Drug Cocktail Selection for the Treatment of Chagas Disease: a Multi-objective Approach

Mateo Torres*, Juan J. Cáceres*, Rubén Jiménez†, Víctor Yubero†,
Celeste Vega†‡, Miriam Rolón‡, Luca Cernuzzi†, Benjamín Barán† and Alberto Paccanaro*

*Centre for Systems and Synthethic Biology, Department of Computer Science, — Royal Holloway, University of London
Egham, Surrey, United Kingdom
†Facultad de Ciencias y Tecnología — Universidad Católica "Nuestra Señora de la Asunción"
Asunción, Paraguay
‡Centro para el Desarrollo de la Investigación Científica
Asunción, Paraguay

***Abstract*—Chagas disease is a parasitic disease, endemic in South America. As of today, there is no effective treatment in its chronic stage. We have recently identified 134 FDA approved drugs with potential antitrypanosomal activity. In this paper, we propose a novel method for selecting combinations of drugs (drug cocktails), to provide a more effective treatment against Chagas disease. We define three measures to evaluate the predicted performance of a cocktail, establishing in this way a mathematical foundation for its analysis. This allows us to model the drug cocktail selection as a multi-objective optimisation problem, that we show can be solved efficiently with state-of-the-art evolutionary algorithms. Our analysis retrieves 57 drug cocktails containing between 2 and 6 drugs. We discuss the improvement of the cocktail selection given by our method, and the application of this approach to the identification of cocktails against other parasitic diseases.**

***Keywords*—Multi-objective optimisation, Drug repurposing, Chagas Disease, Genetic Algorithms.**

1. Introduction and the Problem

Chagas Disease, caused by the protozoan parasite *Trypanosoma cruzi* (*T. cruzi*), is endemic throughout Latin America and has spread to other countries, making it a worldwide issue. About 6 to 7 million people are infected with *T. cruzi* [1], and around 40 million are at risk of infection [2]. Insect vectors known as triatomine bugs are the primary means of human infection. After biting, they leave *T. cruzi* parasites (trypomastigotes) into excretion, usually introduced into the bloodstream through the bite wound or mucous membranes [3].

The disease goes through two different phases: the acute, and the chronic phase. In the acute phase, symptoms are often absent or mild; they may include fever, headache, and enlarged lymph glands. Less than 50% cases develop characteristic symptoms: a skin lesion or a purplish swelling of the lids of one eye [1]. In the chronic phase, parasites lodge mainly in the heart and digestive muscles. This can lead to severe organ pathologies and ultimately death [3].

Two drugs are currently available for the acute phase: Nifurtimox and Benznidazole. This phase, however, often goes undiagnosed due to a lack of proper diagnostic methods, and the inherent absence of symptoms [4]. Clinically, the disease is most commonly encountered in the chronic phase [5] and in this phase treatment is highly limited due to the low potency of the abovementioned drugs against the parasites [6] and the lack of other effective drugs [4].

Our aim is to find drug cocktails with antitrypanosomal effects. Drug cocktails have been found to be effective against a number of diseases, including infective diseases and even for the treatment of diseases caused by parasites, such as malaria [7] (caused by Plasmodium falciparum). However, the prediction of cocktails of drugs effective against Chagas disease is hindered by our limited biological knowledge of *T. cruzi*.

Comparative genomics attempts to harness the biological knowledge of well studied organisms in order to make inferences on less studied organisms; it infers molecular function and behaviour by comparing genome sequences from evolutionary related species [8]. In our case, although very little is known about the biology of *T. cruzi*, there exists abundant knowledge about other (evolutionarily related) model organisms and, in some cases, we even have knowledge about drugs that are effective against these organisms.

We have recently developed a method that exploits concepts from comparative genomics for the prediction of FDA approved drugs which could be effective against *T. cruzi* [9]. Briefly, we begin by predicting which metabolic

978-1-5386-3057-0/17/$31.00

pathways[1] from model organism are present in *T. cruzi*. Then, our method selects FDA approved drugs which target enzymes in model organisms that are evolutionarily related to enzymes in *T. cruzi* pathways, and could therefore be effective at disrupting them.

This method has produced an initial set of 134 FDA approved drugs. At this point, we need a way to define effective drug cocktails, that is, an optimal subset of these drugs which will be effective at disrupting *T. cruzi* pathways. Note that a brute force approach that would experimentally test every possible subset of n drugs would be infeasible even for small values of n, and therefore algorithmic methods for selecting effective drug combinations are needed.

This paper introduces an innovative mathematical approach to quantify potential antitrypanosomal activity of drugs in terms of three biologically motivated measures. This allows us to formulate the drug cocktail selection problem as a multi-objective optimisation problem for which efficient solution methods are available.

In the following sections, we provide a formal definition of the measures, and illustrate how to select cocktails by combining these measures using a multi-objective optimisation approach. The optimisation problem is solved using the Non-dominated Sorting Genetic Algorithm II (NSGA-II) [11], a state-of-the-art genetic algorithm to solve multi-objective problems [12]. Finally, we present the optimisation results, and discuss the predicted drug cocktails.

2. Multi-objective Approach

The comparative genomics approach summarized in the previous section [9] provided us with 146 *T. cruzi* metabolic pathways and 134 drugs which have potential interactions with 121 enzymes from 94 *T. cruzi* pathways. We define the following measures:

1) *Number of enzymes that are being targeted by the cocktail.*
   The notion behind this measure is that if the target protein and the enzyme are similar at molecular level, then it is highly likely that they share similar structure. This will allow the drug to bind to the enzyme, preventing its function. If this is the case, we might be blocking many enzymes and increasing the likelihood of affecting the parasite.
2) *Number of pathways with at least one enzyme targeted by the cocktail.*
   Although we cannot say for sure that every pathway considered in this measure will be disrupted, it is reasonable to argue that when the drug cocktail targets more than one pathway, the chance to disrupt essential pathways is higher. Cocktails that disrupt essential pathways should be more effective against the organism.
3) *Number of pathways which have all their enzymes "covered" by the cocktail.*
   Intuitively, pathways that are fully covered by a drug cocktail are more likely to be disrupted.

The selection of a putative drug cocktail is not a trivial problem for two main reasons. First, a measure of antitrypanosomal activity of a drug cocktail must be established in order to optimise the composition of the cocktail. Second, the amount of possible drug cocktails makes the pairwise comparison an infeasible computational problem.

The relative importance of the 3 measures described above is not evident. Nevertheless, it is desirable to take all of them into account, as they model different aspects that can lead to antitrypanosomal activity. To optimise all measures simultaneously, we decided to model the cocktail selection problem as a multi-objective optimisation problem.

Problems modelled to optimise multi-dimensional objective functions are called Multi-Objective Problems (MOPs) [12]. Intuitively, it is simpler to model a particular aspect without taking into account other (possibly conflicting) factors that would make the function very complex. Without loss of generality, the formal definition of a MOP is presented as a minimisation [12]:

Definition 1. *Multi-Objective Problem:* Let F be a set of M objective functions $\{f_1, f_2, \cdots, f_M\}, f_i : \mathbb{R}^n \Rightarrow \mathbb{R}$, a Multi-Objective Problem is defined as:

$$\begin{aligned} \text{Minimise } y = F(x) &= (f_1(x), f_2(x), \cdots, f_M(x)) \\ x &= (x_1, x_2, \cdots, x_n) \in \mathcal{X} \subseteq \mathbb{R}^n \\ y &= (y_1, y_2, \cdots, y_M) \in \mathcal{Y} \subseteq \mathbb{R}^M \end{aligned} \tag{1}$$

subject to

$$x_i^{(L)} \leq x_i \leq x_i^{(U)} \ \forall \ i \in \{1, 2, \cdots, n\} \tag{2}$$

$$r(x) = (r_1(x), r_2(x), \cdots, r_l(x)) \leq 0 \tag{3}$$

where x is a vector with n decision variables, while y is a M-dimensional objective vector.

Constraints (2) represent lower ($x_i^{(L)}$) and upper ($x_i^{(U)}$) variable bounds that define the decision space $\mathcal{X}$. The objective functions make a multidimensional space called "objective space", termed $\mathcal{Y}$. The r vector is made of l constraint functions that shape the feasible region of $\mathcal{X}$. Solutions that do not satisfy the constraint functions or variable bounds are called "infeasible solutions", and those that satisfy every constraint in (2) and (3) are feasible solutions. The set of all feasible solutions $\mathcal{X}_f$ is known as the feasible region. The domain of every f_i is $\mathcal{X}_f$. For every solution $x \in \mathcal{X}_f$ there exists a point y in the objective space. This defines the feasible objective space $\mathcal{Y}_f$

$$\mathcal{Y}_f = F(\mathcal{X}_f) = \bigcup_{x \in \mathcal{X}_f} \{F(x)\} \tag{4}$$

For two feasible solutions $u, v \in \mathcal{X}_f$, it is said that u dominates v (termed $u \succ v$) if it is not worst in any objective

1. A metabolic pathway is a coordinated sequence of chemical reactions by which cells transform initial source compounds into final target compounds [10]. Enzymes are protein catalysts in charge of the chemical reactions occurring within the pathway. In each step of the pathway, enzymes convert source compounds (substrates) into target compounds (products) by attaching or detaching chemical groups from substrates.

and it is strictly better in at least one objective [12], [13]. Furthermore, given two possible solution $u, v \in \mathcal{X}_f$, it is said that u and v are non-comparable (denoted $u \sim v$) if neither u dominates v ($u \not\succ v$), nor v dominates u ($v \not\succ u$).

Finding an unique solution that optimises every objective is very unlikely. Therefore, the optimisation usually selects a set of non-comparable solutions. This set is known as the Pareto set.

Definition 2. *Pareto set [12]:* For a given MOP, the Pareto set, termed $\mathcal{P}^*$ is defined as the set of all non-dominated feasible solutions:

$$\mathcal{P}^* = \{x \in \mathcal{X}_f | \nexists\ x' \in \mathcal{X}_f \text{ such that } x' \succ x\} \quad (5)$$

Definition 3. *Pareto front [12]:* For a given MOP, the Pareto front, termed $\mathcal{PF}^*$ is defined as the image in objective space of the Pareto set $\mathcal{P}^*$:

$$\mathcal{PF}^* = \{y = F(x) \in \mathcal{Y}_f | x \in \mathcal{P}^*\} \quad (6)$$

Note that every pair of solutions in $\mathcal{P}^*$ is non-comparable, and therefore they are all equally good from a purely multi-objective perspective when no other criterion is used.

2.1. The Optimisation Procedure

Optimising a multi-objective problem is not a trivial endeavour. In our specific case, our measures are discontinuous and the decision space grows exponentially with the cocktail size. Therefore, we decided to use Evolutionary Algorithms as the optimisation procedure.

Inspired in natural evolution, these algorithms simulate the process of natural selection in order to optimise the problem at hand. They use the concepts of *selection*, *crossover* and *mutation* as tools to take a *population* of *genes* through a simulated sequence of *generations* in which they will adapt to the environment imposed by a *fitness function* [14].

We used NSGA-II [11], an evolutionary algorithm that is widely used to solve MOPs [12]. It is based on the idea of using Pareto dominance to achieve elitism, and introduces the concept of *crowding distance* to preserve diversity in the population.

2.2. Encoding of the Drug Cocktails

Every solution of the optimisation problem is a drug cocktail. A drug cocktail x is encoded as a set of n drugs, taken without repetition — i.e. $x \in D^n$, where D is the set of all FDA approved drugs mapped onto *T. cruzi* enzymes. Formally, the domain of the optimisation problem is:

Definition 4. *Drug Search Space:* the search space $\mathcal{X}^n$ is defined by:

$$\mathcal{X}^n = \{x \in D^n\} \quad (7)$$

where a cocktail $x = (x_1, x_2, \cdots, x_n)$ has no repeated drugs $x_i \neq x_j \forall i \neq j$.

Note that this creates a search space for every n. Intuitively, it defines each configuration of the optimisation problem as "choosing the set of n-drug cocktails from D that simultaneously optimises every measure".

2.3. Encoding of the Biological Measures

2.3.1. Number of enzymes. This function only considers how many enzymes are targeted by a drug cocktail. It returns the number of distinct enzymes that are homologous with at least one drug-targeted protein from the input set. Formally:

$$E_{\text{hit}}(x) = \left\| \bigcup_1^n E(x_i) \right\| \quad (8)$$

where $E(x_i) \subseteq E$ is the set of enzymes targeted by drug x_i, and $E = \{e_1, e_2, \cdots, e_\alpha\}$ is the set of *T. cruzi* enzymes. The minimisable version of this measure is the fraction of enzymes not hit by the cocktail:

$$f_1^n(x) = 1 - \frac{E_{\text{hit}}(x)}{\alpha}. \quad (9)$$

This measure is optimised when $f_1^n(x) = 0$, as $E_{\text{hit}}(x) = \alpha$ enzymes are hit by the cocktail — i.e. all of the enzymes are being hit. As E_{hit} decreases, f_1^n will approach 1, effectively transforming our measure into a minimisable function.

2.3.2. Number of pathways. This function verifies how many pathways are being targeted by a drug cocktail. The number increases for every distinct pathway that contains at least one enzyme which is homologous with a drug target from the cocktail. Formally:

$$P_{\text{hit}}(x) = \left\| \bigcup_1^n P(x_i) \right\| \quad (10)$$

where $P(x_i) \subseteq P$ is the set of pathways targeted by drug x_i, $P = \{p_1, p_2, \cdots, p_\beta\}$ is the set of *T. cruzi* pathways, and $p_j \subset E$ are the enzymes that conform the j-th pathway. A pathway p_j is targeted by drug x_i if it contain at least one enzyme from $E(x_i)$. The minimisable version of this measure is the fraction of pathways not hit by the cocktail:

$$f_2^n(x) = 1 - \frac{P_{\text{hit}}(x)}{\beta}. \quad (11)$$

Analogous to the previous case, as $P_{\text{hit}}(x)$ approaches the total number of pathways β, $f_2^n(x)$ approaches 0.

2.3.3. Number of fully covered pathways. This function retrieves the number of pathways fully covered by a drug cocktail. A pathway is fully covered if each one of its enzyme is targeted by a drug in the cocktail. Formally:

$$P_{\text{cov}}(x) = \|\{p_j | p_j = p_j \cap E(x_i)\}\|_{\substack{1 \leq i \leq n \\ 1 \leq j \leq \beta}} \quad (12)$$

The minimisable version of this measure is the fraction of pathways not fully covered by the cocktail:

$$f_3^n(x) = 1 - \frac{P_{\text{cov}}(x)}{\beta}. \quad (13)$$

Similarly to the previous 2 cases, the best case scenario is achieved when $P_{\text{cov}}(x) = \beta$ and $f_3^n(x) = 0$.

2.3.4. Multi-objective Function. We construct the multi-objective function such that minimising it improves the viability of the drug cocktail. The multi-objective function $F^n(x) : \mathcal{X}^n \rightarrow [0,1]^3$ combines the biological measures (Equations 9, 11, and 13) as follows:

$$F^n(x) = (f_1^n, f_2^n, f_3^n) \tag{14}$$

Therefore, the objective space is:

$$\mathcal{Y}^n = \bigcup_{x \in \mathcal{X}^n} \{F^n(x)\} \tag{15}$$

3. Results

The implementation of the proposed algorithm is available at: http://dev.paccanarolab.org/code/PaccanaroLab/chagas-mop-public. Table 1 shows the parameters of the NSGA-II algorithm. A total of 57 cocktails were found. All of these experiments were computed using a MacBook Pro with a 2.5 GHz Intel Core i7 processor, 16 GB 1600 MHz DDR3 of memory, running macOS Sierra.

Population size	Cocktail size	Run time	Selected cocktails
100	2	4.05 s	5
100	3	8.17 s	7
500	3	19.70 s	7
100	4	23.28 s	11
500	4	158.78 s	11
100	5	130.73 s	17
500	5	202.21 s	17
100	6	112.20 s	16
500	6	181.77 s	17

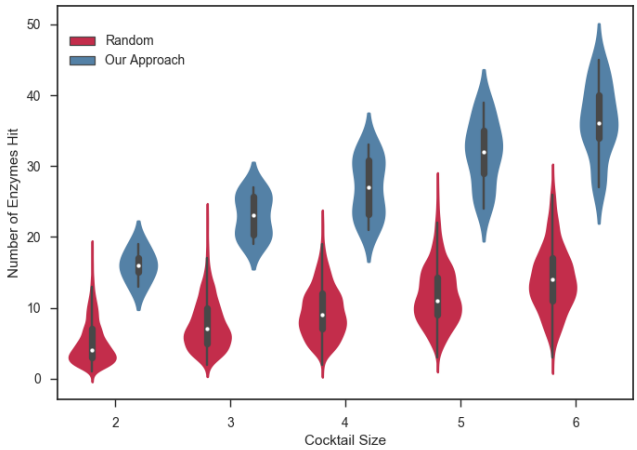
TABLE 1: Parameters used to select drug cocktails. *Population size* is the number of genes allowed to the genetic algorithm to use during the search of the optimal set of cocktails. *Cocktail size* is the number of drugs per cocktail. *Run time* is the time it takes for the evolutionary algorithm to converge to the best solution *Selected cocktails* is the number of cocktails returned by the proposed algorithm.

Figure 1 shows a comparison of the cocktail performance using the three measures separately. Violin plots show the distribution of scores for a random set of cocktails, compared to the scores of the selection given by our method. Our method outperforms the random selection in every measure and cocktail size. Furthermore, as the cocktail size increases, the gap between the randomly selected cocktails and those selected by our method increases significantly.

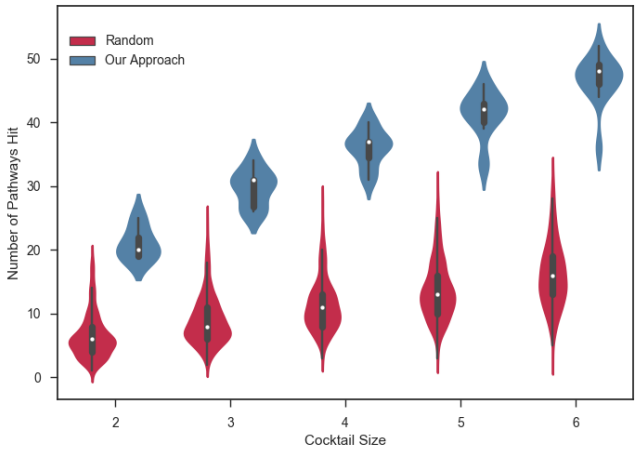
Figure 2 shows the decision space after the algorithm converges for cocktails of 2 drugs. The selected cocktails are highlighted in blue and they represent the best non-dominated drug cocktails for this particular configuration of the problem.

All datasets used as input and the results will be available at https://www.dei.uc.edu.py/proyectos/proyectochagas[2].

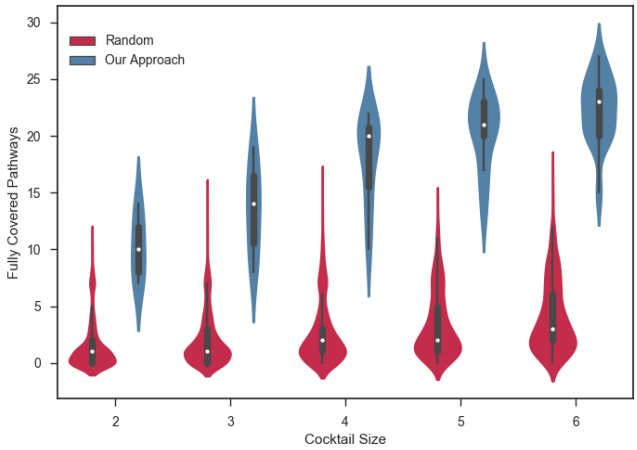
2. This work is part of the CONACyT Project "*Indentificación de Cócteles de drogas para el Tratamiendo de la Enfermedad de Chagas*". Datasets will be released once the project concludes.



(a) Comparison of the distribution of $E_{\text{hit}}(x)$



(b) Comparison of the distribution of $P_{\text{hit}}(x)$



(c) Comparison of the distribution of $P_{\text{cov}}(x)$

Figure 1: Comparison between our approach and randomly selected cocktails. 1000 cocktails were randomly selected from $\mathcal{X}^2$, $\mathcal{X}^3$, $\mathcal{X}^4$, $\mathcal{X}^5$, and $\mathcal{X}^6$ and tested with the 3 measures. Notice that in this plot higher values are better, since we are showing the measures in their initial form as described in Equations 8, 10, and 12.

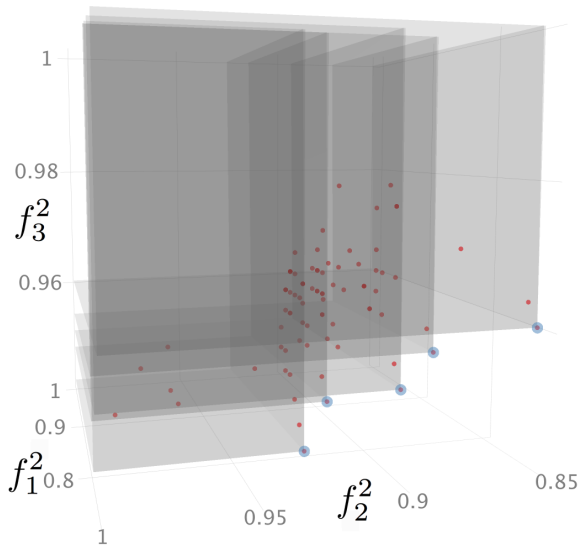


Figure 2: Decision space for cocktails of 2 drugs. The axes f_1^2, f_2^2, and f_3^2 represent the optimisation functions for cocktails of size 2 defined in Equations 9, 11, and 13 respectively. The red dots represent cocktails in $\mathcal{Y}^2$ from the last iteration of the evolutionary algorithm computation. The highlighted dots are the selected cocktails — i.e. the ones that optimise $F^2(x)$. The grey volumes represent the space dominated by solutions in the Pareto front, indicating that the highlighted dots outperform all the red dots.

4. Discussion

The initial pool of drugs we found includes drugs that were already studied for antitrypanosomal activity in the past [15], [16], [17], [18]. This is encouraging evidence that our comparative genomics analysis produces reasonable candidates for drug repurposing. Additionally, the majority of the drugs selected in the drug cocktails were never tested against *T. cruzi*, confirming the novelty of our results.

Our selection of drug cocktails have a varied mechanism of action that increases the likelihood of disrupting essential pathways. Hence, we satisfy our initial objective of establishing a sensible order for the experiments.

In a next stage, several of these cocktails will be assayed using *in vitro* models of pharmacological screenings against *T. cruzi*. The toxicity of the cocktails will also be evaluated on mammalian cells.

The generality of our approach allows it to be reused in many ways. For instance, other neglected parasitic diseases can be targeted using the same rationale. Moreover, it is possible to add further evidence of antitrypanosomal activity, by incorporating other measures to our multi-objective approach.

5. Acknowledgements

This work was funded by the CONACyT Paraguay Grants 14-INV-088 and PINV15-315, and Biotechnology and Biological Sciences Research Council (BBSRC), grants BB/K004131/1, BB/F00964X/1 and BB/M025047/1. The authors thank Dr. Horacio Caniza for insightful discussions regarding the manuscript.

References

[1] W. H. Organization, "Chagas disease fact sheet no. 340," Available at: http://www.who.int/mediacentre/factsheets/fs340/en/, March 2017, accessed: 20.03.2017.

[2] C. J. Schofield, J. Jannin, and R. Salvatella, "The future of chagas disease control," *Trends in parasitology*, vol. 22, no. 12, pp. 583–588, 2006.

[3] I. Ribeiro, A. Sevcsik *et al.*, "New, improved treatments for chagas disease: From the R&D pipeline to the patients," *Plos Neglect Trop D*, vol. 3, no. 7, p. e484, 2009.

[4] J. R. Coura and P. A. Viñas, "Chagas disease: a new worldwide challenge," *Nature*, vol. 465, no. n7301_supp, pp. S6–S7, 2010.

[5] S. R Wilkinson, C. Bot, J. M Kelly, and B. S Hall, "Trypanocidal activity of nitroaromatic prodrugs: current treatments and future perspectives," *Current topics in medicinal chemistry*, vol. 11, no. 16, pp. 2072–2084, 2011.

[6] J. A. Urbina, "Specific chemotherapy of chagas disease: relevance, current limitations and new approaches," *Acta tropica*, vol. 115, no. 1, pp. 55–68, 2010.

[7] F. Nosten and N. J. White, "Artemisinin-based combination treatment of falciparum malaria," *The American journal of tropical medicine and hygiene*, vol. 77, no. 6 Suppl, pp. 181–192, 2007.

[8] J. Touchman, "Comparative genomics," *Nature Education Knowledge*, 2010.

[9] M. Torres, J. J. Cáceres, R. Jiménez, V. Yubero, C. Vega, M. Rolón, L. Cernuzzi, B. Barán, and A. Paccanaro, "Comparative genomics selection of FDA approved drugs with potential antitrypanosomal activity," *Unpublished manuscript and raw data*.

[10] K. Faust, D. Croes, and J. van Helden, "Prediction of metabolic pathways from genome-scale metabolic networks," *Biosystems*, vol. 105, no. 2, pp. 109–121, 2011.

[11] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *Evolutionary Computation, IEEE Transactions on*, vol. 6, no. 2, pp. 182–197, Apr 2002.

[12] C. von Lücken, B. Barán, and C. Brizuela, "A survey on multi-objective evolutionary algorithms for many-objective problems," *Computational Optimization and Applications*, vol. 58, no. 3, pp. 707–756, 2014. [Online]. Available: http://dx.doi.org/10.1007/s10589-014-9644-1

[13] K. Deb, *Multi-Objective Optimization Using Evolutionary Algorithms*. New York, NY, USA: John Wiley & Sons, Inc., 2001.

[14] S. Sivanandam and S. Deepa, *Introduction to genetic algorithms*. Springer Science & Business Media, 2007.

[15] K. Lazardi, J. Urbina, and W. De Souza, "Ultrastructural alterations induced by two ergosterol biosynthesis inhibitors, ketoconazole and terbinafine, on epimastigotes and amastigotes of trypanosoma (schizotrypanum) cruzi." *Antimicrobial agents and chemotherapy*, vol. 34, no. 11, pp. 2097–2105, 1990.

[16] L. J. Goad, R. L. Berens, J. J. Marr, D. H. Beach, and G. G. Holz, "The activity of ketoconazole and other azoles against trypanosoma cruzi: biochemistry and chemotherapeutic action in vitro," *Molecular and biochemical parasitology*, vol. 32, no. 2-3, pp. 179–189, 1989.

[17] M. Cal, J.-R. Ioset, M. A. Fügi, P. Mäser, and M. Kaiser, "Assessing anti-t. cruzi candidates in vitro for sterile cidality," *International Journal for Parasitology: Drugs and Drug Resistance*, vol. 6, no. 3, pp. 165–170, 2016.

[18] C.-K. Chen, S. S. Leung, C. Guilbert, M. P. Jacobson, J. H. McKerrow, and L. M. Podust, "Structural characterization of cyp51 from trypanosoma cruzi and trypanosoma brucei bound to the antifungal drugs posaconazole and fluconazole," *PLoS Negl Trop Dis*, vol. 4, no. 4, p. e651, 2010.