

FABIANO CARVALHO GOMES

PROCURA IN SÍLICO POR GENES DE TRIPANOSOMATÍDEOS COM POSSÍVEL
FUNÇÃO REGULATÓRIA DO SISTEMA COMPLEMENTO HUMANO

Dissertação apresentada como requisito parcial
à obtenção do grau de Mestre em
Bioinformática, pelo Programa de Pós-
Graduação em Bioinformática da Universidade
Federal do Paraná.

Orientadora: Maria Berenice Reynaud Steffens
Co-orientador: Marcel Ivan Ramirez Araya

CURITIBA
2014


TERMO DE APROVAÇÃO

FABIANO CARVALHO GOMES

“Procura in silico por genes de tripanosomatídeos com possível função regulatória do sistema complemento humano”

Dissertação aprovada como requisito parcial para obtenção do grau de Mestre em Bioinformática, pelo Programa de Pós-graduação em Bioinformática, Setor de Educação Profissional e Tecnológica, da Universidade Federal do Paraná, pela seguinte banca examinadora:

Orientadora: 
Prof.^a Dr.^a Maria Berenice Reynaud Steffens

Coorientador: 
Prof. Dr. Marcel Ivan Ramirez Araya


Prof. Dr. Fabricio Klerynton Marchini
Instituto Carlos Chagas – FIOCRUZ/PR


Prof. Dr. Marco Aurelio Schüler de Oliveira
Universidade Federal do Paraná - UFPR

Curitiba, 17 de abril de 2014

AGRADECIMENTO

Agradeço primeiramente aos meus pais (José Alcides e Vera Lúcia), pelo melhor que eles puderam fazer para minha formação como indivíduo, incluindo educação, cidadania, carácter, religião, respeito, humildade, ética, entre diversos outros princípios que nos possibilita sermos quem somos e conviver em sociedade com os demais.

A minha esposa Gisele, que fez a opção em sua vida, de caminhar ao meu lado, me apoiando sempre em meus planos de formação pessoal, profissional e familiar. Com certeza, sem o apoio dessa pessoa maravilhosa eu não teria enfrentado tamanho desafio.

Quero agradecer também a Faculdade Metodista Granbery, onde me formei Bacharel em Sistemas de Informação e a todo povo da EMBRAPA Gado de Leite, que foi onde se iniciou o meu interesse pela pesquisa.

Ao professor Dr^o José Miguel Ortega (Miguelito) que em 2005, me recebeu em seu laboratório de braços abertos e por ter me estimulado ainda mais a seguir a Bioinformática. Foi através do Miguelito, que conheci o amigo Dr^o Adriano Barbosa, que devo todo o meu agradecimento, por ter me estimulado ao máximo para abrir mão do meu trabalho no Rio de Janeiro e ir morar em Curitiba para fazer o mestrado e realizar um sonho.

Agradeço a UFPR e ao Programa de Pós-graduação em Bioinformática por oferecem recursos para que eu pudesse adquirir conhecimentos necessários e ferramentas para atingir os resultados dessa dissertação de mestrado.

Obrigado à C.A.P.E.S. pela bolsa de mestrado oferecida. Espero que o apoio à ciência e a jovens pesquisadores aumentem cada vez mais, com o governo valorizando estes estudantes, pesquisadores e professores. Que este mesmo governo faça com que o Brasil se torne um país mais justo e com menos desigualdade. Uma pátria de paz, ordem, progresso e dignidade para todos os que habitam nela.

Agradeço a minha orientadora professora Dr^a Maria Berenice Steffens pela acolhida, depois de uma fase órfã do mestrado, aceitando ser minha orientadora até o fim dessa jornada. Muito obrigado por todo empenho, preocupação e orientação que recebi de você. É sempre muito bom conhecer pessoas como você.

Ao meu orientador/coorientador chileno Drº Marcel Ivan Ramirez Araya. Obrigado por disponibilizar uma pergunta científica que me aguçou, e posteriormente a isso, por ter me auxiliado na compreensão de todos os passos necessários para que eu conseguisse encontrar a resposta para essa pergunta. Agradeço por amplificar em mim, a prática de estar sempre consciente, entre vários outros exemplos que talvez você não tenha percebido, mas eu os aprendi.

Agradeço a Fundação Oswaldo Cruz e o Instituto Carlos Chagas por oferecer espaço para eu poder estudar e trabalhar quando estava no Rio de Janeiro.

Ao Drº Fernando Bachega com a grande contribuição na elaboração de imagens das estruturas das proteínas nesse trabalho.

Obrigado também ao professor Drº Wanderson, flamenguista e conterrâneo, que acompanhou e colaborou com esse trabalho desde o início.

A todos os alunos do Programa de Pós-graduação que foram parceiros durante essa jornada e confiaram em mim a representação de todos frente ao colegiado desse curso.

Ao grupo BIKE CURITIBA e a outros diversos ciclistas da cidade de Curitiba que me acolheram como um membro da família. Obrigado a Osmar Potulski e Anna Borba, Msc Augusto Machado, Eduardo Emmerick, Ricardo Satto, Bettina Konig e Pedro Luna e outras várias pessoas que seria muito complicado citar o nome agora, pela amizade séria e verdadeira estabelecida durante esse período de estadia em Curitiba.

Gostaria também de me desculpar a todos de qualquer algo que possa ter ocorrido. Se caso algo de desagrado ocorreu, não foi de minha intenção que ocorresse.

Obrigado a todos por este metrado ter valido a pena.

RESUMO

O *Trypanosoma cruzi* é o agente causador da doença de Chagas e possui um ciclo de vida complexo que alterna entre hospedeiros invertebrados e vertebrados. O ciclo do *T. cruzi* no hospedeiro vertebrado inicia-se na via sanguínea, onde ele entra em contato com o sistema complemento, o qual ele evade para penetrar em células e se multiplicar. O sistema complemento é um conjunto de proteínas séricas ativadas em cascata culminando com a lise do patógeno. Este sistema possui três vias de ativação; a via clássica, a via da lectina e a via alternativa. Vários estudos já apresentaram alguns mecanismos utilizados pelo *T. cruzi* para evadir a ação do sistema complemento. O presente trabalho propôs através de uma estratégia de bioinformática buscar novos mecanismos de evasão do *T. cruzi* ao sistema complemento. Com este objetivo desenvolvemos uma base de dados com dois grupos. Um grupo com dados de proteínas humanas de complemento e reguladoras ou inibidoras desse sistema e um outro grupo com informações de proteínas de tripanossomatídeos, coletadas de três bancos de dados proteicos distintos. As informações dos dois grupos foram confrontadas resultando em proteínas dos tripanossomatídeos homólogas a C1qBP (uma reguladora do sistema complemento em humanos). As proteínas homólogas a C1qBP foram submetidas a análises de características biológicas e a fim de identificar as proteínas presentes nas três bases de dados simultaneamente, foram implementadas técnica de mineração de dados. Dentre as proteínas selecionadas, três eram de *T. cruzi*. Os resultados obtidos propõem um novo mecanismo de evasão ao sistema complemento utilizado pelo *T. cruzi*. Este mecanismo atua diretamente com o complexo antígeno anticorpo na ligação com a extremidade gC1q da proteína C1q, inibindo a ativação do componente C1 da via clássica do sistema complemento.

ABSTRACT

The *Trypanosoma cruzi* is the causative agent of Chagas disease, have a complex life cycle that alternates between vertebrate and invertebrate hosts. The parasite need evades the complement system to penetrate into eukaryotic cells for multiplying and produce the infection. The complement system is a group of serum proteins activated in cascade culminating with the lysis of the pathogen. This system has three activation pathways: the classical pathway, lectin pathway and alternative pathway. Several studies already shown some mechanisms used by *Trypanosome cruzi* to evade the action of the complement system. This work propose through a strategy of bioinformatics seek new evasion mechanisms of *T. cruzi* for complement system..We have developed a database with two data groups. The first data group including complement system proteins and regulators and inhibitors to complement. The second group contain informations about trypanosomatid proteins, collected from three different protein databanks. The information of the two groups were compared resulting in proteins homologous to C1qBP trypanosomes. Homologous proteins C1qBP were analyzed for biological characteristics as functional domain, cellular localization and protein structure . In order to identify the proteins present in the three databases we have used simultaneously data mining technique to ensure best coverage between them. Among the selected proteins were three *T. cruzi*. The results proposes a new mechanism to evade complement system used by *T. cruzi*. This mechanism work directly with the antigen antibody complex in connection with gC1q end of C1q protein, inhibiting the activation of the C1 component of the classical pathway of the complement system.

LISTA DE FIGURAS

Figura 1: Ciclo de vida do Trypanosoma cruzi entre hospedeiros invertebrado e vertebrado	17
Figura 2: Organização estrutural da forma tripomastigota do Trypanosoma cruzi.....	19
Figura 3: Ativação do sistema complemento.....	21
Figura 4: Fluxo de de obtenção, anotação, processamento e análise de dados do projeto	38
Figura 5: Resumo dos resultados obtidos em cada etapa da estratégia proposta.....	47
Figura 6: Identificação das 49 proteínas alvos selecionadas e resumo de suas funções. .	48
Figura 7: Resultado do confronto das 49 proteínas contra o NR utilizando o PSI-Blast. ...	50
Figura 8: Proteínas de Tripanossomatídeos homólogas a C1QBP_HUMAN encontradas no NR.....	51
Figura 9: Identificação do domínio funcional das 25 proteínas selecionadas no NR.....	52
Figura 10: Análise comparativa do domínio MAM33 entre C1QBP e proteínas homólogas em tripanossomatídeos.....	53
Figura 11: Predição de localização celular das 21 proteínas homologas a C1QBP.....	54
Figura 12: Alinhamento das 18 proteínas homólogas a C1QBP com domínio MAM33.....	57
Figura 13: Divisão das 18 proteínas homólogas a C1QBP em grupos e subgrupos.....	58
Figura 14: Alinhamento das proteínas do grupo A com similaridade a C1QBP.....	59
Figura 15: Alinhamento das proteínas do grupo B de similaridade com C1QBP	60
Figura 16: Alinhamento das proteínas do grupo C de similaridade com C1QBP	61
Figura 17: Domínio funcional das proteínas homólogas à C1QBP encontradas na base de dados TriTrypDB.....	64
Figura 18: Predição de direcionamento proteico das proteínas homólogas à C1QBP selecionadas em TriTrypDB.....	65
Figura 19: Equivalencia entre as proteínas das bases de dados TriTrypDB e UniProtKB. .	66
Figura 20: Proteínas selecionadas do GenBank consideradas homólogas à C1QBP	69
Figura 21: Identificação de domínio das 61 proteínas selecionadas homologas a C1QBP em TriTrypDB.....	70
Figura 22: Proteínas selecionadas do GenBank consideradas homólogas a C1QBP.....	71
Figura 23: Predição e direcionamento proteico para proteínas homólogas a C1QBP e com MAM33 em GenBank.....	72
Figura 24: Patógenos selecionados do GeneDB que apresentam proteínas homólogas a	

C1QBP e com MAM33.....	74
Figura 25: Análise comparativa do domínio MAM33 das proteínas encontradas em patógenos no GeneDB.....	75
Figura 26: Distribuição das proteínas homólogas a C1QBP e com MAM33 encontradas nas 3 bases de dados simultaneamente.....	76
Figura 27: Distribuição das proteínas homólogas a C1QBP, que possuem domínio MAM33 e direcionamento para mitocôndria encontradas nas 3 bases de dados simultaneamente.....	77
Figura 28: Alinhamento das proteínas comuns nas 3 bases de dados proteicas.....	79
Figura 29: Alinhamento das proteínas do grupo A, comuns nas 3 bases de dados proteicas.....	81
Figura 30: Alinhamento das proteínas do grupo B, comuns nas 3 bases de dados proteicas.....	82
Figura 31: Filograma das 11 proteínas encontradas nas 3 bases de dados proteicas simultaneamente.....	82
Figura 32: Estrutura triméricas da C1QBP e da gC1q.....	84
Figura 33: Alinhamento de aminoácidos de C1QBP e Hq412588.....	86
Figura 34: Alinhamento entre as sequências de C1QBP, K2NJJ1 e Q4DG13.....	87
Figura 35: Estrutura de Q4DG13.....	88
Figura 36: Estrutura dos monômeros das proteínas tipo C1QBP.....	89
Figura 37: Comparação entre os homotrímeros das proteínas C1QBP's.....	91
Figura 38: Concorrência entre o complexo antígeno anticorpo e o domínio MAM33.....	102
Figura 39: Comparação entre os locais de ação entre o domínio MAM33 e calreticulina.....	102

LISTA DE TABELAS

Tabela 1: Comparação entre as sequências modeladas neste trabalho e seus modelos de referência. Em negrito estão evidenciadas as estruturas de modelos escolhidas pelo servidor SWISS-MODEL. As proteínas C1qBP_HUMAN e Hq412588 foram modeladas com base na estrutura humana (1P32), e Q2DG13 e K2NJJ1 foram modeladas baseadas na estrutura da proteína p22 de *T. brucei* (3JV1).....86

LISTA DE QUADROS

Quadro 1: Proteínas localizadas na membrana que regulam a ativação do sistema complemento	23
Quadro 2: Proteínas receptoras de membrana que ativam o sistema complemento.....	24
Quadro 3: Sumário da anotação genômica do <i>Trypanosoma cruzi</i>	27
Quadro 4: Extensão das famílias de gene do <i>Trypanosoma cruzi</i>	28
Quadro 5: Código dos genes das 21 proteínas homólogas a C1qBP com domínio MAM33 buscados no GenBank.....	63
Quadro 6: Proteínas homólogas a C1qBP selecionadas do TriTrypDB.....	64
Quadro 7: Tripanossomatídeos selecionados para busca de proteínas na base de dados do GenBank.....	68
Quadro 8: Espécies de tripanossomatídeos selecionadas com respectivos códigos taxonômico.....	69
Quadro 9: Proteínas com seus respectivos organismos encontradas nas três bases de dados proteicas simultaneamente.....	78

SUMÁRIO

1 INTRODUÇÃO.....	15
1.1 Tripanossomatídeos.....	15
1.2 Leishmania.....	15
1.3 Trypanosoma brucei.....	16
1.4 Trypanosoma cruzi.....	16
1.5 SISTEMA COMPLEMENTO E IMUNIDADE INATA.....	20
1.6 REGULADORES DO SISTEMA COMPLEMENTO.....	22
1.7 EVASÃO DO Trypanosoma cruzi AO SISTEMA COMPLEMENTO.....	24
1.8 BIOINFORMÁTICA.....	25
1.9 PROJETO GENOMA DO Trypanosoma cruzi.....	25
1.10 OBJETIVO.....	29
2 MATERIAIS E MÉTODOS.....	30
2.1 ESTRUTURA DE BANCO DE DADOS.....	30
2.1.1 BANCO DE DADOS DE PROTEÍNAS ALVOS.....	30
2.1.2 BANCO DE DADOS DE TRIPANOSSOMATÍDEOS.....	31
2.2 BASES DE DADOS DE INFORMAÇÃO DE PROTEÍNAS UTILIZADAS.....	31
2.2.1 UNIPROTKB (Uniprot Knowledgebase).....	31
2.2.2 GENBANK	32
2.2.3 TRITRYDB (KINETOPLASTID GENOMIC RESOURCE).....	32
2.2.4 GENEDB.....	33
2.3 PROGRAMAS DE BIOINFORMÁTICA.....	33
2.3.1 BLASTP.....	33
2.3.2 PSI-BLAST (POSITION-SPECIFIC ITERATIVE BASIC LOCAL ALIGNMENT SEARCH TOOL).....	34
2.3.3 PFAM.....	35
2.3.4 SIGNALP.....	35
2.3.5 TARGETP.....	36
2.3.6 PRED-GPI.....	36
2.3.7 CLUSTAL OMEGA.....	36
2.4 RECURSOS COMPUTACIONAIS AUXILIARES	37
2.4.1 MYSQL.....	37
2.4.2 PERL.....	37

2.5 FLUXO DE ANÁLISES REALIZADAS DURANTE A PESQUISA.....	38
2.5.1 SELEÇÃO DE PROTEÍNAS ALVO.....	39
2.5.2 BUSCA DE PROTEÍNAS HOMÓLOGAS UTILIZANDO PSI-BLAST.....	39
2.5.3 BUSCA DE DOMÍNIOS FUNCIONAIS NAS PROTEÍNAS CANDIDATAS.....	39
2.5.4 ESTUDO DE DIRECIONAMENTO PROTEICO DAS PROTEÍNAS CANDIDATAS.	40
2.5.5 SIMILARIDADE ENTRE PROTEÍNAS CANDIDATAS.....	40
2.5.6 DETERMINAÇÃO DOS GENES QUE CODIFICAM ÀS PROTEÍNAS CANDIDATAS	40
2.6 CONSULTA AO TRITRYPDB.....	41
2.6.1 ANÁLISE DE DOMÍNIO FUNCIONAL NO RESULTADO DE TRITRYPDB.....	41
2.6.2 ESTUDO DO DIRECIONAMENTO PROTEICO.....	41
2.6.3 INTEGRAÇÃO DAS INFORMAÇÕES.....	42
2.7 CONSULTA AO GENEDB.....	42
2.8 CONSTRUÇÃO DE TABELA LOCAL COM ORGANISMOS ESPECÍFICOS.....	43
2.8.1 ANOTAÇÃO DE PROTEOMAS E PROTEÍNAS.....	43
2.8.2 COMPARAÇÃO DAS PROTEÍNAS ALVOS COM OS ORGANISMOS DO GENBANK.....	43
2.8.3 FILTRAGEM POR PROTEÍNAS ALVOS RELEVANTES.....	43
2.8.4 ANÁLISE DO DOMÍNIO FUNCIONAL.....	44
2.8.5 ANÁLISE DE DIRECIONAMENTO PROTEICO.....	44
2.9 INTEGRAÇÃO DOS RESULTADOS.....	44
2.9.1 COMPARAÇÃO DOS DOMÍNIOS FUNCIONAIS.....	44
2.9.2 COMPARAÇÃO DO DIRECIONAMENTO PROTEICO.....	45
2.9.3 MODELAGEM COMPARATIVA E CALCULO DA DISTRIBUIÇÃO DE CARGA.....	45
3 RESULTADOS.....	46
3.1 PROTEÍNAS ALVOS.....	48
3.2 DADOS DAS PROTEÍNAS ALVOS.....	49
3.3 BUSCA POR PROTEÍNAS HOMÓLOGAS ÀS PROTEÍNAS ALVOS NA BASE DE DADOS NR.....	49
3.3.1 ANÁLISE DO DOMÍNIO FUNCIONAL DAS PROTEÍNAS SELECIONADAS.....	52
3.3.2 COMPARAÇÃO DO DOMÍNIO MAM33 DAS PROTEÍNAS DE TRIPANOSSOMATÍDEOS COM O DOMÍNIO MAM DE C1QBP_HUMAN.....	53
3.3.3 INVESTIGAÇÃO DE DIRECIONAMENTO PROTEICO DAS PROTEÍNAS HOMÓLOGAS À C1QBP_HUMAN E COM DOMÍNIO MAM33.....	54
3.3.4 ANÁLISE DE SEMELHANÇA ENTRE AS PROTEÍNAS DE	

TRIPANOSOMATÍDEOS.....	55
3.3.5 IDENTIFICAÇÃO DOS GENES CODIFICADORES DAS PROTEÍNAS SELECIONADAS.....	61
3.4 CONSULTA A OUTRAS BASES DE DADOS GENÔMICOS.....	62
3.4.1 BUSCA POR PROTEÍNAS HOMÓLOGAS A C1QBP_HUMAN NA BASE DE DADOS TRITRYPDB.....	63
3.4.1.1 ANÁLISE DE DOMÍNIO FUNCIONAL DAS PROTEÍNAS SELECIONADAS EM TRITRYPDB.....	64
3.4.1.2 BUSCA POR PROTEÍNAS HOMÓLOGAS AS PROTEÍNAS ALVOS NA BASE DE TRITRYDB.....	65
3.4.1.3 INTEGRAÇÃO ENTRE BASE DE DADOS.....	66
3.4.2 SELEÇÃO DE TRYPANOSOMATIDEOS PARA ANALISE DE sequênciA NO GENBANK.....	67
3.4.2.1 COLETA DE DADOS NO GENBANK.....	67
3.4.2.2 CONFRONTO DAS PROTEÍNAS ALVOS CONTRA AS PROTEÍNAS SELECIONADAS DO GENBANK.....	68
3.4.2.3 ANÁLISE DO DOMÍNIO FUNCIONAL DAS PROTEÍNAS.....	70
3.4.2.4 INVESTIGAÇÃO DE DIRECIONAMENTO DAS PROTEÍNAS COM DOMÍNIO MAM33.....	72
3.5 PROTEÍNAS HOMÓLOGAS A C1QBP_HUMAN E COM DOMÍNIO MAM33 EM DIFERENTES PATÓGENOS DA BASE DE DADOS GENEDB.....	72
3.6 INTEGRAÇÃO DOS DADOS COLETADOS NAS TRES BASES DE DADOS NR, TRITRYPDB E GENBANK.....	75
3.6.1 MINERAÇÃO DE DADOS ENTRE BASES DE PROTEÍNAS.....	75
3.6.2 ALINHAMENTO E ANÁLISE DE SEMELHANÇA.....	78
3.6.2.1 AGRUPAMENTO POR SEMELHANÇA DE ALINHAMENTO.....	80
3.7 MODELAGEM COMPARATIVA DAS PROTEÍNAS C1QBP.....	83
4 DISCUSSÃO.....	92
4.1 RECURSOS DE BIOINFORMÁTICA UTILIZADOS.....	92
4.1.1 GERAÇÃO DE BANCO DE DADOS.....	92
4.1.2 FERRAMENTAS DE BIOINFORMÁTICA UTILIZADAS.....	93
4.1.2.1 FERRAMENTAS DE ALINHAMENTO.....	93
4.1.2.2 FERRAMENTAS DE LOCALIZAÇÃO DE DOMÍNIO.....	93
4.1.2.3 FERRAMENTAS DE LOCALIZAÇÃO PROTEICA.....	94
4.1.2.4 FERRAMENTAS DE ALINHAMENTO DE SEQUÊNCIAS.....	95

4.2 MINERAÇÃO DE DADOS E PROTEÍNAS DE Trypanosoma cruzi.....	96
4.3 Trypanosoma cruzi E SEU GENOMA.....	96
4.4 ESTRATÉGIA DE EVASÃO DO COMPLEMENTO POR PATÓGENOS.....	97
4.5 EVASÃO DO COMPLEMENTO POR TRIPANOSOMATÍDEOS.....	99
4.6 INIBIÇÃO DE ATIVIDADE POR ESTRUTURA PROTEICA.....	100
4.7 Trypanosoma cruzi POSSUI C1QBP-LIKE.....	101

1 INTRODUÇÃO

Nos últimos anos, a ciência obteve consideráveis avanços nas áreas da Saúde, sendo um bom indicador destes avanços o aumento de investimento nacional e internacional referente ao controle e eliminação das doenças, planejamentos e ações de saúde pública, fornecimento de medicamentos, pesquisas e orientação da população sobre as doenças (WHO, 2013). Porém apesar dos investimentos realizados, milhares de pessoas morrem a cada ano por doenças que poderiam ser prevenidas e/ou tratadas.

Em 2013 a Organização Mundial de Saúde (WHO) definiu dengue, doença de Chagas, raiva e leishmaniose como doenças negligenciadas de clima tropical.

1.1 Tripanossomatídeos

Trypanosomatidae é uma família que inclui um conjunto de protozoários flagelados, possuindo muitos deles um ciclo de vida com uma alternância entre hospedeiros vertebrados e invertebrados e uma variação de sua forma física a cada estágio do seu ciclo de vida. O *T. cruzi*, *T. brucei* e a *Leishmania* são espécies de *Trypanosomatidae* causadores de doenças que afetam cerca de 550 milhões de pessoas em todo o mundo e devido a isso, estas três espécies são as mais estudadas hoje em dia pela ciência médica (TEIXEIRA; DAROCHA, 2003).

1.2 *Leishmania*

A *Leishmania* é um gênero de protozoários que inclui o agente etiológico causador da leishmaniose e que é transmitida ao homem através de insetos flebótomos como o *Lutzomyia longipalpis*, conhecido popularmente como mosquito-palha. Este protozoário causa doenças que variam de úlceras cutâneas a infecções viscerais fatais em mais de 12 milhões de pessoas em todo o mundo (UENO; WILSON, 2012).

1.3 *Trypanosoma brucei*

O *T. brucei* é o agente causador da tripanossomíase africana em humano (HAT), mais conhecida popularmente como doença do sono. A mosca *Glossina palpalis*, apelidada de mosca tsé-tsé é vetor que transmite a doença para o hospedeiro vertebrado (GOODHEAD *et al.*, 2013).

1.4 *Trypanosoma cruzi*

O *T. cruzi* é o agente etiológico da doença de Chagas (tripanossomíase americana) que é uma importante doença tropical negligenciada (WHO, 2013) causadora de um alto número de morbidade e mortalidade em toda América Latina.

O *T. cruzi* infecta insetos da Família *Reduviidae*, onde *Triatomas infestans* e *Panstrongylus megistus* são as espécies mais presentes no Brasil e são popularmente conhecidos como barbeiros. A doença de Chagas possui uma fase aguda e uma fase crônica. A fase aguda, que é frequente em 1% dos casos, ocorre entre 5 a 14 dias da infecção, podendo levar o indivíduo infectado à morte. Alguns de seus sintomas são; febre, mal-estar, inchaço no olho e inchaço e vermelhidão no local da picada do inseto vetor. A fase crônica pode ocorrer entre 10 a 30 anos após a fase aguda aparente ou inaparente. Alguns de seus sintomas são; constipação, problemas digestivos, dor no abdome e dificuldade ao respirar. A fase crônica sintomática causa parasitemia com lesões típicas no coração e tubo digestivo como; cardiopatia chagásica crômica, megaesôfago e megacólon. Enquanto isso, os pacientes inseridos na fase crônica indeterminada são soro positivos e sem sintomas. O *T. cruzi* é endêmico da América Latina, onde atinge mais de 16 milhões de pessoas infectadas, porém atualmente devido à grande imigração dos povos vem sendo encontrado também em áreas não endêmicas (ACKERMANN *et al.*, 2012; MACHADO *et al.*, 2012). Este protozoário apresenta um ciclo de vida complexo que consiste em quatro formas físicas distintas envolvendo hospedeiros vertebrados e invertebrados (MACHADO *et al.*, 2012) (FIGURA 1).

Doença de Chagas: Ciclo de Vida do *Trypanosoma cruzi*

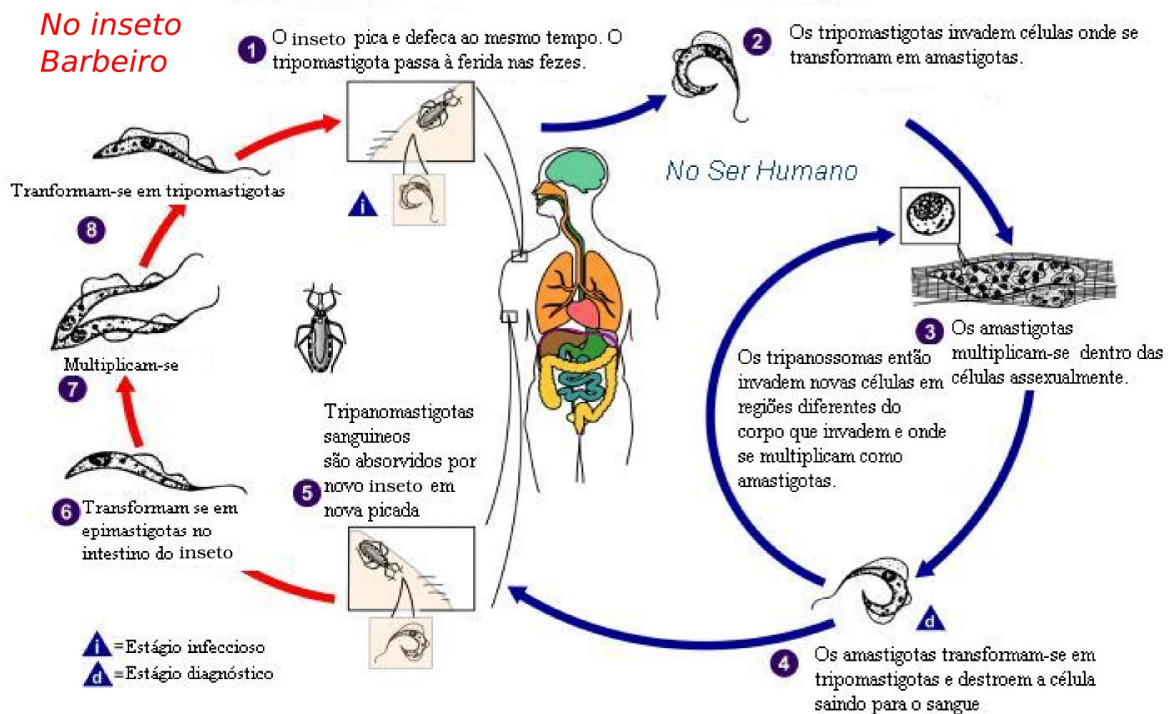


FIGURA 1: Ciclo de vida do *Trypanosoma cruzi* entre os hospedeiros invertebrado e vertebrado: (1) O vetor contaminado com *T. cruzi* após seu repasto sanguíneo elimina fezes contendo o patógeno em sua forma tripomastigota metacíclico, que irá invadir o mamífero através da mucosa oral ou nasal, ferida na pele ou pelo orifício da própria picada do inseto. (2) Circulando pelas vias sanguíneas na forma tripomastigota metacíclico, o *T. cruzi* invade células do hospedeiro e altera sua forma para amastigota, (3) que se multiplica dentro da célula através de múltiplas divisões binárias até (4) alterar sua forma para tripomastigota sanguínea, causando a ruptura da célula e voltando a transitar pelas vias sanguíneas, onde poderão invadir novas células. (5) O vetor ao realizar um repasto sanguíneo pode adquirir formas tripomastigotas sanguínea com o sangue ingerido. (6) Uma vez no organismo do hospedeiro invertebrado, o *T. cruzi* altera sua forma de tripomastigota sanguínea para epimastigota que (7) realiza multiplicação binária no estômago do vetor. (8) No reto do vetor o *T. cruzi* altera sua forma para tripomastigota metacíclico, podendo infectar um mamífero em um novo repasto sanguíneo do vetor, reiniciando o ciclo de vida do *T. cruzi* (MACHADO *et al.*, 2013).

FONTE: Adaptado de Center for Disease Control and Prevention (<http://www.cdc.gov/parasites/chagas/biology.html>, 2009).

O *T. cruzi* apresenta uma biologia celular peculiar e algumas organelas específicas aparecem descritas no esquema de um tripomastigota, apresentado na FIGURA 2. O parasita possui uma mitocôndria que ocupa grande parte do seu citoplasma e que possui em seu interior um cinetoplasto que contém uma complexa rede de DNA. O DNA mitocondrial do *T. cruzi* difere de outros eucariotos, devido ao seu enorme tamanho, forma complexa e grande número de genes. Está composto por *maxicírculos*

concatenados possuidores de genes de rRNA e milhares de *minicírculos* heterogêneos que codificam pequenos RNAs (WESTENBERGER *et al.*, 2006). Também possui um grande flagelo que o acompanha em todas as formas físicas de seu ciclo de vida, estando mais discreto apenas na forma amastigota. (PAES *et al.*, 2011). O *T. cruzi* possui organelas específicas como reservossomo e acidocalcissomo. Reservossomo são encontrados na forma física epimastigota do parasita, são esféricos e possuem um conteúdo elétron-denso e inclusões elétron-lucentes. No reservossomo são depositadas todas as moléculas capturadas pelo processo endocítico do *T. cruzi*. Tem a função de armazenamento de lipídios, proteínas e proteases importantes, e também, possivelmente tem as funções de armazenamento, reciclagem e funções típicas de lisossomos, degradando macromoléculas. Nas formas amastigota e tripomastigota estas organelas são apresentadas na forma de reservossomo-like. Os acidocalcissomos são organelas ricas em cálcio, polifosfato e outros íons. Possuem uma forma esférica de 200 nm e estão distribuídas por toda célula, porém normalmente são encontradas na periferia celular. Tem a função de armazenamento de fosforo e cálcio, manutenção da homeostasia intracelular, manutenção do pH e osmorregulação (TEIXEIRA *et al.*, 2011).

O *T. cruzi* apresenta um genoma diploide e seus genes se encontram organizados em varias famílias multigênicas. A regulação da expressão genica em *T. cruzi* é complexa. O parasita carece de promotores de RNA polimerase II e na maioria dos casos não há introns nos seus genes. Diferindo dos eucariotos superiores o *T. cruzi* não precisa de ativadores para estimular a expressão de seus genes, realizando suas transcrições de forma aleatória, contínua ou basal (GÜNZL, 2010). O processamento do RNA é semelhante com os das bactérias onde a transcrição genica se dá por transcrição policistrônica e há processamento individual ou monocistrônico chamado de *trans-splicing*. Nesse mecanismo existe o processamento de cada gene na forma monocistrônicas através de cortes, polinadenililação e edição de suas extremidades com a adição de um mini exon na extremidade N-terminal constituindo a 5'UTR e a adição de uma molécula PoliA na extremidade C-terminal formando a 3'UTR, resultando em um RNAm (ácido ribonucléico mensageiro), responsável pela síntese protéica (PAES *et al.*, 2011).

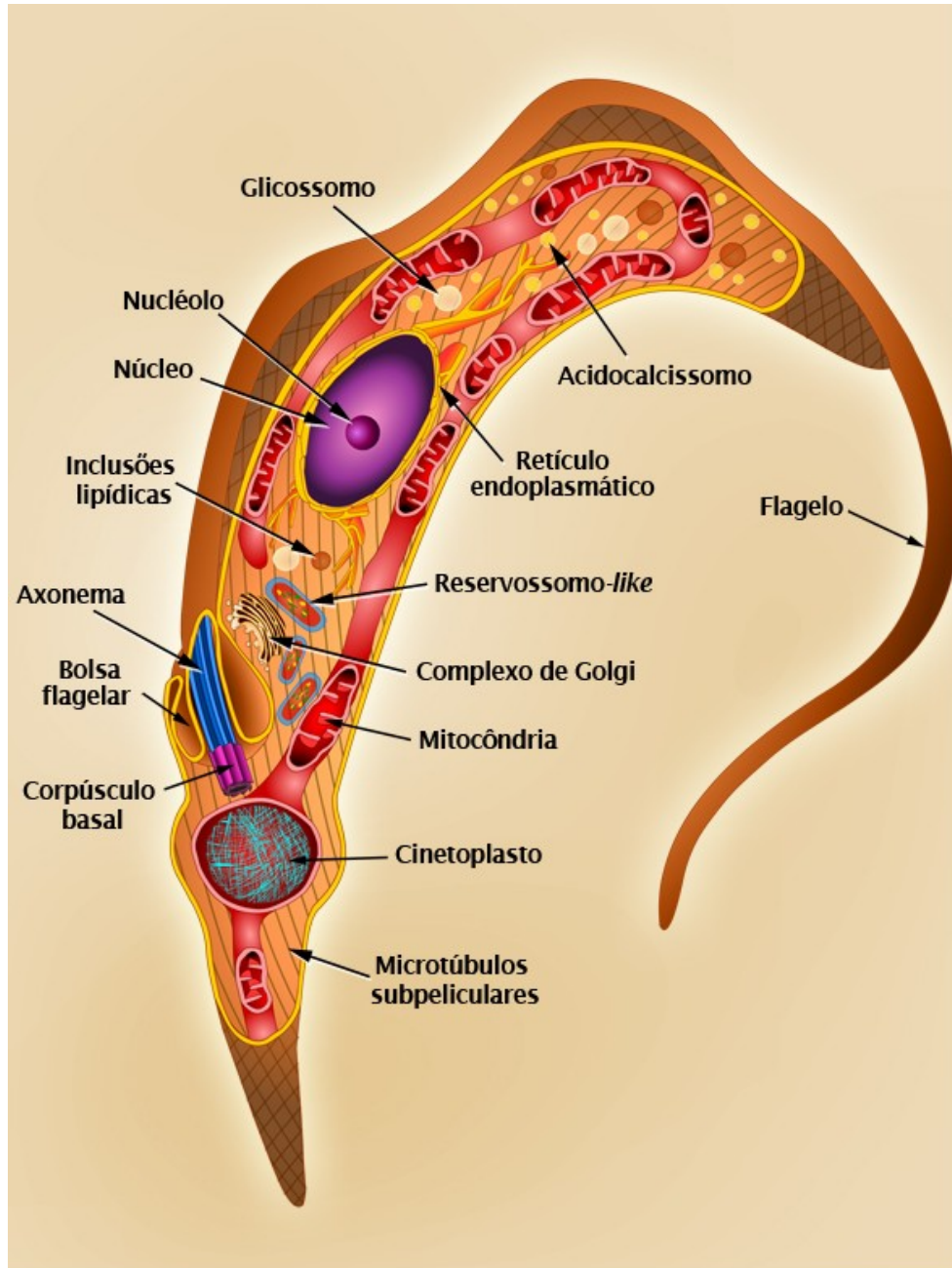


FIGURA 2: Organização estrutural da forma tripomastigota do *Trypanosoma cruzi*. No interior do *T. cruzi* se encontra uma única e grande mitocôndria que se distribui pelo corpo celular. No interior da mitocôndria, próxima ao corpúsculo basal (início do flagelo) é encontrado o cinetoplasto, local onde é armazenado o DNA mitocondrial. O núcleo é alongado e localizado na região central da célula, ao seu arredor e por toda célula são encontrados retículos endoplasmáticos e em uma proporção pequena o complexo de Golgi. O flagelo emerge da bolsa flagelar e adere ao longo do corpo celular tornando-se livre na outra extremidade. Também são encontradas organelas semelhante aos lisossomos chamadas reservossomo-like. O acidocalcissomo são organelas com características ácidas, ricas em cálcio e outro íons, e por isso possui esse nome. Diversas inclusões lipídicas também são encontradas por todo corpo celular do *T. cruzi* (TEIXEIRA *et al.*, 2011).
 FONTE: TEIXEIRA (2011)

1.5 SISTEMA COMPLEMENTO E IMUNIDADE INATA

Quando o parasita infecta hospedeiros mamífero, enfrenta uma série de barreiras biológicas que representam os sistemas de imunidade do hospedeiro. Que tem finalidade de proteger o hospedeiro contra patógenos invasores. Entre os mecanismos mais relevantes são: imunidade local, células NK (*Natural Killer*) e o sistema complemento (LAMBRIS *et al.*, 2008). No primeiro contato o parasita deve enfrentar barreiras físicas do hospedeiro, como pele e superfície mucosas do sistema digestivo e respiratório, assim como células fagocíticas e as células NK. Tanto os neutrófilos como os macrófagos podem reconhecer, diretamente, patógenos invasores, através de estruturas conservadas na membrana como manose e lipopolissacarídeos (CESTARI *et al.*, 2013).

Dentre os três mecanismos de imunidade mais relevantes, o sistema complemento é o primeiro mecanismo a combater a infecção de patógenos, constituído por um conjunto de proteínas séricas ativadas em cascata que culmina na formação de um complexo de ataque à membrana do parasita resultando na lise do mesmo. Há três vias de ativação do sistema complemento: a clássica, a alternativa e a lectina (LAMBRIS *et al.*, 2008; EVANS-OSSES *et al.*, 2013) (FIGURA 3). As proteínas do sistema complemento estão presentes nas no sangue em sua forma inativa ou em um baixo nível de ativação espontânea (CESTARI *et al.*, 2013).

O fragmento C5b presente nas três vias do sistema complemento se fixa na membrana do patógeno se agrupando aos componentes C6, C7 e C8. Esse agrupamento ancorado ao patógeno permite a formação de um poro formado por conjunto de proteínas C9 (Poli-C9) na membrana (ABBAS *et al.*, 2012). Por este poro entram eletrólitos e água que provocam a lise do patógeno (ABBAS *et al.*, 2012).

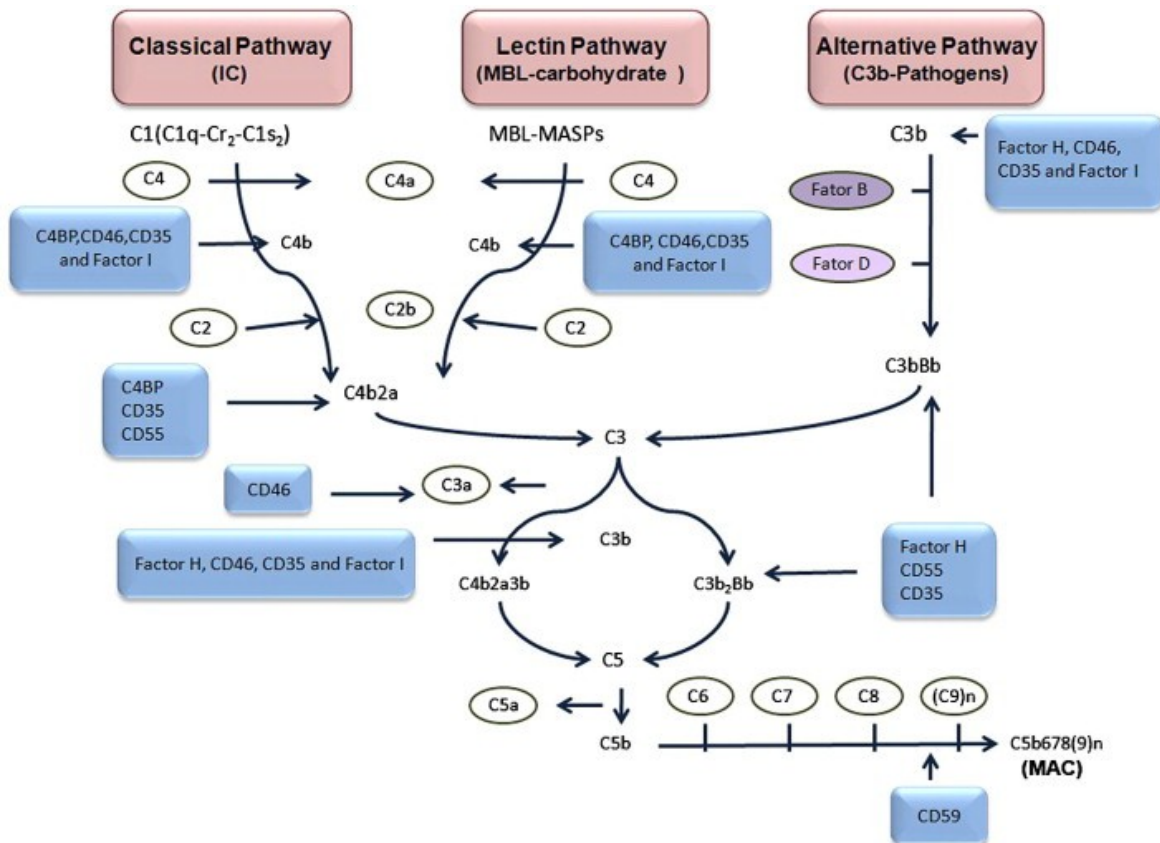


FIGURA 3: Ativação do sistema complemento. Há três vias que ativam o sistema complemento: via da lectina, via clássica e via alternativa. A via da lectina é ativada quando a lectina de ligação a membrana (MBL) ou ficolinas reconhecem as moléculas que contêm carboidrato na superfície do patógeno. Forma-se então um complexo MBL associado à serina protease 2 (MASP2) que cliva C2 e C4, resultando na formação de C3 convertase C4b2a. A C3 convertase pode clivar C3 em C3a e C3b, onde C3b associado com C3 convertase (C4b2a ou C3bBb) forma a C5 convertase C4b2a3b. A C5 convertase pode clivar C5 em C5a e C5b. O fragmento C5b se fixa na superfície do patógeno formando uma ancora e conseqüentemente agrupando-se com C6, C7 e C8 constituindo o complexo de ataque à membrana (MAC), que é formado por várias moléculas C9. O MAC é um poro que irá lisar o patógeno permitindo o fluxo de água e de eletrólitos. A ativação do sistema complemento pela via clássica é dependente da resposta de anticorpos contra o patógeno. Ela ocorre pela ligação do complexo C1 (várias C1q, 2 C1s e 2 C1r) em complexos antígeno-anticorpos que reconhecem o agente patogênico. Uma vez ligado à superfície do patógeno, o complexo C1 é ativado clivando C2 em C2a e C2b e C4 em C4a e C4b gerando então a C3 convertase C4b2a (semelhante à via da lectina). A ativação da via clássica depende exclusivamente de C3b que pode ser formado por hidrólise espontânea de C3 (H₂O) ou pelo C3b gerado pela ativação das vias lectina ou alternativa (da atividade de C3 convertase). Na via alternativa, uma vez que C3b é fixado na superfície do patógeno, a molécula do fator B se liga a C3b. O fator B é clivado por uma molécula solúvel chamada fator D gerando os fragmentos Ba e Bb. O fragmento Bb se junta a C3b formando a C3 convertase da via alternativa na superfície do patógeno (que difere da via clássica, C4b2a). Esta C3 convertase também cliva C3 conforme as outras vias. Uma vez a C3 convertase formada, todas as etapas subsequentes são comuns para todas as três vias (CESTARI *et al.*, 2013; PICOLI *et al.*, 2011).

FONTE: PICCOLI *et al.* (2011)

1.6 REGULADORES DO SISTEMA COMPLEMENTO

O sistema complemento possui capacidade de lisar tanto micro-organismos invasores quanto as células saudáveis do hospedeiro. Devido a isso, vários mecanismos são envolvidos na regulação desse sistema, restringindo desta forma sua ação apenas para objetos específicos. Também uma ativação descontrolada do complemento pode levar à formação do complexo lítico da membrana no próprio tecido e à formação excessiva de mediadores da inflamação. Isso normalmente ocorre devido a ativação ser regulada por várias proteínas plasmáticas e outras ligadas a membrana celular com funções específicas, mantendo um controle rigoroso da ativação (CESTARI *et al.*, 2012). Graças a estas proteínas reguladoras o organismo pode realizar um delicado controle de inibição e ativação da cascata do sistema complemento. Os quadros 1 e 2 apresentam proteínas que possuem características de regulação do sistema complemento (YAMAMOTO; PORTINHO, 2001).

Casos também, como a deficiência genética para a síntese das proteínas envolvidas com o complemento, podem evitar o funcionamento do sistema no organismo quando esta deficiência presente nos primeiros elementos das vias clássica da lectina interrompem ou perturbam o funcionamento do sistema, já que a via alternativa inicia sua ativação com C3b. Por isso a deficiência de C3 ou outros componentes da via efetora do sistema complemento permite o indivíduo a desenvolver doenças meningocócicas com frequência (YAMAMOTO; PORTINHO, 2001).

Deficiência genéticas em proteínas reguladoras do sistema complemento também podem ocorrer. A deficiência do fator I pode diminuir a quantidade de C3 devido o alto consumo de C3b tornando o indivíduo mais suscetível ao doenças bacterianas (YAMAMOTO; PORTINHO, 2001).

Os organismos efetores também utilizam de proteínas reguladoras e inibidoras do sistema complemento para obterem sucesso em suas invasões impossibilitando a ação lítica do sistema complemento sobre sua membrana (MACHADO *et al.*, 2012).

Proteína	PM(Kd)	Distribuição celular	Interação com	Função
Receptor de complemento tipo I	190-280	E,B,G,M,L, Miócito*	C3b, C4b, iC3b	Acelera a dissociação de C3 e C5 convertase da via clássica e da via alternativa. Atua com cofator do Fator I na clivagem de C3b e C4b. Promove a eliminação de imunocomplexos em circulação.
Proteína cofatora de membrana (PCM)	45-70	B, T, N, M, Cels. endoteliais, epiteliais, fibroblastos e miócitos*	C3b, C4b	Atua como cofator do Fator I na clivagem de C3b e C4b.
Fator acelerador da dissociação (DAF)	70	E, L, P, Miócito*	C4b2a, C3bBb	Acelera a dissociação das C3 convertases Previne a associação de C3b com fator B e C4b com C2.
Fator de restrição homólogo (HRF) (C8bp)	65	E, L, M, N, P Miócito§	C8, C9	Interfere na ligação de C8 e C9 prevenindo a formação de CLM em células homólogas.
Inibidor de membrana de lise reativa (MIRL) (CD59) (protectina)	18 Miócito§	E, L, M, N, P	C7, C8 complexo C5b-6.	Bloqueia a ligação de C7 e C8 ao

Kd=Kilodalton. E=eritrócitos. B=linfócitos B. G=granulócitos. M=monócitos. L=leucócitos. T=linfócitos T. N=neutrófilos.

P=plaquetas. CLM=complexo lítico de membrana.

***Fraca expressão. § Forte expressão.**

QUADRO 1: Proteínas localizadas na membrana celular que regulam a ativação do sistema complemento

FONTE: YAMAMOTO (2001)

Proteína	Peso Molecular (kd)	Interação específica com:	Distribuição celular	Função
Receptor tipo 1 (CRI)	190-280	C3b, C4b, iC3b	E, miócito	Promove a eliminação de imunocomplexos da circulação. Acelera a dissociação de C3 e C5 convertases de ambas as vias. Atua como cofator do Fator I na clivagem de C3b e C4b.
			N, M, m	Incrementa a fagocitose mediada pelo receptor Fc. Intervém na fagocitose independente do receptor Fc.
			E	?
			B	?
			T	?
Receptor tipo 2 (CR2)	145	iC3b, C3dg, C3d, C3b, VEB	B	Regulação das funções de linfócitos B. Receptor para VEB.
Receptor tipo 4 (CR4)	260	iC3b	N, M, P	Incrementa a fagocitose mediana pelo receptor Fc. Intervém na fagocitose independentemente do fator
Fc.				
Receptor para C3a/C3b	100	C3a, C4a	B Mastócitos	Ligação de anafilatoxinas C3a, C4a. Liberação de histamina e outros mediadores da inflamação
			Céls. endot., N, M, m, P.	Incremento na permeabilidade celular
				Promove a quimiotaxia.
Receptor Clq	65	Clq	B, M, P, m, céls. endot.	Intermedeia a ligação de imunocomplexos à células fagocitárias.

Kd=kilodaltons. E=eritrócitos. B=linfócitos B. G=granulócitos. M=monócitos. L=leucócitos. T=linfócitos T. N=neutrófilos. P=plaquetas. m=macrófagos. e=eosinófilos. VEB=vírus Epstein-Baar.

QUADRO 2: Proteínas receptoras de membrana que ativam o sistema complemento.
FONTE: YAMAMOTO (2001)

1.7 EVASÃO DO *Trypanosoma cruzi* AO SISTEMA COMPLEMENTO

O *T. cruzi*, para invadir células e se multiplicar dando sequência à infecção causadora da doença, necessita evadir-se a resposta imune inata do mamífero contaminado. O sistema complemento é a primeira barreira de defesa que o *T. cruzi* encontra no hospedeiro vertebrado (CESTARI *et al.*, 2009). Estudos realizados utilizando o fator B e o componente C3 contra *T. cruzi* em suas formas epimastigota e tripomastigota mostraram que C3 adere à superfície da forma epimastigota, porém na forma tripomastigota o fragmento de inativação iC3b foi encontrado em sua superfície (MACHADO *et al.*, 2012; CESTARI *et al.*, 2013). *T. cruzi* é capaz de ativar rapidamente a

via das lectinas do sistema complemento (CESTARI *et al.*, 2009) e tem sido visto que simultaneamente desencadeia mecanismos de resistência para sobreviver (CESTARI *et al.*, 2012, revisado por CESTARI *et al.*, 2013 e EVANS-OSES *et al.*, 2014). Proteínas como calreticulina que inibe o complexo C1 (FERREIRA *et al.*, 2004; VALCK *et al.*, 2010; SÁNCHEZ-VALDÉZ *et al.*, 2014) e T-DAF (*decay-acceleration factor*) regulando C3 convertase (JOINER *et al.*, 1988) também são sintetizadas pelo *T. cruzi* e permitem que o mesmo evada-se da ação do sistema complemento.

Nos últimos anos tem sido mostrado proteínas de *T. cruzi* e mecanismos capazes de inibir a via da lectina do sistema complemento, mostrando mecanismos de evasão para essa eficiente via de controle parasitaria (revisado por CESTARI *et al.*, 2013).

1.8 BIOINFORMÁTICA

A bioinformática originou-se da primeira necessidade da ciência em processar elevada quantidade de dados biológicos, realizando anotações e armazenamento, com o objetivo de gerar informações que auxiliassem nas respostas de suas perguntas biológicas. Com o avanço tecnológico e a abundância de informações biológicas, principalmente com os inúmeros genomas montados nos últimos anos, a bioinformática se tornou peça chave para a administração, manipulação e processamento destes dados, sendo construídos diversos bancos de dados genômicos (UniProtKB, GenBank, etc) e ferramentas que abrangem várias áreas como: análise de prognósticos de radiografias, estudos comportamentais através de análise complexa de dados, genômica comparativa e estrutura de proteínas. Hoje em dia, cada vez mais se encontra a utilização de recursos e estratégias de bioinformática sendo utilizadas para o auxílio de projetos científicos (PROSDOCIMI, 2007; ACKERMANN *et al.*, 2012).

1.9 PROJETO GENOMA DO *Trypanosoma cruzi*

Varias iniciativas de sequenciamento do genoma foram iniciadas nas últimas décadas e, finalmente, em 2005 um consórcio multinacional publicou o sequenciamento

da cepa CL brener (EL-SAYED *et al.*, 2005). O sequenciamento total do genoma deste protozoário revelou que o genoma diploide contém genes que codificam 22.570 proteínas. O genoma do *T. cruzi* possui o maior conteúdo genômico (60 MB de pares de bases) entre os tripanossomatídeos sequenciados até o momento (QUADRO 3). Esse genoma apresenta um grande número de sequências repetitivas e particularmente na grande quantidade de genes que se organizam em famílias multigênicas (QUADRO 4). Os genes de *T. cruzi* apresentaram diferenças entre outros organismo eucariotos em relação reparo DNA e um incomum DNA mitocondrial. Embora haja ausência de várias moléculas de regulação, o *T. cruzi* possui um conjunto grande de proteínas quinases e fosfatase que através de distintas interações realizam o processo de regulação gênica .

Entre as famílias multigênicas constituídas por centos de genes temos a GP85-transialidase e a Família da mucinas, as duas famílias contendo genes codificam para proteínas de superfície envolvidas com interação parasita-célula hospedeira e mecanismos de evasão parasitária. As mucinas constituem uma família multigênica das mais amplas do genoma, sua função principal é proteger o parasita dos mecanismos de defesa do mamífero e do vetor, assegurando seu ponto de ancoragem e invasão de células e tecidos. Famílias multigênicas como DGF-1 e MASP são outros exemplos importantes que formam a complexa biologia de *T. cruzi* (EL-SAYED *et al.*, 2005; PABLOS, DE; OSUNA, 2012).

Mais de 50% do genoma do *T. cruzi* está constituído por sequências repetidas, como retrotransposons e famílias de genes que expressam moléculas de superfície, tais como; *trans-sialidases*, mucinas, gp63s e proteínas de superfícies associadas às mucinas (MASP).

Parameter	Number
<i>The genome</i>	
Size* (bp)	60,372,297
G+C content (%)	51
Sequence scaffolds†	838
Sequence contigs	4,008
Percent coding	58.9
<i>Protein-coding genes</i>	
No. of gene models	23,216
No. of genes‡	22,570
Estimated no. of genes per haploid genome§	~12,000
Pseudogenes	3,590
Mean CDS length (bp)	1,513
Median CDS length (bp)	1,152
G+C content (%)	53.4
Gene density (genes per Mb)	385
<i>Intergenic regions¶</i>	
Mean length (bp)	1,024
G+C content (%)	47
<i>RNA genes</i>	
tRNA	115
rRNA	219
sRNA	192
snRNA	19
snoRNA	1,447
srpRNA	2

QUADRO 3: Sumário da anotação genômica do *Trypanosoma cruzi*: O genoma está dividido entre as colunas (*Parameter*) que refere ao tipo de informação encontrada e a coluna (*Number*) contendo o número de registros ocorreram para cada parâmetro e as regiões (*The genome*) que mostra resultados referente as sequências do DNA do *T.cruzi* e (*RNA - genes*) apresentando dados referentes aos RNA dos genes funcionais encontrados. Legenda: *. Inclusão de regiões *scaffolds* e codificadores de genes maior que 5 kb, †. 784 *scaffolds* + 54 codificadores, ‡. Genes divididos por codificadores contados uma única vez, §. ver detalhes no Quadro 4, ||. Exclusão de genes parciais e pseudogenes, ¶. Regiões entre codificadores de proteínas CDS's.
FONTE: EL-SAYED (2005)

Gene product	Members	Tritryp orthologs
trans-Sialidase (TS)	1430 (693)	<i>Tb</i>
MASP	1377 (433)	No
Mucin	863 (201)	No
Retrotransposon hot spot (RHS) protein	752 (557)	<i>Tb</i>
Dispersed gene family protein 1 (DGF-1)	565 (136)	No
Surface protease (gp63)	425 (251)	<i>Tb + Lm</i>
Mucinlike protein	123	No
Hypothetical	117	<i>Lm+Tb</i>
Hypothetical	93	<i>Lm+Tb</i>
Kinesin, putative	79	<i>Lm+Tb</i>
Protein kinase (CMGC group)	77	<i>Lm+Tb</i>
Protein kinase (several groups)	79	<i>Lm+Tb</i>
Hypothetical protein	42	No
Glycosyltransferase	52	<i>Lm+Tb</i>
RNA helicase (eIF-4a)	47	<i>Lm+Tb</i>
Protein kinase (NEK group)	39	<i>Lm+Tb</i>
MASP-related	38	No
Glycosyltransferase	36	<i>Lm+Tb</i>
Hypothetical	35	<i>Lm+Tb</i>
Amino acid permease	28	<i>Lm+Tb</i>
AAA ATPase	33	<i>Lm+Tb</i>
Protein phosphatase	30	<i>Lm+Tb</i>
Heat shock protein HSP70	21	<i>Lm+Tb</i>
Protein kinase (STE group)	25	<i>Lm+Tb</i>
RNA helicase	23	<i>Lm+Tb</i>
Phosphatidylinositol phosphate kinase-related	23	<i>Lm+Tb</i>
Hypothetical	24	<i>Lm+Tb</i>
Elongation factor 1- γ (EF-1- γ)	22	<i>Lm+Tb</i>
DNA helicase (DNA repair)	21	<i>Lm+Tb</i>
Actin-related	20	<i>Lm+Tb</i>
Cysteine peptidase	20	<i>Lm+Tb</i>

QUADRO 4: Extensão das famílias de genes em *Trypanosoma cruzi*. A coluna (*Gene product*) apresenta os nomes das famílias gênicas, a coluna (*Members*) a lista todas de números de genes (pseudogenes e parentes) e a coluna (*Tritryp orthologs*), a existência desse gene em *Leishmania* e/ou *Trypanosma brucei*.

FONTE: EL-SAYED (2005)

1.10 OBJETIVO

O *T. cruzi* é um parasita intracelular obrigatório, e quando infecta um mamífero deve circular pelas vias sanguíneas até invadir uma célula para se multiplicar e progredir na infecção. Para ter sucesso o parasita evade-se a resposta imune inata, especialmente o sistema do complemento que constitui uma complexa rede de proteínas ativadas para lisar patógenos invasores. Já tem sido mostrada uma série de mecanismos de evasão ao complemento em *T. cruzi* e no presente trabalho utilizamos uma estratégia de bioinformática para procurar nos genomas de tripanossomatídeos genes codificadores de proteínas homólogas de fatores do sistema complemento ou de proteínas regulatórias humanas.

A pergunta central da dissertação pretende responder através de uma análise in silico se *T. cruzi* possui no genoma genes que codificam proteínas do complemento ou proteínas regulatórias. Dessa maneira ao serem expressas poderiam regular o sistema complemento, atuando como um mecanismo de evasão parasitária.

2 MATERIAIS E MÉTODOS

2.1 ESTRUTURA DE BANCO DE DADOS

Para armazenar as informações que seriam geradas em cada etapa do projeto um banco de dados local de proteínas nomeado BLDP foi desenvolvido com tabelas estruturadas para representar o grupo de dados depositados. As tabelas foram criadas a partir de um modelo relacional e normalizadas para facilitar a análise e mineração dos dados.

O Sistema Gerenciador de Banco de Dados (SGBD) utilizado para a criação do BLDP foi o MySQL (mysql.com), um programa gratuito e muito utilizado na área da bioinformática. O motivo de se utilizar esse sistema foi proporcionar um ambiente seguro e eficiente para o armazenamento e recuperação das informações no banco de dados.

2.1.1 BANCO DE DADOS DE PROTEÍNAS ALVOS

Proteínas pertencentes às vias do Sistema Complemento em humanos e proteínas descritas na literatura como inibidoras e ou reguladoras da ação desse sistema formaram o grupo de proteínas alvos. Estas proteínas foram selecionadas através da literatura e informações: 1. FASTA das informações da proteína, 2. Código de identificação do UniProt “*entry*”, 3. Identificador do gene “*GI*”, 4. Nome do organismo a que pertence a proteína e 5. Nome da própria proteína foram coletadas na base de dados do UniProtKB (uniprot.org).

Estas informações foram utilizadas para alimentar as tabelas do BDLP, criadas de forma relacional para permitir as análises e a mineração dos dados nas demais etapas.

2.1.2 BANCO DE DADOS DE TRIPANOSSOMATÍDEOS

Com o objetivo de encontrar em tripanossomatídeos proteínas homólogas às proteínas alvos, foram utilizadas as bases de dados proteicas UniProtKB (uniprot.org), GenBank (ncbi.nlm.nih.gov/genbank), GeneDB (genedb.org) um banco de dados específico de patógenos e TriTrypDB (tritrypdb.org) uma base de dados especializada em tripanossomatídeos. As proteínas selecionadas destas bases foram armazenadas no BDLP e cada uma recebeu um código identificador com base no UniProtKB.

Detalhes das bases de dados utilizadas estão descritos abaixo.

2.2 BASES DE DADOS DE INFORMAÇÃO DE PROTEÍNAS UTILIZADAS

2.2.1 UNIPROTKB (*Uniprot Knowledgebase*)

UniProtKB/Swiss-Prot (UniProtKB/Swiss-Prot) é o produto principal do grupo *Universal Protein Resource* (UniProt), resultante da colaboração entre o *European Bioinformatics Institute* (EBI/EMBL), o *Swiss Institute of Bioinformatics* (SIB) e *Protein Information Resource* (PIR) (uniprot.org).

O UniProtKB contém registros de sequências proteicas de alta qualidade, derivadas de análises, comparações e fusão de todas as sequências disponíveis para uma determinada proteína, bem como uma revisão crítica experimental dos dados previstos. Estas análises são realizadas por fórmulas computacionais somadas a informações biológicas extraídas da literatura, definindo desta forma o processo de curadoria dos dados do UniProtKB. Este sistema revisa regularmente as famílias e grupo de proteínas para mantê-las integradas com informações científicas atuais (uniprot.org/help/biocuration).

O UniProt fornece em um único registro os detalhes das proteínas originadas pelo mesmo gene de determinado organismo, podendo estas serem proteínas derivadas por *splicing* alternativo, polimorfismo e ou mudanças pós-traducionais (CONSORTIUM, 2014).

Outras bases de dados contidas no UniProt são:

- UniMES (*Uniprot Metagenomic and Environmental Sequences*): Armazena sequências de aminoácidos recuperadas diretamente a partir de amostras ambientais. As proteínas previstas a partir desse método são combinadas por classificação automática considerando família proteica, domínio e função.
- UniParc (*UniProt Archives*): Projetado para capturar os dados das proteínas em vários bancos e montar um único repositório de informações contendo todas as sequências de aminoácidos de forma não redundante.
- UniRef (*UniProt Reference Clusters*): Baseado na mescla de 3 bases provindas do UniProtKB: UniRef50, UniRef90 e UniRef100, que contém sequências com graus de identidade 50%, 90% e 100% respectivamente. Facilita a seleção de proteínas através de sequência representativa.

2.2.2 GENBANK

GenBank é um banco de dados de informações genéticas que contém todas as sequências de DNA (ácido desoxirribonucleico) públicas disponíveis. É parte do *International Nucleotide Sequence Database Collaboration*, que é formado pelo *European Molecular Biology Laboratory* (EMBL), *DNA Data Bank of Japan* (DDBJ) e pelo *National Center for Biotechnology Information* (NCBI), que contém o GenBank. Através da integração dos dados das três instituições, suas informações são mantidas sempre atualizadas. Foi concebido para facilitar o acesso da mais vasta e atualizada fonte de sequências de DNA pela comunidade científica (BENSON *et al.*, 2014).

2.2.3 TRITRYDB (KINETOPLASTID GENOMIC RESOURCE)

TriTrypDB é uma base de dados integrada de acesso a dados de genomas de parasitas cinetoplastídeos, ou seja, parasitas do gênero *Trypanosoma* e *Leishmania* (ASLETT *et al.*, 2010), e que apresenta diversos mecanismos de consultas para o auxílio

na investigação de seus dados. Atualmente o TriTrypDB possibilita análises de genes individuais ou cromossômica em seu contexto genômico. Dados coletados dessa base podem ser armazenados e integrados.

É um projeto colaborativo que utiliza da infraestrutura computacional GUS/WDK, desenvolvida pelo *Eukaryotic Pathogen Bioinformatics Resource Center* (eupathdb.org), para integrar a anotação e análise de genomas no GeneDB e em outros bancos de dados de genômica funcional. Anotações são facilmente acrescentadas auxiliando na curadoria de suas informações.

2.2.4 GENEDB

GeneDB é um banco de dados de genomas de patógenos procariotos e eucariotos. É mantido pelo grupo *Pathogen Genomics* do *Wellcome Trust Sanger Institute* (sanger.ac.uk). Nele são armazenados dados de projetos em andamentos e anotações curadas. Atualmente armazena 41 genomas de organismos distintos e possui um processo curadoria manual realizada por biólogos que mantém os dados atualizados através de análises em literatura científica entre outras fontes (LOGAN-KLUMPLER *et al.*, 2012).

2.3 PROGRAMAS DE BIOINFORMÁTICA

2.3.1 BLASTP

Basic Local Alignment Search Tools (BLAST) é um programa cujo objetivo principal é encontrar regiões de similaridade entre sequências de aminoácidos ou sequência de nucleotídeos. O programa compara as sequências submetidas contra uma base de dados de sequências e depois calcula a estatística de similaridade significativa de seus resultados (ALTSCHUL *et al.*, 1990). O BLASTp pertence a família de programas BLAST e sua especificidade é o processamento de sequências de proteínas, ou seja, dado uma

sequência de aminoácidos no formato FASTA, ele busca entre outras sequências, também de aminoácidos no formato FASTA, o nível de similaridade entre elas e apresenta valores significantes entre o alinhamento das duas sequências.

2.3.2 PSI-BLAST (*POSITION-SPECIFIC ITERATIVE BASIC LOCAL ALIGNMENT SEARCH TOOL*)

PSI-BLAST é uma ferramenta de busca por posição-específica de iteração básica em sítios de alinhamento que utiliza uma *Position-Specific Scoring Matrix* (PSSM), ou seja, uma matriz que se renova com o resultado dos alinhamentos de sequências detectadas a cada iteração, utilizando modelo oculto de Markov e BLASTp. A finalidade do PSSM é permitir a realização de novas consultas ao banco de proteínas, e a cada nova consulta realizada sua matriz é atualizada contendo valores referentes ao novo resultado da busca. As iterações do PSSM com a base de proteínas permite o PSI-BLAST detectar as relações de distâncias entre as proteínas.

Segundo Bergman (2008), a tarefa realizada pelo PSI-BLAST é um método de pesquisa por perfil de sequência de proteínas que se acumula fora dos resultados de BLASTp. Apesar de sua primeira iteração ser uma execução simples de BLASTp, em seguida ele gera um novo alinhamento múltiplo com os resultados do BLASTp que possuem as melhores pontuações, tudo isso definido por atributos de configuração ou por PSSM através de alinhamento múltiplo.

O PSSM armazena a cada iteração uma matriz de pontuação de posições do alinhamento padrão de conservação entre as sequências. Quanto maior o nível de conservação de suas posições, maior é sua pontuação e vice versa. Desta forma o PSSM estabelece um novo padrão de conservação podendo substituir a matriz anterior e permitir uma nova busca para detectar novas de sequências que correspondam ao novo padrão. A cada iteração todo esse processo se repete e a matriz contida no PSSM é atualizada com o resultado de alinhamento de novas sequências. Essa atualização do PSSM se repete até atingir a iteratividade desejada ou até a convergência dos resultados (ALTSCHUL *et al.*, 1997).

O processo de geração de perfil iterativo torna o PSI-BLAST mais eficiente em detectar sequências semelhantes distantes do que uma única consulta pelo BLASTp,

isso por combinar a informação subjacente a conservação de uma série de sequências relacionadas no PSSM (ALTSCHUL *et al.*, 1997).

2.3.3 PFAM

PFAM (pfam.sanger.ac.uk) é um banco de dados utilizado para análise de família de proteínas. Hoje, possui mais de 13.000 famílias curadas armazenadas (PUNTA *et al.*, 2012). As famílias são conjuntos de regiões proteicas que partilham um grau significativo de similaridade entre suas sequências, ou seja, homologia. Seu objetivo é determinar em proteínas qual ou quais as funções que elas exercem no organismo. Essa análise é realizada através da comparação de sítios da cadeia polipeptídica de uma dada proteína com as diversas famílias de proteínas anotadas em sua base de dados. Também como retorno o PFAM indica a posição do domínio na proteína e o *score* (pontuação) do domínio encontrado.

2.3.4 SIGNALP

O SignalP é um programa utilizado para predição da possível existência de peptídeo sinal em proteínas. Incorpora uma combinação de redes neurais artificiais treinadas com conjuntos de informações de eucariotos e procariotos que permite indicar a presença e localização de um sítio de clivagem de um peptídeo sinal e também a predição de sequência de aminoácidos transmembrana (NIELSEN *et al.*, 1997). É uma ferramenta versátil que pode ser utilizada localmente, pela web ou por tecnologia de *web service* e para facilitar a visualização, apresenta um gráfico que indica ponto de clivagem e demais informações.

A versão adotada para as análises desse trabalho foi a versão 4.0 que inclui algoritmos de modelo oculto de Markov (PETERSEN *et al.*, 2011) e está disponível no endereço cbs.dtu.dk/services/SignalP.

2.3.5 TARGETP

O TargetP é um programa que possui como finalidade prever a localização celular de proteínas eucarióticas. Sua estratégia de localização baseia-se na presença prevista de sequências na região N-terminal: peptídeo de transporte para cloroplasto (cTP), peptídeo de direcionamento mitocondrial (mTP), ou peptídeo de sinal de secreção, peptídeo sinal (SP). Utiliza o programa SignalP para esta última análise. É um programa mantido pelo Center for *Biological Sequence Analysis* (CBS) e está disponível *on-line* através do link cbs.dtu.dk/services/TargetP/ (EMANUELSSON *et al.*, 2000)

2.3.6 PRED-GPI

Pred-GPI (*GPI-anchor Predictor* – gpcr.biocomp.unibo.it/predgpi) é um programa que possui características semelhantes as do SignalP porém, sua função é prever a existência de âncora de glicosil fosfatidil inositol (ancora-gpi) na sequência polipeptídica das proteínas submetidas. Para prever a existência de ancora-gpi e a posição do ω -site (local de inicialização da ancora-gpi) o programa faz uso de um sistema de predição com base em *Support Vector Machine* (SVM) acoplado a um modelo oculto de Markov. O SVM é uma técnica de inteligência artificial que segundo os conceitos de rede neural, deve ser treinada. Para isso foi selecionado um conjunto de proteínas não redundantes do UniProt para servir de base de treinamento para o SVM (PIERLEONI *et al.*, 2008).

2.3.7 CLUSTAL OMEGA

Clustal Omega (clustal.org/omega/) é um programa que tem a finalidade de realizar *Multiple Sequence Alignment* (MAS). Produz MAS de qualidades e consegue manipular milhares de sequência em pouco tempo. As sequências podem ser submetidas ao Clustal Omega através de um arquivo de texto e serão alinhadas e agrupadas formando uma árvore de guia que direciona um alinhamento progressivo das sequências. O Clustal

Omega utiliza também modelo oculto de Markov para o mecanismo de alinhamento e realiza apenas alinhamento de proteínas (SIEVERS *et al.*, 2011).

2.4 RECURSOS COMPUTACIONAIS AUXILIARES

2.4.1 MYSQL

MySQL (mysql.com) é um SGBD que utiliza a linguagem SQL (Linguagem de Consulta Estruturada) como interface de comunicação com os dados e por se tratar de um software livre, vários usuários no mundo contribuem implementando melhorias e executando testes para seu aperfeiçoamento (NOGUEIRA *et al.*, 2009).

Possui como características relevantes:

- Portabilidade (pode ser executado em várias plataformas)
- Compatibilidade (possui comunicação com diversas linguagens de programação distintas)
- Ótimo desempenho e estabilidade
- Pouca exigência de recurso de hardware
- Boa usabilidade

2.4.2 PERL

Practical Extraction and Report Language (Perl) é uma linguagem de programação desenvolvida especialmente para a manipulação de textos (perldoc.perl.org). Devido a essa característica, Perl se tornou uma das linguagens mais comuns para se desenvolver *script's* (programas baseados em processamento de cadeias de codificação). O Perl é uma linguagem interpretada, o que permite construir e testar programas simples com facilidade. É estável e possui característica multiplataforma, permitindo ser manipulado em diversos sistemas operacionais. Também é muito utilizado na programação de

formulários WWW (Word Wide Web) (www.perl.org).

2.5 FLUXO DE ANÁLISES REALIZADAS DURANTE A PESQUISA

Foi delineado um fluxo específico de coleta de dados e análise com objetivo de filtrar as proteínas encontradas em cada etapa. Este fluxo está apresentado na figura 4 e será detalhado a seguir.

Fluxo de Dados/Análises

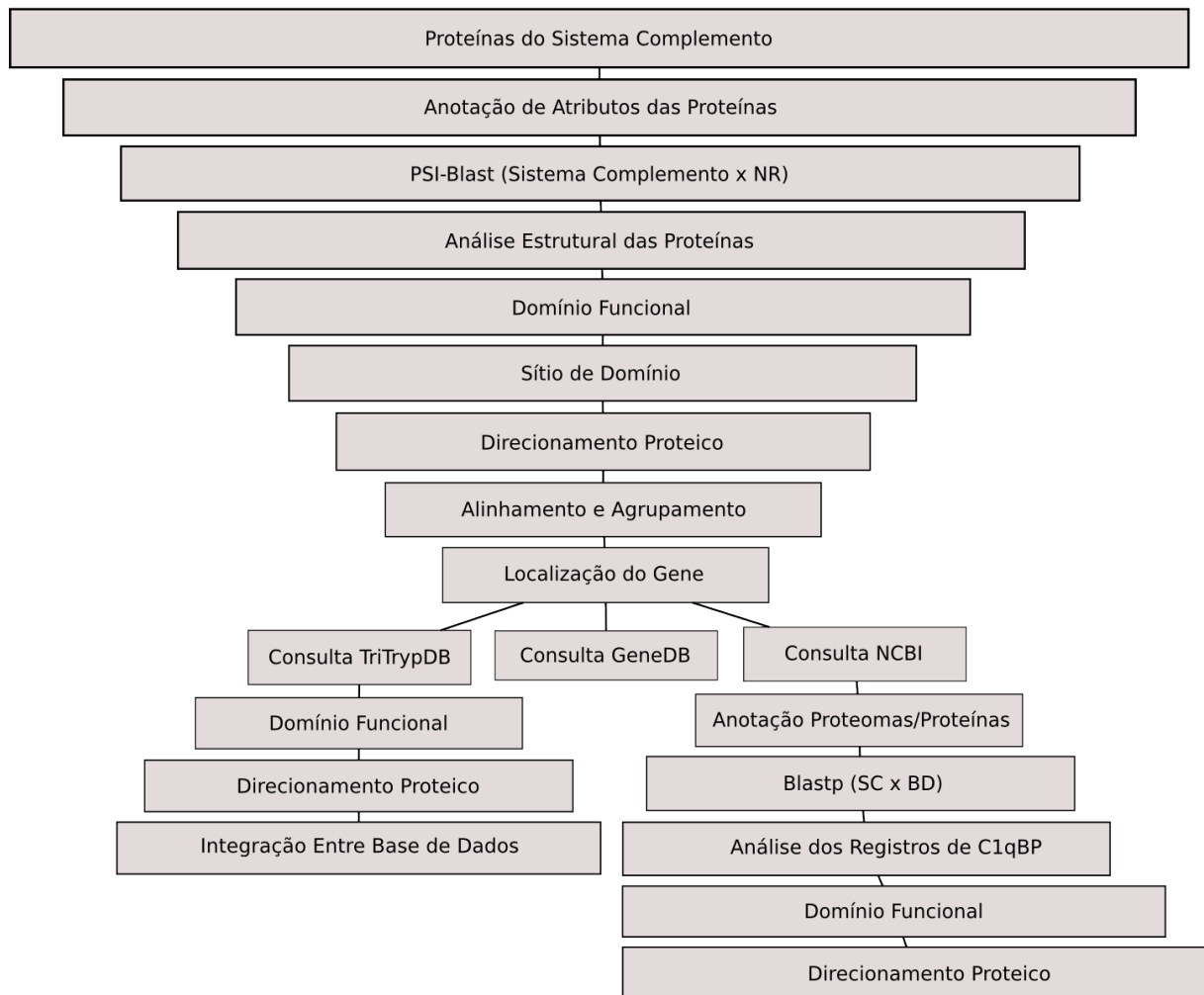


FIGURA 4: Fluxo de obtenção, anotação, processamento, mineração e análise dos dados do projeto.

FONTE: GOMES (2014)

2.5.1 SELEÇÃO DE PROTEÍNAS ALVO

Neste projeto as proteínas alvo são proteínas pertencentes às vias do sistema complemento humano e outras proteínas que regulam a função dessas vias. Com base na literatura foram selecionadas proteínas que passaram a compor a primeira tabela do BDLP. Atributos referentes a estas proteínas FASTA, *entry*, GI, nome do organismo e o nome da própria proteína foram coletados na base de dados do UniProtKB.

2.5.2 BUSCA DE PROTEÍNAS HOMÓLOGAS UTILIZANDO PSI-BLAST

A primeira busca por proteínas de tripanossomatídeos homólogas as proteínas alvo foi por confronto entre as proteínas alvo do BDLP e as proteínas do *Non-redundant protein sequence database* (NR, blast.ncbi.nlm.nih.gov) utilizando PSI-Blast. Foram executadas 20 iterações para cada proteína alvo e para cada retorno, o resultado era filtrado com intuito de averiguar a existência de proteínas que pertencesse a tripanossomatídeos. Foram realizadas anotações para cada etapa das 20 iterações e para cada uma das proteínas alvos. As proteínas selecionadas nesta etapa foram consideradas candidatas a desempenharem função semelhante àquelas realizadas em humanos.

2.5.3 BUSCA DE DOMÍNIOS FUNCIONAIS NAS PROTEÍNAS CANDIDATAS

As proteínas homólogas candidatas encontradas na etapa anterior foram submetidas ao Pfam para se determinar seus domínios funcionais, caso existissem. As proteínas alvo que apresentaram iterações com proteínas candidatas relevantes também foram submetidas ao Pfam para a verificação se também possuíam o mesmo domínio de função. Conseqüentemente, foram anotadas informações referentes à localização do domínio dentro de cada sequência, a pontuação de similaridade da sequência encontrada com a família do domínio definida no Pfam e a sequência que representa o domínio dentro da proteína analisada.

2.5.4 ESTUDO DE DIRECIONAMENTO PROTEICO DAS PROTEÍNAS CANDIDATAS

As proteínas selecionadas conforme descrito no item anterior foram analisadas pelos programas SignalP e TargetP para se determinar a presença ou ausência de peptídeo sinal e/ou sequência transmembrana, ou se a mesma é direcionada para a mitocôndria. Também foram analisadas pelo programa PredGPI para se determinar a presença de âncora-gpi.

Nesta etapa foi analisado o direcionamento proteico de todas as proteínas alvos e foram realizadas as anotações dos resultados dos dois grupos de proteínas (alvos e candidatas).

2.5.5 SIMILARIDADE ENTRE PROTEÍNAS CANDIDATAS

A fim de agrupar as proteínas encontradas que possuíssem similaridade entre si, foi organizado em um arquivo possuindo todas as sequências marcadas para facilitar sua identificação posterior. Estas proteínas foram então submetidas ao Clustal Omega.

O resultado obtido pelo Clustal Omega comparado ao resultado das iterações do PSI-Blast permitiu a formação de grupos e subgrupos que determinaram as aproximações entre elas e também a detecção de duplicidades e repetições de proteínas ou genes.

As proteínas alvos não foram utilizadas para análise nessa etapa e os dados obtidos foram armazenados em novas tabelas que sempre se referenciavam às proteínas relacionadas.

2.5.6 DETERMINAÇÃO DOS GENES QUE CODIFICAM ÀS PROTEÍNAS CANDIDATAS

Os genes relativos que codificam às proteínas homólogas candidatas em tripanossomatídeos filtradas pelas etapas anteriores foram buscados no GenBank e anotadas também no BDLP. A finalidade dessa análise foi descobrir se um mesmo gene é responsável pela transcrição de mais de uma proteína e se são ortólogos ou parálogos.

Até esta etapa, o conjunto de proteínas candidatas encontradas foi refinado e em seguida, visando aumentar o nível de confiabilidade das proteínas selecionadas, o projeto aprofundou-se a análise utilizando outras três bases de dados de proteínas relacionadas com o tema abordado: TryTripDB, GeneDB e NCBI.

Alterações na estratégia de confronto com estas bases foram realizadas, porém a análise das características comparativas das novas proteínas candidatas com as proteínas alvos foram mantidas, como consulta de domínio funcional e direcionamento proteico.

2.6 CONSULTA AO TRITRYPDB

Diferente do confronto realizado utilizando o PSI-Blast contra o banco de dados de proteínas NR, desta vez foi utilizado o BLASTp para buscar proteínas homólogas às proteínas alvos na base de dados do TryTripDB. As proteínas encontradas foram separadas e anotadas no BDLP e tiveram suas características mais relevantes armazenadas.

2.6.1 ANÁLISE DE DOMÍNIO FUNCIONAL NO RESULTADO DE TRITRYPDB

Nesta etapa foi utilizado o Pfam para a análise dos domínios funcionais das proteínas candidatas e das proteínas alvos que possuíam proteínas homólogas a elas. Esta etapa foi semelhante às análises realizadas nas proteínas candidatas encontradas quando o confronto foi realizado com o NR de proteínas utilizando o PSI-Blast.

2.6.2 ESTUDO DO DIRECIONAMENTO PROTEICO

Similar as análises de filtragem realizadas nas proteínas resultantes na comparação com o banco de dados de proteínas NR, as proteínas candidatas que

possuíram o mesmo domínio funcional que as proteínas alvos em TriTrypDB também foram submetidas ao SignalP, TargetP e PredGPI para a predição de seu direcionamento proteico. Estes resultados foram armazenados também no BDLP.

2.6.3 INTEGRAÇÃO DAS INFORMAÇÕES

Como o TriTrypDB e UniProtKB não possuem o mesmo padrão de identificação de seus dados, foi necessário realizar uma equiparação para descobrir qual proteína em um banco representa a mesma em outro banco. Essa análise foi realizada utilizando BLASTp com as proteínas de TriTrypDB em confronto com o base de dados do UniProtKB. Essa integração garantiu o melhor relacionamento entre os dados permitindo o uso de técnicas de mineração de dados.

2.7 CONSULTA AO GENEDB

Como comentado no item 2.2.3, o GeneDB contém todos os organismos do banco TriTrypDB e, desta forma, foram selecionados para a análises no GeneDB apenas os organismos que não pertencem ao grupo de tripanossomatídeos. Estes organismos tiveram seus proteomas confrontados com as proteínas alvos utilizando BLASTp. Apenas as proteínas alvos que possuíram registros no resultado do PSI-Blast foram confrontadas.

As proteínas dos organismos contidos no GeneDB que apresentaram homologia às proteínas alvos foram levadas ao Pfam para a realização de análise de semelhança de domínio funcional. Apenas os organismos que possuem proteínas homólogas às proteínas alvo e o mesmo domínio funcional foram marcados. O objetivo foi identificar outros patógenos que possivelmente pudessem possuir as mesmas características que os tripanossomatídeos candidatos.

2.8 CONSTRUÇÃO DE TABELA LOCAL COM ORGANISMOS ESPECÍFICOS

Através de consulta a literatura, foram selecionados representantes de tripanossomatídeos para a criação de um novo banco com o objetivos de se realizar mais análises. Estas informações foram buscadas no GenBank e passaram pelos processos descritos abaixo.

2.8.1 ANOTAÇÃO DE PROTEOMAS E PROTEÍNAS

Os dados taxonômicos e o proteoma de cada organismo selecionado foi obtido do GenBank. Todos os FASTA's das proteínas dos referidos proteomas foram capturados do GenBank e anotados no BDLP através de *scripts* desenvolvidos em Perl.

2.8.2 COMPARAÇÃO DAS PROTEÍNAS ALVOS COM OS ORGANISMOS DO GENBANK

As proteínas alvos do sistema complemento humano foram comparadas com as proteínas dos organismos selecionados utilizando o programa BLASTp. Os resultados foram armazenados em uma nova tabela do BDLP.

2.8.3 FILTRAGEM POR PROTEÍNAS ALVOS RELEVANTES

Considerando todos os registros resultante do Blastp no item 2.8.2, apenas os registros de proteínas homólogas às proteínas alvos que apareceram no resultado de PSI-Blast (item 2.5.2) avançaram para a nova sequência de análises.

2.8.4 ANÁLISE DO DOMÍNIO FUNCIONAL

As proteínas encontradas em BLASTp que possuem homologia com as mesmas proteínas alvos encontradas em PSI-Blast foram analisadas no Pfam para determinação de domínio funcional. Estes resultados foram anotados no BDLP.

2.8.5 ANÁLISE DE DIRECIONAMENTO PROTEICO

As proteínas selecionadas pela análise descrita no item 2.8.4 foram também submetidas aos programas SignalP, TargetP e Pred-GPI para que seus direcionamentos proteicos e os resultados também foram armazenados no BDLP.

2.9 INTEGRAÇÃO DOS RESULTADOS

Os resultados obtidos a partir dos fluxos de análises distintas para três base de dados (TriTrypDB, UniProt e GenBank) foram analisados através de técnicas de mineração de dados.

2.9.1 COMPARAÇÃO DOS DOMÍNIOS FUNCIONAIS

Foram criadas linhas de programas em SQL para pesquisar no BDLP quais proteínas pertencentes às três bases de dados possuíam os mesmos domínios funcionais.

2.9.2 COMPARAÇÃO DO DIRECIONAMENTO PROTEICO

Foram criadas linhas de programas em SQL para pesquisar no BDLP quais proteínas pertencentes às três bases de dados possuíam o mesmo direcionamento proteico.

As proteínas que passaram por todas as análises desse fluxo de dados são as proteínas de tripanossomatídeos consideradas candidatas exercem algum tipo de regulação do sistema complemento em humanos.

2.9.3 MODELAGEM COMPARATIVA E CALCULO DA DISTRIBUIÇÃO DE CARGA

A modelagem comparativa das proteínas Q2DG13, K2NJJ1 e Hq412588, e a remodelagem C1qBP-humana, foi realizada no servidor SWISS-MODEL (<http://swissmodel.expasy.org>) utilizando a opção de modelagem automática. Neste caso, o servidor procura e alinha as estruturas moldes mais adequadas para construção do modelo comparativo de forma automatizada. As estruturas oligoméricas (homotrímeros) foram geradas no programa PYMOL através da sobreposição de múltiplas cópias dos modelos obtidos com os trímeros das estruturas cristalográficas utilizadas como referência. A atribuição das cargas parciais foi realizada no servidor PDB2PQR versão 1.8 (http://nbc222.ucsd.edu/pdb2pqr_1.8/) utilizando o modelo de cargas padrão do servidor (PARSE). O programa ProPKa (incluso no servidor PDB2PQR) foi utilizado para prever o estado de protonação dos aminoácidos carregados.

3 RESULTADOS

O Sistema Complemento está composto por um conjunto de proteínas séricas que tem a característica de serem ativadas e reguladas em cascatas de reações sequenciais conhecidas como vias de ativação do sistema complemento. Este sistema faz parte do sistema imune inato do organismo e tem as funções de fagocitose, lise celular de patógenos, ativação e recrutamento celulares, depuração de complexos imunes e respostas inflamatórias locais. O nível da resposta desencadeada pelo sistema complemento é controlado por um grupo de proteínas reguladoras que garantem que a ativação seja proporcional à concentração e à duração da presença dos ativadores (ABBAS *et al.*, 2012).

A infecção do organismo humano por tripanossomatídeos desencadeia uma resposta imune que envolve, entre outros mecanismos, a ativação do sistema complemento e que tem o objetivo de controlar ou mesmo eliminar os parasitas (MACHADO *et al.*, 2012). A persistência dos parasitas no organismo e a consequente instalação de patologia resulta de sua capacidade de evasão do sistema imune. Diferentes estudos tem mostrados que, em especial, *T. cruzi* apresenta mecanismos moleculares capazes de permitir evasão para essa via de controle parasitária (Revisado por CESTARI *et al.*, 2013). Um possível mecanismo de evasão poderia envolver a expressão no parasita de uma ou mais proteínas com função similar as proteínas do sistema complemento, que atuariam como bloqueadores competindo com as homólogas presentes no soro normal humano, também a expressão de proteínas reguladoras do complemento (Revisado por CESTARI *et al.*, 2013).

Na sessão anterior foi apresentada a estratégia estabelecida para busca de proteínas homólogas às proteínas do sistema complemento humano em tripanossomatídeos (FIGURA 4). A FIGURA 5 apresenta o resumo dos resultados obtidos.

Fluxo de Dados/Análises

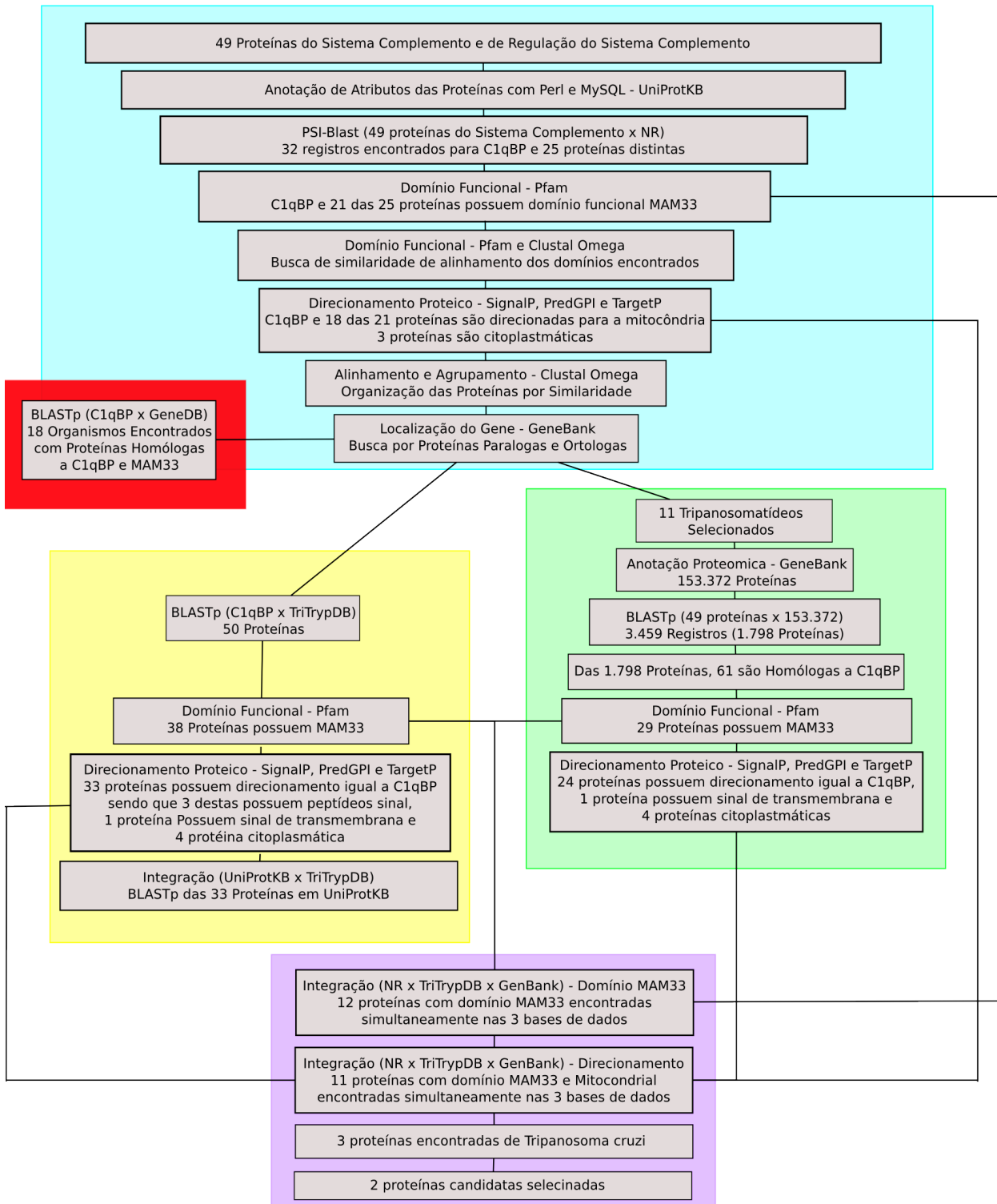


FIGURA 5: Resumo dos resultados obtidos em cada etapa da estratégia proposta para busca de proteínas homólogas às proteínas do sistema complemento humano em tripanossomatídeos. Em azul estão representados os resultados das etapas de análise na base de dados NR, em amarelo o resumo dos resultados obtidos nas análises na base de dados TriTrypDB, em verde os resultados obtidos com análises na base de dados GenBank, em vermelho a etapa por buscas de proteínas homólogas a C1qBP e com domínio MAM33 na base de dados do GenBank e em lilás os resultados obtidos da integração dos resultados das bases de dados NR, TriTrypDB e GenBank.

FONTE: GOMES (2014)

3.1 PROTEÍNAS ALVOS

Proteínas pertencentes às vias de ativação do sistema complemento ou proteínas que estejam relacionadas à regulação dessas vias foram selecionadas na literatura (YAMAMOTO; PORTINHO, 2001; ABBAS *et al*, 2012), e passaram a compor o conjunto de proteínas alvos. Foram escolhidas 49 proteínas (FIGURA 6), sendo 32 proteínas componentes das vias do sistema complemento e 17 vinculadas à sua regulação.

PROTEÍNA	FUNÇÃO
C1q subunit A	Ligante a porção FC da imunoglobina
C1q subunit B	Ligante a porção FC da imunoglobina
C1q subunit C	Ligante a porção FC da imunoglobina
C1r like	Protease sérica proteína análoga a C1r
C1r	Protease sérica, cliva C1s
C1s	Protease sérica, cliva C4 e C2
C2	Protease sérica, faz parte da C3 e C5 convertase
C3	Formada por C3a e C3b
C3a	É uma anafilaxina
C4a	É uma anafilaxina
C4b	Forma ligações com as superfícies ativadoras, faz parte de C3 convertase
C4b prot. alpha	Inibe a formação da C3 convertase acelerando sua degradação
C5	Formada por C5a e C5b
C5a receptor L	Receptor de quimiotaxia e inafilatoxina inflamatória C3a, C4a e C5a
C5a receptor	Receptor de quimiotaxia e inafilatoxina inflamatória C5a
C6	Componente do complexo lítico da membrana
C7	Componente do complexo lítico da membrana
C8 alpha	Componente do complexo lítico da membrana
C8 beta	Componente do complexo lítico da membrana
C8 gama	Componente do complexo lítico da membrana
C9	Componente do complexo lítico da membrana, sofre polimerização
C9 (Fragment)	Fragmento de C9
CD9 antigen	Regula a formação de junção paranodal
Integrin alpha-M	Regula o fragmento C3b
Integrin alpha-X	Realiza a mediação da interação entre as células durante resposta inflamatória
Integrin beta-2	Receptor para ICAM1, ICAM2, ICAM3 e ICAM4
CD59 glycoprotein	Inibe complexo de ataque a membrana, se liga a C8 e C9
C. receptor type 1	Acelera a dissociação de C3 e C5 convertase, atua na clivagem de C3b e C4b
C. receptor type 2	Regula a função de linfócitos
CD97 antigen	Desempenha um papel essencial na migração de leucócitos
C. factor H	Inativa C3b para a formação de C3 e C5 convertase
C1q sub binding prot.	Promove a quimiotaxia. Intermedia a ligação de imunocomplexos à células fagocitárias
C1q like prot. 2	Regular o número de sinapses excitatórias
C1q like prot. 3	Regular o número de sinapses excitatórias
C1q like prot. 4	Regular o número de sinapses excitatórias
C1q receptor	Receptor para C1q
C1q related factor	Regula o número de sinapses excitatórias
C1q 9A	Ativa AMPK, AKT e MAPK p44/42 vias de sinalização
C1q 9B	Interage com CTRP9A e ADIPOQ
C1 inhibitor	Complexo molecular covalente com C1r e C1s, inibindo a atividade deste
Mannan lect. serine 1	Funciona como uma endopeptidase e pode ativar MASP2 ou C2 ou C3
C4b prot. beta	Inibe a ligação de C4b com C2a
Integrin alpha L	Receptor para ICAM1, ICAM2, ICAM3 e ICAM4
decay accelerating factor	Regula a formação de C3 e C5 convertase
factor H protein 4	Acelera a dissociação de C3 convertase de via alternativa
factor H protein 5	Acelera a dissociação de C3 convertase de via alternativa
factor H protein 3	Acelera a dissociação de C3 convertase de via alternativa
factor H protein 1	Acelera a dissociação de C3 convertase de via alternativa
factor H protein 2	Acelera a dissociação de C3 convertase de via alternativa

FIGURA 6: Identificação das 49 proteínas alvos selecionadas e resumo de suas funções
 FONTE: GOMES (2014)

3.2 DADOS DAS PROTEÍNAS ALVOS

Foi criada no Banco de Dados Local de Proteínas (BDLP) uma tabela para o armazenamento dos seguintes dados relativos às proteínas alvos: 1. Pacote FASTA de informações da proteína, 2. Código de identificação do UniProt “*entry*”, 3. Identificador do gene “GI”, 4. Nome do organismo a que pertence a proteína e 5. Nome da própria proteína. Estes dados foram obtidos do UniProtKB com o auxílio de programas “*scripts*” escritos em Perl.

3.3 BUSCA POR PROTEÍNAS HOMÓLOGAS ÀS PROTEÍNAS ALVOS NA BASE DE DADOS NR

Para cada uma das 49 proteínas alvos, foram executadas 20 iterações contra o Banco de Dados de sequências não Redundantes (*Non-redundant protein sequence database* (NR, blast.ncbi.nlm.nih.gov) utilizando o programa PSI-Blast. O padrão de retorno para cada iteração no PSI-Blast foi de 500 proteínas ordenadas por maior pontuação de similaridade (ALSCHUL *et al.*, 1997), porém apenas as proteínas pertencentes ao grupo dos tripanossomatídeos foram contabilizadas e anotadas. A FIGURA 7 apresenta o resultado obtido nestas iterações e os registros das ocorrências de proteínas homólogas.

Observamos que dentre todas as proteínas apresentadas nos resultados do PSI-Blast apenas a proteína C1QBP_HUMAN (proteína de ligação para a subunidade globular do complemento C1q humano) (LIM BL, REID KBM, GHEBREHIWET B, 1996) apresentou homologia com proteínas de tripanossomatídeos. Estas proteínas apareceram apenas nas quatro primeiras iterações que são os principais níveis para se garantir um alto nível de identidade (ALTSCHUL *et al.*, 1997). Ao todo foram registradas 32 proteínas dentre as quais, apenas 25 proteínas são proteínas distintas (FIGURA 8).

NOME CURTO	PROTEINA	ENTRY	ITERAÇÕES																			
			1ª	2ª	3ª	4ª	5ª	6ª	7ª	8ª	9ª	10ª	11ª	12ª	13ª	14ª	15ª	16ª	17ª	18ª	19ª	20ª
C1QA_HUMAN	C1q subunit A	P02745	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1QB_HUMAN	C1q subunit B	P02746	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1QC_HUMAN	C1q subunit C	P02747	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1RL_HUMAN	C1r like	Q9NZP8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1R_HUMAN	C1r	P00736	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1S_HUMAN	C1s	P09871	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO2_HUMAN	C2	P06681	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO3_HUMAN	C3	P01024	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C3AR_HUMAN	C3a	Q16581	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO4A_HUMAN	C4a	POCOL4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO4B_HUMAN	C4b	POCOL5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C4BPA_HUMAN	C4b prot. alpha	P04003	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO5_HUMAN	C5	P01031	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C5ARL_HUMAN	C5a receptor L	Q9P296	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C5AR_HUMAN	C5a receptor	P21730	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO6_HUMAN	C6	P13671	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO7_HUMAN	C7	P10643	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO8A_HUMAN	C8 alpha	P07357	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO8B_HUMAN	C8 beta	P07358	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO8G_HUMAN	C8 gama	P07360	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CO9_HUMAN	C9	P02748	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Q9UGI4_HUMAN	C9 (Fragment)	Q9UGI4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CD9_HUMAN	CD9 antigen	P21926	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ITAM_HUMAN	Integrin alpha-M	P11215	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ITAX_HUMAN	Integrin alpha-X	P20702	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ITB2_HUMAN	Integrin beta-2	P05107	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CD59_HUMAN	CD59 glycoprotein	P13987	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CR1_HUMAN	C. receptor type 1	P17927	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CR2_HUMAN	C. receptor type 2	P20023	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CD97_HUMAN	CD97 antigen	P48960	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
CFAH_HUMAN	C. factor H	P08603	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1QB_HUMAN	C1q sub binding prot.	Q07021	16	9	6	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1QL2_HUMAN	C1q like prot. 2	Q7Z5L3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1QL3_HUMAN	C1q like prot. 3	Q5VWW1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1QL4_HUMAN	C1q like prot. 4	Q86223	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1QR1_HUMAN	C1q receptor	Q9NPY3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1QRF_HUMAN	C1q related factor	O75973	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1T9A_HUMAN	C1q 9A	POC862	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C1T9B_HUMAN	C1q 9B	B2RNN3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
IC1_HUMAN	C1 inhibitor	P05155	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MASP1_HUMAN	Mannan lect. serine 1	P48740	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C4BPB_HUMAN	C4b prot. beta	P20851	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ITAL_HUMAN	Integrin alpha L	P20701	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
DAF_HUMAN	decay accelerating factor	P08174	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FHR4_HUMAN	factor H protein 4	Q92496	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FHR5_HUMAN	factor H protein 5	Q9BXR6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FHR3_HUMAN	factor H protein 3	Q02985	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FHR1_HUMAN	factor H protein 1	Q03591	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
FHR2_HUMAN	factor H protein 2	P36980	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

FIGURA 7: Resultado do confronto das 49 proteínas contra o *Non-redundant protein sequence database* (NR, blast.ncbi.nlm.nih.gov) utilizando o programa PSI-Blast.
 FONTE: GOMES (2014)

Proteínas Encontradas no NR Utilizando PSI-BLAST							
Entry	Source	Organism	Length	Score	Identites	E-Value	Iteração
Q57WG2	P22 protein, putative	Trypanosoma brucei brucei	260	98	40%	0.26	1ª
C9ZSQ9	p22 protein, putative	Trypanosoma brucei gambiense	260	98	40%	0.26	1ª
A4HCI6	Uncharacterized protein	Leishmania braziliensis	248	95	39%	0.59	1ª
Q4QBJ9	Uncharacterized protein	Leishmania major	249	93	39%	1	1ª
E9BFZ8	Uncharacterized protein	Leishmania donovani	255	92	39%	1.3	1ª
E9AVY0	Putative uncharacterized protein	Leishmania mexicana	249	92	39%	1.3	1ª
A4I007	Uncharacterized protein	Leishmania infantum	255	92	39%	1.3	1ª
F9W3J5	WGS project CAEQ00000000 data, annotated contig 1060	Trypanosoma congolense	260	91	34%	1.7	1ª
Q583P6	Kinesin, putative	Trypanosoma brucei brucei	1085	89	21%	2.9	1ª
K2NJJ1	Uncharacterized protein	Trypanosoma cruzi marinkellei	261	89	36%	2.9	1ª
C9ZQV8	Kinesin, putative	Trypanosoma brucei gambiense	1085	89	21%	2.9	1ª
Q4DQ42	Uncharacterized protein	Trypanosoma cruzi	261	87	34%	5	1ª
Q4DFL0	Uncharacterized protein	Trypanosoma cruzi	261	87	34%	5	1ª
K4EA71	Uncharacterized protein	Trypanosoma cruzi	261	87	34%	5	1ª
Q58NS4	Variant surface glycoprotein MITat 1.13	Trypanosoma brucei	499	86	35%	6.5	1ª
B3GVT4	Variant surface glycoprotein	Trypanosoma brucei brucei	499	86	35%	6.5	1ª
Q4E0T1	p22 protein, putative	Trypanosoma cruzi	225	198	21%	7,0E-13	2ª
Q4DG13	p22 protein, putative	Trypanosoma cruzi	225	198	21%	7,0E-13	2ª
A4HI52	Putative p22 protein	Leishmania braziliensis	191	196	19%	1,0E-12	2ª
E9BLJ6	p22 protein, putative	Leishmania donovani	204	195	17%	2,0E-12	2ª,3ª
A4I5C8	Putative p22 protein	Leishmania infantum	204	195	17%	2,0E-12	2ª,3ª
Q4Q7K6	Putative p22 protein	Leishmania major	195	193	18%	3,0E-12	2ª,3ª
E9B0M6	Putative p22 protein	Leishmania mexicana	233	208	17%	5,0E-14	2ª,3ª
G0TWR1	Putative p22 protein	Trypanosoma vivax	226	199	18%	5,0E-13	2ª,3ª
G0UNJ8	Putative p22 protein	Trypanosoma congolense	226	214	19%	1,0E-14	2ª,3ª,4ª

FIGURA 8: Proteínas de Tripanossomatídeos homólogas a proteína C1QBP_HUMAN encontradas na base de dados NR. A pesquisa foi realizada utilizando o programa PSI-Blast e as 49 proteínas alvos humanas listadas na FIGURA 6.

FONTE: GOMES (2014)

Na FIGURA 8 estão apresentadas as seguintes informações das proteínas de tripanossomatídeos homólogas a proteína C1QBP_HUMAN: 1. *Entry* - identificador da proteína no UniProt, 2. *Source* - A especificação da proteína no UniProt, 3. *Organism* - O organismo que ela pertence, 4. *Length* - Seu tamanho em número de aminoácidos, 5. *Score* - Pontuação referente o alinhamento da proteína quando comparada com C1qBP (Este valor é calculado a partir das matrizes de pontuação e das penalidades de *gap* (espaço entre as sequências de aminoácidos) de penalidade. Quanto mais alto este valor, maior é a similaridade), 6. *Identites* - O percentual de identidade quando comparada com C1qBP, 7. *E-Value* – É um número de parâmetro estatístico que representa a significância do alinhamento, quanto menor o valor de *E-Value*, menor é a probabilidade de o alinhamento ter ocorrido por acaso, ou seja, quanto menor seu valor, menos provável que exista uma similaridade aleatória e 8. Iterações em que as proteínas apareceram.

3.3.1 ANÁLISE DO DOMÍNIO FUNCIONAL DAS PROTEÍNAS SELECIONADAS

As 25 proteínas homólogas de tripanossomatídeos foram submetidas ao Pfam com o objetivo de determinarmos seus domínios funcionais. Foi observado que 21 das 25 proteínas de tripanossomatídeos possuem o domínio MAM33, um domínio importante no processo de ligação da C1QBP_HUMAN na subunidade globular de C1Q (LIM BL, REID KBM, GHEBREHIWET B, 1996) (FIGURA 9). As proteínas Q583P6 e C9ZQV8 possuem o domínio Cinesina (*Kinesin*), típico de proteínas motoras envolvidas no transporte de organelas em células de eucariotos (HIROKAWA *et al.*, 2009), e as proteínas Q58NS4 e B3GVT4 possuem o domínio C-terminal de glicoproteínas variáveis de superfície (*Trypanosome variant surface glycoprotein C-terminal domain – Trypan Glycop C*) (RAUSCH *et al.*, 1994). Glicoproteínas variáveis de superfície são expressas por tripanossomatídeos como um recurso para escapar da resposta imune do hospedeiro (RAUSCH *et al.*, 1994). As proteínas com domínio MAM33 foram selecionadas para a continuidade das análises por possuírem o mesmo domínio que a proteína C1QBP_HUMAN, que poderia ser um indicativo de semelhança funcional.

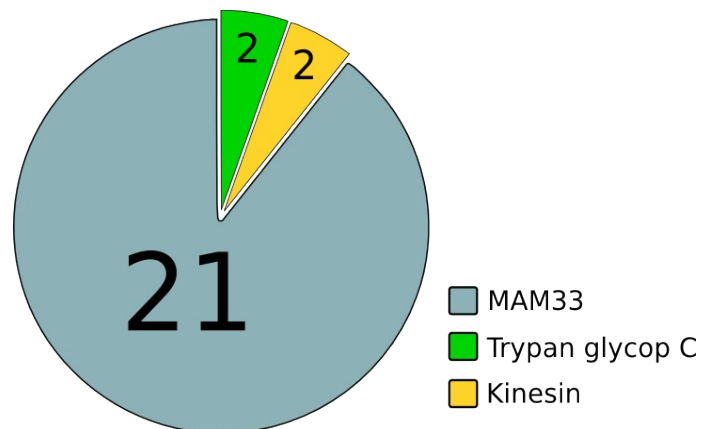


FIGURA 9: Identificação do domínio funcional das 25 proteínas de tripanossomatídeos selecionadas na base de dados NR.
FONTE: GOMES (2014)

3.3.2 COMPARAÇÃO DO DOMÍNIO MAM33 DAS PROTEÍNAS DE TRIPANOSSOMATÍDEOS COM O DOMÍNIO MAM DE C1QBP_HUMAN

Informações específicas para cada uma das 21 proteínas com domínio MAM33 foram obtidas através do Pfam (FIGURA 10) onde todas as 21 proteínas apresentaram o domínio MAM33 (FIGURA 10).

Análise de Comparação Entre os Domínios										
Entry	Length	Domínio Pfam	Envelope		Envelope length	Alignment		Alignment length	Bit score	E-value
			Start	End		Start	End			
Q07021	282	MAM33	86	279	194	86	279	194	259.8	1.6e-77
K2NJJ1	261	MAM33	92	260	169	201	259	59	49.1	5.3e-13
Q4DQ42	261	MAM33	51	260	210	189	259	71	46.6	3.2e-12
Q4DFL0	261	MAM33	51	260	210	186	259	74	47.1	2.2e-12
K4EA71	261	MAM33	51	260	210	189	259	71	46.6	3.2e-12
E9BFZ8	255	MAM33	102	254	153	194	253	60	53.6	2.3e-14
A4I007	255	MAM33	102	254	153	194	253	60	53.6	2.3e-14
A4HC16	248	MAM33	56	247	192	189	246	58	50.7	1.8e-13
Q4QBJ9	249	MAM33	57	248	192	188	247	60	54.9	9.2e-15
E9AVY0	249	MAM33	73	248	176	187	247	61	53.9	1.9e-14
F9W3J5	260	MAM33	50	259	210	57	258	202	45.7	6.2e-12
Q57WG2	260	MAM33	50	259	210	98	258	161	52.7	4.5e-14
C9ZSQ9	260	MAM33	51	259	209	110	258	149	52.3	5.9e-14
Q4E0T1	225	MAM33	104	224	121	117	224	108	59.9	2.7e-16
Q4DG13	225	MAM33	104	224	121	117	224	108	59.7	3.2e-16
G0TWR1	226	MAM33	110	225	116	139	225	87	43	3.9e-11
G0UNJ8	226	MAM33	103	225	123	113	225	113	61.7	7.8e-17
E9BLJ6	204	MAM33	72	203	132	86	202	117	62.6	3.9e-17
A4I5C8	204	MAM33	72	203	132	86	202	117	62.6	3.9e-17
E9BOM6	233	MAM33	94	232	139	103	321	219	64.7	9.4e-18
A4HI52	191	MAM33	69	190	122	79	189	111	65.5	5.2e-18
Q4Q7K6	195	MAM33	57	194	138	74	193	120	64.2	1.3e-17

FIGURA 10: Análise comparativa do domínio MAM33 entre a C1QBP_HUMAN (Q07021) e proteínas homólogas em tripanossomatídeos (a identidade das proteínas e os respectivos organismos estão indicados na FIGURA 8). Coluna 1. Código do UniProtKB (*entry*) referente a proteína, 2. (*Length*) Número de resíduos de aminoácidos, 3. (Domínio Pfam). Tipo de domínio da proteína, 4. (*Envelope Start/End*) Posicionamento dos aminoácidos referente ao início e fim do envelope, 5. (*Envelope length*) Número de resíduos de aminoácidos entre os dois extremos do envelope, 6. (*Alignment Start/End*) Posicionamento dos aminoácidos referente ao início e fim do alinhamento, 7. (*Alignment length*) Número de resíduos de aminoácidos entre os dois extremos do alinhamento, 8. (*Bit score*) pontuação referente ao percentual de similaridade entre domínio encontrado na proteína e o domínio MAM33 e 9. (*E-value*) medida que avalia aproximação de alinhamento entre domínio MAM33

O número de aminoácidos das proteínas selecionadas em relação a C1QBP_HUMAN (Q07021) variou em 21 a 87 resíduos e esta variação se refletiu no tamanho e na posição do domínio na proteína. O *Bit score* referente ao domínio MAM33 presente nas proteínas de tripanossomatídeos variou de 43 a 66,5. Este último valor foi 4 vezes menor que o valor obtido para o domínio MAM33 da proteína Q07021

(C1QBP_HUMAN). O *bit score* indica o quanto a sequência investigada alinha com o modelo da base de dados (PERTEA *et al.*, 2007).

3.3.3 INVESTIGAÇÃO DE DIRECIONAMENTO PROTEICO DAS PROTEÍNAS HOMÓLOGAS À C1QBP_HUMAN E COM DOMÍNIO MAM33

Sabendo-se que a proteína C1QBP-HUMAN apresenta sequência de aminoácidos que a direciona para a mitocôndria (DEDIO *et al.*, 1998), as 21 proteínas homólogas que contêm o domínio MAM33 foram submetidas aos programas SignalP, Pred-GPI e TargetP com a finalidade de se detectar eventuais sequências de aminoácidos indicativas do seu direcionamento proteico para mitocôndria, para membrana plasmática ou, ainda, a possibilidade de serem secretadas. Obtivemos resultado somente com TargetP que retornou 18 proteínas apresentando sequência de aminoácidos que direcionam para a mitocôndria. As outras 3 proteínas não apresentam sinal de endereçamento (FIGURA 11).

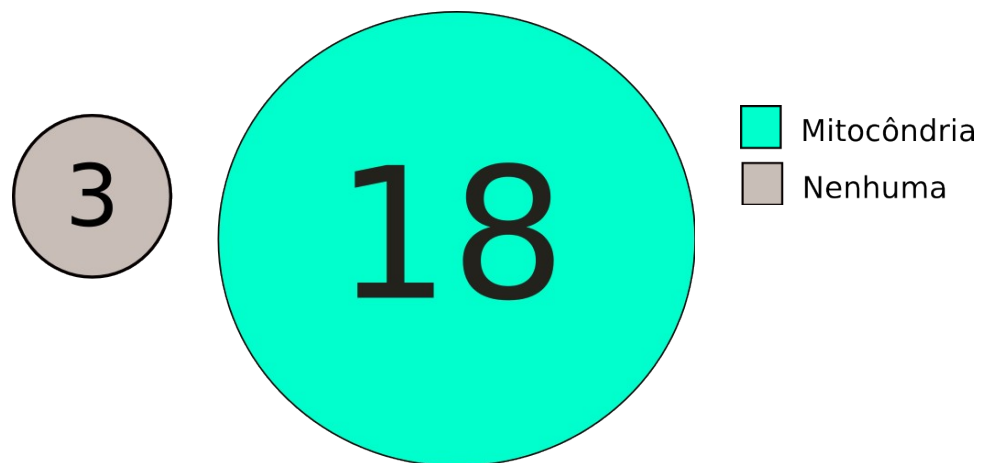


FIGURA 11: Resultado das análises de predição de localização celular das 21 proteínas homólogas a C1QBP_HUMAN.
FONTE: GOMES (2014)

3.3.4 ANÁLISE DE SEMELHANÇA ENTRE AS PROTEÍNAS DE TRIPANOSOMATÍDEOS

As sequências de aminoácidos das 18 proteínas de tripanossomatídeos que possuem o domínio MAM33 e sequência de endereçamento para mitocôndria foram alinhadas com a finalidade de determinarmos o grau de similaridade entre elas como também apresentarmos as regiões dos domínios MAM33 e seus sítios de direcionamento para a mitocôndria (FIGURA 12). Identificamos grupos de proteínas que possuem padrões próximos de alinhamento e integramos o resultado aos níveis de iteração das proteínas à proteína C1QBP_HUMAN (Q07021) obtidos no programa PSI-Blast, que mostra que algumas proteínas aparecem em mais de uma iteração (FIGURA 8). Desta maneira foi possível dividir o conjunto em grupos e subgrupos de proteínas por níveis de similaridades (FIGURA 13). Foram definidos dois níveis de similaridades representados por grupos e subgrupos de proteínas (FIGURA 13). O primeiro nível foi criado a partir de análises realizadas sobre as 18 proteínas e o segundo nível criado por análises realizadas sobre as proteínas de cada grupo do primeiro nível. Foram obtidos 3 grupos e 11 subgrupos com as seguintes informações: 1. (*Entry*) Código da proteína na base de dados do UniProtKB, 2. (*Length*) Quantidade de aminoácidos que compõe a proteína, 3. (Iteração) Iteração do programa PSI-Blast em que a proteína apareceu no resultado, 4. (Sub-grupo) Sub-grupo a que pertence a proteína e 5. (Grupo) Grupo a que pertence a proteína (FIGURA 13).

O Pfam foi utilizado para a identificação do domínio de cada proteína e o programa MitoProt II foi utilizado para a análise de quais aminoácidos pertencem ao direcionamento das proteínas para a mitocôndria. Foi observado nos resultados do programa MitoProt II que as proteínas A4I007, E9BFZ8 e Q4QBJ9 não apresentam direcionamento para a mitocôndria o que é divergente dos resultados obtidos pelo programa TargetP. Isso ocorre possivelmente devido a diferença entre os algoritmos dos dois programas e o banco de dados de treinamento de sua rede de inteligência artificial.

Os alinhamentos referentes aos grupos A, B e C, respectivamente, estão apresentados nas figuras 14, 15 e 16. Nestes alinhamentos é possível perceber mais claramente o elevado número de aminoácidos idênticos que indica elevado grau de conservação.

A4HCI6 -----MSGFRVSFGETLRNAAAGKTDLPARSEPRRH
E9AVY0 -----M-----MSGARAGLGALYRNLIIGSTAELAARSEHRRH
A4I007 MLRRYT-----I-----TLGVRAGLGTSYRNAIASTTELPARSEPRRH
E9BFZ8 MLRRYT-----I-----TLGVRAGLGTSYRNAIASTTELPARSEPRRH
Q4QBJ9 -----M-----MPGARAVLGASYRNAIASTKELPVRSEPRRH
K2NJJ1 MYRLF SRCKTNK-----LRGVFSHALDASIKNEGVAEQRKRHILPRRH
K4EA71 MYRLF SRCETNK-----LKRVSFHALDASIKKEGVAQQRKRQVLP RRH
Q4DQ42 MYRLF SRCKTNK-----LKRVSFHALDASIKNEGVAQQRKRQVLP RRH
C9ZSQ9 MRRVVASRKAIC-----EYRLYSQVSPQGYPVTAATPQQIRRVPFYRH
Q57WG2 MRRVVASRKAIC-----EYRLYSQVSPQGYPVTAATPQQIRRVPFYRH
F9W3J5 MRRMVSHCRPRF-----LTRLCSQVSESKIYSNGDSPKISTRVLPFRH
A4I5C8 -----MSGFD-----RVPSTSAPTALTPRRSASD
E9BLJ6 -----MSGFD-----RVPSTSAPTALTPRRSASD
E9B0M6 MRRTFNKLGLSALRACRGAGAAGAAASAHVSGFA-----RVPSTSAPTALTPRRSASD
G0TWR1 MRRALLCVSRGV-----H-ATPMMAGRF---AALASMVGRPEVSWTHASVLQQRFASC
G0UNJ8 MRRIVLATFGA-----RWAAP-----ISGLSRQPEMGALPGPMALRPVFSQMRLVSN
Q4DG13 MRRVLFSAALSL-----VPKAPLFRRCVPSVAKAASMGP-----LVAMEQRRLASH
Q4E0T1 MRRVLFSAAISF-----VPKAPLFRRCVPSVATAVSMGP-----LVAMEQRRLASH

A4HCI6 RKLYQLTMREEEEGIRDFLPPRPLLPVGVWMEHASGNSRFDL FKNVEIRQCGSEELHII
E9AVY0 RKLYQLTMREEEEGIRDFLPPRPLLPVGVWMEHASGNSRFDL FKNVEIRHCGAEELHVI
A4I007 RKLYQLTMREEYEEGIRDFLPPRPLLPVGVWMEHASGNSRFDL FKNVEVRHCGAEELHAI
E9BFZ8 RKLYQLTMREEYEEGIRDFLPPRPLLPVGVWMEHASGNSRFDL FKNVEVRHCGAEELHAI
Q4QBJ9 RKLYQLTMREEYEEGLRDFLPPQPLLPVGVWMEHASGNSRFDL FKNVEIRHCGSEELHVI
K2NJJ1 RKLHMLTIRENEEGLRDFLPPKPVVPAGWKLEHIKGSNRFDLRKSVEIRDCGKEDLHAV
K4EA71 RKLHMLTIRENEEGLRDFLPPKPVVPAGWKLEHVKGSNRFDLRKSVEIRDCGKEHLHAV
Q4DQ42 RKLHMLTIRENEEGLRDFLPPKPVVPTGWKLEHAKGSNRFDLRKSVEIRDCGKEHLHAV
C9ZSQ9 RKLYDLTTRELNEEAIRDFLPPKPAVPAGWSMEHKIGSCRFDLTKTMGADDSAREDLHVV
Q57WG2 RKLYDLTTRELSEEAIRDFLPPKPAVPAGWSMEHKIGSCRFDLTKTMGADDSAREDLHVV
F9W3J5 RKLYSITLRELHEESLRDFLPPKPAIPAGWTIEHQVGSRRFDLRKVHKADGCTEEDLHVV
A4I5C8 AALADATRRELEEEEMGRSDKPEQPTPPAGWQVVRKPGTCTFDLTKSFEGEEL-----V--
E9BLJ6 AALADATRRELEEEEMGRSDKPEQPTPPAGWQVVRKPGTCTFDLTKSFEGEEL-----V--
E9B0M6 AALADATRRELEEEEMGRSDKPEQPTPTGWQVVRKSGTCTFDLTKSFEDEEL-----V--
G0TWR1 AALANYIRTEIEDEKERAEPKVPKEIPSGWTVHEHEPGNMFFKLCKTYEDEEI-----V--
G0UNJ8 SKLAAATLSEMEDEQQRSEKPAKPELPAGWTLERKPGQMLFTMRKKHEDEEI-----I--
Q4DG13 AALSALRHELEEEQQRSEKPAKPELPAGWTLERKPGQMLFTMRKKHEDEEI-----I--
Q4E0T1 AALSALRHELEEEQQRSEKPAKPELPAGWTLERKPGQMLFTMRKKHEDEEI-----V--

* * : * * * : * * . * : : * * : :

A4HCI6 TLMETKEYEGTYRMDNGEREEQEYLNFGLFMRKKRYPTGGLEFSLT SIDLELVMDGLTIH
E9AVY0 TLMETKQYEGTYRMDNGEREEQEYLNFGLFMRKKRYPAGGLEFSLT SIDLELVMDALTIH
A4I007 TLMETKQYEGTYRMDNGEREEQEYLNFGLFMRKKRYPAGGLEFSLT SIDLELVMDALTIH
E9BFZ8 TLMETKQYEGTYRMDNGEREEQEYLNFGLFMRKKRYPAGGLEFSLT SIDLELVMDALTIH
Q4QBJ9 TLMETKQYEGTYRMDNGEREEQEYLNFGLFMRKKRYPAGGLEFSLT SIDLELVMDALTIH
K2NJJ1 AMMEVKEYEGTYRMDNGEREEYLFFTLFVQKNRFP-GALEFGLT SIDMELVMDSLAIH
K4EA71 AMMEVKEYEGTYRMDNGEREEYLFFTLFVQKNRFP-GALEFGLT SIDMELVMDSLAIH
Q4DQ42 AMMEVKEYEGTYRMDNGEREEYLFFTLFVQKNRFP-GALEFGLT SIDMELVMDSLAIH
C9ZSQ9 ALMEPKKYEQTYRMDNGERNGEYLVFNLFIKKHQHS-GGVEFGLT SIDMELVMDSLVHH
Q57WG2 ALMEPKKYEQTYRMDNGERNGEYLVFNLFIKKHQHS-GGVEFGLT SIDMELVMDSLVHH
F9W3J5 ALMESKKYEETTYRMDNGEREEYLFFTLFIKKSNCN-GGLEFGLT SIDMELVMDTLVHH
A4I5C8 -----VRYSTNQDS-----DKANSHDIFVYITQT--NGQTMQADLSIEEGELV LNNIRFY
E9BLJ6 -----VRYSTNQDS-----DKANSHDIFVYITQT--NGQTMQADLSIEEGELV LNNIRFY
E9B0M6 -----VRYSTNQDS-----DKANSHDIFVYITQK--NGQTMQADLSIEEGELV LNNIRFY
G0TWR1 -----LHFKGTREA-----EGDIFYDFKVFVNG---EKGVMFDLSYDH-EIVVDRVIFL



FIGURA 12: Alinhamento das 18 proteínas homólogas a C1QBP_HUMAN e com domínio MAM33, utilizando o programa Clustal Omega. A identidade das proteínas e os respectivos organismos estão indicados na FIGURA 8. (*) Aminoácidos idênticos; (:) Aminoácidos de alta similaridade; (.) Aminoácidos de baixa similaridade. A sequência de aminoácidos sombreada de azul representam o sítio de direcionamento para a mitocôndria e as sequências sombreadas de laranja representam o domínio.

FONTE: GOMES (2014)

Entry	Length	Iteração	Sub-Grupo	Grupo
K2NJJ1	261	1 ^a	A1	A
Q4DQ42	261	1 ^a	A1	
K4EA71	261	1 ^a	A1	
E9BFZ8	255	1 ^a	A2	
A4I007	255	1 ^a	A2	
A4HCI6	248	1 ^a	A3	
Q4QBJ9	249	1 ^a	A3	
E9AVY0	249	1 ^a	A3	
F9W3J5	260	1 ^a	A4	
Q57WG2	260	1 ^a	A4	
C9ZSQ9	260	1 ^a	A4	
Q4E0T1	225	2 ^a	B1	B
Q4DG13	225	2 ^a	B1	
G0TWR1	226	2 ^a ,3 ^a	B2	
G0UNJ8	226	2 ^a ,3 ^a ,4 ^a	B3	C
E9BLJ6	204	2 ^a ,3 ^a	C1	
A4I5C8	204	2 ^a ,3 ^a	C1	
E9B0M6	233	2 ^a ,3 ^a	C2	

FIGURA 13: Divisão das 18 proteínas homólogas a C1QBP_HUMAN em grupos e subgrupos de similaridades. A identidade das proteínas e os respectivos organismos estão indicados na FIGURA 8.

FONTE: GOMES (2014)



FIGURA 14: Alinhamento das proteínas de tripanossomatídeos com alto grau de similaridade a C1QBP_HUMAN, pertencentes ao grupo A (FIGURA 13). O alinhamento foi feito no Clustal Omega. A identidade das proteínas e os respectivos organismos estão indicados na FIGURA 8. (*) Aminoácidos idênticos; (:) Aminoácidos de alta similaridade; (.) Aminoácidos de baixa similaridade. Os aminoácidos estão coloridos de acordo com a classificação do grupo R. FONTE: GOMES (2014)

```

GOTWR1      MRRALLCVSRGVH-ATPMMAGRFA---ALASMVGRPEVSWTHASVLQQRFASCAALANYI
GOUNJ8      MRRRAIVLATFGARWAAPI-----SGLSRQPEMGALPGMALRPVFSQMRLVSNSKLAAAT
Q4E0T1      MRRVLFSAAISFVPKAPLFHRCVPSVATAVSMG-----PLVAMEQRRLASHAALSSAL
Q4DG13      MRRVLFSAALSLVPKAPLFRRCVPSVAKAASMG-----PLVAMEQRRLASHAALSSAL
                ***.:. .: . :*: . * * : * : * :

GOTWR1      RTEIEDEKERAEKPKPEIPSGWTVHEHPGNMFFKLCKTYEDEEIVLHFKGTREA-EGDI
GOUNJ8      LSEMEDEQQRSEKPEKEMPASWTLDRKIGETFFTMRRTYEDEEIVLQYTGECE-KNGVA
Q4E0T1      RHELEEEQQRSEKPAKPELPAGWTLERKPGQMLFTMRKKHEDEEIVIRCLGEESGDDVV
Q4DG13      RHELEEEQQRSEKPAKPELPAGWTLERKPGQMLFTMRKKHEDEEIIIRCLGEESGDDVV
                *:*:*:*:*:* *:*:*:*:*:* * : * : : : * : : : * : : :

GOTWR1      FYDFKVFVNGEKGVFDSLSDH-EIVVDRVIFLRDAKVAIKQSALGQPHDPFIYPGPKM
GOUNJ8      THTFIVFVCKNKGLVFNMSVEEGEIVLNNVCFRQDAKLAMDSAAESQAKNDLLYGDPDV
Q4E0T1      SLDFDAYITCNNKALVCRMSFESEEVIMGQVSFLDDAKLALDDVEGNQRQWLYKGPKL
Q4DG13      SLDFDAYITCNNKALVCRMSFESEEVIMGQVSFLDDAKLALDDVEGNRKRQWLYKGPKL
                * . : . : * : : : * : : : * * * * * : * : : : * * : :

GOTWR1      EDMEDNVVEAFIRYLEARGVNDELGDFVLHYACWAEQVEYEQWLSDIHKFVA
GOUNJ8      ADLEDSLVEAFASYLEERGVNDDLGNFIERYSYWAEQAEEYEWLGNINKFVS
Q4E0T1      DELDERLVDSLTSYLKDRGVNEDLCRFMEEYFFWAEQAEEYEWLSAINRFVS
Q4DG13      DELDERLVDSLTSYLKDRGVNEDLCRFMEEYFFWAEQAEEYEWLSAINRFV
                : : : : * : : : * * : * * * : * * * * * : * : : * : : :

```

FIGURA 15: Alinhamento das proteínas de tripanossomatídeos com alto grau de similaridade a C1QBP_HUMAN, pertencentes ao grupo B (FIGURA 13). O alinhamento foi feito do Clustal Omega. A identidade das proteínas e os respectivos organismos estão indicados na FIGURA 8. (*) Aminoácidos idênticos; (:) Aminoácidos de alta similaridade; (.) Aminoácidos de baixa similaridade. Os aminoácidos estão coloridos de acordo com a classificação do grupo R.
 FONTE: GOMES (2014)

```

E9BLJ6      -----MSGFDRVPSTSAPTALTVPRRSASDAALADA
A4I5C8      -----MSGFDRVPSTSAPTALTVPRRSASDAALADA
E9B0M6      MRRTFNKLGLSALRACRGAGAAAGAAASAHVSGFARVPSTSAPTALTVPRRSASDAALADA
                                                    :***
*****
*****
E9BLJ6      TRRELEEEMGRSDKPEQPTPPAGWQVVRKPGTCTFDLTKSFEGEELVVRYSSTNQDSDKAN
A4I5C8      TRRELEEEMGRSDKPEQPTPPAGWQVVRKPGTCTFDLTKSFEGEELVVRYSSTNQDSDKAN
E9B0M6      TRRELEEEMGRSDKPEQPTPPAGWQVVRKSGTCTFDLTKSFEDEELVVRYSSTNQDSDKAN
                *****.***** *****
*****
E9BLJ6      SHDIFVYITQTNGQTMQADLSIEEGELVLNNIRFYDEAALAKDTSAAEAEAKRNELYTGPL
A4I5C8      SHDIFVYITQTNGQTMQADLSIEEGELVLNNIRFYDEAALAKDTSAAEAEAKRNELYTGPL
E9B0M6      SHDIFVYITQKNGQTMQADLSIEEGELVLNNIRFYDEAALAKDTSAAEAEAKRNELYTGPL
                *****.*****
*****
E9BLJ6      VHELDYDLLNCVMTYLEKRGVDEKLGFEVVLYSFWAEQQDYEAWLTTMNKFAS
A4I5C8      VHELDYDLLNCVMTYLEKRGVDEKLGFEVVLYSFWAEQQDYEAWLTTMNKFAS
E9B0M6      VHELDYDLLNCVMTYLEKRGVDEKLGFEVVLYSFWAEQQDYEAWLTTMNKFAS
                *****

```

FIGURA 16: Alinhamento das proteínas de tripanossomatídeos com alto grau de similaridade a C1QBP_HUMAN, pertencentes ao grupo C (FIGURA 13). O alinhamento foi feito no Clustal Omega. A identidade das proteínas e os respectivos organismos estão indicados na FIGURA 8. (*) Aminoácidos idênticos; (:) Aminoácidos de alta similaridade; (.) Aminoácidos de baixa similaridade. Os aminoácidos estão coloridos de acordo com a classificação do grupo R.

FONTE: GOMES (2014)

3.3.5 IDENTIFICAÇÃO DOS GENES CODIFICADORES DAS PROTEÍNAS SELECIONADAS

O(s) gene(s) que transcreve(m) a proteína C1QBP_HUMAN (Q07021) e cada uma das 18 proteínas homólogas a C1QBP_HUMAN e com domínio MAM33, nos seus respectivos organismos, foram determinados por pesquisa ao GenBank. O QUADRO 5 contém as informações: 1. *Entry* da proteína, 2. (GI) código dos genes no GenBank e 3. (*Organism*) Organismo a que pertence à proteína e o gene.

Entry	GI	Organism
Q07021	708	Homo sapiens
A4HCI6	322504943	Leishmania braziliensis
A4I007	321399431	Leishmania infantum
A4I5C8	321398530	Leishmania infantum
C9ZSQ9	261329040	Trypanosoma brucei
E9AVY0	322491694	Leishmania mexicana
E9B0M6	322493392	Leishmania mexicana
E9BFZ8	361050152	Leishmania donovani
E9BLJ6	361050143	Leishmania donovani
F9W3J5	343477465	Trypanosoma congolense (strain IL3000)
G0TWR1	340053916	Trypanosoma vivax (strain Y486)
G0UNJ8	342181295	Trypanosoma congolense (strain IL3000)
K2NJJ1	407426008	Trypanosoma cruzi
K4EA71	407859520	Trypanosoma cruzi
Q4DG13	70878176	Trypanosoma cruzi
Q4DQ42	70881613	Trypanosoma cruzi
Q4E0T1	70885557	Trypanosoma cruzi
Q4QBJ9	321399744	Leishmania major
Q57WG2	62175901\7080218	Trypanosoma brucei

QUADRO 5: Código dos genes no GenBank referentes as 21 proteínas homólogas a C1QBP_HUMAN e com domínio MAM33. A identidade das proteínas está indicada na FIGURA 8. FONTE: GOMES (2014)

3.4 CONSULTA A OUTRAS BASES DE DADOS GENÔMICOS

Na análise realizada com o conjunto de proteínas pertencentes às vias de ativação do sistema complemento ou de proteínas que estejam relacionadas à regulação dessas vias, na base de dados NR de proteínas e utilizando o PSI-Blast (item 3.3), apenas a proteína C1QBP_HUMAN reguladora do sistema complemento apresentou homologia com as proteínas de tripanossomatídeos. Com a finalidade de garantir a confiabilidade e qualidade destas informações as bases de dados TriTrypDB e GenBank também foram analisadas com o objetivo de buscar outras proteínas homólogas a C1QBP_HUMAN.

3.4.1 BUSCA POR PROTEÍNAS HOMÓLOGAS A C1QBP_HUMAN NA BASE DE DADOS TRITRYPDB

A proteína C1QBP_HUMAN foi confrontada contra toda base de dados TriTrypDB utilizando o programa Blastp. Ao total foram selecionadas 50 proteínas pertencentes a 6 espécies do gênero *Leishmania* e 4 espécies do gênero *Trypanosoma*, sendo que para este último, foi possível ainda identificar as cepas (QUADRO 6).

Protein (TriTrypDB)	Score	Identities	Length	Strain	Species	Subgenus	Genus	Family
LbrM.22.1360	95	39%	248	<i>braziliensis</i>	<i>Leishmania braziliensis</i>			<i>Trypanosomatidae</i>
LbrM.05.0770	77	30%	340					
LdBPK_221330.1	96	27%	255	<i>donovani</i>	<i>Leishmania onovani</i>			
LdBPK_050790.1	79	30%	387					
LinJ.22.1330	96	27%	255	<i>JPCM5</i>	<i>Leishmania infantum</i>			
LinJ.05.0790	79	30%	390					
LmjF.22.1480	93	39%	249					
LmjF.30.0970	67	32%	195	<i>Friedlin</i>	<i>Leishmania major</i>	<i>Leishmania</i>	<i>Leishmaniinae</i>	
LmjF.05.0790	77	30%	388					
LmxM.22.1480	92	39%	249					
LmxM.29.0970	66	25%	233	<i>mexicana</i>	<i>Leishmania mexicana</i>			
LmxM.05.0790	79	30%	389					
LtaP22.1460	92	39%	255	<i>tarentolae</i>	<i>Leishmania lizard</i>			
LtaP05.0810	52	23%	384					
Tbg972.7.3800	98	40%	260					
Tbg972.6.2180	78	26%	227	<i>gambiense</i>				
Tbg972.7.8330	81	35%	369					
Tb927.7.3470	72	35%	260					
Tb927.6.2420	78	26%	227	<i>TREU927</i>	<i>Trypanosoma brucei</i>	<i>Trypanozoon</i>		
Tb927.7.7080	81	35%	372					
Tb427.07.3470	98	40%	260					
Tb427.06.2420	78	26%	227	<i>Strain 427</i>				
Tb427.07.7080	81	35%	372					
TcIL3000_0_26670	91	34%	260					
TcIL3000_7_5760	84	24%	372	<i>IL3000</i>	<i>Trypanosoma congolense</i>	<i>Nannomonas</i>		
TcIL3000_6_1930	68	26%	226					
TcCLB.509965.290	76	28%	225	<i>esmeraldo</i>			<i>Trypanosoma</i>	
TcCLB.508323.120	87	34%	261					
TcCLB.510615.10	74	30%	342					
TcCLB.509053.70	76	28%	225					
TcCLB.511825.100	88	27%	261	<i>non-esmeraldo</i>	<i>Trypanosoma cruzi</i>	<i>Schizotrypanum</i>		
TcCLB.506789.330	74	30%	341					
Tc_MARK_209	89	36%	261	<i>marinkellei</i>				
Tc_MARK_756	79	30%	180					
TCSYLVIO_001808	88	27%	261	<i>sylvio</i>				
TCSYLVIO_001971	74	30%	404					
TvY486_0703360	84	34%	259	<i>Y486</i>	<i>Trypanosoma vivax</i>	<i>Duttonella</i>		
TvY486_0019440	80	33%	327					

QUADRO 6: Proteínas homólogas a C1QBP_HUMAN selecionadas da base de dados TriTrypDB e os respectivos organismos.

FONTE: GOMES (2014)

3.4.1.1 ANÁLISE DE DOMÍNIO FUNCIONAL DAS PROTEÍNAS SELECIONADAS EM TRITRYPDB

As 50 proteínas selecionadas na base de dados TriTrypDB que possuem características de homologia com a proteína C1QBP_HUMAN foram submetidas ao Pfam para determinação de seus domínios funcionais. Trinta e oito proteínas possuem o domínio funcional MAM33, 3 proteínas o domínio *KINESIN* (3.3.1), 1 proteína o domínio *DUF3720*, 1 proteína o domínio *DRAXIN* e 7 proteínas não apresentaram nenhum domínio funcional (FIGURA 17). O domínio *DUF3720* pertence à família de domínios *DUF* cuja função é desconhecida, já o domínio *DRAXIN* pertence a uma família de proteínas inibidoras da via de sinalização WNT como também de moléculas de orientação axonal (<http://pfam.sanger.ac.uk>). Somente as proteínas contendo o domínio MAM33 foram selecionadas para a próxima análise.

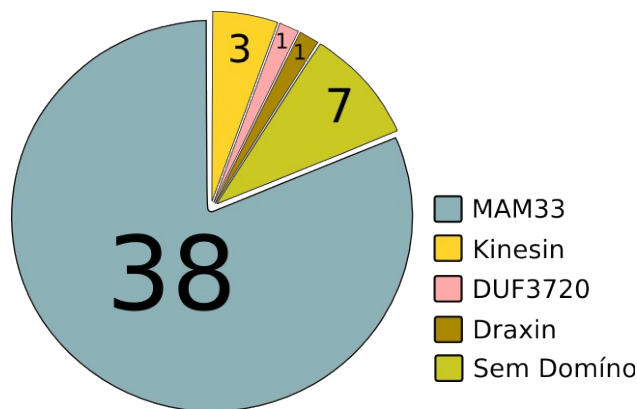


FIGURA 17: Domínio funcional das proteínas homólogas a C1qBP encontradas na base de dados TriTrypDB
 FONTE: GOMES (2014)

3.4.1.2 BUSCA POR PROTEÍNAS HOMÓLOGAS AS PROTEÍNAS ALVOS NA BASE DE TRITRYDB

As 38 proteínas com domínio MAM33 e homólogas a C1qBP_HUMAN selecionadas (3.4.1.1) foram submetidas aos programas SignalP, Pred-GPI e TargetP para predição de sequência de direcionamento celular. Observamos que 33 das 38 proteínas analisadas possuem o mesmo direcionamento proteico que a proteína C1QBP_HUMAN (3.3.3), ou seja, são direcionadas para a mitocôndria (FIGURA 18). Porém, 3 proteínas que apresentaram sequência de direcionamento para a mitocôndria também apresentam sinal de importação para o retículo endoplasmático (peptídeo sinal), 1 proteína apresenta sequência que direciona para a membrana (sinal transmembrana) e 1 proteína não apresenta sequência de direcionamento, provavelmente permanecendo no citoplasma da célula (FIGURA 18).

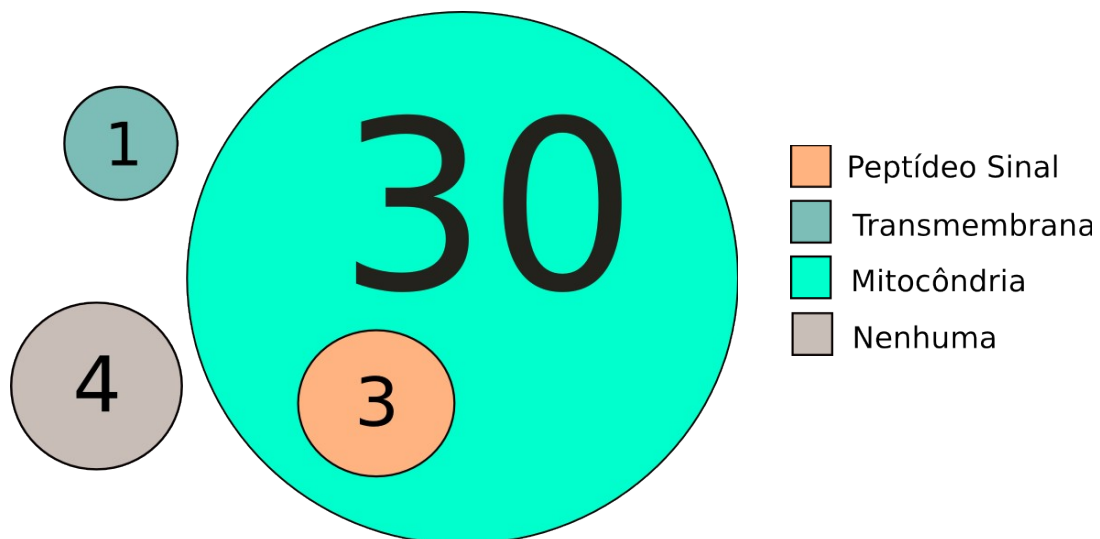


FIGURA 18: Resultado da predição de direcionamento proteico das proteínas homólogas a C1qBP_HUMAN selecionadas na base de dados TriTrypDB e que possuem domínio MAM33.
 FONTE: GOMES (2014)

3.4.1.3 INTEGRAÇÃO ENTRE BASE DE DADOS

O padrão de identificação dos dados na base de dados TriTrypDB difere do padrão utilizado pela base de dados UniProtKB e desta forma foi necessário confirmar a existência das proteínas na base TriTrypDB selecionadas (3.4.1) na base de dados do UniProtKB. Utilizando o programa Blastp, confrontamos todas as proteínas da coluna *Protein (TriTrypDB)* da FIGURA 19 (uma proteína por vez) contra a base de dados do UniProtKB e as proteínas resultantes que apresentam 100% de similaridade foram anotadas na coluna (*Entry*) como proteínas equivalentes. A figura 19 relaciona para cada proteína de TriTrypDB (coluna *Protein (TriTrypDB)*) sua equivalente no UniProtKB (coluna *Entry*).

Foi observado nesta análise que apenas as proteínas LtaP22.1460 e LtaP05.0810 do organismo *L. tarentolae* não apresentaram nenhuma proteína equivalente na base de dados do UniProtKB apesar deste organismo está contido no UniProtKB.

Protein (TriTrypDB)	Entry
LbrM.22.1360	A4HCI6
LdBPK_050790.1	E9B8G9
LdBPK_221330.1	E9BFZ8
LinJ.05.0790	A4HSK1
LinJ.22.1330	A4I007
LmjF.05.0790	Q4QJC3
LmjF.22.1480	Q4QBJ9
LmxM.05.0790	E9AKI5
LmxM.22.1480	E9AVY0
LmxM.29.0970	E9B0M6
LtaP05.0810	
LtaP22.1460	
Tb427.06.2420	Q94R46
Tb427.07.3470	Q57WG2
Tb427.07.7080	Q57XS8
Tb927.6.2420	Q584R4
Tb927.7.3470	Q57WG2
Tb927.7.7080	Q57XS8
Tbg972.6.2180	Q94R46
Tbg972.7.3800	C9ZSQ9
Tbg972.7.8330	C9ZU92
TcCLB.506789.33	Q4DZV
TcCLB.509053.70	Q4DG13
TcCLB.509965.29	Q4E0T
TcCLB.510615.10	Q4CNB4
TcIL3000_0_2667	F9W3J
TcIL3000_6_1930	G0UNJ8
TcIL3000_7_5760	G0UQV1
TCSYLVI0_001808	K4EA71
Tc_MARK_209	K2NJJ1
Tc_MARK_756	K2PDK3
TvY486_0019440	F9WNY2
TvY486_0703360	G0TYF5

FIGURA 19: Equivalência entre as proteínas das bases de dados TriTrypDB e UniProtKB. A identidade das proteínas e os respectivos organismos estão indicados na FIGURA 8.
 FONTE: GOMES (2014)

3.4.2 SELEÇÃO DE TRYPANOSOMATÍDEOS PARA ANÁLISE DE SEQUÊNCIA NO GENBANK

O GenBank foi a terceira base de dados estudada para a busca de proteínas homólogas a proteína C1qBP_HUMAN. Onze representantes de tripanossomatídeos foram selecionados na literatura e seus dados armazenados no BDLP (2.1) (QUADRO 7).

taxonomy	name
353153	Trypanosoma cruzi (strain CL Brener)
85056	Trypanosoma cruzi marinkellei
5693	Trypanosoma cruzi
31285	Trypanosoma brucei gambiense
5702	Trypanosoma brucei brucei
5691	Trypanosoma brucei
5665	Leishmania mexicana
5664	Leishmania major
5671	Leishmania infantum
5661	Leishmania donovani
5660	Leishmania braziliensis

QUADRO 7: Tripanossomatídeos selecionados para busca de proteínas na base de dados do GenBank.

FONTE: GOMES (2014)

3.4.2.1 COLETA DE DADOS NO GENBANK

Foi desenvolvido um *script* em Perl (2.8.1) para capturar no GenBank as informações referentes aos bancos de proteínas e aos códigos FASTA's das proteínas dos tripanossomatídeos selecionados. O código taxonômico de cada organismo e quantidade de proteínas capturadas para cada um deles estão apresentados no QUADRO 8. As 153.372 sequências de resíduos de aminoácidos foram armazenadas no BDLP e utilizadas para a comparação com as 49 proteínas do sistema complemento humano selecionados para este projeto de pesquisa (3.1).

Taxonoma	Organismo	Proteínas
5660	<i>Leishmania braziliensis</i>	8246
5661	<i>Leishmania donovani</i>	8661
5664	<i>Leishmania major</i>	8369
5665	<i>Leishmania mexicana</i>	8194
5671	<i>Leishmania infantum</i>	8278
5691	<i>Trypanosoma brucei</i>	20209
5693	<i>Trypanosoma cruzi</i>	42970
5702	<i>Trypanosoma brucei brucei</i>	9319
31285	<i>Trypanosoma brucei gambiense</i>	9756
85056	<i>Trypanosoma cruzi marinkellei</i>	10127
353153	<i>Trypanosoma cruzi</i> (strain CL Brener)	19243

QUADRO 8: Espécies de tripanossomatídeos selecionados da literatura e seus respectivos códigos taxonômicos.

FONTE: GOMES (2014)

3.4.2.2 CONFRONTO DAS PROTEÍNAS ALVOS CONTRA AS PROTEÍNAS SELECIONADAS DO GENBANK

As 49 proteínas alvos do sistema complemento (3.1) foram confrontadas contra as 153.372 proteínas selecionadas do GenBank (QUADRO 8) utilizando o programa Blastp. Foram gerados 3.459 registros sendo, dentre estes, selecionados 1.798 proteínas distintas. Dentre estas 61 proteínas foram consideradas homólogas a C1QBP_HUMAN (FIGURA 20).

Entry	Length	Organism	Score	E-Value	Identities
Q57WG2	260	Trypanosoma brucei	42.4	0.0001	40
C9ZSQ9	260	Trypanosoma brucei	42.4	0.0001	40
A4HCI6	248	Leishmania braziliensis	41.2	0.0001	39.58
Q4QBJ9	249	Leishmania major	40.4	0.0002	39.58
E9BFZ8	255	Leishmania donovani	40.4	0.0003	39.58
E9AVY0	249	Leishmania mexicana	40	0.0003	39.58
A4I007	255	Leishmania infantum	40.4	0.0003	39.58
Q4DQ42	261	Trypanosoma cruzi	39.3	0.002	34
Q4DFL0	261	Trypanosoma cruzi	39.3	0.002	34
K2NJJ1	261	Trypanosoma cruzi	39.7	0.002	36
K4EA71	261	Trypanosoma cruzi	39.3	0.002	34
B3GVT4	499	Trypanosoma brucei	38.1	0.004	36.21
Q58NS4	499	Trypanosoma brucei	38.1	0.004	36.21
C9ZU92	369	Trypanosoma brucei	37.4	0.005	34.55
Q57XS8	372	Trypanosoma brucei	37.4	0.005	34.55
E9B8G9	387	Leishmania donovani	35.8	0.009	29.09
E9AKI5	389	Leishmania mexicana	35.8	0.01	29.09
A4HSK1	390	Leishmania infantum	35.8	0.01	29.09
A4H4C2	340	Leishmania braziliensis	34.7	0.022	29.09
Q4QJC3	388	Leishmania major	34.7	0.026	29.09
A4HI52	191	Leishmania braziliensis	31.2	0.17	32.65
Q4CNB4	342	Trypanosoma cruzi	33.9	0.17	29.09
Q4DZV7	341	Trypanosoma cruzi	33.5	0.18	29.09
E9B0M6	233	Leishmania mexicana	31.2	0.2	32.65
Q4Q7K6	195	Leishmania major	31.2	0.2	32.65
E9BLJ6	204	Leishmania donovani	31.2	0.21	32.65
A4I5C8	204	Leishmania infantum	31.2	0.21	32.65
K4E9M9	404	Trypanosoma cruzi	33.5	0.22	29.09
K2PDK3	180	Trypanosoma cruzi	32.3	0.29	29.09
Q4DRU6	184	Trypanosoma cruzi (strain CL Brener)	29.6	1.1	31.08
A4I0K5	285	Leishmania infantum	28.9	1.3	34.85
E9BG17	285	Leishmania donovani	28.9	1.3	34.85
Q4E5V6	139	Trypanosoma cruzi	30	1.4	26.44
A4HM84	1714	Leishmania braziliensis	29.3	1.5	39.34
E9BHL9	6525	Leishmania donovani	28.9	1.7	38
Q4Q8R2	1888	Leishmania major	28.9	1.7	36.84
A4I1F3	6530	Leishmania infantum	28.9	1.8	38
Q4CR08	560	Trypanosoma cruzi (strain CL Brener)	29.6	1.8	35.29
E9B465	1433	Leishmania mexicana	28.9	1.9	24.39
D0A956	787	Trypanosoma brucei	29.6	2	41.18
Q4DG13	225	Trypanosoma cruzi	30	2.2	35.29
Q4E0T1	225	Trypanosoma cruzi	29.6	2.6	35.29
Q4CZM2	1425	Trypanosoma cruzi	30	2.9	29.49
Q4E4N0	666	Trypanosoma cruzi (strain CL Brener)	28.9	2.9	38.64
Q4D6V2	337	Trypanosoma cruzi	29.6	3.1	28.75
Q4CN74	560	Trypanosoma cruzi (strain CL Brener)	28.9	3.2	33.33
A4HC93	424	Leishmania braziliensis	27.3	5.1	34.78
Q4QFK9	205	Leishmania major	26.9	5.1	45.83
C9ZXL2	1397	Trypanosoma brucei gambiense	27.3	5.4	27.78
Q38FA4	1397	Trypanosoma brucei brucei	27.3	5.7	27.78
Q4DFS6	337	Trypanosoma cruzi (strain CL Brener)	27.7	5.8	27.5
Q4DU18	143	Trypanosoma cruzi (strain CL Brener)	27.3	6.2	28.57
Q383A1	812	Trypanosoma brucei brucei	26.9	6.3	38.24
Q4Q803	885	Leishmania major	27.3	6.3	35.85
E9B3B8	1317	Leishmania mexicana	27.3	6.4	26.32
D0A7A5	626	Trypanosoma brucei	27.7	6.4	30.95
Q385G9	616	Trypanosoma brucei	28.1	6.4	30.95
E9B5V4	925	Leishmania mexicana	26.9	7.5	35.29
K2N810	337	Trypanosoma cruzi	28.5	7.7	30
A4H642	6733	Leishmania braziliensis	26.9	8.6	30.14
E9AIP3	880	Leishmania braziliensis	26.6	8.9	30.14

FIGURA 20: Proteínas selecionadas do GenBank consideradas homólogas a C1QBP_HUMAN. 1. (Entry) identificador da proteína na base de dados UniProtKB, 2.(Length) quantidade de aminoácidos que compõe a proteína, 3.(Organism) organismo que pertence a proteína, 4.(Score) pontuação atribuída pelo Blastp referente ao alinhamento da proteína quando comparada com C1QBP_HUMAN, 5.(E-Value) número de alinhamentos esperado aleatoriamente e 6.(Identities) Percentual de identidade quando comparada com C1QBP_HUMAN.

FONTE: GOMES (2014)

3.4.2.3 ANÁLISE DO DOMÍNIO FUNCIONAL DAS PROTEÍNAS

As 61 proteínas inicialmente consideradas homólogas a C1QBP_HUMAN foram submetidas ao programa Pfam e observou-se que apenas 29 delas possuem o domínio funcional MAM33 que caracteriza a C1QBP_HUMAN (FIGURA 21).

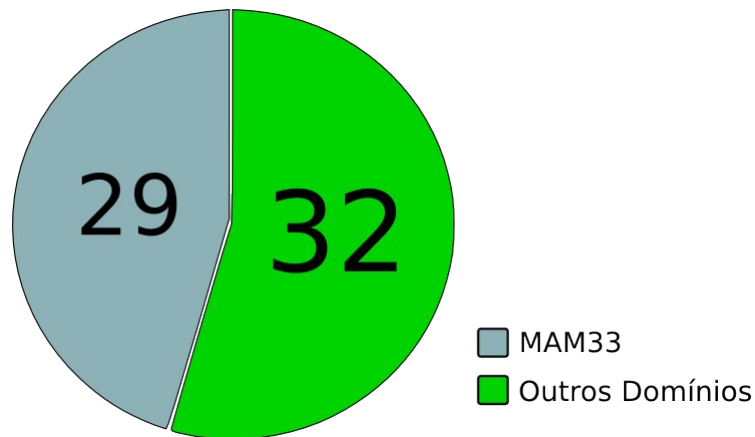


FIGURA 21: Separação das 61 proteínas inicialmente consideradas homólogas a C1QBP_HUMAN que possuem domínio MAM33 e que não possuem o domínio MAM33.
FONTE: GOMES (2014)

Os resultados obtidos no Pfam foram armazenados no BDLP e estão apresentados na figura 22 (coluna *Domain*). É possível perceber que as proteínas que possuem domínio MAM33 possuem também os menores valores de E-Value.

Os outros domínios não foram analisados em detalhe, pois o objetivo nesta etapa foi buscar proteínas homólogas a C1QBP_HUMAN com domínio MAM33 para validarmos os resultados das outras análises.

Entry	Length	Organism	Score	E-Value	Identities	Domain
Q57WG2	260	Trypanosoma brucei	42.4	0.0001	40	MAM33
C9ZSQ9	260	Trypanosoma brucei	42.4	0.0001	40	MAM33
A4HCI6	248	Leishmania braziliensis	41.2	0.0001	39.58	MAM33
Q4QBJ9	249	Leishmania major	40.4	0.0002	39.58	MAM33
E9BFZ8	255	Leishmania donovani	40.4	0.0003	39.58	MAM33
E9AVY0	249	Leishmania mexicana	40	0.0003	39.58	MAM33
A4I007	255	Leishmania infantum	40.4	0.0003	39.58	MAM33
Q4DQ42	261	Trypanosoma cruzi	39.3	0.002	34	MAM33
Q4DFL0	261	Trypanosoma cruzi	39.3	0.002	34	MAM33
K2NJJ1	261	Trypanosoma cruzi	39.7	0.002	36	MAM33
K4EA71	261	Trypanosoma cruzi	39.3	0.002	34	MAM33
B3GVT4	499	Trypanosoma brucei	38.1	0.004	36.21	Trypan_glycop_C
Q58NS4	499	Trypanosoma brucei	38.1	0.004	36.21	Trypan_glycop_C
C9ZU92	369	Trypanosoma brucei	37.4	0.005	34.55	MAM33
Q57XS8	372	Trypanosoma brucei	37.4	0.005	34.55	MAM33
E9B8G9	387	Leishmania donovani	35.8	0.009	29.09	MAM33
E9AKI5	389	Leishmania mexicana	35.8	0.01	29.09	MAM33
A4HSK1	390	Leishmania infantum	35.8	0.01	29.09	MAM33
A4H4C2	340	Leishmania braziliensis	34.7	0.022	29.09	MAM33
Q4QJC3	388	Leishmania major	34.7	0.026	29.09	MAM33
A4HI52	191	Leishmania braziliensis	31.2	0.17	32.65	MAM33
Q4CNB4	342	Trypanosoma cruzi	33.9	0.17	29.09	MAM33
Q4DZV7	341	Trypanosoma cruzi	33.5	0.18	29.09	MAM33
E9B0M6	233	Leishmania mexicana	31.2	0.2	32.65	MAM33
Q4Q7K6	195	Leishmania major	31.2	0.2	32.65	MAM33
E9BLJ6	204	Leishmania donovani	31.2	0.21	32.65	MAM33
A4I5C8	204	Leishmania infantum	31.2	0.21	32.65	MAM33
K4E9M9	404	Trypanosoma cruzi	33.5	0.22	29.09	MAM33
K2PDK3	180	Trypanosoma cruzi	32.3	0.29	29.09	MAM33
Q4DRU6	184	Trypanosoma cruzi (strain CL Brener)	29.6	1.1	31.08	
A4I0K5	285	Leishmania infantum	28.9	1.3	34.85	
E9BGI7	285	Leishmania donovani	28.9	1.3	34.85	
Q4E5V6	139	Trypanosoma cruzi	30	1.4	26.44	
A4HM84	1714	Leishmania braziliensis	29.3	1.5	39.34	Pkinase
E9BHL9	6525	Leishmania donovani	28.9	1.7	38	Kelch_4
Q4Q8R2	1888	Leishmania major	28.9	1.7	36.84	
A4I1F3	6530	Leishmania infantum	28.9	1.8	38	Kelch_4
Q4CR08	560	Trypanosoma cruzi (strain CL Brener)	29.6	1.8	35.29	CoarA
E9B465	1433	Leishmania mexicana	28.9	1.9	24.39	Peptidase_C97I_PUB\ Glyco
D0A956	787	Trypanosoma brucei	29.6	2	41.18	PIF1
Q4DG13	225	Trypanosoma cruzi	30	2.2	35.29	MAM33
Q4E0T1	225	Trypanosoma cruzi	29.6	2.6	35.29	MAM33
Q4CZM2	1425	Trypanosoma cruzi	30	2.9	29.49	
Q4E4N0	666	Trypanosoma cruzi (strain CL Brener)	28.9	2.9	38.64	Cyclin_N
Q4D6V2	337	Trypanosoma cruzi	29.6	3.1	28.75	Ribonuc_red_sm
Q4CN74	560	Trypanosoma cruzi (strain CL Brener)	28.9	3.2	33.33	CoarA
A4HC93	424	Leishmania braziliensis	27.3	5.1	34.78	tRNA_bind
Q4QFK9	205	Leishmania major	26.9	5.1	45.83	
C9ZXL2	1397	Trypanosoma brucei gambiense	27.3	5.4	27.78	RabGAP-TBC
Q38FA4	1397	Trypanosoma brucei brucei	27.3	5.7	27.78	RabGAP-TBC
Q4DFS6	337	Trypanosoma cruzi (strain CL Brener)	27.7	5.8	27.5	Ribonuc_red_sm
Q4DU18	143	Trypanosoma cruzi (strain CL Brener)	27.3	6.2	28.57	
Q383A1	812	Trypanosoma brucei brucei	26.9	6.3	38.24	PIF1
Q4Q803	885	Leishmania major	27.3	6.3	35.85	IQ
E9B3B8	1317	Leishmania mexicana	27.3	6.4	26.32	zf-B_box
D0A7A5	626	Trypanosoma brucei	27.7	6.4	30.95	TruD
Q385G9	616	Trypanosoma brucei	28.1	6.4	30.95	TruD
E9B5V4	925	Leishmania mexicana	26.9	7.5	35.29	
K2N810	337	Trypanosoma cruzi	28.5	7.7	30	Ribonuc_red_sm
A4H642	6733	Leishmania braziliensis	26.9	8.6	30.14	Baculo_11_kDa
E9AIP3	880	Leishmania braziliensis	26.6	8.9	30.14	

FIGURA 22: Proteínas selecionadas do GenBank consideradas homólogas a C1QBP_HUMAN com ênfase na presença do domínio MAM33 (coluna *Domain*).

FONTE: GOMES (2014)

3.4.2.4 INVESTIGAÇÃO DE DIRECIONAMENTO DAS PROTEÍNAS COM DOMÍNIO MAM33

As 29 proteínas selecionadas no item 3.4.2.3 foram submetidas aos programas SignalP, PredGPI e TargetP com a finalidade de se detectar eventuais sequências de aminoácidos que indiquem o seu direcionamento proteico para a membrana plasmática ou, mitocôndria ou ainda, a possibilidade de serem secretadas da célula. Assim como a proteína C1QBP_HUMAN (3.3.3), 24 proteínas possuem seu direcionamento para a mitocôndria (FIGURA 23). Uma proteína apresenta sequência de proteínas transmembrana e 4 proteínas não apresentam sequência de direcionamento.

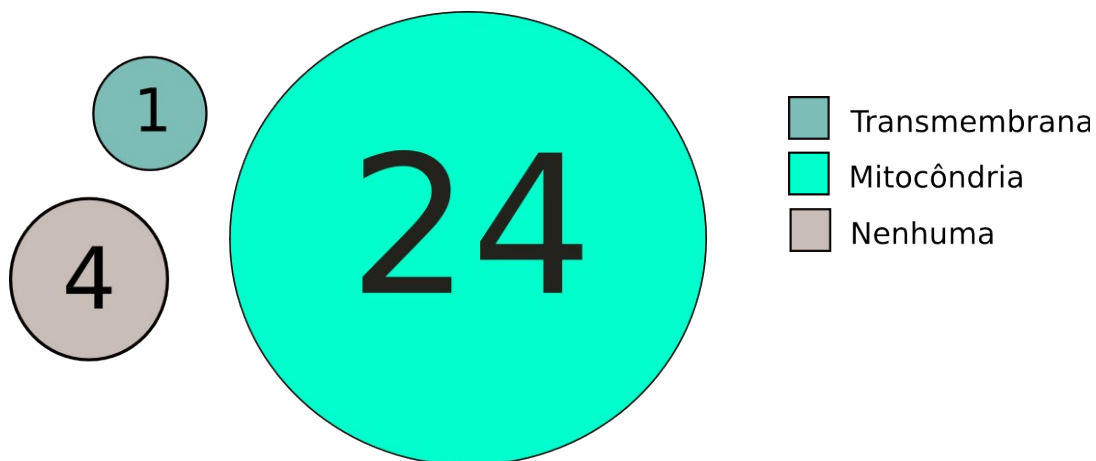


FIGURA 23: Resultado da predição de direcionamento proteico para proteínas homólogas a C1QBP_HUMAN e com domínio MAM33.
 FONTE: GOMES (2014)

3.5 PROTEÍNAS HOMÓLOGAS A C1QBP_HUMAN E COM DOMÍNIO MAM33 EM DIFERENTES PATÓGENOS DA BASE DE DADOS GENEDB

A base de dados GeneDB contém dados referentes a mais de 40 genomas de diversos patógenos, em vários estágios de anotação e curadoria (genedb.org). Esta base foi consultada com o objetivo de encontrarmos em patógenos que não pertencem ao grupo dos tripanossomatídeos proteínas homólogas a C1QBP_HUMAN e que também

possuem o domínio MAM33. Os patógenos selecionados para esta busca estão listados na FIGURA 24. O programa utilizado para o confronto da proteína C1QBP_HUMAN contra as proteínas foi o Blastp e os registros de proteínas homólogas foram armazenados no BDLP. Na FIGURA 24 está apresentada a quantidade de proteínas registradas para cada patógeno. Estas proteínas foram analisadas com o programa Pfam e foi possível observar que 18 das 47 espécies de patógenos analisadas possuem ao menos uma proteína homóloga a C1QBP_HUMAN e com o domínio MAM33 (FIGURA 24). Estes resultados são indicativos de que todos os organismos analisados possuem proteínas com diversos graus de homologia com a C1QBP_HUMAN, mas que nem todas possuem o domínio MAM33. A FIGURA 25 apresenta as proteínas dos organismos que possuem proteínas homólogas a C1QBP_HUMAN e com domínio MAM33.

Organismo	Registros	MAM33
<i>Plasmodium berghei</i>	32	x
<i>Plasmodium chabaudi</i>	43	x
<i>Plasmodium falciparum</i> 3D7	48	x
<i>Plasmodium falciparum</i> IT	39	
<i>Plasmodium knowlesi</i>	30	x
<i>Plasmodium reichenowi</i>	41	x
<i>Plasmodium vivax</i>	30	x
<i>Plasmodium yoelii</i> 17XNL	31	x
<i>Plasmodium yoelii</i> YM	31	x
<i>Bordetella bronchiseptica</i>	4	
<i>Burkholderia cenocepacia</i>	2	
<i>Bacteroides fragilis</i> NCTC 9343	9	
<i>Bordetella parapertussis</i>	3	
<i>Bordetella pertussis</i>	5	
<i>Burkholderia pseudomallei</i>	6	
<i>Bodo saltans</i>	12	x
<i>Bursaphelenchus xylophilus</i>	27	x
<i>Chlamydomonas abortus</i>	5	
<i>Corynebacterium diphtheriae</i>	8	
<i>Campylobacter jejuni</i>	10	
<i>Erwinia carotovora</i>	5	
<i>Echinococcus granulosus</i>	16	x
<i>Echinococcus multilocularis</i>	15	x
<i>Eimeria tenella</i>	12	x
<i>Globodera pallida</i>	9	
<i>Hymenolepis microstoma</i>	13	x
<i>Neospora caninum</i>	30	x
<i>Rhizobium leguminosarum</i>	4	
<i>Staphylococcus aureus</i> EMRSA15	16	
<i>Staphylococcus aureus</i> LGA251	13	
<i>Staphylococcus aureus</i> MRSA252	10	
<i>Staphylococcus aureus</i> MSSA476	11	
<i>Staphylococcus aureus</i> ST398	13	
<i>Staphylococcus aureus</i> TW20	12	
<i>Streptomyces coelicolor</i>	7	
<i>Schistosoma japonicum</i>	13	x
<i>Schistosoma mansoni</i>	28	x
<i>Streptococcus pneumoniae</i> ATCC 700669	10	
<i>Streptococcus pneumoniae</i> D39	6	
<i>Streptococcus pneumoniae</i> OXC141	3	
<i>Streptococcus pneumoniae</i> TIGR4	7	
<i>Streptococcus pyogenes</i>	7	
<i>Salmonella typhi</i>	6	
<i>Salmonella typhimurium</i>	5	
<i>Streptococcus uberis</i>	6	
<i>Theileria annulata</i>	32	
<i>Taenia solium</i>	18	x

FIGURA 24: Organismos patógenos selecionados da base de dados GeneDB e analisados quanto a apresentarem proteínas homólogas a C1QBP_HUMAN e que contenham domínio MAM33. Colunas: (Organismo) – Nome da espécie do organismo, (Registros) – Números de proteínas homólogas a C1QBP_HUMAN que foram encontradas no organismo e (X) – O organismo que possui ao menos uma proteína homóloga a C1qBP com domínio MAM33.

FONTE: GOMES (2014)

Domínio das proteínas do GeneDB Homologas a C1qBP e com Domínio MAM33 (Pfam)											
Protein	Organism	Length	Domain	Envelope		Envelope Length	Alignment		Alignment Length	Bit score	E-value
				Start	End		Start	End			
BS73295.1..pep	Bodo saltans	227	MAM33	51	226	175	52	225	173	61.6	8.2e_17
BS92610.1..pep	Bodo saltans	149	MAM33	11	148	137	58	147	89	32.5	6.7e_08
BUX_s00705.13.1..pep	Bursaphelenchus xylophilus	235	MAM33	48	232	184	48	232	184	153.6	5.3e_45
EgrG_000940000.1..pep	Echinococcus granulosus	242	MAM33	55	240	185	55	239	184	33.9	2.4e_08
Emuj_000940000.1..pep	Echinococcus multilocularis	242	MAM33	55	240	185	55	239	184	34.2	1.9e_08
HmN_000844700.1..pep	Hymenolepis microstoma	242	MAM33	55	239	184	56	235	179	26.4	4.8e_06
NCLIV_051840..pep	Neospora caninum	289	MAM33	173	287	114	180	287	107	58.4	8e_16
PBANKA_101000.1..pep	Plasmodium berghei	252	MAM33	125	250	125	130	250	120	64.7	8.9e_18
PCHAS_101090.1..pep	Plasmodium chabaudi	252	MAM33	125	250	125	130	250	120	65.5	5e_18
PF3D7_1434800.1..pep	Plasmodium falciparum 3D7	252	MAM33	125	250	125	130	250	120	66.5	2.6e_18
PKH_131450.1..pep	Plasmodium knowlesi	254	MAM33	127	252	125	132	252	120	65.8	4.3e_18
PRCDC_1434100.1..pep	Plasmodium reichenowi	252	MAM33	125	250	125	130	250	120	66.5	2.6e_18
PVX_084795.1..pep	Plasmodium vivax	254	MAM33	127	252	125	132	252	120	65.8	4.3e_18
PY17X_1011600.1..pep	Plasmodium yolkii 17XNL	252	MAM33	125	250	125	130	250	120	64.7	8.9e_18
PYYM_1011600.1..pep	Plasmodium yolkii YM	252	MAM33	125	250	125	130	250	120	64.7	8.9e_18
TsM_000611300.1..pep	Taenia solium	242	MAM33	55	240	185	95	239	144	42.8	4.6e_18

FIGURA 25: Análise comparativa do domínio MAM33 entre as proteínas homólogas a C1QBP_HUMAN encontradas em patógenos pertencente a base de dados GeneDB. A coluna (*Protein*) contém o identificador da proteína na base de dados GeneDB, a coluna (*Organism*) o organismo a que a proteína pertence. A coluna (*Length*) a quantidade de aminoácidos da proteína. Coluna (*Domain*) o domínio que a proteína possui. A coluna (*Envelope Start e End*) aponta o início e o fim do envelope onde está contido o domínio. A coluna *Envelope length* mostra o comprimento em aminoácidos do domínio dentro da proteína. A coluna (*Alignment Star e End*) aponta a localização do aminoácido que inicializa e finaliza o domínio e a coluna *Alignment length* mostra a quantidade de aminoácidos dentro do envelope que realmente formam o domínio.

FONTE: GOMES (2014)

3.6 INTEGRAÇÃO DOS DADOS COLETADOS NAS TRES BASES DE DADOS NR, TRITRYPDB E GENBANK

3.6.1 MINERAÇÃO DE DADOS ENTRE BASES DE PROTEÍNAS

Utilizando de técnicas de mineração de dados, foram desenvolvidos programas na linguagem SQL para buscar no conjunto de proteínas selecionadas quais delas eram comuns às três bases de dados (NR, TriTrypDB e GeneBank). A distribuição destas proteínas está representada na FIGURA 26 que resgata as informações das figuras 9, 17 e 21. Doze proteínas homólogas a C1qBP_HUMAN e com domínio MAM33 são comuns as três base de dados, um resultado que pode ser interpretado como um indicativo da confiabilidade da estratégia de busca.

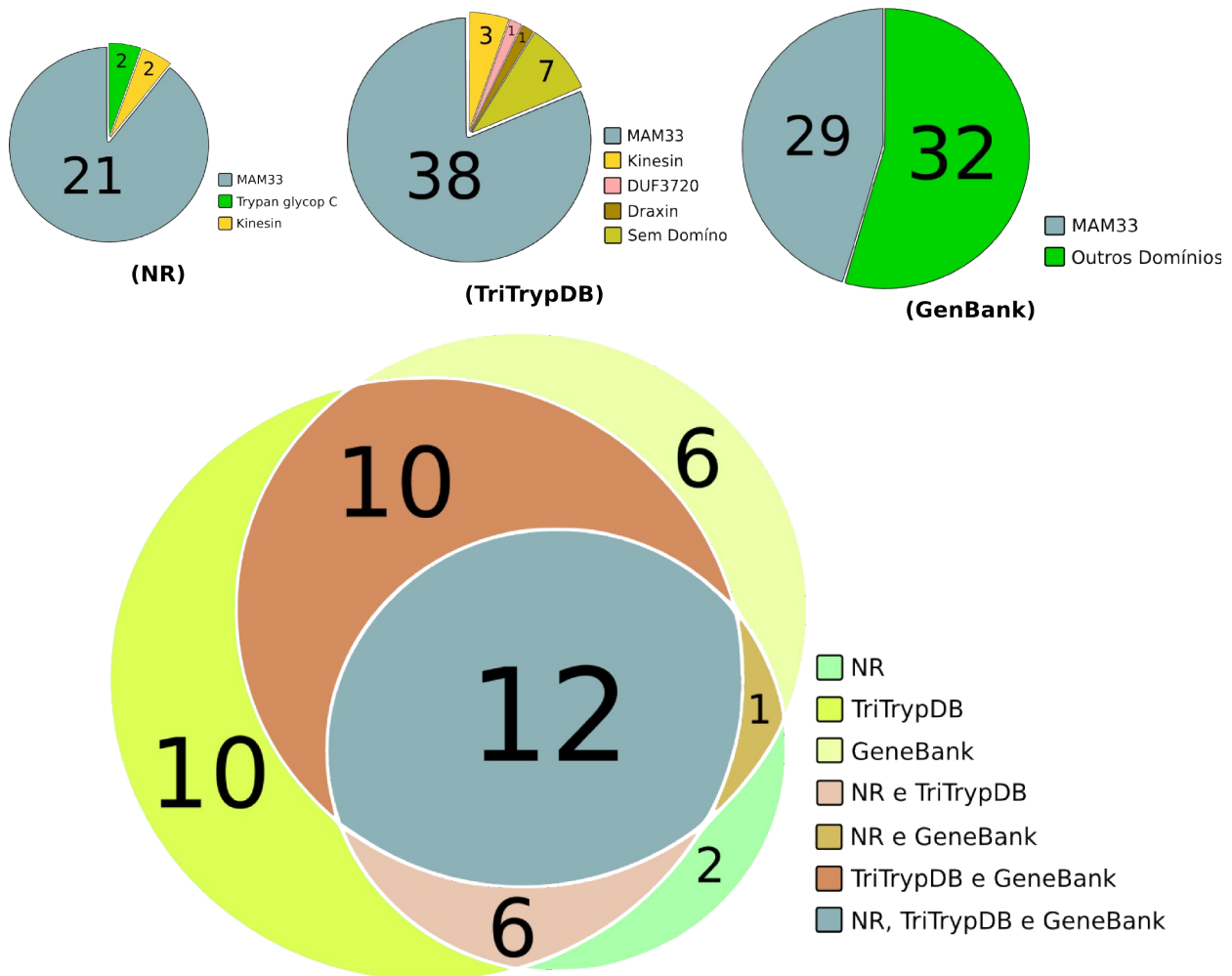


FIGURA 26: Distribuição por conjunto das proteínas homólogas a C1QBP_HUMAN e com domínio MAM33 encontradas nas bases de dados NR, TriTrypDB e GenBank.
 FONTE: GOMES (2014)

Um programa em SQL foi desenvolvido para buscarmos nas bases de dados NR, TriTrypDB e GenBank proteínas homólogas a C1qBP_HUMAN, que contém o domínio MAM33 e que possuem o mesmo direcionamento celular (item 3.3.3). O resultado desta mineração de dados está representado na FIGURA 27 que resgata as figuras 11, 18 e 23. Onze proteínas são comuns às três bases de dados (QUADRO 9) e quando estas foram comparadas com as 12 proteínas que contêm domínio MAM33 observamos que somente a proteína Q4Q7K6 não possui o direcionamento para a mitocôndria. Cinco proteínas foram encontradas em *Trypanosoma* sp e 6 proteínas foram encontradas em *Leishmania* sp. No caso da proteína Q57WG2 foram obtidos dois registros gênicos.

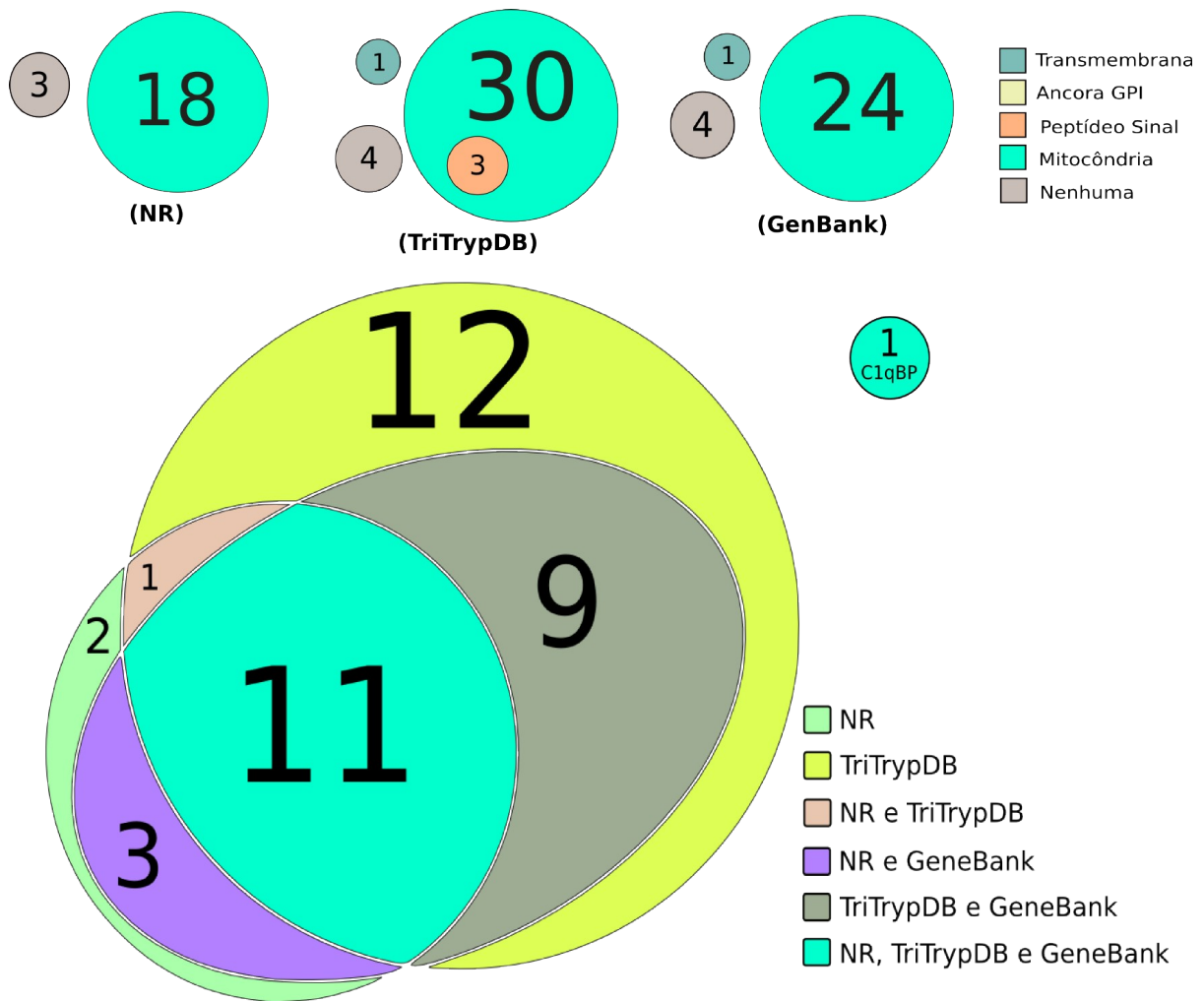


FIGURA 27: Distribuição por conjunto das proteínas homólogas a C1qBP_HUMAN, que possuem o domínio MAM33 e direcionamento para mitocôndria, encontradas nas bases de dados NR, TriTrypDB e GenBank.

FONTE GOMES (2014)

Entry	GI	Organism
A4HCI6	322504943	<i>Leishmania braziliensis</i>
A4I007	321399431	<i>Leishmania infantum</i>
C9ZSQ9	261329040	<i>Trypanosoma brucei</i>
E9AVY0	322491694	<i>Leishmania mexicana</i>
E9B0M6	322493392	<i>Leishmania mexicana</i>
E9BFZ8	361050152	<i>Leishmania donovani</i>
K2NJJ1	407426008	<i>Trypanosoma cruzi</i>
K4EA71	407859520	<i>Trypanosoma cruzi</i>
Q4DG13	70878176	<i>Trypanosoma cruzi</i>
Q4QBJ9	321399744	<i>Leishmania major</i>
Q57WG2	62175901\7080218	<i>Trypanosoma brucei</i>

QUADRO 9: Proteínas e seus respectivos organismos, extraídas por mineração de dados do conjunto de proteínas selecionadas das bases de dados NR, TriTrypDB e GenBank (FIGURA 26).

FONTE: GOMES (2014)

3.6.2 ALINHAMENTO E ANÁLISE DE SEMELHANÇA

As sequências de aminoácidos das 11 proteínas apresentadas no QUADRO 9 foram alinhadas com o objetivo de analisarmos os níveis de similaridade e as separarmos em grupos de maior semelhança . A FIGURA 28 detalha o resultado do alinhamento.

3.6.2.1 AGRUPAMENTO POR SEMELHANÇA DE ALINHAMENTO

O resultado do alinhamento das 11 proteínas apresentado na FIGURA 29 foi analisado e as proteínas foram divididas em dois grupos, A e B, com base na similaridade de alinhamento (FIGURAS 29 e 30). Esta análise mostra que os domínios em ambos os grupos são muito conservados. A relação filogenética entre as proteínas dos grupos A (azul) e B (vermelho) esta apresentada na FIGURA 31. As proteínas estão identificadas pelo seu código do UniProtKB (*Entry*) associado a um grau de identidade para com as outras proteínas. Esta árvore filogenética ou cladograma retangular, gerada no programa Clustal, aproxima as proteínas e não os organismos, ou seja, quanto maior for a pontuação de uma proteína maior sua identidade em relação à outra proteína, e a árvore mostra isso através dos níveis de ramificação. A relação filogenética entre as proteínas dos grupos A (azul) e B (vermelho) está apresentada na figura 31. A conservação dos aminoácidos é o fator mais impactante na construção da árvore e alinhamento (SIEVERS et al., 2011).

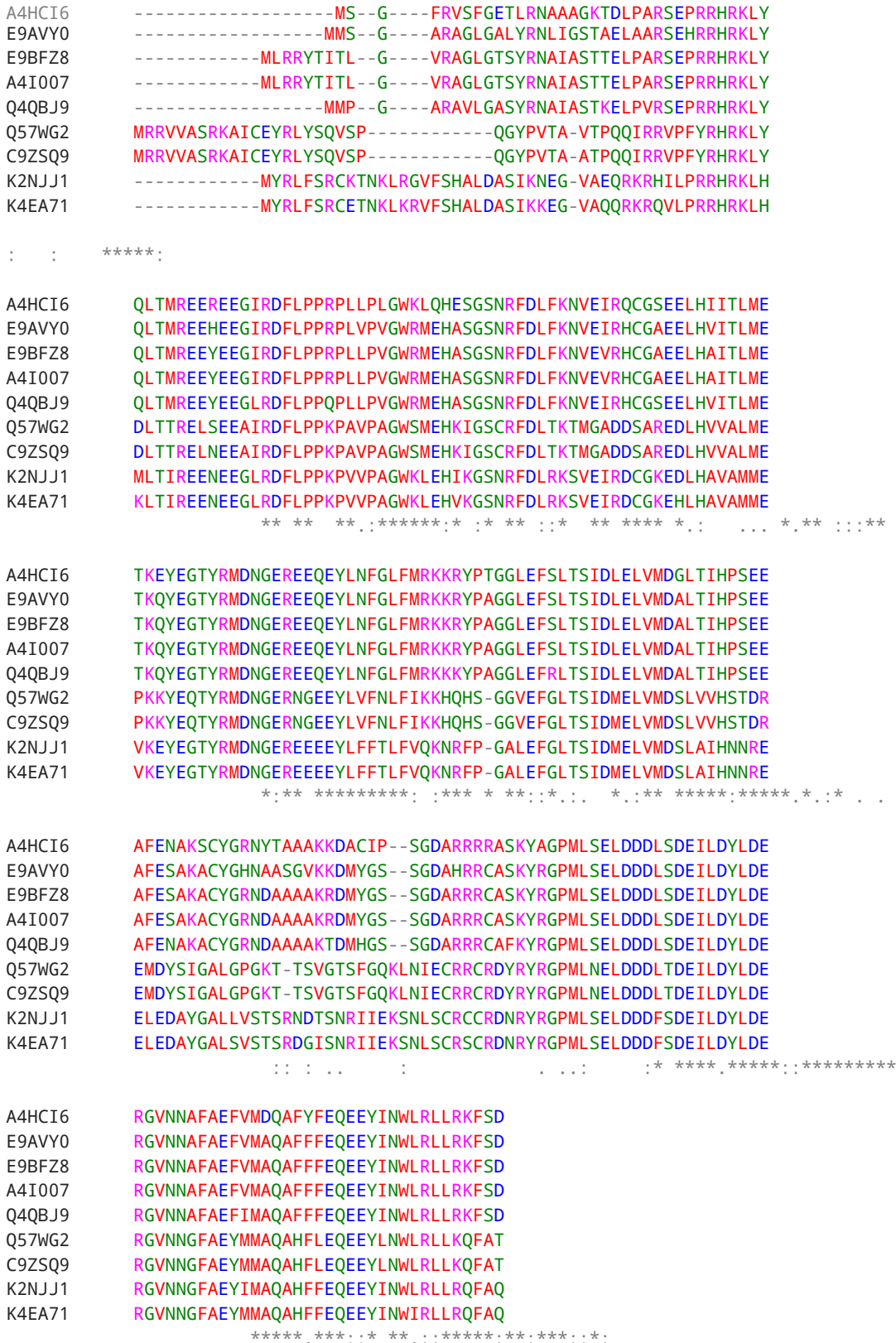


FIGURA 29. Alinhamento de sequencias das 9 proteínas pertencentes ao grupo A e comuns às bases de dados do NR, TriTrypDB e GenBank que são homólogas a C1qBP e que contém MAM33 e mesmo direcionamento para mitocôndria. O alinhamento foi feito no Clustal Omega (*) Aminoácidos idênticos; (:) Aminoácidos de alta similaridade; (.) Aminoácidos de baixa similaridade. Os aminoácidos estão coloridos de acordo com a classificação do grupo R. FONTE: GOMES (2014)

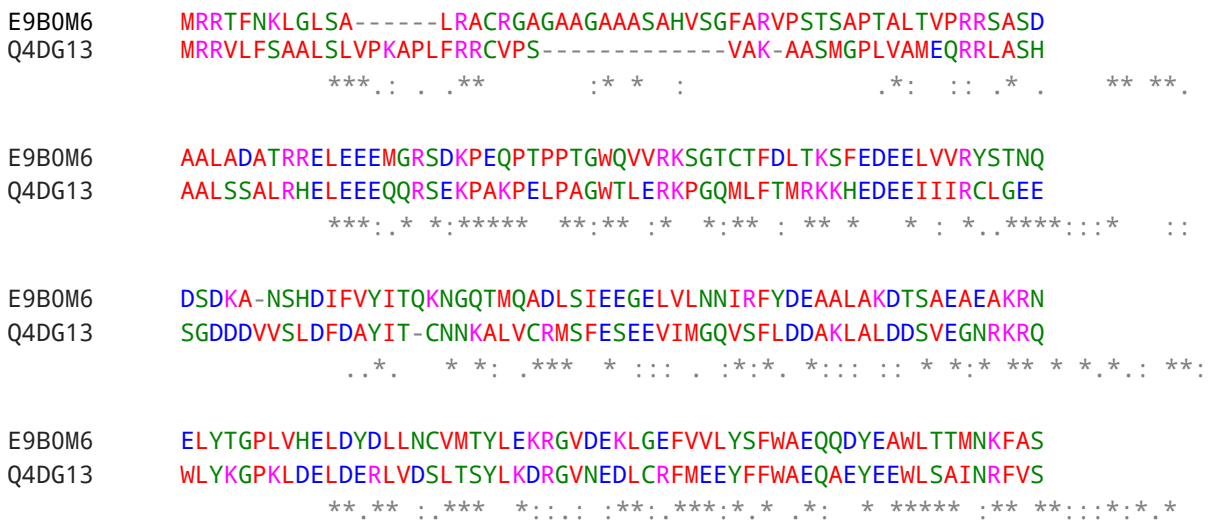


FIGURA 30: Alinhamento de seqüências das 2 proteínas pertencentes ao grupo B e comuns as bases de dados do NR, TriTrypDB e GenBank que são homólogas a C1qBP e que contém MAM33 e mesmo direcionamento para mitocôndria. O alinhamento foi feito no Clustal Omega (*) Aminoácidos idênticos; (:) Aminoácidos de alta similaridade; (.) Aminoácidos de baixa similaridade. Os aminoácidos estão coloridos de acordo com a classificação do grupo R. FONTE: GOMES (2014)

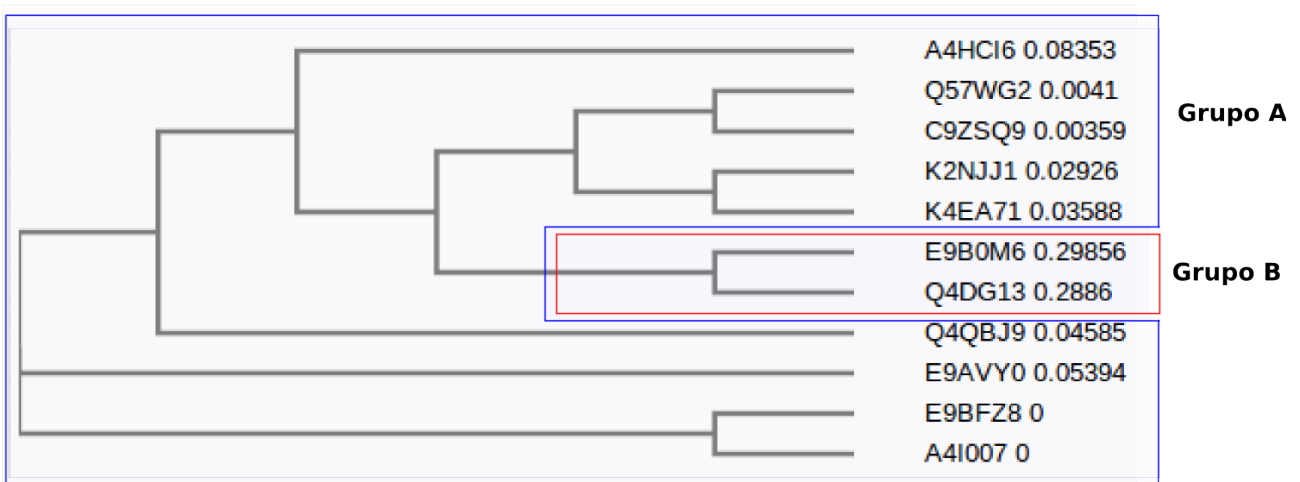


FIGURA 31: Filograma das 11 proteínas (QUADRO 9) apresentando a localização dos dois grupos. A pontuação apresentada após o código Entry de identificação de cada proteína mostra a aproximação entre elas e somado com as análises de alinhamento realizada no item 3.6.2, formaram-se os grupos A e B conforme as figuras 29 e 30 FONTE: GOMES (2014)

3.7 MODELAGEM COMPARATIVA DAS PROTEÍNAS C1QBP

O principal componente do complexo C1, da via clássica do sistema complemento, é a glicoproteína gC1q que é responsável pela ativação proteolítica intramolecular do tetrâmero C1r(2)C1s(2). Além disso, gC1q pode desencadear respostas célula-específicas ao ligar-se, por exemplo, a monócitos, linfócitos, células endoteliais e plaquetas (GHEBREHIWET; PEERSCHKE, 2004). Estas respostas são mediadas pela interação de gC1q com proteínas ligantes ou receptores na superfície celular. Dentre estas está a proteína C1qBP (ou gC1qBP ou gC1q-R) que se liga a cabeça globular de gC1q (LIM BL, REID KBM, GHEBREHIWET B, 1996). Proteínas C1qBP-like estão amplamente distribuídas entre os seres vivos e expressas em vários tipos de células e tecidos que estão envolvidas em diversas respostas celulares mediadas por ligação a receptores (GHEBREHIWET *et al.*, 2001). A proteína que desempenha esta função em seres humanos é a C1QBP_HUMAN (Q07021). Com relação à estrutura, as C1qBPs constituem um grupo de proteínas com enovelamento característico, inicialmente descritas como sendo compostas de sete fitas betas antiparalelas, flanqueadas por uma hélice N-terminal (α A) e duas hélices C-terminais (α B e α C) (JIANG *et al.*, 1999). Porém pequenas variações nesse enovelamento foram observadas com a resolução de novas estruturas (SPREHE *et al.*, 2010; PU *et al.*, 2011). As hélices α A e α B são importantes para os contatos entre subunidades que formam o homotrímero em forma de rosca (unidade biológica, relacionada por um eixo de simetria de ordem 3). Essa estrutura quaternária apresenta um acúmulo de carga negativa em uma das faces da rosca e, supostamente, é importante para a interação com outros complexos proteicos. É sabido que a C1QBP_HUMAN interage com o receptor gC1q o qual também é uma estrutura homotrimérica de mesma simetria. Este por sua vez, possui um acúmulo de carga positiva em uma das extremidades, e é complementar a distribuição de cargas negativas na C1qBP humana. Isso sugere que interação entre essas duas estruturas oligoméricas é predominantemente de natureza eletrostática. A FIGURA 32 mostra a distribuição de carga calculada para a superfície da proteína gC1q (código PDB – 4f3j) e da estrutura da C1QBP_HUMAN (código PDB – 1p32).

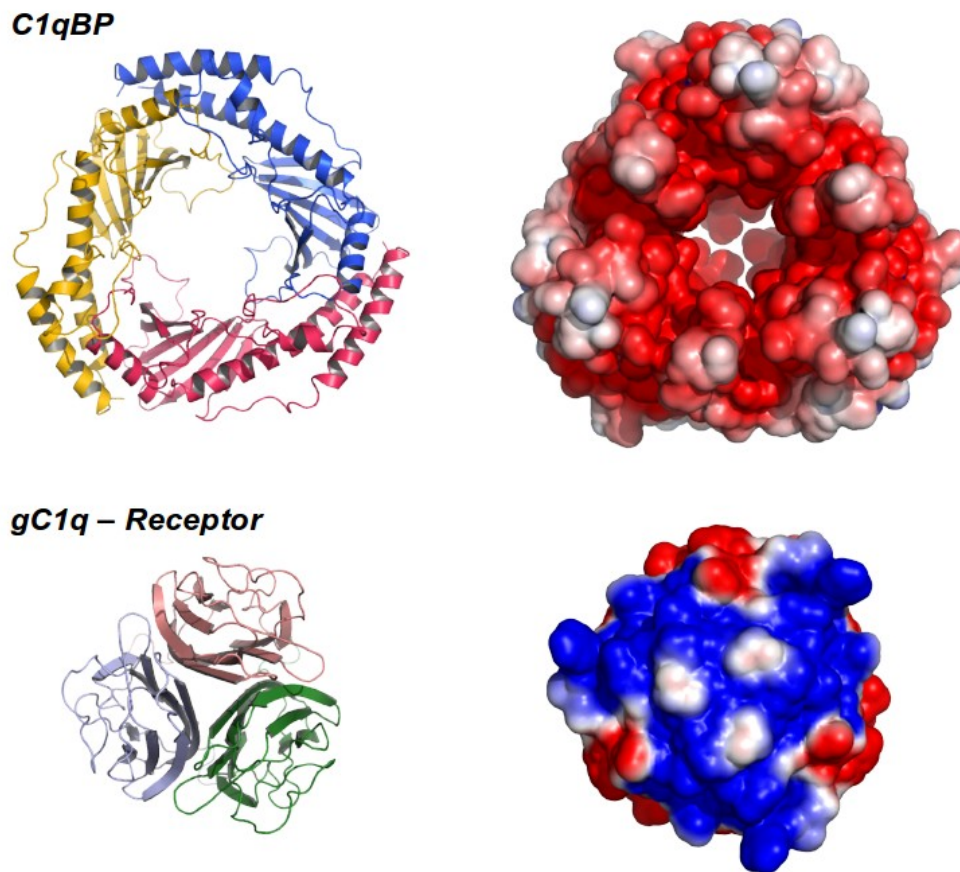


FIGURA 32: Estruturas triméricas da C1qBP_HUMAN (código PDB – 1p32) e da gC1q (código PDB – 4f3j j). Os modelos são representados na forma de *cartoon* evidenciando os elementos de estruturas secundária e também na forma de superfície. As cores vermelho/azul representam acúmulo de carga negativa/positiva em uma dada região, mostrando que existe uma complementariedade de cargas entre as duas estruturas oligoméricas.

FONTE: BACHEGA (2014)

Atualmente, estão depositadas no banco de dados PDB 5 estruturas do tipo C1qBP que atendem ao enovelamento descrito anteriormente, são elas: duas referentes à C1qBP_HUMAN (1P32 e 3RPX), uma referente a proteína p22 de *T. brucei* (3JV1), uma referente à proteína hipotética de *L. major*, homóloga a p32 humana (1YQF), e por fim, a estrutura do Mam33 de *Saccharomyces cerevisiae* (3QV0). Embora existam estruturas cristalográficas da C1qBP humana disponíveis estas apresentam regiões não modeladas que limitam a interpretação do modelo. Uma vez que existam estruturas de homólogos próximos, a modelagem comparativa, ou modelagem por homologia como também é chamada, é uma alternativa para a obtenção de informações estruturais que permitam inferir a interpretação do comportamento biológico/molecular desses sistemas. Neste trabalho a modelagem comparativa foi utilizada para a obtenção dos modelos das proteínas Q2DG13, K2NJJ1, e também para remodelar a C1qBP_HUMAN, predizendo as regiões ausentes na estrutura cristalográfica. Um quarto modelo foi gerado para a

proteína Hq412588 de *Penaeus monodon*, uma espécie de camarão. Este modelo foi gerado com base no trabalho de Yang e colaboradores (2013) que aponta esta proteína como formadora de um complexo com as gC1q. A TABELA 1 mostra uma comparação para as identidades calculadas entre as estruturas disponíveis no banco de dados PDB e as sequências de interesse deste trabalho. Esta etapa do trabalho teve a colaboração do Dr. Jose Fernando Ruggiero Bachega, pós-doutorando no Programa de pós-graduação em Bioinformática (PNPD CAPES).

	1P32 - Humana	3JV1 -<i>T. brucei</i>	1YQF - <i>L. major</i>	3QV0 - <i>S. cerevisiae</i>
C1qBP_HUMAN	100.0	20.9	22.2	22.5
Q2DG13	22.9	49.4	41.2	35.6
K2NJJ1	15.1	34.6	27.2	21.6
Hq412588	33.3	20.5	17.2	23.4

TABELA 1: Comparação entre as sequências modeladas neste trabalho e seus modelos de referência. Em negrito estão evidenciadas as estruturas de modelos escolhidas pelo servidor SWISS-MODEL. As proteínas C1qBP_HUMAN e Hq412588 foram modeladas com base na estrutura humana (1P32), e Q2DG13 e K2NJJ1 foram modeladas baseadas na estrutura da proteína p22 de *T. brucei* (3JV1).

FONTE: BACHEGA (2014)

A escolha da estrutura molde é feita pelo próprio servidor Swiss-Model levando-se em consideração a cobertura e a identidade entre sequência da estrutura molde e sequência a ser modelada. Dessa forma, para as proteínas C1qBP_HUMAN e Hq412588, a estrutura humana (1P32) foi escolhida como molde, com uma cobertura de 90% e 75% respectivamente. A FIGURA 33 mostra o alinhamento entre C1qBP_HUMAN e Hq412588 onde estão representados elementos de estrutura secundária e regiões ausentes no modelo cristalográfico.

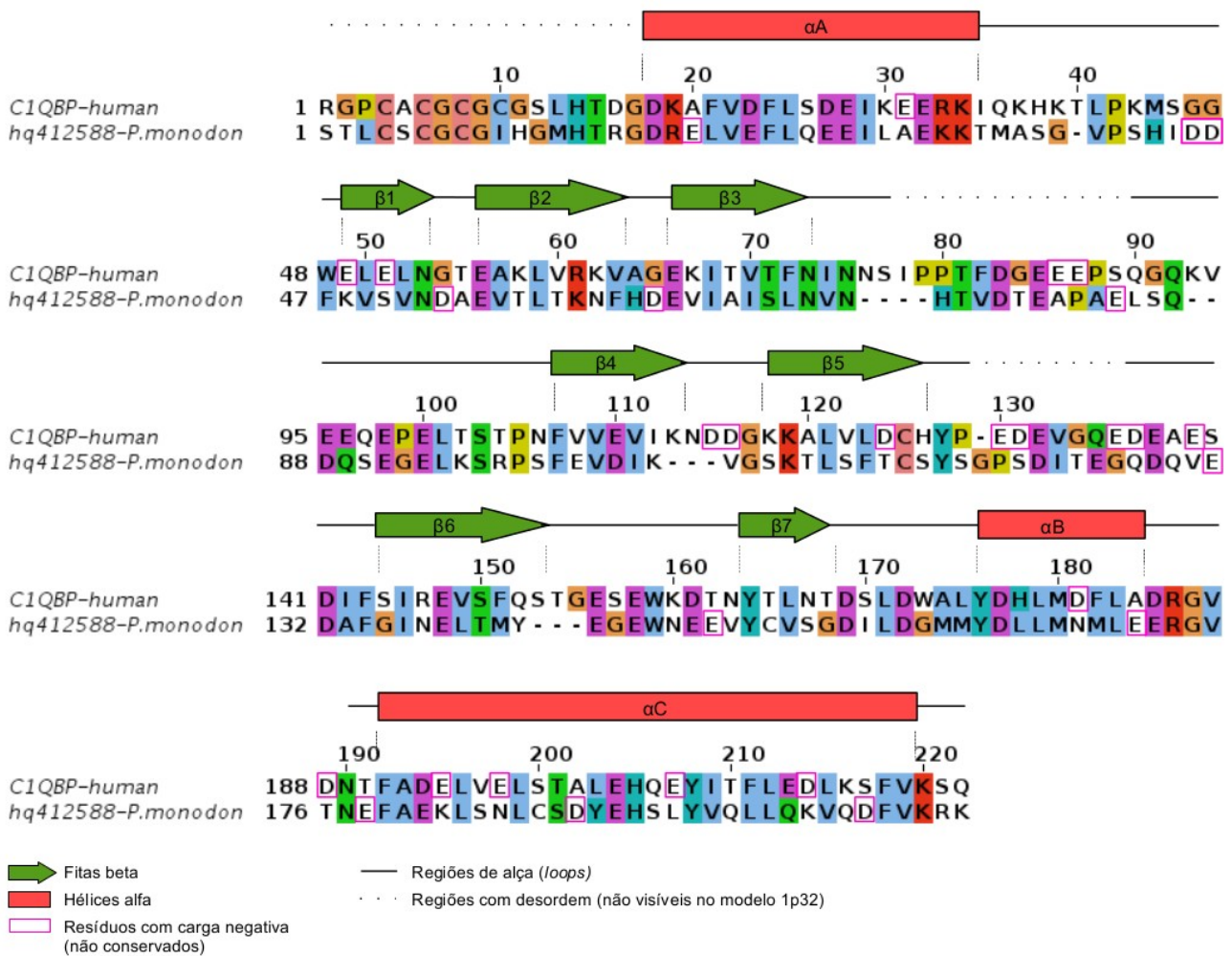


FIGURA 33: Alinhamento de aminoácidos de C1qBP _ HUMAN e Hq412588. Estão representados elementos de estrutura secundária e regiões ausentes no modelo cristalográfico (1p32). FONTE: BACHEGA (2014)

As sequências Q2DG13 e K2NJJ1 foram modeladas baseadas na estrutura da proteína p22 de *T. brucei* (3JV1), com cobertura 99% e 70% respectivamente. A FIGURA 34 mostra o alinhamento entre sequências de aminoácidos das proteínas selecionadas em *T. cruzi* e a estrutura molde de *T. brucei*. Observa-se que a topologia para as proteínas de *T. cruzi*/*T. brucei* é distinta da descrita para a estrutura humana. Nos tripanossomatídeos não é observada a última estrutura tipo beta (fita beta 7), todavia, esta região é substituída pela inserção das duas pequenas hélices, nominadas aqui como $\alpha A2$ e $\alpha A3$, e que antecedem as hélices αB e αC .

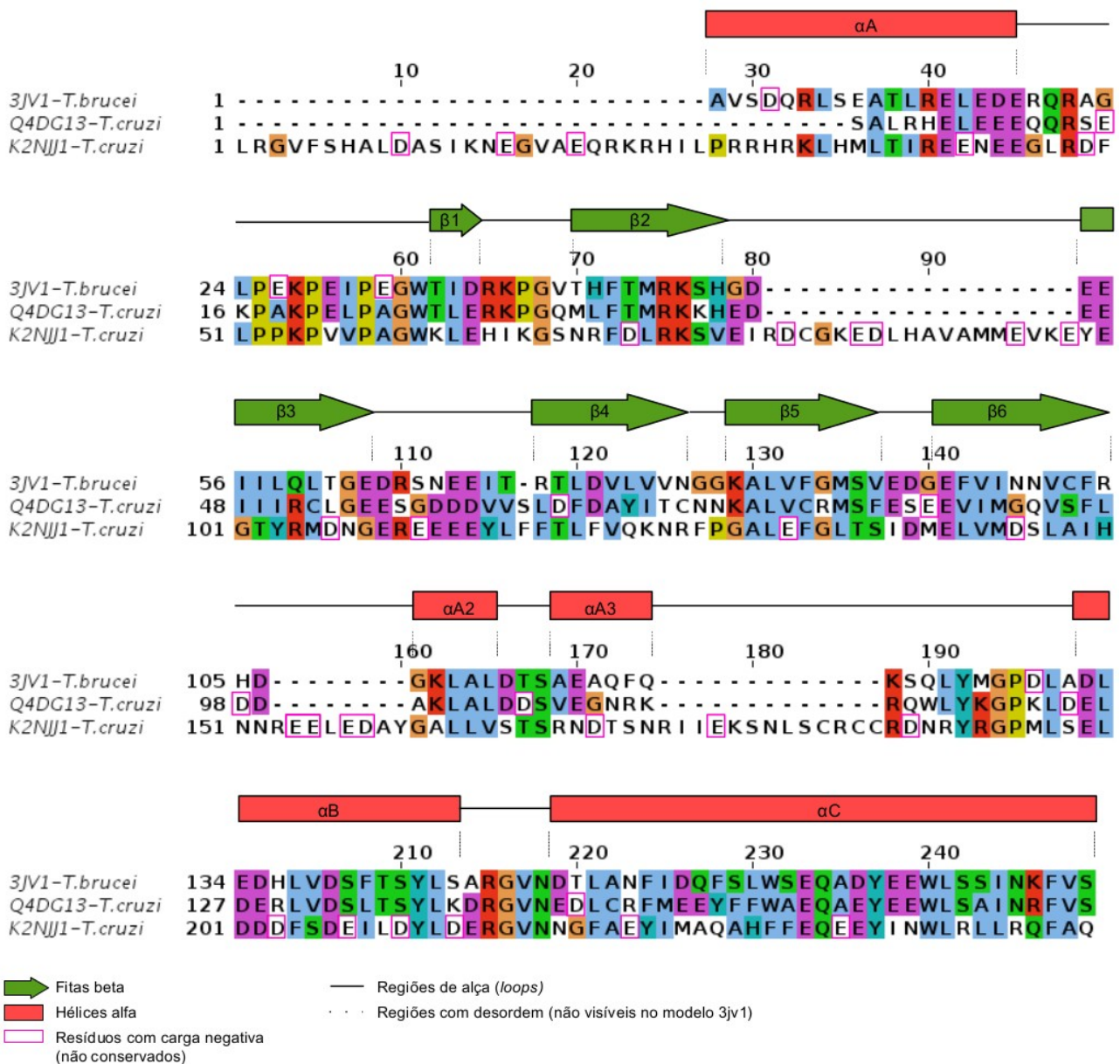


FIGURA 34: Alinhamento entre sequências das proteínas selecionadas em *T. cruzi* e a estrutura molde de *T. brucei*. Note que a topologia para as proteínas de *T. cruzi*/*T. brucei* é distinta da descrita para a estrutura humana. Nos tripanossomatídeos última estrutura tipo beta (beta 7) é substituída pela inserção das duas pequenas hélices, nominadas aqui como $\alpha A2$ e $\alpha A3$, e que antecedem as hélices αB e αC .

FONTE: BACHEGA (2014)

Assim como esperado, os modelos comparativos obtidos são bastante semelhantes, atendendo aos elementos básicos de estrutura secundária que compreendem as C1qBP. O *loop* (alça) que conecta as fitas beta3/beta4 é a região de maior variabilidade entre as proteínas em questão e também a região com menor confiabilidade nos modelos obtidos. Uma estimativa da desordem com base nas

sequências das proteínas C1qBP_HUMAN, Hq412588 e K2NJJ1, mostra que o *loop* que conecta as fitas beta3/beta4 é desordenado para os três casos. Fato que se reflete na estrutura cristalográfica da 1P32, onde esta região não é visível em virtude da desordem local. Nos modelos por comparação obtidos para a C1qBP_HUMAN, Hq412588 e K2NJJ1, esta região foi modelada como *loops* protuberantes (nenhum procedimento extra para a otimização de *loops* foi utilizado). Nas estruturas cristalográficas 3JV1 e 1YQF, o *loop* conectando as fitas beta3/beta4 é uma estrutura visível, de forma que para a proteína Q2DG13 foi possível modelar tal região com melhor precisão. Neste caso, trata-se de um *loop* menor do que o descrito para as sequências C1qBP_HUMAN, Hq412588 e K2NJJ1, porém, em todos os casos, a região em questão conta com a presença de resíduos negativamente carregados, o que contribui para o acúmulo de carga negativa em uma das faces da estrutura em forma de rosca que o homotrímico assume. A FIGURA 35 mostra o modelo comparativo obtido para Q2DG13 e a topologia que essa estrutura

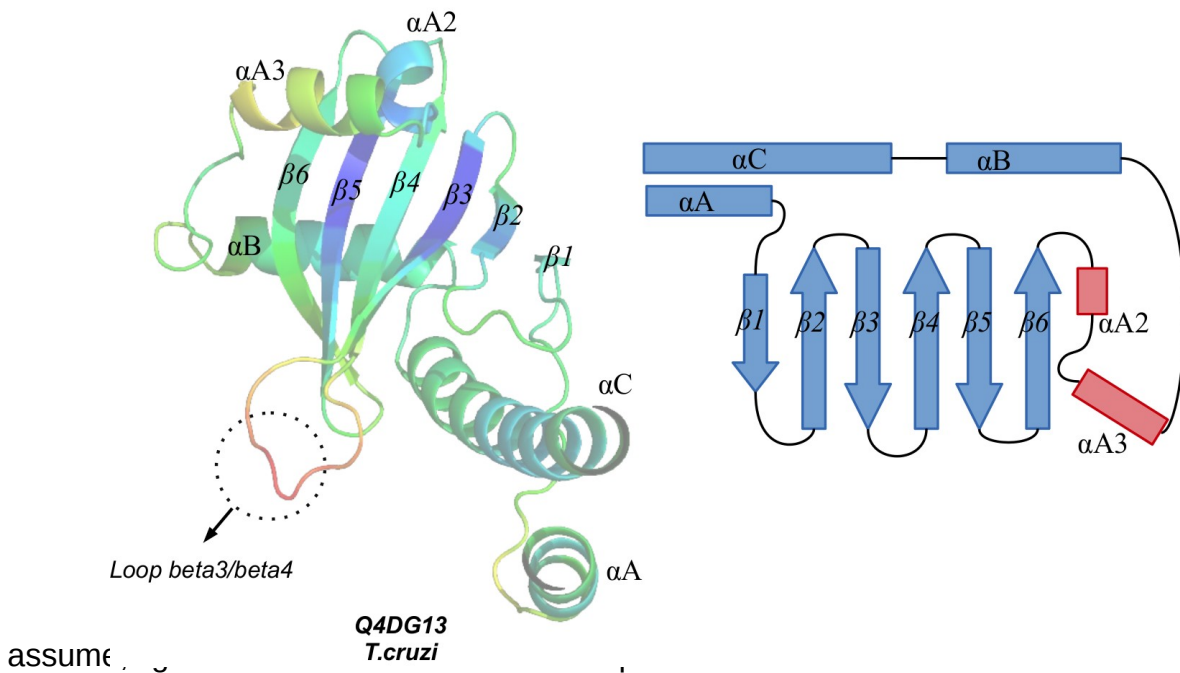


FIGURA 35: Estrutura de Q2DG13 modelada e a topologia que as proteínas Q4DG13, 3JV1 e 1YQF assumem. Na representação em *cartoon*, a escala de cores representa uma estimativa da desordem local no modelo (azul – baixa desordem, vermelho – alta desordem). Embora mais curto, os *loops* que conectam as fitas beta3/beta5 também são regiões de flexíveis quando comparadas ao resto do modelo.

FONTE: BACHEGA (2014)

Uma comparação entre os modelos obtidos e as estruturas cristalográficas de

interesse neste trabalho esta apresentada na FIGURA 36. Uma escala de cores é utilizada para descrever a confiabilidade do modelo (Qmean) gerado para cada região (no caso de C1qBP_HUMAN, Q2DG13, K2NJJ1 e Hq412588) ou o fator de temperatura (*b-factor*) para as estruturas 3JV1 e 1YQF. Nos dois casos a escala de cores pode ser interpretada como uma medida da desordem local do modelo azul, baixa desordem, vermelho – alta desordem.

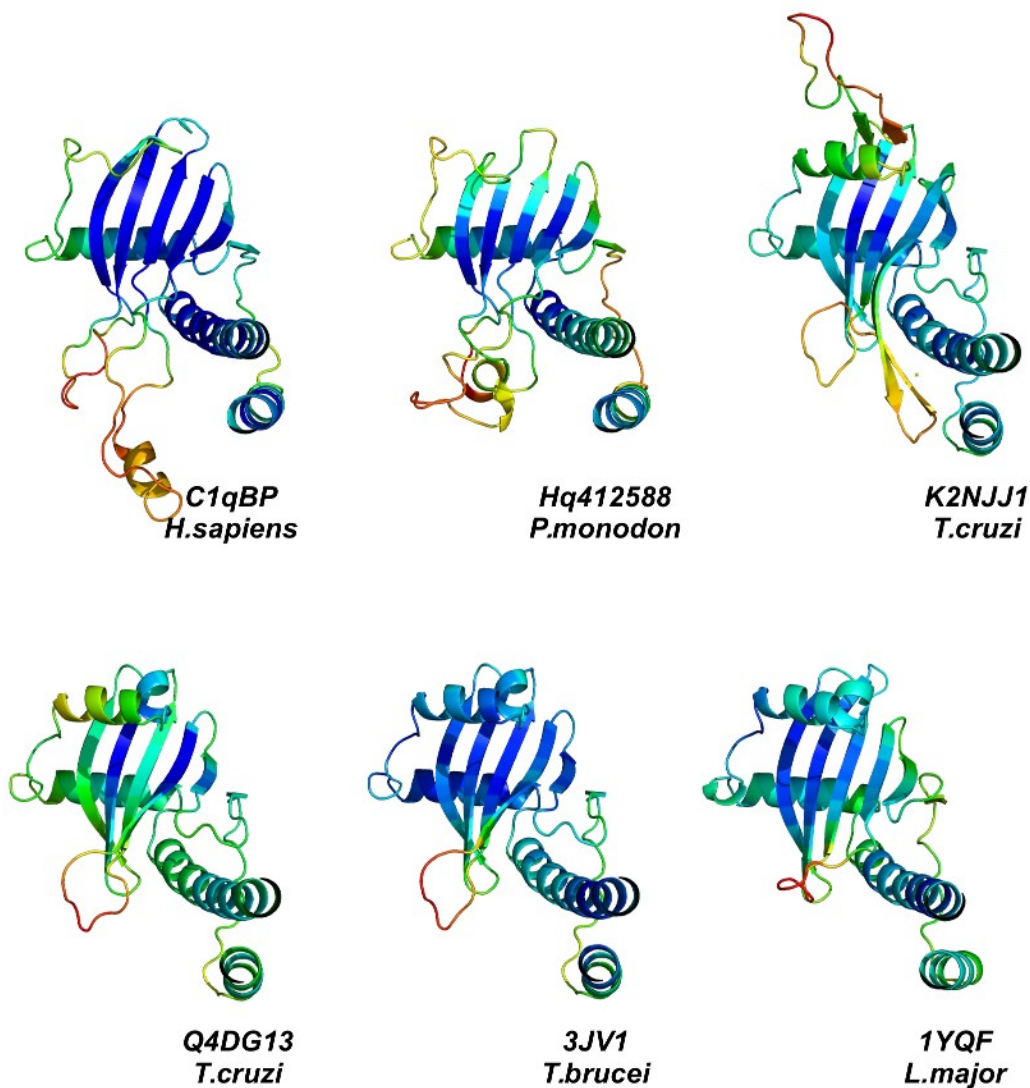


FIGURA 36: Estruturas dos monômeros das proteínas tipo C1qBP modeladas e disponíveis no banco de dados PDB. As estruturas foram espacialmente alinhadas e visualizadas a partir de uma mesma orientação. Na representação em *cartoon*, a escala de cores representa uma estimativa da desordem local no modelo (azul – baixa desordem, vermelho – alta desordem). Os *loops* que conectam as fitas beta3 e beta4 são as regiões de maior desordem nos modelos.

FONTE: BACHEGA (2014)

Embora a modelagem comparativa seja limitada na predição acurada da estrutura

dos *loops* (devido à falta de modelos de referência, e pela própria natureza dos *loops* – desordenados) os modelos comparativos obtidos são razoáveis para prever onde ocorre o acúmulo de carga na superfície dessas estruturas. Uma comparação entre os trímeros (unidade biológica) das C1qBPs de interesse neste trabalho mostra que o acúmulo de carga negativa em uma das faces da estrutura em forma de rosca é conservado em todas elas, sugerindo que a complementariedade de carga exigida na interação com o receptor é satisfeita. A FIGURA 37 mostra a comparação da distribuição de carga na superfície dos homotrímeros gerados para as proteínas de interesse neste trabalho. Não existe uma conservação na posição dos resíduos negativamente carregados (ver alinhamentos), porém em todos os casos as cargas negativas se manifestam predominantemente em uma das faces da estrutura de rosca, reforçando a ideia de que as interações eletrostáticas têm um grande peso na formação dos complexos com os receptores gC1q, tornando remota a possibilidade de que os contatos poderiam ocorrer de outra forma, como por exemplo, por meio de uma interface hidrofóbica. Todavia, a variação nas sequências, tamanho e mobilidade dos *loops*, podem ser fatores importantes para a especificidade e afinidade com que as C1qBP-like ligam-se ao receptor.

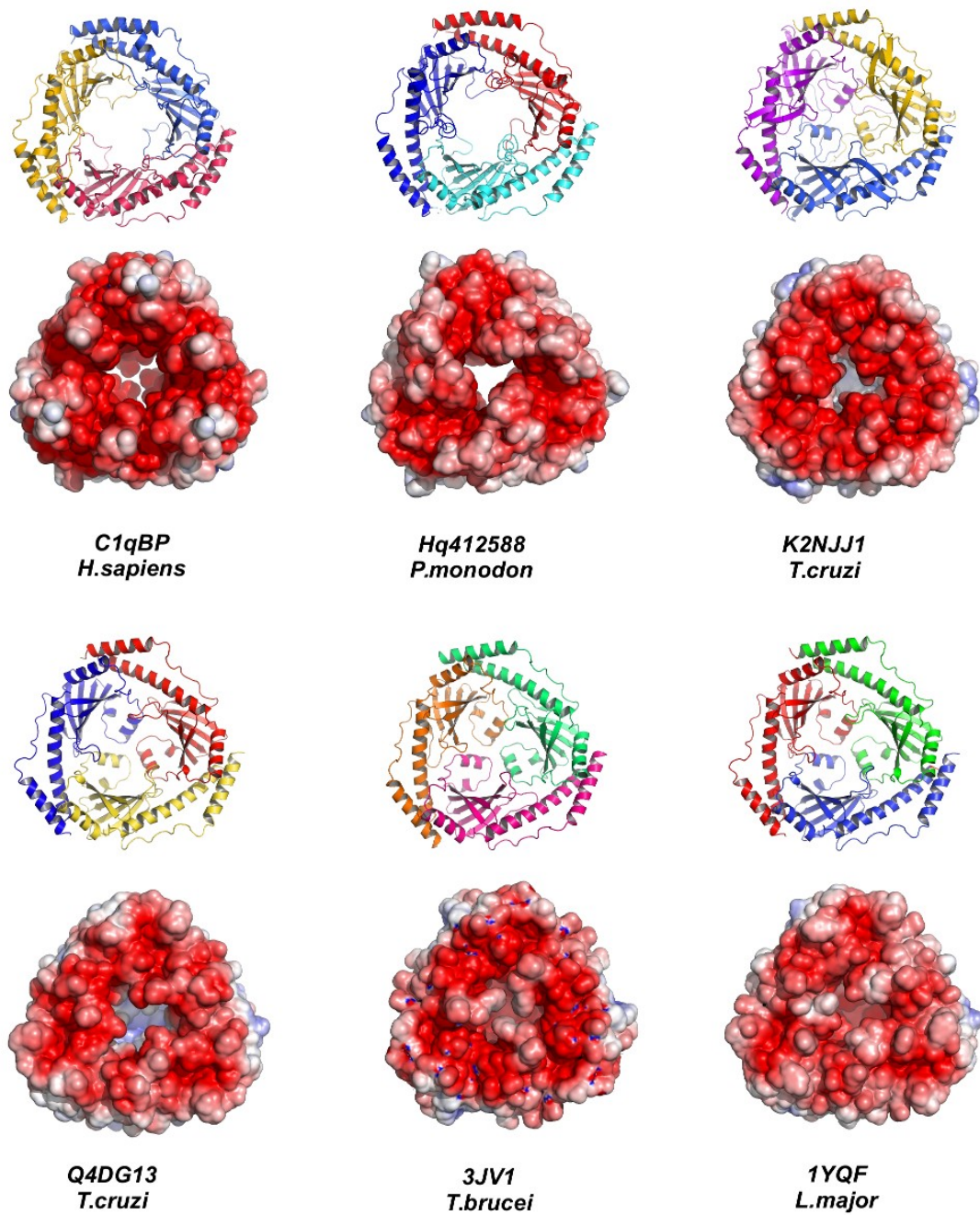


FIGURA 37: Comparação entre os homotrimeros das C1qBPs mostrando que o acúmulo de carga negativa em uma das faces da estrutura em forma de rosca é conservado em todas elas.
FONTE: BACHEGA (2014)

4 DISCUSSÃO

4.1 RECURSOS DE BIOINFORMÁTICA UTILIZADOS

4.1.1 GERAÇÃO DE BANCO DE DADOS

O primeiro passo desse trabalho foi o desenvolvimento de um banco de dados para a inclusão de dois grupos de proteínas que foram utilizados em nosso estudo; um grupo com informações de proteínas relacionadas ao sistema complemento e um segundo grupo com informações de proteínas pertencentes ao genoma de tripanossomatídeos. As proteínas relacionadas ao sistema complemento foram obtidas da base de dados UniProtKB, um banco de dados proteico que contém dados curados de proteínas com alto padrão de qualidade. Já, para as proteínas relacionadas aos tripanossomatídeos, foram utilizadas as bases de dados proteicas NR, TryTripDB e GenBank. Incluímos estas três bases devido a: NR é uma base de dados onde estão contidas informações de proteínas de inúmeros organismos que são atualizadas constantemente e apresentam apenas as últimas versões das informações referentes a cada proteína, dando desta forma, um caráter não redundante, TryTripDB é um banco de dados genômico específico de cinetoplastídeos; e GenBank, um dos principais banco de dados genômicos que inclui organismos eucariotos e procariotos, onde estão contidas diversas versões de informações de genes não curadas e que nos permitiu realizar uma análise de alinhamento local, agregando mais confiabilidade nos resultados atingidos com os bancos de dados anteriores.

Os dois grupos de informações das proteínas relacionadas ao sistema complemento e das proteínas de tripanossomatídeos foram representados por tabelas geradas de forma estruturada no banco de dados, proporcionando uma melhor organização dos dados e garantindo a interação entre os mesmos. Este banco de dados foi desenvolvido no sistema gerenciador de banco de dados (SGBD) MySQL, uma ferramenta gratuita, de boa performance para a execução em pequenos a grandes computadores e que possui código aberto, permitindo que o usuário acrescente novas funcionalidades ou corrija erros encontrados.

4.1.2 FERRAMENTAS DE BIOINFORMÁTICA UTILIZADAS

4.1.2.1 FERRAMENTAS DE ALINHAMENTO

Os programas Blastp e PSI-Blast foram utilizados para a realização dos confrontos de alinhamentos entre as sequências de aminoácidos das proteínas relacionadas ao sistema complemento e as proteínas dos tripanossomatídeos. A diferente especificidade entre estes dois programas está em que, o Blastp retorna o percentual de similaridade entre as proteínas confrontadas, enquanto o programa PSI-Blast, utilizando do Blastp e da *Position-Specific Scoring Matrix* (PSSM) retorna, através de níveis de iterações, a aproximação entre as proteínas pesquisadas (FIGURA 7). Para o processamento da PSSM o PSI-Blast utiliza o modelo oculto de Markov (HMM), que se trata de um processamento matemático realizado sobre os dados de forma transparente e com alto desempenho, garantindo mais agilidade no processamento de um grande volume de dados.

A maioria dos trabalhos realizados em bioinformática utilizam ferramentas de similaridade entre alinhamentos para determinar a presença de determinados(s) gene(s) e/ou a identidade de gene(s) e proteína(s). Os valores resultantes destas ferramentas são expressos através de percentuais de similaridade e de um valor conhecido como *e-value*. Em nossa estratégia de trabalho, utilizamos destas ferramentas apenas o critério de detecção de proteínas, independente do seu percentual de similaridade ou valor de *e-value*. Nosso interesse era selecionar a maior quantidade de proteínas e posteriormente avaliar as características biológicas como, domínio funcional e localização celular, para então considerá-las ortólogas às proteínas do sistema complemento.

4.1.2.2 FERRAMENTAS DE LOCALIZAÇÃO DE DOMÍNIO

O programa Pfam foi utilizado para a definição de domínios funcionais das proteínas selecionadas das bases de dados proteicas. A análise de domínio funcional é realizada pelo confronto de uma proteína contra uma família proteica. Uma família

proteica é um conjunto de sequências de proteínas, que realizam a mesma função. A busca do domínio funcional de uma proteína se diferencia radicalmente do alinhamento entre proteínas realizado pelo Blastp, pois enquanto o Blastp confronta duas sequências de aminoácidos já definidas, o Pfam compara uma proteína contra um padrão definido entre várias proteínas pertencente uma determinada família. Esse padrão de comparação, criado através das proteínas pertencentes a uma determinada família é construído por HMM (*hidden Markov model*).

A relação entre duas proteínas que possuem o mesmo domínio funcional pode ser muito mais relevante do que a de proteínas com alto percentual de similaridade e/ou bom valor de *e-value* entre elas, porém sem o mesmo domínio funcional (HAUBOLD; WIEHE, 2004; PARSONS *et al.*, 2011).

Ao submeter as proteínas selecionadas das três bases de dados no programa Pfam, observamos que a maioria destas proteínas apresentaram o domínio MAM33 (FIGURAS 9, 17 e 21), domínio este que também está presente na proteína C1QBP_HUMAN. Os outros domínios apresentados não tinham relação com o sistema complemento.

O MAM33 é um domínio constituído por 204 aminoácidos. Nas proteínas homólogas a C1QBP_HUMAN, selecionadas nesse trabalho, o domínio MAM33 variou entre 58 a 202 aminoácidos (FIGURA 10), para a menor e maior sequência encontrada, estando este domínio sempre localizado na região C-terminal. Encontramos também em outros patógenos, proteínas homólogas a C1QBP_HUMAN que apresentam o domínio MAM33 (FIGURA 24 e 25), o que indica que a presença do gene que produz proteínas homólogas a C1QBP_HUMAN, seja um evento frequente em eucariotos.

Os diferentes tamanhos do domínio MAM33 das proteínas selecionadas não garantem se a proteína realiza a função referente ao domínio. Apenas através de uma demonstração *in vitro* ou *in vivo* é possível confirmar a função da proteína.

4.1.2.3 FERRAMENTAS DE LOCALIZAÇÃO PROTEICA

Para determinarmos a localização celular das proteínas obtidas das três bases de dados (FIGURAS 11, 18 e 23), ferramentas de predição de localização celular foram utilizadas: SignalP, TargetP e Pred-GPI. O SignalP tem a capacidade de predizer a

presença de peptídeo sinal e sinal de transmembrana na cadeia polipeptídica da proteína. O TargetP prediz peptídeos que direcionam as proteínas para organelas, tais como a mitocôndria, o cloroplasto ou o retículo endoplasmático. Para a análise de peptídeo sinal, o TargetP utiliza internamente o programa SignalP. O Pred-GPI realiza a predição de ancora de glicosilfosfatidil inositol (ancora-gpi) nas proteínas, uma estrutura que insere a proteína na membrana plasmática. Os três programas utilizam técnicas de inteligência artificial, treinados cada um, por um banco de sequências de proteínas que definem o modelo de comparação da aplicação. Porém, o SignalP e TargetP que são fomentados pela CBS, utilizam estratégias de redes neurais enquanto Pred-GPI estratégias de *Support Vector Machine* (SVN). Os programas SignalP e TargetP também utilizam de HMM para manipular os dados analisados através de suas redes neurais. Proteínas que possuem o mesmo domínio e localização celular garantem ainda mais a possibilidade de executarem a mesma função no mesmo local no organismo (SILBER; PEREIRA, 2012).

O interesse em detectar a localização celular das proteínas foi apenas para identificar se elas eram direcionadas para a mitocôndria, como C1QBP_HUMAN. Outras proteínas direcionadas para outros locais não foram descartadas, pois também possuem a mesma função que C1QBP_HUMAN.

4.1.2.4 FERRAMENTAS DE ALINHAMENTO DE SEQUÊNCIAS

O Clustal Omega é uma ferramenta específica para realizar alinhamento entre sequência de aminoácidos de proteínas, apresenta o resultado entre o pareamento de sequências de forma clara e bastante representativa (exemplo FIGURA 28). Pontua os aminoácidos de alinhamento por conservação, alta propriedade de similaridade ou baixa propriedade de similaridade. O alinhamento do Clustal Omega diverge do alinhamento do Blastp, pois enquanto o Clustal realiza seu alinhamento de forma global, o Blastp realiza diversos alinhamentos entre duas proteínas, por trechos de melhor similaridade, onde o mesmo trecho pode ser comparado várias vezes com outros trechos de outra proteína.

Em nossas análises, os resultados do Clustal Omega juntamente com os resultados do PSI-Blast permitiram a criação de agrupamentos de 18 proteínas baseados no tamanho dos motivos e a iteração onde apareceu a proteína confrontada ao banco NR. (FIGURA 13). O clustal nos permitiu construir posteriormente uma árvore filogenética

quando alcançamos o grupo de 11 proteínas obtidas das diferentes bases de dados. (FIGURA 31).

4.2 MINERAÇÃO DE DADOS E PROTEÍNAS DE *Trypanosoma cruzi*

Após a seleção das proteínas homólogas a C1QBP_HUMAN, nas bases de dados NR, TryTripDB e GenBank, e caracterização das mesma quanto ao seu domínio funcional e localização celular, buscamos através de técnica de mineração de dados, quais proteínas se apresentavam nas três bases simultaneamente (FIGURA 26) e que possuíam as mesmas características biológicas (FIGURA 27). Esta análise nos permitiu identificar 11 proteínas que apareciam nas três bases de dados (QUADRO 9).

Dentre as 11 proteínas encontradas na mineração de dados, 3 pertencem ao *T. cruzi* (K2NJJ1, K4EA71 e Q4DG13) e devido a alta similaridade no alinhamento das sequências de K2NJJ1 e K4EA71 (FIGURA 28), apenas as proteínas K2NJJ1 e Q4DG13 foram selecionadas para a etapa de análise de estrutura e foram consideradas candidatas para análise *in vitro*.

4.3 *Trypanosoma cruzi* E SEU GENOMA

O *T. cruzi* é um organismo que possui um genoma diplóide com características interessantes. A quantidade total de DNA é variável entre clones (DVORAK *et al.*, 1982). Curiosamente, foi visto que entre cepas e clones de uma mesma cepa também há variações no conteúdo de DNA (MCDANIEL; DVORAK, 1993). Além disso, vários grupos de pesquisa mostraram através de cariótipo molecular e ensaios de polimorfismo por restrição enzimática (RFLP- *restriction fragment length polymorphism*) que o genoma deste tripanossomatídeo tem a capacidade de trocar, ganhar ou perder o conteúdo gênico com bastante plasticidade (Gibson; Miles, 1986; Engman *et al.*, 1987; Dietrich *et al.*, 1990; Campetella *et al.*, 1992, Cano *et al.*, 1995). Troca genética em tripanossomatídeos já havia sido demonstrada *in vitro* em *T. brucei* (JENNI *et al.*, 1986) e evidências de híbridos naturais também já foram descritas em *Leishmania* (KELLY *et al.*, 1991; BELLI *et al.*, 1994). Experimentos com dois clones de *T. cruzi* que completaram o ciclo de vida inteiro do parasita mostraram que houve fusão de genótipos parentais, perda de alelos e

recombinação homóloga, dando uma forte evidência da existência de troca genética em *T. cruzi* (GAUNT *et al.*, 2003). A propriedade de troca genética e de plasticidade do genoma confere alta diversidade ao parasita. Em um esforço de consenso científico a comunidade em parasitologia adaptou recentemente a classificação de cepas em seis maiores genótipos ou DTUs (*Discrete Typing Units*) TcI-Tc VI (ZINGALES *et al.*, 2009).

Diferente de outros tripanossomatídeos que infectam mamíferos, o *T. cruzi* pode invadir células hospedeiras não fagocíticas (EL-SAYED *et al.*, 2005), um evento que depende necessariamente da mobilização de cálcio celular, assim como visto para as bactérias *Shigella* e *Salmonella*. Entre outras possíveis explicações, foi sugerido que este mecanismo poderia ter sido adquirido pela transferência horizontal de genes ou uma evolução convergente de proteínas de superfície não homólogas de *T. cruzi* (OPPERDOES; MICHELS, 2007; SILVA *et al.*, 2013).

Outro possível mecanismo utilizado pelo *T. cruzi* para evadir à resposta imune do hospedeiro é a mímica. Um exemplo desse mecanismo é a família de genes FL-160 CRP que codifica proteínas capazes de imitar um antígeno do tecido nervoso, estando diretamente envolvida na geração de auto-imunidade em relação ao tecido nervoso por mimetismo molecular (MATHIEU-DAUDÉ *et al.*, 2008). Esse cenário é observado no estágio crônico da doença de Chagas (VOORHIS, VAN, 1993).

Possivelmente, o ponto mais crítico e relevante dos nossos resultados seja entender o motivo da presença de genes homólogos a C1QBP_HUMAN no genoma parasitário. Acima foi discutida a enorme variabilidade genética entre as diferentes cepas e tem sido mostrado até uma grande diversidade nos parasitas no mesmo organismo hospedeiro (LLEWELLYN *et al.*, 2012). Uma possível justificativa para a existência de genes homólogos a C1QBP_HUMAN no genoma do *T. cruzi*, poderia estar relacionado com eventos de recombinação ou troca genética, originada através de genes ancestrais adquiridos por transferência genética horizontal, originando domínios com muita variabilidade de tamanhos, assim como muitos pseudogenes formados pelos eventos genéticos de plasticidade do genoma. Tanto a transferência horizontal genética, quanto a mímica de uma proteína reguladora do complexo C1, são hipóteses de como proteínas homólogas a C1QBP_HUMAN com domínio MAM33 foram introduzidas no genoma de *T. cruzi* e usadas por ele para a evasão da via clássica do sistema complemento.

4.4 ESTRATÉGIA DE EVASÃO DO COMPLEMENTO POR PATÓGENOS

Estratégias de evasão imune são muitas vezes focadas no sistema do complemento que compõe a peça central da imunidade inata e é considerada a primeira linha de defesa contra microrganismos patogênicos. Muitos patógenos parecem ter desenvolvido rotas paralelas para fugir do complemento, e vários mecanismos de evasão são compartilhados não somente entre os membros do mesmo gênero, mas também entre diversos organismos, tais como bactérias, leveduras, vírus, fungos e parasitas (LAMBRIS *et al.*, 2008). Apesar do grande número de proteínas dos patógenos capazes de reconhecer componentes do complemento, seus mecanismos de ação podem ser resumidos a algumas estratégias bem sucedidas: mimetizando ou recrutando reguladores do complemento, a inibição ou modulação por meio da interação direta com proteínas do complemento e inativação por meio de degradação enzimática (LAMBRIS *et al.*, 2008). Além de tais estratégias diretas, existem outras indiretas como a parede celular de bactérias gram-positivas que impede a lise por não permitir a formação do complexo de ataque à membrana, conhecido como MAC (JOINER *et al.*, 1988). Outra estratégia amplamente usada é a aquisição de inibidores do complemento circulantes no plasma para evitar o ataque do complemento (KRAICZY *et al.*, 2006). É uma estratégia comum em bactérias e fungos que também tem sido descrita para vírus e parasitos (INAL, 2004). A deposição eficiente de fatores do complemento na superfície da maioria dos patógenos requer a ativação do complemento. A fim de inibir a via clássica e das lectinas, muitos microrganismos têm desenvolvido a capacidade de se ligar a proteína do complemento C4bP, que é um inibidor chave de fase fluida destas vias. Proteção contra a via alternativa ocorre devido à captura do principal inibidor desta via o fator H e suas proteínas relacionadas enquanto MAC pode ser inibido pela vitronectina. O número de agentes patogênicos capazes de se ligar ou produzir inibidores do complemento está em constante crescimento e pode-se especular que a maioria dos patógenos que sobrevivem em contato com superfícies mucosas e no sangue é capaz de se proteger por este mecanismo. Outra estratégia de evasão do complemento por patógenos que tem se mostrado muito eficaz é a de capturar e inativar C3 que é a molécula central de todas as três vias de ativação do complemento (BLOM *et al.*, 2009).

A degradação dos componentes do complemento em fragmentos menores e não funcionais é a principal tarefa de proteases ativas do complemento. Os patógenos também inativam o complemento por intermédio da ação de proteases que degradam componentes do complemento cruciais como C3 e C5. Assim, um número de proteases

patogênicas como a SepB do *S. pyogenes* cliva C3 que é uma proteína estabilizante da C3 convertase e intensificadora da via alternativa. *S. pyogenes* também produz a peptidase C5a do complemento que cliva C5a e, portanto, impede a sua capacidade de atrair e ativar neutrófilos (TERAO, 2008).

4.5 EVASÃO DO COMPLEMENTO POR TRIPANOSOMATÍDEOS

Existem vários mecanismos distintos de evasão do complemento pelos tripanossomatídeos. O mais comum deles é a expressão de receptores do complemento que inibem a formação da C3 convertase e aqui seria classificada a expressão de C1QBP nos parasitas (CESTARI *et al.*, 2009). Já foi mostrado que as formas tripomastigotas metacíclicas apresentam em sua superfície o receptor CRIT (*Complement C2 receptor inhibitor trispanning*) que interage com o componente C2 não permitindo a formação da C3 convertase (CESTARI *et al.*, 2009). Este receptor é altamente conservado em *Schistosoma sp.* Outra molécula envolvida na inibição do complemento por *T. cruzi* é a calreticulina (CRT) que está localizada na superfície dos tripomastigotas. Ela inibe a via clássica por se ligar a caudas de colágeno presentes na molécula de C1q e interfere na via das lectinas, pois impede o reconhecimento de manose pela (lectina ligadora de manose). Não sabemos em que estágio do parasita a proteína C1QBP estaria sendo expressa, no entanto já foram delineados experimentos para avaliar sua expressão nos diferentes estágios assim como seu possível papel como uma proteína reguladora/inibidora do complemento.

Uma proteína chamada GP160, é encontrada na forma tripomastigota de *T. cruzi* e é conhecida como uma proteína reguladora do complemento. Ela dissocia as C3 convertases da via clássica e da via alternativa por se ligar a C3b e C4b. Outra estratégia empregada pelo *T. cruzi* para evasão do complemento se baseia na estabilização e inibição das C3 convertases por intermédio de microvesículas derivadas da membrana plasmática de células do hospedeiro. Essas microvesículas se ligam a C4bC2a e impedem sua atividade na superfície do parasito (CESTARI *et al.*, 2012).

Recentemente foi mostrado que formas metacíclicas sensíveis ao complemento seriam capazes de usar MBL circulante no plasma para invadir rapidamente as células eucariotas e assim evadir o complemento (EVANS-OSSES *et al.*, 2014).

Quando a proteína C1QBP_HUMAN é expressa ela inibe o sistema complemento e esse mecanismo de inibição é algo que deve ser investigado. Uma hipótese é que ela participe da resposta imunológica adaptativa durante a infecção crônica na doença de chagas e que a lise parasitária provoque a exposição das proteínas mitocondriais C1QBP, que poderiam então bloquear C1. A inibição da ativação da via clássica do complemento permitiria a invasão das células eucarióticas pelas formas tripomastigota, a manutenção a sobrevivência parasitária e o prosseguimento da infecção.

4.6 INIBIÇÃO DE ATIVIDADE POR ESTRUTURA PROTEICA

Tentativas globais para análises e seleção de compostos inibitórios de atividades enzimáticas aumentaram com a disponibilidade de vários genomas completos no final da década de 1990 e se expandiram pela década de 2000, quando foram anotados os genomas de *Leishmania* e de espécies de *Trypanosoma* (EL-SAYED *et al.*, 2005)(BARRETT; CROFT, 2014).

Inúmeras literaturas das décadas 1990 e 2000 estão repletas de descrições de proteínas que possuem sua atividade inibida por uma variedade de moléculas, possuindo algumas destas moléculas, propriedades físicoquímicas com grande potencial para sua utilização como fármacos. A resolução de estruturas de proteínas tornou possível a predição de ligantes em sítios ativos e inúmeros estudos foram realizados para encontrar tais moléculas e utilizá-las na geração/seleção de mais compostos regulatórios ou inibitórios (BARRETT; CROFT, 2014).

Nos tripanossomatídeos, várias enzimas envolvidas no metabolismo de um composto incomum chamado tripanotiona, foram vistas como alvo de drogas úteis, dada sua presença exclusiva nestes protozoários e ausentes no homem (GILBERT, 2013). No entanto, apesar de diversos trabalhos realizados, ainda não foi desenvolvida uma droga eficiente contra estes patógenos (GILBERT, 2013). Triagens de novos alvos, realizadas até recentemente não conseguiram desempenhar todo o seu potencial, com algumas exceções; incluindo novos inibidores de diidrofalato redutases, *topoisomerase* de DNA e *C14- α -esterol demethylases* em fungos (BARRETT; CROFT, 2014). A cadeia de transporte de elétrons, essencial em biosínteses de pirimidina em *Plasmodium*, também foi alvo de compostos da família de quinolonas (BARRETT; CROFT, 2014; STOCKS *et al.*,

2014).

A produção de moléculas regulatórias de complemento poderiam ser alternativas terapêuticas em pacientes chagásicos crônicos se identificados claramente os mecanismos de evasão parasitária utilizados para persistir na infecção.

4.7 *Trypanosoma cruzi* POSSUI C1QBP-LIKE

Os resultados de nosso trabalho indicam que o *T. cruzi* possui proteína homóloga a C1QBP_HUMAN e com domínio MAM33 conservado. C1QBP_HUMAN possui a função de regular o complexo C1 em sua extremidade gC1q através da interação proteína-proteína que envolve o domínio MAM33. gC1q possui a estrutura de uma cabeça globular que interage com o complexo antígeno anticorpo quando este se encontra fixado e ativado na superfície do patógeno (PU *et al.*, 2011), portanto entendemos que o domínio MAM33 das proteínas C1QBP-like concorrem diretamente com o complexo antígeno anticorpo, o que possivelmente inibiria a ativação da via clássica do sistema complemento (FIGURA 38).

O *T. cruzi* também expressa proteínas que possuem o domínio calreticulina. Esse domínio também regula o complexo C1, porém diferentemente do domínio MAM33 da C1QBP, a regulação realizada por calreticulina atinge outra extremidade do complexo C1, inibindo que C1s clive C2 e C4 para a formação de C3 convertase (FIGURA 39) e inibindo a via clássica e a via das lectinas (FERREIRA *et al.*, 2004).

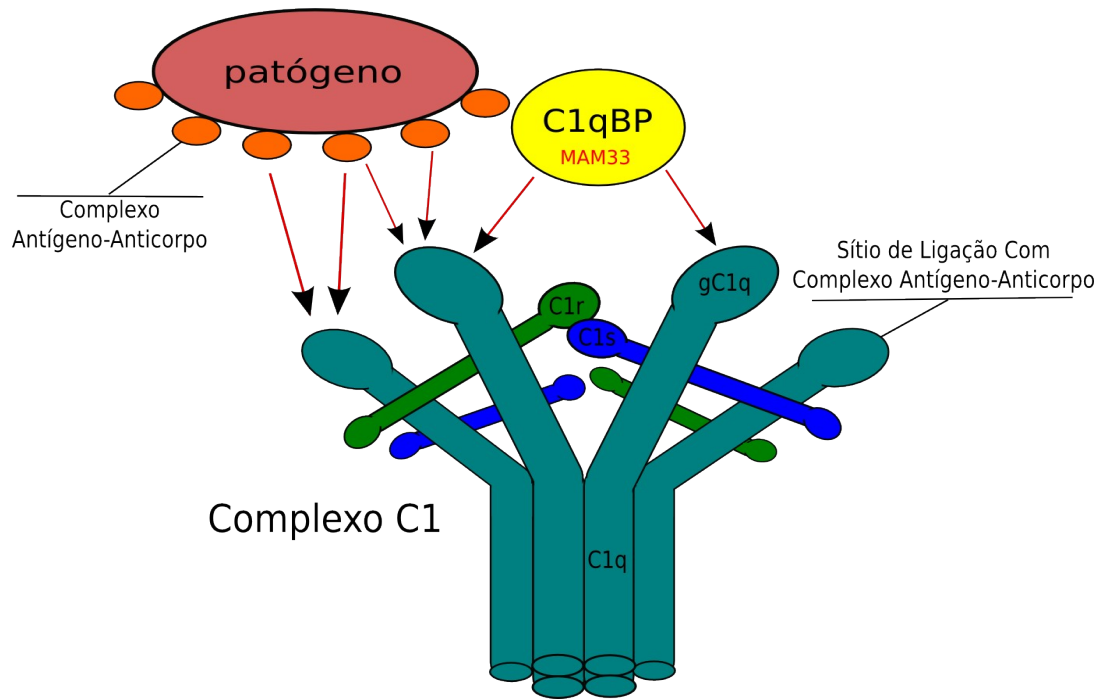


FIGURA 38: Concorrência entre o complexo antígeno anticorpo e o domínio MAM33 de C1QBP para atração de gC1q da proteínas C1q do complexo C1 do sistema complemento. O domínio MAM33 possui características para ligar-se em gC1q, que é por sua vez, o sítio de ligação com o complexo antígeno anticorpo e primeiro componente da formação da via clássica do sistema complemento.

FONTE: GOMES (2014)

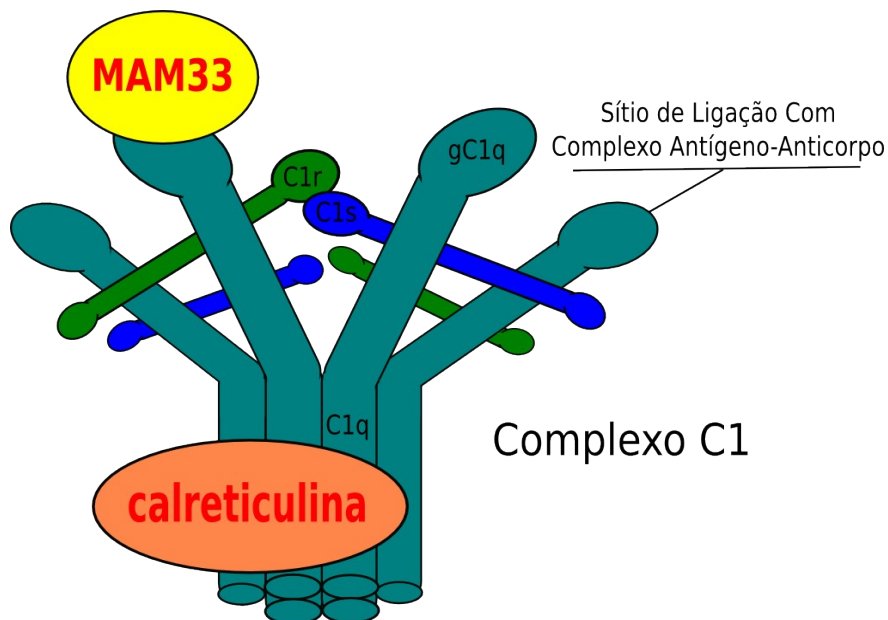


FIGURA 39: Comparação entre os locais de ação dos domínios MAM33 e calreticulina. Os dois domínios são inibidores do complexo C1 da via clássica do sistema complemento, porém enquanto o domínio MAM33 regula o complexo C1 pelo seu sítio de ligação com o complexo antígeno anticorpo, o domínio calreticulina inibe a proteína C1s de clivar C2 e C4 para a formação de C3 convertase.

FONTE: GOMES (2014)

Nossos resultados apresentaram uma abordagem *in silico* para procura de genes de complemento nos genomas de tripanossomatídeos com ênfase em *T. cruzi*. Através de uma estratégia de bioinformática conseguimos identificar um gene alvo que estaria possivelmente associado a um novo mecanismo de evasão a resposta imune do sistema complemento utilizado pelo *T. cruzi*. A relevância do mecanismo assim como a descrição da origem deste gene no genoma, de sua expressão nos diferentes estágios e a possibilidade de alternativas terapêuticas ao mecanismo de evasão são perguntas em aberto, para serem abordadas do ponto de vista genético, biológico e imunológico.

REFERÊNCIAS

ABBAS, A. K; LICHTMAN, A. H; PILLAI, S **Imunologia celular e molecular**. Rio de Janeiro: Elsevier, 2011.

ACKERMANN, A. A; PANUNZI, L. G.; COSENTINO, R. O.; SÁNCHEZ, D. O.; AGÜERO, F. A genomic scale map of genetic diversity in *Trypanosoma cruzi*. **BMC genomics**, v. 13, p. 736, 2012. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3545726&tool=pmcentrez&rendertype=abstract>>. Acesso em: 7/3/2014.

ALTSCHUL, S. F.; MADDEN, T. L.; SCHÄFFER, A A; *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. **Nucleic acids research**, v. 25, n. 17, p. 3389–402, 1997. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=146917&tool=pmcentrez&rendertype=abstract>>. .

ALTSCHUL, S.; GISH, W.; MILLER, W. Basic local alignment search tool. **Journal of molecular ...**, v. 215, p. 403–410, 1990. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022283605803602>>. Acesso em: 23/3/2014.

ASLETT, M.; AURRECOECHEA, C.; BERRIMAN, M.; *et al.* TriTrypDB: a functional genomic resource for the Trypanosomatidae. **Nucleic acids research**, v. 38, n. Database issue, p. D457–62, 2010. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2808979&tool=pmcentrez&rendertype=abstract>>. Acesso em: 24/1/2014.

BARRETT, M. P.; CROFT, S. L. Emerging paradigms in anti-infective drug design. **Parasitology**, v. 141, n. 1, p. 1–7, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24401336>>. Acesso em: 29/3/2014.

BELLI, A. A.; MILES, M. A.; KELLY, J. M. A putative *Leishmania panamensis*/*Leishmania braziliensis* hybrid is a causative agent of human cutaneous leishmaniasis in Nicaragua. **Parasitology**, v. 109 (Pt 4, p. 435–42, 1994. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/7800411>>. Acesso em: 2/4/2014.

BENSON, D. A; CLARK, K.; KARSCH-MIZRACHI, I.; *et al.* GenBank. **Nucleic acids research**, v. 42, n. 1, p. D32–7, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24217914>>. Acesso em: 24/1/2014.

BLOM, A. M.; HALLSTRÖM, T.; RIESBECK, K. Complement evasion strategies of pathogens-acquisition of inhibitors and beyond. **Molecular immunology**, v. 46, n. 14, p. 2808–17, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19477524>>. Acesso em: 2/4/2014.

CESTARI, I.; ANSA-ADDO, E.; DEOLINDO, P.; INAL, J. M.; RAMIREZ, M. I. Trypanosoma cruzi immune evasion mediated by host cell-derived microvesicles. **Journal of immunology (Baltimore, Md. : 1950)**, v. 188, n. 4, p. 1942–52, 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22262654>>. Acesso em: 6/3/2014.

CESTARI, I. DOS S.; KRARUP, A.; SIM, R. B.; INAL, J. M.; RAMIREZ, M. I. Role of early lectin pathway activation in the complement-mediated killing of Trypanosoma cruzi. **Molecular immunology**, v. 47, n. 2-3, p. 426–37, 2009. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0161589009006981>>. Acesso em: 18/3/2014.

CESTARI, I.; EVANS-OSES, I.; SCHLAPBACH, L. J.; MESSIAS-REASON, I. DE; RAMIREZ, M. I. Mechanisms of complement lectin pathway activation and resistance by trypanosomatid parasites. **Molecular immunology**, v. 53, n. 4, p. 328–34, 2013. Elsevier Ltd. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/23063472>>. Acesso em: 26/7/2013.

CONSORTIUM, T. U. Activities at the Universal Protein Resource (UniProt). **Nucleic acids research**, v. 42, n. 1, p. D191–8, 2014. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24253303>>. Acesso em: 20/1/2014.

DEDIO, J.; JAHNEN-DECHENT, W.; BACHMANN, M.; MÜLLER-ESTERL, W. The multiligand-binding protein gC1qR, putative C1q receptor, is a mitochondrial protein. **Journal of immunology (Baltimore, Md. : 1950)**, v. 160, n. 7, p. 3534–42, 1998. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/9531316>>.

DVORAK, J. A.; HALL, T. E.; CRANE, M. S.; *et al.* Trypanosoma cruzi: flow cytometric analysis. I. Analysis of total DNA/organism by means of mithramycin-induced fluorescence. **The Journal of protozoology**, v. 29, n. 3, p. 430–7, 1982. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/6182288>>. Acesso em: 2/4/2014.

EL-SAYED, N. M.; MYLER, P. J.; BARTHOLOMEU, D. C.; *et al.* The genome sequence of Trypanosoma cruzi, etiologic agent of Chagas disease. **Science (New York, N.Y.)**, v. 309, n. 5733, p. 409–15, 2005. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/16020725>>. Acesso em: 19/3/2014.

EMANUELSSON, O.; NIELSEN, H.; BRUNAK, S.; HEIJNE, G. VON. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. **Journal of molecular biology**, v. 300, n. 4, p. 1005–16, 2000. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0022283600939032>>. Acesso em:

22/1/2014.

EVANS-OSES, I.; MESSIAS-REASON, I. DE; RAMIREZ, M. I. The emerging role of complement lectin pathway in trypanosomatids: molecular bases in activation, genetic deficiencies, susceptibility to infection, and complement system-based therapeutics. **TheScientificWorldJournal**, v. 2013, p. 675898, 2013. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3595680&tool=pmcentrez&rendertype=abstract>>.

EVANS-OSES, I.; MOJOLI, A.; BELTRAME, M. H.; *et al.* Differential ability to resist to complement lysis and invade host cells mediated by MBL in R4 and 860 strains of *Trypanosoma cruzi*. **FEBS letters**, v. 588, n. 6, p. 956–961, 2014. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0014579314000969>>. Acesso em: 27/2/2014.

FERREIRA, V.; VALCK, C.; SÁNCHEZ, G.; *et al.* The classical activation pathway of the human complement system is specifically inhibited by calreticulin from *Trypanosoma cruzi*. **Journal of immunology (Baltimore, Md. : 1950)**, v. 172, n. 5, p. 3042–50, 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/14978109>>.

GAUNT, M. W.; YEO, M.; FRAME, I. A.; *et al.* Mechanism of genetic exchange in American trypanosomes. **Nature**, v. 421, n. 6926, p. 936–9, 2003. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/12606999>>. Acesso em: 2/4/2014.

GHEBREHIWET, B.; LIM, B. L.; KUMAR, R.; FENG, X.; PEERSCHKE, E. I. gC1q-R/p33, a member of a new class of multifunctional and multicompartamental cellular proteins, is involved in inflammation and infection. **Immunological reviews**, v. 180, p. 65–77, 2001. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/11414365>>. Acesso em: 28/3/2014.

GHEBREHIWET, B.; PEERSCHKE, E. I. B. cC1q-R (calreticulin) and gC1q-R/p33: ubiquitously expressed multi-ligand binding cellular proteins involved in inflammation and infection. **Molecular immunology**, v. 41, n. 2-3, p. 173–83, 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15159063>>. Acesso em: 2/4/2014.

GILBERT, J. A. New hormone stimulates pancreatic β -cell proliferation. **The lancet. Diabetes & endocrinology**, v. 1 Suppl 1, p. s12, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24622586>>. Acesso em: 2/4/2014.

GOODHEAD, I.; CAPEWELL, P.; BAILEY, J. Whole-genome sequencing of *trypanosoma brucei* reveals introgression between subspecies that is associated with virulence. **mBio**, v. 4, n. 4, p. 1–8, 2013. Disponível em: <<http://mbio.asm.org/content/4/4/e00197-13.short>>. Acesso em: 13/3/2014.

GÜNZL, A. The pre-mRNA splicing machinery of trypanosomes: complex or simplified? **Eukaryotic cell**, v. 9, n. 8, p. 1159–70, 2010. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2918933&tool=pmcentrez&rendertype=abstract>>. Acesso em: 13/3/2014.

HAUBOLD, B.; WIEHE, T. Comparative genomics: methods and applications. **Die Naturwissenschaften**, v. 91, n. 9, p. 405–21, 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15278216>>. Acesso em: 27/3/2014.

HIROKAWA, N.; NODA, Y.; TANAKA, Y.; NIWA, S. Kinesin superfamily motor proteins and intracellular transport. **Nature reviews. Molecular cell biology**, v. 10, n. 10, p. 682–96, 2009. Nature Publishing Group. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19773780>>. Acesso em: 21/1/2014.

INAL, J. M. Parasite interaction with host complement: beyond attack regulation. **Trends in parasitology**, v. 20, n. 9, p. 407–12, 2004. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/15324730>>. Acesso em: 2/4/2014.

JENNI, L.; MARTI, S.; SCHWEIZER, J.; *et al.* Hybrid formation between African trypanosomes during cyclical transmission. **Nature**, v. 322, n. 6075, p. 173–5, 1986. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/3724860>>. Acesso em: 2/4/2014.

JIANG, J.; ZHANG, Y.; KRAINER, A. R.; XU, R.-M. Crystal structure of human p32, a doughnut-shaped acidic mitochondrial matrix protein. **Proceedings of the National Academy of Sciences**, v. 96, n. 7, p. 3572–3577, 1999. Disponível em: <<http://www.pnas.org/cgi/doi/10.1073/pnas.96.7.3572>>. Acesso em: 23/3/2014.

JOINER, A. K.; DASILVA, W. D.; RIMOLDI, M. T.; *et al.* Biochemical Characterization of a Factor Produced by Trypomastigotes of. **The Journal of biological chemistry**, v. 263, n. 23, p. 11327–11335, 1988.

KELLY, J. M.; LAW, J. M.; CHAPMAN, C. J.; EYS, G. J. VAN; EVANS, D. A. Evidence of genetic recombination in *Leishmania*. **Molecular and biochemical parasitology**, v. 46, n. 2, p. 253–63, 1991. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/1656255>>. Acesso em: 2/4/2014.

KRAICZY, P.; ROSSMANN, E.; BRADE, V.; *et al.* Binding of human complement regulators FHL-1 and factor H to CRASP-1 orthologs of *Borrelia burgdorferi*. **Wiener klinische Wochenschrift**, v. 118, n. 21-22, p. 669–76, 2006. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17160605>>. Acesso em: 2/4/2014.

LAMBRIS, J.; RICKLIN, D.; GEISBRECHT, B. Complement evasion by human pathogens. **Nature Reviews Microbiology**, v. 6, n. 2, p. 1–22, 2008. Disponível em: <<http://www.nature.com/nrmicro/journal/v6/n2/abs/nrmicro1824.html>>. Acesso em: 21/3/2014.

LIM BL, REID KBM, GHEBREHIWET B, P. E. The Binding Protein for Globular Heads of Complement C1q, gC1qR. FUNCTIONAL EXPRESSION AND CHARACTERIZATION AS A NOVEL VITRONECTIN BINDING FACTOR. **Journal of Biological Chemistry**, v. 271, n. 43, p. 26739–26744, 1996. Disponível em: <<http://www.jbc.org/cgi/doi/10.1074/jbc.271.43.26739>>. Acesso em: 11/2/2014.

LLEWELLYN, T.; ZHENG, H.; LIU, X.; XU, B.; PATEL, K. P. Median preoptic nucleus and subfornical organ drive renal sympathetic nerve activity via a glutamatergic mechanism within the paraventricular nucleus. **American journal of physiology. Regulatory, integrative and comparative physiology**, v. 302, n. 4, p. R424–32, 2012. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3293509&tool=pmcentrez&rendertype=abstract>>. Acesso em: 2/4/2014.

LOGAN-KLUMPLER, F. J.; SILVA, N. DE; BOEHME, U.; *et al.* GeneDB--an annotation database for pathogens. **Nucleic acids research**, v. 40, n. Database issue, p. D98–108, 2012. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3245030&tool=pmcentrez&rendertype=abstract>>. Acesso em: 24/1/2014.

MACHADO, F. S.; DUTRA, W. O.; ESPER, L.; *et al.* Current understanding of immunity to *Trypanosoma cruzi* infection and pathogenesis of Chagas disease. **Seminars in immunopathology**, v. 34, n. 6, p. 753–70, 2012. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3498515&tool=pmcentrez&rendertype=abstract>>. Acesso em: 14/3/2014.

MATHIEU-DAUDÉ, F.; LAFAY, B.; TOUZET, O.; *et al.* Exploring the FL-160-CRP gene family through sequence variability of the complement regulatory protein (CRP) expressed by the trypomastigote stage of *Trypanosoma cruzi*. **Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases**, v. 8, n. 3, p. 258–66, 2008. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/18296127>>. Acesso em: 26/7/2013.

MCDANIEL, J. P.; DVORAK, J. A. Identification, isolation, and characterization of naturally-occurring *Trypanosoma cruzi* variants. **Molecular and biochemical parasitology**, v. 57, n. 2, p. 213–22, 1993. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/8433713>>. Acesso em: 2/4/2014.

NIELSEN, H.; ENGELBRECHT, J.; BRUNAK, S.; HEIJNE, G. VON. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. **Protein engineering**, v. 10, n. 1, p. 1–6, 1997. Disponível em:

<<http://www.ncbi.nlm.nih.gov/pubmed/10065837>>.

OPPERDOES, F. R.; MICHELS, P. A. M. Horizontal gene transfer in trypanosomatids. **Trends in parasitology**, v. 23, n. 10, p. 470–6, 2007. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/17826337>>. Acesso em: 2/4/2014.

PABLOS, L. M. DE; OSUNA, A. Multigene families in *Trypanosoma cruzi* and their role in infectivity. **Infection and immunity**, v. 80, n. 7, p. 2258–64, 2012. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3416482&tool=pmcentrez&rendertype=abstract>>. Acesso em: 12/3/2014.

PAES, L. S.; MANTILLA, B. S.; BARISÓN, M. J.; WRENGER, C.; SILBER, A. M. The Uniqueness of the *Trypanosoma cruzi* Mitochondrion: Opportunities to Target New Drugs Against Chagas Disease. **Current Pharmaceutical Design**, v. 17, p. 2074–2099, 2011. Disponível em: <<http://www.ingentaconnect.com/content/ben/cpd/2011/00000017/00000020/art00007>>. Acesso em: 13/3/2014.

PARSONS, M.; MYLER, P. J.; BERRIMAN, M.; ROOS, D. S.; STUART, K. D. Identity crisis? The need for systematic gene IDs. **Trends in parasitology**, v. 27, n. 5, p. 183–4, 2011. Elsevier Ltd. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21474380>>. Acesso em: 27/3/2014.

PERTEA, M.; MOUNT, S. M.; SALZBERG, S. L. A computational survey of candidate exonic splicing enhancer motifs in the model plant *Arabidopsis thaliana*. **BMC bioinformatics**, v. 8, p. 159, 2007. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1892810&tool=pmcentrez&rendertype=abstract>>. Acesso em: 2/4/2014.

PETERSEN, T. N.; BRUNAK, S.; HEIJNE, G. VON; NIELSEN, H. SignalP 4.0: discriminating signal peptides from transmembrane regions. **Nature methods**, v. 8, n. 10, p. 785–6, 2011. Nature Publishing Group. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21959131>>. Acesso em: 23/1/2014.

PICCOLI, A. K.; ALEGRETTI, A. P.; SCHNEIDER, L.; LORA, P. S.; XAVIER, R. M. Expression of complement regulatory proteins CD55, CD59, CD35, and CD46 in rheumatoid arthritis. **Revista brasileira de reumatologia**, v. 51, n. 5, p. 503–10, 2011. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21953001>>. Acesso em: 2/4/2014.

PIERLEONI, A.; MARTELLI, P. L.; CASADIO, R. PredGPI: a GPI-anchor predictor. **BMC bioinformatics**, v. 9, p. 392, 2008. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2571997&tool=pmcentrez&rendertype=abstract>>. Acesso em: 24/1/2014.

PROSDOCIMI, F. C o l i b. ,2007.

PU, Y.-G.; JIANG, Y.-L.; YE, X.-D.; *et al.* Crystal structures and putative interface of *Saccharomyces cerevisiae* mitochondrial matrix proteins Mmf1 and Mam33. **Journal of structural biology**, v. 175, n. 3, p. 469–74, 2011. Elsevier Inc. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21600990>>. Acesso em: 26/7/2013.

RAUSCH, S.; SHAYAN, P.; SALNIKOFF, J.; REINWALD, E. Sequence determination of three variable surface glycoproteins from *Trypanosoma congolense*. Conserved sequence and structural motifs. **European journal of biochemistry / FEBS**, v. 223, n. 3, p. 813–21, 1994. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/8055958>>. Acesso em: 2/4/2014.

SÁNCHEZ-VALDÉZ, F. J.; PÉREZ BRANDÁN, C.; RAMÍREZ, G.; *et al.* A Monoallelic Deletion of the TcCRT Gene Increases the Attenuation of a Cultured *Trypanosoma cruzi* Strain, Protecting against an In Vivo Virulent Challenge. **PLoS neglected tropical diseases**, v. 8, n. 2, p. e2696, 2014. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3923724&tool=pmcentrez&rendertype=abstract>>. Acesso em: 14/3/2014.

SANTOS, J.; SILVA, J. SGBD MySQL. **SQL magazine**, p. 4–6, 2009. Disponível em: <<http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:SGBD+MySQL#5>>. Acesso em: 23/3/2014.

SIEVERS, F.; WILM, A.; DINEEN, D.; *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. **Molecular systems biology**, v. 7, n. 539, p. 539, 2011. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3261699&tool=pmcentrez&rendertype=abstract>>. Acesso em: 21/1/2014.

SILBER, A. M.; PEREIRA, C. A. Assignment of putative functions to membrane “hypothetical proteins” from the *Trypanosoma cruzi* genome. **The Journal of membrane biology**, v. 245, n. 3, p. 125–9, 2012. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/22354180>>. Acesso em: 27/3/2014.

SILVA, N. DE F.; LAMEIRA, J.; ALVES, C. N.; MARTÍ, S. Computational study of the mechanism of half-reactions in class 1A dihydroorotate dehydrogenase from *Trypanosoma cruzi*. **Physical chemistry chemical physics : PCCP**, v. 15, n. 43, p. 18863–71, 2013. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/24084894>>. Acesso em: 2/4/2014.

SPREHE, M.; FISK, J. C.; MCEVOY, S. M.; READ, L. K.; SCHUMACHER, M. A. Structure of the *Trypanosoma brucei* p22 protein, a cytochrome oxidase subunit II-specific

RNA-editing accessory factor. **The Journal of biological chemistry**, v. 285, n. 24, p. 18899–908, 2010. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2881812&tool=pmcentrez&rendertype=abstract>>. Acesso em: 19/3/2014.

STOCKS, M. E.; OGDEN, S.; HADDAD, D.; *et al.* Effect of water, sanitation, and hygiene on the prevention of trachoma: a systematic review and meta-analysis. **PLoS medicine**, v. 11, n. 2, p. e1001605, 2014. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3934994&tool=pmcentrez&rendertype=abstract>>. Acesso em: 24/3/2014.

TEIXEIRA, A. R. L.; HECHT, M. M.; GUIMARO, M. C.; SOUSA, A. O.; NITZ, N. Pathogenesis of chagas' disease: parasite persistence and autoimmunity. **Clinical microbiology reviews**, v. 24, n. 3, p. 592–630, 2011. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3131057&tool=pmcentrez&rendertype=abstract>>. Acesso em: 21/1/2014.

TEIXEIRA, S. M. R.; DAROCHA, W. D. Control of gene expression and genetic manipulation in the Trypanosomatidae. **Genetics and molecular research : GMR**, v. 2, n. 1, p. 148–58, 2003. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/12917811>>.

TERAO, Y. [Molecular analyses of the development mechanisms of severe Streptococcus pyogenes infections]. **Nihon saikingaku zasshi. Japanese journal of bacteriology**, v. 63, n. 2-4, p. 391–8, 2008. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/19317228>>. Acesso em: 2/4/2014.

UENO, N.; WILSON, M. E. Receptor-mediated phagocytosis of Leishmania: implications for intracellular survival. **Trends in parasitology**, v. 28, n. 8, p. 335–44, 2012. Elsevier Ltd. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3399048&tool=pmcentrez&rendertype=abstract>>. Acesso em: 13/2/2014.

VALCK, C.; RAMÍREZ, G.; LÓPEZ, N.; *et al.* Molecular mechanisms involved in the inactivation of the first component of human complement by Trypanosoma cruzi calreticulin. **Molecular immunology**, v. 47, n. 7-8, p. 1516–21, 2010. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0161589010000313>>. Acesso em: 18/3/2014.

VOORHIS, W. C. VAN. FL-160 proteins of Trypanosoma cruzi are expressed from a multigene family and contain two distinct epitopes that mimic nervous tissues. **Journal of Experimental Medicine**, v. 178, n. 2, p. 681–694, 1993. Disponível em: <<http://jem.rupress.org/cgi/content/long/178/2/681>>. Acesso em: 29/3/2014.

WESTENBERGER, S. J.; CERQUEIRA, G. C.; EL-SAYED, N. M.; *et al.* Trypanosoma cruzi mitochondrial maxicircles display species- and strain-specific variation and a

conserved element in the non-coding region. **BMC genomics**, v. 7, p. 60, 2006. Disponível em: <<http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1559615&tool=pmcentrez&rendertype=abstract>>. Acesso em: 12/2/2014.

WHO. Neglected tropical diseases. **SIXTY-SIXTH WORLD HEALTH ASSEMBLY**, , n. May 2013, p. 1–4, 2013.

YAMAMOTO, I.; PORTINHO. SISTEMA COMPLEMENTO: ATIVAÇÃO, REGULAÇÃO E DEFICIÊNCIAS CONGÊNITAS E ADQUIRIDAS. **Associação Brasileira de Medicina**, v. 47, n. 1, p. 41–51, 2001.

ZINGALES, B.; ANDRADE, S. G.; BRIONES, M. R. S.; *et al.* A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI. **Memórias do Instituto Oswaldo Cruz**, v. 104, n. 7, p. 1051–4, 2009. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/20027478>>. Acesso em: 2/4/2014.