Bioorganic & Medicinal Chemistry 22 (2014) 1568–1585

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Antiprotozoan lead discovery by aligning *dry* and *wet* screening: Prediction, synthesis, and biological assay of novel quinoxalinones

Miriam A. Martins Alho [a], Yovani Marrero-Ponce [b,d,c,*], Stephen J. Barigye [b], Alfredo Meneses-Marcel [b], Yanetsy Machado Tugores [b], Alina Montero-Torres [b], Alicia Gómez-Barrio [f], Juan J. Nogal [f], Rory N. García-Sánchez [f,h], María Celeste Vega [f,g], Miriam Rolón [f], Antonio R. Martínez-Fernández [f], José A. Escario [f], Facundo Pérez-Giménez [d], Ramón Garcia-Domenech [d], Norma Rivera [i,j], Ricardo Mondragón [i,j], Mónica Mondragón [i,j], Froylán Ibarra-Velarde [k], Atteneri Lopez-Arencibia [l], Carmen Martín-Navarro [l], Jacob Lorenzo-Morales [l], Maria Gabriela Cabrera-Serra [l], Jose Piñero [l], Jan Tytgat [e], Roberto Chicharro [m], Vicente J. Arán [n]

[a] *CIHIDECAR (CONICET), Departamento de Química Orgánica, Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires, C1428EGA Buenos Aires, Argentina*
[b] *Unit of Computer-Aided Molecular 'Biosilico' Discovery and Bioinformatic Research (CAMD-BIR Unit), Faculty of Chemistry-Pharmacy, Universidad Central 'Marta Abreu' de Las Villas, Santa Clara 54830, Villa Clara, Cuba*
[c] *Environmental and Computational Chemistry Group, Facultad de Química Farmacéutica, Universidad de Cartagena, Cartagena de Indias, Bolivar, Colombia*
[d] *Unidad de Investigación de Diseño de Fármacos y Conectividad Molecular, Departamento de Química Física, Facultad de Farmacia, Universitat de València, Spain*
[e] *Laboratory of Toxicology, University of Leuven (KULeuven), Campus Gasthuisberg, O&N2, PO Box 922, Herestraat 49, 3000 Leuven, Belgium*
[f] *Departamento de Parasitología, Facultad de Farmacia, Universidad Complutense, 28040 Madrid, Spain*
[g] *Centro para el Desarrollo de la Investigación Científica (CEDIC), Fundación Moisés Bertoni/Díaz Gill Medicina Laboratorial, Pai Perez 1165, Asunción, Paraguay*
[h] *Laboratorio de Investigación de Productos Naturales Antiparasitarios de la Amazonía, Universidad Nacional de la Amazonía Peruana, Pasaje Los Paujiles s/n, A A.H H Nuevo San Lorenzo, San Juan Bautista, Iquitos, Peru*
[i] *Departamento de Bioquímica, Centro de Investigaciones y Estudios Avanzados del IPN, Av. Instituto Politécnico Nacional No 2508, Col. San Pedro Zacatenco, México DF 07360, Mexico*
[j] *Departamento de Microbiología y Parasitología, Facultad de Medicina, UNAM, México DF 04510, Mexico*
[k] *Department of Parasitology, Faculty of Veterinarian Medicinal and Zootecnic, UNAM, Mexico DF 04510, Mexico*
[l] *Instituto Universitario de Enfermedades Tropicales y Salud Pública de Canarias, University of La Laguna, Avda. Astrofísico Fco. Sánchez, S/N, 38203 La Laguna, Tenerife, Canary Islands, Spain*
[m] *Instituto de Química Orgánica General, CSIC, c/Juan de la Cierva 3, 28006 Madrid, Spain*
[n] *Instituto de Química Médica, CSIC, c/Juan de la Cierva 3, 28006 Madrid, Spain*

ARTICLE INFO

Article history:
Received 29 May 2012
Revised 13 January 2014
Accepted 21 January 2014
Available online 31 January 2014

Keywords:
In silico study
TOMOCOMD–CARDD software
Non-stochastic and stochastic linear indices
Classification model
Machine learning-based QSAR
Antiprotozoan database
In vitro assay
Antimalarial
Antitrypanosomal
Antitoxoplasma
Antitrichomonas
Leishmanicide
Cytotoxicity

ABSTRACT

Protozoan parasites have been one of the most significant public health problems for centuries and several human infections caused by them have massive global impact. Most of the current drugs used to treat these illnesses have been used for decades and have many limitations such as the emergence of drug resistance, severe side-effects, low-to-medium drug efficacy, administration routes, cost, etc. These drugs have been largely neglected as models for drug development because they are majorly used in countries with limited resources and as a consequence with scarce marketing possibilities. Nowadays, there is a pressing need to identify and develop new drug-based antiprotozoan therapies. In an effort to overcome this problem, the main purpose of this study is to develop a QSARs-based ensemble classifier for antiprotozoan drug-like entities from a heterogeneous compounds collection. Here, we use some of the **TOMOCOMD-CARDD** molecular descriptors and linear discriminant analysis (LDA) to derive individual linear classification functions in order to discriminate between antiprotozoan and non-antiprotozoan compounds as a way to enable the computational screening of virtual combinatorial datasets and/or drugs already approved. Firstly, we construct a wide-spectrum *benchmark* database comprising of 680 organic chemicals with great structural variability (254 of them antiprotozoan agents and 426 to drugs having other clinical uses). This series of compounds was processed by a *k*-means cluster analysis in order to design training and predicting sets. In total, seven discriminant functions were obtained, by using the whole set of atom-based linear indices. All the LDA-based QSAR models show accuracies above 85% in the training set and values of Matthews correlation coefficients (***C***) vary from 0.70 to 0.86. The external

* Corresponding author. Tel./fax: +34 963543156.
E-mail addresses: ymarrero77@yahoo.es, ymponce@gmail.com (Y. Marrero-Ponce), escario@farm.ucm.es (J.A. Escario), vjaran@iqm.csic.es (V.J. Arán).
URL: http://www.uv.es/yoma/ (Y. Marrero-Ponce).

http://dx.doi.org/10.1016/j.bmc.2014.01.036
0968-0896/

validation set shows rather-good global classifications of around 80% (92.05% for best equation). Later, we developed a multi-agent QSAR classification system, in which the individual QSAR outputs are the inputs of the aforementioned fusion approach. Finally, the fusion model was used for the identification of a novel generation of lead-like antiprotozoan compounds by using ligand-based virtual screening of 'available' small molecules (with synthetic feasibility) in our 'in-house' library. A new molecular *subsystem* (quinoxalinones) was then theoretically selected as a promising lead series, and its derivatives subsequently synthesized, structurally characterized, and experimentally assayed by using in vitro screening that took into consideration a battery of five parasite-based assays. The chemicals **11**(**12**) and **16** are the most active (*hits*) against apicomplexa (sporozoa) and mastigophora (flagellata) *subphylum* parasites, respectively. Both compounds depicted good activity in every protozoan in vitro panel and they did not show unspecific cytotoxicity on the host cells. The described technical framework seems to be a promising QSAR-classifier tool for the molecular discovery and development of novel classes of broad—antiprotozoan—spectrum drugs, which may meet the dual challenges posed by drug-resistant parasites and the rapid progression of protozoan illnesses.

1. Introduction

Diseases caused by tropical parasites affect hundreds of millions of people worldwide, mainly distributed in tropical and subtropical regions. In fact, parasitic diseases have been one of the most significant public health problems for centuries with noteworthy mortality and devastating social and economic consequences. Parasites belonging to *phylum protozoa* are the most important causal pathogens and cause several human infections with globally massive impact. For instance, malaria (*Plasmodium* spp.),[1] leishmaniasis (*Leishmania* spp.),[2] trypanosomiasis (*Trypanosoma brucei* [sleeping sickness][3] and *Trypanosoma cruzi* [Chagas disease][4]) as well as giardiasis[5]/amebiasis[6] (*Giardia lamblia*/*Entamoeba histolytica*) are among the main neglected parasitic diseases with great social impact. Trichomoniasis, one of the most common sexually transmitted diseases (with around 120 million vaginitis infections worldwide every year) caused by the flagellate protozoa *Trichomonas vaginalis*, is increasingly recognized as an important infection in women and men.[7]Other serious disease caused by a related apicomplexan parasite, *Toxoplasma gondii*, has gained increasing relevance in immunocompromised patients, such as patients with transplants, cancer, or AIDS, and in congenitally infected infants.[8]

Although most of the current anti-protozoan drugs are well known and broadly used in medical treatments, most of them are decades old and have many limitations, including the emergence of drug resistance, severe side-reactions (toxicity), low-to-medium efficacy, limitations in the routes of administration, price and other important inconveniences. These drawbacks of current antiprotozoan chemotherapy make the search for new drugs an urgent need. However, the development of such drugs has been largely neglected because they are intended for the treatment of pathologies that mainly affect poor people in regions of the world with limited resources and with scarce marketing possibilities, particularly in today's post-merger climate.

Nevertheless, the search for antiprotozoan compounds is now on the desktop of medicinal chemists and great efforts to reinvigorate the drug development pipeline for these diseases are being addressed by new consortia of scientists from the academy and industry, which are driven in large part by support from major philanthropies.[9] Recently, using whole-organism screening with compounds derived from libraries containing drugs already approved for human use (with other therapeutic use, but 'off-label' like antiparasitic efficacy), a few *hits* were identified in diversity screening assays against *T. brucei*, *Plasmodium falciparum* and *leishmania*.[10–13] In this '*trial-and-error*' search for antiprotozoan drug-like compounds a lot of chemicals had to be experimentally screened (>15,000) and the efficacy of this process was very low, yielding only 3 (and 20 additional in a second study), 19, and 40 know drugs with efficacy equal to or greater than that of the drugs used currently against leishmania-, malaria- or trypanosoma-reference (control) compounds, respectively.[10–13] In addition to the low efficiency of this type of drug discovery landscape, the usually *expensive* and *time consuming* approaches impose on us the necessity to develop alternative and more rational techniques in the classical—*trial and error*—screenings.

In order to reduce costs, pharmaceutical companies have to find new technologies in the quest of new chemical entities (NCE), where an in silico 'virtual' world of data, analysis and computer-aided molecular design can be seen as an adequate alternative to the 'real' world of synthesis and screening of compounds in the laboratory. By such means, '*the expensive commitment to actual synthesis and bioassay is made only after exploring the initial concepts with computational models and screens*'. In silico screening is now incorporated in all areas of lead discovery; from target identification and library design, to *hit* analysis and compound profiling. This theoretical(*dry*)-to-experimental(*wet*) integration procedure will be used here in order to find predictive models that permit the 'rational' identification of new antiprotozoan drug-like compounds.

1.1. Background-review of TOMOCOMD–CARDD method in drug discovery for parasitic diseases: meeting the challenge

Some of our research teams have previously reported several antimicrobial-chemoinformatic studies to drive the selection of novel chemicals as promising NCEs. In these studies, the **TOMOCOMD–CARDD** (acronym of **TO**pological **MO**lecular **COM**putational **D**esign–**C**omputer-**A**ided '**R**ational' **D**rug **D**esign) method[14] and linear discriminant analysis (LDA), have been used in order to parameterize molecules in a database and for developing classification functions, respectively. The LDA is one of most important and simple (supervised, linear and parametric) patter recognition technique that can be used to determine which variables discriminate between two or more naturally occurring groups. The **TOMOCOMD–CARDD** approach is a novel scheme the rational—in silico—molecular design and Quantitative Structure Activity/Property Relationships (QSAR/QSPR).[15–19] It calculates several new families of 2D,3D-Chiral (2.5) and 3D (geometric and topographic) non-stochastic and (simple and double) stochastic (as well as canonical forms) atom- and bond-based molecular descriptors (MDs), denominated quadratic, linear and bilinear indices in analogy to the quadratic, linear and bilinear mathematical.[15–19] For instance, the **TOMOCOMD–CARDD** strategy has been used for the in silico screening of novel molecular *subsystems* having a desired activity against *Trichomonas vaginalis*.[19–21] It was also