

Introduction to Bioinformatics

Ernane Rosa Martins
(Organizador)

 **Atena**
Editora
2019



Ernane Rosa Martins
(Organizador)

Introduction to Bioinformatics

Atena Editora
2019

2019 by Atena Editora

Copyright © da Atena Editora

Editora Chefe: Profª Drª Antonella Carvalho de Oliveira

Diagramação e Edição de Arte: Geraldo Alves e Natália Sandrini

Revisão: Os autores

Conselho Editorial

- Prof. Dr. Alan Mario Zuffo – Universidade Federal de Mato Grosso do Sul
Prof. Dr. Álvaro Augusto de Borba Barreto – Universidade Federal de Pelotas
Prof. Dr. Antonio Carlos Frasson – Universidade Tecnológica Federal do Paraná
Prof. Dr. Antonio Isidro-Filho – Universidade de Brasília
Profª Drª Cristina Gaio – Universidade de Lisboa
Prof. Dr. Constantino Ribeiro de Oliveira Junior – Universidade Estadual de Ponta Grossa
Profª Drª Daiane Garabeli Trojan – Universidade Norte do Paraná
Prof. Dr. Darllan Collins da Cunha e Silva – Universidade Estadual Paulista
Profª Drª Deusilene Souza Vieira Dall’Acqua – Universidade Federal de Rondônia
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Prof. Dr. Fábio Steiner – Universidade Estadual de Mato Grosso do Sul
Prof. Dr. Gianfábio Pimentel Franco – Universidade Federal de Santa Maria
Prof. Dr. Gilmei Fleck – Universidade Estadual do Oeste do Paraná
Profª Drª Girlene Santos de Souza – Universidade Federal do Recôncavo da Bahia
Profª Drª Ivone Goulart Lopes – Istituto Internazionele delle Figlie de Maria Ausiliatrice
Profª Drª Juliane Sant’Ana Bento – Universidade Federal do Rio Grande do Sul
Prof. Dr. Julio Candido de Meirelles Junior – Universidade Federal Fluminense
Prof. Dr. Jorge González Aguilera – Universidade Federal de Mato Grosso do Sul
Profª Drª Lina Maria Gonçalves – Universidade Federal do Tocantins
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Profª Drª Paola Andressa Scortegagna – Universidade Estadual de Ponta Grossa
Profª Drª Raissa Rachel Salustriano da Silva Matos – Universidade Federal do Maranhão
Prof. Dr. Ronilson Freitas de Souza – Universidade do Estado do Pará
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista
Prof. Dr. Urandi João Rodrigues Junior – Universidade Federal do Oeste do Pará
Prof. Dr. Valdemar Antonio Paffaro Junior – Universidade Federal de Alfenas
Profª Drª Vanessa Bordin Viera – Universidade Federal de Campina Grande
Profª Drª Vanessa Lima Gonçalves – Universidade Estadual de Ponta Grossa
Prof. Dr. Willian Douglas Guilherme – Universidade Federal do Tocantins

Dados Internacionais de Catalogação na Publicação (CIP) (eDOC BRASIL, Belo Horizonte/MG)

I61	Introduction to bioinformatics [recurso eletrônico] / Organizador Ernane Rosa Martins. – Ponta Grossa (PR): Atena Editora, 2019. Formato: PDF Requisitos de sistema: Adobe Acrobat Reader Modo de acesso: World Wide Web Inclui bibliografia. ISBN 978-85-7247-113-8 DOI 10.22533/at.ed.138191202 1. Bioinformática. 2. Inteligência artificial. I. Martins, Ernane Rosa. CDD 570.285
-----	---

Elaborado por Maurício Amormino Júnior – CRB6/2422

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores.

2019

Permitido o download da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

www.atenaeditora.com.br

APRESENTAÇÃO

A bioinformática é um campo interdisciplinar, que busca analisar, interpretar e processar dados biológicos, com foco na aplicação de técnicas computacionais intensivas, tais como: métodos computacionais, teoria de grafos, inteligência artificial, algoritmos matemáticos, reconhecimento de padrões, mineração de dados, algoritmos de aprendizado de máquina, processamento de imagens e simulação computacional. Como um campo interdisciplinar, a bioinformática combina diversas áreas do conhecimento, como: engenharia, matemática, física, química, estatística, ciência da computação e biologia, entre outras.

A coletânea “*Introduction to bioinformatics*” é um livro composto por 6 capítulos que abordam assuntos atuais, tais como: o adenocarcinoma gástrico que é uma malignidade com elevada incidência e mortalidade no mundo; o vírus zika (VZIK) que é um Arbovirus que pertence à família Flaviviridae; As H^+ -ATPases que são proteínas integrais da membrana plasmática que têm a capacidade de utilizar a energia química da hidrólise de ATP para expulsar os prótons para o ambiente extracelular, atuando na manutenção da homeostase iônica e transporte de solutos; o vírus da família Geminiviridae que tem sido intensamente estudado devido à gravidade das doenças causadas em várias culturas importantes como: feijão, algodão, milho, tomate e mandioca.

Espero que os capítulos deste livro possam contribuir efetivamente na disseminação dos conhecimentos relevantes da bioinformática, proporcionando uma visão ampla sobre este campo de conhecimento.

Assim, desejo a todos uma excelente leitura.

Ernane Rosa Martins

SUMÁRIO

CAPÍTULO 1 1

ANÁLISE DE lncRNAs EM NPCs DE HAMSTER GOLDEN SÍRIO (*Mesocricetus auratus*) RECÉM-NASCIDOS INFECTADOS PELO VÍRUS ZIKA

Jardel Fabio Lopes Ferreira
Samir Mansour Moraes Casseb
Karla Fabiane Lopes de Melo
Carlos Alberto Marques de Carvalho
Gustavo Moraes Holanda
Paloma Daguer Ewerton dos Santos
Suellen de Almeida Machado
Francisco Canindé Ferreira de Luna
Walter Felix Franco Neto
Livia Carício Martins
Ana Cecília Ribeiro Cruz
Pedro Fernando da Costa Vasconcelos

DOI 10.22533/at.ed.1381912021

CAPÍTULO 2 11

ANÁLISE *IN SILICO* DA FARMACOCINÉTICA E FARMACODINÂMICA DO COMPOSTO BENZOTIAZÓLICO COM POTENCIAL ANTITUMORAL CONTRA LINHAGEM DE ADENOCARCINOMA GÁSTRICO

Felipe Pantoja Mesquita
Luina Benevides Lima
Julio Paulino Daniel
Adrhyan Jullyanne de Sousa Portilho
Lais Lacerda Brasil de Oliveira
Emerson Lucena da Silva
Eliza de Lucas Chazin
Thatyana Rocha Alves Vasconcelos
Maria Elisabete Amaral de Moraes
Raquel Carvalho Montenegro

DOI 10.22533/at.ed.1381912022

CAPÍTULO 3 23

ANÁLISE PRIMÁRIA DE TRANSCRIPTOMA DE TECIDO TESTICULAR DE HAMSTERS (*MESOCRICETUS AURATUS*) INFECTADOS COM VÍRUS ZIKA

Walter Felix Franco Neto
Samir Mansour Moraes Casseb
Karla Fabiane Lopes de Melo
Wallax Augusto Silva Ferreira
Ana Paula Sousa Araujo
Jardel Fabio Lopes Ferreira
Taiana Andrade Freitas
Milene Ferreira Silveira
Livia Carício Martins
Pedro Fernando da Costa Vasconcelos

DOI 10.22533/at.ed.1381912023

CAPÍTULO 4	32
CARACTERIZAÇÃO FILOGENÉTICA DA FAMÍLIA MULTIGÊNICA DA H ⁺ -ATPASE DE MEMBRANA PLASMÁTICA EM MONOCOTILEDÔNEAS DA ORDEM POALES	
Lyndefânia Melo de Sousa Clesivan Pereira dos Santos Thais Andrade Germano Moacíria de Souza Lemos Stelamaris de Oliveira Paula Rafael de Souza Miranda José Helio Costa	
DOI 10.22533/at.ed.1381912024	
CAPÍTULO 5	40
CLADISTIC ANALISYS IN GEMINIVIRIDAE: AN EVIDENCE OF MULTISPECIFICITY FOR CULTIVARS HOSTS	
Rafael Trindade Maia Aparecida Yasmim Silva de Azevedo Maria Bartira Chaves de Souza Silva Ana Verônica Silva do Nascimento	
DOI 10.22533/at.ed.1381912025	
CAPÍTULO 6	50
DESENVOLVIMENTO DE FRAMEWORK PARA CRIAÇÃO DE MODELOS COMPUTACIONAIS DE CÉLULA COMPLETA	
Frederico Chaves Carvalho Paulo Eduardo Ambrósio	
DOI 10.22533/at.ed.1381912026	
CAPÍTULO 7	63
IN-SILICO DETOXIFICATION EVIDENCE OF THE HERBICIDE BISPYRIBAC SODIUM BY A TEORETHICAL MODEL OF GLUTATHIONE S-TRANSFERASE TAU 5 FROM <i>Oryza sativa</i> L.	
Vinícius Costa Amador Ravenna Lins Rodrigues Felipe de Oliveira França Rafael Trindade Maia	
DOI 10.22533/at.ed.1381912027	
CAPÍTULO 8	73
INVESTIGAÇÃO IN SILICO DA EFICÁCIA DE FÁRMACOS ANTIVIRAIS NA INIBIÇÃO DA NS5 DO VÍRUS DA ZIKA	
Henriqueta Monalisa Farias Rafael de Lima Medeiros Franklin de Ferreira Farias Nóbrega Rafael Trindade Maia	
DOI 10.22533/at.ed.1381912028	
SOBRE O ORGANIZADOR	85

ANÁLISE DE LNCRNAS EM NPCS DE HAMSTER GOLDEN SÍRIO (*Mesocricetus auratus*) RECÉM-NASCIDOS INFECTADOS PELO VÍRUS ZIKA

Jardel Fabio Lopes Ferreira

Instituto Evandro Chagas
Ananindeua – PA

Samir Mansour Moraes Casseb

Instituto Evandro Chagas
Ananindeua – PA

Karla Fabiane Lopes de Melo

Escola Superior da Amazônia
Belém – PA

Carlos Alberto Marques de Carvalho

Centro Universitário Metropolitano da Amazônia –
UNIFAMAZ – PA

Gustavo Moraes Holanda

Universidade do Estado do Pará – UEPA – PA

Paloma Daguer Ewerton dos Santos

Instituto Evandro Chagas – IEC – PA

Suellen de Almeida Machado

Instituto Evandro Chagas
Ananindeua – PA

Francisco Canindé Ferreira de Luna

Universidade Federal do Pará
Belém – PA

Walter Felix Franco Neto

Instituto Evandro Chagas
Ananindeua – PA

Lívia Carício Martins

Instituto Evandro Chagas
Ananindeua – PA

Ana Cecília Ribeiro Cruz

Instituto Evandro Chagas
Ananindeua – PA

Pedro Fernando da Costa Vasconcelos

Instituto Evandro Chagas
Ananindeua – PA

RESUMO: O Vírus Zika (VZIK) é um arbovírus pertencente à família *Flaviviridae*, gênero *flavivirus* e necessita de vetores artrópodes do gênero *Aedes* (*Ae.aegypti* e *Ae.albopictus*) para manter seu ciclo de replicação viral. Possui RNA de fita simples com polaridade positiva com informações para codificação de proteínas estruturais e não estruturais. O VZIK vem se destacando a partir de epidemias nas Américas e, principalmente, pela capacidade em causar agravos neurológicos severos como a microcefalia em crianças recém-nascidas e síndrome de Guillian-Barré em adultos. Além disso, foi relatada a infertilidade em homens infectados com o vírus, o que evidencia o tropismo por diferentes órgãos. Portanto, tentar elucidar a relação vírus-hospedeiro através de RNAs longos não-codificantes de proteínas (lncRNAs) a partir de transcriptoma dos animais infectados utilizando RNA-seq, é de suma importância para o meio acadêmico científico, uma vez que esses RNAs são reguladores da expressão gênica e estão intimamente relacionados a processos fisiológicos celulares importantes, como apoptose e proliferação celular.

1 | INTRODUÇÃO

O Vírus Zika (VZIK) teve seu isolamento na África 1947, em Uganda na floresta de Zika, originando o nome dado a esse arbovírus. O isolamento do VZIK foi a partir de amostras coletadas de macaco do gênero *Rhesus* que servia de estudo para o Vírus da Febre Amarela (VFA) na região.

O VZIK possui genoma completo de aproximadamente 10.5 a 11kb de tamanho, formado por RNA de fita simples com polaridade positiva, responsável por conter informação para codificação de proteínas estruturais e não-estruturais sendo essas últimas responsáveis principalmente pela evasão contra a resposta imune do hospedeiro.

Pertencente à família *Flaviviridae* e gênero *flavivirus*, assim como outros arbovírus importantes para a saúde pública mundial como o Vírus Dengue (VDEN) e VFA, o VZIK necessita de vetores artrópodes do gênero *Aedes* para manter seu ciclo de replicação, sendo *Ae.aegypti* e *Ae.albopictus* seus principais vetores, podendo infectar células animais hospedeiras e ocasionar uma gama de doenças em humanos.

O termo arbovírus provém do inglês *arthropod-borne* que foi designado em 1962, para descrever os arbovírus que infectam animais vertebrados a partir de vetores artrópodes hematófagos. As principais regiões que concentram uma grande variedade de vetores artrópodes são as de climas temperadas e tropicais, possibilitando um ambiente propício para reprodução desses vetores e causando sua disseminação no mundo.

2 | FAMÍLIA FLAVIVIRIDAE E GÊNERO FLAVIVIRUS

A família *Flaviviridae* e gênero *flavivirus* contém a maioria dos arbovírus que causam problemas de saúde pública e social em todo o mundo, chamando atenção da Organização da Mundial da Saúde (OMS) para tomar medidas preventivas contra epidemias. Essa classe de família e gênero é formada por vírus que possuem RNA de fita simples com polaridade positiva, chamando atenção para outros arbovírus que além do VZIK, são capazes de ocasionar agravos neurológicos como o Vírus West Nile (VWN) e Vírus da Encefalite Japonesa (VEJ).

O gênero *Flavivirus* possui dois agrupamentos importantes, sendo os vírus transmitidos por carrapatos e os vírus transmitidos por mosquitos. A diversidade de vetores artrópodes de acordo com as características das doenças por eles provocadas e ampla distribuição geográfica, são fatores importantes para considerar que há necessidade de medidas preventivas uma vez que inúmeras populações mundiais residem em áreas endêmicas.

O material genético é lido no sentido 5' e 3', possuindo regiões que não são traduzidas (RNT) de aproximadamente 100 a 500 nucleotídeos. Seu genoma pode variar de 10.5kb a 11kb formando proteínas não-estruturais que auxiliam no processo de replicação e proteínas estruturais, importantes nas estratégias virais de infecção.

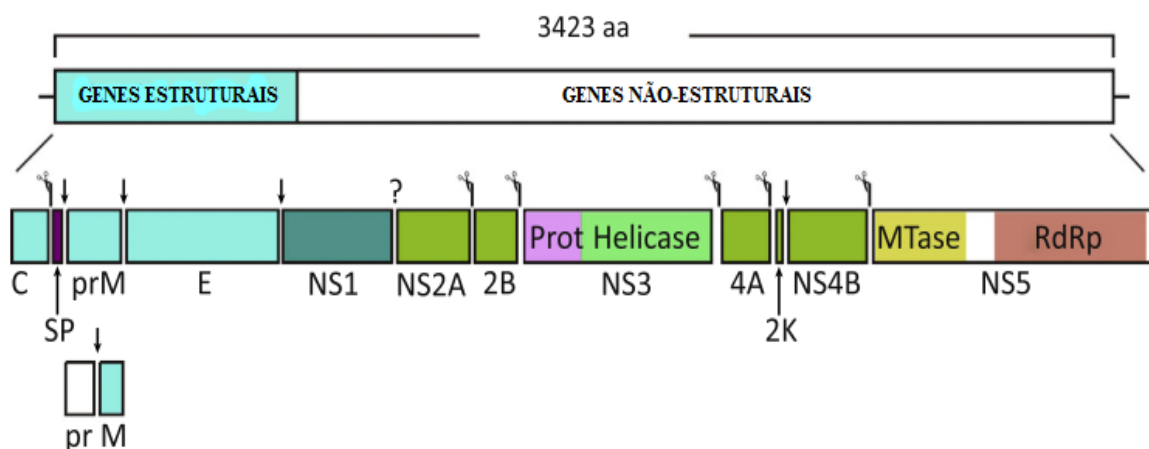
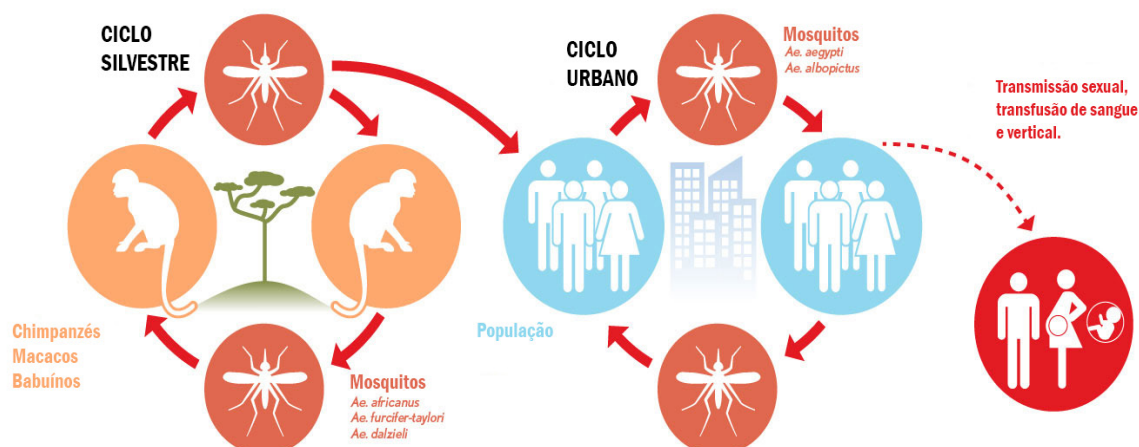


FIGURA 1: Genoma completo de flavivírus, demonstrando a codificação de genes relacionados a morfologia viral e os genes não-estruturais que estão responsáveis por auxiliar o processo de replicação e evasão imunológica da célula hospedeira.

FONTE: Adaptado de SHI and GAO, 2017.

3 | DINÂMICA DE TRANSMISSÃO DO VÍRUS ZIKA

Para que ocorra a transmissão do VZIK para seres humanos, ocorre primeiramente o ciclo silvestre e posterior urbano, através das picadas de mosquitos fêmeas infectadas com o vírus, sendo *Aedes aegypti* e *Aedes albopictus*, os principais vetores, assemelhando-se a forma de transmissão dos VDEN e Vírus Chikungunya. Os mosquitos do gênero *Aedes* estão confinados nas regiões tropicais e subtropicais, limitando assim o potencial de transmissão do VZIK.



Sources: CDC, PLOS, Reuters Credits: David Foster, Laurie Garrett, Doug Halsey, Gabriella Meltzer

No entanto, estudos demonstraram transmissão perinatal e potencial risco para a transmissão por transfusão sanguínea. Aliado a isso, a transmissão sexual do VZIK foi sugerida por Faye e colaboradores (2011) e posteriormente o vírus foi detectado em amostras de sêmen e urina de um paciente com o quadro de hematospemia que havia apresentado sintomas da infecção duas semanas antes da coleta.

A presença de VZIK na urina, em títulos mais elevados por um período mais longo que em amostras de soro, sugere sua adequação como amostra para detecção viral e diagnóstico da infecção, sobretudo, depois de intervalos maiores após o início dos sintomas (entre 10 ou mais de 20 dias).

4 | MANIFESTAÇÕES CLÍNICAS

As manifestações clínicas resultantes da infecção pelo VZIK, estão relacionadas com um grande número de pacientes assintomáticos, aproximadamente 80%, tendo como contrastes as manifestações sintomáticas intimamente relacionadas com outras infecções causadas por flavivirus, como VDEN e VCHIK. O VZIK é incubado por cerca de 12 dias, os pacientes sintomáticos apresentam febre, erupção cutânea mialgia, artralgia, conjuntivite, distúrbios gastrointestinais e dor de cabeça.

A apresentação clínica da infecção por VZIK geralmente não é específica e pode ser confundida com outras arboviroses, mais comumente DENV e CHIKV. Em outubro de 2015, casos de microcefalia começaram a ser investigados no nordeste do Brasil para verificar a hipótese de uma associação com o VZIK.

A microcefalia geralmente resulta do desenvolvimento anormal do cérebro. Suas consequências a longo prazo dependem das anomalias fundamentais do cérebro e podem variar de atrasos leves no desenvolvimento a déficit intelectual e motor, como paralisia cerebral. Além das infecções congênitas, a mesma pode resultar de anormalidades cromossômicas, exposição a drogas, álcool ou outras toxinas ambientais, fusão prematura dos ossos do crânio (craniossinostose) e determinados distúrbios metabólicos.

5 | EPIDEMIOLOGIA DO VÍRUS ZIKA

A distribuição geográfica dos arbovírus se estende pelos continentes e por todas as regiões, em especial nas regiões tropicais, devido às condições favoráveis. Apesar de hospedeiros e vetores existirem durante todo o ano, o ciclo de transmissão é interrompido em países de clima temperado na estação de inverno.

O vírus pode ser rapidamente disseminado, uma vez que uma única fêmea

de *Aedes aegypti* pode infectar vários hospedeiros por meio de suas alimentações curtas (CONSOLI & OLIVEIRA, 1998). No início do ano de 2015, métodos moleculares confirmaram a circulação do VZIK na Bahia (CAMPOS; BANDEIRA; SARDI, 2015) e em São Paulo, logo em seguida, nos estados do Rio Grande do Norte, Alagoas, Maranhão, Pará e Rio de Janeiro.

Aliado a isso, durante o surto da infecção pelo VZIK reportado recentemente no nordeste do Brasil no início do ano de 2015, as autoridades em saúde pública receberam relatos de médicos a cerca de um aumento no número de recém-nascidos com microcefalia. Logo em seguida, o Ministério da Saúde do Brasil (MS) confirmou um crescimento na prevalência de nascimentos com microcefalia no nordeste do país, comparado às estimativas registradas anteriormente (cerca de 0,5/10.000 nascidos vivos), que são baseadas na revisão das certidões de nascimento e incluem as descrições das maiores anomalias congênitas. Desde então, o MS estabeleceu um registro de microcefalia no Brasil, e em novembro de 2015, divulgou o aumento no número de casos de microcefalia.

Com frequência, as infecções por VZIK se tornaram uma realidade, com consequências inimagináveis para uma infecção causada por arbovírus. Embora a ocorrência de microcefalia e síndrome de Guillain-Barré seja atribuída ao VZIK, ainda há muito a ser investigado. Em 1º de fevereiro de 2016, a OMS declarou agrupamento de microcefalia e VZIK, principalmente observado nas regiões brasileiras, como “Emergência de Saúde Pública de Interesse Internacional” (PHEIC), isto é indicativo da extensão da emergência global.

6 | RNAs LONGOS NÃO-CODIFICANTES (lncRNAs)

Os lncRNAs são RNAs com 200 nucleotídeos ou mais que são sintetizados abundantemente por diversos tipos celulares. Apesar do conhecimento voltado aos lncRNAs ainda ser módico, é elucidado que eles são encarregados de promover a regulação celular, como desenvolvimento, diferenciação e metabolismo.

As alterações fisiopatológicas ocasionadas pela infecção viral nas células do hospedeiro, estão relacionadas com a expressão diferencial de RNAs não-codificantes de proteínas responsáveis pela regulação gênica celular, como os lncRNAs (*long non-coding RNAs*).

Embora a patogênese do VZIK hoje seja bem elucidada, muitos estudos ainda são necessários para uma compreensão de forma mais completa de como ocorre a interação vírus-hospedeiro. Diante disso, uma vasta área de transcritos são alvos importantes para ajudar a compreender tal interação, uma vez que o vírus utiliza de vias diferenciadas que estão relacionadas com importantes mecanismos fisiológicos celulares - afim de evadir da resposta imune e manter seu ciclo de replicação - sendo ativadas a partir da regulação gênica através de seus mediadores, como os lncRNAs

7 | MATERIAL E MÉTODOS

7.1 Preparação da biblioteca e sequenciamento de RNA

As bibliotecas de RNA-seq foram preparadas de acordo com o protocolo sugerido pelo guia de preparação de amostras TruSeq da Illumina. Resumidamente, 3 a 5 µg de RNA total foram usados para o sequenciamento. O RNA restante foi usado para a construção de ambas as cadeias de cDNA. A biblioteca foi então sequenciada utilizando protocolos e reagentes padrão da Illumina em instrumentação NextSeq 550 (Illumina). Os arquivos Fastq foram gerados pelo software CASAVA 1.8 offline. Análise de controle de qualidade adicional foi realizada usando FastQC (Ref). A qualidade média da sequência por base para todas as amostras foi superior a 32.

7.2 Análise dos dados

As leituras foram alinhadas ao genoma de referência do *Mesocricetus auratus* disponível no NCBI usando o programa Tophat (Ref) com os seguintes parâmetros “-G *mesocricetus.gtf* -l 2000 -i 40 --no-coverage-search UCSC/sacCer3/Bowtie2Index/genome”. Limites de transcrição e quantificação foram determinados usando o programa cufflinks (Ref). A visualização de dados foi realizada no programa R Studio (<http://www.R-project.org/>).

Os transcritos foram considerados novos se eles não se sobrepusessem com transcritos anotados ou UTRs das extremidades fossem maiores de 50 pb diferentes dos RNAs não codificantes previamente relatados SUTs entre outros.

8 | RESULTADOS E DISCUSSÃO

8.1 Identificação dos lncRNAs

Para a identificação de lncRNAs que foram submetidas cerca de 193 milhões de *reads* aos filtros de controle de qualidade para os efeitos aleatórios da variabilidade deixando apenas transcritos que estavam presentes ou eram secundários para RNA-Seq. Para as leituras filtradas, a montagem *de novo* foi realizada, e 193.846 contigs foram gerados com um valor N50 igual a 1,171 e taxa efetiva de leituras mapeadas de 84,66%. Estes contigs foram usados como sequências de referência para a identificação de lncRNA putativos em Hamster.

A dificuldade em caracterizar lncRNA em *Mesocricetus auratus* a falta de um robusto banco de dados lncRNA e a baixa conservação do genoma entre hamsters, exigiram um processamento em vários estágios dos dados com objetivos específicos. A primeira foi a remoção dos transcritos codificadores de proteínas (exclusão de transcritos que codificassem proteínas conhecidas). Como resultado, acima de 70K contigs foram anotados (valor E <10⁻⁵) que foram descartados de análise posterior.

Em segundo lugar, um filtro altamente rigoroso foi aplicado para cobertura contig.

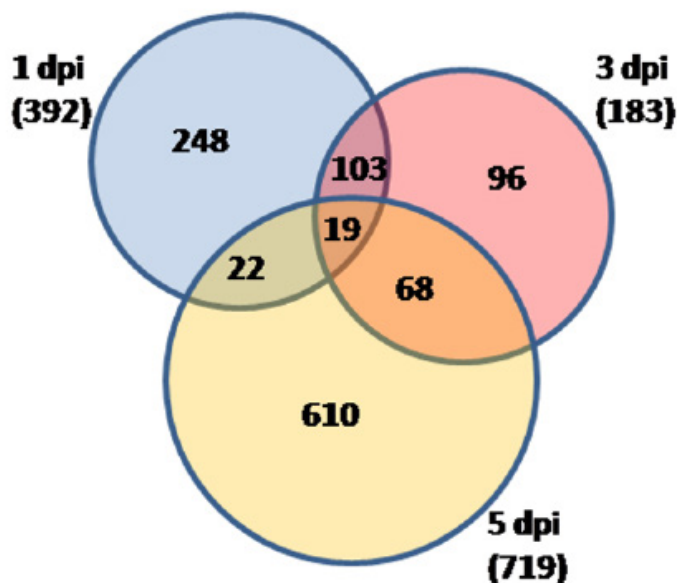


Figura 3: Representação da expressão diferencial de lncRNAs

Este filtro permite minimizar o viés da montagem *de novo* e descarta os contigs com baixa cobertura (12k contigs com alta cobertura <50 reads / pb). Como novos transcritos de codificação de proteína poderiam se associar a novos transcritos não-codificantes no processo de predição, foi utilizado o Calculador do Potencial de Codificação (CPC) entre 28-30 em para avaliar o potencial de codificação de proteínas de novos transcritos e remover transcritos putativos de codificação de proteínas. Em seguida, inspecionamos a presença de quadros de leitura abertos (ORF) > 200 e a presença de domínios conservados de proteína putativa entre todos os quadros de tradução possíveis. Com base no valor de CPC e na previsão de ORF, os contigs não mostraram evidência de codificação de domínios potenciais e putativos conservados dentro das suas sequências de aminoácidos previstas. As transcrições restantes foram mapeadas contra o genoma de referência para garantir a qualidade da montagem. Desse conjunto, 5.635 não apresentaram sobreposição parcial de locos preditos e representam potenciais lncRNAs de *Mesocricetus auratus*.

8.2 Caracterização dos lncRNAs

Para examinar se os lncRNAs identificados neste estudo exibem características típicas observadas em outros mamíferos (principalmente Hamsters), analisamos características como potencial de codificação, comprimento de sequência, conteúdo de GC e conservação de sequências com espécies intimamente relacionadas. Como os lncRNAs eram estritamente definidos por sua incapacidade de codificar proteína, determinamos a probabilidade de codificação de nossos lncRNAs recém-identificados e os comparamos com os conhecidos lncRNA, 3'UTR, 5'UTR e mRNA codificador de proteínas. Descobrimos que, à semelhança de outras sequências não codificantes,

como o conhecido lncRNA, 3'UTR e 5'UTR, nossos novos transcritos de lncRNA têm probabilidade de codificação extremamente baixa quando comparados ao mRNA codificador de proteínas. Além disso, descobrimos que os novos e conhecidos lncRNAs (fornecidos por anotação genômica de *Mesocricetus auratus*) eram mais curtos do que os transcritos codificadores de proteínas. O comprimento médio dos novos e conhecidos lncRNAs foi de 815 pb e 137727pb respectivamente, enquanto o mRNA codificador de proteínas tem um comprimento médio de 3220 pb.

Semelhante aos relatos anteriores (LIU; DING, 2017; OUYANG; HU; CHEN, 2016), observamos que os lncRNAs identificados neste estudo apresentavam um conteúdo GC menor que os mRNAs codificadores de proteínas. Por exemplo, o conteúdo médio GC de novos lncRNA e mRNA foi de 38,1% e 45,2%, respectivamente. Por outro lado, o conteúdo GC médio da sequência 5'UTR e 3'UTR foi de 43,1% e 34,6%, respectivamente. No geral, o conteúdo de GC da sequência não codificadora foi relativamente menor que a sequência de codificação.

Ao Determinar os lncRNAs que foram altamente conservados para outras espécies relacionados as espécies de estudo, semelhante ao estudo anterior (CULSHAW; MONGKOLSAPAYA; SCREATON, 2018; OUYANG; HU; CHEN, 2016; WANG et al., 2017), observamos sequencias com alto grau de semelhança. Ao agrupar esses lncRNAs altamente conservados, descobrimos que a maioria deles compartilhava semelhanças de sequência com Hamster domésticos, sugerindo que eles eram presumivelmente específicos do gênero. Embora os lncRNAs do *Mesocricetus auratus* compartilhe alta similaridade, principalmente *Mus musculus*, em comparação com os lncRNAs totais. Apenas 11% (1.258 dos 11.376 lncRNAs) apresentaram alta similaridade com espécies evolutivamente mais relacionados.

REFERÊNCIAS

BASARAB, M.; BOWMAN, C.; AARONS, E. J.; e CROPLEY, I. Zika Virus. **BMJ** 352:i1049,2016

BUATHONG, R.; HERMANN, L.; THAISOMBOONSUL, B.; RUTVISUTTINUNT, W.; KLUNGTHONG, C.; CHINNAWIROTPISAN, P.; et al. Detection of Zika Virus Infection in Thailand, 2012-2014. **Am. J. Trop. Med. Hyg.** 93: 380-383. 2015

CASSEB, A. R.; CASSEB, L. M. N.; SILVA, S. P. VASCONCELOS, P. F. C. Arbovírus: Importante Zoonose na Amazônia Brasileira. **Veterinária e Zootecnia**, v.20, n.3, p. 9-21, 2013.

CHAMBERS, T. J., HAHN, C. S., GALLER, R.; RICE, C. M. Flavivirus: genome

organization, expression and replication. **Revista de Microbiologia**.44: p. 649-688, 1990.

CULSHAW, A.; MONGKOLSAPAYA, J.; SCREATON, G. The immunology of Zika Virus. **F1000Research**, v. 7, n. 0, p. 203, 2018.

DICK, G. W.; KITCHEN, S. F.; HADDOW, A. J. Zika virus. I. Isolations and serological specificity. **Trans R Soc Trop Med Hyg**, n. 5 v. 46, p. 509-20, 1952.

- GOURINAT, A. C.; O'CONNOR, O.; CALVEZ, E.; GOARANT, C.; DUPONT-ROUZEYROL, M; Detection of zika virus in urine. **Emerging Infectious Diseases**, v. 21, n. 1, p. 84–86, 2015.
- GRITSUN, T.S., TUPLIN, A.K., GOULD, E.A., 2006. Origin, evolution and function of flavivirus RNA in untranslated and coding regions: implications for virus transmission. In: Kalitzky, M., Borowski, P. (Eds.), *Flaviviridae: Pathogenesis, Molecular Biology and Genetics*. **Horizon Scientific Press**, pp. 47–99
- HOLBROOK, R. M. Historical perspectives on Flavivirus Research. **Viruses**. **9**: 97, 2017
- LIU, W.; DING, C. Roles of LncRNAs in Viral Infections. **Frontiers in Cellular and Infection Microbiology**, v. 7, n. May, 26 maio 2017.
- McGRAW, E. A.; O'NEILL, S. L. Beyond insecticides: new thinking on an ancient problem. **Nature Reviews Microbiology**, v.11, n.3, p. 181-193, 2013.
- MERCER, T. R.; MATTICK, J. S. Structure and function of long noncoding RNAs in epigenetic regulation. **Nature Structural & Molecular Biology**, v.20, n.3, p. 300-307, 2013.
- MOTA O. T. M; Moquito-transmitted viruses - the great Brazilian Challenge **Brazilian Journal Of Microbiology**, 2016
- MUSSO, D.; NHAN, T.; ROBIN, E.; ROCHE, C.; BIERLAIRE, D.; ZISOU, K.; SHAN, A. Y.; CAO-LORMEAU, V. M.; BROULT, J.; Potential for Zika virus transmission through blood transfusion demonstrated during an outbreak in French Polynesia, November 2013 to February 2014. **Euro Surveill**, n.14 v.19, p. 1-3, 2014
- NEUFELDT, C. J.; CORTESE, M.; ACOSTA, E. G.; BARTENSCHLAGER, R. Rewiring cellular networks by members of the Flaviviridae Family. **Nature Reviews Microbiology**, v.16, n.3, p. 125-142, 2018.
- OUYANG, J.; HU, J.; CHEN, J.-L. lncRNAs regulate the innate immune response to viral infection. **Wiley Interdisciplinary Reviews: RNA**, v. 7, n. 1, p. 129–143, jan. 2016.
- PAPAIT, R.; KUNDERFRANCO, P.; STIRPARO, G. G.; LATRONICO, M. V. G.; CONDORELLI, G. Long Noncoding RNA: a New Player of Heart Failure?. **Journal of Cardiovascular Translational Research**, v.6, n.6, p. 876-883, 2013.
- REGO, S.; PALÁCIOS, M.; Ética, saúde global e a infecção pelo vírus zika: uma visão a partir do Brasil. **Revista Bioética**, n.3 v.24, p. 430-434, 2016.
- SCHULER-FACCINI, L. et al. Possível associação entre a infecção pelo vírus zika e a microcefalia — Brasil, 2015. **Weekly**, v. 65, n. 3, 2016.
- Strauss, J.H., Strauss, E.G. Evolution of RNA viruses. **Microbiol.** v.42, 657–683, 1988.
- THOMAS, S. M.; OBERMAYR, U.; FISCHER, D.; KREYLLING, J.; BEIERKUHNLEIN, C. Low-temperature threshold for egg survival of a post-diapause and non-diapause European aedine strain, *Aedes albopictus* (Diptera: Culicidae). **Parasites and Vectors**, v.5, n.1, p. 1-7, 2012.
- WANG, Z.; WANG, P.; AN, J. Zika virus and Zika fever. **Virologica Sinica**, v. 31, n. 2, p. 103–109, 2016.
- WANG, A. et al. Zika virus genome biology and molecular pathogenesis. **Emerging Microbes & Infections**, v. 6, n. 3, p. e13, 2017.

ZANLUCA, C.; MELO, V. C. A.; MOSIMANN, A. L. P.; SANTOS, G. I. V.; LUZ, K.; First report of autochthonous transmission of Zika virus in Brazil. **Memorias do Instituto Oswaldo Cruz**, v. 110, n. 4, p. 569–572, 2015

ANALISE *IN SILICO* DA FARMACOCINÉTICA E FARMACODINÂMICA DO COMPOSTO BENZOTIAZÓLICO COM POTENCIAL ANTITUMORAL CONTRA LINHAGEM DE ADENOCARCINOMA GÁSTRICO

Felipe Pantoja Mesquita

Universidade Federal do Ceará, Núcleo de
Pesquisa e Desenvolvimento em Medicamentos,
Faculdade de Medicina
Fortaleza-Ceará

Luina Benevides Lima

Universidade Federal do Ceará, Núcleo de
Pesquisa e Desenvolvimento em Medicamentos,
Faculdade de Medicina
Fortaleza-Ceará

Julio Paulino Daniel

Universidade Federal do Ceará, Núcleo de
Pesquisa e Desenvolvimento em Medicamentos,
Faculdade de Medicina
Fortaleza-Ceará

Adrhyann Jullyanne de Sousa Portilho

Universidade Federal do Ceará, Núcleo de
Pesquisa e Desenvolvimento em Medicamentos,
Faculdade de Medicina
Fortaleza-Ceará

Lais Lacerda Brasil de Oliveira

Universidade Federal do Ceará, Núcleo de
Pesquisa e Desenvolvimento em Medicamentos,
Faculdade de Medicina
Fortaleza-Ceará

Emerson Lucena da Silva

Universidade Federal do Ceará, Núcleo de
Pesquisa e Desenvolvimento em Medicamentos,
Faculdade de Medicina
Fortaleza-Ceará

Eliza de Lucas Chazin

Universidade Federal Fluminense, Departamento

de Química Orgânica, Instituto de Química.
Niterói – Rio de Janeiro

Thatyana Rocha Alves Vasconcelos

Universidade Federal Fluminense, Departamento
de Química Orgânica, Instituto de Química
Niterói - Rio de Janeiro

Maria Elisabete Amaral de Moraes

Universidade Federal do Ceará, Núcleo de
Pesquisa e Desenvolvimento em Medicamentos,
Faculdade de Medicina
Fortaleza-Ceará

Raquel Carvalho Montenegro

Universidade Federal do Ceará, Núcleo de
Pesquisa e Desenvolvimento em Medicamentos,
Faculdade de Medicina
Fortaleza-Ceará

RESUMO: O adenocarcinoma gástrico é uma malignidade com elevada incidência e mortalidade no mundo. Buscando novas alternativas de tratamento, esse trabalho teve como objetivo avaliar a atividade antitumoral do benzotiazol AFN01 na linhagem de adenocarcinoma gástrico ACP03. Além disso, realizar a predição computacional da farmacocinética e farmacodinâmica dessa molécula, bem como o *docking* molecular com um alvo farmacológico relevante encontrado na predição. Para tanto, foi realizado o ensaio do MTT para a avaliação do potencial citotóxico e

a microscopia de fluorescência para medir a capacidade de indução de morte celular, com os fluoróforos Diacetato de Fluoresceína, Hoechst 33342 e Iodeto de Propídeo. O resultado da curva concentração-resposta revelou um efeito citotóxico promissor na linhagem com valores de CI50 em torno de 17,24 μM para 24h, 8,23 μM para 48h e 1,39 μM para 72h de tratamento. A AFN01 foi capaz de induzir apoptose nas concentrações de 1 e 2 μM . As análises computacionais identificaram propriedades farmacocinéticas relevantes como capacidade de absorção intestinal positiva, inibição do CYP1A2, CYP2C9 e CYP3A4, inibição fraca de hERG e toxicidade oral aguda classe III. A predição farmacodinâmica identificou alvos moleculares relacionados à atividade antitumoral. A enzima TDP1 foi selecionada para o *docking* molecular, indicando uma energia de ligação de -7.5 kcal/mol com o benzotiazol e importantes interações com resíduos do sítio catalítico da enzima. Em conclusão, o composto benzotiazólico AFN01 se mostrou um excelente candidato para o tratamento do adenocarcinoma gástrico, com bons parâmetros farmacocinéticos preditos e possível alvo farmacológico relevante do ponto de vista molecular.

PALAVRAS-CHAVE: FARMACOCINÉTICA; FARMACODINÂMICA; BENZOTIAZOL

ABSTRACT: Gastric adenocarcinoma has high incidence and mortality rates worldwide. The aim of this study was to evaluate the antitumor activity of benzothiazole AFN01 against ACP03 gastric adenocarcinoma cell line. Furthermore, computational prediction of the pharmacokinetics and pharmacodynamics of this molecule and the molecular docking with the relevant pharmacological target found were performed. The MTT assay was performed to evaluate the cytotoxic potential of the compound and fluorescence microscopy was used to measure cell death induction with Fluorescein Diacetate, Hoechst 33342 and Propidium Iodide fluorophores. The result of the concentration-response curve obtained by MTT assay revealed a relevant cytotoxic effect against the lineage with IC50 values around 17.24 μM at 24 h, 8.23 μM at 48 h and 1.39 μM at 72 h of exposure. AFN01 was able to induce apoptosis at 1 and 2 μM . Computational analyses identified relevant pharmacokinetic properties such as positive intestinal absorption, inhibition of CYP1A2, CYP2C9 and CYP3A4, as well as weak inhibition of hERG and class III acute oral toxicity. The pharmacodynamic prediction identified molecular targets related to antitumor activity. The enzyme TDP1 was selected for molecular docking, indicating that benzothiazole has a binding energy of -7.5 kcal/mol and important interactions with catalytic site residues of the enzyme. In conclusion, the AFN01 benzothiazole compound is an excellent candidate for the treatment of gastric adenocarcinoma, with good predicted pharmacokinetic parameters and a potentially new molecular target.

KEYWORDS: PHARMACOKINETICS; PHARMACODYNAMICS; BENZOTHIAZOLE

1 | INTRODUÇÃO

O adenocarcinoma gástrico é uma das malignidades originadas no estômago com elevada incidência e mortalidade no mundo. Apesar de um aparente declínio em

sua incidência, este ainda tem um prognóstico ruim e é muitas vezes diagnosticado em estágios avançados, o que leva ao aumento nas taxas de mortalidade (CARCAS, 2014; GABR *et al.*, 2014). Nesse sentido, o câncer gástrico exige atenção continuada e novas pesquisas relacionadas à prevenção, detecção precoce e novas opções terapêuticas.

Dependendo do tamanho e localização do tumor primário, a terapia mais utilizada é a ressecção cirúrgica com gastrectomia total ou subtotal. As terapias alvo dirigidas têm sido importantes para o aprimoramento das intervenções terapêuticas (CARCAS, 2014). Vários estudos vêm sendo desenvolvidos na intenção de ampliar as opções de tratamentos disponíveis com o intuito de melhorar a resposta farmacológica e a sobrevida dos pacientes (CARCAS, 2014; QIN, 2014).

Nesse contexto, os benzotiazóis constituem uma classe importante de agentes terapêuticos. Compostos derivados dessa molécula já demonstraram diversos efeitos farmacológicos, incluindo atividades antimicrobiana, antiinflamatória e antitumoral (GABR *et al.*, 2014). Nogueira e colaboradores (2010) sintetizaram onze hidrazonas contendo o núcleo benzotiazólico, dentre essas, o composto (*E*)-2-((2-(benzo[d]tiazol-2-il)hidrazono)metil)-4-nitrofenol (AFN01) foi o único que demonstrou citotoxicidade nas linhagens câncer de colón (HCT-8), de mama (MDAMB-435) e leucemia (HL-60). Além disso, no estudo de Vasconcelos e colaboradores (2018) foi avaliado a atividade deste composto em linhagens de melanoma maligno, comprovando seu potencial antitumoral, indutor de apoptose e inibidor de migração e invasão *in vitro*.

Com base na necessidade de novos fármacos como alternativas ao tratamento do adenocarcinoma gástrico, esse trabalho teve por objetivo testar a atividade antitumoral do benzotiazol (*E*)-2-((2-(benzo[d]tiazol-2il)hidrazono)metil)-4-nitrofenol na linhagem de adenocarcinoma gástrico do tipo difuso ACP03, realizar a predição computacional da farmacocinética e farmacodinâmica dessa molécula, bem como o *docking* molecular com o alvo farmacológico relevante encontrado na predição.

2 | METODOLOGIA

2.1 Cultivo celular

A linhagem de adenocarcinoma gástrico utilizada neste estudo foi estabelecida no Laboratório de Citogenética Humana situado na Universidade Federal do Pará (UFPA), a partir de um tumor primário de adenocarcinoma gástrico do tipo intestinal de um paciente natural do Brasil (LEAL *et al.*, 2009). As células foram cultivadas em Dulbecco's Modified Eagle's Medium (DMEM, Gibco®), suplementado com 10% de Soro Fetal Bovino (SFB, Gibco®), 100 U / mL de penicilina e 100 µg / mL de estreptomicina (Gibco®). A linhagem celular foi armazenada em estufa com ambiente contendo 5% de CO₂ a 37°C.

2.2 Ensaio MTT

Acitotoxicidade do benzotiazolAFN01 foi avaliada na linhagem de adenocarcinoma gástrico ACP03 e determinada por meio do ensaio do MTT (BERRIDGE *et al.*, 1996). A linhagem foi semeada na densidade de 3×10^3 células por poço em placas de 96 poços e incubados por 24 h. Em seguida, a solução estoque de AFN01 (10 mM) foi diluída em meio de cultura para obtenção de uma curva de concentração (10 μ M a 0,156 μ M). O controle negativo foi tratado com o solvente DMSO (0,01%) e o controle positivo com doxorubicina por 24, 48 e 72 horas. A densidade ótica foi medida por espectrofotometria no comprimento de onda de 550 nm. Os valores de CI50 (concentração inibitória de 50% do crescimento celular) e IC95% (intervalo de confiança de 95%) foram obtidos a partir de uma regressão não linear sigmoidal seguindo a equação $Y=100/(1+10^{((\text{LogIC50}-X)*\text{HillSlope}))}$.

2.3 Estudo do padrão de morte celular

A capacidade de indução de morte celular foi realizada através da análise morfológica das células por microscopia de fluorescência utilizando os fluoróforos Diacetato de Fluoresceína (DAF), Hoechst 33342 (HO) e Iodeto de Propídeo (IP). O método se baseia na revelação diferencial das células viáveis, apoptóticas e necróticas frente a esses fluoróforos. O DAF é lipossolúvel e cora o citoplasma em verde, enquanto que o HO lipossolúvel intercala no DNA corando o núcleo em azul. Por outro lado, o IP é um intercalante de DNA hidrossolúvel e só marca o núcleo se houver perda da integridade da membrana plasmática (CHAN *et al.*, 2015).

Dessa maneira, as células viáveis possuem membrana íntegra em verde e apresentam núcleo uniforme corado em azul. As células em apoptose (membrana ainda íntegra) apresentam manchas azuis brilhantes com condensação da cromatina e fragmentação nuclear, por vezes são marcadas por IP na apoptose tardia ou secundária em que ocorre o afrouxamento da membrana permitindo que o corante incorpore no núcleo. Há ainda alterações na membrana celular em decorrência da formação de corpos apoptóticos. Já as células em necrose são coradas inteiramente em vermelho, devido a perda da integridade plasmática, e apresentam-se com tamanho celular aumentado (GALLUZZI *et al.*, 2018).

2.4 Análise por ferramentas computacionais

2.4.1 Desenho do ligante

A estrutura bidimensional (2D) do composto foi desenhada no programa MarvinSketch®. A partir do modelo, obteve-se o SMILES (Simplified Molecular Input Line Entry Specification), uma forma de representar a molécula com caracteres comuns.

2.4.2 Predição farmacocinética

As propriedades de similaridade a fármacos (*drug-likeness*), baseadas nas Regras de Lipinski (LIPINSKI *et al.*, 1997; LIPINSKI, 2000), também foram preditas. Essa regra caracteriza fármacos de melhor absorção como aqueles com $\log P \leq 5$ (log do coeficiente de partição da molécula), peso molecular ≤ 500 Da (PM), número de aceptores de ligação de hidrogênio ≤ 10 (ALH), número de doadores de ligação de hidrogênio ≤ 5 (DLH) e área de superfície polar ≤ 140 Å (ASP). Além disso, uma regra adicional, proposta por Veber, diz que o número de ligações rotacionáveis (NLR) não deve ser superior a 10, visto que a molécula deve adotar uma conformação fixa para interagir com o alvo, passar pelas membranas, e ao mesmo tempo adotar a conformação ativa (VEBER *et al.*, 2002). A violação de duas ou mais dessas regras podem indicar possíveis problemas na biodisponibilidade oral. Para tanto, o *SMILES* foi utilizado no servidor molinspiration.

A ferramenta admetSAR foi usada para calcular os parâmetros de absorção, dispersão, biotransformação e toxicidade teóricos dos benzotiazóis (CHENG *et al.*, 2012). Esta informação é útil para a condução de estudos que necessitam de uma avaliação do risco para humanos. O método da ferramenta é baseado na similaridade estrutural com outros compostos de parâmetros farmacocinéticos já conhecidos e descritos em modelos biológicos, presentes no banco de dados do servidor. Para a realização deste método, o *SMILES* da molécula estudada foi submetido no servidor admetSAR (CHENG *et al.*, 2012)

2.4.3 Avaliação do perfil farmacodinâmico

Os possíveis alvos terapêuticos da molécula AFN01 foram avaliados por meio do servidor SwissTarget Prediction. Esse servidor compara a similaridade dos compostos com uma biblioteca de 280.000 compostos ativos, em mais de 2.000 alvos farmacológicos, em organismos diferentes (GFELLER; BASSANI-STERBERG, 2018). Novamente, o *SMILES* da molécula foi submetido à ferramenta online, e os alvos foram classificados segundo uma função de probabilidade.

2.4.4 Preparo do ligante

Com o uso do programa Avogadro[®], a estrutura 2D informada pelo *SMILES* foi convertida no esqueleto tridimensional (3D), seguindo a correta estereoquímica. Em seguida, todas foram submetidas a otimização geométrica do Avogadro[®] e salvas em “.mol2”. O *software* AutoDock Tools (ADT) foi utilizado para converter as moléculas no formato “.mol2” em “.pdbqt”, contendo somente os hidrogênios polares e com as definições de ligações de rotação livre ou rígida.

2.4.5 Preparo do alvo farmacológico

Após a avaliação dos alvos farmacodinâmicos, foi feita a busca das estruturas cristalográficas dos potenciais receptores. O *Protein Data Bank* (PDB) foi utilizado para busca e download das estruturas no formato “.pdb”. O *software* PyMol® foi utilizado para visualização e manipulação dos arquivos, a fim de remover moléculas de água, ligantes nativos, outros reagentes e adicionar os átomos de hidrogênios.

2.4.6 Ancoragem (*docking*) molecular

Os arquivos do receptor e do ligante foi convertido, respectivamente, do formato “.pdb” para “.pdbqt”, por meio do *software* Auto Dock Tools (ADT). O mesmo programa também foi utilizado para visualização do tamanho e centro do *grid box*, que foi anotado para a definição das configurações do *docking* com o *software* Vina® (TROTT; OLSON, 2010). Além do tamanho nos eixos x, y e z, e das coordenadas do centro da caixa (nos mesmos eixos), o parâmetro de *exhaustiveness* (exaustividade) também foi definido no arquivo de configuração de cada *docking*. O valor padrão para a exaustividade é de 8, mas o valor escolhido para este trabalho foi de 150, permitindo um maior tempo de busca da conformação e aumentando a probabilidade de encontrar o mínimo global.

3 | RESULTADOS E DISCUSSÃO

A partir do ensaio do MTT foi obtida uma curva concentração-resposta com um efeito citotóxico relevante contra a linhagem ACP03 com valores de CI50 [Intervalo de confiança com 95%] em torno de 17,24 μM [14,92 – 19,92] para 24h, 8,23 μM [7,02 – 9,65] para 48h e 1,39 μM [1,14 – 1,69] para 72h. O resultado pode ser visualizado na Figura 1. Nogueira e colaboradores (2010) mostraram que o benzotiazol AFN01 obteve uma atividade anticâncer *in vitro* contra células de câncer de colon (HCT-8), mama (MDAMB-435) e leucemia (HL-60), sendo a linhagem leucêmica a mais sensível ao composto com um CI50 de 1,66 μM . Portanto, os resultados encontrados neste estudo mostram que o composto AFN01 é significativamente sensível contra a linhagem de câncer gástrico do tipo intestinal com uma potência de 1,39 μM após 72 horas de tratamento.

Este potencial citotóxico observado na linhagem de câncer gástrico é de extrema importância, visto que já foi descrito na literatura que esta linhagem celular possui uma expressiva quimiorresistência contra alguns fármacos utilizados rotineiramente na clínica, como a cisplatina (CI50 de 28,62 μM) e irinotecano (CI50 de 32,07 μM) (PINTO *et al.*, 2015). A citotoxicidade dos benzotiazóis descrita na literatura é variável. Alguns trabalhos mostram que os derivados de benzotiazóis podem ter uma potência em torno de 20 μM e outros mostram uma potência baixa em torno de 2 μM , condizente com o efeito mostrado neste presente trabalho.

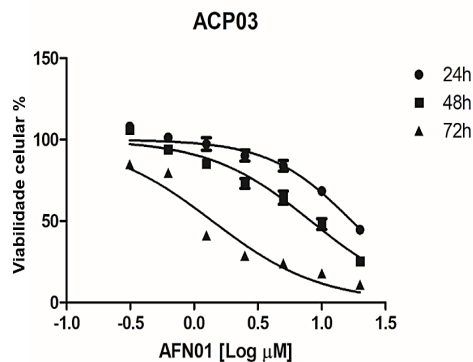


Figura 1: **Efeito citotóxico da AFN01 na linhagem ACP03.** A atividade citotóxica foi avaliada através do ensaio do MTT em três tempos distintos de tratamento, foi realizada uma curva de concentração da substância teste (10-0.156 μM).

Com o intuito de esclarecer o padrão de morte desencadeado pelo efeito citotóxico do composto AFN01, a análise por microscopia de fluorescência mostrou que o composto AFN01 foi capaz de induzir de forma significativa apoptose nas concentrações de 1 e 2 μM ($P < 0.05$), assim como a doxorrubicina, utilizada como controle positivo (Figura 2). Este efeito pró-apoptótico observado é consistente com o observado na literatura com outros benzotiazóis que foram identificados como bloqueadores do ciclo celular e indutores de morte celular por apoptose (WANG *et al.*, 2011; SINGH *et al.*, 2016).

A apoptose é um importante mecanismo de defesa para as células, impedindo a proliferação desordenada em condições normais e, portanto, tem sido considerada um alvo para o desenvolvimento de novos fármacos anticâncer (PFEFFER; SINGH, 2018). Portanto, por ser capaz de induzir a morte celular por apoptose em baixas concentrações, o benzotiazol AFN01 pode prover uma nova abordagem terapêutica.

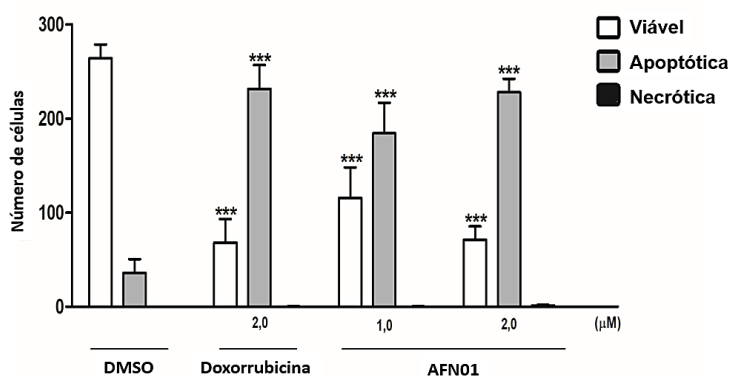


Figura 2: **Análise do mecanismo de morte celular.** Números de células viáveis, apoptóticas e necróticas. A doxorrubicina (2 μM) foi usado como controle positivo. Os dados são apresentados como a média \pm desvio padrão de três experimentos independentes. As amostras tratadas foram comparadas com o controle negativo DMSO, analisados estatisticamente usando ANOVA seguido por teste de Bonferroni. Diferenças significativas: *** $p < 0.001$.

As análises computacionais são de extrema importância para o desenvolvimento de novos fármacos para o mercado, pois fornecem informações úteis que direcionam os estudos a um baixo custo e sem utilização de animais de laboratório (YAN *et al.*, 2017). Primeiramente, a avaliação do perfil físico-químico pelo programa Molinspiration (tabela 1) identificou que a molécula AFN01 se adequou os parâmetros propostos por Lipinski e Veber (LIPINSKI *et al.*, 1997; LIPINSKI, 2000; Veber *et al.*, 2002). Portanto, este resultado indica que a molécula é passível de boas características farmacocinéticas quanto a absorção e biodisponibilidade.

Propriedades físico-químicas	
LogP	3,66
ASP	103,33
PM	314,33
ALH	7
DLH	2
NLR	4

Tabela 1 – Parâmetros preditos pelo servidor Molinspiration.

Na análise pelo servidor admetSAR foram identificadas propriedades farmacocinéticas relevantes, como capacidade de absorção intestinal positiva, confirmando a possibilidade de administração oral do composto. A inibição do CYP1A2, CYP2C9 e CYP3A4 também foi encontrada, importante na questão do metabolismo de fármacos e interações medicamentosas. Outro parâmetro avaliado foi a inibição de canais iônicos de potássio hERG que mostrou ser fraca, indicando uma baixa probabilidade de cardiotoxicidade (YU *et al.*, 2015). A presença de toxicidade oral aguda classe III (LD50 entre 500 e 5000 mg/kg) corrobora com a hipótese de que o benzotiazol AFN01 é um composto seguro e aceitável para o consumo humano (BOIK; NEWMAN, 2008). Entretanto, ainda há a necessidade dos estudos biológicos *in vivo* para a confirmação de tais resultados encontrados no modelo *in silico*.

A predição farmacodinâmica permitiu a triagem de alvos moleculares de importância para a atividade antitumoral. Nesta análise, os alvos farmacológicos são triados baseados na similaridade estrutural dos ligantes em estudo com ligantes com alvos já conhecidos. Dentre diversos alvos encontrados, a enzima Tiroxil-DNA fosfodiesterase-1 (TDP1) foi um dos alvos com maior probabilidade de ligação com a AFN01 (tabela 2).

Target	Uniprot ID	Gene Code	Probability
Microtubule-associated protein tau	P10636	MAPT	0.68
Tyrosyl-DNA phosphodiesterase 1	Q9NUW8	TDP1	0.63
Arachidonate 5-lipoxygenase	P09917	ALOX5	0.46
Arachidonate 15-lipoxygenase	P16050	ALOX15	0.46
Arachidonate 12-lipoxygenase, 12S-type	P18054	ALOX12	0.46
Arachidonate 15-lipoxygenase B	O15296	ALOX15B	0.46
Arachidonate 12-lipoxygenase, 12R-type	O75342	ALOX12B	0.46
Epidermis-type lipoxygenase 3	Q9BYJ1	ALOXE3	0.46
Focal adhesion kinase 1	Q05397	PTK2	0.29
Protein-tyrosine kinase 2-beta	Q14289	PTK2B	0.29

Tabela 2 - Predição dos alvos farmacológicos utilizando o servidor SwissTarget Prediction.

A TDP1 é uma enzima importante relacionada com mecanismos de reparo do DNA, atuando em conjunto com a enzima Topoisomerase-1, remove lesões na ponta 3' da fita de DNA e protege contra uma possível degradação durante o reparo (FLETT *et al.*, 2018). Após essa identificação, a enzima TDP1 foi selecionada para o *docking* molecular. O resultado mostrou que o benzotiazol possui uma energia de ligação de -7.5 kcal/mol e importantes interações com resíduos do sitio catalítico da enzima (Figuras 3 e 4). Interações como ligações de hidrogênio convencionais compõem a maior parte da interação entre o ligante AFN01 e a proteína TDP1, favorecendo a estabilidade da molécula dentro da cavidade catalítica. Estes dados são importantes no que diz respeito ao entendimento do mecanismo de ação do composto para produzir o efeito anticâncer esperado. Ao inibir a enzima TDP1, o composto AFN01 impede o mecanismo de reparo do DNA da célula tumoral, direcionando a célula para a via de morte celular programada (MEISENBERG *et al.*, 2015).

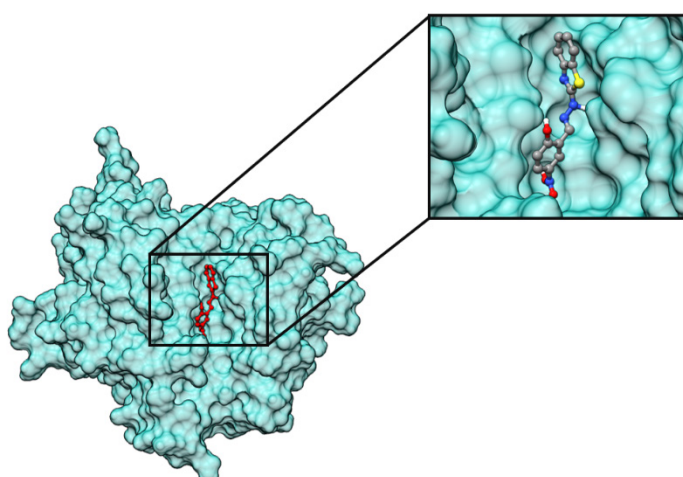


Figura 3: **Docking molecular da molécula AFN01.** A estrutura da proteína TDP1 está representada em azul e a molécula em vermelho. Ao lado, a ampliação da imagem mostra com mais detalhes o ligante acomodado na cavidade catalítica da enzima.

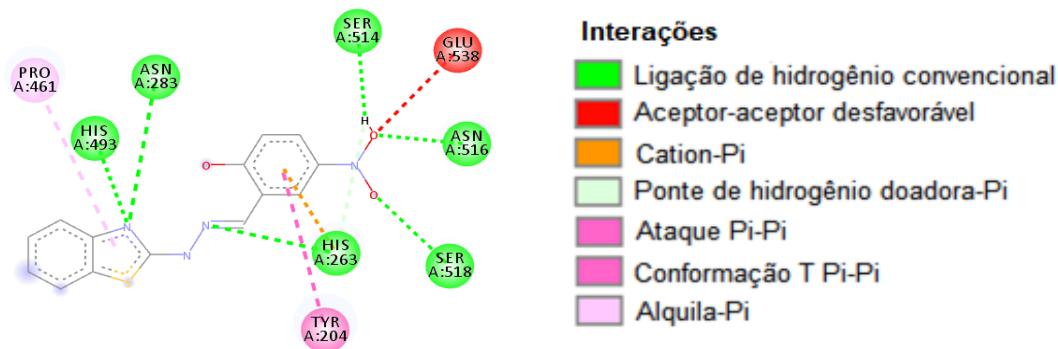


Figura 4: Representação das interações da molécula e do ligante.

4 | CONSIDERAÇÕES FINAIS

Em conclusão, o composto benzotiazólico AFN01 é um excelente candidato para o tratamento do adenocarcinoma gástrico, com parâmetros farmacocinéticos preditos aceitáveis e possível alvo farmacológico relevante do ponto de vista molecular. Entretanto, faz-se necessário a realização de mais ensaios para a confirmação dos efeitos celulares e moleculares, bem como os aspectos toxicológicos em modelos animais, para que então o possível fármaco possa seguir para as fases clínicas.

5 | AGRADECIMENTOS

Agradecemos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPQ) e Financiadora de Estudos e Projetos (FINEP) por tornarem possível que este trabalho fosse realizado.

REFERÊNCIAS

- BERRIDGE, M. V.; TAN, A. S.; MCCOY, K. D.; WANG, R. **The biochemical and cellular basis of cell proliferation assays that use tetrazolium salts.** *Biochemical*, v. 4, p. 14–19, jan. 1996.
- BOIK, J. C.; NEWMAN, R. A. **Structure-activity models of oral clearance, cytotoxicity, and LD50: a screen for promising anticancer compounds.** *BMC Pharmacol.* v. 8, n. 12, 2008.
- CARCAS, L. P. **Gastric cancer review.** *Journal of Carcinogenesis*, v. 13, n. :14, dez. 2014.
- CHAN, L. L.; KUKSIN, D.; LAVERTY, D. J.; SALDI, S.; QIU, J. **Morphological observation and analysis using automated image cytometry for the comparison of trypan blue and fluorescence-based viability detection method.** *Cytotechnology*, v. 67, n. (3), p. 461-473, mar. 2015.
- CHENG, F. LI, W.; ZHOU, Y.; SHEN, J.; WU, Z.; LIU, G.; LEE, P. W.; TANG, Y. **AdmetSAR: A comprehensive source and free tool for assessment of chemical ADMET properties.** *Journal of Chemical Information and Modeling*, v. 52, n. 11, p. 3099–3105, nov. 2012.

- FLETT, F. J.; RUKSENAITE, E.; ARMSTRONG, L. A.; BHARATI, S.; CARLONI, R.; MORRIS, E. R.; MACKAY, C. L.; INTERTHAL, H.; RICHARDSON, J. M. **Structural basis for DNA 3'-end processing by human tyrosyl-DNA phosphodiesterase 1**. *Nat Commun.*, v. 8, n. 24, jan. 2018.
- GABR, M. T.; EL-GOHARY, N. S.; EL-BENDARY, E. R.; EL-KERDAWY, M. M. **New series of benzothiazole and pyrimido[2,1-b]benzothiazole derivatives: synthesis, antitumor activity, EGFR tyrosine kinase inhibitory activity and molecular modeling studies**. *Medicinal Chemistry Research*, v. 24, n. 2, p. 860–878, fev. 2015.
- GALLUZZI, L.; VITALE, I. et al. **Molecular mechanisms of cell death: recommendations of the Nomenclature Committee on Cell Death 2018**. *Cell Death and Differentiation*, v. 25, p. 486-541, 2018.
- GFELLER, D.; BASSANI-STERBERG, M. **Predicting Antigen Presentation—What Could we Learn from a Million Peptides?**. *Frontiers in immunology*, v. 9, n. 2, p. 1716, jul. 2018.
- LEAL, M. F.; MARTINS DO NASCIMENTO, J. L.; DA SILVA, C. E.; VITA LAMARÃO, M.F.; CALCAGNO, D.Q.; KHAYAT, A. S.; ASSUMPÇÃO, P.P.; CABRAL, I. R.; DE ARRUDA CARDOSO, S. M.; BURBANO, R. R. **Establishment and conventional cytogenetic characterization of three gastric cancer cell lines**. *Cancer Genet Cytogenet.*, v. 195, n. 1, p. 85-91, nov. 2009.
- LIPINSKI, C. A. **Drug-like properties and the causes of poor solubility and poor permeability**. *Journal of Pharmacological and Toxicological Methods*, v. 44, n. 1, p. 235–249, mar. 2000.
- LIPINSKI, C. A.; LOMBARDO, F.; DOMINY, B. W.; FEENEY, P. J. **Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings**. *Advanced Drug Delivery Reviews*, v. 23, p. 3–25, jan. 1997.
- MEISENBERG, C.; GILBERT, D. C.; CHALMERS, A.; HALEY, V.; GOLLINS, S.; WARD, S.E.; EL-KHAMISY, S.F. **Clinical and cellular roles for TDP1 and TOP1 in modulating colorectal cancer response to irinotecan**. *Mol Cancer Ther.*, v. 14, n. 2, p. 575-85, fev. 2015.
- NOGUEIRA, A. F.; AZEVEDO, E.C.; FERREIRA, V. F.; ARAUJO, A. J.; SANTOS, E. A.; PESSOA, C.; COSTA-LOTUFO, L. V.; MONTENEGRO, R. C.; MORAES, M. O.; VASCONCELOS, T. R. A. **Synthesis and Antitumor Evaluation of (E)-2-Benzothiazole Hydrazones**. *Letters in Drug Design & Discovery*, v. 7, n. 8, p. 551-555, mai. 2010.
- PFEFFER, C. M.; SINGH, A. T. **Apoptosis: a target for anticancer therapy**. *International journal of molecular sciences*, v.19, n. 2, p. 448, fev. 2018.
- PINTO, L. C; SOARES, B. M.; PINHEIRO J. J.; RIGGINS, G. J.; ASSUMPÇÃO, P. P.; BURBANO, R. M.; MONTENEGRO, R. C. **The anthelmintic drug mebendazole inhibits growth, migration and invasion in gastric cancer cell model**. *Toxicology in vitro*, v. 29, n. 8, p. 2038-2044, dez. 2015.
- QIN, S. **Phase III study of apatinib in advanced gastric cancer: A randomized, double-blind, placebo-controlled trial**. *Journal of Clinical Oncology*, v. 32, n. 15 (Suppl.), p. 4003-4003, 2014.
- SINGH, M.; MODI, A.; NARAVAN, G.; SINGH, S. K. **Benzothiazole derivatives bearing amide moiety: potential cytotoxic and apoptosis-inducing agents against cervical cancer**. *Anti-Cancer Drugs*, v. 27, n. 6, p. 519-532, jul. 2016.
- TROTT, O.; OLSON, A. J. **AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading**. *Journal of computational chemistry*, v. 31, n. 2, p. 455-461, jan. 2010.
- VASCONCELOS, Z. S.; RALPH, A. C. L.; CALCAGNO, D. Q.; DOS SANTOS BARBOSA, G.;

PEDROSA, T. N.; ANTONY, L. P.; SMITH, M. A. C.; CHAZIN E. L.; VASCONCELOS, T. R. A. ; MONTENEGRO, R. C.; DE VASCONCELLOS, M. C. **Anticancer potential of benzothiazolic derivative (E)-2-((2-(benzo [d] thiazol-2-yl) hydrazono) methyl)-4-nitrophenol against melanoma cells.** *Toxicology in Vitro*, v. 50, n. 3, p. 225-235, ago. 2018.

VEBER, D. F.; JOHNSON, S.R.; CHENG, H. Y.; SMITH, B. R.; WARD, K. W.; KOPPLE, K. D. **Molecular properties that influence the oral bioavailability of drug candidates.** *Journal of Medicinal Chemistry*, v. 45, n. 12, p. 2615–2623, 2002.

WANG, Z. SHI, X.H.; WANG, J.; ZHOU, T.; XU, Y. Z.; HUANG, T. T.; LI, Y. F.; ZHAO, Y. L.; YANG, L.; YANG, S. Y.; YU, L. T.; WEI, Y. Q. **Synthesis, structure–activity relationships and preliminary antitumor evaluation of benzothiazole-2-thiol derivatives as novel apoptosis inducers.** *Bioorg Med Chem Lett*, v. 21, n. 4, p. 1097-101, fev. 2011.

YAN, G.; WANG, X.; CHEN, Z.; WU, X.; PAN, J.; HUANG, Y.; WAN, G.; YANG, Z. **In-silico ADME studies for new drug discovery: from chemical compounds to Chinese herbal medicines.** *Curr Drug Metab.*, v. 18, n. 6, p. 535-539, jul. 2017.

YU, Z.; IJZERMAN, A. P.; HEITMAN, L. H. **Kv11.1 (hERG)-induced cardiotoxicity: a molecular insight from a binding kinetics study of prototypical Kv11.1 (hERG) inhibitors.** *Br J Pharmacol.* v. 172, n. 3, p. 940-955, 2015.

ANALISE PRIMARIA DE TRANSCRIPTOMA DE TECIDO TESTICULAR DE HAMSTERS (MESOCRICETUS AURATUS) INFECTADOS COM VÍRUS ZIKA

Walter Felix Franco Neto

Instituto Evandro Chagas
Ananindeua – Pará

Samir Mansour Moraes Casseb

Instituto Evandro Chagas
Ananindeua – Pará

Karla Fabiane Lopes de Melo

Instituto Evandro Chagas
Ananindeua – Pará

Wallax Augusto Silva Ferreira

Instituto Evandro Chagas
Ananindeua – Pará

Ana Paula Sousa Araujo

Instituto Evandro Chagas
Ananindeua – Pará

Jardel Fabio Lopes Ferreira

Instituto Evandro Chagas
Ananindeua – Pará

Taiana Andrade Freitas

Instituto Evandro Chagas
Ananindeua – Pará

Milene Ferreira Silveira

Instituto Evandro Chagas
Ananindeua – Pará

Livia Carício Martins

Instituto Evandro Chagas
Ananindeua – Pará

Pedro Fernando da Costa Vasconcelos

Instituto Evandro Chagas
Ananindeua – Pará

RESUMO: O *vírus zika* (VZIK) é um Arbovirus que pertence à família Flaviviridae, gênero flavivirus, seu primeiro isolamento foi feito de amostras de macaco *Rhesus* na floresta de Zika em Uganda (1947). Durante o período de 2014 – 2015 ocorreu uma grande epidemia desse vírus associada a casos graves, como a microcefalia e síndrome de Guillain – barré, além disso, outras formas de transmissão também foram identificadas, como a transmissão sexual. Além de outras patologias associadas a infecção pelo vírus, que são: cegueira e infertilidade, porém, estas ainda são pouco esclarecidas. No âmbito dos avanços tecnológicos de análise genômica ocorridos nas últimas décadas, possibilitou-se mensurar o perfil de expressão dos genes que estão relacionados a diversas alterações, logo, embasar de forma específica as estratégias para o desenvolvimento de medidas profiláticas e terapêuticas. O objetivo deste trabalho foi observar alteração no nível dos transcritos durante a infecção pelo vírus Zika. Para isso, fizemos uma análise primária de sequências de testículos oriundas de um projeto experimental de infecção do *Vírus Zika* em hamsters dourados, cujo organismo possui uma reposta imune definida, além de demonstrar alterações semelhantes descrita em humanos, essas sequências foram analisadas através da plataforma online Galaxy Project e foram encontradas a alta e baixa expressão de

genes no decimo dia de infecção, sugerindo a possível a possível influencia do vírus.

PALAVRAS CHAVES: Zika; hamster dourado; testículo; transcriptoma.

1 | INTRODUÇÃO

O vírus *Zika* (VZIK) é um arbovirus, sendo este grupo de vírus transmitido por artrópodes (Casseb et al.,2013). Este vírus vem tornando-se um dos arbovírus de maior atenção da comunidade científica devido as manifestações causadas por sua infecção, principalmente envolvendo neonatos, onde foram descritas crianças com microcefalia, proveniente de uma afinidade do vírus a infectar células neurais progenitoras prejudicando o crescimento do cérebro, além de poder causar a síndrome de Guillain-Barré (OLIVEIRA, et al 2016) (MINER, et al 2017).

Outros aspectos também foram de destaque para este vírus, como a transmissão por vias pouco comuns entre os arbovírus, onde é possível a transmissão por relações sexuais e por fluidos como a lagrima, e também causar lesões nos órgãos que contemplam esses sistemas; dependendo da evolução do quadro dos hospedeiros infectados e da resposta a infecção, há a possibilidade de cegueira e infertilidade (MINER, et al 2017) (OLIVEIRA, et al 2016).

O VZIK foi isolado pela primeira vez na floresta de Zika em Uganda em 1947, através do sangue de um macaco *Rhesus* que apresentava um quadro febril.(DICK, et al 1952) VZIK pertence à família Flaviviridae do gênero flavivirus, é um vírus envelopado com formato icosaédrico, além de possuir um genoma com sentido positivo que codifica 3 proteínas estruturais (prM, M e E) e 7 proteínas não estruturais (NS1, NS2A, NS2B, NS3, NS4A, NS4B e NS5), semelhante a outros flavivirus conhecidos como os vírus: *Vírus Febre da Amarela* (VFA), VDEN, *Vírus Ilhéus* (VILH), *Vírus Rocio* (VROC), *Vírus da Encefalite de Saint Louis* (VESL), *Vírus do Nilo Ocidental* (VNO) e *Vírus da Encefalite Japonesa* (VEJ) (LOPEZ-DENMAN, et al 2017) (SIROHI, et al 2016).

Durante infecções virais em mamíferos são desencadeadas diversas respostas antivirais, e mensurar essas respostas são de grande importância para o entendimento das patogêneses e em consequência, fundamentar medidas a serem adotadas para a terapêutica dessas infecções. Para isso, técnicas de biologia molecular associadas a ferramentas de bioinformática permitem analisar a expressão do RNA mensageiro (mRNA) através de RT-qPCR, técnicas de *arrays* e transcriptomas (TRAPNELL, et al 2010) (GU, et al 2017) (MOSER, et al 2018).

Diante disso, este trabalho busca demonstrar através da técnica de transcriptoma o perfil de expressão dos mRNAs dos testículos durante a infecção pelo Vírus Zika, na tentativa de elucidar a influencia da infecção sobre os processos gênicos do órgão, visto que foram descritos caso de lesões, atrofia e até infertilidade (MOSER, et al 2018).

2 | TRANSCRIPTOMA

Originalmente o termo transcriptoma foi mencionada na década de 1990, onde foram feitas as primeiras tentativas de construção e análise do mesmo, em 1991 foi publicado a primeira captura de transcriptoma com um total de 609 sequências de mRNA de uma amostra de cérebro humano, hoje contamos com plataformas que chegam a gerar cerca de 20 Terabytes (TB) de dados de sequenciamento (TRAPNELL, et al 2010). Isso permite que sejam desenvolvidos projetos mais complexos, abordando desde vias metabólicas celulares a influência de fatores abióticos sobre a expressão de genes, bem elucidado para pesquisas relacionadas a melhoramento na qualidade de plantações. Em humanos a aplicação dos estudos de transcriptomas, estão sendo voltadas para o diagnóstico de doenças, como o câncer e para o entendimento da resposta do sistema imunológico contra agentes infecciosos como bactérias, fungos, vírus e outros parasitas (TRAPNELL, et al 2010).

A primeira técnica de transcriptoma desenvolvida foi a de SAGE (análise seriada de expressão genica) que posteriormente foi melhorada para os microarrays, técnica baseada em mensurar a abundância de transcritos hibridizados para sondas específicas; e RNA-seq mensura a abundância através de bibliotecas de cDNA podendo permitir uma cobertura suficiente para transcritos precisos para humanos que é possível com a tecnologia de sequenciamento Solexa/Illumina (San Diego, EUA).

Para se obter bons resultados de RNA-seq, técnica utilizada nesse trabalho, as amostras precisam passar por algumas etapas, são elas: extração; enriquecimento e controle de qualidade. A extração pode ser feita por diversas metodologias, as mais utilizadas são extração por coluna e beads magnéticas. Durante o processamento dessas amostras é recomendado para melhorar a qualidade da extração, utilizar técnicas para eliminação de interferentes como proteinase K que bloqueia a ação de RNase, diminuindo a perda de material, além da utilização de DNase para eliminação de DNA. Quanto ao enriquecimento das amostras deve ser seguido o protocolo do fabricante, respeitando as etapas a serem seguidas de sucessivos testes de controle de qualidade, como a verificação de integridade do material, formação de dímeros, quantidade de adaptadores, fatores esses, que afetam na qualidade do sequenciamento e na análise de bioinformática.

3 | METODOLOGIA

3.1 Extração do RNA

As amostras que foram utilizadas são oriundas de um projeto maior intitulado “infecção experimental do vírus zika em hamsters dourados (*Mesocricetus auratus*)” do grupo de transmissão sexual, previamente aprovado pela comissão de ética no uso de animais (CEUA) do Instituto Evandro Chagas-IEC com o número de registro nº

24/2016.

Os hamsters macho, foram inoculados via peritoneal com a cepa do vírus zika (H818308) cepa Asiática, do estoque da seção de Arbovirologia e Febres Hemorrágicas do Instituto Evandro Chagas. Para este trabalho foram observados 72hpi a 240hpi, visto que a 72hpi há o pico de carga viral registrado, e aos 240hpi foram observadas lesões testiculares nas replicatas submetidas a necropsias. Essas amostras posteriormente foram maceradas para o preparo para extração pelo método semi automatizada com captura por beads magnéticas com o kit tissue simply RNA (Promega) e processadas na plataforma maxwell 16lev (Promega).

3.2 Construção da biblioteca e sequenciamento

Os mRNAs extraídos foram submetidos para teste quantitativos e qualitativos. No aparelho Qubit 2.0 (Thermo Fisher, EUA) foi analisado a quantidade de RNA por ng/ μ L, depois foi feita a análise da qualidade do fragmento na plataforma bioanalyzer 2100 (Agilent Technology, EUA) que indica o valor de integridade do RNA (RIN) sendo necessário para próxima etapa o valor igual ou acima de 8 seguindo os pré-requisitos do kit de preparo da biblioteca genômica. O preparo da biblioteca genômica foi feito utilizando o TruSeq Small RNA Library Preparation Kits (Illumina, EUA), seguindo o protocolo descrito pelo fabricante e após o preparo da amostra a biblioteca foi sequenciada na plataforma Nextseq 550 (Illumina, EUA).

3.3 Filtragem no Trimmomatic

O Trimmomatic é uma ferramenta de pré-processamento necessária para análise de sequenciamento, ela tem como função eliminar sequência que por ventura podem interferir na interpretação de dados. A origem dessas sequencias podem ser de adaptadores utilizados na etapa de pcr na construção da biblioteca, que durante o sequenciamento se anelaram as layer, podendo assim comprometer a qualidade das amostras (BOLGER, et al 2014).

A leitura de sequenciamento deste trabalho foi feita por amostras pareadas gerando dois arquivos R1 (primeiro par) e R2 (segundo par).

Os arquivos de saída deste programa geraram dois arquivos correspondentes para cada par das amostras pareadas, essas amostras foram submetidas a um teste de qualidade feita na ferramenta FASTQC (BOLGER, et al 2014).

3.4 Alinhamento na ferramenta HISAT

Após a filtragem, os arquivos foram alinhados utilizando a ferramenta HISAT (Indexação Hierárquica para Alinhamento de Transcrições), que alinha os dados de forma hierárquica seguindo esquema de indexação baseado na transformação de Burrows-Wheeler e no índice Ferragina-Manzini (FM) (KIM, et al 2015).

3.5 Reconstrução do Transcriptoma e Anotações

A reconstrução do transcriptoma e anotações foram feitas nas seguintes plataformas: Stringtie, foi utilizado para reconstruir os transcritos, calcula o número de transcritos e o seu respectivo tamanho. Os transcritos reconstruídos posteriormente foram concatenados com a ferramenta Stringmerge (PERTEA, et al 2015); A ferramenta *Featurescounts*, foi usada para a contagem de leituras encontradas no alinhamento, com objetivo de otimizar as análises estatísticas segundo descrito por LIAO, et al 2014.

3.6 Análise Diferencial

Foi utilizado o método Deseq2 para análise diferencial de dados de contagem geradas no *Featurescounts*, com o objetivo de estimar a retração para dispersões e mudanças de dobras (LOVE, et al 2014).

4 | RESULTADO E DISCUSSÃO

Ao observarmos os dados gerados neste trabalho foi possível detectar que o número de leituras totais limpas não possui uma grande diferença entre as amostras. Sendo que mais de 96% das leituras mapeadas para o genoma hospedeiro com cerca de 80% para regiões gênicas e 20% para áreas intergênicas do genoma do hospedeiro.

A análise e comparação dos perfis de expressão de mRNA em diferentes momentos após a infecção VZIK revelou que, no total, 1332 genes tiveram alterações de 2 ou mais vezes em qualquer direção (Figura 01).

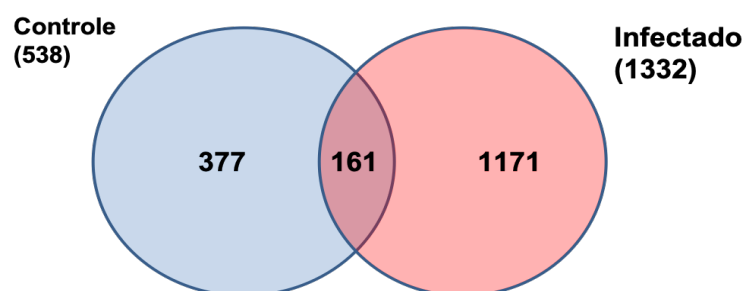


Figura 1 - Diagrama de Venn representando o número de genes codificadores expressos diferencialmente entre o período anterior e posterior à infecção. Sendo que o número de genes expressos é significativamente maior em relação ao controle não infectado.

Estudos recentes como os vem demonstrando que durante a infecção por Zika o número de mRNAs expressos e de duas a três vezes maior que das amostras controles. Este fenômeno vem sendo descrito como necessário para uma regulação da replicação viral, isto é, facilitando a replicação viral, uma vez que mais genes acabam por ser regulados positiva ou negativamente (CULSHAW; MONGKOLSAPAYA; SCREATON, 2018; SINGH et al., 2018; STASSEN et al., 2018).

No entanto ao compararmos com outros vírus do mesmo gênero, como Dengue

e Febre amarela, é importante ressaltar que autores vem demonstrando uma maior expressão diferenciada de mRNA para codificações de proteínas chaves para a defesa celular durante a infecção por Flavivírus (CASSEB et al., 2016; DOMINGO et al., 2012; SILVA et al., 2011). Desta maneira a relação dos mRNAs com a infecções por Zika tende a se comportar de forma semelhante à de outros vírus com genomas relacionado a eles, no entanto suas funções durante esse processo ainda não estão completamente esclarecida.

É importante ressaltar também que os tecidos testiculares não são órgãos muito comuns para a infecção por Flavivirus, no entanto o Zika parece possuir uma maior afinidade as células que compõem este tecido, tanto que STASSEN, et al 2018 já haviam descrito este fenômeno em seus estudos. Além disso, outros estudos já vem descrevendo a transmissão sexual deste vírus (JAMIL; WAHEED; DURRANI, 2016; OSTER et al., 2016).

Além destes resultados foi possível também estimar a dispersão da expressão entre 3 dias após a infecção por zika (dpi) para o 10 dpi (Figura 2).

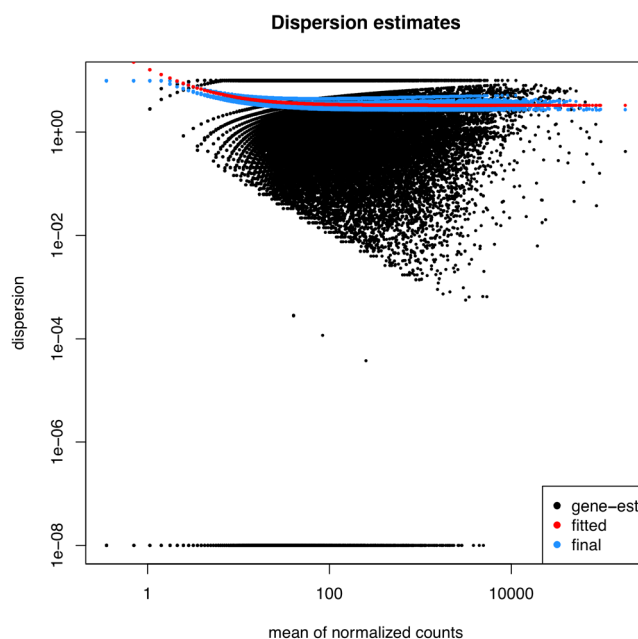


Figura 2: Gráfico de dispersão entre o 10 dpi

Surpreendentemente o Zika vem demonstrando uma grande afinidade pelo tecido testicular e esta regulação gênica vem sendo observada em vários estudos por todo o mundo (DASTI, 2016; LEE et al., 2018; WIKAN; SMITH, 2016). Em nosso estudo foi possível uma oscilação de expressão entre dias após infecção, desta maneira colaborando com a possibilidade deste vírus regular o processo celular e assim mantendo sua replicação mais eficiente por um maior tempo dentro das células testiculares (Figura 03).

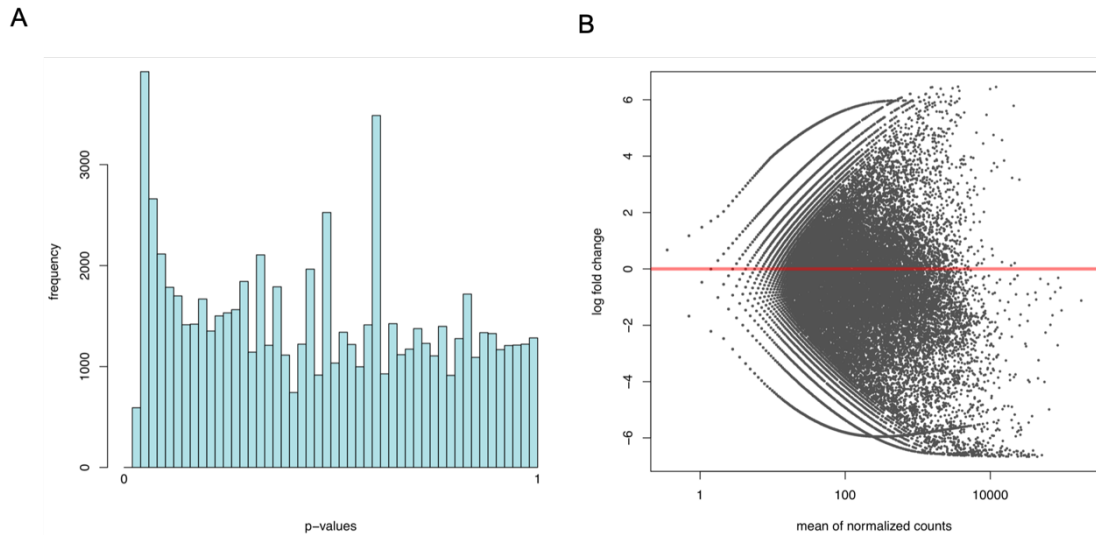


Figura 3 - (A) Histograma de frequência de mRNAs entre 3 e 10 dpi, onde é demonstrado a frequência e o valor de “p”. (B) Gráfico de dispersão dos mRNAs entre o 3 e 10 dpi, a linha vermelha demonstra os genes que não sofreram variação de expressão entre as duas amostras

Em geral, nossos resultados mostraram grandes mudanças no transcriptoma anaeróbico após a infecção por VZIK, tanto na codificação quanto nos RNAs não-codificantes. No entanto podemos afirmar que a expressão do mRNAs ao longo dos dias após infecção demonstra que o Zika consegue regular a célula para manter sua eficiência de replicação por um longo tempo neste tipo celular, desta maneira levando a uma possível relação da infecção viral com mudanças nas gonadas masculinas, portanto sendo possível causar esterilidade em hamsters dourados (*Mesocricetus auratus*).

REFERÊNCIAS

- BOLGER, Anthony M; LOHSE, Marc; USADEL, Bjoern. Trimmomatic: a flexible trimmer for Illumina sequence data. **Bioinformatics**, v. 30, n. 15, p. 2114–2120, 2014. Disponível em: <<http://dx.doi.org/10.1093/bioinformatics/btu170>>.
- CARLIN, Aaron F; VIZCARRA, Edward A; BRANCHE, Emilie; *et al.* Deconvolution of pro- and antiviral genomic responses in Zika virus-infected and bystander macrophages. **Proceedings of the National Academy of Sciences of the United States of America**, v. 115, n. 39, p. E9172–E9181, 2018. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/30206152>>.
- DICK, G W A; KITCHEN, S F; HADDOW, A J. Zika Virus (I). Isolations and serological specificity. **Transactions of The Royal Society of Tropical Medicine and Hygiene**, v. 46, n. 5, p. 509–520, 1952. Disponível em: <[http://dx.doi.org/10.1016/0035-9203\(52\)90042-4](http://dx.doi.org/10.1016/0035-9203(52)90042-4)>.
- GU, Se Hun; SONG, Dong Hyun; LEE, Daesang; *et al.* Whole-genome sequence analysis of Zika virus, amplified from urine of traveler from the Philippines. **Virus genes**, v. 53, n. 6, p. 918–921, 2017. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/28795266>>.
- KIM, Daehwan; LANGMEAD, Ben; SALZBERG, Steven L. HISAT: a fast spliced aligner with low memory requirements. **Nature Methods**, v. 12, p. 357, 2015. Disponível em: <<http://dx.doi.org/10.1038/nmeth.3317>>.

- LIAO, Yang; SMYTH, Gordon K; SHI, Wei. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. **Bioinformatics**, v. 30, n. 7, p. 923–930, 2014. Disponível em: <<http://dx.doi.org/10.1093/bioinformatics/btt656>>.
- LOPEZ-DENMAN, Adam J; MACKENZIE, Jason M. The importance of the Nucleus during Flavivirus Replication. **Viruses**, v. 9, n. 1, p. 14, 2017. Disponível em: <<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5294983/>>.
- LOVE, Michael I; HUBER, Wolfgang; ANDERS, Simon. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. **Genome Biology**, v. 15, n. 12, p. 550, 2014. Disponível em: <<https://doi.org/10.1186/s13059-014-0550-8>>.
- MINER, Jonathan J; DIAMOND, Michael S. Zika Virus Pathogenesis and Tissue Tropism. **Cell Host & Microbe**, v. 21, n. 2, p. 134–142, 2017. Disponível em: <<http://dx.doi.org/10.1016/j.chom.2017.01.004>>.
- MOSEER, Lindsey A; OLDFIELD, Lauren M; FEDOROVA, Nadia; *et al.* Whole-Genome Sequences of Zika Virus FLR Strains after Passage in Vero or C6/36 Cells. **Genome announcements**, v. 6, n. 4, 2018. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/29371358>>.
- OLIVEIRA, Consuelo Silva de; VASCONCELOS, Pedro Fernando da Costa. Microcephaly and Zika virus. **Jornal de Pediatria**, v. 92, p. 103–105, 2016.
- PERTEA, Mihaela; PERTEA, Geo M; ANTONESCU, Corina M; *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. **Nature Biotechnology**, v. 33, p. 290, 2015. Disponível em: <<http://dx.doi.org/10.1038/nbt.3122>>.
- SIROHI, Devika; CHEN, Zhenguang; SUN, Lei; *et al.* The 3.8 Å resolution cryo-EM structure of Zika virus. **Science**, v. 352, n. 6284, p. 467 LP-470, 2016. Disponível em: <<http://science.sciencemag.org/content/352/6284/467.abstract>>.
- TRAPNELL, Cole; WILLIAMS, Brian A; PERTEA, Geo; *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. **Nature Biotechnology**, v. 28, p. 511, 2010. Disponível em: <<http://dx.doi.org/10.1038/nbt.1621>>.
- CASSEB, S. M. M. *et al.* down-regulated in human cells infected with dengue virus 4, and play a role in viral pathogenesis. **Genetic and Molecular Research**, v. 15, n. 2, 2016.
- CULSHAW, A.; MONGKOLSAPAYA, J.; SCREATON, G. The immunology of Zika Virus. **F1000Research**, v. 7, n. 0, p. 203, 2018.
- DASTI, J. I. Zika virus infections: an overview of current scenario. **Asian Pacific Journal of Tropical Medicine**, v. 9, n. 7, p. 621–625, 2016.
- DOMINGO, C. *et al.* Advanced yellow fever virus genome detection in point-of-care facilities and reference laboratories. **Journal of clinical microbiology**, v. 50, n. 12, p. 4054–60, dez. 2012.
- JAMIL, Z.; WAHEED, Y.; DURRANI, T. Z. Zika virus, a pathway to new challenges. **Asian Pacific Journal of Tropical Medicine**, v. 9, n. 7, p. 626–629, 2016.
- LEE, I. *et al.* Probing Molecular Insights into Zika Virus-Host Interactions. **Viruses**, v. 10, n. 5, p. 233, 2 maio 2018.
- OSTER, A. M. *et al.* Update: Interim Guidance for Prevention of Sexual Transmission of Zika Virus - United States, 2016. **MMWR. Morbidity and mortality weekly report**, v. 65, n. 12, p. 323–5, 2016.

PINTO JUNIOR, V. L.; LUZ, K.; PARREIRA, R. P. Vírus Zika : Revisão para Clínicos Zika Virus : A Review to Clinicians. **Acta Med Port.**, v. 28, n. 6, p. 760–765, 2015.

SILVA, B. M. et al. The dengue virus nonstructural protein 1 (NS1) increases NF- κ B transcriptional activity in HepG2 cells. **Archives of Virology**, v. 156, n. 7, p. 1275–1279, 2011.

SINGH, P. K. et al. Determination of system level alterations in host transcriptome due to Zika virus (ZIKV) Infection in retinal pigment epithelium. **Scientific Reports**, v. 8, n. 1, p. 11209, 25 dez. 2018.

STASSEN, L. et al. Zika Virus in the Male Reproductive Tract. **Viruses**, v. 10, n. 4, p. 198, 2018.

WIKAN, N.; SMITH, D. R. Zika virus: history of a newly emerging arbovirus. **The Lancet Infectious Diseases**, v. 16, n. 7, p. e119–e126, 2016.

CARACTERIZAÇÃO FILOGENÉTICA DA FAMÍLIA MULTIGÊNICA DA H⁺-ATPASE DE MEMBRANA PLASMÁTICA EM MONOCOTILEDÔNEAS DA ORDEM POALES

Lyndefânia Melo de Sousa

Universidade Federal do Ceará, Departamento de Bioquímica e Biologia Molecular, Laboratório de Genômica Funcional e Bioinformática
Fortaleza - Ceará

Clesivan Pereira dos Santos

Universidade Federal do Ceará, Departamento de Bioquímica e Biologia Molecular, Laboratório de Genômica Funcional e Bioinformática
Fortaleza - Ceará

Thais Andrade Germano

Universidade Federal do Ceará, Departamento de Bioquímica e Biologia Molecular, Laboratório de Genômica Funcional e Bioinformática
Fortaleza - Ceará

Moacília de Souza Lemos

Universidade Federal do Ceará, Departamento de Bioquímica e Biologia Molecular, Laboratório de Genômica Funcional e Bioinformática
Fortaleza - Ceará

Stelamaris de Oliveira Paula

Universidade Federal do Ceará, Departamento de Bioquímica e Biologia Molecular, Laboratório de Genômica Funcional e Bioinformática
Fortaleza - Ceará

Rafael de Souza Miranda

Universidade Federal do Piauí, Departamento de Biologia.
Bom Jesus – Piauí.

José Helio Costa

Universidade Federal do Ceará, Departamento de Bioquímica e Biologia Molecular, Laboratório de

Genômica Funcional e Bioinformática
Fortaleza - Ceará

RESUMO: As H⁺-ATPases são proteínas integrais da membrana plasmática que têm a capacidade de utilizar a energia química da hidrólise de ATP para expulsar os prótons para o ambiente extracelular, atuando na manutenção da homeostase iônica e transporte de solutos. O presente trabalho objetivou identificar, caracterizar e analisar a distribuição filogenética da família multigênica H⁺-ATPase de membrana plasmática em monocotiledôneas da ordem Poales. Para este fim, os membros gênicos foram identificados e anotados em espécies *da ordem Poales* através de buscas em bancos de genomas disponíveis no NCBI e *Phytozome*, utilizando a ferramenta BLAST. As sequências foram alinhadas pelo algoritmo MUSCLE e agrupadas pelo método *Neighbor joining*, a partir de 1000 replicatas, através do programa MEGA 7.0. Os resultados revelaram que as sequências continham entre 2 e 21 éxons formando janelas de leitura aberta (ORFs) apresentando entre 2.571 a 3.327 pb e proteínas contendo entre 857 a 976 aminoácidos. A análise filogenética agrupou as sequências em quatro clados distintos identificados como clado I-IV. O clado I, por apresentar apenas uma sequência representativa de cada espécie foi nomeada de

H⁺-ATPase1. Os clados II, III e IV foram nomeadas de H⁺-ATPase2 (2a, 2a1 ou 2a2, 2b ou 2b1 ou 2b2 e 2c), H⁺-ATPase3 (3a ou 3a1 ou 3a2, 3b, ou 3b1 ou 3b2, ou 3c) e H⁺-ATPase4 (4a ou 4a1 ou 4a2 ou 4b) respectivamente, por apresentarem duas ou mais sequências da mesma espécie. Esses resultados fornecem suporte para estudos que visem identificar a função específica de cada membro gênico entre espécies.

PALAVRAS-CHAVES: Anotação gênica, caracterização gênica, distribuição filogenética.

ABSTRACT: The H⁺-ATPases are integral proteins of the plasma membrane that have the capacity to use the chemical energy of the hydrolysis of ATP to expel the protons to the extracellular environment, acting in the maintenance of ionic homeostasis and transport of solutes. The present work aimed to identify, characterize and analyze the phylogenetic distribution of the multigenic H⁺-ATPase family of plasma membrane in monocotyledons of the order Poales. For this purpose, gene members were identified and annotated in species of the order Poales through searches in genomes database available in the NCBI and Phytozome, using the tool BLAST. The sequences were aligned by the MUSCLE algorithm and grouped by the Neighbor joining method, from 1000 replicates, through the MEGA 7.0 program. The results revealed that the sequences contained between 2 and 21 exons forming open reading frames (ORFs) ranging from 2.571 to 3.327 bp and proteins containing between 857 and 976 amino acids. Phylogenetic analysis grouped the sequences into four distinct clades identified as clade I-IV. Clade I, because it presented only one representative sequence of each species, was named H⁺-ATPase 1. Clades II, III and IV were named as H⁺-ATPase2 (2a, 2a1 ou 2a2, 2b ou 2b1 ou 2b2 e 2c), H⁺-ATPase3 (3a ou 3a1 ou 3a2, 3b, ou 3b1 ou 3b2, ou 3c) e H⁺-ATPase4 (4a ou 4a1 ou 4a2 ou 4b), respectively, as they present two or more sequences of the same species. These results provide support for studies aimed at identifying the specific function of each gene member between species.

KEYWORDS: Gene annotation, gene characterization, phylogenetic distribution

1 | INTRODUÇÃO

As H⁺-ATPases de membrana plasmática são proteínas integrais que atuam no bombeamento de prótons do citoplasma para o ambiente extracelular utilizando a energia química da hidrólise do ATP gerando força próton motriz que é fundamental para a atividade de um grande grupo de transportadores secundários que movem íons e metabólitos contra gradiente de concentração (PALMGREN, 2001; WADOWIKOWSKA; KLOBUS, 2016). Em plantas, essas bombas de prótons estão relacionadas a diferentes processos fisiológicos tais como, a absorção de nutrientes, a regulação do pH intracelular, a abertura estomática, o crescimento celular e a manutenção da homeostase iônica (PALMGREN, 2001; DUBY; BOUTRY, 2009; TODA *et al.*, 2016; MORSOMME; BOUTRY, 2000). Além disso, as H⁺-ATPases estão associadas a mecanismos de aclimação das plantas a condições de estresse. Estudos conduzidos nas espécies

Sorghum bicolor, *Cucumis sativus* e *Aeluropus littoralis* demonstraram um aumento no perfil de expressão gênica dessas H⁺-ATPases quando submetidas a condições de estresse salino e hídrico (MIRANDA *et al.*, 2017; WADOWIKOWSKA; KLOBUS, 2016; OLFATMIRI *et al.*, 2014). Embora essas proteínas apresentem papéis fisiológicos importantes, estudos sobre a sua caracterização gênica e distribuição filogenética ainda são escassos, limitando-se a poucas espécies. Dentre elas destacam-se as eudicotiledôneas *Arabidopsis thaliana* (GAXIOLA *et al.*, 2007; ARANGO *et al.*, 2003; HARPET *et al.*, 1990) e *Cucumis sativus* (WADOWIKOWSKA; KLOBUS, 2016) bem como as monocotiledôneas *Oryza sativa* (BAXTER, 2003) e *Zea mays* (SANTI *et al.*, 2003). Entretanto, nesses estudos não há conexão de nomenclatura e classificação filogenética entre os genes das diferentes espécies. Dessa forma, o presente trabalho objetivou identificar, caracterizar e analisar a distribuição filogenética da família multigênica das H⁺-ATPases de membrana plasmática em monocotiledôneas da ordem Poales a fim de estabelecer uma classificação filogenética dos membros gênicos dentro dessa ordem.

2 | MATERIAL E MÉTODOS

2.1 Identificação e caracterização dos genes da família multigênica H⁺-ATPase de membrana plasmática.

A identificação dos genes da família multigênica H⁺-ATPase de membrana plasmática foi realizada através da ferramenta BLAST (*Basic Local Alignment Search Tool*) disponível no GenBank do *National Center for Biotechnology Information* (NCBI) (ALTSCHUL *et al.*, 1997). A sequência proteica da H⁺-ATPase de membrana plasmática do arroz (ID: XP_015620234.1) foi utilizada como referência para a busca de genes homólogos nos genomas das espécies *Aegilops tauschii*; *Brachypodium distachyon*; *Oryza sativa*; *Panicum hallii*; *Leersia perrieri*; *Setaria viridis*; *Zea mays*; *Zoysia japonica* e *Ananas comosus* todas pertencentes a ordem Poales e depositados nos bancos de dados *Whole Genome Shotgun* (WGS), *Refseq representative genomes* do GenBank (<https://www.ncbi.nlm.nih.gov/>) e *Phytozome* (<https://phytozome.jgi.doe.gov/pz/portal.html>). Os genes da alga *Chlamydomonas reinhardtii* foram também identificados para serem utilizados como grupo externo. Deve-se ressaltar que das sequências homólogas encontradas apenas as que tiveram identidade acima de 45% foram selecionadas. No processo de identificação, estes genes foram manualmente anotados para a dedução dos cDNAs utilizando como referência as seqüências de mRNA das H⁺-ATPase de membrana plasmática depositadas nos bancos de dados RNA, tais como, *Refseq_RNA* (*Reference RNAsequence*), *EST* (*Expressed Sequence Tags*) e *TSA* (*Transcriptome Shotgun Assembly*). Em seguida, os cDNAs obtidos foram traduzidos em seqüências de aminoácidos com o auxílio da ferramenta de tradução disponível

no servidor da web EXPASY (web.expasy.org/translate). As proteínas deduzidas foram validadas através da comparação com sequências homólogas, depositadas no banco de proteínas (*Non-redundant*). O processo de anotação foi realizado de acordo com o método proposto por Costa *et al.* (2014), o que também possibilitou a identificação do número de éxons, íntrons e janela de leitura aberta (ORFs).

2.2 Alinhamento de sequências e análise filogenética

O alinhamento das proteínas deduzidas, objetivando a análise filogenética, foi realizado através do algoritmo MUSCLE usando o método de agrupamento UPGMA (*Unweighted Pair Group Method using Arithmetic averages*). A análise filogenética foi construída através do método *Neighbor joining*, (SAITOU & NEI, 1987) utilizando como modelo de substituição o *Poisson model*. Além disso, foram usados valores de bootstraps gerados a partir de 1000 replicatas, através do programa MEGA 7.0. (KUMAR *et al.*, 2016).

3 | RESULTADOS E DISCUSSÃO

3.1 Identificação e anotação dos genes da H⁺-ATPase de membrana plasmática na ordem Poales.

No presente estudo foram identificadas e anotadas um total de 103 sequências gênicas que codificam bombas de prótons H⁺-ATPase de membrana plasmática da ordem Poales. O número de genes identificados variou de 9 genes nas espécies *Aegilops tauschii* e *Brachypodium distachyon* a 12 genes em *Zea mays* (ver figura 1). A análise dos genomas de *Arabidopsis thaliana* e *Oriza sativa* revelou a presença de pelo menos onze e dez genes, respectivamente, que codificam as proteínas H⁺-ATPase de membrana plasmática (ARANGO *et al.*, 2003; BAXTER *et al.*, 2003), o que corrobora com a quantidade de genes encontrados no presente trabalho. Entretanto, com relação a *Zea mays*, apenas 4 genes haviam sido relatados anteriormente (SANTI *et al.*, 2003), portanto, nosso estudo revela a presença de outros 8 genes no genoma dessa espécie.

A caracterização gênica mostrou que as sequências continham entre 2 e 21 éxons formando janelas de leitura aberta (ORFs) apresentando entre 2.571 a 3.327 pares de base (pb) e proteínas contendo entre 857 a 976 aminoácidos. Resultados semelhantes foram anteriormente demonstrados nas espécies *Oryza Sativa*, *Cucumis sativus* e *Solanum lycopersicum* (ARANGO *et. al.*, 2003; WADOWIKOWSKA; KLOBUS, 2016; LIU *et al.*, 2016). A análise das proteínas deduzidas revelou que o ponto isoelétrico teórico variou de 5.62 a 7.05, enquanto que a massa molecular apresentou valores entre 93,56 a 106,56 kDa reforçando resultados descritos em estudos anteriores (PALMGREEN,2001).

Caracterização gênica								
Família	Tribo	Espécie	Quantidade de genes	Éxons	ORFs (pb)	Número de aminoácidos	Ponto Isoelétrico (PI)	Peso Molecular (kDa)
Poaceae	Oryzae	<i>Oryza sativa</i>	10	03 a 21	2.5.80 a 2.874	859 a 957	5.63 a 6.43	93,76 a 105,53
		<i>Leersia Perriere</i>	10	04 a 21	2.571 a 2.904	859 a 967	5.80 a 6.42	93,76 a 105,86
	Triticeae	<i>Aegilops tauschii</i>	9	03 a 21	2.955 a 3.534	945 a 976	5.62 a 6.5	103,37 a 106,52
	Brachypodieae	<i>Brachypodium distachyon</i>	9	03 a 21	2.841 a 3.327	946 a 976	5.67 a 6.19	103,49 a 106,32
	Andropogoneae	<i>Sorghum bicolor</i>	11	02 a 21	2.574 a 3.195	857 a 974	5.81 a 6.64	93,56 a 106,41
		<i>Zea mays</i>	12	02 a 21	2.574 a 3.597	857 a 972	5.65 a 6.49	93,76 a 106,34
	Paniceae	<i>Setaria viridis</i>	10	02 a 21	2.636 a 3089	871 a 968	5.72 a 6.54	95,02 a 105,74
		<i>Panicum halli</i>	10	03 a 21	2.574 a 2.925	857 a 974	5.68 a 6.46	93,79 a 106,26
	Zoysieae	<i>Zoysia japonica</i>	11	03 a 21	2.853 a 2.874	950 a 956	5.79 a 6.87	103,91 a 105,22
	Bromeleaceae		<i>Ananas comosus</i>	10	12 a 21	2.850 a 3.105	949 a 971	5.86 a 7,05

Figura 1 - Caracterização dos genes da H⁺ -ATPase em espécies da ordem Poales.

3.2 Análise filogenética

A análise filogenética revelou que as sequências eram distribuídas em quatro clados distintos identificados como clado I-IV. No clado I, as sequências foram nomeadas de H⁺ -ATPase1 unicamente, tendo em vista a existência de apenas uma sequência representativa em cada espécie. Já nos clados II, III e IV as sequências foram nomeadas de H⁺ -ATPase2 (2a, 2a1 ou 2a2, 2b ou 2b1 ou 2b2 e 2c), H⁺ -ATPase3 (3a ou 3a1 ou 3a2, 3b, ou 3b1 ou 3b2, ou 3c) e H⁺ -ATPase4 (4a ou 4a1 ou 4a2 ou 4b), respectivamente, por apresentarem duas ou mais sequências representativas na mesma espécie. As espécies *Zea mays*, *Sorghum bicolor* e *Panicum halli* apresentaram proteínas em todos os clados, exceto clado I, indicando que os genes da H⁺ -ATPase1 foram deletados na subfamília Panicoideae. Ademais, o maior número de genes nos clados II, III e IV sugere possíveis processos de duplicação. Nesse sentido, a espécie *Ananas comosus* apresentou genes duplicados em todos os clados citados anteriormente.

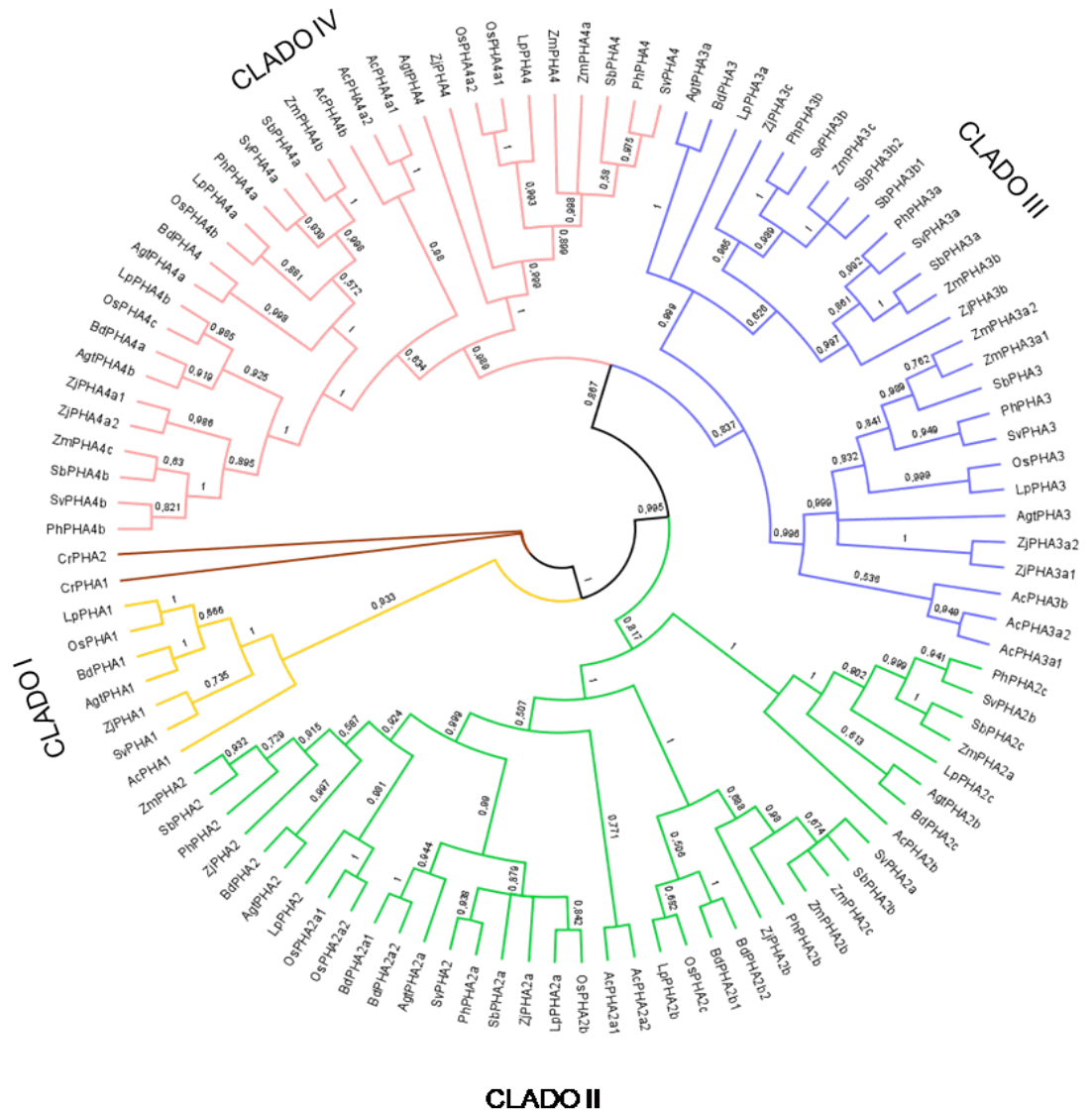


Figura 2. Análise filogenética da H⁺ -ATPase. As sequências proteicas das H⁺ -ATPase foram agrupadas em quatro clados distintos nomeados de I a IV. As cores, amarelo, verde, azul e rosa foram utilizadas no processo de identificação dos clados I, II, III e IV respectivamente. A cor marrom foi utilizada para identificar a alga *Chlamydomonas reinhardtii* utilizada como grupo EXTERNO.

4 | CONCLUSÃO

A análise da caracterização gênica e da distribuição filogenética dos membros da família multigênica H⁺ -ATPase na ordem Poales fornece uma classificação confiável dentro do grupo das monocotiledôneas, possibilitando a identificação de genes ortólogos. Esta classificação dará suporte a estudos que visem identificar a função específica de cada membro gênico entre espécies.

REFERENCIAS

- 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 research**, v. 25, p. 3389-3402, 1997.
- ARANGO, M.; GEVAUDANT, F.; OUFATTOLE, M.; BOUTRY, M. The plasma membrane proton pump ATPase: the significance of gene subfamilies. **Planta**, v. 216, p. 355–365, 2003.
- BAXTER, I.; TCHIEU, J.; SUSSMAN, M.; BOUTRY, M.; PALMGREN, M.; GRIBSKOV, M.; HARPER, J.; AXELSEN, K. Genomic comparison of P-type ATPase ion pumps in Arabidopsis and rice. **Plant Physiology**, v. 132, p. 618-828, 2003.
- COSTA, J. H.; MCDONALD, A. E.; ARNHOLDT-SCHMITT, B.; MELO, D. F. A classification scheme for alternative oxidases reveals the taxonomic distribution and evolutionary history of the enzyme in angiosperms. **Mitochondrion**, v. 19, p. 172-183, 2014.
- DUBY, G.; BOUTRY, M. The plant plasma membrane proton pump ATPase: a highly regulated P-type ATPase with multiple physiological roles. **Pflügers Archiv**, v. 457, p. 645-55, 2009.
- GAXIOLA, R. A.; PALMGREN, M. G.; SCHUMACHER, K. Plant proton pumps. **FEBS Letters**, v. 581, p. 2204-2214, 2007.
- HARPER, J. F.; MANNEY, L.; DEWITT, N. D.; YOO, M. H.; SUSSMAN, M. R. The Arabidopsis thaliana plasma membrane H⁺-ATPase multigene family. Genomic sequence and expression of a third isoform. **Journal of Biological Chemistry**, v. 265, p. 13601-13608, 1990.
- KUMAR, S.; STECHER, G.; TAMURA, K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. **Molecular biology and evolution**, v. 33, p. 1870-1874, 2016.
- LIU, J.; LIU, J.; CHEN, A.; JI, M.; CHEN, J.; YANG, X.; GU, M.; QU, H.; XU, G. Analysis of tomato plasma membrane H⁺-ATPase gene family suggests a mycorrhiza-mediated regulatory mechanism conserved in diverse plant species. **Mycorrhiza**, v. 26, p. 645-656, 2016.
- MIRANDA, R.S.; MESQUITA, R.O.; COSTA, J.H.; ALVAREZ-PIZARRO, J.C.; PRISCO, J.T.; GOMES-FILHO, E. Integrative Control Between Proton Pumps and SOS1 Antiporters in Roots is Crucial for Maintaining Low Na⁺ Accumulation and Salt Tolerance in Ammonium-Supplied Sorghum bicolor. **Plant and Cell Physiology**, v. 58, p. 522-536, 2017.
- MORSOMME, P.; BOUTRY, M. The plant plasma membrane H⁺-ATPase: structure, function and regulation. **Biochimica et Biophysica Acta**, v. 1465, p. 1-16, 2000.
- OLFATMIRI, H.; ALEMZADEH, A.; ZAKIPOUR, Z. Up-regulation of plasma membrane H⁺-ATPase under salt stress may enable *Aeluropus litoralis* to cope with stress. **Molecular Biology Research Communications**, v. 3, p. 67-75, 2014.
- PALMGREN, M. Plant plasma membrane H⁺-ATPase: powerhouse for nutrient uptake. **Annual review of plant physiology and plant molecular biology**, v. 52, p. 817–845, 2001.
- SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Molecular biology and evolution**, v. 4, p. 406-425, 1987.
- SANTI, S.; LOCCI, G.; MONTE, R.; PINTON, R.; VARANINI, Z. Induction of nitrate uptake in maize roots: expression of a putative high affinity nitrate transporter and plasma membrane H⁺-ATPase isoforms. **Journal of Experimental Botany**, v. 54, p.1851-1864, 2003.

TODA, Y.; WANG, Y.; TAKAHASHI, A.; KAWAI, Y.; TADA, Y.; YAMAJI, N.; FENG, M.A.J.; ASHIKARI, M.; KINOSHITA, T. *Oryza sativa* H⁺-ATPase (OSA) is involved in the regulation of dumbbell-shaped guard cells of rice. **Plant and Cell Physiology**, v. 57, p. 1220-1230, 2016.

WDOWIKOWSKA, A.; KLOBUS, G. The plasma membrane proton pump gene family in cucumber. **Acta physiologiae plantarum**, v. 38, p. 135-145, 2016.

CLADISTIC ANALYSIS IN GEMINIVIRIDAE: AN EVIDENCE OF MULTISPECIFICITY FOR CULTIVARS HOSTS

Rafael Trindade Maia

Universidade Federal de Campina Grande, Centro de Desenvolvimento do Semiárido Sumé, Paraíba. E-mail: rafael.rafatrin@gmail.com

Corresponding author

Aparecida Yasmim Silva de Azevedo

Universidade Federal de Campina Grande, Centro de Desenvolvimento do Semiárido. Sumé, Paraíba.

Maria Bartira Chaves de Souza Silva

Universidade Federal de Campina Grande, Centro de Desenvolvimento do Semiárido. Sumé, Paraíba.

Ana Verônica Silva do Nascimento

Universidade Federal de Campina Grande, Centro de Desenvolvimento do Semiárido. Sumé, Paraíba.

ABSTRACT: In recent years the *Geminiviridae* family virus has been intensively studied due the severity of the diseases caused in several cultures of economic importance; like bean, cotton, corn, tomato and cassava. The objective of this work was to do a cladistic inference of *Begomovirus* populations through computational tools. Viral genome sequences were obtained from NCBI (<https://www.ncbi.nlm.nih.gov/>), totaling 297 sequences from different countries. The sequences were aligned with the Bioedit program, using the ClustalW algorithm.

After the alignment, the cladogram was obtained through the Maximum Likelihood method in the MEGA program, with bootstrap analysis of 1000 replicates. ModelTest was used to select the most suitable evolutionary model (Reversible G + I) for the data set. The results showed that the bootstrap values in the clades ranged from 9 to 100. Pepper, potato and watermelon virus sequences were used with outgroup, but they were grouped into clades shared by the viruses obtained in the Tomato strain, revealing an evidence that the mosaic virus in these cultivars belong to the same species and same genus.

KEYWORDS: Maximum Likelihood; Mosaic virus; Tomato.

1 | INTRODUCTION

The *Geminiviridae* family is structurally characterized by the geminated morphology of the viral particle, with 18-30 nm and genetically by having a single-stranded circular single-stranded DNA molecule (ssDNA). Each of the molecules has 2500-3000 nucleotides (nt) encapsidated by a single structural protein that is arranged in the form of 22 capsomers forming two incomplete icosahedrons which confers the geminated appearance of the viral particles, typical characteristic of this family (BROWN et al., 2011).

The *Geminiviridae* family is composed by seven genus: *Begomovirus*, *Mastrevirus*, *Curtovirus*, *Becurtovirus*, *Turncurtovirus*, *Eragrovirus* and *Topocovirus* (2016). In addition, some representatives of the *Begomovirus* genus are associated with DNA satellites called alpha and beta-satellite. *Begomoviruses* (*Geminiviridae* family) and their associated satellite DNAs, known as alphasatellites and betasatellites, form complexes that cause devastating diseases in agricultural systems (LEKE et al., 2015). Except the *Begomovirus* genus, which includes viruses with a mono or bisegmented genome, all other genus of the family include only viruses with a monosegmented genome (PASSOS, 2016).

The DNA-A has five or six genes, one or two in the viral sense (CP and, in Old World viruses, AV2) and four in the complementary sense (Rep, TrAP, REn and AC4). The role of Rep (Replication-associated protein) in the coupling to the origin of replication for the initiation of DNA synthesis. The TrAP (Trans-Activating Protein) protein is responsible for activating transcription of the CP and NSP genes, as well as acting in the suppression of plant defense responses. REn (Replication-Enhancer protein) enhances replication, increasing the concentration of viral DNA. AC4 protein is related to suppression of gene silencing and, in monosegmented *Begomoviruses*, in viral movement. In some species an AC5 ORF is present, a recent study has shown that the encoded product plays an important role in suppressing plant defense responses. The CP protein is responsible for the viral encapsidation and specificity with the insect vector and, in the monosegmented *Begomoviruses*, it acts on the viral movement. The AV2 gene is present only in the *Begomovirus* of the Old World and its function is related to the viral movement (BRIDDON et al., 2010, PASSOS, 2016, VANITHARANI et al., 2004).

According to Passos (2016), of the three new genera, established by the ICTV, the *Becurtovirus* presents two species: Beet curly top Iran virus (BCTIV) and Spinach curly top Arizona virus (SCTAV). The transmission of BCTIV is by spittlebugs, while the SCTAV has not yet had the vector identified. *Eragrostis curvula* streak virus (ECSV) and Turnip curly top virus (TCTV), respectively, with the first one with no known vector and the second one transmitted by *Eragrovirus* and *Turncurtovirus*, spittlebugs Varsani et al. (2014) reports that a difference observed between the genus *Becurtovirus* and *Eragrovirus* and the other viruses in the family refers to the conserved sequence associated with the viral replication start site. Generally *Geminiviruses* have the TAATATTAC nine-nucleotide, however, viruses of both genus have the TAAGATTCC sequence.

Although the family *Geminiviridae* presents great diversity of genus, one of these has been highlighting in the recent research, for triggering several diseases that affect the tomato, belonging to the genus *Begomovirus*. It is currently the most relevant for the high incidence of viruses, with reports of losses up to 100% in tomato crops (BRINDDON et al., 2015).

Begomovirus Genus

Viruses of the *Begomovirus* genus has mostly of the genome divided into two single-stranded circular DNA components, called DNA A and DNA B, each with approximately 2.6 kbp. DNA-A encodes all the proteins necessary for the transcription, replication and encapsidation of both DNAs, whereas DNA-B contributes with the necessary functions for the movement of the virus in the plant and the development of symptoms (ROJAS et al., 2005). They have an intergenic region of approximately 200 bp, highly conserved between the two components of the same virus, which is essential for the recognition of the replication and transcription processes of the two genomes, thus maintaining the fidelity of the bipartite genome (BROWN et al., 2011 FAUQUET et al., 2005, JESKE, 2009).

The DNA-A has five or six genes, one or two in the viral sense (CP and, in Old World viruses, AV2) and four in the complementary sense (Rep, TrAP, REn and AC4). The role of Rep (Replication-associated protein) in the coupling to the origin of replication for the initiation of DNA synthesis. The TrAP (Trans-Activating Protein) protein is responsible for activating transcription of the CP and NSP genes, as well as acting in the suppression of plant defense responses. REn (Replication-Enhancer protein) enhances replication, increasing the concentration of viral DNA. AC4 protein is related to suppression of gene silencing and, in monosegmented *Begomoviruses*, in viral movement. In some species an AC5 ORF is present, a recent study has shown that the encoded product plays an important role in suppressing plant defense responses. The CP protein is responsible for the viral encapsidation and specificity with the insect vector and, in the monosegmented *Begomoviruses*, it acts on the viral movement. The AV2 gene is present only in the *Begomovirus* from Old World and its function is related to the viral movement (BRIDDON ET AL., 2010, PASSOS, 2016 AND VANITHARANI et al., 2004).

DNA-B has the NSP genes in the viral sense and MP in the complementary sense. Nuclear Shuttle Protein (NSP) is responsible for transporting viral DNA from the nucleus to the cytoplasm, and MP (Movement Protein) protein transports the viral cell-to-cell DNA through the plasmodesmas. In monopartic *Begomoviruses*, intracellular transport is performed by CP. Movement protein (MP) is encoded by the BC1 ORF in bipartite *Begomoviruses* or by V2 in single-pores, although the two ORFs do not share sequence identity. This protein is responsible for cell-to-cell viral movement, increasing the limit of exclusion of the plasmodesms (NOUEIRY et al., 1994; RÉGO, 2016; UNSELD et al., 2001).

Begomovirus replication

Begomoviruses are transmitted by *Bemisia tabaci* in a persistent circulatory way. Viral particles are acquired via stylet during the feeding process in infected plants, enter

the esophagus and the filter chamber, and are subsequently transported through the wall of the intestine to the hemolymph where they circulate until they enter the salivary gland, from which are transmitted to new plants during insect feeding (CZOSNEK et al., 2002; GHANIM et al., 2007; GILBERTSON et al., 2015).

DNA replication occurs in three phases: initiation, elongation and termination. The *Geminivirus* Rep catalyzes the initiation and termination of circle-cycle replication by cleavage and binding of the viral DNA at a conserved site within the viral genome. Similar to a number of small DNA viruses, geminiviruses do not encode their own DNA polymerases and therefore depend on host polymerases and associated factors (collectively, the so-called host replication) for viral DNA synthesis during the elongation step. In healthy plants, the availability of host repliastomy is heavily regulated by cell cycle and developmental controls, which must be reprogrammed before geminiviruses can replicate their genomes. The virions penetrate the plant cells during the feeding process of the virulent insect-vector. Inside the cell, the genetic material is routed to the nucleus. It is unclear whether the virus moves into the encapsulated or uncapsidated nucleus, but it is believed that CP is involved in this process, interacting with the host's transport chain (Gafni and Epel, 2002; HANLEY-BOWDOIN et al., 2013).

The replicative cycle of *Geminiviruses* can be subdivided into two functionally distinct stages, characterized by specific events. The first stage of the infectious cycle involves the conversion of the circular genomic ssDNA into a covalently closed circular supercoiled dsDNA intermediate called replicative form I (RFI), which is completed with the exclusive participation of the proteins of the infected cells since the dsDNA is the transcriptionally active template. The second stage consists of the use of dsDNA as a template for the ssDNA amplification by the rolling circle mechanism, in which the Rep viral protein is required. It is responsible for the initiation reaction involving a cut within the conserved nonanucleotide (5'-TAATATT - AC-3') in all *Geminiviruses*, located in the loop of the staple structure present in the intergenic region, this protein has several biochemical activities, including single and double stranded DNA binding property and DNA binding in specific sequences, ATP hydrolysis and initiation of the circle-cycle replication mechanism. After its inception the factors necessary to complete the rolling circle phase are of cellular origin. Later, the production and encapsidation of mature circular genomic ssDNA into viral particles occurs (FERNANDES, 2010, HANLEY-BOWDOIN et al., 2013).

These mechanisms present high complexity and according to the genus it can present different modes of transmission and replication. Therefore, the study of phylogeny presents one of the effective tools in the identification and characterization of these virus.

PHYLOGENY

The world is made up of a huge variety of organisms. In order to study and understand this diversity, it became necessary to group beings by their common characteristics, that is, to classify them. There are several criteria that can provide a basis for a classification system for any set of entities. Therefore, the fundamental principle of any attempt to classify objects or organisms was to speed communication between people who use information about these organisms (CABRAL, 2011).

The first attempt of classification according to MIYAKI et al. (2001) occurred from the studies of the species of Charles Darwin and from this the concept of phylogeny arose after the concept of ancestry between species, through its first diagram published, representing the similarity between species. Thus, the phylogenies are nothing more than the indication of the supposed ancestral relations for a set of species. The term molecular phylogeny is evident, which is understood as the study of the evolutionary relations between organisms by the use of molecular data, such as sequences of nucleic acids and proteins, or other molecular markers. The logic of phylogenetic inference for molecular and morphological characters is identical, but the two have different properties, methods, and concepts. The main problems faced in the analysis of morphological data are also faced with molecular data (LIMA, 2003; RIDLEY, 2004).

The evolutionary inferences made from molecular data allows an inference about the evolutionary relations between species. Several results obtained from the molecular phylogeny have aided in the structuring and the current conception of the taxonomic classification of the species (SAKAMOTO, 2016).

The phylogenetic trees are constructed through data and parameters, being these defined like graphs that they have a hierarchical structure. In these trees, nodes are called taxonomic units, which may represent, depending on the data analyzed, species, populations, genes or proteins. The nodes are classified in terminals (leaves), when these are in the end of the tree, or in internal, when of them depart one or more descending branches. The terminal nodes represent the same samples used for inference of the tree and are therefore also called operational taxonomic units (OTUs), which in turn correspond to the basic unit (species, population, gene or protein) to be studied and compared. Internal nodes represent evolutionary events that depict the divergence of the taxonomic unit under analysis, such as speciation events, if the taxonomic unit is population, or gene duplication events, if the taxonomic unit is a gene or a protein. As the determination of internal nodes are products of a phylogenetic inference, we also call them hypothetical taxonomic units (HTUs). The branches are elements that connect the nodes and their size represents the estimated time of the evolutionary relation between the taxonomic units. The smaller the distance of the branches between the taxonomic units in comparison, the closer they are to evolutionarily (SOKAI, 1966).

Phylogenetic trees may also assume different conformations depending on the arrangement of the branches along the tree. These different conformations that a tree can assume are called topologies. The concept of topology is of great importance in phylogenetic studies because it represents the basis of all interpretation of evolutionary histories among the samples under analysis. (SAKAMOTO, 2016).

METHODOLOGY

Blast algorithm was used to search *Geminivirus* genome sequences. A total of 297 entries were selected from NCBI genbank database (<https://www.ncbi.nlm.nih.gov/genbank/>) from different countries around the world. Multiple sequence alignments were prepared with Bioedit (HALL et al. 1999) program (ClustalW algorithm). After alignment, Modeltest was run for the sequence set and the evolutionary model was chosen (Reversible G + I). The Maximum Likelihood method to build the cladogram in the MEGA 7.0 software (<http://www.megasoftware.net>). The Maximum Likelihood method is a point estimation method, which consists in estimating the unknown parameters so that the probability of observing the Y data is greater (or maximum). Thus, a (point) estimator of θ at any function (Y) is one that makes use of the information contained in a population sample $f(Y; \theta)$ to obtain a set of numbers that can be considered to represent approximately the unknown value of the parameters in θ . Thus, the estimation takes place from the realization of the function $\theta(Y)$ for a given sample (GUJARATI AND PORTER, 2011).

Due to these features, this method has the best statistical properties, under large samples (consistency, non-bias, efficiency and normal distribution). However, in practice its use requires the use of numerical methods, starting with an initial solution arbitrated by the user numerical calculation and a possibly large number of interactions (MILAGRES, 2015).

RESULTS AND DISCUSSION

The bootstrap values ranged from 9 to 100. Although we get some uncertain branches in our analysis based on bootstrap support (<70), it is clear that here in these results we can conclude that *Geminivirus (Begomovirus)* displays a multispecificity for interesting economic crops. Here, in the partial tree (Fig 1) it is clear to conclude that the *Begomovirus* does not display a great specificity for cultivars. Some studies determine that *Geminiviridae* are classified by genome organization, hosts distribution and vector specificity.

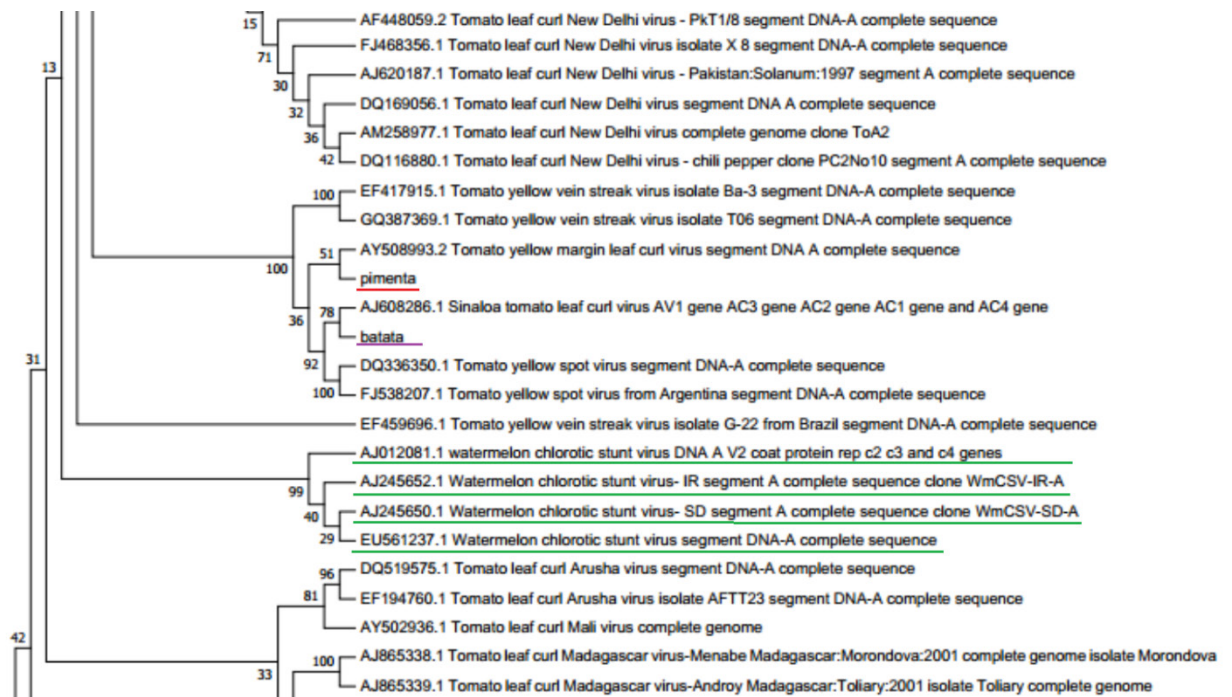


Figure 1. Partial view of the *Geminiviridae* tree. Labeled in red is the pepper sequence, while potato are labeled in purple and watermelon in green.

This is an interesting result, and corroborates with the multispecificity for hosts. We can also see, by accessing the genbank code (AY508993.2), that *Geminiviridae* from pepper is more related to a strain obtained from Merida region (Venezuela country), while potato seems to be more related to a strain from Sinaloa, Mexico. The watermelon sequences has a more closed relationship with a *Tomato* strain from Brazil.

In another segment of the *Geminiviridae* tree (Figure 2), the watermelon grouped in an adjacent clade shared by pepper and potato. However, they grouped with different strains from those observed in Figure 1. The most closed related clade to Watermelon was a clade constituted by Ramie (*Boehmeria nivea*) strains from China. The pepper and potato strains shared a clade with *Tomato* sequence obtained from Sinaloa strains. Although we found some relevant insights about the *Geminivirus* multispecificity, many aspects of its distribution and phylogeny are still uncertainly. Future studies of genetic population and distribution of these strains should be applied to improve a better understand of this virus family.

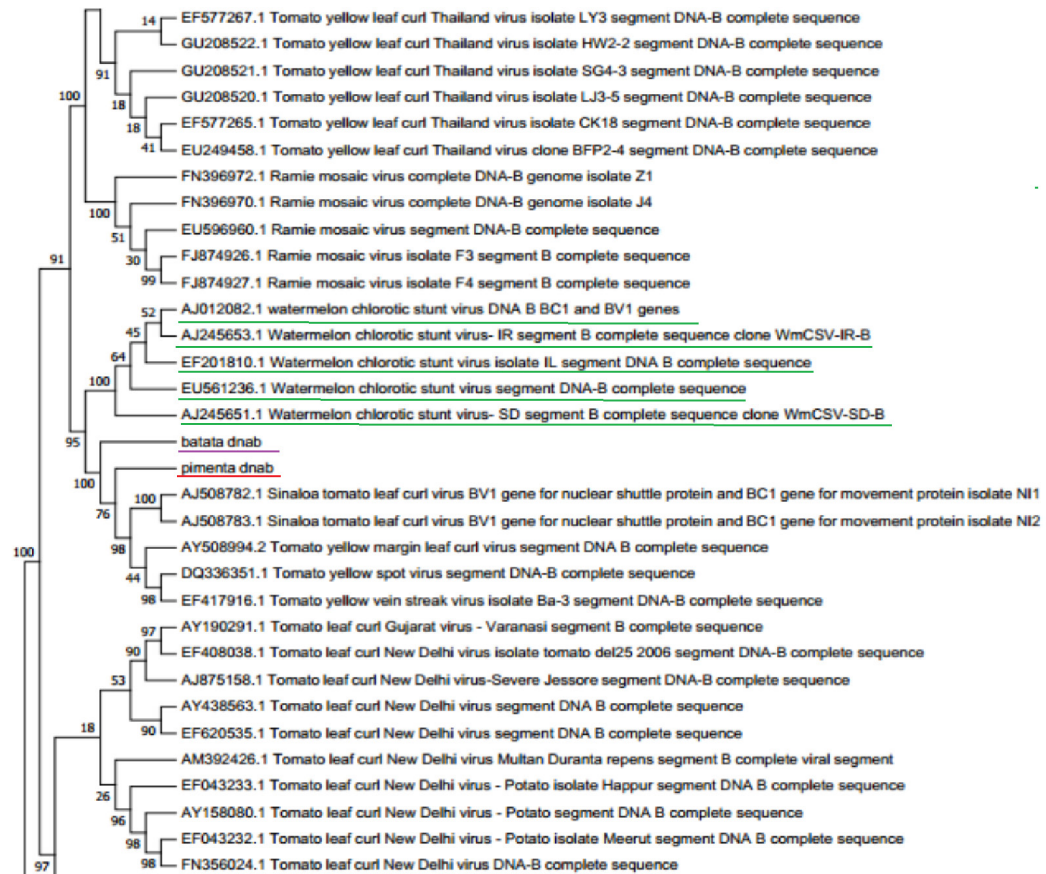


Figure 2. Partial tree view. Watermelon is labeled in green; potato in purple and pepper in red.

CONCLUSIONS

In the context of our results, and considering the *Geminiviridae* distribution, we can conclude that *Begomovirus* does not present a straight specificity to its hosts. It means, as discussed before, that the *Geminiviridae* viruses specificity is more associated with the vector than the plant hosts, which represents an evidence of multispecificity for host infection. These results will be useful for future crop management and mosaic disease control.

REFERÊNCIAS

- BRIDDON, R. W.; PATIL, B. L.; BAGEWADI, B.; NAWAZ-UL-REHMAN, M. S.; FAUQUET, C. M. Distinct evolutionary histories of the DNA-A and DNA-B components of bipartite begomoviruses. **BMC Evolutionary Biology**, v. 10, n. 1, p. 97-108, 2010.
- BRIDDON, R. W. *Geminiviridae*. In: Encyclopedia Life Sciences (eLS), **John Wiley & Sons**, 1–12, Ltd: Chichester, 2015.
- BROWN, J. K.; FAUQUET, C. M.; BRIDDON, R. W.; ZERBINI, M.; MORIONES, E.; NAVASCASILLLO, J. *Geminiviridae*. In: KING A.M.Q., ADAMS M.J., CARSTENS E.B., LEFKOWITZ E.J. (eds). *Virus Taxonomy - Ninth Report of the International Committee on Taxonomy of Viruses*. Associated Press, Elsevier Inc.; London, UK. pp. 351– 373, 2012.
- CABRAL, T. S. Filogenia molecular de fungos gasteroides das ordens Phallales e Geastrales (Phallomycetidae). Dissertação apresentada ao Programa de Pós-graduação em Ciências Biológicas da Universidade Federal do Rio Grande do Norte, 2011.

CZOSNEK, H.; GHANIM, M. The circulative pathway of begomoviruses in the whitefly vector *Bemisia tabaci*: Insights from studies with Tomato yellow leaf curl virus. **Annals of Applied Biology**, v. 140, p. 215-232, 2002.

FERNANDES, N. A. N. **Variabilidade genômica e geográfica de espécies de Begomovirus em tomateiro e em dois gêneros de plantas daninhas do Brasil**. Dissertação apresentada ao Programa de Pós-Graduação em Fitopatologia da Universidade de Brasília, 2010.

GAFNI, Y.; EPEL, B. L. The role of host and viral proteins in intra- and inter-cellular trafficking of geminiviruses. **Physiol. Mol. Plant Pathol.** 2002, 60: 231-241.

GILBERTSON, R.L.; BATUMAN, O.; WEBSTER, C.G.; ADKINS, S. Role of the insect supervectors *Bemisia tabaci* and *Frankliniella occidentalis* in the emergence and global spread of plant viruses. **Annual Review of Virology**, v. 2, p. 67-93, 2015.

GHANIM, M.; SOBOL, I.; GHANIM, M.; CZOSNEK, H. Horizontal transmission of begomoviruses between *Bemisia tabaci* biotypes. **Arthropod-Plant Interactions**, v. 1, p. 195-204, 2007.

GARCIA, M. **Uma filogenia mitocondrial de metazoários**. 212f. Petrópolis, RJ. Dissertação (Mestrado em Modelagem Computacional com Ênfase em Bioinformática) - Programa de Pós-graduação em Modelagem Computacional, Universidade Federal do Rio de Janeiro, 2007.

GUJARATI, D. N.; PORTER, D. C. **Econometria Básica**. 5ª ed. Editora: AMGH, Rio de Janeiro, 2011.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. **Nucleic Acids Symposium Series**, 41, 95-98. 1999.

HANLEY-BOWDOIN, L.; BEJARANO, E. R.; ROBERTSON, D.; MANSOOR, S. Geminiviruses: masters at redirecting and reprogramming plant processes. **Nat Rev Microbiol.** 2013 Nov;11(11):777-88.

LEKE, W.; MIGNOUNA, D.; BROWN, J.; KVARNHEDEN, A. Begomovirus disease complex: Emerging threat to vegetable production systems of West and Central Africa. **Agric. Food Secur.** 2015.

LIMA, J. P. M. **Estudos taxonômicos moleculares no táxon Phaseoleae DC. (Leguminosae, Papilionoideae) utilizando sequências de DNA ribossômico (rDNA)**. 151 f. (Dissertação de Mestrado), Universidade Federal do Ceará, Fortaleza, 2003.

MILAGRES, T.G. P. S. Otimização da função de verossimilhança de modelos GAS com aplicações em séries de vento. Relatório apresentado ao Departamento de Engenharia Elétrica da PUC – Rio, 2015.

MIYAKI, C. Y., RUSSO, C. A. de M. & PEREIRA, S. L. Reconstrução Filogenética. Introdução e o Método da Máxima Parcimônia. In: MATIOLI, S. R. (ed.) **Biologia Molecular e Evolução**. p. 97-107. Holos editora. Ribeirão Preto, SP. 2001.

NOUEIRY, A. O.; LUCAS, W. J.; GILBERTSON, R. L. Two Proteins of a Plant DNA Virus Coordinate Nuclear and Plasmodesmal Transport. **Cell**, v.76, n. 5, p. 925-932, 1994.

PASSOS, L. S. **Diversidade de espécies e filogenia de begomovírus que infectam plantas não-cultivadas nos estados do Ceará e Piauí**. Dissertação apresentada ao Programa de Pós-Graduação em Genética e Melhoramento da Universidade Federal do Piauí, 2016.

RÊGO, C. M. **Diversidade genômica de Begomovírus em tomateiros resistente (brs sena) e susceptível (h-9553)**. Dissertação apresentada ao Programa de Pós- Graduação em Fitopatologia da Universidade de Brasília , 2016.

SAITOU, N.; NEI, M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. **Mol Biol Evol.** 1987 Jul;4(4):406-25.

SAKAMOTO, T. **Ferramentas para análise filogenética e de distribuição taxonômica de genes ortólogos.** Tese apresentada ao Programa Interunidades de Pós-Graduação em Bioinformática do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, 2016.

SANDES, E. F. O.; MELO, A. C. M. A. Algoritmos paralelos exatos e otimizações para alinhamento de sequências biológicas longas em plataformas de alto desempenho. **XXXVI Congresso da Sociedade Brasileira de Computação**, 2016.

SILVA, J. C. F. **Data warehouse enriquecido com métodos de aprendizado de máquina para a família Geminiviridae.** Dissertação apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Viçosa, 2016.

SOKAL, R. R. Numerical Taxonomy. **Scientific American**, v. 215, p. 106–116, 1966.

SOUZA, L. H. T. **Análises da entropia do fluxo de informação em redes de interação proteica associadas as ferramentas de enriquecimento funcional revelam genes de interesse prognóstico em glioblastomas.** Dissertação apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal do Rio Grande do Sul, 2015.

UNSELD, S.; HOHNLE, M.; RINGEL, M.; FRISCHMUTH, T. Subcellular targeting of the coat protein of African cassava mosaic geminivirus. **Virology** 286: 373-383, 2001.

VARSANI, A.; NAVAS-XASTILLO, J.; MORIONES, E.; HERNÁNDEZ-ZEPEDA, C.; IDRIS, A.; BROWN, J. K.; ZERBINI, F. M.; MARTIN, D. P. Establishment of three new genera in the family Geminiviridae: Becurtovirus, Eragrovirus and Turncurtovirus. **Archives of Virology**, v. 159, n. 8, p. 2193-2203, 2014.

VANITHARANI, R.; CHELLAPPAN, P.; PITA, J. S.; FAUQUET, C. M. Differential roles of AC2 and AC4 of cassava geminiviruses in mediating synergism and suppression of posttranscriptional gene silencing. **Journal of Virology** 78:9487-9498, 2004.

DESENVOLVIMENTO DE FRAMEWORK PARA CRIAÇÃO DE MODELOS COMPUTACIONAIS DE CÉLULA COMPLETA

Frederico Chaves Carvalho

Programa de Pós-graduação em Modelagem Computacional em Ciência e Tecnologia –
Universidade Estadual de Santa Cruz
Ilhéus – Bahia

Paulo Eduardo Ambrósio

Programa de Pós-graduação em Modelagem Computacional em Ciência e Tecnologia –
Universidade Estadual de Santa Cruz
Ilhéus – Bahia

RESUMO: Propostas para a criação modelos computacionais de células completas que levem em consideração a função dos genes são relativamente recentes. Tais modelos buscam representar todos os processos bioquímicos intracelulares de maneira a fornecer uma maneira rápida e eficiente de obter resultados simulados confiáveis e comparáveis aos obtidos com métodos *in vitro* ou *in vivo*. O desenvolvimento desses modelos tem ajudado a consolidar o conhecimento atual da biologia, e fornecido subsídios para avanços científicos mais rápidos em áreas como a medicina e a bioengenharia. No entanto, a alta complexidade envolvida em sua construção é uma das principais barreiras para sua popularização. Visando simplificar e acelerar o processo de produção de novos modelos, nos propomos a criar um framework que represente a formalização da metodologia

seguida pelos principais modelos atuais, e guie o usuário no processo de criação e simulação dos modelos.

PALAVRAS-CHAVE: Desenvolvimento de software, Modelagem computacional, Modelos de célula completa, Ferramenta computacional.

ABSTRACT: Attempts to create whole-cell computational models that take into consideration the function of genes are relatively recent. Such models aim to represent all the intracellular biochemical processes in a way that allows fast and efficient achievement of fast and reliable results, comparable to the ones obtained from *in vitro* or *in vivo* experiments. The development of these models has been important not only to consolidate the current knowledge of Biology, but also as a way to provide subsidy for faster scientific advancements in areas such as medicine and bioengineering. However, the high complexity involved in their construction is one of the main barriers to their popularization. To simplify and accelerate the process of model building, we are creating a framework as a tool that guides the user through the methodology used to create and simulate the main current whole-cell models.

KEYWORDS: Software development, Computer modelling, Whole-cell models, Computational tool.

1 | INTRODUÇÃO

O desenvolvimento de ferramentas computacionais especializadas é um passo importante para garantir a acessibilidade e popularização de determinadas metodologias e técnicas que, de outra maneira, não seriam possíveis de serem utilizadas pelo público que mais se beneficiaria de seu uso.

Nas engenharias, por exemplo, é comum encontrar ferramentas computacionais que podem ser utilizadas para acelerar e aperfeiçoar avanços científicos. Softwares como AutoCAD, ANSYS e COMSOL são frequentemente utilizados como auxiliares no processo de projeto e investigação, não só conferindo mais celeridade, como também reduzindo os custos envolvidos. Isso ocorre pois tais softwares permitem a obtenção de resultados através de simulações computacionais, dispensando a necessidade de construção de protótipos físicos, ou realização de experimentos laboratoriais onerosos.

Por outro lado, quando observamos os métodos utilizados pela biologia e áreas afins, como a farmacologia e a medicina, percebemos que existem poucas alternativas tecnológicas com a mesma capacidade. Apesar da existência de técnicas como virtual screening e sequenciamento, por exemplo, partes importantes das ciências biológicas ainda se sustentam puramente em observações e experimentos laboratoriais para produzir resultados e avançar.

Nesse contexto, modelos computacionais tem se mostrado bastante promissores quando utilizados como ferramentas para auxiliar no processo de investigação científica na Biologia (FALL et al. 2002). Seu uso permite que cientistas sejam capazes de testar hipóteses e observar fenômenos que, de outra forma, seriam inviáveis ou impossíveis de serem observados. Em particular, os modelos computacionais de célula completa são de especial interesse para investigações que buscam compreender melhor o funcionamento de células isoladas, que podem ser desde simples bactérias até células humanas, importantes para pesquisas sobre o câncer, por exemplo.

Modelos de célula completa são modelos computacionais que buscam prever o fenótipo (características observáveis e comportamentos de um organismo) através do conjunto de dados genotípicos (código genético) e bioquímicos de uma célula (KARR et al. 2012). Para tanto, esses modelos são criados a partir da representação computacional de diversos comportamentos celulares, que ocorrem em várias escalas, desde as interações entre as menores moléculas até a forma de estruturas maiores, como a membrana plasmática.

As tentativas de criação de modelos computacionais que representam sistemas biológicos não são recentes. Ainda na década de 1950 (TURING, 1952), Allan Turing criou o primeiro modelo biológico a ser representado matematicamente e calculado utilizando um computador. Tratava-se do modelo que ilustrava suas teorias a respeito das bases químicas da morfogênese. Este modelo, assim como a maioria dos que o seguiram, utilizavam puramente equações químicas e matemáticas para representar aspectos importantes do sistema descrito. Tal abordagem resultava na criação de

modelos simplificados, mas que descreviam os sistemas com precisão suficiente para testar hipóteses e realizar observações.

O primeiro modelo de célula completa a levar em consideração o papel dos genes no ciclo de vida de uma célula foi o E-CELL (TOMITA, 1999). O Projeto iniciado em 1996, na Universidade de Keio (Japão) tinha como objetivo modelar a bactéria *Mycoplasma genitalium*, cujo genoma havia sido sequenciado em 1995 (FRASER et al. 1995). Apesar de considerar o papel dos genes no fenótipo da bactéria, o modelo representava apenas 127 dos 525 genes da *M. genitalium*, o que foi o suficiente para que o modelo pudesse replicar alguns comportamentos celulares conhecidos, como o rápido pico na concentração de ATP intracelular, que ocorre quando a célula inicia um processo de morte por falta de alimento.

Cerca de 16 anos depois, o segundo modelo da mesma bactéria foi criado por um grupo de pesquisadores da Universidade de Stanford, nos Estados Unidos (KARR et al. 2012). O novo modelo representa a função de todos os genes da *M. genitalium*, bem como a quase totalidade de seus fenômenos bioquímicos, sendo capaz de replicar computacionalmente uma série de comportamentos celulares, alguns até então pouco estudados.

Para a construção do modelo, a equipe precisou reunir e analisar dados experimentais disponíveis em aproximadamente 900 artigos científicos, extraindo cerca de 1900 parâmetros que foram utilizados na modelagem (HAYES, 2013). O modelo final é constituído de 28 submodelos, que descrevem distintos processos celulares, cada um com inputs e outputs específicos. Um dos maiores desafios nessa abordagem inovadora foi a integração dos submodelos em um único modelo, já que nem todos os módulos possuíam a mesma natureza, podendo eles ser matemáticos, booleanos, probabilísticos, baseados em regras, etc. (KARR et al. 2012).

Por se tratar de um modelo extremamente complexo e que abrange eventos que ocorrem em várias dimensões simultaneamente, é natural que o modelo seja computacionalmente custoso. Nesse aspecto, vale ressaltar que as primeiras simulações do modelo necessitaram de aproximadamente 1 dia para completar (GOLDBERG, CHEW, KARR, 2016). Uma versão paralelizada do algoritmo de simulação foi capaz de reduzir esse tempo para cerca de 9 horas, quando executado por um cluster (HAYES, 2013).

Testes conduzidos utilizando o modelo confirmaram seu potencial como ferramenta para aceleração de pesquisas científicas. Através da realização de experimentos *in silico*, a mesma equipe que construiu o modelo foi capaz de prever efeitos de parâmetros cinéticos e de determinadas enzimas sobre o ciclo de vida da célula, podendo retardar ou limitar o crescimento celular. Tais previsões foram posteriormente confirmadas e validadas através de experimentos laboratoriais (SANGHIVI et al. 2013).

Tais êxitos alcançados levam a crer que, uma vez que modelos suficientemente precisos sejam criados, será possível utilizá-los em diversas áreas do conhecimento que possuem interface com a biologia. Por exemplo, através do uso de modelos de

célula completa, farmacêuticos poderão desenvolver medicamentos mais eficientes, e personalizados para o genoma de cada paciente; e bioengenheiros serão capazes de projetar microrganismos para funções específicas, como a produção de biocombustíveis ou detecção de doenças.

Apesar do sucesso alcançado com o modelo de 2012, e as potencialidades do uso de modelos computacionais de células completas nas pesquisas científicas, a popularização de tais modelos ainda parece ser uma realidade distante. O alto volume e heterogeneidade dos dados, as dificuldades em traduzi-los em algoritmos computacionais e a falta de ferramentas projetadas especificamente para tal finalidade são alguns dos principais desafios que têm afastado diversos cientistas da construção desses modelos.

Tendo em mente tais barreiras, uma das possíveis soluções é o desenvolvimento de uma ferramenta computacional que possa ser utilizada com facilidade por pesquisadores sem conhecimentos avançados de computação. É desejável que a ferramenta criada seja o mais fiel possível à metodologia utilizada para a construção do modelo construído por Karr et al. (2012), uma vez que este representa o estado da arte dos modelos de célula completa. Dessa maneira, seria possível produzir outros modelos com qualidade semelhante, e em menor tempo, contanto que haja disponibilidade de dados completos a respeito da célula a ser modelada.

Quanto aos dados, vale ressaltar que a popularização e os constantes avanços nas técnicas de high-throughput tem propiciado um significativo aumento na quantidade de informações disponíveis a respeito das mais diversas células e processos celulares. Atualmente, diversos bancos de dados contam com conjuntos de dados suficientemente completos para desenvolver modelos mais simples.

Neste capítulo, descrevemos o processo de criação de um framework para desenvolvimento de modelos de célula completa, cujo objetivo é permitir que cientistas que possam se beneficiar com o uso desses modelos tenham condições de criá-los de maneira rápida e fácil, sem necessitar de conhecimentos avançados de computação, ou de uma equipe interdisciplinar para isso.

Nas seções subsequentes serão detalhadas as bases teóricas por trás da criação de modelos de células completas, a arquitetura e especificações do framework, e por fim, os primeiros resultados alcançados, bem como os próximos passos no desenvolvimento.

2 | BASES TEÓRICAS DO FRAMEWORK

Quando se estuda uma célula, por mais simples que esta seja, é possível identificar que trata-se de uma composição de diversos subsistemas que atuam em conjunto de maneira a possibilitar todas as funções que sustentam a vida da célula. Cada um desses subsistemas possui suas peculiaridades e componentes específicos,

que podem ou não estar presentes em outros subsistemas. Das interações entre distintos subsistemas surgem ainda propriedades emergentes, que devem ser levadas em consideração quando se modela uma célula.

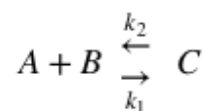
A heterogeneidade observada entre os subsistemas tem reflexos na quantidade e qualidade dos dados disponíveis sobre cada um deles. Por exemplo, já existem métodos experimentais que permitem identificar e quantificar com precisão concentrações de determinados compostos em uma célula, e suas constantes de velocidade de reação. Por outro lado, ainda é um desafio identificar com alta precisão, e de maneira quantitativa, a ação de fatores de transcrição.

Para traduzir essa complexidade inerente às características celulares em linguagem computacional, é necessário recorrer à utilização de diversos paradigmas diferentes de modelagem (COVERT, 2006). Enquanto as reações entre moléculas podem ser modeladas como um sistema de equações diferenciais ordinárias, redes reguladoras de genes, que representam a ação de fatores de transcrição, são melhor representadas computacionalmente através de modelos booleanos.

É possível classificar os diferentes tipos de modelos que podem ser utilizados para tal finalidade em três grandes grupos: modelos baseados em redes; modelos baseados em regras; e modelos estatísticos (KLIP et al., 2016). Cada um desses grupos requer diferentes tipos de input e geram outputs distintos, fazendo-os mais ou menos apropriado como solução para modelar determinados sistemas.

Os modelos baseados em redes, por exemplo, são apropriados quando os sistemas sendo modelados podem ser descritos como um conjunto de nós interconectados por relações bem definidas. Um exemplo disso são redes metabólicas, que possuem metabólitos como nós, e reações químicas como conexões entre os nós. Para representa-las, podemos utilizar sistemas de equações diferenciais ordinárias, equações diferenciais parciais ou equações diferenciais estocásticas, por exemplo

Para ilustrar esse tipo de modelo, podemos utilizar a seguinte reação:



Para a reação acima, podemos elaborar a seguinte equação diferencial ordinária, que representa a taxa com a qual a concentração de C (representada por [C]) varia com o tempo:

$$\frac{d[C]}{dt} = k_1[A][B] - k_2[C]$$

O mesmo procedimento pode ser repetido para encontrarmos a variação das concentrações das espécies A e B nesse mesmo sistema. É importante ressaltar que, nesse caso, estamos aproximando o modelo através da consideração de que este representa uma solução bem homogeneizada, onde não há gradientes de concentração, ou desigualdade na distribuição espacial das moléculas. Se quisermos

incluir dados relativos à espacialidade nos modelos, o que pode ser útil ao levar em consideração fenômenos como a difusão, por exemplo, podemos recorrer ao uso de equações diferenciais parciais para isso, como a equação abaixo:

$$\frac{\partial C}{\partial t} = D \nabla^2 C$$

Os modelos baseados em redes podem ainda ser do tipo booleanos, redes de Petri, ou estocásticos.

O segundo grupo de modelos são os modelos baseados em regras, ou modelos baseados em agentes. Assim como os primeiros, estes possuem elementos claramente definidos, porém, a relação entre eles ocorre através de um conjunto de regras, e não reações ou interações que podem ser descritas por equações. Esse tipo de modelo é de grande utilidade quando se quer modelar sistemas de sinalização celular levando em consideração tanto dados temporais como espaciais, por exemplo.

Por fim, os modelos estatísticos, incluídos no último grupo, são modelos mais apropriados para trabalhar com modelos criados a partir de grande volume de dados experimentais. Esses modelos tem se tornado mais comuns e importantes à medida que o volume de dados se torna maior, já que através deste tipo de modelagem é possível estabelecer relações de causalidade e frequências de ocorrência de determinados eventos, por exemplo.

Do ponto de vista computacional, há ainda um outro desafio relativo não à construção, mas à simulação de modelos de célula completa. Por se tratar de modelos de alta demanda computacional, a construção de um modelo não otimizado pode resultar em tempos de simulação elevados. Para resolver este problema em potencial, além da otimização do modelo e do algoritmo de simulação, uma outra técnica pode ser utilizada: a paralelização dos algoritmos.

Paralelizar um algoritmo significa permitir que sua execução seja dividida entre diferentes unidades de processamento. Tendo em mente que a maioria dos computadores atuais possuem algum grau de descentralização em suas unidades de processamento, essa técnica ganha ainda mais importância. Além disso, através da aplicação de técnicas de paralelização, permite-se que os modelos possam ser simulados eficientemente em um cluster, acelerando ainda mais a obtenção de resultados em simulações de maior complexidade.

Diversos padrões estão disponíveis para auxiliar o processo de paralelização de algoritmos. Dentre eles, os mais conhecidos, e utilizados, hoje são MPI e OpenMP. Nas últimas duas décadas, uma nova técnica de paralelização vem ganhando bastante notoriedade, que é o uso de placas gráficas (GPU) como coprocessadores. Essa técnica também pode ser aplicada de maneira a auxiliar na simulação de modelos de célula completa. Para a utilização de GPUs, os principais padrões atuais são CUDA (restritos e otimizados para placas gráficas da marca NVIDIA), e o OpenCL (irrestritos, porém com menor otimização em alguns casos).

É importante observar que, apesar de existirem formas de utilizar os padrões supracitados com qualquer linguagem de programação, esses foram desenvolvidos para implementação utilizando C ou Fortran, de maneira que a utilização dessas linguagens geralmente resulta em códigos com maior grau de eficiência.

3 | ARQUITETURA E FUNCIONAMENTO DO FRAMEWORK

Para alcançar o objetivo de ser uma ferramenta de fácil utilização para cientistas que não possuem conhecimentos avançados de programação, o framework foi planejado para guiar o usuário de forma simples e natural pelas diversas etapas do processo de criação de um modelo de célula completa.

Sua interface gráfica foi projetada de maneira a refletir a metodologia adotada como padrão para construção dos modelos de célula completa (KARR et al. 2012). Com a finalidade de estabelecer uma sequência lógica a ser seguida durante a criação de um modelo utilizando o framework, este foi dividido em 5 módulos. São eles:

- Inserção de Dados
- Modelagem
- Definições de Simulação
- Definições de Tecnologia
- Visualização de Resultados

Visando melhorar a usabilidade, a interface foi desenvolvida de maneira a permitir que o usuário possa alternar entre os módulos como achar conveniente, podendo voltar e inserir novos dados caso necessário, por exemplo. Além disso, tendo em vista o tempo necessário para a construção de um modelo, o usuário pode salvar o estado atual do modelo a qualquer momento, e continuar posteriormente a construção.

A Figura 1 mostra um esquema simplificado da arquitetura do Framework. Nela, as setas denotam o fluxo dos dados desde suas fontes (experimentos, bancos de dados e literatura) até a obtenção dos resultados finais. Na seção “Processamento”, é possível observar também a forma como a paralelização, cujos parâmetros serão definidos no módulo “Definições de Tecnologia”.

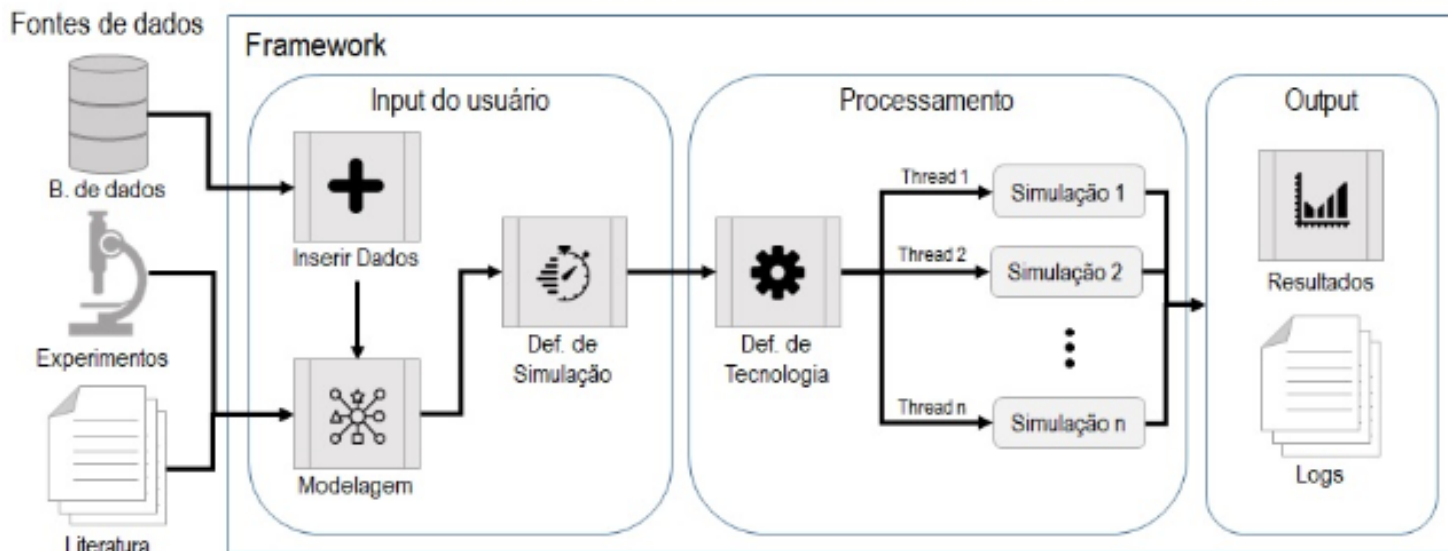


Figura 1: Arquitetura e fluxo de dados no framework

As funcionalidades básicas da interface (clique de botões, mudança de aba, inserção de dados, etc.) estão sendo desenvolvidos na linguagem Python. Essa escolha deve-se principalmente à simplicidade com que o código pode ser desenvolvido utilizando-se de bibliotecas disponíveis nesta linguagem, como numpy, pandas e matplotlib. Além disso, como a interface gráfica em si está sendo desenvolvida em Qt, Python oferece algumas soluções interessantes à organização dos arquivos, como a possibilidade de criar um arquivo à parte para a implementação das funções da interface, não alterando assim o arquivo principal, que pode ser atualizado sem necessidade de reescrita total do código.

Por não se tratar de um ponto crítico para o desempenho ou para o tempo total de simulação do modelo desenvolvido, não há necessidade de implementar técnicas de paralelização às funcionalidades básicas da interface. Por fim, a opção por essa linguagem permite implementar visualização de gráficos dinâmicos de maneira mais simples, através de extensões disponíveis na biblioteca matplotlib, que permitem integrá-la ao Qt.

Buscando melhorar a experiência do usuário com o framework, a tradução das informações biológicas em código é feita automaticamente, à medida que o usuário procede pelas etapas da modelagem. Para isso, foi implementada uma abordagem simples na qual o framework trabalha com a construção de 5 arquivos durante a construção do modelo e simulação do modelo.

O primeiro arquivo consiste em uma lista de classes, com atributos e métodos predefinidos. Cada tipo de input esperado (e permitido ao usuário) é representado por uma classe. Por exemplo, a classe “Proteínas” tem como atributos: “sequência”, “concentração”, “localização” (opcional), “gradiente” (opcional), “meia vida”, “estrutura 3D” (opcional), etc. Os métodos pré-definidos são “degradação”, “reação”, “interação”, “transporte ativo”, “mutação aleatória” (estatisticamente ativado) e “difusão”. Usuários mais avançados podem editar este arquivo de maneira a definir novos métodos que

reflitam comportamentos desejados, além dos pré-definidos.

O segundo arquivo guarda todos os dados fornecidos pelo usuário como objetos das classes apropriadas. Esse arquivo é escrito à medida que o usuário fornece informações ao programa, nos módulos “Inserir Dados” e “Modelagem”. Sempre que o usuário decidir criar uma nova molécula, ele preencherá um formulário, e ao salvar, as informações entradas serão devidamente organizadas nesse arquivo.

O terceiro arquivo corresponde à reunião de todos os algoritmos que podem ser aplicados para a simulação do modelo construído. O(s) algoritmo(s) utilizado(s) dependerão das entradas fornecidas pelo usuário no módulo “Definições de Simulação”. Do ponto de vista do desempenho em geral da simulação, este é o arquivo mais importante. Sua função dele é processar todos os inputs de acordo com as opções estabelecidas no módulo “Definições de Tecnologia”, onde o usuário poderá controlar, até certo ponto, o grau de paralelização do modelo criado, além de introduzir informações a respeito da máquina ou cluster que processará a simulação, permitindo assim que os pontos críticos de paralelismo, já pré-definidos nos algoritmos sejam otimizados.

Por padrão, a paralelização se dará da seguinte maneira: cada submodelo será inicializado como um processo, e cada uma das reações, regras ou interações entre os componentes será processado por uma thread diferente, quando possível. Caso o usuário opte por utilizar também placas gráficas para o processamento, submodelos adjacentes (que possuem maior grau de interação entre si) poderão ser atribuídos ao mesmo processo, e serão processados por threads da mesma GPU.

O quarto arquivo contém os outputs da simulação. Nele, os dados referentes ao andamento da simulação serão gravados a cada passo concluído, que representam intervalos de tempo definidos pelo usuário. Como outputs, o usuário poderá definir no módulo “Definições de Simulação”, e visualizá-los no módulo “Resultados”. A estrutura do arquivo permite que o usuário tenha acesso a toda a evolução temporal da simulação, e visualize eventos celulares importantes com maior precisão.

O quinto arquivo, contendo os registros (logs) da simulação serão salvos em um arquivo à parte, e poderão ser consultados através de uma seção da interface. Neste arquivo, o usuário terá acesso a informações complementares da simulação, como as definições, parâmetros de entrada selecionados, critérios de parada e alertas de possíveis erros e sua provável origem.

4 | PRIMEIROS RESULTADOS

Uma vez definida a metodologia para criação de modelos e os requisitos destacados na seção anterior, o desenvolvimento do framework foi iniciado pela construção de sua interface gráfica. No estágio atual de desenvolvimento, todos os módulos da interface gráfica já estão desenhados, porém somente os dois primeiros

estão funcionais.

O primeiro módulo, denominado “Inserção de dados” (Figura 2) permite que o usuário forneça todos os dados necessários à criação do modelo. Com o intuito de facilitar o processo de inserção de dados, o usuário pode optar por importar dados diretamente de bancos de dados. Todos os dados inseridos ficam registrados para posterior uso no módulo “Modelagem”.

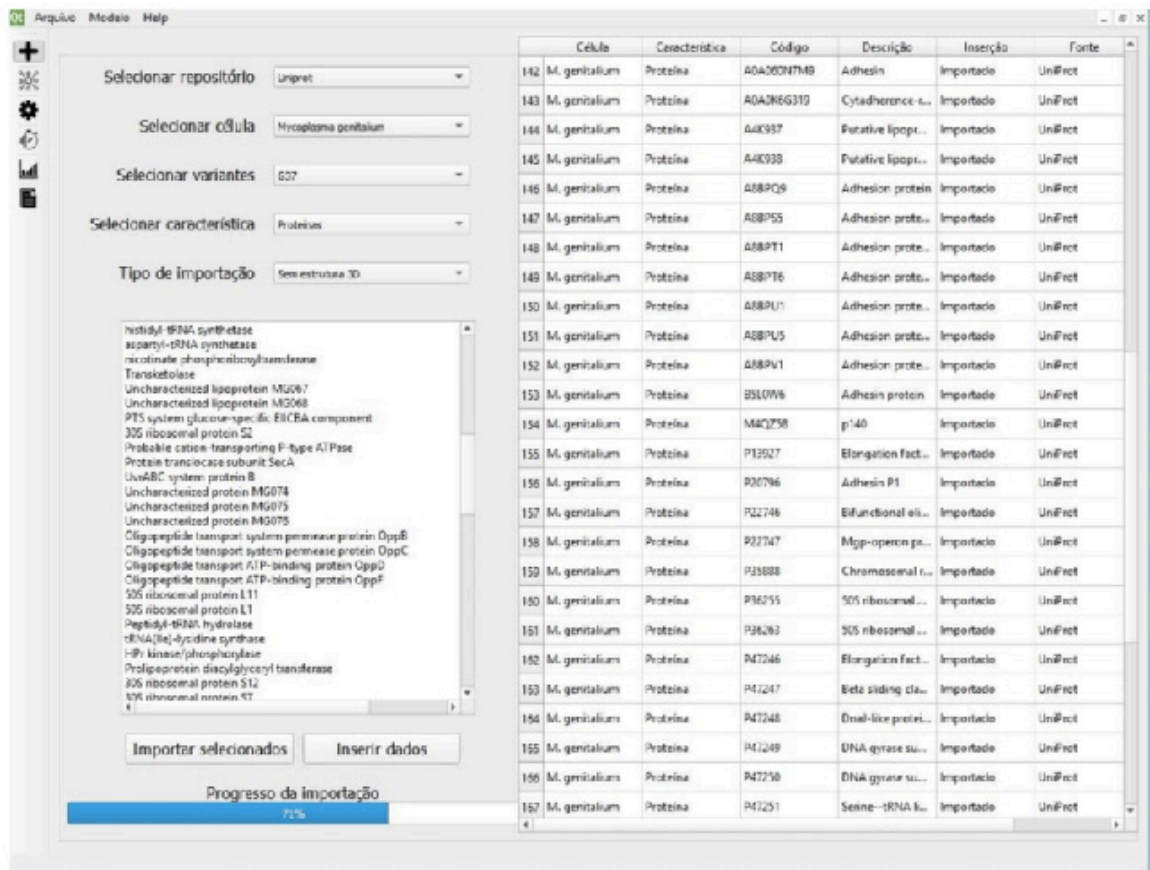


Figura 2: Exemplo da utilização da função "importar dados" do módulo 1. Essa função atualmente permite que usuário obtenha dados para construção do modelo diretamente de bancos de dados suportados (UniProt e WholeCellKB). Para importar dados, o usuário deverá selecionar o repositório, o tipo de célula, a variante (caso mais de um tipo esteja disponível), os parâmetros e o tipo de importação. Esta última opção permite escolher entre importar dados relativos à estrutura 3D, quando disponíveis, dados de interação, e sequências, o que facilita o processo de modelagem.

O módulo “Modelagem” (Figura 3) permite que o usuário construa seu modelo utilizando os dados inseridos no módulo anterior. A interface foi projetada de maneira possibilitar que o modelo seja criado visualmente, onde o usuário define o tipo de modelo, os elementos participantes e as relações entre eles. Dessa forma é possível construir submodelos utilizando o paradigma de modelagem apropriado, devendo o usuário definir as regras de integração dos submodelos para a obtenção de um modelo final coerente.

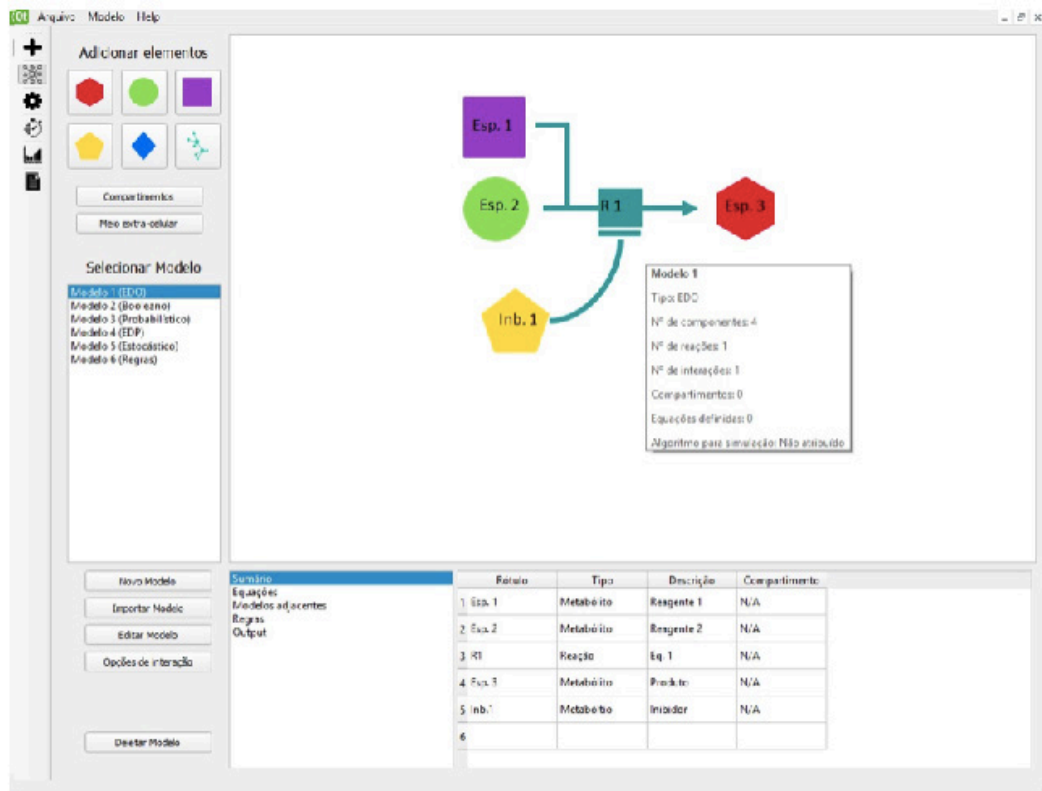


Figura 3: Exemplo de utilização do módulo "Modelagem". Neste módulo, o usuário pode utilizar formas geométricas para representar componentes do sistema modelado, e estabelecer a relação entre eles através de equações ou regras. Neste exemplo, o modelo criado é do tipo EDO, e representa uma reação genérica sob a ação de um inibidor. Clicando sobre qualquer uma das espécies, inibidor ou reação representada, o usuário pode alterar concentrações, parâmetros ou equações que governam a reação, por exemplo. Clicando sobre as linhas é possível estabelecer as relações entre as espécies representadas, de maneira a gerar equações pré-definidas automaticamente, agilizando o processo de criação de modelos.

O terceiro módulo ("Definições de Simulação") permite que o usuário defina os algoritmos a serem utilizados, a resolução da simulação (definição da escala de tempo e passo), variáveis a serem monitoradas, outputs, etc. Neste módulo o usuário pode ainda definir eventos que deverão ocorrer quando determinadas condições forem atendidas, a exemplo de perturbações devem ser introduzidas após a decorrência de certo tempo. Dessa forma, será possível simular experimentos com maior precisão dos resultados.

O quarto módulo permite definir e analisar critérios de paralelização, como quantidade de processos ou threads utilizados, como eles serão utilizados, e se a simulação deve ou não utilizar placas gráficas como coprocessadores, quando disponíveis. Por se tratar de um módulo que exige conhecimentos avançados, todas as configurações vem com valores pré-definidos recomendáveis que, permitem que a simulação seja paralelizada de forma genérica para reduzir os tempos de simulação de maneira razoável. Usuários avançados poderão ajustar critérios que exigem maior conhecimento de computação de alto desempenho, como balanceamento de carga, alocação de memória e divisão das threads e processos.

O último módulo ("Resultados") permite analisar os resultados graficamente. Nele o usuário pode ter acesso a toda a evolução temporal das variáveis previamente

selecionadas, de maneira que o efeito de todos os eventos definidos possam ser visualizados tanto visual quanto quantitativamente. Em suma, esse módulo tem o objetivo de permitir uma análise aprofundada dos resultados da simulação.

5 | PRÓXIMOS PASSOS

Além da implementação das funcionalidades dos módulos restantes, as próximas etapas do desenvolvimento do framework incluem melhorias no suporte a diferentes bancos de dados, suporte a modelos SBML, criação e validação de um primeiro modelo de célula completa utilizando o framework, e publicação do framework.

Uma vez implementadas todas as funções do framework, será necessário proceder à validação do framework e de modelos criados por esta ferramenta. Para isso, um primeiro modelo será desenvolvido e simulado com o framework, e seus resultados comparados ao modelo atual da bactéria *Mycoplasma Genitalium* e a dados provenientes de literatura, que não tenham sido utilizados na construção do modelo.

Após validado, o framework será disponibilizado através da internet com licença Open Source, juntamente com sua documentação e manual, de maneira a permitir que pesquisadores nas áreas de biologia, farmácia, química, medicina, biomedicina, etc. tenham acesso a esta ferramenta em suas pesquisas, e possam produzir modelos computacionais utilizando-os como auxiliares em suas investigações.

Levando em consideração que nem todas as informações de interesse podem ser encontradas em um único, ou poucos bancos de dados, é desejável que o framework seja capaz de importar dados de uma maior variedade de repositórios. Dessa forma, o usuário poderá obter dados de maneira mais simples e rápida, acelerando ainda mais o processo de construção de modelos.

Outro aspecto importante é a possibilidade de compartilhar modelos, estejam eles prontos ou em desenvolvimento. Dessa maneira, será possível a criação de modelos de forma colaborativa. Por ser um formato já estabelecido, é interessante que o framework implemente uma forma de traduzir os arquivos referentes aos dados do modelo, e suas configurações, em um único arquivo SBML, que possa ser publicado em plataformas de compartilhamento de modelos, por exemplo.

6 | AGRADECIMENTOS

Os autores deste trabalho agradecem à FAPESB (Fundação de Amparo à Pesquisa do Estado da Bahia) pelo financiamento do projeto, através da concessão da bolsa BOL0029/2018.

Os ícones utilizados para a produção dos diagramas e do Framework foram produzidos pelos usuários “Freepik” e “geotatah”, e disponibilizados gratuitamente em flaticon.com.

REFERÊNCIAS

- COVERT, M. W. **Integrated Regulatory and Metabolic Models**. In: KRIETE, A.; EILS, R. *Computational System Biology*. San Diego: Elsevier Inc., 2006. cap. 10.
- COVERT, M. W. **Simulating a living cell**. *Scientific American*, v. 310, n. 1, p. 44–51, 2014.
- FALL, C.P. et al. **Computational Cell Biology**. New York: Springer, 2002.
- FRASER, C. M. *et al.* **The Minimal Gene Complement of Mycoplasma genitalium**. *Science*, v. 270, n. 5235, p. 397–404, 1995.
- FREDDOLINO, P. L.; TAVAZOIE, S. **The dawn of virtual cell biology**. *Cell*, v. 150, n. 2, p. 248–250, 2012.
- GOLDBERG, A. P.; CHEW, Y. H.; KARR, J. R. **Toward Scalable Whole-Cell Modeling of Human Cells**. *Proceedings of the 2016 annual ACM Conference on SIGSIM Principles of Advanced Discrete Simulation - SIGSIM-PADS '16*, p. 259–262, 2016.
- HAYES, B. **Imitation of life**. *American Scientist*, v. 101, n. 1, p. 10–15, 2013.
- HOOPS, S. *et al.* **COPASI - A COmplex PATHway Simulator**. *Bioinformatics*, v. 22, n. 24, p. 3067–3074, 2006.
- HUCKA, M., et al. **The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models**. *Bioinformatics*, v. 19, n. 4, p. 524–531, 2003.
- KARR, J. R. *et al.* **A whole-cell computational model predicts phenotype from genotype**. *Cell*, v. 150, n. 2, p. 389–401, 2012.
- KARR, J. R.; TAKAHASHI, K.; FUNAHASHI, A. **The principles of whole-cell modeling**. *Current Opinion in Microbiology*, v. 27, p. 18–24, 2015.
- KLIPP, E.; LIEBERMEISTER, W.; WIERLING, C.; AXEL, K. **Systems Biology**. 2. ed. Weinheim: Wiley-VCH, 2012.
- SZIGETI, B. et al. **A blueprint for human whole-cell modeling**. *Current Opinion in Systems Biology*, v. 7, p. 8–15, 2018.
- TOMITA, M. et al. **E-CELL: Software environment for whole-cell simulation**. *Bioinformatics*, v. 15, n. 1, p. 72–84, 1999.
- TOMITA, M. **Whole-cell simulation: A grand challenge of the 21st century**. *Trends in Biotechnology*, v. 19, n. 6, p. 205–210, 2001.
- TURING, A. M. **The chemical basis of morphogenesis**. *Bulletin of Mathematical Biology*, v. 237, n. 641, p. 37–72, 1952.
- WALTEMATH, D. et al. **Toward Community Standards and Software for Whole-Cell Modeling**. *IEEE Transactions on Biomedical Engineering*, v. 63, n. 10, p. 2007–2014, 2016.

IN-SILICO DETOXIFICATION EVIDENCE OF THE HERBICIDE BISPYRIBAC SODIUM BY A THEORETICAL MODEL OF GLUTATHIONE S-TRANSFERASE TAU 5 FROM *Oryza sativa* L.

Vinícius Costa Amador

Universidade Federal Rural de Pernambuco,
Departamento de Agronomia, Recife-PE.

Ravenna Lins Rodrigues

Universidade Federal de Campina Grande, Centro
de Desenvolvimento Sustentável do Semiárido.
Sumé-PB.

Felipe de Oliveira França

Universidade Federal de Campina Grande, Centro
de Desenvolvimento Sustentável do Semiárido.
Sumé-PB.

Rafael Trindade Maia

Universidade Federal de Campina Grande, Centro
de Desenvolvimento Sustentável do Semiárido.
Sumé-PB.

ABSTRACT: Rice (*Oryza sativa* L.) is one of the most important crops in the world, and it has been the primary source of nutritional layout in developing countries in Asian. Despite this remarkable importance, there are few studies about the development of techniques that minimize the potential problems inherent of this grain cultivation, such as competition with weeds, making necessary the use of herbicides. The glutathione S-transferases (GSTs) superfamily confers to rice protection against biotic and abiotic stress, and herbicide resistance. However, the three-dimensional structure of a GST Tau class, is unsolved.

The objectives of this work were to develop a reliable comparative model for the s-transferase glutathione class Tau 5 from rice, and simulate docking interactions, against herbicides bentazon and metsulfuron. Results showed that the predicted model is reliable and has structural quality. Ramachandran plot set 95,4% of the residues in the most favored regions. All complexes showed negative binding energies values; and bispyribac sodium docked to the glutathione tripeptide, and it represents an in-silico evidence of glutathione conjugation with this herbicide.

KEYWORDS: Molecular Modeling, Docking, OsGST, Herbicide.

1 | INTRODUCTION

Rice (*Oryza sativa* L.) is one of the most important crops in world, being 90% of the mundial rice cultivation arising from Asia, supplying 60% of the rice mundial demand (DOGARA; JUMARE, 2014). This means an importance against its impact on the agrobusiness, since its nutritional value consists in the primary source of nutritional layout in developing countries in Asian (DOGARA; JUMARE, 2014). According to FAO, and the International Institute of Rice Research (IRRI), rice needs more production to supply the alimentation demand,

since for the year 2050 it is estimated that the demand increases at least 60% in relation to 2016 production level (ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA ALIMENTACIÓN Y LA AGRICULTURA, 2016). However, there is a few studies about development of techniques with the potential to minimize the problems inherent to the cultivation of this cereal, and competition with weeds, (SANTOS, 2006), without despising the damage caused by many herbicides.

Among detoxification enzymes, the most known and studied are glutathione transferases, which conjugates these xenobiotics and turn them into a water soluble complex (WILCE e PARKER, 1994; KREUZ, TOMMASINI, MARINOIA, 1992). This enzymes confers to rice catalitic action, and protection agains biotic and abiotic stress including herbicides resistance/tolerance (FROVA, 2003, 2006), and its main reaction consists in the conjugation of the tripeptide glutathione to a hydrophobic compound, making it more soluble and less (MARRS, 1996), maintaining cellular homeostasis. Tau class GSTs are extremely important as they are involved in the metabolization to a broad spectrum of important commercial herbicides (EDWARDS, DIXON, WALBOT, 2000). A study made by (Lajmanovich et al., 2013) suggest that Bispyribac Sodium has a correlation with Glutathion S-trasnferase expression. The aim of this study was to construct a theoretical model for a tau 5 *Oryza sativa* glutathione S-transferase (*OsGSTU5*) and perform docking simulations against Bispyribac Sodium herbicide.

2 | MATERIAL AND METHODS

2.1 Data mining

At first, the primare *OsGSTU5* (*Oryza sativa* L. glutathiona S-transferase tau 5) sequence were obtained in *.fasta* extension in NCBI data bank, a good quality structure model were searched for template using the “BLASTp” (Basic Local Alignment Search Tool for proteins - <http://blast.ncbi.nlm.nih.gov/Blast.cgi>) in *.pdb* extension from PDB (Protein Data Bank, <http://www.rcsb.org/>) data bank.

2.2 Homology modeling and model validation

The *OsGSTU5* primary sequence was obtained from NCBI database (<https://www.ncbi.nlm.nih.gov/>). The SWISS-MODEL server Automated Mode tool (<http://swissmodel.expasy.org/>) was used to generate the model (ARNOLD et al., 2006; KIEFER et al., 2009) The model validation process used the Ramachandran graph analysis through the PROCHECK program (LASKOWSKI et al., 1993) to verify the stereochemical quality of the structure. Local quality was accessed by ANOLEA (MELO, 1998) and GROMOS force fields (van GUNSTEREN et al., 1998). All the generated docked complexes were visualized with Visual Molecular Dynamics software (SURHONE et al., 2010).

2.3 Protein pockets identification, docking simulation and anchor residues

identification

The protein pockets identification where using the ghecom 1.0 finder server used to find multi-scale pockets on protein surfaces using mathematical morphology (KAWABATA 2007, 2010).

The as ligands structure (Table 1.0) were obtained from the ZINC database (<http://zinc.docking.org/>) in *.mol2* extension files. These files were converted to *.pdbqt* in Autodock 4.2.1 (MORRIS et al., 2009) (<https://www.chpc.utah.edu/documentation/software/autodock.php>), polar hydrogens were removed and their molecules were assigned with the Gasteiger parameters (GASTAIGER, 1980).

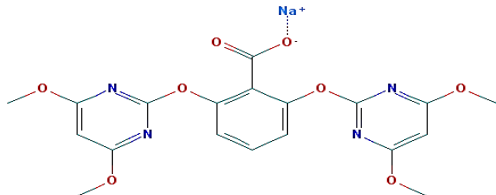
Name	Zinc ID	Class	HRAC	2D
Bispyribac sodium	ZINC 04098944	Ácido Pirimidiniloxibenzóico	B	

Table 1. Herbicide used in docking simulation

The *OsGSTU5* theoretical model, was converted to *.pdbqt* file in Autodock, hydrogens and Kollman parameters were added (WEINER et al., 1984). The GTX was treated as a cofactor. Docking simulations were run on the Autodock 1.5.6 program (MORRIS et al., 2009) and the Lamarckian genetic algorithm (LGA) was chosen. The simulations had the following parameters: 10,000 replicates, energy analyzes per 1,500,000 and 27,000 generations, population size of 150 and mutation rates and crossing-over of 0.02 and 0.08 respectively. The 10 conformations were generated that were ranked based on the lowest energy and important residues interaction were analyzed in the VMD (SURHONE et al., 2010) (<http://www.ks.uiuc.edu/Research/vmd/>).

3 | RESULTS AND DISCUSSION

3.1 Data mining

The template used was a *crystal structure of a glutathione S-transferase PtGSTU30* from *Populus Trichocarpa*, in complex with GSH (PDB ID: 5J4U) presented 1.249 Å resolution (THOM et al., 2002), and high 55.09% identity value, revealing a homology between the *OsGSTU5* and *PtGSTU30* proteins appropriate for the modeling what was perceived by the results with other Tau class proteins used in researches (KILILI et al., 2004).

Target Protein		Template data				
Protein	ID	Protein	(PDB-ID)	Metod	Identy	Resolution Å
<i>OsGSTU5</i>	AAG32470.1	PtGSTU30	5j4u	XRD	55.09	1.249

Table 2. Target and template proteins data description

3.2 Sequences alignment and conserved regions identification

The Figure 1 shows conserved and semiconserved regions between *OsGSTU5* and PtGSTU30 (template) sequences (labeled), and some important anchor residue in the catalitic pocket HIS 51 (green arrow) other important residues like LYS53, LYS 111 and LYS 112 (red arrows) were not conserved residues.

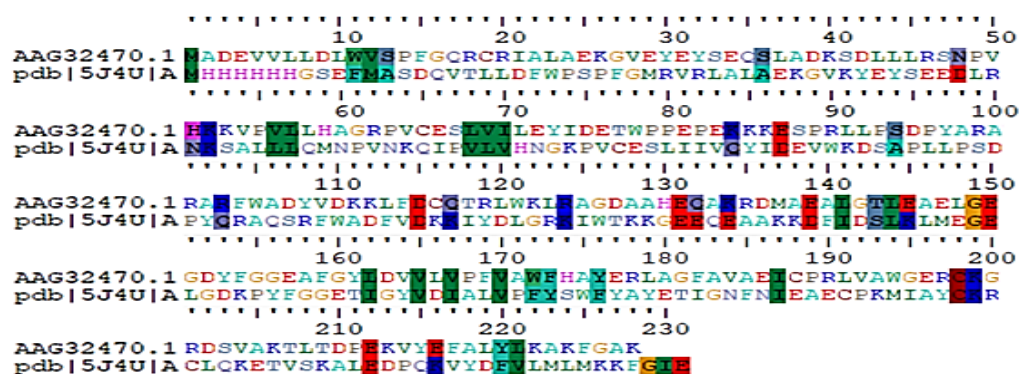


Figure 1. Sequences alignment, identification of conserved and semiconserved regions (labeled) and anchor residues (arrows)

3.3 Theoretical model construction, quality evaluation and validation

According to Laskowski (1993, 2012), a reliable predicted model are supposed to display over 90% of residues in core regions of Ramachandran plot (A, B and L), for Ho (2005) some residues as glycine and proline has predictable and distinct distribution on the Ramachandran plot, as they present different stereochemical patterns. The stereochemical quality was accessed considering the Laskowski (2012) critters, the Ramachandran plot showed 95.4% of the residues (black squares and triangles) in regions that were more favorable (red), 4.1% in allowed regions (yellow), 0.5% in generously allowed regions (cream) and 0.0% in regions (white) as shown in Figure 2, a model validation results by Maia and Nadvorny (2014) had 100.0% of the residues in allowed regions.

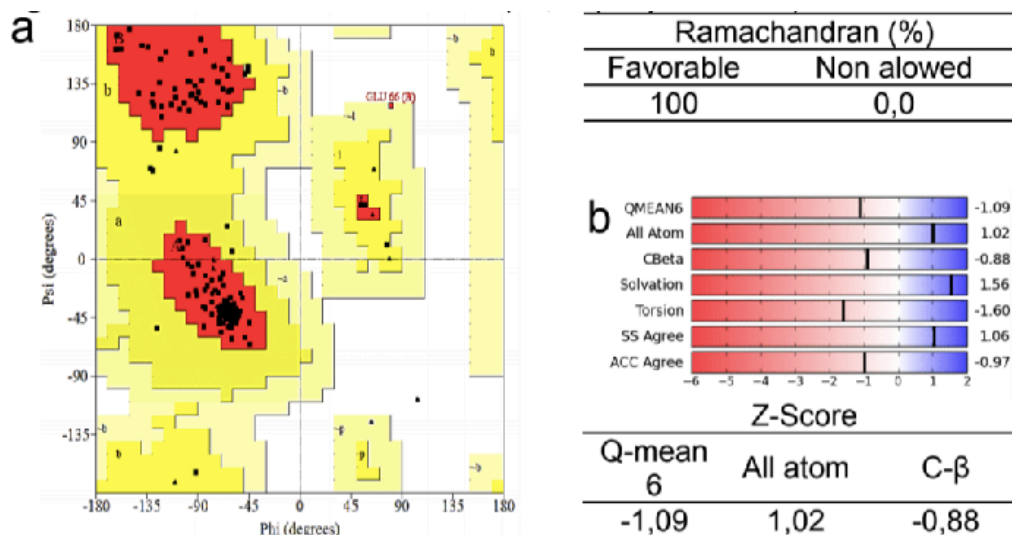


Figure 2. Model validation a. Ramachandran plot, b. quality mean for 6 parameters.

Quality mean for 6 parameters Z-score (Qmean6) was between 1 and 2 and considered appropriated, meaning that the evaluation of the theoretical native structure protein in comparison with experimental models of the similar size (residues) in databanks had its stereochemical and atom parameters average considered good, showed by the red arrow/star situated in Figure 3.

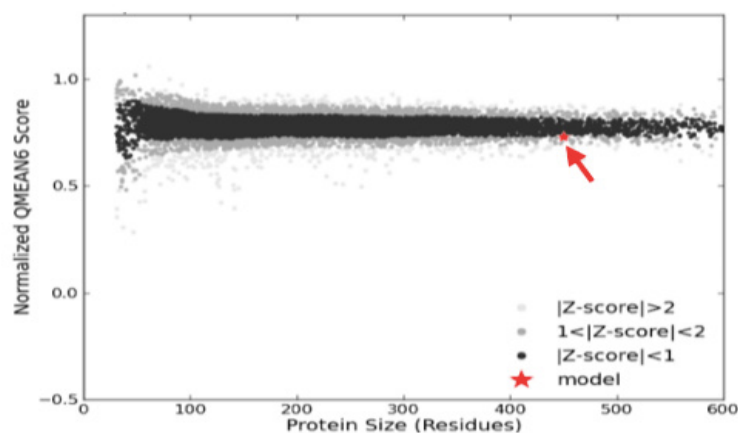


Figure 3. Theoretical model quality comparison among experimental models quality.

The results from ANOLEA (MELO, 1998), and for GROMOS (van GUNSTEREN; BERENDSEN, 1990), were generally negative values, revealing a model with stable energy, and negative values for majority of the residues, corroborating with the results showed by Hamid (2013), Maia and Nadvorny (2014). The generated and validated theoretical model had the atomic coordinate considered satisfactory for its native structure and appropriated for Docking test.

3.4 Docking and anchor residues and protein pockets identification

The protein pockets were predicted (Figure 4. and Figure 4. A) by a multi-scale, mathematical morphology algorithm (KAWABATA 2007, 2010). The docking anchor

residues were compared through the Ghecom results, and was perceived that they belonged to the best pocket (in red), the catalytic cleft (Figure 4.B), near the detoxifying subunit.

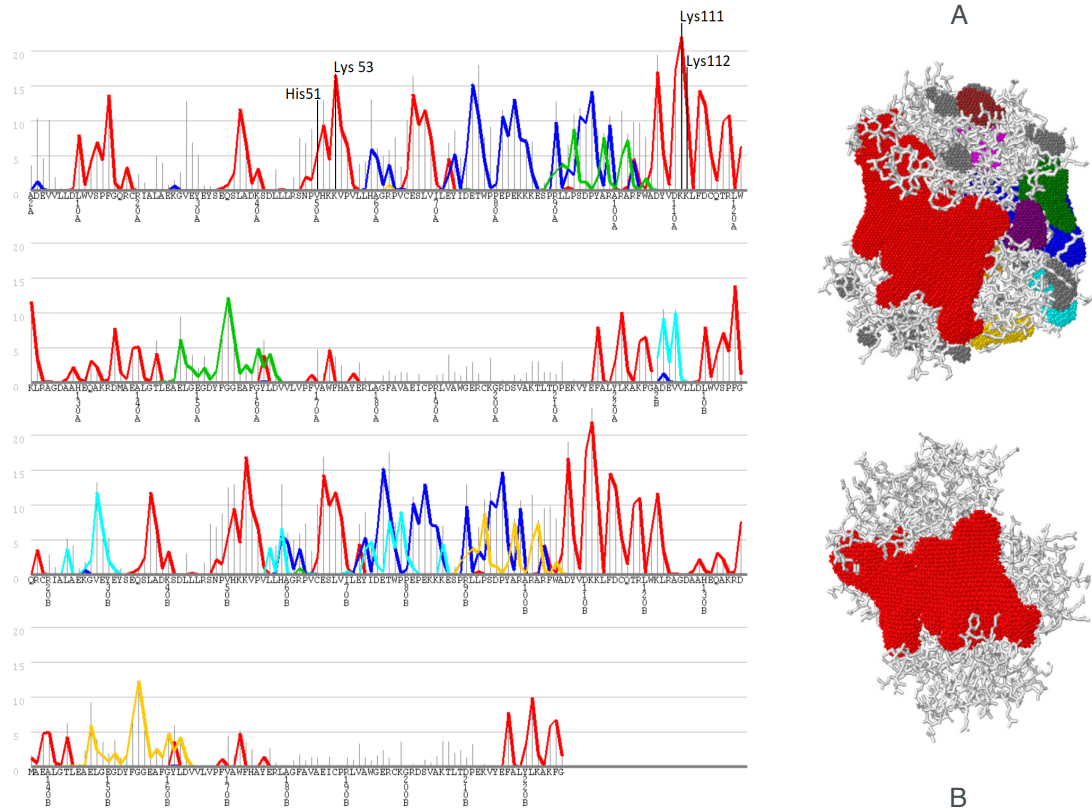


Figure 4. Pockets and aminoacids sequence description, and anchor residues described
Figure4A All ranked pockets, Figure4 B Best pocket described (red)

The docking results for the *Bispyribac Sodium* “2,6-bis[(4,6-dimethoxypyrimidin-2-yl)oxy]benzoic acid”, processed by the autodock program ranked ten possible complexes based on intermolecular energy scores, binding energy, and H-bonds that reveals atom (and residues) of the docked region binding of the ligand and the protein (MORRIS *et al.* 2009) that present favorable interaction information of the model described in Table 3.

Bispyribac Sodium Herbicide "2,6-bis[(4,6-dimethoxypyrimidin-2-yl)oxy]benzoic acid"

Protein	Ranking	Sistem Energy	Inhibition Constant (Ki)	Binding Atoms	Image
				Ligand_Protein residue_distance Å	
	1	-4,71	354,22	<O>O:C1_ LYS112:NZ1_3,68 <O>O:O8_ LYS53:HZ3_2,70 <O>O:O8_ HIS51:ND1_3,16	
U5	2	-4,01	1,16	<O>O:O8_ HIS51:HD1_2,66 <O>O:O8_ LYS53:HZ1_1,73 <O>O:C1_ GSH1:H12_2,52 <O>O:C1_ LYS111:HZ2_2,52	
	3	-3,62	2,21	<O>O:C17_ LYS112:HZ1_2,98 <O>O:O8_ LYS111:HZ2_1,98 <O>O:O3_ GSH2:H11_2.13	

Table 3. Best three dockings ranked by the autodock, with descriptions of the interaction.

For clarification the non GSH ligand atoms distribution on the protein topology is described in the Figure 5, and can be perceived they belonging to the catalytic cleft of the protein (in White).

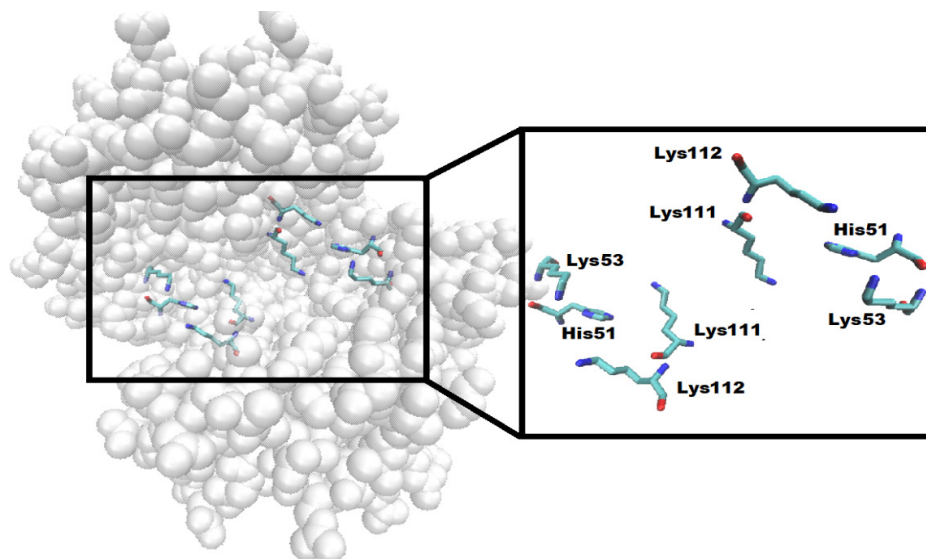


Figure 5. Description of anchor residues location.

The docking results for herbicide Bispyribac Sodium and *OsGSTU5* are displayed on the Table 3. and suggest that Lysins (LYS) situated on the catalytic pocket are important residues for the interaction, representing anchor residues. The anchor residues seem to be three lysines at positions 53, 111 and 112 and a histidine at the 51 position, located on the catalytic H-site of the enzyme (Figure 5) and a GSH bound listed in Table 3, the results of the docking still reveals that the herbicide can dock to U5 protein, in the tripeptide glutathione (GSH) subunit showed in Figure 3, suggesting a possible detoxifying process (KILILI, 2004). It is probably that *OsGSTU5* metabolizes bispyribac sodium herbicide through GSH-conjugation (YAMAMOTO et al., 2012). A work done by Lajmanovich and Junges (2013) and another done by Lu (2013) showed a correlation between Bispyribac Sodium and the superfamily of GST expression, it is possible that *OsGSTU5* is one of the superfamily members that suffer the same expression effect indicating the correlation between the herbicide metsulfuron and *OsGSTU5*.

4 | CONCLUSIONS

The results lead to conclude that the theoretical model developed (*OsGSTU5*), presents quality and is a representative native model suitable to the docking test. Our insights could be used to appoint possible molecular markers, for further marker-assisted selection tolerance/resistance to herbicide in plant studies.

ACKNOWLEDGMENT

BCT (Biologia Computacional e Teórica) CDSA (Centro de Desenvolvimento Sustentável do Semiárido CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) NemPe (Núcleo de estudos em melhoramento genético de plantas de

Pernambuco).

REFERENCES

ARNOLD, Konstantin, et al. **The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling.** *Bioinformatics*, v. 22, n. 2, p. 195-201, 2006.

SANTOS, HG dos et al. **Cultivo do arroz de terras altas no estado de Mato Grosso.** Embrapa Arroz e feijão, 2006.

DOGARA, Abdulrahman Mahmoud; JUMARE, Aisha Ishaq. **Origin, distribution and heading date in cultivated rice.** *Int J Plant Biol Res*, v. 2, n. 1, p. 1008, 2014.

EDWARDS, Robert; DIXON, David P.; WALBOT, Virginia. **Plant glutathione S-transferases: enzymes with multiple functions in sickness and in health.** *Trends in plant science*, v. 5, n. 5, p. 193-198, 2000.

FROVA, Carla. **Glutathione transferases in the genomics era: new insights and perspectives.** *Biomolecular engineering*, v. 23, n. 4, p. 149-169, 2006.

FROVA, Carla. **The plant glutathione transferase gene family: genomic structure, functions, expression and evolution.** *Physiologia plantarum*, v. 119, n. 4, p. 469-479, 2003.

GASTEIGER, Johann; MARSILI, Mario. **Iterative partial equalization of orbital electronegativity—a rapid access to atomic charges.** *Tetrahedron*, v. 36, n. 22, p. 3219-3228, 1980.

HAMID, Azzmer Azzar Abdul et al. **Molecular modelling and functional studies of the non-stereospecific α -haloalkanoic acid Dehalogenase (DehE) from *Rhizobium* sp. RC1 and its association with 3-chloropropionic acid (β -chlorinated aliphatic acid).** *Biotechnology & Biotechnological Equipment*, v. 27, n. 2, p. 3725-3736, 2013.

HO, Bosco K.; BRASSEUR, Robert. **The Ramachandran plots of glycine and pre-proline.** *BMC structural biology*, v. 5, n. 1, p. 14, 2005.

KAWABATA, Takeshi; GO, Nobuhiro. **Detection of pockets on protein surfaces using small and large probe spheres to find putative ligand binding sites.** *Proteins: Structure, Function, and Bioinformatics*, v. 68, n. 2, p. 516-529, 2007.

KAWABATA, Takeshi. **Detection of multiscale pockets on protein surfaces using mathematical morphology.** *Proteins: Structure, Function, and Bioinformatics*, v. 78, n. 5, p. 1195-1211, 2010.

KIEFER, Florian et al. **The SWISS-MODEL Repository and associated resources.** *Nucleic acids research*, v. 37, n. suppl_1, p. D387-D392, 2008.

KILILI, Kimiti G. et al. **Differential roles of tau-class glutathione S-transferases in oxidative stress.** *Journal of Biological Chemistry*, 2004.

KREUZ, Klaus; TOMMASINI, Roberto; MARTINOIA, Enrico. **Old enzymes for a new job (herbicide detoxification in plants).** *Plant Physiology*, v. 111, n. 2, p. 349, 1996.

LAJMANOVICH, Rafael C. et al. **Individual and mixture toxicity of commercial formulations containing glyphosate, metsulfuron-methyl, bispyribac-sodium, and picloram on *Rhinella arenarum* tadpoles.** *Water, Air, & Soil Pollution*, v. 224, n. 3, p. 1404, 2013.

- LASKOWSKI, Roman A. et al. **PROCHECK: a program to check the stereochemical quality of protein structures.** Journal of applied crystallography, v. 26, n. 2, p. 283-291, 1993.
- LASKOWSKI, R. A.; MACARTHUR, M. W.; THORNTON, J. M. **PROCHECK: validation of protein-structure coordinates.** 2006.
- MAIA, Rafael Trindade; NADVORNY, Daniela. **Molecular docking of Anopheles gambiae and Aedes aegypti glutathione S-Transferases Epsilon 2 (GSTE2) against usnic acid: an evidence of glutathione conjugation.** Brazilian Archives of Biology and Technology, v. 57, n. 5, p. 689-694, 2014.
- MARRS, Kathleen A. **The functions and regulation of glutathione S-transferases in plants.** Annual review of plant biology, v. 47, n. 1, p. 127-158, 1996.
- MELO, Francisco; FEYTMANS, Ernest. **Assessing protein structures with a non-local atomic interaction energy1.** Journal of molecular biology, v. 277, n. 5, p. 1141-1152, 1998.
- MORRIS, Garrett M. et al. **AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.** Journal of computational chemistry, v. 30, n. 16, p. 2785-2791, 2009.
- ORGANIZACIÓN DE LAS NACIONES UNIDAS PARA LA ALIMENTACIÓN Y LA AGRICULTURA. **El estado mundial de la agricultura y la alimentación. Organización de las Naciones Unidas para la Agricultura y la Alimentación - (FAO),** p. 192, 2016.
- Surhone LM, Timpledon MT, Marseken SF. **Visual Molecular Dynamics [J].** Version 1.9.3 [software] 2010. Betascript Publishing. 2010. Available: <https://www.ks.uiuc.edu/Research/vmd/>.
- THOM, Russell et al. **Structure of a tau class glutathione S-transferase from wheat active in herbicide detoxification.** Biochemistry, v. 41, n. 22, p. 7008-7020, 2002.
- VAN GUNSTEREN, Wilfred F.; BERENDSEN, Herman JC. **Computer simulation of molecular dynamics: Methodology, applications, and perspectives in chemistry.** Angewandte Chemie International Edition in English, v. 29, n. 9, p. 992-1023, 1990.
- VAN GUNSTEREN, Wilfred F, et al. **GROMOS force field.** Schleyer P.V.R. et al. (Orgs.). Encyclopedia of computational chemistry. 1998. p. 1211–1216.
- WEINER, Scott J. et al. **A new force field for molecular mechanical simulation of nucleic acids and proteins.** Journal of the American Chemical Society, v. 106, n. 3, p. 765-784, 1984.
- WILCE, Matthew CJ; PARKER, Michael W. **Structure and function of glutathione S-transferases.** Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology, v. 1205, n. 1, p. 1-18, 1994.
- YAMAMOTO, Kohji et al. **Structural basis for catalytic activity of a silkworm Delta-class glutathione transferase.** Biochimica et Biophysica Acta (BBA)-General Subjects, v. 1820, n. 10, p. 1469-1474, 2012.

INVESTIGAÇÃO IN SILICO DA EFICÁCIA DE FÁRMACOS ANTIVIRAIS NA INIBIÇÃO DA NS5 DO VÍRUS DA ZIKA

Henriqueta Monalisa Farias

Universidade Federal de Campina Grande
Sumé – Paraíba – henriquetamonalisa@gmail.com

Rafael de Lima Medeiros

Universidade Federal de Campina Grande
Cuité – Paraíba - rafaelufcg@gmail.com

Franklin de Ferreira Farias Nóbrega

Universidade Federal de Campina Grande
Sumé – Paraíba - franklinnobrega@yahoo.com.br

Rafael Trindade Maia

Universidade Federal de Campina Grande
Sumé – Paraíba rafael.rafatrin@gmail.com
Autor para correspondência

RESUMO: A febre Zika tornou-se uma das maiores preocupações na saúde pública mundial e do Brasil, principalmente depois da recente descoberta de sua associação com casos de microcefalia. Até o momento não existe vacina disponível para o ZIKV e o tratamento é sintomático. Em uma recente pesquisa foi elucidada a estrutura tridimensional da ZIKV NS5 RdRp, uma proteína associada à RNA polimerase e importante ativadora do processo de replicação do vírus, com esta informação, veio a possibilidade de tornar esta enzima um alvo de novos fármacos. Neste âmbito, estudos teórico-computacionais são de grande valia para sinalizar possíveis inibidores desta enzima que

funcionem com antivirais, logo, o artigo teve como objetivo a investigação da eficácia de fármacos já descritos na inibição da NS5 RdRp do vírus da Zika. A metodologia seguiu os seguintes passos: Análise da estrutura tridimensional da NS5, bem como seu sítio ativo; busca na literatura e nos bancos de dados de moléculas por possíveis inibidores; simulações de *docking* molecular e investigação de interações nos complexos proteína-ligantes e/ou proteína-proteína gerados. Buscou-se destacar quatro parâmetros considerados mais importantes na avaliação para o melhor complexo de cada *docking*: A Constante de Inibição Estimada (CIE), a Energia Intermolecular Final (EIF) e a Energia Livre de Ligação (ELL) estimada pelo programa. Os fármacos testados foram os antivirais: Acyclovir, Indinavir e Valgacyclovir. A pesquisa obteve êxito quanto o objetivo proposto, tendo sido capaz de identificar valores de interação entre os fármacos testados e a proteína responsável pela replicação viral do ZIKV NS5 RdRp.

PALAVRAS-CHAVE: Saúde Pública, Arbovírus, Docking Molecular.

ABSTRACT: Zika fever has become one of the biggest concerns in world public health and Brazil, especially after the recent discovery of its association with cases of microcephaly. Until now, there is no vaccine available for

ZIKV, and the treatment is symptomatic. In a recent research was solved the three-dimensional structure of ZIKV NS RdRp X-Ray crystallography, a protein associated with RNA polymerase and important activator of the virus replication process, with this information, occurred the possibility of making this enzyme a target for new drugs. In this context, theoretical-computational studies are of great value to indicate possible inhibitors for this enzyme that work as antiviral, so the article aimed to investigate the efficacy of drugs already described in the inhibition of NS5 RdRp of the Zika virus. The methodology adopted the following steps: analysis of the three-dimensional structure of the NS5 and its active site; search by potential inhibitors in the literature and databases of molecules; perform molecular docking simulations and analysis of interactions in protein-ligands and/or protein-protein complexes generated. Four parameters, considered to be important in the evaluation for the best complex of each docking, were used: The estimated inhibition constant (EIC), the Final Intermolecular Energy (FIE) and the Free Binding Energy (FBE) estimated by the program. The drugs tested were the antivirais: Acyclovir, Indinavir and Valgacyclovir. The research was successful in terms of the proposed objective and was able to identify interaction values between the drugs tested and the protein responsible for viral replication of the ZIKV NS5 RdRp.

KEYWORDS: Public health, Arbovirus, Molecular docking.

1 | INTRODUÇÃO

No Brasil e no mundo, doenças relacionadas com vetores transmissores de epidemias ainda preocupam e são responsáveis por um alto número de mortalidade e mobilidade humana. O agravante disto é antecedido por alguns fatores, como: ciclo de vida do mosquito, condições ambientais e sanitárias onde estes se desenvolvem e hospedeiros e/ou reservatórios que são diretamente responsáveis por encabeçar a cadeia de transmissão das doenças (CARVALHO & SOUZA, 2016). Focando no Brasil, infelizmente o país ainda vive uma época de carência em termos de saneamento básico, dez anos após a incisão da lei de saneamento básico (Lei 11.445/2007) tem-se 49,07% da população brasileira sem acesso a coleta de esgoto e 57,3% sem esgoto tratado (G1, 2017), um número considerável e agravante em termos de proliferação de doenças virais.

O Zika Vírus (ZIKV) é um arbovírus transmitido pela picada dos mosquitos *Aedes aegypti* e *Aedes albopictus*, e recentes evidências apontam também o mosquito *Culex quinquesfasciatus* como vetor. Também pode ser transmitida por relações sexuais e transfusão sanguínea, mesmo com parceiros e doadores assintomáticos. Os meios de transmissão saliva, urina ou leite materno ainda não foram confirmados. O primeiro registro desta doença foi em 1947 na Floresta Zika, em Uganda, detectado inicialmente em macacos. Em 1954, na Nigéria, foram registrados os primeiros casos humanos (Zanluca et al., 2015).

A confirmação de transmissão da febre causada pelo Zika vírus surgiu no Brasil em meados de abril de 2015, apresentou-se inicialmente nos estados do Rio Grande

do Norte e da Bahia na região Nordeste do país. A semana Epidemiológica (SE) alertou no mesmo ano sobre a sinalização de 19 unidades da Federação sobre a confirmação laboratorial da doença, porém não foram divulgados dados quantitativos. Já em 2016 a SE divulgou não só as unidades (13) como também os dados quantitativos que se mostraram alarmantes: 35.505 casos da doença na região Sudeste, seguido pelo Nordeste, 30.286 casos e no geral foram 91.387 casos de Zika no Brasil (CARVALHO & SOUZA, 2016).

A epidemia causada pelo ZIKV começa a se propagar entre 3 a 12 dias após a picada e tem como sintoma marcante a febre baixa (entre 37,8° e 38,5°C), que vem acompanhada se sintomas gerais, dores articulares, conjuntivite, erupções cutâneas com coceiras, dor muscular e dor de cabeça. Apesar dos sintomas parecem ser triviais surgiram casos mais severos no Brasil e Polinésia Francesa ligados a microcefalia, uma doença neurológica congênita (síndrome de Guillain-Barré) tem sido catalogada. Dentro desse contexto, com o aumento exponencial de casos surgiu a preocupação dos órgãos públicos em respeito à saúde mundial, além de imprescindibilidade em vigiar síndromes neurológicas em pacientes febris agudos, também, existe a necessidade de mais estudos relacionados ao Vírus da zika e pesquisas voltadas a cessar o culminio de seu efeito sobre a população (ZANLUCA, 2015).

Até o momento não existe vacina disponível para o ZIKV e o tratamento é sintomático, ou seja, não há tratamento específico para a doença, apenas medicamentos para amenizar os sintomas. No entanto, diversos estudos focam em desenvolver vacinas e medicamentos específicos. Em uma recente pesquisa foi elucidada a estrutura tridimensional da ZIKV NS5 RdRp, uma proteína associada à RNA polimerase e importante ativadora do processo de replicação (GODOY, A. et al. 2017).

A estrutura tridimensional de uma proteína alvo, preferivelmente com um ligante para a possível elucidção do seu sitio receptor é o segredo para o desenho de fármacos com base na estrutura, uma vez definida a estrutura do sitio receptor pode ser utilizada para determinar um farmacóforo para a triagem virtual em bibliotecas de compostos e em pesquisas de docking molecular que podem ser usados para aclarar e aperfeiçoar a estrutura dos compostos líderes (Willis, 2002).

Com a elucidção da estrutura da ZIKV NS5 RdRp, proteína associada à replicação viral, surgem possibilidades de tornar esta enzima um alvo de novos fármacos, intentando a busca de inibidores e experimentos *in silico*, como o docking molecular (atracamento), podem ser extremamente úteis para a sinalização de possíveis moléculas com potencial para inibição da proteína NS5, logo, este trabalho teve como objetivo selecionar *in silico* possíveis inibidores da proteína de replicação NS5.

2 | MÉTODOS/METODOLOGIA

2.1 Estrutura da ZIKV NS5 RdRp

As coordenadas atômicas da ZIKV NS5 RdRp foram obtidas no banco de dados *PDB-Protein Data Bank* através do código de acesso 5U04 (Godoy et al. 2017). A estrutura (Figura 1) foi analisada visualmente pelo *software VMD-Visual Molecule Dynamics* (Humphrey et al., 1996).



Figura 1 – Visualização do ZIKV NS5 RdRp

Fonte – Autores.

2.2 Seleção Das Moléculas

Foi realizada uma revisão de literatura, com a perspectiva de melhor desenvolvimento de pesquisa e elucidar possível moléculas ligantes que possam vir a mostrar bons resultados computacionais quanto a sua inibição a NS5. A pesquisa se deu nos sites de busca acadêmica Google acadêmico onde foi possível visitar banco de informações como SciELO, PubMed, Lilacs, Medline entre outras plataformas.

Com relação a estrutura dos ligantes foram feitas buscas na literatura e em bancos de dados de pequenas moléculas, como o *ZINC Database*, *Pubchem* e também o banco do *Chempider* (h. Foram gerados arquivos na extensão *.pdb* e as estruturas dools compostos a serem testados forma analisadas visualmente pelo *VMD-Visual Molecule Dynamics* e no *Discovery Studio*.

2.3 Docking molecular

Os experimentos de docking proteína-ligante ZIKV NS5 RdRp +Ligante foram realizados no programa Autodock1.5.6 (Morris et al., 2009), no qual ocorreu um pré-

dimensionamento da grade cúbica de energia de volume 126 x 126 x 126 Å, que foi centrada no receptor. Utilizou-se o algoritmo genético Lamarckiano (LGA). Dez conformações ranqueadas por energias foram geradas para cada complexo proteína-ligante, após organizadas, coletadas fez-se a análise através da opção Autodock tools. As conformações obtidas organizadas, coletadas e analisadas através da opção Autodock tools. Os melhores complexos de cada simulação foram selecionados para análise visual no VMD (Humphrey et al., 1996).

2.4 Compostos Testados

As análises foram feitas a partir de fármacos já conhecidos no mercado pela sua aplicação antiviral, porém nenhum deles possui trabalhos na literatura sobre esta função aplicada diretamente ao Zika vírus.

2.5 Acyclovir

É análogo acíclico da desoxiguanosina. Seu mecanismo de ação está esclarecido. Pela ação catalítica das timidino-quinases dos vírus *Herpes simplex* e *Varicella zoster*, as quais o fármaco se liga 200 vezes mais firmemente que a enzima celular, o aciclovir é convertido ao monofosfato de aciclovir, um nucleotídeo, que é convertido ao difosfato pela guanilato-quinase celular e ao trifosfato pela guanilato-quinase celular e ao trifosfato por várias enzimas celulares. O trifosfato interfere com a DNA polimerase dos vírus citados e inibe a replicação do DNA viral, o que acarreta a morte dos vírus. Esta ação é seletiva; para inibir o crescimento das células do hospedeiro seria necessária concentração 3.000 vezes maior do que a atingida do vírus (Anvisa, 2015). A molécula (Figura 2) pode ser obtida através de <http://zinc.docking.org/substance/1530555>.

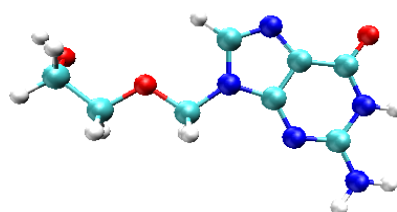


Figura 2 – Visualização do Acyclovir no software VMD-Visual Molecule Dynamics.

Fonte – Autores.

2.6 Valgacyclovir

É um pró-fármaco do ganciclovir (Figura 3) resultante da mistura de dois diastereômeros. Estes são convertidos ao ganciclovir através de esterases hepáticas e intestinais. Sua ação antiviral resulta da inibição da síntese do DNA viral mediada pelo trifosfato de ganciclovir. Comercializado como cloridrato (Anvisa, 2016). A molécula pode ser obtida através de <http://zinc.docking.org/substance/1543916>.

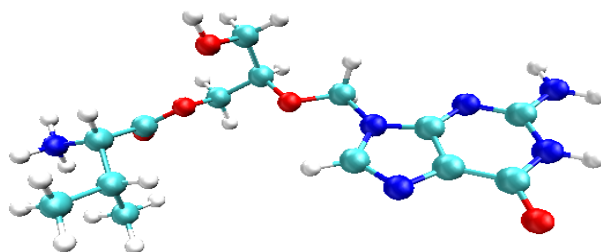


Figura 3 – Visualização do Valganciclovir no software *VMD-Visual Molecule Dynamics*.

Fonte – Autores.

2.7 Indinavir

Pertencente a categoria dos inibidores de protease, ao contrário zidovudina, didanosina, zalcitabina e estavudina, que são inibidores da transcriptase reversa (Anvisa, 2016). O sulfato de indinavir possui a seguinte estrutura química, contendo cinco centros quirais (Figura 4). A molécula pode ser obtida através de <http://zinc.docking.org/substance/22448696>

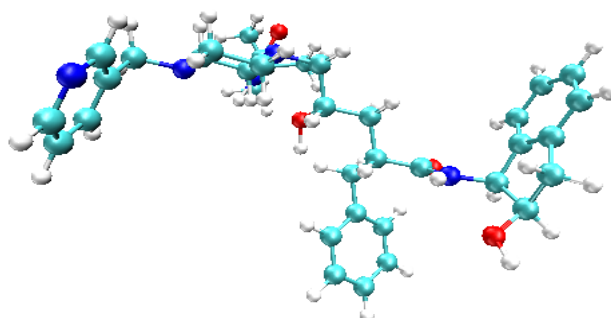


Figura 4 – Visualização da Indinavir no software *VMD-Visual Molecule Dynamics*.

Fonte – Autores.

3 | RESULTADOS E DISCUSSÃO

As energias obtidas nos cálculos de *Docking* foram calculadas a partir de grades localizadas no sítio ativo da proteína, a escolha do algoritmo lamarkiano foi fundamental para que os ligantes buscassem conformações que fossem favoráveis para seu atracamento. No programa Autodock4 a proteína foi previamente considerada rígida em todas as simulações de *Docking*, enquanto os ligantes flexíveis, ao fim, foram obtidos dez conformações de cada interação proteína-ligante testados neste trabalho. Destacou-se três parâmetros que foram considerados mais importantes para a busca do objetivo proposto, estes foram: A Constante de Inibição Estimada (mM), A Energia intermolecular Final e a Energia livre de ligação (kcal/mol) estimada pelo programa. As tabelas a seguir irão expor o resultado destes parâmetros.

3.1 Acyclovir

Os resultados de *Docking* Molecular do Acyclovir com a NS5 retratam o quanto os parâmetros podem variar de acordo com a conformação/complexo utilizado. A Constante de inibição expressa se mostrou bastante elevada em todos os complexos, variando de 393.06 mM a 997.97mM. Com relação a Energia Livre de Ligação e Energia Intermolecular Final, em praticamente todos os complexos, exceto no complexo 4 (onde o programa não apresentou valores para Ki), as energias foram negativas, evidenciando através do Docking Molecular, que a interação proteína-ligante utiliza menos energia em sua conformação conjunta do que separadas. Na tabela 1, é apresentado os resultados dos cálculos de docking do Acyclovir a partir das estruturas obtidas nesta pesquisa. Na tabela 1, é apresentado os resultados dos cálculos de docking do Acyclovir a partir das estruturas obtidas nesta pesquisa.

Complexo	Energia Livre de Ligação	Energia Intermolecular	Ki
1	-0.07	-1.86	888.17
2	-0.17	-1.96	749.35
3	-0.37	-2.16	535.47
4	+0.13	-1.66	-
5	-0.37	-2.16	539.31
6	-0.35	-2.14	554.31
7	-0.11	-1.90	826.13
8	-0.11	-1.90	833.49
9	-0.01	-1.80	997.97
10	-0.55	-2.34	393.06

Tabela 1 - KI ou Constante de inibição expressa (CIE) em mM, Energia estimada livre (EEL) em kcal/mol e energia intermolecular final (EIF) estimada pelo programa AutoDock4.

Fonte – Autores.

Após a realização dos cálculos de docking do Acyclovir, utilizando o programa Discovery Studio, foi possível identificar as ligações ocorridas entre os aminoácidos de maior proximidade, apresentando ligações de hidrogênio com os aminoácidos Ser703, Asn704 e Glu707. Como pode-se ver na Figura 5.

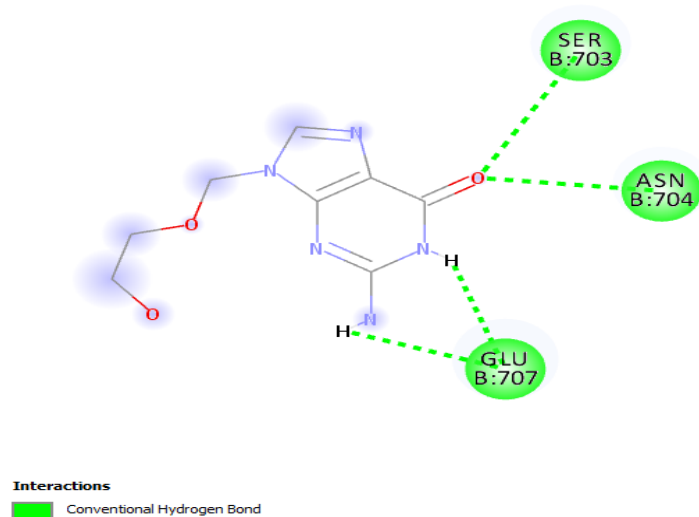


Figura 5 – Visualização 2D do composto Acyclovir e os resíduos da NS5 de maior interação.

Fonte – Autores.

Ainda no uso do software Discovery Studio (BIOVIA DISCOVERY STUDIO VISUALIZER, 2014) foi possível analisar a nuvem eletrostática do receptor (NS5), onde é possível notar que o local de atracamento está situado em uma região de interação neutra. Próximo ao local de encontro, também é possível notar uma região muito eletronegativa, um dos aminoácidos responsáveis é o Glu 707. Essas afirmações podem ser observadas na Figura 6.

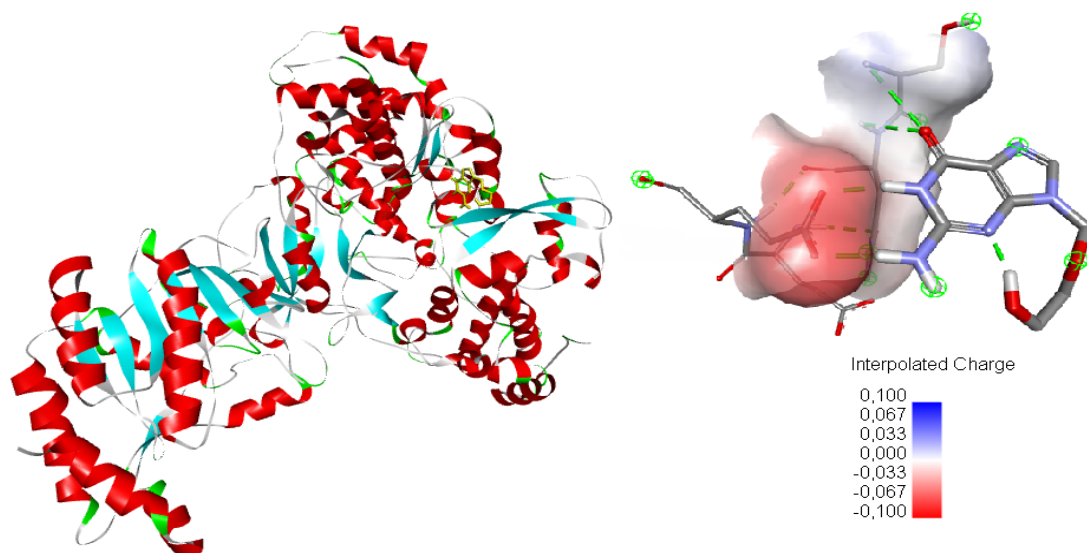


Figura 6 – Proteína NS5 RdRp com fármaco Acyclovir (em amarelo) e a nuvem eletrostática da interação receptor com o acyclovir.

Fonte – Autores.

3.2 Indinavir

O Docking Molecular do fármaco indinavir com a Proteína NS5 RdRp apresentou altas constantes de inibição, variando de 8.13mM a 652.40mM. A Energia Livre de ligação foi negativa em basicamente todos os complexos, exceto o complexo 8. Enquanto a Energia Intermolecular Final apresentou resultados expressivamente

negativos. Logo, os parâmetros evidenciam uma boa interação ligante-proteína entre as moléculas propostas (Tabela 2).

Complexo	Energia Livre de Ligação	Energia Intermolecular	Ki
1	-1.27	-5.45	117.02
2	-0.60	-4.78	362.50
3	-0.25	-4.43	651.89
4	-2.47	-6.65	15.46
5	-1.69	-5.86	57.98
6	-0.98	-5.16	189.90
7	-2.85	-7.03	8.13
8	+2.85	-3.81	-
9	+0.36	-4.43	652.40
10	-0.64	-4.82	336.90

Tabela 2 - KI ou Constante de inibição estimada (CIE) em mM, Energia estimada livre (EEL) em kcal/mol e energia intermolecular final (EIF) estimada pelo programa AutoDock4.

Fonte – Autores.

Após a realização dos cálculos de docking do Indinavir, utilizando Discovery Studio, foi possível identificar as ligações ocorridas entre os aminoácidos de maior proximidade, apresentando ligação de hidrogênio com os aminoácidos Trp702, ligações Pi com Asp534, Thr 700, Cys 849, Val 708, Pro 709. Como pode-se ver na Figura 7.

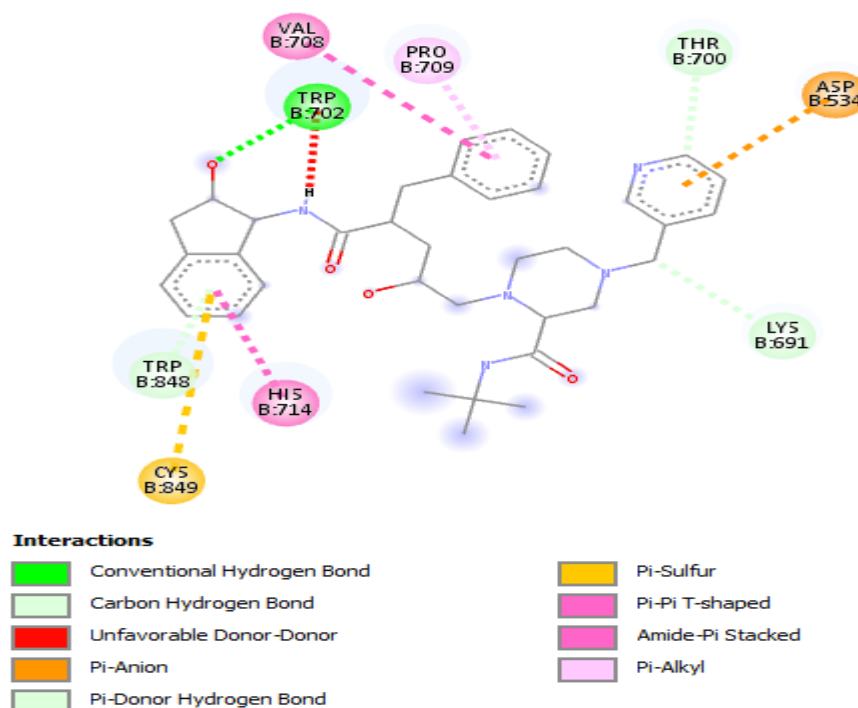


Figura 7 – Visualização 2D do composto Indinavir e os resíduos da NS5 de maior interação.

Fonte – Autores.

Utilizando o Discovery Studio foi possível analisar a nuvem eletrostática do receptor, onde é possível notar que o local de atracamento está situado em uma região de interação neutra. As regiões mais próximas não apresentam picos eletrostáticos.

Essas afirmações podem ser observadas na Figura 8.

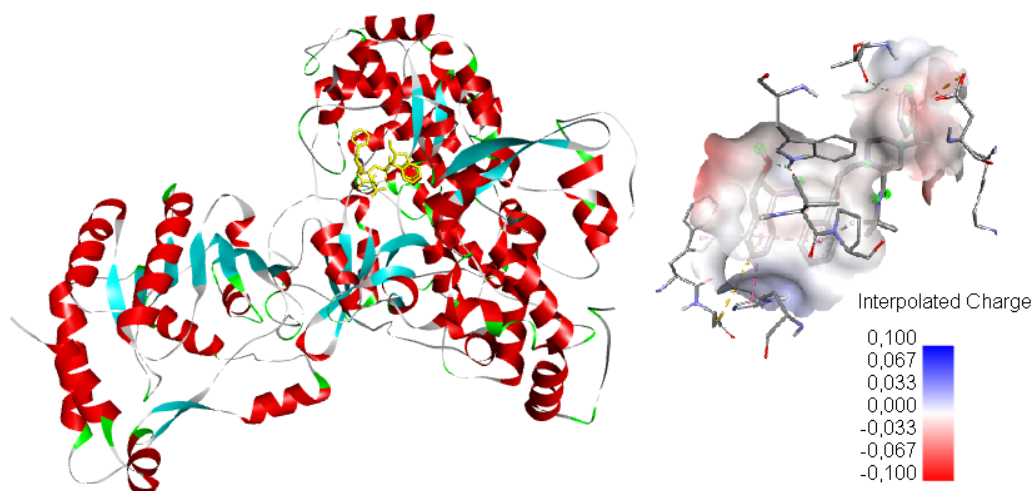


Figura 8 - Proteína NS5 RdRp com fármaco Indinavir (em amarelo) e a nuvem eletrostática da interação receptor com o Indinavir.

Fonte – Autores.

3.3 Valgacyclovir

O Valgacyclovir demonstrou resultados interessantes. Nos oito primeiros complexos o programa não apontou valores para constantes de inibição e nestes mesmos complexos, a Energia Livre de Ligação se mostrou positiva. Entretanto, A Energia Intermolecular Final de todos os complexos se apresentaram negativas, enquanto, as constantes de inibição apontada entre os complexos nove e dez variaram entre 100.23mM a 467.82mM (Tabela 3).

Complexo	Energia Livre de Ligação	Energia Intermolecular	Ki
1	+0.78	-2.80	-
2	+0.85	-2.75	-
3	+0.60	-2.98	-
4	+1.47	-2.11	-
5	+0.99	-2.59	-
6	+0.13	-3.45	-
7	+2.00	-1.58	-
8	+1.99	-1.59	-
9	-1.36	-4.94	100.23
10	-0.45	-4.03	467.82

Tabela 3 – KI ou Constante de inibição expressa (CIE) em mM, Energia estimada livre (EEL) em kcal/mol e energia intermolecular final (EIF) estimada pelo programa AutoDock4.

Fonte – Autores.

Após a realização dos cálculos de docking do Valgacyclovir, utilizando Discovery Studio, foi possível identificar as ligações ocorridas entre os aminoácidos de maior proximidade, apresentando ligação de hidrogênio com os aminoácidos Ala 533 e Asp 692, Atração eletrostática com Asp 535, Asp 534. Como pode-se ver na Figura 9.

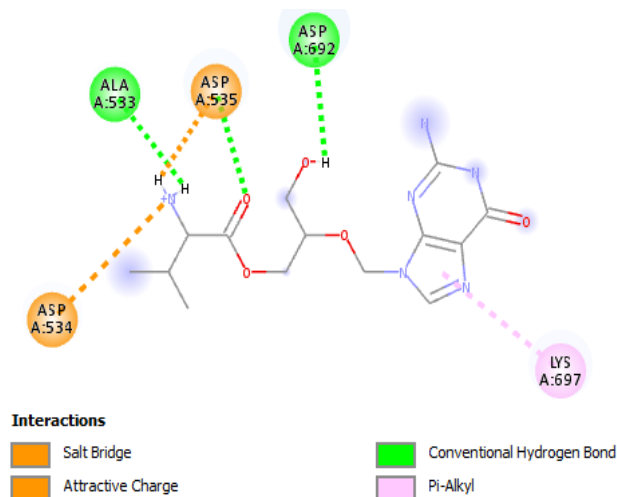


Figura 9– Visualização 2D do composto Valgacyclovir e os resíduos da NS5 de maior interação.

Fonte – Autores.

Utilizando o Discovery Studio foi possível analisar a nuvem eletrostática do receptor, onde é possível notar que o local de atracamento está situado em uma região dentro da nuvem. As regiões mais próximas apresentam picos eletrostáticos. Essas afirmações podem ser observadas na Figura 10.

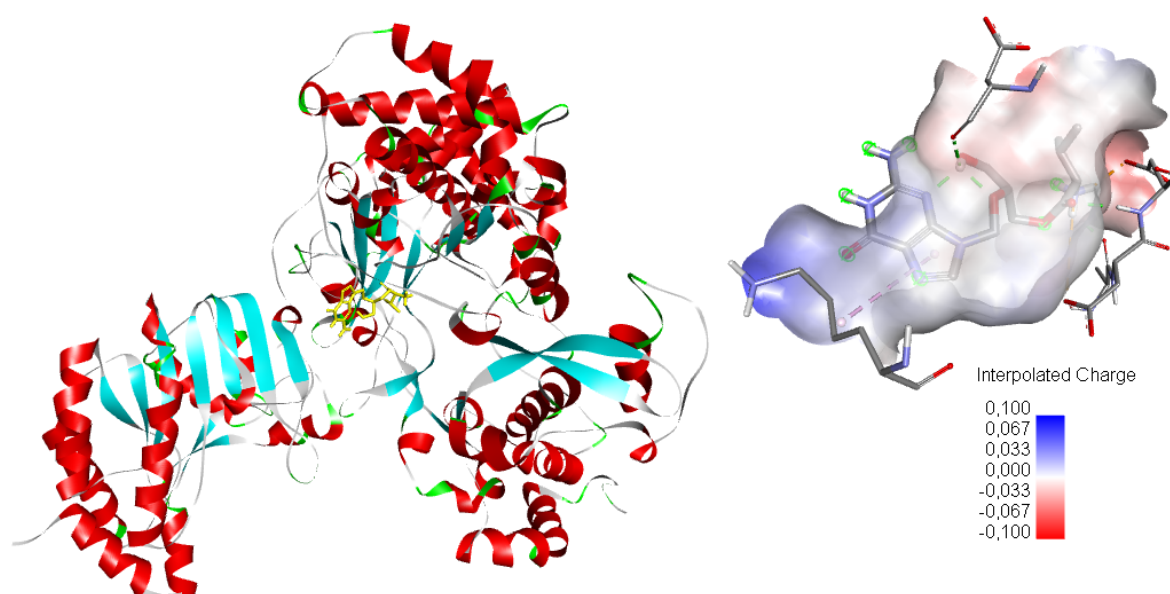


Figura 10 - Proteína NS5 RdRp com fármaco Valgacyclovir (em amarelo) e a nuvem eletrostática da interação receptor com o Valgacyclovir.

Fonte – Autores.

4 | CONSIDERAÇÕES FINAIS

Os estudos *in silico* trouxeram consigo ferramentas computacionais altamente eficientes para a resolução de problemas farmacológicos, dentre essas ferramentas, é possível destacar o Docking molecular, que para a realização deste trabalho vem sendo indispensável, por unir conceitos matemáticos físico-químicos, termodinâmicos

e químicos propriamente ditos.

Os três resultados em destaque mostraram uma interação direta com o sítio ativo da proteína por meio de ligações covalentes e pontes de Hidrogênio, além de excelentes constantes de inibição e suas energias (intermolecular final e livre de ligação) apontarem uma alta taxa de negatividade o que indica menor perda de energia na união dessas moléculas. Os três compostos em questão demonstraram evidências na inibição da NS5 RdRp do ZIKV.

REFERÊNCIAS

Anvisa. Pharlab Indústria Farmacêutica. 2015. **Aciclovir**. Disponível em: http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=9641322015&pIdAnexo=2929512. Acesso em: Jun de 2018.

Anvisa. Produtos Roche Químicos e Farmacêuticos S.A. 2016. Disponível em: http://www.anvisa.gov.br/datavisa/fila_bula/frmVisualizarBula.asp?pNuTransacao=27709202016&pIdAnexo=4202882. Acesso em: Jun de 2018.

BIOVIA DISCOVERY STUDIO VISUALIZER. 2014. Disponível em: <http://accelrys.com/products/datasheets/discovery-studio-visualizer.pdf>. Acesso em: jul de 2018.

CARVALHO, C. D. S; SOUZA, Z.H. **Reflexão Acerca Da Incidência Dos Casos De Dengue, Chikungunya E Zika No Brasil**. I Colóquio Estadual De Pesquisa Multidisciplinar, 2016, Unifimes – Centro Universitário De Mineiros – Goiás, 2016.

G1. **Saneamento Melhora, Mas Metade Dos Brasileiros Segue Sem Esgoto No País**. São Paulo, 2017. Disponível. Em: <https://G1.Globo.Com/Economia/Noticia/Saneamento-Melhora-Mas-Metade-Dos-Brasileiros-Segue-Sem-Esgoto-No-Pais.Ghtml> .

GODOY, A. et al. **Crystal structure of Zika virus NS5 RNA-dependent RNA polymerase**. *Nature Communications*. VL - 8, 2017.

HUMPHREY, W. et al. **“VMD - Visual Molecular Dynamics”**. *J. Molec. Graphics*, vol. 14, p. 33-38. 1996.

WILLIS, R.C., (2002), **“Surveying the binding site”**, *Modern Drug Discovery*, September, pages 28-34

ZANLUCA C. et al. **First report of autochthonous transmission of Zika virus in Brazil**. *Mem Inst Oswaldo Cruz*, 110:569–72, 2015.

SOBRE O ORGANIZADOR

ERNANE ROSA MARTINS Doutorado em andamento em Ciência da Informação com ênfase em Sistemas, Tecnologias e Gestão da Informação, na Universidade Fernando Pessoa, em Porto/Portugal. Mestre em Engenharia de Produção e Sistemas pela PUC-Goiás, possui Pós-Graduação em Tecnologia em Gestão da Informação pela Anhanguera, Graduação em Ciência da Computação pela Anhanguera e Graduação em Sistemas de Informação pela Uni Evangélica. Atualmente é Professor de Informática do Instituto Federal de Educação, Ciência e Tecnologia de Goiás - IFG (Câmpus Luziânia), ministrando disciplinas nas áreas de Engenharia de Software, Desenvolvimento de Sistemas, Linguagens de Programação, Banco de Dados e Gestão em Tecnologia da Informação. Pesquisador do Núcleo de Inovação, Tecnologia e Educação (NITE).

Agência Brasileira do ISBN
ISBN 978-85-7247-113-8

