



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE CIÊNCIAS
DEPARTAMENTO DE BIOQUÍMICA E BIOLOGIA MOLECULAR
BACHARELADO EM BIOTECNOLOGIA

IAN NUNES VALENÇA

ANÁLISE COMPUTACIONAL DAS INTERAÇÕES DA LACTOFERRINA BOVINA
COM PROTEÍNAS DO ZIKA VÍRUS

FORTALEZA

2017

IAN NUNES VALENÇA

ANÁLISE COMPUTACIONAL DAS INTERAÇÕES DA LACTOFERRINA BOVINA
COM PROTEÍNAS DO ZIKA VÍRUS

Monografia apresentada ao Programa de Graduação em Biotecnologia do Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharel em Biotecnologia.

Orientador: Dr. João Hermínio Martins da Silva

FORTALEZA

2017

IAN NUNES VALENÇA

ANÁLISE COMPUTACIONAL DAS INTERAÇÕES DA LACTOFERRINA COM
PROTEÍNAS DO ZIKA VÍRUS

Monografia apresentada ao Programa de Graduação em Biotecnologia do Departamento de Bioquímica e Biologia Molecular da Universidade Federal do Ceará, como requisito parcial à obtenção do título de bacharel em Biotecnologia.

Aprovada em: 17 /11 / 2017.

BANCA EXAMINADORA

Dr. João Hermínio Martins da Silva
Fundação Oswaldo Cruz – Ceará (Fiocruz-CE)

Prof.^a. Dr.^a Daniele de Oliveira Bezerra de Souza
Universidade Federal do Ceará (UFC)

Me. Disraeli Cavalcante Araújo Vasconcelos
Fundação Oswaldo Cruz – Ceará (Fiocruz – CE)

A Deus.

Aos meus pais, Virgínia e Tião,
Familiares e Amigos.

AGRADECIMENTOS

À Fundação Oswaldo Cruz Ceará pelos recursos cedidos para desenvolver o meu trabalho.

Ao meu orientador, Dr. João Hermínio, por ter me aceito como aluno e me proporcionado um aprendizado que levarei para a vida.

Ao Me. Disraeli por sempre estar disposto a ajudar e por ter aceito prontamente o meu convite para compor a minha banca.

A prof.^a Daniele por ter sido sempre uma professora muito compreensiva e competente, e por ter me dado grande satisfação em fazer parte da minha banca.

À coordenação do curso de Biotecnologia, que sempre faz o possível e o impossível pelos alunos.

À minha família, que é meu bem mais precioso.

Aos meus pais, que sempre me apoiaram e providenciaram tudo que foi preciso para que eu trilhasse meu caminho até aqui.

As minhas tias Judith, Zélia, Glorinha e Bela que abriram as portas das suas casas para mim e me deram suporte durante todo período da graduação.

Aos meus amigos da Biotecnologia sem exceção, todos contribuíram de uma forma ou de outra para o meu aprendizado e vivência na universidade, em especial Bia Ripardo, Karol, Danilo, Isabel e Jorge com quem eu dividi os momentos mais difíceis, e, sobretudo, com quem compartilhei os melhores momentos.

Aos meus amigos do Intercâmbio, em especial Fernanda, Roberta, Catarina e Pedro, que viveram comigo muitos desafios e conquistas e agora fazem parte da minha família.

“Seja mais forte que a sua melhor desculpa. ”

(Lucas Colaço)

RESUMO

O vírus da Zika, pertencente à família *Flaviviridae*, foi responsável por um surto nas américas que causou uma preocupação global de que pudéssemos estar à beira de uma crise de saúde. A falta de pesquisa sobre o Zika vírus durante os mais de 60 anos nos quais se conhece o vírus nos deixou com poucas opções no combate a essa recente epidemia. Pesquisadores O foco do presente trabalho se deteve em análises computacionais dos cristais de três proteínas não-estruturais (NS1, NS3, NS5), uma proteína estrutural do Zika vírus (Envelope) e da lactoferrina bovina, uma proteína encontrada no leite que recentemente demonstrou um forte poder inibitório em testes virais com o Zika vírus. Dessa forma visou-se avaliar, por meio da técnica do *docking* molecular e do *alanine scanning*, como a lactoferrina interagia com essas proteínas virais. Inicialmente, os sítios de ligação das proteínas virais foram identificados por meio do programa *metaPocket* e as melhores cavidades foram escolhidas para o *docking*. O *docking* molecular foi realizado com o auxílio do programa *HADDOCK* e as interações resultantes entre os sítios de ligação das proteínas virais com a lactoferrina foram analisadas. Os complexos ainda foram analisados quanto a formação de ligações de hidrogênio e pontes salinas através do servidor *PDBePISA*. Dentre os complexos, aquele formado entre a lactoferrina e a proteína envelope demonstrou os melhores resultados em comparação com os demais, o que torna a proteína envelope um provável alvo da lactoferrina dentro das células *in vivo*.

Palavras-chave: Zika vírus, *Docking* molecular, Lactoferrina.

ABSTRACT

The Zika virus, belonging to the *Flaviviridae* family, was responsible for an outbreak in the Americas causing a global concern that we could be on the verge of a health crisis. The lack of research on the Zika virus during more than 60 years in which the virus is known has left us with few options in the fight against this recent epidemic. The present work focused on computational analyzes of four Zika virus proteins: three non-structural proteins (NS1, NS3, NS5), a structural protein (Envelope); and bovine lactoferrin, a protein found in milk that recently demonstrated a strong inhibitory effect in viral tests with the Zika virus. In this way, the aim was to evaluate, through the technique of molecular docking and alanine scanning, how lactoferrin interacted with these viral proteins. Initially, viral protein binding sites were identified through the metaPocket program and the best pockets were chosen for docking. Molecular docking was performed using the HADDOCK program and the resulting interactions between the viral protein binding sites with lactoferrin were analyzed. The complexes were also analyzed for the formation of hydrogen bonds and salt bridges through the PDBePISA server. Among the complexes, the one formed between lactoferrin and envelope protein showed the best results in comparison to the others, which makes the envelope protein a probable target of lactoferrin within the cells in vivo.

Keywords: Zika virus, Molecular Docking, Lactoferrin

LISTA DE FIGURAS

Figura 1 – Ciclo de vida de mosquitos do gênero <i>Aedes</i>	13
Figura 2 – Representação das proteínas NS1, NS3 e NS5 em cadeias.....	18
Figura 3 – Representação das proteínas envelope e lactoferrina em cadeias.....	29
Figura 4 – Perfil eletrostático das proteínas NS1, NS3 e NS5.....	31
Figura 5 – Perfil eletrostático das proteínas envelope e lactoferrina.....	33
Figura 6 – Representação das cavidades da proteína NS1.....	34
Figura 7 – Representação das cavidades da proteína NS3.....	36
Figura 8 – Representação das cavidades da proteína NS5.....	37
Figura 9 – Representação das cavidades da proteína envelope.....	38
Figura 10 – Complexo LACTF- NS1 pós- <i>docking</i>	41
Figura 11 – Complexo LACTF- NS3 pós- <i>docking</i>	43
Figura 12 – Complexo LACTF- NS5 pós- <i>docking</i>	44
Figura 13 – Complexo LACTF- envelope pós- <i>docking</i>	45

SUMÁRIO

1	INTRODUÇÃO	12
1.1	Caracterização do problema.....	12
1.2	Zika Vírus.....	13
1.2.1	<i>Epidemiologia</i>	13
1.2.3	<i>Patogênese</i>	16
1.2.4	<i>Sintomas</i>	17
1.3	Lactoferrina.....	18
1.4	Abordagem computacional.....	19
1.4.1	<i>Docking molecular</i>	19
1.4.2	<i>Escaneamento computacional de Alanina ponto-a-ponto de interfaces proteína-proteína</i>	20
2	OBJETIVOS	21
2.1	Objetivo geral.....	21
2.2	Objetivos específicos.....	21
3	METODOLOGIA	22
3.1	Perfil eletrostático das proteínas virais e Lactoferrina.....	22
3.2	<i>Docking</i> computacional proteína-proteína.....	22
3.2.1	<i>Identificação dos resíduos ativos da lactoferrina e proteínas virais</i>	23
3.2.2	<i>Identificação dos resíduos passivos da lactoferrina e proteínas virais</i>	24
3.2.3	<i>HADDOCK</i>	24
3.3	Mutagênese <i>in silico</i> : <i>Alanine scanning</i>	26
4	RESULTADOS E DISCUSSÃO	28
4.1	Análise dos cristais das proteínas virais e lactoferrina.....	28
4.2	Análise do perfil eletrostático das proteínas virais do Zika vírus e da lactoferrina.....	31
4.3	Determinação dos sítios de ligação dos cristais das proteínas do ZIKV... ..	34
4.3.1	<i>Cavidades encontradas para o cristal 5YI3 da proteína NS1</i>	34
4.3.2	<i>Cavidades encontradas para o cristal 5KQR da proteína NS3</i>	37
4.3.3	<i>Cavidades encontradas para o cristal 5JHM da proteína NS5</i>	38
4.3.4	<i>Cavidades encontradas para o cristal 5JMT da proteína NS3</i>	38

4.4	Identificação dos resíduos ativos da lactoferrina e das proteínas virais..	39
4.5	<i>Docking</i> entre a lactoferrina e as proteínas virais.....	40
4.3	Escaneamento computacional de alanina ponto-a-ponto dos complexos <i>dockados</i>	45
5	CONCLUSÃO	46
6	REFERÊNCIAS	47

1 INTRODUÇÃO

1.1 Caracterização do problema

Zika vírus (ZIKV), um vírus da família *flaviviridae* (gênero *Flavivirus*), atualmente está causando uma epidemia nas Américas de proporções nunca antes vistas. Historicamente, o ZIKV se espalhou a partir da África equatorial e Ásia para as ilhas do Pacífico, durante meados do ano 2000 até o início de 2010, invadindo a América Central e do Sul em 2015 e alcançando a América do Norte em 2016. Apesar da infecção pelo ZIKV geralmente não apresentar sintomas e ser considerada uma doença branda, ela foi recentemente relacionada com o aumento de casos de doenças neurológicas como a microcefalia e a síndrome de Guillain-Barré, que afetam principalmente os recém-nascidos. (NUGENT *et al*, 2017) Devido a contínua expansão geográfica do vírus e do seu mosquito vetor, ZIKV impõe uma séria ameaça à saúde pública ao redor do mundo. Contudo, não existem vacinas ou terapias antivirais disponíveis até o momento e abordagens computacionais como as utilizadas neste trabalho tem o papel de auxiliar no processo de descobrimento de novos fármacos ou proteínas terapêuticas.

1.2 Zika Vírus

O ZIKV é um arbovírus constituído por um RNA de fita simples de senso positivo pertencente à família *Flaviviridae*, gênero *Flavivirus*. Este vírus está relacionado com os vírus da dengue, da febre amarela, do Oeste do Nilo e da Encefalite Japonesa. Seu genoma viral possui 10.794 nucleotídeos que codificam 3.419 aminoácidos que resultam em três proteínas estruturais (C, PrM/M e E) e sete proteínas não estruturais (NS1, NS2A, NS2B, NS3, NS4A, NS4B e NS5) (ABUSHUK *et al*, 2016). O vírion possui aproximadamente 40nm de diâmetro com 5~10nm nas projeções da superfície. Foi primeiramente isolado de um macaco rhesus em Uganda em 1947 e pode ser encontrado em mosquitos, primatas não-humanos e humanos. O ZIKV apresenta duas principais linhagens: uma linhagem Asiática e outra Africana. Inicialmente casos em humanos eram esporádicos e geralmente se tratava de uma enfermidade branda. Contudo, grandes epidemias ocorreram na ilha Yap, na Micronésia, com 59 casos confirmados em 2007, na Polinésia Francesa com 28.000 casos (11% da população) em 2013, e no Brasil em 2015. Em fevereiro de 2016 um comitê da Organização Mundial da Saúde (OMS) declarou que o Zika vírus e as desordens neurológicas associadas a ele representavam uma

ameaça à saúde pública de preocupação internacional. Em um mesmo relatório, a OMS relatou que o Ministério da Saúde no Brasil estimava que entre 497.593 e 1.482.701 de casos de infecção por Zika ocorreram desde o final de 2014. Os quadros clínicos associados com essa infecção mudaram significativamente e agora incluem múltiplos casos de Guillain-Barré na Polinésia Francesa e Colômbia e microcefalia no Brasil. (NUGENT *et al*, 2017)

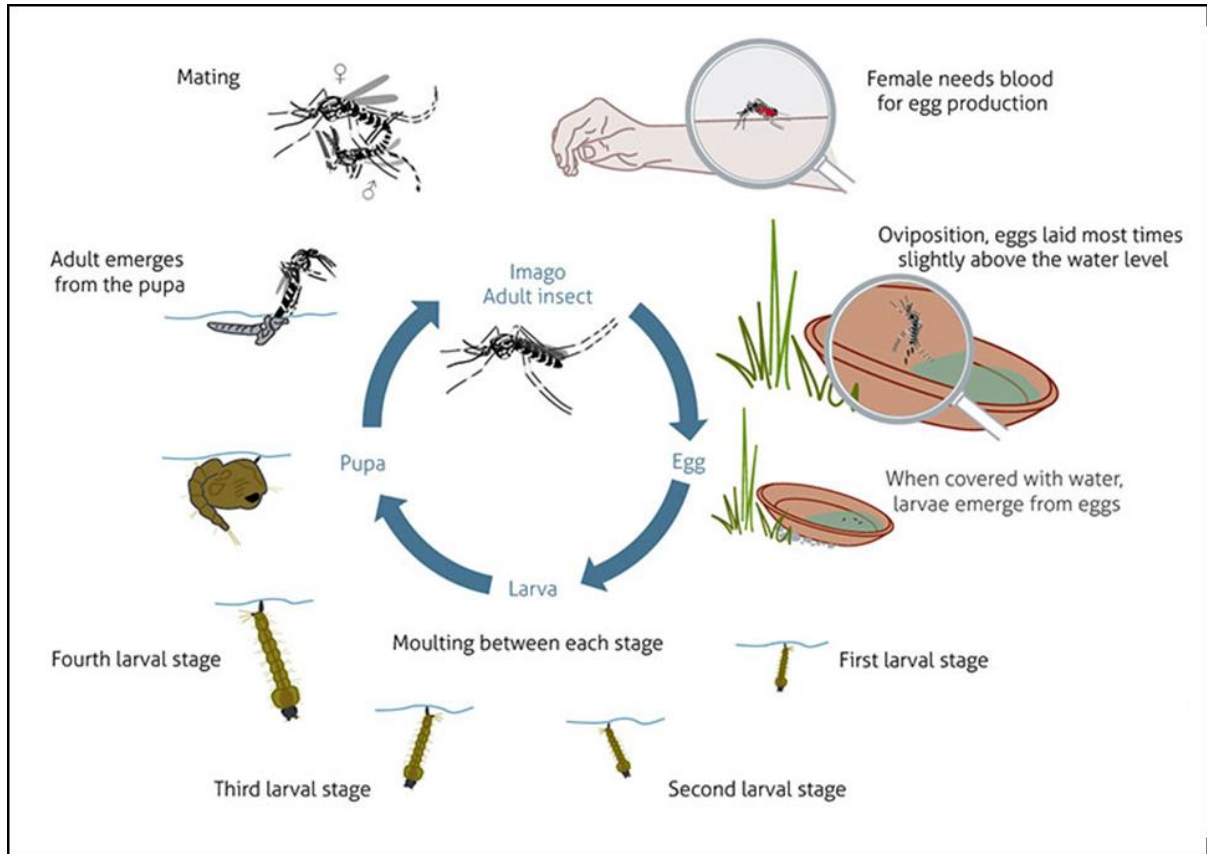
1.2.1 Epidemiologia

O vírus foi primeiro identificado nas Américas em março de 2015, quando uma epidemia começou no Brasil. O Ministério da Saúde então estimou 1.3 milhão de casos suspeitos de infecção ocorridos até dezembro do mesmo ano. O ZIKV foi identificado na Colômbia em outubro de 2015 e, até março de 2016, 50.000 casos de infecção foram contabilizados no país (PETERSEN *et al*, 2016). Nesse mesmo período o vírus já podia ser encontrado em pelo menos 33 países e territórios nas Américas. A explicação para essa rápida disseminação é baseada em vários fatores, que incluem o aumento da população mundial, o crescimento populacional de mosquitos viáveis como vetores e o aumento do transporte de mercadorias entre os países que permite a migração de mosquitos de uma região para outra antes inacessível. A falta de imunidade nas populações dessas regiões também contribui para o desenvolvimento de epidemias. Além disso, reações cruzadas com vírus relacionados, como o vírus da dengue, e mutações no genoma do Zika podem contribuir para a frequência e severidade clínica dessas infecções. Um estudo sobre o genoma do ZIKV de várias regiões sugeriu que houve uma única introdução do vírus nas Américas entre maio e dezembro de 2013 (FARIA *et al*, 2016). Uma análise das mutações do genoma viral e identificou quatro substituições de aminoácidos envolvendo a região do receptor viral, a proteína envelope viral e a proteína M. Essas mutações aumentaram o potencial de atracamento viral, entrada nas células, maturação e saída das células e explica parcialmente o aumento da disseminação e mudanças na patogênese associadas com essa infecção. ZIKV infecta alguns, mas não todos os mosquitos *Aedes*, e a viabilidade do vetor tem um importante papel nessas epidemias (PETERSEN *et al*, 2016).

O ciclo de vida do mosquito consiste em quatro fases. Mosquitos fêmeas adultos depositam seus ovos dentro de recipientes contendo água; os ovos são resistentes e podem permanecer viáveis por 8 meses antes de ressecarem completamente. Os ovos eclodem dando origem às larvas, que se transformam em pupas e posteriormente em mosquitos adultos. Quando a fêmea adulta se alimenta de sangue ela deposita seus ovos e o ciclo recomeça. O ciclo de vida

completo leva em torno de 8-10 dias em média, mas pode variar de acordo com as condições ambientais, como temperatura e disponibilidade de água.

Figura 1 – Ciclo de vida de mosquitos do gênero *Aedes*.



Fonte: CDC (Centers for Disease Control and Prevention).

O ZIKV é primariamente transmitido à humanos através da picada de um mosquito *Aedes* fêmea infectado. O vírus se espalha quando um mosquito pica uma pessoa infectada. O vírus permanece ativo no interior do mosquito e é transmitido para outras pessoas através de subsequentes picadas. Conseqüentemente, a taxa de infecção em humanos irá depender do tamanho da população de mosquitos, que por sua vez depende de múltiplos fatores ambientais, e as condições ideais geralmente variam entre as espécies de mosquitos. Essa dinâmica complexa sugere que existe uma variabilidade significativa entre a taxa de infecção de mosquitos e a taxa de transferência para o próximo humano hospedeiro. Guzzeta et al desenvolveram um modelo que estima a probabilidade de uma epidemia, seus cálculos indicaram que a variável crítica é a probabilidade de um mosquito ficar infectado após picar um hospedeiro virêmico. O potencial para uma epidemia aumenta quando essa probabilidade excede 25% (GUZZETA *et al*, 2016).

O ZIKV foi isolado de várias espécies de *Aedes*, e a transmissão desse vírus pelos mosquitos *Aedes aegypti* e *Aedes albopictus* é a maior preocupação nos Estados Unidos; a distribuição dos mosquitos *Aedes* é crítica para o desenvolvimento de epidemias, e o sistema de viagem moderno permite o movimento de mosquitos para novos lugares. Ambas espécies, *aegypti* e *albopictus*, foram responsáveis pelas epidemias causadas por outros arbovírus e são bem estabelecidas do Brasil aos Estados Unidos, além de serem consideradas agressivas e preferirem viver próximas à pessoas. Chouin-Carneiro estudou a suscetibilidade dos mosquitos *Aedes aegypti* e *Aedes albopictus* ao Zika vírus. As condições experimentais para esse estudo envolveram exposição dos mosquitos a sangue infectado e um periódico sacrifício dos mosquitos nos dias 4, 7 e 14 para determinar se a infecção tinha ou não ocorrido e em qual parte o vírus havia se alojado no corpo dos mosquitos. Nenhum vírus foi encontrado na saliva durante os dias quatro e sete após a exposição. No dia 14, entretanto, a taxa de infecção dos mosquitos *Aedes aegypti* foi de 76.7%, a taxa de disseminação para o abdômen e tórax foi de 60.7% e recuperados da saliva foi de 21.4% (CHOUIN *et al*, 2016). Esse estudo da viabilidade do vetor indica que a taxa de infecção varia de uma espécie de mosquito para outra e varia em dentro de uma particular espécie dependendo de sua origem geográfica. Di Luca *et al* também estudou infecções em duas espécies de mosquitos *Aedes*. Taxas de Infecção, disseminação, e transmissão forem maiores em mosquitos *Aedes aegypti* do que em *Aedes albopictus* (DI LUCA *et al*, 2016).

O ZIKV também pode ser transmitido de várias outras formas além do ciclo mosquito-humano-mosquito atualmente ocorrendo nos ambientes urbanos nas Américas. Essas formas incluem transmissão sexual e transmissão materno-fetal. Partículas virais e RNA do Zika vírus foram detectados em esperma vários dias após o aparecimento dos sintomas, e a transmissão foi documentada principalmente por homens com sintomas no momento ou logo depois do ato sexual.

D'Ortenzio *et al* relataram a transmissão do vírus da Zika de um homem para uma mulher em uma cidade na França sem uma população estabelecida de mosquitos *Aedes*. ZIKV foi isolado nos dias 18 e 24 após uma viagem ao Brasil. A mulher teve resultados positivos do ZIKV em amostras de saliva e urina no terceiro dia após o aparecimento dos sintomas (D'ORTENZIO *et al*, 2016). Harrower *et al* também relataram transmissão sexual do ZIKV. O vírus foi identificado no sêmen 76 dias após o aparecimento dos sintomas (HARROWER *et al*, 2016). Conseqüentemente, consistente e cuidadoso uso de camisinha ou abstinência por longos períodos seriam necessários para prevenir a transmissão sexual após a infecção. Esse modo de transmissão não foi relatado em outras infecções por arbovírus, e os diagnósticos em regiões

com co-circulação dos vírus da dengue e Chikungunya são difíceis uma vez que todos os três vírus podem causar quadros clínicos parecidos.

Embora nenhum caso de transmissão através do leite materno, urina ou saliva tenham sido relatados, o leite materno em mães infectadas tem altos níveis de partículas virais e vírus viáveis foram isolados de urina/saliva de pacientes infectados no Brasil. Esperma e óvulos preservados para futura utilização através de reprodução assistida também possuem riscos teóricos de transmissão uma vez que a criopreservação dificilmente destruirá os vírus, e atualmente nenhum teste tem sido realizado nesses produtos.

1.2.3 Patogênese

Existe informação limitada sobre a patogênese das infecções do Zika vírus. Lazear et al estudaram essa patogênese em ratos e demonstraram que os animais poderiam ser infectados com injeções subcutâneas, intraperitoneais e intravenosas do vírus. Ratos que produziram uma quantidade pequena de interferons alfa/beta e ratos que não possuíam receptores desenvolveram doenças neurológicas e morreram com ZIKV. Esses animais desenvolveram progressivas síndromes neurológicas que iniciavam com parcial fraqueza dos membros traseiros, seguida de parcial paralisia desses membros, completa paralisia dos membros traseiros e então completa paralisia de todos os quatro membros. O maior contingente viral foi encontrado no cérebro, medula espinhal e testículos (LAZEAR *et al*, 2016). Esse modelo animal é particularmente importante quando se considera as recentes observações em infecções humanas que incluem transmissões por contato sexual, desenvolvimento da síndrome de Guillain-Barré e o desenvolvimento da microcefalia congênita severa em recém-nascidos de mulheres infectadas com o ZIKV.

Acredita-se que a maioria dos arbovírus se replicam dentro das células dendríticas da pele do hospedeiro antes de se espalharem para os linfonodos regionais e posteriormente para a corrente sanguínea. Além disso, o vírus passa por um período de incubação intrínseca de 4 a 5 dias no hospedeiro humano, infectando outros insetos vetores durante a sucção sanguínea onde o ZIKV irá passar por um período de incubação extrínseca de 8 a 12 dias e depois se disseminará pela saliva desses vetores infectando outros hospedeiros (HAYES, 2009).

1.2.4 Sintomas

O quadro clínico da infecção por Zika pode variar desde condições assintomáticas, que ocorre em 80% dos casos, até um quadro febril que se assemelha aos sintomas da dengue com febre baixa, conjuntivite bilateral, erupções cutâneas, dor de cabeça, dor retro orbital e artrite ou artralgia com edema nas articulações das mãos e dos pés que podem durar de três a cinco dias, mas, em alguns pacientes, podem persistir ou recorrer por mais de 30 dias após o início da infecção. Existem quadros mais severos que foram relatados apresentando febre alta, arrepios e hipotensão. Além disso, sintomas relacionados ao trato digestório também podem estar presentes incluindo náusea, vômito, diarreia, constipação e dor abdominal. Todo esse quadro clínico anterior pode ser confundido com uma infecção por Chikungunya ou dengue. Alguns sintomas, no entanto, podem ajudar a distinguir a febre por Zika das demais, como edemas mais proeminentes nas extremidades, dores de cabeça menos severas e, em comparação com o Chikungunya por exemplo, as crises de artralgia não são muito intensas. Além disso, infecções por Zika, diferentemente dos quadros de dengue, não levam às complicações hemorrágicas. Contudo, complicações neurológicas como a síndrome de Guillain-Barré tem sido reportadas (MCCARTHY, 2016).

Apesar da infecção pelo vírus da Zika ser considerado uma doença de moderada intensidade e de baixa ou nenhuma letalidade, a maior preocupação que a atual epidemia do vírus levanta está relacionada com os casos de microcefalia em bebês nascidos de mães que foram infectadas com o ZIKV. A microcefalia é uma condição onde a cabeça do bebê é muito menor que o esperado. Durante a gravidez, a cabeça do bebê cresce à medida que seu cérebro se desenvolve, portanto, se algum dano ou falha de desenvolvimento cerebral ocorrer isso será refletido no tamanho de sua cabeça. Pesquisadores do Centro de controle e prevenção de doenças dos Estados Unidos anunciaram que evidências suficientes já foram acumuladas para se concluir que a infecção pelo Zika vírus durante a gravidez é uma das causas de microcefalia e outros defeitos cerebrais considerados severos em fetos (SCHIMDT, 2016).

1.3 Lactoferrina

Lactoferrina (Lf) é uma glicoproteína de 80-kDa de aproximadamente 700 aminoácidos da família das transferrinas que apresenta uma alta afinidade de ligação a íons de ferro. A Lf possui uma alta identidade entre as espécies e é considerada uma proteína multifuncional que é largamente distribuída no leite, assim como em outras secreções como saliva e lágrima. Ela é secretada pelos neutrófilos no sangue e tecidos inflamados. Lf age diretamente como um antimicrobiano, limita a proliferação e a adesão de micróbios (bactérias, vírus e parasitas) e/ou os elimina. Esse efeito da Lf é o resultado de sua habilidade de sequestrar ferro em fluidos biológicos e desestabilizar as membranas de microrganismos. Os metais aos quais a Lf se liga são os íons Fe^{2+} ou Fe^{3+} (HASTOGI *et al*, 2014).

Por causa de sua versatilidade, a Lactoferrina surgiu como um potencial medicamento contra as infecções causadas pelo ZIKV e Chikungunya. Em testes antivirais a Lactoferrina demonstrou um forte efeito inibitório dose-dependente tanto em células infectadas pelo ZIKV quanto por Chikungunya, alcançando uma inibição máxima de aproximadamente 80% utilizando uma concentração de 1.0 mg/ml com diferentes valores de IC_{50} . Foi ainda constatado que a atividade inibitória da proteína se deu em um estágio de pré-entrada, antes do vírus entrar nas células, possivelmente por ligação, mas também em um estágio de pós-entrada nas células, possivelmente afetando a replicação (CARVALHO *et al*, 2017).

1.4 Abordagem computacional:

1.4.1: Docking molecular

Para um melhor entendimento da função biológica da proteína, o conhecimento da sua estrutura tridimensional é crucial. A resolução das estruturas de proteínas é feita principalmente através de dois métodos: cristalografia de raio-X e ressonância magnética nuclear (RMN). De acordo com as estatísticas do PDB (<http://www.rcsb.org/pdb/>), aproximadamente 11655 estruturas por NMR e 110550 estruturas por raio-X foram resolvidas e depositadas até a presente data.

Embora muitos métodos estejam disponíveis para o estudo de complexos de proteínas, apenas algumas dessas técnicas providenciam informação de alta resolução em um nível atômico. Raio-X e RMN encontram dificuldades quando lidam com complexos de proteínas. No caso do raio-X, a dinâmica da formação do complexo faz a cristalização difícil, enquanto que a

limitação do tamanho na RMN é o maior problema quando consideramos complexos de alto peso molecular. Embora as técnicas de RMN permitam o mapeamento da interface proteína-proteína, o que permite que os aminoácidos que estão envolvidos na formação do complexo sejam identificados facilmente, ela não providencia nenhuma informação a respeito da orientação de uma proteína em relação a outra. Por conta disso, essa informação dificilmente se torna diretamente aproveitável no processo de cálculo da estrutura.

Uma abordagem mais recente em relação a esses métodos experimentais, e que já está bem estabelecida, se baseia no *docking* para o estudo de complexos protéicos a um nível estrutural. Existem agora programas que operam basicamente através de duas estratégias distintas: *ab initio* ou *docking* dirigido (DE VRIES; VAN DIJK; BONVIN, 2010). A primeira abordagem fixa a primeira molécula e rotacional a segunda em torno de seu centro de massa ou a gira em torno da primeira, assumindo diversas configurações ao longo do tempo, para as quais um escore é calculado com base em parâmetros como as interações eletrostáticas, de van der Waals e complementariedade de superfícies, por exemplo. Já o *docking* dirigido utiliza dados experimentais – obtidos a partir de RMN, por exemplo – para orientar o atracamento das duas moléculas. Este trabalho utilizou a segunda abordagem para descrever o modo de ligação entre a lactoferrina e as proteínas virais.

1.4.2 Escaneamento computacional de Alanina ponto-a-ponto de interfaces Proteína-Proteína

Interações proteína-proteína são componentes chave de todos os processos de transdução de sinal, mediando a integração de rotas lineares em complexas redes de interações reveladas pelos estudos genômicos. Ferramentas que racionalmente alteram e interferem nas interações proteicas oferecem grandes promessas em ajudar a dissecar a função de conectividades nessas redes. A habilidade em alterar as interações entre as proteínas requer um entendimento dos determinantes de afinidade e especificidade nas interfaces proteicas.

“*Alanine-scanning mutagenesis*” é um poderoso método para analisar importantes interações em interfaces proteína-proteína. O Escaneamento da alanina mede o efeito da deleção de uma cadeia lateral de um aminoácido, após o átomo de carbono C_{β} , na afinidade de um complexo proteína-proteína. Substituições individuais de muitos aminoácidos pela alanina gera um mapa de quais interações são críticas para a estabilidade do complexo e quais não são. Clarkson e Wells chamaram esses resíduos geneticamente importantes de “*hot*

spots” no pioneiro trabalho deles sobre a ligação do hormônio do crescimento humano com seu receptor, onde apenas uma pequena fração dos resíduos da interface contam para a maior parte da energia de ligação (KORTEMME *et al*, 2004). Subseqüentes estudos sugerem que a presença de “*hot spots*” comportando apenas uma fração área total da interface é uma propriedade geral da maioria dos complexos proteína-proteína.

2. OBJETIVOS

2.1 Objetivo Geral

Descrever as características estruturais e os potenciais sítios de ligação dos complexos resultantes do *docking* molecular entre a lactoferrina e as proteínas NS1, NS3, NS5 e GPE do Zika vírus e determinar os complexos de maior estabilidade.

2.1 Objetivos específicos

1. Selecionar as estruturas cristalizadas das proteínas não-estruturais NS1, NS3 e NS5, GPE e a Lactoferrina;
2. Definir as cavidades *drogáveis* dos cristais das proteínas alvo, bem como o perfil eletrostático das mesmas;
3. Realizar a simulação das interações *in silico* da Lactoferrina com as proteínas NS1, NS3, NS5 e GPE do Zika vírus utilizando o servidor HADDOCK no sentido de encontrar o modo de ligação entre essas estruturas;
4. Analisar os complexos de menor energia gerados pelo *docking*, bem como as interações entre as proteínas.

3. METODOLOGIA

3.1 Perfil eletrostático das proteínas virais e Lactoferrina

A habilidade de avaliar o perfil eletrostático de biomoléculas possibilita prever o comportamento de um ligante em uma região específica da proteína. Assim, o entendimento das interações eletrostáticas é essencial para o desenvolvimento de moléculas com potencial terapêutico.

O software APBS (*Adaptive Poisson-Boltzman Solver*) foi utilizado para a análise do perfil eletrostático predominante nas cavidades dos cristais das proteínas virais NS1, NS3, NS5 e Envelope e a da proteína Lactoferrina.

O APBS é um pacote de software que utiliza a plataforma do programa *Pymol* para a visualização de seus resultados. Ele calcula, por meio da equação de Poisson-Boltzmann (PBE), as propriedades eletrostáticas de moléculas com dezenas a milhões de átomos (BAKER *et al.*, 2001).

Para a análise dos cristais das proteínas foi estabelecido um campo de perfil eletrostático entre -3 e $+3kT/e$ (onde k representa a constante de Boltzmann, T a temperatura e e a energia), o que possibilitou uma melhor visualização dos perfis positivos e negativos.

Basicamente, as regiões da proteína que tendem ao espectro vermelho apresentam um perfil eletrostático predominantemente negativo enquanto as que apresentam uma tendência ao espectro azul são predominantemente positivas e as regiões em branco apresentam um perfil neutro.

3.2 Docking computacional proteína-proteína

Os testes se basearam em *dockings* proteína-proteína entre a Lactoferrina bovina e as proteínas NS1, NS3, NS5 e Glicoproteína E do Zika vírus feitos no programa HADDOCK versão 2.2, onde é necessário fazer o upload das estruturas em formato “pdb” das proteínas do vírus e da Lactoferrina individualmente e identificar os resíduos ativos – ou seja, aqueles que sabidamente fazem contato direto com a proteína parceira – e passivos de cada molécula. Tais informações são convertidas pelo servidor em Restrições Ambíguas de Interação (RAI), que direcionam todo o processo de *docking*: para cada resíduo ativo, uma RAI é definida entre este resíduo e todos os resíduos ativos e passivos da outra proteína, e todos os cálculos do processo

são desenvolvidos a partir de então (VRIES; VAN DIJK; BONVIN, 2010). Dessa forma, é necessário definir quais são os resíduos ativos e passivos de cada uma das proteínas envolvidas.

3.2.1 Identificação dos resíduos ativos da Lactoferrina e proteínas virais

Para o *docking* considerou-se como sendo ativos os resíduos identificados pelo servidor MetaPocket 2.0. Esse programa permite a predição de possíveis sítios de ligação na superfície da proteína estudada que, frequentemente, é o ponto inicial para a identificação da função de uma proteína e desenvolvimento de drogas baseadas em sua estrutura. Além disso, a detecção de sítios de ligação bem definidos é um pré-requisito para o *docking* proteína-proteína. Nas últimas décadas, vários métodos computacionais para a predição desses sítios de ligação foram desenvolvidos, e o MetaPocket 2.0 (HUANG, 2009) utiliza 4 deles: LIGSITE^{cs} (HUANG AND SCHROEDER, 2006) PASS (BRADY AND STOUTEN, 2000), Q-SiteFinder (LAURIE AND JACKSON, 2005) and SURFNET (LASKOWSKI, 1995). Estes programas combinados geram um aprimoramento da chance de sucesso da predição. O LIGSITE^{cs} seleciona os eventos superfície-solvente-superfície mais precisos usando a superfície Connolly da proteína (CONNOLLY, 1983) ao invés de selecionar eventos proteína-solvente-proteína. SURFNET e PASS também utilizam abordagens geométricas para a detecção de bolsos (a idéia central é que se existe uma esfera que separa dois átomos, mas que não contém nenhum átomo, esse espaço constitui um bolso, ou *pocket*). Primeiramente, uma esfera é posicionada de modo que dois determinados átomos estão posicionados de lados opostos da esfera. Se o interior da esfera contiver qualquer outro átomo, ela é reduzida em tamanho até que não contenha nenhum átomo. Somente esferas com um raio de 1 a 4 Å são mantidas. O resultado desse procedimento é um número de grupos separados de esferas interconectadas, chamados de regiões *gap*, que se situam tanto dentro como na superfície da proteína, e que correspondem as cavidades (Huang and Schroeder, 2006). PASS utiliza sondas esféricas para completar as cavidades camada por camada. Primeiramente, um revestimento inicial da proteína com sondas esféricas é calculado. Cada sonda tem uma contagem de profundidade que conta o número de átomos em um espaço de 8 Å de distância. Somente sondas com uma contagem acima de um limite são mantidas. Esse procedimento é repetido até que uma camada não produza nenhuma sonda inserida em cavidades. Então cada sonda é assinalada com um peso que é proporcional ao número de esferas na vizinhança e a sua profundidade. Um pequeno número sítios ativos (ASP) são então selecionados através da identificação de sondas centrais que contêm muitas esferas com uma

contagem de profundidade elevada. Finalmente, os sítios ativos que foram escolhidos são ordenados de acordo com o peso das sondas. O último método utilizado, o Q-SiteFinder (LAURIE AND JACKSON, 2005), complementa os cálculos com uma abordagem energética, ao invés de geométrica como os outros três métodos anteriores. No Q-SiteFinder a superfície da proteína é coberta com uma camada de sondas de metil ($-\text{CH}_3$) de modo a calcular as interações de energias de van der Waals entre a proteína e as sondas. As sondas com interações energéticas favoráveis são mantidas, e os aglomerados dessas sondas são ordenados de acordo com o número de sondas em um único aglomerado. Os grupos maiores e mais energeticamente favoráveis são ordenados nas primeiras colocações e são considerados potenciais sítios de ligação.

3.2.2 Identificação dos resíduos passivos da Lactoferrina e proteínas virais

Resíduos passivos de uma proteína são aqueles que potencialmente farão contato direto com a outra proteína. Apesar de, em teoria todos os resíduos de uma proteína poderem ser considerados passivos, neste trabalho economizou-se tempo computacional considerando-se como passivos apenas os resíduos que estejam expostos em um raio de até 6.5 Å de qualquer resíduo ativo – encontrados anteriormente. Esse procedimento é feito automaticamente pelo servidor HADDOCK 2.0 se o usuário selecionar esta opção na *Easy Interface*. (VRIES; VAN DIJK; BONVIN, 2010).

3.2.3 HADDOCK

No HADDOCK, os dados experimentais são inseridos na forma de resíduos ativos ou passivos que são convertidos pelo programa em *Ambiguous Interaction Restraints* que irão guiar o *docking*. O protocolo do *docking* irá consistir de três estágios, que são: uma randomização de orientações e minimização de energia de corpo-rígido, um refinamento semi-flexível na torsão do ângulo espacial e um refinamento final no solvente explícito. Após cada uma dessas etapas as estruturas recebem uma pontuação e são ranqueadas e, então, as melhores estruturas são escolhidas para a próxima etapa. O HADDOCK *score* é uma soma ponderada das energias de van der Waals, eletrostáticas, dessolvatação e violação de restrição junto com a área superficial não-exposta.

No estágio de randomização, as duas proteínas são posicionadas a 150 Å de distância uma da outra e cada proteína é aleatoriamente rotacionada ao redor dos seus centros de massa. A minimização de energia de corpo rígido é então executada: primeiro, quatro ciclos de otimização da orientação são realizados na qual cada proteína no seu turno é permitida rotacionar para minimizar sua função energia intermolecular. Então as rotações e translações são executadas e as duas proteínas são acopladas pela minimização de energia de corpo-rígido. Usualmente cerca de 1000 complexos são calculados nesse estágio. As melhores 200 soluções em termos de energia intermolecular irão para a próxima fase.

O segundo estágio consiste de três refinamentos de recozimento simulados. O primeiro tem 1000 passos de 9 fentossegundos (fs) e leva o sistema de 2000 a 50 Kelvin (K). Aqui as duas proteínas são consideradas corpos rígidos e suas respectivas orientações são otimizadas. Na segunda simulação de recozimento, as cadeias laterais na interface são liberadas para se mover. Na terceira simulação, as cadeias laterais e a cadeia central na interface são liberadas para se mover permitindo alguns rearranjos conformacionais. As estruturas resultantes são então submetidas a 200 etapas de uma minimização de energia mais intensa.

O terceiro estágio consiste de um leve refinamento em uma cavidade de 8 Å preenchida com moléculas de água TIP3P (VRIES; VAN DIJK; BONVIN, 2010). Para a integração das equações de movimento nesse caso são usados passos de 2fs. O sistema é primeiramente aquecido até 300 K em 500 pasos com restrições de posições em todos os átomos, exceto os átomos das cadeias laterais flexíveis na interface proteína-proteína. Então 5000 passos de dinâmica molecular são realizados a 300 K com restrição de posições apenas nos átomos pesados fora da interface. Na fase final de resfriamento, de 1000 passos, as restrições de posição são limitadas aos átomos o esqueleto peptídico fora da interface.

Portanto, ao final do *docking*, todas as estruturas são agrupadas de acordo com o emparelhamento do desvio quadrático médio (RMSD, do inglês *Root Mean Square Deviation*) em relação aos esqueletos peptídicos na interface. Isso significa que cada grupo formado contém pelo menos duas estruturas com RMSD de interface menor que 1 Å. Os grupos são então analisados e classificados de acordo com o seu HADDOCK score, que é simplesmente uma soma ponderada entre as energias de van der Waals, eletrostáticas, energias de desolvatação e de violação de restrições junto com uma média da área superficial não-exposta (VRIES; VAN DIJK; BONVIN, 2010).

Todo procedimento citado acima foi realizado automaticamente, sem interferências do usuário, pois usou-se a “*Easy Interface*” do Haddock, que permite apenas o envio das estruturas em formato “pdb” e a escolha de resíduos ativos e passivos. A avaliação da

performance de cada complexo proteico foi feita com base nos seguintes parâmetros: HADDOCK escore, tamanho do grupo – o número de estruturas finais que compõem cada grupo é uma avaliação da probabilidade de que aquela conformação das duas moléculas aconteça em condições *in vivo*; então pode-se afirmar que um grupo de tamanho maior tem maior probabilidade de acontecer em detrimento de um menor – e RMSD em relação a estrutura de menor energia.

Os complexos gerados pelo *docking* foram ainda analisados pelo servidor PDBePISA (KRISSEL; HENRICK, 2007) em busca de interações polares das interfaces das proteínas.

3.3 Mutagênese *in silico*: Alanine scanning

Para a realização do *alanine scanning* os modelos dos complexos proteína-proteína utilizados nesse trabalho, gerados pelo servidor HADDOCK 2.0, foram inseridos na plataforma ROSETTA em arquivos no formato pdb. O campo referente as mutações de alanina obtidas experimentalmente foi deixado em branco, e um escaneamento automático de todos os resíduos de interface para cada um dos complexos foi feito.

O escaneamento computacional da alanina usa uma simples função de energia livre que calcula os efeitos das mutações de alanina na energia livre de ligação de um complexo proteína-proteína. A função, mostrada abaixo, consiste de uma combinação linear de um potencial Lennard-Jones (E_{LJattr}) para descrever as interações do empacotamento atômico, um modelo de solvatação implícito (G_{sol}), um potencial de ligação de hidrogênio dependente de orientação ($E_{HB(sc-sc)}$ e $E_{HB(sc-bb)}$) um potencial derivado de estruturas proteicas de alta resolução, termos estatísticos de aproximação do esqueleto dependente do tipo de aminoácido com probabilidades de rotâmeros (E_{aa}) e uma estimativa de estados de energia desdobrados (W).

$$\Delta G = W_{attr} E_{LJattr} + W_{rep} E_{LJrep} + W_{HB(sc-bb)} E_{HB(sc-bb)} + W_{HB(sc-sc)} E_{HB(sc-sc)}$$

$$W_{sol} G_{sol} + W_{\phi/\psi} E_{\phi/\psi}(aa) + \sum_{aa=1}^{20} n_{aa} E_{aa}^{ref}$$

Quando o escaneamento da alanina é aplicado a um complexo sem a existência prévia de dados experimentais, como foi o caso nesse estudo, um algoritmo automaticamente identifica todos os resíduos de interface de uma interface proteína-proteína. Um resíduo de interface é definido como (i) um resíduo que tem ao menos um átomo no interior de uma esfera com um raio de 4 Å de um átomo pertencente a proteína parceira no complexo proteico, ou (ii) um resíduo que se tornou significativamente inserido em um complexo, medido por um aumento no número de átomos de C_β no interior de uma esfera com raio de 8 Å ao redor de um átomo C_β de um resíduo de interesse. O programa então substitui cada um dos resíduos de interface individualmente com resíduos de alanina e computa o efeito dessa mutação sobre a energia livre de ligação do complexo. Para efeitos de comparação de dados experimentais, um *hot-spot* identificado corretamente significa um resíduo com um valor de $\Delta\Delta G_{\text{ligação}}$ esperado e observado maior que 1 kcal/mol.

Para esse trabalho os resíduos “*hot-spots*” foram definidos como aqueles nos quais as mutações da alanina tiveram efeitos desestabilizadores no $\Delta\Delta G_{\text{ligação}}$ de mais de 1 kcal/mol.

4 RESULTADOS E DISCUSSÃO

4.1 Análise dos cristais das proteínas virais e Lactoferrina

A visualização da estrutura dos cristais das proteínas permitiu uma melhor compreensão dos vários níveis organizacionais e da função das proteínas do ZIKV e da lactoferrina. Essas estruturas foram visualizadas pelo *Pymol 1.8.6.2* e representadas tanto por suas cadeias, quanto por suas estruturas secundárias.

Entre os cristais das proteínas não-estruturais do ZIKV, o cristal 5IY3 da proteína NS1, formado por 352 resíduos de aminoácidos, apresentou duas cadeias simétricas e opostas na molécula (indicadas pelas cores verde e azul) e uma predominância de folhas-beta em sua estrutura secundária (representadas pela cor amarela) (Figura 2).

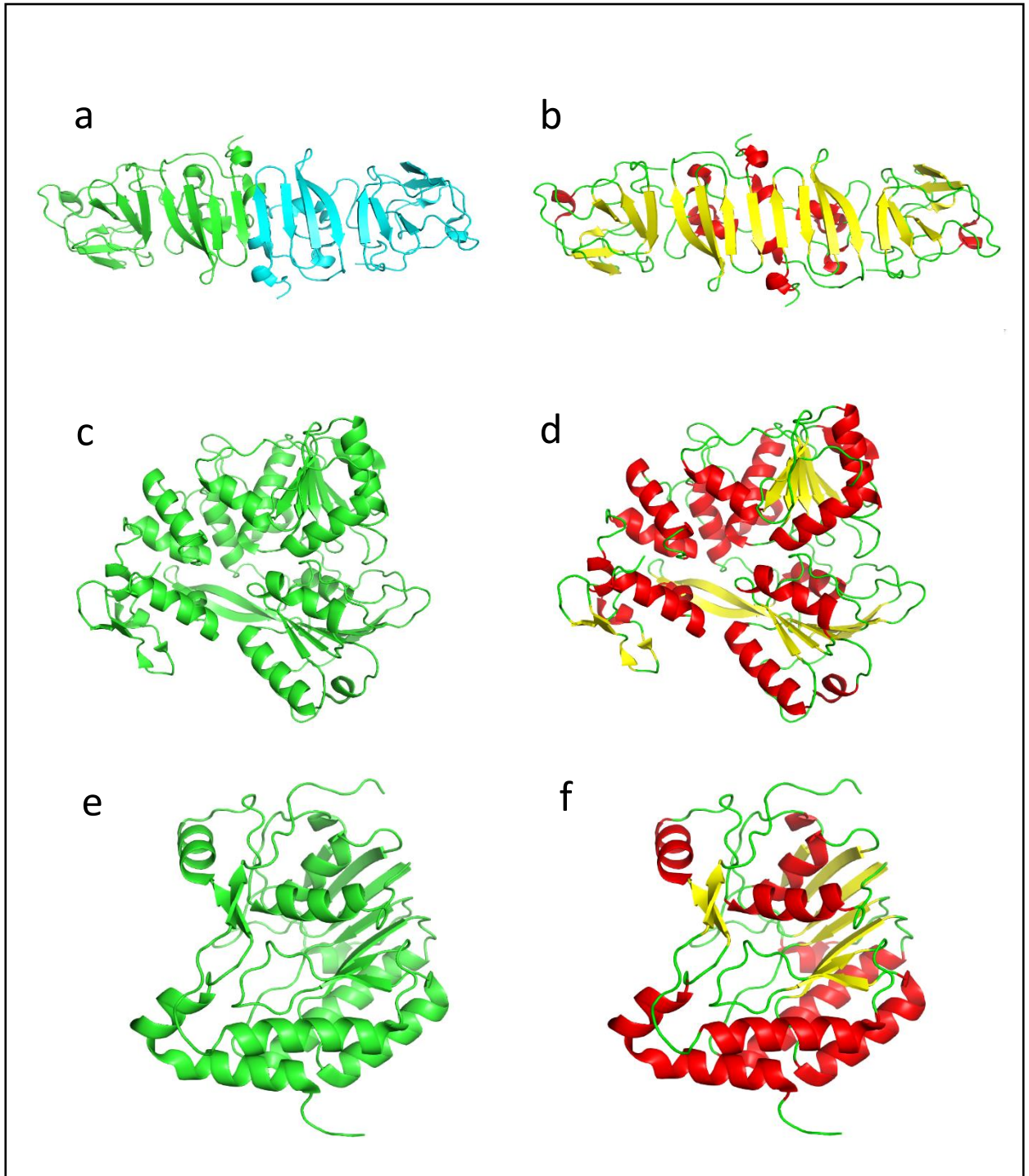
Os outros dois cristais de proteínas não-estruturais obtidos do *Protein Data Bank* (PDB) (BERMAN *et al*, 2000), o 5JMT, correspondente a proteína helicase NS3, de 442 aminoácidos, e o 5KQR, da proteína polimerase NS5, de 263 aminoácidos, são formadas por uma única cadeia e possuem predominância de alfa-hélices na estrutura secundária.

Todos as três estruturas (NS1, NS3, NS5) foram obtidas por meio de difração de raio-x com resoluções de 2.2, 1.8 e 1.33 Å respectivamente. Nenhum dos arquivos dos cristais depositados no PDB possuíam moléculas ligadas à algum ponto da estrutura que pudesse indicar potenciais sítios de ligação.

Para o estudo da lactoferrina foi utilizado o cristal 4OQO, formado por duas cadeias de 342 aminoácidos cada uma. A estrutura bi-lobular dessa proteína é projetada para duplicar sua capacidade sequestradora de íons de ferro. As duas subunidades são conectadas por um peptídeo helicoidal de 10 resíduos, e cada subunidade contém um sítio de ligação à átomos de ferro. Na estrutura bi-lobular da lactoferrina (Figura 3), a interface entre os dois lóbulos providencia uma estabilidade adicional a proteína (HASTOGI *et al*, 2014).

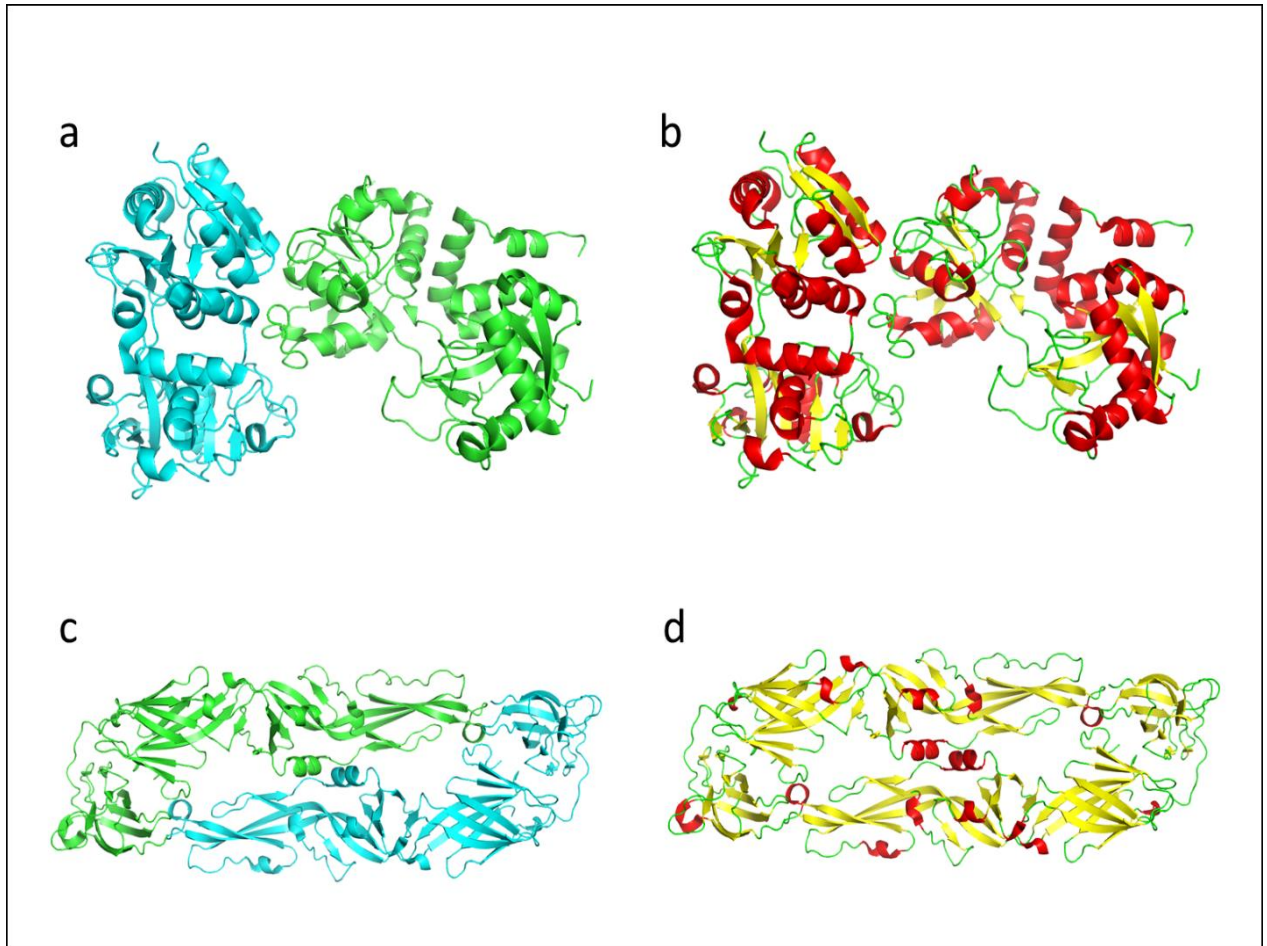
A GPE, uma proteína estrutural, foi estudada através do cristal 5JHM. Essa proteína também é formada por duas cadeias que possuem, cada uma, 406 resíduos de aminoácidos. Essas duas subunidades são ligadas fracamente por um único resíduo de volta em cada lado, o que torna sua estrutura relativamente fraca (Figura 3).

Figura 2 - Representação das proteínas NS1, NS3 e NS5 em cadeias indicadas pela cor verde e azul nos cristais (a), (c) e (e) respectivamente. Nos cristais (b), (d) e (f) representação da estrutura secundária das proteínas NS1, NS3 e NS5 respectivamente, onde as alfas hélices estão indicadas em vermelho, as folhas beta em amarelo e as alças em verde.



Fonte: Elaborado pelo autor (2017)

Figura 3 – Representação da lactoferrina e proteína envelope em cadeias indicadas pela cor verde e azul nos cristais (a) e (c) respectivamente. Nos cristais (b) e (d) representação da estrutura secundária da lactoferrina e proteína envelope respectivamente, onde as alfas hélices estão indicadas em vermelho, as folhas betas em amarelo e as alças em verde.



Fonte: elaborado pelo autor (2017).

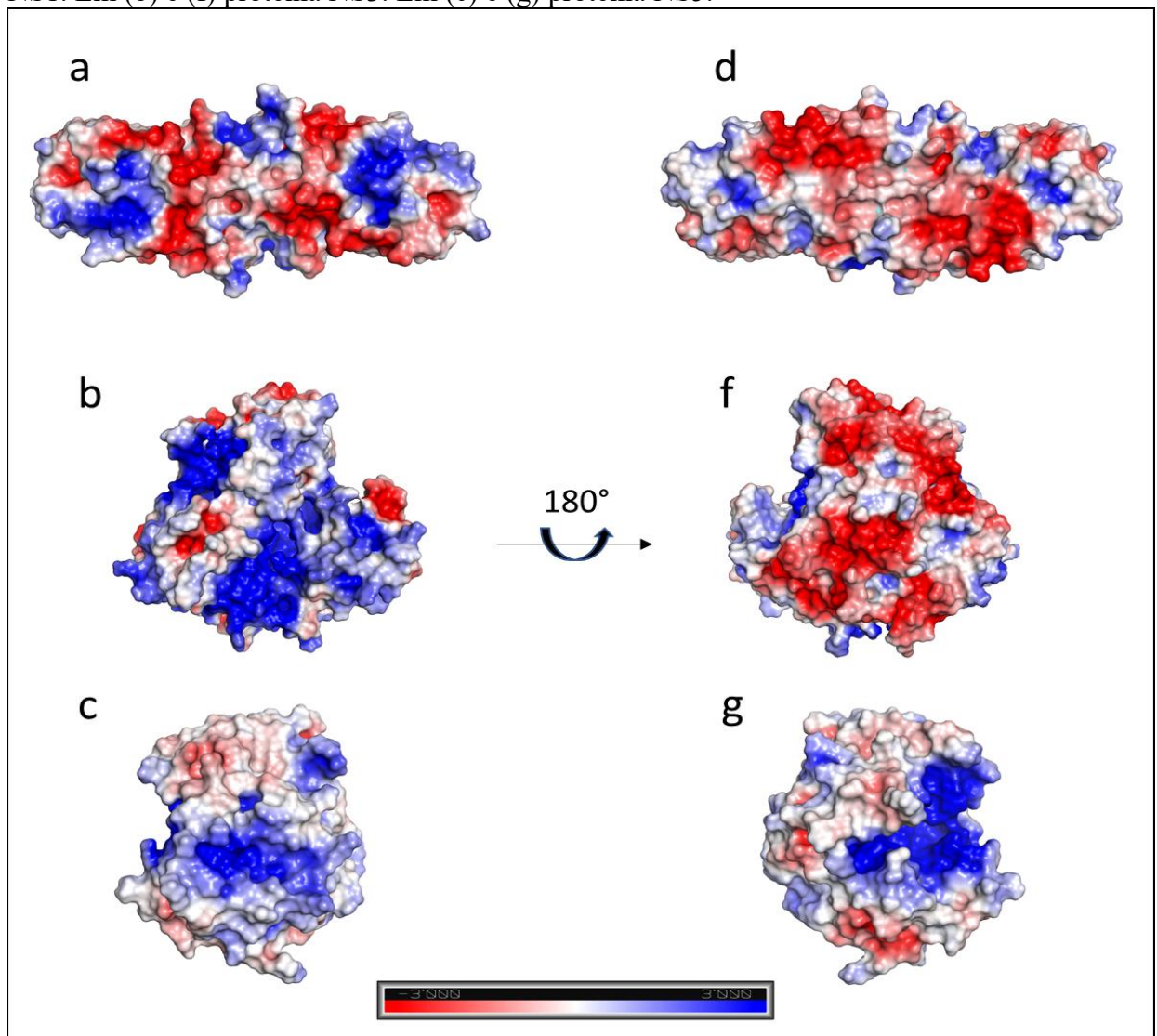
4.2 Análise do perfil eletrostático das proteínas virais do Zika Vírus e da lactoferrina

O perfil eletrostático gerado para cada uma das proteínas com o auxílio do *software APBS* estão evidenciados nas figuras 3 e 4. O campo de perfil eletrostático foi ajustado para ± 3 kT/e de modo que fosse mais evidente o contraste entre as regiões positivas, negativas e neutras das superfícies das proteínas.

A proteína NS1 apresentou uma significativa área da superfície carregada, nas porções centrais da proteína as cargas se distribuíram de uma forma heterogênea e carregadas e em direção as regiões distais houve uma predominância de cargas negativas em ambos os lados da proteína. Quando analisada juntamente com as cavidades foi percebido que o sítio de ligação MPT3 se encontrava em uma região de cargas predominantemente positivas, enquanto as outras duas foram encontradas em regiões fortemente negativas. Quando

comparada aos perfis eletrostáticos de proteínas NS1 de outros flavivírus a superfície conhecida por ser mais ativa na proteína NS1 (figura 4) apresenta um perfil destoante dos demais, como por exemplo da proteína NS1 do vírus da dengue e do Oeste do Nilo, o que pode refletir em diferentes propriedades de ligação a anticorpos com propriedades inibitórias nesses outros flavivírus (KIM *et al*, 2016).

Figura 4 – Perfil eletrostático das proteínas NS1, NS3 e NS5 do ZIKV. Em (a) e (d) proteína NS1. Em (b) e (f) proteína NS3. Em (c) e (g) proteína NS5.



Fonte: elaborado pelo autor

Já para o perfil eletrostático da proteína NS3 houve uma clara predominância de cargas significativamente positivas em uma face e cargas negativas na face oposta. No entanto, quando analisado com as cavidades encontradas pelo *metaPocket* o hemisfério positivo foi o

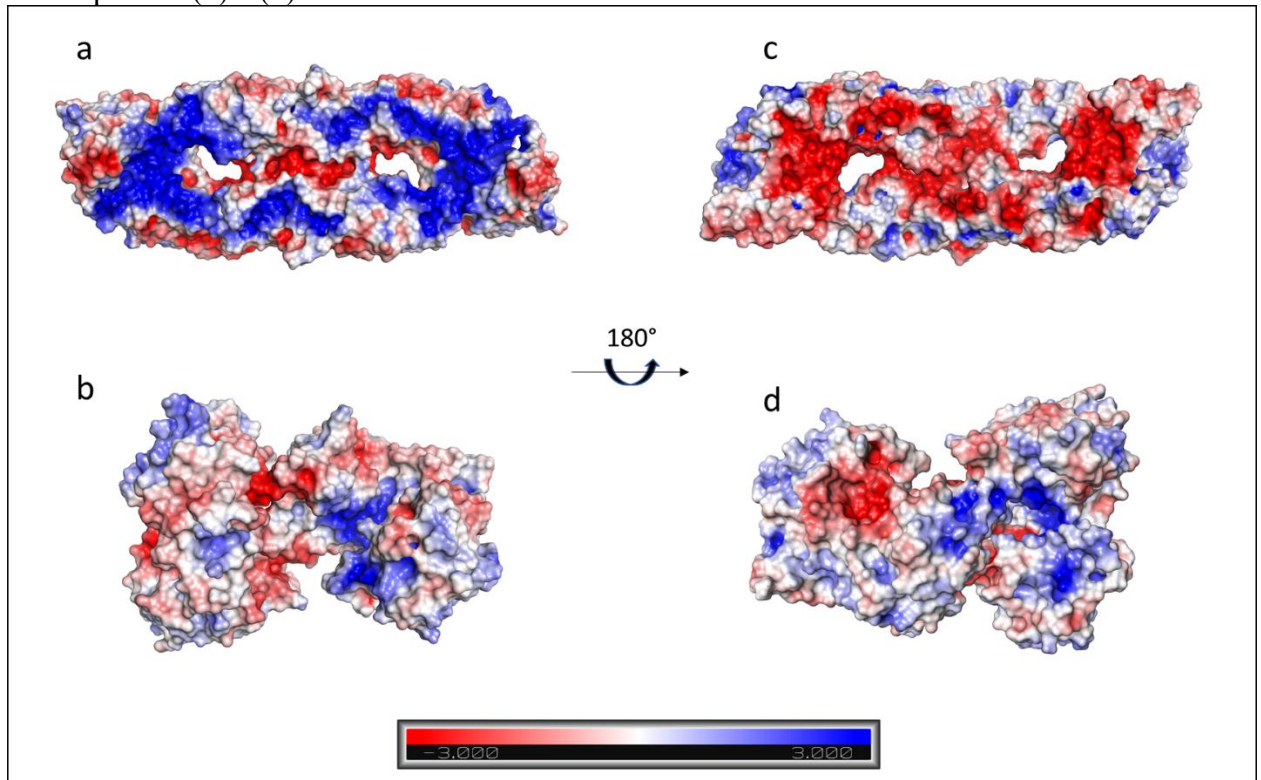
único que apresentou sítios de ligação, o que sugere que potenciais moléculas e proteínas que se liguem a essa proteína deverão possuir sítios de ligação negativos.

A proteína NS5, no entanto, destoa das outras proteínas analisadas quando se observa que ela apresentou uma considerável área de superfície neutra, uma pequena área de carga negativa e de baixa intensidade e duas grandes regiões de forte potencial eletropositivo em cada um dos lados da proteína. A localização das cavidades encontradas pelo *metaPocket* na proteína NS5 se deu precisamente nas duas regiões positivas.

O perfil eletrostático da proteína envelope exhibe um padrão que pode se relacionar com sua função na constituição do complexo que envolve o vírus. Cada uma das faces da proteína possui uma predominância de apenas uma das cargas, assim como foi visto para a proteína NS3. O lado positivo, quando visualizado no complexo formado por várias repetições da proteína envelope, se posiciona voltado para o lado exterior, enquanto que a face negativa está voltada para o interior. Dessa forma, os sítios de ligação encontrados na proteína envelope estão voltados para a superfície exterior, embora alguns não estejam completamente expostos.

Por último, o perfil eletrostático calculado para a lactoferrina demonstrou uma heterogeneidade de cargas de baixa intensidade pela superfície da proteína, com uma área neutra significativa e algumas regiões positivas e negativas de maior intensidade. Como não foi possível a recuperação das cavidades da lactoferrina através do servidor *metaPocket*, a visualização dos sítios de ligação só foi possível nos complexos formados após o *docking*.

Figura 4 – Perfil eletrostático das proteínas envelope e lactoferrina. Em (a) e (c) proteína envelope. Em (b) e (d) lactoferrina.



Fonte: Elaborado pelo autor

4.3 Determinação dos sítios de ligação dos cristais das proteínas do ZIKV

Os sítios de ligações dos cristais das proteínas NS1, NS3, NS5 e Envelope, preditos pelo servidor *metaPocket* 2.0, correspondem às três principais regiões de cada proteína, ou seja, aquelas com os maiores números de resíduos ativos próximos identificados pelo programa.

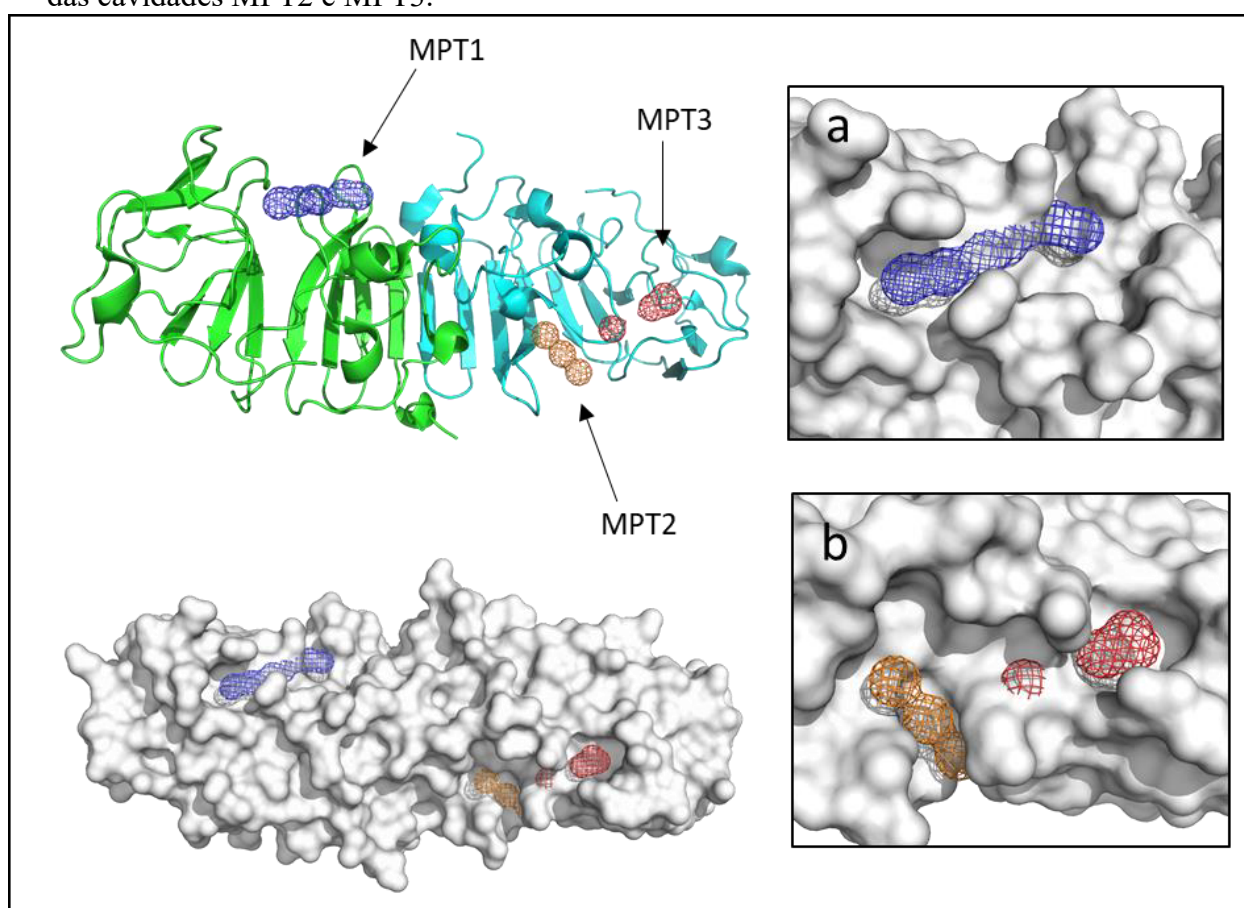
4.3.1 Cavidades encontradas para o cristal 5IY3 da proteína NS1

As imagens resultantes da análise com o *metaPocket* foram geradas representando o cristal tanto por suas cadeias (em verde e azul), quanto por sua superfície para garantir a melhor visualização das cavidades encontradas, possibilitando assim, a escolha das melhores regiões para o *docking* molecular.

Como determinado na metodologia do presente trabalho, os três sítios de ligação em cada uma das proteínas foram denominados: MPT1, MPT2 e MPT3, indicadas por estruturas em malha de cores padronizadas em azul, laranja e vermelho, respectivamente.

As cavidades encontradas na proteína NS1 pelo *metaPocket* foram todas consideradas para o *docking*, visto que as três cavidades apresentaram um nível significativo de exposição na superfície da proteína. Curiosamente, duas das cavidades encontradas, MPT2 e MPT3, estão situadas em regiões muito próximas, formando praticamente um único sítio de ligação que é equivalente em posição a região da cavidade MPT1 na cadeia oposta (Figura 6).

Figura 6 – Representação das cavidades da proteína NS1. No canto superior esquerdo, representação em cadeia da proteína NS1 com as cavidades MPT1, MPT2 e MPT3 em azul, laranja e vermelho respectivamente. No canto inferior esquerdo, representação em superfície da proteína NS1 com as cavidades inseridas, com detalhes em (a) da cavidade MPT1 e em (b) das cavidades MPT2 e MPT3.



Fonte: elaborado pelo autor

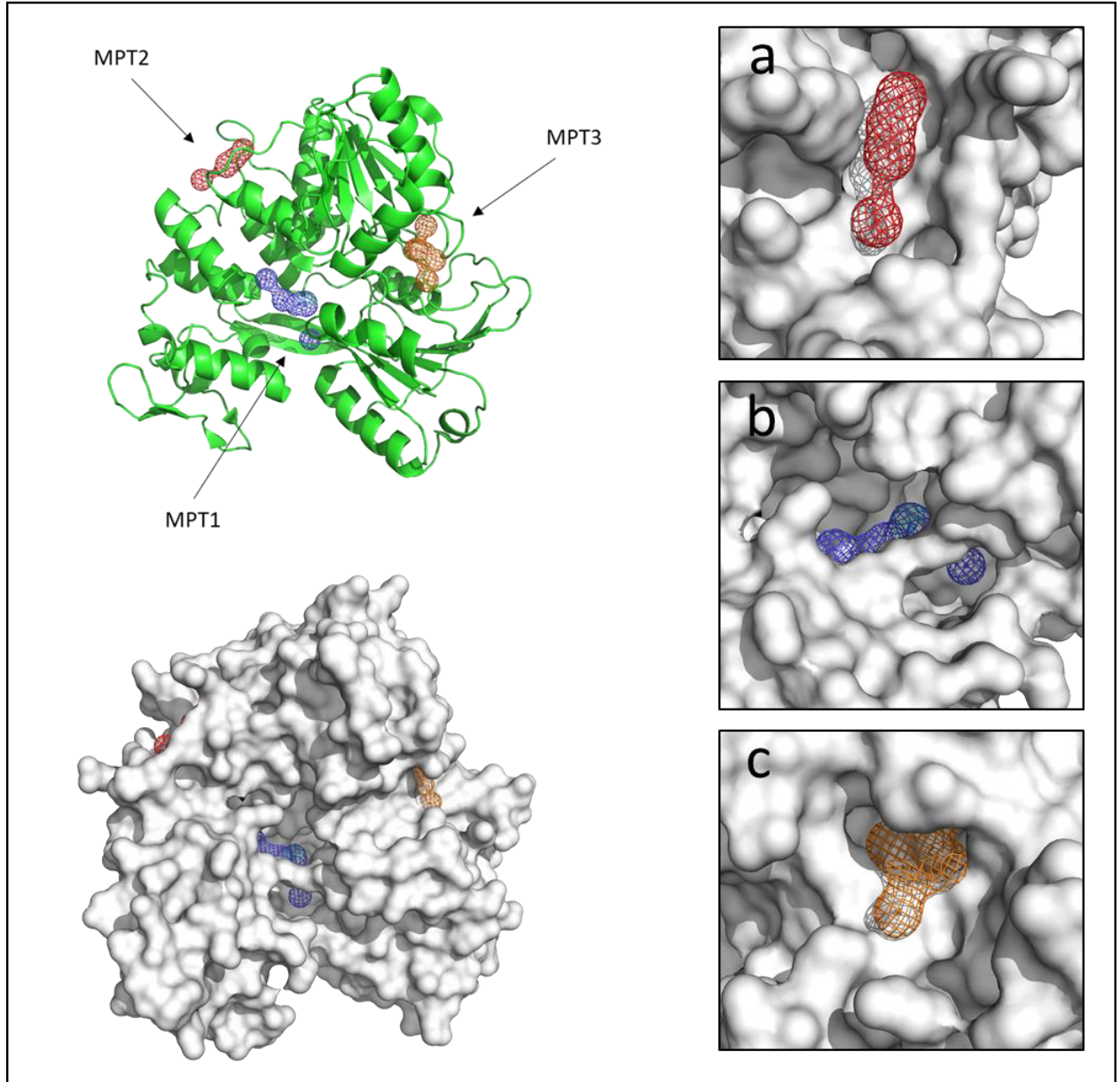
4.3.2 Cavidades encontradas para o cristal 5JMT da proteína NS3

Entre as cavidades encontradas na proteína NS3, apenas duas foram consideradas para o *docking*. A cavidade MPT1 apresentou suas coordenadas no interior da proteína, visualmente inacessível para que o atracamento com uma outra proteína fosse viável.

As cavidades MPT2 e MPT3, no entanto, foram encontradas em regiões expostas da proteína e, dessa forma, ambas tiveram seus resíduos ativos anotados para que fossem usados no *docking* posteriormente.

Além disso, a região MPT3 em outros cristais da proteína NS3 disponíveis na plataforma PDB, coincide com pontos de ligação a moléculas de ATP e alguns íons, o que também sugere uma importante atividade dessa região (Figura 7).

Figura 7 - Representação das cavidades da proteína NS3. No canto superior esquerdo, representação em cadeia da proteína NS3 com as cavidades MPT1, MPT2 e MPT3 em azul, laranja e vermelho respectivamente. No canto inferior esquerdo, representação em superfície da proteína NS3 com as cavidades inseridas, com detalhes em (a) da cavidade MPT2, em (b) da cavidade MPT1 e em (c) da cavidade MPT3.

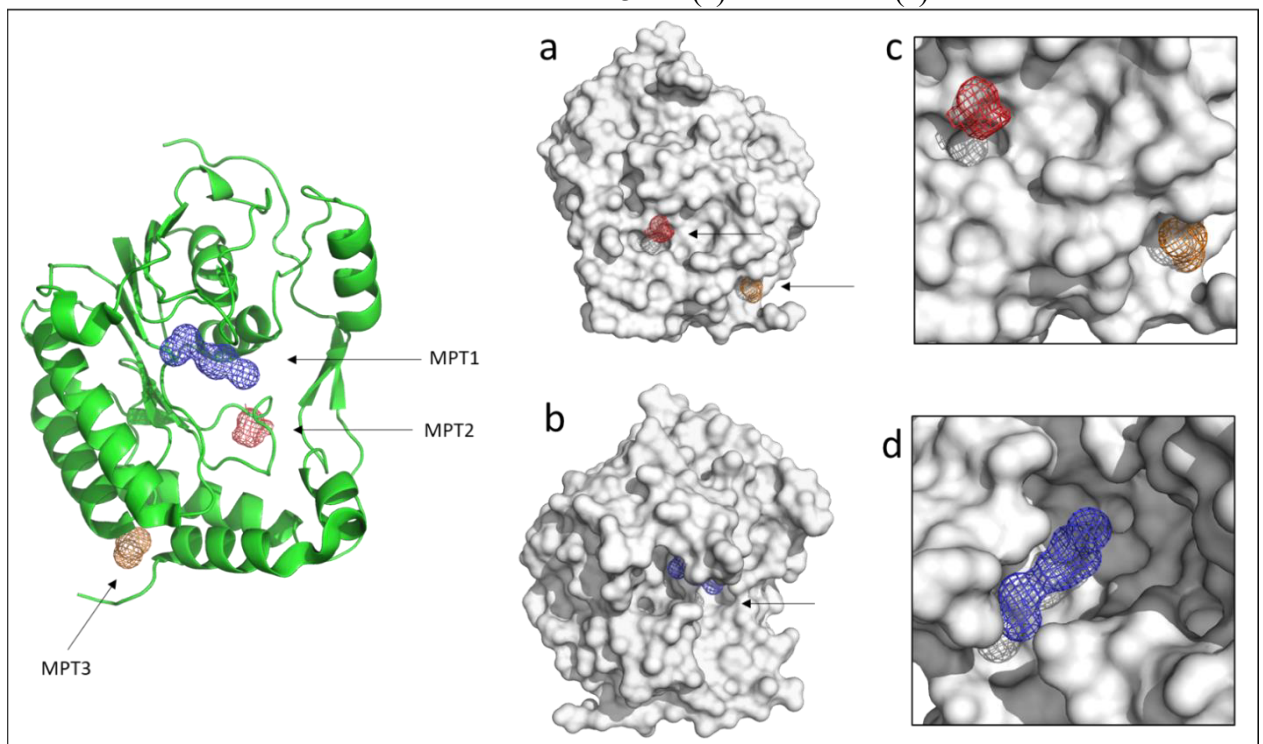


Fonte: elaborado pelo autor

4.3.3 Cavidades encontradas para o cristal 5KQR da proteína NS5

O terceiro cristal analisado foi o 5KQR correspondente a proteína NS5. As três cavidades encontradas foram consideradas viáveis para o *docking*, mesmo tendo em vista que a cavidade MPT1 não se encontrava exposta da melhor forma. Em contraste, as outras duas cavidades apresentaram boa exposição na superfície da proteína e, como foi visto com duas cavidades da proteína NS3, as cavidades MPT2 e MPT3 foram encontradas em regiões próximas, o que pode indicar uma favorável região alvo para neutralizar essa proteína (Figura 8).

Figura 8 – Representação das cavidades da proteína NS5. A esquerda, representação em cadeia da proteína NS1 com as cavidades MPT1, MPT2 e MPT3 em azul, vermelho e laranja respectivamente. Em (a) e (b) representação em superfície da proteína NS5 com as cavidades inseridas. Detalhes das cavidades MPT2 e MPT3 em (c) e MPT1 em (a).



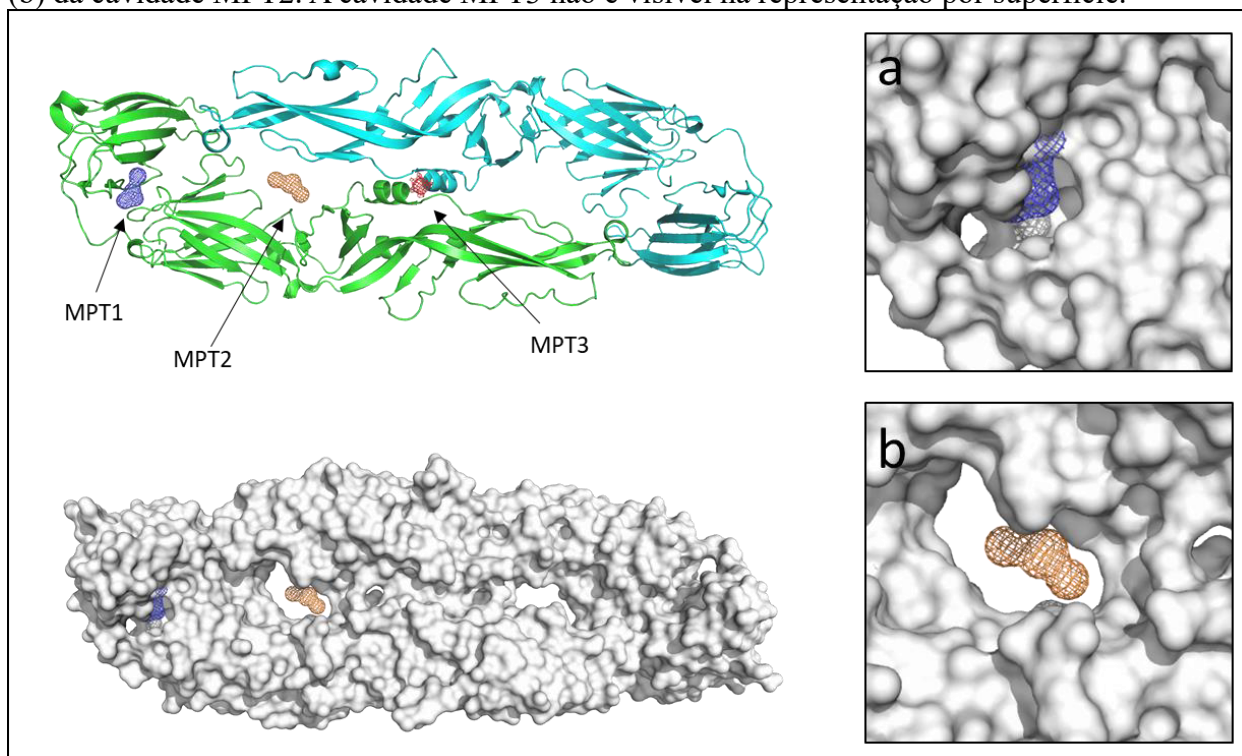
Fonte: elaborado pelo autor

4.3.4 Cavidades encontradas para o cristal 5JHM da proteína Envelope

Por último, o cristal 5JHM da proteína envelope foi analisado pelo *metaPocket*. A cavidade MPT3 foi identificada no interior da proteína, de forma que na representação por superfície ela não demonstrou nenhuma visibilidade da malha, e, portanto, não foi considerada

para o *docking*. Dessa forma, somente duas das três cavidades foram classificadas como acessíveis, tendo, então, seus resíduos ativos anotados para o *docking* (Figura 9).

Figura 9 - Representação das cavidades da proteína envelope. No canto superior esquerdo, representação em cadeia da proteína envelope com as cavidades MPT1, MPT2 e MPT3 em azul, laranja e vermelho respectivamente. No canto inferior esquerdo, representação em superfície da proteína envelope com as cavidades inseridas, com detalhes em (a) da cavidade MPT1, em (b) da cavidade MPT2. A cavidade MPT3 não é visível na representação por superfície.



Fonte: elaborado pelo autor.

4.4 Identificação dos resíduos ativos da lactoferrina e das proteínas NS1, NS3, NS5 e Glicoproteína E

A tabela abaixo lista todos os resíduos ativos identificados pelo servidor *metaPocket* 2.0 e que foram usados como dados de entrada no servidor HADDOCK para que o *docking* fosse realizado. Esses resíduos compõem os três sítios de ligação encontrados em cada uma das proteínas. Em geral, foi observado um grande número de resíduos ativos para todas as proteínas, com exceção da proteína NS1, que apresentou apenas 32 resíduos. Em contraste, a lactoferrina e a proteína NS3 apresentaram 183 e 207 resíduos ativos respectivamente.

<i>Proteína alvo</i>	<i>PDB ID</i>	<i>Resíduos ativos</i>
<i>Lactoferrina</i>	4OQO	347,348,350,351,352,353,354,355,356,357,359,360,373,374,375,376,377,378,379,382,383,386,388,392,393,394,395,396,397,398,399,401,413,414,415,416,417,418,419,420,421,421,422,423,424,425,428,429,430,431,432,433,434,440,441,442,443,446,455,459,460,461,462,463,464,465,466,467,468,473,481,483,484,489,493,494,495,496,497,503,504,505,506,508,509,510,511,512,513,514,515,516,517,518,519,520,521,522,523,524,525,526,527,528,531,533,534,535,536,537,538,539,544,545,546,547,549,550,551,552,553,554,555,556,557,558,559,560,561,562,563,564,565,566,567,568,569,570,580,581,582,583,584,585,589,590,591,592,593,594,595,625,626,629,630,633,634,635,636,637,638,639,640,641,642,643,644,646,664,665,669,671,672,673,674,675,676,677,678,679,680,681,683.
<i>NSI</i>	5IY3	207,208,209,210,211,232,235,236,237,240,253,256,257,258,259,260,261,261,289,291,292,293,294,313,314,315,316,317,318,319,334,336.
<i>NS3</i>	5JMT	2,3,177,178,179,180,192,193,194,195,196,197,198,199,200,201,202,203,204,207,224,225,226,227,228,230,231,232,234,235,244,245,246,247,262,263,264,265,267,268,269,270,271,272,273,274,275,276,277,278,282,285,286,288,289,290,291,292,293,294,300,304,307,309,315,316,317,318,319,320,321,322,323,327,330,331,357,358,362,363,364,365,366,367,369,370,385,386,387,388,389,390,392,393,394,395,396,398,399,400,401,402,403,407,408,409,410,411,412,413,414,415,416,417,418,419,420,421,428,429,430,431,432,442,446,447,448,449,451,452,454,455,456,458,459,460,461,462,463,466,469,477,478,479,480,481,482,483,484,486,487,489,493,495,496,497,498,499,500,501,502,503,504,509,510,511,512,513,513,514,523,524,525,526,527,528,529,530,531,533,534,536,537,538,540,541,542,543,545,572,573,574,575,597,598,599,600,601,602,603,604,605,606,607,608,609,617
<i>NS5</i>	5KQR	2,3,4,5,6,7,8,9,10,11,29,32,33,34,35,37,41,55,56,57,58,59,60,61,64,79,81,82,83,84,85,86,87,104,105,106,109,110,111,112,130,131,132,133,146,147,148,149,150,157,159,160,161,162,163,164,182,184,187,189,190,192,193,196,208,210,211,212,213,215,216,218,220,230,233,234,235,237,238
<i>GPE</i>	5JHM	1,2,3,4,5,6,7,8,26,27,28,29,30,38,39,40,44,45,46,47,48,61,62,63,64,97,98,99,100,101,102,103,108,109,122,123,140,143,144,145,146,163,164,183,206,207,208,209,210,211,217,220,221,222,223,240,241,242,244,246,247,248,249,250,251,252,253,254,255,256,257,258,259,260,261,262,264,265,266,267,268,269,270,271,272,273,274,275,276,277,278,279,280,283,284,285,286,300,301,302,303,304,305,306,312,322,329,330,331,332,338,340,341,343,344,358,360,361,362,363,364,365,366,367,372,373,374,375,376,378,387,388,389

Fonte: elaborado pelo autor

4.5 Docking entre a Lactoferrina e as proteínas NS1, NS3, NS5 e proteína Envelope

Os resultados dos *dockings* proteína-proteína são mostrados na tabela abaixo, mostrando que nenhum dos complexos obteve os melhores resultados nos três parâmetros (HADDOCK score, Tamanho do grupo e RMSD) simultaneamente. No entanto, levando-se em conta o tamanho do grupo como parâmetro de maior significância o complexo LACTF-Envelope pode ser considerado o de maior estabilidade, seguido pelos complexos LACTF-NS5 e LACTF-NS3.

Para que houvesse uma segunda base de comparação, as mesmas proteínas foram acopladas através do servidor ClusPro, que também utiliza o tamanho do grupo como método principal de ranqueamento dos complexos gerados. Desse modo, dentre os modelos gerados por esse programa para cada um dos complexos, o melhor modelo para o complexo LACTF-Envelope coincidiu com o mesmo posicionamento das proteínas do complexo produzido com o servidor HADDOCK, apresentando uma região de ligação muito semelhante. Esse modelo possui um grupo com 30 membros e as menores energias entre todos os complexos.

Tabela 1 – Avaliação de cada *docking* proteína-proteína realizado.

<i>Complexo dockado</i>	<i>HADDOCK score</i>	<i>Tamanho do grupo</i>	<i>RMSD (Å)</i>
<i>LACTF- Envelope</i>	650.2 +/- 13.6	43	29.8 +/- 0.1
<i>LACTF - NS1</i>	216.0 +/- 7.4	17	24.9 +/- 0.2
<i>LACTF – NS3</i>	435.1 +/- 21.6	23	1.2 +/- 0.9
<i>LACTF – NS5</i>	420.7 +/- 35.2	27	10.2 +/- 0.5

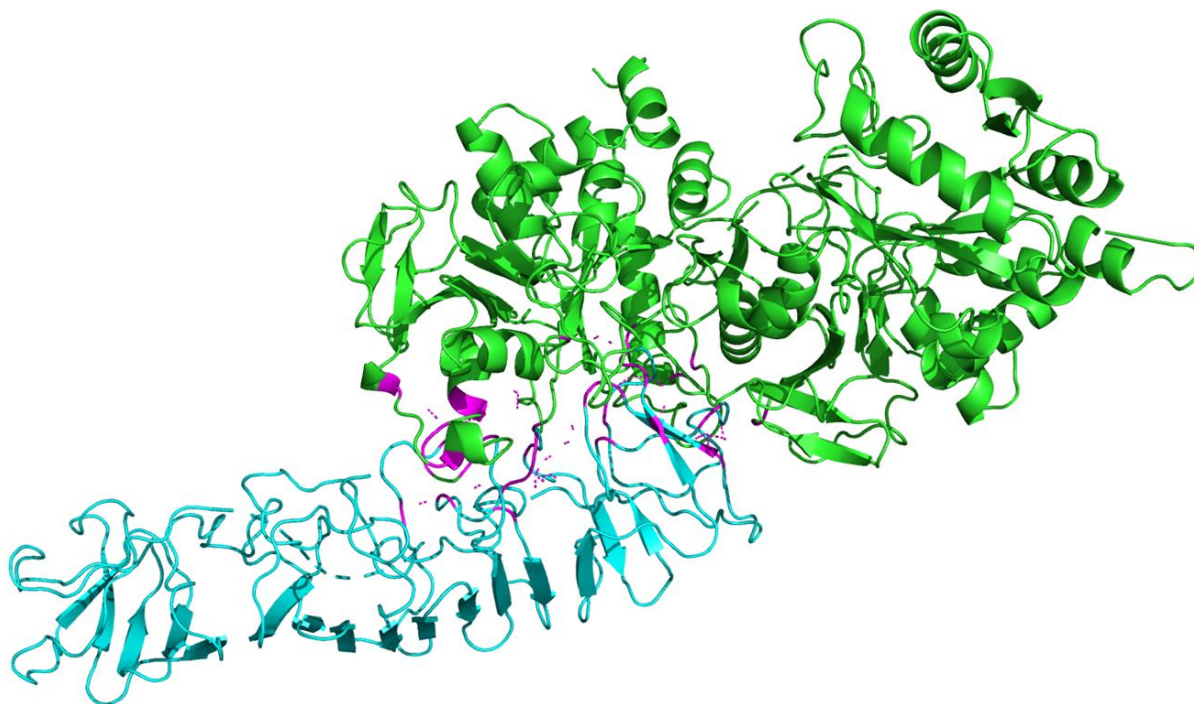
Fonte: elaborado pelo autor

As figuras a seguir mostram os complexos pós-*docking* revelando detalhes sobre o posicionamento das proteínas. Através do servidor PDBePISA, diversas ligações hidrogênio e pontes salinas foram encontradas em todos os complexos formados entre a Lactoferrina e as proteínas virais que serão descritas para cada um dos complexos.

Para o complexo LACTF-NS1 (Lactoferrina-NS1) foram identificadas 31 ligações hidrogênio, sendo a menor delas, de 1.62 Å, formada entre a glutamina 642 da Lactoferrina e a lisina 341 da NS1, e a maior, de 3.85 Å, formada entre a histidina 427 da

Lactoferrina e a metionina 266 da NS1. Para o mesmo complexo também foram identificadas 16 pontes salinas, sendo a menor delas, de 2.64 Å, formada entre o ácido glutâmico 557 da Lactoferrina e a arginina 263 da proteína NS1 e a maior de, 3.94 Å, formada entre o ácido aspártico 227 da Lactoferrina e a arginina 340 da NS1 (Figura 10).

Figura 10 – Complexo LACTF- NS1 pós-*docking*, com os resíduos formadores de ligações em violeta, e representação das ligações de hidrogênio em linhas violetas pontilhadas.



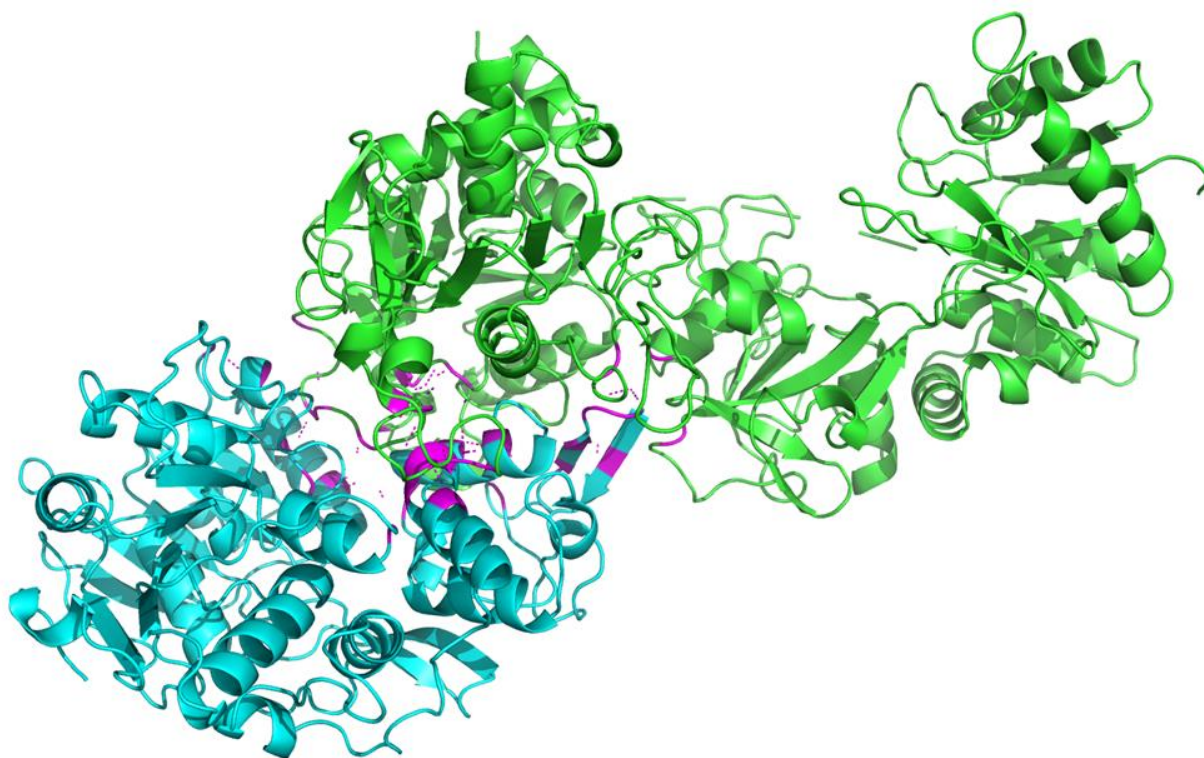
Fonte: elaborado pelo autor com auxílio do programa PyMol.

Para o complexo LACTF-NS3 foram identificadas 34 ligações de hidrogênio: a menor delas, de 1.57 Å, formada entre o ácido glutâmico 562 da Lactoferrina e a lisina 363 da proteína NS3, e a maior, de 3.32 Å, formada entre a asparagina 525 pertencente a Lactoferrina e a histidina 410 da NS3. Já para as pontes salinas entre as 19 identificadas a menor encontrada, de 2.62 Å, formada entre a lisina 640 da e o ácido glutâmico 399, pertencentes a Lactoferrina e a proteína NS3 respectivamente (Figura 11).

No complexo LACTF-NS5 foram identificadas 38 ligações de hidrogênio: a menor delas, de 1.58 Å, formada entre o ácido glutâmico 666 da Lactoferrina e a lisina 43 da NS5 e a maior, de 3.56 Å, formada entre o ácido glutâmico 438 e a histidina 108, pertencentes a Lactoferrina e a proteína NS5 respectivamente. Foram identificadas 15 pontes salinas sendo que a menor delas, de 2.61 Å, é formada pelo mesmo ácido glutâmico e histidina que formam

a menor ligação de hidrogênio nesse complexo, e a maior, de 3.85 Å, é formada entre um resíduo do ácido glutâmico 590 da Lactoferrina e uma arginina 123 da proteína NS5 (Figura 11).

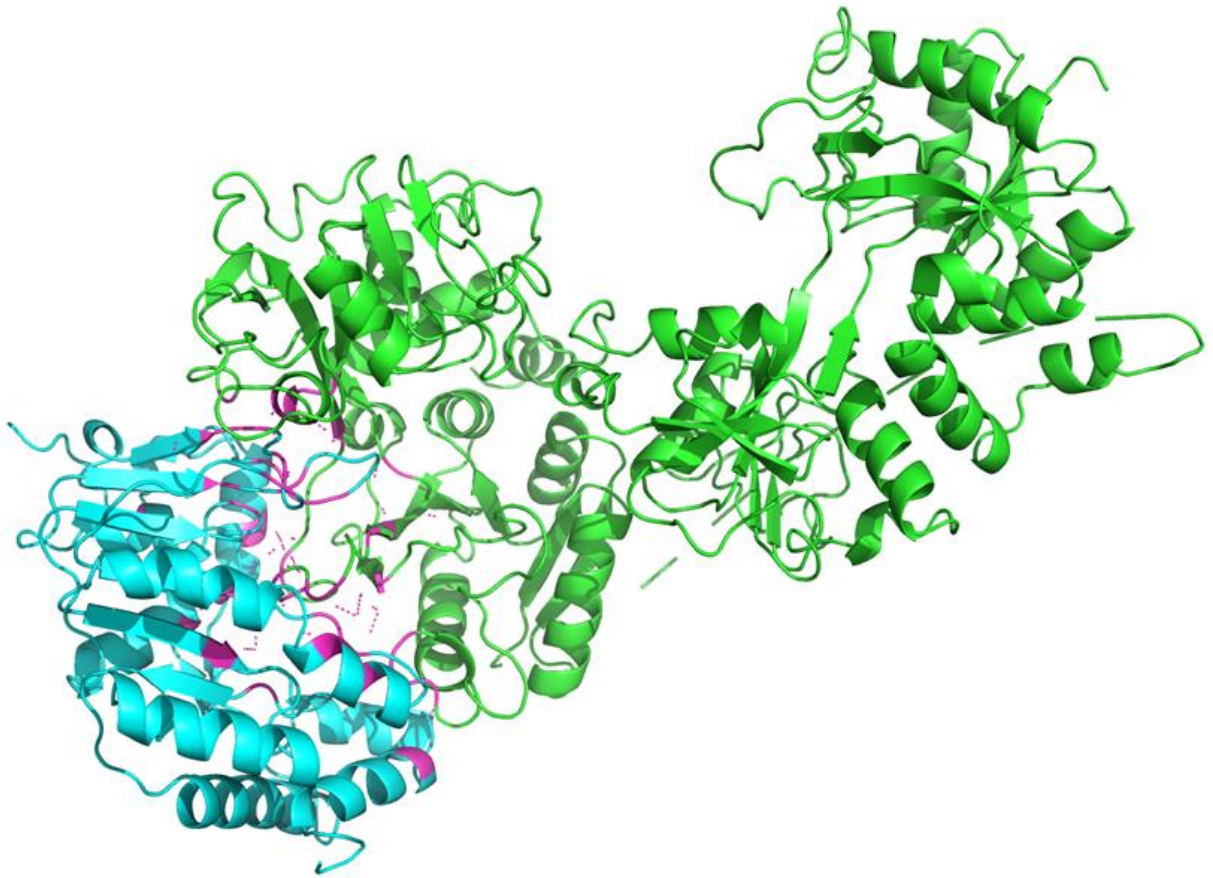
Figura 11 – Complexo LACTF- NS3 pós-*docking*, com os resíduos formadores de ligações em violeta, e representação das ligações de hidrogênio em linhas violetas pontilhadas.



Fonte: elaborado pelo autor com auxílio do programa PyMol.

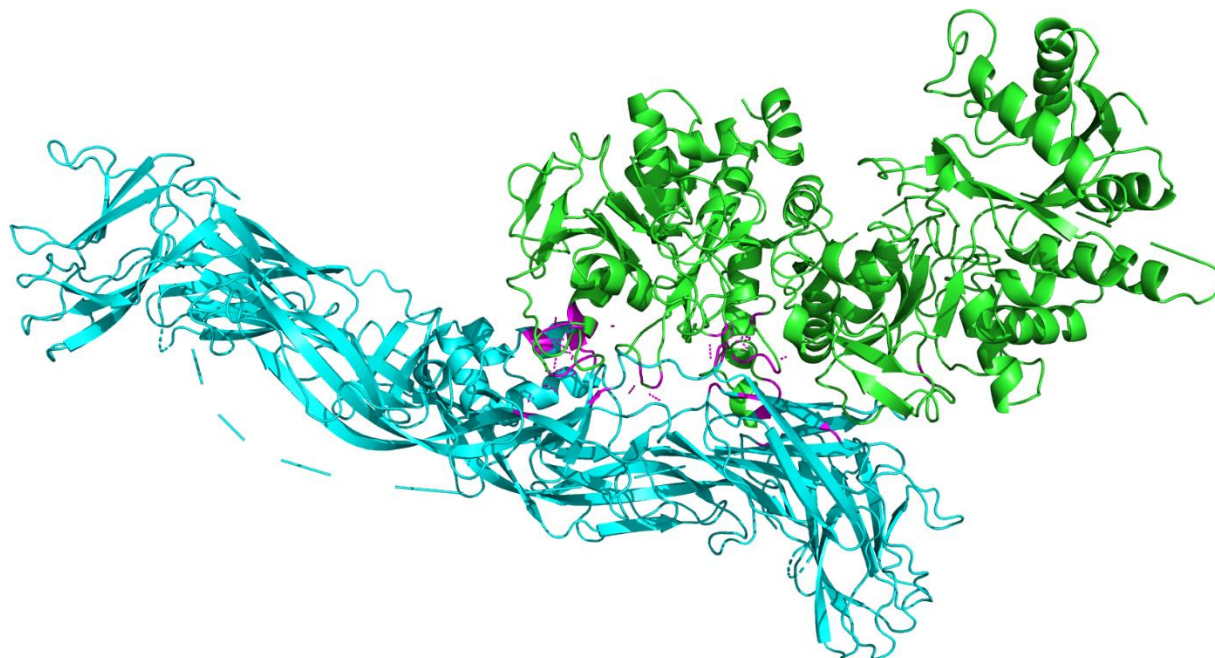
Por último, foram identificadas 28 ligações de hidrogênio no complexo LACTF-GPE (Glicoproteína E). A menor das ligações, de 1.55 Å, formada entre uma lisina 627 da Lactoferrina e um ácido glutâmico 26 da GPE e a maior, de 3.71 Å, formada entre uma leucina 430 da Lactoferrina e uma alanina 255 da GPE. Além disso, 26 pontes salinas foram identificadas, o maior número de pontes salinas entre os complexos. Dentre elas, três ligações diferentes apresentaram a menor distância entre os átomos: 2.59 Å. Já o maior comprimento foi compartilhado por duas ligações diferentes: uma entre o ácido glutâmico 103 da lactoferrina e a arginina 160 da GPE e, a outra, entre a lisina 640 da Lactoferrina e o ácido glutâmico 26 da GPE. Praticamente todos os resíduos da LACTF que participaram das ligações de hidrogênio também formaram pontes salinas com a GPE (Figura 12).

Figura 11 – Complexo LACTF- NS5 pós-*docking*, com os resíduos formadores de ligações em violeta e representação das ligações de hidrogênio em linhas violetas pontilhadas.



Fonte: elaborado pelo autor com auxílio do programa PyMol.

Figura 12 – Complexo LACTF- Envelope pós-*docking*, com os resíduos formadores de ligações em violeta e representação das ligações de hidrogênio em linhas violetas pontilhadas.



Fonte: elaborado pelo autor com auxílio do programa PyMol.

Como resumido na Tabela 2, o complexo proteína Envelope-LACTF apresentou a menor ligação de hidrogênio como também a menor ponte salina, que nesse caso se repetiu em três aminoácidos diferentes do complexo. Embora tenha apresentado o menor número de ligações de hidrogênio no total, o complexo Envelope-LACTF possui o maior número de pontes salinas dentre os 4 complexos.

Tabela 2 – Avaliação dos complexos quanto as ligações de hidrogênio e pontes salinas formadas

COMPLEXOS	TOTAL DE LIGAÇÕES DE HIDROGÊNIO	TOTAL DE PONTES SALINAS	MENOR LIGAÇÃO DE hidrogênio (Å)	MENOR PONTE salina (Å)
<i>NS1-LACTF</i>	31	16	1.62	2.64
<i>NS3-LACTF</i>	34	19	1.57	2.62
<i>NS5-LACTF</i>	38	15	1.58	2.61
<i>ENVELOPE- LACTF</i>	28	26	1.55	2.59

Fonte: elaborado pelo autor

4.3 Escaneamento computacional de alanina ponto-a-ponto dos complexos *dockados*

Os resíduos energeticamente importantes (“*hot spots*”) encontrados nos quatro complexos estudados através do servidor Robetta, em sua maioria, foram resíduos previamente identificados como resíduos ativos e que também formaram ligações de hidrogênio e pontes salinas evidenciadas pelo servidor PDBePISA. O resíduo de triptofano 556 e a histidina 427, pertencentes a Lactoferrina, demonstraram constituir resíduos particularmente importantes no sítio ativo que formou ligações com as proteínas virais. Cada um dos dois resíduos participou diretamente do acoplamento de pelo menos três dos quatro complexos formados, assim como alguns dos resíduos mais próximos.

5 CONCLUSÃO

Com esse trabalho os sítios de ligação das proteínas virais NS1, NS3, NS5, envelope e da lactoferrina foram identificados, o que permitiu uma análise do grau de interação entre essas proteínas através do *docking*. A proteína envelope mostrou-se como o mais provável alvo da lactoferrina nas células quando comparada com as demais proteínas, particularmente, por ter demonstrado um resultado consistente em mais de uma plataforma de *docking*. No entanto, ainda não é possível afirmar que os resultados positivos obtidos com a lactoferrina em testes de laboratório sejam devidos a inativação de uma das proteínas utilizadas nesse trabalho, muito embora seja provável. Tendo em vista que a lactoferrina representa uma recente abordagem contra o ZKV e o seu mecanismo molecular de ação ainda é desconhecido, trabalhos *in silico* podem ajudar nesse processo de elucidação.

Como perspectivas futuras, outras abordagens computacionais podem se somar aos resultados desse trabalho. A dinâmica molecular pode oferecer uma visão de maior precisão do comportamento desses complexos em um ambiente mais próximo às condições *in vivo*. Além disso, seria possível dar continuação a esse trabalho utilizando as proteínas estruturais e não-estruturais do Zika vírus que não foram utilizadas até então.

REFERÊNCIAS

- ABUSHOUK, A. I.; NEGIDA, A.; AHMED, H. An updated review of Zika virus. **Journal of Clinical Virology**, v. 84, p. 53–58, 2016.
- BACHMAN, J. **Site-Directed Mutagenesis**. [s.l.] Elsevier Ltd. v. 529, 2013.
- BERMAN, H. M.; WESTBROOK, J.; FENG, Z.; GILLILAND, G.; BHAT, T. N.; WEISSIG, H.; SHINDYALOV, I. N.; BOURNE, P. E. The protein data bank. **Nucleic acids research**, v. 28, n. 1, p. 235 – 242, 2000.
- CARVALHO, C. A. M.; CASSEB, S. M. M.; GONÇALVES, R. B.; SILVA, E. V. P.; GOMES, A. M. O.; VASCONCELOS, P. F. C. Bovine lactoferrin activity against chikungunya and zika viruses. **Journal of General Virology**, v. 98, n. 7, p. 1749–1754, 2017.
- CHOUIN-CARNEIRO, T.; VEGA-RUA, A.; VAZEILLE, M *et al.* Differential Susceptibilities of *Aedes aegypti* and *Aedes albopictus* from the Americas to Zika Virus. **PLoS Neglected Tropical Diseases**, v. 10, n. 3, p. 1–11, 2016.
- DE VRIES, S. J.; VAN DIJK, M.; BONVIN, A. M. J. J. The HADDOCK web server for data-driven biomolecular docking. **Nature Protocols**, v. 5, n. 5, p. 883–897, 2010. Disponível em: <<http://www.nature.com/doifinder/10.1038/nprot.2010.32>>.
- DELVECCHIO, R.; HIGA, L. M.; PEZZUTO, P *et al.* Chloroquine inhibits Zika Virus infection in different cellular models. **bioRxiv**, p. 51268, 2016. Disponível em: <<http://biorxiv.org/lookup/doi/10.1101/051268>>.
- DOMINGUEZ, C.; BOELEN, R.; BONVIN, A. M. J. J. HADDOCK: A protein-protein docking approach based on biochemical or biophysical information. **Journal of the American Chemical Society**, v. 125, n. 7, p. 1731–1737, 2003.
- EKINS, S.; MIETCHEN, D.; COFFEE, M *et al.* Open drug discovery for the Zika virus.

F1000Research, v. 5, n. 0, p. 150, 2016.

FARIA, N. R.; AZEVEDO, R. d. S. d. S.; KRAEMER, M. U. G. *et al.* Zika virus in the Americas: Early epidemiological and genetic findings. **Science**, v. 352, n. 6283, p. 345–349, 2016.

GUZZETTA, G.; POLETTI, P.; MONTARSI, F. *et al.* S. Assessing the potential risk of Zika virus epidemics in temperate areas with established *Aedes albopictus* populations. **Eurosurveillance**, v. 21, n. 15, p. 30199, 2016.

HARROWER, J.; KIEDRZYNSKI, T.; BAKER, S *et al.*. Sexual transmission of Zika Virus and persistence in Semen, New Zealand, 2016. **Emerging Infectious Diseases**, v. 22, n. 10, p. 1855–1857, 2016.

HAYES, E. B. Zika virus outside Africa. **Emerging Infectious Diseases**, v. 15, n. 9, p. 1347 – 1350, 2009.

KIM, J.; KWON, J.; KIM, M.; DO, J.; LEE, D.; HAN, H. Low-dielectric-constant polyimide aerogel composite films with low water uptake. **Polymer Journal**, v. 48, n. 7, p. 829 – 834, 2016.

KORTEMME, T.; KIM, D. E.; BAKER, D. Computational Alanine Scanning of Protein-Protein Interfaces. **Science Signaling**, v. 2004, n. 219, p. p12-p12, 2004.

MCCARTHY, M. US health officials investigate sexually transmitted Zika virus infections. **Bmj**, v. 1180, n. February, p. i1180, 2016.

NUGENT, E. K.; NUGENT, A. K.; NUGENT, R.; NUGENT, K. Zika Virus: Epidemiology, Pathogenesis and Human Disease. **American Journal of the Medical Sciences**, v. 353, n. 5, p. 466 – 473, 2017.

PETERSEN, L. R.; JAMIESON, D. J.; POWERS, A. M.; HONEIN, M. A. Zika Virus. **New England Journal of Medicine**, v. 374, n. 16, p. 1552–1563, 2016

RASTOGI, N.; SINGH, A.; PANDEY, S. N *et al.* Structure of the iron-free true C-terminal half of bovine lactoferrin produced by tryptic digestion and its functional significance in the gut. **FEBS Journal**, v. 281, n. 12, p. 2871–2882, 2014.

SCHMIDT, C. W. Zika in the United States how are we preparing? **Environmental Health Perspectives**, v. 124, n. 9, p. A158 – A165, 2016.