CHARLES UNIVERSITY

FACULTY OF PHARMACY IN HRADEC KRÁLOVÉ

DEPARTMENT OF PHARMACEUTICAL CHEMISTRY AND DRUG ANALYSIS

Creation and analysis of in-house database of pyrazine derivatives with potential antimicrobial activity

(Diploma Thesis)

Head of Department: Prof. PharmDr. Martin Doležal, Ph.D.

Supervisor: PharmDr. Jan Zitko, Ph.D.

Candidate: Legae G. B. Kebakuile

Hradec Králové, 2018

Declaration

I declare that this thesis is my original work. All literature and other resources which were used during the preparation of this review are listed in bibliography and properly cited.

Signed: . . . . . Dated: . . . . . .

Acknowledgement

I would hereby like to show my utmost appreciation and gratitude to my mentor PharmDr. Jan Zitko, Ph.D. for his patience, guidance and motivation during the creation of this work. I would also like to thank my parents Margaret and Edison Kebakuile for always believing in me.

This work was supported by project SVV 260 401.

ABSTRACT

In the early phases of drug design and development, scientists must overcome many challenges involved in identifying potential drug-like or lead-like compounds. This has led to the need of creating large sets of chemical data which will aid in improving the identification of pharmacophores and active compounds. Various scientific fields especially pharmacology, medicinal chemistry and biochemistry have begun to employ the use of computer sciences to aid in the screening for potential leads with more specificity with regards to drug-like compounds' or substances' bioactivity. The emphasis of this project was to create a database containing a collection of pyrazine compounds synthesized overtime in the Faculty of Pharmacy in Hradec Kralove (Charles University) with the aim of having antimycobacterial (and possible antibacterial and antifungal) activity, and further utilize this database to predict descriptors important for pharmacokinetic and bioavailability properties. This project seeks to demonstrate how certain molecular descriptors can be used as reliable chemoinformation to determine the likeliness or possibility of developing a lead-like or drug-like compound by utilizing computer software. An in-house database of 623 compounds saved in SMILES format was created and used in demonstrating quantitative structure-activity relationships (QSAR) and in evaluating and analyzing whether optimal lead-like or drug-like compounds are being produced. The database can be used to guide future synthesis with regards to CADD (Computer aided drug design).

TABLE OF CONTENTS

ABSTRACT 4

List of important abbreviations 7

1. INTRODUCTION 9

2. AIM 10

3. exper ..... **Error! Bookmark not defined.**

3.1 Molecular descriptors 11

3.2 File formats used to store molecular structures 13

3.2.1 Connection tables/adjacency matrices 13

3.2.2 Linear string notation 14

3.3 Molecular descriptors utilized in the database 19

3.4 Molecular descriptors used for predicting kinetics and bioavailability 20

3.5 Rules and descriptors which are used to define drug-like and lead-like molecules 21

3.6 Drug-likeness versus lead-likeness 22

3.7 Role of in-house databases in drug design and development 25

4. EXPERIMENTAL PART 29

4.1 Procedure, software and sources used to create and analyze the database 29

4.2 Stepwise illustration of how the database was created (using compound No. 291) 31

5. RESULTS 35

5.1 Statistics of the database 35

5.2 GRAPHS SHOWING THE IMPORTANT DESCRIPTOR STATISTICS OF THE DATABASE 36

5.2.1 MW 36

5.2.2 HBA 37

5.2.3 HBD 38

5.2.4 NRot 39

5.2.5 PSA 40

5.2.6 logP 41

5.3 Statistics of Descriptors 42

5.4 Outputs of the database 45

6. CONCLUSION 47

APPENDIX I 48

Works Cited 52

List of important abbreviations

CADD	Computer aided drug design
clogP	fragment based or calculated logP
CMR	Calculated Molar Refractivity
DT	Diploma thesis
HBA	Hydrogen-bond acceptor
HBD	Hydrogen-bond donor
HIV	human immunodeficiency virus
HTS	high-throughput screening
InChI	International Chemical Identifier
InChIKey	hashed International Chemical Identifier
LogP	base 10 logarithm of partition coefficient
Mtbc	*Mycobacterium tuberculosis* strand H37Rv
MIC	minimum inhibitory concentration
MDR-TB	multidrug-resistant tuberculosis
MW	molecular weight
NRot	Number of rotatable bonds
NMR	nuclear magnetic resonance
POA	Pyrazinoic acid
PSA	polar surface area
PZA	Pyrazinamide
QSAR	quantitative structure-activity relationships
SMARTS	SMILES Arbitrary Target Specification
SMILES	Simplified Molecular-Input Line Entry System

TAACF	Tuberculosis Antimicrobial Acquisition and Coordinating Facility
TB	tuberculosis
XDR-TB	extensively drug-resistant tuberculosis

1. INTRODUCTION

Tuberculosis (TB) continues to be a major health issue despite it being treatable and curable. Approximately 10.4million people were diagnosed with TB in 2016, with the disease being the top cause of death from a single infectious agent (1). TB is also the leading cause of death in human immunodeficiency virus (HIV)-positive individuals, causing a quarter of all HIV-related deaths. Drug-resistance has posed a great challenge in controlling the disease. Multidrug-resistant tuberculosis (MDR-TB) strains (that is those resistant to the most potent anti-TB drugs rifampicin and isoniazid) and extensively drug-resistant TB (XDR-TB) strains (those strains not susceptible to second-line anti-tuberculosis drugs like fluoroquinolones) have begun to emerge, causing even more detriment (2). Pyrazinamide (PZA) was first used against TB in 1952, and was unique because it reduced the treatment course from a typical 9-12 months to 6 months. This action is attributed to its ability to kill certain dormant microorganisms at acidic conditions that are not suited for other anti-TB agents.

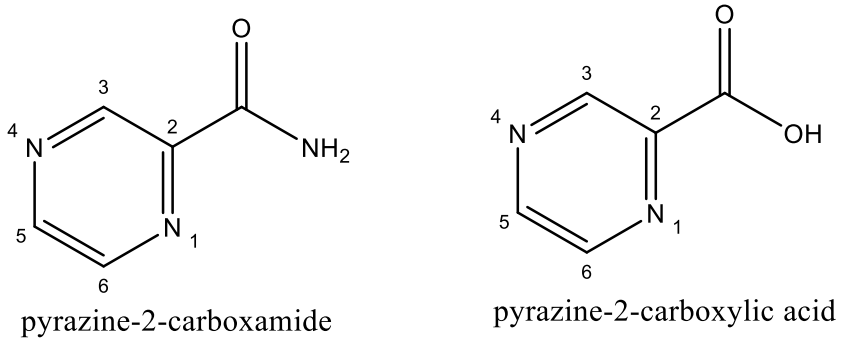


Figure 1: Structures and systematic names of Pyrazinamide and Pyrazinoic acid

PZA is the active form of pyrazinoic acid (POA). They exert their action by multiple mechanisms. Firstly by inhibiting the specific mycobacterial enzyme Fatty Acid Synthase I (FAS I) leading to the disruption of membrane function (3). PZA must first cross the mycobacterial cell envelope, via active transport. It then accumulates within the cellular environment and blocks translational processes involved in protein synthesis. (4)

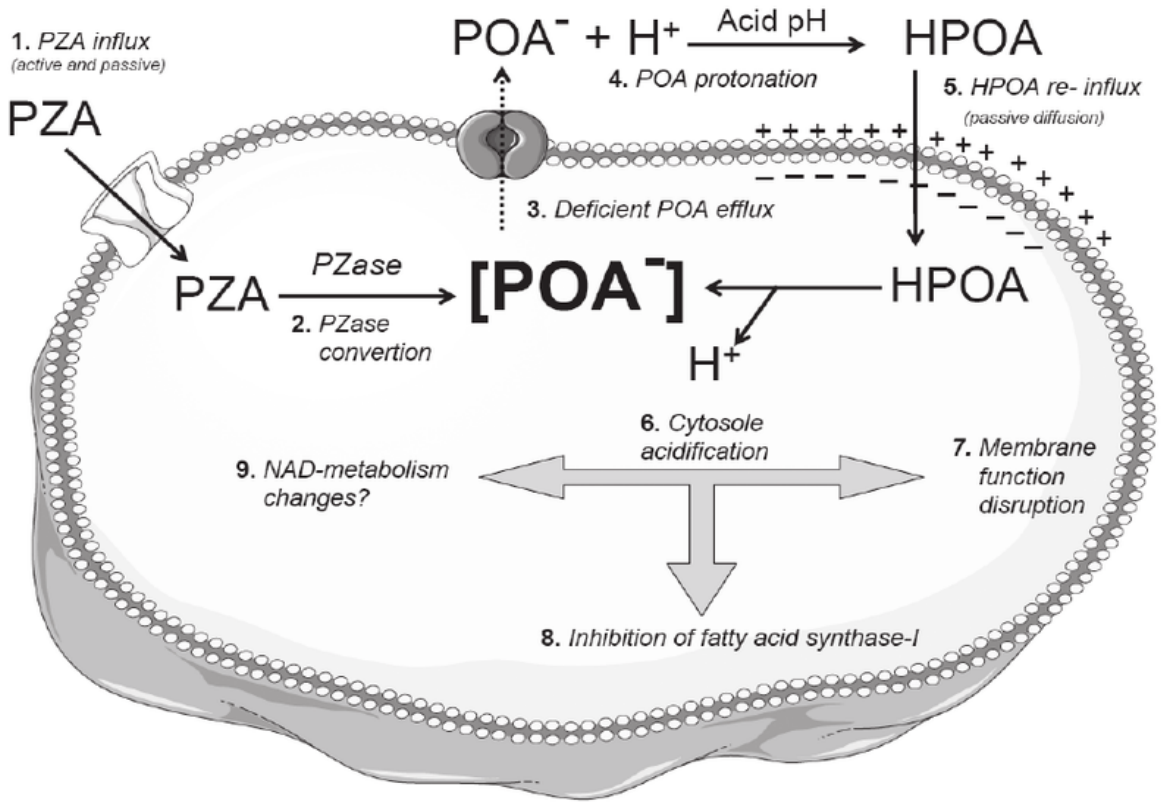


Figure 2: Illustration of the different mechanisms of action of PZA and POA (5)

2. AIM

The main aim of this thesis was to create an electronic in-house database of pyrazine derivatives possessing antimicrobial activity that have been published by teams in the Faculty of Pharmacy (Charles University, Hradec Kralove). Furthermore, it must be possible to use the database in demonstrating quantitative structure-activity relationships (QSAR) and to evaluate and analyze whether optimal **lead-like** or **drug-like** compounds are being produced. The resulting database must also be usable in guiding future syntheses.

3.1 Molecular descriptors

The structure of a chemical compound is the main interest for chemists as this directly related to the biological activity. Molecular descriptors, essentially, give an idea to the features of a chemical entity. They relate the structure of the molecule to the properties it possesses. There are various types of molecular descriptors according to (6). Those that are used commonly are constitutional, surface, molecular connectivity, electrostatic, shape, geometry, quantum-chemical, physico-chemical and hybrid.

3.1.1 Constitutional molecular descriptors inform you about a compound's chemical composition. For example: the number of atoms in the molecule, the number of bonds in a molecule, or the number atoms/groups of type "x".

3.1.2 Topological/surface molecular descriptors are very important is they are used in determining the solubility and permeability of a drug according to its surface properties. PSA (Polar Surface Area) describes total surface area of the molecule represented by polar atoms like oxygen, nitrogen and halogens. Other topological descriptors include SAS (Solvent-Accessible Surface), van der Waals surface and MS (Molecular Surface area).

3.1.3 Quantum descriptors, which are derived computationally (7), cover all geometric and electronic features of a molecule, instead of its empiric ones. These include Lewis acid and Lewis base properties, charge transfer characteristics, hydrogen bonding ability, polarizability, polarity, steric factors, and lipophilicity.

Amongst physico-chemical properties employed in structure-activity relationship and prediction of pharmacokinetic properties are boiling and melting point, dipole moment, molar refractivity and water partition coefficient. From the point of view of designing drugs, the most relevant molecular descriptors are the molecule's solubility and its 1-octanol/water partition coefficient. Solubility directly affects the oral bioavailability of a drug. Furthermore key physico-chemical

properties that are associated with hydrophobicity and lipophilicity include solubility, hydrogen bonding capacity, and the state of ionization. These properties in turn strongly influence permeability through biological membranes, thereby, affecting absorption, distribution, metabolism and elimination (8).

It is also worth noting that molecular descriptors can be classed in various manners (9). The basis can be by origin (i.e. experimental versus calculated). Experimentally obtained descriptors include: logP, aqueous solubility and Abraham's hydrogen-bond acidity/basicity parameters (10). Calculated molecular descriptors are assessed in-silico from different dimensional molecular models. The basis can be on the described object (i.e. global, local or field description). Global descriptors describe the entire molecule (e.g. molecular surface). Local descriptors are for particular atoms or fragments of the molecule (e.g. bond polarizability and atomic charges). Field descriptors are detail molecular fields in the area surrounding the molecule such as electrostatic potential. Finally, classification can be according to the dimensionality (1D, 2D or 3D) of the structure. 1-Dimensional descriptors include constitutional counts and molecular weight. 2-Dimensional descriptors are topology-based (that is to say they are obtained from the molecular graph representation (branching degree, shape, steric effects). 3-Dimensional descriptors are obtained through quantum mechanics. Examples include LUMO (Lowest occupied molecular orbital energy) HOMO (Highest occupied molecular orbital energy), IP (Ionization Potential), ΔE (Protonation Energy). 2D and 3D descriptors are especially useful in identifying and studying the pharmacophore and lipophilicity potential.

3.2 File formats used to store molecular structures

3.2.1 Connection tables/adjacency matrices

Connection tables are employed to store information regarding the structure of a molecule. The first table (called the atom table) contains two fields, one identifying the atom in discussion and the other identifies that atom (e.g. N, O, S, C, P, Cl etc). The second table (called the bond table) has three fields, of which, two of those indicate which atoms the bond connects and one field indicates the bonder order (i.e. single, double or triple). Further fields may need to be added to the tables to denote chirality/stereochemistry (11).

With regard to adjacency matrices, topology (graph theory) is used whereby a square matrix represents a finite graph. The atoms in the molecule are usually assigned a type describing their chemical identity. This type can be a mnemonic symbol or integral number like '12' or 'Csp2'. "The type reflects not only an element but also a particular arrangement of bonds formed by the atom, and its formal charge". Also, the type of atom may depend on neighboring atoms in the molecule. The majority of molecular modeling systems assigned different types, for example, to the amine, ammonium, imine, amide and other nitrogens. Likewise, bonds are designated types: single, double, aromatic, etc. (12)

Aside from real chemical atoms and bonds, most systems introduce pseudo/virtual atoms and bonds which can be used to indicate important molecular features like geometry and orientation. The relation amongst atoms should be given to fully specify molecules to the computer. This information is made up of two parts: "the specification of bonds and specification of geometry (the spatial relation between atoms)" (12).

Concerning connection tables and adjacency matrices fall several file formats which are coded using these methods. CML (Chemical Mark-up Language) supports concepts such as reaction schemes and spectra (13). PDB (Protein Data Bank) is a database containing 3D images of

molecules and crystals, especially large biomolecules like nucleic acids, obtained via x-ray crystallography and NMR (nuclear magnetic resonance) spectroscopy. The original format of such a file was pdb, more recent formats are mmCIF (macromolecular crystallographic information file) and PDBML (14).

Molfile is a format type containing information about atoms, connectivity (and coordinates) of a molecule. Multiple molfiles can be joined and put together with more information about the compounds. Those molfiles are what then build SDfiles (Structure-data files), basically SDfiles wrap the information about molecules contained in molfiles as molfiles can be fragments. RDfiles (reaction-data files) have the same concept as SDfiles but with a more general format which may include reactions and molecules, together with their related data.

3.2.2 Linear string notation

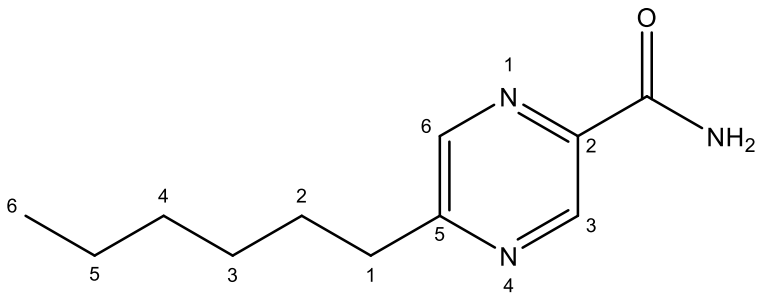
3.2.2.1 SMILES

SMILES (Simplified Molecular-Input Line Entry System) is a line-notation that describes the structure of the chemical species and allows conversion between 2D and 3D models. A SMILES string is composed of two different parts, the syntax specification and the semantic specification. The syntax specification specifies how the atoms, bonds, parentheses, digits etc are represented, and the semantic describes how those symbols are interpreted as a sensible molecule. Every atom that is not a hydrogen is represented by its atomic symbol enclosed within brackets. It is assumed by convention that hydrogens comprise the remaining part of an atom's lowest normal valence while formal charges are expressed using a + or – sign. In aromatic compounds, the atoms are specified by the lowercase atomic symbol, and the bonds are indicated by '-' (single), '=' (double) and '#' (triple). Branched systems are denoted by enclosing within parentheses. A cyclic structure is depicted by disjoining a ring at a single or aromatic bond then numbering the atoms adjacent to the break with a number. Tautomeric structures (that is their bonds and mobile hydrogens) do not have definitions in SMILE notation, therefore, tautomers must be clearly specified as separate structures (15).

An example of SMILES using molecules from the prepared database

No. 63

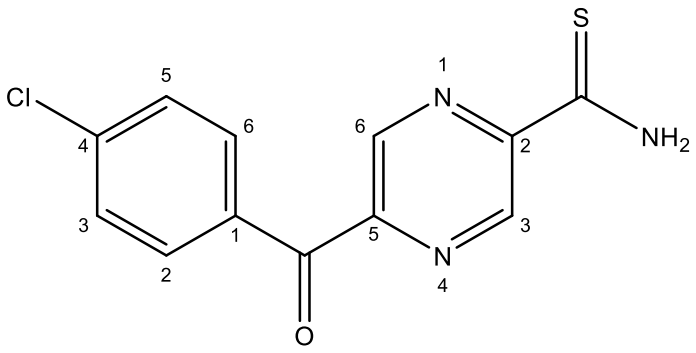
O=C(C1=NC=C(CCCCCC)N=C1)N



5-hexylpyrazine-2-carboxamide

No. 336

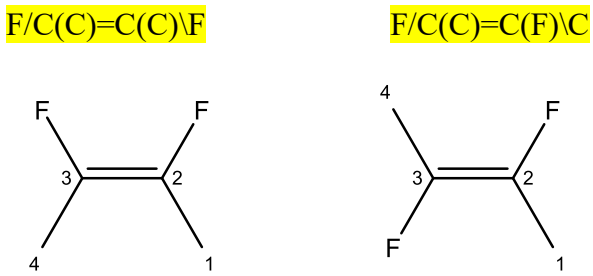
S=C(C1=NC=C(C(C2=CC=C(Cl)C=C2)=O)N=C1)N



5-(4-chlorobenzoyl)pyrazine-2-carbothioamide

Figure 3: chemical structures of Compounds No. 63 and 336 from the database

Even though the order of priority of atoms/groups connected to stereogenic centers does not matter when it comes to this type of notation, Optical configuration can still be expressed in SMILES (16). Clockwise (S) chirality is shown by '@@' while anticlockwise (R) chirality is shown by '@' after the symbol of the first atom connected the stereocenter (i.e. the order depends solely on the writer of the SMILE). Geometric configuration of atoms around double bonds can be denoted within a SMILE using the symbols '/' and '\' in pairs with the idea in mind that the symbols represent bonds pointing above or below the double bond (17). Using these symbols, whether an atom is above or below is in relation to the carbon atom not the double. See illustration below.



(*Z*)-2,3-difluorobut-2-ene (*E*)-2,3-difluorobut-2-ene

Figure 4: Illustration of cis-(*Z*) and trans-(*E*) configuration using SMILES

It is important to note that in practice, the syntax and semantics are usually mixed together in the code that implements a SMILES analyzer or reader (15) and that different versions of a SMILE can represent a single molecule. Software such as ChemDraw can be used to produce SMILES.

3.2.2.2 SMARTS

SMILES Arbitrary Target Specification (SMARTS) uses direct rules extended from SMILES and allows for specification of sub-structural patterns and atoms. This affords one the ability to search for SMILES as substructures. Over and above the SMILES naming conventions, SMARTS includes logical operators, such as "AND" (&), "OR" (,), and "NOT" (!), and special bond and atomic symbols permitting flexibility to chemical names.

3.2.2.3 InChI

Another well-established line notation is the InChI (International Chemical Identifier) string developed by IUPAC (International Union of Pure and Applied Chemistry). InChI is a unique, non-proprietary chemical identifier produced, using software, from the chemical structure drawn on a computer. There are fundamental features governing the design of an InChI, name (18). Firstly, an approach based on the structure is used. Anyone should be able to produce the identifier from only the structural formula of the chemical entity. Secondly, the identifier/label must be unique to one substance, guaranteeing that each time in application the substance will be labeled the same. Thirdly, since it is non-proprietary, the source code of the InChI must be openly and freely accessible, hence the programming software. An InChI must be applicable to the whole organic chemistry field, significantly to inorganic entities and of course a general usability in all of chemistry. Fifthly, it should be possible to create the same InChI for structures drawn using different conventions and styles (to a reasonable extent), particularly those represented by mesomers (stereochemically superimposable mirror image isomers). The sixth critical point in generating an InChI is that an "hierarchical approach allowing encoding of molecular structure with different levels of granularity, this depends on the software. A specific necessity is the capability of including or excluding stereochemical, tautomer and isotope information. Finally, an identifier that has some "default" controls to ensure interoperability (interusability) in large databases. It is worth noting that InChI may be generated for combinations of unbound components, not just single structure. These may be viewed as equimolar mixture representations. (18) See the InChI of compounds number 63 and 336 in the database below obtained via ChemDraw.

No. 63: 1S/C11H17N3O/c1-2-3-4-5-6-9-7-14-10(8-13-9)11(12)15/h7-8H,2-6H2,1H3,(H2,12,15)

No. 336: 1S/C12H8ClN3OS/c13-8-3-1-7(2-4-8)11(17)9-5-16-10(6-15-9)12(14)18/h1-6H,(H2,14,18)

To ease interoperability and information sharing, especially when handling large volumes of data, a standard InChI which differentiates chemical entities at a “level of stereochemistry, connectivity and isotopic composition” was developed in 2009. Connectivity is the valence-bond connectivity that is invariant to tautomers, whereby different tautomers have matching hydrogen layer/connectivity. Stereochemistry is stereogenic atom and bond configuration where absolute stereo or no stereo is permitted, and unknown configurations are regarded as undefined. Composition of isotopes is, when specified, the mass number of the isotopic atoms. (18)

3.2.2.4 InChIKey

InChIKey is a compact, 27-character chemical identifier based on InChI. It is more practical and convenient for internet searches and databases used for indexing. The InChIKey is required to be in upper case and must not be shortened or changed for it to be accepted by all search engines. An example of InChIKey of compounds number 63 and 336 in the database is given beneath, obtained using ChemDraw. As the InChI cannot be reconstructed from InChIKey, it is a must that an InChIKey is always linked to the original InChI in order to get back the original structure. This is done through a resolver which acts as a look up service such as PubChem (19), NCI (National Cancer Institute) (20), ChemSpider (21) and UniChem (22).

No. 36: QWNFJQMIKHKSEU-UHFFFAOYSA-N

No. 336: LTBULBXTNYJLBA-UHFFFAOYSA-N

In creating this database, the molecules were first stored as molecular structures in the internal ChemDraw format. Following this, the structures were converted into SMILES notation. This was done to create a smaller sized file which allowed better manipulation of the database by preventing freezing/lagging. More importantly, SMILES do ‘carry’ enough data about the specific molecular entity that can be exported to all chemical software.

3.3 Molecular descriptors utilized in the database

1. Number of Atoms is the total number of atoms that make up the molecule
2. Number of Heavy Atoms is any atom of any element besides hydrogen
3. Molar refractivity reflects arrangements of ions' electron shells in molecules, informing on the electronic polarization of ions. It shows the changes in the properties caused by deformation or polarization of the ions' electron shells due to influence of the electric fields of adjacent ions (23)
4. CMR (Calculated Molar Refractivity) is used to estimate the molecular volume and steric bulk of a compound (23)
5. HBA (Hydrogen-bond acceptor) are heteroatoms with lone electron pair(s) that can to form a hydrogen-bond (24)
6. HBD (Hydrogen-bond donor) is a heteroatom that is covalently bound to at least one hydrogen.
7. Number of Rotatable Bonds are the bonds that are able to freely rotate
8. PSA (Polar-surface Area) describes the total hydrogen bonding capacity of a molecule. This is the van der Waals surface arising from all nitrogen and oxygen atoms together with all the hydrogens attached to them. It has been used to determine intestinal absorption. (24)
9. LogP (Partition Coefficient) is the octanol/water partition constant. P is the ratio of activity or concentration of a substance in a mixture of two immiscible solvents, an aqueous (water) phase and an organic phase (octanol) at equilibrium. This mimics a biological environment and is used to study a substance's affinity for the aqueous (hydrophilic) or organic (lipophilic/hydrophobic) phase. Lipophilicity, for organic substances, can be described in terms of the partition coefficient (log P), the intrinsic lipophilicity of functional groups and carbon skeleton in the molecule without dissociation or ionization.
10. cLogP (fragment based logP) is puts together values for fragments of a structure and correction factors depending on the particular way the parts are combined.
11. LogS is the solubility of a substance in water, this directly affects the absorption and distribution. Solubility itself is measured in moles per liter, but the logarithmic form has no units.
12. pKa is the measure of acid/base strength or acid dissociation/ionization constant. The more negative the pKa the stronger the acid

13. Molecular weight (MW) is the mass of a molecule (sum of all elements from the molecular formula).

3.4 Molecular descriptors used for predicting kinetics and bioavailability

The set of optimal descriptors include molecular mass (MW), molecular surface area (MSA), molecular volume (MV), molecular refractivity (MR), total hydrogen count (HC), partition coefficient (clogP), rotatable bonds (NRot), polar surface area (PSA) and solubility index (logS). (24)

3.5 Rules and descriptors which are used to define drug-like and lead-like molecules

Lipinski's 'Rule of Five', also called the '**Pfizer rule'**, is the main approach used to determine potential leads or drugs. Based on oral formulations, this rule focuses mainly on the permeation or absorption of a drug/molecule because this is the first pharmacokinetic barrier that a drug must overcome to elicit a pharmacological action. Classes of compounds that are substrates to biological transporters are excluded from this rule (25). This rule was applied to already know drugs vetted by different professional bodies (26). The 'rule of five' states that the absorption of a molecule will be likely impaired when the molecule's octanol/water partition coefficient is greater than 5, when its molecular weight is greater than 500, the molecule possesses more than 5 Hydrogen bond donor groups and lastly the molecule has over 10 Hydrogen bond acceptor groups. (27).

The 'Pfizer rule' dictates that a molecule must have a balance in hyrdophobic-hydrophilic characteristics. This character is profoundly influenced by two physicochemical criteria: solubility in water (which is fundamental to delivering the drug) and hydrophobicity (which affects absorption, transport and distribution of the drug). The aqueous solubility is expressed as the molecular descriptor logS, the log units of molar solubility. The base 10 logarithm of octanol/water partition coefficient of a chemical substance is key in quantitative structure-activity relationship studies. LogP is linked to hydrophobicity of organic compounds, and this plays a critical part in setting the tone of absorption, distribution, metabolism and elimination/clearance processes. In particular "drug-membrane interactions, drug transport, biotransformation, distribution, accumulation, protein and receptor binding are all related to drug hydrophobicity" (27). A higher number of hydrogen bonds lessen partitioning from the aqueous phase into the bilayer lipid membrane resulting in permeability being affected (therefore H-bonds must be minimized). (27)

High bioavailabilty is a crucial matter that, when it comes to the most preferred administration route, orally, is difficult to achieve. Bioavailability is the percentage of drug which reaches the central circulation from the point of administration unchanged. If the molecule has unfavorable 'rule of five' character then its bioavailability will be predictably poor due to the disadvantageous physicochemical properties. Compounds with PSA less than 140Å and number of rotatable bonds

less than 10 will have good bioavailability (28). It is important to note, as described by (24), PSA and the number of rotatable bonds are important discriminants for compounds with oral activity. A greater number of rotatable bonds in a molecule correspond with a reduced rate of permeation. The sum of polar atoms in a molecule (PSA) determines drug-permeability; the lesser the PSA the better/increased the permeation. Molecular entities with PSA below 90Å are able to cross the blood-brain barrier (29), where those below 60Å have a biovailability above 90% (30) .

3.6 Drug-likeness versus lead-likeness

A **drug** is deemed as an already patented compound that causes an expected or intended physiological effect. More specifically according to the World Health Organization a drug is "any chemical agent that alters the biochemical physiological processes of tissues or organisms" (31). **Drug-likeness** refers to products that share such property or similarity. A **lead**, however, is described as a prototype chemical entity that meets certain criteria in the process of drug discovery (32). Using various screens, biological activity against a known target should be determined, pharmacokinetic properties must be validated. Moreover, the compound must be in series with others to show structure-activity relationship during analysis; this matters greatly to medicinal chemists during lead optimization (33). An alternative concept that is becoming more widely accepted is that **drug-like** and **lead-like** compounds require different physicochemical properties.

A lead can be a drug or result in multiple drugs. **Lead-likeness** is determined by a strict criterion called the "**Rule-of-Three**" or **Ro3**. Surmised by Astex Technologies, specific limits regarding selected molecular descriptors must be observed in fragment-based screening (32). The limits to observe are MW below 300, HBD, HBA and clogP less than or equal to 3 respectively. Note that clogP (compound or fragment logP) (34) is determined by using data from full compounds, or fragments that have been deduced experimentally followed by quantitative structure-activity relationship modelling. The clogP differs from logP in that logP is generally predicted via software rather than through practical experiments (34). Furthermore, a PSA of not more than 60 and NRot (number of rotatable bonds) not exceeding 3 have been suggested as useful selection criteria of

promising fragments. Additionally, the presence of a single charge (secondary or tertiary amine preferred) is an advantage.

It can be said that Lipinski's Ro5 can be used to determine drug-likeness, whereas the Ro3 parameters from T. I. Oprea et al (of the AstraZeneca group) (35) (36) are more suited for lead-likeness determination.

In the figure below, the values of lead-likeness are based on Oprea's work. In the original finding of Lipinski's Rule, NRot (number of rotatable bonds) and PSA (Polar Surface Area) were not defined. Later on in 2002, work by D. L. Veber et al (through GlaxoSmithKline) elaborated on this. (24)

LEAD-LIKENESS	DRUG-LIKENESS
Physiochemical properties typical of good lead compounds in target-driven drug discovery programs that employ biochemical assays: • MW<300 • HBA≤3 • HBD≤3 • logP≤3 • PSA≤60 • NRot≤3	Physiochemical properties that improve probability of success in drug development by addressing issues of absorption and bioavailability: • MW <500 • logP ≤5 • HBD ≤5 • HBA ≤10 • PSA ≤140 • NRot ≤10

Figure 5: The properties of lead-likeness as compared with the properties of drug-likeness adapted from Rishton (2003) (37)

Permeability of a molecule or drug through the mycobacterial cell wall is an imperative determinant of its efficacy. So far it has been a challenge to reliably predict the permeability of

potential antimycobacterial agents. MycPermCheck, a bioinformatics tool, has been developed to estimate the likelihood or probability that a small organic entity will permeate the mycobateria's cell envelopes (38). MycPermCheck (Wuezberg, Germany) can be accessed online for free. The program allows entry of the molecule into the database as a CSV (comma-separated values) file format. The input must contain pre-calculated descriptors from a different type of software, for example Schrodinger software (Schrodinger, LLC, new Yourk, USA), in order to generate the CSV file utilized by the program.

The latest edition of software is MycPermCheck 1.1 (39). The results output are the name of the compound, its predicted permeability and whether the specific data (that is, the 5 descriptors calculated and used by the program to process the final probability) falls within the specific limits/ranges (see figure below). This type of permeability prediction provides a more realistic manner of estimating the quality of a drug-likeness, rather than only following more "traditional" cut offs such as Ro3 and Lipinski's rule (40).

Descriptors calculated by Schrodinger software	**Definition**	**Range of optimal values of permeable molecules**
FOSA	Represents hydrophobic part that is solvent-accessible (i.e. saturated carbons and the attached hydrogens)	90.8-272.23Å^2
logP	the calculated octanol/water partition coefficient	2.779-4.479
PISA	The π-interacting part of the solvent accessible surface area	205.16-355.49Å^2

HBA	Number of hydrogen-bond acceptors, calculated as average number of configurations in the molecule	3.750-6.000
glob	Globularity descriptor is the ratio of generic spherical surface to molecule surface NB 1.0=spherical molecule	0.794-0.839

Figure 6: Explanation of the main descriptors used in MycPermCheck and their optimal values for permeability prediction, derived from Merget et al, 2012 (38)

3.7 Role of in-house databases in drug design and development

After identifying and validating a target, the next step in drug development is identifying a **hit** and lead molecules. A 'hit' molecule is one that possesses a desired activity found through screening a large number of molecules or compounds. The hit molecule's activity is then confirmed by retesting (41). Molecular libraries play an important role in pre-clinical stages of design in both academic research and big pharmaceutical industries. Various methods of screening for hits and leads are applied through chemical/molecular databases, according to (42). The main methods are High-throughput screening (HTS), Focused screening, Fragment screening, Physiological screening, NMR screening, Structure-aided drug design and Virtual screening. The creation of chemical compound libraries is what facilitates the screening process to identify hit molecules.

In-house databases allow for pooling of the data contained in them into subsets for more particular screening, this may be in silico (virtual) testing or real "wet screening". These databases are what hold information about the compounds in libraries. The focus can be based upon previous knowledge from literature or preceding drug patents (i.e. patented chemical classes or groups with a known activity) (43). They can be used by companies for compound sourcing and even for analyzing the uniqueness of compounds. This is especially important regarding commercial.

Table 1: Explanation of different of screening methods to identify hits or leads adapted from (42)

Method	Description
HTS (High-throughput screening)	a quick, automated biochemical and pharmacokinetic assay catering best for libraries of less than 1000 chemical entities. The automation means that robots handle each step of the process at different stations, from mixing reagents to the last step of detection output
Focused screens	based on likening compounds which have been identified before as hitting specific biologic targets to new compounds with similar structures. This type of screening, however, can miss new hits although it may be a cheap way to discovering novel compounds
Fragment screen	involves the soaking of tiny molecules into crystals to get compounds with low millimolar (mM) activity that may then be used to build bigger molecules. It is absolutely necessary to have a crystal structure, and the smaller fragments can be conjoined to improve potency
NMR (nuclear magnetic resonance) screening	similar to Fragment screening as it also involves crystallizing small molecules or fragments into targets with known structure to search for hits with a low millimolar (mM) activity that may then be used to build bigger compounds. NMR technology is used to determine the structure of a molecule and the functional groups in it.
Physiological screens	are better suited for small pools of compounds and aim to replicate tissue environment as closely as possible to assess the effects the drug's effects at tissue site instead of cellular or subcellular effects.
Virtual screening	utilizes docking of virtual compound banks against the x-ray of the protein or ligand as the basis of designing new compounds. By determining the orientation of a hit-and-

	target complex, scientists can begin the development of completely new chemical entities based on existing ones, or use the results of a virtual screen as a starting point for a focused screen, without having to use costly large database screens. Pharmacophore searching has also begun to gain acceptance
Structural-aided drug design	employs the use of crystalline structures. Such methods study the macromolecular 3D structure of RNA or proteins (the targets of drugs), to identify key sites and interactions important for biological functions. Big pharmaceutical companies use this method to supplement other screening techniques

The above described screens have contributed greatly to the development of different types of databases for analyzing chemical biology activity of molecular entities as explained by (41). In this particular instance, the type of database is deemed a "chemical interaction database" which gathers, categorizes and organize the results. The database constitutes two parts, the first being a large collection of chemical structures and the second is a set of results from assays that correlate the compounds to certain biological activities. PubChem (19), STITCH (44), ChemDB (45) and Norvatis' Avalon (46) are well known examples of chemical interaction databases

In this project, the determining drug class is pyrazine compounds. The database contains a collection of structures of the in-house pyrazine compounds which were analyzed for activity against *Mycobacterium tuberculosis* H37Rv *(Mtb)* and further, but to a minimal extent, possible antifungal and antibacterial activity. The structures are stored as different formats.

It can be shown how in-house database contribute to multiple scientific fields such as medicinal chemistry, molecular modeling, structural biology and biochemistry that are involved in drug-development stages from the point of synthesizing compounds until the pre-clinical testing phase. (47)

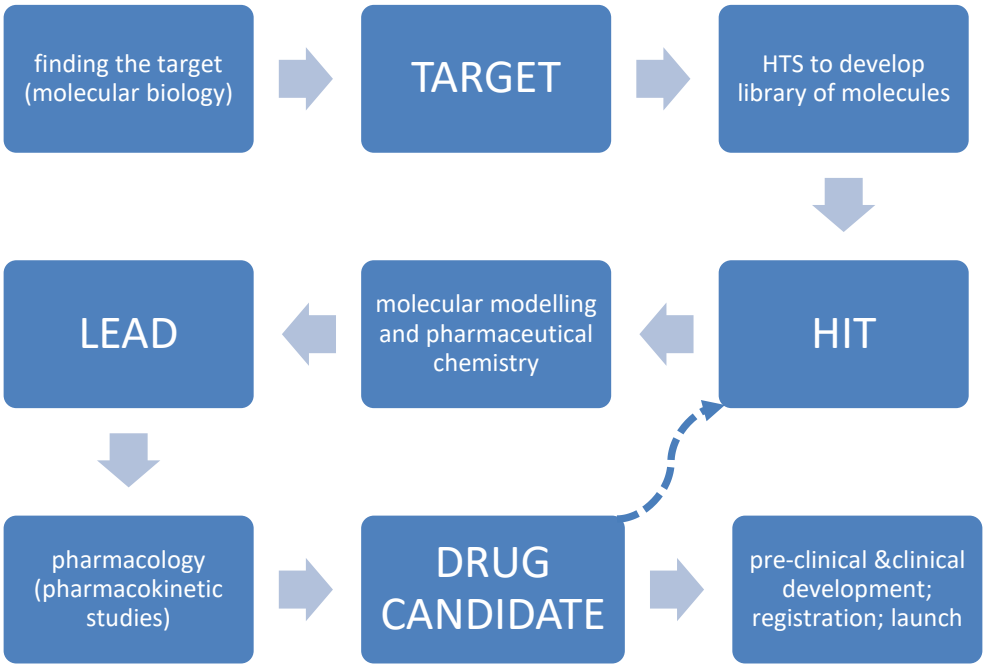


Figure 7: Diagrammatic representation of how in-house database play a role in drug discovery and development, adapted from (47). The dotted line shows how drug candidates can become new hits which then re-enter the cycle.

3. EXPERIMENTAL PART

4.1 Procedure, software and sources used to create and analyze the database

In creation of this database, the first step was to pool together compounds that had already been published in the Faculty of Pharmacy of Charles University, through collaboration of various departments from the year 2000. The source of these compounds was various works that included rigorous thesis carried out under different undergraduate and doctoral studies and scientific experiments by various team members. Works led by M. Dolezal, O. Jandourek, J. Zitko, B. Servusova, L. Semelkova, J. Jampilek & P. Palek who were all affiliated with the faculty. A detailed list of the works used is provided below as Appendix I. Also included were diploma thesis works related to this project from M. Halirova, A. Mindlova,

The data-mining approach proved challenging in certain instances. Generally, a profound problem was there were different experimental methods used to determine MIC or activity. Another concern was that certain compounds are very sensitive to *in vitro* testing conditions like pH and the growth medium used, especially for pyrazinamide. For example, pyrazinamide shows activity when tested at 37°C in low/acidic pH about 5.5 (48), whereas activity at neutral pH is undetectable (49).

During this data mining, mainly pyrazine-like compounds or derivatives with some activity against *Mtb* were selected. The activity had been determined by testing for the MIC (Minimum Inhibitory Concentration) of the compound in micromoles per liter µmol/L, after a series of microdilutions tests against the microbe. Where MIC values were not reported or available, the percentage of inhibition of microbial growth at 6.25µg/mL concentration of the tested antimicrobial was opted for (i.e. TAACF values). The Tuberculosis Antimicrobial Acquisition and Coordinating Facility (TAACF) provides high end pre-clinical screening and efficacy testing services to researchers free of charge. The TAACF was established in 1994 to foster discovery of novel anti-tuberculosis, established by the National Institute of Allergy and Infectious Diseases (NIAID) based in

Maryland, USA. The TAACF program itself ended in 2010 officially, however NIAID still continues to provide its resources worldwide (50) (51).

Besides *Mycobacterium tuberculosis* M37Rv *(Mtb)*, the other mycobacterial species that were tested against the antimicrobial molecules were *Mycobacterium kansasii, Mycobacterium avium* and *Mycobacterium smegmatis.* Furthermore, other bacteria were tested for susceptibility; these included gram-negative bacteria (*Escherichia coli, Klebsiella pneumoniae* and *Pseudomonas aeruginosa*) gram-positive bacteria (*Staphylococcus aureus,* MRSA [methicillin resistant *Staphylococcus aureus*], *Staphylococcus epidermidis* and *Enterococcus faecalis*). Finally, antifungal activity was also undertaken, testing being done on *Candida albicans, Candida tropicalis, Candida krusei, Candida glabrata, Trichosporon beigelii, Aspergillus fumigatus, Absidia corymbifera* and *Trichophyton mentagrophytes*. For bacteria and fungi, the IC80 (80% Inhibited visible growth of microbial population) value after a period of 24 hours was chosen. The data for activity against gram-negative bacteria was of peculiar interest since a few compounds have expressed this activity, hence any compounds showing such properties was recorded in the database.

The computer software that was chosen to create the database was Microsoft® Excel 2016 spreadsheet (Washington, USA) and ChemDraw Professional 15.1 (PerkinElmer®, Massachusetts, USA). Besides statistical applications, the spreadsheet software has various capabilities that optimized data screening and has an add-in ChemDraw function that adds chemical intelligence to Excel. ChemDraw Professional is registered under PerkinElmer®, a technology company that is dedicated to innovative detection, imaging and informatics which are useful in fields such as life science and diagnostics globally.

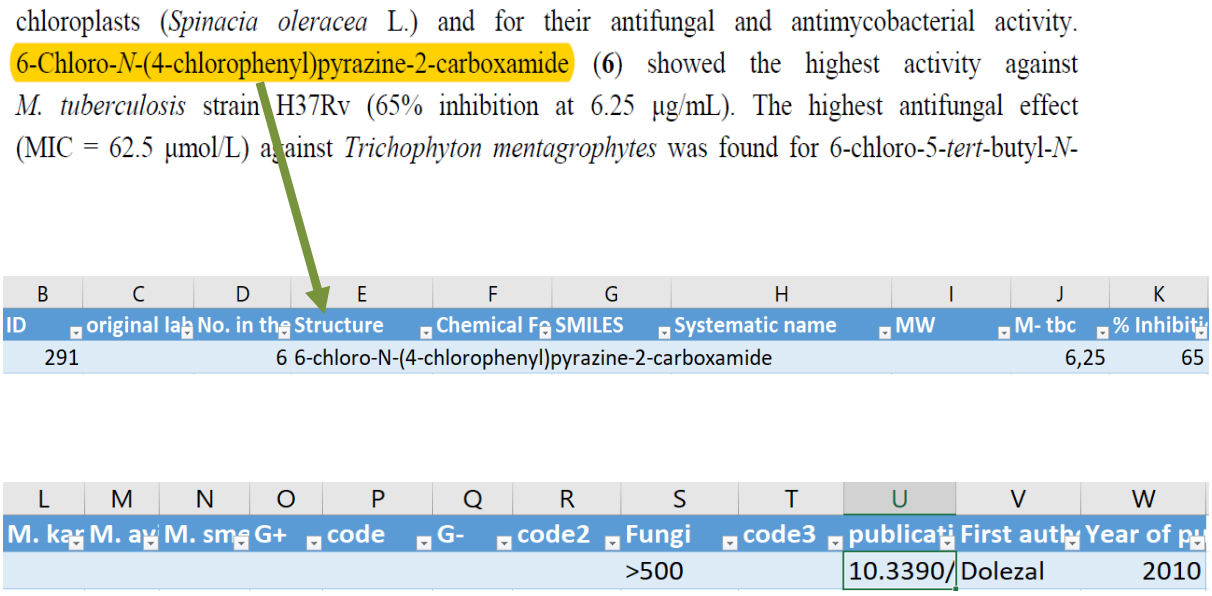
ChemDraw Professional 15.1 (Massachusetts, USA) is a drawing tool that is has capability to effectively and quickly draw molecules, reactions, biological entities and pathways to be utilized in documents; searching databases; to accurately generate names from structures (and vice versa, which was very important for this project) and in predicting properties and spectra of molecules.

In this project it was used to process various types of chemical structure/formula (SMILES and chemical formulae).

ChemDraw for Excel plugin implements the spreadsheet's sorting, analysis and organization functions to manipulate and enrich sets of compounds, scientific data and explore structure-activity relationships. This was very crucial as it is what made it possible to accurately calculate the molecular descriptors for each compound.

4.2 Stepwise illustration of how the database was created (using compound No. 291)

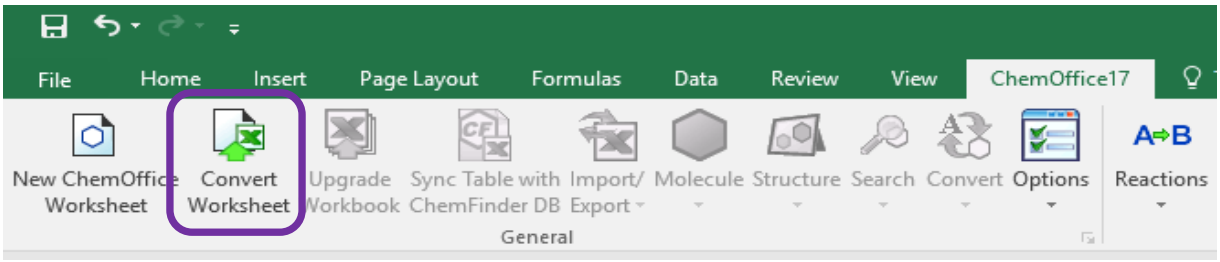
Step 1: Manual extraction of molecule from publication: copy and paste name of compound into spreadsheet. Enter all relevant information to identify the source (i.e. main author, original lab code, number or code in original publication, DOI and year of publication, MIC or activity against *Mtb* and other specified pathogens)


chloroplasts (*Spinacia oleracea* L.) and for their antifungal and antimycobacterial activity. 6-Chloro-*N*-(4-chlorophenyl)pyrazine-2-carboxamide (**6**) showed the highest activity against *M. tuberculosis* strain H37Rv (65% inhibition at 6.25 µg/mL). The highest antifungal effect (MIC = 62.5 µmol/L) against *Trichophyton mentagrophytes* was found for 6-chloro-5-*tert*-butyl-*N*-

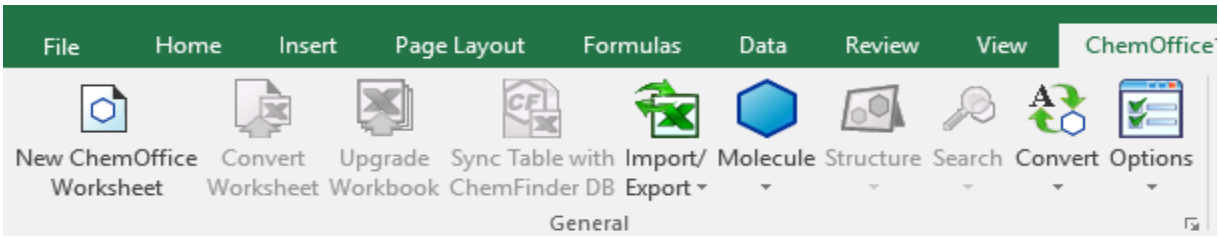
B	C	D	E	F	G	H	I	J	K
ID	original lab	No. in the	Structure	Chemical Fo	SMILES	Systematic name	MW	M- tbc	% Inhibiti
291		6	6-chloro-N-(4-chlorophenyl)pyrazine-2-carboxamide					6,25	65

L	M	N	O	P	Q	R	S	T	U	V	W
M. ka	M. av	M. sm	G+	code	G-	code2	Fungi	code3	publicati	First auth	Year of pu
							>500		10.3390/	Dolezal	2010

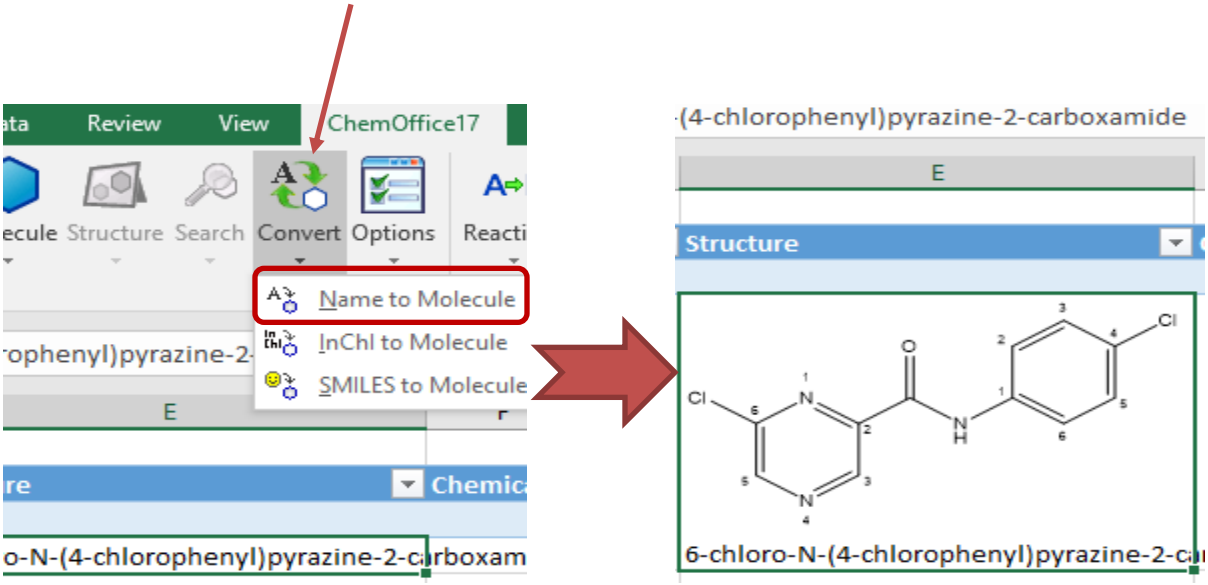
Step 2: Convert spreadsheet to ChemOffice worksheet



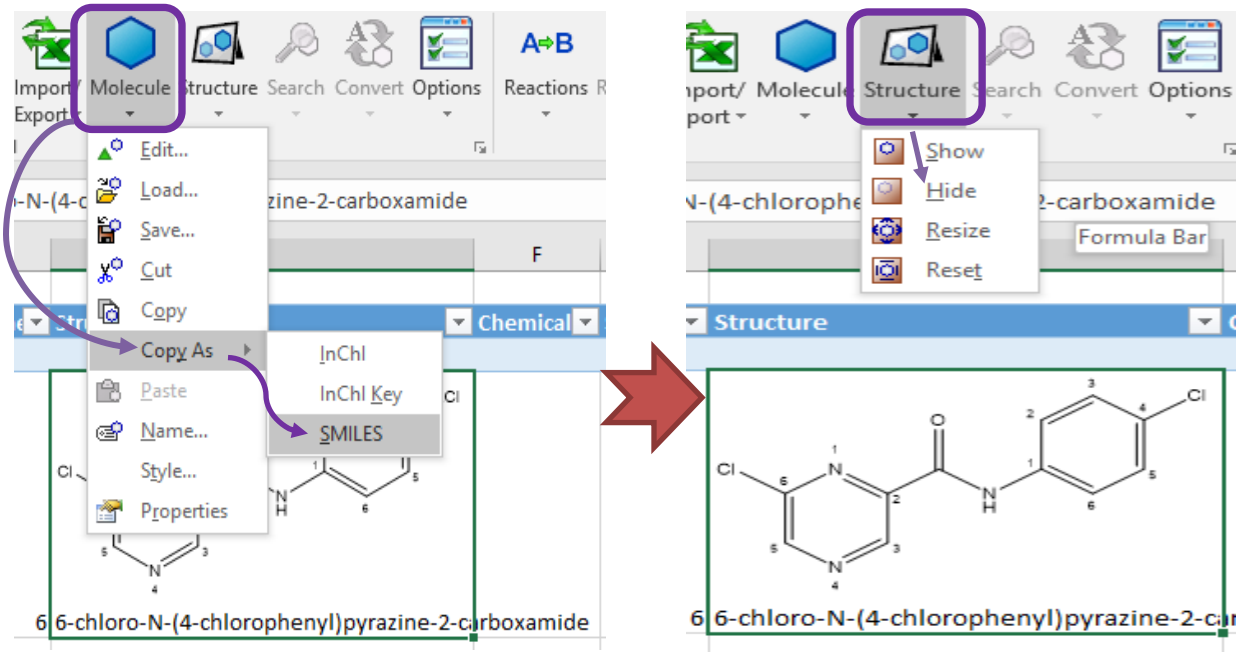
after all compounds were collected, the normal spreadsheet was converted to a ChemDraw worksheet to allow the utilization of ChemOffice functions.



Step 3: convert the systematic name to molecular structures using ChemOffice functions

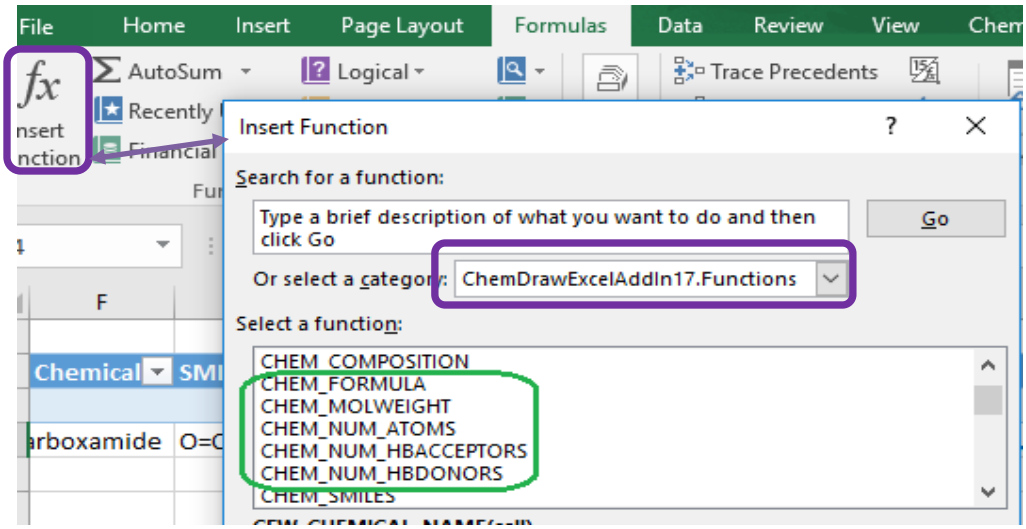


Step 4: The structure is then converted to SMILES using ChemDraw Professional functions. Each compound was also checked for spelling errors by converting each name to a structural formula. The structures (SMILES) were used to regenerate the systematic name of each compound. The structure was then hidden to allow easier manipulation of the file.



Another way to determine SMILES is shown below.

Step 5: following the correction of structures and spellings of names the molecular descriptors were calculated using the built in ChemOffice functions (including but not limited to the chemical formula, SMILES, MW, HBA, HBD, logP, clogP, PSA, and number total of atoms in the molecule).



Here is the complete record of compound 91 from the database

1	ID	ginal lab cc	in the arti	Structure	Chemical I	SMILES	Systemati	MW	M- tbc	% Inhibitic	M. kansas	M. avium	M. smeg
292	291		6	6-chloro-N	C11H7Cl2N	O=C(C1=N	6-chloro-N	268,097	6,25	65			

G+	code	G-	code2	Fungi	code3	publicatio	First autho	Year of pu	Number of	Number o
				>500		10.3390/m	Dolezal	2010	24	17

Molecular Re	CMR	HBA	HBD	Rot_Bond	PSA	LogP	cLogP	LogP (priva	LogS (priva	pKa (privat
65,298	6,6286	3	1	3	53,82	2,0479	2,934692	3,86401	-4,46699	None

After calculating the molecular descriptors, the database was checked for duplications by sorting using the MW. 16 duplicates were removed. If compounds with identical composition and multiple tests results were found the entity with a lower MIC was kept in the database.

5. RESULTS

5.1 Statistics of the database

1. Total number of compounds: 623

2. Active compounds: 91

   A compound is designated as Active if it has a reported MIC of 6.25µg/mL or TAACF value greater than 90%.

   a. Compounds with MIC≤6.25µg/mL: 87
   b. Compounds with TAACF≥90%: 4

3. Inactive compounds: 532

   A compound designated as Inactive if it has a reported MIC of over 6.25µg/mL, TAACF values less than 90% or if results at the time of data collection were still pending

   a. Compounds with MIC>6.25µg/mL: 450
   b. Compounds TAACF<90%: 81
   c. Compounds with no available data: 1

5.2 GRAPHS SHOWING THE IMPORTANT DESCRIPTOR STATISTICS OF THE DATABASE

5.2.1 MW

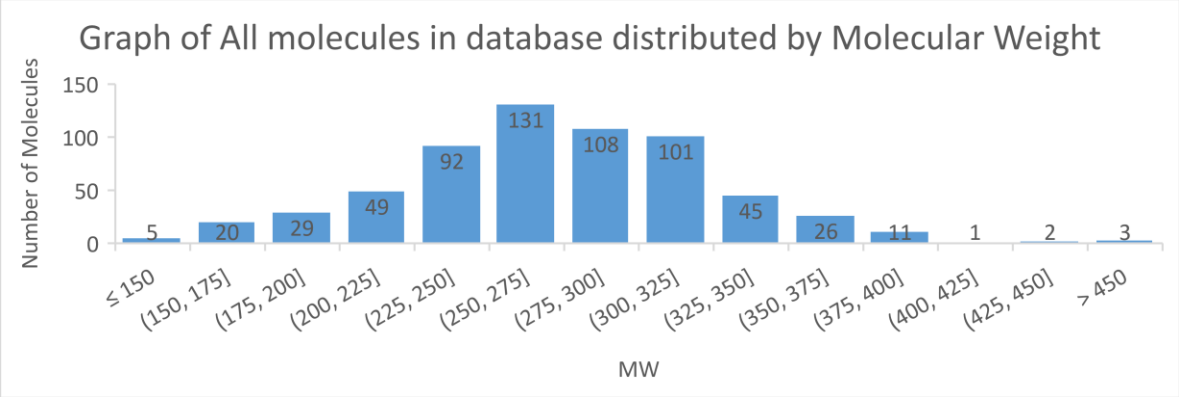


Figure 8: All molecules distributed by MW

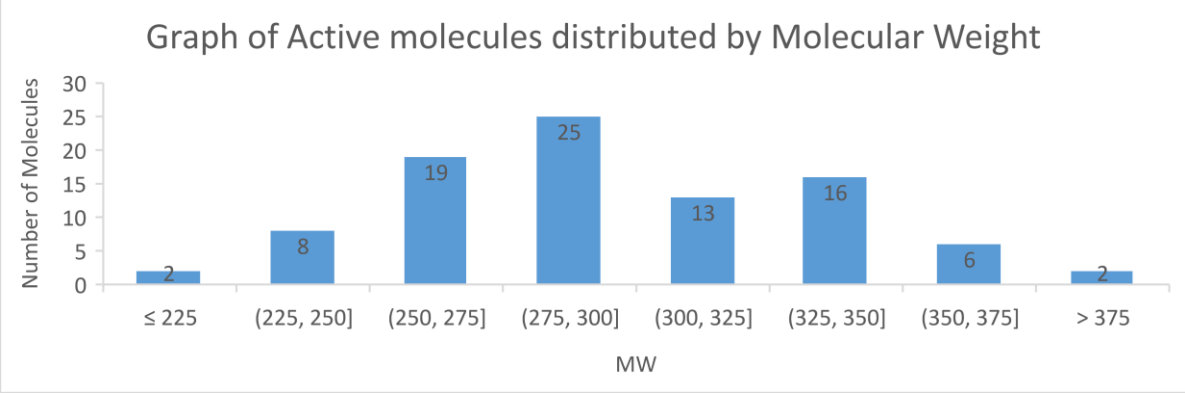


Figure 9: Active molecules distributed by MW

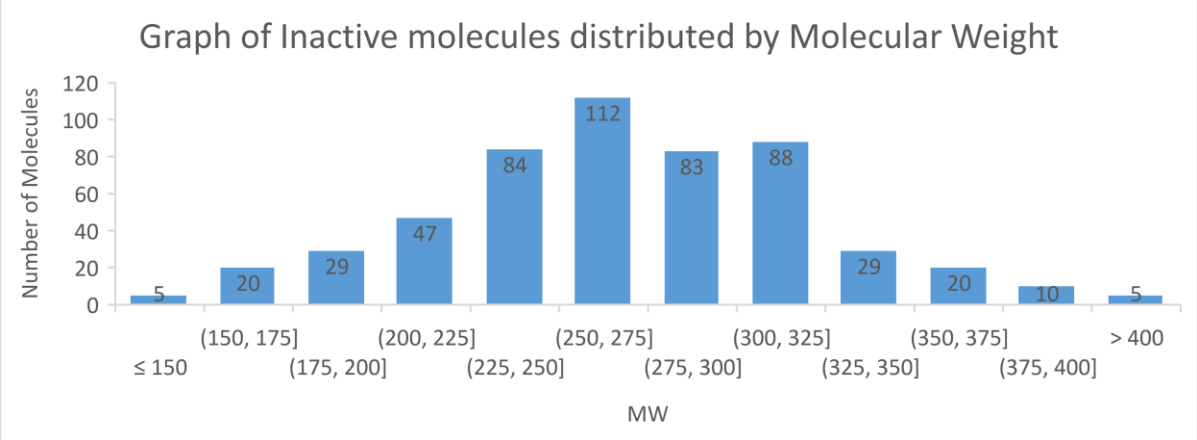


Figure 10: Inactive molecules distributed by MW

5.2.2 HBA

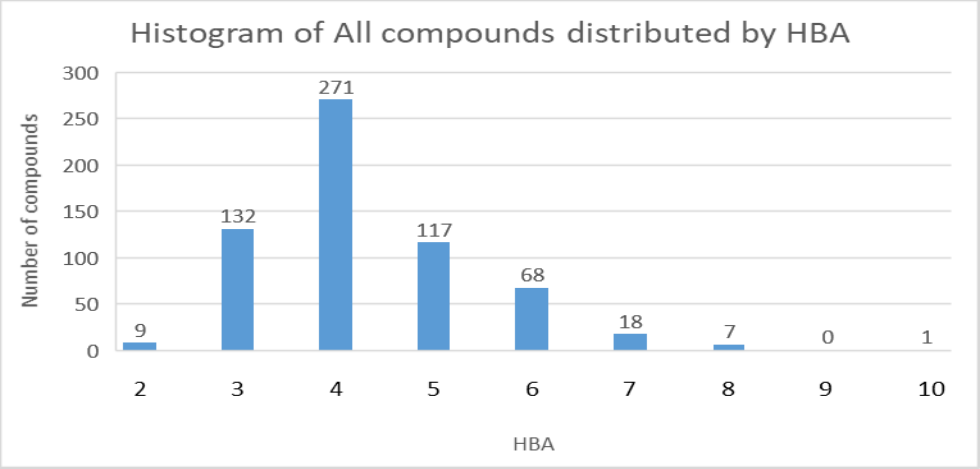


Figure 11: All compounds distributed by number of HBA

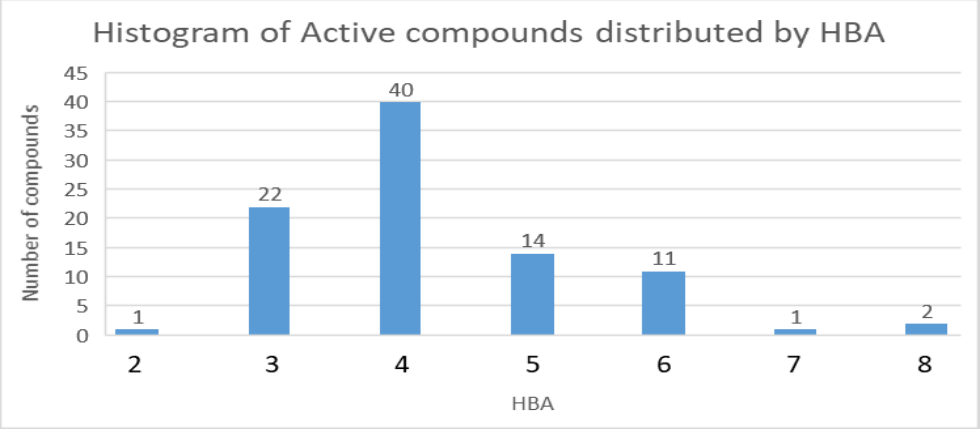


Figure 12: Active compounds distributed by number of HBA

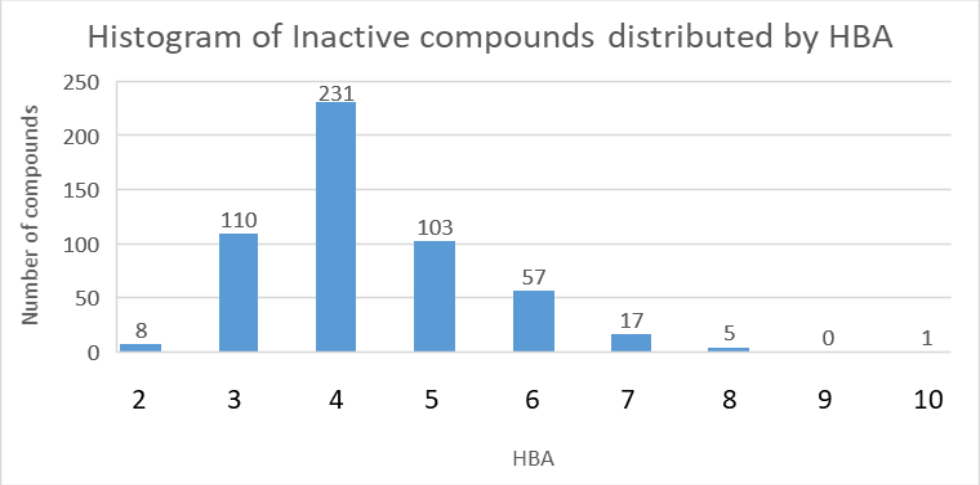


Figure 13: Inactive compounds distributed by number of HBA

5.2.3 HBD

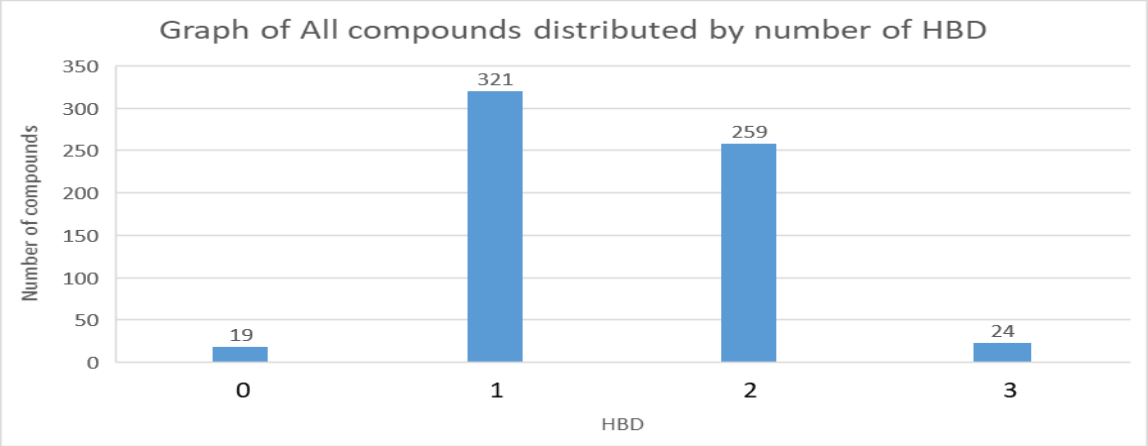


Figure 14: All compounds distributed by number of HBD

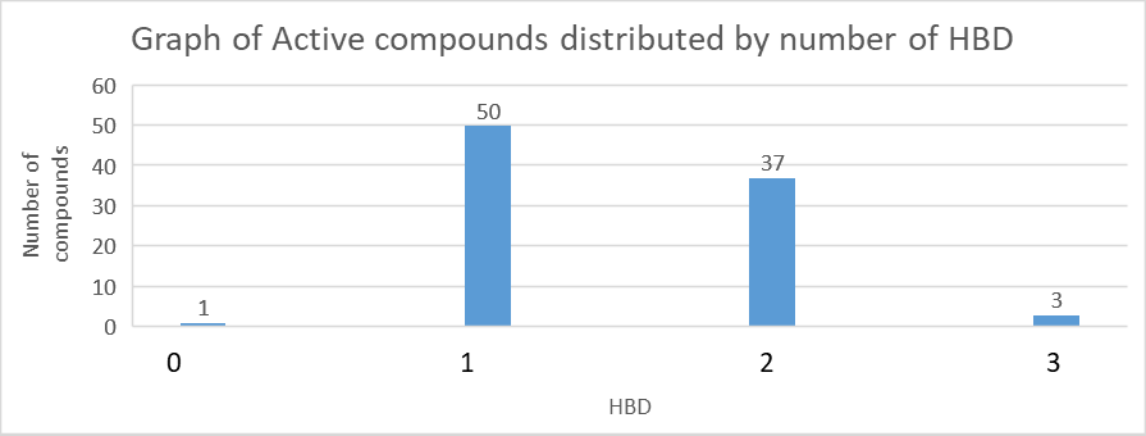


Figure 15: Active compounds distributed by number of HBD

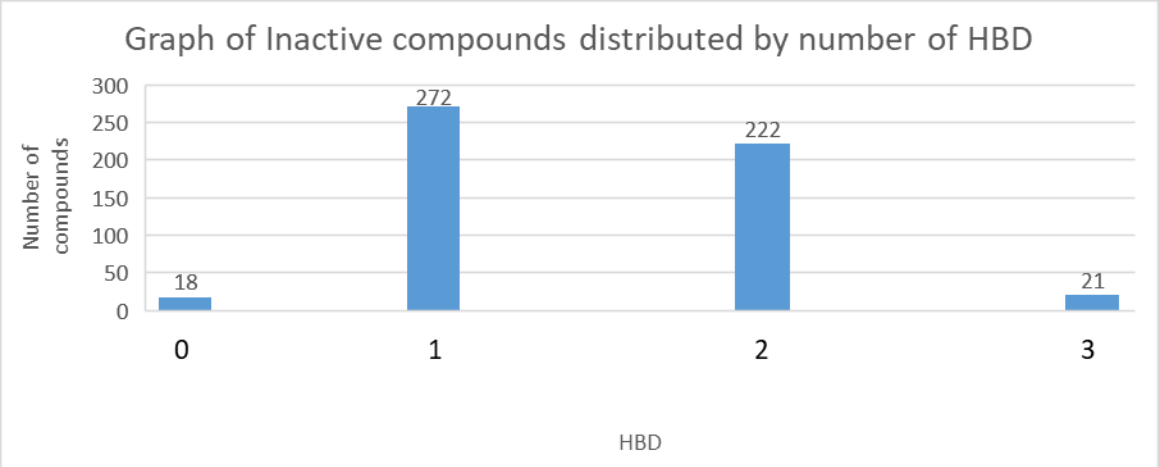


Figure 16: Inactive compounds distributed by number of HBD

5.2.4 NRot

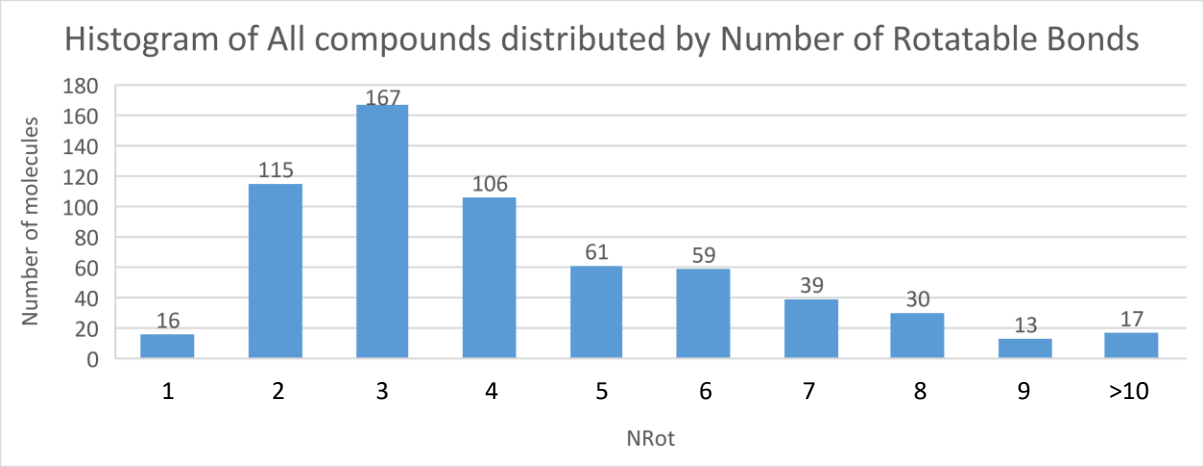


Figure 17: All compounds distributed by NRot

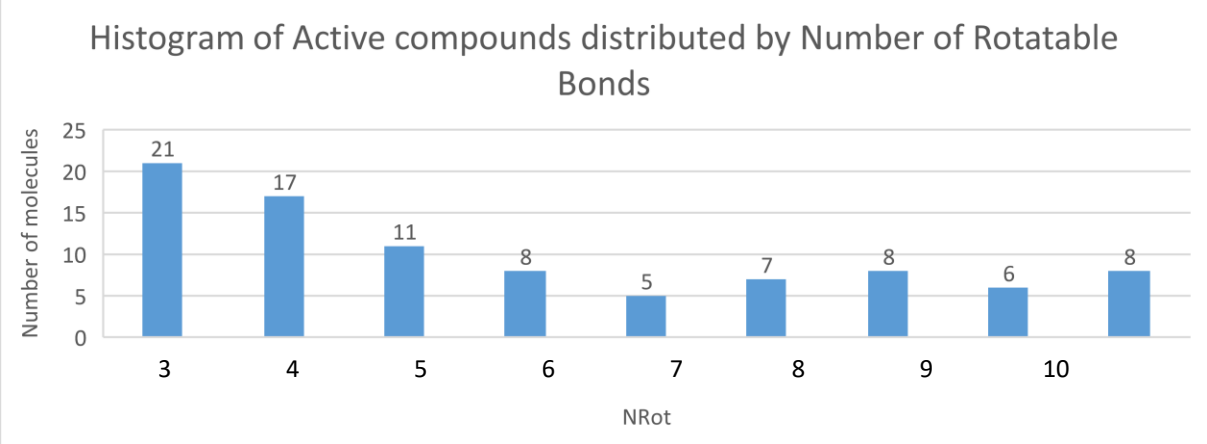


Figure 18: Active compounds distributed by NRot

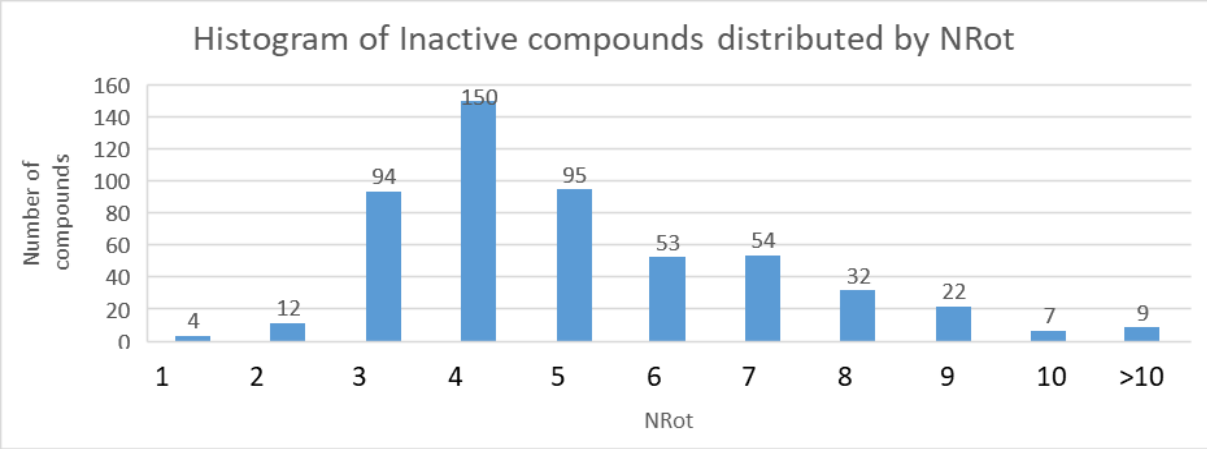


Figure 19: Inactive compounds distributed by NRot

5.2.5 PSA

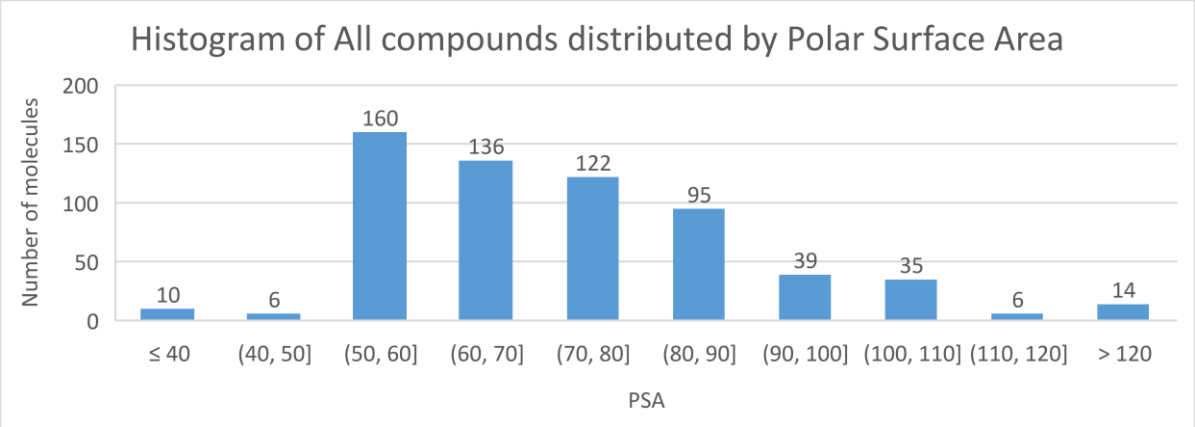


Figure 20: All compounds distributed by PSA

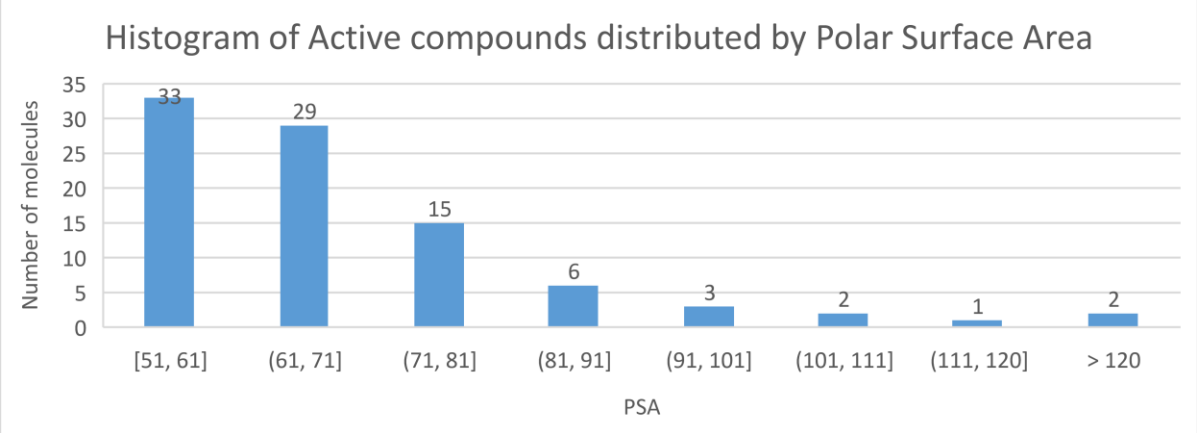


Figure 21: Active compounds distributed by PSA

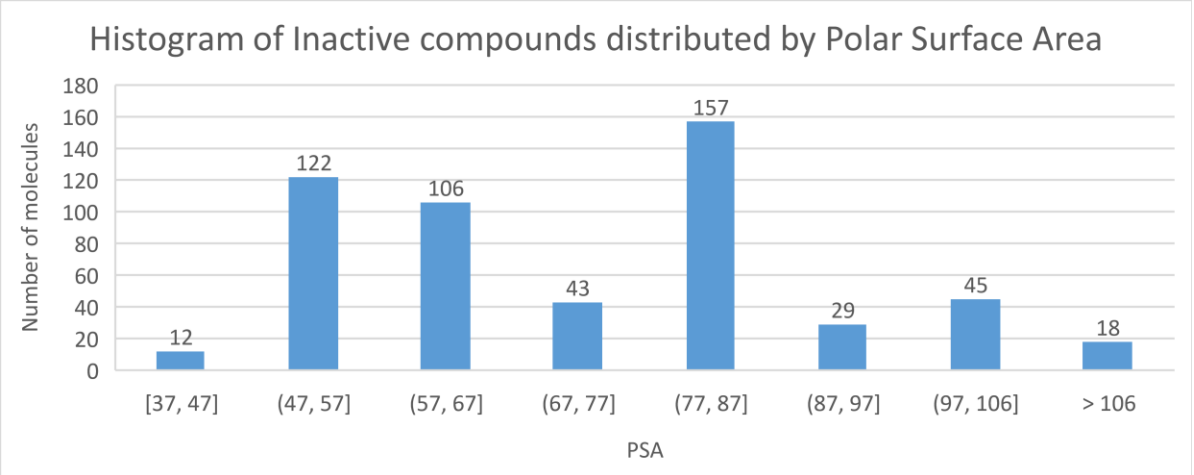


Figure 22: Active compounds distributed by PSA

5.2.6 logP

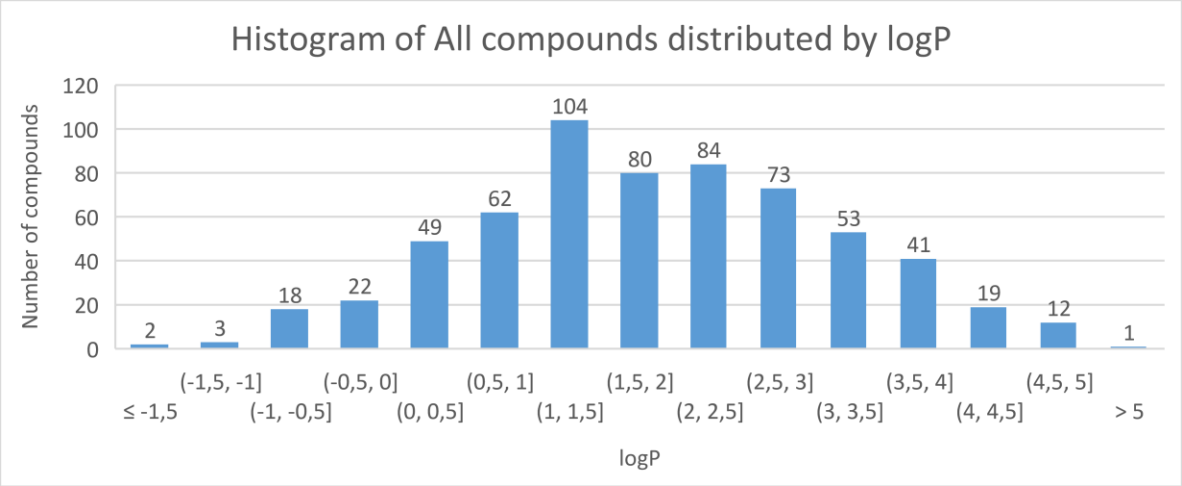


Figure 23: All compounds distributed by logP

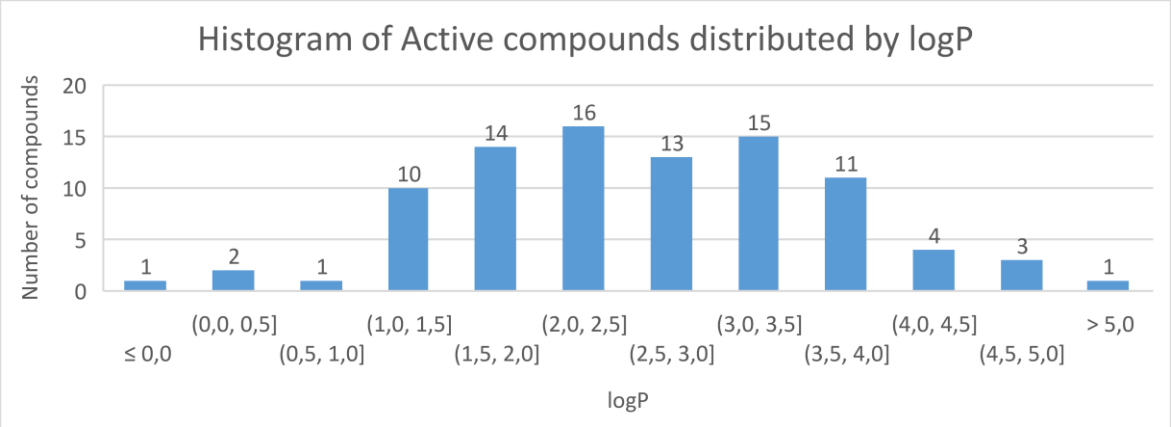


Figure 24: Active compounds distributed by logP

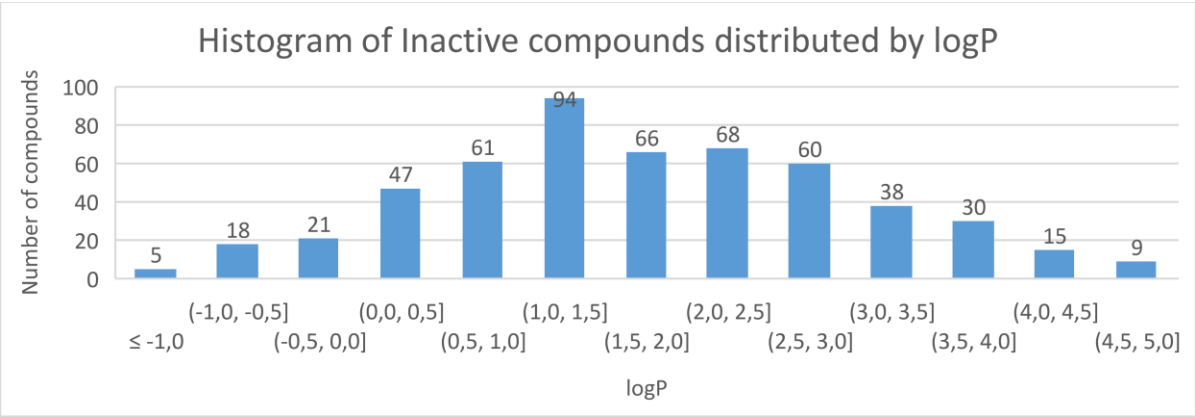


Figure 25: Inactive compounds distributed by logP

5.3 Statistics of Descriptors

Table 2: Statistics of the descriptors

Descriptor	Minimum	Maximum	Mean	Mean SD	Median
MW	**207.28**	**429.69**	**295.86**	**35.10**	**289.22**
	123.16	*463.55*	*269.16*	*53.94*	*268.10*
	123.16	463.55	273.06	52.89	270.34
Number of Atoms	**24**	**56**	**35.10**	**8.95**	**33**
	12	*64*	*31.57*	*6.89*	*30*
	12	64	32.08	7.33	30
Number of Heavy Atoms	**15**	**26**	**20.44**	**2.77**	**20**
	9	*32*	*18.68*	*3.38*	*18*
	9	32	18.94	3.36	19
Molecular Refractivity	**57.78**	**104.62**	**78.40**	**12.44**	**77.20**
	34.97	*113.80*	*70.47*	*13.25*	*69.96*
	34.97	113.80	71.62	13.42	70.37
CMR	**5.92**	**10.65**	**7.85**	**1.24**	**7.53**
	3.50	*11.42*	*7.07*	*1.34*	*7.06*
	3.50	11.42	7.18	1.35	7.07
HBA	**2**	**8**	**4.25**	**1.15**	**4**
	2	*10*	*4.32*	*1.14*	*4*
	2	10	4.31	1.14	4
HBD	**0**	**3**	**1.46**	**0.58**	**1**
	0	*3*	*1.46*	*0.63*	*1*
	0	3	1.46	0.62	1
NRot	**3**	**13**	**6.03**	**2.82**	**5**
	1	*17*	*5.11*	*2.09*	*5*
	1	17	5.24	2.23	5
PSA	**50.74**	**142.93**	**68.28**	**16.77**	**65.85**
	36.75	*163.16*	*74.03*	*19.64*	*74.05*
	36.75	163.16	73.19	19.35	68.42
LogP	**-0.17**	**5.64**	**2.59**	**1.09**	**2.58**
	-2.18	*4.96*	*1.75*	*1.28*	*1.65*
	-2.18	5.46	1.87	1.29	1.75
cLogP	**0.40**	**6.17**	**3.33**	**1.41**	**3.06**
	-0.68	*7.95*	*2.08*	*1.36*	*2.19*
	-1.01	7.95	2.48	1.41	2.30

KEY: Actives | *Inactives* | All Cmpds

Looking at the average molecules, there is a great difference between the size of active compounds compared to inactive ones. The average MW of active compounds is much larger than that of inactive compounds.

With regard to HBA and HBD, there is no significant difference in either the mean or median values of all the compounds (whether active or inactive). This implies that these descriptors are not determinants in our setup.

The logP values of Active compounds are higher than of the inactive compounds by almost a whole unit. Active compounds have a logP of approximately 2.6.

4. DISCUSSION OF RESULTS

According to our results, 91 out of 623 compounds were found to be 'active', that is about 14% of the molecules in the database. 87 had MIC≤6.25µg/mL 4 reported TAACF≥90%.

Table 3: Comparison of average molecule from whole database to the limits/range of Ro3 and Lipinski's Rule

Descriptor	MW	HBA	HBD	NRot	PSA	logP
Mean molecule	273.06	4.31	1.46	5.24	73.19	1.87
Median molecule	270.34	4	1	5	65.85	1.75
Ro3	<300	<3	<3	<3	<60	<3
Ro5	<500	<10	<5	<10	<140	<5

Regarding the Ro3, the average molecule produced slightly exceeds the limit of number of HBA and the PSA. In relation to the Ro5, the molecules are within all limits. Therefore it can be concluded that drug-like compounds have been produced more than lead-like compounds. Based on Lipinski's rule optimal molecules are being produced.

Table 4: Comparison of average molecule of whole database to the limits/range of MycPermCheck

Descriptor	FOSA	logP	PISA	HBA	Globularity
Mean molecule	null	1.87	null	4. 306581	null
Median molecule	null	1.75	null	4	null
MycPermCheck	90.8-272.23	2.779-4.479	205.16-355.49	3.750-6.000	0.794-0.839

NB 'null' means the values were not obtained during preparation of the database

Two descriptors calculated from the database that relate to MycPermCheck are logP and HBA. These values fall with the defined limits. Thus, based only on these two descriptors it can be said that the average molecule can successfully permeate a mycobacterial cell.

5.4 Outputs of the database

This database has proven useful in pooling together data from an extended period time as previously there was no actual electronic repository of synthesized or tested entities in the Department of Pharmaceutical Chemistry. Creating an electronic library will allow an easier, less laborious way to find those molecules from the cited team leaders that are useful (together with their related data in one place).

This database has applied the cheminformatics approach of keeping together and retrieving all experimental data regarding a compound and heterogeneous entities electronically (52). In future this could be advantageous in keeping track of which exact compounds were produced to avoid redundancy of recreating compounds that had already been made before or that proved to be futile. Also large datasets within the database can sub-grouped to allow more critical analysis and evaluation or screening (53).

The database can further be extrapolated to other chemical software. This would allow it to be applied in modelling studies in computer aided drug design (CADD) especially ligand based drug design (LBDD) (54). LBDD focuses on known ligands for a target so as to establish a relationship between physicochemical properties of the compound and its activities (i.e. structure-activity relationship). This can further be implemented in pharmacophore and/or similarity searching, whereby the part of a molecule responsible for pharmacological action is identified or analogous chemicals that are physicochemically similar as input compound can be found (55). LBDD techniques are used without the 3-dimensional information of a receptor, only the information about how molecules bind with the target is relied upon. 3D QSAR (3D quantitative structure-activity relationships) and pharmacophore modeling are critical methods and commonly used in LBDD.

On the other hand, SBDD (Structure based drug design) studies the 3D structure of biological macromolecular targets with the intent of creating their inhibitors (56). The structure can be obtained through NMR spectroscopy or x-ray crystallography. The most commonly employed

method of SBDD is SILCS (Site Identification by Ligand Competitive Saturation) which uses molecular dynamics to characterize binding patterns of 3-dimensional functional groups by simulating the target molecule in an aqueous solution of different fragments. The in-house database that was developed could be used in such a manner (57).

Quantitative structure-activity relationships immensely contribute to hit-to-lead optimization (42). Thus, a fundamental application of this database could in hit-lead-optimization. This could be achieved by recommending possible alterations to molecular structure and by reducing the number of molecules that are to be synthesized practically. (58). Optimization aims to improve the activity, selectivity and physicochemical properties of a hit by chemical modification.

A hit is a compound with a certain desired activity found through screening large numbers of compounds or molecules. A lead, however, is a prototype compound or molecule that possesses bioactivity against a known target with valid pharmacokinetic properties and structure-activity relationship to a series of other compounds (42).

A single compound (ID number 376) was found to possess activity against gram-negative *P. aeruginosa.*

5. CONCLUSION

A digital library of pyrazine-derivatives with antimicrobial activity, prominently focusing on activity against *Mycobacterium tuberculosis* H37Rv *(Mtb)*, was successfully created. The compounds were further successfully analyzed in order to elucidate their quantitative structure-activity relationship.

Evaluation of the results based on specific scientific rules showed that the average molecule produced is more drug-like than it is lead-like. This means that the molecular entities that have been produced thus far are more close to known anti-tuberculosis drugs than novel prototypes.

APPENDIX I

First Author	Title	DOI/type of work/publication	year of publication
M. Dolezal	Substituted N-Phenylpyrazine-2-carboxamides: Synthesis and Antimycobacterial Evaluation	10.3390/molecules14104180	2009
M. Dolezal	Synthesis, Antimycobacterial, Antifungal and Photosynthesis-Inhibiting Activity of Chlorinated N-phenylpyrazine-2-carboxamides	10.3390/molecules15128567	2010
M. Dolezal	Substituted 5-aroylpyrazine-2-carboxylic acid derivatives: synthesis and biological activity	Il Farmaco 58 (2003) 1105-1111	2003
M. Dolezal	Substituted Pyrazinecarboxamides: Synthesis and Biological Evaluation	10.3390/11040242	2006
M. Halirova	5-alkylpyrazine-2-carboxylic acid derivatives as potential antiinfectives	DT (Charles University Faculty of Pharmacy, Hradec Kralove)	2017
J. Jampilek	Antimicrobial Evaluation of Some Arylsulfanylpyrazinecarboxylic Acid Derivatives	10.2174/157340607780620635	2007
O. Jandourek	New Potentially Active Pyrazinamide Derivatives Synthesized Under Microwave Conditions	10.3390/molecules19079318	2014
O. Jandourek	Synthesis of Novel Pyrazinamide Derivatives	10.3390/molecules22020223	2017

	Based on 3-Chloropyrazine-2-carboxamide and Their Antimicrobial Evaluation		
A. Mindlova	Pyrazine derivatives as potential drugs	DT (Charles University Faculty of Pharmacy, Hradec Kralove)	2016
P. Palek	Synthesis, Antimycobacterial and Antifungal Evaluation of 3-Arylaminopyrazine-2,5-dicarbonitriles	10.1002/ardp.200700119	2008
L. Semelkova	Synthesis and Biological Evaluation of N-Alkyl-3-(alkylamino)-pyrazine-2-carboxamides	10.3390/molecules20058687	2015
L. Semelkova	n/a	lab	n/a
L. Semelkova	3-Substituted N-Benzylpyrazine-2-carboxamide Derivatives: Synthesis, Antimycobacterial and Antibacterial Evaluation	10.3390/molecules22030495	2017
L. Semelkova	Design, Synthesis, Antimycobacterial Evaluation, and In Silico Studies of 3-(Phenylcarbamoyl)-pyrazine-2-carboxylic Acids	10.3390/molecules22091491	2017
B. Servusova	Alkylamino derivatives of pyrazinamide: Synthesis and antimycobacterial evaluation	10.1016/j.bmcl.2013.12.054	2014
B. Servusova	Synthesis and Antimicrobial Evaluation of 6-Alkylamino-N-phenylpyrazine-2-carboxamides	10.1111/cbdd.12536	2015

B. Servusova	Synthesis and antimycobacterial evaluation of N-substituted 5-chloropyrazine-2-carboxamides	10.1016/j.bmcl.2013.04.021	2013
B. Servusova-Vanaskova	Alkylamino derivatives of N-benzylpyrazine-2-carboxamide: synthesis and antimycobacterial evaluation	10.1039/c5md00178a	2015
O. Valasek	Pyrazine derivatives as potential drugs	DT (Charles University Faculty of Pharmacy, Hradec Kralove)	2016
J. Zitko	Synthesis and antimycobacterial evaluation of N-substituted 3-aminopyrazine-2,5-dicarbonitriles	10.1016/j.bmcl.2011.12.129	2012
J. Zitko	n/a	lab	n/a
J. Zitko	Synthesis and antimycobacterial properties of N-substituted 6-amino-5-cyanopyrazine-2-carboxamides	10.1016/j.bmc.2010.12.054	2011
J. Zitko	Synthesis and antimycobacterial evaluation of 5-alkylamino-N-phenylpyrazine-2-carboxamides	10.1016/j.bmc.2014.11.014	2015
J. Zitko	Synthesis and anti-infective evaluation of 5-amino-N-phenylpyrazine-2-carboxamides	Čes. slov. Farm. 2015	2015
J. Zitko	Synthesis and antimycobacterial evaluation of pyrazinamide derivatives with benzylamino substitution	10.1016/j.bmcl.2012.11.052	2013
J. Zitko	Synthesis, Antimycobacterial Activity and In Vitro Cytotoxicity of	10.3390/molecules181214807	2013

	5-Chloro-N-phenylpyrazine-2-carboxamides		
J. Zitko	Design, synthesis and antimycobacterial activity of hybrid molecules combining pyrazinamide with 4-phenylthiazol-2-amine scaffold	10.1039/C8MD00056E	2018

Works Cited

1. World Health Organization. *Global tuberculosis report 2017.* [Online] [Cited: August 26, 2018.] http://apps.who.int/iris/bitstream/handle/10665/259366/9789241565516-eng.pdf;jsessionid=D761B52F57452D4F21D285A9E70AE04B?sequence=1.

2. *A Note on Derivatives of Isoniazid, Rifampicin, and Pyrazinamide Showing Activity Against Resistant Mycobacterium tuberculosis.* Unissa, A. N., Hanna, L. E. and Soumaya, Swaminathan. 4, s.l. : John Wiley & Sons, 2015, Chemical Biology & Drug Design, Vol. 87, pp. 537-550.

3. *Synthesis of Novel Pyrazinamide Derivatives Based on 3-Chloropyrazine-2-carboxamide and Their Antimicrobial Evaluation.* Jandourek, O., et al. 223, 2017, MDPI, Vol. 22.

4. *Pyrazinamide inhibits trans-translation in Mycobacterium tuberculosis: a potential mechanism for shortening the duration of tuberculosis chemotherapy.* Shi, W., Zhang, X., Jiang, X., Ruan, H., Barry, C. E., Wang, H., Zhang, Y., Zhang, W. 6049, s.l. : HHS Public Access, 2011, Science, Vol. 333, pp. 1630-1632.

5. *Pyrazinamide - Pharmaceutical, biochemical and pharmacological properties and reappraisal of its role in the chemotherapy of tuberculosis.* Momekov, G., et al. 1, 2014, Pharmacia, Vol. 61, pp. 38-67.

6. Karthikeyan, Muthukumarasamy and Viyas, Renu. *Practical Chemoinformatics.* s.l. : Springer India, 2014. p. 533.

7. Winiwarter, S., et al. *Use of Molecular Descriptors for Absorption, Distribution, Metabolism, and Excretion Predictions.* s.l. : Elsevier, 2007. pp. 531-544.

8. Smith, Dennis, et al. *Pharmacokinetics and Metabolism in Drug Design.* Singapore : Wiley-VCH Verlag GmbH, 2012. Online ISBN:9783527600212.

9. Gragos, H., Marcou, G. and Varnek, A. [Online] [Cited: April 2, 2018.] http://bigchem.eu/sites/default/files/School1_Horvat.pdf.

10. *A refined model for prediction of hydrogen bond acidity and basicity parameters from quantum chemical molecular descriptors.* Devereux, M., Popelier, P. L. and M., McLay. I. 2009, s.l. : Royal Society of Chemistry, 2009, Physical Chemistry Chemical Physics, Vol. 10, pp. 1595-1603.

11. Belford, Robert. Introduction to Connection Tables. *LibreTexts.* [Online] MindTouch. [Cited: April 5, 2018.] https://chem.libretexts.org/LibreTexts/University_of_Arkansas_Little_Rock/Spring_2017_Cheminfomatics_CHEM_4399%2F%2F5399/2.2%3A_Chemical_Representations_on_Computer%3A_Part_II/2.2.1._Introduction_to_Connection_Tables.

12. *Three-Dimensional Quantitative Structure-Activity Relationships. 2. Conformational Mimicry and Topographical Similarity of Flexible Molecules.* Labanowski, J, et al. 4, Weinheim : WILEY-VCH Verlag GmbH & Co. KGaA, 1986, Molecular Informatics, Vol. 5, pp. 138-152.

13. *Development of chemical markup language (CML) as a system for handling complex chemical content.* Murray-Rust, P., Rzepa, H. S. and Wright, M. 2001, New Journal of Chemistry, pp. 618-634.

14. Westbrook, J., et al. PDBML: the representation of archival macromolecular structure data in XML. *Bioinformatics Volume 21, Issue 7.* April 01, 2005, pp. 988-992.

15. *Computational Methods in Drug Discovery.* Sliwoski, G., et al. 2014, Pharmacological Reviews, pp. 334-395.

16. *Automated Identification and Classification of Stereochemistry: Chirality and Double Bond Stereoisomerism.* Teixeira, A. L, Leal, J. P. and Falcao, A. O. 2013, CoRR.

17. James, Craig A. OpenSMILES specification. *OPENSMILES.* [Online] 2017. [Cited: August 12, 2018.] http://opensmiles.org/opensmiles.html.

18. *InChI, the IUPAC International Chemical Identifier.* Heller, S. R., et al. 2015, Journal of Cheminformatics.

19. Information, National Center for Biotechnology. PubChem. *PubChem.* [Online] [Cited: August 16, 2018.] https://pubchem.ncbi.nlm.nih.gov/.

20. Chemical Identifier Resolver. [Online] 2009. [Cited: August 21, 2018.] https://cactus.nci.nih.gov/chemical/structure.

21. [Online] Royal Society of Chemistry, 2015. [Cited: August 21, 2018.] http://www.chemspider.com/.

22. UniChem. [Online] European Bioinformatics Institute. [Cited: April 08, 2018.] https://www.ebi.ac.uk/unichem/.

23. *Quantitative structure-activity relationship studies on nitrofuranyl antitubercular agents.* Hevener, K. E., et al. 17, s.l. : NIH Public Access, September 1, 2008, Bioorganic Medicinal Chemistry, Vol. 16, pp. 8042-8053.

24. *Molecular properties that influence the oral bioavailability of drug candidates.* Veber, D. F., et al. 12, Pennsylvania : American Chemical Society, 2002, Journal of Medicinal Chemistry, Vol. 45, pp. 2615-2623.

25. *Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings.* Lipinski, C. A., et al. 1997, s.l. : Elsevier Science, 1996, Advanced Drug delivery Reviews, Vol. 23, pp. 3-25. PII of original article: S0169-409X(96)00423-1.

26. *Understanding drug-likeness.* Ursu, O., et al. 1, s.l. : John Wiley & Sons Ltd, 2011, Computational Molecular Science, Vol. 2011, pp. 760-781.

27. Schneider, G. *Prediction of Drug-like properties.* Texas : Landes Bioscience, 2000-2013. Madame Curie Bioscience Database [Internet].

28. *Identification of New Molecular Entities (NMEs) as Potential Leads against Tuberculosis from Open Source Compound Repository.* Kotapalli, S. S., et al. 12, 2015, PLoS ONE, Vol. 10.

29. *Structure−Brain Exposure Relationships.* Hitchcock, S. A. and Pennington, L. D. 26, Thousand Oaks : American Chemical Society, 2006, Journal of Medicinal Chemistry, Vol. 49, pp. 7559-7583.

30. *Polar molecular surface properties predict the intestinal absorption of drugs in humans.* Palk, K., et al. 5, Uppsala : Springer Science, 1997, Pharmaceutical Research, Vol. 14, pp. 568-571.

31. World Health Organization. *Lexicon of alcohol and drug terms published by the World Health Organization.* [Online] [Cited: April 16, 2018.] http://www.who.int/substance_abuse/terminology/who_lexicon/en/.

32. *A 'Rule of Three' for fragment-based lead discovery?* Congreve, M., et al. 19, Cambridge : Elsevier Science Ltd, October 1, 2003, Drug Discovery Today, Vol. 8, pp. 876-877.

33. *Molecular recognition: the fragment approach in lead generation.* Fattori, D. 5, s.l. : Elsevier Science Ltd, March 1, 2004, Drug Discovery Today, Vol. 9, pp. 229-238.

34. *Calculation of molecular lipophilicity: State-of-the-art and comparison of log P methods on more than 96,000 compounds.* Mannhold, R., et al. 2009, s.l. : Wiley InterScience , 2008, Journal of Pharmaceutical Sciences, Vol. 98, pp. 861-893.

35. *Is There a Difference between Leads and Drugs? A Historical Perspective.* Oprea, T. I., et al. 3-6, 2001, Journal of Chemical Information and Computer Science, Vol. 41, pp. 1308-1315. PubMed ID: 11604031 .

36. *The Design of Leadlike Combinatorial Libraries.* Teague, S. J., et al. 24, Weinheim : WILEY-VCH Verlag GmbH, 1999, Vol. 38, pp. 3955-3957.

37. *Nonleadlikeness and leadlikeness in biochemical screening.* Rishton, G. M. 2, s.l. : Elsevier Science Ltd, January 15, 2003, Drug Discovery Today, Vol. 8, p. 94.

38. *MycPermCheck: The Mycobacterium tuberculosis permeability prediction tool for small molecules.* Merget, B., et al. 1, Oxford : Oxford Univeristy Press, 2012, Bioinformatics, Vol. 29, pp. 62-68.

39. Merget, Benjamin. MycPermCheck. *MycPermCheck.* [Online] 1.1, Stroiffer Lab, 2012. [Cited: August 21, 2018.] http://www.mycpermcheck.aksotriffer.pharmazie.uni-wuerzburg.de/.

40. *Quantifying the chemical beauty of drugs.* Bickerton, R., et al. 2, s.l. : Macmillan Publishers Ltd, 2012, Nature Chemistry, Vol. 4, pp. 90-98.

41. *AN OVERVIEW OF COMPUTATIONAL LIFE SCIENCE DATABASES & EXCHANGE FORMATS OF RELEVANCE TO CHEMICAL BIOLOGY RESEARCH.* Hall, A. S., et al. 2013, Combinatorial Chemistry & High Throuput Screening, pp. 189-198.

42. *Principles of early drug discovery.* Hughes, J., et al. 2011, British Journal of Pharmacology, pp. 239-1249.

43. *Selecting, Acquiring, and Using Small Molecule Libraries for High-Throughput Screenin.* Dandapani, S., et al. s.l. : HHS Public Access, 2012, Current Protocols in Chemical Biology, Vol. 4, pp. 177-191.

44. STITCH. *STITCH.* [Online] Institute of Genomics and Bioinformatics. [Cited: August 20, 2018.] http://stitch.embl.de/.

45. ChemDB Chemoinformatics Portal. [Online] Institute of Genomics and Bioinformatics, University of California, 2012. [Cited: August 20, 2018.] http://cdb.ics.uci.edu/.

46. Ertl, Peter. [Online] [Cited: April 12, 2018.] http://infochim.u-strasbg.fr/CS3/program/material/Ertl.pdf.

47. *OSIRIS, an Entirely in-House Developed Drug Discovery Informatics System.* Sander, T., et al. 2, Allschwil : American Chemical Society, 2009, Journal of Chemical Information and Modeling, Vol. 49, pp. 232-246.

48. *Synthesis and antimycobacterial properties of N-substituted.* Zitko, J., et al. s.l. : Elsevier Ltd, 2011, Journal of Bioorganic and Medicinal Chemistry, Vol. 19, pp. 1471-1476.

49. *Pyrazinamide Is Active against Mycobacterium tuberculosis Cultures at Neutral pH and Low Temperature.* den Hertog, A. L., et al. 8, 2016, Antimicrobial Agents and Chemotherapy, Vol. 60, pp. 4956-4960.

50. National Institue of Allergy and Infectious Disease. *National Institue of Health (NIH).* [Online] [Cited: August 24, 2018.] https://www.niaid.nih.gov/diseases-conditions/tuberculosis-tb.

51. *Antituberculosis Activity of the Molecular Libraries Screening Center Network Library.* Maddry, J. A., et al. 5, Edinburgh : Elsevier Ltd, September 2009, Tuberculosis (Edinburgh), Vol. 89, pp. 354-363.

52. *A Study on Cheminformatics and its Applications on Modern Drug Discovery.* Firdaus Begam, B. and Shatheesh Kumar, J. 2012, Coimbatore : Elsevier Ltd, 2012, Vol. 38, pp. 1264-1275.

53. *AN OVERVIEW OF COMPUTATIONAL LIFE SCIENCE DATABASES & EXCHANGE FORMATS OF RELEVANCE TO CHEMICAL BIOLOGY RESEARCH.* Hall, A. S., et al. 3, 2013, Combinatorial Chemistry & High Throughput Screening, Vol. 16, pp. 189-198.

54. *Computer-Aided Drug Design Methods.* Yu, W. & MacKerell, A. D. 2017, Methods in Molecular Biology, pp. 85-106.

55. *Computer-Aided Drug Design Applied to Marine Drug Discovery: Meridianins as Alzheimer's Disease Therapeutic Agents.* Llorach-Prares, L., et al. 12, 2017, Marine Drugs, Vol. 15.

56. *Structure and Ligand Based Drug Design Strategies in the Development of Novel 5-LOX Inhibitors.* Aparoy, P., Kumar Reddy, K. and Reddanna, P. 22, s.l. : Bentham Science Publishers, Current Medicinal Chemistry, Vol. 19, pp. 3763-3778.

57. *Computational methods in drug discovery.* Leelananda, S. P. and Lindert, S. 2016, Beilstein Journal of Organic Chemistry, Vol. 12, pp. 2694-2718.

58. *Chemistry-Driven Hit-To-Lead Optimization Guided by Structure-Based Approaches.* Hoffer, L., et al. Weinheim : Wiley-VCH Verlag GmbH & Co. KGaA, 2018, Molecular Informatics, Vol. 37.

59. Lexicon of alcohol and drug terms published by the World Health Organization. *World Health Organization.* [Online] [Cited: April 16, 2018.] http://www.who.int/substance_abuse/terminology/who_lexicon/en/.

60. *An overview of heavy-atom derivatization of protein crystals.* Pike, A. C. W., Garman, E. F., Krojer, T., von Delft, F., & Carpenter, E. P. Pt 3, 2016, Acta Crystallographica. Section D, Structural Biology, Vol. 72, pp. 303-318.