

Análise *in silico* para detecção de mutações na região SRS29B de  
*Toxoplasma gondii* com implicações na sensibilidade de  
imunoensaios

Gabriel Nogueira Araujo

Campos dos Goytacazes, RJ

Agosto de 2022

Análise *in silico* para detecção de mutações na região SRS29B de  
*Toxoplasma gondii* com implicações na sensibilidade de  
imunoensaios

Gabriel Nogueira Araujo

Dissertação apresentada ao Centro de Biociências e Biotecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para a obtenção do título de Mestre em Biociências e Biotecnologia.

Orientador: Dr<sup>a</sup> Alba Lucínia Peixoto Rangel

Campos dos Goytacazes, RJ

Agosto de 2022

Análise *in silico* para detecção de mutações na região SRS29B de *Toxoplasma gondii* com implicações na sensibilidade de imunoenaios

Gabriel Nogueira Araujo

Dissertação apresentada ao Centro de Biociências e Biotecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro, como parte das exigências para a obtenção do título de Mestre em Biociências e Biotecnologia.

Trabalho aprovado. Campos dos Goytacazes, RJ, 30 de agosto de 2022:

---

**Dra. Andrea Cristina Vetö Arnholdt**  
LBR/CBB/UENF

---

**Dr. Jorge Hernandez Fernandez**  
LQFPP/CBB/UENF

---

**Dr. Leandro Martínez**  
IQM/UNICAMP

---

**Dra. Drª Alba Lucinia Peixoto Rangel**  
LBR/CBB/UENF  
Orientadora

Campos dos Goytacazes, RJ  
Agosto de 2022

# Agradecimentos

Agradeço aos parceiros do Laboratório de Biologia do Reconhecer pelo direcionamento, compartilhamento da experiência na área e pelo desenvolvimento de pesquisas que serviram de alicerce para o desenvolvimento deste trabalho. À minha orientadora Dra. Alba, que indicou o tema da pesquisa e acreditou no potencial da metodologia proposta que foge um pouco do que é normalmente realizado pela equipe. Agradeço à instituição UENF e aos professores e os administradores do PGBB, que organizaram as aulas remotas com sucesso no período de crise sanitária e humanitária que vivemos durante o período do curso. Agradeço aos pesquisadores que se dedicaram aos estudos parasitológicos sobre *T. gondii*, reunindo e divulgando os dados estatísticos e laboratoriais que possibilitaram a realização deste trabalho. Agradeço ao CENAPAD/SP e à equipe responsável, da UNICAMP, pela plataforma de computação de alto desempenho disponibilizada que foi fundamental ao desenvolvimento desta pesquisa. Agradeço aos membros da banca pela disponibilidade e interesse. Agradeço à CAPES pelo apoio financeiro durante o desenvolvimento deste trabalho.



*“ when, clutching our crystals and nervously consulting our horoscopes, our critical faculties in decline, unable to distinguish between what feels good and what’s true, we slide, almost without noticing, back into superstition and darkness. ”*

*(Carl Sagan)*

## Resumo

*Toxoplasma gondii* é um parasito apicomplexo, amplamente distribuído e de grande importância médica. O Brasil apresenta grande prevalência de cepas não-clonais de *T. gondii*. Essa característica populacional é observada com maior intensidade em certos municípios do país, como Campos dos Goytacazes, e diverge da dominância de perfis clonais deste parasita encontrados na Europa e América do Norte. A infecção por *T. gondii* pode ser diagnosticada indiretamente, através de métodos sorológicos, sendo o ensaio de ELISA (Enzyme Linked Immuno Sorbent Assay) o método sorológico mais comumente utilizado. Trabalhos anteriores do nosso grupo relatam amostras de indivíduos que apresentavam lesões oculares com morfologia característica de toxoplasmose ocular, porém com sorologia não reagente para a infecção por *T. gondii*. Adicionalmente, amostras de indivíduos, que mesmo sem sinais clínicos da doença e sorologia não reagente, apresentaram material genético do parasita. Esses dados podem indicar uma baixa sensibilidade no diagnóstico sorológico por meio de uso de kits comerciais ou ELISA in house que não contemplam linhagens polimórficas de *T. gondii* em regiões altamente endêmicas e com grande diversidade genética deste parasita, como no caso de Campos dos Goytacazes. A proteína SRS29B é uma das mais utilizadas como antígeno de sensibilização em imunodiagnósticos comerciais. Neste sentido, o presente trabalho buscou determinar se substituições de aminoácidos na proteína de *T. gondii* SRS29B poderiam estar presentes em variantes de *T. gondii* endêmicas em Campos dos Goytacazes, comprometendo a eficácia de kits comerciais usualmente utilizados para diagnóstico sorológico de infecções por *T. gondii*, em amostras de indivíduos infectados por cepas polimórficas deste parasita. Para tanto, foram aplicados métodos *in silico* como a comparação das trocas de aminoácidos na sequência protéica, predição de epítomos lineares, e modificação de estruturas protéicas de referência através de substituições observáveis e não observáveis na natureza, com subsequente análise de dinâmica molecular para mensuração dos potenciais efeitos de substituições na imunocompatibilidade de SRS29B. Os experimentos *in silico* realizados neste trabalho utilizaram sequências de SRS29B disponíveis publicamente e não puderam apontar uma determinada substituição de aminoácido potencialmente responsável por alterações em epítomos lineares ou na estrutura da proteína. A pipeline desenvolvida neste trabalho pode analisar sequências de isolados de *T. gondii* endêmicos de Campos dos Goytacazes que podem estar disponíveis no futuro.

**Palavras-chave:** *Toxoplasma gondii*, toxoplasmose, imunodiagnóstico, srs29b, dinâmica molecular

## Abstract

*Toxoplasma gondii* is a globally distributed apicomplexan parasite with medical relevance. Brazil shows a large prevalence of non-clonal *T. gondii* strains. This populational feature is even more intense in certain geographical regions of the country, such as Campos dos Goytacazes. Meanwhile, most of the strains found in North America and Europe are clonal. Infections caused by *T. gondii* can be diagnosed indirectly by using serological methods. ELISA (Enzyme Linked Immuno Sorbent Assay) is the most used of such methods. Previous studies done by our group reported blood samples with nonreactive serology against *T. gondii*, but collected from individuals that presented ocular lesions with morphology associated with ocular toxoplasmosis. Additionally, samples from individuals that had no clinical manifestations of the disease and non reactive serology were found to contain DNA from the parasite. This data can indicate a low sensitivity in commercial kits for immunoassays or in house ELISA that do not contemplate highly polymorphic *T. gondii* lineages which are present in regions such as Campos dos Goytacazes. SRS29B is commonly used as sensitization antigen in commercial immunodiagnosis kits. In this sense, this research tried to determine if aminoacid substitutions in SRS29B could be present in *T. gondii* variants that are endemic in Campos dos Goytacazes and compromise the efficacy of commercial immunoassay kits on patients infected by such variants. *In silico* methods applied include comparison of aminoacid substitution in protein sequences and linear epitope prediction. Finally, aminoacid replacement in reference protein structures using natural and artificial substitutions followed by molecular dynamics simulations will help to determine the potential impact of tested substitutions to SRS29B's immunocompatibility. The *in silico* experiments carried in this project used publicly available SRS29B sequences from various isolates as input and were unable to identify a particular aminoacid substitution that is likely responsible for linear epitope or protein structure changes. The pipeline that was developed for this project can analyze sequencing data of *T. gondii* variants that are endemic in Campos dos Goytacazes which may be available in the future.

**Keywords:** *Toxoplasma gondii*, toxoplasmosis, immunodiagnosis, srs29b, molecular dynamics

# Lista de ilustrações

Figura 1 – Anticorpos . . . . .	15
Figura 2 – Interação molecular entre SRS29B e um anticorpo monoclonal. . . . .	16
Figura 3 – Variação da concentração sorológica de anticorpos ao longo do tempo. . . . .	20
Figura 4 – Imunoensaio ELISA . . . . .	22
Figura 5 – Métricas de sensibilidade e especificidade para imunoensaios . . . . .	23
Figura 6 – Mecanismo de como o <i>antigenic drift</i> afeta a sensibilidade de imunoensaios. . . . .	25
Figura 7 – Distribuição geográfica de genótipos de <i>T. gondii</i> . . . . .	27
Figura 8 – Alinhamento de uma região upstream ao gene SRS29B. . . . .	33
Figura 9 – Resumo da parte estrutural da metodologia. . . . .	37
Figura 10 – Estrutura de SRS29B conforme a cadeia F de 1YNT. . . . .	38
Figura 12 – Trocas de aminoácido observadas na proteína SRS29B de <i>T. gondii</i> . . . . .	42
Figura 11 – Alinhamento das sequências naturais de SRS29B com as sequências descritas em patentes. . . . .	43
Figura 13 – Predição de Epítomos lineares em SRS29B . . . . .	44
Figura 14 – RMSD das simulações moleculares do grupo de estruturas de SRS29B. . . . .	48
Figura 15 – RMSF das simulações moleculares do grupo de estruturas de SRS29B. . . . .	49
Figura 16 – Ângulo entre os domínios das simulações moleculares do grupo <i>natural</i> de estruturas de SRS29B. . . . .	50
Figura 17 – SASA das simulações moleculares do grupo de estruturas de SRS29B. . . . .	51
Figura 18 – Número de resíduos em estruturas secundárias das simulações moleculares do grupo de estruturas de SRS29B. . . . .	52
Figura 19 – Estruturas alinhadas das simulações moleculares do grupo <i>natural</i> de estruturas de SRS29B. . . . .	53
Figura 20 – RMSD das simulações moleculares do grupo de estruturas de SRS29B. . . . .	54
Figura 21 – Ângulo entre os domínios das simulações moleculares do grupo <i>natural</i> de estruturas de SRS29B. . . . .	54
Figura 22 – Gráficos de Ramachandran para as trajetórias das variantes naturais de SRS29B. . . . .	55
Figura 23 – Prevalência de IgG para toxoplasmose e proporção de isolados atípicos em alguns países. . . . .	59

# Lista de tabelas

Tabela 1	– Genotipagem PCR-RFLP de isolados locais de <i>T. gondii</i> . . . . .	30
Tabela 2	– Genotipagem PCR-RFLP de isolados de <i>T. gondii</i> de interesse. . . . .	31
Tabela 3	– Bancos de dados consultados durante o projeto. . . . .	36
Tabela 4	– Parâmetros utilizados na dinâmica molecular executada no GROMACS. . .	40
Tabela 5	– Patentes que detalham antígenos de sensibilização para diagnóstico de toxoplasmosse. . . . .	41
Tabela 7	– Estruturas obtidas através de substituições de aminoácidos em SRS29B. . .	45
Tabela 8	– Ocorrência das variantes observadas em SRS29B nos isolados analisados. .	46
Tabela 9	– Efeitos de substituições individuais de aminoácidos na estabilidade da proteína, de acordo com algoritmos de predição. . . . .	47

# Lista de abreviaturas e siglas

AS	Antígeno de Sensibilização
CDS	Coding Sequence - Parte Codificante da Sequência de um Gene
SASA	Solvent Accessible Surface Area - Área de Superfície Acessível ao Solvente
TO	Toxoplasmose Ocular
Fc	Fragment-crystallizable
HIV	Human Immunodeficiency Virus - Vírus da imunodeficiência humana
SSR	Simple sequence repeat - Repetição de sequência simples
CNV	Copy number variation - Variação do número de cópias
SASA	Solvent accessible surface area - Área de superfície acessível ao solvente
RMSD	Root-mean-square deviation (of atomic positions)
RMSF	Root-mean-square fluctuation

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>12</b>
<b>2</b>	<b>IMUNOGLOBULINAS</b>	<b>14</b>
<b>2.1</b>	<b>Estrutura das imunoglobulinas</b>	<b>14</b>
<b>2.2</b>	<b>Antígenos</b>	<b>15</b>
2.2.1	Epítomos	16
<b>2.3</b>	<b>Interação entre imunoglobulina e antígeno</b>	<b>17</b>
<b>2.4</b>	<b>Especificidade de imunoglobulinas</b>	<b>18</b>
<b>2.5</b>	<b>Atividade imunológica</b>	<b>18</b>
<b>3</b>	<b>IMUNODIAGNÓSTICOS</b>	<b>20</b>
<b>3.1</b>	<b>Importância dos Imunodiagnósticos</b>	<b>20</b>
<b>3.2</b>	<b>Metodologia de imunodiagnósticos</b>	<b>21</b>
<b>3.3</b>	<b>Antígenos de sensibilização</b>	<b>23</b>
<b>3.4</b>	<b>Métricas para imunodiagnósticos</b>	<b>24</b>
<b>4</b>	<b><i>TOXOPLASMA GONDII</i></b>	<b>26</b>
<b>4.1</b>	<b>Variabilidade genotípica do parasito</b>	<b>27</b>
4.1.1	Variabilidade heterogênea	28
4.1.2	Isolados de Campos dos Goytacazes	28
4.1.3	Variações na proteína SRS29B	32
4.1.4	Variações intrônicas em SRS29B	32
<b>5</b>	<b>JUSTIFICATIVA</b>	<b>34</b>
<b>6</b>	<b>OBJETIVOS</b>	<b>35</b>
<b>6.1</b>	<b>Objetivo geral</b>	<b>35</b>
<b>6.2</b>	<b>Objetivos específicos</b>	<b>35</b>
<b>7</b>	<b>METODOLOGIA</b>	<b>36</b>
<b>7.1</b>	<b>Coleta de dados</b>	<b>36</b>
<b>7.2</b>	<b>Coleta de informações sobre imunoensaios em patentes</b>	<b>36</b>
<b>7.3</b>	<b>Análise de sequências protéicas</b>	<b>36</b>
<b>7.4</b>	<b>Análise de estruturas protéicas</b>	<b>37</b>
7.4.1	Obtenção das estruturas moleculares	37
7.4.2	Execução da dinâmica molecular	38
7.4.3	Análise das trajetórias moleculares	39

7.4.4	Análises estruturais alternativas . . . . .	39
<b>8</b>	<b>RESULTADOS . . . . .</b>	<b>41</b>
<b>8.1</b>	<b>Coleta de informações sobre imunoenaios em patentes . . . . .</b>	<b>41</b>
<b>8.2</b>	<b>Análise de epítomos lineares . . . . .</b>	<b>42</b>
<b>8.3</b>	<b>Análise de epítomos estruturais . . . . .</b>	<b>44</b>
<b>9</b>	<b>DISCUSSÃO . . . . .</b>	<b>56</b>
<b>9.1</b>	<b>Possíveis causas das baixas sensibilidades observadas . . . . .</b>	<b>56</b>
9.1.1	Janela de produção de imunoglobulinas . . . . .	56
9.1.2	Falhas metodológicas em imunoenaios . . . . .	57
9.1.3	Incompatibilidade de antígenos de sensibilização . . . . .	57
9.1.4	Isolados não clonais de <i>T. gondii</i> no mundo . . . . .	58
<b>9.2</b>	<b>Impacto de mutações em epítomos . . . . .</b>	<b>59</b>
9.2.1	Epítomos lineares . . . . .	60
9.2.2	Epítomos conformacionais . . . . .	60
<b>9.3</b>	<b>Mutações relevantes para a sensibilidade de imunoenaios . . . . .</b>	<b>61</b>
<b>9.4</b>	<b>Imunodiagnósticos utilizados atualmente . . . . .</b>	<b>63</b>
<b>9.5</b>	<b>Desenvolvendo imunodiagnósticos mais precisos . . . . .</b>	<b>63</b>
<b>10</b>	<b>PERSPECTIVAS FUTURAS . . . . .</b>	<b>66</b>
<b>11</b>	<b>CONCLUSÃO . . . . .</b>	<b>68</b>
<b>12</b>	<b>REFERÊNCIAS . . . . .</b>	<b>69</b>
	<b>REFERÊNCIAS . . . . .</b>	<b>69</b>



# 1 Introdução

A toxoplasmose é a doença causada pelo protozoário *Toxoplasma gondii*. Ela é altamente prevalente nas populações humanas, mantendo uma média mundial de 30% (BIGNA et al., 2020). As regiões de maior prevalência de toxoplasmose são observadas principalmente na América do Sul (ROSTAMI et al., 2019) e, especialmente no Brasil (DUBEY et al., 2012), onde algumas dessas regiões atingem mais de 80% da população, como é o caso de Campos dos Goytacazes.

Na maioria dos casos, a infecção acontece após a ingestão de alimentos contaminados. Esses alimentos incluem hortaliças contaminadas que não foram higienizadas adequadamente, carnes de animais infectados que não foram bem cozidas e águas contaminadas não filtradas e não fervidas (TENTER; HECKEROTH; WEISS, 2000). Após um curto período de incubação, inicia-se a toxoplasmose aguda que é geralmente assintomática e dura cerca de uma semana. Essa etapa é seguida de uma fase crônica, aonde o parasito se abriga em cistos intracelulares. Acredita-se que a fase crônica estende-se por toda a vida do hospedeiro (ZHAO; EWALD, 2020). Ainda não existe um fármaco capaz de eliminar o parasito na fase crônica. Diversos mecanismos de evasão imunológica são atribuídos a *T. gondii*, mas muitos deles ainda não são bem compreendidos. A toxoplasmose tende a causar quadros mais graves em indivíduos imunocomprometidos, que podem evoluir para inflamações neurológicas e oculares, que podem levar à morte (KHAN; KHAN, 2017).

No século XIX, doenças infecciosas como a gripe, cólera e o sarampo eram a maior causa de morte e debilidade prematura, podendo até dizimar populações inteiras (HOLMES et al., 2017). Os avanços tecnológicos na biologia e na medicina trouxeram métodos eficazes de diagnóstico, prevenção e tratamento para essas condições e a mortalidade declinou acentuadamente, principalmente nos países desenvolvidos. Atualmente, grande parte das doenças infecciosas de alta mortalidade estão controladas através de medidas como a vacinação em massa (GUALANO et al., 2019) e o desenvolvimento e ampla utilização dos antibióticos (SENSA-KOVIC; SMITH, 2001). Devido a certa especificidade dos antibióticos e manejo de possíveis efeitos colaterais, o diagnóstico é fundamental para o tratamento adequado da maioria das doenças infecciosas. Isso inclui os casos de toxoplasmose em pacientes imunocomprometidos. O principal método de diagnóstico para esse tipo de doença é o imunodiagnóstico. Essa classe de protocolos laboratoriais tem como objetivo determinar se houve contato prévio do paciente com um determinado patógeno. Isso pode ser realizado através da detecção de anticorpos específicos contra o patógeno em amostras sorológicas do paciente, ou através da detecção de determinados antígenos nesse mesmo tipo de amostra. A sensibilidade é uma métrica para diagnósticos que indica a proporção de casos positivos corretamente identificados como positivos. Os resultados de algum outro método de diagnóstico cuja alta precisão é conhecida são utilizados como

referência (BINNEY; HYDE; BOSSUYT, 2021).

Trabalhos anteriores do nosso grupo revelam indivíduos com lesão ocular característica de toxoplasmose ocular com teste sorológico não reagente para imunoglobulinas anti-*T. gondii* (PEIXOTO-RANGEL et al., 2008) e amostras sorológicas da população geral de Campos dos Goytacazes com presença de DNA de regiões genômicas do parasito como SAG1, SAG2 e NTS1 livres de anticorpos contra *T. gondii* pelos imunoenaios comerciais (MARTINS, 2015).

Os testes imunodiagnósticos de detecção de IgG anti-*T. gondii* utilizados nesses pacientes de Campos incluem o *Toxonostika*, da *Biomerrieux*, e o *Axsym*, da *Abbott Laboratories*. A sensibilidade indicada pelo fabricante no manual desses imunodiagnósticos é de 99.99%. Porém, no caso dos estudos citados acima, a sensibilidade observada parece ser menor que a sensibilidade apresentada pelos manuais dos imunoenaios.

Imunoenaios dependem da afinidade entre seus antígenos de captura (ASs) e os anticorpos que podem estar presentes na amostra, pois apenas anticorpos que interagem com esses antígenos são detectados. A literatura reconhece divergência de resultados, principalmente entre imunoenaios que utilizam diferentes antígenos de sensibilização, quando utilizados em um mesmo conjunto de amostras (MURATA et al., 2019). Isso garante certa especificidade ao diagnóstico, considerando a grande variedade de anticorpos presente em cada amostra sorológica, considerando que um grande repertório de imunoglobulinas é produzido contra diferentes antígenos de cada patógeno (ZHAN et al., 2021) e até mesmo contra antígenos do próprio organismo (NAGELE et al., 2012). Os ASs são moléculas presentes no parasito, purificadas à partir de culturas *in vitro* ou sintetizadas através de vetores de expressão gênica. Durante a realização do imunoenasão, ASs devem ser reconhecidos pelos mesmos anticorpos que o hospedeiro produz após ser exposto ao agente infeccioso. A proteína de *T. gondii* SRS29B, anteriormente conhecida como SAG1, é uma proteína de superfície altamente imunogênica (WANG; YIN, 2014). Ela é amplamente estudada e amplamente utilizada como antígeno de sensibilização em imunodiagnósticos. Esses fatores determinaram SRS29B como a proteína analisada neste trabalho, que busca investigar se a causa das divergências entre a sensibilidade declarada pelo fabricante dos imunoenaios comerciais e a sensibilidade observada em imunoenasios realizados anteriormente pelo grupo em pacientes de Campos dos Goytacazes, além do inventário oftalmológico, pode envolver a variabilidade genotípica das cepas de *T. gondii* encontradas nessa região. Para validação inicial dessa hipótese, um sistema *in silico* para análise do efeito de substituições de aminoácidos de ocorrência natural na proteína SRS29B de *T. gondii* foi desenvolvido, tendo como principais etapas a predição de epítomos lineares e experimentos de dinâmica molecular. Além disso, a possibilidade da ocorrência dessas substituições em isolados de *T. gondii* endêmicos na cidade de Campos dos Goytacazes foi avaliada como causa das baixas sensibilidades observadas em imunodiagnósticos que utilizam antígenos de cepas clonais como ASs. Com base nas informações obtidas, bases teóricas para o desenvolvimento de um imunoenasão mais sensível para pacientes infectados por cepas não-clonais de *T. gondii* foram determinadas.

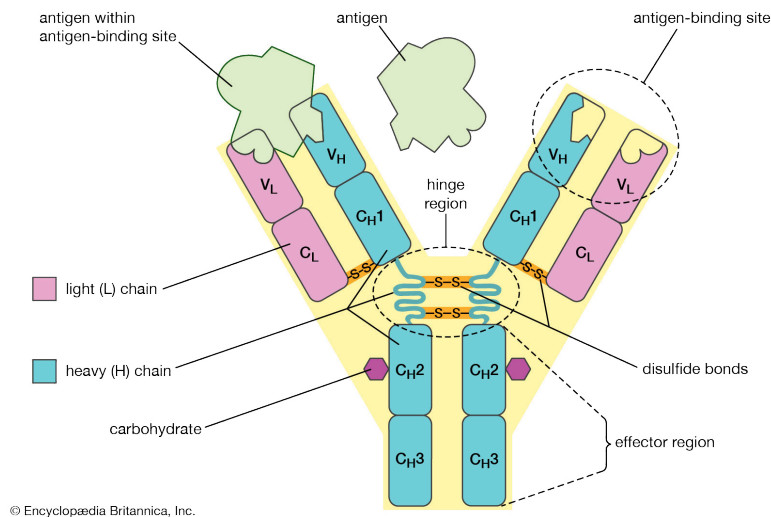
## 2 Imunoglobulinas

A produção de imunoglobulinas (Ig) é um aspecto da atividade imunológica dos vertebrados. Igs são proteínas que possuem uma região (parátipo) que possui afinidade com um sítio específico (epítipo) de uma determinada molécula (antígeno), que geralmente é uma proteína ou um carboidrato (MEGHA; MOHANAN, 2020). Existem diversas classes de Igs, e exemplos que são produzidos por *H. sapiens* incluem IgM e IgG. Classes de Igs são também conhecidas como isótipos. No caso de organismos que produzem diversas classes de Igs, cada isótipo pode desempenhar funções específicas. A IgA humana atua principalmente em mucosas enquanto a IgE é especializada na ativação de mastócitos e basófilos, por exemplo. Imunoglobulinas possuem uma região constante, ou Fc (*fragment-crystallizable*), que é comum a todos os anticorpos de sua classe, e uma região variável, ou Fab (*fragment antigen-binding*), que apresenta uma sequência de aminoácidos exclusiva e conseqüentemente uma estrutura molecular específica e complementar ao antígeno alvo (NEZLIN, 2019).

### 2.1 Estrutura das imunoglobulinas

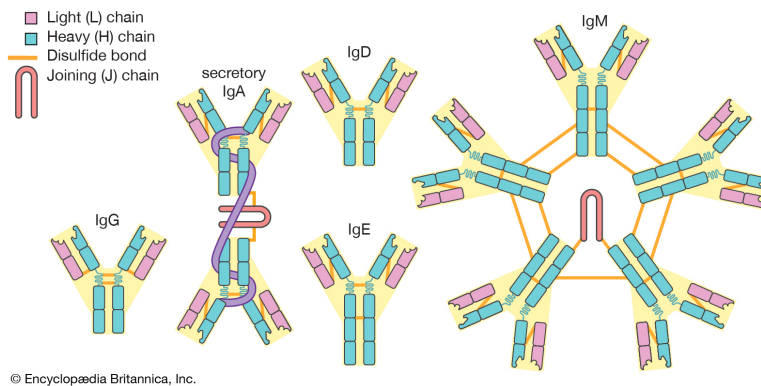
Estruturalmente, Igs são formadas por duas cópias de uma cadeia pesada (H) e duas cópias de uma cadeia leve (L). De outra forma, podemos classificar as regiões das Igs em termos de variabilidade estrutural entre diferentes imunoglobulinas. A porção Fab da Ig possui grande variabilidade nos domínios variáveis (V) que são específicos para ligação ao antígeno alvo. Já a porção Fc diferencia os subtipos de imunoglobulinas produzidos por animais da mesma espécie e também de espécies diferentes. Em Fab temos domínios V formados por cadeias pesadas ( $V_H$ ) e por cadeias leves ( $V_L$ ). Em Fc temos domínios constantes também formados por ambas cadeias  $C_H$  e  $C_L$ . A região efetora da imunoglobulina (Porção Fc) serve como sítio de ligação a outros componentes do sistema imunológico, e localiza-se na extremidade oposta às regiões variáveis. O anticorpo inclui uma “dobradiça” que liga a região responsável pela especificidade ao antígeno (Fab) com a região efetora (Fc). A “dobradiça” possui uma flexibilidade estrutural causada pela ausência de estruturas secundárias na região. Essa flexibilidade facilita o acesso da região efetora pelos outros componentes do sistema imunológico. Pontes dissulfeto unem as cadeias leves às cadeias pesadas, inclusive pela região da “dobradiça”. A porção Fc possui vários domínios de cadeia pesada, denominados  $C_{H1}$ ,  $C_{H2}$  e  $C_{H3}$ . Alguns isótipos apresentam  $C_{H4}$ .

Esse arranjo pode ser visto na Figura 1a. Cada um desses dois domínios únicos de Fab possui três regiões hipervariáveis, que podem ser observados como protuberâncias na estrutura proteica que possuem cerca de 10 aminoácidos de extensão. Essas regiões são comparativamente mais diversificadas entre diferentes Igs que as demais regiões da Fab, sendo fundamentais para afinidade da imunoglobulina a um antígeno específico.



© Encyclopædia Britannica, Inc.

(a) Estrutura geral de um complexo antígeno-anticorpo.



© Encyclopædia Britannica, Inc.

(b) Comparativo estrutural das diferentes classes de anticorpos presentes em humanos.

Figura 1 – Anticorpos

Fonte: Os editores da Encyclopaedia Britannica (2020)

## 2.2 Antígenos

Qualquer molécula capaz de despertar uma resposta imunológica pode ser chamada de antígeno. Anticorpos são produzidos, em geral, contra epítopos de moléculas encontradas no organismo hospedeiro, em geral, moléculas provenientes de organismos patogênicos ou estranho ao próprio, mas também podem ser produzidos contra moléculas do próprio organismo (NAGELE et al., 2012). O reconhecimento do antígeno pelos receptores de antígenos dos linfócitos B e a subsequente produção de imunoglobulinas com especificidade para essa molécula são etapas fundamentais da resposta imune adaptativa (KANELLOPOULOS; OJCIUS, 2019). A maior parte dos antígenos são proteínas, que são moléculas de alto peso molecular. Entre os antígenos de menor massa existem os haptenos, que são proteínas capazes de induzir a produção de anticorpos apenas quando ligadas a determinadas proteínas. Os anticorpos produzidos como parte da resposta imunológica aos haptenos podem se ligar à forma livre ou ligada dessas moléculas. Exemplos de haptenos incluem o corante de alimentos anilina, a toxina vegetal urushiol e o fármaco

maco utilizado para hipertensão hidralazina. Esses haptenos causam problemas de saúde quando ligam-se à proteína maior e despertam reações imunológicas. Ligantes comuns para haptenos em humanos são as proteínas séricas dos grupos globulinas, albuminas e ovoalbuminas. Pela incapacidade em estimular a produção de anticorpos por si próprios, os haptenos são considerados antígenos incompletos.

### 2.2.1 Epítomos

Um único antígeno pode ter diversos epítomos, que são regiões do antígeno que efetivamente interagem com as imunoglobulinas. A representação da interação entre SRS29B e um anticorpo monoclonal pode ser vista na Figura 2.

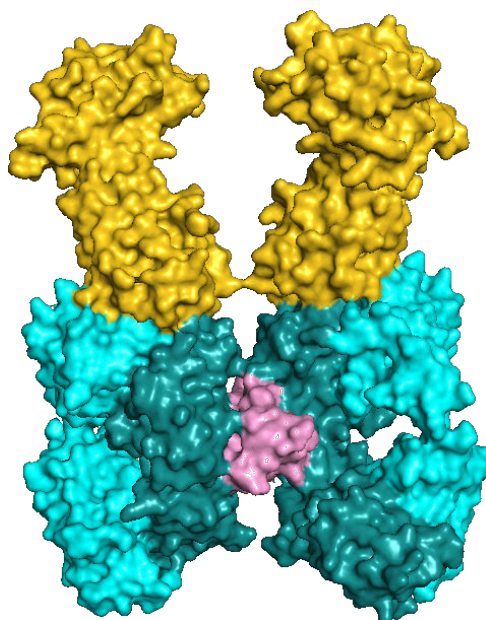


Figura 2 – Interação molecular entre SRS29B e um anticorpo monoclonal.

Nota: SRS29B em amarelo, cadeias leves em azul claro e cadeias pesadas em azul mais escuro.

Fonte: Estrutura 1YNT, disponível na RCSB Graille et al. (2005).

O epítopo possui entre 4 e 12 aminoácidos de extensão (BUUS et al., 2012). Epítomos podem estar contidos em uma parte linear da sequência de aminoácidos, sendo assim denominados epítomos lineares, ou envolverem a estrutura tridimensional da proteína quando os aminoácidos que interagem com o anticorpo são descontínuos na cadeia protéica. Esses epítomos descontínuos são os epítomos não-lineares. Epítomos lineares são determinados pela sequência de aminoácidos e as estruturas secundárias da proteína. Determinados aminoácidos e estruturas secundárias são mais presentes, estatisticamente, em regiões do antígeno que atuam como epítomos lineares (DALKAS et al., 2013). Existem ferramentas *in silico* capazes de prever epítomos lineares com até 80% de precisão, como faz o *B-Cell Epitope Oracle* (WANG et al., 2010) e o *BepiPred* (JESPERSEN et al., 2017). Dessa forma, é sabido que existem alguns padrões de aminoácidos que são mais propensos ao reconhecimento por linfócitos B além da inferência estatística, e esse

padrões podem ser reconhecidos por algoritmos como o classificador *Random Forest* (RIGATTI, 2017), utilizado no *BepiPred*. Já a determinação dos epítomos não-lineares requer informações sobre a estrutura tridimensional da proteína, sendo uma operação mais complexa que pode envolver simulações de interações moleculares, conhecidas como *docking*, envolvendo Igs e seus antígenos. Entender os epítomos não-lineares é um grande desafio da imunoinformática, uma área do conhecimento que envolve a aplicação de técnicas da bioinformática na imunologia (TOMAR; DE, 2014).

Através de seus fatores físicos e químicos e de características do sistema imunológico dos animais, um epítomo pode ser mais propenso a despertar a resposta imunológica humoral em comparação a outros epítomos. O epítomo mais imunogênico é chamado de epítomo dominante (YAMANO et al., 2011). Considerando *T. gondii*, anticorpos são produzidos contra proteínas de superfície, como SRS29B, e também contra proteínas secretadas pelo parasito, como GRA7, GRA8 e ROP1.

## 2.3 Interação entre imunoglobulina e antígeno

A interação entre anticorpo e antígeno depende da complementaridade estrutural entre o parátomo e o epítomo (SELA-CULANG; KUNIK; OFRAN, 2013). Forças atômicas como ligações de hidrogênio, a força de Wan der Waals e atração eletrostática contribuem para a afinidade entre as duas partes. As ligações de hidrogênio possuem comprimento entre 1.6 e 2.0 Å e ocorrem quando os átomos envolvidos de cada estrutura situam-se em distância compatível. A título de comparação, um átomo de hidrogênio possui diâmetro de cerca de 1.0 Å, de modo que as margens para o comprimento das ligações de hidrogênio são relativamente estreitas (SIGALA et al., 2015). Ligações de hidrogênio são feitas por um átomo de hidrogênio e outro átomo altamente eletronegativo, como o oxigênio (O). As pontes salinas também são importantes na estabilização do complexo anticorpo-antígeno (SINHA et al., 1999). Dessa forma, a complementaridade estrutural entre parátomo e epítomo determina quais forças atuarão entre os átomos das duas moléculas, determinando assim o grau de afinidade entre as partes, que é inversamente proporcional à reversibilidade da reação.

A especificidade aos antígenos é determinada pela região variável do anticorpo, chamada CDR (*Complementarity-determining regions*) (KAPINGIDZA; KOWAL; CHRUSZCZ, 2020). Essa parte é diferente entre anticorpos de mesma classe que possuem antígenos diferentes como alvo. No entanto, modificações em outras partes também podem influenciar na especificidade ao antígeno. A estrutura formada quando uma imunoglobulina liga-se a seu antígeno-alvo é chamada de imunocomplexo.

## 2.4 Especificidade de imunoglobulinas

Os linfócitos B humanos são capazes de produzir um número maior que  $10^{10}$  anticorpos diferentes. Esse repertório é obtido com o auxílio da recombinação de grupos de genes das cadeias variáveis de imunoglobulinas. Esse processo acontece no baço e envolve determinados tipos de linfócitos T e B, e resulta na produção de anticorpos altamente específicos a um determinado antígeno (WANG et al., 2020).

Moléculas como carboidratos de grande massa com padrões estruturais repetidos (polissacarídeos) conseguem despertar uma resposta imunológica humoral de linfócitos B sem o auxílio de linfócitos T. Esses antígenos são chamados de T-independentes. Essa resposta imunológica não produz imunoglobulinas de alta afinidade ou células de memória. Para a maior parte dos antígenos, inclusive os protéicos, é preciso uma ativação dos linfócitos B por um linfócito T, sendo esses os chamados antígenos T-dependentes.

Cada organismo que é capaz de produzir anticorpos mantém uma população base de linfócitos T e linfócitos B. Esses linfócitos nunca foram expostos a antígenos, sendo chamados de *naive*. Eles estão presentes mesmo em organismos que nunca foram expostos a antígenos exógenos (REES, 2020; GREEF et al., 2019). A primeira etapa da produção de anticorpos contra antígenos T-dependentes acontece quando a célula dendrítica apresenta o antígeno previamente capturado para um linfócito T CD4<sup>+</sup> (LAMAN; CLAASSEN; NOELLE, 2018). A demonstração de afinidade do linfócito B ao antígeno acontece quando esse linfócito é capaz de reconhecer esse antígeno e apresentá-lo via MHC Classe II ao linfócito T auxiliar que demonstrou previamente afinidade a esse mesmo antígeno. Como os epítomos de linfócitos T dependem do complexo MHC (PETERS; NIELSEN; SETTE, 2020) enquanto os epítomos de linfócitos B não dependem, esses dois tipos celulares geralmente possuem afinidade a epítomos diferentes do antígeno.

## 2.5 Atividade imunológica

A população de linfócitos B que participa do processo de maturação nos centros germinais atinge afinidade média crescente ao antígeno (MOLARI et al., 2020). Esses linfócitos B são expostos a uma série de sinalizadores dos linfócitos T CD4<sup>+</sup> que possuem afinidade ao mesmo antígeno. Um dos efeitos dessa interação mediada por CD40 é a ativação do fator de transcrição *Bcl-6* nos linfócitos B. Esses linfócitos possuem vida longa, apresentam alta capacidade migratória e alta afinidade aos antígenos. As alterações estruturais resultantes incluem a troca do isótipo do anticorpo produzido. Esse é um marco na atividade imunológica dos linfócitos B mediada por linfócitos T.

Em humanos, a troca de isótipos comum na resposta contra bactérias, vírus e protozoários é a troca de IgM para IgG. O isótipo IgG ainda possui as subclasses IgG<sub>1</sub>, IgG<sub>2</sub>, IgG<sub>3</sub> e

IgG<sub>4</sub>, onde cada subclasse apresenta nível médio de avidéz, particularidades estruturais, funções específicas, e momento da resposta imunológica em que são produzidos (VIDARSSON; DEKKERS; RISPENS, 2014). Anticorpos secretados ficam livres no plasma. A porção constante Fc dos anticorpos é reconhecida pelos receptores Fc $\gamma$ R, presentes na superfície de certas células imunológicas como basófilos, macrófagos e células NK. Assim, Fc $\gamma$ R funciona também como uma ponte entre a imunidade adaptativa e a imunidade inata. Considerando que a região Fab liga-se ao antígeno e a região Fc localiza-se em outra extremidade do polipeptídeo da imunoglobulina, o complexo antígeno-anticorpo expõe a região Fc. Dessa forma, os Fc $\gamma$ R de células fagocíticas e citotóxicas possuem acesso espacial aos domínios Fc de Igs complexadas (KAPINGIDZA; KOWAL; CHRUSZCZ, 2020). A porção Fc também serve de ponto de ancoragem de componentes da via clássica do sistema complemento, como C1q. A instalação do sistema complemento resulta na lise da membrana do patógeno, causando a sua destruição (ERP et al., 2018). A presença física do anticorpo pode impedir a função original da proteína alvo, bloqueando receptores (MINOR et al., 1984), enzimas (WANG et al., 2008) e interações com outras moléculas (CHEN et al., 2020). Essa atuação pode inibir a atividade do patógeno, facilitando o controle da infecção.

Imunoglobulinas cujos parátomos reconhecem epítomos presentes no próprio organismo existem naturalmente em organismos saudáveis (NAGELE et al., 2012). Certas doenças autoimunes como a artrite reumatóide foram associadas com a atividade dessa população de Igs (POULSEN et al., 2020).



### 3 Imunodiagnósticos

O imunoensaio é um conjunto de técnicas laboratoriais que utilizam a afinidade natural entre imunoglobulinas e seus antígenos para verificar a presença de uma determinada molécula em uma amostra. No contexto do diagnóstico clínico, essa técnica é utilizada para a detecção de imunoglobulinas com especificidade a antígenos de determinado patógeno em amostras sorológicas do paciente. Considera-se que a presença das imunoglobulinas indica a ocorrência do contato prévio do organismo com o patógeno. O diagnóstico por imunoensaio é chamado de imunodiagnóstico.

Existem diferentes classes de anticorpos que são produzidas em momentos distintos da infecção, de modo que as curvas de concentração sorológica variam ao longo do tempo após o primeiro contato com o agente infeccioso. Essas curvas podem ser visualizadas na Figura 3. As classes IgM e IgG são bastante utilizadas no diagnóstico clínico.

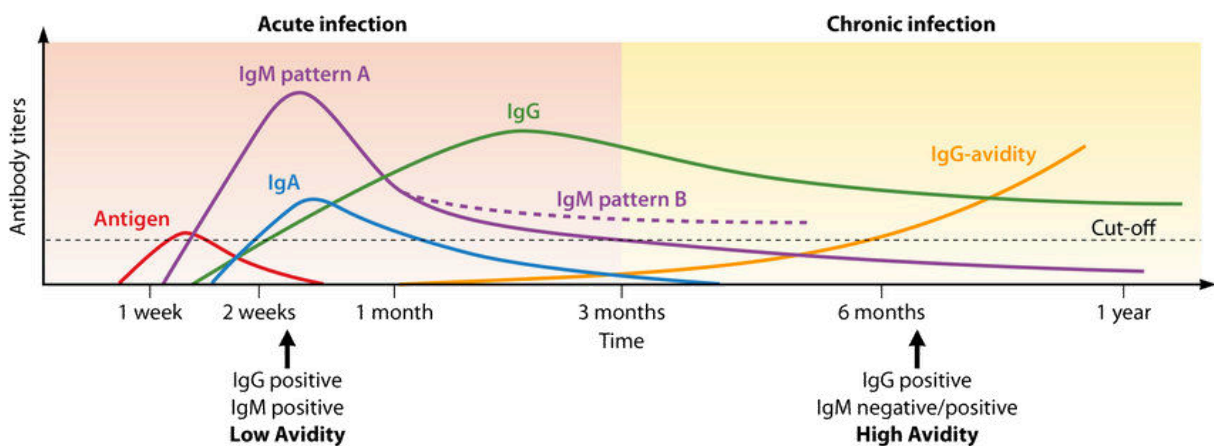


Figura 3 – Variação da concentração sorológica de anticorpos ao longo do tempo.

Fonte: Teimouri et al. (2020)

#### 3.1 Importância dos Imunodiagnósticos

O imunoensaio é amplamente utilizado para o diagnóstico de doenças infecciosas (ANDREOTTI et al., 2003). O diagnóstico sorológico da toxoplasmose, no contexto clínico, é realizado principalmente em gestantes, já que a ocorrência da fase aguda da doença durante a gestação pode resultar em complicações no desenvolvimento do sistema nervoso central do feto, como hidrocefalia ou morte (LOPES et al., 2007). O diagnóstico correto permite que o tratamento, que geralmente envolve a sulfadiazina (anti-microbiano) e a pirimetamina (anti-malárico), seja iniciado, considerando que esses fármacos não são bem tolerados (GUARALDO et al., 2017). Além disso, alguns desses medicamentos, como a sulfadiazina, possuem ação antibacteriana e seu uso indiscriminado favorece o surgimento de cepas bacterianas resistentes a antibióticos

(KOPMANN et al., 2012; SENGUPTA; CHATTOPADHYAY; GROSSART, 2012). Pacientes imunocomprometidos que estão na fase crônica da toxoplasmose estão sujeitos à reativação da doença, resultando em encefalite e outras complicações potencialmente fatais (LUFT; REMINGTON, 1992). Nesses casos, o diagnóstico correto é necessário para a avaliação do risco na administração de imunossupressores e para profilaxia farmacológica nos casos soropositivos.

## 3.2 Metodologia de imunodiagnósticos

Introduzido em 1959, o radioimunoensaio (RIA - *Radioimmunoassay*) permaneceu durante anos como o único método de imunoensaio para a detecção de anticorpos em amostras. Laboriosa, essa metodologia envolve marcadores radioativos que oferecem riscos associados à exposição durante o manuseio e preparação dos reagentes. A metodologia exigia equipamento especializado para proteção do operador e visualização dos resultados. Apesar das desvantagens, o RIA permanece até hoje como um dos tipos de imunoensaio com menor custo por amostra e com maior precisão de diagnóstico.

A técnica que permitiu ampla utilização dos imunoensaios foi a ELISA (*Enzyme Linked Immunosorbent Assay*). A técnica foi definida em 1971 por pesquisadores da Universidade de Estocolmo, através da união de técnicas desenvolvidas anteriormente. O novo método destacou-se pela simplicidade de execução, baixo custo, e segurança ao não utilizar marcadores radioativos que eram comuns nas metodologias utilizadas na época.

As reações do imunoensaio ELISA acontecem em uma placa de polipropileno de fundo chato com 96 poços, cujo material tem capacidade de adsorção. As etapas desse ensaio consistem em adsorver compostos nos poços na sensibilização com o antígeno e detecção de anticorpo presente na amostra, que depende da interação entre essas moléculas. Após cada uma dessas etapas, ocorre uma etapa de lavagem para remoção de moléculas não adsorvidas. Ao final de todas essas etapas, qualquer molécula que passou pelo poço e não formou complexos com as moléculas fixadas na parede poço já foram descartadas. Anticorpos secundários com marcadores que se tornam fluorescentes na presença de um substrato que é também adicionado no poço são geralmente utilizados para a detecção das moléculas de imunoglobulinas com especificidade ao antígeno. Se a molécula de interesse é um anticorpo, como no caso do exemplo, serão utilizados anticorpos secundários com afinidade à porção Fc desse anticorpo. Se a molécula de interesse é uma proteína, os anticorpos secundários empregados precisam ter afinidade a essa proteína.

O ELISA possui ainda diversas variações, como o método direto, indireto e sanduíche. O método a ser utilizado depende do objetivo do ensaio. A variante mais utilizada para imunodiagnósticos é o ELISA indireto, que detecta anticorpos nas amostras. Uma representação do funcionamento do imunoensaio ELISA em suas diferentes vertentes pode ser vista na Figura 4.

Com base nos princípios do ELISA, imunoensaios comerciais como o Vidas e o AxSYM utilizam técnicas ligeiramente diferentes como o MEIA (*Microparticle Enzyme Immunoassay*) e

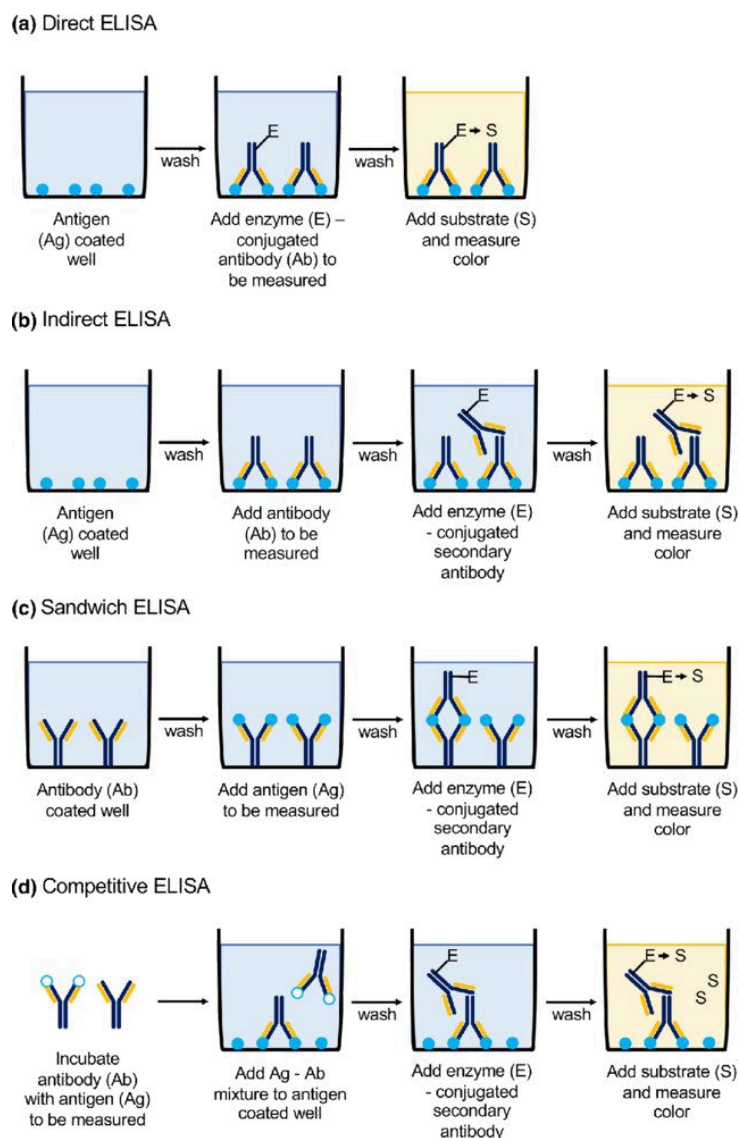


Figura 4 – Imunoensaio ELISA

Nota: Esquemas mostrando o funcionamento das diferentes variações do imunoensaio ELISA. O mais utilizado em imunoensaios é o ELISA indireto (b).

Fonte: Boguszewska et al. (2019)

CMIA (*Chemiluminescence Microparticle Immunoassay*). Esses métodos derivados diferem-se do ELISA no método utilizado para detectar os complexos compostos pelas proteínas fixadas nos poços e proteínas presentes na amostra.

Os imunoensaios modernos para toxoplasmose possuem sensibilidade maior que 99%, conforme descrito nos manuais dos produtos e em estudos independentes (CALDERARO et al., 2008). Estudos que determinam as métricas do diagnóstico consistem na comparação do resultado dos imunoensaios testados com resultados de diagnósticos realizados por um método referencial de alta eficácia, e que geralmente é de realização mais complexa. Esse método referencial é chamado de padrão-ouro. A detecção de material genético do agente infeccioso na amostra sorológica por meio do PCR é geralmente usada como padrão-ouro para imunodiagnós-

ticos. A comparação dos resultados do imunoensaio com os resultados do método referencial é realizada utilizando o mesmo grupo de amostras sorológicas. Através dos cálculos mostrados na Figura 5, obtém-se as métricas de sensibilidade e especificidade do diagnóstico.

		<b>Diagnóstico pelo padrão-ouro</b>	
		<b>+</b> (PO+)	<b>—</b> (PO-)
<b>Diagnóstico pelo imunoensaio avaliado</b>	<b>+</b>	<b>Positivos Verdadeiros</b>	<b>Falsos Positivos</b>
	<b>—</b>	<b>Falsos Negativos</b>	<b>Negativos Verdadeiros</b>
		<b>Sensibilidade = PV / PO+</b>	<b>Especificidade = NV / PO-</b>

Figura 5 – Métricas de sensibilidade e especificidade para imunoensaios

Nota: Descrição dos métodos de cálculo da sensibilidade e especificidade de imunodiagnósticos.

### 3.3 Antígenos de sensibilização

Antígenos de sensibilização são responsáveis pelo reconhecimento da imunoglobulina alvo na amostra sorológica. Qualquer antígeno do parasito que induza uma resposta imunológica do hospedeiro pode ser utilizado como antígeno de sensibilização.

A primeira etapa do processo de desenvolvimento de imunoensaios envolve a identificação de proteínas capazes de interagir com os anticorpos de interesse, as candidatas a ASs, e a confirmação experimental dessa afinidade. Essa etapa inicial pode ser realizada de modo *in silico* através da utilização de ferramentas como a docagem molecular e a dinâmica molecular (CAN et al., 2020). O antígeno de sensibilização é um componente fundamental do imunoensaio e afeta diretamente suas métricas de sensibilidade e especificidade.

A descrição dos antígenos de sensibilização que foram utilizados em determinado imunoensaio geralmente não estão presentes nos manuais de operação dos produtos. Esse tipo de informação é considerada propriedade intelectual e é protegida por patentes. A documentação das patentes descreve detalhes sobre ASs, incluindo a sequência de RNA mensageiro capaz de sintetizar o AS. Patentes podem ser consultadas através do serviço Google Patents. Os ASs podem ser adaptados para uso em vacinas, além dos imunodiagnósticos, e essa possibilidade

geralmente é informada nos registros desse tipo de tecnologia para proteger os interesses do inventor.

### 3.4 Métricas para imunodiagnósticos

Imunodiagnósticos podem indicar resultados incorretos, e isso acontece através de diversos possíveis mecanismos. Existem dois tipos de resultados incorretos, os falsos negativos e os falsos positivos. Na amostra, a presença de substâncias que alteram a mensurabilidade do resultado ou alteram a afinidade de anticorpos podem resultar em interferência no imunoensaio (TATE; WARD, 2008). Essas interferências podem levar a leituras distorcidamente altas ou baixas, prejudicando a precisão do resultado.

A variabilidade genética apresentada por agentes patogênicos, no caso das doenças infecciosas, pode interferir na precisão dos resultados (ASCOLI, 2021). Anticorpos produzidos pelo hospedeiro contra uma determinada versão do antígeno utilizado como AS podem não ter afinidade suficiente ao AS utilizado no imunoensaio. Isso pode acontecer devido a variações entre o antígeno do parasito e o AS presente no imunoensaio. O *antigenic drift* é um fenômeno evolucionário aonde o parasito desenvolve variações em antígenos imunogênicos para escapar da atividade imunológica do hospedeiro, obtendo assim maior sucesso reprodutivo. Esse fenômeno pode explicar uma maior variação genética em proteínas altamente antigênicas de organismos parasitas. Tendo sido primeiramente observado em vírus (SCHILD et al., 1974), o *antigenic drift* é também observado em protozoários (VANBUSKIRK et al., 2004). Essas variações antigênicas costumam ser estudadas em agentes infecciosos de importância clínica (HARVEY et al., 2021), e pode afetar resultados de imunodiagnósticos (LIPPI; ADELI; PLEBANI, 2021).

Um diagrama que ilustra como a variação antigênica pode afetar o resultado de imunoensaios pode ser visto na Figura 6.

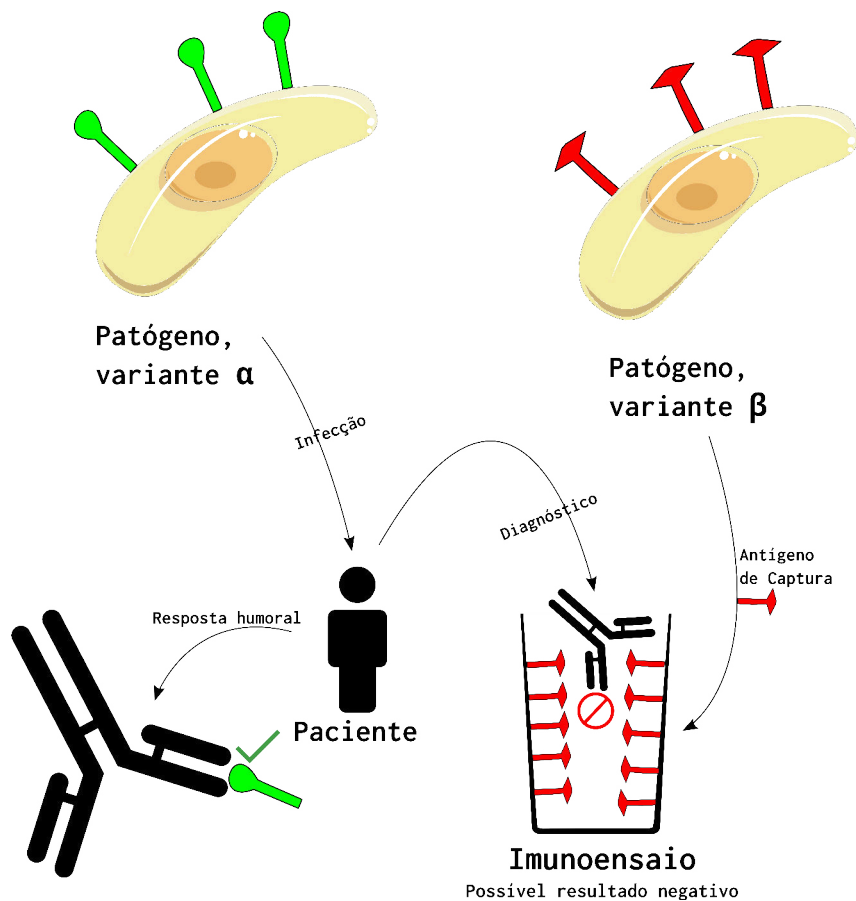


Figura 6 – Mecanismo de como o *antigenic drift* afeta a sensibilidade de imunoensaios.

Nota: O antígeno da variante  $\alpha$  do agente infeccioso é diferente do antígeno da variante  $\beta$  devido a substituições de aminoácidos. O hospedeiro infectado pela variante  $\alpha$  produz anticorpos contra essa variante. Enquanto isso, o imunoensaio utiliza antígenos de sensibilização similares aos antígenos da variante  $\beta$ . A baixa afinidade entre o anticorpo do paciente e o antígeno de sensibilização do imunoensaio pode resultar em resultados falsos negativos no imunoensaio.

## 4 *Toxoplasma gondii*

*Toxoplasma gondii* é um parasita amplamente disseminado em populações humanas e animais em diversas regiões do planeta. A infecção por *T. gondii* geralmente é assintomática, mas causa complicações ocasionais que são mais problemáticas em hospedeiros imunocomprometidos. A resolução da infecção primária envolve uma mudança de estágio de vida do parasita. A forma de rápido metabolismo da fase aguda, o taquizoíta, transforma-se no bradizoíta, que é uma forma de baixa atividade metabólica que abriga-se em cistos intracelulares, caracterizando a fase crônica da doença (CERUTTI; BLANCHARD; BESTEIRO, 2020). Esse estágio é altamente persistente e acredita-se que dura pelo resto da vida do hospedeiro.

*T. gondii* é um protozoário que pertence ao filo apicomplexa, formado por endoparasitas intracelulares obrigatórios de animais. A maioria dos apicomplexa apresenta um plastídeo chamado apicoplasto, que possui seu próprio material genético. O plastídeo possui quatro membranas celulares e assim acredita-se que seja remanescente de uma alga vermelha que sofreu processo de endossimbiose secundária (DOOREN; STRIEPEN, 2013). Essa organela apresenta um genoma circular e codifica enzimas relacionadas ao metabolismo de ácidos graxos que são fundamentais ao parasito (MCFADDEN, 2010). O filo apicomplexa inclui outros organismos causadores de doenças em humanos. Exemplos incluem *Plasmodium falciparum*, o agente patogênico causador da malária (GARCIA, 2010), e *Criptosporidium parvum*, que causa a criptosporidiose e é uma importante causa de mortalidade em humanos com menos de 24 meses de idade (KING; TYLER; HUNTER, 2018), e *Babesia spp.*, causadores da babesiose, que é uma das parasitoses mais comuns em mamíferos, incluindo gado e humanos, transmitida pela pulga (UILENBERG, 2006). Já o *Hammondia hammondi* possui ciclo de vida bastante similar ao *T. gondii*, mas não infecta humanos (WALZER et al., 2013).

Os hospedeiros definitivos de *T. gondii* consistem nos membros da família *Felidae*, abrangendo felinos de grande porte como onças e tigres, além dos gatos domésticos (INNES, 2009). O conceito de hospedeiro definitivo deriva-se do fato do parasito reproduzir-se sexualmente apenas nas células do epitélio intestinal dos felinos. Essa especificidade acontece devido a uma deficiência na enzima delta-6-desaturase observada em membros dessa família de animais. Essa enzima é presente em todos os mamíferos e degrada o ácido linoléico proveniente da dieta em ácidos graxos menores. Devido a essa deficiência enzimática, o epitélio intestinal dos membros da família *Felidae* apresenta alta concentração de ácido linoléico. Esse ácido graxo é necessário, por motivos desconhecidos, ao mecanismo de reprodução sexual de *T. gondii* (GENOVA et al., 2019).



### 4.1 Variabilidade genotípica do parasito

Eventos de recombinação genômica seguidos de longos períodos de expansão clonal ao longo da história evolutiva desse organismo são responsáveis pela considerável variabilidade genotípica observada atualmente (SIBLEY; AJIOKA, 2008). Um mapa geral dessa variação pode ser observado na Figura 7.

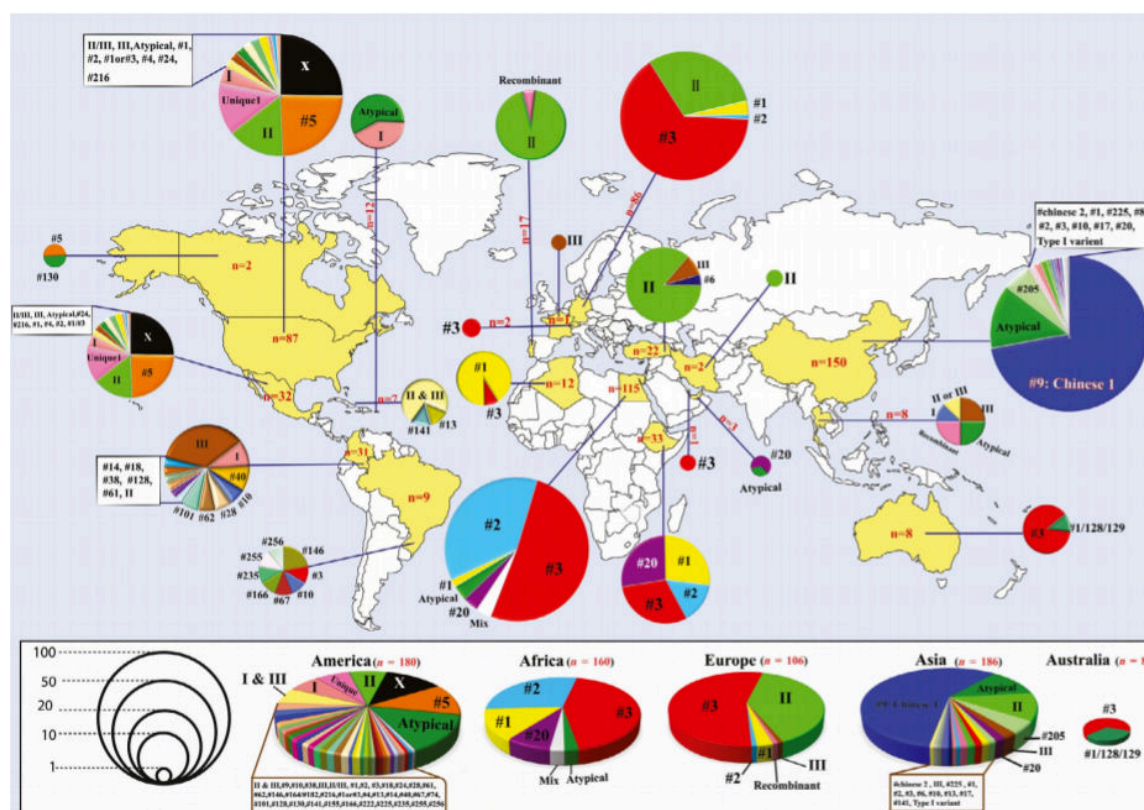


Figura 7 – Distribuição geográfica de genótipos de *T. gondii*

Nota: Distribuição geográfica mundial de genótipos de isolados de *T. gondii* extraídos de felinos. Países destacados em amarelo possuem isolados genotipados em quantidade indicada pela legenda em vermelho. Os gráficos circulares indicam a distribuição de genótipos para os isolados de cada país. Países mostrados em branco não possuem dados de genotipagens de *T. gondii* disponíveis.

Fonte: Amouei et al. (2019)

O sistema de genotipagem em loci PCR-RFLP proposto por Su et al. (2009) é amplamente utilizado para a genotipagem de isolados de *T. gondii* em escala internacional. Utilizando esse sistema de 10 marcadores, isolados oriundos de Campos dos Goytacazes foram genotipados como parte de um trabalho anterior do grupo (VIEIRA, 2015). Uma tabela disponível publicamente compila resultados da genotipagem para isolados originários de diversas regiões do planeta (SU et al., 2009). Esses isolados são classificados em haplogrupos, onde cada conjunto possível de marcadores PCR-RFLP representa um haplogrupo. Atualmente esse sistema de classificação inclui cerca de 250 haplogrupos, identificados através da genotipagem intensiva de isolados mundiais que é feita a mais de 20 anos (desde 1995), por diversos grupos de pesquisa (AJZENBERG, 2015). O propósito desse esforço coletivo é determinar a distribuição geográfica



e variação genotípica do protozoário, em escala mundial, sem recorrer ao sequenciamento genômico. Considerando os custos decrescentes do sequenciamento genômico conforme a tecnologia avança, em breve o sequenciamento dos isolados pode substituir os marcadores PCR-RFLP para este propósito. A genotipagem através de microssatélites SSR (*simple sequence repeat*), é outra análise genotípica bastante utilizada na determinação de diferenças clonotípicas em *T. gondii*, e consiste na contagem de CNVs em regiões específicas do genoma do parasito. Marcadores SSRs são menos utilizados em genotipagens de *T. gondii* que os RFLP devido ao maior custo e tempo de execução (YANG et al., 2012).

#### 4.1.1 Variabilidade heterogênea

De acordo com as genotipagens realizadas em escala global (Figura 7), *T. gondii* apresenta baixa variabilidade genética em regiões como a América do Norte e Eurásia, enquanto apresenta alta variabilidade em países da América do Sul e África. Análises e modelagem geográfica dos isolados encontrados apontam para uma origem sul americana para esse parasito, mais especificamente na região que hoje situa-se a Colômbia, de onde as variantes ancestrais teriam se espalhado a cerca de 1.5 milhões de anos atrás (BERTRANPETIT et al., 2016). As regiões mais próximas do ponto de origem de *T. gondii* apresentam maior diversidade genotípica desse parasito. A história evolutiva do parasito aponta para um alternância entre períodos de expansão clonal e eventos de recombinação sexuada (SILVA; LANGONI; MEGID, 2017).

#### 4.1.2 Isolados de Campos dos Goytacazes

O método PCR-RFLP diferencia regiões através da comparação das faixas observadas na eletroforese em gel com referências que representam cada grupo clonal. Faixas idênticas indicam similaridade (segundo o método) entre as variantes comparadas. Assim, o PCR-RFLP pode classificar cada região como similar a isolados de cada grupo clonal, ou ainda não similar a esses grupos, sendo esses alelos denominados “atípicos” e identificadas pelo prefixo *u*.

Os isolados de Campos dos Goytacazes apresentam mais loci PCR-RFLP atípicos que a média geral dos isolados conhecidos, principalmente nas regiões SAG1 (SRS29B) e c228, conforme pode ser visto na Tabela 1. Isolados de outras regiões de alta prevalência de isolados não-clonais estão na Tabela 2. Além disso, é possível observar a mistura de alelos clonais nos isolados endêmicos em Campos, principalmente os alelos clonais I e III. A recombinação e a presença de alelos atípicos resulta em algumas combinações de regiões RFLP nunca antes vistas em isolados de *T. gondii*. Dessa forma, dois desses isolados endêmicos em Campos são os primeiros isolados a se enquadrar em seus próprios haplogrupos. Os genótipos podem ser observados na Tabela 1.

Analisando a Tabela 1, observamos que 38% dos haplogrupos encontrados em Campos dos Goytacazes apresentam alelos atípicos em SAG1 (SRS29B), enquanto 14% dos haplogrupos

encontrados mundialmente apresentam alelos atípicos nesse marcador (SU et al., 2009). O único alelo atípico observado em SAG1 é o  $u-1$ . Como não há dados de sequenciamento para a região SAG1 desses isolados de Campos, não se sabe se há variação genética no gene SAG1 entre isolados que compartilham SAG1 <sub>$u-1$</sub> .

Tabela 1 – Genotipagem PCR-RFLP de isolados locais de *T. gondii*.

Isolados	SAG1	SAG2	Alt. SAG2	SAG3	BTUB	GRA6	c22-8	c29-2	L358	PK1	Apico	Genótipo #
TgCkBr 31/56/365	II or III	III	III	III	III	III	III	III	III	III	III	#2
TgCkBr 55/79/86/87/367	I	I	I	III	I	II	u-1	I	I	I	I	#6
TgCkBr 57/64	I	I	II	III	III	III	I	III	I	II	III	#11
TgCkBr 82/90	I	III	III	III	III	III	III	I	III	III	III	#14
TgCkBr 81	u-1	I	II	III	III	III	u-1	I	I	III	I	#17
TgCkBr 28/33/50/52/58/30/60	I	III	III	III	III	III	I	I	I	u-1	I	#19
TgCkBr 95/371	I	III	III	III	III	III	I	I	I	III	III	#21
TgCkBr 27/38/44/51/65/66/78/80/65	u-1	I	II	III	III	III	u-1	I	III	III	III	#22
TgCkBr 41/42/49/60/62	u-1	I	II	III	I	III	u-1	I	I	I	I	#33
TgCkBr 59/30/34/67/16/328/11/68/70	I	I	I	III	I	III	II	I	III	I	III	#36
TgCkBr 36/32/84/85	I	II	II	III	III	III	u-1	I	I	III	I	#37
TgCkBr 75/76/92	u-1	I	II	III	III	III	III	III	I	III	I	#40
TgCkBr 46/3/64	u-1	I	II	III	I	III	II	I	I	I	I	#51
TgCkBr 96	u-1	I	II	III	III	III	II	I	I	III	I	#53
TgCkBr 40/47	I	I	I	III	III	II	u-1	I	I	I	I	#59
TgCkBr 89	I	I	II	III	III	III	u-1	I	I	III	I	#65
TgCkBr 93/94	I	III	III	III	III	II	I	III	I	II	I	#69
TgCkBr 26/69	I	III	III	III	III	III	II	I	III	III	I	#71
TgCkBr 48/88	u-1	I	II	III	III	III	II	I	I	III	III	#75
TgCkBr 54	I	I	I	III	I	III	II	I	I	I	III	#82
TgCkBr 61	I	I	II	I	III	II	u-1	I	I	III	I	#93
TgCkBr 37	I	I	II	III	III	II	u-1	I	I	III	I	#107
TgCkBr 63	I	I	II	III	III	III	u-1	III	I	III	I	#110
TgCkBr 26	I	I	u-1	III	III	III	u-1	I	III	III	I	#111
TgCkBr 45	u-1	I	II	III	III	III	II	I	III	III	I	#135
TgCkBr 74	u-1	III	III	III	III	III	III	I	III	III	III	#138
TgCkBr 12	u-1	I	II	III	III	III	u-1	I	I	III	III	#247 (NOVO)
TgCkBr 46	I	III	III	III	III	III	u-1	I	III	u-2	III	#248 (NOVO)

Fonte: Vieira (2015)

Tabela 2 – Genotipagem PCR-RFLP de isolados de *T. gondii* de interesse.

Código BRC do Isolado	Genótipo #	Local de Origem	SAG1	SAG2	alt.SAG2	SAG3	BTUB	GRA6	c22-8	c29-2	L358	PK1	Apico
TgCatBr18	119	Brasil	I	III	III	III	III	II	u-1	I	I	u-1	I
TgCatBr25	47	Brasil	I	III	III	III	III	II	u-1	I	I	II	I
TgCatBr34	104	Brasil	I	I	II	III	I	III	u-1	I	I	III	I
TgCatBr40	92	Brasil	I	I	II	I	III	II	II	I	I	II	I
TgCatBr57	108	Brasil	I	I	II	III	III	III	II	I	I	III	I
TgCatBr81	124	Brasil	I	III	III	III	III	III	II	III	I	u-1	I
TgCkBr037	107	Brasil	I	I	II	III	III	II	u-1	I	I	III	I
TgCkBr061	93	Brasil	I	I	II	I	III	II	u-1	I	I	III	I
TgCkBr147	17	Brasil	u-1	I	II	III	III	III	u-1	I	I	III	I
TgCkBr155	76	Brasil	u-1	III	III	III	III	III	u-1	I	I	III	I
TgDgBr18	106	Brasil	I	I	II	III	III	II	u-1	I	I	II	I
TgCkBr059	36	Brasil	I	I	I	III	I	III	II	I	III	I	III
TgCkBr186	88	Brasil	I	I	I	III	III	III	II	I	III	I	III
TgCatBr38	136	Brasil	u-1	I	u-1	III	III	III	II	I	I	III	I
TgCatBr64	111	Brasil	I	I	u-1	III	III	III	u-1	I	III	III	I
TgCkBr013	63	Brasil	I	I	II	III	III	III	I	III	I	II	I
TgCkBr045	135	Brasil	u-1	I	II	III	III	III	II	I	III	III	I
TgCkBr173	81	Brasil	I	I	I	III	I	II	u-1	I	I	III	III
TgCkBr177	109	Brasil	I	I	II	III	III	III	III	I	III	III	III
TgCkBr178	134	Brasil	u-1	I	II	III	I	III	II	III	III	I	III
GUY-BAS1 (TgH18006)	100	Guiana Francesa	I	I	II	I	III	III	u-1	I	III	III	I
GUY-KOE (TgH18002)	60	Guiana Francesa	I	I	II	I	III	III	III	I	III	III	I
GUY-MAT (TgH18003)	95	Guiana Francesa	I	I	II	I	III	III	II	I	I	III	I
GUY-RUB (TgH00002)	98	Guiana Francesa	I	I	II	I	III	III	III	III	I	III	III
GUY-JAG1 (TgA18001)	197	Guiana Francesa	I	II	II	II	II	II	II	II	I	II	I
TgCtCo05	61	Colômbia	I	I	II	III	I	III	II	III	I	u-2	I
TgCtCo15	101	Colômbia	I	I	II	III	I	III	II	I	I	u-2	I
TgDgCo06	44	Colômbia	I	I	II	III	I	III	II	I	I	u-3	I
TgCtPRC04	9	China	u-1	II	II	III	III	II	II	III	II	II	I

Fonte: Su et al. (2012)

### 4.1.3 Variações na proteína SRS29B

A proteína SAG1 (*Surface Antigen 1* - Antígeno de Superfície 1) foi o primeiro antígeno de superfície descoberto em *T. gondii*, é o antígeno imunodominante do parasito, e foi inicialmente batizado como p30 (RODRIGUEZ et al., 1985). Com a descoberta de novos antígenos relacionados, p30 foi renomeada para SAG1 e a família se tornou conhecida por SRS (*SAG-Related Sequence*) Sequência relacionada a SAG (JUNG; LEE; GRIGG, 2003). Algumas proteínas do grupo SRS possuíam o prefixo SAG, como SAG2 e SAG3. Outras possuíam o prefixo SRS, como SRS9 e SRS2 (WASMUTH et al., 2012). Mais tarde houve uma renomeação geral dessas proteínas, e SAG1 foi renomeada para SRS29B. Dessa forma, os três nomes SRS29B, SAG1 e p30 coexistem na literatura, designando a mesma proteína.

O gene SRS29B codifica o mRNA da proteína SRS29B e é localizado no cromossomo VIII de *T. gondii*. A proteína sofre modificações pós translacionais, incluindo a clivagem de porções N-terminal e C-terminal da sequência de aminoácidos. A sequência de aminoácidos traduzida possui 336aa, enquanto a estrutura protéica que é encontrada na membrana do protozoário possui 252 aminoácidos. Substituições de aminoácidos podem levar a alterações estruturais na proteína, principalmente quando a substituição envolve resíduos de cargas diferentes (ZHOU; PANG, 2018). Diferenças nas estruturas primárias e secundárias em diferentes variantes SRS29B já demonstraram induzir variação nos níveis de produção de diferentes subclasses de IgG (KATO et al., 2006).

### 4.1.4 Variações intrônicas em SRS29B

Considerando os genomas públicos de isolados de *T. gondii*, é possível observar uma variação no número de repetições CNV (*Copy number variation*) de uma sequência no íntron de SRS29B, conforme mostra o alinhamento de seqüências genômicas na Figura 8. CNVs intrônicos podem afetar a expressão do gene em que se encontram (ROSE, 2008). CNVs na região promotora de SRS29B já foram relacionados a diferentes níveis de produção da proteína em conjunto com diferentes níveis de virulência. Em um estudo, variantes com 5 repetições em um determinado fragmento de 27pb localizado na região promotora de SRS29B produziram quatro vezes mais SRS29B e eram mais virulentas em camundongos, em comparação com as variantes que apresentavam 4 repetições daquele fragmento (WINDECK; GROSS, 1996). Essa observação pode estar ligada ao papel de SRS29B no processo de invasão da célula hospedeira (WINDECK; GROSS, 1996).

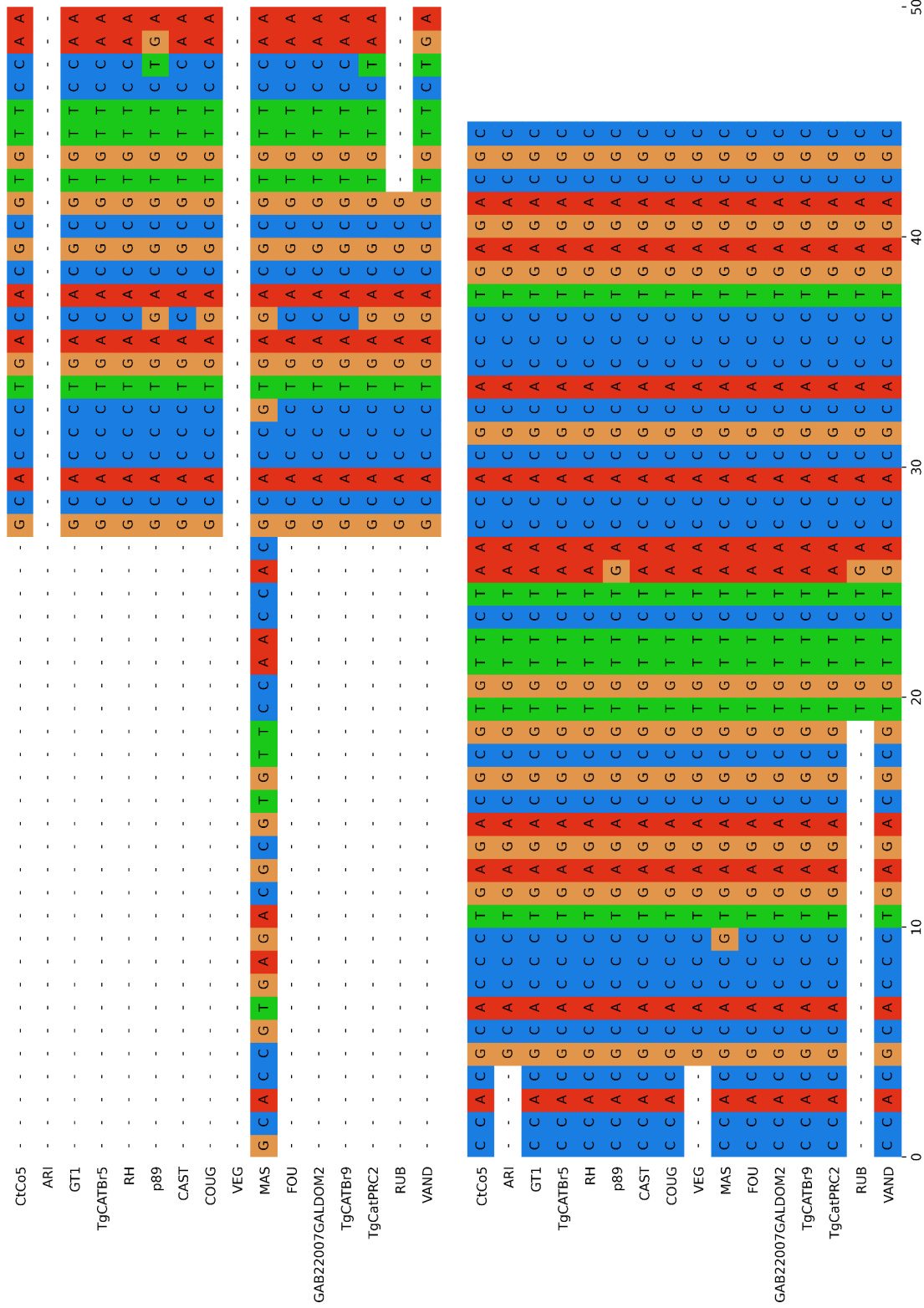


Figura 8 – Alinhamento de uma região upstream ao gene SRS29B.

Nota: A região genômica mostrada é a 2662325:2664915 do cromossomo VIII, tendo como referência o genoma da variante TGME49.

## 5 Justificativa

A toxoplasmose pode causar complicações clínicas em indivíduos imunocomprometidos (WANG et al., 2016) e indivíduos imunocompetentes (BASIT et al., 2018), além de possíveis alterações subclínicas em toda a população (FLEGR; HORÁČEK, 2019). O imunoensaio é o principal método de diagnóstico da toxoplasmose, tanto para propósitos clínicos quanto epidemiológicos (DARD et al., 2016). Tendo sido observada a inesperada baixa sensibilidade em imunodiagnósticos comerciais para toxoplasmose em pacientes no município de Campos dos Goytacazes (PEIXOTO-RANGEL et al., 2008; MARTINS, 2015), este trabalho contempla a hipótese de que a variabilidade genética das cepas endêmicas no município contribui para a baixa sensibilidade observada nos imunoensaios. Considera-se que um diagnóstico falso negativo pode retardar o tratamento adequado de pacientes verdadeiramente infectados. Isso pode resultar em danos oculares e neurológicos permanentes, causados pela toxoplasmose.

A proteína de *T. gondii* SRS29B (SAG1) é amplamente utilizada como antígeno de sensibilização em imunoensaios comerciais, inclusive nos imunoensaios utilizados nos trabalhos anteriores do grupo que indicaram a baixa sensibilidade destes em pacientes residentes em Campos. Sendo também uma das poucas proteínas com estruturas disponíveis na RCSB, SRS29B foi a escolhida dentre as proteínas utilizadas em imunoensaios de diagnóstico de toxoplasmose para compor as análises estruturais realizadas neste trabalho.

Existem outras regiões geográficas que assim como Campos dos Goytacazes, são endêmicas de variantes altamente polimórficas de *T. gondii*. Essas regiões estão localizadas principalmente na América do Sul. É possível que a baixa sensibilidade em imunoensaios observada em Campos ocorra nessas regiões. Dessa forma, apesar deste trabalho ser focado nos isolados de *T. gondii* de Campos e nos pacientes dessa região, seus resultados podem beneficiar outros grupos populacionais.

## 6 Objetivos

### 6.1 Objetivo geral

Neste sentido, o presente trabalho buscou determinar, através de experimentos *in silico*, se substituições de aminoácidos na proteína SRS29B, uma vez presentes em variantes de *T. gondii* endêmicas em Campos dos Goytacazes, poderiam comprometer a eficácia de kits comerciais, usualmente utilizados para diagnóstico sorológico de infecções por *T. gondii*, em amostras de indivíduos infectados por cepas polimórficas deste parasita.

### 6.2 Objetivos específicos

1. Determinar a possibilidade da ocorrência de substituições de aminoácidos em SRS29B que afetem imunocompatibilidade de antígenos de sensibilização baseados nessa proteína.
2. Determinar se as substituições disruptivas podem ocorrer em isolados endêmicos em Campos dos Goytacazes.
3. Reunir e organizar sequências protéicas de antígenos de sensibilização utilizados em imunoenaios comerciais e experimentais, através da consulta em patentes.
4. Desenvolver uma *pipeline* (sequência de etapas metodológicas) *in silico* (em ambiente computacional), para determinar quais aminoácidos dos antígenos de sensibilização são mais críticos para sua imunocompatibilidade, ou seja, mais determinantes para a afinidade com imunoglobulinas específicas e que assim possuem maior probabilidade de afetar as métricas de sucesso de imunoenaios quando sofrem substituições em populações do patógeno.



## 7 Metodologia

### 7.1 Coleta de dados

Diferentes bancos de dados de livre acesso descritos na Tabela 3 serão utilizados para obtenção das informações necessárias ao desenvolvimento deste trabalho. Essa informação inclui especificações técnicas de imunoenaios comerciais, sequências protéicas de diversos isolados e estruturas protéicas para simulação molecular. A coleta dos dados será realizada através dos websites de cada banco, com auxílio dos respectivos sistemas de busca.

Tabela 3 – Bancos de dados consultados durante o projeto.

Banco de Dados	Endereço Eletrônico	Tipo de informação
ToxoDB	<a href="https://toxodb.org/">https://toxodb.org/</a>	Sequências genômicas do parasito.
IEDB	<a href="http://www.iedb.org/">http://www.iedb.org/</a>	Epítomos confirmados experimentalmente.
RCSB	<a href="https://www.rcsb.org/">https://www.rcsb.org/</a>	Estruturas protéicas em arquivos PDB.
Pubmed	<a href="https://pubmed.gov/">https://pubmed.gov/</a>	Embasamento teórico na literatura.
Google Patents	<a href="https://patents.google.com/">https://patents.google.com/</a>	Busca de patentes por palavras-chave.

Nota: Bancos de dados consultados durante o projeto e seus respectivos endereços eletrônicos.

### 7.2 Coleta de informações sobre imunoenaios em patentes

Patentes relacionadas a imunoenaios foram procuradas no portal Google Patents. O termo de busca utilizado foi o “(C07K14/45) (A61K39/002)”. Esse termo contém dois códigos que são categorias de patentes e representam os termos “Toxoplasma” e “Protozoa antigens”, respectivamente. Alguns documentos de patentes mostram as sequências protéicas de interesse dentro de imagens ou com formatação fora dos padrões, impossibilitando a extração do conteúdo das sequências através da leitura direta dos *pdfs* utilizando o módulo de *Python PyPDF2*. Nesses casos, a ferramenta de reconhecimento óptico *Tesseract* (SMITH, 2007) foi empregada na extração textual. Scripts *in-house* (desenvolvidos especificamente para este trabalho) foram utilizados como etapa final de normalização da formatação das sequências em todos os casos.

### 7.3 Análise de sequências protéicas

Um script em *Python* foi desenvolvido para mostrar todas as substituições de aminoácidos observadas em um grupo de isolados de *T. gondii*. Considerando um determinado gene de interesse, é realizada a tradução protéica da região codificante (CDS) do gene apresentada por cada isolado. Em seguida, é montada uma imagem que mostra a sequência consenso em

primeiro plano, destacando as posições de aminoácido que apresentam variação, considerando as sequências de todos os isolados analisados.

A predição de epítomos lineares foi realizada através do *software* BepiPred 2.0 (JESPERSEN et al., 2017). A ferramenta foi executada através de requisições ao servidor da IEDB, que oferece esse serviço.

## 7.4 Análise de estruturas protéicas

A metodologia da etapa de análise estrutural descrita nesta seção está resumida na Figura 9.

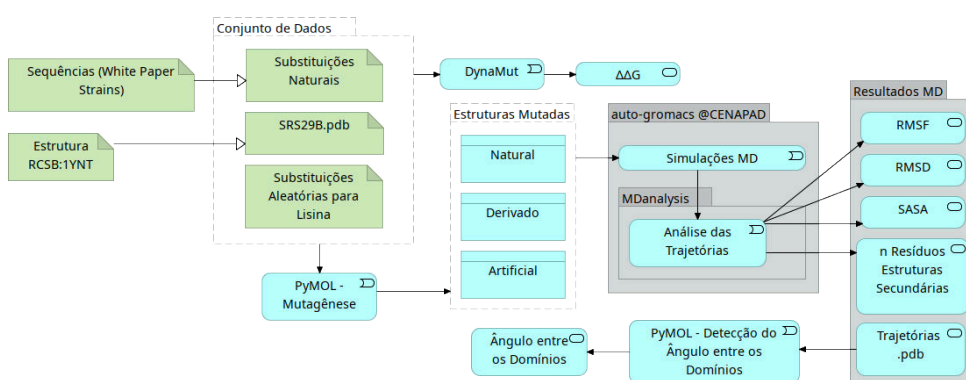


Figura 9 – Resumo da parte estrutural da metodologia.

Nota: *auto-gromacs* refere-se ao script disponibilizado em (ARAUJO; BANSAL; MERUGU, 2021). *@CENAPAD* refere-se à parte da metodologia executada no CENAPAD.

### 7.4.1 Obtenção das estruturas moleculares

A proteína SRS29B de *T. gondii* foi obtida do banco de dados RCSB, através do código 1YNT. Essa estrutura contém duas cópias de SRS29B, representadas nas cadeias F e G da estrutura. O RMSD pode ser utilizado para mensurar variações gerais entre estruturas protéicas. O RMSD entre a cadeia F e G de 1YNT é de 0.031Å, que representa uma diferença estrutural irrelevante. A cadeia F foi escolhida para as análises realizadas neste trabalho. A Figura 10 mostra a estrutura original de SRS29B.

As estruturas simuladas foram divididas em três grupos. O grupo *Natural* consiste nas variantes naturalmente observadas de SRS29B. O grupo *Derivada* consiste em combinações arbitrárias de mutações observadas naturalmente. O grupo *Artificial* contém estruturas com mutações não observadas naturalmente. Essas mutações foram criadas através de substituições aleatórias para o aminoácido lisina (K). As sequências naturais consideradas correspondem às variantes de SRS29B observadas nos bancos de dados de seqüências da ToxoDB.

Para a aplicação das mutações na estrutura original de SRS29B extraída da entrada 1YNT da RCSB foi utilizado o protocolo de mutagênese do software PyMOL. Esse assistente

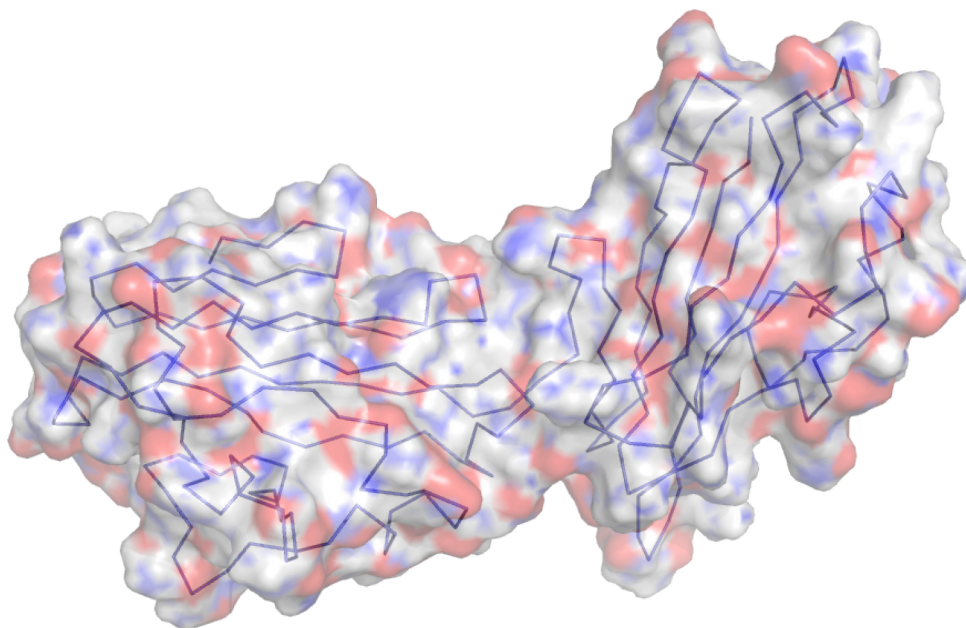


Figura 10 – Estrutura de SRS29B conforme a cadeia F de 1YNT.

Nota: A linha azul representa a cadeia principal da proteína.

de mutagênese substitui o aminoácido automaticamente e escolhe o rotâmero mais adequado dentre os que possui em sua biblioteca. As mutações em estruturas foram numeradas de acordo com a cadeia de aminoácidos da estrutura, não do transcrito primário.

#### 7.4.2 Execução da dinâmica molecular

Cada estrutura obtida foi submetida a uma simulação de dinâmica molecular independente, utilizando o software GROMACS (SPOEL et al., 2005). Um script em Python para coordenar todas as etapas do processo de DM foi desenvolvido à partir de um script encontrado no GitHub (ARAUJO; BANSAL; MERUGU, 2021). Como primeiro passo da simulação, cada estrutura foi dissolvida em água numa caixa cúbica, através da ferramenta *gmx solvate*. Após isso, íons  $\text{Cl}^-$  e  $\text{Na}^+$  foram adicionados até a solução atingir a concentração de NaCl a 0.15M. A quantidade desses íons é calculada automaticamente pelo *gmx genion*, que também inclui os íons  $\text{Cl}^-$  ou  $\text{Na}^+$  necessários para a neutralização do sistema. O próximo passo consiste na etapa de minimização de energia (EM), uma simulação de curta duração para encontrar um conjunto de coordenadas que representa a conformação de menor energia, ou ao menos de energia razoavelmente baixa, para a proteína. Segue-se a primeira etapa de equilíbrio, outra DM de curta duração utilizando o *ensemble* NVT, que mantém constante o número de átomos, o volume e a temperatura do sistema. A segunda etapa de equilíbrio é novamente uma DM de curta duração, dessa vez utilizando o *ensemble* NPT, que mantém constante o número de átomos, a pressão e a temperatura do sistema. Após a preparação descrita, seguiu-se a simulação de dinâmica molecular principal, que é mais longa que as anteriores e cujas trajetórias são analisadas. O *ensemble* NVT é utilizado nessas simulações.

Esses passos executados seguem o protocolo descrito no tutorial *Lizozyme in Water* (LEMKUL, 2018), exceto pela concentração de soluto. O campo de força *Amber ff14SB* foi utilizado nas simulações (MAIER et al., 2015). Uma modificação nesse campo de força precisou ser aplicada para que funcionasse com as estruturas utilizadas neste trabalho. Ela consiste na correção de uma incompatibilidade com determinados estilos de nomeação de átomos em arquivos *.pdb*. A incompatibilidade pode ter sido introduzida na conversão desse campo de força, que é nativo do Amber, para o GROMACS (ZHANG, 2017). As simulações de dinâmica molecular com o software GROMACS foram realizadas com os parâmetros listados na Tabela 4.

Todas as estruturas foram submetidas a simulações principais de 400ns de duração. Para maior embasamento estatístico, o grupo de estruturas *Natural* passou por uma etapa adicional de simulações, cada uma com 4000ns de duração.

A estrutura computacional do CENAPAD da UNICAMP foi empregada para a realização das simulações moleculares e análise das trajetórias. O ambiente utilizado foi o *Lovelace*, pela disponibilidade de GPUs.

### 7.4.3 Análise das trajetórias moleculares

Scripts em Python foram desenvolvidos para coordenar a análise das trajetórias moleculares das simulações realizadas (ARAUJO; BANSAL; MERUGU, 2021), utilizando a biblioteca *MDAnalysis* (MICHAUD-AGRAWAL et al., 2010) para o cálculo de RMSD e RMSF, além da criação do gráfico de Ramachandran para validação dos ângulos diedrais de cada aminoácido das estruturas simuladas. Análises de estruturas secundárias foram feitas pelo *DSSP*, aplicado nas trajetórias através da ferramenta *gmx do\_dssp*. O cálculo de área de superfície acessível ao solvente SASA (*Solvent accessible surface area*) foi realizado utilizando o programa *free-sasa* (MITTERNACHT, 2016) em cada *frame* das trajetórias moleculares. O alinhamento de estruturas resultantes das simulações foi feito no *Pymol*. O seletor *resi 1-129* foi utilizado para alinhar apenas o primeiro domínio de SRS29B. A medida dos ângulos entre os dois domínios de SRS29B foi feita com o auxílio de uma automação para *Pymol* previamente descrita (HOLDER, 2010).

### 7.4.4 Análises estruturais alternativas

O serviço DynaMut (RODRIGUES; PIRES; ASCHER, 2018) foi utilizado para a predição da tendência de mutações individuais em desestabilizar ( $\Delta\Delta G$ ) a estrutura de SRS29B. O arquivo *.pdb* contendo SRS29B, conforme extraída da estrutura original 1YNT, foi enviado para a plataforma junto com a lista de substituições de aminoácidos. O servidor remoto demorou cerca de um mês para processar as 25 mutações enviadas.

Tabela 4 – Parâmetros utilizados na dinâmica molecular executada no GROMACS.

Etapa	Nome do parâmetro	Valor
(todos)	force field	amber14sb_parmbsc1_cufix
(todos)	water model	tip3p
NVT+NPT	nsteps	500000
NVT+NPT+MD	nstlist	30
NVT+NPT+MD	cutoff-scheme	Verlet
NVT+NPT+MD	rcoulomb	1.0
NVT+NPT+MD	rvdw	1.0
NVT+NPT+MD	pbc	xyz
NVT+NPT	define	-DPOSRES
NVT+NPT+MD	integrator	md
NVT+NPT	dt	0.001
NVT+NPT	nstxout	5000
NVT+NPT	nstvout	5000
NVT+NPT	nstenergy	5000
NVT+NPT	nstlog	50000
NVT	continuation	no
NVT+NPT+MD	constraint_algorithm	lincs
NVT+NPT	constraints	h-bonds
NVT+NPT+MD	lincs_iter	1
NVT+NPT+MD	lincs_order	4
NVT+NPT+MD	ns_type	grid
NVT+NPT+MD	DispCorr	EnerPres
NVT+NPT+MD	coulombtype	PME
NVT+NPT+MD	pme_order	4
NVT+NPT+MD	fourierspacing	0.16
NVT+NPT+MD	tcoupl	V-rescale
NVT+NPT+MD	tc-grps	Protein Non-Protein
NVT+NPT	tau_t	0.1 0.1
NVT+NPT	ref_t	300 300
NVT	pcoupl	no
NVT	gen_vel	yes
NVT	gen_temp	300
NVT	gen_seed	-1
NPT+MD	continuation	yes
NPT+MD	pcoupl	Parrinello-Rahman
NPT+MD	pcoupltype	isotropic
NPT+MD	tau_p	2.0
NPT+MD	ref_p	1.0
NPT+MD	compressibility	4.5e-5
NPT	refcoord_scaling	com
NPT+MD	gen_vel	no
MD	nsteps	200000000
MD	dt	0.002
MD	nstxout	50000
MD	nstvout	50000
MD	nstenergy	50000
MD	nstlog	5000000
MD	nstxout-compressed	500000
MD	compressed-x-grps	System
MD	constraints	none
MD	tau_t	0.1 0.1
MD	ref_t	300 300

Nota: As siglas da coluna etapa representam cada etapa de simulação que integra a metodologia da dinâmica molecular. NVT representa a primeira simulação de equilíbrio do sistema, que utiliza o *ensemble* NVT. NPT representa a segunda simulação de equilíbrio do sistema, que utiliza o *ensemble* NPT. MD representa a simulação principal, que utiliza o *ensemble* NVT.

## 8 Resultados

### 8.1 Coleta de informações sobre imunoenaios em patentes

Uma relação das patentes encontradas e os antígenos de sensibilização descritos nos respectivos documentos pode ser vista na Tabela 5. Um resumo da eficácia de alguns imunoenaios em análises heterogêneas é mostrado na Tabela 6.

Tabela 5 – Patentes que detalham antígenos de sensibilização para diagnóstico de toxoplasmose.

Antígenos de Captura	Requerente	Patente	Ano
GRA6	Koolen et al	EP0710724A2	1995
SRS29B	Biomerieux	US6372897B1	1995
GRA7	Jacobs et al	WO1997027300A1	1997
GRA8	Abbott Laboratories	WO1999061906A2	1999
SRS29B	Abbott Laboratories	WO2004031358A2	2004
SRS29B MIC2 MIC3 GRA3 GRA7 M2AP	Gargano et al	WO2006094665A1	2006
ROP1 GRA6 GRA7 GRA8	FIOCRUZ/PR	WO2018170571A1	2018

Nota: Patentes pesquisadas e obtidas no Google Patents. Patentes referenciadas pelo código de registro mundial, quando possível. A data representa a publicação mais antiga do documento. Não é possível comparar diretamente a eficácia de cada um desses imunoenaios através das informações disponíveis.

Tabela 6 – Imunoenaios comerciais, antígenos de sensibilização utilizados e sensibilidades.

Imunoensaio	Fabricante	Método	Antígenos de Captura	Sensibilidade
KRIEGER et al	FIOCRUZ/PR	ELISA	ROP1;GRA7;GRA8	100%
Vidas	Biomérieux	ELFA	SRS29B; GRA6	< S
Architect	Abbott	CMIA	SRS29B; GRA8	< S
AxSYM®	Abbott	MEIA	SRS29B*	< S

Nota: Os testes com os imunoenaios FIOCRUZ/PR, Vidas e Architect foram por KRIEGER, ZANCHIN e BASCHIROTTO (2018), enquanto imunoensaio AxSYM foi testado por Martins (2015). Valores de sensibilidade foram calculados neste trabalho.

Conforme mostrado na Tabela 6, o imunoensaio AxSYM (Abbott) utiliza SAG1 recombinante como AC. Antígenos recombinantes são produzidos através de DNA extraído do patógeno e posteriormente inserido com auxílio de um vetor de expressão gênica em um outro organismo como *Escherichia coli*, que produzirá a proteína (FUMAGALLI et al., 2018).

A companhia farmacêutica Abbott Laboratories desenvolveu o conjunto de ASs descrito na patente WO2004/031358, publicada primeiramente em 1999. A invenção consiste em um coquetel de antígenos composto por diversas variantes da proteína SRS29B. No processo de

desenvolvimento do imunoenensaio houve uma otimização das métricas de sensibilidade e especificidade através de tentativas com diferentes fragmentos de SRS29B. Esse processo viabilizou métricas aceitáveis de sensibilidade e especificidade para o imunoenensaio, apesar da utilização de SRS29B como único AC. A patente mencionada descreve as sequências genômicas e proteicas das proteínas presentes no coquetel de antígenos. As 23 sequências de variantes de SRS29B utilizadas como ASs nessa patente podem ser visualizadas e comparadas com as 6 variantes naturais conhecidas de SRS29B na Figura 11. Além das sequências de referência que são as variantes catalogadas da proteína, esse alinhamento mostra a sequência principal do antígeno de sensibilização, denominada *p30*, que é a mais longa e abrange toda a extensão do transcrito primário de SRS29B. O alinhamento também apresenta outras sequências, mais curtas e com variações em relação a *p30*. Para gerar a diversidade encontrada nesses peptídeos curtos foram utilizados recursos como a deleção ou substituição de aminoácidos individuais, e a eliminação de porções N-terminais e/ou C terminais.

Não foi possível determinar quais sequências de ASs estão presentes em quais kits de imunodiagnóstico comerciais. Todos os manuais de imunodiagnósticos comerciais contém uma lista de patentes relevantes ao produto, mas as listas observadas não incluem patentes relacionadas a ASs.

## 8.2 Análise de epítomos lineares

Considerando as 65 sequências para SRS29B disponíveis na ToxoDB (KISSINGER et al., 2003), observa-se 6 sequências únicas de nucleotídeos para SRS29B. Considerando a região do produto final de SRS29B, após as clivagens, existem 4 sequências únicas de aminoácidos. Essas sequências são diferenciadas por 6 substituições de aminoácidos, conforme mostrado na Figura 12. Todas as substituições observadas representam mudanças de carga na cadeia lateral.



Figura 12 – Trocas de aminoácido observadas na proteína SRS29B de *T. gondii*

Nota: Trocas de aminoácido observadas na proteína SRS29B de *T. gondii*, considerando as cepas disponíveis na ToxoDB. Resíduos sublinhados fazem parte da estrutura PDB disponível (RCSB: 1YNT). Resíduos em roxo fazem parte de epítomos experimentalmente detectados segundo o banco IEDB. Resíduos destacados por blocos representam substituições apresentadas por ao menos um isolado do grupo observado, enquanto as cores indicam a prevalência da troca no grupo de isolados conforme descrito na legenda dentro da figura.

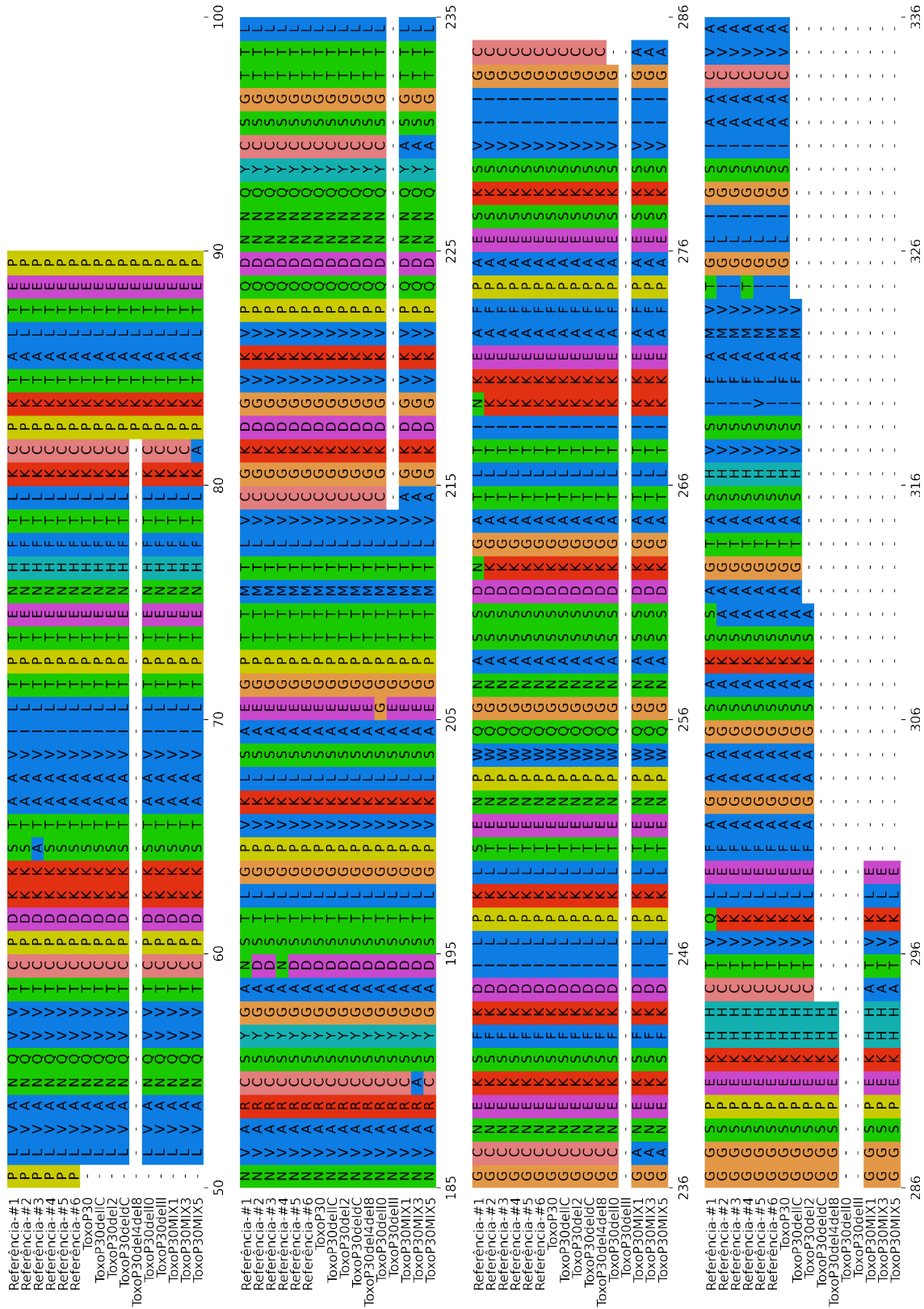


Figura 11 – Alinhamento das sequências naturais de SRS29B com as sequências descritas em patentes.

Nota: Comparação entre as 6 sequências conhecidas de SRS29B (nomeadas *Referência*) e as sequências descritas em Abbott Laboratories (1999).

Fonte: Adaptado de Abbott Laboratories (1999).



A predição de epítomos lineares indica a probabilidade de cada resíduo da proteína compor algum epítomo linear. É possível observar algumas substituições naturais de aminoácidos em áreas de alta probabilidade de ocorrência de epítomos lineares em SRS29B, conforme mostrado na Figura 13. Quanto aos epítomos lineares preditos para cada variante, é possível observar nas regiões próximas aos aminoácidos 7 e 155 que determinadas variantes divergem quanto à presença de epítomos lineares, considerando a linha de corte na pontuação de 0.5, conforme a opção padrão do *software*.

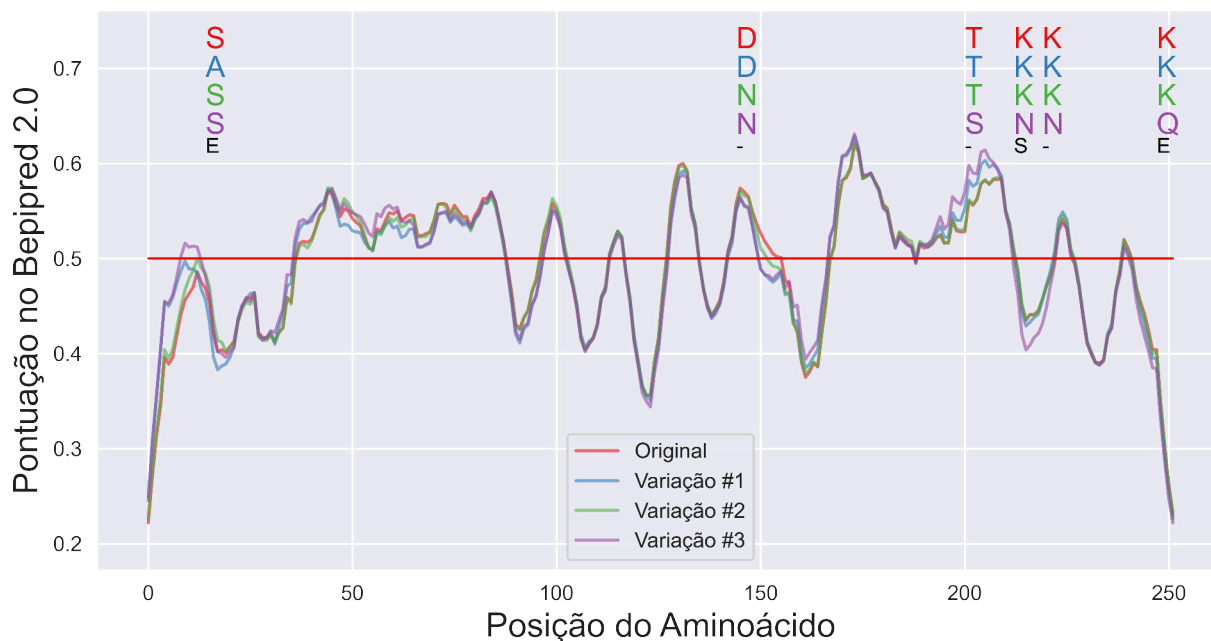


Figura 13 – Predição de Epítomos lineares em SRS29B

Nota: Pontuação atribuída pelo software BepiPred 2.0 para SRS29B mostrada em gráfico 2D contendo a posição de cada aminoácido no eixo X e a pontuação atribuída pelo BepiPred no eixo Y. Pontuações acima da linha dos 0.5 pontos indicam alta probabilidade do resíduo compor um epítomo linear de célula B. Cada linha do gráfico representa uma versão da proteína encontrada nos isolados considerados. Letras representam mutações específicas de cada sequência, mostrada em posição correspondente no eixo X. Letras pretas representam a estrutura secundária a que pertence o resíduo: fita (E), volta (S), ou indeterminada (-).

### 8.3 Análise de epítomos estruturais

A Tabela 7 resume as estruturas geradas para SRS29B e as mutações contidas em cada uma delas, através das metodologias aqui aplicadas.

A Tabela 8 indica três variantes que são representadas por poucos isolados: S15A com três isolados, I270V com dois, e F271L representada apenas pelo isolado RUB.

A predição de  $\Delta\Delta G$  realizada pelo DynaMut indica que diferentes mutações afetam a estabilidade estrutural de SRS29B em diferentes magnitudes. Um  $\Delta\Delta G$  negativo indica estabilização estrutural quando a substituição é realizada, uma tendência para a conservação da estrutura da proteína. Enquanto isso, o  $\Delta\Delta G$  positivo indica uma perda de estabilidade para a substituição, uma tendência à alteração da estrutura proteica. Conforme a Tabela 9, S15A é

Tabela 7 – Estruturas obtidas através de substituições de aminoácidos em SRS29B.

Nome	Variações	IP	Gravy
Original		7.117	-0.164
Natural 1	S15A	7.117	-0.154
Natural 2	D145N	7.791	-0.164
Natural 3	D145N; T201S; K213N; K220N; K248Q	5.961	-0.16
Derivada 1	S15A; D145N	7.791	-0.154
Derivada 2	T201S; K213N	6.381	-0.163
Derivada 3	K220N; K248Q	5.965	-0.161
Derivada 4	T201S; K213N; K220N; K248Q	5.657	-0.16
Artificial 1	V19K; L21K; T22K	8.308	-0.24
Artificial 1-1	V19K	7.789	-0.196
Artificial 1-2	L21K	7.789	-0.195
Artificial 1-3	T22K	7.789	-0.177
Artificial 2	A128K; S131K; S132K	8.308	-0.212
Artificial 3	I51K	7.789	-0.198
Artificial 4	R129K	7.116	-0.162
Artificial 5	P240K	7.789	-0.173
Artificial 6	V133K	7.789	-0.196
Artificial 7	A4K; Q6K; V8K	8.308	-0.221
Artificial 8	S230K; I233K; C235K	8.339	-0.235
Artificial 9	S46K; N48K; Q50K	8.308	-0.18

a única mutação natural com  $\Delta\Delta G$  positivo e maior que 1.0 kcal/mol, demonstrando capacidade de alterar a conformação da proteína (CAPRIOTTI et al., 2008). D145N, K220N e K248Q apresentam  $\Delta\Delta G$  positivo, mas de menor valor. As demais mutações naturais apresentam  $\Delta\Delta G$  negativo e de baixa magnitude. Mutações artificiais como S131K e S132K possuem  $\Delta\Delta G$  acima de 1.0 kcal/mol. Uma limitação do DynaMut é a ausência de predição para grupos de mutações, avaliando somente o impacto de mutações individuais.

A variação de RMSD nas estruturas simuladas são mostradas na Figura 14. Algumas proteínas simuladas apresentaram mudanças conformacionais significativas. Essas mudanças são observadas após os 200ns de simulação. A variante original de SRS29B mostrou pouca flutuação de RMSD, conforme esperado para a estrutura sem substituições. Nas simulações de 400ns, a variante caracterizada pelas substituições V19K, L21K e T22K demonstrou  $\Delta\text{RMSD} > 5\text{\AA}$ , indicando alteração estrutural significativa em relação à variante *Original*, conforme mostrado na Figura 14. Nas simulações de 4000ns realizadas para o grupo *Natural* (Figura 20), todas as estruturas demonstraram variação  $\Delta\text{RMSD} > 5\text{\AA}$  em algum momento.

Os gráficos de RMSF mostrados na Figura 15 mostram que não há picos destacados de alta movimentação em regiões específicas da proteína. Ao invés disso, existem múltiplos picos moderados que ocorrem ao longo de toda a cadeia proteica e podem ser observados em todas as estruturas analisadas.

Quanto ao número de resíduos envolvidos em estruturas secundárias, não se observa variação significativa ao longo das simulações, conforme a Figura 18. A ausência de variação

Tabela 8 – Ocorrência das variantes observadas em SRS29B nos isolados analisados.

Variante	Isolados Portadores	Genótipos SAG1 (PCR-RFLP)
S15S	59 isolados.	II/III (6); u-1 (6); I (21); Desconhecido (26)
S15A	GUY-DOS; GUY-MAT; GUY-2003-MEL	I (1); Desconhecido (2)
D145N	18 isolados.	II/III (6); u-1 (5); Desconhecido (7)
D145D	44 isolados.	u-1 (1); I (22); Desconhecido (21)
T201S	12 isolados.	II/III (6); u-1 (2); Desconhecido (4)
T201T	50 isolados.	u-1 (4); I (22); Desconhecido (24)
K213N	12 isolados.	II/III (6); u-1 (2); Desconhecido (4)
K213K	50 isolados.	u-1 (4); I (22); Desconhecido (24)
K220N	12 isolados.	II/III (6); u-1 (2); Desconhecido (4)
K220K	50 isolados.	u-1 (4); I (22); Desconhecido (24)
K248Q	12 isolados.	II/III (6); u-1 (2); Desconhecido (4)
K248K	50 isolados.	u-1 (4); I (22); Desconhecido (24)
A261S	12 isolados.	II/III (6); u-1 (2); Desconhecido (4)
A261A	50 isolados.	u-1 (4); I (22); Desconhecido (24)
I270I	60 isolados.	II/III (6); u-1 (6); I (21); Desconhecido (27)
I270V	COUG; GUY-2004-JAG1	I (1); Desconhecido (1)
F271F	61 isolados.	II/III (6); u-1 (6); I (21); Desconhecido (28)
F271L	RUB	I (1)
I275T	18 isolados.	II/III (6); u-1 (5); Desconhecido (7)
I275I	44 isolados.	u-1 (1); I (22); Desconhecido (21)

Nota: Informação sobre genótipos extraídas de Su et al. (2012). Alguns genótipos indicados como *Desconhecido* podem estar disponíveis na literatura.

significativa nesses gráficos indica que não há disrupção global das estruturas secundárias das estruturas analisadas.

Os gráficos de Ramachandran mostram que a grande maioria dos ângulos  $\Phi$  e  $\Psi$  observados nas trajetórias MD são ângulos favoráveis (Figura 22).

Observando o ângulo entre os dois domínios de SRS29B ao longo das simulações, nota-se alta correlação entre o  $\Delta$ RMSD e o ângulo entre os domínios ao longo das simulações.

A Tabela 10 resume os experimentos de dinâmica molecular conduzidos neste trabalho. Observa-se que algumas variantes não induzem alterações estruturais na proteína, enquanto outras induzem. K220N e K248Q participaram de duas estruturas simuladas cada uma, onde mostraram resultados diferentes em cada simulação.

Tabela 9 – Efeitos de substituições individuais de aminoácidos na estabilidade da proteína, de acordo com algoritmos de predição.

Mutação	Tipo da Mutação	$\Delta$ Estabilidade (kcal/mol)	Predição $\Delta\Delta G$ (kcal/mol)
S132K	Artificial	0.409	1.734
V19K	Artificial	0.438	1.585
S131K	Artificial	1.071	1.511
S15A	Natural	0.105	1.325
Q50K	Artificial	0.055	0.497
A128K	Artificial	0.27	0.39
C235K	Artificial	0.959	0.364
D145N	Natural	-0.245	0.279
K220N	Natural	-0.048	0.239
Q6K	Artificial	0.009	0.226
T22K	Artificial	-0.031	0.079
S46K	Artificial	-0.016	0.039
K248Q	Natural	-0.243	0.025
P240K	Artificial	-0.722	-0.06
L21K	Artificial	0.053	-0.06
N48K	Artificial	-0.065	-0.146
K213N	Natural	-0.107	-0.159
S230K	Artificial	-0.241	-0.244
T201S	Natural	0.034	-0.324
A4K	Artificial	-0.146	-0.451
V133K	Artificial	0.036	-0.583
I51K	Artificial	0.098	-0.616
R129K	Artificial	-0.475	-0.833
V8K	Artificial	0.21	-1.12
I233K	Artificial	-0.234	-1.971

Nota: Substituições ordenadas pela magnitude prevista para  $\Delta\Delta G$ .

Tabela 10 – Métricas finais da análise de mutações naturais em SRS29B.

Variante	Troca de Carga	Estrutura Secundária	$\Delta$ RMSD MD > 5.0 Å	$\Delta\Delta G > 0.2$
S15A	Sim	Hélice	Sim	Sim
D145N	Sim	-	Sim	Não
K220N	Sim	Folha	Sim + Não	Não
K248Q	Sim	Hélice	Sim + Não	Não
K213N	Sim	-	Não	Não
T201S	Não	-	Não	Não

Nota: Resumo dos atributos mais relevantes das mutações naturais de SRS29B, considerando todas as etapas da análise conduzida neste trabalho. A coluna  $\Delta$ RMSD MD > 5.0 Å indica resultados para mais de uma simulação, quando aplicável.

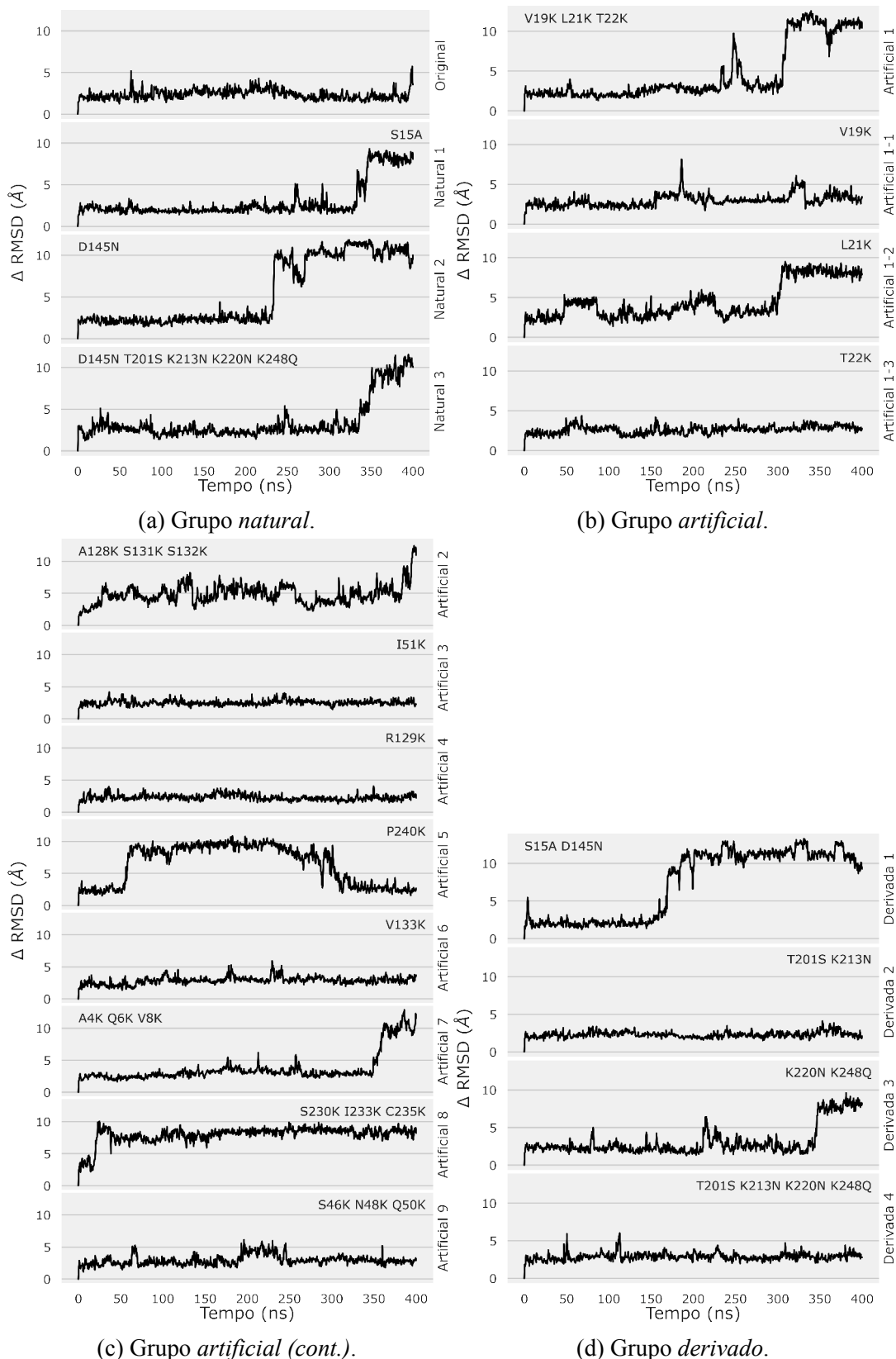


Figura 14 – RMSD das simulações moleculares do grupo de estruturas de SRS29B.

Nota: Variação de RMSD observada em simulações de dinâmica molecular de variantes mutadas da proteína SRS29B utilizando a entrada RCSB:1YNT como modelo. Cada linha do gráfico representa uma simulação, mostrando a métrica RMSD entre a conformação do frame e a conformação inicial da proteína. Simulações possuem 400ns de duração e utilizam o campo de força Amber ff14SB (MAIER et al., 2015) no software GROMACS. Os identificadores de mutação indicam quais mutações estão presentes nas estruturas representadas em cada gráfico.

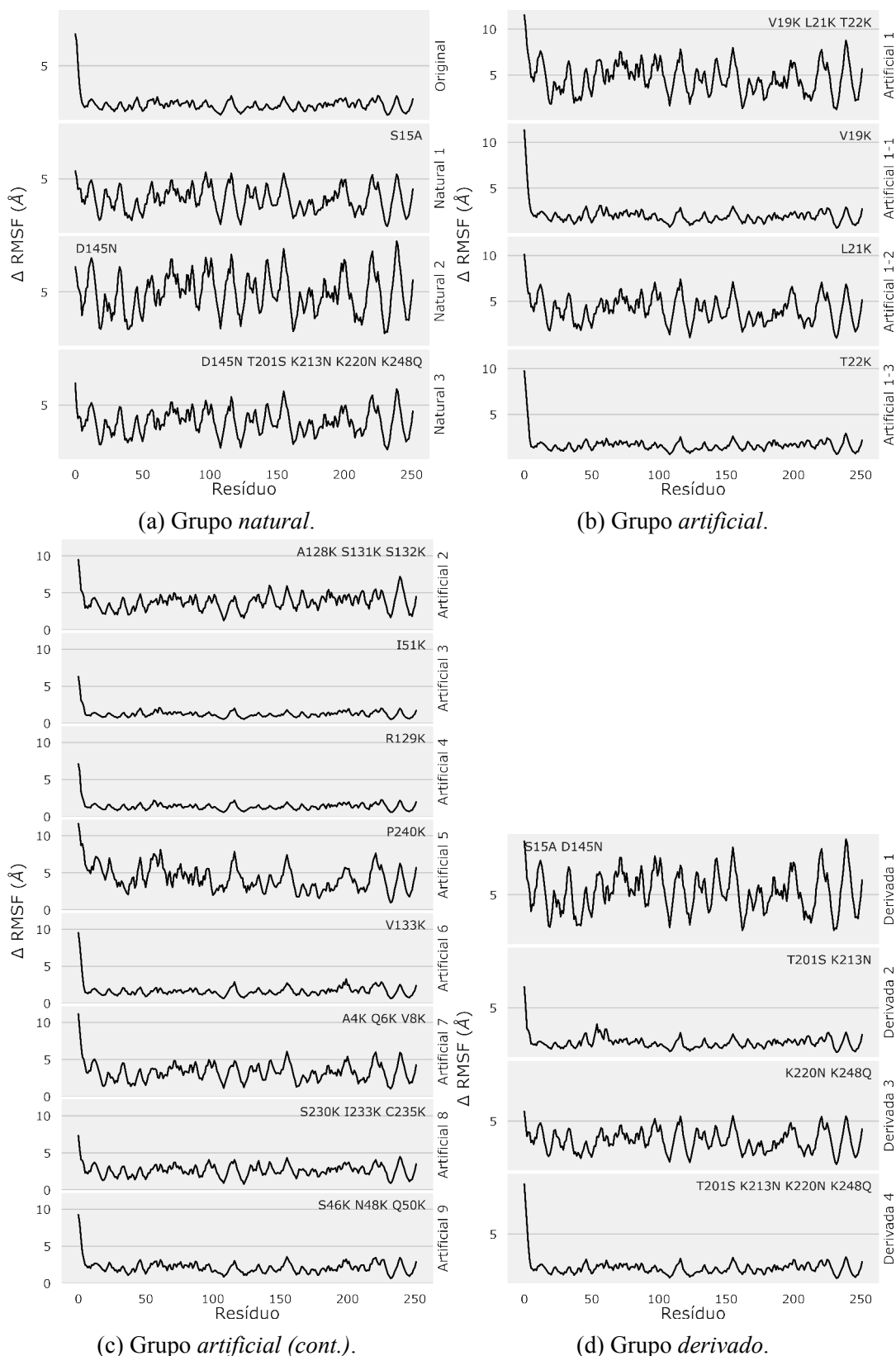


Figura 15 – RMSF das simulações moleculares do grupo de estruturas de SRS29B.

Nota: Variação de RMSF observada em simulações de dinâmica molecular de variantes mutadas da proteína SRS29B utilizando a entrada RCSB:1YNT como modelo. Cada linha no gráfico representa uma estrutura simulada. Simulações possuem 400ns de duração, o trecho entre 360ns e 380ns foi utilizado para a análise de RMSF. Os identificadores de mutação indicam quais mutações estão presentes nas estruturas representadas em cada gráfico.

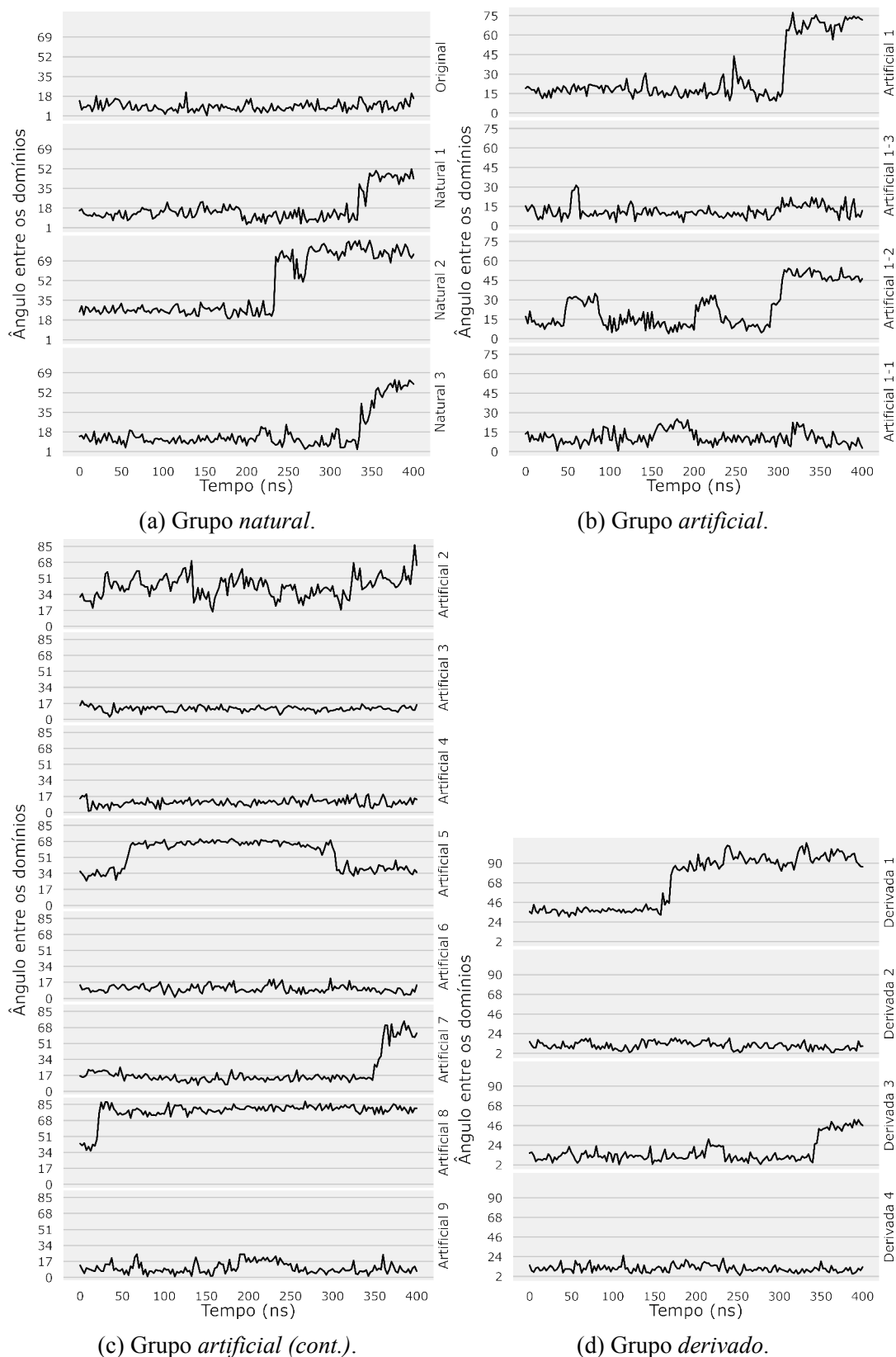


Figura 16 – Ângulo entre os domínios das simulações moleculares do grupo *natural* de estruturas de SRS29B.

Nota: Variação do ângulo medido entre os dois domínios de SRS29B observada em simulações de dinâmica molecular de variantes mutadas da proteína SRS29B utilizando a entrada RCSB:1YNT como modelo. Cada linha do gráfico representa uma simulação, mostrando a métrica RMSD entre a conformação do frame e a conformação inicial da proteína. Simulações possuem 400ns de duração e utilizam o campo de força Amber ff14SB (MAIER et al., 2015) no software GROMACS. Os identificadores de mutação indicam quais mutações estão presentes nas estruturas representadas em cada gráfico.

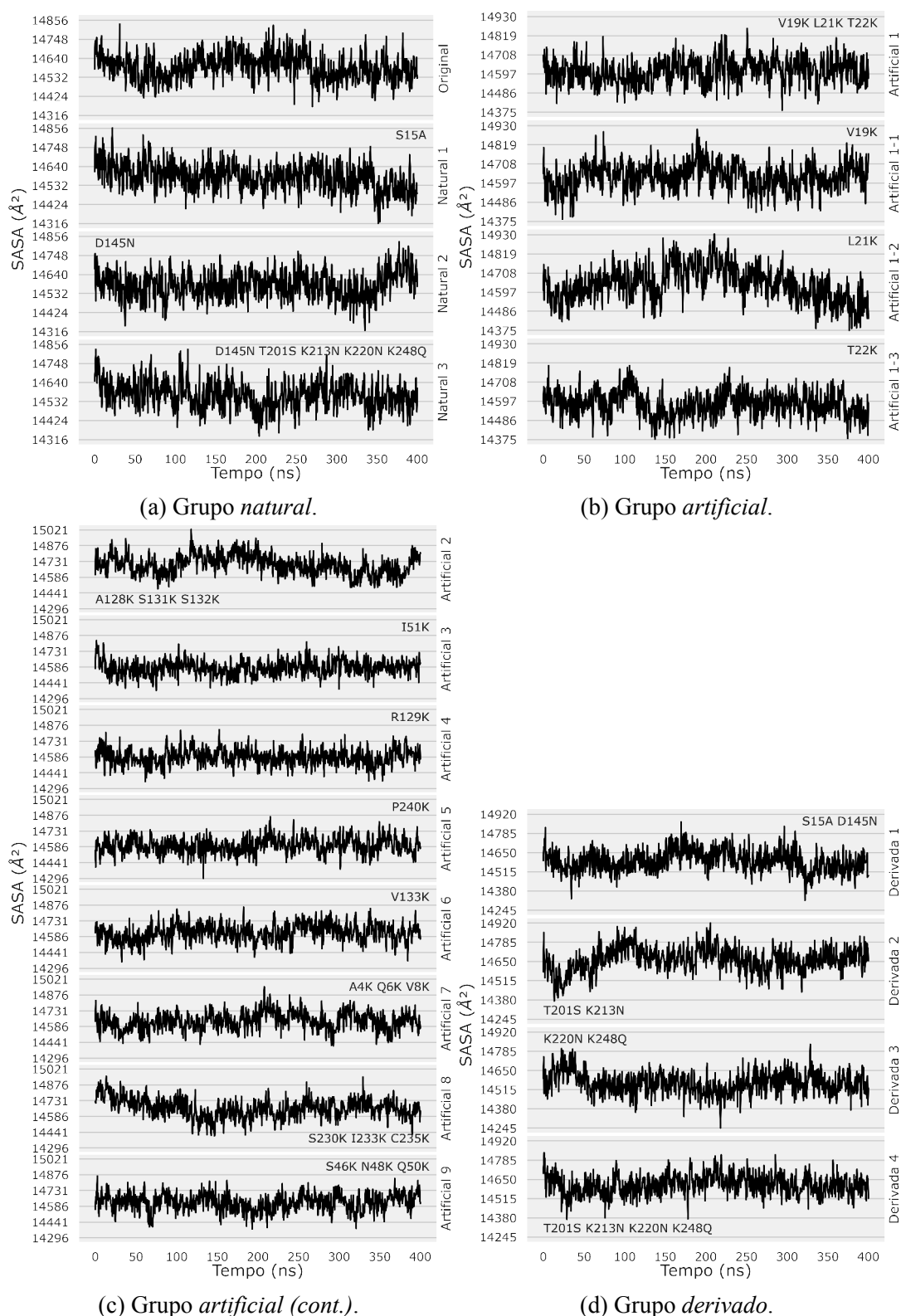


Figura 17 – SASA das simulações moleculares do grupo de estruturas de SRS29B.

Nota: Superfície total acessível ao solvente ao longo de trajetórias moleculares de variantes mutadas da proteína SRS29B utilizando a entrada RCSB:1YNT como modelo. Cada gráfico representa uma simulação. Simulações possuem 400 nanossegundos de duração e utilizam o campo de força Amber ff14SB (MAIER et al., 2015) no software GROMACS. Os identificadores de mutação indicam quais mutações estão presentes nas estruturas representadas em cada gráfico.



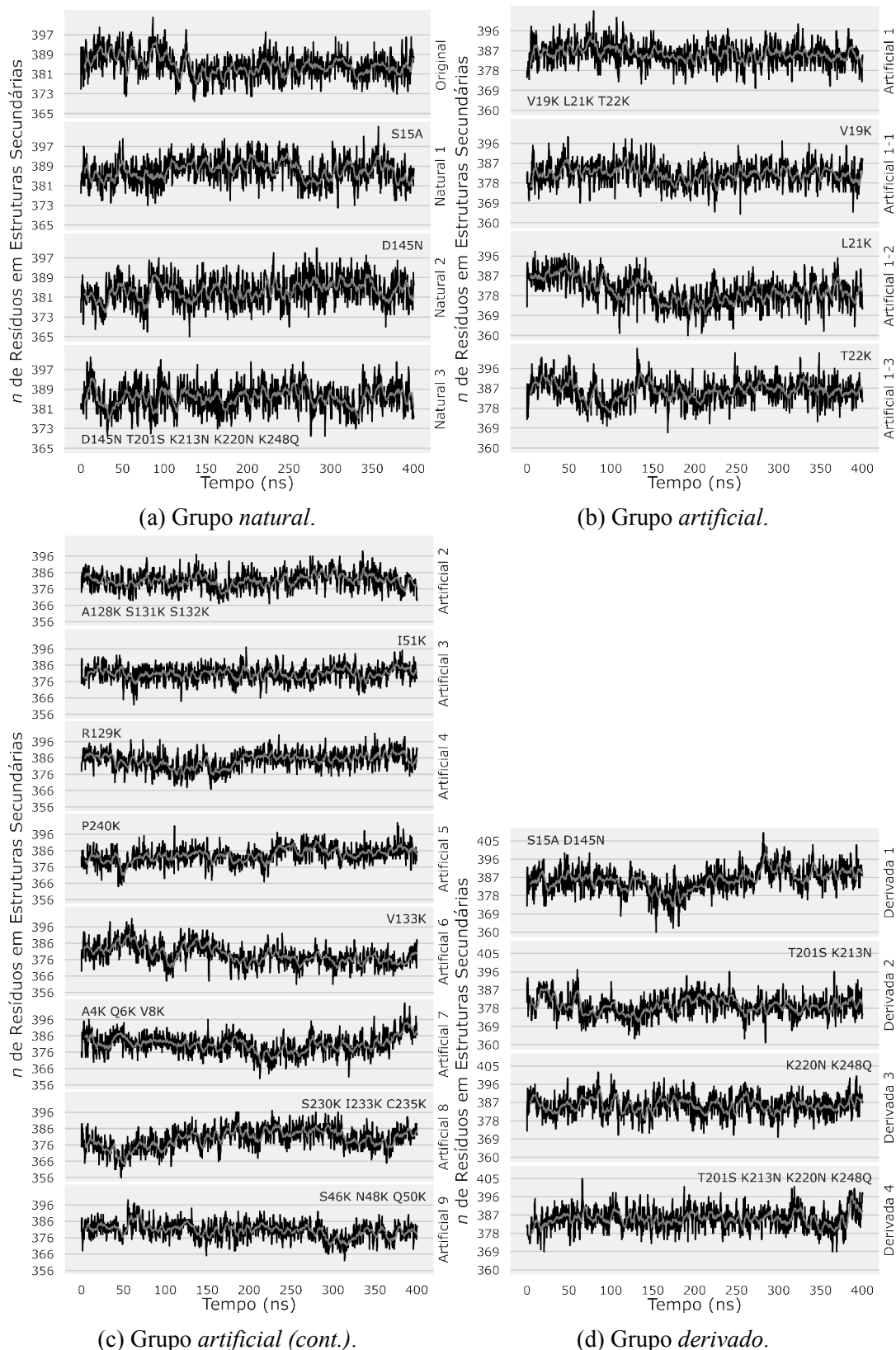


Figura 18 – Número de resíduos em estruturas secundárias das simulações moleculares do grupo de estruturas de SRS29B.

Nota: Número de resíduos envolvidos em estruturas secundárias ao longo simulações de dinâmica molecular de variantes mutadas da proteína SRS29B utilizando a entrada RCSB:1YNT como modelo. Linhas cinzas representam a média móvel dos 10 frames anteriores. Cada gráfico representa uma simulação. Simulações possuem 400 nanossegundos de duração e utilizam o campo de força Amber ff14SB (MAIER et al., 2015) no software GROMACS. Os identificadores de mutação indicam quais mutações estão presentes nas estruturas representadas em cada gráfico.

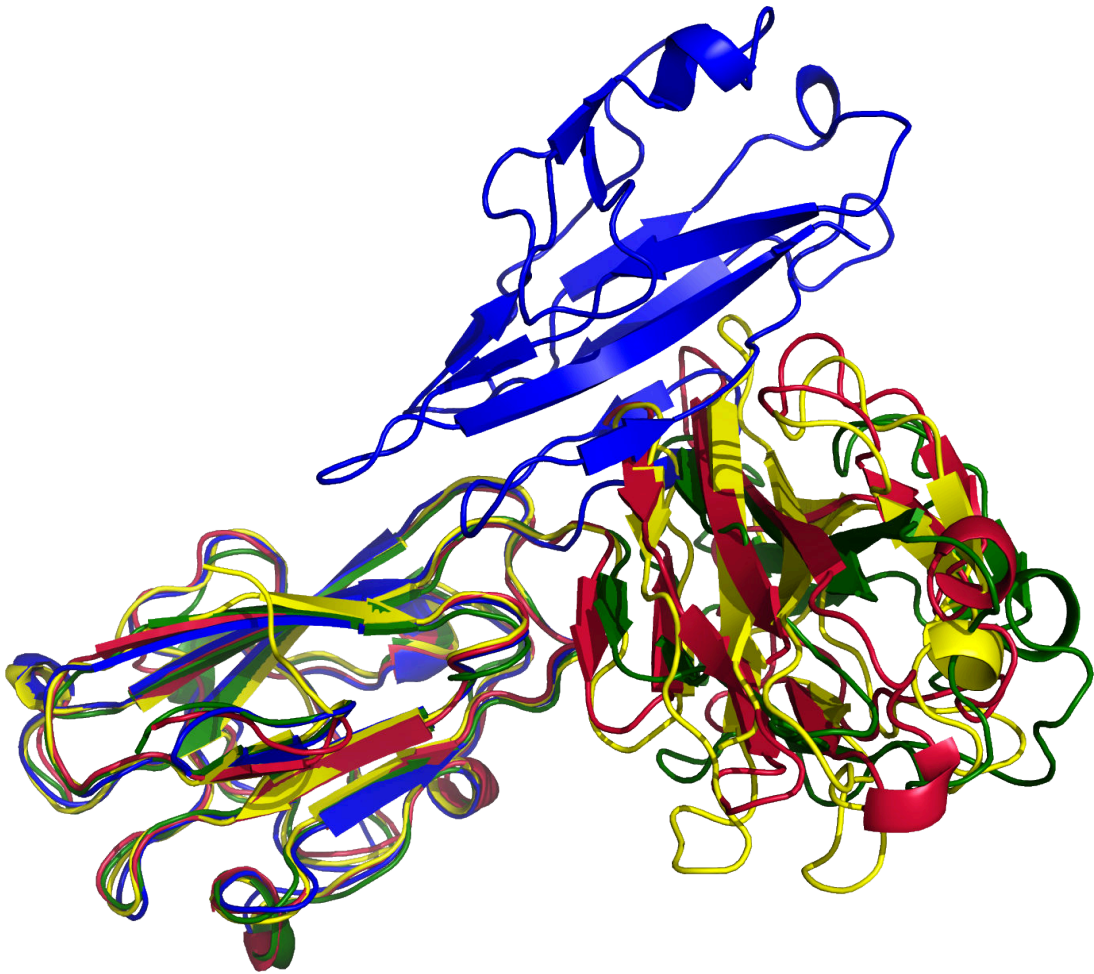


Figura 19 – Estruturas alinhadas das simulações moleculares do grupo *natural* de estruturas de SRS29B.

Nota: Alinhamento estrutural de todos os membros do grupo. Cada estrutura representa a média das posições atômicas considerando os frames entre 360ns e 380ns da trajetória molecular. Em vermelho: estrutura *Original*, verde: *Natural 1*, azul: *Natural 2*, amarelo: *Natural 3*. Esta imagem não retrata toda a variação observada ao longo das trajetórias, apenas ilustra a variação do ângulo entre os dois domínios de SRS29B conforme observada no conjunto de trajetórias de 400ns do grupo *Natural*. Os identificadores de mutação indicam quais mutações estão presentes nas estruturas representadas em cada gráfico.

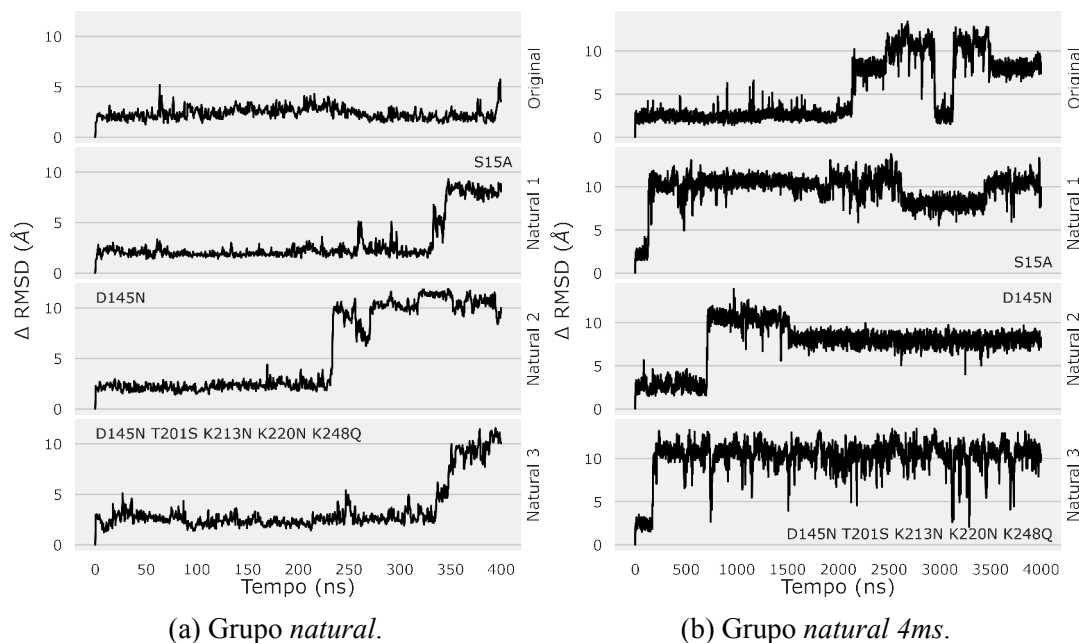


Figura 20 – RMSD das simulações moleculares do grupo de estruturas de SRS29B.

Nota: Variação de RMSD observada em simulações de dinâmica molecular de variantes mutadas da proteína SRS29B utilizando a entrada RCSB:1YNT como modelo. Cada linha do gráfico representa uma simulação, mostrando a métrica RMSD entre a conformação do frame e a conformação inicial da proteína. Simulações possuem 4000ns de duração e utilizam o campo de força Amber ff14SB (MAIER et al., 2015) no software GROMACS. Os identificadores de mutação indicam quais mutações estão presentes nas estruturas representadas em cada gráfico.

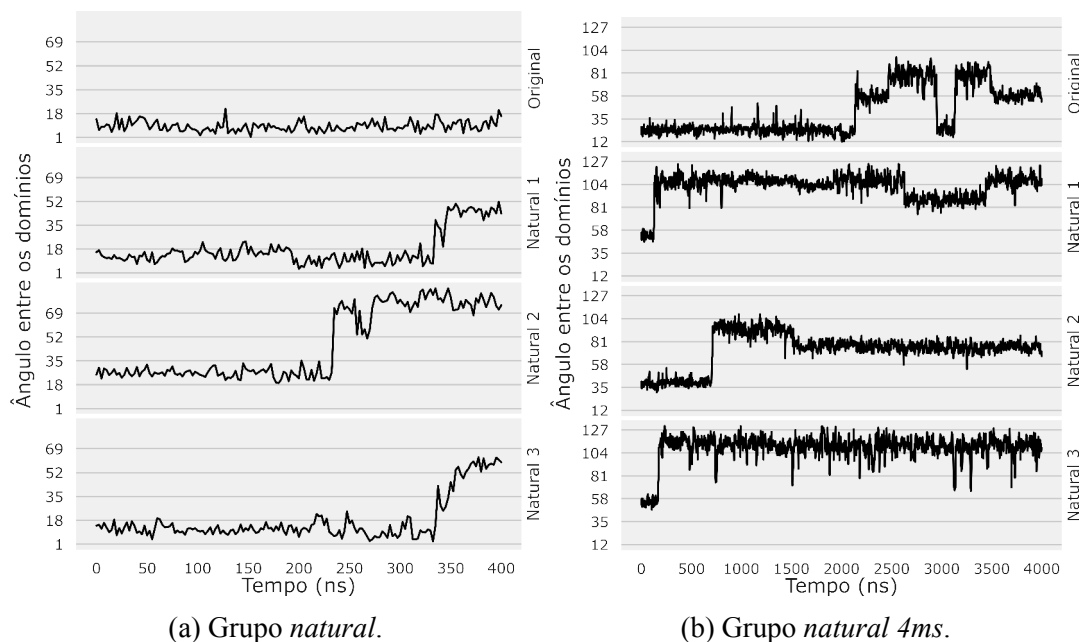


Figura 21 – Ângulo entre os domínios das simulações moleculares do grupo *natural* de estruturas de SRS29B.

Nota: Variação do ângulo medido entre os dois domínios de SRS29B observada em simulações de dinâmica molecular de variantes mutadas da proteína SRS29B utilizando a entrada RCSB:1YNT como modelo. Cada linha do gráfico representa uma simulação, mostrando a métrica RMSD entre a conformação do frame e a conformação inicial da proteína. Simulações possuem 400ns de duração e utilizam o campo de força Amber ff14SB (MAIER et al., 2015) no software GROMACS. Os identificadores de mutação indicam quais mutações estão presentes nas estruturas representadas em cada gráfico.

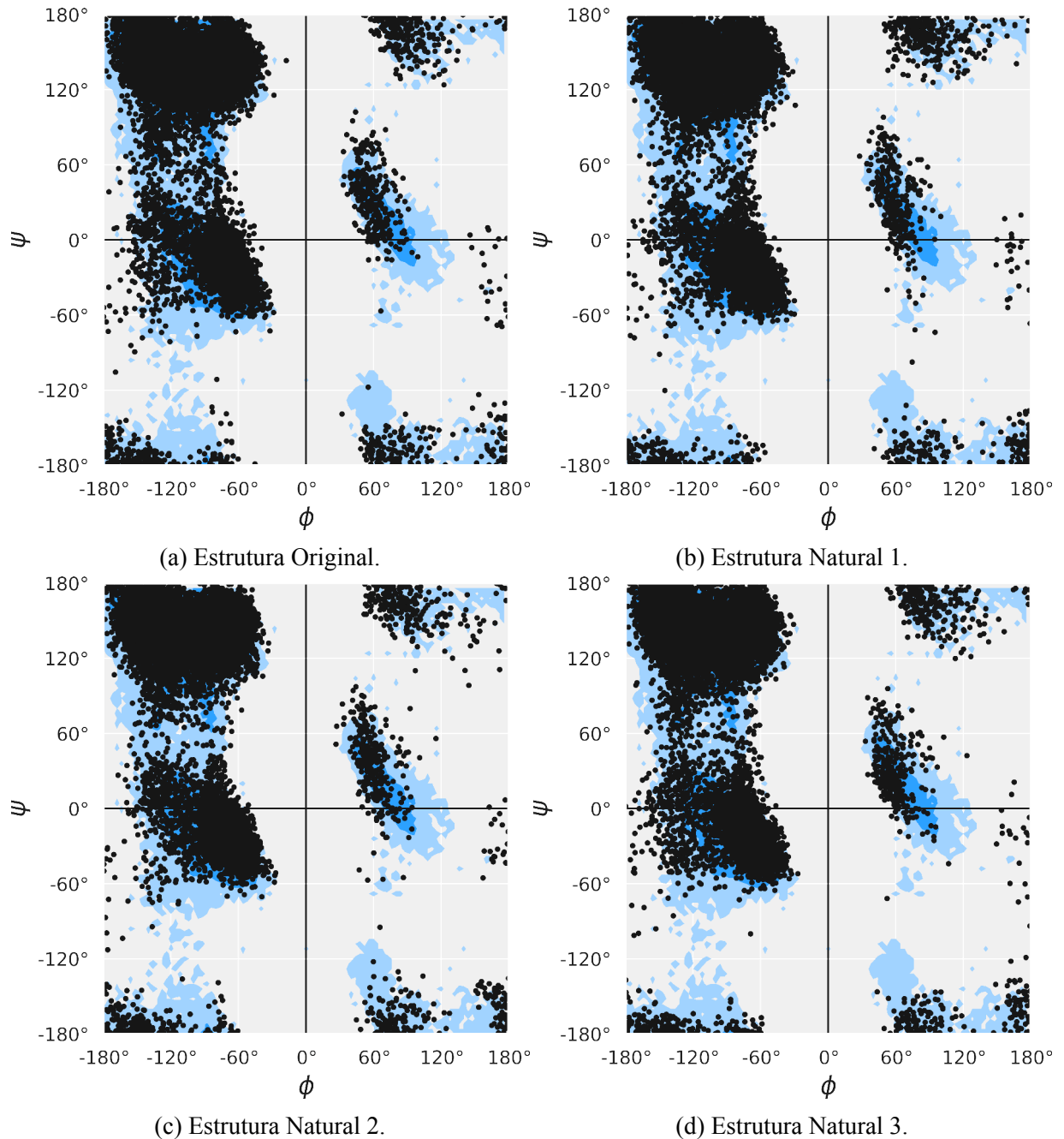


Figura 22 – Gráficos de Ramachandran para as trajetórias das variantes naturais de SRS29B.

Nota: Gráficos de Ramachandran para a trajetória molecular de 400ns da estrutura *Original* do grupo *Natural*. Os ângulos  $\Phi$  e  $\Psi$  de cada resíduo de cada frame são indicados por pontos pretos no plot. Zonas em azul indicam ângulos favoráveis e naturalmente observadas para os diferentes tipos de aminoácidos, sobrepostos.

## 9 Discussão

### 9.1 Possíveis causas das baixas sensibilidades observadas

No trabalho anterior do grupo foi observada uma discrepância entre os resultados dos diagnósticos por imunoenaios e os resultados dos diagnósticos por PCR. A proporção de resultados positivos pelos imunoenaios comerciais ficou abaixo da esperada. Análises adicionais indicaram a presença de DNA de *T. gondii* em 22 de 24 (91,66%) amostras soronegativas no imunoenamo Axsym (Abbott). Na detecção de DNA por PCR, considerou-se a presença de ao menos um dos marcadores do conjunto ITS1, NTS2, B1 e SAG3 (MARTINS, 2015). Diversos fatores poderiam explicar a divergência de sensibilidade observada entre os dois métodos de diagnóstico. Alguns deles serão discutidos neste capítulo.

#### 9.1.1 Janela de produção de imunoglobulinas

As imunoglobulinas são produzidas pelo hospedeiro em janelas de tempo específicas, considerando o primeiro contato com o patógeno como o marco zero dessa escala temporal. Diferentes patógenos induzem o hospedeiro a padrões ligeiramente diferentes de produção de imunoglobulinas. Os anticorpos do isótipo IgM tornam-se indetectáveis nas amostras após um certo tempo do início da infecção. IgMs possuem relativa baixa afinidade pelo patógeno e são substituídos, no decorrer da resposta imunológica, por outros isótipos com mais especificidade, como IgG. Supõe-se que o IgG produzido contra um determinado patógeno continue sendo produzido por tempo indeterminado, mantendo-se em níveis detectáveis nas amostras sorológicas durante esse período, que pode estender-se por toda a vida do hospedeiro. Essa longa produção de IgG pode ser ainda mais acentuada no caso da toxoplasmose, que é uma infecção em que não ocorre a eliminação completa do parasito do organismo hospedeiro, podendo ainda ocorrer reativações ocasionais da infecção que são contidas pelo sistema imunológico. No entanto, não há a confirmação científica tanto da permanência de *T. gondii* quanto de quantidades detectáveis de IgG circulante contra *T. gondii* por toda a vida de todos os pacientes. Dessa forma, pacientes que foram infectados pelo parasito muito tempo antes da coleta das amostras podem apresentar níveis indetectáveis de IgG. Alternativamente, há um período de latência entre o momento da infecção e o momento em que os níveis de IgG atingem níveis detectáveis. Amostras coletadas entre esses pontos devem ser negativas para IgG, podendo apresentar níveis detectáveis de DNA do parasito. Esses pacientes devem ser positivos para IgM, considerando que o período de latência de IgM é mais curto que o de IgG.

### 9.1.2 Falhas metodológicas em imunoenaios

Falhas metodológicas na execução do imunoenensaio incluem a utilização de reagentes expirados, erros no cálculo de diluição das amostras (DALY et al., 2004), ou ainda erros no cálculo do *cutoff* dos níveis de absorbância em sistemas que utilizam tal mecanismo. A categoria ainda inclui erros de manejo das amostras, que pode levar a contaminações. Os sistemas de imunoenaios automatizados como o AxSym e Architect minimizam a possibilidade de contaminações ao reduzir o manuseio das amostras (ISMAIL, 2009).

### 9.1.3 Incompatibilidade de antígenos de sensibilização

A variabilidade genética do parasito pode influenciar a imunocompatibilidade de seus antígenos. Esse fenômeno se chama *antigenic drift* e é uma estratégia evolutiva do parasito para evadir o sistema imune de seus hospedeiros. O *antigenic drift* de alguns vírus são bem investigados na literatura (YEWDELL, 2021). Principalmente aqueles que demandam esforços de imunização coletiva de monitoramento epidemiológico pelas autoridades de saúde pública. Os protozoários em geral evoluem em escalas de tempo maiores que os vírus. Isso diminui a importância da vigilância em tempo real da variação de antígenos nesse tipo de organismo, mesmo considerando os protozoários de importância clínica maior que *T. gondii*, como *T. cruzi* e *P. falciparum*. Dessa forma, a variação de antígenos encontrada em protozoários refere-se mais à variação já existente que ainda não foi catalogada do que à variação que ocorre no espaço de um ano entre as variantes do vírus *Influenza A*, por exemplo.

A precisão dos imunoenaios depende da alta afinidade entre anticorpos e antígenos (BOURASSA et al., 2021). Considerando os imunoenaios comerciais investigados neste trabalho, não é possível determinar quais imunoenaios comerciais utilizam ASs descritos em quais patentes, e ainda quais subconjuntos de ASs descritos nas patentes são efetivamente utilizados nos imunoenaios. Consideramos que tanto a Biomerriéux e a Abbott utilizam os ASs descritos em suas próprias patentes. Em alguns casos, a informação sobre qual proteína do parasito foi utilizada como AS em determinado imunoenensaio pode ser obtida em artigos publicados (LIU et al., 2019; REYNOSO-PALOMAR; MORENO-GÁLVEZ; VILLA-MANCERA, 2019; BACHAN et al., 2017).

A proteína SRS29B é reconhecida desde a década de 90 como o antígeno imunodominante de *T. gondii* (CESBRON-DELAUW et al., 1994), sendo bem caracterizada e bastante utilizada como antígeno de sensibilização em imunoenaios para diagnóstico de toxoplasmose, conforme mostrado na Tabela 5. Essa proteína é muitas vezes utilizada como único AS em um imunoenensaio. Apesar de variantes do parasito sem a proteína SRS29B ( $\Delta$ SRS29B) preservarem sua virulência (CHARLES; CALLEGAN; BLADER, 2007), a proteína indica possuir função no processo de invasão celular ao promover aderência entre o taquizoíto e a célula alvo através de sua afinidade com proteínas de membrana do hospedeiro (WANG; YIN, 2014). SRS29B é altamente expressa em todos os isolados de *T. gondii* que foram submetidos a essa análise

(BOOTHROYD et al., 1998). Talvez pelo papel na invasão celular, SRS29B apresenta alta conservação de sua sequência proteica em relação a outras proteínas de *T. gondii*. Apesar disso, a Tabela 6 inclui um imunoenensaio que não utiliza SRS29B como antígeno de sensibilização e esse apresentou a maior sensibilidade quando utilizado em pacientes brasileiros. Isso pode estar relacionado com uma diversidade desconhecida no gene SRS29B em isolados endêmicos do Brasil. Em diagnósticos por PCR realizados anteriormente pelo grupo em amostras de pacientes de Campos dos Goytacazes, foi observado que o marcador SAG1 falhou em amplificar o material genético em algumas amostras (MARTINS, 2015). Isso aconteceu mesmo em amostras onde outros marcadores como NTS2 e B1 amplificaram a amostra. Essa observação indica uma certa variação genotípica na região SAG1 (SRS29B) das cepas endêmicas da região norte fluminense.

Outras proteínas de *T. gondii* também são utilizadas como AS em imunoenensaio, como GRA7, GRA8. Algumas combinações de AS se mostram mais eficientes para detecção de IgM, outras para detecção de IgG (LIU et al., 2011).

#### 9.1.4 Isolados não clonais de *T. gondii* no mundo

A distribuição de genótipos de *T. gondii* varia pelas diferentes regiões geográficas do planeta. De maneira geral, podemos falar em regiões que possuem maior representatividade de isolados clonais e regiões onde predominam os isolados não-clonais. Cada região apresenta sua própria proporção de isolados não-clonais. Essas duas métricas estão representadas para alguns países na Figura 23.

A análise de prevalência mostrada nessa figura baseia-se em imunoenensaio realizados no subgrupo populacional das gestantes. Essa população é rotineiramente submetida a diagnósticos para toxoplasmose por questões de saúde do feto e considera-se que representa a população geral. É possível observar no gráfico três grupos distintos de países. O *Grupo 2* reúne países que apresentam as mais altas proporções de isolados não-clonais em conjunto com as menores taxas de prevalência de *T. gondii*. É importante notar que certos países possuem relativa pouca diversidade de genótipos do parasito, mesmo tendo alta proporção de isolados atípicos. No caso da China, grande parte de seus isolados pertencem ao *Genótipo #9*, apelidado de *Chinese 1*. Isolados desse haplogrupo apresentam alta virulência em murinos (GAO et al., 2017). Não se sabe existem divergências genotípicas importantes entre isolados de um mesmo haplogrupo. Conforme mostrado na subseção 4.1.2, o locus SAG1 (SRS29B) dos isolados do *Genótipo #9* apresenta o alelo *u-1*, que é atípico. No entanto, não são observadas substituições de aminoácidos exclusivas desse isolado em SRS29B. Uma análise de genética populacional colocou o grupo *Chinese 1* como mais próximo das variantes clonais<sub>II</sub> do que das outras variantes clonais ou mesmo dos isolados atípicos (AMOUEI et al., 2019).

Uma hipótese pode ser levantada ao analisar a situação dos países do *Grupo 2*. Nesses países, a baixa incidência de toxoplasmose observada pode estar relacionada com a falta de sensibilidade dos imunoenensaio em pacientes infectados por isolados não clonais. Outra hipótese



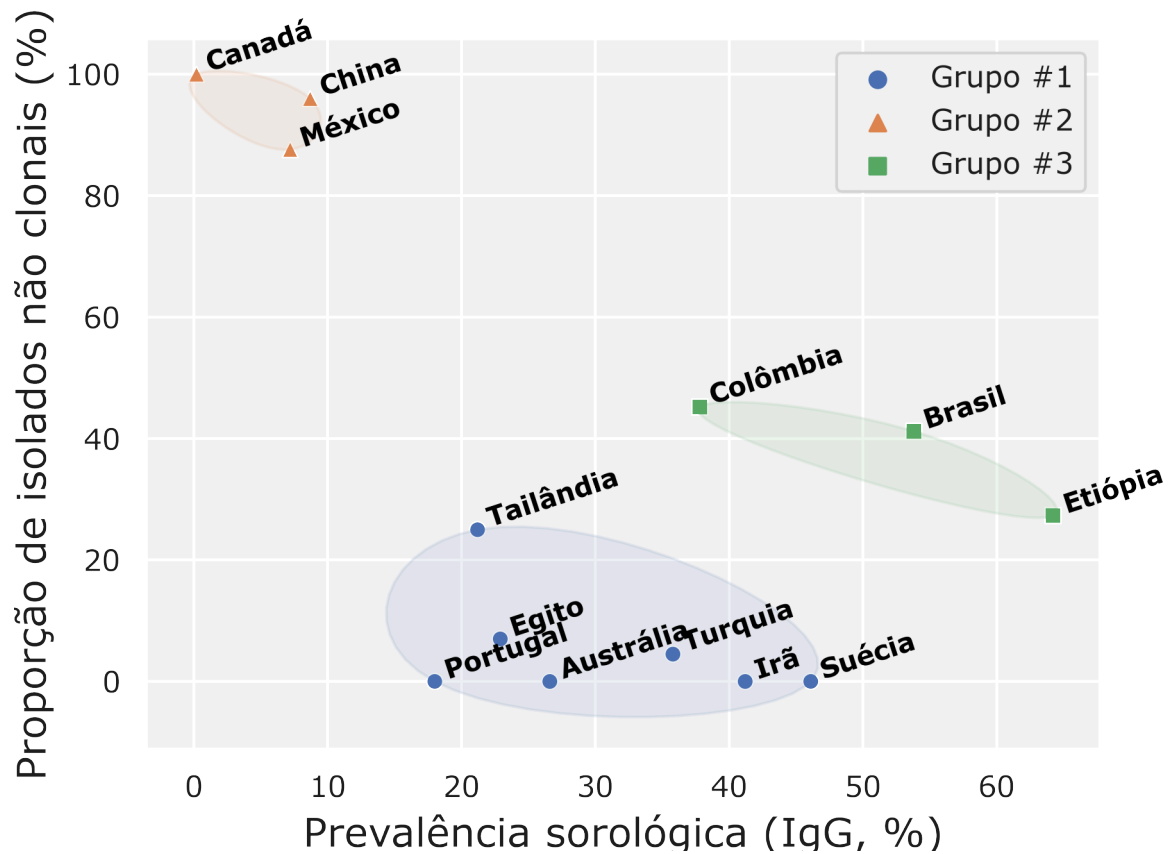


Figura 23 – Prevalência de IgG para toxoplasmose e proporção de isolados atípicos em alguns países.

Nota: Prevalência de IgG em gestantes, compiladas por Bigna et al. (2020). Proporção de isolados atípicos em cada país obtidos em Amouei et al. (2019). Os países que estão contidos em ambas as fontes de dados são mostrados no gráfico.

é que poucos imunoenaios são realizados nesses países para fins epidemiológicos. Dentre esses países, o México apresenta grandes alterações entre estudos realizados em regiões diferentes, com prevalências entre cerca de 2% e 47% (HERNÁNDEZ-CORTAZAR et al., 2013). Estudos de prevalência utilizando métodos mais precisos que o imunoenensaio, como a amplificação de fragmentos PCR, precisam ser realizados nas populações dos países do *Grupo 2* para validar ou não essa hipótese.

## 9.2 Impacto de mutações em epítomos

Epítomos com mutações em seu domínio podem escapar da afinidade de anticorpos específicos aos epítomos originais (HOSSAIN; UEDA, 2016). Essa variação genética nas proteínas de superfície é reconhecida como parte de uma estratégia evolutiva de patógenos para a evasão dos mecanismos imunológicos do hospedeiro (DEITSCH; LUKEHART; STRINGER, 2009).



### 9.2.1 Epítomos lineares

Substituições de aminoácidos afetam apenas epítomos lineares que abrangem a posição da mutação. Diversos epítomos lineares de células B para SRS29B estão disponíveis na IEDB. Observa-se que toda a extensão do produto final de SRS29B está coberta por esse tipo de epítomo (Figura 12).

O programa BepiPred 2.0 foi utilizado para a predição de epítomos lineares. Ele indica a probabilidade de cada aminoácido da sequência fazer parte de um epítomo de células B (JESPERSEN et al., 2017). Esse software utiliza classificadores *Random Forest* treinados em datasets de sequências protéicas e epítomos experimentais fornecidos pelo IEDB. Existem 6 pontos onde substituições são observadas em SRS29B, todos eles contidos em epítomos lineares experimentalmente verificados. Muitos deles envolvem substituições entre aminoácidos com cargas diferentes. Essas substituições possuem maior probabilidade de afetar esse tipo de epítomo, considerando o papel da atração eletrostática para as interações moleculares.

Não há uma metodologia *in silico* específica para verificar o impacto de substituições na antigenicidade de epítomos lineares. No entanto, existem variações nas curvas de predição para as diferentes sequências analisadas (Figura 13). Isso pode indicar alterações na imunogenicidade dos epítomos lineares preditos para SRS29B, mesmo que fracas.

### 9.2.2 Epítomos conformacionais

O antígeno SAG1/SRS29B de *T. gondii* está representado no banco de dados RCSB pela entrada 1YNT, que mostra duas proteínas SRS29B conjugadas a um anticorpo em resolução de 3.5Å (GRAILLE et al., 2005). A estrutura foi obtida pela cristalização do complexo, sendo considerada uma fiel representação da estrutura real. SRS29B possui dois domínios globulares, cada um representando cerca de metade da proteína. Essa divisão não é aparente ao olhar a superfície da estrutura. Não existem epítomos conformacionais experimentalmente identificados de SRS29B no catálogo da IEDB. Assim, alterações conformacionais vistas em dinâmica molecular foram utilizadas como *proxy* (variável substituta) para a detecção de alterações na imunocompatibilidade da proteína. Considera-se a correlação entre alterações conformacionais e de imunocompatibilidade e a dificuldade em realizar experimentos *in silico* mais apropriados para determinação de imunocompatibilidade, que envolveriam docking molecular (simulações de interações intermoleculares) entre antígeno e Igs e síntese em larga escala de estruturas de Igs.

Diversas métricas podem ser extraídas de trajetórias moleculares resultantes de experimentos de dinâmica molecular. O RMSD (*Root-mean-square deviation (of atomic positions)*) indica a mudança média nas posições dos átomos centrais (carbono alfa) de cada aminoácido da proteína. Considerando a posição inicial como referência, essa métrica pode ser calculada para cada quadro de uma trajetória molecular. Dessa forma, essa análise informa a variação geral da

estrutura molecular ao longo da simulação. Variações significativas de RMSD ( $\Delta\text{RMSD} > 5\text{\AA}$ ) ao longo do gráfico indicam alterações estruturais significativas. No contexto deste trabalho, as mutações presentes em variantes que apresentam  $\Delta\text{RMSD}$  significativo são consideradas como disruptivas da estrutura de SRS29B. A estabilidade estrutural apresentada pela variante sem mutações indica ausência de grandes problemas com a metodologia dos experimentos de dinâmica molecular realizados neste trabalho. Isso inclui validação dos parâmetros da simulação descritos na Tabela 4.

Para comparar os gráficos de RMSD com as predições de  $\Delta\Delta G$ , é preciso considerar quais mutações estão presentes em cada estrutura. S15A é a única mutação presente na variante *Natural 1*, que apresenta  $\Delta\text{RMSD}$  significativo. Em concordância com essa observação, S15A é uma das mutações naturais que apresenta maior  $\Delta\Delta G$  de acordo com o *DynaMut*. S15A consiste numa troca de um aminoácido polar para um hidrofóbico, um conhecido fator para mudanças estruturais em proteínas (RIDGLEY et al., 2014). D145N, por sua vez, representa uma troca de aminoácido carregado negativamente para um polar sem carga. Essa substituição foi predita como estabilizadora pelo *DynaMut*, e não alterou a conformação de SRS29B nas trajetórias moleculares de forma significativa. No geral, observa-se uma concordância entre os resultados da dinâmica molecular e a predição de  $\Delta\Delta G$  realizada pelo *DynaMut*. Mutações descritas como desestabilizadoras pelo *DynaMut* integram estruturas que apresentam grandes alterações conformacionais nas trajetórias MD.

Uma outra métrica é a SASA (*Surface accessible solvent area*). Um valor alto de SASA para determinado aminoácido significa que ele está espacialmente disponível para os ligantes, que é um dos fatores de predição do papel de um aminoácido em epítomos (DALL'ANTONIA et al., 2013). Não foram observadas alterações significativas de SASA nas estruturas analisadas.

É importante lembrar que todas as mudanças conformacionais relevantes observadas ocorrem após os 200ns de simulação. Essas movimentações seriam indetectáveis em simulações de duração menor que 200ns. O protocolo de dinâmica molecular não reproduz perfeitamente as movimentações moleculares observadas na natureza (CHILDERS; DAGGETT, 2018). Assim, evidências levantadas através desses experimentos não podem ser consideradas conclusivas e precisam de validação na bancada.

### 9.3 Mutações relevantes para a sensibilidade de imunoenaios

As alterações estruturais significativas ( $\Delta\text{RMSD} > 5\text{\AA}$ ) observadas em algumas estruturas estão diretamente relacionadas ao ângulo medido entre os dois domínios de SRS29B (Figura 16). O movimento entre as duas cadeias acontece na região do aminoácido 129, e determina as principais alterações de RMSD observadas na proteína. Quanto a esse ângulo, uma das conformações observadas é semelhante à conformação de SRS29B em 1YNT, apresentada por todas as simulações que não demonstram  $\Delta\text{RMSD}$  significativo. Outras conformações observadas

dentro do grupo *Natural* podem ser vistas na Figura 19.

Conforme mostrado na subseção 4.1.2, a mutação S15A ocorre em apenas 3 isolados.

Considerando as simulações de 4000ns realizadas no grupo *Natural*, observa-se que todo o conjunto de estruturas apresenta  $\Delta\text{RMSD} > 5\text{\AA}$ , inclusive a estrutura *Original*. A similaridade de trajetórias entre estruturas mutadas e a não mutada indica que não há mutação relevante para a conformação de SRS29B entre as variantes analisadas. Esse resultado contraria o que foi observado nas simulações de 400ns do mesmo grupo, aonde a estrutura *Original* manteve-se sem alterações significativas por toda a trajetória. É interessante notar que no conjunto de 4000ns, a variante *Original* manteve-se até os 2200ns na faixa não significativa de RMSD, de forma diferente das outras três estruturas do grupo.

*GUY-DOS* e *GUY-MAT* são isolados provenientes de pacientes humanos na Guiné Francesa (CARME et al., 2002). *GUY-2003-MEL* é um isolado similar, tendo sido descrito um ano depois. Todos esses três isolados não apresentam marcadores PCR-RFLP atípicos, mas possuem misturas de marcadores, uma característica das cepas recombinantes. A Guiana Francesa é um país localizado na região Amazônica, e os isolados endêmicos da região são conhecidos pela alta virulência (SIMON et al., 2019) e pelo não enquadramento em grupos clonais. A patogenicidade desses isolados originou o termo clínico *Amazonian Toxoplasmosis*, devido aos graves sintomas apresentados por indivíduos imunocomprometidos e infectados por tais variantes. Esses isolados, portadores de S15A, apresentam o locus PCR-RFLP SAG1<sub>I</sub>. De acordo com estudos, a maioria dos padrões de aminoácidos que descaracterizam estruturas secundárias estão localizados em alfa hélices (IMAI; MITAKU, 2004). S15A é a única substituição natural que acontece em resíduo que participa de estrutura secundária do tipo.

É interessante notar que os isolados SAG1<sub>I</sub> geralmente não possuem a mutação S15A. Dessa forma, os marcadores PCR-RFLP não são determinantes para imunocompatibilidade entre antígenos. O amplicon considerado nesse tipo de genotipagem é de cerca de 100bp, enquanto a CDS de SRS29B engloba aproximadamente 300pb. Além disso, as enzimas de restrição definem cada genótipo PCR-RFLP através do reconhecimento de alvos de menos de 10pb. Dessa forma, a capacidade da genotipagem PCR-RFLP em discriminar SNPs é limitada (SU et al., 2004).

Nenhum isolado brasileiro ou norte fluminense apresenta genótipo PCR-RFLP semelhante aos isolados da Guiné Francesa. Dessa forma, não existem indícios da presença de isolados portadores de S15A em Campos dos Goytacazes. Um isolado altamente relacionado com as variantes responsáveis pela *Amazonian Toxoplasmosis*, que também apresentava alta virulência, já foi encontrado no estado brasileiro de São Paulo (PENA et al., 2020).

Considerando a presença de S15A em um restrito grupo de isolados da Guiana Francesa, não podemos excluir a existência de substituições de aminoácidos igualmente raras ou até presentes exclusivamente em variantes de *T. gondii* endêmicas em Campos dos Goytacazes. As substituições de aminoácidos eventualmente descobertas através do sequenciamento dessas va-

riantes de Campos poderiam analisadas pela metodologia *in silico* desenvolvida neste trabalho, cujo código está livremente disponível na internet (ARAUJO, 2022). Dessa forma, a contribuição das variantes de Campos na imunocompatibilidade da proteína SRS29B pode ser melhor compreendida.

## 9.4 Imunodiagnósticos utilizados atualmente

O coquetel de antígenos descrito em uma das patentes publicadas pela Abbot Laboratories que utiliza SRS29B como antígeno de sensibilização. De acordo com o documento da patente, esse coquetel foi desenvolvido através da diversificação de sequências peptídicas pelo recorte da sequência e deleção e substituição de aminoácidos. Os critérios para diversificação das sequências são desconhecidos. A sequência principal do imunoenensaio é identificada como *Toxo P30* no documento é quase idêntica à sequência presente na estrutura apelidada de *Original* neste trabalho. No entanto, *Toxo P30* contém o peptídeo C-terminal, que não existe na estrutura *Original*. Comparando as demais sequências da patente da Abbott com as mutações naturais de SRS29B, nota-se que a diversificação empregada para esses ASs não corresponde a nenhum isolado conhecido de *T. gondii*. Devemos considerar que esses ASs utilizados atualmente foram desenvolvidos na década de 1990, nos primórdios da era do sequenciamento do DNA genômico, quando a tecnologia disponível era primitiva e de alto custo. Atualmente, existe uma grande base de dados de sequências de *T. gondii*. A utilização de variantes de ocorrência natural das proteínas utilizadas como ASs seria um método mais moderno para o desenvolvimento de imunoenensaios. Adicionalmente, poderia haver um esforço de sequenciamento de isolados de regiões endêmicas de variantes não clonais de *T. gondii*. Considerar a grande diversidade ainda não catalogada desses protozoários pode possibilitar o desenvolvimento de coquetéis de ASs mais abrangentes.

## 9.5 Desenvolvendo imunodiagnósticos mais precisos

Para criar um imunodiagnóstico mais sensível em pacientes de Campos dos Goytacazes, uma possibilidade seria a utilização simultânea de diversas versões de SRS29B, abrangendo as variantes que podem apresentar imunocompatibilidade divergente conforme discutido neste trabalho.

Outra possibilidade é a utilização de outras proteínas como ASs. Além de SRS29B, diversas outras proteínas de *T. gondii* demonstram ser imunogênicas quando purificadas e expostas ao soro de murinos previamente expostos ao patógeno, conforme mostrado na Tabela 11 (DÖŞKAYA et al., 2018). A imunogenicidade é o requisito principal para a utilização de determinada proteína como AS, considerando que os imunoenensaios atuam através da detecção de anticorpos produzidos pelo hospedeiro contra essas moléculas, na amostra sorológica. Para determinar as proteínas que seriam menos sensíveis a variantes polimórficas do parasito, as sequências dos genes foram obtidas no portal ToxoDB. Considerando os 65 isolados do protozoário cujas

sequências são disponíveis publicamente, a porcentagem da sequência que não sofre mutações foi extraída para cada gene conforme compilado na Tabela 11. Nessa tabela, é possível observar que muitas das proteínas imunogênicas já foram utilizadas previamente como ASs em imunoenaios para diagnóstico de toxoplasmose e aparecem na Tabela 5, como GRA3, GRA6, GRA7, GRA8, ROP1 e MIC2. SRS29B não foi testada por Döşkaya et al. (2018), talvez por ser amplamente reconhecida como um antígeno dominante. Enquanto isso, proteínas relacionadas a SRS29B como SRS29A e SRS29C possuem atividade imunológica confirmada naquele trabalho (DöşKAYA et al., 2018).

Algumas dessas proteínas podem combinar a alta imunogenicidade de SRS29B com uma baixa variabilidade polimórfica em diferentes populações do parasito. É importante notar que a própria SRS29B é uma das proteínas menos variáveis entre os isolados analisados. É possível que, embora seja pouco variável no grupo de isolados globais, SRS29B seja altamente variável no grupo de isolados endêmicos em regiões de alta atipicidade, ou especificamente na região de Campos dos Goytacazes. Talvez a utilização de proteínas além de SRS29B seja fundamental para obtenção de métricas de sensibilidade satisfatórias em pacientes de regiões de alta variabilidade genotípica de *T. gondii*.

Tabela 11 – Proteínas imunogênicas de *Toxoplasma gondii*

Nome do gene	Produto transcrito	Região constante (%)	<i>n</i> mutações
TGME49_319340	hypothetical protein	97.72	61
TGME49_270240	MAG1	97.55	31
TGME49_233450	SAG-related sequence SRS29A	97.37	17
TGME49_293730	DHHC zinc finger DCP	97.18	148
TGME49_233460	SAG-related sequence SRS29B	97.1	18
TGME49_258660	roptry protein ROP6	96.98	56
TGME49_290010	hypothetical protein	96.88	16
TGME49_201780	microneme protein MIC2	96.86	66
TGME49_227620	dense granule protein GRA2	96.8	48
TGME49_291890	microneme protein MIC1	96.66	110
TGME49_233530	hypothetical protein	96.59	129
TGME49_233480	SAG-related sequence SRS29C	96.39	78
TGME49_222370	SAG-related sequence SRS13	96.28	121
TGME49_227280	dense granule protein GRA3	95.97	146
TGME49_215980	hypothetical protein	95.83	23
TGME49_203310	dense granule protein GRA7	95.78	182
TGME49_267680	microneme protein MIC12	95.72	328
TGME49_204050	subtilisin SUB1	95.69	219
TGME49_226660	hypothetical protein	95.68	349
TGME49_201390	hypothetical protein	95.66	84
TGME49_310780	dense granule protein GRA4	95.51	79
TGME49_270250	dense granule protein GRA1	95.51	74
TGME49_286450	dense granule protein GRA5	95.49	69
TGME49_254720	dense granule protein GRA8	95.44	71
TGME49_257080	3'5'-cyclic nucleotide phosphodiesterase DCP	95.4	380
TGME49_221710	TBC DCP	95.23	378
TGME49_275490	hypothetical protein	95.13	417
TGME49_313020	STAS DCP	95.13	412
TGME49_239740	dense granule protein GRA14	95.06	116
TGME49_305020	hypothetical protein	95.04	421
TGME49_213388	hypothetical protein	94.97	314
TGME49_275440	dense granule protein GRA6	94.96	147
TGME49_309590	roptry protein ROP1	93.82	261
TGME49_288380	heat shock protein HSP90	93.73	435
TGME49_306310	RecF/RecN/SMC N terminal DCP	93.73	578
TGME49_210095	hypothetical protein	93.65	106

Nota: A relação de proteínas imunogênicas foi obtida através do contato das proteínas purificadas com o soro murino previamente sensibilizado por Döşkaya et al. (2018). O gene de SRS29B foi adicionado à lista por se tratar de uma conhecida proteína imunogênica. Valores de variação foram computados neste trabalho com base nas 65 sequências de *T. gondii* disponíveis na ToxoDB. A área constante mede a porcentagem de loci do gene onde não existem mutações conhecidas. O *n* de mutações considera a quantidade de loci do gene onde mutações são observadas.

## 10 Perspectivas futuras

Conforme a tecnologia do sequenciamento genômico avança, observamos a diminuição progressiva dos custos envolvidos nesse processo. Espera-se que isso favoreça a adoção do sequenciamento completo de genomas de isolados como principal ferramenta para mensuração de variabilidade genética de *T. gondii*, substituindo a genotipagem por marcadores RFLP que é amplamente adotada atualmente.

O problema da indisponibilidade de estruturas protéicas de *T. gondii* no RCSB pode ser solucionada através da modelagem *de novo* utilizando o software AlphaFold (JUMPER et al., 2021). Ele teve seu acesso liberado ao público no segundo semestre de 2021. Esse programa é capaz de determinar a conformação de proteínas à partir da sequência de aminoácidos, com precisão satisfatória. Dessa forma, em trabalhos futuros será possível obter estruturas de proteínas tradicionalmente usadas como ASs em imunodiagnósticos para *T. gondii* como GRA6, GRA7 e GRA8. A disponibilidade de estruturas protéicas e ampla utilização em imunoenaios foram fatores determinantes para a escolha de SRS29B como proteína a ser analisada neste trabalho. O acesso a essas estruturas permite estender a análise que foi feita em SRS29B neste trabalho a outras proteínas. Isso permite a avaliação de imunodiagnósticos já existentes que utilizam essas proteínas como ASs, e pode guiar o desenvolvimento de novos imunodiagnósticos.

Uma lista contendo dados importantes que hoje não estão disponíveis, mas que futuramente podem complementar a *pipeline* deste trabalho é apresentada abaixo.

- Sequências de mais isolados do grupo *Chinese 1*, além de TgPRC4.
- Sequências de isolados de Campos dos Goytacazes.
- Sequências de *T. gondii* obtidas através de PCR em amostras de pacientes com serologia negativa para a toxoplasmose.
- Estruturas moleculares de GRA7, GRA8 e ROP1 (podem ser geradas pelo *AlphaFold*).

A maior disponibilidade de informação genômica sobre variantes do do parasito permite a avaliação *in silico* do potencial de interação entre ASs e anticorpos produzidos contra diferentes variantes de proteínas de *T. gondii*. Isso deve beneficiar especialmente os pacientes de regiões geográficas que são endêmicas de populações heterogêneas e/ou atípicas do parasito. Além da possibilidade da melhoria na precisão dos diagnósticos para a toxoplasmose, os efeitos da doença podem ser melhor compreendidos, principalmente quando causados por esses isolados atípicos.

A combinação da predição de epítomos lineares (*BepiPred 2.0*), predição de  $\Delta\Delta G$  (*DynaMut*), execução da dinâmica molecular (*GROMACS*) e análise de trajetórias simuladas (*mdanalysis*) pode ser aplicada em dados de sequenciamento de *T. gondii* disponibilizados futuramente. Adicionalmente, essa *pipeline* pode ser utilizada para aprimorar imunodiagnósticos de qualquer doença infecciosa. Essa metodologia é caracterizada pela análise simplificada das trajetórias de cada simulação, utilizando apenas séries temporais simples como RMSD e RMSF. O foco é a execução em larga escala de simulações para um grande número estruturas e substituições. O principal cenário de uso seria em casos similares ao que foi observado em Campos dos Goytacazes, quando imunoenaios apresentam baixa sensibilidade apenas em pacientes de uma região geográfica onde as variantes endêmicas do patógeno são pouco conhecidas. Como as estruturas dos antígenos podem ser determinadas pelo *AlphaFold*, a única informação necessária seria o sequenciamento da região imunogênica de interesse para um número suficiente de isolados do patógeno. Apesar dessa metodologia se basear em métodos preditivos ao invés de resultados experimentais, o acesso a um servidor de computação de alto desempenho é o único requisito material para a sua execução.

Desenvolver imunodiagnósticos sensíveis e específicos é um desafio. Embora existam outras metodologias de diagnóstico mais precisas, o imunodiagnóstico destaca-se pelo baixo custo e facilidade de operação. Assim, imunodiagnósticos são acessíveis aos sistemas de saúde dos países em desenvolvimento. Muitos desses países não possuem estrutura de saneamento básico e de saúde, possuindo alta prevalência de doenças infecciosas como malária e leishmaniose. Além disso, grande parte dos países em desenvolvimento estão localizados em regiões de clima tropical, que possuem maior atividade de agentes infecciosos (SATTENSPIEL, 2000).



# 11 Conclusão

As patentes mais recentes que descrevem antígenos de sensibilização potencialmente utilizados em imunodiagnósticos comerciais para toxoplasmose (IgG) foram desenvolvidas há mais de duas décadas. Não é possível determinar se essas patentes são de fato utilizadas em imunodiagnósticos comercializados atualmente.

Quatro variações únicas da proteína SRS29B foram identificadas através de sequências disponíveis publicamente, e são diferenciáveis por seis mutações. Predições de epítomos lineares para essas sequências não indicam variação significativa no conjunto.

Análises das trajetórias obtidas nos experimentos de dinâmica molecular de 400ns realizados em estruturas mutadas de SRS29B indicam  $\Delta\text{RMSD} > 5\text{\AA}$  em algumas variantes, enquanto a estrutura *Original*, que não possui mutações, manteve seu  $\Delta\text{RMSD}$  pouco alterado ao longo trajetória.

As mutações envolvidas em estruturas que apresentam  $\Delta\text{RMSD} > 5\text{\AA}$  na dinâmica molecular de 400ns também apresentam  $\Delta\Delta G$  positivo e significativo na predição pelo DynaMut.

Simulações adicionais de 4000ns para o grupo *Natural* mostram relativa uniformidade em todas as trajetórias de estruturas do grupo.

A alteração do ângulo entre os dois domínios de SRS29B é o principal fator responsável pelas alterações conformacionais observadas em todos os experimentos de dinâmica molecular.

Algumas das substituições observadas naturalmente, como S15A, ocorrem em um pequeno grupo de isolados. Substituições de aminoácidos em SRS29B podem ocorrer sem que o genótipo PCR-RFLP da região SAG1 do isolado seja afetado.

A metodologia para avaliação preliminar do efeito estrutural de mutações em antígenos desenvolvida durante este trabalho está disponível para uso público (ARAUJO, 2022). Esta *pipeline* pode auxiliar a analisar dados de sequenciamento de *T. gondii* obtidos futuramente. Limitada pela disponibilidade de estruturas protéicas e de sequências dos genes correspondentes, ela pode ser utilizada para analisar qualquer proteína relevante aos imunoenaios, de *T. gondii* ou de outros organismos patogênicos.

## 12 Referências

- ABBOTT LABORATORIES. *Toxoplasma gondii* antigens, p35, and uses thereof. 1999. Citado na página 43.
- AJZENBERG, D. 1995-2015: it is time to celebrate 20 years of (intensive) genotyping of toxoplasma gondii strains. *Future microbiology*, v. 10, 2015. Citado na página 27.
- AMOUEI, A. et al. A systematic review of toxoplasma gondii genotypes and feline: Geographical distribution trends. *Transboundary and emerging diseases*, v. 67, 2019. Citado 3 vezes nas páginas 27, 58 e 59.
- ANDREOTTI, P. E. et al. Immunoassay of infectious agents. *BioTechniques*, v. 35, 2003. Citado na página 20.
- ARAUJO, G. N. *Pipeline para análise de antígenos de sensibilização*. 2022. Disponível em: <<https://github.com/Gab0/toxo-immunodiagnosis-protocol>>. Citado 2 vezes nas páginas 63 e 68.
- ARAUJO, G. N.; BANSAL, U.; MERUGU, R. R. *autoGromacs*. 2021. Disponível em: <<https://github.com/Gab0/auto-gromacs>>. Citado 3 vezes nas páginas 37, 38 e 39.
- ASCOLI, C. A. Could mutations of sars-cov-2 suppress diagnostic detection? *Nature biotechnology*, v. 39, 2021. Citado na página 24.
- BACHAN, M. et al. High seroprevalence of toxoplasma gondii in goats in jharkhand state of india. *Veterinary parasitology, regional studies and reports*, v. 12, 2017. Citado na página 57.
- BASIT, K. A. et al. Toxoplasmosis in an immunocompetent patient. *Pakistan journal of medical sciences*, v. 34, 2018. Citado na página 34.
- BERTRANPETIT, E. et al. Phylogeography of toxoplasma gondii points to a south american origin. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases*, v. 48, 2016. Citado na página 28.
- BIGNA, J. J. et al. Global, regional, and country seroprevalence of toxoplasma gondii in pregnant women: a systematic review, modelling and meta-analysis. *Scientific reports*, v. 10, 2020. Citado 2 vezes nas páginas 12 e 59.
- BINNEY, N.; HYDE, C.; BOSSUYT, P. M. On the origin of sensitivity and specificity. *Annals of internal medicine*, v. 174, 2021. Citado na página 13.
- BOGUSZEWSKA, K. et al. Review: immunoassays in dna damage and instability detection. *Cellular and molecular life sciences : CMLS*, v. 76, 2019. Citado na página 22.
- BOOTHROYD, J. C. et al. The surface of toxoplasma: more and less. *International journal for parasitology*, v. 28, 1998. Citado na página 58.
- BOURASSA, L. et al. A sars-cov-2 nucleocapsid variant that affects antigen test performance. *Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology*, v. 141, 2021. Citado na página 57.

- BUUS, S. et al. High-resolution mapping of linear antibody epitopes using ultrahigh-density peptide microarrays. *Molecular & cellular proteomics : MCP*, v. 11, 2012. Citado na página 16.
- CALDERARO, A. et al. Evaluation of toxoplasma gondii immunoglobulin g (igg) and igm assays incorporating the newvidia analyzer system. *Clinical and vaccine immunology : CVI*, v. 15, 2008. Citado na página 22.
- CAN, H. et al. In silico discovery of antigenic proteins and epitopes of sars-cov-2 for the development of a vaccine or a diagnostic approach for covid-19. *Scientific reports*, v. 10, 2020. Citado na página 23.
- CAPRIOTTI, E. et al. A three-state prediction of single point mutations on protein stability changes. *BMC bioinformatics*, v. 9 Suppl 2, 2008. Citado na página 45.
- CARME, B. et al. Severe acquired toxoplasmosis in immunocompetent adult patients in french guiana. *Journal of clinical microbiology*, v. 40, 2002. Citado na página 62.
- CERUTTI, A.; BLANCHARD, N.; BESTEIRO, S. The bradyzoite: A key developmental stage for the persistence and pathogenesis of toxoplasmosis. *Pathogens (Basel, Switzerland)*, v. 9, 2020. Citado na página 26.
- CESBRON-DELAUW, M. F. et al. Similarities between the primary structures of two distinct major surface proteins of toxoplasma gondii. *The Journal of biological chemistry*, v. 269, 1994. Citado na página 57.
- CHARLES, E.; CALLEGAN, M. C.; BLADER, I. J. The sag1 toxoplasma gondii surface protein is not required for acute ocular toxoplasmosis in mice. *Infection and immunity*, v. 75, 2007. Citado na página 57.
- CHEN, X. et al. Human monoclonal antibodies block the binding of sars-cov-2 spike protein to angiotensin converting enzyme 2 receptor. *Cellular & molecular immunology*, v. 17, 2020. Citado na página 19.
- CHILDERS, M. C.; DAGGETT, V. Validating molecular dynamics simulations against experimental observables in light of underlying conformational ensembles. *The journal of physical chemistry. B*, v. 122, 2018. Citado na página 61.
- DALKAS, G. A. et al. Cation- $\pi$ , amino- $\pi$ ,  $\pi$ - $\pi$ , and h-bond interactions stabilize antigen-antibody interfaces. *Proteins*, v. 82, 2013. Citado na página 16.
- DALL'ANTONIA, F. et al. Structure of allergens and structure based epitope predictions. *Methods (San Diego, Calif.)*, v. 66, 2013. Citado na página 61.
- DALY, D. S. et al. Evaluating concentration estimation errors in elisa microarray experiments. *BMC bioinformatics*, v. 6, 2004. Citado na página 57.
- DARD, C. et al. Relevance of and new developments in serology for toxoplasmosis. *Trends in parasitology*, v. 32, 2016. Citado na página 34.
- DEITSCH, K. W.; LUKEHART, S. A.; STRINGER, J. R. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nature reviews. Microbiology*, v. 7, 2009. Citado na página 59.

- DOOREN, G. G. van; STRIEPEN, B. The algal past and parasite present of the apicoplast. *Annual review of microbiology*, v. 67, 2013. Citado na página 26.
- DUBEY, J. P. et al. Toxoplasmosis in humans and animals in brazil: high prevalence, high burden of disease, and epidemiology. *Parasitology*, v. 139, 2012. Citado na página 12.
- DöşKAYA, M. et al. Discovery of new toxoplasma gondii antigenic proteins using a high throughput protein microarray approach screening sera of murine model infected orally with oocysts and tissue cysts. *Parasites & vectors*, v. 11, 2018. Citado 3 vezes nas páginas 63, 64 e 65.
- ERP, E. A. van et al. Fc-mediated antibody effector functions during respiratory syncytial virus infection and disease. *Frontiers in immunology*, v. 10, 2018. Citado na página 19.
- FLEGR, J.; HORÁČEK, J. Negative effects of latent toxoplasmosis on mental health. *Frontiers in psychiatry*, v. 10, 2019. Citado na página 34.
- FUMAGALLI, M. J. et al. Enzyme-linked immunosorbent assay using recombinant envelope protein 2 antigen for diagnosis of chikungunya virus. *Virology journal*, v. 15, 2018. Citado na página 41.
- GAO, J.-M. et al. Genetic analyses of chinese isolates of toxoplasma gondii reveal a new genotype with high virulence to murine hosts. *Veterinary parasitology*, v. 241, 2017. Citado na página 58.
- GARCIA, L. S. Malaria. *Clinics in laboratory medicine*, v. 30, 2010. Citado na página 26.
- GENOVA, B. M. D. et al. Intestinal delta-6-desaturase activity determines host range for toxoplasma sexual reproduction. *PLoS biology*, v. 17, 2019. Citado na página 26.
- GRAILLE, M. et al. Crystal structure of the complex between the monomeric form of toxoplasma gondii surface antigen 1 (sag1) and a monoclonal antibody that mimics the human immune response. *Journal of molecular biology*, v. 354, 2005. Citado 2 vezes nas páginas 16 e 60.
- GREEF, P. C. de et al. The naive t-cell receptor repertoire has an extremely broad distribution of clone sizes. *eLife*, v. 9, 2019. Citado na página 18.
- GUALANO, M. R. et al. Knowledge, attitudes and beliefs towards compulsory vaccination: a systematic review. *Human vaccines & immunotherapeutics*, v. 15, 2019. Citado na página 12.
- GUARALDO, L. et al. Ocular toxoplasmosis: adverse reactions to treatment in a brazilian cohort. *Transactions of the Royal Society of Tropical Medicine and Hygiene*, v. 112, 2017. Citado na página 20.
- HARVEY, W. T. et al. Sars-cov-2 variants, spike mutations and immune escape. *Nature reviews. Microbiology*, v. 19, 2021. Citado na página 24.
- HERNÁNDEZ-CORTAZAR, I. et al. Toxoplasmosis in mexico: epidemiological situation in humans and animals. *Revista do Instituto de Medicina Tropical de Sao Paulo*, v. 57, 2013. Citado na página 59.
- HOLDER, T. *Orientation, displacement and angle measurments of helices and domains*. 2010. Disponível em: <[https://pymolwiki.org/index.php/Angle\\_between\\_domains](https://pymolwiki.org/index.php/Angle_between_domains)>. Citado na página 39.

- HOLMES, K. K. et al. Major infectious diseases: Key messages from disease control priorities, third edition. *Disease Control Priorities, Third Edition (Volume 6): Major Infectious Diseases*, The World Bank, p. 1-27, Nov 2017. Disponível em: <[http://dx.doi.org/10.1596/978-1-4648-0524-0\\_ch1](http://dx.doi.org/10.1596/978-1-4648-0524-0_ch1)>. Citado na página 12.
- HOSSAIN, M. G.; UEDA, K. Investigation of a novel hepatitis b virus surface antigen (hbsag) escape mutant affecting immunogenicity. *PloS one*, v. 12, 2016. Citado na página 59.
- IMAI, K.; MITAKU, S. Mechanisms of secondary structure breakers in soluble proteins. *Biophysics (Nagoya-shi, Japan)*, v. 1, 2004. Citado na página 62.
- INNES, E. A. A brief history and overview of toxoplasma gondii. *Zoonoses and public health*, v. 57, 2009. Citado na página 26.
- ISMAIL, A. A. A. Interference from endogenous antibodies in automated immunoassays: what laboratorians need to know. *Journal of clinical pathology*, v. 62, 2009. Citado na página 57.
- JESPERSEN, M. C. et al. Bepipred-2.0: improving sequence-based b-cell epitope prediction using conformational epitopes. *Nucleic acids research*, v. 45, 2017. Citado 3 vezes nas páginas 16, 37 e 60.
- JUMPER, J. et al. Highly accurate protein structure prediction with alphafold. *Nature*, v. 596, 2021. Citado na página 66.
- JUNG, C.; LEE, C. Y.-F.; GRIGG, M. E. The srs superfamily of toxoplasma surface proteins. *International journal for parasitology*, v. 34, 2003. Citado na página 32.
- KANELLOPOULOS, J. M.; OJCIUS, D. M. Development of humoral immunity. *Biomedical journal*, v. 42, 2019. Citado na página 15.
- KAPINGIDZA, A. B.; KOWAL, K.; CHRUSZCZ, M. Antigen-antibody complexes. *Sub-cellular biochemistry*, v. 94, 2020. Citado 2 vezes nas páginas 17 e 19.
- KATO, M. et al. Reactivity of synthetic sag1 (p30) peptide sequences with rh, s273 and beverley strain-induced anti-toxoplasma gondii antibodies. *Pathobiology : journal of immunopathology, molecular and cellular biology*, v. 74, 2006. Citado na página 32.
- KHAN, K.; KHAN, W. Congenital toxoplasmosis: An overview of the neurological and ocular manifestations. *Parasitology international*, v. 67, 2017. Citado na página 12.
- KING, P.; TYLER, K. M.; HUNTER, P. R. Anthroponotic transmission of cryptosporidium parvum predominates in countries with poorer sanitation: a systematic review and meta-analysis. *Parasites & vectors*, v. 12, 2018. Citado na página 26.
- KISSINGER, J. C. et al. Toxodb: accessing the toxoplasma gondii genome. *Nucleic acids research*, v. 31, 2003. Citado na página 42.
- KOPMANN, C. et al. Abundance and transferability of antibiotic resistance as related to the fate of sulfadiazine in maize rhizosphere and bulk soil. *FEMS microbiology ecology*, v. 83, 2012. Citado na página 21.
- KRIEGER, M. A.; ZANCHIN, N. I. T.; BASCHIROTTI, P. T. *Método para imunodiagnóstico de toxoplasmose aguda e crônica*. 2018. Citado na página 41.

- LAMAN, J. D.; CLAASSEN, E.; NOELLE, R. J. Functions of cd40 and its ligand, gp39 (cd40l). *Critical reviews in immunology*, v. 37, 2018. Citado na página 18.
- LEMKUL, J. A. *Lysozyme in Water*. 2018. Disponível em: <<http://www.mdtutorials.com/gmx/lysozyme/index.html>>. Citado na página 39.
- LIPPI, G.; ADELI, K.; PLEBANI, M. Commercial immunoassays for detection of anti-sars-cov-2 spike and rbd antibodies: urgent call for validation against new and highly mutated variants. *Clinical chemistry and laboratory medicine*, 2021. Citado na página 24.
- LIU, L. et al. rrop2(186-533): a novel peptide antigen for detection of igm antibodies against toxoplasma gondii. *Foodborne pathogens and disease*, v. 9, 2011. Citado na página 58.
- LIU, X.-Y. et al. Toxoplasma gondii oocyst-driven infection in pigs, chickens and humans in northeastern china. *BMC veterinary research*, v. 15, 2019. Citado na página 57.
- LOPES, F. M. R. et al. Toxoplasma gondii infection in pregnancy. *The Brazilian journal of infectious diseases : an official publication of the Brazilian Society of Infectious Diseases*, v. 11, 2007. Citado na página 20.
- LUFT, B. J.; REMINGTON, J. S. Toxoplasmic encephalitis in aids. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, v. 15, 1992. Citado na página 21.
- MAIER, J. A. et al. ff14sb: Improving the accuracy of protein side chain and backbone parameters from ff99sb. *Journal of chemical theory and computation*, v. 11, 2015. Citado 6 vezes nas páginas 39, 48, 50, 51, 52 e 54.
- MARTINS, L. M. Identificação de biomarcadores associados ao desenvolvimento de toxoplasmose ocular. 2015. Citado 5 vezes nas páginas 13, 34, 41, 56 e 58.
- MCFADDEN, G. I. The apicoplast. *Protozoa*, v. 248, 2010. Citado na página 26.
- MEGHA, K. B.; MOHANAN, P. V. Role of immunoglobulin and antibodies in disease management. *International journal of biological macromolecules*, v. 169, 2020. Citado na página 14.
- MICHAUD-AGRAWAL, N. et al. Mdanalysis: a toolkit for the analysis of molecular dynamics simulations. *Journal of computational chemistry*, v. 32, 2010. Citado na página 39.
- MINOR, P. D. et al. Monoclonal antibodies which block cellular receptors of poliovirus. *Virus research*, v. 1, 1984. Citado na página 19.
- MITTERNACHT, S. Freesasa: An open source c library for solvent accessible surface area calculations. *F1000Research*, v. 5, 2016. Citado na página 39.
- MOLARI, M. et al. Quantitative modeling of the effect of antigen dosage on b-cell affinity distributions in maturing germinal centers. *eLife*, v. 9, 2020. Citado na página 18.
- MURATA, F. H. A. et al. Evaluation of serological and molecular tests used for the identification of *Toxoplasma gondii* infection in patients treated in an ophthalmology clinic of a public health service in são paulo state, brazil. *Frontiers in cellular and infection microbiology*, v. 9, 2019. Citado na página 13.

- NAGELE, E. P. et al. Natural igg autoantibodies are abundant and ubiquitous in human sera, and their number is influenced by age, gender, and disease. *PloS one*, v. 8, 2012. Citado 3 vezes nas páginas 13, 15 e 19.
- NEZLIN, R. Dynamic aspects of the immunoglobulin structure. *Immunological investigations*, v. 48, 2019. Citado na página 14.
- OS EDITORES DA ENCYCLOPAEDIA BRITANNICA. 2020. Citado na página 15.
- PEIXOTO-RANGEL, A. L. et al. Investigation of immunogenetic and environmental risk factors predominant in the development of toxoplasmic retinochoroiditis. *TOXOPLASMA CENTENNIAL CONGRESS: FROM DISCOVERY TO PUBLIC HEALTH MANAGEMENT*, 2008. Citado 2 vezes nas páginas 13 e 34.
- PENA, H. F. J. et al. Toxoplasma gondii isolated from a brazilian patient with rare pulmonary toxoplasmosis has a novel genotype and is closely related to amazonian isolates. *Parasitology research*, v. 120, 2020. Citado na página 62.
- PETERS, B.; NIELSEN, M.; SETTE, A. T cell epitope predictions. *Annual review of immunology*, v. 38, 2020. Citado na página 18.
- POULSEN, T. B. G. et al. Identification of novel native autoantigens in rheumatoid arthritis. *Biomedicines*, v. 8, 2020. Citado na página 19.
- REES, A. R. Understanding the human antibody repertoire. *mAbs*, v. 12, 2020. Citado na página 18.
- REYNOSO-PALOMAR, A.; MORENO-GÁLVEZ, D.; VILLA-MANCERA, A. Prevalence of toxoplasma gondii parasite in captive mexican jaguars determined by recombinant surface antigens (sag1) and dense granular antigens (gra1 and gra7) in elisa-based serodiagnosis. *Experimental parasitology*, v. 208, 2019. Citado na página 57.
- RIDGLEY, D. M. et al. The role of protein hydrophobicity in conformation change and self-assembly into large amyloid fibers. *Biomacromolecules*, v. 15, 2014. Citado na página 61.
- RIGATTI, S. J. Random forest. *Journal of insurance medicine (New York, N.Y.)*, v. 47, 2017. Citado na página 17.
- RODRIGUES, C. H.; PIRES, D. E.; ASCHER, D. B. Dynamut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic acids research*, v. 46, 2018. Citado na página 39.
- RODRIGUEZ, C. et al. Major surface protein of toxoplasma gondii (p30) contains an immunodominant region with repetitive epitopes. *European journal of immunology*, v. 15, 1985. Citado na página 32.
- ROSE, A. B. Intron-mediated regulation of gene expression. *Current topics in microbiology and immunology*, v. 326, 2008. Citado na página 32.
- ROSTAMI, A. et al. Global prevalence of latent toxoplasmosis in pregnant women: a systematic review and meta-analysis. *Clinical microbiology and infection : the official publication of the European Society of Clinical Microbiology and Infectious Diseases*, v. 26, 2019. Citado na página 12.



SATTENSPIEL, L. Tropical environments, human activities, and the transmission of infectious diseases. *American journal of physical anthropology*, Suppl 31, 2000. Citado na página 67.

SCHILD, G. C. et al. Antigenic variation in current influenza A viruses: evidence for a high frequency of antigenic 'drift' for the Hong Kong virus. *Bulletin of the World Health Organization*, v. 51, 1974. Citado na página 24.

SELA-CULANG, I.; KUNIK, V.; OFRAN, Y. The structural basis of antibody-antigen recognition. *Frontiers in immunology*, v. 4, 2013. Citado na página 17.

SENGUPTA, S.; CHATTOPADHYAY, M. K.; GROSSART, H.-P. The multifaceted roles of antibiotics and antibiotic resistance in nature. *Frontiers in microbiology*, v. 4, 2012. Citado na página 21.

SENSAKOVIC, J. W.; SMITH, L. G. Oral antibiotic treatment of infectious diseases. *The Medical Clinics of North America*, v. 85, 2001. Citado na página 12.

SIBLEY, L. D.; AJIOKA, J. W. Population structure of *Toxoplasma gondii*: clonal expansion driven by infrequent recombination and selective sweeps. *Annual review of microbiology*, v. 62, 2008. Citado na página 27.

SIGALA, P. A. et al. Determination of hydrogen bond structure in water versus aprotic environments to test the relationship between length and stability. *Journal of the American Chemical Society*, v. 137, 2015. Citado na página 17.

SILVA, R. C. da; LANGONI, H.; MEGID, J. Adaptive and genetic evolution of *Toxoplasma gondii*: a host-parasite interaction. *Revista da Sociedade Brasileira de Medicina Tropical*, v. 50, 2017. Citado na página 28.

SIMON, S. et al. Virulence of atypical *Toxoplasma gondii* strains isolated in French Guiana in a murine model. *Parasite (Paris, France)*, v. 26, 2019. Citado na página 62.

SINHA, N. et al. Understanding antibody-antigen associations by molecular dynamics simulations: detection of important intra- and inter-molecular salt bridges. *Cell Biochemistry and Biophysics*, v. 47, 1999. Citado na página 17.

SMITH, R. An overview of the Tesseract OCR engine. *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007) Vol 2*, IEEE, Sep 2007. ISSN 1520-5363. Disponível em: <<http://dx.doi.org/10.1109/ICDAR.2007.4376991>>. Citado na página 36.

SPOEL, D. V. D. et al. GROMACS: fast, flexible, and free. *Journal of Computational Chemistry*, v. 26, 2005. Citado na página 38.

SU, C. et al. Typing single-nucleotide polymorphisms in *Toxoplasma gondii* by allele-specific primer extension and microarray detection. *Methods in Molecular Biology (Clifton, N.J.)*, v. 270, 2004. Citado na página 62.

SU, C. et al. Globally diverse *Toxoplasma gondii* isolates comprise six major clades originating from a small number of distinct ancestral lineages. *Proceedings of the National Academy of Sciences of the United States of America*, v. 109, 2012. Citado 2 vezes nas páginas 31 e 46.

SU, C. et al. Moving towards an integrated approach to molecular detection and identification of *Toxoplasma gondii*. *Parasitology*, v. 137, 2009. Citado 2 vezes nas páginas 27 e 29.



- TATE, J.; WARD, G. Interferences in immunoassay. *The Clinical biochemist. Reviews*, v. 25, 2008. Citado na página 24.
- TEIMOURI, A. et al. Role of toxoplasma gondii igg avidity testing in discriminating between acute and chronic toxoplasmosis in pregnancy. *Journal of clinical microbiology*, v. 58, 2020. Citado na página 20.
- TENTER, A. M.; HECKEROTH, A. R.; WEISS, L. M. Toxoplasma gondii: from animals to humans. *International journal for parasitology*, v. 30, 2000. Citado na página 12.
- TOMAR, N.; DE, R. K. Immunoinformatics: a brief review. *Methods in molecular biology (Clifton, N.J.)*, v. 1184, 2014. Citado na página 17.
- UILENBERG, G. Babesia—a historical overview. *Veterinary parasitology*, v. 138, 2006. Citado na página 26.
- VANBUSKIRK, K. M. et al. Antigenic drift in the ligand domain of plasmodium vivax duffy binding protein confers resistance to inhibitory antibodies. *The Journal of infectious diseases*, v. 190, 2004. Citado na página 24.
- VIDARSSON, G.; DEKKERS, G.; RISPENS, T. Igg subclasses and allotypes: from structure to effector functions. *Frontiers in immunology*, v. 5, 2014. Citado na página 19.
- VIEIRA, F. P. Contaminação ambiental por oocistos de toxoplasma gondii e toxoplasmose de veiculação hídrica sob a perspectiva da vulnerabilidade de aquíferos. 2015. Citado 2 vezes nas páginas 27 e 30.
- WALZER, K. A. et al. Hammondia hammondi, an avirulent relative of toxoplasma gondii, has functional orthologs of known t. gondii virulence genes. *Proceedings of the National Academy of Sciences of the United States of America*, v. 110, 2013. Citado na página 26.
- WANG, J. et al. Neutralizing antibodies to therapeutic enzymes: considerations for testing, prevention and treatment. *Nature biotechnology*, v. 26, 2008. Citado na página 19.
- WANG, Y. et al. B cell development and maturation. *Advances in experimental medicine and biology*, v. 1254, 2020. Citado na página 18.
- WANG, Y. et al. Determinants of antigenicity and specificity in immune response for protein sequences. *BMC bioinformatics*, v. 12, 2010. Citado na página 16.
- WANG, Y.; YIN, H. Research progress on surface antigen 1 (sag1) of toxoplasma gondii. *Parasites & vectors*, v. 7, 2014. Citado 2 vezes nas páginas 13 e 57.
- WANG, Z.-D. et al. Toxoplasma gondii infection in immunocompromised patients: A systematic review and meta-analysis. *Frontiers in microbiology*, v. 8, 2016. Citado na página 34.
- WASMUTH, J. D. et al. Integrated bioinformatic and targeted deletion analyses of the srs gene superfamily identify srs29c as a negative regulator of toxoplasma virulence. *mBio*, v. 3, 2012. Citado na página 32.
- WINDECK, T.; GROSS, U. Toxoplasma gondii strain-specific transcript levels of sag1 and their association with virulence. *Parasitology research*, v. 82, 1996. Citado na página 32.

YAMANO, K. et al. Gal $\alpha$ 1-4gal $\beta$ 1-3galnac is the dominant epitope of em2 antigen, the mucin-type glycoprotein from echinococcus multilocularis. *Parasitology research*, v. 111, 2011. Citado na página 17.

YANG, W. et al. Review on the development of genotyping methods for assessing farm animal diversity. *Journal of animal science and biotechnology*, v. 4, 2012. Citado na página 28.

YEWDELL, J. W. Antigenic drift: Understanding covid-19. *Immunity*, v. 54, 2021. Citado na página 57.

ZHAN, Q. et al. Human immune repertoire in hepatitis b virus infection. *World journal of gastroenterology*, v. 27, 2021. Citado na página 13.

ZHANG. 2017. Disponível em: <<http://zhenglz.blogspot.com/2017/05/fixing-bugs-in-ffl4sb-port-for-gromacs.html>>. Citado na página 39.

ZHAO, X.-Y.; EWALD, S. E. The molecular biology and immune control of chronic toxoplasma gondii infection. *The Journal of clinical investigation*, v. 130, 2020. Citado na página 12.

ZHOU, H.-X.; PANG, X. Electrostatic interactions in protein structure, folding, binding, and condensation. *Chemical reviews*, v. 118, 2018. Citado na página 32.