Search for a platelet-activating factor receptor in the *Trypanosoma cruzi* proteome: a potential target for Chagas disease chemotherapy

Daniel Fábio Kawano[1]/+, Vinicius Barreto da Silva[1], Daniel Macedo de Melo Jorge[2], Carlos Henrique Tomich de Paula da Silva[1], Ivone Carvalho[1]

[1]Departamento de Ciências Farmacêuticas, Faculdade de Ciências Farmacêuticas de Ribeirão Preto [2]Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Av. do Café s/n, 14040-903 Ribeirão Preto, SP, Brasil

Chagas disease (CD) causes the highest burden of parasitic diseases in the Western Hemisphere and is therefore a priority for drug research and development. Platelet-activating factor (PAF) causes the CD parasite Trypanosoma cruzi *to differentiate, which suggests that the parasite may express PAF receptors. Here, we explored the* T. cruzi *proteome for PAF receptor-like proteins. From a total of 23,000 protein sequences, we identified 29 hypothetical proteins that are predicted to have seven transmembrane domains (TMDs), which is the main characteristic of the G protein-coupled receptors (GPCRs), including the PAF receptor. The TMDs of these sequences were independently aligned with domains from 25 animal PAF receptors and the sequences were analysed for conserved residues. The conservation score mean values for the TMDs of the hypothetical proteins ranged from 31.7-44.1%, which suggests that if the putative* T. cruzi *PAF receptor is among the sequences identified, the TMDs are not highly conserved. These results suggest that* T. cruzi *contains several GPCR-like proteins and that one of these GPCRs may be a PAF receptor. Future studies may further validate the PAF receptor as a target for CD chemotherapy.*

Key words: PAF - GPCR - *Trypanosoma cruzi* - Chagas disease - bioinformatics - proteome

Chagas disease (CD), or American trypanosomiasis, is a tropical parasitic disease that affects approximately 10 million people in Latin America (WHO 2010). CD is caused by the flagellated protozoan *Trypanosoma cruzi*, which is often transmitted to humans and other mammals by blood-sucking insects, such as the triatomines (Urbina et al. 2003). Transmission can also occur via blood transfusion (Young et al. 2007), organ transplantation (Kun et al. 2009), ingestion of contaminated foods or fluids (Benchimol Barbosa 2006) or congenital acquisition (Dorn et al. 2007). CD kills approximately 10,000 people every year (WHO 2010) and has a direct impact on workforce productivity in Latin America (Conteh et al. 2010).

Although CD causes the highest burden of parasitic diseases in the Western Hemisphere (Bern & Montgomery 2009), current treatments are highly toxic and often ineffective, particularly for the chronic stage of the disease (Trouiller et al. 2002). Strikingly, no drug has emerged as an effective candidate for clinical trials in the last 30 years (Moreira et al. 2009). Therefore, the identification of an enzyme or receptor that is essential for parasite growth and/or survival in vivo may facilitate the development of safer and more effective drugs (Wang 1997).

Platelet-activating factor (PAF) is a potent lipid mediator that plays a role in several human physiological processes, including inflammation, platelet aggregation and immune response (Harayama et al. 2008). In *T. cruzi*, PAF proved to trigger cell differentiation to the infective and replicative parasite forms, suggesting the existence of specific molecular recognition sites that can be explored as potential drug targets (Rodrigues et al. 1996, Ambrósio et al. 2003). This hypothesis is supported by the fact that PAF effects in *T. cruzi* are totally abrogated in the presence of the PAF receptor antagonist WEB 2086, possibly acting via a seven transmembrane domain (TMD) receptor, a G protein-coupled receptor (GPCR), as observed in mammals (Rodrigues et al. 1996). Additional studies also demonstrated that PAF was only active on intact parasite cells, suggesting the existence of PAF-mediated intracellular signalling mechanisms (Rodrigues et al. 1999).

Although endogenous PAF has been identified in *T. cruzi* and immunofluorescence assays indicate that PAF binding sites are present on the parasite membrane, attempts to isolate *T. cruzi* PAF receptors have failed (Gomes et al. 2006). We hypothesised that these PAF receptors proteins are encoded by the *T. cruzi* genome, which has already been sequenced (Atwood 3rd et al. 2005).

Recent advances in bioinformatics have facilitated the discovery of new drug targets (Chen & Chen 2008, Alves-Ferreira et al. 2009). Here, we evaluated bioinformatic approaches that were previously used to discover new GPCRs in fungi, plants and invertebrates (Kulkarni et al. 2005, Gookin et al. 2008, Kamesh et al. 2008) to identify PAF receptor-like proteins in the *T. cruzi* proteome.

MATERIALS AND METHODS

The PAF receptor sequences from 25 different animal species were retrieved from Uniprot (uniprot.org/) and ENSEMBL (ensembl.org/) (Supplementary data).

Financial support: FAPESP
+ Corresponding author: dkawano@usp.br
Received 8 July 2011
Accepted 8 November 2011

The *T. cruzi* proteome was downloaded from TriTrypDB (tritrypdb.org/) (Aslett et al. 2010).

Phylogenetic analyses - The PAF receptor sequences were converted to Molecular Evolutionary Genetics Analysis (MEGA) file format with the MEGA 4.0 software (Tamura et al. 2007) and then aligned using Multiple Sequence Alignment by Log-Expectation (MUSCLE) (Edgar 2004). The phylogenetic analyses were performed with MEGA with the neighbour-joining algorithm (Saitou & Nei 1987). The bootstrap support values for the optimal trees were calculated using 1,000 replicates with heuristic search settings that were identical to the original search settings (Felsenstein 1985). The evolutionary distances were calculated with the Jones-Taylor-Thornton matrix-based method (Jones et al. 1992) with respect to the units of the number of amino acid (AA) mutations per site. The indels were treated as missing data.

PAF receptor conserved residue analysis - The residues that are totally conserved among PAF receptors were analysed with MEGA through multiple sequence alignments. The sequences that correspond to the seven membrane-spanning TMDs of the PAF receptors were inferred by alignment to the published human PAF receptor sequence (Gui et al. 2007).

ConSurf-conservation analysis (Ashkenazy et al. 2010) was performed using the tertiary structure of the human PAF receptor as the three-dimensional scaffold (PDB code 2B0X). Multiple sequence alignments of the 25 PAF receptors were computed using MUSCLE (Edgar 2004), and the Bayesian method (Mayrose et al. 2004) was used to calculate the conservation scores with the Jones-Taylor-Thornton matrix-based method (Jones et al. 1992) as a model for protein substitutions.

Search for PAF receptor homologues - The search for PAF receptor homologues was performed with the Basic Local Alignment Search Tool (BLAST)P algorithm (Altschul et al. 1990) in the National Center for Biotechnology Information (NCBI) BLAST package version 2.2.24. The search was optimised to detect orthologues by selecting soft filtering (chosen in the standalone NCBI version of BLAST with the -F ''m S'' option) and the Smith-Waterman algorithm was used to produce the final alignments and calculate the scores (-s T option) (Moreno-Hagelsieb & Latimer 2008). The *T. cruzi* proteome was used as query in BLASTP against PAF receptor sequences with an e-value threshold of 10 (default). To decrease the possibility of missing a putative PAF receptor sequence, we selected all sequences that were above the threshold value for conservational analyses, not just the reciprocal best hits, which are typically considered for orthologue detection.

T. cruzi hypothetical protein conservation analysis - The *T. cruzi* sequences that were obtained from the BLAST searches were classified by protein description and the ConPred II software was used to select proteins that have a predicted transmembrane topology. The program predicts TMDs based on a consensus approach by combining the results of several methods, including KKD, TMpred, TopPred II, DAS, TMAP, MEMSAT 1.8, SOSUI, TMHMM 2.0 and HMMTOP 2.0 (Arai et al. 2004). The theoretical molecular weights of these proteins were predicted using the Sequence Manipulation Suite tool (bioinformatics.org/sms2/protein_mw.html).

The sequences of hypothetical proteins that were predicted to contain seven TMDs were manually edited to separate the transmembrane regions, which were then independently aligned with the corresponding TMDs of the 25 PAF receptors using MUSCLE (Edgar 2004). A conservation score for each AA position in the TMD multiple sequence alignment was computed with the SCORECONS web server with default parameters (ebi.ac.uk/thornton-srv/databases/cgi-bin/valdar/scorecons_server.pl) (Valdar 2002). The method generates conservation scores for individual residues based on normalised substitution matrices or multiple alignments of a set of sequences (Valdar 2002). To reduce the apparent weight assigned to the PAF receptor sequences in the final score, we highlighted conservation in each AA position of the hypothetical proteins and rescaled the scores between 0-99, as previously proposed by Haubertin et al. (2006): $Rscore = 99 - [99 \times (S - Lscore)/\Delta score]$, where *Rscore* is the rescaled score, *S* is the score calculated with SCORECONS, *Lscore* is the lowest score for a given AA position and $\Delta score$ is the amplitude of all of the scores for the AA residue. The *Rscores* that were derived from gapped positions in the *T. cruzi* hypothetical protein sequences with respect to the scaffold were defined as zero. With the rescaled scores from each AA residue, the weighted means were calculated for each TMD using the following formula, $\bar{x} = \sum(x.w)/\sum w$, where $\bar{x}$ is the weighted mean, *x* the *Rscore* value for a given AA residue and *w* is the corresponding conservation weight (score) calculated with SCORECONS when only the multiple sequence alignment of the 25 PAF receptor sequences was used as the input. The final conservation scores were calculated by the weighted mean of the rescaled scores regarding all AA residues in the seven TMDs predicted for the given hypothetical protein.

RESULTS

We assessed the evolutionary relationship of PAF receptors from 25 different animal species (Supplementary data). To the best of our knowledge, these are the only PAF receptor sequences that are currently available. The phylogenetic tree (Fig. 1) clearly indicates that these PAF receptors are split into two main branches. One cluster contains the mammals and the amphibian/bird clades and the second cluster exclusively contains fish. The mammalian lineage can be further subdivided into two branches, where the opossum and platypus are separated from the mammals. The human, orangutan and chimpanzee PAF receptors were sub-grouped together.

The multiple sequence alignment of the PAF receptors (Supplementary data) shows that 77 AAs are completely conserved among the 25 species. The majority of these residues (approximately 82%) are located in the TMDs (Fig. 2A, Table I). When residues that are partially conserved among species are analysed (shown in pink in Fig. 2B), the alignment clearly shows that these residues are also in the TMDs, except for the fourth domain, which is the most variable TMD (Table I).

To identify *T. cruzi* PAF receptor protein sequences, the 23,031 total *T. cruzi* protein sequences available in TriTrypDB (Aslett et al. 2010) were compared to other the 25 PAF receptor sequences via a BLAST search with an *E*-value threshold of 10. A total of 264 parasite protein sequences were selected, of which 138 were identified as potential PAF receptor proteins. A comparison of these hypothetical proteins with the PAF receptors yields similar BLAST best score values that range from 17-29. The hypothetical protein Tc00.1047053511507.60 presented the highest number of BLAST hits: 24 from the 25 PAF receptor sequences (Table II).

Of the 138 hypothetical *T. cruzi* proteins identified, 29 sequences were predicted to contain seven TMDs with chain lengths that ranged from 305-1,410 AA. TMD sequence of these hypothetical proteins was analysed for AA conservation and the resulting mean values for the seven TMDs ranged from 31.7-44.1%. When the TMDs were analysed independently, no *T. cruzi* hypothetical protein exhibited a weighted mean value above 60% (Table III).

DISCUSSION

Several groups have previously proposed that PAF exerts a biological effect on *T. cruzi* by binding to transmembrane receptors (Rodrigues et al. 1996, Ambrósio et al. 2003, Gomes et al. 2006). PAF receptors are GPCRs, which are transmembrane signalling proteins that are evolutionarily conserved among eukaryotic organisms (Römpler et al. 2007). The occurrence of GPCRs and G-protein signalling dates back 1.2 billion years, which was

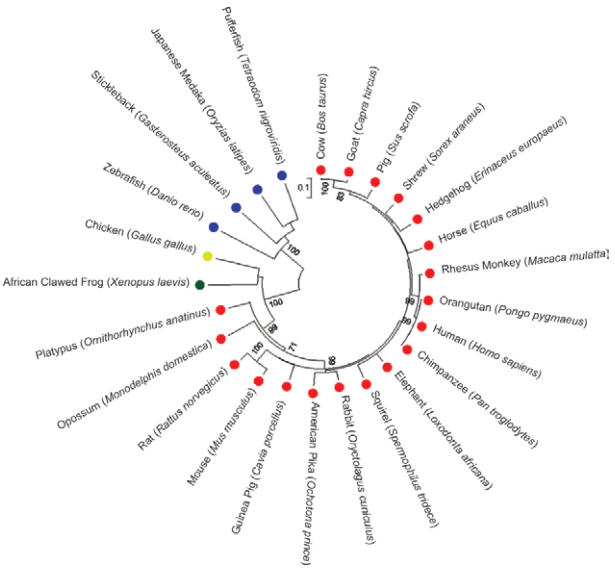

Fig. 1: evolutionary relationships of platelet-activating factor (PAF) receptors from 25 animal species. The evolutionary relationship of PAF receptors was inferred using the neighbour-joining method (Saitou & Nei 1987). The bootstrap consensus tree inferred from 1,000 replicates (Felsenstein 1985) is taken to represent the evolutionary history of the taxa analyzed. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Jones-Taylor-Thornton matrix-based method (Jones et al. 1992) and are in the units of the number of amino acid (AA) substitutions per site. The analysis involved 25 AA sequences. All positions containing gaps and missing data were eliminated. There were a total of 281 positions in the final dataset. Evolutionary analyses were conducted in Molecular Evolutionary Genetics Analysis 4.0 software (Tamura et al. 2007).

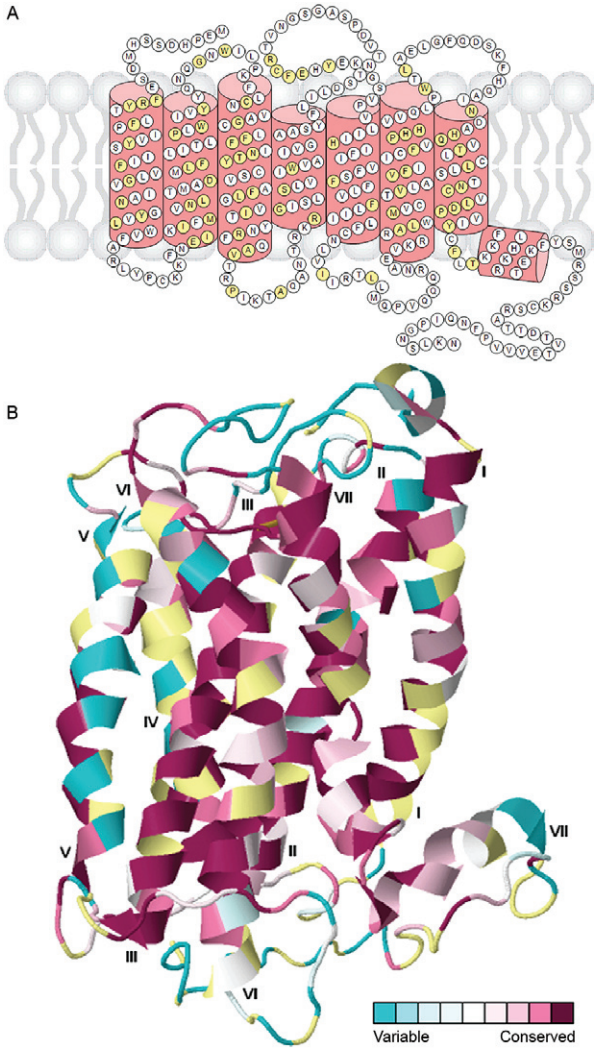

Fig. 2A: conservation pattern of platelet-activating factor (PAF) receptors from 25 animal species. Residues totally conserved among receptors are highlighted in yellow. Multiple sequence alignments were performed using Multiple Sequence Alignment by Log-Expectation (MUSCLE) (Edgar 2004) and the analysis of conserved residues was carried out in Molecular Evolutionary Genetics Analysis (MEGA) 4.0 software (Tamura et al. 2007). The two-dimensional representation of human PAF receptor was adapted from Gui et al. (2007); B: conservation profile for the tertiary structure of PAF receptor (PDB code 2B0X) using ConSurf-conservational analysis. Protein was visualized using FirstGlance in Jmol with colour-coded conservation scores. The conserved and variable residues are coloured according to the conservation scores. Sites, for which the inferred conservation level was assigned with low confidence, are coloured light yellow. Multiple sequence alignments of the 25 PAF receptors were computed using MUSCLE and the Bayesian method (Mayrose et al. 2004) was applied for the calculations of the conservation scores using the Jones-Taylor-Thornton matrix-based method (Jones et al. 1992) as the model of substitution for proteins.

TABLE I

Conserved residues in the transmembrane domains of platelet-activating factor (PAF) receptors from 25 animal species analyses of residues that are totally conserved among PAF receptors were carried out in Molecular Evolutionary Genetics Analysis (Tamura et al. 2007) and residues partially conserved were computed using ConSurf-conservation analysis (Ashkenazy et al. 2010)

Domain	Residues totally conserved (n)	Residues partially conserve (n)	Total of residues (n)
I	10	18	30
II	12	23	30
III	13	21	33
IV	4	7	23
V	3	11	26
VI	10	24	31
VII	11	24	31

before plants, fungi and animals emerged from a common ancestor (Schöneberg et al. 2007). Thus, it seems reasonable that *T. cruzi* may also express a PAF receptor.

Current theory regarding receptor structure and function evolution suggests that the structural diversity of the receptor between different species is a result of evolution, which is characterised by the continuous accumulation of mutations. However, if receptor function is essential for survival of the organism, some structural conservation is required to ensure protein function (Schöneberg et al. 2007).

Our studies regarding PAF receptor evolution indicate that the TMDs have the highest density of conserved residues. This observation is consistent with molecular modelling studies, which suggest that the ligand-binding site of the human PAF receptor is fundamentally composed of TMD residues (Gui et al. 2007). Moreover, the TMDs tend to be conserved among GPCRs, while the extra and intracellular loop regions and *C*-terminus are typically more divergent (Herz & Thomsen 2007).

Based on similarity to known PAF receptor protein sequences, we identified 29 hypothetical seven TMD proteins in the *T. cruzi* proteome. To the best of our knowledge, these proteins represent the first putative GPCRs identified in this parasite. Although the evolutionary origins of the encoding genes remain unknown, previous studies have indicated that horizontal gene transfer (the process where genetic material is transferred between distinct evolutionary lineages) has occurred in parasitic protozoa (Richards et al. 2003); moreover, viral infection may explain the origin of several mammalian genes, including genes that code for membrane receptors (Yu et al. 1996).

An alternative hypothesis is that putative GPCRs have been present since the beginning of trypanosome evolution. For example, the discovery of specific GPCRs in the protozoan *Dictyostelium discoideum* that were previously thought to be only in animals suggests that these genes arose before the divergence of animals and fungi and were subsequently lost in fungi (Prabhu & Eichinger 2006).

Our evolutionary analysis of previously identified PAF receptors indicates that the fish receptor sequences are grouped into a distinct clade from the other PAF receptors. *Danio rerio*, *Gasterosteus aculeatus*, *Oryzias latipes* and *Tetraodon nigroviridis* are teleost fishes, some of the most primitive vertebrates (Verburg-van Kemenade et al. 2009), thus pointing to the use of a PAF receptor sequence from a fish (instead of from a human) as the reference to calculate the percentages of un-gapped AA residues in the conservation analyses of hypothetical proteins. However, in view of the considerable evolutionary distance between GPCRs from animals and their putative counterparts in *T. cruzi* (Wilkie & Kinch 2005), we believe that the advantage of using a PAF receptor sequence from a fish as the reference could be assumed as insignificant. Besides, the fact that both PAF and its antagonist WEB 2086 are active in human and *T. cruzi* (Rodrigues et al. 1996) suggests some conservation between these species, but the same cannot be stated for fishes and *T. cruzi*.

A comparison of the identified *T. cruzi* seven TMD proteins and the human PAF receptor sequence resulted in relatively similar low BLAST scores. However, the BLAST algorithm typically does not provide high scores among phylogenetically distant species because genomes with a large evolutionary distance typically have a low protein sequence similarity (Kim et al. 2008). Thus, it would be optimal to use PAF receptor sequences from more primitive organisms, such as *Dictyostelium* or *Entamoeba*, which have genes that encode GPCRs (Wilkie & Kinch 2005). However, because no PAF receptor sequence from a lower eukaryote is currently available, a more precise relationship cannot be established.

Conservation analyses of the TMD AAs of the hypothetical *T. cruzi* proteins did not allow us to identify a protein sequence that is orthologous to the human PAF receptor. Although the weighted mean conservation scores that we calculated cannot be assumed to be absolute percentages of AA conservation relative to the PAF receptors, we expect that a putative PAF receptor orthologue would exhibit a higher number of conserved motifs; consequently, this should be expressed as a considerably higher conservation score compared to the remaining non-orthologous proteins. Nevertheless, if the putative *T. cruzi* PAF receptor is among the sequences we identified, this receptor does not show a high degree of conservation in the TMDs compared to animal PAF receptors. This is apparent in the protozoan *D. discoideum*, where the seven TMD receptors show little sequence similarity to the corresponding GPCRs in animals (Lagerström & Schiöth 2008).

The 29 putative seven TMD proteins identified in this study represent a starting point for additional biochemical, physiological and molecular biology studies

TABLE II

Hypothetical proteins of *Trypanosoma cruzi* detected via Basic Local Alignment Search Tool (BLAST) search against platelet-activating factor (PAF) receptors from 25 animal species prediction of transmembrane domains (TMDs) was carried out with ConPred II software (Arai et al. 2004)

	BLAST						
Hypothetical protein	Hits	Best score	e-value	PAF	Length (AA)	Weight (kDa)	TMDs
Tc00.1047053511573.49	3	29	0.002	Rat	235	27.26	6
Tc00.1047053509027.50	19	27	0.041	Rat	534	59.23	6
Tc00.1047053503487.70	17	26	0.031	Elephant	382	41.66	8
Tc00.1047053511127.300	2	26	0.071	Squirrel	534	59.42	6
Tc00.1047053510493.10	5	25	0.057	Mouse	240	25.82	6
Tc00.1047053508801.10	12	25	0.10	Opossum	372	41.71	8
Tc00.1047053508641.200	5	24	0.14	Elephant	305	32.55	7
Tc00.1047053506869.20	3	24	0.20	Platypus	458	52.67	8
Tc00.1047053510519.100	3	24	0.20	Platypus	458	52.64	8
Tc00.1047053508815.80	3	24	0.23	Chicken	570	64.11	9
Tc00.1047053510311.20	13	24	0.30	Frog	801	91.08	6
Tc00.1047053508873.530	10	24	0.35	Platypus	1,071	115.86	6
Tc00.1047053511629.40	4	23	0.12	Mouse	240	25.84	6
Tc00.1047053511285.80	16	23	0.18	Pika	247	27.23	6
Tc00.1047053509649.40	16	23	0.19	Pika	247	27.25	6
Tc00.1047053508547.140	1	23	0.26	Zebrafish	482	53.43	8
Tc00.1047053507943.70	3	23	0.27	Medaka	405	43.59	9
Tc00.1047053506681.24	8	23	0.35	Horse	341	38.37	8
Tc00.1047053508215.6	4	23	0.40	Platypus	494	55.71	7
Tc00.1047053506925.450	3	23	0.40	Opossum	681	77.39	6
Tc00.1047053503463.20	4	23	0.41	Platypus	494	55.69	7
Tc00.1047053510725.70	4	23	0.46	Opossum	486	54.09	6
Tc00.1047053509099.130	16	23	0.50	Human	822	91.69	8
Tc00.1047053510885.60	19	23	0.72	Guinea pig	1,151	131.20	6
Tc00.1047053508159.26	3	23	1.1	Guinea pig	1,278	137.16	4
Tc00.1047053511183.315	2	23	1.1	Medaka	1,085	115.00	4
Tc00.1047053510305.79	10	23	1.2	Medaka	1,410	156.58	7
Tc00.1047053507941.110	3	22	0.44	Medaka	385	41.43	9
Tc00.1047053506825.24	7	22	0.48	Horse	341	38.44	7
Tc00.1047053510729.290	20	22	0.62	Platypus	335	37.98	7
Tc00.1047053506467.4	3	22	0.67	Stickleback	340	38.27	7
Tc00.1047053504137.150	8	22	0.68	Stickleback	340	38.52	7
Tc00.1047053505193.74	4	22	0.69	Stickleback	340	38.21	7
Tc00.1047053509075.31	3	22	0.70	Mouse	540	58.70	5
Tc00.1047053509561.70	11	22	0.72	Pika	306	34.20	6
Tc00.1047053510103.30	17	22	0.81	Opossum	594	65.97	5
Tc00.1047053506009.60	4	22	0.82	Guinea pig	416	45.27	6
Tc00.1047053506179.30	1	22	0.97	Chicken	600	66.99	8
Tc00.1047053507663.80	1	22	0.99	Chicken	600	66.77	8

↳

	BLAST						
Hypothetical protein	Hits	Best score	e-value	PAF	Length (AA)	Weight (kDa)	TMDs
Tc00.1047053509719.56	2	22	1.2	Zebrafish	724	82.56	7
Tc00.1047053511461.40	5	22	1.3	Pig	552	62.88	6
Tc00.1047053508535.30	4	22	1.3	Pig	553	63.02	7
Tc00.1047053508443.50	3	22	1.5	Pufferfish	630	70.30	5
Tc00.1047053509157.110	3	22	1.5	Pufferfish	630	70.46	4
Tc00.1047053507821.101	2	22	1.5	Rat	994	106.93	4
Tc00.1047053507209.60	16	22	1.6	Guinea pig	819	91.35	8
Tc00.1047053508153.920	3	22	1.6	Platypus	611	68.86	1
Tc00.1047053506529.150	1	22	1.8	Zebrafish	1,151	131.49	6
Tc00.1047053511921.71	9	22	1.9	Opossum	1,157	124.18	7
Tc00.1047053509849.30	4	22	2.2	Shrew	860	97.79	7
Tc00.1047053511185.150	1	22	3.2	Guinea pig	2,247	243.13	-
Tc00.1047053508891.80	1	22	4.5	Mouse	1,935	216.30	6
Tc00.1047053507615.60	2	21	0.76	Frog	284	32.16	6
Tc00.1047053508277.210	21	21	0.81	Stickleback	319	36.83	6
Tc00.1047053511495.20	4	21	0.81	Elephant	324	36.45	6
Tc00.1047053508489.20	4	21	0.82	Elephant	324	36.39	6
Tc00.1047053508799.290	4	21	0.83	Pufferfish	221	24.45	6
Tc00.1047053506753.30	19	21	0.87	Elephant	341	37.88	7
Tc00.1047053511305.40	4	21	0.88	Chicken	301	32.92	5
Tc00.1047053509945.24	22	21	0.89	Stickleback	319	36.79	6
Tc00.1047053507257.180	2	21	0.94	Medaka	327	35.92	5
Tc00.1047053510355.190	20	21	0.97	Elephant	341	37.92	7
Tc00.1047053511507.60	24	21	1.1	Shrew	311	35.11	7
Tc00.1047053507883.50	20	21	1.1	Pika	319	35.46	5
Tc00.1047053503897.120	11	21	1.1	Pika	306	34.20	5
Tc00.1047053510001.31	4	21	1.1	Medaka	317	34.99	6
Tc00.1047053503647.40	1	21	1.1	Chicken	409	45.33	9
Tc00.1047053508741.20	2	21	1.2	Medaka	319	34.70	9
Tc00.1047053506195.50	15	21	1.3	Pika	319	35.29	5
Tc00.1047053508041.20	12	21	1.4	Goat	528	60.25	6
Tc00.104705350997.73	3	21	1.4	Platypus	351	38.97	8
Tc00.1047053506297.340	12	21	1.5	Goat	528	60.22	6
Tc00.1047053507603.220	3	21	1.5	Chicken	444	49.64	6
Tc00.1047053506661.140	3	21	1.5	Pika	486	54.79	6
Tc00.1047053509269.19	7	21	1.7	Pufferfish	471	53.80	8
Tc00.1047053509429.280	5	21	1.9	Chicken	552	62.15	8
Tc00.1047053504021.90	1	21	2.3	Guinea pig	600	65.97	9
Tc00.1047053507083.100	3	21	2.5	Chicken	662	72.90	8
Tc00.1047053507677.60	1	21	2.5	Guinea pig	600	65.87	9
Tc00.1047053508317.70	4	21	2.6	Shrew	856	97.38	7
Tc00.1047053503735.30	1	21	2.6	Horse	844	95.54	7
Tc00.1047053509901.100	3	21	2.9	Goat	662	72.94	8

↳

	BLAST						
Hypothetical protein	Hits	Best score	e-value	PAF	Length (AA)	Weight (kDa)	TMDs
Tc00.1047053508121.15	2	21	3.1	Chicken	1,027	111.03	4
Tc00.1047053510507.60	2	21	4.0	Cow	1,130	127.09	5
Tc00.1047053503919.60	1	21	6.1	Squirrel	1,579	177.58	5
Tc00.1047053507723.70	1	21	9.1	Rabbit	2,960	326.30	-
Tc00.1047053509799.90	4	20	1.4	Frog	310	35.33	7
Tc00.1047053509431.10	5	20	1.5	Cow	297	33.22	5
Tc00.1047053508739.30	5	20	1.6	Pika	331	36.04	8
Tc00.1047053503909.90	23	20	1.7	Mouse	311	35.14	6
Tc00.1047053508241.104	2	20	1.7	Stickleback	295	33.58	6
Tc00.1047053511693.130	6	20	1.8	Squirrel	383	43.59	8
Tc00.1047053506565.29	2	20	2.4	Chicken	425	49.18	9
Tc00.1047053509245.29	1	20	2.4	Opossum	315	35.83	8
Tc00.1047053507969.30	6	20	2.5	Opossum	470	53.77	7
Tc00.1047053504029.90	4	20	2.7	Guinea pig	351	38.96	8
Tc00.1047053509799.100	7	20	2.8	Stickleback	340	38.66	7
Tc00.1047053511245.24	3	20	2.8	Pika	486	54.56	6
Tc00.1047053510311.80	4	20	3.3	Platypus	655	72.93	8
Tc00.1047053511801.4	1	20	3.7	Goat	634	70.95	6
Tc00.1047053511001.50	3	20	4.3	Shrew	494	54.09	7
Tc00.1047053503593.60	1	20	4.4	Medaka	654	72.40	5
Tc00.1047053506127.120	1	20	4.5	Medaka	654	72.17	5
Tc00.1047053511279.20	2	20	4.7	Pika	709	79.47	6
Tc00.1047053508723.24	1	20	4.9	Frog	559	63.22	7
Tc00.1047053509791.60	1	20	4.9	Frog	559	63.25	7
Tc00.1047053511923.71	10	20	5.0	Mouse	1,103	120.80	7
Tc00.1047053509897.132	9	20	5.4	Pika	1,129	119.93	5
Tc00.1047053508741.260	1	20	5.5	Elephant	814	90.46	8
Tc00.1047053511003.160	2	20	6.2	Rat	1,034	114.04	6
Tc00.1047053504033.140	1	20	6.4	Chicken	1,013	113.12	6
Tc00.1047053511787.30	1	20	6.9	Rat	1,267	134.46	7
Tc00.1047053510861.94	1	20	8.0	Horse	845	95.95	7
Tc00.1047053509899.70	1	20	9.0	Pufferfish	1,271	135.75	6
Tc00.1047053508347.140	1	19	2.8	Frog	284	32.08	6
Tc00.1047053504765.20	17	19	2.9	Pig	269	30.04	6
Tc00.1047053504137.140	4	19	3.6	Frog	310	35.44	7
Tc00.1047053505757.10	10	19	3.7	Rat	259	28.18	5
Tc00.1047053506669.10	10	19	3.7	Rat	259	28.18	5
Tc00.1047053510885.20	3	19	3.9	Horse	296	32.15	6
Tc00.1047053506529.110	2	19	3.9	Horse	296	32.08	6
Tc00.1047053511837.40	1	19	4.1	Medaka	290	32.68	8
Tc00.1047053511531.44	2	19	4.7	Platypus	405	44.45	5
Tc00.1047053509741.20	1	19	5.2	Stickleback	356	39.50	4
Tc00.1047053507681.50	1	19	5.3	Platypus	516	57.64	7

↳

Hypothetical protein	BLAST				Length (AA)	Weight (kDa)	TMDs
	Hits	Best score	e-value	PAF			
Tc00.1047053507509.80	2	19	5.7	Cow	573	63.50	6
Tc00.1047053506871.130	2	19	5.9	Cow	573	63.64	6
Tc00.1047053507395.10	8	19	6.4	Elephant	405	45.36	6
Tc00.1047053510087.30	7	19	6.4	Elephant	403	45.13	5
Tc00.1047053510265.14	1	19	7.5	Rabbit	655	72.97	8
Tc00.1047053511471.10	1	19	8.7	Rabbit	796	86.81	1
Tc00.1047053430895.10	11	18	4.0	Rat	259	28.24	5
Tc00.1047053509253.30	1	18	5.1	Stickleback	265	30.50	5
Tc00.1047053508207.120	2	18	6.0	Rat	224	25.28	5
Tc00.1047053511683.20	2	18	6.1	Medaka	318	34.60	9
Tc00.1047053507625.200	1	18	6.1	Stickleback	355	39.56	9
Tc00.1047053402857.10	1	18	7.9	Stickleback	321	34.93	8
Tc00.1047053506513.14	1	17	9.0	Guinea pig	222	25.39	6

AA: amino acid.

TABLE III

Conservation analysis of amino acid (AA) residues in the transmembrane domains (TMDs) of *Trypanosoma cruzi* hypothetical proteins performed in SCORECONS web server (Valdar 2002) conservation scores for each TMD are expressed as the weighted mean of rescaled scores for each AA position. Percentages of un-gapped AA residues in the seven TMDs were calculated in comparison to the sequence of human PAF receptor in the multiple sequence alignment containing each hypothetical protein and the sequence of 25 PAF receptors.

Hypothetical protein	Length (AA)	Un-gapped TMDs residues (%)	Conservation scores (%)							
			Mean (7 TMDs)	Domains						
				I	II	III	IV	V	VI	VII
Tc00.1047053508215.6	494	89.1	44.1	55.5	49.3	43.5	44.5	35.3	33.1	45.8
Tc00.1047053508723.24	559	83.7	42.9	39.1	38.3	52.4	50.7	44.3	38	38.7
Tc00.1047053509791.60	559	83.7	42.8	39.1	38.2	52.4	50.7	44.3	38	38.7
Tc00.1047053506753.30	341	87.8	42.8	51.1	36.4	41.9	46,3	42.7	28.4	52.5
Tc00.1047053505193.74	340	91.2	42.5	52.1	38.3	45.5	31.3	34.9	44.9	48.2
Tc00.1047053503463.20	494	87.1	42.5	44.7	48.6	43.5	44.5	36.3	33.1	45.8
Tc00.1047053504137.150	340	89.8	42.2	44.3	30.1	42.4	51.3	52.5	47.8	31
Tc00.1047053506467.4	340	91.2	42.2	52.1	38.3	45.5	30.2	34.9	44.9	46.9
Tc00.1047053509799.100	340	89.8	42.1	44.3	30.1	42.4	51.3	52.5	46.7	31
Tc00.1047053508641.200	305	87.8	41.7	48.3	37.1	40.6	41.6	46.5	50.3	28.9
Tc00.1047053511787.30	1,267	89.1	41	29	38.6	36.6	37	51.1	43.7	51.2
Tc00.1047053510305.79	1,410	88.4	40.9	44.1	34.2	31.8	41.6	40	45.5	49.1
Tc00.1047053510861.94	845	95.2	40.8	39.4	35.9	34.7	47.5	42.4	41.4	45.5
Tc00.1047053511923.71	1,103	81.6	40.6	39.9	39.5	40.1	40.8	44.4	25.4	53.3
Tc00.1047053511921.71	1,157	89.8	40.6	39.9	44.9	38.6	30.4	50.5	24.4	53.3
Tc00.1047053503735.30	844	92.5	40.6	39.4	38.4	33	47.5	42.4	41.4	43.4

↳

Hypothetical protein	Length (AA)	Un-gapped TMDs residues (%)	Conservation scores (%)							
			Mean (7 TMDs)	Domains						
				I	II	III	IV	V	VI	VII
Tc00.1047053510355.190	341	87.8	40.5	51.1	36.4	25.9	46.3	41.1	32.2	50.8
Tc00.1047053509719.56	724	86.4	40.2	34.7	36.7	46	43.5	42.2	30.9	47.2
Tc00.104705350831 7.70	856	86.4	40.1	31.4	38.6	57.3	43	40.4	40.2	30.3
Tc00.1047053509849.30	860	86.4	38.6	47.1	31	38.4	44.6	40.4	40.2	30.3
Tc00.1047053510729.290	335	90.5	38.5	48.7	31.7	43.2	41.7	33.8	36.8	33.7
Tc00.1047053504137.140	310	90.5	38.2	41.9	28.3	44.1	31	36.5	32.8	50.7
Tc00.1047053509799.90	310	93.2	37.7	47.9	28.3	44.1	32.5	33	32.8	44.1
Tc00.1047053507969.30	470	93.2	37.7	36.5	27.2	59.4	39.6	31.8	34.3	35.3
Tc00.1047053511001.50	494	86.4	37.7	33.2	34.6	26.4	41.2	34.9	43.4	50.3
Tc00.1047053507681.50	516	89.8	36.7	48.1	26.2	34.3	34.9	32	41.8	39.4
Tc00.1047053506825.24	341	87.1	34.7	40.1	28.1	25.9	38	42.6	31.3	38
Tc00.1047053511507.60	311	85.7	34.5	31.7	27.4	33.4	36.4	43.8	35.1	35.2
Tc00.1047053508535.30	553	87.1	31.7	39.5	19.5	10.7	42.4	43.4	31.7	38

to identify surface proteins in *T. cruzi*. None of these 29 sequences or any other sequence derived from hypothetical proteins in the parasite proteome display a high degree of similarity with the human PAF receptor. However, the identification of the *T. cruzi* receptor may theoretically enable the design of PAF antagonists that interact with the *T. cruzi*, but not human PAF receptor. Molecules that affect the growth/differentiation of parasites can subsequently be tested as selective and effective drugs for the treatment of neglected diseases (Hammarton et al. 2003).

The limitations of this study include the ability of the BLAST algorithm to correctly identify the putative PAF receptor sequence and the ability to correctly predict TMD structures. To minimise these limitations, we selected a BLASTP algorithm that has been optimised to detect orthologues (Moreno-Hagelsieb & Latimer 2008) and used a consensus approach to properly detect the TMDs (Arai et al. 2004).

Additional limitations of this work may be related to the nature of the proteomic data analysed. In contrast to the kinetoplastid pathogens *Leishmania major* and *Trypanosoma brucei*, the *T. cruzi* assembled genome sequences are considered to be highly repetitive because the CL Brener reference strain is a hybrid of the "non-Esmeraldo-like" and "Esmeraldo-like" lineages; consequently, each chromosome is comprised of two homologous chromosomes (Alves-Ferreira et al. 2009, Weatherly et al. 2009). Therefore, the predicted *T. cruzi* proteome has some inherent restrictions regarding computational gene prediction and annotation that may hamper the identification of a potential unpredicted and/or un-annotated PAF receptor. Additional studies regarding the identification of PAF receptor homologues in the complete *T. cruzi* genome (not only against the predicted proteome) are needed, as they may overcome these limitations.

ACKNOWLEDGEMENTS

To Prof. Marcelo Dias Baruffi, for the continuous support throughout the course of this project.

REFERENCES

Altschul SF, Gish W, Miller W, Meyers EW, Lipman DJ 1990. Basic local alignment search tool. *J Mol Biol 215*: 403-410.

Alves-Ferreira M, Guimarães ACR, Capriles PVSZ, Dardenne LE, Degrave WM 2009. A new approach for potential drug target discovery through *in silico* metabolic pathway analysis using *Trypanosoma cruzi* genome information. *Mem Inst Oswaldo Cruz 104*: 1100-1110.

Ambrósio L, Aliberti JC, Guaratini T, Lopes NP, Machado JBM, Kato MJ, Tedesco AC, Yamauchi LML, Silva JS, Dias-Baruffi M 2003. Interaction between platelet-activating factor (PAF) and *Trypanosoma cruzi*: biological action and molecular aspects of a putative receptor. *Rev Inst Med Trop 45*: 74-75.

Arai M, Mitsuke H, Ikeda M, Xia J-X, Kikuchi T, Satake M, Shimizu T 2004. ConPred II: a consensus prediction method for obtaining transmembrane topology models with high reliability. *Nucleic Acids Res 32*: 390-393.

Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res 38*: W529-W533.

Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP, Fischer S, Gajria B, Gao X, Gardner MJ, Gingle A, Grant G, Harb OS, Heiges M, Hertz-Fowler C, Houston R, Innamorato F, Iodice J, Kissinger JC, Kraemer E, Li W, Logan FJ, Miller JA, Mitra S, Myler PJ, Nayak V, Pennington

C, Phan I, Pinney DF, Ramasamy G, Rogers MB, Roos DS, Ross C, Sivam D, Smith DF, Srinivasamoorthy G, Stoeckert JR CJ, Subramanian S, Thibodeau R, Tivey A, Treatman C, Velarde G, Wang H 2010. TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res 38*: 457-462.

Atwood 3rd JA, Weatherly DB, Minning TA, Bundy B, Cavola C, Opperdoes FR, Orlando R, Tarleton RL 2005. The *Trypanosoma cruzi* proteome. *Science 309*: 473-476.

Benchimol Barbosa PR 2006. The oral transmission of Chagas disease: an acute form of infection responsible for regional outbreaks. *Int J Cardiol 112*: 132-133.

Bern C, Montgomery SP 2009. An estimate of the burden of Chagas disease in the United States. *Clin Infect Dis 49*: 52-54.

Chen YPP, Chen F 2008. Identifying targets for drug discovery using bioinformatics. *Expert Opin Ther Targets 12*: 383-389.

Conteh L, Engels T, Molyneux DH 2010. Socioeconomic aspects of neglected tropical diseases. *Lancet 375*: 239-247.

Dorn P, Perniciaro L, Yabsley MJ, Roellig DM, Balsamo G, Diaz J, Wesson D 2007. Autochthonous transmission of *Trypanosoma cruzi*. *Emerg Infect Dis 13: 505-507.*

Edgar RC 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res 32*: 1792-1797.

Felsenstein J 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution 39*: 783-791.

Gomes MT, Monteiro RQ, Grillo LA, Leite-Lopes F, Stroeder H, Ferreira-Pereira A, Alviano CS, Barreto-Bergter E, Neto HC, Cunha e Silva NL, Almeida IC, Soares RM, Lopes AH 2006. Platelet-activating factor-like activity isolated from *Trypanosoma cruzi*. *Int J Parasitol 36*: 165-173.

Gookin TE, Kim J, Assmann SM 2008. Whole proteome identification of plant candidate G-protein coupled receptors in *Arabidopsis*, rice and poplar: computational prediction and *in vivo* protein coupling. *Genome Biol 9*: RI20.1-RI20.26.

Gui C, Zhu W, Chen G, Luo X, Liew OW, Puah CM, Chen K, Jiang H 2007. Understanding the regulation mechanisms of PAF receptor by agonists and antagonists: molecular modeling and molecular dynamics simulation studies. *Proteins: Struct Funct Bioinf 67*: 41-52.

Hammarton TC, Mottram JC, Doerig C 2003. The cell cycle of parasitic protozoa: potential for chemotherapeutic exploitation. In L Meijer, A Jézéquel, M Roberge, *Progress in cell cycle research*, vol. V, Springer, New York, p. 91-101.

Harayama T, Shindou H, Ogasawara R, Suwabe A, Shimizu T 2008. Identification of a novel noninflammatory biosynthetic pathway of platelet-activating factor. *J Biol Chem 283*: 11097-11106.

Haubertin DY, Madaoui H, Sanson A, Guérois R, Orlowsk S 2006. Molecular dynamics simulations of *E. coli* MsbA transmembrane domain: formation of a semipore structure. *Biophys J 91*: 2517-2531.

Herz JM, Thomsen WJ 2007. Receptors for drugs: discovery in the post-genomic era. In J Swarbrick, *Encyclopedia of pharmaceutical technology*, vol. V, 3rd ed., Informa Health Care, New York, p. 3108-3127.

Jones DT, Taylor WR, Thornton JM 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci 8*: 275-282.

Kamesh N, Aradhyam GK, Manoj N 2008. The repertoire of G protein-coupled receptors in the sea squirt *Ciona intestinalis*. *BMC Evol Biol 8*: 129.

Kim S, Kang J, Chung YJ, Li J, Ryu KH 2008. Clustering orthologous proteins across phylogenetically distant species. *Proteins 71*: 1113-1122.

Kulkarni RD, Thon MR, Pan H, Dean RA 2005. Novel G-protein-coupled receptor-like proteins in the plant pathogenic fungus *Magnaporthe grisea*. *Genome Biol 6*: R24.1-R24.14.

Kun H, Moore A, Mascola L, Steurer F, Lawrence G, Kubak B, Radhakrishna S, Leiby D, Herron R, Mone T, Hunter R, Kuehnert M 2009. Transmission of Trypanosoma cruzi by heart transplantation. *Clin Infect Dis 48*: 1534-1540.

Lagerström MC, Schiöth HB 2008. Structural diversity of G protein coupled receptors and significance for drug discovery. *Nat Rev Drug Discov 7*: 339-357.

Mayrose I, Graur D, Ben-Tal N, Pupko T 2004. Comparison of site-specific rate-inference methods for protein sequences: empirical Bayesian methods are superior. *Mol Biol Evol 21*: 1781-1791.

Moreira DR, Leite AC, Dos Santos RR, Soares MB 2009. Approaches for the development of new anti-*Trypanosoma cruzi* agents. *Curr Drug Targets 10*: 212-231.

Moreno-Hagelsieb G, Latimer K 2008. Choosing BLAST options for better detection of orthologs as reciprocal best hits. *Bioinformatics 24*: 319-324.

Prabhu Y, Eichinger L 2006. The *Dictyostelium* repertoire of seven transmembrane domain receptors. *Eur J Cell Biol 85*: 937-946.

Richards TA, Hirt RP, Williams BA, Embley TM 2003. Horizontal gene transfer and the evolution of parasitic protozoa. *Protist 154*: 17-32.

Rodrigues CO, Dutra PML, Barros FS, Souto-Padron T, Meyer-Fernandes JR, Lopes AHCS 1999. Platelet-activating factor induction of secreted phosphatase activity in *Trypanosoma cruzi*. *Biochem Biophys Res Commun 266*: 36-42.

Rodrigues CO, Dutra PML, Souto-Padron T, Cordeiro RSB, Lopes AHCS 1996. Effect of platelet-activating factor on cell differentiation of *Trypanosoma cruzi*. *Biochem Biophys Res Commun 223*: 735-740.

Römpler H, Stäubert C, Thor D, Schulz A, Hofreiter M, Schöneberg T 2007. G protein-coupled time travel: evolutionary aspects of GPCR research. *Mol Interv 7*: 17-25.

Saitou N, Nei M 1987. The neighbour-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol 4*: 406-425.

Schöneberg T, Hofreiter M, Schulz A, Römpler H 2007. Learning from the past: evolution of GPCR functions. *Trends Pharmacol Sci 28*: 117-121.

Tamura K, Dudley J, Nei M, Kumar S 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol 24*: 1596-1599.

Trouiller P, Olliaro P, Torreele E, Orbinski J, Laing R, Ford N 2002. Drug development for neglected diseases: a deficient market and a public-health policy failure. *Lancet 359*: 2188-2194.

Urbina JA, Payares G, Sanoja C, Molina J, Lira R, Brener Z, Romanha AJ 2003. Parasitological cure of acute and chronic experimental Chagas disease using the long-acting experimental triazole TAK-187: activity against drug-resistant *Trypanosoma cruzi* strains. *Int J Antimicrob Ag 21*: 39-48.

Valdar WS 2002. Scoring residue conservation. *Proteins 48*: 227-241.

Verburg-van Kemenade BML, Stolte EH, Metz JR, Chadzinska M 2009. Neuroendocrine-immune interactions in teleost fish. In NJ Bernier, G Kraak, AP Farrell, CJ Brauner, *Fish neuronendocrinology*, vol. XXVIII, 1st ed., Academic Press, London, p. 313-364.

Wang CC 1997. Validating targets for antiparasite chemotherapy. *Parasitology 114*: 31-44.

Weatherly DB, Boehlke C, Tarleton RL 2009. Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. *BMC Genomics 10*: 1-13.

WHO - World Health Organization 2010. *Working to overcome the global impact of neglected tropical diseases: first WHO report on neglected tropical diseases*, WHO, Geneva, p. 1-172.

Wilkie TM, Kinch L 2005. New roles for Gα and RGS proteins: communication continues despite pulling sisters apart. *Curr Biol 15*: 843-854.

Young C, Losikoff P, Chawla A, Glasser L, Forman E 2007. Transfusion-acquired *Trypanosoma cruzi* infection. *Transfusion 47*: 540-547.

Yu DC, Wang AL, Wang CC 1996. Amplification, expression and packaging of a foreign gene by giardiavirus in *Giardia lamblia*. *J Virol 70*: 8752-8757.