

MONOGRAFIA

MAPEAMENTO ANTIGÊNICO DA PROTEÍNA NS1 DE *FLAVIVIRUS*:
UMA ABORDAGEM *IN-SILICO*

Ana Clara Gomes de Souza

Ouro preto, MG

2021

Ana Clara Gomes de Souza

Mapeamento da proteína NS1 de *Flavivirus*: uma abordagem *in-silico*

TCC apresentado a banca avaliadora na Universidade Federal de Ouro Preto como requisito parcial a obtenção do título de Graduação em Ciências Biológicas.

Área de concentração: Microbiologia.

Orientador: Prof. Breno de Mello Silva

Co-orientador: Ricardo Lemes Gonçalves

Ouro Preto, MG

2021

SISBIN - SISTEMA DE BIBLIOTECAS E INFORMAÇÃO

S729m Souza, Ana Clara Gomes de.
Mapeamento da Proteína NS1 de flavivírus [manuscrito]: uma
abordagem in-sílico. / Ana Clara Gomes de Souza. - 2021.
56 f.

Orientador: Prof. Dr. Breno de Mello Silva.
Coorientador: Dr. Ricardo Lemes Gonçalves.
Monografia (Bacharelado). Universidade Federal de Ouro Preto.
Instituto de Ciências Exatas e Biológicas. Graduação em Ciências
Biológicas .

1. Flavivírus. 2. Proteínas. 3. Bioinformática. I. Silva, Breno de Mello.
II. Gonçalves, Ricardo Lemes. III. Universidade Federal de Ouro Preto. IV.
Título.

CDU 579

Bibliotecário(a) Responsável: Celina Brasil Luiz - CRB6-1589



FOLHA DE APROVAÇÃO

Ana Clara Gomes de Souza

Mapeamento antigênico da proteína NS1 de *Flavivirus*: uma abordagem *in silico*

Monografia apresentada ao Curso de Ciências Biológicas da Universidade Federal de Ouro Preto como requisito parcial para obtenção do título de Bacharel em Ciências Biológicas

Aprovada em 27 de agosto de 2021

Membros da banca

Dr. Breno de Mello Silva - Orientador - Universidade Federal de Ouro Preto
Msc. Ricardo Lemes Gonçalves - Co-orientador - Universidade Federal de Ouro Preto
Dr. Rodrigo Araújo Lima Rodrigues - Universidade Federal de Minas Gerais
Dr. - Leandro Márcio Moreira - Universidade Federal de Ouro Preto

Breno de Mello Silva, orientador do trabalho, aprovou a versão final e autorizou seu depósito na Biblioteca Digital de Trabalhos de Conclusão de Curso da UFOP em 23/08/2023



Documento assinado eletronicamente por **Breno de Mello Silva, PROFESSOR DE MAGISTERIO SUPERIOR**, em 23/08/2023, às 22:12, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site http://sei.ufop.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0579897** e o código CRC **E37DE55B**.

AGRADECIMENTOS

Aos meus pais Carlos e Jacqueline, por tanto amor e dedicação quanto a minha
vida,
Ao meu avô José Emílio, pelo auxílio na formação do meu bem mais precioso:
conhecimento,
As minhas avós Maria da Consolação e Maria Sérgia, por todo o amor e pelas
orações,
Aos meus tios(as) paternos, e maternos Giovanni e Jennifer pelos conselhos e
ombro amigo,
Ao meu namorado Yann por todo o companheirismo durante a minha
trajetória,
Aos meus amigos Karol, Alzira, Ubiratan, Raquel, João, Laisy e Diego, pelos
momentos de felicidade proporcionados,
Ao meu orientador Breno e ao meu co-orientador Ricardo por todo o
conhecimento trocado, pela compreensão e paciência.

“Construí amigos, enfrentei derrotas, venci obstáculos, bati na porta da vida e
disse-lhe:

Não tenho medo de vive-la”

(Augusto Cury)

SUMÁRIO

1. Introdução	10
1.1 Arbovírus no Brasil	10
1.2 <i>Flavivirus</i>	10
1.3 A replicação dos <i>Flavivirus</i>	12
1.4 A proteína não estrutural 1(NS1)	14
1.5 Sistema imune e epítomos	15
1.6 Bioinformática	16
1.7 Caracterização e importância do trabalho	17
2. Objetivos	18
2.1 Geral	18
2.2 Específicos	18
3. Materiais e métodos	19
4. Resultados e discussão	26
4.1 GenBank e alinhamento das sequências inter-específicas	26
I. Regiões não conservadas	27
II. Regiões conservadas	29
4.2 Variações intra-específicas	31
I. Domínio β-rool	34
II. Domínio Wing	35
III. Domínio β-ladder	36
4.3 Antigênicidade da NS1	37
4.4 Predição de peptídeos imunogênicos e epítomos lineares para células B	38
I. Bepipred	38
II. IMED	39
III. NTCcell	40
4.5 Modelagem estrutural	41
4.6 Epítomos conformacionais para células B	44
I. Ellipro	44
4.7 Variações de superfície da NS1	46
4.8 Modelagem de epítomos	47
4.9 Conclusão e perspectivas futuras	50
5. Referências	52

RESUMO

Atualmente, o número de doenças provocadas por *Flavivirus* é significativo. No ano de 2020 foram notificados 931.903 casos possíveis de infecções por *Dengue virus* (DENV) e 6.805 casos possíveis de infecções por *Zika virus* (ZIKV). Apesar do alto impacto causado por estes casos, pouco se sabe sobre as prováveis variações de proteínas entre as diversas linhagens circulantes no Brasil, o que pode ser um empecilho na criação de novas vacinas e testes. O alvo do estudo que se segue é a proteína NS1, uma vez que além de ter função na montagem da estrutura viral, está ligada a diversas funções na infecção como romper vasos sanguíneos e causar uma cascata de infecção celular através da ativação de enzimas do organismo presentes no glicocalix, tropismo determinado por células endoteliais e vazamento de vasos e células para o extravasamento do vírus. O trabalho consiste no alinhamento de sequências da NS1 correspondentes ao vírus DENV, ZIKV e YFV, com a finalidade de uma observação de regiões conservadas e polimórficas na proteína, bem como análise das variações intra e inter específicas entre os resíduos e predição de epítomos imunogênicos, o trabalho usou da bioinformática para obter grande parte das conclusões. As variações foram descritas, a proteína foi modelada *IN-SILICO* através de programas computacionais, houve a predição de epítomos para células T e B e também foram feitas imagens de variações na superfície da NS1 para as espécies. Os resultados quanto à variabilidade da proteína foram interessantes, uma vez que houve o registro de diversas regiões polimórficas e não polimórficas em uma curta quantidade de resíduos, a similaridade entre as proteínas foi de 28,6%, enquanto a identidade foi de 50,6%. Houveram alterações na quantidade de mutações e variabilidade entre as espécies de acordo com as regiões da proteína, o que pode ter ocorrido devido a exposição ao meio e a funcionalidade da mesma. A NS1 se mostrou antigênica e teve grande quantidade de epítomos encontrados nos servidores de predição, sendo 96 epítomos para células T e 136 para células B. O trabalho poderá auxiliar no direcionamento da realização de novas vacinas e testes diagnósticos para o mercado.

Palavras chave: *Flavivirus*, NS1, Bioinformática.

ABSTRACT

The number of diseases caused by *Flavivirus* is significant nowadays. Although, for example, in 2020 931,903 possible cases of DENV and 6,805 probable cases of ZIKV were reported, despite the high impact caused by this number of cases, little is known about the likely variations of proteins among the various strains circulating in Brazil, which can be an obstacle in the creation of new vaccines and tests. The target of the study is the NS1 protein that has a role in the assembly of the viral capsule and activates enzymes present in glycocalyx in cells from blood vessels, causing endothelial cells and leakage of vessels and cells for virus extravasation. The work consists of aligning NS1 sequences corresponding to the DENV, ZIKV, and YFV viruses to observe conserved and polymorphic islands in the protein and analyze Intra and interspecific variations between residues, and predicting immunogenic epitopes. The work used bioinformatics to obtain most of the conclusions. Variations were described, the protein was modeled in-silico using computer programs, epitopes were predicted for T and B cells, and NS1 surface variations were also imaged for the species. The results regarding protein variability were interesting since there was a record of several polymorphic and non-polymorphic islands in a short number of residues, the similarity between the proteins was 28.6%. In comparison, the identity was 50.6 %, there were changes in the number of mutations and variability between species according to the protein regions, which may have occurred due to exposure to the medium and its functionality. NS1 proved to be antigenic and had many epitopes found in prediction servers, with 96 epitopes for T cells and 136 for B cells. The work may help guide the realization of new vaccines and diagnostic tests for the market.

Key words: *Flavivirus*, NS1, Bioinformatic.

LISTA DE FIGURAS

FIGURA1- a-Organização genômica dos Flavivirus. b- Organização da poliproteína na membrana.....	12
FIGURA2- a-Ciclo de replicação viral. b- bolha de replicação.....	13
FIGURA3- Formas e domínios da proteína NS1..	14
FIGURA4- Etapa de Threading..	22
FIGURA5- Etapa de montagem estrutural	22
FIGURA6- Etapa de seleção e refinamento.....	23
FIGURA7- Sequências de NS1 coletadas do GenBank..	26
FIGURA8- Alinhamento das proteínas consenso para cada espécie..	27
FIGURA9- a- Modelagem NS1 monômero. b- Modelagem NS1 dímero...	41
FIGURA10- Variação de cargas elétricas na superfície, face externa.....	44
FIGURA11- Variação de resíduos na superfície, face externa.....	45
FIGURA 11- Peptídeos imunogênicos modelados.....	46
FIGURA 2- Modelagem dos peptídeos imunogênicos conservados entre as espécies.....	46

LISTA DE TABELAS

TABELA1- Tabela de variação intra-específica	30
TABELA2- Antigênicidade da NS1.....	35
TABELA 3- Epítapos Imed.....	36
TABELA 4- Epítapos NTCcell.	37
TABELA 6- Epítapos Bepipred.....	38
TABELA 5- Validação das proteínas modeladas.....	41
TABELA 7- Epítapos Ellipro.....	41

LISTA DE ABREVIATURAS

aa.- Aminoácido
ACC- Transformação de auto-covariância cruzada
C- Carbono
C α - Carbono alfa
C β - Carbono beta
DENV- Dengue virus
EJ- Encefalite japonesa vírus
H- Hidrogênio
NS1- Proteína não estrutural 1
NS2A- Proteína não estrutural 2A
NS2B- Proteína não estrutural 2B
NS3- Proteína não estrutural 3
NS4A- Proteína não estrutural 4A
NS4B- Proteína não estrutural 4B
NS5- Proteína não estrutural 5
N- Nitrogênio
O- Oxigênio
PDB- Banco de dados de proteína
PH- Potencial hidrogeniônico
prM- Proteína de membrana
prE- Proteína de envelope
RNA- Ácido Ribonucleico
YFV- *Yellow fever virus*
ZIKV- *Zika virus*

1. Introdução

1.1 Arbovírus no Brasil

Arbovírus são um grupo de vírus que tem artrópodes como vetores, sendo assim, pode-se inferir que esses vírus necessitam de um artrópode para que seu ciclo de infecção seja completo e sua espécie seja perpetuada através da transmissão, que é finalizada quando o mosquito pica o vertebrado(ELISA, 2014). As doenças causadas por arbovírus estão se tornando cada vez mais presentes na sociedade, uma vez que com a urbanização o processo de contaminação vírus/artrópode/vertebrado se torna mais fácil, dentre as famílias de vírus presentes neste grupo temos *Togaviridae*, *Bunyaviridae*, e *Flaviviridae*(ELISA, 2014).

De acordo com o boletim epidemiológico de número 41 emitido pelo Ministério da Saúde em maio de 2020 foram notificados 931.903 casos de Dengue virus 1(DENV1), Dengue virus 2(DENV2), Dengue virus 3(DENV3), Dengue virus 4(DENV4) , sendo 28,2 casos por cem mil habitantes, também foram notificados 6.805 casos de Zika vírus(ZIKV), sendo 3,2 casos por cem mil habitantes(BRASIL. MINISTÉRIO DA SAÚDE, 2020), ainda existem episódios que não contam com testes válidos ou mesmo situações onde o infectado não procura ou recebe atendimento médico, estes são dados confirmam a incidência de arboviroses no Brasil bem como a necessidade de se pesquisar sobre os vírus e as proteínas virais para um possível tratamento que melhore a qualidade de recuperação do doente ou evite a infecção.

O alvo deste trabalho são DENV1, DENV2, DENV3 e DENV4, *Yellow fever virus*(YFV) e ZIKV que são espécies do gênero *Flavivirus* e da família *Flaviviridae*, no gênero *Flavivirus* existem 39 espécies virais consideradas arbovírus de um total de 53 espécies (ELISA, 2014).

1.2 *Flavivirus*

Os *Flavivirus* são compostos por vírions classicamente esféricos com proteínas envelope recobrando sua face externa, o genoma é constituído por RNA simples com 9-12Kb e de fita positiva, após a leitura do RNA é formada uma poliproteína que irá gerar proteínas estruturais e não estruturais, no genoma viral existem duas extremidades não codificantes 3' e 5', apesar dessas regiões não serem

codificadas elas realizam funções primordiais na replicação viral (KASHIWABARA, 2007; SIMMONDS et al., 2017).

A constituição dos *Flavivirus* se dá por três proteínas estruturais responsáveis pela estrutura externa viral e penetração do vírus na célula, sendo elas proteína de capsídeo, envelope e membrana e sete proteínas não estruturais responsáveis pela replicação e infecção do vírus no organismo, sendo elas proteína não estrutural 1(NS1), NS2(A,B), NS3, NS4(A,B) e NS5(MUKHOPADHYAY; KUHN; ROSSMANN, 2005).

A poliproteína formada pela leitura do RNA viral é clivada em diversas porções, essa clivagem pode ocorrer através de enzimas presentes no organismo como helicases e proteases ou através da proteína NS3 (NEUFELDT et al., 2018). Cada proteína não estrutural apresentará uma função específica para que a replicação, bem como a transmissão e infecção efetiva do vírus sejam feitas de forma eficiente, a NS1 é responsável pela organização correta das proteínas na formação de uma nova cepa, a NS2 é dividida em A e B e sua função é juntar a NS3 ao complexo proteico, a NS4 também é dividida em A e B e auxilia na construção da curvatura da membrana, além de se juntar a proteína NS3 com função desconhecida, a NS5 é responsável pela adição de grupos fosfato e metil na extremidade 5' do RNA, todo o genoma, bem como as poli-proteínas se encontram ilustrados na figura 1 a e b (NEUFELDT et al., 2018).

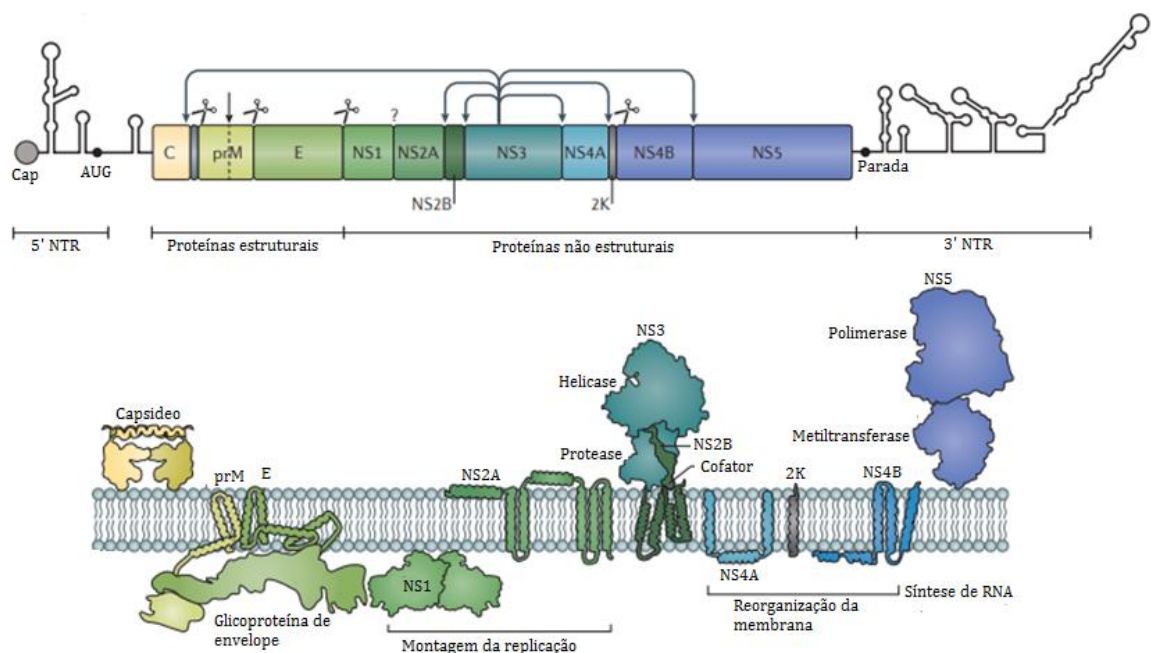


FIGURA1: a- Organização genômica dos Flavivirus. O genoma se inicia com uma região cap tipo 1 na região não codificante 5', seguido de um códon de iniciação e das proteínas estruturais, logo após temos as proteínas não estruturais e um códon de parada na região 3' não codificante. As setas representam clivagens por protease sinal, a seta com pontilhados indica clivagem por protease furina, a interrogação tem a protease de clivagem desconhecida para as poliproteínas de DENV, a região 2K é um peptídeo sinal para a inserção da proteína NS4B na membrana, a proteína NS2B auxilia a protease (NS3). **b- Organização da poliproteína na membrana.** No lúmen do retículo se encontram as proteínas estruturais, a NS1 assim como a proteína NS2 auxiliam na produção de vírions, a NS2B em específico recruta a NS3 na membrana do retículo, tem a principal função enzimática com proteases e helicases, a NS4A auxilia na curvatura da membrana enquanto a NS4B se associa a NS3 com atividade enzimática, a NS5 tem atividade na metilação do genoma viral. Adaptado de:(NEUFELDT et al., 2018).

O gênero *Flavivirus* é composto por muitas espécies, a relação que se tem é a de que linhagens que infectam mosquitos e vertebrados é distante filogeneticamente da linhagem que infecta carrapatos, pesquisas indicam que a ordem de proximidade se inicia em YFV, passa por EJ(*Encefalite japonesa*) e DENV, onde o DENV4 está mais próximo do DENV2 assim como o DENV1 está mais próximo do DENV3, o ZIKV é recente, entretanto já se mostra ramificado no clado com origem Asiática, Africana oriental e ocidental, a cepa que chegou a América tem descendência Asiática(LI et al., 2017; ZANOTRO et al., 1996).

1.3 A replicação dos *Flavivirus*

A replicação do vírus se inicia com a endocitose mediada por proteínas receptoras, como a lectina do tipo C (CD2019), receptor de manose, e membros da família de proteínas TIM e TAM, a endocitose é dependente de vias metabólicas específicas de clatrina, logo após a penetração do vírion na célula o PH ácido do endossomo faz com que as proteínas estruturais se fundam a parede endossomica, permitindo a exposição do RNA viral, processo ilustrado na figura2(ASNIS; CRUPI, 2005; NEUFELDT et al., 2018).

O material viral será replicado no retículo endoplasmático onde as endomembranas formarão organelas de replicação invaginadas repletas de vesículas virais, ilustradas na figura2 a e b, a replicação ocorre nas organelas pois elas aumentam fatores bioquímicos e de proteção que irão favorecer o vírus, além de coordenar as etapas e defender o antígeno do sistema imune, junto das organelas estão as membranas convolutas, com função ainda desconhecida entretanto com suspeita de

participar da maturação de poliproteínas, quem transformará a molécula de RNA em uma poliproteína dentro das organelas é o ribossomo (NEUFELDT et al., 2018).

Existem suspeitas de que a curvatura da vesícula no retículo endoplasmático seja ajustada por proteínas virais NS4(a-b), NS2a e NS1, somado a fatores do hospedeiro (NEUFELDT et al., 2018). Estudos indicam que a proteína de capsídeo é essencial para a montagem do vírion, uma vez que ela é capaz de ordenar clivagens e a organização das proteínas prM e prE através de sua estrutura com peptídeos sinal (TAN et al., [s.d.]).

Os *Flavivirus* movimentam as vias metabólicas do infectado e usam proteínas do mesmo para se promover, as peptidases clivam a poliproteína viral e as chaperonas auxiliam no dobramento dessas proteínas, sendo usadas em algumas partes da replicação, o que existe é uma rede de modulação que permite com que novos vírions sejam produzidos (NEUFELDT et al., 2018).

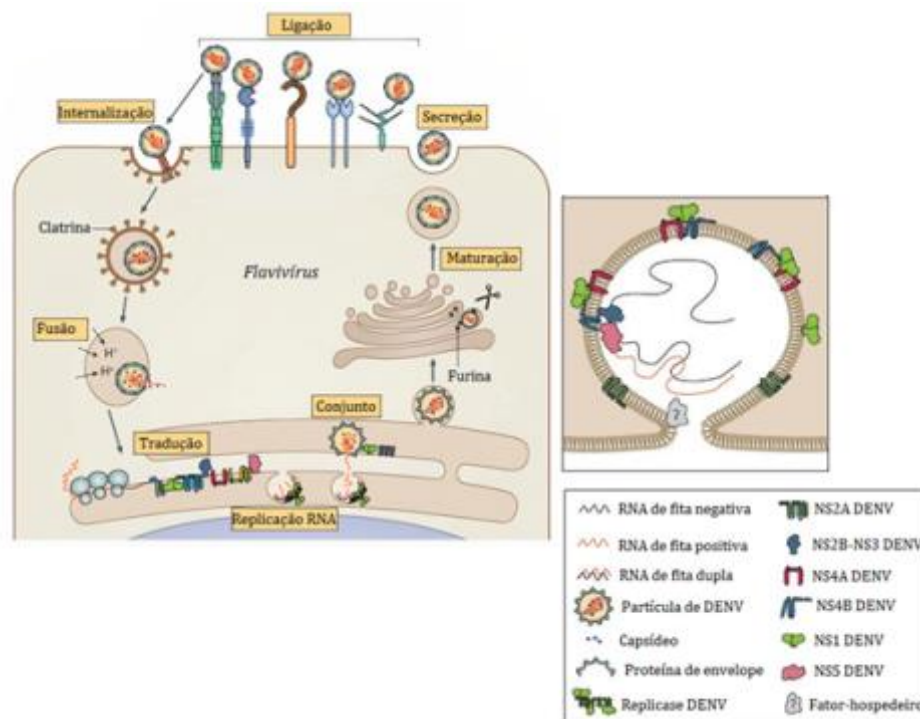


FIGURA2: a- Ciclo de replicação viral. A partícula viral se liga aos receptores celulares e é internalizada, uma série de receptores é ligada ao vírion para que a internalização seja específica, a via de entrada da partícula de Flavivirus é dependente de clatrina, a acidez do endossomo faz com que a partícula se fusione a sua parede liberando o RNA+ que é reconhecido pelo ribossomo, no ribossomo irá ocorrer a tradução do RNA em uma poliproteína bem como a replicação do RNA viral, a maturação da partícula viral ocorre em vias secretoras mediadas por furina no retículo endoplasmático, a última etapa se dá pela secreção do vírion. **b- bolha de replicação.** Local onde novas poliproteínas serão formadas e

seguirão para a formação de vírions, as proteínas estruturais funcionam ordenadamente para que o processo ocorra com sucesso. Adaptado de:(NEUFELDT et al., 2018)

1.4 A proteína não estrutural 1(NS1)

Além de mostrar função na montagem da partícula viral, recentemente foi descoberto através de estudos com NS1 recombinante produzida em mamíferos e análises de resistência elétrica trans-endotelial que a proteína tem capacidade de romper a barreira endotelial de células, podendo alterar a carga viral em certos órgãos a partir da espécie de vírus infectante resultando no tropismo viral, foi dado que a NS1 também tem capacidade de causar vazamento vascular em células cerebrais, hepáticas, dérmicas e pulmonares, processo realizado com estudos de fluorescência, o extravasamento é responsável por disseminar vírions(PUERTA-GUARDO et al., 2019).

O gene da NS1 tem 1.056 nucleotídeos e formará 352 aminoácidos, o seu peso molecular varia de 46-55kDa(GLASNER et al., 2018). A proteína é segmentada em 3 principais domínios que são β -roll^(aa.1-30), β -ladder^(aa.31-180) e Wing^(aa.181-352) demonstrados na figura3, o primeiro domínio é considerado o mais conservado pois é o mais funcional dentre toda a proteína, este domínio apresenta caráter hidrofóbico devido as interações que ele fará com a membrana lipídica, os outros dois domínios são considerados anfipáticos e entram contato direto com o sangue e fluídos corpóreos, conseqüentemente são mais susceptíveis a mutações (NEUFELDT et al., 2018).

A sintetização da proteína ocorre inicialmente formando monômeros que passam por dimerização, transformação que contribuirá para a associação da proteína a membranas celular, depois de dimerizada a proteína pode sofrer oligomerizações e se mostrar de formas variadas, prevalecendo a hexamérica, forma que é lançada na corrente sanguínea devido a sua solubilidade externa(GLASNER et al., 2018).

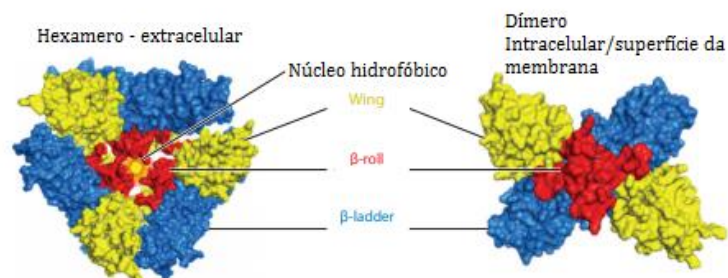


FIGURA3: Formas e domínios da proteína NS1. NS1 demonstrada nas formas hexamérica, que fica em circulação corpórea extracelular e dimérica, que fica associada a membrana. As

subunidades β -roll, Wing e β -ladder estão representadas pelas cores vermelho, amarelo e azul. No centro do hexâmero existem os núcleos hidrofóbicos conhecidos como balsas lipídicas. Adaptado de: (GLASNER et al., 2018).

A NS1 apresenta uma média de 30% de identidade e 70% de similaridade entre os DENV(1-4), YFV e ZIKV(PUERTA-GUARDO et al., 2019) e é considerada uma proteína antigênica já que está em alta circulação na corrente sanguínea durante a fase aguda da doença (HILGENFELD, 2016). Sua similaridade entre as espécies, tal como seu poder antigênico fazem com que essa proteína seja um bom alvo de estudos para vacina e teste diagnóstico.

1.5 Sistema imune e epítomos

Sistema imune é dado por um conjunto de células especializadas em combater antígenos ou moléculas estranhas ao organismo, o mecanismo de proteção acontece de forma ordenada porém complexa, onde diversas vias são acionadas e funcionarão em conjunto com o objetivo final de cessar a infecção causada pelo agente estranho(SILVA; FERNADEZ; TEVA, 2010).

Dentre as células compositoras do sistema imune, temos os leucócitos, considerados células primordiais, responsáveis pelo primeiro estímulo ao organismo, macrófagos, responsáveis pela fagocitose de corpos estranhos, células dendríticas, responsáveis pela apresentação do corpo estranho ao sistema imune, granulócitos, formados pelos neutrófilos, eosinófilos e basófilos atuando na resposta inata e sobre a alergia, mastócitos, atuando na parede de vasos sanguíneos(SILVA; FERNADEZ; TEVA, 2010).

Os linfócitos, células responsáveis pela resposta contra *Flavivirus*, são divididos em B e T e podem se diferenciar em diversas ocasiões com a finalidade de matar células infectadas, ativar outras células do sistema imune e efetivar a secreção de anticorpos, existem também os receptores de célula B e T, que participam da resposta humoral e resposta mediada por células respectivamente(SILVA; FERNADEZ; TEVA, 2010).

Por fim tem-se as células de memória, que permitirão com que o organismo infectado consiga ter uma melhor resposta imune na segunda infecção pelo antígeno junto das células efetoras, que sobrevivem algum tempo pós infecção para que o quadro não retorne(SILVA; FERNADEZ; TEVA, 2010).

Epítopo é considerado um conjunto de aminoácidos de uma proteína responsável por estimular resposta imune (GETZOFF; TAINER; LERNER, 1988), os epítotos podem ser classificados como contínuos quando os aminoácidos são sequenciais e geralmente mais curtos e como descontínuos quando não são sequenciais na proteína, mas se unem por dobramentos expostos da estrutura, geralmente são mais longos (“© 1986 Nature Publishing Group”, 1986).

Estudos indicam que em uma infecção causada por *Dengue virus*, as principais linhas de resposta humoral são resultantes de células T CD4+ e CD8+, ressaltando que a resposta mediada por CD8+ apresenta-se em maior porcentagem “(75%)”, dado que dois tipos de resposta ocorrem predominantemente na infecção por *Flavivirus*, foi descoberto uma segregação de subtipos onde a resposta do tipo T CD4+ se identificou com proteínas não estruturais e CD8+ com proteínas estruturais (RIVINO et al., 2013).

1.6 Bioinformática

A bioinformática é a ciência que une moléculas biológicas a equações computacionais, seu uso foi iniciado na década de 1970 com um artigo de simulação molecular para proteínas, a medida em que a tecnologia dos computadores e máquinas avançou os estudos de moléculas por estes meios também cresceu, atualmente as possibilidades são enormes, fazendo com que grandes questões sejam resolvidas por meio do computador, sejam elas totalmente computacionais ou parte dos experimentos de bancada (Bioinformática : da Biologia à Flexibilidade Molecular, 2014).

O presente trabalho utiliza uma série de ferramentas da bioinformática para uma geração relativamente rápida de resultados com baixos custos, pois os gastos para a pesquisa computacional são bem menores quando comparados à experimentação em bancada. Consideramos a bioinformática como pontapé inicial para próximos estudos que tenham comprovação com experimentos, esperamos que as informações aqui apresentadas fomentem a busca por uma vacina ou teste diagnóstico para vírus negligenciados que afetam grande parte da população.

A modelagem molecular utilizada no trabalho teve como objetivo a construção de uma proteína que pudesse ser utilizada como base para testes em bancada, uma estrutura válida e o mais próximo do que é encontrado na Natureza foi confeccionada.

O refinamento é um ponto importante na dinâmica molecular, uma vez que as estruturas vindas dos servidores de modelagem molecular nem sempre são válidas quanto as interações eletrostáticas e ao arranjo das moléculas, isso ocorre devido aos

campos de força aplicado pelas plataformas, o refinamento permite com que os imprevistos vindos das plataformas possam ser corrigidos e conseqüentemente a estrutura seja otimizada (BAKER, 2015).

O processo de refinamento pode variar de acordo com a metodologia do servidor, podendo ser de otimização de energia, rearranjo de cadeias laterais, comparação com cristais e PDBs de banco de dados, entre outros.

1.7 Caracterização e importância do trabalho

No estudo coletamos todas as cepas de NS1 para os DENV(1-4), ZIKV e YFV reportadas no banco de dados GenBank até o ano de 2018, consecutivamente analisamos todas as sequências obtidas e pontuamos variações e blocos consenso, construímos um modelo de proteína validado a partir da modelagem molecular, analisamos a antigênicidade e fizemos a predição de epítomos para cada um dos vírus tratados através de servidores computacionais.

É relevante ressaltar que a estrutura molecular da NS1 construída obteve bons parâmetros na plataforma de validação, indicando que a estrutura é 100% idêntica ao que é encontrado no banco de dados feito com base na comparação com o cristal da proteína, entretanto nos bancos de dados atuais existem poucas estruturas ótimas de NS1, assim além de ter previsões ideais a partir da estrutura construída será possível contribuir com diversos trabalhos futuros.

O trabalho se faz importante uma vez que a NS1 é uma proteína antigênica, como demonstrado nos resultados, além disso atualmente ainda não existe um teste diagnóstico ou mesmo vacina que identifiquem ou protejam o indivíduo hospedeiro de forma efetiva, já que além da proteína ter altas porcentagens de identidade e similaridade entre as espécies, também mostra grande quantia de variações intra específicas, por isso é necessário que ocorra uma caracterização palpável para que os próximos passos possam ser a criação de métodos concretos.

2. Objetivos

2.1 Geral

Realizar análise antigênica da proteína não estrutural 1(NS1) através da bioinformática, utilizando todas as sequências depositadas na plataforma GenBank e localizadas no Brasil, com a finalidade de expandir conhecimento nos estudos de criação de vacinas pan-*Flavivirus* e testes diagnósticos diferenciais.

2.2 Específicos

- Coleta e análise das sequências de NS1 por meio do alinhamento FASTA.
- Localização de blocos conservados ou variantes inter e intra-específicos.
- Análise antigênica e localização de epítomos e peptídeos imunogênicos.
- Modelagem e refinamento molecular.
- Busca por epítomos candidatos a vacina ou teste diagnóstico.

3. Materiais e métodos

3.1 Coleta das sequências.

As sequências foram coletadas através da plataforma GenBank(BENSON et al., 2010) para as espécies *Dengue vírus 1*, *Dengue vírus 2*, *Dengue vírus 3*, *Dengue vírus 4*, *Yellow Fever vírus* e *Zika vírus* entre os anos de 1994-2018, foram coletadas apenas as linhagens indexadas no Brasil e com a proteína NS1 sequenciada completa.

3.2 Triagem das sequências.

Na triagem as proteínas foram separadas em arquivos FASTA de acordo com a espécie viral, com uma sub-secção por regiões do Brasil sendo Norte, Nordeste, Sul, Sudeste e Centro-Oeste, as linhagens com localização indevida em outros países, junto daquelas que tinham a NS1 incompleta foram descartadas, as linhagens que não estavam com a região do Brasil identificada foram fracionadas com a nomenclatura “região desconhecida”.

3.3 Alinhamento das sequências.

As sequências intra-específicas foram alinhadas com o servidor MEGA-X(KUMAR et al., 2018) utilizando o método de coloração ClustalW(THOMPSON; HIGGINS; GIBSON, 1994).

O funcionamento da ferramenta ClustalW ocorre em três passos distintos que se complementarão, o primeiro passo é a construção de uma matriz responsável por numerar distâncias entre os aminoácidos da sequência submetida, os valores obtidos dentro da matriz são percentuais e descrevem a proximidade entre duas sequências, no segundo passo é construída uma árvore guia onde pares de linhagens da mesma espécie e similares são alinhados na medida em que a árvore filogenética se ramifica, formando uma ordem lógica de grupos mais próximos e mais distantes, os grupos próximos fornecem pesos menores e os grupos distantes pesos maiores, o terceiro passo é a construção de um alinhamento com base na árvore construída no passo anterior, o ClustalW irá dar preferência para o conjunto de resíduos com peso menor, esse conjunto será denominado consenso(KUMAR et al., 2018; THOMPSON; HIGGINS; GIBSON, 1994).

3.4 Descrição dos polimorfismos

A partir do alinhamento intra-específico foi possível observar os pontos de concordância e variância entre as linhagens, bem como descreve-los.

3.5 Criação de sequência consenso otimizada.

Uma sequência consenso otimizada foi selecionada para cada espécie viral recorrida no trabalho através da ferramenta EMBOSS, que utiliza da coloração ClustalW para definir a sequência de resíduos dominantes entre as proteínas submetidas através de uma linha de comando(RICE; LONGDEN; BLEASBY, 2000).

3.6 Alinhamento das proteínas e descrição das variações inter-específicas observadas.

Os alinhamentos foram realizados pelo MEGA-X(KUMAR et al., 2018) através da coloração ClustalW(THOMPSON; HIGGINS; GIBSON, 1994), após a obtenção e alinhamento das sequências consenso otimizadas foi possível observar, bem como descrever as variações entre as espécies virais estudadas.

3.7 Análise de antigênicidade da NS1.

Foi realizada a análise de antigênicidade da proteína para cada espécie trabalhada através da plataforma VaxiJen(DOYTCHINOVA; FLOWER, 2007).

O VaxiJen usa para predição de antigênicidade o método de transformação de auto-covariância cruzada(ACC)(WOLD et al., 1993), que transforma sequências peptídicas em matrizes de dados padrão, as médias de conversão de peptídeos para matriz são baseadas em valores Z, definidos como hidrofobicidade, tamanho e polaridade dos aminoácidos, o conjunto de variáveis fornecido pela combinação de aminoácidos diferentes propõe características únicas para cada sequência, os resultados obtidos nos cálculos de transformação de ACC são comparados com um conjunto de dados localizados no próprio servidor, e formam vetores com base na similaridade e especificidade das sequências, a partir da comparação o grau de antigênicidade do peptídeo é proposto.

3.8 Obtenção de peptídeos imunogênicos

A predição de peptídeos de célula T foi feita pelos servidores IMed(GARCIA-BORONAT et al., 2008), Bepipred (RICE; LONGDEN; BLEASBY, 2000) e NTCCell(VITA et al., 2019).

O método da plataforma IMed consiste na quantificação da variabilidade e diversidade das sequências, seu principal objetivo é determinar variações que impedem a ligação anticorpo-epítipo, além de selecionar aminoácidos conservados que possam gerar uma possível resposta imune celular. A plataforma é manipulada com a entrada de múltiplas sequências de alinhamento (MSA) no formato fasta, o cálculo de variabilidade é realizado com 3 principais índices que irão medir a variabilidade do resíduo e a frequência com que o aminoácido se repete em certa posição(GARCIA-BORONAT et al., 2008).

Além de uma busca na literatura a plataforma NTCCell faz a identificação de epítomos de classe MHCI através de uma análise da imunogenicidade por predição de cada peptídeo, a análise tem como fatores cruciais a posição e o tipo de aminoácido que compõe a sequência, bem como a afinidade de ligação entre certa classe de aminoácidos e células T, dentre os resultados demonstrados no artigo escrito por Calis et al. aminoácidos como Lisina, Metionina e Triptofano mostraram mais afinidade de ligação com o complexo MHCI, assim como aqueles que se encontraram em posições centralizadas no peptídeo, também foi evidenciado que resíduos aromáticos e com cadeias laterais grandes são mais reconhecidos pelo sistema imune, dado comprovado com a análise da porcentagem de aminoácidos mais frequentes em epítomos validados para resposta. A estatística de Kullback-Leibler foi utilizada para medir a correspondência entre os perfis de MHC e peptídeos, a imunogenicidade do conjunto de epítomos ideal foi medida pelo servidor e 66% dos peptídeos selecionados foram imunogênicos(CALIS et al., 2013).

A interface Bepipred2.0 é capaz de identificar epítomos conformacionais de célula B, os peptídeos são definidos a partir do raio de distância que se encontram do anticorpo, calculado pela interface através de cristais, também são feitos cálculos de probabilidade com o algoritmo *Random Forest Regression* que usa o tamanho, carga, volume e acessibilidade relativa de cada aminoácido para a predição de epítomos(RICE; LONGDEN; BLEASBY, 2000).

3.9 Modelagem e refinamento estrutural.

A modelagem estrutural do monômero foi feita pelo servidor I-TASSER (ROY; KUCUKURAL; ZHANG, 2010), enquanto o dímero foi modelado pelo GRAMM-X (TOVCHIGRECHKO; VAKSER, 2006). Foram realizados dois processos complementares de refinamento, pelos servidores GalaxyRefineComplex (KO et al., 2012) e ModRefiner (XU; ZHANG, 2011). A validação das proteínas resultantes do refinamento foi feita pelo servidor Molprobity (CHEN et al., 2010).

A construção de uma proteína em uma das melhores plataformas conhecidas(I-TASSER) consiste na etapa 1 ou de Threading ilustrada na figura4, o objetivo da primeira etapa é identificar proteínas resolvidas e com sequências homólogas a proteína submetida em um banco de dados através de um BLAST demonstrado na primeira imagem da figura, um conjunto de alinhamentos homólogos é criado e a estrutura secundária é prevista, a mesma estrutura é submetida em um banco de PDBs que se encontra dentro da ferramenta e está ilustrado na terceira imagem da figura,

templates são criados a partir do PDB selecionado que melhor se aproxima da estrutura secundária prevista ilustrado na imagem quatro da figura e novamente selecionados através da qualidade dos modelos com base em um cálculo estatístico da melhor energia e alinhamento ilustrado na imagem cinco (ROY; KUCUKURAL; ZHANG, 2010).



FIGURA4: Etapa de Threading. Sequências com estruturas homólogas a proteína alvo foram obtidas, a partir destas foi feita a previsão de estrutura secundária, templates iniciais foram construídos e selecionados.

A etapa 2 ou de Montagem estrutural representada pela figura5 busca modelar as regiões que se alinham bem com as regiões mal alinhadas pelo método ab initio, formando uma estrutura que se completa, para otimizar a etapa as cadeias são reduzidas a átomos Carbono alfa(C α) e centros de massa da cadeia lateral, ilustrado na imagem um da figura5. A modelagem se baseia no método de simulação de Monte Carlo, que faz diversas combinações em diversas temperaturas, a energia também é uma variável. Os princípios da simulação buscam aconselhar os termos fornecidos pelos PDBs, como cadeia lateral e hidrofobicidade, as restrições fornecidas pela sequência/composição de cada resíduo assim como a sua ocupação espacial, a proteína é dobrada a fim de respeitar as ligações e o raio de movimentação dos resíduos, as estruturas construídas em temperaturas menores são selecionadas devido a menor energia livre, centros de proteínas selecionadas são identificados pela sua coordenada ilustrado na imagem dois da figura5 (ROY; KUCUKURAL; ZHANG, 2010).

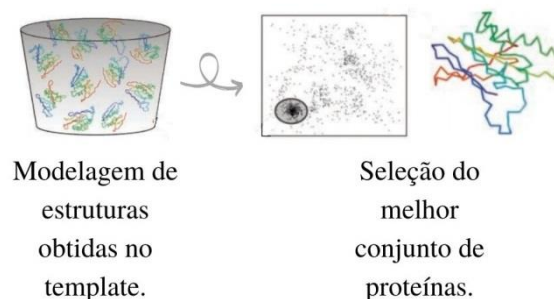


FIGURA5: Etapa de montagem estrutural. Os templates selecionados passaram pela modelagem estrutural e as melhores proteínas representantes da NS1 foram escolhidas.

Na etapa 3 ou etapa de seleção e refinamento dos modelos mostrada na figura 6 é feita uma nova simulação com as proteínas selecionadas, com o objetivo de remover as colisões entre resíduos e refinar as estruturas, as proteínas de menor energia são selecionadas e o modelo geral é gerado com a otimização de ligações de hidrogênio (ROY; KUCUKURAL; ZHANG, 2010).



FIGURA6: Etapa de seleção e refinamento. Após a seleção dos primeiros modelos uma segunda peneira escolheu os melhores, as estruturas obtidas foram refinadas e um modelo final foi selecionado.

A etapa 4 ou observação funcional baseada na estrutura consiste na consulta da proteína obtida comparada com estruturas já conhecidas e presentes no banco de dados da plataforma (ROY; KUCUKURAL; ZHANG, 2010).

O servidor de refinamento GalaxyRefine otimiza as áreas menos precisas da proteína, o processo se inicia a partir da seleção de regiões não estáveis, identificadas por faixas mal conformadas do banco de dados de PDB's usado, a localização é feita através de múltiplos alinhamentos entre a estrutura submetida e o banco de dados, também é feito um cálculo de raiz quadrada média por resíduo, quando três resíduos consecutivos ressaltam valores que extrapolam os limites estipulados a região é considerada como variável (KO et al., 2012).

Um cálculo de energia foi adotado para as regiões variáveis definidas, as faixas instáveis com menor índice de energia foram selecionadas através do MODELLERCSA (CHERAGA, 1999) que construiu modelos homólogos por espaço de anelamento conformacional. Assim a modelagem das não estáveis é feita através de um alinhamento seguido da construção da estrutura tridimensional (KO et al., 2012).

O ModRefine faz o refinamento global da proteína em duas etapas, a primeira etapa é feita após a identificação de traços de carbono alfa onde a plataforma irá adicionar os átomos nas cadeias principais e realizará uma otimização de energia de baixa resolução do conjunto, a segunda etapa consiste na adição dos átomos das

cadeias laterais e minimização da energia em alta resolução do conjunto obtido nas duas etapas (XU; ZHANG, 2011).

Quando a cadeia principal é incompleta a sua integralização se dá por análises e tabelamento dos quatro carbonos alfa consecutivos existentes seguidos de C e N terminais, um template de estruturas obtidas do servidor PISCES (WANG; JR, 2003) é utilizado como base para completar a proteína incompleta, além do cálculo de ângulos internos e de torção e a distância entre os átomos para alocação de novos carbonos alfa, a minimização de energia da cadeia principal tem como objetivo refinar a estrutura fisicamente, para isso os constituintes(N, O, C α , H, C e C β) e as cadeias laterais foram consideradas, vale ressaltar que para as cadeias laterais apenas centros de posição média foram construídos com base em um conjunto de ângulos formado por cada aa., a conformação da proteína é definida quando um conjunto de movimentos realizado pelos átomos mostra menor energia (XU; ZHANG, 2011).

Para a adição da cadeia lateral os rotâmeros são considerados, a escolha se baseia na estatística de acordo com os ângulos de rotação que melhor se adequam a estrutura, respeitando o aa. seguinte, uma nova minimização de energia é feita após adição das cadeias laterais, conferindo flexibilidade a estrutura, a minimização supõe a conformação correta do conjunto cadeia principal e cadeias laterais (XU; ZHANG, 2011).

3.10 Obtenção de peptídeos de células B.

Os peptídeos de células B foram realizados após a modelagem, utilizando nossa estrutura em PDB(protein data bank) através do servidor Ellipro (PONOMARENKO et al., 2008).

O servidor Ellipro é capaz de definir epítomos com três algoritmos base, sendo eles o método de Thornton que se baseia na forma elíptica da proteína submetida para definir a porcentagem de átomos constituintes, seguido da identificação dos átomos mais protuberantes pelos valores de IP responsáveis por indicar os átomos protuberantes na NS1, uma vez que com a maior exposição a acessibilidade do sistema imune é maior, valor do índice de protuberância é dado a partir do cálculo de porcentagem: $\text{total de átomos} / \text{átomos protuberantes} \times 100$ o resultando fornece um conjunto de epítomos válidos, após a obtenção dos resíduos protuberantes é calculada uma distância R entre os centros de massa, a partir dessa distância os resíduos são

agrupados formando um epítopo (PONOMARENKO et al., 2008; THORNTON et al., 1986).

4. Resultados e discussão

4.1 GenBank e alinhamento das sequências inter-específicas

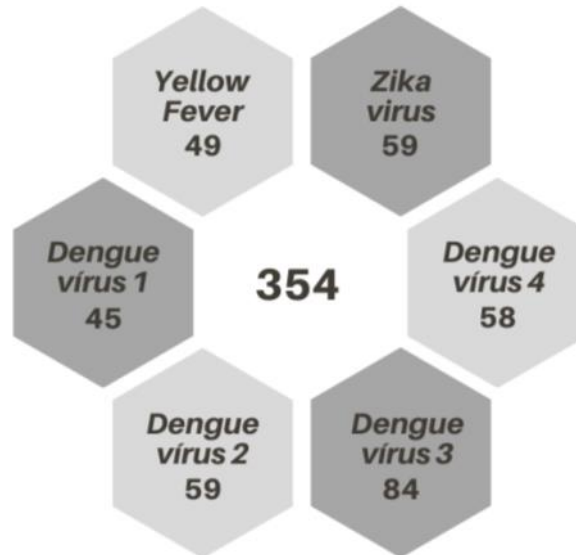


FIGURA7: Sequências de NS1 coletadas do GenBank. Sequências coletadas e validadas correspondentes a cada espécie recorrida no Trabalho, no centro existe a quantidade total de proteínas coletadas.

As sequências da NS1 foram coletadas na plataforma GenBank (BENSON et al., 2010) entre os anos de 2017 e 2018, todas as sequências recolhidas estavam endereçadas no Brasil, resultando em 354 proteínas completas de DENV1, DENV2, DENV3, DENV4, ZIKV e YFV indicadas na figura7, o vírus com mais linhagens foi o DENV3 enquanto o DENV1 foi o vírus com menos linhagens encontradas. Os genomas com a proteína incompleta foram excluídos do trabalho, uma vez que a falha poderia causar lacunas nas análises. As sequências de NS1 coletadas foram segregadas em arquivos fasta de acordo com a espécie correspondente, para cada grupo estudado foi criada uma sequência consenso otimizada da NS1 por meio do programa EMBOSS(RICE; LONGDEN; BLEASBY, 2000).

As proteínas consenso otimizadas de DENV(1-4), ZIKV e YFV ilustradas na figura8 foram alinhadas com a ferramenta de alinhamento ClustalW(THOMPSON; HIGGINS; GIBSON, 1994) do software MEGA-X(KUMAR et al., 2018), onde foi definida a coloração para cada aminoácido de acordo com seu grupo químico.

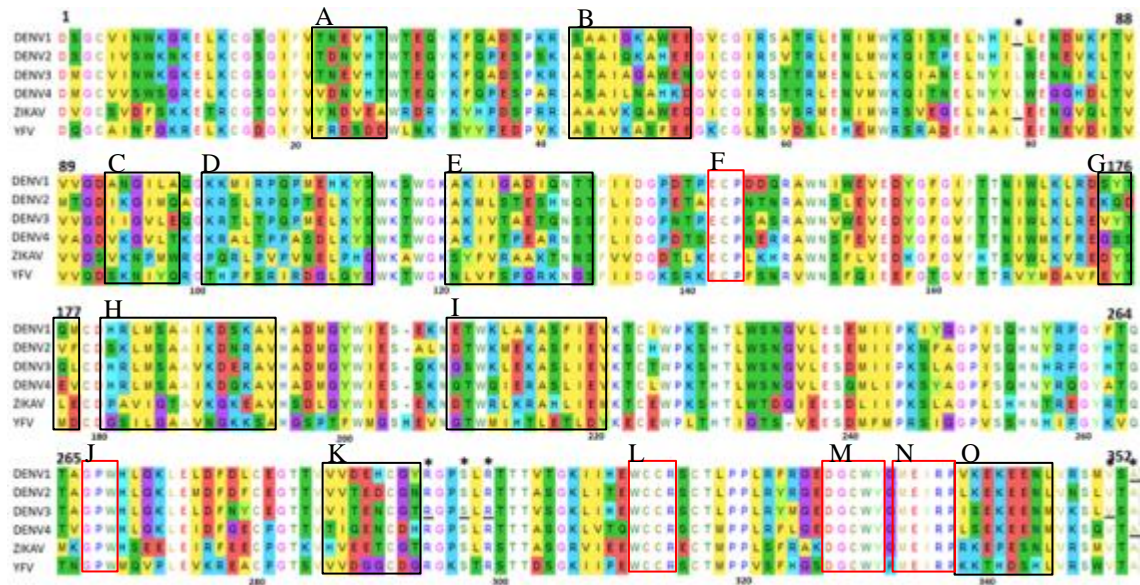


FIGURA8: Alinhamento das proteínas consenso para cada espécie. Alinhamento da proteína NS1, cada linha corresponde a um vírus estudado com indicação da espécie a esquerda, a direita da imagem existem números correspondentes aos resíduos da sequência(1-352), as regiões com quadrados coloridos são variáveis, enquanto as com letras coloridas são conservadas entre as espécies, os resíduos com asterisco indicam variação intra-específica em uma região conservada inter-específica.

As NS1 dos DENV(1-4), ZIKV e YFV tem identidade de 28,1% e similaridade de 50,6%, similaridade foi o termo utilizado para indicar regiões que variaram a estrutura do aminoácido entretanto não variaram a característica química, no alinhamento existem regiões com alta porcentagem de variabilidade nos domínios β -ladder⁽³¹⁻¹⁸⁰⁾ e Wing⁽¹⁸¹⁻³⁵²⁾ que estão em contato com o sistema imune, o domínio β -rool⁽¹⁻³⁰⁾ é hidrofóbico, responsável pelo ancoramento da proteína na membrana, além de carregar ácidos graxos das balsas lipídicas no hexâmero e é o menos variante.

A baixa variabilidade do β -rool pode ser explicada devido a sua não exposição ao organismo, além de ter função bem definida para o vírus, a informação é retratada por (NEUFELDT et al., 2018) onde o autor reafirma que o domínio β -rool⁽¹⁻³⁰⁾ tem função no ancoramento da NS1 na membrana do organismo, na interação de proteínas com a camada lipídica e estará no centro do hexâmero carregando as balsas lipídicas, tornando a área de difícil acesso para o sistema imune.

I. Regiões não conservadas

As regiões não conservadas foram definidas como um conjunto de pelo menos seis resíduos sequentes com variação que modificou o grupo químico de pelo menos dois aminoácidos por posição, elas estão entre os resíduo22-27 representado por A,

resíduo43-52 representado por B, resíduo93-98 representado por C, resíduo101-114 representado por D, resíduo121-129 representado por E, resíduo174-178 representado por G, resíduo181-194 representado por H, resíduo212-221 representado por I, resíduo286-293 representado por K e por último resíduo338-345 representado por O e circulado na figura8 com a cor preta, totalizando em 10 regiões.

De acordo com os resultados obtidos por (GONÇALVES et al., 2019) em um estudo com DENV1 e ZIKV a NS1 tem uma estrutura rígida devido a quantidade de ligações dissulfeto existentes, entretanto existem regiões flexíveis que estão relacionadas com a exposição ao solvente ou a funcionalidade do resíduo, a região 1-23 no monômero tem alta taxa de flexibilidade, porém na forma dimérica torna-se mais estável para cumprir com a sua função de ancorar-se com membranas, no alinhamento esta foi uma região pouco variável intra e interespecífica. A região 108-128 tem alta flexibilidade e está em contato com o solvente na forma de dímero (GONÇALVES et al., 2019), ao analisar o alinhamento pode-se observar que 108-128 é um local com alta variação intra específica demonstrada pelo quadro D, o resíduo 130 está exposto, e é ponto conservado de N glicosilação.

Somnake observou em seu trabalho realizado com DENV que o resíduo 130, além de fazer parte de um sítio de glicosilação está relacionado a estrutura hexamérica na proteína, razão por qual é conservado (SOMNUKE et al., 2011), no alinhamento da figura8 este resíduo foi conservado entre todas as espécies.

Os resíduos 19-21 da NS1 estão relacionados ao suporte estrutural das folhas β nos vírus DENV e ZIKV (GONÇALVES et al., 2019), nos resultados obtidos na figura8 os mesmos resíduos não mostram pouca variabilidade, sendo elas apenas entre aminoácidos de mesmo grupo químico indicando a conservação da região.

O resíduo 201W se mostra presente nos DENV e contribui para que a sua hidrofobicidade seja levemente maior quando comparada ao vírus ZIKV, isso pode proporcionar maior sensibilidade a flexibilidade da NS1 devido a dificuldade de entrada de solventes quando comparada a outras espécies(GONÇALVES et al., 2019), o resíduo 201 é conservado para os DENV, ZIKV e YFV, também não ocorrem variações intra-específicas para este resíduo.

O número de variações na cadeia polipeptídica das NS1's pode indicar funcionalidades para cada vírus de acordo com a sua sequência, fornecendo funções conservadas a cada espécie, informação relatada por Puerta-Guardo, 2019 onde a NS1

tem o papel de modular o tropismo viral diferenciado, no qual a proteína tem capacidade de romper a barreira endotelial de proteínas, o estudo foi feito através de uma membrana elétrica e mostrou que os DENV causam rompimento sistêmico, enquanto o ZIKVV acomete células cerebrais e de placenta e o YFV acometeu principalmente células hepáticas(PUERTA-GUARDO et al., 2019). Song et al., 2016 obteve resultados semelhantes no seu trabalho, onde utilizou 10 *Flavivirus* para a construção de uma árvore filogenética baseada na NS1 e obteve grupos diferentes sendo os principais DENV(1-4), YFV e *Tick-borne Encephalitis vírus* e ZIKV, assim foi possível observar que a NS1 tem características específicas para cada agrupamento(SONG et al., 2016).

As regiões da figura8 marcadas com asterisco apontam variação inter-especifica em locais onde não existe variação no alinhamento intra-especifico, as variações ocorrem nos resíduo79L-79I(Dv1) em 1 linhagem, resíduo79L-79S(ZIKV) em 1 linhagem, resíduo294R-249S(Dv3) em 1 linhagem, resíduo297S-297A(Dv3) em 2 linhagens, resíduo299R-299T (Dv3) em 1 linhagem, resíduo350V-350A(Dv3) em 2 linhagens, resíduo352 A-352S(Dv1) em 1 linhagem e resíduo352A-352T(Dv4) em 1 linhagem, de todas as mutações citadas 63,5% são entre grupos quimicos não sinonimos, a maioria mudando de hidrofobico para polar neutro, apesar de existirem poucas linhagens com variação em regiões conservadas ainda é possível que essa variação seja considerada devido ao baixo valor amostral. As mutações citadas acima podem se encaixar na ideia publicada por Song de que a NS1 varia entre as espécies e que isso pode acarretar funcionalizações diferentes (SONG et al., 2016), o conceito de pressão seletiva também é valido quando se trata de variações .

II. Regiões conservadas

As regiões conservadas da NS1 estão entre os aminoácidos 142-144 representado por F, 267-269 representado por J, 311-314 representado por L, 327-331 representado por M e 333-337 representado por N, demarcadas com retângulos vermelhos na figura8, essas regiões não contém mais do que 4resíduos consecutivos estáveis, podendo influênciar na construção de vacinas e teste diagnóstico, entretanto como citado em (EDSALL, 1946) as variações que contemplam resíduos homólogos podem ser válidas, uma vez que não indicam alterações bruscas, é possível que diversos epítomos com características químicas iguais possam se ligar a uma gama de anticorpos.

No artigo publicado por Edeling, 2014 com base no estudo da cadeia C-terminal de 9 *Flavivirus* incluindo DENV(1-4) e YFV foram observadas três regiões conservadas entre os resíduos 172-352, indicando ser uma região preservada devido a participação nas ligações de sulfeto que estabilizam a proteína e no ancoramento da proteína na membrana, as regiões se encontravam entre os resíduos 281-338, 181-233 e 263-386(EDELING, 2014), no trabalho essas regiões mostraram áreas conservadas sendo elas J, L, M, N e O e não conservadas sendo elas G, H, I e K em mesma proporção, apontando que por mais que a cadeia esteja exposta ao solvente e propicia a variações, ainda sim mostra áreas funcionais e conservadas.

Para (EDELING, 2014) o aminoácido prolina na posição 250 é conservado entre os *Flavivirus*, através de experimentos foi descoberto que variações nesse resíduo desestabilizam o dímero, no alinhamento realizado nesse trabalho o resíduo 250 encontra-se conservado entre todas as espécies.

Os resíduos 118, 122, 123 e 124 do ZIKV estão estritamente relacionados com a interação da proteína com a membrana de acordo com (HILGENFELD, 2016) e por isso devem ser estáveis entre as linhagens, entretanto o trabalho mostrou variação para 2 linhagens de Y122H, sendo de um grupo aromático não polar para um grupo positivo de cadeia fechada na posição 122, o polimorfismo na posição pode acarretar mudanças na estabilidade estrutural, possibilidade que ainda não possui nenhum apontamento na literatura.

O resíduo 115W no ZIKV é conservado e esta intimamente relacionado com a produção de vírions maduros e infecciosos (HILGENFELD, 2016), não foram encontrados polimorfismos nesse resíduo para as linhagens coletadas no trabalho.

Ensaio realizados por (AKEY et al., 2015) mostraram que mutações entre as posições 159 e 162 de DENV2 foram deletérias, onde estas, também não foram identificadas nas sequencias aqui analisadas, sendo relatada como região importante para a replicação viral.

De acordo com (EDELING, 2014) os resíduos 107, 183-185, 182, 189, 206, 210, 229, 231-233, 250 e 254 são conservados nos DENV(1-4), já que participam das ligações entre dímeros, sendo que uma variação no resíduo da posição 250 foi capaz de desestabilizar os dímeros na formação do hexâmero, nenhum dos resíduos citados sofreu variação nas espécies utilizadas no trabalho.

O resíduo 207 deve ser conservado entre os DENV pois tem importância na secreção e dobramento da NS1 por chaperonas (SOMNUKE et al., 2011), não foram encontradas variações para o resíduo dentro de cada uma das espécies.

Os resíduos 230, 232 e 234 de YFV formam uma ponte de hidrogênio que confere estabilidade as ligações monoméricas (MENEZES et al., 2019), após análise da tabela de variações é possível afirmar que os resíduos não são modificados por linhagens, confirmando a sua importância para a NS1.

4.2 Variações intra-específicas

Posição			DENV1	DENV2	DENV3	DENV4	YFV	ZIKV
	C.	M.	45	59	84	58	59	49
			%					
B-root	5	I	V		31,5		1,75	
		V	A			2,3		
	6	I	V					
	8	F	L					2,2
	9	S	G				3,4	
	11	R	K				3,5	
	17	S	N	2,2				
	21	I	V				3,4	
	22	I	V				1,75	
	27	T	A			1,19		
B-ladder	30	N	S				17,2	
	38	S	F		1,75			
	48	H	R	31,1				
		A	S				3,4	
	71	S	V			1,19		
	77	H	Y	42,2				
		Y	H			1,19		
	79	L	I	20				
L		S				10,3		

84	V	I		1,75				
	I	V			2,3			
92	D	N	26					
93	A	I	2,2					
	I	T			1,75			
94	S	P					13,7	
	N	D	2,2					
96	I	T			4,76			
	I	V					10,3	
98	A	I	6,6					
	E	K			1,19			
103	T	S				1,75		
	P	S					3,4	
105	R	K	17,7					
	T	A				1,75		
108	A	V				10,7		
109	S	N				1,75		
110	P	G				3,5		
121	A	V						
	A	S		1,75	2,3			
122	K	R				3,5		
	Y	H						4,2
123	I	M			3,5			
125	S	P		8,7				
128	I	T	2,2					
	A	T				1,75		
129	Q	R			1,19			
131	I	T	6,6					
	Q	N				3,5		
135	I	V				1,75		
139	N	S			3,5			
141	I	T	2,2					
142	I	Y						

Wing	146	D	E/G	40/2,2				
		T	A		3,5			
		A	T			1,19		
		K	E					4,2
		152	S	F			1,75	
		153	F	L			10,5	4,2
		175	Y	H	40		1,19	
	177	M	I				10,3	
	188	I	V		3,5			
	190	D	N		1,75			
	191	K	R				4,2	
	212	M	V		1,75			
	214	R	K			10,5		
	218	T	A				1,75	
	222	T	A				12,2	
	224	I	V/T	16/2,2				
	233	T	A				4,2	
	240	D	E				3,5	
	242	I	V	31,1				
246	I	T	4,4					
	N	V		3,5				
	S	A			1,75			
247	V	M	2,2					
251	I	F	2,2					
256	H	Y			2,3			
	F	I				1,75		
275	M	D		1,75				
284	S	I						
	S	N				26,8		
287	I	V				10,3		
288	T	S			2,3			
290	N	D				10,5		
	N	G				10,3		

292	G	E		12,2				
293	Y	V/T	31,1/4,4					
	N	K		12,2				
294	R	S			1,19			
297	S	A			2,3			
299	R	T			1,19			
307	I	V					10,3	
324	V	I			1,19			
	R	W						2,2
332	P	S					17,2	
338	K	R						
339	S	N			2,3	1,75	10,3	
342	N	E					10,3	
343	D	S					3,4	
349	M	V	2,2					19,1
350	V	A			2,3			
351	S	A	2,2					
	T	A				8,7		
352	A	S	2,2					
	A	T				1,75		

TABELA1: Tabela de variação intra-específica. A tabela indica os resíduos que sofrem variação, bem como a porcentagem da variação para cada espécie do trabalho, a esquerda existem a delimitação das regiões da proteína, seguido da indicação do resíduo, sendo C. conservado assim como a variação sofrida indicada por M., logo após existem a espécie viral e a porcentagem para cada espécie.

A tabela1 mostra as variações e as porcentagens com que elas ocorrem para cada espécie, no total foram observadas 105 variantes., a região com menor quantia de variantes é a β -rool, já os domínios β -ladder e Wing obtiveram quantia semelhante de variações.

I. Domínio β -rool

A região β -rool tem 10,4% de variantes do total, sendo 11 (homólogas e 2 não homólogas), as variações químicas neste domínio contemplam as espécies YFV em 3,4% das linhagens e DENV3 em 1,19% das linhagens, as variações apenas físicas contemplam o DENV2 na posição 5 para 31,5% das linhagens, seguido do YFV com

17,2% na posição 30. o restante das porcentagens variam entre 1,19% e 3,5% para todos os vírus citados, a explicação para a discrepância entre a porcentagem de variações entre mesmo grupo químico ou grupos químicos diferentes está na funcionalidade da região e no seu baixo contato com os anticorpos.

De acordo com BROWN et al., 2016, o domínio β -rool tem uma sequência de resíduos estritamente relacionada ao ancoramento da NS1 na membrana e ligação hexamérica, especialmente os resíduo27-30 , curiosamente os resíduos deste domínio mostram polimorfismos para DENV3 27T-27A trocando de um grupo polar neutro para um grupo hidrofóbico com porcentagem de 1,19%(1linhagem), e YFV 30N-30S permanecendo no mesmo grupo químico, mas alterando a estrutura física do aminoácido afetando 17,2%(8linhagens), as mutações podem em último caso, implicar na variação da dinâmica estrutural da cadeia polipeptídica, podendo até mesmo causar alterações em suas funções.

Os resíduos 1, 5, 20 e 22 localizados no domínio β -rool do YFV são responsáveis por fazer ligações de ponte salina com outros resíduos dos domínio β -ladder e Wing com finalidade de estabilizar a proteína na forma dimérica, simulações realizadas por (MENEZES et al., 2019) indicam que mutações nessas regiões poderiam aumentar a perturbação da proteína quando comparada a energia de estado natural no estudo citado anteriormente não ocorrem variações inter-específicas nos locais citados, curiosamente ocorrem variações intra-específicas em I5V e V5A para os DENV2, DENV3 e DENV4 e em I22V para o DENV4, apesar das variações serem dentro do mesmo grupo químico, podem indicar mudança na estrutura do aminoácido.

II. Domínio Wing

O domínio Wing tem 45,7% do total de variantes , sendo 17 variações homólogas e 31 variações não homólogas, neste domínio existe um aumento na quantidade de variações. Na análise de variações nos domínios foram indicadas detalhadamente apenas as variações que afetaram mais que 10% do total de linhagens, totalizando em 31.

No DENV1 as variações físicas são nos resíduo L79I em 9 linhagens e resíduo R105K em 8 linhagens, nos DENV2 e DENV3 nenhuma variação acomete mais do que 10% do total de sequências, no DENV4 ocorre variação nos A108V para 6 linhagens e resíduo F153L em 6 linhagens, no YFV ocorre variação nos resíduo I96V

para 5 linhagens e resíduo I177M para 5 linhagens, para o ZIKVV não foram encontradas mutações consideráveis.

As variações químicas do domínio Wing ocorreram em valores notáveis, para o DENV1 o resíduo H48R varia de um grupo aromático carregado positivamente para um grupo de cadeia aberta positiva afetando 31,1% das linhagens, resíduo H77Y mudando de um grupo aromático positivo para aromático neutro em 42,2% das linhagens, resíduo D92N mudando de um grupo carregado negativo para polar neutro e afetando 26,6% das linhagens e resíduo Y175H mudando de um grupo aromático neutro para aromático positivo em 40% das linhagens, para o DENV2 houve variação apenas no resíduo S125P mudando de um grupo polar neutro para um grupo positivo de cadeia aberta em 4 linhagens, o DENV3 não obtivemos mutações entre grupos químicos distintos consideráveis, assim como para o DENV4 e ZIKVV, o YFV mostra mutação no resíduo S93P mudando de um grupo polar neutro para um positivo de cadeia aberta em 11,8% das linhagens.

Os resíduos 123 e 124 da NS1 de ZIKVV e DENV(1-4) são relatados como fatores importantes na associação com a membrana, por isso devem ser hidrofóbicos (SONG et al., 2016) na tabela de variações esses resíduos se mantiveram conservados com as características citadas sem que houvesse variações relevantes.

III. Domínio β -ladder

A região β -ladder mostra 43,9% do total de variações, entretanto poucas espécies mostram mais do que 10% de variação em suas linhagens. Apenas uma variação física do DENV1 foi observada no resíduo I242V em 31,1% das linhagens, nos DENV2, DENV3 e DENV4 não existem mutações homólogas significativas., no YFV foram observadas variações no resíduo S248N em 30,5% das linhagens, no ZIKV a variação ocorre para o resíduo M349V afetando 20,4% das linhagens.

Variações não homólogas não foram encontradas entre as sequências do DENV1, contrapondo com o domínio Wing onde foi o vírus apresentou variações, já o DENV2 apresentou apenas uma variação no resíduo N342E afetando 10,1% das linhagens de um grupo polar neutro para um carregado negativo, o DENV3 não apresentou mutações significativas, o sorotipo do DENV4 apresentou uma variação significativa no resíduo N290D afetando 10,3% das linhagens trocando de um grupo polar neutro para um carregado negativamente. Além da presença de variações nos sorotipos dos DENV's foi observada a variação 332P-332S no YFV, mudando de

positivo de cadeia aberta para polar neutro e afetando 8 linhagens, o ZIKV apresenta apenas uma variação no resíduo T222A mudando de um grupo polar neutro para hidrofóbico e afetando 14,2% das sequências avaliadas pelo estudo.

4.3 Antigênicidade da NS1

A proteína NS1 é comum entre os *Flavivirus*, assim ela pode ser recorrida na construção de um método diagnóstico ou vacinal, de acordo com as análises das sequências dos vírus DENV(1-4), YFV e ZIKV existem locais conservados ou com baixa porcentagem de polimorfismos, os mesmos que poderão ser aplicados nas metodologias citadas acima. Para confirmar locais reconhecidos pelo sistema imune fizemos a predição de antigênicidade para todas as NS1's dos vírus estudados.

Os resultados demonstrados até o momento indicam que a NS1 é boa candidata tanto na construção de testes diagnósticos como de vacinas, uma vez que ela mostra áreas conservadas inter e intra específicas e também áreas não conservadas entre as espécies, para que os métodos citados acima possam funcionar é necessário que a sequência de peptídeos seja antigênica.

A antigênicidade da NS1 foi medida pelo VaxiJen(DOYTCHINOVA; FLOWER, 2007), servidor de predição de antígenos com base nas propriedades físico-químicas da proteína, os resultados se encontram na tabela2, um score de 0,4 foi admitido pelo servidor para definição de antigênicidade, valores preditos acima do score dão resultado de antigênicidade positivo.

Espécie	Vaxijen Score
DENV1	0,52
DENV2	0,6
DENV3	0,49
DENV4	0,52
YFV	0,57
ZIKV	0,47

TABELA2: Antigênicidade da NS1. A tabela indica os valores de antigênicidade da proteína NS1 medido pelo VaxiJen (DOYTCHINOVA; FLOWER, 2007) para cada espécie, o valor mínimo para que uma proteína seja considerada antigênica é de 0,4.

De acordo com os resultados obtidos no VaxiJen a NS1 é uma proteína antigênica para todas as espécies presentes no estudo, diversos estudos certificam a capacidade da NS1 de estimular o sistema imune, como (HERTZ et al., 2018) que

observou a estimulação de IgE e IgM nos DENV(1-4) pela NS1 em camundongos, Jiang identificou anticorpos anti-NS1 através de ELISA no soro de coelhos infectados(JIANG et al., 2010) e (NAGAR; SAVARGAONKAR; ANVIKAR, 2020) que identificou resposta de anticorpos IgM, IgE e anti-NS1 contra a proteína, a antigênicidade contra a NS1 é provocada uma vez que a proteína é secretada em grandes quantidades durante a fase aguda da infecção, informação confirmada por(HERTZ et al., 2018).

4.4 Predição de peptídeos imunogênicos e epítomos lineares para células B

I. Bepipred

	DENV1		DENV2		DENV3		DENV4		YFV		ZIKV	
	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final
1	26	51	23	50	26	40	26	39	25	37	26	40
2	93	130	93	130	93	130	116	129	93	119	93	131
3	140	150	140	150	141	150	140	150	123	132	137	148
4	173	179	228	241	173	179	172	179	138	147	172	179
5	228	243	248	272	228	243	228	242	224	242	227	242
6	248	270	289	318	248	283	248	272	248	283	249	282
7	272	282	338	346			291	318	291	318	290	318
8	290	318					338	345	338	348	338	348
9	338	348										

TABELA 3: Epítomos Bepipred. Sequência de epítomos fornecidos pelo servidor Bepipred, com a quantia total de epítomos, resíduos inicial e final da sequência e espécie indexadas.

De acordo com (KAM et al., 2019) os epítomos das regiões 17-34 e 257-274 foram encontrados em soros de pacientes infectados com ZIKVV através de um teste de ELISA, os epítomos são considerados diferenciais e tem similaridade de 44,4% e 61,1% respectivamente com os DENV, nosso trabalho mostrou ambos epítomos para os vírus ZIKV e DENV na tabela6, as identidades encontradas no trabalho foram de 44,7% e 61,8% afirmando os dados ilustrados por KAM.

O epítomo 17-34 relatado por (KAM et al., 2019) mostrou altos índices de reconhecimento de anticorpos em mulheres grávidas, podendo ser considerado um marcador de infecção para gestantes, a questão a ser discutida leva em conta a região em que se encontra o epítomo, uma vez que como citado anteriormente não está exposta ao sistema imune, duas alternativas podem ser utilizadas, a primeira de que a

conformação dos vírus fornece pequenos espaços para entrada de solventes (GONÇALVES et al., 2019) e a segunda de que a proteína foi encontrada em forma monomérica pelo sistema de defesa em alguma etapa do processo oligomerização, curiosamente o epítipo mostrou maiores porcentagens de identificação em grávidas, abrindo espaço para pesquisas relacionadas a modulação do sistema imune durante o período gestacional.

II. IMED

A plataforma IMed foi uma das selecionadas para predição de peptídeos para células T, o *input* na plataforma foi feito com múltiplas sequências de alinhamento(MAS) das NS1 de cada espécie viral construídas no trabalho, o output foi um conjunto de aminoácidos consequentes com baixa variabilidade e alta validade na ligação de anticorpos de célula T, os resultados encontrados são mostrados na tabela3(GARCIA-BORONAT et al., 2008). A plataforma foi optada pois sua metodologia é direcionada principalmente para epítipos vacinais contra patógenos com alta variabilidade, o que acontece para a proteína NS1 de *Flavivirus*.

	DENV1		DENV2		DENV3		DENV4		YFV		ZIKV	
	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final
1	11	23	11	19	11	23	11	24	11	22	15	23
2	38	46	38	47	38	46	38	48	30	48	38	48
3	51	57	51	62	51	57	51	57	55	62	53	59
4	85	91	71	77	84	98	74	80	84	92	83	92
5	167	173	103	113	156	164	84	98	119	125	101	110
6	175	181	154	164	167	181	103	113	140	146	119	125
7	187	196	173	181	183	196	175	181	167	175	131	137
8	212	229	216	229	213	225	187	196	179	186	139	146
9	242	250	247	254	232	239	213	229	247	262	151	173
10	269	295	256	262	241	250	232	250	268	289	175	186
11	308	321	281	289	269	279	261	267	308	323	189	200
12			308	321	281	288	304	323	325	331	213	219
13					303	321			341	348	238	253
14											281	289
15											308	315

16										317	323
----	--	--	--	--	--	--	--	--	--	-----	-----

TABELA 4: Epítomos Imed. Epítomos preditos pelo servidor Imed, sendo indexadas a quantidade de epítomos, a espécie e os resíduos inicial e final da sequência.

III. NTCcell

Epítomos lineares para o complexo MHCI foram buscados na base de dados IEDB através do servidor NTCcell, a seleção dos epítomos pelo NTCcell se dá por uma busca de epítomos antigênicos para vírus, bactéria e cancerígenos. O *input* foi feito com a sequência consenso de cada uma das seis espécies (VITA et al., 2019).

O método foi selecionado uma vez que a NS1 é responsável por estimular grande parte do sistema imune através de células TCD4⁺ e TCD8⁺ (GLASNER et al., 2018).

	DENV1		DENV2		DENV3		DENV4		YFV		ZIKV	
	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final
1	24	32	21	32	24	32	24	32	24	32	24	32
2	125	133	125	133	125	133	108	116	89	98	192	203
3	163	171	150	158	163	171	150	158	145	153	269	277
4	285	293	163	171	323	331	163	171	167	177		
5							323	331	177	184		
6									252	260		

TABELA 5: Epítomos NTCcell. Epítomos indicados pela plataforma NTCcell, sendo indexadas a numeração dos epítomos, vírus correspondentes e resíduos inicial e final da sequência.

96 epítomos de células T foram encontrados, o domínio β -rool demonstrou 12 epítomos, o domínio β -ladder mostrou 49 epítomos e o Wing 35 epítomos.

No estudo realizado por Reynolds através da reprodução de uma infecção de ZIKV em camundongos foi observado a presença de dois epítomos para NS1 encontrados por células TCD4⁺, sendo eles localizados nas regiões 171-190 e 271-290 (REYNOLDS et al., 2018), o epítopo 171-190 não foi localizado nas tabelas 3 e 4, entretanto o epítopo de posição 10 na tabela 3 para ZIKV está próximo da região citada, indicando que a mesma tem movimentação de anticorpos, o epítopo 271-290 não foi demonstrado nas tabelas 3 e 4, o resíduo de posição 14 na tabela 3 também se encontra próximo aos resultados obtidos por Reynolds, mostrando a possível antigênicidade do local.

4.5 Modelagem estrutural.

Uma vez que a proteína NS1 foi dada como antigênica e ocorreu a predição de epítomos para Células T o próximo passo foi realizar a modelagem estrutural.

A modelagem estrutural da NS1 foi feita em dois passos, o primeiro passo foi a construção de um monômero por via da plataforma I-TASSER(ROY; KUCUKURAL; ZHANG, 2010), seguido da dimerização pelo GRAMM-X(TOVCHIGRECHKO; VAKSER, 2006) e refinamentos pelas plataformas GalaxyRefineComplex(KO et al., 2012) e ModRefiner(XU; ZHANG, 2011). A modelagem foi feita com o objetivo de elucidar a proteína, otimizando a visualização da estrutura, facilitando a predição de epítomos confirmacionais e a análise de propriedades como área de acessibilidade ao solvente além de semelhanças entre as proteínas de cada espécie, a proteína modelada está elucidada na figura 9 a e b.

A plataforma I-TASSER fornece uma abordagem composta na construção molecular utilizando a *de novo*, que consiste na busca por modelos homólogos da proteína submetida e segmentação que busca resíduos semelhantes a proteína submetida em estruturas não homólogas, mas com dobras semelhantes, em um banco de dados de proteínas modeladas, o método foi classificado como o melhor para a previsão de estruturas 3D em diversos concursos realizados nos últimos anos (ROY; KUCUKURAL; ZHANG, 2010).

O servidor GRAMM-X foi recorrido para a construção da interação da NS1 no formato de dímero, a capacidade de modificar a estrutura da proteína de acordo com as necessidades estimulando um modelo próximo da realidade é uma vantagem no uso dessa ferramenta, o servidor utiliza uma projeção de grades calculadas pelo potencial de Lennard Jones para melhor suavização e projeção da estrutura(TOVCHIGRECHKO; VAKSER, 2006).

As etapas de refinamento foram buscadas com o objetivo de ter estruturas validadas e o mais próximo do que encontramos na realidade, para que os ajustes na proteína recém modeladas fossem feitos com eficiência escolhemos dois servidores, o servidor GalaxyRefine(KO et al., 2012) foi usado para o refinamento de regiões imprecisas de complexos proteicos, enquanto o ModRefine(XU; ZHANG, 2011) se baseia na estrutura completa da proteína para performar o refinamento estrutural. Foram usados dois servidores pois possuem metodologias complementares, assim obtivemos estruturas com ótimos padrões, mostradas na figura 9.

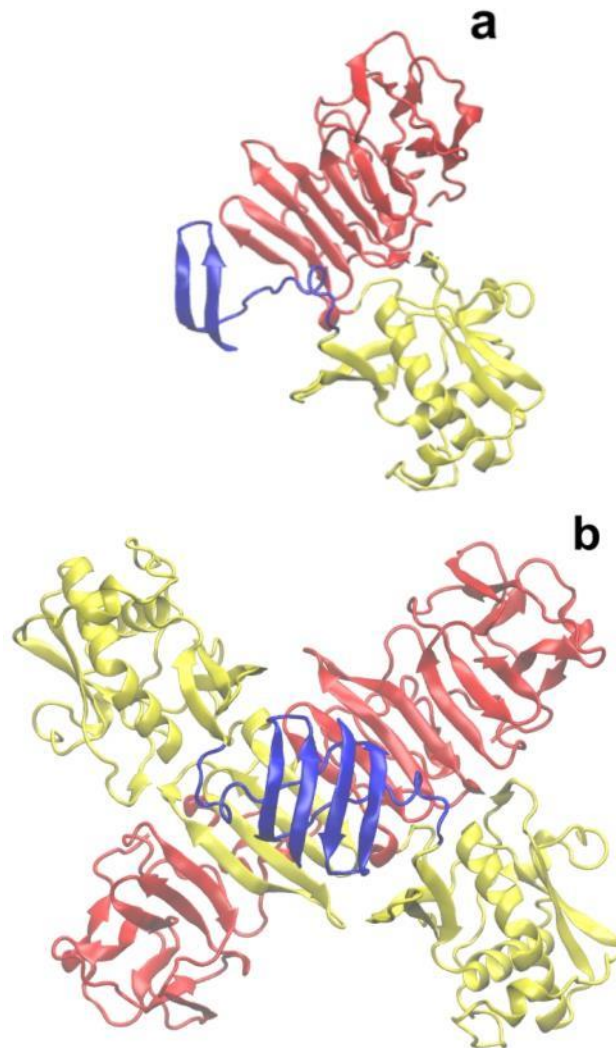


FIGURA9: a- Modelagem NS1 monômero. Monômero da proteína modelado com os domínios indicados pelas cores azul, amarelo e vermelho, sendo respectivamente β -roll, β -ladder e Wing. **b- Modelagem NS1 dímero.** Proteína NS1 modelada em forma de dímero com os domínios representados pelas respectivas cores, no dímero os domínios do monômero são duplicados.

O servidor MolProbity é amplamente utilizado na validação de estruturas de proteína e DNA, a vantagem de uso está na velocidade do processamento e na análise fina de contato de todos os átomos, os principais parâmetros calculados nesse servidor são Clashscore que analisa o número de conflitos por 1000 átomos próximos respeitando uma distância inicial, Ramashandran que calcula a localização exata de cada átomo na cadeia central e a sua liberdade de rotação, a pontuação geral MolProbity score analisa um conjunto de estatísticas de qualidade da proteína através do logaritmo do clashscore, ramachandra e resoluções dos rotameros com base na cristalografia da proteína original(CHEN et al., 2010). Os resultados obtidos pelo servidor após a modelagem e refinamentos estão na tabela5.

	DENV1	DENV2	DENV3	DENV4	ZIKV	YFV
Clashscore	16,98	12,43	9,83	12,59	10,2	14,39
Original	42 th	57 th	72 th	60 th	72 th	52 th
Clashscore	12,87	11,32	7,84	12,11	11,17	9,77
GalaxyRefine	61 th	65 th	82 th	62 th	66 th	73 th
Clashscore	0	0	0	0	0	0
ModRefine	100 th	100 th	100 th	100 th	100 th	100 th
MolProbity	3,04	3,00	2,93	3,10	2,96	3,00
Original	21 nd	23 nd	23 nd	19 nd	26 nd	23 nd
MolProbity	1,83	1,63	1,49	1,79	1,98	1,57
GalaxyRefine	84 nd	91 th	95 nd	85 nd	90 nd	93 nd
MolProbity	0,72	0,65	0,77	0,80	1,01	0,91
ModRefine	100 nd	100 th	100 nd	100 nd	100 nd	100 nd

TABELA 6: Validação das proteínas modeladas. Validação das proteínas pelo servidor MolProbity, com os principais parâmetros oferecidos pelo servidor relacionados as espécies virais.

Como citado anteriormente, o MolProbity score é o parâmetro geral da ferramenta, onde seu resultado é dado em percentil, e, neste caso, quanto mais próximo o score tende a 0, maior o percentil apresentado, conseqüentemente, maior é a proximidade da estrutura modelada e do cristal, a estrutura original mostrou percentis ruins quando comparados com a refinada, principalmente no DENV4 onde a proteína deslocou da posição 19 para a posição 100 sendo totalmente idêntica a cristalografia, o segundo refinamento foi necessário, uma vez que nenhum resultado mostrado pelo GalaxyRefine chegou a 100. Ramashandran é o conjunto de resultados referente a localização e rotação dos átomos da proteína, os modelos refinados chegaram as melhores coordenadas, atingindo o percentil 100 e se mostrando idêntica a todas as estruturas de NS1 existentes nos bancos de dados, vale ressaltar que a quantia de NS1s publicada em banco de dados é pequena, além de que grande parte das estruturas encontradas nesses locais estão incompletas, assim fica concluído que a proteína construída tem grande importância no cenário atual.

4.6 Epítomos conformacionais para células B

I. Ellipro

O preditor Ellipro foi usado para definir epítomos de células B conformacionais, seu uso complementou os resultados trazidos pelo Bepipred.

	DENV1		DENV2		DENV3		DENV4		YFV		ZIKV	
	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final	Inicial	Final
1	11	16	11	16	36	58	36	56	36	58	36	58
2	14	18	36	56	69	95	69	95	69	95	69	94
3	37	58	69	94	100	149	100	148	100	152	104	149
4	69	95	100	174	237	241	160	164	237	241	160	164
5	100	113	277	352	258	267	263	267	258	270	257	261
6	119	149	388	408	278	320	278	320	278	322	263	270
7	172	175	421	446	324	352	325	352	324	352	278	320
8	257	267	452	501	388	410	388	408	388	410	325	352
9	278	320	609	622	421	445	421	447	421	504	388	408
10	325	352	629	673	452	501	452	499	512	516	421	445
11	389	410	676	704	421	445	615	619	610	622	451	500
12	421	446			388	410	630	674	630	674	609	612
13	452	504			610	622	677	704			615	619
14	609	622			630	674					630	672
15	630	672			677	704					677	704
16	676	704										

TABELA 7: Epítomos Ellipro. Epítomos preditos pelo servidor Ellipro, com quantidade de epítomos indexada, espécies correspondentes e resíduos inicial e final.

Foram encontrados 123 epítomos para célula B, sendo 9 para o domínio β -rool, 38 para o domínio β -ladder e 76 para o domínio Wing.No trabalho realizado por (VITA et al., 2019) alguns epítomos para ZIKV foram preditos e validados através de microarray, sendo localizados nas regiões 50-65, 75-88, 108-129, 206-215, 275-290, 311-330 e 341-351, no trabalho mostramos epítomos na tabela7 nas posições 2 para o epítomo 75-88 em todos os vírus, 3 para o epítomo 108-129 para todos os vírus exceto DENV1, o epítomo 275-290 foi identificado na posição 9 da tabela para DENV1, 5

para DENV2, 6 para DENV3, DENV4 e YFV e 7 para ZIKV, confirmando a possibilidade de antigênicidade das predições feitas no trabalho.

O estudo de (JAYATHILAKA et al., [s.d.]) mostrou que durante uma infecção secundária de DENV o perfil de anticorpos mudou, o estudo foi baseado nos DENV1 e DENV2, o perfil de anticorpos em uma primeira infecção para DENV1 foi entre os resíduos 297-311, 318-334, 341-353, enquanto em uma versão hemorrágica o perfil contemplou os resíduos 73-107 e 115-137, já para o DENV2 o perfil de anticorpos da primeira infecção foi entre os resíduos 62-102, 177-210, 236-280 e 289-306, o perfil analisado na versão hemorrágica mostrou os resíduos 33-111, 115-141, 146-163. Dentre os epítomos da primeira infecção para DENV1 listamos na tabela6 297-311 registrado na posição 8 e 341-353 localizado na posição 10 da tabela7, para DENV2 os epítomos 62-102 localizado na posição 3 da tabela7, 236-280 e 289-306 registrado nas posições 4, 5 e 6 da tabela6, parte dos epítomos descrito no trabalho foram detectadas na primeira infecção, o que pode ser explicado pela predição estrutural, que usa o PDB se aproximando o máximo possível da proteína no organismo, os epítomos encontrados no soro de pacientes e não relatado nas predições podem estar relacionados a eficiência do método preditivo, assim como a diferença entre as sequências submetidas e as linhagens circulantes, para a infecção hemorrágica o epítomo 73-107 foi listado na posição 2 da tabela6 e nas posições 4 e 5 da tabela7, enquanto para DENV2 os epítomos 33-111 e 115-141 foram identificados respectivamente nas posições 1 da tabela6 e 2 da tabela7, grande parte dos epítomos localizados na segunda infecção pelo estudo de Jayathilaka foram encontrados nos servidores utilizados.

Pode-se concluir que a maioria dos peptídeos imunogênicos se encontraram nas regiões β -ladder e Wing, confirmando a hipótese de que essas são boas regiões para alvos vacinais e de teste diagnóstico, é válida uma busca na literatura, bem como descrição do que já foi indicado pelo sistema imune para que um futuro método seja validado de forma efetiva.

4.7 Variações de superfície da NS1.

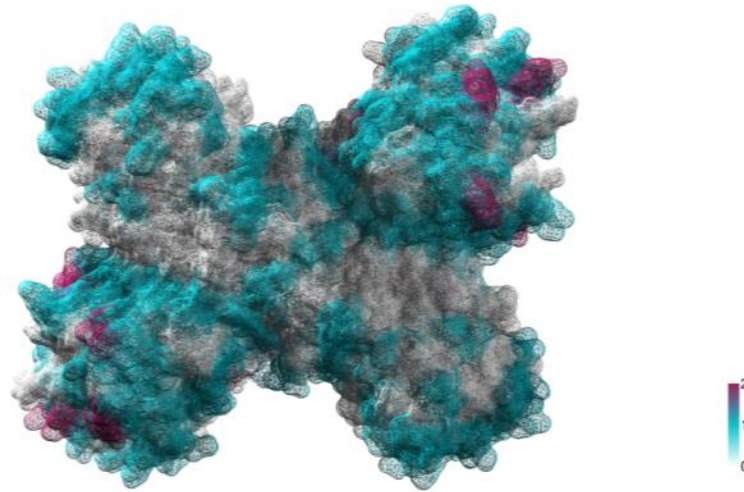


FIGURA10: Variação de cargas elétricas na superfície, face externa.

A figura10 mostra as variações de cargas elétricas na NS1 em sua face externa, as regiões demarcadas em cinza indicam que a proteína mantém as mesmas cargas entre os vírus DENV, YFV e ZIKV, as regiões em azul indicam mudança em um nível de carga, sendo negativa ou positiva para neutra, as regiões em roxo indicam mudança em dois níveis de carga, sendo de negativo para positivo ou vice-versa. O domínio β -ladder mostra-se mais conservado, enquanto as extremidades do domínio Wing estão com maior variabilidade entre as cargas, no geral a proteína não mostra mudanças bruscas, isso pode ocorrer devido a sua funcionalidade na infecção, uma vez que alterações consideráveis poderão interferir na interação da proteína com diversos fatores do organismo.

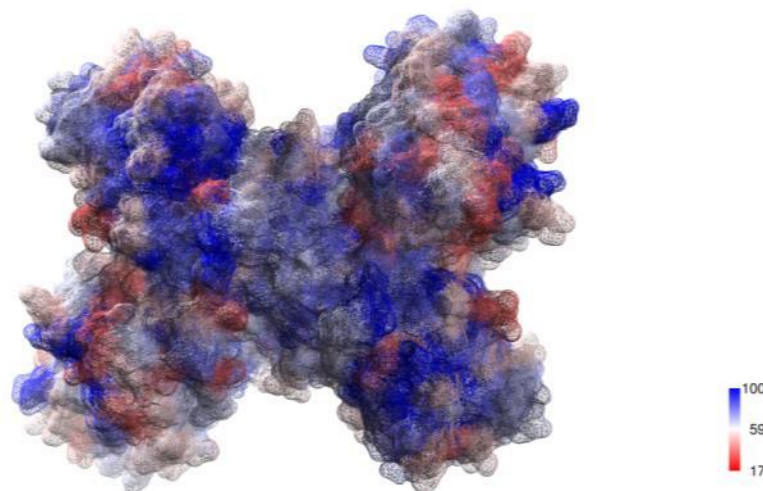


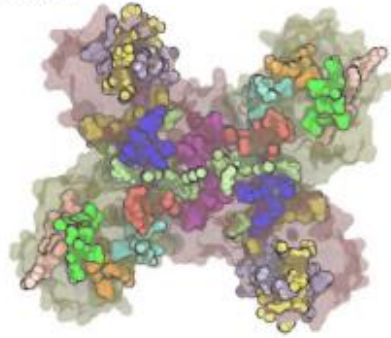
Figura 11: Variação de resíduos na superfície, face externa.

A figura11 mostra a identidade de resíduos que compõem a NS1, a proteína esta sendo vista externamente, a coloração vermelha indica baixa identidade entre os resíduos e aparece com maior frequência no domínio Wing, onde está localizada a região C-terminal, a coloração branca indica resíduos com identidade média entre os vírus DENV, YFV e ZIKV e está mais concentrada no domínio β -ladder assim como na sombra do domínio β -rool indicada na região central da imagem, as regiões em azul são idênticas entre as proteínas das espécies e estão representadas nos domínios β -ladder e β -rool. A região Wing é a que menos mostra identidade entre os vírus estudados, isso deve ocorrer devido a alta incidência de ataque de anticorpos, além de sua baixa funcionalidade em um contexto de infecção viral, o domínio β -rool é o mais conservado entre as espécies.

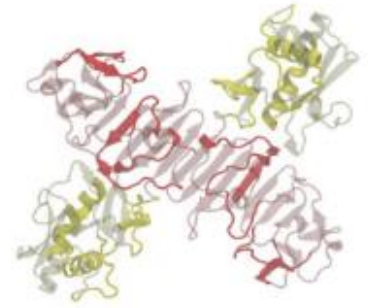
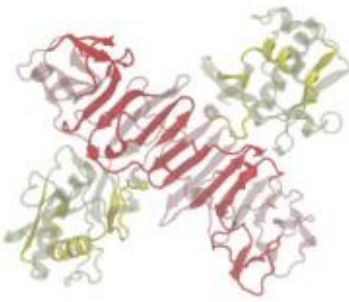
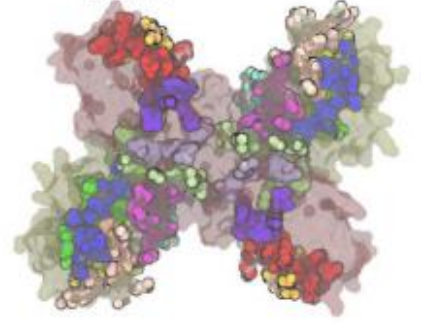
4.8 Modelagem de epítomos

Foi realizada a modelagem dos epítomos encontrados para os servidores que indicaram peptídeos imunogênicos e lineares para células B para todas as espécies estudadas, elucidada na figura13, os peptídeos que se encontravam em diferentes métodos foram reduzidos a apenas um, pode-se perceber que a proteína NS1 além de ser imunogênica apresenta grande quantidade de epítomos para os domínios que estão em contato com o sistema imune(β -ladder e Wing), o que indica que ela pode ser identificada tanto para testes diagnóstico quanto para vacinas através do encontro antígeno/anticorpo, os peptídeos para células B não foram indicados, uma vez que grande parte destes contemplaram regiões já indicadas pelos peptídeos citados no início do texto.

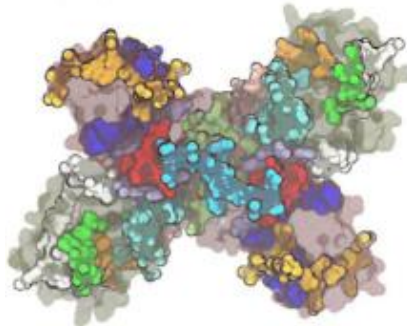
Denv1



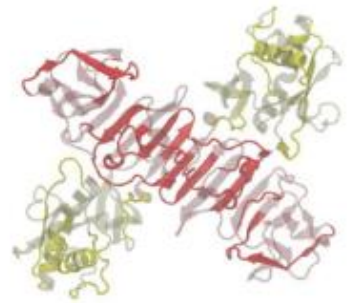
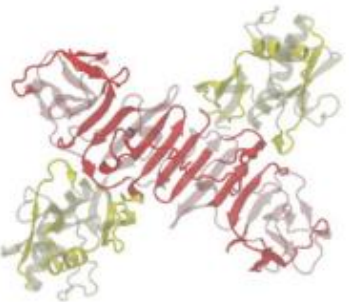
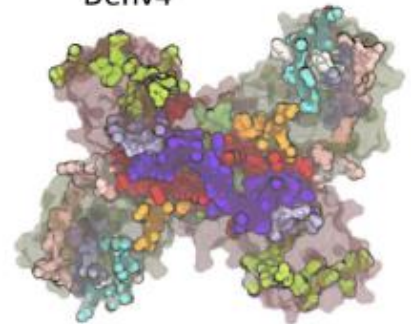
Denv2



Denv3



Denv4



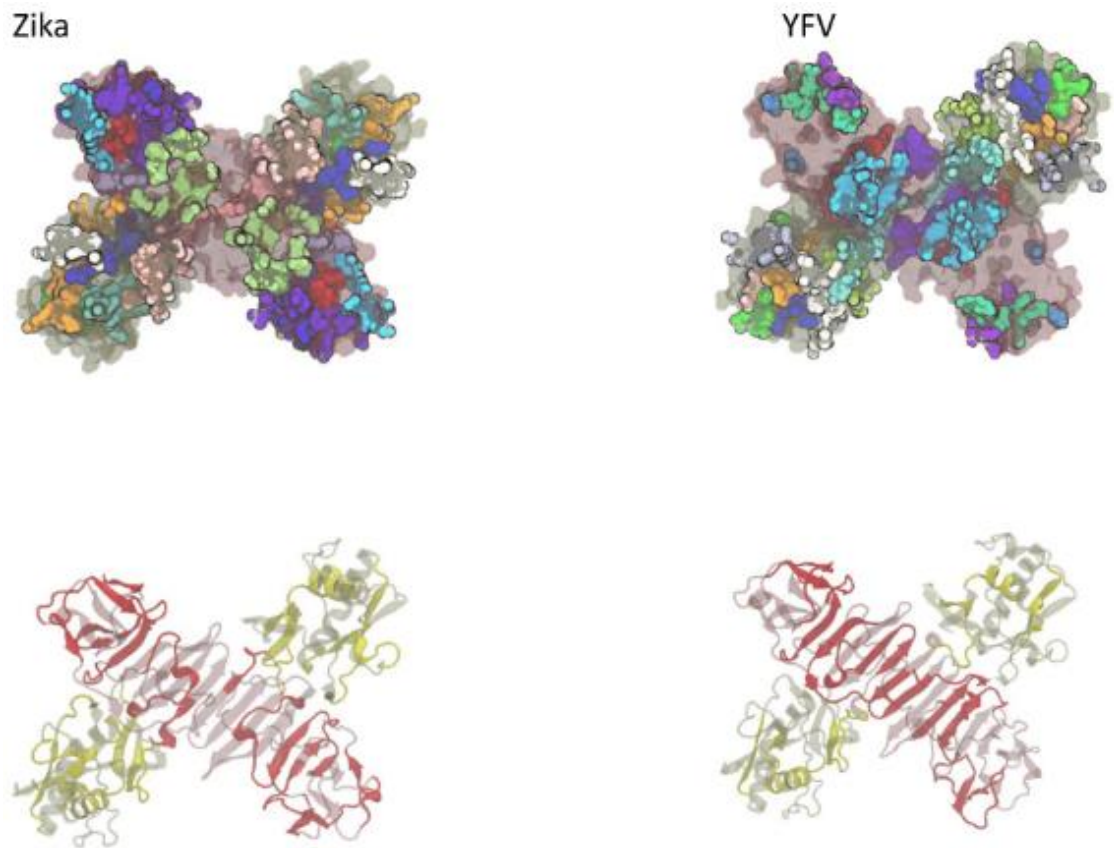


Figura 12- Peptídeos imunogênicos modelados. Modelagem molecular de peptídeos na proteína NS1, as espécies correspondentes a imagem se encontram à esquerda, na figura tridimensional cada cor representa um epítipo encontrado após seleção e os domínios estão representados nas cores azul e vermelha, abaixo de cada imagem tridimensional está a imagem bidimensional onde os peptídeos estão representados nas cores vermelho e amarelo e os domínios com cores translúcidas.

Após a modelagem estrutural dos peptídeos encontrados foi feita uma busca e seleção de peptídeos em bloco conservado para todas as espécies a fim de obter possíveis candidatos a vacina, ilustrados na figura 14, os principais epítipos encontrados estão nas regiões 38-46, 218-225, 281-288, 308-321 e mostram baixa quantia de variações.

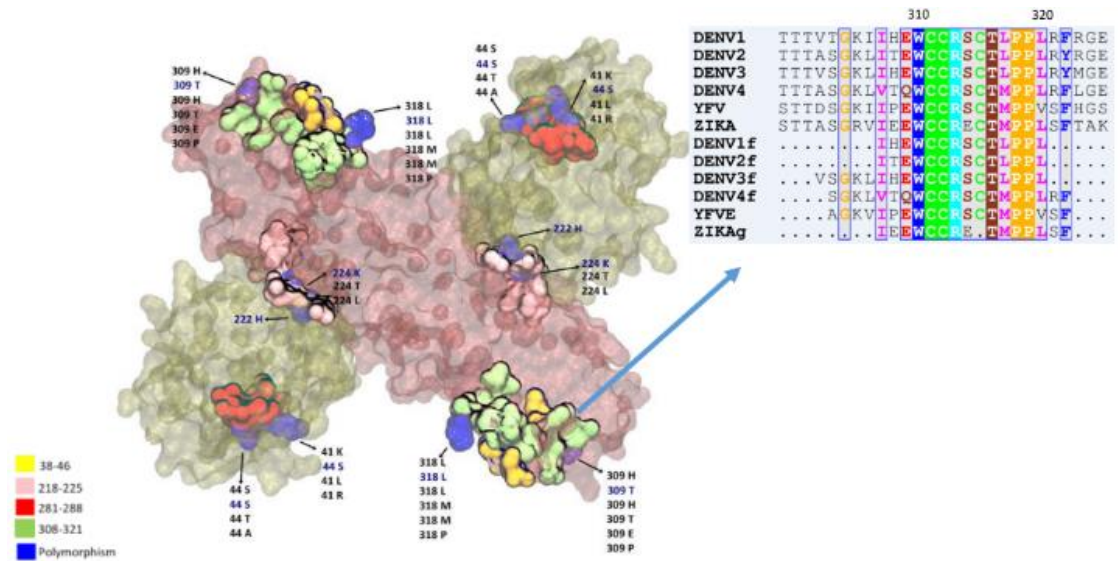


Figura 13- Modelagem dos peptídeos imunogênicos conservados entre as espécies. Epítomos conservados entre as espécies, cada cor representa um epítomo, sendo a região indicada a esquerda da imagem, os pontos em azul mostram variações dentro de um peptídeo. No alinhamento a direita da imagem está alinhado o epítomo da região 308-321, as seis primeiras sequências do alinhamento são as sequências consenso, logo em seguida as espécies estão acompanhadas de uma letra que indica a nomenclatura dada ao epítomo.

Os epítomos com menos variações foram encontrados nas regiões 281-288 e 218-225, apesar de poucas variações a sequência peptídica destes epítomos foi pequena. O melhor epítomo encontrado foi considerado da 308-321, já que além de mostrar sequência grande e bem exposta também mostrou a maior parte das variações como apenas físicas, não alterando a característica química do aminoácido.

4.9 Conclusão e perspectivas futuras

Obtivemos a topologia completa da NS1 para todas as espécies abordadas, o que ainda não existia na literatura, a partir da modelagem foi possível encontrar epítomos conformacionais para células B.

A partir da busca por epítomos encontramos 4 estruturas candidatas a uma vacina pan-flavivírus, para que o estudo seja atualizado e possa continuar em testes de bancada é necessário que ocorra uma atualização das sequências de NS1 depositadas no GenBank de 2018 a 2021 a fim de encontrar novas variações.

Com as regiões variáveis apontadas é possível o estudo de sequências para a criação de um teste diagnóstico diferencial entre as seis espécies de *Flavivirus* trabalhadas, evitando ao máximo testes com resposta cruzada.

Fica concluído que a modelagem molecular, assim como a bioinformática são ótimas ferramentas para se usar antes de ir para a bancada, dado que a quantidade de dados obtidas é tão boa quanto a qualidade destes dados.

Para perspectivas futuras podem ser estruturados diversos projetos, como:

- Publicação dos trabalhos gerados até o momento em revistas científicas especializadas e apresentação em eventos.
- Uso do mapeamento imunogênico para gerar plataformas vacinais pan-*Flavivirus* e plataformas de diagnóstico intra e inter específicos.
- Publicação das estruturas tridimensionais com alta qualidade produzidas somada ao seu uso para trabalhos de virtual *screening* elucidando ligações com estruturas e anticorpos.
- Construção de uma chimera vacinal pan-flavivírus com os principais epitopos selecionados no trabalho, prosseguindo com testes em bancada para a validação da efetividade da estrutura.

Alguns trabalhos partindo das análises feitas na monografia foram iniciados, como a seleção dos melhores peptídeos imunogênicos de células T e sua caracterização físico-química a fim de analisar as diferenças entre os epítomos estruturalmente.

5. Referências

- © 1986 Nature Publishing Group. 1986.
- AKEY, D. L. et al. association and regions of interaction with the immune system. v. 343, n. 6173, p. 881–885, 2015.
- ASNIS, D. S.; CRUPI, R. Flaviviridae. **Emerging Neurological Infections**, p. 239–260, 2005.
- BAKER, C. M. Polarizable force fields for molecular dynamics simulations of biomolecules. **Wiley Interdisciplinary Reviews: Computational Molecular Science**, v. 5, n. 2, p. 241–254, 2015.
- BENSON, D. A. et al. GenBank. v. 38, n. November 2009, p. 46–51, 2010.
- Bioinformática : da Biologia à Flexibilidade Molecular**. [s.l: s.n.].
- BRASIL. MINISTÉRIO DA SAÚDE. Monitoramento dos casos de arboviroses urbanas transmitidas pelo Aedes. **Boletim Epidemiológico Arboviroses**, v. 51, n. 24, p. 1–13, 2020.
- CALIS, J. J. A. et al. Properties of MHC Class I Presented Peptides That Enhance Immunogenicity. v. 9, n. 10, 2013.
- CHEN, V. B. et al. research papers MolProbity : all-atom structure validation for macromolecular crystallography research papers. p. 12–21, 2010.
- CHERAGA, H. A. A. S. annealing and an off-lattice united-residue force field : Application to the 10-55 fragment of staphylococcal protein A and to apo calbindin D9K. v. 96, n. March, p. 2025–2030, 1999.
- DOYTCHINOVA, I. A.; FLOWER, D. R. VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. **BMC Bioinformatics**, v. 8, p. 1–7, 2007.
- EDELING. Structural basis of Flavivirus NS1 assembly and antibody recognition. v. 111, n. 11, p. 4285–4290, 2014.
- EDSALL, J. T. Erwin brand. n. December, 1946.
- ELISA, R. Características gerais e epidemiologia dos arbovírus emergentes no Brasil. v. 55, n. 43, p. 55–64, 2014.
- GARCIA-BORONAT, M. et al. PVS : a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. n. July, p. 35–41, 2008.

GETZOFF, E. D.; TAINER, J. A.; LERNER, A. N. D. R. A. *The Chemistry and Mechanism of Antibody Binding to Protein Antigens*. 1988.

GLASNER, D. R. et al. The Good, the Bad, and the Shocking: The Multiple Roles of Dengue virus Nonstructural Protein 1 in Protection and Pathogenesis. **Annual Review of Virology**, v. 5, n. 1, p. 227–253, 2018.

GONÇALVES, R. L. et al. Dynamic behavior of Dengue and ZIKV víruses NS1 protein reveals monomer – monomer interaction mechanisms and insights to rational drug design monomer – monomer interaction mechanisms and insights to rational drug design. **Journal of Biomolecular Structure and Dynamics**, v. 0, n. 0, p. 1–11, 2019.

HERTZ, T. et al. HHS Public Access. v. 198, n. 10, p. 4025–4035, 2018.

HILGENFELD, R. with many faces. v. 35, n. 24, p. 2631–2633, 2016.

JAYATHILAKA, D. et al. secondary dengue infection. **Nature Communications**, n. 2018, [s.d.].

JIANG, L. et al. Selection and identification of B-cell epitope on NS1 protein of Dengue virus type 2. **Virus Research**, v. 150, n. 1–2, p. 49–55, 2010.

KAM, Y. et al. ZIKV-Specific NS1 Epitopes as Serological Markers of Acute ZIKV Vírus Infection. v. 220, p. 203–212, 2019.

KASHIWABARA, A. Y. MYOP: um arcabouço para predição de genes {\it ab initio}. 2007.

KO, J. et al. GalaxyWEB server for protein structure prediction and refinement. v. 40, n. May, p. 294–297, 2012.

KUMAR, S. et al. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. v. 35, n. May, p. 1547–1549, 2018.

LI, Y. et al. ZIKV and Flaviviruses Phylogeny Based. v. 36, n. 2, p. 1–8, 2017.

MENEZES, G. D. L. et al. Mutation of critical residues reveals insights of yellow fever vírus nonstructural protein 1 (NS1) stability and its formation. **Journal of Biomolecular Structure and Dynamics**, v. 0, n. 0, p. 1–10, 2019.

MUKHOPADHYAY, S.; KUHN, R. J.; ROSSMANN, M. G. A structural perspective of the Flavivirus life cycle. **Nature Reviews Microbiology**, v. 3, n. 1, p. 13–22, 2005.

NAGAR, P. K.; SAVARGAONKAR, D.; ANVIKAR, A. R. Detection of

Dengue virus-Specific IgM and IgG Antibodies through Peptide Sequences of Envelope and NS1 Proteins for Serological Identification. v. 2020, 2020.

NEUFELDT, C. J. et al. Rewiring cellular networks by members of the Flaviviridae family. **Nature Reviews Microbiology**, v. 16, n. 3, p. 125–142, 2018.

PONOMARENKO, J. et al. ElliPro: A new structure-based tool for the prediction of antibody epitopes. **BMC Bioinformatics**, v. 9, p. 1–8, 2008.

PUERTA-GUARDO, H. et al. Flavivirus NS1 Triggers Tissue-Specific Vascular Endothelial Dysfunction Reflecting Disease Tropism. **Cell Reports**, v. 26, n. 6, p. 1598- 1613.e8, 2019.

REYNOLDS, C. J. et al. T cell immunity to ZIKV vírus targets immunodominant epitopes that show cross-reactivity with other Flaviviruses. n. December 2017, p. 1–12, 2018.

RICE, P.; LONGDEN, I.; BLEASBY, A. EMBOSS: The European Molecular Biology Open Software Suite. **Trends in Genetics**, v. 16, n. 6, p. 276–277, jun. 2000.

RIVINO, L. et al. Differential Targeting of Viral Components by CD4+ versus CD8+ T Lymphocytes in Dengue virus Infection. **Journal of Virology**, v. 87, n. 5, p. 2693–2706, 2013.

ROY, A.; KUCUKURAL, A.; ZHANG, Y. I-TASSER : a unified platform for automated protein structure and function prediction. v. 5, n. 4, p. 725–738, 2010.

SILVA, V. L.; FERNADEZ, J. C. C.; TEVA, A. Capítulo 1: Imunologia. **Conceitos e Métodos para a Formação de Profissionais em Laboratórios de Saúde**, v. 1, p. 19–124, 2010.

SIMMONDS, P. et al. ICTV vírus taxonomy profile: Flaviviridae. **Journal of General Virology**, v. 98, n. 1, p. 2–3, 2017.

SOMNUKE, P. et al. N-linked glycosylation of Dengue virus NS1 protein modulates secretion, cell-surface expression, hexamer stability, and interactions with human complement. **Virology**, v. 413, n. 2, p. 253–264, 2011.

SONG, H. et al. BRIE F COMMUNICATIONS ZIKV vírus NS1 structure reveals diversity of electrostatic surfaces among Flaviviruses. v. 23, n. 5, p. 3–6, 2016.

TAN, T. Y. et al. Capsid protein structure in ZIKV vírus reveals the fl avivírus assembly process. **Nature Communications**, n. 2020, p. 1–13, [s.d.].

THOMPSON, J. D.; HIGGINS, D. G.; GIBSON, T. J. CLUSTAL W :

improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. v. 22, n. 22, p. 4673–4680, 1994.

THORNTON, J. M. et al. Location of “continuous” antigenic determinants in the protruding regions of proteins. **The EMBO journal**, v. 5, n. 2, p. 409–413, 1986.

TOVCHIGRECHKO, A.; VAKSER, I. A. GRAMM-X public web server for protein – protein docking. v. 34, p. 310–314, 2006.

VITA, R. et al. The Immune Epitope Database (IEDB): 2018 update. **Nucleic Acids Research**, v. 47, n. D1, p. D339–D343, 2019.

WANG, G.; JR, R. L. D. BIOINFORMATICS APPLICATIONS NOTE PISCES : a protein sequence culling server. v. 19, n. 12, p. 1589–1591, 2003.

WOLD, S. et al. DNA and peptide sequences and chemical processes multivariately modelled by principal component analysis and partial least-squares projections to latent structures. v. 217, p. 239–253, 1993.

XU, D.; ZHANG, Y. Improving the physical realism and structural accuracy of protein models by a two-step atomic-level energy minimization. **Biophysical Journal**, v. 101, n. 10, p. 2525–2534, 2011.

ZANOTRO, P. M. D. E. A. et al. Population dynamics of Flaviviruses revealed by molecular phylogenies. v. 93, n. January, p. 548–553, 1996.