

Review

The Elusive *Trypanosoma cruzi* Disperse Gene Protein Family (DGF-1)

José Luis Ramírez^{1,2}

1. Centro de Biotecnología, Instituto de Estudios Avanzados (Mincyt)

2. Instituto de Biología Experimental Universidad Central de Venezuela Caracas, Venezuela; ramjose-luis@gmail.com

Abstract: Chaga's disease caused by *Trypanosoma cruzi* infections is included in the group of neglected diseases, and efforts to develop new therapeutic or immunoprevention approaches have not been successful. After the publication of the *T. cruzi* genome, the number of molecular and biochemical studies on this parasite have increased considerably, many of which are focused on families of variant-surface-proteins, especially the trans-sialidases, mucins and mucin-associated proteins. The disperse gene protein 1 family (DGF-1) is one of the most abundant families in the *T. cruzi* genome, however, the large gene size, high copy numbers, and low antibody titers detected in infected humans make it an unattractive study target. However, here we argue that given the ubiquitous presence in all *T. cruzi* species, and physicochemical characteristics, the DGF-1 gene family may play an important role in host-parasite interactions.

Keywords: Chagas Disease; surface proteins; disperse protein family; *Trypanosoma cruzi*

1. Introduction

Trypanosoma cruzi, the etiological agent of American Trypanosomiasis or Chagas, has proven to be a big challenge for those working on the molecular biology and genomic of the parasite. The repetitive nature of the *T. cruzi* genome, the presence of cell hybridization events, variations in ploidy, and chromosomal size polymorphisms have led to heated debates about the nature of its population dynamics [1]. Nonetheless, *T. cruzi* is a pathogenic entity that depending on geographic locations and routes of infection, the symptoms in the human host vary [2]. Chagas is currently reported from the south of the United States to Argentina, and due to human migration, cases are registered in non-endemic regions of Europe and Asia [3]. Attempts to develop new effective drugs or vaccines have failed, and the only drugs for treatment are nifurtimox and benznidazole. The parasite's resilience and variability rely on its very dynamic genome, making it necessary to search for new ways to control this parasite [4]. The Disperse Gene Family (DGF-1) ranked fifth among the most repeated gene families of the *T. cruzi* CL Brener genome with 565 gene copies and 136 pseudogenes. Additionally, members of the family are among the largest genes of the parasite [5]. Despite having been discovered more than 30 years ago [6], the role of DGF-1 remains a mystery. Here, we summarize the most relevant findings concerning this gene family, providing some clues about its potential function, and discuss the incongruences between the DGF-1 transcriptome and proteome results [7].

2. The discovery of the DGF-1 family

Wincker et. al. (1990) [6] discovered the DGF-1 family while working on the genome of the *Didelphis marsupialis* *T. cruzi* strain Dm28, reporting 220 members ranging in length from 10 to 12 Kbp, and representing about 1% of the parasite's genome. In Northern Blot experiments of pulse-field resolved chromosomes, it was found that the DGF-1 genes were spread throughout the genome and did not have internally repeated sequences. Interestingly, these authors postulated that given the dispersion of the family and the large

gene size, they could participate in inter- and intrachromosomal recombination events generating chromosomal size polymorphisms. In a second work [8], the same authors registered abundant DGF-1 transcripts in the replicative epimastigote forms of the parasite, and sequenced a 10 Kbp gene that was dubbed DGF-1.1. Based on the presence of cysteine rich motifs spread along the gene, they hinted at a potential role of this protein as a receptor, and the presence of two tripeptides RGD suggested potential interactions with the host cell. In DNA hybridization experiments with other trypanosomatids such as *Phytomonas* sp., *Leptomonas samueli*, *Blastocrithidia* sp., *Crithidia fasciculata* and *Trypanosoma rangeli*, these authors reported that the family was exclusive of *T. cruzi* strains. However, DGF-1 genes were later found in whole genome sequencing of *T. rangeli* strains Sc-58, Coachi and M80 [9,10] and more recently in the African stercoraria trypanosomes *T. theileri* [11] and *T. grayi* [12].

3. General molecular characteristics of DGF-1 proteins

After the description of the first DGF-1 gene (DGF-1.1) [6], other members of the family were characterized [13- 14-15] whose most relevant features are: presence of eight to nine transmembrane hydrophobic helices at their C-end that might serve as anchors to membranes, six epidermal growth factor-1 like (EGF-1) and one EGF-2-like signatures regularly spaced approximately 400 aa apart; and lectin binding motifs. There are also integrin-like sequences suggesting a cell surface location [13-15], and a possible role either interacting with other cells, or as signal transduction receptors. These integrin-like motifs are closer to the primitive integrins (proteoglycans) of *Saccharomyces cerevisiae*, *Entamoeba histolytica*, *Candida albicans* and *Dyctiostelium*, having in common the RGD tripeptide mentioned above [8-16]. Approximately half of the family members have a canonical signal peptide, and all members lack a GPI anchor site [13]. A topology model of DGF-1.2 inserted in the cell membrane summarizing some features based on sequence homology searches, to which we added new information derived from recent advances in structural biology [17-18] is presented in Figure 1. The new structural and biological information revealed that the extracellular part of the protein has a sequence and structural identity with interactive domains such as pectin lyases, phage tail spikes, or receptors associated with virulence-factors. The general topology also resembles those of well-studied receptors, with an external cytoplasmic interactive region, and an internal cytoplasmic region which may be a signaling transmitting region [19-20].

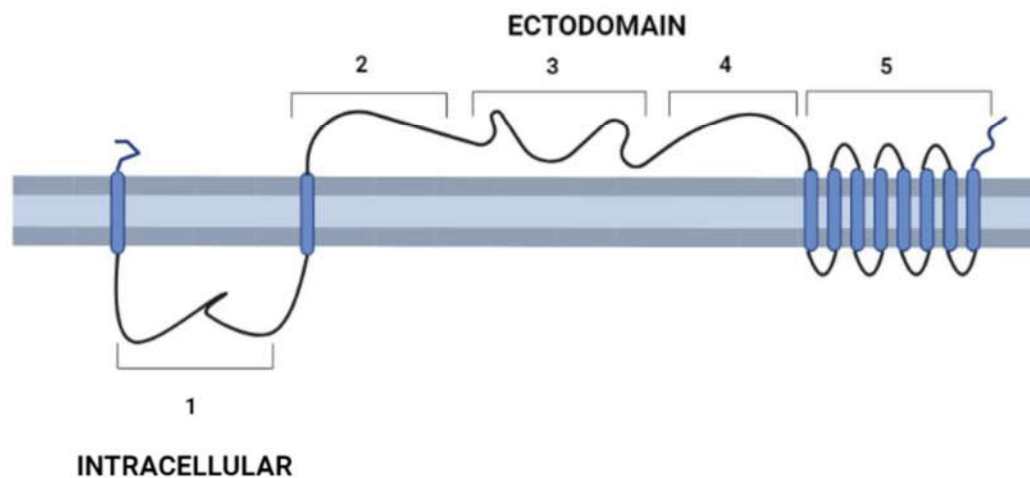


Figure 1. Schematic representation of *T. cruzi*'s DGF-1.2 protein in the cell membrane. Cell membrane cytoplasmic-extracellular topology predicted by UNIPRO. Numbers mark domains predicted by different structure-prediction software. β -sheet structure (aa 60-140) identity to Human Galectin 7 2-3 (SWISSPRO); pectin-lyase superfamily IPRO11050 (aa 882-987) (InterPro); Regions 2, 3, and 4 with several structural predictions based in sequence identity, structure similarity, and amino acids folding: Complement component C9 receptor (aa 1189-1227) (aa 1584-1622) (SWISS-PRO); Notch ligand (Panther prediction) (aa 1192-1621); pectine lyase-fold (aa 1670-1898) (AlphaFold-InterPro); arabinofuranosyltransferase aftd2 from mycobacteria (aa1993-2170) structure prediction confidence 96% (Phyre-2); 4. Threonine-rich region (aa 2800-2927) PROSITE, and structure of tailspike protein gp49 from pseudomonas phage2 lka1 88% confidence (Phyre-2), epitopes from reference 58 are located in this region; 5. Transmembrane helices region DGF1_14, DGF_1.5 (InterPro). Blue cylinders: Transmembrane domains. Graph made with BioRender.

4. DGF-1 genealogy

In their exhaustive study of the DGF-1 family in the CL Brener strain, Kawashita et.al (2009) [13] based on homology studies and the distribution of putative functional domains, divided the family in two main groups that included most gene members (66 and 51, respectively). The phylogenetic analysis also suggested that the family might have expanded by gene duplication events, but given the sequence similarity among members, these events were relatively recent or subjected to homogenization events. Also, the presence of parallel edges in the two main groups indicates that the family has experienced reticulate events such as recombination or gene conversion [13]. Two members of the family were clearly outgroups, namely XP 807429 (non-Esmeraldo-like haplotype type I group) and XP 816326 (Esmeraldo-like haplotype type II group), both lacking the pectin-lyase motif but having an immunoglobulin-like fold, and a cysteine-proteinase inhibitor domain not present in the rest of the family. As mentioned before, members of the DGF-1 family have been found in *T. rangeli*, *T. theliri* and *T. gayi*, and using genomic data they were clustered in a distinct phylogenetic group [11].

Jackson et. al. (2016) [21] proposed that DGF-1 genes appear to be innovations from more ancestral genes present in free-living organisms included in the genus *Bodo*, generically designated by as Bodonin genes. Bodonin is defined as a multicopy gene family of transmembrane glycoproteins. The typical Bodonin gene has seven transmembrane

anchors at the C-terminal, preceded by an extracellular domain, which in turn is preceded by an intracellular domain. The overall organization of Bonodin and DGF-1.2 genes is shown in Figure 2.

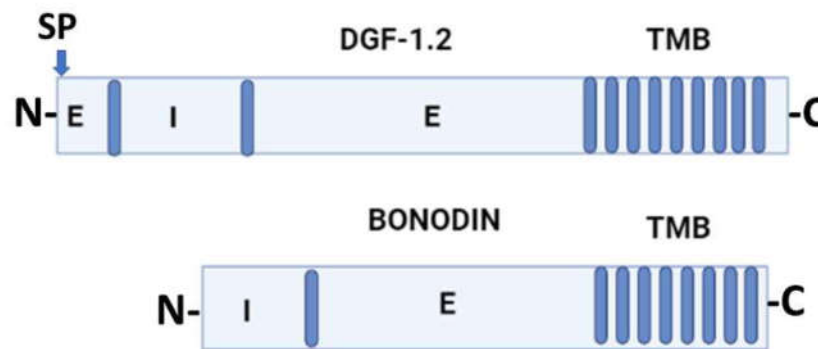


Figure 2. Schematic representation of *T. cruzi* DGF-1.2 protein and *Bodo saltans* Bonodin gene (adapted from reference 17). Barrels represent transmembrane regions; SP, signal peptide; N and C protein terminals; I, intracellular regions; E, extracellular regions. Graph made with BioRender.

5. Chromosomal distribution and DGF-1 copy numbers

DGF-1 gene organization was also revealed by Olivares et. al. (2000) [22] who while analyzing sequences of large *T. cruzi* BAC recombinants reported DGF-1 members intermingled with the L1Tc retrotransposon. Later, Kim et al. (2005) [14] reported DGF-1 copies in the subtelomeric regions of BAC-telomere recombinants of the CLBrener strain. Interestingly, as in the case of Olivares et. al. (2000) [22], the DGF-1 copies were surrounded by genes and pseudogenes of the trans-sialidases, Retrotransposon Hotspot Sequence (RHS) families, and retrotransposon elements. Despite this seemingly unstable environment the DGF-1 copies were uninterrupted ORFs.

The subtelomeric location of DGF-1 has been confirmed in different *T. cruzi* strains [23-24-25] however, there is a considerable variation in copy numbers among strains [23]. Berna et. al. (2018) [25] found that DGF-1 genes were often clustered with other repeated families in what they called “disrupted chromosomal regions”. Following this idea, then subtelomeres are part of these regions, and are likely to be involved in *T. cruzi* chromosomal gene variation dynamics [14-24-26]. A very interesting observation made by de Bezerra de Araujo et. al. [27] is the presence of chromosomal replication origins in subtelomeric DGF-1 members; this location could produce frontal collisions with the transcriptional machinery running in the direction of the telomere, and cause chromosomal instability [28], chromosomal breaks, and recombination events leading to genetic variability.

6. Possible roles for DGF-1 genes

Our homology and structural searches indicate that DGF-1 genes interact with other proteins as receptors or bidirectional signal transducers; this function is likely to be

regulated by cyclic AMP or nucleoside analogs [29-30] and non-cAMP kinases. This assumption is stressed by the results of Bao et. al. (2008) [31] using two-hybrids experiments where they found one cAMP site and two putative “non-typical” phosphorylation sites in DGF-1 peptides of the *T. cruzi* HO 3/15 strain. Moreover, Atwood et.al. (2007) [32] in a glycoproteome study detected a glycopeptide that hybridized with many DGF-1 genes, and asserted that without this type of technology these proteins would not have been detected. These features together with the *in-silico* predictions highlight the possibility that DGF-1 genes interact with other proteins, and that these interactions are regulated.

In silico studies in DGF-1

Azuaje et. al. (2007) [33] tested the behavior of Trans-sialidases and DGF-1 gene families of *T. cruzi* CL Brener using mutagenesis pressure simulations. The task involved taking members of each family, subjecting them to repeated cycles of mutations, and then compare the outcome with the repertoire of real sequences from the *T. cruzi* genome [5]. By these means, they observed that DGF-1 members, contrary to those of the trans-sialidase family, were able to generate variation at very low mutation rates or gene conversion events. In other words, DGF-1 genes displayed a very robust behavior against more dramatic changes involving intrafamily exchanges, and highlighted the possibility that DGF-1's pseudogenes may contribute to the generation of genetic variability or serve as buffers to avoid it. The contributions of pseudogenes in a dynamic exchange with the DGF-1 genes may also explain why Kawashita et. al. (2009) [13] could not detect differences in codon usage frequencies between pseudogenes and non-pseudogenes.

In a different work, this time exploring the robustness of genes of the trans-sialidase and DGF-1 families to undergo mutations causing changes at codon level, Azuaje et. al. (2007) [34] found no significant changes among members of the TS family except for the MVar1, and the DGF-1 family. The first, was the most volatile member of the TS family, while the second stood out as a family with very low volatility. In other words, MVar1 had the highest capability to vary under the immunological pressure exerted by the host, whilst DGF-1 had the lowest. The results in volatility agreed with Shannon's entropy studies done by Kawashita et. al. (2009) [13] where despite the presence of four blocks of high variability (high entropy), in general DGF-1 sequences were very well conserved. Furthermore, when these high variable blocks were tested for positive selection, a neutral evolution and purifying selection for some of these blocks were found. In summary, evidence contradicts the idea that the DGF-1 family is a relevant part of the parasite variable surface determinants, and its role is more likely related to cell-to-cell interactions or signal transduction.

Gonzalez et.al. (2009) [35] using machine learning techniques reported that *T. cruzi* adhesins such as GP82 and Tc8 [36] evolved differently from bacterial or fungi adhesins. In the case of *T. cruzi* adhesins, they bind to host cell receptors triggering responses in both directions but avoiding a strong host immunological response. In the case of DGF-1, when classifiers derived from *T. cruzi* adhesins were used, all members of the family were identified as adhesin-like proteins. In Kawashita et. al (2009) [13] as well as Gonzalez et. al. (2009) [35] XP 807429 and XP 816326 were identified as outgroup sequences.

Pseudogenes deserve a separate discussion, since their presence in such amounts in the CL Brener *T. cruzi* genome cannot be ignored, if we assume as good readings the total number of pseudogenes for the large families they totalize approximately 2000 copies, and some of them are transcribed, polyadenylated and even found in polysomes [37]. Experiments in mammalian and *Drosophila* demonstrated that degenerated genes are transcribed [38-39] and that they can regulate parental genes through siRNAs. In addition, some processed pseudogenes seem to have evolved into primate microRNA genes [40]. *T. cruzi* lacks the activity of essential genes for the siRNA machinery [41-42] thus the possibility is open to a participation as generators or suppressors of variability, or as regulators of gene expression by unknown means.

7. Expression of DGF-1 genes

Chandaa et. al. (2007) [43] placed DGF-1 proteins among the highest expressed genes in Trypanosomatids, and linked these genes to significant contributions to *T. cruzi* codon/aminoacid genomics variations. However, there are significant discrepancies between transcriptome and proteome studies.

Translation

At protein level, evidence gathered from several laboratories confirmed the presence of DGF-1 proteins in all *T. cruzi* developmental stages [13-15-44-45]. Nonetheless, in the first whole genome proteome study in *T. cruzi* CL Brener [46], no hits were found for these proteins. As mentioned before, a whole glycoproteome approach [28] reported a glycopeptide matching several members of the DGF-1 family. Using a different approach, Kawashita et al. (2009) [13] biotinylated *T. cruzi* trypomastigotes surface proteins and after recovering the proteins bound or not to streptavidin columns, used a monoclonal antibody against a DGF-1 peptide to examine both fractions. The assay revealed that a 250 KDa band was present in both fractions, and the streptavidin bound fraction had a more complex band pattern with additional lower MW bands. Interestingly, when these authors used the same antibody to probe the proteins secreted into the conditioned medium, they found similar band patterns, thus suggesting that DGF-1 proteins are shed into the medium and processed to generate lower molecular bands. In a trypomastigote-amastigote differentiation follow-up, both axenic or infected HeLa and Myoblast cell, we probed *T. cruzi* cells with an immunopurified polyclonal antibody prepared against a peptide from the DGF-1.2 protein [15]. In this experiment we found that the DGF-1.2 protein was mainly expressed in amastigote forms and the highest expression occurred at 14 hr in the cell free system, and at 18 hr in the infected mammalian cells. During trypomastigote-amastigote transition the fluorescent particles containing DGF-1.2 progressed towards the inner side of the parasite cell membrane, and later released into the medium or inside the mammalian host cells. When we analyzed the cell-free differentiation medium using Southern Blot experiments after 20 hr of differentiation, we detected protein bands above 220 KDa as well as lower MW bands. When these bands were analyzed by MADIT-OF mass-spectrometry, all of them presented matches with DGF-1 genes. These results indicated that DGF-1 gene expression is regulated during the differentiation from trypomastigotes to amastigotes, and that the protein is processed into lower MW bands. Using a totally different approach consisting of cell-fractionation of *T. cruzi* trypomastigotes, followed by one-dimensional gel electrophoresis and LC-MS/MS, Ulrich et al (2011) [44] identified different peptides that matched at least 39 members of the DGF-1 family, suggesting that these proteins were located in an organelle population different from acidocalcisomes, glycosomes, reservosomes, lipid droplets, or endocytic vesicles in the different stages of *T. cruzi* and provided definitive evidence that DGF-1 proteins are also expressed in epimastigote forms. Consistent with these results, Brossas et. al. (2017) [45] studying the *T. cruzi* secretome during the trypomastigotes-amastigotes transition in Vero cells, registered 18 DGF-1 proteins by Nano LC-MS/MS technology. The authors did not state whether these were free proteins or if they were included in vesicles. In our unpublished EM results with gold labelled DGF-1.2, we detected these proteins included in vesicles smaller than 22 μ shed by all *T. cruzi* developmental forms.

In conclusion, DGF-1 proteins are expressed at all developmental stages of *T. cruzi* but in particular during trypomastigote-amastigote differentiation, and eventually released from amastigotes as small bodies inside the mammalian cell.

Transcription

In most organisms there is little agreement between transcriptome and proteome studies [47-48]. Some authors argue that the discrepancies are mostly due to technical problems that will eventually be resolved with more advanced proteomic techniques, but at least in developmental stages of *Leishmania* the correlation was nearly 60% [49]. In

Brossas et. al. (2017) [45] proteomic analysis of the *T. cruzi* secretome of trypomastigote forms on CL Brener and VD strains, DGF-1 was second among the top proteins detected. Contrarily, Callejas-Hernández et al. (2019) [50] in their whole transcriptome analysis of trypomastigote forms of the *T. cruzi* Silvio X10 strain did not report any DGF-1 transcripts.

In most assays focused on the detection of RNAs by qPCR and proteins using specific antibodies, there is better agreement [13-15-44-45-51] than in whole proteome and transcriptome techniques. Most proteomic techniques will underestimate or miss proteins with modifications like lipidation, glycosylation etc., and *in silico* prediction reveals that DGF-1 has many of these modification sites. Due to recent advances in DNA sequencing techniques plus the affordability of this type of studies, the number of *T. cruzi* transcriptomics reports have increased considerably [50] (not all mentioned here). On the other hand, most transcriptome techniques rely on the ability to trap polyadenylated RNAs after cell lysis including or excluding the nucleus. As Pastro et. al. (2017) [52] demonstrated there is a nuclear accumulation of some RNAs in *T. cruzi*, which at a given differentiation stage are transported to the cytoplasm for translation. Additionally, trypanosomatids genes are polycistronically transcribed and polyadenylated (including pseudogenes). However, not all RNAs are translated, a fact that is defined in some cases by the regulatory control sequences in their 3'UTRs [53-54] or by sequestration in mRNA-binding-protein bodies (RBP) [55]. Scmidin et. al. (2015) [56] claimed a better agreement between transcription and translation in *T. cruzi* by isolating polysomal mRNAs. A clear example of this type of problem is revealed when we compare the GenSeq transcriptome results of Berna et. al. (2017) [57] with the follow-up study of DGF-1.2 protein by Lander et.al. (2010) mentioned above. In the first case, the authors did not observe DGF-1 mRNA expression differences in all *T. cruzi* developmental stages, whereas in the second a step-by-step approach revealed a clear time-regulated expression of DGF-1.2 during trypomastigote-amastigote transition. Berna et. al. (2017) [57] warned that their results should be taken cautiously given the possibility of miss-assembling and annotation problems in the *T. cruzi* Dm28c strain genome. A different problem in most transcriptome works is the lack of synchronicity of *T. cruzi* cells, a simple observation of parasites in culture, or parasites recovered from Triatoma bugs confirm the extensive pleomorphism exhibited by *T. cruzi*'s cells. Hence, it appears, without denying the utility of whole approaches, that at their current stage they lack the precision to study the fine tuning of *T. cruzi* differentiation processes.

8. Conclusions

Despite the wealth of information provided by the new omics and advances in immunology, the role of a prominent and ubiquitous protein family like DGF-1 remains unknown.

The large number of copies, and gene sizes of the DGF-1 family represent an energetic burden for any parasitic organism. In addition, DGF-1 genes appear to be very robust with little variation among duplicated genes, suggesting that the function of most members of the family ought to be essential for the parasite. Is this stability related to the function of the protein per se, or/and does it have to do with the location of replication origins inside these genes?

It would be tempting to speculate that, given that *T. cruzi* groups with African sterocorarian parasites which also have DGF-1 families, these proteins may play a role in contaminative transmission. Nonetheless, *Trypanosome rangeli*, a typical salivarian parasite, has many putative DGF-1 gene copies, but so far, the DGF-1 proteins reported for this organism are rather small [9].

In vitro differentiation experiments have revealed that the DGF-1 protein is released inside the mammalian host cells and might interact, affecting their defenses [15]. Our attempts to obtain an immunological protection in mice injected with a DGF-1.2 have failed, and this fact agrees with the failure to block the parasite invasion reported by Kawashita et. al. (2009) [13]. However, both peptides used in these experiments (aa 779-1030 and 876-

1117, respectively) faced the intracellular domain of the protein (Fig 1), and it is likely that by using peptides based on the protein ectdomain we might obtain a better immunological response. Most infectivity and immunological studies in *T. cruzi* focus on surface proteins like trans-sialidases, mucin, and mucin-associated protein families to which the parasite developed a confounding strategy exposing simultaneously multiple immunogenic determinants to avoid an effective response [58] thus, we ought to redirect our attention to targets not involved in this type of interaction.

An example supporting this idea is the recent discovery that the *Trypanosoma brucei* invariant protein gp65 (ISG65) is a C3b complement receptor [59]. This receptor is an important part of innate immune system, and its role is central to the activation of the alternative, classical and lectin pathways of complement. Therefore, among other possibilities, this interaction could block the cascade of events that lead to complement parasite lysis, thus allowing the parasite to pass the first host immunological barrier.

Proteins associated with *T. cruzi*'s infectivity are mostly expressed in the infective forms of the parasite, i.e., metacyclic forms from the insect guts, intracellular amastigote replicative forms, and blood metacyclic forms, all of which seem to be resistant to complement mediated lysis. In the acute phase of the disease, once amastigotes burst out of the host cells they are exposed to the bloodstream where they activate the complement system and bind to the C9 terminal part of the complement cascade, but at the same time the C5b-9 complexes bind to the cell membranes but do not get internalized.

Iida et al. (1989) [60] and Ramírez-Tolosa and Ferreira (2017) [61] have proposed that *T. cruzi* amastigote-complement resistance can be accounted for the "presence of unidentified specific inhibitors, homologous to a host factor that will prevent the incorporation of C8 and C9 on amastigotes" (Iida et al., 1989) [60]. An alternative explanation by the same authors is the presence of hydrophobic domains on the parasite surface molecules (or that they are released by the parasite) serving as non-specific "traps" of nascent C5b-7 complexes. Could the hydrophobic domains in DGF-1 be some of these traps? Is DGF-1 one of the many factors preventing the action of the host complement system or is it a signal transduction receptor? Additionally, in a comprehensive computational screening of the *T. cruzi* proteome aimed to identify eight proteins for a multiple-epitope vaccine, DGF-1 epitopes were included in the final vaccine. For this selection, epitopes were chosen according to their predicted high antigenic and immunogenic MHC class I, MHC class II and B cell [62].

9. Future Directions

The information described herein suggests that the DGF-1 family participates in basic signaling or the interaction between the parasite and the mammalian host non-related to the evasion by typical surface protein variants of the trans-sialidases or mucin families. Although the large gene size and multiple copies of the DGF-1 family are important obstacles for molecular approaches, we believe it deserves further studies.

Acknowledgements: This publication of this work was possible by the support of Fundacion Empresas Polar, Caracas, Venezuela. To Mr. Ralph Wilk for revising the English version

References

1. Tibayrenc, M., Ayala, F.J. (2022). Microevolution and subspecific taxonomy of *Trypanosoma cruzi*. Infect. Genet. Evol. 103:105344. doi: 10.1016/j.meegid.2022.105344
2. [https://www.who.int/news-room/fact-sheets/detail/chagas-disease-\(american-trypanosomiasis\)](https://www.who.int/news-room/fact-sheets/detail/chagas-disease-(american-trypanosomiasis))
3. Gómez-Ochoa, S.A., Rojas, L.Z., Echeverría, L.E., Muka, T. and Franco, O.H., 2022. Global, Regional, and National Trends of Chagas Disease from 1990 to 2019: Comprehensive Analysis of the Global Burden of Disease Study. Global Heart, 17(1), p.59. DOI: <http://doi.org/10.5334/gh.1150>
4. García-Huertas, P., Cardona-Castro, N. (2021) Advances in the treatment of Chagas disease: Promising new drugs, plants and targets. Biomedicine & Pharmacotherapy, Volume 142:112020 <https://doi.org/10.1016/j.biopha.2021.112020>.
5. El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C. Nilsson, D., Aggarwal, G., Tran, Anh-Nhi, Ghedin, E, Worthey, E.A, Delcher, A.L, Blandin, G.et. al. (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science. 309:409-415. doi: 10.1126/science.1112631

6. Wincker, P., Roizes, G., Goldenberg, S. (1990). Characterization of a *Trypanosoma cruzi* specific nuclear repeated sequence, Mol. Biochem. Parasitol. 41:147-152. [https://doi.org/10.1016/0166-6851\(90\)90105-U](https://doi.org/10.1016/0166-6851(90)90105-U)
7. Chávez, S., Urbaniak, M.D., Benz, C., Smircich, P., Garat, B., Sotelo-Silveira, J.R., Duhagon, M.A. (2021). Extensive Translational Regulation through the Proliferative Transition of *Trypanosoma cruzi* Revealed by Multi-Omics. mSphere. 6(5):e0036621. doi: 10.1128/mSphere.00366-21.
8. Wincker, P., Cristina Murto-Dovales, A., Goldenberg, S. (1992). Nucleotide sequence of a representative member of a *Trypanosoma cruzi* dispersed gene family, Mol. Biochem. Parasitol. 55:217-220, [https://doi.org/10.1016/0166-6851\(92\)90142-7](https://doi.org/10.1016/0166-6851(92)90142-7).
9. Stoco, P.H., Wagner, G., Talavera-Lopez, C., Gerber, A., Zaha, A., Thompson, C.E., Bartholomeu, D.C., Lückemeyer, D.D., Bahia, D., Loreto, E., et al., (2014). Genome of the avirulent human-infective trypanosome *Trypanosoma rangeli*. PLoS Negl Trop Dis. 18;8(9): e3176. doi: 10.1371/journal.pntd.0003176
10. Bradwell, K.R., Koparde, V.N., Matveyev, A.V. Serraneo, M.G, Joao, J.M.P., Parikh, H., Huang, B., Lee, V., Espinosa-Alvarez O, et al. (2018). Genomic comparison of *Trypanosoma conorhini* and *Trypanosoma rangeli* to *Trypanosoma cruzi* strains of high and low virulence. BMC Genomics. 19(1):770. doi: 10.1186/s12864-018-5112-0.
11. Kelly, S., Ivens, A., Mott, G.A. O'Neill, M., Emms, D., Macleod, O., Voorheis, P., Tyler, K., Clark, M., Matthews J, et. al. (2017). An Alternative Strategy for *Trypanosome* Survival in the Mammalian Bloodstream Revealed through Genome and Transcriptome Analysis of the Ubiquitous Bovine Parasite *Trypanosoma* (Megatrypanum) *theileri*. Genome Biol Evol. 9:2093-2109. doi: 10.1093/gbe/evx152.
12. Kelly, S., Ivens, A., Manna, P., Gibson, W. Field, M.C. (2014). A draft genome for the African crocodilian trypanosome *Trypanosoma grayi*. Sci Data 1, 140024 <https://doi.org/10.1038/sdata.2014.24>.
13. Kawashita, S. Y., da Silva, C.V., Mortara, R.A., Burleigh, B., Briones, M.R.S. (2009). Homology, paralogy and function of DGF-1, a highly dispersed *Trypanosoma cruzi* specific gene family and its implications for information entropy of its encoded proteins, Mol. Biochem. Parasitol. 165:19-31 <https://doi.org/10.1016/j.molbiopara.2008.12.010>.
14. Kim, D., Chiurillo, M.A., El-Sayed, N., Santos, M.R.M., Johns, K., Porcile, P.E., Andersson, B., Myler, P., da Silveira, J.F., Ramírez, J.L. (2005). Telomere and subtelomere of *Trypanosoma cruzi* chromosomes are enriched in (pseudo)genes of retrotransposon hot spot and trans-sialidase-like gene families: the origins of *T. cruzi* telomeres. Gene 346:153-61. doi: 10.1016/j.gene.10.014.
15. Lander, N., Bernal, C., Diez, N., Añez, N., Docampo, R., Ramirez, J.L. (2010). Localization and developmental regulation of a disperse gene family 1 protein in *Trypanosoma cruzi*. Infect. Immun 78:231-241.
16. Ouaisi, M.A. (1988) Role of the RGD sequence in parasite adhesion to host cells. Parasitol. Today 4:169-173
17. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko A, et. al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. <https://doi.org/10.1038/s41586-021-03819-2>
18. Baek, M., DiMaio, F., Anishchenko, I., Bauparas, J., Ovchinnikov, S. Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science 373:871-876. doi:10.1126/science.abj8754
19. Bhavé, G., Nadin, B.M, Brasier, D.J., Glauner, K.S., Heinemann, G.F, Farzana, K. Karim, F., Gereau, R.W. (2003). Membrane Topology of a Metabotropic Glutamate Receptor J. Biol. Chem. 278: 30294–30301.
20. Gunnar von Heijne (2006). Membrane-protein topology Nature Rev, Mol Cell Biol 7:909-918 doi:10.1038/nrm2063.
21. Jackson, A.P., Otto, T.D., Aslett, M. Armstrong, S.D., Bringaud, F., Schlasch, A., Hartley, C., Sanders, M., Wastling, J.M., Dacks, J.B. et. al. (2016). Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism Current Biology 26:161–172 <http://dx.doi.org/10.1016/j.cub.2015.11.055>.
22. Olivares, M., Thomas, M.C., Lopez-Barajas, A., Requena, J.M., Garcias-Perez, J.L., Alonso, C, Lopez, M.C. (2000). Genome clustering of the *Trypanosoma cruzi* non-long terminal L1Tc retrotransposon with defined intersperse repeated DNA elements. Electrophoresis 21:2973-2982.
23. Callejas-Hernández, F., Rastrojo, A., Poveda, C., Girones, N., Fresno, M. (2018). Genomic assemblies of newly sequenced *Trypanosoma cruzi* strains reveal new genomic expansion and greater complexity. Sci. Rep. 8, 14631 <https://doi.org/10.1038/s41598-018-32877-2>.
24. Talavera-Lopez, C., Messenger, L.A., Lewis, M.D., Yeo, M., Reis-Cunha, J.L., Machado Matos, G., Bartholomeu, D.C., Calzada, J.E., Saldaña, A., Ramírez, J.D., Guhl, F, et. al. (2021). Repeat-driven Generation of antigenic diversity in a major human pathogen, *Trypanosoma cruzi* Front. Cell. Infect Microbiol. 11:614665 doi: 103389/fcimb.2021.61466
25. Berná, L., Rodríguez, M., Chiribao, M.L., Parodi-Talice, A., Pita, S. Rijo, G., Alvarez-Valin, F., Robello, C. (2018). Expanding an expanded genome: long-read sequencing of *Trypanosoma cruzi*. Microb. Genomics. 4(5). DOI: 10.1099/mgen.0.000177.
26. Ramirez, J.L. (2020) An Evolutionary View of *Trypanosoma cruzi* Telomeres. Front. Cell. Infect. Microbiol. 9:439. doi: 10.3389/fcimb.2019.00439
27. Bezerra de Araujo, C., Pinheiro Chagas da Cunha, J., Inada, T., Damasceno, J., Lima, A.R.J., Iraiwa, P., Marques, C., Gonçalves, E., Nishiyama-Junior, M.Y., McCulloch, R., et. al. (2020) Replication origin location might contribute to genetic variability in *Trypanosoma cruzi* BMC Genomics 21:414 <https://doi.org/10.1186/s12864-020-06803-8>.
28. Saha, A., Nanavaty, V.P., Li, B. (2019). Telomere and Subtelomere R-loops and Antigenic variation in Trypanosomes, J. Mol. Biol., <https://doi.org/10.1016/j.jmb.2019.10.025>.
29. Sunil Laxman, S., Riechers, A., Sadilek, M. (2006). Hydrolysis products of cAMP analogs cause transformation of *Trypanosoma brucei* from slender to stumpy-like forms. PNAS 103: 19194–19199. doi:10.1073.pnas.0608971103.

30. Bachmaier, S., Volpato Santos, Y., Kramer, S., Githure, G.B., Klockner, T., Pepperi, J., Baums, C., Schenk, R., Schwede, F., Genieser, H.G. et al. (2019). Nucleoside analogue activators of cyclic AMP-independent protein kinase A of *Trypanosoma*. *Nat Commun* 10, 1421 <https://doi.org/10.1038/s41467-019-09338-z>
31. Bao, Y., Weiss, L.M., Braunstein, V.L. (2008) The Role of Protein Kinase A in *Trypanosoma cruzi*. *Infect Imm.* 76:4757-4763 76. doi:10.1128/IAI.00527-08.
32. Atwood, J.A. 3rd, Minning, T., Ludolf, F., Nuccio, A., Wheatherly, D.B., Alvarez-Manilla, G. Tarleton, R., Orlando, R. (2006) Glycoproteomics of *Trypanosoma cruzi* trypomastigotes using subcellular fractionation, lectin affinity, and stable isotope labeling. *J. Proteome Res.* 5:3376-84. doi: 10.1021/pr060364b.
33. Azuaje, F.J., Ramirez, J.L., Da Silveira, J.F. (2007). *In silico*, biologically-inspired modelling of genomic variation generation in surface proteins of *Trypanosoma cruzi*. *Kinetoplastid Biol Dis.* 10; 6:6. doi: 10.1186/1475-9292-6-6.
34. Azuaje, F.J., Ramirez, J.L., Da Silveira, J.F. (2007). An exploration of the genetic robustness landscape of surface protein families in the human protozoan parasite *Trypanosoma cruzi*. *IEEE Trans Nanobioscience.* 6:223-8. doi: 10.1109/tnb.2007.903482. PMID: 17926780.
35. Gonzalez, A. M., Azuaje, F. J., Ramirez, J.L., Dorronsoro, J.R. (2009). Machine Learning Techniques for the Automated Classification of Adhesin-Like Proteins in the Human Protozoan Parasite *Trypanosoma cruzi*, *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 6:695-702, doi: 10.1109/TCBB.2008.125.
36. Ramirez, M. I., de Cassia Ruiz, R., Araya, J.E, J.F.Da S ilveira, Yoshida, N. (1993). Involvement of the Stage-Specific 82-Kilodalton Adhesion Molecule of *Trypanosoma cruzi* Metacyclic Trypomastigotes in Host Cell Invasion. *Infect. Imm.* 61: 3636-3641.
37. Wen, Y.Z., Zheng, L.L., Qu, L.H., Ayala, F.J., Lun, Z-R (2012). Pseudogenes are not pseudo any more. *RNA Biol.* 9:27-32.
38. Harrison, P., Zheng, D., Zhang, Z. Carriero, N., Gerstein, M. (2005). Transcribed processed pseudogenes in the human genome: an intermediate form of expressed retrosequence lacking protein-coding ability. *Nucleic Acids Res.* 33: 2374-83 DOI: 10.1093/nar/gki531
39. Sasidharan, R., Gerstein, M. (2008). Protein fossils live on as RNA. *Nature.* 453:729-31. doi: 10.1038/453729a
40. Devor, E.J. (2006) Primate MicroRNAs miR-220 and miR-492 Lie within Processed Pseudogenes *J. Hered.* 97:186-190 doi:10.1093/jhered/esj022
41. Garcia Silva, M.R., Tosar, J.P., Frugier, M., Pantano, S., Bonilla, B., Esteban, L., Serra, E., Rovira, C., Robello, C., Cayota, A. (2010) Cloning, characterization and subcellular localization of a *Trypanosoma cruzi* argonaute protein defining a new subfamily distinctive of trypanosomatids. *Gene.* 466:26-35. doi:10.1016/j.gene. 06.012
42. Zheng, Y., Cai, X., and Bradley, J.E. (2013) microRNAs in parasites and parasite infection, *RNA Biology*, 10:371-379, DOI: 10.4161/rna.23716
43. Chanda, I., Panb, A., Kumar Saha, S. (2007). Comparative codon and amino acid composition analysis of Tritryps-conspicuous features of *Leishmania major* *FEBS Letters* 581 5751–5758.
44. Ulrich, P.N., Jimenez, V., Park, M., Martins, V.P., Altwood II, J., Collins, D., Rohloff, P., Tarleton, R., Moreno, S.N., Orlando, R. et. al. (2011) Identification of Contractile Vacuole Proteins in *Trypanosoma cruzi*. *PLoS ONE* 6(3): e18013. doi:10.1371/journal.pone.0018013
45. Brossas, J.Y., Gulin, J.E.N., Bisio, M.M.C., Chapelle, M., Marinach-Patrice, C., Bordessoules, M., Palazon Ruiz, G., Vion, J., Paris, L., Altcheh, J., et. al. (2017). Secretome analysis of *Trypanosoma cruzi* by proteomics studies. *PLoS One.*;12(10):e0185504. doi:10.1371/journal.pone.0185504
46. Atwood, J.A. 3rd, Weatherly, D.B., Minning, T.A., Bundy, B., Cavola, C., Oppendoes, F.R. Orlando, R., Tarleton, R.L. (2005). The *Trypanosoma cruzi* proteome. *Science* 309:473-6. doi:10.1126/science.1110289.
47. Ghazalpour, A., Bennett, B., Petyuk, V.A., Orozco, L., Agopian, R., Mungrue, I.N., Farber, C.R., Sinsheimer, J., Kang, H.M., Furlotte, N. et. al. (2011) Comparative Analysis of Proteome and Transcriptome Variation in Mouse. *PLoSGenet* 7(6): e1001393. doi:10.1371/journal.pgen.1001393.
48. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, A, Chen, W., Selbach, M. (2011) Global quantification of mammalian gene expression control. *Nature* 473:337–342 <https://doi.org/10.1038/nature10098>
49. Cortazzo da Silva, L., Aoki, J.I. and Floeter-Winter, L.M. (2022). Finding Correlations Between mRNA and Protein Levels in *Leishmania* Development: Is There a Discrepancy? *Front. Cell. Infect. Microbiol.* 12:852902. doi: 10.3389/fcimb.2022.852902.
50. Callejas-Hernández, F., Gutierrez-Nogues, A., Rastrojo, A. (2019). Analysis of mRNA processing at whole transcriptome level, transcriptomic profile and genome sequence refinement of *Trypanosoma cruzi*. *Sci. Reports* 9:17376 <https://doi.org/10.1038/s41598-019-53924-6>
51. Ibarrola-Vannucci, A.K., De Pablos, L.M., Moreira, R.L., Cornet-Gomez, A., Cruz-Bustos, T., Schijman, A.G., Ramírez, J.L., Vélchez, S., Osuna, A. (2021). Characterization and functional analysis of the proteins Prohibitin 1 and 2 in *Trypanosoma cruzi*. *PLoS Negl. Trop. Dis.* 15(4): e0009322.
52. Pastro, L., Smircich, P., Di Paolo, A., Becco, L., Duhagon, L.A, Sotello-Silveira, J, Garat, B. (2017) Nuclear Compartmentalization Contributes to Stage-Specific Gene Expression Control in *Trypanosoma cruzi*. *Front. Cell Dev. Biol.* 5:8. doi: 10.3389/fcell.2017.00008.
53. Clayton, C. (2019). Regulation of gene expression in trypanosomatids: living with polycistronic transcription. *Open Biol.* 9:190072 doi: 10.1098/rsob.190072
54. Kramer, S. and Carrington, M. (2011). Trans-acting proteins regulating mRNA maturation, stability and translation in trypanosomatids *Trends Parasitol.* 27:1471-1492.

-
55. de Pablos, L.M., Ferreira, T.R., Dowle, A.A., Forrester, S., Parry, E., Newlin, K., Walrat, P.B. (2019). The mRNA-bound Proteome of *Leishmania mexicana*: Novel Genetic Insight into an Ancient Parasite. *Mol. Cell. Proteomics* 18:1271–1284
 56. Smircich, P., Eastman, G., Bispo, S., Duhagon, M.A., Guerra-Slompo, E.P., Garat, B., Goldenberg, S., Munroe, D.J., Dallagiovanna, B., Holetz, F. et. al. (2015). Ribosome profiling reveals translation control as a key mechanism generating differential gene expression in *Trypanosoma cruzi*. *BMC Genomics* 16:443 doi: 10.1186/s12864-015-1563-8
 57. Berná, L., Chiribao, M. L., Greif, G. (2017). Transcriptomic analysis reveals metabolic switches and surface remodeling as key processes for stage transition in *Trypanosoma cruzi*. *Peer J.*, 5, e3017. <https://doi.org/10.7717/peerj.3017>
 58. Kahn, S., Van Voorhis, W.C. and Eisen, H. (1990) The major 85-kd surface antigen of the mammalian form of *Trypanosoma cruzi* is encoded by a large heterogeneous family of simultaneously expressed genes, *J. Exp. Med.* 172:589-597.
 59. Macleod, O.J.S., Cook, A.D., Webb, H., Crow, M., Burns, R., Redpath, M., Seisenberger, S., Trevor, C.E., Peacock, L., Schwede, A. et. al. (2022). Invariant surface glycoprotein 65 of *Trypanosoma brucei* is a complement C3 receptor. *Nat. Commun.*13:5085 <https://doi.org/10.1038/s41467-022-32728-9>
 60. Iida, K., Whitlow, M. B., Nussenzweig, V. (1989). Amastigotes of *Trypanosoma cruzi* escape destruction by the terminal complement components. *J. Exp. Med.* 169:881-891. doi: 10.1084/jem.169.3.881.
 61. Ramírez-Tolosa, G., Ferreira, A. (2016). *Trypanosoma cruzi* evades the Complement System as an Efficient Strategy to Survive in the Mammalian Host: The Specific Roles of Host/Parasite Molecules and *Trypanosoma cruzi* Calreticulin. *Frontiers in Microbiology*, 8. <https://doi.org/10.3389/fmicb.2017.01667>
 62. Rawal, K., Sinha, R., Abbasi, B.A., Chaudary, A., Nath, S.K., Kumari, P., Preeti, P., Saraf, D., Singh, S., Mishra, K., et. al. (2021). Identification of vaccine targets in pathogens and design of a vaccine using computational approaches. *Sci Rep.* 2021;11(1):17626. doi:10.1038/s41598-021-96863-x