



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIENCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUIMICA FUNDAMENTAL
PROGRAMA DE POS-GRADUACAO EM QUIMICA

DANILO FERNANDES COELHO

ENGENHARIA, CARACTERIZAÇÃO ESTRUTURAL E AVALIAÇÃO
IMUNOLÓGICA DE ANTÍGENOS DE DENV, ZIKV E HIV-1

Recife

2018

DANILO FERNANDES COELHO

ENGENHARIA, CARACTERIZAÇÃO ESTRUTURAL E AVALIAÇÃO
IMUNOLÓGICA DE ANTÍGENOS DE DENV, ZIKV E HIV-1

Tese de doutorado apresentada à
coordenação do curso de Pós-graduação
em Química, como um dos requisitos para a
obtenção do título de Doutor em Química.

Área de concentração: Química Teórica

Orientador: Prof. Dr. Roberto Dias Lins

Recife
2018

Catálogo na fonte
Bibliotecário Jefferson Luiz Alves Nazareno CRB4-1758

C672 Coêlho, Danilo Fernandes.
Engenharia, caracterização estrutural e avaliação imunológica de
antígenos de DENV, ZIKV e HIV-1 / Danilo Fernandes Coêlho . – 2018.
194f. : fig., tab.

Orientador: Roberto Dias Lins
Tese (Doutorado) – Universidade Federal de Pernambuco. CCEN.
Química Fundamental. Recife, 2018.
Inclui referências e apêndices.

1. Química teórica. 2. Zika. 3. Dengue. 4. HIV-1. I. Lins, Roberto Dias .
(Orientador). II. Título.

540.1

CDD (22. ed.)

UFPE-FQ 2018-60

DANILO FERNANDES COELHO

***ENGENHARIA, CARACTERIZACAO ESTRUTURAL E
AVALIACAO IMUNOLOGICA DE ANTIGENOS DE DENV,
ZIKV E HIV-1.***

Tese apresentada ao Programa de Pós-Graduação no Departamento de Química Fundamental da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutor em Química.

Aprovado em: 09/08/2018.

BANCA EXAMINADORA

Prof. Roberto Dias Lins Neto (Orientador)

Instituto Aggeu Magalhães
Fiocruz

Prof. Richard Charles Garratt

Instituto de Física de São Carlos
Universidade de São Paulo

Prof. Diogo Ardaillon Simões

Departamento de Bioquímica
Universidade Federal de Pernambuco

Profa. Isabelle Freire Tabosa Viana

Instituto Aggeu Magalhães
Fiocruz

Prof. Rafael Dhalia

Instituto Aggeu Magalhães

AGRADECIMENTOS

À CAPES pelo apoio financeiro para realização desse trabalho.

Ao Departamento de Química Fundamental e ao Instituto Aggeu Magalhães pelo suporte técnico e infraestrutura.

Ao professor Dr. Roberto Dias Lins pela orientação, dedicação e pelos ensinamentos.

Aos colegas do BIOMAT-DQF no nome da Prof^a Thereza pelos eventuais debates durante essa jornada.

Aos colegas do BIOMAT-FIOCRUZ Carlos, Matheus, Marjorie, Maíra, Madson, Laércio, Catarina e Bruno que colaboraram na construção e debate deste trabalho.

Aos colegas de bancada Bruna, Deborah e Fernando pelas preciosas colaborações nos experimentos e ensinamentos na área de Biologia Molecular.

Aos pesquisadores da Fiocruz-PE, em especial Dr. Rafael Dhalia, Dr. Ernesto Marques e Dr^a. Marli Tenório, pelas discussões e colaborações.

Ao Dr. Ricardo Durães pela contribuição e colaboração no alinhamento e seleção de epítomos putativos de Zika.

À Dr^a. Isabelle Viana pelos ensinamentos, colaborações e discussões que foram essenciais para conclusão deste trabalho.

À Universidade de Pittsburgh, em nome do Dr. Ernesto Marques e Dr. Eduardo Nascimento, por terem me recebido e colaborado na realização deste trabalho.

À minha família pelo carinho e apoio de sempre. À Carolina pelo incentivo e compreensão.

Aos colegas do LAC-UFPE pelos momentos de descontração.

A todos que de alguma forma estiveram presentes em minha jornada e contribuíram para a realização deste trabalho.

RESUMO

Utilizando métodos de engenharia de proteínas e simulação por dinâmica molecular (MD), o propósito deste trabalho foi o desenvolvimento *in silico* e caracterização de proteínas/peptídeos para uso como antígenos de diagnóstico ou como biomarcador. Os vírus alvos neste estudo foram os vírus da dengue (DENV), zika (ZIKV) e HIV-1. Através da estratégia de *threading* seguida do *de novo design* e análise por MD, elementos estruturais da proteína NS1 de DENV foram transplantados para a Top7. Foram desenvolvidas 3 quimeras: 2S20-30_Top7, 7S22-24_Top7 e 8S22-25_Top7. As três proteínas foram avaliadas imunologicamente por ensaios de ELISA e os resultados sugerem que dois antígenos são reconhecidos por anticorpos humanos anti-DENV1 e 2 e reagem moderadamente com anti-DENV-3 e 4, reproduzindo dados da literatura. A metodologia aplicada permite uma triagem evitando o uso dispendioso de alocação computacional e insumos de laboratório. Na segunda parte do trabalho, foi avaliada a estabilidade estrutural da proteína Top7-2F5 após reação de biotinylation. Análises de dicroísmo circular e de MD indicam que o processo de biotinylation não compromete a estrutura nativa da Top7-2F5. Os dados também indicam diferentes propriedades eletrostáticas que induzem a uma menor capacidade de ser reconhecida pelo mAb 2F5 e induz maior solubilidade das proteínas biotiniladas. Na terceira parte, através do método *MotifGraft*, seguido de análise de MD, dois epítomos da proteína envelope (E) e um da NS4b de ZIKV foram carregados nas proteínas FNIII e TL1A. Também foi identificado um peptídeo baseado na proteína E de ZIKV, com baixo grau de similaridade com os demais flavivírus. A parte C-terminal do peptídeo conserva conformação nativa, possibilitando reconhecimento por anticorpos anti-E-ZIKV. O peptídeo foi utilizado para obtenção de anticorpos policlonais específicos para ZIKV para o desenvolvimento de um ensaio de ELISA sanduíche.

Palavras-chave: Zika. Dengue. HIV-1. Dinâmica molecular. Rosetta.

ABSTRACT

Using protein engineering and molecular dynamics (MD) methods, the purpose of this work was to develop proteins/peptide *in silico* and to experimentally characterize their use as potential diagnostic antigens or as a biomarker. The target viruses in this study were dengue virus (DENV), zika virus (ZIKV) and HIV-1. Through the threading strategy followed by de novo design and MD analysis, structural elements of the DENV NS1 protein were transplanted to Top7 scaffold. Three chimeras were developed: 2S20-30_Top7, 7S22-24_Top7 and 8S22-25_Top7. The three proteins were immunologically evaluated by ELISA assays and the results suggest that two antigens are recognized by anti-DENV1 and 2 human antibodies, but react moderately with anti-DENV-3 and 4, according to literature data. The applied methodology allows screening of new antigens, avoiding the expensive use of computational allocation and laboratory work. In the second part of the work, the structural stability of the Top7-2F5 protein was evaluated after biotinylation reaction. Results of circular dichroism and MD indicate that the biotinylation process does not compromise the native structure of Top7-2F5-Biotin. The data also indicate different electrostatic properties that induce a lower capacity to be recognized by mAb 2F5 and induce greater solubility of the biotinylated proteins. In the third part, protein engineering was used to develop peptides and proteins harboring ZIKV epitopes for diagnostic and / or vaccine purposes. Using MotifGraft method, followed by MD analysis, two epitopes of envelope protein (E) and one of ZIKV NS4b were loaded onto the FNIII and TL1A proteins. A peptide based on the ZIKV E protein was also identified, with a low degree of similarity to the other flaviviruses. The C-terminal part of the peptide conserves native conformation, allowing recognition by anti-E-ZIKV antibodies. The peptide was used to obtain polyclonal antibodies specific for ZIKV for the development of a sandwich ELISA assay.

Keywords: Zika. Dengue. HIV-1. Molecular dynamics. Rosetta.

LISTA DE TABELAS

Tabela 1 -	Sequências da NS1 de DENV selecionadas a partir do gráfico de hidropatia de Kyte-Doolittle.....	82
Tabela 2 -	Proteínas construídas usando a Top7 como <i>scaffold</i> através do protocolo de <i>threading</i>	84
Tabela 3 -	Sequência de aminoácidos das proteínas construídas utilizado os epítomos 2, 7 e 8 e a Top7 como <i>scaffold</i> . A região do epítomo encontra-se destacada em negrito.	86
Tabela 4 -	Sequências selecionadas para o <i>design</i> de antígenos de ZIKV.	126
Tabela 5 -	Proteína quimérica desenhada contendo o epítomo de NS4b de ZIKV.....	132
Tabela 6 -	Proteínas desenvolvidas usando os epítomos E2 e E3 de proteína E de ZIKV. A região do epítomo se destaca em negrito, a região deletada do scaffold em vermelho.....	143

LISTA DE FIGURAS

- Figura 1 - Aminoácidos naturalmente encontrados nos seres vivos. Podem ser separados em grupos de acordo com suas propriedades físico-químicas. Retirado de (Verli, 2014)..... 29
- Figura 2 - Modelo do funil de enovelamento de uma proteína. As proteínas têm uma superfície energética com muitas estruturas de alta energia, desenoveladas e apenas algumas estruturas enoveladas de baixa energia. Retirado de (Dill e Maccallum, 2012)..... 30
- Figura 3 - Algoritmo de *de novo design* utilizado pelo Rosetta. (a) Bibliotecas de fragmentos com ângulos ϕ e ψ que capturam o espaço conformacional acessível. (b) A combinação de fragmentos diferentes de forma randômica e (c) posterior minimização num processo de Monte Carlo/Metropolis resulta em uma estrutura de mínima energia. Retirado de (Kaufmann *et al.*, 2010)..... 32
- Figura 4 - Comparação do modelo projetado computacionalmente (azul) com a estrutura de raios X resolvida (vermelho) da Top7. Retirado de (Kuhlman *et al.*, 2003)..... 33
- Figura 5 - Caracterização bioquímica de nucleocapsídeos sintéticos. (a) Modelo desenhado computacionalmente. (b) Os nucleocapsídeos sintéticos encapsulam seus próprios genomas enquanto se montam. (c) Microscopia eletrônica dos capsídeos obtidos experimentalmente. Retirado de (Butterfield *et al.*, 2017) 34
- Figura 6 - Método computacional para projetar proteínas com flexibilidade total do *backbone* desenvolvido por Correia e colaboradores. Método foi utilizado no desenvolvimento de uma vacina baseado no epítipo do vírus RSV. A imunização de macacos rhesus induziu anticorpos neutralizantes de longa duração de RSV. Retirado de (Correia *et al.*, 2014)..... 35

Figura 7 -	Representação em <i>cartoon</i> de estruturas das quimeras Top7-HIV-1, desenhadas por Viana e colaboradores, após 50 ns de simulação de dinâmica molecular. As regiões do epítipo estão em amarelo. Retirado de (Viana, I. <i>et al.</i> , 2013).	37
Figura 8 -	(a) Valores de <i>score</i> de hidropatia para cada tipo de aminoácido. Os valores foram obtidos de (Kyte e Doolittle, 1982). (b) representação esquemática de como funciona o algoritmo de Kyte-Doolittle; o valor final é atribuído ao aminoácido do meio da janela.	39
Figura 9 -	Comparação entre energia de Van der Waals (Lennard-Jones) tradicional (vermelho) e método do Rosetta (azul). (A) Termo repulsivo da energia de van der Waals; (B) termo de atração e; (C) Energia eletrostática de Coulomb clássica (vermelho) em comparação com Rosetta <i>fa_elec</i> (azul). Dados representam interações com o átomo de nitrogênio do <i>backbone</i> . Adaptado de (Alford <i>et al.</i> , 2017).	42
Figura 10 -	Modelo de solvatação de Lazaridis-Karplus. O Rosetta usa dois termos de energia para avaliar a dessolvatação de cadeias laterais de proteínas: (A) um termo isotrópico (<i>fa_sol</i>) e (B) um termo anisotrópico. Adaptado de (Alford <i>et al.</i> , 2017).	45
Figura 11 -	Modelo de ligação de hidrogênio dependente da orientação. (A) Graus de liberdade avaliados pelo termo de ligação de hidrogênio: distância entre os átomos <i>acceptor-doador</i> d_{HA} ; o ângulo entre átomo base, o <i>acceptor</i> e o hidrogênio, θ_{BAH} ; o ângulo entre o <i>acceptor</i> , hidrogênio e doador, θ_{AHD} ; e o ângulo diédrico correspondente à rotação em torno da ligação base- <i>aceitador</i> , ϕ_{B2BAH} . (D - F) Contribuições de cada grau de liberdade para a energia de ligações de hidrogênio. Adaptado de (Alford <i>et al.</i> , 2017).	46
Figura 12 -	(A) Escaneamento e substituição das cadeias laterais no <i>template</i> . (B) ângulos χ das cadeias laterais otimizados durante o empacotamento. (C) biblioteca de rotâmeros com as possíveis conformações de cada cadeia lateral.	50

Figura 13 -	O <i>AbinitioRelax</i> se baseia em (A) uma biblioteca de fragmentos que vão (B) se encaixando através da substituição dos valores (D) dos ângulos ψ e ϕ , formando um (C) funil de enovelamento. Adaptado de (Kaufmann <i>et al.</i> , 2010).....	52
Figura 14 -	Fluxograma demonstrando o passo a passo do protocolo de <i>MotifGraft</i> . Adaptado de (Silva <i>et al.</i> , 2016).....	53
Figura 15 -	Exemplo de rotina criada para o <i>MotifGraft</i> utilizando o <i>RosettaScripts</i>	55
Figura 16 -	Estrutura cristalográfica da NS1 de DENV-1 (PDB ID: 4OIG). (a) Face estruturada e (b) face não-estruturada.	71
Figura 17 -	Gráfico de hidropatia de Kyte-Doolittle para a região dos resíduos 172-352 da NS1 de DENV-1.....	81
Figura 18 -	Representação <i>cartoon</i> das proteínas utilizadas como <i>scaffolds</i> . Folhas- β estão representadas como setas na cor vermelha, hélices- α como espirais na cor azul e regiões não estruturadas na cor branca.....	83
Figura 19 -	RMSD em função do Rosetta <i>score</i> para a quimera 2S20-30_Top7 construída utilizando o epítipo 2 e a Top7. A linha em vermelho indica a energia mínima global do perfil de enovelamento, neste caso de -248 REU.	85
Figura 20 -	Rosetta <i>Score</i> em função do RMSD das proteínas 7S22-24_Top7 e 8S22-25_Top7. O RMSD foi calculado em relação às estruturas de menor energia de cada <i>ensemble</i> , e ambas formam um funil de enovelamento, indicando serem estruturas nativas e proteínas estáveis.	87
Figura 21 -	Representação <i>cartoon</i> das estruturas de menor energia das proteínas (A) 2S20-30_Top7, (B) 7S22-24_Top7 e (C) 8S22-25_Top7. Folhas- β estão representadas como setas na cor vermelha, hélices- α como espirais na cor azul e regiões não estruturadas na cor branca. As regiões correspondentes aos epítipos estão destacadas em amarelo.....	88
Figura 22 -	(A) RMSD dos átomos do <i>backbone</i> da proteína Top7 em função do tempo, convergindo para um RMSD médio de 0,15 nm. (B) Conteúdo de estrutura secundária da Top7 em função	

	do tempo de simulação, indicando que não há perda considerável de conteúdo estruturado e a proteína se mantém estável.	89
Figura 23 -	(A) RMSD dos átomos do <i>backbone</i> da quimera 2S20-30_Top7 (vermelho) convergem para um valor de 0.25 nm, enquanto que o backbone do epítipo (azul) converge para 0.4 nm. (B) Conteúdo de estrutura secundaria da quimera em função do tempo de simulação, indicando a perda parcial de sua primeira hélice- α , na região onde o epítipo foi alocado (destacado em amarelo).....	90
Figura 24 -	(A) RMSD dos átomos do <i>backbone</i> da quimera 7S22-24_Top7 (vermelho) convergem para um valor de 0.25 nm, enquanto que o backbone do epítipo (azul) converge para 0.35 nm. (B) Conteúdo de estrutura secundaria da quimera em função do tempo de simulação, não exhibe nenhuma perda de motivo estrutural. Região do epítipo destacada em amarelo.....	91
Figura 25 -	(A) RMSD dos átomos do <i>backbone</i> da quimera 8S22-25_Top7 (vermelho) convergem para um valor de 0.28 nm, enquanto que o backbone do epítipo (azul) converge para 0.4 nm. (B) Conteúdo de estrutura secundaria da quimera em função do tempo de simulação, não exhibe nenhuma perda de motivo estrutural, exceto flutuações na região do epítipo destacada em amarelo.....	92
Figura 26 -	Representação <i>cartoon</i> das quimeras (A) 2S20-30_Top7, (B) 7S22- 24_Top7, e (C) 8S22-25_Top7, após 100 ns de simulação. Folhas- β estão representadas como setas na cor vermelha, hélices- α como espirais na cor azul e regiões não estruturadas na cor branca. As regiões correspondentes aos epítipos estão destacadas em amarelo.....	92
Figura 27 -	SDS-PAGE das quimeras obtidas por expressão bacteriana e purificadas por cromatografia de afinidade. (A) Top7 original, obtida com alto rendimento. (B) Quimera 2S20-30_Top7 obtida com baixas taxas de rendimento, visto que durante a MD observou-se perda de estrutura nativa. (C) A 7S22-24_Top7	

também foi obtida em grande quantidade, indicando ser uma proteína bastante solúvel. (D) 8S22-25_Top7 com rendimento moderado. Bandas de proteínas provenientes de bactéria são observadas nas eluições das proteínas de baixo e médio rendimento..... 94

Figura 28 - (A) Ensaio de ELISA do antígeno 2S20-30_Top7, exibindo uma distinção clara e significativa entre as populações positivas e negativas para DENV ($p < 0.05$), mas não se observa essa diferença quando compara-se os sinais da 2S20-30_Top7 com os da Top7 original. (B) Curva ROC do ensaio, com AUC = 0.96. 95

Figura 29 - (A) Ensaio de ELISA do antígeno 7S22-24_Top7, com bom desempenho diagnóstico, exibindo distinção significativa entre as populações positivas e negativas para DENV, e em comparação aos sinais da Top7 ($p < 0.05$). (B) Curva ROC do ensaio com a proteína 7S22-24_Top7, indicando sensibilidade de 71.43 % e especificidade de 100%, e AUC = 0.7857..... 96

Figura 30 - Especificidade dos sorotipos de anticorpos humanos anti-NS1 obtidos de pacientes com infecção secundária de DENV-2. Células Vero foram infectadas com DENV-1, DENV-2, DENV-3, DENV-4 ou JEV (vírus encefalite japonesa). A reatividade de cada um dos mAbs com células infectadas com cada vírus foi então examinada por microscopia de fluorescência. Adaptado de (Omokoko *et al.*, 2014). 97

Figura 31 - (A) Ensaio de ELISA do antígeno 8S22-25_Top7 exibindo distinção significativa entre as populações positivas e negativas para DENV, e em comparação aos sinais da Top7 ($p < 0.05$) (Figura 3.16 A). (B) O ensaio obteve sensibilidade de 69.23 % e especificidade de 100% e AUC = 0.7692. 98

Figura 32 - (A) Proteína quimérica baseada na Top7 contendo o epítopo 2F5 construída por Viana e colaboradores (Viana, I. *et al.*, 2013). (B) Representação da reação da Top7-2F5 com a molécula de biotina. A parte linear terminando no grupo

	carboxila é responsável pela reação com aminas primárias presentes em cadeias de proteínas.....	101
Figura 33 -	(A) Imunorreatividade das proteínas Top7-2F5 e Top7-2F5-Biotina foram avaliadas em diferentes períodos de tempo. Ambas as proteínas possuem imunorreatividade semelhantes quando utilizadas assim que purificadas, com uma leitura ligeiramente inferior para Top7-E2F5-Biotina. (B) Concentração das proteínas Top7-2F5 e Top7-2F5-Biotina em diferentes períodos de tempo. Dados obtidos pela colaboradora Dr ^a Isabelle Viana (Coelho, In preparation).	102
Figura 34 -	Espectro de CD das proteínas (A) Top7 e Top7-Biotina, e (B) Top7-2F5 e Top7-2F5-Biotina. Todos os quatro espectros são característicos de proteínas α / β . Os espectros revelam duas bandas negativas características do conteúdo de hélice α em 208 nm e 222 nm e uma banda negativa correspondente ao conteúdo em folha β entre 216 e 218 nm. A comparação entre os espectros das proteínas Top7 e Top7-2F5 com suas versões biotiniladas sugerem semelhanças no <i>folding</i> e manutenção da estrutura das proteínas.....	108
Figura 35 -	Valores de RMSD médios em função da temperatura de simulação para cada proteína. Após biotinilação, os valores dos RMSD sistematicamente foram menores do que para as proteínas nativas, indicando uma estabilidade estrutural conferida pela biotina.....	110
Figura 36 -	A flutuação média quadrática posicional (RMSF) para os átomos do <i>backbone</i> . (A) Na Top7 nativa, as regiões de maiores flutuações atômicas correspondem a regiões de <i>turn/loops</i> e motivos em α -hélice. Os valores de RMSF aumentam com o aumento da temperatura, Por outro lado, o RMSF da Top7-Biotina permaneceu abaixo de 0,2 nm independentemente da temperatura. (B) Na Top7-2F5, os maiores valores de RMSF corresponderam a regiões de <i>turn/loops</i> e motivos em α -hélice, com valores de RMSF aumentando com o aumento da temperatura. Na Top7-2F5-biotina os valores de RMSF a foram	

geral ligeiramente inferiores quando comparados com os da Top7-2F5. O *loop* do epítopo 2F5 torna-se mais flexível na proteína Top7-2F5-Biotina devido a uma biotinilação no resíduo 25LYS localizada no meio do epítopo. 112

Figura 37 - Perfis de estrutura secundária dependentes do tempo (DSSP) para os quatro sistemas, indicando que todas as estruturas secundárias foram bem mantidas em todas as trajetórias a 300 K. À medida que a temperatura de simulação aumenta, as proteínas nativas (Top7 e Top7-2F5) perdem conteúdo estrutural. Em contraste, o perfil de DSSP das proteínas biotiniladas sofrem menos perturbações com o aumento da temperatura, indicando que a biotina confere estabilidade estrutural..... 114

Figura 38 - Distribuição populacional dos valores de RMSD entre os átomos do *backbone* do epítopo na proteína (A) Top7-2F5 e (B) Top7-2F5-Biotina. Os valores são em relação à conformação do epítopo quando complexado ao anticorpo mAb 2F5 (PDB 1TJI), indicando que o epítopo não-biotinilado possui uma conformação preferencial, enquanto que o biotinilado possui duas conformações. 115

Figura 39 - Superfície de potencial eletroestático das proteínas Top7-2F5 e Top7-2F5-Biotina. A localização do epítopo está destacada em verde. A Top7-2F5-Biotin mostra um perfil notavelmente mais negativo, enquanto que a Top7-2F5 exibe uma distribuição de regiões positiva e negativa mais homogênea..... 116

Figura 40 - Representação em *cartoon* da estrutura da NS1 de ZIKV. As cadeias de proteínas são mostradas em branco e cinza. Os epítomos putativos são mapeados como a superfície de estrutura de arame e cor codificados como E1 em amarelo, E2 em verde, E3 em ciano, E4 em azul, E5 em laranja, E6 em vermelho e E7 em magenta. Retirado de (Freire *et al.*, 2017). 120

Figura 41 - Representação <i>cartoon</i> da proteína Envelope do ZIKV (PDB ID 5GZN). Domínio I (azul), domínio II (vermelho) e domínio III (amarelo).	121
Figura 42 - Motivos estruturais das sequências selecionadas como epítomos putativos de ZIKV. (A) Resíduos 144-166 da proteína E, epítopo E1; (B) Resíduos 158-168 da proteína E, epítopo E2; e (C) Resíduos 364-376 da proteína E, epítopo E3.	126
Figura 43 - Representação <i>cartoon</i> das proteínas <i>scaffolds</i> utilizadas para o desenho de quimeras contendo o epítopo NS4b. (A) Top7 (PDB ID 1QYS) e (B) Fibronectina tipo III humana (PDB ID 1FNA).	127
Figura 44 - Representação <i>cartoon</i> da Top7 contendo o epítopo de NS4b entre os resíduos 22 e 23. (A) Estrutura inicial obtida com o protocolo KIC, mostrando a região do epítopo não-estruturada e completamente exposta ao solvente. (B) Estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítopo adota uma conformação de folha- β e desestabiliza a hélice- α adjacente ao epítopo.....	127
Figura 45 - Análise da MD da proteína Top7 contendo epítopo de NS4b entre os resíduos 22 e 23. (A) O RMSD (gráfico superior) permanece com valor médio de 0.3 nm durante toda a simulação, e o RMSF (gráfico inferior) indica uma proteína com baixa flexibilidade, exceto na região do epítopo. (B) O conteúdo de estrutura secundária durante a simulação indica mudança de conformação nativa do epítopo (destacado em amarelo) , ao mesmo tempo que desestabiliza a hélice- α adjacente ao epítopo.....	128
Figura 46 - Representação <i>cartoon</i> da Top7 contendo o epítopo de NS4b entre os resíduos 42 e 44. (A) Estrutura inicial mostrando a região do epítopo não-estruturada e completamente exposta ao solvente. (B) Estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítopo mantém a sua conformação não-estruturada mas desestabilizando completamente a hélice- α adjacente ao epítopo e parcialmente a segunda folha- β da proteína.....	129

- Figura 47 - Análise da MD da proteína Top7 contendo epítipo de NS4b entre os resíduos 42 e 44. (A) O RMSD (gráfico superior) adota valores altos a partir de 25 ns de simulação, sugerindo uma desestabilização da estrutura. O RMSF (gráfico inferior) indica duas regiões de alta flexibilidade da proteína, uma sendo o epítipo e a outra a hélice- α adjacente. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa do epítipo (destacado em amarelo), ao mesmo tempo que desestabiliza a hélice- α adjacente ao epítipo..... 130
- Figura 48 - Representação *cartoon* da FNIII contendo o epítipo de NS4b. (A) Estrutura inicial mostrando a região do epítipo não-estruturada e completamente exposta ao solvente. (B) Estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação não-estruturada e nenhuma perturbação estrutural no *scaffold*. 131
- Figura 49 - Análise da MD da proteína FNIII contendo epítipo de NS4b entre a 2ª e 3ª folha- β . (A) O RMSD (gráfico superior) adota valores relativamente baixos, sugerindo uma proteína sem perturbação estrutural. O RMSF (gráfico inferior) indica apenas uma região de alta flexibilidade da proteína, onde foi colocado o epítipo. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítipo (destacado em amarelo), como de todo o *scaffold*.... 131
- Figura 50 - Proteína TL1A humana utilizada como *scaffold*. Pertence à família de citocinas capazes de provocar morte de células cancerígenas. PDB ID 2RJL..... 132
- Figura 51 - Representação *cartoon* da Top7 contendo o epítipo E2 após alinhamento com o *MotifGraft*. (A) Estrutura inicial mostrando a região da folha- β onde o epítipo foi transplantado (destacado em amarelo) e (B) estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação nativa, porém desestabilização da hélice- α adjacente. 133

- Figura 52 - Análise da MD da proteína Top7 contendo epítipo E2 de ZIKV na 2ª folha- β . (A) O RMSD (gráfico superior) assume valores acima de 0.3 nm, sugerindo uma mudança de conformação da proteína, corroborada pelos altos valores de RMSF (gráfico inferior) entre os resíduos 20 e 40. (B) O conteúdo de estrutura secundária durante a simulação indica desenovelamento parcial da primeira hélice- α , assim como instabilidade da folha- β onde o epítipo foi transplantado (destacado em amarelo). 134
- Figura 53 - Representação *cartoon* da FNIII contendo o epítipo E2 após alinhamento com o *MotifGraft*. (A) Estrutura inicial mostrando a região da folha- β /*Coil* onde o epítipo foi transplantado (destacado em amarelo) e (B) estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação nativa, e o *scaffold* não apresenta nenhuma mudança considerável de estrutura secundária. 135
- Figura 54 - Análise da MD da proteína FNIII contendo o epítipo E2 da proteína E de ZIKV na 4ª folha- β . (A) O RMSD adota valores baixos, com média de 0.25 nm, sugerindo uma proteína estruturalmente estável. O RMSD do epítipo (curva vermelha) também sugere baixa mudança de conformação da estrutura nativa. O RMSF indica apenas uma região de alta flexibilidade da proteína, onde se encontra uma região em *loop*. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítipo (destacado em amarelo), como de todo o *scaffold*. 135
- Figura 55 - Gráfico comparando a superfície de acessibilidade ao solvente (SASA) entre o epítipo em sua conformação nativa (preto) e durante a trajetória de MD transplantado na FNIII (vermelho). Os resíduos estão tão expostos como quando em sua estrutura nativa, indicando estarem aptos para reconhecimento. Os resíduos N, A e V não foram inseridos no *scaffold*. 136
- Figura 56 - Representação *cartoon* da TL1A contendo o epítipo E3 após alinhamento com o *MotifGraft*. (A) Estrutura inicial mostrando a região da folha- β /*Coil* onde o epítipo foi transplantado

(destacado em amarelo) e (B) estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítopo mantém a sua conformação nativa, e o *scaffold* apresenta apenas uma mudança de posição do *loop* central. 137

Figura 57 - Análise da MD da proteína TL1A contendo o epítopo E3 da proteína E de ZIKV. (A) O RMSD (gráfico superior) adota valores relativamente altos, sugerindo alguma modificação estrutural. O RMSF (gráfico inferior) indica duas regiões de alta flexibilidade da proteína, onde se encontram regiões predominantemente em *loop*, cuja alta mobilidade contribui para os altos valores de RMSD. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítopo (destacado em amarelo), como de todo o *scaffold*..... 138

Figura 58 - Representação *cartoon* da TL1A contendo o epítopo E3 após exclusão do *loop* central. (A) Estrutura inicial mostrando a região da folha- β /*Coil* onde o epítopo foi transplantado (destacado em amarelo) e (B) estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítopo mantém a sua conformação nativa..... 139

Figura 59 - Análise da MD da proteína TL1A contendo o epítopo E3 da proteína E de ZIKV após exclusão do *loop* central. (A) O RMSD (gráfico superior) mostra valores mais baixos se comparado com a presença do *loop*. O RMSF (gráfico inferior) indica apenas uma região de alta flexibilidade da proteína, onde se encontram predominantemente *loops*. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítopo (destacado em amarelo), como de todo o *scaffold*..... 139

Figura 60 - Representação *cartoon* da TL1A contendo o epítopo E3 após exclusão do *loop* central e mutações. (A) Estrutura inicial mostrando a região da folha- β /*Coil* onde o epítopo foi transplantado (destacado em amarelo) e (B) estrutura final da

	proteína após 100 ns de simulação de MD, mostrando que o epítopo mantém a sua conformação nativa.....	140
Figura 61 -	Análise da MD da proteína TL1A contendo o epítopo E3 da proteína E de ZIKV após exclusão do <i>loop</i> central e mutações. (A) O RMSD (gráfico superior, curva preta) mostra valores um pouco altos, porém o epítopo se mantém bastante semelhante à sua configuração nativa (curva vermelha). O RMSF (gráfico inferior) indica apenas uma região de alta flexibilidade da proteína, onde se encontram predominantemente <i>loops</i> . (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítopo (destacado em amarelo), como de todo o <i>scaffold</i>	141
Figura 62 -	Gráfico comparando SASA do epítopo E3 em sua conformação nativa (preto) e durante a trajetória de MD transplantado na TL1A (vermelho). Os resíduos estão tão expostos como quando em sua estrutura nativa, indicando estarem aptos para reconhecimento.	142
Figura 63 -	Representação esquemática das vacinas de ChAdOx1 contra ZIKV desenvolvidas por López-Camacho e colaboradores. A vacina prME codifica a região pré-membrana, proteína E, e região trans-membrana; a prME Δ TM exclui apenas a região trans-membrana; a Env exclui apenas a região pré-membrana; e a Env Δ TM exclui ambas as regiões pré-membrana e trans-membrana. Retirado de (López-Camacho <i>et al.</i> , 2018).	144
Figura 64 -	A sequência do peptídeo E1 de ZIKV (Q91KX7) alinhado com regiões homólogas das proteínas E de DENV1 (Q7TGE4), DENV2 (AHB63923), DENV3 (ABA25838), DENV4 (Q80L18), WNV (Q91R00) e YFV (Q89286).	145
Figura 65 -	(A) Representação do motivo estrutural da sequência do peptídeo E1 selecionada na estrutura nativa da proteína E de ZIKV. (B) SASA de cada resíduo do peptídeo na estrutura nativa (barras pretas) em comparação com a acessibilidade ao longo da trajetória de MD (barras vermelhas), indicando a maior acessibilidade no peptídeo isolado.	146

- Figura 66 - (A) O conteúdo de estrutura secundária durante a simulação indica uma transição entre conformações não-estruturadas e conformações em β -hairpin (B) com preferência por conformações sem estrutura definida, como indicado pela estrutura mais representativa após análise de clusters..... 147
- Figura 67 - Análise de RMSD do peptídeo E3 de ZIKV ao longo da trajetória de MD. A região N-terminal possui 3 populações igualmente distribuídas com RMSD em torno de 0.2, 0.3 e 0.4 nm, enquanto que a região C-terminal majoritariamente possui RMSD de 0.2 nm..... 147
- Figura 68 - Dois coelhos (R27 e R34) foram imunizados e os seus soros testados por *western blot* para atestar especificidade na ligação à proteína E de ZIKV (sE). Como esperado, os soros retirados dos animais antes da imunização (pré-imune PI) não reconhecem o envelope. O anticorpo monoclonal mAb 0302156 foi usado como controle positivo. Dados obtidos por colaboradores da Universidade de Glasgow, Escócia e publicados em (López-Camacho *et al.*, 2018). 148
- Figura 69 - (A) Ensaio de MN do ZIKV em camundongos BALB/c 4 semanas e 16 semanas após vacinação. As vacinas prME e prME Δ TM mostraram atividade de neutralização, enquanto que as vacinas Env e Env Δ TM não demonstraram nenhum poder de neutralização. Os resultados mostram a média de três repetições biológicas, com duplicatas e linhas pretas representando a média (n = 6). Análise ANOVA de dois fatores seguida pela comparação múltipla de Sidak foi realizada. **** p <0,0001, *** p <0,0003. (B) A neutralização do vírus também foi avaliada em camundongos CD1, após 4 semanas da vacinação, mostrando novamente que a vacina prME Δ TM é capaz de neutralizar o ZIKV. Adaptado de (López-Camacho *et al.*, 2018). 149
- Figura 70 - Comparação da % de infectividade de ZIKV entre nosso ensaio de MN e o PRNT tradicional em células Vero. O anticorpo monoclonal altamente neutralizante EDE1 C8 foi utilizado em

diferentes concentrações. As linhas representam o ajuste não-linear. Barras de erro são o SD da média de três amostras de réplicas. O ensaio de MN acoplado ao ELISA sanduíche desenvolvidos se mostraram bastante robustos, com uma relação sinal-ruído >10 e um fator Z de 0.8. Adaptado de (López-Camacho *et al.*, 2018). 150

LISTA DE ABREVIACÕES

C	Capsídeo
CHIKV	Chikungunya vírus
DENV	Vírus da dengue
DSSP	<i>Database of secondary structure of proteins</i>
E	Proteína de envelope
ELISA	Ensaio de imunoabsorção enzimática
GBS	Síndrome de Guillain-Barré
HIV	Vírus da imunodeficiência humana
IPTG	Indutor β -1-D-isopropil-tiogalactopiranosida
KIC	Algoritmo <i>kinematic loop closure</i>
mAb 2F5	Anticorpo monoclonal 2F5
MC	Método Monte Carlo
MCM	Método Monte Carlo/Metropolis
MD	Dinâmica molecular
MM	Mecânica molecular clássica
MN	Ensaio de microneutralização
MPER	Região externa proximal à membrana do HIV
mRNA	RNA mensageiro
NPT	<i>Ensemble</i> isotérmico-isobárico
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
NVT	<i>Ensemble</i> isocórico
OMS	Organização Mundial da Saúde
PBC	Condições periódicas de contorno
PDB	Banco de Dados de Proteínas
prM	Proteína pré-membrana
QM	Mecânica quântica

REF	Função de energia do Rosetta
REF15	Função de energia 2015 do Rosetta
REU	Unidade de energia do Rosetta
RMN	Ressonância magnética nuclear
RMSD	<i>Root mean square deviation</i> , desvio quadrático médio
RMSF	Flutuação quadrática média
RSV	Vírus sincicial respiratório
RT-qPCR	Reação da transcriptase reversa seguida da reação em cadeia da polimerase em tempo real
SASA	Superfície de acessibilidade ao solvente
WNV	West Nile vírus
ZIKV	Vírus da zika

SUMÁRIO

1	INTRODUÇÃO	28
1.1	HIPÓTESE	28
1.2	ENGENHARIA DE PROTEÍNAS E DINÂMICA MOLECULAR	28
2	FUNDAMENTAÇÃO TEÓRICA	38
2.1	HIDROPATIA DE KYTE-DOOLITTLE	38
2.2	PREDIÇÃO DE ESTRUTURA DE PROTEÍNAS	39
2.2.1	Função de energia do Rosetta	40
2.2.1.1	Interações De Van Der Waals	41
2.2.1.2	Interações de Coulomb	43
2.2.1.3	Termo de solvatação	44
2.2.1.4	Termo de ligação de hidrogênio	46
2.2.1.5	Termo de Ramachandran	47
2.2.1.6	Termo de design do backbone	47
2.2.1.7	Termo de conformação da cadeia lateral	47
2.2.2	Threading	48
2.2.3	De novo design	50
2.2.4	Motifgraft	52
2.3	MECÂNICA MOLECULAR	55
2.3.1	Potencial de interação de ligação	57
2.3.2	Potencial de ângulo de ligação	57
2.3.3	Potencial de diedro impróprio	58
2.3.4	Potencial torsional	58
2.3.5	Interações De Van Der Waals	59
2.3.6	Interações eletrostáticas	60
2.4	DINÂMICA MOLECULAR	62
2.4.1	Algoritmo de integração <i>leap-frog</i>	64
2.4.2	Controle de temperatura	66
2.4.3	Controle de pressão	67
2.4.4	Condições periódicas de contorno	67
2.5	ENSAIOS DE ELISA	68
3	DESENVOLVIMENTO IN SILICO DE ANTÍGENOS RECOMBINANTES BASEADOS EM NS1 DE DENGUE	70

3.1	OBJETIVOS	73
3.1.1	Objetivos específicos	73
3.2	METODOLOGIA	73
3.2.1	Seleção das sequências de NS1	73
3.2.2	Threading.....	74
3.2.3	De novo design	74
3.2.4	Dinâmica molecular	75
3.3	PREPARAÇÃO DOS VETORES PARA EXPRESSÃO	76
3.3.1	Preparação do vetor de expressão	76
3.3.2	Preparação dos genes sintéticos	77
3.3.3	Subclonagem dos genes no vetor pRSET A	78
3.4	EXPRESSÃO E PURIFICAÇÃO DAS PROTEÍNAS.....	79
3.5	AVALIAÇÃO IMUNOLÓGICA POR ENSAIO DE ELISA	79
3.6	RESULTADOS E DISCUSSÃO	80
3.6.1	Sequências de NS1 selecionadas	80
3.6.2	Quimeras baseadas na TOP7	83
3.6.2.1	Threading e de novo design	83
3.6.2.2	Dinâmica molecular	88
3.6.2.3	Expressão bacteriana e avaliação imunológica por ELISA	93
3.7	CONCLUSÕES	98
4	AVALIAÇÃO ESTRUTURAL DA TOP7-2F5 BIOTINILADA PARA DETECÇÃO DE ANTICORPOS ANTI-2F5 DE HIV-1	100
4.1	OBJETIVOS	103
4.1.1	Objetivos específicos	103
4.2	METODOLOGIA	104
4.2.1	Dinâmica molecular	104
4.2.2	Expressão das proteínas e biotinylação	105
4.2.3	Dicroísmo circular	106
4.3	RESULTADOS E DISCUSSÃO	107
4.3.1	Estabilidade estrutural e dinâmica conformacional	107
4.3.2	Perfil eletroestático da proteína biotinilada	115
4.4	CONCLUSÕES	117

5	DESENVOLVIMENTO <i>IN SILICO</i> DE ANTÍGENOS RECOMBINANTES BASEADOS NAS PROTEÍNAS ENVELOPE E NS4B DE ZIKA.....	118
5.1	OBJETIVOS.....	122
5.1.1	Objetivos específicos	122
5.2	METODOLOGIA	122
5.2.1	Alinhamento sequencial.....	122
5.2.2	Motifgraft.....	123
5.2.3	Dinâmica molecular	124
5.3	RESULTADOS E DISCUSSÃO	125
5.3.1	Sequências selecionadas	125
5.3.2	Proteínas quiméricas contendo epítipo NS4b	126
5.3.3	Proteínas quiméricas contendo epítipos do envelope	132
5.3.4	Peptídeo baseado em envelope de ZIKV para ensaio de ELISA sanduíche	143
5.4	CONCLUSÕES	151
6	CONSIDERAÇÕES FINAIS	152
	REFERÊNCIAS	153
	APÊNDICE A - ARTIGO PUBLICADO EM PERIÓDICO INDEXADO DE CIRCULAÇÃO INTERNACIONAL REFERENTE AO TRABALHO DA TESE.....	164
	APÊNDICE B - PUBLICAÇÃO RELACIONADA AOS RESULTADOS DESTA TESE E QUE ESTÁ EM PROCESSO DE SUBMISSÃO	176
	APÊNDICE C - ARTIGOS PUBLICADOS REFLETINDO APENAS PROJETOS PARALELOS AO LONGO DA TRAJETÓRIA DO DOUTORADO.....	177

1 INTRODUÇÃO

1.1 HIPÓTESE

O uso de técnicas de engenharia de proteínas associadas a cálculos de dinâmica molecular constitui uma abordagem robusta para o desenvolvimento e triagem de antígenos artificiais imunorreativos.

1.2 ENGENHARIA DE PROTEÍNAS E DINÂMICA MOLECULAR

Proteínas são biomoléculas de natureza polimérica que participam de processos químicos, podendo atuar como catalisador, agente de reconhecimento, sinalização e defesa do organismo, dentre outras funções. São obtidas a partir da polimerização de aminoácidos (resíduos), que ocorrem naturalmente em 20 diferentes moléculas (Figura 1) e que podem ser agrupados pelas semelhanças nas suas propriedades. São moléculas bastante versáteis, sendo capazes de adotar diferentes arranjos tridimensionais após seu enovelamento. Por sua vez, a geometria da proteína, juntamente com a apresentação específica de cargas e grupos moleculares em sua superfície, determinam a sua função (Pantazes *et al.*, 2011). Uma determinada sequência de aminoácidos (estrutura primária) terá sempre a mesma estrutura tridimensional final (estrutura terciária) (exceto em condições extremas de pH e temperatura). Por outro lado, diferentes combinações de resíduos com propriedades semelhantes podem levar a uma mesma estrutura final. De fato, sabe-se que a estrutura terciária é mais conservada ao longo da evolução que a estrutura primária, de modo que proteínas com identidade muito baixa entre suas sequências podem possuir estruturas muito semelhantes (Creighton, 1990; Verli, 2014). Em outras palavras, a estrutura tridimensional da proteína é determinada pela sua estrutura primária, i.e. a sequência de aminoácidos que compõem a sua cadeia.

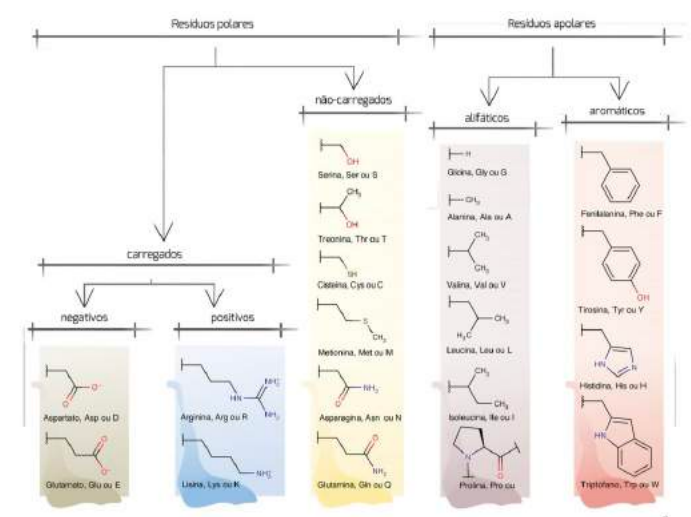


Figura 1 - Aminoácidos naturalmente encontrados nos seres vivos. Podem ser separados em grupos de acordo com suas propriedades físico-químicas. Retirado de (Verli, 2014)

As aplicações das proteínas envolvem seu uso em terapias, diagnósticos, catálise industrial, detecção de contaminantes químicos, etc. Encontrar uma forma de otimizar as proteínas para o uso em aplicações específicas é de grande importância. Para isso, devem ser compreendidos os fundamentos de enovelamento de proteínas e, com isso, ser possível modificar proteínas já existentes ou até mesmo desenhar novas proteínas a partir do zero. O enovelamento é um processo de automontagem e o seu mecanismo de atuação é conhecido como “o problema do enovelamento das proteínas” (Creighton, 1990). De acordo com a Hipótese de Anfinsen, a estrutura final (nativa) que a proteína adota é a estrutura mais estável termodinamicamente, ou seja, o estado de menor energia, e depende apenas da sequência primária de aminoácidos (Anfinsen, 1973; Dill *et al.*, 2008; Huang *et al.*, 2016). Os processos mais importantes que impulsionam as cadeias proteicas para o seu estado nativo são: 1) o empacotamento de resíduos hidrofóbicos no interior da proteína, de modo a minimizar seu contato com o solvente polar; 2) a exposição dos aminoácidos polares e/ou carregados, que têm preferência por interagir com as moléculas do solvente e possíveis íons; 3) minimizar o volume ocupado pela proteína e, por sua vez, maximizar as interações de van der Waals; 4) formação de ligações de hidrogênio intramolecular e entre resíduos polares e moléculas do solvente.

(Dill e Maccallum, 2012; Huang *et al.*, 2016). Como consequência dessas considerações, as proteínas adotam estruturas secundárias em formas de α -hélices e folhas- β , nas quais os grupos carbonila e amida do *backbone* formam ligações de hidrogênio, mantendo a estrutura compacta e as cadeias laterais hidrofóbicas longe do solvente.

O enovelamento segue o princípio da frustração mínima: durante o processo, interações favoráveis ao estado nativo são mantidas em relação às interações desfavoráveis. Dessa forma, estados intermediários, parcialmente enovelados, progressivamente se organizam para formar o estado nativo. Como o estado nativo possui poucas conformações de menor energia se comparado aos estados intermediários, a superfície de energia de enovelamento possui a forma de um funil rugoso (Figura 2) (Levy e Onuchic, 2006; Dill e Maccallum, 2012).

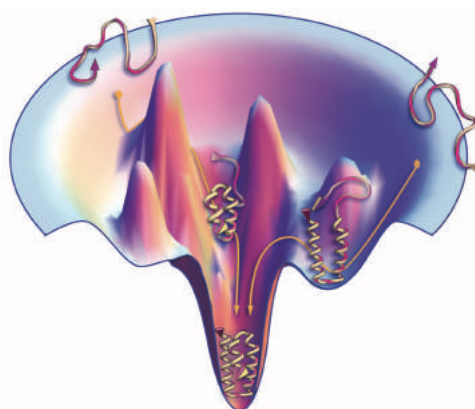


Figura 2 - Modelo do funil de enovelamento de uma proteína. As proteínas têm uma superfície energética com muitas estruturas de alta energia, desenoveladas e apenas algumas estruturas enoveladas de baixa energia. Retirado de (Dill e Maccallum, 2012).

O *design* de proteínas consiste em determinar uma sequência de aminoácidos que adotarão uma estrutura alvo específica e realizarão uma função desejada. Isto requer poder prever a estrutura da nova proteína a partir da sua estrutura primária. As primeiras predições acuradas de estruturas eram modelagens comparativas, ou por homologia. Neste método, a estrutura desconhecida era modelada baseado na similaridade de sua

sequência com proteínas de estruturas já conhecidas (*templates*) (Martí-Renom *et al.*, 2000). A modelagem por homologia consiste basicamente em quatro etapas: 1) encontrar os *templates*, geralmente estruturas conhecidas relacionadas à sequência alvo; 2) alinhar a sequência com os *templates*; 3) construir um modelo; e 4) avaliar o modelo (Baker e Sali, 2001). A modelagem comparativa é possível, pois evolutivamente estrutura é mais conservada do que sequência (Creighton, 1990). Porém, a qualidade do modelo obtido depende fortemente do quão similares são a proteína alvo e o *template* e da sensibilidade da metodologia utilizada para o alinhamento (Chothia e Lesk, 1986; Martí-Renom *et al.*, 2000; Baker e Sali, 2001).

Devido às limitações do método comparativo, novas ferramentas computacionais para predição de estrutura de proteínas vêm sendo desenvolvidas. O método de *de novo design* assume a Hipótese de Anfinsen e utiliza algoritmos de busca e otimização, se baseando numa função de energia e num processo iterativo de *design*/relaxamento da estrutura, determinando a estrutura mais provável como a de menor energia (Gainza *et al.*, 2016). O problema agora passa a ser encontrar uma função de energia adequadamente precisa, bem como métodos para amostragem no espaço de possíveis estruturas e sequências de proteínas (Huang *et al.*, 2016). Modelos de maior acurácia e resolução requerem funções de energia fisicamente mais realistas e melhores algoritmos de busca, que por sua vez requerem maior custo computacional (Bradley *et al.*, 2005; Gainza *et al.*, 2016). Dentre os programas desenvolvidos para este problema destaca-se o Rosetta, que é um pacote de software que inclui algoritmos para modelagem computacional e análise de estruturas de proteínas. Este software tem permitido avanços científicos notáveis em biologia computacional, incluindo design *de novo* de proteína, desenho de enzima, ancoragem de ligante e previsão de estrutura de macromoléculas biológicas e complexos macromoleculares (Rohl *et al.*, 2004; Huang *et al.*, 2016). Para métodos de amostragem, o Rosetta se baseia em fragmentos curtos de proteínas conhecidas, num processo de montagem randômico com estratégia de Monte Carlo (MC) e minimização de energia baseada em algoritmos de gradiente (Figura 3) (Rohl *et al.*, 2004; Kaufmann *et al.*, 2010). O Rosetta utiliza uma função de energia baseada em

dados experimentais e estatístico derivados do *Protein Data Bank* (PDB). Combinando termos de energias tradicionais de mecânica molecular e potenciais estatísticos, a função de energia permitiu que a Rosetta atingisse vários marcos na predição de estrutura de proteínas (Alford *et al.*, 2017). A função de energia possui alta correlação entre dados experimentais de energia livre de mutação $\Delta\Delta G$ e valores preditos pelo Rosetta ($R=0.994$) (Park *et al.*, 2016).

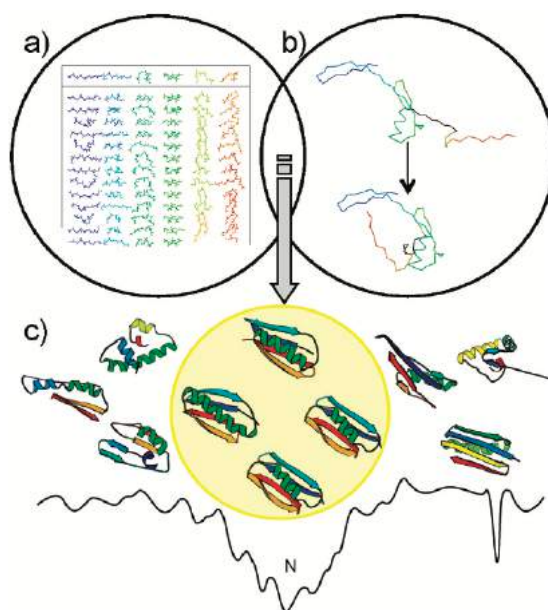


Figura 3 - Algoritmo de *de novo design* utilizado pelo Rosetta. (a) Bibliotecas de fragmentos com ângulos ϕ e ψ que capturam o espaço conformacional acessível. (b) A combinação de fragmentos diferentes de forma randômica e (c) posterior minimização num processo de Monte Carlo/Metropolis resulta em uma estrutura de mínima energia. Retirado de (Kaufmann *et al.*, 2010)

Para demonstrar o potencial computacional do Rosetta, os desenvolvedores criaram *in silico* uma proteína com topologia e sequência não encontradas na natureza, chamada Top7 (Kuhlman *et al.*, 2003). A determinação experimental da estrutura por cristalografia de raios X da Top7 demonstrou desvio quadrático médio de 1,2 Å em relação ao modelo computacional criado (Figura 4).

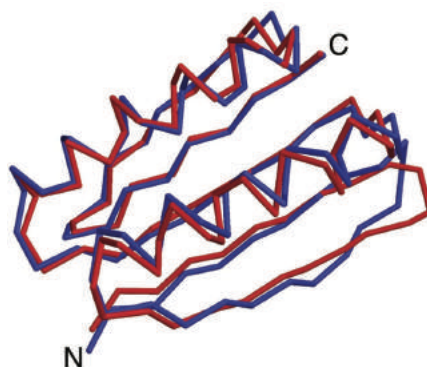


Figura 4 - Comparação do modelo projetado computacionalmente (azul) com a estrutura de raios X resolvida (vermelho) da Top7. Retirado de (Kuhlman *et al.*, 2003).

O avanço, tanto nos métodos e algoritmos utilizados pelo Rosetta, como na tecnologia de hardware e poder de cálculo de computadores, tem permitido também avanços em pesquisas na área de engenharia de proteínas. Novas enzimas têm sido desenvolvidas para uso em indústrias de processos químicos e farmacêuticos (Kiss *et al.*, 2013). Na área de química analítica e ambiental, novos sensores são desenvolvidos baseados na interação proteína-ligante para detecção de produtos tóxicos no meio ambiente, como o composto fentanil, derivado da morfina (Bick *et al.*, 2017). Abordagens computacional e experimental integradas permitiram o rápido desenvolvimento de dezenas de milhares de novos peptídeos com características de anticorpos e fármacos, proporcionando proteção imunológica contra vírus ou pequenas moléculas tóxicas (Chevalier *et al.*, 2017). Sabendo que a repetida administração de uma proteína terapêutica pode levar à ativação de células B e à produção de anticorpos contra essa proteína, desenvolveu-se métodos para o *design* de proteínas com imunogenicidade reduzida, porém mantendo a sua função e estabilidade (King *et al.*, 2014). Recentemente, utilizando todo o poder de predição do Rosetta, Butterfield e colaboradores puderam criar partículas com características virais, ou nucleocapsídeos sintéticos, formados por proteínas icosaédricas desenhadas *in silico* e que podem encapsular seus próprios mRNA genômicos (Figura 5) (Butterfield *et al.*, 2017).

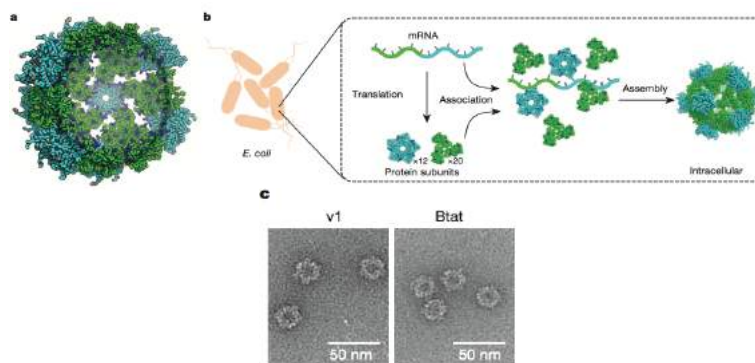


Figura 5 - Caracterização bioquímica de nucleocapsídeos sintéticos. (a) Modelo desenhado computacionalmente. (b) Os nucleocapsídeos sintéticos encapsulam seus próprios genomas enquanto se montam. (c) Microscopia eletrônica dos capsídeos obtidos experimentalmente. Retirado de (Butterfield *et al.*, 2017)

A área de engenharia de proteínas também vem sendo utilizada para desenvolvimento de novas vacinas. As estratégias mais utilizadas no desenvolvimento de uma nova vacina já são conhecidas há muito tempo: utilização de partículas semelhantes a vírus recombinantes; o vírus vivo ou vivo e atenuado; o vírus inativado; vacinas bacterianas compostas de proteínas de superfície de bactérias, toxinas ou polissacarídeos (Correia *et al.*, 2014). Uma estratégia que vem sendo utilizada recentemente consiste no planejamento de proteínas sintéticas, solúveis e estáveis, que permitam a exposição conformacional nativa do epítipo de interesse. Utilizando a técnica de “*side-chain grafting*”, Correia e colaboradores (Correia *et al.*, 2010) descreveram o método de “*epitope-scaffolding*”, onde foram capazes de transplantar o epítipo 4E10 do vírus da imunodeficiência humana (HIV) em proteínas carreadoras que irão servir para a estabilização conformacional e apresentação imunológica. Apesar do sucesso em produzir proteínas solúveis, estáveis e com grande afinidade pelo anticorpo monoclonal, a resposta imunológica não foi capaz de neutralizar o HIV. Num trabalho posterior, Correia e colaboradores (Correia *et al.*, 2011) utilizaram a técnica de “*flexible-backbone remodeling*” que permite otimizar a conformação do esqueleto da proteína após a sua modificação. Ao tentar reduzir o tamanho da proteína escolhida, no intuito de otimizar o imunogenicidade do epítipo transplantado, conseguiram obter novas estruturas estáveis e solúveis, mas sem sucesso em aumentar a resposta imune. Como prova de conceito,

Correia e colaboradores (Correia *et al.*, 2014) desenvolveram uma proteína carreadora de um epítipo do vírus sincicial respiratório e demonstraram que esta vacina baseada em proteínas é capaz de induzir resposta imune de longa duração em macacos rhesus (Figura 6).

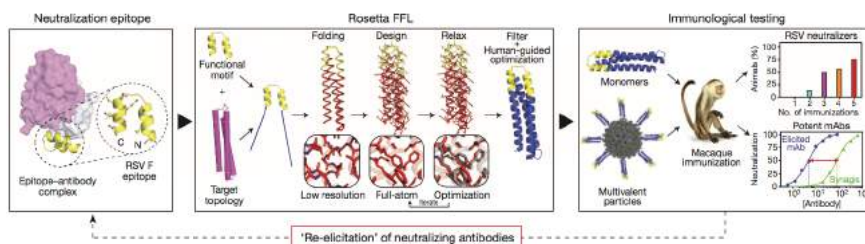


Figura 6 - Método computacional para projetar proteínas com flexibilidade total do *backbone* desenvolvido por Correia e colaboradores. Método foi utilizado no desenvolvimento de uma vacina baseado no epítipo do vírus RSV. A imunização de macacos rhesus induziu anticorpos neutralizantes de longa duração de RSV. Retirado de (Correia *et al.*, 2014).

A qualidade dos modelos criados computacionalmente depende de forma direta da qualidade da função de energia utilizada. O maior desafio para essas técnicas é que a energia de um sistema não pode ser computada com perfeita acurácia (Huang *et al.*, 2016). Por exemplo, o Rosetta possui algumas falhas já conhecidas pela comunidade científica: a amostragem é estocástica e, portanto, nem toda trajetória irá amostrar um mínimo regional; a função de energia é heurística e simplificada para maior eficiência de cálculo, de modo que não consegue descrever de forma totalmente realística as forças envolvidas; por fim, devido às simplificações, os mínimos da função de energia não são garantia de descrever estados biologicamente relevantes (Bender *et al.*, 2016). Outras limitações bem conhecidas para métodos de modelagem em geral são o uso de tratamento implícito do solvente, a necessidade de melhorias nas estratégias de amostragem, no tratamento de contribuições eletrostáticas e a dependência de dados estruturais presentes no PDB (Cunha *et al.*, 2015). Em particular, a água é importante para a estrutura, estabilidade, dinâmica e função das proteínas, atuando no enovelamento, mediando as interações durante o reconhecimento molecular, e contribui para a estabilização entálpica ou entrópica (Bellissent-Funel *et al.*,

2016). Consequentemente, a água não deve ser tratada como agente de interação implícito, mas sim como um componente integral e ativo dos sistemas biomoleculares, onde tem papéis dinâmicos e estruturais (Levy e Onuchic, 2006). Além disso, a identificação dos princípios estruturais fundamentais à estabilidade das proteínas é um passo crucial para o projeto racional de novas funcionalidades (Soares *et al.*, 2010). Apenas um número muito limitado de propriedades de um sistema biomolecular é realmente acessível a medições experimentais, e tais técnicas além de caras enfrentam limitações como a restrição do ambiente físico-químico comum para a determinação estrutural de proteínas e agregação e a insolubilidade limitam o uso em solução por espectroscopia de ressonância magnética nuclear (RMN) (Van Gunsteren *et al.*, 2006; Soares *et al.*, 2010).

Como os processos que desempenham um papel fundamental no comportamento dos sistemas biomoleculares são impulsionados por interações interatômicas fracas e não-ligadas, e a faixa de temperatura está basicamente entre as temperaturas ambiente e fisiológica, podem ser modelados no nível atômico e molecular e descritos pelas leis da mecânica estatística clássica (Van Gunsteren *et al.*, 2006). Simulações de dinâmica molecular (MD) são amplamente utilizadas para obter informações sobre a evolução temporal de propriedades moleculares em nível atômico, utilizando representações explícitas tanto do soluto como do solvente, e sob variadas condições ambientais que não podem ser totalmente testadas em nível atômico apenas por métodos experimentais (Soares *et al.*, 2010). Assim, como a síntese de biomoléculas é trabalhosa e muitas vezes cara, o uso em paralelo de diferentes técnicas de triagem computacional, como MD e *de novo design*, representa um complemento ao experimento (Van Gunsteren *et al.*, 2006; Bhardwaj *et al.*, 2016).

Baseados nisso, Boschek e colaboradores utilizaram técnicas de “*epitope-scaffolding*” e de simulação por MD para projetar uma proteína baseada na Top7, inserindo uma sequência peptídica derivada de um anticorpo específico para CD4 e demonstraram que a proteína se liga especificamente (Boschek *et al.*, 2009). Também utilizando metodologia de MD, Viana e colaboradores (Viana, I. *et al.*, 2013) desenvolveram de modo *in*

silico seis proteínas quiméricas contendo diferentes epítomos de HIV-1 (Figura 7). Quatro das seis proteínas desenhadas mostraram boa performance imunogênica e conseguiram diferenciar soros positivos dos soros negativos para HIV-1. Em um estudo posterior, a proteína exibindo o epítopo 2F5 mostrou afinidade contra o anticorpo monoclonal anti-2F5 superior ao peptídeo 23-MPER que é o atual padrão utilizado para diagnóstico de HIV (Viana, 2014).

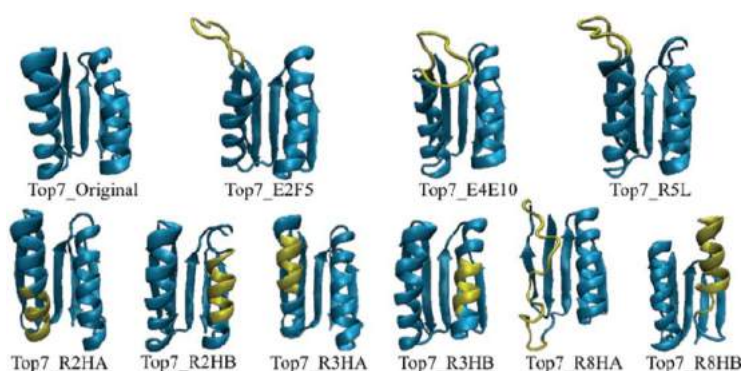


Figura 7 - Representação em *cartoon* de estruturas das quimeras Top7-HIV-1, desenhadas por Viana e colaboradores, após 50 ns de simulação de dinâmica molecular. As regiões do epítopo estão em amarelo. Retirado de (Viana, I. et al., 2013).

Utilizando metodologias de engenharia de proteínas e simulação por dinâmica molecular, o propósito deste trabalho é o desenvolvimento *in silico* e caracterização de proteínas/peptídeos para uso como potenciais antígenos de diagnóstico ou como método de identificação de anticorpos. Os vírus alvos neste estudo foram vírus da dengue, zika e HIV-1. Primeiro será apresentada uma fundamentação teórica das técnicas utilizadas neste trabalho. Em seguida, o uso destas técnicas para o estudo de moléculas específicas para cada doença será abordado em capítulos separados.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 HIDROPATIA DE KYTE-DOOLITTLE

Durante o enovelamento de uma proteína, se observa a tendência de cadeias laterais hidrofílicas estarem expostas ao solvente. As cadeias hidrofóbicas, por sua vez, tendem a se voltar para o interior da proteína. Entretanto, dados experimentais mostram que nem sempre essa tendência é obedecida. O efeito que as cadeias laterais vizinhas exercem também deve ser levado em consideração. Jack Kyte e Russell Doolittle descreveram um método sistemático visando avaliar as tendências hidrofílicas e hidrofóbicas de cadeias polipeptídicas (Kyte e Doolittle, 1982).

Para cada aminoácido determina-se um valor de hidropatia, que é calculado através da combinação de dois valores obtidos da literatura (Figura 8 A). O primeiro é a energia livre de transferência do aminoácido da água no estado líquido para a água no estado de vapor. O segundo é a tendência de o aminoácido estar exposto ou enterrado. Essa medida é feita através da investigação de aminoácidos em uma série de proteínas globulares. O método então seleciona segmentos ou janelas da sequência primária da proteína, que podem ter tamanho de 7 a 21 aminoácidos, começando pela parte N-terminal. Faz-se a soma dos valores de hidropatia de todos os aminoácidos da janela e atribui-se o valor da soma para o aminoácido central (Figura 8 B). A janela seguinte é escolhida avançando um aminoácido na sequência e repete-se os cálculos até cobrir toda a sequência da proteína. Ao final obtém-se um gráfico onde é possível visualizar o caráter hidropático da proteína em toda sua extensão. Valores muito negativos indicam regiões hidrofílicas. Uma janela de tamanho 9 é a ideal para previsão de regiões fortemente hidrofóbicas voltadas para o interior da proteína e de segmentos hidrofílicos voltados para o exterior (Kyte e Doolittle, 1982).

A

Aminoácido	Código 1 letra	Score de hidropatia
Isoleucina	I	4.5
Valina	V	4.2
Leucina	L	3.8
Fenilalanina	F	2.8
Cisteína	C	2.5
Metionina	M	1.9
Alanina	A	1.8
Glicina	G	-0.4
Treonina	T	-0.7
Serina	S	-0.8
Triptofano	W	-0.9
Tirosina	Y	-1.3
Prolina	P	-1.6
Histidina	H	-3.2
Ácido Glutâmico	E	-3.5
Glutamina	Q	-3.5
Ácido aspártico	D	-3.5
Asparagina	N	-3.5
Lisina	K	-3.9
Arginina	R	-4.5

B

· · · G K N F D Y T Y T V T T E S E L Q K V L N E L M D · · ·

Figura 8 - (a) Valores de score de hidropatia para cada tipo de aminoácido. Os valores foram obtidos de (Kyte e Doolittle, 1982). (b) representação esquemática de como funciona o algoritmo de Kyte-Doolittle; o valor final é atribuído ao aminoácido do meio da janela.

2.2 PREDIÇÃO DE ESTRUTURA DE PROTEÍNAS

Em biologia molecular, o problema do enovelamento de proteínas versa sobre como a sequência primária de aminoácidos de uma proteína dita a sua estrutura tridimensional (Dill *et al.*, 2008). De acordo com a hipótese termodinâmica de Anfinsen “a estrutura nativa de uma proteína é a estrutura termodinamicamente mais estável; ela depende apenas da sequência de aminoácidos e das condições da solução, não dependendo da rota cinética de enovelamento” (Anfinsen, 1973). Portanto, prever a estrutura nativa de uma proteína equivale a prever a estrutura termodinamicamente estável. Existem diversas técnicas de predição de estrutura de proteína na literatura. Algumas delas estão presentes no pacote Rosetta e seus detalhes serão apresentados neste trabalho. Mas primeiro será discutido o componente que faz do Rosetta uma ferramenta bastante eficaz e versátil: a sua função de energia.

2.2.1 Função de energia do Rosetta

Uma função de energia acurada é muito importante para que os modelos simulados estejam o mais próximos possíveis da estrutura real. No caso de modelos biomoleculares, em específico proteínas, procura-se qual seria a sua conformação nativa. Como mencionado anteriormente, procurar pela estrutura nativa equivale a procurar pela estrutura de menor energia. Portanto, a qualidade do modelo final depende da qualidade da função de energia que avalia as estruturas preditas (Gainza *et al.*, 2016). Em geral essas funções são combinações de vários termos, onde cada termo computa um tipo de energia em função de diferentes graus de liberdade (ou variáveis), e no final devem ser capazes de diferenciar entre uma estrutura não-enovelada, uma estrutura enovelada e a estrutura-nativa (Alford *et al.*, 2017).

O Rosetta é um pacote de ferramentas computacionais voltadas para problemas de biologia molecular, como o desenvolvimento de vacinas (Correia *et al.*, 2014), nano-biomateriais auto-montáveis (Huang *et al.*, 2016; Butterfield *et al.*, 2017) e desenvolvimento de novos anticorpos (Weitzner *et al.*, 2017), dentre outros. Sendo desenvolvido desde o início da década de 2000, sua função de energia tem sofrido muitas mudanças com o tempo, de modo a sempre melhorar a sua acurácia (Rohl *et al.*, 2004; Kellogg *et al.*, 2011; Park *et al.*, 2016; Alford *et al.*, 2017).

A função de energia do Rosetta (REF) considera a energia de uma proteína como sendo a combinação linear de vários termos energéticos. Cada termo depende de um grau de liberdade Θ_i , da identidade de cada aminoácido aa_i e possui o seu peso na combinação w_i .

$$\Delta E_{total} = \sum w_i E_i(\Theta_i, aa_i) \quad (\text{eq. 2.1})$$

Até recentemente, a REF usava uma unidade arbitrária (REU, *Rosetta energy unit*), pois alguns de seus termos não eram calibrados com dados experimentais e o uso de termos estatísticos tornava a interpretação de energia confusa (Rohl *et al.*, 2004; Alford *et al.*, 2017). A versão mais recente,

chamada de REF15, foi parametrizada a partir de estruturas de alta resolução de proteínas e dados termodinâmicos obtidos experimentalmente, de modo que ao comparar seus resultados com dados experimentais, obtém-se correlação de Pearson de 74% (Park *et al.*, 2016; Alford *et al.*, 2017). Dessa forma, assim como em outros campos de força, a unidade da REF15 passou a ser expressa em kcal/mol.

O Rosetta utiliza um algoritmo de amostragem de Monte Carlo/Metropolis (MCM). A amostragem do *backbone* se dá por uma etapa de baixa resolução (*coarse grained*), enquanto que amostragem de cadeias laterais é de alta resolução (*full atom*). A otimização do *backbone* em larga escala é feita baseada em fragmentos de 3 e 9 aminoácidos, derivados do PDB. Refinamentos locais do *backbone* se dão por pequenas perturbações dos ângulos ϕ e ψ . A amostragem das cadeias laterais também utiliza informações derivadas do PDB para criar uma biblioteca de rotâmeros.

A seguir serão discutidos os termos mais importantes da REF15. Porém, é importante ressaltar que outros termos também compõem a função e mais detalhes podem ser obtidos da literatura (Rohl *et al.*, 2004; Park *et al.*, 2016; Alford *et al.*, 2017).

2.2.1.1 Interações De Van Der Waals

As forças de Van der Waals correspondem a interações de curto alcance devido à presença de dipolos induzidos ou permanentes nos átomos ou moléculas. O Rosetta modela essas interações utilizando o potencial de Lenard-Jones (eq. 2.2), que calcula a energia de interação entre dois átomos i e j em função da soma dos raios atômicos, σ_{ij} , e da distância entre os átomos, d_{ij} e ϵ_{ij} é o valor mínimo do potencial:

$$E_{vdw}(i,j) = \epsilon_{i,j} \left[\left(\frac{\sigma_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{d_{ij}} \right)^6 \right] \text{ (eq 2.2)}$$

De modo a aumentar a eficiência dos cálculos, na parametrização da REF, esse potencial é dividido em duas partes: uma repulsiva (fa_rep), e uma atrativa (fa_atr).

Devido à contribuição energética alta do termo $1/d_{ij}^{12}$ à pequenas distâncias, suaviza-se o componente repulsivo (fa_rep, eq. 2.3) substituindo o potencial por um termo linear quando $d_{ij} \leq 0.6\sigma$. Os termos m_{ij} e b_{ij} dependem do par de átomos e são ajustados de modo a manter a função derivável em $0.6\sigma_{ij}$. A função então passa suavemente para uma forma modificada do potencial padrão de Lenard-Jones até $d_{ij} = \sigma_{ij}$ quando a função passa a ser zero e se mantém nula para $d_{ij} > \sigma_{ij}$ (Figura 9 A).

$$E_{fa_rep}(i,j) = \sum_{i,j} w_{ij} \begin{cases} m_{ij}d_{ij} + b_{ij} & d_{ij} \leq 0.6\sigma_{ij} \\ \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{d_{ij}} \right)^6 + 1 \right] & 0.6\sigma_{ij} < d_{ij} \leq \sigma_{ij} \\ 0 & d_{ij} \geq \sigma_{ij} \end{cases} \quad \text{eq. 2.3}$$

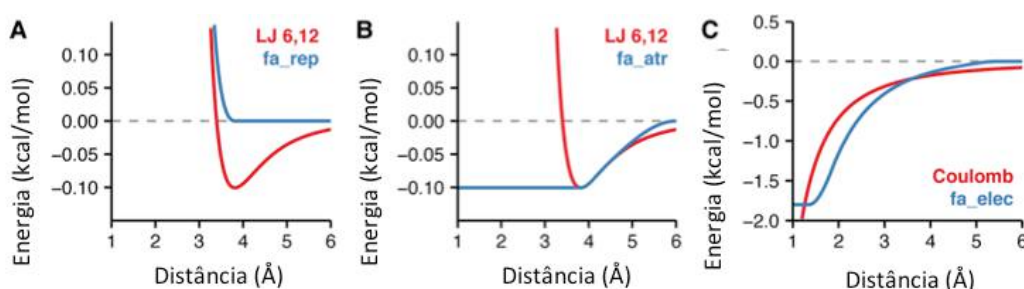


Figura 9 - Comparação entre energia de Van der Waals (Lennard-Jones) tradicional (vermelho) e método do Rosetta (azul). (A) Termo repulsivo da energia de van der Waals; (B) termo de atração e; (C) Energia eletrostática de Coulomb clássica (vermelho) em comparação com Rosetta fa_elec (azul). Dados representam interações com o átomo de nitrogênio do backbone. Adaptado de (Alford *et al.*, 2017).

O termo atrativo (fa_atr, eq. 2.4) possui valor truncado em $-\varepsilon_{ij}$ enquanto $d_{ij} \leq \sigma_{ij}$ e então transita para o potencial de Lenard-Jones quando $\sigma_{ij} < d_{ij} \leq 4.5 \text{ Å}$. Para distancias maiores do que 6.0 Å as interações de Van der Waals são negligenciáveis e para efeito de aumentar a eficiência dos cálculos, a função é truncada novamente no valor zero quando $6.0 \text{ Å} < d_{ij}$. Para garantir que a função seja derivável, um termo polinomial cúbico é utilizado na transição $4.5 \text{ Å} < d_{ij} \leq 6.0 \text{ Å}$ (Figura 9 B).

$$E_{fa_{atr}}(ij) = \sum_{i,j} w_{ij} \begin{cases} -\varepsilon_{ij} & d_{ij} \leq \sigma_{ij} \\ \varepsilon_{ij} \left[\left(\frac{\sigma_{ij}}{d_{ij}} \right)^{12} - 2 \left(\frac{\sigma_{ij}}{d_{ij}} \right)^6 \right] & \sigma_{ij} < d_{ij} \leq 4.5 \text{ \AA} \\ f_{poli}(d_{ij}) & 4.5 \text{ \AA} < d_{ij} \leq 6.0 \text{ \AA} \\ 0 & 6.0 \text{ \AA} < d_{ij} \end{cases} \quad \text{eq. 2.4}$$

Cada termo (fa_rep e fa_atr) é multiplicado por um valor w_{ij} de modo a ponderar contribuições muito altas em átomos separados por menos do que 4 ligações (eq. 2.5). A contribuição energética de átomos separados por 3 ou menos ligações é levada em consideração no termo relacionado aos ângulos torcionais, discutidos posteriormente.

$$w_{ij} = \begin{cases} 0 & n_{ij}^{ligacoes} \leq 3 \\ 0.2 & n_{ij}^{ligacoes} = 4 \\ 1 & n_{ij}^{ligacoes} \geq 5 \end{cases} \quad \text{eq. 2.5}$$

2.2.1.2 Interações de Coulomb

Para avaliar as interações eletrostáticas, o Rosetta utiliza a expressão de Coulomb (eq. 2.6), que descreve a energia de interação entre duas cargas pontuais q_i e q_j , imersas num meio com constante dielétrica ϵ e separadas por uma distância d_{ij} . Como a constante dielétrica depende do meio e possui valores diferentes para o solvente e para o núcleo hidrofóbico da proteína, o valor de ϵ não pode ser um valor constante. Dessa forma, é substituído por uma função que aumenta de $\epsilon = 6$ para $\epsilon = 80$ de forma sigmoidal no intervalo $0 \text{ \AA} \leq d_{ij} \leq 4 \text{ \AA}$. Dessa forma, com o aumento da distância entre os átomos, leva-se em consideração a presença de água.

$$E_{Coulomb} = \frac{c_0 q_i q_j}{\epsilon} \frac{1}{d_{ij}} \quad (\text{eq. 2.6})$$

Da mesma forma que nas interações de Van der Waals, a função é truncada no intervalo $d_{ij} < 1.45 \text{ \AA}$ no valor constante de $E_{elec}(d_{min})$ para evitar valores altos de repulsão/atração em pequenas distâncias. Como o valor da

função de Coulomb tende a 0 a partir de uma certa distância, por motivos de eficiência, a função é truncada com valor zero acima de $d_{\max} = 5.5 \text{ \AA}$. Para garantir que a função se mantenha contínua e derivável, a expressão de Coulomb é modificada subtraindo o termo $1/d_{\max}^2$ e usa-se funções polinomiais cúbicas para fazer as transições entre os intervalos (eq. 2.7, Figura 9 C).

$$E_{fa_elec} = \sum_{i,j} w_{ij} \begin{cases} E_{elec}(i,j, d_{min}) & d_{ij} < 1.45 \text{ \AA} \\ f_{poli}^{elec,low}(d_{ij}) & 1.45 \text{ \AA} \leq d_{ij} < 1.85 \text{ \AA} \\ \frac{c_0 q_i q_j}{\epsilon(d_{ij})} \left(\frac{1}{d_{ij}^2} - \frac{1}{d_{\max}^2} \right) & 1.85 \text{ \AA} \leq d_{ij} < 4.5 \text{ \AA} \\ f_{poli}(d_{ij}) & 4.5 \text{ \AA} \leq d_{ij} < 5.5 \text{ \AA} \\ 0 & 5.5 \text{ \AA} \leq d_{ij} \end{cases} \quad \text{eq. 2.7}$$

Assim como na expressão de Van der Waals, os termos são multiplicados por pesos que dependem da conectividade entre os átomos (eq. 2.5).

2.2.1.3 Termo de solvatação

Para computar as energias de solvatação, o Rosetta utiliza o modelo de exclusão de Lazaridis-Karplus (Lazaridis e Karplus, 1999), que consiste em representar o solvente de modo implícito. De acordo com esse modelo, para cada grupo que compõe a proteína a energia de solvatação é dada como a energia quando o grupo está totalmente exposto ao solvente menos a energia quando a solvatação do grupo diminui devido à presença de outros grupos ao redor. A REF15 desmembra o termo de solvatação em dois: um isotrópico e outro anisotrópico. O modelo anisotrópico leva em consideração a preferência da organização das moléculas de água ao redor de átomos polares e para mais detalhes deve-se consultar a referência (Alford *et al.*, 2017).

O termo isotrópico de solvatação considera uma distribuição uniforme das moléculas de água ao redor dos átomos (Figura 10 A). Baseia-se na função f_{desolv} (eq. 2.8) que descreve a energia necessária para desolvatar um átomo i quando na presença do átomo j (Alford *et al.*, 2017). O valor da função depende da distância entre os átomos d_{ij} , da energia livre de

transferência da fase vapor para água ΔG_i^{free} , da soma dos raios atômicos σ_{ij} , e do volume atômico do átomo j, V_j . Para evitar a sobreposição dos átomos à pequenas distâncias, a função é truncada no valor $f(\sigma_{ij})$ no intervalo $d_{ij} \leq \sigma_{ij} - 0.3 \text{ \AA}$; a função tende a zero à medida que a distância entre os átomos aumenta, e por isso seu valor é truncado em zero quando $6 \text{ \AA} < d_{ij}$ de forma a acelerar os cálculos. Para garantir que a função seja contínua e derivável, utiliza-se as funções polinomiais f_{poli}^{low} e f_{poli}^{high} nos pontos de transições (eq. 2.9).

$$f_{desolv} = -V_j \frac{\Delta G_i^{free}}{2\pi^{2/3} \lambda_i \sigma_{ij}^2} e^{-\left(\frac{d_{ij} - \sigma_{ij}}{\lambda_i}\right)^2} \quad \text{eq. 2.8}$$

$$g_{desolv} \begin{cases} f_{desolv}(i,j,\sigma_{ij}) & d_{ij} \leq \sigma_{ij} - 0.3 \text{ \AA} \\ f_{poli}^{low}(i,j,d_{ij}) & \sigma_{ij} - 0.3 \text{ \AA} < d_{ij} \leq \sigma_{ij} + 0.2 \text{ \AA} \\ f_{desolv}(i,j,d_{ij}) & \sigma_{ij} + 0.2 \text{ \AA} < d_{ij} \leq 4.5 \text{ \AA} \\ f_{poli}^{high}(i,j,d_{ij}) & 4.5 \text{ \AA} < d_{ij} \leq 6 \text{ \AA} \\ 0 & 6 \text{ \AA} < d_{ij} \end{cases} \quad \text{eq. 2.9}$$

A energia total de solvatação isotrópica (eq. 2.10) é dada como a soma entre as energias do átomo j desolvatando o átomo i e vice-versa, escalonado pelo termos da eq. 2.5.

$$E_{fa_sol} = \sum_{ij} w_{ij} [g_{desolv}(i,j) + g_{desolv}(j,i)] \quad \text{eq. 2.10}$$

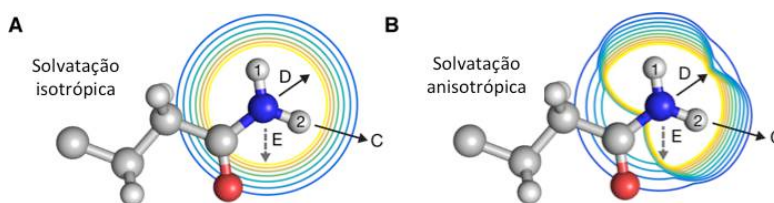


Figura 10 - Modelo de solvatação de Lazaridis-Karplus. O Rosetta usa dois termos de energia para avaliar a dessolvatação de cadeias laterais de proteínas: (A) um termo isotrópico (fa_sol) e (B) um termo anisotrópico. Adaptado de (Alford et al., 2017).

2.2.1.4 Termo de ligação de hidrogênio

A ligação de hidrogênio é uma interação parcialmente covalente que depende da geometria para maximizar o *overlap* entre os orbitais do átomo doador de densidade eletrônica e o átomo de hidrogênio (Alford *et al.*, 2017). Na REF15, a energia da ligação de hidrogênio é uma função que depende de 4 graus de liberdade (eq. 2.11, Figura 11): 1) a distância entre o átomo de hidrogênio H e o átomo acceptor A, d_{HA} ; 2) o ângulo formado pelo átomo acceptor A, átomo de hidrogênio H e o átomo doador de hidrogênio D, θ_{AHD} ; 3) o ângulo formado pelo átomo base B do átomo acceptor, átomo acceptor A e o átomo de hidrogênio H, θ_{BAH} ; e 4) o ângulo diedro formado pelo átomo de hidrogênio H, átomo acceptor A, átomo base B e o átomo base B₂, ϕ_{B_2BAH} . Os termos E_{hbond}^{HA} , E_{hbond}^{AHD} , E_{hbond}^{BAH} são polinomiais, cuja forma depende dos tipos dos átomos A e D. O termo $E_{hbond}^{B_2BAH}$ depende do tipo de hibridização (ρ) do átomo A, e é uma combinação entre uma função polinomial e uma função do cosseno (Alford *et al.*, 2017). A expressão final é multiplicada por pesos dependentes dos tipos de átomos H e A, w_H e w_A .

$$E_{hbond} = \sum_{H,A} w_H w_A f(E_{hbond}^{HA}(d_{HA}) + E_{hbond}^{AHD}(\theta_{AHD}) + E_{hbond}^{BAH}(\theta_{BAH}) + E_{hbond}^{B_2BAH}(\rho, \phi_{B_2BAH}, \theta_{BAH})) \quad \text{eq. 2.11}$$

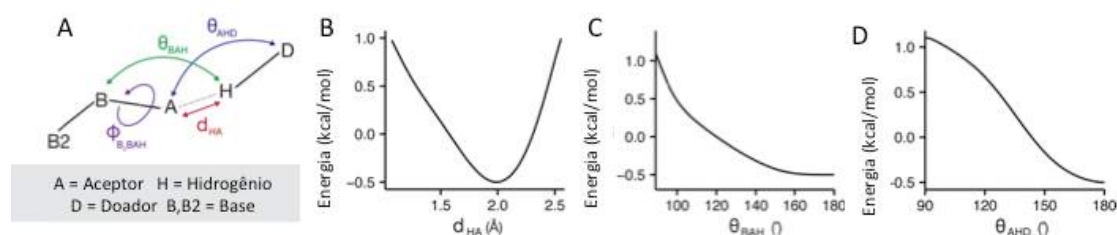


Figura 11 - Modelo de ligação de hidrogênio dependente da orientação. (A) Graus de liberdade avaliados pelo termo de ligação de hidrogênio: distância entre os átomos acceptor-doador d_{HA} ; o ângulo entre átomo base, o acceptor e o hidrogênio, θ_{BAH} ; o ângulo entre o acceptor, hidrogênio e doador, θ_{AHD} ; e o ângulo diédrico correspondente à rotação em torno da ligação base-aceitador, ϕ_{B_2BAH} . (D – F) Contribuições de cada grau de liberdade para a energia de ligações de hidrogênio. Adaptado de (Alford *et al.*, 2017).

2.2.1.5 Termo de Ramachandran

Para avaliar os ângulos ϕ e ψ , define-se um termo de energia baseado num mapa de Ramachandran obtido de um banco de dados de estruturas cristalográficas de alta resolução, para cada tipo de aminoácido. A energia é computada pela conversão das probabilidades de ocorrência dos ângulos em cada aminoácido em energia (eq. 2.12, (Alford *et al.*, 2017)).

$$E_{rama} = \sum_i -kT \ln[P(\phi_i, \psi_i | aa_i)] \quad \text{eq. 2.12}$$

2.2.1.6 Termo de design do backbone

Calcula-se também um termo energético que leva em consideração qual a probabilidade de que um aminoácido específico, aa_i , esteja localizado numa posição do *backbone* com conformação ϕ, ψ . Em outras palavras, qual a propensão de aa_i em relação aos outros 19 aminoácidos canônicos (Alford *et al.*, 2017). De forma análoga ao termo de Ramachandran, as probabilidades de cada aminoácido é derivada de um banco de dados de alta resolução e convertidas em energia pela relação abaixo

$$E_{p_aa_pp} = \sum_r -kT \ln \left[\frac{P(\phi, \psi | aa)P(aa)}{P(aa_r) \sum_{aa'} P(\phi, \psi | aa')} \right] \quad \text{eq. 2.13}$$

2.2.1.7 Termo de conformação da cadeia lateral

As cadeias laterais das proteínas assumem conformações específicas chamadas de rotâmeros. Os rotâmeros são conformações discretas devido às barreiras de rotação ao redor das ligações químicas. Baseado numa distribuição de probabilidade para cada rotâmero derivada de numa biblioteca de rotâmeros dependente da conformação do *backbone* (Shapovalov e Dunbrack, 2011), a REF15 descreve a energia das conformações das cadeias laterais convertendo essas probabilidades em valores energéticos.

O termo depende de três fatores probabilísticos: 1) probabilidade de observar um rotâmero específico dado um aminoácido aa_i e ângulos ϕ, ψ ; (2)

probabilidade de observar um ângulo χ não terminal dado um rotâmero rot_r e aminoácido aa_i ; e 3) observar um ângulo χ_T terminal

$$E_{fa_dun} = \sum_r -\ln[P(rot_r|\phi_r, \psi_r, aa_r)] + \sum_{k < T_r} \frac{1}{2} \left(\frac{\chi_{k,r} - \mu_{\chi_k}(\phi_r, \psi_r | rot_r, aa_r)}{\sigma_{\chi_k}(\phi_r, \psi_r | rot_r, aa_r)} \right)^2 - \ln[P(\chi_{T_r,r}|\phi_r, \psi_r, rot_r, aa_r)] \quad \text{eq. 2.14}$$

Tendo discutido os termos principais da REF, iremos discutir agora as técnicas do Rosetta que fazem uso da sua função de energia.

2.2.2 Threading

O *threading* tradicional (Jones *et al.*, 1992) é um método de predição de estrutura de proteína considerado *template based*, pois se baseia em estruturas já resolvidas. Uma biblioteca de estruturas conhecidas de proteínas é derivada de um banco de dados, geralmente o PDB. Cada estrutura é considerada uma possível conformação da proteína a ser estudada; as cadeias laterais dos *templates* são ignoradas. A sequência de aminoácidos da proteína a ser predita é então substituída em cada *template* e, através de uma função de energia, faz-se um ranking de energia com todas as possíveis estruturas. A de menor energia é considerada como a estrutura mais provável para aquela sequência de aminoácidos.

No desenho de uma nova proteína, algumas vezes o objetivo é apenas modificar posições na sequência de aminoácidos de uma proteína já conhecida. Nesse caso, o *template* é a própria proteína nativa e o *threading* é realizado em cima dela. Para o *design* das proteínas desse trabalho foi utilizado o protocolo *pmut_scan_parallel* (Kuhlman e Baker, 2000; Kuhlman *et al.*, 2003), incluso no software Rosetta. O protocolo consiste nas seguintes etapas (Figura 12):

1. Ler a sequência primária da proteína a ser modelada e substituir as cadeias laterais no *template*, que neste trabalho é a proteína carreadora nativa;

2. Otimizar o empacotamento das cadeias laterais através de pequenas rotações, mantendo o *backbone* fixo;
3. Calcular a energia a cada otimização da geometria através de uma função de energia;
4. Analisar a diferença de energia entre a proteína modelada e a proteína *template* nativa ($\Delta\Delta G$).

Para otimizar a geometria das cadeias laterais, usa-se uma biblioteca de rotâmeros (Dunbrack Roland e Cohen Fred, 2008) para cada aminoácido (Figura 12 C). Essa biblioteca consiste nos valores mais prováveis dos ângulos χ das cadeias laterais. Ela é construída através de uma análise estatística bayesiana em cima dos valores obtidos experimentalmente do PDB. A otimização ocorre através de um procedimento de amostragem de Monte Carlo em que cada etapa consiste em trocar um rotâmero por outro em uma posição escolhida aleatoriamente. A nova configuração será mantida ou descartada de acordo com o critério de MCM (Metropolis *et al.*, 1953):

$$\begin{cases} \text{Se } \Delta E < 0 \rightarrow \text{configuração aceita,} \\ \text{Se } \Delta E > 0 \rightarrow \text{probabilidade } \alpha = e^{\frac{-\Delta E}{k_B T}} \end{cases} \quad \text{eq. 2.15}$$

onde, ΔE é a diferença de energia entre a nova configuração e a configuração anterior; a probabilidade α é dada pela distribuição de Boltzmann, onde k_B é a constante de Boltzmann e T é a temperatura. Se $\alpha \geq \xi$, onde ξ é um número aleatório entre 0 e 1, então a nova configuração é mantida. Caso contrário mantém-se a configuração anterior. Uma nova configuração de maior energia não é descartada para evitar cair em mínimos globais, pois o funil de enovelamento é rugoso. Ao final do ciclo, a configuração de menor energia é considerada a estrutura mais provável.

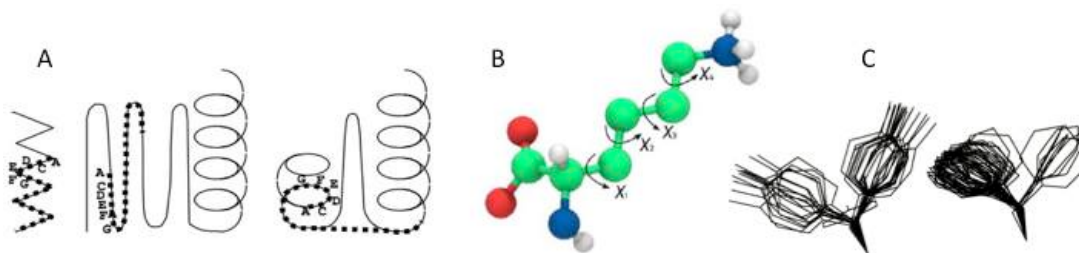


Figura 12 - (A) Escaneamento e substituição das cadeias laterais no *template*. (B) ângulos χ das cadeias laterais otimizados durante o empacotamento. (C) biblioteca de rotâmeros com as possíveis conformações de cada cadeia lateral.

2.2.3 De novo design

O método *de novo* design consiste na predição da estrutura de uma proteína a partir da sua sequência primária de aminoácidos. A princípio, os métodos de *de novo* design eram considerados técnicas *ab initio* ou *template free*, pois não utilizavam nenhuma estrutura conhecida como base. À medida que as técnicas foram ficando mais sofisticadas, as fronteiras entre diferentes categorias têm se tornado menos claras. Existem métodos hoje que são considerados híbridos entre *template free* e *template based*. Esses métodos não comparam a sequência alvo a uma proteína conhecida, mas a fragmentos, sequências curtas de aminoácidos, cuja estrutura é conhecida do PDB (Floudas *et al.*, 2006). Eles se baseiam na premissa de que durante o enovelamento, uma interação local favorável ao se formar é mantida. Esse contato localizado entre duas partes diferentes da proteína cria uma tendência, trazendo outros contatos em proximidade espacial. O processo se repete até o enovelamento completo da proteína.

O programa Rosetta possui o protocolo *AbInitioRelax*, que é um método de *de novo* de predição de estrutura de proteínas baseado em fragmentos (Rohl *et al.*, 2004). A amostragem das possíveis conformações é baseada na distribuição de estruturas homólogas de 3 ou 9 aminoácidos retiradas do PDB (Figura 13). Essa biblioteca de fragmentos representa a gama de estruturas locais acessíveis para todos os segmentos curtos da cadeia a ser predita. Estruturas compactas são então montadas

aleatoriamente através da combinação destes fragmentos, usando método MC.

O *AbInitioRelax* representa as configurações através de um espaço de torção, onde se utiliza como coordenadas os ângulos Φ , Ψ e ω do *backbone*, ao invés das coordenadas cartesianas usuais (Figura 13 D). O método é dividido em duas etapas: uma modelagem de baixa resolução apenas do *backbone* (*ab initio*), seguida de um refinamento de alta resolução da estrutura como um todo (*Relax*). Na etapa *ab initio* as cadeias laterais são representadas por um centroide localizado no centro de massa. Essa etapa é dividida nos seguintes passos:

1. A sequência a ser predita está inicialmente totalmente estendida;
2. 9 resíduos são escolhidos aleatoriamente e um fragmento contendo os 9 resíduos é escolhido da biblioteca de fragmentos;
3. Os ângulos Φ , Ψ e ω do fragmento são substituídos no *backbone* da cadeia estendida;
4. A energia da nova configuração é avaliada através de uma função de energia de baixa resolução;
5. A nova configuração é aceita de acordo com o critério de MCM na equação 2.15;
6. Por padrão, são realizados 28.000 ciclos de substituição de fragmentos de 9 resíduos, seguidos de 8.000 ciclos de substituições de fragmentos de 3 resíduos.

Ao final da etapa *ab initio* obtém-se um modelo de baixa resolução do *backbone*. É necessário agora refinar a estrutura, levando em consideração todos os átomos explicitamente. A etapa *Relax* de alta resolução é dividida nos seguintes passos:

1. Os ângulos Φ e Ψ são levemente modificados (2° a 3°), mantendo a conformação global intacta;
2. As cadeias laterais são ajustadas utilizando bibliotecas de rotâmeros, de forma análoga a descrita na metodologia do *pmut_scan_parallel*;

3. A conformação escolhida aleatoriamente nos passos anteriores é então submetida a uma minimização convencional (Press *et al.*, 1992), levando a um mínimo local;
4. O mínimo local é avaliado de acordo com o critério de MCM da equação 2.15

Ao final do procedimento obtém-se um modelo de alta resolução da estrutura. Para uma predição confiável, é aconselhável realizar de 20.000 a 50.000 modelos para cada proteína. Uma pequena fração desses modelos estará no mínimo de energia que é considerado próximo da energia da estrutura nativa. Caso haja alguma estrutura de referência para comparação, é possível obter um gráfico de energia em função do desvio quadrático médio (*root mean square deviation*, RMSD). O RMSD é uma medida da diferença geométrica entre um ou mais estruturas e a estrutura de referência.

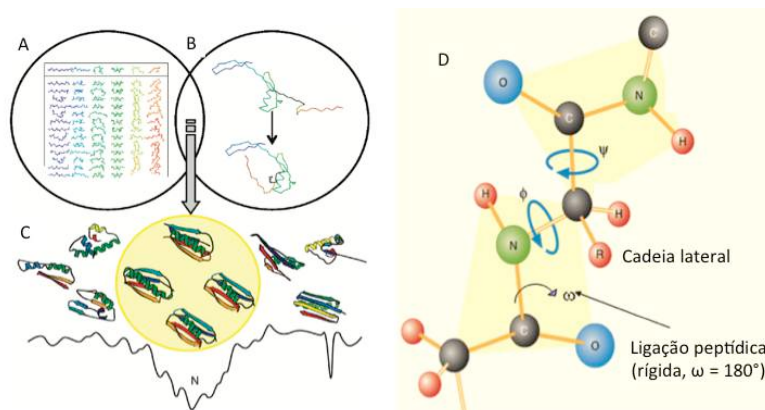


Figura 13 - O *AbinitioRelax* se baseia em (A) uma biblioteca de fragmentos que vão (B) se encaixando através da substituição dos valores (D) dos ângulos ψ e ϕ , formando um (C) funil de enovelamento. Adaptado de (Kaufmann *et al.*, 2010).

2.2.4 Motifgraft

Como já mencionado, o método de “*epitope-scaffolding*” consiste em transplantar epítopos de estrutura conhecida em outra proteína de modo a maximizar a exposição do epítipo na sua conformação nativa. Este procedimento pode ser realizado através da ferramenta *MotifGraft* presente

no Rosetta (Silva *et al.*, 2016). O procedimento do *MotifGraft* consiste nos seguintes passos (Figura 14):

1. Definir qual o epítopo a ser transplantado;
2. Definir uma biblioteca de estruturas cristalográficas proteicas que servirão como possíveis proteínas carreadoras;
3. Alinhar estruturalmente o epítopo à proteína carreadora de modo a determinar qual a melhor região para o transplante;
4. Transplantar o epítopo através do “*side-chain grafting*” ou “*backbone-grafting*”;
5. Redesenhar a proteína carreadora do ao redor do epítopo de modo a garantir o enovelamento da nova proteína;
6. Inspeccionar as estruturas obtidas (energia, sequência, estrutura terciária, etc.) de modo a escolher a mais estável.

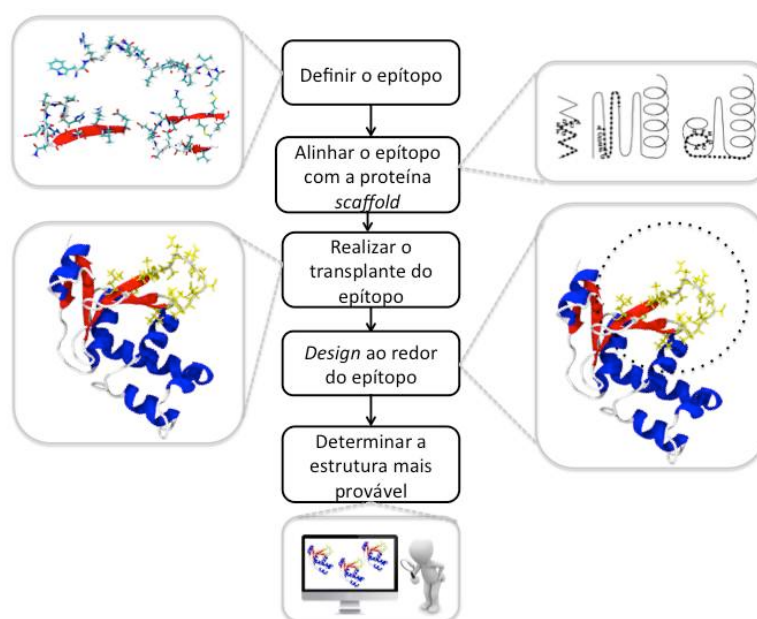


Figura 14 - Fluxograma demonstrando o passo a passo do protocolo de *MotifGraft*. Adaptado de (Silva *et al.*, 2016).

Na etapa de alinhamento entre o epítopo e o *scaffold* algumas condições de contorno podem ser ajustadas: pode-se limitar o RMSD entre os *backbones* a um valor máximo (e.g. 2 Å); ao realizar o transplante, pode ocorrer a superposição entre átomos do epítopo e da proteína carreadora,

por isso pode-se limitar esse valor também a um número máximo (e.g. 15 átomos conflitantes). Se o alinhamento entre as estruturas não obedecer a esses critérios, é descartado e passa-se para o próximo *scaffold*.

Após o alinhamento pode-se escolher entre manter o *backbone* do *scaffold* e transplantar apenas as cadeias laterais do epítopo (*side-chain grafting*) ou descartar os elementos estruturais do carreador, substituindo-o totalmente pelo motivo estrutural do epítopo (*backbone-grafting*). Sugere-se que o *side-chain grafting* deva ser tentado em primeiro lugar, uma vez que realiza o mínimo de alterações para o *scaffold*, aumentando assim as chances de obter estruturas com *folding* correto durante a validação experimental (Silva *et al.*, 2016).

O transplante do epítopo na proteína carreadora provoca uma mudança na sequência de aminoácidos e, conseqüentemente, muda as interações que antes mantinham o *scaffold* na sua estrutura nativa. Dessa forma, para garantir que a nova proteína manterá o *folding*, realiza-se um *design* ao redor do epítopo, de modo a aumentar as interações favoráveis e modificar o núcleo hidrofóbico da proteína o mínimo possível.

A energia de um sistema não pode ser perfeitamente computada devido à falhas nas funções de energia (Huang *et al.*, 2016; Liu e Chen, 2016; Park *et al.*, 2016) e a REF15 não foge à regra. Duas falhas conhecidas do Rosetta são: 1) posicionar aminoácidos polares no núcleo hidrofóbico; e 2) criar regiões enterradas com alto teor de alaninas (Silva *et al.*, 2016). Dessa forma, ao final do procedimento de *MotifGraft*, deve-se inspecionar as estruturas utilizando critérios como: menor energia, qualidade do núcleo hidrofóbico e probabilidade da proteína realizar o *folding* correto. Neste último caso, pode-se realizar uma predição de *de novo design* para investigar se há a formação do funil de enovelamento para a nova proteína.

O *MotifGraft* é uma ferramenta que deve ser utilizada com o *RosettaScripts* (Fleishman *et al.*, 2011). O *RosettaScripts* é uma linguagem em XML que permite ao usuário criar rotinas, especificando quais ferramentas de modelagem devem ser aplicadas aos arquivos de *input*, em um ordem definida (Figura 15). As ferramentas do *RosettaScripts* se enquadram em quatro categorias: 1) *Movers*, que são as ferramentas que modificam uma estrutura de algum modo (e.g. *MotifGraft*); 2) Filtros, que

avaliam propriedades de uma estrutura e permitem ou não continuar o protocolo de acordo com o *threshold* definido (e.g. RMSD); 3) *TaskOperations*, que controlam os graus de liberdade dos Movers, como por exemplo a amostragem das cadeias laterais; e 4) *ScoreFunctions*, que determina a função de energia que irá avaliar a energia das estruturas de *output* (Fleishman *et al.*, 2011; Bender *et al.*, 2016). Combinando estes quatro elementos, os usuários podem controlar de maneira mais fácil as ferramentas de modelagem e são capazes de criar rotinas de amostragem, modelagem e pontuação de acordo com sua necessidade.

```
<ROSETTASCRIPTS>
  <SCOREFXNS>
    <ScoreFunction name="ref2015" weights="ref2015.wts" />
  </SCOREFXNS>
  <RESIDUE_SELECTORS>
  </RESIDUE_SELECTORS>
  <TASKOPERATIONS>
    <RestrictResiduesToRepacking name="epitope_nodesign"
      residues="109,110,111,112,113,114,115,116,117,118,119,120,121" />
    <DesignAround name="design_around" design_shell="10.0"
      resnums="109,110,111,112,113,114,115,116,117,118,119,120,121" repack_shell="15.0" allow_design="1"
      resnums_allow_design="0" />
  </TASKOPERATIONS>
  <FILTERS>
    <Rmsd name="rmsd" threshold="1.5" superimpose="1" confidence="1" />
    <ExposedHydrophobics name="hydrophobic_sasa" sasa_cutoff="20" threshold="-1" confidence="0" />
    <BuriedUnsathbonds name="unsat" jump_number="0" confidence="0" />
  </FILTERS>
  <MOVERS>
    <MotifGraft name="motif_grafting" context_structure="gly.pdb" motif_structure="epitope.pdb"
      RMSD_tolerance="1.5" NC_points_RMSD_tolerance="1.5" clash_score_cutoff="15" clash_test_residue="GLY"
      hotspots="1:2:3:4:5:6:7:8:9:10:11:12" combinatory_fragment_size_delta="0:0" max_fragment_replacement_size_delta="0:0"
      full_motif_bb_alignment="1" graft_only_hotspots_by_replacement="1" revert_graft_to_native_sequence="0" />
    <PackRotamersMover name="redesign" scorefxn="ref2015" />
    <MinMover name="minimization" bb="1" chi="1" jump="0" tolerance="0.001" max_iter="1000" />
  </MOVERS>
  <APPLY_TO_POSE>
  </APPLY_TO_POSE>
  <PROTOCOLS>
    <Add mover="motif_grafting" />
    <Add mover="redesign" />
    <Add mover="minimization" />
    <Add filter="rmsd" />
    <Add mover="redesign" />
    <Add mover="minimization" />
    <Add filter="rmsd" />
    <Add filter="hydrophobic_sasa" />
    <Add filter="unsat" />
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

Figura 15 - Exemplo de rotina criada para o *MotifGraft* utilizando o *RosettaScripts*

2.3 MECÂNICA MOLECULAR

Sistemas moleculares podem ser computacionalmente modelados através de técnicas baseadas principalmente em métodos de mecânica quântica (QM) ou de mecânica molecular clássica (MM). Os métodos QM lidam com os elétrons do sistema, através da resolução da equação de Schrödinger. Os métodos QM permitem maior precisão comparados às técnicas clássicas, além de fornecerem dados sobre a estrutura eletrônica. Porém, ao levar em consideração os elétrons do sistema, os cálculos

consomem grande quantidade de tempo e memória computacional. A MM ignora o movimento eletrônico e calcula a energia do sistema em função da posição nuclear, simplificação conhecida como aproximação de Born-Oppenheimer (Leach, 2001). Movimentos moleculares, como o estiramento de ligações químicas, a deformação de um ângulo de ligação ou a torção de um diedro são descritos como sistemas massa-mola. Essa aproximação permite a modelagem de sistemas com um número significativo de átomos a um menor custo computacional. Em muitos casos a MM fornece resultados tão acurados quanto a QM. Entretanto, a MM não pode fornecer propriedades que dependem da distribuição eletrônica da molécula.

As técnicas baseadas em MM necessitam de um conjunto de parâmetros que descrevam a energia potencial do sistema. Esse conjunto de parâmetros, associado a uma função empírica de energia potencial, é conhecido como campo de força. Na sua forma funcional, o campo de força pode ser interpretado como a soma de termos de energia potencial. No caso do campo de força GROMOS (Oostenbrink *et al.*, 2004) utilizado neste trabalho, pode-se separar os termos para interações entre átomos ligados e não-ligados:

$$V^{\text{Total}} = V^{\text{lig}} + V^{\text{não-lig}} \quad \text{eq. 2.16}$$

As interações entre átomos ligados são a soma dos termos do potencial de ligação V^b , potencial angular V^θ , potencial de diedro improprio V^ξ e potencial torcional V^φ :

$$V^{\text{lig}} = V^b + V^\theta + V^\xi + V^\varphi \quad \text{eq. 2.17}$$

As interações não-ligadas são a soma dos termos de van der Waals (Lennard-Jones) V^{LJ} e do potencial eletrostático de Coulomb V^{C} :

$$V^{\text{não-lig}} = V^{\text{LJ}} + V^{\text{C}} \quad \text{eq. 2.18}$$

Os termos serão descritos nas próximas seções, com as informações e mais detalhes sendo obtidos de (Oostenbrink *et al.*, 2004; Van Gunsteren *et al.*, 2006).

2.3.1 Potencial de interação de ligação

No campo de força GROMOS a energia potencial devido às interações de ligação covalente é calculada como a soma sobre todas as ligações N_b e depende dos parâmetros K_b e b_0

$$V^b = \sum_{n=1}^{N_b} \frac{1}{4} K_{b_n} [b_n^2 - b_{0_n}^2]^2 \quad \text{eq. 2.19}$$

onde K_b é a constante de força da n -ésima ligação, b_0 é a posição de equilíbrio da n -ésima ligação e b_n é o comprimento real da n -ésima ligação. Os parâmetros K_b e b_0 são definidos para cada tipo de ligação e seus valores foram derivados de dados experimentais.

2.3.2 Potencial de ângulo de ligação

Assim como as interações de ligação, a energia potencial devido ao ângulo da ligação covalente é calculada como a soma sobre todos os ângulos N_θ e depende dos parâmetros K_θ e θ_0 :

$$V^\theta = \sum_{n=1}^{N_\theta} \frac{1}{2} K_{\theta_n} [\cos \theta_n - \cos \theta_{0_n}]^2 \quad \text{eq. 2.20}$$

onde θ_n é o valor real do n -ésimo ângulo, θ_0 é o ângulo de equilíbrio e K_θ é a constante de força do ângulo. Os valores de K_θ e θ_0 são definidos para cada tipo de ângulo de ligação e foram derivados de dados experimentais.

2.3.3 Potencial de diedro impróprio

Interações de diedro impróprio, também conhecidas como potencial fora-do-plano, são usadas para manter um conjunto de quatro átomos numa configuração específica. A energia potencial devido a essas interações é calculada como uma soma sobre N_ξ centros de interação de diedros impróprios, com parâmetros K_ξ e ξ_0 :

$$V^\xi = \sum_{n=1}^{N_\xi} \frac{1}{2} K_{\xi_n} [\xi_n - \xi_{0n}]^2 \quad \text{eq. 2.21}$$

onde ξ_n é o ângulo impróprio real definido por 4 átomos, K_ξ é a constante que define a altura da barreira de rotação e ξ_0 é o ângulo de equilíbrio para o diedro.

2.3.4 Potencial torsional

A energia potencial proveniente de interações de torção é calculada como uma soma sobre N_φ ângulos diedros de torção, utilizando os parâmetros K_φ , δ e m :

$$V^\varphi = \sum_{n=1}^{N_\varphi} K_{\varphi_n} [1 + \cos(\delta_n) \cos(m_n \varphi_n)] \quad \text{eq. 2.22}$$

onde δ_n é a diferença de fase, que é restrito a 0 ou π (o valor depende se for um máximo ou mínimo de energia), m_n é a multiplicidade do ângulo diedro de torção e φ_n é o valor real do ângulo diedro. Os parâmetros K_φ , δ e m são definidos sobre os tipos de ângulo diedro próprios. Os valores são selecionados de modo que os perfis de energia torcionais da mecânica quântica fossem reproduzidos pelo potencial torcional.

2.3.5 Interações De Van Der Waals

De forma semelhante ao Rosetta, as interações de van der Waals são calculadas como a soma sobre todos os pares de átomos não-ligados que interagem, usando uma função de Lennard-Jones 12-6 com os parâmetros C12 e C6:

$$V^{LJ} = \sum_{\text{pares } i,j} \left(\frac{C_{12ij}}{r_{ij}^{12}} - \frac{C_{6ij}}{r_{ij}^6} \right) \quad \text{eq. 2.23}$$

onde r_{ij} é a distância entre os átomos não-ligados i e j , e os termos C6 e C12 correspondem à atração e repulsão entre os átomos, respectivamente. Os parâmetros C_{12ij} e C_{6ij} para o par de átomos i, j são obtidos a partir de C_{12ii} , C_{12jj} , e C_{6ii} , C_{6jj} , definidos no GROMOS para cada tipo de átomo, usando as combinações geométricas:

$$C_{12ij} = \sqrt{C_{12ii} \cdot C_{12jj}} \quad \text{eq. 2.24}$$

$$C_{6ij} = \sqrt{C_{6ii} \cdot C_{6jj}} \quad \text{eq. 2.25}$$

Em geral, além dos átomos covalentemente ligados (primeiros vizinhos), os átomos separados por duas ligações (segundos vizinhos) também são excluídos da soma. Esses átomos já interagem através de interações ligadas e, portanto, seus pares não são contabilizados. Além disso, é delimitada uma região esférica do espaço, truncando o potencial em um raio de corte esférico. Dessa forma, só são levadas em consideração as interações que estão dentro do raio de corte, diminuindo tempo de cálculo. Adicionalmente, evita-se que partículas interajam com suas imagens nas condições periódicas de contorno. A simplificação pode ser feita, pois as interações do potencial de LJ são de curto alcance, decaindo rapidamente na ordem de r^{-6} ou r^{-12} .

2.3.6 Interações eletrostáticas

O tratamento das interações eletrostáticas de longo alcance podem ser divididas em duas contribuições. A primeira consiste numa soma sobre todos os pares de átomos que interagem, usando uma interação de Coulomb V^C , que depende das cargas parciais q_i em cada átomo e da distância r_{ij} entre os átomos:

$$V^C = \sum_{\text{pares } i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \frac{1}{r_{ij}} \quad \text{eq. 2.26}$$

onde ϵ_0 é a permissividade do vácuo e ϵ_1 é a permissividade do meio. Os valores dos parâmetros empregados são determinados a partir de resultados de cálculos de mecânica quântica e/ou por ajustes a dados experimentais.

Átomos covalentemente ligados (primeiros vizinhos) e átomos separados por duas ligações (segundos vizinhos) em geral são excluídos da soma. Além disso, as interações de Coulomb são frequentemente truncadas por raio de corte esférico. Porém, as interações eletrostáticas apresentam potencial que decai na ordem de r^{-1} . Devido ao maior alcance, o truncamento do potencial eletrostático poderá induzir erros significativos em simulações que apresentam espécies carregadas (Tironi *et al.*, 1995). Dessa forma, correções na truncagem das contribuições de longo alcance são necessárias. Dentre elas, a contribuição campo de reação é frequentemente utilizada juntamente com o campo de força GROMOS. Neste caso são levados em conta mais dois termos: a contribuição do campo de reação V^{RF} e um termo do campo de reação independente da distância V^{RFc} .

O V^{RF} é calculado como a interação do átomo i com o meio além do raio de corte R_{rf} , representado como um dielétrico contínuo uniforme devido à presença do átomo j :

$$V^{RF} = \sum_{\text{pares } i,j} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_1} \frac{-\frac{1}{2}C_{rf}r_{ij}^2}{R_{rf}^3} \quad \text{eq. 2.27}$$

onde

$$C_{rf} = \frac{(2\varepsilon_1 - 2\varepsilon_2)(1 + \kappa R_{rf}) - \varepsilon_2(\kappa R_{rf})^2}{(\varepsilon_1 + 2\varepsilon_2)(1 + \kappa R_{rf}) + \varepsilon_2(\kappa R_{rf})^2} \quad \text{eq. 2.28}$$

e ε_1 é a permissividade dielétrica do meio dentro de R_{rf} , ε_2 é a permissividade do meio além de R_{rf} e κ é uma constante que depende da força iônica do dielétrico além do raio de corte.

O último termo a ser considerado V^{RFc} é uma constante que é adicionada para anular a interação quando a distância entre os átomos i e j for igual à distância do raio de corte:

$$V^{RFc} = \sum_{\text{pares } i,j} \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_1} \frac{-\left(1 - \frac{1}{2}C_{rf}\right)}{R_{rf}} \quad \text{eq.2.29}$$

Assim, o potencial eletrostático total do sistema consiste na soma das equações 2.26, 2.27 e 2.29.

As simulações computacionais procuram entender e prever o comportamento e propriedades de sistemas líquidos puros, soluções, sólidos ou moléculas imersas num meio. Os dados experimentais são obtidos através de medições realizadas em sistemas macroscópicos, com enormes quantidades de átomos ou moléculas. A superfície de energia desses sistemas possui grande número de mínimos e uma quantificação completa do seu perfil de energia é impraticável. As técnicas de simulação permitem estudar as propriedades de sistemas macroscópicos através de réplicas menores do sistema. Nos próximos tópicos serão discutidos alguns princípios gerais envolvidos na técnica de simulação utilizada neste trabalho, a dinâmica molecular. As informações foram tiradas de (Leach, 2001; Cramer, 2004) de onde mais detalhes podem ser obtidos.

2.4 DINÂMICA MOLECULAR

Em geral, as propriedades de um sistema dependem das posições e dos momentos das N partículas que compõem o sistema. O valor instantâneo de uma propriedade A é escrita como $A(\mathbf{p}^N(t), \mathbf{r}^N(t))$, onde $\mathbf{p}^N(t)$ é o momento da partícula N no instante t , e $\mathbf{r}^N(t)$ é a posição da partícula N no instante t . O conjunto de todos os valores do par $(\mathbf{p}^N, \mathbf{r}^N)$ compõe o espaço de fase do sistema. Ao fazer uma medida experimental o que se obtém na verdade é uma média de A sobre o tempo de medição, conhecida como média temporal. À medida que o tempo de integração sobre A aumenta, mais próximo da realidade é o valor da média temporal:

$$A_{\text{méd}} = \lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_{\tau=0}^{\tau} A(\mathbf{p}^N(t), \mathbf{r}^N(t)) dt \quad \text{eq. 2.30}$$

A dificuldade é que para sistemas macroscópicos o número de partículas é da ordem de 10^{23} e é impraticável determinar as configurações de toda essa quantidade de átomos e moléculas. Para contornar esse problema, métodos computacionais utilizam a mecânica estatística, desenvolvida por Boltzmann e Gibbs. Nesse caso, um sistema evoluindo no tempo é substituído por um grande número de réplicas do sistema e são consideradas simultaneamente. O conjunto de réplicas é conhecido como *ensemble* e a média temporal nesse caso é substituída pela média do *ensemble*:

$$\langle A \rangle = \iint A(\mathbf{p}^N, \mathbf{r}^N) \rho(\mathbf{p}^N, \mathbf{r}^N) d\mathbf{p}^N d\mathbf{r}^N \quad \text{eq. 2.31}$$

onde $\langle A \rangle$ é o valor esperado da propriedade A , uma média sobre todas as réplicas do *ensemble* gerados na simulação. $\rho(\mathbf{p}^N, \mathbf{r}^N)$ é a densidade de probabilidade do *ensemble*, ou seja, a probabilidade de encontrar uma configuração com momento $\mathbf{p}^N$ e posições $\mathbf{r}^N$. O valor esperado, ou média do *ensemble*, é obtida da integração sobre todas as possíveis configurações do sistema. Sobre condições de número de partículas, volume e temperatura

constantes (*ensemble* NVT), a densidade de probabilidade é obtida da distribuição de Boltzmann:

$$\rho(\mathbf{p}^N, \mathbf{r}^N) = \frac{e^{-E(\mathbf{p}^N, \mathbf{r}^N)/k_B T}}{Q} \quad \text{eq. 2.32}$$

onde $E(\mathbf{p}^N, \mathbf{r}^N)$ é a energia, k_B é a constante de Boltzmann, T é a temperatura e Q é a função de partição do sistema:

$$Q = \iint e^{-E(\mathbf{p}^N, \mathbf{r}^N)/k_B T} d\mathbf{p} d\mathbf{r} \quad \text{eq. 2.33}$$

A distribuição de Boltzmann mostra que um ponto no espaço de fase com uma energia muito alta tem baixa probabilidade de ser ocupado e, portanto, contribui pouco para o valor esperado da propriedade A . Então, o que é necessário para avaliar as equações 2.30 e 2.31 é uma metodologia para escolher os pontos de fase mais importantes.

Se o sistema for iniciado num ponto de fase de baixa energia, é intuitivo considerar que a sua evolução no tempo irá amostrar regiões importantes do espaço de fases. Nesse caso, uma forma razoável de computar o valor esperado de A é calculando o valor da propriedade periodicamente a cada passo t_i :

$$\langle A \rangle = \frac{1}{M} \sum_i^M A(\mathbf{p}^N, \mathbf{r}^N) \quad \text{eq. 2.34}$$

onde M é o número de vezes que a propriedade é amostrada. A forma de como seguir essa trajetória no espaço de fases é o tema central da MD.

Na MD, o sistema é representado por uma caixa de simulação contendo moléculas do soluto e do solvente. As sucessivas configurações do sistema são geradas pela integração das equações diferenciais da segunda lei de Newton, para cada átomo do sistema:

$$\vec{F} = m \frac{d^2 \vec{r}}{dt^2} = m \vec{a} \quad \text{eq. 2.35}$$

O resultado é uma trajetória que especifica como a posição e as velocidades das partículas do sistema variam com o tempo. Para encontrar o valor da força resultante em cada átomo basta utilizar a relação entre energia potencial e força numa dada direção q :

$$F_q = - \frac{\partial U}{\partial q} \quad \text{eq. 2.36}$$

A ideia essencial é particionar a integração das equações de movimento em pequenos estágios separados por um intervalo de tempo fixo δt e seguir os seguintes passos:

1. Calcula-se a força total em cada partícula no instante t ;
2. A partir da força calcula-se a aceleração. Combinando com a velocidade e a posição no instante t , determina-se a velocidade e a posição de cada átomo no instante $t + \delta t$;
3. Durante o intervalo de tempo δt a força que atua na partícula é constante;
4. Calcula-se as novas energias de interação e conseqüentemente as forças resultantes em cada átomo na nova posição;
5. Determina-se as acelerações na nova posição. Combinando com a velocidade e a posição, determina-se a velocidade e a posição no instante $t + 2\delta t$ e assim sucessivamente.

Nos próximos tópicos serão discutidas algumas fundamentações da MD.

2.4.1 Algoritmo de integração *leap-frog*

Existem diversos algoritmos de integração das equações de movimento que são utilizadas nas técnicas de MD. A maioria leva em consideração a expansão de Taylor das propriedades dinâmicas, como por exemplo a expansão no vetor posição

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}(t)\delta t + \frac{1}{2!}\mathbf{a}(t)\delta t^2 + \frac{1}{3!}\left.\frac{d^3\mathbf{r}(\tau)}{d\tau^3}\right|_{\tau=t}\delta t^3 + \dots \quad \text{eq. 2.34}$$

Um dos métodos mais utilizados é o algoritmo de Verlet, que considera a expansão em $(t + \delta t)$ e $(t - \delta t)$. Somando as duas expressões, os termos derivativos de ordem-ímpar se cancelam, resultando na expressão abaixo:

$$\mathbf{r}(t + \delta t) = 2\mathbf{r}(t) - \mathbf{r}(t - \delta t) + \mathbf{a}(t)\delta t^2 \quad \text{eq. 2.35}$$

Dessa forma, para qualquer partícula, cada posição no passo subsequente é determinada pela posição atual, pela posição no passo anterior e pela aceleração que é determinada pela força. O algoritmo de Verlet não possui um termo explícito para a velocidade, tornando difícil simulações em que seja necessário controlar a temperatura do sistema.

O algoritmo de integração *leapfrog* é uma modificação do algoritmo de Verlet que leva em consideração a propagação da posição e da velocidade de forma acoplada. Fazendo a expansão para o vetor posição em torno de $t + \delta t/2$ e $t - \delta t/2$ e subtraindo a segunda da primeira, teremos:

$$\mathbf{r}(t + \delta t) = \mathbf{r}(t) + \mathbf{v}\left(t + \frac{1}{2}\delta t\right)\delta t \quad \text{eq. 2.36}$$

De forma similar é possível mostrar que:

$$\mathbf{v}\left(t + \frac{1}{2}\delta t\right) = \mathbf{v}\left(t - \frac{1}{2}\delta t\right) + \mathbf{a}(t)\delta t \quad \text{eq. 2.37}$$

No método *leapfrog* as posições dependem das velocidades computadas meio passo fora de fase. Dessa forma, é possível escalonar as velocidades para controlar a temperatura do sistema.

A configuração inicial do sistema pode ser obtida de dados experimentais, modelos teóricos ou uma combinação dos dois. Para as velocidades iniciais dos átomos utiliza-se a distribuição de Maxwell-Boltzmann na temperatura inicial de interesse.

2.4.2 Controle de temperatura

Em muitas simulações é desejável controlar ou até mesmo manter a temperatura do sistema num valor constante. No caso de simulações em condições fisiológicas, a temperatura deve ser $T = 310$ K. Vários métodos de controlar a temperatura existem na literatura e são amplamente utilizados. Um desses métodos é o esquema de reescalonamento de velocidades, utilizando uma amostragem canônica (Bussi *et al.*, 2007). De acordo com a mecânica estatística, a temperatura está relacionada com o momento de cada partícula que compõe o sistema:

$$T(t) = \frac{1}{(3N-n)k_B} \sum_{i=1}^N \frac{|\mathbf{p}_i(t)|^2}{m_i} \quad \text{eq. 2.38}$$

onde N é o total de átomos, $(3N - n)$ é o grau de liberdade do sistema. Na sua formulação simples, o método de reescalonamento de velocidade consiste em multiplicar as velocidades de todas as partículas por um fator α :

$$\alpha = \sqrt{\frac{K_t}{K}} \quad \text{eq. 2.39}$$

Esse fator de reescalonamento é calculado de modo a forçar uma distribuição canônica da energia cinética, onde K é a energia cinética total do sistema e K_t é a energia cinética de referência escolhida estocasticamente da distribuição de equilíbrio canônico para a energia cinética:

$$\bar{P}(K_t)dK_t \propto K_t^{(N_f/2-1)} e^{-\beta K_t} dK_t \quad \text{eq. 2.40}$$

onde $\bar{P}$ é a densidade de probabilidade de K_t , N_f é o grau de liberdade e β é a temperatura inversa. A expressão final de α vai depender da função K_t escolhida e de um parâmetro τ que possui dimensão de tempo e indica a escala temporal do reescalonamento. Então é só multiplicar as velocidades pelo fator de reescalonamento a cada passo da simulação.

2.4.3 Controle de pressão

Muitas das medições experimentais são realizadas em condições de pressão e temperatura constantes. Dessa forma, simulações em *ensemble* isotérmico-isobárico (NPT) são mais relevantes em relação a dados experimentais. Assim como um sistema macroscópico real mantém a pressão variando o volume, uma simulação no *ensemble* isotérmico-isobárico também mantém a pressão constante através de uma mudança de volume. Um dos métodos mais utilizados é o barostato de Berendsen (Berendsen *et al.*, 1984). Nesse caso a pressão é mantida constante por um escalonamento do volume, acoplando o sistema a um “banho de pressão”, em analogia a um banho térmico. O fator de reescalamento é dado por:

$$\lambda = 1 - \kappa \frac{\delta t}{\tau_p} (P - P_{\text{banho}}) \quad \text{eq. 2.41}$$

onde τ_p é a constante de acoplamento e tem unidade de tempo, P_{banho} é a pressão do “banho”, P é a atual pressão do sistema e κ é a compressibilidade do solvente. O reescalamento ocorre a cada passo multiplicando o volume por λ .

2.4.4 Condições periódicas de contorno

Ao representar sistemas macroscópicos através de caixas de simulação, limita-se a quantidade de átomos entre 10^5 e 10^6 , o que ainda é bem pequeno se comparado com o número de Avogrado ($6,023 \times 10^{23}$). Em sistemas pequenos, uma grande parte das moléculas estão próximas o suficiente das bordas e sofrem influência sob o que está “por trás” das paredes. Nesses casos, os efeitos de borda e superfície não podem ser negligenciados (Van Gunsteren *et al.*, 2006). Esses efeitos são minimizados utilizando condições periódicas de contorno (PBC). A caixa de simulação que contém o sistema molecular é cercada por infinitas cópias dela mesma. Se

uma partícula deixar a caixa de simulação, ela é substituída por outra idêntica que entra pelo lado oposto. O aspecto chave das PBC's é que nenhuma partícula irá interagir com o vácuo dentro do alcance do raio de corte das interações não-ligadas (Cramer, 2004). Além disso, a massa, número de partículas e a energia total da caixa se mantêm constantes.

2.5 ENSAIOS DE ELISA

ELISA (*enzyme-linked immunosorbent assay*) é uma técnica de ensaio baseada em placas projetada para detectar e quantificar substâncias como peptídeos, proteínas, anticorpos e hormônios. Em um ELISA, um antígeno deve ser imobilizado em uma superfície sólida e depois complexado com um anticorpo que está ligado a uma enzima. A detecção é realizada avaliando a atividade da enzima conjugada através da incubação com um substrato para produzir um produto mensurável. O elemento mais crucial da estratégia de detecção é uma interação anticorpo-antígeno altamente específica.

Os ELISAs são tipicamente realizados em placas de poliestireno de 96 poços, onde se ligam passivamente anticorpos ou proteínas. Uma enzima de detecção pode ser ligada diretamente ao anticorpo primário ou introduzida através de um anticorpo secundário que reconhece o anticorpo primário.

Depois de fixar o antígeno, é necessário ligar um marcador acoplado a anticorpo, para detectar posteriormente a presença do antígeno esperado. Esse marcador pode ser fluorescente ou pode ser uma enzima que catalize uma reação que produza um produto mensurável. E essa ligação se dá por anticorpos específicos ao antígeno, por isso a precisão do teste. O marcador de detecção pode ser ligado diretamente ao anticorpo primário ou introduzido por meio de um anticorpo secundário que reconhece o anticorpo primário.

O ELISA indireto faz o uso de dois anticorpos: o anticorpo primário e o anticorpo secundário. Primeiramente, a amostra é preparada com o anticorpo primário, que possui especificidade com o antígeno a ser detectado. Caso o antígeno a ser procurado esteja presente na amostra, o anticorpo primário irá se ligar ao seu alvo. Em seguida, haverá uma lavagem para retirar os anticorpos que não se ligaram ao antígeno teste. A seguir, coloca-se os

antígenos secundários que possuem especificidade apenas para os anticorpos primários. Caso os anticorpos primários tenham se ligado ao antígeno, haverá ligação entre anticorpo primário e anticorpo secundário e isso que irá permitir a detecção.

3 DESENVOLVIMENTO IN SILICO DE ANTÍGENOS RECOMBINANTES BASEADOS EM NS1 DE DENGUE

A dengue é uma doença viral aguda sistêmica que se estabeleceu globalmente em ambos os ciclos de transmissão endêmicas e epidêmicas. Cerca de 390 milhões de pessoas são infectadas a cada ano, dos quais 96 milhões apresentam manifestação aparente da doença (Bhatt *et al.*, 2013). O vírus da dengue (DENV) é transmitido pelos mosquitos *Aedes aegypti* e *Aedes albopictus* e é encontrado em 4 diferentes sorotipos: DENV-1, -2, -3 e -4 (Omokoko *et al.*, 2014). A área geográfica em que a transmissão da dengue ocorre se expandiu nas ultimas décadas e todos os 4 sorotipos circulam na África, Ásia e Américas (Guzman *et al.*, 2010).

O diagnóstico laboratorial precoce é importante para garantir a implementação rápida de um tratamento clínico adequado. O DENV pode ser detectado através do isolamento do vírus, amplificação do RNA por PCR em tempo real (RT-qPCR) ou em imunoenaios para detectar antígenos virais (Hunsperger *et al.*, 2014). Porém, o isolamento do vírus requer condições laboratoriais para o meio de cultura celular, enquanto que o RT-PCR não é viável em condições de baixo orçamento. Atualmente, o teste de diagnóstico de dengue mais amplamente utilizado continua a ser a captura de IgM anti-DENV através de Enzyme-Linked Immunosorbent Assay (ELISA) (Innis *et al.*, 1989; Hunsperger *et al.*, 2014). Porém, trata-se de um método que carece de sensibilidade e especificidade na fase aguda da doença (Peeling *et al.*, 2010; Hunsperger *et al.*, 2014). Novos métodos de diagnóstico de dengue vêm sendo desenvolvidos nos últimos anos. Dentre esses métodos, a detecção da proteína NS1 merece destaque. Entretanto, o método não é capaz de distinguir entre os diferentes sorotipos e sua sensibilidade varia dependendo do tipo de infecção – se primária ou secundária (Cordeiro, 2012).

O vírus da dengue pertence ao gênero flavivírus, dentro da família *Flaviridae* e seu RNA codifica três proteínas estruturais – capsídeo (C), pre-membrana/membrana (prM/M) e envelope (E) – e sete proteínas não-estruturais (NS1, NS2A, NS2B, NS3, NS4A, NS4B e NS5) (Guzman *et al.*, 2010). A NS1 (Figura 16) é uma glicoproteína de 352 resíduos que é secretada pelas células infectadas e não participa da formação do virion

(Omokoko *et al.*, 2014). A proteína pode ser encontrada na membrana das células infectadas ou solúvel na corrente sanguínea (Chuang *et al.*, 2013). Na fase aguda de infecção já é detectada a presença de NS1 no sangue e a mesma é capaz de ativar o sistema complemento (Kurosu *et al.*, 2007).

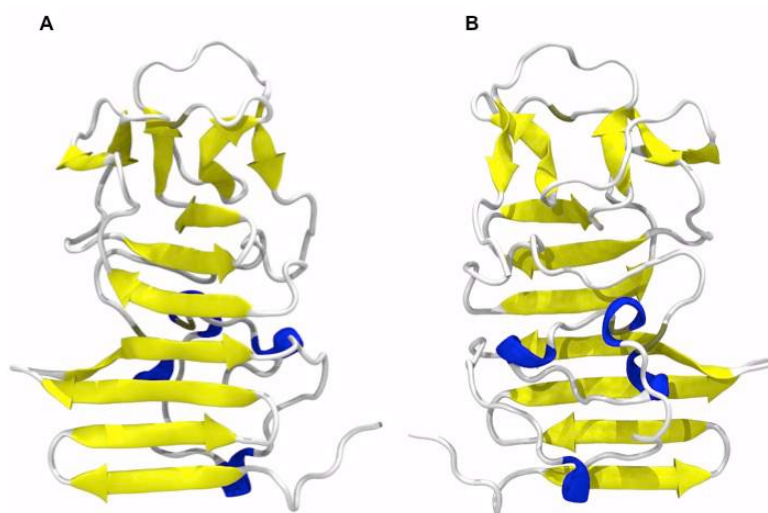


Figura 16 - Estrutura cristalográfica da NS1 de DENV-1 (PDB ID: 4OIG). (a) Face estruturada e (b) face não-estruturada.

A infecção com um sorotipo de DENV fornece proteção contra esse sorotipo específico pro resto da vida, mas fornece apenas proteção de curto prazo contra infecção dos outros sorotipos (Aguiar *et al.*, 2016; Whitehead, 2016). Estudos epidemiológicos determinaram que a imunidade preexistente a um sorotipo de DENV é um fator de risco numa infecção secundária devido ao fenômeno de *antibody-dependent enhancement* (ADE) (Bhatt *et al.*, 2013; Whitehead, 2016). Na ADE os anticorpos heterotípicos não-neutralizantes criados na primeira infecção ligam-se aos vírus sem neutralizá-los, aumentando a eficiência da infecção. Por isso, no desenvolvimento de uma vacina contra dengue é necessário que a mesma tenha alta eficácia contra todos os 4 sorotipos.

Atualmente a única vacina licenciada contra dengue demonstrou diferentes eficácias contra os quatro sorotipos. O Instituto Sanofi Pasteur propôs uma vacina tetravalente viva, recombinante e atenuada, composta de 3 doses, com eficácia de 50% contra DENV-1, 35% contra DENV-2, 78.4% contra DENV-3 e 75.3% contra DENV- 4, em testes fase 3 com crianças

asiáticas (Capeding *et al.*, 2014). A mesma vacina obteve eficácia de 50.3% para DENV-1, 42.3% para DENV-2, 74.0% para DENV-3 e 77.7% para DENV-4 em testes fase 3 com crianças da América Latina (Villar *et al.*, 2015). No entanto, a eficácia da vacina para casos confirmados de dengue foi menor em indivíduos soronegativos do que em indivíduos soropositivos (Aguiar *et al.*, 2017). Os resultados sugerem altas taxas de proteção em indivíduos previamente imunizados parcialmente contra a dengue (soropositivos), mas taxas altas de internações acompanhadas de infecções por dengue em indivíduos soronegativos vacinados, i. e. nunca infectados com dengue antes da vacinação (Aguiar *et al.*, 2017). Isto ocorreu pois os indivíduos soronegativos vacinados apresentaram ADE devido aos anticorpos não-protetores provenientes da vacina. Pode ser relevante o fato da vacina da Sanofi não conter proteínas não-estruturais derivadas de DENV, sendo substituídas por proteínas de febre amarela, e que isso pode ter impedido a indução de células T protetoras (Halstead, 2017).

Recentemente, uma nova vacina está sendo desenvolvida pelo órgão americano National Institute of Allergy and Infectious Diseases (NIH) em parceria com o Instituto Butantan e encontra-se já em testes fase 3. É também uma vacina tetravalente viva e atenuada, porém difere da vacina da Sanofi por utilizar apenas uma dose e conter as proteínas não-estruturais em sua composição (Whitehead, 2016).

Com o intuito de expor regiões imunogênicas da proteína NS1, nós propomos um novo protocolo, utilizando a engenharia de proteínas para o planejamento de proteínas quiméricas com potências de diagnóstico de dengue. Através da estratégia de “*epitope-scaffolding*”, elementos estruturais da NS1 foram transplantados para uma proteína aceptora mais robusta, conhecida como *scaffold*. A identificação das sequências antigênicas da NS1 pode ser feita através de dados experimentais (Omokoko *et al.*, 2014) e de cálculos de hidrofobicidade (Kyte e Doolittle, 1982). Para o scaffold deve-se escolher proteínas que apresentem marcante estabilidade estrutural e térmica. Neste trabalho foi escolhida a proteína Top7 (Kuhlman *et al.*, 2003), pois já é conhecida na literatura o seu uso para desenvolvimento de novos antígenos (Boschek *et al.*, 2009; Soares *et al.*, 2010; Viana, I. *et al.*, 2013).

Para avaliar a estabilidade estrutural das quimeras, três ferramentas computacionais foram utilizadas: *threading*, *de novo design* e dinâmica molecular. Em seguida, as proteínas quiméricas foram também avaliadas quanto à sua performance imunológica, através de ensaios de ELISA.

3.1 OBJETIVOS

Desenhar e avaliar proteínas quiméricas capazes de expor e estabilizar os motivos estruturais de epítomos putativos de NS1 como potenciais candidatos a antígenos de diagnóstico utilizando a estratégia de “*epitope-scaffolding*” em conjunto com técnicas computacionais e experimentais.

3.1.1 Objetivos específicos

- Selecionar epítomos putativos da NS1;
- Transplantar os epítomos para as proteínas *scaffolds* e avaliar energeticamente a sua estabilidade ;
- Simular as quimeras mais estáveis e estruturalmente favoráveis através de dinâmica molecular;
- Produzir as proteínas estáveis em sistema procarioto e avaliar as imunorreatividades através de ensaios de ELISA;

3.2 METODOLOGIA

3.2.1 Seleção das sequências de NS1

A partir da sequência primária de aminoácidos da NS1 DENV-1, foram selecionadas regiões hidrofílicas que tendem a estar expostas ao solvente e aptas para reconhecimento por anticorpos. Para a identificação dessas sequências foi utilizado o gráfico de hidropatia de Kyte-Doolittle (Johnson *et al.*; Kyte e Doolittle, 1982), obtido com uma janela de tamanho 9 e a

sequência de aminoácidos 178 a 352 da NS1 DENV-1 (sequência descrita no PDB ID 4OIG). Além das sequências identificadas pelo gráfico de hidropatia, a região descrita experimentalmente por Omokoko e colaboradores (Omokoko *et al.*, 2014) também foi utilizada.

3.2.2 Threading

Para investigar em quais regiões das proteínas carreadoras cada sequência da NS1 escolhida pode ser acomodada, foi realizado um *threading* em toda a extensão de cada proteína carreadora. Através do protocolo *pmut_scan_parallel* (Kuhlman *et al.*, 2003) do programa Rosetta v3.5 foram realizados os *side-chain grafting*. A estabilidade relativa entre a proteína nativa e a mutante foi analisada através dos valores de $\Delta\Delta G$ calculados a partir da função de energia do programa. A amostragem do rotâmero χ_1 e dos rotâmeros χ_2 , χ_3 e χ_4 foram incrementados de ± 4 e ± 1 desvios-padrões, respectivamente. Todos os resíduos foram considerados enterrados para utilizar os rotâmeros extras. Os parâmetros utilizados na linha de comando foram: `-mutants_list mutante_file -ex1:level 4 -ex2 -ex3 -ex4 -use_input_sc -ignore_unrecognized_res -no_his_his_pairE -mute basic core -s native_scaffold.pdb -output_mutant_structures -extrachi_cutoff 1 -multi_cool_annealer 10 -DDG_cutoff 1000`.

3.2.3 De novo design

Para a análise de *de novo design* foi utilizado o protocolo *AbInitioRelax* (Rohl *et al.*, 2004; Raman *et al.*, 2009) do programa Rosetta v3.5. A fim de uma melhor convergência em direção à estrutura nativa, foram realizadas 20.000 simulações para cada proteína. As bibliotecas de 3 e 9 fragmentos foram obtidas do servidor Robetta (<http://robetta.bakerlab.org>). Para aumentar a amostragem conformacional, os ciclos foram incrementados por um fator de 10. A amostragem dos rotâmeros χ_1 e χ_2 foram incrementados de ± 1 desvio padrão. A contribuição do raio de giro no cálculo de energia foi redistribuído por um fator de 0.5. Os parâmetros utilizados na linha de

comando foram: `-in:file:fasta fasta_file -in:file:native native_scaffold.pdb -in:file:frag3 frag3_lib -in:file:frag9 frag9_lib -nstruct 1 -out.pdb -abinitio::increase_cycles 10 -relax::fast -abinitio:relax -ex1 -ex2 -no_filters -abinitio::rg_reweight 0.5 -output_all -overwrite`. Para computar o RMSD foi utilizada a estrutura cristalográfica das proteína Top7 nativa (PDB ID 1QYS).

3.2.4 Dinâmica molecular

As proteínas quiméricas consideradas viáveis nas etapas anteriores foram analisadas por MD, através do programa Gromacs 4.6.5 (Hess *et al.*, 2008). As coordenadas iniciais das proteínas foram obtidas das estruturas escolhidas no *de novo design*. As proteínas foram centradas numa caixa cúbica de dimensões 8.0 x 8.0 x 8.0 nm e explicitamente solvatadas usando o modelo de água SPC (Berendsen *et al.*, 1981). Condições periódicas de contorno foram utilizadas nas direções -x, -y e -z. Simulações de dinâmica molecular de 100 ns de duração foram realizadas usando o *ensemble* NPT e campo de força atomístico GROMOS 54A7 (Schmid *et al.*, 2011), com parâmetros dos íons do campo de força GROMOS 53A6 (Oostenbrink *et al.*, 2004). As ligações envolvendo hidrogênio foram restringidas usando o método LINCS (Hess *et al.*, 1997). As temperaturas dos solutos e do solvente foram acopladas separadamente a um termostato usando o esquema de reescalonamento de velocidade (Bussi *et al.*, 2007), com temperatura de referência de 310 K e tempo de relaxação de 0.5 ps. A pressão foi mantida constante isotropicamente a 1.013 bar utilizando um barostato de Berendsen (Berendsen *et al.*, 1984), com tempo de relaxação de 0.5 ps e compressibilidade (κ_T) de $4.5 \times 10^{-5} \text{ bar}^{-1}$. Um raio de corte de 1.4 nm foi utilizado para interações não-ligadas de curto alcance. Interações de longo alcance foram levadas em conta usando o método de reação de campo (Tironi *et al.*, 1995), com $\epsilon=66$ além do raio de corte de 1.4 nm. Contra-íons foram adicionados aos sistemas para assegurar a eletroneutralidade e uma concentração salina de 0.150 mM. Os sistemas foram inicialmente minimizados energeticamente usando 5.000 passos do algoritmo *steepest*

descent. Em seguida, a integração da dinâmica foi realizada usando um tempo de integração de 2 fs baseado no algoritmo *leap-frog* (Hockney, 1970).

3.3 PREPARAÇÃO DOS VETORES PARA EXPRESSÃO

As sequências de DNA que codificam as proteínas selecionadas ao final das análises computacionais foram comercialmente sintetizadas (GenScript). Os DNA's foram otimizadas visando maximizar sua expressão em *Escherichia coli*. Foram levados em consideração parâmetros como uso de códons, estrutura secundária do mRNA, distribuição do conteúdo de GC, sequências longas repetitivas de DNA e motivos repetitivos de DNA para permitir um aumento significativo nos níveis de expressão das proteínas recombinantes em células procarióticas. As sequências otimizadas foram, em seguida, fusionadas a sítios de clonagem: 5' (Nco I e Nhe I) e 3' (Xho I e Hind III).

3.3.1 Preparação do vetor de expressão

Foi utilizado o vetor de expressão pRSET A, que é um vetor concebido para alto nível de expressão e purificação de proteínas a partir de genes clonados em *E. coli* e que permite fusionar a proteína a uma cauda de 6 histidinas (6X His-tag) na porção N-terminal para purificação por cromatografia de afinidade. O vetor pRSET A foi gentilmente cedido pelo Dr. André Filipe Pastor da Silva/ Lavite – IAM/FIOCRUZ-PE.

Uma alíquota de 50 µL de células competentes *E. coli* NEB5α foi transformada por choque térmico com ~100 ng do vetor pRSETA. A reação de transformação foi incubada durante 30 minutos no gelo, sendo em seguida submetida a choque térmico em 37°C por 5 minutos e esfriada novamente no gelo. Após transformação, as células foram semeadas em placa de meio LB sólido contendo ampicilina (LB-AMP) a concentração de 50 mg·mL⁻¹ e incubadas a 37 °C por 16h. Uma colônia foi inoculada em 250 mL de meio líquido LB-AMP 100 µg·mL⁻¹ e mantida a 37 °C sob agitação por 16 h para crescimento. A cultura foi centrifugada a 10.000 RPM/4 °C por 10 minutos e o

DNA plasmidial foi extraído com o kit QIAprep Midiprep (Qiagen), de acordo com as recomendações do fabricante.

O produto da Midiprep foi quantificado por espectrofotometria, através do NanoPhotometer Pearl UV/Vis (Implen). O DNA obtido foi linearizado através de digestão com as enzimas Nhe I e Hind III (Biolabs Inc.). A reação foi realizada a volume final de 50 μ L contendo: 1X tampão de digestão 2.1, 8 μ g do vetor e 50 Unidades - U de cada enzima. A digestão foi realizada à 37 °C, durante 4 horas. Para confirmar a linearização, os produtos da digestão, juntamente com o vetor intacto, foram fracionados em gel de agarose 1 %, corado com brometo de etídio. Identificada a linearização, o fragmento foi então purificado do gel utilizando o kit comercial GMXTM PCR DNA and Gel Band Purification kit (GE Healthcare), de acordo com as recomendações do fabricante, e quantificado por espectrofotometria.

3.3.2 Preparação dos genes sintéticos

As sequências de DNA otimizadas foram comercialmente obtidas no vetor pUC57 (GenScript®). Alíquotas de 50 μ L de células competentes *E. coli* NEB5 α foram transformadas por choque térmico com ~100 ng de cada DNA. As reações de transformação foram incubadas durante 30 minutos no gelo, sendo em seguida submetidas a choque térmico em 37°C por 5 minutos e esfriadas novamente no gelo. Após transformação, as células foram semeadas em placa de meio LB-AMP sólido e incubadas a 37 °C por 16h. Uma colônia de cada construção foi inoculada em 250 mL de meio líquido LB-AMP 100 μ g·mL⁻¹. Os meios inoculados foram mantidos a 37 °C sob agitação por 16 h para crescimento. As culturas bacterianas foram centrifugadas a 10.000 RPM/4 °C por 10 minutos e os DNA's plasmidiais foram extraídos com o kit QIAprep Midiprep (Qiagen), de acordo com as recomendações do fabricante e quantificados por espectrofotometria, através de NanoPhotometer Pearl UV/Vis (Implen).

Os DNA's foram submetidos à reação de digestão para identificação dos genes e posterior extração e purificação. As reações de digestão foram realizadas com as enzimas Nhe I e Hind III (Biolabs Inc.), com volume final de

50 µL cada, contendo: 1X tampão de digestão 2.1, 8 µg de cada vetor e 50 Unidades - U de cada enzima. As digestões foram realizadas à 37 °C, durante 4 horas. Os produtos das digestões foram fracionados em gel de agarose 1% corado com brometo de etídio e a identidade de cada sequência foi analisada através do padrão de bandas gerado. Os fragmentos relativos a cada gene otimizado foram extraídos do gel e purificados utilizando o kit comercial GMX™ PCR DNA and Gel Band Purification kit (GE Healthcare), de acordo com as recomendações do fabricante, e quantificados por espectrofotometria.

3.3.3 Subclonagem dos genes no vetor pRSET A

Os genes purificados foram subclonados no vetor pRSET A, previamente purificado. Para as reações de ligação entre os genes e o vetor de expressão, foi fixada uma quantidade de 30 ng/reação do vetor de expressão e volume final de 10 µl. Foram utilizadas três diferentes quantidades dos genes: proporções de 1:1, 3:1 e 5:1 em relação ao vetor. Em cada reação foram utilizados: 1 µL de 10X Ligase Buffer, 400 U de T4 DNA Ligase (Biolabs®) e água Milli-Q em quantidade suficiente para completar os 10 µl. Como controle negativo, foram utilizadas reações contendo apenas o vetor linearizado, sem os genes sintéticos. As reações de ligação foram realizadas por 18 h / 16 °C, e foram em seguida utilizadas para a transformação de células competentes DH5α (Invitrogen®), visando à confirmação das clonagens. Cinco colônias provenientes de cada ligação foram selecionadas para preparação de DNA plasmidial, utilizando o kit QIAprep Midiprep (Qiagen). A identidade de cada gene foi analisada através de provas de digestão, utilizando as enzimas Nhe I e Hind III. As reações de digestão foram realizadas à 37 °C, durante 6 horas, e os produtos foram fracionados em gel de agarose 1% corado com brometo de etídio.

3.4 EXPRESSÃO E PURIFICAÇÃO DAS PROTEÍNAS

Expressão bacteriana das proteínas quiméricas foi realizada utilizando *Escherichia coli* BL21 (Invitrogen®), transformada com os vetores pRSET A contendo os genes das respectivas proteínas. As expressões em larga escala foram realizadas à temperatura de 37 °C e concentração de 1 mM do indutor β -1-D-isopropil-tiogalactopiranosida (IPTG), em 1000 mL de meio LB contendo 100 mg·mL⁻¹ de ampicilina. Após as 4h de expressão, os meios foram centrifugados a 10.000 RPM / 4 °C por 10min. Os *pellets* foram ressuspensos em 10 mL de tampão de lise contendo: 50 mM de Tris-HCl, NaCl 300 mM, imidazol 20 mM e pH 8.0. As células foram sonicadas em banho de gelo usando seis pulsos de 30s cada, com intervalo de 60s entre os pulsos. Os lisados foram então centrifugados a 10.000 RPM / 4 °C por 10 min, adicionados de inibidor de protease na concentração 1X, para remover os restos celulares.

As proteínas foram purificadas por cromatografia de afinidade, utilizando um equipamento de cromatografia líquida de proteína rápida (*fast protein liquid chromatography*, FPLC) Äkta Pure da GE Health Care®, com coluna de níquel de 5 mL Hi-Trap His. As proteínas fusionadas à cauda de histidina foram eluídas em Tampão C (50 mM Tris-HCl pH 8.0, NaCl 300 mM, imidazol 0.5 M) e a quantificação das mesmas foi realizada por espectrofotometria e correção pelo coeficiente de extinção molar, de cada uma das sequências de aminoácidos.

3.5 AVALIAÇÃO IMUNOLÓGICA POR ENSAIO DE ELISA

Placas de 96 poços (Costar®) foram revestidas com 16 ou 8 pmol/poço de cada proteína durante 16 h a 4 °C. Os revestimentos foram bloqueados com leite desnatado a 5% (m/v) (Bio-Rad) em tampão PBS-T (1X PBS com Tween-20 a 0,05% (v/v) durante 15 minutos em temperatura ambiente. Em seguida, os anticorpos primários (soros de pacientes infectados com DENV-3 e não infectados) foram diluídos na proporção de 1:100 em tampão de ensaio (5% m/v de leite desnatado em PBS-T) e adicionados à placa durante 1 h à temperatura ambiente. Foram utilizados 16

soros de pacientes positivos (4 para DENV-1; 4 para DENV-2, 3 para DENV-3; 1 para DENV-4 e 4 para DENV-3/4) e 4 soros de pacientes negativos para DENV.

As placas foram lavadas cinco vezes com PBS-T para remover qualquer excesso de anticorpo não ligado. Em seguida, diluiu-se anticorpo secundário anti-IgG humano conjugado com peroxidase (Jackson ImmuneResearch) a uma concentração de 1:10000 ou 1: 20000 em tampão de ensaio, dependendo de qual proteína, e adicionou-se aos poços. As placas foram incubadas durante 1 h à temperatura ambiente e posteriormente lavadas com PBS-T. Foi realizada uma incubação com o substrato TMB-KPL durante 30 minutos à temperatura ambiente, seguida da adição de HCl 1 N para parar a reação e posterior detecção colorimétrica.

As densidades ópticas foram obtidas em um comprimento de onda de 450 nm (OD450nm) num espectrofotômetro de microplacas (BioTek). Todas as amostras foram testadas em duplicata e apenas valores cuja variabilidade intra-ensaio foi inferior a 20% foram levados em consideração.

3.6 RESULTADOS E DISCUSSÃO

3.6.1 Sequências de NS1 selecionadas

De acordo com Kyte-Doolittle, regiões com tendência para exposição ao solvente são aquelas que possuem valores negativos no gráfico de hidropatia (Kyte e Doolittle, 1982). O gráfico de hidropatia de Kye-Doolittle para a NS1 DENV-1 (Figura 17) mostra que se trata de uma proteína com vários picos de *scores* negativos, indicando ser uma proteína globular. A fim de selecionar regiões hidrofílicas da proteína, foram escolhidos 6 picos negativos correspondentes às sequências da Tabela 1. A região descrita por Omokoko (Omokoko *et al.*, 2014) como a mais provável de conter um epítipo também foi selecionada. Estas foram as sequências utilizadas para o *threading*.

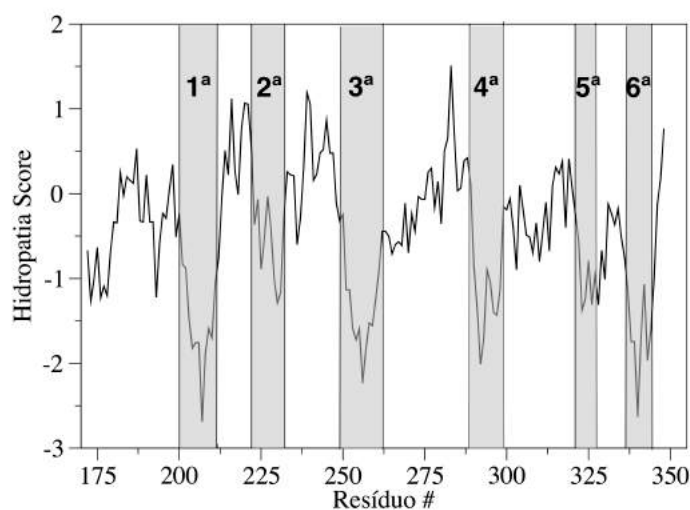


Figura 17 - Gráfico de hidropatia de Kyte-Doolittle para a região dos resíduos 172-352 da NS1 de DENV-1.

Deve-se escolher proteínas robustas como carreadoras de epítomos, que possam aguentar grandes mutações sem um ganho energético muito alto. Neste trabalho foram utilizadas 4 proteínas (Figura 18): a Top7, uma proteína sintética desenhada computacionalmente pelos desenvolvedores do programa Rosetta (Kuhlman *et al.*, 2003); a Sso7d, uma proteína dos microrganismos extremófilos *Sulfolobus solfataricus* (Knapp *et al.*, 1996); a Tioredoxina A, uma proteína envolvida em processos de oxidação (Stefankova *et al.*, 2005); e a Proteína GB1, uma proteína de microrganismo *Streptococcal* (Gallagher *et al.*, 1994).

Tabela 1 - Sequências da NS1 de DENV selecionadas a partir do gráfico de hidropatia de Kyte-Doolittle

Epítopo	Sequência	Posição na NS1	Motivo
1	YWIESEK	200-206	Folha β
2	TCIWPKSHTLW	222-232	Não-estruturado
3	GPISQHNYPGYFT	249-262	Não-estruturado
4	EHCGNRGPSLR	289-299	Não-estruturado
5	LRFKGE	321-326	Folha β
6	VKEKEEN	338-344	Não-estruturado
7 ^a	SHTLWSNGVLESE	228-240	Não-estruturado
8 ^a	WPKSHTLWSNGVLESEMIIPK	225-245	Não-estruturado

^a Sequências descritas por (Omokoko *et al.*, 2014)

Neste trabalho foi possível apenas avaliar experimentalmente proteínas desenhadas utilizando a Top7 como proteína carreadora. Dessa forma, serão discutidos apenas os resultados para três quimeras baseadas na Top7. Porém, vale ressaltar que através do protocolo utilizado neste trabalho, combinando métodos de engenharia de proteína com dinâmica molecular, foi possível avaliar mais de 1.100 quimeras por métodos computacionais. Dessa forma, o protocolo se mostrou uma poderosa ferramenta de triagem para seleção de candidatos a antígenos de DENV.

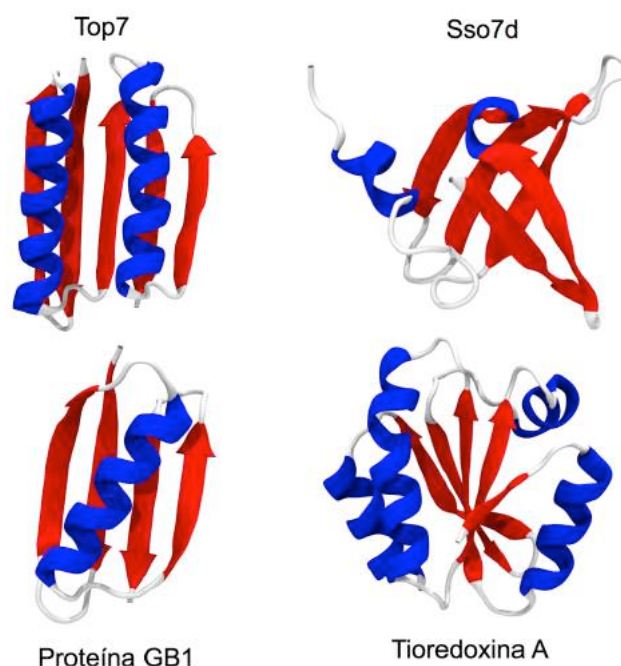


Figura 18 - Representação *cartoon* das proteínas utilizadas como *scaffolds*. Folhas- β estão representadas como setas na cor vermelha, hélices- α como espirais na cor azul e regiões não estruturadas na cor branca.

3.6.2 Quimeras baseadas na TOP7

3.6.2.1 Threading e de novo design

Através do *threading*, foram analisadas todas as sequências da Tabela 1 em todas as possíveis posições na Top7, totalizando 488 proteínas. As proteínas nativas em geral são apenas de 5 a 10 kcal·mol⁻¹ mais estáveis do que seus estados desnaturados (Creighton, 1990; Dill *et al.*, 2008). Como essa barreira energética de desnaturação é baixa, pequenas mutações nas proteínas nativas podem desestabilizá-las por completo. Dessa forma, foram consideradas viáveis as estruturas quiméricas cuja diferença de energia em relação ao *scaffold* nativo é menor ou igual a 5 unidades de *score* do Rosetta (REU). Também só foram consideradas mutações que mantivessem, total ou parcialmente, o motivo estrutural do epítipo na NS1 nativa, i.e. folhas β foram alocadas apenas em regiões folhas β da Top7, e regiões não estruturadas similarmente. O resultado do *threading* para a Top7 encontra-se na Tabela 2.

Tabela 2 - Proteínas construídas usando a Top7 como *scaffold* através do protocolo de *threading*.

Código	Epítopo	Posição na Top7	$\Delta\Delta G^*$
1S15-21_Top7	1	15-21	2,995
1S16-22_Top7	1	16-22	4,714
1S19-25_Top7	1	19-25	-1,329
1S75-81_Top7	1	75-81	4,243
2S20-30_Top7	2	20-30	2,712
5S4-9_Top7	5	4-9	0,676
5S8-13_Top7	5	8-13	4,699
5S13-18_Top7	5	13-18	1,469
5S79-84_Top7	5	79-84	3,814
7S22-30_Top7	7 ^a	22-30	3,525
7S20-28_Top7	7 ^b	20-28	0,836
7S51-59_Top7	7 ^b	51-59	4,327

^a Foram utilizados apenas os resíduos 230-238

^b Foram utilizados apenas os resíduos 232-240

* Unidade Rosetta score (REU)

Da Tabela 2 apenas a quimera 2S20-30_Top7, destacada em negrito, pôde ser avaliada experimentalmente. Por isso, as demais análises serão discutidas apenas para essa proteína.

O *de novo design* possui uma função de energia mais refinada do que o *threading*. O mínimo de energia global para o funil de enovelamento da Top7 nativa é de -255 REU (Cunha *et al.*, 2015). Tomando esse valor como referência, e mantendo o critério energético ($\Delta\Delta G \leq 5$ REU), foram consideradas viáveis as estruturas com mínimo global ≤ -250 REU no funil de enovelamento. Para obtenção do funil de enovelamento, o RMSD foi calculado em relação ao *backbone* da Top7 nativa a fim de investigar a manutenção da conformação global do *scaffold*.

No caso da molécula 2S20-30_Top7, o *de novo design* indica uma proteína com mínimo global de aproximadamente -248 REU, apenas 2

unidades de *score* acima do máximo estabelecido (Figura 19). Além disso, o funil de enovelamento é bastante semelhante ao da Top7 original (Cunha *et al.*, 2015). É possível observar duas populações conformacionais, sendo a mais populosa com valores de RMSD abaixo de 5 Å. O RMSD da estrutura de menor energia amostrada foi de aproximadamente 1,7 Å, sendo bastante semelhante ao *scaffold* nativo (Figura 21 A). Considerando a alta similaridade estrutural e o funil de enovelamento estar bem próximo do *cut-off*, decidiu-se que valeria à pena avaliar essa estrutura por MD. A sequência da 2S20-30_Top7 encontra-se na Tabela 3.

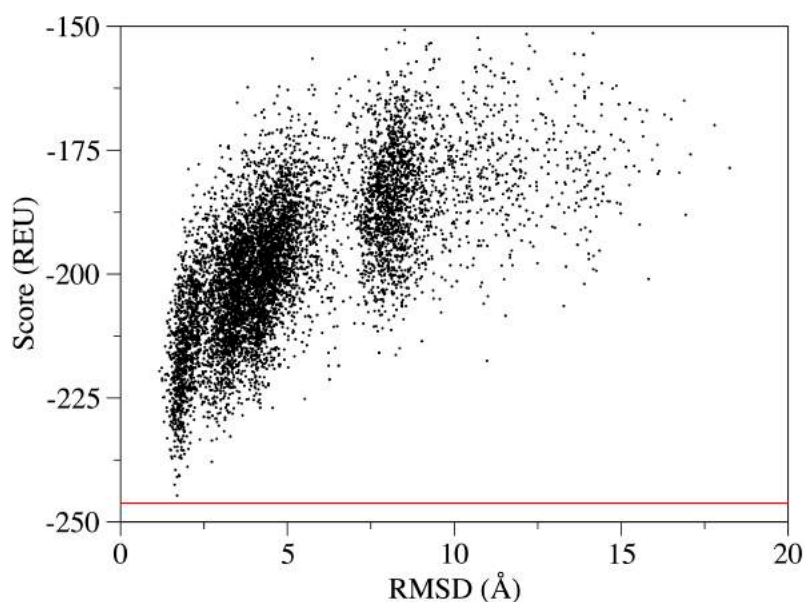


Figura 19 - RMSD em função do Rosetta *score* para a quimera 2S20-30_Top7 construída utilizando o epítipo 2 e a Top7. A linha em vermelho indica a energia mínima global do perfil de enovelamento, neste caso de -248 REU.

Outras duas proteínas quiméricas foram construídas utilizando os epítopos 7 e 8 da Tabela 1. Para tal, as sequências de aminoácidos dos epítopos foram inseridos entre os aminoácidos T22 e E24 da Top7 (Tabela 3).

Tabela 3 - Sequência de aminoácidos das proteínas construídas utilizando os epítomos 2, 7 e 8 e a Top7 como *scaffold*. A região do epítopo encontra-se destacada em negrito.

Código	Sequência
Top7	DIQVQVNIDDNGKNFDYTYTVTTESELQKVLNELMDYIKKQGAKRV RISITARTKKEAEKFAAILIKVFAELGYNDINVTFDGDTVTVEGQL
2S20-30_Top7	DIQVQVNIDDNGKNFDYTY TCTWPKSHTLWL NELMDYIKKQGAKR VRISITARTKKEAEKFAAILIKVFAELGYNDINVTFDGDTVTVEGQL
7S22-24_Top7	DIQVQVNIDDNGKNFDYTYVT HTLWSNGVLESE LQKVLNELMDYI KKQGAKRVRISITARTKKEAEKFAAILIKVFAELGYNDINVTFDGDTV TVEGQL
8S22-25_Top7	DIQVQVNIDDNGKNFDYTYTV WPKSHTLWSNGVLESE MIIPKLQKV LNELMDYIKKQGAKRVRISITARTKKEAEKFAAILIKVFAELGYNDIN VTFDGDTVTVEGQL

O *de novo* design foi realizado para essas duas quimeras, porém não é possível comparar os funis de enovelamento com a Top7 original devido ao número de aminoácidos ser diferente. Entretanto, para cada simulação, podemos comparar todas as estruturas amostradas ao modelo de menor energia. Se o gráfico formar um funil de enovelamento com mínimos de energia locais ou globais, assume-se que as baixas energias são indicativas de estruturas nativas (Silva *et al.*, 2016; Weitzner *et al.*, 2017). No caso da quimera 7S22-24_Top7, o funil possuiu mínimo global de -256 REU (Figura 20 A). Para a proteína 8S22-25_Top7 o mínimo global foi de -269 REU (Figura 20 B). Ambas as quimeras possuem um perfil de enovelamento no formato de funil, indicando que as estruturas de menor energia são representativas da estrutura nativa.

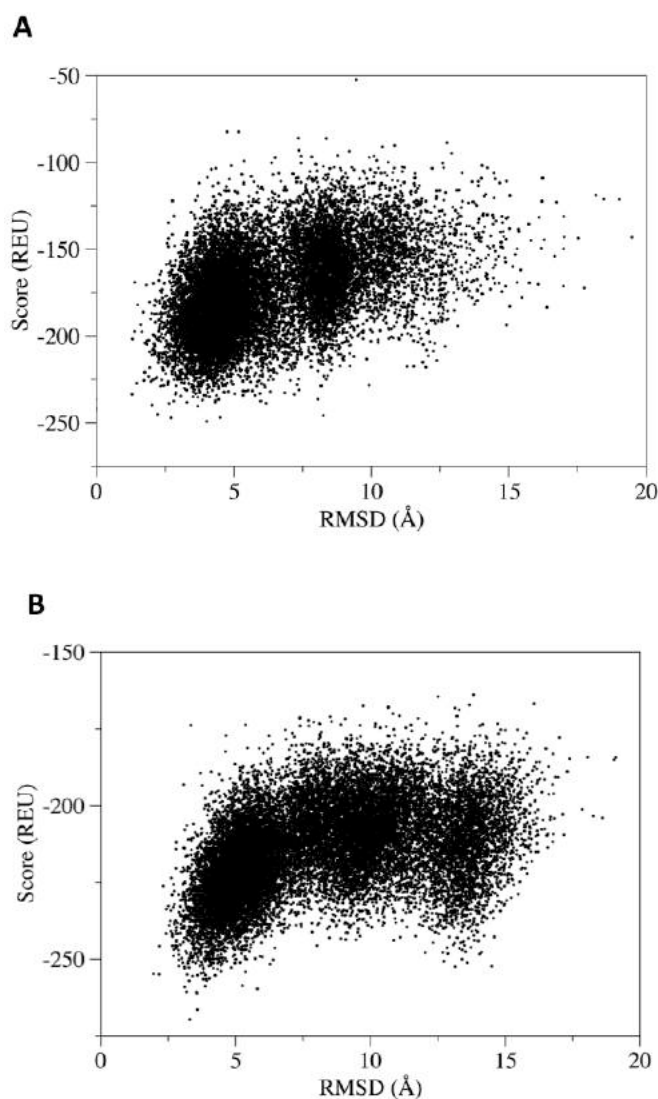


Figura 20 - Rosetta Score em função do RMSD das proteínas 7S22-24_Top7 e 8S22-25_Top7. O RMSD foi calculado em relação às estruturas de menor energia de cada *ensemble*, e ambas formam um funil de enovelamento, indicando serem estruturas nativas e proteínas estáveis.

Uma inspeção visual das estruturas simuladas de menor energia para as quimeras 2S20-30_Top7, 7S22-24_Top7 e 8S22-25_Top7 (Figura 21) mostra que apesar da inclusão dos epítomos, houve manutenção da conformação global do *scaffold*.

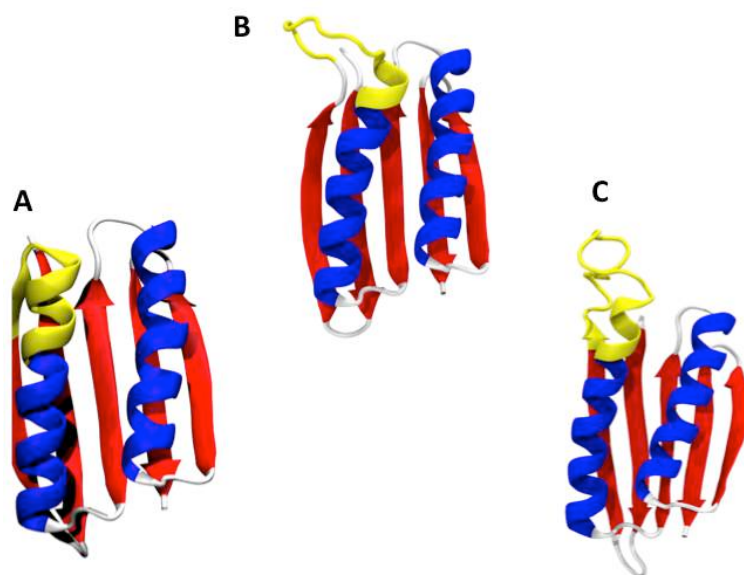


Figura 21 - Representação *cartoon* das estruturas de menor energia das proteínas (A) 2S20-30_Top7, (B) 7S22-24_Top7 e (C) 8S22-25_Top7. Folhas- β estão representadas como setas na cor vermelha, hélices- α como espirais na cor azul e regiões não estruturadas na cor branca. As regiões correspondentes aos epítomos estão destacadas em amarelo.

3.6.2.2 Dinâmica molecular

Simulações de MD foram realizadas para investigar se as conformações nativas do epítipo e do *scaffold* se mantêm. Para avaliar a estabilidade estrutural, foram analisados os valores de RMSD dos átomos dos *backbones* em relação às estruturas iniciais ao longo do tempo. Os conteúdos de estruturas secundária ao longo do tempo de simulação também foram analisados, utilizando o software DSSP (*Database of Secondary Structure in Proteins*) (Kabsch e Sander, 1983). A Top7 nativa será utilizada como referência para as análises.

A Top7 nativa possui valores de RMSD que variam entre 0.1 e 0.2 nm (Figura 22 A), que correspondem a uma torção das hélices α (Cunha *et al.*, 2015). Analisando o seu conteúdo de estrutura secundária dependente do tempo, observam-se flutuações no conteúdo de suas folhas- β e da segunda hélice- α (Figura 22 B). Apesar disso, não há perda considerável de conteúdo estruturado e a proteína se mantém estável, como já observado em estudos anteriores (Viana, I. *et al.*, 2013; Cunha *et al.*, 2015).

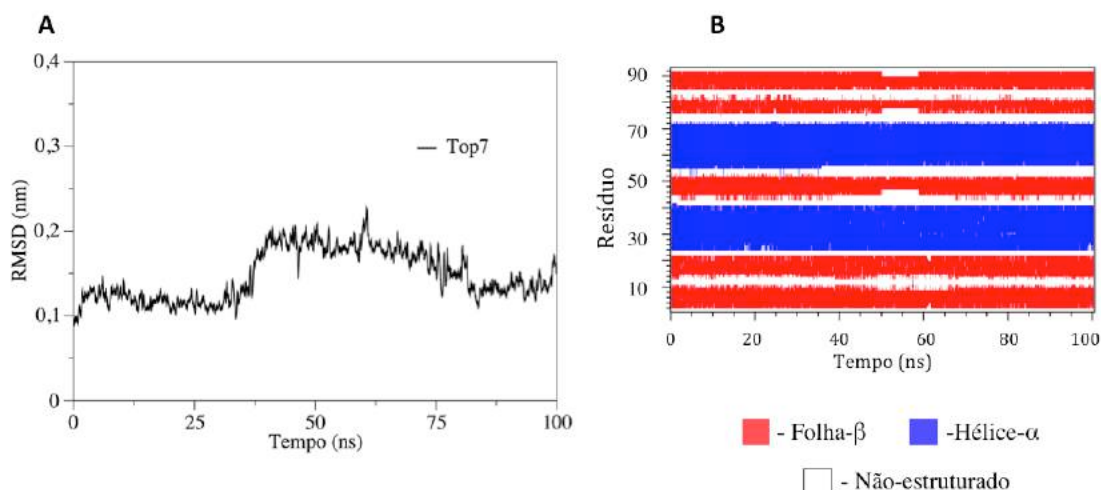


Figura 22 - (A) RMSD dos átomos do *backbone* da proteína Top7 em função do tempo, convergindo para um RMSD médio de 0,15 nm. (B) Conteúdo de estrutura secundária da Top7 em função do tempo de simulação, indicando que não há perda considerável de conteúdo estruturado e a proteína se mantém estável.

A proteína 2S20-30_Top7 apresenta um perfil de RMSD estável, porém se distancia da sua configuração inicial, convergindo para um valor em torno de 0,25 nm (Figura 23 A), quase o dobro do *scaffold* original. Analisando o diagrama de estrutura secundária ao longo do tempo (Figura 23 B) fica evidente que o motivo dos valores de RMSD mais altos é a perda parcial de sua primeira hélice- α . Essa é a região onde o epítipo foi alocado e a perda de estrutura não é uma surpresa, já que o mínimo global no funil de enovelamento para esta quimera ficou acima do *cut-off* estabelecido. Para avaliar se o epítipo mantém uma conformação próxima à nativa, o RMSD dos átomos do *backbone* do epítipo foi calculado em relação à estrutura nativa na NS1 (PDB 4OIG) (Figura 23 A, curva azul). A análise indica que o epítipo converge para uma configuração que se distancia de aproximadamente 0.4 nm em relação à estrutura nativa. O fato de perder conteúdo estrutural e um alto valor de RMSD entre o epítipo na quimera e sua estrutura nativa pode implicar numa baixa performance imunogênica.

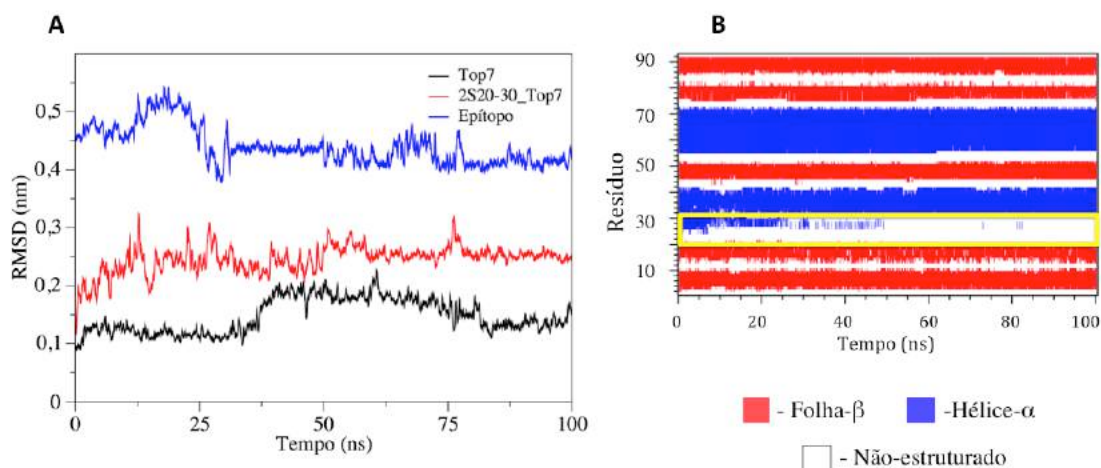


Figura 23 - (A) RMSD dos átomos do *backbone* da quimera 2S20-30_Top7 (vermelho) convergem para um valor de 0.25 nm, enquanto que o backbone do epítipo (azul) converge para 0.4 nm. **(B)** Conteúdo de estrutura secundária da quimera em função do tempo de simulação, indicando a perda parcial de sua primeira hélice- α , na região onde o epítipo foi alocado (destacado em amarelo).

Analisando a proteína construída através da inserção do epítipo 7, verifica-se que se trata de uma proteína bem estável estruturalmente (Figura 24 A). Os valores de RMSD para essa proteína convergem para aproximadamente 0,25 nm. Apesar dos valores médios de RMSD serem mais altos do que da Top7, estamos analisando uma proteína em que o epítipo inserido possui uma conformação de *loop* de alta flexibilidade. O gráfico de estrutura secundária da 7S22-24_Top7 (Figura 24 B) não exibe nenhuma perda de motivo estrutural além das flutuações já observadas na Top7. A região onde o epítipo 7 foi inserido se mantém estável durante toda a simulação, com RMSD de 0.35 nm em relação à sua estrutura nativa (Figura 24 A, curva azul).

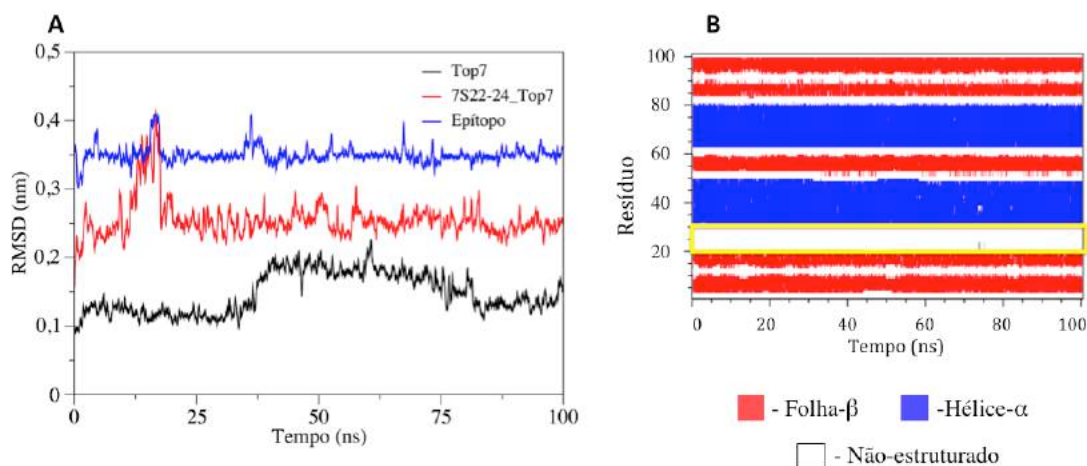


Figura 24 - (A) RMSD dos átomos do *backbone* da quimera 7S22-24_Top7 (vermelho) convergem para um valor de 0.25 nm, enquanto que o *backbone* do epítipo (azul) converge para 0.35 nm. (B) Conteúdo de estrutura secundária da quimera em função do tempo de simulação, não exibe nenhuma perda de motivo estrutural. Região do epítipo destacada em amarelo.

Analisando os dados de MD da quimera 8S22-15_Top7, observa-se que os valores de RMSD convergem para 0.28 nm (Figura 25 A). Da mesma forma que a 7S22-24_Top7, os valores de RMSD são mais altos por ser uma proteína mais extensa e o epítipo inserido estar na forma de *loop*. O gráfico de DSSP não indica perda de estrutura secundária do *scaffold*, porém a região central do epítipo 8 flutua entre uma hélice- α e um *loop* (Figura 25 B). A perda da estrutura nessa região não se torna um problema, pois o próprio epítipo é uma região não estruturada. Analisando o RMSD em relação ao epítipo nativo, observam-se valores acima de 0.4 nm, o que novamente pode prejudicar o reconhecimento do antígeno pelos anticorpos.

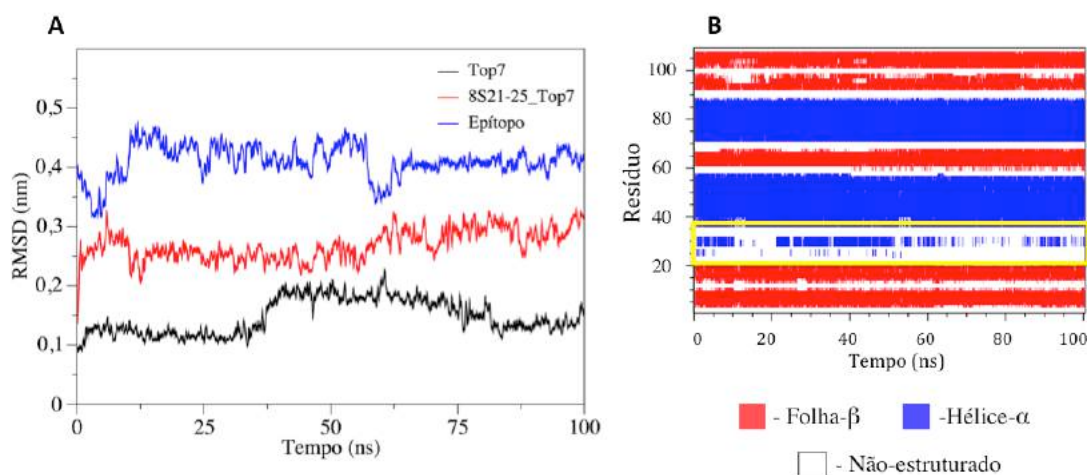


Figura 25 - (A) RMSD dos átomos do *backbone* da quimera 8S22-25_Top7 (vermelho) convergem para um valor de 0.28 nm, enquanto que o backbone do epítipo (azul) converge para 0.4 nm. **(B)** Conteúdo de estrutura secundária da quimera em função do tempo de simulação, não exibe nenhuma perda de motivo estrutural, exceto flutuações na região do epítipo destacada em amarelo.

Das mais de 1.100 quimeras analisadas pelo protocolo de *threading* + *de novo design* utilizando a Top7 como *scaffold*, três foram avaliadas por MD, uma apresentou perda de estrutura na região do *loop*, e duas se mostraram estáveis (Figura 26). Essas proteínas constituem candidatos para avaliação imunológica e serão discutidas na próxima seção.

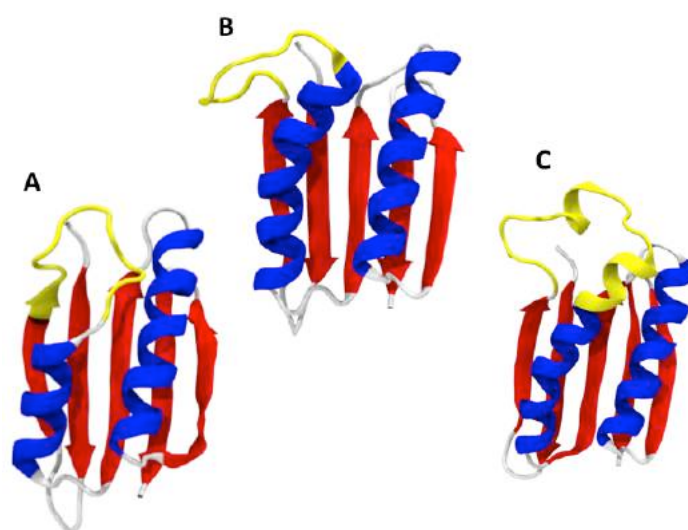


Figura 26 - Representação *cartoon* das quimeras (A) 2S20-30_Top7, (B) 7S22-24_Top7, e (C) 8S22-25_Top7, após 100 ns de simulação. Folhas- β estão representadas como setas na cor vermelha, hélices- α como espirais na cor

azul e regiões não estruturadas na cor branca. As regiões correspondentes aos epítomos estão destacadas em amarelo.

3.6.2.3 Expressão bacteriana e avaliação imunológica por ELISA

As sequências de DNA das proteínas Top7, 2S20-30_Top7, 7S22-24_Top7 e 8S22-25_Top7 otimizadas para expressão em *E. coli* foram subclonadas no vetor de expressão pRSET A. As proteínas foram expressas em sistema procarioto e purificadas por cromatografia de afinidade (Figura 27). A Top7 original é uma proteína globular e teve alto rendimento de expressão, sendo obtida a concentrações de 1 g/L. A químera 7S22-24_Top7 também aparenta ser globular, também com rendimento de 1 g/L. Já as proteínas 2S20-30_Top7 e 8S22-25_Top7 obtiveram taxas de rendimento mais baixas, com concentrações abaixo de 1 mg/L. No caso da construção 2S20-30_Top7 o baixo rendimento pode ser explicado pela perda de estrutura secundária observada durante. A exposição do núcleo hidrofóbico pode acarretar numa diminuição da solubilidade da proteína.

Devido ao fenômeno de competição por afinidade na coluna, pode-se observar que à medida que a concentração da proteína quimérica diminuía, aumentava a presença de outras bandas nas eluições.

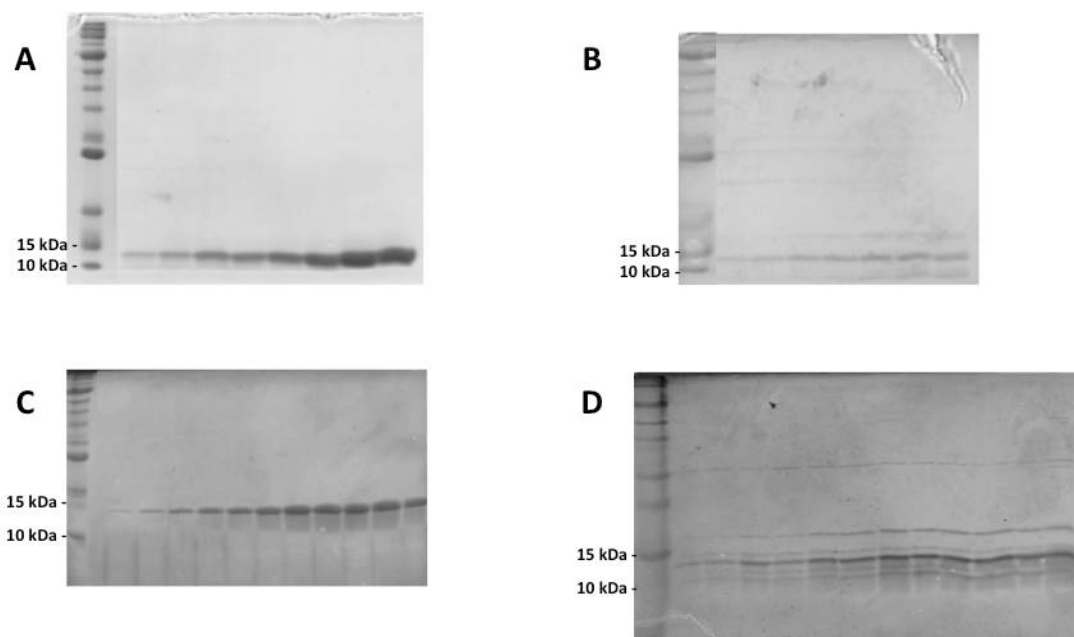


Figura 27 - SDS-PAGE das quimeras obtidas por expressão bacteriana e purificadas por cromatografia de afinidade. (A) Top7 original, obtida com alto rendimento. (B) Quimera 2S20-30_Top7 obtida com baixas taxas de rendimento, visto que durante a MD observou-se perda de estrutura nativa. (C) A 7S22-24_Top7 também foi obtida em grande quantidade, indicando ser uma proteína bastante solúvel. (D) 8S22-25_Top7 com rendimento moderado. Bandas de proteínas provenientes de bactéria são observadas nas eluições das proteínas de baixo e médio rendimento.

Após purificação e quantificação, a imunorreatividade das proteínas foi avaliada por ensaio de ELISA indireto. Foram utilizados 16 soros de pacientes positivos (4 para DENV-1; 4 para DENV-2, 3 para DENV-3; 1 para DENV-4 e 4 para DENV-3/4), coletados na fase convalescente e testados positivos para IgG total. Foram também utilizados 4 soros de pacientes negativos para DENV. Os ensaios foram realizados em duplicatas. Os sinais de OD relacionados à proteína Top7 original podem ser considerados como ruído de ligação inespecífica (Viana, I. et al., 2013), e foram utilizados para fins de comparação. Curvas ROC (*Receiver Operating Characteristic*) foram obtidas para cada ensaio a fim de avaliar o desempenho do potencial de diagnóstico de cada antígeno. Cada curva foi gerada plotando-se a taxa de verdadeiro positivo (sensibilidade) contra a taxa de falso-positivo (especificidade). As linhas pontilhadas nos gráficos de ELISA indicam o *cut-off* abaixo do qual uma amostra é considerada negativa para o antígeno. Os valores dos *cut-offs* foram determinados através das curvas ROC para cada

ensaio, de modo a ter o maior valor do par especificidade/sensibilidade. Adicionalmente, teste-t de Student não pareados foram realizados para determinar a significância estatística ($p < 0.05$) da diferença entre as medianas dos grupos DENV + e DENV -, e testes pareados foram realizados entre os grupos DENV +/-Quimera e DENV +/-Top7 para determinar a significância estatística ($p < 0.05$) da diferença entre o sinal inespecífico da proteína carreadora.

O ensaio para a proteína 2S20-30_Top7 foi realizado com concentração de 8 pmol/poço do antígeno e diluição de 1:10000 do anticorpo secundário. Apesar da proteína aparentemente ter uma boa performance imunológica, exibindo uma distinção clara e significativa entre as populações positivas e negativas para DENV ($p < 0.05$), não se observa essa diferença quando compara-se os sinais da 2S20-30_Top7 com os da Top7 original ($p = 0.2319$) (Figura 28). A performance imunológica comparável ao do *scaffold* pode ser explicada pela perda de estrutura secundária prevista para o antígeno 2S20-30_Top7 nas simulações de MD, o que pode levar a várias novas ligações inespecíficas (Viana, I. *et al.*, 2013).

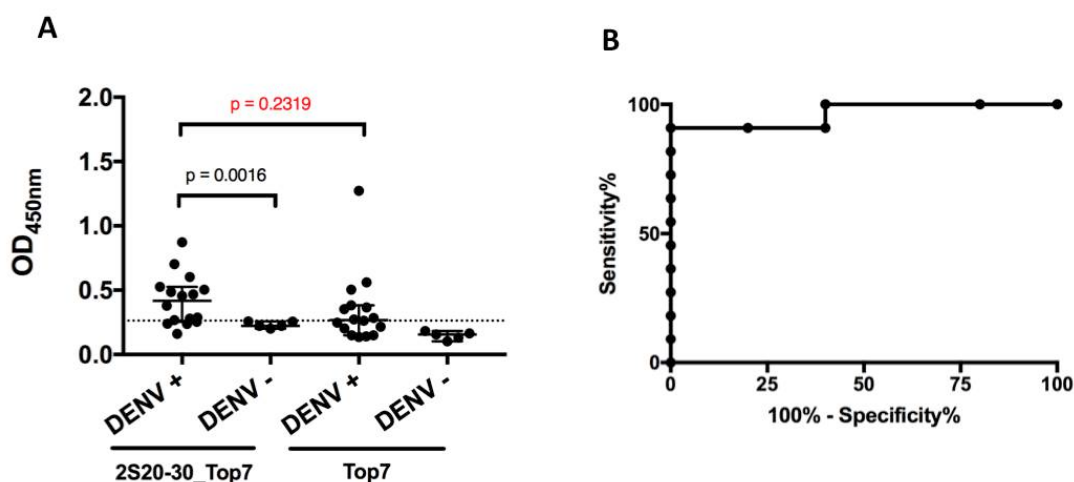


Figura 28 - (A) Ensaio de ELISA do antígeno 2S20-30_Top7, exibindo uma distinção clara e significativa entre as populações positivas e negativas para DENV ($p < 0.05$), mas não se observa essa diferença quando compara-se os sinais da 2S20-30_Top7 com os da Top7 original. (B) Curva ROC do ensaio, com AUC = 0.96.

O ensaio para a proteína 7S22-24_Top7 foi realizado com concentração de 16 pmol/poço do antígeno e diluição de 1:20000 do anticorpo secundário. A proteína quimérica demonstrou bom desempenho diagnóstico, exibindo distinção significativa entre as populações positivas e negativas para DENV, e em comparação aos sinais da Top7 ($p < 0.05$) (Figura 29 A). O resultado corrobora as previsões computacionais, estabelecendo uma relação direta entre a manutenção da estrutura geral da proteína e a manutenção das propriedades imunológicas (Viana, I. *et al.*, 2013). O ensaio obteve sensibilidade de 71.43 % e especificidade de 100% (Figura 29 B). A baixa sensibilidade do ensaio pode ser explicada pelo fato da proteína quimérica ter sido construída utilizando um epítopo que reage fortemente com anticorpos provenientes de DENV-1 e 2, enquanto que reage moderadamente com anticorpos de DENV-3 e 4 (Figura 30) (Omokoko *et al.*, 2014). Além disso, nem todos os indivíduos devem ter anticorpos específicos para a região da NS1 utilizada.

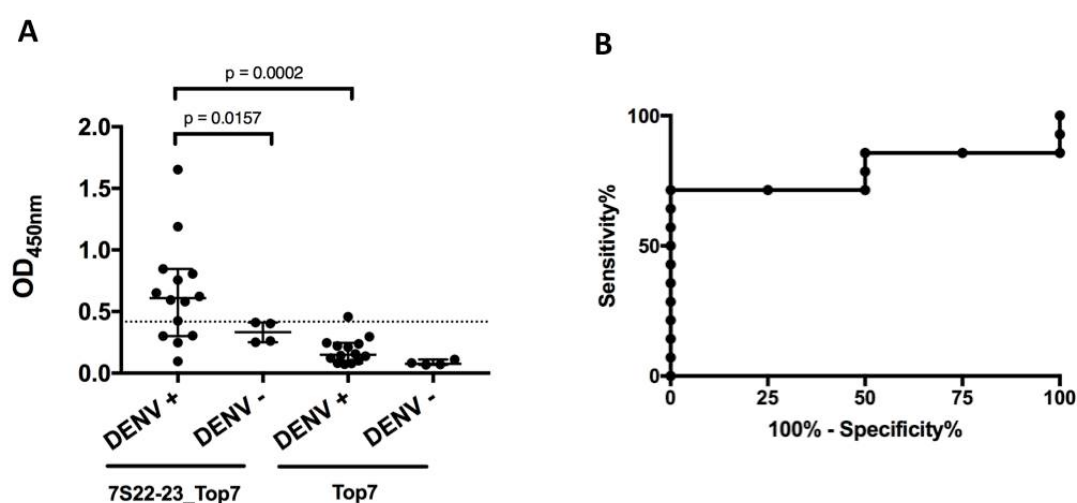


Figura 29 - (A) Ensaio de ELISA do antígeno 7S22-24_Top7, com bom desempenho diagnóstico, exibindo distinção significativa entre as populações positivas e negativas para DENV, e em comparação aos sinais da Top7 ($p < 0.05$). (B) Curva ROC do ensaio com a proteína 7S22-24_Top7, indicando sensibilidade de 71.43 % e especificidade de 100%, e AUC = 0.7857.

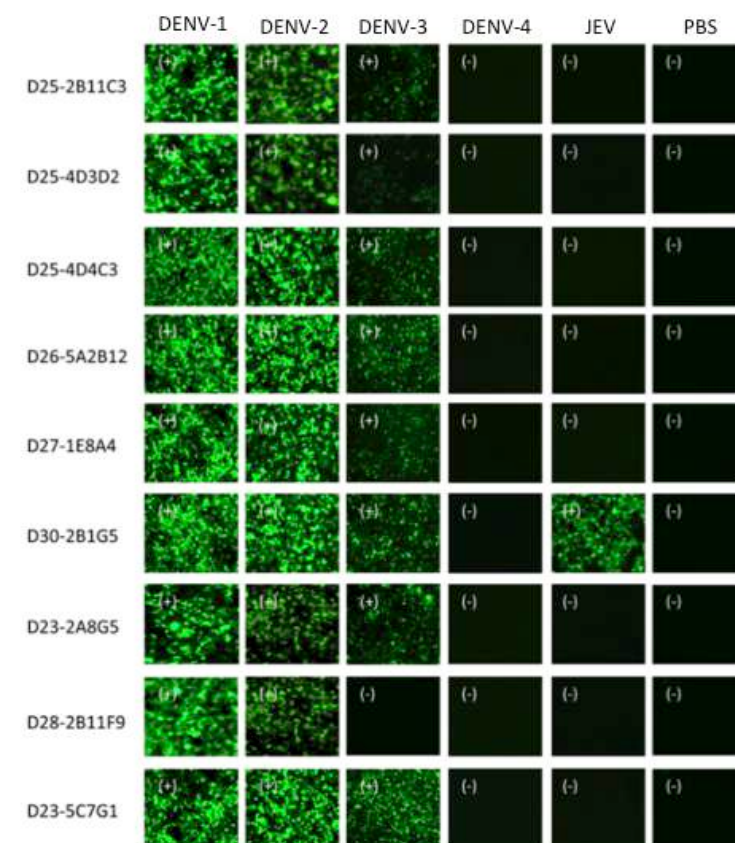


Figura 30 - Especificidade dos sorotipos de anticorpos humanos anti-NS1 obtidos de pacientes com infecção secundária de DENV-2. Células Vero foram infectadas com DENV-1, DENV-2, DENV-3, DENV-4 ou JEV (vírus encefalite japonesa). A reatividade de cada um dos mAbs com células infectadas com cada vírus foi então examinada por microscopia de fluorescência. Adaptado de (Omokoko *et al.*, 2014).

O ensaio para a proteína 8S22-25_Top7 foi realizado com concentração de 8 pmol/poço do antígeno e diluição de 1:20.000 do anticorpo secundário. O epítipo 8 utilizado na quimera 8S22-25_Top7 possui os mesmos aminoácidos do epítipo 7, diferenciando deste por ter alguns aminoácidos a mais (Tabela 3). Dessa forma, espera-se um comportamento semelhante à proteína 7S22-24_Top7. De fato, a proteína 8S22-25_Top7 também demonstrou bom desempenho diagnóstico, exibindo distinção significativa entre as populações positivas e negativas para DENV, e em comparação aos sinais da Top7 ($p < 0.05$) (Figura 31 A). O ensaio obteve sensibilidade de 69.23 % e especificidade de 100% (Figura 31 B). Novamente, a baixa sensibilidade do ensaio pode ser explicada pelo fato da proteína quimérica ter sido construída utilizando um epítipo que reage

fortemente com anticorpos provenientes de DENV-1 e 2, enquanto que reage moderadamente com anticorpos de DENV-3 e 4.

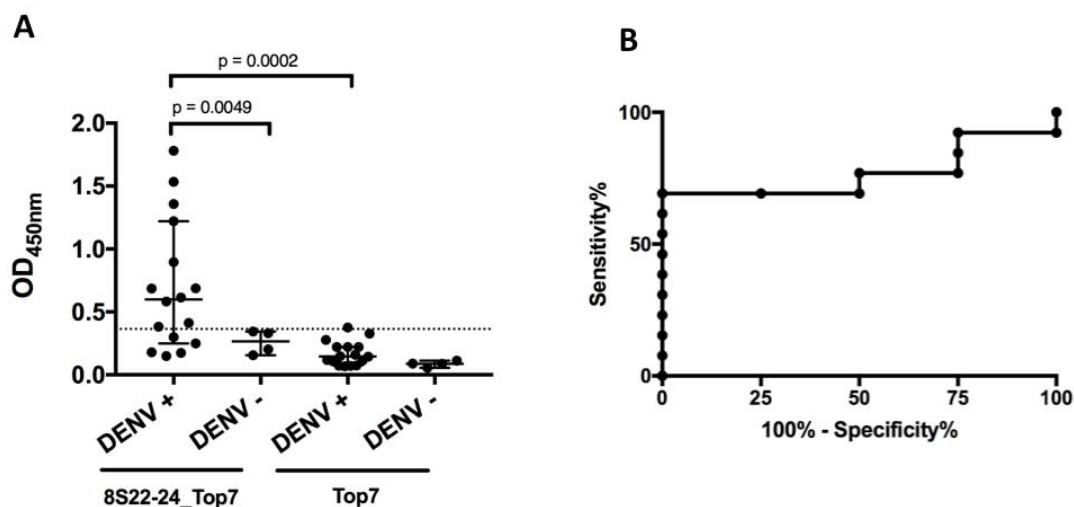


Figura 31 - (A) Ensaio de ELISA do antígeno 8S22-25_Top7 exibindo distinção significativa entre as populações positivas e negativas para DENV, e em comparação aos sinais da Top7 ($p < 0.05$) (Figura 3.16 A). **(B)** O ensaio obteve sensibilidade de 69.23 % e especificidade de 100% e AUC = 0.7692.

3.7 CONCLUSÕES

A metodologia aplicada nesse trabalho, o *threading* seguido do *de novo design*, mostrou ser eficaz na construção de proteínas quiméricas contendo epítomos, neste caso da proteína NS1 de dengue. Com esse conjunto de técnicas é possível avaliar milhares de proteínas em curto espaço de tempo. Apenas as construções viáveis energeticamente são então avaliadas por métodos mais dispendiosos, como a dinâmica molecular. Dessa forma, evita-se o gasto desnecessário de alocação computacional com estruturas não viáveis. Por consequência, utilizando esse método de triagem, evita-se o uso dispendioso de insumos de laboratório para testes cegos de proteínas possivelmente imunorreativas sem nenhuma informação prévia.

Foram analisadas 3 quimeras construídas utilizando a Top7 como *scaffold*, das quais duas (7S22-24_Top7 e 8S22-25_Top7) foram consideradas estáveis e conseguem manter tanto a sua conformação global

como a conformação do epítopo, enquanto que a 2S20-30_Top7 demonstrou perda de estrutura secundária. Apesar disso, todas as três proteínas foram expressas em sistema procarioto e purificadas por cromatografia de afinidade. Como esperado, a 2S20-30_Top7 foi obtida com baixo rendimento, assim como a 8S22-25_Top7. Já a 7S22-23_top7 se mostrou uma proteína globular e com alto rendimento de expressão.

Os ensaios de ELISA para avaliação imunológica das proteínas quiméricas sugerem que não há diferença significativa na imunorreatividade da 2S20-30_Top7 em comparação com a Top7. Esse resultado corrobora com a hipótese de Viana e colaboradores de que para que a proteína seja imunorreativa, a inserção do epítopo não deve afetar a estrutura secundária da proteína carreadora (Viana, I. *et al.*, 2013). Os demais antígenos desenhados demonstraram bom desempenho diagnóstico, exibindo distinção significativa entre as populações positivas e negativas para DENV, e em comparação aos sinais da Top7 ($p < 0.05$). As sensibilidades abaixo de 100% dos ensaios podem ser explicadas pelo fato das proteínas quiméricas terem sido construídas utilizando epítopos que reagem fortemente com anticorpos provenientes de DENV-1 e 2, enquanto que reagem moderadamente com anticorpos de DENV-3 e 4 (Omokoko *et al.*, 2014). Além disso, nem todos os indivíduos devem ter anticorpos específicos para a região da NS1 utilizada.

Dessa forma, pode-se concluir que utilizando o conjunto de técnicas de engenharia de proteínas (*threading* + *de novo design*) em conjunto com análises de dinâmica molecular, foi possível desenhar *in silico* três proteínas expondo epítopos de NS1 de DENV, cujas imunorreatividades estão de acordo com dados da literatura.

4 AVALIAÇÃO ESTRUTURAL DA TOP7-2F5 BIOTINILADA PARA DETECÇÃO DE ANTICORPOS ANTI-2F5 DE HIV-1

Anticorpos são os biomarcadores mais comuns utilizados para detectar e quantificar ligação e reconhecimento de antígenos específicos (Boschek *et al.*, 2009; Viana, I. *et al.*, 2013). Em infecções naturais, a produção de anticorpos é tipicamente induzida por proteínas imunogênicas. As estratégias terapêuticas baseiam-se frequentemente na utilização de antígenos, por exemplo, partículas virais, vírus atenuados, domínios protéicos, para imitar a resposta imune do hospedeiro após a exposição ao patógeno e induzir a produção de anticorpos protetores.

No entanto, abordagens convencionais têm falhado para o vírus da imunodeficiência humana (HIV). O vírus do HIV desenvolveu um arsenal de truques moleculares para evitar ou atenuar as respostas imunes (Burton, 2010), incluindo a exposição transiente de epítomos cruciais capazes de elicitar a produção de anticorpos neutralizantes. Um desses epítomos críticos é conhecido como 2F5. O 2F5 consiste numa sequência contínua e conservada de 9 aminoácidos (ELDKWASLW), que se encontra no ectodomínio conhecido como região externa proximal à membrana (MPER) da proteína gp41 de envelope do vírus HIV (Ofek *et al.*, 2004). Entretanto, os anticorpos elicitados contra esse epítopo, conhecidos como anti-2F5, são produzidos por apenas uma minoria de indivíduos infectados após 2-3 anos de exposição ao vírus (Kwong e Mascola, 2012). O anticorpo anti-2F5 é um dos anticorpos anti-HIV mais poderosos e amplamente protetores descritos até agora, podendo neutralizar mais de 50% dos painéis de vírus isolados e proteger modelos de primatas após o desafio (Muster *et al.*, 1993; Hessel *et al.*, 2010). Apesar de ser um alvo importante para produção de vacinas, todas as tentativas profiláticas baseadas na exposição do epítopo 2F5 não conseguiram obter imunidade protetora a longo prazo (Bomsel *et al.*, 2011; Serrano *et al.*, 2014; Safrit *et al.*, 2016), refletindo a dificuldade de mimetizar a estrutura nativa do epítopo.

Para superar esta limitação, Viana e colaboradores desenvolveram previamente uma proteína quimérica baseada na Top7 contendo o epítopo 2F5 (Viana, I. *et al.*, 2013). A nova proteína, denominada Top7-2F5 (Figura

32 A), demonstrou ser especificamente reconhecida pelo respectivo anticorpo monoclonal (mAb 2F5) por meio de ensaios de ELISA (Viana, I. *et al.*, 2013). Como o mAb 2F5 é um anticorpo neutralizante, a proteína quimérica detém o potencial de ser usada como um antígeno vacinal baseado na estrutura nativa do MPER do HIV-1.

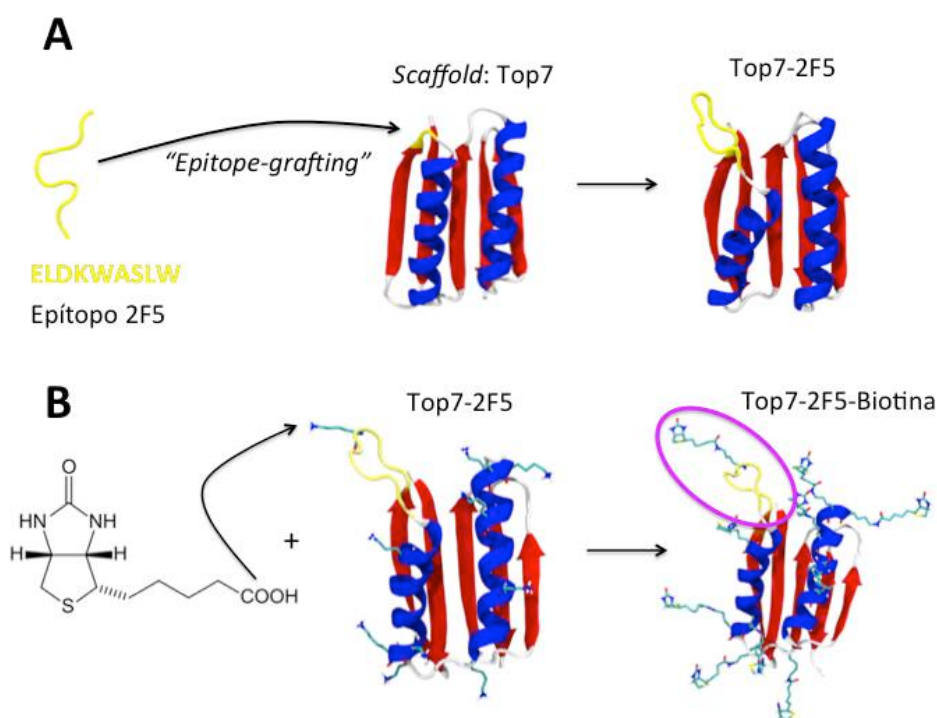


Figura 32 - (A) Proteína quimérica baseada na Top7 contendo o epítipo 2F5 construída por Viana e colaboradores (Viana, I. *et al.*, 2013). (B) Representação da reação da Top7-2F5 com a molécula de biotina. A parte linear terminando no grupo carboxila é responsável pela reação com aminas primárias presentes em cadeias de proteínas.

O principal requisito para uma vacina protetora contra o HIV é que o antígeno projetado seja capaz de ativar as células B virgens nos centros germinativos e/ou estimular as células B de memória para a evolução e maturação dos anticorpos neutralizantes. Neste caso, para avaliar se uma vacina focada no epítipo 2F5 consegue ativar células B, é necessário o desenvolvimento prévio e teste de uma plataforma para detectar e quantificar células B que expressem anticorpos anti-2F5. Com esse objetivo, a Top7-2F5 foi covalentemente ligada a moléculas de biotina (Figura 32 B) para servir de

biomarcador e detectar a ligação específica ao receptor de células B por sinal de fluorescência através de citometria de fluxo.

Embora seja sabido que o processo de biotinilação afete minimamente as propriedades físico-químicas de uma molécula, isso não é necessariamente sempre verdade, de modo que uma biomolécula marcada com biotina deve sempre ser caracterizada antes de seu uso propriamente dito (Bayer e Wilchek, 1990). A fim de verificar se a Top7-2F5 biotinilada (Top7-2F5-Biotina) mantém sua imunorreatividade com o mAb 2F5, a reação da proteína biotinilada com o anticorpo monoclonal foi previamente avaliada por meio de ELISA (Figura 33, dados obtidos pela colaboradora Dr^a Isabelle Viana) (Coelho, In preparation).

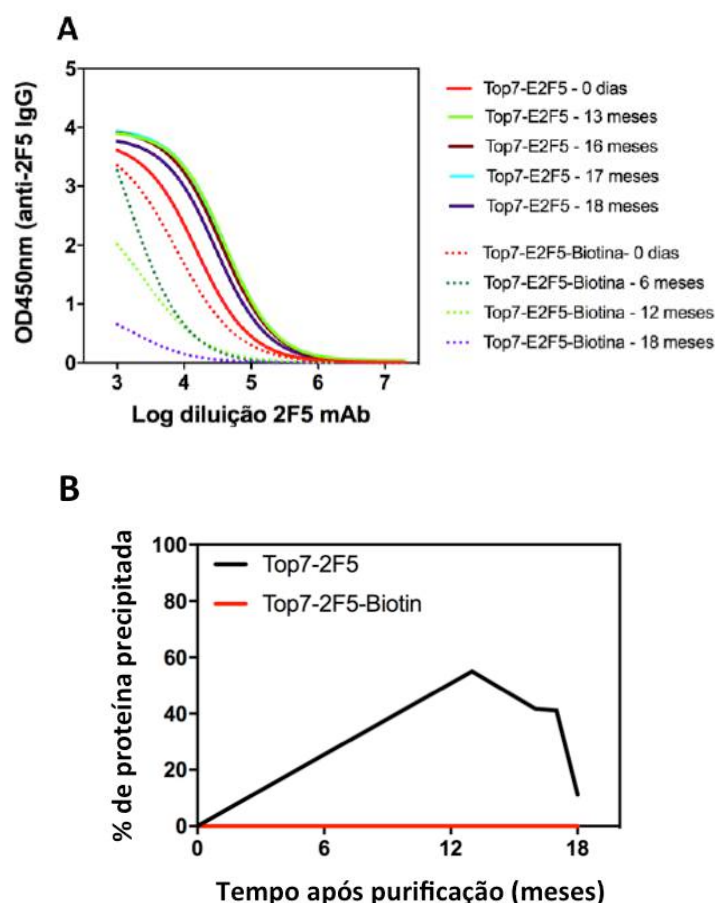


Figura 33 - (A) Imunorreatividade das proteínas Top7-2F5 e Top7-2F5-Biotina foram avaliadas em diferentes períodos de tempo. Ambas as proteínas possuem imunorreatividade semelhantes quando utilizadas assim que purificadas, com uma leitura ligeiramente inferior para Top7-E2F5-Biotina. **(B)** Concentração das proteínas Top7-2F5 e Top7-2F5-Biotina em diferentes períodos de tempo. Dados obtidos pela colaboradora Dr^a Isabelle Viana (Coelho, In preparation).

Os dados indicam que ambas as proteínas Top7-2F5 e Top7-2F5-Biotina possuem imunorreatividade semelhante quando utilizadas assim que purificadas. Porém, a leitura da Top7-2F5-Biotina é inferior e diminui significativamente em função do tempo de armazenamento, enquanto permanece inalterada para a proteína não biotinilada até 18 meses. Outra curiosidade observada foi que enquanto a proteína não-biotinilada se agrega e precipita ao longo do tempo, a proteína biotinilada não mostrou agregação em função do tempo (dados obtidos pela colaboradora Dr^a Isabelle Viana, Figura 33 B) (Coelho, In preparation).

Como a reação de biotinilação ocorre em amins primárias, isto é ainda mais preocupante no caso do epítipo 2F5, pois um resíduo de lisina se encontra no meio do epítipo. Portanto, os efeitos na estabilidade e na imunorreatividade da proteína Top7-2F5 biotinilada devem ser avaliados. A fim de verificar as consequências da biotinilação da Top7-2F5 para uso como biomarcador de células B que expressem anticorpo anti-2F5, neste trabalho foram investigadas as bases moleculares para estrutura e propriedades da proteína Top7-2F5 biotinilada através de técnicas computacional e experimental.

4.1 OBJETIVOS

Caracterizar a estabilidade estrutural da proteína Top7-2F5 biotinilada e as consequências da biotinilação na imunorreatividade contra o mAb 2F5 por meio de técnicas experimentais e computacionais. O *scaffold* nativo Top7 será usado para fins de comparação.

4.1.1 Objetivos específicos

- Determinar estabilidade estrutural da Top7-2F5 e sua versão biotinilada, Top7-2F5-Biotina, juntamente com a estrutura nativa Top7, assim como sua versão biotinilada, Top7-Biotina, por espectroscopia de dicroísmo circular (CD);

- Analisar a estabilidade estrutural e propriedades das proteínas Top7-2F5, Top7-2F5-Biotina, Top7 e Top7-Biotina por meio de simulações de dinâmica molecular;
- Avaliar a interferência da biotinilação no resíduo LYS do epítipo 2F5 e no reconhecimento pelo anticorpo mAb 2F5.

4.2 METODOLOGIA

4.2.1 Dinâmica molecular

As coordenadas iniciais para a proteína Top7 nativa foram obtidas do PDB 1QYS. Para a construção do Top7-2F5, os resíduos T22 e T23 do Top7 foram removidos e a sequência do epítipo 2F5 em uma conformação de *loop* foi adicionada ao *scaffold* (Viana, I. et al., 2013). A modelagem do sistema foi realizada com o software Swiss-PDB Viewer 4.0.1 (Guex e Peitsch, 1997).

A topologia molecular da biotina foi construída baseada nos parâmetros recuperados do campo de força GROMOS53A6 (Oostenbrink *et al.*, 2004). Uma simulação MD curta de 15 ns foi realizada e a geometria molecular dos anéis da biotina foi observada na conformação de barco, de acordo com dados experimentais de difração de raios-X (Chivers *et al.*, 2011). A parametrização foi realizada pelo integrante do nosso grupo Matheus Ferraz.

As versões biotiniladas da Top7 e Top7-2F5 foram construídas ligando todas as aminas primárias disponíveis (resíduos LYS e N-terminal) com o grupo carboxílico da biotina. Todas as estruturas foram centradas em uma caixa cúbica de 10 x 10 x 10 nm e explicitamente solvatadas usando o modelo de água SPC (Berendsen *et al.*, 1981). Condições periódicas de contorno foram usadas nas direções x, y e z. O método LINCS foi usado para restringir todas as ligações envolvendo átomos de hidrogênio (Hess *et al.*, 1997). As temperaturas das proteínas e do restante do sistema foram acopladas separadamente a um termostato utilizando o esquema *velocity rescale* (Bussi *et al.*, 2007) em três diferentes temperaturas de referência:

300 K, 340 K e 380 K e constante de acoplamento de 0.2 ps. A pressão foi mantida constante em 1 bar de forma isotrópica utilizando um barostato de Berendsen (Berendsen *et al.*, 1984) com constante de acoplamento de 0.2 ps e compressibilidade (κ_T) de $4.5 \times 10^{-5} \text{ bar}^{-1}$. Um raio de corte de 1.4 nm foi utilizado para interações não-ligadas e interações de longo alcance foram contabilizadas utilizando o método *reaction-field* (Tironi *et al.*, 1995) com $\epsilon=66$ além do raio de corte. Contra-íons foram adicionados para garantir uma carga neutra do sistema e uma concentração de NaCl de 0,150 mM.

Os sistemas foram inicialmente minimizados usando 5.000 passos do algoritmo *steepest descent*. A integração foi realizada usando tempo de integração de 2 fs baseada no algoritmo *leapfrog* (Hockney, 1970). O equilíbrio térmico de todos os sistemas foi atingido usando *ensemble* NVT na respectiva temperatura por 20 ps, seguida por um equilíbrio de pressão em *ensemble* NPT por 1 ns. Todas as etapas de equilíbrio foram realizadas com restrições de posição para todos os átomos. Simulações MD foram realizadas usando o *ensemble* NPT com tempo total de 100 ns usando o campo de força GROMOS 53A6 (Oostenbrink *et al.*, 2004) com aceleração em GPU e esquema de raio de corte Verlet (Páll e Hess, 2013).

Todas as simulações foram realizadas e analisadas usando o pacote GROMACS 4.6.5 (Hess *et al.*, 2008), DSSP (Database of Secondary Structure in Proteins) (Kabsch e Sander, 1983) e análises visuais usando o programa Visual Molecular Dynamics (VMD) 1.8.7. Estruturas representativas das proteínas obtidas a partir dos últimos 50 ns de cada trajetória e empregando a ferramenta *g_cluster* do conjunto GROMACS, alinhando os átomos do *backbone* com um raio de corte de 0,14 nm. Cada *cluster* conformacional único foi usado para realizar análises de superfície potencial eletrostática usando o software MEAD v2.2 (Bashford e Gerwert, 1992).

4.2.2 Expressão das proteínas e biotinylação

As sequências de DNA que codificam as proteínas Top7 e Top7-2F5 foram inseridas entre dois sítios de clonagem múltiplos (5' Hind III e Nhe I e 3' NcoI e XhoI) com cauda de histidina N-terminal e otimizadas para tradução e

expressão em sistema bacteriano. Os genes otimizados foram sintetizados por um fornecedor comercial (GeneScript®) e inseridos no vetor de expressão pRSET A (Invitrogen®).

A expressão bacteriana das proteínas foi realizada com a linhagem *Escherichia coli* BL21 *Star* (DE3) pLysS (Invitrogen®) contendo os vectores de expressão. As células foram cultivadas em meio LB suplementado com 100 mg / mL de ampicilina (LB-AMP) para uma OD600 de 0.5 e a expressão proteica foi induzida com 1 mM de IPTG a 37 °C, 4 h e 225 RPM. Em seguida, os *pellets* bacterianos foram recolhidos por centrifugação a 4 °C, 30 min, 5000 xg e ressuspensos em tampão de lise. As células foram lisadas por sonicação em gelo, com amplitude ajustada em 40%, utilizando 6 pulsos de 30 s em intervalos de 60 s. O lisado foi clarificado por centrifugação e as proteínas recombinantes foram purificadas por cromatografia de afinidade com coluna de matriz de níquel (Ni-NTA - QIAGEN).

Antes da biotinylation, as proteínas recombinantes foram dialisadas, para remover quaisquer vestígios de amins primárias que pudessem diminuir a eficiência da reação de biotinylation. O procedimento de biotinylation foi realizado utilizando o kit Thermo Scientific™ Pierce™ EZ-Link Sulfo-NHS-Biotina, seguindo as instruções do fabricante. Resumidamente, preparou-se uma solução de 10 mM de biotina e adicionou-a a cada solução proteica, sempre mantendo a biotina em excesso molar de 20 vezes. As reações de biotinylation ocorreram por 2 h no gelo e as proteínas marcadas foram purificadas por diálise para um tampão contendo NaCl 300 mM, Tris-HCl 50 mM, pH 8.0.

As etapas de expressão e purificação proteica, assim como de biotinylation, foram realizadas em colaboração com a Dr^a Isabelle Viana, à época trabalhando como *pós-doc* na Universidade de Pittsburgh.

4.2.3 Dicroísmo circular

As proteínas foram diluídas em tampão A, contendo 300 mM NaCl, 50 mM Tris-HCl, pH 8.0 buffer (Buffer A) a uma concentração de 100 µg/mL e dialisadas por 12h a 4°C contra 5 tampões contendo diferentes

concentrações decrescentes de NaCl e Tris-HCl: tampão B - 225 mM NaCl, 50 mM Tris-HCl, pH 8.0; tampão C - 150 mM NaCl, 40 mM Tris-HCl, pH 8.0; tampão D - 75 mM NaCl, 30 mM Tris-HCl, pH 8.0; tampão E - 45 mM NaCl, 25 mM Tris-HCl, pH 8.0; tampão F - 30 mM NaCl, 20 mM Tris-HCl, pH 8.0. As amostras dialisadas foram concentradas usando colunas Vivaspin, 10,000 MWCO (Sartorius Stedim Biotech) e quantificadas por espectrofotometria. A concentração final de cada proteína foi de 10 μ M.

Os espectros foram obtidos de 260 a 200 nm, com tempos médios de 5 segundos, tamanho do passo de 1 nm e largura de banda de 2 nm a 25 ° C. Os espectros foram corrigidos em relação ao branco e a elipticidade molar em relação à linha de base a 260 nm. Os curvas foram suavizadas pelo método de Stavitsky-Golay. Os dados de dicroísmo circular foram obtidos num espectrômetro Olis DSM17.

4.3 RESULTADOS E DISCUSSÃO

4.3.1 Estabilidade estrutural e dinâmica conformacional

A estabilidade estrutural da Top7-2F5 e sua versão biotinilada, Top7-2F5-Biotina, juntamente com o *scaffold* nativo Top7 e sua versão biotinilada, Top7-Biotina, foi avaliada por espectroscopia de dicroísmo circular (CD) à temperatura ambiente. Os espectros de CD foram coletados de 200 a 260 nm. Porém, há uma banda de absorção em 190 nm de folhas- β e, portanto, os dados permitem apenas uma comparação qualitativa do conteúdo de estrutura secundária. Dessa forma, os espectros foram utilizados para avaliar a similaridade de conteúdo de estrutura secundária e do enovelamento entre as proteínas nativas e biotiniladas.

Todas as quatro proteínas exibiram um espectro de CD característico de proteínas α / β (Figura 34), em acordo com os espectros de CD relatados anteriormente para Top7 e Top7-2F5 não-biotiniladas (Viana, Isabelle *et al.*, 2013). Os quatro espectros revelam duas bandas negativas características do conteúdo de hélice α em 208 nm e 222 nm e uma banda negativa correspondente ao conteúdo em folha β entre 216 e 218 nm. A comparação

entre os espectros da Top7 e da Top7-Biotina mostra que o conteúdo de estrutura secundária de ambos os sistemas é comparável em magnitude (Figura 34 A). Da mesma forma, nenhuma mudança significativa de elipticidade em 208, 222 e 216-218 nm foi observada quando se comparam os sistemas Top7-2F5 e Top7-2F5-Biotina (Figura 34 B). Os dados sugerem uma notável semelhança no *folding* das quatro proteínas analisadas, indicando que a biotinilação não interfere estruturalmente no enovelamento.

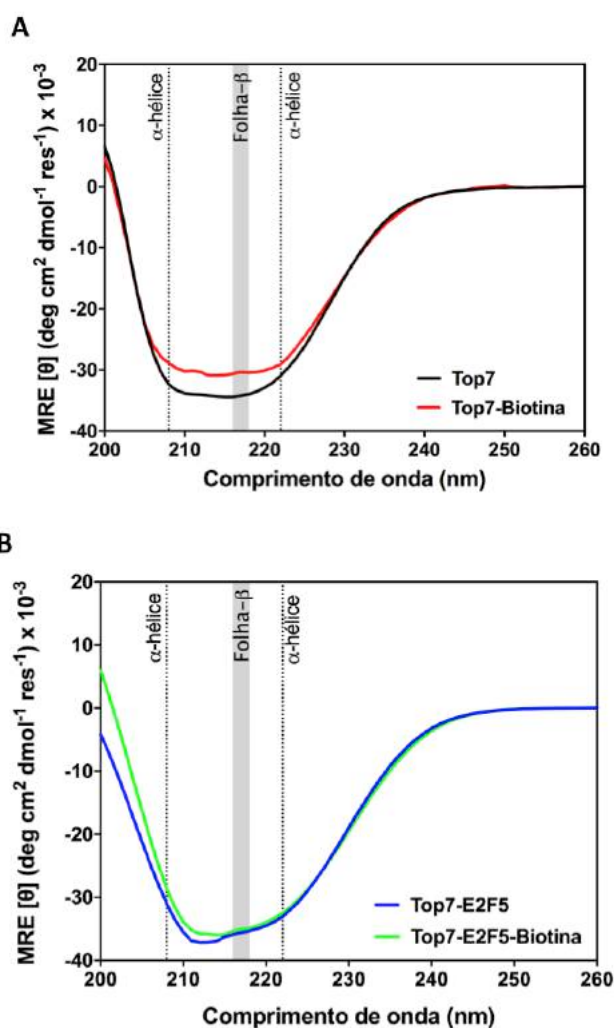


Figura 34 - Espectro de CD das proteínas (A) Top7 e Top7-Biotina, e (B) Top7-2F5 e Top7-2F5-Biotina. Todos os quatro espectros são característicos de proteínas α / β . Os espectros revelam duas bandas negativas características do conteúdo de hélice α em 208 nm e 222 nm e uma banda negativa correspondente ao conteúdo em folha β entre 216 e 218 nm. A comparação entre os espectros das proteínas Top7 e Top7-2F5 com suas versões biotiniladas sugerem semelhanças no *folding* e manutenção da estrutura das proteínas.

Em seguida, a base molecular da estabilidade estrutural das proteínas após biotinylation foi computacionalmente investigada por simulações de MD de 100 ns em três diferentes temperaturas: 300 K, 340 K e 380 K. O desvio quadrático médio (RMSD) para a posição de todos os átomos do *backbone* foi calculado nos últimos 50 ns de cada trajetória (Figura 35), e os resultados estão de acordo com estudos anteriores (Viana, I. *et al.*, 2013; Cunha *et al.*, 2015). Os dados mostram que as proteínas Top7 e Top-2F5 mantiveram sua estrutura nativa em 300 K, com um RMSD médio de $0,19 \pm 0,04$ nm e $0,24 \pm 0,02$ nm, respectivamente (Figura 35, losangos preto e azul, respectivamente). À medida que a temperatura da simulação aumenta, o RMSD médio assume valores maiores: $0,35 \pm 0,04$ nm e $0,47 \pm 0,07$ nm a 340 K e 380 K, respectivamente, para Top7 (Figura 35, gráfico acima, losangos pretos) e $0,47 \pm 0,04$ nm e $0,50 \pm 0,06$ nm a 340 K e 380 K, respectivamente, para Top7-2F5 (Figura 35, gráfico abaixo, losangos azuis). Isto é esperado devido aos efeitos da energia térmica acarretando em desenovelamento (Day et al., 2002). Portanto, os valores mais elevados de RMSD sugerem que as proteínas não-biotiniladas estudadas sofrem mudanças conformacionais substanciais e não mantêm o *folding* da estrutura nativa a temperaturas mais altas.

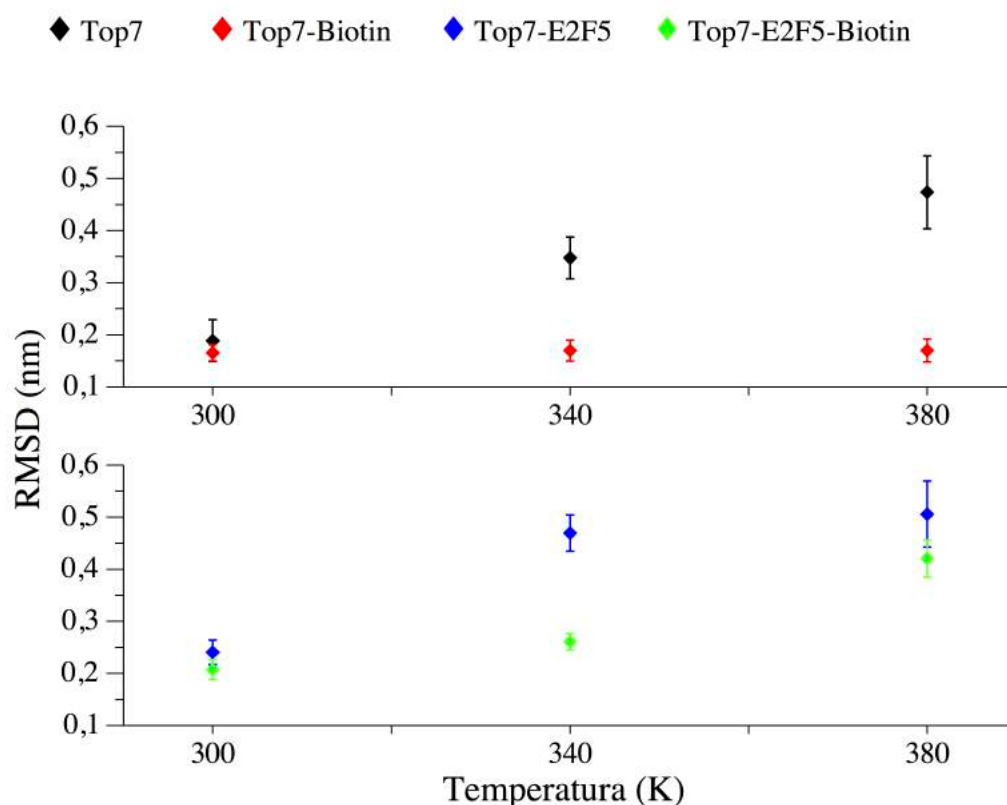


Figura 35 - Valores de RMSD médios em função da temperatura de simulação para cada proteína. Após biotinição, os valores dos RMSD sistematicamente foram menores do que para as proteínas nativas, indicando uma estabilidade estrutural conferida pela biotina.

Por outro lado, a Top7-biotina apresenta consistentemente baixos valores de RMSD, independentemente da temperatura de simulação: $0,16 \pm 0,02$ nm, $0,17 \pm 0,02$ nm e $0,17 \pm 0,02$ nm a 300 K, 340 K e 380 K, respectivamente (Figura 35, gráfico acima, losangos vermelhos). Um comportamento semelhante é observado para o Top7-2F5-biotina (Top7-2F5-Biotina), que exibe RMSD médio de $0,21 \pm 0,02$ nm, $0,26 \pm 0,02$ nm e $0,42 \pm 0,04$ nm a 300 K, 340 K e 380 K, respectivamente (Figura 35, gráfico abaixo, losangos verdes). Com exceção da Top7-2F5-Biotina a 380 K, os baixos valores de RMSD sugerem que ambas as proteínas biotiniladas permanecem com estruturas próximas à nativa mesmo em temperaturas relativamente altas, apontando para a maior estabilidade dos sistemas biotinilados em comparação com suas versões não-biotiniladas.

A flutuação média quadrática posicional (RMSF) para os átomos do *backbone* também foi calculada para todos os sistemas nos últimos 50 ns

(Figura 36). Para a proteína Top7 nativa, as regiões de maiores flutuações atômicas em todas as temperaturas simuladas correspondem a regiões de *turn/loops* e motivos em α -hélice (Figura 36 A, curvas pretas). Os valores de RMSF aumentam com o aumento da temperatura, com as flutuações nas regiões de α -hélice a 380 K alcançando valores três vezes maiores do que o observado em 300 K. Por outro lado, o RMSF da Top7-Biotina (Figura 36 A, curvas vermelhas) permaneceu abaixo de 0,2 nm independentemente da temperatura.

Similarmente, para a proteína Top7-2F5 (Figura 36 B, curvas azuis), os maiores valores de RMSF corresponderam a regiões de *turn/loops* e motivos em α -hélice, incluindo a região do epítipo 2F5. Como esperado, os valores de RMSF aumentam com o aumento da temperatura e foram notavelmente mais altos na região do epítipo a 380 K, onde o RMSF atingiu valores três vezes maiores do que o observado a 300 K. A análise das flutuações posicionais da Top7-2F5-biotina (Figura 36 B, curvas verdes) revelou que os valores de RMSF a 300 K foram em geral ligeiramente inferiores quando comparados com os da Top7-2F5 não-biotinilada à mesma temperatura. Curiosamente, ainda à 300 K, o *loop* do epítipo 2F5 torna-se mais flexível na proteína Top7-2F5-Biotina. Isto deve ser uma consequência da biotinilação no resíduo 25LYS localizada no meio do epítipo. Dessa forma, a maior flexibilidade do epítipo após biotinilação representa uma barreira para o reconhecimento do anticorpo, como observado nos resultados de ELISA (Figura 33). Em acordo com o que foi observado para Top7 e sua versão biotinilada, os valores de RMSF para Top7-2F5-Biotina permaneceram abaixo de 0,25 nm ao longo do tempo de simulação, independentemente da temperatura. Estes dados sugerem que a biotinilação da Top7 e Top7-2F5 conferem estabilidade estrutural às proteínas.

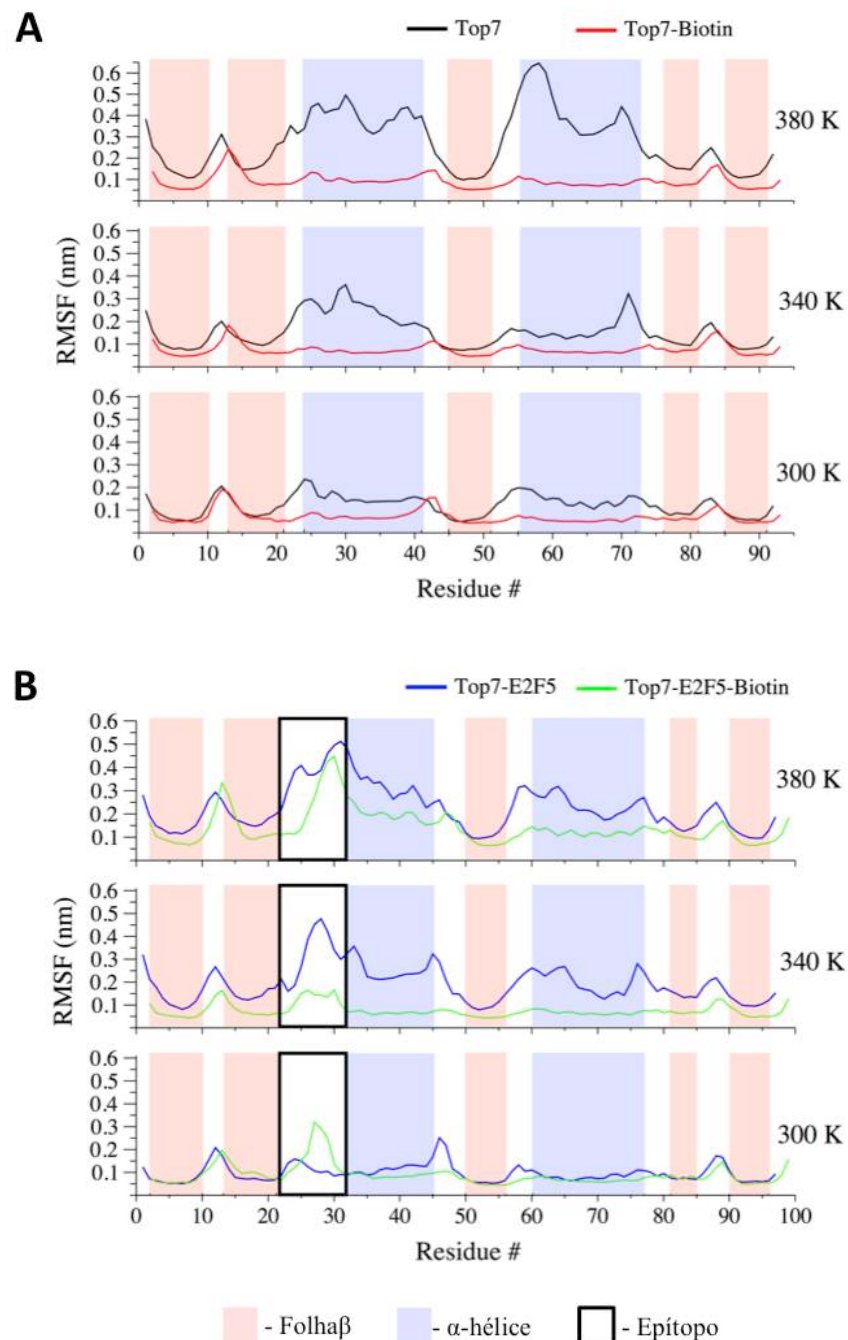


Figura 36 - A flutuação média quadrática posicional (RMSF) para os átomos do *backbone*. (A) Na Top7 nativa, as regiões de maiores flutuações atômicas correspondem a regiões de *turn/loops* e motivos em α -hélice. Os valores de RMSF aumentam com o aumento da temperatura. Por outro lado, o RMSF da Top7-Biotina permaneceu abaixo de 0,2 nm independentemente da temperatura. (B) Na Top7-2F5, os maiores valores de RMSF corresponderam a regiões de *turn/loops* e motivos em α -hélice, com valores de RMSF aumentando com o aumento da temperatura. Na Top7-2F5-biotina os valores de RMSF foram geral ligeiramente inferiores quando comparados com os da Top7-2F5. O *loop* do epítipo 2F5 torna-se mais flexível na proteína Top7-2F5-Biotina devido a uma biotinilação no resíduo 25LYS localizada no meio do epítipo.

Os perfis de estrutura secundária dependentes do tempo (DSSP) para os quatro sistemas foram analisados (Figura 37) e corroboram os dados anteriormente discutidos. Os conteúdos de estrutura secundária de todas as estruturas foram bem mantidos em todas as trajetórias a 300 K. Para a proteína Top7, pequenas perturbações no conteúdo helicoidal foram observadas a 340 K. No entanto, conforme a temperatura aumenta para 380 K, mudanças mais dramáticas na estrutura secundária são observadas. Ocorre uma desestabilização permanente de ambas as α -hélices, assim como o surgimento de conteúdo de folha β , indicando o desenovelamento dessa proteína. Em contraste, o perfil de DSSP da Top7-Biotina não mostra qualquer ruptura parcial ou permanente da estrutura secundária a qualquer temperatura, sugerindo que a biotinilação fornece estabilidade estrutural independentemente da temperatura de simulação.

Observou-se comportamento semelhante para a Top7-2F5, cuja versão não biotinilada exibiu desestabilização notável e crescente da primeira hélice α após o aumento da temperatura e surgimento de conteúdo de folha β na região do epítipo. Por outro lado, Top7-2F5-Biotina segue a mesma tendência observada para Top7-Biotina, sem qualquer ruptura aparente do conteúdo da estrutura secundária nas temperaturas analisadas e perda de conteúdo helicoidal apenas observada na temperatura mais alta (380 K).

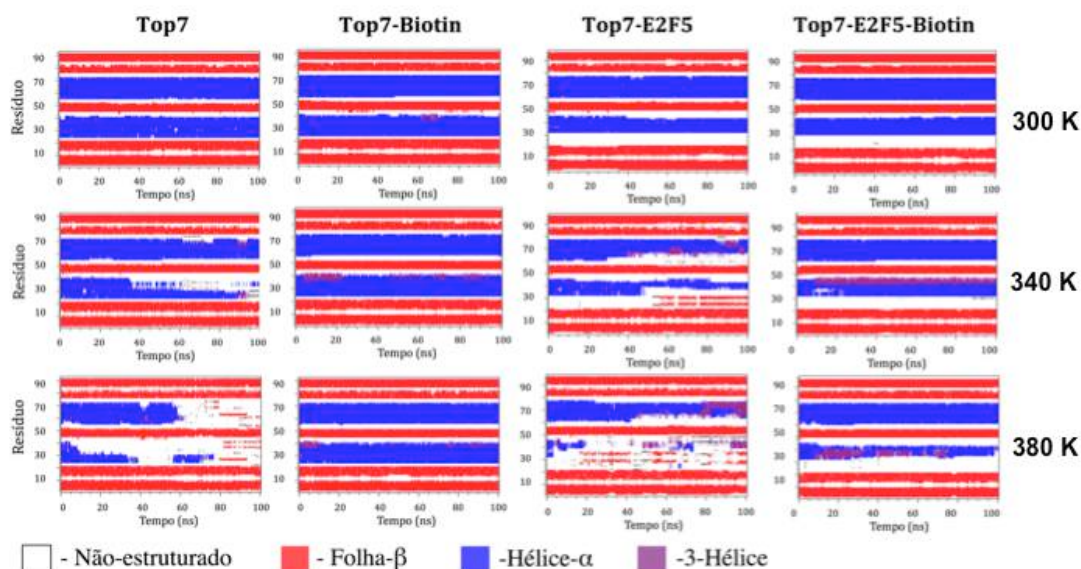


Figura 37 - Perfis de estrutura secundária dependentes do tempo (DSSP) para os quatro sistemas, indicando que todas as estruturas secundárias foram bem mantidas em todas as trajetórias a 300 K. À medida que a temperatura de simulação aumenta, as proteínas nativas (Top7 e Top7-2F5) perdem conteúdo estrutural. Em contraste, o perfil de DSSP das proteínas biotiniladas sofre menos perturbações com o aumento da temperatura, indicando que a biotina confere estabilidade estrutural.

Comparando o *backbone* do epítipo nas proteínas Top7-2F5 e Top7-2F5-Biotina com a sua estrutura cristalográfica complexada com o anticorpo mAb 2F5 (PDB ID: 1TJI), pode-se observar comportamentos distintos (Figura 38). Quando na Top7-2F5, o epítipo exibe apenas uma conformação principal, com RMSD de cerca de 0.38 nm (Figura 38 A). Por outro lado, quando o epítipo possui a 25LYS biotinilada, apresenta duas populações conformacionais, com RMSD de cerca de 0.36 e 0.40 nm (Figura 38 B).

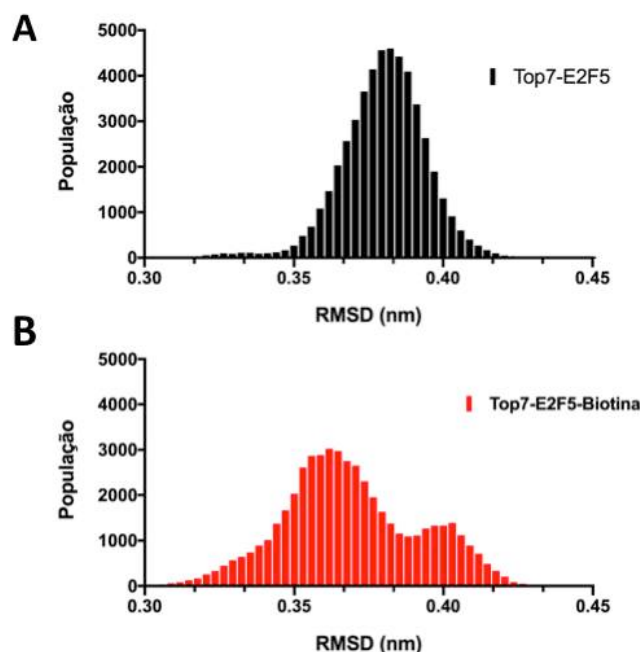


Figura 38 - Distribuição populacional dos valores de RMSD entre os átomos do *backbone* do epítipo na proteína (A) Top7-2F5 e (B) Top7-2F5-Biotina. Os valores são em relação à conformação do epítipo quando complexado ao anticorpo mAb 2F5 (PDB 1TJI), indicando que o epítipo não-biotinilado possui uma conformação preferencial, enquanto que o biotinilado possui duas conformações.

Os efeitos estruturais devido à presença da biotina no epítipo, observados anteriormente nas análises de RMSF, e nas diferentes distribuições de conformações em relação à estrutura nativa do epítipo sugerem que deve haver interferência na interação do antígeno com o anticorpo devido à mudança de conformação dos sítios de ligação (Sinha *et al.*, 2002), como observado nos dados de ELISA (Figura 33).

4.3.2 Perfil eletroestático da proteína biotinilada

Sabe-se também que as propriedades eletrostáticas são importantes para a reatividade e especificidade dos anticorpos (Sinha *et al.*, 2002). Dessa forma, os potenciais de superfície eletrostática foram calculados e plotados na superfície molecular de estruturas representativas da Top7-2F5 e Top7-2F5-Biotina (Figura 39). Uma divergência significativa pode ser observada, pois enquanto a Top7-2F5-Biotina mostra um perfil notavelmente mais negativo (Figura 39 à direita), a Top7-2F5 exibe uma distribuição de regiões

positiva e negativa mais homogênea (Figura 39 à esquerda). Além disso, há uma mudança de perfil eletroestático na região do epítipo 2F5 (Figura 39, destacada em verde). Essa mudança deve contribuir para a diferença de afinidade entre antígeno e anticorpo observada nos dados de ELISA (Figura 33), devido à interferência nos contatos intermoleculares complementares baseados em interações de carga na interface de ligação.

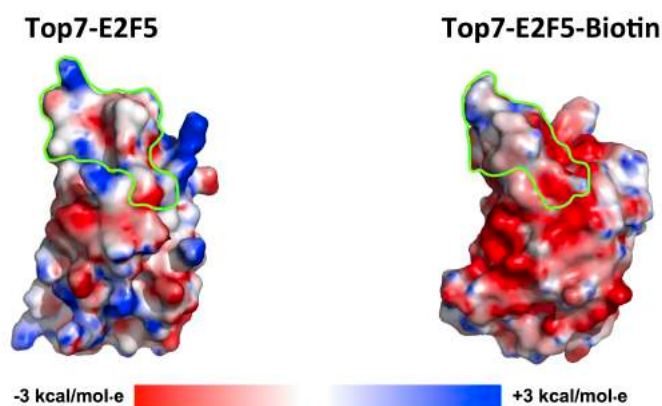


Figura 39 - Superfície de potencial eletroestático das proteínas Top7-2F5 e Top7-2F5-Biotina. A localização do epítipo está destacada em verde. A Top7-2F5-Biotin mostra um perfil notavelmente mais negativo, enquanto que a Top7-2F5 exibe uma distribuição de regiões positiva e negativa mais homogênea.

A mudança dramática na carga líquida da proteína também altera as interações de curto e longo alcance e pode causar uma alteração na resistência à agregação da proteína (O'brien *et al.*, 2016). Adicionalmente, essa carga líquida notavelmente negativa pode acarretar na melhora da solubilidade das proteínas Top7-Biotina e Top7-2F5-Biotina, prevenindo a sua agregação e precipitação (Boschek *et al.*, 2009; Kramer *et al.*, 2012; Perchiacca *et al.*, 2014). Embora os efeitos da temperatura e do tempo não possam ser diretamente comparados, as evidências de estabilização estrutural oferecidas pelas simulações de MD indicam que mesmo um aumento modesto na temperatura acarreta em perda de conteúdo helicoidal e aumento do teor de folhas para as proteínas não-biotiniladas (Figura 37). Esse é um efeito típico de agregação prévia e que não ocorre nas proteínas biotiniladas em temperaturas até 380K, para o mesmo período de simulação. Portanto, a biotinilação das proteínas não apenas confere estabilidade estrutural, evitando desenovelamento das proteínas, como a mudança de

perfil eletroestático melhora a solubilidade das proteínas e previne agregação e precipitação, em acordo com o observado experimentalmente (Figura 33 B).

4.4 CONCLUSÕES

Os espectros de CD indicam que o processo de biotinylação não compromete o *folding* e, portanto, a Top7-Biotina e a Top7-2F5-Biotina são proteínas com estruturas similares às nativas.

Os resultados computacionais indicam que a Top7-2F5-Biotina tem maior flexibilidade e mudanças nas propriedades eletrostáticas na região do *loop* do epítipo devido à presença de uma biotina no resíduo 25LYS. Estes resultados podem explicar a menor capacidade de serem reconhecidos pelo mAb 2F5. A razão para a afinidade da Top7-2F5-Biotina diminuir como uma função do tempo permanece ainda incerta.

Adicionalmente, a mudança nas propriedades eletrostáticas também melhora a solubilidade das proteínas biotiniladas, como observado experimentalmente, ao se medir a taxa de precipitação ao longo do tempo. Além disso, a carga negativa acentuada da proteína biotinilada impede que as moléculas se aproximem e agreguem, fazendo com que a taxa de precipitação diminua.

Levando em conta as propriedades estruturais e eletrostáticas observadas, a Top7-2F5-Biotina constitui uma proteína estruturalmente muito estável e, apesar da afinidade ser menor do que a Top7-2F5 nativa, pode ser utilizada como biomarcador de células B que expressam anticorpos anti-2F5.

5 DESENVOLVIMENTO *IN SILICO* DE ANTÍGENOS RECOMBINANTES BASEADOS NAS PROTEÍNAS ENVELOPE E NS4B DE ZIKA

O vírus da Zika (ZIKV), assim como o DENV, é um vírus do gênero *Flavivírus*, pertencente à família *flaviridae*. O ZIKV foi isolado pela primeira vez em 1947 na floresta Zika em Uganda e um ano depois, também foi isolado de mosquitos *Aedes* na mesma floresta, evidenciando um potencial ciclo de transmissão envolvendo primatas (Priyamvada *et al.*, 2017). Existe um consenso de que *Aedes aegypti* e *Aedes albopictus* são os principais vetores do ZIKV em áreas urbanas.

A primeira infecção humana com ZIKV foi detectada na Nigéria em 1953, mas apenas 13 casos naturalmente adquiridos foram relatados durante décadas (Petersen *et al.*, 2016). Em 2007, foi relatado um surto de uma doença caracterizada por erupções cutâneas, febre, conjuntivite e dores nas articulações na Ilha de Yap, localizada na Micronésia (Duffy *et al.*, 2009). Testes de RT-PCR detectaram RNA do ZIKV no soro dos pacientes, porém nenhum outro arbovírus, sendo este o primeiro surto documentado da doença. Em 2013 um segundo surto foi detectado na Polinésia Francesa, com uma suspeita de aproximadamente 6.000 casos da doença (Van-Mai *et al.*, 2014). O *Aedes aegypti* foi o principal vetor de transmissão neste caso, e pela primeira vez complicações neurológicas da Síndrome de Guillain-Barré (GBS) foram associadas ao ZIKV (Carod-Artal, 2018).

A Organização Mundial da Saúde (OMS) recebeu os primeiros relatos de infecção no Brasil em maio de 2015, com os primeiros casos da doença ocorrendo na Bahia e se espalhando para pelo menos 14 estados no meses subsequentes (Kindhauser *et al.*, 2016; Petersen *et al.*, 2016). Dados epidemiológicos mostram que o Haiti atuou como intermediário para a chegada e estabelecimento do ZIKV no Brasil, de onde o vírus se espalhou posteriormente para outras regiões (Faria *et al.*, 2017). Em julho do mesmo ano, foi relatada a associação entre a infecção pelo ZIKV e a síndrome de Guillain-Barré em adultos brasileiros e posteriormente de casos de microcefalia em bebês cujas mães haviam sido expostas ao ZIKV durante a gravidez (De Araújo *et al.*, 2016; Kindhauser *et al.*, 2016). Diante dos fatos, investigadores da Polinésia Francesa identificaram retrospectivamente um

aumento no número de anomalias fetais, incluindo microcefalia, após o surto do ZIKV naquele país (Petersen *et al.*, 2016). No início de 2016 a OMS declarou que a infecção pelo ZIKV associada à microcefalia e outros distúrbios neurológicos constituía uma emergência em Saúde Pública de preocupação internacional (Kindhauser *et al.*, 2016). Há evidências de que o RNA do ZIKV pode ser detectado em fluidos corporais (plasma, urina, saliva, sêmen) até 8 meses após infecção, e o vírus na sua forma infecciosa pode permanecer por até 30 dias no sêmen de homens infectados (Barzon *et al.*, 2018; Mead *et al.*, 2018).

O RNA do ZIKV codifica uma poliproteína única que é clivada em três proteínas estruturais (Capsídeo (C), pré-membrana (prM) e envelope (E), que formam a partícula viral) e sete proteínas não-estruturais (NS1, NS2A, NS2B, NS3, NS4A, NS4B e NS5) (Freire *et al.*, 2017). A estrutura de cryo-EM do ZIKV revela que o vírus tem uma estrutura organizacional quase idêntica ao DENV (Sirohi *et al.*, 2016). Estudos têm demonstrado ocorrência de reatividade cruzada entre DENV e ZIKV, com anticorpos contra um vírus reagindo com as proteínas homólogas do outro (Bosch *et al.*, 2017; Priyamvada *et al.*, 2017). O mapeamento da sequência na proteína NS1 de ZIKV revelou dois epítomos conformacionais e um linear (Figura 40), no entanto, os resíduos próximos aos epítomos não são conservados e conferem um potencial de superfície eletrostática divergente nas regiões de epítopo dos ZIKV e DENV2 (Freire *et al.*, 2017). Atualmente, existem poucos kits de sorologia disponível comercialmente, para a detecção de ZIKV (David *et al.*, 2017; Inbios International, 2017). Mas a especificidade e sensibilidade de tais ensaios é prejudicada pela similaridade estrutural entre ZIKV e DENV e a consequente reatividade cruzada, que leva a resultados falso-positivos (David *et al.*, 2017; Granger *et al.*, 2017). Dessa forma, para diferenciar infecções de ZIKV e DENV, deve-se utilizar métodos mais sofisticados, como o ensaio Trioplex RT-PCR em tempo real (Santiago *et al.*, 2018). Portanto, há a necessidade de se desenvolver técnicas mais baratas que tenham sensibilidade e especificidade para diferenciar flavivírus.

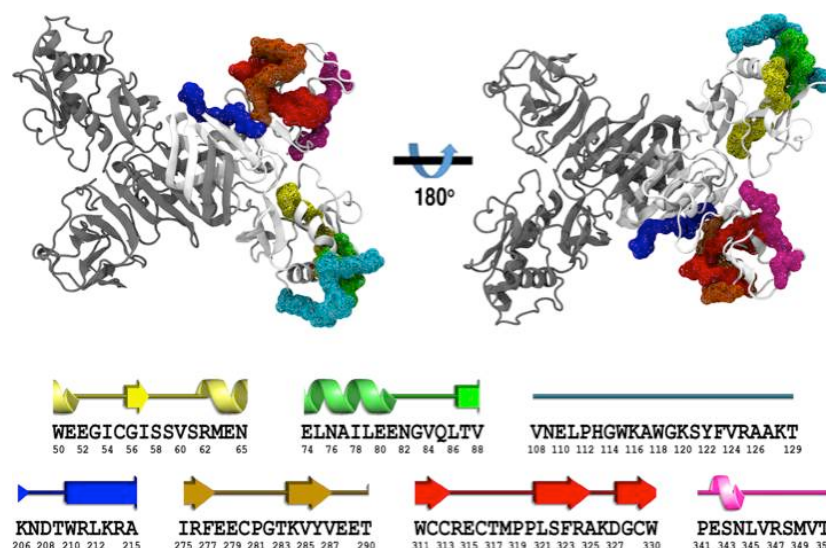


Figura 40 - Representação em *cartoon* da estrutura da NS1 de ZIKV. As cadeias de proteínas são mostradas em branco e cinza. Os epítomos putativos são mapeados como a superfície de estrutura de arame e cor codificados como E1 em amarelo, E2 em verde, E3 em ciano, E4 em azul, E5 em laranja, E6 em vermelho e E7 em magenta. Retirado de (Freire *et al.*, 2017).

Atualmente, não existem vacinas ou drogas antivirais licenciadas para prevenir ou tratar Zika. A forma de prevenção se dá por medidas de erradicação do vetor. Algumas estratégias de desenvolvimento de vacina têm se baseado no uso da proteína E de ZIKV. A proteína E é formada por cerca de 500 aminoácidos, e possui um ectodomínio de 400 aminoácidos, subdividido em três domínios (DI, DII e DIII) (Figura 41). Wang e colaboradores demonstraram que anticorpos humanos anti-ZIKV-E adquiridos durante a infecção são capazes de neutralizar o vírus *in-vitro* e conferir proteção pós-exposição a camundongos *in vivo* (Wang *et al.*, 2016). De forma semelhante, Zhao e colaboradores desenvolveram anticorpos monoclonais de camundongos anti-ZIKV-E capazes de neutralizar cepas Africana, Asiática e Americana do ZIKV (Zhao *et al.*, 2016). Barba-Spaeth e colaboradores demonstraram que anticorpos adquiridos contra proteína E de DENV também são capazes de neutralizar ZIKV (Barba-Spaeth *et al.*, 2016). Entretanto, alguns estudos também demonstraram que a capacidade de ADE de anticorpos induzidos por DENV pode se estender ao ZIKV (Priyamvada *et al.*, 2017). Dado que o DENV também é endêmico em muitas regiões infectadas por ZIKV, o desenvolvimento de anticorpos específicos para ZIKV, que não reajam de forma cruzada com outros flavivírus, é desejável para

evitar qualquer risco de ADE (Wang *et al.*, 2016; López-Camacho *et al.*, 2018). Rockstroh e colaboradores demonstraram que o *design* de uma proteína baseada no envelope, de modo a diminuir a identidade entre os flavivírus, pode conferir especificidade (Rockstroh *et al.*, 2017). Isso enfatiza as possíveis vantagens do desenvolvimento de vacinas baseadas em epítomos.

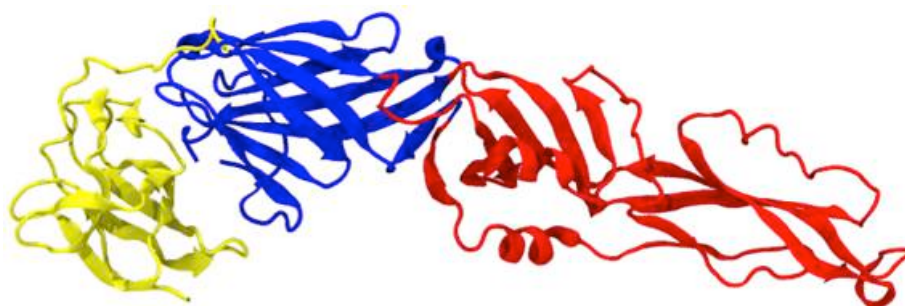


Figura 41 - Representação *cartoon* da proteína Envelope do ZIKV (PDB ID 5GZN). Domínio I (azul), domínio II (vermelho) e domínio III (amarelo).

Outra estratégia que vem sendo utilizada é o desenvolvimento de vacinas de DNA. Larocca e colaboradores produziram uma vacina de DNA, expressando as proteínas prM e E de ZIKV, proporcionando proteção contra o ZIKV (Larocca *et al.*, 2016). López-Camacho e colaboradores demonstraram a eficácia de uma vacina de DNA utilizando o adenovírus chimpanzé com deficiência de replicação (ChAdOx1) como vetor da vacina (López-Camacho *et al.*, 2018). Em teste fase 1, Tebas e colaboradores concluíram que uma vacina de DNA, que codifica as proteínas E e prM do ZIKV, após três doses, induziu respostas imunes anti-ZIKV (Tebas *et al.*, 2017). Outra vacina também em fase I, desenvolvida por Gaudinski e colaboradores, composta por um plasmídeo que também codifica as proteínas prM e E de ZIKV, se mostrou uma vacina segura de duas doses e com resposta positiva de anticorpos detectáveis (Gaudinski *et al.*, 2018). Entretanto, nenhuma das vacinas desenvolvidas até o momento conseguiu induzir imunidade duradoura nem impedir ADE com DENV.

Uma das formas de se avaliar a eficácia de uma vacina é analisar a capacidade de neutralização do ZIKV pelos anticorpos produzidos *in vitro*.

Para isso, realiza-se um ensaio de microneutralização. No entanto, é difícil realizar uma comparação precisa da eficácia da vacina com outras vacinas relatadas, pois os títulos de neutralização variam de um ensaio para outro (López-Camacho *et al.*, 2018). Dessa forma, faz-se necessário o desenvolvimento de um ensaio robusto o suficiente para ser transferido inter laboratórios, e que serão altamente informativos para selecionar de forma eficiente os candidatos a vacina contra zika em ensaios pré-clínicos.

Este trabalho se propôs a utilizar a engenharia de proteínas para desenvolver peptídeos e proteínas quiméricas de epítomos do ZIKV para fins de diagnóstico e/ou vacinais.

5.1 OBJETIVOS

Desenho *in silico* de antígenos de ZIKV como potenciais candidatos a vacina e/ou antígenos de diagnóstico utilizando métodos de engenharia de proteínas em conjunto com análise por dinâmica molecular.

5.1.1 Objetivos específicos

- Selecionar epítomos putativos de ZIKV;
- Utilizar os epítomos na forma de peptídeos ou transplantar os epítomos para as proteínas *scaffolds* através do *MotifGraft*;
- Analisar estabilidade estrutural dos peptídeos/quimeras através de dinâmica molecular;

5.2 METODOLOGIA

5.2.1 Alinhamento sequencial

Para selecionar um epítomo de ZIKV proveniente da proteína E, a sequência da proteína E de ZIKV (Q91KX7) foi alinhada com proteínas E de DENV1 (Q7TGE4), DENV2 (AHB63923), DENV3 (ABA25838), DENV4

(Q80L18), WNV (Q91R00) e YFV (Q89286), através do algoritmo de Pesquisa de Alinhamento Local Básico de Proteína-Proteína (BLASTp). A região de menor identidade com os demais flavivírus foi selecionada.

Uma segunda análise mais robusta e abrangente foi realizada pelo colaborador do nosso grupo Dr. Ricardo Durães, do Departamento de Virologia do Instituto Aggeu Magalhães, Fiocruz - PE. De forma breve, um alinhamento foi realizado utilizando todas as sequências disponíveis de ZIKV e DENV-1-4 no GenBank (n = 10.834) e no Resource Database Virus and Gogen Resource (n = 24.284) (ViPR, disponível em <https://www.viprbrc.org>), totalizando 35.118 sequências. Visando encontrar peptídeos específicos para serem utilizados na discriminação entre o ZIKV e os 4 sorotipos de DENV, as sequências foram alinhadas pelos algoritmos de alto rendimento alocados no programa MUSCLE v3.8.31 (Edgar, 2004) e em seguida, foram analisadas sequências consenso e graus de conservação com o programa Jalview v.2 (Waterhouse *et al.*, 2009). As sequências peptídicas preditas foram adicionalmente analisadas com o algoritmo BLASTp contra as sequências de DENV, CHIKV, vírus West Nile (WNV) e febre amarela (YF) contidas nos bancos de dados de sequências de proteínas. Como critério de escolha, foram levados em consideração os graus de homologia e identidade entre as sequências, e o valor de E (quanto mais próximo de zero, mais significativo o resultado).

5.2.2 Motifgraft

Para desenhar as proteínas quiméricas carreadoras dos epítipo de ZIKV, foi utilizada a abordagem *MotifGraft* (Silva *et al.*, 2016) do programa Rosetta v3.7 (Leaver-Fay *et al.*, 2011), com auxílio de *scripts* em formato XML do *RosettaScripts* (Fleishman *et al.*, 2011).

Para a construção da biblioteca de *scaffolds*, foram obtidas 1248 estruturas de proteínas do PDB (www.rcsb.org) com as seguintes características: 1) estruturas cristalográficas determinadas por raio-X com resolução < 2.5 Å; 2) proteínas que foram obtidas por expressão em *E. coli*; 3) estruturas monoméricas; e 4) ausências de ligantes. Todas as estruturas

foram minimizadas utilizando o *MinMover* com valor de 1.5 Å de RMSD em relação à estrutura inicial como ponto de corte.

Os motivos estruturais dos epítomos foram obtidos de estruturas cristalográficas de alta resolução depositadas no PDB. Os *backbones* dos epítomos foram computacionalmente alinhados aos *scaffolds*, utilizando um RMSD de 1.5 Å como valor de corte (acima disso o alinhamento era descartado). Após o alinhamento passar pelo filtro de RMSD, as cadeias laterais do epítomo foram transplantadas para o *scaffold* (*side-chain grafting*). Após o transplante, realizou-se um *re-design* dos aminoácidos num raio de 10 Å ao redor do epítomo, de modo estabilizar energeticamente a nova proteína quimérica.

5.2.3 Dinâmica molecular

Coordenadas iniciais para cada antígeno foram obtidas dos outputs do Rosetta na etapa de engenharia de proteína. Cada proteína foi centralizada numa caixa cúbica de 8.0 x 8.0 x 8.0 nm e solvatadas explicitamente usando modelo de água SPC (Berendsen *et al.*, 1981) com condições periódicas de contorno nas direções x, y e z. O método LINCS foi usado para restringir ligações de hidrogênio (Hess *et al.*, 1997). As temperaturas dos solutos e dos solventes foram acopladas separadamente a um termostato usando o esquema de reescalonamento de velocidade (*velocity rescale scheme*) (Bussi *et al.*, 2007) com temperatura de referência de 310 K e tempo de relaxamento de 0.2 ps. A pressão foi mantida constante no valor de 1.013 bar por escalonamento de coordenadas isotrópico usando um barostato de Berendsen (Berendsen *et al.*, 1984) com constante de relaxamento de 0.2 ps e compressibilidade (κ_T) de $4.5 \times 10^{-5} \text{ bar}^{-1}$. Um raio de corte de 1.4 nm foi utilizado para interações não-ligadas de pequeno alcance. Interações eletrostáticas de longo alcance foram contabilizadas usando o método de *reaction field* (Tironi *et al.*, 1995) com $\epsilon=66$ além do raio de corte de 1.4 nm. Contra íons foram adicionados para garantir a eletroneutralidade do sistema e a uma concentração de 150 mM de NaCl. Os sistemas foram inicialmente minimizados usando 5.000 passos com o algoritmo *steepest descent*.

Integração foi realizada usando um tempo de integração de 2 fs baseado no algoritmo *leapfrog* (Hockney, 1970). Os sistemas foram aquecidos até a temperatura de simulação usando um *ensemble* NVT por 100 ps seguido de uma pressurização num *ensemble* NPT por 500 ps. Termalização e pressurização foram realizadas com restrição de posição para todos os átomos do soluto. As simulações de dinâmica molecular atomísticas foram realizadas em *ensemble* NPT numa trajetória de 100 ns utilizando o campo de força GROMOS 53A6 (Oostenbrink *et al.*, 2004). As simulações e posterior análises foram realizadas com o programa Gromacs 4.6.7 (Hess *et al.*, 2008), DSSP (Database of Secondary Structure in Proteins) software (Kabsch e Sander, 1983) e análises visuais através do programa Visual Molecular Dynamics software (VMD) 1.8.7 (Humphrey *et al.*, 1996).

5.3 RESULTADOS E DISCUSSÃO

5.3.1 Sequências selecionadas

Após o alinhamento das sequências, foram selecionadas as regiões de menor identidade entre ZIKV e os outros flavivírus. Na Tabela 4 estão as 4 regiões selecionadas: 1 sequência na proteína NS4b (NS4b) e 3 na proteína E (E1, E2 e E3). As sequências identificadas pelo Dr Ricardo Durães foram NS4b, E2 e E3, enquanto que o peptídeo E1 foi identificado independentemente pelo nosso grupo.

Os epítomos putativos E2 e E3 possuem motivos estruturais nativos distribuídos entre regiões não-estruturadas e regiões folhas- β , como mostra a Figura 42. Estas foram as sequências utilizadas para o *MotifGraft*. No caso da NS4b não é possível demonstrar a região nativa do epítopo selecionado, pois não há até o momento estrutura resolvida para esta proteína.

Tabela 4 - Sequências selecionadas para o *design* de antígenos de ZIKV.

Proteína NS4b			
NS4b	SHLMGRREEGATIGFS	Resíduos 14-29	Não-estruturado
Proteína E			
E1	HGSQHSGMIVNDTGHETDENRAK	Resíduos 144-166	Coil//3-Hélice/Folha- β
E2	HETDENRAKVE	Resíduos 158-168	Coil//Folha- β
E3	VITESTENSKMML	Resíduos 364-376	Coil//Folha- β

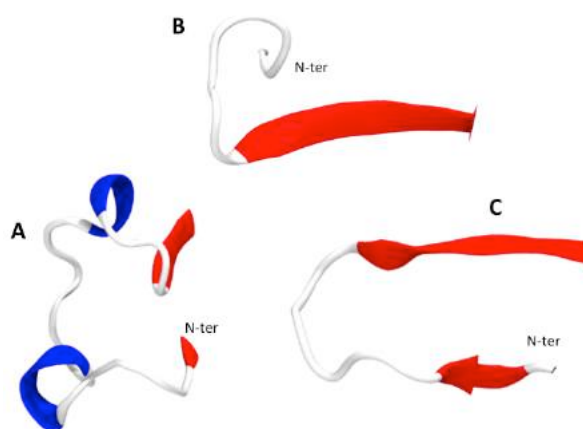


Figura 42 - Motivos estruturais das sequências selecionadas como epítipos putativos de ZIKV. (A) Resíduos 144-166 da proteína E, epítipo E1; (B) Resíduos 158-168 da proteína E, epítipo E2; e (C) Resíduos 364-376 da proteína E, epítipo E3.

5.3.2 Proteínas quiméricas contendo epítipo NS4b

Não há até o presente momento nenhuma estrutura cristalográfica da proteína NS4B de ZIKV na qual possamos nos basear para o desenho de proteínas quiméricas. Porém, sabe-se que a NS4B se trata de uma proteína transmembrana e cuja região em que estamos interessados (resíduos 14-29) se trata de uma porção não-estruturada que fica exposta ao solvente (Bryan *et al.*, 2015; Li *et al.*, 2016). Para realizar a modelagem de *loop* nas proteínas *scaffolds*, utilizamos o algoritmo *Kinematic Closure Confirmation* (KIC) (Mandell *et al.*, 2009). Como *scaffold*, utilizamos duas proteínas já

conhecidas na literatura por serem utilizadas como carreadoras de epítipo: a Top7 e a fibronectina tipo III humana (FNIII) (Figura 43).

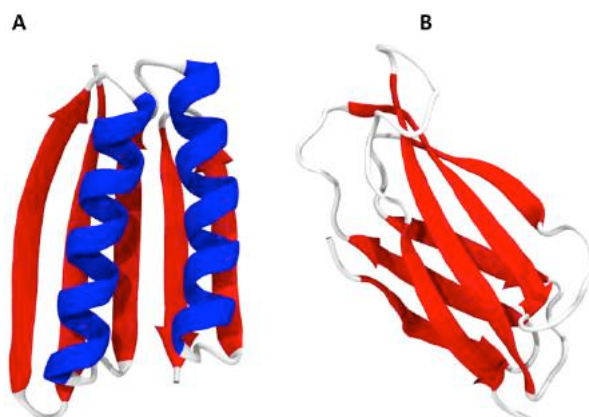


Figura 43 - Representação *cartoon* das proteínas *scaffolds* utilizadas para o desenho de quimeras contendo o epítipo NS4b. (A) Top7 (PDB ID 1QYS) e (B) Fibronectina tipo III humana (PDB ID 1FNA).

A sequência de NS4b selecionada possui 16 aminoácidos, mas a Top7 não possui uma região não-estruturada tão longa. Dessa forma, foi preciso incluir os *loops* na estrutura da proteína de forma a manter o epítipo exposto ao solvente. Foram escolhidas as regiões entre os aminoácidos 22-23, de forma que o *loop* foi criado entre a segunda folha- β e primeira α -hélice (Figura 44 A); e entre os resíduos 42-44, que fica entre a primeira α -hélice e terceira folha- β (Figura 46 A).

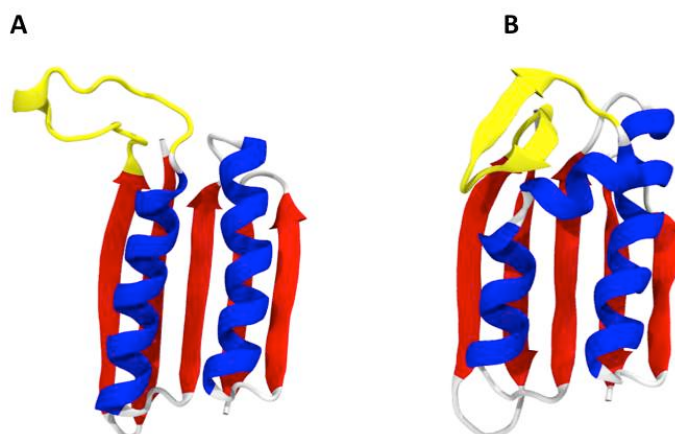


Figura 44 - Representação *cartoon* da Top7 contendo o epítipo de NS4b entre os resíduos 22 e 23. (A) Estrutura inicial obtida com o protocolo KIC, mostrando a região do epítipo não-estruturada e completamente exposta ao solvente. (B) Estrutura final da proteína após 100 ns de simulação de MD,

mostrando que o epítopo adota uma conformação de folha- β e desestabiliza a hélice- α adjacente ao epítopo.

Para avaliar a estabilidade das proteínas quiméricas foi realizada uma MD em condições fisiológicas. No caso da quimera construída com a Top7 e entre os resíduos 22 e 23, o RMSD teve valor médio de aproximadamente 0.3 nm (Figura 45 A, gráfico superior). O gráfico de RMSF indica que o *scaffold* não possui muita flexibilidade, o que se esperaria da Top7 (Figura 45 A, gráfico inferior), exceto na região onde o epítopo foi construído. Os baixos valores de RMSD e RMSF seriam indícios de que a presença do epítopo não desestabiliza o *scaffold*, apesar da região do epítopo ser bem flexível. Porém, ao analisar o conteúdo de estrutura secundária da proteína ao longo da simulação (Figura 45 B) e a sua estrutura final (Figura 44 B), fica evidente que o epítopo que inicialmente era uma região não-estruturada passa a adotar uma conformação em folha- β . Essa mudança não apenas desvia o epítopo de sua conformação nativa, como também provoca uma desestabilização na hélice- α adjacente ao epítopo. Dessa forma, conclui-se que não é possível a construção de uma proteína estável contendo o epítopo de NS4b de ZIKV, baseada na Top7 e entre os resíduos 22 e 23.

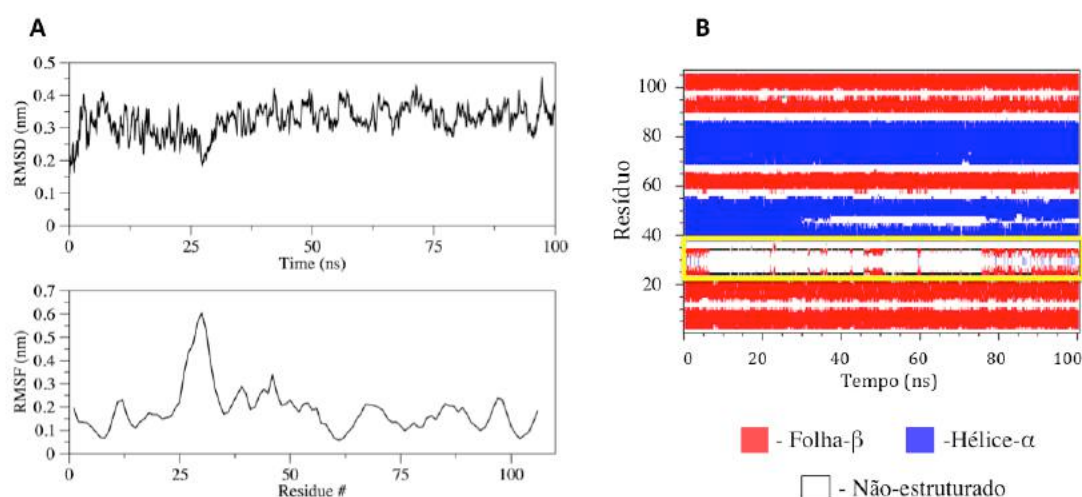


Figura 45 - Análise da MD da proteína Top7 contendo epítopo de NS4b entre os resíduos 22 e 23. (A) O RMSD (gráfico superior) permanece com valor médio de 0.3 nm durante toda a simulação, e o RMSF (gráfico inferior) indica uma proteína com baixa flexibilidade, exceto na região do epítopo. (B) O conteúdo de estrutura secundária durante a simulação indica mudança de conformação nativa do epítopo (destacado em amarelo), ao mesmo tempo que desestabiliza a hélice- α adjacente ao epítopo.

Analizando a proteína construída com o mesmo epítipo, porém inserindo o *loop* entre os resíduos 42 e 44 da Top7, observam-se altos valores de RMSD (Figura 47 A, gráfico superior), indicando desvio da configuração inicial e uma desestabilização da proteína. O gráfico de RMSF (Figura 47 A, gráfico inferior) indica duas regiões bastante flexíveis da proteína construída: a região do epítipo, já esperado por se tratar de um *loop*; e a região de hélice- α adjacente ao epítipo, sugerindo que este motivo estrutural se desestabiliza. Ao analisar o conteúdo de estrutura secundária da proteína ao longo da simulação (Figura 47 B) e a sua estrutura final (Figura 46 B), observa-se que apesar de o epítipo manter a sua conformação nativa, há uma desestabilização na hélice- α adjacente, como sugerido pelas análises de RMSF. Adicionalmente, a folha- β adjacente a esta hélice também sofre desorganização. Dessa forma, conclui-se também que não é possível a construção de uma proteína estável contendo o epítipo de NS4b de ZIKV, baseada na Top7 e entre os resíduos 42 e 44.

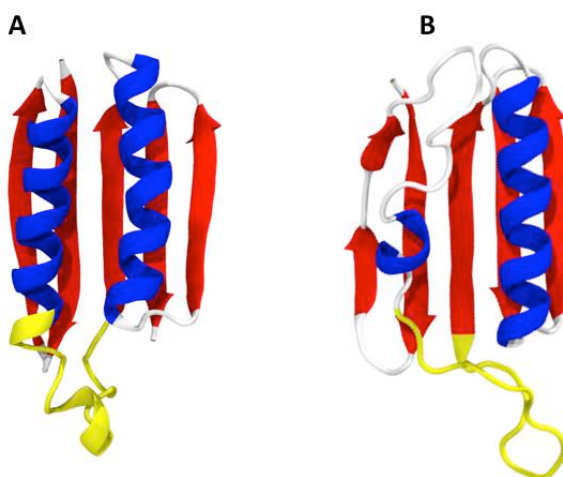


Figura 46 - Representação *cartoon* da Top7 contendo o epítipo de NS4b entre os resíduos 42 e 44. (A) Estrutura inicial mostrando a região do epítipo não-estruturada e completamente exposta ao solvente. (B) Estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação não-estruturada mas desestabilizando completamente a hélice- α adjacente ao epítipo e parcialmente a segunda folha- β da proteína.

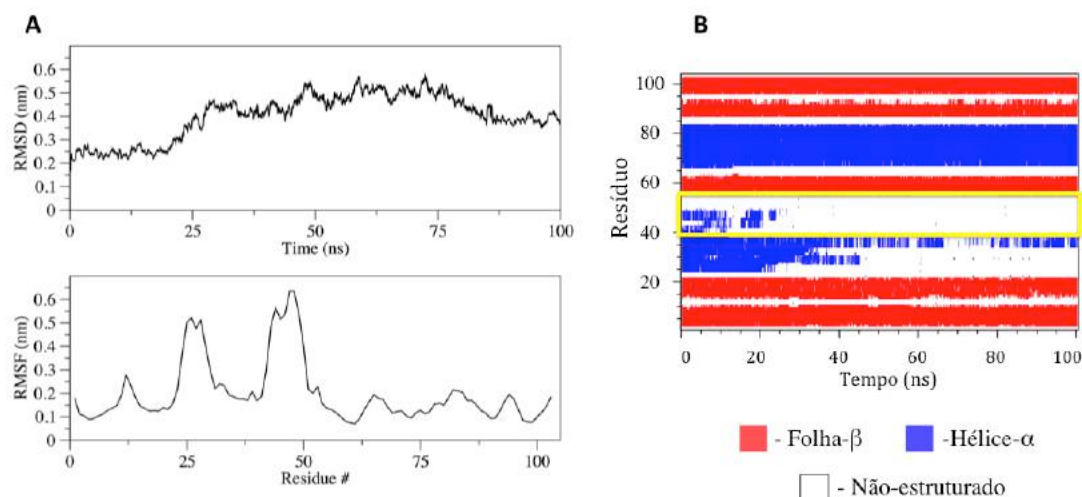


Figura 47 - Análise da MD da proteína Top7 contendo epítipo de NS4b entre os resíduos 42 e 44. (A) O RMSD (gráfico superior) adota valores altos a partir de 25 ns de simulação, sugerindo uma desestabilização da estrutura. O RMSF (gráfico inferior) indica duas regiões de alta flexibilidade da proteína, uma sendo o epítipo e a outra a hélice- α adjacente. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa do epítipo (destacado em amarelo), ao mesmo tempo que desestabiliza a hélice- α adjacente ao epítipo.

Utilizando a FNIII como *scaffold* foi construída uma proteína contendo o epítipo no *loop* entre a 2^a e 3^a folha- β (Figura 48). Analisando o RMSD durante a simulação (Figura 49 A, gráfico superior), observam-se valores médios de 0.3 nm, com um aumento repentino no final. Este aumento súbito se deve a uma mudança de direção do *loop* do epítipo, como pode ser observado nas Figuras 5.9 A e B, e não deve ser interpretado como uma desestabilização da proteína. O gráfico de RMSF indica que o *scaffold* não possui muita flexibilidade (Figura 49 A, gráfico inferior), exceto na região onde o epítipo foi construído, novamente esperado por se tratar de uma região não-estruturada. Os baixos valores de RMSD e RMSF são indícios de que a presença do epítipo não desestabiliza o *scaffold* e a proteína desenhada mantém uma estrutura próxima à nativa.

Ao analisar o conteúdo de estrutura secundária da proteína ao longo da simulação (Figura 49 B) e a sua estrutura final (Figura 48 B), novamente se observa a manutenção não apenas da estrutura em *loop* do epítipo como de toda a estrutura do *scaffold*. Dessas análises, pode-se concluir que a proteína quimérica utilizando FNIII como *scaffold* e o epítipo NS4b de ZIKV é

uma estrutura viável e candidata para avaliação experimental de sua imunorreatividade. A identidade entre a FNIII nativa e a FNIII-NS4b-ZIKV contendo o epítipo de NS4b é de 86% (Tabela 5).

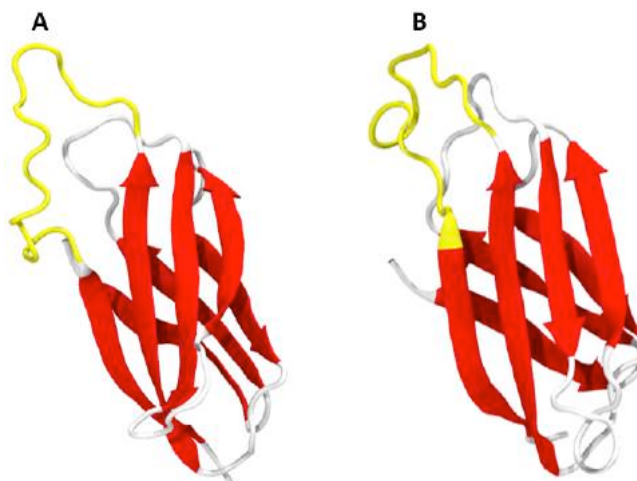


Figura 48 - Representação *cartoon* da FNIII contendo o epítipo de NS4b. (A) Estrutura inicial mostrando a região do epítipo não-estruturada e completamente exposta ao solvente. (B) Estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação não-estruturada e nenhuma perturbação estrutural no *scaffold*.

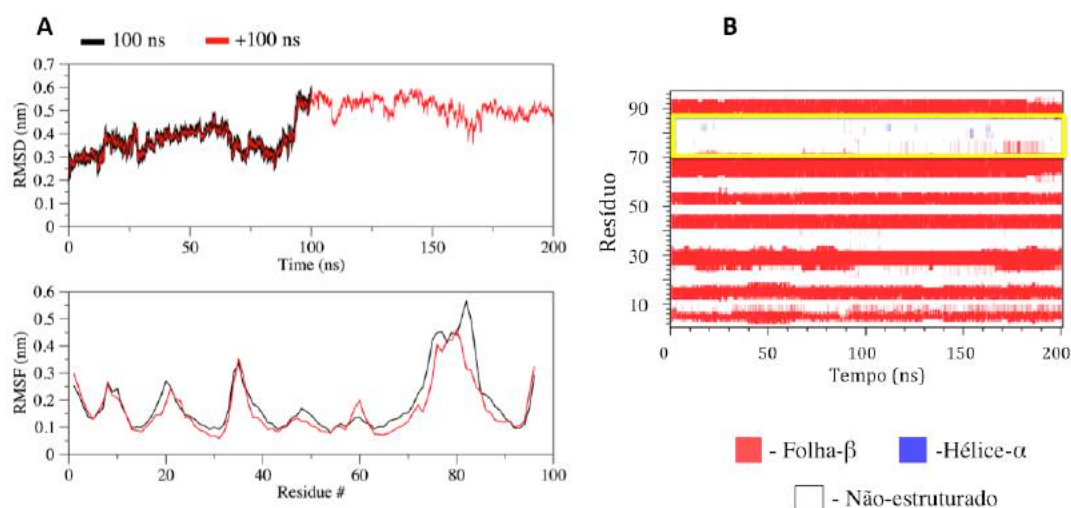


Figura 49 - Análise da MD da proteína FNIII contendo epítipo de NS4b entre a 2ª e 3ª folha- β . (A) O RMSD (gráfico superior) adota valores relativamente baixos, sugerindo uma proteína sem perturbação estrutural. O RMSF (gráfico inferior) indica apenas uma região de alta flexibilidade da proteína, onde foi colocado o epítipo. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítipo (destacado em amarelo), como de todo o *scaffold*.

Tabela 5 - Proteína quimérica desenhada contendo o epítipo de NS4b de ZIKV

Código	Sequência
FNIII	RDLEVVAATPTSLISWDAPAVTVRYRITYGETGGNSPVQEFTVP GSKSTATISGLKPGVDYTITVYAVTGRGDSPASSKPISINYRTEI
FNIII-NS4b- ZIKV	RDLEVVAATPTSLISWDAPAVTVRYRITYGETGGNSPVQEFTVP GSKSTATISGLKPGVDYTITVYAVT SHLMGRREEGATIGFS ISINYR TEI

5.3.3 Proteínas quiméricas contendo epítipos do envelope

As coordenadas dos epítipos E2 e E3 utilizadas para alinhamento no *MotifGraft* foram retiradas da estrutura cristalográfica de alta resolução da proteína envelope de ZIKV (PDB 5GZN). Nesta etapa, das 1248 estruturas utilizadas na biblioteca de *scaffolds*, apenas 3 obtiveram alinhamentos favoráveis para o *graft* dos epítipos: a Top7 e FNIII para o epítipo E2; e a TL1A (Figura 50) que é uma proteína humana do grupo das citocinas, capaz de provocar a morte de células tumorais, para o epítipo E3.



Figura 50 - Proteína TL1A humana utilizada como *scaffold*. Pertence à família de citocinas capazes de provocar morte de células cancerígenas. PDB ID 2RJL.

Utilizando a Top7 como *scaffold*, o epítipo E2 pôde ser transplantado na segunda folha- β da proteína (Figura 51 A). Analisando os resultados de MD, observa-se que o RMSD (Figura 52 A, gráfico superior) assume valores

acima de 0.3 nm a partir de 25 ns, sugerindo uma grande mudança de conformação da proteína. O gráfico de RMSF (Figura 52 A, gráfico inferior) indica que a região entre os resíduos 20 e 40 a proteína apresenta uma alta flexibilidade, que pode estar relacionada com a mudança de conformação sugerida pelo gráfico de RMSD. O conteúdo de estrutura secundária da proteína ao longo da simulação (Figura 52 B) e a sua estrutura final (Figura 51 B) demonstram claramente que apesar do epítipo manter a sua conformação nativa, há um desenovelamento parcial da primeira hélice- α da proteína. Conclui-se que não é possível a construção de uma proteína estável contendo o epítipo de E2 da proteína E de ZIKV baseada na Top7.

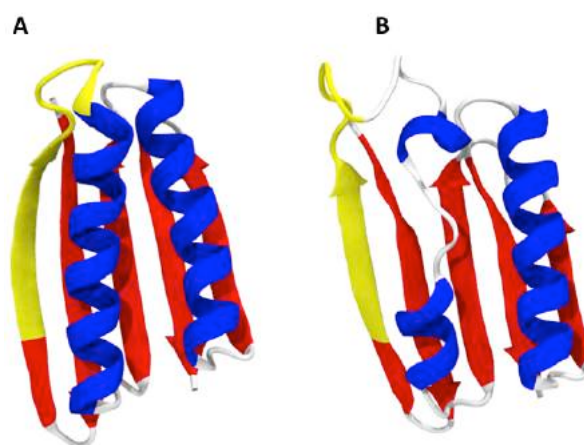


Figura 51 - Representação *cartoon* da Top7 contendo o epítipo E2 após alinhamento com o *MotifGraft*. (A) Estrutura inicial mostrando a região da folha- β onde o epítipo foi transplantado (destacado em amarelo) e (B) estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação nativa, porém desestabilização da hélice- α adjacente.

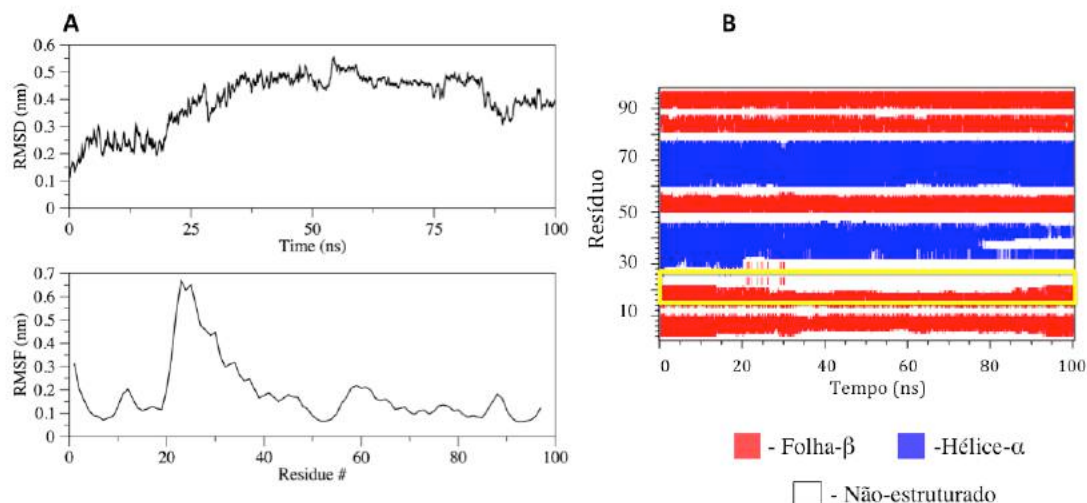


Figura 52 - Análise da MD da proteína Top7 contendo epítipo E2 de ZIKV na 2ª folha- β . (A) O RMSD (gráfico superior) assume valores acima de 0.3 nm, sugerindo uma mudança de conformação da proteína, corroborada pelos altos valores de RMSF (gráfico inferior) entre os resíduos 20 e 40. (B) O conteúdo de estrutura secundária durante a simulação indica desenovelamento parcial da primeira hélice- α , assim como instabilidade da folha- β onde o epítipo foi transplantado (destacado em amarelo).

Alinhando o epítipo E2 à FNIII, foi possível construir uma proteína contendo o epítipo na 4ª folha- β (Figura 53). Analisando o RMSD durante a simulação (Figura 54 A, gráfico superior, curva preta), observam-se valores baixos, convergindo para aproximadamente 0.3 nm, sem nenhuma indicação de mudança brusca de conformação da proteína. O gráfico de RMSF (Figura 54 A, gráfico inferior) indica apenas uma região de alta flexibilidade, entre os resíduos 70 e 80, mas que se trata de um *loop* de onde se espera ter grande mobilidade. Os baixos valores médios de RMSD e RMSF são indícios de que a presença do epítipo não desestabiliza o *scaffold* e a proteína desenhada mantém a sua estrutura nativa. Ao analisar o conteúdo de estrutura secundária da proteína ao longo da simulação (Figura 54 B) e a sua estrutura final (Figura 53 B), observa-se a manutenção da estrutura nativa tanto do epítipo como de toda a estrutura do *scaffold*.

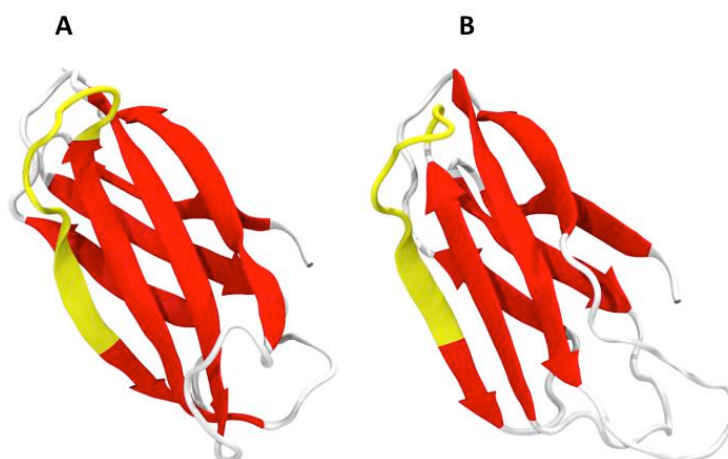


Figura 53 - Representação *cartoon* da FNIII contendo o epítipo E2 após alinhamento com o *MotifGraft*. (A) Estrutura inicial mostrando a região da folha- β /Coil onde o epítipo foi transplantado (destacado em amarelo) e (B) estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação nativa, e o *scaffold* não apresenta nenhuma mudança considerável de estrutura secundária.

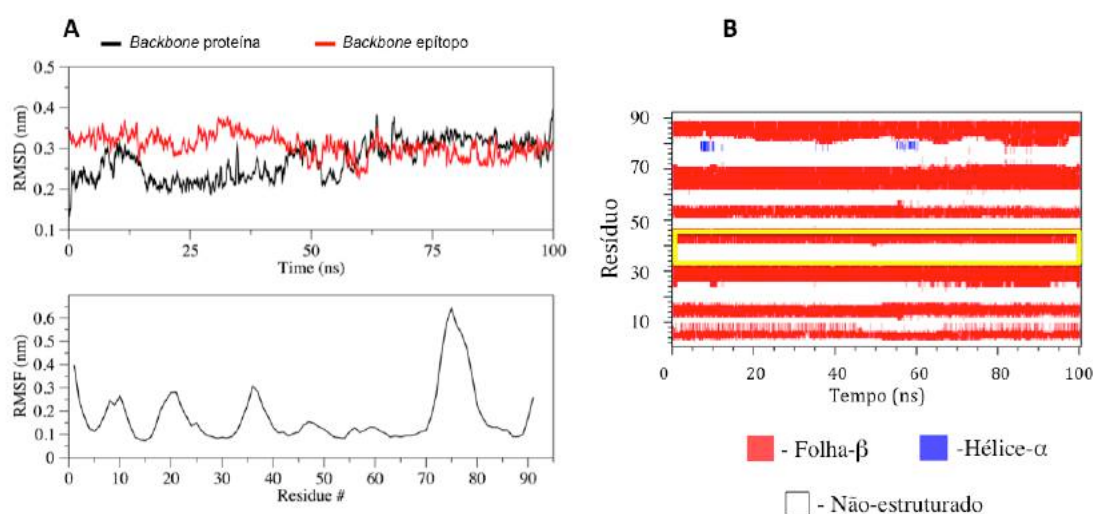


Figura 54 - Análise da MD da proteína FNIII contendo o epítipo E2 da proteína E de ZIKV na 4ª folha- β . (A) O RMSD adota valores baixos, com média de 0.25 nm, sugerindo uma proteína estruturalmente estável. O RMSD do epítipo (curva vermelha) também sugere baixa mudança de conformação da estrutura nativa. O RMSF indica apenas uma região de alta flexibilidade da proteína, onde se encontra uma região em *loop*. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítipo (destacado em amarelo), como de todo o *scaffold*.

Sabendo se tratar de uma proteína estável, decidimos comparar a estrutura do epítipo na quimera com a sua configuração nativa (Figura 42 A).

Os átomos do *backbone* do epítipo na quimera foram alinhados com o *backbone* da estrutura nativa ao longo da trajetória (Figura 54 A, gráfico superior, curva vermelha). O RMSD do epítipo converge para um valor de 0.3 nm, sugerindo similaridade com a conformação nativa. A área de superfície exposta ao solvente (SASA) dos resíduos do epítipo também foi calculada durante a trajetória e comparadas com a exposição na estrutura nativa (Figura 55). Como observado nas barras em vermelho, os resíduos na proteína quimérica estão tão ou até mais expostos em comparação com a estrutura cristalográfica da proteína E, indicando que o epítipo está apto a ser reconhecido por anticorpos. Aqui vale ressaltar que os aminoácidos N, A e V, que estão minimamente expostos na proteína E, não foram transplantados para o *scaffold*, pois não devem interferir no reconhecimento, e para evitar maior perturbação no *scaffold*. Dessas análises, pode-se concluir que a proteína quimérica utilizando FNIII como *scaffold* e o epítipo E2 da proteína E de ZIKV é uma estrutura estável e candidata para avaliação experimental de suas propriedades imunológicas. A identidade entre a FNIII nativa e a FNIII-E2-ZIKV contendo o epítipo E2 é de 90% (Tabela 6).

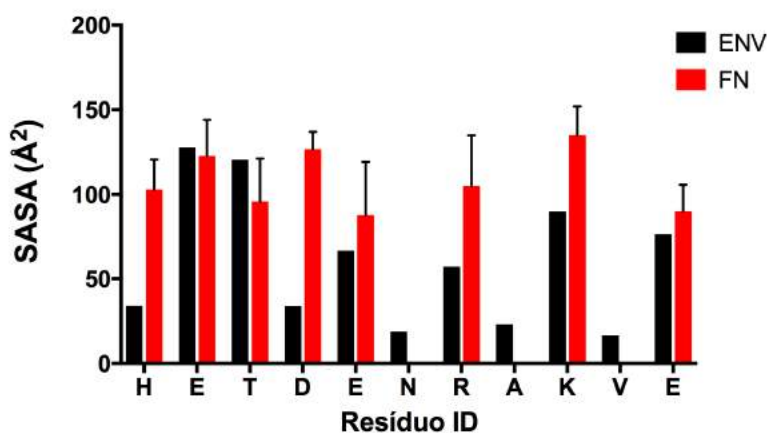


Figura 55 - Gráfico comparando a superfície de acessibilidade ao solvente (SASA) entre o epítipo em sua conformação nativa (preto) e durante a trajetória de MD transplantado na FNIII (vermelho). Os resíduos estão tão expostos como quando em sua estrutura nativa, indicando estarem aptos para reconhecimento. Os resíduos N, A e V não foram inseridos no *scaffold*.

Para o epítipo E3, foi possível alinhar e construir uma proteína utilizando a TL1A como *scaffold* (Figura 56 A). Durante a simulação de MD a proteína aparenta adotar uma conformação diferente da inicial, sugerido pelos altos valores de RMSD, e com regiões de alta flexibilidade, como indicado pelo gráfico de RMSF (Figura 57 A). Porém, ao analisar o conteúdo de estrutura secundária da proteína ao longo da simulação (Figura 57 B) e a sua estrutura final (Figura 56 B), não se observa perda considerável de estrutura secundária. Por outro lado, há uma mudança de posição do *loop* localizado na parte superior central da proteína, como indicado nas Figura 56 A e B. Este *loop* possui uma alta mobilidade, o que faz com que tanto o RMSD e o RMSF da proteína adotem valores altos. Apesar disso, a proteína consegue manter a sua estrutura nativa, assim como o epítipo.

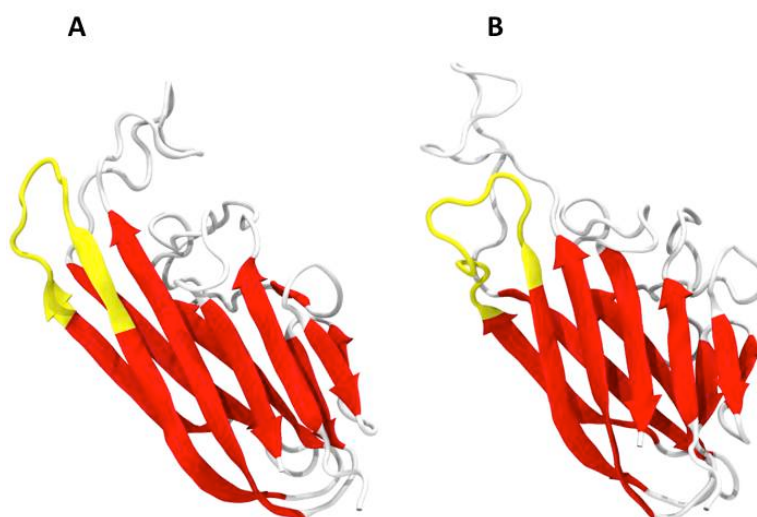


Figura 56 - Representação *cartoon* da TL1A contendo o epítipo E3 após alinhamento com o *MotifGraft*. (A) Estrutura inicial mostrando a região da folha- β /*Coil* onde o epítipo foi transplantado (destacado em amarelo) e (B) estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação nativa, e o *scaffold* apresenta apenas uma mudança de posição do *loop* central.

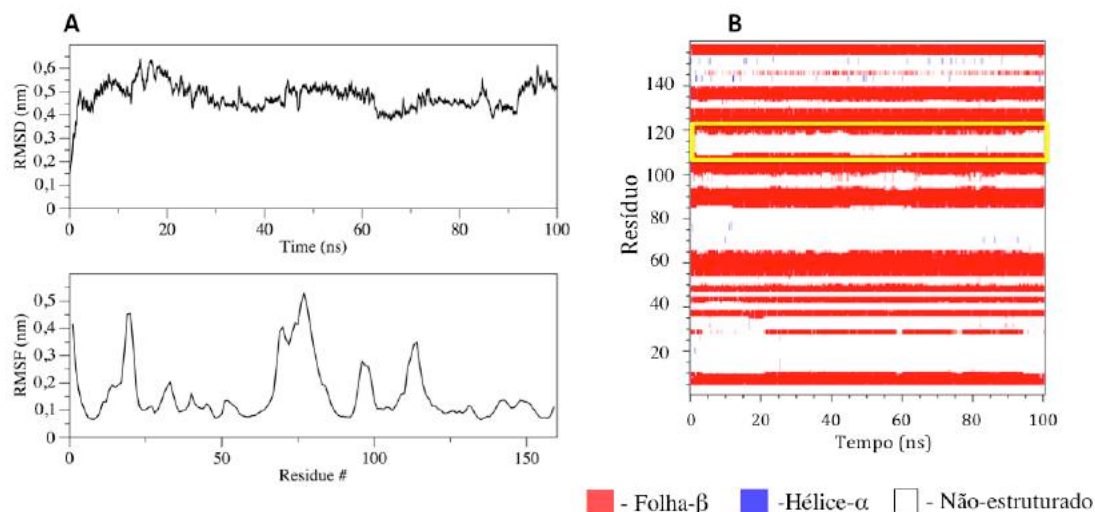


Figura 57 - Análise da MD da proteína TL1A contendo o epítipo E3 da proteína E de ZIKV. (A) O RMSD (gráfico superior) adota valores relativamente altos, sugerindo alguma modificação estrutural. O RMSF (gráfico inferior) indica duas regiões de alta flexibilidade da proteína, onde se encontram regiões predominantemente em *loop*, cuja alta mobilidade contribui para os altos valores de RMSD. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítipo (destacado em amarelo), como de todo o *scaffold*.

Sabendo da alta mobilidade do *loop* central da proteína e que estando próximo ao epítipo poderia dificultar o reconhecimento pelo anticorpo, a próxima etapa foi redesenhar a proteína excluindo o *loop* em questão (Figura 58 A). Analisando a trajetória de MD dessa nova proteína, observa-se que tanto os valores de RMSD como os valores de RMSF (Figura 59 A) ficam bem mais baixo, indicando que a proteína passa a ser menos flexível e consegue manter a sua estrutura nativa. Ao analisar o conteúdo de estrutura secundária da proteína ao longo da simulação (Figura 59 B) e a sua estrutura final (Figura 58 B), observa-se a manutenção da estrutura nativa tanto do epítipo como de toda a estrutura do *scaffold*.

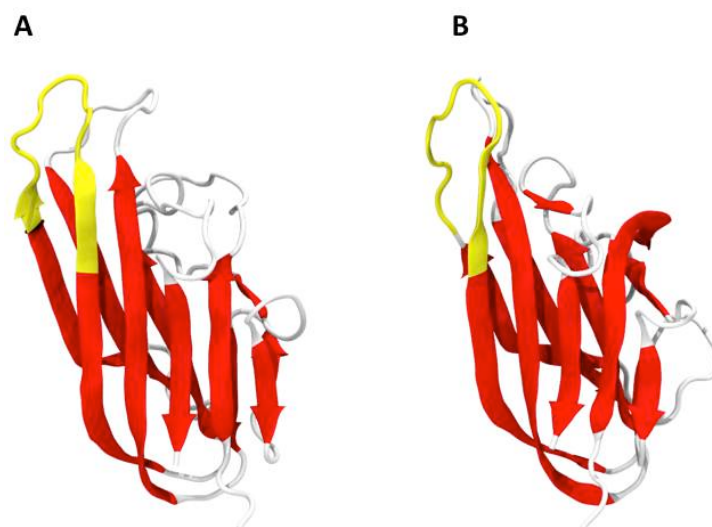


Figura 58 - Representação *cartoon* da TL1A contendo o epítipo E3 após exclusão do *loop* central. (A) Estrutura inicial mostrando a região da folha- β /*Coil* onde o epítipo foi transplantado (destacado em amarelo) e (B) estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação nativa.

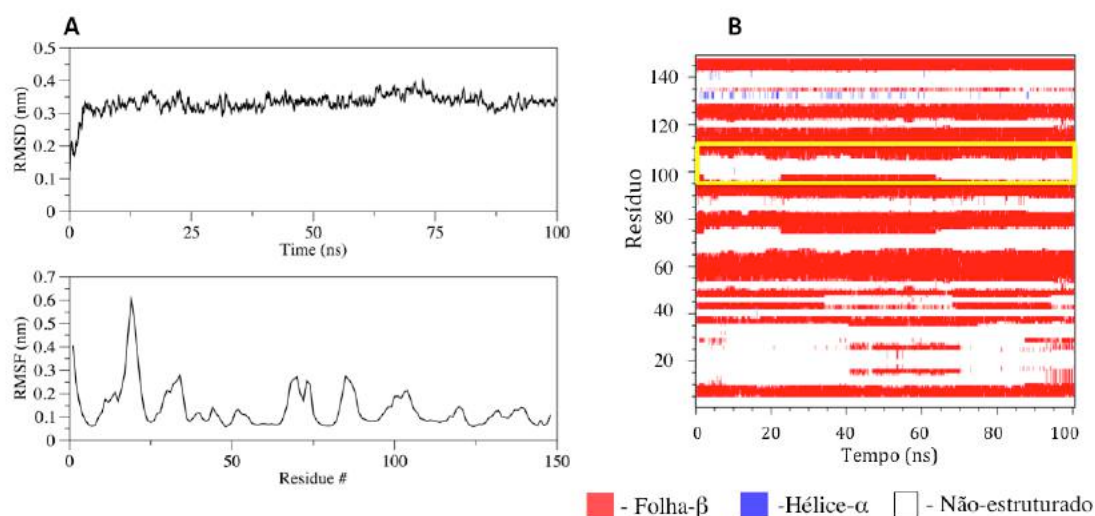


Figura 59 - Análise da MD da proteína TL1A contendo o epítipo E3 da proteína E de ZIKV após exclusão do *loop* central. (A) O RMSD (gráfico superior) mostra valores mais baixos se comparado com a presença do *loop*. O RMSF (gráfico inferior) indica apenas uma região de alta flexibilidade da proteína, onde se encontram predominantemente *loops*. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítipo (destacado em amarelo), como de todo o *scaffold*.

A TL1A está diretamente envolvida na patogênese de doenças autoimunes e interage diretamente com receptores DcR3 (Zhan *et al.*, 2009).

Fizemos mutações pontuais na proteína, de modo a diminuir a sua interação com os receptores DcR3, de acordo com o trabalho de Zhan e colaboradores (Zhan *et al.*, 2009). As seguintes mutações foram realizadas: E53A, E55A, G57D, Y121F, E174A. A nova proteína contendo as mutações foi também analisada através de MD.

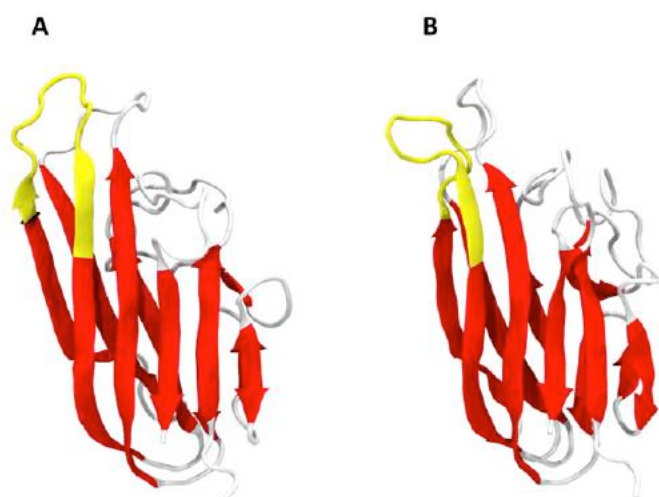


Figura 60 - Representação *cartoon* da TL1A contendo o epítipo E3 após exclusão do *loop* central e mutações. (A) Estrutura inicial mostrando a região da folha- β /Coil onde o epítipo foi transplantado (destacado em amarelo) e (B) estrutura final da proteína após 100 ns de simulação de MD, mostrando que o epítipo mantém a sua conformação nativa.

Após as mutações, a proteína TL1A contendo o epítipo E3 passa a adotar valores de RMSD maiores, convergindo para aproximadamente 0.4 nm (Figura 61 A, gráfico superior, curva preta). Os valores de RMSF ficam no mesmo patamar de antes das mutações (Figura 61 A, gráfico inferior), com um aumento da flexibilidade na região N-terminal desestruturada entre os resíduos 10 e 25. Ao analisar o conteúdo de estrutura secundária da proteína ao longo da simulação (Figura 61 B) e a sua estrutura final (Figura 60 B), observa-se a manutenção da estrutura nativa tanto do epítipo como de toda a estrutura do *scaffold*. Novamente, a fim de comparar o epítipo com sua estrutura nativa, os átomos do *backbone* do epítipo na quimera foram alinhados com o *backbone* da estrutura nativa ao longo da trajetória (Figura 61 A, gráfico superior, curva vermelha). O RMSD do epítipo permanece baixo durante a toda simulação, em torno de aproximadamente 0.2 nm. Os

valores de SASA dos resíduos do epítipo (Figura 62) indicam que os resíduos na proteína quimérica estão bem expostos em comparação com a estrutura cristalográfica da proteína E, indicando que o epítipo está apto a ser reconhecido por anticorpos. Dessa forma, pode-se concluir que a proteína quimérica utilizando a TL1A modificada como *scaffold* e o epítipo E3 da proteína E de ZIKV é uma estrutura estável e candidata para avaliação experimental de suas propriedades imunológicas.

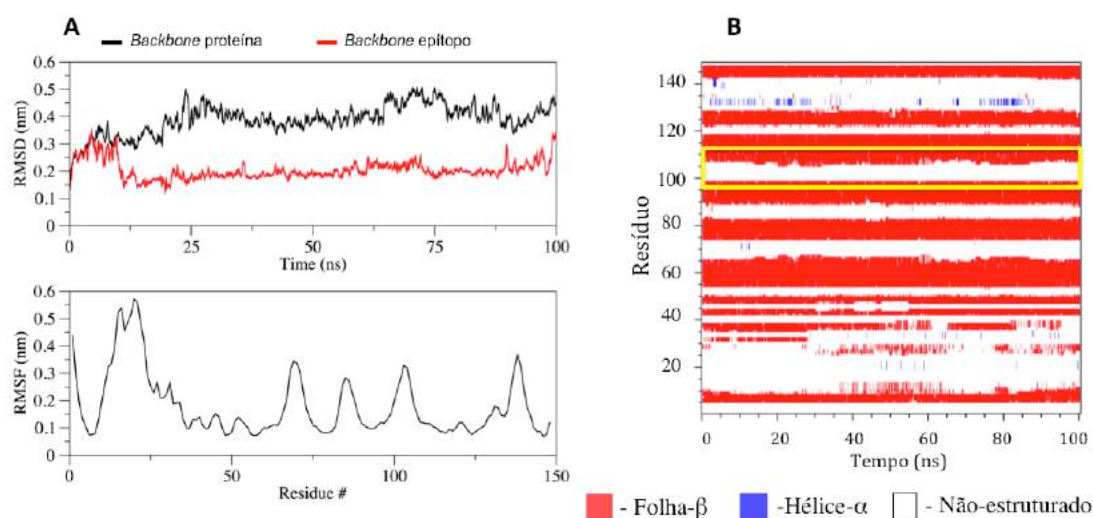


Figura 61 - Análise da MD da proteína TL1A contendo o epítipo E3 da proteína E de ZIKV após exclusão do *loop* central e mutações. (A) O RMSD (gráfico superior, curva preta) mostra valores um pouco altos, porém o epítipo se mantém bastante semelhante à sua configuração nativa (curva vermelha). O RMSF (gráfico inferior) indica apenas uma região de alta flexibilidade da proteína, onde se encontram predominantemente *loops*. (B) O conteúdo de estrutura secundária durante a simulação indica manutenção da conformação nativa tanto do epítipo (destacado em amarelo), como de todo o *scaffold*.

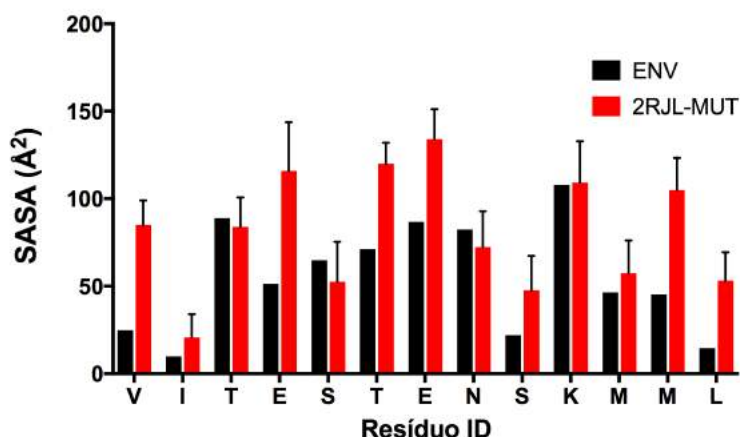


Figura 62 - Gráfico comparando SASA do epítipo E3 em sua conformação nativa (preto) e durante a trajetória de MD transplantado na TL1A (vermelho). Os resíduos estão tão expostos como quando em sua estrutura nativa, indicando estarem aptos para reconhecimento.

A identidade entre a proteína TL1A-E3-ZIKV, contendo o epítipo 3, e a TL1A nativa é de 69% (Tabela 6), já levando em consideração os 12 aminoácidos a menos do que o *scaffold* nativo. Como discutido, foram realizadas mutações para inibir a possível interação da proteína desenhada com os receptores DcR3. Porém, a TL1A nativa também interage com receptores DR3 (Zhan *et al.*, 2009). A nossa hipótese é de que após as modificações realizadas, a nova proteína desenhada terá baixa interação com os receptores. Validações experimentais dessa hipótese estão sendo realizadas. Caso a hipótese seja nula, desenvolveremos uma nova geração da proteína, mantendo a região do epítipo e o *core* hidrofóbico da proteína e redesenhando os demais resíduos.

Tabela 6 - Proteínas desenvolvidas usando os epítomos E2 e E3 de proteína E de ZIKV. A região do epítopo se destaca em negrito, a região deletada do scaffold em vermelho.

Código	Sequência
FNIII	RDLEVVAATPTSELLISWDAPAVTVRYRITYGETGGNSPVQEFTVP GSKSTATISGLKPGVDYTITVYAVTGRGDSPASSKPISINYRTEI
FNIII-E2-ZIKV	RDLEVVAATPTSELLISWDAPAVTVRYRITYG HETDEPRQKFEVPG SKSTATISGLKPGVDYTITVYAVTGRGDSPASSKPISINYRTEI
TL1A	GDKPRAHLTVVRQTPTQHFKNQFPALHWEHELGLAFTKNRMNYT NKFLIPESGDYFIYSQVTFRGMTSESSEI RQAGRPNKPDS ITVVIT KVTD SYPEPTQLLMGTSVSEVGSNWFQPIYLGAMFSLQEGDKLM VNVSDISLVDYTKEDKTFFGAFL
TL1A-E3-ZIKV	GDKPRAHLTVVRQTPTQDFKNQNPALHWEHELGLAFTKNRMNYT NKFLIPESGDYFIYGQVTFRGQSDKQSEFIYVITKVTD SYPEPTQL LSA VITESTENSKMML TLGDMFSLQEGDKLMGNVGDITAVDHT DERRAHFGAFL

5.3.4 Peptídeo baseado em envelope de ZIKV para ensaio de ELISA sanduíche

Ao se desenvolver uma nova vacina, é crucial que se ateste a sua eficácia, por meio de medidas de seu poder de neutralização do patógeno em questão. Uma das formas de se medir essa eficácia *in vitro* é por meio de ensaios de microneutralização (MN). Outros ensaios tradicionais, como os PRNT50 e FRNT50 também são adequados. Entretanto, é difícil realizar uma comparação precisa da eficácia de uma vacina com outras já relatadas, pois os títulos de neutralização variam de um ensaio para outro (López-Camacho *et al.*, 2018).

Um ensaio de MN consiste em detectar anticorpos neutralizantes no soro de animais ou indivíduos infectados ou vacinados contra um determinado agente infeccioso. No caso de uma vacina, quanto maior o título de anticorpos neutralizantes, mais eficaz é a vacina. A combinação de um ensaio de MN com um ensaio de ELISA para detectar anticorpos

neutralizantes pode produzir resultados altamente sensíveis e específicos para detectar anticorpos neutralizantes do vírus em estudo (Rowe *et al.*, 1999; Network, 2011).

Fazendo uma triagem na proteína E de ZIKV, pudemos identificar uma região de baixa similaridade com os demais flavivírus, o epítipo E1 da Tabela 4. Sabendo se tratar de uma região da proteína E de ZIKV, Lopez-Camacho e colaboradores desenvolveram quatro vacinas diferentes contra ZIKV que incluíam a expressão da proteína E (Figura 63). As vacinas foram desenvolvidas utilizando o vetor adenovírus de chimpanzé deficiente em replicação (ChAdOx1) como plataforma: a vacina prME que codifica a região pré-membrana, proteína E, e região trans-membrana; a prME Δ TM que exclui apenas a região trans-membrana; a Env que exclui apenas a região pré-membrana; e a Env Δ TM que exclui ambas as regiões pré-membrana e trans-membrana (López-Camacho *et al.*, 2018). Para atestar a eficácia de suas vacinas, foi necessário desenvolver um novo teste de ELISA sanduíche acoplado a um ensaio de MN, baseado no peptídeo E1.

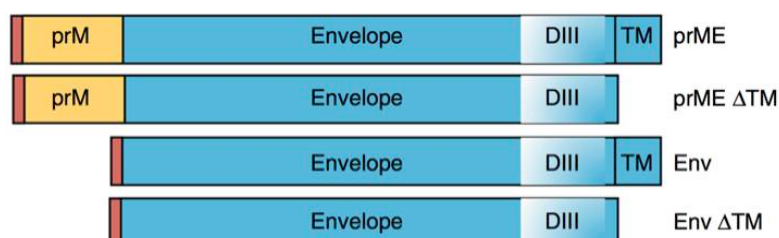


Figura 63 - Representação esquemática das vacinas de ChAdOx1 contra ZIKV desenvolvidas por López-Camacho e colaboradores. A vacina prME codifica a região pré-membrana, proteína E, e região trans-membrana; a prME Δ TM exclui apenas a região trans-membrana; a Env exclui apenas a região pré-membrana; e a Env Δ TM exclui ambas as regiões pré-membrana e trans-membrana. Retirado de (López-Camacho *et al.*, 2018).

O peptídeo E1 se encontra entre os resíduos 144 e 166 da proteína E de ZIKV, sendo composto por 23 aminoácidos e majoritariamente uma região não-estruturada da proteína (Figura 42 A). A identidade de E1 com seus respectivos homólogos nos demais flavivírus (Figura 64) fica abaixo de 40%, e é uma região da proteína E que está exposta ao solvente e apta para o

reconhecimento por anticorpos (Figura 65). O E1 foi simulado por MD para analisar a sua capacidade de mimetizar a estrutura nativa quando isolado.

Z I K V	H - G S Q H S G M I V N D T G H E T D E N R A K	
D E N V 1	H T G D Q H - - - - - Q V G N E S T E H G T T	I D : 3 4 %
D E N V 2	H S G E E H A - - V G N D T G K H G K E - - - -	I D : 3 9 %
D E N V 3	H T G D Q H Q - - V G N D T Q G V T - - - - -	I D : 3 4 %
D E N V 4	H N G D T H A - - - - - V G N D T S N H G V T	I D : 2 6 %
W N V	H - G P T T V E S H G N Y S T Q V G A T Q A G R	I D : 2 6 %
Y F	H V G A K Q E N W - - - - - N T D I K T L K	I D : 2 1 %

 - IDÊNTICO	 - SIMILAR	 - DIFERENTE
--	---	---

Figura 64 - A sequência do peptídeo E1 de ZIKV (Q91KX7) alinhado com regiões homólogas das proteínas E de DENV1 (Q7TGE4), DENV2 (AHB63923), DENV3 (ABA25838), DENV4 (Q80L18), WNV (Q91R00) e YFV (Q89286).

O mapa de DSSP do E1 ao longo da simulação sugere que o peptídeo transaciona entre conformações não-estruturadas e conformações em β -*hairpin* (Figura 66 A). Há porém uma preferência por conformações sem estrutura definida, como indicado pela estrutura mais representativa após análise de clusters, na Figura 66 B. Além disso, por ser uma cadeia de apenas 23 aminoácidos, as cadeias laterais estão sistematicamente mais expostas ao solvente do que na estrutura nativa (Figura 65 B). Calculando o RMSD dos átomos do *backbone* do peptídeo em relação à estrutura nativa, ao longo da simulação, pode-se observar que a parte N-terminal (resíduos 1 a 12) se comporta de maneira diferente da parte C-terminal (resíduos 13 a 23) (Figura 67). Enquanto a região N-terminal possui 3 populações igualmente distribuídas com RMSD em torno de 0.2, 0.3 e 0.4 nm, a região C-terminal majoritariamente possui RMSD de 0.2 nm. Corroborando a nossa hipótese, sabe-se que os resíduos da região C-terminal (T156, H158, E159 e K166) estão envolvidos num epítipo conformacional reconhecido por anticorpos anti-ZIKV (Wang *et al.*, 2016).

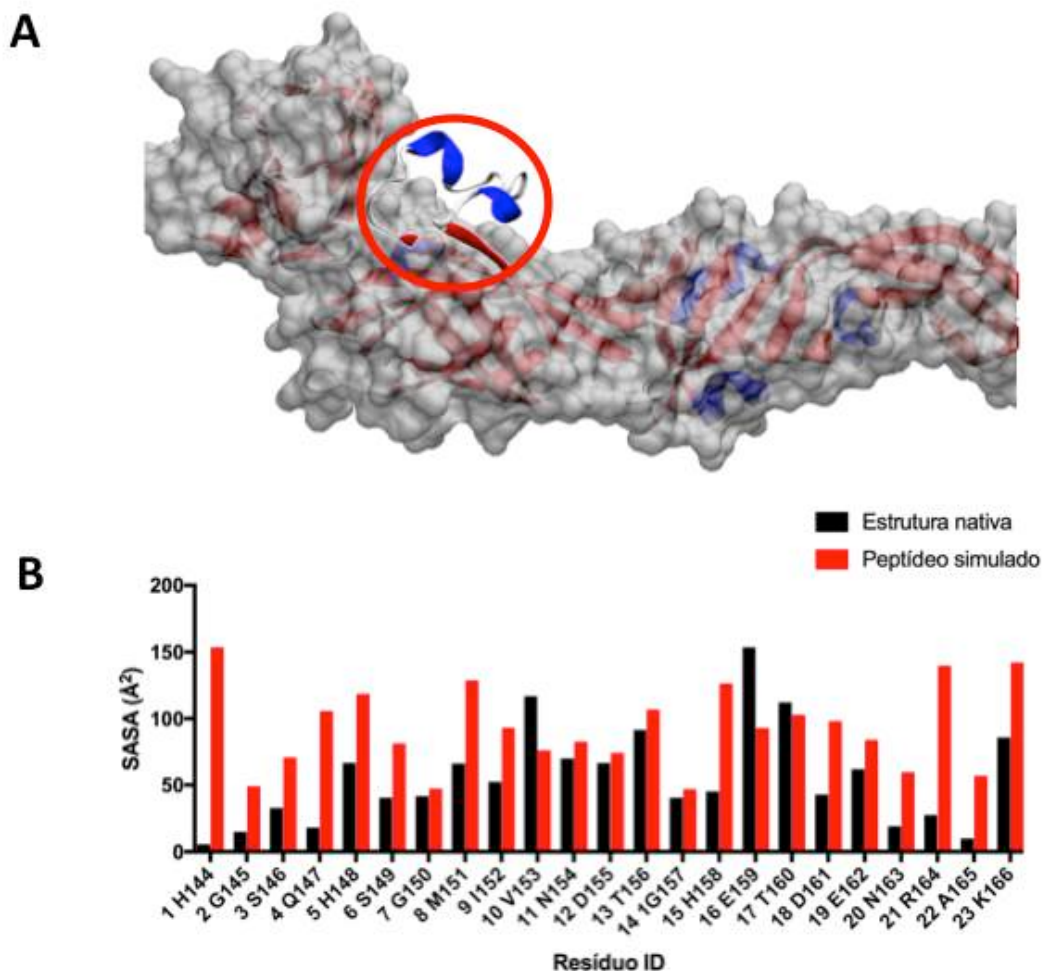


Figura 65 - (A) Representação do motivo estrutural da sequência do peptídeo E1 selecionada na estrutura nativa da proteína E de ZIKV. **(B)** SASA de cada resíduo do peptídeo na estrutura nativa (barras pretas) em comparação com a acessibilidade ao longo da trajetória de MD (barras vermelhas), indicando a maior acessibilidade no peptídeo isolado.

Das análises podemos concluir que os aminoácidos estão mais expostos ao solvente no peptídeo e que há uma preferência por conformações não-estruturadas, assim como na estrutura nativa da região. Além disso, há uma população com baixo valor de RMSD, em torno de 0.2 nm em relação à estrutura nativa na região C-terminal, região esta que está envolvida em reconhecimento de anticorpos. Portanto, concluímos que o peptídeo proposto pode ser utilizado para obtenção de anticorpos anti-E de ZIKV. Tais anticorpos serão utilizados para o desenvolvimento de um novo ensaio de ELISA sanduíche a fim de atestar a eficácia das 4 vacinas desenvolvidas por Lopez-Camacho e colaboradores.

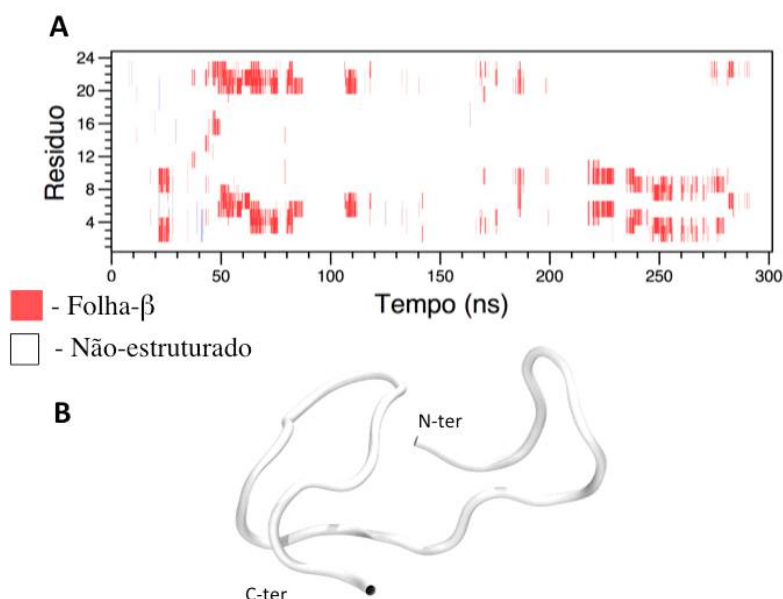


Figura 66 - (A) O conteúdo de estrutura secundária durante a simulação indica uma transição entre conformações não-estruturadas e conformações em β -*hairpin* (B) com preferência por conformações sem estrutura definida, como indicado pela estrutura mais representativa após análise de clusters.

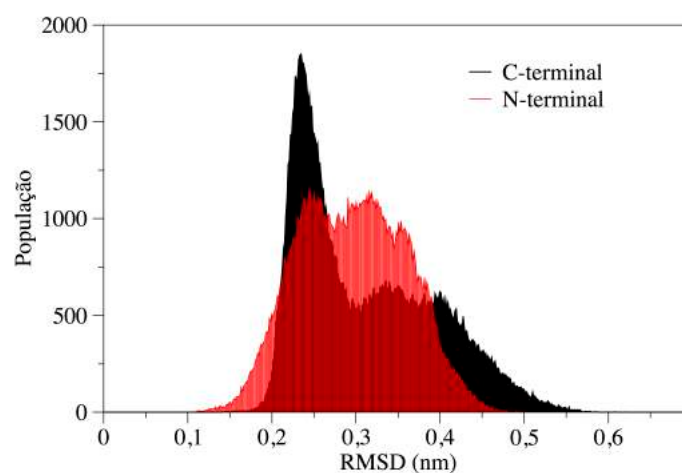


Figura 67 - Análise de RMSD do peptídeo E3 de ZIKV ao longo da trajetória de MD. A região N-terminal possui 3 populações igualmente distribuídas com RMSD em torno de 0.2, 0.3 e 0.4 nm, enquanto que a região C-terminal majoritariamente possui RMSD de 0.2 nm.

Colaboradores da Universidade de Glasgow utilizaram o peptídeo E1 para imunizar dois coelhos, dos quais foram obtidos soros policlonais anti-E ZIKV. Os dois soros policlonais, R27 e R34, foram testados por ELISA (dados

não disponibilizados) e *western blot* e reconheceram especificamente a proteína E de ZIKV (Figura 68). O soro R34 foi escolhido para purificação de IgG total por cromatografia de afinidade com proteína G.

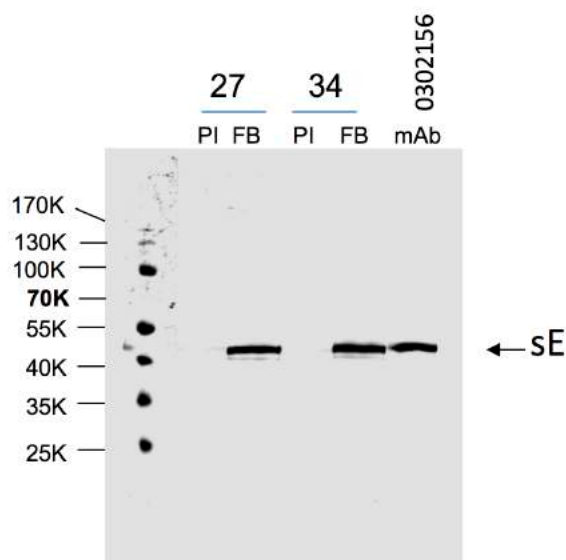


Figura 68 - Dois coelhos (R27 e R34) foram imunizados e os seus soros testados por *western blot* para atestar especificidade na ligação à proteína E de ZIKV (sE). Como esperado, os soros retirados dos animais antes da imunização (pré-imune PI) não reconhecem o envelope. O anticorpo monoclonal mAb 0302156 foi usado como controle positivo. Dados obtidos por colaboradores da Universidade de Glasgow, Escócia e publicados em (López-Camacho *et al.*, 2018).

Os anticorpos policlonais R34 foram então utilizados como anticorpos de detecção num ensaio de ELISA sanduíche acoplado a um ensaio de MN para quantificar os níveis de proteína E em células infectadas *in vitro* por ZIKV na presença de anticorpos de camundongo induzidos pelas 4 vacinas desenvolvidas por Lopez-Camacho e colaboradores (López-Camacho *et al.*, 2018). A quantidade de proteína E correlaciona-se com o nível de infectividade do vírus, ou seja, quanto menor a quantidade de proteína E, menos o vírus se replicou e maior o poder de neutralização do anticorpo induzido pela vacina. Os ensaios também foram realizados por nossos colaboradores na Universidade de Glasgow na Escócia.

Através do ensaio de MN acoplado ao ELISA sanduíche, foi possível estabelecer que as vacinas prME e prME Δ TM mostraram atividade de

neutralização, enquanto que as vacinas Env e Env Δ TM demonstraram baixo ou nenhum poder de neutralização (Figura 69 A e B).

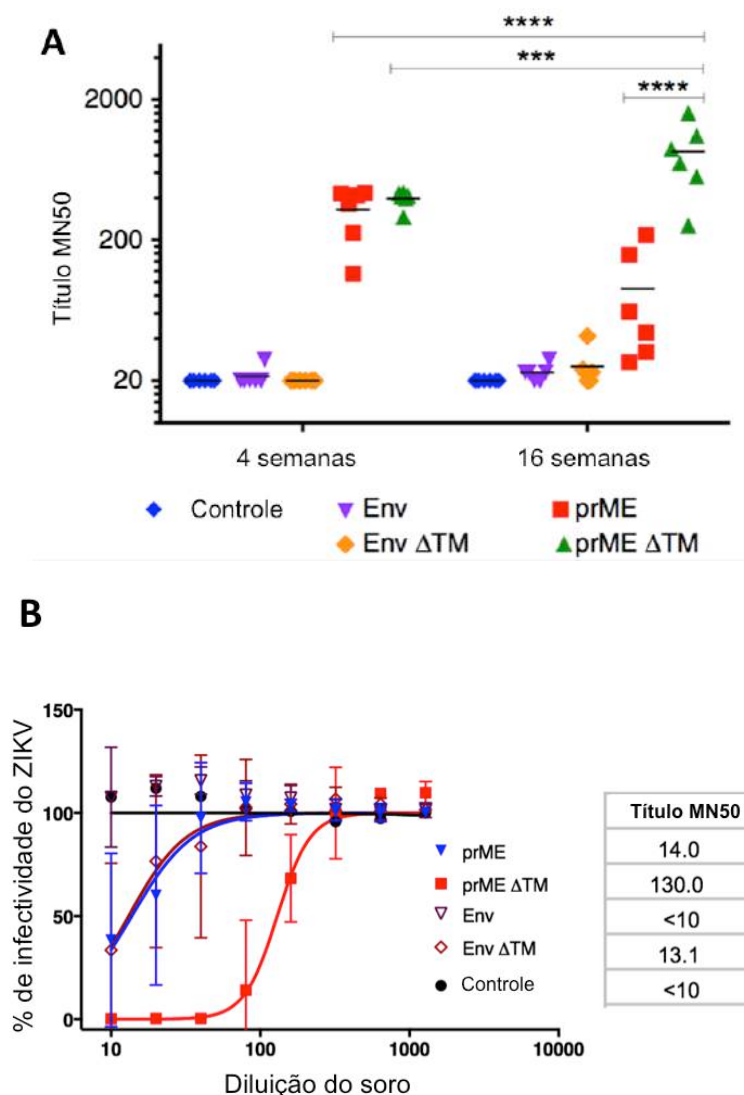


Figura 69 - (A) Ensaio de MN do ZIKV em camundongos BALB/c 4 semanas e 16 semanas após vacinação. As vacinas prME e prME Δ TM mostraram atividade de neutralização, enquanto que as vacinas Env e Env Δ TM não demonstraram nenhum poder de neutralização. Os resultados mostram a média de três repetições biológicas, com duplicatas e linhas pretas representando a média ($n = 6$). Análise ANOVA de dois fatores seguida pela comparação múltipla de Sidak foi realizada. **** $p < 0,0001$, *** $p < 0,0003$. **(B)** A neutralização do vírus também foi avaliada em camundongos CD1, após 4 semanas da vacinação, mostrando novamente que a vacina prME Δ TM é capaz de neutralizar o ZIKV. Adaptado de (López-Camacho *et al.*, 2018).

Os resultados indicam um efeito benéfico da inclusão da região prM como parte de uma vacina e destacam a importância de eliminar a região TM

do envelope para manter boas respostas de memória. Tais resultados são consistentes com dados obtidos através de outros ensaios realizados por López-Camacho e colaboradores (ELISA indireto comercial, RT-qPCR, ELISPOT) que já indicavam a maior eficiência da vacina prME Δ TM (López-Camacho *et al.*, 2018).

O ensaio de MN acoplado ao ELISA sanduíche, se mostrou bastante robusto, com uma relação sinal-ruído >10 e um fator Z de 0.8 (López-Camacho *et al.*, 2018), um valor ideal para ensaios de triagem de alto rendimento (Zhang *et al.*, 1999). Para certificar a eficácia do ensaio, nossos colaboradores realizaram uma comparação entre o ensaio desenvolvido e um ensaio PRNT tradicional (Poggianella *et al.*, 2015), demonstrando que o ensaio acoplado é mais rigoroso na detecção da infecção pelo ZIKV do que o PRNT. Dessa forma, o ensaio de MN acoplado ao ELISA sanduíche usando o soro R34 específico do peptídeo E1 é altamente informativo para selecionar candidatos a vacina contra ZIKV de forma eficiente na fase pré-clínica.

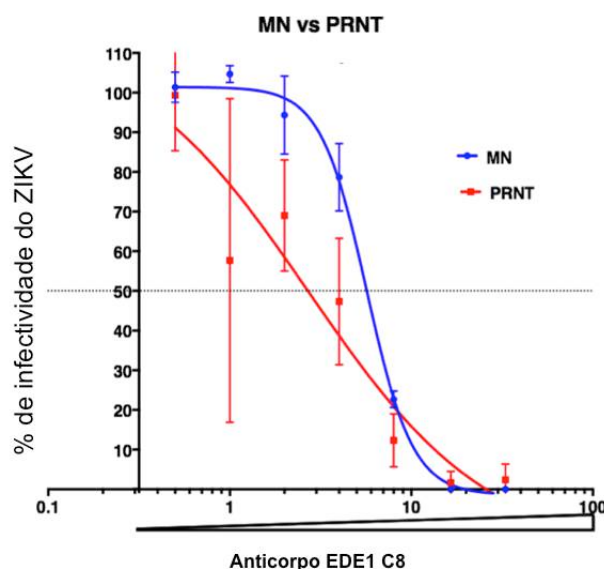


Figura 70 - Comparação da % de infectividade de ZIKV entre nosso ensaio de MN e o PRNT tradicional em células Vero. O anticorpo monoclonal altamente neutralizante EDE1 C8 foi utilizado em diferentes concentrações. As linhas representam o ajuste não-linear. Barras de erro são o SD da média de três amostras de réplicas. O ensaio de MN acoplado ao ELISA sanduíche desenvolvidos se mostraram bastante robustos, com uma relação sinal-ruído >10 e um fator Z de 0.8. Adaptado de (López-Camacho *et al.*, 2018).

5.4 CONCLUSÕES

Através da metodologia de *MotifGraft*, seguida de análises de dinâmica molecular, foi possível realizar uma triagem numa biblioteca de milhares de proteínas candidatas a carreadora para os epítomos putativos de ZIKV identificados por métodos de alinhamento. Três proteínas quiméricas contendo epítomos das proteínas E e NS4b de ZIKV foram desenhadas *in silico*, utilizando os scaffold FNIII e TL1A. Tais proteínas estão sendo obtidas por meio de expressão em meio bacteriano, para posterior avaliação imunológica.

Foi também proposto um peptídeo baseado na proteína E de ZIKV e de baixa identidade com regiões homólogas do demais flavivírus, utilizado para o desenvolvimento de um ensaio de micro neutralização acoplado a ELISA sanduíche. De acordo com simulações de MD, o peptídeo tem preferência por conformações sem estrutura definida, mantendo as cadeias laterais expostas ao solvente. Análise de RMSD indica que a parte C-terminal do peptídeo, sabidamente parte de um epítomo conformacional, conserva conformação nativa, possibilitando reconhecimento por anticorpos anti-E-ZIKV. O peptídeo foi capaz de elicitar anticorpos policlonais de coelho específicos de proteína E de ZIKV que foram utilizados como anticorpos de detecção num ensaio de ELISA sanduíche acoplado ao um ensaio de microneutralização. O ensaio desenvolvido obteve resultados comparáveis aos ensaios tradicionais da literatura, sendo uma ferramenta altamente informativa para seleção de candidatos a vacina contra ZIKV na fase pré-clínica.

6 CONSIDERAÇÕES FINAIS

Ao longo da trajetória desta tese, o foco principal foi aplicar metodologias teórico-computacional e experimental para criar e caracterizar moléculas com aplicações biológicas. Foram utilizados conceitos de engenharia de proteína, como *threading*, *de novo design* e *MotifGraft*. A dinâmica molecular clássica foi utilizada para refinar os resultados obtidos com as ferramentas de desenho de antígenos. Em conjunto com a validação experimental, essas ferramentas se mostraram uma abordagem robusta para triagem no desenvolvimento de novos antígenos, evitando consumo demasiado de insumos laboratoriais.

Os resultados obtidos mostram que fomos capazes de transplantar epítopos de dengue em proteínas carreadoras que reagem com anticorpos humanos, conseguindo reproduzir resultados experimentais da literatura. Através da dinâmica molecular e de dados de dicroísmo circular fomos capazes de estabelecer as bases moleculares que conferem à proteína Top7-2F5 estabilidade, menor taxa de agregação e precipitação. Também foi possível indicar a possível causa da menor afinidade da Top7-2F5-Biotina pelo mAb.

Por fim, fomos capazes de desenhar proteínas contendo epítopos de ZIKV com potencial para uso *in vitro*. Além disso, foi identificado um peptídeo capaz de induzir anticorpos policlonais específicos para a proteína do envelope de ZIKV. Por sua vez, esses anticorpos possibilitaram o desenvolvimento de um novo ensaio de ELISA sanduíche, constituindo uma abordagem para avaliação de candidatos a vacina contra ZIKV na fase pré-clínica.

REFERÊNCIAS

AGUIAR, M.; HALSTEAD, S. B.; STOLLENWERK, N. Consider stopping dengvaxia administration without immunological screening. **Expert Review of Vaccines**, v. 16, n. 4, p. 301-302, 2017/04/03 2017.

AGUIAR, M.; STOLLENWERK, N.; HALSTEAD, S. B. The Impact of the Newly Licensed Dengue Vaccine in Endemic Countries. **PLOS Neglected Tropical Diseases**, v. 10, n. 12, p. e0005179, 2016.

ALFORD, R. F. et al. The Rosetta All-Atom Energy Function for Macromolecular Modeling and Design. **Journal of Chemical Theory and Computation**, v. 13, n. 6, p. 3031-3048, 2017/06/13 2017.

ANFINSEN, C. B. Principles that Govern the Folding of Protein Chains. **Science**, v. 181, n. 4096, p. 223, 1973.

BAKER, D.; SALI, A. Protein Structure Prediction and Structural Genomics. **Science**, v. 294, n. 5540, p. 93, 2001.

BARBA-SPAETH, G. et al. Structural basis of potent Zika–dengue virus antibody cross-neutralization. **Nature**, v. 536, p. 48, 06/23/online 2016.

BARZON, L. et al. Virus and Antibody Dynamics in Travelers With Acute Zika Virus Infection. **Clinical Infectious Diseases**, v. 66, n. 8, p. 1173-1180, 2018.

BASHFORD, D.; GERWERT, K. Electrostatic calculations of the pKa values of ionizable groups in bacteriorhodopsin. **Journal of Molecular Biology**, v. 224, n. 2, p. 473-486, 1992/03/20/ 1992.

BAYER, E. A.; WILCHEK, M. Protein biotinylation. In: WILCHEK, M. e BAYER, E. A. (Ed.). **Methods in Enzymology**: Academic Press, v.184, 1990. p.138-160. ISBN 0076-6879.

BELLISSENT-FUNEL, M.-C. et al. Water Determines the Structure and Dynamics of Proteins. **Chemical Reviews**, v. 116, n. 13, p. 7673-7697, 2016/07/13 2016.

BENDER, B. J. et al. Protocols for Molecular Modeling with Rosetta3 and RosettaScripts. **Biochemistry**, v. 55, n. 34, p. 4748-4763, 2016/08/30 2016.

BERENDSEN, H. J. C. et al. Molecular dynamics with coupling to an external bath. **The Journal of Chemical Physics**, v. 81, n. 8, p. 3684, 1984.

BERENDSEN, H. J. C. et al. Interaction Models for Water in Relation to Protein Hydration. In: PULLMAN, B. (Ed.). **Intermolecular Forces**: Springer

Netherlands, v.14, 1981. cap. 21, p.331-342. (The Jerusalem Symposia on Quantum Chemistry and Biochemistry). ISBN 978-90-481-8368-5.

BHARDWAJ, G. et al. Accurate de novo design of hyperstable constrained peptides. **Nature**, v. 538, p. 329, 09/14/online 2016.

BHATT, S. et al. The global distribution and burden of dengue. **Nature**, v. 496, n. 7446, p. 504-7, Apr 25 2013.

BICK, M. J. et al. Computational design of environmental sensors for the potent opioid fentanyl. **eLife**, v. 6, p. e28909, 2017/09/19 2017.

BOMSEL, M. et al. Immunization with HIV-1 gp41 subunit virosomes induces mucosal antibodies protecting nonhuman primates against vaginal SHIV challenges. **Immunity**, v. 34, n. 2, p. 269-80, Feb 25 2011.

BOSCH, I. et al. Rapid antigen tests for dengue virus serotypes and Zika virus in patient serum. **Science Translational Medicine**, v. 9, n. 409, 2017.

BOSCHEK, C. B. et al. Engineering an ultra-stable affinity reagent based on Top7. **Protein Eng Des Sel**, v. 22, n. 5, p. 325-32, May 2009.

BRADLEY, P.; MISURA, K. M. S.; BAKER, D. Toward High-Resolution de Novo Structure Prediction for Small Proteins. **Science**, v. 309, n. 5742, p. 1868, 2005.

BRYAN, D. C.; RICHARD, A. S.; RAYMOND, F. S. Predicting Zika virus structural biology: Challenges and opportunities for intervention. **Antiviral Chemistry and Chemotherapy**, v. 24, n. 3-4, p. 118-126, 2015/08/01 2015.

BURTON, D. R. Scaffolding to build a rational vaccine design strategy. **Proceedings of the National Academy of Sciences of the United States of America**, v. 107, n. 42, p. 17859-17860, 10/11 2010.

BUSSI, G.; DONADIO, D.; PARRINELLO, M. Canonical sampling through velocity rescaling. **The Journal of Chemical Physics**, v. 126, n. 1, p. 014101, 2007.

BUTTERFIELD, G. L. et al. Evolution of a designed protein assembly encapsulating its own RNA genome. **Nature**, v. 552, p. 415, 12/13/online 2017.

CAPEDING, M. R. et al. Clinical efficacy and safety of a novel tetravalent dengue vaccine in healthy children in Asia: a phase 3, randomised, observer-masked, placebo-controlled trial. **The Lancet**, v. 384, n. 9951, p. 1358-1365, 2014.

CAROD-ARTAL, F. J. Neurological complications of Zika virus infection. **Expert Review of Anti-infective Therapy**, p. null-null, 2018.

CHEVALIER, A. et al. Massively parallel de novo protein design for targeted therapeutics. **Nature**, v. 550, p. 74, 09/27/online 2017.

CHIVERS, CLAIRE E. et al. How the biotin–streptavidin interaction was made even stronger: investigation via crystallography and a chimaeric tetramer. **Biochemical Journal**, v. 435, n. Pt 1, p. 55-63, 2011.

CHOTHIA, C.; LESK, A. M. The relation between the divergence of sequence and structure in proteins. **The EMBO Journal**, v. 5, n. 4, p. 823-826, 1986.

CHUANG, Y. C. et al. Re-evaluation of the pathogenic roles of nonstructural protein 1 and its antibodies during dengue virus infection. **J Biomed Sci**, v. 20, p. 42, 2013.

COELHO, D. F. F., M. V. F.; MARQUES, E. T. A.; LINS, R. D.; VIANA, I. F. T. **In vitro detection of B-cells producing HIV-1 2F5 antibodies by a computer designed protein**. In preparation

CORDEIRO, M. T. Laboratory diagnosis for dengue. **Revista do Instituto de Medicina Tropical de São Paulo**, v. 54, p. 10-12, 2012.

CORREIA, B. E. et al. Computational protein design using flexible backbone remodeling and resurfacing: case studies in structure-based antigen design. **J Mol Biol**, v. 405, n. 1, p. 284-97, Jan 7 2011.

CORREIA, B. E. et al. Computational design of epitope-scaffolds allows induction of antibodies specific for a poorly immunogenic HIV vaccine epitope. **Structure**, v. 18, n. 9, p. 1116-26, Sep 8 2010.

CORREIA, B. E. et al. Proof of principle for epitope-focused vaccine design. **Nature**, v. 507, n. 7491, p. 201-6, Mar 13 2014.

CRAMER, C. J. **Essentials of computational chemistry : theories and models**. 2 ed. John Wiley & Sons Ltd., 2004. 607.

CREIGHTON, T. E. Protein folding. **Biochemical Journal**, v. 270, n. 1, p. 1-16, 1990.

CUNHA, K. C. et al. Assessing protein conformational sampling and structural stability via de novo design and molecular dynamics simulations. **Biopolymers**, v. 103, n. 6, p. 351-61, Jun 2015.

DAVID, S. et al. Evaluation of 5 Commercially Available Zika Virus Immunoassays. **Emerging Infectious Disease journal**, v. 23, n. 9, p. 1577, 2017.

DE ARAÚJO, T. V. B. et al. Association between Zika virus infection and microcephaly in Brazil, January to May, 2016: preliminary report of a case-control study. **The Lancet Infectious Diseases**, v. 16, n. 12, p. 1356-1363, 2016/12/01/ 2016.

DILL, K. A.; MACCALLUM, J. L. The Protein-Folding Problem, 50 Years On. **Science**, v. 338, n. 6110, p. 1042, 2012.

DILL, K. A. et al. The protein folding problem. **Annu Rev Biophys**, v. 37, p. 289-316, 2008.

DUFFY, M. R. et al. Zika Virus Outbreak on Yap Island, Federated States of Micronesia. **New England Journal of Medicine**, v. 360, n. 24, p. 2536-2543, 2009/06/11 2009.

DUNBRACK ROLAND, L.; COHEN FRED, E. Bayesian statistical analysis of protein side - chain rotamer preferences. **Protein Science**, v. 6, n. 8, p. 1661-1681, 2008.

EDGAR, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. **Nucleic Acids Research**, v. 32, n. 5, p. 1792-1797, 2004.

FARIA, N. R. et al. Establishment and cryptic transmission of Zika virus in Brazil and the Americas. **Nature**, v. 546, p. 406, 05/24/online 2017.

FLEISHMAN, S. J. et al. RosettaScripts: A Scripting Language Interface to the Rosetta Macromolecular Modeling Suite. **PLOS ONE**, v. 6, n. 6, p. e20161, 2011.

FLOUDAS, C. A. et al. Advances in protein structure prediction and de novo protein design: A review. **Chemical Engineering Science**, v. 61, n. 3, p. 966-988, 2// 2006.

FREIRE, M. C. L. C. et al. Mapping Putative B-Cell Zika Virus NS1 Epitopes Provides Molecular Basis for Anti-NS1 Antibody Discrimination between Zika and Dengue Viruses. **ACS Omega**, v. 2, n. 7, p. 3913-3920, 2017/07/31 2017.

GAINZA, P.; NISONOFF, H. M.; DONALD, B. R. Algorithms for protein design. **Current Opinion in Structural Biology**, v. 39, p. 16-26, 2016/08/01/ 2016.

GALLAGHER, T. et al. Two Crystal Structures of the B1 Immunoglobulin-Binding Domain of Streptococcal Protein G and Comparison with NMR. **Biochemistry**, v. 33, n. 15, p. 4721-4729, 1994/04/01 1994.

GAUDINSKI, M. R. et al. Safety, tolerability, and immunogenicity of two Zika virus DNA vaccine candidates in healthy adults: randomised, open-label, phase 1 clinical trials. **The Lancet**, v. 391, n. 10120, p. 552-562, 2018.

GRANGER, D. et al. Serologic Testing for Zika Virus: Comparison of Three Zika Virus IgM ELISAs and Initial Laboratory Experiences. **Journal of Clinical Microbiology**, 2017.

GUDEX, N.; PEITSCH, M. C. SWISS-MODEL and the Swiss-Pdb Viewer: An environment for comparative protein modeling. **ELECTROPHORESIS**, v. 18, n. 15, p. 2714-2723, 1997.

GUZMAN, M. G. et al. Dengue: a continuing global threat. **Nat Rev Microbiol**, v. 8, n. 12 Suppl, p. S7-16, Dec 2010.

HALSTEAD, S. B. Dengvaxia sensitizes seronegatives to vaccine enhanced disease regardless of age. **Vaccine**, v. 35, n. 47, p. 6355-6358, 2017/11/07/ 2017.

HESS, B. et al. LINCS: A linear constraint solver for molecular simulations. **Journal of Computational Chemistry**, v. 18, n. 12, p. 1463-1472, 1997.

HESS, B. et al. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. **Journal of Chemical Theory and Computation**, v. 4, n. 3, p. 435-447, 2008/03/01 2008.

HESSELL, A. J. et al. Broadly Neutralizing Monoclonal Antibodies 2F5 and 4E10 Directed against the Human Immunodeficiency Virus Type 1 gp41 Membrane-Proximal External Region Protect against Mucosal Challenge by Simian-Human Immunodeficiency Virus SHIV(Ba-L). **Journal of Virology**, v. 84, n. 3, p. 1302-1313, 2010.

HOCKNEY, R. W. The potential calculation and some applications. In: ALDER, B.; FERNBACH, S., et al (Ed.). **Methods in Computational Physics**. New York/London: Academic Press, 1970.

HUANG, P.-S.; BOYKEN, S. E.; BAKER, D. The coming of age of de novo protein design. **Nature**, v. 537, p. 320, 09/14/online 2016.

HUMPHREY, W.; DALKE, A.; SCHULTEN, K. VMD: Visual molecular dynamics. **Journal of Molecular Graphics**, v. 14, n. 1, p. 33-38, 1996/02/01/ 1996.

HUNSPERGER, E. A. et al. Evaluation of commercially available diagnostic tests for the detection of dengue virus NS1 antigen and anti-dengue virus IgM antibody. **PLoS Negl Trop Dis**, v. 8, n. 10, p. e3171, Oct 2014.

INBIOS INTERNATIONAL, I. ZIKV Detect™ IgM Capture ELISA Kit. 2017. Disponível em: < <http://www.inbios.com/zikv-detecttm-igm-capture-elisa-kit-usa/> >. Acesso em: 04/24.

INNIS, B. L. et al. An Enzyme-Linked Immunosorbent Assay to Characterize Dengue Infections Where Dengue and Japanese Encephalitis Co-Circulate. **The American Journal of Tropical Medicine and Hygiene**, v. 40, n. 4, p. 418-427, 1989.

JOHNSON, S.; MCCORD, R. P.; ROBINSON, L. KYTE DOOLITTLE HYDROPATHY PLOT. Disponível em: <

<http://gcat.davidson.edu/DGPB/kd/kyte-doolittle.htm> >. Acesso em: 6 de maio de 2014.

JONES, D. T.; TAYLOR, W. R.; THORNTON, J. M. A new approach to protein fold recognition. **Nature**, v. 358, p. 86, 07/02/online 1992.

KABSCH, W.; SANDER, C. Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. **Biopolymers**, v. 22, n. 12, p. 2577-2637, 1983.

KAUFMANN, K. W. et al. Practically Useful: What the Rosetta Protein Modeling Suite Can Do for You. **Biochemistry**, v. 49, n. 14, p. 2987-2998, 2010/04/13 2010.

KELLOGG, E. H.; LEAVER-FAY, A.; BAKER, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. **Proteins**, v. 79, n. 3, p. 830-8, Mar 2011.

KINDHAUSER, M. K. et al. Zika: the origin and spread of a mosquito-borne virus. **Bulletin of the World Health Organization**, v. 94, n. 9, p. 675-686C, 2016.

KING, C. et al. Removing T-cell epitopes with computational protein design. **Proceedings of the National Academy of Sciences**, v. 111, n. 23, p. 8577, 2014.

KISS, G. et al. Computational Enzyme Design. **Angewandte Chemie International Edition**, v. 52, n. 22, p. 5700-5725, 2013.

KNAPP, S. et al. Thermal Unfolding of the DNA-binding Protein Sso7d from the Hyperthermophile *Sulfolobus solfataricus*. **Journal of Molecular Biology**, v. 264, n. 5, p. 1132-1144, 12/20/ 1996.

KRAMER, RYAN M. et al. Toward a Molecular Understanding of Protein Solubility: Increased Negative Surface Charge Correlates with Increased Solubility. **Biophysical Journal**, v. 102, n. 8, p. 1907-1915, 08/18/received 01/30/accepted 2012.

KUHLMAN, B.; BAKER, D. Native protein sequences are close to optimal for their structures. **Proceedings of the National Academy of Sciences**, v. 97, n. 19, p. 10383, 2000.

KUHLMAN, B. et al. Design of a novel globular protein fold with atomic-level accuracy. **Science**, v. 302, n. 5649, p. 1364-8, Nov 21 2003.

KUROSU, T. et al. Secreted complement regulatory protein clusterin interacts with dengue virus nonstructural protein 1. **Biochem Biophys Res Commun**, v. 362, n. 4, p. 1051-6, Nov 3 2007.

KWONG, P. D.; MASCOLA, J. R. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. **Immunity**, v. 37, n. 3, p. 412-25, Sep 21 2012.

KYTE, J.; DOOLITTLE, R. F. A simple method for displaying the hydropathic character of a protein. **Journal of Molecular Biology**, v. 157, n. 1, p. 105-132, 1982/05/05/ 1982.

LAROCCA, R. A. et al. Vaccine protection against Zika virus from Brazil. **Nature**, v. 536, p. 474, 06/28/online 2016.

LAZARIDIS, T.; KARPLUS, M. Effective energy function for proteins in solution. **Proteins: Structure, Function, and Bioinformatics**, v. 35, n. 2, p. 133-152, 1999.

LEACH, A. **Molecular Modelling: Principles and Applications** 2^a ed. Prentice Hall, 2001. 784.

LEAVER-FAY, A. et al. Rosetta3: An Object-Oriented Software Suite for the Simulation and Design of Macromolecules. In: JOHNSON, M. L. e BRAND, L. (Ed.). **Methods in Enzymology**: Academic Press, v.487, 2011. p.545-574. ISBN 0076-6879.

LEVY, Y.; ONUCHIC, J. N. WATER MEDIATION IN PROTEIN FOLDING AND MOLECULAR RECOGNITION. **Annual Review of Biophysics and Biomolecular Structure**, v. 35, n. 1, p. 389-415, 2006/06/01 2006.

LI, Y. et al. Secondary Structure and Membrane Topology of the Full-Length Dengue Virus NS4B in Micelles. **Angewandte Chemie International Edition**, v. 55, n. 39, p. 12068-12072, 2016.

LIU, H.; CHEN, Q. Computational protein design for given backbone: recent progresses in general method-related aspects. **Current Opinion in Structural Biology**, v. 39, p. 89-95, 2016/08/01/ 2016.

LÓPEZ-CAMACHO, C. et al. Rational Zika vaccine design via the modulation of antigen membrane anchors in chimpanzee adenoviral vectors. **Nature Communications**, v. 9, n. 1, p. 2441, 2018/06/22 2018.

MANDELL, D. J.; COUTSIAS, E. A.; KORTEEMME, T. Sub-angstrom accuracy in protein loop reconstruction by robotics-inspired conformational sampling. **Nature Methods**, v. 6, p. 551, 08/01/online 2009.

MARTÍ-RENOM, M. A. et al. Comparative Protein Structure Modeling of Genes and Genomes. **Annual Review of Biophysics and Biomolecular Structure**, v. 29, n. 1, p. 291-325, 2000/06/01 2000.

MEAD, P. S. et al. Zika Virus Shedding in Semen of Symptomatic Infected Men. **New England Journal of Medicine**, v. 378, n. 15, p. 1377-1385, 2018/04/12 2018.

METROPOLIS, N. et al. Equation of State Calculations by Fast Computing Machines. **The Journal of Chemical Physics**, v. 21, n. 6, p. 1087, 1953.

MUSTER, T. et al. A conserved neutralizing epitope on gp41 of human immunodeficiency virus type 1. **Journal of Virology**, v. 67, n. 11, p. 6642-6647, November 1, 1993 1993.

NETWORK, W. G. I. S. Serological diagnosis of influenza by microneutralization assay. In: (Ed.). **Manual for the laboratory diagnosis and virological surveillance of influenza**, 2011. cap. 2.G, p.63-78. ISBN 978 92 4 154809 0.

O'BRIEN, C. J. et al. Modulating non-native aggregation and electrostatic protein–protein interactions with computationally designed single-point mutations. **Protein Engineering, Design and Selection**, v. 29, n. 6, p. 231-243, 2016.

OFEK, G. et al. Structure and mechanistic analysis of the anti-human immunodeficiency virus type 1 antibody 2F5 in complex with its gp41 epitope. **J Virol**, v. 78, n. 19, p. 10724-37, Oct 2004.

OMOKOKO, M. D. et al. A highly conserved region between amino acids 221 and 266 of dengue virus non-structural protein 1 is a major epitope region in infected patients. **Am J Trop Med Hyg**, v. 91, n. 1, p. 146-55, Jul 2014.

OOSTENBRINK, C. et al. A biomolecular force field based on the free enthalpy of hydration and solvation: the GROMOS force-field parameter sets 53A5 and 53A6. **J Comput Chem**, v. 25, n. 13, p. 1656-76, Oct 2004.

PÁLL, S.; HESS, B. A flexible algorithm for calculating pair interactions on SIMD architectures. **Computer Physics Communications**, v. 184, n. 12, p. 2641-2650, 2013/12/01/ 2013.

PANTAZES, R. J.; GRISEWOOD, M. J.; MARANAS, C. D. Recent advances in computational protein design. **Current Opinion in Structural Biology**, v. 21, n. 4, p. 467-472, 2011/08/01/ 2011.

PARK, H. et al. Simultaneous Optimization of Biomolecular Energy Functions on Features from Small Molecules and Macromolecules. **Journal of Chemical Theory and Computation**, v. 12, n. 12, p. 6201-6212, 2016/12/13 2016.

PEELING, R. W. et al. Evaluation of diagnostic tests: dengue. **Nature Reviews Microbiology**, v. 8, n. 12, p. S30-S37, 2010.

PERCHIACCA, J. M.; LEE, C. C.; TESSIER, P. M. Optimal charged mutations in the complementarity-determining regions that prevent domain antibody aggregation are dependent on the antibody scaffold. **Protein Engineering, Design and Selection**, v. 27, n. 2, p. 29-39, 2014.

PETERSEN, L. R. et al. Zika Virus. **New England Journal of Medicine**, v. 374, n. 16, p. 1552-1563, 2016/04/21 2016.

POGGIANELLA, M. et al. Dengue E Protein Domain III-Based DNA Immunisation Induces Strong Antibody Responses to All Four Viral Serotypes. **PLOS Neglected Tropical Diseases**, v. 9, n. 7, p. e0003947, 2015.

PRESS, W. H. et al. Minimization or maximization of functions. In: (Ed.). **Numerical Recipes in FORTRAN 77: The art of scientific computing**. 2 ed. Cambridge: Cambridge University Press, 1992. cap. 10, p.387-448.

PRIYAMVADA, L. et al. Humoral cross-reactivity between Zika and dengue viruses: implications for protection and pathology. **Emerging Microbes & Infections**, v. 6, p. e33, 05/10/online 2017.

RAMAN, S. et al. Structure prediction for CASP8 with all-atom refinement using Rosetta. **Proteins**, v. 77 Suppl 9, p. 89-99, 2009.

ROCKSTROH, A. et al. Specific detection of dengue and Zika virus antibodies using envelope proteins with mutations in the conserved fusion loop. **Emerging Microbes & Infections**, v. 6, p. e99, 11/08/online 2017.

ROHL, C. A. et al. Protein Structure Prediction Using Rosetta. In: (Ed.). **Methods in Enzymology**: Academic Press, v.383, 2004. p.66-93. ISBN 0076-6879.

ROWE, T. et al. Detection of Antibody to Avian Influenza A (H5N1) Virus in Human Serum by Using a Combination of Serologic Assays. **Journal of Clinical Microbiology**, v. 37, n. 4, p. 937-943, 1999.

SAFRIT, J. T. et al. Status of vaccine research and development of vaccines for HIV-1. **Vaccine**, v. 34, n. 26, p. 2921-2925, Jun 3 2016.

SANTIAGO, G. A. et al. Performance of the Triplex real-time RT-PCR assay for detection of Zika, dengue, and chikungunya viruses. **Nature Communications**, v. 9, n. 1, p. 1391, 2018/04/11 2018.

SCHMID, N. et al. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. **European Biophysics Journal**, v. 40, n. 7, p. 843-856, 2011/07/01 2011.

SERRANO, S. et al. Structure and immunogenicity of a peptide vaccine, including the complete HIV-1 gp41 2F5 epitope: implications for antibody recognition mechanism and immunogen design. **J Biol Chem**, v. 289, n. 10, p. 6565-80, Mar 7 2014.

SHAPOVALOV, MAXIM V.; DUNBRACK, ROLAND L., JR. A Smoothed Backbone-Dependent Rotamer Library for Proteins Derived from Adaptive

Kernel Density Estimates and Regressions. **Structure**, v. 19, n. 6, p. 844-858, 2011.

SILVA, D.-A.; CORREIA, B. E.; PROCKO, E. Motif-Driven Design of Protein–Protein Interfaces. In: STODDARD, B. L. (Ed.). **Computational Design of Ligand Binding Proteins**. New York, NY: Springer New York, 2016. p.285-304. ISBN 978-1-4939-3569-7.

SINHA, N. et al. Differences in Electrostatic Properties at Antibody–Antigen Binding Sites: Implications for Specificity and Cross-Reactivity. **Biophysical Journal**, v. 83, n. 6, p. 2946-2968, 2002/12/01/ 2002.

SIROHI, D. et al. The 3.8 Å resolution cryo-EM structure of Zika virus. **Science**, v. 352, n. 6284, p. 467-70, Apr 22 2016.

SOARES, T. A. et al. Molecular basis of the structural stability of a Top7-based scaffold at extreme pH and temperature conditions. **Journal of Molecular Graphics and Modelling**, v. 28, n. 8, p. 755-765, 2010/06/01/ 2010.

STEFANKOVA, P.; KOLLAROVA M FAU - BARAK, I.; BARAK, I. Thioredoxin - structural and functional complexity. **General Physiology and Biophysics**, v. 24, p. 3-11, 2005.

TEBAS, P. et al. Safety and Immunogenicity of an Anti–Zika Virus DNA Vaccine — Preliminary Report. **New England Journal of Medicine**, 2017.

TIRONI, I. G. et al. A generalized reaction field method for molecular dynamics simulations. **The Journal of Chemical Physics**, v. 102, n. 13, p. 5451, 1995.

VAN GUNSTEREN, W. F. et al. Biomolecular Modeling: Goals, Problems, Perspectives. **Angewandte Chemie International Edition**, v. 45, n. 25, p. 4064-4092, 2006.

VAN-MAI, C.-L. et al. Zika Virus, French Polynesia, South Pacific, 2013. **Emerging Infectious Disease journal**, v. 20, n. 6, p. 1084, 2014.

VERLI, H. **Bioinformática : da biologia à flexibilidade molecular**. 1. Porto Alegre: Sociedade Brasileira de Bioquímica e Biologia Molecular, 2014. 282.

VIANA, I. et al. **Influence of Scaffold Stability and Electrostatics on Top7-Based Engineered Helical HIV-1 Epitopes**. 2013. 94-103.

VIANA, I. et al. De novo design of immunoreactive conformation- specific HIV-1 epitopes based on Top7 scaffold. **RSC Advances**, v. 3, n. 29, p. 11790-11800, 2013.

VIANA, I. F. T. **Desenho in silico e caracterização estrutural e imunológica de proteínas sintéticas carreadoras de epítopos**

específicos de HIV-1. 2014. 122 Doutorado (Doutorado). Pós-Graduação em Biociências e Biotecnologia, Instituto Carlos Chagas, Curitiba.

VILLAR, L. et al. Efficacy of a tetravalent dengue vaccine in children in Latin America. **N Engl J Med**, v. 372, n. 2, p. 113-23, Jan 8 2015.

WANG, Q. et al. Molecular determinants of human neutralizing antibodies isolated from a patient infected with Zika virus. **Science Translational Medicine**, v. 8, n. 369, p. 369ra179, 2016.

WATERHOUSE, A. M. et al. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. **Bioinformatics**, v. 25, n. 9, p. 1189-1191, 2009.

WEITZNER, B. D. et al. Modeling and docking of antibody structures with Rosetta. **Nature Protocols**, v. 12, p. 401, 01/26/online 2017.

WHITEHEAD, S. S. Development of TV003/TV005, a single dose, highly immunogenic live attenuated dengue vaccine; what makes this vaccine different from the Sanofi-Pasteur CYD™ vaccine? **Expert Review of Vaccines**, v. 15, n. 4, p. 509-517, 2016/04/02 2016.

ZHAN, C. et al. Biochemical and Structural Characterization of the Human TL1A Ectodomain. **Biochemistry**, v. 48, n. 32, p. 7636-7645, 2009/08/18 2009.

ZHANG, J.-H.; CHUNG, T. D. Y.; OLDENBURG, K. R. A Simple Statistical Parameter for Use in Evaluation and Validation of High Throughput Screening Assays. **Journal of Biomolecular Screening**, v. 4, n. 2, p. 67-73, 1999/04/01 1999.

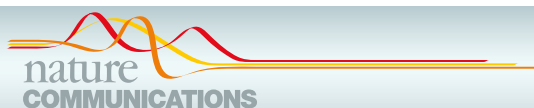
APÊNDICE A - ARTIGO PUBLICADO EM PERIÓDICO INDEXADO DE CIRCULAÇÃO INTERNACIONAL REFERENTE AO TRABALHO DA TESE

Título: Rational Zika vaccine design via the modulation of antigen membrane anchors in chimpanzee adenoviral vectors

Autores: César López-Camacho, Peter Abbink, Rafael A. Larocca, Wanwisa Dejnirattisai, Michael Boyd, Alex Badamchi-Zadeh, Zoë R. Wallace, Jennifer Doig, Ricardo Sanchez Velazquez, **Roberto Dias Lins Neto, Danilo F. Coelho**, Young Chan Kim, Claire L. Donald, Ania Owsianka, Giuditta De Lorenzo, Alain Kohl, Sarah C. Gilbert, Lucy Dorrell, Juthathip Mongkolsapaya, Arvind H. Patel, Gavin R. Screaton, Dan H. Barouch, Adrian V.S. Hill, Arturo Reyes-Sandoval.

Publicado no periódico: Nature Communications, 2018.

doi: 10.1038/s41467-018-04859-5



ARTICLE

DOI: 10.1038/s41467-018-04859-5

OPEN

Rational Zika vaccine design via the modulation of antigen membrane anchors in chimpanzee adenoviral vectors

César López-Camacho¹, Peter Abbink², Rafael A. Larocca², Wanwisa Dejnirattisai³, Michael Boyd², Alex Badamchi-Zadeh², Zoë R. Wallace⁴, Jennifer Doig⁵, Ricardo Sanchez Velazquez⁵, Roberto Dias Lins Neto⁶, Danilo F. Coelho⁶, Young Chan Kim¹, Claire L. Donald⁵, Ania Owsianka⁵, Giuditta De Lorenzo⁵, Alain Kohl⁵, Sarah C. Gilbert⁷, Lucy Dorrell⁴, Juthathip Mongkolsapaya^{3,8}, Arvind H. Patel⁵, Gavin R. Screaton⁹, Dan H. Barouch², Adrian V.S. Hill⁷ & Arturo Reyes-Sandoval¹

Zika virus (ZIKV) emerged on a global scale and no licensed vaccine ensures long-lasting anti-ZIKV immunity. Here we report the design and comparative evaluation of four replication-deficient chimpanzee adenoviral (ChAdOx1) ZIKV vaccine candidates comprising the addition or deletion of precursor membrane (prM) and envelope, with or without its transmembrane domain (TM). A single, non-adjuvanted vaccination of ChAdOx1 ZIKV vaccines elicits suitable levels of protective responses in mice challenged with ZIKV. ChAdOx1 prME ΔTM encoding prM and envelope without TM provides 100% protection, as well as long-lasting anti-envelope immune responses and no evidence of in vitro antibody-dependent enhancement to dengue virus. Deletion of prM and addition of TM reduces protective efficacy and yields lower anti-envelope responses. Our finding that immunity against ZIKV can be enhanced by modulating antigen membrane anchoring highlights important parameters in the design of viral vectored ZIKV vaccines to support further clinical assessments.

¹The Jenner Institute, Nuffield Department of Medicine, University of Oxford, The Henry Wellcome Building for Molecular Physiology, Roosevelt Drive, Oxford OX3 7BN, UK. ²Center for Virology and Vaccine Research, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA. ³Division of Immunology and Inflammation, Department of Medicine, Hammersmith Campus, Imperial College London, London W12 0NN, UK. ⁴Nuffield Department of Medicine and Oxford NIHR Biomedical Research Centre, University of Oxford, Oxford OX3 7FZ, UK. ⁵MRC-University of Glasgow Centre for Virus Research, University of Glasgow, Glasgow, G61 1QH Scotland, UK. ⁶Aggeu Magalhães Institute, Oswaldo Cruz Foundation, 50670-465 Recife, Brazil. ⁷The Jenner Institute, Nuffield Department of Medicine, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, UK. ⁸Dengue Hemorrhagic Fever Research Unit, Office for Research and Development, Faculty of Medicine Siriraj Hospital, Mahidol University, Bangkok 10700, Thailand. ⁹Division of Medical Sciences, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, UK. Correspondence and requests for materials should be addressed to A.R.-S. (email: arturo.reyes@ndm.ox.ac.uk)

Zika virus (ZIKV) has caused a remarkably explosive outbreak in the Americas, with rapid spread to more than 70 countries¹. Its association of ZIKV with Guillain-Barré syndrome², microcephaly in newborns³ and person-to-person transmission⁴ underlines the need for an efficacious vaccine that provides long-lasting anti-ZIKV immunity. Currently, no licensed Zika vaccines or antivirals are available to prevent or treat infection. However, exceptional progress has been made since the World Health Organization declared ZIKV a public health emergency, notably the description of the ZIKV atomic structure^{5,6}, the expansion of sequenced genomes of several ZIKV isolates and in particular the development of pre-clinical ZIKV challenge models in mice^{7,8} and in non-human primates^{9,10}. These models have proven useful to explore efficacy of many ZIKV vaccine developments and many vaccine platforms such as subunit envelope protein, virus-like particles (VLP), live attenuated virus, inactivated virus, naked DNA vaccines, liposome-encapsulated RNA vaccines and viral vectored-based vaccines^{11–15}. Equally important, attempts to optimise antigen secretion and presentation to immune cells include (1) addition or replacement of ZIKV envelope transmembrane domain to that of other flavivirus, (2) a great variety of different signal sequences to improve antigen secretion and (3) addition or replacement of ZIKV prM for those of other flaviviruses. However, it is unknown to which extent these subtle and yet non-unified modifications may play a role at maintaining long-lasting anti-ZIKV immunity.

In the present study we utilise the clinically validated replication-deficient chimpanzee adenovirus vector (ChAdOx1) as vaccine platform to express the *pre-membrane* (prM) and *envelope* genes of ZIKV. We demonstrate that immune responses vary upon the modulation of membrane anchors, in particular the transmembrane (TM) domain of the envelope and prM. The ChAdOx1 ZIKV vaccine candidates substantially reduce levels of viraemia in a challenge with a Brazilian ZIKV isolate in mice, with highest efficacy resulting in viral vectors expressing the ZIKV prM and envelope with deletion of the TM domain. Importantly, no evidence of *in vitro* antibody-dependent enhancement (ADE) to dengue is identified. Owing to the good safety and immunogenicity profile of the ChAdOx1 platform in humans and their suitability for high-scale production under Good Manufacturing Practices (GMP), the ChAdOx1 ZIKV vaccine is a robust candidate for further clinical assessment.

Results

Antigen design in DNA and ChAdOx1 ZIKV vaccines. Because the Asian lineage of ZIKV is circulating in the Americas, efforts have focused on developing an Asian lineage-based vaccine^{7,9,10}. By November 2015, before the World Health Organization declared ZIKV an emerging threat, only 5 complete genomes of Asian lineage were available (Fig. 1a). A year later, at least 47 Asian ZIKV sequences were deposited in the GenBank (Supplementary Fig. 1). Here we used a consensus-based approach to design cassettes for DNA vaccines and chimpanzee adenoviral vectored vaccines carrying Asian ZIKV sequences. Percentage identity of our consensus sequence versus ZIKV genomes was calculated (Fig. 1b) and a gene cassette was synthesised containing the prM followed by the envelope (E) transgene (prME) using the Asian consensus sequence (Fig. 1c). We designed modified versions of this cassette comprising envelope with inclusion or deletion of upstream prM (prME or E, respectively) and addition or deletion of the nucleotides encoding the envelope C-terminal transmembrane domain, a.a. 729–794 (Δ TM) of ZIKV (prME Δ TM or Env Δ TM; Fig. 1c); and cloned them into DNA vaccine plasmids and into ChAdOx1 vaccine platforms.

The expected size of the cloned cassettes was verified (Fig. 1d) and ZIKV envelope expression in mammalian cells was confirmed by western immunoblotting (Fig. 1e). We developed four DNA-based vaccines and four replication-deficient chimpanzee adenoviral vectored ZIKV vaccines (ChAdOx1 ZIKV).

ZIKV immune responses after vaccination. We performed immunogenicity assays to identify the most immunogenic candidates (Fig. 2). BALB/c mice immunised with a DNA vaccine using a homologous Prime-Boost (P-B) regimen (Fig. 2a, left) showed modest anti-ZIKV envelope antibody responses upon both prime and boost (Fig. 2b, left). In contrast, BALB/c mice immunised with a single, non-adjuvanted vaccine dose of ChAdOx1 (Fig. 2a, right) showed anti-ZIKV envelope antibodies at 2 weeks and peaked at 4 weeks (Fig. 2b, right). Interestingly, our results indicated that the concomitant inclusion of prM and Δ TM deletion elicited the highest anti-ZIKV envelope antibody titres peaking at 4 months post immunisation (Fig. 2b, right). Importantly, the durability of the anti-ZIKV antibody responses was maintained in the prME Δ TM-immunised group for up to 9 months, whilst lower antibody titres were observed in groups receiving the other vaccine candidates (Fig. 2b, right). Vaccine immunogenicity was also assessed in CD1 mice, confirming the capacity of the ChAdOx1 vaccines to induce antibody responses in outbred animals, which was sustained at high levels of antibody titres over a year, elicited by the same single and non-adjuvanted dose of ChAdOx1 prME Δ TM (Supplementary Fig. 3a).

T-cell responses were quantified by ex vivo interferon- γ (IFN γ) ELISPOT. BALB/c mice receiving DNA vaccines showed modest T-cell frequencies after stimulation with a pool of ZIKV envelope peptides both, at 2 weeks each after prime and boost, with the Env Δ TM group being the most efficient T-cell inducer (Fig. 2c, left). In contrast, a single immunisation of ChAdOx1 ZIKV vaccines induced far more robust T-cell responses than DNA vaccines as early as 2 weeks after prime and remained detectable for at least 3 months (Fig. 2c, right). Overall, T-cell responses to ChAdOx1 vaccines expressing Env, Env Δ TM and their prM counterparts were similar at week 2, with a trend by ChAdOx1 prME towards induction of lower T-cell responses. Interestingly, the prME Δ TM candidate maintained sustained responses at 2 weeks and 3 months after immunisation, indicating its ability to induce a good T-cell memory response.

Efficacy of ChAdOx1 ZIKV vaccines in a ZIKV challenge model. We used an Asian ZIKV strain to challenge mice⁷ for the assessment of efficacy of the various ChAdOx1 ZIKV vaccines. BALB/c mice were immunised with non-adjuvanted ChAdOx1 vaccines and challenged 4 weeks later with 100 plaque-forming units (PFU) of a ZIKV isolated from Brazil (ZIKV-BR). Vaccine efficacy was measured in blood by ZIKV viral load (VL) using quantitative reverse transcriptionPCR (RT-qPCR) at several time points during 7 days (Fig. 3a). Both naive and ChAdOx1 control groups (Supplementary Fig. 3b) displayed the typical onset of viraemia with a well-established peak at d3, and all infected mice were able to clear ZIKV by d7 (Fig. 3b, naive). Mice vaccinated with the rest of the vaccine candidates (prME, Env and Env Δ TM) substantially decreased VL but did not afford sterile protection to all animals in those groups. Importantly, the ChAdOx1 prME Δ TM vaccine conferred complete protection against ZIKV-BR challenge with no detectable viraemia at any time point (Fig. 3b). Additional parameters reflecting vaccine efficacy were assessed, such as an analysis of 50-fold reduction in viraemia since day 1 after ZIKV challenge and reduction of viraemia peak at day 3 (Fig. 3c). These parameters can inform on the various capacity levels of ChAdOx1 ZIKV vaccines to confer

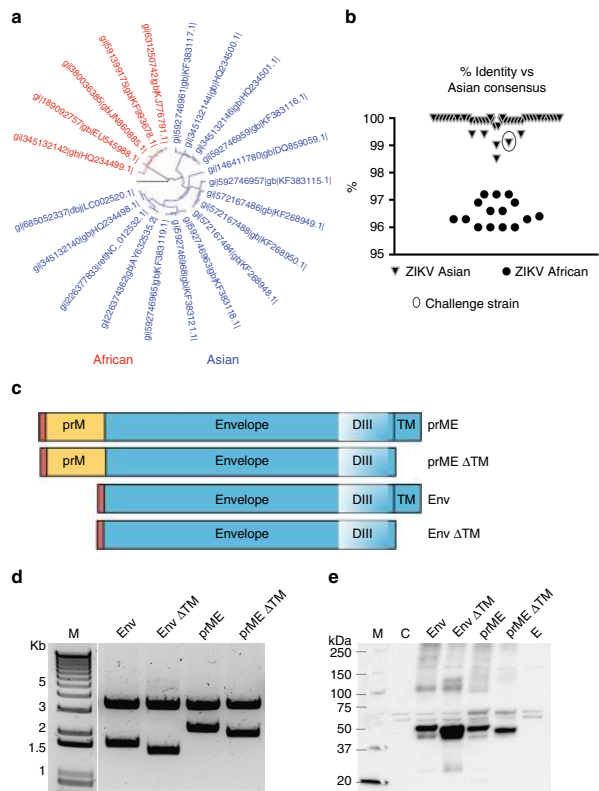


Fig. 1 Zika vaccine design. **a** A phylogenetic tree for ZIKV genomes up to November 2015; blue and red labels represent the African and Asian lineages of ZIKV, respectively. **b** Conservation homology of an Asian consensus sequence versus all genomic sequences depicted in Supplementary Fig. 1. The circle represents the ZIKV-BR strain used for the challenge experiment. **c** Schematic representation of the Zika immunogen versions used in this study; red box represents the tPA leading sequence. **d** Restriction enzyme analysis of the plasmid DNA vaccine constructs that were further cloned into ChAdOx1 vector; a 3.3 kb band size represents the DNA vaccine plasmid back bone. **e** Expression of ZIKV immunogens by western immunoblotting using an anti-ZIKV envelope antibody. M molecular marker, C control, E empty plasmid. Uncropped images are shown in Supplementary Fig. 2

protection in this challenge model, indicating that the ChAdOx1 prME Δ TM is superior to the other vaccine candidates.

As in previous reports, anti-ZIKV envelope antibodies prior to the mouse challenge were quantified by enzyme-linked immunosorbent assay (ELISA) and reciprocal titres were calculated for each vaccinated group (Supplementary Fig. 4). This provided confirmation under different laboratory conditions and in a blinded ZIKV challenge study that the deletion of TM played an important role in the production of high antibody titres able to bind the ZIKV structural antigen. A comparison of the levels of ZIKV envelope antibody titres against the ZIKV-BR VL from all animals and all vaccine candidates under the study (Fig. 3d) revealed that an endpoint titre ≥ 3 was indicative of the strongest reduction of VL after the ZIKV-BR challenge (Fig. 3d, left), whereas prME and Env (Fig. 3d, right) presented endpoint titres < 3 that were related to a partial vaccine protection, although with a substantial decrease of VL of several log fold.

ZIKV antigen secretion by membrane-anchoring components.

Although the anti-envelope antibodies elicited by ChAdOx1 prME were detectable, those induced by the counterpart vaccine ChAdOx1 prME Δ TM were significantly higher (Fig. 3e, left). Similar findings were observed in the ChAdOx1 Env Δ TM vaccine with greater antibody-binding activity to ZIKV envelope when compared to the ChAdOx1 Env (Fig. 3e, right). These data suggest that presence of the envelope transmembrane domain impairs antibody production elicited by our adenoviral vector. In an attempt to explain the differential immunogenicity among the ChAdOx1 ZIKV vaccine candidates, the subcellular localisation of the endogenously expressed vaccine antigen was explored by immunofluorescence (IF) using a monoclonal anti-flavivirus antibody 48 h after transfection (Fig. 3f). We detected differential staining with brighter positive signals across the cytoplasmic compartment in the vaccines expressing prME and Env, when compared to their counterparts prME Δ TM, that showed positive signal mainly

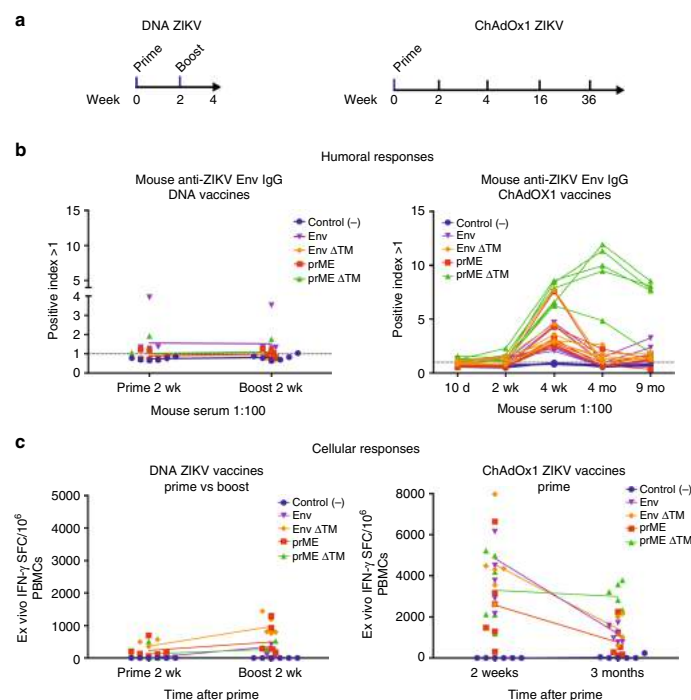


Fig. 2 Immune responses elicited by DNA versus ChAdOx1 vaccines. **a** For ZIKV DNA vaccines, BALB/c mice ($n = 6$ /group) were immunised intramuscularly (i.m.) with a dose of $100 \mu\text{g}/\text{mouse}$, followed by a DNA Boost 2 weeks thereafter. For ChAdOx1 ZIKV vaccines, a single dose of 10^8 i.u./mouse was i.m. administered. Blood samples were obtained at several time points for either ELISA or ELISPOT assay. **b** Humoral responses elicited by DNA Prime and Prime-Boost after 2 weeks (left) and by a single immunisation of ChAdOx1 ZIKV vaccines at 10 days, and 2, 4, 16 and 36 weeks (right). Antibody responses were quantified by ELISA with plates coated with ZIKV envelope protein. Positive index was calculated as depicted by ELISA kit manufacturers. **c** PBMCs-IFN γ -producing cells from DNA Prime-Boost after 2 weeks (left) and ChAdOx1 ZIKV vaccines at 2 weeks and 3 months after single immunisation were quantified by ELISPOT. 20-mer peptides spanning the ZIKV envelope protein ($10 \mu\text{g}/\text{ml}$) were used for stimulation. Line colours and shapes represent mice vaccinated with each vaccine. Antibody production by ELISA and cellular responses by ELISPOT were replicated three times in the laboratory. A positive control sample was included to validate the ELISA assay

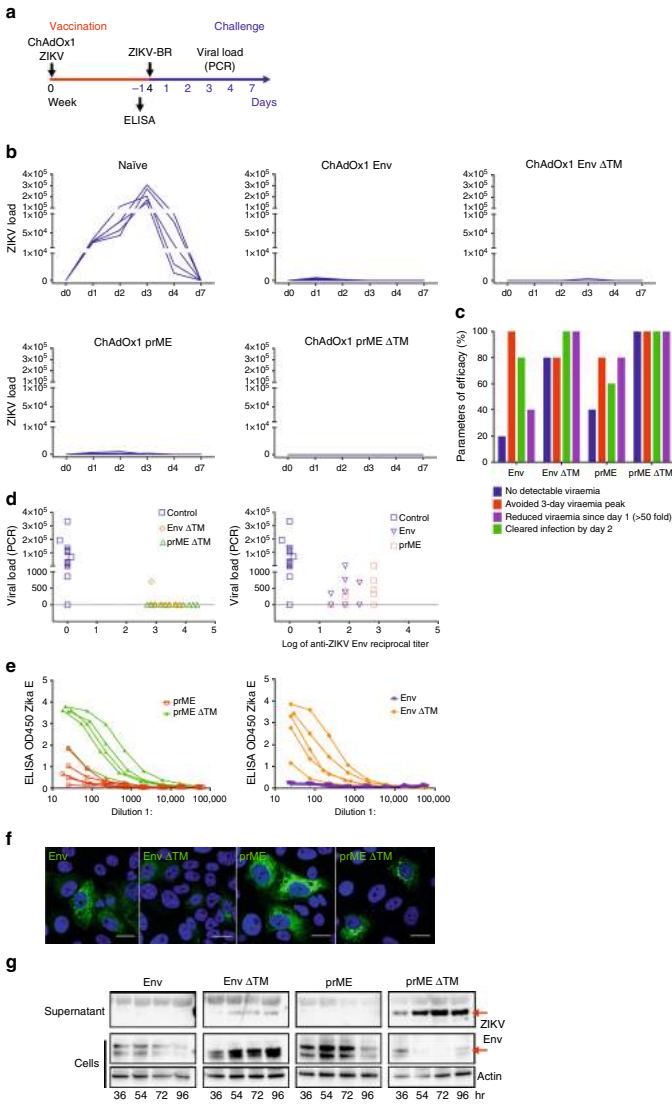
around the nucleus and with punctuate patterns across the cytoplasmic compartment and for Env ΔTM , with less bright staining around the nucleus. Furthermore, we have assessed the presence of the ZIKV envelope antigen in transfected cells in both, the intracellular compartment and the supernatant (Fig. 3g and Supplementary Figs. 5, 9 and 10). HEK-293 cells expressing prME ΔTM and Env ΔTM antigens secreted ZIKV envelope to the culture supernatant with higher efficiency than their counterparts Env and prME, with prME ΔTM secreting with highest efficiency in vitro (Fig. 3g, supernatant fraction, red arrows), while having the lowest levels of envelope in the cell fractions. For prME, we detected ZIKV envelope only in concentrated supernatant at 40 h but not at 24 h after transfection, whereas the Env construct containing the TM domain did not support secretion of the antigen to the supernatant at these time points (Supplementary Figs. 9, 11). These results suggest that antigen secretion in our adenoviral vaccine platform is modulated by membrane-anchoring components of the ZIKV structural proteins.

ZIKV-neutralising capacity in vaccinated mice sera. We tested BALB/c mice sera obtained 4 and 16 weeks after a single ChAdOx1 ZIKV vaccination to assess ZIKV antibody neutralisation activity in vitro. The method we used is based on the high-throughput microneutralisation (MN) assay described in several studies that have evaluated the efficacy of experimental vaccines against ZIKV in mice and macaques^{9,16–18}, as well as in the validated neutralisation assay used for a phase 1 clinical Zika DNA vaccine¹⁹, and recently in three phase 1 clinical trials of a purified inactivated ZIKV vaccine²⁰, with small modifications. Briefly, 50–100 PFU of the Brazilian ZIKV-PE243²¹ strain previously mixed with different threefold dilutions of mouse serum (1:20–1:43 740) are incubated with Vero cells in a 96-well format. After a 3-day incubation, cells are lysed and the levels of ZIKV envelope protein from infected cells are quantified by measurement of optical density using a highly sensitive sandwich ELISA format. For the sandwich ELISA, a rabbit antiserum R34 was developed to recognise a region between a.a. 147 and 161 that

is unique to each flavivirus and that forms a conformational exposed loop (Supplementary Fig. 6).

At 4 weeks post vaccination, sera from the ChAdOx1 prME and ChAdOx1 prME ΔTM-vaccinated group showed neutralisation activity against ZIKV-PE243 with MN50s ranging from 114 to 432 and from 292 to 434, respectively (Fig. 4a). In comparison,

at 16 weeks after the ChAdOx1 ZIKV vaccination, MN50s were surprisingly maintained at high levels in the prME ΔTM group with MN50s ranging from 253 to 1600, while the ChAdOx1 prME showed lower MN50s of 30 to 89. These results indicate a beneficial effect of including the prM region as part of an adenoviral vectored vaccine and it highlights the importance of



deleting the 3' TM region of the envelope to maintain good memory responses. Virus neutralisation was also assessed in outbred CD1 4 months after a ChAdOx1 ZIKV vaccination (Supplementary Fig. 3c), showing that ChAdOx1 prME Δ TM vaccine is able to induce MN50 values of 113 to 150, whereas ChAdOx1 prME group showed MN50 values of 10.38 to 18.23. Taken together, this confirms that Δ TM is an important factor to maintain neutralising antibody activity over longer periods of time.

Lack of evidence of ADE to dengue 2 after ZIKV vaccination.

Next, we tested whether antibodies induced by ChAdOx1 prME Δ TM vaccine cross-react with the envelope protein from all dengue virus (DENV) serotypes (DENV1-4), by ELISA. Our results indicated that ZIKV antibodies are able to bind DENV (Fig. 4b). We then addressed the possibility of ADE to DENV infection being induced by ZIKV vaccination, using a well-established *in vitro* assay for ADE to DENV2²² (Fig. 4c), as it is the most difficult serotype to protect against with the only licensed dengue vaccine^{23,24} and the most suitable serotype to address ADE *in vitro*. We assessed the ability of sera from ChAdOx1 ZIKV-vaccinated mice groups to promote ADE in the human myeloid cell line U937, which is less permissive to infection by DENV in the absence of ADE²². DENV2 was pre-incubated with a titration of mouse sera at 4 weeks after a single immunisation with ChAdOx1 ZIKV vaccines, and the DENV2-serum complex was used to infect U937 cells. As positive controls, we used anti-DENV2 antibodies that leads to *in vitro* ADE by DENV2 (Fig. 4c, top left and Supplementary Fig. 8) and as a negative control we used sera from the ChAdOx1 control vaccine (Fig. 4c, top right). None of the sera obtained from ChAdOx1-vaccinated groups induced enhancement of DENV2 infection regardless presence or absence of TM, whereas there was a 50-fold ADE increase when using an anti-DENV2 antibody (Fig. 4c, bottom left and right). These results demonstrate that antibodies raised to ZIKV in ChAdOx1 ZIKV-vaccinated mice do not promote ADE of DENV2 *in vitro* despite cross-reacting with envelope of all DENV serotypes.

Discussion

We have developed a ChAdOx1 ZIKV vaccine with the ability to provide 100% protective efficacy against a ZIKV-BR isolate. In our hands, the addition of prM and deletion of TM from ZIKV envelope are important factors in driving improved protective immunity, while promoting long-term memory responses by enhancing and maintaining antibody production as well as T-cell responses, when compared to previously reported vaccines. Our vaccine design is different from the previously published ZIKV vaccines^{7,9,10}. Larocca et al.⁷ and Abbink et al.⁹ showed that addition of the TM from envelope was important to elicit full protective efficacy in vaccines carrying a

Japanese encephalitis virus (JEV) leader sequence and a 93 a.a. deletion of pr from prM. Importantly, the impact of Δ TM in the immunogenicity and efficacy was not assessed in the adenoviral platform RhAd52-prM-Env, only in DNA-based vaccines. We found that DNA vaccines showed differential immunogenicity when compared to the adenoviral counterparts and hence the importance of constructing four adenoviral vaccine candidates and assessing their protective capabilities. Dowd et al.¹⁰ developed a DNA vaccine consisting of a JEV leader sequence followed by a full prM and the full ZIKV envelope; or with a replacement of the endogenous stem and TM from ZIKV with those from JEV, however the authors did not assess the impact of the Δ TM. Richner et al. have developed an mRNA-based vaccine in lipid nanoparticles encoding the full prME of ZIKV and using a leader sequence from the human *IgE* gene. Results from this mRNA-based vaccine show that two doses elicited strong and protective antibody responses¹⁵. Alternatively, Pardi et al.¹³ achieved neutralising antibodies in a single intradermal dose with an mRNA-based nanoparticle vaccine, this time using the full prME from ZIKV, including a leader sequence from major histocompatibility complex (MHC) class II. However, neither of the mRNA-based ZIKV vaccine studies assessed the impact of a Δ TM deletion in their candidates to study its differential immunogenicity. On the other hand, Kim et al.²⁵ have reported a protective two-dose regimen with a human adenoviral vaccine (Ad5) encoding the prME of ZIKV with a deletion of envelope transmembrane domain, being replaced with a T4 fibrin foldon trimerization domain, as well as a human leader sequence, both to improve antigen secretion.

It has been shown that an optimal leader sequence is protein-dependent²⁶ and differences observed in these studies may be also attributed to the use of alternative leader sequences (such as those from JEV, MHCII and IgE). We have used a non-viral leader sequence (tissue plasminogen activator; tPA)²⁷, which has been used safely in humans for many vectored vaccines²⁸ and which is unlikely to induce responses in humans²⁹. Regarding previous vaccine developments, it is important to take into account pre-existing humoral and cellular cross-reactive immune responses to human adenovirus; as well as when using JEV sequences to improve antigen secretion (leader sequence and/or JEV transmembrane domains), as they might have a detrimental effect on the efficient ZIKV antigen production, i.e., Asian populations exposed to or vaccinated against JEV. Moreover, different vaccine platforms such as DNA, mRNA-based vaccines, VLPs or adenoviral vectors, ZIKV strains used for the immunogen design, as well as different routes of administration may also account for different immunogenicity and efficacy levels achieved. In this study, we have used a non-lethal challenge model in BALB/c mice and we have measured vaccine efficacy by means of partial or complete reduction of viraemia by PCR. However, we do not rule out the possibility of ZIKV presence in key organs such as brain, liver, spleen, reproductive tract, etc. Further studies, such as the

Fig. 3 Assessment of protective efficacy induced by ChAdOx1 ZIKV vaccines. **a** BALB/c mice ($n = 5$) immunised with a single i.m. injection of ChAdOx1 ZIKV vaccines and naive mice were intravenously challenged with 10^5 VP of ZIKA-BR at week 4 after prime. Pre-challenge serum samples were collected for ELISA as shown. **b** Viral load in ZIKV-challenged groups was monitored for 7 days in sera to follow the onset of viraemia. Each blue line represents one mouse. **c** Assessment of various parameters of viraemia in groups vaccinated with ChAdOx1 ZIKV vaccines (100% = 6 out of 6 mice). **d** Pre-challenge reciprocal ELISA titres (Supplementary Fig. 4) were plotted against global ZIKV VL after ZIKV challenge and a comparison was made between ChAdOx1 Env Δ TM versus ChAdOx1 prME Δ TM (left) and ChAdOx1 prME versus ChAdOx1 Env (right). **e** ELISA OD 450 reads of serially diluted serum samples were plotted to assess the impact of Δ TM in antibody production. Antibody production by ELISA in pre-challenge sera was replicated two times. **f** Immunofluorescence analysis of Vero cells expressing the ZIKV immunogens (green) to assess subcellular staining and distribution. Antigen was detected using a commercial anti-flavivirus antibody. Blue (DAPI) represents the nucleus and scale bar is 20 μ m. Cell transfections were performed in technical duplicates in two biological replicates. **g** Kinetics of ZIKV envelope antigen expression in HEK-293 cells by western blot. Antigen was detected using an anti-ZIKV monoclonal antibody in both, non-concentrated supernatant and cell extracts. Red arrows, ZIKV envelope. Actin was used as a loading control. Expression of antigen was verified in three biological replicates. Uncropped blots are shown in Supplementary Fig. 5

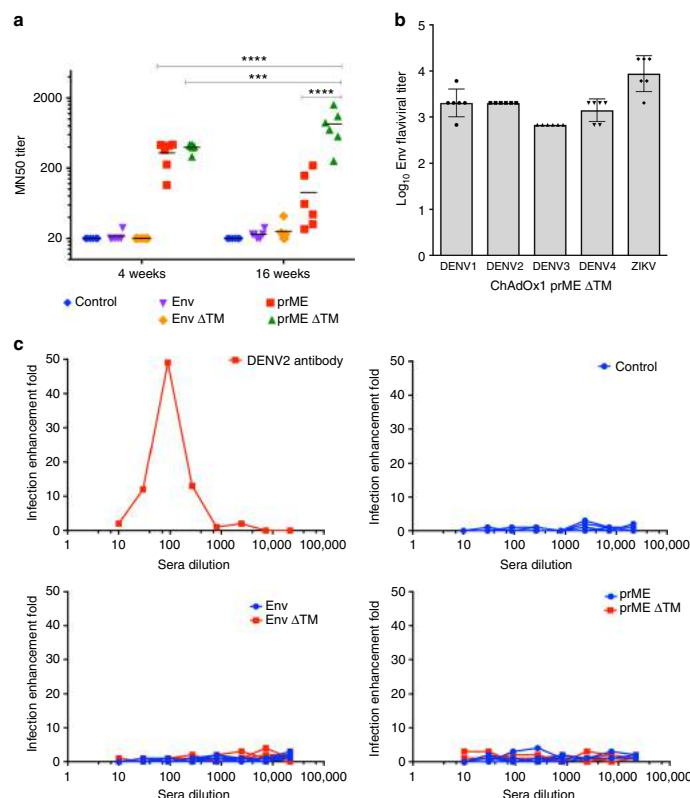


Fig. 4 ZIKV neutralisation assay and antibody-dependent enhancement (ADE) to DENV2 infection. **a** Microneutralisation titre 50 (MN50) assay in Vero cells was assessed upon incubation of ZIKV-BR in the presence of serially diluted sera from BALB/c mice vaccinated with ChAdOx1 ZIKV vaccines at 4 weeks and 4 months. Results show the average of three biological replicates, with duplicates and black lines represents the mean ($n = 6$). A two-way ANOVA followed by Sidak's Multiple comparison was performed. **** $p < 0.0001$, *** $p < 0.0003$. Supplementary Table 1 shows the multiple comparison between all the groups. **b** ELISA cross-reactivity of DENV1-4 in sera from mice vaccinated with the highly protective ChAdOx1 prME ΔTM vaccine. Bars represent the average and error is the s.d. **c** Infection of U937 cells by a DENV2 strain in the presence of sera from mice vaccinated with ChAdOx1 ZIKV candidates. Infection enhancement fold was calculated as the ratio of focus-forming units (FFU) of an anti-DENV2 monoclonal antibody to that of the FFU in the presence of serial dilutions of mouse sera collected 4 weeks after ChAdOx1 ZIKV vaccination. Each coloured line and shape represents a mouse for each group. $n = 6$ /vaccinated group. Data represent three technical replicates, with a positive control anti-DENV2 antibody

use of a highly susceptible ZIKV challenge model in A129 mice, are needed to explore ZIKV presence in those organs after a single and non-adjuvanted ChAdOx1 ZIKV vaccination. Another correlate of vaccine efficacy is assessing the ZIKV neutralisation of sera from vaccinated mice in vitro. For this we have used a MN assay, highly similar to that performed in phase I clinical trials of several Zika vaccines in the pipeline^{20,30,31}. Furthermore, MN is a validated assay used previously in several publications to assess ZIKV vaccine efficacy in mice and non-human primates^{9,16–18}, as well as other flaviviruses¹⁹. Other neutralisation assays, such as traditional PRNT50, FRNT50 and reporter viral particle assay (using green fluorescent protein, Luciferase and mCherry as reporters) have also proven to be suitable. However,

it is challenging to perform an accurate comparison of vaccine efficacy between other reported vaccine candidates, as neutralisation 50 (NT50) titres vary from one assay to another (different ZIKV strains, different ZIKV purification methodologies and different cell lines [i.e., Vero and BHK-2] and different detection methods). Nonetheless, we have performed a fair head-to-head comparison between our high-throughput MN assay versus a traditional PRNT assay. Our results demonstrate that our MN assay is more stringent at detecting ZIKV infection than PRNT (Supplementary Fig. 3d). Therefore, the MN50 values in our study are moderate when compared to PRNT. Similar observation has been reported by Dowd et al.¹⁰ in which MN50 titres are actually lower than those of a parallel FRNT50 assay.

However, MN assays together with ZIKV challenge models are both highly informative to efficiently down-select Zika vaccine candidates in the preclinical setting.

The risk of inducing ADE³², a process in which vaccine- or infection-induced cross-reactive antibodies facilitate DENV infections³³ by promoting viral entry via Fcγ receptors, adds another level of complexity in the design of a safe and efficacious ZIKV vaccine. ZIKV and DENV co-circulate in endemic regions making ADE a concern as it is not known whether antibodies elicited by ZIKV vaccines may in turn exacerbate other flaviviral infections such as dengue. A protective vaccine that does not induce ADE while providing long-lasting immunity to ZIKV would be a major breakthrough to halt the epidemics. Here we demonstrate that anti-ZIKV antibodies did not induce ADE to DENV2 upon immunisation with ChAdOx1 ZIKV vaccines in vitro. In this study, we incubated mouse sera with ZIKV and then inoculated this mixture onto human U937 cells expressing a human Fc receptor. Although we have reacted the mouse monoclonal antibody (mAb; 4G2) with DENV2 and observed the ADE phenomena in human U937 cells (Supplementary Fig. 7), it is possible that mouse Fc could present a degree of incompatibility with the human Fc receptor in U937 cells. Additional in vivo experiments should be carried out to completely rule out any enhancing effect of anti-ZIKV antibodies to DENV. In vitro ADE to ZIKA virus has been demonstrated in human sera positive for DENV that cross-reacted with ZIKV²². Richner et al.¹⁵ have also shown that a mutation of the fusion loop of ZIKV envelope decreased the risk of ADE to DENV2 in their mRNA-based vaccines in vitro. A recent study, interestingly, shows that DENV antibodies did not increase viremia and disease after a ZIKV challenge in macaques, although ADE to ZIKV was detected in vitro³⁴. Similarly, cross-reactive human mAbs by ZIKV or DENV directed primarily and cross-reactive to EDI/II can mediate heterologous ADE, while in vivo observations indicate that the same cross-reactive anti-ZIKV mAbs can lead to ADE and enhance infection by DENV, while the opposite effect for ZIKV ADE does not take place³⁵. Therefore, assessing ADE both in vitro and in vivo should be taken into account and mechanistically explored in current and ongoing ZIKV vaccine developments.

Chimpanzee adenovirus vaccine platforms such as ChAdOx1 offer a good safety profile in phase I clinical trials in humans^{36–40}; efficient biomanufacture of millions of doses of non-replicating adenoviral vectors is possible using GMP-approved cell lines. They avoid pre-existing immunity to human adenoviruses, and a single non-adjuvanted dose of ChAdOx1 is enough to achieve remarkable breadth, durability and potency of both humoral and cellular immune responses. A ChAdOx1 ZIKV vaccine engineered to positively impact immune responses over longer periods of time, by means of deletion of membrane anchors, is a valid platform to be explored in further clinical trials.

Methods

Animals. Female inbred BALB/c (H-2d) and outbred CD1 (ICR) mice (6–8 weeks) were used for the assessment of immunogenicity ($n = 6$ mice/group). ZIKV challenge was performed in BALB/c mice ($n = 5$ /group). Mice were purchased from Envigo (UK) and from Charles River (USA). The experimental design took into account the 3R reduction (replacement, reduction and refinement). No randomisation was used in this work.

Immunisation of mice. For DNA vaccines encoding the ZIKV virus antigens, mice received a prime of 100 µg of DNA vaccine/mouse in endotoxin-free phosphate-buffered saline (PBS); 2 weeks after the prime a boost injection was performed using the same concentration of DNA vaccine. For adenoviral vaccines, mice were vaccinated with a single dose of chimpanzee adenoviral vectored vaccine (ChAdOx1) encoding the ZIKV antigens at a dose of 1×10^8 infectious units. All DNA vaccines and viral vector vaccines were administered intramuscularly and diluted

in endotoxin-free PBS. Empty DNA vaccine and a ChAdOx1 expressing an unrelated malarial antigen (cCSP) were used as control vaccines.

Ethics statement. All animals and procedures were used in accordance with the terms of the UK Home Office Animals Act Project License. Procedures were approved by the University of Oxford Animal Care and Ethical Review Committee (PPL 30/2414). Animal experiments performed at Harvard were approved by the BIDMC Institutional Animal Care and Use Committee.

Bioinformatics analysis of ZIKV genomes and proteins. Complete ZIKV genomes deposited in the GenBank in November 2015 (no less than 10.5 kb) were used to construct a phylogenetic tree, which allowed us to further identify and design our vaccine transgene candidates. Further phylogenetic trees were built for genome sequences deposited in the gene bank between April and October 2016. Alignment analysis for both nucleotides and protein sequences, consensus analysis and phylogenetic trees were produced by focusing only in the structural and non-structural coding sequences. Geneious software was used to generate the phylogeny trees using a Neighbour-Joining method. Sequence alignments between zika, dengue and yellow fever proteins allowed us to annotate the ZIKV consensus sequences generated in November 2015 using the following features (start codon, transmembrane domains and known flaviviruses enzymatic cleavage sites).

Transgene design. For the Asian Lineage vaccine design, prM and envelope coding sequence (translated to 670 a.a.) from the Asian ZIKV consensus sequence was codon optimised, and a tPA leading nucleotide sequence was added before the 5' end of prM. To improve initiation of translation a Kozak consensus sequence was included before the 5' end of tPA. Finally, the transgene design included the required enzymatic restriction sites to allow the in-frame cloning of the transgene between the cytomegalovirus (CMV) promoter and the polyA sequence region contained in our shuttle and expression vector (pMono). Synthetic gene cassette was produced by GeneArt® and was named prME.

The GeneArt® prME plasmid was used as template to further generate the following versions by PCR subcloning, including restriction sites, stop codons or other features as follows:

- prME with the TM domain of envelope deleted (prME ΔTM):
Forward primer AGAAGAGGATCGAAGCTTGCCATCACCAGAGAGGCCA
GGC
Reverse primer GCATGCTCGAGCGGCGCTCATCACATTCTCTGGCTC
CCCCG
- Full envelope (Env):
Forward primer GAAACCGGTAAGCTTATGCGGTGTATCGGCGGTGTCC
Reverse primer ACCGGTCAGACGTACTCGA
- Envelope with TM deleted (Env ΔTM):
Forward primer GAAACCGGTAAGCTTATGCGGTGTATCGGCGGTGTCC
Reverse primer CACCAATGCGGCGGCTCGAGCTATTACAGAGAGTTCA
GGGCTCCG

DNA vaccine production. The prME and prME ΔTM were digested with *KpnI* and *NofI* to allow in-frame ligation between the CMV promoter and the poly(A) regions contained in the DNA vaccine plasmid (pMono). The remaining PCR-subcloning transgenes (Env and Env ΔTM) were digested with *AgeI* and *NofI* to allow in-frame ligation between the tPA leader sequence and the poly(A) regions of pMono. pMono was digested accordingly to generate the following DNA vaccines: pMono prME; pMono prME ΔTM; pMono Env; and pMono Env ΔTM. Zika DNA plasmid vaccines were expanded in *Escherichia coli* and Endo-free Giga Prep (Qiagen) was used for plasmid purifications. DNA vaccines were verified by restriction analysis and 5' and 3' flanking sequencing. In addition, DNA vaccines were transfected in HEK-293 and Vero cells to verify the correct expression of the transgenes.

Adenoviral Zika vaccine production. All of the DNA constructs required for ChAdOx1 were cloned in two steps. In the first step, unique *KpnI* and *NofI* sites were used to insert the synthetic transgenes into an adenovirus entry plasmid. The transgene was placed upstream of BGH poly(A) transcription termination sequence and under the control of the CMV promoter. The entry plasmid containing attL regions sequences was recombined with those attR regions contained in the destination vector ChAdOx1 using an in vitro Gateway reaction (LR Clonase II system, Invitrogen). Successfully recombined ChAdOx1 ZIKV plasmids were verified by DNA sequencing using flanking primers (promoter forward primer and poly-(A) reverse primer). Standard cell biology and virology techniques were performed to generate the non-replicative adenoviral vectors³⁶.

Cell culture. HEK-293 and Vero cells (ATCC®, CRL-1573TM and ATCC CCL81TM, respectively) were grown in Dulbecco's modified Eagle's media (DMEM) supplemented with 10% fetal bovine serum, 1% L-glutamine and 1% non-essential amino acids. Cell number and viability were calculated by trypan blue

staining using the Countess Automated Cell Counter (Life Technologies). Both cell lines are described in the NCBI Biosample database.

Immunofluorescence microscopy. Transfected Vero cells grown on coverslips were fixed in 4% paraformaldehyde in 1× PBS (10 min, room temperature), permeabilised in 0.1% PBS-Triton X-100 for 20 min at room temperature (RT) and blocked in 2.5% goat serum/0.5% bovine serum albumin (BSA) in PBS for 45 min at RT. The primary antibody (anti-flavivirus group antigen D1-4G2-4-15, 1/250, Merck Millipore) was diluted in 1% BSA in PBS and incubated with samples for 1 h at RT. After successive washes in PBS, secondary antibody coupled to Alexa Fluor 488 was added for 1 h at RT in the dark. Samples were washed in PBS with a last wash in distilled water and mounted in Prolong-DAPI mounting reagent (Life Technologies). Cells were acquired on a LSM 880 inverted confocal microscope using Zen 2.1 software (Zeiss). Images were analysed using ImageJ software (National Institutes of Health).

Western blot analysis. Cells were washed twice with ice-cold PBS and harvested in cold 1× PBS. 1× and 4× Laemmli buffer were added to cells and supernatants, respectively, and boiled at 100 °C for 5 min. Equal amounts of total cell extracts and 15 µl of cell supernatants were resolved by SDS/PAGE and transferred to polyvinylidene fluoride membranes. In some cases, supernatants were 10× concentrated using Amicon Ultra-0.5 ml centrifugal filters. Blots were blocked with 1× PBS-Tween-5% milk and incubated with anti-ZIKV envelope seropositive mice sera (1:500) or a mAb versus ZIKV envelope (AZ1176, Clone 0302156), Aalto Bio Reagents), followed by incubation with horseradish peroxidase (HRP)-conjugated secondary antibody (1:5000). Chemiluminescence (Perkin-Elmer Life Sciences, Boston, MA) was visualised using the BioRad ChemiDoc SRS device. Uncropped images of the blots are shown in Supplementary Figs. 2, 5, 9, 10 and 11).

Whole IgG ELISA. RecombivirusTM mouse anti-ZIKV envelope protein IgG ELISA kits (Alpha Diagnostic International, RV-403120-1) were used under the manufacturer protocol. Briefly, 96-well plates coated with ZIKV envelope protein were equilibrated with 300 µl of kit working wash buffer. Serial dilutions (threefold) of sera from vaccinated mice were added. Diluted sera were incubated at RT for 1 h, and after four times washing buffer incubations, 100 µl/well of anti-mouse IgG HRP conjugate working solution was added for 30 min at room temperature. Plates were washed five times and developed 15 min at room temperature with 100 µl of tetramethyl benzidine (TMB) substrate, then stopped by the addition of 100 µl of stop solution. Absorbance was measured at 450 nm on a microplate reader. ELISA optical densities (ODs) were compared between all vaccinated groups at different sera dilutions. Alternatively, a relative positive index was calculated following the manufacturer RecombivirusTM Mouse Anti-Zika virus ELISA kit. Briefly, positive index was calculated by the OD mean plus two times the s.d. of the control/naive sera. Then, we divided the OD of each sample by the positive index. Values below 1 are considered negative and values above 1 are considered positive. Relative positive index then was plot for every vaccinated mouse per group.

For pre-challenge ELISA, endpoint titres were defined as the highest reciprocal serum dilution that yielded an absorbance greater than twofold over background values, as calculate elsewhere^{8–10}.

DENV1-4 ELISA: Dengue Env-2 45 kDa (Den021), Dengue Env-1 45 kDa (Den022), Dengue Env-3 45 kDa (Den023) and Dengue Env-4 45 kDa (Den024) from Dundee Cell Products Ltd. were used to coat Maxiisorp NUNC 96-well plates (2 µg/ml) and ELISAs performed to determine the reciprocal titre of the sera dilution.

Ex vivo IFN γ ELISPOT assay. ELISPOTs were carried out using peripheral blood mononuclear cells (PBMCs) isolated from the blood. Briefly, MAIP ELISPOT plates (Millipore) were coated an anti-mouse IFN γ mAb, after 1 h blocking with complete DMEM media (10% fetal calf serum). Isolated PBMCs (using ACK buffer solution) were plated alongside with 20-mer specific peptides overlapped by 10 a.a. (10 µg/ml) and 2.5×10^5 splenocytes from naive mice per well. After 16 h incubation, cells were discarded and plates washed with PBS. Following this, 50 µl of biotinylated anti-mouse IFN γ mAb (1:1000 in PBS) was added to each well and incubated for 2 h. After washing, plates were incubated with 50 µl of ALP (1:1000 in PBS) reagent for 1 h. After another washing step, developing solution (BIORAD) was used. Once spots were visible, the reaction was stopped by washing plate off with water. Spots were acquired using an ELISPOT reader. Spot-forming cells/ 10^6 PBMCs producing IFN γ were calculated.

ZIKV challenge model. ZIKV challenge was performed as described⁷. Briefly, naive and vaccinated BALB/c mice ($n = 5$ /group) were infected at week 4 post immunisation by the intravenous route with 10^5 viral particles (10^5 PFU) of ZIKV-BR. VIs following ZIKV challenge were quantitated by RT-PCR at days 1, 2, 3, 4 and 7. Sample size was determined to achieve 80% power to detect significant differences in protective efficacy. Vaccines were administered in a blind experiment. After the experimental outcome of the ZIKV challenge and the pre-challenge ELISAs, samples were decoded.

Reverse transcription-PCR. VIs were assessed as depicted⁷. Briefly, RNA was extracted from serum with a QIAcube HT (Qiagen). RNA was purified using the RNA clean and concentrator kit (Zymo Research), and RNA quality and concentration was assessed by the BIDMC Molecular Core Facility of Harvard University. Log dilutions of the RNA standard were reverse-transcribed and included with each RT-PCR assay. VIs were calculated as virus particles/ml. Assay sensitivity was 100 copies/ml. The infectivity of virus in peripheral blood from ZIKV-challenged mice was confirmed by PFU assays. Specific primers were used to amplify a region contained in the Capsid of the ZIKV genome.

Polyclonal anti-ZIKV E antiserum R34. We identified a short region (a.a. 147–161) around the N154 glycosylation site of the envelope protein that is unique to each flavivirus and that forms an exposed loop (the so called 150 loop) on the surface of the envelope protein (Supplementary Fig. 6a, b), and then raised a rabbit antiserum (R34) to it which specifically recognised the ZIKV envelope protein (Supplementary Fig. 6c). A peptide corresponding to amino acid residues 144–166 of ZIKV envelope (HGSQHSQMIVNDTGHTDENRAK) with an additional C-terminal cysteine residue was synthesised by Isca Biochemicals (UK) and directionally conjugated to keyhole limpet haemocyanin (KLH) using maleimide chemistry. Two rabbits were immunised with the peptide-KLH conjugate combined with Freund's adjuvant, following a standard protocol. This work was carried out by Dundee Cell Products (UK). The immune sera were tested by western blot, ELISA and IF for specific reactivity with ZIKV E protein, and R34 was selected as the better antiserum. Total IgG was purified from R34 antiserum by protein G affinity chromatography.

ZIKV MN assay. Vero cells were seeded in 96-well plates at 7×10^3 /well and incubated at 37 °C in 5% CO₂ for 1 day to form a confluent monolayer. Aliquots of Brazilian ZIKV-PE243²¹ at 100 PFU in 15 µl were mixed with an equal volume of twofold dilutions of sera in medium and incubated at 37 °C for 1 h. The cells were infected with the serum/virus mixture (30 µl/well), incubated at 37 °C for 1 h, and then 100 µl of medium was added to each well. At day 3 post infection, cells were lysed in lysis buffer (20 mM Tris-HCl [pH 7.4], 20 mM iodoacetamide, 150 mM NaCl, 1 mM EDTA, 0.5% Triton X-100 and Complete[®] protease inhibitors) and the viral E protein quantitated by sandwich ELISA (below). The amount of envelope detected correlates with the level of virus infectivity. Values representing % of ZIKV infectivity relative to the control (i.e., virus not pre-incubated with immune sera) were plotted using Graphpad Prism 6, and nonlinear regression (curve fit) performed for the data points using Log (inhibitor) versus response (variable slope) to determine MN50 titres. The MN50 titre was defined as the serum dilution that neutralised >50% of ZIKV. This format resulted in a very robust MN assay, with a signal-to-noise ratio of >10 and a Z-factor of 0.8, which is optimal value for high-throughput screening assays. A comparison between our MN assay versus a traditional PRNT assay⁴¹ was performed, using a series of twofold dilutions of a highly potent monoclonal neutralising antibody to ZIKV (EDE1 C8)⁴².

Sandwich ELISA to assess ZIKV infectivity. Immulon 2 HB plate (Thermo Scientific[®]) 96-well plates were coated with 3 µg/well of purified pan-flavivirus MAb D1-4G2-4-15 (ATCC[®] HB112TM) in PBS and incubated overnight at RT. The antibody was removed and the wells blocked for 2 h at RT with 2% skimmed milk powder in PBS containing 0.02% Tween-20 (PBST). After washing with PBST, ZIKV-infected cell lysates were added and incubated for 1 h at RT. Wells were washed with PBST, incubated with R34 IgG at 6 µg/ml in PBST for 1 h at RT and washed again. The antibodies bound to ZIKV envelope protein were detected using anti-rabbit HRP conjugate (Abcam) and TMB (Sigma) substrate. Absorbance at 450 nm was measured in a microplate reader.

In vitro assessment of ADE. Serially diluted mAb or mice sera samples were incubated with virus DENV2 strain 16681, $Moi = 2$ for 1 h at 37 °C before adding to U937 cells. After 3-day incubation, supernatants were harvested and viral titres determined by focus-forming units. Fold enhancement was calculated by comparison to viral titres in the presence or absence of antibody²². Antibodies used as positive control are mouse anti-Flavivirus envelope protein antibody (4G2) and anti-envelope dimer Epitope 1 [752-2C8].

Statistics. Graphing and statistical analysis were performed using GraphPad Prism version 7.00 for Mac, GraphPad Software, La Jolla California USA, www.graphpad.com. For pre-challenge ELISA, the analysis used was the one-way analysis of variance (ANOVA), followed by Turkey's multiple comparisons test. For the ZIKV neutralisation assay, a two-way ANOVA followed by Sidak's multiple comparison was performed. All graphs were plotted with bars and errors, representing the mean and the s.d. Assumptions of data are normal distribution, homogeneity, independent and groups must have same sample size.

Data availability. The authors declare that the data supporting the findings of this study are shown in the article and in the Supplementary section and are available from the authors upon request.

Received: 11 January 2017 Accepted: 23 May 2018

Published online: 22 June 2018

References

- Dick, G. W., Kitchen, S. F. & Haddow, A. J. Zika virus. I. Isolations and serological specificity. *Trans. R. Soc. Trop. Med. Hyg.* **46**, 509–520 (1952).
- Oehler, E. et al. Zika virus infection complicated by Guillain-Barre syndrome—case report, French Polynesia, December 2013. *Eur. Surveill.* **19**, 20720 (2014).
- Ventura, C. V., Maia, M., Bravo-Filho, V., Gois, A. L. & Belfort, R. Jr. Zika virus in Brazil and macular atrophy in a child with microcephaly. *Lancet* **387**, 228 (2016).
- Swaminathan, S., Schlager, R., Lewis, J., Hanson, K. E. & Couturier, M. R. Fatal Zika virus infection with secondary nonsexual transmission. *N. Engl. J. Med.* **375**, 1907–1909 (2016).
- Kostyuchenko, V. A. et al. Structure of the thermally stable Zika virus. *Nature* **533**, 425–428 (2016).
- Sirohi, D. et al. The 3.8 Å resolution cryo-EM structure of Zika virus. *Science* **352**, 467–470 (2016).
- Larocca, R. A. et al. Vaccine protection against Zika virus from Brazil. *Nature* **536**, 474–478 (2016).
- Yockey, L. J. et al. Vaginal exposure to Zika virus during pregnancy leads to fetal brain infection. *Cell* **166**, 1247–1256.e4 (2016).
- Abbink, P. et al. Protective efficacy of multiple vaccine platforms against Zika virus challenge in rhesus monkeys. *Science* **353**, 1129–1132 (2016).
- Dowd, K. A. et al. Rapid development of a DNA vaccine for Zika virus. *Science* **354**, 237–240 (2016).
- Chahal, J. S. et al. An RNA nanoparticle vaccine against Zika virus elicits antibody and CD8+T cell responses in a mouse model. *Sci. Rep.* **7**, 252 (2017).
- Diamond, M. S. & Coyne, C. B. Vaccines in 2017: closing in on a Zika virus vaccine. *Nat. Rev. Immunol.* **18**, 89–90 (2018).
- Pardi, N. et al. Zika virus protection by a single low-dose nucleoside-modified mRNA vaccination. *Nature* **543**, 248–251 (2017).
- Prow, N. A. et al. A vaccinia-based single vector construct multi-pathogen vaccine protects against both Zika and chikungunya viruses. *Nat. Commun.* **9**, 1230 (2018).
- Richner, J. M. et al. Modified mRNA vaccines protect against Zika virus infection. *Cell* **169**, 176 (2017).
- Abbink, P. et al. Durability and correlates of vaccine protection against Zika virus in rhesus monkeys. *Sci. Transl. Med.* **9**, ea04163 (2017).
- Xu, K. et al. Recombinant chimpanzee adenovirus vaccine AdC7-M/E protects against Zika virus infection and testis damage. *J. Virol.* **92**, e01722–17 (2018).
- Yi, G. et al. A DNA vaccine protects human immune cells against Zika virus infection in humanized mice. *EBioMedicine* **25**, 87–94 (2017).
- Thomas, S. J. et al. A phase II, randomized, safety and immunogenicity study of a re-derived, live-attenuated dengue virus vaccine in healthy adults. *Am. J. Trop. Med. Hyg.* **88**, 73–88 (2013).
- Modjarrad, K. et al. Preliminary aggregate safety and immunogenicity results from three trials of a purified inactivated Zika virus vaccine candidate: phase 1, randomised, double-blind, placebo-controlled clinical trials. *Lancet* **391**, 563–571 (2017).
- Donald, C. L. et al. Full genome sequence and sRNA interferon antagonist activity of Zika virus from Recife, Brazil. *PLoS Negl. Trop. Dis.* **10**, e0005048 (2016).
- Dejnirattisai, W. et al. Dengue virus sero-cross-reactivity drives antibody-dependent enhancement of infection with Zika virus. *Nat. Immunol.* **17**, 1102–1108 (2016).
- Guy, B., Briand, O., Lang, J., Saville, M. & Jackson, N. Development of the Sanofi Pasteur tetravalent dengue vaccine: one more step forward. *Vaccine* **33**, 7100–7111 (2015).
- Villar, L. et al. Efficacy of a tetravalent dengue vaccine in children in Latin America. *N. Engl. J. Med.* **372**, 113–123 (2015).
- Kim, E. et al. Preventative vaccines for Zika virus outbreak: preliminary evaluation. *EBioMedicine* **13**, 315–320 (2016).
- Golden, A. et al. Effect of promoters and signal sequences on the production of secreted HIV-1 gp120 protein in the baculovirus system. *Protein Expr. Purif.* **14**, 8–12 (1998).
- Luo, M. et al. Immunization with plasmid DNA encoding influenza A virus nucleoprotein fused to a tissue plasminogen activator signal sequence elicits strong immune responses and protection against H5N1 challenge in mice. *J. Virol. Methods* **154**, 121–127 (2008).
- Tameris, M. D. et al. Safety and efficacy of MVA85A, a new tuberculosis vaccine, in infants previously vaccinated with BCG: a randomised, placebo-controlled phase 2b trial. *Lancet* **381**, 1021–1028 (2013).
- Satti, I. et al. Safety and immunogenicity of a candidate tuberculosis vaccine MVA85A delivered by aerosol in BCG-vaccinated healthy adults: a phase 1, double-blind, randomised controlled trial. *Lancet Infect. Dis.* **14**, 939–946 (2014).
- Gaudinski, M. R. et al. Safety, tolerability, and immunogenicity of two Zika virus DNA vaccine candidates in healthy adults: randomised, open-label, phase 1 clinical trials. *Lancet* **391**, 552–562 (2017).
- Tebas, P. et al. Safety and immunogenicity of an anti-Zika virus DNA vaccine—preliminary report. *N. Engl. J. Med.* <https://doi.org/10.1056/NEJMoa1708120> (2017).
- Halstead, S. B. Dengue antibody-dependent enhancement: knowns and unknowns. *Microbiol. Spectr.* **2** <https://doi.org/10.1128/microbiolspec.AID-0022-2014> (2014).
- Halstead, S. B., Mahalingam, S., Marovich, M. A., Ubol, S. & Mosser, D. M. Intrinsic antibody-dependent enhancement of microbial infection in macrophages: disease regulation by immune complexes. *Lancet Infect. Dis.* **10**, 712–722 (2010).
- Pantoja, P. et al. Zika virus pathogenesis in rhesus macaques is unaffected by pre-existing immunity to dengue virus. *Nat. Commun.* **8**, 15674 (2017).
- Stettler, K. et al. Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. *Science* **353**, 823–826 (2016).
- Antrobus, R. D. et al. Clinical assessment of a novel recombinant simian adenovirus ChAdOx1 as a vectored vaccine expressing conserved Influenza A antigens. *Mol. Ther.* **22**, 668–674 (2014).
- Borthwick, N. et al. Vaccine-elicited human T cells recognizing conserved protein regions inhibit HIV-1. *Mol. Ther.* **22**, 464–475 (2014).
- Ewer, K. et al. A monovalent chimpanzee adenovirus ebola vaccine boosted with MVA. *N. Engl. J. Med.* **374**, 1635–1646 (2016).
- Ewer, K. J. et al. Protective CD8+T cell immunity to human malaria induced by chimpanzee adenovirus-MVA immunisation. *Nat. Commun.* **4**, 2836 (2013).
- Green, C. A. et al. Chimpanzee adenovirus- and MVA-vectored respiratory syncytial virus vaccine is safe and immunogenic in adults. *Sci. Transl. Med.* **7**, 300ra126 (2015).
- Poggianella, M. et al. Dengue E protein domain III-based DNA immunisation induces strong antibody responses to all four viral serotypes. *PLoS Negl. Trop. Dis.* **9**, e0003947 (2015).
- Barba-Spaeth, G. et al. Structural basis of potent Zika-dengue virus antibody cross-neutralization. *Nature* **536**, 48–53 (2016).

Acknowledgements

We would like to thank the Jenner Institute's Vector Core Facility for producing and supplying the recombinant viral vectors. A.R.-S., A.V.S.H. and S.C.G. are Jenner Investigators and an Oxford Martin Fellows. R.D.L.N. and D.F.C. were supported by Capes, Facepe, CNPq and LNCC. This report is independent research funded by the UK Department of Health and Social Care through Innovate UK "New vaccines for global epidemics: development and manufacture" grant No. 972216 (A.R.-S.), and also funded from an ODA budget (Global Health (ODA), 16/107/05—Design, development and GMP manufacture of a Zika vaccine) (A.H.P. and A.R.-S.). The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health and Social Care. We also acknowledge funding by the UK Medical Research Council [MC_UU_12014 (A.H.P. and A.K.) and MR/N017552/1 (A.K.)]. Juthathip Mongkolsapaya is supported by an MRC-Newton Fund grant, Gavin Screaton is a Wellcome Trust Senior Investigator.

Author contributions

A.R.-S. and C.L.-C. directed the project and commissioned the work. C.L.-C. designed, constructed and characterised the vaccines. C.L.-C., P.A. and R.A.L. designed and performed the animal experiments. M.B., A.B.-Z., P.A. and R.A.L. performed the ZIKV challenge model and the RT-PCR viral loads. C.L.-C., P.A. and R.A.L. performed ELISA assays and analysed ZIKV challenge data. W.D. and J.M. designed, performed and analysed in vitro ADE experiments. C.L.-C. performed ELISPOTs, cell culture, transfections and western blots. Y.C.K. performed animal work and assisted with ELISA experiments. Z.R.W. performed confocal microscopy. A.H.P., A.K., A.O., R.S.V., C.L.D., J.D. and G.D.L. designed and performed the microneutralisation assay. R.D.L.N. and D.F.C. designed the anti-ZIKV envelope R34 antiserum. A.H.P., A.K., L.D., G.R.S., D.H.B., S.C.G. and A.S.H. provided vital characterised reagents and conceptual support. All authors read and commented the manuscript. C.L.-C. and A.R.-S. wrote the manuscript.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-04859-5>.

Competing interests: A.R.-S. and C.L.-C. are co-inventors of the Zika vaccines described in this manuscript, filed by Oxford University Innovation Limited in the International Patent Application No. PCT/GB2017/052220 Zika Vaccine; A.V.S.H. and S.C.G. are co-inventors on a patent application (WO/2012/172277) on the ChAdOx1 viral vector filed by Oxford University Innovation. The remaining authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018

APÊNDICE B - PUBLICAÇÃO RELACIONADA AOS RESULTADOS DESTA TESE E QUE ESTÁ EM PROCESSO DE SUBMISSÃO

Título: *In vitro* detection of B-cells producing HIV-1 2F5 antibodies by a computer designed protein.

Autores: Danilo F. Coêlho, Matheus V. F. Ferraz, Gabriel G. Slade, Ronaldo J. Oliveira, Ernesto T. A. Marques, Roberto D. Lins, Isabelle F. T. Viana.

APÊNDICE C - ARTIGOS PUBLICADOS REFLETINDO APENAS PROJETOS PARALELOS AO LONGO DA TRAJETÓRIA DO DOUTORADO.

Título: Mapping Putative B-Cell Zika Virus NS1 Epitopes Provides Molecular Basis for Anti-NS1 Antibody Discrimination between Zika and Dengue Viruses.

Autores: Marjorie C. L. C. Freire, Laércio Pol-Fachin, **Danilo F. Coêlho**, Isabelle F. T. Viana, Tereza Magalhaes, Marli T. Cordeiro, Nico Fischer, Felix F. Loeffler, Thomas Jaenisch, Rafael F. França, Ernesto T. A. Marques, **Roberto D. Lins**.

Publicado no periódico: ACS Omega, 2017.

doi: 10.1021/acsomega.7b00608



Mapping Putative B-Cell Zika Virus NS1 Epitopes Provides Molecular Basis for Anti-NS1 Antibody Discrimination between Zika and Dengue Viruses

Marjorie C. L. C. Freire,[†] Laércio Pol-Fachin,^{†,‡} Danilo F. Coêlho,^{†,‡} Isabelle F. T. Viana,[†] Tereza Magalhães,[†] Marli T. Cordeiro,[†] Nico Fischer,^{§,⊥} Felix F. Loeffler,^{||,⊥} Thomas Jaenisch,^{§,⊥} Rafael F. Franca,[†] Ernesto T. A. Marques,^{*,†,‡} and Roberto D. Lins^{*,†,‡,§}

[†]Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife 50740-465, Brazil

[‡]Department of Fundamental Chemistry, Federal University of Pernambuco, Recife 50740-670, Brazil

[§]Department of Infectious Diseases, Section Clinical Tropical Medicine, Heidelberg University Hospital, INF 324, Heidelberg 69120, Germany

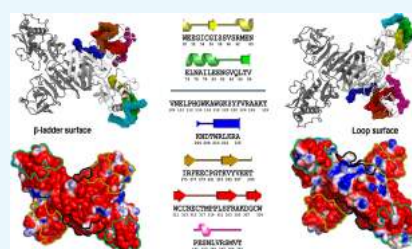
^{||}Department of Biomolecular Systems, Max Planck Institute of Colloids and Interfaces, Am Mühlenberg 1, Potsdam 14476, Germany

[⊥]HEiKA - Heidelberg Karlsruhe Research Partnership, Heidelberg University, Karlsruhe Institute of Technology (KIT), Karlsruhe 76131, Germany

^{*}Center for Vaccine Research, University of Pittsburgh, Pittsburgh 15261, Pennsylvania, United States

Supporting Information

ABSTRACT: B-cell epitope sequences from Zika virus (ZIKV) NS1 protein have been identified using epitope prediction tools. Mapping these sequences onto the NS1 surface reveals two major conformational epitopes and a single linear one. Despite an overall average sequence identity of ca. 55% between the NS1 from ZIKV and the four dengue virus (DENV) serotypes, epitope sequences were found to be highly conserved. Nevertheless, nonconserved epitope-flanking residues are responsible for a dramatically divergent electrostatic surface potential on the epitope regions of ZIKV and DENV2 serotypes. These findings suggest that strategies for differential diagnostics on the basis of short linear NS1 sequences are likely to fail due to immunological cross-reactions. Overall, results provide the molecular basis of differential discrimination between Zika and DENVs by NS1 monoclonal antibodies.



INTRODUCTION

Given the recent outbreak of Zika virus (ZIKV) in the Americas in 2015 and the association with microcephaly cases^{1,2} and neurological disorders, such as the Guillain-Barré syndrome, ZIKV was recently placed as a public health emergency of international concern by the World Health Organization.³ ZIKV is a flavivirus of the family *Flaviviridae*, transmitted by *Aedes* sp. mosquitoes,⁴ and closely related to dengue virus (DENV), West Nile virus, Japanese encephalitis virus, and yellow fever virus (YFV).⁵ To date, there is no prophylactic treatment nor vaccine available against ZIKV and disease control is limited to vector eradication strategies. The presumptive diagnosis is typically clinical, whereas confirmatory laboratory tests include classic virus isolation on cell culture and viral RNA detection by reverse transcriptase PCR (RT-PCR) in serum, saliva, and/or urine samples within the first 5 or 6 days of infection.^{6,7} ZIKV serology is usually performed by enzyme-linked immunosorbent assay (ELISA), using full-length viral

proteins or linear peptides, however, due to high cross-reactivity of IgM and IgG antibodies between ZIKV and other related flavivirus, especially in endemic areas where cocirculation exists, confirmatory tests are necessary. Currently, confirmation testing is performed by plaque reduction neutralization test (PRNT), according to previously published protocols.^{8–11} At present, there are only a very few commercially available serology kits for ZIKV antibody detection; moreover, the specificity of these kits remains elusive, given the high cross-reactivity already reported among individuals who experienced consecutive flavivirus infections.^{12,13} On the basis of this, PRNT remains the “gold standard” for ant flavivirus differentiation.¹⁰ PRNT, however, is a high-cost technique that requires highly specialized

Received: May 15, 2017

Accepted: July 13, 2017

Published: July 25, 2017

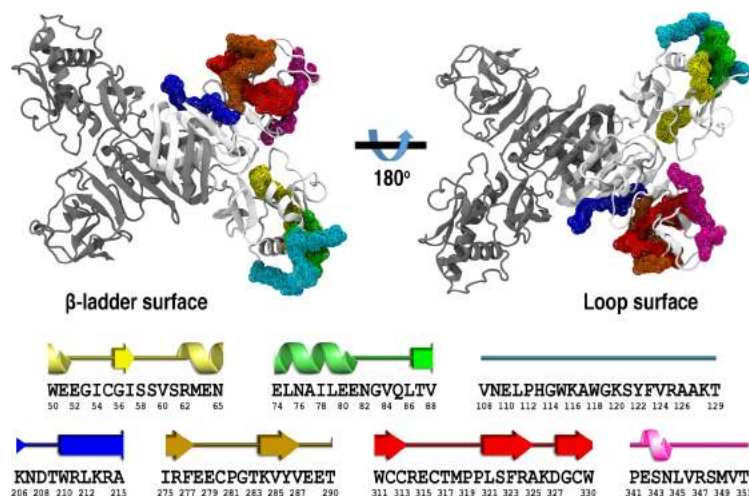


Figure 1. Cartoon representation of the ZIKV NS1 homodimer structure. Protein chains are shown in white and gray. Putative epitopes are mapped as the wireframe surface and color coded as E1 in yellow, E2 in green, E3 in cyan, E4 in blue, E5 in orange, E6 in red, and E7 in magenta. The corresponding secondary structure assignment for each putative epitope is shown at the bottom, with their respective amino acid sequence (PDB ID 5K6K).

laboratories and special regulations due to live virus manipulations. Although new protocols using recombinant viruses¹⁴ or reporter virus particles¹¹ have been developed, these are still not available for ZIKV routine diagnosis.¹⁵ In this scenario, there is an urgent need for a low-cost unequivocal serological diagnostic method for ZIKV to be able to overcome the high cross-reactivity with other flaviviruses, such as DENV and yellow fever (especially on vaccinated individuals).¹⁶

The ZIKV positive-sense RNA genome comprises a single open reading frame encoding a unique polyprotein that is cleaved into three structural proteins (Capsid (C), premembrane (prM), and envelope (E)), which form the virus particle, and seven nonstructural (NS) proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5), which perform essential functions in genome replication, polyprotein cleavage, and modulation of cellular processes to viral advantage.^{17,18} Among the NS proteins, NS1 glycoprotein is a multifunctional virulence factor.^{19,20} Within infected cells, glycosylated NS1 is found as a membrane-associated dimer after translocation into the endoplasmic reticulum lumen, in which it colocalizes with dsRNA and the viral replication complex, being essential for viral genome replication.¹⁸ Infected cells also secrete NS1 as a hexameric lipoprotein,²¹ which interacts with complement-system proteins and has several immune-modulatory functions.¹⁸ This conformational dynamism of NS1 is also employed by the virus during the infection cycle to evade antiviral cellular immune responses from the host.²² The homodimer structure of ZIKV NS1 is presented in Figure 1 and encompasses a highly structured continuous extended β -sheet ladder on one surface. On the opposite face, there is a predominance of a nonstructured loop medley and a flexible loop, comprising the residues from 108 to 129.¹⁸ The loop

surface of flaviviruses NS1 has been previously suggested to play a role in interactions of this protein with host factors and antibodies.²³

Besides playing a crucial role in the pathogenesis of the ZIKV infection, NS1 emerges as a potential candidate for differential diagnostics.¹⁶ NS1 is considered the major antigenic marker for flavivirus infections, and it has been used for the detection of DENV in patient sera.^{22,24–30} Despite structural conservation of NS1 among flaviviruses, the average sequence identity between ZIKV and DENV NS1 proteins is about 55%,^{16,31} and ZIKV-specific monoclonal antibodies (mAbs) have been isolated,¹⁶ showing that this protein contains regions capable of being specifically recognized. Additionally, some differences between the NS1 of ZIKV and other flaviviruses have been reported. A negatively charged glutamate residue located at position 315 of ZIKV NS1 has been pointed out as exclusive for ZIKV, being absent in other flaviviruses including DENV serotypes which have a serine residue at this position. As a consequence, differences are observed in the electrostatic surface potential of ZIKV NS1 compared to that of other flaviviruses.³¹ Another recent work reported the crystal structure of full-length ZIKV NS1 and revealed a region formed by the residues 108–129 on the wing-domain flexible loop, which had not been previously described in other flaviviruses.¹⁸ Nevertheless, complete mapping of B-cell ZIKV NS1 epitopes have not been reported. In the present work, we have assigned ZIKV NS1 putative B-cell epitopes and mapped the identified regions onto the corresponding three-dimensional structure. Comparisons of the identified B-cell ZIKV NS1 epitopes to their homologous regions in NS1 from DENV2 on the basis of sequence identity analysis and conformation assessment through molecular dynamics (MD)

simulations provided a solid understanding of the molecular basis for differential antibody discrimination.

RESULTS AND DISCUSSION

Epitope Prediction. Detection of antigen and pathogen specific antibodies is among the most used approaches for clinical diagnostics. Antigen specific antibody detection is used in the diagnosis of several diseases, and its application continues to expand. In addition, antibodies are the most effective mechanism of protection provided by the adaptive immune system. A critical and at many times limiting aspect for eliciting highly specific antibodies is the structural transiency of certain conformational B-cell epitopes. In addition, important epitopes are often cryptic, frequently buried within the protein, and only exposed when the virus binds to its receptor.³² Thus, when searching for B-cell epitopes in a given protein, it is important that they are located in accessible regions to allow proper recognition and interaction with their respective antibodies. Therefore, to perform *in silico* epitope prediction, the ZIKV NS1 protein structure (PDB ID: 5K6K)¹⁸ was used. Although the ZIKV NS1 structure is presented in a dimeric form, only the B chain was used in the epitope prediction analyses, as the three-dimensional coordinates for residues 113–120 have not been experimentally assigned for the chain A. Most methods of epitope prediction use properties related to surface exposure, such as flexibility or hydrophobicity.³³ In this work, we have chosen three algorithms to assess the robustness of our epitope prediction rationale. We combined ElliPro that is based on the geometrical properties of the protein structure (antigenicity, solvent accessibility, and flexibility) and can predict linear and conformational epitopes,³⁴ EpiTope that implements a machine-learning algorithm that was trained to discern antigenic features within a given protein and predicts the most probable conformational epitope regions,³⁵ and DiscoTope that is based on amino acid statistics, spatial information, and surface accessibility and can predict residues that are part of conformational epitopes recognized by B-cells.³⁶ A previous study has pointed out that consensus results from the three servers would improve the specificity and accuracy of epitope prediction, resulting in a lower number of false positive results and thus generating greater confidence.³⁷ Considering consensus overlapping regions by at least two servers, 7 putative ZIKV NS1 regions have been selected, with epitope lengths ranging from 10 to 22 residues (Table 1). The identified regions include sequences that encompass the E315 residue and residues 108–129, which have been previously described due to the presence of exclusive features in ZIKV that were not found in other flaviviruses.^{18,31}

Table 1. ZIKV NS1 Regions Identified *In Silico* as Putative Epitopes

ZIKV NS1 epitope	residue interval	sequence length	amino acid sequence
E1	50–65	16	WEEGICGISSVSRMEN
E2	74–88	15	ELNAILEENGVLTV
E3	108–129	22	VNELPHGWKAWGKSYFVRAAKT
E4	206–215	10	KNDTWRLKRA
E5	275–290	16	IRFECECPGTVVVEET
E6	311–330	20	WCRECTMPPLSFRAKDGCW
E7	341–351	11	PESNLVRSMT

Upon comparing the results obtained by each server, ElliPro predicted a higher number of B-cell epitopes, followed by EpiTope, whereas DiscoTope predicted only three putative epitopes (Table 2). The greater performance of ElliPro may be

Table 2. Overview of the *In Silico* B-Cell Epitope Predictions for the Seven Identified Regions and Their Classification into Putative Conformational or Linear Epitopes^{a,b}

ZIKV NS1 epitope	ElliPro	EpiTope	DiscoTope
CE1			
E1	+	+	–
E2	+	+	–
E3	+	+	+
LE1			
E4	+	–	+
CE2			
E5	+	+	–
E6	+	+	–
E7	+	+	+

^a+, Prediction contains at least one residue from the consensus interval, as described in Table 1. ^b–, Did not predict any residue within the consensus interval, as described in Table 1.

related to an algorithm that takes into account structural information as well as the identification of regions in the protein that protrude from the globular surface. DiscoTope presented a lower performance possibly because it uses a sequence-based prediction wherein individual residues are listed. Therefore, this suggests that at least for the NS1 protein of ZIKV, sequence information alone may lead to a poor prediction performance. The selected regions were mapped onto the ZIKV NS1 structure along with their corresponding secondary structure content, as shown in Figure 1. Among the seven identified putative epitopes, the structural mapping (Figure 1) suggests that those regions comprise two conformational epitopes (CE1 and CE2) and one linear epitope (LE1). CE1 consists of E1, E2, and E3 identified regions; CE2 consists of E5, E6, and E7 identified regions; whereas LE1 corresponds to the E4 region. The same epitope prediction scheme was carried out for the NS1 protein from DENV2 (PDB ID 4O6B³⁸), as it is the only serotype structure available in the PDB. All three algorithms predicted the seven corresponding regions in the same manner as putative epitopes (data not shown for conciseness). This is not unexpected given that the two proteins share a nearly identical three-dimensional structure.

ZIKV Peptide Microarray. To assess the reactivity of the identified ZIKV NS1 epitopes with specific IgG antibodies, a 15-mer peptide library spanning the entire ZIKV NS1 protein was tested against seven confirmed convalescent ZIKV positive patients and 6 ZIKV and DENV naive human serum samples (control) in a peptide microarray. The ZIKV NS1 peptides were tested against each individual sample. Figure 2A shows the schematic representation of the assay, whereas Figure 2B shows the fluorescence readings for the putative epitopes. Values are presented as an average of individual measurements for each serum along with the corresponding standard error of the mean (SEM). SEM is a statistical method that takes into account the sample size and standard deviation to provide accuracy of the mean. (Given the small sample size (7 ZIKV positives and 6 controls), results from most statistical analyses, such as *t*-test, cannot be applicable with confidence). Accession numbers for

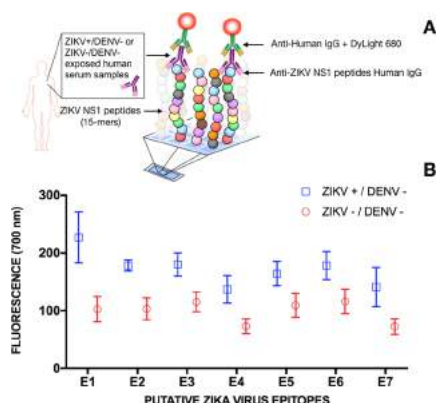


Figure 2. IgG response from sera of 7 confirmed ZIKV positive convalescent patients and 6 confirmed ZIKV and DENV negative patients to seven 15-mer peptides, whose sequences correspond to the predicted ZIKV NS1 epitopes. (A) The experimental detection scheme of the ZIKV NS1 15-mer peptide array. (B) Anti-ZIKV NS1 IgG detection for each ZIKV putative epitope expressed as the fluorescence average over the patients' serum samples along with their corresponding standard mean error and statistical significance.

the sequences used in the peptide microarray are listed in Table S1.

The results of the ZIKV peptide microarray technique show that all of the seven *in silico* predicted regions were recognized by IgG antibodies present in human serum samples from ZIKV-infected individuals. Although larger differences between positive and negative serum samples are observed for E1–E4, discrimination was possible for all cases. Inference about the amplitude of the differences cannot be made due to a small sample size and the fact that (intrinsically flexible) peptides are used as markers for antibody recognition and therefore may only be present in a recognizable conformation, part of the time. Nevertheless, these results confirm the immunogenicity of

the identified regions and thus validate the computational prediction approach. (It is worth mentioning that for the limitations stated above, it is out of the scope of this work to draw conclusions regarding ZIKV immune response to these epitopes; this assay was used with the sole purpose of validating the computational predictions).

Sequence Identity Analysis. The identification of B-cell epitopes containing low sequence identity between ZIKV and DENV1–4 is of particular interest for the development of differential diagnostic tools between the two infections. Conversely, epitopes exhibiting high sequence identities between the two viruses and capable of eliciting cross-neutralization can potentially be used in multivalent vaccine formulation strategies. Alignment of the identified ZIKV NS1 putative epitopes with DENV NS1 revealed a considerably high identity with at least one strain (Figure 3). Therefore, all seven peptide sequences are unlikely to be successful for differential diagnosis between ZIKV and DENV patients. Interestingly, those seven regions comprise 31% of the entire NS1 protein and by excluding them from the alignment, the overall identity falls to ca. 16% (from ca. 55 to 39%, Figure 3). We consider sequence-bias prediction unlikely because the sequences of the seven epitopes are very different from each other. The evolutionary implications of having conserved B-cell epitopes between the two viruses are unknown. It suggests that these epitopes might be involved in important infection mechanisms because in principle, a lower identity would aid ZIKV to evade immune response from patients previously infected by DENV.

Nevertheless, it has been reported that anti-ZIKV NS1 mAbs were able to be virus type-specific, not showing cross-reactivity with DENV NS1.¹⁶ We have observed that individual linear sequences are conserved and clustered together in two regions, possibly forming two conformational epitopes. The high identities of the linear sequences of the epitopes of ZIKV are likely to cross-react with antibodies anti-DENV and suggest that such differential antibody recognition takes place via conformational epitopes. Although one cannot overrule an eventual nonpredicted epitope on NS1 from ZIKV and DENV, we consider it unlikely given that the three algorithms used predicted epitopes on the same regions of the two proteins. Therefore, we postulate that such discrimination at the molecular-level is likely explained by either a contrasting

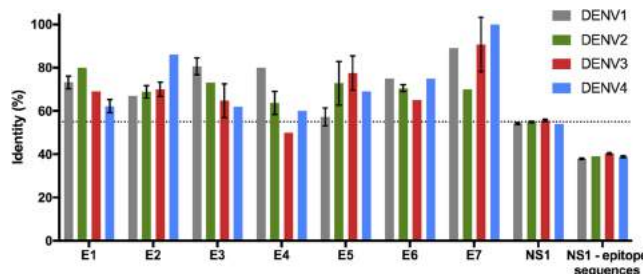


Figure 3. Sequence identity comparison between ZIKV and DENV1–4 NS1 proteins. (Standard deviation bars are presented for the 10 most prevalent DENV strains. NS1-epitope sequences correspond to the entire NS1 sequence excluding the seven *in silico* predicted putative epitope regions). A dotted line represents the average identity between the full-length NS1 of the Brazilian strain with the 10 most prevalent strains of DENV1–4.

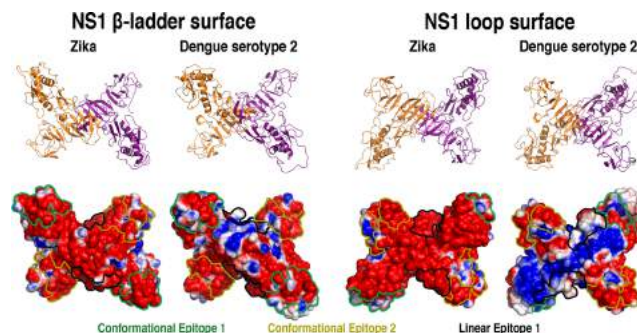


Figure 4. Cartoon representations and electrostatic surface potential on both faces of glycosylated ZIKV and DENV2 NS1 proteins. The locations of the predicted epitopes are color-contoured (green: CE1, yellow: CE2, and black: LE1). Negatively charged regions are shown in red, positively charged regions in blue, and neutral regions in white. (Plotted potential ranged from -2 to $+2$ $\text{kJ mol}^{-1} \text{e}^{-1}$.)

tertiary arrangement of those regions on the protein surface and/or a distinct charge signature profile.

Epitopes' Conformational Arrangement and Electrostatic Surface Profiles. To gain insights into the dynamics and conformational profiles of ZIKV and DENV NS1 proteins, MD simulations were performed using the dimeric form of the protein from both viruses. The DENV serotype 2 was chosen as the representative of DENV strains, as it is the only dimeric structure currently available in the Protein Data Bank (PDB). On the basis of the MD results, no significant differences could be observed by comparing the global structural protein dynamics (Figure S1). The root-mean-square deviation, radius of gyration (R_g), and root-mean-square fluctuation for both systems showed equivalent behaviors for the NS1 dimers and each monomeric unit within the dimer. These findings are in agreement with the work by Brown et al.¹⁸ and Song et al.,³¹ showing that in its protein fold and domain arrangement, ZIKV NS1 is virtually identical to DENV2 NS1. Still, a small per-residue solvent accessibility (Figure S2) and fluctuational (Figure S1) differences may be observed in some of the seven predicted epitope regions, when comparing ZIKV and DENV2 NS1. However, differences in solvation on the epitope regions are subtle (Figure S3) and do not warrant a differential tertiary arrangement of such epitopes onto ZIKV and DENV2 NS1 surfaces.

Because the putative epitopes were shown to be conserved but not the remaining residues, the electrostatic surface potentials were calculated and plotted onto the molecular surface of representative frames of ZIKV and DENV2 NS1 extracted from the MD simulations (Figure 4). A significant divergence between ZIKV and DENV2 NS1 can be observed. Despite approximately 55% of sequence identity between the proteins, ZIKV NS1 shows a remarkably negatively charged profile, whereas DENV2 presents extensively positively charged regions, especially in the loop region. Accordingly, the same charge-distribution pattern was described by Brown et al. when evaluating the electrostatic surface potential of entire NS1 proteins from ZIKV and DENV.¹⁸ Because the outer face of NS1 is more likely to elicit an immune response, regardless of NS1 being on the cell surface or being secreted, this difference in electrostatic distribution between ZIKV and DENV has implications for immune system recognition and vaccine

development and points toward different modes of flavivirus pathogenesis.

Most of these regions, however, occur away from the predicted epitopes (Figure 4) and shall be involved in other processes beyond antibody recognition. In addition, significant differences in the electrostatic potentials are also observed for the predicted epitopes (Figure 4). CE1 contains a higher amount of positively charged surface in the β -ladder view in DENV2 and significantly different conformational arrangements in the two proteins in the loop surface due to the Asn130-linked glycan orientation. For CE2, although the spatial distribution is quite similar in both surface views, the charge distribution differs between ZIKV and DENV2. LE1 also shows significant differences both in terms of conformation and electrostatic potential, being more exposed in the loop surfaces of both proteins but containing a considerably more positively charged surface in DENV2. Altogether, CE1 appears to have most conformational and charge-distribution differences between ZIKV and DENV2 (Figure 4) while being highly solvent-exposed in both (Figure S2), especially in the region between residues 108 and 129 (Figure S3).

CONCLUSIONS

Putative ZIKV NS1 B-cell epitopes were mapped onto the NS1 protein surface. Our results revealed that: (i) ZIKV NS1 immunogenic regions are comprised by two conformational epitopes and a linear one; (ii) epitope sequences were found to be conserved between ZIKV and DENV; (iii) despite the high epitope identity and structural similarity, ZIKV and DENV2 NS1 proteins display a loop–surface interface with significantly different electrostatic potentials; (iv) these differences are suggested to explain the molecular basis for unique binding properties to host factors and specific protective monoclonal antibody recognition therefore providing a lead for rational epitope-focused design of diagnostic tools and structure-based vaccine antigens for ZIKV infection.

MATERIALS AND METHODS

Epitope Prediction. Atomic coordinates for the Brazilian ZIKV NS1 strain obtained experimentally (PDB ID: 5K6K¹⁸) were used in three different structure-based tools for epitope

prediction: ElliPro,³⁴ EpiTope,³⁵ and DiscoTope.³⁶ Additionally, the solvent-accessible surface area was calculated using the postprocessing tool *g_sas* within GROMACS 5.1.2 package,³⁹ to evaluate the exposure of each putative epitope to the solvent.

NS1 ZIKV Peptide Microarray. The publicly-available whole proteome sequences of 15 ZIKV strains (see accession numbers on Table S1) were retrieved from the NCBI database and translated into 15-mer peptides with a peptide-peptide overlap of 12 amino acids. The whole proteome array data will be published elsewhere. Peptide arrays were produced by the company PEPperPRINT GmbH (Heidelberg, Germany) in a laser printing process on glass slides, coated with a poly(ethylene glycol) methacrylate/poly(methyl methacrylate) graft copolymer, which were functionalized with a β Ala- β Ala-linker. In brief, a layer of amino acid particles containing Fmoc-amino acid pentafluorophenyl esters was printed layer after layer onto the functionalized glass slides, with intermittent melting (i.e., coupling) steps at 90 °C and chemical washing and capping steps.⁴⁰ Peptides were generated in duplicates on the arrays, which were screened for IgG responses in human sera.

Immunostaining of Peptide Microarrays. Peptide microarrays were placed in incubation trays (PEPperPRINT GmbH; Heidelberg, Germany) and blocked for 30 min at room temperature at 120 rpm orbital shaking, with western blot blocking buffer MB-070 (Rockland). Then, sera were diluted by 1:1000 in phosphate-buffered saline (PBS) buffer with 0.05% Tween 20, pH 7.4 (PBS with Tween 20 (PBST)), and 10% blocking buffer, incubating the sera for 16 h at 4 °C and 50 rpm orbital shaking. Peptide microarrays were washed three times with PBST, followed by incubation with a 1:2500 dilution of the secondary antibody (goat anti-human IgG Fc specific DyLight 680; Rockland) together with the control antibody diluted by 1:500 (anti-c-Myc antibody, PEPperPRINT, Germany) for 30 min at room temperature and 120 rpm orbital shaking. The peptide microarrays were washed for 3 $\times$ 10 s with PBST and rinsed with deionized water. After drying in a stream of air, fluorescent images were acquired using an Odyssey Imaging System (LI-COR) at 700 nm, with a resolution of 21 μ m and a scanning sensitivity of 7. Image analysis and quantification was performed with PepSlide Analyzer software (Sicasys Software GmbH; Heidelberg, Germany).

Samples used for the peptide array experiments were collected from individuals with acute febrile illnesses enrolled in a prospective cohort study from May 2015 to May 2016. The cohort was established in an urgent health care clinic in the Recife metropolitan region as part of the International Research Consortium on Dengue Risk Assessment, Management, and Surveillance-IDAMS.^{41,42} The ages of patients from whom sera were used for the assays varied from 9 to 57 years, wherein 16 were females and 13 were males. Sample collection was performed on the first day of recruitment (Day 1 – acute sample), which following the IDAMS protocol corresponds to the period within the first 72 h of the febrile period, and in the convalescent phase (days 10–30 after recruitment – convalescent sample). For molecular viral diagnosis, quantitative real-time PCR (qRT-PCR) for DENV and ZIKV was performed following modified previously reported protocols.^{43,44} As for serology, samples were assayed for antiDENV IgM and IgG and antiZIKV IgM through ELISA. The Panbio Dengue Capture ELISA was used for the antiDENV IgM and IgG assays, following the manufacturer's protocol. The antiZIKV IgM MAC-ELISA protocol was that of the Centers

for Diseases Control and Prevention (CDC).^{43,45} According to assay results, each sample was classified as: (1) *ZIKV Positive/DENV Negative*, if they were positive for ZIKV qRT-PCR and negative for DENV qRT-PCR in the acute phase and/or positive for antiZIKV IgM with titers > 2 times those for antiDENV IgM in the convalescent phase in the MAC-ELISA, and negative for antiDENV IgG in the acute phase in the Panbio assay; (2) *ZIKV Negative/DENV Negative*, if they were negative for DENV and ZIKV qRT-PCRs in the acute phase and negative for antiDENV and antiZIKV IgM in the acute phase in MAC-ELISA and negative for antiDENV IgG in both acute and convalescent phases in the Panbio assay. The Panbio indirect antiDENV IgM and IgG test was not performed. Additionally, one serum sample used in this study was collected from a 21 year old male individual enrolled in the prospective cohort study PRONEX established in the city of Recife, Brazil. The sample was collected on day 2 after the onset of symptoms and its serological status determined as negative for DENV IgM and IgG ELISAs and positive in the ZIKV IgM ELISA.

Sequence Identity Analysis. ZIKV NS1 selected regions' sequences were aligned with DENV serotypes (DENV1, DENV2, DENV3, and DENV4) using the basic local alignment search tool provided by National Center for Biotechnology Information – NCBI⁴⁶ to obtain their sequence identity content. The same procedure was done with the complete sequence and for the sequence excluding the putative epitopes. Accession codes: DENV1 (AHK09916.1, AHK09915.1, AKA43760.1, AHK09902.1, AHK09902.1, AHK09909.1, AHK09914.1, AHK09919.1, AMO26073.1, AKK23335.1), DENV2 (AAD11533.1, AAA42947.1, AAL58460.1, AAD11532.1, AEZ68639.1, AJP06043.1, AAD11531.1, AAR17682.1, CAA35218.1, AHK09925.1), DENV3 (AHK09941.1, AAB52248.2, AHK09929.1, AAC04352.2, AAB52247.1, AAR17683.1, AHK09939.1, AAB52244.2, AAB2245.2, AHK09940.1), DENV4 (AHK09946.1, AHG23296.1, AKQ00032.1, AKQ00028.1, AAW51420.1, AHG23273.1, AHG23297.1, AHG23295.1, AEA50927.1, ACL99035.1).

NS1 Glycosylation, Modeling, and MD Simulations. NS1 structures for Dengue serotype 2 and ZIKV were retrieved from PDB ID 4O6B³⁸ and 5K6K,¹⁸ respectively. The missing loops in those structures were modeled using ModLoop tool,^{47,48} whereas the N-glycosylation motifs located at Asn130 (containing Asn130-Asn131-Ser132 in ZIKV and Asn130-Gln131-Thr132 in DENV2) and Asn207 (containing Asn207-Asp208-Thr209 in both proteins) were modified using glycosciences modeling tools,⁴⁹ respectively, with a complex-type and with a high-mannose type oligosaccharide⁵⁰ (Figure S3). The electrostatic potential surfaces were calculated using APBS.⁵¹ All MD calculations and remaining analyses were performed using GROMACS simulation suite, version 5.1.2.³⁹ All atomistic simulations were performed using GROMOS 54A7 force field for the protein part⁵² and GROMOS 53A6_{GLYC} parameters set for the glycan moieties.^{53,54} The glycosylated NS1 structures were explicitly solvated in 18 nm $\times$ 18 nm $\times$ 18 nm cubic boxes, using periodic boundary conditions in x, y, and z directions and the SPC water model,⁵⁵ employing at least 14 Å distance from the outside of each structure and the box edge. To neutralize the charge of these systems, Na⁺ counterions were added, as necessary. The LINCS method⁵⁶ was applied to constrain covalent bond lengths, allowing an integration step of 2 fs after an initial energy minimization using steepest descent algorithm. A short-range cutoff radius of 1.4 nm was used for all

prediction: ElliPro,³⁴ EpiTope,³⁵ and DiscoTope.³⁶ Additionally, the solvent-accessible surface area was calculated using the postprocessing tool *g_sas* within GROMACS 5.1.2 package,³⁹ to evaluate the exposure of each putative epitope to the solvent.

NS1 ZIKV Peptide Microarray. The publicly-available whole proteome sequences of 15 ZIKV strains (see accession numbers on Table S1) were retrieved from the NCBI database and translated into 15-mer peptides with a peptide-peptide overlap of 12 amino acids. The whole proteome array data will be published elsewhere. Peptide arrays were produced by the company PEPperPRINT GmbH (Heidelberg, Germany) in a laser printing process on glass slides, coated with a poly(ethylene glycol) methacrylate/poly(methyl methacrylate) graft copolymer, which were functionalized with a β Ala- β Ala-linker. In brief, a layer of amino acid particles containing Fmoc-amino acid pentafluorophenyl esters was printed layer after layer onto the functionalized glass slides, with intermittent melting (i.e., coupling) steps at 90 °C and chemical washing and capping steps.⁴⁰ Peptides were generated in duplicates on the arrays, which were screened for IgG responses in human sera.

Immunostaining of Peptide Microarrays. Peptide microarrays were placed in incubation trays (PEPperPRINT GmbH; Heidelberg, Germany) and blocked for 30 min at room temperature at 120 rpm orbital shaking, with western blot blocking buffer MB-070 (Rockland). Then, sera were diluted by 1:1000 in phosphate-buffered saline (PBS) buffer with 0.05% Tween 20, pH 7.4 (PBS with Tween 20 (PBST)), and 10% blocking buffer, incubating the sera for 16 h at 4 °C and 50 rpm orbital shaking. Peptide microarrays were washed three times with PBST, followed by incubation with a 1:2500 dilution of the secondary antibody (goat anti-human IgG Fc specific DyLight 680; Rockland) together with the control antibody diluted by 1:500 (anti-c-Myc antibody, PEPperPRINT, Germany) for 30 min at room temperature and 120 rpm orbital shaking. The peptide microarrays were washed for 3 $\times$ 10 s with PBST and rinsed with deionized water. After drying in a stream of air, fluorescent images were acquired using an Odyssey Imaging System (LI-COR) at 700 nm, with a resolution of 21 μ m and a scanning sensitivity of 7. Image analysis and quantification was performed with PepSlide Analyzer software (Sicasys Software GmbH; Heidelberg, Germany).

Samples used for the peptide array experiments were collected from individuals with acute febrile illnesses enrolled in a prospective cohort study from May 2015 to May 2016. The cohort was established in an urgent health care clinic in the Recife metropolitan region as part of the International Research Consortium on Dengue Risk Assessment, Management, and Surveillance-IDAMS.^{41,42} The ages of patients from whom sera were used for the assays varied from 9 to 57 years, wherein 16 were females and 13 were males. Sample collection was performed on the first day of recruitment (Day 1 – acute sample), which following the IDAMS protocol corresponds to the period within the first 72 h of the febrile period, and in the convalescent phase (days 10–30 after recruitment – convalescent sample). For molecular viral diagnosis, quantitative real-time PCR (qRT-PCR) for DENV and ZIKV was performed following modified previously reported protocols.^{43,44} As for serology, samples were assayed for antiDENV IgM and IgG and antiZIKV IgM through ELISA. The Panbio Dengue Capture ELISA was used for the antiDENV IgM and IgG assays, following the manufacturer's protocol. The antiZIKV IgM MAC-ELISA protocol was that of the Centers

for Diseases Control and Prevention (CDC).^{43,45} According to assay results, each sample was classified as: (1) *ZIKV Positive/DENV Negative*, if they were positive for ZIKV qRT-PCR and negative for DENV qRT-PCR in the acute phase and/or positive for antiZIKV IgM with titers > 2 times those for antiDENV IgM in the convalescent phase in the MAC-ELISA, and negative for antiDENV IgG in the acute phase in the Panbio assay; (2) *ZIKV Negative/DENV Negative*, if they were negative for DENV and ZIKV qRT-PCRs in the acute phase and negative for antiDENV and antiZIKV IgM in the acute phase in MAC-ELISA and negative for antiDENV IgG in both acute and convalescent phases in the Panbio assay. The Panbio indirect antiDENV IgM and IgG test was not performed. Additionally, one serum sample used in this study was collected from a 21 year old male individual enrolled in the prospective cohort study PRONEX established in the city of Recife, Brazil. The sample was collected on day 2 after the onset of symptoms and its serological status determined as negative for DENV IgM and IgG ELISAs and positive in the ZIKV IgM ELISA.

Sequence Identity Analysis. ZIKV NS1 selected regions' sequences were aligned with DENV serotypes (DENV1, DENV2, DENV3, and DENV4) using the basic local alignment search tool provided by National Center for Biotechnology Information – NCBI⁴⁶ to obtain their sequence identity content. The same procedure was done with the complete sequence and for the sequence excluding the putative epitopes. Accession codes: DENV1 (AHK09916.1, AHK09915.1, AKA43760.1, AHK09902.1, AHK09902.1, AHK09909.1, AHK09914.1, AHK09919.1, AMO26073.1, AKK23335.1), DENV2 (AAD11533.1, AAA42947.1, AAL58460.1, AAD11532.1, AEZ68639.1, AJP06043.1, AAD11531.1, AAR17682.1, CAA35218.1, AHK09925.1), DENV3 (AHK09941.1, AAB52248.2, AHK09929.1, AAC04352.2, AAB52247.1, AAR17683.1, AHK09939.1, AAB52244.2, AAB2245.2, AHK09940.1), DENV4 (AHK09946.1, AHG23296.1, AKQ00032.1, AKQ00028.1, AAW51420.1, AHG23273.1, AHG23297.1, AHG23295.1, AEA50927.1, ACL99035.1).

NS1 Glycosylation, Modeling, and MD Simulations. NS1 structures for Dengue serotype 2 and ZIKV were retrieved from PDB ID 4O6B³⁸ and 5K6K,¹⁸ respectively. The missing loops in those structures were modeled using ModLoop tool,^{47,48} whereas the N-glycosylation motifs located at Asn130 (containing Asn130-Asn131-Ser132 in ZIKV and Asn130-Gln131-Thr132 in DENV2) and Asn207 (containing Asn207-Asp208-Thr209 in both proteins) were modified using glycosciences modeling tools,⁴⁹ respectively, with a complex-type and with a high-mannose type oligosaccharide⁵⁰ (Figure S3). The electrostatic potential surfaces were calculated using APBS.⁵¹ All MD calculations and remaining analyses were performed using GROMACS simulation suite, version 5.1.2.³⁹ All atomistic simulations were performed using GROMOS 54A7 force field for the protein part⁵² and GROMOS 53A6_{GLYC} parameters set for the glycan moieties.^{53,54} The glycosylated NS1 structures were explicitly solvated in 18 nm $\times$ 18 nm $\times$ 18 nm cubic boxes, using periodic boundary conditions in x, y, and z directions and the SPC water model,⁵⁵ employing at least 14 Å distance from the outside of each structure and the box edge. To neutralize the charge of these systems, Na⁺ counterions were added, as necessary. The LINCS method⁵⁶ was applied to constrain covalent bond lengths, allowing an integration step of 2 fs after an initial energy minimization using steepest descent algorithm. A short-range cutoff radius of 1.4 nm was used for all

nonbonded interactions. Long-range electrostatic interactions were calculated through the reaction-field method,⁵⁷ with $\epsilon = 66$. The system was coupled to a thermostat with a reference temperature of 310 K using the velocity-rescale scheme,⁵⁸ with a relaxation time of 0.4 ps. Pressure was maintained constant at 1 bar by isotropic coordinate scaling using Parrinello–Rahman barostat,^{59,60} with a coupling constant of 1.0 ps and a compressibility of 4.5×10^{-5} bar. A 1 ns MD simulation with positional restraints on all heavy atoms, with a force constant of 1000 kJ mol^{-1} , was performed as an equilibration step and was not taken into account to calculate the average ensemble properties. Subsequently, all systems were further extended to 200 ns (without positional restraints). Representative structures were selected from the equilibrated part of trajectories (last 50 ns) by clustering all the frames generated within this time scale (employing *g_cluster* tool from GROMACS suite, with a 0.4 nm cutoff). Thus, each of ZIKV and DENV2 NS1 MD simulations provided a single conformational cluster and the representative frames used for electrostatic surface potential analyses.

■ ASSOCIATED CONTENT

■ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acsomega.7b00608.

MD structural data analyses; glycosylation sequence used in NS1 proteins; and accession numbers for the ZIKV sequences used in the peptide array (PDF)

■ AUTHOR INFORMATION

Corresponding Authors

*E-mail: marques@pitt.edu. Tel +1 (412) 624.4440 (E.T.A.M.).

*E-mail: roberto.lins@cpqam.fiocruz.br. Tel +55 (81) 2123.7848 (R.D.L.).

ORCID

Roberto D. Lins: 0000-0002-3983-8025

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by grants from FACEPE, CAPES, CNPq and CuraZika Foundation. Computational resources were provided by the Brazilian National Supercomputing Center (LNCC).

■ REFERENCES

- (1) Lover, A. A. Zika virus and microcephaly. *Lancet Infect. Dis.* **2016**, *16*, 1331–1332.
- (2) de Araújo, T. V.; Rodrigues, L. C.; de Alencar Ximenes, R. A.; de Barros Miranda-Filho, D.; Montarroyos, U. R.; de Melo, A. P.; Valongueiro, S.; de Albuquerque, M. F.; Souza, W. V.; Braga, C.; Filho, S. P.; Cordeiro, M. T.; Vazquez, E.; Di Cavalcanti Souza Cruz, D.; Henriques, C. M.; Bezerra, L. C.; da Silva Castanha, P. M.; Dhalia, R.; Marques-Júnior, E. T.; Martelli, C. M. Association between Zika virus infection and microcephaly in Brazil, January to May, 2016: preliminary report of a case-control study. *Lancet Infect. Dis.* **2016**, *16*, 1356–1363.
- (3) Lessler, J.; Chaisson, L. H.; Kucirka, L. M.; Bi, Q.; Grantz, K.; Salje, H.; Carcelen, A. C.; Ott, C. T.; Sheffield, J. S.; Ferguson, N. M.; Cummings, D. A.; Metcalf, C. J.; Rodriguez-Barraquer, I. Assessing the global threat from Zika virus. *Science* **2016**, *353*, No. aaf8160.

- (4) Chouin-Carreiro, T.; Vega-Rua, A.; Vazeille, M.; Yebakima, A.; Girod, R.; Goindin, D.; Dupont-Rouzeyrol, M.; Lourenco-de-Oliveira, R.; Failloux, A. B. Differential Susceptibilities of *Aedes aegypti* and *Aedes albopictus* from the Americas to Zika Virus. *PLoS Neglected Trop. Dis.* **2016**, *10*, No. e0004543.
- (5) Lazear, H. M.; Diamond, M. S. Zika Virus: New Clinical Syndromes and Its Emergence in the Western Hemisphere. *J. Virol.* **2016**, *90*, 4864–4875.
- (6) Dawes, B. E.; Smalley, C. A.; Tiner, B. L.; Beasley, D. W.; Milligan, G. N.; Reece, L. M.; Hombach, J.; Barrett, A. D. Research and development of Zika virus vaccines. *npj Vaccines* **2016**, *1*, No. 16007.
- (7) Waggoner, J. J.; Pinsky, B. A. Zika Virus: Diagnostics for an Emerging Pandemic Threat. *J. Clin. Microbiol.* **2016**, *54*, 860–867.
- (8) Johnson, A. J.; Martin, D. A.; Karabatsos, N.; Roehrig, J. T. Detection of anti-arboviral immunoglobulin G by using a monoclonal antibody-based capture enzyme-linked immunosorbent assay. *J. Clin. Microbiol.* **2000**, *38*, 1827–1831.
- (9) Martin, D. A.; Muth, D. A.; Brown, T.; Johnson, A. J.; Karabatsos, N.; Roehrig, J. T. Standardization of immunoglobulin M capture enzyme-linked immunosorbent assays for routine diagnosis of arboviral infections. *J. Clin. Microbiol.* **2000**, *38*, 1823–1826.
- (10) Kuno, G. Serodiagnosis of flaviviral infections and vaccinations in humans. *Adv. Virus Res.* **2003**, *61*, 3–65.
- (11) Maeda, A.; Maeda, J. Review of diagnostic plaque reduction neutralization tests for flavivirus infection. *Vet. J.* **2013**, *195*, 33–40.
- (12) Wong, S. J.; Furuya, A.; Zou, J.; Xie, X.; Dupuis Li, A. P.; Kramer, L. D.; Shi, P.-Y. A Multiplex Microsphere Immunoassay for Zika Virus Diagnosis. *EBioMedicine* **2017**, *16*, 136–140.
- (13) Sumita, L. M.; Rodrigues, J. P.; Ferreira, N. E.; Felix, A. C.; Souza, N. C. S.; Machado, C. M.; Júnior, H. F. d. A. Detection of human anti-zika virus IgG by elisa using an antigen from in vitro infected vero cells: preliminary results. *Rev. Inst. Med. Trop. Sao Paulo* **2016**, *58*, 89.
- (14) Johnson, B. W.; Kosoy, O.; Hunsperger, E.; Beltran, M.; Delorey, M.; Guirakhoo, F.; Monath, T. Evaluation of chimeric Japanese encephalitis and dengue viruses for use in diagnostic plaque reduction neutralization tests. *Clin. Vaccine Immunol.* **2009**, *16*, 1052–1059.
- (15) Musso, D.; Gubler, D. J. Zika Virus. *Clin. Microbiol. Rev.* **2016**, *29*, 487–524.
- (16) Stettler, K.; Beltramello, M.; Espinosa, D. A.; Graham, V.; Cassotta, A.; Bianchi, S.; Vanzetta, F.; Minola, A.; Jaconi, S.; Mele, F.; Foglierini, M.; Pedotti, M.; Simonelli, L.; Dowall, S.; Atkinson, B.; Percivalle, E.; Simmons, C. P.; Varani, L.; Blum, J.; Baldanti, F.; Camerini, E.; Hewson, R.; Harris, E.; Lanzavecchia, A.; Sallusto, F.; Corti, D. Specificity, cross-reactivity, and function of antibodies elicited by Zika virus infection. *Science* **2016**, *353*, 823–826.
- (17) Kuno, G.; Chang, G. J. Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. *Arch. Virol.* **2007**, *152*, 687–696.
- (18) Brown, W. C.; Akey, D. L.; Konwerski, J. R.; Tarrasch, J. T.; Skiniotis, G.; Kuhn, R. J.; Smith, J. L. Extended surface for membrane association in Zika virus NS1 structure. *Nat. Struct. Mol. Biol.* **2016**, *23*, 865–867.
- (19) Muller, D. A.; Young, P. R. The flavivirus NS1 protein: molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral Res.* **2013**, *98*, 192–208.
- (20) Watterson, D.; Modhiran, N.; Young, P. R. The many faces of the flavivirus NS1 protein offer a multitude of options for inhibitor design. *Antiviral Res.* **2016**, *130*, 7–18.
- (21) Gutsche, I.; Coulibaly, F.; Voss, J. E.; Salmon, J.; d'Alayer, J.; Ermonval, M.; Larquet, E.; Charneau, P.; Krey, T.; Megret, F.; Guittet, E.; Rey, F. A.; Flamand, M. Secreted dengue virus nonstructural protein NS1 is an atypical barrel-shaped high-density lipoprotein. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 8003–8008.
- (22) Young, P. R.; Hilditch, P. A.; Bletchly, C.; Halloran, W. An antigen capture enzyme-linked immunosorbent assay reveals high

levels of the dengue virus protein NS1 in the sera of infected patients. *J. Clin. Microbiol.* **2000**, *38*, 1053–1057.

(23) Edeling, M. A.; Diamond, M. S.; Fremont, D. H. Structural basis of Flavivirus NS1 assembly and antibody recognition. *Proc. Natl. Acad. Sci. U.S.A.* **2014**, *111*, 4285–4290.

(24) Muller, D. A.; Young, P. R. The flavivirus NS1 protein: molecular and structural biology, immunology, role in pathogenesis and application as a diagnostic biomarker. *Antiviral Res.* **2013**, *98*, 192–208.

(25) Cecchetto, J.; Fernandes, F. C.; Lopes, R.; Bueno, P. R. The capacitive sensing of NS1 Flavivirus biomarker. *Biosens. Bioelectron.* **2017**, *87*, 949–956.

(26) Cavalcanti, I. T.; Guedes, M. I.; Sotomayor, M. D.; Yamanaka, H.; Dutra, R. F. A label-free immunosensor based on recordable compact disk chip for early diagnostic of the dengue virus infection. *Biochem. Eng. J.* **2012**, *67*, 225–230.

(27) Dias, A. C. M. S.; Gomes-Filho, S. L.; Silva, M. M.; Dutra, R. F. A sensor tip based on carbon nanotube-ink printed electrode for the dengue virus NS1 protein. *Biosens. Bioelectron.* **2013**, *44*, 216–221.

(28) Figueiredo, A.; Vieira, N. C.; Dos Santos, J. F.; Janegitz, B. C.; Aoki, S. M.; Junior, P. P.; Lovato, R. L.; Nogueira, M. L.; Zucolotto, V.; Guimarães, F. E. Electrical detection of dengue biomarker using egg yolk immunoglobulin as the biological recognition element. *Sci. Rep.* **2015**, *5*, No. 7865.

(29) Silva, M. M. S.; Dias, A.; Cordeiro, M.; Marques, E.; Goulart, M.; Dutra, R. A thiophene-modified screen printed electrode for detection of dengue virus NS1 protein. *Talanta* **2014**, *128*, S05–S10.

(30) Su, C.-C.; Wu, T.-Z.; Chen, L.-K.; Yang, H.-H.; Tai, D.-F. Development of immunochips for the detection of dengue viral antigens. *Anal. Chim. Acta* **2003**, *479*, 117–123.

(31) Song, H.; Qi, J.; Haywood, J.; Shi, Y.; Gao, G. F. Zika virus NS1 structure reveals diversity of electrostatic surfaces among flaviviruses. *Nat. Struct. Mol. Biol.* **2016**, *23*, 456–458.

(32) Viana, I. F.; Soares, T. A.; Lima, L. F.; Marques, E. T.; Krieger, M. A.; Dhalia, R.; Lins, R. D. De novo design of immunoreactive conformation-specific HIV-1 epitopes based on Top7 scaffold. *RSC Adv.* **2013**, *3*, 11790–11800.

(33) Flower, D. R. Designing immunogenic peptides. *Nat. Chem. Biol.* **2013**, *9*, 749–753.

(34) Ponomarenko, J.; Bui, H.-H.; Li, W.; Fusseder, N.; Bourne, P. E.; Sette, A.; Peters, B. ElliPro: a new structure-based tool for the prediction of antibody epitopes. *BMC Bioinf.* **2008**, *9*, No. S14.

(35) Rubinstein, N. D.; Mayrose, I.; Martz, E.; Pupko, T. EpiTope: a web-server for predicting B-cell epitopes. *BMC Bioinf.* **2009**, *10*, No. 287.

(36) Haste Andersen, P.; Nielsen, M.; Lund, O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. *Protein Sci.* **2006**, *15*, 2558–2567.

(37) Borley, D. W.; Mahapatra, M.; Paton, D. J.; Esnouf, R. M.; Stuart, D. I.; Fry, E. E. Evaluation and use of in-silico structure-based epitope prediction with foot-and-mouth disease virus. *PLoS One* **2013**, *8*, No. e61122.

(38) Akey, D. L.; Brown, W. C.; Dutta, S.; Konwerski, J.; Jose, J.; Jurkiw, T. J.; DelProposto, J.; Ogata, C. M.; Skiniotis, G.; Kuhn, R. J.; Smith, J. L. Flavivirus NS1 structures reveal surfaces for associations with membranes and the immune system. *Science* **2014**, *343*, 881–885.

(39) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

(40) Stadler, V.; Felgenhauer, T.; Beyer, M.; Fernandez, S.; Leibe, K.; Guttler, S.; Groning, M.; König, K.; Torralba, G.; Hausmann, M.; Lindentruth, V.; Nesterov, A.; Block, I.; Pipkorn, R.; Poustka, A.; Bischoff, F. R.; Breitling, F. Combinatorial synthesis of peptide arrays with a laser printer. *Angew. Chem., Int. Ed. Engl.* **2008**, *47*, 7132–7135.

(41) Jaenisch, T.; Sakuntabhai, A.; Wilder-Smith, A. DengueTools, Dengue research funded by the European Commission-scientific strategies of three European dengue research consortia. *PLoS Neglected Trop. Dis.* **2013**, *7*, No. e2320.

(42) Jaenisch, T.; Tam, D. T.; Kieu, N. T.; Van Ngoc, T.; Nam, N. T.; Van Kinh, N.; Yacoub, S.; Chanpheaktra, N.; Kumar, V.; See, L. L.; Sathar, J.; Sandoval, E. P.; Alfaro, G. M.; Laksono, I. S.; Mahendradhata, Y.; Sarker, M.; Ahmed, F.; Caprara, A.; Benevides, B. S.; Marques, E. T.; Magalhaes, T.; Brasil, P.; Netto, M.; Tami, A.; Bethencourt, S. E.; Guzman, M.; Simmons, C.; Quyen, N. T.; Merson, L.; Dung, N. T.; Beck, D.; Wirths, M.; Wolbers, M.; Lam, P. K.; Rosenberger, K.; Wills, B. Clinical evaluation of dengue and identification of risk factors for severe disease: protocol for a multicentre study in 8 countries. *BMC Infect. Dis.* **2016**, *16*, 120.

(43) Lanciotti, R. S.; Kosoy, O. L.; Laven, J. J.; Panella, A. J.; Velez, J. O.; Lambert, A. J.; Campbell, G. L. Chikungunya virus in US travelers returning from India, 2006. *Emerging Infect. Dis.* **2007**, *13*, 764–767.

(44) Lanciotti, R. S.; Kosoy, O. L.; Laven, J. J.; Velez, J. O.; Lambert, A. J.; Johnson, A. J.; Stanfield, S. M.; Duffy, M. R. Genetic and serologic properties of Zika virus associated with an epidemic, Yap State, Micronesia, 2007. *Emerging Infect. Dis.* **2008**, *14*, 1232–1239.

(45) Cordeiro, M. T.; Brito, C. A.; Pena, L. J.; Castanha, P. M.; Gil, L. H.; Lopes, K. G.; Dhalia, R.; Meneses, J. A.; Ishigami, A. C.; Mello, L. M.; Alencar, L. X.; Guarines, K. M.; Rodrigues, L. C.; Marques, E. T. Results of a Zika Virus (ZIKV) Immunoglobulin M-Specific Diagnostic Assay Are Highly Correlated With Detection of Neutralizing Anti-ZIKV Antibodies in Neonates With Congenital Disease. *J. Infect. Dis.* **2016**, *214*, 1897–1904.

(46) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **1997**, *25*, 3389–3402.

(47) Fiser, A.; Do, R. K.; Sali, A. Modeling of loops in protein structures. *Protein Sci.* **2000**, *9*, 1753–1773.

(48) Fiser, A.; Sali, A. ModLoop: automated modeling of loops in protein structures. *Bioinformatics* **2003**, *19*, 2500–2501.

(49) Lütke, T.; Bohne-Lang, A.; Loss, A.; Goetz, T.; Frank, M.; von der Lieth, C.-W. GLYCOSCIENCES.de: an Internet portal to support glycomics and glycobiology research. *Glycobiology* **2006**, *16*, 71R–81R.

(50) Pryor, M. J.; Wright, P. J. Glycosylation mutants of dengue virus NS1 protein. *J. Gen. Virol.* **1994**, *75*, 1183–1187.

(51) Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037–10041.

(52) Schmid, N.; Eichenberger, A. P.; Choutko, A.; Riniker, S.; Winger, M.; Mark, A. E.; van Gunsteren, W. F. Definition and testing of the GROMOS force-field versions 54A7 and 54B7. *Eur. Biophys. J.* **2011**, *40*, 843–856.

(53) Pol-Fachin, L.; Rusu, V. H.; Verli, H.; Lins, R. D. GROMOS 53A6GLYC, an Improved GROMOS Force Field for Hexopyranose-Based Carbohydrates. *J. Chem. Theory Comput.* **2012**, *8*, 4681–4690.

(54) Pol-Fachin, L.; Verli, H.; Lins, R. D. Extension and validation of the GROMOS 53A6(GLYC) parameter set for glycoproteins. *J. Comput. Chem.* **2014**, *35*, 2087–2095.

(55) Berendsen, H.; Grigera, J.; Straatsma, T. The missing term in effective pair potentials. *J. Phys. Chem.* **1987**, *91*, 6269–6271.

(56) Hess, B.; Bekker, H.; Berendsen, H. J.; Fraaije, J. G. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(57) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. A generalized reaction field method for molecular dynamics simulations. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

(58) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, No. 014101.

(59) Parrinello, M.; Rahman, A. Polymorphic transitions in single crystals: A new molecular dynamics method. *J. Appl. Phys.* **1981**, *52*, 7182–7190.

(60) Nosé, S.; Klein, M. Constant pressure molecular dynamics for molecular systems. *Mol. Phys.* **1983**, *50*, 1055–1076.

Titulo: Detection of IgG3 antibodies specific to the human immunodeficiency virus type 1 (HIV-1) p24 protein as marker for recently acquired infection

Autores: I. F. T. Viana, **D. F. Coêlho**, M. L. Palma, E. J. M. Nascimento, G. Gu, L. F. O. Lima, L. Foti, M. A. Krieger, C. Pilcher, C. E. Calzavara-Silva, R. B. Mailliard, C. R. Rinaldo, R. Dhalia, E. T. A. Marques.

Publicado no periódico: Epidemiology and Infection, 2018.
doi: 10.1017/S0950268818001218



Original Paper

Cite this article: Viana IFT *et al* (2018). Detection of IgG3 antibodies specific to the human immunodeficiency virus type 1 (HIV-1) p24 protein as marker for recently acquired infection. *Epidemiology and Infection* 1–8. <https://doi.org/10.1017/S0950268818001218>

Received: 21 December 2017

Revised: 8 April 2018

Accepted: 17 April 2018

Key words:

Early infection marker; ELISA; HIV/AIDS; recent infection

Author for correspondence:

Isabelle F. T. Viana, isabelle.viana@cpqam.fiocruz.br

Detection of IgG3 antibodies specific to the human immunodeficiency virus type 1 (HIV-1) p24 protein as marker for recently acquired infection

I. F. T. Viana^{1,2,3}, D. F. Coêlho^{1,2,4}, M. L. Palma³, E. J. M. Nascimento^{2,3}, G. Gu^{2,5}, L. F. O. Lima^{6,7}, L. Foti^{6,8}, M. A. Krieger^{6,8}, C. Pilcher⁹, C. E. Calzavara-Silva¹⁰, R. B. Mailliard³, C. R. Rinaldo³, R. Dhalia¹ and E. T. A. Marques^{1,2,3}

¹Aggeu Magalhães Institute, Oswaldo Cruz Foundation/FIOCRUZ, Brazil; ²Center for Vaccine Research, University of Pittsburgh, USA; ³Infectious Diseases and Microbiology Department, University of Pittsburgh, USA; ⁴Fundamental Chemistry Department, Federal University of Pernambuco, Recife, Brazil; ⁵Tsinghua University, School of Medicine, Beijing, China; ⁶Carlos Chagas Institute, Fiocruz, Brazil; ⁷Technological Institute of Paraná/TECPAR, Brazil; ⁸Molecular Biology Institute of Paraná/IBMP, Brazil; ⁹UCSF School of Medicine, USA and ¹⁰René Rachou Research Center, FIOCRUZ, Brazil

Abstract

Reducing the risk of human immunodeficiency virus type 1 (HIV-1) transmission is still a public health priority. The development of effective control strategies relies on the quantification of the effects of prophylactic and therapeutic measures in disease incidence. Although several assays can be used to estimate HIV incidence, these estimates are limited by the poor performance of these assays in distinguishing recent from long-standing infections. To address such limitation, we have developed an assay to titrate p24-specific IgG3 antibodies as a marker of recent infection. The assay is based on a recombinant p24 protein capable to detect total IgG antibodies in sera using a liquid micro array and enzyme-linked immunosorbent assay. Subsequently, the assay was optimised to detect and titrate anti-p24 IgG3 responses in a panel of sequential specimens from seroconverters over 24 months. The kinetics of p24-specific IgG3 titres revealed a transient peak in the 4 to 5-month period after seroconversion. It was followed by a sharp decline, allowing infections with less than 6 months to be distinguished from older ones. The developed assay exhibited a mean duration of recent infection of 144 days and a false-recent rate of ca. 14%. Our findings show that HIV-1 p24-specific IgG3 titres can be used as a tool to evaluate HIV incidence in serosurveys and to monitor the efficacy of vaccines and other transmission control strategies.

Introduction

Reducing human immunodeficiency virus type 1 (HIV-1) transmission and increasing patients' life expectancy still remain a public health priority and a major challenge [1]. Several strategies have shown to be able to block HIV-1 transmission. However, the implementation of these strategies has only provided a limited reduction of the HIV-1 incidence. Aiming to reach this goal, the proper monitoring of the changing patterns of HIV transmission under prophylactic and therapeutic interventions is of utmost importance. Accurate distinguishing of recent from chronic infections allows for precise estimates of the disease's incidence rate. It is also crucial to evaluate the efficacy of HIV intervention and prevention measures and serves as an objective reference for allocating HIV-related health care resources [2]. It is well known that HIV antigenemia occurs during the initial stages of infection and antigen detection assays are a useful public health index. Nevertheless, that does not reflect the rate of current HIV transmission, as antigenemia can happen late in the course of infection, i.e., as progression to AIDS occurs [3]. In this regard, laboratory tests capable of detecting recently-acquired HIV-1 infection and also distinguishing it from the long-standing infections have been extensively developed. Despite that, a reliable method to estimate HIV-1 incidence remains as an immediate need in the community [4].

The current gold standard approach to identify recent infections involves the longitudinal follow-up and repeated testing of uninfected individuals [5]. However, several difficulties such as the logistical challenges of following a cohort of individuals at risk for acquiring HIV infection and the need for testing a large number of HIV negative individuals, lead to inaccuracy of modelled-based estimates. These practices are also time-consuming and resource intensive, therefore representing important bottlenecks for broader routine HIV-1 screening. Recent efforts around incidence estimation have focused on the detection of recent seroconversions among HIV-1 antibody positive individuals. The incidence estimate is based on the

maturation of evolving antibodies that usually follows a predictable pattern [6], antibody titres and antibody avidity [7]. It relies on the assumption that the HIV-1-specific antibody response evolves following exposure to the virus towards a progressive increase in concentration and affinity [6–9]. An individual is then classified as recently infected (i.e., infected within a defined mean duration of recent infection (MDRI)) if the biomarker value is below the predetermined assay threshold. Although these methods based on biomarkers of recent infections are extensively used to determine the incidence, their recent performance assessment [9] revealed the need for optimisation in order to reduce the tendency to produce substantial false-recent rates (FRR). The main causes are susceptibility to variation in immune responses that are associated with different HIV subtypes, viral load levels and antiretroviral therapy (ART), which can then result in overestimated HIV-1 incidence [5]. These confounding variables may contribute to inaccurate incidence estimates and therefore have profound consequences for evaluating the impact of HIV prevention or surveillance program measures [1].

Due to the challenges associated with the current tests for recent HIV infection, there is an urgent need to improve the performance of current assays or to identify new biomarkers that can reliably distinguish recent from long-standing infections, thereby yielding low FRRs and acceptable MDRI [10]. In this scenario, differences in the kinetics of HIV-1-specific IgG subclass antibody responses during infection have been addressed as a potential strategy to estimate the relative timing of HIV seroconversion with less variation and misclassification. A customised multiplex assay directed against different HIV-1 antigens described a decline in HIV-1-specific IgG3 antibody responses during recent infection, contrasting with the stably elevated levels of HIV-1-specific IgG1 throughout the disease progression [11]. The shorter duration of antigen-specific IgG3 may reflect inherent differences in the subclass recycling and durability [12] or on the dependency of an effective T-helper-cell response [13, 14]. On that account, estimates of peak IgG3 antibody concentrations and the half-life of these antibodies during recent HIV-1 infection may be applied as part of an algorithm to substitute or complement the methods now in use to define recent infection and estimate HIV-1 incidence [11].

This report describes, for the first time, the performance characteristics of an affordable, simple and versatile enzyme-linked immunosorbent assay (ELISA) to measure the titre of anti-p24 HIV-1 IgG3 antibodies in a retrospective cohort of HIV-1 infected individuals whose sera samples comprised post-seroconversion times ranging from 1 to 24 months. It is based on a developed protocol for a designed recombinant p24 antigen that was able to quantitatively detect HIV-1-specific total IgG and IgG3 antibodies responses at the analysed post-seroconversion window times. Assay refinement is also described, including estimation of cut-off values and corresponding MDRI and FRR. Our results show that the combination of the assays to measure total IgG and IgG3 antibody titres against HIV-1 p24 could be applied for proper assessment of cohort patients' status and for sample classification in research settings. Given the public health importance of accurately estimating HIV-1 incidence, the importance of the method described here is strengthened by its capacity to detect recent infection and discriminate recent from late infection using a single antigen.

Methods

A schematic representation of the methodology is presented in Fig. S1.

HIV-1 p24 DNA sequence selection, optimization and cloning

99 amino acid sequences from American viral isolates encoding the HIV-1 capsid (p24) protein were downloaded from the National Center for Biotechnology Information – NCBI (<https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html>) (Table S1). The sequences were aligned using the Basic Local Alignment Search Tool version 2.2.32 provided by NCBI [15, 16] to generate the consensus sequence. A reverse search was then carried out to identify the primary viral isolate with the closest sequence match to the consensus sequence (access number: ACA49289.1). The selected DNA sequence encoding the full-length p24 protein is composed of 228 amino acid residues and was flanked by two multiple cloning sites (5' *Hind* III, *Xho* I, *Nde* I, *Spe* I and 3' *Sap* I, *Not* I, *Nco* I). The corresponding DNA sequence was optimised for codon usage, reduced messenger RNA secondary structures, distribution of GC content and removal of repetitive sequences and motifs aiming enhanced mRNA transcription, stability and translation in prokaryotic system using the software LETO 1.0 (Entelechon®). The artificially optimised gene (access number: MG201975) was further synthesised by a commercial supplier (GeneScript®) and cloned into the expression vector pRSETA (Invitrogen®).

Recombinant protein production

Bacterial expression of the recombinant capsid p24 protein was carried out in *Escherichia coli* BL21 Star™ (DE3) pLysS (Invitrogen®) strain transformed with the expression vector. A transformant colony harbouring the expression plasmid was grown in Luria-Bertani medium supplemented with 100 µg/ml of ampicillin to an OD_{600nm} of 0.5 (ca. 2×10^8 cells/ml) and protein expression was induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) at 37 °C for 4 h. The bacterial pellet was harvested by centrifugation at 5000 g/4 °C/20 min and resuspended in lysis buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20 mM Imidazole) supplemented with 1X protease inhibitor cocktail (Roche). Bacterial lysis was performed by sonication at 4 °C in a Vibracell VCX 750 Sonicator (amplitude setting of 40%) by 6 pulses of 30 s each with 60 s interval between the pulses. The lysate was clarified by centrifugation at 8000 g/4 °C/30 min and the recombinant protein was purified through affinity chromatography with a nickel matrix column (Ni-NTA – QIAGEN). Protein concentration was determined by spectrophotometry.

Human serum samples

Cryopreserved serum obtained from 45 HIV-1 positive participants and 235 HIV-1 negative participants were provided by the Brazilian National Panel for Blood Screening Quality Control. All sera were tested by the Technology Institute for Immunobiologicals (Bio-Manguinhos/Oswaldo Cruz Foundation-FIOCRUZ, Brazil) and the National Institute for Health Quality Control of the Oswaldo Cruz Foundation. Cryopreserved serum was also obtained from 48 HIV-1 positive participants whom plasma HIV RNA levels varied from 55 to 290 000 copies/ml and 48 HIV-1 negative participants and provided by the Multicenter AIDS Cohort Study (MACS). The seroconversion panel was composed of sera samples from 11 participants who had seroconverted based on HIV antibody screening and obtained from the Consortium for the Evaluation and Performance of HIV

Incidence Assays (CEPHIA). The samples included date back several weeks before antibody seroconversion and continued up to 6 months after the detection of HIV antibodies. CEPHIA has also provided 150 serial serum samples from 30 HIV-1 positive participants obtained from 1 month up to 24 months post-seroconversion. All HIV-1 reactive samples used in this study were obtained from individuals infected with clade B virus and who were not under ART and who had not progressed to AIDS. All serum samples were aliquoted, heat inactivated (56 °C for 40 min), allowed to cool to room temperature and then stored at -80 °C until use.

Liquid microarray (LMA) assay

For the covalent coupling of the chimerical protein to microspheres, paramagnetic carboxylated microspheres (Luminex Corp, Austin, USA) were coated with the purified proteins at 50 µg/ml through a two-step carbodiimide coupling, according to the manufacturer's instructions. The protein-coated microspheres concentration was determined using the MultisizerTM 3 Coulter Counter (Beckman Coulter) and the microspheres were stored at 4 °C in the dark until use. LMA standard protocol. Immunoreactions were performed in low binding 96-well plates (SARSTEDT) and all incubations were performed at 37 °C on a microplate shaker (set at 300 RPM) and protected from light. A blocking step using *E. coli* lysate was performed as previously described [17]. For the wash steps, a Hydroflex plate washer with a magnetic support (TECAN, Dürban, NC, USA) was used. Sera samples were diluted 1:200 in 1X PBS containing 1% BSA, 0.05% Tween-20 and the assays were conducted using standard procedures described by the manufacturer (Luminex Corp, Austin-TX, USA). Two thousand and five hundreds antigen-coated microspheres (in a final volume of 50 µl) were added to each well of a 96-well plate and incubated with 50 µl of each serum for 30 min. Following two washes in 100 µl of assay buffer (1X PBS containing 1% BSA, 0.05% Tween-20, 100 mM Tris) the microspheres were incubated with goat anti-human IgG conjugated to phycoerythrin (MOSS Substrates) diluted 1:1000 (0.66 µg/ml) in assay buffer for 30 min. The microspheres were washed twice and re-suspended in 100 µl of assay buffer before fluorescence measurement. The reporter fluorescence, expressed as median fluorescence intensity (MFI), was determined with a Luminex 200 instrument using xPonent 4.1 software, acquiring 100 independent events per well.

LMA data analysis

The diagnostic performance evaluation of the recombinant antigen was performed using the receiver-operating characteristics curve (ROC) analysis [18, 19] built in the Prism 6 for Mac OS X (version 6.0d; GraphPad Software Inc.). The determined area under the curve (AUC), sensibility and specificity values for the immunological assay were used as a standard metric. Additionally, two-tailed unpaired *t*-tests and one-way analysis of variance (ANOVA) were used to compare the MFI values obtained for the p24 protein against the HIV-1 panels. Pearson's correlation was used to assess the co-variation among the groups analysed in the ELISA assays. Differences were considered statistically significant when $P < 0.05$. All statistical tests were conducted at the 95% confidence interval. False positive rate = FP/(FP + TN).

ELISA for detection of p24-specific total IgG and IgG3

High binding, half area 96-well polystyrene plates (Costar; Lowell, MA, USA) were coated overnight at 4 °C with 1 µg/ml (for anti-p24 total IgG detection) or 2.3 µg/ml (for anti-p24 IgG3 detection) of p24 recombinant protein diluted in 0.2 M carbonate/bicarbonate buffer (Pierce, IL, USA). Plates were blocked with skimmed milk (Bio-Rad) at 5% (w/v) in PBS-T buffer [1X PBS with 0.05% (v/v) Tween 20] for 15 min at room temperature. Serum samples were diluted in assay buffer [5% (w/v) skimmed milk in PBS-T] and added to plates. All samples were incubated for 2 h at room temperature. Plates were washed five times with PBS-T, followed by incubation with the horseradish peroxidase (HRP)-linked antibody against total IgG or IgG3 diluted in assay buffer. Plates were incubated for 1 h at room temperature, washed four times with PBS-T and allowed to soak in the same buffer for 5 min at room temperature. An incubation of tetramethylbenzidine TMB-KPL substrate (Pierce, IL, USA) for 30 min at room temperature, followed by 1N HCl was used for the colorimetric detection of titres endpoint. Optical densities at a wavelength of 450 nm (OD_{450nm}) were read using a microplate spectrophotometer (BioTek; Winooski, VT, USA). Standard curves were generated using pools of HIV-1 positive individuals diluted over the range 1:5000–1:5120000 (for p24-specific total IgG detection) or 1:25–1:25 600 (for p24-specific IgG3 detection). All samples were tested in duplicate and the intra-assay variability was below 20% (a variation of up to 25% at the lower limit of quantification (bottom) is recommended as default acceptance criteria for accuracy and inter-batch precision [20]. Four-parameter variable slope non-linear regression was calculated using log-transformed optical density values on Prism 6 for Mac OS X. The curves were run on each plate and used to determine the reproducibility of the assay.

For the IgG3 assay, the precision of the assay was also evaluated by running in each plate assay, three points of the standard curve in duplicate each: the OD_{450nm} at the lower limit of detection (LLD), the approximate midrange of the calibration curve and the upper limit of the curve (ULC). To be accepted as the lowest and the highest values of OD_{450nm} on the standard curve, LLD and ULC need to meet the following criteria: (1) the OD_{450nm} is at least 5 times higher than the OD_{450nm} of the background (HIV-1 negative sera); (2) the response should be reproducible with a mean coefficient of variation (CV) of $< 10\%$. The CV is defined as the ratio of the standard deviation to the mean $[(s.d./mean) \times 100]$. The ULC is selected as the highest plateau of the calibration curve and is evaluated using the same protocol as for LLD. The analysis of intra- and inter-assay precision was carried out 6 times on separate occasions, with the acceptance criteria being a mean CV of $< 0\%$.

Results

Immunoreactivity analysis of the recombinant p24 protein by singleplex LMA assay

The strategy for developing a recent infection assay for HIV-1 that allows the quantification of total IgG and IgG3 antibody titres directed against HIV-1 p24 protein is based on the design of an antigen capable of mimicking the structural content and reactivity of the native protein. The DNA sequence coding for the HIV-1 p24 protein was optimised for enhanced bacterial expression, produced as a N-terminally 6x His-tagged protein and purified under non-denaturing conditions by affinity chromatography.

SDS-PAGE and western-blot analysis showed that the recombinant p24 protein is purified as a single protein of 24 kDa and no aggregates are present in the purified sample (Fig. S2). The experimental Far-UV circular dichroism (CD) spectrum for the p24 protein was acquired from 200 to 260 nm and the stability of the p24 protein was confirmed by monitoring the heat-induced unfolding curve (Fig. S3).

The immunoreactivity and specificity of the recombinant p24 protein were firstly assessed against 45 HIV-1 positive and 317 HIV-1 negative human serum samples (the latter including negative samples from healthy individuals ($n = 82$), positive samples for Hepatitis B ($n = 79$), Hepatitis C ($n = 40$), Chagas Disease ($n = 38$), HTLV-1/2 ($n = 48$) and Syphilis ($n = 30$)) in a singleplex LMA (Fig. 1a). The results revealed that the p24 protein is strongly reactive against total IgG antibodies present in the HIV-1 positive samples group and exhibited a false-positive rate of 4.02%, hence assuring that cross-reactivity was not significant. These data were also used to determine the diagnostic performance of the p24 protein through receiver operating characteristics (ROC) analysis (Fig. 1b). The AUC for the p24 protein was 0.9984, while sensibility and specificity values corresponded to 100% and 95.98%, respectively.

Development of an ELISA assay to detect and titrate HIV-1 p24-specific total IgG antibodies in patient serum samples

Aiming to establish a tool to allow simple measurements of incidence estimation and samples cohort classification, IgG titre profiles were monitored against p24 protein as recent infection markers for HIV-1 infection. A developed ELISA was used to detect and quantify p24-specific total IgG antibodies in clinical serum samples. The reliability and reproducibility of this assay were evaluated by using a pool of 30 HIV-1 positive serum samples from the MACS (Fig. S4a). Using a panel of 48 serum samples from chronically infected HIV-1 participants and 48 serum

samples from healthy participants (enrolled in the MACS), our p24-based total IgG detection assay allowed a clear distinction between the two sample sets (Fig. 2a). The developed assay was subsequently tested against 22 paired serum samples from 11 HIV-1 participants enrolled in the Consortium for the Evaluation and Performance of HIV Incidence Assays (CEPHIA). The paired samples comprised the last negative sample before seroconversion and a positive sample collected at least 6 months (and up to 12 months) after seroconversion. Our results revealed that the developed assay detected anti-p24 IgG specifically after seroconversion, revealing titres ranging from 774 to 5626 in samples from long-standing HIV-1 infection (Fig. 2b).

Kinetics of p24-specific total IgG and IgG3 antibodies in sequential samples and assessment of recent infection test parameters

To determine the kinetics of p24-specific IgG3 antibodies appearance early after seroconversion, a second assay to detect and titrate HIV-1 p24-specific IgG3 antibodies in human serum samples was developed (calibration curves assays are shown in the Fig. S4b). The response curves for p24-specific total IgG and IgG3 antibodies in prospective sequential sera samples from 30 seroconverted participants enrolled in the CEPHIA cohort were compared (Fig. 3). Serum samples were collected over the period of 1–24 months post-seroconversion and distributed in five groups (1–4, 5–8, 9–12, 13–18 and 19–24 months). According to our results, the total IgG antibody titres are overall sustained at high levels throughout the observed time (ranging from 314 up to 889 000) (Fig. 3a). Compiled kinetics data from all 30 seroconverters show that the majority of individuals presents an increase in p24-specific total IgG levels over time; however, there is considerable heterogeneity among individual responses, including some that ramp up slowly or even decline slightly over time (Fig. 3b). These data indicate that p24-specific total

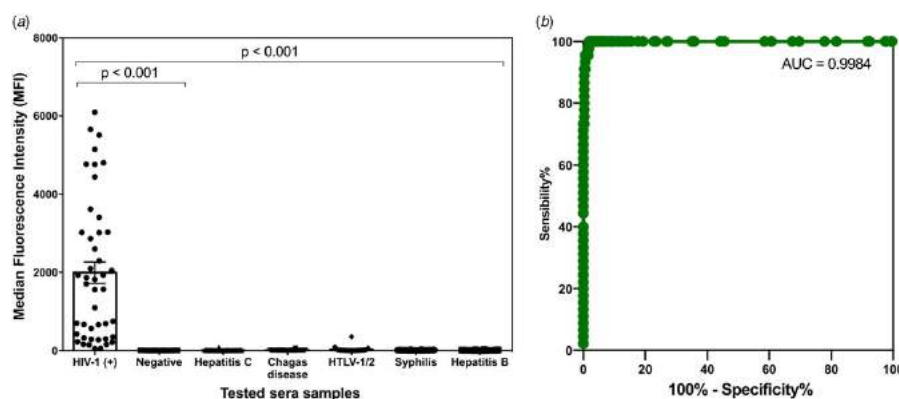


Fig. 1. Immunoreactivity and diagnostic performance assessment of the designed HIV-1 p24 protein. (a) – Median fluorescence intensity (MFI) for the p24 protein against HIV-1 positive and negative (control) human serum samples and against Hepatitis B and C, Chagas Disease, Syphilis and HTLV-1/2 positive serum samples. The results are presented as the median values over a 100 events for individual samples (dots) along with the corresponding standard deviation ($P < 0.001$). Mean values and correspondent standard deviations are shown. The P -values of two-tailed t -test and one-way ANOVA are indicated above each analysis. (b) – ROC curve analysis of the HIV-1 p24 protein. The paired results for sensitivity and specificity were plotted as points in a ROC space and the trade-off between these measures for different discrimination cut-offs are graphically represented.

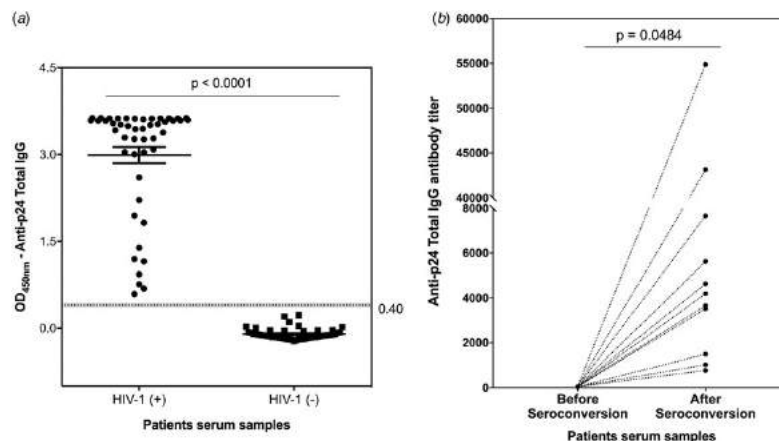


Fig. 2. Detection and titration of p24-specific total IgG antibodies in clinical samples by ELISA assay. (a) – Scatter plot representation of p24-specific IgG antibodies detection in serum samples from 48 healthy participants and 48 HIV-1 chronically infected individuals. The results are shown as mean values over all measurements and corresponding standard error of the mean (S.E.M.). Dots represent individual measurements. A dashed line indicates the cut-off value and the *P*-value of two-tailed *t*-test is indicated above. (b) – Titration of p24-specific total IgG antibodies in paired samples from 11 HIV-1 positive individuals comprising the last negative sample and first positive sample within 6–12 months after seroconversion. Each sample was tested in duplicate and the mean values are shown. Dotted lines connect paired samples and the *P*-value of two-tailed *t* test is indicated above the graph.

IgG response does not show correlation with progression throughout the early stages after seroconversion, hence not allowing inference on time after patient seroconversion. In contrast, the IgG3 titre profile against p24 exhibited a transient reactivity peak observed within the first 4 months post-seroconversion, with titres ranging from 83 to 2527, which then declined (Fig. 3c). This finding is being reported for the first time and indicates that the p24-specific IgG3 response may be used to differentiate recent from long-standing HIV-1 infections. The selected cut-off value (titre: 190) is displayed in the graph as a dotted black line and it was chosen based on the bottom end of the 95% confidence

interval for the reactivity of the first group measurement (recent samples, 1–4 months after seroconversion). Therefore, recent infection is defined as any value above the defined cut-off value.

Assuming uniformly distributed seroconversion events over the time intervals [0, 730 days], the MDRI was estimated from the proportion of all seroconverters testing recent according to the defined titre cut-off. Figure 4a shows the development of p24-specific IgG3 antibody response for each subject as a function of days after-seroconversion. According to the results, the value for MDRI on the developed assay is 144 days (95% CI 132–156). In addition, among the 119 specimens from

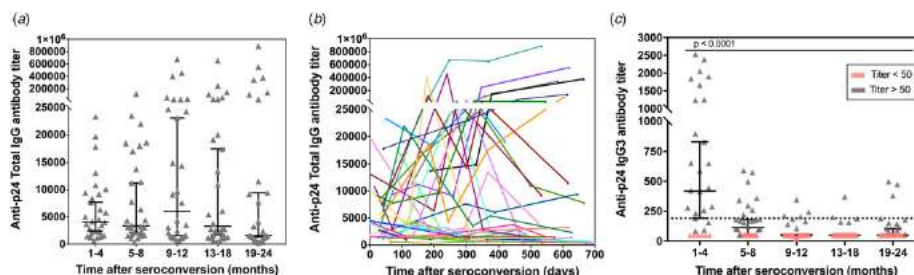


Fig. 3. p24-specific total IgG and IgG3 antibody titres as measured by our ELISA assay in 150 sequential serum samples from 30 HIV-1 positive and recently seroconverted individuals. The total IgG (a) and IgG3 (c) response towards p24 was determined by ELISA in 150 serum samples from 30 HIV-1 positive individuals who recently seroconverted. The samples were divided into five groups based on the time post-seroconversion and over a total period of 24 months. The antibody titres are plotted vs. the intervals after seroconversion in months. Individual measurements are shown (grey symbols) along with the median (black lines) and corresponding 95% confidence interval. The black dotted line in (c) represents the cut-off value (titre: 190) selected based on the final end of the 95% confidence interval of the mean of first time-point. (b) – Titration of p24-specific total IgG antibodies in relation to days since seroconversion. Each line represents a single individual with sequential specimens collected over time.

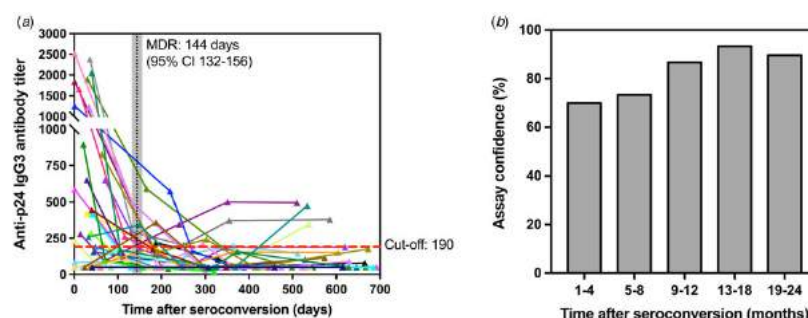


Fig. 4. Kinetics of p24-specific IgG3 antibodies as measured by our ELISA assay in 150 sequential serum samples from 30 HIV-1 positive and recently seroconverted individuals in relation to time since seroconversion (days). (a) – Each line represents a single individual with sequential specimens collected over time post-seroconversion (in days). Cut-off (titre = 190) is shown as a red dashed line, whereas the mean duration of recent infection (144 days) and corresponding confidence interval (95% CI 132–156) are shown as a vertical black line and shaded grey area, respectively. (b) – Confidence levels of the developed assay are shown as the frequency of individuals correctly classified as recent (1–4 months) and non-recent (5–24 months) according to the assay cut-off described in (a).

HIV-1-infected individuals that were known to have seroconverted for >4 months, only 17 specimens were classified as recent on the assay, resulting in a false-recent rate (FRR) of 14.2% for the developed test. The empirical frequency of individuals correctly classified as recent (in the 1–4 months group) and non-recent (in the 5–8, 9–12, 13–18 and 19–24 months groups) is also shown in Figure 4b as a confidence measure of the developed recent infection assay.

Discussion

Intensive efforts by HIV-control programs from many countries have made a significant impact on reducing transmission of HIV-1 infection [6]. This achievement is largely due to the development of laboratory-based incidence assays. Although some HIV-1 incidence assays are commercially available, the establishment of an easy-to-perform assay that is capable of accurately discriminate early from long-term HIV-1 infections in resource-constrained settings is still critically needed for both public health improvement and epidemiological surveillance. In the present work, a designed recombinant p24 protein was applied as an antigen to two different assay formats to detect p24-specific IgG responses: singleplex LMA and ELISA. We have also developed a simple and reproducible ELISA assay to assess the kinetics of p24-specific immunoglobulin 3 (IgG3) responses over the early stages after seroconversion with reduced inter-run variations. The data has shown that anti-p24 IgG3 antibody titres exhibit a unique early and transient peak after seroconversion. This distinct behaviour enabled the use of elevated anti-p24 IgG3 titres as an early infection marker capable of differentiating recent from long-standing HIV-1 infections. Therefore, the IgG3 capture ELISA assay was also used to establish the optimal cut-off threshold value for recent infection analyses.

In summary, the folding and conformational stability of a designed p24 recombinant protein was confirmed by CD spectroscopy (Fig. S1). The recombinant protein was used to develop a singleplex LMA to detect specific total IgG antibodies in serum samples from 45 HIV-1 chronically infected participants as a preliminary immunoreactivity evaluation method. The results showed that our p24-based antigen exhibits the high power of

discrimination between positive and negative sample sets and exhibited 100% of sensibility and 98.95% of specificity (Fig. 1). In agreement with the LMA data, the developed p24-based ELISA assay allowed the detection of p24-specific total IgG antibodies in all serum samples from 48 HIV-1 chronically infected participants from the MACS cohort (Fig. 2a). The antibody titration assay, designed in the same ELISA format, showed the presence of p24-specific total IgG titres up to 55 000 in samples from 11 seroconverters (seroconversion time >6 months) (Fig. 2b).

In addition to p24-specific total IgG antibodies, the IgG subclasses responses triggered by this protein can mark the progression throughout the early infection phase, facilitating the identification of recent infections and stratification of participants in clinical cohorts. During the HIV-1 infection, the antibody response to Gag appears at an average time of 18 days after seroconversion, being primarily directed against p24 and p17 proteins and leading to the appearance of IgG antibodies of the IgG3 subclass, followed by IgG1, IgG2 and IgG4 [21–23]. Previous studies suggest that IgG subclass responses to HIV-1 Gag proteins may vary according to infection time and progression status [24, 25]. IgG3 is one of the second most predominant subclasses in the antibody response against HIV [21, 22, 25] and, although anti-p24 IgG3 is known to appear within the first 6 months after seroconversion, the time point when the response reaches the maximum and then declines lacks to be defined. The development of a new recent infection assay described here was geared to address this limitation. Due to their early appearance in the context of the HIV-1 infection, IgG3 antibody titres directed against p24 were evaluated as an early infection marker in this study. Using our developed ELISA assay to titrate p24-specific total IgG antibodies as a platform, a modified version of this assay was set up to titrate p24-specific IgG3 antibodies in a cohort of sequential serum samples from untreated HIV-1 positive individuals collected over the period of 24 months after seroconversion (Figs 3 and 4). Monitoring of the calibration curves over 20 plates demonstrated that our assay is consistent and shows inter- and intra-assay variability <20% (Fig. S2), which meet the intra- and inter-assay precision determined by the FDA guidance acceptance criteria [26].

Our findings reveal that, while the p24-specific total IgG antibodies titres are maintained markedly high in untreated HIV-1

individuals, IgG3 antibody titres are found at high levels only during a limited early period after seroconversion (Figs 3 and 4), regardless of the presence of high plasma viral loads (above 10 000 copies/ml of blood) (data not shown for conciseness). After a single transient peak of IgG3 titres during the first 5 months after seroconversion, the response declines, probably reflecting the antibody deficient recycling, which leads to a shorter half-life [12, 27] and a decrease in the IgG3 antibody levels. Although out of the scope of this paper, it is worth mentioning that IgG3 responses are predominantly linked to T-helper 1 (TH1) responses [28, 29] and the presence of IgG3 in the first 4–5 months after seroconversion may also be a marker of well-preserved Th1 response in the early stages of infection. It is also important mentioning that, although individuals with long-standing infections under ARR present a reestablishment of normal T-cell counts, those individuals did not present measurable levels of anti-p24 IgG3 antibodies in our assays (data not shown for conciseness). We have also demonstrated that titration of p24-specific IgG3 antibodies has a MDRI of 144 days (95% CI 132–156), which is between 4 and 5 months and sufficiently long to detect new infections for good incidence estimates (Fig. 4a). Our results are comparable with the results from Duong and colleagues [6], who have reported a Lag-Avidity EIA with a MDRI of 141 days (95% CI 119–160) with a low CV. On the other hand, among 119 individuals tested in this study with long-standing HIV-1 infections (seroconversion time >144 days), the misclassification rate (false-recent rate – FRR) was 14.2%, which is considerably higher than the LAg-Avidity EIA developed by Dung and colleagues (0.2%) and the BED assay (2.9%) [7]. However, it is important to note that precision of seroconversion time and frequency of sample collection may affect FRR values. Therefore, testing of more seroconverters and additional longitudinal specimens representing different populations will be required to further refine our findings as part of future studies by our group.

In conclusion, in the current study, we have demonstrated the potential of a recombinant p24 protein-based ELISA assay for fast and reproducible titration of p24-specific IgG3 antibodies in human sera to accurately differentiate recent from long-standing infections. Our application in a longitudinal specimen set suggests that a decrease in IgG3 titre against p24 reflects entering in the non-recent phase of HIV-1 infection. This assay may be useful for better allocation of scarce resources as well as for research and public health applications, specifically for research cohort stratification purposes and similar studies (such as the design and efficacy evaluation of therapeutic and prophylactic vaccines).

Supplementary material. The supplementary material for this article can be found at <https://doi.org/10.1017/S0950268818001218>

Acknowledgements. The authors would like to thank the Multicenter AIDS Cohort Study team for its valuable cooperation with this project. Dr Seth Horn, Department of Chemistry, University of Pittsburgh, is acknowledged for providing access to the circular dichroism facility. This work was supported by the Coordination for the Improvement of Higher Education Personnel (I.E.T.V., Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES, grant number 9848-12-6), Brazil; by the National Institutes of Health – NIH in a partnership with the Brazilian National Council for Scientific and Technological Development (I.E.T.V., E.T.A.M., Conselho Nacional de Desenvolvimento Científico e Tecnológico, grant number PA-14-328); by the NIH (C.R.R., R.B.M., grant number U01 AI035041, University of Pittsburgh Multicenter AIDS Cohort Study (MACS)); by the Molecular Biology Institute of Paraná – IBMP, Brazil (L.F.O.L., L.F., M.A.K., no specific grant); and by the Center for Vaccine Research at the University of Pittsburgh, USA (M.L.P., D.F.C., E.J.M.N., G.G., no specific grant). The other authors received no specific

grant from any funding agency, commercial or not-for-profit sectors to perform this research.

Authors contribution. I.E.T.V. designed and produced the p24 protein, acquired the CD spectra, designed and performed Luminex and ELISA assays against p24-specific total IgG and IgG3 antibodies, performed statistical analysis and wrote the paper; D.F.C. produced the p24 protein, performed ELISA against p24-specific total IgG and IgG3 antibodies and wrote the paper; M.L.P. performed ELISA against p24-specific total IgG and IgG3 antibodies and wrote the paper; E.J.M.N. designed and performed ELISA against p24-specific total IgG and IgG3 antibodies; G.G. performed SDS-PAGE and western-blot analysis; L.F.O.L. and L.F. designed and performed LMA against p24-specific total IgG antibodies; M.A.K. selected serum samples for LMA analysis; C.P. selected and provided the CEPHIA panel samples; C.E.C.S. designed the DNA sequence coding for the p24 protein; R.M. and C.R. selected the MACS samples for analysis; R.D. optimised the DNA sequence coding for the p24 protein and designed LMA experiments; E.T.A.M. designed experiments and coordinated study.

Compliance with ethical standards

Conflicts of interest. The authors declare that they have no conflict of interest.

Ethical Approval. The human serum samples used in the LMA and ELISA experiments were commercially acquired or provided by the CEPHIA consortium. All the human serum samples from the MACS cohort reported by the authors in this article were provided by the MACS team after approval by the Ethical Committee of the University of Pittsburgh.

References

1. Curtis KA *et al.* (2013) Evaluation of a multiplex assay for estimation of HIV-1 incidence. *PLoS ONE* 8(5), e64201.
2. Park SY *et al.* (2014) Developing high-throughput HIV incidence assay with pyrosequencing platform. *Journal of Virology* 88(5), 2977–2990.
3. WHO U. (2011) When and how to use assays for recent infection to estimate HIV incidence at a population level. In Press W (ed.), *WHO Library Cataloguing-in-Publication Data*. Switzerland, pp. 1–49.
4. Busch MP *et al.* (2010) Beyond detuning: 10 years of progress and new challenges in the development and application of assays for HIV incidence estimation. *Aids (London, England)* 24(18), 2763–2771.
5. Moyo S *et al.* (2015) Identifying recent HIV infections: from serological assays to genomics. *Viruses* 7(10), 5508–5524.
6. Duong YT *et al.* (2012) Detection of recent HIV-1 infection using a new limiting-antigen avidity assay: potential for HIV-1 incidence estimates and avidity maturation studies. *PLoS ONE* 7(3), e33328.
7. Parekh BS *et al.* (2002) Quantitative detection of increasing HIV type 1 antibodies after seroconversion: a simple assay for detecting recent HIV infection and estimating incidence. *AIDS Research and Human Retroviruses* 18(4), 295–307.
8. Keating SM *et al.* (2012) Lower-sensitivity and avidity modifications of the vitros anti-HIV 1 + 2 assay for detection of recent HIV infections and incidence estimation. *Journal of Clinical Microbiology* 50(12), 3968–3976.
9. Kassanjee R *et al.* (2014) Independent assessment of candidate HIV incidence assays on specimens in the CEPHIA repository. *Aids (London, England)* 28(16), 2439–2449.
10. Incidence Assay Critical Path Working G (2011) More and better information to tackle HIV epidemics: towards improved HIV incidence assays. *PLoS Medicine* 8(6), e1001045.
11. Yates NL *et al.* (2011) Multiple HIV-1-specific IgG3 responses decline during acute HIV-1: implications for detection of incident HIV infection. *Aids (London, England)* 25(17), 2089–2097.
12. Stapleton NM *et al.* (2011) Competition for FcRn-mediated transport gives rise to short half-life of human IgG3 and offers therapeutic potential. *Nature Communications* 2, 599.
13. Khalife J *et al.* (1988) Isotypic restriction of the antibody response to human immunodeficiency virus. *AIDS Research and Human Retroviruses* 4(1), 3–9.

14. **Binley JM *et al.*** (1997) Differential regulation of the antibody responses to Gag and Env proteins of human immunodeficiency virus type 1. *Journal of Virology* **71**(4), 2799–2809.
15. **Altschul SF *et al.*** (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**(17), 3389–3402.
16. **Altschul SF *et al.*** (2005) Protein database searches using compositionally adjusted substitution matrices. *FEBS Journal* **272**(20), 5101–5109.
17. **Crestani S *et al.*** (2016) Enhanced target-specific signal detection using an *Escherichia coli* lysate in multiplex microbead immunoassays with *E. coli*-derived recombinant antigens. *Journal of Immunological Methods* **435**, 17–26.
18. **Hanley JA and McNeil BJ** (1982) The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* **143**(1), 29–36.
19. **Zweig MH and Campbell G** (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry* **39**(4), 561–577.
20. **DeSilva B *et al.*** (2003) Recommendations for the bioanalytical method validation of ligand-binding assays to support pharmacokinetic assessments of macromolecules. *Pharmaceutical Research* **20**(11), 1885–1900.
21. **Tomaras GD and Haynes BF** (2009) HIV-1-specific antibody responses during acute and chronic HIV-1 infection. *Current Opinion in HIV and AIDS* **4**(5), 373–379.
22. **Wilson KM *et al.*** (2004) Incidence immunoassay for distinguishing recent from established HIV-1 infection in therapy-naïve populations. *Aids (London, England)* **18**(17), 2253–2259.
23. **Ferrante A, Beard LJ and Feldman RG** (1990) IgG subclass distribution of antibodies to bacterial and viral antigens. *Pediatric Infectious Disease Journal* **9**(Suppl. 8), S16–S24.
24. **Banerjee K *et al.*** (2010) IgG subclass profiles in infected HIV type 1 controllers and chronic progressors and in uninfected recipients of Env vaccines. *AIDS Research and Human Retroviruses* **26**(4), 445–458.
25. **McDougal JS *et al.*** (1987) Antibody response to human immunodeficiency virus in homosexual men. Relation of antibody specificity, titer, and isotype to clinical status, severity of immunodeficiency, and disease progression. *Journal of Clinical Investigation* **80**(2), 316–324.
26. **Administration UFaD** (2001) Guidance for industry bioanalytical method validation. Available at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>.
27. **Vidarsson G, Dekkers G and Rispens T** (2014) IgG subclasses and allotypes: from structure to effector functions. *Frontiers in Immunology* **5**, 520.
28. **Snapper CM and Mond JJ** (1993) Towards a comprehensive view of immunoglobulin class switching. *Immunology Today* **14**(1), 15–17.
29. **Germann T *et al.*** (1995) Interleukin-12 profoundly up-regulates the synthesis of antigen-specific complement-fixing IgG2a, IgG2b and IgG3 antibody subclasses in vivo. *European Journal of Immunology* **25**(3), 823–829.