

# Trypanosoma cruzi Disperse Gene Protein Family (DGF-1)

Subjects: **Biotechnology & Applied Microbiology**

Contributor: José Luis Ramírez

Chagas 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 has increased considerably, many of which are focused on families of variant surface proteins, especially 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.

chagas disease

surface proteins

disperse protein family

Trypanosoma cruzi

## 1. Introduction

*Trypanosoma cruzi*, the etiological agent of American Trypanosomiasis or Chagas disease, has proven to be a big challenge for those working on the molecular biology and genomics 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, and 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 also 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.

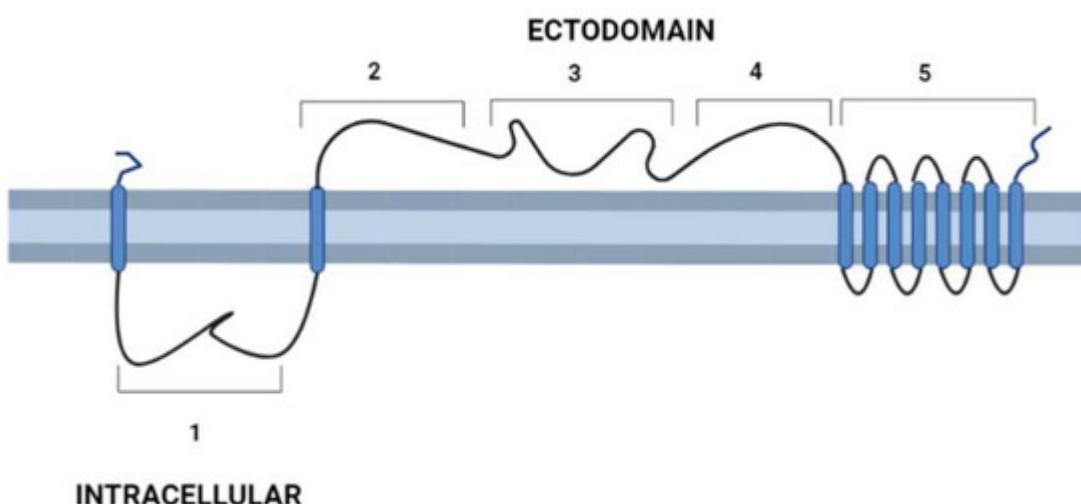
## 2. The Discovery of the DGF-1 Family

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 with 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 [7], 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., *Critidium 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 [8][9] and more recently in the African stercoraria trypanosomes *T. theileri* [10] and *T. grayi* [11].

### 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 [12][13][14], whose most relevant features are: the 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 signature regularly spaced approximately 400 aa apart; and lectin binding motifs. There are also integrin-like sequences suggesting a cell surface location [12][13][14] and a possible role either interacting with other cells or as signal transduction receptors. These integrin-like motifs are closer to the primitive integrins (protointegrins) of *Saccharomyces cerevisiae*, *Entamoeba histolytica*, *Candida albicans*, and *Dictyostelium*, having in common the RGD tripeptide mentioned above [6][15]. Approximately half of the family members have a canonical signal peptide, and all members lack a GPI anchor site [12]. A topology model of DGF-1.2 inserted in the cell membrane summarizing some features based on sequence homology searches, to which researchers added new information derived from recent advances in structural biology [16][17], 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 [18][19].

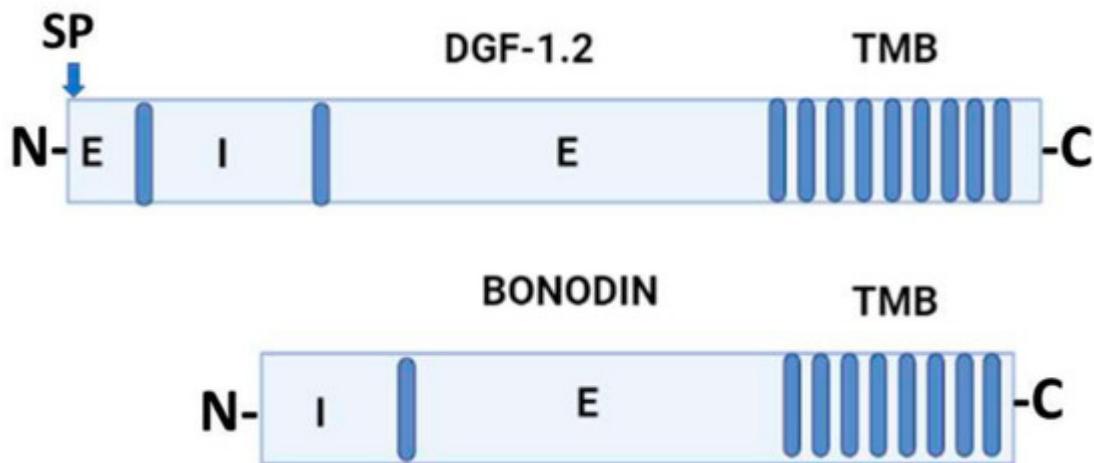


**Figure 1.** Schematic representation of *T. cruzi*'s DGF-1.2 protein in the cell membrane. Cell membrane cytoplasmic-extracellular topology predicted by UNIPRO. Numbers mark domains predicted by different structure-prediction software. 1.  $\beta$ -sheet structure (aa 60–140) identity to Human Galectin 7 2–3 (SWISSPRO); pectin-lyase superfamily IPRO11050 (aa 882–987) (InterPro); Regions 2, 3, and 4 with several structural predictions based on sequence identity, structure similarity, and amino acid 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 (aa 1993–2170) 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 [20] 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) [12], based on homology studies and the distribution of putative functional domains, divided the family into two main groups that included most gene members (66 and 51, respectively). The phylogenetic analysis also suggested that the family might have expanded through 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 [12]. 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 [10].

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 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 the Bodonin 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 [16]. 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 intermingling with the L1Tc retrotransposon. Later, Kim et al. (2005) [13] 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], but 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 [13][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.

## References

1. Tibayrenc, M.; Ayala, F.J. Microevolution and subspecific taxonomy of *Trypanosoma cruzi*. *Infect. Genet. Evol.* 2022, 103, 105344.

2. Chagas Disease. Available online: [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)) (accessed on 13 April 2022).
3. Gómez-Ochoa, S.A.; Rojas, L.Z.; Echeverría, L.E.; Muka, T.; Franco, O.H. Global, Regional, and National Trends of Chagas Disease from 1990 to 2019: Comprehensive Analysis of the Global Burden of Disease Study. *Glob. Heart* 2022, 17, 59.
4. García-Huertas, P.; Cardona-Castro, N. Advances in the treatment of Chagas disease: Promising new drugs, plants and targets. *Biomed. Pharmacother.* 2021, 142, 112020.
5. El-Sayed, N.M.; Myler, P.J.; Bartholomeu, D.C.; Nilsson, D.; Aggarwal, G.; Tran, A.-N.; Ghedin, E.; Worthey, E.A.; Delcher, A.L.; Blandin, G.; et al. The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. *Science* 2005, 309, 409–415.
6. Wincker, P.; Roizes, G.; Goldenberg, S. Characterization of a *Trypanosoma cruzi* specific nuclear repeated sequence. *Mol. Biochem. Parasitol.* 1990, 41, 147–152.
7. Wincker, P.; Cristina Murto-Dovales, A.; Goldenberg, S. Nucleotide sequence of a representative member of a *Trypanosoma cruzi* dispersed gene family. *Mol. Biochem. Parasitol.* 1992, 55, 217–220.
8. 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. Genome of the avirulent human-infective trypanosome *Trypanosoma rangeli*. *PLoS Negl. Trop. Dis.* 2014, 18, e3176.
9. 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.; Ortiz, P.A.; et al. Genomic comparison of *Trypanosoma conorhini* and *Trypanosoma rangeli* to *Trypanosoma cruzi* strains of high and low virulence. *BMC Genom.* 2018, 19, 770.
10. Kelly, S.; Ivens, A.; Mott, G.A.; O'Neill, M.; Emms, D.; Macleod, O.; Voorheis, P.; Tyler, K.; Clark, M.; Matthews, J.; et al. 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.* 2017, 9, 2093–2109.
11. Kelly, S.; Ivens, A.; Manna, P.; Gibson, W.; Field, M.C. A draft genome for the African crocodilian trypanosome *Trypanosoma grayi*. *Sci. Data* 2014, 1, 140024.
12. Kawashita, S.Y.; da Silva, C.V.; Mortara, R.A.; Burleigh, B.; Briones, M.R.S. 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.* 2009, 165, 19–31.
13. 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. 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* 2005, 346, 153–161.

14. Lander, N.; Bernal, C.; Diez, N.; Añez, N.; Docampo, R.; Ramirez, J.L. Localization and developmental regulation of a disperse gene family 1 protein in *Trypanosoma cruzi*. *Infect. Immun.* 2010, 78, 231–241.

15. Ouaissi, M.A. Role of the RGD sequence in parasite adhesion to host cells. *Parasitol. Today* 2018, 4, 169–173.

16. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* 2021, 596, 583–589.

17. Baek, M.; DiMaio, F.; Anishchenko, I.; Bauparas, J.; Ovchinnikov, S.; Lee, G.R.; Wang, J.; Cong, Q.; Kinch, L.N.; Schaeffer, R.D. Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 2021, 373, 871–876.

18. Bhave, G.; Nadin, B.M.; Brasier, D.J.; Glauner, K.S.; Heinemann, G.F.; Farzana, K.; Karim, F.; Gereau, R.W. Membrane Topology of a Metabotropic Glutamate Receptor. *J. Biol. Chem.* 2003, 278, 30294–30301.

19. von Heijne, G. Membrane-protein topology. *Nat. Rev. Mol. Cell Biol.* 2006, 7, 909–918.

20. Rawal, K.; Sinha, R.; Abbasi, B.A.; Chaudary, A.; Nath, S.K.; Kumari, P.; Preeti, P.; Saraf, D.; Singh, S.; Mishra, K.; et al. Identification of vaccine targets in pathogens and design of a vaccine using computational approaches. *Sci. Rep.* 2021, 11, 17626.

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. Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Curr. Biol.* 2016, 26, 161–172.

22. Olivares, M.; Thomas, M.C.; Lopez-Barajas, A.; Requena, J.M.; Garcias-Perez, J.L.; Alonso, C.; Lopez, M.C. Genome clustering of the *Trypanosoma cruzi* non-long terminal L1Tc retrotransposon with defined intersperse repeated DNA elements. *Electrophoresis* 2000, 21, 2973–2982.

23. Callejas-Hernández, F.; Rastrojo, A.; Poveda, C.; Girones, N.; Fresno, M. Genomic assemblies of newly sequenced *Trypanosoma cruzi* strains reveal new genomic expansion and greater complexity. *Sci. Rep.* 2018, 8, 14631.

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.; et al. Repeat-driven Generation of antigenic diversity in a major human pathogen, *Trypanosoma cruzi*. *Front. Cell. Infect. Microbiol.* 2021, 11, 614665.

25. Berná, L.; Rodriguez, M.; Chiribao, M.L.; Parodi-Talice, A.; Pita SRijo, G.; Alvarez-Valin, F.; Robello, C. Expanding an expanded genome: Long-read sequencing of *Trypanosoma cruzi*. *Microb. Genomics*. 2018, 4, e000177.
26. Ramirez, J.L. An Evolutionary View of *Trypanosoma cruzi* Telomeres. *Front. Cell. Infect. Microbiol.* 2020, 9, 439.
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. Replication origin location might contribute to genetic variability in *Trypanosoma cruzi*. *BMC Genom.* 2020, 21, 414.
28. Saha, A.; Nanavaty, V.P.; Li, B. Telomere and Subtelomere R-loops and Antigenic variation in Trypanosomes. *J. Mol. Biol.* 2019, 41, 4167.

---

Retrieved from <https://encyclopedia.pub/entry/history/show/95141>