DISSERTATIONES
CHIMICAE
UNIVERSITATIS
TARTUENSIS
166

BIRGIT VIIRA

Design and modelling in early drug development in targeting HIV-1 reverse transcriptase and Malaria

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
166

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS
166

BIRGIT VIIRA

Design and modelling in early drug development in targeting HIV-1 reverse transcriptase and Malaria

UNIVERSITY OF TARTU
Press

Institute of Chemistry, Faculty of Science and Technology, University of Tartu, Estonia

This Dissertation is accepted for the commencement of the Degree of Doctor of Philosophy in Molecular Engineering on June 15, 2017 by the Doctoral Committee of the Institute of Chemistry, University of Tartu.

Supervisor: Dr. Uko Maran
Institute of Chemistry, University of Tartu

Opponent: Professor Tom Solmajer, Scientific Counsellor, Laboratory for Molecular Modeling, National Institute of Chemistry, Ljubljana and Faculty of Pharmacy, University of Ljubljana, Slovenia.

Commencement: August 29, 2017 at 12:00, Ravila 14a, room 1021

This work has been partially supported by Graduate School of Functional materials and technologies receiving funding from the European Regional Development Fund in University of Tartu, Estonia

ISSN 1406-0299
ISBN 978-9949-77-515-6 (print)
ISBN 978-9949-77-516-3 (pdf)

University of Tartu Press
www.tyk.ee

TABLE OF CONTENTS

LIST OF ORIGINAL PUBLICATIONS

The present thesis is based on four articles:

I. **Viira, B.**; Selytina, A.; García-Sosa, A.T.; Karonen, M.; Sinkkonen, J.; Merits, A.; Maran, U. Design, discovery, modelling, synthesis, and biological evaluation of novel and small, low toxicity *s*-triazine derivatives as HIV-1 non-nucleoside reverse transcriptase inhibitors. *Bioorg. Med. Chem.* **2016**, 24(11), 2519–29. DOI: 10.1016/j.bmc.2016.04.018

II. **Viira, B.**; García-Sosa, A.T.; Maran, U. Chemical Structure and Correlation Analysis of HIV-1 NNRT and NRT Inhibitors and Database-Curated, Published Inhibition Constants with Chemical Structure in Diverse Datasets. *J. Mol. Graph. Model*. **2017**, *In Press.* DOI: 10.1016/j.mgm.2017.06.019.

III. **Viira, B**.; Gendron, T.; Lanfranchi, D.D.; Cojean, S.; Horvath, D.; Marcou, G.; Varnek, A.; Maes, L.; Maran, U.; Loiseau P.M.; Davioud-Charvet, E. In Silico for Antimalarial Structure-Activity Knowledge and Discovery of Novel Antimalarial Curcuminoids. *Molecules.* **2016**, 21(7), 853. DOI:10.3390/molecules21070853.

IV. Sidorov, P; **Viira, B.**; Davioud-Charvet, E.; Maran, U.; Marcou, G.; Horvath, D.; Varnek, A. QSAR modeling and chemical space analysis of antimalarial compounds. *J. Comp. Aided. Mol.* **2017**, 31(5), 441–451. DOI: 10.1007/s10822-017-0019-4

Thesis author contribution to the articles:

I. Performed synthesis of *s*-triazines and took part in discovery of novel, small and low toxicity *s*-triazine derivatives. Performed HPLC-MS and NMR (nuclear magnetic resonance) analysis and wrote the synthesis part and contributed to the analysis/discovery part of the manuscript.

II. Designed and curated focused HIV-1 RT chemically diverse data set and analyzed this data using hierarchical classification of scaffolds and derived QSAR models for the HIV-1 NNRTIs and NRTIs subsets. I wrote most of the manuscript (data curation and respective data analysis).

III. Designed focused database and curated data for antimalarial compounds from in-house and public data sources. Participated in model building process and wrote the database design portion of the manuscript.

IV. Designed focused database for antimalarial compounds from in-house and public data sources and participated in writing the manuscript.

LIST OF ABBREVATIONS

2,6-DATHTP	2,6-diaryltetrahydrothiopyran-4-ones
AIDS	Acquired Immuno Deficiency Syndrome
DAA	diaryllideneacetone
DNA	deoxyribonucleic acid
EC_{50}	half-maximal effective concentration
ED_{50}	half-maxima effective dose
eq	equivalent
FDA	Food and Drug Administration
HIV	human immunodeficiency virus
HIV-1, HIV-2	human immunodeficiency virus type-1, type-2
HPLC	High-Performance Liquid Chromatography
IC	information content
IC_{50}	half-maximal inhibitory concentration
ISIDA	In Silico design and Data Analysis
K_i	equilibrium dissociation constant determined from an inhibition assay
LIBSVM	Library for Support Vector Machines
NNRTI	non-nucleoside reverse transcriptase inhibitor
NRTI	nucleoside (nucleotide)-analogue reverse transcriptase inhibitor
R^2	squared coefficient of correlation
Ref	reference
RNA	ribonucleic acid
RT	Reverse Transcriptase, reverse transcription
rt	room temperature
s^2	squared standard error of regression
SVM	Support Vector Machine

INTRODUCTION

The process of developing a new drug from scratch is time consuming and costly, taking approximately 10–15 years, and costs around 1 billion USD [1]. This process has been conventionally divided into five stages: (a) basic research, (b) lead target and lead compound(s) discovery, (c) preclinical development, (d) clinical development and (e) filing to drug administration agency. On this pipeline, the basic research stage (a) includes collection and investigation of available data and information about different targets, pathways, known drugs/ bioactive compounds and respective measured bioactivities to find a potential target and new compound(s) to work forward with. The lead target(s) selection of lead discovery stage (b) is a crucial step for helping to identify the most promising approaches before starting to look more into potential drug candidates. The search of the lead compound(s) follows the successful selection of lead target(s) (b). The potential lead compounds have five general requirements [2]. At first, compound must be able to absorb into the bloodstream. For second, it must be able to locate and selectively bind to the proper site of the body. Thirdly, a compound must metabolize efficiently and effectively. Fourth, it must be easily extracted from the body after eliciting the desired functional response from the target. Last and most importantly, the compound must be non-toxic in *in vitro* and *in vivo* tests. In the preclinical development (c), the third stage in drug development pipeline, the goal is to make sure that the found lead compound(s) is (are) ready for clinical development. The clinical development (d) mostly deals with the safety and efficacy issues on patients and includes: safety testing in small group of healthy volunteers (Phase I), safety and efficiency assessing in a small group of patients (Phase II), and safety and efficacy demonstration in a large group of patients (Phase III). After finishing all previously mentioned stages (a-d) successfully, the manufacturing part (e) of the new medicine follows. In developing a new drug, the basic research (a) and lead discovery stages (b) are the challenging steps and form the ground for the success in a drug discovery pipeline. This sets the focus of the present thesis on these first two stages (a, b), i.e. how to collect, curate and prepare data for the computational study, and how to discover and modify lead compound(s).

The essential starting point in early drug development is the collection of existing knowledge about small molecule compounds, with indispensable steps of data curation and verification. Exponential growth of data entries and increase of data heterogeneity and complexity is making data assembly from different sources and diverse experimental protocols a more and more complex problem. The current thesis focuses on two highly prevalent infections affecting many regions in the world: malaria and human immunodeficiency virus 1. While working with anti-HIV-1 (human immunodeficiency virus type-1) compounds, the focus was two-fold. First, design, discovery, modelling, synthesis and biological evaluation of novel *s*-triazine derivatives (article I) and second, on chemical structure analysis of HIV-1 RT (reverse transcriptase, reverse

transcription) inhibitors with available equilibrium binding affinity data; revealing the structures of their scaffold trees and allowing to build QSAR models on this structurally diverse data (article II). When working with antimalarial compounds, the focus was at first on the data curation and focused database assembly (article III, IV) that allowed designing predictive models, which in turn allowed to suggest new chemical structures of the antimalarial curcuminoids, which were successfully externally validated (article III). In following chapters, general overview about HIV-1 and Malaria together with currently known drugs is given, followed by the overview of methods used in the current thesis. Then the research described in the articles will be summarized, along with the major conclusions.

1. LITERATURE OVERVIEW

1.1. Human immunodeficiency virus

The HIV (human immunodeficiency virus) targets the immune system and virus infection progresses to an advanced stage that causes AIDS (Acquired Immuno Deficiency Syndrome). The first case of AIDS was reported in USA in 1981 within young homosexual men, who succumbed and eventually died due to the weakened immune system [3,4,5,6]. The HIV can be transmitted in multiple ways: unprotected sexual intercourse with an infected person, transfusion of contaminated blood and sharing of contaminated needles, syringes or other sharp instruments, and between a mother and her infant during pregnancy, childbirth and breastfeeding. It has been estimated by the WHO (World Health Organization) and UNAIDS (Joint United Nations Program on HIV and AIDS) that 36.7 million people are living with HIV around the world, particularly in Africa, being the epicenter of the epidemic [7]. Successful discovery of the HIV was made in two laboratories, by Robert Gallo and co-workers in USA [8,9] and by Luc Montaigner's group in France [10]. In 2008, Luc Montagnier and Françoise Barré-Sinoussi received the Nobel Prize for their discovery of HIV. There are two different types of HIV: HIV-1 and HIV-2 (human immuno-deficiency virus type-2). The HIV-2 has found to have the lower viral loads compared to the HIV-1 in infected individuals and lower transmission rates [11]. This could be the reason why the HIV-2 is not a pandemic virus like the HIV-1. Up to now, there is no vaccine for HIV-1 and the treatment should rely on suppressing its infection cycle with antiviral compounds. Till today, 28 different such compounds have been licensed as anti-HIV drugs [12]. The FDA (Food and Drug Administration) has approved 24 drugs for treatment of HIV-1 infections, that are divided into six distinct classes based on their molecular mechanism and resistance profiles [12]: (i) NRTIs (nucleoside(nucleotide)-analogue reverse transcriptase inhibitors), (ii) NNRTIs (non-nucleoside reverse transcriptase inhibitors), (iii) integrase inhibitors, (iv) protease inhibitors, (v) fusion inhibitors and (vi) co-receptor antagonists. In medical treatment, the monotherapy was used in early 1990s, nowadays, more common is the dual therapy and the triple drug combinational therapy HAART (Highly Active Anti-retriviral Therapy). [13,14,15] The latter was seminal in reducing the morbidity and mortality related with HIV-1 infection by suppressing dramatically viral load and resulting in a significant reconstitution of the immune system [16,17,18].

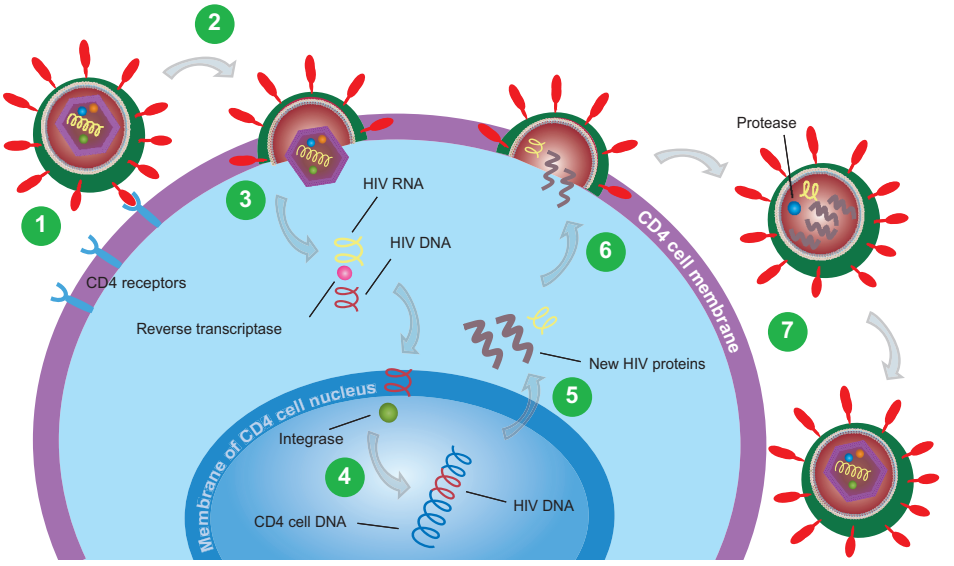


Figure 1. Schematic representation of the HIV life cycle. 1 – Binding; 2 – Fusion, 3 – Reverse transcription, 4 – Integration, 5 – Replication, 6 – Assembly, 7 – Budding (explanation in the text). Figure adapted from https://aidsinfo.nih.gov/education-materials/fact-sheets/19/73/the-hiv-life-cycle.

1.1.1. HIV infection cycle

The HIV strikes aggressively the immune system by attacking and killing CD4 (cluster of differentiation 4) cells, which play an important role in fighting infections in body. The HIV life cycle has seven stages (**Figure 1**). At the first stage, the HIV binds to the CD4 receptors and co-receptors. Followed by the second fusion stage, where the HIV envelope joins the host CD4 cell membrane. In the third stage, reverse transcription of HIV RNA (ribonucleic acid) to DNA (deoxyribonucleic acid) takes place. Integration is the fourth stage, where the HIV releases integrase, using it to integrate HIVs viral DNA into the DNA of the CD4 cell. Integration is followed by replication (fifth stage), where HIV begins to make long chains of HIV proteins. To create a new infectious virion, new HIV proteins and HIV RNA assembles into non-infections HIV as sixth stage. Budding is the seventh stage, where newly formed non-infectious HIV buds itself out of the host CD4 cell and follows protease-mediated maturation to create an infectious viral particle. [19,20]. All mentioned stages are potential targets for antiretroviral drug discovery.

1.1.2. HIV-1 reverse transcriptase

Reverse transcriptase enzyme (revertase, RNA-dependent DNA polymerase, RdDp – RNA-dependent DNA polymerization) is responsible for synthesis of

double-stranded DNA from the single-stranded viral RNA genome. Therefore, reverse transcription is a key step in the life cycle of retroviruses. Within current thesis, it is also a target used in studies. The HIV-1 reverse transcriptase enzyme is heterodimer, consisting two subunits: p51 (440 amino acids) and p66 (560 amino acids). This enzyme exhibits two different enzymatic activities: a DNA polymerase activity, using RNA or DNA as a template and an endo-nucleolytic RNase H (ribonuclease H) activity, destroying RNA in RNA:DNA heteroduplexes [21]. Currently, there are 13 HIV-1 RT antiretroviral agents approved by the U.S. FDA, classified into two different therapeutic groups: NRTIs (nucleotide/nucleoside analogue RT inhibitors) and NNRTIs (non-nucleoside analogue RT inhibitors).

Nucleoside (nucleotide) – analogue reverse transcriptase inhibitors: The nucleoside-analogue reverse transcriptase inhibitors were the first developed and proved to suppress HIV-1 replication as antiretroviral agents. In the mid-1960s, AZT (*azidothymidine* or *zidovudine*) was initially synthesized as a potential anticancer agent, but failed and was put aside at that time [22,23]. About twenty years later, AZT was found to block the reverse transcription step of HIV-1 life cycle and was the first antiretroviral drug approved by FDA for clinical use [22,24,25]. Currently, there are eight available and approved NRTIs (see **Figure 2**) [26]. The NRTIs are nucleoside analogues (dNTPs – naturally occurring deoxyribonucleosides) and targeted to terminate the DNA chain synthesis, due to their lack of 3'-OH group on their sugar moiety [27]. The NRTIs act as prodrugs because they need to be converted to their active metabolites by cellular enzymes to compete with normal nucleotides for binding to the catalytic site of the RT [28]. The NRTIs are also known to be rather toxic due to the additional ability to inhibit the cellular DNA polymerases [29].

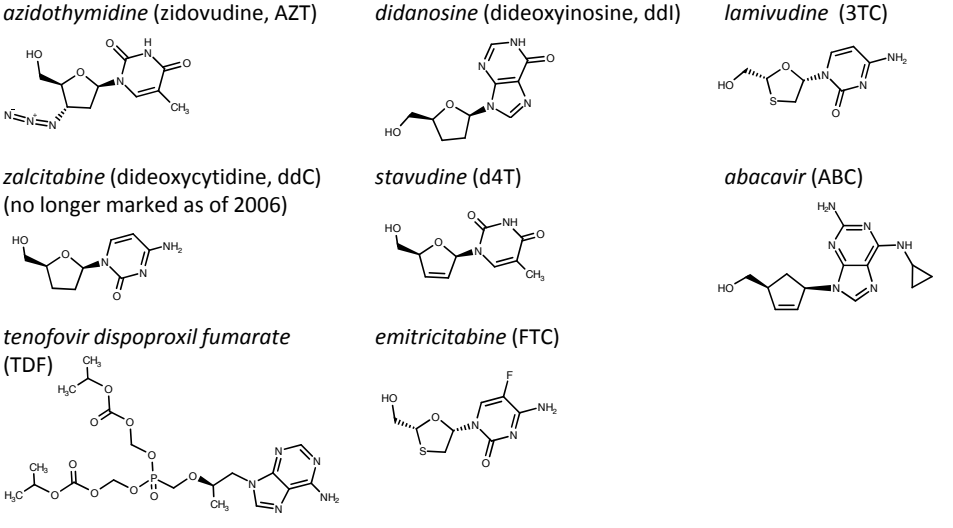


Figure 2. Structure of NRTIs approved by the FDA.

Non-nucleoside reverse transcriptase inhibitors: *Nevirapine* is the first-generation NNRTIs discovered in early 1990s. Currently, there are four available and approved NNRTIs (see **Figure 3**). [26] The NNRTIs are relatively small (<600 Da) hydrophobic compounds with high variety of structures (**Figure 3**) [21]. They all share the same mode of action, by binding directly and non-competitively to an allosteric pocket site of HIV-1 RT, resulting in conformational changes in the enzyme and polymerization inhibition [30]. As opposed to the NRTIs, the NNRTIs do not inhibit other RT lentiviruses like HIV-2 and SIV (simian immunodeficiency virus), therefore, NNRTIs are more selective compared to NRTIs [31,32]. The NNRTI resistance mutations can emerge relatively quickly. Therefore, NNRTIs are mostly used in combination with NRTIs, using the advantage of non-overlapping mechanisms and mutation sites. [27].

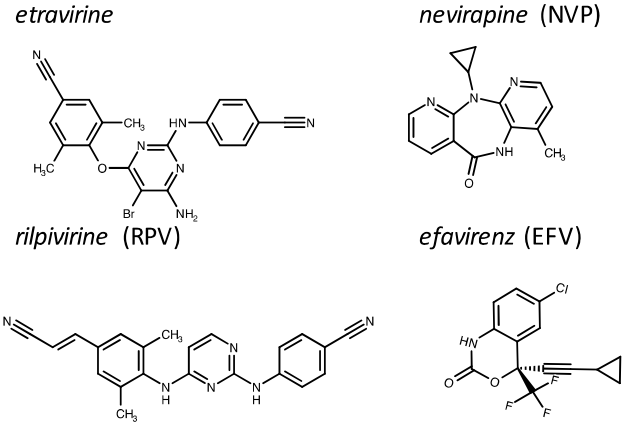
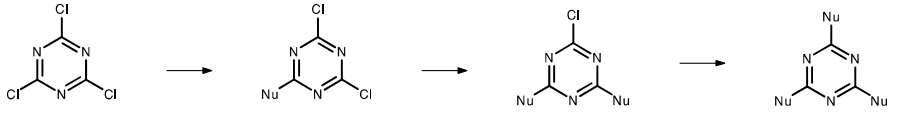


Figure 3. Structure of NNRTIs approved by the FDA.

In addition to the four approved NNRTI molecules (**Figure 3**), there are about 30 different structural classes of compounds reported as potential NNRTIs [33]: for example HEPT (1-[(2-hydroxyethoxy)methyl]-6-(phenylthio)thumine) derivatives [34], TSAO (2',5'-bis-*O*-(tert-butyldimethylsilyl)-3'-spiro-5"-(amino-1",2"-oxathiole-2",2"-dioxide)pyrimidine) derivatives [35, 36], benzothiazine dioxides [37], N1,N3-disubstituted uracils [38], substituted S-DABOs (2-alkylsulfanyl-6-benzyl-3,4-dihydropyrimidin-4(3H)-ones) [39], 2-adamantyl-substituted thiazolidin-4-ones [40], lectins [41], indolobenzothiazepinenones [42] and many others [43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56].

Despite the large number of already known compounds and compound families, new compounds are searched for. This thesis has focused on *s*-triazine (1,3,5-triazines) derivatives that have not been extensively analyzed as NNRTIs. The *s*-triazines have been studied and used other in areas. They have found application in wide range, for example in anticancer therapy [57], estrogen receptor mo-dulators [58], antibacterials [59, 60], antimicrobals [60, 61, 62, 63,

64, 65, 66] and tumor growth inhibitors [67]. Due to large interest, different methods, for example solid phase [16] and combinatorial [9] synthesis, have been developed to synthesize the 1,3,5-triazines with diverse substitution pattern. The cyanuric chloride is the most common and practical starting reagent for synthesis of substituted 1,3,5-triazines (**Figure 4**) [68]. The nucleophilic substitution of each chloride could be controlled by taking into consideration the decrease in reactivity with increasing number of substituents (**Figure 4**) [69, 70, 71, 72].



cyanyric chloride

Figure 4. Stepwise nucleophilic substitution of 1,3,5-cyanyric chloride.

1.2. Malaria

Malaria is a tropical disease, caused by *Plasmodium* parasites and affects about 3.2 billion people around the world. Possible symptoms are high fevers, headache, muscle pain, chills, vomiting and in case of cerebral malaria, neurological complications may occur (brain injury and coma). The rise of morbidity and mortality is caused by the late discovery and lack of appropriate treatment of malaria [73]. Altogether, there are five parasite species known to cause malaria: *P. falciparum, P. vivax, P. ovale, P. malarie and P. knowlesi.* The most prevalent and dangerous, *P. falciparum*, is responsible for the severe form of the disease. The infected female *Anopheles mosquitoes* spread the parasites to people through bites. The infants, children under five years of age, pregnant women, non-immune migrants, mobile populations, travelers and patients with HIV/AIDS are population groups at considerably higher risk of contracting malaria and developing severe disease. [7] Currently, most blood infections could be treated with existing antimalarial drugs, but resistance is causing serious obstacles. The vaccine for malaria it is still under trial, therefore, the main strategies are to use LLINs (long-lasting insecticidal bed nets), RDTs (rapid diagnostic tests) and ACT (artemisinin-based combination therapy) as first-line treatment in endemic regions [74,75].

1.2.1. The life cycle of malaria

Malaria can be transmitted from infected *Anopheles* mosquitos to humans, and the other way around (**Figure 5**). The malaria life cycle begins when the infected *Anopheles* mosquito bites a human (1). As a result, hundreds of sporozoites enter the bloodstream and move to the liver within 30 minutes to infect liver

cells (2). In liver, the parasites start reproducing rapidly and develop into schizonts (3), which rupture and release thousands of individual merozoites into the bloodstream (4). Mereozoites invade erythrocytes and multiply further and restart a new blood cycle every 48 hours (5), infecting more red blood cells, leading to the clinical symptoms of the disease and potential death if not treated. After passing more than ten days, some merozoites develop into gametocytes (6). Another mosquito sucks blood from an infected human (7), they take up the gametocytes, and the cycle continues. The gametocytes are ingested into the mosquito's stomach and mature into gametes (8). The male gamete and female gamete fuse together and produce a zygote (9). The zygotes elongate into ookinetes and burrow through the stomach wall. Then the ookinete develop into an oocyst (11). The oocysts grow and rupture, releasing sporozoites (12). The sporozoites migrate to salivary glands and are ready to be injected and renew the cycle. [76,77]

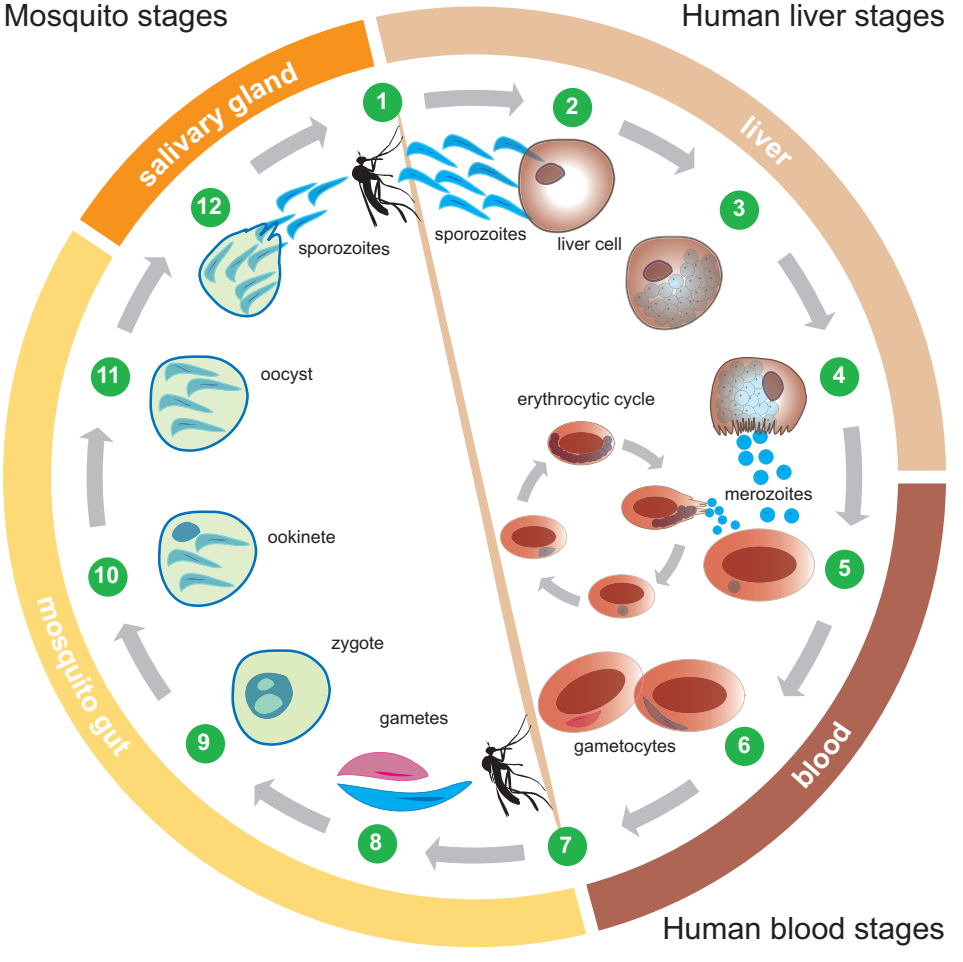


Figure 5. Schematic representation of the life cycle of the Malaria (explanation in the text).

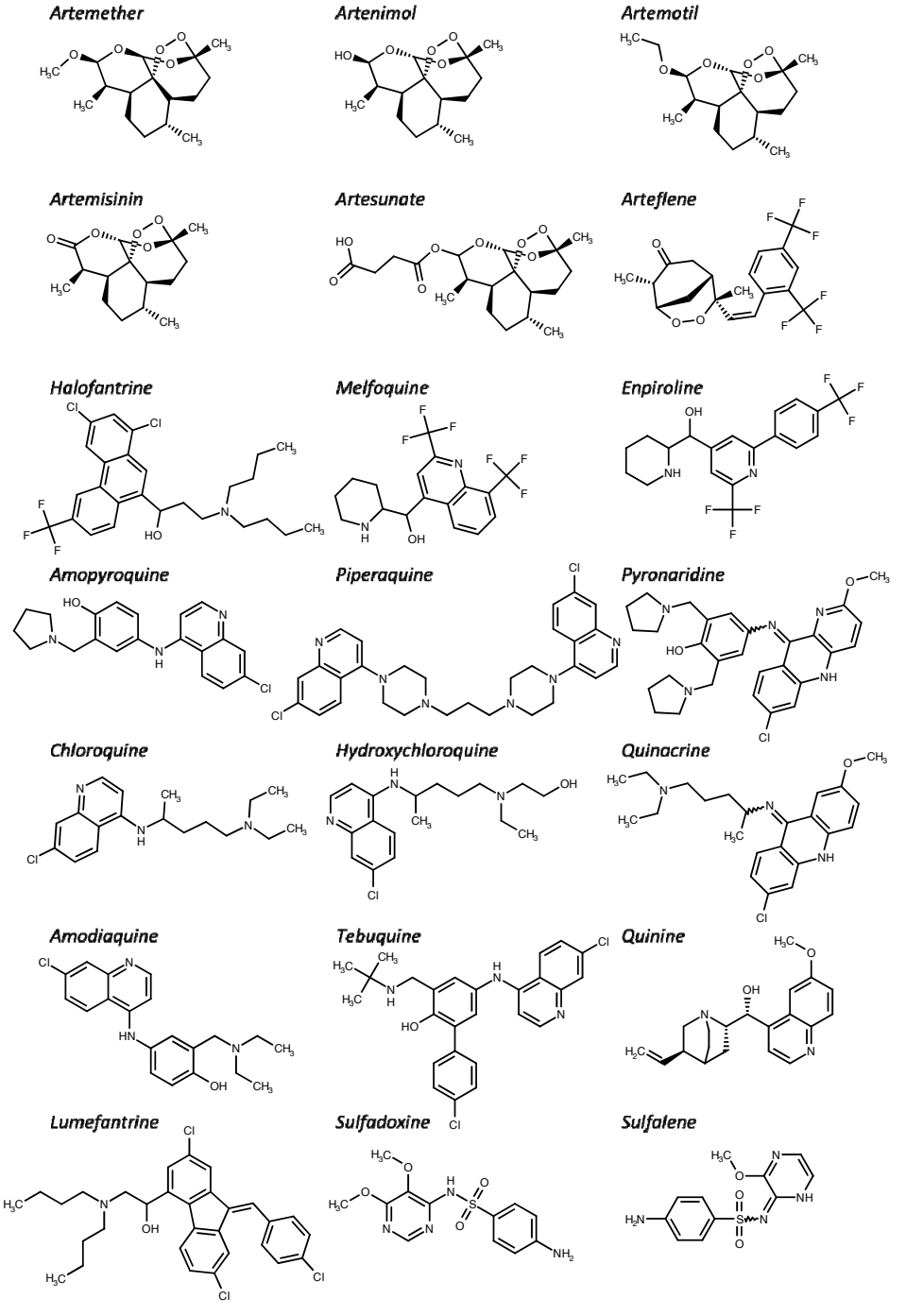


Figure 6. Available and approved antimalarial drugs.

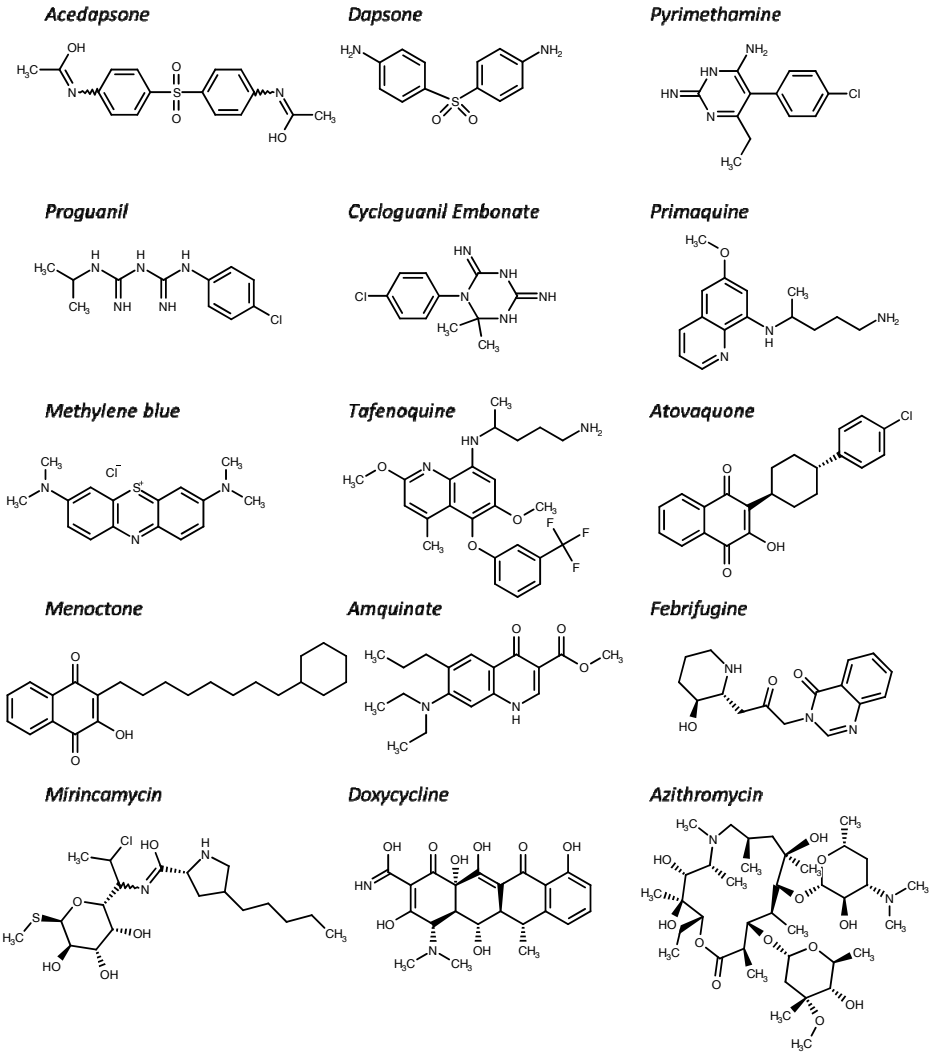
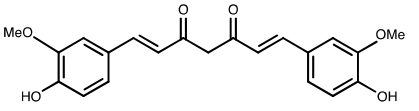


Figure 6. Continues

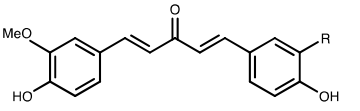
Available drugs for malaria. During the erythrocytic cycle (stage 5 in **Figure 5**), the host hemoglobin is digested and it leads to the production of free heme as a toxic product [78]. The free heme is detoxified as side product into the malarial pigment called hemozoin by the parasite [79,80]. This cycle is an important target pathway for numerous antimalarial drugs, mostly for compounds belonging to the aminoquinoline series [81]. *Quinine* (**Figure 6**) was the first known drug to be efficient against malaria and targeted the erythrocyte stage of the parasite. This was also the starting point for drugs, such as *chloroquine*, *quinacrine* and *primaquine* [82]. There is also *atovaquone*, which is one of the most widely used naphtoquinone drugs [83]. Currently, the WHO recommends using ATC (*artemisinin* combinational therapy). The ATC is targeted in

preventing gametocytogenesis (stages 6 in **Figure 5**) [84]. Also, other mechanisms exist and for example, curcumin, a natural product, has been suggested to kill malarial parasites *via* induced ROS (reactive oxygen species) [85,86] and in this way contribute to ACTs as a valuable compound [87]. In total, 36 antimalarial drugs are approved and available (**Figure 6**).

Known antimalarial curcuminoids and its analogues: Amongst the antimarial compounds, the current thesis focuses largely on curcuminoids (particularly DAAs – diaryllideneacetones) and 2,6-DATHTP (2,6-diaryltetrahydrothiopyran-4-one) series (**Figure 7**). The DAAs were reported to display an effect on several biological activities [88, 89, 90, 91], from anti-inflammatory, anticancer, antioxidant to antiparasitic activities and also antimalarial activities [87, 92, 93, 94, 95, 96]. Curcumin is the base of turmeric spice and the major component of extracts of *Curcuma longa*. Demethoxycurcumin, bisdemethoxycurcumin and monocarbonyl curcuminoids (curcuminoid DAA analogues, R=H, or R=OMe, **Figure 7**) can also be extracted from *Curcuma longa*. [97]. The second series of interest were 2,6-DATHTP derivatives, which are acting as prodrug of DAA and therefore have possibilities as antimalarial candidates. Both DAA and 2,6-DATHTP series were synthesized by dr. Davioud-Charvet's group [98, 99, 100, 101, 102].



Curcumin



Diarylideneacetone (DAA) derivatives

R=H, OMe

Figure 7. Structures of natural curcumin and diarylideneacetone (DAA) derivatives that have been identified in *Curcuma* extracts.

1.3. Link between HIV-1 and malaria

About 36.7 million people are living with HIV-1, and 214 million malaria cases have been estimated worldwide in 2015 [103]. Therefore, both infections are critical global health problems, particularly in developing countries. The interaction between HIV-1 and Malaria is bidirectional [104], meaning that one infection will influence other infection, making it worse or/and the treatment more complicated. Observations on HIV-1 infected pregnant women [105,106] and adults [107,108] show that people with HIV-1 infection are at more risk of having faster development of malaria symptoms and developing severe malaria, than people without HIV-1 infection. Due to the increase of cellular immunosuppression, the immune response to malaria degrades and thus hamper to preventing malaria infection and suppression of parasitaemia [108]. The impact of malaria on HIV-1 is not so clearly distinguished, because immune cells must

be activated, to measure increased HIV-1 viral load [109]. The viral load measurement is performed to detect the amount of HIV in organism and track the progress of the infection. Hoffmann *et al.* [110] and Kublin *et al.* [111] have shown that malaria affects HIV-1 by increasing HIV viral load and accelerates disease progression.

1.4. Data curation and systemization

We are living in the era of chemical data expansion. The vast amount of chemistry data is available from databases that are today mostly online. Just to name a few that have open access: the NIST (National Institute of Standards and Technology) [112], the NCI (National Cancer Institute) [113], the NLM (National Library of Medicine) [114], ChemBank [115], PubChem [116], ChemSpider [117], ChEMBL [118], and DrugBank [119]. Private companies have also created many such databases: including CambridgeSoft [120], SRC (Syracuse Research Corporation) [121], SciFinder [122], etc. Many of these databases are also well annotated, i.e. different bits of data are connected to form information, making the data increasingly useful to aid in the early stages of drug discovery and design, through the application of different computational methods [123].

Due to the vast amount of data available, crucial is the data systemization and curation relative to the research task. Therefore, creating a specialized and focused database or data set gathered from various (or even single) public and private data sources has become increasingly important task in early stage of drug discovery. Moreover, in large amounts of available data, some random errors produced by human entry and/or systematic errors generated by incorrect translation are inevitable, and therefore, the data curation process is an essential step in data pretreatment [124, 125]. Therefore, data curation and systematization has become an emerging issue and research area. It has been extensively discussed and solutions have been provided by several authors [126, 127, 128]. The correct data and the data curation are important in any research area and for computational modelers as well, because the correct data has a substantial impact on model outcome [124].

While systematizing data and generating databases and data sets fit for purpose, it is utmost important that the obtained information about chemical structure and linked experimental data is accurate. Even small errors and non-compatibility in data could lead into propagated and further significant errors. Therefore, reliability of chemical structure and data quality must be assessed. In the literature, one can find proposed workflows for data curation [126, 127, 128]. Analyzing these works one can see that they are composed of two bigger blocks: systematizing-curating data, and filtering data to make it fit for the modelling task. The main components of data systematizing-curation are assessing the reliability of data quality by identifying, characterizing compounds of interest and their experimentally data, using the original source and

information available. This follows a filtering process, where inorganics, organometallics, counter ions and mixtures are removed, remained structures are normalized and then duplicates removed, etc., making the data set fit for the modelling. Finally, and ideally, the obtained data should be double-checked manually, if size of the dataset allows. In this thesis, the data curation is extensively used to build data sets and databases for both HIV-1 and malaria computational studies. Much of the emphasis in this work has been in comparing experimental protocols, in order to understand what type experimental data can be used together, i.e. what data is compatible with each other. In addition to in-house data, the ChEMBL database as single source of external public data has been used in course of research in three articles (II, III, IV) of present thesis.

1.5. Hierarchical classification of scaffolds

The increasing amount of compound collections with biochemical information relevant to drug discover, has opened a need for methods for analyzing and systemizing molecular structures in order to understand the structural diversity and accompanying structural hierarchy of datasets in a systematic manner. Classification of chemical structures and understanding relationships between chemicals allows understanding and reducing the complexity of large compound libraries, by grouping similar or related molecules. Properties of chemicals are mostly determined by different functional groups attached and molecules sharing same scaffold could often share a common synthetic pathway(s) [129]. Therefore, taking advantage of the knowledge of what kind of effect on a property of interest different scaffolds hold and knowing the synthesis pathways, the structures could be modified in a way that the scaffold can be changed, but its desirable properties will be preserved. Several methods and related computer programs for analyzing and visualizing chemical libraries have been presented in the scientific literature. For example, heat-maps and tree-maps are used to present clustered chemical structure data in eXplorer [130]. Radial clustergrams use adjacency of nodes to present parent-child relationships by displaying properties in the chemical structure tree [131]. HierS is an approach developed for compound clustering and analyzing high-throughput screening results by grouping molecular frameworks hierarchically based on the ring systems in the scaffolds that are obtained when all linker bonds are removed [132]. Leadscope Inc. has developed an analysis tool for complex data that uses a manually built, hierarchically sorted dictionary of cyclic and acyclic fragments to analyze structural data sets [133]. In HiTSEE (High-Throughput Screening Exploration Environment), the software focus is on structural similarities in the neighborhood of selected compounds rather than scaffolds. This allows also detecting activity cliffs by spotting big changes in size within a given cluster [134]. CheS-Mapper allows dividing large datasets with small compounds into clusters by arranging them based on their 3D similarity [135]. The SOMs (self-

organizing maps) are used to cluster compounds in ChemSpaceShuttle, implementing a non-linear encoder network and non-linear partial least squares for projection of high-dimensional descriptor vectors into a three-dimensional space [136]. The search and filtering of SMARTs, flags (threshold schemas) and molecular descriptors are used in Screening Assistant 2 allowing to integrate external sparse data in a flexible way via integrating data mining queries and management, interactive visualization, analysis of scaffolds [137]. The SARANEA (structure activity relationship and selectivity analysis) visualizes network of molecular graphs. The main feature of this tool is the calculation and visualization (through color) of a 'cliff index', describing the shifts between neighbors and mainly focused on the exploration of a large set of compounds [138]. The four level hierarchical clustering of compound structure, molecular and topological framework and terminal rings and bonds in SARConnect provides an interface for visualization molecular relationships *via* connecting targets, activities, and compounds from the major internal and external sources [139]. The two-side clustering analysis on drugs and targets based on isomorphism, substructure and fingerprint similarity searches allows to visualize the 2D small molecule structures and is implemented in DrugViz [140].

In the current thesis, a hierarchical classification method of chemical structures implemented in Scaffold Hunter [141] has been used. The unique "parent" scaffolds for each molecule are generated by pruning all terminal side chains by preserving double bonds directly attached to the core structure. The 13 predefined rules, which reflect the following synthetic and medicinal rationales, were used to stepwise simplify scaffolds by removing a single ring form the larger "child" scaffolds to generate smaller 'parent' scaffolds (with the lowest number of acyclic link bonds). The pruning of the molecules (ring system) will start with removing smaller rings first, those with the least number of heteroatoms. If the number of heteroatoms is equal, the priority of heteroatoms to retain is N > O > S. Rings with ≥ 12 atoms should not be removed until there are still smaller rings present, because a macrocycle is considered to be the most characteristic ring system occurring in a molecule. Bridged rings, spiro rings and nonlinear ring fusion patterns should be retained with preference, keeping in mind that bridged ring systems are retained with preference over spiro ring systems. When molecules contain linker(s), then firstly, the rings will be removed where the linker is attached to a ring heteroatom at either end of the linker. From a fully aromatic ring system, the rings should not be removed in a way that the resulting system is not aromatic any more. In case of a mix of aromatic and nonaromatic ring systems, the nonaromatic rings should be retained with priority. Finally, according to the tiebreaking rule, the compound whose canonical SMILES (simplified molecular-input line-entry system) based on the Molinspiration SMILES canonizer [142] has the lower rank in alphabetical order (A has a higher rank than B, and so on), will be removed first. The procedure terminates when a single 'parent' ring is obtained. As a result, hierarchical arrangement of "child" and 'parent" scaffolds' yield branches, which are connected to each other and combined to form a tree with compounds

which are connected to each other and combined to form a tree with compounds families. Branches can be annotated with available biological affinity constants and analyzed to see structure and activity relationships in a more detailed way.

All of the above mentioned computational methods and tools allow making drug discovery and development faster and more economical, and making promising scaffolds, and their hierarchy more apparent. For example, the hierarchical classification method described in more detail above has proven its utility. Wetzel *et al.* have shown that information rich datasets obtained from biochemical or biological screening of large compound libraries can be efficiently analyzed and navigated using hierarchical classification of scaffolds [143]. This method can also be used for linking chemical and biological space to define suitable starting points that guide the synthesis of compound collections with biological relevance [144].

1.7. Quantitative or qualitative structure-activity relationships

The purpose of a (Q)SAR (quantitative or quantitative structure-activity relationship) is to find and study relationships between physical or chemical properties or biological activity and the chemical structure of compounds when the structure has been expressed in the form of molecular descriptors. When this relationship is established and mechanistically explained, the final outcome will be predictive model(s) that allow to (computationally) estimate the activity or property of existing and novel compound(s). The schematic workflow displayed in **Figure 8** explains the key steps in a (Q)SAR development and validation or application. The figure has been vertically organized into two groups, data preparation and computational (automated) steps. Horizontal view presents model development and model validation or application workflow. The process starts with **data preparation** (known also as data pre-treatment) that typically includes data collection, systematization and curation (see Chapter 1.4), followed by the training and validation set formation. Evaluation of the data quality and design of the training sets are crucial steps in the model building (see Chapter 1.4). It is utmost important to make sure that provided experimental values are measured by using well-standardized assays, particularly, if the collected data is from different research groups and literature sources making data curation in that respect important step in data preparation for modelling [126, 127, 145, 146].

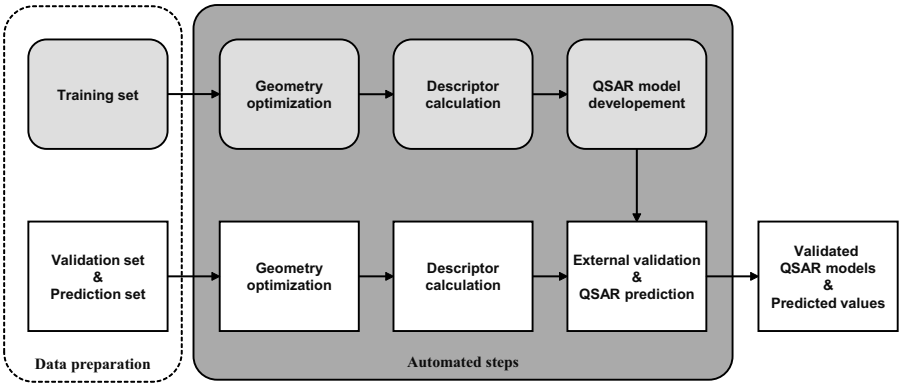


Figure 8. Schematic presentation for successful development of valid the QSARs (adopted from [147])

The next step for both horizontal workflows is **geometry optimization** that, depending on the need, can be accomplished on multiple levels of theory depending on the information one needs for the calculation of molecular descriptors and can therefore range from molecular mechanics calculations to quantum mechanical calculations. The geometry optimization can be skipped when only two-dimensional structures are needed for calculation of molecular descriptors.

The calculation of **molecular descriptors** is the next step of the workflow. Molecular descriptors encode numerical information about the chemical structure. They can be classified into five groups: constitutional, topological, geometrical, electrostatic/charge-related and quantum chemical descriptors depending of the information they capture and level of the theory they have been used to be developed [148, 149, 150]. Molecular descriptors can be calculated for the whole molecule what is one of the most common approaches in modelling of bulk properties of chemicals. The descriptors can be calculated also for atoms and pairs of atoms being useful for modelling specific sites in the molecules and mostly in congeneric data series. The structural composition of a chemical can also be coded in larger fractions of chemical structures, *i.e.* fragments that are useful in analyzing substitution pattern of molecules, but also very diverse data sets.

Molecular descriptors form knowledge space about structural information about molecules in the data set. It is an essential task to find in this knowledge space an appropriate descriptor or set of them that correlate with the property or activity under interest. The link between the molecular structure and the corresponding property/activity is usually achieved by means of statistical methods in a **model development** process. This process is usually the combination of a descriptors selection algorithm and a mathematical representation of model, often working hand-in-hand. Descriptors from a large pool can be selected in multiple ways ranging from forward and backward selection to

genetic algorithms. A typical statistical method for mathematical representation of QSAR is the MLR (multi-linear relationship) method that is easy to understand and the most commonly used method to provide insight to chemical phenomena *via* analyzing relationships between experimental values and molecular descriptors. However, there are also several supervised and unsupervised data processing machine learning algorithms, such as kNN (k-Nearest Neighbors), SVM (Support Vector Machines), RF (Random Forests), DT (Decision Trees), PCA (Principle Component Analysis), PLS (Partial Least Squares), etc. that allow exploring complex relationships between structure and activity [151, 152].

Once the model is obtained, reliability and fitting ability of the predictive model should be verified. This is done using the internal validation of LOO (leave-one-out) and LMO (leave-many-out) cross-validation techniques [153] or using **external validation** with a prediction set as indicated by the second workflow (**Figure 8**). In case of LOO and LMO, the activities of the systematically or randomly excluded compounds are predicted using the developed model and gained predicted results are compared with experimental values. While using external validation, the prediction results characterized with the R^2_{ext} (square-root of correlation coefficient for the external validation set) must be comparable with the model R^2 (square-root of correlation coefficient) value and they should be preferably similar (close to each other), the developed model to be valid. If the R^2_{ext} value is considerably higher than R^2 value, then the model has been poorly represented by the training set. A component of model validation are also model diagnostics techniques that are carried out for the training, validation and also external validation sets in order to understand the applicability domain of the derived model. Simple and applicable to all types of the models is to analyze the ranges of molecular descriptors used in the models and also ranges of property or activity under the study. For the regression type of models, according to the literature, the most commonly used and visually appealing is juxtaposition of leverages and standardized residuals on one plot, commonly known as Williams plot. The leverage for compound is calculated from the matrix of molecular descriptors included into the model [154]. The standardized residuals are calculated from the difference between experimental and predicted value, divided by the standard error of the regression (s) of the training set.

For the modelling of HIV-1 data within the given thesis (Chapter 2.4), the above described workflow computational steps include quantum chemical calculations implemented in MOPAC (Molecular Orbital PACkage) 6.0 [155], using eigenvector following geometry optimization algorithm [156] with AM1 (Austin Model 1) parametrization [157]. The molecular descriptors were calculated with the CODESSA (COmprehensive DEscriptors for Structural and Statistical Analysis) 2.20 package [159] and were extended with $logP_{OW}$ (logarithm of octanol-water partition coefficient) values provided by the PubChem database (XlogP3 algorithm). The BMLR (Best Multi-Linear Regression) method that combines the forward selection of descriptors and

multi-linear regression was used to find relationships between the activities and molecular descriptors [148, 158, 159]. The mathematical representation of the model is simple and allows easy interpretation of the relationship between studied endpoint and chemical structure. The method identifies the orthogonal pairs of descriptor and extends the best correlations with the addition of new descriptors in step forward manner. In other words, the two-parameter regression models are built with the orthogonal pairs and certain number of models with the highest R^2 are selected. New descriptors are added to each previous selected model and the models with higher number of parameters are built. The best presentation of the activity within calculated descriptors is the model with the highest R^2 value. [148] The final MLR model that is close to an optimal description of the dataset within given set of descriptors. It was assessed using several statistical parameters: R^2 (the squared coefficient of correlation), R^2_{cv} (leave-one-out cross-validated squared coefficient of correlation), R^2_{ext} (squared coefficient of correlation of external validation), Y-scrambled (1000 randomization steps) R^2_{scr} (squared coefficient of correlation), F (Fisher's criterion), s^2 (the squared standard error of the regression), etc.

For the modelling of curated antimalarial data (Chapter 2.1.1), the computational steps of the above workflow skip the geometry optimization step as the descriptors are calculated from the 2D representation of the molecular structure. The chemical structures were described with the 39 different ISIDA (In Silico design and Data Analysis) (for in silico design and data analysis) fragment descriptor series [160]. ISIDA have been used to develop new methods and original software tools for structure-property modelling and computer-aided design of new compounds [161, 162, 163]. The 39 descriptor schemes in ISIDA, correspond to different fragmentation strategies, starting from atom type to force field type-based atom coloring, that have been previously showed to be applicable to model a variety of different biological properties [163]. The activity values (IC_{50}, EC_{50}, etc) in antimalarial dataset were heterogenic, which limited the possibility to use regression type models and directed to the classification models while seeking relationships between activity and molecular descriptors. The models were selected using the evolutionary optimizer tool [164], designed for the LIBSVM (library for support vector machine algorithm [165]. The evolutionary optimizer produced for each data set a family of SVM classification models by simulating the Darwinian competition of parameter configurations (so called "chromosomes"). The models are ranked by a "fitness" score, which is an expression of the mean balanced accuracy of classification achieved during the repeated, three-fold cross-validation. Three-fold cross-validation means that the training set has been selected randomly and will be used for the model development and then challenged to predict the remaining compounds, kept out during the training (calibration) phase.

2. RESEARCH RESULTS

The research results are presented in the conventional order of drug discovery and design during the basic research (a) and small molecule compound(s) discovery (b) stages. Research has been grouped into three major parts as carried out by the author of the thesis: the collection, curation and preparation of data for the computational study; discovery and modification of bioactive compounds; understanding and analyzing known chemical space for new discoveries and new *in silico* predictive models to facilitate new discoveries.

2.1. Curating in-house and literature data

During the research the tasks for the analysis, systematization, curation and filtering of data were varied. In one task, existing in-house data was curated and particularly measurement protocols were systematically analyzed and grouped, providing next step to the data curation, mapping in-house data with the database data with aid to extend the structural diversity of data for modelling purposes (Article III, IV). In the second task, the data in the public database was curated for the additional grouping and filtered for the analysis of structural space and *in silico* model development (Article II).

2.1.1. Consolidation of in-house and literature data for the design of Malaria inhibitors

The database or dataset for *in silico* predictive modelling requires the training sets to be as large and as diverse as possible. However, the particular problem in antimalarial research is not the scarcity, but the heterogeneity of the data, and one part of it is diversity and variability of experimental protocols used for the measurements. This makes it meaningful to assess the feasibility of fusing various protocol-specific sets into larger training sets. For that, one needs to understand protocols well and this forms the core of current data curation task. A challenge consisted of curating and fusing the various data sources, the in-house data and literature data from the malaria subset of ChEMBL. To cope with this challenge, the strategy was introduced that allowed compiling coherent training sets where compound structures are associated to the respective antimalarial activity measurements.

In-house protocols and corresponding chemical structures: The in-house source of the antimalarial data against the human pathogen *Plasmodium falciparum* in cell culture was measured at EDC's (Dr. Elisabeth Davioud-Charvet) laboratory (Article III), being the perfect starting point as the analysis protocols are in this case well known. The data structure was analyzed, systematized into five different data tables and organized into a database using

Instant JChem [166]. In this database, each chemical structure was associated with one or several experimentally determined antimalarial activity values. The structure of the tables and connections between the tables are available in **Figure 9** and explained in more detail in the Appendix. All Structures, Properties (one for in-house and one for ChEMBL data), Protocols and References organized into the database showed that antimalarial activity has been measured in-house with eleven protocols for 266 recently synthesized compounds. The filtration of duplicates and salts left 182 compounds fit for further data consolidation and modelling tasks.

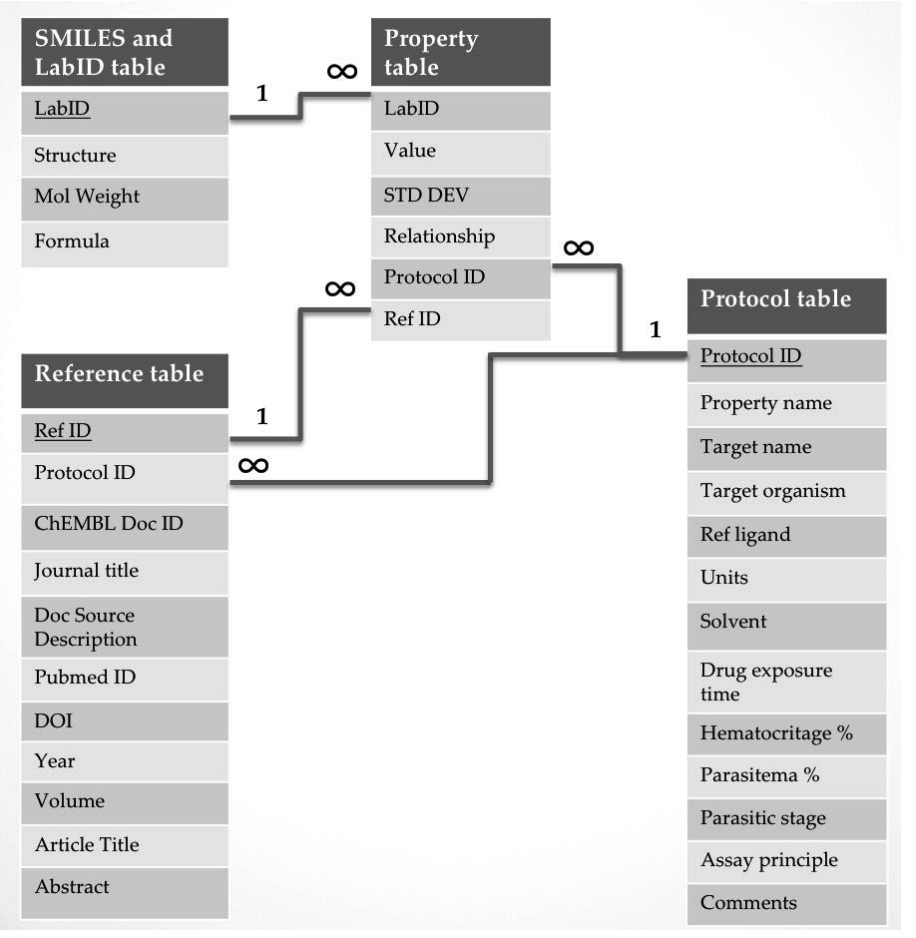


Figure 9. Connections between data tables.

When organizing data from multiple measurements, the coincidence of experimental protocols becomes vital, because the conditions of antimalarial protocols tend to vary. In order to distinguish protocols, six relevant parameters were compared: strain of *Plasmodium falciparum* (Target organism), Drug exposure

time, Parasitic stage, Assay principle, Hematocrit % and Parasitema %. Number of compounds measured for these protocols ranged from 4 to 69. This data was solid and originated form one laboratory. Parameters relevant for the protocol were set for the purpose of potential expansion with data that is available from the literature and allows wider structural coverage for planned consensus modelling.

Comparable protocols and data from ChEMBL: The most challenging in exploiting the antimalarial literature data is how to interpret the various protocols used for testing. The ChEMBL electronic database [118] (version 20) was used to increase the structural diversity and number of compounds in data series corresponding to the protocols used in-house. The structures and corresponding experimental data was retrieved with two different data queries. The purpose of the first query was to search for compounds that have available activity values measured using similar protocol conditions in comparison with in-house protocols. Only compounds were considered that have been reported with IC_{50} values in the nano-molar range in order for them to be comparable with obtained in-house data. The sixty-two compound series with 2295 compounds match the query in ChEMBL database and were added into the combined EDC-ChEMBL antimalarial database. This primary search query was based on EDC in-house protocols, focused on protocol parameters, and therefore narrowed the search significantly.

The second search query was broader, to fully exploit the malaria initiative data from ChEMBL and was therefore focused purely on the target, *Plasmodium falciparum* (Target ID: CHEMBL364), and therefore, search results included also EC_{50}, ED_{50} values, etc., in addition to IC_{50}. This resulted in 249,658 compounds with 400,176 measured activity values. The retrieved compounds were divided into 2,900 different experimental assay-based series. From the second query, the majority of data series were rather small (<50 compounds) and where therefore discarded. Thirty series containing more than 50 compounds were carefully studied for their data quality, keeping only dose-response-based activity values, and series that were redundant or with too few active compounds were discarded. Seven series fulfilled these criteria and were used in further work (CHEMBL730080, CHEMBL896244, CHEMBL896245, CHEMBL1038869, CHEMBL1038870, CHEMBL730081 and CHEMBL73061, in Article III).

After merging EDC in-house and ChEMBL data series into the database, the structures were studied, chemically unstable structures were rejected, and the remaining structures were standardized using a virtual screening workflow installed on a web server realized with the ChemAxon toolkit [167]. This work-flow was designed to remove compounds with heavy metal(s), molecules with high molecular weight, salts, N-oxides with split formal charges, convert com-pounds to their predicted most stable tautomeric form and 5- and 6-membered aromatic rings to the “basic” aromatic forms, etc. After standardization of structures, duplicates were identified and removed. This activity completes the first stage in performing curation, standardization and organization of data that

was previously present in-house and based on the protocol information was complemented with the literature data (Article III, see Figure 3:a and accompanying description).

The analysis of protocols and structural content of data series suggests merging the experimental data series into training sets. For this, two different steps were performed (Article III, Figure 3: b, c). In the first step, the six key parameters of in-house and ChEMBL protocols were compared and data series were merged if the protocol conditions overlapped (Article III, Figure 3: b). Five protocols shared the same protocol conditions and therefore could be merged. In the second step (Article III, Figure 3: c), the experimental values of most commonly used reference compounds (*artemisinin*, *atovaquone* and *chloroquine*) were compared. If their values are nearly equal (within expected experimental error, 0.5 log units for *p*IC/*p*EC), the series were further merged and the data sets' chemical diversity was extended. Comparing the experimental values of reference compounds in each protocol also showed if there were any protocol parameters, out of six key parameters, which might not have a significant effect on the measured values and therefore could not be considered further. The comparison of experimental values of reference compounds resulted in two main conclusions. Firstly, the difference in Parasitema % does not have a significant influence on the measured bioactivity values of reference compounds compared to other five key parameters. Secondly, if assays were different (SYBRGreen or 3H-hypoxanthine), and all remaining conditions are same (Target organism, Drug exposure time, Hematocrit %, Parasitic stage), then they can be merged. As result, the 20 series were merged into 10 training sets pairwise.

The total number of all merged series with the same protocol conditions and unmerged series formed together 30 training sets. Each of them included more than 50 molecules and span a significant activity range (i.e., contain both active and inactive) (Article III, Figure 3d). For seventeen of these training sets, successful consensus SVM classification models were obtained discriminating compounds that have a significant probability to be active under the specific conditions of the antimalarial test associated with each set (see further details in Chapter 2.2.2).

2.1.2. Curating and filtering database data for modelling HIV-1 RT structure space

In order to understand the structural space of HIV-1 RT inhibitors and derive *in silico* predictive models for estimating their equilibrium binding affinity, the data for HIV-1 was extracted from the ChEMBL database (version 20) (Article II). The primary search focused on the target "human immunodeficiency virus type 1 reverse transcriptase", yielding 3 282 different structures with 7 187 records of measured bioactivities (IC_{50}, ED_{50}, K_i, % inhibition, etc.). Extensive spreadsheet analysis of retrieved records made clear that several compounds

lacked available measured values (371), i.e., the column for the numerical value was left empty; or the value relationship was not always exact (1 179), i.e., it was higher or lower than a given value, or not available at all. While assembling data set with various measured bioactivities, the outcome is rather diverse and values from different methods or assays are not comparable with each other. The comparison of experimental methods or assay details can be very tricky, because not all details of methods or assays are always available. The precise K_i (equilibrium dissociation constant determined from an inhibition assay) values (i.e., the measured value equals a given numerical value) in the nM units were chosen as K_i values measured with different methods or assays are comparable. After applying previously mentioned steps, 234 K_i values out of 5 637 data points with experimental values remained in the data set, showing that impression on quantity of data can be deceptive. From the 234 data points, only the ones with the highest confidence score by ChEMBL (9 – direct single protein target assigned) were used. When one compound had more than one value with the same confidence score, then the median of experimental values was used. Unfortunately, the clear distinction between NRTI's and NNRTI's is not present in the ChEMBL and this requires extra data curation steps from the original literature. Indeed, the structures in ChEMBL allow a first premonition. However, to assure, that all compounds were grouped correctly, all the original references were evaluated also for the correctness of experimental values. This data was supplemented with two recently discovered new *s*-triazine derivative NNRTI inhibitors (Article I: compounds **18a** and **18b** in the original publication). The final dataset included altogether 71 compounds, with 39 NNRTIs (Article II, Table 1) and 32 NRTIs (Article II, Table 2). From these tables, it could be seen that, in the case of NRTIs, the majority of compounds have more than one measured value with the same confidence score and the situation was the opposite for the NNRTIs. From this, one can make general conclusion, that NRTIs have more repeated measurements with different methods or assays to gain Ki value, than NNRTIs. Hierarchical clustering of chemical structures of these subsets is discussed in Chapter 2.3 and descriptive and predictive QSAR-s in Chapter 2.4.

2.2. Discovery and modification of bioactive compounds

2.2.1. Discovery, design and synthesis of *s*-triazines as HIV-1 RT inhibitors

Previously carried out in silico screening by our research group [168] proposed several new chemicals as inhibitors for HIV-1 RT. A collaboration to experimentally validate these chemicals resulted in Article I. Biological tests of sixteen compounds revealed compound **8** (in Article I, Table 1) as a new potent HIV-1 RT inhibitor. The compound **8** belongs to the symmetric *s*-triazines and has not been described previously in HIV-1 research in any publication or patent. It

contains a diaryltriazine core substituted by three aminoalkyl groups, with two N-*p*-tolyl groups and one N-methyltetrahydrofuran group. The purity of the commercial sample **8** (n.p.- non-purified) was verified by HPLC-MS (High-Performance Liquid Chromatography-Mass Spectrometry). Three additional distinctive components (in Article I, Figure S2, compounds **17**, **18**, **19**) were detected, which together represented approximately 11% of the mass of the sample. These components were purified to homogeneity with HPLC and collected. Compounds **17** and **18** were collected together as a mixture, because they were failed to separate well under the used conditions/equipment and the amount of these compounds were considerably small. To verify the anti-retroviral activity of compound **8**, all of the obtained components of the sample (purified compound **8**(p), **19**, and mixture **17**/**18**) were tested for their cytotoxicity and antiviral activity. Testing revealed that the mixture of **17**/**18** was found to be the most active (in Article I, Figure S3) and also lacked any cytotoxic effect, even at highest concentration (in Article I, Figure S4). These results allow to conclude that instead of **8**, components **17** and/or **18** must represent a potent HIV-1 RT inhibitor with a good therapeutic index. This finding also provided the grounds to speculate that the components discovered in the sample could be side-products from synthesis or intermediates and have structures similar to that of compound **8**. Experimental and analytical testing and verification of the original components **17**/**18** and **19** were complicated due to the low amount (approximately 0.2 mg) in the commercial sample. Discovery of these compounds initiated the need for in house synthesis of predicted*s*-triazine inhibitors (Article I). Considering the symmetric nature of the compound **8** core and the molar masses from the mass spectra of the sample components **17**/**18** and **19**, and the synthesis pathways, possible structures were suggested (in Article I, Table 3). The molar mass for component **18** was 307.4 g/mol, but considering the possible substitutions, it could match two structures that differ in one substituent (**18a**: $-NH_3^+$ or **18b**: –OH). This conclusion led to the use of high-resolution LC-ESI-QTOF-MS (liquid chromatography-electronspray ionization-quadrupole-time of flight-mass spectrometry) to separate the original mixture **17**/**18** and determine the structure for **18**. The latter showed that compound **18** must have a structure that contains two amino-tolyl and one hydroxyl group, i.e., compound **18b**. These proposed structures were not reported before and neither were they commercially available. Possible methods for the synthesis of substituted 1,3,5-triazines were adopted from the literature and modified (Chapter 1.1.2). The functionalization of the less expensive reagent cyanuric chloride was selected as a promising approach. It has been shown that nucleophilic substitution of each chloride is successfully controlled by taking advantage of the decrease in reactivity with increase of the number of substituents. The 2,4,6-substituted 1,3,5-triazines were obtained from cyanuric chloride by sequential substitution of the chloride atom using nitrogen and oxygen centered nucleophiles (**Figure 10**), using different solvents and substituents to achieve the desired products. The current synthesis scheme enabled to obtain compounds **8**, **17**, **18a**, **18b** and **19** (in

Article I, Table 3) in large quantities and with considerable yield. Mono- and di-substituted derivatives of **8** were prepared by the reaction of cyanuric chloride with the corresponding aminoalkyl group in DCM (dichloromethane) using di-isopropyl amine as a base at ambient temperature (**Figure 10** methods i and ii). In order to obtain the monosubstituted compound, the reaction was carried out at 0 °C or below. To substitute the first chlorine with an amino group, the solution was kept at rt (room temperature), and the reaction finished rapidly (**Figure 10** method iii). The substitution of the remaining chloride atom(s) required higher temperature and longer reaction time (**Figure 10** methods iv-vi). The yields of the obtained compounds were between 60–99% (Article I). Toxicity and antiretroviral activity of the synthesized compounds were analysed. The obtained results were consistent with previously obtained experimental data for sample components (in Article I, Section 2.2) and effective binding to the active site and structural interactions that determine these interactions were also verified computationally (in Article I, Section 2.4).

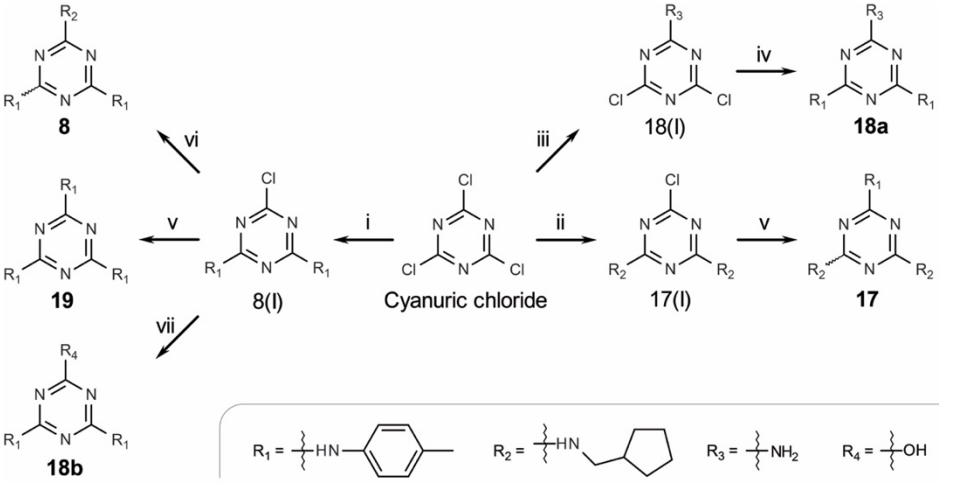


Figure 10. General synthetic pathways used for the synthesis of 2,4,6-trisubstituted-1,3,5-triazines. i) cyanyric chloride, p-toluene (2 eq), diisopropylamine (2.85 eq), DCM, 0 °C to rt; ii) cyanuric chloride, tetrahydrofurane-4-amine (2 eq), diisopropylamine (2.85 eq), DCM, 0 °C to rt; iii) cyanyric choride, NH_4OH (1 eq), DCM, rt; iv) 2,4-dichloro-1,3,5-triazine, tetrahydrofurane-4-amine (2.5 eq), 1,4-dioxane, reflux; v) 2-chloro-1,3,5-triazine, p-toluene (1.5 eq), 1,4-dioxane, reflux; vi) 2-chloro-1,3,5-triazine, tetrahydrofurane-4-amine (1.5 eq), 1,4-dioxane, reflux; and vii) 2-chloro-1,3,5-triazine, H_2O (0.5 eq), dest.H2O, TFA, reflux.

2.2.2. Classification models for antimalarial compounds and experimental validation of predictions

The 30 training sets resulting from the data curation and mixing from the two resources (described in Chapter 2.1.1) were subjected for building qualitative predictive models (methods in Chapter 1.6). The original experimental quantitative values were converted into binary classes of "inactive = 1" and "active = 2". Separation in each training set was done by picking the best cut-off values on the log scale of 7.5–6.0 and separating active compounds from inactive. The goal was to obtain the threshold for the active compounds between 1/3 and 1/4 of the entire set. The cutoffs ranging from the micro-molar scale were increased stepwise by1/2 log until the previously set goal was achieved.

The series of 39 different ISIDA fragment descriptor sets were calculated for each training set to find out which descriptor strategy would be the most suitable for defining the antimalarial activity. The LIBSVM support vector machine algorithm together with the evolutionary optimizer tool was used on the 30 training sets, to select the actual descriptor space for modelling. The "fitness" score, that gives the expression of mean balanced accuracy of classification achieved during the repeated three-fold cross-validation, was used to evaluate the models. The obtained models are consensus models, because they included several independent predictive equations, each calibrated on random subsets of the training set and then used to predict the remaining compounds kept out during the training (calibration) phase. The sets, which had fitness scores 0.7 and higher, were considered as sufficient. The seventeen sets out of initial 30 achieved this criterion (in Article III, Table 3, and Table S3 for more detailed information about experimental protocol). Each consensus model returns a real value likelihood (percentage) of the candidate to qualify as active, i.e. a "vote" towards the category in which a predicted molecule is assigned. The likelihood of around 50% is inconclusive, because this means that, out of the individual distinct equations composing the consensus model, those predicting "active" and "inactive" are roughly equal. The closer the likelihood is to 100%, the higher the trust is in the hypothesis that compound will be active.

The 17 models were applied in virtual screening of 72 in-house compounds: 35 symmetrical and dissymmetrical diarylideneacetone derivatives (DAA, see Article III, Table S1 for results) and 37 2,6-diaryltetrahydrothiopyran-4-ones (2,6-DATHTP, see Article III, Table S2 for results). Compounds were considered active if at least one model gave the likelihood to be active >70% or more than one model predicted the likelihood to be active >50% (see Article III, Tables S1 and S2 for assignments). According to this approach, 40 compounds (26 DAAs and 14 2,6-DATHTPs) were predicted to be active and 32 compounds non-active against the 3D7 *P. falciparum* strain. From these compounds, 31 compounds with best predicted properties (12 DAA and 19 2,6-DATHTP) including 27 with *in silico* status "A – active", and four "I – inactive", were tested for both their antimalarial potencies against *P. falciparum* and cytotoxicity against human fibroblasts MRC-5 strain (in Article III, Table 1, 2).

Out of predicted active 27 compounds, 17 were confirmed by the experiment (in Article III, Table 1 and 2). The correct predictions were obtained when at least one of the 17 used consensus models "voted" the compounds to be active by a large majority of >70%. The alternative selection scheme, voted active by at least two models at a low majority >50% of compounds, was less successful. Positive is also the fact that compounds predicted as inactive were indeed found as inactive in the experiment, providing extra validation for the computational approach used.

2.3. Hierarchical classification of HIV-1 NNRTIs and NRTIs

The HIV-1 NRTI's and NNRTI's structures curated and filtered (Chapter 2.1.2, and Article II) were analyzed with the hierarchical classification method, using the scaffold tree algorithm [129] implemented in the software Scaffold Hunter [141] (Chapter 1.5). This research was initiated to understand the chemical landscape in two series of inhibitors and, most importantly, how our new discovered *s*-triazines (in Article I, compounds **18a** and **18b**) position on this landscape.

For 39 **NNRTIs,** hierarchical classification showed that current set is very diverse with ten different structural 'parent types': oxazepanone, diazepnanone, piperazinone, pyrazine, oxazinanone, diazinanone, pyridine, pyrrole, thiazole and triazine (in Article II, Table 1 and Figure 1). These NNRTI 'parent types' include at least one nitrogen atom and cycles vary from 5–7 atoms. Annotated activity constants for 39 compounds are from μM up to nM range. For detailed analysis of each branch, see Article II. As general conclusions, the identified scaffold branches and corresponding structures could be divided into four structural groups: PBO (pyrrolobenzoaxepine) derivatives [169] (oxazepanone branch); QXPTs (quinoxalinylethylpyridylthioureas) [170] (diazepanone, piperazinone, pyrazine, oxazinanone and diazinanone branch); halopyridyl and thiazolyl thiourea derivatives (pyrimidine and thiazole branch) and AAP-BHAPs ((alkylamino)piperidine bis(heteroaryl)piperazine analogues) [171] (pyrrole branch). Two new compounds, **38** and **39** (in Article I compounds **18a** and **18b** respectively) form a separate branch and have a distinct difference compared with other curated NNRTIs, meaning that they a new family of symmetric NNRT inhibitors has been identified by the research carried out in the Article I. Their structural parameters common for typical lead compounds (molar weight of close to 300, etc.) allow to conclude that they could be further modified by adding beneficial substituents. When comparing their structure to those of the most active compounds, then an additional cycle with longer linker (ethylthiourea) could easily be added as a fourth cycle, which may increase their potency. Also possible is to modify their parent type by reducing the nitrogen atoms from three to two (pyrazine) or increasing the cycle by adding an additional carbon or even oxygen (oxazepanone) to make a ring with 7 or 8 atoms.

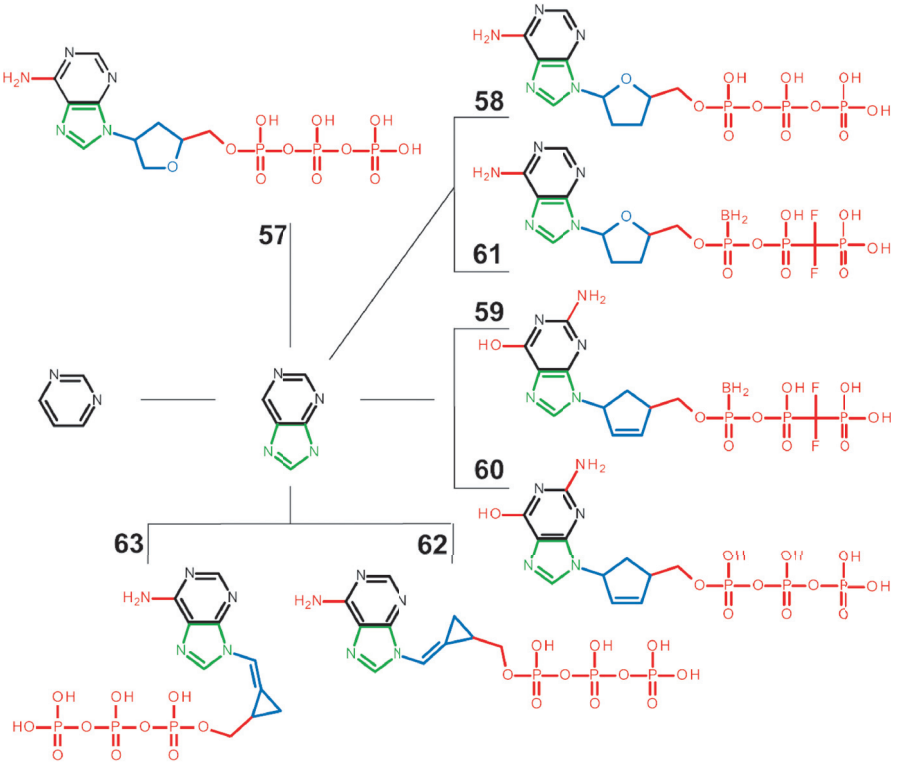


Figure 11. Pyrimidine branch: Color scheme shows the structural hierarchy from parent structure (black) to full compound (black, green, blue, red) and intermediate scaffolds.

The **NRTI** subset is less diverse in comparison with the NNRTI subset and includes four different 'parent types' (in Article I, Table 2 and Figure 14): uracil, pyrimidine, pyrimidione and imidazole. Named 'parent types' involve at least two nitrogen atoms and measured activity constants range from μM to nM. The detailed analysis of each branch is provided in the Article II. The example on **Figure 11** depicts pyrimidine branch, providing color scheme of eliminated groups: at first eliminating functional groups (in red), then first ring (in blue), then second ring (in green), until the parent type structure remains (in black). The distribution of compounds in identified branches is out of balance in terms of number of compounds in branch. This allows to conclude that to now, uracil type of compounds have been most studied. NRTIs being nucleoside/nucleotide analogues, do not have too much "room" to vary in a structural sense, as they should mimic the naturally occurring nucleoside/nucleotide mechanism of action.

2.4. QSAR-s for HIV-1 NNTRIs and NRTIs

Data sets assembled from the ChEMBL database for HIV-1 NNRT and NRT inhibitors are unique because of at least two main reasons. They are in chemical terms very diverse and are characterized by binding affinity equilibrium constants. This, in addition to hierarchical classification and analysis of structural landscape, provides opportunity to derive QSAR models for both types of inhibitors that have not been present before in currently available literature (see Introduction to Article II).

From the hierarchical classification analysis (in Article II, Table 1), one could see that NNRTI's dataset is rather diverse but unbalanced towards structural patterns in the oxazepanone, piperazinone, and pyrazine branches. This gave a challenge for establishing training and validation sets in the way that structural diversity is best covered in the training set. It lead to a chemical structure intuitive approach with four rules, giving approximately 80:20 ratio between training and validation sets that is optimal for the given number of compounds in the dataset. Rules are as follows: (i) each parent structure with only one compound remains in the training set; (ii) for the remaining parent structures, experimental values are aligned from largest to smallest; (iii) for parent compound groups with seven or fewer compounds only one compound from the middle of the alignment was selected to the validation set. (iv) for the two remaining cases, two compounds from the edges of the alignment were selected to the validation set. This approach is essential, because the dataset was chemically diverse and numerically small, where a structural diversity should be represented first and foremost in the training data set. As a result, 31 compounds were listed in the training set and 8 in the validation set (in Article II, Table 1).

$$\log K_{i(NNRTI)} = -12.303\ (\pm 2.075) N_{rel} - 1.637\ (\pm 0.427)\ \epsilon_{LUMO} + 0.628\ (0.295)\ {}^{2}IC + 3.507\ (\pm 1.149)$$
$$R^2 = 0.7080,\ R^2_{cv} = 0.6317,\ R^2_{ext} = 0.7248,\ R^2_{scr} = 0.1,\ s^2 = 0.2230,\ n_{tr} = 31,\ n_{val} = 8 \quad (1)$$

The QSAR model (Equation (1)) for NNRTIs include three molecular descriptors [148, 172]: energy of the lowest unoccupied molecular orbital (ϵ_{LUMO}), relative number of nitrogen atoms (N_{rel}) and average information content (order 2) (^{2}IC). The relationship of binding affinity *vs.* molecular descriptor enables to analyze trend in more detailed. The first two descriptors indicate that compounds with lower LUMO (Lowest Unoccupied Molecular Orbital) energy and smaller content of nitrogen have the higher binding affinity (in Article II, Figure 12: c). All compounds in the NNRTI dataset have at least one nitrogen present in the molecular ring and the number of nitrogen atoms goes up to eight when considering substituents and sidechains. The energy of the lowest unoccupied molecular orbital and the relative number of nitrogen atoms, both have also substantial impact on binding affinity. The nitrogen atom is involved

in electrostatic interactions with hydrophilic amino acids and the LUMO energy has impact on hydrophobic interactions. Both explanations are relevant because the NNRTI binding pocket contains a hydrophobic sub-pocket with aromatic residues Tyr181 (Tyrosine 181), Tyr188 (Tyrosine 188), Phe227 (Phenylalanine 227) and Trp229 (Tryptophan 229) and hydrophilic amino acid residues such as Lys101 (Lysine 101) or Lys103 (Lysine 103) [173]. The third descriptor is topological and has a positive contribution, meaning that compounds with more diverse structural content and bonding pattern have higher binding affinity. In general, the prediction quality of the model shows good stability, particularly towards the external validation set (in Article II, Figure 12).

A similar approach with NNRTIs was used in compiling the training and validation sets as for the NRTIs. For NRTIs, the data points were well-distributed over the scale of experimental values and almost half of the dataset are uracil-based compounds (in Article II, Table 2). 26 NRTIs remained in the training set and 6 NRTIs were separated to the validation set (in Article II, Table 2).

$$\log K_{i(NRTI)} = -35.133\ (\pm 5.479)\ {}^{2}SIC + 0.766\ (\pm 0.015)\ \text{XlogP3} - 13.518\ (4.400)\ V_M/X_{max}Y_{max}Z_{max} + 44.418\ (\pm 5.494)$$
$$R^2=0.7000,\ R^2_{cv}=0.5875,\ R^2_{ext}=0.7445,\ R^2_{scr}=0.12,\ s^2=0.3812,\ n_{tr}=26,\ n_{val}=6 \quad (2)$$

The QSAR model for NRTIs (Equation (2), in Article II, Figure 18) also includes three molecular descriptors: ${}^{2}SIC$ (average structural information content) [172]; XlogP3 (calculated logarithm of octanol-water partition coefficient) and $V_M/X_{max}Y_{max}Z_{max}$ (molecular volume / XYZ Box) [148]. The first and third descriptors have negative correlation with the binding affinity, meaning that compounds with less diverse content and less compact compounds have higher binding affinity. The second descriptor has a positive correlation, showing that, if hydrophobicity of a compound increases, the binding affinity also increases. Comparison of the prediction quality of the model indicates stability, for external validation set R^2_{ext} being higher and for the cross-validated correlation coefficient being smaller in comparison with the R^2; both are strongly influenced by the size of the datasets and distribution of data points (in Article II, Figure 18).

3. CONCLUSION

Current thesis focused on two highly prevalent infections affecting many regions in the world: malaria and human immunodeficiency virus 1. Developing a new drug from scratch is time consuming and costly. This has been conventionally divided into five stages: basic research, lead target and lead compound(s) discovery, preclinical development, clinical development and filing to drug administration agency. Research within present thesis focuses on experimental and computational areas in early drug discovery processes.

In the case of HIV-1, the focus was two-fold. First, based on the earlier multi-objective in silico screening, novel *s*-triazine derivatives (**8**, **17**, **18**, **18a**, **18b**, **19**) were designed, discovered, and synthesized, where findings where supported by the modelling tasks and biological evaluation. The most potent compound **18b** proved to be non-toxic and effective in both cell-based and enzyme experiments. The small molecular size, potent ligand efficiencies, and measured low toxicities permit further exploration of its derivatives. Second, these identified *s*-triazines motivated to analyze the publicly available HIV-1 RT inhibitors to find out how the discovered compounds position on the landscape of NNRTI inhibitors. For this, local focused datasets were created for HIV-1 NNRTIs and NRTIs based on retrieved data from in-house and from online ChEMBL database. Extracted data was carefully filtered, curated and grouped into 32 NRTIs and 39 NNRTIs.

The generated HIV-1 RT inhibitor datasets were analyzed using a hierarchical classification method to discover the structural hierarchy in existing HIV-1 NNRTIs and NRTIs. The NNRTIs and NRTIs were dealt as separate subsets. Analysis of the HIV-1 NNRTIs resulted in ten different common structural 'parent types' and the HIV-1 NRTIs were divided into four different 'parent types'. Previously discovered *s*-triazines formed a separate structural 'parent type' group. Each 'parent type' group was analyzed and examined by means of binding affinity equilibrium constants (K_i) assigned to each individual structure. The scaffolds affecting the potency and stability were highlighted.

The unique datasets allowed the development of descriptive and predictive QSAR models for both HIV-1 NNRTIs and NRTIs groups. The obtained QSAR models include three different, mechanistically relevant molecular descriptors. Analysis of the trends in descriptors revealed important structural parameters and how they influence the binding affinity of molecules. Such QSAR models for data series with binding affinity equilibrium constants have not been published up to now and can aid in predicting new compounds.

To find new promising antimalarial compounds, the experimental anti-*Plasmodium* data from in-house data source and from online ChEMBL database were used. Data from both sources was extensively curated, systematically analyzing experimental protocols and three common compounds. Also, chemical structures were carefully curated, fused and filtered, and finally, a procedure was developed to group different data series into 30 data sets for the

modelling. Out of 30 data series, 17 showed promising models when developed with the support vector machine classification consensus approach. Statistical significance of these 17 models allowed consensus predictions for virtual screening of the series of 72 curcuminoids and 2,6-DATHTPs synthesized in-house. The top ranking positive predictions (27) together with a few predicted as inactive (4) were then submitted to experimental *in vitro* antimalarial testing. A large majority of the predicted compounds (17) showed antimalarial activity, but not those predicted as inactive, thus also experimentally validating the in silico screening approach.

4. SUMMARY IN ESTONIAN

Disain ja modelleerimine HIV-1 pöördtranskriptaasi ja Malaaria ravimite väljatöötamise varajases faasis

Käesolev uurimus keskendub kahele ohtlikule nakkushaigusele: malaaria ja inimese immuunpuuduikkuse viiruse tüübile 1 (HIV-1). Uue ravimi väljatöötamine algusest lõpuni on pikk, keeruline ning kulukas protsess, mida võib tinglikult jagada viieks etapiks: baasuuringud; sihtmärgi ja lähteühendi(te) valimine; eelkliinilised uuringud; kliinilised uuringud; vajalike dokumentide esitamine ravimiametisse. Käesolev töö hõlmab endas kahe esimest ravimiarenduse etappi, nn. ravimite väljatöötamise varajases faasi, jooksul tehtavaid tegevusi.

HIV-1 uurimisel oli kaks põhisuunda. Esmalt teostati uudsete *s*-triasiini derivaatide (**8**, **17**, **18a**, **18b**, **19**) disainimine, avastamine, ja süntees, tuginedes eelnevalt tehtud virtuaalsõelumise tulemustele. Läbiviidud raku- ning ensüümkatsete tulemuste põhjal osutus ühend **18b** saadud ainetest kõige tõhusamaks HIV-1 NNRT inhibiitoriks. Ühendi **18b** väike molaarmass, tugev ligandi efektiivsus ning mõõdetud madal toksilisus soodustavad antud ühendi edasist modifitseerimist. Teiseks suunaks oli HIV-1 inhibiitorite laiem uurimine, et kindlaks teha kas ühendid **18b** ja **18a** moodustavad HIV-1 RT struktuuriruumis unikaalse ühendite grupi. Selleks kasutati olemasolevaid avalikult kättesaadavaid HIV-1 inhibiitorite andmeid. Esmalt eraldati ChEMBLI andmebaasist HIV-1 RT inhibiitorite andmed, kuhu lisati ka kohalikud uudsed *s*-triasiini derivaadid, et võrrelda neid teiste olemasolevate HIV-1 inhibiitoritega. Saadud andmeseeriate hoolika kureerimise ja filtreerimise tulemusena jagunesid ühendid kahe gruppi: 32 NRT ja 39 NNRT inhibiitorit. Andmeseeriates olevate struktuuride vahelisi seoseid analüüsiti kasutades hierarhilise klassifitseerimise meetodit, mis jaotas struktuurid korrapäraselt gruppidesse. Analüüsi tulemusena saadi, et NNRTI andmeseeria jaguneb kümneks erinevaks 'vanema tüübi' grupiks millel on iseloomulik baasstruktuur millest järgnevad keemilised ühendid tulenevad ning NRTI sisaldab nelja erinevat 'vanema tüübi' gruppi. Eelnevalt avastatud *s*-triasiinid moodustasid eraldi seisva 'vanema tüübi' grupi. Leitud 'vanema tüübi' gruppe uuriti ning analüüsiti, lisades juurde ka vastavad mõõdetud seondumise afiinsuse tasakaalukonstandid (K_i) ning toodi välja struktuurid, mis omavad olulist rolli afiinsuse ning stabiilsuse seisukohast. Unikaalsed HIV-1 NNRTI ja NRTI andmeseeriad omakorda võimaldasid arendada kirjeldavaid ja ennustavaid QSAR mudeleid. Loodud QSAR mudelid sisaldasid kolme erinevat mehhanistlikult olulist deskriptorit. Leitud mudelite ning nendes sisalduvate deskriptorite analüüs tõi esile olulised struktuuri parameetrid ning nende mõju seondumisafiinsusele uuritavates ühendites.

Uudsete anti-malaaria ravimikanditaatide leidmiseks koostati fokusseeritud andmebaas eksperimentaalsete anti-*Plasmodium* andmetega. Andebaas sisaldas nii asutusesisesed, kui ka ChEMBLI veebiandmebaasis olevaid andmeid. Saadud andmete põhjaliku ulatusliku kureerimise, filtreerimise ning andmeseeriate ühendamise tulemusena saadi 30 modelleeritavat andmeseeriat, mis sisaldasid

igaüks viiskümmend ja enam keemilist ühendit. Nendel andmeseeriatele tuletati klassifitseerimise nn. üksmeele prognoosmudelid kasutades tugivektor masinate algoritmi. Prognoosmudelite ennustusvõimekuse analüüs näitas, et 17 mudelit, mille ennustusvõime on rohkem kui 70% ja rohkem kui 50% aktiivsete ühendite poole, on kasutatavad juba sünteesitud ühendite järjestamiseks enne kontroll-eksperimendi teostamist. Prognoosivõimeliste mudelitega teostati üksmeele ennustusi asutusesiseselt olemasolevatele *curcuminoidide* seerjale (DAA) ning nende analoogidele (2,6-DATHTP). Ennustustäpsuse järgi järjestatud ühenditest valiti välja kõrgema järjestusega aktiivsed ja ka mitteaktiivsed ühendid millele teostati eksperimentaalne valideerimine *in vitro* katsetega. *In vitro* kastete tulemused andsid 17 potentsiaalset malaaria vastast eelravimit.

REFRENCES

1. Myers, S.; Baker, A. Drug discovery – an operating model for a new era. *Biotech. Nature.* **2001**, 19, 727–730.
2. Hefti, F. F. Requirements for a lead compound to become a clinical candidate. *BMC Neuroscience.* **2008**, 9, 1–17.
3. Brennan, R. O.; Durack, D. T. Gay compromise syndrome. *Lancet*. **1981**, 2, 1338–1339.
4. Friedman-Kien, A. E. Dissemniated Kaposi's sarcoma syndrome in young homosexual men. *J. Am. Acad. Dermatol.* **1981**, 5, 468–471.
5. Gottlieb, M. S.; Schroff, R.; Schanker, H. M.; Weismann, J. D.; Fan, P. T.; Wolf, R. A.; Saxon, A. Pneumocystis *carinii* pneumoonia and mucosal candidiasis in previously healthy homosexual men. *N. Eng. J. Med*. **1981**, 305, 1425–1431.
6. Siegal, F. P; Lopez, C.; Hammer, G. S.; Brown, A. E.; Kornfeld, S. J.; Gold, J.; Hassett, J.; Hirschman, S. Z.; Cunningham-Rundles, C.; Adelsberg, B. R.; Parham, D. M.; Siegal, M.; Cunningham-Rundles, S.; Armstrong, D. Severe acquired immunodeficiency in male homosexuals, manifested by chronic perianal ulcerative herpes simplex lesion. *N. Eng. J. Med*. **1981**, 305, 1439–1444.
7. World Health Organization. HIV/AIDS, Fact sheet n, 360, uptated July 2016. www.who.int/mediacentre/factsheets/fs360/en/ (accessed Sept 29, 2015)
8. Poiesz, B. J.; Ruscetti, F. W.; Gazdar, A. F.; Bunn, P. A.; Minna, J. D.; Gallo, R. C. Detection and isolating of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc. Natl. Acad. Sci.* **1980**, 77, 7415–7419.
9. Gallo, R. C.; Salahuddin, S. Z.; Popovic, M.; Shrearer, G. M.; Kaplan, M.; Haynes, B. F.; Palker, T. J.; Redfield, R.; Oleske, J.; Safai, B.; *et al*. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science.* **1984**, 224, 497–500.
10. Barré-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Daquet, C.; Axel-Blin, C.; Veezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science.* **1983**, 220, 868–71.
11. Sharp, P. M.; Hahn, B. H. Origins of HIV and the AIDS Pandemic. *Cold. Harb. Perspect. Med.* **2011**, 1:a006841.
12. AIDSinfo: Guideline for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. http://aidsinfo.nih.gov/contentfiles/lvguidelines/AdultantAdolescentGL.pdf (accessed April 10, 2017)
13. Collier, A. C.; Coombs, R. W.; Schoenfeld, D. A.; Bassett, R. L.; Timpone, J.; Baruch, A.; Jones, M.; Facey, K.; Whitacre, C.; McAuliffe, V. J.; Friedman, H. M.; Reichman, R. C.; Hooper, C.; Corey, L. Treatment of human immunodeficiency virus infection with saquinavir, zidovudine, and zalcitabine. AIDS Clinical Trials Group. *N. Engl. J. Med.* **1996**, 334, 1011–1017.
14. D'Aquila, R. T.; Hughes, M. D.; Johnson, V. A.; Fischl, M. A.; Sommadossi, J. P.; Liou, S. H.; Timpone, J.; Myers, M.; Basgoz, N.; Niu, M.; Hirsch, M. S. Nevirapine, zidovudine, and didanosine compared with zidovudine and didanosine in patients with HIV-1 infection. A randomized, double-blind, placebo-controlled trial. National Institute of Allergy and Infectious Diseases AIDS Clinical Trials Group Protocol 241 Investigators. *Ann. Intern. Med.* **1996**, 124, 1019–1030.

15. Staszewski, S.; Miller, V.; Rehmet, S.; Stark, T.; De Crée, J.; De Brabander, M.; Peeters, M.; Andries, K.; Moeremans, M.; De Raeymaeker, M.; Pearce, G.; Van den Broeck, R.; Verbiest, W.; Stoffels, P. Virological and immunological analysis of a triple combination pilot study with loviride, lamivudine and zidovudine in HIV-1-infected patients. *AIDS.* **1996**, 10, 1–7.
16. Autran, B.; Carcelain, G.; Li, T. S.; Blanc, C.; Mathez, D.; Tubiana, R.; Katlama, C.; Debre, P.; Leibowitch, J. Positive effects of combined antiretroviral theraphy on CD4+ T cell homeostasis and function in advanced HIV disease. *Science.* **1997**, 277, 112–116.
17. Komanduri, K. V.; Viswanathan, M. N.; Wieder, E. D.; Schmidt, D. K.; Bredt, B. M.; Jacobson, M. A.; McCune, J. M. Restoration of cytomegavirus-specific CD4+ T-lymphocyte responses after ganciclovir and highly active antiretroviral theraphy in individuals infected with HIV-1. *Nat. Med.* **1998**, 4, 953–956.
18. Lederman, M. M.; Connick, E.; Landay, A.; Kuritzles, D. R.; Spitzler, J.; St. Clair, M.; Kotzin, B. L.; Fox, L.; Chiozzi, M. H.; Leonard, J. M.; Rousseau, F.; Wade, M.; Roe, J. D.; Martinez, A.; Kessler, H. Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine, and ritonavir: results of AIDS Clinical Trials Group Protocol 315. *J. Infect. Dis.* **1998**, 178, 70–79.
19. AIDSinfo Home Page. https://aidsinfo.nih.gov/education-materials/fact-sheets/19/73/the-hiv-life-cycle. (accessed Oct 3, 2016).
20. Endelman, A.; Cherepanov, P. The structural biology of HIV-1: mechanistic and therapeutic insights. *Nat. Rev. Microbiol.* **2012**, 10, 279–290.
21. Sluis-Cremer, N.; Temiz, N. A.; Bahar, I. Conformational Changes in HIV-1 Reverse Transcriptase Induced by Nonnucleoside Reverse Transcriptase Inhibitor Binding. *Curr. HIV. Res.* **2004**, 2, 323–332.
22. Mitsuya, H.; Weinhold, K. J.; Furman, P. A.; St. Clair, M. H.; Lehrman, S. N.; Gallo, R. C.; Bolognesi, D.; Barry, D. W.; Broder, S. 3'-Azido-3'-deoxythymidine (BW A509U): an antiviral agent that inhibits the infectivity and cytopathic effect of human T-lymphotropic virus type III/lymphadenopathy-associated virus in vitro. *Proc. Natl. Sci. USA.* **1985**, 82, 7096–7100.
23. Broder, S. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. *Antiviral. Res.* **2010**, 85, 1–38.
24. Klecker, R. W. Jr.; Collins, J. M.; Yarchoan, R.; Thomas, R.; Jenkins, J. F.; Broder, S.; Myers, C. E. Plasma and cerebrospinal fluid pharmacokinetics of 3'–azido-3'-deoxythymidine: a novel pyrimidine analog with potential application for the treatment of patients with AIDS and related diseases. *Clin. Pharmacol. Ther.* **1987**, 41, 407–412.
25. Fischl, M. A.; Richman, D. D.; Grieco, M. H.; Gottlieb, M. S.; Volderding, P. A.; Laskin, O. L.; Leedom, J. M.; Groopman, J. E., Mildvan, D.; Schooley, R. T.; Jackson, G. G.; Durack, D. T.; King, D. and The AZT Collaborative Working Group. The Efficacy of Azidothymidine (AZT) in the Treatment of Pateints with AIDS and AIDS-related Complex. *N. Engl. J. Med.* **1987**, 317, 185–191.
26. AIDSinfo Drug Database. https://aidsinfo.nih.gov/drugs/search/searchterm/2/1 (accessed Oct 3, 2016)
27. Das, K.; Arnold, E. HIV-1 reverse transcriptase and antiviral drug resistance. *Perspect. Med.* **2013**, 1, 111–118.

28. De Béthune, M. P. Non-nucleoside reverse transcriptase inhibitors (NNRTIs), their discovery development, and use in the treatment of HIV-1 infection: A review of the last 20 years (1989–2009). *Antivir. Res.* **2010**, 85, 75–90.
29. Kocnor, C. A.; Lewis, W. Nucleoside reverse transcriptase inhibitor toxicity and mitochondrial DNA, Expert. Opin. Drug. Metab. Toxicol. **2010**, 6, 1493–1594.
30. De Clercq, E. The role of non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the therapy of HIV-1 infection. *Antivir. Res*. **1998**, 38, 153–179.
31. Kohlstaedt, L. A.; Wang, J.; Friedman, J. M.; Rice, P. A.; Steitz, T. A. Crystal structures at 3.5 A resolution of HIV-1 reverse transcriptase complexes with an inhibitor. *Science.* **1992**, 26, 1783–1790.
32. Witvrouw, M.; Pannecouque, C.; Van Laethem, K.; Desmyter, J.; De Clercq, E.; Vandamme, A. M. Activity of non-nucleoside reverse transcriptase inhibitors against HIV-2 and SIV. *AIDS.* **1999**, 13, 1477–1483.
33. Prajapati, D. G.; Ramajayam, R.; Yadav, M. R.; Giridhar, R. The search for potent, small molecule NNRTIs: A review. *Bioorg. Med. Chem.* **2009**, 17, 5744–5762.
34. Baba, M.; Shigeta, S.; Yuasa, S.; Takashima, H.; Sekiya, K.; Ubasawa, M.; Tanaka, H.; Miyasaka, T.; Walker, R. T.; De Clercq, E. Preclinical Evaluation of MKC-442, a Highly Potent and Specific Inhibitor of Human Immunodeficiency Virus Type 1 In Vitro. *Antimicrob. Agents. Chemother.* **1994**, 38, 688–692.
35. Balzarini, J.; Pérez- Pérez, M. J.; San-Félix, A.; Schols, D.; Perno, C. F.; Vandamme, A. M.; Camarasa, M. J.; De Clercq, E. 2',5'-Bis-O-(tert-butyldimethylsilyl)-3'-spiro-5"-(4"-amino-1",2"-oxathiole-2",2"-dioxide)pyrimidine (TSAO) nucleoside analogues: Highly selective inhibitors of human immunodeficiency virustype 1 that are targeted at the viral reverse transcriptase. *Proc. Natl. Acad. Sci.* **1992**, 89, 4392–4396.
36. Balzarini, J.; Pérez- Pérez, M. J.; San-Félix, A.; Velazquez, S.; Camarasa, M. J.; De Clercq, E. [2' ,5 '-Bis-O-(tert-Butyldimethylsilyl)]-3' -Spiro-5"-(4"-Amino-1",2"-Oxathiole-2",2"-Dioxide) (TSAO) Derivatives of Purine and Pyrimidine Nucleosides as Potent and Selective Inhibitors of Human Immunodeficiency Virus Type 1. *Antimicrob. Agents. Chemother.* **1992**, 36, 1073–1080.
37. Brzozowski, Z.; Saczewski, F.; Neamati, N. Synthesis and anti-HIV-1 activity of a novel series of 1,4,2-benzodithiazine-dioxides. *Bioorg. Med. Chem. Lett.* **2006**, 16, 5298–5302.
38. Novikov, M. S.; Valuev-Elliston, V. T.; Babkov, D. A.; Paramonova, M. P.; Ivanov, A. V.; Gavryushov, S. A.; Khandazhinskaya, A. L.; Kochetkov, S. N.; Pannecouque, C.; Andrei, G.; Snoeck, R.; Balzarini, J.; Seley-Radtke, K. L. N^1,N^3-disubstituted uracils as nonnucleosides inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem.* **2013**, 21, 1150–1158.
39. Wang, Y. P.; Chen, F. E.; De Clercq, E.; Balzarini, J.; Pannecouque, C. Synthesis and in vitro anti-HIV evaluation of a new series of 6-arylmethyl-substituted S-DABOs as potential non-nucleoside HIV-1 reverse transcriptase inhibitors. *Eur. J. Med. Chem.* **2009**, 44, 1016–1023.
40. Balzarini, J.; Orzeszko, B.; Maurin, J. K.; Orzeszko, A. Synthesis and anti-HIV studies of 2-adamantyl-substituted thiazolidin-4-ones. *Eur. J. Med. Chem.* **2007**, 42, 993–1003.
41. Akkouh, O.; Ng, T. B.; Singh, S. S.; Yin, C.; Dan, X.; Chan, Y. S.; Pan, W.; Cheung, R. C. F. Lectins with Anti-HIV Activity: A Review. *Molecules.* **2015**, 20, 648–668.

42. Silvestri, R.; Artico, M.; Bruno, B.; Massa, S.; Novellino, E.; Greco, G.; Marongiu, M. E.; Pani, A.; De Montis, A.; La Colla, P. Synthesis and Biological Evaluation of 5*H*-Indolo [3,2-b][1,5]Benzothiazepine Derivatives, Designed as Conformationally Constrained Analogues of the Human Immunodeficiency Virus Type 1 Reverse Transcriptase Inhibitor L-737,126. *Antivir. Chem. Chemoth.* **1988**, 9, 139–148.
43. Merluzzi, V. J.; Hargrave, K. D.; Labadia, M.; Grozinger, K.; Skoog, M.; Wu, J. C.; Eckner, K.; Hattox, S.; Adams, J.; Rosehthal, A. S.; Faanes, R.; Eckner, R. J.; Koup, R. A.; Sullivan, J. L. Inhibition of HIV-1 Replication by a Nonnucleoside Reverse Transcriptase Inhibitor. *Science.* **1990**, 250, 1141–1143.
44. Goldman, M. E.; Nunberg, J. H.; O'Brien, A. J.; Quintero, J. C.; Schleif, W. A.; Freunds, K. F.; Gaul, S. L.; Saari, W. S.; Wai, J. S.; Hoffman, J. M.; Anderson, P. S.; Hupe, D. J.; Emini, E. A.; Stern, A. M. Pyridione derivatines: Specific human immunodeficiency virus type 1 reverse transcriptase inhibitors with antiviral activity. *Proc. Natl. Acad. Sci.* **1991**, 88, 6863–6867.
45. Dueweke, T. J.; Poppe, S. M.; Romero, D. L.; Swaney, S. M.; So, A. G.; Downey, K. M.; Althaus, I. W.; Reusser, F.; Busso, M.; Resnick, L.; Mayers, D. L.; Lane, J.; Aristoff, P. A.; Thomas, R. C.; Tarpley, W. G. U-90152, a Potent Inhibitor of Human Immunodeficiency Virus Type 1 Replication. *Antimic. Angents. Chemoth.* **1993**, 37, 1127–1131.
46. Pauwels, R.; Andries, K.; Debyser, Z.; Van Daele, P.; Schols, D.; Stoffels, P.; De Vreese, K.; Woestenborghs, R.; Vandamme, A. M.; Janssen, C. G.; Anne, J.; Cauwenberghs, G.; Desmyter, J.; Heykants, J.; Janssen, M. A.; De Clercq, E.; Janssen, P. A. Potent and highly selective human immunodeficiency virus type 1 (HIV-1) inhibition by a series of a a-anilinophenylacetamide derivatives targeted at HIV-1 reverse transcriptase. *Proc. Natl. Acad. Sci.* **1993**, 90, 1711–1715.
47. Ahgren, C.; Backro, K.; Bell, F. W.; Cantrell, A. S.; Clemens, M.; Colacino, J. M.; Deeter, J. B.; Engelhardt, J. A.; Hogberg, M.; Jaskunas, S. R.; Johansson, N. G.; Jordan, C. L.; Kasher, J. S.; Kinnick, M. D.; Lind, P.; Lopez, C.; Morin, Jr. J. M.; Muesing, M. A.; Noreen, R.; Oberg, B.; Paget, C. J.; Palkowitz, J. A.; Parrish, C. A.; Pranc, P.; Rippy, M. K.; Rydergard, C.; Sahlberd, C.; Swanson, S.; Ternansky, R. J.; Unge, T.; Vasileff, R. T.; Vrang, L.; West, S. J.; Zhang, H.; Zhou, X. X. The PETT Series, a New Class of Potent Nonnucleoside Inhibitors of Human Immunodeficiency Virus Type 1 Reverse Transcriptase. *Antimicrob. Agents. Chemother.* **1995**, 39, 1329–35.
48. Buckheit Jr., R. W.; Snow, M. K.; Fliakas-Boltz, V.; Kinjerski, T. L.; Russell, J. D.; Pallansch, L. A.; Brouwer, W. G.; Yang, S. S. Highly Potent Oxathiin Carboxanilide Derivatives with Efficacy against Nonnucleoside Reverse Transcriptase Inhibitor-Resistant Human Immunodeficiency Virus Isolates. *Antimicrob. Agents. Chemother.* **1997**, 41, 831–837.
49. Morningstar, M. L.; Roth, T.; Farnsworth, D. W.; Smith, M. K.; Watson, K.; Buckheit, R. W.; Das, K.; Zhang, W.; Arnold, E.; Julias, J. G.; Hughes, S. H.; Michejda, C. J. Synthesis, Biological Activity, and Crystal Structure of Potent Nonnucleoside Inhibitors of HIV-1 Reverse Transcriptase That Retain Activity against Mutant Forms of the Enzyme. *J. Med. Chem.* **2007**, 50, 4003–4015.
50. Famiglini, V.; Coluccia, A.; Brancale, A.; Pelliccia, S.; La Regina, G.; Silvestri, R. Arylsulfone-based HIV-I non-nucleoside reverse transcriptase inhibitors. *Future. Med. Chem.* **2013**, 5, 2141–2156.

51. De Clercq, E. Dancing with chemical formulae of antivirals: a personal account. *Biochem. Pharmacol.* **2013**, 86(6), 711–725.
52. Li, D.; Zhan, P.; Liu, H.; Pannecouque, C.; Balzarini, J.; De Clercq, E. Synthesis and biological evaluation of pyrazine derivatives as novel HIV-1 NNRTIs. *Bioorg. Med. Chem.* **2013**, 21, 2128–2134.
53. La Regina, G.; Coluccia, A.; Brancale, A.; Piscitelli, F.; Gatti, V.; Maga, G.; Samuele, A.; Pannecouque, C.; Schols, D.; Balzarini, J.; Novellino, E.; Silvesteri, R. Indolylarylsulfones as HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors: New Cyclic Substituents at Indole-2-carboxamide. *J. Med. Chem.* **2011**, 54, 1587–1598.
54. La Regina, G.; Coluccia, A.; Brancale, A.; Piscitelli, F.; Famiglini, V.; Cosconati, S.; Maga, G.; Samuele, A.; Gonzales, E.; Clotet, B.; Schols, D.; Esté, J. A.; Novellino, E.; Silvesteri, R. New nitrogen containing substituents at the indole-2-carboxamide yield high potent and broad spectrum indolylarylsulfone HIV-1 non-nucleoside reverse transcriptase inhibitors. *J. Med. Chem.* **2012**, 55, 6634–6638.
55. Rotili, D.; Samuele, A.; Tarantino, D.; Ragno, R.; Musmuca, I.; Ballante, F.; Botta, G.; Morera, L.; Pierini, M.; Cirilli, R.; Nawrozkij, M. B.; Gonzalez, E.; Clotet, B.; Artico, M.; Esté, J. A.; Maga, G.; Mai, A. 2-(Alkyl/aryl)amino-6-benzylpyrimidin-4(3H)-ones as inhibitors of wild-type and mutant HIV-1: enantioselectivity studies. *J. Med. Chem.* **2012**, 55, 3558–3562.
56. Rawal, R. K.; Tripathi, R.; Katti, S. B.; Pannecouque, C.; De Clercq, E. Design and synthesis of 2-(2,6-dibromophenyl)-3-heteroaryl-1,3-thiazolidin-4-ones as anti-HIV agents. *Eur. J. Med. Chem*. **2008**, 43, 28002–2806.
57. Menicagli, R.; Samaritani, S.; Signore, G.; Vaglini, F.; Via, L. D. In Vitro Cytotoxic Activities of 2-Alkyl-4,6-diheteroalkyl-1,3,5-triazines: New Molecules in Anticancer Research. *J. Med. Chem.* **2004**, 47, 4649–4652.
58. Henke, B. R.; Consler, T. G.; Go, N.; Hale, L. A.; Hohman, D. R.; Jones, S. A.; Lu, A. T.; Moore, L. B.; Moore, J. T.; Orband-Miller, L. A.; Robinett, R. G.; Shearin, J.; Spearing, P. K.; Stewart, E. L.; Turnbull, P. S.; Weaver, S. L.; Williams, G.; Wisely, B.; Hambert, M. H. A New Series of Estrogen Receptor Modulators That Display Selectivity for Estrogen Receptor β. *J. Med. Chem.* **2002**, 45, 5492–5505.
59. Pandey, V. K.; Tusi, S.; Tusi, Z.; Joshi, M.; Bajpai, S. Synthesis and biological activity 2,4,6-s-triazines. *Acta. Pharm.* **2004**, 54, 1–12.
60. Srinivas, K.; Srinivas, U.; Jayathirtha Rao, V.; Bhanuprakasj, K.; Hara Kishore, K.; Murty, U. S. N. Synthesis and antibacterial activity of 2,4,6-tri substituted *s*-triazines. *Bioorg. Med. Chem. Lett.* **2005**, 15, 1121–1123.
61. Lübbers, T.; Angehrn, P.; Gmünder, H.; Herzig, S.; Kulhanek, J. Design, Synthesis, and Structure-Activity Relationship Studies of ATP Analogues as DNA Gyrase Inhibitors. *Bioorg. Med. Chem. Lett.* **2000**, 10, 821–826.
62. Ghaib, A.; Ménager, S.; Vérité, P.; Lafont, O. Synthesis of variously 9,9-dialkylated octahydropyrimido [3,4-α]-*s*-triazines with potential antifungal activity. *Il. Farmaco*. **2002**, 57, 109–116.
63. Lebreton, S.; Newcombe, N.; Bradley, M. Antibacterial single-bead screening. *Tetrahedron.* **2003**, 59, 10213–10222.
64. McKay, G.; Reddy, R.; Arhin, F.; Belley, A.; Lehoux, D.; Moeck, G.; Sarmiento, I.; Parr, T. R.; Gros, P.; Pelletier, J.; Far, A. R. Triaminotriazine DNA helicase inhibitors with antibacterial activity.*Bioorg. Med. Chem. Lett.* **2006**, 16, 1286–1290.

65. Koc, Z. E.; Bingol, H.; Saf, A. O.; Torlak, E.; Coskun, A. Synthesis of novel tripodal-benzimidazole from 2,4,6-tris[p-formylphenoxy)-1,3,5-triazine: Structural, electrochemical and antimicrobial studies. *J. Hazard. Mater.* **2010**, 183, 251–255.
66. Kumar. A.; Menon, S. K. Fullerene derivatized s-triazine analogues as anti-microbial agents. *Eur. J. Med. Chem.* **2009**, 44, 2178–2183.
67. Chang, C. F.; Huang, C. Y.; Huang, Y. C.; Lin, K. Y.; Lee, Y. J.; Wang, C. J. Total synthesis of (±)-Armepavines and (±)-Nuciferines From (2-Nitroethenyl)benzene Derivatives. *Synthetic. Commun.* **2010**, 40, 3452–3466.
68. Blotny, G. Recent applications of 2,4,6-trichloro-1,3,5-triazine and its derivatives in organic synthesis. *Tetrahedron.* **2006**, 62, 9507–9522, doi: 10.1016/j.tet.2006.07.039.
69. Thurston, J. T.; Dudley, J. R.; Kaiser, D. W.; Hechenbleikner, I.; Schaefer, F. C.; Holm-Hansen, D. Cyanuric Chloride Derivates. I. Aminochloro-*s*-triazines. *J. Am. Chem. Soc.* **1951**, 73, 2981–2983.
70. Kaiser, D. W.; Thurston, J. T.; Dudley, J. R.; Schaefer, F. C.; Hechenbleikner, I.; Holm-Hansen, D. Cyanyric Chloride Derivates. II. Substituted Melamines. *J. Am. Chem. Soc.* **1951**, 73, 2984–2986.
71. Dudley, J. R.; Thurston, J. T.; Schaefer, F. C.; Holm-Hansen, D.; Clarence, J.; Adams, P. Cyanuric Chloride Derivates. III. Alkoxy-*s*-triazines. *J. Am. Chem. Soc.* **1951**, 73, 2986–2990.
72. Alfonso, C. A.; Lourenço N. M. T.; Rosatella, A. A. Synthesis of 2,4,6-Tri-substituted-1,3,5-Triazines. *Molecules.* **2006**, 11, 81–102.
73. Kain, K. C.; Harrington, M. A.; Tennyson, S.; Keystone, J. S. Imported malaria: prospective analysis of problems in diagnosis and management. *Clin. Infect. Dis.* **1998**, 27, 142–149.
74. Gonçalves, D.; Hunziker, P. Transmission-blocking strategies: the roadmap from laboratory bench to community. *Malar. J.* **2016**, 15:95, doi: 10.1186/s12936-016-1163-3.
75. Landier, J.; Parker, D. M.; Thu, A. M.; Carrara, V. I.; Lwin, K. M.; Bonnington, C. A.; Pukrittayakamee, S.; Delmas, G.; Nosten, F. H. The role of early detection and treatment in malaria elimination. *Malar. J.* **2016**, 15:363, doi: 10.1186/s12936-016-1399-y.
76. Winzeler, E. A. Malaria research in the post-genomic era. *Nature.* **2008**, 9, 751–56.
77. Gazzinelli, R. T.; Kalantari, P.; Fitzgerald, K. A.; Golenbock, D. T. Innate sensing of malaria parasites. *Nature.* **2014**, 14, 744–757.
78. Cohen, S. N.; Phifer, K. O.; Yielding, K. L. Complex formation between chloroquine and ferrihaemic acid *in vitro*, and its effect on antimalarial action of chloroquine. *Nature.* **1964**, 202, 805–806.
79. Egan, T. J. Physico-chemical aspects of hemozoin (malaria pigment) structure and formation. *J. Inorg. Biochem.* **2002**, 91, 19–26.
80. Egan, T. J. Recent advances in understanding the mechanism of hemozoin (malaria pigment) formation. *J. Inorg. Biochem.* **2008**, 102, 1288–1299.
81. De Villiers, K. A.; Egan, T. J. Recent advances in the discovery of haem-targeting drugs for malaria and schistosomiasis. *Molecules*. **2009**, 14, 2868–2887.
82. Belorgey, D.; Lanfranchi, D. A.; Davioud-Charvet, E. 1,4-Naphthoquinones and Others NADPH-Dependent Glutathione Reductase-Catalyzed Redox Cyclers as Antimalarial Agents. *Curr. Pharm. Des.* **2013**, 19, 2512–2528.
83. Hyde. J. E. Drug-resistant malaria – an insight. *FEBS. J.* **2007**, 274, 4688–4698.

84. White, N. J.; Ashley, E. A.; Recht, J.; Delves, M.; Ruecker, A.; Smithuis, F. M.; Eziefula, A. C.; Bousema, T.; Drakeley, C.; Cjotivanich, K.; Imwong, M.; Pukrittayakamee, S.; Prachumsri, J.; Chu, C.; Andolina, C.; Bancone, G.; Hien, T. T.; Mayxay, M.; Taylor, W. R. J.; von Seidlein, L.; Price, R. N.; Barnes, K.; Djimdé, A.; ter Kuile, F.; Gosling, R.; Chen, I.; Dhorda, M. J.; Stepniewska, K.; Gieerin, P.; Woodrow, C. J.; Dondorp, A.; Day, N. P. J.; Nosten, F. H. Assessment of the therapeutic responses to gametocytocidal drugs in *Plasmodium falciparum* malaria. *Malar. J.* **2014**, 13:483.
85. Gandhy, S. U.; Kim, K.; Larsen, L.; Rosengren, R. J.; Safe, S. Curcumin and synthetic analogs induce reactive oxygen species and decreases specifity protein (Sp) transcription factors by targeting microRNAs. *BMC. Cancer.* **2012**, 12, 564–75.
86. Norrato, G. D.; Jutooru, I.; Safe, S.; Angel-Morales, G.; Mertens-Talcott, S. U. The drug resistance suppression induced by curcuminoids in colon cancer SW-480 cell is mediated by reactive oxygen species-induced distruption of the microRNA-27a-ZBTB10-Sp axis. *Mol. Nutr. Food. Res.* **2013**, 57, 1638–48.
87. Nandakumar, D. N.; Nagaraj, V. A.; Vathsala, P. G.; Rangarajan, P.; Padmanaban, G. Curcumin-Artemisinin Combination Therapy for Malaria. *Antimicrob. Agents. Chemother.* **2006**, 50, 1859–60.
88. Dinkova-Kostova, A. T.; Massiah, M. A.; Bozak, R. E.; Hicks, R. J.; Talalay, P. Potency of Micheal reaction acceptors as inducers of enzymes that protect against carciogenesis depends on their reactivity with sulhydryl groups. *Proc. Natl. Acad. Sci. USA.* **2001**, 98, 3404–3409.
89. Weber, W. M.; Hunsaker, L. A.; Abcouwer, S. F.; Deck, L. M.; vander Jagt, D. L. Anti-oxidant activities of curcumin nd related enones. *Bioorg. Med. Chem.* **2005**, 13, 3811–3820.
90. Carroll, R. E.; Benya, R. V.; Turgeon, D. K.; Vareed, S.; Neumann, M.; Rodriguez, L.; Kakarala, M.; Carpenter, P. M.; McLaren, C.; Meyskens, F. L.; et al. Phase IIa Clinical trial of curcumin for the prevention of colorectal neoplasia. *Cancer. Prev. Res.* **2011**, 4, 354–364.
91. Wang, Y.; Xiao, J.; Zhou, H.; Yang, S.; Wu, X.; Jiang, C.; Zhao, Y.; Liang, D.; Li, X.; Liang, G. A novel monocharbonyl analogue of curcumin, (1E,4E)-1,5-Bis(2,3-dimethoxyphenyl)penta-1,4-dien-3-one, induced cancer cell H460 apoptosis via activation of endoplasmic reticulum stress signalling pathway. *J. Med. Chem.* **2011**, 54, 3768–3778.
92. Munigunti, R.; Gathiaka, S.; Acevedo, O.; Sahu, R.; Tekwani, B.; Calderón, A. I. Determination of antiplasmodial activity and binding affinity of curcumin and demethoxycurcumin towards PfTrxR. *Nat. Prod. Res.* **2014**, 28, 359–364.
93. Reddy, R. C.; Vatsala P. G.; Keshamouni, V. G.; Padmanab, G.; Rangarajan, P. N. Curcumin for malaaria therapy. *Biochem. Biophys. Res. Commun.* **2005**, 326, 472–474.
94. Mishra, S.; Karmodiya, K.; Surolia, A. Synthesis and exploration of novel curcumin analogues as anti-malarial agents. *Biochem. Med. Chem.* **2008**, 16, 2894–2902.
95. Shukla, A.; Singh, A.; Pathak, L. P.; Shrivastava, N.; Tripathi, P. K.; Singh, M. P.; Singh, K. Inhibition of *P. falciparum* PEATP6 by curcumin and its derivatives: A bioinformatics study. *Cell. Mol. Biol.* **2012**, 58, 182–186.
96. Manohar, S.; Khan, S. I.; Kandi, S. K.; Raj, K.; Sun, G.; Yang, X.; Calderon Molina, A. D.; Ni, N.; Wang, B.; Rawat, D. S. Synthesis, antimalarial activity and

cytotoxic potential of new monocarbonyl analogues of curcumin. *Bioorg. Med. Chem. Lett.* **2013**, 23, 112–116.
97. Masuda, T.; Jitoe, A. T.; Isobe, J.; Nakatani, N.; Yonemori, S. Anti-oxidative and anti-inflammatory curcumin-related phenolics from rhizomes of *Curcuma domestica. Phytochem.* **1993**, 32, 1557–1560.
98. Grendron, T.; Davioud-Charvet, E.; Thomas, J. J. M. Versatile Synthesis of Dissymetric Diarylideneacetones via a Palladium-Catalyzed Coupling-Isomeration Reaction. *Synthesis*. **2012**, 44, 3829–3835, *Erratum: Synthesis,* **2013***, 45, 1270.*
99. Gendron, T.; Kessedjian, H.; Davioud-Charvet, E.; Lanfranchi, D. A. Diastereo-selective synthesis of 2,6-diaryltetrahydrothiopyran-4-ones by phase-transfer catalysis. *Eur. J. Org. Chem.* **2015**, 8, 1790–1796.
100. Davioud-Charvet, E.; Wenzel, I. N.; Müller, T. J. J.; Hanquet, G.; Lanfranchi, D. A.; Leroux, F.; Gendron, T. Dibenzylidene- and heteroarylideneacetone derivatives as kinetoplastideae parasiticides and their preparation, pharmaceutical compositions and use in the treatment of tryptanosomiasis and leishmaniasis. *PTC Int. Appl.* WO 2011033115 A2, 24 March **2011**.
101. Wenzel, N. Synthesis and Mechanism of Antiparasitic Mannich Base Derivatives Affecting the Redox Equilibrium of Trypanosomes and Malaria Parasites. *Ph.D. Thesis,* Heidelberg University, Heidelberg Germany, 21 September **2009**.
102. Gendron, T. Synthesis and Evaluation of the Antiparasitic Activity of Diarylideneacetones and Their Related Thiopyranone and S-Oxide Prodrugs. *Ph.D. Thesis*, Strasbourg University, Strasbourg, 23 November **2012**.
103. World Health Organization Home Page. http://www.who.int/en (accessed Nov 31, 2016).
104. Alemu, A.; Shiferaw, Y.; Addis, Z.; Mathewos, B.; Birhan, W. Effect of malaria on HIV/AIDS transmission and progression. *Parasite. Vector.* **2013**, 6, 18–25.
105. Steketee, R. W.; Wirima, J. J.; Bloland, P. B.; Chilima, B.; Hermin, J. H.; Chitsulo, L.; Breman, J. G. Impairment of a pregnant woman's acquired ability to limit *Plamsodium falciparum* by infection with human immunodeficienct virus type-1. *Am. J. Trop. Med. Hyg.* **1996**, 55, 42–49.
106. Steketee, R. W.; Wirima, J. J.; Slutsker, L.; Breman, J. G.; Heymann, D. L. Comparability of treatment groups and risk factors for parasitemia at the first antenatal clinic visit in a study of malaria treatment and prevention in pregnancy in rural Malawi. *Am. J. Trop. Med. Hyg.* **1996**, 55, 17–23.
107. French, N.; Nakiyingi, J.; Lugada, E.; Watera, C.; Whitworth, J. A.; Gilks, C. F. Increasing rates of malarial fever with deteriorating immune status in HIV-1-infected Uganda adults. *AIDS*. **2001**, 15, 899–906.
108. Withworth, J.; Morgan, D.; Quigley, M.; Smith, A.; Mayanja, B.; Eotu, H.; Omoding, N.; Okongo, M.; Malamba, S.; Ojwiya, A. Effect of HIV-1 and increasing immunosuppression on malaria parasitaemia and clinical episodes in adults in rural Uganda: a cohort study. *Lancet.* **2000**, 356, 1051–1056.
109. Worku, S.; Bjorkman, A.; Troye-Blomberg, M. Lymphocyte activation and subset redistribution in the peripheral blood in acute malaria illness: distinct T cell patterns in plasmodium falciparum and P. vivax infections. *Clin. Exp. Immunol.* **1997**, 108, 34–41.
110. Hoffmann, I. F.; Jere, C. S.; Taylor, T. E.; Munthali, P.; Dyer, J. R.; Wirima, J. J.; Rogerson, S. J.; Kumwenda, N.; Eron, J. J.; Fiscus, S. A.; Chakraborty, H.; Taha, T. E.; Cohen, M. S.; Molyneux, M. E. The effect of Plasmodium falciparum malaria on HIV-1 RNA blood plasma concentration. *AIDS*. **1999**, 13, 487–494.

111. Kublin, J. G.; Patnaik, P.; Jere, C. S.; Miller, W. C.; Hoffman, I. F.; Chimiya, N.; Pendame, R.; Taylor, T. E.; Molyneux, M. E. Effect of Plasmodium falciparum malaria on concentration of HIV-1-RNA in the blood of adults in rural Malawi: a prospective cohort study. *Lancet*. **2005**, 365, 233–240.
112. National Institute of Standards and Technology, NIST Chemistry WebBook. http://webbook.nist.gov/chemistry (accessed Nov 31, 2016).
113. National Cancer Institute. http:cactus.nci.nih.gov/ncidb2/download.html (accessed Nov 31, 2016).
114. National Library of Medicine, ChemIDPlus Advanced. http://chem.sis.nlm.nih.gov/chemidplus (accessed Nov 31, 2016).
115. ChemBank Home Page. http://chembank.broadinstitute.org (accessed Nov 31, 2016).
116. PubChem Home Page. http://pubchem.ncbi.nlm.nih.gov (accessed Nov 31, 2016).
117. ChemSpider Home Page. http://chemspider.com (accessed Nov 31, 2016).
118. ChEMBL Home Page. http://ebi.ac.uk (accessed Nov 31, 2016).
119. DrugBank Home Page. http://drugbank.ca (accessed Nov 31, 2016).
120. Cambridge Soft, ChemFinder 10.0 ed. http//www.cambridgesoft.com (accessed Oct 31, 2016).
121. Syracuse Research Corporation, EPI Suite 3.12 ed. http//www.syrres.com (accessed Oct 31, 2016).
122. SciFinder Home Page. https://scifinder.cas.org (accessed Nov 31, 2016).
123. Yang, S. S.; Lu, J. Y.; Kong, X. Q.; Liang, Z. J.; Lou, C.; Jiang, H. Computational drug discovery. *Acta. Pharmacologica. Sinica.* **2012**, 33, 1131–1140.
124. Young, D.; Martin, Todd.; Venkatapathy, R.; Harten, P. Are the Chemical Structures in Your QSAR Correct? *QSAR. Comb. Sci.* **2008**, 28, 1337–1345.
125. Williams, A.J.; Erkins, S. A quality alert and call for improved curation of public chemistry databases. *Drug. Discov. Today.* **2011**, 16, 747–750.
126. Ruusmann, V.; Maran, U. From data point timelines to a well curated datset, data mining of experimental data and chemical structure data from scientific articles, problems and possible solutions. *J. Comput. Aided. Mol. Des.* **2013**, 27, 583–603.
127. Fourches, D.; Muratov, E.; Tripsha, A. Trust, But Verify: On the Importance of Chemical Structure Curation in Chemoinformatics and QSAR Modeling Research. *J. Chem. Inf. Model.* **2010**, 50, 1189–1204.
128. Fourches, D.; Muratov, E.; Tropsha, A. Trust, but Verify II: A Practical Guide to Chemogenomics Data Curation. *J. Chem. Inf. Model*. **2016**, 56, 1243–1252.
129. Schuffenhauer, A.; Ertl, P.; Roggo, S.; Wetzel, S.; Koch, M. A.; Haldmann, H. The scaffold tree – visualization of the scaffold universe by hierarchical scaffold classification. *J. Chem. Inf. Model*. **2007**, 47, 47–58.
130. Kibbey, C.; Clavet, A. Molecular Property eXplorer: a novel approach to visualizing SAR using tree-maps and heatmaps, *J. Chem. Model*. **2005**, 45, 523–32.
131. Agrafirotis, D. K.; Bandyopadhyay, D.; Farnum, M. Radial clustergrams: visualizing the aggregate properties of hierarchical clusters, *J. Chem. Inf. Model.* **2007**, 48, 69–75.
132. Wilkens, S. J.; Janes, J.; Su, A. I. HierS: hierarchical scaffold clustering using topological chemical graphs, *J. Med. Chem*. **2005**, 48, 3182–3193.
133. Roberts, G.; Myatt, G. J.; Johnson, W. P; Cross, K. P.; Jr. Blower, P. E. LeadScope: software for exploring large sets of screening data, *J. Chem. Inf. Comput. Sci*. **2000**, 40, 1302–1314.

134. Strobelt, H.; Bertini, E.; Braun, J.; Deussen, O.; Groth, U.; Mayer, T. U.; Merfhor, D. HiTSEE KNIME: a visualization tool for hit selection and analysis in high-throughput screening experiments for the KNIME platform, *BMC. Bioinformatics*. **2012**, 13, doi: 10.1186/1471-2015-13-S8-S4.
135. Gütlein, M.; Karwath, A.; Kramer, S. CheS-Mapper – Chemical Space Mapping and Visualization in 3D, *J. Cheminform*. **2012**, 4(1):7, doi: 10.1186/1758-2946-4-7.
136. Givehci, A.; Dietrich, A.; Wrede, P.; Schneider, G. ChemSpaceShuttle: A tool for data mining in drug discovery by classification, projection, and 3D visualization, *QSAR Comb. Sci.* **2003**, 22 549–559.
137. Le Guilloux, V.; Arrault, A.; Colliandre, L.; Bourg, S.; Vayer, P.; Morin-Allory, L. Mining collections of compounds with Screening Assistant 2. *J. Chemoinf.* **2012**, 4:20, doi: 10.1186/1758-2946-4-20.
138. Lounkine, E.; Wawer, M.; Wassermann, A. M.; Bajorath, J. SARANEA: a freely available program to mine structure-activity and structure-selectivity relationship information in compound data sets, *J. Chem. Inf. Model.* **2010**, 50, 68–78.
139. Eriksson, M.; Nilsson, I.; Kogej, T.; Southan, C.; Johansson, M.; Tyrchan, C.; Muresan, S.; Blomberg, N.; Bjäreland, M. SARConnect: A Tool to Interrogate the Connectivity Between Proteins, Chemical Structures and Activity Data. *Mol. Inform.* **2012**, 31, 555–568.
140. Xiong, B.; Liu, K.; Wu, J.; Burk, D. L.; Jiang, H.; Shen, J. DrugViz: a Cytoscape plugin for visualizing and analyzing small molecule drugs in biological networks. *Bioinformatics*. **2008**, 24, 2117–2118.
141. Scaffold Hunter Home Page. http://scaffoldhunter.sourceforge.net (accessed Oct 19, 2016).
142. Molinspiration Products and Services Home Page. www.molinspiration.com (accessed Aug 25, 2016)
143. Wetzel, S.; Klein, K.; Renner, S.; Rauh, D.; Opera, T. I.; Mutzel, P.; Haldmann, H. Interactive exploration of chemical space with Scaffold Hunter. *Nat. Chem.* **2009**, 5, 581–583.
144. Bon, R. S.; Waldmann, H. Bioactivity-Guided Navigation of Chemical Space. *Acc. Chem. Res*. **2010**, 43, 1103–1114.
145. Dearden, J. C.; Cronin, M. T.; Kaiser, K. L. How not to develop a quantitative structure-activity or structure-property relationship (QSAR/QSPR). *SAR. QSAR. Environ. Res.* **2009**, 20, 241–266.
146. Johnson, S. R. The trouble with QSAR (or how I learned to stop worrying and embrace fallacy). *J. Chem. Inf. Model.* **2008**, 48, 25–26.
147. Maran, U.; Sild, S.; Massatorta, P.; Casalegno, M.; Benfenati, E.; Romberg, M. Grid computing for the estimation of toxicity: acute toxicity on fathead minnow (Pimephales promelas), In Distributed, High-Performance and Grid Computing in Computational Biology, Dubitzky, W.; Schuster, A.; Sloot, P.M.A.; Schroeder, M.; Eds.; Springer-Verlag Berlin Heidelberg, **2007**, 4360, 60–74.
148. Karelson, M. Molecular descriptors in QSAR/QSPR. Wiley- Interscience publication, **2000**.
149. Todeschini, R.; Consonni, V.; Eds. Handbook of Molecular Descriptors, Wiley-IbraceVCHrbrace Verlang IbraceGmbHrbrace, **2000**.
150. Todeschini, R.; Consonni, V.; Mannhold, R.; Kubinyi, H.; Folkers, G. Molecular Descriptors for Chemoinformatics, Methods and Principles in Medicinal Chemistry, *Wiley*. **2009**.
151. Livingstone, D. A practical guide to scientific data analysis, Wiley. **2009**.

152. Leach, A. R.; Gillet, V. J. An introduction to chemoinformatics. Springer. **2007.**
153. Hawkins, D. M.; Basak, S. C.; Mills, D. Assessing model fit by cross-validation. *J. Chem. Inf. Comput. Sci.* **2003**, 43, 579–586.
154. Atkinson, A. C. Plots, Transformation, Regression. Clarendon Press. Oxford, UK, 1991; pp 282.
155. Stewart, J. J. MOPAC: a Semiepirical Moleclar Orbital Program. *J. Comput.–Aided. Mol. Des*. **1990**, 4, 1–45.
156. Baker, J. An algorithm for the location of transition states. *J. Comput. Chem*. **1986**, 7, 385–395.
157. Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. P. AM1:a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc*. **1985**, 107, 3902–3909.
158. Draper, N. R.; Smith, H. Applied regression analysis, Wiley, New York, **1966**.
159. Katritzky, A. R.; Lobanov, V. S.; Karelson, M. Codessa reference manual 2.0, Gainesville, FL, **1994**.
160. Solov'ev, P.; Varnek, A.; Wipff, G. Modeling of Ion Complexation and Extraction Using Substructural Molecular Fragments. *J. Chem. Inf. Comput. Sci*. **2000**, 40, 847–858.
161. Solov'ev, V. P.; Varnek, A.; Wipff, G. Modeling of Ion Complexation and Extraction Using Substructural Molecular Fragments. *J. Chem. Inf. Comput. Sci.* **2000**, 40, 847–858.
162. Sidorov, P.; Gaspar, G.; Marcou, G.; Varnek, A.; Horvath, D. Mappability of drug-like space: Towards a polypharmacologically competent map of drug-relevant compounds. *J. Comput. Aided. Mol. Des*. **2015**, 29, 1087–1108.
163. Ruggiu, F.; Marcou, G.; Varnek, A.; Horvath, D. ISIDA Property-Labelled Fragment Descriptors. *Mol. Inf*. **2010**, 29, 855–868.
164. Horvath, D.; Brown, J.B.; Marcou, G.; Varnek, A. An Evolutionary Optimiser of libsvm Models. *Challenges* **2014**, *5*, 450–472.
165. Chang, C. C.; Lin, C. J. LIBSVM: A library for support vector machines. *J. ACM. TIST*. **2001**, 2, doi: 10.1145/1961189.1961199.
166. Instant JChem Home Page. www.chemaxon.com/products/instant-jchem/ (accessed Oct 19, 2016).
167. ChemAxon Home Page. http://www.chemaxon.com/products/standardizer (accessed May 15, 2016).
168. García-Sosa, A. T.; Sild, S.; Takkis, K.; Maran, U. Combined Approach Using Ligand Efficiency, Cross-Docking, and Antitarget Hits for Wild-Type and Drug-Resistant Y181C HIV-1 Reverse Transcriptase. *J. Chem. Inf. Model*. **2011**, 51, 2595–2611.
169. Campiani, G.; Morelli, E.; Fabbrini, M.; Nacci, V.; Greco, G.; Novellino, E.; Ramuno, A.; Maga, G.; Spadari, S.; Caliendo, G.; Bergamini, A.; Faggioli, E.; Uccella, I.; Bolacchi, F.; Marini, S.; Coletta, M.; Nacca, A.; Caccia, S. Pyrrolobenzoxazepinone Derivatives as Non-Nucleoside HIV-1 RT Inhibitors: Further Structure−Activity Relationship Studies and Identification of More Potent Broad-Spectrum HIV-1 RT Inhibitors with Antiviral Activity. *J. Med. Chem*. **1999**, 42, 4462–4470.
170. Campiani, G.; Aiello, F.; Fabbrini, M.; Morelli, E.; Ramunno, A.; Armaroli, S.; Nacci, V.; Garofalo, A.; Greco, G.; Novellino, E.; Maga, G.; Spadari, S.; Bergamini, A.; Ventura, L.; Bongiovanni, B.; Capozzi, M.; Bolacchi, F.; Marini, S.; Coletta, M.; Guiso, G.; Caccia, S. Quinoxalinylethylpyridylthioureas (QXPTs) as Potent Non-

Nucleoside HIV-1 Reverse Transcriptase (RT) Inhibitors. Further SAR Studies and Identification of a Novel Orally Bioavailable Hydrazine-Based Antiviral Agent. *J. Med. Chem*. **2001**, 44, 305–315.

171. Romero, D. L.; Olmsted, R. A.; Poel, T. J.; Morge, R. A.; Biles, C.; Keiser, B. J.; Kopta, L. A.; Friis, J. M.; Hosley, J. D.; Stefanski, K. J.; Wishka, D. G.; Evans, D. B.; Morris, J.; Stehle, R. G.; Sharma, S. K.; Yagi, Y.; Voorman, R. L.; Adams, W. J.; Tarpley, W. G.; Thomas, R. C. Targeting Delavirdine/Atevirdine Resistant HIV-1: Identification of (Alkylamino)piperidine-Containing Bis(heteroaryl)-piperazines as Broad Spectrum HIV-1 Reverse Transcriptase Inhibitors. *J. Med. Chem*. **1996**, 39, 3769–3789.
172. Maran, U.; Sild, S.; Tulp, I.; Takkis, K.; Moosus, M. Molecular Descriptors from Two-Dimensional Chemical Structure, in: M.T.D. Cronin. J.C. Madden (Eds.), In Silico Toxicology: Principles and Applications, Royal Society of Chemistry, **2010**, pp. 148–192.
173. Ren, J.; Esnouf, R.; Garman, E.; Somers, D.; Ross, C.; Kirby, I.; Keeling, J.; Darby, G.; Jones, Y.; Stuart, D.; Stammers, D. High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nature*. **1995**, 2, 293–302.

APPENDIX 1

Property table for in-house and ChEMBL data	
Column label	**Information type**
LabID	Origin of the compound: in the case of in-house data it was identified by person who synthesized the compound in case of ChEMBL data, the ChEMBL ID without letters ”CHEMBL” was assigned
Value – log(IC50[M])	Measured numerical value
STD DEV	Standard deviation value
relationship	Relationship of the value, if it is bigger/smaller/equal to given value (>; <; =)
Protocol ID	Protocol used to measure the experimental value
Ref ID	The source ID, from where the value was taken

Protocol table for in-house and ChEMBL data	
Column label	**Information type**
Protocol ID	Protocol used to measure the given/available value
Protocol name	Type of measurement
Property name	Type of property measured
Target name	The target mechanism
Target organism	Parasite strain used
Ref ligand	Compound used as reference
Units	The unit of measured value
Solvent	Solvent used
Drug exposure time	How long was the drug exposure time
Hematocritage %	Amount of hematocritage used
Parasitema %	Amount of parasitema used
Parasitic stage	Parasitic stage used: synchronized or unsynchronized
Assay principle	Assay used
Comments	For any additional comments

SMILES and LabID table for in-house and ChEMBL data	
Column label	**Information type**
Structure	Compound structure
Mol weight	Molecular weight of the compound
Formula	Compound formula
LabID	Origin of the compound: in the case of in-house data it was identified by person who synthesized the compound incase of ChEMBL data, the ChEMBL ID without letters ”CHEMBL” was assigned

Reference table for in-house and ChEMBL data	
Column label	**Information type**
Ref ID	The source ID, from where the value was taken
Protocol ID	Protocol used to measure the given/available value
Journal title	in which journal it was published
Doc source description	what type of source the data was taken from
Pubmed ID	Pubmed ID if available
DOI	Digital Object Identifier if available
Year	Year of data publication
Volume	Volume of source article
Article title	Title of source article
Abstract	Abstract of source article

ACKNOWLEDGEMENTS

First, I am grateful to my supervisor Dr. Uko Maran for inviting me to join his workgroup and supporting, guiding me during my doctoral studies. I would like to acknowledge colleagues, collaborators and the co-authors I have pleasure to work together with and of course thank my lunch companions for their patience to listen and for creating always positive and fun atmosphere.

I am thankful to Dr. Elisabeth Davioud-Charvet and Prof. Aleksandre Varnek for the opportunity to work in their laboratories during my doctoral studies.

Finally, special thanks to my family and friends for loving, supporting, motivating and being there for me through good and intense times.

PUBLICATIONS

CURRICULUM VITAE

Name:	Birgit Viira
Date of birth:	August 15, 1987
Citizenship:	Estonian
Address:	University of Tartu, Institute of Chemistry Ravila 14a, 50411, Tartu, Estonia
E-mail:	birgit.viira@ut.ee

Education:

2012 to present	University of Tartu, PhD student in Molecular Engineering
2010–2012	University of Tartu, MSc in chemistry
2007–2010	University of Tartu, BSc in chemistry

Professional career:

2016 to present	Chemist, Chair of Molecular Technology, Chemicum, University of Tartu, Tartu, Estonia
2011 to 2012	Chemist, TBD Biodiscovery Inc.

Stipends:

2016	COST CM1307 STMS stipend by European Cooperation in Science and Technology Short-term scientific mission for 2-month visit at the University of Strasbourg
2015	COST CM1307 STMS stipend by European Cooperation in Science and Technology Short-term scientific mission for 2-month visit at the University of Strasbourg
2014	DoRa Activity 6 mobility stipend by European Social Fund's Doctoral Studies and Internationalisation Programme DoRa for 5-month visit at the University of Strasbourg.
2014	COST CM1307 STMS stipend by European Cooperation in Science and Technology Short-term scientific mission for 3-month visit at the University of Strasbourg

List of publications:

1. **Viira, B.**; Selytina, A.; García-Sosa, A.T.; Karonen, M.; Sinkkonen, J.; Merits, A.; Maran, U. Design, discovery, modelling, synthesis, and biological evaluation of novel and small, low toxicity *s*-triazine derivatives as HIV-1 non-nucleoside reverse transcriptase inhibitors. *Bioorg. Med. Chem.* **2016**, 24(11), 2519–29. DOI: 10.1016/j.bmc.2016.04.018
2. **Viira, B**.; Gendron, T.; Lanfranchi, D.D.; Cojean, S.; Horvath, D.; Marcou, G.; Varnek, A.; Maes, L.; Maran, U.; Loiseau P.M.; Davioud-Charvet, E. In Silico for Antimalarial Structure-Activity Knowledge and Discovery of Novel Antimalarial Curcuminoids. *Molecules.* **2016**, 21(7), 853. DOI:10.3390/molecules21070853.

3. Ivan, T.; Enkvist, E.; **Viira, B.**; Manoharan, G.B.; Raudaru, G.; Pflug, A.; Alam, K.A.; Zaccolo, M.; Engh, R.A.; Uri, A. Bifunctional Ligands for Inhibition of Tight-Binding Protein-Protein Interactions. Bioconjug. Chem., **2016**, 27(8), 1900–10.
4. Sidorov, P; **Viira, B.**; Davioud-Charvet, E.; Maran, U.; Marcou, G.; Horvath, D.; Varnek, A. QSAR modeling and chemical space analysis of antimalarial compounds. *J. Comp. Aided. Mol.* **2017**, 31(5), 441–451. DOI 10.1007/s10822-017-0019-4.
5. **Viira, B.**; García-Sosa, A.T.; Maran, U. Chemical Structure and Correlation Analysis of HIV-1 NNRT and NRT Inhibitors and Database-Curated, Published Inhibition Constants with Chemical Structure in Diverse Datasets. *J. Mol. Graph. Model*. **2017***, In Press.* DOI: 10.1016/j.mgm.2017.06.019.

ELULOOKIRJELDUS

Nimi:	Birgit Viira
Sünniaeg:	15. August, 1987
Kodakondsus:	Eesti
Aadress:	Tartu Ülikool, keemia instituut Ravila 14a, 50411, Tartu, Eesti
E-mail:	birgit.viira@ut.ee

Haridus:

2012– …	Tartu Ülikool, doktoriõpe Molekulaartehnoloogias
2010–2012	Tartu Ülikool, MSc keemias
2007–2010	Tartu Ülikool, BSc keemias

Teenistuskäik:

2016–…	Keemik, Molekulaartehnoloogia õppetool
2011–2012	Keemik, TBD Biodiscovery OÜ

Stipendiumid:

2016	COST CM1307 STMS stipend by European Cooperation in Science and Technology Short-term scientific mission. Kaks kuud Strasbourgi Ülikoolis Elisabeth Davioud-Charvet ja Alexandre Varnek töögrupis.
2015	COST CM1307 STMS stipend by European Cooperation in Science and Technology Short-term scientific mission. Kaks kuud Strasbourgi Ülikoolis Elisabeth Davioud-Charvet ja Alexandre Varnek töögrupis.
2014	Mobiilsusstipendium Euroopa Liidu Euroopa Sotsiaalfond programme DoRa raames. Viis kuud Strasbourgi Ülikoolis Elisabeth Davioud-Charvet ja Alexandre Varnek töögrupis.
2014	COST CM1307 STMS stipend by European Cooperation in Science and Technology Short-term scientific mission. Kolm kuud Strasbourgi Ülikoolis, Dr. Elisabeth Davioud-Charvet ja Prof. Alexandre Varnek töögrupis.

Publikatsioonide loetelu:

1. **Viira, B.**; Selytina, A.; García-Sosa, A.T.; Karonen, M.; Sinkkonen, J.; Merits, A.; Maran, U. Design, discovery, modelling, synthesis, and biological evaluation of novel and small, low toxicity *s*-triazine derivatives as HIV-1 non-nucleoside reverse transcriptase inhibitors. *Bioorg. Med. Chem.* **2016**, 24(11), 2519–29. DOI: 10.1016/j.bmc.2016.04.018
2. **Viira, B**.; Gendron, T.; Lanfranchi, D.D.; Cojean, S.; Horvath, D.; Marcou, G.; Varnek, A.; Maes, L.; Maran, U.; Loiseau P.M.; Davioud-Charvet, E. In Silico for Antimalarial Structure-Activity Knowledge and Discovery of Novel Antimalarial Curcuminoids. *Molecules.* **2016**, 21(7), 853. DOI:10.3390/molecules21070853.

3. Ivan, T.; Enkvist, E.; **Viira, B.**; Manoharan, G.B.; Raudaru, G.; Pflug, A.; Alam, K.A.; Zaccolo, M.; Engh, R.A.; Uri, A. Bifunctional Ligands for Inhibition of Tight-Binding Protein-Protein Interactions. Bioconjug. Chem., **2016**, 27(8), 1900–10.
4. Sidorov, P; **Viira, B.**; Davioud-Charvet, E.; Maran, U.; Marcou, G.; Horvath, D.; Varnek, A. QSAR modeling and chemical space analysis of antimalarial compounds. *J. Comp. Aided. Mol.* **2017**, 31(5), 441–451. DOI 10.1007/s10822-017-0019-4.
5. **Viira, B.**; García-Sosa, A.T.; Maran, U. Chemical Structure and Correlation Analysis of HIV-1 NNRT and NRT Inhibitors and Database-Curated, Published Inhibition Constants with Chemical Structure in Diverse Datasets. *J. Mol. Graph. Model*. **2017**, *In Press.* DOI: 10.1016/j.mgm.2017.06.019.

DISSERTATIONES CHIMICAE UNIVERSITATIS TARTUENSIS

1. **Toomas Tamm.** Quantum-chemical simulation of solvent effects. Tartu, 1993, 110 p.
2. **Peeter Burk.** Theoretical study of gas-phase acid-base equilibria. Tartu, 1994, 96 p.
3. **Victor Lobanov.** Quantitative structure-property relationships in large descriptor spaces. Tartu, 1995, 135 p.
4. **Vahur Mäemets.** The ^{17}O and ^{1}H nuclear magnetic resonance study of H_2O in individual solvents and its charged clusters in aqueous solutions of electrolytes. Tartu, 1997, 140 p.
5. **Andrus Metsala.** Microcanonical rate constant in nonequilibrium distribution of vibrational energy and in restricted intramolecular vibrational energy redistribution on the basis of slater's theory of unimolecular reactions. Tartu, 1997, 150 p.
6. **Uko Maran.** Quantum-mechanical study of potential energy surfaces in different environments. Tartu, 1997, 137 p.
7. **Alar Jänes.** Adsorption of organic compounds on antimony, bismuth and cadmium electrodes. Tartu, 1998, 219 p.
8. **Kaido Tammeveski.** Oxygen electroreduction on thin platinum films and the electrochemical detection of superoxide anion. Tartu, 1998, 139 p.
9. **Ivo Leito.** Studies of Brønsted acid-base equilibria in water and non-aqueous media. Tartu, 1998, 101 p.
10. **Jaan Leis.** Conformational dynamics and equilibria in amides. Tartu, 1998, 131 p.
11. **Toonika Rinken.** The modelling of amperometric biosensors based on oxidoreductases. Tartu, 2000, 108 p.
12. **Dmitri Panov.** Partially solvated Grignard reagents. Tartu, 2000, 64 p.
13. **Kaja Orupõld.** Treatment and analysis of phenolic wastewater with microorganisms. Tartu, 2000, 123 p.
14. **Jüri Ivask.** Ion Chromatographic determination of major anions and cations in polar ice core. Tartu, 2000, 85 p.
15. **Lauri Vares.** Stereoselective Synthesis of Tetrahydrofuran and Tetrahydropyran Derivatives by Use of Asymmetric Horner-Wadsworth-Emmons and Ring Closure Reactions. Tartu, 2000, 184 p.
16. **Martin Lepiku.** Kinetic aspects of dopamine D_2 receptor interactions with specific ligands. Tartu, 2000, 81 p.
17. **Katrin Sak.** Some aspects of ligand specificity of P2Y receptors. Tartu, 2000, 106 p.
18. **Vello Pällin.** The role of solvation in the formation of iotsitch complexes. Tartu, 2001, 95 p.
19. **Katrin Kollist.** Interactions between polycyclic aromatic compounds and humic substances. Tartu, 2001, 93 p.

20. **Ivar Koppel.** Quantum chemical study of acidity of strong and superstrong Brønsted acids. Tartu, 2001, 104 p.
21. **Viljar Pihl.** The study of the substituent and solvent effects on the acidity of OH and CH acids. Tartu, 2001, 132 p.
22. **Natalia Palm.** Specification of the minimum, sufficient and significant set of descriptors for general description of solvent effects. Tartu, 2001, 134 p.
23. **Sulev Sild.** QSPR/QSAR approaches for complex molecular systems. Tartu, 2001, 134 p.
24. **Ruslan Petrukhin.** Industrial applications of the quantitative structure-property relationships. Tartu, 2001, 162 p.
25. **Boris V. Rogovoy.** Synthesis of (benzotriazolyl)carboximidamides and their application in relations with *N*- and *S*-nucleophyles. Tartu, 2002, 84 p.
26. **Koit Herodes.** Solvent effects on UV-vis absorption spectra of some solvatochromic substances in binary solvent mixtures: the preferential solvation model. Tartu, 2002, 102 p.
27. **Anti Perkson.** Synthesis and characterisation of nanostructured carbon. Tartu, 2002, 152 p.
28. **Ivari Kaljurand.** Self-consistent acidity scales of neutral and cationic Brønsted acids in acetonitrile and tetrahydrofuran. Tartu, 2003, 108 p.
29. **Karmen Lust.** Adsorption of anions on bismuth single crystal electrodes. Tartu, 2003, 128 p.
30. **Mare Piirsalu.** Substituent, temperature and solvent effects on the alkaline hydrolysis of substituted phenyl and alkyl esters of benzoic acid. Tartu, 2003, 156 p.
31. **Meeri Sassian.** Reactions of partially solvated Grignard reagents. Tartu, 2003, 78 p.
32. **Tarmo Tamm.** Quantum chemical modelling of polypyrrole. Tartu, 2003. 100 p.
33. **Erik Teinemaa.** The environmental fate of the particulate matter and organic pollutants from an oil shale power plant. Tartu, 2003. 102 p.
34. **Jaana Tammiku-Taul.** Quantum chemical study of the properties of Grignard reagents. Tartu, 2003. 120 p.
35. **Andre Lomaka.** Biomedical applications of predictive computational chemistry. Tartu, 2003. 132 p.
36. **Kostyantyn Kirichenko.** Benzotriazole – Mediated Carbon–Carbon Bond Formation. Tartu, 2003. 132 p.
37. **Gunnar Nurk.** Adsorption kinetics of some organic compounds on bismuth single crystal electrodes. Tartu, 2003, 170 p.
38. **Mati Arulepp.** Electrochemical characteristics of porous carbon materials and electrical double layer capacitors. Tartu, 2003, 196 p.
39. **Dan Cornel Fara.** QSPR modeling of complexation and distribution of organic compounds. Tartu, 2004, 126 p.
40. **Riina Mahlapuu.** Signalling of galanin and amyloid precursor protein through adenylate cyclase. Tartu, 2004, 124 p.

41. **Mihkel Kerikmäe.** Some luminescent materials for dosimetric applications and physical research. Tartu, 2004, 143 p.
42. **Jaanus Kruusma.** Determination of some important trace metal ions in human blood. Tartu, 2004, 115 p.
43. **Urmas Johanson.** Investigations of the electrochemical properties of polypyrrole modified electrodes. Tartu, 2004, 91 p.
44. **Kaido Sillar.** Computational study of the acid sites in zeolite ZSM-5. Tartu, 2004, 80 p.
45. **Aldo Oras.** Kinetic aspects of dATPαS interaction with $P2Y_1$ receptor. Tartu, 2004, 75 p.
46. **Erik Mölder.** Measurement of the oxygen mass transfer through the air-water interface. Tartu, 2005, 73 p.
47. **Thomas Thomberg.** The kinetics of electroreduction of peroxodisulfate anion on cadmium (0001) single crystal electrode. Tartu, 2005, 95 p.
48. **Olavi Loog.** Aspects of condensations of carbonyl compounds and their imine analogues. Tartu, 2005, 83 p.
49. **Siim Salmar.** Effect of ultrasound on ester hydrolysis in aqueous ethanol. Tartu, 2006, 73 p.
50. **Ain Uustare.** Modulation of signal transduction of heptahelical receptors by other receptors and G proteins. Tartu, 2006, 121 p.
51. **Sergei Yurchenko.** Determination of some carcinogenic contaminants in food. Tartu, 2006, 143 p.
52. **Kaido Tämm.** QSPR modeling of some properties of organic compounds. Tartu, 2006, 67 p.
53. **Olga Tšubrik.** New methods in the synthesis of multisubstituted hydrazines. Tartu. 2006, 183 p.
54. **Lilli Sooväli.** Spectrophotometric measurements and their uncertainty in chemical analysis and dissociation constant measurements. Tartu, 2006, 125 p.
55. **Eve Koort.** Uncertainty estimation of potentiometrically measured ph and pK_a values. Tartu, 2006, 139 p.
56. **Sergei Kopanchuk.** Regulation of ligand binding to melanocortin receptor subtypes. Tartu, 2006, 119 p.
57. **Silvar Kallip.** Surface structure of some bismuth and antimony single crystal electrodes. Tartu, 2006, 107 p.
58. **Kristjan Saal.** Surface silanization and its application in biomolecule coupling. Tartu, 2006, 77 p.
59. **Tanel Tätte.** High viscosity $Sn(OBu)_4$ oligomeric concentrates and their applications in technology. Tartu, 2006, 91 p.
60. **Dimitar Atanasov Dobchev**. Robust QSAR methods for the prediction of properties from molecular structure. Tartu, 2006, 118 p.
61. **Hannes Hagu**. Impact of ultrasound on hydrophobic interactions in solutions. Tartu, 2007, 81 p.
62. **Rutha Jäger.** Electroreduction of peroxodisulfate anion on bismuth electrodes. Tartu, 2007, 142 p.

63. **Kaido Viht.** Immobilizable bisubstrate-analogue inhibitors of basophilic protein kinases: development and application in biosensors. Tartu, 2007, 88 p.
64. **Eva-Ingrid Rõõm.** Acid-base equilibria in nonpolar media. Tartu, 2007, 156 p.
65. **Sven Tamp.** DFT study of the cesium cation containing complexes relevant to the cesium cation binding by the humic acids. Tartu, 2007, 102 p.
66. **Jaak Nerut.** Electroreduction of hexacyanoferrate(III) anion on Cadmium (0001) single crystal electrode. Tartu, 2007, 180 p.
67. **Lauri Jalukse.** Measurement uncertainty estimation in amperometric dissolved oxygen concentration measurement. Tartu, 2007, 112 p.
68. **Aime Lust.** Charge state of dopants and ordered clusters formation in CaF_2:Mn and CaF_2:Eu luminophors. Tartu, 2007, 100 p.
69. **Iiris Kahn**. Quantitative Structure-Activity Relationships of environmentally relevant properties. Tartu, 2007, 98 p.
70. **Mari Reinik.** Nitrates, nitrites, N-nitrosamines and polycyclic aromatic hydrocarbons in food: analytical methods, occurrence and dietary intake. Tartu, 2007, 172 p.
71. **Heili Kasuk.** Thermodynamic parameters and adsorption kinetics of organic compounds forming the compact adsorption layer at Bi single crystal electrodes. Tartu, 2007, 212 p.
72. **Erki Enkvist.** Synthesis of adenosine-peptide conjugates for biological applications. Tartu, 2007, 114 p.
73. **Svetoslav Hristov Slavov**. Biomedical applications of the QSAR approach. Tartu, 2007, 146 p.
74. **Eneli Härk.** Electroreduction of complex cations on electrochemically polished Bi(*hkl*) single crystal electrodes. Tartu, 2008, 158 p.
75. **Priit Möller.** Electrochemical characteristics of some cathodes for medium temperature solid oxide fuel cells, synthesized by solid state reaction technique. Tartu, 2008, 90 p.
76. **Signe Viggor.** Impact of biochemical parameters of genetically different pseudomonads at the degradation of phenolic compounds. Tartu, 2008, 122 p.
77. **Ave Sarapuu.** Electrochemical reduction of oxygen on quinone-modified carbon electrodes and on thin films of platinum and gold. Tartu, 2008, 134 p.
78. **Agnes Kütt.** Studies of acid-base equilibria in non-aqueous media. Tartu, 2008, 198 p.
79. **Rouvim Kadis.** Evaluation of measurement uncertainty in analytical chemistry: related concepts and some points of misinterpretation. Tartu, 2008, 118 p.
80. **Valter Reedo.** Elaboration of IVB group metal oxide structures and their possible applications. Tartu, 2008, 98 p.
81. **Aleksei Kuznetsov.** Allosteric effects in reactions catalyzed by the cAMP-dependent protein kinase catalytic subunit. Tartu, 2009, 133 p.

82. **Aleksei Bredihhin.** Use of mono- and polyanions in the synthesis of multisubstituted hydrazine derivatives. Tartu, 2009, 105 p.
83. **Anu Ploom.** Quantitative structure-reactivity analysis in organosilicon chemistry. Tartu, 2009, 99 p.
84. **Argo Vonk.** Determination of adenosine A_{2A}- and dopamine D_1 receptor-specific modulation of adenylate cyclase activity in rat striatum. Tartu, 2009, 129 p.
85. **Indrek Kivi.** Synthesis and electrochemical characterization of porous cathode materials for intermediate temperature solid oxide fuel cells. Tartu, 2009, 177 p.
86. **Jaanus Eskusson.** Synthesis and characterisation of diamond-like carbon thin films prepared by pulsed laser deposition method. Tartu, 2009, 117 p.
87. **Marko Lätt.** Carbide derived microporous carbon and electrical double layer capacitors. Tartu, 2009, 107 p.
88. **Vladimir Stepanov.** Slow conformational changes in dopamine transporter interaction with its ligands. Tartu, 2009, 103 p.
89. **Aleksander Trummal.** Computational Study of Structural and Solvent Effects on Acidities of Some Brønsted Acids. Tartu, 2009, 103 p.
90. **Eerold Vellemäe.** Applications of mischmetal in organic synthesis. Tartu, 2009, 93 p.
91. **Sven Parkel.** Ligand binding to 5-HT_{1A} receptors and its regulation by Mg^{2+} and Mn^{2+}. Tartu, 2010, 99 p.
92. **Signe Vahur.** Expanding the possibilities of ATR-FT-IR spectroscopy in determination of inorganic pigments. Tartu, 2010, 184 p.
93. **Tavo Romann**. Preparation and surface modification of bismuth thin film, porous, and microelectrodes. Tartu, 2010, 155 p.
94. **Nadežda Aleksejeva.** Electrocatalytic reduction of oxygen on carbon nanotube-based nanocomposite materials. Tartu, 2010, 147 p.
95. **Marko Kullapere.** Electrochemical properties of glassy carbon, nickel and gold electrodes modified with aryl groups. Tartu, 2010, 233 p.
96. **Liis Siinor.** Adsorption kinetics of ions at Bi single crystal planes from aqueous electrolyte solutions and room-temperature ionic liquids. Tartu, 2010, 101 p.
97. **Angela Vaasa.** Development of fluorescence-based kinetic and binding assays for characterization of protein kinases and their inhibitors. Tartu 2010, 101 p.
98. **Indrek Tulp.** Multivariate analysis of chemical and biological properties. Tartu 2010, 105 p.
99. **Aare Selberg.** Evaluation of environmental quality in Northern Estonia by the analysis of leachate. Tartu 2010, 117 p.
100. **Darja Lavõgina.** Development of protein kinase inhibitors based on adenosine analogue-oligoarginine conjugates. Tartu 2010, 248 p.
101. **Laura Herm.** Biochemistry of dopamine D_2 receptors and its association with motivated behaviour. Tartu 2010, 156 p.

102. **Terje Raudsepp.** Influence of dopant anions on the electrochemical properties of polypyrrole films. Tartu 2010, 112 p.
103. **Margus Marandi.** Electroformation of Polypyrrole Films: *In-situ* AFM and STM Study. Tartu 2011, 116 p.
104. **Kairi Kivirand.** Diamine oxidase-based biosensors: construction and working principles. Tartu, 2011, 140 p.
105. **Anneli Kruve.** Matrix effects in liquid-chromatography electrospray mass-spectrometry. Tartu, 2011, 156 p.
106. **Gary Urb.** Assessment of environmental impact of oil shale fly ash from PF and CFB combustion. Tartu, 2011, 108 p.
107. **Nikita Oskolkov.** A novel strategy for peptide-mediated cellular delivery and induction of endosomal escape. Tartu, 2011, 106 p.
108. **Dana Martin.** The QSPR/QSAR approach for the prediction of properties of fullerene derivatives. Tartu, 2011, 98 p.
109. **Säde Viirlaid.** Novel glutathione analogues and their antioxidant activity. Tartu, 2011, 106 p.
110. **Ülis Sõukand.** Simultaneous adsorption of Cd^{2+}, Ni^{2+}, and Pb^{2+} on peat. Tartu, 2011, 124 p.
111. **Lauri Lipping.** The acidity of strong and superstrong Brønsted acids, an outreach for the "limits of growth": a quantum chemical study. Tartu, 2011, 124 p.
112. **Heisi Kurig.** Electrical double-layer capacitors based on ionic liquids as electrolytes. Tartu, 2011, 146 p.
113. **Marje Kasari.** Bisubstrate luminescent probes, optical sensors and affinity adsorbents for measurement of active protein kinases in biological samples. Tartu, 2012, 126 p.
114. **Kalev Takkis.** Virtual screening of chemical databases for bioactive molecules. Tartu, 2012, 122 p.
115. **Ksenija Kisseljova**. Synthesis of aza-β^3-amino acid containing peptides and kinetic study of their phosphorylation by protein kinase A. Tartu, 2012, 104 p.
116. **Riin Rebane.** Advanced method development strategy for derivatization LC/ESI/MS. Tartu, 2012, 184 p.
117. **Vladislav Ivaništšev.** Double layer structure and adsorption kinetics of ions at metal electrodes in room temperature ionic liquids. Tartu, 2012, 128 p.
118. **Irja Helm.** High accuracy gravimetric Winkler method for determination of dissolved oxygen. Tartu, 2012, 139 p.
119. **Karin Kipper.** Fluoroalcohols as Components of LC-ESI-MS Eluents: Usage and Applications. Tartu, 2012, 164 p.
120. **Arno Ratas.** Energy storage and transfer in dosimetric luminescent materials. Tartu, 2012, 163 p.
121. **Reet Reinart-Okugbeni**. Assay systems for characterisation of subtype-selective binding and functional activity of ligands on dopamine receptors. Tartu, 2012, 159 p.

122. **Lauri Sikk.** Computational study of the Sonogashira cross-coupling reaction. Tartu, 2012, 81 p.
123. **Karita Raudkivi.** Neurochemical studies on inter-individual differences in affect-related behaviour of the laboratory rat. Tartu, 2012, 161 p.
124. **Indrek Saar.** Design of GalR2 subtype specific ligands: their role in depression-like behavior and feeding regulation. Tartu, 2013, 126 p.
125. **Ann Laheäär.** Electrochemical characterization of alkali metal salt based non-aqueous electrolytes for supercapacitors. Tartu, 2013, 127 p.
126. **Kerli Tõnurist.** Influence of electrospun separator materials properties on electrochemical performance of electrical double-layer capacitors. Tartu, 2013, 147 p.
127. **Kaija Põhako-Esko.** Novel organic and inorganic ionogels: preparation and characterization. Tartu, 2013, 124 p.
128. **Ivar Kruusenberg.** Electroreduction of oxygen on carbon nanomaterial-based catalysts. Tartu, 2013, 191 p.
129. **Sander Piiskop.** Kinetic effects of ultrasound in aqueous acetonitrile solutions. Tartu, 2013, 95 p.
130. **Ilona Faustova**. Regulatory role of L-type pyruvate kinase N-terminal domain. Tartu, 2013, 109 p.
131. **Kadi Tamm.** Synthesis and characterization of the micro-mesoporous anode materials and testing of the medium temperature solid oxide fuel cell single cells. Tartu, 2013, 138 p.
132. **Iva Bozhidarova Stoyanova-Slavova.** Validation of QSAR/QSPR for regulatory purposes. Tartu, 2013, 109 p.
133. **Vitali Grozovski**. Adsorption of organic molecules at single crystal electrodes studied by *in situ* STM method. Tartu, 2014, 146 p.
134. **Santa Veikšina**. Development of assay systems for characterisation of ligand binding properties to melanocortin 4 receptors. Tartu, 2014, 151 p.
135. **Jüri Liiv.** PVDF (polyvinylidene difluoride) as material for active element of twisting-ball displays. Tartu, 2014, 111 p.
136. **Kersti Vaarmets.** Electrochemical and physical characterization of pristine and activated molybdenum carbide-derived carbon electrodes for the oxygen electroreduction reaction. Tartu, 2014, 131 p.
137. **Lauri Tõntson**. Regulation of G-protein subtypes by receptors, guanine nucleotides and Mn^{2+}. Tartu, 2014, 105 p.
138. **Aiko Adamson.** Properties of amine-boranes and phosphorus analogues in the gas phase. Tartu, 2014, 78 p.
139. **Elo Kibena**. Electrochemical grafting of glassy carbon, gold, highly oriented pyrolytic graphite and chemical vapour deposition-grown graphene electrodes by diazonium reduction method. Tartu, 2014, 184 p.
140. **Teemu Näykki.** Novel Tools for Water Quality Monitoring – From Field to Laboratory. Tartu, 2014, 202 p.
141. **Karl Kaupmees.** Acidity and basicity in non-aqueous media: importance of solvent properties and purity. Tartu, 2014, 128 p.

142. **Oleg Lebedev**. Hydrazine polyanions: different strategies in the synthesis of heterocycles. Tartu, 2015, 118 p.
143. **Geven Piir.** Environmental risk assessment of chemicals using QSAR methods. Tartu, 2015, 123 p.
144. **Olga Mazina.** Development and application of the biosensor assay for measurements of cyclic adenosine monophosphate in studies of G protein-coupled receptor signalinga. Tartu, 2015, 116 p.
145. **Sandip Ashokrao Kadam.** Anion receptors: synthesis and accurate binding measurements. Tartu, 2015, 116 p.
146. **Indrek Tallo.** Synthesis and characterization of new micro-mesoporous carbide derived carbon materials for high energy and power density electrical double layer capacitors. Tartu, 2015, 148 p.
147. **Heiki Erikson.** Electrochemical reduction of oxygen on nanostructured palladium and gold catalysts. Tartu, 2015, 204 p.
148. **Erik Anderson.** *In situ* Scanning Tunnelling Microscopy studies of the interfacial structure between Bi(111) electrode and a room temperature ionic liquid. Tartu, 2015, 118 p.
149. **Girinath G. Pillai.** Computational Modelling of Diverse Chemical, Bio-chemical and Biomedical Properties. Tartu, 2015, 140 p.
150. **Piret Pikma.** Interfacial structure and adsorption of organic compounds at Cd(0001) and Sb(111) electrodes from ionic liquid and aqueous electrolytes: an *in situ* STM study. Tartu, 2015, 126 p.
151. **Ganesh babu Manoharan.** Combining chemical and genetic approaches for photoluminescence assays of protein kinases. Tartu, 2016, 126 p.
152. **Carolin Siimenson**. Electrochemical characterization of halide ion adsorption from liquid mixtures at Bi(111) and pyrolytic graphite electrode surface. Tartu, 2016, 110 p.
153. **Asko Laaniste.** Comparison and optimisation of novel mass spectrometry ionisation sources. Tartu, 2016, 156 p.
154. **Hanno Evard.** Estimating limit of detection for mass spectrometric analysis methods. Tartu, 2016, 224 p.
155. **Kadri Ligi.** Characterization and application of protein kinase-responsive organic probes with triplet-singlet energy transfer. Tartu, 2016, 122 p.
156. **Margarita Kagan.** Biosensing penicillins' residues in milk flows. Tartu, 2016, 130 p.
157. **Marie Kriisa.** Development of protein kinase-responsive photo-luminescent probes and cellular regulators of protein phosphorylation. Tartu, 2016, 106 p.
158. **Mihkel Vestli.** Ultrasonic spray pyrolysis deposited electrolyte layers for intermediate temperature solid oxide fuel cells. Tartu, 2016, 156 p.
159. **Silver Sepp**. Influence of porosity of the carbide-derived carbon on the properties of the composite electrocatalysts and characteristics of polymer electrolyte fuel cells. Tartu, 2016, 137p.
160. **Kristjan Haav**. Quantitative relative equilibrium constant measurements in supramolecular chemistry. Tartu, 2017, 158 p.

161. **Anu Teearu**. Development of MALDI-FT-ICR-MS methodology for the analysis of resinous materials. Tartu, 2017, 205 p.
162. **Taavi Ivan**. Bifunctional inhibitors and photoluminescent probes for studies on protein complexes. Tartu, 2017, 140 p.
163. **Maarja-Liisa Oldekop**. Characterization of amino acid derivatization reagents for LC-MS analysis. Tartu, 2017, 147 p.
164. **Kristel Jukk**. Electrochemical reduction of oxygen on platinum- and palladium-based nanocatalysts. Tartu, 2017, 250 p.
165. **Siim Kukk**. Kinetic aspects of interaction between dopamine transporter and *N*-substituted nortropane derivatives. Tartu, 2017, 107 p.