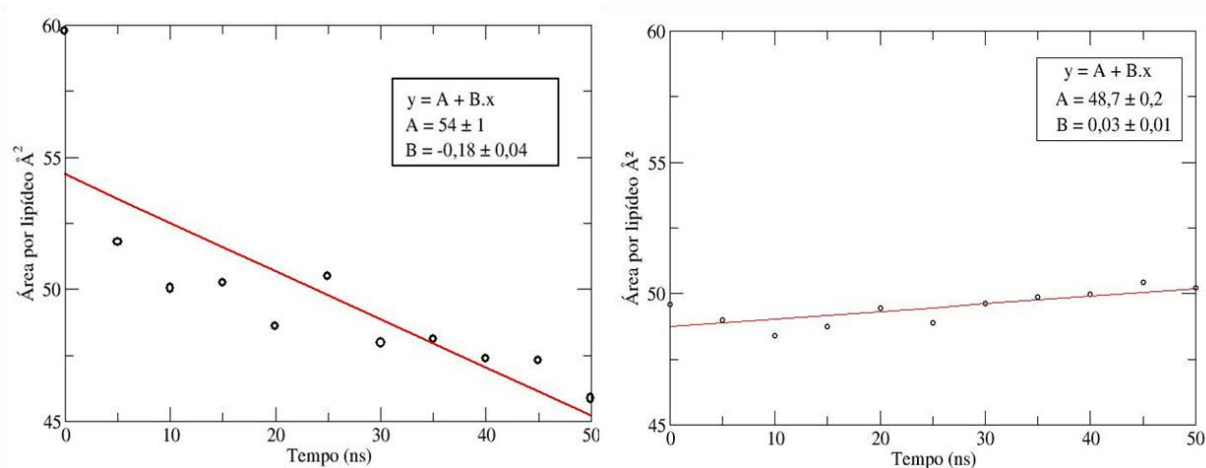
**Figura 32: Parâmetro de ordem das cadeias Palmitoil e Oleoil da bicamada POPC e dos sistemas 1 e 2.** A linha contínua representa o sistema da bicamada; a linha pontilhada o sistema bicamada-poro 1 e a tracejada bicamada-poro 2 (VMD). Os pontos circulares e as linhas em preto representam os valores do PO das cadeias Oleoil e os pontos quadrados e as linhas em vermelho as cadeias Palmitoil.

Na análise da área por lipídeo por tempo simulado (**figura 33**), observa-se que na presença das cinco estruturas (pentâmero), a área por lipídeo torna-se menor. A área por lipídeo média para os últimos 50 ns de simulação do sistema 1 foi de  $48 \pm 3 \text{ \AA}^2$  e para o sistema 2  $48,7 \pm 0,2 \text{ \AA}^2$ , valores estes menores que a do sistema com uma proteína SH inserida na membrana ( $57,3 \pm 0,6 \text{ \AA}^2$ ).

Na análise da flutuação dos valores da área por lipídeo, o ajuste linear do sistema bicamada-poro1 mostra que o coeficiente linear  $B = -0,18 \pm 0,04 \text{ \AA}^2 \text{ ns}^{-1}$  indica que ainda não ocorreu uma estabilidade na área por lipídeo como no sistema proteína SH – bicamada, o



mesmo comportamento foi observado no sistema 2 (VMD), no entanto o valor do coeficiente linear foi menor comparado ao sistema 1 ( $B = 0,0 \pm 0,01 \text{ \AA}^2 \text{ ns}^{-1}$ ). Este resultado sugeri uma maior estabilidade do sistema 2 uma vez que há uma orientação mais adequada dos monômeros em relação ao sistema 1.



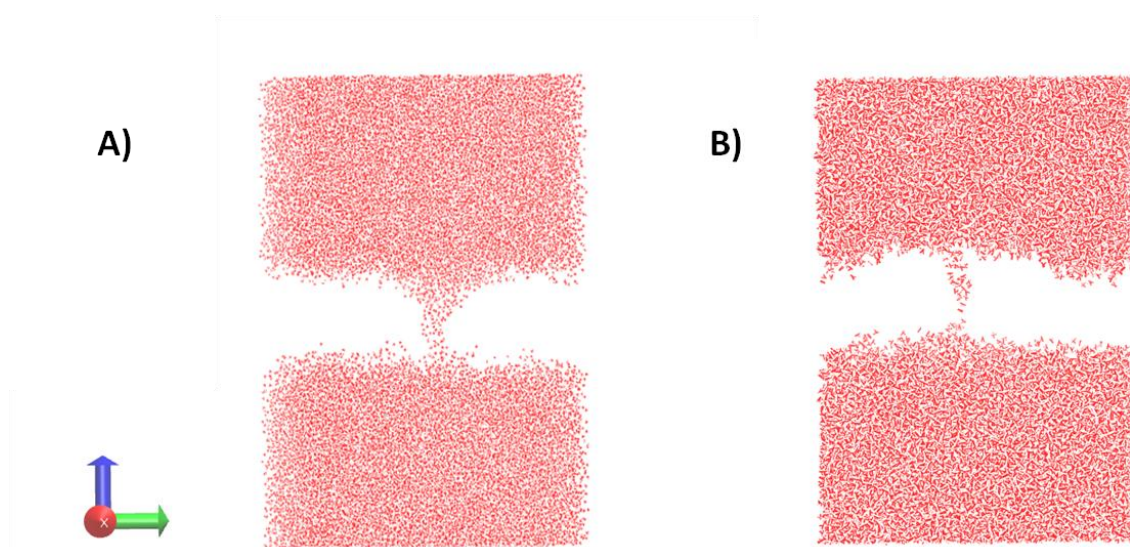
**Figura 33: Área por lipídeo dos sistemas bicamada-poro e bicamada-poro (VMD).** O ajuste linear foi realizado para análise da flutuação dos valores da área por lipídeo por tempo de simulação. O coeficiente linear para o sistema 1 foi de  $B = -0,18 \pm 0,04$ , indicando que o sistema ainda apresenta uma pequena flutuação nos valores da área por lipídeo. O coeficiente linear para o sistema 2 foi de  $B = 0,03 \pm 0,01$ , indicando uma maior estabilidade do sistema.

Os resultados da área por lipídeo e do parâmetro de ordem mostram que assim como na inserção da proteína SH, a inserção do pentâmero na bicamada de POPC diminui a área por lipídeo, conferindo as cadeias menor mobilidade e um maior alinhamento, conforme pode ser observado no aumento do parâmetro de ordem para esses sistemas. Este comportamento de proporcionar uma maior estabilidade da estrutura da membrana pode ser interpretado como uma estratégia do vírus para prolongar seu ciclo de replicação, uma vez que estudos mostram que a presença da proteína SH tem diminuído o efeito citopático e apoptoses em linhagens de células epteliais do pulmão<sup>36</sup>.

Outra característica analisada foi se a atividade do poro poderia ser visualizada com a orientação dos oligômeros. Para verificar a passagem de moléculas de água através do poro foi calculado a barreira de energia livre da penetração da água pela equação abaixo:

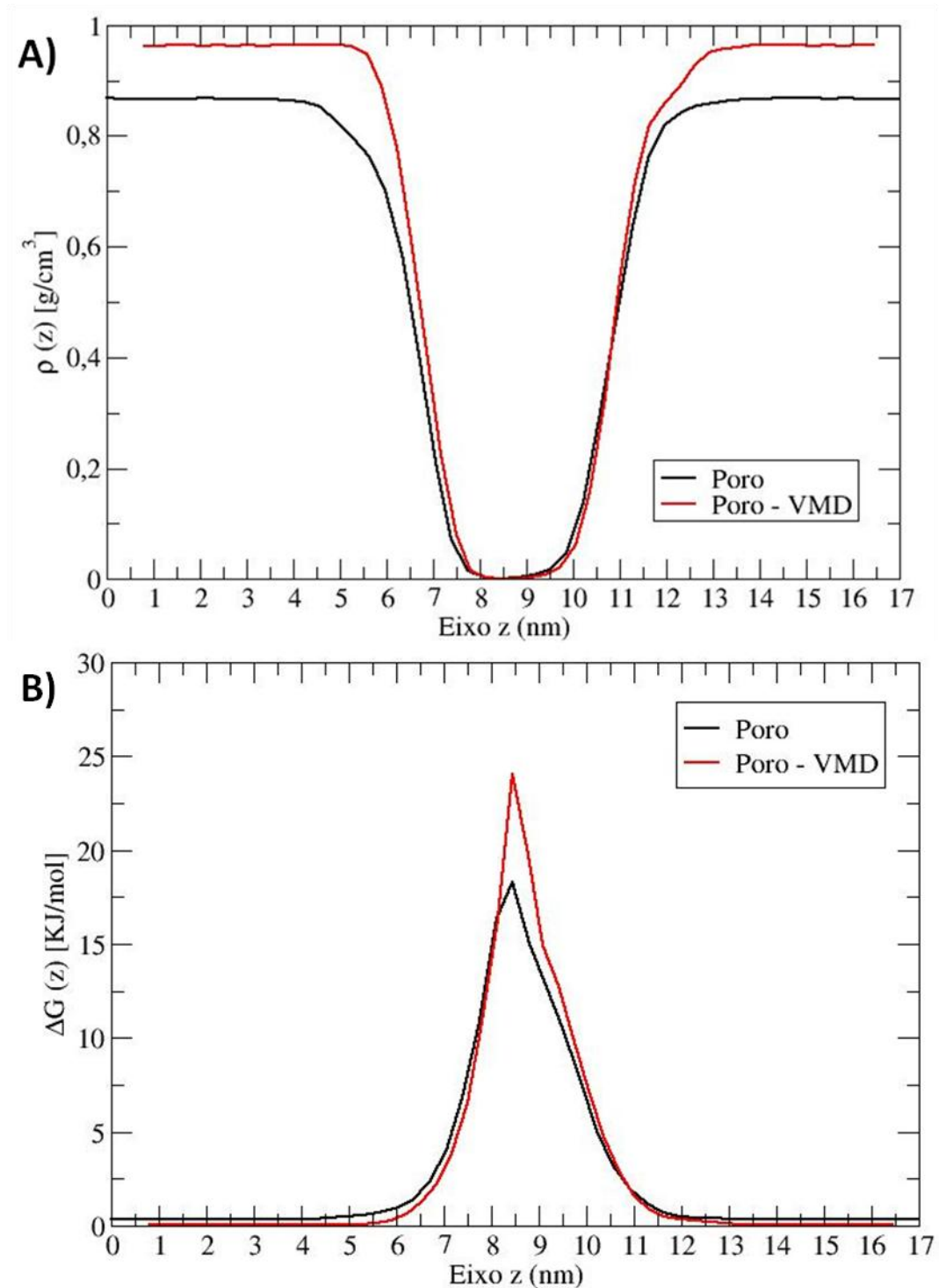
$$\Delta G(z) = -KT \ln \rho_{\text{água}}(z),$$

onde  $\rho_{\text{água}}(z)$  é a densidade de água normalizada ao longo do eixo normal à bicamada obtida da simulação e  $KT$  é a constante de *Boltzmann* vezes a temperatura da simulação (310 K). Como pode ser visto na **figura 34**, durante a simulação, observa-se a passagem de algumas moléculas de água pelo interior do poro dos dois sistemas simulados. Esta atividade não era observada nos outros sistemas (bicamada e bicamada-proteína SH). A passagem de moléculas de águas e íons pelo canal formado pela proteína SH já foi observado em vários experimentos, e corrobora com a característica da proteína SH pertencer ao grupo das vioporinas<sup>72</sup>. Estudos realizados por Gan e seus colaboradores mostram também que a atividade do canal da proteína SH é dependente dos resíduos de histidinas, H22 e H51, que devem estar na forma protonada para que ocorra a passagem de moléculas<sup>38</sup>. Como citado anteriormente, todas as simulações foram realizadas em pH 7.0. Desta forma, nestes sistemas os resíduos de histidinas, H22 e H51, estão na forma protonada, colaborando com os resultados que indicam a dependência do funcionamento do poro da proteína SH com o estado de protonação dos resíduos de histidinas.



**Figura 34:** Imagem obtida das simulações dos sistemas contendo a estrutura pentamérica da proteína SH. **A)** Sistema 1 contendo a estrutura da pentâmero da proteína SH gerado pelo I-TASSER. **B)** Sistema 2 contendo a estrutura do pentâmero da proteína SH gerado pelo VMD. Podemos observar a passagem de moléculas de água nos dois sistemas, confirmando a atividade da estrutura pentamérica.

Durante a simulação, observou-se a passagens de moléculas de água conforme mostrado na **figura 34**. Analisando os resultados da **figura 35**, observa-se que a barreira de energia para os dois sistemas estão bem próximos e é da ordem de 18,2 KJ/mol e 24 KJ/mol para os sistemas 1 e 2, respectivamente. A barreira apresentada pelos sistemas apresentou-se menor comparado com resultados de simulações de um sistema contendo 12 hexapeptídeos antimicrobianos em bicamada de DPPC (dipalmitoilfosfadidilcolina), onde a barreira apresentada foi de  $28,7 \pm 1,7$  KJ/mol<sup>73</sup>. Os resultados com os peptídeos antimicrobianos mostraram que a sua presença afeta a barreira de água permitido a passagem dessas moléculas<sup>73</sup>. Com base neste resultado, verifica-se que a presença da estrutura pentamérica, assim como o peptídeo antimicrobiano, favorece a passagem de água através da membrana.



**Figura 35: Perfil da densidade de água e energia livre da barreira de água ao longo do eixo normal da membrana para os dois sistemas simulados durante 120 ns à 310K. A)** Em preto densidade de água do sistema com a estrutura do pentâmero construído pelo modelo da proteína SH gerado pelo I-TASSER, e em vermelho densidade de água do sistema com a estrutura do pentâmero construído pelo modelo da proteína SH gerado pelo VMD. **B)** Em preto a energia livre da barreira de água do sistema com a estrutura do pentâmero construído pelo modelo da proteína SH gerado pelo I-TASSER, e em vermelho a energia livre da barreira de água de água do sistema com a estrutura do pentâmero construído pelo modelo da proteína SH gerado pelo VMD.

Na **tabela 4** estão listados os dados obtidos da simulação do sistema bicamada-poro construído com o modelo da proteína SH gerado pelo I-TASSER (sistema 1) e os dados da simulação do sistema bicamada-poro construído com o modelo da proteína SH gerado pelo VMD (sistema 2).

**Tabela 4: Dados dos últimos 50 ns de simulação dos sistemas 1 (I-TASSER) e 2 (VMD).** Tempo (ns): intervalos de tempo da simulação para aquisição de dados em ns; (POPC): número de moléculas de lipídeos POPC; (SOL): número de moléculas de água; (Proteína): número de proteína; (CL-): número de íons cloreto (x,y,z): dimensões x, y e z final da caixa de simulação em nm; (Ap): Área por lipídeo em Å<sup>2</sup> ; (P.O. palmitoil): parâmetro de ordem as cadeia palmitoil; (P.O. oleoil): Parâmetro de ordem da cadeia oleoil.

Tempo (ns)	70-120	70 – 120
POPC	341	341
SOL	45864	39811
Proteína	5	5
CL-	5	5
x, y, z (nm)	(10.4, 10.4, 17)	(10.8, 10.8, 15.6)
Ap (Å <sup>2</sup> )	48±3	48,7±0,2
PO palmitoil	0,40±0,1	0,38±0,1
PO oleoil	0,32±0,1	0,30±0,1

## 8. CONCLUSÃO

A metodologia utilizada apresentou-se eficiente e muito satisfatória no estudo da proteína SH e formação de seus oligômeros. Predições de estruturas secundárias da proteína SH indicou a presença de uma alfa-hélice compreendida entre os aminoácidos 20-42. Esta mesma região apresentou-se como altamente hidrofóbica nos teste de hidrofobicidade indicando ser um domínio transmembrânico. Os mesmo resultados de estruturas secundárias foram identificados nos modelos terciários da proteína SH, e a hidrofobicidade do domínio central foi confirmada nas simulações de dinâmica molecular em que, quando em solução, a proteína SH muda sua estrutura linear para uma conformação globular.

As modificações pós-traducionais foram identificadas nas regiões terminais da proteína, estas modificações estão envolvidas principalmente na sua distribuição celular durante a infecção viral. Sítios de interação localizados nas regiões terminais da proteína indicam que a proteína SH interage com outras proteínas virais, entre elas a F e G, sugerindo estar ligada ao mecanismo de fusão celular e também com proteínas do hospedeiro, uma vez que esta proteína é necessária para a inibição da TNF- $\alpha$  induzida por apoptose.

Os resultados obtidos da simulação da bicamada POPC construída reproduz um ambiente bastante semelhante às membranas celulares, onde a proteína SH desempenha sua função. Durante o tempo de simulação do sistema proteína SH - bicamada ocorreu uma estabilidade do domínio central da proteína SH quando inserida na membrana, no entanto as regiões terminais permaneceram desestruturadas. Nas simulações com a presença da estrutura pentamérica, observa-se o início da estruturação das regiões terminais e a presença de ligações de hidrogênio entre os resíduos terminais da proteína SH. Estes resultados indicam que essas interações entre os monômeros são essenciais para estruturação completa desta proteína.

A presença das cinco estruturas na bicamada resultou num decréscimo considerável da área por lipídeo, conferindo as cadeias menor mobilidade e um maior alinhamento, ou seja, não afetou a estabilidade da membrana, sugerindo que a membrana é necessária para a estruturação e funcionamento do poro. Esta estabilidade da membrana proporcionada pela presença da proteína SH sozinha ou em sua forma pentamérica, além de necessária para a estruturação da forma quaternária da proteína, aponta ser uma estratégia do vírus para prolongar seu ciclo de replicação durante a infecção viral.

A passagem de moléculas de água pelo poro ocorreu nas simulações com a presença da estrutura pentamérica. Esta atividade do poro da proteína SH ocorreu em ambiente onde os resíduos de histidinas H22 e H51 apresentaram-se na forma protonada, indicando a dependência desta atividade com o pH do meio. No entanto, esta atividade pode também estar relacionada com algum tipo de interação com outras proteínas virais ou do hospedeiro ou por meio das sinalizações decorrentes de mudanças pós-traducionais.

Com base nestas análises, foi possível propor uma estrutura terciária e quaternária da proteína SH e, com as análises da formação pentamérica da proteína, analisar sua influência no ambiente composto pela bicamada e apresentar uma melhor compreensão da sua função durante a infectividade viral.

## 9. REFERÊNCIAS BIBLIOGRÁFICAS

1. Thomazelli, L. M. *et al.* Surveillance of eight respiratory viruses in clinical samples of pediatric patients in Southeast Brazil. *Jornal de Pediatria* **83**, 422–428 (2007).
2. Hinman, A. R. Global progress in infectious disease control. *Vaccine* **16**, 1116–1121 (1998).
3. Manoha, C., Espinosa, S., Aho, S.-L., Huet, F. & Pothier, P. Epidemiological and clinical features of hMPV, RSV and RVs infections in young children. *J. Clin. Virol.* **38**, 221–226 (2007).
4. Pitrez, P. M. C. *et al.* Rhinovirus and acute bronchiolitis in young infants. *Jornal de Pediatria* **81**, 417–420 (2005).
5. Welliver, T. P. *et al.* Severe human lower respiratory tract illness caused by respiratory syncytial virus and influenza virus is characterized by the absence of pulmonary cytotoxic lymphocyte responses. *J. Infect. Dis.* **195**, 1126–1136 (2007).
6. Sullender, W. M. Respiratory syncytial virus genetic and antigenic diversity. *Clin. Microbiol. Rev.* **13**, 1–15, table of contents (2000).
7. Walsh, E. E., Schlesinger, J. J. & Brandriss, M. W. Protection from respiratory syncytial virus infection in cotton rats by passive transfer of monoclonal antibodies. *Infect Immun* **43**, 756–758 (1984).
8. Parveen, S. *et al.* Genetic variability in the G protein gene of group A and B respiratory syncytial viruses from India. *J. Clin. Microbiol.* **44**, 3055–3064 (2006).
9. Knipe, D. M. & Howley, P. M. *Fields' Virology*. (Lippincott Williams & Wilkins, 2007).
10. Ogra, P. L. Respiratory syncytial virus: the virus, the disease and the immune response. *Paediatr Respir Rev* **5 Suppl A**, S119–126 (2004).
11. Hacking, D. & Hull, J. Respiratory syncytial virus--viral biology and the host response. *J. Infect.* **45**, 18–24 (2002).
12. Ghildyal, R., Mills, J., Murray, M., Vardaxis, N. & Meanger, J. Respiratory syncytial virus matrix protein associates with nucleocapsids in infected cells. *J. Gen. Virol.* **83**, 753–757 (2002).
13. Dupuy, L. C., Dobson, S., Bitko, V. & Barik, S. Casein kinase 2-mediated phosphorylation of respiratory syncytial virus phosphoprotein P is essential for the transcription elongation activity of the viral polymerase; phosphorylation by casein kinase 1 occurs mainly at Ser(215) and is without effect. *J. Virol.* **73**, 8384–8392 (1999).
14. Cuesta, I., Geng, X., Asenjo, A. & Villanueva, N. Structural phosphoprotein M2-1 of the human respiratory syncytial virus is an RNA binding protein. *J. Virol.* **74**, 9858–9867 (2000).
15. Cartee, T. L. & Wertz, G. W. Respiratory Syncytial Virus M2-1 Protein Requires Phosphorylation for Efficient Function and Binds Viral RNA during Infection. *J Virol* **75**, 12188–12197 (2001).
16. Mason, S. W. *et al.* Interaction between Human Respiratory Syncytial Virus (RSV) M2-1 and P Proteins Is Required for Reconstitution of M2-1-Dependent RSV Minigenome Activity. *J Virol* **77**, 10670–10676 (2003).
17. Collins, P. L., Hill, M. G., Cristina, J. & Grosfeld, H. Transcription elongation factor of respiratory syncytial virus, a nonsegmented negative-strand RNA virus. *Proc Natl Acad Sci USA* **93**, 81–85 (1996).
18. Jin, H., Cheng, X., Zhou, H. Z., Li, S. & Seddiqui, A. Respiratory syncytial virus that lacks open reading frame 2 of the M2 gene (M2-2) has altered growth characteristics and is attenuated in rodents. *J. Virol.* **74**, 74–82 (2000).
19. Evans, J. E., Cane, P. A. & Pringle, C. R. Expression and characterisation of the NS1 and NS2 proteins of respiratory syncytial virus. *Virus Res.* **43**, 155–161 (1996).



20. Collins, P. L. & Melero, J. A. Progress in understanding and controlling respiratory syncytial virus: still crazy after all these years. *Virus Res.* **162**, 80–99 (2011).
21. L González-Reyes, M. B. R.-A. Cleavage of the human respiratory syncytial virus fusion protein at two distinct sites is required for activation of membrane fusion. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 9859–64 (2001).
22. García, O. *et al.* Evolutionary pattern of human respiratory syncytial virus (subgroup A): cocirculating lineages and correlation of genetic and antigenic changes in the G glycoprotein. *J. Virol.* **68**, 5448–5459 (1994).
23. Johnson, P. R., Spriggs, M. K., Olmsted, R. A. & Collins, P. L. The G glycoprotein of human respiratory syncytial viruses of subgroups A and B: extensive sequence divergence between antigenically related proteins. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 5625–5629 (1987).
24. Cane, P. A., Matthews, D. A. & Pringle, C. R. Identification of variable domains of the attachment (G) protein of subgroup A respiratory syncytial viruses. *J. Gen. Virol.* **72** (Pt 9), 2091–2096 (1991).
25. Bukreyev, A. *et al.* The secreted form of respiratory syncytial virus G glycoprotein helps the virus evade antibody-mediated restriction of replication by acting as an antigen decoy and through effects on Fc receptor-bearing leukocytes. *J. Virol.* **82**, 12191–12204 (2008).
26. Murphy, B. R. *et al.* Effect of age and preexisting antibody on serum antibody response of infants and children to the F and G glycoproteins during respiratory syncytial virus infection. *J Clin Microbiol* **24**, 894–898 (1986).
27. Collins, P. L. & Mottet, G. Membrane orientation and oligomerization of the small hydrophobic protein of human respiratory syncytial virus. *J. Gen. Virol.* **74** (Pt 7), 1445–1450 (1993).
28. Collins, P. L., Olmsted, R. A. & Johnson, P. R. The small hydrophobic protein of human respiratory syncytial virus: comparison between antigenic subgroups A and B. *J. Gen. Virol.* **71** (Pt 7), 1571–1576 (1990).
29. Gan, S. W., Ng, L., Lin, X., Gong, X. & Torres, J. Structure and ion channel activity of the human respiratory syncytial virus (hRSV) small hydrophobic protein transmembrane domain. *Protein Sci* **17**, 813–820 (2008).
30. Torres, J. *et al.* Model of a putative pore: the pentameric alpha-helical bundle of SARS coronavirus E protein in lipid bilayers. *Biophys. J.* **91**, 938–947 (2006).
31. Rixon, H. W. M. *et al.* The small hydrophobic (SH) protein accumulates within lipid-raft structures of the Golgi complex during respiratory syncytial virus infection. *J. Gen. Virol.* **85**, 1153–1165 (2004).
32. Olmsted, R. A. & Collins, P. L. The 1A protein of respiratory syncytial virus is an integral membrane protein present as multiple, structurally distinct species. *J. Virol.* **63**, 2019–2029 (1989).
33. Techaarpornkul, S., Barretto, N. & Peeples, M. E. Functional Analysis of Recombinant Respiratory Syncytial Virus Deletion Mutants Lacking the Small Hydrophobic and/or Attachment Glycoprotein Gene. *J. Virol.* **75**, 6825–6834 (2001).
34. He, B., Lin, G. Y., Durbin, J. E., Durbin, R. K. & Lamb, R. A. The SH Integral Membrane Protein of the Paramyxovirus Simian Virus 5 Is Required To Block Apoptosis in MDBK Cells. *J. Virol.* **75**, 4068–4079 (2001).
35. Wilson, R. L. *et al.* Function of Small Hydrophobic Proteins of Paramyxovirus. *J. Virol.* **80**, 1700–1709 (2006).
36. Fuentes, S., Tran, K. C., Luthra, P., Teng, M. N. & He, B. Function of the respiratory syncytial virus small hydrophobic protein. *J. Virol.* **81**, 8361–8366 (2007).
37. Gan, S. W., Ng, L., Lin, X., Gong, X. & Torres, J. Structure and ion channel activity of the human respiratory syncytial virus (hRSV) small hydrophobic protein transmembrane domain. *Protein Sci.* **17**, 813–820 (2008).

38. Gan, S.-W. *et al.* The small hydrophobic protein of the human respiratory syncytial virus forms pentameric ion channels. *J. Biol. Chem.* **287**, 24671–24689 (2012).
39. Benson, D. A. *et al.* GenBank. *Nucleic Acids Res.* **41**, D36–42 (2013).
40. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
41. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).
42. Rost, B., Yachdav, G. & Liu, J. The PredictProtein server. *Nucleic Acids Research* **32**, W321–W326 (2004).
43. Buchan, D. W. A. *et al.* Protein annotation and modelling servers at University College London. *Nucleic Acids Res* **38**, W563–W568 (2010).
44. Zhang, Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* **9**, 40 (2008).
45. Abola, E. E., Bernstein, F. C. & Koetzle, T. F. *Protein Data Bank*. (Brookhaven National Lab., Upton, NY (USA), 1984). at <[http://www.osti.gov/energycitations/product.biblio.jsp?osti\\_id=6020246](http://www.osti.gov/energycitations/product.biblio.jsp?osti_id=6020246)>
46. Floudas, C. a. Computational methods in protein structure prediction. *Biotechnology and Bioengineering* **97**, 207–213 (2007).
47. Steinbach, P. J. & Brooks, B. R. Protein hydration elucidated by molecular dynamics simulation. *PNAS* **90**, 9135–9139 (1993).
48. Karplus, M. & McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nature Structural & Molecular Biology* **9**, 646–652 (2002).
49. Oostenbrink, C., Villa, A., Mark, A. E. & Van Gunsteren, W. F. A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6. *Journal of Computational Chemistry* **25**, 1656–1676 (2004).
50. Berger, O., Edholm, O. & Jähnig, F. Molecular dynamics simulations of a fluid bilayer of dipalmitoylphosphatidylcholine at full hydration, constant pressure, and constant temperature. *Biophysical Journal* **72**, 2002–2013 (1997).
51. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **4**, 435–447 (2008).
52. Verlet, L. Computer ‘Experiments’ on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Physical Review* **159**, 98–103 (1967).
53. Beeman, D. Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics* **20**, 130–139 (1976).
54. Pande, V. S., Grosberg, A. Y. & Tanaka, T. On the theory of folding kinetics for short proteins. *Folding and Design* **2**, 109–114 (1997).
55. Whitford, P. C. *et al.* An all-atom structure-based potential for proteins: Bridging minimal models with all-atom empirical forcefields. *Proteins: Structure, Function, and Bioinformatics* **75**, 430–441 (2009).
56. Taketomi, H., Ueda, Y. & Gō, N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int. J. Pept. Protein Res.* **7**, 445–459 (1975).
57. Sobolev, V., Wade, R. C., Vriend, G. & Edelman, M. Molecular docking using surface complementarity. *Proteins* **25**, 120–129 (1996).
58. Noel, J. K., Whitford, P. C., Sanbonmatsu, K. Y. & Onuchic, J. N. SMOG@ctbp: simplified deployment of structure-based models in GROMACS. *Nucleic Acids Res.* **38**, W657–661 (2010).
59. Allen, W. J., Lemkul, J. A. & Bevan, D. R. GridMAT-MD: a grid-based membrane analysis tool for use with molecular dynamics. *J Comput Chem* **30**, 1952–1958 (2009).

60. Róg, T., Murzyn, K. & Pasenkiewicz-Gierula, M. Molecular dynamics simulations of charged and neutral lipid bilayers: treatment of electrostatic interactions. *Acta Biochimica Polonica* **50**, (2003).
61. Berendsen, H., Postma, J., van Gunsteren, W. & Hermans, J. Interaction models for water in relation to protein hydration. *Intermolecular Forces* 331–342 (1981).
62. Berendsen, H. J. C., Postma, J. P. M., van Gunsteren, W. F., DiNola, A. & Haak, J. R. Molecular dynamics with coupling to an external bath. *The Journal of Chemical Physics* **81**, 3684–3690 (1984).
63. Hess, B., Bekker, H., Berendsen, H. J. C. & Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *Journal of Computational Chemistry* **18**, 1463–1472 (1997).
64. Essmann, U. *et al.* A smooth particle mesh Ewald method. *The Journal of Chemical Physics* **103**, 8577–8593 (1995).
65. Clementi, C., Nymeyer, H. & Onuchic, J. N. Topological and energetic factors: what determines the structural details of the transition state ensemble and ‘en-route’ intermediates for protein folding? an investigation for small globular proteins. *Journal of Molecular Biology* **298**, 937–953 (2000).
66. Humphrey, W., Dalke, A. & Schulten, K. VMD: visual molecular dynamics. *J Mol Graph* **14**, 33–38, 27–28 (1996).
67. Feldman, S. A., Crim, R. L., Audet, S. A. & Beeler, J. A. Human respiratory syncytial virus surface glycoproteins F, G and SH form an oligomeric complex. *Arch. Virol.* **146**, 2369–2383 (2001).
68. Low, K.-W., Tan, T., Ng, K., Tan, B.-H. & Sugrue, R. J. The RSV F and G glycoproteins interact to form a complex on the surface of infected cells. *Biochem. Biophys. Res. Commun.* **366**, 308–313 (2008).
69. Lin, Y., Bright, A. C., Rothermel, T. A. & He, B. Induction of apoptosis by paramyxovirus simian virus 5 lacking a small hydrophobic gene. *J. Virol.* **77**, 3371–3383 (2003).
70. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133–7155 (1990).
71. Murzyn, K., Róg, T. & Pasenkiewicz-Gierula, M. Phosphatidylethanolamine-Phosphatidylglycerol Bilayer as a Model of the Inner Bacterial Membrane. *Biophysical Journal* **88**, 1091–1103 (2005).
72. Gonzalez, M. E. & Carrasco, L. Viroporins. *FEBS Lett.* **552**, 28–34 (2003).
73. Appelt, C., Eisenmenger, F., Kuhne, R., Schmieder, P. & Soderhall, J. A. Interaction of the Antimicrobial Peptide Cyclo(RRWRF) with Membranes by Molecular Dynamics Simulations. *Biophys J* **89**, 2296–2306 (2005).

## Anexo I

### Trabalhos realizados

Durante o período de mestrado, além do trabalho de pesquisa desenvolvido já descrito, outros três trabalhos foram desenvolvidos dos quais geraram os artigos: “*Interaction model*

*between HRSV G-protein and flavonoids*”, *“Molecular Model of Interaction between HRSV Fusion Protein and Inhibitors of Viral Infection”* e *“Biophysical and structural characterization of the recombinant human eIF3L”*. Os manuscritos *“Interaction model between HRSV G-protein and flavonoids”* e *“Molecular Model of Interaction between HRSV Fusion Protein and Inhibitors of Viral Infection”* foram desenvolvidos juntamente com Haroldo Lima Pimentel Cravo e Mariana Pela Sabbag sob orientação da Profª Dra. Fátima Pereira de Souza. O artigo *“Biophysical and structural characterization of the recombinant human eIF3L”*, foi realizado em colaboração com a Dra. Ana Thereza Moraes sob orientação do Prof. Dr. Mauricio L Nogueira. Os três trabalhos abordaram aplicações das ferramentas de bioinformática e modelagem molecular no estudo funcional e estrutural de proteínas. Até a presente data, os trabalhos já foram submetidos, sendo que *“Biophysical and structural characterization of the recombinant human eIF3L”* e *“Interaction model between HRSV G-protein and flavonoids”* foram aceitos, e o terceiro aguardando resposta.