Identification of active targets in bioassays of Ivermectin similar compounds

Sacha LEVY, Arthur MOREAU, Anaïs NGUYEN-GOUMENT, Nicolas TONNEAU

Abstract:

Phenotypic-based screening strategy is getting popular in pharmaceutical research to jumpstart drug discovery effort with already active compound on disease related phenotype. However, these screening techniques use no "a priori" knowledge on the mechanism of action i.e. the protein targets the drug candidate interfering with. As such there is a need for an *in silico* target identification protocol from which a target profile of an cell-active compound can be deduced from. This approach is based on compound similarity with compounds already screened against known targets, meaning higher the similarity, better the prediction. We tested our methods on known drugs for which the target is known. We then applied our software for the identification of nuclear targets involved in antibiotics resistant pathway for the polyketide antibiotics Ivermectin. We successfully predict the existence of such nuclear receptors and provide insights for such biological pathways using PuBChem Bioassay database

Introduction:

Computational biology or bioinformatics is the science of using biological data to develop algorithms and relations among various biological systems. There are large amounts of chemical and biological data available, this is why it is important for a large range of applications to be able to sort and organize these data.

We focused on the ability to find quickly the known targets of a given small molecule (ligand), for application in drug discovery and assessments of toxicity. In this article, we propose an informatic method to automatize target research from a ligand query. Indeed, it is possible to find this information manually by browsing the different databases websites and using their search tools manually but this can be a tedious work due to the great amount and variance of results.

In this review, the investigated approach is interesting to complement phenotypic screening, used in drug discovery, which has been getting a lot of support during the last years [1]. Phenotypic screening can identify active compounds against a disease-related phenotype but does not give any indication of the target or the drug's action mechanism. One of the main challenges of early-stage drug discovery is to find this mechanism of action. Therefore, being able to find quickly the known targets of such compound may help to understand it's way of action.

Experimental target-based bioassay data from public domain database such as PubChem [2] has been used in order to find all target activities of a specific compound. Combining it with similarity search tool from a query molecule allows to broaden research to proteins that are known to be targets to similar compounds.

Potential targets of interest represent all targets that are shown by the bioassays to be

active when confronted with the query and similar compounds.

We aim to test this in silico target finding approach in continuation of a project concerning drug resistance with respect to a natural polyketide, Ivermectin (IVM), a well-known antiparasitic drug a broad spectrum of activity, high efficacy as well as a wide margin of safety [3].

Ivermectin binds selectively and with high affinity to glutamate-gated chloride ion channels in invertebrate muscle and nerve cells of the microfilaria. This binding causes an increase in the permeability of the cell membrane to chloride ions and results in hyperpolarization of the cell, leading to paralysis and death of the parasite. Ivermectin is also known to bind nuclear receptor in human and parasites [4]. Our previous work was to try to determine the targets of IVM, more specifically to understand its binding with nuclear receptors. The purpose of the work presented in this article is to identify active targets for a given compound, taking IVM and its similar compounds as an example.

Materials and Methods:

Our goal is to speed up the search of targets by automatizing it informatically. In order to do this, we created a python code that does all the following steps automatically starting from an input of the user that we will call the query.

As preliminary work, we tried to look for the targets manually, to be able to know the steps we would have to take to create our program. During our research, we noticed that PubChem's search tool's URL addresses were built very specifically. Each URL includes a constant part and a variable one. The variable part defines the kind of search tool that was used and the item ID (compound, bioassay, etc.) used as the query. We took advantage of this to be able to access all the pages automatically in our python program by adding directly the text string corresponding to the search tool we wanted to use, and the ID of our query compound into the URL. We call this method logical URL completion.

The program starts with the input of a small target molecule in SMILE format. SMILE format is the conventional way of writing molecules in bioinformatics. In order to broaden the search and get more potential results, we first use the similarity search.

During our approach to validate the model, we searched IVM similar compounds in PubChem database to find other compounds SMILE similar to the query.

The first thing to do is to find the ID of the compound for which we want to find targets, so we can use it as a query for the search tools. Afterwards, logical URL completion has been utilized, as described above, to access the list of similar compounds using PubChem's similarity search tool. Urllib library has been used to download the data (Annexe).

With a similar approach, we downloaded all the bioassays in PubChem, using the bioassay search tool for each of the similar compounds found with the previous step.

The next step consists in the filtering of all the downloaded bioassays by the python program which keeps in a separate file only the relevant bioassays displaying an activity toward a target, represented by "non NULL" and "active target".

Finally, we got a bioassay file which only contains active targets of IVM and similar compounds.

An interesting part during the research was the origin of the organism. Indeed, the origin of the target's organism might be a criteria for the selection of a drug compound. We may want to find only organism-specific targets, or to keep the broad results. To deal with this parameter, we decided to add it into the bioassays file. It is performed by a simple function that can either take all the results or select them only according to the wanted origin.

After collecting all bioassays of IVM similar compounds which correspond to the

criteria, we searched the gene of each active bioassay. Targets' gene data has been stored, which could help for developing new drugs. This data allow to find orthologous genes to the target's genes. Orthologous genes are genes in different species that evolved from a common ancestral gene by speciation. Normally, orthologs retain the same function in the course of evolution. Identification of orthologs is critical for reliable prediction of gene function in newly sequenced genomes. Finding orthologous genes can help to find uses of drugs known to work in some specific organisms into organisms in which this drug would have a similar target and mechanism of action.

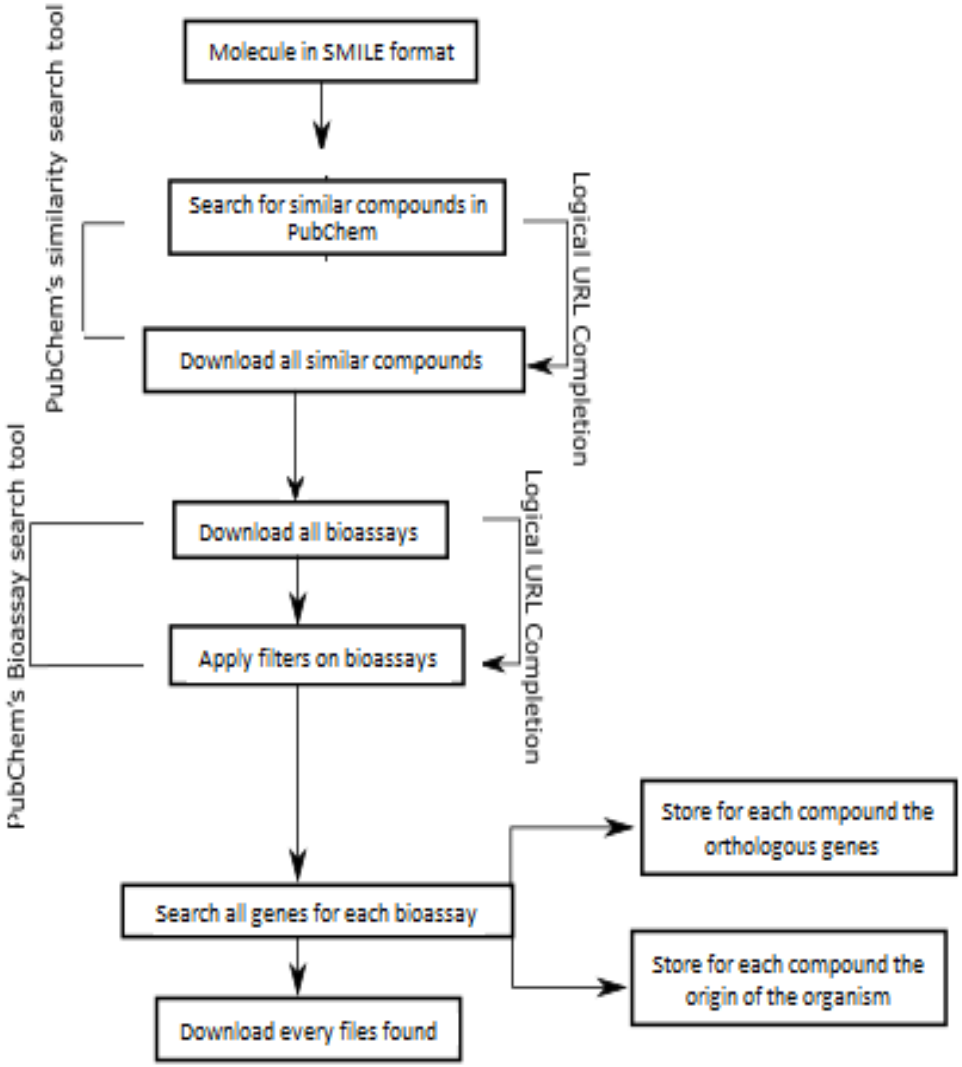


Figure 1: Schematic functions of the script

Moreover, we used a PDB tool to identify targets based on their 3D structure and their documented ligands into the database. To do so we used the XML query the PDB is based on.

The XML query is a formatted in a way to ask the system a specific question. So, for each compound found, a query was formulated in order to look for 3D similar compound with a percentage of similarity that can be set by the user depending on his research. For our tests, we set the similarity at 95%. The output is then the protein-ligand complexes (including these compounds) documented in the PDB. Thus, the output is a list of protein targets of the compounds.

Results:

The program we made has been tested with the IVM molecule as a query.

Once the program has been fed with IVM's SMILE, it returned a folder containing all the similar compounds in PubChem named by their compound ID. Each of these compound folders contain a list of bioassays, each one giving information about one target of that specific compound. Each bioassay gives the following information: activity of the compound, activity value, cid, sid, aid, aid name, target name and target URL.

For each one of these compounds we looked for 3D structures and protein-ligand complexes in the PDB as explained previously.

So, for each compound a list of PDB id was saved along with the Structure Data File of the compound.

It is very interesting to see that by using this platform and this method we find newly studied compound for IVM resistance like the Farnesoid X Receptor [5] (pdb ID: 4WVD).

On addition, all active bioassays from PubChem are stored in a CSV file. Here in this particular example the file is named: all_active_bioassay.csv file (Figure 2 and 4).

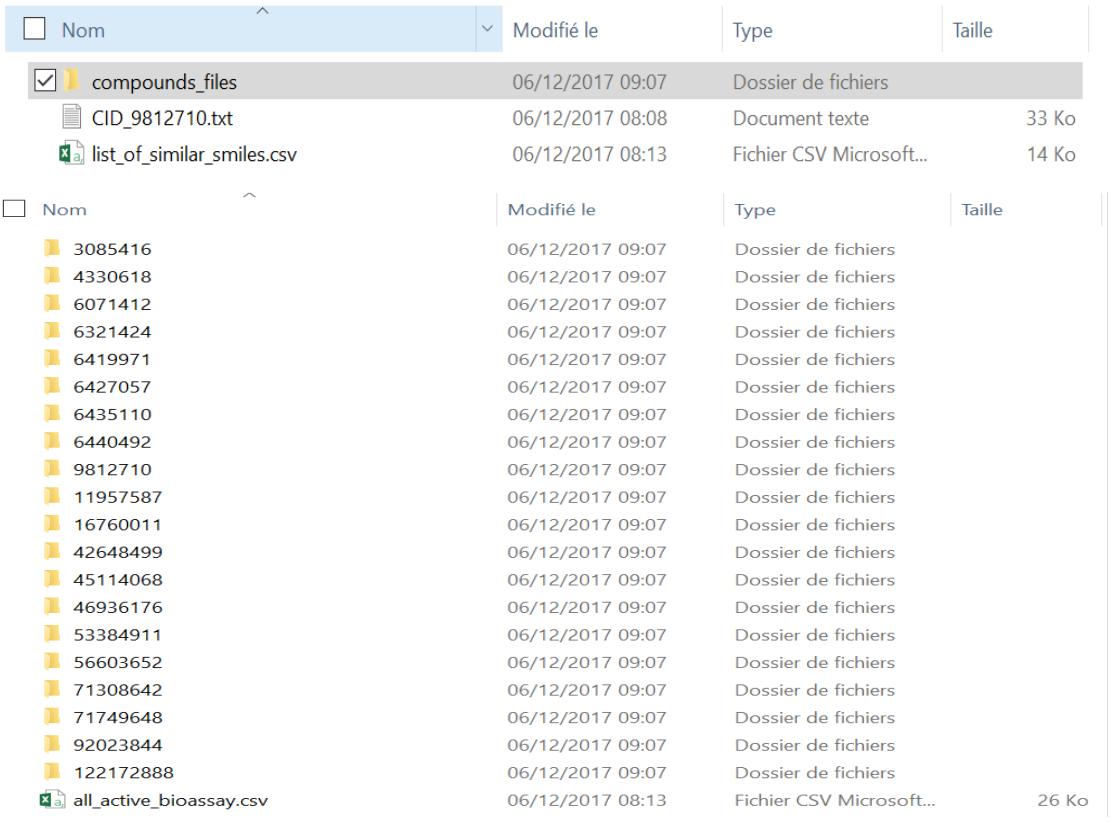


Figure 2: Compounds folders named by compound ID containing bioassays for each of them

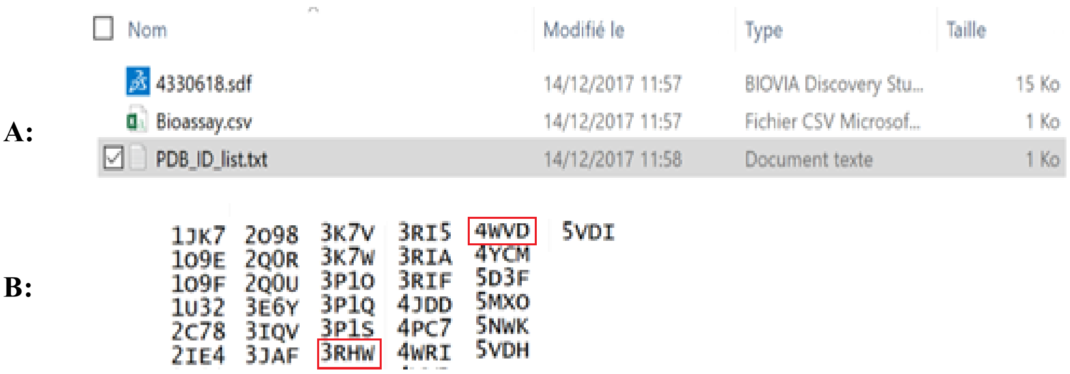
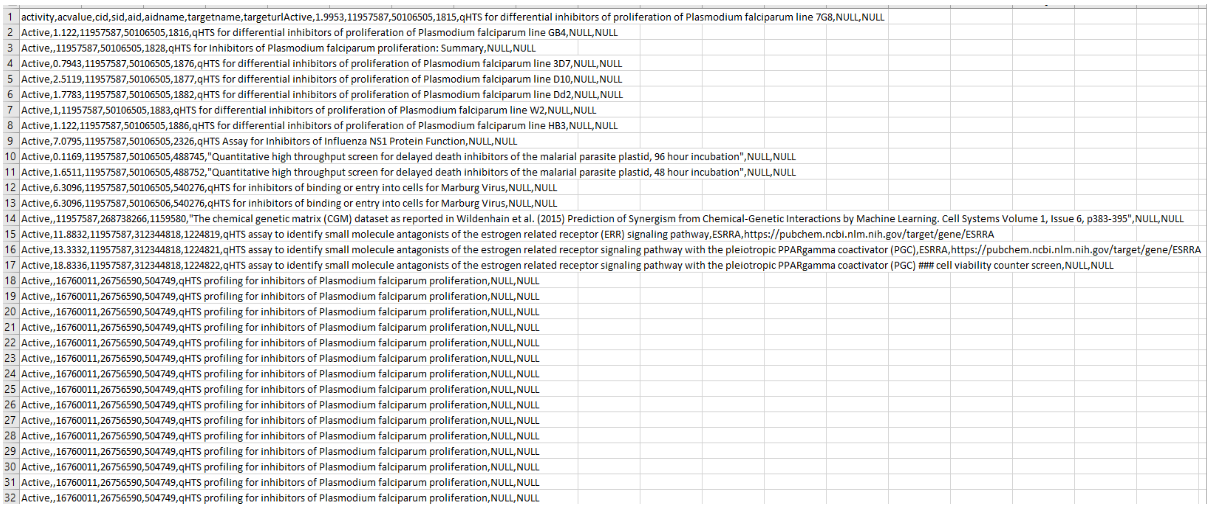


Figure 3: **A:** Example folder containing the files downloaded for one compound. **B:** PDB id of targets of compounds with 95% similarity to IVM (IVM included)



1 activity,acvalue,cid,sid,aid,aidname,targetname,targeturlActive,1.9953,11957587,50106505,1815,qHTS for differential inhibitors of proliferation of Plasmodium falciparum line 7G8,NULL,NULL
2 Active,1.122,11957587,50106505,1816,qHTS for differential inhibitors of proliferation of Plasmodium falciparum line GB4,NULL,NULL
3 Active,,11957587,50106505,1828,qHTS for Inhibitors of Plasmodium falciparum proliferation: Summary,NULL,NULL
4 Active,0.7943,11957587,50106505,1876,qHTS for differential inhibitors of proliferation of Plasmodium falciparum line 3D7,NULL,NULL
5 Active,2.5119,11957587,50106505,1877,qHTS for differential inhibitors of proliferation of Plasmodium falciparum line D10,NULL,NULL
6 Active,1.7783,11957587,50106505,1882,qHTS for differential inhibitors of proliferation of Plasmodium falciparum line Dd2,NULL,NULL
7 Active,1,11957587,50106505,1883,qHTS for differential inhibitors of proliferation of Plasmodium falciparum line W2,NULL,NULL
8 Active,1.122,11957587,50106505,1886,qHTS for differential inhibitors of proliferation of Plasmodium falciparum line HB3,NULL,NULL
9 Active,7.0795,11957587,50106505,2326,qHTS Assay for Inhibitors of Influenza NS1 Protein Function,NULL,NULL
10 Active,0.1169,11957587,50106505,488745,"Quantitative high throughput screen for delayed death inhibitors of the malarial parasite plastid, 96 hour incubation",NULL,NULL
11 Active,1.6511,11957587,50106505,488752,"Quantitative high throughput screen for delayed death inhibitors of the malarial parasite plastid, 48 hour incubation",NULL,NULL
12 Active,6.3096,11957587,50106505,540276,qHTS for inhibitors of binding or entry into cells for Marburg Virus,NULL,NULL
13 Active,6.3096,11957587,50106506,540276,qHTS for inhibitors of binding or entry into cells for Marburg Virus,NULL,NULL
14 Active,,11957587,268738266,1159580,"The chemical genetic matrix (CGM) dataset as reported in Wildenhain et al. (2015) Prediction of Synergism from Chemical-Genetic Interactions by Machine Learning. Cell Systems Volume 1, Issue 6, p383-395",NULL,NULL
15 Active,11.8832,11957587,312344818,1224819,qHTS assay to identify small molecule antagonists of the estrogen related receptor (ERR) signaling pathway,ESRRA,https://pubchem.ncbi.nlm.nih.gov/target/gene/ESRRA
16 Active,13.3332,11957587,312344818,1224821,qHTS assay to identify small molecule antagonists of the estrogen related receptor signaling pathway with the pleiotropic PPARgamma coactivator (PGC),ESRRA,https://pubchem.ncbi.nlm.nih.gov/target/gene/ESRRA
17 Active,18.8336,11957587,312344818,1224822,qHTS assay to identify small molecule antagonists of the estrogen related receptor signaling pathway with the pleiotropic PPARgamma coactivator (PGC) ### cell viability counter screen,NULL,NULL
18 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
19 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
20 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
21 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
22 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
23 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
24 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
25 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
26 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
27 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
28 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
29 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
30 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
31 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL
32 Active,,16760011,26756590,504749,qHTS profiling for inhibitors of Plasmodium falciparum proliferation,NULL,NULL

Figure 4: Extract of CSV file containing active bioassays

Figure 5: Orthologous genes folder

As represented in Figure 3, where only relevant bioassays were kept, bioassays results can be filtered and sorted. We were interested in finding targets, so we filtered bioassays and only kept the one showing the activity of the compound toward a target. In order to select relevant active bioassays, we only keep bioassay for which target activities is written as "active" and target name as "non null".

Orthologous genes are also obtained, as explained previously, in a folder containing every orthologous genes found during the computation.

Discussion:

We finally tested our program with IVM as a query. The results we expected were forecasted by manual research on databases and background from our previous work done last year about the targets of IVM.

The manually searched IVM similar compounds that we expected were present in our program's output among other results. Thus, our results are in adequation with our expectations.

We downloaded the bioassays of all IVM similar compounds for the selected criteria. We decided to sort and select all active and non-null targets manually and we put all these data in a file. Again, those results were confirmed by our program which selected the same non-null active targets. We were able to find in these results the FXR receptor (PDB id: 4WVD) [6] and *C. elegans* glutamate-gated chloride channel (PDB id: 3RHW) [7], or in other words the targets we worked on during our previous work about Ivermectin. Thus, once again, our expectations are in adequation with our results.

We didn't have any background on orthologous gene, so we only checked our results versus a quick manual search. Results appeared to be as expected, but further confrontation might be done in the future. In fact, it often gives more results than those obtained manually because as humans, we are not as methodical as this tool. This is also a reason why we must ensure that all our data is well stored and ordered as it should remain easy for the user to interpret the results and navigate in the files.

The results we presented were expected and were to be expected as our automatized tool only searches for known data in databases. We only fetch existing and already known data, so there was little surprise to have in the outcome of our program considering it working as intended. This tool now still allows us to quickly fetch and find among a tremendous amount of data every known targets to any drug, and compounds similar to it, and quickly find about orthologous genes to deduce similar targets.

During this work we focused on the results present in PubChem's bioassay library but also implemented a path to search into the Protein Data Bank as well. We can then manually confront the results coming from each platform. In the future, we intend to improve our tool to make our extracted data more homogeneous in order to fuse all the results and automatize a comparison so that results are flagged depending whether they come from either or both the PubChem and the PDB. Our expected final tool would then in a very simplified way look like figure 6.

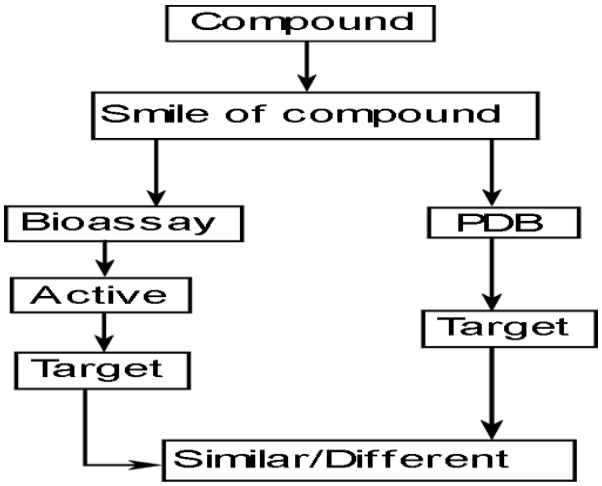


Figure 6: Simplified schematic of the program

Bibliography:

[1] Wagner B. “The resurgence of phenotypic screening in drug discovery and development”, Expert Opinion on Drug Discovery, Volume 11, 2016 - Issue 2

[2] Kim S, Thiessen PA, Bolton EE, Chen J, Fu G, Gindulyte A, Han L, He J, He S, Shoemaker BA, Wang J, Yu B, Zhang J, Bryant SH. PubChem Substance and Compound databases. Nucleic Acids Research. 2016 Jan 4; 44(D1):D1202-13. Epub 2015 Sep 22 [PubMed PMID: 26400175] doi: 10.1093/nar/gkv951.

[3] Canga A.G., Prieto A.M.S., Liébana M.J.D., Martínez N.F., Vega M.S., Vieitez J.J.G. The pharmacokinetics and interactions of ivermectin in humans—A mini-review. AAPS J. 2008;10:42–46. doi: 10.1208/s12248-007-9000-9.

[4] "Ivermectin". The American Society of Health-System Pharmacists. https://www.drugs.com/monograph/ivermectin.html

[5] Hsu C-W, Hsieh J-H, Huang R, et al., “Differential modulation of FXR activity by chlorophacinone and ivermectin analogs.” Toxicology and applied pharmacology. 2016;313:138-148. doi:10.1016/j.taap.2016.10.017.

[6] Jin L, Feng X, Rong H, Pan Z, Inaba Y, Qiu L, Zheng W, Lin S, Wang R, Wang Z, Wang S, Liu H, Li S, Xie W, Li Y., “The antiparasitic drug ivermectin is a novel FXR ligand that regulates metabolism.” Nature Communications. 2013;4:1937

[7] Hibbs RE, Gouaux E. “Principles of activation and permeation in an anion-selective Cys-loop receptor.” Nature. 2011;474(7349):54-60.

Annexes:

pubchempy: "*PubChemPy provides a way to interact with PubChem in Python. It allows chemical searches by name, substructure and similarity, chemical standardization, conversion between chemical file formats, depiction and retrieval of chemical properties.*" *https://media.readthedocs.org/pdf/pubchempy/latest/pubchempy.pdf*

xml.etree.ElementTree: "*XML is an inherently hierarchical data format, and the most natural way to represent it is with a tree. ET has two classes for this purpose - ElementTree represents the whole XML document as a tree, and Element represents a single node in this tree. Interactions with the whole document (reading and writing to/from files) are usually done on the ElementTree level. Interactions with a single XML element and its sub-elements are done on the Element level.*" *https://docs.python.org/3/library/xml.etree.elementtree.html*

parse xml element:*"Loads an external XML section into this element tree. source is a file name or file object. parser is an optional parser instance. If not given, the standard XMLParserparser is used. Returns the section root element." https://docs.python.org/3/library/xml.etree.elementtree.html*

urllib: "This module provides a high-level interface for fetching data across the World Wide Web. In particular, the **urlopen()** function is similar to the built-in function **open()**, but accepts Universal Resource Locators (URLs) instead of filenames. Some restrictions apply — it can only open URLs for reading, and no seek operations are available." https://docs.python.org/2/library/urllib.html

easygui: *"EasyGUI is a module for very simple, very easy GUI programming in Python. EasyGUI is different from other GUI generators in that EasyGUI is NOT event-driven. Instead, all GUI interactions are invoked by simple function calls.*
EasyGui provides an easy-to-use interface for simple GUI interaction with a user. It does not require the programmer to know anything about tkinter, frames, widgets, callbacks or lambda." http://easygui.sourceforge.net/