

Accepted Manuscript

Investigating the structure-activity relationships of *N'*-[(5-nitrofur-2-yl) methylene] substituted hydrazides against *Trypanosoma cruzi* to design novel active compounds

Fanny Palace-Berl, Kerly Fernanda Mesquita Pasqualoto, Bianca Zingales, Carolina Borsoi Moraes, Mariana Bury, Caio Haddad Franco, Adelson Lopes da Silva Neto, João Sussumu Murayama, Solange Lessa Nunes, Marcelo Nunes Silva, Leoberto Costa Tavares



PII: S0223-5234(17)31016-4

DOI: [10.1016/j.ejmech.2017.12.011](https://doi.org/10.1016/j.ejmech.2017.12.011)

Reference: EJMECH 9989

To appear in: *European Journal of Medicinal Chemistry*

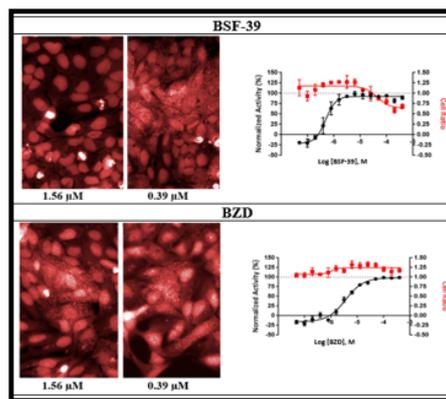
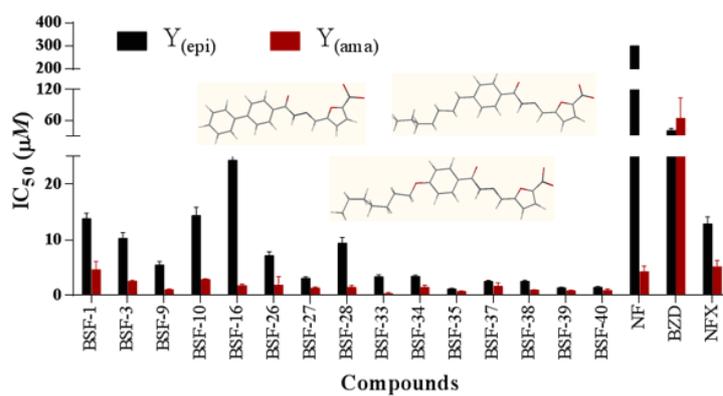
Received Date: 23 August 2017

Revised Date: 29 November 2017

Accepted Date: 2 December 2017

Please cite this article as: F. Palace-Berl, K.F.M. Pasqualoto, B. Zingales, C.B. Moraes, M. Bury, C.H. Franco, A.L. da Silva Neto, Joã.Sussumu. Murayama, S.L. Nunes, M.N. Silva, L.C. Tavares, Investigating the structure-activity relationships of *N'*-[(5-nitrofur-2-yl) methylene] substituted hydrazides against *Trypanosoma cruzi* to design novel active compounds, *European Journal of Medicinal Chemistry* (2018), doi: 10.1016/j.ejmech.2017.12.011.

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.



Investigating the structure-activity relationships of *N'*-[(5-nitrofur-2-yl) methylene] substituted hydrazides against *Trypanosoma cruzi* to design novel active compounds

Fanny Palace-Berl,^{a*} Kerly Fernanda Mesquita Pasqualoto,^b Bianca Zingales,^c Carolina Borsoi Moraes,^d Mariana Bury,^c Caio Haddad Franco,^d Adelson Lopes da Silva Neto,^a João Sussumu Murayama,^a Solange Lessa Nunes,^c Marcelo Nunes Silva,^c Leoberto Costa Tavares^a

^aDepartment of Biochemical and Pharmaceutical Technology, Faculty of Pharmaceutical Sciences, University of São Paulo, SP, Brazil; ^bInnovation and Industrial Development Laboratory, Butantan Institute, SP, Brazil; ^cDepartment of Biochemistry, Chemistry Institute, University of São Paulo, SP, Brazil; ^dLaboratório Nacional de Biotecnologia (LNBio), Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas, Brazil.

ABSTRACT

Chagas disease, caused by the protozoan *Trypanosoma cruzi*, is a neglected chronic tropical infection endemic in Latin America. New and effective treatments are urgently needed because the two available drugs - benznidazole (BZD) and nifurtimox (NFX) - have limited curative power in the chronic phase of the disease. We have previously reported the design and synthesis of *N'*-[(5-nitrofur-2-yl) methylene] substituted hydrazides that showed high trypanocidal activity against axenic epimastigote forms of three *T. cruzi* strains. Here we show that these compounds are also active against a BZD- and NFX-resistant strain. Herein, multivariate approaches (hierarchical cluster analysis and principal component analysis) were applied to a set of thirty-six formerly characterized compounds. Based on the findings from exploratory data analysis, novel compounds were designed and synthesized. These compounds showed two- to three-fold higher trypanocidal activity against epimastigote forms than the previous set and were 25 to 30-fold more active than BZD. Their activity was also evaluated against intracellular amastigotes by high content screening (HCS). The most active compounds

*Corresponding author address: Av. Prof. Lineu Prestes, 580, São Paulo, SP 05508-000, Brazil. Phone: +55 11 30913693; +55 11 3815 6386; Cel.: +1 519 830 6827; e-mail: palaceberlf@usp.br

(BSF-38 to BSF-40) showed a selective index (SI') greater than 200, in contrast to the SI' values of reference drugs (NFX, 16.45; BZD, > 3), and a 70-fold greater activity than BZD. These findings indicate that nitrofurans compounds designed based on the activity against epimastigote forms show promising trypanocidal activity against intracellular amastigotes, which correspond to the predominant parasite stage in the chronic phase of Chagas disease.

Keywords: Nitrofurans; *Trypanosoma cruzi*; amastigote intracellular forms; exploratory data analysis; chemometric approaches; Chagas disease; structure-activity relationships.

1 INTRODUCTION

Chagas disease (CD) is a potentially life-threatening illness to around eight million infected people in 21 Latin American countries. The disease is emergent in non-endemic countries introduced by extensive global migrations and perpetuated by means of congenital transmission [1].

The protozoan *Trypanosoma cruzi*, which is the etiological agent of CD, has three main developmental forms: the epimastigote, encountered in the gut of triatomine vectors and easily cultured in the laboratory; the trypomastigote, encountered in the blood of infected hosts and that has the capacity to invade mammalian cells; and the amastigote, which multiplies in the cytoplasm of infected cells. *T. cruzi* strains show a remarkable genotypic and phenotypic heterogeneity. At the present, the parasite strains are partitioned into six lineages or discrete typing units (DTUs), TcI–TcVI [2], which have distinct, but not exclusive ecological and epidemiological associations [3].

Current treatment options for CD are limited to two nitroheterocyclic drugs: benznidazole (BZD) and nifurtimox (NFX). Although both drugs are quite effective in curing patients in the acute phase and cases of congenital transmission, the efficacy of

both drastically diminishes in the chronic phase, which is the prevalent clinical presentation encountered [4]. The reasons for the marked difference in the drug activity in the two phases of CD are unclear, but they may be related to unfavorable pharmacokinetic properties, such as relatively short terminal half-life and reduced tissue penetration, which might limit their action on the intracellular amastigotes, which prevail in the chronic phase [5]. On the other hand, the acute phase is characterized by high parasitemia and abundance of trypomastigote bloodstream forms. Drug performance can also be affected by individual characteristics of the immunological response of the patient [6]. One additional factor implicated in the low rates of cure of BZD and/or NFX in some patients (acute or chronic phases) is the occurrence of *T. cruzi* strains naturally resistant to these compounds [7, 8]. It is worth noting that the resistance or sensitivity to the available drugs is an intrinsic characteristic of a given strain, and is verified in the three developmental stages of the strain (epimastigote, trypomastigote and amastigote) [9-11].

The information outlined above indicates the urgent need for efficacious new drugs active against all phases of CD and all parasite strains. In a previous study, we have reported the designing and synthesis of *N'*-[(5-nitrofuran-2-yl) methylene] substituted hydrazides against *T. cruzi*, primarily based on molecular modifications of nifuroxazide (NF), a nitroheterocyclic drug [12, 13] (Figure 1). In those studies, the anti-*T. cruzi* activity was assessed against epimastigote forms of three parasite strains representatives of DTUs TcI, TcII and TcV, which have high prevalence in patients of different Latin American regions [3]. Most of the synthesized compounds showed trypanocidal activity four- to ten-fold higher than BZD.

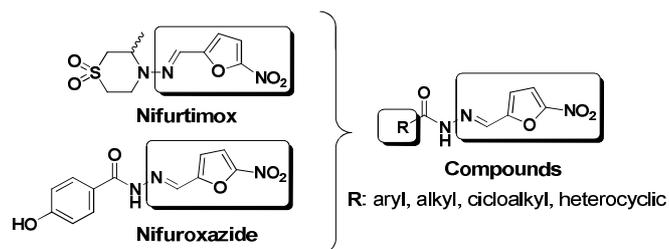


Figure 1. Nifurtimox (NFX) and nifuroxazide (NF; lead compound) chemical structures and structural scaffold of the designed compounds. Similar structural portions of NFX and NF (rectangles) were maintained in the designed compounds. The molecular modifications in R correspond to aryl, alkyl, cycloalkyl, and heterocyclic groups.

Herein, we have investigated the activity of the previously reported compounds [12, 13] against a fourth strain, Colombiana, which is considered resistant to both BZD and NFX [7]. Structure-activity relationships were established to a set of thirty-six N'-[(5-nitrofurane-2-yl) methylene] analogues employing multivariate chemometric techniques in order to design novel and more active compounds against *T. cruzi*. The explored compounds have in common the (5-nitrofurane-2-yl)methanamine moiety, that is similar to the pharmacophoric group of NFX (rectangles in Figure 1). The 5-nitrofurane group was disclosed as an important structural fragment for anti-trypanosome activity [14, 15]. In addition, nitroheterocyclic compounds have been reported as pro-drugs due to the nitro group reduction process that is required to undergo biological response [16, 17]. These information about the 5-nitrofurane group led us to further investigate N'-[(5-nitrofurane-2-yl) methylene] analogues against trypanosomatids.

Multivariate methods are often applied when more than one variable needs to be observed simultaneously. As complementary methods to recognize patterns in the investigated set of compounds, hierarchical cluster screening (HCS) and principal component analysis (PCA) were employed [18, 19]. Based on the findings of these

analyses, structure-activity relationships were established. This allowed the designing and synthesis of novel compounds that showed a considerable increase in trypanocidal activity against trypomastigotes and intracellular amastigote forms, the latter evaluated by high content screening (HCS).

2 RESULTS AND DISCUSSION

2.1 Biological activity against epimastigote forms of *T. cruzi*

The biological activity of thirty-six *N'*-[(5-nitrofur-2-yl) methylene] substituted hydrazides against the Colombiana strain, which belongs to *T. cruzi* DTU TcI [2], was determined, following protocols previously reported [13]. This activity is expressed as the inhibitory concentration IC₅₀, that corresponds to the drug concentration that inhibits 50% of the epimastigote growth (Table 1; See also Supplementary Material: Dose-response curve of anti-*T. cruzi* activity – epimastigote forms, Colombiana strain, p.S3). The activity of eighteen *N'*-[(5-nitrofur-2-yl) methylene] substituted hydrazides against Silvio X10 c11, Y and Bug 2149 c10 strains has been reported elsewhere [12, 13]. Compounds BSF-14 and BSF-31 showed low anti-epimastigote activity and their IC₅₀ values could not be determined even with the maximum concentration used in the assay, dictated by the compound solubility.

According to Table 1, eighteen compounds out of thirty-six (50%) were more active than NFX, and thirty-two compounds (89%) were more active than BZD for the Colombiana strain. The activity of the same compounds was evaluated against epimastigotes of the Silvio X10 c11 strain (also DTU TcI), indicating that four compounds out of thirty-six (11%) were more active than NFX and twenty-five compounds out of thirty-six (70%), more active than BZD [13] (See Supplementary Material: Table 1S, Biological activity of compounds against four strains of *T. cruzi*).

Comparison of the data of both strains evidences that the Colombiana strain is more susceptible to the compounds than the Silvio strain, in contrast to what is reported for the reference drugs, BZD and NFX, for which Silvio is more sensitive than Colombiana [7].

Furthermore, it is noteworthy that 4-lipophilic-substituted benzene compounds show high anti-parasitic activity. This feature could be related to the presence of a lipophilic pocket in a putative receptor or to facilitated membrane permeation. In contrast, hydrophilic substituents did not show increment of activity. Interestingly, compound BSF-23 [-C₆H₄-4-N(CH₃)₂] exhibited a huge loss of biological activity when compared to its bioisostere BSF-28 [-C₆H₄-4-*tert*-C₄H₉]. This characteristic was previously reported [13], and was mainly attributed to the solvent accessible surface area of hydrophobic atoms (ASA_H), which showed high values for the most active compound (BSF-28) and low values for less active (BSF-23). Thus, the hydrophobicity of 4-substituted benzene seems to be directly related to the anti-*T. cruzi* activity. These observations match with the computational analysis presented below.

In face of the promising anti-*T. cruzi* activity of the compounds and aiming at predicting more active compounds, the experimental data were used as dependent variables to establish qualitative structure-activity relationships.

Table 1. Biological activity of *N'*-[(5-nitrofurán-2-yl) methylene] substituted hydrazides against epimastigote forms of *T. cruzi* Colombiana strain.

Cpd	Substituents (R)	Colombiana (TcI) IC ₅₀ (μM) (mean ±SD)	Cpd	Substituents (R)	Colombiana (TcI) IC ₅₀ (μM) (mean ±SD)
BSF-1^a	-CH ₃	9.52 ±1.03	BSF-19^a	-C ₆ H ₄ -4-Cl	8.64 ±0.92
BSF-2	-CH ₂ CN	60.61 ±6.17	BSF-20^a	-C ₈ H ₁₇	8.07 ±0.78
BSF-3^a	-C ₅ H ₁₁	7.58 ±0.76	BSF-21^a	-C ₆ H ₄ -4- <i>iso</i> -C ₃ H ₇	6.08 ±0.65
BSF-4	-OC ₂ H ₅	7.07 ±0.74	BSF-22^a	-C ₆ H ₄ -4-C ₃ H ₇	7.25 ±0.75

BSF-5	2-furyl	13.14	±1.07	BSF-23^a	-C ₆ H ₄ -4-N(CH ₃) ₂	85.53	±9.36
BSF-6^a	cyclopentyl	9.67	±0.93	BSF-24	-C ₆ H ₄ -4-OC ₂ H ₅	11.55	±1.15
BSF-7^a	-O- <i>tert</i> -C ₄ H ₉	7.41	±0.73	BSF-25	-C ₆ H ₄ -3-NO ₂	40.64	±4.24
BSF-8^a	-C ₆ H ₄	11.30	±1.14	BSF-26	2-naphthyl	7.58	±0.28
BSF-9	2-thiophene	15.12	±1.32	BSF-27^a	-C ₆ H ₄ -4-C ₄ H ₉	2.45	±0.27
BSF-10^a	cyclohexyl	10.19	±0.78	BSF-28^a	-C ₆ H ₄ -4- <i>tert</i> -C ₄ H ₉	6.10	±0.64
BSF-11^a	-C ₆ H ₄ -3-CH ₃	8.53	±0.80	BSF-29^a	-C ₆ H ₄ -4-OC ₃ H ₇	5.26	±0.55
BSF-12^a	-CH ₂ -C ₆ H ₄	10.06	±0.99	BSF-30	-C ₆ H ₄ -3,4-OCH ₃	26.59	±2.63
BSF-13^a	-C ₆ H ₄ -4-CH ₃	11.92	±1.21	BSF-31	-C ₆ H ₄ -4-sulfamoyl	ND	
BSF-14	-C ₆ H ₄ -4-NH ₂	ND		BSF-32	-C ₆ H ₄ -3CF ₃	9.98	±1.06
BSF-15	-5-4-CH ₃ -1,2,3-thiadiazole	14.24	±1.45	BSF-33^a	-C ₆ H ₄ -4-NHC ₄ H ₉	3.94	±0.43
BSF-16^a	-C ₆ H ₄ -4-CN	20.24	±2.04	BSF-34^a	-C ₆ H ₄ -4-OC ₄ H ₉	2.86	±0.31
BSF-17	-C ₆ H ₄ -C ₂ H ₅	6.55	±0.69	BSF-35^a	-C ₆ H ₄ -4-C ₆ H ₄	3.06	±0.23
BSF-18^a	-C ₆ H ₄ -4-OCH ₃	8.08	±0.85	BSF-36^a	-C ₆ H ₄ -2-OC ₆ H ₄	5.79	±0.43
NF		187.78	±9.91				
BZD		47.91	±4.96				
NFX		9.93	±1.01				

a: The anti-*T. cruzi* activity of these compounds was previously reported against Silvio X10 c11, Y and Bug 2149 c110 strains [12, 13]. IC₅₀: compound concentration that causes a reduction of 50% of parasite growth, compared to controls. IC₅₀ values presented as the mean and the standard deviation (SD), correspond to triplicates from at least two independent experiments. Errors are in a range of 10%. ND: not determined in the maximum concentration used in the assay, without precipitation of compound. Cpd: compounds, NF: nifuroxazide, BZD: benznidazole, NFX: nifurtimox.

2.2 Molecular modelling approach, calculation and selection of descriptors

The crystallographic structure of NF was used as template to construct the three-dimensional (3D) molecular models of each compound in their neutral form, as described [12, 13]. The lowest-energy conformer of each compound was selected in the energy equilibrium region of the conformational ensemble profile (CEP) from molecular dynamics (MD) simulations. The root means square deviation (RMSD) was computed as a criterion to verify whether the structural integrity was maintained after simulations by comparing the atomic positions of the selected conformers from CEP and those of the energy-minimized structures (See Figure 1S, Supplementary Material). The RMSD values found for the lowest-energy conformer of all compounds and the NF template were lower than 0.07 Å. The pharmacophore (5-nitro-2-furan) and the *N*-

acylhydrazone moiety were considered in the structures overlay in order to emphasise the pharmacophore moiety structural integrity for the investigated compounds. Otherwise, higher RMSD values would indicate loss of structural integrity during MD simulations regarding the crystallographic structure used as template [20]. The total potential energy (E_{TOTAL}) was calculated considering the summation of the following energy contributions: contribution of axial strain energy (E_{STRETCH}), energy input of angular deformation (E_{BEND}), contribution of torsional strain energy (E_{TORS}), energy input of interactions of 1-4 type (E_{1-4}), energy input of van der Waals energy (E_{vdW}), contribution of electrostatic energy (E_{CHARGE}), including the intramolecular energy of solvation (E_{solv}) and the intramolecular energy of hydrogen bonding (E_{Hb}). The values obtained for each energy contribution and for total potential energy can be found in Table 2S of Supplementary Material. The total potential energy values ranged from 4.22 to 67.86 kcal.mol⁻¹.

Molecular properties, or descriptors, of different nature (thermodynamic, hydrophobic, electronic, topological, geometric, spectrograph and steric properties) were calculated for each compound, using the methods listed in Table 3S (Supplementary Material). The values of the calculated properties can be found in Table 4S (Supplementary Material).

The activity values (expressed as \log_1/IC_{50}) against epimastigote forms of the four strains were considered as the dependent variable. Because compounds BSF-14 and BSF-31 did not show activity against *T. cruzi*, they were discarded. Thirty-four compounds plus NF were considered to carry out the analysis ($n = 35$). The activity ranges were evaluated for each strain and, then, common intervals were assigned as follows: \log_1/IC_{50} values from 6.00 to 5.40 as highly active compounds (H); \log_1/IC_{50}

values from 5.40 to 4.60 as moderately active (M); $\log 1/IC_{50}$ values from 4.60 to 3.50 as low activity compounds (L).

A matrix composed by 35 rows (samples or compounds) and 56 columns (descriptors plus biological activity values) was used as input for studies of structure-property/activity relationships. The descriptors values were autoscaled [21] due to the different orders of magnitude among the molecular properties considered in the analysis. A preliminary variable selection was carried out using as criteria (i) the Pearson's linear correlation coefficient between descriptors and biological activity (cutoff = 0.35), and (ii) data distribution through scatter plots (biological activity, pIC_{50} values, versus each calculated molecular property or descriptor) [18]. These findings are presented in Figure 2S of Supplementary Material.

2.3 Exploratory data analysis (Hierarchical cluster analysis, HCA, and principal components analysis, PCA)

The unsupervised HCA aims at grouping a data set considering similar attributes. Euclidean distance and complete linking method were used to systematically cluster the samples [18, 22]. The similarity index (SI), generated by the data set, varies between 0 and 1, in which 1 corresponds to the maximum similarity. The similarity matrix generates the dendrogram, which is a map shaped tree constructed from the data distances [18, 19, 21]. According to HCA, 35 samples (34 compounds and NF) were grouped into two major clusters, A and B, sharing 33.9 % and 61.2 % similarity, respectively (Figure 2A). Cluster A grouped more hydrophilic compounds with low anti-trypanosomal activity. Compound BSF-23 [-C₆H₄-4-N(CH₃)₂] was grouped with BSF-21 [-C₆H₄-4-*iso*-C₃H₇] into cluster B, probably due to topological similarities. Compounds containing alkyl and heterocyclic substituents (BSF-1 to BSF-7; BSF-9,

BSF-10, BSF-12 and BSF-15) grouped mostly in cluster A. Only one alkyl-substituted compound, BSF-20 [-C₈H₁₇], was grouped into cluster B likely due to its bulky highly lipophilic substituent. The highly active compounds were grouped in the sub-cluster B' sharing 80.7 % similarity. Compound BSF-27 [-C₆H₄-4-C₄H₉] (highly active) was grouped with BSF-28 [-C₆H₄-4-*terc*-C₄H₉] (moderately active) into sub-cluster B'' probably because they are constitutional isomers, but they differ regarding molecular volume.

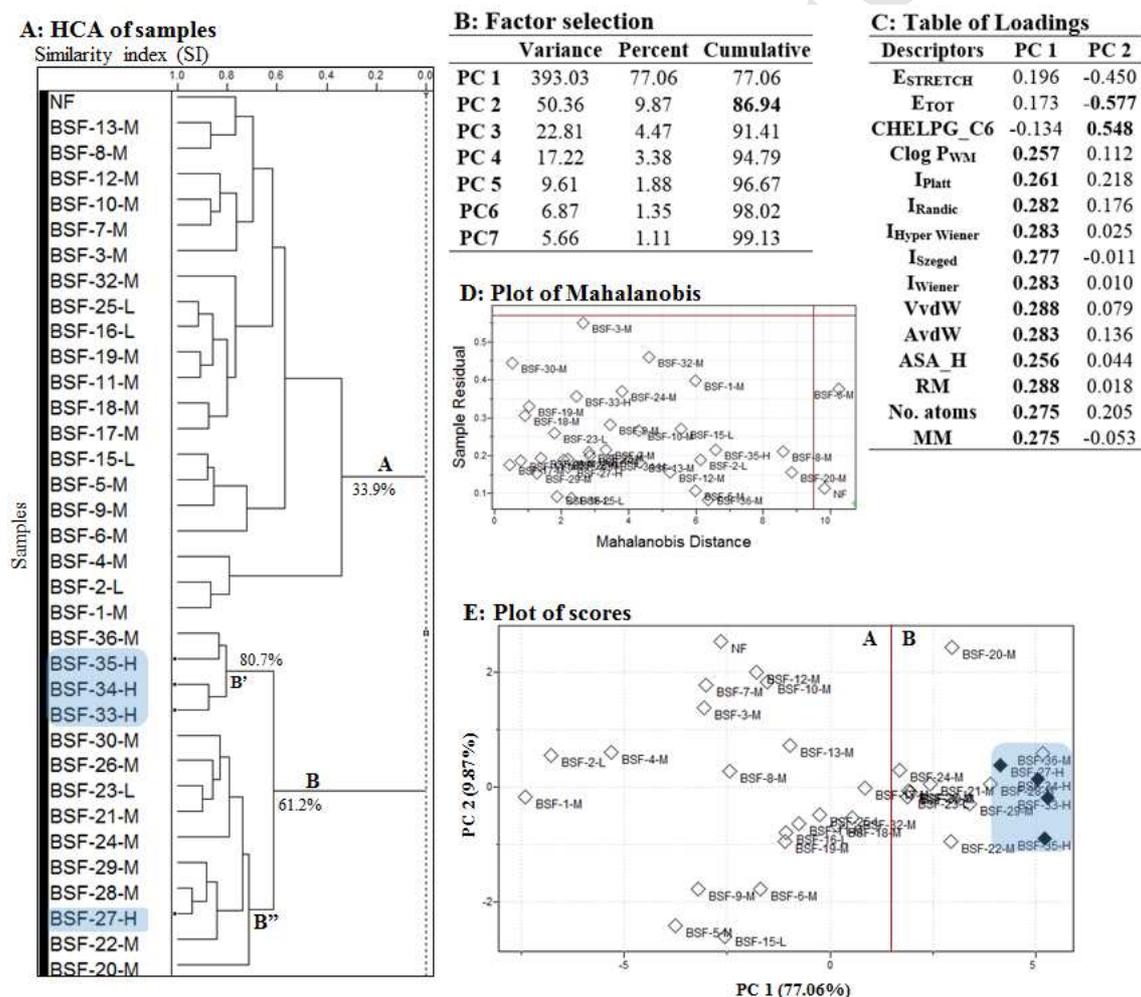


Figure 2. HCA findings: (A) Dendrogram of samples in which the compounds were grouped into two main clusters: A (SI = 0.339; less bulky, hydrophilic, aliphatic, heterocyclic, cycloalkyl substituents) and B (SI = 0.612; bulky and hydrophobic substituents). SI: similarity index; H: highly active compounds against the *T. cruzi*

epimastigote forms (in blue); M: moderately active compounds; L: low activity compounds. PCA findings: (B) Factors selection table: the two first factors, or PCs, explained 86.94 % of total variance from the original data; (C) Loadings' table for the two first factors, PC1 and PC2: high loadings values are in bold; (D) Outliers' diagnosis: samples residual *versus* Mahalanobis distance; (E) Plot of scores for the two first factors, PC1 *versus* PC2.

PCA has the purpose of correlating descriptors in order to find other variables, not only to reduce the descriptors' number, but also to group redundant descriptors in a new set of variables, orthogonal, called principal components (PC) or factors, which describe the information contained in the original data set [18, 19, 21]. In regard to our data set, the first two factors have described approximately 87 % of the total variance from the original data, being 77 % discriminated by PC1 (Figure 2B). In the table of loadings (Figure 2C) are highlighted the descriptors which most influenced the compounds' discrimination. In PC1, the loading values for stereochemical and geometric (ASA_H, A_{vdW} and V_{vdW}), topological (I_{Platt}, I_{Randic}, I_{Hyper Wiener}, I_{Szeged}, I_{Wiener}), hydrophobic (ClogP_{WM}), hydrophobic/stereochemistry (RM), and constitutional (n^o of atoms and MM) properties were higher and similar ranging from 0.25 to 0.29 (absolute values). For PC2, thermodynamic (E_{TOT}; -0.57) and electronic (CHELPG_C6; 0.55) properties showed the highest loading values.

Regarding the score plot (Figure 2E), PCA and HCA had complementary findings. Compounds with higher values in PC1, or Factor 1 (77 %), correspond to the same compounds grouped into cluster B (dendrogram of samples, HCA). Of note, bulkier and more hydrophobic substituents seem to improve the activity against *T. cruzi* epimastigotes.

The diagnosis of possible outliers was carried through plotting sample residual values *versus* Mahalanobis distance [23]. The confidence interval used was 95 %, which

weighted method) [24-27] and ASA_H (solvent accessible surface area of all hydrophobic atoms) properties can be visualized in Figure 4, comparing the more active compounds (BSF-27, BSF-33, BSF-34, BSF-35) with the NF lead compound and NFX drug (Figure 4).

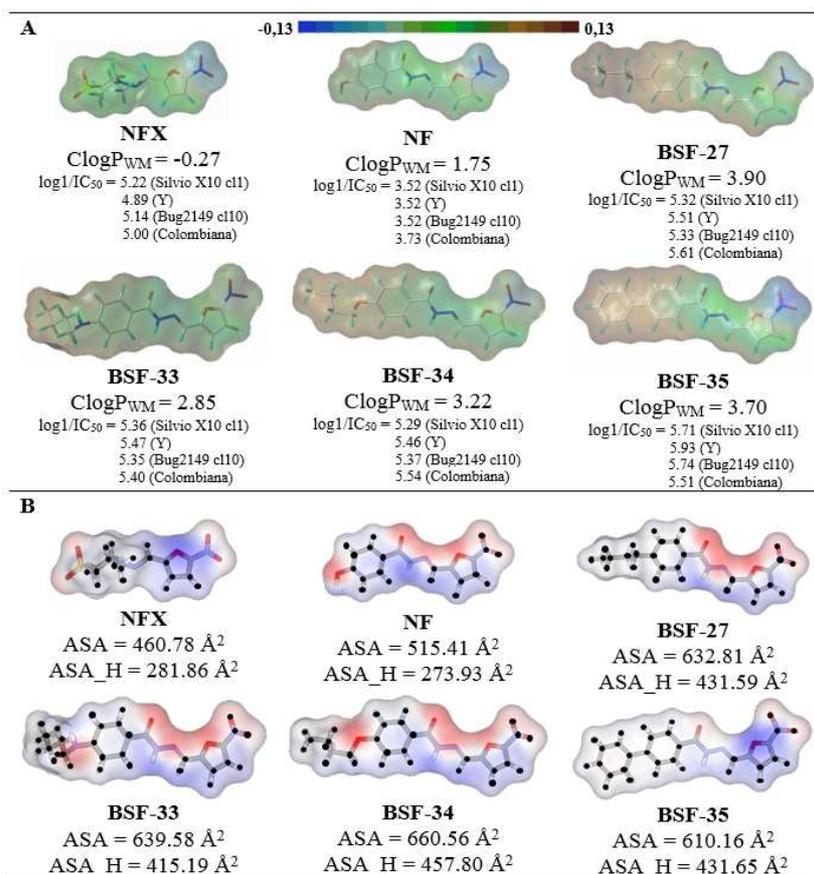


Figure 4. Representation of the calculated hydrophobic and steric/geometric descriptors for the most active compounds (BSF-27, BSF-33, BSF-34, BSF-35), NF (nifuroxazide) and NFX (nifurtimox): (A) map of lipophilic potential, MLP, calculated using the SYBYL 8.0 package. More hydrophobic regions are shown in brown (0.13) and more hydrophilic regions in blue (-0.13); the ClogP values were calculated using the Marvin program [27], considering equal weights to the methods of VISVANADHAN et al, 1989 Klopman et al, 1993 and PHYSPROP © database [24-27]. The compounds are presented in the stick model, where carbon atoms are in light gray, oxygen in red, nitrogen in blue, sulphur in yellow, and hydrogen in green. (B) ASA_H was calculated using ViewerLite 5.0 software [28] employing a probe of 1.4 Å radii (water); the

surfaces are displayed as transparent format. The black spots correspond to hydrophobic atoms ($|q_i| < 0.125$; $|q_i|$ is the absolute value of the partial charge of the atom). The compounds are showed in stick model (carbon atoms in gray, oxygen in red, nitrogen in blue, sulphur in yellow and hydrogen in white).

The lipophilicity property is widely investigated in SAR and QSAR approaches. The descriptors of such nature are commonly related to biological activity, since they may express the permeation of a molecule through biological tissues/membranes, as well as interactions with transporter proteins and enzymes [29]. The biological activity against *T. cruzi* has been related to hydrophobic descriptors in previous studies considering compounds of similar structures [13, 30, 31]. In this regard, comparing BSF-8 [-C₆H₄] and BSF-13 [-C₆H₄-4-CH₃] with the novel 4-alkyl substituted benzene compounds, we verified that the biological activity increases proportionally with the size of the alkyl chain. This suggests the presence of a lipophilic pocket in a putative receptor of the parasite that would better accommodate compounds with greater hydrophobicity.

Maps of lipophilic potential (MLP) were calculated onto a molecular Connolly surface and interpreted according to a color scheme, which ranged from brown (0.13, more lipophilic region) to blue (-0.13; more hydrophilic region). Regarding the most active compounds, 4-alkyl-substituted benzenes have contributed to increase compound lipophilicity (brown region on the molecular surface of BSF-27, BSF-33, BSF-34, BSF-35; see Figure 4A). The ASA_H descriptor indicated that the solvent accessible surface area of hydrophobic atoms is bigger for the most active compounds (BSF-27, BSF-33, BSF-34, BSF-35) than for NF and NFX.

2.4 Designing of new compounds, synthesis and biological activity against epimastigote forms

Based on the structure-activity/property findings, we designed four new compounds (See Supplementary material, Figure 3S: Structure elucidation of 5-nitro-2-furfuriliden derivatives - ^1H and ^{13}C NMR spectra of new compounds). Three compounds (BSF-38 to BSF-40) were designed aiming at increasing the activity against epimastigote forms, and the fourth compound (BSF-37) was designed to evaluate the contribution of benzene ring to the activity. The compounds were synthesized as shown in Scheme 1 (Figure 5A), and their biological activity was evaluated against epimastigote forms of the four *T. cruzi* strains (Figure 5B).

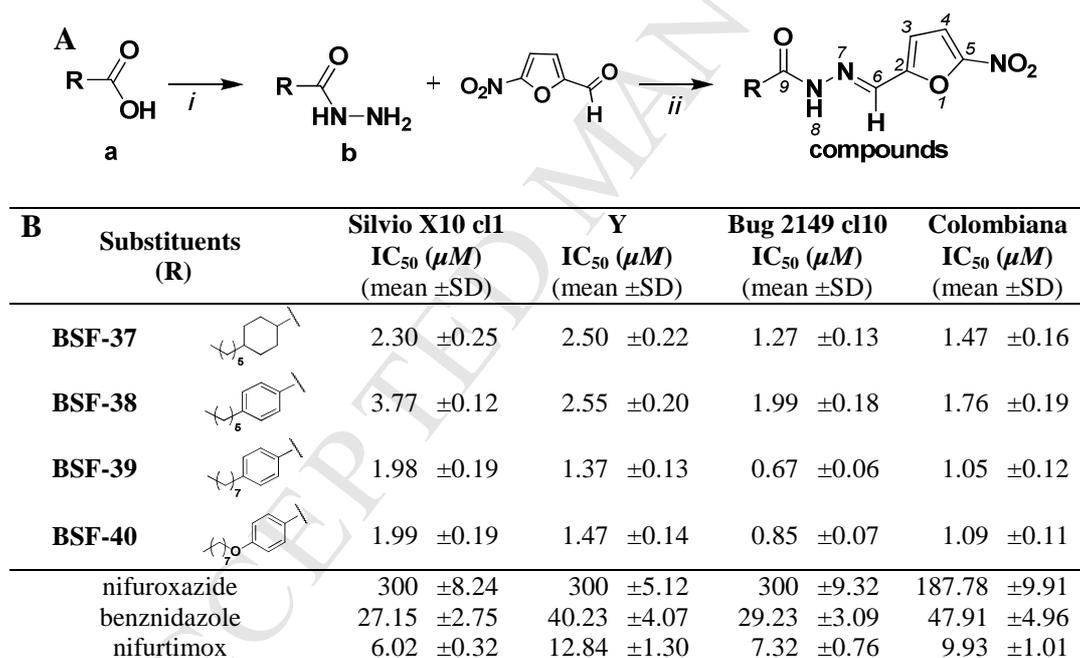


Figure 5. (A) Scheme 1: synthetic route of 5-nitro-2-furfuriliden derivatives. Reaction conditions: (i) CH_3OH , H_2SO_4 /reflux/4 h; N_2H_4 80% in H_2O /75 °C/30 min; (ii) 5-Nitro-2-furaldehyde, $\text{C}_2\text{H}_5\text{OH}$ /r.t./6 - 8 h. (B) Biological activity of new compounds against epimastigote forms of four *T. cruzi* strains. IC₅₀ value: compound concentration causing 50% reduction in the parasite growth in comparison to controls. IC₅₀ values, presented as mean and standard deviation (SD), correspond to triplicates from at least two independent experiments. Errors are in a range of 10%.

The data indicate that all new compounds show increased trypanocidal activity against epimastigote forms of the four strains as compared to the starting compounds (Table 1S of Supplementary Material: Biological activity of compounds against four strains of *T. cruzi*). Regardless the phenotypic differences among the strains [2], the substitution of benzene ring (BSF-38) by cyclohexyl (BSF-37) has no impact on the anti-epimastigote activity. Furthermore, the increase of the aliphatic chain of the 4-substituted 6-membered rings (BSF-37 to BSF-40) improves the activity against the epimastigote forms of the four strains.

2.5 Biological activity against trypomastigote forms of *T. cruzi*.

As mentioned above, the acute phase of CD is characterized by high parasitemia due to circulating trypomastigote forms. Thus, we evaluated the activity of the new compounds (BSF-37 to BSF-40) and the reference drug NFX against this developmental stage. Trypomastigotes of the Colombiana strain (10^6 /mL) were incubated in the absence or presence of the drugs. Percentages of inhibition of trypomastigote viability promoted by the compounds, in relation to the control incubated in the absence of drug were calculated (Table 2).

Table 2. Anti-*T. cruzi* activity of the compounds and reference drug against trypomastigote forms of the Colombiana strain.

	Trypomastigotes (10^5 /mL) ^a	Inhibition of viability (%) ^b
BSF-37	0.1	98.5
BSF-38	0.1	98.5
BSF-39	0.1	98.5
BSF-40	0.1	98.5
NFX	6.3	7.3
Control	6.8	---

Compounds were incubated at 1 μ M final concentration for 20h incubation at 37 °C; a: values represent the mean of two independent experiments; b: inhibition of viability in relation to the control.

After 20 h incubation in the absence of drug, but in the presence of 0.5% DMSO, we observed 32% inhibition of viable trypomastigote in the control, regarding the initial parasite concentration of 10^6 parasites/mL. This is due to the presence of DMSO and to the fact that trypomastigotes are non-dividing cells. NFX at $1 \mu\text{M}$ concentration inhibited only 7.3% of parasite viability, whereas the new compounds promoted a drastic inhibition of 98.5%.

2.6 Biological activity against intracellular amastigote forms of *T. cruzi*.

Drug analysis against the intracellular amastigote form has been considered as the most relevant assay to study compounds with anti-*T. cruzi* activity [32], since it more closely represents the *T. cruzi* tissue infection that leads to the development of CD in humans.

Exploratory data analysis (HCA method) was used to assist the selection of compounds to be screened against the amastigote forms. Thereby, compounds were selected to represent each cluster in the dendrogram (Figure 2A), including the novel four compounds designed posteriorly. Compounds containing substituents of different chemical nature were included in the sample to evaluate whether they have the same tendency against both epimastigote and amastigote forms of *T. cruzi*.

The analysis against amastigote forms was carried out using a high content screening (HCS). This approach is applied in studies of cell signalling, cell physiology, *in vitro* toxicological tests, enzyme targets, among others [32]. HCS evaluates the antiparasitic activity and cytotoxicity toward host cells based on the determination of parameters such as number of host cells, ratio of infected cells and number of amastigotes in infected cells.

The assays were carried out with the Y strain and the human osteosarcoma cell line U2OS as host cell [33]. The anti-amastigote activity (IC_{50}), the activity on U2OS host cells (cytotoxicity, CC_{50}) and the selectivity toward amastigotes (selectivity index, SI') were determined (Table 3).

It was observed that the most active compounds were very selective, with SI' values greater than 200, in contrast to the values of reference drugs: SI' for NFX =16.45 and BZD > 3. Furthermore, it was observed high correlation between results obtained for the same compounds against epimastigote (Table 1S of Supplementary Material: Biological activity of compounds against four strains of *T. cruzi*) and amastigote forms of the Y strain, reaching a correlation value of 0.786 (*Spearman's* non-parametric correlation method). Considering results obtained against amastigote forms, the same behavior was observed for BSF-37 and BSF-38 compounds ($P = 0.137$; statistical analysis – *Student's t*-test, unpaired) meaning that they do not have significant differences of activity on amastigote forms. The novel designed compounds having the benzene ring (BSF-38 to BSF-40) had similar anti-amastigote activity values ($P = 0.694$, one-way variance analysis, one way-ANOVA, with 95% of confidence interval).

Table 3. Anti-*T. cruzi* activity and host cell toxicity of the compounds and reference drugs against the amastigote form of the Y strain

Compounds	Y		U2OS CC_{50}		SI'	Max. Activity (%)
	IC_{50} (μ M)		(μ M)			
	(mean \pm SD)		(mean \pm SD)			
BSF-1	4.62	\pm 1.48	39.21	\pm 0.66	8.49	98.68
BSF-3	2.61	\pm 0.15	19.29	\pm 7.39	7.39	114.96
BSF-9	1.06	\pm 0.11	57.18	\pm 9.12	53.94	94.53
BSF-10	2.90	\pm 0.09	25.43	\pm 1,01	8.77	119.43
BSF-16	1.77	\pm 0.27	21.39	\pm 3.90	12.08	95.51
BSF-26	1.87	\pm 1.54	64.20	\pm 12.32	34.33	97.05
BSF-27	1.33	\pm 0.19	19.55	\pm 7.47	14.70	114.96
BSF-28	1.43	\pm 0.40	>200		>140	98.37

BSF-33	0.35	±0.20	12.04	±1.90	34.40	101.80
BSF-34	1.46	±0.39	46.62	±11.80	31.93	114.96
BSF-35	0.71	±0.11	109.07	±23.95	153.62	97.80
BSF-37	1.66	±0.62	>200		> 120	102.60
BSF-38	0.96	±0.10	>200		>208	104.24
BSF-39	0.82	±0.12	>200		>244	103.06
BSF-40	0.88	±0.30	>200		>227	98.26
NF	4.26	±1.02	45.69	±23.60	10.73	91.94
BZD	64.34	±39.52	>200		> 3	85.16
NFX	5.18	±1.15	85.19	± 21.08	16.45	96.60

Y: amastigote form of the Y strain. IC₅₀: concentration of compound which inhibits 50% of amastigotes compared to the control. CC₅₀: concentration of compound which inhibits 50% of U2OS cells, compared to control. U2OS: *Homo sapiens* osteosarcoma cells. SI': selectivity index [SI' = CC₅₀ U2OS / IC₅₀ of Y]. Max. Activity: Maximum Activity of the inhibition of amastigotes, compared to controls. Data presented as mean and standard deviation (SD) of triplicates and at least two independent assays. BZD: benznidazole; NFX: nifurtimox; NF: nifuroxazide (lead compound).

Among the most active compounds, compound BSF-39 [-C₆H₄-4-C₇H₁₅] showed the best selectivity to the amastigotes in relation to the host cell. Representative images obtained for infected cells treated with this compound at the concentration of 1.56 μM showed that the typical morphology of U2OS cell was maintained and no amastigote forms were present (Figure 6). Few intracellular parasites were still observed when BSF-39 was added at 0.39 μM concentration (Figure 6). The data obtained with the reference drug BZD at 1.56 μM indicated the presence of abundant intracellular parasites (Figure 6) (See also Figure 4S of Supplementary Material). The dose-response curves obtained for each compound selected in the assays against the amastigote forms are found in Figure 5S of Supplementary Material.

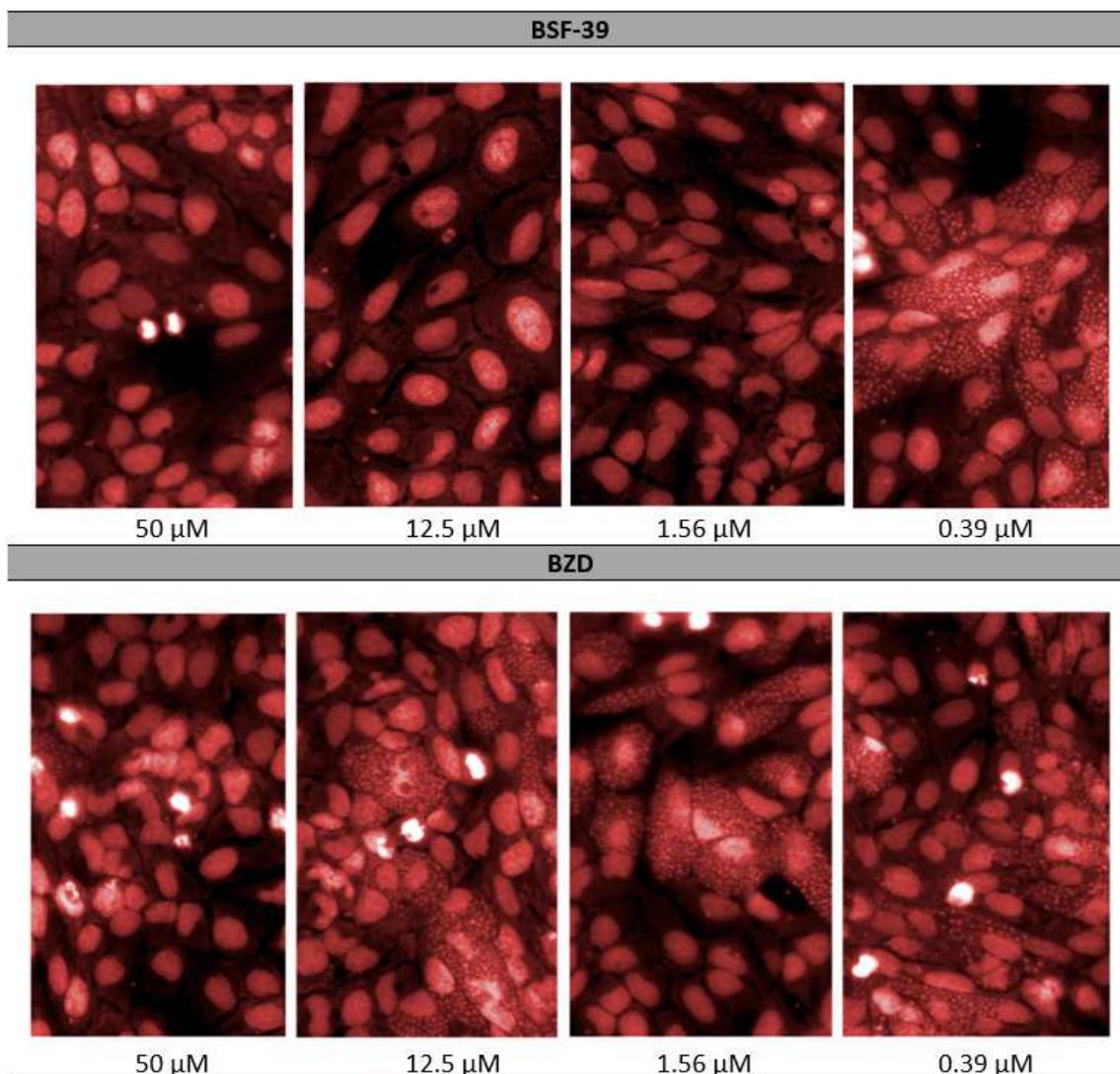


Figure 6. Images obtained by high content screening (HCS) for BSF-39 [-C6H4-4-C7H15]- and benznidazole-treated infected U2OS cells at 50, 12.5, 1.56 and 0.39 μM concentrations. Assays performed with the Y strain, in U2OS cells and incubation for 72 h. Host cell and parasite DNA are stained with Draq5 (Biofocus).

3 CONCLUSION

Based on the findings from exploratory data analysis, four new compounds were designed and synthesized. and their activity was evaluated against trypomastigotes and intracellular amastigotes. Our findings indicate that the novel compounds are active and selective for the three developmental forms of *T. cruzi*: epimastigotes, trypomastigotes and intracellular amastigotes. These results stimulate further studies on N'-[(5-

nitrofuranyl) methylene] substituted hydrazides as potential drug candidates to treat Chagas disease.

4 EXPERIMENTAL SECTIONS

4.1 Chemistry

NMR spectra were recorded on a Bruker ADPX Advanced (300 MHz) spectrometer employing DMSO- d_6 solutions with tetramethylsilane as internal standard. Melting points were determined using Büchi M-560 apparatus and elemental analysis was performed on a Perkin-Elmer 24013 CHN Elemental Analyzer. Synthesis and anti-*T. cruzi* active of compounds BSF-13, BSF-18 [12]; BSF-1, BSF-3, BSF-6, BSF-7, BSF-8, BSF-10, BSF-11, BSF-12, BSF-16, BSF-19, BSF-20, BSF-21, BSF-22, BSF-23, BSF-27, BSF-28, BSF-29, BSF-33, BSF-34, BSF-35, BSF-36 [13], was previously reported by our group.

4.1.1 General procedure for the synthesis of hydrazides (b): carboxylic acid (I) (0.02 mol) was refluxed for 4 hours in 20.0 mL (0.49 mol) of anhydrous methanol and 0.5 mL (0.01 mol) of sulphuric acid. The reaction mixture was cooled down to room temperature and the hydrazine hydrate 80% (v/v) (10.0 mL, 0.13 mol) was added. The system was maintained by vigorously stirring for more 30 minutes in reflux. After this period, the mixture was maintained at low temperature to give II and was purified from ethyl acetate. The hydrazide intermediate of compounds 1 and 2 were commercially obtained (Sigma-Aldrich, purity of 97%) [34, 35].

4.1.2 General procedure for the synthesis of nitrofurans compounds: were synthesized by refluxing 5-nitro-2-furaldehyde 98% (5 mmol) and hydrazides II (5

mmol) in water, sulphuric acid, acetic acid, and methanol (8:7:8:20 v/v) for 1 h. After cooling, the mixture was poured into cold water to precipitate and purified by recrystallization from acetonitrile [36]. Compounds with high water solubility were synthesized in 10 mL of ethanol PA (5 mmol) with 5-nitro-2-furaldehyde 98% (1 mmol) and hydrazides II (1 mmol) at room temperature and vigorous stirring for 6-8 h [37]. Compounds BSF-26, BSF-37, BSF-38, BSF-39, BSF-40 was identified as a new chemical entity and is described following.

4.1.2.1 4-penthyll-*N'*-[(5-nitrofuranyl)methylene] cyclohexane carbohydrazide

(BSF-37): light yellow solid (91%); mp 169 – 170 °C. ¹H NMR (DMSO-*d*₆, 300 MHz): δ (ppm): 11.67/11.50 (s, 1H, H8), 8.17/7.93 (s, 1H, H6), 7.76 (d, 1H, *J*_(4,3) = 3.8 Hz, H4), 7.18 (d, 1H, *J*_(3,4) = 3.2 Hz, H3), 3.07/2.18 (t, 1H, *J* = 12 Hz, H10), 1.79 (d, 4H, *J* = 10.5 Hz, H11, H15), 1.47-1.19 (m, 11H, H3, H18, H19 – Hax. e H16, H17 Hax./eq.), 0.95-0.84 (m, 5H, H18, H19 (Heq.), H20 (CH₃)); ¹³C NMR {H} (DMSO- *d*₆, 75 MHz): δ (ppm): 177.3/172.2 (C9), 152.0 (C2), 151.9 (C5), 133.8/130.5 (C6), 114.6 (C4), 114.0 (C3), 43.1/36.4 (C10), 36.7 (C16), 36.6 (C13), 32.0 (C12), 31.9 (C14), 31.6 (C18), 28.8 (C11), 28.3 (C15), 25.9 (C17), 22.1 (C19), 13.9 (C20); Anal. Calcd. for (C₁₇H₂₅N₃O₄): C, 60.88; H, 7.51; N, 12.53. Found: C, 60.58; H, 7.20; N, 12.16.

4.1.2.2 4-penthyll-*N'*-[(5-nitrofuranyl)methylene]benzohydrazide (BSF-38):

Yellow solid (η 94%); mp 200-201 °C. ¹H NMR (DMSO-*d*₆, 300 MHz): δ (ppm): 12.15 (s, 1H, H8), 8.41 (s, 1H, H6), 7.84 (d, 2H, *J* = 7.7 Hz, H11, H15), 7.79 (d, 1H, *J* = 3.0 Hz, H4), 7.36 (d, 2H, *J* = 7.9 Hz, H12, H14), 7.26 (d, 1H, *J* = 3.3 Hz, H3), 2.65 (t, 2H, *J* = 7.5 Hz, H16), 1.60 (t, 2H, *J* = 6.9 Hz, H17), 1.30 (d, 4H, H18, H19), 0.86 (t, 3H, *J* = 6.6 Hz, H20); ¹³C NMR {H} (DMSO- *d*₆, 75 MHz): δ (ppm): 163.3 (C9), 151.9 (C2, C5), 147.0 (C13), 135.1 (C6), 130.2 (C10), 128.4 (C11, C15), 127.8 (C12, C14), 115.0 (C4), 114.6

(C3), 34.9 (C16), 30.8 (C17), 30.3 (C18), 21.9 (C19), 13.8 (C20); Anal. Calcd. for (C₁₇H₁₉N₃O₄, 329.35 g): C, 62.0; H, 5.81; N, 12.76. Found: C, 61.94; H, 5.73; N, 12.45.

4.1.2.3 4-heptyl-*N'*-[(5-nitrofur-2-yl)methylene]benzohydrazide (BSF-39):

Yellow solid (η 92%); mp 178-179 °C. ¹H NMR (DMSO-*d*₆, 300 MHz): δ (ppm): 12.15 (s, 1H, H8), 8.42 (s, 1H, H6), 7.85 (d, 2H, *J*=7.6 Hz, H11, H15), 7.79 (d, 1H, *J*=3.0 Hz, H4), 7.35 (d, 2H, *J*=7.7 Hz, H12, H14), 7.26 (s, 1H, H3), 2.65 (t, 2H, *J*=7.2 Hz, H16), 1.59 (s, 2H, H17), 1.28 (s, 8H, H18-H21), 0.86 (s, 3H, H22); ¹³C NMR {H} (DMSO-*d*₆, 75 MHz): δ (ppm): 163.3 (C9), 151.9 (C2), 151.8 (C5), 147.0 (C13), 135.2 (C6), 130.2 (C10), 128.4 (C11, C15), 127.9 (C12), 127.8 (C14), 115.0 (C4), 114.6 (C3), 35.0 (C16), 31.2 (C20), 30.6 (C17), 28.5 (C18), 28.4 (C19), 22.0 (C21), 13.8 (C20); Anal. Calcd. for (C₁₉H₂₃N₃O₄, 357.470 g): C, 63.85; H, 6.49; N, 11.76. Found: C, 63.58; H, 6.39; N, 11.38.

4.1.2.4 4-heptyloxy-*N'*-[(5-nitrofur-2-yl)methylene]benzohydrazide (BSF-40):

Yellow solid (η 88%); mp 172-173 °C. ¹H NMR (DMSO-*d*₆, 300 MHz): δ (ppm): 12.09 (s, 1H, H8), 8.41 (s, 1H, H6), 7.91 (d, 2H, *J*=8.8 Hz, H11, H15), 7.79 (d, 1H, *J*=3.9 Hz, H4), 7.24 (d, 1H, *J*=3.9 Hz, H3), 7.06 (d, 2H, *J*=8.8 Hz, H12, H14), 4.05 (t, 2H, *J*=6.3 Hz, H17), 1.78-1.71 (m, 2H, H18), 1.42-1.28 (m, 8H, H19-H22), 0.87 (s, 3H, *J*=6.9 Hz, H23); ¹³C NMR {H} (DMSO-*d*₆, 75 MHz): δ (ppm): 161.8 (C9), 160.0 (C13), 152.0 (C2), 151.8 (C5), 134.7 (C6), 129.9 (C11), 129.8 (C15), 124.5 (C10), 114.8 (C4), 114.6 (C3), 114.2 (C12, C14), 67.8 (C17), 31.2 (C21), 28.5 (C18), 28.4 (C20), 25.4 (C19), 22.0 (C22), 13.9 (C23); Anal. Calcd. for (C₁₉H₂₃N₃O₅, 373.40 g): C, 61.11; H, 6.21; N, 11.25. Found: C, 60.99; H, 5.96; N, 10.87.

4.2 Biological activity against epimastigote forms of *T. cruzi* strains

In vitro anti-*T. cruzi* activity assays were performed against the epimastigote forms of Silvio X10 c11, Y, Bug 2149 c110 and Colombiana strains, as described elsewhere [13]. Exponentially growing epimastigotes (1.0×10^7 parasites/mL) were incubated in 96-well, flat-bottom tissue culture plates in 200 μ L LIT-FCS medium with different drug concentrations for 72 h at 28 °C. The nitro derivatives, BZ and NFX standard drugs, and the lead compound NF were dissolved in DMSO and diluted into LIT-FCS medium to obtain a concentration from 0.4 to 300.0 μ M. DMSO concentration in the medium did not exceed 1.0%. After the incubation period, the amount of viable parasites was determined from the absorbance measurements in a microplate reader (Biochrom – EX Read 400, England), at 562 nm wavelength. The assays were performed in triplicate and, at least, in two independent experiments. The anti-*T. cruzi* activity was calculated by using the formula: $PGI = \{1 - [(A_p - A_{pb}) / (A_c - A_{cb})]\} \times 100$, where: PGI = percentage of growth inhibition; $A_p = A_{562}$ of the culture at a given compound concentration after 72 h incubation; $A_{pb} = A_{562}$ of the medium at a given compound concentration without parasites (blank of compounds in each concentration); $A_c = A_{562}$ of the culture in the absence of any compound; ; $A_{cb} = A_{562}$ of the LIT-FCS medium (blank). The inhibitory concentration IC_{50} , corresponding to the drug concentration that inhibited 50% parasite growth, was calculated from the percentage of growth inhibition using nonlinear model growth/sigmoid dose-response for each case using, at least, six concentration values [38, 39].

4.3 Molecular modelling approach, calculation and selection of descriptors

The 3D molecular models of the compounds were built up using the Cartesian coordinates of crystallized structure of the lead compound NF (LEQTAC code, $R_{factor} = 0.11$)[40] retrieved from the Cambridge Structural Database (CSD) as reference

geometry [41]. Each molecular model was energy-minimized in the MM+ force field [42], without any constraints. The partial atomic charges were calculated employing the AM1 (Austin Model 1) semiempirical method [43]. The energy minimization of the 3D molecular models was also performed employing the MOLSIM 3.2 program[44] by applying the steepest descent and, subsequently, the conjugated gradient method, using a convergence criterion of 0.01 kcal/mol. The energy-minimized structures were the input to the MD simulations of 1ns (1,000,000 steps; step size of 1fs) at 301 K (28 °C), the same temperature of the biological assay. Dielectric constant of 3.5 was used to simulate the environment of the biological membranes. It was assigned a fictitious atomic mass of 5000 u.m.a. to some atoms position in order to maintain the structural integrity of the molecular models during simulation, based on the ^1H and ^{13}C NMR and NF crystallized structure. Trajectory files were recorded every 20 simulation steps to generate 50,000 conformations. The lowest-energy conformation was selected from the equilibrium region of conformational ensemble profile (CEP) and compared to the initial energy-minimized molecular model to verify whether the structural integrity was maintained after simulation. The RMSD value was used as criterion employing the Hyperchem 8.0 program.[41] The energy minimization procedure was made in MOLSIM 3.2 program[44], applying the steepest descent and conjugated gradient methods and the convergence criterion was 0.01 kcal/mol. At this time, some thermodynamic properties were obtained as follows: the total potential energy from minimization step (E_{total}); the intramolecular energy contribution of solvation (E_{solv}), applying the hydration shell model proposed by Forsythe and Hopfinger (1973) [45]; and, the intramolecular energy contribution of hydrogen bonding (E_{Hb}). The resulting 3D molecular model for each compound was the input structure to the calculation of physicochemical and structural properties, which were explored in this study.

The electronic properties as Charges from Electrostatic Potentials using a Grid based method (CHELPG) [46], dipole moment (total and x, y, z) and frontier molecular orbital energy (E_{HOMO} and E_{LUMO}) were calculated using the B3LYP 6.311++(d,p) method [47-49] implemented in the Gaussian 03W program [50].

The Marvin 5.2.1_1 program [27] was used to calculate steric, topological/geometric, and lipophilic properties. ClogP_{WM} was calculated by the weighted method, assigning equal weight for each method [24-26].

Further detailed information regarding all descriptors, methods, and respective software used to perform these calculations are listed in table 3S of Supplementary Material.

The independent variables or descriptors or molecular properties of different nature were calculated and generated a matrix containing 56 columns (independent variables or descriptors plus biological activity ($\log 1/C$)) and 36 rows corresponding to the number of compounds.

Procedure to select the most representative molecular properties: two filters were used as criteria to previously select the independent variables: (i) the Pearson correlation coefficient and (ii) a visual inspection of scatter plots for biological activity versus each variable or descriptor or molecular property [18]. The Pearson correlation coefficient value of 0.35 was formerly used as cut off for selecting the calculated independent variables. Regarding the visual inspection, only variables presenting uniform distribution and linear tendency with the biological data were selected to compose the final matrix.

The final matrix was used as input for the exploratory data analysis. Due to the distinct magnitude orders among the calculated variables, the autoscaling procedure was applied as a preprocessing method [21].

4.4 Exploratory data analysis[18, 19]

4.4.1 Hierarchical cluster analysis, HCA

HCA was performed using the Pirouette 3.11 program [22], employing the complete linkage method and Euclidean distance. In HCA, distances between pairs of samples (or variables) are calculated and compared. When distances between samples are relatively small, this implies that the samples are similar. The calculated distances between samples were set on a similarity matrix whose elements are called similarity indices, ranging between 0 and 1, where 1 is equivalent to a maximum similarity. The results are usually presented as a two-dimensional chart, called dendrogram [22].

The multivariate distance d_{ab} between two samples vectors, a and b, was determined by computing differences at each of the m variables:

$$d_{ab} = \left[\sum_i^m (x_{aj} - x_{bj})^M \right]^{1/M}$$

M is the order of the distance, and here corresponds to the Euclidean distance (M = 2).[22] Because of inter-sample distances can vary with the type and number of measurements, it is customary to transform them onto a somewhat more standard scale of similarity, where d_{max} is the largest distance in the data set:

$$\text{similarity}_{ab} = 1 - \frac{d_{ab}}{d_{max}}$$

HCA was carried out for samples and variables. In the first case, biological activity was considered as dependent variable. In the second option, the biological activity is considered as independent variable [18, 19, 22].

4.4.2 Principal Components Analysis, PCA

The exploratory analysis of PCA was also carried out employing the Pirouette 3.11 software [22]. For PCA, the biological activity was considered as the dependent variable. The data were decomposed into two matrices, one of scores related to the samples, and another of loadings, related to the variables [18, 19].

The new set of axes generates the PCs, or factors, into which are the information related to the original descriptors. Thus, the number of PCs that explain most of the variability in the data set can be determined, considering that these PCs are uncorrelated and mutually orthogonal variables built up as simple linear combinations from the original data. In this exploratory data analysis, PCA was run up to ten factors or PCs. The outliers' diagnosis, implemented in Pirouette 3.11 software [22], was also performed through the Mahalanobis distance [51].

4.5 Biological activity against trypomastigote forms of *T. cruzi*

Trypomastigote forms of the Colombiana strain were obtained from the supernatant of infected LLC-MK₂ cells. After centrifugation at 7,000 rpm at 4 °C, the parasites (10⁶ cells/mL) were resuspended in DME medium supplemented with 2% FCS and added to 96-well, flat-bottom tissue culture plates in the absence or presence of 1 μM of different compounds for 20 h at 37 °C in a 5% CO₂ atmosphere. DMSO concentration in the medium was 0.5%. After incubation, the viability of the parasites was examined by motility under a light microscope. The number of live parasites was quantified in a Neubauer chamber. The percentage of inhibition of trypomastigote viability promoted by the compounds was calculated in relation to the control incubated in the absence of drug. The assay was performed in two independent experiments.

4.6 Biological activity against amastigote forms of *T. cruzi*

Test and reference compounds were diluted in the proper volume of 100% DMSO, to prepare 10 mM stock solutions. Aliquots of all solutions were kept frozen at $-80\text{ }^{\circ}\text{C}$, protected from light, and submitted for a maximum of 3 cycles of freezing-thawing.

The HCS assay was performed against intracellular amastigote as described in Moraes *et al.*, 2014 [15]. Briefly: on day 1, U2OS cells were seeded in black μClear 384 well tissue culture treated polystyrene plates (Greiner Bio-One) at 700 cells in 40 μL of high glucose DME medium (Hyclone) supplemented with 10% of FBS and a penicillin/streptomycin solution (Gibco Thermo Fisher Scientific). After 24 h of microplate incubation at $37\text{ }^{\circ}\text{C}$, 5% CO_2 and controlled humidity, trypomastigotes of Y strain were harvested from infected LLC-MK₂ cells and added to the U2OS-containing 384 microplate at 2,800 trypomastigotes in 10 μL of low glucose DME medium (supplemented as above) per well.

On day 3, 10 μL of compound solution (14 points in a 2-fold dilution scheme) were transferred onto the assay plates containing *T. cruzi*-infected U2OS cells, reaching the highest concentration of 200 μM and 1% DMSO for each compound, and the final assay volume of 60 μL . Each compound concentration was tested in doublets, and each experiment was performed in duplicate. At the assay endpoint, microplates were fixed with 4% PFA and stained with 5 μM Draq5 (Biofocus). Images were acquired using the Operetta High Content System (Perkin Elmer) with the objective lens of 20x WD magnification. The high content analysis was performed using the Harmony software (Perkin Elmer). Several parameters, such as host cells number, infection ratio and number of parasites per infected cell, were determined. For the purpose of this study, the infection ratio (IR) was defined as the ratio between the total number of infected cells from the well and the total number of cells from the same well. The raw data for

IR values were normalized to negative (infected cells, DMSO-mock treated) and positive (not infected cells) controls to determine the normalized antiparasitic activity. Activity values were processed with the Graphpad Prism software, version 7, for generation of sigmoidal dose-response (variable slope) nonlinear curve fitting and determination of EC_{50} and CC_{50} values by interpolation.

For the purpose of this study, IC_{50} was defined as the compound concentration corresponding to 50% normalized activity (reduction of infection) after 72 h of compound incubation. Potency relates to the IC_{50} values – the more potent the compound, the lower is its IC_{50} – whereas efficacy relates to the maximum observed activity of a compound – the more efficacious the compound, the closer its maximum activity is from 100%. Cytotoxicity was measured in terms of CC_{50} value, defined as the compound concentration corresponding to reduction of 50% in the number of cells in comparison with negative controls (DMSO-treated). Both for IC_{50} and CC_{50} values were determined by interpolation to the dose-response curves. The selectivity index was defined as the ratio between the CC_{50} and the IC_{50} values obtained for a given compound.

Supplementary material available

Dose-response curve of anti-*T. cruzi* activity – epimastigote forms, Colombiana strain; Biological activity of compounds against four strains of *T. cruzi*; Conformational ensemble profile (CEP) from molecular dynamics (MD) simulation of each compound; Values of the thermodynamic parameters obtained for the lowest-energy conformer of each compound from MD simulation; Descriptors, methods, and software used to perform calculations for 5-nitro-2-furfuriliden derivatives; Independent variables obtained from literature and descriptors calculated for each compound; Pearson's

correlation coefficient and scatter plots of each descriptor *versus* compounds' biological activities; Structure elucidation of 5-nitro-2-furfuriliden derivatives (^1H and ^{13}C NMR spectra of new compounds); Images obtained in the assay against the amastigote form of $Y_{(\text{ama-R})}$ strain; Dose-response curves obtained in the assays against the amastigote form through the high content screening (HCS).

ACKNOWLEDGEMENTS

The authors thank Professor Maria Júlia Manso Alves of the Laboratory of Biochemical Parasitology, Department of Biochemistry, Institute of Chemistry, University of São Paulo (USP) for providing the epimastigote forms of *T. cruzi* Y strain and LIT medium. This work was supported by grants of Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (grants number 2013/13333-8 and 2014/06061-4) and Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (grants number 304793/2009-4 and 472739/2013-1). CHF, MB and SLN are recipient of fellowships from CNPq (grants number 870448/1997-8, 163670/2015-3, and 501189/2013-0, respectively).

REFERENCES

- [1] J.R. Coura, P.A. Viñas, Chagas disease: a new worldwide challenge, *Nature*, 465 (2010) S6-S7.
- [2] B. Zingales, S.G. Andrade, M.R.S. Briones, D.A. Campbell, E. Chiari, O. Fernandes, F. Guhl, E. Lages-Silva, A.M. Macedo, C.R. Machado, A new consensus for *Trypanosoma cruzi* intraspecific nomenclature: second revision meeting recommends TcI to TcVI, *Memórias do Instituto Oswaldo Cruz*, 104 (2009) 1051-1054.
- [3] B. Zingales, M.A. Miles, C.B. Moraes, A. Luquetti, F. Guhl, A.G. Schijman, I. Ribeiro, Drug discovery for Chagas disease should consider *Trypanosoma cruzi* strain diversity, *Memórias do Instituto Oswaldo Cruz*, 109 (2014) 828-833.
- [4] A. Rassi, J.M. de Rezende, American trypanosomiasis (Chagas disease), *Infectious disease clinics of North America*, 26 (2012) 275-291.

- [5] J.A. Urbina, Specific chemotherapy of Chagas disease: relevance, current limitations and new approaches, *Acta tropica*, 115 (2010) 55-68.
- [6] M.D. Lewis, J.M. Kelly, Putting infection dynamics at the heart of Chagas disease, *Trends in parasitology*, 32 (2016) 899-911.
- [7] L.S. Filardi, Z. Brener, Susceptibility and natural resistance of *Trypanosoma cruzi* strains to drugs used clinically in Chagas disease, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 81 (1987) 755-759.
- [8] S.G. Andrade, A. Rassi, J.B. Magalhaes, F. Ferriolli Filho, A.O. Luquetti, Specific chemotherapy of Chagas disease: a comparison between the response in patients and experimental animals inoculated with the same strains, *Transactions of the Royal Society of Tropical Medicine and Hygiene*, 86 (1992) 624-626.
- [9] B. Zingales, *Trypanosoma cruzi* genetic diversity: Something new for something known about Chagas disease manifestations, serodiagnosis and drug sensitivity, *Acta tropica*, xx (2017) xxxx-xxxx. <http://dx.doi.org/10.1016/j.actatropica.2017.09.017>
- [10] S. Croft, In vitro screens in the experimental chemotherapy of leishmaniasis and trypanosomiasis, *Parasitology Today*, 2 (1986) 64-69.
- [11] M. Moreno, D.A. D'ávila, M.N. Silva, L. Galvão, A.M. Macedo, E. Chiari, E.D. Gontijo, B. Zingales, *Trypanosoma cruzi* benzimidazole susceptibility in vitro does not predict the therapeutic outcome of human Chagas disease, *Memórias do Instituto Oswaldo Cruz*, 105 (2010) 918-924.
- [12] F. Palace-Berl, S.D. Jorge, K.F.M. Pasqualoto, A.K. Ferreira, D.A. Maria, R.R. Zorzi, L. de Sá Bortolozzo, J.Â.L. Lindoso, L.C. Tavares, 5-Nitro-2-furfuriliden derivatives as potential anti-*Trypanosoma cruzi* agents: Design, synthesis, bioactivity evaluation, cytotoxicity and exploratory data analysis, *Bioorg. Med. Chem.*, 21 (2013) 5395-5406.
- [13] F. Palace-Berl, K.F.M. Pasqualoto, S.D. Jorge, B. Zingales, R.R. Zorzi, M.N. Silva, A.K. Ferreira, R.A. de Azevedo, S.F. Teixeira, L.C. Tavares, Designing and exploring active N'-[(5-nitrofuranyl) methylene] substituted hydrazides against three *Trypanosoma cruzi* strains more prevalent in Chagas disease patients, *European journal of medicinal chemistry*, 96 (2015) 330-339.
- [14] P. Hernández, R. Rojas, R.H. Gilman, M. Sauvain, L.M. Lima, E.J. Barreiro, M. González, H. Cerecetto, Hybrid furoxanyl N-acylhydrazone derivatives as hits for the development of neglected diseases drug candidates, *European journal of medicinal chemistry*, 59 (2013) 64-74.
- [15] C.B. Moraes, M.A. Giardini, H. Kim, C.H. Franco, A.M. Araujo-Junior, S. Schenkman, E. Chatelain, L.H. Freitas-Junior, Nitroheterocyclic compounds are more efficacious than CYP51 inhibitors against *Trypanosoma cruzi*: implications for Chagas disease drug discovery and development, *Scientific reports*, 4 (2014). DOI: 10.1038/srep04703
- [16] S.R. Wilkinson, M.C. Taylor, D. Horn, J.M. Kelly, I. Cheeseman, A mechanism for cross-resistance to nifurtimox and benzimidazole in trypanosomes, *Proceedings of the National Academy of Sciences*, 105 (2008) 5022-5027.
- [17] B.S. Hall, C. Bot, S.R. Wilkinson, Nifurtimox activation by trypanosomal type I nitroreductases generates cytotoxic nitrile metabolites, *Journal of Biological Chemistry*, 286 (2011) 13088-13095.
- [18] M. Ferreira, Multivariate QSAR, *J. Braz. Chem. Soc.*, 13 (2002) 742-753.
- [19] K.R. Beebe, R.J. Pell, M.B. Seasholtz, *Chemometrics: a practical guide*, Wiley-Interscience, New York, 1998.

- [20] J.C. Baber, D.C. Thompson, J.B. Cross, C. Humblet, GARD: a Generally Applicable Replacement for RMSD, *Journal of chemical information and modeling*, 49 (2009) 1889-1900.
- [21] M. Ferreira, A.M. Antunes, M.S. Melgo, P.L.O. Volpe, *Quimiometria I: calibração multivariada, um tutorial*, *Química Nova*, 22 (1999) 724-731.
- [22] PIROUETTE 3.11, Infometrix Inc.: Woodinville, WA, 1990-2003.
- [23] P.C. Mahalanobis, On the generalized distance in statistics, *Proceedings of the National Institute of Sciences of India*, New Delhi, 1936, pp. 49-55.
- [24] V.N. Viswanadhan, A.K. Ghose, G.R. Revankar, R.K. Robins, Atomic physicochemical parameters for three dimensional structure directed quantitative structure-activity relationships. 4. Additional parameters for hydrophobic and dispersive interactions and their application for an automated superposition of certain naturally occurring nucleoside antibiotics, *Journal of chemical information and modeling*, 29 (1989) 163-172.
- [25] G. Klopman, J.-Y. Li, S. Wang, M. Dimayuga, Computer automated log P calculations based on an extended group contribution approach, *Journal of chemical information and modeling*, 34 (1994) 752-781.
- [26] PHYSPROP database. <http://www.syrres.com/what-we-do/databaseforms.aspx?id=386>.
- [27] Marvin Beans, version 5.2.1_1; ChemAxon Ltd.: 1998-2010, Budapest, Hungary.
- [28] VIEWERLITE 5.0. Accelrys Inc.: 2000.
- [29] C.G. Wermuth, *The practice of medicinal chemistry*, Academic Press, Oxford, 2008.
- [30] S.D. Jorge, F. Palace-Berl, K.F. Mesquita Pasqualoto, M. Ishii, A.K. Ferreira, C.M. Berra, R.V. Bosch, D.A. Maria, L.C. Tavares, Ligand-based design, synthesis, and experimental evaluation of novel benzofuroxan derivatives as anti-*Trypanosoma cruzi* agents, *European Journal Medicinal Chemistry*, 64 (2013) 200-214.
- [31] F. Palace-Berl, S.D. Jorge, K.F.M. Pasqualoto, A.K. Ferreira, D.A. Maria, R.R. Zorzi, L. de Sá Bortolozzo, J.Â.L. Lindoso, L.C. Tavares, 5-Nitro-2-furfuriliden derivatives as potential anti-*Trypanosoma cruzi* agents: Design, synthesis, bioactivity evaluation, cytotoxicity and exploratory data analysis, *Bioorganic & medicinal chemistry*, 21 (2013) 5395-5406.
- [32] J.M. Zock, Applications of high content screening in life science research, *Combinatorial chemistry & high throughput screening*, 12 (2009) 870-876.
- [33] J. Ponten, E. Saksela, Two established in vitro cell lines from human mesenchymal tumours, *International journal of cancer*, 2 (1967) 434-447.
- [34] S.D. Jorge, F. Palace-Berl, A. Masunari, C.A. Cechinel, M. Ishii, K.F.M. Pasqualoto, L.C. Tavares, Novel benzofuroxan derivatives against multidrug-resistant *Staphylococcus aureus* strains: Design using Topliss' decision tree, synthesis and biological assay, *Bioorganic & medicinal chemistry*, 19 (2011) 5031-5038.
- [35] C.-K. Li, Y.-J. Ma, L.-H. Cao, Synthesis of Novel 3-Acetyl-2-aryl-5-(3-aryl-1-phenyl-pyrazol-4-yl)-2, 3-dihydro-1, 3, 4-oxadiazoles, *Journal of the Chinese Chemical Society*, 56 (2009) 182-185.
- [36] L.C. Tavares, T.C. Penna, A.T. Amaral, Synthesis and biological activity of nifuroxazide and analogs, *Bollettino chimico farmaceutico*, 136 (1997) 244-249.
- [37] C.F. Da Costa, A.C. Pinheiro, M.V. De Almeida, M. Lourenço, M.V. De Souza, Synthesis and antitubercular activity of novel amino acid derivatives, *Chemical biology & drug design*, 79 (2012) 216-222.
- [38] M. Ishii, S.D. Jorge, A.A. Oliveira, F. Palace-Berl, I.Y. Sonehara, K.F.M. Pasqualoto, L.C. Tavares, Synthesis, molecular modeling and preliminary biological

evaluation of a set of 3-acetyl-2, 5-disubstituted-2, 3-dihydro-1, 3, 4-oxadiazole as potential antibacterial, anti-*Trypanosoma cruzi* and antifungal agents, *Bioorganic & Medicinal Chemistry*, 19 (2011) 6292-6301.

[39] H. Cerecetto, R. Di Maio, M. González, M. Risso, P. Saenz, G. Seoane, A. Denicola, G. Peluffo, C. Quijano, C. Olea-Azar, 1, 2, 5-Oxadiazole N-oxide derivatives and related compounds as potential antitrypanosomal drugs: structure-activity relationships, *Journal of Medicinal Chemistry*, 42 (1999) 1941-1950.

[40] B. Pniewska, M. Januchowski, Structural investigation of nifuroxazide, p-hydroxy-N',-(5-nitrofurfurylidene) benzhydrazide, *Polish Journal of Chemistry*, 72 (1998) 2629-2634.

[41] HYPERCHEM Program *for Windows*; Hypercube, Inc.: Gainesville, FL, 2008.

[42] N.L. Allinger, Conformational analysis. 130. MM2. A hydrocarbon force field utilizing V1 and V2 torsional terms, *Journal of the American Chemical Society*, 99 (1977) 8127-8134.

[43] M.J. Dewar, E.G. Zoebisch, E.F. Healy, J.J. Stewart, Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model, *Journal of the American Chemical Society*, 107 (1985) 3902-3909.

[44] D. Doherty, MOLSIM: Molecular Mechanics and Dynamics Simulation Software-User's Guide, version 3.2. The Chem21 Group Inc.: Lake Forest.1997.

[45] K. Forsythe, A. Hopfinger, The Influence of Solvent on the Secondary Structures of Poly (L-alanine) and Poly (L-proline), *Macromolecules*, 6 (1973) 423-437.

[46] C.M. Breneman, K.B. Wiberg, Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis, *Journal of Computational Chemistry*, 11 (2004) 361-373.

[47] A.D. Becke, Density-functional exchange-energy approximation with correct asymptotic behavior, *Physical review A*, 38 (1988) 3098.

[48] C. Lee, W. Yang, R.G. Parr, Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density, *Physical review B*, 37 (1988) 785.

[49] G. Petersson, M.A. Al-Laham, A complete basis set model chemistry. II. Open-shell systems and the total energies of the first-row atoms, *The Journal of chemical physics*, 94 (1991) 6081-6090.

[50] GAUSSIAN 03W-revision B.02 *for Windows*, version 6; Gaussian Inc.: Pittsburgh, PA, 1995-2003.

[51] P.C. Mahalanobis, On tests and measures of group divergence, *Journal of the Asiatic Society of Bengal*, 26 (1930) 541-588.

Research Highlights (EJMC2017):

> A set of 36 *N'*-(5-nitrofur-2-yl) methylene] substituted hydrazides, previously reported, were active against the benznidazole-resistant Colombiana strain. > Multivariate approaches (hierarchical cluster analysis and principal component analysis) were applied to established structure-activity relationship > Four new compounds were designed and synthesized >. They were highly active against epimastigotes, trypomastigotes and intracellular amastigotes and showed low cytotoxicity to the host cells.